APPENDIX 11.3: PROVING THE GENERALITY OF $(m_1 - m_2) \pm t_{\alpha/2}(s_{m_1-m_2})$

The derivation in this appendix has been shown in the appendixes of Chapters 6 and 7. The only thing that changes here is the way that t is calculated. (This is the last of these derivations that we'll consider.)

By definition we know that (1–α)100% of the t-distribution lies in the interval $-t_{\alpha/2}$ to $t_{\alpha/2}$. We can express this as

$$\Pr(-t_{\alpha/2} < t < t_{\alpha/2}) = 1-\alpha. \qquad (11.A3.1)$$

If we insert the definition of t into equation 11.A3.1, we obtain the following:

$$\Pr\left(-t_{\alpha/2} < \frac{(m_1 - m_2) - \mu_{m_1-m_2}}{s_{m_1-m_2}} < t_{\alpha/2}\right) = 1-\alpha. \qquad (11.A3.2)$$

Multiplying all three terms by $s_{m_1-m_2}$ and then subtracting $m_1 - m_2$ from all three terms leaves us with this:

$$\Pr(-t_{\alpha/2}(s_{m_1-m_2}) - (m_1 - m_2) < -\mu_{m_1-m_2} < t_{\alpha/2}(s_{m_1-m_2}) - (m_1 - m_2)) = 1-\alpha. \qquad (11.A3.3)$$

This transformation leaves $-\mu_{m_1-m_2}$ in the center of the inequality. If we now multiply all three terms by –1 (to make $\mu_{m_1-m_2}$ positive), we obtain

$$\Pr(t_{\alpha/2}(s_{m_1-m_2}) + (m_1 - m_2) > \mu_{m_1-m_2} > -t_{\alpha/2}(s_{m_1-m_2}) + (m_1 - m_2)) = 1-\alpha. \qquad (11.A3.4)$$

When we rearrange the statistic $m_1 - m_2$ and the margin of error $t_{\alpha/2}(s_{m_1-m_2})$, we obtain

$$\Pr((m_1 - m_2) + t_{\alpha/2}(s_{m_1-m_2}) > \mu_{m_1-m_2} > (m_1 - m_2) - t_{\alpha/2}(s_{m_1-m_2})) = 1-\alpha. \qquad (11.A3.5)$$

Equation 11.A3.5 shows that

$$(m_1 - m_2) \pm t_{\alpha/2}(s_{m_1-m_2}) \qquad (11.A3.6)$$

will contain $\mu_{m_1-m_2}$, with probability 1–α. This means that $\mu_{m_1-m_2}$ will be within (1–α)100% of intervals computed as $(m_1 - m_2) \pm t_{\alpha/2}(s_{m_1-m_2})$.

The beauty of this small demonstration is the generality of what it shows. No matter what n_1, n_2, μ_1, μ_2, and σ^2 are, (1–α)100% of confidence intervals computed as in equation 11.A3.6 will capture $\mu_{m_1-m_2}$.

APPENDIX 11.4: WELCH-SATTERTHWAITE CORRECTION FACTOR

In Chapter 11, we discussed using the t-distribution when we are able to assume that $\sigma_1^2 = \sigma_2^2$. When we are not able to make this assumption, then there is good and bad news. The good news is that we compute $s_{m_1-m_2}$ as follows:

$$s_{m_1-m_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}. \qquad (11.A4.1)$$

That's pretty straightforward. More good news is that the sampling distribution of

$$t = \frac{(m_1 - m_2) - (\mu_1 - \mu_2)}{s_{m_1-m_2}}$$

will be a t-distribution. The bad news is that it will not be associated with $n_1 + n_2 - 2$ degrees of freedom. Rather, an adjustment to the degrees of freedom must be made to account for this change, and the needed equation is a little complex:

$$df_{WS} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}\right)}. \qquad (11.A4.2)$$

We call the adjusted degrees of freedom df_{WS} because equation 11.A4.2 is called the Welch-Satterthwaite correction factor, after the statisticians who developed it. The effect of the correction is to reduce the degrees of freedom somewhat. Reducing the degrees of freedom makes $t_{\alpha/2}$ larger than it would otherwise be.

We will reexamine the data from the riddle study described in Chapter 11 (Table 11.3 is reproduced in Table 11.A4.1) without the assumption that $\sigma_1^2 = \sigma_2^2$.

TABLE 11.A4.1 ■ Riddle Study Data

	Quiet (*q*)	Noisy (*n*)
n	11	11
m	12	9
s^2	24.4	15.6

From this information, we can compute $s_{m_1-m_2}$ as follows:

$$s_{m_1-m_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{24.4}{11} + \frac{15.6}{11}} = 1.9069.$$

To put a confidence interval around the difference between the two means, we would have to adjust the degrees of freedom associated with $t_{\alpha/2}$. This requires evaluating the Welch-Satterthwaite correction as follows:

$$\begin{aligned} df_{\text{WS}} &= \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}\right)} \\ &= \frac{\left(\frac{24.4}{11} + \frac{15.6}{11}\right)^2}{\left(\frac{595.36}{121(10)} + \frac{243.36}{121(10)}\right)} \\ &= 19.08. \end{aligned}$$

Please note that s^4 is the square of the variance, s^2. That is, $s^4 = (s^2)^2$.

The $t_{\alpha/2}$ value associated with df_{WS} cannot be taken from the *t*-table because there are fractional degrees of freedom. In addition, although the **T.INV.2T** function in Excel accepts fractional degrees of freedom, it simply ignores the fractional part and rounds down to the next integer; this is not much help. Fortunately, **R** allows us to compute $t_{\alpha/2}$ for fractional degrees of freedom. Typing the following command into **R**,

```
qt(p = .975, df = 19.08),
```

yields $t_{\text{WS}} = 2.09243$, which is slightly larger than the 2.086 that we would have obtained with $11 + 11 - 2 = 20$ degrees of freedom. The Welch-Satterthwaite corrected 95% confidence interval around the difference between the two means will be

$$\begin{aligned} \text{CI} &= (m_1 - m_2) \pm t_{\text{WS}}(s_{m_1-m_2}) \\ &= (12 - 9) \pm 2.09243(1.9069) \\ &= [-0.99, 6.99]. \end{aligned}$$

In general, the Welch-Satterthwaite correction for unequal variances will increase the width of our confidence intervals. However, the extent of the increase depends on the size of the difference between the two sample variances; the larger the difference, the greater the increase. If sample variances and sample sizes are identical, then there will be no correction to the degrees of freedom.

APPENDIX 11.5: PLANNING THE MARGIN OF ERROR

Specifying the Desired Margin of Error

We saw previously that when estimating μ when σ is known, the sample size required to obtain an *moe* of $f * \sigma$ is

$$n = (z_{\alpha/2}/f)^2.$$

When estimating μ when σ is unknown, we can use the same formula but add 3 to *n*, to get the approximate sample size needed to achieve an *moe* that is $f * \sigma$, on average.

We are now faced with determining the required sample size to achieve an *moe* of $f * \sigma$ when there are two independent groups of scores. As before, the solution is very easy if we can tolerate slight imprecision. To determine our required sample sizes, we will assume two groups of equal size, drawn from distributions having the same standard deviation, σ. To achieve an average *moe* of $f * \sigma$, both groups should have a sample size of

$$n = 2(z_{\alpha/2}/f)^2 + 1.$$

If we would like to achieve a precision of $.2 * \sigma$, for example, then we would need to have

$$n = 2(z_{\alpha/2}/f)^2 + 1 = 2(1.96/.2)^2 + 1 = 193.08$$

scores in each sample. Of course, 193.08 scores would be difficult to obtain, so we round up to 194.

Planning a Study

In previous chapters, we considered the Prescott Pharmaceuticals Company's (PPC) interest in attentional focus (AF) in adolescents with attention deficit disorder (ADD). They now wish to test a new drug aimed at improving AF scores in adolescents with ADD. Their study will have two independent groups.

One group will be adolescents with ADD who receive no treatment, and the other will be adolescents with ADD who undergo 4 months of treatment with a new drug. PPC would like the precision of their 95% confidence interval to be .1(σ), on average (i.e., $f = .1$).

They offered \$50,000.00 for someone to carry out this study, and your research showed that it would cost \$95.00 to obtain AF scores from each adolescent. To figure out how much profit you would make if you were to accept this contract, you need to determine the number of participants required to achieve the desired *moe*. You start by estimating the number that would be required if σ were known then add 1 to the result. When we fill numbers into the expression above we obtain the following:

$$n = 2(z_{\alpha/2}/f)^2 + 1 = 2(1.96/.1)^2 + 1 = 769.32.$$

Again, we need to round up to 770 to achieve a whole number. So, we need 770 individuals *in each group* for a total of 1,540. At a rate of \$95.00 per measure, it would cost you 1,540 * \$95 = \$146,300.00 to conduct the study. It would be a very bad idea to accept this contract.

Precision Implied by Sample Size

Now that we know how to derive the sample size required to achieve a desired precision, we can turn this around and ask what precision we achieve with a given sample size, $n = n_1 = n_2$. The following formula provides this answer:

$$f = \frac{(z_{\alpha/2})\sqrt{2}}{\sqrt{n-1}}.$$

In the riddle example from Chapter 11, we noted that the precision of the 95% confidence interval around $m_1 - m_2$ was $f = 0.88$. This was calculated as follows:

$$f = \frac{(z_{\alpha/2})\sqrt{2}}{\sqrt{n-1}} = \frac{1.96 * 1.4142}{\sqrt{11-1}} = 0.88.$$

We can look at this slightly differently and ask about the relationship between the margin of error, $moe = t_{\alpha/2}(s_{m_1-m_2})$, and our best estimate of σ, which is s_{pooled}. In the riddle example, $t_{\alpha/2}(s_{m_1-m_2}) = 3.98$ and $s_{\text{pooled}} = 4.47$. Therefore, an estimate of f is given by

$$f = \frac{t_{\alpha/2}(s_{m_1-m_2})}{s_{\text{pooled}}} = \frac{3.98}{4.47} = 0.89.$$

Both methods show that our estimate is rather imprecise, with the margin of error being close to a full standard deviation, σ.

APPENDIX 11.6: USING G*POWER TO COMPUTE POST HOC POWER AND SENSITIVITY

Prospective power analysis is conducted before running an experiment or quasi-experiment to determine the sample size required to have a high probability of rejecting the null hypothesis for a given effect size, δ. Unfortunately, power calculations are rarely conducted before running an experiment. However, in a sort of postmortem, researchers may ask about the power of their experiment based on data they've collected. This postmortem requires assuming that d is equal to δ.

In Figure 11.A6.1, we use G*Power to estimate the power of our experiment, assuming d equals δ. We will work with the riddle experiment from Chapter 11, in which there were 11 participants in each group and δ was estimated to be $d = .67$. Figure 11.A6.1 shows that we have chosen *t*-tests from the drop-down list under Test family. From the drop-down list under Statistical test, we have chosen Means: Difference between two independent means (two groups), which is the topic of this chapter. The Type of power analysis is Post hoc: Compute achieved power - given α, sample size, and effect size. "Post hoc" is the Latin term for "after the event." So, after the experiment has been run, we're trying to determine the power of the experiment assuming that $\delta = d$. Therefore, in the Input parameters panel, we describe the experiment as one-tailed, with an effect size of .67, $\alpha = .05$, and the two sample sizes as 11. When we click on the **Calculate** button, G*Power performs its calculations and shows that the power of the experiment is approximately .45. That is, if δ really were equal to .67, then the experiment we've run would only reject the null hypothesis 45% of the time when two samples are drawn from two distributions whose means are separated by $\delta = .67$.

We can also ask the following: for a one-tailed test with $\alpha = .05$ and $n_1 = n_2 = 11$, what would δ have to be to yield power = .8? This question addresses the *sensitivity* of the experiment. The smaller the effect size that can yield power(100)% correct rejections of the null hypothesis, the more sensitive the test. The sensitivity of the test can be determined using G*Power as shown in Figure 11.A6.2.

We have chosen t-tests from the drop-down list under Test family, and we chose Means: Difference between two independent means (two groups) from the drop-down list

FIGURE 11.A6.1 ■ Post Hoc Power Computations

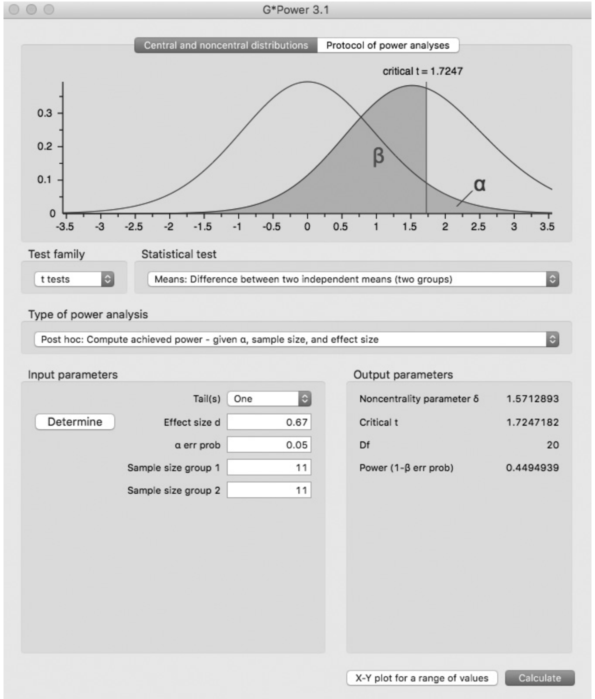
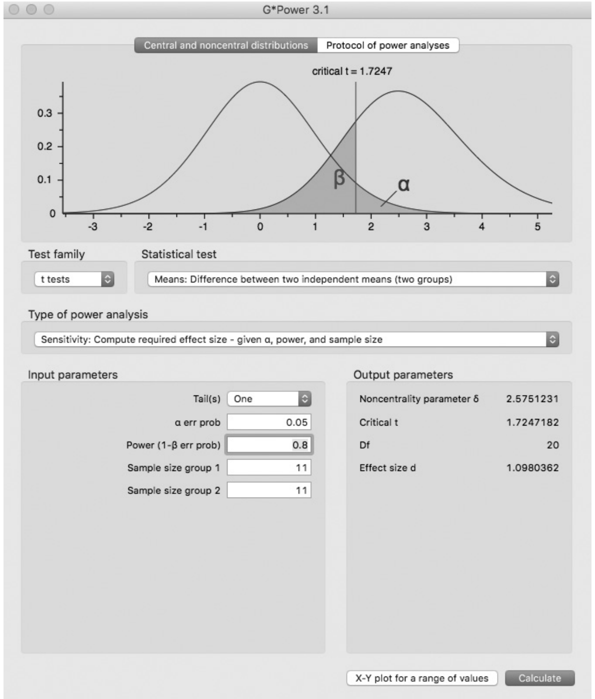


FIGURE 11.A6.2 ■ Sensitivity Computations



under Statistical test, just as in Figure 11.A6.1. The Type of power analysis is Sensitivity: Compute required effect size - given α, sample size, and effect size. Therefore, in the Input parameters panel, we describe the experiment as one-tailed, with power = .8, and the two sample sizes as 11. When we click on the **Calculate** button, G*Power performs its calculations and shows the value of δ required to achieve the desired power. The analysis shows that our experiment would require δ = 1.1 to achieve power = .8. Therefore, one might say that our hypothetical experimenter's choice of sample size implies that she was interested in effect sizes of 1.1 or greater.