

Writing up your results – Guidelines based on APA style

In a results section, your goal is to report the results of the data analyses used to test your hypotheses. To do this, you need to identify your data analysis technique, report your test statistic, and provide some interpretation of the results. Each analysis you run should be related to your hypotheses. And if you analyze data that is exploratory, you need to indicate this.

If your results are complicated – you have many conditions and/or many dependent measures, adding a table of figure can be helpful. See the APA Publication manual (5th edition) for examples.

In reporting the results of statistical tests, report the descriptive statistics, such as means and standard deviations, as well as the test statistic, degrees of freedom, obtained value of the test, and the probability of the result occurring by chance (p value). Test statistics and p values should be rounded to two decimal places (if you are providing precise p-values for future use in meta-analyses, 3 decimal places is acceptable). All statistical symbols (sample statistics) that are not Greek letters should be italicized (*M*, *SD*, *t*, *p*, etc.).

When reporting a significant difference between two conditions, indicate the direction of this difference, i.e. which condition was more/less/higher/lower than the other condition(s). Assume that your audience has a professional knowledge of statistics. Do not explain how or why you used a certain test unless it is unusual (i.e., such as a non-parametric test).

In the sections below, there are examples and tips for how to report your test statistics. We will cover many of the tests you are learning in this class and will likely use regularly in the future. We will also touch upon data analysis techniques that you will learn in other classes (i.e., regression).

p values

There are two ways to report p values. One way is to use the alpha level (the a priori criterion for the probability of falsely rejecting your null hypothesis), which is typically .05 or .01. Example: $F(1, 24) = 44.4, p < .01$. You may also report the exact p value (this is the preferred option if you want to make your data convenient for individuals conducting a meta-analysis on the topic). For example: $t(33) = 2.10, p = .03$. If your exact p value is less than .001, it is conventional to state merely $p < .001$. If you report exact p values, state early in the results section the alpha level used as a significance criterion for your tests. For example: “We used an alpha level of .05 for all statistical tests.”

What if your results are in the predicted direction but not significant? If your p-value is .10 or less, you can say your results were *marginally* significant. Example: Results indicated a marginally significant preference for pecan pie ($M = 3.45, SD = 1.11$) over cherry pie ($M = 3.00, SD = .80$), $t(5) = 1.25, p = .08$. If your p-value is over .10, you can say your results revealed a non-significant trend in the predicted direction. Example: Results indicated a non-significant trending in the predicted direction indicating a preference for pecan pie ($M = 3.45, SD = 2.11$)

over cherry pie ($M = 3.00$, $SD = 2.80$), $t(5) = 1.25$, $p = .26$. Note: this is one example where disciplines vary but this is what is acceptable in Social Psychology.

Descriptive Statistics

Mean and Standard Deviation are most clearly presented in parentheses:

- The sample as a whole was relatively young ($M = 19.22$, $SD = 3.45$).
- The average age of students was 19.22 years ($SD = 3.45$).

Percentages are also most clearly displayed in parentheses with no decimal places:

- Nearly half (49%) of the sample was married.

t-tests

There are several different research designs that utilize a t-test for the statistical inference testing. The differences between one-sample t-tests, related measures t-tests, and independent samples t-tests are so clear to the knowledgeable reader that most journal editors eliminate the elaboration of which type of t-test has been used. Additionally, the descriptive statistics provided will identify further which variation was employed. It is important to note that we assume that all p-values represent two-tailed tests unless otherwise noted and that independent samples t-tests use the pooled variance approach (based on an equal variances assumption) unless otherwise noted:

- There was a significant effect for gender, $t(54) = 5.43$, $p < .001$, with men receiving higher scores than women.
- Results indicate a significant preference for pecan pie ($M = 3.45$, $SD = 1.11$) over cherry pie ($M = 3.00$, $SD = .80$), $t(15) = 4.00$, $p = .001$.
- The 36 study participants had a mean age of 27.4 ($SD = 12.6$) were significantly older than the university norm of 21.2 years, $t(35) = 2.95$, $p = 0.01$.
- Students taking statistics courses in psychology at the University of Washington reported studying more hours for tests ($M = 121$, $SD = 14.2$) than did UW college students in general, $t(33) = 2.10$, $p = .034$.
- The 25 participants had an average difference from pre-test to post-test anxiety scores of -4.8 ($SD = 5.5$), indicating the anxiety treatment resulted in a significant decrease in anxiety levels, $t(24) = -4.36$, $p = .005$ (one-tailed).
- The 36 participants in the treatment group ($M = 14.8$, $SD = 2.0$) and the 25 participants in the control group ($M = 16.6$, $SD = 2.5$), demonstrated a significance difference in performance ($t[59] = -3.12$, $p = .01$); as expected, the visual priming treatment inhibited performance on the phoneme recognition task.
- UW students taking statistics courses in Psychology had higher IQ scores ($M = 121$, $SD = 14.2$) than did those taking statistics courses in Statistics ($M = 117$, $SD = 10.3$), $t(44) = 1.23$, $p = .09$.
- Over a two-day period, participants drank significantly fewer drinks in the experimental group ($M = 0.667$, $SD = 1.15$) than did those in the wait-list control group ($M = 8.00$, $SD = 2.00$), $t(4) = -5.51$, $p = .005$.

ANOVA and post hoc tests

ANOVAs are reported like the t test, but there are two degrees-of-freedom numbers to report. First report the between-groups degrees of freedom, then report the within-groups degrees of

freedom (separated by a comma). After that report the F statistic (rounded off to two decimal places) and the significance level.

One-way ANOVA:

- The 12 participants in the high dosage group had an average reaction time of 12.3 seconds ($SD = 4.1$); the 9 participants in the moderate dosage group had an average reaction time of 7.4 seconds ($SD = 2.3$), and the 8 participants in the control group had a mean of 6.6 ($SD = 3.1$). The effect of dosage, therefore, was significant, $F(2,26) = 8.76$, $p = .012$.
- An one way analysis of variance showed that the effect of noise was significant, $F(3,27) = 5.94$, $p = .007$. Post hoc analyses using the Scheffé post hoc criterion for significance indicated that the average number of errors was significantly lower in the white noise condition ($M = 12.4$, $SD = 2.26$) than in the other two noise conditions (traffic and industrial) combined ($M = 13.62$, $SD = 5.56$), $F(3, 27) = 7.77$, $p = .042$.
- Tests of the four a priori hypotheses were conducted using Bonferroni adjusted alpha levels of .0125 per test (.05/4). Results indicated that the average number of errors was significantly lower in the silence condition ($M = 8.11$, $SD = 4.32$) than were those in both the white noise condition ($M = 12.4$, $SD = 2.26$), $F(1, 27) = 8.90$, $p = .011$ and in the industrial noise condition ($M = 15.28$, $SD = 3.30$), $F(1, 27) = 10.22$, $p = .007$. The pairwise comparison of the traffic noise condition with the silence condition was non-significant. The average number of errors in all noise conditions combined ($M = 15.2$, $SD = 6.32$) was significantly higher than those in the silence condition ($M = 8.11$, $SD = 3.30$), $F(1, 27) = 8.66$, $p = .009$.

Multiple Factor (Independent Variable) ANOVA

- There was a significant main effect for treatment, $F(1, 145) = 5.43$, $p < .01$, and a significant interaction, $F(2, 145) = 3.13$, $p < .05$.
- The cell sizes, means, and standard deviations for the 3x4 factorial design are presented in Table 1. The main effect of Dosage was marginally significant ($F[2,17] = 3.23$, $p = .067$), as was the main effect of diagnosis category, $F(3,17) = 2.87$, $p = .097$. The interaction of dosage and diagnosis, however, has significant, $F(6,17) = 14.2$, $p = .0005$.
- Attitude change scores were subjected to a two-way analysis of variance having two levels of message discrepancy (small, large) and two levels of source expertise (high, low). All effects were statistically significant at the .05 significance level. The main effect of message discrepancy yielded an F ratio of $F(1, 24) = 44.4$, $p < .001$, indicating that the mean change score was significantly greater for large-discrepancy messages ($M = 4.78$, $SD = 1.99$) than for small-discrepancy messages ($M = 2.17$, $SD = 1.25$). The main effect of source expertise yielded an F ratio of $F(1, 24) = 25.4$, $p < .01$, indicating that the mean change score was significantly higher in the high-expertise message source ($M = 5.49$, $SD = 2.25$) than in the low-expertise message source ($M = 0.88$, $SD = 1.21$). The interaction effect was non-significant, $F(1, 24) = 1.22$, $p > .05$.
- A two-way analysis of variance yielded a main effect for the diner's gender, $F(1,108) = 3.93$, $p < .05$, such that the average tip was significantly higher for men ($M = 15.3\%$, $SD = 4.44$) than for women ($M = 12.6\%$, $SD = 6.18$). The main effect of touch was non-significant, $F(1, 108) = 2.24$, $p > .05$. However, the interaction effect was significant, $F(1, 108) = 5.55$, $p < .05$, indicating that the gender effect was greater in the touch condition than in the non-touch condition.

Chi Square

Chi-Square statistics are reported with degrees of freedom and sample size in parentheses, the Pearson chi-square value (rounded to two decimal places), and the significance level:

- The percentage of participants that were married did not differ by gender, $X^2(1, N = 90) = 0.89, p > .05$.
- The sample included 30 respondents who had never married, 54 who were married, 26 who reported being separated or divorced, and 16 who were widowed. These frequencies were significantly different, $X^2(3, N = 126) = 10.1, p = .017$.
- As can be seen by the frequencies cross tabulated in Table xx, there is a significant relationship between marital status and depression, $X^2(3, N = 126) = 24.7, p < .001$.
- The relation between these variables was significant, $X^2(2, N = 170) = 14.14, p < .01$. Catholic teens were less likely to show an interest in attending college than were Protestant teens.
- Preference for the three sodas was not equally distributed in the population, $X^2(2, N = 55) = 4.53, p < .05$.

Correlations

Correlations are reported with the degrees of freedom (which is N-2) in parentheses and the significance level:

- The two variables were strongly correlated, $r(55) = .49, p < .01$.

Regression analyses

Regression results are often best presented in a table. APA doesn't say much about how to report regression results in the text, but if you would like to report the regression in the text of your Results section, you should at least present the standardized slope (beta) along with the t-test and the corresponding significance level. (Degrees of freedom for the t-test is N-k-1 where k equals the number of predictor variables.) It is also customary to report the percentage of variance explained along with the corresponding F test.

- Social support significantly predicted depression scores, $b = -.34, t(225) = 6.53, p < .01$. Social support also explained a significant proportion of variance in depression scores, $R^2 = .12, F(1, 225) = 42.64, p < .01$.

Tables

- APA style tables do not contain any vertical lines
- There are no periods used after the table number or title.
- When using columns with decimal numbers, make the decimal points line up.
- I use MS Word tables to create my tables

Table 1.

Correlations Between Measures

Measure	Second-order belief	Factual-deception	Self-presentation
Age	0.763*	0.631**	0.842**
Second-order belief		0.724**	0.775**

Note. * $p < .01$, ** $p < .001$