

(Answers calculated by RNC — *caveat emptor*. Thanks to MRFA for catching errors.) In some of these examples I'll quote exact p values from a computer, rather than just saying ' $p < 0.05$ '. If I say ' $p = .03$ ', your tables would show that $p < .05$, but not that $p < .01$. If I say ' $p = .125$ ', your tables would show that the answer is not significant at $p = .1$ (i.e. $p > .1$)... and so on.

- Q1 0.62% ($Z = -2.5$, corresponding to $p = 0.0062$)
- Q2 No (preliminary F test of variances, $F_{14,14} = 1.366$, NS; unpaired t test assuming equal variances, $t_{28} = 1.1397$, NS, exact $p = .264$)
- Q3 (a) 0, -2.5, -5, 5, 1.4
(b) 30, 70, 69.5, 24.2, 66.5, 46.7
(c) 137
(d) 0
(e) 57
- Q4 No; $\chi^2 = 2.134$, NS, $p = .344$. (Note that the expected values for left and right turns are not equal — there might be some bias that promotes left or right turning independent of the rats' breeding. That's why there are several rat groups. Calculate the expected values as usual for a 2×2 contingency table.)
- Q5 Correlating the rank order of the amounts with the rank order of their arrival time (Spearman's correlation coefficient) gives $r_s = -0.732$ ($n = 15$), $p < .01$ two-tailed.
- Q6 (a) mean 45, SD 15
(b) 81.8%
(c) between 34.9 and 55.1 (to 1 dp)
- Q7 Yes (paired t test, $t_{19} = 2.27$, exact $p = .035$).
- Q8 No ($r = -0.19$, $t_{10} = -0.613$, NS).
- Q9 No (preliminary F test to check homogeneity of variances $F_{7,14} = 1.36$, NS; unpaired two-sample t test assuming equal variances $t_{21} = 1.71$, NS, exact $p = .59$).
- Q10 (a) [only 100,000 people above IQ of] 143
(b) [num. people with IQ between 95 and 105] 13,055,860 (should really round this to 3 sf!)
(c) [... between 60 and 70] 945,982 (should really round this to 3 sf!)
- Q11 Yes. A preliminary F test gives $F_{9,11} = 2.982$, two-tailed $p = .09$. Whether you decide this is OK and proceed to an equal-variances-assumed unpaired t test ($t_{20} = 2.7999$, $p = .011$ two-tailed), or prefer a t test not based on the equal-variances assumption ($t'_9 = 2.6696$, $p = .026$ two-tailed) — which is probably better (most people use two-tailed $\alpha = 0.1$ for preliminary F tests) — you still find a significant difference even with two-tailed t tests.
- Q12 Yes! $\chi^2_{12} = 78.72$, $p = 7.2 \times 10^{-12}$.
This test assumes equal independence of observations (that no student got more than one degree, for example), normality (no expected values very small and no dramatic skew of row/column values — this looks OK), and inclusion of non-occurrences (that we've included all the students and universities that we studied). The first of these is the one that we'd have to be most careful about checking!
- Q13 No significant difference; Mann–Whitney $U_{7,13} = 44$, NS.
- Q14 Yes; $\chi^2_4 = 15.52$, $p = .0037$.
- Q15 Rearrange the data into a 2×2 contingency table: {scratching/not scratching} versus {tooth-baring/not tooth-baring}. Then we find $\chi^2_1 = 6.88$, $p = .0087$, so yes, the two are 'connected'. The way they're connected seems to be that they're unlikely to occur together!
- Q16 Three examples of (potentially) relevant tests:
(a) Are husbands' preferences correlated with those of their wives? No ($r = 0.310$, $t_{24} = 1.60$, NS).
(b) Do the husbands' preferences vary more or less than the wives'? There is a significant difference in variance ($F_{25,25} = 6.29$, $p < .001$) — husbands' preferences vary more.
(c) Are husbands' preferences (mean 2.61) significantly different from their wives' (mean 2.42)? No (paired t test, $t_{25} = 0.385$, NS).
(d) Is there a relationship between sex and number of children? Perhaps I should rephrase that... is there a relationship between gender and desired number of children? Yes. Categorize people by gender and by

number of children desired (e.g. '0', '1', '2', '3', '4', 'more than 4') and perform a contingency χ^2 test. If you use these categories, you find $\chi^2_5 = 15.81, p = .0074$. However, this approach assumes independence of husbands' and wives' preferences, which is perhaps questionable (the answer to (a) just tells you there's no overall linear correlation, which isn't the same thing).

Thanks to MRFA for (d), which I didn't think of. You could also apply nonparametric versions of some of the above (simply a different strategy — these additional tests don't provide new information, so running a parametric and a nonparametric version of the same test wouldn't 'count' as two tests). The only rather artificial extra question I can come up with is to use the order information, e.g. 'is there a relationship between the mean number of children preferred by a couple and the order in which the couples were asked?' (no; correlation between the couple order and the rank order of the couple's mean preference, $r_s = -0.326, p > .1$ two-tailed).

Q17 Yes; $t_9 = 3.21, p = .011$ by paired t test.

Q18 Not quite: correlation $r = -0.507, t_{13} = -2.122, p = .054$.

Q19 No, unless you were in a particularly suspicious mood: $\chi^2_5 = 10.36, p = .066$.

Q20 Yes; Wilcoxon signed-rank $T_{15} = 21, p < .05$ two-tailed.

Q21 Yes. You have to calculate 'hits' from what you know (misses and totals). Then $\chi^2_3 = 17.27, p = .00062$.

Q22 A bit nasty, this, and **way** beyond exam standard. You'd need to calculate *which* normal distribution *best* fits the data, and then see whether there's a significant difference between the actual values and the values predicted by this normal distribution. You could do it like this:

(a) Take the dioptre measurement as your variable. So you have 11 observations with value -5 , 29 observations with value -4 ... up to 5 observations with value $+5$. There are a total of $n = 1000$ observations. The sum of the observations is $(11 \times -5) + (29 \times -4) + \dots + (5 \times +5)$: we have $\sum x = -174$. So the mean is

$$\bar{x} = \frac{\sum x}{n} = \frac{-174}{1000} = -0.174. \text{ Now we need the SD:}$$

$$\sum x^2 = (11 \times (-5)^2) + (29 \times (-4)^2) + \dots + (5 \times 5^2) = 2322$$

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{2322 - \frac{(-174)^2}{1000}}{999}} = 1.515$$

(b) The second phase is to determine the expected values for a normally-distributed variable with a mean of -0.174 and an SD of 1.515 if we measured them in a categorical way. It would be reasonable to suppose that the '0D' category includes people measuring -0.5 to $+0.5D$; the '1D' category includes people from $+0.5$ to $+1.5D$, and so on. So we can calculate Z scores for each boundary between dioptre categories, work out the proportion we expect to find in each category, and multiple by 1000 to get the number of observations we expect. That gives us expected values like this:

-5D	-4D	-3D	-2D	-1D	0D	+1D	+2D	+3D	+4D	+5D
1.93	11.90	48.27	128.35	224.13	257.05	193.63	95.79	31.10	6.62	0.92

(c) Finally, we run a χ^2 test with these as the expected values. Since we have made the observed and expected values agree as to n , \bar{x} , and s_x , we have lost *three* degrees of freedom (rather than a 'conventional' χ^2 test, in which we make them agree only as to n , and lose only one df). So $df = k - 3$ (where k is the number of categories) = 8 . We find that $\chi^2_8 = 534.4, p = 2.87 \times 10^{-110}$, which we could call 'very significant indeed'!

The lens powers *do* deviate significantly from a normal distribution.

Out of interest, here are the histograms of the observed values and corresponding expected values based on a normal distribution having the same mean and standard deviation. (Trivia: the posh name for this type of deviation from normality — too peaked in the middle — is *leptokurtic*.)

