Reporting the results of statistical analyses

- Generally speaking you should state exactly what statistical test was conducted and describe what the results showed in terms of the variables of interest.

- The actual test statistics are usually reported in brackets at an appropriate place within that sentence. The general form is something like the following:

  (test statistic = value, degrees of freedom (if appropriate), n = value (if appropriate), *p* = value, one- or two-tailed (if appropriate)

- Note that you should always note whether the results of the analyses support your hypotheses or not.

- Note that the numbers used in these examples are made up and are for the purposes of illustration only.

- Note that the wording of the interpretation of the test result is only a guide and it is not intended that you should always copy it verbatim.

Test for differences between two groups/conditions

With tests for differences you should mention the **direction** of any significant differences between groups or conditions. The results of the non-parametric equivalents of the t-tests (i.e., Mann-Whitney and Wilcoxon sign rank) should be reported in much the same way but with the appropriate test statistic substituted. Note however, that the non-parametric tests refer to differences in medians rather than means.

Unrelated (aka independent samples, between-subjects) t-test

As predicted, an unrelated t-test revealed that participants in the noise condition recalled significantly fewer words than participants in the no noise condition ($t = 6.14$, df = 15, p = 0.003, one-tailed).

Related (aka paired-samples, within-subjects, repeated measures) t-test

Contrary to our hypothesis, a related t-test showed that there was no significant difference between participants' reaction times for the caffeine and no caffeine conditions ($t = 0.93$, p = 0.54, one-tailed).

Test for differences between two or more groups/conditions

With tests for differences you should mention the **direction** of any significant differences between groups or conditions that are discovered as a result of conducting further **multiple comparison** tests. You should point out whether these were **planned** (a priori) or **unplanned** (post hoc).

The results of the non-parametric equivalents of the one-way ANOVAs (i.e., Kruskal-Wallis and Friedman) should be reported in much the same way but with the appropriate test statistic

substituted. Note however, that the non-parametric tests refer to differences in medians rather than means.

Multiple comparison tests for the non-parametric equivalents of ANOVA can be conducted using the non-parametric equivalents of the unrelated and paired-samples t-tests (i.e. Mann-Whitney, Wilcoxon sign rank).

One-way unrelated ANOVA

A one-way unrelated ANOVA indicated that, as predicted, the level of processing had a significant effect on word recall ($F(2,25) = 6.54$, $p = 0.002$). Planned Bonferroni multiple comparison tests found that the mean number of words recalled in the Category condition was significantly greater than the Rhyming condition ($p = 0.03$, one-tailed) which was in turn significantly greater than the Capitals condition. ($p = 0.008$, one-tailed).

The amount of information that can be reported for the results of multiple comparison tests will depend on the type of test(s) used. As a minimum you should report the p value associated with the difference.

One-way repeated measures ANOVA

One-way repeated ANOVA results are reported in much the same way as the unrelated ANOVA results. However, note that because SPSS does not allow you to carry out multiple comparisons in the same way, so you will need to do a series of paired-sample t-tests if you wish to conduct multiple comparisons, regardless of whether they are planned or unplanned. You should correct for multiple testing by dividing the usual criterion for significance ($p < 0.05$) by the number of multiple comparisons you are carrying out. This revised p value should be used as the criterion for statistical significance for each multiple comparison result.

A one-way repeated measures ANOVA indicated that, as predicted, the type of stimuli had a significant effect on participant reaction times ($F(2,14) = 5.89$, $p = 0.002$). Post hoc multiple comparison paired-samples t-tests found that the mean reaction times for both the Incongruent ($t = 3.69$, $df = 14$, $p = 0.003$, two-tailed) and Rhyming conditions ($t = 2.93$, $df = 14$, $p = 0.009$, two-tailed) were significantly greater than for the Baseline condition. There was no significant difference between the mean reaction times for the Incongruent and Rhyming conditions ($t = 1.36$, $df = 14$, $p = 0.23$, two-tailed).

Factorial ANOVA

A 2 x 2 unrelated factorial ANOVA revealed that, as predicted, there was no significant main effect of paper type ($F(1,44) = 132.44$, $p = 0.34$) and no significant main effect of paranormal belief ($F(1,44) = 156.89$, $p = 0.67$) on the quality rating assigned to the experimental report. However, as predicted, there was a significant paper type x paranormal belief interaction effect ($F(1,44) = 78.44$, $p = 0.03$). The results of planned unrelated t-tests revealed two significant simple effects: believers gave significantly higher ratings than non-believers on the pro-ESP report ($t = 2.54$, $df = 22$, $p = 0.03$, one-tailed); non-believers gave significantly higher ratings than believers on the anti-ESP report ($t = 2.89$, $df = 22$, $p = 0.02$, one-tailed)

Two-way repeated measures and mixed factorial ANOVAs should be reported in much the same way.

Measures of association

Chi-square

A 2 x 2 chi-square test revealed that, as predicted, there was a significant association between smoking and drinking ($\chi^2 = 12.12$, df = 1, p = 0.006); drinkers are more likely to be smokers than non-drinkers.

Correlation between two variables

With correlations you should mention the **magnitude** and **direction** of any relationship.

Pearson product-moment correlation

As predicted, the Pearson's product-moment correlation test showed that there was a significant moderate positive correlation (r = 0.56, n = 20, p = 0.003, one-tailed) between environmental temperature and ice-cream sales.

Spearman's rho correlation

Contrary to hypothesis one, the Spearman's rho correlation test showed that there was no significant relationship between environmental temperature and ice-cream sales; the effect size was almost zero and negative ($r_s = - 0.04$, n = 20, p = 0.65, two-tailed).

There's also a very useful chapter in a new edition of a statistics for psychology textbook:

Howitt, D., & Cramer, D. (2000). Reporting significance levels succinctly. In *An introduction to statistics in psychology: A complete guide for students* (2nd ed.) (pp. 167-171). Harlow, England: Prentice Hall.

Dr Simon Sherwood
11th Feb 2002