NST 1B Experimental Psychology
Statistics practical 4
χ^2

Rudolf Cardinal & Mike Aitken
2 / 3 March 2004; Department of Experimental Psychology
University of Cambridge

Handouts (!):

- Answers to Examples 4 (from last time)
- Handout 5 (χ^2)
- Examples 5 (χ^2)
- Examples 6 (past paper questions)
- Examples 7 (mixed)

pobox.com / ~rudolf / psychology

These slides are on the web.
No need to scribble frantically.

pobox.com / ~rudolf / psychology

The χ^2 test: for categorical data

(1) "Goodness of fit" test

Categorical data

100 people choose between chocolate and garibaldi biscuits.
Every person falls into one of **two categories:** chocolate or garibaldi.
This is **categorical data.**

If people choose at chance, we'd expect a 50:50 split.
— **EXPECTED values** under the hypothesis 'choose at chance'.

Suppose 65 choose chocolate and 35 choose garibaldi.
— **OBSERVED values.**

Do the observed values differ significantly from the expected values? Are the data (*O*) a *good fit* to the 'model' (*E*)?

Null hypothesis: observed values do *not* differ from the expected values (the model *is* a good fit to the data).

χ^2 test with 1 categorical variable, 2 categories

100 people choose between chocolate and garibaldi biscuits

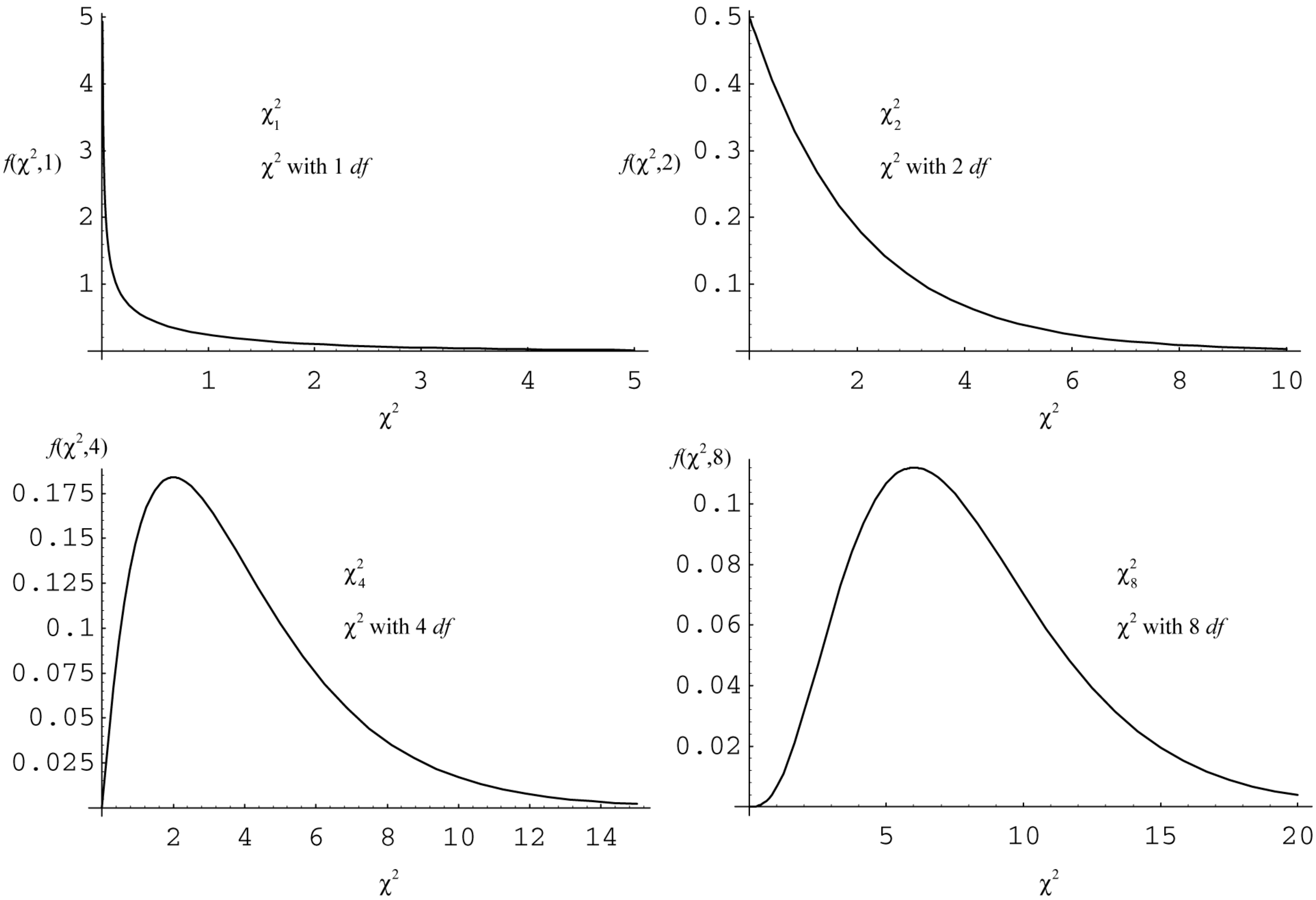
Expected (*E*): 50 chocolate, 50 garibaldi.
Observed (*O*): 65 chocolate, 35 garibaldi.

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

If *O* values are close to *E* values, χ^2 is small (close to zero).
If *O* values are very far from *E* values, χ^2 is big.
If χ^2 is big enough, we will reject the null hypothesis.
For our biscuit example:

$$\chi^2 = \frac{(65-50)^2}{50} + \frac{(35-50)^2}{50} = 9$$

Distribution of χ^2 depends on number of degrees of freedom



Degrees of freedom for a goodness-of-fit χ^2 test

100 people choose between chocolate and garibaldi biscuits.
Observed (*O*): 65 chocolate, 35 garibaldi.
Expected (*E*): 50 chocolate, 50 garibaldi.

$n = 100$
We have made sure that the expected values add up to the same n as the observed values. Therefore, we lose one *df*.

_df_ = categories – 1

For our biscuit example: two categories (chocolate, garibaldi), so **1 _df_.** We could write

$$\chi_1^2 = \frac{(65-50)^2}{50} + \frac{(35-50)^2}{50} = 9$$

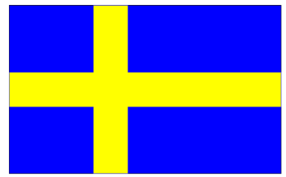
Critical value of χ^2 for 1 *df* and $\alpha = 0.05$ is 3.84 (from tables). Ours is larger. So we **reject** the null hypothesis (the **model, _E_,** of a 50:50 split is not a good fit to the data); preferences differed from chance.

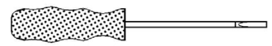
χ^2 test with 1 categorical variable, >2 categories

Suppose an Ikea factory makes only rough-hewn pine chair backs, chair seats, and chair legs. We **sample** 50 items at random from the thousands in the warehouse. If they are making items in the correct ratio, there should be one back and one seat for every 4 legs.

Expected (*E*): 8.33 backs, 8.33 seats, 33.33 legs.

Observed (*O*): 10 backs, 6 seats, 34 legs.





As always,

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Here,

$$\chi^2 = \frac{(10-8.33)^2}{8.33} + \frac{(6-8.33)^2}{8.33} + \frac{(34-33.33)^2}{33.33} = 1$$

df = categories – 1 = 3 – 1 = 2. Critical value of χ^2 for 2 *df* and α = 0.05 is 5.99 (from tables). Ours is **not** larger. So we **don't** reject the 1:1:4 model; the factory's doing OK.

Aside: is the χ^2 test one-tailed or two-tailed?

We always test to see if our χ^2 is *bigger* than a critical value (χ^2 is never negative). So the process of testing it is one-tailed.

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

However, the way we **calculate** χ^2 is inherently two-tailed; χ^2 will get larger whether $O > E$ or $E > O$. So a χ^2 test always performs a **two-tailed** test on our data. The α in the χ^2 tables is therefore effectively a two-tailed α.

Q. Where does the χ^2 test come from?

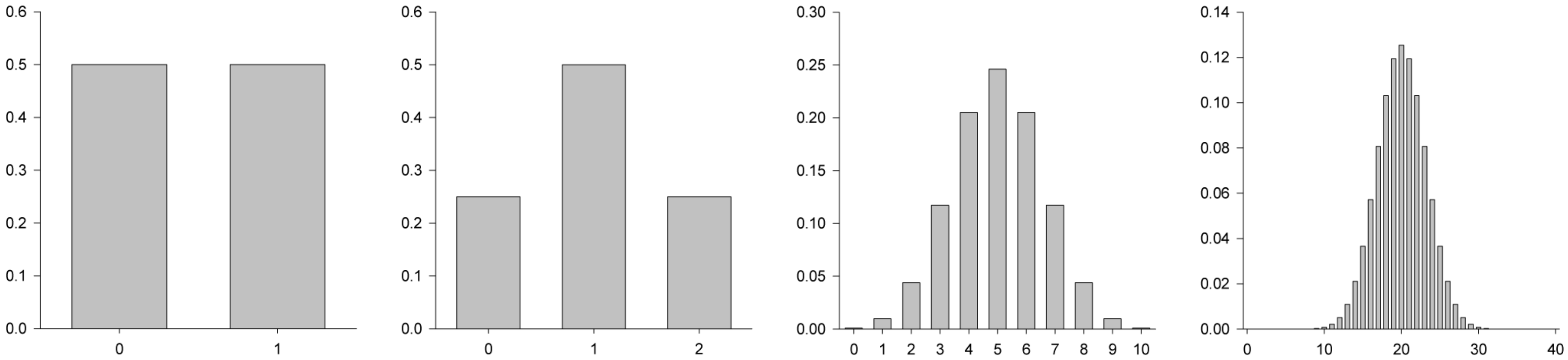
A. It's in the handout, if you're interested. You don't need to know. We'll glance at it briefly.

The binomial distribution: prelude to χ^2

• **Multiple events (trials). One of two** things can happen on each event (trial).

• Example: flip a coin *n* times. Possible outcomes on each trial: heads or tails.
• Probability of heads = *p*. Probability of tails = *q* = 1 – *p*.
• If the coin is fair, *p* = *q* = 0.5

• How many heads would we expect to see in *n* coin flips? How likely are we to see 7 heads in 20 flips if the coin is fair? These questions are answered by the **binomial distribution.** (Details in the handout; you don't need to know.)
• As *n* increases, the binomial distribution starts to look like the **normal** distribution. (Below: distribution of total number of heads in 1, 2, 10, and 40 coin flips.)



Logic of χ^2: only for the mathematically inclined!

2 categories, e.g. *n* coin flips

- *n* independent events
- Each event can fall into 2 possible categories (H = heads, T = tails).
- Null hypothesis: the probability that any event falls into category H is *p*; probability of category T is *q* (= 1 – *p*).
- Under null hypothesis, can calculate probability of a certain number of H events and a certain number of T events using the **binomial** distribution.
- For large *n*, easier to use the **normal** approximation to the binomial distribution.
- Likelihood can therefore be described by a single *z* score. Square to ensure ≥0: get z^2.
- $\chi_1^2 = z^2$
- So χ^2 with 1 *df* is the normal approximation to the binomial distribution.

>2 categories, e.g. *n* die rolls

- *n* independent events
- *k* possible categories (A, B, C, ...).
- Null hypothesis: $P(\mathrm{A}) = p$; $P(\mathrm{B}) = q$; $P(\mathrm{C}) = r$ etc.; $p + q + r + \ldots = 1$.
- Under null hypothesis, can calculate probability of a certain number of A / B / C / ... events using the **multinomial** distribution.
- Approximate multinomial with *k*–1 **different** normal distributions.
- Square and add. Likelihood is therefore described by a Σz^2 score.
- $\chi_2^2 = z_A^2 + z_B^2$ and $\chi_{k-1}^2 = \sum_{i=1}^{k-1} z_i^2$
- So χ^2 with *k*–1 *df* is the normal approximation to the multinomial.

Applying χ^2 to your data

GOODNESS-OF-FIT EXAMPLE FROM PRACTICAL (1)

A very simple example indeed…

Reasoning practical.

In Group A, 51 subjects attempted the *Missionaries & Cannibals* problem. Of those, **9** subjects solved it in under ten minutes, and **42** didn't. Suppose we had reason to believe that ten minutes was the median time to solve this problem (i.e. that half would solve faster, and half slower or not at all). Are your data consistent with this hypothesis?

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$df = \text{categories} - 1$$

GOODNESS-OF-FIT EXAMPLE FROM PRACTICAL (2)

In Group A, 51 subjects attempted the *M&C* problem. Of those, **9** subjects solved it in under ten minutes, and **42** didn't. Suppose we had reason to believe that ten minutes was the median time to solve this problem (i.e. that half would solve faster, and half slower or not at all). Are your data consistent with this hypothesis?

Observed: 9 fast, 42 slow.
Expected under hypothesis: 25.5 fast, 25.5 slow.

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \left(\frac{(9-25.5)^2}{25.5} \right) + \left(\frac{(42-25.5)^2}{25.5} \right)$$

$$= 10.676 + 10.676$$

$$= 21.35$$

$$df = 2 - 1 = 1$$

For 1 *df*, critical value of χ^2 for $\alpha = 0.05$ is 3.84.

Our χ^2 is larger: **reject** hypothesis.

The χ^2 test: for categorical data

(2) "Contingency" test

χ^2 test with 2 categorical variables (contingency test) — 1

Now suppose we have *two* categorical variables. Pugh (1983) examined the decisions of US juries in 358 rape trials. Each trial could be classified according to two categorical variables:

- was the defendant found **guilty** or **not guilty?**
- did the defence allege the victim was **at fault** or **not?**

In the UK, jury research is restricted by the Contempt of Court Act 1981 and this study would have been illegal.

Obtained values	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	153	24	*177*
Victim portrayed as high-fault	105	76	*181*
Total	*258*	*100*	*358*

Did the two variables influence each other?

Was there a *contingency* between them?

Was the conviction rate different for 'low-fault' and 'high-fault'?

This table is called a *contingency table.*

χ^2 test with 2 categorical variables (contingency test) — 2

Obtained values	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	153	24	*177*
Victim portrayed as high-fault	105	76	*181*
Total	*258*	*100*	*358*

If the two variables **did not** influence each other **(null hypothesis),** what values (*E*) would we expect? **Not this:**

WRONG expected values	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	89.5	89.5	*179*
Victim portrayed as high-fault	89.5	89.5	*179*
Total	*179*	*179*	*358*

We know that, **overall,**

- the victim was portrayed as low-fault in 49% (177 / 358) of cases
- the defendant was found guilty in 72% of cases (258 / 358).

So our expected values should keep those proportions but have no *interrelationship (contingency)* between the two variables...

χ^2 test with 2 categorical variables (contingency test) — 3

Obtained values	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	153	24	*177*
Victim portrayed as high-fault	105	76	*181*
Total	*258*	*100*	*358*

So our expected values should look like this:

Expected values	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	127.559	49.441	*177*
Victim portrayed as high-fault	130.441	50.559	*181*
Total	*258*	*100*	*358*

The row and column totals are now the same as before, so the ‘guilty’ proportion and the ‘low fault’ proportion are the same, but there is **no contingency:** 72% of defendants in ‘low-fault’ cases are convicted, and so are 72% of defendants in ‘high-fault’ cases (etc.).

$$E(row_i, column_j) = \frac{R_i C_j}{n}$$

χ^2 test with 2 categorical variables (contingency test) — 4

Now we can calculate χ^2 exactly as before:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Obtained values	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	153	24	*177*
Victim portrayed as high-fault	105	76	*181*
Total	*258*	*100*	*358*

Expected values	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	127.559	49.441	*177*
Victim portrayed as high-fault	130.441	50.559	*181*
Total	*258*	*100*	*358*

$$\chi^2 = \frac{(153-127.559)^2}{127.559} + \frac{(105-130.441)^2}{130.441} + \frac{(24-49.441)^2}{49.441} + \frac{(76-50.559)^2}{50.559}$$

$$= 35.9$$

We have made *E* agree with *O* as to (1) *n*, (2) proportion low-fault, and (3) proportion guilty, so we have lost 3 *df* and only have 1 left.

In general, for a contingency table,

df = (rows – 1) × (columns – 1)

χ^2 test with 2 categorical variables (contingency test) — 5

So we know $\chi^2 = 35.9$ and $df = 1$.
Critical value of χ^2 for 1 df and $\alpha = 0.05$ is 3.84 (from tables).
Ours is larger. In fact, $p < 0.001$ (critical value 10.83 for $\alpha = 0.001$).
So we **reject** the null hypothesis (the **model, *E*,** is not a good fit to the data).

There *was* a relationship between the defence's portrayal of the victim and the conviction rate; the defendant was less likely to be convicted if the defence portrayed the victim as being at fault (58% convicted = 105 / 181) than if they didn't (86% = 153 / 177).

Obtained values	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	153	24	*177*
Victim portrayed as high-fault	105	76	*181*
Total	*258*	*100*	*358*

Expected values	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	127.559	49.441	*177*
Victim portrayed as high-fault	130.441	50.559	*181*
Total	*258*	*100*	*358*

CONTINGENCY EXAMPLE FROM PRACTICAL (1)

Reasoning practical again.

In Group A (who tried *Missionaries & Cannibals*, then the counter-moving problem), 51 subjects attempted the *M&C* problem. Of those, **9** subjects solved it in under ten minutes, and **42** didn't.

In Group B (who did counter-moving before *M&C*), another 51 subjects attempted the problem. **36** solved it fast, and **15** didn't.

Did the proportion of people solving the *M&C* problem fast differ across groups?

$$E(row_i, column_j) = \frac{R_i C_j}{n} \qquad \chi^2 = \sum \frac{(O-E)^2}{E}$$

$$\boldsymbol{df = (\text{rows} - 1) \times (\text{columns} - 1)}$$

CONTINGENCY EXAMPLE FROM PRACTICAL (2)

Obtained values	Solved fast	Slow or not at all	Total
Group A	9	42	*51*
Group B	36	15	*51*
Total	*45*	*57*	*102*

Expected values	Solved fast	Slow or not at all	Total
Group A	22.5	28.5	*51*
Group B	22.5	28.5	*51*
Total	*45*	*57*	*102*

$$E(row_i, column_j) = \frac{R_i C_j}{n}$$

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \left(\frac{(9-22.5)^2}{22.5}\right) + \left(\frac{(42-28.5)^2}{28.5}\right) + \left(\frac{(36-22.5)^2}{22.5}\right) + \left(\frac{(15-28.5)^2}{28.5}\right)$$

$$= 8.1 + 6.395 + 8.1 + 6.395$$

$$= 28.99$$

df **= (rows – 1) × (columns – 1) = 1 × 1 = 1**

For 1 *df*, critical value of χ^2 for $\alpha = 0.05$ is 3.84. **Reject** null hypothesis.

Assumptions of the χ^2 test: IMPORTANT

The χ^2 test is simple to use, but it is perhaps the most commonly misused statistical test. There are several ways to cock up. The test assumes:

• **Independence of observations.** In the examples so far, one person's chocolate / garibaldi choice didn't affect another's; one court case didn't affect another. If this isn't true, can't use a χ^2 test.

  • **Mustn't** analyse data from several subjects when there are multiple observations per subject. Need one observation per subject. **Most common cock-up?**

  • Can analyse data from *only* one subject — then all observations are equally independent — but conclusions only apply to that subject.

• **Normality.** There shouldn't be any very small expected frequencies, or the data won't approximate a normal distribution (required by the underlying maths). Rule of thumb: **no *E* value less than 5.** (Possible to go <5 under some special circumstances — see handout — but can never have an *E* value of 0 or you can't calculate χ^2!)

• **Inclusion of non-occurrences.** (See next slide.)

Include non-occurrences!

Suppose we ask 20 men and 20 women whether they supported the sale of alcohol in petrol stations. 17 men say yes; 11 women say yes. Do men's preferences differ from women's?

This is **wrong;** it omits information about nay-sayers.

Obtained values	Men	Women
Support booze	17	11

$\chi^2 = 1.29, NS$ ✗

This is correct:

Obtained values	Men	Women
Yes to booze	17	11
No	3	9

$\chi^2 = 4.29, p < 0.05$ ✓

Should be easy to understand. The first (incorrect) table could equally represent this situation, which represents a completely different pattern of male / female preference:

Obtained values	Men	Women
Yes to booze	17	11
No	1983	9

Don't analyse proportions! Analyse the actual numbers.

Obtained values	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	153	24	*177*
Victim portrayed as high-fault	105	76	*181*
Total	*258*	*100*	*358*

✓ $\chi^2 = 35.9$ ✓

↓ *aargh...*

WRONG: Observed values as proportions of total	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	42.7%	6.7%	*49.4%*
Victim portrayed as high-fault	29.3%	21.2%	*50.6%*
Total	*72.1%*	*27.9%*	*100%*

✗

WRONG: Expected values as proportions of total	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	35.6%	13.8%	*49.4%*
Victim portrayed as high-fault	36.5%	14.1%	*50.6%*
Total	*72.1%*	*27.9%*	*100%*

✗

$\chi^2 = 10.1$ ✗

Analysing percentages pretends that you had exactly $n = 100$ events. Unless you did, the answer will be wrong. If $n > 100$, your calculated χ^2 will be too low (as here); if $n < 100$, your calculated χ^2 will be too high.