Reporting Results of Common Descriptive and Inferential Statistics in APA Format

The Results sections of any manuscript (APA format or non-APA format) are intended to report the quantitative and sometimes qualitative results of descriptive summaries and inferential tests performed to support or refute a hypothesis. Results sections should not be discussions of the relevance of the empirical findings and should avoid interpreting the results with respect to the hypothesis or current theories; that's what the Discussion section is for. Thus, the Results section should avoid describing the outcome of a study as "good" or "bad," "consistent" or "inconsistent" with a hypothesis, "unexpected" or "expected," and "interesting" or "uninteresting." Instead the Results section should present the inferential analyses and descriptive statistics in as concise a way as possible so the reader knows exactly what evidence is being used to support or to refute a hypothesis. It's all about the numbers!

As far as descriptive statistics, you can report them directly within the text of the Results sections (examples are later in the document). However, if your study is complicated, such as a multilevel single-factor experimental design, a factorial design, or a correlational study; it is often good to present your descriptive statistics as a table or in a figure (graph). (See Chapter 5 of the APA Manual.) Remember, a picture's worth a thousand words!

When you report the results of an inferential statistic, you should report:

- the test statistic (*F*-ratio, *t*-value, *z*-score, etc.)
- the degrees of freedom (remember, there are many ways to calculate these)
- the error term (if necessary—non-parametric tests and the correlations usually do not include this)
- and the p-value (a priori or a posteriori alpha level: α)

Importantly, statistical tests included in the results section should relate to the hypothesis being evaluated. For example, you would not include ANOVA if your hypothesis calls for assessing a correlation between variables. When you begin a Results section it is good to state the hypothesis, which is generally done descriptively with words rather than symbolically ($\mu_1 = \mu_0$). After your hypothesis has been stated you can indicate whether the results support or refute the hypothesis and provide the descriptive and inferential statistics as support.

A point on formatting the statistical symbols: The values of the test statistic, means, and error values are generally rounded to two decimal places, whereas the p-values are rounded to two or three decimal places. All statistical symbols that are not Greek letters should be italicized (*M*, *SD*, *t*, *r*, *p*), Greek symbols are not (χ, α, β).

A point on p-values. You can report the p-value as the chosen alpha level that is used for rejecting or retaining the null hypothesis, that is, the a priori criterion for the probability of falsely rejecting a true null hypothesis, which is generally .05 or .01. For example: *t*(28) = 2.99, *SEM* = 10.50, *p* < .05. Alternatively, you may report the exact p-value that is provided after running an inferential test in a software program (SPSS, R, Minitab), that is, the a posteriori probability that the result obtained or one more extreme is consistent with the null hypothesis. For example: *t*(28) = 2.99, *SEM* = 10.50, *p* = .0057.[1] If you do report the a posteriori probability and the value is less than .001, it is customary to report *p* < .001. It is also acceptable to state, up front, the alpha level used as the criterion for statistical significance in all inferential tests. If you do this you do not need to report the p-value for each test. For example: "An alpha level of .01 was used for all statistical tests."

When reporting non-significant results, the p-value is generally reported as the a posteriori probability of the test-statistic. For example: *t*(28) = 1.10, *SEM* = 28.95, *p* = .268.

1 When reporting the a posterior probability you can report the p-value as being equal to the probability, as in the example here, or less than the probability when rounded to three places. In this case, the p-value would be written as *p* < .006.

Examples[2]

Reporting a significant one-sample t-test:

The number of hours spent watching Family Guy per week (*M* = 11.00, *SD* = 2.00) was significantly different from 10.00 hours per week, *t*(49) = 2.36, *SEM* = 0.42, *p* < .05.

Reporting a significant independent-samples (independent-groups) t-test:

Results show that students in the basic study group (*M* = 2.70, *SD* = 0.47) had a significantly lower GPA at the end of their freshman year than students in the group study condition (*M* = 2.93, *SD* = 0.23), *t*(18) = 2.81, *SEM* = 0.081, *p* < .05.

Reporting a significant correlated-samples (dependent groups) t-test:

The results suggested a statistically significant preference for bread with the butter side down (*M* = 6.75, *SD* = 1.45) than bread with the butter-side up (*M* = 3.25, *SD* = 1.10), *t*(19) = 3.15, *SEM* = 1.11, *p* = .005.

Reporting a significant Pearson correlation:

There was a significant correlation between percent of a nation's gross domestic product spent on the military and percentage of a nation's population that believes in the Theory of Evolution, *r*(31) = -0.63, *p* < .001.

Reporting an omnibus oneway ANOVA, with post-hoc tests:

The analysis of variances showed that the effect of group significantly influenced anxiety, *F*(2, 57) = 5.00, *MSE* = 100.25, *p* = .009. Post hoc analyses were conducted using Tukey's post-hoc test. Based on a Tukey's value of *CD* = 2.50, the anxiety in the drug group (*M* = 5.25, *SD* = 1.80) was significantly less than in the placebo group (*M* = 8.35, *SD* = 2.68) and the control group (*M* = 8.10, *SD* = 1.69). The anxiety in the placebo group and the drug group did not differ significantly.

Reporting multiple a priori comparisons:

Tests of three a prior hypotheses were assessed using three independent groups t-Tests with a Bonferroni adjusted alpha level of .016 per test (.05/3). Results suggest that the anxiety that was experienced in the drug group (*M* = 5.25, *SD* = 1.80) was significantly less than the anxiety experienced in the placebo group (*M* = 8.35, *SD* = 2.68), *t*(38) = 3.25, *SEM* = 0.95, *p* = .002, and control group (*M* = 8.10, *SD* = 1.69), *t*(38) = 2.72, *SEM* = 1.05, *p* = .009. However, the placebo group and the control group did not significantly differ in anxiety, *t*(38) = 0.65, *SEM* = 0.38, *p* = .519.

2 Normally, Results sections are double-spaced, the pages have 1" margins, and the section begins with the title 'Results'. For brevity, this formatting is not included here and these examples should be viewed as reporting only the inferential statistics (not the hypotheses).

Reporting a factorial ANOVA, with non-significant interaction:

Test completion times, in minutes, were submitted to a two-way ANOVA with two levels of room noise (noisy, quiet) and two levels of room temperature (hot, cold). The main effect of room noise was significant, $F(1, 39) = 12.65$, $MSE = 101.55$, $p = .001$, suggesting that the test completion time in the noisy group ($M = 3.55$, $SD = 1.20$) was greater than the quiet group ($M = 2.78$, $SD = 0.92$). The main effect of room temperature was significant, $F(1, 39) = 5.85$, $MSE = 101.55$, $p = .02$, suggesting that the test completion time in the cold group ($M = 4.04$, $SD = 1.25$) was greater than the hot group ($M = 3.58$, $SD = 0.99$). The interaction was not significant, $F(1, 39) = 1.12$, $MSE = 101.55$, $p = .296$.

Reporting a factorial ANOVA, with significant interaction:

Test error rates were submitted to a two-way ANOVA with two levels of room noise (noisy, quiet) and two levels of room temperature (hot, cold). The main effect of room noise was significant, $F(1, 39) = 8.66$, $MSE = 5.65$, $p = .005$, suggesting that number of errors in the noisy group ($M = 5.75$, $SD = 2.98$) was greater than the quiet group ($M = 3.89$ $SD = 1.89$). The main effect of room temperature was not significant, $F(1, 39) = 0.89$, $MSE = 5.65$, $p = .351$. However, the interaction was significant, $F(1, 39) = 9.55$, $MSE = 5.65$, $p = .003$, suggesting that effect of room noise was greater in the hot room than in the cold room.

Reporting a chi-square test of independence:

A chi-square test of independence examined the relationship between religion (Catholic, Protestant, Mormon, Secular/Atheist) and belief in the Theory of Evolution (Yes, No). The relationship between these variables was statistically significant, χ^2 (3, $N = 2000$) = 9.64, $p < .01$. Results show the Mormons were less likely to believe in the Theory of Evolution and Secularists/Atheists were more likely to believe in the Theory of Evolution.

Reporting a chi-square goodness of fit test:

A chi-square goodness of fit test was used to determine whether the six Star Wars movies (Phantom Menace, Attack of the Clones, Revenge of the Sith, New Hope, Empire Strikes Back, Return of the Jedi) were equally preferred. Preference for the six Star Wars movies was not equally distributed, χ^2 (5, $N = 250$) = 21.36, $p < .001$.