7.6 Template and APA-Style Write-Up

Next we develop APA-style paragraphs describing the results for both examples. First is a paragraph describing the results of the independent *t* test for the cholesterol example, and this is followed by dependent *t* test for the swimming example.

**Independent *t* Test**

Recall that our graduate research assistant, Marie, was working with JoAnn, a local nurse practitioner, to assist in analyzing cholesterol levels. Her task was to assist JoAnn with writing her research question (*Is there a mean difference in cholesterol level between males and females?*) and generating the test of inference to answer her question. Marie suggested an independent samples *t* test as the test of inference. A template for writing a research question for an independent *t* test is presented as follows:

Is there a mean difference in [dependent variable] between [group 1 of the independent variable] and [group 2 of the independent variable]?

It may be helpful to preface the results of the independent samples *t* test with information on an examination of the extent to which the assumptions were met (recall there are three assumptions: normality, homogeneity of variances, and independence). This assists the reader in understanding that you were thorough in data screening prior to conducting the test of inference.

An independent samples *t* test was conducted to determine if the mean cholesterol level of males differed from females. The assumption of normality was tested and met for the distributional shape of the dependent variable (cholesterol level) for females. Review of the S-W test for normality (*SW = .931, df = 8, p = .525*) and skewness (.000) and kurtosis (-1.790) statistics suggested that normality of cholesterol levels for females was a reasonable assumption. Similar results were found for male cholesterol levels. Review of the S-W test for normality (*SW = .949, df = 12, p = .617*) and skewness (.000) and kurtosis (-1.446) statistics suggested that normality of males cholesterol levels was a reasonable assumption. The boxplots suggested a relatively normal distributional shape (with no outliers) of cholesterol levels for both males and females. The Q-Q plots and histograms suggested some minor nonnormality for both male and female cholesterol levels. Due to the small sample, this was anticipated. Although normality indices generally suggest the assumption is met, even if there are slight departures from normality, the effects on Type I and Type II errors will be minimal given the use of a two-tailed test (e.g., Glass, Peckham, & Sanders, 1972; Sawilowsky & Blair, 1992). According to Levene’s test, the homogeneity of variance assumption was satisfied (*F = 3.2007, p = .090*). Because there was no random assignment of the individuals to gender, the assumption of independence was not met, creating a potential for an increased probability of a Type I or Type II error.
It is also desirable to include a measure of effect size. Recall our formula for computing the effect size, \( \delta \), presented earlier in the chapter. Plugging in the values for our cholesterol example, we find an effect size \( \delta \) of \(-1.1339\), which is interpreted according to Cohen's (1988) guidelines as a large effect:

\[
d = \frac{\bar{Y}_1 - \bar{Y}_2}{s_p} = \frac{185 - 215}{26.4575} = -1.1339
\]

Remember that for the two-sample mean test, \( d \) indicates how many standard deviations the mean of sample 1 is from the mean of sample 2. Thus, with an effect size of \(-1.1339\), there are nearly one and one-quarter standard deviation units between the mean cholesterol levels of males as compared to females. The negative sign simply indicates that group 1 has the smaller mean (as it is the first value in the numerator of the formula; in our case, the mean cholesterol level of females).

Here is an APA-style example paragraph of results for the cholesterol level data (remember that this will be prefaced by the paragraph reporting the extent to which the assumptions of the test were met).

As shown in Table 7.3, cholesterol data were gathered from samples of 12 males and 8 females, with a female sample mean of 185 (SD = 19.09) and a male sample mean of 215 (SD = 30.22). The independent \( t \) test indicated that the cholesterol means were statistically significantly different for males and females (\( t = -2.4842, df = 18, p = .023 \)). Thus, the null hypothesis that the cholesterol means were the same by gender was rejected at the .05 level of significance. The effect size \( \delta \) (calculated using the pooled standard deviation) was \(-1.1339\). Using Cohen's (1988) guidelines, this is interpreted as a large effect. The results provide evidence to support the conclusion that males and females differ in cholesterol levels, on average. More specifically, males were observed to have larger cholesterol levels, on average, than females.

Parenthetically, notice that the results of the Welch \( t' \) test were the same as for the independent \( t \) test (Welch \( t' = -2.7197 \), rounded \( df = 18, p = .014 \)). Thus, any deviation from homogeneity of variance did not affect the results.

---

**Dependent \( t \) Test**

Marie, as you recall, was also working with Mark, a local swimming coach, to assist in analyzing freestyle swimming time before and after swimmers participated in an intensive training program. Marie suggested a research question (Is there a mean difference in swim time for the 50-meter freestyle event before participation in an intensive training program as compared to swim time for the 50-meter freestyle event after participation in an intensive training program?) and assisted in generating the test of inference (specifically the dependent \( t \) test) to answer her question. A template for writing a research question for a dependent \( t \) test is presented as follows.

Is there a mean difference in [paired sample 1] as compared to [paired sample 2]?
Inferences About the Difference Between Two Means

It may be helpful to preface the results of the dependent samples $t$ test with information on the extent to which the assumptions were met (recall there are three assumptions: normality, homogeneity of variance, and independence). This assists the reader in understanding that you were thorough in data screening prior to conducting the test of inference.

A dependent samples $t$ test was conducted to determine if there was a difference in the mean swim time for the 50 meter freestyle before participation in an intensive training program as compared to the mean swim time for the 50 meter freestyle after participation in an intensive training program. The assumption of normality was tested and met for the distributional shape of the paired differences. Review of the S-W test for normality ($SW = .956$, $df = 10$, $p = .734$) and skewness (.248) and kurtosis (.050) statistics suggested that normality of the paired differences was reasonable. The boxplot suggested a relatively normal distributional shape, and there were no outliers present. The Q-Q plot and histogram suggested minor nonnormality. Due to the small sample, this was anticipated. Homogeneity of variance was tested by reviewing the ratio of the raw score variances. The ratio of the smallest (posttest = 13.111) to largest (pretest = 17.778) variance was less than 1.4; therefore, there is evidence of the equal variance assumption. The individuals were not randomly selected; therefore, the assumption of independence was not met, creating a potential for an increased probability of a Type I or Type II error.

It is also important to include a measure of effect size. Recall our formula for computing the effect size, $d$, presented earlier in the chapter. Plugging in the values for our swimming example, we find an effect size $d$ of 2.3146, which is interpreted according to Cohen's (1988) guidelines as a large effect:

$$
\text{Cohen } d = \frac{\overline{d}}{s_d} = \frac{5}{2.1602} = 2.3146
$$

With an effect size of 2.3146, there are about two and a third standard deviation units between the pretraining mean swim time and the posttraining mean swim time.

Here is an APA-style example paragraph of results for the swimming data (remember that this will be prefaced by the paragraph reporting the extent to which the assumptions of the test were met).

From Table 7.4, we see that pretest and posttest data were collected from a sample of 10 swimmers, with a pretest mean of 64 seconds ($SD = 4.22$) and a posttest mean of 59 seconds ($SD = 3.62$). Thus, swimming times decreased from pretest to posttest. The dependent $t$ test was conducted to determine if this difference was statistically significantly different from 0, and the results indicate that the pretest and posttest means were statistically different ($t = 7.319$, $df = 9$, $p < .001$). Thus, the null hypothesis that the freestyle swimming means were the same at both points in time was rejected at the .05 level of significance. The effect size $d$ (calculated as the mean difference divided by the standard
deviation of the difference) was 2.3146. Using Cohen’s (1988) guidelines, this is interpreted as a large effect. The results provide evidence to support the conclusion that the mean 50 meter freestyle swimming time prior to intensive training is different than the mean 50 meter freestyle swimming time after intensive training.

7.7 Summary

In this chapter, we considered a second inferential testing situation, testing hypotheses about the difference between two means. Several inferential tests and new concepts were discussed. New concepts introduced were independent versus dependent samples, the sampling distribution of the difference between two means, the standard error of the difference between two means, and parametric versus nonparametric tests. We then moved on to describe the following three inferential tests for determining the difference between two independent means: the independent t test, the Welch t' test, and briefly the Mann–Whitney–Wilcoxon test. The following two tests for determining the difference between two dependent means were considered: the dependent t test and briefly the Wilcoxon signed ranks test. In addition, examples were presented for each of the t tests, and recommendations were made as to when each test is most appropriate. The chapter concluded with a look at SPSS and G’Power (for post hoc power) as well as developing an APA-style write-up of results. At this point, you should have met the following objectives: (a) be able to understand the basic concepts underlying the inferential tests of two means, (b) be able to select the appropriate test, and (c) be able to determine and interpret the results from the appropriate test. In the next chapter, we discuss inferential tests involving proportions. Other inferential tests are covered in subsequent chapters.

Problems

Conceptual Problems

7.1 We test the following hypothesis:

\[ H_0: \mu_1 - \mu_2 = 0 \]

\[ H_1: \mu_1 - \mu_2 \neq 0 \]

The level of significance is .05 and \( H_0 \) is rejected. Assuming all assumptions are met and \( H_0 \) is true, the probability of committing a Type I error is which one of the following?

a. 0
b. 0.05
c. Between .05 and .95
d. 0.95
e. 1.00