Part 1: quick summary

I'm not a statistics expert, so *caveat emptor*. If you spot any mistakes or have suggestions to make this document more useful, please let me know (at *rudolf@pobox.com*). Thanks to Mike Aitken (MRFA) for helpful comments!

1.1 Overview of this document

- First, in Part 1, we'll summarize what most people want to know to get going — how to choose and perform an ANOVA. Nobody reads ANOVA theory before starting to analyse, much as statisticians may complain about this, so we might as well be pragmatic. This can be combined with Part 3, which talks about common things that are required in ANOVA analysis, and Part 5, which shows how to perform an ANOVA in SPSS.
- Then, in Part 2, we'll cover what ANOVA does and what it assumes — things people should have known before running an ANOVA but probably didn't.
- In Part 3, we'll walk through what most people need to do to complete an ANOVA analysis.
- In Part 4, we'll look at experimental design and analysis issues, such as how to analyse changes from baseline, and when and how to perform *post hoc* tests.
- In Part 5, we'll look at how to use SPSS to perform different ANOVAs.
- In Part 6, we'll cover complex theory that most people will never need.
- In Part 7, we'll look at a variety of ANOVA models that can be used for different experimental designs. These will range from the very simple (one-way ANOVA) through the very useful (mixed designs with both between- and within-subject factors) to the very complicated. This material is for *reference*.
- In Part 8, we'll revise mathematics that is touched on occasionally elsewhere, and cover very advanced mathematics that underpins computer calculations of complex ANOVAs.
- In Part 9, there's a glossary.

1.2 Background knowledge

This handout is aimed at graduate students who need to perform analysis of variance (ANOVA). Covering the theory of ANOVA is one thing; putting it into practice in psychology and neuroscience research unfortunately means using the technique at a level at which even statisticians debate the proper methods. This is depressing to the beginner; I hope this handout helps. It's also a reminder to me of information I've collected about different ANOVA designs. It covers simple ANOVA and also some complex techniques that are not often used but rather powerful. It assumes a basic knowledge of statistics. Explicit coverage of the background knowledge can be found in my NST IB Psychology handouts, available at

www.pobox.com / ~rudolf / psychology

and coverage of exploratory data analysis (EDA) and ANOVA can be found in Mike Aitken's NST II Psychology handouts, available at

foxfield.psychol.cam.ac.uk / stats / default.html

1.3 Quick summary: choosing and performing an ANOVA

We'll presume your experiment was sensibly designed and free of confounds. No amount of analysis will fix a bad design. Now, **the purpose of ANOVA is to predict a single dependent variable on the basis of one or more predictor variables,**

and to establish whether those predictors are good predictors or not. Therefore you need to do the following:

- Identify your dependent variable.
- Identify your predictor variables.
- Establish whether your predictor variables are **discrete** (e.g. sham / lesion, sham / core / shell, day 1 / 2 / 3) or **continuous** (e.g. body mass). We will call discrete variables **factors,** and continuous variables **covariates.** The number of discrete values that a factor can take is known as the number of **levels** of that factor.
- For most psychology designs, the key unit of 'relatedness' is usually the subject. It then suffices to establish whether your predictor variables are **between-subjects** variables (e.g. operative group; every subject is only measured at one level of the factor, such as 'lesion') or **within-subjects** variables (e.g. test day; each subject is measured at more than one level of the factor, such as 'day 1', 'day 2', and so on).
- You should now be able to identify your design (e.g. 'one between-subjects factor and two within-subjects factors') using this document. The sections giving detail on each design also give the **SPSS syntax.**
- You should check that the assumptions of ANOVA are met — for example, do you need to **transform** your dependent variable (by taking the square root, arcsine, logarithm, etc.) before analysis?
- Run the ANOVA.
- If your ANOVA has within-subjects factors, check **Mauchly's test of sphericity,** which your software should have done for you. If the Mauchly test is 'significant' (small *p* value), one of the assumptions behind within-subjects ANOVA has been violated. Don't use the normal *df*; use the corrected *df* — either with the Greenhouse–Geisser (conservative) or Huynh–Feldt (better; Myers & Well, 1995, p. 248) correction. Your software should provide both.
- Interpret the results. You may need to perform **further analyses** *post hoc* to explain main effects or interactions that you find.

I use a notation for describing ANOVA models in which factors are written with their number of levels as a subscript, covariates are written with 'cov' as a subscript, S denotes subjects, factors / covariates in brackets with 'S' are within-subjects predictors, and unbracketed factors / covariates are between-subjects predictors. An ANOVA with one between-subjects factor (A) and two within-subjects factors (U, V) might be written like this:

$$\text{dependent variable} = A \times (U \times V \times S)$$

As a more concrete example of this notation, suppose you measured locomotor activity (dependent variable) in two groups of rats (sham / lesion). Each rat was tested on six occasions: following one of three drug treatments (saline / low-dose cocaine / high-dose cocaine), and in one of two rooms (hot / cold). We assume the testing order for within-subjects factors was appropriately counterbalanced to avoid order effects (see handouts at www.pobox.com / ~rudolf / psychology). We could write this design as:

$$\text{locomotor activity} = \text{Group}_2 \times (\text{Drug}_3 \times \text{Room}_2 \times S)$$

In this document, I will try to use A, B, C… as labels for between-subjects factors and U, V, W… as labels for within-subjects factors, since it gets hard to read otherwise when there are both between- and within-subjects factors in a design.

Designs with both between-subjects and within-subjects factors are called 'mixed' or 'nested' designs (Keppel, 1991, p. 563): variability due to subjects is 'nested' within variability due to the between-subjects factor(s), because each subject is only tested at one level of the between-subjects factor(s).

If you have units of relatedness other than ‘subject’ (e.g. ‘plot of land’), but you only have one level of relatedness, you can merely think of your design in the same between- / within-subject terms.

If you have multiple levels of relatedness, you will need a complex or **hierarchical** design (Myers & Well, 1995, chapter 10); you should aim to understand the principles behind the designs discussed in this document. At the end we’ll cover some hierarchical designs, but this is hard stuff.

Part 2: understanding the basics of ANOVA

2.1 The basic logic and assumptions of ANOVA

2.1.1 The underlying model

After Howell (1997, ch. 11). Suppose that the average height of UK adults is 175 cm, that of adult females is 170 cm, and that of adult males is 180 cm. So 'maleness' contributes, on average, +5 cm to an adult's height (compared to the mean of all adults), and 'femaleness' contributes, on average, –5 cm. Suppose we take a given adult male. We could break his height down into three components: 175 cm for being an adult, 5 cm for being a male, and some other component that represents this individual's 'uniqueness', since there is of course variation in the heights of adult men. We could write this model as

$$\text{height} = 175 \text{ cm} + 5 \text{ cm} + \text{uniqueness}$$

or in more general terms

$$\text{height}_{\text{individual male}} = \mu + \tau_{\text{male}} + \varepsilon_{\text{individual}}$$

where μ is the overall mean (175 cm), τ_{male} is the contribution for being a male, and $\varepsilon_{\text{individual}}$ is a particular individual's unique contribution. We have written an expression for our dependent variable (height) in terms of predictor variables (the grand mean and a *factor*, sex) and unpredicted variability. Let's extend that principle.

2.1.2 An example: data and a structural model

Suppose 50 subjects are assigned to five groups. Each group reads a list of words in a different way: one was asked to count the number of letters in each word, one to think of a rhyme for each word, one to give an adjective that could be used with each word, one to form a vivid image of each word, and one to memorize each word for later recall. Later, all groups were asked to recall all the words they could remember. In ANOVA terminology, we have a single factor Group with five levels (Group_1, Group_2, … Group_5). Here are some results (Howell, 1997, p. 301):

No. words recalled	Group_1 Counting	Group_2 Rhyming	Group_3 Adjective	Group_4 Imagery	Group_5 Memorize	Total
One	9	7	11	12	10	
number,	8	9	13	11	19	
one	6	6	8	16	14	
subject	8	6	6	11	5	
	10	6	14	9	10	
	4	11	11	23	11	
	6	6	13	12	14	
	5	3	13	10	15	
	7	8	10	19	11	
	7	7	11	11	11	
total	$T_1 = 70$	$T_2 = 69$	$T_3 = 110$	$T_4 = 134$	$T_5 = 120$	$\sum X = 503$
n	$n_1 = 10$	$n_2 = 10$	$n_3 = 10$	$n_4 = 10$	$n_5 = 10$	$N = 50$
mean	$\bar{x}_1 = 7$	$\bar{x}_2 = 6.9$	$\bar{x}_3 = 11$	$\bar{x}_4 = 13.4$	$\bar{x}_5 = 12$	$\bar{x} = 10.06$
SD	$s_1 = 1.83$	$s_2 = 2.13$	$s_3 = 2.49$	$s_4 = 4.50$	$s_5 = 3.74$	4.01
variance	$s_1^2 = 3.33$	$s_2^2 = 4.54$	$s_3^2 = 6.22$	$s_4^2 = 20.27$	$s_5^2 = 14$	$s^2 = 16.06$

For this data, we can specify a model, just as we did before. Let

- X_{ij} represent the score of person j in condition (group) i
- μ represent the overall mean score
- μ_i represent the mean of scores in condition i
- τ_i represent the degree to which the mean of condition i deviates from the overall mean (the contribution of condition i), i.e. $\tau_i = \mu_i - \mu$

- ε_{ij} represent the amount by which person *j* in condition *i* deviates from the mean of his or her group (the 'uniqueness' of person *j* in condition *i*), i.e. $\varepsilon_{ij} = X_{ij} - \mu_i$

Since it's obvious that

$$X_{ij} = \mu + (\mu_i - \mu) + (X_{ij} - \mu_i)$$

it follows that

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

2.1.3 The null hypothesis

We will test the null hypothesis that there is no difference between the various groups (conditions). We can state that null hypothesis like this:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu$$

In other words, the null hypothesis is that all means are equal to each other and to the grand mean (μ), and that all treatment (group) effects are zero.

2.1.4 The assumptions of ANOVA

If μ_i represents the population mean of condition *i* and σ_i^2 represents the population variance of this condition, analysis of variance is based on certain assumptions about these population parameters.

1. Homogeneity of variance

We assume that each of our populations has the same variance:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_e^2$$

The term σ_e^2 (where *e* stands for error) represents the **error variance** — the variance unrelated to any treatment (condition) differences. We would expect homogeneity of variance if the effect of any treatment is to add or subtract a constant to everyone's score — without a treatment the variance would be σ_e^2, and if you add a constant to a variable, the variance of that variable doesn't change.

2. Normality

We assume that the scores for each condition are normally distributed around the mean for that condition. (Since ε_{ij} represents the variability of each person's score around the mean of that condition, this assumption is the same as saying that error is normally distributed within each condition — sometimes referred to as the assumption of the 'normal distribution of error'.)

3. Independence of error components (≈ independence of observations)

We also assume that the observations are independent — technically, that the error components (e_{ij}) are independent. For any two observations within an experimental treatment, we assume that knowing how one of these observations stands relative to the treatment (or population) mean tells us nothing about the other observation. Random assignment of subjects to groups is an important way of achieving this. To deal with observations that are *not* independent — for example, observations that are correlated because they come from the same subjects — we need to account specifically for the sources of 'relatedness' to make sure that the residual error components

are independent; this is why we need within-subjects (repeated measures) designs for this sort of situation. But we'll ignore that for the moment.

2.1.5 The logic of ANOVA

Since we have assumed that the distribution of the scores for each condition have the same shape (are normally distributed) and have the same variance (homogeneity of variance), they can only differ in their means. Now if we measure the variance of any *one* condition, such as s_1^2, that variance will be an estimate of the common population variance σ_e^2 (remember, we assumed $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_e^2$, that is, homogeneity of variance). In each case, our sample variance estimates a population variance:

$$\sigma_1^2 \doteq s_1^2;\ \sigma_2^2 \doteq s_2^2;\ \ldots\ \sigma_5^2 \doteq s_5^2$$

(where $\doteq$ denotes 'is estimated by'). Because of our homogeneity of variance assumption, each of these sample variances is an estimate of σ_e^2:

$$\sigma_e^2 \doteq s_1^2;\ \sigma_e^2 \doteq s_2^2;\ \ldots\ \sigma_e^2 \doteq s_5^2$$

To improve our estimate of σ_e^2, we can **pool** the five estimates by taking their mean (if $n_1 = n_2 = n_3 = n_4 = n_5 = n$), and thus

$$\sigma_e^2 \doteq s^2 = \bar{s}_i^2 = \frac{\sum s_i^2}{a}$$

where a is the number of treatments — in this case, 5. (If the sample sizes were not equal, we would still average the five estimates, but we would weight them by the number of degrees of freedom for each sample, so variance estimates from larger samples would get more weight.) This gives us an estimate of the population variance that is referred to as $\mathbf{MS_{error}}$ ('mean square error'), sometimes called $\mathbf{MS_{within}}$, or $\mathbf{MS_{subjects\ within\ groups}}$, or $\mathbf{MS_{S/groups}}$ ('mean square for subjects within groups'). This is true **regardless of whether H_0 is true or false.** For the example above, our pooled estimate of σ_e^2 will be

$$\sigma_e^2 = \frac{3.33 + 4.54 + 6.22 + 20.27 + 14.00}{5} = 9.67$$

Now let us **assume that H_0 is true.** In this case, our five samples of 10 cases may be thought of as five independent samples from the same population (or, equivalently, five samples from five identical populations). The Central Limit Theorem (see handouts at www.pobox.com/~rudolf/psychology) states that the variance of means drawn from the same population is equal to the variance of the population divided by the sample size. If H_0 is true, therefore, the variance of our five sample means estimates σ_e^2 / n:

$$\frac{\sigma_e^2}{n} \doteq s_{\bar{x}}^2$$

and so

$$\sigma_e^2 \doteq n s_{\bar{x}}^2$$

This is therefore a second estimate of σ_e^2, referred to as $\mathbf{MS_{treatment}}$ or $\mathbf{MS_{group}}$. On the other hand, **if H_0 is false,** this will *not* be a good estimate of σ_e^2. So we have found that MS_{error} estimates σ_e^2 whether H_0 is true or false, but $MS_{treatment}$ only estimates σ_e^2 if H_0 is true. Therefore, if our two estimates of σ_e^2, $MS_{treatment}$ and MS_{error}, are similar, this is evidence that H_0 is true; if they are very different, this is evidence

that H_0 is false. We will compare the two variance estimates with an ***F* test,** which is designed specifically for comparing two variances (see handouts at www.pobox.com/~rudolf/psychology): $F = \mathrm{MS}_{\text{treatment}} / \mathrm{MS}_{\text{error}}$. If our F statistic is very different from 1, we will reject the null hypothesis that the two variances ($\mathrm{MS}_{\text{treatment}}$ and $\mathrm{MS}_{\text{error}}$) are the same, and hence reject the null hypothesis of the ANOVA.

2.1.6 Expected mean squares (EMS)

Let's formalize that. We've defined the **treatment effect** τ_i as $\mu_i - \mu$, the difference between the mean of treatment i (μ_i) and the grand mean (μ). We will also define σ_τ^2 as the variance of the true population's means ($\mu_1, \mu_2, \ldots \mu_a$):

$$\sigma_\tau^2 = \frac{\sum(\mu_i - \mu)}{a-1} = \frac{\sum \tau_i^2}{a-1}$$

Technically, this is not actually the variance — since we are working with parameters, not statistics, we should have divided by a rather than $a - 1$ if we wanted the variance. However, we can think of it as a variance without much problem.

We can then define, without proof, the **expected value** of the mean square terms:

$$E(\mathrm{MS}_{\text{error}}) = \sigma_e^2$$
$$E(\mathrm{MS}_{\text{treatment}}) = \sigma_e^2 + \frac{n\sum \tau_i^2}{a-1} = \sigma_e^2 + n\sigma_\tau^2$$

where σ_e^2 is the variance within each population and σ_τ^2 is the variance of the population means (μ_j). So if H_0 is true, $\sigma_\tau^2 = 0$, so $E(\mathrm{MS}_{\text{treatment}}) = E(\mathrm{MS}_{\text{error}})$, but if H_0 is false, $E(\mathrm{MS}_{\text{treatment}}) > E(\mathrm{MS}_{\text{error}})$.

2.2 The calculations behind a simple one-way ANOVA (one between-subjects factor)

Let's go back to the results in the table we saw earlier and conduct an ANOVA.

2.2.1 Calculations with means (conceptual) or totals (for manual calculation only)

Most ANOVA calculations are based on **sums of squares.** Remember that a variance is a sum of squared deviations from the mean (a 'sum of squares') divided by the number of degrees of freedom. We work with sums of squares because they are *additive*, whereas mean squares and variances are only additive if they are based on the same number of degrees of freedom.

Purely for convenience, Howell (1997) tends to do the calculations in terms of treatment **totals** rather than treatment means. In the table above, we have defined T_i as the total for treatment i. Totals are linearly related to means ($T = n\bar{x}$). If you multiple a variable by a constant, you multiply the variance of that variable by the square of the constant. So since $T = n\bar{x}$, we can see that

$$s_T^2 = n^2 s_{\bar{x}}^2$$
$$s_{\bar{x}}^2 = \frac{s_T^2}{n^2}$$

We saw earlier that if H_0 is true, $\sigma_e^2 \doteq n s_{\bar{x}}^2$; therefore, if H_0 is true,

$$\sigma_e^2 \doteq \frac{s_T^2}{n}$$

On the other hand, though calculating sums of squares may be easier in terms of treatment totals, **conceptually** it is *much* easier to think in terms of means. We'll present both for a while — first the definition in terms of means, and then, in brackets, the formula in terms of totals. **Ignore what's in brackets** unless you're doing the calculations by hand. Eventually we'll just show the calculations in terms of means. After all, you'll be using a computer for the hard work.

2.2.2 Calculating SS_{total}, $SS_{treatment}$, and SS_{error}

First, we calculate $\mathbf{SS_{total}}$ ('total sum of squares') — the sum of squares of all the observations (the *summed squared deviations of each observation from the overall mean*), regardless of which treatment group the observations came from.

$$\mathrm{SS}_{\text{total}} = \sum (x - \bar{x})^2 \left[= \sum x^2 - \frac{(\sum x)^2}{N} \right]$$

Now we calculate $\mathbf{SS_{treatment}}$. This represents the *summed squared deviations of the treatment mean from the mean of all treatment means, summed over each data point*. (Or, in terms of totals, the summed squared deviations of each total [T_j] from the mean of the treatment totals [$\overline{T}$], all divided by the number of observations per total.)

$$\mathrm{SS}_{\text{treatment}} = \sum n(\bar{x}_i - \bar{x})^2 \left[\begin{array}{l} = \sum n\left(\frac{T_i}{n} - \frac{\overline{T}}{n}\right)^2 = \sum \frac{n}{n^2}\left(T_i - \overline{T}\right)^2 \\ = \frac{\sum (T_i - \overline{T})^2}{n} \\ = \frac{\sum T_i^2 - \frac{(\sum T_i)^2}{a}}{n} = \frac{\sum T_i^2}{n} - \frac{(\sum T_i)^2}{na} \\ = \frac{\sum T_i^2}{n} - \frac{(\sum x)^2}{N} \end{array} \right]$$

where *a* is the number of treatments, *n* is the number of observations per treatment, and *N* is the total number of observations (= *na*).

Now we can calculate $\mathbf{SS_{error}}$. This represents the sum of the squared deviations of each point from its group mean. Since $SS_{total} = SS_{treatment} + SS_{error}$, the quick way to obtain SS_{error} is by subtraction:

$$\mathrm{SS}_{\text{error}} = \sum (x - \bar{x}_i)^2 = \mathrm{SS}_{\text{total}} - \mathrm{SS}_{\text{treatment}}$$

Alternatively, we could have calculated SS_{error} by working out an SS for each group separately and adding them up:

$$\begin{aligned} \mathrm{SS}_{\text{group 1}} &= \sum (x_1 - \bar{x}_1)^2 = (9-7)^2 + (8-7)^2 + \ldots + (7-7)^2 \\ \mathrm{SS}_{\text{group 2}} &= \sum (x_2 - \bar{x}_2)^2 = (7-6.9)^2 + (9-6.9)^2 + \ldots + (7-6.9)^2 \\ &\ldots \\ \mathrm{SS}_{\text{error}} &= \mathrm{SS}_{\text{group 1}} + \mathrm{SS}_{\text{group 1}} + \ldots + \mathrm{SS}_{\text{group 5}} \end{aligned}$$

Both approaches give the same answer.

2.2.3 Degrees of freedom

If there are N observations in total, $\boldsymbol{df_{total} = N - 1}$**.** If there are a treatments, $\boldsymbol{df_{treatment} = a - 1}$**.** We can calculate the degrees of freedom for error like this:

$$\boldsymbol{df_{error} = df_{total} - df_{treatment}}$$

Alternatively, we could calculate df_{error} as the sum of the degrees of freedom within each treatment; if there are n observations in each of a treatments, there are $n - 1$ degrees of freedom within each treatment, and so $df_{error} = a(n - 1)$. This gives the same answer (since $df_{total} - df_{treatment} = [N - 1] - [a - 1] = [na - 1] - [a - 1] = na - a = a[n - 1]$).

2.2.4 Mean squares

Mean squares are easy; just divide each SS by the corresponding number of *df*.

2.2.5 The F test

From the definitions of EMS above,

$$\frac{E(\mathrm{MS}_{\text{treatment}})}{E(\mathrm{MS}_{\text{error}})} = \frac{\sigma_e^2 + n\sigma_\tau^2}{\sigma_e^2}$$

We can therefore calculate an F statistic

$$F = \frac{\mathrm{MS}_{\text{treatment}}}{\mathrm{MS}_{\text{error}}}$$

and it is distributed as $F_{a-1,\ a(n-1)}$ — that is, as $F_{treatment\ df,\ error\ df}$, under the null hypothesis. So we can look up critical value of F in tables. If it's 'significant' (unlikely given the null hypothesis), we reject the null hypothesis and say that the treatment did influence our dependent variable.

A very complicated aside: if H_0 is true and $\sigma_\tau^2 = 0$, although $\frac{E(\mathrm{MS}_{\text{treatment}})}{E(\mathrm{MS}_{\text{error}})} = \frac{\sigma_e^2 + n\sigma_\tau^2}{\sigma_e^2}$ and therefore under the null hypothesis $\frac{E(\mathrm{MS}_{\text{treatment}})}{E(\mathrm{MS}_{\text{error}})} = 1$, and so you'd think $E\left(\frac{\mathrm{MS}_{\text{treatment}}}{\mathrm{MS}_{\text{error}}}\right) = 1$, the expected value of F under the null hypothesis is actually $E(F) = \frac{df_{\text{error}}}{df_{\text{error}} - 2}$ (Frank & Althoen, 1994, pp. 470, 513). I don't fully understand that; I suspected that the difference was that $\frac{E(\mathrm{MS}_{\text{treatment}})}{E(\mathrm{MS}_{\text{error}})} \neq E\left(\frac{\mathrm{MS}_{\text{treatment}}}{\mathrm{MS}_{\text{error}}}\right)$ because $E(XY) = E(X)E(Y)$ only if X and Y are independently distributed. MRFA has since pointed out the real reason: under the null hypothesis, $\mathrm{MS}_{\text{error}}$ is asymmetrically distributed. For asymmetric distributions, $E(1/X) \neq 1/E(X)$, so $E(1/\mathrm{MS}_{\text{error}}) \neq 1/E(\mathrm{MS}_{\text{error}})$. It's akin to the reasoning behind using a t test rather than a Z test when you estimate the population standard deviation σ using the sample standard deviation s: even though $E(s) = E(\sigma)$, $E(1/s) \neq E(1/\sigma)$.

2.2.6 ANOVA summary table

ANOVA results are presented in a summary table like this:

Source	d.f.	SS	MS	*F*
Treatment	a–1	$SS_{treatment}$	$SS_{treatment} / df_{treatment}$	$MS_{treatment} / MS_{error}$
Error (S / treatments)	$a(n$–1)	SS_{error}	SS_{error} / df_{error}	
Total	N–1 = an – 1	SS_{total}	SS_{total} / df_{total} [= s^2]	

Remember that 'S / treatments' denotes 'subjects within treatments'; this is the source of all our error variability in this example. Anyway, for our example, we can now calculate all the SS:

$$SS_{total} = \sum (x - \bar{x})^2 = \sum x^2 - \frac{(\sum x)^2}{N} = (9^2 + 8^2 + \ldots + 11^2) - \frac{503^2}{50} = 786.82$$

$$SS_{treatment} = \sum n(\bar{x}_i - \bar{x})^2 = \frac{\sum T_i^2}{n} - \frac{(\sum x)^2}{N} = \frac{(70^2 + 69^2 + \ldots + 120^2)}{10} - \frac{503^2}{50} = 351.52$$

$$SS_{error} = SS_{total} - SS_{treatment} = 435.30$$

so our ANOVA table looks like this:

Source	d.f.	SS	MS	*F*
Treatment	4	351.52	87.88	9.09
Error (S / treatments)	45	435.30	9.67	
Total	49	786.82	16.06	

Our *F* has (4, 45) degrees of freedom. We could write $F_{4,45} = 9.09$, and look this up to find an associated *p* value ($p = 0.00002$).

2.2.7 $SS_{treatment}$ for unequal sample sizes

What if our group sizes were not equal? Previously we had defined

$$SS_{treatment} = \sum n(\bar{x}_i - \bar{x})^2$$

$$\left[= \frac{\sum n^2(\bar{x}_i - \bar{x})^2}{n} = \frac{\sum (n\bar{x}_i - n\bar{x})^2}{n} = \frac{\sum (T_i - \bar{T})^2}{n} \right.$$

$$= \frac{\sum T_i^2 - \frac{(\sum T_i)^2}{a}}{n} = \frac{\sum T_i^2}{n} - \frac{(\sum T_i)^2}{na}$$

$$\left. = \frac{\sum T_i^2}{n} - \frac{(\sum x)^2}{N} \right]$$

which applies when all groups have *n* observations. If the group sizes are unequal, we simply multiply the deviation of each score from its treatment mean by the number of scores in that treatment group (so the larger one sample is, the more it contributes to $SS_{treatment}$):

$$SS_{treatment} = \sum n_i(\bar{x}_i - \bar{x})^2$$

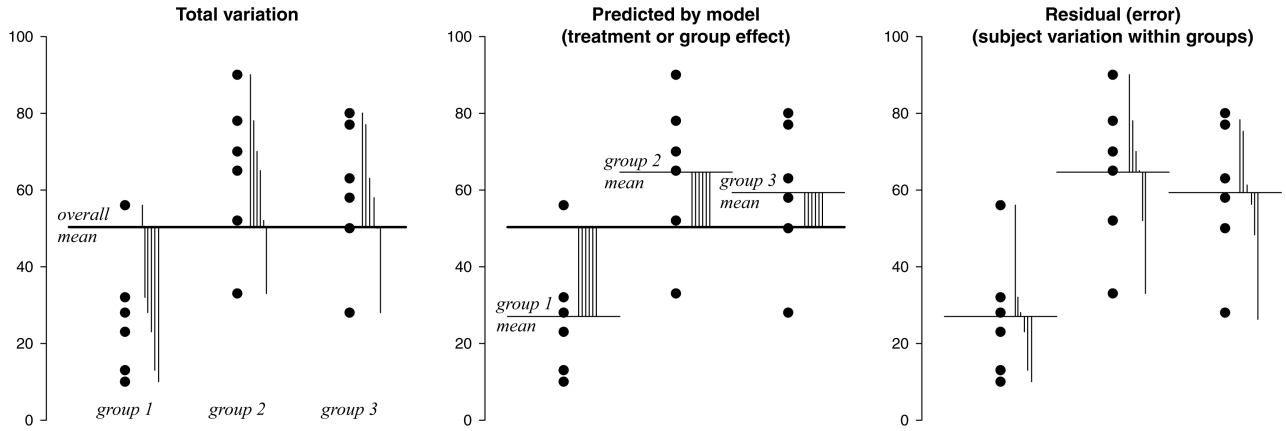
$$\left[= \frac{\sum n_i^2(\bar{x}_i - \bar{x})^2}{n_i} = \frac{\sum (n_i\bar{x}_i - n_i\bar{x})^2}{n_i} = \frac{\sum (T_i - \bar{T})^2}{n_i} \right.$$

$$\left. = \frac{\sum T_i^2}{n_i} - \frac{(\sum x)^2}{N} \right]$$

2.2.8 Pictorial representation

What the ANOVA technique has done is to partition *total variation* from the overall mean (SS_{total}) into variation from the overall mean accounted for or predicted by the treatment or group difference ($SS_{treatment}$ or SS_{groups}) and further variation within the

groups due to inter-subject variability (SS_{error} or $SS_{S/groups}$). If the variation attributable to the model is large, compared to the error variability, we will reject the null hypothesis.



$$SS_{total} = \sum (x - \bar{x})^2$$
$$df_{total} = N - 1$$

The sum of squares is the sum of the squared lengths of the vertical lines (deviations from the mean).

$$SS_{treatment} = \sum n_i (\bar{x}_i - \bar{x})^2$$
$$df_{treatment} = a - 1$$

Do you see now why we've been multiplying the deviations by the group size to find $SS_{treatment}$?

$$SS_{error} = \sum (x - \bar{x}_i)^2$$
$$df_{error} = N - a$$

Another way to look at ANOVA is this: the hypothesis test we have performed effectively compares two models (Myers & Well, 1995, p. 440-1): one (restricted) model allows for the effects of a mean only — all other variability is 'error'; the other (full) model allows for the effects of a mean and a treatment (and everything else is error). If the full model accounts for significantly more variability than the restricted model, we reject the null hypothesis that the treatment has no effect.

2.2.9 Relating SS calculations to the structural model

Note that our structural model was this:

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij}$$
$$\tau_i = \mu_i - \mu$$
$$\varepsilon_{ij} = X_{ij} - \mu_i$$

and our SS were these:

$$SS_{total} = \sum (x - \bar{x})^2$$
$$SS_{treatment} = \sum n(\bar{x}_i - \bar{x})^2$$
$$SS_{error} = \sum (x - \bar{x}_i)^2 = SS_{total} - SS_{treatment}$$

See the similarity? We can prove that the one follows from the other. This is not something we have to do routinely, but it demonstrates how the sums of squares (SS) are derived directly from the model. Our model was this:

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

or

$$X_{ij} = \mu + (\mu_i - \mu) + (X_{ij} - \mu_i)$$

Rearranging to express the left-hand side as a deviation of each score from the overall mean:

$$X_{ij} - \mu = (\mu_i - \mu) + (X_{ij} - \mu_i)$$

Squaring each side:

$$(X_{ij} - \mu)^2 = (\mu_i - \mu)^2 + (X_{ij} - \mu_i)^2 + 2(\mu_i - \mu)(X_{ij} - \mu_i)$$

Summing over *i* and *j*:

$$\sum_i^a \sum_j^n (X_{ij} - \mu)^2 = n\sum_i^a (\mu_i - \mu)^2 + \sum_i^a \sum_j^n (X_{ij} - \mu_i)^2 + 2\sum_i^a \sum_j^n (\mu_i - \mu)(X_{ij} - \mu_i)$$

The far-right term is actually zero:

$$\begin{aligned}\sum_i^a \sum_j^n (\mu_i - \mu)(X_{ij} - \mu_i) &= \sum_i^a (\mu_i - \mu) \times \sum_j^n (X_{ij} - \mu_i) \\ &= \sum_i^a (\mu_i - \mu) \times 0 \\ &= 0\end{aligned}$$

… since the sum of deviations of all observations about their mean is zero. So we're left with:

$$\sum_i^a \sum_j^n (X_{ij} - \mu)^2 = n\sum_i^a (\mu_i - \mu)^2 + \sum_i^a \sum_j^n (X_{ij} - \mu_i)^2$$

$$\mathrm{SS}_{\text{total}} = \mathrm{SS}_{\text{A}} + \mathrm{SS}_{\text{error}}$$

The degrees of freedom are similarly related:

$$df_{\text{total}} = df_{\text{A}} + df_{\text{error}}$$

2.3 Regression ANOVA: the other way to understand the basic logic

2.3.1 Linear regression in terms of sums of squares

Suppose that in some way we can measure the risk of a heart attack (call it *Y*) in many 50-year-old men. If we then want to predict the risk of a heart attack in an unknown 50-year-old man, our best guess would be the mean risk of a heart attack ($\bar{y}$). If we call our predicted variable $\hat{Y}$, and a predicted individual value $\hat{y}$, then our best guess could be written

$$\hat{y} = \bar{y}$$

We could also write it like this:

$$y = \bar{y} + \varepsilon$$

where ε represents 'error variability' or natural variation. The error in our best guess would be the same as the natural variability in *Y* — it would be described by some way by the standard deviation of *Y*, s_Y, or the variance, s_Y^2. The sample variance (which estimates the population variance), remember, is

$$s_Y^2 = \frac{\sum (y - \bar{y})^2}{n - 1}$$

This variance, like any variance, is the *sum of squared deviations about the mean* divided by the *number of degrees of freedom* that the variance estimate is based on. Because they are conveniently additive, we could write the variability in our estimate just in terms of the sum of squared deviations about the mean — the **sum of squares:**

$$SS_Y = SS_{total} = \sum (y - \bar{y})^2$$

This is the *total* variability in cholesterol, so it's sometimes written SS_{total}. Now suppose we also measure cholesterol levels (X) for each of our subjects. We now have (x, y) pairs (cholesterol and heart attack risk) for each subject. We could **predict *Y* from *X*** using linear regression. We would call the predicted variable $\hat{Y}$, and we'd call an individual predicted value $\hat{y}$. A standard linear regression (see handouts at www.pobox.com/~rudolf/psychology) will give us this equation:

$$\hat{Y} = a + bX$$

where a is the intercept and b is the slope. We could also write our model like this:

$$y = \hat{y} + \varepsilon = a + bx + \varepsilon$$

Now our best guess of the heart attack risk of a new subject should be rather better than $\hat{y} = \bar{y}$; if we measure our new subject's cholesterol as well, we can make what should be a better prediction:

$$\hat{y} = a + bx$$

The error in *this* prediction will related to the deviations between the predicted value, $\hat{y}$, and the actual value, y. We could write this either in terms of a variance…

$$s^2_{residual} = \frac{\sum (y - \hat{y})^2}{n - 2}$$

… or as a sum of squares:

$$SS_{residual} = \sum (y - \hat{y})^2$$

If cholesterol is somehow linearly related to heart attack risk, the error in our prediction, which was SS_{total}, has now been *reduced* to SS_{error}. Therefore, the amount of variability in Y that we have *accounted for* by predicting it from X, which we can write as $SS_{regression}$ or SS_{model} or $SS_{\hat{Y}}$, is based on the difference between the predicted values and the overall mean:

$$SS_{model} = \sum (\hat{y} - \bar{y})^2$$

It's also true that

$$\sum (y - \bar{y})^2 = \sum (\hat{y} - \bar{y})^2 + \sum (y - \hat{y})^2$$
$$SS_{total} = SS_{model} + SS_{residual}$$

and that

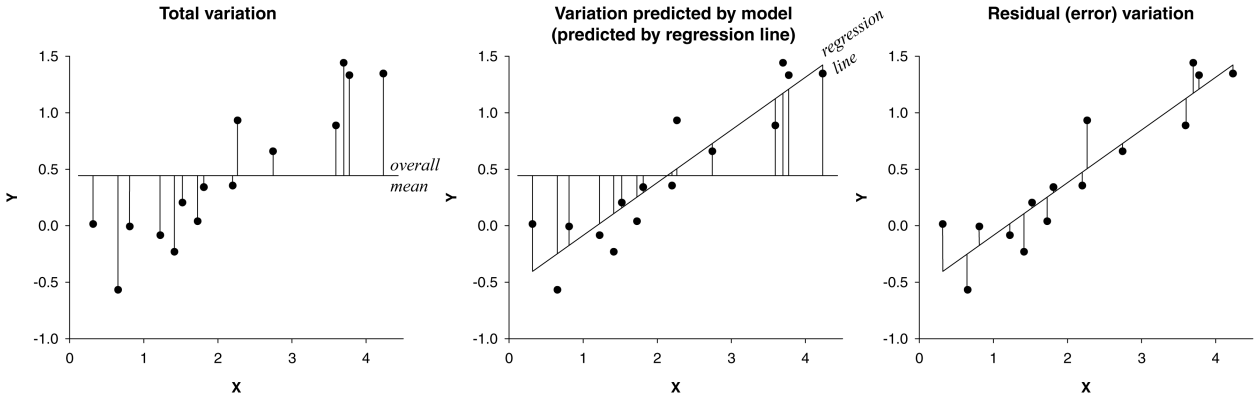
$$df_{total} = df_{model} + df_{residual}$$
$$n - 1 = 1 + (n - 2)$$

Since we have already calculated the overall mean, and the regression line always passes through the overall mean, the regression model has one *df* (its slope). That is, people vary in their cholesterol levels (SS_X), they vary in their heart attack risk (SS_Y = SS_{total}), a certain amount of the variability in their heart attack risk is predictable from their cholesterol ($SS_{\hat{Y}} = SS_{model}$), and a certain amount of variability is left over after you've made that prediction ($SS_{residual}$ = SS_{error}). Incidentally, the *propor-*

tion of the total variability in *Y* that's accounted for by predicting it from *X* is also equal to r^2:

$$r^2 = \frac{\mathrm{SS}_{\hat{Y}}}{\mathrm{SS}_Y} = \frac{\mathrm{SS}_{\text{model}}}{\mathrm{SS}_{\text{total}}}$$

We can illustrate $\mathrm{SS}_{\text{total}}$, $\mathrm{SS}_{\text{model}}$, and $\mathrm{SS}_{\text{residual}}$ like this:



$$\mathrm{SS}_{\text{total}} = \sum (y - \bar{y})^2 \qquad df_{\text{total}} = N - 1$$

$$\mathrm{SS}_{\text{model}} = \sum (\hat{y} - \bar{y})^2 \qquad df_{\text{model}} = 1$$

$$\mathrm{SS}_{\text{error}} = \sum (y - \hat{y})^2 \qquad df_{\text{error}} = N - 2$$

What would it mean to alter $\mathrm{SS}_{\text{model}}$ and $\mathrm{SS}_{\text{residual}}$? If you pulled all of the scores 'further away' from the regression line (if a point is above the regression line, move it up; if it's below, move it down) without changing the slope of the regression line, you'd increase $\mathrm{SS}_{\text{error}}$ without altering $\mathrm{SS}_{\text{model}}$. If you altered the slope of the regression line but moved the individual scores up or down to keep them the same distance from the line, you'd increase $\mathrm{SS}_{\text{model}}$ without changing $\mathrm{SS}_{\text{residual}}$.

2.3.2 Linear regression as an ANOVA

We can use this way of writing a linear regression model to express linear regression as an ANOVA. If there is no correlation between *X* and *Y*, then predicting *Y* from *X* won't be any better than using $\bar{y}$ as our estimate of a value of *y*. So we could obtain a measure of the total variability in *Y*:

$$\mathrm{MS}_Y = \mathrm{MS}_{\text{total}} = s_Y^2 = \frac{\sum (y - \bar{y})^2}{n - 1} = \frac{\mathrm{SS}_{\text{total}}}{df_{\text{total}}}$$

and we could similarly obtain

$$\mathrm{MS}_{\text{model}} = s_{\hat{Y}}^2 = \frac{\sum (\hat{y} - \bar{y})^2}{1} = \frac{\mathrm{SS}_{\text{model}}}{df_{\text{model}}}$$

$$\mathrm{MS}_{\text{residual}} = \mathrm{MS}_{\text{error}} = s_{\text{residual}}^2 = \frac{\sum (y - \hat{y})^2}{n - 2} = \frac{\mathrm{SS}_{\text{residual}}}{df_{\text{residual}}}$$

If the null hypothesis is true and there is no correlation between *X* and *Y*, then some of the variation in *Y* will, by chance, fit a linear model, and contribute to $\mathrm{SS}_{\text{model}}$. The rest will not, and will contribute to $\mathrm{SS}_{\text{residual}}$. The corresponding MS values, once we have divided by the *df*, will be measuring the same thing — the variability of *Y*. That is, under the null hypothesis, $E(\mathrm{MS}_{\text{model}}) = E(\mathrm{MS}_{\text{error}})$. On the other hand, if there is a correlation, and *Y* varies consistently with *X*, then $\mathrm{SS}_{\text{model}}$ will contain variation due to this effect as well as variation due to other things (error), but $\mathrm{SS}_{\text{residual}}$ will only contain variation due to other things (error). Therefore, if the null hypothesis is false, $E(\mathrm{MS}_{\text{model}}) > E(\mathrm{MS}_{\text{error}})$. We can therefore compare $\mathrm{MS}_{\text{model}}$ to $\mathrm{MS}_{\text{error}}$ with an *F* test;

if they are significantly different, we reject the null hypothesis. Our ANOVA table would look like this:

Source	d.f.	SS	MS	*F*
Model	1	SS_{model}	SS_{model} / df_{model}	MS_{model} / MS_{error}
Error (residual)	*N*–2	SS_{error}	SS_{error} / df_{error}	
Total	*N*–1	SS_{total}		

where N is the total number of (x, y) observations. To calculate a regression ANOVA by hand, SS_{total} can be calculated as $s_Y^2(n-1)$ and SS_{model} can be calculated as $r^2 \times SS_{total}$.

2.4 Factors versus covariates

We've seen that we can perform an ANOVA to predict our dependent variable using a **discrete variable, or factor,** that has several 'levels' — as when we asked whether word recall differed between five groups that had read the same word list in different ways. We saw a pictorial representation of a three-group example. We've also seen that we can perform an ANOVA to predict our dependent variable using a **continuous variable, or covariate,** as in our linear regression example, and we've seen a pictorial representation of that.

The mathematical technique of ANOVA does not 'care' whether our predictor variables are discrete (factors) or continuous (covariates). We'll see that in Part 6 when we look at the idea of a **general linear model** (p. 84).

However, the way most people *use* covariates is slightly different from the way they use factors. If you are running an experiment, you do not generally *assign* subjects to different values of a continuous variable (covariate) — you assign subjects to different levels of a factor, with several subjects per level (group). Therefore, real-life covariates are generally things that you *measure* rather than things that you *manipulate*. As a consequence, most people use covariates and analysis of covariance (ANCOVA) as a way to **increase the power** of ANOVA — if you can account for some of your 'error' variability by using a covariate to predict your dependent variable, there is less 'error' variability and therefore there may be more power to detect effects of the factors that you're interested in.

2.5 Assumptions of ANOVA involving covariates

Take a common design involving covariates: a design with one between-subjects factor and one between-subjects covariate. Suppose you have 100 children at your disposal. You measure their IQ. Then you randomly assign 50 children to receive the standard method of maths teaching, and 50 children to receive a new method. This represents the between-subject factor Method, with two levels. After some time, you measure their mathematical problem-solving ability. But you suspect that their IQ may also play a part in determining their final score, not just the teaching method — IQ may be contributing to the 'error' (unmeasured) variability in the scores of your two groups. So you enter IQ as a covariate into your ANOVA model. This covariate may therefore account for some of the previously-unmeasured variability, reducing your 'error' term, and increasing the power to detect an effect of teaching method.

If you use ANCOVA in this way, there are a few assumptions (Myers & Well, 1995, pp. 439-440; Howell, 1997, p. 587):

- that the relationship between the covariate and the dependent variable is **linear;**
- that the regression slopes relating the covariate to the dependent variable are the same in both groups — **homogeneity of regression.**

This is the design discussed in §7.12.1 (p. 138). The second assumption is directly testable, and the method for testing it is discussed in §7.12.2 (p. 144).

A final assumption in this sort of design is this:

- that the covariate and the treatment are **independent** (Myers & Well, 1995, p. 451). If this is not the case, interpretation is very difficult. Using *X* as a covariate removes the component of *Y* predictable from *X*. If the treatment influences *X* or is otherwise predictable from *X*, performing an ANCOVA will not simply remove nuisance variability from *Y*; it will remove part of the effect of the treatment itself. For example, if you had measured IQ at the end of the experiment and the teaching method actually influenced IQ, interpretation would be very hard; similarly, it would be hard to interpret if you had assigned high-IQ students to one teaching method and low-IQ students to another. This can also be a problem in situations when you are using (for example) patient groups and IQ (if the patients have a different IQ to the controls), or sex and body mass (males have a higher body mass).

2.6 ANOVA with two between-subjects factors

We can extend our basic one-way ANOVA to two factors. Suppose we have two factors, one with two levels and one with five levels; this design would be called a 2 × 5 factorial. Suppose we repeat our previous experiment (Howell, 1997, p. 403) but for young and old subjects. Factor A is age (young versus old); factor B is task type (counting, rhyming, adjective, imagery, intentional). Suppose our results look like this:

No. words recalled	**B_1 Counting**	**B_2 Rhyming**	**B_3 Adjective**	**B_4 Imagery**	**B_5 Memorize**	Total
A_1 old	9	7	11	12	10	
	8	9	13	11	19	
	6	6	8	16	14	
	8	6	6	11	5	
	10	6	14	9	10	
	4	11	11	23	11	
	6	6	13	12	14	
	5	3	13	10	15	
	7	8	10	19	11	
	7	7	11	11	11	
total	T_{A1B1} = 70	T_{A1B2} = 69	T_{A1B3} = 110	T_{A1B4} = 134	T_{A1B5} = 120	T_{A1} =503
A_2 young	8	10	14	20	21	
	6	7	11	16	19	
	4	8	18	16	17	
*This dotted line encloses one **cell.** This is a very important term to understand!*	6	10	14	15	15	
	7	4	13	18	22	
	6	7	22	16	16	
	5	10	17	20	22	
	7	6	16	22	22	
	9	7	12	14	18	
	7	7	11	19	21	
	T_{A2B1} = 65	T_{A2B2} = 76	T_{A2B3} = 148	T_{A2B4} = 176	T_{A2B5} = 193	T_{A2} = 658
	T_{B1} = 135	T_{B2} = 145	T_{B3} = 258	T_{B4} = 310	T_{B5} = 313	$T = \Sigma x$ = 1161

Note our definition of **cell** — one particular (A, B) condition, such as A_2B_1 (shown here with a dotted line around it).

2.6.1 Structural model and terminology (main effects, interactions, simple effects)

Our ANOVA must allow for the effects of factor A, and factor B. It should also allow the possibility that A and B **interact** — that the effect of factor A depends on the level of factor B, or vice versa. For example, suppose that young people are generally better, regardless of task type; we would call this a **main effect** of factor A (age). A main effect is an effect of a factor *regardless* of (ignoring) the other factor(s). Suppose that the 'memorize' condition gives better recall than the 'counting' condition, regardless of age; we would call this a main effect of factor B (task type).

On the other hand, perhaps young people have a particular advantage in the 'memorize' condition but not in other conditions; this would be an **interaction** between A and B, written 'A × B' or sometimes 'AB'. We may also define, for later, the term **simple effect:** this is an effect of one factor *at only one level* of another factor. For example, if the 'memorize' condition gives better performance than the 'adjective' condition considering young people only, this is a simple effect of factor B (task type) *at the 'young' level of factor A (age).*

We can specify a model, just as we did before:

$$X_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

where

- X_{ijk} = the score of person k in condition A_iB_j
- μ = the overall mean score
- α_i = the degree to which the mean of condition A_i deviates from the overall mean (= the contribution of condition A_i), i.e. $\alpha_i = \mu_{A_i} - \mu$. By this definition, $\sum \alpha_i = 0$.
- β_j = the degree to which the mean of condition A_i deviates from the overall mean (= the contribution of condition B_j), i.e. $\beta_j = \mu_{B_j} - \mu$. By this definition, $\sum \beta_j = 0$.
- $\alpha\beta_{ij}$ = the degree to which the mean of condition A_iB_j deviates from what you'd expect based on the overall mean and the separate contributions of A_i and B_j (= the interaction A × B), i.e. $\alpha\beta_{ij} = \mu_{A_iB_j} - (\mu + \alpha_i + \beta_j)$. By this definition, $\sum_i \alpha\beta_{ij} = \sum_j \alpha\beta_{ij} = 0$.
- ε_{ijk} = the 'error' or amount by which person k in condition A_iB_j deviates from the mean of his or her group (the 'uniqueness' of person k in condition A_iB_j), i.e. $\varepsilon_{ijk} = X_{ijk} - (\mu_j + \alpha_i + \beta_j + \alpha\beta_{ij})$. By our usual assumption of normal distribution of error, ε_{ijk} is normally distributed with mean 0 and variance σ_e^2.

2.6.2 Expected mean squares

Although we won't derive it, the EMS terms are:

Source	E(MS)
A	$\sigma_e^2 + nb\sigma_A^2$
B	$\sigma_e^2 + na\sigma_B^2$
AB (A × B)	$\sigma_e^2 + n\sigma_{AB}^2$
Error	σ_e^2

(Note that these EMS values assume that the factors are **fixed factors;** see p. 31.) So we should be able to form F ratios based on the error term. For example, if the null hypothesis that factor A has no effect is true, $\mu_{A1} = \mu_{A2} = 0$, so $\sigma_A^2 = 0$ and $E(\mathrm{MS_A}) = E(\mathrm{MS_{error}})$. If this null hypothesis is false, $E(\mathrm{MS_A}) > E(\mathrm{MS_{error}})$. So the ratio

$$\frac{E(\mathrm{MS_A})}{E(\mathrm{MS_{error}})} = \frac{\sigma_e^2 + nb\sigma_A^2}{\sigma_e^2}$$

can be tested using an F test with df_A and df_{error} degrees of freedom.

2.6.3 Degrees of freedom

There are n = 10 subjects per (A, B) condition (per cell), so N = 100 observations in all. Therefore, df_{total} = 99. By our usual rules, df_A = 1 and df_B = 4 (one less than the

number of levels). The **interaction term,** written 'A × B' or 'AB', represents the possibility that the effects of factors A and B represent each other. The *df* for an interaction term A × B is always the product of df_A and df_B — in our example, 4. So our total *df* are partitioned like this:

$$df_{total} = df_A + df_B + df_{AB} + df_{error}$$
$$99 = 1 + 4 + 4 + df_{error}$$

so we have 90 error *df* in our example.

2.6.4 Sums of squares

Similarly,

$$SS_{total} = SS_A + SS_B + SS_{AB} + SS_{error}$$

SS_{total} is calculated exactly as before: the sum of squared deviations of *every* observation from the grand mean.

$$SS_{total} = \sum (x - \bar{x})^2 \left[= \sum x^2 - \frac{(\sum x)^2}{n} \right]$$

The SS for factor A is calculated exactly as we would if this were a one-way ANOVA without the other factor. The same's true for SS_B. That is, we take the *sum of the squares of the deviations of the means of each A condition (A_1, A_2…) from the overall mean, summed over every observation.* (In terms of totals, it's the sum of the squares of the deviations of the totals of each A condition — A_1, A_2, … — from the overall mean total, divided by the number of observations on which each mean was based.) In our example, since there are 2 A conditions and each is made up of *n* observations per cell and 5 cells (= *b* = levels of B) per A condition, there are *nb* observations contributing to each A condition mean. So:

$$SS_A = \sum nb(\bar{x}_A - \bar{x})^2 \left[= \frac{\sum (T_A - \bar{T})^2}{nb} = \frac{\sum T_A^2}{nb} - \frac{(\sum x)^2}{N} \right]$$
$$SS_B = \sum na(\bar{x}_B - \bar{x})^2 \left[= \frac{\sum (T_B - \bar{T})^2}{na} = \frac{\sum T_B^2}{na} - \frac{(\sum x)^2}{N} \right]$$

To find the interaction term SS_{AB}, we calculate an intermediate value, SS_{cells}, which measures the variability of the cell means. Since cell variability can be due to A, B, or AB, we can see that $SS_{cells} = SS_A + SS_B + SS_{AB}$, and therefore calculate SS_{AB} this way. SS_{cells} is the *sum of the squares of the deviations of individual cell means from the grand mean, summed over each observation.* (In terms of totals, it's the sum of the squares of the deviations of individual cell totals from the grand mean total, divided by the number of observations that contributed to each cell mean — i.e. the number of observations per cell.) Whew.

$$SS_{cells} = \sum n(\bar{x}_{AB} - \bar{x})^2 \left[= \frac{\sum (T_{AB} - \bar{T})^2}{n} = \frac{\sum T_{AB}^2}{nb} - \frac{(\sum x)^2}{N} \right]$$
$$SS_{AB} = SS_{cells} - (SS_A + SS_B)$$

To find the error term, we know that

$$SS_{total} = SS_A + SS_B + SS_{AB} + SS_{error} = SS_{cells} + SS_{error}$$

so we can find SS_{error} by subtraction. Alternatively, we could calculate SS_{error} as the *grand sum of the sums of the squares of the deviations of individual observations from their cell means.*

2.6.5 Relating SS calculations to the structural model

Note that our structural model was this:

$$
\begin{aligned}
X_{ijk} &= \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \\
\alpha_i &= \mu_{A_i} - \mu \\
\beta_j &= \mu_{B_j} - \mu \\
\alpha\beta_{ij} &= \mu_{A_i B_j} - (\mu + \alpha_i + \beta_j) \\
\varepsilon_{ijk} &= X_{ijk} - (\mu + \alpha_i + \beta_j + \alpha\beta_{ij})
\end{aligned}
$$

and our SS were these:

$$
\begin{aligned}
\mathrm{SS}_{\mathrm{total}} &= \sum (x - \bar{x})^2 \\
\mathrm{SS}_{\mathrm{A}} &= \sum nb(\bar{x}_A - \bar{x})^2 \\
\mathrm{SS}_{\mathrm{B}} &= \sum na(\bar{x}_B - \bar{x})^2 \\
\mathrm{SS}_{\mathrm{AB}} &= \sum n(\bar{x}_{AB} - \bar{x})^2 - (\mathrm{SS}_{\mathrm{A}} + \mathrm{SS}_{\mathrm{B}}) \\
\mathrm{SS}_{\mathrm{error}} &= \mathrm{SS}_{\mathrm{total}} - (\mathrm{SS}_{\mathrm{A}} + \mathrm{SS}_{\mathrm{B}} + \mathrm{SS}_{\mathrm{AB}})
\end{aligned}
$$

See the similarity?

2.6.6 ANOVA table

We've ended up with this:

Source of variation	d.f.	SS	MS	*F*
Between cells	$df_A + df_B + df_{AB}$	$\mathrm{SS}_{\mathrm{cells}}$		
A	$a-1$	SS_{A}	$\mathrm{SS}_{\mathrm{A}} / df_{\mathrm{A}}$	$\mathrm{MS}_{\mathrm{A}} / \mathrm{MS}_{\mathrm{error}}$
B	$b-1$	SS_{B}	$\mathrm{SS}_{\mathrm{B}} / df_{\mathrm{B}}$	$\mathrm{MS}_{\mathrm{B}} / \mathrm{MS}_{\mathrm{error}}$
AB (A × B)	$(a-1)(b-1)$	$\mathrm{SS}_{\mathrm{AB}}$	$\mathrm{SS}_{\mathrm{AB}} / df_{\mathrm{AB}}$	$\mathrm{MS}_{\mathrm{AB}} / \mathrm{MS}_{\mathrm{error}}$
Within cells (= error = S / cells)	$ab(n-1)$	$\mathrm{SS}_{\mathrm{error}}$	$\mathrm{SS}_{\mathrm{error}} / df_{\mathrm{error}}$	
Total	$N-1 = abn - 1$	$\mathrm{SS}_{\mathrm{total}}$	$\mathrm{SS}_{\mathrm{total}} / df_{\mathrm{total}}$ [$= s^2$]	

2.7 Within-subjects (repeated measures) ANOVA

Principle: if a set of measurements are more correlated than we would expect by chance, we must account for this correlation. We can say that these measurements come from the same 'subject' (in psychological terminology), or that this measure was 'repeated'.

Suppose we have one within-subjects factor. Call it U. Let's suppose we've measured all our subjects in three conditions (U_1 hot, U_2 warm, U_3 cold), once each, and have counterbalanced appropriately to avoid nasty order effects. All we have to do is to partition the sum of squares so as to account for the fact that we've measured subjects several times each…

2.7.1 Structural model

Our structural model is either one of these two:

$$X_{ij} = \mu + \pi_i + \alpha_j + \varepsilon_{ij} \text{ (Model 1: 'additive')}$$

$$X_{ij} = \mu + \pi_i + \alpha_j + \pi\alpha_{ij} + \varepsilon_{ij} \text{ (Model 2: 'nonadditive')}$$

where

- X_{ij} is the dependent variable for subject i in condition U_j
- μ is the overall mean

- π_i is the contribution from a particular person or subject (subject *i*, or S_i): $\pi_i = \mu_{S_i} - \mu$.
- α_j is the contribution from a particular level (level *j*) of the factor U: $\alpha_j = \mu_{U_j} - \mu$.
- $\pi\alpha_{ij}$ is the contribution from the interaction of subject *i* with treatment *j*: $\pi_i = \mu_{S_i U_j} - (\mu + \pi_i + \alpha_j)$. This interaction would reflect that the subjects responded differently to the different levels of U.
- ε_{ij} is everything else (the experimental error associated with subject *i* in condition *j*). In Model 1, this will be $\varepsilon_{ij} = X_{ij} - (\mu + \pi_i + \alpha_j)$. In Model 2, this will be $\varepsilon_{ij} = X_{ij} - (\mu + \pi_i + \alpha_j + \pi\alpha_{ij})$.

These two models differ in the presence or absence of $\pi\alpha_{ij}$, the interaction of U with a particular person (Howell, 1997, pp. 452-454). Including it makes for a realistic model — it is likely that subjects do not all respond equally to all conditions (levels of U). However, if we measure each person in each condition once, we will not be *able* to measure differences in the way subjects respond to different conditions *independently of other sources of error* such as measurement error. (To do that, we'd need to measure subjects more than once per condition, and then we'd need a different model again!) This is another way of saying that the S × U interaction is confounded with — is! — the 'error' term.

2.7.2 Degrees of freedom

We partition the *df* like this:

$$\begin{aligned} df_{\text{total}} &= df_{\text{between subjects}} + df_{\text{within subjects}} \\ df_{\text{within subjects}} &= df_{\text{U}} + df_{\text{error S×U}} \end{aligned}$$

Therefore

$$\begin{aligned} df_{\text{total}} &= df_{\text{between subjects}} + df_{\text{U}} + df_{\text{error}} \\ df_{\text{total}} &= N - 1 \\ df_{\text{between subjects}} &= s - 1 \\ df_{\text{U}} &= u - 1 \\ df_{\text{error}} &= df_{\text{total}} - df_{\text{between subjects}} \end{aligned}$$

where *s* is the number of subjects, *u* is the number of levels of *U*, and *N* is the total number of observations (= *su*). We could also write $df_{\text{between subjects}}$ as df_{S}, which you sometimes see.

2.7.3 Sums of squares

Similarly, we can partition the SS like this:

$$\begin{aligned} \text{SS}_{\text{total}} &= \text{SS}_{\text{between subjects}} + \text{SS}_{\text{within subjects}} \\ \text{SS}_{\text{within subjects}} &= \text{SS}_{\text{U}} + \text{SS}_{\text{error S×U}} \\ \text{SS}_{\text{total}} &= \text{SS}_{\text{between subjects}} + \text{SS}_{\text{U}} + \text{SS}_{\text{error}} \end{aligned}$$

We can define our SS as usual…

$$\begin{aligned} \text{SS}_{\text{total}} &= \sum (x - \bar{x})^2 \\ \text{SS}_{\text{U}} &= \sum s(\bar{x}_U - \bar{x})^2 \\ \text{SS}_{\text{between subjects}} &= \sum u(\bar{x}_S - \bar{x})^2 \end{aligned}$$

where *s* is the number of subjects and *u* is the number of levels of *U*. $\bar{x}_U$ represents the mean for a particular level of *U* (across subjects), and $\bar{x}_S$ represents the mean for a particular subject (across levels of *U*). Our total number of observations will be $N = su$.

2.7.4 EMS and ANOVA summary table

The EMS depend on which model we use:

Source of variation	Model 1: E(MS)	Model 2: E(MS)
Between subjects (S)	$\sigma_e^2 + u\sigma_S^2$	$\sigma_e^2 + u\sigma_S^2$
U	$\sigma_e^2 + s\sigma_U^2$	$\sigma_e^2 + \sigma_{US}^2 + s\sigma_U^2$
Error	σ_e^2	$\sigma_e^2 + \sigma_{US}^2$

This means that in Model 2 it's rather hard to do a proper *F* test for the 'between subjects' factor, since there's no term whose *E*(MS) is identical to $E(\text{MS}_{\text{between subjects}})$ except for the presence of σ_S^2, the relevant variance for the between-subjects factor. On the other hand, who cares. If this term were significant, all it would tell us is that subjects are different, which is hardly earth-shattering. Either way, we have no problem testing *U*: the proper way to test for an effect of *U* is to do an *F* test comparing MS_U to MS_{error}.

If Model 1 is true — if subjects respond equally to the treatments; if the effects are additive — we will have more *power* to detect effects of *U*, since if the null hypothesis (that *U* has no effect) is false,

$$\frac{E(\text{MS}_{\text{U-model 1}})}{E(\text{MS}_{\text{error-model 1}})} = \frac{\sigma_e^2 + s\sigma_U^2}{\sigma_e^2} > \frac{\sigma_e^2 + \sigma_{US}^2 + s\sigma_U^2}{\sigma_e^2 + \sigma_{US}^2} = \frac{E(\text{MS}_{\text{U-model 2}})}{E(\text{MS}_{\text{error-model 2}})}$$

and the bigger the ratio of MS_U to MS_{error}, the bigger the *F* ratio, and the more likely the effect is to be 'significant' (Myers & Well, 1995, p. 244; Howell, 1997, pp. 452-454).

You may be thinking 'the calculations for the two models are exactly the same in practice, so why all this fuss?' You'd be right — unless you wanted to estimate the proportion of variance accounted for by a particular term (Myers & Well, 1995, pp. 242, 252-255). See p. 112.

2.8 Assumptions of within-subjects ANOVA: Mauchly, Greenhouse–Geisser, etc.

2.8.1 Short version

1. **Any ANOVA involving within-subjects factors has a potential problem. There is an assumption known as 'sphericity [of the covariance matrix]'. If this assumption is violated, Type I error rates will be inflated (if the null hypothesis is true, you will get too many results that you will declare 'significant' than you should).**

Mauchly's test of sphericity checks for this. A significant Mauchly's test means that the assumption is likely to have been violated. But it's not a very good test (see below), so we should probably ignore it.

2. **Correct the *df* for any term involving a within-subjects factor, and the corresponding error *df*, by multiplying them both by a correction factor.** The correction factor is known as 'epsilon' (ε). If the sphericity assumption is not violated, $\varepsilon = 1$ (so applying the correction changes nothing). You do not need to correct any terms that have only between-subjects factors. And you can never

violate the sphericity assumption for a within-subjects factor that has only 2 levels.

3. **Use either the Greenhouse–Geisser or the Huynh–Feldt epsilon.** The Greenhouse–Geisser one (sometimes written $\hat{\varepsilon}$) is probably a bit too conservative; the Huynh–Feldt one (sometimes written $\tilde{\varepsilon}$) is better (Myers & Well, 1995, p. 248; Howell, 1997, p. 465) — but more detail below.

SPSS reports Mauchly's test and both the G–G and H–F corrections whenever you run a within-subjects ANOVA using its menus.

Just to confuse you, there are actually several different approaches:

- NOT THE BEST: Look at the results of Mauchly's test; apply a correction (G–G or H–F) if and only if Mauchly's test is significant for a factor that's part of the term in question, indicating a violation of the sphericity assumption. This is *not* ideal, because Mauchly's test isn't very reliable (Myers & Well, 1995, p. 246; Howell, 1997, p. 466, and see below).
- NOT THE BEST: Always use the Greenhouse–Geisser correction. Too conservative.
- **Good and simple: Always use the Huynh–Feldt correction.** This is not totally ideal because the H–F procedure tends to overestimate sphericity (be a bit too optimistic) (see refs in Field, 1998), but it's pretty good; Myers & Well (1995, p. 248) recommend it.
- OK but awkward: use the average of the $\hat{\varepsilon}$ and $\tilde{\varepsilon}$.
- **Good: Look at the estimated epsilons (G–G $\hat{\varepsilon}$ and H–F $\tilde{\varepsilon}$); if they're in the region of 0.75 or higher (in some textbooks, if $\hat{\varepsilon} \geq 0.75$) use the H–F $\tilde{\varepsilon}$; if below, use the G–G $\hat{\varepsilon}$** (Howell, 1997, pp. 465-466).

Of course, if you really want to avoid Type I errors, you'd be predisposed to using the G–G correction (conservative); if you'd rather avoid Type II errors, you'd be predisposed to using the H–F correction (more liberal).

2.8.2 Long version

Sphericity is the assumption of *homogeneity of variance of difference scores* (Myers & Well, 1995, p. 244-250); see also www-staff.lboro.ac.uk/~hutsb/Spheric.htm. Suppose we test 5 subjects at three levels of A. We can therefore calculate three sets of difference scores ($A_3 - A_2$), ($A_2 - A_1$), and ($A_3 - A_1$), for each subject. Sphericity is the assumption that the variances of these difference scores are the same. Here are two examples:

Data set A: exhibiting sphericity (homogeneity of variance of difference scores)

Subject	A_1	A_2	A_3	difference $A_3 - A_2$	difference $A_2 - A_1$	difference $A_3 - A_1$
S_1	21.05	7.214	26.812	19.598	-13.836	5.762
S_2	6.915	29.599	16.366	-13.233	22.684	9.451
S_3	3.89	21	41.053	20.053	17.11	37.163
S_4	11.975	12.401	18.896	6.495	0.426	6.921
S_5	31.169	34.786	31.872	-2.914	3.617	0.703
mean	15.00	21.00	27.00	6.00	6.00	12.00
variance	124.00	132.00	100.00	**208.00**	**208.00**	**208.00**

Data set B: exhibiting nonsphericity

Subject	A_1	A_2	A_3	difference $A_3 - A_2$	difference $A_2 - A_1$	difference $A_3 - A_1$
S_1	1.7	3.9	6	2.1	2.2	4.3
S_2	4.4	6.5	14.5	8	2.1	10.1
S_3	7.8	13.3	18.6	5.3	5.5	10.8
S_4	6.6	9.4	14.5	5.1	2.8	7.9
S_5	9.1	15.2	23.5	8.3	6.1	14.4
mean	5.92	9.66	15.42	5.76	3.74	9.50
variance	8.56	21.79	41.46	**6.38**	**3.65**	**13.92**

In general, if there are *a* treatment levels, there are $\frac{a(a-1)}{2}$ difference scores, and it is assumed that they all have the same variance.

Obviously, the sphericity assumption cannot be violated if the within-subjects factor has less than 3 levels.

The sphericity assumption will be met if there is no S × A interaction (if there is **additivity**). In this case, any difference score is *exactly the same* over subjects, so there is zero variance in the difference scores. However, sphericity can be met without additivity, as shown above (that is to say, additivity is sufficient but not necessary for sphericity).

Another condition that is sufficient (but not necessary) for sphericity is **compound symmetry.** This requires homogeneity of the population treatment variances:

$$\sigma_{A1}^2 = \sigma_{A2}^2 = \ldots$$

and homogeneity of the population covariances:

$$\rho_{A1,A2}\sigma_{A1}\sigma_{A2} = \rho_{A1,A3}\sigma_{A1}\sigma_{A3} = \rho_{A2,A3}\sigma_{A2}\sigma_{A3} = \ldots$$

where $\rho_{A1,A2}$ is the population correlation between the A1 and A2 scores, and $\rho_{A1,A2}\sigma_{A1}\sigma_{A2}$ is their covariance (see handouts at pobox.com/~rudolf/psychology). The variance sum law tells us that the variance of a difference between two variables is

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho_{XY}\sigma_X\sigma_Y$$

and so if the two conditions above are met, the variances of the difference scores will all be the same. Howell (1997, p. 455) explains why the term 'compound symmetry' is applied to this situation, using a matrix that illustrates variances and covariances between A_1, A_2, and A_3 (this is illustrated under *covariance matrix* in the Glossary on p. 214). However, the explanation is not as clear as Myers & Well's. Yet data set A shown above exhibits sphericity without compound symmetry (that is, although the variances of difference scores are identical, i.e. sphericity is true, the variances of the individual scores are not the same and nor are the covariances for pairs of treatments).

Myers & Well (1995, p. 246) don't like Mauchly's test because it tends to give 'significant' results (suggesting a problem) even in situations when sphericity holds — that is, using Mauchly's test is a conservative approach.

The three things you can do about violations of sphericity are (1) the usual *F* test with adjusted degrees of freedom, as suggested above (after Box, 1954); (2) multivariate ANOVA (MANOVA) (see p. 92); (3) tests of planned contrasts (see p. 75). See Myers & Well (1995, pp. 246-252).

2.9 Missing data in designs involving within-subjects factors

If some data are lost for a particular subject, you have a problem. You can either assume the 'additive' model discussed above — that the effect of within-subjects factors are the same for all subjects — and **estimate** the missing value (Myers & Well, 1995, p. 256-8). Every time you estimate a value, you reduce the *df* for the relevant error term by 1. If you don't assume the additive model, you can't estimate the value, and you may then have to **throw out all data for that subject. SPSS does the latter by default.**

2.10 Mixed ANOVA (with both between-subjects and within-subject factors)

We will illustrate the simplest mixed design here: one between-subjects factor and one within-subjects factor. General principles of more complicated within-subjects models are given by Keppel (1991, pp. 491-496), and laid out in Part 7.

Suppose we take three groups of rats, $n = 8$ subjects per group ($s = 24$ subjects total). We give one group treatment A_1, one group treatment A_2, and one group treatment A_3 ($a = 3$). One subject only experiences one treatment. Note that $s = an$. Then we measure every subject's performance at six time points $U_1 \ldots U_6$ ($u = 6$). We have $N = su = anu = 8 \times 3 \times 6 = 144$ observations in total.

We first partition the total variation into *between-subjects* variability and *within-subjects* variability.

The between-subjects variability can be attributed to either the effect of the treatment group (A), or differences between subjects in the same group ('S within A' or 'S / A'). (This notation indicates that there is a different group of subjects at each level of the between-subjects factor, A; we could not measure simply 'subject variation independent of the effects of A' since no subjects ever serve in more than one group, or level of A.) So we have these sources of between-subjects variability:

A
S / A

The within-subjects variability can be attributed to either the effects of the time point (U), or an interaction between the time point and the drug group (U × A), or an interaction between the time point and the subject-to-subject variability, which again we can only measure *within* a drug group (U × S / A). So we have these sources of within-subject variability:

U
U × A
U × S / A

2.10.1 Structural model

Following Myers & Well (1995, p. 295-6):

$$X_{ijk} = \mu + \alpha_i + \pi_{j/i} + \beta_k + \alpha\beta_{ik} + \pi\beta_{jk/i} + \varepsilon_{ijk}$$

where

- X_{ijk} is the dependent variable for subject j in group A_i and condition U_k
- μ is the overall mean
- α_i is the contribution from a particular level (level i) of factor A: $\alpha_i = \mu_{A_i} - \mu$. By this definition, $\sum \alpha_i = 0$.
- $\pi_{j/i}$ is the contribution from a particular person or subject (subject j), who only serves *within* condition A_i ('subject within group', or S / A): $\pi_{j/i} = \mu_{S_j/A_i} - \mu$.

(There is no straightforward interaction of A with S: every subject is only measured at one level of A, so this term would be indistinguishable from the subject-only effect $\pi_{j/i}$.)

- β_k is the contribution from a particular level (level *k*) of factor U: $\beta_k = \mu_{U_k} - \mu$. By this definition, $\sum \beta_j = 0$.
- $\alpha\beta_{ik}$ is the contribution from the interaction of A_i and U_k: $\alpha\beta_{ik} = \mu_{A_i U_k} - (\mu + \alpha_i + \beta_k)$. By this definition, $\sum_i \alpha\beta_{ik} = \sum_k \alpha\beta_{ik} = 0$.
- $\pi\beta_{jk/i}$ is the contribution from the interaction of U_k with subject *j*, which can only be measured within one level of A (it's the 'SU / A' term): $\pi\beta_{jk/i} = \mu_{S_j U_k / A_i} - (\mu + \pi_{j/i} + \beta_k)$. By this definition, $\sum_k \pi\beta_{jk/i} = 0$.

  (There is no straightforward three-way A × U × S interaction: every subject is only measured at one level of A, so this term would be indistinguishable from the SU / A effect $\pi\beta_{jk/i}$.)
- ε_{ijk} is everything else (the experimental error associated with measuring person *j* — who always experiences treatment A_i — in condition U_k): $\varepsilon_{ijk} = X_{ijk} - (\mu + \alpha_i + \pi_{j/i} + \beta_k + \alpha\beta_{ik} + \pi\beta_{jk/i})$.

  Note that we cannot actually measure ε_{ijk} independent of the SU / A term if we only have one measurement per subject per level of *U*; this term simply contributes to the within-subjects error term (see below).

2.10.2 Degrees of freedom

We can partition the *df* like this:

$$\begin{aligned} df_{\text{total}} &= df_{\text{between subjects}} + df_{\text{within subjects}} \\ df_{\text{between subjects}} &= df_{\text{A}} + df_{\text{S/A}} \\ df_{\text{within subjects}} &= df_{\text{U}} + df_{\text{U×A}} + df_{\text{U×S/A}} \end{aligned}$$

So now we can calculate all our *df*. (Often, $df_{\text{between subjects}}$ and $SS_{\text{between subjects}}$ are simply written df_{S} and SS_{S}.)

$$\begin{aligned} df_{\text{total}} &= N - 1 \\ df_{\text{A}} &= a - 1 \\ df_{\text{U}} &= u - 1 \\ df_{\text{U×A}} &= df_{\text{A}} \times df_{\text{U}} \\ df_{\text{between subjects}} &= s - 1 \\ df_{\text{S/A}} &= df_{\text{between subjects}} - df_{\text{A}} \\ df_{\text{U×S/A}} &= df_{\text{within subjects}} - (df_{\text{A}} + df_{\text{U×A}}) \\ &= (df_{\text{total}} - df_{\text{between subjects}}) - (df_{\text{A}} + df_{\text{U×A}}) \end{aligned}$$

2.10.3 Sums of squares

The partitioning is always exactly the same as for the *df*:

$$\begin{aligned} SS_{\text{total}} &= SS_{\text{between subjects}} + SS_{\text{within subjects}} \\ SS_{\text{between subjects}} &= SS_{\text{A}} + SS_{\text{S/A}} \\ SS_{\text{within subjects}} &= SS_{\text{U}} + SS_{\text{U×A}} + SS_{\text{U×S/A}} \end{aligned}$$

So

$$SS_{\text{total}} = SS_{\text{A}} + SS_{\text{S/A}} + SS_{\text{U}} + SS_{\text{U×A}} + SS_{\text{U×S/A}}$$

We have two different 'error' terms, one for the between-subjects factor and one for the within-subjects factor (and its interaction with the between-subjects factor), so

we can't just label them 'SS_{error}'. But we could rewrite the total like this if we wanted:

$$SS_{total} = SS_A + SS_{error\text{-}between} + SS_U + SS_{U\times A} + SS_{error\text{-}within}$$

Now we can calculate the SS. Remember, each SS must be made up of *N* components, because there are *N* observations. Take the example of SS_A: we calculate this by summing over *a* means (namely $\bar{x}_{A_1}, \bar{x}_{A_2}, \ldots \bar{x}_{A_a}$). But each mean is based on (or, if you prefer, contributes to) $N/a = su/a = anu/a = nu$ individual scores; we therefore multiply our deviations by *nu* to get the total SS_A.

$$\begin{aligned}
SS_{total} &= \sum (x - \bar{x})^2 \\
SS_{between\ subjects} &= \sum u(\bar{x}_S - \bar{x})^2 \\
SS_A &= \sum nu(\bar{x}_A - \bar{x})^2 \\
SS_{S/A} &= SS_{between\ subjects} - SS_A \\
SS_U &= \sum s(\bar{x}_U - \bar{x})^2 \\
SS_{U\times A} &= \sum n(\bar{x}_{UA} - \bar{x})^2 \\
SS_{U\times S/A} &= SS_{within\ subjects} - (SS_U + SS_{U\times A}) \\
&= (SS_{total} - SS_{between\ subjects}) - (SS_U + SS_{U\times A})
\end{aligned}$$

Just to make it clear how many scores each mean is based on:

	Subject	**U_1**	**U_2**	**U_3**	**U_4**	**U_5**	**U_6**	***u* = 6**
A_1	**S_1**	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	$\bar{x}_A$ means are
	S_2	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	based on *nu* =
	S_3	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	48 scores
	S_4	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
	S_5	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
	S_6	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
	S_7	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
	S_8	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
A_2	**S_9**	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
	S_{10}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
	S_{11}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	$\bar{x}_S$ means are
	S_{12}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	based on *u* = 6
	S_{13}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	scores
	S_{14}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
	S_{15}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
	S_{16}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
A_3	**S_{17}**	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	$\bar{x}_{UA}$ means are
	S_{18}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	based on *n* = 8
	S_{19}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	scores
	S_{20}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
	S_{21}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
	S_{22}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
	S_{23}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
	S_{24}	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	
			$\bar{x}_U$ means are based					
a = 3	*s* = 24		on *an* = *s* = 24 scores					*N* = *su* = *anu* =
	au = 24							144

2.10.4 ANOVA table

Source	d.f.	SS	*F*
Between subjects (S):	$s-1 = an - 1$		
A	$a-1$	SS_A	$MS_A / MS_{S/A}$
error S / A	$(an-1)-(a-1) = a(n-1)$	$SS_{S/A}$	
Within subjects:	$(N-1)-(s-1) = an(u-1)$		
U	$u-1$	SS_U	$MS_U / MS_{U\times S/A}$
U × A	$(u-1)(a-1)$	$SS_{U\times A}$	$MS_{U\times A} / MS_{U\times S/A}$
error U × S / A	$a(u-1)(n-1)$	$SS_{U\times S/A}$	
Total	$N-1 = aun - 1$	SS_{total}	

where a is the number of levels of factor A, etc., N is the total number of observations (= aun), n is the number of subjects per group (per level of A), and s is the total number of subjects.

2.11 Fixed and random factors

When we consider ANOVA factors we must distinguish **fixed** factors, which contain all the levels we are interested in (e.g. sex: male v. female) and **random** factors, where we have sampled some of the possible levels at random (e.g. subjects). Random factors can be thought of as those whose levels might change; if we repeated the experiment, we might pick different subjects.

Sometimes the fixed / random distinction is pretty much inherent in the factor — Subjects is usually a random factor, for example. But sometimes whether a factor is fixed or random really does depend on the study. Howell (1997, p. 330) uses the example of painkillers. If we are asked to study the relative efficacy of the UK's *four most popular* over-the-counter painkillers, we have no choice in which painkillers we study. If we were asked to repeat the study, we would use the same four painkillers. Painkillers would be a fixed factor. If, on the other hand, we were asked to compare several painkillers to see if 'one brand is as good as the next', we might select a few painkillers *randomly* from the dozens on offer. In this case, where our sample is intended to be representative of painkillers in general but where it is an arbitrary and non-exclusive sample, we would consider Painkiller to be a *random* factor. Further examples are given by Keppel (1991, p. 485 and Appendix C), and by Myers & Well (1995, pp. 270-1).

When we test effects involving a random factor, we often have to test effects against an interaction term. Examples are given in the consideration of within-subjects designs (which involve random factors, since Subjects is a random factor). The determination of appropriate error terms is discussed later in the section on expected mean squares (EMS) (p. 73), which are different for fixed and random factors.

Part 3: practical analysis

3.1 Reminder: assumptions of ANOVA

1. Homogeneity of variance

We assume that each of our groups (conditions) has the same variance.

- **How to check?** In SPSS, Levene's test (Levene, 1960) checks this assumption. To obtain it, choose **Options → Homogeneity tests** from the ANOVA dialogue box. If Levene's test produces a 'significant' result, the assumption is violated — there is heterogeneity of variance. This is a Potentially Bad Thing. Consider transformation of the data (see below, p. 34). You can also plot the standard deviation (and variances) versus the means of each level of a factor by choosing **Options → Spread vs. level plot.**
- Unequal *n*s exaggerate the consequences of heterogeneity of variance — a Bad Thing (p. 33) (see also Myers & Well, 1995, p. 105-106).

2. Normality

We assume that the scores for each condition are normally distributed around the mean for that condition. (This assumption is the same as saying that error is normally distributed within each condition.)

- **How to check?** You can inspect the data to get an idea whether the data are normally distributed in each condition. In SPSS, choose **Analyze → Descriptive Statistics → Explore.** This gives you get stem-and-left plots, boxplots, and so on. In the dialogue box, tick **'Both'** to get statistics and plots; click **Plots → Normality plots with tests.** This produces a **Q–Q plot** — a plot of each score against its expected *z* value (the value it would have if the distribution were normal — calculated as the deviation of the score from the mean, divided by the standard deviation of the scores). If this produces a straight line, the data are normally distributed. You also get the **Kolmogorov–Smirnov test with Lilliefors correction** (Lilliefors, 1967) applied to a normal distribution, and the **Shapiro–Wilk** test (Shapiro & Wilk, 1965) — if these are significant, your data are *not* normally distributed — a Bad Thing. Consider transformation of the data (see below, p. 34).

3. Independence of observations

We also assume that the observations are independent — technically, that the error components (ε) are independent. For any two observations within an experimental treatment, we assume that knowing how one of these observations stands relative to the treatment (or population) mean tells us nothing about the other observation. Random assignment of subjects to groups is an important way of achieving this. We must **account** for any non-independence of observations — for example, observations that are correlated because they come from the same subjects — by adding factors (e.g. Subject) that account for the non-independence. Introducing factors to account for non-independence of observations makes the error terms independent again, and we're OK. However, these designs — known as within-subject or repeated measures designs — have their own assumptions, listed below.

- A statistics package can't check this assumption for you! It depends on your experiment.

3.2 Reminder: assumption of ANOVA with within-subject factors

Sphericity

Any ANOVA involving within-subjects factors assumes **sphericity** (discussed earlier). If this assumption is violated, Type I error rates will be inflated (if the null hypothesis is true, you will get too many results that you will declare 'significant' than you should). A simple plan of action:

- Look at Mauchly's test of sphericity. A significant Mauchly's test means that the assumption is likely to have been violated.
- When the assumption has been violated for a particular within-subjects factor, correct the *df* for any term involving the within-subjects factor, and the corresponding error *df*, by multiplying them both by epsilon (ε).
- Use either the Greenhouse–Geisser or the Huynh–Feldt epsilon. The Greenhouse–Geisser one (sometimes written $\hat{\varepsilon}$) is probably a bit too conservative; the Huynh–Feldt one (sometimes written $\tilde{\varepsilon}$) is better (Myers & Well, 1995, p. 248; Howell, 1997, p. 465).

SPSS reports Mauchly's test and both the G–G and H–F corrections whenever you run a within-subjects ANOVA using its menus.

You never need to correct any terms that have only between-subjects factors. And you can never violate the sphericity assumption for a within-subjects factor that has only 2 levels.

3.3 Consequences of violating the assumptions of ANOVA

- **Independence of observations.** If there are correlations between scores that are not taken account of by the ANOVA model, Type I error rates can be inflated (Myers & Well, 1995, p. 69, 101).

- **Normality.** The Type I error rate is not affected much by sampling from non-normal populations unless the samples are quite small and the departure from normality extremely marked (Myers & Well, 1995, pp. 69, 101). This is the effect of the central limit theorem: the distribution of means and their differences will tend to be normal as *n* increases, even when the distribution of the parent population is not. Things are pretty OK even when the dependent variable is discretely (rather than continuously) distributed (Myers & Well, 1995, p. 101). However, there are **nonparametric** alternatives to ANOVA which may sometimes be better when the normality assumption is violated — such as the Kruskal–Wallis *H* test (Myers & Well, 1995, p. 102-105). For repeated-measures designs, there are others: Friedman's chi-square (χ_F^2), the rank-transformation *F* test (F_r), the Wilcoxon signed-rank test, and Cochran's Q test (Myers & Well, 1995, pp. 271-280).

- **Homogeneity of variance.** If the two sample sizes are equal, there is little distortion to Type I error rates unless *n* is very small and the ratio of the variances is quite large. There's generally not a problem if the ratio of the largest to the smallest variance is no more than 4:1, and sometimes even bigger discrepancies can be tolerated. However, when *n*s are unequal, there's more of a problem. Whether the Type I error rate goes up or down depends on relationship between the sample size and the population variance: if the larger group has the larger variance, the test is conservative, but if the smaller group has the larger variance, the test is liberal — too many Type I errors — and sometimes the Type I error rate gets really high (Myers & Well, 1995, pp. 69-71, 105-110). The two strategies are to use an *alternative test* or to *transform the data* to improve the homogeneity of variance (see p. 34). The alternative tests include the Welch and Brown–Forsythe modified *F* tests (Myers & Well, 1995, pp. 106-109; Howell, 1997, pp. 321-323).

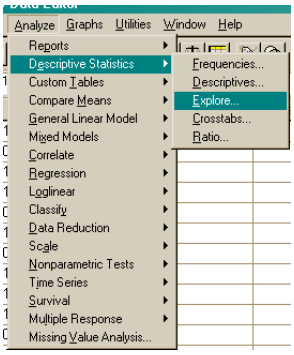
- **Sphericity.** Violations are a problem. We've discussed the solutions elsewhere (p. 25).

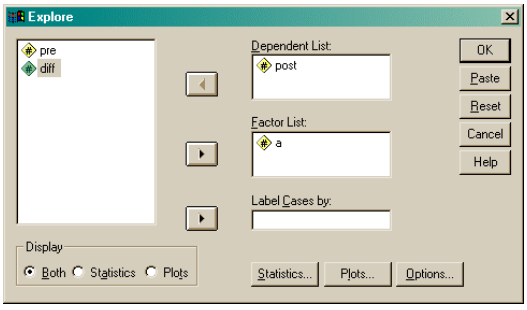
3.4 Exploratory data analysis, transformations, and residuals

3.4.1. Plot your data

It's a very good idea to plot your data before analysing it. Although there are formal tests for things like homogeneity of variance, and the other assumptions of an ANOVA, the tests won't describe the distribution of your data or show you if there are outliers. See also Howell (1997, chapter 2).

In SPSS, you can choose **Analyze → Descriptive Statistics → Descriptives** to get simple descriptive statistics, or **Analyze → Descriptive Statistics → Explore** for a very comprehensive set of options, including descriptive statistics, histograms, stem-and-leaf plots, Q–Q plots (see p. 9; are the data normally distributed?), boxplots (also known as box-and-whisker plots, showing outliers), and so on — with your data broken down by a factor. For example, to analyse 'Post' scores broken down by levels of the factor 'A', I might do this:





3.4.2. Outliers

Outliers can cause substantial problems with parametric statistical tests (e.g. Myers & Well, 1995, p. 15). If you find one, check that the datum has been entered correctly — if not, re-enter it, or if you can't, throw it out. If it was entered correctly, you may consider **removing the outlier.** There is a danger here — we can't simply throw away data we don't like (see Myers & Well, 1995, p. 419), and maybe this is a valid measurement, in which case we shouldn't be chucking it out. But sometimes it represents something we're not interested in. If a reason for the outlier can be established (data mis-entered, equipment broken, subject fell asleep, etc.) then it may be corrected or removed as appropriate. We can always use **nonparametric tests,** which are much less sensitive to outliers. How do we define an outlier? With a boxplot, one convention is to regard *points more than 3 box widths (3 × interquartile range) from the box* as outliers. Another is to consider points outside the whiskers as outliers (Tukey's original suggestion), but this throws away many — too many — data points. Finally, another approach is to define outliers as points **>2 standard deviations from the group mean.**

Another technique related to outlier removal is the use of **trimmed samples.** Rather than transforming your data to achieve homogeneity of variance (see below), another approach to 'heavy-tailed' samples (fairly flat distributions with a lot of observations in the tail — posh name *platykurtic*) is to trim the sample. For example, with 40 cases per sample, a 5% trimmed sample is the sample with the top two and the bottom two observations removed (5% removed from each tail). When comparing several groups, as in ANOVA, each sample would be trimmed by the same percentage. However, there is a special technique required for the ANOVA: the MS_{error} should be based on the variance of the corresponding 'Winsorized' sample — one in which the values you removed are replaced by copies of the next-most-extreme datum (Howell, 1997, p. 329). To my knowledge, this isn't a very common technique.

3.4.3. Transformations

Transformations can be used (1) to transform skewed distributions into something closer to a normal distribution; (2) to reduce heterogeneity of variance; (3) to remedy 'non-additivity' in within-subject designs. A transformation that achieves one purpose well may not be equally suited to other purposes, although transformations that equate variances do tend to give more normally distributed scores (Myers & Well, 1995, p. 109). We'll focus on transformations designed to achieve homogeneous variances. Such transformations can be derived if the relationship between μ_j (the group mean) and σ_j^2 (the group variance) is known. Here are some examples, and a general rule (Myers & Well, 1995, pp. 109-110; Howell, 1997, pp. 323-329):

- **If the data are proportions,** such as 'percent correct', the scores in the population are binomially distributed; the variance can be written as $\sigma_j^2 = k\mu_j(1-\mu_j)$ where *k* is a constant. The appropriate transformation is the **arcsine** transformation: $Y' = \arcsin\sqrt{Y}$. For example, if a datum (*Y* value) is 0.5, the transformed value Y' is $\arcsin\sqrt{0.5} = 45^\circ$; you would use the value 45 for analysis. (Or you could do it in radians; it doesn't matter: π radians = 180°.) Your data should be in the range 0–1; if your data are percentages (97%), analyse them as proportions (0.97). The arcsine transformation stretches out both tails (numbers near to 0 and 1) relative to the middle (numbers near to 0.5).

- **In general…** Plot $\log(\hat{\sigma}_j)$, the log of the standard deviation of each group, against $\log(\overline{Y}_j)$, the log of the mean of each group. If this relation is approximately a straight line, find its slope. The appropriate transformation would be $Y' = Y^{(1-\text{slope})}$. If 1 – slope = 0, take the log of each score instead of raising it to a power.

- **If the data are markedly skewed, or the standard deviation is proportional to the mean,** you will often find that the slope is 1 and a **log** transformation is appropriate. Reaction times may be amenable to this transformation. It is also applicable when the scores themselves are standard deviations. You can use any base for the logarithm (10 is simple, but you could use *e* or anything else). You can't find logarithms of zero or of negative numbers, so if your data are negative it is permissible to add a constant before taking the log: $Y' = \log(Y + k)$. If you have near-zero values, use $Y' = \log(Y + 1)$ rather than $Y' = \log(Y)$.

- **Variance is proportional to the mean.** Consider taking the square root of each datum: $Y' = \sqrt{Y} = Y^{0.5}$. The square-root transformation compresses the upper tail of the distribution. If your scores are small (e.g. <10), you may find that $Y' = \sqrt{Y + 0.5}$ or even $Y' = \sqrt{Y} + \sqrt{Y + 1}$ works better for equating variances.

- **The reciprocal transformation,** $Y' = \frac{1}{Y} = Y^{-1}$, is also useful if the data are positively skewed (a few very large values at the upper end of the distribution).

> Indeed, it may often make a lot of sense to use it — particularly in the example of transforming *reaction times* or latencies to *reaction speeds*.

Don't apply a transformation unless you need to, or it makes theoretical sense. A major problem with transformations is interpreting subsequent analyses. Sometimes transformations make excellent sense, such as in the time-to-speed transformation. Or you might have a theoretical reason to think that Y is a power function of some variable X: $Y = aX^b$. Then analysing $\log(Y)$ and $\log(X)$ would make sense, because their relationship would then be linear and ANOVA techniques are built around linear relationships. And if there is a clear relationship between group means and standard deviations, the appropriate transformation will tend to give a more powerful statistical test. But sometimes transformations that improve heterogeneity of variance don't help you theoretically — you may discover that group 1 makes more $\sqrt{\text{lever presses} + 0.5}$ than group 2, and then have to interpret that in terms of the real world of lever presses. And if relative distances between means are of interest, problems can crop up: Myers & Well (1995, p. 110) give the example of comparing two teaching methods for high- and low-ability subjects. Even if the difference between the two teaching methods were the same for the high- and low-ability groups on the original data scale, the transformation might well produce a different result on the new scale; conversely one method might have more of an advantage for low-ability subjects on the original data scale, but again the results might be quite different on the transformed scale.

If you transform your data, it is only fair that you plot the transformed data in your figures, since that's what you analysed (especially if your figures show indices of variability, such as error bars, and / or make some claims as to significant differences between conditions). You may also choose to report the group means converted back to 'real' units. But be aware that this can be a little misleading. For example, if a group of six rats makes (16, 28, 38, 96, 55, 5) lever presses (mean = 39.67) and you analyse the square-root transformed data (4, 5.29, 6.16, 9.80, 7.42, 2.24), you will find that the mean of the transformed data is 5.82. But 5.82 is the square root of 33.87 — so converting the mean of the transformed data back to the original scale (by applying the reverse of the transformation) doesn't give you the untransformed mean.

If you do need to transform, it is perfectly permissible to 'shop around', trying out several transformations until you find one that does a good job of reducing heterogeneity of variance (Howell, 1997, p. 329). But it is *not* permissible to shop around until you find a transformation that gives you a significant result! You are trying to optimize the data *so that the ANOVA is valid* — you are not trying to 'optimize' the ANOVA result.

3.4.4. Plot your residuals

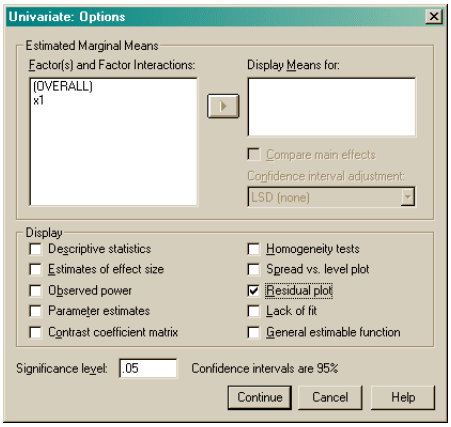
You should always plot the distribution of the **residuals** from any analysis — $y - \hat{y}$, or what's left over after you've predicted your dependent variable. Are the residuals normally distributed? If not, you should do something. Remember that an assumption of ANOVA was that the error variance was normally distributed (p. 9). If your residuals are not normally distributed, your p values don't mean what you hope. You can

- transform your dependent variable (p. 35)
- add another predictor (p. 51)

Why might non-normal residuals suggest that adding another predictor would be a good idea? Well, normal (Gaussian) residuals are what you'd expect if 'error' was in fact made up of a load of independent things of roughly equal importance (e.g. measurement error, room temperature fluctuations, background noise variations, time of day variations, subject alertness…); remember that the central limit theorem tells us that the distribution of a sum of a set of identically distributed random variables approaches the normal distribution. For a given standard deviation, the normal distribution has the *maximum uncertainty* (in information-theoretic terms, conveys

the maximum information). So normally distributed residuals help to suggest that there's no other major non-normally-distributed predictor you should add in.

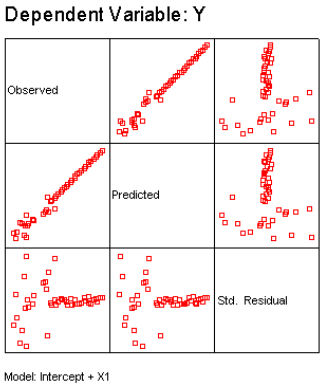
In SPSS, you can choose **Options → Residual plot** for any ANOVA:



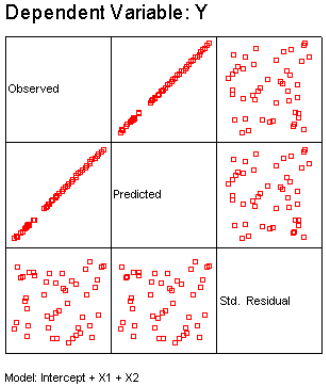
The three types of plot you get are:

- observed Y values (y) against predicted Y values ($\hat{y}$) — there'll be a correlation if your model is any good;
- observed Y values (y) against residuals ($y - \hat{y}$) — there'll be a correlation, since the two aren't independent (since $y = \hat{y} + \text{residual}$);
- predicted Y values ($\hat{y}$) against residuals ($y - \hat{y}$) — the two should be independent.

These plots are not terribly helpful. SPSS uses standardized residuals in these plots. (A *standardized residual* is a residual divided by an estimate of its standard deviation.) In SPSS's output there are two copies of each plot — one arranged with one variable on the x axis and the other on the y axis, and the other flipped (mirrored around the $x = y$ line). Here I've faked some data where Y depends on two other variables X_1 and X_2 (both continuous, for this example, i.e. covariates, but they could equally be factors), which happen themselves to correlate. This is what SPSS's residual plots look like:

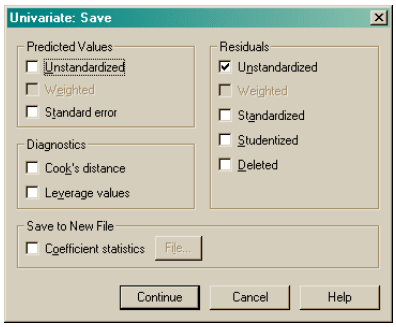


Not a very useful plot. Here, only X_1 is used as a predictor variable (model: Y = constant + b_1X_1). There is a correlation between observed and predicted Y values, meaning that our model is doing some predicting. The main thing to look at is the 'standardized residual' versus 'predicted Y' plot. Does that look like a random scatterplot? No. That suggests there's a further relationship between X_1 and Y that's not captured by a linear relationship.

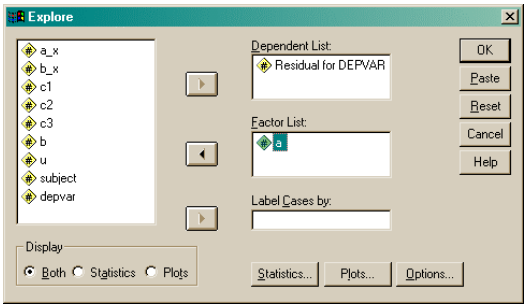


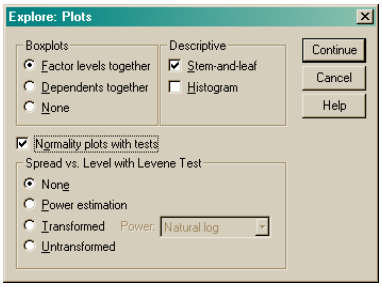
Not a very useful plot. When we improve our model by including X_2 as a predictor (model: Y = constant + b_1X_1 + b_2X_2), the 'standardized residual' versus 'predicted Y' plot looks more like a scatterplot. Whatever part of Y is not predicted (the residual) now appears to be uncorrelated with the predicted part, which is good — our model is doing a better job.

However, this residual analysis is not ideal — it doesn't give us a very clear indication of whether the residuals are normally distributed. What you can also do is to **save the residuals** from any ANOVA. Choose the **Save** dialogue box and choose the appropriate option, such as the unstandardized (raw) residuals:

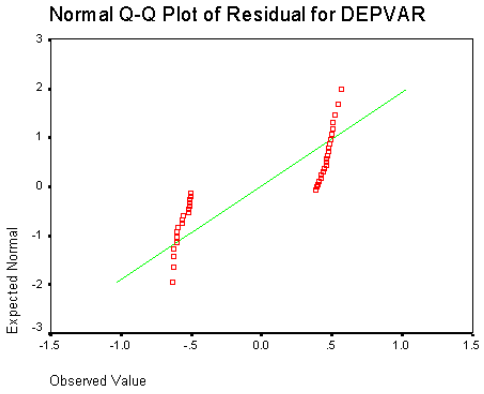


When the ANOVA is run, new column(s) are created with the residuals in. (If you run an ANOVA with within-subjects factors in the usual way using *Analyze → General Linear Model → Repeated Measures*, with one subject per row, you get one residual column for every data column in your input. This dialogue box can also be used to save the predicted values. In syntax, you can specify `/SAVE = PRED RESID` to get both.) Once you've obtained your residuals, you can check them for normality: **Analyze → Descriptive Statistics → Explore;** tick **'Both'** to get statistics and plots; click **Plots → Normality plots with tests.** This produces a Q–Q plot (if this produces a straight line, the data are normally distributed) and the Kolmogorov–Smirnov and Shapiro–Wilk tests (if significant, your residuals are non-normal); see p. 32 for explanation of these. To examine the residual distribution for several groups separately, enter the grouping factor into the **Factor List:**

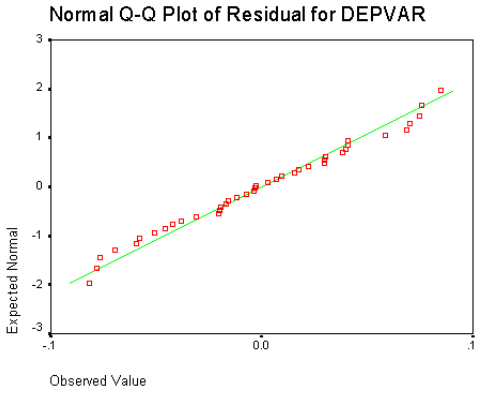




As an example, I created some data in which the data were created from the sum of contributions from factor A (two levels), factor B (two levels), and random noise. If we analyse with an ANOVA that only has factor A in it, saving and plotting the residuals as described above, we get a Q–Q plot of the residuals that looks like the left-hand side of the figure below — clearly not normal. This might suggest to us that we need to include another predictor. If we now include factor B in the ANOVA and replot the new residuals, we get the right-hand version, in which the residuals are normally distributed. That meets the assumptions of the ANOVA, and we can feel a bit happier that we haven't 'left anything out' of the analysis.



Dependent variable was caused by factors A and B, but only factor A was entered into the analysis. Residuals are not normally distributed.



Factors A and B are both entered into the analysis. Residuals are normally distributed.

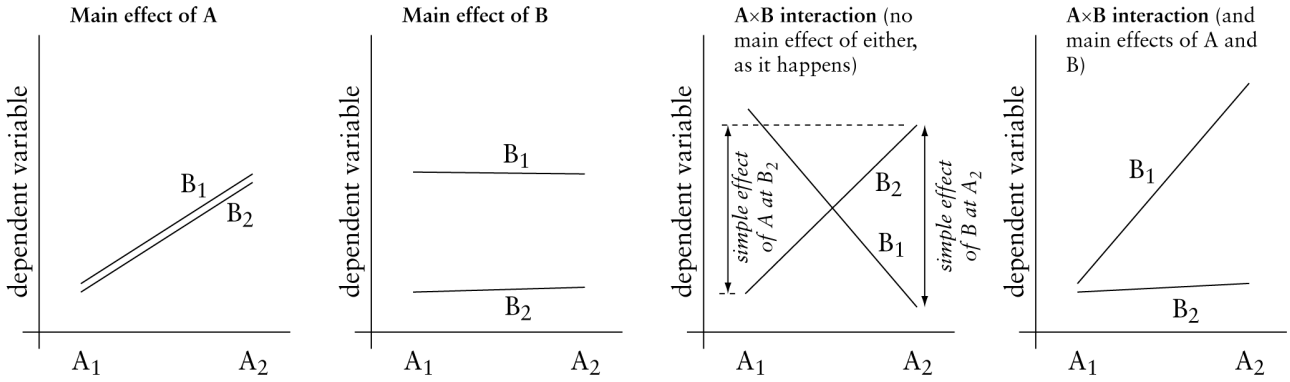
Finally, **residuals that are outliers (large)** for a group reflect data points that are outliers, so residual plots are another way to spot outliers (see also Myers & Well, 1995, p. 414).

3.5 Further analysis: main effects, interactions, post hoc tests, simple effects

Plot your data. With any reasonably complex experiment, you can't interpret the data until you've plotted it…

3.5.1. Simple effects

A reminder of what main effects, interactions, and simple effects refer to (see p. 20). It's easiest to visualize with a two-factor ANOVA. A **main effect** of A means that the A means (A_1, A_2, …A_a) are not all equal. Similarly for a main effect of B. An **interaction** means that the effects of A are not the same at all levels of B (equivalently, that the effects of B are not the same at all levels of A).



Suppose we have a between-subjects factor A (group: A_1 = control, A_2 = drugged) and a within-subjects factor U (task condition: U_1 = hot room, U_2 = cold room). We analyse our data and find an interaction. We may want to ask questions about **simple effects:** was there an effect of drug on performance in a hot room (simple effect of A at U_1, also written A/U_2)? Was there an effect of drug on performance in a cold room (A/U_2)? Was there an effect of room temperature on the control group (simple effect of U at A_1, written U/A_1)? On the drugged group (U/A_2)?

There are two ways of running simple effects analysis. The first and simplest is **only to analyse the data that's relevant.** So to ask about A/U_2, we'd only analyse the U_2 (cold room) data, this time with a one-factor ANOVA — we've dropped out the U factor. Similarly, if we had started with a three-way ANOVA (A × B × C), we would have run a two-way ANOVA to establish effects such as A/C_1, B/C_1, and

$A \times B / C_1$ (the last one is sometimes called a 'simple interaction'). This is easy and generally recommended (Myers & Well, 1995, p. 304). It can be applied to between- and within-subjects factors.

It is possible to obtain a more powerful test of the simple effects. This involves calculating the MS for the simple effect just as before, but testing it not against the MS_{error} for the sub-analysis (the one-factor ANOVA in our A × U example), but against the MS_{error} for the *original*, full ANOVA — known as the **pooled error** term. If you want to do this, you have to do it by hand: $F_{df\text{-factor}/df\text{-pooled-error}} = MS_{factor}/MS_{pooled\ error}$. Similarly, you can use the pooled error term for multiple comparisons between treatment means, if your factors have >2 levels. **However,** you shouldn't do this for within-subjects simple effects, as corrections for violations of the sphericity assumption are inadequate (Howell, 1997, p. 468). Furthermore, if there is some heterogeneity of variance, there can also be substantial problems (Myers & Well, 1995, pp. 134-136, 304-305). So it's simplest and probably best to ignore this technique — just run a simpler ANOVA on a subset of your data.

3.5.2. Determining the effects of a factor with >2 levels

If you discover that you have a significant main effect of a factor A with 2 levels, you know what it means: $\mu_{A_1} \neq \mu_{A_2}$. So you only have to look at the means to work out if $\mu_{A_1} > \mu_{A_2}$ or $\mu_{A_1} < \mu_{A_2}$. But if you have five levels, a significant main effect merely means that the null hypothesis

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

has been rejected. But what does that mean? There are all sorts of alternatives:

$$\mu_1 = \mu_2 = \mu_3 = \mu_4 \neq \mu_5$$
$$\mu_1 = \mu_2 = \mu_3 \neq \mu_4 = \mu_5$$
$$\mu_1 = \mu_2 = \mu_3 \neq \mu_4 \neq \mu_5$$
$$\ldots$$

This is where we would use *post hoc* comparisons among treatment means. There are two *types* of *post hoc* tests. One kind tests all possible **pairwise comparisons.** For 5 levels, we can compare μ_1 and μ_2, μ_1 and μ_3… up to μ_4 and μ_5. For 5 comparisons, there are ${}^5_2C = 10$ possible pairwise comparisons. The other type of test groups the means into **homogeneous subsets,** and tells you something like 'μ_1, μ_3, and μ_4 fall into one subset [are all the same]… μ_2 and μ_5 fall into another subset [are the same]… the subsets differ from each other'.

But we must be careful.

3.5.3. Post-hoc tests: the problem

The upshot: if you collect your data, look at it, and wonder 'are those two points significantly different?', you need to use a *post hoc* test — because your eye has already selected particular points to compare, which influences the likelihood of finding a 'significant difference'…

It's beautifully put by www.statsoft.nl/textbook/stglm.html: "Sometimes we find effects in an experiment that were not expected. Even though in most cases a creative experimenter will be able to explain almost any pattern of means, it would not be appropriate to analyse and evaluate that pattern as if one had predicted it all along. The problem here is one of capitalizing on chance when performing multiple tests *post hoc*, that is, without *a priori* hypotheses. To illustrate this point, let us consider the following 'experiment'. Imagine we were to write down a number between 1 and 10 on 100 pieces of paper. We then put all of those pieces into a hat and draw 20 samples (of pieces of paper) of 5 observations each, and compute the means (from the numbers written on the pieces of paper) for each group. How likely do you

think it is that we will find two sample means that are significantly different from each other? It is very likely! Selecting the extreme means obtained from 20 samples is very different from taking only 2 samples from the hat in the first place, which is what the test via the contrast analysis [known as an *a priori* test or planned contrast] implies. Without going into further detail, there are several so-called *post hoc* tests that are explicitly based on the first scenario (taking the extremes from 20 samples), that is, they are based on the assumption that we have chosen for our comparison the most extreme (different) means out of *k* total means in the design. Those tests apply 'corrections' that are designed to offset the advantage of *post hoc* selection of the most extreme comparisons. Whenever one finds unexpected results in an experiment one should use those *post hoc* procedures to test their statistical significance."

In general, we can define the *per-test* Type I error rate (α, also called the error rate per contrast) and the *familywise* Type I error rate (α_{FW}), the probability of making *at least one* Type I error rate when performing a 'family' of multiple comparisons.

3.5.4. The special case of three groups: multiple t tests are OK

There's a special case in which multiple uncorrected *t* tests are OK — when you have three groups (Howell, 1997, p. 370) *and you have a significant main effect* for your factor. This isn't widely appreciated. The ANOVA *F* test assesses the null hypothesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

If we have a significant main effect, then we've already rejected this null hypothesis. That means that one of the following must be true:

$$\mu_1 \neq \mu_2 = \mu_3$$
$$\mu_1 = \mu_2 \neq \mu_3$$
$$\mu_1 \neq \mu_2 \neq \mu_3$$

If we run a complete set of (3) uncorrected *t* tests, we will choose one of these three conclusions. But no conclusion involves us judging that there are more than two inequalities (significant differences between individual means). And we know that there is at least one inequality, since we've rejected the overall null hypothesis. So we can make *at most one Type I error*. Therefore, the probability of making that Type I error (choosing $\mu_1 \neq \mu_2 \neq \mu_3$ when one of the other two is correct) is the plain α for each test, and no further correction is necessary.

3.5.5. Otherwise… a variety of post hoc tests

For between-subjects factors, SPSS provides too many options in its *Post Hoc* box:

Equal variances assumed

- ***LSD** (Fisher's least significant difference). Uncorrected multiple *t* tests, except that the test is only performed when an ANOVA has rejected the overall null hypothesis, i.e. shown that 'something's going on' (Myers & Well, 1995, p. 188; Howell, 1997, p. 369-370). $\alpha_{FW} = 1 - (1- \alpha)^k$ when *k* independent tests are performed, and $\alpha_{FW} \leq 1 - (1- \alpha)^k$ when the tests are not independent (Myers & Well, 1995, p. 177). Only suitable for ≤3 levels of a factor — in which case it's the most powerful test — but **don't** use it otherwise.
- ***Bonferroni *t* procedure.** Occasionally called the Dunn procedure. Makes use of the Bonferroni inequality: $\alpha_{FW} \leq k\alpha$, or more generally, $\alpha_{FW} \leq \sum_i \alpha_i$ where α_i is the probability of a Type I error for the *i*th contrast (Myers & Well, 1995, p. 179). This is derived from the 'proper' version, $\alpha_{FW} \leq 1 - (1- \alpha)^k$, by noting that for small values of α (and 0.05 is small), $(1-\alpha)^k \approx 1 - k\alpha$. Therefore, each contrast is tested at $\alpha = \alpha_{FW} / k$. For example, if four tests are to be performed ($k = 4$) and we desire $\alpha_{FW} = 0.05$, then each test is per-

formed at $\alpha = 0.0125$. Quick to do. Additionally, we don't have to have all the αs equal. If we're much more interested in one of our four comparisons, we could allocate $\alpha = 0.03$ to it, and $\alpha = 0.0067$ to each of the others (Myers & Well, 1995, p. 181). Can be used for testing k planned contrasts.

- *†**Šidák (or Dunn–Šidák, or Sidak).** Since $\alpha_{FW} = 1 - (1- \alpha)^k$, this procedure solves for α [$\alpha = 1 - (1- \alpha_{FW})^{1/k}$] so as to get α_{FW} to be what you want (typically 0.05). Like the Bonferroni correction, but more accurate (i.e. it's correct). See also Howell (1997, p. 364).
- *†**Scheffé.** See Myers & Well (1995, p. 183) and Howell (1997, p. 379). Controls α_{FW} against all possible linear contrasts (see p. 75), not just pairwise contrasts. Consequently, very conservative.
- †**REGWF (Ryan–Einot–Gabriel–Welsch *F* test).** No idea; somehow similar to the REGWQ.
- †**REGWQ (Ryan–Einot–Gabriel–Welsch) range test.** A compromise between the Newman–Keuls (liberal) and Tukey HSD (conservative) (Howell, 1997, p. 378). This test does not require the overall F for groups to be significant as it controls the familywise error rate independently and test different hypotheses from the overall ANOVA, with different power (Howell, 1997, p. 351). Recommended (Howell, 1997, p. 378) except for unequal cell sizes (SPSS help).
- †**SNK (Student–Newman–Keuls, a.k.a. Newman–Keuls).** Not often used. Poor control of α_{FW} (Myers & Well, 1995, p. 188; Howell, 1997, p. 372-377) unless there are only three means to be compared, in which case it's OK.
- *†**Tukey HSD (honestly significant difference).** Similar to the Newman–Keuls test except that it fixes α_{FW} properly (Howell, 1997, p. 377).
- †**Tukey's-b.** Tukey's test as a range test? Not sure.
- †**Duncan's multiple range test.** Not often used. Poor control of α_{FW} (Myers & Well, 1995, p. 188).
- *†**Hochberg's GT2.** Less powerful variant of Tukey's; see SPSS help.
- *†**Gabriel's pairwise comparisons test.** 'A more powerful version of Hochberg's GT2 when cell sizes are unequal; may become liberal when the cell sizes vary' (SPSS help).
- †**Waller–Duncan *t* test.** 'Uses a Bayesian approach. Uses the harmonic mean of the sample size when the sample sizes are unequal' (SPSS help). That doesn't tell you much.
- **Dunnett's test for comparing treatment groups with a control group.** Sometimes we are interested in comparing each of the a–1 treatment groups to a control group, and less interested in comparing them to each other. For this case, since no two of the set of contrasts are orthogonal, the Bonferroni approach would be conservative (see pp. 76, 77). This test does not require the overall F for groups to be significant as it controls the familywise error rate independently and test different hypotheses from the overall ANOVA, with different power (Howell, 1997, p. 351).

Equal variances not assumed

- *†**Tamhane's T2.** 'Conservative pairwise comparisons, based on a *t* test' (SPSS help).
- *†**Dunnett's T3.** No idea. Range test.
- *†**Games–Howell.** 'Sometimes liberal' (SPSS help).
- *†**Dunnett's C.** No idea. Range test.

* Pairwise comparison test.
† Homogeneous subset test.
A range test is one based on a Studentized range statistic q, a modification of the t statistic (Howell, 1997, p. 370-372).

The important tests are summarized by Myers & Well (1995, p. 186). You can do most of what you want with the **Sidak** correction for pairwise comparisons, **Dunnett's test** when you're comparing treatment groups to a control group, and perhaps the **REGWQ** as a homogeneous subset test.

Pick your *post hoc* tests in advance: it is not valid to run all sorts of tests and then pick the 'most significant'. I suggest uncorrected *t* tests (Fisher's LSD) for three groups, the Sidak correction for >3 groups, and Dunnett's test for comparing treatment groups to a control group if you're more interested in that comparison than in differences between the treatment groups. If you would like a homogeneous subset test, then the Tukey HSD test is popular but the REGWQ is perhaps better. **Tukey's HSD, REGWQ, Dunnett's, and the Sidak test don't even require the overall *F* test from the ANOVA to be significant** (Howell, 1997, pp. 351, 364, 377), although the 3-group Fisher LSD does.

SPSS doesn't let you perform many of those tests on within-subjects factors, for good reason — many of them aren't valid (see Howell, 1997, p. 471). However, you can choose 'Display means for…' in the *Options* box and tick 'Compare main effects' with either no correction for multiple comparisons (LSD) — only valid if the factor has only 3 levels — or a Bonferroni or Sidak correction. The facility to compare means with a Sidak correction and to run further ANOVAs on subsets of your data is enough to analyse any between / within-subjects design, unless you also want to run specific contrasts (see p. 75).

3.6 Drawing pictures: error bars for different comparisons

Much of this is reproduced from www.pobox.com / ~rudolf / psychology, except the section on ANOVA.

3.6.1. Error bars for t tests: between-subjects comparisons

In brief:

- The standard error of the mean (SEM) conveys the precision with which the population mean was estimated. (It depends on the SD and the sample size, *n*.) Every mean (e.g. every group) has its own SEM.
- It is appropriate to use it as an error bar for between-subjects comparisons.
- It is the most common error bar you see published.
- The convention is to plot ±1 SEM — that is, your error bars extend above the mean by 1 SEM and below the mean by 1 SEM.

- Alternatives include the standard deviation (SD), which measures the variability of observations about their mean and is independent of *n*, and confidence intervals (CI); these show the range of values within which the population mean probably lies, and depend on the SD and *n*.

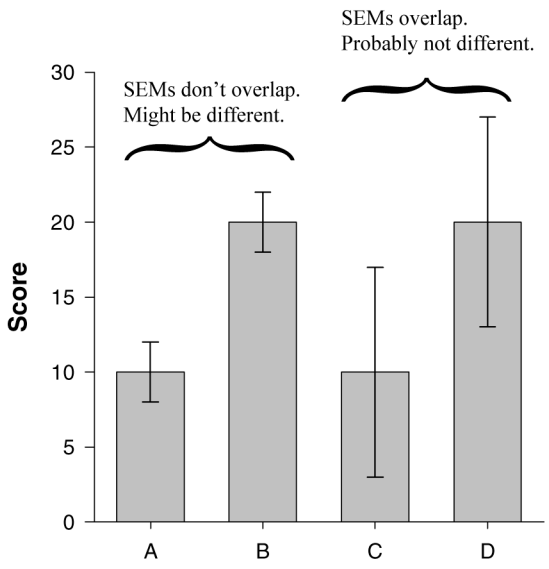
The SEM is frequently used as an index of variation when people publish data. They may quote a measurement of '25.4 ± 1.2 g', or display a datum on a graph with a value of 25.4 units and error bars that are each 1.2 units long. These 'variation' indices could be one of several things — mean ± SD, mean ± 95% CI, mean ± SEM… The paper should state somewhere which one is being used, but usually it's the SEM. Why? First, it's smaller than the SD, so it conveys an impression of improved precision (remember that **accuracy** is how close a measurement is to a 'true' value and **precision** is how well it is defined; thus, 2.5000000003×10^8 m·s^{-1} is a more precise but far less accurate measurement of the speed of light than 3.0×10^8 m·s^{-1}). In fact, using the SEM is perfectly fair and correct: the precision of an estimator is generally measured by the standard error of its sampling distribution (Winer, 1971, p. 7). Secondly — more importantly — if the SEM error bars of two groups overlap, it's very unlikely that the two groups are significantly different. (This is explained somewhat in the figure.) The opposite isn't necessarily true, though — **just because two sets of error bars don't overlap doesn't mean they are significantly different** (they have to 'not overlap' by a certain amount, and that depends on the sample size, and so on).

3.6.2. Error bars for t tests: within-subjects comparisons

In brief:

- SEMs are misleading for within-subjects comparisons. Use the standard error of the difference (SED) for the relevant comparison instead.
- SEDs are also appropriate for between-subjects comparisons.
- SEDs are not 'attached' to a particular mean, so the convention is to plot a 'free-floating' error bar that is 2 SED long, and label it. (The reader can use it as a mental ruler to make comparisons between the relevant means.)

For **within-subjects** comparisons, SEMs calculated for each condition are highly misleading (see figure below). For this comparison — indeed, for any comparison — the SED is an appropriate index of comparison, because that's what the *t* test is based on (*t* = difference between means / SED). So **if the difference between two means is greater than twice the SED, *t* > 2.** And for a healthy *n*, *t* > 2 is significant at the two-tailed $\alpha = 0.05$ level (have a quick glance at your tables of critical values of *t*).

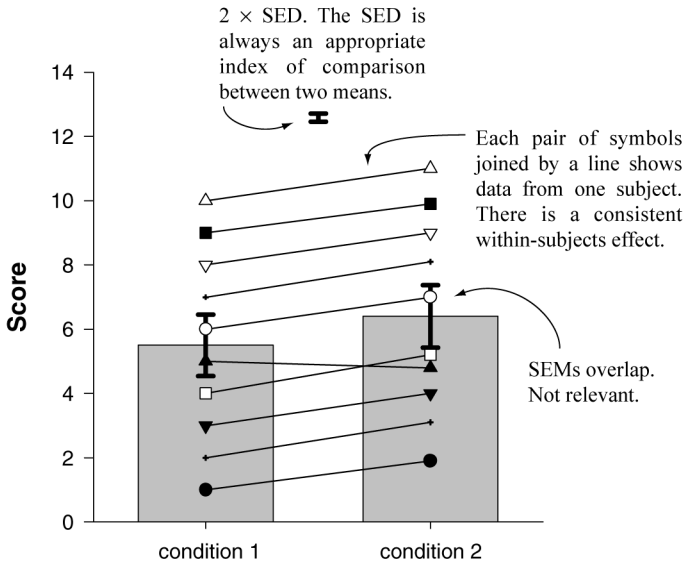


Unrelated groups
(between-subjects comparison)

Height of bar = mean
Error bar = ±1 SEM (1 SEM above, 1 SEM below the mean).

If the *n*s and SEMs of two groups are the same, then *t* = (difference between means) divided by (√2 × SEM). And if the SEMs of the two groups are the same and the SEMs overlap, then the means differ by <2 × SEM, so $t < 2 / \sqrt{2} = 1.4$. And $t < 1.4$ is never significant even at the 0.1 level.

So for independent groups, if the SEM error bars overlap, there's probably not a significant difference.



Related groups
(within-subjects comparison)

For within-subject comparisons, the SEM of each condition is not helpful. The vertical bars show group means; their error bars show ±1 SEM. You would think that the groups don't differ. But in fact, the same subjects were tested in condition 1 and condition 2. The subjects all scored very differently, but there is a consistent improvement from condition 1 to condition 2. If we ran a paired-sample *t* test on the difference scores, we would find a highly significant difference between the two conditions. The appropriate index of variation to compare the two conditions is the standard error of the difference between means (SED), shown at the top.

Another way of plotting these data would just be to plot the difference scores, with their SEM; readers could then visually compare that mean to zero. However, that would not show the baseline scores.

The SED is always an appropriate index of comparison; a *t* test is calculated as (difference between means) divided by (appropriate SED). But different comparisons require different SEDs. If your error bars don't convey the right impression, consider using SEDs (as in the top-right example; you could say "the error bar is 2 × the standard error of the difference for the comparison between …").

The SED is therefore a very good index of variation that can be used to make visual comparisons directly, particularly if you draw error bars that are 2SED long — if the means to be compared are further apart than the length of this bar, there's a good chance the difference is significant. However, it's a bit more work to calculate the SED, which is why you don't see it very often.

If you want to work out an SED, just choose the appropriate *t* test and calculate the denominator of the *t* test. For between-group comparisons where the group SEMs are SEM_1 and SEM_2, you'll see that $SED = \sqrt{(SEM_1^2 + SEM_2^2)}$.

To summarize, for within-subject changes:

1. The mean within-subject change equals the difference of the group means.
2. The variance of the within-subject change may differ greatly from the variance of any one condition (group).
3. Present within-subject changes when the baseline varies a lot, or you want to show variance of the within-subject measure.
4. Present group means when the baseline matters.

3.6.3. Error bars for an ANOVA

In brief:

- SEDs are always appropriate.
- Use $SED = \sqrt{\frac{2MS_{error}}{n}}$ if all groups are the same size.
- Use $SED = \sqrt{\frac{MS_{error}}{n_1} + \frac{MS_{error}}{n_2}}$ if there are two groups being compared and they are of unequal size.
- This means there may be a different SED for each comparison of two means. In SPSS, you can obtain these using pairwise comparisons for interaction effects (see p. 62). However, most people want to plot a 'single' SED. For this purpose, if there are >2 groups of unequal size, *I think* the most appropriate one to use is $SED = \sqrt{\frac{2MS_{error}}{\bar{n}_h}}$ where $\bar{n}_h$ is the harmonic mean of the group sizes (see p. 213 and also p. 70). For two groups, that reduces to the formula above.
- In an ANOVA with several factors, there may be are several different SEDs, corresponding to several different MS_{error} terms. Although you can plot the most relevant one(s), the most common convention is to plot the SED from the *highest interaction shown in your graph* (so if your graph shows factors A and B, you would plot the SED from the A × B interaction).
- The convention is to plot a 'free-floating' error bar that is 2 SED long, and label it as such.

A *t* test is directly related to an ANOVA: $F_{1,k} = t_k^2$ and $t_k = \sqrt{F_{1,k}}$. And a *t* test has this general formula:

$$t = \frac{\text{quantity}}{\text{standard error of that quantity}}$$

For example, a one-sample *t* test has the formula

$$t = \frac{\text{mean} - \text{test value}}{\text{standard error of the mean (SEM)}}$$

and a two-sample *t* test has the formula

$$t = \frac{\text{mean}_2 - \text{mean}_1}{\text{standard error of the difference between means (SED)}}$$

For a single sample, the SEM (the standard deviation of all sample means of a given sample size *n*) is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \text{ with corresponding variance } \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$$

For two independent samples, the SED (the standard deviation of the set of differences between pairs of sample means) is

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\mathrm{SEM}_1^2 + \mathrm{SEM}_2^2} \quad \text{with corresponding variance } \sigma^2_{\bar{x}_1-\bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

In an ANOVA with one factor and two groups, since we assume homogeneity of variance, our best estimate of the variances of two groups σ_1^2 and σ_2^2 is a weighted ('pooled') average of the two group variances (Myers & Well, 1995, pp. 65-66):

$$\hat{\sigma}^2_{\text{pooled}} = \frac{df_1\hat{\sigma}_1^2}{df_1 + df_2} + \frac{df_2\hat{\sigma}_2^2}{df_1 + df_2} = \frac{\mathrm{SS}_1 + \mathrm{SS}_2}{df_1 + df_2} = \frac{\mathrm{SS}_{\text{error}}}{df_{\text{error}}} = \mathrm{MS}_{\text{error}}$$

So $\mathrm{MS}_{\text{error}}$ is an approximation to σ^2. In fact, we knew that already (see pp. 9, 10). In general, the *standard error of an estimate* (Myers & Well, 1995, pp. 500-1), $\hat{\sigma}_e$, which estimates the standard deviation of the error variability ε, is

$$\hat{\sigma}_e = \sqrt{\frac{\mathrm{SS}_{\text{error}}}{df_{\text{error}}}} = \sqrt{\mathrm{MS}_{\text{error}}}, \text{ or } \hat{\sigma}_e^2 = \mathrm{MS}_{\text{error}}$$

and therefore for a comparison between two groups, the SED is given by

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\mathrm{MS}_{\text{error}}}{n_1} + \frac{\mathrm{MS}_{\text{error}}}{n_2}}$$

For equal group sizes, with n observations per group, this simplifies:

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{2\mathrm{MS}_{\text{error}}}{n}}$$

SPSS provides SEM and SED estimates for any given comparison when you choose *Options → Estimated Marginal Means* for a factor or set of factors, or if you use the `/EMMEANS = TABLES(factor)` syntax (see illustrated example on p. 56→). **But note that there is no 'one' SED appropriate for all comparisons.** If you have >2 groups, and their sizes are unequal, the SED for comparing group 1 to group 2 may be different for that for comparing group 1 to group 3. And in a multi-factor ANOVA, the SED for comparisons involving factor A will differ from the SED for comparisons between A × B subgroups. As we saw above, the convention is to plot the SED from the highest-order interaction.

3.7 Summarizing your methods: a guide for thesis-writing and publication

The following is an extract from my PhD thesis methods (which proved perfectly publishable: e.g. Cardinal *et al.*, 2003), with comments in square brackets.

> Data… were… analysed with [computer package, e.g. SPSS], using principles based on Howell (1997) [or other appropriate textbook]. Graphical output was provided by [computer package, e.g. Excel 97 and Sigmaplot 2001]. All graphs show group means and error bars are ±1 SEM unless otherwise stated.
>
> **Transformations.** Skewed data, which violate the distribution requirement of analysis of variance, were subjected to appropriate transformations (Howell, 1997, section 11.9). Count data ([e.g.] lever presses and locomotor activity counts), for which variance

increases with the mean, were subjected to a square-root transformation. Homogeneity of variance was verified using Levene's test.

Analysis of variance. Behavioural data were subjected to analysis of variance (ANOVA) using a general linear model, using SPSS's Type III sum-of-squares method. Missing values were not estimated but excluded from analysis [⇒ *subjects for whom some data were missing were omitted entirely; SPSS's default*]. All tests of significance were performed at $\alpha = .05$; full factorial models were used unless otherwise stated. ANOVA models are described using a form of Keppel's (1982) notation; that is, *dependent variable* = $A_2 \times (B_5 \times S)$ where A is a between-subjects factor with two levels and B is a within-subjects factor with five levels; S denotes subjects.

For repeated measures analyses, Mauchly's (1940) test of sphericity of the covariance matrix was applied and the degrees of freedom corrected to more conservative values using the Huynh–Feldt epsilon $\tilde{\varepsilon}$ (Huynh & Feldt, 1970) for any terms involving factors in which the sphericity assumption was violated.

[**Better approach, now I've learned more (see p. 25):** Degrees of freedom for terms involving within-subjects factors were corrected using the Greenhouse–Geisser epsilon $\hat{\varepsilon}$ (Greenhouse & Geisser, 1959) where the sphericity assumption was violated substantially ($\hat{\varepsilon} < 0.75$) or the Huynh–Feldt epsilon $\tilde{\varepsilon}$ (Huynh & Feldt, 1970) when the sphericity assumption was violated minimally ($\hat{\varepsilon} \geq 0.75$).]

[**Pretty good and simple approach (see p. 25):** Degrees of freedom for terms involving within-subjects factors were corrected using the Huynh–Feldt epsilon $\tilde{\varepsilon}$ (Huynh & Feldt, 1970).]

Thus, the same analysis with and without sphericity correction would be reported as follows:

Uncorrected: $F_{10,160} = 2.047$, $p = .032$

Corrected: $F_{4.83,77.3} = 2.047$, $\hat{\varepsilon} = 0.483$, $p = .084$

[Journals used to quibble about non-integer df because they were ignorant; such quibbling is less common these days. If you quote non-integer df, though, state the correction factor so people can work out the original df.]

Post-hoc tests. Significant main effects of interest were investigated using pairwise comparisons with a Sidak correction. This is based on the observation that $\alpha_{\text{familywise}} = 1 - (1 - \alpha_{\text{each test}})^n$ when n tests are performed; the correction was made such that $\alpha_{\text{familywise}} = .05$.

Where main effects were found for between-subjects factors with three or more levels, *post hoc* comparisons were performed with the REGWQ range test (familywise $\alpha = 0.05$), or Dunnett's test in situations where several experimental treatments were compared with a single control group. These tests do not require the overall F for groups to be significant as they control the familywise error rate independently and test different hypotheses from the overall ANOVA, with different power (Howell, 1997, p. 351). *[I was clearly rambling a bit here!]*

Where significant interactions were found following factorial analysis of variance, simple effects of a priori interest were calculated by one-way ANOVA and tested by hand against the pooled error term ($F = \text{MS}_{\text{factor}} / \text{MS}_{\text{pooled error}}$; critical values of F based on df_{factor} and $df_{\text{pooled error}}$). Multiple comparisons for simple effects were performed as described above but using the pooled error term.

Where significant interactions were found following repeated measures analysis, a pooled error term was used to test between-subjects simple effects of *a priori* interest, but separate error terms (i.e. plain one-way ANOVA) were used for within-subjects factors as sphericity corrections are inadequate if a pooled error term is used (Howell, 1997, p. 468). [*These days I wouldn't use the pooled error term at all, and would just use plain one-way ANOVA; see Myers & Well (1995, pp. 304-5).*]

Add any other special procedures you used! For example, you might add this:

… dependent variables were checked for normality by inspection of Q–Q plots (which plot scores against their expected values under a normal distribution) and using the Kolmogorov–Smirnov test with Lilliefors correction (Lilliefors, 1967) [and/or] Shapiro–Wilks test (Shapiro & Wilk, 1965).

Part 4: pitfalls and common issues in experimental design

4.1 Time in within-subjects (repeated measures) designs

There's nothing inherently special about 'time' as a within-subjects factor — you only get that impression from books that distinguish 'repeated measures' (implying time) from designs that are logically equivalent to within-subjects designs, e.g. in agriculture. As always, the sphericity assumption should be checked; time also represents a continuous factor, so that trend analysis (p. 80) involving it may be meaningful. And counterbalancing is often vital to avoid order effects. That's about it.

4.2 Analysis of pre-test versus post-test data

A very common design is as follows. Subjects are randomly assigned to groups (levels of A), such as A_1 and A_2. They are tested; the treatment (A_1 or A_2) is applied; they are retested. Since subjects were randomly assigned to groups, there are no systematic group differences in the pre-test scores. The post-test scores reflect the effects of the treatment. There are several ways to analyse this sort of design (Myers & Well, 1995, pp. 305-306, p. 454; also Howell, 1997, p. 606-7):

1. **Analysis of covariance** (p. 138). When its assumptions are met, this is the most powerful. Basically, this assumes that the post-test scores are linear functions of the pre-test scores. (It is often also assumed that the slopes of these functions are the same at each level of A, but see p. 138). The analysis takes advantage of this relationship, reducing error variability in the post-test scores by removing variability accounted for by the pre-test scores.

2. **Analysis of difference scores.** For each subject, the pre-test score is subtracted from the post-test scores; a one-factor ANOVA (using factor A) is then performed on these scores. The approach assumes that the effect of each treatment is to add a constant to the pretest score. Because this model is less likely to be true than that assumed by the analysis of covariance, it will generally be a less powerful test.

3. **Analysis of post-test scores only.** This approach is valid, but ignores information (the pre-test scores) that could help to reduce error variance, and therefore will be less powerful than those above.

4. **Analysis using a mixed design: A as a between-subjects factor, P as pre-test versus post-test.** This is frequently done. However, it will be a very conservative test of the main effect of A — it doesn't take account of the information that the pre-test scores cannot be affected by A. A better test for A would be that given by the A × P interaction, which is identical to that obtained by performing a one-way ANOVA on the difference scores — and as we saw above, an analysis of covariance is generally better.

If the subjects haven't been randomly assigned to levels of A, then the analysis (or the interpretation) can be much more difficult. If you don't understand the principles of multiple regression with correlated variables, don't go there — just analyse the post-test scores (Myers & Well, 1995, p. 306). Or understand the tricky stuff (Parts 6 & 7)…

4.3 Observing subjects repeatedly to increase power

An example: low-*n* experiment where subjects are precious. The dependent variable is change in blood pressure in response to a conditioned stimulus (CS). Two CSs are used: one signalling a high-incentive, tasty food, and the other signalling a low-

incentive, less-preferred food. Furthermore, subjects are tested following administration of a drug or placebo. The response of each subject to each CS is observed 6 times, to reduce the measurement error or increase power somehow (the experimenter feels that more observations should give more power, but can't verbalize exactly how). Presentation order is suitably counterbalanced. The original data layout is shown below. How should this be analysed to maximize power?

Subject (S)	Incentive (A)	Drug (B)	Observation (C)	Dependent variable
1	Low	Placebo	1	*datum*
1	Low	Placebo	2	*datum*
1	Low	Placebo	3	*datum*
1	Low	Placebo	4	*datum*
1	Low	Placebo	5	*datum*
1	Low	Placebo	6	*datum*
1	High	Placebo	1	*datum*
1	High	Placebo	2	*datum*
1	High	Placebo	3	*datum*
…	…	…	…	…
1	Low	Drug	1	*datum*
1	Low	Drug	2	*datum*
…	…	…	…	…
1	High	Drug	1	*datum*
1	High	Drug	2	*datum*
…	…	…	…	…
2	Low	Placebo	1	*datum*
2	Low	Placebo	2	*datum*
…	…	…	…	…
3 subjects	2 levels	2 levels	6 observations per level	72 observations

We have these possible factors, even if we do not use them all: subject (S), which is a random factor (see p. 31); incentive (A), which is a fixed factor; drug (B), which is a fixed factor; perhaps observation (C), which we'll consider to be a fixed factor. We seek to test the effects of A (does the response to a high incentive CS differ from that to a low incentive CS?), B (does the response of a drugged subject differ from that of a non-drugged subject?), and A × B (does the effect of incentive alter as a result of the drug?) with maximum power.

Consider the options:

1. A and B are used as factors. No 'subject' term is entered, so it's effectively a between-subjects design. **Wrong.** This is pseudoreplication; we are pretending that we have 18 independent observations per AB combination. In fact, we have 3 subjects per AB combination with 6 observations per subject — and those observations are likely to be correlated, because they come from the same subject. We must take account of this fact. Indeed, to do so is likely to improve our power, by accounting for differences between subjects. Remember the key assumption of ANOVA: that the error components (ε) are *independent*.

2. A, B, and S are used as factors. This is a design with two within-subjects factors. There are 6 observations per 'cell' (per ABS combination). We are assuming that there is no correlation between observations *beyond* that attributable to them coming from the same subject / A / B combination. Somewhat related to within-subjects ANCOVA (Bland & Altman, 1995a) (see p. 152). **Valid.**

3. A, B, C, and S are used as factors. This is a design with three within-subjects factors. We have a factor of 'observation number'. This may mean very little to us (we wouldn't be interested in effects attributable to it), but we include it in the hope that it removes some variability, reducing our error variability and improving our power. We have one observation per cell. **Valid.**

4. We take the mean of the 6 observations per subject per AB combination. We now have $N = 12$ observations rather than $N = 72$, but we expect the means to

be more accurate estimators of the true effect on each subject. We analyse them with A, B, and S as factors. We have one observation per cell. **Valid.**

So of designs 2–4, which is optimal? They'll all give **identical** answers! Observing a subject more than once in the *same* condition simply improves the precision with which the subject is measured in that condition. You can use that more precise mean directly (design 4), or let the ANOVA maths work out the means for each condition (designs 2 and 3). The variability that you reduce by measuring the subject repeatedly is the variability *about the mean for that subject in that condition*, not the variability associated with measuring the effect of factors A or B. Try it and see.

See also the *CRD with subsampling* and *RCB with subsampling* agricultural designs (p. 186→).

4.4 'It's significant in this subject…'

Words to strike fear into your heart. The scenario runs like this. An experimenter using precious subjects assigns them to sham or lesion groups. Each is measured repeatedly for its response to a stimulus paired with food (CS^+) and to a neutral stimulus (CS^0). Let's say we have ten CS^+ observations and ten CS^0 observations per subject.

It is, of course, completely valid to perform a *t* test or equivalent ANOVA to ask whether the effect of CS (CS^+ versus CS^0) is significant *for that subject*. Note that you might use an *unpaired ('between-subjects')* analysis, since the CS^+ data and the CS^0 data are not related *beyond* the fact that they come from the same subject (which is now your experimental 'universe') — unless there is some further factor that pairs data points within each subject. (One such factor might be 'trial pair', if one trial pair has one CS^+ and one CS^0 presented close to each other in time.) However, the conclusions of such a test **apply only to that subject.** You could *not* generalize it to others ('subjects in general with such-and-such a lesion').

I've seen arguments that run like this: "We compared a CS^+ and a CS^0 for each subject to obtain a measurement of CS reactivity [a single number per subject]. We compared these CS reactivity scores pre-operatively and post-operatively. The lesion significantly reduced CS reactivity scores in 2 out of 4 lesioned subjects [note within-one-subject significance tests]. None of the 4 sham-operated subjects showed a significant change in CS reactivity scores." The implication that one is presumably meant to draw is the lesion reduced CS reactivity. There are at least two fallacies here:

- The significance tests for individual subjects don't tell you that the change was significant for a *group*.
- (Ignoring the previous point for a moment…) 'The change in reactivity scores was significant for group A but not for group B; therefore group A differed from group B.' This is a common statistical fallacy. There might have been a decrease in scores for one group ($p = 0.04$) but not the other ($p = 0.06$) — that does *not* mean that the two groups differed. That test would require examination of the lesion × (pre-post) interaction — or, better (as we saw above), an analysis of covariance with pre-operative scores as the covariate.
- Even if you used 'significant or not' as a dichotomy — and it would be an artificial dichotomy (using a criterion p value as a cut-off, rather than a genuine dichotomy such as sex; see Howell, 1997, p. 286), the test across groups would then be a χ^2 contingency test with two variables (sham versus lesion; changed versus unchanged). For this specific example, $\chi_1^2 = 2.67$, $p = 0.1$, NS.

4.5 Should I add / remove a factor? Full versus reduced models

Omitting relevant variables and including irrelevant variables can both alter your estimate of effects of other variables (Myers & Well, 1995, pp. 519-521). **Including irrelevant variables** isn't too bad — this doesn't bias the estimate of the proportion of variability accounted for by your other predictors, but it does use up error degrees of freedom, reducing the power to detect effects of other variables. **Omitting relevant variables** is worse; it can substantially bias the estimates of the effects of the other terms. As a simple example, suppose your data contain a main effect of A and a main effect of B, but no interaction. If you were to analyse these data using a model with just A and AB terms (and no B term), you've omitted a relevant variable, and you can get a 'spurious' interaction.

There are various formal ways to work out the 'best' set of predictor variables to use if you have a lot of potential predictors (e.g. forward selection, backward elimination, and stepwise regression; see Myers & Well, 1995, p. 516-518), but they are primarily of use in descriptive (correlative, non-experimental) research and none of them removes the need to think carefully about your experimental design.

People commonly **neglect potentially important predictors** (Myers & Well, 1995, pp. 100-101, 149-151), such as who tested the subjects, because they're not of interest, or they weren't thought about. These are poor reasons. A good reason to remove a predictor from an ANOVA is that you have evidence that it isn't contributing to the prediction. If so, then by removing it, you may increase the power to detect other effects. A good rule is to **include all the potentially relevant predictors initially, and consider removing a term if (a) you have *a priori* reason to think it's irrelevant and (b) the term is not significant at the $\alpha = 0.25$ level** (Myers & Well, 1995, pp. 100-101, 151).

Note that a non-normal distribution of residuals (p. 36) may also suggest the need to add another predictor (or to transform the dependent variable).

For example, suppose we have a three-way ANOVA (factors A, B, and C). The experimenter is primarily interested in the effects of A and B. The analysis shows that none of the A×C, B×C, A×B×C terms are significant at the $\alpha = 0.25$ level, but the main effect of C is significant at $\alpha = 0.25$. The plan would then be to drop out those interactions, so you're left with A, B, A×B, and C.

Dropping out terms that are genuinely not contributing helps, because it increases the error *df* (which increases power); the *df* and any variability attributable to the term joins (is 'pooled with') the error *df* and error variability. You hope that the error *df* go up but the error variability doesn't — which should be the case if the term wasn't contributing to the prediction. But if your term is actually contributing, then pooling its variability as part of the error term also increases the *E*(MS) of the error term, negatively biasing all your other *F* tests — making it *less* likely that you'll detect other effects that you're interested in (Myers & Well, 1995, pp. 149-151).

This argument also applies to the experimental design technique of **blocking** (Myers & Well, 1995, pp. 157-158). Suppose we want to test the effect of different types of teaching method (A) on reading skill (Y) in children, and subjects are randomly assigned to the levels of A. If there is considerable individual variation (variability among subjects within groups — the error term for the ANOVA) we may have low power to detect effects of A. One way to deal with this is to **block** the subjects. We would divide them into groups based on their performance on some variable, X (perhaps IQ?), that we believe to be highly correlated with Y. Suppose we used five blocks: block B_1 would contain the children with the highest X scores, block B_2 would have the next-highest X scores, and so on. Then we would randomly assign the members of block B_1 to the different A conditions. We have made our one-factor ANOVA into a two-factor ANOVA; we hope that this reduces the within-block inter-subject variability, and therefore increases the power to detect effects of A. In general, blocking is intended to reduce error variability (which increases power). Of course, it uses up error *df* (which reduces power). Therefore, to get the best power,

you should choose the number of blocks based on N, a, and the correlation (ρ) between X and Y (see Myers & Well, 1995, pp. 157-158).

4.6 Should I add / remove / collapse over levels of a factor?

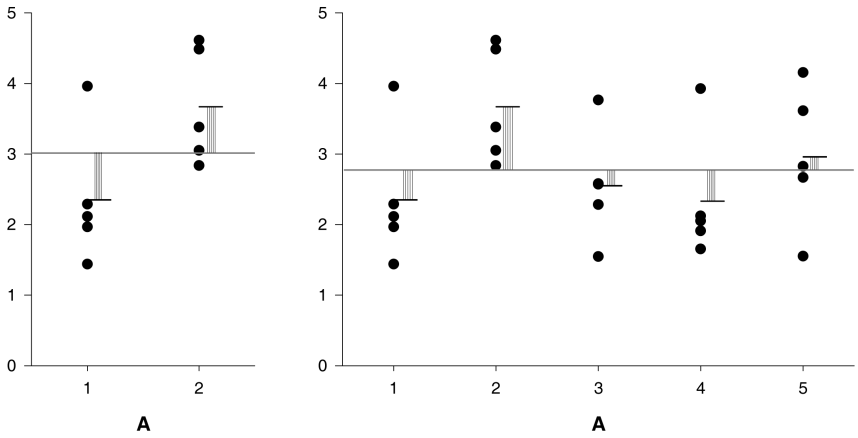
The key thing to remember is this:

$$F = \frac{\mathrm{MS}_{\text{predictor}}}{\mathrm{MS}_{\text{error}}} = \frac{\mathrm{SS}_{\text{predictor}} \times df_{\text{error}}}{\mathrm{SS}_{\text{error}} \times df_{\text{predictor}}}$$

The more levels a factor has, the larger its $df_{\text{predictor}}$, so on its own this will reduce the F statistic, and therefore the power to detect the effect of this factor. On the other hand, if adding a level increases $\mathrm{SS}_{\text{predictor}}$, power goes up. And, all other things being equal, adding more observations increases power because it increases df_{error}. Let's illustrate this with a couple of examples:

4.6.1. Adding and removing levels by adding new observations

Taking new observations at further levels of a factor can reduce power:

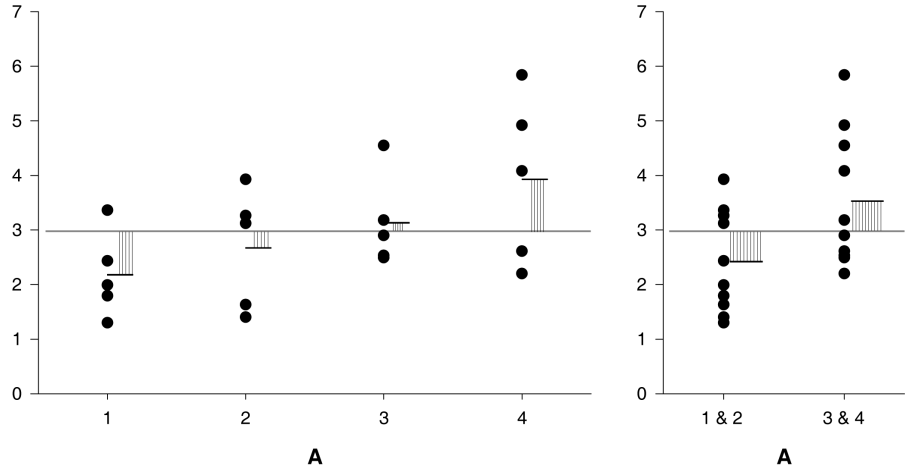


Left: *the dependent variable is measured at only two levels of A (n = 5 per group). There is a significant effect of A (MS_A = 4.349, MS_{error} = 0.793, $F_{1,8}$ = 5.486, p = 0.047).* ***Right:*** *three more groups have been measured. Even though the original data is unchanged, the effect of A is now not significant (MS_A = 1.581, MS_{error} = 0.808, $F_{4,20}$ = 1.958, p = .14).* ***Vertical lines*** *represent contributions to SS_A; specifically, SS_A is the sum of the squares of these vertical lines (deviations of group means from the overall mean).*

Equally, it's very easy to imagine a situation in which a non-significant effect with a few levels becomes a significant effect when subjects are measured at more levels — a very easy example would be a drug measured at 0 and 0.1 mg doses, where 0.1 mg is below the effective dose; if 10 and 100 mg doses are added to the study, the effect of the drug might emerge.

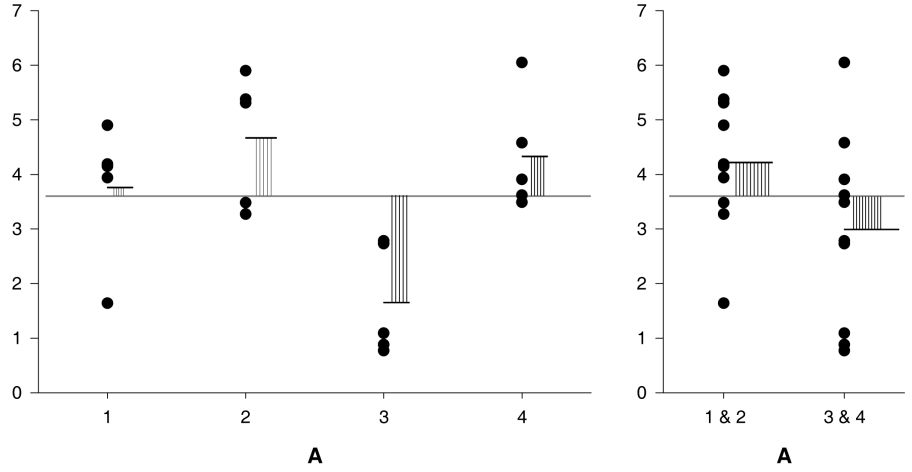
4.6.2. Collapsing over or subdividing levels

Collapsing over levels with similar means increases power:



Left: *there is not a significant effect of A (SS_A = 8.324, MS_A = 2.775, MS_{error} = 1.216, $F_{3,16}$ = 2.281, p = 0.118).* ***Right:*** *if we collapse over levels by combining levels 1 and 2, and levels 3 and 4, there is a significant effect of A (SS_A = 6.118, MS_A = 6.118, MS_{error} = 1.204, $F_{1,18}$ = 5.082, p = 0.037).*

But collapsing over levels can have the opposite effect, if you collapse over levels with dissimilar means:



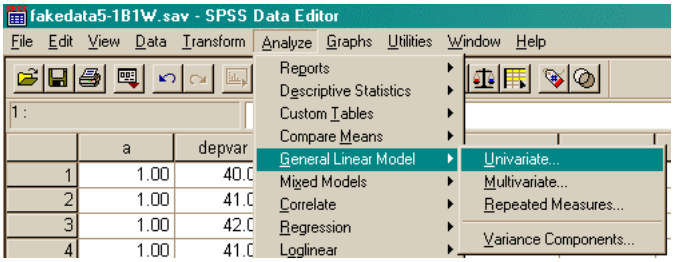
Left: *there is a significant effect of A (SS_A = 27.549, MS_A = 259.613, MS_{error} = 20.507, $F_{3,16}$ = 7.165, p = 0.003).* ***Right:*** *if we collapse over levels in the same way as before, we reduce the sum of squares and there is no longer a significant effect of A (SS_A = 7.522, MS_A = 7.522, MS_{error} = 2.252, $F_{1,18}$ = 3.34, p = 0.084).*

Part 5: using SPSS for ANOVA

5.1 Running ANOVAs using SPSS

5.1.1 Analysis of variance

You can perform ANOVAs from the **Analyze → General Linear Model** menu (below).



'Univariate' analyses a single dependent variable. It will easily handle between-subjects designs.

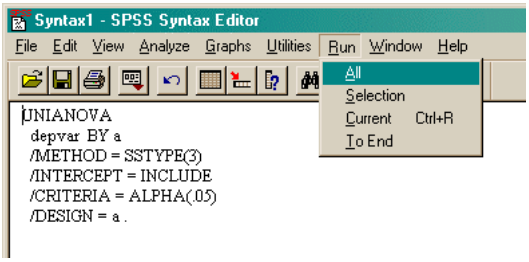
- Fill in your between-subjects factors as **fixed** factors and add any between-subjects covariates (by default these will not interact with any factors).
- It will also handle within-subjects designs if your data is in a 'one column, one variable' format — simply enter Subject as a random factor and enter all the 'real' factors as fixed factors. However, this way of doing within-subjects analysis may be slow and will not include Mauchly's test or the Greenhouse–Geisser or Huynh–Feldt corrections (explained above; see p. 25). Furthermore, it will get the analysis of *mixed* models (models that have both between-subjects and within-subjects factors) wrong unless you enter a custom model in the 'Models' dialogue box.

The easier way of analysing simple designs that include within-subjects factors is with the **Repeated Measures** option; this requires that your data is in a 'one row, one subject' format. This option also allows you to include between-subjects factors and between-subjects covariates.

The **'Multivariate'** option is used for analysing multiple dependent variables (multivariate analysis of variance: MANOVA), and we will not cover it.

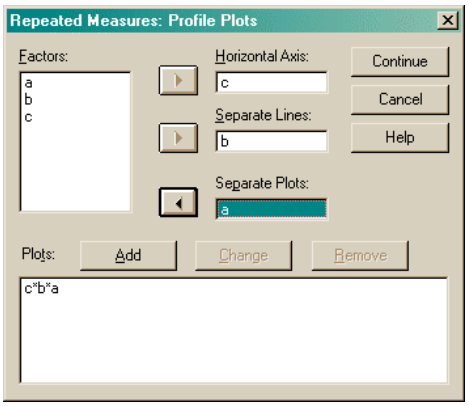
5.1.2 Syntax

Whenever you see an **'OK'** button to begin an analysis in SPSS, there will also be a **'Paste'** button that will not run the analysis, but will copy the syntax for the analysis into a syntax window (opening one if necessary). This allows you to edit the syntax, if you want to do something complicated; it also allows you to save syntax so that you can run large multi-step analysis time after time with the minimum of effort. You can even include syntax to load data from a file, or retrieve it from an ODBC-compatible database. The **Run** menu of a syntax window allows you to run all of a syntax file, or part that you have selected.



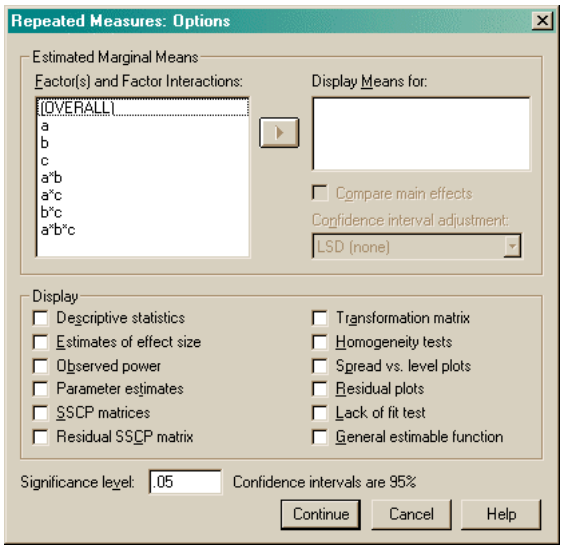
5.1.3 Plots

SPSS can produce sketch plots along with its ANOVA output. Click the **Plots** option of an ANOVA dialogue box and fill in the gaps. Click **Add** to add your plot to the list once you've assembled its components.

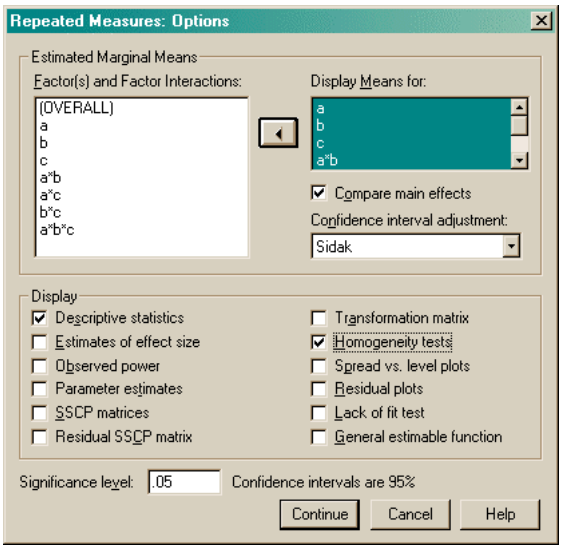


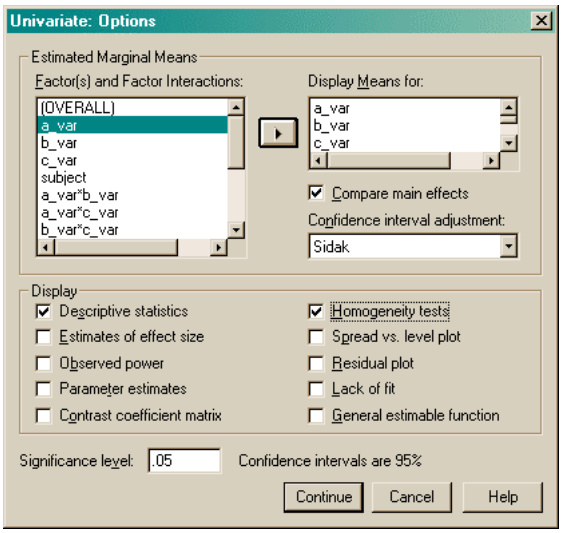
5.1.4 Options, including homogeneity-of-variance tests

All the ANOVA dialogue boxes also allow you to set **Options.** By default, no options are ticked:



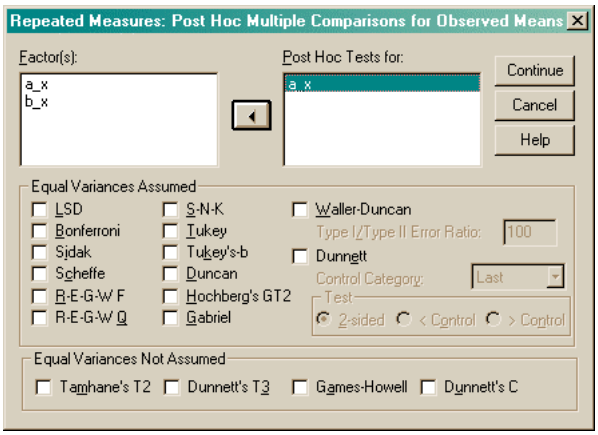
I find it useful to include descriptive statistics (including means and SEMs for all levels of factors and interactions). I tend reflexively to compare main effects using a Sidak correction. It's certainly worthwhile including **homogeneity tests** to check the assumptions of the ANOVA; SPSS will perform Levene's test for homogeneity of variance (significant = heterogeneous = a problem) if you tick this box. The options menu for the **'Univariate'** analysis looks slightly different:





5.1.5 Post hoc tests

SPSS will allow you to specify *post hoc* tests for between-subjects factors in the **'Post hoc'** dialogue box:



It won't allow you to specify post-hoc tests for within-subjects factors, **mainly because most *post hoc* tests are not suitable for use with within-subjects factors** (see Howell, 1997, p. 471). SPSS tries hard not to let you do something daft. The simplest and usually best thing to do is to run a separate within-subjects ANOVA for the data you want to perform a within-subjects *post hoc* test on.

5.2 Interpreting the output

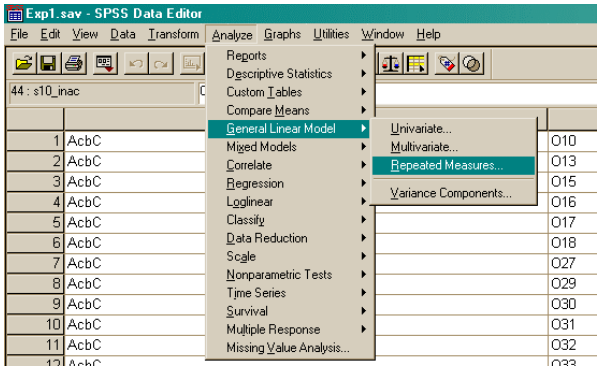
Let's look at a real and fairly complicated analysis. It involves four factors. Rats were received either lesions of the nucleus accumbens core (AcbC) or sham surgery. Each group was further divided into three (delay = 0, 10, or 20 s). All rats were placed in operant chambers with two levers present throughout each session. One lever (Inactive) did nothing. The other (Active) delivered a single food pellet. In the 'delay = 0' group, that pellet was delivered immediately. For the 'delay = 10s' group, the pellet was delivered after a 10 s delay, and for the 'delay = 20s' group, after 20 s. They were trained for 14 sessions each. These are our factors:

Factor	Between-subjects (B) or within-subjects (W)	Number of levels	Levels
Lesion	B	2	sham, AcbC
Delay	B	3	0, 10, 20 s
Lever	W	2	active, inactive
Session	W	14	1–14

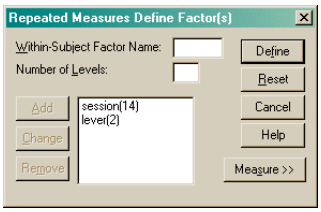
The data is entered into SPSS in 'one subject, one row' format (see p. 122), like this:

Subject	*Lesion*	*Delay*	*S1_Active*	*S1_Inactive*	*S2_Active*	*S2_Inactive …*
O1	sham	0	*datum*	*datum*	*datum*	*datum*
O2	sham	0	*datum*	*datum*	*datum*	*datum*
…						
O19	sham	10	*datum*	*datum*	*datum*	*datum*
…						
O48	AcbC	20	*datum*	*datum*	*datum*	*datum*

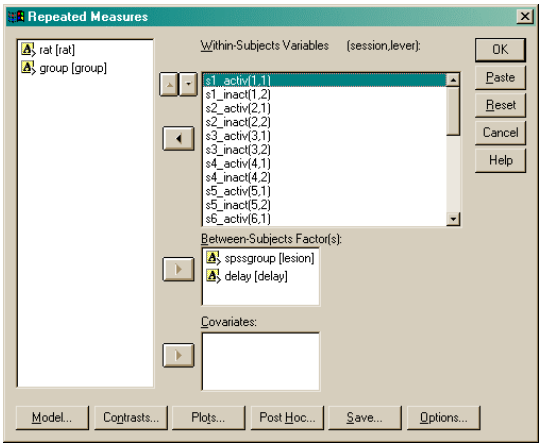
We have within-subjects factors and we have the data in one-subject-one-row format, so we choose **Analyze → General Linear Model → Repeated Measures:**



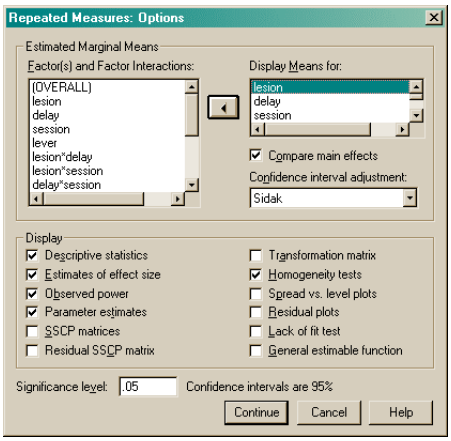
We declare the within-subjects factors:



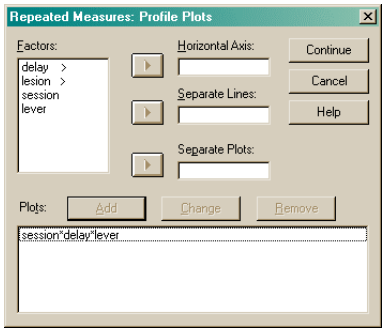
Now we fill in the between-subjects factors and assign individual columns to appropriate levels of the within-subjects factors:



It's important to ensure that the within-subjects level assignments are correct — so 's5_inact' is labelled as (5,2), and the dialogue box tells us that this refers to (session, lever) — so it's going to be level 5 of session and level 2 of lever. This is correct. So we proceed to set appropriate options. I'm going to tick loads of things so we can interpret a fairly full output:

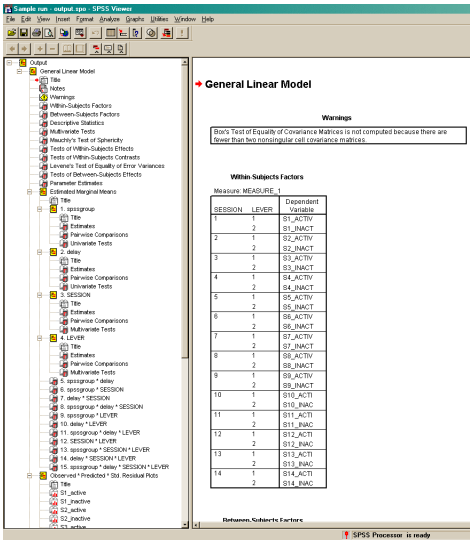


I wouldn't normally tick 'estimates of effect size', 'observed power', or 'parameter estimates'. We can also set up some plots:



SPSS doesn't do very good graphs, and it'll only plot three factors at once. So this plot has *session* on the horizontal axis, *delay* on separate lines, and *lever* on separate plots. (The data will be collapsed across *lesion*, which means this graph won't give us any indication of how the sham / lesion groups differed — not very helpful!)

OK. Now we could **Paste** the syntax for this command into the syntax editor to save it and / or fiddle with it, or just click **OK** to run the analysis. Let's run the analysis. We get a lot of stuff…



It's huge! Let's look at them one by one.

- **Title.** Says 'General Linear Model'.
- **Notes.** None.

- **Warnings.** Tells you what it couldn't do. Sometimes this information is helpful; here, it's not very comprehensible and we ignore it.

- **Within-subjects factors.** This tells you what you told it. It lists all your within-subjects factors and tells you which column of data has been matched to each level of the factor(s). If this is wrong, the rest of your analysis will be meaningless, so it's worth checking.

- **Between-subjects factors.** The same, but for between-subjects factors. It also gives you the number of subjects in each condition. Check this — it may not always be what you expect. If a subject has missing data somewhere, SPSS will default to chucking the subject out completely.

Between-Subjects Factors

		N
spssgroup	AcbC	20
	sham	24
delay	0	14
	10	15
	20	15

- **Descriptive statistics.** Since we asked for this in the *Options*, we get a long list of cell means:

Descriptive Statistics

	spssgroup	delay	Mean	Std. Deviation	N
S1_Active	AcbC	0	.8324	.43428	6
		10	.6586	.14757	7
		20	.5414	.29284	7
		Total	.6697	.31313	20
	sham	0	.3306	.23671	8
		10	.8046	.11218	8
		20	.6402	.26392	8
		Total	.5918	.28700	24
	Total	0	.5457	.41122	14
		10	.7364	.14598	15
		20	.5941	.27236	15
		Total	.6272	.29820	44
S1_Inactive	AcbC	0	.6297	.17563	6
		10	.5566	.26822	7
		20	.6829	.28929	7
		Total	.6227	.24541	20

- **Multivariate tests.** Ignore 'em; we're not analysing multiple dependent variables. We're analysing one (lever-pressing), predicted by four predictors (factors). So skip this.

- **Mauchly's test of sphericity.** For every within-subjects factor and interaction of within-subjects factors, SPSS performs Mauchly's test of sphericity. **If it's significant** ('Sig.' column = $p < 0.05$), then you should multiply your *df* by the Huynh–Feldt epsilon $\tilde{\varepsilon}$ listed by it. For example, the Session factor has violated the sphericity assumption and will have its *df* multiplied by $\tilde{\varepsilon} = 0.287$, while the Session × Lever interaction will have its *df* multiplied by $\tilde{\varepsilon} = 0.464$. The Lever factor has *not* violated the sphericity assumption. Sometimes you can tell because the 'Sig.' column has a *p* value that's >0.05. Here, there's no *p* value — but since $\tilde{\varepsilon} = 1$, we know that there's no problem anyway.

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
SESSION	.000	465.481	90	.000	.232	.287	7.692E-02
LEVER	1.000	.000	0	.	1.000	1.000	1.000
SESSION * LEVER	.000	288.316	90	.000	.355	.464	7.692E-02

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept+LESION+DELAY+LESION * DELAY
Within Subjects Design: SESSION+LEVER+SESSION*LEVER

- **Tests of within-subjects effects.** This is one of the **important** bits. There's a set of rows for every within-subjects factor, or interaction involving a within-subjects factor.

- There's a columns corresponding to the SS ('Type III Sum of Squares' — SPSS has a few ways of calculating the SS and you almost certainly want Type III, which is the default).
- It gives you the *df*. The top row ('sphericity assumed') gives you the normal *df*. The subsequent rows give you the *df* multiplied by the various correction factors listed in the results of Mauchly's test, including the Huynh–Feldt epsilon $\tilde{\varepsilon}$.
- The MS is the SS divided by the *df*.
- The *F* ratio is the MS for the term divided by the MS for the corresponding error term. It's always the same, no matter whether you use the Huynh–Feldt correction or not. For example, the *F* for Session (80.11) is the MS for SESSION (9.045, 38.99, 31.475, or 117.584, depending on the *df* correction) divided by the MS for 'Error(SESSION)' (.113, .487, .393, or 1.468, depending on the *df* correction).
- The 'Sig.' column is the *p* value for the *F* ratio, assessed on the relevant number of degrees of freedom. It may vary depending on whether or not you need to use the Huynh–Feldt correction.
- **In this example,** Lever doesn't require any correction, so we would report $F_{1,38} = 678$, $p < 0.001$ for the effect of Lever. However, Session requires a Huynh–Feldt correction, as we saw above, so we would report $F_{3.736,141.958} = 80.11$, $\tilde{\varepsilon} = 0.287$, $p < 0.001$. If you correct the *df*, it's good practice to report $\tilde{\varepsilon}$ so readers can work out the original *df* (which tells them something about your analysis).
- **Partial eta-squared** is a column that only appeared because we ticked *Estimates of effect size* in the Options. For details of η^2_{partial}, see p. 102.
- **Noncent(rality) parameter** and **Observed power** only appeared because we ticked *Observed power*. The observed power is the probability that the *F* test would detect a *population* difference between the two groups equal to that implied by the *sample* difference (SPSS, 2001, p. 476). The noncentrality parameter is used to calculate this (Howell, 1997, pp. 334-5).

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
SESSION	Sphericity Assumed	117.584	13	9.045	80.110	.000	.678	1041.429	1.000
	Greenhouse-Geisser	117.584	3.016	38.990	80.110	.000	.678	241.590	1.000
	Huynh-Feldt	117.584	3.736	31.475	80.110	.000	.678	299.270	1.000
	Lower-bound	117.584	1.000	117.584	80.110	.000	.678	80.110	1.000
SESSION * LESION	Sphericity Assumed	3.007	13	.231	2.049	.016	.051	26.633	.945
	Greenhouse-Geisser	3.007	3.016	.997	2.049	.111	.051	6.178	.515
	Huynh-Feldt	3.007	3.736	.805	2.049	.095	.051	7.653	.580
	Lower-bound	3.007	1.000	3.007	2.049	.161	.051	2.049	.287
SESSION * DELAY	Sphericity Assumed	8.216	26	.316	2.799	.000	.128	72.766	1.000
	Greenhouse-Geisser	8.216	6.031	1.362	2.799	.014	.128	16.880	.868
	Huynh-Feldt	8.216	7.471	1.100	2.799	.008	.128	20.910	.920
	Lower-bound	8.216	2.000	4.108	2.799	.073	.128	5.597	.518
SESSION * LESION * DELAY	Sphericity Assumed	2.046	26	7.870E-02	.697	.867	.035	18.124	.636
	Greenhouse-Geisser	2.046	6.031	.339	.697	.653	.035	4.204	.268
	Huynh-Feldt	2.046	7.471	.274	.697	.684	.035	5.208	.302
	Lower-bound	2.046	2.000	1.023	.697	.504	.035	1.394	.159
Error(SESSION)	Sphericity Assumed	55.776	494	.113					
	Greenhouse-Geisser	55.776	114.598	.487					
	Huynh-Feldt	55.776	141.958	.393					
	Lower-bound	55.776	38.000	1.468					
LEVER	Sphericity Assumed	412.256	1	412.256	678.473	.000	.947	678.473	1.000
	Greenhouse-Geisser	412.256	1.000	412.256	678.473	.000	.947	678.473	1.000
	Huynh-Feldt	412.256	1.000	412.256	678.473	.000	.947	678.473	1.000
	Lower-bound	412.256	1.000	412.256	678.473	.000	.947	678.473	1.000
LEVER * LESION	Sphericity Assumed	4.866	1	4.866	8.008	.007	.174	8.008	.787
	Greenhouse-Geisser	4.866	1.000	4.866	8.008	.007	.174	8.008	.787
	Huynh-Feldt	4.866	1.000	4.866	8.008	.007	.174	8.008	.787
	Lower-bound	4.866	1.000	4.866	8.008	.007	.174	8.008	.787
LEVER * DELAY	Sphericity Assumed	68.174	2	34.087	56.099	.000	.747	112.198	1.000
	Greenhouse-Geisser	68.174	2.000	34.087	56.099	.000	.747	112.198	1.000
	Huynh-Feldt	68.174	2.000	34.087	56.099	.000	.747	112.198	1.000
	Lower-bound	68.174	2.000	34.087	56.099	.000	.747	112.198	1.000
LEVER * LESION * DELAY	Sphericity Assumed	6.278	2	3.139	5.166	.010	.214	10.332	.796
	Greenhouse-Geisser	6.278	2.000	3.139	5.166	.010	.214	10.332	.796
	Huynh-Feldt	6.278	2.000	3.139	5.166	.010	.214	10.332	.796
	Lower-bound	6.278	2.000	3.139	5.166	.010	.214	10.332	.796
Error(LEVER)	Sphericity Assumed	23.090	38	.608					
	Greenhouse-Geisser	23.090	38.000	.608					
	Huynh-Feldt	23.090	38.000	.608					
	Lower-bound	23.090	38.000	.608					
SESSION * LEVER	Sphericity Assumed	68.423	13	5.263	65.086	.000	.631	846.119	1.000

- **Tests of within-subjects contrasts.** Well, we didn't ask for this explicitly and we're not interested in any specific contrasts at the moment, so we'll ignore this.

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	SESSION LEVER	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
SESSION	Linear	90.232	1	90.232	163.668	.000	.812
	Quadratic	25.762	1	25.762	117.355	.000	.755
	Cubic	.427	1	.427	3.176	.083	.077
	Order 4	.347	1	.347	1.966	.169	.049
	Order 5	.720	1	.720	6.745	.013	.151
	Order 6	1.390E-03	1	1.390E-03	.021	.885	.001
	Order 7	1.078E-02	1	1.078E-02	.285	.596	.007
	Order 8	3.208E-02	1	3.208E-02	.918	.344	.024
	Order 9	1.396E-02	1	1.396E-02	.541	.467	.014
	Order 10	8.725E-03	1	8.725E-03	.342	.562	.009
	Order 11	1.708E-02	1	1.708E-02	.633	.431	.016
	Order 12	3.351E-03	1	3.351E-03	.120	.731	.003
	Order 13	8.174E-03	1	8.174E-03	.234	.631	.006
SESSION * LESION	Linear	6.717E-02	1	6.717E-02	.122	.729	.003
	Quadratic	2.059	1	2.059	9.381	.004	.198
	Cubic	.538	1	4.883	.853	.385	

- **Levene's test of equality of error variances.** A more important one: tests whether the variances of the various data columns differs across groups (defined by the between-subjects factors). This tests the **homogeneity of variance** assumption of ANOVA. The results here aren't ideal — we have a few violations of this assumption (where $p < 0.05$). For example, the variability of 'session 2, active lever' responses isn't the same across all six between-subjects groups (sham-0, sham-10, sham-20, AcbC-0, AcbC-10, AcbC-20). These data have in fact already been square-root transformed to try to improve matters, but there is still a violation of the homogeneity of variance assumption in 7 of the 28 data columns. We have to make a judgement about the robustness of ANOVA in these circumstances (and the alternative analytical techniques available); although significant, the variances don't in fact differ by huge amounts if you look at the descriptive statistics (for example, the session 2 / active lever responses have SDs that range from 0.407 to 1.001 — a 2.5-fold difference, which isn't the end of the world as ANOVA is reasonably robust to that level of violation; see p. 33).

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
S1_Active	1.329	5	38	.273
S1_Inactive	1.214	5	38	.321
S2_Active	6.407	5	38	.000
S2_Inactive	1.504	5	38	.212
S3_Active	4.806	5	38	.002
S3_Inactive	.633	5	38	.676
S4_Active	2.869	5	38	.027
S4_Inactive	1.099	5	38	.377
S5_Active	.986	5	38	.439
S5_Inactive	4.147	5	38	.004
S6_Active	1.688	5	38	.161
S6_Inactive	4.508	5	38	.003
S7_Active	2.189	5	38	.076
S7_Inactive	4.871	5	38	.002
S8_Active	1.739	5	38	.149
S8_Inactive	3.271	5	38	.015
S9_Active	2.349	5	38	.059
S9_Inactive	1.337	5	38	.270
S10_Active	.900	5	38	.491
S10_Inactive	2.327	5	38	.061
S11_Active	1.102	5	38	.375
S11_Inactive	2.338	5	38	.060
S12_Active	1.878	5	38	.121
S12_Inactive	1.838	5	38	.129
S13_Active	2.034	5	38	.096
S13_Inactive	1.124	5	38	.364
S14_Active	1.185	5	38	.335
S14_Inactive	.838	5	38	.531

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+LESION+DELAY+LESION * DELAY
Within Subjects Design:
SESSION+LEVER+SESSION*LEVER

- **Tests of between-subjects effects.** The other **important** bit that everyone will want to look at. And very easy to interpret. We can see that there's a significant effect of delay ($F_{2,38} = 19.357$, $p < 0.001$) and although there's no main effect of lesion ($F < 1$, NS), there is a lesion × delay interaction ($F_{2,38} = 5.887$, $p = 0.006$). Of course, we'd want to interpret all the within-subjects factors and the

complex interactions too (for example, this data set has a 4-way session × lever × lesion × delay interaction).

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	1924.574	1	1924.574	2234.805	.000	.983	2234.805	1.000
LESION	.463	1	.463	.538	.468	.014	.538	.110
DELAY	33.339	2	16.670	19.357	.000	.505	38.714	1.000
LESION * DELAY	10.139	2	5.069	5.887	.006	.237	11.773	.848
Error	32.725	38	.861					

a. Computed using alpha = .05

- **Parameter estimates.** Not really very useful unless we're doing some regression analysis, so it probably wasn't worth ticking it for this analysis!

Parameter Estimates

Dependent Variable	Parameter	B	Std. Error	t	Sig.	95% Cont Lower Bound
S1_Active	Intercept	.640	.092	6.982	.000	.45
	[LESION=AcbC]	-9.88E-02	.134	-.736	.466	-.37
	[LESION=sham]	0[b]	.	.	.	
	[DELAY=0]	-.310	.130	-2.388	.022	-.57
	[DELAY=10]	.164	.130	1.267	.213	-9.818E-0
	[DELAY=20]	0[b]	.	.	.	
	[LESION=AcbC] * [DELAY=0]	.601	.194	3.096	.004	.20
	[LESION=AcbC] * [DELAY=10]	-4.72E-02	.190	-.249	.805	-.43
	[LESION=AcbC] * [DELAY=20]	0[b]	.	.	.	

- **Estimated marginal means.** These can be useful. SPSS gives the means for the various levels of each factor (or interaction). I also ticked 'Compare main effects… with a Sidak adjustment' in the *Options*. This gives us some quick post-hoc tests. If you have a factor with only two levels (e.g. Lesion), this tells you nothing more than the ANOVA did. But for factors with >2 levels, it can be useful. Here are the means for Delay, which it is certainly valid to perform *post hoc* tests on (since it was significant in the ANOVA, above). We see the mean (across all other variables) for Delay ('Estimates'), and then it compares pairs of delays (0 v. 10, 0 v. 20, 10 v. 20) ('Pairwise comparisons'). We also get the **standard error of the mean (SEM)** for each mean and the **standard error of the difference between means (SED)** for every pairwise comparison (see p. 43→). Finally, it repeats the overall *F* test from the ANOVA (not very helpfully; 'Univariate Tests').

Tip: pairwise comparisons for interactions

Top tip: by default, SPSS only performs pairwise comparisons for factors, and not interactions. If we were to *Paste* the syntax for this analysis, we'd see this sort of thing:

```
/EMMEANS = TABLES(lesion) COMPARE ADJ(SIDAK)
/EMMEANS = TABLES(delay) COMPARE ADJ(SIDAK)
/EMMEANS = TABLES(session) COMPARE ADJ(SIDAK)
/EMMEANS = TABLES(lever) COMPARE ADJ(SIDAK)
/EMMEANS = TABLES(lesion*delay)
/EMMEANS = TABLES(lesion*session)
/EMMEANS = TABLES(delay*session)
…
```

Note that the main effects have `COMPARE` and `ADJ(SIDAK)` on them, but the interactions don't. If you want, you can add that in syntax! Like this:

```
/EMMEANS = TABLES(lesion) COMPARE ADJ(SIDAK)
/EMMEANS = TABLES(delay) COMPARE ADJ(SIDAK)
```

```
/EMMEANS = TABLES(session) COMPARE ADJ(SIDAK)
/EMMEANS = TABLES(lever) COMPARE ADJ(SIDAK)
/EMMEANS = TABLES(lesion*delay) COMPARE(lesion) ADJ(SIDAK)
/EMMEANS = TABLES(lesion*delay) COMPARE(delay) ADJ(SIDAK)
…
```

You can't just put `COMPARE`, because SPSS wouldn't know whether to compare Lesion differences for each level of Delay, or Delay differences for each level of Lesion. So you specify one other thing; for example, **`COMPARE(lesion)`** would compare Lesion groups at each level of Delay. You can specify both kinds of comparison, as I did above. The output also gives you the **standard error of the difference** for each comparison (see p. 45). Finally, you can specify a Sidak correction to the tests by adding **`ADJ(SIDAK)`**, or similarly for Bonferroni if you really want to. This can be extended to higher-order interactions; you specify the factor you want to be compared at all possible combinations of the other factors.

Estimates

Measure: MEASURE_1

delay	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	1.454	.047	1.358	1.550
10	1.269	.045	1.177	1.361
20	1.047	.045	.956	1.139

Pairwise Comparisons

Measure: MEASURE_1

(I) delay	(J) delay	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
0	10	.186*	.066	.022	2.168E-02	.349
	20	.407*	.066	.000	.243	.571
10	0	-.186*	.066	.022	-.349	-2.168E-02
	20	.221*	.064	.004	6.112E-02	.382
20	0	-.407*	.066	.000	-.571	-.243
	10	-.221*	.064	.004	-.382	-6.112E-02

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Sidak.

Univariate Tests

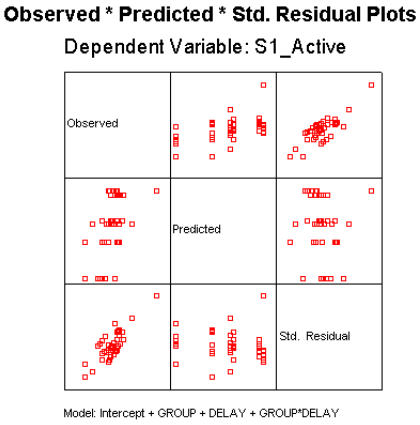
Measure: MEASURE_1

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Contrast	1.191	2	.595	19.357	.000	.505	38.714	1.000
Error	1.169	38	3.076E-02					

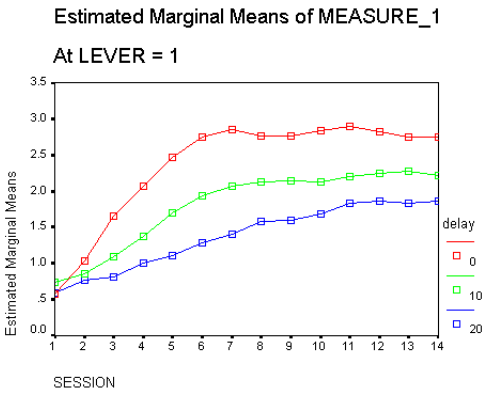
The F tests the effect of delay. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Computed using alpha = .05

- **Observed * predicted * std. residual plots.** SPSS's residual plots are a little bit incomprehensible; see p. 36 for explanations.

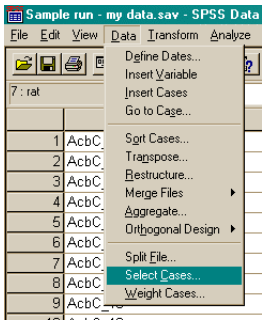


- **Profile plots.** Finally, we get some not-so-pretty graphs:

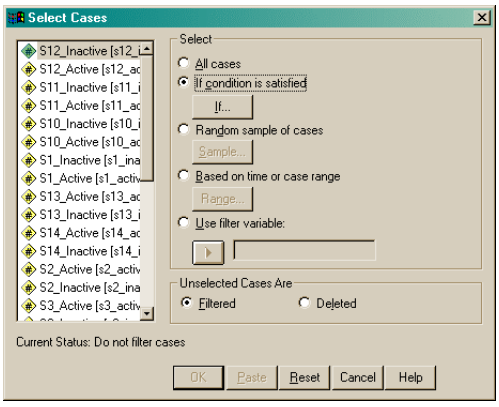


5.3. Further analysis: selecting cases

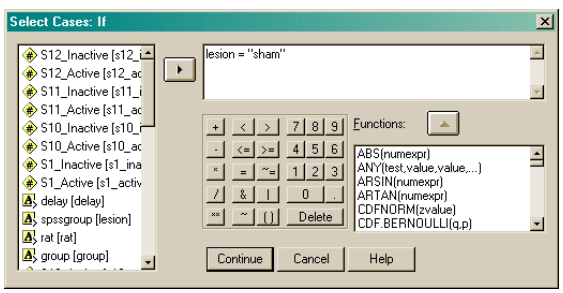
In this situation, we'd want to do **further analysis,** especially since we have a hugely complex 4-way interaction. We might want to find out if there are effect of Lesion or Delay if we only consider Active lever responses — easy, we just run another repeated-measures ANOVA on the Active lever data only, without the Lever factor. We might also want to see if there is an effect of delay / session / lever in the shams alone. For this we might want to **restrict the cases** analysed by SPSS. Choose **Data → Select cases:**



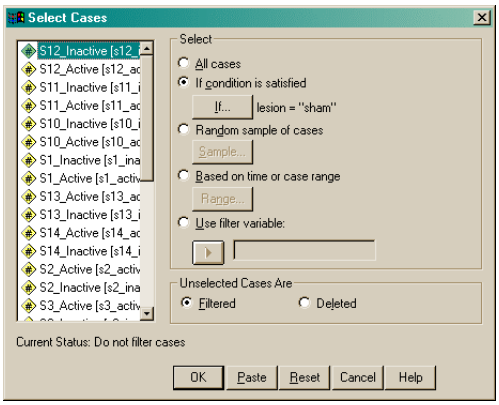
Then click **If…**



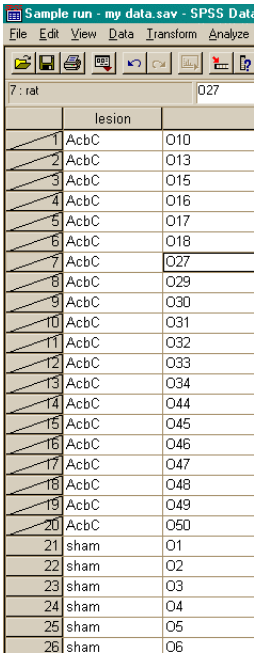
We only want to select cases **if** the **lesion** variable is equal to **"sham":**



Click **Continue** and the condition is entered into the previous dialogue box:

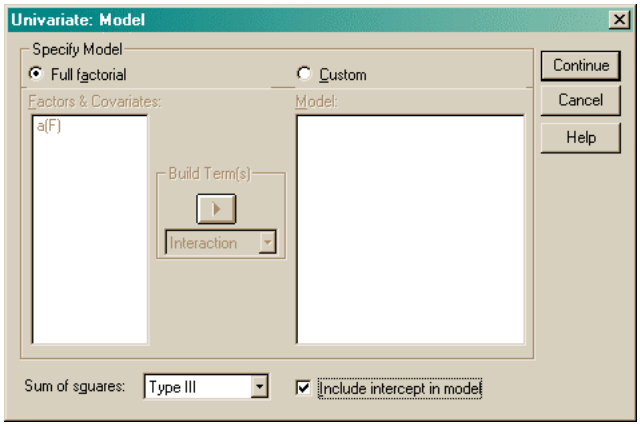


Click OK. You'll now find that all cases (rows) that don't match your criterion are crossed out, and won't be analysed:



5.4 The 'intercept', 'total', and 'corrected total' terms

When you run an ANOVA with SPSS, by default it includes the intercept term. To turn this on / off with the menus, click on the 'Model' button:



You can then choose to 'Include intercept in model' or not. In syntax, you can add the command

```
/INTERCEPT = EXCLUDE
```

or

```
/INTERCEPT = INCLUDE
```

What does this do? Let's illustrate with some sample data for an ANOVA with a single between-subjects factor with two levels:

A	Dependent variable	$\bar{x}_A$	Overall mean $\bar{x}$	$(x-\bar{x})^2$	$(\bar{x}_A-\bar{x})^2$	$(x-\bar{x}_A)^2$	$\bar{x}^2$	x^2	$\bar{x}_A^2$
A_1	10.00	6.55	8.77	1.51	4.93	11.90	76.91	100.00	42.90
A_1	14.00	6.55	8.77	27.35	4.93	55.50	76.91	196.00	42.90
A_1	8.00	6.55	8.77	0.59	4.93	2.10	76.91	64.00	42.90
A_1	7.00	6.55	8.77	3.13	4.93	0.20	76.91	49.00	42.90
A_1	2.00	6.55	8.77	45.83	4.93	20.70	76.91	4.00	42.90
A_1	10.00	6.55	8.77	1.51	4.93	11.90	76.91	100.00	42.90
A_1	1.00	6.55	8.77	60.37	4.93	30.80	76.91	1.00	42.90
A_1	3.00	6.55	8.77	33.29	4.93	12.60	76.91	9.00	42.90
A_1	2.00	6.55	8.77	45.83	4.93	20.70	76.91	4.00	42.90
A_1	8.50	6.55	8.77	0.07	4.93	3.80	76.91	72.25	42.90
A_2	14.29	10.99	8.77	30.47	4.93	10.89	76.91	204.20	120.78
A_2	18.49	10.99	8.77	94.48	4.93	56.25	76.91	341.88	120.78
A_2	12.46	10.99	8.77	13.62	4.93	2.16	76.91	155.25	120.78
A_2	11.63	10.99	8.77	8.18	4.93	0.41	76.91	135.26	120.78
A_2	6.66	10.99	8.77	4.45	4.93	18.75	76.91	44.36	120.78
A_2	14.02	10.99	8.77	27.56	4.93	9.18	76.91	196.56	120.78
A_2	5.66	10.99	8.77	9.67	4.93	28.41	76.91	32.04	120.78
A_2	7.06	10.99	8.77	2.92	4.93	15.44	76.91	49.84	120.78
A_2	6.37	10.99	8.77	5.76	4.93	21.34	76.91	40.58	120.78
A_2	13.26	10.99	8.77	20.16	4.93	5.15	76.91	175.83	120.78
$a = 2$	$n = 10$ per group $N = an$			436.78 = SS_{total} as usually calculated	98.57 = SS_A	338.22 = SS_{error}	1538.26 = $SS_{intercept}$	1975.04 = SS_{total} with intercept included	1636.83 = SS_{model} with intercept as part of model

If you run this analysis with the intercept included, SPSS prints this:

Tests of Between-Subjects Effects

Dependent Variable: DEPVAR

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	98.568[a]	1	98.568	5.246	.034
Intercept	1538.258	1	1538.258	81.867	.000
A	98.568	1	98.568	5.246	.034
Error	338.216	18	18.790		
Total	1975.042	20			
Corrected Total	436.784	19			

a. R Squared = .226 (Adjusted R Squared = .183)

- Here, its SS_{total} is $\sum x^2$; its df_{total} is N.
- The intercept itself (the grand mean) has $SS_{intercept} = N\bar{x}^2$ with $df_{intercept} = 1$.
- The 'corrected total', $SS_{corrected\ total} = SS_{total} - SS_{intercept}$ is what we normally think of as SS_{total}, namely $\sum(x-\bar{x})^2$, with the usual df of $N - 1$.
- The effect of A is given by $SS_A = \sum n(\bar{x}_A - \bar{x})^2$, $df_A = a - 1$.
- The 'corrected model' models the effects of the factor(s), A, *ignoring* the effect of the intercept (the grand mean). If you have more than one factor, the 'corrected model' term is the sum of all their effects: $SS_{corrected\ model} = SS_{total} - SS_{intercept} - SS_{error}$.
- The error is calculated as usual: $SS_{error} = \sum(x-\bar{x}_A)^2$, $df_{error} = (N - 1) - (a - 1)$.

Incidentally, **the *F* test on the intercept term ($MS_{intercept}/MS_{error}$) tests the null hypothesis that the grand mean is zero.** If you run an ANOVA with no factors other than the intercept (or with a factor with only one level, which SPSS will let you do), it is equivalent to a one-sample *t* test comparing all *N* observations to zero; as for any *t* test, $F_{1,k} = t_k^2$ and $t_k = \sqrt{F_{1,k}}$.

If you don't include the intercept, you get this:

Tests of Between-Subjects Effects

Dependent Variable: DEPVAR

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Model	1636.826[a]	2	818.413	43.556	.000
A	1636.826	2	818.413	43.556	.000
Error	338.216	18	18.790		
Total	1975.042	20			

a. R Squared = .829 (Adjusted R Squared = .810)

In other words, when you exclude the intercept, the model models the effects of the factor(s), A, *and the intercept*, together, without distinguishing the two. In this case, it calculates

- SS_{total} is $\sum x^2$; its df_{total} is N.
- The model (intercept plus effect of A) has $SS_{model} = \sum n\bar{x}_A^2$, $df = a = 2$ (df = 2 because there are two $\bar{x}_A$ means and one overall $\bar{x}$ mean).
- SS_A is calculated *without* considering the difference between the effect of A and the grand mean as we would usually do, so $SS_A = SS_{model}$ for this one-factor case.
- The error is calculated as usual: $SS_{error} = \sum (x - \bar{x}_A)^2$, $df_{error} = (N - 1) - (a - 1)$.

It should be fairly clear that **you probably want to 'include' the intercept** when running ANOVAs in SPSS. **This is the default.**

Part 6: advanced topics — harder things about ANOVA

6.1 Rules for calculating sums of squares

6.1.1 Partitioning sums of squares

Sums of squares are partitioned exactly as degrees of freedom (see below, p. 68). This requires a **structural model.** We've seen several examples of this, and many more are discussed in Part 7.

6.1.2 General rule for calculating sums of squares

- Every SS corresponds to a term in the structural model that represents the *difference* between two quantities *P* and *Q*.
- Every SS is the *summed squared deviation* of *P* from *Q*.
- If a term contributing to the SS is based on *n* observations, multiply its contribution by *n*.

For example, for two between-subjects factors *A* and *B*, the structural model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

and if there are *a* levels of *A*, *b* levels of *B*, and *n* subjects (independent observations) per *AB* combination, the SS are

Term	Sum of squares
μ	$SS_{intercept} = N\bar{y}^2$ — generally ignored
$\alpha_i = \mu_{A_i} - \mu$	$SS_A = \sum nb(\bar{y}_A - \bar{y})^2$ — each $\bar{y}_A$ mean based on *nb* scores
$\beta_j = \mu_{B_j} - \mu$	$SS_B = \sum na(\bar{y}_B - \bar{y})^2$ — each $\bar{y}_B$ mean based on *na* scores
$\alpha\beta_{ij} = \mu_{A_iB_j} - (\mu + \alpha_i + \beta_j)$	$SS_{AB} = \sum n(\bar{y}_{AB} - \bar{y})^2 - (SS_A + SS_B)$ — each $\bar{y}_{AB}$ mean based on *n* scores
$\varepsilon_{ijk} = Y_{ijk} - (\mu + \alpha_i + \beta_j + \alpha\beta_{ij})$	$SS_{error} = \sum (y - \bar{y})^2 - (SS_A + SS_B + SS_{AB}) = SS_{total} - (SS_A + SS_B + SS_{AB})$
	$SS_{total} = \sum (y - \bar{y})^2 = SS_A + SS_B + SS_{AB} + SS_{error}$
	$SS_{grand\ total\ including\ intercept} = SS_{total} + SS_{intercept} = \sum y^2$ — generally ignored

We first saw the general technique for deriving these SS equations on p. 15 (and another is on p. 159): we rearrange the structural model to give $Y_{ijk} - \mu$ on the left-hand side, expand out the definition of all the terms, simplify, square both sides of the equation (so we have SS_{total} on the left-hand side), and eliminate a number of terms that sum to zero.

The expected value of the squared terms in the structural model are directly related to the *E*(MS), discussed below (p. 73); for example, $E(\varepsilon_{ijk}^2) = \sigma_e^2$; $E(\alpha_i^2) = \sigma_e^2 + nb\sigma_A^2$.

6.2 Rules for calculating degrees of freedom

From Keppel (1991, pp. 207-214). For any source of variance:

- The *df* equal the number of different observations on which each sum of squares is based, minus the number of constraints operating on these observations. (This is the definition of *df* in general: the number of independent

observations, or the number of observations minus the number of constraints.)

For between-subjects designs:

- The main effect of a factor with a levels has $a - 1$ *df*. So $df_A = a - 1$ and $df_B = b - 1$.
- The main effect of a covariate has 1 *df* (since its effect is represented by a straight line, which can be determined by two parameters, but the line is constrained to pass through the overall mean, so the one *df* represents the line's slope; it's thus akin to a factor with two levels).
- The *df* for an A × B interaction, where A has a levels and B has b levels, is the product of the two separate *df*s, i.e. $df_{A\times B} = (a - 1)(b - 1)$.
- The total number of *df*s is the number of observations N minus 1, i.e. $(N - 1)$.
- The error or residual *df* is df_{total} minus the sum of everything else.

We partition *df*s in exactly the same way as SSs. For example, for an A × B × S design,

$$SS_{total} = SS_A + SS_B + SS_{A\times B} + SS_{error}$$
$$df_{total} = df_A + df_B + df_{A\times B} + df_{error}$$

For within-subjects and mixed designs, most of the above still holds, but we don't have just a single 'error' term. Taking 'groups' to refer to groups of subjects defined by between-subjects factors:

- $df_{\text{between subjects}} = \text{total subjects} - 1$
- $df_{\text{within subjects}} = df_{total} - df_{\text{between subjects}}$
- $df_{\text{subjects within groups}} = df_{\text{between subjects}} - df_{\text{groups}}$
- $df_{\text{WS factor} \times \text{subjects within groups}} = df_{\text{within subjects}} - df_{\text{WS factor}} - df_{\text{WS factor} \times \text{groups}}$

If a group is defined by the between-subjects factor A, we would write 'subjects within groups' as 'S / A'. For example, if we have the design A × (U × S) with a between-subjects factor A with 3 levels, $n = 8$ subjects per group (24 subjects total), and a within-subjects factor U with 6 levels, we would be able to calculate:

$df_{total} = N - 1 = anu - 1 = (3 \times 8 \times 6) - 1 = 143$
$df_{\text{between subjects}} = \text{total subjects} - 1 = 24 - 1 = 23$
$df_A = a - 1 = 3 - 1 = 2$
$df_{S/A} = df_{\text{between subjects}} - df_A = 23 - 2 = 21$
$df_{\text{within subjects}} = df_{total} - df_{\text{between subjects}} = 143 - 23 = 120$
$df_U = u - 1 = 6 - 1 = 5$
$df_{U\times A} = df_U \times df_A = 2 \times 5 = 10$
$df_{U\times S/A} = df_{\text{within subjects}} - df_U - df_{U\times A} = 120 - 5 - 10 = 105$

We partition sums of squares in exactly the same way as *df*s (described for this particular design in more detail later), like this:

$SS_{total} = SS_{\text{between subjects}} + SS_{\text{within subjects}}$
$SS_{\text{between subjects}} = SS_A + SS_{S/A}$
$SS_{\text{within subjects}} = SS_U + SS_{U\times A} + SS_{U\times S/A}$

You can see that this exactly mirrors the *df* partitioning shown above (with suitable simple arithmetic rearrangement).

6.3 Nasty bit: unequal group sizes and non-orthogonal sums of squares

This can be very complicated. So far we've assumed that equal-sized experimental groups have been sampled from equal-sized treatment populations. If this is not the case, we can have problems. Firstly, unequal *n*s exaggerate the problem of heteroge-

neity of variance (see Myers & Well, 1995, pp. 105-106) (and see p. 33). Secondly, they can really screw up an ANOVA.

6.3.1 Proportional cell frequencies

If we have unequal population sizes and the sample sizes reflect the ratios of their sizes — and, if there is >1 factor, the inequalities are in consistent proportions across those factors — we're OK. For example, suppose (Myers & Well, 1995, p. 151) we know that Labour, Conservative, and Liberal Democrat supporters are present in our population in the ratio 4:3:3, and we know that two-thirds of each group voted in the last election. We could quite reasonably run experiments on them with the following numbers of subjects:

	Labour	Conservative	Lib Dem
Voted	24	18	18
Did not vote	12	9	9

No huge problem here. Suppose we use two between-subjects factors *A* and *B* again, as above. Suppose there there are *a* levels of *A* and *b* levels of *B*. But now suppose there are n_i observations for condition A_i, n_j observations for condition B_j, and n_{ij} observations for condition A_iB_j. Since every SS has a contribution from every observation it's based on, the formulae are still very simple:

Term	Sum of squares
μ	$SS_{intercept} = N\bar{y}^2$ — generally ignored
$\alpha_i = \mu_{A_i} - \mu$	$SS_A = \sum_i n_i(\bar{y}_{A_i} - \bar{y})^2$ — since $\bar{y}_{A_i}$ is based on n_i scores
$\beta_j = \mu_{B_j} - \mu$	$SS_B = \sum_j n_j(\bar{y}_{B_j} - \bar{y})^2$ — since $\bar{y}_{B_j}$ is based on n_j scores
$\alpha\beta_{ij} = \mu_{A_iB_j} - (\mu + \alpha_i + \beta_j)$	$SS_{AB} = \sum_{ij} n_{ij}(\bar{y}_{A_iB_j} - \bar{y})^2 - (SS_A + SS_B)$ — since $\bar{y}_{A_iB_j}$ is based on n_{ij} scores
$\varepsilon_{ijk} = Y_{ijk} - (\mu_j + \alpha_i + \beta_j + \alpha\beta_{ij})$	$SS_{error} = \sum(y - \bar{y})^2 - (SS_A + SS_B + SS_{AB}) = SS_{total} - (SS_A + SS_B + SS_{AB})$
	$SS_{total} = \sum(y - \bar{y})^2 = SS_A + SS_B + SS_{AB} + SS_{error}$
	$SS_{grand\ total\ including\ intercept} = SS_{total} + SS_{intercept} = \sum y^2$ — generally ignored

6.3.2 Disproportionate cell frequencies — a problem

Here's an example (from Howell, 1997, p. 430): experimenters test the number of errors made by sober and by drunk people on a simulated driving test. Two experiments divide up the work, testing half the subjects in their Michigan lab and half in their Arizona lab. They have absolutely no reason to think that the choice of state makes any difference. These are their results:

Number of errors	Sober	Drunk	
Michigan	13, 15, 14, 16, 12	18, 20, 22, 19, 21, 23, 17, 18, 22, 20	Michigan mean = 18.0
	(*n* = 5, mean = 14)	(*n* = 10, mean = 20)	
Arizona	13, 15, 18, 14, 10, 12, 16, 17, 15, 10, 14	24, 25, 17, 16, 18	Arizona mean = 15.9
	(*n* = 11, mean = 14)	(*n* = 5, mean = 20)	
	Sober mean = 14	Drunk mean = 20	

It appears that drunk subjects make more errors than sober subjects, which makes sense, but it also looks like Michigan subjects make more errors than Arizona subjects. But clearly that's an Alcohol effect masquerading as a State effect — the Michigan lab tested a higher proportion of its subjects while drunk. The two factors

are *correlated*, thanks to disproportionate cell frequencies — if you knew whether a subject was drunk or sober, you could guess better than chance which state the subject came from. What can we do? We can use **unweighted means.** When we calculated the Michigan mean, we calculated it as a **weighted mean** (where M = Michigan, S = sober, D = drunk in the formula):

$$\begin{aligned}\bar{y}_M &= \frac{\sum y_{M,S} + \sum y_{M,D}}{n_M} \\ &= \frac{n_{M,S}\bar{y}_{M,S} + n_{M,D}\bar{y}_{M,D}}{n_M} \\ &= \frac{5\times 14 + 10\times 20}{15} = 18\end{aligned}$$

This is *weighted* in the sense that the contribution of individual cell means ($\bar{y}_{M,S}$ and $\bar{y}_{M,D}$) is weighted by the sample sizes ($n_{M,S}$ and $n_{M,D}$). An *unweighted mean* (or, more properly, an **equally weighted mean**) is what you get when you simply average the cell means, ignoring the number of subjects in each cell. That would give us a Michigan mean of $(14 + 20)/2 = 17$, and an Arizona mean exactly the same. In an unweighted-means analysis, each cell mean contributes equally to the calculation of each of the sums of squares. In the calculation, we calculate an average cell size (the *harmonic mean* of the cells sizes; see revision maths chapter, p. 213) and use that average *n* as if every cell had that many subjects (Howell, 1997, pp. 430-435).

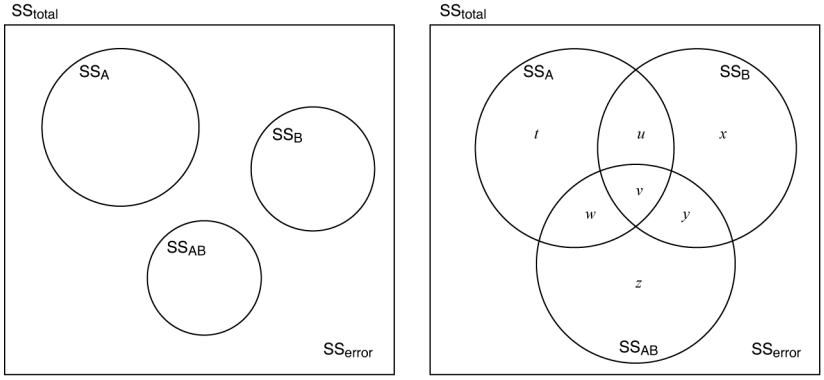
This is a specific example of a general problem — when the effects of two or more effects (or interactions) are *not fully independent.* The example shown above is fairly common (the effects of one factor, State, are *partly* correlated with the effects of another, Alcohol, because one state tested a higher proportion of drunks). It may be easier to visualize the problem with an even more extreme example — one in which two factors A and B are *completely* correlated. Consider this particularly stupid set of data collected as part of an $A_2 \times B_2 \times S$ design (Myers & Well, 1995, p. 153):

	A_1	A_2
B_1	no observations	18 12 11 7 14 6 7 6
B_2	10 14 8 7 2 10 1 3	no observations

Let's calculate the SS. Each observation makes one contribution to the SS, as usual, so we should define n_i as the number of observations at level A_i, n_j as the number of observations at level B_j, and n_{ij} as the number of observations at A_jB_j. Then

$$SS_{total} = \sum (y - \bar{y})^2 = 322.00$$
$$SS_A = \sum_i n_i (\bar{y}_{A_i} - \bar{y})^2 = 42.25$$
$$SS_B = \sum_j n_j (\bar{y}_{B_j} - \bar{y})^2 = 42.25$$
$$SS_{AB} = \sum_{ij} n_{ij} (\bar{y}_{A_i B_j} - \bar{y})^2 - (SS_A + SS_B) = -42.25$$
$$SS_{error} = SS_{total} - (SS_A + SS_B + SS_{AB}) = 279.75$$

Pretty stupid; we have a negative SS_{AB}! The problem is that the effects of A and B in this design are **not orthogonal;** the main effects of A and B are perfectly correlated (simply because there are only observations for A_1B_2 and A_2B_1; the effects of A and B are **confounded**). If we added two A_1B_1 and two A_2B_2 observations, the effects of A and B are now not perfectly correlated, but they are still correlated. The problem can be illustrated like this:



Orthogonal sums of squares *Nonorthogonal sums of squares*

If we calculate SS_A in the usual way, it consists of $t+u+v+w$. On the other hand, if we adjust it for the contribution of the other main effect B, it would consist of $t+w$. Or we could adjust it for the contribution of B and AB, in which case the adjusted SS_A would consist only of t. Similar options exist for the other sources of variance. The appropriate choice probably depends on the importance the experimenter attaches to the various factors (Myers & Well, 1995, p. 155). See also Howell (1997, pp. 578-582).

This also means that **the order you enter terms into a computer analysis can affect the results.** On some packages, an ANOVA with the sources of variance being A, B, and A × B gives you a different answer from an ANOVA with the sources of variance being B, A, and A × B. The **default method in SPSS does not care about the order** — it's what SPSS refers to as the 'Type III' sum of squares. I think (Myers & Well, 1995, p. 155) that this method uses area t for SS_A, area x for SS_B, and z for SS_{AB}. This is probably what you want — it is certainly appropriate for the case when there is **chance variation in cell frequencies,** such as when subjects drop out at random (Myers & Well, 1995, p. 155). It is also the method approximated by the 'unweighted (equally weighted) means' solution described above (Howell, 1997, p. 582).

In general, whenever cell frequencies (*n*s in each cell) are equal or proportional (meaning that for each cell, $n_{ij} = n_i n_j / N$), the sums of squares are orthogonal (unless the experiment itself has been mis-designed and confounds two variables). But **whenever cell frequences are disproportionate, the sums of squares are nonorthogonal** (Myers & Well, 1995, p. 154; Howell, 1997, pp. 429-435 and 578-579).

This problem occurs whenever predictor variables are themselves correlated (see also Myers & Well, 1995, pp. 555-563). ANOVA with equal cell frequencies is exactly equivalent to multiple regression with uncorrelated categorical variables (Myers & Well, 1995, p. 536), and ANOVA with disproportionate cell frequencies implies that the factors are correlated. This is easy to see: if our Autism × Sex ex-

periment has 8 male autistics, 2 female autistics, 2 male controls, and 8 female controls (disproportionate cell frequencies), you can make a better-than chance guess as to whether a subject is male or female if you know whether they're autistic or not — the two factors are correlated. It is, of course, possible to have a middle ground — unequal but proportionate cell frequencies (see above, p. 70, for an example), which still involves orthogonal sums of squares.

6.4 Expected mean squares (EMS) and error terms

First we need to consider the **sampling fraction** for fixed and random factors (fixed and random factors are defined on p. 31). If we have factor A with a levels and it is a fixed factor, we have sampled all the levels. We can say that the maximum number of levels of A is $a_{max} = a$, and the sampling fraction $a / a_{max} = 1$. On the other hand, if our factor is a random factor, a_{max} is likely to be very large, so $a / a_{max} = 0$, approximately. Take the example of subjects: we presume that our s subjects are sampled from a very large population, $s_{max} \approx \infty$, so the sampling fraction $s / s_{max} = 0$.

It is possible to have sampling fractions between 0 and 1 (Howell, 1997, p. 423) — but you will have to work out some messy EMSs yourself. Software packages such as SPSS assume that the sampling fraction is 1 for fixed factors and 0 for random factors.

6.4.1 Rules for obtaining expected mean squares (EMS)

From Myers & Well (1995, p. 299). Let's list the rules with an illustrative example. Suppose we have one between-subjects factor A with 3 levels. There are 6 subjects *per level* of the between-subjects factor ($n = 6$). There are 4 levels of a within-subjects factor B.

1. Decide for each independent variable, including Subjects, whether it is **fixed or random.** Assign a letter to designate each variable. Assign another letter to represent the number of levels of each variable. (In our example, the variables are designated *A*, *B*, and *S*; the levels are *a*, *b*, and *n* respectively. *A* and *B* are fixed and *S* is random.)

2. Determine the **sources of variance** (SS) from the structural model. (We've already seen what this produces for our example design, when we discussed it earlier: SS_{total} is made up of $SS_A + SS_{S/A} + SS_B + SS_{AB} + SS_{SB/A}$. These are our sources of variance.)

3. List σ_e^2 as part of each EMS.

4. For each EMS, list the null hypothesis component — that is, the component corresponding directly to the source of variance under consideration. (Thus, we add $nb\sigma_A^2$ to the EMS for the *A* line, and $b\sigma_{S/A}^2$ to the EMS for the S/A line.) Note that a component consists of three parts:

   - A coefficient representing the number of scores at each level of the effect (for example, *nb* scores at each level of *A*, or *b* scores for each subject).
   - σ^2

     [Myers & Well (1995, pp. 299) use σ_A^2 if *A* is a random factor, and θ_A^2 if *A* is a fixed factor; Howell (1997, p. 423) doesn't, and I think it's clearer not to.]
   - As subscripts, those letters that designate the effect under consideration.

5. Now add to each EMS all components whose subscripts contain all the letters designating the source of variance in question. (For example, since the subscript SB/A contains the letters *S* and *A*, add $\sigma_{SB/A}^2$ to the EMS for the S/A line.)

6. Next, examine the components for each source of variance. If a slash (/) appears in the subscript, define only the letters to the left of the slash as 'essential'. If there are several slashes, only the letters preceding the leftmost slash are essential. If there is no slash, all letters are essential.

7. Among the essential letters, ignore any that are necessary to designate the source of variance. (If the source of variance is *A*, for example, then when considering $n\sigma^2_{AB}$, ignore the *A*. If the source is S/A, then when considering the $\sigma^2_{SB/A}$ component, *S* and *B* are essential subscripts and *S* is to be ignored.) If any of the remaining (non-ignored) essential letters designate fixed variables, delete the entire component from the EMS.

An example:

Term	EMS so far

Step 1: identify variables and numbers of levels.

A, a (between-subjects factor)
B, b (within-subjects factor)
S, n (number of subjects per group)

Step 2: identify sources of variance.

A
S / A
B
BA
SB / A

Step 3: List σ^2_e as part of each EMS.

A	σ^2_e
S / A	σ^2_e
B	σ^2_e
BA	σ^2_e
SB / A	σ^2_e

Step 4: list the null hypothesis component.

A	$\sigma^2_e + nb\sigma^2_A$
S / A	$\sigma^2_e + b\sigma^2_{S/A}$
B	$\sigma^2_e + an\sigma^2_B$
BA	$\sigma^2_e + n\sigma^2_{BA}$
SB / A	$\sigma^2_e + \sigma^2_{SB/A}$

Step 5: add all components whose subscripts contain all the letters designating the source of variance in question.

A	$\sigma^2_e + nb\sigma^2_A + b\sigma^2_{S/A} + n\sigma^2_{BA} + \sigma^2_{SB/A}$
S / A	$\sigma^2_e + b\sigma^2_{S/A} + \sigma^2_{SB/A}$
B	$\sigma^2_e + an\sigma^2_B + n\sigma^2_{BA} + \sigma^2_{SB/A}$
BA	$\sigma^2_e + n\sigma^2_{BA} + \sigma^2_{SB/A}$
SB / A	$\sigma^2_e + \sigma^2_{SB/A}$

Steps 6 and 7: for each component, define 'essential' letters; ignore any that are part of the designation of the source of variance; if any remaining essential letters contain fixed factors, delete the component.

A	$\sigma_e^2 + nb\sigma_A^2 + b\sigma_{S/A}^2$
S / A	$\sigma_e^2 + b\sigma_{S/A}^2$
B	$\sigma_e^2 + an\sigma_B^2 + \sigma_{SB/A}^2$
BA	$\sigma_e^2 + n\sigma_{BA}^2 + \sigma_{SB/A}^2$
SB / A	$\sigma_e^2 + \sigma_{SB/A}^2$

6.4.2 Choosing an error term

A mean square qualifies as an error term for testing an effect if its E(MS) matches the $E(\text{MS}_{\text{effect}})$ in all respects except the null-hypothesis component (Keppel, 1991, p. 568). In our example above, therefore, we'd test MS_{A} against $\text{MS}_{\text{S/A}}$, and we'd test both MS_{B} and MS_{BA} against $\text{MS}_{\text{SB/A}}$.

6.4.3 Pooling error terms

When we have random factors in a model, important variables are often tested against an interaction term. Since interaction terms have few *df* (and since power depends on *F* being large when the null hypothesis is false, and since *F* is the ratio of $\text{MS}_{\text{effect}}$ to MS_{error}, and since MS_{error} is $\text{SS}_{\text{error}} / df_{\text{error}}$), this means we may have poor power to detect such effects.

One possibility is to **test** interaction terms in a full model with a conservative criterion, like this (Howell, 1997, p. 425). If there is an interaction ($p < 0.05$), we declare that there's an interaction. If there isn't ($0.05 < p < 0.25$), we just look at the results for other terms. But if there is no interaction ($p > 0.25$), we **remove** the interaction term from the model. In the example above, if we found that the AB interaction was not significant ($p > 0.25$), we could remove any terms including it and its *df* would contribute to the within-subjects error term, which might increase power to detect effects of B (see p. 51).

6.5 Contrasts

See Howell (1997, pp. 354-369); Myers & Well (1995, chapter 6).

6.5.1. Linear contrasts

Linear contrasts are comparisons between linear combinations of different groups. Suppose we want to know whether students are more bored on Wednesdays than other weekdays, because Wednesday is statistics day, and whether they're more bored on weekdays than weekends. We could measure their boredom on all days of the week, and use DayOfWeek as a factor (with 7 levels) in an ANOVA. If this turned up significant, we would know that all days were not the same — but it wouldn't answer our original questions. We can do that with linear contrasts.

In general, a linear contrast is a linear combination of a set of treatment means. Each mean μ_j is *weighted* by a weight w_j:

$$L = w_1\mu_1 + w_2\mu_2 + \ldots + w_k\mu_k = \sum_j w_j\mu_j$$

$$\text{such that } \sum_j w_j = 0$$

In our example, suppose μ_1 is the Monday mean, μ_2 is the Tuesday mean, and so on. Our 'Wednesdays versus other weekdays' question can be written as a linear contrast:

$$L = \frac{\mu_{\text{Mon}} + \mu_{\text{Tue}} + \mu_{\text{Thu}} + \mu_{\text{Fri}}}{4} - \mu_{\text{Wed}}$$

or

$$L = +\frac{1}{4}\mu_{\text{Mon}} + \frac{1}{4}\mu_{\text{Tue}} - 1\mu_{\text{Wed}} + \frac{1}{4}\mu_{\text{Thu}} + \frac{1}{4}\mu_{\text{Fri}} + 0\mu_{\text{Sat}} + 0\mu_{\text{Sun}}$$

Equivalently (multiply everything up to get whole numbers):

$$L = +1\mu_{\text{Mon}} + 1\mu_{\text{Tue}} - 4\mu_{\text{Wed}} + 1\mu_{\text{Thu}} + 1\mu_{\text{Fri}} + 0\mu_{\text{Sat}} + 0\mu_{\text{Sun}}$$

If the Wednesday mean is the same as the mean of the other weekdays, we expect that $L = 0$. So our null hypothesis is that $L = 0$. If a statistical test rejects this null hypothesis (shows that L deviates from 0 more than chance alone would predict), we would conclude that Wednesdays were different from other weekdays. Our 'weekdays versus weekends' question could be written as a different linear contrast:

$$L = +\frac{1}{5}\mu_{\text{Mon}} + \frac{1}{5}\mu_{\text{Tue}} + \frac{1}{5}\mu_{\text{Wed}} + \frac{1}{5}\mu_{\text{Thu}} + \frac{1}{5}\mu_{\text{Fri}} - \frac{1}{2}\mu_{\text{Sat}} - \frac{1}{2}\mu_{\text{Sun}}$$

Again, if the null hypothesis (weekdays the same as weekends) is true, the expected value of L is 0. Comparisons between individual pairs of means can also be accomplished with linear contrasts — for example, Sunday versus Monday (the 'back to work' effect?):

$$L = +1\mu_{\text{Mon}} + 0\mu_{\text{Tue}} + 0\mu_{\text{Wed}} + 0\mu_{\text{Thu}} + 0\mu_{\text{Fri}} - 0\mu_{\text{Sat}} - 1\mu_{\text{Sun}}$$

For any contrast,

$$\text{SS}_{\text{contrast}} = \frac{L^2}{\sum_j w_j^2 / n_j}$$

All linear contrasts have 1 *df* per contrast. The significance test of a contrast is given by $F = \text{MS}_{\text{contrast}} / \text{MS}_{\text{error}}$.

6.5.2. Type I error rates with planned contrasts

If we ran pairwise comparison *post hoc* tests on our days-of-the-week example, we'd make ${}_2^7C = 21$ pairwise comparisons, so if we used $\alpha = 0.05$ per comparison, our familywise α_{FW} would be a huge 0.66. We'd run the risk of falsely declaring all sorts of differences significant. But our experiment was only designed to answer three questions: Wednesdays v. other weekdays, weekdays v. weekends, and Sundays v. Mondays. So if we only ask these questions, which we had in mind *a priori*, we could never declare the 'Monday v. Tuesday' difference significant. Ask fewer questions, less chance of a Type I error.

In general, the methods of controlling for Type I errors are the same in principle for *a priori* and *post hoc* tests. The differences are simply (1) that we generally ask fewer questions *a priori*, and (2) when we perform *post hoc* tests we often focus on the differences that look biggest — which is logically equivalent to performing all possible comparisons (visually) and then selecting the biggest for statistical testing. Since this has a high likelihood of a Type I error, such data-guided *post hoc* tests must be corrected as if we were making all possible comparisons (because actually we are). As Myers & Well (1995, p. 179) put it, 'the effective size of a family of *post hoc* contrasts is determined not by the number of contrasts actually tested but by those that conceivably might have been tested, had the data suggested it was worth doing so'.

When we specify in advance (*a priori*) which comparisons we're interested in, we can specify the Type I error rate per contrast (EC or α) or per family of contrasts (EF or α_{FW}). What should constitute a 'family' of contrasts? All the contrasts an experi-

menter ever runs? All that are published in a single paper? Most people would say no; although that would result in a very low Type I error rate, it would lead to a high Type II error rate (low power) — missing real differences. There are two serious candidates for a 'family' (Myers & Well, 1995, p. 178). They are (1) all the contrasts made in a single experiment; (2) all the contrasts associated with a single source of variance in a single experiment. Suppose your experiment has three factors, A, B, and C. By the first criterion, all contrasts in your A × B × C design together constitute one family. By the second criterion, there are seven families (involving A, B, C, AB, AC, BC, and ABC). Myers & Well (1995, p. 178) recommend the second criterion as a reasonable compromise between Type I and Type II errors.

Once you've decided how many contrasts are in a family, you can reduce your EC (α), or increase your p values, to obtain the desired EF (α_{FW}). For example, you could use the **Bonferroni** or **Sidak** corrections discussed above; these are simple (though the Bonferroni is over-conservative, so I prefer the Sidak). If you run k contrasts that are independent (orthogonal, see p. 77), $\alpha_{FW} = 1 - (1- \alpha)^k$, so the Sidak correction is spot on. If your contrasts are not independent, $\alpha_{FW} < 1 - (1- \alpha)^k$ (Myers & Well, 1995, p. 177) but it is hard to calculate α_{FW} exactly, so just use the Šidák or Bonferroni correction and at worst your tests will be conservative.

Planned contrasts may be conducted whether or not the overall *F* tests from the ANOVA are significant (Myers & Well, 1995, p. 179). In fact, you could run them instead of the usual ANOVA, but you are recommended to run the ANOVA too (Myers & Well, 1995, pp. 179, 196). Why? (1) Because our theories are rarely good enough that we are willing to forgo checking whether unanticipated effects are present in the data with *post hoc* tests, suitably controlled for Type I error. (2) The ANOVA carries additional information, for example about the effect size; see p. 97→. Note also that the ANOVA may give a different result from a family of *post hoc* tests, since the power of the ANOVA is that of the 'maximum contrast' (Myers & Well, 1995, p. 196), which may not be obvious or interesting (e.g. it may reflect a linear combination of groups that you wouldn't have thought about in advance, such as 0.3×Mon + 0.7×Tue – 0.4×Wed – 0.6×Sat).

6.5.3. Orthogonal contrasts

Contrasts are *orthogonal* if the questions they ask are *independent*. This is one set of 6 orthogonal contrasts for our days-of-the-week example, showing how you can break down a set of means into a set of orthogonal contrasts:

(Mon, Tue, Wed, Thu, Fri) v. (Sat, Sun)
(Mon, Tue) v. (Wed, Thu, Fri) (Sat) v. (Sun)
(Mon) v. (Tue) (Wed) v. (Thu, Fri)
(Thu) v. (Fri)

All these are independent of each other. But these two are not independent:

(Mon) v. (Tue) (Mon) v. (Wed)

There are many possible sets of orthogonal contrasts (some of them involving odd fractional combinations of day means, which might not be very meaningful experimentally!). For any complete set of orthogonal contrasts, $SS_{treatment} = \sum SS_{contrast}$, and $df_{treatment} = \sum df_{contrast}$. So for our days-of-the-week example, we would need 6 orthogonal contrasts for a complete set; the set of 6 shown above is one complete set.

Formally, two contrasts $L_1 = \sum_j w_{j(1)}\mu_j$ and $L_2 = \sum_j w_{j(2)}\mu_j$ are orthogonal if, for equal sample sizes, $\sum_j w_{j(1)}w_{j(2)} = 0$ (Howell, 1997, p. 361). The more general condition, for unequal sample sizes, is $\sum_j \frac{w_{j(1)}w_{j(2)}}{n_j} = 0$ (Myers & Well, 1995, p. 176).

There's no reason that we should test only orthogonal contrasts — we test the contrasts that ask the questions we're interested in.

6.5.4. Linear contrasts in SPSS

In SPSS, to run linear contrasts other than very specific ones (such as comparing all groups separately to the last one), you need to specify the design in syntax using the `/CONTRAST()=SPECIAL()` or `/LMATRIX` command. For a between-subjects ANOVA of a dependent variable (depvar) with one factor (Day, 7 levels), you can specify your contrasts like this:

```
UNIANOVA
  depvar BY day
  /CONTRAST (day)=Special(0.25  0.25 -1     0.25  0.25  0     0
                          0.2   0.2   0.2   0.2   0.2  -0.5  -0.5
                          1     0     0     0     0     0    -1)
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /PRINT = TEST(LMATRIX)
  /DESIGN = day .
```

The `/PRINT…` command makes your custom contrast matrix appear under the heading *Custom Hypothesis Tests*, followed by the results (significance values for each test), followed by the sum of squares for the contrast. In this example you can see that contrast L1 is 'Wednesdays v. other weekdays', L2 is 'weekdays v. weekends', and L3 is 'Sundays v. Mondays'. All are significant in this example.

Custom Hypothesis Tests

Contrast Coefficients (L' Matrix)

	DAY Special Contrast		
Parameter	L1	L2	L3
Intercept	.000	.000	.000
[DAY=1.00]	.250	.200	1.000
[DAY=2.00]	.250	.200	.000
[DAY=3.00]	-1.000	.200	.000
[DAY=4.00]	.250	.200	.000
[DAY=5.00]	.250	.200	.000
[DAY=6.00]	.000	-.500	.000
[DAY=7.00]	.000	-.500	-1.000

The default display of this matrix is the transpose of the corresponding L matrix.

Contrast Results (K Matrix)

DAY Special Contrast			Dependent Variable DEPVAR
L1	Contrast Estimate		.315
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		.315
	Std. Error		.067
	Sig.		.000
	95% Confidence Interval for Difference	Lower Bound	.177
		Upper Bound	.453
L2	Contrast Estimate		-1.184
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		-1.184
	Std. Error		.050
	Sig.		.000
	95% Confidence Interval for Difference	Lower Bound	-1.287
		Upper Bound	-1.080
L3	Contrast Estimate		-1.134
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		-1.134
	Std. Error		.085
	Sig.		.000
	95% Confidence Interval for Difference	Lower Bound	-1.308
		Upper Bound	-.959

Test Results

Dependent Variable: DEPVAR

Source	Sum of Squares	df	Mean Square	F	Sig.
Contrast	10.403	3	3.468	191.080	.000
Error	.508	28	.018		

Alternatively, you can use the LMATRIX syntax, which allows you to specify any linear combination of any number of factors or interactions (SPSS, 2001, pp. 478-9). It may help to read the GLM section to understand this (p. 84→, especially p. 93→). For our simple example the syntax would be:

```
UNIANOVA
```

```
depvar BY day
/LMATRIX = "Wed_v_otherweekday"
           day 0.25  0.25 -1     0.25  0.25  0     0
/LMATRIX = "weekday_v_wkend"
           day 0.2   0.2   0.2   0.2   0.2  -0.5  -0.5
/LMATRIX = "sun_v_mon"
           day 1     0     0     0     0     0    -1
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/CRITERIA = ALPHA(.05)
/PRINT = TEST(LMATRIX)
/DESIGN = day .
```

or to put all the tests into one matrix as before,

```
UNIANOVA
  depvar BY day
  /LMATRIX = day 0.25  0.25 -1     0.25  0.25  0     0;
             day 0.2   0.2   0.2   0.2   0.2  -0.5  -0.5;
             day 1     0     0     0     0     0    -1
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /PRINT = TEST(LMATRIX)
  /DESIGN = day .
```

If you want to obtain separate sums of squares for each contrast (reasons for which are given below), you can use the version with several `/LMATRIX` commands — you get one 'Test Results' box with one sum of squares for each `/LMATRIX` command. (It's also possible to work out $SS_{contrast}$ from the 'contrast estimate' L given in the results and the weight coefficients printed in the L matrix, using $SS_{contrast} = \frac{L^2}{\sum_j w_j^2 / n_j}$, but this is rather a pain.)

If you specify nonorthogonal contrasts, like this:

```
UNIANOVA
  depvar BY a
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /LMATRIX = "contrast1" a -1 +1 0 0 0
  /LMATRIX = "contrast2" a -1 0 +1 0 0
  /LMATRIX = "bothtogether" a -1 +1 0 0 0;
                            a -1 0 +1 0 0
  /PRINT = TEST(LMATRIX)
  /DESIGN = a .
```

then you will find that $SS_{contrast1} + SS_{contrast2} \neq SS_{bothtogether}$. For a discussion of correlated (nonorthogonal) predictors, see above and pp. 70 and 97.

6.5.5. Contrasts in multifactor designs — an overview

The same principles can be applied to any contrast, even involving multiple factors (Myers & Well, 1995, pp. 188-185). Suppose we have two factors: therapy type (A: control CON, analytic therapy AT, behaviour therapy BT, cognitive therapy CT) and patient diagnosis (B: unipolar depression D, schizophrenia S, manic depression M). We measure some sort of dependent variable. We find a main effect of A, a main effect of B, and an AB interaction. We can therefore reject these null hypotheses:

$$\alpha_{CON} = \alpha_{AT} = \alpha_{BT} = \alpha_{CT} = 0$$
$$\beta_D = \beta_S = \beta_M = 0$$
$$\alpha\beta_{CON,D} = \ldots = \alpha\beta_{CT,M} = 0$$

We can ask further questions using contrasts. Does the mean of control subjects differ from the mean of all the therapy populations? That would be a single contrast:

$$L_1 : \mu_{CON} - \frac{\mu_{AT} + \mu_{BT} + \mu_{CT}}{3}$$

Call this (control) versus (all other treatments) the 'treatment effect'. Does the treatment effect vary over clinical populations? That would involving seeing if three contrasts differ:

$$T = \mu_{CON} - \frac{\mu_{AT} + \mu_{BT} + \mu_{CT}}{3}$$

$$T_D = \mu_{CON,D} - \frac{\mu_{AT,D} + \mu_{BT,D} + \mu_{CT,D}}{3}$$

$$T_S = \mu_{CON,S} - \frac{\mu_{AT,S} + \mu_{BT,S} + \mu_{CT,S}}{3}$$

$$T_M = \mu_{CON,S} - \frac{\mu_{AT,M} + \mu_{BT,M} + \mu_{CT,M}}{3}$$

$$H_0 : T_D = T_S = T_M = T$$

This is harder but possible (Myers & Well, 1995, pp. 190-5); it involves testing a sum of squares based on the deviations of T_D, T_S, and T_M from the overall treatment effect T. And so on. An SPSS-based example of this sort of thing is given on p. 93→.

6.6 Trend analysis: the effects of quantitative factors

6.6.1. Trends

Factors are categorical variables. But some categories are qualitative (male / female; bipolar / schizophrenic / control) and some are quantitative (session 1 / 2 / 3…, stimulus height 7 cm / 9 cm / 11 cm…). How can we ask quantitative questions about the relationship between stimulus height and our dependent variable? Well, if the predictor variables are continuous (covariates), you can ask things like 'is my dependent variable a linear function of the predictor?' (simple ANCOVA = linear regression, see p. 135) or 'is my dependent variable a quadratic function of the predictor?' (polynomial ANCOVA, see p. 88→). But with categorical predictors (factors), you use **trend analysis** (see Myers & Well, 1995, chapter 7; Howell, 1997, pp. 386-396). Obviously, this technique requires that the levels of the factor are in some sort of order.

We can accomplish this using contrasts but with particular weights for our contrast coefficients. For example, returning to our days-of-the-week example, taking just the weekdays, we can ask:

	Mon	Tue	Wed	Thu	Fri
Do people get happier during the week?	–2	–1	0	+1	+2
A **linear trend.**					
Are people happier in the middle of the week?	+2	–1	–2	+1	+2
A **quadratic trend,** an example of a non-linear (curved) trend.					

So for our linear example, we could test the contrast

$$L = -2\mu_{Mon} - 1\mu_{Tue} + 0\mu_{Wed} + 1\mu_{Thu} + 2\mu_{Fri}$$

$$H_0 : L = 0$$

The contrast coefficients shown above would be valid if (1) the values of the factor are equally spaced, as they are for days of the week, and (2) each mean is based on the same number of scores. If not, see below.

One common approach to trend testing is to ask what set of trends explain the data well (Myers & Well, 1995, pp. 209-216). Here we would be guided by our theories. Suppose (Myers & Well, 1995, p. 204) we are performing a generalization expermient; we train subjects that an 11" stimulus predicts electric shock. We might expect that an 11" stimulus would elicit a substantial skin conductance response, which would generalize somewhat to 9" and 13" stimuli, but less so to 7" and 15" stimuli. This would be an inverted-U-shaped curve, and such a curve can be described by a quadratic equation ($y = a + bx^2$, where $b < 0$). So the responses to 7/9/11/13/15" stimuli might be something like 1/4/9/4/1 units. We might also expect that larger stimuli cause more of a response — a straight line relationship between stimulus size and response, which can be described by a linear equation ($y = a + bx$). So if this were the only thing influencing responding, responding for the 7/9/11/13/15" stimuli might be something like 1/2/3/4/5 units. Overall, if these two effects are independent, we might expect an asymmetric inverted-U curve, the sum of the other two effects ($y = b_0 + b_1x + b_2x^2$) — in this example, 2/6/12/8/6 units.

We can perform an ANOVA to ask if the stimuli differ. Suppose they do — the effect of the factor is significant. We know that taking full account of our factor, A, can explain a certain amount of variance: SS_A, such that $SS_{total} = SS_A + SS_{error}$. Applying Occam's razor, it's common to ask first whether a straight line (a linear trend) can explain the data well. Suppose we obtain a sum of squares for our linear contrast, SS_{linear}. We can see if this accounts for a significant amount of variability: $F_{linear} = MS_{linear} / MS_{error}$. So does the effect of A include something *over and above* a linear component? Well, $SS_A = SS_{linear} + SS_{nonlinear}$ (and, of course, $df_A = df_{linear} + df_{nonlinear} = 1 + df_{nonlinear}$). So we can calculate an F test to see if there's anything 'substantial' in that nonlinear component: $F_{df\text{-nonlinear}/\text{df-error}} = MS_{nonlinear} / MS_{error}$. **This is an *F* test for the *lack of fit* of the linear model** (see also Myers & Well, 1995, p. 411) — we know how much variability A accounts for overall; the question is, what component of that is linear and what is not. If this isn't significant, our linear model does a good enough job — we stop. If it is, we can add in a quadratic trend. We would now have $SS_A = SS_{linear} + SS_{quadratic} + SS_{higher\text{-}order}$. We can test $SS_{higher\text{-}order}$ to see if we should add any other predictors (SS_{cubic}…) and carry on until the 'leftovers' no longer contain anything significant. However, if your theory predicts certain components (e.g. linear and quadratic), you shouldn't perform tests that you're not interested in (Myers & Well, 1995, p. 216).

If you have a groups, then you can fit at most a polynomial of order a–1. So if you have 5 groups, you can only fit a linear (x^1), quadratic (x^2), cubic (x^3), and quartic (x^4) trend; you haven't got enough data to fit a quintic (x^5) trend. So in general, the most complex polynomial equation that can be fitted with a groups is

$$\hat{Y}_j = b_0 + b_1 X_j + b_2 X_j^2 + \ldots + b_p X_j^p + \ldots + b_{a-1} X_j^{a-1}$$

To apply this technique, the trends (SS_{linear}, $SS_{quadratic}$, …) have to be independent of each other, or *orthogonal*, so that their sums of squares add up to SS_A.

If (1) the values of the factor are equally spaced, as they are for days of the week, and (2) each mean is based on the same number of scores, coefficients can easily be generated for a set of orthogonal polynomials (Myers & Well, 1995, pp. 209-216 and Table D7; Howell, 1997, p. 391 and Appendix Polynomial). It is possible to derive coefficients when the *n*s are not equal and/or the groups are not equally spaced (Myers & Well, 1995, pp. 211, 227-229) but it is much simpler to use standard linear and/or nonlinear regression techniques, treating the predictor as a continuous variable (see pp. 82, 88→, 135).

6.6.2. Trend analysis in SPSS

In SPSS, polynomial contrasts can be done easily. In the example of an ANOVA with a factor A, specify this:

```
UNIANOVA
  depvar BY a
```

```
/CONTRAST (a)=Polynomial
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/CRITERIA = ALPHA(.05)
/PRINT = TEST(LMATRIX)
/DESIGN = a .
```

You can also specify the contrast coefficients by hand. For a factor with five levels, equally spaced, with equal *n*, you could use:

```
UNIANOVA
  depvar BY a
  /LMATRIX = "a linear"     a -2 -1  0  1  2
  /LMATRIX = "a quadratic"  a  2 -1 -2 -1  2
  /LMATRIX = "a cubic"      a -1  2  0 -2  1
  /LMATRIX = "a quartic"    a  1 -4  6 -4  1
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /PRINT = TEST(LMATRIX)
  /DESIGN = a .
```

To use the ANOVA dialogue box (Analyze → General Linear Model → …), choose **Contrasts,** set the contrast for your factor to 'Polynomial', and click **change.** To get the LMATRIX printout, choose **Options → Contrast coefficient matrix.** Other forms of contrast (and the lack-of-fit test described above for 'is there anything significant left over that needs to be accounted for?') can be specified by hand using the syntax outlined above (p. 77).

For one-way ANOVA, better output is obtained from **Analyze → Compare means → One-way ANOVA.** Click **Contrasts → Polynomial,** and enter the order of the polynomial. You may also want **Options → Means plot.** The output looks like this:

ANOVA

ILLNESS

			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		6791.540	4	1697.885	20.779	.000
	Linear Term	Contrast	174.845	1	174.845	2.140	.147
		Deviation	6616.695	3	2205.565	26.992	.000
	Quadratic Term	Contrast	6100.889	1	6100.889	74.663	.000
		Deviation	515.806	2	257.903	3.156	.047
	Cubic Term	Contrast	389.205	1	389.205	4.763	.032
		Deviation	126.601	1	126.601	1.549	.216
	4th-order Term	Contrast	126.601	1	126.601	1.549	.216
Within Groups			7762.700	95	81.713		
Total			14554.240	99			

6.6.3.How trend analysis relates to multiple regression or polynomial ANCOVA

Trend analysis described how well linear, quadratic, etc., components fit the means of each group. Suppose A is your factor (five levels: 7", 9", 11", 13", and 15" stimuli). Your dependent variable is *Y*: you have 20 observations per level (100 subjects). You could treat A as a factor, as we've seen, or as a continuous variable.

- If you performed a linear regression or ANCOVA with your predictor variable having one of the values 7, 9, 11, 13, and 15 for all the subjects, and your independent variable being *Y*, you would find that your $SS_{regression}$ was the same as the SS_{linear} from the ANOVA contrast.
- If you performed a linear regression with your predictor variable having the values 7^2, 9^2, 11^2, 13^2, and 15^2 and your independent variable being *Y*, you would *not* obtain $SS_{quadratic}$. Trend analysis assumes that the centre of the quadratic function is in the middle of your groups. In our example, the 'middle value' is 11. So the quadratic trend analysis calculates $\hat{y} = a + b(x-11)^2$, not $\hat{y} = a + bx^2$. If you obtain $SS_{regression}$ with $(x - 11)^2$ as your predictor, you *will* obtain $SS_{quadratic}$ from the ANOVA contrast. You'll also obtain the same answer if you use the quadratic trend coefficients (2, –1, –2, 1, 2) as your predictor (x^2) values.
- You'd think that SS_{cubic} is the $SS_{regression}$ you obtain with the regression model $\hat{y} = a + b(x-11)^3$. But the cubic trend analysis coefficients for five

groups are (–1, 2, 0, –2, 1), so that's what you need to use as your predictor values to obtain SS_{cubic}. I'd initially thought they'd be something like (–8, –1, 0, 1, 8) — but the problem is that these values are *not orthogonal* to the other (linear, quadratic) components. Specifically, (–8, –1, 0, 1, 8) is not orthogonal to the linear component. If you use (–1, 2, 0, –2, 1) your cubic predictor values, you do obtain SS_{cubic}. This is the cubic component *over and above any linear component.*

- If you put all these predictors into a multiple regression, you get the correct SS for each component *as long as the predictors are orthogonal*; otherwise, they start to reduce each other's SS (see pp. 70 and 97).
- Each of the SSs should be compared to the overall MS_{error} from the ANOVA to get the same *F* values for all the components. The multiple regression approach can never measure the 'group means' for different values of A in the way the ANOVA does, so it can never measure the 'lack of fit'. Its $SS_{error\text{-}multiple\text{-}regression}$ reduces towards $SS_{error\text{-}ANOVA}$ as you put in more predictors and the prediction gets better. To work out whether it's worth putting another predictor in, you would have to compare the multiple regression R^2 values for models with and without the extra predictor (see p. 86→). This is one advantage of trend analysis — you begin by knowing how much variability the group means of your factor account for (which the multiple regression doesn't), and you try to work out what polynomial components contribute to that.

6.7 How computers perform complex ANOVAs: the general linear model (GLM)

6.7.1 The basic idea of a GLM, illustrated with multiple regression

Following Howell (1997, p. 567), suppose we want to solve a multiple-regression equation to predict Y with three predictors (variables X_1, X_2, and X_3). Our equation is:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3$$

or, written out for an individual observation:

$$Y_i = b_0 + b_1 X_{i,1} + b_2 X_{i,2} + b_3 X_{i,3} + e_i$$

where i stands for a particular observation (labelled from 1 to n) and e_i is the error associated with each observation. We could write that using vector (matrix) notation (see revision chapter on matrices, p. 196):

$$\mathbf{y} = \mathbf{b_0} + b_1 \mathbf{x_1} + b_2 \mathbf{x}_2 + b_3 \mathbf{x_3} + \mathbf{e}$$

where $\mathbf{y}$, $\mathbf{x_1}$, $\mathbf{x_2}$, $\mathbf{x_3}$ are $n \times 1$ vectors of data, $\mathbf{e}$ is an $n \times 1$ vector of errors, and $\mathbf{b_0}$ is an $n \times 1$ vector whose elements are the intercept. This can be further reduced to

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}$$

where there are p predictor variables, $\mathbf{X}$ is an $n \times (p + 1)$ matrix of predictors, the first column of which contains only ones, and $\mathbf{b}$ is a $(p + 1) \times 1$ matrix of regression coefficients — like this:

$$\mathbf{y} = \begin{bmatrix} 1 & X_{1,1} & X_{1,2} & X_{1,3} \\ 1 & X_{2,1} & X_{2,2} & X_{2,3} \\ 1 & X_{\ldots,1} & X_{\ldots,2} & X_{\ldots,3} \\ 1 & X_{n,1} & X_{n,2} & X_{n,3} \end{bmatrix} \times \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \ldots \\ e_n \end{bmatrix}$$

Solving a multiple regression equation then becomes the problem of solving $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$ for $\mathbf{b}$ so as to minimize the sum of squares of the residuals, $\sum (Y_i - \hat{Y})$ or $\sum e_i^2$. When this is solved, $\mathbf{b}$ contains the correct regression coefficients.

Three things are worth noting. Firstly, the *multiple regression coefficient* R is the correlation between Y and $\hat{Y}$, and its square is the proportion of variability accounted for by the overall regression:

$$R^2 = \frac{\mathrm{SS}_{\text{regression}}}{\mathrm{SS}_Y}$$

Secondly, the contribution of *individual* predictors may be easy to specify (if the predictors themselves aren't correlated, in which case $r^2_{\text{each predictor}}$ represents the proportion of total variation explained by each predictor and $R^2 = \sum r^2_{\text{each predictor}}$) or rather tricky to specify (if the predictors are correlated); see Myers & Well (1995, pp. 505-508). And just as r for a sample can be adjusted (to r_{adj}) to provide a better estimate of ρ for the underlying population, R^2 can also be adjusted according to the sample size (Myers & Well, 1995, p. 508-9) (see p. 98). Other issues regarding multiple regression are discussed by Howell (1997, ch. 15).

Thirdly, the method of solving this matrix equation is pretty damn complicated when there are several predictors. It's illustrated for linear regression (and even more complicated cases) at

www.mathworld.com/LeastSquaresFitting.html
www.mathworld.com/NonlinearLeastSquaresFitting.html

and a general proof is given on p. 204, but we'll just leave SPSS to do it for us.

6.7.2 Using a GLM for simple ANOVA: the design matrix

How can we represent ANOVA in this way? Suppose we take our old favourite, the one-way ANOVA with a single between-subjects factor, A. Our equation for this is

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

where τ_i is the effect of level *i* of factor A. This symbol τ_i represents τ_1, τ_2, τ_3 … τ_a (if there are *a* levels of factor A) but for one subject we are only interested in the contribution of one level of A. We can accomplish this with something called a **design matrix.** The design matrix, **X**, will have $a + 1$ columns and as many rows as there ae subjects. Suppose (after Howell, 1997, p. 567) there are 6 subjects, 3 levels of A, and 2 subjects per level. Then our design matrix looks like this (the 'S' or Subject column is purely for explanation and isn't part of the matrix):

$$\begin{array}{cc} S & \begin{array}{cccc} \mu & A_1 & A_2 & A_3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} & \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

So subjects 1 and 2 experienced treatment A_1, subjects 3 and 4 experienced treatments A_2, and subjects 5 and 6 experienced treatments A_3. All subjects experienced the effect of the overall mean, so the first column is full of ones. We can now define our treatment matrix and write the whole thing in matrix form:

$$\mathbf{y} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} + \begin{bmatrix} e_{1,1} \\ e_{1,2} \\ e_{2,1} \\ e_{2,2} \\ e_{3,1} \\ e_{3,2} \end{bmatrix}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{e}$$

Solving this equation for $\boldsymbol{\tau}$ so as to minimize $\sum e^2$ gives us the treatment effects (μ, τ_1, τ_2, τ_3) we're after.

However, for practical use, it's common to alter the design matrix slightly. Firstly, the μ column has no variance, so it can't go into a standard multiple regression analysis, so we remove it. Secondly, the A_3 column is redundant: if a subject isn't in A_1 or A_2, we know it's in A_3 (i.e. there are only 2 *df* for A), so we remove that too. Finally, to make our treatment effects matrix give treatment effects that are relative to the overall mean (μ), the mean of each column must be zero (corresponding to the ANOVA requirement that $\sum \tau_i = 0$). We can achieve this by scoring a subject 1 in column A_i if the subject is a member of treatment A_i, scoring –1 if the subject is a member of the last (*a*th) treatment, and scoring 0 otherwise. (This is sometimes

called *sigma-restricted parameterization*, since the columns sum to zero, while the original form is called the *overparameterized model*, since it contains redundant information. It is possible to analyse using the overparameterized model; see www.statsoft.nl/textbook/stglm.html.) Anyway, this process gives us this revised design matrix, which carries all the necessary information:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ -1 & -1 \\ -1 & -1 \end{bmatrix}$$

6.7.3 Example of a GLM for a one-way ANOVA

So suppose we have these data (one datum per subject):

A_1	A_2	A_3	A_4
8	5	3	6
9	7	4	4
7	3	1	9

To analyse them with a GLM, we use a set of matrices like this (one row per subject):

$$\begin{bmatrix} 8 \\ 9 \\ 7 \\ 5 \\ 7 \\ 3 \\ 3 \\ 4 \\ 1 \\ 6 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{bmatrix} \times \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} + \mathbf{e}$$

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{e}$$

$$\mathbf{Y} = \mathbf{X\tau} + \mathbf{e}$$

The regression coefficient matrix can be called **b** (as it was for multiple regression) or **τ** (as it was for ANOVA). The overall R^2 will represent $\frac{SS_{model}}{SS_{total}}$, and testing it for significance is the same as testing the effect of A for significance. The intercept in the regression model will equal the grand mean (Howell, 1997, p. 571-2).

6.7.4 GLM for two-way ANOVA and beyond

Let's move up to a two-way ANOVA, with between-subjects factors A and B (Howell, 1997, p. 572). Our full model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

We can easily deal with the α_i and β_j terms in a design matrix. To represent the interaction, $\alpha\beta_{ij}$, we use the fact that an interaction represents a multiplicative effect of

the two variables. Let's start with a **2 × 2** design. Our design matrix, once coded using sigma-restricted parameterization, would look like this:

$$\mathbf{X} = \begin{array}{c} \\ a_1b_1 \\ a_1b_2 \\ a_2b_1 \\ a_2b_2 \end{array} \begin{array}{c} \begin{array}{ccc} A_1 & B_1 & AB_{11} \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix} \end{array}$$

This matrix has one row per AB combination, but in actual use we'd have to replicate the rows so that there was **one row per subject.** So if there were five subjects in the a_1b_2 condition, for example, there'd have to be five rows whose coefficients were [1 –1 –1]. In this matrix the first column represents the main effect of A, as it distinguishes those subjects who received treatment A_1 and those who received A_2. The second column represents the main effect of B, distinguishing B_1 from B_2. The third column is the AB interaction. Its elements are obtained by multiplying the corresponding elements of the first two columns. As always, we have as many columns per effect as we have degrees of freedom for that effect ($df_A = 1$; $df_B = 1$; $df_{AB} = 1$). There are no '0' entries because with only two levels of each variable, a subject is either in the first or the last (–1) level.

Now consider a 2 × 3 factorial ($A_2 \times B_3$). We now have $df_A = 1$, $df_B = 2$, and $df_{AB} = 2$. So for the **full model,** we obtain the following matrix (again, we'd need to ensure that we had one row per subject in the 'real thing'):

$$\mathbf{X} = \begin{array}{c} \\ a_1b_1 \\ a_1b_2 \\ a_1b_3 \\ a_2b_1 \\ a_2b_2 \\ a_2b_3 \end{array} \begin{array}{c} \begin{array}{ccccc} A_1 & B_1 & B_2 & AB_{11} & AB_{12} \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 & -1 \\ -1 & -1 & -1 & 1 & 1 \end{bmatrix} \end{array}$$

This simply applies the principles outlined above for the A_1, B_1, and B_2 columns; the AB_{11} column is the product of the A_1 column and the B_1 column, while the AB_{12} column is the product of A_1 and B_2.

Running an ANOVA like this gives us an overall R^2. Since we know that $SS_{regression} = SS_{model} = SS_Y \times R^2 = SS_A + SS_B + SS_{AB}$, and $SS_{residual} = SS_{error} = SS_Y(1 - R^2)$, we can calculate our SS_{model} and SS_{error}, we know our df_{model} (= $df_A + df_B + df_{AB}$) and df_{error}, and therefore we can calculate an F test for the whole model (= MS_{model} / MS_{error}). However, this doesn't tell us what proportion of the effect is attributable to A, B, or AB. To partition the variance, we must recalculate the regression for a number of **reduced models.** We might call the sum of squares for the full model that we've just looked at $\mathbf{SS_{regression–\alpha,\beta,\alpha\beta}}$. If we dropped the interaction columns (AB_{11} and AB_{12}), we'd be deleting the predictors containing information about the interaction but we'd retain the predictors containing information about α and β; we'd call the resulting sum of squares $\mathbf{SS_{regression–\alpha,\beta}}$. If we only used the A_1, AB_{11} and AB_{12} columns, our model would only account for α and $\alpha\beta$; we'd obtain $\mathbf{SS_{regression–\alpha,\alpha\beta}}$. If we only used the B_1, B_2, AB_{11} and AB_{12} columns, our model would only account for β and $\alpha\beta$; we'd obtain $\mathbf{SS_{regression–\beta,\alpha\beta}}$. Once we've calculated these, we can say that

$$\begin{aligned} SS_{AB} &= SS_{regression_{\alpha,\beta,\alpha\beta}} - SS_{regression_{\alpha,\beta}} \\ SS_A &= SS_{regression_{\alpha,\beta,\alpha\beta}} - SS_{regression_{\beta,\alpha\beta}} \\ SS_B &= SS_{regression_{\alpha,\beta,\alpha\beta}} - SS_{regression_{\alpha,\alpha\beta}} \end{aligned}$$

For example, if the interaction term accounts for any of the variance in Y, then removing the interaction term should lead to a decrease in the variation in Y account-

able for; that decrease is equal to the variation attributable to the interaction. And so on for the other terms. Note that if the predictors are in some way intercorrelated, these sums of squares may not add up to SS_{total} (see section above on disproportionate cell means with a Venn diagram, p. 70); that's fine (Howell, 1997, p. 583-5). This method is the one that assigns SS_A = area *t*, SS_B = area *x*, SS_{AB} = area *z* in the Venn diagram above (see p. 70→), which is often what you want (Myers & Well, 1995, p.155).

Finally, to test these effects (to obtain *F* statistics for the effects of A, B, and AB), we need to know how to **compare one model to another.** And this is very simple (Myers & Well, 1995, p. 441 and 512-4; Howell, 1997, p.578). We can use any of the following equivalent statements. If we have a Full and a Reduced model (with *f* and *r* predictors respectively),

$$F_{\left(df_{\text{error(R)}}-df_{\text{error(F)}}\right),df_{\text{error(F)}}} = \frac{\left(\text{SS}_{\text{error(R)}}-\text{SS}_{\text{error(F)}}\right)\div\left(df_{\text{error(R)}}-df_{\text{error(F)}}\right)}{\text{SS}_{\text{error(F)}}\div df_{\text{error(F)}}}$$

$$F_{\left(df_{\text{model(F)}}-df_{\text{model(R)}}\right),df_{\text{error(F)}}} = \frac{\left(\text{SS}_{\text{model(F)}}-\text{SS}_{\text{model(R)}}\right)\div\left(df_{\text{model(F)}}-df_{\text{model(R)}}\right)}{\text{SS}_{\text{error(F)}}\div df_{\text{error(F)}}}$$

$$F_{f-r,N-f-1} = \frac{\left(N-f-1\right)\left(R_f^2-R_r^2\right)}{\left(f-r\right)\left(1-R_f^2\right)}$$

The second formulation is perhaps the clearest from the point of view of ANOVA; the third is the most useful when you have a multiple regression coefficient R^2 for each model. So to test the effect of A, we calculate the full model to obtain $SS_{\text{regression}-\alpha,\beta,\alpha\beta}$, a reduced model to obtain $SS_{\text{regression}-\beta,\alpha\beta}$, and test the difference between them as above. But this simplifies a bit — for example, take the effect of A:

$$\begin{aligned}\text{SS}_\text{A} &= \text{SS}_{\text{regression}\,\alpha,\beta,\alpha\beta} - \text{SS}_{\text{regression}\,\beta,\alpha\beta}\\ &= \text{SS}_{\text{regression(full)}} - \text{SS}_{\text{regression(reduced)}}\end{aligned}$$

$$F_{\left(df_{\text{model(F)}}-df_{\text{model(R)}}\right),df_{\text{error(F)}}} = \frac{\left(\text{SS}_{\text{model(F)}}-\text{SS}_{\text{model(R)}}\right)\div\left(df_{\text{model(F)}}-df_{\text{model(R)}}\right)}{\text{SS}_{\text{error(F)}}\div df_{\text{error(F)}}}$$

$$\begin{aligned}F_{df_\text{A},df_{\text{error(F)}}} &= \frac{\text{SS}_\text{A}\div df_\text{A}}{\text{SS}_{\text{error(F)}}\div df_{\text{error(F)}}}\\ &= \frac{\text{MS}_\text{A}}{\text{MS}_{\text{error(F)}}}\end{aligned}$$

6.7.5 An overview of GLM designs

We've seen that a **one-way** ANOVA uses this design matrix:

$$\mathbf{X} = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array}\begin{array}{c} \begin{array}{ccc} X_0 & X_1 & X_2 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix} \end{array}$$

This form of the matrix keeps the first 'grand mean' column (X_0) but uses sigma-restricted coding for the A factor. As usual, the duplication of rows necessary to get 'one datum, one row', is not shown — if there were one subject (S1) in condition A_1, two subjects (S2, S3) in condition A_2, and one subject (S4) in condition A_3 that would make the final matrix look like this:

$$\mathbf{X} = \begin{array}{c} \\ A_{1,S1} \\ A_{2,S2} \\ A_{2,S3} \\ A_{3,S4} \end{array} \begin{array}{c} \begin{array}{ccc} X_0 & X_1 & X_2 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix} \end{array}$$

If we used the overparameterized model to represent A, the matrix is simpler. This is the reduced form (ignoring the fact that we'll eventually need one row per subject):

$$\mathbf{X} = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} \begin{array}{c} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

A **two-way ANOVA with no interaction terms** might look like this (left-hand version in sigma-parameterized form; right-hand version in overparameterized version):

$$\mathbf{X} = \begin{array}{c} \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2 \end{array} \begin{array}{c} \\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \end{array} \text{ or } \mathbf{X} = \begin{array}{c} \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2 \end{array} \begin{array}{c} \\ \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix} \end{array}$$

A **two-way ANOVA with the usual interaction term** looks like this, with an X_3 column (the interaction term) that is the product of the X_1 (A) and X_2 (B) columns:

$$\mathbf{X} = \begin{array}{c} \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2 \end{array} \begin{array}{c} \begin{array}{cccc} X_0 & X_1 & X_2 & X_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \end{array} \text{ or } \mathbf{X} = \begin{array}{c} \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2 \end{array} \begin{array}{c} \\ \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

In the overparameterized form, there's a grand mean column, then two columns for the two levels of A, then two columns for the two levels of B, then four columns for the possible values of the AB interaction.

In a **fractional factorial design,** columns are omitted from a full factorial design. We saw an example above, in which the interaction was omitted from a 2×2 factorial design. Similarly, you might choose to run a $2 \times 2 \times 2$ ANOVA but to ignore the 3-way interaction. The appropriate matrix is shown below (overparameterized version); it has 1 grand mean column, 2 columns for A, 2 columns for B, 2 columns for C, 4 columns for AB, 4 columns for AC, and 4 columns for BC.

$$\mathbf{X} = \begin{array}{c} \\ A_1B_1C_1 \\ A_1B_1C_2 \\ A_1B_2C_1 \\ A_1B_2C_2 \\ A_2B_1C_1 \\ A_2B_1C_2 \\ A_2B_2C_1 \\ A_2B_2C_2 \end{array} \begin{array}{c} \\ \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

In a **nested design,** variability due to one factor is 'nested' within variability due to another factor. For example, if one were to administer four different tests to four school classes (i.e. a between-groups factor with four levels), and two of those four classes are in school A, whereas the other two classes are in school B, then the levels of the first factor (four different tests) would be nested in the second factor (two dif-

ferent schools). In the design, nested variables never appear as main effects. For example, if we have a factor A (3 levels) and a factor B (2 levels) nested within A, our overparameterized matrix has one grand mean column, 3 columns for A, and 6 columns for the effect of B nested within A ['B / A' or 'B(A)'].

$$\mathbf{X} = \begin{array}{c} \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2 \\ A_3B_1 \\ A_3B_2 \end{array} \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Overparameterized models are always used to represent nested designs, as the sigma-restricted coding method has difficulty dealing with the design (see www.statsoft.nl / textbook / stglm.html).

A **simple regression** design, with a single continuous predictor variable, is easy to code. If there were three Y data points (dependent variable) and the corresponding values of the predictor variable X were 7, 4, and 9, then the design matrix for the regression $Y = b_0 + b_1X$ would be:

$$\mathbf{X} = \begin{array}{c} X_0 \quad X_1 \end{array} \begin{bmatrix} 1 & 7 \\ 1 & 4 \\ 1 & 9 \end{bmatrix}$$

A simple **quadratic regression** such as $Y = b_0 + b_1X^2$ would be coded simply by squaring the relevant values:

$$\mathbf{X} = \begin{array}{c} X_0 \quad X_1 \end{array} \begin{bmatrix} 1 & 49 \\ 1 & 16 \\ 1 & 81 \end{bmatrix}$$

Multiple regressions, such as $Y = b_0 + b_1P + b_2Q + b_3R$, are coded just as simple regressions. In a **factorial regression** design, combinations (products) of the predictors are included in the design. If the predictors are P and Q, then the full factorial design would include P, Q, and their interaction (P by Q), represented by the product of P and Q scores for each case. So the equation would be $Y = b_0 + b_1P + b_2Q + b_3PQ$. Factorial regression designs can also be **fractional,** in which you omit some of the higher-order effects from the design. An example would be a design with three predictors that omitted the three-way interaction: $Y = b_0 + b_1P + b_2Q + b_3R + b_4PQ + b_5PR + b_6QR$. **Polynomial regressions** contain main and higher-order effects for the predictors but do not include interactions. For example, the second-degree polynomial design for three predictors would include main (first-order) effects, quadratic (second-order) effects, but not interactions: $Y = b_0 + b_1P + b_2P^2 + b_3Q + b_4Q^2 + b_5R + b_6R^2$. There are many other possible designs.

Analysis of covariance refers to a design containing both categorical predictors (factors) and continuous predictors (covariates). Traditionally, however, the term has referred specifically to designs in which the first-order effects (only) of one or more continuous predictors are taken into account when assessing the effects of one or more factors. For example, suppose a researcher wants to assess the influence of a factor A with 3 levels on some outcome, and measurements on a continuous predictor C, known to covary with the outcome, are available. If the data are:

$$\begin{matrix} C \\ \begin{bmatrix} 7 \\ 4 \\ 9 \\ 3 \\ 6 \\ 8 \end{bmatrix} \end{matrix} \quad \begin{matrix} group \\ \begin{bmatrix} A_1 \\ A_1 \\ A_2 \\ A_2 \\ A_3 \\ A_3 \end{bmatrix} \end{matrix}$$

then the design matrix would be

$$\mathbf{X} = \begin{matrix} X_0 & X_1 & X_2 & X_3 \end{matrix} \begin{bmatrix} 1 & 7 & 1 & 0 \\ 1 & 4 & 1 & 0 \\ 1 & 9 & 0 & 1 \\ 1 & 3 & 0 & 1 \\ 1 & 6 & -1 & -1 \\ 1 & 8 & -1 & -1 \end{bmatrix} \text{ or } \mathbf{X} = \begin{matrix} X_0 & X_1 & X_2 & X_3 & X_4 \end{matrix} \begin{bmatrix} 1 & 7 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 & 0 \\ 1 & 9 & 0 & 1 & 0 \\ 1 & 3 & 0 & 1 & 0 \\ 1 & 6 & 0 & 0 & 1 \\ 1 & 8 & 0 & 0 & 1 \end{bmatrix}$$

In the left-hand (sigma-restricted) model, the equation is $Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3$ and the coefficients b_2 and b_3 represent the effects of A, controlling for the effects of *C*. The b_1 coefficient represents the effects of *C* controlling for *A*.

This traditional analysis is inappropriate when the categorical and continuous predictors interact in influencing the dependent variable. The appropriate design is the **separate slope design,** which includes the factor × covariate interaction. For the situation above, the overparameterized matrix that includes the main effect of *A* and the $A \times C$ interaction would be:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & 7 & 0 & 0 \\ 1 & 1 & 0 & 0 & 4 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 9 & 0 \\ 1 & 0 & 1 & 0 & 0 & 3 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 6 \\ 1 & 0 & 0 & 1 & 0 & 0 & 8 \end{bmatrix}$$

Separate slope designs omit the main effects of *C*. Overparameterized matrices are always used for separate slope designs, since the sigma-restricted model runs into problems (www.statsoft.nl/textbook/stglm.html). The **homogeneity of slopes** design can be used to test whether the covariate and factor interact, and thus whether the traditional ANCOVA or the separate slope design is better. This one does include the main effect of *C*:

$$\mathbf{X} = \begin{bmatrix} 1 & 7 & 1 & 0 & 0 & 7 & 0 & 0 \\ 1 & 4 & 1 & 0 & 0 & 4 & 0 & 0 \\ 1 & 9 & 0 & 1 & 0 & 0 & 9 & 0 \\ 1 & 3 & 0 & 1 & 0 & 0 & 3 & 0 \\ 1 & 6 & 0 & 0 & 1 & 0 & 0 & 6 \\ 1 & 8 & 0 & 0 & 1 & 0 & 0 & 8 \end{bmatrix}$$

'Mixed' ANOVA and ANCOVA models are those that contain *random effects*, rather than fixed effects, for one or more factors. The difference is only in how effects are tested. When computers perform tests for designs that include random (rather than fixed) factors, they have to work out the appropriate error term for every effect in the model. In a fixed-effect design, between-subjects effects are always tested using the mean squared residual as the error term. But in mixed-model de-

signs, between-subjects effects are tested using relevant error terms based on the covariation of random sources of variation in the design. Computers do this with something called the 'denominator synthesis' approach of Satterthwaite (1946); details are at

www.statsoft.nl / textbook / stglm.html [covers much GLM theory]
www.statsoft.nl / textbook / stvarcom.html

Remember, a mean square qualifies as an error term for testing an effect if its $E(\text{MS})$ matches the $E(\text{MS}_{\text{effect}})$ in all respects except the null-hypothesis component (Keppel, 1991, p. 568).

Within-subjects (repeated measures) designs can be analysed by coding 'Subject' as a set of columns (Myers & Well, 1995, pp. 569-572). If there are n subjects, there must be n–1 'S' columns (sigma-restricted parameterization form of the matrix) or n columns (overparameterized form); similarly, any interactions involving S can be coded.

Within-subjects (repeated measures) designs can also be analysed by constructing new dependent variables — for example, if subjects are tested at time 1 and time 2, a new 'difference between the two times' variable can be constructed and analysed. These techniques can be extended to multiple levels of a within-subjects factor and multiple factors using special techniques based on multivariate analysis (see below), or by considering 'Subjects' as a (random) factor in its own right and working out the relationship between the other factors. For example, a very common example is a design with one between-subjects factor and one within-subjects factor, written A $\times$ (U $\times$ S); variation due to subjects is nested within variation due to A (or, for shorthand, S is nested within A), because each subject is only tested at one level of the between-subjects factor. The disadvantage with the latter technique is that it does not take account of the potentially major problem of correlation between differences between levels of a within-subjects factor, known as the *sphericity* problem (see below and p. 25→).

6.7.6 A hint at multivariate analysis: MANOVA

The **Y** matrix, so far an $n \times 1$ vector of n observations of a single Y variable, can be replaced by an $n \times m$ matrix of n observations of m different Y variables. In this case, the **b** vector similarly has to be replaced by a matrix of coefficients. The advantage is that you can then analyse linear combinations of several *dependent* variables, which may themselves be correlated; one application is to measure the strength of the relationships between predictor and dependent variables independent of the dependent variable interrelationships. For example, if we give students one of two textbooks and measure their performance on maths *and* physics (two dependent variables), we might want to ask whether the textbooks affected performance, and if so, whether a textbook improved maths, physics, or both — yet students' performance on maths and physics tests may be related. Some of the theory is discussed at

www.statsoft.nl / textbook / stglm.html [general GLM theory]
www.statsoft.nl / textbook / stanman.html#multivariate

A multivariate approach can also be used for **within-subjects** (repeated measures) designs. The bonus is that the **sphericity** problem (q.v.) is bypassed altogether. Essentially, the problem of sphericity relates to the fact that the comparisons involved in testing within-subjects factors with >2 levels may or may not be independent of each other, and if they're not, then the ANOVA results will be wrong unless we account for this. For example, if subjects learn some material and are tested at times 1, 2, and 3, then subjects who learn most between time 1 and time 2 (contrast: time 2 – time 1) may learn least between time 2 and time 3 (contrast: time 3 – time 2), so the two contrasts are *not independent*. ANOVA assumes that all contrasts are independent (orthogonal). It's easy to see what that means if you had a factor A: 'male or not' and a factor B: 'female or not' — if you entered both factors into an ANOVA, both factors would account for equal variance (since they ask the same question — are not orthogonal) and if you partitioned out this variance you'd get the wrong answer

(since you'd be partitioning out the same thing twice). This is the problem that within-subjects contrasts can run into. Correction procedures such as the Greenhouse–Geisser and Huynh–Feldt procedure attempt to deal with this. But a multivariate analysis automatically deals with correlations between dependent variables, so you don't have to worry about the problem. Sometimes MANOVA can't be used because it requires a bit more data. Sometimes repeated-measured ANOVA and MANOVA give different answers — but this means that the differences between levels of the repeated-measures factors (e.g. time 1 v. time 2; time 2 v. time 3) are correlated across subjects in some way, and that may itself be of interest.

6.7.7 Linear contrasts with a GLM

GLMs make it easy to specify linear combinations of effects to test as contrasts. For example, if you had measured subjects on each of the 7 days of the week, and you wanted to ask whether the dependent variable was different on weekdays and weekends, you could use the contrast

$$-\frac{1}{5}\text{Mon}-\frac{1}{5}\text{Tue}-\frac{1}{5}\text{Wed}-\frac{1}{5}\text{Thu}-\frac{1}{5}\text{Fri}+\frac{1}{2}\text{Sat}+\frac{1}{2}\text{Sun}$$

This contrast would be zero if the mean weekend score and the mean weekday score were the same, so it's an appropriate contrast. If your design matrix looked like this:

$$\mathbf{X}=\begin{array}{c}\\ \textit{Mon}\\ \textit{Tue}\\ \textit{Wed}\\ \textit{Thu}\\ \textit{Fri}\\ \textit{Sat}\\ \textit{Sun}\end{array}\begin{bmatrix}1&1&0&0&0&0&0&0\\1&0&1&0&0&0&0&0\\1&0&0&1&0&0&0&0\\1&0&0&0&1&0&0&0\\1&0&0&0&0&1&0&0\\1&0&0&0&0&0&1&0\\1&0&0&0&0&0&0&1\end{bmatrix}$$

then a suitable contrast matrix might look like this:

$$\mathbf{L}=\begin{bmatrix}0&-\frac{1}{5}&-\frac{1}{5}&-\frac{1}{5}&-\frac{1}{5}&-\frac{1}{5}&+\frac{1}{2}&+\frac{1}{2}\end{bmatrix}$$

This would be equally appropriate:

$$\mathbf{L}=\begin{bmatrix}0&2&2&2&2&2&-5&-5\end{bmatrix}$$

It works like this: you solve the usual GLM, $\mathbf{Y}=\mathbf{Xb}+\mathbf{e}$, to find the parameter estimates **b.** Then you calculate $L=\mathbf{Lb}$ to estimate the value of your contrast. You can then test it for significance; its sum of squares is given by the usual $\text{SS}_{\text{contrast}}=\frac{L^2}{\sum_j w_j^2/n_j}$ where w_j are the weights in the **L** matrix and n_j are the corresponding group sizes, and $\text{MS}_{\text{contrast}} = \text{SS}_{\text{contrast}}$ is compared to MS_{error}. For details, see

www.statsoft.nl / textbook / stglm.html#testing

6.7.8. GLMs in SPSS

If you run an ANOVA in SPSS, how can you see the design matrix? SPSS doesn't show you this directly, but it will show you **parameter estimates** — that is, the **b** matrix. And it labels each row of the **b** matrix with a description of the relevant column of the corresponding design matrix (**X**). To obtain this, either use the

```
/PRINT = PARAMETER
```

option or, from the menus, *Options* → *Parameter estimates*. You'll get something like this:

Parameter Estimates

Dependent Variable: DEPVAR

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	81.304	3.283	24.768	.000	74.433	88.175
C	1.342	.434	3.092	.006	.434	2.250
[A=1.00]	-77.483	1.841	-42.095	.000	-81.336	-73.631
[A=2.00]	0[a]	.	.	.	.	.
[B=1.00]	-85.667	1.063	-80.611	.000	-87.891	-83.442
[B=2.00]	0[a]	.	.	.	.	.
[A=1.00] * [B=1.00]	83.350	.868	96.058	.000	81.534	85.166
[A=1.00] * [B=2.00]	0[a]	.	.	.	.	.
[A=2.00] * [B=1.00]	0[a]	.	.	.	.	.
[A=2.00] * [B=2.00]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

The **design matrix** is specified by the `/DESIGN` command — try clicking **Paste** instead of OK when youre about to run any ANOVA and you will see the `/DESIGN` command it was going to use. Similarly, if you add `/PRINT=TEST(LMATRIX)`, you see a contrast for every term in the design matrix, which shows you the columns present in the design matrix. For example, with a two-way ANOVA, $A_2 \times B_2$, you get this:

Intercept

Parameter	Contrast L1
Intercept	1.000
[A=1.00]	.500
[A=2.00]	.500
[B=1.00]	.500
[B=2.00]	.500
[A=1.00] * [B=1.00]	.250
[A=1.00] * [B=2.00]	.250
[A=2.00] * [B=1.00]	.250
[A=2.00] * [B=2.00]	.250

A

Parameter	Contrast L2
Intercept	.000
[A=1.00]	1.000
[A=2.00]	-1.000
[B=1.00]	.000
[B=2.00]	.000
[A=1.00] * [B=1.00]	.500
[A=1.00] * [B=2.00]	.500
[A=2.00] * [B=1.00]	-.500
[A=2.00] * [B=2.00]	-.500

B

Parameter	Contrast L4
Intercept	.000
[A=1.00]	.000
[A=2.00]	.000
[B=1.00]	1.000
[B=2.00]	-1.000
[A=1.00] * [B=1.00]	.500
[A=1.00] * [B=2.00]	-.500
[A=2.00] * [B=1.00]	.500
[A=2.00] * [B=2.00]	-.500

A * B

Parameter	Contrast L6
Intercept	.000
[A=1.00]	.000
[A=2.00]	.000
[B=1.00]	.000
[B=2.00]	.000
[A=1.00] * [B=1.00]	1.000
[A=1.00] * [B=2.00]	-1.000
[A=2.00] * [B=1.00]	-1.000
[A=2.00] * [B=2.00]	1.000

The default display of this matrix is the transpose of the corresponding L matrix.
Based on Type III Sums of Squares.

Rather than simply using the null hypothesis $\mathbf{Lb} = 0$, SPSS can also test custom hypotheses with non-zero expected values for the contrast: $\mathbf{Lb} = k$, or for multiple contrasts simultaneously, with more than one row for the **L** matrix, $\mathbf{Lb} = \mathbf{K}$. This can be specified with the `/LMATRIX` and `/KMATRIX` subcommands (SPSS, 2001, p. 478-481).

For an $A_3 \times B_3$ ANOVA, the default contrasts look like this:

Intercept

Parameter	Contrast L1
Intercept	1.000
[A=1.00]	.333
[A=2.00]	.333
[A=3.00]	.333
[B=1.00]	.333
[B=2.00]	.333
[B=3.00]	.333
[A=1.00] * [B=1.00]	.111
[A=1.00] * [B=2.00]	.111
[A=1.00] * [B=3.00]	.111
[A=2.00] * [B=1.00]	.111
[A=2.00] * [B=2.00]	.111
[A=2.00] * [B=3.00]	.111
[A=3.00] * [B=1.00]	.111
[A=3.00] * [B=2.00]	.111
[A=3.00] * [B=3.00]	.111

A

Parameter	Contrast L2	Contrast L3
Intercept	.000	.000
[A=1.00]	1.000	.000
[A=2.00]	.000	1.000
[A=3.00]	-1.000	-1.000
[B=1.00]	.000	.000
[B=2.00]	.000	.000
[B=3.00]	.000	.000
[A=1.00] * [B=1.00]	.333	.000
[A=1.00] * [B=2.00]	.333	.000
[A=1.00] * [B=3.00]	.333	.000
[A=2.00] * [B=1.00]	.000	.333
[A=2.00] * [B=2.00]	.000	.333
[A=2.00] * [B=3.00]	.000	.333
[A=3.00] * [B=1.00]	-.333	-.333
[A=3.00] * [B=2.00]	-.333	-.333
[A=3.00] * [B=3.00]	-.333	-.333

B

Parameter	Contrast L5	Contrast L6
Intercept	.000	.000
[A=1.00]	.000	.000
[A=2.00]	.000	.000
[A=3.00]	.000	.000
[B=1.00]	1.000	.000
[B=2.00]	.000	1.000
[B=3.00]	-1.000	-1.000
[A=1.00] * [B=1.00]	.333	.000
[A=1.00] * [B=2.00]	.000	.333
[A=1.00] * [B=3.00]	-.333	-.333
[A=2.00] * [B=1.00]	.333	.000
[A=2.00] * [B=2.00]	.000	.333
[A=2.00] * [B=3.00]	-.333	-.333
[A=3.00] * [B=1.00]	.333	.000
[A=3.00] * [B=2.00]	.000	.333
[A=3.00] * [B=3.00]	-.333	-.333

A * B

Parameter	Contrast L8	Contrast L9	Contrast L11	Contrast L12
Intercept	.000	.000	.000	.000
[A=1.00]	.000	.000	.000	.000
[A=2.00]	.000	.000	.000	.000
[A=3.00]	.000	.000	.000	.000
[B=1.00]	.000	.000	.000	.000
[B=2.00]	.000	.000	.000	.000
[B=3.00]	.000	.000	.000	.000
[A=1.00] * [B=1.00]	1.000	.000	.000	.000
[A=1.00] * [B=2.00]	.000	1.000	.000	.000
[A=1.00] * [B=3.00]	-1.000	-1.000	.000	.000
[A=2.00] * [B=1.00]	.000	.000	1.000	.000
[A=2.00] * [B=2.00]	.000	.000	.000	1.000
[A=2.00] * [B=3.00]	-1.000	-1.000	.000	-1.000
[A=3.00] * [B=1.00]	-1.000	.000	-1.000	.000
[A=3.00] * [B=2.00]	.000	-1.000	.000	-1.000
[A=3.00] * [B=3.00]	1.000	1.000	1.000	1.000

The default display of this matrix is the transpose of the corresponding L matrix.
Based on Type III Sums of Squares.

The contrasts shown above — the default contrasts that examine the main effects of A and B and the AB interaction — could be specified by hand like this:

```
GLM DEPVAR BY A B
/LMATRIX = "Intercept"
                   all  1
                        1/3    1/3    1/3
                        1/3    1/3    1/3
                        1/9    1/9    1/9
                        1/9    1/9    1/9
                        1/9    1/9    1/9
```

```
/LMATRIX = "A"
                      a  1      0     -1
                      b  0      0      0
                    a*b  1/3    1/3    1/3
                         0      0      0
                        -1/3   -1/3   -1/3;
                      a  0      1     -1
                      b  0      0      0
                    a*b  0      0      0
                         1/3    1/3    1/3
                        -1/3   -1/3   -1/3
/LMATRIX = "B"
                      a  0      0      0
                      b  1      0     -1
                    a*b  1/3    0     -1/3
                         1/3    0     -1/3
                         1/3    0     -1/3;
                      a  0      0      0
                      b  0      1     -1
                    a*b  0      1/3   -1/3
                         0      1/3   -1/3
                         0      1/3   -1/3
/LMATRIX = "AxB"
                      a  0      0      0
                      b  0      0      0
                    a*b  1      0     -1
                         0      0      0
                        -1      0      1;
                      a  0      0      0
                      b  0      0      0
                    a*b  0      1     -1
                         0      0      0
                         0     -1      1;
                      a  0      0      0
                      b  0      0      0
                    a*b  0      0      0
                         1      0     -1
                        -1      0      1;
                      a  0      0      0
                      b  0      0      0
                    a*b  0      0      0
                         0      1     -1
                         0     -1      1
/DESIGN = A, B, A*B.
```

(You have to use `1/3` rather than `0.333` to avoid rounding errors; if the coefficients don't add up to 1 for each contrast matrix you won't get an answer.) Having seen how the general technique works, we can test advanced contrasts:

```
/LMATRIX = "B1 vs B2 at A1"
                      B  1     -1      0
                    A*B  1     -1      0
                         0      0      0
                         0      0      0
```

This would be more powerful than just analysing the A_1 data and applying a B_1 v. B_2 contrast — the $MS_{contrast}$ would be the same, but the contrast specified above uses the overall (pooled) MS_{error}, making it more powerful (more error *df*).

```
/LMATRIX = "B1 vs (B2+B3)"
                      A  0      0      0
                      B  1     -1/2   -1/2
                    A*B  1/3   -1/6   -1/6
                         1/3   -1/6   -1/6
                         1/3   -1/6   -1/6
```

Finally, a really complex one. Suppose B_1 is a control condition and B_2 and B_3 are two different selective serotonin reuptake inhibitor drugs. Therefore, (B_1) versus (B_2 and B_3) might represent an 'SSRI treatment effect' (call it *T*) that we're interested in. Suppose that A_1 and A_2 are depressives and schizophrenics. If we want to compare the SSRI effect between depressives (T_{A1}) and schizophrenics (T_{A2}), we could follow this logic:

$$T_{A1} = \mu_{B1,A1} - \frac{\mu_{B2,A1} + \mu_{B3,A1}}{2}$$

$$T_{A2} = \mu_{B1,A2} - \frac{\mu_{B2,A2} + \mu_{B3,A2}}{2}$$

$$H_0 : T_{A1} = T_{A2}$$

$$H_0 : T_{A1} - T_{A2} = 0$$

$$H_0 : \mu_{B1,A1} - \frac{\mu_{B2,A1} + \mu_{B3,A1}}{2} - \mu_{B1,A2} + \frac{\mu_{B2,A2} + \mu_{B3,A2}}{2} = 0$$

Having calculated our null hypothesis, we can specify the contrast:

```
/LMATRIX = "(B1 vs (B2+B3)) - at A1 versus A2"
                   A*B  1     -1/2   -1/2
                       -1      1/2    1/2
                        0      0      0
```

Hope I've got that right; it seems to work.

6.8 Effect size

Whether a contribution is *significant* or not does not tell you whether that significant contribution is *large*. If you have high power (large *n*), you may be able to measure significant small effects. And if you have lower power (small *n*), you may 'miss' (fail to declare as significant) large effects. **To ask about effect size is to ask not just whether the effect of a predictor variable is statistically significant, but *how big (important)* its effect is.**

In general, when we are predicting a dependent variable *Y* by one or more predictor variables, be they continuous (ANCOVA, multiple regression) or discrete (ANOVA factors), we can ask to what extent a given term (main effect, interaction, etc.) contributes to the prediction of the dependent variable. We've already seen that this can be complicated, especially if the predictors are themselves correlated — effect size is a fairly complex topic (Winer, 1971, pp. 405-415; Keppel, 1991, pp. 64-68, 221-224, 437-440; Myers & Well, 1995, pp. 111-113, 252-256, 504-509; Howell, 1997, pp. 330-334, 426-429, 544-546).

We'll start by examining effect size in the context of multiple regression (predicting *Y* from X_1, X_2, and so on), because it's the simplest conceptually. In general, effect size can refer to **the size of the change in *Y* that follows a certain change in a predictor** (regression slope) or **the proportion of variation in *Y* explicable by a predictor** (equivalent to r^2 in simple linear regression).

6.8.1. Effect size in the language of multiple regression

A reminder of what the 'significance' of a predictor means

Assuming you use the usual (SPSS Type III) way of partitioning sums of squares with correlated predictors, the significance test of each predictor reflects **whether that predictor contributes to the prediction *over and above* all the other predictors in the model** (see also sections on correlated predictors earlier: p. 70→ and p. 86→). This is *not* effect size.

Interpreting the effects of individual predictors: the regression slope, b

The computerized results will give us individual slope parameters for each of the effects in our model (in SPSS, tick **Options → Parameter estimates**). Remember that a multiple regression equation looks like this:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots$$
$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}$$

The parameters are the values of *b*. The first, b_0, is the intercept (grand mean). The others reflect the effects of all the other predictors. However, there are problems of interpretation of the individual slope parameters b_j (Myers & Well, 1995, p. 522; Howell, 1997, pp. 510-532 and 544-546). It is tempting to think that if we were to change X_j by one unit, *Y* would change by b_j units — this would be true of simple linear regression (with one predictor). However, a regression coefficient b_j does not reflect the total effect of X_j on *Y*. Rather, it reflects the *direct* effect of X_j on *Y* — the rate of change of *Y* with X_j *holding all of the other variables in the equation constant.* If the various predictors (X_1, X_2, …) are mutually correlated (known as **collinearity** or **multicollinearity**), it may often not make a great deal of sense to ask this question — for example, if we were predicting car crash fatalities by drivers' annual mileage and drivers' annual fuel consumption, it's not clear what it would mean to change annual mileage while holding fuel consumption constant. When we ask about the consequences of changing X_j, we must be concerned not only with the direct effect but also with the *indirect effects* — the effects on *Y* that occur because of changes in the other variables. Given a valid causal model, **path analysis** can be used to calculate the total effect (direct + indirect effects) of changing a variable. However, if the model is incomplete or invalid, we would have to establish the ef-

fects of changing X_j experimentally, by manipulating it without confounding it with the other variables, and observing the results.

Standardized regression slope, β (= r)

The standardized regression slope, β_j, is simply the b_j that you would obtain if both the dependent variable Y and the predictor X_j were standardized — that is, transformed so that they have a mean of 0 and a standard deviation of 1 (Howell, 1997, pp. 244, 517-8, 544-6). If $b = 0.75$, then a one unit increase in X would be reflected in an 0.75 unit increase in Y. If $\beta = 0.75$, then a one standard deviation increase in X would be reflected in an 0.75 standard deviation increase in Y. It's easy to calculate β. If b_j and s_j are the regression slope and standard deviation of a predictor X_j, then

$$\beta_j = \frac{b_j s_j}{s_Y}$$

Bear in mind that slopes are related to r: for simple linear regression,

$$b = r\frac{s_Y}{s_X}$$

and so $b = r$ when both variables are standardized (Howell, 1997, p. 242), and $\beta = r$ at all times.

However, with multiple predictors, the problem with β_j is just the same as for b_j: it reflects the change in Y associated with a change in X_j *holding all other predictors constant*, and if the predictors are correlated this may not make much sense.

Overall R^2 and R^2_{adj}: how good is the whole model?

The computerized results of an ANOVA, ANCOVA, or other GLM will give an *overall* R^2, which reflects the proportion of total Y variance predicted by all the predictors together, i.e. $SS_{regression} / SS_{total}$. (Alternatively, we could say that R is the correlation between the dependent variable and the best linear combination of the predictors.) R^2 can also be adjusted (downwards) to give R^2_{adj}, a better estimate of the corresponding population parameter (Myers & Well, 1995, p. 508-509; Howell, 1997, p. 521), and SPSS will do that automatically. If there are N observations and p continuous predictors:

$$R^2_{adj} = 1 - (1 - R^2)\left(\frac{N-1}{N-1-p}\right)$$

If you are using predictors with >1 *df* per predictor (e.g. factors with >2 levels), you need a more general form of this equation, which I believe is

$$R^2_{adj} = 1 - (1 - R^2)\left(\frac{df_{\text{total}}}{df_{\text{error}}}\right)$$

Assessing the importance of individual predictors: $r^2_{\text{semipartial}}$ — a good one

Let's move on to a better measure (Myers & Well, 1995, pp. 505-508; Howell, 1997, pp. 528-531, 544-546). When we predict Y from p predictor variables, $R^2_{Y.1,2\ldots p}$ (or simply R^2) is the proportion of variability in Y accounted for by the regression on all p predictors. If the p predictors are not mutually correlated (Myers & Well, 1995, p. 505), $SS_{regression}$ can be partitioned into nonoverlapping components from each of the predictors:

$$\begin{aligned} SS_{\text{regression}} &= SS_{Y.1} + SS_{Y.2} + \ldots + SS_{Y.p} \\ &= r^2_{Y.1}SS_Y + r^2_{Y.2}SS_Y + \ldots + r^2_{Y.p}SS_Y \end{aligned}$$

where $SS_{Y.1}$ is the proportion of variability of Y accounted for by the predictor X_j, and $r^2_{Y.j}$ is the correlation between X_j and Y. Since

$$SS_{\text{regression}} = R^2_{Y.1,2\ldots p}SS_Y$$

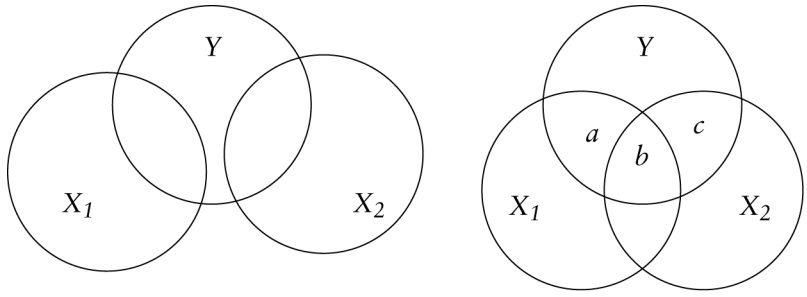
it follows that for uncorrelated predictors,

$$\begin{aligned} R^2_{Y.1,2\ldots p} &= r^2_{Y.1} + r^2_{Y.2} + \ldots r^2_{Y.p} \\ &= \sum_j r^2_{Y.j} \end{aligned}$$

If the predictors are correlated, we must use this (Myers & Well, 1995, p. 506):

$$R^2_{Y.1,2\ldots p} = \frac{\sum_j r^2_{Y.j} b_j \hat{\sigma}_j}{\hat{\sigma}_Y}$$

where b_j is the regression coefficient of X_j in the multiple regression equation and $\hat{\sigma}_j$ and $\hat{\sigma}_Y$ are the standard deviations of X_j and Y, respectively. The **increase in R^2 when X_2 is added to a regression equation that already contains X_1 is $r^2_{Y.(2|1)}$, the square of the semipartial correlation coefficient** (Myers & Well, 1995, p. 486 and 507). Here's a visual interpretation, in which the area of each circle represents the total variability of a given variable:



Uncorrelated predictors Correlated predictors

$$\begin{aligned} R^2_{Y.1} &= a + b \\ R^2_{Y.2} &= b + c \\ R^2_{Y.1,2} &= a + b + c \end{aligned} \qquad \begin{aligned} r^2_{Y.(2|1)} &= c \\ r^2_{Y.(1|2)} &= a \end{aligned}$$

You could also say that the semipartial correlation $r_{Y.(2|1)}$ is the correlation of Y with that part of X_2 that is independent of X_1 (Howell, 1997, p. 528). In general, $r^2_{Y.(p+1|1,2\ldots p)}$ is the increase in R^2 that follows from adding X_{p+1} to a regression equation that already includes X_1, X_2, … X_p. That is,

$$\begin{aligned} R^2_{Y.1,2\ldots p+1} &= R^2_{Y.1,2\ldots p} + r^2_{Y.(p+1|1,2\ldots p)} \\ r^2_{Y.(p+1|1,2\ldots p)} &= R^2_{Y.1,2\ldots p+1} - R^2_{Y.1,2\ldots p} \end{aligned}$$

This would seem to be a useful measure. Howell (1997, p. 544-6) agrees, stating that 'when the main goal is prediction rather than explanation, this is probably the best

measure of "importance".' If the computer package doesn't give it directly (and SPSS doesn't), it can easily be calculated (Howell, 1997, p. 546):

$$r^2_{Y.(i|1,2\ldots\text{everything except } i\ldots p)} = \frac{F_i(1-R^2_{Y.1,2\ldots p})}{N-p-1}$$

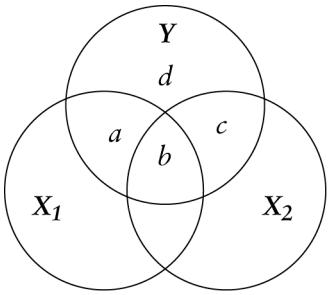
where p is the total number of predictors, $r^2_{Y.(i|1,2\ldots\text{everything except } i\ldots p)}$ is the squared semipartial correlation for predictor i, F_i is the F test for predictor i (use $F = t^2$ if your stats package reports t instead), $R^2_{Y.1,2\ldots p}$ is the overall R^2 (with predictor i included), and N is the total number of observations. Note that this means that the F statistics in an ANOVA are in the same order as the the squared semipartial correlation coefficients (within an ANOVA, you could say that 'bigger $F \Rightarrow$ more important'). If you're using factors as predictors (i.e. predictors with >1 *df* per predictor), I rather suspect that Howell's formula should be rewritten like this:

$$r^2_{Y.(i|1,2\ldots\text{everything except } i\ldots p)} = \frac{F_i(1-R^2_{Y.1,2\ldots p})}{df_{\text{error}}}$$

But if you're having trouble working out a formula, you can always fall back to the position of running the ANOVA with and without a particular term, and calculating the difference between the two overall R^2 values.

Partial and semipartial correlations

It's easy to be confused by the difference between partial and semipartial correlations. We've just seen what the semipartial correlation is (Howell, 1997, pp. 526-531). Let's go back to the Venn diagram:

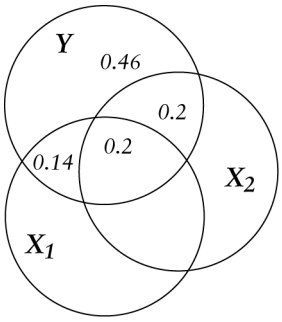


The squared semipartial correlation coefficient $r^2_{Y.(2|1)}$ is the proportion of the variability in Y explained by X_2 over and above what's explained by X_1. The squared *partial* correlation coefficient $r^2_{Y.2|1}$ is the proportion of the variability in Y explained by X_2 *relative* to that *not* explained by X_1. In our Venn diagram, the two look like this:

Overall prediction of models	Squared semipartial	Squared partial
$R^2_{Y.1} = a+b$		
$R^2_{Y.2} = b+c$	$r^2_{Y.(2\|1)} = c$	$r^2_{Y.2\|1} = \frac{c}{c+d}$
$R^2_{Y.1,2} = a+b+c$	$r^2_{Y.(1\|2)} = a$	$r^2_{Y.1\|2} = \frac{a}{a+d}$
$1-R^2_{Y.1,2} = d$		

Suppose that $R^2_{Y.2} = 0.4$, the squared semipartial $r^2_{Y.(2|1)} = 0.2$ and the squared partial $r^2_{Y.2|1} = 0.3$. That would mean that X_2 explains 40% of the variability in Y if it's the only predictor, that X_2 explains 20% of the variability in Y once X_1 has been taken into account (semipartial), and that X_2 explains 30% of the variability in Y that X_1

failed to explain (partial). That would reflect this situation (areas denote variability; figures are proportions of the total variability of *Y*; sorry if it's not quite to scale):



Another definition of partial correlation

If r_{xy} is the correlation between *X* and *Y*, then $r_{xy|z}$, the partial correlation between *X* and *Y* with the effects of *Z* partialed out, is the correlation between *X*|*Z* and *Y*|*Z*, where $X \mid Z = X - \hat{X}$ is the residual that results when *X* is regressed on *Z*, and $Y \mid Z = Y - \hat{Y}$ is the residual that results when *Y* is regressed on *Z* (Myers & Well, 1995, p. 483). It's possible to obtain $r_{xy|z}$ from the simple correlations between each of the variables:

$$r_{xy|z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}$$

For example, suppose we look at 48 US states and measure population, motor vehicle deaths, and whether or not the state has enacted seat belt legislation — the last being a dichotomous variable, but that's OK (Myers & Well, 1995, p. 483). There's a positive correlation between deaths and belt legislation (+0.309), which might seem worrying. However, $r_{\text{deaths,population}} = +0.928$ and $r_{\text{belts,population}} = +0.345$ — larger states have more deaths, and larger states are more likely to have seat belt legislation. The partial correlation $r_{\text{deaths,belts|population}} = -0.032$, indicating a small but negative relationship between seat belt laws and motor vehicle deaths once the effects of population have been partialled out.

Another definition of semipartial correlation

The semipartial correlation coefficient $r_{y(x|z)}$ is the correlation between *Y* and *X*|*Z*, where $X \mid Z = X - \hat{X}$ is the residual that results when *X* is regressed on *Z*. It too can be calculated from the simple regression coefficients:

$$r_{y(x|z)} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)}}$$

6.8.2. Effect size in the language of ANOVA

The effect size in the context of ANOVA is the same thing as the effect size in multiple regression (since both are simply instances of a GLM), but people tend to use different terminology. A helpful discussion of some different measures of effect size is given at web.uccs.edu / lbecker / SPSS / glm_effectsize.htm.

Difference between level means

This is simple. If you have a factor (e.g. Sex: Male / Female) and you establish through an ANOVA that its effect is significant, you have an instant measure of its effect: the difference between μ_{male} and μ_{female}. You can extend this approach to multiple factors and to interactions. For example, for the data shown below, we can state

the effect sizes very simply. **Overall mean:** The overall mean ('being a male or female 10- or 10-year-old') is 137 cm. **Main effects:** Maleness contributes +3.5 cm and femaleness contributes –3.5 cm (or, maleness contributes +7 cm compared to femaleness). Being a 10-year-old contributes –33.5 cm; being a 20-year-old contributes +33.5 cm (or, being a 20-year-old contributes +67 cm relative to being a 10-year-old). **Interactions:** if the overall mean contributes 137 cm, being a male contributes +3.5 cm, and being a 20-year-old contributes +33.5 cm, we'd expect 20-year-old males to be 174 cm, but in fact they're 177 cm, so the interaction term (being a male 20-year-old) contributes an extra +3 cm on top of the main effects. And so on.

Height	Male	Female	mean
10-year-old	104 cm	103 cm	103.5 cm
20-year-old	177 cm	164 cm	170.5 cm
mean	140.5 cm	133.5 cm	137 cm

Effect size measures related to the difference between means — perhaps best to skip this bit!

There are lots of these, most designed to facilitate calculation of **power.** For a situation with two groups with the same standard deviation, we can measure the difference between means $\mu_2 - \mu_1$. We can 'standardize' that by dividing by the standard deviation to produce **d,** often called the 'effect size':

$$\mathbf{d} = \frac{\mu_2 - \mu_1}{\sigma}$$

This number **d** can be combined with knowledge of the sample size *n* to calculate $\delta = \mathbf{d}\sqrt{n}$, which in turn can be used to calculate power (Myers & Well, 1995, pp. 113-116; Howell, 1997, p. 216-226). Cohen (1988) more or less arbitrarily called **d** = 0.2 a small effect (the means differ by 0.2 of a standard deviation), 0.5 a medium effect, and 0.8 a large effect. Similar principles can be applied to ANOVA (Howell, 1997, p. 334-340), but the notation is a bit different. If there are *k* levels for a factor, the standardized measure of effect size is

$$\phi' = f = \frac{\sigma_{\text{treatment}}}{\sigma_{\text{error}}} = \sqrt{\frac{\frac{\sum_i (\mu_i - \mu)^2}{k}}{\sigma^2_{\text{error}}}}$$

This can then be combined with knowledge of the sample size to calculate $\phi = \phi'\sqrt{n}$, which in turn can be used to calculate power. This can be extended to factorial designs (Myers & Well, 1995, pp. 147-149). And just as correlation slopes *b* were related to r^2 in the language of regression, ϕ' (also written *f*) is related to η^2 (see below) in the language of ANOVA (Winer *et al.*, 1991, p. 124):

$$f^2 = \frac{\eta^2}{1 - \eta^2}$$

δ and ϕ are also known as *noncentrality parameters* (Winer *et al.*, 1991, pp. 126-140; Howell, 1997, pp. 220, 334-5). This refers to the fact that if there *is* an effect (if the null hypothesis is false), the distribution of *F* statistics isn't the plain *F* distribution (as it would be if the null hypothesis were true), but a shifted (noncentral) *F* distribution. The noncentrality parameters measure effect size by how much the distribution is shifted.

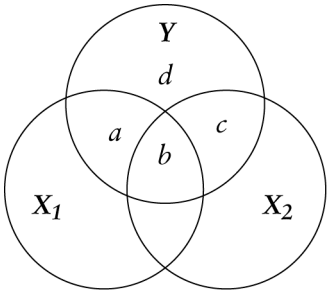
Assessing the importance of individual predictors: η^2

Eta-squared is given by

$$\eta^2 = \frac{SS_{effect}}{SS_{total}}$$

η^2 represents the proportion of total variation accounted for by a factor; equivalently, the proportion by which error is reduced when you use the factor to predict the dependent variable (Howell, 1997, p. 332). Eta itself, η, is called the correlation ratio (Winer *et al.*, 1991, p. 123), although η^2 is also sometimes called the correlation ratio (Howell, 1997, p. 331).

If you only have one predictor, $\eta^2 = R^2$. If you have more than one predictor and they're correlated, η^2 depends on how you calculate SS_{effect}. Assuming you use the usual (SPSS Type III) method (see p. 70→), the SS for predictor X_2 in the diagram below is area *c*, and SS_{total} (SS_Y) is area $a + b + c + d$.



So for our usual (SPSS Type III) sums-of-squares method, the η^2 for X_2 is

$$\eta^2_2 = \frac{SS_{effect}}{SS_{total}} = \frac{c}{a+b+c+d} = c = r^2_{Y.(2|1)}$$

so η^2 is the squared semipartial correlation coefficient, it seems to me. If you calculate η^2 by hand in SPSS, remember that what we normally refer to as SS_{total}, $\sum (y - \bar{y})^2$, is labelled 'corrected total' by SPSS. (Its 'total' is $\sum y^2$, which we're not interested in.)

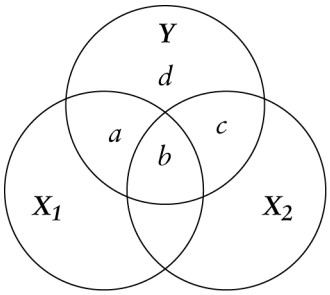
Assessing the importance of individual predictors: $\eta^2_{partial}$ — not very helpful

One measure of the importance of individual predictors is the *partial eta-squared* coefficient, which is something that SPSS gives you (tick **Options → Estimates of effect size**). We've just seen what η^2 is (above). The *partial* eta-squared is 'an overestimate of the effect size in an *F* test' (SPSS, 2001, p. 475). Specifically, it's this:

$$\eta^2_{partial} = \frac{df_{effect} \times F}{df_{effect} \times F + df_{error}}$$
$$= \frac{SS_{effect}}{SS_{effect} + SS_{error\ term\ for\ that\ effect}}$$

The top formula is from SPSS (2001, p. 475) and the second from web.uccs.edu / lbecker / SPSS / glm_effectsize.htm. I'm not sure if it's particularly useful, especially as the partial eta-squared terms sum to more than one ($\sum \eta^2_{partial} > 1$), which is pretty daft. In terms of our Venn diagram, the $\eta^2_{partial}$ for X_2 is:

$$\eta^2_{partial_2} = \frac{SS_{treatment}}{SS_{treatment} + SS_{error}} = \frac{c}{c+d} = r^2_{Y.2|1}$$



so η^2_{partial} is the squared partial correlation coefficient, it seems to me. Therefore, I'll ignore it.

Another one: ω^2

When a factor A predicts a dependent variable Y, omega-squared (ω^2) for A is defined as the proportion of the total variance in *Y* attributable to the effects of A (Myers & Well, 1995, p. 113). In general, the estimated ω^2, written $\hat{\omega}^2$, is

$$\hat{\omega}^2 = \frac{\hat{\sigma}^2_{\text{A}}}{\hat{\sigma}^2_{\text{Y}}}$$

For a **fixed** (not a random) effect A, ω^2 is estimated by

$$\hat{\omega}^2 = \frac{\text{SS}_{\text{A}} - df_{\text{A}} \times \text{MS}_{\text{error}}}{\text{MS}_{\text{error}} + \text{SS}_{\text{total}}}$$

(Formula from web.uccs.edu/lbecker/SPSS/glm_effectsize.htm.) For random effects, such as in **within-subjects** (repeated measures) designs, the definition of $\hat{\omega}^2$ depends on the specific ANOVA model (Myers & Well, 1995, pp. 252-256; Howell, 1997, pp. 426-429), and sometimes it cannot be estimated exactly (Myers & Well, 1995, p. 254).

And another: the intraclass correlation ρ_I

The intraclass correlation coefficient is a measure of association between the independent and dependent variables for a random-effects model (Howell, 1997, p. 334) (web.uccs.edu/lbecker/SPSS/glm_effectsize.htm); for an effect A, it's

$$\rho_I = \frac{\text{MS}_{\text{A}} - \text{MS}_{\text{error}}}{\text{MS}_{\text{A}} + df_{\text{A}} \times \text{MS}_{\text{error}}}$$

The squared intraclass correlation, ρ^2_I, is a version of ω^2 for the random model.

Which one to use?

Although η^2 is perhaps the simplest, it does have a problem (Howell, 1997, pp. 333-334). When it's applied to *population* data, it's correct; when applied to *samples* (as we normally do), it's biased as a measure of the underlying population effect size. So $\hat{\omega}^2$ is generally preferred when we want an estimate of the effect size in the population — the way it's calculated takes account of sample size appropriately (so $\hat{\omega}^2$ will always be smaller than η^2 or η^2_{partial}). On the other hand, SPSS doesn't produce it, which is a bit of a shame, and it's laborious to calculate by hand. So for a quick idea, η^2 is perhaps easiest. This also has an advantage over η^2_{partial} in that it's additive (the η^2 values sum to 1, while the η^2_{partial} values can sum to >1) and is therefore perhaps easier to conceptualize and interpret.

Part 7: specific designs

For designs 1–17, all factors other than 'subject' factors are assumed to be fixed. If you wish to use other random factors, see Myers & Well (1995, p. 262) or just tell SPSS that's what you want and trust it to sort out the maths.

Design (BS = between-subjects; WS = within-subjects)	**Description (in most economical format; S = subjects; 'cov' subscript = covariate)**	**Between-subjects factor(s) or covariate(s)**	**Within-subjects factors(s) or covariates)**
1 – One BS factor Includes step-by-step instructions for performing between-subjects analysis in SPSS	$A \times S$	A	–
2 – Two BS factors	$A \times B \times S$	A, B	–
3 – Three BS factors	$A \times B \times C \times S$	A, B, C	–
4 – One WS factor	$(U \times S)$	–	U
5 – Two WS factors	$(U \times V \times S)$	–	U, V
6 – Three WS factors	$(U \times V \times W \times S)$	–	U, V, W
7 – One BS and one WS factor Includes step-by-step instructions for performing within-subjects (repeated measures) analysis in SPSS	$A \times (U \times S)$	A	U
8 – Two BS factors and one WS factor	$A \times B \times (U \times S)$	A, B	U
9 – One BS factor and two WS factors	$A \times (U \times V \times S)$	A	U, V
10 – Higher-order designs along the same principles and summary of designs 1–9	See text	See text	See text
11 – One BS covariate (linear regression)	$C_{cov} \times S$	C_{cov}	–
12 – One BS covariate and one BS factor	$C_{cov} \times A \times S$	C_{cov}, A	–
13 – One BS covariate and two BS factors	$C_{cov} \times A \times B \times S$	C_{cov}, A, B	–
14 – Two or more BS covariates (multiple regression)	$C_{cov} \times D_{cov} \times \ldots \times S$	C_{cov}, D_{cov}, …	–
15 – Two or more BS covariates and one or more BS factors	e.g. $C_{cov} \times D_{cov} \times A \times B \times S$	C_{cov}, D_{cov}, A, B, etc.	–
16 – One WS covariate	$(C_{cov} \times S)$	–	C_{cov}
17 – One WS covariate and one BS factor	$A \times (C_{cov} \times S)$	A	C_{cov}
18 – Hierarchical designs	See text (complex)	See text (complex)	See text (complex)
19 – Latin square designs	See text (complex)	See text (complex)	See text (complex)
20 – Agricultural designs	See text (complex)	See text (complex)	See text (complex)

7.1 One between-subjects factor

Alternative names

- One-way ANOVA
- Completely randomized design (CRD)

Example

Subjects are assigned at random to drug treatments A1, A2, or A3 (*completely randomized design; single factor with three levels*) and their reaction time is measured on some task (*dependent variable*). Does the drug treatment affect performance?

A researcher wishes to test the effectiveness of four fertilizers (A1, A2, A3, A4). He divides his field into sixteen plots (equivalent to 'subjects' or 'replications') and randomly assigns fertilizer A1 to four replications, A2 to four replications, and so on.

Notes

For two levels of the factor, this is equivalent to an unpaired (independent sample) *t* test. Treatments (levels of the factor) are assigned at random to subjects (replications). For full details, see Howell (1997, chapter 11).

Model description (S = subjects)

depvar = A × S

Model

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

where

- Y_{ij} is the dependent variable for subject *j* experiencing level *i* of the factor
- μ is the overall mean
- α_i is the contribution from a particular level (level *i*) of the factor: $\alpha_i = \mu_i - \mu$ and $\sum_i \alpha_i = 0$. The null hypothesis is that all values of α_i are zero.
- ε_{ij} is everything else (the 'uniqueness' of subject *j* in condition *i*, 'error', 'individual variation', etc.): $\varepsilon_{ij} = Y_{ij} - \mu_i$. We assume ε_{ij} is normally distributed with mean 0 and variance σ_e^2.

Sources of variance

Analysis of variance discards constant terms (like μ) and examines the *sources of variability (variance).* Writing this in terms of sums of squares (SS),

$$SS_{total} = SS_A + SS_{error}$$

where SS_{total} is the total variability, SS_{factor} is the variability attributable to the factor, and SS_{error} is the 'error' variability (everything that's left over). Alternatively, we could write

$$SS_{total} = SS_A + SS_{S/A}$$

because our total variability is made up of variability due to factor A, and variability due to inter-subject differences within each level of A ('S within A', or 'S / A').

ANOVA table

In all cases, the mean square (MS) is the sum of squares (SS) for a particular row divided by the degrees of freedom (d.f.) for the same row. Assuming the same number of subjects *n* for each level of factor A, we have

Source	d.f.	SS	*F*
A	$a-1$	SS_A	MS_A / MS_{error}
Error (S / A)	$a(n-1)$	SS_{error}	
Total	$N-1 = an-1$	SS_{total}	

where *a* is the number of levels of factor A, *N* is the total number of observations (subjects), and *n* is the number of subjects (or 'replications') *per level* of factor A. Note that the error is sometimes written 'S / A', i.e. 'subjects within A'.

SPSS technique

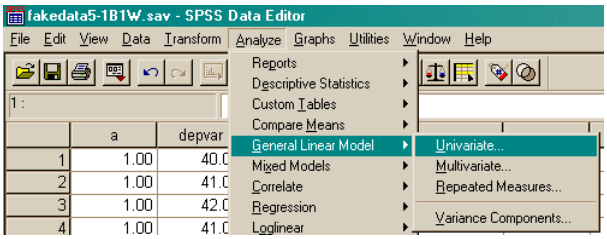
Data layout:

depvar	*A*
datum	*level_1*
datum	*level_1*
datum	*level_1*
datum	*level_2*
datum	*level_2*
…	

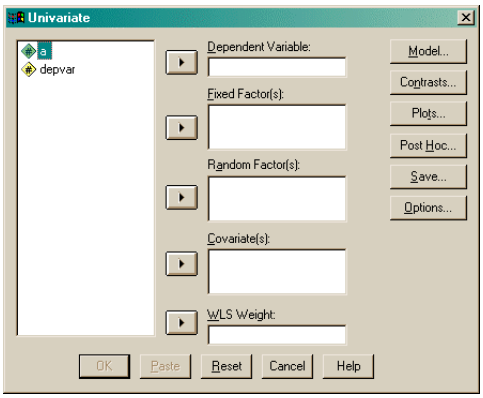
Syntax:

```
UNIANOVA
  depvar BY A
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = A .
```

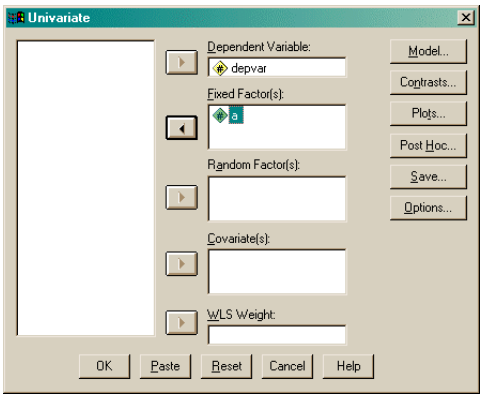
Using the menus, choose **Analyze → General Linear Model → Univariate.**



We now see this:



Our dependent variable is *depvar*; our (fixed) factor is A:



Once everything else is OK, click ‘OK’ to run the analysis, or ‘Paste’ to copy the syntax for the analysis to a syntax window.

7.2 Two between-subjects factors

Alternative names

Two-way ANOVA

- Factorial ANOVA
- $a \times b$ factorial ANOVA (where *a* and *b* are the number of levels of factors A and B; e.g. '2 × 5 factorial')
- Factorial, completely randomized design ANOVA

Example

Subjects are assigned at random to a high-arousal (A1) or a low-arousal (A2) situation, and are also given drug (B1) or placebo (B2) (*completely randomized design; 2 × 2 factorial ANOVA).* Their performance is measured on a task (*dependent variable*). Does the arousal situation (A) or the drug (B) affect performance, and does the effect of the drug depend on arousal (A × B interaction)?

Notes

A factorial design is one in which every level of every factor is paired with every level of every other factor (Howell, 1997, p. 401).

Model description (S = subjects)

depvar = A × B × S

Model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

where

- Y_{ijk} is the dependent variable in condition A_i, B_j for subject *k*
- μ is the overall mean
- α_i is the contribution from level *i* of factor A (A_i): $\alpha_i = \mu_{A_i} - \mu$ and $\sum \alpha_i = 0$.
- β_j is the contribution from level *j* of factor B (B_j): $\beta_j = \mu_{B_j} - \mu$ and $\sum \beta_j = 0$.
- $\alpha\beta_{ij}$ is the contribution from the interaction of level *i* of factor A and level *j* of factor B — that is, the degree to which the mean of condition A_iB_j deviates from what you'd expect based on the overall mean and the separate contributions of A_i and B_j (= the interaction A × B), i.e. $\alpha\beta_{ij} = \mu_{A_iB_j} - (\mu + \alpha_i + \beta_j)$. By this definition, $\sum_i \alpha\beta_{ij} = \sum_j \alpha\beta_{ij} = 0$.
- ε_{ijk} is everything else (the 'uniqueness' of subject *k* in condition *i* of factor A and condition *j* of factor B, 'error', 'individual variation', etc.): $\varepsilon_{ijk} = Y_{ijk} - (\mu_j + \alpha_i + \beta_j + \alpha\beta_{ij})$. By our usual assumption of normal distribution of error, ε_{ijk} is normally distributed with mean 0 and variance σ_e^2.

Sources of variance

As before, we consider only the sources of *variation* for the ANOVA analysis:

$$SS_{total} = SS_A + SS_B + SS_{A\times B} + SS_{error}$$

where

- SS_{total} is the total variability
- SS_A is the variability attributable to factor A
- SS_B is the variability attributable to factor B
- $SS_{A\times B}$ is the variability attributable to the interaction
- SS_{error} is the 'error' variability (everything that's left over). This is sometimes written $SS_{S/AB}$ (indicating variability due to inter-subject variation within A × B combinations).

ANOVA table

In all cases, the mean square (MS) is the sum of squares (SS) for a particular row divided by the degrees of freedom (d.f.) for the same row. Assuming the same number of subjects *n* for each cell (combination of one level of factor A and one level of factor B), we have

Source	d.f.	SS	*F*
A	$a-1$	SS_A	MS_A / MS_{error}
B	$b-1$	SS_B	MS_B / MS_{error}
A × B	$(a-1)(b-1)$	$SS_{A\times B}$	$MS_{A\times B} / MS_{error}$
Error (S / AB)	$ab(n-1)$	SS_{error}	

Total	N–1 = abn–1	SS_{total}

where a is the number of levels of factor A, N is the total number of observations (subjects), and n is the number of subjects (or 'replications') per cell.

SPSS technique

Data layout:

depvar	*A*	*B*
datum	*level_1*	*level_1*
datum	*level_1*	*level_1*
datum	*level_1*	*level_2*
datum	*level_1*	*level_2*
datum	*level_2*	*level_1*
datum	*level_2*	*level_1*
datum	*level_2*	*level_2*
datum	*level_2*	*level_2*
…		

Syntax:

```
UNIANOVA
  depvar BY a b
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = a b a*b .
```

Using the menus, choose **Analyze → General Linear Model → Univariate.** Enter A and B as between-subjects factors.

7.3 Three between-subjects factors

Alternative names

- $a \times b \times c$ factorial ANOVA (where *a*, *b* and *c* are the number of levels of factors A, B, and C; e.g. '$2 \times 5 \times 3$ factorial')
- Factorial, completely randomized design ANOVA

Example

Subjects have their prefrontal cortex destroyed (A1) or not (A2) or have a special prefrontal cortex augmenter fitted (A3), are assigned at random to a high-arousal (B1) or a low-arousal (B2) situation, and are also given drug (C1) or placebo (C2) (*completely randomized design; 3 × 2 × 2 factorial ANOVA).* Their performance is measured on a task (*dependent variable*). Do factors A, B, or C affect performance? Do they interact?

Notes

Model description (S = subjects)

depvar = A × B × C × S

Model

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \varepsilon_{ijkl}$$

where

- X_{ijkl} is the dependent variable in condition A_i, B_j, C_k for subject *l*
- μ is the overall mean
- α_i is the contribution from level *i* of factor A: $\alpha_i = \mu_{A_i} - \mu$
- β_j is the contribution from level *j* of factor B: $\beta_j = \mu_{B_j} - \mu$
- γ_k is the contribution from level *k* of factor C: $\gamma_k = \mu_{C_k} - \mu$
- $\alpha\beta_{ij}$ is the contribution from the interaction of level *i* of factor A and level *j* of factor B: $\alpha\beta_{ij} = \mu_{A_iB_j} - (\mu + \alpha_i + \beta_j)$
- $\alpha\gamma_{ik}$ is the contribution from the interaction of level *i* of factor A and level *k* of factor C: $\alpha\gamma_{ik} = \mu_{A_iC_k} - (\mu + \alpha_i + \gamma_k)$
- $\beta\gamma_{jk}$ is the contribution from the interaction of level *j* of factor B and level *k* of factor C: $\beta\gamma_{jk} = \mu_{B_jC_k} - (\mu + \beta_j + \gamma_k)$
- ε_{ijkl} is everything else (the 'uniqueness' of subject *l* in condition *i* of factor A and condition *j* of factor B and condition *k* of factor C, 'error', 'individual variation', etc.): $\varepsilon_{ijk} = Y_{ijk} - (\mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk})$.

Sources of variance

As before, we consider only the sources of *variation* for the ANOVA analysis:

$$SS_{total} = SS_A + SS_B + SS_C + SS_{A\times B} + SS_{A\times C} + SS_{B\times C} + SS_{A\times B\times C} + SS_{error}$$

where

- SS_{total} is the total variability
- SS_A is the variability attributable to factor A
- SS_B is the variability attributable to factor B
- SS_C is the variability attributable to factor C
- $SS_{A\times B}$ is the variability attributable to the A × B interaction
- $SS_{A\times C}$ is the variability attributable to the A × C interaction
- $SS_{B\times C}$ is the variability attributable to the B × C interaction
- $SS_{A\times B\times C}$ is the variability attributable to the A × B × C interaction
- SS_{error} is the 'error' variability (everything that's left over). This is sometimes written $SS_{S/ABC}$ (indicating variability due to inter-subject variation within A × B × C combinations).

ANOVA table

In all cases, the mean square (MS) is the sum of squares (SS) for a particular row divided by the degrees of freedom (d.f.) for the same row. Assuming the same number of subjects *n* for each cell (combination of one level of factor A, one level of factor B, and one level of factor C) we have

Source	d.f.	SS	*F*
A	a–1	SS_A	MS_A / MS_{error}
B	b–1	SS_B	MS_B / MS_{error}
C	c–1	SS_C	MS_C / MS_{error}
A × B	(a–1)(b–1)	$SS_{A\times B}$	$MS_{A\times B} / MS_{error}$
A × C	(a–1)(c–1)	$SS_{A\times C}$	$MS_{A\times C} / MS_{error}$
B × C	(b–1)(c–1)	$SS_{B\times C}$	$MS_{B\times C} / MS_{error}$
A × B × C	(a–1)(b–1)(c–1)	$SS_{A\times B\times C}$	$MS_{A\times B\times C} / MS_{error}$
Error (S / ABC)	abc(n–1)	SS_{error}	
Total	N–1 = $abcn$–1	SS_{total}	

where a is the number of levels of factor A (etc.), N is the total number of observations (subjects), and n is the number of subjects (or 'replications') per cell.

SPSS technique

Data layout:

depvar	*A*	*B*	*C*
datum	*level_1*	*level_1*	*level_1*
datum	*level_1*	*level_1*	*level_1*
datum	*level_1*	*level_2*	*level_1*
datum	*level_1*	*level_2*	*level_1*
datum	*level_2*	*level_1*	*level_1*
datum	*level_2*	*level_1*	*level_1*
datum	*level_2*	*level_2*	*level_1*
datum	*level_2*	*level_2*	*level_1*
datum	*level_1*	*level_1*	*level_2*
datum	*level_1*	*level_1*	*level_2*
datum	*level_1*	*level_2*	*level_2*
datum	*level_1*	*level_2*	*level_2*
datum	*level_2*	*level_1*	*level_2*
datum	*level_2*	*level_1*	*level_2*
datum	*level_2*	*level_2*	*level_2*
datum	*level_2*	*level_2*	*level_2*
…			

Syntax:

```
UNIANOVA
  depvar BY a b c
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = a b c a*b a*c b*c a*b*c .
```

Using the menus, choose **Analyze → General Linear Model → Univariate.** Enter A, B, C as between-subjects factors.

7.4 One within-subjects factor

Alternative names

- Repeated-measures ANOVA (with one factor)
- Randomized complete block (RCB) design (with one factor)
- Single-factor within-subjects design

Examples

Twenty students have their digit span tested on dry land (U1) and then those *same* students have a further digit span test when they are diving in a dry suit in the Pacific Ocean (U2). Does their location affect performance?

A researcher wishes to test the effectiveness of four fertilizers (U1, U2, U3, U4). He divides his orchard into four *blocks* (equivalent to 'subjects') to account for variations across the orchard (e.g. southern sunny block, northern cool block, eastern morning sun block, western evening sun block). He divides each block into four plots and assigns fertilizers U1–U4 to each plot at random, so that each block has all four fertilizers in it.

Notes

Described in detail by Howell (1997, chapter 14). Total variation is first partitioned into variation *between* subjects and variation *within* subjects. Variation within subjects is then subdivided into variation between *treatments* (levels of our factor) and *error.*

We're not particularly interested in variation between subjects, but accounting for it allows us to isolate the effect of our factor more accurately.

If our factor has only two levels, this is equivalent to a two-sample paired *t* test.

Model description (S = subjects)

depvar = (U × S)

Model

Either

$$Y_{ij} = \mu + \pi_i + \alpha_j + \varepsilon_{ij} \quad \textbf{(additive model)}$$

where

- Y_{ij} is the dependent variable for subject *i* in condition U_j
- μ is the overall mean
- π_i is the contribution from a particular person or subject (subject *i*, or S_i): $\pi_i = \mu_{S_i} - \mu$
- α_j is the contribution from a particular level (level *j*) of the factor U: $\alpha_j = \mu_{U_j} - \mu$
- ε_{ij} is everything else (the experimental error associated with subject *i* in condition *j*): $\varepsilon_{ij} = X_{ij} - (\mu + \pi_i + \alpha_j)$.

or, perhaps better,

$$Y_{ij} = \mu + \pi_i + \alpha_j + \pi\alpha_{ij} + \varepsilon_{ij} \quad \textbf{(nonadditive model)}$$

where

- $\pi\alpha_{ij}$ is the contribution from the interaction of subject *i* with treatment *j*:
- in this case, ε_{ij} would be redefined as $\varepsilon_{ij} = Y_{ij} - (\mu + \pi_i + \alpha_j + \pi\alpha_{ij})$.

However, if we measure each person in each condition once, we will not be *able* to measure differences in the way subjects respond to different conditions ($\pi\alpha_{ij}$) *independently of other sources of error* (ε_{ij}). (To do that, we'd need to measure subjects more than once, and then we'd need a different model again!) This is another way of saying that the S × U interaction is confounded with — is! — the 'error' term. Therefore, the calculations do not differ for the two models (Myers & Well, 1995, p. 242); the only difference is if you want to estimate ω^2, the proportion of variance accounted for by a particular term (Myers & Well, 1995, pp. 252-255).

Sources of variance

Analysis of variance discards constant terms (like μ) and examines the *sources of variability (variance).* Writing this in terms of sums of squares (SS),

$$SS_{total} = SS_{subjects} + SS_U + SS_{error}$$

where SS_{total} is the total variability, SS_U is the variability attributable to the (within-subjects) factor U, and SS_{error} is the 'error' variability (everything that's left over).

This equation can be used to represent both models described above (with or without the subject × factor interaction), since, to repeat, the subject × factor interaction *is* the error term in this design (with only one score per cell) and cannot be separated from 'error'; see Howell (1997, p. 452-4).

ANOVA table

In all cases, the mean square (MS) is the sum of squares (SS) for a particular row divided by the degrees of freedom (d.f.) for the same row. Assuming one observation per cell, we have

Source	d.f.	SS	*F*
Between subjects (S)	n–1	$SS_{subjects}$	$MS_{subjects} / MS_{error}$
U	u–1	SS_U	MS_U / MS_{error}
Error (S × U)	(n–1)(u–1)	SS_{error}	
Total	N–1 = un – 1	SS_{total}	

where u is the number of levels of factor U, N is the total number of observations (= un), and n is the number of subjects.

SPSS technique 1

One row, one subject:

Ulevel1	*Ulevel2*	*Ulevel3*	…
datum	*datum*	*datum*	…
datum	*datum*	*datum*	…
datum	*datum*	*datum*	…
…			

Syntax:

```
GLM
  u1 u2 u3
  /WSFACTOR = u 3 Polynomial
  /METHOD = SSTYPE(3)
  /CRITERIA = ALPHA(.05)
  /WSDESIGN = u .
```

SPSS won't report the 'between-subjects' effects (the one based on $SS_{subjects}$, which we're not particularly interested in). It'll report something else (I'm not sure what…) as 'Between-Subjects Effects: Intercept', and the within-subjects effect that we *are* interested in as 'Within-Subjets: U'.

It will also report Mauchly's test of sphericity of the covariance matrix, together with Greenhouse–Geisser and Huynh–Feldt corrections for use if the assumption of sphericity is violated.

Using the menus, choose **Analyze → General Linear Model → Repeated Measures.** Define the within-subjects factor (with its number of levels). Then you can assign individual variables (e.g. *Ulevel1*) to appropriate levels of the factor. For a worked example, see p. 122.

SPSS technique 2

One column, one variable:

depvar	*subject*	*U*
datum	*subj_1*	*level_1*
datum	*subj_1*	*level_2*
datum	*subj_1*	*level_3*
datum	*subj_2*	*level_1*
datum	*subj_2*	*level_2*
datum	*subj_2*	*level_3*
datum	*subj_3*	*level_1*
datum	*subj_3*	*level_2*
datum	*subj_3*	*level_3*
…		

Syntax:

```
GLM depvar BY subject u
  /RANDOM = subject
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = subject u .
```

SPSS will report the within-subjects effect as 'Between-Subjects: U' (since it doesn't know that anything's a within-subjects effect!). It'll report the $SS_{subjects}$ term (the difference between subjects) as 'Between-Subjects: SUBJECT'. It'll report the same 'Intercept' term as before.

Mauchly's test is not reported; neither are the G–G and H–F corrections. To obtain these, use technique 1 instead.

You could also use this:

```
GLM depvar BY subject u
  /RANDOM = subject
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = subject u subject*u .
```

… but as we've said, the Subject × U interaction is confounded with error in this design, and SPSS simply won't give you a result for it. All other answers will be the same.

Using the menus, choose **Analyze → General Linear Model → Univariate.** Enter U as a **fixed** factor; enter Subject as a **random** factor.

7.5 Two within-subjects factors

Alternative names

- Repeated-measures ANOVA (with two factors)
- Randomized complete block (RCB) design (with two factors)
- Two-factor within-subjects design
- Split-block design

Example

Twenty students have their digit span tested on dry land when sober (U1 V1) and then those *same* students have a further digit span test when they're on dry land and sober (U1 V2), when they are diving in a dry suit in the Pacific Ocean and sober (U2 V1) and when they're drunk and diving (U2 V2). Don't try this at home, kids. Does their location or sobriety affect performance? Do these two factors interact?

A researcher wishes to test the effectiveness of three fertilizers (U1, U2, U3) and three tree thinning techniques (V1, V2, V3). He divides his national park forest into four *blocks* (equivalent to 'subjects') to account for variations across the park (e.g. mountainous conifers, lowland deciduous, timber-harvested forest, volcanic ash area). He divides each block into nine plots and assigns fertilizers U1–U3 and thinning techniques V1–V3 to each plot at random but such that every block contains every combination of fertilizer and thinning treatment once.

Notes

Model description

depvar = (U × V × S)

Model

There are two alternative models (see Howell, 1997, p.486-7, which describes the problem for three within-subjects factors; this is merely a simpler case). The first, simpler model, is this, in which the Subject term doesn't interact with anything:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \pi_k + \varepsilon_{ijk}$$

where

- Y_{ijk} is the dependent variable for subject k in condition U_i, V_j
- μ is the overall mean
- α_i is the contribution from a particular level (level i) of factor U: $\alpha_i = \mu_{U_i} - \mu$
- β_j is the contribution from a particular level (level j) of factor V: $\beta_j = \mu_{V_j} - \mu$
- π_k is the contribution from a particular person or subject (subject k): $\pi_k = \mu_{S_k} - \mu$
- ε_{ijk} is everything else (the experimental error associated with subject k in condition U_iV_j): $\varepsilon_{ijk} = Y_{ijk} - (\mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \pi_k)$

The second, probably better model, is this, which allows the Subject term to interact with the other variables (i.e. accounts for the fact that different treatments may affect different subjects in different ways):

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \pi_k + \alpha\pi_{ik} + \beta\pi_{jk} + \alpha\beta\pi_{ijk} + \varepsilon_{ijk}$$

where

- $\alpha\pi_{ik}$ is the contribution from the interaction of subject k with treatment U_i: $\alpha\pi_{ik} = \mu_{S_kU_i} - (\mu + \alpha_i + \pi_k)$
- $\beta\pi_{jk}$ is the contribution from the interaction of subject k with treatment V_j: $\beta\pi_{jk} = \mu_{S_kV_j} - (\mu + \beta_j + \pi_k)$
- $\alpha\beta\pi_{ijk}$ is the contribution from the interaction of subject k with the treatment combination U_iV_j: $\alpha\beta\pi_{ijk} = \mu_{S_kU_iV_j} - (\mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \pi_k + \alpha\pi_{ij} + \beta\pi_{jk})$
- in this case, we would redefine the error term: $\varepsilon_{ijk} = Y_{ijk} - (\mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \pi_k + \alpha\pi_{ik} + \beta\pi_{jk} + \alpha\beta\pi_{ijk})$

However, this more complex model does have a problem: since we have included the Subject term as a variable that interacts with everything, we now only have one score per cell, and we have no residual left for estimating error (ε_{ijkl}). However, as it happens (Howell, 1997, pp. 487-

8), we can use the sum of squares for the U × S term ($\alpha\pi_{il}$) as an error estimate for the U term (α_i), the sum of squares for V × S as an error estimate for the V term, and so on. **The full model is usually preferable (Howell, 1997, p. 487).**

Sources of variance

Either the reduced model

$$SS_{total} = SS_{subjects} + SS_U + SS_V + SS_{U\times V} + SS_{error}$$

or the full model

$$SS_{total} = SS_{subjects} + SS_U + SS_V + SS_{U\times V} + SS_{U\times S} + SS_{V\times S} + SS_{U\times V\times S}$$

ANOVA table

In all cases, the mean square (MS) is the sum of squares (SS) for a particular row divided by the degrees of freedom (d.f.) for the same row. Assuming one observation per cell, we have either

Source	d.f.	SS	*F*
Between subjects	n–1	$SS_{subjects}$	
U	u–1	SS_U	MS_U / MS_{error}
V	v–1	SS_V	MS_V / MS_{error}
U × V	(u–1)(v–1)	$SS_{U\times V}$	$MS_{U\times V} / MS_{error}$
Error	(n–1)(uv–1)	SS_{error}	
Total	N–1 = uvn – 1	SS_{total}	

or, with the full model:

Source	d.f.	SS	*F*
Between subjects (S)	n–1	SS_S	
U	u–1	SS_U	$MS_U / MS_{U\times S}$
error U × S	(u–1)(n–1)	$SS_{U\times S}$	
V	v–1	SS_V	$MS_V / MS_{V\times S}$
error V × S	(v–1)(n–1)	$SS_{V\times S}$	
U × V	(u–1)(v–1)	$SS_{U\times V}$	$MS_{U\times V} / MS_{U\times V\times S}$
error U × V × S	(u–1)(v–1)(n–1)	$SS_{U\times V\times S}$	
Total	N–1 = uvn – 1	SS_{total}	

where u is the number of levels of factor U, etc., N is the total number of observations (= uvn), and n is the number of subjects.

SPSS technique 1

One row, one subject:

U1V1	*U2V1*	*U1V2*	*U2V2*
datum	*datum*	*datum*	*datum*
datum	*datum*	*datum*	*datum*
datum	*datum*	*datum*	*datum*
…			

Syntax:

```
GLM
  u1v1 u1v2 u2v1 u2v2
  /WSFACTOR = u 2 Polynomial v 2 Polynomial
  /METHOD = SSTYPE(3)
  /CRITERIA = ALPHA(.05)
  /WSDESIGN = u v u*v .
```

This will give you the 'full model' answer (see above), in which the 'Subject' factor is allowed to interact with everything in full.

Using the menus, choose **Analyze → General Linear Model → Repeated Measures.** Define the within-subjects factors (with their numbers of levels). Then you can assign individual variables (e.g. *U1V1*) to appropriate levels of the factors. For a worked example, see p. 122.

SPSS technique 2

One column, one variable:

Subject	*U*	*V*	*depvar*
1	*1*	*1*	*datum*
1	*2*	*1*	*datum*
1	*1*	*2*	*datum*
1	*2*	*2*	*datum*
2	*1*	*1*	*datum*
2	*2*	*1*	*datum*
2	*1*	*2*	*datum*
2	*2*	*2*	*datum*
…			

To get the 'reduced' model (see above):

```
GLM depvar BY subject u v
  /RANDOM = subject
  /METHOD = SSTYPE(3)
  /CRITERIA = ALPHA(.05)
  /DESIGN = u v u*v subject .
```

To get the 'full' model, matching SPSS's usual within-subjects technique (see above):

```
GLM depvar BY subject u v
  /RANDOM = subject
  /METHOD = SSTYPE(3)
  /CRITERIA = ALPHA(.05)
  /DESIGN = u v u*v subject u*subject v*subject u*v*subject .
```

As usual with this technique, Mauchly's test is not reported; neither are the G–G and H–F corrections. To obtain these, use technique 1 instead.

Using the menus, choose **Analyze → General Linear Model → Univariate.** Enter U, V as **fixed** factors; enter Subject as a **random** factor.

7.6 Three within-subjects factors

Alternative names

- Repeated-measures ANOVA (with three factors)
- Randomized complete block (RCB) design (with three factors)
- Three-factor within-subjects design

Example

Oh, it gets boring making these up. A set of subjects are all tested in every combination of three treatments ($U_1 \ldots U_u$, $V_1 \ldots V_v$, $W_1 \ldots W_w$).

Notes

In the agricultural version, this is what an RCB design might look like:

Block 1			Block 2			Block 3		
U1 V1 W3	U2 V1 W1	U1 V1 W1	U1 V2 W3	U2 V1 W2	U1 V2 W1	U1 V2 W2	U1 V1 W3	U1 V1 W1
U1 V1 W2	U2 V2 W3	U2 V2 W2	U2 V1 W3	U2 V2 W3	U1 V1 W1	U2 V2 W2	U2 V1 W2	U2 V2 W3
U2 V2 W1	U1 V2 W2	U1 V2 W3	U2 V2 W2	U1 V1 W3	U2 V2 W1	U2 V2 W1	U1 V1 W2	U2 V1 W1
U2 V1 W2	U1 V2 W1	U2 V1 W3	U2 V1 W1	U1 V1 W2	U1 V2 W2	U2 V1 W3	U1 V2 W1	U1 V2 W3

Randomized complete block design with three blocks.
Factors are U (2 levels), V (2 levels), W (3 levels).
Every block is treated with all 12 combinations of W, V, and U in full factorial fashion.
The treatments are randomized within the 12 divisions of each block.

In our terminology, the agricultural 'block' is the psychological 'subject': each subject experiences each combination of the factors U, V, and W.

Model description

depvar = (U × V × W × S)

Model

There are two alternative models (see Howell, 1997, p.486-8). The first, simpler model, is this, in which the Subject term doesn't interact with anything:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \pi_l + \varepsilon_{ijkl}$$

where

- Y_{ijkl} is the dependent variable for subject l in condition U_i, V_j, W_k
- μ is the overall mean
- α_i is the contribution from a particular level (level i) of factor U
- β_j is the contribution from a particular level (level j) of factor V
- γ_k is the contribution from a particular level (level k) of factor W
- $\alpha\beta_{ij}$, $\alpha\gamma_{ik}$, $\beta\gamma_{jk}$, and $\alpha\beta\gamma_{ijk}$ are the contributions from the UV, UW, VW, and UVW interaction terms
- π_l is the contribution from a particular person or subject (subject l)
- ε_{ijkl} is everything else (the experimental error associated with subject l in condition $U_iV_jW_k$).

The second, probably better model, is this, which allows the Subject term to interact with the other variables (i.e. accounts for the fact that different treatments may affect different subjects in different ways):

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \pi_l + \alpha\pi_{il} + \beta\pi_{jk} + \gamma\pi_{kl} + \alpha\beta\pi_{ijl} + \alpha\gamma\pi_{ikl} + \beta\gamma\pi_{jkl} + \alpha\beta\gamma\pi_{ijkl} + \varepsilon_{ijkl}$$

where

- $\alpha\pi_{il}$ is the contribution from the interaction of subject *l* with treatment U_i
- $\beta\pi_{jk}$ is the contribution from the interaction of subject *l* with treatment V_j
- $\gamma\pi_{kl}$ is the contribution from the interaction of subject *l* with treatment W_k
- $\alpha\beta\pi_{ijl}$ is the contribution from the interaction of subject *l* with the treatment combination U_iV_j
- $\alpha\gamma\pi_{ikl}$ is the contribution from the interaction of subject *l* with the treatment combination U_iW_k
- $\beta\gamma\pi_{jkl}$ is the contribution from the interaction of subject *l* with the treatment combination V_jW_k
- $\alpha\beta\gamma\pi_{ijkl}$ is the contribution from the interaction of subject *l* with the treatment combination $U_iV_jW_k$

For exact specification of each of these components (e.g. $\alpha_i = \mu_{U_i} - \mu$) see the previous model (p. 115→); it's just the same but with more terms.

However, this more complex model does have a problem: since we have included the Subject term as a variable that interacts with everything, we now only have one score per cell, and we have no residual left for estimating error (ε_{ijkl}). However, as it happens (Howell, 1997, pp. 487-8), we can use the sum of squares for the U × S term ($\alpha\pi_{il}$) as an error estimate for the U term (α_i), the sum of squares for V × S as an error estimate for the V term, and so on. **The full model is usually preferable (Howell, 1997, p. 487).**

Sources of variance

Either the reduced model

$$SS_{total} = SS_{subjects} + SS_U + SS_V + SS_W + SS_{U\times V} + SS_{U\times W} + SS_{V\times W} + SS_{U\times V\times W} + SS_{error}$$

or the full model

$$SS_{total} = SS_{subjects} + SS_U + SS_V + SS_W + SS_{U\times V} + SS_{U\times W} + SS_{V\times W} + SS_{U\times V\times W} + SS_{U\times S} + SS_{V\times S} + SS_{W\times S} + SS_{U\times V\times S} + SS_{U\times W\times S} + SS_{V\times W\times S} + SS_{U\times V\times W\times S}$$

ANOVA table

In all cases, the mean square (MS) is the sum of squares (SS) for a particular row divided by the degrees of freedom (d.f.) for the same row. Assuming one observation per cell, we have either

Source	d.f.	SS	*F*
Between subjects	$n-1$	$SS_{subjects}$	
U	$u-1$	SS_U	MS_U / MS_{error}
V	$v-1$	SS_V	MS_V / MS_{error}
W	$w-1$	SS_W	MS_W / MS_{error}
U × V	$(u-1)(v-1)$	$SS_{U\times V}$	$MS_{U\times V} / MS_{error}$
U × W	$(u-1)(w-1)$	$SS_{U\times W}$	$MS_{U\times W} / MS_{error}$
V × W	$(v-1)(w-1)$	$SS_{V\times W}$	$MS_{V\times W} / MS_{error}$
U × V × W	$(u-1)(v-1)(w-1)$	$SS_{U\times V\times W}$	$MS_{U\times V\times W} / MS_{error}$
Error	$(n-1)(uvw-1)$	SS_{error}	
Total	$N-1 = uvwn - 1$	SS_{total}	

or in the 'full' version:

Source	d.f.	SS	*F*
Between subjects	$n-1$		
U	$u-1$	SS_U	$MS_U / MS_{U\times S}$
error U × S	$(u-1)(n-1)$	$SS_{U\times S}$	
V	$v-1$	SS_V	$MS_V / MS_{V\times S}$
error V × S	$(v-1)(n-1)$	$SS_{V\times S}$	
W	$w-1$	SS_W	$MS_W / MS_{W\times S}$
error W × S	$(w-1)(n-1)$	$SS_{W\times S}$	
U × V	$(u-1)(v-1)$	$SS_{U\times V}$	$MS_{U\times V} / MS_{U\times V\times S}$
error U × V × S	$(u-1)(v-1)(n-1)$	$SS_{U\times V\times S}$	
U × W	$(u-1)(w-1)$	$SS_{U\times W}$	$MS_{U\times W} / MS_{U\times W\times S}$
error U × W × S	$(u-1)(w-1)(n-1)$	$SS_{U\times W\times S}$	

V × W	$(v-1)(w-1)$	$SS_{V\times W}$	$MS_{V\times W} / MS_{V\times W\times S}$
error V × W × S	$(v-1)(w-1)(n-1)$	$SS_{V\times W\times S}$	
U × V × W	$(u-1)(v-1)(w-1)$	$SS_{U\times V\times W}$	$MS_{U\times V\times W} / MS_{U\times V\times W\times S}$
error U × V × W × S	$(u-1)(v-1)(w-1)(n-1)$	$SS_{U\times V\times W\times S}$	
Total	$N-1 = uvwn - 1$	SS_{total}	

where *u* is the number of levels of factor U, etc., *N* is the total number of observations (= *uvwn*), and *n* is the number of subjects.

SPSS technique 1 One row, one subject:

U1V1W1	*U2V1W1*	*U1V2W1*	*U2V2W1*	*U1V1W2*	*U2V1W2*	*U1V2W2*	*U2V2W2* *(etc.)*
datum	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*
datum	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*
datum	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*	*datum*
…							

Syntax:

```
GLM
  u1v1w1 u1v1w2 u1v2w1 u1v2w2 u2v1w1 u2v1w2 u2v2w1 u2v2w2
  /WSFACTOR = u 2 Polynomial v 2 Polynomial w 2 Polynomial
  /METHOD = SSTYPE(3)
  /CRITERIA = ALPHA(.05)
  /WSDESIGN = u v w u*v u*w v*w u*v*w .
```

This will give you the 'full model' answer (see above), in which the 'Subject' factor is allowed to interact with everything in full. This layout doesn't allow you to use the 'reduced' model, as far as I can see.

Using the menus, choose **Analyze → General Linear Model → Repeated Measures.** Define the within-subjects factors (with their numbers of levels). Then you can assign individual variables (e.g. *U1V1W1*) to appropriate levels of the factors. For a worked example, see p. 122.

SPSS technique 2 One column, one variable:

Subject	*U*	*V*	*W*	*depvar*
1	*1*	*1*	*1*	*datum*
1	*2*	*1*	*1*	*datum*
1	*1*	*2*	*1*	*datum*
1	*2*	*2*	*1*	*datum*
2	*1*	*1*	*1*	*datum*
2	*2*	*1*	*1*	*datum*
2	*1*	*2*	*1*	*datum*
2	*2*	*2*	*1*	*datum*
1	*1*	*1*	*2*	*datum*
1	*2*	*1*	*2*	*datum*
1	*1*	*2*	*2*	*datum*
1	*2*	*2*	*2*	*datum*
2	*1*	*1*	*2*	*datum*
2	*2*	*1*	*2*	*datum*
2	*1*	*2*	*2*	*datum*
2	*2*	*2*	*2*	*datum*
…				

To get the 'reduced' model (see above):

```
GLM depvar BY subject u v w
  /RANDOM = subject
  /METHOD = SSTYPE(3)
  /CRITERIA = ALPHA(.05)
  /DESIGN = u v w u*v u*w v*w u*v*w subject .
```

To get the 'full' model, matching SPSS's usual within-subjects technique (see above):

```
GLM depvar BY subject u v w
```

```
/RANDOM = subject
/METHOD = SSTYPE(3)
/CRITERIA = ALPHA(.05)
/DESIGN = [subject]
            u u*subject
             v v*subject
             w w*subject
             u*v u*v*subject
             u*w u*w*subject
             v*w v*w*subject
             u*v*w subject u*v*w*subject .
```

As usual with this technique, Mauchly's test is not reported; neither are the G–G and H–F corrections.

Using the menus, choose **Analyze → General Linear Model → Univariate.** Enter A, B, C as **fixed** factors; enter Subject as a **random** factor.

7.7 One between- and one within-subjects factor

Alternative names

- Split-plot design (Keppel, 1991)
- Mixed two-factor within-subjects design (Keppel, 1991)
- Repeated measures analysis using a split-plot design (SPSS, 2001, p. 464)
- Univariate mixed models approach with subject as a random effect (SPSS, 2001, p. 464)

Example

We take three groups of rats, n = 8 per group (s = 24). We give one group treatment A1, one group treatment A2, and one group treatment A3. (One subject only experiences one treatment.) Then we measure every subject's performance at six time points U1…U6.

Notes

We first partition the total variation into *between-subjects* variability and *within-subjects* variability.

The between-subjects variability can be attributed to either the effect of the treatment group (A), or differences between subjects in the same group ('S within A' or 'S / A'). (This notation indicates that there is a different group of subjects at each level of the between-subjects factor, A; we could not measure simply of 'subject variation independent of the effects of A' since no subjects ever serve in more than one group, or level of A. SPSS uses the alternative notation of S(A).) So we have these sources of between-subjects variability:

A
S / A

The within-subjects variability can be attributed to either the effects of the time point (U), or an interaction between the time point and the drug group (U × A), or an interaction between the time point and the subject-to-subject variability, which again we can only measure *within* a drug group (U × S / A). So we have these sources of within-subject variability:

U
U × A
U × S / A

Model description

depvar = A × (U × S)

Model

Following Myers & Well (1995, p. 295-6):

$$Y_{ijk} = \mu + \alpha_i + \pi_{j/i} + \beta_k + \alpha\beta_{ik} + \pi\beta_{jk/i} + \varepsilon_{ijk}$$

where

- Y_{ijk} is the dependent variable for subject j in group A_i and condition U_k
- μ is the overall mean
- α_i is the contribution from a particular level (level i) of factor A: $\alpha_i = \mu_{A_i} - \mu$
- $\pi_{j/i}$ is the contribution from a particular person or subject (subject j), who only serves *within* condition A_i ('subject within group', or S / A): $\pi_{j/i} = \mu_{S_j/A_i} - \mu$
- (There is no straightforward interaction of A with S: every subject is only measured at one level of A, so this term would be indistinguishable from the subject-only effect $\pi_{j/i}$.)
- β_k is the contribution from a particular level (level k) of factor U: $\beta_k = \mu_{U_k} - \mu$
- $\alpha\beta_{ik}$ is the contribution from the interaction of A_i and U_k: $\alpha\beta_{ik} = \mu_{A_iU_k} - (\mu + \alpha_i + \beta_k)$
- $\pi\beta_{jk/i}$ is the contribution from the interaction of U_k with subject j, which can only be measured within one level of A (it's the 'SU / A' term): $\pi\beta_{jk/i} = \mu_{S_jU_k/A_i} - (\mu + \pi_{j/i} + \beta_k)$
- (There is no straightforward three-way A × U × S interaction: every subject is only measured at one level of A, so this term would be indistinguishable from the SU / A effect $\pi\beta_{jk/i}$.)
- ε_{ijk} is everything else (the experimental error associated with measuring person j — who always experiences treatment A_i — in condition U_k):
  $\varepsilon_{ijk} = Y_{ijk} - (\mu + \alpha_i + \pi_{j/i} + \beta_k + \alpha\beta_{ik} + \pi\beta_{jk/i})$.

Note that we cannot actually measure ε_{ijk} independent of the SU / A term if we only have one measurement per subject per level of U.

Sources of variance

$$SS_{total} = SS_{between\ subjects} + SS_{within\ subjects}$$
$$SS_{between\ subjects} = SS_A + SS_{S/A}$$
$$SS_{within\ subjects} = SS_U + SS_{U\times A} + SS_{U\times S/A}$$

So

$$SS_{total} = SS_A + SS_{S/A} + SS_U + SS_{U\times A} + SS_{U\times S/A}$$

We have two different 'error' terms, one for the between-subjects factor and one for the within-subjects factor (and its interaction with the between-subjects factor), so we can't just label them 'SS_{error}'. But we could rewrite the total like this if we wanted:

$$SS_{total} = SS_A + SS_{error\text{-}between} + SS_U + SS_{U\times A} + SS_{error\text{-}within}$$

ANOVA table

Source	d.f.	SS	*F*
Between subjects (S):	$s-1 = an-1$		
A	$a-1$	SS_A	$MS_A / MS_{S/A}$
error S / A	$(an-1)-(a-1) = a(n-1)$	$SS_{S/A}$	
Within subjects:	$(N-1)-(s-1) = an(u-1)$		
U	$u-1$	SS_U	$MS_U / MS_{U\times S/A}$
U × A	$(u-1)(a-1)$	$SS_{U\times A}$	$MS_{U\times A} / MS_{U\times S/A}$
error U × S / A	$a(u-1)(n-1)$	$SS_{U\times S/A}$	
Total	$N-1 = aun-1$	SS_{total}	

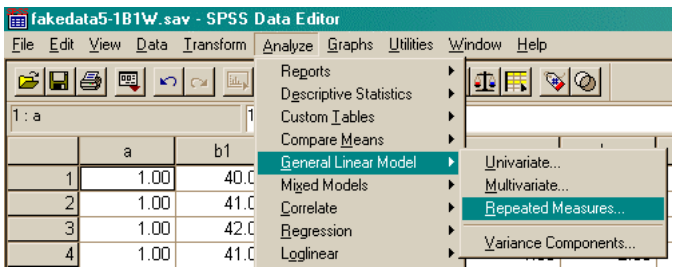
where *a* is the number of levels of factor A, etc., *N* is the total number of observations (= *aun*), *n* is the number of subjects per group (per level of A), and *s* is the total number of subjects (= *an*).

SPSS technique 1

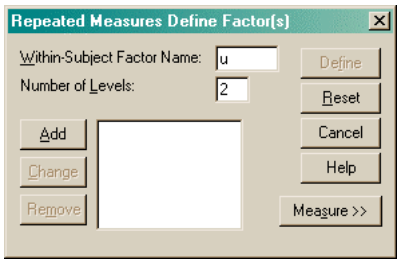
One subject, one row:

A	*U1*	*U2*
1	*datum*	*datum*
1	*datum*	*datum*
1	*datum*	*datum*
…		
2	*datum*	*datum*
2	*datum*	*datum*
2	*datum*	*datum*
…		

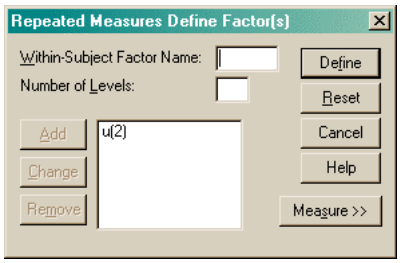
Using the menus, choose **Analyze → General Linear Model → Repeated Measures.** Define the within-subjects factor (with its number of levels). Then you can assign individual variables (e.g. *U1*) to appropriate levels of the factors, and assign the between-subjects factor.



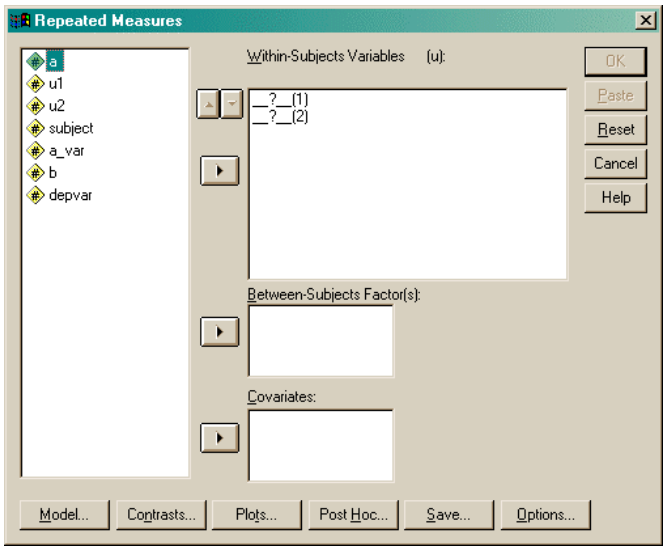
Here's where we fill in the list of within-subjects factors and the number of levels. Type them in and click 'Add'.



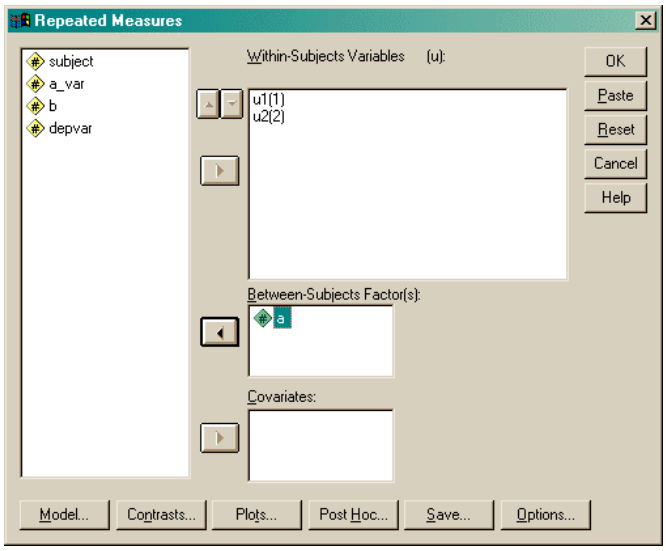
They appear in the list.



If we had more within-subjects factors, we could add them too. Once we've finished, we click 'Define'.



We can now fill in the variables (U1, U2) corresponding to the levels of factor U; we can also define A as a between-subjects factor.



Once everything else is OK, click 'OK' to run the analysis, or 'Paste' to copy the syntax for the analysis to a syntax window. This analysis produces the following syntax:

```
GLM
  u1 u2 BY a
  /WSFACTOR = u 2 Polynomial
  /METHOD = SSTYPE(3)
  /CRITERIA = ALPHA(.05)
  /WSDESIGN = u
  /DESIGN = a .
```

SPSS technique 2

One column, one variable:

A	*Subject*	*U*	*depvar*
1	*1*	*1*	*datum*
1	*1*	*2*	*datum*
1	*2*	*1*	*datum*
1	*2*	*2*	*datum*
1	*3*	*1*	*datum*
1	*3*	*1*	*datum*
…			
2	*14*	*1*	*datum*
2	*14*	*2*	*datum*
2	*15*	*1*	*datum*
2	*15*	*2*	*datum*
…			

Syntax:

```
GLM depvar BY A subject U
  /RANDOM = subject
  /DESIGN = A subject*A
              U U*A U*subject*A.
```

or alternatively

```
GLM depvar BY A subject U
  /RANDOM = subject
  /DESIGN = A subject(A)
            U U*A U*subject(A).
```

(This syntax is an example on page 464 of the SPSS 11.0 Syntax Reference Guide PDF.) It tests MS_A against $MS_{subject \times A}$, and it tests the others (MS_U and $MS_{U \times A}$) against what it simply calls MS_{error}. As usual with this technique, Mauchly's test is not reported; neither are the G–G and H–F corrections. The underlined bit is optional, since this is the same as the residual error and won't be fully calculated, but including it won't change the answers for any other factor.

Not entirely trivial to accomplish with the SPSS menus. Using the menus, choose *Analyze* → *General Linear Model* → *Univariate.* Enter A, U as *fixed* factors; enter Subject as a *random* factor. Since SPSS will get the model wrong for 'mixed' models (by including S and U × S terms), you then need to edit the *Model* directly before running the analysis. Untick *'Full factorial'* by ticking *'Custom'*. Enter the desired terms (in this case the between subjects term A, the error term S / A which you enter as S × A, the within-subject bits U, U × A, and if you want, the error term U × S / A which you enter as U × S × A, though that's optional).

7.8 Two between-subjects factors and one within-subjects factor

Alternative names

Example

Fat men, thin men, fat women, and thin women (A_1B_1, A_2B_1, A_1B_2, and A_2B_2) all have their blood pressure measured in the morning (U_1) and in the evening (U_2). Does blood pressure depend on any of these factors, or on a combination of them? Obesity and sex are between-subjects variables; time of day is a within-subject variable.

Notes

We first partition the total variation into *between-subjects* variability and *within-subjects* variability.

The between-subjects variability can be attributed to either the effect of the between-subjects factors (A, B, A × B), or differences between subjects in the same group ('S within group', or in Keppel's notation, since a group is specified by a unique combination of A and B, 'S / AB'). So we have these sources of between-subjects variability:

A
B
A × B
S / AB (between-subjects error)

The within-subjects variability can be attributed to either the effects of the within-subjects factor (U), or some form of interaction between U and the between-subjects factors (U × A, U × B, U × A × B), or an interaction between U and the subject-to-subject variability, which again we can only measure *within* a 'group' (U × S / AB). So we have these sources of within-subject variability:

U
U × A
U × B
U × A × B
U × S / AB (within-subjects error)

Model description

depvar = A × B × (U × S)

Model

I made this up, but I got it right for a change (Myers & Well, 1995, p. 308):

$$\begin{aligned} Y_{ijkl} &= \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \pi_{k/ij} \\ &\quad + \gamma_l + \alpha\gamma_{il} + \beta\gamma_{jl} + \alpha\beta\gamma_{ijl} + \pi\gamma_{kl/ij} + \varepsilon_{ijkl} \end{aligned}$$

where

- Y_{ijkl} is the dependent variable for subject k in condition A_i, B_j, U_k
- μ is the overall mean
- α_i is the contribution from a particular level (level i) of factor A: $\alpha_i = \mu_{A_i} - \mu$
- β_j is the contribution from a particular level (level j) of factor B: $\beta_j = \mu_{B_j} - \mu$
- $\alpha\beta_{ij}$ is the contribution from the interaction of A_i and B_j: $\alpha\beta_{ij} = \mu_{A_iB_j} - (\mu + \alpha_i + \beta_j)$
- $\pi_{k/ij}$ is the contribution from a particular person or subject (subject k), who is measured only in condition A_iB_j (this is the S / AB term): $\pi_{k/ij} = \mu_{S_j/A_iB_j} - \mu$
- γ_l is the contribution of level l of factor U: $\gamma_l = \mu_{U_l} - \mu$
- $\alpha\gamma_{il}$, $\beta\gamma_{jl}$, and $\alpha\beta\gamma_{ijl}$ represent the A_i / U_l, B_j / U_l, and A_i / B_j / U_l interaction contributions, respectively: $\alpha\gamma_{il} = \mu_{A_iU_l} - (\mu + \alpha_i + \gamma_l)$; $\beta\gamma_{jl} = \mu_{B_jU_l} - (\mu + \beta_j + \gamma_l)$; and $\alpha\beta\gamma_{ijl} = \mu_{A_iB_jU_l} - (\mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \gamma_l + \alpha\gamma_{il} + \beta\gamma_{jl})$.
- $\pi\gamma_{kl/ij}$ represents the interaction of U_l with subject k (who only experiences condition A_iB_j) — the U×S / AB term: $\pi\gamma_{kl/ij} = \mu_{S_kU_l/A_iB_j} - (\mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \pi_{k/ij} + \gamma_l + \alpha\gamma_{il} + \beta\gamma_{jl} + \alpha\beta\gamma_{ijl})$
- ε_{ijk} is everything else (the experimental error associated with measuring person k, who al-

ways experiences treatment A_i, in condition U_j):
$\varepsilon_{ijkl} = Y_{ijkl} - (\mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \pi_{k/ij} + \gamma_l + \alpha\gamma_{il} + \beta\gamma_{jl} + \alpha\beta\gamma_{ijl} + \pi\gamma_{kl/ij})$. Of course, this cannot be measured independently of the U×S / AB term (since there is only one observation in condition $A_iB_jS_kU_l$).

Sources of variance

$$SS_{total} = SS_{between\text{-}subjects} + SS_{within\text{-}subjects}$$
$$SS_{between\text{-}subjects} = SS_A + SS_B + SS_{A\times B} + SS_{error\text{-}between}$$
$$SS_{within\text{-}subjects} = SS_U + SS_{U\times A} + SS_{U\times B} + SS_{U\times A\times B} + SS_{error\text{-}within}$$

ANOVA table

Source	d.f.	SS	*F*
Between subjects (S):	*abn*–1		
A	*a*–1	SS_A	$MS_A / MS_{S/AB}$
B	*b*–1	SS_B	$MS_B / MS_{S/AB}$
A × B	(*a*–1)(*b*–1)	$SS_{A\times B}$	$MS_{A\times B} / MS_{S/AB}$
error S / AB	*ab*(*n*–1)	$SS_{S/AB}$	
Within subjects:	*abn*(*u*–1)		
U	*u*–1	SS_U	$MS_U / MS_{U\times S/AB}$
U × A	(*u*–1)(*a*–1)	$SS_{U\times A}$	$MS_{U\times A} / MS_{U\times S/AB}$
U × B	(*u*–1)(*b*–1)	$SS_{U\times B}$	$MS_{U\times B} / MS_{U\times S/AB}$
U × A × B	(*u*–1)(*a*–1)(*b*–1)	$SS_{U\times A\times B}$	$MS_{U\times A\times B} / MS_{U\times S/AB}$
error U × S / AB	*ab*(*u*–1)(*n*–1)	$SS_{U\times S/AB}$	
Total	*N*–1 = *abun* – 1	SS_{total}	

where *a* is the number of levels of factor A, etc., *N* is the total number of observations (= *abun*), and *n* is the number of subjects per group (where a group is defined by the combination of factors A and B).

SPSS technique 1 One subject, one row:

A	*B*	*U1*	*U2*	*U3…*
1	1	*datum*	*datum*	*datum…*
1	1	*datum*	*datum*	*datum…*
1	2	*datum*	*datum*	*datum…*
1	2	*datum*	*datum*	*datum…*
2	1	*datum*	*datum*	*datum…*
2	1	*datum*	*datum*	*datum…*
…				

Syntax:

```
GLM
  u1 u2 u3 BY a b
  /WSFACTOR = u 3 Polynomial
  /METHOD = SSTYPE(3)
  /CRITERIA = ALPHA(.05)
  /WSDESIGN = u
  /DESIGN = a b a*b .
```

Using the menus, choose **Analyze → General Linear Model → Repeated Measures.** Define the within-subjects factor (with its number of levels). Then you can assign individual variables (e.g. *U1*) to appropriate levels of the factors, and assign the between-subjects factors.

SPSS technique 2 One column, one variable:

A	*B*	*Subject*	*U*	*depvar*
1	*1*	*1*	*1*	*datum*
1	*1*	*1*	*2*	*datum*
1	*1*	*1*	*3*	*datum*
1	*1*	*2*	*1*	*datum*
1	*1*	*2*	*2*	*datum*
1	*1*	*2*	*3*	*datum*

1	*2*	*3*	*1*	*datum*
1	*2*	*3*	*2*	*datum*
1	*2*	*3*	*3*	*datum*
…				

Syntax:

```
GLM depvar BY a b subject u
  /RANDOM = subject
  /DESIGN = a b a*b subject*a*b
              u u*a u*b u*a*b u*subject*a*b.
```

An alternative syntax is this:

```
GLM depvar BY a b subject u
  /RANDOM = subject
  /DESIGN = a b a*b subject(a*b)
            u u*a u*b u*a*b u*subject(a*b).
```

As usual with this technique, Mauchly's test is not reported; neither are the G–G and H–F corrections. The underlined bit is optional, since this is the same as the residual error and won't be fully calculated, but including it won't change the answers for any other factor.

Not entirely trivial to accomplish with the SPSS menus. Using the menus, choose *Analyze → General Linear Model → Univariate.* Enter A, B, U as *fixed* factors; enter Subject as a *random* factor. Since SPSS will get the model wrong for 'mixed' models (by including S and all sorts of terms in which the between-subjects factors interact with S), you then need to edit the *Model* directly before running the analysis. Untick *'Full factorial'* by ticking *'Custom'*. Enter the desired terms (in this case the between-subjects bits A, B, A × B, the error term S / AB which you enter as S × A × B, the within-subjects bits U, U × A, U × B, U × A × B, and optionally the error term U × S / AB which you enter as U × S × A × B).

7.9 One between-subjects factor and two within-subjects factors

Alternative names

Example

Rats are given a brain lesion (A_1) or a sham operation (A_2). They are repeatedly offered two levers; one delivers small, immediate reward, and the other delivers large, delayed reward. Their preference for the large, delayed reward is assessed (dependent variable) at different delays (U_1, U_2, … U_5). Furthermore, they are tested hungry (V_1) or sated (V_2). All subjects experience all combinations of U and V, suitably counterbalanced, but one subject is only ever in one A group.

Notes

We first partition the total variation into *between-subjects* variability and *within-subjects* variability.

The between-subjects variability can be attributed to either the effect the between-subjects factor (A), or differences between subjects in the same group ('S within group', or 'S / A'). So we have these sources of between-subjects variability:

A
S / A (between-subjects error)

The within-subjects variability can be attributed to either the effects of the within-subjects factors (B, C, B × C), or some form of interaction between the within-subjects factors and the between-subjects factor (B × A, C × A, B × C × A), or an interaction between the within-subjects factors and the subject-to-subject variability (B × S / A, C × S / A, B × C × S / A) where 'S / A' again refers to subject variability *within* a 'group' (defined by the between-subjects factor, A). So we have these sources of within-subject variability:

U
U × A
U × S / A (within-subjects error term for the preceding two factors)
V
V × A
V × S / A (within-subjects error term for the preceding two factors)
U × V
U × V × A
U × V × S / A (within-subjects error term for the preceding two factors)

Model description

depvar = A × (U × V × S)

Model

This would be rather tedious to write out (see Myers & Well, 1995, p. 312); follow the principles in the previous model, which was for A × (U × S). The models always start with the overall mean (μ). Then the between-subject factors (here, α), and their interactions (here, none), are added. Then there's subject term (π), which is **nested** within levels of A. Then there are the within-subject factors (β, γ), and their interactions ($\beta\gamma$). Then for the full model all within-subject factors and interactions interact with the subject term, which itself is nested within A (to give $\beta\pi$, $\gamma\pi$, $\beta\gamma\pi$). Finally there's the ε term.

Sources of variance

$$SS_{total} = SS_{between\text{-}subjects} + SS_{within\text{-}subjects}$$
$$SS_{between\text{-}subjects} = SS_A + SS_{error\text{-}between}$$
$$SS_{within\text{-}subjects} = SS_U + SS_{U\times A} + SS_{U\times S/A} + SS_V + SS_{V\times A} + SS_{V\times S/A} + SS_{U\times V} + SS_{U\times V\times A} + SS_{U\times V\times S/A}$$

ANOVA table

Source	d.f.	SS	*F*
Between subjects:	$an-1$		
A	$a-1$	SS_A	$MS_A / MS_{S/A}$
error S / A	$a(n-1)$	$SS_{S/A}$	
Within subjects:	$an(uv-1)$		
U	$u-1$	SS_U	$MS_U / MS_{U\times S/A}$
U × A	$(u-1)(a-1)$	$SS_{A\times U}$	$MS_{A\times U} / MS_{U\times S/A}$
error U × S / A	$a(u-1)(n-1)$	$SS_{U\times S/A}$	

V	v–1	SS_V	$MS_V / MS_{V\times S/A}$
V × A	$(v–1)(a–1)$	$SS_{V\times A}$	$MS_{V\times A} / MS_{V\times S/A}$
error V × S / A	$a(v–1)(n–1)$	$SS_{V\times S/A}$	
U × V	$(u–1)(v–1)$	$SS_{U\times V}$	$MS_{U\times V} / MS_{U\times V\times S/A}$
U × V × A	$(v–1)(a–1)(u–1)$	$SS_{U\times V\times A}$	$MS_{U\times V\times A} / MS_{U\times V\times S/A}$
error U × V × S / A	$a(u–1)(v–1)(n–1)$	$SS_{U\times V\times S/A}$	
Total	$N–1 = auvn – 1$	SS_{total}	

where a is the number of levels of factor A, etc., N is the total number of observations (= $auvn$), and n is the number of subjects per group (where group is defined by factor A).

SPSS technique 1 One row, one subject:

A	*U1V1*	*U1V2*	*U2V1*	*U2V2*
1	*datum*	*datum*	*datum*	*datum*
1	*datum*	*datum*	*datum*	*datum*
1	*datum*	*datum*	*datum*	*datum*
2	*datum*	*datum*	*datum*	*datum*
2	*datum*	*datum*	*datum*	*datum*
2	*datum*	*datum*	*datum*	*datum*
…				

Syntax:

```
GLM
  u1v1 u1v2 u2v1 u2v2 BY a
  /WSFACTOR = u 2 Polynomial v 2 Polynomial
  /METHOD = SSTYPE(3)
  /PRINT = DESCRIPTIVE HOMOGENEITY
  /CRITERIA = ALPHA(.05)
  /WSDESIGN = u v u*v
  /DESIGN = a .
```

Using the menus, choose **Analyze → General Linear Model → Repeated Measures.** Define the within-subjects factors (with their numbers of levels). Then you can assign individual variables (e.g. *U1V1*) to appropriate levels of the factors, and assign the between-subjects factor.

SPSS technique 2 One column, one variable:

A	*Subject*	*U*	*V*	*depvar*
1	*1*	*1*	*1*	*datum*
1	*1*	*1*	*2*	*datum*
1	*1*	*2*	*1*	*datum*
1	*1*	*2*	*2*	*datum*
1	*2*	*1*	*1*	*datum*
1	*2*	*1*	*2*	*datum*
1	*2*	*2*	*1*	*datum*
1	*2*	*2*	*2*	*datum*
…				
2	*3*	*1*	*1*	*datum*
2	*3*	*1*	*2*	*datum*
2	*3*	*2*	*1*	*datum*
2	*3*	*2*	*2*	*datum*

Syntax:

```
UNIANOVA
  depvar BY a subject u v
  /RANDOM = subject
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /PRINT = DESCRIPTIVE HOMOGENEITY
  /CRITERIA = ALPHA(.05)
  /DESIGN = a subject*a
              u u*a u*subject*a
```

```
              v v*a v*subject*a
              u*v u*v*a u*v*subject*a .
```

Incidentally, the notation `Subject(A)` will be accepted as equivalent to `Subject*A` in these sorts of designs; feel free to use this alternative form if it seems clearer:

```
UNIANOVA
  depvar BY a subject u v
  /RANDOM = subject
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /PRINT = DESCRIPTIVE HOMOGENEITY
  /CRITERIA = ALPHA(.05)
  /DESIGN = a subject(a)
              u u*a u*subject(a)
              v v*a v*subject(a)
           u*v u*v*a u*v*subject(a) .
```

Not entirely trivial to accomplish with the SPSS menus. Using the menus, choose *Analyze → General Linear Model → Univariate.* Enter A, B, U as *fixed* factors; enter Subject as a *random* factor. Since SPSS will get the model wrong for 'mixed' models (by including S and all sorts of terms in which the between-subjects factors interact with S), you then need to edit the *Model* directly before running the analysis. Untick *'Full factorial'* by ticking *'Custom'*. Enter the desired terms, as listed above; the method is explained further in the section on the 'two-between, one-within' model.

7.10 Other ANOVA designs with between and/or within-subjects factors

The models above can be extended along the same principles. See Keppel (1991), pp. 491–496. A full map of all the 'error' terms is given on p. 493; an expanded version showing all terms is presented here. For any term, the appropriate error term is the next error term in the list. The different error terms needed for partial and full within-subjects models are discussed by Howell (1997, pp. 487-488). Only full models are presented for designs involving between-subject factors.

			Between-subjects factors			
			None	**1 factor (A)**	**2 factors (A, B)**	**3 factors (A, B, C)**
Within-subjects factors	**None**	**Design:**	**–**	**A × S**	**A × B × S**	**A × B × C × S**
		Terms:	–	A error [= S / A]	A B A × B error [= S / AB]	A B C A × B A × C B × C A × B × C error [= S / ABC]
	1 factor (U)	**Design:**	**(U × S)**	**A × (U × S)**	**A × B × (U × S)**	**A × B × C × (U × S)**
		Terms:	between subjects term [S] U error [= U × S]	*between subjects:* A error S / A *within subjects:* U U × A error U × S / A	*between subjects:* A B A × B error S / AB *within subjects:* U U × A U × A × B error U × S / AB	*between subjects:* A B C A × B A × C B × C A × B × C error S / ABC *within subjects:* U U × A U × B U × C U × A × B U × A × C U × B × C U × A × B × C error U × S / ABC
	2 factors (U, V)	**Design:**	**(U × V × S)**	**A × (U × V × S)**	**A × B × (U × V × S)**	**A × B × C × (U × V × S)**
		Terms:	***simpler model:*** between-subjects term [S] U V U × V error ***full model (preferable):*** between-subjects term [S] (no corresponding error term) U error U × S V error V × S U × V error U × V × S	*between subjects:* A error S / A *within subjects:* U U × A error U × S / A V V × A error V × S / A U × V U × V × A error U × V × S / A	*between subjects:* A B A × B error S / AB *within subjects:* U U × A U × B U × A × B error U × S / AB V V × A V × B V × A × B error V × S / AB U × V U × V × A U × V × B U × V × A × B error U × V × S / AB	*between subjects:* A B A × B A × B × C error S / ABC *within subjects:* U U × A U × B U × A × B error U × S / AB V V × A V × B V × A × B error V × S / AB U × V U × V × A U × V × B U × V × A × B error U × V × S / AB

	3 factors (U, V, W)	**Design:**	**(U × V × W × S)**	**A × (U × V × W × S)**	**A × B × (U × V × W × S)**	**A × B × C × (U × V × W × S)**
		Terms:	***simpler model:*** between-subjects term [S] U V U × V U × W V × W U × V × W error ***full model (preferable):*** between-subjects term [S] (no corresponding error term) U error U × S V error V × S W error W × S U × V error U × V × S U × W error U × W × S V × W error V × W × S U × V × W error U × V × W × S	*between subjects:* A error S / A *within subjects:* U U × A error U × S / A V V × A error V × S / A W W × A error W × S / A U × V U × V × A error U × V × S / A U × W U × W × A error U × W × S / A V × W V × W × A error V × W × S / A U × V × W U × V × W × A error U × V × W × S / A	*between subjects:* A B A × B error S / AB *within subjects:* U U × A U × B U × A × B error U × S / AB V V × A V × B V × A × B error V × S / AB W W × A W × B W × A × B error W × S / AB U × V U × V × A U × V × B U × V × A × B error U × V × S / AB U × W U × W × A U × W × B U × W × A × B error U × W × S / AB V × W V × W × A V × W × B V × W × A × B error V × W × S / AB U × V × W U × V × W × A U × V × W × B U × V × W × A × B error U × V × W × S / AB	*between subjects:* A B C A × B A × C A × B × C error S / ABC *within subjects:* U U × A U × B U × C U × A × B U × A × C U × B × C U × A × B × C error U × S / ABC V V × A V × B V × C V × A × B V × A × C V × B × C V × A × B × C error V × S / ABC W W × A W × B W × C W × A × B W × A × C W × B × C W × A × B × C error W × S / ABC U × V U × V × A U × V × B U × V × C U × V × A × B U × V × A × C U × V × B × C U × V × A × B × C error U × V × S / ABC U × W U × W × A U × W × B U × W × C U × W × A × B U × W × A × C U × W × B × C U × W × A × B × C error U × W × S / ABC V × W V × W × A V × W × B V × W × C V × W × A × B V × W × A × C V × W × B × C V × W × A × B × C error V × W × S / ABC U × V × W U × V × W × A U × V × W × B U × V × W × C U × V × W × A × B U × V × W × A × C U × V × W × B × C U × V × W × A × B × C error U × V × W × S / ABC

7.11 One between-subjects covariate (linear regression)

Alternative names

- Analysis of covariance (ANCOVA) — though traditionally this term isn't applied to a design with no other factors
- Linear regression

Example

You measure subjects' income (dependent variable) and want to predict it in the basis of their IQ. Every subject contributes an single (IQ, income) pair of values. This is basic linear regression. In regression terminology we would be trying to predict the dependent variable *Y* from the another, predictor variable *X* — i.e. solving the regression equation

$$\hat{Y} = bX + a$$

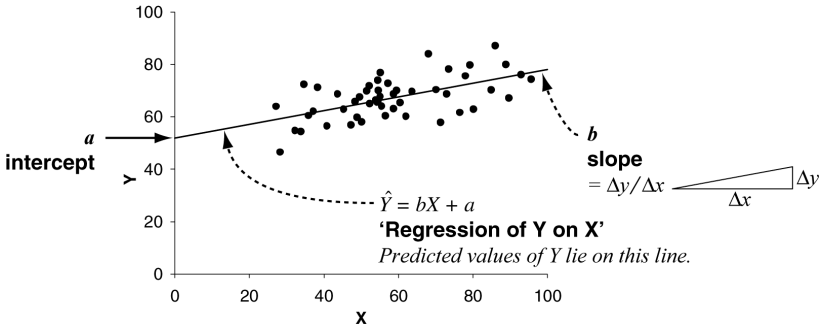
$$\text{where } b = \frac{\text{cov}_{XY}}{s_X^2} = r\frac{s_Y}{s_X} = r\frac{\sqrt{SS_Y}}{\sqrt{SS_X}}$$

$$\text{and } a = \bar{y} - b\bar{x}$$

where $\hat{Y}$ is the predicted value of *Y* (see also Myers & Well, 1995, p. 387). Alternatively, we could write this:

$$Y = bX + a + \varepsilon$$

where ε symbolizes the error or residual. The equation represents, of course, this:



Or we could lay out the equation so as to be extensible to multiple regression (which we'll look at later):

$$\hat{Y} = b_0 + b_1 X$$

$$Y = b_0 + b_1 X + \varepsilon$$

In ANCOVA terminology, the predictor variable is the covariate, which we'll call *C*. So we could first rewrite the simple linear regression equation with the letters we'll use from now on:

$$\hat{Y} = a + bC \text{ where } a = \overline{Y} - b\overline{C}$$

and now write it as a prediction for specific values of *Y* and *C*, namely Y_i and C_i:

$$Y_i = a + bC_i + \varepsilon \text{ where } a = \overline{Y} - b\overline{C}$$

and now write it terms of the means of *Y* ($= \overline{Y} = \mu$) and *C* ($\overline{C}$):

$$\begin{aligned} Y_i &= \mu - b\overline{C} + bC_i + \varepsilon \\ &= \mu + b(C_i - \overline{C}) + \varepsilon \end{aligned}$$

(Compare Myers & Well, 1995, p. 436.) We'll use this below. It helps to distinguish between the **predicted value of *Y* based on the covariate** [which is $\hat{Y}_i = a + bC_i = \mu + b(C_i - \overline{C})$] and the **contribution of the covariate,** which is the deviation of the covariate-predicted value of *Y*

from the overall mean of *Y* [which is therefore $c_i = b(C_i - \overline{C})$]. Obviously, $c_i = \hat{Y}_i - \mu$.

Note also that the *proportion* of the total variability in *Y* that's accounted for by predicting it from *C* is equal to r^2:

$$r^2 = \frac{SS_{\hat{Y}}}{SS_Y} = \frac{SS_{model}}{SS_{total}}$$

and the SS attributable to the model (SS_{model} or $SS_{regression}$ or SS_{reg}) can be written

$$\begin{aligned} SS_{reg} &= \sum(\hat{Y}_i - \overline{Y})^2 \\ &= r^2 SS_Y \\ &= b^2 SS_C \end{aligned}$$

Notes

Model description depvar = C_{cov} + S

(I've made that up, as Keppel doesn't have a specific notation for models including covariates.)

Model $Y_i = \mu + c_i + \varepsilon_i$

where

- Y_i is the dependent variable for subject *i*
- μ is the overall mean
- c_i is the contribution of the covariate for subject *i*: $c_i = b(C_i - \overline{C}) = \hat{Y}_i - \mu$ where *b* is the regression coefficient, C_i is the value of the covariate for subject *i*, $\overline{C}$ is the overall mean value of the covariate, and $\hat{Y}_i$ is the value of Y_i predicted by on the basis of the covariate.
- ε_i is everything else (the error, residual, 'individual variation', etc.): $\varepsilon_i = Y_i - (\mu + c_i)$

Sources of variance $SS_{total} = SS_{reg} + SS_{error}$

The SS_{reg} is given by $SS_{reg} = \sum(c_i)^2 = \sum b(C_i - \overline{C}) = \sum(\hat{Y}_i - \mu) = r^2 SS_Y = b^2 SS_C$ (Myers & Well, 1995, p. 393). It's the sum of the squared contributions of the covariate, which is to say the sum of the squared deviations between the covariate-predicted value and overall mean.

ANOVA table Covariates have 1 degree of freedom.

Source	d.f.	SS	*F*
C_{cov} (regression)	1	SS_C	MS_C / MS_{error}
Error	*N*–2	SS_{error}	
Total	*N*–1	SS_{total}	

where *N* is the number of subjects.

SPSS technique Data layout:

C	*depvar*
datum	*datum*
datum	*datum*
datum	*datum*
…	

Either run the analysis as a regression:

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
```

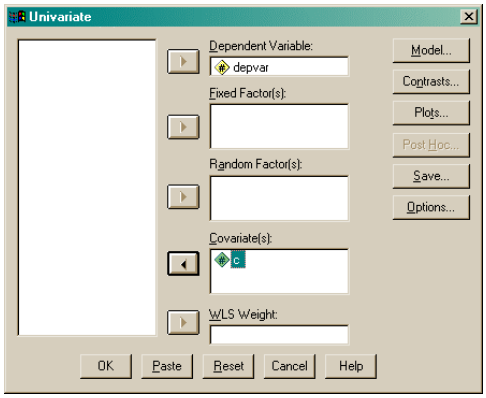
```
  /DEPENDENT depvar
  /METHOD=ENTER c .
```

… or as an ANCOVA (note use of **WITH** for covariates, rather than **BY** for factors):

```
UNIANOVA
  depvar  WITH c
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /PRINT = PARAMETER
  /DESIGN = c .
```

This will also give you r^2 for the model. The `/PRINT = PARAMETER` syntax also gives you *b*; you can combine $\sqrt{r^2}$ with the sign of *b* to calculate *r*.

Using the menus, choose **Analyze → General Linear Model → Univariate;** enter the dependent variable and the covariate in the appropriate boxes.



To get parameter (*b*) estimates as well, choose *Options → Parameter estimates*.

7.12 One between-subjects covariate and one between-subjects factor

7.12.1. The covariate and factor do not interact

Alternative names

- Analysis of covariance (ANCOVA)
- Analysis of covariance (ANCOVA) assuming homogeneity of regression
- Traditional ANCOVA

Example

After Howell (1997, p. 585). Suppose we are interested in whether small cars are easier to handle. We can compare driving proficiency using three cars: small, medium, and large (A_1, A_2, A_3). One driver is tested in only one car. We have three groups of drivers to test, but they vary considerably in their driving experience (C_{cov}). We have arranged matters so the mean driving experience is the same in each group. If driving experience has a very large effect on performance, we may be unable to detect an effect of car type. So we can 'partial out' the effect of driving experience (C_{cov}), increasing our power to detect an effect of car type (A).

More controversially, suppose that the mean level of driving experience was *not* the same for the three groups. Then performing an analysis of covariance is like asking what the effect of car type was *had the groups not differed* on the covariate. This may not make sense; see Howell (1997, pp. 596-7) and Myers & Well (1995, pp. 449-454). For example, if you measure the effect of a drug on three-year-old and five-year-old children and covary for body weight, it may make little sense to ask what the effect on three-year-olds would be if they weighed the same as five-year-olds — they don't. **Statistically** controlling for the covariate is not the same as **experimentally** controlling for the covariate (Myers & Well, 1995, p. 452).

Even worse is the situation when you measure the covariate *after* the treatment (factor) has been applied and the treatment has affected the covariate; it's then pretty difficult to interpret an analysis of covariance meaningfully. See Howell (1997, pp. 596-7).

Notes

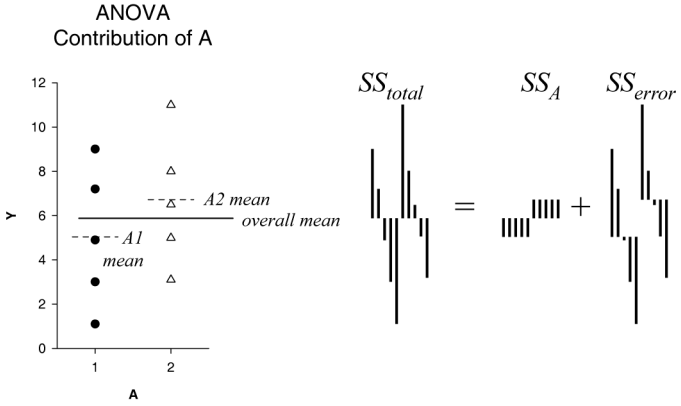
Howell tends to refer to covariates as things that are accounted for or partialled out *in advance* of consideration of other factors (Howell, 1997, p. 587; p. 606). This implies that the covariate × factor interaction is *not* included, except to check the assumption of homogeneity of regression. This is a traditional meaning of ANCOVA; see the GLM section (p. 88→) for a full explanation. SPSS refers to covariates in the sense of 'continuous predictor variables' (as opposed to factors, which are discrete predictor variables) but does follow Howell's approach when you use covariates in its 'full model' mode. I will refer to covariates in the sense of continuous predictor variables and try to make it explicit when covariates interact with factors or do not.

This model assumes that the covariate is *independent* of the experimental treatments. (If not, see the 'interacting' version below.)

Let's take these data:

A	A_1	A_1	A_1	A_1	A_1	A_2	A_2	A_2	A_2	A_2
depvar (Y)	1.1	3	4.9	7.2	9	3.1	5	6.5	8	11

We might run a one-way ANOVA on it, using our standard partitioning of variance:



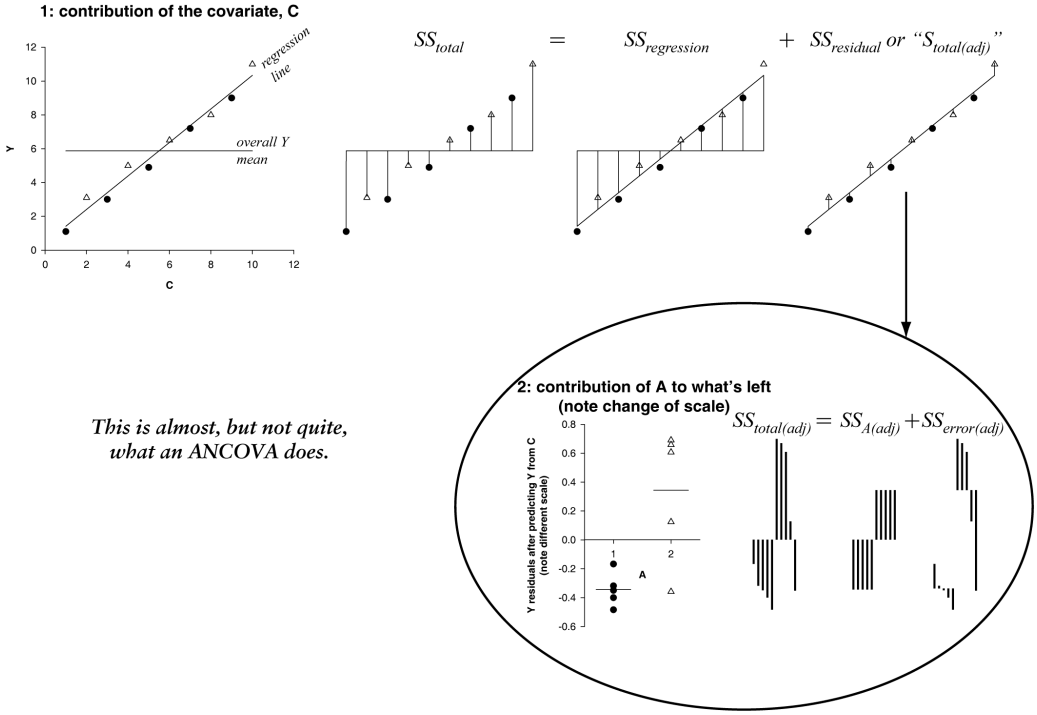
But suppose we also have information about a covariate C:

A	A_1	A_1	A_1	A_1	A_1	A_2	A_2	A_2	A_2	A_2
C	1	3	5	7	9	2	4	6	8	10
depvar (Y)	1.1	3	4.9	7.2	9	3.1	5	6.5	8	11

We might be able to get a much more powerful test of the effects of A if we removed the effect of C. We could, for example, correlate Y with C for all 10 data points, obtain predicted values of Y based on C, obtain the **residuals** and see what effect A has on those. We could therefore split the SS like this:

$$SS_{total} = SS_{regression(overall)} + SS_{residual}$$
$$\text{where } SS_{residual} = SS_A + SS_{error}$$

That'd look like this:



This is **almost** what one-way ANCOVA does. However, the regression line used is not quite the 'overall' regression (Myers & Well, 1995, pp. 436-439). To see why, consider these data:

A	A_1	A_1	A_1	A_1	A_1	A_2	A_2	A_2	A_2	A_2
C	1	3	5	7	9	2	4.3	6	8	10
depvar (Y)	1.1	3.5	4.9	6	9	8.1	10.5	11.5	16	16

Here, if we calculated the regression line using all the data lumped together, we wouldn't get as good a fit as if we fitted separate regression lines for each A group (one line for A_1, another for A_2). But the ANCOVA model we are using assumes **homogeneity of regression** — that is, that the A_1 and A_2 data may have different intercepts but they have the same slope. How do we estimate this slope? Apparently (Myers & Well, 1995, p. 438) the best estimate of what's called the **pooled within-group slope** is this:

$$b_{S/A} = \sum_i \frac{SS_{C/A_i}}{SS_{S/A(C)}} b_{A_i} = \frac{SS_{C/A_1}}{SS_{S/A(C)}} b_{A_1} + \frac{SS_{C/A_2}}{SS_{S/A(C)}} b_{A_2} + \ldots$$

where b_{A_i} is the slope calculated just for observations in group A_i

and SS_{C/A_i} is the variance of C for observations in group A_i

$$\text{and } SS_{S/A(C)} = \sum_i SS_{C/A_i}$$

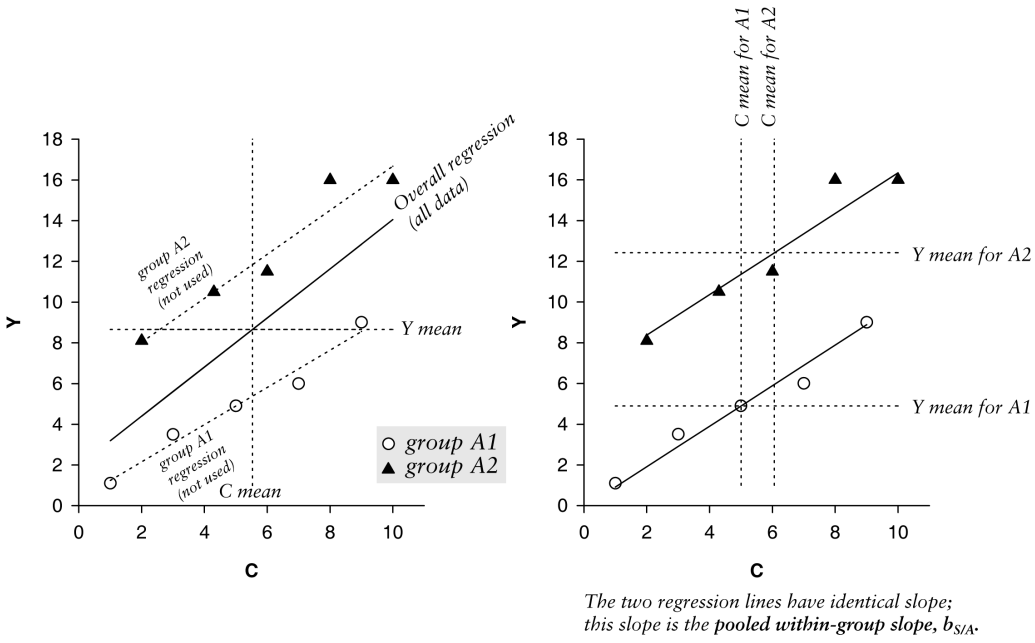
For example, with the data set above,

$$b_{A_1} = 0.915;\ b_{A_2} = 1.081$$
$$SS_{C/A_1} = 40;\ SS_{C/A_2} = 38.872$$
$$SS_{S/A(C)} = 40 + 38.872 = 78.872$$
$$b_{S/A} = \frac{40}{78.872}0.915 + \frac{40}{78.872}1.081 = 0.997$$

We can then calculate the sum of squares for linear regression within groups, $SS_{reg(S/A)}$, by summing the variabilities accounted for by the regressions with the common slope in each of the groups (Myers & Well, 1995, p. 439):

$$SS_{reg(S/A)} = b_{S/A}^2 SS_{C/A_1} + b_{S/A}^2 SS_{C/A_2} + \ldots$$
$$= b_{S/A}^2 SS_{C/A(C)}$$

… in this case, $SS_{reg(S/A)} = (0.997)^2 \times 78.872 = 78.377$. Since the within-group regression line will pass through the within-group mean points $\{\overline{C}_{A_i}, \overline{Y}_{A_i}\}$, we can sketch the situation:

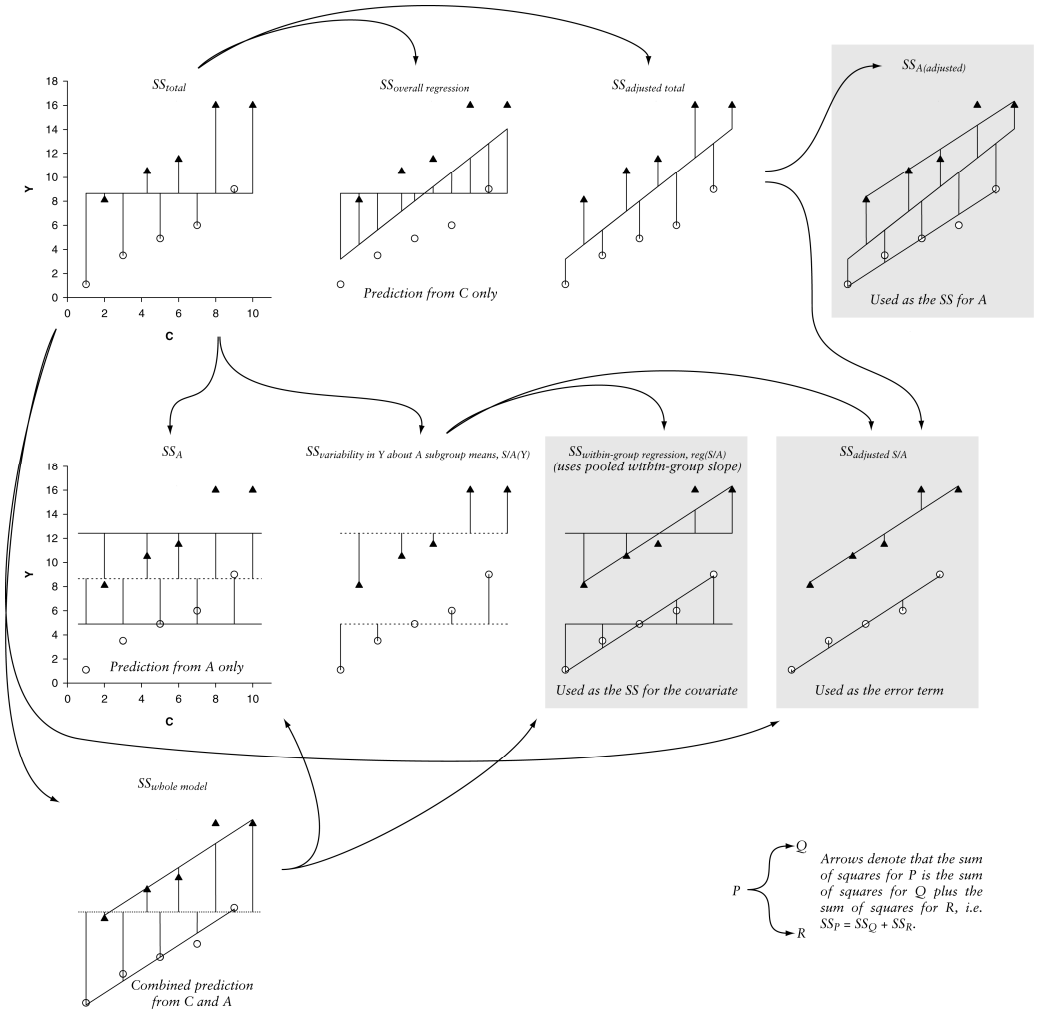


The two regression lines have identical slope; this slope is the ***pooled within-group slope, $b_{S/A}$.***

Finally, we can partition the variance like this:

$$SS_{total} = SS_{overall\ regression} + SS_{adjusted\ total}$$
$$SS_{total} = SS_A + SS_{S/A}$$
$$SS_{S/A} = SS_{within\text{-}group\ regression,\ reg(S/A)} + SS_{adjusted\ S/A}$$
$$SS_{adjusted\ total} = SS_{adjusted\ A} + SS_{adjusted\ S/A}$$

which looks like this (!):



As a result, the quoted $SS_{covariate}$ (= $SS_{reg(S/A)}$), quoted SS_A (= $SS_{A,adjusted}$), and quoted error (= $SS_{adjusted\ S/A}$) won't add up to SS_{total}.

Model description

depvar = $C_{cov} + A \times S$

(I've used the notation '+' to separate out things that don't interact with anything… this seems reasonably consistent.)

Model

Essentially, the model is

$$Y_{ij} = \mu + c_i + \alpha_j + \varepsilon_{ij}$$

where

- Y_{ij} is the dependent variable for subject *i* in condition A_j
- μ is the overall mean of *Y*
- c_i is the contribution of the covariate for subject *i*
- α_j is the contribution from a particular level (level *j*) of factor A
- ε_{ij} is everything else (the error in measuring subject *i* in condition *j*, residual, 'individual variation', etc.): $\varepsilon_i = Y_i - (\mu + c_i + \alpha_j)$

And everyone claims this is their model (Myers & Well, 1995, p. 436; Howell, 1997, pp. 588-590); see also Keppel (1991, pp. 308-317). However, what's actually going on is a bit more sophisticated — there's are *two definitions* for c_i and α_j, depending on what we want to test. What actually happens is this (best explained by Myers & Well, 1995, pp. 440-444; but also by Howell, 1997, pp. 590-1):

- We can view any ANOVA hypothesis test as a *comparison of two models.* For example, a simple one-way ANOVA is a comparison of a *full model* that incorporates the effect of a

factor A ($Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$) with a *restricted model* that doesn't — in this case, the restricted model is $Y_i = \mu + \varepsilon_i$.

- **Contrasting two models.** The correct way of contrasting a full (F) model and a restricted (R) model is to use this *F* test (Myers & Well, 1995, p. 441):

$$F_{(df_{\text{error(R)}} - df_{\text{error(F)}}),df_{\text{error(F)}}} = \frac{\left(\text{SS}_{\text{error(R)}} - \text{SS}_{\text{error(F)}}\right) \div \left(df_{\text{error(R)}} - df_{\text{error(F)}}\right)}{\text{SS}_{\text{error(F)}} \div df_{\text{error(F)}}}$$

Or, we could rewrite that, since $\text{SS}_{\text{total}} = \text{SS}_{\text{model}} + \text{SS}_{\text{error}}$ and $df_{\text{total}} = df_{\text{model}} + df_{\text{error}}$:

$$F_{(df_{\text{model(F)}} - df_{\text{model(R)}}),df_{\text{error(F)}}} = \frac{\left(\text{SS}_{\text{model(F)}} - \text{SS}_{\text{model(R)}}\right) \div \left(df_{\text{model(F)}} - df_{\text{model(R)}}\right)}{\text{SS}_{\text{error(F)}} \div df_{\text{error(F)}}}$$

For a one-way ANOVA, this formula reduces to $F = \text{MS}_{\text{A}} / \text{MS}_{\text{S/A}}$, our usual formula for testing the effects of A — see p. 86→ in the section on GLMs. An alternative formulation uses the R^2 values for each model (Howell, 1997, p. 578): if *f* and *r* are the number of predictors in the full and reduced models,

$$F_{f-r,N-f-1} = \frac{(N - f - 1)\left(R_f^2 - R_r^2\right)}{(f - r)\left(1 - R_f^2\right)}.$$

- Now we apply that principle to ANCOVA.
- **To test the effects of the factor A,** one model is calculated testing just the effect of the covariate C. That model is our usual regression ANOVA model, $Y_i = \mu + c_i + \varepsilon_i$, where μ is the overall mean and c_i is the contribution of the covariate, calculated using the *overall* regression ($c_i = b(C_i - \overline{C})$) — since in this model we have no information about which level of A a given subject is at, so we can't calculate the pooled within-groups slope yet. Then we calculate another model including the factor A. That model is $Y_{ij} = \mu + c_i + \alpha_j + \varepsilon_{ij}$, where α_j is the *extra* contribution of the factor. And knowledge of that factor allows us to improve our regression as well, because it allows us to calculate two regression lines with the same slope (the pooled within-groups slope, $b_{\text{S/A}}$) but different intercepts (Myers & Well, 1995, p. 442). So the extra contribution is $\alpha_j = \mu_{A_j} + b_{S/A}(C_{ij} - \overline{C}_{A_j}) - (\mu + c_i)$. We compare those two models.
- **To test the effects of the covariate C,** one model is calculated testing just the effect of the factor A. That model is our usual one-way ANOVA model $Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$, where μ is the overall mean and α_j is the contribution from a particular level (level *j*) of the factor ($\alpha_j = \mu_{A_j} - \mu$). Then we calculate another model including the covariate C. That model is $Y_{ij} = \mu + c_i + \alpha_j + \varepsilon_{ij}$, where c_i is the *extra* contribution of the covariate, using the pooled within-groups slope (i.e. *using* the information about which subject is at which level of factor A), i.e. $c_i = b_{S/A}(C_{ij} - \overline{C}_{A_i})$. We compare those two models.
- **The complicated picture above shows this.** The top row — partitioning SS_{total} into $\text{SS}_{\text{overall regression}}$, $\text{SS}_{\text{A(adjusted)}}$, and an error term, corresponds to testing the effects of A *over and above those of the covariate.* The middle row — partitioning SS_{total} into SS_{A}, $\text{SS}_{\text{within-group regression}}$, and an error term, corresponds to testing the effects of C *over and above those of the factor.*
- **Since the covariate and the factor may be correlated (provide mutual information), the questions 'what does A do?' and 'what does C do?' are *not independent*;** we therefore ask 'what does A do, over and above the effects of C?' and 'what does C do, over and above the effects of A?'

Sources of variance See above.

ANOVA table

Covariates account for 1 degree of freedom.

Source	d.f.	SS	*F*
C_{cov}	1	$SS_{reg(S/A)}$	$MS_{reg(S/A)} / MS_{S/A,adjusted}$
A	$a - 1$	$SS_{A,\ adjusted}$	$MS_{A,adjusted} / MS_{S/A,adjusted}$
Error	$N - a - 1$	$SS_{S/A,adjusted}$	
Total	$N - 1$	SS_{total}	

where N is the number of subjects and a the number of levels of factor A.

Note that the SS components for C, A, and error do not add up to SS_{total}. This is confusing; the method of partitioning is described above.

Correlation coefficient from ANCOVA

See discussion under the 'one within-subjects covariate' model (p. 152) for details of how to obtain correlation coefficients (r, r^2) and parameter estimates (b) from ANCOVA.

SPSS technique

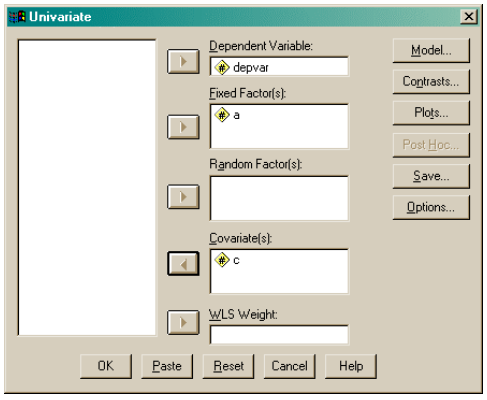
Data layout:

C	*A*	*depvar*
datum	*1*	*datum*
datum	*1*	*datum*
datum	*1*	*datum*
datum	*2*	*datum*
datum	*2*	*datum*
datum	*2*	*datum*
…		

Syntax:

```
UNIANOVA
  depvar BY a  WITH c
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = c a .
```

Using the menus, choose **Analyze → General Linear Model → Univariate.** Enter A as a fixed factor. Enter C_{cov} as a covariate.



Note that the interaction term ($C_{cov} \times A$) is not included in this model — see below for a version with the interaction.

7.12.2. The covariate and factor interact

Alternative names

- Analysis of covariance (ANCOVA) allowing covariate × factor interaction
- Analysis of covariance (ANCOVA): full model to check homogeneity of regression
- Homogeneity of slopes design ANCOVA (see p. 88→)

Example

Rats receive sham surgery (A_1) or lesions of the nucleus accumbens core (A_2). They are then trained in a task in which they may press a lever freely; each lever press produces a pellet some time later. For each rat, we measure the mean time between pressing the lever and receiving the pellet (C_{cov}; one value per subject). This is a continuous variable. We also measure their learning speed (dependent variable). Does the learning speed depend on the delay each rat experienced (main effect of C_{cov})? Does the learning speed depend on the group they were in (main effect of A)? Does the way the learning speed depends on the delay depend in turn on which group they were in (C_{cov} × A interaction)?

Note the interpretative difficulties (discussed above) that can plague any ANCOVA if you don't think things through very carefully.

Notes

Allows the covariate to interact with the factor — that is, allows for the possibility that the effects of the factor differ depending on the value of the covariate, or (equivalently) that the effects of the covariate differ depending on the level of the factor. See above for a non-interaction version.

Howell (1997, pp. 587-590) discusses the approach to a standard ANCOVA that assumes homogeneity of regression (that the regression coefficients are equal across levels of the factor, i.e. that there is no covariate × factor interaction). We discussed this 'reduced model' ANCOVA in above (p. 138). Howell (1997, pp. 587-590) uses the 'full' model, which includes the interaction term, to test the assumption of homogeneity of regression before using the 'reduced model'. However, there are times when we are interested in the interaction term for its own sake (see Example above).

Model description

depvar = $C_{cov} \times A \times S$

Model

$$Y_{ij} = \mu + c_i + \alpha_j + c\alpha_{ij} + \varepsilon_{ij}$$

where

- Y_{ij} is the dependent variable for subject *i* in condition A_j
- μ is the overall mean
- c_i is the contribution of the covariate for subject *i*
- α_j is the contribution from a particular level (level *j*) of factor A
- $c\alpha_{ij}$ is the interaction of the covariate for subject *i* with level *j* of factor A
- ε_{ij} is everything else (the error in measuring subject *i* in condition *j*, residual, 'individual variation', etc.): $\varepsilon_i = Y_i - (\mu + c_i + \alpha_j + c\alpha_{ij})$

Just as before, we can't define c_i, α_j and so on in just one way, since they may be correlated. We'll have to ask what the covariate contributes *over and above* the factor, and so on.

The test for the interaction term (Myers & Well, 1995, p. 447; Howell, 1997, p. 588-590) involves the comparison of a full model in which the regression slopes can differ for each group, or level of A (so the regression slopes are b_{A_j}):

$$Y_{ij} = \mu + \alpha_j + b_{A_j}(C_{ij} - \overline{C}_{A_j}) + \varepsilon_{ij}$$

and a restricted model in which each group has the same slope:

$$Y_{ij} = \mu + \alpha_j + b(C_{ij} - \overline{C}_{A_j}) + \varepsilon_{ij}$$

Approach 1: testing the homogeneity of regression assumption. Test the interaction term as above (i.e. perform an ANCOVA including the factor × covariate assumption). If the interaction term is not significant, the slopes don't differ. Drop the interaction term out of the model and perform your usual ANCOVA (factor, covariate, no interaction) safe in the knowledge that the

assumption of homogeneity of regression is valid. This is why most textbooks test this interaction (Myers & Well, 1995, p. 450; Howell, 1997, p. 588-590).

Approach 2: asking about the factor × covariate assumption for its own sake. Perform the full analysis with the interaction; interpret that directly. Interpretation of any main effects in the presence of an interaction may be tricky, as it is in factorial ANOVA (Myers & Well, 1995, p. 450).

Sources of variance

SS_C, SS_A, $SS_{C\times A}$, SS_{error}… but these may not be independent, so they won't necessarily add up to SS_{total} — see above.

ANOVA table

Covariates account for 1 degree of freedom.

Source	d.f.	SS	*F*
C_{cov}	1	SS_C	MS_C / MS_{error}
A	$a - 1$	SS_A	MS_A / MS_{error}
$C_{cov} \times A$	$a - 1$	$SS_{C\times A}$	$MS_{C\times A} / MS_{error}$
Error	$N - 2a$	SS_{error}	
Total	$N - 1$	SS_{total}	

where N is the number of subjects and a the number of levels of factor A.

Correlation coefficient from ANCOVA

See discussion under the 'one within-subjects covariate' model (p. 152) for details of how to obtain correlation coefficients (r, r^2) and parameter estimates (b) from ANCOVA.

SPSS technique

Data layout:

C	*A*	*depvar*
datum	*1*	*datum*
datum	*1*	*datum*
datum	*1*	*datum*
datum	*2*	*datum*
datum	*2*	*datum*
datum	*2*	*datum*
…		

Syntax:

```
UNIANOVA
  depvar BY a  WITH c
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = c a c*a .
```

Note that the interaction term ($C_{cov} \times A$) is included.

Not entirely trivial to accomplish with the SPSS menus. Using the menus, choose *Analyze → General Linear Model → Univariate.* Enter C as a covariate. Enter A as a fixed factor. By default, SPSS will not include the $C_{cov} \times A$ interaction. So you need to edit the *Model* directly before running the analysis. Untick *'Full factorial'* by ticking *'Custom'*. Enter the desired terms (in this case C, A, C × A).

7.13 One between-subjects covariate and two between-subjects factors

Alternative names

- Factorial analysis of covariance (factorial ANCOVA)

Example

Suppose we are again interested in whether small cars are easier to handle. We can compare driving proficiency using three cars: small, medium, and large (A_1, A_2, A_3). One driver is tested in only one car. We have three groups of male drivers (B_1), and three groups of female drivers (B_2), which we assign to our three cars in a standard factorial design. We also want to account for variation in driving experience (C_{cov}; one value per subject).

Notes

There's nothing to stop you including covariate × factor interactions in your model, though we won't present them here.

The general linear model will also be perfectly happy for you to include covariate × covariate interactions, if you think that's meaningful. Think carefully, though; this would be a complex design! We won't present that here.

More detailed discussion of this design is given by Myers & Well (1995, pp. 457-459).

Model description (S = subjects)

depvar = $C_{cov} + A \times B \times S$

Model

$$Y_{ijk} = \mu + c_i + \alpha_j + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijk}$$

where

- Y_{ijk} is the dependent variable for subject i in condition A_j, B_k
- μ is the overall mean
- c_i is the contribution from covariate C for subject i
- α_j is the contribution from a particular level (level j) of factor A
- β_k is the contribution from a particular level (level k) of factor B
- $\alpha\beta_{jk}$ is the contribution from the interaction of level j of factor A and level k of factor B
- ε_{ijk} is everything else (the 'uniqueness' of subject i in condition j of factor A and condition k of factor B, 'error', 'individual variation', etc.).

However, since the predictors may be correlated, there is no 'unique' way to define the contributions of each of these components (see above).

Sources of variance

As the sources of variance may not be independent, the components (SS_C, SS_A, SS_B, SS_{AB}, SS_{error}) may not add up to SS_{total}; see above.

ANOVA table

Source	d.f.	SS	*F*
C_{cov}	1	SS_C	MS_C / MS_{error}
A	a–1	SS_A	MS_A / MS_{error}
B	b–1	SS_B	MS_B / MS_{error}
A × B	(a–1)(b–1)	$SS_{A\times B}$	$MS_{A\times B} / MS_{error}$
Error	$ab(n–1)$–1	SS_{error}	
Total	N–1 = abn–1	SS_{total}	

where a is the number of levels of factor A, etc., N is the total number of observations (subjects), and n is the number of subjects (or 'replications') per cell.

Correlation coefficient from ANCOVA

See discussion under the 'one within-subjects covariate' model (p. 152) for details of how to obtain correlation coefficients (r, r^2) and parameter estimates (b) from ANCOVA.

SPSS technique

Data layout:

depvar	*A*	*B*	*C*
datum	*level_1*	*level_1*	*datum*
datum	*level_1*	*level_1*	*datum*
datum	*level_1*	*level_2*	*datum*
datum	*level_1*	*level_2*	*datum*
datum	*level_2*	*level_1*	*datum*
datum	*level_2*	*level_1*	*datum*
datum	*level_2*	*level_2*	*datum*
datum	*level_2*	*level_2*	*datum*
…			

Syntax:

```
UNIANOVA
  depvar BY a b  WITH c
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = c a b a*b .
```

Using the menus, choose **Analyze → General Linear Model → Univariate.** Enter A and B as between-subjects factors. Enter C as a covariate.

7.14 Two or more between-subjects covariates (multiple regression)

Alternative names

- Multiple regression
- Multiple linear regression

Example

Suppose we want to predict marks in undergraduate exams on the basis of A-Level points (A_{cov}) and IQ (B_{cov}).

Notes

See Howell (1997, p. 510 on) for a discussion of multiple regression, and Howell (1997, pp. 605-606) for a discussion of the use of multiple covariates.

A standard multiple regression solves the equation

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_p X_p$$

where b_0 is the intercept and b_1, b_2, … b_p represent the regression coefficients (slopes) for the predictors X_1, X_2, … X_p respectively. In general, as for linear regression, this equation is solved so as to perform least-squares regression, i.e. to minimize

$$\sum (Y - \hat{Y})^2$$

However, if the two covariates are themselves **correlated,** there will be a problem of interpretation of effects involving one or other of them (because we will have **non-orthogonal sums of squares,** as discussed earlier in the context of unequal group sizes; see p. 70→ and p. 97→).

Model description (S = subjects)

$C_{cov} + D_{cov} + \ldots \times S$
For the two-covariate case, $C_{cov} + D_{cov} + S$.

Model

To achieve standard multiple regression, in the two-predictor case, the multiple regression equation above leads us to this model in our usual ANOVA notation:

$$Y_i = \mu + c_i + d_i + \varepsilon_i$$

where

- Y_i is the dependent variable for subject *i*
- μ is the overall mean
- c_i is the contribution from covariate C for subject *i*
- d_j is the contribution from covariate D for subject *i*
- ε_i is everything else (the error in measuring subject *i*, residual, 'individual variation', etc.).

However, since the predictors may be correlated, there is no 'unique' way to define the contributions of each of these components (see above).

The $C_{cov} \times D_{cov}$ interaction is **not** included for conventional multiple linear regression.

Sources of variance

For the two-covariate case, if the covariates are independent, then $SS_{total} = SS_C + SS_D + SS_{error}$. But if the covariates are themselves correlated, the contributions of each won't necessarily add up to the total (Myers & Well, 1995, pp. 505-508).

ANOVA table

Covariates account for 1 degree of freedom each. For the two-covariate case,

Source	d.f.	SS	*F*
C_{cov}	1	SS_C	MS_C / MS_{error}
D_{cov}	1	SS_D	MS_D / MS_{error}
Error	$N - 3$	SS_{error}	
Total	$N - 1$	SS_{total}	

where *N* is the number of subjects.

Correlation coefficients and parameter estimates

See discussion under the 'one within-subjects covariate' model (p. 152) for details of how to obtain correlation coefficients (r, r^2) and parameter estimates (b) from ANCOVA. See discussion of effect size (p. 97→) to see how to interpret them.

SPSS technique

Data layout:

C	*D*	*depvar*
datum	*datum*	*datum*
datum	*datum*	*datum*
datum	*datum*	*datum*
datum	*datum*	*datum*
datum	*datum*	*datum*
datum	*datum*	*datum*
…		

Either perform the analysis as a multiple linear regression (Analyze → Regression → Linear; enter C and D as the independent variables), which gives this syntax:

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT depvar
  /METHOD=ENTER c d .
```

Or run it as an ANOVA (Analyze → General Linear Model → Univariate; enter C and D as covariates), which gives this syntax:

```
UNIANOVA
  depvar  WITH c d
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = c d .
```

Note that the interaction term ($C_{cov} \times D_{cov}$) is *not* included. You could include it if you wanted — the software won't complain — but you'd have to think very carefully about its interpretation.

7.15 Two or more between-subjects covariates and one or more between-subjects factors

Alternative names

- Factorial analysis of covariance (factorial ANCOVA) with multiple covariates

Example

Modifying Howell's (1997, pp. 605-6) example slightly, suppose we want to look at the effect of two teaching styles (A) and two classroom temperatures (B) on student performance using a factorial design. We might also want to partial out the effect of age (C_{cov}) and IQ (D_{cov}). No problem — statistically, at least.

Notes

There's nothing to stop you including covariate × factor interactions in your model, though we won't present them here.

The general linear model will also be perfectly happy for you to include covariate × covariate interactions, if you think that's meaningful. Think carefully, though; this would be a complex design! We won't present that here.

As in the previous design, if the two covariates are **correlated,** there will be a problem of interpretation (because we will have **non-orthogonal sums of squares,** as discussed earlier in the context of unequal group sizes; see p. 70→ and p. 97→).

Designs with more than one covariate are briefly discussed by Myers & Well (1995, p. 459), as is **polynomial ANCOVA** (Myers & Well, 1995, p. 460); see also p. 88→.

Model description (S = subjects)

Following our example, we'll illustrate a two-covariate, two-factor model:

$$\text{depvar} = C_{cov} + D_{cov} + A \times B \times S$$

Model

$$Y_{ijk} = \mu + c_i + d_i + \alpha_j + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijk}$$

where

- Y_{ijk} is the dependent variable for subject i in condition A_j, B_k
- μ is the overall mean
- c_i is the contribution from covariate C for subject i
- d_i is the contribution from covariate D for subject i
- α_j is the contribution from a particular level (level j) of factor A
- β_k is the contribution from a particular level (level k) of factor B
- $\alpha\beta_{jk}$ is the contribution from the interaction of level j of factor A and level k of factor B
- ε_{ijk} is everything else (the 'uniqueness' of subject i in condition j of factor A and condition k of factor B, 'error', 'individual variation', etc.).

However, since the predictors may be correlated, there is no 'unique' way to define the contributions of each of these components (see above).

Sources of variance

As the sources of variance may not be independent, the components (SS_C, SS_D, SS_A, SS_B, SS_{AB}, SS_{error}) may not add up to SS_{total}; see above.

ANOVA table

Source	d.f.	SS	*F*
C_{cov}	1	SS_C	MS_C / MS_{error}
D_{cov}	1	SS_D	MS_D / MS_{error}
A	a–1	SS_A	MS_A / MS_{error}
B	b–1	SS_B	MS_B / MS_{error}
A × B	(a–1)(b–1)	$SS_{A\times B}$	$MS_{A\times B} / MS_{error}$
Error	ab(n–1)–2	SS_{error}	
Total	N–1 = abn–1	SS_{total}	

where a is the number of levels of factor A, etc., N is the total number of observations (subjects), and n is the number of subjects (or 'replications') per cell.

Correlation coefficients and effect sizes

See discussion under the 'one within-subjects covariate' model (p. 152) for details of how to obtain correlation coefficients (r, r^2) and parameter estimates (b) from ANCOVA. See discussion of effect size above (p. 97→) to see how to interpret them.

SPSS technique

Data layout:

depvar	*A*	*B*	*C*	*D*
datum	*level_1*	*level_1*	*datum*	*datum*
datum	*level_1*	*level_1*	*datum*	*datum*
datum	*level_1*	*level_2*	*datum*	*datum*
datum	*level_1*	*level_2*	*datum*	*datum*
datum	*level_2*	*level_1*	*datum*	*datum*
datum	*level_2*	*level_1*	*datum*	*datum*
datum	*level_2*	*level_2*	*datum*	*datum*
datum	*level_2*	*level_2*	*datum*	*datum*
…				

Syntax:

```
UNIANOVA
  depvar BY a b  WITH c d
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = c d a b a*b .
```

Using the menus, choose **Analyze → General Linear Model → Univariate.** Enter A and B as between-subjects factors. Enter C and D as covariates.

7.16 One within-subjects covariate

Alternative names

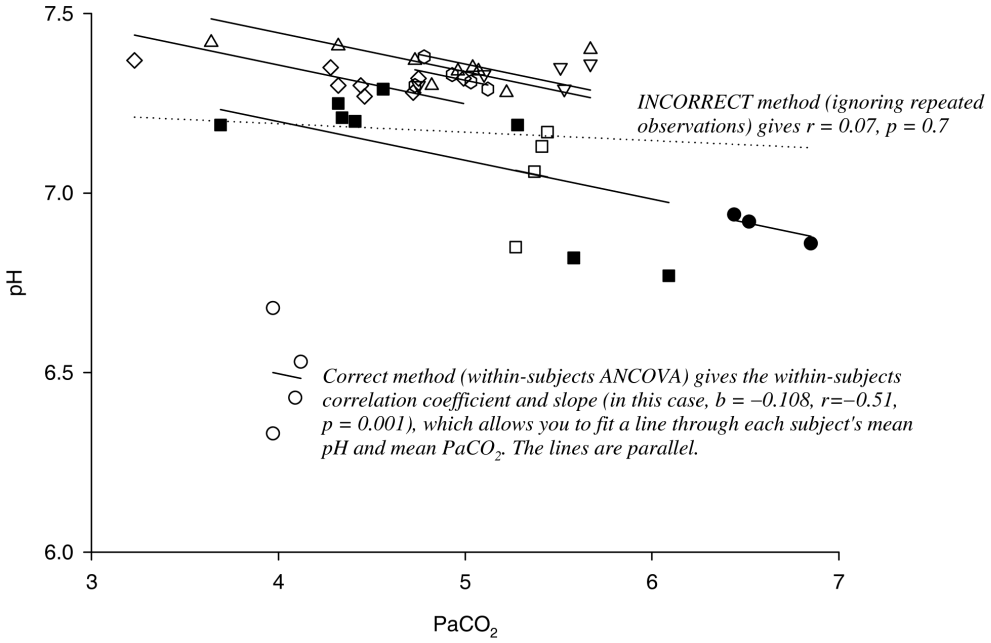
- Multiple regression with the covariate and Subject as predictors

Example

We measure gastric pH and $PaCO_2$ (partial pressure of arterial carbon dioxide) for a group of 8 subjects, making several measurements on each subject so we end up with 47 measurements (see Bland & Altman, 1995a). Is there a relationship between $PaCO_2$ and pH? We must **not** analyse this as if there were 47 independent observations. Subjects may vary widely in their gastric pH and arterial $PaCO_2$, yet there may be a consistent relationship *within* each subject between the two, and this is what we want to look at.

Notes

I've largely made up the model and sources of variance here, so I hope it's correct. It does match Bland & Altman's answer. Note that it is logically identical to the model we looked at earlier with one between-subjects covariate and one between-subjects factor (the version in which the covariate and the factor do not interact), except that our factor is now 'subjects' itself; the only difference is that subjects is a random, not a fixed, factor. Data from Bland & Altman (1995a); originally from Boyd *et al.* (1993).



Model description (S = subjects)

depvar = (C_{cov} + S)

Model

$$Y_i = \mu + c_i + \pi_i + \varepsilon_i$$

where

- Y_i is the dependent variable for subject *i*
- μ is the overall mean
- c_i is the contribution from covariate C for subject *i*
- π_i is the average contribution from a particular subject (subject *i*)
- ε_{ijk} is everything else (measurement error, intra-subject variation, etc.).

Sources of variance

$SS_{total} = SS_{subjects} + SS_C + SS_{error}$

ANOVA table

Source	d.f.	SS	*F*
Between subjects	$s - 1$	$SS_{subjects}$	$MS_{subjects} / MS_{error}$
C	1	SS_C	MS_C / MS_{error}
Error	$N - 1 - s$	SS_{error}	
Total	$N - 1$	SS_{total}	

where *N* is the total number of observations and *s* is the number of subjects.

Correlation coefficient from ANCOVA

Note also that since if we are predicting a variable Y (calling the prediction $\hat{Y}$) we can express r^2 in terms of sums of squares:

$$r^2 = \frac{SS_{\hat{Y}}}{SS_Y} = \frac{SS_{\hat{Y}}}{SS_{\hat{Y}} + SS_{residual}}$$

(see *Correlation & Regression* handout at www.pobox.com/~rudolf/psychology). If we re-write this for our present case, C is the thing that makes the prediction. The total within-subjects variation is what we're left with after we've accounted for between-subjects variation (= $SS_{total} - SS_{subjects} = SS_C + SS_{error}$) and the variation accounted for by the prediction from C is SS_C. So the proportion of the within-subjects variation accountable for by C is:

$$r^2 = \frac{SS_C}{SS_C + SS_{error}}$$

This allows us to work out the within-subjects correlation coefficient from the ANCOVA table. To obtain r itself, take the square root of r^2 and combine it with the sign of the regression coefficient. To obtain regression coefficients in SPSS, tick **Parameter estimates** in the ANOVA **Options** dialogue box, or add `/PRINT = PARAMETER` to your SPSS syntax. The regression coefficient (slope) will appear in the 'B' column and the row corresponding to the covariate.

SPSS technique

Data layout:

subject	*C*	*depvar*
1	*datum*	*datum*
1	*datum*	*datum*
1	*datum*	*datum*
2	*datum*	*datum*
2	*datum*	*datum*
3	*datum*	*datum*
3	*datum*	*datum*
3	*datum*	*datum*
3	*datum*	*datum*
3	*datum*	*datum*
4	*datum*	*datum*
…		

Syntax:

```
UNIANOVA
  depvar BY subject  WITH c
  /RANDOM = subject
  /METHOD = SSTYPE(3)
  /PRINT = PARAMETER
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = c subject .
```

Using the menus, select **Analyze → General Linear Model → Univariate.** Enter Subject as a random factor and C as a covariate.

7.17 One within-subjects covariate and one between-subjects factor

7.17.1. The covariate and factor do not interact

Alternative names

Example

We give a drug (A_1) or placebo (A_2) to two groups of subjects to see if it affects their secretion of growth hormone (dependent variable). The drug's effects are known to last for days, and we know that time of day (C) also affects growth hormone secretion — we believe there is a linear relationship between time of day measured in a certain way and growth hormone levels. Each subject only experiences either the drug or the placebo, but we measure each subject repeatedly at several different time points. We wish to 'partial out' the effects of time of day to have a better chance of finding an effect of the drug.

(Note that our experimental design must ensure that there is no systematic relationship between A and C, or interpretation will be well-nigh impossible — for example, it would be vital not to measure the drug group in the evening and the placebo group in the morning.)

Notes

Model description (S = subjects)

depvar = A × (C_{cov} + S) [but with no C_{cov} × A term in the model]

Model

I would guess either this:

$$Y_{ij} = \mu + \alpha_i + \pi_{j/i} + c_j + \varepsilon_{ij}$$

where

- Y_{ij} is the dependent variable for subject *j* in condition A_i
- μ is the overall mean
- α_i is the contribution from level *i* of factor A
- $\pi_{j/i}$ is the average contribution from a particular subject (subject *j*), who is only measured in condition A_i
- c_{jk} is the contribution from covariate C for subject *j*
- ε_{ij} is everything else (measurement error, intra-subject variation, etc.).

or this:

$$Y_{ij} = \mu + \alpha_i + \pi_{j/i} + c_j + \pi c_{j/i} + \varepsilon_{ij}$$

where

- $\pi c_{j/i}$ is the interaction of the covariate C with subject *j* (who is only measured in condition $A_{i)}$
- ε_{ij} is redefined as 'everything else' in this new model

Should we include the subject × covariate interaction, C × S / A (allowing a different regression slope for the covariate for each subject)? Maybe that depends on the situation. Obviously, to include it, we must have multiple measurements for each subject. One approach, I suppose, would be to test the full model and proceed to the simpler model if the subject × covariate interaction doesn't contribute significantly. Including it will improve the power to detect effects of C probably at the expense of power to detect effects of A (see below).

Sources of variance

The sources of variation (SS_A, $SS_{error\text{-}between}$, SS_C, perhaps $SS_{C\times S/A}$, and $SS_{error\text{-}within}$) may not be independent and may therefore not add up to SS_{total}.

ANOVA table

If the effects of A and C are uncorrelated, the ANOVA table will look like this:

Source	d.f.	SS	*F*
Between subjects:	$s-1 = an-1$		
A	$a-1$	SS_A	$MS_A / MS_{S/A}$
error (S / A)	$a(n-1)$	$SS_{S/A}$	
Within subjects:	$(N-1)-(s-1) = N-s$		
C_{cov}	1	SS_C	$MS_C / MS_{error\text{-}within}$

error-within	$N–s–1$	$SS_{error\text{-}within}$	

Alternative for within subjects (in the model with the C × S / A term):

C_{cov}	1	SS_C	$MS_C / MS_{error\text{-}within}$
$C_{cov} \times S / A$	$a(n–1)$	$SS_{C\times S/A}$	$MS_{C\times S/A} / MS_{error\text{-}within}$
error	$N–s–a(n–1)–1$	$SS_{error\text{-}within}$	
Total	$N–1$	SS_{total}	

where a is the number of levels of factor A and N is the total number of observations (= aun), n is the number of subjects per group (where 'group' is defined by factor A), and s is the total number of subjects.

SPSS technique

Data layout:

A	*Subject*	*C*	*depvar*
1	*1*	*datum*	*datum*
1	*1*	*datum*	*datum*
1	*1*	*datum*	*datum*
1	*2*	*datum*	*datum*
1	*2*	*datum*	*datum*
1	*2*	*datum*	*datum*
…			
2	*7*	*datum*	*datum*
2	*7*	*datum*	*datum*
2	*7*	*datum*	*datum*
2	*7*	*datum*	*datum*
2	*8*	*datum*	*datum*
2	*8*	*datum*	*datum*
…			

Syntax (using the notation `subject(a)` rather than the functionally equivalent `subject*a` for the term S / A):

```
GLM
  depvar BY a subject  WITH c
  /RANDOM = subject
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /PRINT = DESCRIPTIVE HOMOGENEITY
  /CRITERIA = ALPHA(.05)
  /DESIGN = a subject(a) c c*subject(a) .
```

or not…

Choose whether or not to include the C × S / A term… If you do include it, the C × S / A term is calculated and itself assessed against the residual MS_{error}, whereas otherwise C × S / A is part of the error term. This inevitably reduces the residual MS_{error} and will therefore improve power to detect effects of C (either as an effect of C or a C × S / A interaction), probably at the expense of power to detect the effect of A.

One thing worth noticing: SPSS assesses MS_A against a linear combination of $MS_{S/A}$ and the residual (what it calls MS_{error}). You might think that it should be assessed only against $MS_{S/A}$ — and this is what it will do if A and C are totally uncorrelated. It's possible to force SPSS to do this at any time with a custom hypothesis test using the syntax `/TEST = a VS subject(a)`. But this may not be a good idea, because if A and C are partially correlated, SPSS tries to sort things out. It calculates its error terms using Satterthwaite's (1946) denominator synthesis approach. If A and C are pretty much uncorrelated, you'll find that the linear combination it uses as its error term is heavily weighted towards $MS_{S/A}$ (e.g. $0.97 \times MS_{S/A} + 0.03 \times MS_{error}$). If they're correlated, the weighting will change (e.g. $0.239 \times MS_{S/A} + 0.761 \times MS_{error}$). And if A and C are substantially correlated, your interpretation may be very difficult in any case.

In any case, the easiest way to think about the calculations going on in this sort of analysis is to view each test as a comparison of two models (see section on GLMs, p. 84→). For example, assuming we're using the usual method (SPSS's Type III sums of squares) for partialling out

the effects of mutually correlated predictors, the test of the effect of C is a test of the difference between a full model, including C, and a restricted model including every effect but C, and so on for all the other terms.

Not entirely trivial to accomplish with the SPSS menus. Using the menus, choose *Analyze → General Linear Model → Univariate.* Enter A as a fixed factor; enter Subject as a random factor; enter C as a covariate. Since SPSS will not give you the correct model by default (it will include S), you then need to edit the *Model* directly before running the analysis. Untick *'Full factorial'* by ticking *'Custom'*. Enter the desired terms as listed in the ANOVA table.

7.17.2. The covariate and factor interact

Alternative names

Examples

- One of my examples, so I do hope it's appropriate. Subjects are assigned to two groups (A_1 brain lesion, A_2 sham). They are given a task in which they have a choice between two levers. Lever A delivers a single food pellet with probability $p = 1$. Lever B delivers four pellets, but with a probability that ranges from 1 to 0.0625; the probability changes in steps and the rats have an opportunity to experience the probability currently in force before they choose between levers A (small, certain reward) and B (large, uncertain reward). The dependent variable is the proportion of trials on which they choose lever B. We could analyse these with two factors: A (group: lesion / sham; between subjects) and B (probability: 1 / 0.5 / 0.25 / 0.125 / 0.0625; within subjects). But since delivery of the large reward is under the control of a random process, the probability *experienced* by the subjects may not always match the *programmed* probability (e.g. if they have 10 trials and the programmed probability is 0.5, it's perfectly possible that they get 3 rewarded and 7 unrewarded trials, giving an experienced probability of only 0.3). So rather than using programmed probability as a within-subjects factor, we could use experienced probability as a within-subjects covariate (call it C). We can then ask whether the probability influenced choice (main effect of C), whether the lesion influenced choice (main effect of A), and whether the lesion influenced the effect of probability (A × C interaction).

- Subjects are assigned to two groups (A_1 brain lesion, A_2 sham). They respond freely on two levers, left and right, to receive food pellets. Both levers deliver food with an element of randomness. Rats are tested for several sessions. Across sessions, the relative number of pellets delivered by each lever varies. For each session, we calculate the proportion of responses allocated to the left lever — the relative response distribution (dependent variable) — and the proportion of the total number of pellets that were earned by responding on the left lever — the relative reinforcer distribution (C). Both are continuous, rather than discrete, variables. Did the reinforcer distribution influence responding (main effect of C)? Did the lesion influence responding (main effect of A)? Did the lesion influence the way the animals responded to the reinforcer distribution (interaction between C and A)?

Notes

In terms of the model, this is logically equivalent to the 'one between-subjects factor, one within-subjects factor' design discussed earlier (q.v.). The computerized ANOVA process used by SPSS, based on a general linear model (GLM), does not care whether predictor variables are discrete (factors) or continuous (covariates), except in the way that it builds its default model (which we need to override here).

Model description (S = subjects)

depvar = $A \times (C_{cov} \times S)$

Model

Well, I'm making this up again; I would guess the full model would be essentially the same as the 'one between, one within' design discussed earlier (q.v.):

$$Y_{ijk} = \mu + \alpha_i + \pi_{j/i} + c_k + \alpha c_{ik} + \pi c_{jk/i} + \varepsilon_{ijk}$$

where

- Y_{ijk} is the dependent variable for subject j in condition A_i
- μ is the overall mean
- α_i is the contribution from a particular level (level i) of factor A

- $\pi_{j/i}$ is the contribution from a particular person or subject (subject *j*), who only serves *within* condition A_i ('subject within group', or S / A)
- (There is no straightforward interaction of A with S: every subject is only measured at one level of A, so this term would be indistinguishable from the subject-only effect $\pi_{j/i}$.)
- c_k is the contribution from the covariate C for subject *j* (call it C_j for the moment)
- αc_{ik} is the contribution from the interaction of A_i and C_j
- πc_{jk} is the contribution from the interaction of C_k with subject *j* (who only serves within condition A_i) — **if you choose to include it (see above)**
- ε_{ijk} is everything else (the experimental error associated with measuring person *j*, who always experiences treatment A_i, with covariate contribution C_k).

Sources of variance

If A and C are uncorrelated, we could partition the variance like this:

$$SS_{total} = SS_{between\ subjects} + SS_{within\ subjects}$$
$$SS_{between\ subjects} = SS_A + SS_{S/A}$$
$$SS_{within\ subjects} = SS_C + SS_{C\times A} + SS_{C\times S/A} + SS_{error}$$

or if you don't include $SS_{C\times S/A}$, you'd just write the within-subjects bit like this:

$$SS_{within\ subjects} = SS_C + SS_{C\times A} + SS_{error}$$

However, if A and C are correlated, the sources of variance will not be independent and will not add up to SS_{total}.

ANOVA table

If A and C are uncorrelated, the ANOVA table would look like this:

Source	d.f.	SS	*F*
Between subjects:	$s-1 = an-1$		
A	$a-1$	SS_A	$MS_A / MS_{S/A}$
error S / A	$a(n-1)$	$SS_{S/A}$	
Within subjects:	$N-s$		
C_{cov}	1	SS_C	$MS_C / MS_{error\text{-}within}$
$C_{cov} \times A$	$a-1$	$SS_{C\times A}$	$MS_{C\times A} / MS_{error\text{-}within}$
$C_{cov} \times S / A$	$a(n-1)$	$SS_{C\times S/A}$	$MS_{C\times S/A} / MS_{error\text{-}within}$
error-within	$N-s-an$	$SS_{error\text{-}within}$	
Within subjects in a model that doesn't include $C_{cov} \times S / A$:			
C_{cov}	1	SS_C	$MS_C / MS_{error\text{-}within}$
$C_{cov} \times A$	$a-1$	$SS_{C\times A}$	$MS_{C\times A} / MS_{error\text{-}within}$
error-within	$N-s-a$	$SS_{error\text{-}within}$	
Total	$N-1$	SS_{total}	

where *a* is the number of levels of factor A and *N* is the total number of observations (= *aun*), *n* is the number of subjects per group (where 'group' is defined by factor A), and *s* is the total number of subjects.

SPSS technique

Data layout:

A	*Subject*	*C*	*depvar*
1	*1*	*datum*	*datum*
1	*1*	*datum*	*datum*
1	*1*	*datum*	*datum*
1	*2*	*datum*	*datum*
1	*2*	*datum*	*datum*
1	*2*	*datum*	*datum*
…			
2	*7*	*datum*	*datum*
2	*7*	*datum*	*datum*
2	*7*	*datum*	*datum*
2	*7*	*datum*	*datum*

2	*8*	*datum*	*datum*
2	*8*	*datum*	*datum*
…			

Syntax:

```
GLM
  depvar BY a subject  WITH c
  /RANDOM = subject
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /PRINT = DESCRIPTIVE HOMOGENEITY
  /CRITERIA = ALPHA(.05)
  /DESIGN = a subject(a)
            c c*a c*subject(a) .
```

or not…

Choose whether or not to include the C × S / A term… If you do include it, the C × S / A term is calculated and itself assessed against the residual MS_{error}, whereas otherwise C × S / A is part of the error term. This inevitably reduces the residual MS_{error} and improves power to detect terms involving C (that is, C, C × A, and C × S / A), probably at the expense of power to detect the effect of A.

Note that SPSS calculates its error terms using appropriate linear combinations to deal with any correlation between A and C (see above).

Not entirely trivial to accomplish with the SPSS menus. Using the menus, choose *Analyze → General Linear Model → Univariate.* Enter A as a fixed factor; enter Subject as a random factor. Since SPSS will not give you the correct model by default (it will include S and not include C × A), you then need to edit the *Model* directly before running the analysis. Untick *'Full factorial'* by ticking *'Custom'*. Enter the desired terms as above.

7.18 Hierarchical designs: two or more levels of 'relatedness' in measurement

7.18.1. Subjects within groups within treatments (S/G/A)

Alternative names

- Split-split plot design
- Double-split design
- Doubly-nested design
- Hierarchical design
- Bloody complicated

Example

The simplest hierarchical design (Myers & Well, 1995, pp. 321): *subjects* (S) are tested in *groups* (G). Different groups are assigned to different levels of some *treatment* (A). One subject is only ever in one group, and one group is only ever in one treatment. This design can be written S / G / A (subjects *within* groups *within* treatments). Specific examples:

- Primary school pupils are taught in classes. We assign several classes to one teaching method, several other classes to a different teaching method, and so on. One pupil is only ever in one class; one class only ever uses one teaching method. The design is pupils *within* classes *within* teaching methods. Pupil and class are random factors.
- Different methods of rearing rats might be compared, with each rearing method being applied to several litters of rats. (Rats within a litter are related genetically, so we should take into account this potential source of correlation between scores of rats from the same litter. Stating the same thing in a different way, two randomly-selected rats may differ not only because they are different individuals, or because they experienced different treatments, but because they come from different litters.) The design is rats *within* litters *within* rearing methods. Rat and litter are random factors.

Model

An individual score might be represented as Y_{ijk}, where
i = 1, 2, … a (number of treatment levels)
j = 1, 2, … g (number of groups within a treatment level)
k = 1, 2, … n (number of subjects in a group)

Then

$$Y_{ijk} = \mu + \alpha_i + \gamma_j + \varepsilon_{ijk}$$

or

$$Y_{ijk} = \mu + (\mu_i - \mu) + (\mu_{ij} - \mu_i) + (Y_{ijk} - \mu_{ij})$$
$$Y_{ijk} - \mu = (\mu_i - \mu) + (\mu_{ij} - \mu_i) + (Y_{ijk} - \mu_{ij})$$

where

- Y_{ijk} is the dependent variable in condition A_i, G_j for subject k
- μ is the overall mean
- α_i is the contribution from level i of factor A (A_i): $\alpha_i = \mu_{A_i} - \mu$ and $\sum \alpha_i = 0$.
- γ_j is the contribution from level j of group G in condition A_i (G_{ij}) relative to the mean of A_i: $\gamma_j = \mu_{G_j} - \mu_{A_i}$ and $\sum \gamma_j = 0$.
- ε_{ijk} is everything else (the deviation of subject k from its group mean G_{ij}): $\varepsilon_{ijk} = Y_{ijk} - \mu_{ij}$.

If we sum and square both sides (and eliminate cross-product terms that sum to zero), we get:

$$\sum_i \sum_j \sum_k (Y_{ijk} - \mu)^2 = ng \sum_i (\mu_i - \mu)^2 + n \sum_i \sum_j (\mu_{ij} - \mu_i)^2 + \sum_i \sum_j \sum_k (Y_{ijk} - \mu_{ij})^2$$
$$SS_{total} = SS_A + SS_{G/A} + SS_{S/G/A}$$

Sources of variance

Subject and group are random factors; A is a fixed factor. We'd write the model like this:

$$SS_{total} = SS_{between\text{-}groups} + SS_{within\text{-}groups}$$
$$SS_{between\text{-}groups} = SS_A + SS_{G/A}$$
$$SS_{within\text{-}groups} = SS_{S/G/A}$$

so

$$SS_{total} = SS_A + SS_{G/A} + SS_{S/G/A}$$

We would state G / A as 'group within A', and S / G / A as 'subject within group within A', or simply 'subject within group'. Similarly,

$$df_{\text{total}} = df_{\text{A}} + df_{\text{G/A}} + df_{\text{S/G/A}}$$

ANOVA table

Source	d.f.	SS	*F*
Between groups:	ag–1		
A	a–1	SS_A	$MS_A / MS_{G/A}$
G / A	$a(g$–1)	$SS_{G/A}$	$MS_{G/A} / MS_{S/G/A}$
Within groups:			
S / G / A	$ag(n$–1)	$SS_{S/G/A}$	
Total	N–1 = agn – 1	SS_{total}	

where N is the total number of observations and a, g, and n are as defined above. Note that the error term for A is G / A, and the error term for G / A is S / G / A.

SPSS technique

A	*G*	*Subject*	*depvar*
1	1	1	*datum*
1	1	2	*datum*
1	1	3	*datum*
1	1	4	*datum*
1	2	5	*datum*
1	2	6	*datum*
1	2	7	*datum*
1	2	8	*datum*
2	3	9	*datum*
2	3	10	*datum*
2	3	11	*datum*
2	3	12	*datum*
2	4	13	*datum*
2	4	14	*datum*
2	4	15	*datum*
2	4	16	*datum*
…			

It doesn't matter if you use the same identifiers to code groups within different levels of A. For example, you can call the A_1 groups '1' and '2' and the A_2 groups '3' and '4', as I've done above — or you can call the A_1 groups '1' and '2' and the A_2 groups '1' and '2' again. Since the design 'knows' that groups are nested within levels of A, it doesn't care about how you label them. (Of course, each group must have a unique name *within* each level of A.)

```
GLM depvar BY a g subject
  /RANDOM = g subject
  /DESIGN = a g(a) subject(g(a)) .
```

Further notes

It's a common mistake to use an experiment with this kind of design but not to put the 'Group' factor into the analysis. People often analyse these kinds of data only taking into account the A factor. That will generally **overestimate** the *F* ratio for A (give a lower *p* value than it should) (Myers & Well, 1995, pp. 325-7). On the other hand, if Group has few *df*, the value of MS_G (= SS_G / df_G) will be large and we will have low power to detect effects of A. The alternative model is to **ignore** the effect of G (what most people do without thinking about it):

$$SS_{\text{total}} = SS_{\text{A}} + SS_{\text{S/A}}$$

where $SS_{S/A}$ is the pool of G / A and S / G / A. This is what you get when you run a one-way ANOVA, ignoring the effect of G. In general, $E(MS_{S/A})$ is less than $E(MS_{G/A})$, so you're more likely to find a 'significant' effect of A (Myers & Well, 1995, p. 326). Myers & Well (1995, pp. 151, 327) recommend that you only pool (ignore the effect of G) when you've already run an analysis with G included and this preliminary test of the effect of G was not significant at the α = .25 level, *and* you have prior reason to believe that the things you're pooling over reflect only chance variability (in this case, that you have prior reason to think that groups don't differ systematically).

As Myers & Well (1995, p. 339) put it, 'wishing some variance component [e.g. G] to be zero does not make it so, and the price of wrongly assuming that the component is zero is ordinarily a Type 1 error in testing treatment effects of interest [i.e. declaring the effect of A to be significant when it isn't].

If you can't legitimately pool, then you need to have a high value of g (many groups), so you get high $df_{G/A}$ and therefore low $MS_{G/A}$, and therefore good power to detect effects of A (which uses $MS_{G/A}$ as its error term). This should be fairly obvious, although many people fail to realize it: if one primary school class is taught using one method and another is taught using another method, is a difference in class means due to different methods (A) or to difference in the personal interactions within the two classes (G)? They're confounded.

7.18.2. Groups versus individuals

If you need to compare the effects of *being in a group* ('group' condition) to the effect of *not being in a group* ('individual' condition), there is a special analytical technique (Myers & Well, 1995, pp. 327-9). For example, if 15 students study a topic individually, while another 15 students study the topic in five discussion groups of three, you can analyse the effect of being in a group. This is a fairly common problem in social psychology.

7.18.3. Adding a further within-group, between-subjects variable (S/GB/A)

Example

Subjects (S) are part of groups (G). Within each group, subjects are either anxious or not (anxiety: B). Sets of groups are given different treatments (A). So G is crossed with B (all groups have anxious and non-anxious subjects; anxious and non-anxious subjects are found in all groups) but subjects are nested within GB (a subject is only part of one group and is either anxious or not) and groups are nested within treatments. The model can be written S / GB / A (or S / BG / A).

Model

An individual score might be represented as Y_{ijkl}, where

i = 1, 2, … a (number of A treatment levels)
j = 1, 2, … b (number of B levels within a group, or within a treatment level)
k = 1, 2, … g (number of groups within a treatment level)
l = 1, 2, … n (number of subjects in a group)

Then

$$Y_{ijk} = \mu + \alpha_i + \gamma_{j/i} + \beta_j + \alpha\beta_{ij} + \gamma\beta_{kj/i} + \varepsilon_{ijkl}$$

There are no interactions involving subjects (because subjects cross with none of the other three variables: one subject only ever experiences one level of G, B, and A). G does not cross with A, so there is no AG or ABG term.

Sources of variance

Subject and group are random factors; A and B are fixed factors. We'd write the model like this:

$$SS_{total} = SS_{between\text{-}groups} + SS_{within\text{-}groups}$$
$$SS_{between\text{-}groups} = SS_A + SS_{G/A}$$
$$SS_{within\text{-}groups} = SS_B + SS_{AB} + SS_{GB/A} + SS_{S/GB/A}$$

so

$$SS_{total} = SS_A + SS_{G/A} + SS_B + SS_{AB} + SS_{GB/A} + SS_{S/GB/A}$$

We would state G / A as 'group within A', and S / G / A as 'subject within group within A', or simply 'subject within group'. Similarly,

$$df_{total} = df_A + df_{G/A} + df_B + df_{AB} + df_{GB/A} + df_{S/GB/A}$$

ANOVA table

Source	d.f.	SS	*F*
Between G:	ag–1		
A	a–1	SS_A	$MS_A / MS_{G/A}$
G / A	$a(g$–1)	$SS_{G/A}$	
Within G:	$ag(bn$–1)		
B	b–1	SS_B	$MS_A / MS_{GB/A}$
AB	(a–1)(b–1)	SS_{AB}	$MS_{AB} / MS_{GB/A}$
GB / A	$a(g$–1)(b–1)	$SS_{GB/A}$	$MS_{GB/A} / MS_{S/GB/A}$
S / GB / A	$gba(n$–1)	$SS_{S/GB/A}$	
Total	N–1 = $agbn$ – 1	SS_{total}	

where N is the total number of observations and a, g, and n are as defined above. Note that the error term for A is G / A, and the error term for G / A is S / G / A.

SPSS technique

A	*G*	*B*	*Subject*	*depvar*
1	1	1	1	*datum*
1	1	1	2	*datum*
1	1	2	3	*datum*
1	1	2	4	*datum*
1	2	1	5	*datum*
1	2	1	6	*datum*
1	2	2	7	*datum*
1	2	2	8	*datum*
2	3	1	9	*datum*
2	3	1	10	*datum*
2	3	2	11	*datum*
2	3	2	12	*datum*
2	4	1	13	*datum*
2	4	1	14	*datum*
2	4	2	15	*datum*
2	4	2	16	*datum*
…				

See the notes about group coding above.

```
GLM depvar BY a g b subject
  /RANDOM = g subject
  /DESIGN = a g(a)
            b a*b g*b(a) subject(g*b(a)) .
```

That seems to work (Myers & Well, 1995, p. 332, but note their typo for the F value for the effect of B).

7.18.4. Adding a within-subjects variable (US/GB/A)

Example

We take the previous model to begin with: subjects (S) are part of groups (G). Within each group, subjects are either anxious or not (anxiety: B). Sets of groups are given different treatments (A). Now we measure each subject four times (trial: U). U is crossed with S (since every subject experiences all four trials). So our design can be written US / GB / A (Myers & Well, 1995, p. 333).

Model

See sources of variance below, which follow directly from the model and are easier to grasp.

Sources of variance

The previous model describes the between-subjects variability. We just need to add within-subjects terms — U, and the interaction of U with each of the between-subjects sources from the last model:

$SS_{total} = SS_{between\text{-}groups} + SS_{within\text{-}groups}$
$SS_{between\text{-}groups} = SS_A + SS_{G/A}$
$SS_{within\text{-}groups} = SS_{within\text{-}groups\text{-}between\text{-}subjects} + SS_{within\text{-}subjects}$
$SS_{within\text{-}groups\text{-}between\text{-}subjects} = SS_B + SS_{AB} + SS_{GB/A}$

$$SS_{within\text{-}subjects} = SS_U + SS_{UA} + SS_{UG/A} + SS_{UB} + SS_{UAB} + SS_{UGB/A} + SS_{US/GB/A}$$

ANOVA table

We have *g* groups at each of *a* levels of A. Within each group, there are *b* levels of B and *n* subjects at each of those levels. So we have *bn* subjects in each of *ag* groups, for a total of *agbn* subjects. Each subject provides one score at *u* levels of U — *agbnu* scores in all.

Source	d.f.	SS	*F*
Between G:	$ag–1$		
A	$a–1$	SS_A	$MS_A / MS_{G/A}$
G / A	$a(g–1)$	$SS_{G/A}$	
Within G, between S:	$ag(bn–1)$		
B	$b–1$	SS_B	$MS_A / MS_{GB/A}$
AB	$(a–1)(b–1)$	SS_{AB}	$MS_{AB} / MS_{GB/A}$
GB / A	$a(g–1)(b–1)$	$SS_{GB/A}$	$MS_{GB/A} / MS_{S/GB/A}$
S / GB / A	$gba(n–1)$	$SS_{S/GB/A}$	
Within S:	$agbn(u–1)$		
U	$u–1$	SS_U	$MS_U / MS_{UG/A}$
UA	$(u–1)(a–1)$	SS_{UA}	$MS_{UA} / MS_{UG/A}$
UG / A	$(u–1)a(g–1)$	$SS_{UG/A}$	$MS_{UG/A} / MS_{UGB/A}$
UB	$(u–1)(b–1)$	SS_{UB}	$MS_{UB} / MS_{UGB/A}$
UAB	$(u–1)(a–1)(b–1)$	SS_{UAB}	$MS_{UAB} / MS_{UGB/A}$
UGB / A	$(u–1)a(g–1)(b–1)$	$SS_{UGB/A}$	$MS_{UGB/A} / MS_{US/GB/A}$
US / GB / A	$(u–1)gba(n–1)$	$SS_{US/GB/A}$	
Total	$N–1 = agbnu – 1$	SS_{total}	

Top tip: to check your *df* add up to the total, it's quick to use Mathematica®. For example, `Simplify[(u-1) + (u-1)(a-1) + (u-1)a(g-1) + (u-1)(b-1) + (u-1)(a-1)(b-1) + (u-1)a(g-1)(b-1) + (u-1)g*b*a(n-1)]` gives `a b g n (-1 + u)`. When you really can't work out the appropriate error terms, you can enter the model into SPSS and see what it used.

SPSS technique

A	*G*	*B*	*Subject*	*U*	*depvar*
1	1	1	1	1	*datum*
1	1	1	1	2	*datum*
1	1	1	1	3	*datum*
1	1	1	1	4	*datum*
1	1	1	2	1	*datum*
1	1	1	2	2	*datum*
1	1	1	2	3	*datum*
1	1	1	2	4	*datum*
1	1	2	3	1	*datum*

… and so on. Just the same as the previous example but with the new U column.

See the notes about group coding above.

```
GLM depvar BY a g b subject u
  /RANDOM = g subject
  /DESIGN = a g(a)
            b a*b g*b(a) subject(g*b(a))
            u u*a u*g(a) u*b u*a*b u*g*b(a) u*subject(g*b(a)) .
```

7.18.5. Nesting within-subjects variables, such as V/US/A

Example

We have five experienced subjects and five novice subjects (factor A for experience; between-subjects factor; fixed factor; $a = 2$; $n = 5$; total of $an = 10$ subjects). Every subject is required to solve 12 problems, of which 4 are easy, 4 are of intermediate difficulty, and 4 are hard (factor U for difficulty; factor V for problem; $u = 3$; $v = 4$). This is almost the same as a **one between, two within** design except that V is nested within U, not crossed with it (Myers & Well, 1995, p. 334-338).

Model

See sources of variance below, which follow directly from the model and are easier to grasp.

Sources of variance

We start by partitioning into between-subjects and within-subjects variability:

$SS_{total} = SS_{between\text{-}subjects} + SS_{within\text{-}subjects}$
$SS_{between\text{-}subjects} = SS_{A} + SS_{S/A}$

To partition the within-subjects variability, we can first view the design as involving *uv* levels of 'stimuli'. That is, in general, we begin partitioning within-subjects variability by using our smallest experimental units. We also cross *stimuli* with all the between-subject sources:

$$SS_{within\text{-}subjects} = SS_{stimuli} + SS_{stimuli \times A} + SS_{stimuli \times S/A}$$

We now partition the variability due to stimuli and its interactions:

$$SS_{stimuli} = SS_{U} + SS_{V/U}$$

… and cross those with A and S / A in turn:

$$SS_{stimuli \times A} = SS_{AU} + SS_{AV/U}$$
$$SS_{stimuli \times S/A} = SS_{SU/A} + SS_{SV/AU}$$

We can partition the *df* in the same way. Actual values for the *df*s are in square brackets:

$df_{total}\ [abcn–1] = df_{between\text{-}subjects}\ [an–1] + df_{within\text{-}subjects}\ [an(uv–1)]$
$df_{between\text{-}subjects} = df_{A}\ [a–1] + df_{S/A}\ [a(n–1)]$
$df_{within\text{-}subjects} = df_{stimuli}\ [uv–1] + df_{stimuli \times A}\ [(a–1)(uv–1)] + df_{stimuli \times S/A}\ [a(n–1)(uv–1)]$
$df_{stimuli} = df_{U}\ [u–1] + df_{V/U}\ [u(v–1)]$
$df_{stimuli \times A} = df_{AU}\ [(a–1)(u–1)] + df_{AV/U}\ [u(a–1)(v–1)]$
$df_{stimuli \times S/A} = df_{SU/A}\ [a(n–1)(u–1)] + df_{SV/AU}\ [au(n–1)(v–1)]$

A way of checking the design is to list all factors, random and fixed, noting any nesting. We have four: **A, S / A, U, V / U.** Now we consider all possible cross products of these factors. We write 'no' next to them if it's not legitimate to cross them — for example, if S is nested in A, it cannot also cross with it.

A × S / A	*No*
A × U	**AU**
A × V / U	**AV / U**
S / A × U	**SU / A**
S / A × V / U	**SV / AU**
C × V / U	*No*

The four factors we started with plus the four cross-products generated above are the terms of interest. We should also consider crossing more than two factors, but in this design no legitimate terms would turn up (for example, A × U × V / U is not legitimate because V cannot be nested within U and still cross with it). Once we've specified our factors, we can enter them into SPSS's design syntax.

ANOVA table

Source	d.f.	SS	*F*
Between S:	an–1		
A	a–1	SS_A	$MS_A / MS_{S/A}$
S / A	$a(n–1)$	$SS_{S/A}$	$MS_{S/A} / MS_{SU/A}$
Within S:	$an(uv–1)$		
U	u–1	SS_U	$MS_U / MS_{SU/A}$
AU	$(a–1)(u–1)$	SS_{AU}	$MS_{AU} / MS_{SU/A}$
V / U	$u(v–1)$	$SS_{V/U}$	$MS_{V/U} / MS_{SV/AU}$
AV / U	$u(a–1)(v–1)$	$SS_{AV/U}$	$MS_{AV/U} / MS_{SV/AU}$
SU / A	$a(n–1)(u–1)$	$SS_{SU/A}$	$MS_{SU/A} / MS_{SV/AU}$
SV / AU	$au(n–1)(v–1)$	$SS_{SV/AU}$	
Total	$N–1 = abcn–1$	SS_{total}	

The *F* ratios depend on which factors are treated as fixed and which as random (because that determines the EMS values); the ratios presented above are for when S is random and A, V, and U are all fixed. Actually, our example suggests that V, which we write in full as V / U ('specific

problem of a certain difficulty') should be random; in that situation, the appropriate error term must be synthesized as a linear combination of other terms. It seems that SPSS and BDMP8V do this in slightly different ways (Myers & Well, 1995, p. 337, versus SPSS analysis of the same data).

SPSS technique

A	*Subject*	*U*	*V*	*depvar*
1	1	1	1	*datum*
1	1	1	2	*datum*
1	1	1	3	*datum*
1	1	1	4	*datum*
1	1	2	1	*datum*
1	1	2	2	*datum*
1	1	2	3	*datum*
1	1	2	4	*datum*
1	1	3	1	*datum*
1	1	3	2	*datum*
1	1	3	3	*datum*
1	1	3	4	*datum*
1	2	1	1	*datum*
1	2	1	2	*datum*
1	2	2	3	*datum*
1	2	2	4	*datum*
…				
2	6	1	1	*datum*
2	6	1	2	*datum*
2	6	1	3	*datum*
2	6	1	4	*datum*
…				

```
GLM depvar BY a subject u v
  /RANDOM = subject
  /DESIGN = a subject(a)
            u v(u) a*u a*v(u) subject*u(a) subject*v(a*u) .
```

If V is a random factor too, you'd want `/RANDOM = subject v`, and so on.

7.18.6. The split-split plot design

Alternative names

- Split-split plot, completely randomized design
- Pretty awful

Example

(1) An agricultural example (Winer *et al.*, 1991, pp. 368-9). An orchard is divided into plots. Each level of factor A is assigned at random to *n* plots, so there are *an* plots in total. Each of the plots is then divided into *b* subplots, and the *b* levels of factor B are assigned to them at random. Finally, each of the subplots is divided into *c* sub-subplots, and the *c* levels of factor C are assigned to them at random. Thus the experimental unit for A is the whole plot, the experimental unit for B is the subplot, and the experimental unit for C is the sub-subplot.

Since the sub-subplots are nested within the subplots, and the subplots are nested within the whole plots, factor C is nested under the subplots and factor B is nested under the whole plots. Factor A is partially confounded with groups of whole plots.

(2) A rat example. Rats are implanted with dialysis probes in *either* the medial prefrontal cortex (A_1) or orbitofrontal cortex (A_2). They are then assigned to triplets. One rat in each triplet chooses between two levers offering alternative reinforcers in a task (B_1). Another (B_2) is offered only the lever chosen by the master rat. A third (B_3) is given the reinforcer chosen by the master rat, without any opportunity to press a lever. Finally, all rats are dialysed at five time points ($C_1 \ldots C_5$).

Data from different levels of factor A (probe site) are unrelated. Data from different levels of factor B (choice type) may be related to each other, because they all come from the same triplet. Data from different levels of factor C (time) may be related to each other, because

they all come from the same rat. However, we cannot wholly distinguish rat individuality from the effects of choice type.

This design is equivalent to the agricultural one: Triplet ≡ Plot, and Rat ≡ Subplot. As before, A (lesion) is the whole-plot factor (a triplet either gets medial prefrontal or orbitofrontal probes), B (choice type) is the subplot factor (within a triplet, a rat is either a master, lever-yoked or reinforcer-yoked rat), and C (time) is the sub-subplot factor (every rat gets dialysed at five time points, so the 'sub-subplot' is the combination of a particular rat at a particular time).

Model

$$Y_{ijkm} = \mu + \alpha_i + \pi_{m(i)} + \beta_j + \alpha\beta_{ij} + \pi'_{m(ij)} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{ik} + \alpha\beta\gamma_{ijk} + \pi''_{m(ijk)} + \varepsilon_{ijkm}$$

where

- Y_{ijkm} is the value of an observation in condition A_i, $plot_m$, B_j, and C_k
- μ is the grand mean
- α_i is the contribution of A_i
- β_j is the contribution of B_j
- γ_k is the contribution of C_k
- $\alpha\beta_{ij}$, $\alpha\gamma_{ik}$, $\beta\gamma_{jk}$ and is the contribution of the A_iB_j, A_iC_k, and B_jC_k interactions, respectively
- $\pi_{m(i)}$ is the contribution of plot *m* (which only ever experiences A_i)
- $\pi'_{m(ij)}$ is the contribution of the subplot in plot *m* that experiences A_iB_j
- $\pi''_{m(ijk)}$ is the contribution of the sub-subplot in plot *m* that experiences $A_iB_jC_k$
- ε_{ijkm} is everything else (error)

Sources of variance

For our rat example, we'd call triplet 'plot' and rat 'subplot' (and consider them as random factors, while the others are fixed factors). We'd write the model like this:

$$SS_{total} = SS_{between\text{-}plots} + SS_{within\text{-}plots}$$
$$SS_{between\text{-}plots} = SS_A + SS_{plot\times A}$$
$$SS_{within\text{-}plots} = SS_{between\text{-}subplots\text{-}within\text{-}plots} + SS_{within\text{-}subplots}$$
$$SS_{between\text{-}subplots\text{-}within\text{-}plots} = SS_B + SS_{B\times A} + SS_{B\times plot/A}$$
$$SS_{within\text{-}subplots} = SS_C + SS_{C\times A} + SS_{C\times B} + SS_{C\times A\times B} + SS_{within\text{-}subplot\ error\ C\times plot/AB}$$

ANOVA table

Source	d.f.	SS	*F*
Between plots:	$an-1$		
A	$a-1$	SS_A	$MS_A / MS_{plot\times A}$
error plot × A ('whole-plot residual')	$a(n-1)$	$SS_{plot\times A}$	
Within plots, between subplots:	$an(b-1)$		
B	$b-1$	SS_B	$MS_B / MS_{B\times plot/A}$
B × A	$(b-1)(a-1)$	$SS_{A\times B}$	$MS_{A\times B} / MS_{B\times plot/A}$
error B × plot / A ('subplot residual')	$a(b-1)(n-1)$	$SS_{B\times plot/A}$	
Within subplots:	$abn(c-1)$		
C	$c-1$	SS_C	$MS_C / MS_{C\times plot/AB}$
C × A	$(c-1)(a-1)$	$SS_{C\times A}$	$MS_{C\times A} / MS_{C\times plot/AB}$
C × B	$(c-1)(b-1)$	$SS_{C\times B}$	$MS_{C\times B} / MS_{C\times plot/AB}$
C × A × B	$(c-1)(a-1)(b-1)$	$SS_{C\times A\times B}$	$MS_{C\times A\times B} / MS_{C\times plot/AB}$
error C × plot / AB ('sub-subplot residual')	$ab(c-1)(n-1)$	$SS_{C\times plot/AB}$	
Total	$N-1 = abcn-1$	SS_{total}	

where a is the number of levels of factor A, etc., N is the total number of observations (= $abcn$), and n is the number of subjects. The F ratios above assume that Plot is random and A, B, C are fixed.

For the rat example, simply read 'triplet' instead of 'plot' and 'rat' instead of 'subplot'.

Result! Agrees with Winer (1991, p. 369, although there are typos in his ANOVA table; 'within sub-subplots' is certainly a mistake).

SPSS technique

A	*Plot*	*B*	*C*	*depvar*
1	1	1	1	*datum*
1	1	1	2	*datum*
1	1	1	3	*datum*
1	1	1	4	*datum*
1	1	1	5	*datum*
1	1	2	1	*datum*
1	1	2	2	*datum*
1	1	2	3	*datum*
1	1	2	4	*datum*
1	1	2	5	*datum*
1	1	3	1	*datum*
1	1	3	2	*datum*
1	1	3	3	*datum*
1	1	3	4	*datum*
1	1	3	5	*datum*
1	2	1	1	*datum*
1	2	1	2	*datum*
1	2	1	3	*datum*
1	2	1	4	*datum*
1	2	1	5	*datum*
…				
2	8	1	1	*datum*
2	8	1	2	*datum*
2	8	1	3	*datum*
2	8	1	4	*datum*
2	8	1	5	*datum*
…				

It doesn't matter whether you specify unique labels for nested factors or not — what I mean by this is that you can code 'plot' from 1, 2… for the A_1 condition and carry on counting (8, 9, …) for the A_2 condition, or you can start numbering 'plot' from 1 again in the A_2 condition. Since the design 'knows' that plot is nested within A (one plot only gets one level of A), it won't get confused.

```
GLM depvar BY plot a b c
  /RANDOM = plot
  /DESIGN = a plot*a
            b b*a b*plot(a)
            c c*a c*b c*a*b .
```

Top tip: when faking data to analyse complex models, ensure that you don't over- or under-specify your model! MRFA pointed out that I had been stupid in my initial attempt at this example, which included a 'rat' (subplot) term: because a triplet × B [choicetype] combination uniquely specifies a rat in this example, there's no 'room' in the design for a 'rat' term.

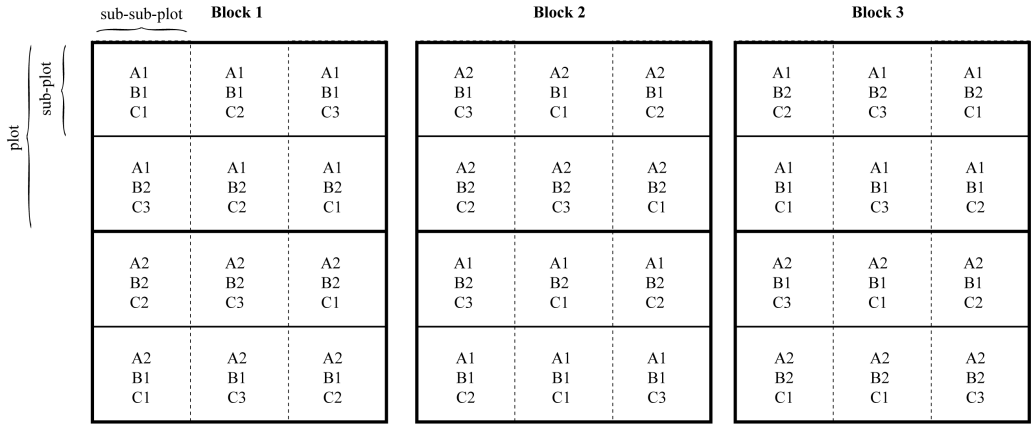
7.18.7. Three levels of relatedness

Alternative names

- Split-split plot, randomized complete block (RCB) design
- Horrendous

Examples

(1) The standard agricultural example: a randomized complete block design (RCBD) with blocks (also known as *replicates*), plots (A), subplots (B), and sub-sub-plots (C). Suppose A has two levels, B has two levels, and C has three levels. This would be a description of a field laid out like this:



Split-split plot, randomized complete block design.
The field is split into blocks.
Each block is split into two plots and factor A (2 levels) is assigned at random to the plots.
Each plot is split into two sub-plots and factor B (2 levels) is assigned at random to the sub-plots.
Each sub-plot is split into three sub-sub-plots and factor C (3 levels) is assigned at random to the sub-sub-plots.

Compare the RCB three-factor agricultural design illustrated in our consideration of the three-within-subject-factor design (U × V × W × S) (p. 118). 'Similarity' or 'relatedness' in agriculture often refers to geographical nearness; in the (U × V × W × S) design discussed earlier, adjacent mini-plots of land were likely to be similar by virtue of coming from the same block, but there was no other consistent relationship between geographical nearness and the factors U, V, or W. This design is a bit different. You can see here that two adjacent individual experimental units (the sub-sub-plots) are most likely be related by virtue of coming from the same Block, quite likely to be related by virtue of having the same value of the A factor, not quite as likely to be related on the B factor, and least likely to be related on the C factor.

Another way of putting it: blocks are crossed with A (all blocks experience all levels of A). Plots are nested within A (one plot only gets one level of A). Plots are crossed with B (all plots experience all levels of B). Subplots are nested within B (one subplot only gets one level of B). Sub-subplots are nested within C (one sub-subplot only experiences one level of C).

(2) Another agricultural example (Prescott *et al.*, 1999). Four blocks were used, spread across a forest (top-level factor: Block); the experiment was replicated across these blocks. Each block was divided into four plots, which were each fertilized with a different fertilizer, assigned to the plots at random. Small bags of leaf litter are placed in these plots (litter placement factor, or 'fertilizer that the litter is placed in': A_1, A_2, A_3, A_4). The bags themselves came either from the same plot or one of the other three plots in the same block (litter source factor, or 'fertilizer that the litter came from': B_1, B_2, B_3, B_4). The litter mass is then measured at different time points ($C_1 \ldots C_5$).

Notes

- This is different to a split-split plot design based on a completely randomized design (CRD), which doesn't have the 'block' factor.

- See www.ndsu.nodak.edu / ndsu / horsley / spspplot.pdf, the only worked example I've been able to find. That also says:

'The split-split plot arrangement is especially suited for three-or-more-factor experiments where different levels of precision are required for the factors evaluated. Three plot sizes correspond to the three factors: the largest plot for the main factor, the intermediate size plot for the subplot factor, and the smallest plot for the sub-subplot factor. There are three levels of precision with the main plot factor receiving the lowest precision, and the sub-subplot factor receiving the highest precision.'

Sources of variance Let's call blocks (replicates) R, the plot treatment A, the subplot treatment B, and the sub-subplot treatment C. Replicate will be a random factor; the others will be fixed. We'd write the model like this:

$$SS_{total} = SS_{between\text{-}replicates} + SS_{within\text{-}replicates}$$

$SS_{between\text{-}replicates} = SS_R$
$SS_{within\text{-}replicates} = SS_{between\text{-}plots\text{-}within\text{-}replicates} + SS_{within\text{-}plots}$
$SS_{between\text{-}plots\text{-}within\text{-}replicates} = SS_A + SS_{R\times A}$
$SS_{within\text{-}plots} = SS_{between\text{-}subplots\text{-}within\text{-}plots} + SS_{within\text{-}subplots}$
$SS_{between\text{-}subplots\text{-}within\text{-}plots} = SS_B + SS_{B\times A} + SS_{R\times B/A}$
$SS_{within\text{-}subplots} = SS_C + SS_{C\times A} + SS_{C\times B} + SS_{C\times A\times B} + SS_{within\text{-}subplot\text{-}error\ C\times R/AB}$

ANOVA table

Source	*df*	SS	F
Between replicates (R):			
R	$r-1$	SS_R	$MS_R / MS_{R\times A}$
Within replicates, between plots:			
A	$a-1$	SS_A	$MS_A / MS_{R\times A}$
error R × A	$(r-1)(a-1)$	$SS_{R\times A}$	
Within plots, between subplots:			
B	$b-1$	SS_B	$MS_B / MS_{R\times B/A}$
B × A	$(b-1)(a-1)$	$SS_{B\times A}$	$MS_{B\times A} / MS_{R\times B/A}$
error R × B / A	$a(r-1)(b-1)$	$SS_{R\times B/A}$	
Within subplots:			
C	$c-1$	SS_C	$MS_C / MS_{R\times C/AB}$
C × A	$(c-1)(a-1)$	$SS_{C\times A}$	$MS_{C\times A} / MS_{R\times C/AB}$
C × B	$(c-1)(b-1)$	$SS_{C\times B}$	$MS_{C\times B} / MS_{R\times C/AB}$
C × A × B	$(c-1)(a-1)(b-1)$	$SS_{C\times A\times B}$	$MS_{C\times A\times B} / MS_{R\times C/AB}$
error R × C / AB	$ab(r-1)(c-1)$	$SS_{R\times C/AB}$	
Total	$rabc-1$	SS_{total}	

SPSS technique

Rep	*A*	*B*	*C*	*depvar*
1	1	1	1	*datum*
1	1	1	2	*datum*
1	1	1	3	*datum*
1	1	2	1	*datum*
1	1	2	2	*datum*
1	1	2	3	*datum*
1	2	1	1	*datum*
1	2	1	2	*datum*
1	2	1	3	*datum*
1	2	2	1	*datum*
1	2	2	2	*datum*
1	2	2	3	*datum*
2	1	1	1	*datum*
2	1	1	2	*datum*
…				

You don't even need explicit 'plot', 'subplot', or 'sub-subplot' labels; all that information is contained in the design and the A / B / C factor labels.

```
GLM depvar BY r a b c
  /RANDOM = r
  /DESIGN = r a r*a
            b b*a b*r(a)
            c c*a c*b c*a*b .
```

7.19 Latin square designs

There are two approaches to Latin squares. One (the simplest) is to use a Latin square as an **experimental design** technique to ensure that some factor (e.g. time, order) is not confounded with experimental treatments. The other (more advanced but far preferable) is to do this, but also to use information about this factor (e.g. time, order) in the **analysis** — to take account of variability attributable to this factor to reduce the error variability and increase the power to detect effects of the treatment of interest. This can be much more complicated than I first thought!

For this section, I will abandon my previous convention of A, B… representing between-subjects factors and U, V… representing within-subjects factors, because this makes it easier to compare complex designs to the original sources.

7.19.1. Latin squares in experimental design

Here's an example of the 'novice' (experimental design only) approach that I've used (e.g. Cardinal *et al.*, 2003). Rats had intracranial cannulae implanted in their nucleus accumbens. They responded on a lever that delivered a stimulus previously paired with reinforcement (a conditioned reinforcer). Before the session, they were given intra-accumbens amphetamine at one of four doses (0, 3, 10, 20 μg per hemisphere). As I put it:

> Doses were counterbalanced in a Latin square design to eliminate differential carryover effects and separated by 24 h. The Latin square was of a digram-balanced design (Keppel, 1991, p. 339), in which each condition immediately precedes and follows the other conditions once (e.g. 1234, 3142, 2413, 4321).

What I meant was that if '1' represents one dose (0 μg), '2' represents the second, '3' the third, and '4' the fourth, the design looked like this:

	Day 1	Day 2	Day 3	Day 4
Pattern 1	1	2	3	4
Pattern 2	3	1	4	2
Pattern 3	2	4	1	3
Pattern 4	4	3	2	1

There were more than 4 subjects, so I allocated them to the four patterns at random. The idea is that the order of treatments 1–4 was counterbalanced appropriately. The square is a **Latin square** — an *n* by *n* grid containing the numbers 1 to *n* arranged in such a way that no row and no column contains the same number twice. If I had given all the subjects the treatments in the order 4, 3, 2, 1, and I found that treatment 4 gave higher responding than treatment 1, I wouldn't know if that was due to the difference in drug doses or the fact that with time, responding declines generally (extinction), or some other effect left over from the previous day's dose. So the Latin square **counterbalances for order.** There are good and bad Latin squares. The one above is **'digram-balanced',** which is good — every condition immediately precedes and follows the other conditions once. The one below is **cyclic,** which isn't so good:

	Day 1	Day 2	Day 3	Day 4
Pattern 1	1	2	3	4
Pattern 2	2	3	4	1
Pattern 3	3	4	1	2
Pattern 4	4	1	2	3

because in this design dose 1 is nearly always preceded by dose 4, and nearly always followed by dose 4 — clearly not as good as the digram-balanced one. The digram-balanced version controls for **sequence effects** better. However, digram balancing can only be done if there is an even number of treatment conditions (Keppel, 1991,

p. 339). Otherwise, there are procedures for selecting a random Latin square (Winer *et al.*, 1991, p. 674; Myers & Well, 1995, p. 346).

Anyway, back to the example. When I analysed these data, I ignored the 'day' factor. I simply took all the 'dose 1' scores, all the 'dose 2' scores, and so on, and entered the data with a within-subjects factor of Dose. This wasn't optimal — I could have used information about the Day factor as well. That could be more **efficient** (Myers & Well, 1995, p. 351), because it would remove variability attributable to Days to give better power to detect effects of Dose. Let's see how that can be done.

7.19.2. The analysis of a basic Latin square

Example

We test five monkeys (Myers & Well, 1995, p. 344) on discrimination learning under five different drug doses on five different test days. We use this Latin square (S = subject = R = row, C = column = day in this example, A = drug dose).

	C_1	**C_2**	**C_3**	**C_4**	**C_5**
S_1	A_1	A_2	A_4	A_3	A_5
S_2	A_3	A_1	A_5	A_2	A_4
S_3	A_4	A_3	A_1	A_5	A_2
S_4	A_2	A_5	A_3	A_4	A_1
S_5	A_5	A_4	A_2	A_1	A_3

Notes

See Myers & Well (1995, chapter 11); Winer (1991, chapter 11).

The Latin square analysis is potentially more efficient than the simple within-subjects analysis (ignoring Day) for the following reasons (Myers & Well, 1995, p. 351). The error term for the within-subjects (A × S) analysis, $MS_{S\times A}$, will be larger than the error term for the Latin square analysis as long as MS_C is larger than the Latin-square error term MS_{error}. However, the Latin square error term has fewer *df*, which reduces power. The relative contribution of the two effects can be calculated (Myers & Well, 1995, pp. 351-2).

When using Latin squares to counterbalance for order, it is vital that the position in the order (Day, in the example) *does not interact* with the treatment (Drug, in the example) (Keppel, 1991, p. 336-9; Winer *et al.*, 1991, p. 682). If one dose has a different effect when it's given first in the order to when it's given third in the order, we'd have to be very careful of the interpretation. It's worth **plotting treatment means against order** to check this assumption. If the effect of different doses reverses on different days, it's very hard to analyse or interpret (Keppel, 1991, p. 338) and we may be reduced to analysing only the data from the first test, which is uncontaminated by any prior effects, but which may have rather low statistical power.

We've seen that one major use of Latin squares is to **counterbalance order effects,** as shown here. But they have other uses. Latin squares were first used in agriculture to control for two **nuisance variables** (assigned to the rows and columns, with the assumption that the treatment effects do not interact with the row and column effects) (Winer *et al.*, 1991, p. 680). They may be extended to deal with three nuisance variables using a Greco–Latin (Graeco–Latin) square, in which two orthogonal Latin squares (Winer *et al.*, 1991, p. 674) are used; one is given Greek letters, the other Roman (Latin) letters, and the two are superimposed (Winer *et al.*, 1991, pp. 680-1). This principle can be extended to four or more nuisance variables. It's also possible to use Latin squares to extract **partial information from confounded factorial designs** (Winer *et al.*, 1991, p. 682). Latin squares are a special case of **fractional factorial designs** (Winer *et al.*, 1991, pp. 585, 683), in which not all the treatment conditions of a factorial design are examined (see also GLM notes about fractional factorial designs, p. 88→).

Model

An **additive** model assumes that main effects are additive, and don't interact — i.e. that the A and C do not interact. The model is:

$$Y_{ijk} = \mu + \eta_i + \alpha_j + \gamma_k + \varepsilon_{ijk}$$

where μ is the grand mean, η_i is the effect of row i (in this example, subject S_i), α_j is the

effect of treatment A_j, and γ_k is the effect of column k (in this example, day k).

Sources of variance For this additive model, $SS_{total} = SS_{row} + SS_{column} + SS_A + SS_{error}$

ANOVA table Since the number of rows, columns, and treatments is the same,

Source	d.f.	SS	*F*
Row (subject)	a–1	SS_R	MS_R / MS_{error}
Column	a–1	SS_C	MS_C / MS_{error}
A	a–1	SS_A	MS_A / MS_{error}
Error	(a–1)(a–2)	SS_{error}	
Total	N–1 = a^2–1	SS_{total}	

SPSS technique Data layout:

S	*C*	*A*	*depvar*
1	*1*	*1*	*datum*
1	*2*	*2*	*datum*
1	*3*	*4*	*datum*
1	*4*	*3*	*datum*
1	*5*	*5*	*datum*
2	*1*	*3*	*datum*
2	*2*	*1*	*datum*
2	*3*	*5*	*datum*
2	*4*	*2*	*datum*
2	*5*	*4*	*datum*
…			

Syntax:

```
UNIANOVA
  depvar BY s c a
  /RANDOM = s
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = s c a .
```

If C is a random factor, simply add it to the `/RANDOM` list. In general, substitute **R**ow for **S**ubject for any suitable Latin square.

Missing values If we assume the additive model, then it's possible to estimate missing scores (Myers & Well, 1995, p. 352) to allow analysis. Of course, our error *df* are reduced when we do that.

Nonadditive model If the additivity assumption (above) isn't realistic, you can use a nonadditive model. The full model adds in the S × C, S × A, A × C, and S × C × A terms. However, it is what we might call very complex indeed (Myers & Well, 1995, pp. 352-356); I certainly don't understand it.

7.19.3. A × B interactions in a single Latin square

Example

We assign not only an A treatment but also a B treatment to each cell of the Latin square. This can be analysed **provided** that all possible AB combinations appear exactly once in each row and column. For example (Myers & Well, 1995, pp. 356-7):

	C_1	**C_2**	**C_3**	**C_4**
S_1	A_1B_2	A_2B_1	A_1B_1	A_2B_2
S_2	A_2B_1	A_2B_2	A_1B_2	A_1B_1
S_3	A_2B_2	A_1B_1	A_2B_1	A_1B_2
S_4	A_1B_1	A_1B_2	A_2B_2	A_2B_1

Notes

Model

$$Y_{ijkm} = \mu + \eta_i + \alpha_j + \beta_k + \alpha\beta_{jk} + \gamma_m + \varepsilon_{ijkm}$$

where μ is the grand mean, η_i is the effect of row *i* (in this example, subject S_i), α_j is the effect of treatment A_j, β_k is the effect of treatment B_k, $\alpha\beta_{jk}$ is the A × B interaction, and γ_m is the effect of column *m*.

Sources of variance

$SS_{total} = SS_{row} + SS_{column} + SS_A + SS_B + SS_{AB} + SS_{error}$

ANOVA table

Since the number of rows, columns, and AB conditions is the same,

Source	d.f.	SS	*F*
Row (subject)	*ab*–1	SS_R	MS_R / MS_{error}
Column	*ab*–1	SS_C	MS_C / MS_{error}
A	*a*–1	SS_A	MS_A / MS_{error}
B	*b*–1	SS_B	MS_B / MS_{error}
A × B	(*a*–1)(*b*–1)	SS_{AB}	MS_{AB} / MS_{error}
Error	(*a*–1)(*a*–2)	SS_{error}	
Total	N–1 = $(ab)^2$–1	SS_{total}	

SPSS technique

Data layout:

S	*C*	*A*	*B*	*depvar*
1	*1*	*1*	*2*	*datum*
1	*2*	*2*	*1*	*datum*
1	*3*	*1*	*1*	*datum*
1	*4*	*2*	*2*	*datum*
2	*1*	*2*	*1*	*datum*
2	*2*	*2*	*2*	*datum*
2	*3*	*1*	*2*	*datum*
2	*4*	*1*	*1*	*datum*
…				

Syntax:

```
UNIANOVA
  depvar BY s c a b
  /RANDOM = s
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = s c a b a*b .
```

If C is a random factor, simply add it to the `/RANDOM` list.

7.19.4. More subjects than rows: (a) using several squares

Example

In the first example above, we had five treatments and were therefore limited to five rows (subjects). If we want to run more subjects, which will increase power, one way is to use several different squares. This approach has an advantage: if there are interactions with order, using several different squares increases the chance that positive and negative interaction effects will cancel each other. Suppose (Myers & Well, 1995, p. 357) we have 12 subjects being tested on four tasks (A_1–A_4) requiring different types of motor skill. Each task is performed on a different day (C). Three 4×4 Latin squares are constructed (see Myers & Well, 1995, pp. 346-348), and subjects are assigned at random to the 12 rows. The design looks like this:

Square

		C₁	**C₂**	**C₃**	**C₄**
	S₁	A_1	A_3	A_4	A_2
Q₁	**S₂**	A_3	A_4	A_2	A_1
	S₃	A_4	A_2	A_1	A_3
	S₄	A_2	A_1	A_3	A_4

		C₁	**C₂**	**C₃**	**C₄**
	S₅	A_2	A_1	A_3	A_4
Q₂	**S₆**	A_4	A_3	A_1	A_2
	S₇	A_3	A_2	A_4	A_1
	S₈	A_1	A_4	A_2	A_3

		C₁	**C₂**	**C₃**	**C₄**
	S₉	A_2	A_1	A_4	A_3
Q₃	**S₁₀**	A_1	A_2	A_3	A_4
	S₁₁	A_4	A_3	A_1	A_2
	S₁₂	A_3	A_4	A_2	A_1

Notes

Subjects (S) are nested within squares (Q). We assume that S and Q are random factors, while A and C are fixed.

Model

Either this model:

$$Y_{ijkm} = \mu + \eta_{i/m} + \alpha_j + \gamma_k + \pi_m + \alpha\pi_{jm} + \gamma\pi_{km} + \varepsilon_{ijkm}$$

where μ is the grand mean, $\eta_{i/m}$ is the effect of subject *i* (within square *m*), α_j is the effect of A_j, γ_k is the effect of column *k*, $\alpha\pi_{jm}$ allows for the possibility that treatment effects depend on the square (AQ interaction), and $\gamma\pi_{km}$ allows for the possibility that column effects depend on the square (CQ interaction)…

… or, if the full model produces no evidence for AQ or CQ interactions, this reduced model, which pools the AQ and CQ terms into the error term to increase power:

$$Y_{ijkm} = \mu + \eta_{i/m} + \alpha_j + \gamma_k + \pi_m + \varepsilon_{ijkm}$$

Sources of variance

Either this (for the first model):

$$SS_{total} = SS_{S/Q} + SS_A + SS_C + SS_Q + SS_{AQ} + SS_{CQ} + SS_{error}$$

or this (for the reduced model):

$$SS_{total} = SS_{S/Q} + SS_A + SS_C + SS_Q + SS_{error}$$

ANOVA table

For the full model:

Source	d.f.	SS	*F*
Squares (Q)	q–1	SS_Q	$MS_Q / MS_{S/Q}$
S / Q	$q(a$–1)	$SS_{S/Q}$	$MS_{S/Q} / MS_{error}$
C	a–1	SS_C	MS_C / MS_{error}
A	a–1	SS_A	MS_A / MS_{error}

C × Q	$(a-1)(q-1)$	SS_{CQ}	MS_{CQ} / MS_{error}
A × Q	$(a-1)(q-1)$	SS_{AQ}	MS_{AQ} / MS_{error}
Error	$q(a-1)(a-2)$	SS_{error}	
Total	$N-1 = qa^2-1$	SS_{total}	

For the reduced model:

Source	d.f.	SS	*F*
Squares (Q)	$q-1$	SS_{Q}	$MS_{Q} / MS_{S/Q}$
S / Q	$q(a-1)$	$SS_{S/Q}$	$MS_{S/Q} / MS_{error}$
C	$a-1$	SS_{C}	MS_{C} / MS_{error}
A	$a-1$	SS_{A}	MS_{A} / MS_{error}
Error	$(qa-2)(a-1)$	SS_{error}	
Total	$N-1 = qa^2-1$	SS_{total}	

SPSS technique

Data layout:

Q	*S*	*C*	*A*	*depvar*
1	*1*	*1*	*1*	*datum*
1	*1*	*2*	*3*	*datum*
1	*1*	*3*	*4*	*datum*
1	*1*	*4*	*2*	*datum*
1	*2*	*1*	*3*	*datum*
1	*2*	*2*	*4*	*datum*
1	*2*	*3*	*2*	*datum*
1	*2*	*4*	*1*	*datum*
…				
2	*5*	*1*	*2*	*datum*
2	*5*	*2*	*1*	*datum*
2	*5*	*3*	*3*	*datum*
2	*5*	*4*	*4*	*datum*
…				

Syntax for the full model:

```
UNIANOVA
  depvar BY q c a s
  /RANDOM = s q
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = q s(q) c a c*q a*q .
```

For the reduced model:

```
UNIANOVA
  depvar BY q c a s
  /RANDOM = s q
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = q s(q) c a .
```

7.19.5. More subjects than rows: (b) using the same square several times (replicating a single Latin square)

Example

As above, but now you decide to use a single 4 × 4 square and assign *n* subjects to each row of the square. If *n* = 3, your design might look like this:

Subjects	**C1**	**C2**	**C3**	**C4**
S1, S2, S3	A2	A4	A3	A1
S4, S5, S6	A1	A3	A2	A4
S7, S8, S9	A4	A2	A1	A3
S10, S11, S12	A3	A1	A4	A2

Notes

See Myers & Well (1995, pp. 364-368, 374-375), who point out that this design is **frequently used but frequently analysed improperly.**

Should you use replicated squares, or several squares (as on p. 174)? Myers & Well (1995, p. 371) suggest that *several squares is better* — experimenters tend to replicate squares purely for simplicity. Anyway, let's see how you analyse replicated squares now.

Model

This is the simple way:

$$Y_{ijkm} = \mu + \pi_m + \eta_{i/m} + \alpha_j + \gamma_k + \varepsilon_{ijkm}$$

where μ is the grand mean, π_m is the effect of row *m*, $\eta_{i/m}$ is the effect of subject *i* (within row *m*), α_j is the effect of A_j, and γ_k is the effect of column *k*.

Sources of variance

That would give these sources of variance:

$SS_{total} = SS_{between\text{-}subjects} + SS_{within\text{-}subjects}$
$SS_{between\text{-}subjects} = SS_{row} + SS_{subjects\text{-}within\text{-}row(S/R)}$
$SS_{within\text{-}subjects} = SS_A + SS_C + SS_{A\times S/R}$

Complicated bit

However, there are some extra finesses: we can partition the data another way.

- There are a^2 cell means in the Latin square. If you account for main effects of A, C, and R, you're left with what's called the *between-cells error* or residual. It has $(a^2–1) – (a–1) – (a–1) – (a–1) = (a–1)(a–2)$ *df.*
- Then you have the *within-cells residual*, which is equivalent to S × A (nested within R), or S × C (nested within R), which are the same thing (since within one row, a subject's level of A completely determines its level of C). This has $a(n–1)(a–1)$ *df.*

Now…

- Variation among row means (SS_R) reflects different effects of A × C combinations. In other words, if there is an A × C interaction, part of its effect will be reflected in MS_R.
- Part of any A × C interaction effect will also be reflected in what's left in the cell mean variability after you've accounted for main effects of A, C, and R — the between-cells error (MS_{bce}). So any A × C interaction would contribute to MS_{bce}.
- So both MS_R and MS_{bce} partially reflect effects of A × C.

This picture would give this model:

$$Y_{ijkm} = \mu + \eta_{i/m} + \alpha_j + \gamma_k + \alpha\gamma_{jk} + \varepsilon_{ijkm}$$

and this partitioning:

$SS_{total} = SS_{between\text{-}subjects} + SS_{within\text{-}subjects}$
$SS_{between\text{-}subjects} = SS_{subjects\text{-}within\text{-}row(S/R)} + \text{some-part-of-}SS_{AC}$
$SS_{within\text{-}subjects} = SS_A + SS_C + \text{some-part-of-}SS_{AC} + SS_{A\times S/R}$

ANOVA table

This is for the simple way of doing things:

Source	d.f.	SS	*F*
R	a–1	SS_R	$MS_Q / MS_{S/R}$
S / R	$n(a$–1)	$SS_{S/R}$	$MS_{S/R} / MS_{error}$
C	a–1	SS_C	MS_C / MS_{error}
A	a–1	SS_A	MS_A / MS_{error}
Error	(a–1)(an–2)	SS_{error}	
Total	N–1 = na^2–1	SS_{total}	

This is for the complex way:

Source	d.f.	SS	*F*
R (AC′)	a–1	SS_R	$MS_Q / MS_{S/R}$
S / R	$n(a$–1)	$SS_{S/R}$	$MS_{S/R} / MS_{wce}$
C	a–1	SS_C	MS_C / MS_{wce}
A	a–1	SS_A	MS_A / MS_{wce}
Between-cells error (AC′)	(a–1)(a–2)	SS_{bce}	MS_{bce} / MS_{wce}
Within-cells error = S×A / R = S×C / R	$a(n$–1)(a–1)	SS_{wce}	
Total	N–1 = na^2–1	SS_{total}	

The rows labelled AC′ give estimates for the effect of the AC interaction, based on partial information. The between-cells error SS_{bce} is calculated as $SS_{RC} - SS_A$ (that is, calculate the row × column interaction and subtract SS_A).

SPSS technique

Data layout:

R	*S*	*C*	*A*	*depvar*
1	*1*	*1*	*2*	*datum*
1	*1*	*2*	*3*	*datum*
1	*1*	*3*	*4*	*datum*
1	*1*	*4*	*1*	*datum*
1	*2*	*1*	*2*	*datum*
1	*2*	*2*	*3*	*datum*
1	*2*	*3*	*4*	*datum*
1	*2*	*4*	*1*	*datum*
…				
2	*5*	*1*	*2*	*datum*
2	*5*	*2*	*1*	*datum*
2	*5*	*3*	*3*	*datum*
2	*5*	*4*	*4*	*datum*
…				

… simple but less powerful way

Run this:

```
UNIANOVA
  depvar BY r c a s
  /RANDOM = s
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = r s(r) c a .
```

That's it.

… complex and more powerful way

This is pretty complicated. **First,** run this to get the R × C interaction sum of squares (all the sums of squares are correct, but this is the only one you need).

```
UNIANOVA
  depvar BY r c s
  /RANDOM = s
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
```

```
/DESIGN = r s(r) c r*c .
```

Then, run this to get everything else. (This gives you correct answers for all sums of squares, *df*s, and MSs. But you can improve on the *F* ratios by using a different error term…)

```
UNIANOVA
  depvar BY r c a s
  /RANDOM = s
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = r s(r) c a .
```

Next, calculate

$SS_{\text{between-cells-error}} = SS_{R\times C} - SS_A$
$SS_{\text{within-cells-error}} = SS_{\text{error-from-second-ANOVA-in-which-A-was-included}} - SS_{\text{between-cells-error}}$

Calculate the corresponding MS by hand. The *df* for these error terms (which you need to work out the MS) are in the ANOVA table above.

Finally, test MS_C and MS_A against $MS_{\text{within-cells-error}}$ by hand.

If you want, you can also test MS_R (against $MS_{S/R}$) and $MS_{\text{between-cells-error}}$ (against $MS_{\text{within-cells-error}}$) as estimates of the effect of the A × C interaction, based on partial information.

Complex caveat

If C is a random, rather than a fixed factor (Myers & Well, 1995, pp. 366-7), things become more complicated, since C should be tested against MS_{wce} but A must be tested against MS_{bce}, but this has poor power; Myers & Well recommend that if the effect of MS_{bce} isn't significant itself that you use MS_{wce} or the pooled MS_{error} to test A and C.

7.19.6. Between-subjects designs using Latin squares (fractional factorial designs)

Example

Suppose (Winer *et al.*, 1991, p. 687; Myers & Well, 1995, p. 372) we want to compare the effects of three teaching methods (A_1–A_3). To increase the power, we decide to block subjects on the basis of previous experience with the subject (R) and on the basis of ability as measured by a pretest (C). For this full-factorial design, we would need $3 \times 3 \times 3 = 27$ cells with *n* subjects in each. Instead, we reduce the labour by using a Latin-square design with only 9 cells: R would be the rows, and C the columns. The design might look like this, with *n* subjects per cell:

	C_1	**C_2**	**C_3**
R_1	A_2	A_3	A_1
R_2	A_3	A_1	A_2
R_3	A_1	A_2	A_3

Notes

This is very similar to the usual agricultural use of Latin squares.

See also Winer (1991, p. 687-691).

Model

If it assumed that there are no interactions between R, C, and A:

$$Y_{ijkm} = \mu + \alpha_j + \beta_k + \gamma_m + \varepsilon_{ijkm}$$

where μ is the grand mean, α_j is the effect of treatment A_j, β_k is the effect of treatment R_k, and γ_m is the effect of column *m*.

Sources of variance

$SS_{total} = SS_A + SS_R + SS_C + SS_{between\text{-}cell\text{-}error} + SS_{within\text{-}cell\text{-}residual}$

where $SS_{between\text{-}cell\text{-}error}$ includes all sources of variation due to treatment effects which are not predictable from the sum of main effects (e.g. interactions… which you hope aren't there; see below).

ANOVA table

Source	d.f.	SS	*F*
R	a–1	SS_R	MS_R / MS_{wce}
C	a–1	SS_C	MS_C / MS_{wce}
A	a–1	SS_A	MS_A / MS_{wce}
Between-cells error	(a–1)(a–2)	SS_{bce}	MS_{bce} / MS_{wce}
Within-cells error = S / ABC	$a^2(n–1)$	SS_{wce}	
Total	N–1 = na^2–1	SS_{total}	

The between-cells error SS_{bce} is calculated as $SS_{RC} - SS_A$ (that is, calculate the row × column interaction and subtract SS_A).

Caveat

This model is appropriate if additivity can be assumed (if there are no interactions between R, C, and A). And if so, $SS_{between\text{-}cell\text{-}error}$ will *not* be substantially larger than SS_{wce}. One way to test this (Winer *et al.*, 1991, p. 687-8) is to look at the *F* test on MS_{bce}. If it's significant, then the assumptions behind the model are not appropriate, and if this is not an appropriate model — if there are interaction effects — then it's very hard to analyse the data (Winer *et al.*, 1991, p. 690; Myers & Well, 1995, p. 373).

SPSS technique

Data layout (with an unnecessary Subject column to make things clearer):

R	*C*	*A*	*S*	*depvar*
1	*1*	*2*	*1*	*datum*
1	*1*	*2*	*2*	*datum*
1	*1*	*2*	*3*	*datum*
1	*1*	*2*	*4*	*datum*
1	*2*	*3*	*5*	*datum*
1	*2*	*3*	*6*	*datum*
1	*2*	*3*	*7*	*datum*

```
1   2   3   8      datum
…
2   1   3   13     datum
2   1   3   14     datum
2   1   3   15     datum
…
```

Run this:

```
UNIANOVA
  depvar BY r c a
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = r c a .
```

That gives you SS_R, SS_C, SS_A, and SS_{total}. But to get SS_{bce} and SS_{wce}, you have to run this to obtain SS_{RC}:

```
UNIANOVA
  depvar BY r c
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = r c r*c .
```

Then calculate $SS_{bce} = SS_{RC} - SS_A$ and $SS_{wce} = SS_{\text{error-from-first-ANOVA-including-A-factor}} - SS_{bce}$ by hand and complete the ANOVA table.

7.19.7. Several-squares design with a between-subjects factor

Example

We saw how to use a design with several Latin squares above (p. 174). We had a within-subjects factor A. Let's add a between-subjects factor B with *b* levels. We have *q* squares per level of B, and *a* subjects in each squares with *a* scores for each subject. If $a = 4$, $b = 2$ and $q = 2$, we might have this:

Square

		C_1	C_2	C_3	C_4
	S_1	A_4	A_1	A_3	A_2
B_1 Q_1	S_2	A_3	A_2	A_4	A_1
	S_3	A_1	A_3	A_2	A_4
	S_4	A_2	A_4	A_1	A_3

		C_1	C_2	C_3	C_4
	S_5	A_4	A_3	A_2	A_1
B_1 Q_2	S_6	A_2	A_1	A_4	A_3
	S_7	A_1	A_4	A_3	A_2
	S_8	A_3	A_2	A_1	A_4

		C_1	C_2	C_3	C_4
	S_9	A_1	A_4	A_3	A_2
B_2 Q_3	S_{10}	A_3	A_2	A_4	A_1
	S_{11}	A_4	A_1	A_2	A_3
	S_{12}	A_2	A_3	A_1	A_4

		C_1	C_2	C_3	C_4
	S_{13}	A_2	A_1	A_3	A_4
B_2 Q_4	S_{14}	A_1	A_2	A_4	A_3
	S_{15}	A_4	A_3	A_2	A_1
	S_{16}	A_3	A_4	A_1	A_2

This example based on Myers & Well (1995, p. 361), though their original has several numerical errors in their fourth square, which isn't even Latin (some A values appear twice in a column).

Notes

Model

$$Y_{ijkm} = \mu + \beta_k + \pi_{p/k} + \eta_{i/p/k} + \alpha_j + \gamma_m + \alpha\beta_{jk} + \beta\gamma_{km} + \alpha\pi_{jp/k} + \gamma\pi_{mp/k} + \varepsilon_{ijkmp}$$

where *i* index subjects (within squares within levels of B), *j* indexes the level of A, *k* indexes the level of B, *m* indexes the level of C, and *p* indexes the square (within a level of B).

Subject and Square are assumed to be random; A, B, and C are assumed to be fixed effects.

Sources of variance

I think it's this (based on Myers & Well, 1995, p. 362):

$SS_{total} = SS_{between\text{-}squares} + SS_{within\text{-}squares}$
- $SS_{between\text{-}squares} = SS_B + SS_{Q/B}$
- $SS_{within\text{-}squares} = SS_{between\text{-}subjects\text{-}within\text{-}squares} + SS_{within\text{-}subjects}$
  - $SS_{between\text{-}subjects\text{-}within\text{-}squares} = SS_{S/Q/B}$
  - $SS_{within\text{-}subjects} = SS_A + SS_C + SS_{AB} + SS_{BC} + SS_{AQ/B} + SS_{CQ/B} + SS_{within\text{-}subject\text{-}error}$

We could also note that $SS_{between\text{-}subjects} = SS_B + SS_{Q/B} + SS_{S/Q/B}$ (Myers & Well, 1995, p. 363).

But if $p > .25$ for the interaction terms AQ/B and CQ/B, it would be reasonable to pool those error terms:

$$SS_{within\text{-}subjects} = SS_A + SS_C + SS_{AB} + SS_{BC} + SS_{pooled\text{-}within\text{-}subject\text{-}error}$$

ANOVA table

For the full model (note that Myers & Well, 1995 cock the df right up):

Source	d.f.	SS	*F*
B	b–1	SS_B	$MS_B / MS_{Q/B}$
Q / B	$b(q$–1)	$SS_{Q/B}$	$MS_{Q/B} / MS_{S/Q/B}$
S / Q / B	$bq(a$–1)	$SS_{S/Q/B}$	$MS_{S/Q/B} / MS_{error}$
A	a–1	SS_A	$MS_A / MS_{AQ/B}$
C	a–1	SS_C	$MS_C / MS_{CQ/B}$
AB	(a–1)(b–1)	SS_{AB}	$MS_{AB} / MS_{AQ/B}$
BC	(b–1)(a–1)	SS_{BC}	$MS_{BC} / MS_{CQ/B}$
AQ / B	$b(a$–1)(q–1)	$SS_{AQ/B}$	$MS_{AQ/B} / MS_{error}$
CQ / B	$b(a$–1)(q–1)	$SS_{CQ/B}$	$MS_{CQ/B} / MS_{error}$
Error	$bq(a$–1)(a–2)	SS_{error}	
Total	bqa^2–1	SS_{total}	

For the pooled error model:

Source	d.f.	SS	*F*
B	b–1	SS_B	$MS_B / MS_{Q/B}$
Q / B	$b(q$–1)	$SS_{Q/B}$	$MS_{Q/B} / MS_{S/Q/B}$
S / Q / B	$bq(a$–1)	$SS_{S/Q/B}$	$MS_{S/Q/B} / MS_{error}$
A	a–1	SS_A	MS_A / MS_{error}
C	a–1	SS_C	MS_C / MS_{error}
AB	(a–1)(b–1)	SS_{AB}	MS_{AB} / MS_{error}
BC	(b–1)(a–1)	SS_{BC}	MS_{BC} / MS_{error}
Error (pooled)	$b(a$–1)(aq–2)	SS_{error}	
Total	bqa^2–1	SS_{total}	

SPSS technique

Since Myers & Well's (1995, pp. 361-3) numerical example is wrong, I have no way of verifying this against some gold standard.

Data format:

B	*Q*	*S*	*C*	*A*	*depvar*
1	*1*	*1*	*1*	*4*	*datum*
1	*1*	*1*	*2*	*1*	*datum*
1	*1*	*1*	*3*	*3*	*datum*
1	*1*	*1*	*4*	*2*	*datum*
1	*1*	*2*	*1*	*3*	*datum*
1	*1*	*2*	*2*	*2*	*datum*
1	*1*	*2*	*3*	*4*	*datum*
1	*1*	*2*	*4*	*1*	*datum*
…					
1	*2*	*5*	*1*	*4*	*datum*
1	*2*	*5*	*2*	*3*	*datum*
1	*2*	*5*	*3*	*2*	*datum*
1	*2*	*5*	*4*	*1*	*datum*
…					
2	*3*	*9*	*1*	*1*	*datum*
2	*3*	*9*	*2*	*4*	*datum*
2	*3*	*9*	*3*	*3*	*datum*
2	*3*	*9*	*4*	*2*	*datum*
…					

Full model syntax:

```
UNIANOVA
  depvar BY b q s c a
  /RANDOM = q s
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = b q(b) s(q(b)) a c a*b b*c a*q(b) c*q(b) .
```

Pooled error model syntax, I presume, is this:

```
UNIANOVA
  depvar BY b q s c a
  /RANDOM = q s
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = b q(b) s(q(b)) a c a*b b*c .
```

7.19.8. Replicated-squares design with a between-subjects factor

Example

We saw how to use a design with a replicated Latin square above (p. 176). We had a within-subjects factor A. Let's add a between-subjects factor B with *b* levels. We have one Latin square with *a* rows, with *n* subjects for each row and therefore *bn* subjects per level of B. If *a* = 4, *b* = 2 and *n* = 2, we might have this:

B_1	**B_2**				
Subjects	**Subjects**	**C_1**	**C_2**	**C_3**	**C_4**
S_1, S_2	**S_9, S_{10}**	A_4	A_2	A_1	A_3
S_3, S_4	**S_{11}, S_{14}**	A_3	A_1	A_4	A_2
S_5, S_6	**S_{13}, S_{15}**	A_2	A_4	A_3	A_1
S_7, S_8	**S_{15}, S_{16}**	A_1	A_3	A_2	A_4

Notes

See Myers & Well (1995, pp. 368-370, 374-375), who point out that this design is **frequently used but frequently analysed improperly.**

Model

$$Y_{ijkm} = \mu + \eta_{i/kp} + \alpha_j + \beta_k + \gamma_m + \alpha\beta_{jk} + \alpha\gamma_{jm} + \beta\gamma_{km} + \alpha\beta\gamma_{jkm}\varepsilon_{ijkmp}$$

where *i* indexes the subject (within a row × B combination; *i* = 1…*n*), *j* indexes the level of A (*j* = 1…*a*), *m* indexes the level of C (*m* = 1…*a*), and *p* indexes the row within the square (*p* = 1…*a*).

Subject is assumed to be a random factor; the rest are fixed.

Sources of variance

$SS_{total} = SS_{between\text{-}subjects} + SS_{within\text{-}subjects}$
$SS_{between\text{-}subjects} = SS_B + SS_R + SS_{BR} + SS_{S/BR}$
$SS_{within\text{-}subjects} = SS_C + SS_A + SS_{AB} + SS_{BC} + SS_{between\text{-}cell\text{-}residual} + SS_{B\times between\text{-}cell\text{-}residual} + SS_{within\text{-}cell\text{-}residual}$

where $SS_{between\text{-}cell\text{-}residual} = SS_{CR} - SS_A$
and $SS_{B\times between\text{-}cell\text{-}residual} = SS_{BCR} - SS_{AB}$.

ANOVA table

Source	d.f.	SS	*F*
B	*b*–1	SS_B	$MS_B / MS_{S/BR}$
R (AC′)	*a*–1	SS_R	$MS_R / MS_{S/BR}$
BR (ABC′)	(*b*–1)(*a*–1)	SS_{BR}	$MS_{BR} / MS_{S/BR}$
S / BR	*ab*(*n*–1)	$SS_{S/BR}$	$MS_{S/BR} / MS_{wce}$
C	*a*–1	SS_C	MS_C / MS_{wce}
A	*a*–1	SS_A	MS_A / MS_{wce}
AB	(*a*–1)(*b*–1)	SS_{AB}	MS_{AB} / MS_{wce}
BC	(*a*–1)(*b*–1)	SS_{BC}	MS_{BC} / MS_{wce}
Between-cells error (AC′)	(*a*–1)(*a*–2)	SS_{bce}	MS_{bce} / MS_{wce}
B × betw.-cells error (ABC′)	(*a*–1)(*a*–2)(*b*–1)	$SS_{B\times bce}$	$MS_{B\times bce} / MS_{wce}$
Within-cells error = S×A / BR = S×C / BR	*ab*(*a*–1)(*n*–1)	SS_{wce}	
Total	N–1 = bna^2–1	SS_{total}	

As before, some terms give estimates of interactions based on partial information; they're labelled with a prime (′) symbol above. Again, there's a *df* error in Myers & Well (1995, p. 370).

SPSS technique

Data layout:

B	*R*	*S*	*C*	*A*	*depvar*
1	*1*	*1*	*1*	*4*	*datum*
1	*1*	*1*	*2*	*2*	*datum*
1	*1*	*1*	*3*	*1*	*datum*
1	*1*	*1*	*4*	*3*	*datum*

```
1  1  2  1  4  datum
1  1  2  2  2  datum
1  1  2  3  1  datum
1  1  2  4  3  datum
1  2  3  1  3  datum
1  2  3  2  1  datum
1  2  3  3  4  datum
1  2  3  4  2  datum
…
2  1  9  1  4  datum
2  1  9  2  2  datum
2  1  9  3  1  datum
2  1  9  4  3  datum
…
```

SPSS syntax:

```
UNIANOVA
  depvar BY b r c a s
  /RANDOM = s
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = b r b*r s(b*r)
            c a a*b b*c .
```

That'll give you all the SS except SS_{bce}, $SS_{B\times bce}$, and SS_{wce}. To get those, obtain SS_{CR} and SS_{BCR} from this syntax:

```
UNIANOVA
  depvar BY b r c s
  /RANDOM = s
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = b r b*r s(b*r)
            c c*b c*r b*c*r .
```

and calculate

$$SS_{bce} = SS_{CR} - SS_{A}$$

and

$$SS_{B\times bce} = SS_{BCR} - SS_{AB}$$

Finally, calculate

$$SS_{wce} = SS_{\text{error-from-first-ANOVA-including-A}} - SS_{bce} - SS_{B\times bce}$$

and use it to test the relevant terms by hand.

7.20 Agricultural terminology and designs, compared to psychology

In psychology, the most important factor in experimental design is often the *subject*, because this accounts for much correlation between observations. If you have two groups of subjects and give the two groups the two treatments, you account for much (you hope all) of the expected correlation between any two subjects by specifying the 'treatment' factor in your analysis. (Of course, that may not be the case — if one group were all men and the other all women, you'd have confounded sex and treatment; another way of saying that is that correlations between individual subjects' scores may be due to them being members of the same sex rather than having experienced a particular treatment.) On the other hand, if you measure subjects more than once, you can expect high correlations between observations from the same subject — much more so than between observations from different subjects. So you need to account for intra-subject correlation, which you do by specifying a Subject factor (by performing a within-subjects analysis). Much psychological research boils down to asking 'is this a between-subjects or a within-subjects factor?'

However, many ANOVA techniques originated in agricultural research, so it often happens that when you want an example of an advanced design, the only ones you find are agricultural. And in agriculture, sources of correlation don't come from 'subjects', but from things like geographical proximity. If you want to see whether fertilizer A works better than fertilizer B, you'd want to give fertilizer A to a set of plants (obviously not just one) and fertilizer B to another set of plants. But it would be pretty daft to spray fertilizer A on the sunny south-facing side of your field and to fertilizer B under the shady oak tree. Agricultural designs and analyses revolve around these sorts of ideas.

This overview of agricultural teminology is principally from Tangren (2002).

Completely randomized design (CRD)

Your smallest experimental unit (sometimes called the 'subject' or 'replication') is a small plot of land with a plant or plants in it. Each experimental unit produces a single value of the dependent variable.

You have four fertilizers (A–D; factor T for treatment; $t = 4$). You give each to four experimental units ('subjects') ($n = 4$ per group) *at random*. Adjacent subjects could potentially have the same treatment. Here's one possible layout, where A–D are treatments and 1–4 are subjects within each treatment (a single 'subject' is underlined):

```
A1  B1  C1  A2
D1  A3  D2  C2
B2  D3  C3  B3
C4  A4  B4  D4
```

The appropriate ANOVA is equivalent to a design with **one between-subjects factor** (p. 106). If t is the number of treatments and r is the number of replications per treatment:

Source	*df*	SS	F
T	t–1	SS_T	MS_T / MS_{error}
error	$t(r$–1)	SS_{error}	
Total	tr–1	SS_{total}	

CRD with subsampling

The same as a CRD, except that you take three samples per plant (or small plot of plants, or whatever your previous basic unit was; plant = replication). Treatments are assigned at random to the plants. For example, if the treatments are A–D, the plants (replications) are 1–4 and the subsamples are a–c, we could get this:

```
A1a A1b A1c B2a B2b B2c C3a C3b C3c B4a B4b B4c
B1a B1b B1c A2a A2b A2c C2a C2b C2c A4a A4b A4c
C1a C1b C1c B3a B3b B3c A3a A3b A3c C4a C4b C4c
```

A single plant / plot / whatever is underlined. The idea is that you get a better idea of your measurement error (within-plant variability), so you can remove this to get a better esti-

mate of your between-plant variability. The ANOVA looks like this:

$$SS_{total} = SS_T + SS_{between\text{-}plant\text{-}variability} + SS_{within\text{-}plant\text{-}variability}$$

Source	*df*	SS	F
T	t–1	SS_T	MS_T / MS_E
experimental error E = replication / T	$t(r–1)$	SS_E	MS_E / MS_S
sampling error S = 'error'	$tr(s–1)$	SS_S	
Total	trs–1	SS_{total}	

where r is the number of replications per treatment and s is the number of subsamples per replication. For example, see

www.stat.wisc.edu / ~clayton / stat850 / Handouts / crdwsubsamp.pdf

No routine psychology equivalent? Except that it is a way to analyse situations in which you have **one between-subjects factor** and you have **multiple observations per subject.**

To run this analysis in SPSS, the data can be laid out like this:

T	*Rep*	*depvar*
1	*1*	*subsample_1_datum*
1	*1*	*subsample_2_datum*
1	*1*	*subsample_3_datum*
1	*2*	*subsample_1_datum*
1	*2*	*subsample_2_datum*
1	*2*	*subsample_3_datum*
…		
2	*5*	*subsample_1_datum*
2	*5*	*subsample_2_datum*
2	*5*	*subsample_3_datum*
…		

and analysed using this syntax:

```
UNIANOVA
  depvar BY trt rep
  /RANDOM = rep
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = trt rep(trt) .
```

To achieve this using the SPSS menus, you have to enter a custom model (because you don't want the Replication factor in there as a main effect; you just want replication / T).

You might think this was a good way to analyse designs in which you measure a subject (replication) several times. And indeed, this is a valid way to analyse such data. Except… this design gives identical answers to taking a mean for every subject (replication) and analysing those means by one-way ANOVA using T as the only factor! See p. 48.

Randomized complete block (RCB) design

'The standard design for agricultural experiments' (Tangren, 2002). The orchard is divided into units called *blocks* to account for any variation across the field (sunny bit, shady bit, etc.). Treatments are then assigned at random to the plants in the blocks, one treatment per plant (or small plot of plants). Each block experiences each treatment. If the blocks are I–IV and the treatments are A–D, we might have this:

```
Block I     A   B   C   D
Block II    D   A   B   C
Block III   B   D   C   A
Block IV    C   A   B   D
```

Source	*df*	SS	F
Block B	*b*–1	SS_B	MS_B / MS_E
Treatment T	*t*–1	SS_T	MS_T / MS_E
error E	(*t*–1)(*b*–1)	SS_E	
Total	*tb*–1	SS_{total}	

Equivalent to a design with **one within-subjects factor** (p. 112) (Block ≡ Subject; Treatment ≡ WS factor).

RCB with subsampling

The layout is the same as an RCB, but each plant (or plot) is sampled several times. For example (a single plant — subsampled basic unit — is underlined):

```
Aa Ab Ac  |  Ba Bb Bc  |  Ca Cb Cc  |  Ba Bb Bc
Ba Bb Bc  |  Aa Ab Ac  |  Ba Bb Bc  |  Aa Ab Ac
Ca Cb Cc  |  Ca Cb Cc  |  Aa Ab Ac  |  Ca Cb Cc
-------------------------------------------------
Block I   |  Block II  |  Block III |  Block IV
```

Source	*df*	SS	F
Block B	*b*–1	SS_B	MS_B / MS_E
Treatment T	*t*–1	SS_T	MS_T / MS_E
experimental error E	(*t*–1)(*b*–1)	SS_E	MS_E / MS_S
sample error S	*tb*(*s*–1)	SS_S	
Total	*tb*–1	SS_{total}	

where *b* is the number of blocks, *t* is the number of treatments and *s* is the number of subsamples per plot. For example, see

www.stat.wisc.edu / ~clayton / stat850 / Handouts / crdwsubsamp.pdf

No routine psychology equivalent? Except that it is a way to analyse situations in which you have **one within-subjects factor** (p. 112) (Block ≡ Subject; Treatment ≡ WS factor) and you have **multiple observations per level of the within-subjects factor per subject.**

To run this analysis in SPSS, the data can be laid out like this:

Block	*T*	*depvar*
1	*1*	*subsample_1_datum*
1	*1*	*subsample_2_datum*
1	*1*	*subsample_3_datum*
1	*2*	*subsample_1_datum*
1	*2*	*subsample_2_datum*
1	*2*	*subsample_3_datum*
…		
2	*5*	*subsample_1_datum*
2	*5*	*subsample_2_datum*
2	*5*	*subsample_3_datum*
…		

and analysed using this syntax:

```
UNIANOVA
  depvar BY t block
  /RANDOM = block
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = t block t*block .
```

To achieve this using the SPSS menus, choose Analyze → General Linear Model → Univariate. Enter T as a fixed factor and Block as a random factor.

You might think this was a good way to analyse designs in which you measure a subject (replication) several times at each level of a within-subjects factor. And indeed, this is a

valid way to analyse such data. Except… this design gives identical answers to taking a mean for every subject / factor combination and analysing those means using a straightforward within-subjects design with T as the only factor! Compare p. 48.

Latin square

Used to control for variation in two different directions, the *row* direction and the *column* direction. Each treatment appears once per row and once per column. There are the same number of rows as columns as treatments (call that number *r*). For example:

Column	1	2	3	4
Row I	A	B	C	D
Row II	C	D	A	B
Row III	D	C	B	A
Row IV	B	A	D	C

Source	*df*	SS	F
Row R	r–1	SS_R	MS_R / MS_E
Column C	r–1	SS_C	MS_C / MS_E
Treatment T	r–1	SS_T	MS_T / MS_E
experimental error E	(r–1)(r–2)	SS_E	
Total	r^2–1	SS_{total}	

Directly equivalent to Latin square designs used in psychology (p. 170→).

CRD factorial

Two treatments are combined — for example, fertilizer (of type A or B) and pesticide (of type a or b) are combined to give treatment combinations Aa, Ab, Ba, Bb. Each combination is then randomly assigned to replications, with *r* replications per treatment combination. For example, with a 2 × 2 design and 4 replications (plants, plots, whatvever) per treatment, you might have the following layout (a single plant / plot is underlined):

Aa1	Ba1	Ab1	Aa2
Bb1	Aa3	Bb2	Ab2
Ba2	Bb3	Ab3	Ba3
Ab4	Aa4	Ba4	Bb4

Equivalent to a design with **two between-subjects factors** (p. 108). So the table is obvious:

Source	*df*	SS	F
first factor F	f–1	SS_F	MS_F / MS_E
second factor S	s–1	SS_S	MS_S / MS_E
F × S	(f–1)(s–1)	SS_{FS}	MS_{FS} / MS_E
error E	fs(r–1)	SS_E	
Total	fsr–1	SS_{total}	

RCB factorial

Orchard is divided into blocks. Every block gets all possible combinations of the two factors, as above (assigned at random within each block). For example:

Block IV	Aa	Ba	Ab	Bb
Block III	Bb	Aa	Ba	Ab
Block II	Ba	Bb	Ab	Aa
Block I	Ab	Aa	Ba	Bb

Equivalent to a design with **two within-subjects factors** (p. 115) (Block ≡ Subject; Treatment A and Treatment B are WS factors).

Source	*df*	SS	F
Block B	b–1	SS_B	MS_B / MS_E
first factor F	f–1	SS_F	MS_F / MS_E
second factor S	s–1	SS_S	MS_S / MS_E
F × S	(f–1)(s–1)	SS_{FS}	MS_{FS} / MS_E
error E	(fs–1)(b–1)	SS_E	
Total	fsb–1	SS_{total}	

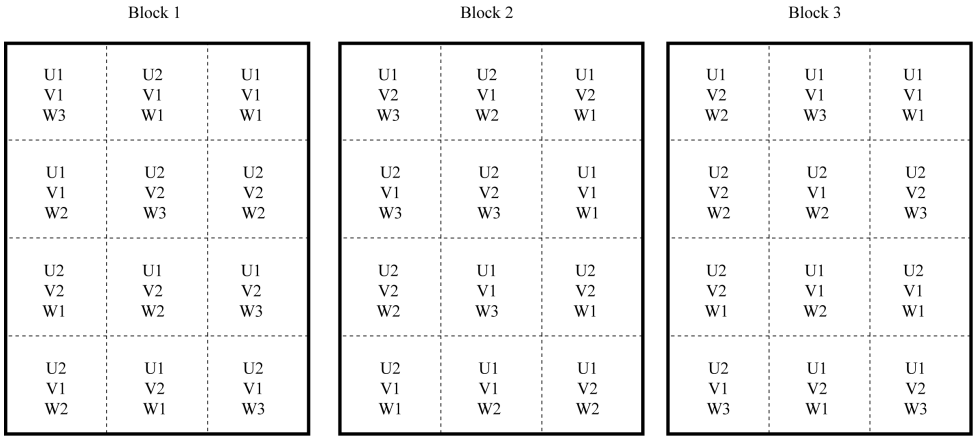
RCB 3-way factorial

Simply an extension of an RCB 2-way factorial (see above) to 3 factors. Therefore equivalent to a design with **three within-subjects factors** (p. 118) (Block ≡ Subject). If our factor levels are A–C (first factor), 1–2 (second factor), a–b (third factor), we might have this:

```
Block I     Block II     Block III
C1a  B2a    C2a  C1a     B2a  B1a
B1a  A2a    C1b  B1a     B2b  A2a
A2b  A1a    B2b  A2a     A2b  A1a
B1b  B2b    A1b  A1a     C1b  C2b
A1b  C2a    B1b  A2b     C1a  A1b
C2b  C1b    C2b  B2a     B1b  C2a
```

Source	*df*	SS	F
Block B	b–1	SS_B	MS_B / MS_E
first factor X	x–1	SS_X	MS_X / MS_E
second factor Y	y–1	SS_Y	MS_Y / MS_E
third factor Z	z–1	SS_Z	MS_Z / MS_E
X × Y	(x–1)(y–1)	SS_{XY}	MS_{XY} / MS_E
X × Z	(x–1)(z–1)	SS_{XZ}	MS_{XZ} / MS_E
Y × Z	(y–1)(z–1)	SS_{YZ}	MS_{YZ} / MS_E
X × Y × Z	(x–1)(y–1)(z–1)	SS_{XYZ}	MS_{XYZ} / MS_E
error E	(xyz–1)(b–1)	SS_E	
Total	$xyzb$–1	SS_{total}	

Here's a picture (partly for comparison to a split–split plot, see below):



Randomized complete block design with three blocks.
Factors are U (2 levels), V (2 levels), W (3 levels).
Every block is treated with all 12 combinations of W, V, and U in full factorial fashion.
The treatments are randomized within the 12 divisions of each block.

Split plot on a CRD

The main experimental units of a CRD (termed *main plots*) are divided further into *subplots* to which another set of treatments are assigned at random. For example, suppose we have pesticides A–C (main treatment), four plots (replications) per treatment (12 plots in total), each divided into three subplots, and three fertilizers a–c (subplot treatment). We could have this:

```
A1a A1b A1c B1c B1b B1a A2b A2c A2a C1a C1c C1b
C2c C2a C2b A3b A3c A3a B2c B2a B2b C3b C3a C3c
B3b B3c B3a A4a A4c A4b C4c C4a C4b B4a B4b B4c
```

One plot (a plot is underlined) only experiences one main treatment, but experiences all three subplot treatments.

Source	*df*	SS	F
plot treatment T	t–1	SS_T	MS_T / MS_{Em}
error, main plots (Em)	t(r–1)	SS_{Em}	

subplot treatment S	s–1	SS_S	MS_S / MS_{Es}
S × T	(t–1)(s–1)	SS_{ST}	MS_{ST} / MS_{Es}
error, subplots (Es)	t(r–1)(s–1)	SS_{Es}	
Total	trs–1	SS_{total}	

Equivalent to a design with **one between-subjects factor and one within-subjects factor** (p. 122) (plot treatment ≡ BS factor, subplot treatment ≡ WS factor; 'plot' ≡ Subject).

Split plot on an RCB

The orchard is divided into units called *blocks* to account for any variation across the field (sunny bit, shady bit, etc.). The blocks are then divided into *plots*. Treatments (e.g. pesticides) are then assigned at random to the plots in the blocks, one treatment per plot. Each block experiences each treatment. The plots are then divided into *subplots* and a further set of treatments (e.g. fertilizer) are applied to the subplots, assigned at random. If the blocks are I–IV, the main plot treatments are A–C, and the subplot treatments are a–c, we might have this:

```
Block-I   Block-II   Block-III   Block-IV
Aa Ab Ac  Bc Bb Ba   Ab Ac Aa    Ca Cc Cb
Cc Ca Cb  Ab Ac Aa   Bc Ba Bb    Aa Ac Ab
Bb Bc Ba  Cb Ca Cc   Cc Ca Cb    Ba Bb Bc
```

A main plot is underlined. The number of blocks is the number of replications.

Source	*df*	SS	F
block B	b–1	SS_B	MS_B / MS_{Em}
plot treatment T	t–1	SS_T	MS_T / MS_{Em}
error, main plots (Em)	(t–1)(b–1)	SS_{Em}	
subplot treatment S	s–1	SS_S	MS_S / MS_{Es}
S × T	(t–1)(s–1)	SS_{ST}	MS_{ST} / MS_{Es}
error, subplots (Es)	t(b–1)(s–1)	SS_{Es}	
Total	tbs–1	SS_{total}	

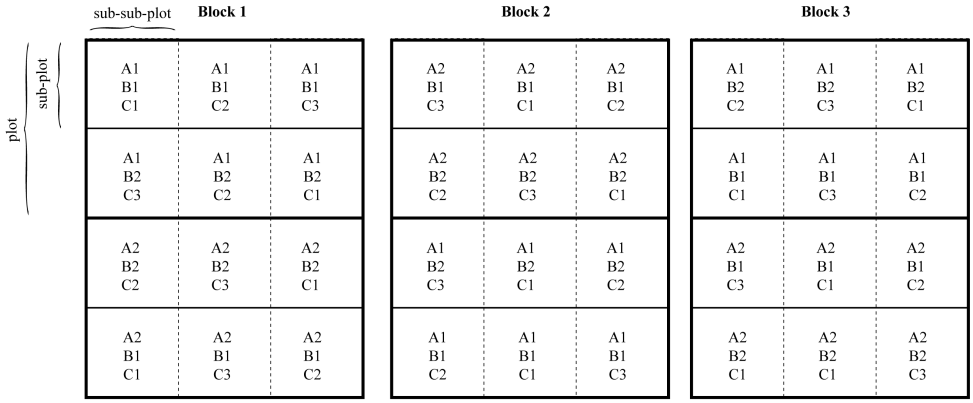
This is a **hierarchical design** (p. 159→). The 'relatedness' factors are Block (plots are related if they come from the same block) and Plot (subplots are related if they come from the same plot).

Split-split plot on an RCB

The orchard is divided into *blocks*. The blocks are then divided into *plots*. Treatments (T, e.g. pesticides) are then assigned at random to the plots in the blocks, one treatment per plot. Each block experiences each treatment. The plots are then divided into *subplots* (or split plots) and a further set of treatments (S, e.g. fertilizer) are assigned at random to the subplots. The subplots are then further subdivided into *split-subplots* (or sub-subplots, or split-split plots) and a third set of treatments (U, e.g. pruning technique) are assigned at random to the split-subplots. If the blocks are I–III, the main plot treatments are A–B, the subplot treatments are 1–2, and the split-subplot treatments are a–c, we might have this:

```
Block I        Treatment A    +    Treatment B
            1a 1b 1c 2c 2b 2a + 2b 2c 2a 1a 1c 1b
Block II       Treatment B    +    Treatment A
            1c 1a 1b 2b 2c 2a + 2c 2a 2b 1b 1a 1c
Block III      Treatment A    +    Treatment B
            2b 2c 2a 1a 1c 1b + 1c 1a 1b 2a 2b 2c
```

Here's a picture:



Split-split plot, randomized complete block design.
The field is split into blocks.
Each block is split into two plots and factor A (2 levels) is assigned at random to the plots.
Each plot is split into two sub-plots and factor B (2 levels) is assigned at random to the sub-plots.
Each sub-plot is split into three sub-sub-plots and factor C (3 levels) is assigned at random to the sub-sub-plots.

Compare this to an RCB 3-way factorial (see above).

Source	*df*	SS	F
Between blocks:			
block B	b–1	SS_B	MS_B / MS_{Em}
Within blocks, between plots:			
plot treatment T	t–1	SS_T	MS_T / MS_{Em}
error, main plots (Em)	(t–1)(b–1)	SS_{Em}	
Within plots, between subplots:			
subplot treatment S	s–1	SS_S	MS_S / MS_{Es}
S × T	(s–1)(t–1)	SS_{ST}	MS_{ST} / MS_{Es}
error, subplots (Es)	t(b–1)(s–1)	SS_{Es}	
Within subplots:			
split-subplot treatment U	u–1	SS_U	MS_U / MS_{Eu}
U × T	(u–1)(t–1)	SS_{UT}	MS_{UT} / MS_{Eu}
U × S	(u–1)(s–1)	SS_{ST}	MS_{ST} / MS_{Eu}
U × S × T	(u–1)(s–1)(t–1)	SS_{UST}	MS_{UST} / MS_{Eu}
error, split-subplots (Eu)	ts(b–1)(u–1)	SS_{Es}	
Total	tbs–1	SS_{total}	

This is a **hierarchical design with *three* levels of 'relatedness'** (p. 159→). They are Block (plots are related if they come from the same block), Plot (subplots are related if they come from the same plot), and Subplot (split-subplots are related if they come from the same subplot). This is one hierarchical level more than the basic split-split plot design (based on a CRD rather than an RCB), discussed above.

Split block

Two sets of treatments are randomized *across* each other in strips in an otherwise RCB design. So the orchard is divided into blocks, and the blocks are divided in an North–South direction and an East–West direction. One treatment (pesticide, A–C) is assigned randomly to the blocks in the North–South direction, so each block experiences all treatments. The other treatment (fertilizer, 1–2) is assigned randomly to the blocks in an East–West direction; again, each block experiences all treatments. It might look like this:

Block I	Block II	Block III
A1 A2	C2 C1	B1 B2
B1 B2	A2 A1	C1 C2
C1 C2	B2 B1	A1 A2

Source	*df*	SS	F
block B	b–1	SS_B	$MS_B / MS_{T\times B}$
treatment T	t–1	SS_T	$MS_T / MS_{T\times B}$

T × B	$(t–1)(b–1)$	$SS_{T\times B}$	
cross-treatment C	$c–1$	SS_C	$MS_C / MS_{C\times B}$
C × B	$(c–1)(b–1)$	$SS_{C\times B}$	
C × T	$(c–1)(t–1)$	$SS_{C\times T}$	$MS_{C\times T} / MS_E$
error, E	$(t–1)(c–1)(b–1)$	SS_E	
Total	$tcb–1$	SS_{total}	

Equivalent to a design with **two within-subjects factors** (Block ≡ Subject). Compare the full model for two within-subjects factors discussed earlier (p. 115).

Pseudoreplication

The researcher applies one treatment to all the trees in a row, the next treatment to all the trees in the next row, etc. There are (say) 4 trees per row.

```
Row I     A  A  A  A
Row II    B  B  B  B
Row III   C  C  C  C
Row IV    D  D  D  D
```

The researcher hoped that the experiment was being replicated by having four trees per row, but the researcher has cocked up. Row is confounded with treatment, so we can't analyse this. (The split-block design is one way of applying treatments to whole rows, properly.)

There are frequent psychology equivalents, but that's not a good thing.

Regression applied to a CRD

The orchard is divided into plots. Each plot has a certain amount of fertilizer applied — treated as a *continuous* variable. Treatments are assigned to plots at random. If the experimenter uses 0, 1, 2, and 5 kg of fertilizer, and has four plots per fertilizer condition (replicates a–d), the orchard might look like this:

```
5a  2a  0a  5b
1a  5c  1b  0b
2b  1c  0c  2c
0d  5d  2d  1d
```

The researcher expects a linear relationship between fertilizer amount and the dependent variable.

Source	*df*	SS	F
regression R	1	SS_R	MS_R / MS_E
error E	$tr–2$	SS_E	
Total	$tr–1$	SS_{total}	

My comment: of course, there is no absolute requirement to have four plots with 0 kg, four plots with 2 kg, and so on; you could have one plot with 0 kg, one with 0.5 kg, one with 1 kg…

Equivalent to simple (i.e. between subjects) linear regression (see p. 135).

Regression, comparing trends from different treatments (applied to a CRD)

The orchard is divided into plots. One treatment factor (pesticide A or B) is crossed with a continuously-measured treatment (fertilizer: 1, 2, 5 kg). There are four plots (replications a–d) per pesticide / fertilizer combination. So we might have this:

```
5Aa   2Aa   5Ba   5Ab   2Ba   1Ba
1Bb   1Aa   2Bb   2Ab   1Bc   1Ab
5Ac   2Bc   2Ac   1Ac   5Bb   5Bc
1Ad   5Bd   5Ad   1Bd   2Bd   2Ad
```

Source	*df*	SS	F
treatment T	$t–1$	SS_T	MS_T / MS_E
regression R	1	SS_R	MS_R / MS_E
T × R	$t–1$	SS_{TR}	MS_{TR} / MS_E
error E	$t(qr–2)$	SS_E	

Total	*tqr*–1	SS_{total}

where *q* is the number of levels of the continuously-measured thing that you're using as a linear predictor (fertilizer, in this example). The T × R interaction measures whether the regression slope differs across treatments.

Equivalent to ANCOVA with one between-subjects covariate and one between-subjects factor, in which the covariate and factor interaction is included (p. 144).

ANCOVA (applied to a CRD)

The orchard is divided into plots, and treatments A–D are applied to the plots at random (this is a CRD). Then an independent factor (e.g. soil nitrogen) will be *measured* for each plot. Suppose there are four replications (1–4; four plots for each level of the treatment). We might then have this layout:

```
A1  B1  C1  A2
D1  A3  D2  C2
B2  D3  C3  B3
C4  A4  B4  D4
```

We'll also measure the covariate (nitrogen) in each plot. This is the ANOVA table:

Source	*df*	SS	F
covariate C	1	SS_C	MS_C / MS_E
adjusted treatment T	*t*–1	SS_T	MS_T / MS_E
error E	*t*(*r*–1)–1	SS_E	
Total	*tr*–1	SS_{total}	

where *r* is the number of replications per treatment. The treatment effect is adjusted for the effects of the covariate.

Equivalent to ANCOVA with one between-subjects covariate and one between-subjects factor, in which the covariate and factor interaction is *not* included (p. 138).

RCB repeated at locations

We have three orchards, widely separated — a *location* factor. We divide each orchard into *blocks*. We assign the levels of our treatment to plots within those blocks, each treatment once per block. (The number of blocks is the number of replications.) For example, if our treatments are A–C, we might have this:

```
Location 1          Location 2
Block I  II  III    Block I  II  III
      A   B   C           B   C   B
      B   A   B           A   B   C
      C   C   A           C   A   A

            Location 3
            Block I  II  III
                  A   C   A
                  C   B   C
                  B   A   B
```

Source	*df*	SS	F
location L	*l*–1	SS_L	MS_L / MS_{El}
error for locations, El	*l*(*b*–1)	SS_{El}	
treatment T	*t*–1	SS_T	MS_T / MS_E
T × L	(*t*–1)(*l*–1)	SS_{TL}	MS_{TL} / MS_E
error E	*l*(*t*–1)(*b*–1)	SS_E	
Total	*ltb*–1	SS_{total}	

Comment: this is again equivalent to a design with **one between-subjects factor and one within-subjects factor** (p. 122) (Block ≡ Subject). Location is the 'between-blocks' factor and Treatment is the 'within-blocks' factor. Therefore, it's also analytically equivalent to the 'split plot on a CRD' design above.

RCB repeated in time

Merely one example of repeating a design in time… The orchard is divided into *blocks*; treatments are assigned at random to plots within those blocks (each treatment once per block, so the number of blocks is the number of replications) and everything is measured three times.

```
t=1        Block I    A  B  C  D  E  F
           Block II   F  A  E  B  D  C
           Block III  C  B  F  A  D  E

t=2        Block I    A  B  C  D  E  F
           Block II   F  A  E  B  D  C
           Block III  C  B  F  A  D  E

t=3        Block I    A  B  C  D  E  F
           Block II   F  A  E  B  D  C
           Block III  C  B  F  A  D  E
```

The appropriate ANOVA depends on the effect of time. The following assumes that there is no more correlation between samples taken closer together in time than between those taken further apart in time (a 'split-plot in time').

Source	*df*	SS	F
block B	$b-1$	SS_B	MS_B / MS_{Em}
treatment T	$t-1$	SS_T	MS_T / MS_{Em}
error, main (Em)	$(t-1)(b-1)$	SS_{Em}	
time Z	$z-1$	SS_Z	MS_Z / MS_E
time × block (Z × B)	$(z-1)(b-1)$	SS_{ZB}	MS_{ZB} / MS_E
time × treatment (Z × T)	$(z-1)(t-1)$	SS_{ZT}	MS_{ZT} / MS_E
error, E	$(z-1)(t-1)(b-1)$	SS_E	
Total	$btz-1$	SS_{total}	

where z is the number of times that measurements are taken.

Comment 1: equivalent to a design with **two within-subjects factors** (p. 115). If Block ≡ Subject, then Treatment is a within-subjects factor and Time is another within-subjects factor. The 'error, main (Em)' term is Treatment × Block, and the 'error, E' term is Time × Treatment × Block. The design above is the same as the full model for two within-subjects factors discussed earlier (p. 115), except that the agricultural design as quoted here (Tangren, 2002) tests Z against Z×T×B rather than Z×B, which is a bit odd. Compare the split-block design above.

Comment 2: the assumption that there is 'no more correlation between samples taken closer together in time than between those taken further apart in time' is a (strong) version of the assumption of **sphericity** that we've met before in the context of within-subjects designs (p. 25). Time is a within-subjects factor that frequently leads to violations of the sphericity assumption.

8 Mathematics, revision and advanced

8.1 Matrices

Before we can examine a general linear model, it helps to understand matrix notation.

8.1.1 Matrix notation

OK, a quick reminder… This is mostly from Myers & Well (1995, Appendix C) with some additional notes from www.mathworld.com. A plain number, or a symbol that represents one, is called a **scalar** (e.g. 12, –3.5, *c*, *x*). A **vector** is a one-dimensional array of elements, e.g.

$$\mathbf{u} = \begin{bmatrix} 5 \\ 13 \\ 2 \\ -4 \\ 17 \end{bmatrix} \text{ or } \mathbf{v} = \begin{bmatrix} 3 & 28 & 19 & -8 & 0 & 4 \end{bmatrix}$$

Here, we would call **u** a **column vector** and **v** a **row vector**. A **matrix** is a two-dimensional array:

$$\mathbf{Y} = \begin{bmatrix} 2 & 5 & 8 \\ 3 & 1 & 6 \\ 2 & 4 & 1 \\ 7 & 2 & 3 \\ 9 & 6 & 5 \end{bmatrix}$$

(More generally, a scalar is a 0-rank **tensor;** a vector is a 1-rank tensor, having one 'index'; a matrix is a 2-rank tensor; and so on.)

Matrices are frequently denoted with bold-face type. The number of rows and columns is referred to as the **order** of the matrix; the matrix **Y** has order 5×3 (rows $\times$ columns). So **u** is a 5×1 matrix and **v** is a 1×6 matrix. We can refer to an element by using subscripts in the format $\mathbf{element}_{\mathbf{row,column}}$. For example, if we take this matrix:

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & & a_{r,c} & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

then $a_{r,c}$ refers to the element in the *r*th row and the *c*th column of **A**. Sometimes the comma is missed out (a_{rc}).

The **transpose** of matrix **A** is written $\mathbf{A'}$ or $\mathbf{A}^{\mathbf{T}}$. The transpose of a matrix is obtained by swapping the rows and columns. So the transpose of **Y** is

$$\mathbf{Y}^{\mathbf{T}} = \begin{bmatrix} 2 & 3 & 2 & 7 & 9 \\ 5 & 1 & 4 & 2 & 6 \\ 8 & 6 & 1 & 3 & 5 \end{bmatrix}$$

A matrix with equal numbers of rows and columns is called a **square matrix.** A matrix **A** such that $\mathbf{A} = \mathbf{A}^{\mathbf{T}}$ is called a **symmetric matrix.** In a symmetric matrix, like this:

$$\begin{bmatrix} 7 & 2 & 8 \\ 2 & 9 & 3 \\ 8 & 3 & 4 \end{bmatrix}$$

for every element, $a_{ij} = a_{ji}$. If this is true, then the elements are symmetrical about the **major (leading) diagonal** of the matrix, which is the diagonal that extends from the top left to the bottom right. Matrices that have nonzero elements along their major diagonals but only zeros as off-diagonal elements are called **diagonal matrices.**

The **identity matrix** is a special square, diagonal matrix that has 1s along the major diagonal and 0s elsewhere, such as the 3×3 identity matrix:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

8.1.2 Matrix algebra

- **Equality. A** = **B** if $a_{ij} = b_{ij}$ for all *i* and *j*. That is, for two matrices to be equal they must have the same order and identical elements.

- **Addition.** Two matrices may be added if and only if they are of the same order. **C** = **A** + **B** if $c_{ij} = a_{ij} + b_{ij}$ for all *i* and *j*. For example,

$$\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} + \begin{bmatrix} g & h \\ i & j \\ k & l \end{bmatrix} = \begin{bmatrix} a+g & b+h \\ c+i & d+j \\ e+k & f+l \end{bmatrix}$$

- **Subtraction.** Two matrices may be added if and only if they are of the same order. **C** = **A** – **B** if $c_{ij} = a_{ij} - b_{ij}$ for all *i* and *j*. For example,

$$\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} - \begin{bmatrix} g & h \\ i & j \\ k & l \end{bmatrix} = \begin{bmatrix} a-g & b-h \\ c-i & d-j \\ e-k & f-l \end{bmatrix}$$

- **Scalar multiplication.** To multiply a matrix by a scalar, multiple every element in the matrix by the scalar. For example,

$$x\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} xa & xb & xc \\ xd & xe & xf \\ xg & xh & xi \end{bmatrix}$$

- It is not possible to add a scalar to a matrix or to subtract a scalar from a matrix.

- **Matrix multiplication.** To multiply matrix **A** by matrix **B**, giving the result **AB = A × B**, there must be the same number of *columns* in **A** as there are *rows* in **B**. The simplest case is multiplying a row by a column vector, which gives a *scalar* product:

$$\begin{bmatrix} a & b & c \end{bmatrix} \times \begin{bmatrix} d \\ e \\ f \end{bmatrix} = ad + be + cf$$

In general, the product **C** of two matrices **A** and **B** is defined by

$$c_{ik} = a_{ij}b_{jk}$$

where j is summed over for all possible values of i and k (this short-hand notation is known as Einstein summation). We could expand that formula:

$$c_{ik} = \sum_j a_{ij}b_{jk}$$

The number of columns in A must equal the number of rows in B. If you multiple an $x \times y$ matrix by a $y \times z$ matrix, you get an $x \times z$ matrix. For example,

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \times \begin{bmatrix} j & k \\ l & m \\ n & o \end{bmatrix} = \begin{bmatrix} aj+bl+cn & ak+bm+co \\ dj+el+fn & dk+em+fo \\ gj+hl+in & gk+hm+io \end{bmatrix}$$

Not all matrices may be multiplied by each other. Matrix multiplication is **not commutative: AB is not necessarily the same as BA.** (If **A** and **B** are interpreted as linear transformations, then **AB** is the linear transformation in which **B** is applied first, and then **A**.) In fact, if **AB** is defined, **BA** may not even be defined, if the number of rows and columns do not match appropriately.

However, matrix multiplication **is associative: A(BC) = (AB)C = ABC.**

Matrix multiplication is also **distributive: A(B+C) = AB + AC.**

Multiplication by the identity matrix leaves the original matrix unchanged: **IA = AI = A.** Note that the order of the identity matrix that premultiplies **A** (**IA**) does not have to be the same as the order of the identity matrix that postmultiplies it (**AI**), as in this example:

$$\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$$

Matrix multiplication is useful in expressing systems of simultaneous equations. Suppose

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$\mathbf{k} = \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 9 & 11 & 7 \\ 1 & 8 & 12 \\ 4 & 4 & 9 \end{bmatrix}$$

then the matrix equation **Dx = k** indicates that

$$\begin{aligned} 9x+11y+7z &= 1 \\ x+8y+12z &= 4 \\ 4x+4y+9z &= 7 \end{aligned}$$

so the matrix equation represents a set of three simultaneous scalar equations.

- **More obscure ways of multiplying matrices.** There are, of course, other ways to multiply matrices; the one discussed above is the 'ordinary' matrix product (www.sciencedaily.com / encyclopedia / matrix_multiplication). Another is the Hadamard product. For two matrices of the same dimension ($m \times n$), the Hadamard product $\mathbf{A}\cdot\mathbf{B}$ is given by $(\mathbf{A}\cdot\mathbf{B})_{i,j} = \mathbf{A}_{i,j} \times \mathbf{B}_{i,j}$. It's rarely used in linear algebra. There's another, too; if $\mathbf{A}$ is an $n \times p$ matrix and $\mathbf{B}$ is an $m \times q$ matrix, the Kronecker product $\mathbf{A}\otimes\mathbf{B}$ (also known as the direct product or the tensor product) is an $mn \times pq$ matrix:

$$\mathbf{A}\otimes\mathbf{B} = \begin{bmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \cdots & a_{1,p}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \cdots & a_{2,p}\mathbf{B} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,1}\mathbf{B} & a_{n,2}\mathbf{B} & \cdots & a_{n,p}\mathbf{B} \end{bmatrix}$$

  We won't mention these further.

8.1.3 The inverse of a matrix

Dividing a scalar b by another scalar a is equivalent to multiplying b by $1/a$ or a^{-1}, the reciprocal or **inverse** of a. The product of a and its inverse, $a^{-1}\cdot a = a\cdot a^{-1} = 1$. Analogously, a **square matrix A is said to have an inverse** if we can find a matrix $\mathbf{A}^{-1}$ such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

This is handy for solving systems of simultaneous equations; if the equation $\mathbf{Ax} = \mathbf{k}$ represents a system of scalar equations (discussed above), then we can solve the equations by premultiplying both sides of the equation by $\mathbf{A}^{-1}$**:**

$$\begin{aligned} \mathbf{A}^{-1}\mathbf{A}\mathbf{x} &= \mathbf{A}^{-1}\mathbf{k} \\ \mathbf{I}\mathbf{x} &= \mathbf{A}^{-1}\mathbf{k} \\ \mathbf{x} &= \mathbf{A}^{-1}\mathbf{k} \end{aligned}$$

Not all matrices have inverses. **Matrices that have inverses are called nonsingular; matrices that do not have inverses are called singular.** Only square matrices can have inverses, but not all square matrices do.

A matrix will have an inverse only if its rows and columns are **linearly independent.** This is true if no row can be expressed as a linear combination of the other rows, and no column can be expressed as a linear combination of the other columns. (If one row is twice another, for example, the rows are linearly *dependent* and the matrix will have no inverse.)

Calculating the inverse of a matrix can be hard. To find the inverse of a 2×2 matrix, there is a simple formula:

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

where $|\mathbf{A}|$ is called the **determinant** of $\mathbf{A}$; clearly, the inverse is only defined if the determinant is non-zero. So a matrix is singular if its determinant is zero. To find the determinant or inverse of a 3×3 matrix or higher, see www.mathworld.com.

8.1.4. Matrix transposition

See

www.mathworld.com / Transpose.html
planetmath.org / encyclopedia / Transpose.html

- As we saw above, the transpose of a matrix is what you get when you swap all elements a_{ij} with a_{ji}.
- $(\mathbf{A}^{\mathbf{T}})^{\mathbf{T}} \equiv \mathbf{A}$
- $(c\mathbf{A})^{\mathbf{T}} \equiv c\mathbf{A}^{\mathbf{T}}$ where c is a constant
- If $\mathbf{A}$ is invertible, then $(\mathbf{A}^{\mathbf{T}})^{-1} \equiv (\mathbf{A}^{-1})^{\mathbf{T}}$
- $(\mathbf{A}+\mathbf{B})^{\mathbf{T}} \equiv \mathbf{A}^{\mathbf{T}}+\mathbf{B}^{\mathbf{T}}$

  Pretty obvious: $\mathbf{C} = \mathbf{A} + \mathbf{B}$ if $c_{ij} = a_{ij} + b_{ij}$ for all i and j. Therefore, $\mathbf{C}^{\mathbf{T}}$ has $c_{ji} = a_{ij} + b_{ij}$. But $\mathbf{A}^{\mathbf{T}}$ has members a_{ji} and $\mathbf{B}^{\mathbf{T}}$ has members b_{ji}, so $\mathbf{D} = \mathbf{A}^{\mathbf{T}} + \mathbf{B}^{\mathbf{T}}$ has members $d_{ij} = a_{ji} + b_{ji}$. Swap the letters i and j over, and the definition of $\mathbf{D}$ is the same as that of $\mathbf{C}^{\mathbf{T}}$; therefore, $(\mathbf{A}+\mathbf{B})^{\mathbf{T}} \equiv \mathbf{A}^{\mathbf{T}}+\mathbf{B}^{\mathbf{T}}$.

- $(\mathbf{AB})^{\mathbf{T}} \equiv \mathbf{B}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}$; the transpose of a product is the product of the transposes in reverse order. Proof:

$$\begin{aligned}(\mathbf{B}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}})_{ij} &\equiv (b^{\mathbf{T}})_{ik}(a^{\mathbf{T}})_{kj} \\ &\equiv b_{ki}a_{jk} \\ &\equiv a_{jk}b_{ki} \\ &\equiv (\mathbf{AB})_{ji} \\ &\equiv (\mathbf{AB})^{\mathbf{T}}_{ij}\end{aligned}$$

  where Einstein summation has been used to sum over repeated indices implicitly; in Einstein's notation, for example,

$$a_i a_i \equiv \sum_i a_i a_i$$

  and

$$a_{ik} a_{ij} \equiv \sum_i a_{ik} a_{ij}$$

  (see www.mathworld.com/EinsteinSummation.html).

- $\mathbf{A}^{\mathbf{T}}\mathbf{B} \equiv (\mathbf{B}^{\mathbf{T}}\mathbf{A})^{\mathbf{T}}$ and $\mathbf{AB}^{\mathbf{T}} \equiv (\mathbf{BA}^{\mathbf{T}})^{\mathbf{T}}$. These follow directly from the preceding results, since

$$\begin{aligned}\mathbf{A}^{\mathbf{T}}\mathbf{B} &\equiv \mathbf{A}^{\mathbf{T}}(\mathbf{B}^{\mathbf{T}})^{\mathbf{T}} \\ &\equiv (\mathbf{B}^{\mathbf{T}}\mathbf{A})^{\mathbf{T}}\end{aligned}$$

8.2. Calculus

8.2.1. Derivatives

Remember that a derivative of a function $f(x)$, written in one of these ways:

$$f'(x) \equiv \frac{df}{dx} \equiv \frac{d}{dx} f(x)$$

is the rate of change of f with respect to whatever parameters it may have (that is, with respect to x). Formally,

$$f'(x) \coloneqq \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

8.2.2. Simple, non-trigonometric derivatives

$$\frac{d}{dx}ax^n = anx^{n-1} \quad \text{(the power rule)}$$

$$\frac{d}{dx}\ln x = \frac{1}{x}$$

$$\frac{d}{dx}e^x = e^x$$

$$\frac{d}{dx}a^x = \frac{d}{dx}e^{\ln a^x} = \frac{d}{dx}e^{x\ln a} = (\ln a)e^{x\ln a} = (\ln a)a^x$$

8.2.3. Rules for differentiation

Derivatives of sums are equal to the sum of derivatives:

$$\frac{d}{dx}f(x)+\ldots+h(x) = f'(x)+\ldots+h'(x)$$

If c is a constant,

$$\frac{d}{dx}cf(x) = cf'(x)$$

The product rule:

$$\frac{d}{dx}f(x)g(x) = f(x)g'(x)+f'(x)g(x)$$

The chain rule:

$$\frac{dy}{dx} = \frac{dy}{du}.\frac{du}{dx} = \frac{\frac{dy}{du}}{\frac{dx}{du}}$$

8.2.4. Derivatives of a vector function

The derivative of a vector function

$$\mathbf{F}(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_k(x) \end{bmatrix}$$

is given by

$$\frac{d\mathbf{F}}{dx} = \begin{bmatrix} \frac{df_1}{dx} \\ \frac{df_2}{dx} \\ \vdots \\ \frac{df_k}{dx} \end{bmatrix}$$

8.2.5. Partial derivatives

If a function has several parameters, such as $f(x,y)$, we can define the *partial derivative*. This is the derivative when all parameters except the variable of interest are held constant during the differentiation. The partial derivative of $f(x,y)$ with respect to x is written

$$\frac{\partial f}{\partial x} \equiv \frac{\partial}{\partial x}f(x,y) \equiv f_x$$

Formally,

$$D_i f(\mathbf{a}) = \frac{\partial f}{\partial a_i} = \lim_{h\to 0} \frac{1}{h}\left(f\begin{pmatrix} a_1 \\ \vdots \\ a_i + h \\ \vdots \\ a_n \end{pmatrix} - f(\mathbf{a}) \right) = \lim_{h\to 0} \frac{f(\mathbf{a} + h\vec{e}_i) - f(\mathbf{a})}{h}$$

where $\vec{e}_i$ is called the 'standard basis vector' of the *i*th variable (this is a vector with a 1 in position *i* and zeros in every other position, I would infer). Calculating partial derivatives with respect to *x* is easy: you treat everything except *x* as being a constant. For example, if

$$f = x^2 + 2xy + y^2 + y^3 z$$

then

$$\frac{\partial f}{\partial x} = 2x + 2y$$
$$\frac{\partial f}{\partial y} = 2x + 2y + 3y^2 z$$
$$\frac{\partial f}{\partial z} = y^3$$

8.2.6. The chain rule for partial derivatives

The general form of the chain rule, using partial derivatives, is:

$$\frac{df}{ds} = \sum_i \frac{\partial f}{\partial x_i} \frac{dx_i}{ds}$$

See

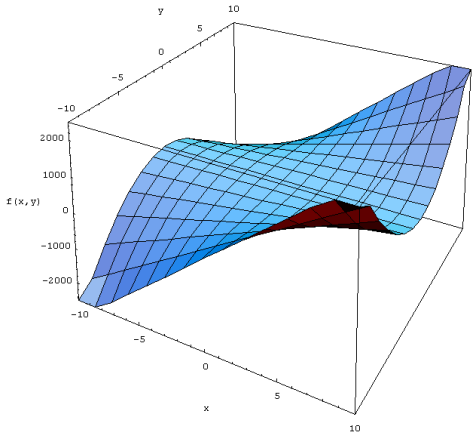
planetmath.org / encyclopedia / Derivative2.html
planetmath.org / encyclopedia / PartialDerivative.html
www.mathworld.com / Derivative.html
phy.asu.edu / phy501-shumway / notes / lec1.pdf

8.2.7. Illustrations of partial derivatives

Suppose we have the function $f = 2x + 3xy^2$. Its partial derivatives with respect to *x* and *y* are:

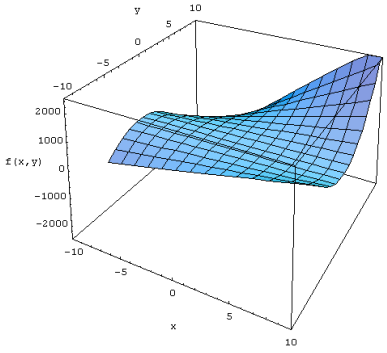
$$\frac{\partial f}{\partial x} = 2 + 3y^2$$
$$\frac{\partial f}{\partial y} = 6xy$$

We can illustrate the whole function:

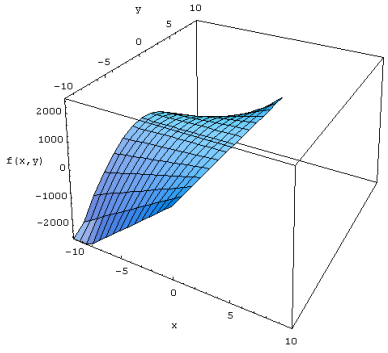


Plot of $f = 2x + 3xy^2$

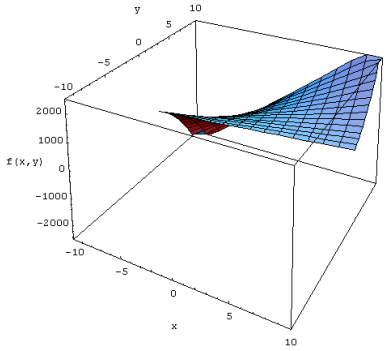
and some partial derivatives:



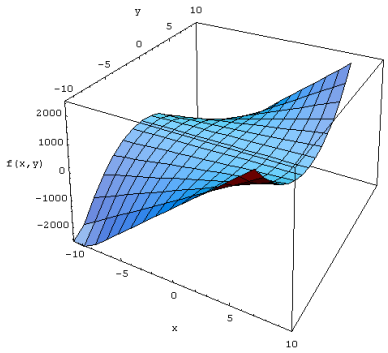
$\frac{\partial f}{\partial x} = 2 + 3y^2$, shown at $y = -5$, where $\frac{\partial f}{\partial x} = 77$



$\frac{\partial f}{\partial y} = 6xy$, shown at $x = 0$, where $\frac{\partial f}{\partial y} = 0$



$\frac{\partial f}{\partial x} = 2 + 3y^2$, shown at $y = +3$, where $\frac{\partial f}{\partial x} = 29$



$\frac{\partial f}{\partial y} = 6xy$, shown at $x = 7$, where $\frac{\partial f}{\partial y} = 42y$

If you're wondering how you'd find the direction in which a ball would roll down this slope (the direction in which the gradient is maximum), and the gradient in that direction, that's given by the vector gradient ('grad'), denoted $\nabla f \equiv \text{grad}(f)$. Details at www.mathworld.com/Gradient.html.

8.3. Solving a GLM (an overdetermined system of equations) (advanced)

Solving a GLM is the the problem of solving $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$ for **b** so as to minimize the sum of squares of the residuals, $\sum (Y_i - \hat{Y})$ or $\sum e_i^2$. When this is solved, **b** contains the correct regression coefficients. Note that we can write an expression for **e**:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}$$
$$\mathbf{e} = \mathbf{y} - \mathbf{Xb}$$

The error (residual) sum of squares can be written like this:

$$\mathrm{SS}_{\text{error}} = \sum e_i^2$$
$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$$
$$\mathbf{e^T} = \begin{bmatrix} e_1 & e_2 & \dots & e_n \end{bmatrix}$$
$$\mathbf{e^T e} = e_1^2 + e_2^2 + \dots + e_n^2 = \sum e_i^2 = \mathrm{SS}_{\text{error}}$$

We can also write it like this:

$$\begin{aligned}
\mathrm{SS}_{\text{error}} &= \mathbf{e^T e} \\
&= (\mathbf{y} - \mathbf{Xb})^{\mathbf{T}} (\mathbf{y} - \mathbf{Xb}) && \text{from definition of } \mathbf{e} \text{ above} \\
&= (\mathbf{y^T} - (\mathbf{Xb})^{\mathbf{T}})(\mathbf{y} - \mathbf{Xb}) && \text{using } (\mathbf{A}+\mathbf{B})^{\mathbf{T}} \equiv \mathbf{A^T} + \mathbf{B^T} \\
&= \mathbf{y^T y} - \mathbf{y^T (Xb)} - \mathbf{(Xb)^T y} + (\mathbf{Xb})^{\mathbf{T}} \mathbf{Xb} && \text{multiplying out} \\
&= \mathbf{y^T y} - ((\mathbf{Xb})^{\mathbf{T}} \mathbf{y})^{\mathbf{T}} - \mathbf{b^T X^T y} + (\mathbf{Xb})^{\mathbf{T}} \mathbf{Xb} && \text{using } \mathbf{A^T B} \equiv (\mathbf{B^T A})^{\mathbf{T}} \\
&= \mathbf{y^T y} - (\mathbf{b^T X^T y})^{\mathbf{T}} - \mathbf{b^T X^T y} + \mathbf{b^T X^T Xb} && \text{using } (\mathbf{AB})^{\mathbf{T}} \equiv \mathbf{B^T A^T} \text{ twice} \\
&= \mathbf{y^T y} - \mathbf{b^T X^T y} - \mathbf{b^T X^T y} + \mathbf{b^T X^T Xb} && \text{because each term in the sum is a real number, and hence equal to its transpose} \\
&= \mathbf{y^T y} - 2\mathbf{b^T X^T y} + \mathbf{b^T X^T Xb}
\end{aligned}$$

To minimize the sum of squares, we solve so that the *partial derivative* of the sum of squares with respect to the model parameters (**b**) is zero. To do this, we will need to use an partial derivative analogue of the product rule for differentiation, which is

$$\frac{d}{dx} f(x)g(x) = f(x)g'(x) + f'(x)g(x)$$

The vector **b** is a set of parameters b_1, b_2, … b_i … b_n. We differentiate with respect to each b_i. The partial derivative of **b** with respect to b_i is a vector with a 1 in the *i*th position and 0 in every other position (see section on partial derivatives, and derivatives of a vector function). We call that vector $\vec{e}_i$ (the 'standard basis vector'); I will use this notation to avoid confusion with the error vector **e**.

$$\frac{\partial}{\partial b_i}\mathbf{b} = \frac{\partial}{\partial b_i}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_i \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \vec{e}_i$$

Similarly,

$$\frac{\partial}{\partial b_i}\mathbf{b}^{\mathbf{T}} = \vec{e}_i^{\mathbf{T}}$$

Armed with this notation, we can obtain the partial derivative of SS_{error}, which we wish to be equal to zero:

$$\begin{aligned}
SS_{error} &= \mathbf{y}^{\mathbf{T}}\mathbf{y} - 2\mathbf{b}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y} + \mathbf{b}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} \\
\frac{\partial SS_{error}}{\partial b_i} &= \frac{\partial}{\partial b_i}\mathbf{y}^{\mathbf{T}}\mathbf{y} - 2\mathbf{b}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y} + \mathbf{b}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} \\
&= 0 - 2\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y} + \left(\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} + \mathbf{b}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\vec{e}_i\right) \\
&= 0 - 2\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y} + \vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} + \left(\vec{e}_i^{\mathbf{T}}\left(\mathbf{b}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{\mathbf{T}}\right)^{\mathbf{T}} \\
&= 0 - 2\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y} + \vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} + \left(\vec{e}_i^{\mathbf{T}}\left(\left(\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{\mathbf{T}}\mathbf{b}\right)^{\mathbf{T}}\right)^{\mathbf{T}}\right)^{\mathbf{T}} \\
&= 0 - 2\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y} + \vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} + \left(\vec{e}_i^{\mathbf{T}}\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{\mathbf{T}}\mathbf{b}\right)^{\mathbf{T}} \\
&= 0 - 2\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y} + \vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} + \left(\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b}\right)^{\mathbf{T}} \\
&= 0 - 2\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y} + \vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} + \vec{e}_i^{\mathbf{T}}\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{\mathbf{T}}\mathbf{b} \\
&= -2\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y} + 2\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} \\
&= 0
\end{aligned}$$

Let's do this in full.

← The first term ($\mathbf{y}^{\mathbf{T}}\mathbf{y}$) contains no terms involving b_i so is treated as a constant. The second is simple. The third has two terms involving b_i, namely $\mathbf{b}^{\mathbf{T}}$ and $\mathbf{b}$, so we use the product rule, differentiating with respect to each in turn.

For the expansion of the right-hand term, we use $(\mathbf{AB})^{\mathbf{T}} \equiv \mathbf{B}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}$ in the form $\mathbf{AB} \equiv (\mathbf{B}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}})^{\mathbf{T}}$.

Next, from $\mathbf{A}^{\mathbf{T}}\mathbf{B} \equiv (\mathbf{B}^{\mathbf{T}}\mathbf{A})^{\mathbf{T}}$ it follows that $(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{\mathbf{T}} \equiv \mathbf{X}^{\mathbf{T}}\mathbf{X}$.

Each term is a real number, and therefore equal to its transpose.

Rearranging:

$$\begin{aligned}
-2\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y} + 2\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} &= 0 \\
2\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} &= 2\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y} \\
\vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} &= \vec{e}_i^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y}
\end{aligned}$$

This says that the *i*th element of $\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b}$ is equal to the *i*th element of $\mathbf{X}^{\mathbf{T}}\mathbf{y}$. Since that is true for all values of *i*, we have the equality

$$\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} = \mathbf{X}^{\mathbf{T}}\mathbf{y}$$

These equations (since the things in the expression above are matrices, they represent more than one equation) are known as the *normal equations* of the linear least-squares problem. If we can find the inverse of $\mathbf{X}^{\mathbf{T}}\mathbf{X}$, and remembering that matrix multiplication is associative — **A(BC) ≡ (AB)C ≡ ABC** — we can derive this expression for **b**:

$$\begin{aligned}
(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} &= (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y} \\
\mathbf{b} &= (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}
\end{aligned}$$

Therefore, our optimal model **b** that minimizes the SS_{error} is given by

$$\mathbf{b} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}$$

Magic. Of course, the solution can only be found if $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ is invertible (which may not be the case if your design matrix contains linearly dependent columns, as with overparametrized design matrices).

For terse versions of these derivations, see

- www.me.psu.edu / sommer / workarea / least_squares.doc
- www.stat.wisc.edu / p / stat / course / st333-larget / public / html / matrix.pdf

8.4. Singular value decomposition to solve GLMs (very advanced)

Singular value decomposition (SVD) is a method that can solve arbitrary GLMs — ones in which we have more information that we need (as is the case in ANOVA), and also ones in which we have exactly the right amount of information, and ones in which we have insufficient information.

When we solve a GLM, we normally have more measurements than we'd need to determine the values of our predictors — the model is $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$, it's *overdetermined* ($\mathbf{e} \neq \mathbf{0}$), and we solve it by minimizing the sum of squares of the errors ($\mathbf{e^T e}$). We can often solve it using the normal equations given above ($\mathbf{X^T Xb} = \mathbf{X^T y}$, or $\mathbf{b} = (\mathbf{X^T X})^{-1}\mathbf{X^T y}$).

When we solve a simple set of equations that are *exactly* determined, we solve $\mathbf{y} = \mathbf{Xb}$ (giving $\mathbf{b} = \mathbf{X^{-1}y}$). This is equivalent to the method for an overdetermined problem, except that $\mathbf{e} = \mathbf{0}$ (our predictions are exact and there is no residual error).

What happens if we don't have enough information? Then our model $\mathbf{y} = \mathbf{Xb}$ is *underdetermined.* Yet if we make *assumptions* about the world, we can still get useful information out.

For example, suppose we're performing a CT scan. We scan a single slice of the body. We want to find a set of X-ray absorbances **b**, one absorbance per voxel. We know which voxels each X-ray beam passes through (**X**), and we know the sum of absorbances for each beam (**y**), assuming some radiation manages to get through (if the X-ray beam is completely absorbed, the maths is harder, which may be why metal causes funny streaky shadows on CT scans). I would guess that CT scans are normally overdetermined, or perhaps exactly determined (though I reckon probably not — it'd be easier to design a machine that made overdetermined scans and the results would probably better, although the price is a bit of time and a bit of unnecessary X-ray radiation). What happens if we had an undetermined situation — like trying to interpret 3D structure from an antero-posterior (AP) and a lateral chest X-ray only? Or like shooting a CT scan from too few directions?

We could assume that tissue is homogeneous unless we receive better information. That corresponds to minimizing the sum of squares of **b** ($\sum b_i^2$). A very simple example: suppose $x + y = 10$. This has an infinite number of solutions. But the one that minimizes $x^2 + y^2$ is $x = 5$, $y = 5$. In general, we may wish to minimize both $\sum e_i^2$ and $\sum b_i^2$. A general technique for this is called **singular value decomposition (SVD).** I won't present it in full, because I don't understand it in full, but it goes like this.

8.4.1. Eigenvectors and eigenvalues

If **A** is a matrix and if there is a column vector $\mathbf{X} \neq \mathbf{0}$ such that

$$\mathbf{AX_R} = \lambda\mathbf{X_R}$$
$$\text{or } (\mathbf{A} - \lambda\mathbf{I})\mathbf{X_R} = 0 \text{ where } \mathbf{I} \text{ is the identity matrix}$$

for some scalar λ, then λ is called the *eigenvalue* of **A** with the corresponding (right) *eigenvector* $\mathbf{X_R}$. (German: 'eigen' = appropriate, innate, own, peculiar.) That is, *an eigenvector is a vector whose direction is unchanged by the transformation* **A**; it is merely stretched by a factor (the eigenvalue). For example, if the matrix represents rotation, it has no eigenvectors. If it represents reflection in a plane, then every vector lying in that plane is an eigenvector, with eigenvalue 1, and any vector perpendicular to the plane will be an eigenvector with eigenvalue –1; these are the only eigenvectors. If the matrix represents 2D reflection (reflection in a line), then vectors lying along that line will be eigenvectors with eigenvalue 1, and vectors perpen-

dicular to that line will be eigenvectors with eigenvalue –1; these are the only eigenvectors. If the matrix represents simultaneous enlargement parallel to the X axis by a factor of *a*, parallel to the Y axis by a factor of *b*, and parallel to the Z axis by a factor of *c*, with $a \neq b \neq c$, so the matrix looks like

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$

then vectors along either the X axis, the Y axis, or the Z axis will be eigenvectors (and these are the only eigenvectors), and their eigenvalues will be *a*, *b*, and *c* respectively. To find eigenvalues, note that if $(\mathbf{A} - \lambda\mathbf{I})\mathbf{X_R}$ and $\mathbf{X_R} \neq \mathbf{0}$ then $(\mathbf{A} - \lambda\mathbf{I})$ must be singular, so solve $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ to get the eigenvalues, and thus the eigenvectors.

Less commonly used: The left eigenvector is a row vector that satisfies $\mathbf{X_L A} = \lambda\mathbf{X_L}$ or $(\mathbf{A} - \lambda\mathbf{I})\mathbf{X_L} = 0$, where **I** is the identity matrix. The eigenvalues for the left and right eigenvectors are the same, although the left and right eigenvectors themselves need not be. When people use the term 'eigenvector' on its own they generally mean 'right eigenvector'.

A square matrix **A** can often be decomposed ('diagonalised') into its eigenvalues and eigenvectors, which are linearly independent. That is,

$$\mathbf{A} = \mathbf{PDP}^{-1}$$

where **P** is a matrix of eigenvectors and **D** is a diagonal matrix of eigenvalues.

See

www.mathworld.com / Eigenvalue.html
www.mathworld.com / Eigenvector.html
www.mathworld.com / EigenDecomposition.html

8.4.2. Singular value decomposition

Any $m \times n$ matrix **X** can be decomposed into

$$\mathbf{X} = \mathbf{USV}^{\mathbf{T}}$$

where

- **U** is an $m \times m$ orthogonal matrix (a matrix **M** is orthogonal if $\mathbf{MM}^{\mathbf{T}} = \mathbf{I}$, i.e. if $\mathbf{M}^{\mathbf{T}} = \mathbf{M}^{-1}$); the columns of **U** are the eigenvectors of $\mathbf{AA}^{\mathbf{T}}$.
- **V** is an $n \times n$ orthogonal matrix; the columns of **V** are the eigenvectors of $\mathbf{A}^{\mathbf{T}}\mathbf{A}$.
- **S** is an $m \times n$ matrix containing a diagonal matrix (a matrix that has nonzero elements along its major diagonal but only zeros elsewhere) with real, non-negative elements σ_i (where *i* is from 1 to the minimum of *m* and *n*) in descending order:

  $$\sigma_1 > \sigma_2 > \ldots > \sigma_{\min(m,n)} > 0$$

  The σ_i elements themselves (the 'singular values') are square roots of eigenvalues from $\mathbf{AA}^{\mathbf{T}}$ or $\mathbf{A}^{\mathbf{T}}\mathbf{A}$. To create **S**, we first create a diagonal matrix $\boldsymbol{\Sigma}$ containing these σ_i elements:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{\min(m,n)} \end{bmatrix}$$

Then we pad it with zeros to make **S** an $m \times n$ matrix:

$$\mathbf{S} = \begin{bmatrix} \mathbf{\Sigma} \\ \mathbf{0} \end{bmatrix} \text{ if } m \geq n \text{ and } \mathbf{S} = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \end{bmatrix} \text{ if } m < n$$

Once we've found the matrices such that $\mathbf{X} = \mathbf{USV^T}$, we can then solve our problem. Since **U** and **V** are orthogonal, their inverses are equal to their transposes. Since **S** is diagonal, its inverse is the diagonal matrix whose elements are the reciprocal of the elements of **S**.

$$\mathbf{X^{-1}} = (\mathbf{USV^T})^{-1} = (\mathbf{V^T})^{-1}\mathbf{S^{-1}}\mathbf{U^{-1}} = \mathbf{VS^{-1}U^T}$$

where the diagonal elements of $\mathbf{S^{-1}}$ are $1/S_i$ [that is, $\mathbf{S^{-1}}$ = diag$(1/S_i)$]. Therefore, since $\mathbf{y} = \mathbf{Xb}$, we have $\mathbf{b} = \mathbf{X^{-1}y}$ and hence

$$\mathbf{b} = \mathbf{VS^{-1}U^Ty}$$

It is possible to solve equations even if the matrices are singular or close to singular using this technique: when you obtain $\mathbf{S^{-1}}$, by taking the values $1/S_i$, if S_i is smaller than a threshold value (the *singularity threshold*) you replace $1/S_i$ with 0. That is, SVD finds the least squares best compromise solution of the linear equation system. For details and proof, see Press *et al.* (1992, pp. 59-70, 676-680) and

rkb.home.cern.ch / rkb / AN16pp / node265.html
www.mlahanas.de / Math / svd.htm

8.4.3. An underdetermined set of equations: the role of expectations

(RNC, April 2004.) Alternatively, we might have prior expectations — in our radiological example, we expect to find a heart, we expect that ribs curve round the side, and so on. We might say that we'd like to interpret the data to fit our expectations as far as possible. If our prior expectations are **p**, then this would correspond to minimizing the sum of squares of (**b** – **p**). We can say that **b** = **p** + **d**, where **d** represents the deviation from prior expectations. Thus,

$$\begin{aligned} \mathbf{y} &= \mathbf{Xb} \\ &= \mathbf{X(p+d)} \\ &= \mathbf{Xp} + \mathbf{Xd} \\ \mathbf{y} - \mathbf{Xp} &= \mathbf{Xd} \end{aligned}$$

The usual singular value decomposition $\mathbf{X} = \mathbf{USV^T}$ is used to solve $\mathbf{y} = \mathbf{Xb}$ for **b**, minimizing the sum of squares of **b** when the system is underdetermined; the solution is given by $\mathbf{b} = \mathbf{VS^{-1}U^Ty}$. In the present case, we use the same decomposition of **X** and simply rewrite to solve for **d**, minimizing its sum of squares:

$$\mathbf{d} = \mathbf{VS^{-1}U^T(y - Xp)}$$

and therefore since $\mathbf{b} = \mathbf{p} + \mathbf{d}$,

$$\mathbf{b} = \mathbf{p} + \mathbf{VS^{-1}U^T(y - Xp)}$$

8.5 Random variables, means, and variances

8.5.1 Summation

If we have n scores we could denote them x_1, x_2, … x_n. Their sum can be written in the following ways:

$$x_1 + x_2 + \ldots + x_n = \sum_{i=1}^{n} x_i$$
$$= \sum_{i} x_i$$

The following are easily proved. If c is a constant, then

$$\sum cx = c\sum x$$
$$\sum_{i=1}^{n} c = nc$$

The summation sign operates like a multiplier on quantities within parentheses. For example:

$$\sum_{i=1}^{n}(x_i - y_i) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i$$
$$\sum (x-y)^2 = \sum x^2 + \sum y^2 + 2\sum xy$$

8.5.2 Random variables; definition of mean and variance

A **random variable (RV)** is a measurable or countable quantity that can take any of a range of values and which has a **probability distribution** associated with it, i.e. there is a means of giving the probability of the variable taking a particular value. If the values an RV can take are real numbers (i.e. an infinite number of possibilities) then the RV is said to be **continuous;** otherwise it is **discrete.** The probability that a discrete RV X has the value x is denoted $P(x)$. We can then define the mean or **expected value:**

$$E[X] = \sum xP(x)$$

and the **variance:**

$$\begin{aligned} Var[X] &= E[(x - E[X])^2] \\ &= \sum (x - E[X])^2 P(x) \\ &= \sum (x^2 - 2xE[X] + (E[X])^2)P(x) \\ &= \sum x^2 P(x) - \sum 2xP(x)E[X] + \sum (E[X])^2 P(x) \\ &= \sum x^2 P(x) - 2E[X]E[X] + (E[X])^2 \sum P(x) \\ &= \sum x^2 P(x) - 2(E[X])^2 + (E[X])^2 \\ &= E[X^2] - (E[X])^2 \end{aligned}$$

and the **standard deviation, σ:**

$$\sigma^2 = Var[X]$$

8.5.3 Continuous random variables

For a continuous random variable X, the probability $P(x)$ of an exact value x occurring is zero, so we must work with the probability density function (PDF), $f(x)$. This is defined as

$$P(a \le x \le b) = \int_a^b f(x)dx$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$
$$\forall x : f(x) \geq 0$$

($\forall x$ means 'for all values of x'). The mean or expected value E[X] is defined as

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

The variance, Var[X] is given by

$$Var[X] = \int_{-\infty}^{\infty} x^2 f(x)dx - (E[X])^2$$

The cumulative distribution function (CDF, also known as the 'distribution function' or 'cumulative density function'), $F(a)$, is given by

$$F(a) = \int_{-\infty}^{a} f(x)dx$$

i.e.

$$F(a) = P(x \leq a)$$
$$P(a \leq x \leq b) = F(b) - F(a)$$

8.5.4 Expected values

If X is a random variable and c is a constant, $E(X)$ denotes the expected value of X.

$$E(c) = c$$
$$E(cX) = cE(X)$$

$E()$ acts like a multiplier. For example:

$$E(X + Y) = E(X) + E(Y)$$
$$E(X + c) = E(X) + E(c) = E(X) + c$$
$$E[(X + Y)^2] = E(X)^2 + E(Y)^2 + 2E(XY)$$

If X and Y are independently distributed, then

$$E(XY) = E(X)E(Y)$$

8.5.5 The sample mean and SD are unbiased estimators of μ and σ²

We will use X to denote the random variable, x for an individual value of that random variable, $\bar{x}$ for the sample mean, s_X^2 for the sample variance (sometimes written $\hat{\sigma}_X^2$), μ_X for the population mean, and σ_X^2 for the population variance. First, the mean:

$$E(\bar{x}) = E\left(\frac{\sum x}{n}\right) = \frac{1}{n}E(\sum x) = \frac{1}{n}\sum E(X) = \frac{1}{n}nE(X) = E(X) = \mu$$

Now the standard deviation (Myers & Well, 1995, p. 592). Consider first the numerator (the sum of squares) [N.B. line 3 uses the fact $\sum(x - \mu) = n(\bar{x} - \mu)$]:

$$
\begin{aligned}
E[\sum(x-\bar{x})^2] &= E\sum\left[(x-\mu)-(\bar{x}-\mu)\right]^2 \\
&= E[\sum(x-\mu)^2 + \sum(\bar{x}-\mu)^2 - 2(\bar{x}-\mu)\sum(x-\mu)] \\
&= E[\sum(x-\mu)^2 + n(\bar{x}-\mu)^2 - 2n(\bar{x}-\mu)^2] \\
&= E[\sum(x-\mu)^2 - n(\bar{x}-\mu)^2] \\
&= \sum E(x-\mu)^2 - nE(\bar{x}-\mu)^2
\end{aligned}
$$

The average squared deviation of a quantity from its average is a variance; that is,

$$E(x-\mu)^2 = \sigma_X^2$$

and, by the Central Limit Theorem,

$$E(\bar{x}-\mu)^2 = \sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$$

Therefore,

$$
\begin{aligned}
E[\sum(x-\bar{x})^2] &= n\sigma_X^2 - \frac{n\sigma_X^2}{n} \\
&= (n-1)\sigma_X^2
\end{aligned}
$$

Hence

$$E\left(\frac{\sum(x-\bar{x})^2}{n-1}\right) = E(s_X^2) = \sigma_X^2$$

8.5.6 Variance laws

If X and Y are two random variables with variances $V(X) = \sigma_X^2$ and $V(Y) = \sigma_Y^2$, and c is a constant, then

$$
\begin{aligned}
V(c) &= 0 \\
V(X+c) &= V(X) = \sigma_X^2 \\
V(cX) &= c^2V(X) = c^2\sigma_X^2
\end{aligned}
$$

$$
\begin{aligned}
V(X+Y) &= \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho_{XY}\sigma_X\sigma_Y \\
V(X-Y) &= \sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho_{XY}\sigma_X\sigma_Y
\end{aligned}
$$

where ρ is the correlation between X and Y; $\rho_{XY}\sigma_X\sigma_Y$ is also known as the covariance:

$$\mathrm{cov}_{XY} = \rho_{XY}\sigma_X\sigma_Y$$

Therefore, if X and Y are **independent,**

$$
\begin{aligned}
\rho_{XY} &= 0 \\
\mathrm{cov}_{XY} &= 0 \\
V(X+Y) &= \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 \\
V(X-Y) &= \sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2
\end{aligned}
$$

8.5.7 Distribution of a set of means: the standard error of the mean

See Frank & Althoen (1994, pp. 281-289). Let X_1, X_2, … X_N be a set of sample means. Then $\bar{X}$ is the mean of all those sample means. First we derive the **density function** of $\bar{X}$.

If we sample n values from a random variable, calling them x_1, x_2… x_n, then their mean is

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

or

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

Likewise, for a set of n random variables X_1, X_2… X_n,

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n$$

Let

$$W_i = \frac{1}{n}X_i$$

then

$$\bar{X} = W_1 + W_2 + \cdots + W_n$$

If X_1, X_2… X_n are independent and identically distributed, as when observations are independent, then W_1, W_2… W_n are likewise independent and identically distributed. The mean $\bar{X}$ can therefore be expressed as the *sum* of n independent, identically distributed random variables, W_i.

The Central Limit Theorem tells us that if W_1, W_2, … W_n are independent, identically distributed random variables and $Y = W_1 + W_2 + \ldots + W_n$, then the probability density function of Y approaches the normal distribution

$$\frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{\frac{-(y-\mu_y)^2}{2\sigma_Y^2}}$$

as $n \to \infty$.

Next we derive the expected value of the sample mean, $E(\bar{X})$. (We saw one derivation above; this is a fuller version.) Since

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

it follows that

$$E(\bar{X}) = E\left(\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right) = \frac{1}{n}E(X_1 + X_2 + \cdots + X_n)$$

From the Algebra of Expectations, the expected value of a sum is equal to the sum of the expected values. So if $E(X_1) = \mu_1$, $E(X_2) = \mu_2$, $E(X_n) = \mu_n$, etc., then

$$E(\bar{X}) = \frac{1}{n}E(\mu_1 + \mu_2 + \cdots + \mu_n)$$

Let us suppose the population mean is μ. Since the distributions of X_1, X_2, … X_n are all identical to the population distribution, it follows that all n random variables have the same expected value:

$$\mu_1 = \mu_2 = \cdots = \mu_n = \mu$$

So

$$E(\overline{X}) = \frac{1}{n} E(\mu + \mu + \cdots + \mu) = \frac{1}{n} n\mu = \mu$$

So the expected value of the sample mean (the mean of a set of sample means) is equal to the population mean.

How about the **variance** of $\overline{X}$?

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

So

$$V(\overline{X}) = V\left(\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right)$$

When you factor a constant out of a variance, it's squared:

$$V(\overline{X}) = \frac{1}{n^2} V(X_1 + X_2 + \cdots + X_n)$$

The variance of a sum of *n* independent random variables is the sum of the individual variances. If $V(X_1) = \sigma_1^2$, $V(X_2) = \sigma_2^2$, …, $V(X_n) = \sigma_n^2$, then

$$V(X_1 + X_2 + \cdots + X_n) = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$$

so

$$V(\overline{X}) = \frac{1}{n^2}\left(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2\right)$$

and since the variables representing our *n* observations all have the same distribution as the parent population, they must all have the same variance, namely σ^2, the population variance. So

$$V(\overline{X}) = \frac{1}{n^2}\left(N\sigma^2\right) = \frac{\sigma^2}{n}$$

So for samples of *n* independent observations, the variance of the sample means is equal to the population variance divided by the sample size:

$$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}$$

and so the standard deviation of the sample means (the *standard error of the mean*) is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

8.6 The harmonic mean

The harmonic mean of *n* observations $x_1, x_2, \ldots x_n$ is

$$\overline{X}_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}$$

9 Glossary

Symbols:

$\Rightarrow$	implies
$\equiv$	is equivalent to
$\bar{x}$	mean of a set of values of x
ε	error
$\hat{\varepsilon}$	Greenhouse–Geisser correction (see p. 25)
$\tilde{\varepsilon}$	Huynh–Feldt correction (see p. 25)
μ	mean
ρ	population correlation
r	sample correlation
r_{xy} or $r_{x.y}$	correlation between x and y
$r_{y.a,b,c}$	multiple correlation between y and (a, b, c)
$r_{y.(x\|z)}$	semipartial correlation between y and x, having partialled out z (see p. 100)
$r_{y.x\|z}$	partial correlation between y and x, having partialled out z (see p. 100)
$\sum$	sum of (see p. 209)
σ_X	population standard deviation of X
s_X	sample standard deviation of X
σ_X^2	population variance of X
s_x^2	sample variance of X

- **Additive model.** In *within-subjects* ANOVA, a structural model that assumes the effects of within-subjects treatments are the same for all subjects.
- **ANCOVA.** Analysis of covariance: an ANOVA that uses a *covariate* as a *predictor variable*.
- **ANOVA.** Analysis of *variance*. See p. 8→ for an explanation of how it works.
- ***A priori* tests.** Tests planned in advance of obtaining the data; compare *post hoc tests*.
- **Balanced ANOVA.** An ANOVA is said to be balanced when all the cells have equal n, when there are no missing cells, and if there is a *nested design*, when the nesting is balanced so that equal numbers of levels of the nested factor appear in the levels of the factor(s) that they are nested within. This greatly simplifies the computation.
- **Between-subjects** (factor or covariate). If each subject is only tested at a single level of an independent variable, the independent variable is called a between-subjects factor. Compare *within-subjects*.
- **Carryover effects.** See *within-subjects.*
- **Categorical predictor variable.** A variable measured on a nominal scale, whose categories identify class or group membership, used to predict one or more dependent variables. Often called a *factor*.
- **Continuous predictor variable.** A continuous variable used to predict one or more dependent variables. Often called a *covariate*.
- **Covariance matrix.** If you have three variables x, y, z, the covariance matrix, denoted $\sum$, is $\Sigma = \begin{array}{c} \\ x \\ y \\ z \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} \sigma_x^2 & \mathrm{cov}_{xy} & \mathrm{cov}_{xz} \\ \mathrm{cov}_{xy} & \sigma_y^2 & \mathrm{cov}_{yz} \\ \mathrm{cov}_{xz} & \mathrm{cov}_{yz} & \sigma_z^2 \end{bmatrix} \end{array}$ where cov_{xy} is the covariance of x and y ($= \rho_{xy}\sigma_x\sigma_y$ where ρ_{xy} is the correlation between x and y and σ_x is the variance of x). Obviously, $\mathrm{cov}_{xx} = \sigma_x^2$. It is sometimes used to check for **compound symmetry** of the covariance matrix, which is a fancy way of saying

$\sigma_x^2 = \sigma_y^2 = \sigma_z^2$ (all numbers on the leading diagonal the same as each other). and $\text{cov}_{xy} = \text{cov}_{yz} = \text{cov}_{xz}$ (all numbers not on the leading diagonal the same as each other). If there is compound symmetry, there is also *sphericity*, which is what's important when you're running ANOVAs with *within-subjects factors*. On the other hand, you can have sphericity without having compound symmetry; see p. 25→.

- **Conservative.** Apt to give *p* values that are too large.
- **Contrast.** See *linear contrast.*
- **Covariate.** A continuous variable (one that can take any value) used as a *predictor variable*.
- **Degrees of freedom (*df*).** Estimates of parameters can be based upon different amounts of information. The number of independent pieces of information that go into the estimate of a parameter is called the degrees of freedom (d.f. or *df*). Or, the number of observations free to vary. For example, if you pick three numbers at random, you have 3 *df* — but once you calculate the sample mean, $\bar{x}$, you only have two *df* left, because you can only alter two numbers freely; the third is constrained by the fact that you have 'fixed' $\bar{x}$. Or, the number of measurements exceeding the amount absolutely necessary to measure the 'object' (or parameter) in question. To measure the length of a rod requires 1 measurement. If 10 measurements are taken, then the set of 10 measurements has 9 *df*. In general, the df of an estimate is the number of independent scores that go into the estimate minus the number of parameters estimated from those scores as intermediate steps. For example, if the population variance σ^2 is estimated (by the sample variance s^2) from a random sample of n independent scores, then the number of degrees of freedom is equal to the number of independent scores (n) minus the number of parameters estimated as intermediate steps (one, as μ is estimated by $\bar{x}$) and is therefore $n - 1$.
- **Dependent variable.** The variable you measure, but do not control. ANOVA is about predicting the value of a single dependent variable using one or more *predictor variables*.
- **Design matrix.** The matrix in a *general linear model* that specifies the experimental design — how different factors and covariates contribute to particular values of the dependent variable(s).
- **Doubly-nested design.** One in which there are two levels of nesting (see *nested design*). Some are described on p. 159→.
- **Error term.** To test the effect of a predictor variable of interest with an ANOVA, the variability attributable to it ($\text{MS}_{\text{variable}}$) is compared to variability attributed to an appropriate 'error term' (MS_{error}), which measures an appropriate *error variability*. The error term is valid if the *expected mean square* for the variable, $E(\text{MS}_{\text{variable}})$, differs from $E(\text{MS}_{\text{error}})$ only in a way attributable solely to the variable of interest.
- **Error variability** (or error variance, σ_e^2). Variability among observations that cannot be attributed to the effects of the independent variable(s). May include measurement error but also the effects of lots of irrelevant variables that are not measured or considered. It may be possible to reduce the error variability by accounting for some of them, and designing our experiment accordingly. For example, if we want to study the effects of two methods of teaching reading on children's reading performance, rather than randomly assigning all our students to teaching method 1 or teaching method 2, we could split our children into groups with low / medium / high intelligence, and randomly allocate students from each level of intelligence to one of our two teaching methods. If intelligence accounts for some of the variability in reading ability, accounting for it in this way will reduce our error variability. *Within-subjects* designs take this principle further (but are susceptible to *carryover effects*).
- **Expected mean square (EMS).** The value a mean square (MS) would be expected to have if the null hypothesis were true.
- ***F* ratio.** The ratio of two variances. In ANOVA, the ratio of the *mean square (MS)* for a *predictor variable* to the MS of the corresponding *error term*.

- **Factor.** A discrete variable (one that can take only certain values) used as a *predictor variable.* A categorical predictor. Factors have a certain number of *levels.*
- **Factorial ANOVA.** An ANOVA using factors as predictor variables. The term is often used to refer to ANOVAs involving more than one factor (compare *one-way ANOVA*). Factorial designs range from the completely randomized design (subjects are randomly assigned to, and serve in only one of several different treatment conditions, i.e. completely between-subjects design), via mixed designs (both between-subjects and within-subjects factors) to completely within-subjects designs, in which each subject serves in every condition.
- **Fixed factor.** A *factor* that contains all the levels we are interested in (e.g. the factor 'sex' has the levels male and female). Compare *random* factor and see p. 31.
- **Gaussian distribution.** Normal distribution.
- **General linear model.** A general way of predicting one or more *dependent variables* from one or more *predictor variables*, be they categorical or continuous. Subsumes regression, multiple regression, ANOVA, ANCOVA, MANOVA, MANCOVA, and so on.
- **Greenhouse–Geisser correction / epsilon.** If the *sphericity assumption* is violated in an ANOVA involving within-subjects factors, you can correct the *df* for any term involving the WS factor (and the *df* of the corresponding error term) by multiplying both by this correction factor. Often written $\hat{\varepsilon}$, where $0 < \hat{\varepsilon} \leq 1$. Originally from Greenhouse & Geisser (1959).
- **Heterogeneity of variance.** Opposite of *homogeneity of variance.* When variances for different treatments are *not* the same.
- **Hierarchical design.** One in which one variable is *nested* within a second, which is itself nested within a third. A doubly-nested design (such as the split-split plot design) is the simplest form of hierarchical designs. They're complex.
- **Homogeneity of variance.** When a set of variances are all equal. If you perform an ANOVA with a factor with *a* levels, the homogeneity of variance assumption is that $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_a^2 = \sigma_e^2$, where σ_e^2 is the *error variance*.
- **Huynh–Feldt correction / epsilon.** If the *sphericity assumption* is violated in an ANOVA involving within-subjects factors, you can correct the *df* for any term involving the WS factor (and the *df* of the corresponding error term) by multiplying both by this correction factor. Often written $\tilde{\varepsilon}$, where $0 < \tilde{\varepsilon} \leq 1$. Originally from Huynh & Feldt (1970).
- **Independent variable.** The variables thought to be influencing the *dependent variable(s).* In experiments, independent variables are manipulated. In correlational studies, independent variables are observed. (The advantage of the experiment is the ease of making causal inferences.)
- **Interaction.** There is an interaction between factors A and B if the effect of factor A depends on the level of factor B, or vice versa. For example, if your dependent variable is engine speed, and your factors are 'presence of spark plugs (Y / N)' (A) and 'presence of petrol (Y / N)' (B), you will find an interaction such that factor A only influences engine speed at the 'petrol present' level of B; similarly, factor B only influences engine speed at the 'spark plugs present' level of B. This is a binary example, but interactions need not be. Compare *main effect, simple effect.*
- **Intercept.** The contribution of the grand mean to the observations. See p. 65. The *F* test on the intercept term ($MS_{intercept} / MS_{error}$) tests the null hypothesis that the grand mean is zero.
- **Level** (of a factor). One of the values that a discrete predictor variable (factor) can take. For example, the factor Weekday might have five levels — Monday, Tuesday, Wednesday, Thursday, Friday. We might write the factor as $Weekday_5$ in descriptions of ANOVA models (as in 'Tedium = $Drowsiness_2 \times Weekday_5 \times S$'), or write the levels themselves as $Weekday_1$ …$Weekday_5$.
- **Levene's test** (for heterogeneity of variance). Originally from Levene (1960). Tests the assumption of *homogeneity of variance*. If Levene's test produces a 'significant' result, the assumption of homogeneity of variance cannot be made (this is generally a Bad Thing and suggests that you might need to transform your data to improve the situation; see p. 34).

- **Liberal.** Apt to give *p* values that are too small.
- **Linear contrasts.** Comparisons between linear combinations of different groups, used to test specific hypotheses. See p. 75→.
- **Linear regression.** Predicting *Y* from *X* using the equation of a straight line: $\hat{Y} = bX + a$. May be performed with *regression ANOVA*.
- **Logistic regression.** See Howell (1997, pp. 548-558). A logistic function is a sigmoid (see www.mathworld.com). If your dependent variable is dichotomous (categorial) but ordered ('flight on time' versus 'flight late', for example) and you wish to predict this (for example, by pilot experience), a logistic function is often better than a straight line. It reflects the fact that the dichotomy imposes a cutoff on some underlying continuous variable (e.g. once your flight delay in seconds — continuous variable — reaches a certain level, you classify the flight as late — dichotomous variable). Dichotomous variables can be converted into variables suitable for linear regression by converting the *probability* of falling into one category, *P*(flight late), into the *odds* of falling into that category, using $\text{odds} = \frac{P(A)}{P(\neg A)}$, and then into the *log odds*, using the natural (base *e*) logarithm $\log_e(\text{odds}) = \ln(\text{odds})$. The probability is therefore a logistic function of the log odds: $\text{probability} = \frac{e^{\ln(\text{odds})}}{1 + e^{\ln(\text{odds})}}$, so performing a linear regression on the log odds is equivalent to performing a logistic regression on probability. This is pretty much what logistic regression does, give or take some procedural wrinkles. Odds ratios (likelihood ratios), the odds for one group divided by the odds for another group, emerge from logistic regression in the way that slope estimates emerge from linear regression, but the statistical tests involved are different. Logistic regression is a computationally iterative task; there's no simple formula (the computer works out the model that best fits the data iteratively).
- **Main effect.** A main effect is an effect of a factor *regardless* of the other factor(s). Compare *simple effect; interaction.*
- **MANCOVA.** Multivariate analysis of covariance; see *MANOVA* and *ANCOVA*.
- **MANOVA.** Multivariate ANOVA — ANOVA that deals with multiple *dependent variables* simultaneously. Not covered in this document. For example, if you think that your treatment has a bigger effect on dependent variable Y_2 than on variable Y_1, how can you see if that is the case? Certainly not by making categorical decisions based on *p* values (significant effect on Y_1, not significant effect on Y_2 — this wouldn't mean that the effect on Y_1 and Y_2 were significantly different!). Instead, you should enter Y_1 and Y_2 into a MANOVA.
- **Mauchly's test** (for sphericity of the covariance matrix). Originally from Mauchly (1940). See *sphericity*, *covariance matrix*, and p. 25.
- **Mean square (MS).** A *sum of squares (SS)* divided by the corresponding number of *degrees of freedom (df)*, or number of independent observations upon which your SS was based. This gives you the mean 'squared deviation from the mean', or the 'mean square'. Effectively, a *variance*.
- **Mixed model.** An ANOVA model that includes both *between-subjects* and *within-subjects* predictor variables. Alternatively, one that includes both *fixed* and *random* factors. The two uses are often equivalent in practice, since Subjects is usually a random factor.
- **Multiple regression.** Predicting a dependent variable on the basis of two or more continuous variables. Equivalent to ANOVA with two or more *covariates*.
- **Nested design.** An ANOVA design in which variability due to one factor is 'nested' within variability due to another factor. For example, if one were to administer four different tests to four school classes (i.e. a between-groups factor with four levels), and two of those four classes are in school A, whereas the other two classes are in school B, then the levels of the first factor (four different tests) would be nested in the second factor (two different schools). A very common example is a design with one between-subjects factor and one within-subjects factor, written A × (U × S); variation due to subjects is nested within variation due to A (or, for short-hand, S is nested within A), because each subject is only tested at one level of the between-subjects factor(s). We might write this S / A ('S is nested within A'); SPSS uses the alternative notation of S(A). See also *doubly-nested design*.

- **Nonadditive model.** In *within-subjects* ANOVA, a structural model that allows that the effects of within-subjects treatments can differ across subjects.
- **Null hypothesis.** For a general discussion of null hypotheses, see handouts at www.pobox.com/~rudolf/psychology. In a one-way ANOVA, when you test the main effect of a factor A with *a* levels, your null hypothesis is that $\mu_1 = \mu_2 = \ldots = \mu_a$. If you reject this null hypothesis (if your *F* ratio is large and significant), you conclude that the effects of all *a* levels of A were not the same. But if there are >2 levels of A, you do not yet know which levels differed from each other; see *post hoc tests.*
- **One-way ANOVA.** ANOVA with a single between-subjects factor.
- **Order effects.** See *within-subjects.*
- **Overparameterized model.** A way of specifying a *general linear model* design matrix in which a separate predictor variable is created for each group identified by a factor. For example, to code Sex, one variable would be created in which males score 1 and females score 0, and another variable would be created in which males score 0 and females score 1. These two variables contain mutually redundant information: there are more predictor variables than are necessary to determine the relationship of a set of predictors to a set of dependent variables. Compare *sigma-restricted model.*
- **Planned contrasts.** *Linear contrasts* run as *a priori tests.*
- **Polynomial ANCOVA.** An *ANCOVA* in which a nonlinear term is used as a *predictor variable* (such as x^2, x^3…, rather than the usual x). See Myers & Well (1995, p. 460).
- ***Post hoc* tests.** Statistical tests you run after an ANOVA to examine the nature of any main effects or interactions you found. For example, if you had an ANOVA with a single *between-subjects factor* with three *levels*, sham/core/shell, and you found a main effect of this factor, was this due to a difference between sham and core subjects? Sham and shell? Shell and core? Are all of them different? There are many *post hoc* tests available for this sort of purpose. However, there are statistical pitfalls if you run many post-hoc tests; you may make Type I errors (see handouts at www.pobox.com/~rudolf/psychology) simply because you are running lots of tests. *Post hoc* tests may include further ANOVAs of subsets of your original data — for example, after finding a significant Group × Difficulty *interaction*, you might ask whether there was a *simple effect* of Group at the 'easy' level of the Difficulty factor, and whether there was a *simple effect* of Group at the 'difficult' level of the Difficulty factor (see pp. 20, 39→).
- **Power of an ANOVA.** Complex to work out. But things that increase the expected *F* ratio for a particular term if the null hypothesis is false increase power, and $F = \frac{\mathrm{MS}_{\mathrm{predictor}}}{\mathrm{MS}_{\mathrm{error}}} = \frac{\mathrm{SS}_{\mathrm{predictor}} \times df_{\mathrm{error}}}{\mathrm{SS}_{\mathrm{error}} \times df_{\mathrm{predictor}}}$. Bigger samples contribute to a larger *df* for your error term; this therefore decreases $\mathrm{MS}_{\mathrm{error}}$ and increases the expected *F* if the null hypothesis is false, and this therefore increases your power. The larger the ratio of $E(\mathrm{MS}_{\mathrm{treatment}})$ to $E(\mathrm{MS}_{\mathrm{error}})$, the larger your power. Sometimes two different structural models give you different EMS ratios; you can use this principle to find out which is more powerful for detecting the effects of a particular effect (see p. 73→). For references to methods of calculating power directly, see p. 102.
- **Predictor variable.** Factors and covariates: things that you use to predict your dependent variable.
- **Pseudoreplication.** What you do when you analyse correlated data without accounting for the correlation. A Bad Thing — entirely Wrong. For example, you could take 3 subjects, measure each 10 times, and pretend that you had 30 independent measurements. No, no, no, no, no. Account for the correlation in your analysis (in this case, by introducing a Subject factor and using an appropriate ANOVA design with a within-subjects factor).
- **Random factor.** A *factor* whose levels we have sampled at random from many possible alternatives. For example, Subjects is a random factor — we pick our subjects out of a large potential pool, and if we repeat the experiment, we may use different subjects. Compare *fixed factor* and see p. 31.

- **Regression ANOVA.** Performing linear regression using ANOVA. A simple linear regression is an ANOVA with a single *covariate* (i.e. ANCOVA) and no other factors.
- **Repeated measures.** Same as within-subjects. 'Repeated measures' is the more general term — within-subjects designs involve repeated measurements of the same subject, but things other than subjects can also be measured repeatedly. In general, within-subjects / repeated-measures analysis is to do with accounting for *relatedness* between sets of observations above that you'd expect by chance. Repeated measurement of a subject will tend to generate data that are more closely related (by virtue of coming from the same subject) than data from different subjects.
- **Robust.** A test that gives correct *p* values even when its assumptions are violated to some degree ('this test is fairly robust to violation of the normality assumption…').
- **Sequence effects.** See *within-subjects.*
- **Sigma-restricted model.** A way of specifying a *general linear model* in which a categorical variable with *k* possible levels is coded in a design matrix with *k* – 1 variables. The values used to code membership of particular groups sum to zero. For example, to code Sex, one variable would be created in which males score +1 and females –1. Compare *overparameterized model.*
- **Simple effect.** An effect of one factor *considered at only one level* of another factor. A simple effect of A at level 2 of factor B is written 'A at B_2' or 'A / B_2'. See *main effect, interaction,* and pp. 20, 39→.
- **Source of variance (SV).** Something contributing to variation in a dependent variable. Includes *predictor variables* and *error variability*.
- **Sphericity assumption.** An important assumption applicable to *within-subjects (repeated measures)* ANOVA. Sphericity is the assumption of *homogeneity of variance of difference scores.* Suppose we test 5 subjects at three levels of A. We can therefore calculate three sets of difference scores ($A_3 - A_2$), ($A_2 - A_1$), and ($A_3 - A_1$), for each subject. Sphericity is the assumption that the variances of these difference scores are the same. See p. 25→.
- **Standard deviation.** The square root of the *variance.*
- **Structural model.** An equation giving the value of the *dependent variable* in terms of *sources of variability* including *predictor variables* and *error variability.*
- **Sum of squares (SS).** In full, the sum of the squared deviations from the mean. See *variance*. Sums of squares are used in preference to actual variances in ANOVA, because sample sums of squares are additive (you can add them up and they still mean something) whereas sample variances are not additive unless they're based on the same number of *degrees of freedom*.
- ***t* test, one-sample.** Equivalent to testing $MS_{intercept} / MS_{error}$ with an ANOVA with no other factors (odd as that sounds). $F_{1,k} = t_k^2$ and $t_k = \sqrt{F_{1,k}}$. See *intercept.*
- ***t* test, two-sample, paired.** Equivalent to ANOVA with one within-subjects factor with two levels. $F_{1,k} = t_k^2$ and $t_k = \sqrt{F_{1,k}}$.
- ***t* test, two-sample, unpaired.** Equivalent to ANOVA with one between-subjects factor with two levels. $F_{1,k} = t_k^2$ and $t_k = \sqrt{F_{1,k}}$.
- **Variance.** To calculate the variance of a set of observations, take each observation and subtract it from the mean. This gives you a set of deviations from the mean. Square them and add them up. At this stage you have the sum of the squared deviations from the mean, also known as the *sum of squares (SS).* Divide by the number of independent observations you have (*n* for the population variance; *n*–1 for the sample variance; or, in general, the number of *degrees of freedom*) to get the variance. See the Background Knowledge handouts at www.pobox.com / ~rudolf / psychology.
- **Within-subjects** (factor or covariate). See also *repeated measures*. If a score is obtained for every subject at each level of an independent variable, the independent variable is called a within-subjects factor. See also *between-subjects.* The advantage of a within-subjects design is that the different treatment conditions are automatically matched on many irrelevant variables — all those that

are relatively unchanging characteristics of the subject (e.g. intelligence, age). However, the design requires that each subject is tested several times, under different treatment conditions. Care must be taken to avoid *order*, *sequence* or *carryover* effects — such as the subject getting better through practice, worse through fatigue, drug hangovers, and so on. If the effect of a treatment is permanent, it is not possible to use a within-subjects design. You could not, for example, use a within-subjects design to study the effects of parachutes (versus no parachute) on mortality rates after falling out of a plane.

10 Further reading

- A very good statistics textbook for psychology is Howell (1997).
- Abelson (1995) examines statistics as an technique of argument and is very clear on the logical principles and some of the philosophy of statistics.
- Keppel (1991) is a fairly hefty tome on ANOVA techniques. Winer (1991) is another monster reference book. Neither are for the faint-hearted.
- Myers & Well (1995) is another excellent one. Less fluffy than Howell (1997) but deals with the issues head on.

There is also an excellent series of Statistics Notes published by the British Medical Journal, mostly by Bland and Altman. A list is available at

www.mbland.sghms.ac.uk / pbstnote.htm

and the articles themselves are available online from

www.bmj.com

This series includes the following:

- The problem of the 'unit of analysis' (Altman & Bland, 1997). Correlation and regression when repeated measurements are taken, and the problem of pseudoreplication (Bland & Altman, 1994a). The approach one should take to measure correlation within subjects (Bland & Altman, 1995a) and correlation between subjects (Bland & Altman, 1995b).
- Why correlation is utterly inappropriate for assessing whether two ways of measuring something agree (Bland & Altman, 1986).
- Generalization and extrapolation (Altman & Bland, 1998).
- Why to randomize (Altman & Bland, 1999b), how to randomize (Altman & Bland, 1999a), and how to match subjects to different experimental groups (Bland & Altman, 1994b).
- Blinding (Day & Altman, 2000; Altman & Schulz, 2001).
- Absence of evidence is not evidence of absence — about power (Altman & Bland, 1995).
- Multiple significance tests: the problem (Bland & Altman, 1995c).
- Regression to the mean (Bland & Altman, 1994e; Bland & Altman, 1994d).
- One-tailed and two-tailed significance tests (Bland & Altman, 1994c).
- Transforming data (Bland & Altman, 1996b) and how to calculate confidence intervals with transformed data (Bland & Altman, 1996c; Bland & Altman, 1996a).
- ANOVA, briefly (Altman & Bland, 1996), and the analysis of interaction effects (Altman & Matthews, 1996; Matthews & Altman, 1996a; Matthews & Altman, 1996b).
- Comparing estimates derived from separate analyses (Altman & Bland, 2003).
- Dealing with differences in baseline by ANCOVA (Vickers & Altman, 2001).

Finally, there's an excellent on-line textbook (StatSoft, 2002):

www.statsoft.nl / textbook / stathome.html

11 Bibliography

Abelson, R. P. (1995). *Statistics As Principled Argument*, Lawrence Erlbaum, Hillsdale, New Jersey.

Altman, D. G. & Bland, J. M. (1995). Absence of evidence is not evidence of absence. *British Medical Journal* **311:** 485.

Altman, D. G. & Bland, J. M. (1996). Comparing several groups using analysis of variance. *British Medical Journal* **312:** 1472-1473.

Altman, D. G. & Bland, J. M. (1997). Statistics notes. Units of analysis. *British Medical Journal* **314:** 1874.

Altman, D. G. & Bland, J. M. (1998). Generalisation and extrapolation. *British Medical Journal* **317:** 409-410.

Altman, D. G. & Bland, J. M. (1999a). How to randomise. *British Medical Journal* **319:** 703-704.

Altman, D. G. & Bland, J. M. (1999b). Statistics notes. Treatment allocation in controlled trials: why randomise? *British Medical Journal* **318:** 1209.

Altman, D. G. & Bland, J. M. (2003). Interaction revisited: the difference between two estimates. *British Medical Journal* **326:** 219.

Altman, D. G. & Matthews, J. N. (1996). Statistics notes. Interaction 1: Heterogeneity of effects. *British Medical Journal* **313:** 486.

Altman, D. G. & Schulz, K. F. (2001). Statistics notes: Concealing treatment allocation in randomised trials. *British Medical Journal* **323:** 446-447.

Bland, J. M. & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **i:** 307-310.

Bland, J. M. & Altman, D. G. (1994a). Correlation, regression, and repeated data. *British Medical Journal* **308:** 896.

Bland, J. M. & Altman, D. G. (1994b). Matching. *British Medical Journal* **309:** 1128.

Bland, J. M. & Altman, D. G. (1994c). One and two sided tests of significance. *British Medical Journal* **309:** 248.

Bland, J. M. & Altman, D. G. (1994d). Regression towards the mean. *British Medical Journal* **308:** 1499.

Bland, J. M. & Altman, D. G. (1994e). Some examples of regression towards the mean. *British Medical Journal* **309:** 780.

Bland, J. M. & Altman, D. G. (1995a). Calculating correlation coefficients with repeated observations: Part 1--Correlation within subjects. *British Medical Journal* **310:** 446.

Bland, J. M. & Altman, D. G. (1995b). Calculating correlation coefficients with repeated observations: Part 2--Correlation between subjects. *British Medical Journal* **310:** 633.

Bland, J. M. & Altman, D. G. (1995c). Multiple significance tests: the Bonferroni method. *British Medical Journal* **310:** 170.

Bland, J. M. & Altman, D. G. (1996a). Transformations, means, and confidence intervals. *British Medical Journal* **312:** 1079.

Bland, J. M. & Altman, D. G. (1996b). Transforming data. *British Medical Journal* **312:** 770.

Bland, J. M. & Altman, D. G. (1996c). The use of transformation when comparing two means. *British Medical Journal* **312:** 1153.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effect of inequality of variance and of correlation of errors in the two-way classification. *Annals of Mathematical Statistics* **25:** 484-498.

Boyd, O., Mackay, C. J., Lamb, G., Bland, J. M., Grounds, R. M. & Bennett, E. D. (1993). Comparison of clinical information gained from routine blood-gas analysis and from gastric tonometry for intramural pH. *Lancet* **341:** 142-146.

Cardinal, R. N., Parkinson, J. A., Djafari Marbini, H., Toner, A. J., Bussey, T. J., Robbins, T. W. & Everitt, B. J. (2003). Role of the anterior cingulate cortex in the control over behaviour by Pavlovian conditioned stimuli in rats. *Behavioral Neuroscience* **117:** 566-587.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. First edition, Academic Press, New York.

Day, S. J. & Altman, D. G. (2000). Statistics notes: blinding in clinical trials and other studies. *British Medical Journal* **321:** 504.

Field, A. P. (1998). A bluffer's guide to sphericity. *Newsletter of the Mathematical, Statistical and computing section of the British Psychological Society* **6:** 13-22.

Frank, H. & Althoen, S. C. (1994). *Statistics: Concepts and Applications*, Cambridge, Cambridge University Press.

Greenhouse, S. W. & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika* **24:** 95-112.

Howell, D. C. (1997). *Statistical Methods for Psychology*. Fourth edition, Wadsworth, Belmont, California.

Huynh, H. & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measures designs have exact *F*-distributions. *Journal of the American Statistical Association* **65:** 1582-1589.

Keppel, G. (1982). *Design and analysis: a researcher's handbook*. Second edition, Englewood Cliffs: Prentice-Hall, London.

Keppel, G. (1991). *Design and analysis: a researcher's handbook*. Third edition, Prentice-Hall, London.

Levene, H. (1960). Robust tests for the equality of variance. In *Contributions to probability and statistics* (Oklin, I., ed.). Stanford University Press, Palo Alto, California.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association* **62:** 399-402.

Matthews, J. N. & Altman, D. G. (1996a). Interaction 3: How to examine heterogeneity. *British Medical Journal* **313:** 862.

Matthews, J. N. & Altman, D. G. (1996b). Statistics notes. Interaction 2: Compare effect sizes not P values. *British Medical Journal* **313:** 808.

Mauchly, J. W. (1940). Significance test for sphericity of a normal *n*-variate distribution. *Annals of Mathematical Statistics* **11:** 204-209.

Myers, J. L. & Well, A. D. (1995). *Research Design and Statistical Analysis*, Lawrence Erlbaum, Hillsdale, New Jersey.

Prescott, C. E., Kabzems, R. & Zabek, L. M. (1999). Effects of fertilization on decomposition rate of Populus tremuloides foliar litter in a boreal forest. *Canadian Journal of Forest Research* **29:** 393-397.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in C*. Second edition, Cambridge University Press, Cambridge, UK.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin* **2:** 110–114.

Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52:** 591-611.

SPSS (2001). SPSS 11.0 Syntax Reference Guide (spssbase.pdf).

StatSoft (2002). *Electronic Statistics Textbook* (http://www.statsoft.com/textbook/stathome.html), Tulsa, OK.

Tangren, J. (2002). A Field Guide To Experimental Designs (http://www.tfrec.wsu.edu/ANOVA/, 2004). Washington State University, Tree Fruit Research and Extension Center.

Vickers, A. J. & Altman, D. G. (2001). Statistics notes: Analysing controlled trials with baseline and follow up measurements. *British Medical Journal* **323:** 1123-1124.

Winer, B. J. (1971). *Statistical principles in experimental design*. Second edition, McGraw-Hill, New York.

Winer, B. J., Brown, D. R. & Michels, K. M. (1991). *Statistical Principles in Experimental Design*, McGraw-Hill, New York, NY.