Suppose that a population of young adults has a mean score of $\mu = 8$ on a memory test, with $\sigma = 2$. A random sample of 25 older adults is taken and they are given the same memory test. We will assume that the population of older adults from which this sample was taken have the same standard deviation as the population of young adults ($\sigma = 2$). We will test the null hypothesis that the older adult population has the same mean as the population of young adults ($\mu = 8$) using a one-sample $z$ test. The alternative hypothesis is a directional one, with the expectation that the older adult population mean is less than the mean for young adults. We will set $\alpha = .05$.

(a) Draw a diagram of the distribution of sample means for the population of older adults, assuming that the null hypothesis is true. We will assume that the distribution of scores in the older adult population is normal, which means that the distribution of sample means will also be normal. Draw a vertical line in the distribution of sample means indicating the location of the mean, and label its value on the horizontal axis. Determine the rejection criterion for the $z$ test that will be carried out. That is, what is the smallest value of the sample mean that must be found if the null hypothesis is to be rejected? Draw a vertical line in the distribution of sample means to show this location and label its value on the horizontal axis.

(b) Draw another diagram representing the distribution of sample means, but this time assuming that the null hypothesis is false and that the real population mean for older adults is 7, rather than 8. Draw a vertical line representing this population mean and label its value. Next, draw the vertical line representing the rejection criterion that was found in part (a) and label its value.

(c) Indicate in the second diagram the region of rejection (the area under the curve...
corresponding to results that allow rejection of $H_0$), and use the standard normal
distribution ($z$) table to compute the power of this $z$ test to detect an effect assuming the
real population mean is 7.
(d) What would the power of this $z$ test have been if $\alpha$ were set at .01?

2. A researcher is planning to test the effectiveness of a memory-enhancing drug on lab
animals. One group of animals is to receive the drug and another is to receive a placebo.
The researcher is going to test the null hypothesis against a nondirectional alternative
hypothesis (even though the direction of the effect is predicted to be enhancement) and
plans to set $\alpha$ at .05. Answer the following questions using the procedure described in class
and the special table for statistical effect size.
(a) How many animals must be used in each group to have power of at least .80 to detect
an effect size of half a standard deviation?
(b) How many animals would be needed in each group if the researcher were to change $\alpha$
to .10?

3. A researcher interested in the effect of alcohol on performance in a perceptual-motor
task randomly assigned a sample of 40 subjects to two groups of 20. One group received a
drink of alcohol and the other received a drink of a neutral substance prior to performing
the task. The following data were obtained. Higher scores indicate better performance.

<table>
<thead>
<tr>
<th></th>
<th>Alcohol</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>22.8</td>
<td>26.3</td>
</tr>
<tr>
<td>$s$</td>
<td>9.2</td>
<td>8.8</td>
</tr>
<tr>
<td>$n$</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

(a) Compute an independent-samples $t$ test and verify that the result is not significant
with $\alpha = .05$ and a nondirectional alternative hypothesis.
(b) Compute the observed effect size ($d$) using the pooled estimate of variance to get an
estimated value for population standard deviation.
(c) Using Cohen's benchmark values, how would you describe this effect size?
(d) Assuming that the observed effect size computed in part (b) is the effect size that truly
exists in the population, what is the power of this study to obtain a significant $t$ ratio?
(e) Given the power of the study to detect the effect size computed in part (b), is it safe to
conclude that such an effect size does not exist in the population? Explain.
(f) On the basis of this study's result, is it safe to conclude that an effect size ($d$) of 1.0 does
not exist in the population? Explain.
(g) Use R to compute how large the sample size for each group would have to be so that
the study would have power of .80 to detect a significant effect when the true population
effect size is .4 (which is close to the effect size actually obtained in the study).

4. A researcher sets out to test the hypothesis that children from rural areas may perform
differently from urban children on standardized tests of academic achievement. The
national average on a standardized achievement test for first grade children dwelling in
urban areas is 40 and the standard deviation is 12. To test her hypothesis, the researcher draws a random sample of 16 first grade children from rural areas and administers the achievement test to these children. The researcher plans to apply a one-sample z test to the data using $\alpha = .05$. Assume that the true population mean for this test among rural first grade children is 35.

(a) What is the power of this researcher's study to detect this effect? Determine your answer by using the standard normal distribution.

(b) Suppose that the researcher used $\alpha = .01$ instead of .05. What would be the power of this study to detect the effect?

5. When conducting statistical tests of hypotheses, researchers have commonly adopted a significance level of .05. This means that there is a .05 probability of rejecting a true or valid null hypothesis. Why do researchers not commonly adopt a lower significance level, such as .01, to reduce the chances of making such an error? Consider what you found in question 3.

6. A researcher plans to test the idea that there is a correlation between self-rated physical attractiveness and attitude ratings of self-confidence. He wants to measure enough subjects so that the power to detect a true correlation coefficient of .5 will be .80, using an $\alpha$ value of .05 and a nondirectional alternative hypothesis. How many subjects will he need?

7. A researcher is planning a study to determine whether male and female university students differ in their amount of use of the internet. The researcher plans to obtain a random sample of female and male students from the local university and to ask each subject how many hours per week they spend connected to the internet. Assuming that the researcher uses a significance level of .05, how many subjects should be included in each group to provide a .9 probability of detecting a difference of half a standard deviation between the means of the female and male populations?

8. In a large-scale medical study, researchers are testing for a possible correlation between a certain type of cancer and amount of chocolate consumed. The researchers include 1,500 subjects in their sample. How much power do they have to detect a correlation of .10, assuming a nondirectional test and $\alpha = .05$?

9. A researcher carries out a study comparing a control group to an experimental group with 16 subjects in each group. Assuming the researcher uses an independent-samples $t$ test with a nondirectional alternative hypothesis and $\alpha = .05$, how large must the actual effect size be in order for this study to have power of .80 to detect it? Draw two normally distributed populations of raw scores with equal variance that reflects this effect size.

10. A researcher plans to test a sample of 20 subjects using a design in which all subjects are tested in each of two conditions. A related-samples $t$ test with a nondirectional alternative hypothesis and $\alpha = .01$ (note the unusual level for $\alpha$) is to be used to test for a
difference between the means in the two conditions. Assume that the researcher wishes to know the level of power for this study to detect a medium effect size (i.e., \( d = 0.5 \)). To compute the power estimate, the researcher needs to estimate the strength of the correlation between scores in the two conditions. Suppose this correlation is estimated to be \( .45 \). What is the power estimate under this assumption? Now suppose that the correlation is much stronger: \( .90 \). What is the power estimate now? What can you say about how statistical power for the related-samples \( t \) test varies with respect to the size of the correlation between conditions? What would power have been if the correlation had been estimated to be \( 0 \)? Compare this estimate with the power estimate for an independent-samples \( t \) test with 20 subjects in each of two groups, a nondirectional alternative hypothesis, and \( \alpha = .01 \).

11. When computing an independent-samples \( t \) test, it is best to have a situation in which one can assume equal variance in the two populations of raw scores. Keeping in mind how this \( t \) test changes when one cannot assume equal variance, how does statistical power change when equal variance cannot be assumed, relative to when it can be assumed? Identify the cause of this change. To help you see what is happening, compare the following two scenarios. In the first case, one can safely assume homogeneity of variance, but in the second case one cannot. Compute a \( t \) test with \( \alpha = .05 \) using a nondirectional alternative hypothesis for each case. Watch carefully for the differences between the two cases. They have been set up so that only a few critical elements distinguish the cases. Notice also that you are provided with sample variances, not sample standard deviations.

<table>
<thead>
<tr>
<th></th>
<th>Case 1</th>
<th>Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Group 2</td>
<td></td>
</tr>
<tr>
<td>( M )</td>
<td>30.0</td>
<td>30.0</td>
</tr>
<tr>
<td>( s^2 )</td>
<td>16.0</td>
<td>5.0</td>
</tr>
<tr>
<td>( n )</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

12. A scientist proposes that people with alcohol dependency tend to be sensitive to alcohol-related words and concepts. To test this idea, she sets up a color-word Stroop task in which subjects with an alcohol dependency are asked to name aloud as quickly as possible the color in which a series of words is printed. Some of the words are alcohol related (e.g., vodka, beer) and the other words are neutral (e.g., grind, soap). The researcher measures for each subject the average color-naming time for alcohol-related words and for neutral words. Suppose that she finds the following results for a sample of 16 subjects.

<table>
<thead>
<tr>
<th></th>
<th>Alcohol</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M )</td>
<td>925.2</td>
<td>906.3</td>
</tr>
<tr>
<td>( s )</td>
<td>67.9</td>
<td>79.3</td>
</tr>
</tbody>
</table>

\( t(15) = 2.58 \) (Note: this is a related-samples \( t \) test)

(a) Is this effect significant with \( \alpha = .05 \) and a nondirectional alternative hypothesis?
(b) Compute the observed effect size using the formula for Cohen's \( d \): \( d = \frac{M_1 - M_2}{s_p} \)
(c) The observed effect size is the effect size actually obtained in the study. Determine the amount of power this study had to produce a significant effect, assuming the true effect size in the population is equal to the effect size observed with this sample. To compute power for this test, you will need an estimate of the correlation between conditions, because this is a related-samples t test. For these data, the researcher found that scores in the alcohol condition and scores in the neutral condition had a correlation of $r = .93$.

(d) Now suppose that a smaller effect size had been obtained, so that the data were as follows (note that sample standard deviations and the correlation between conditions are very similar to what we saw in the first case).

<table>
<thead>
<tr>
<th></th>
<th>Alcohol</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>925.2</td>
<td>914.1</td>
</tr>
<tr>
<td>$s$</td>
<td>67.9</td>
<td>79.7</td>
</tr>
</tbody>
</table>

Is this $t$ ratio significant?

(e) What is the observed effect size in this case, and what is the power of this study to detect an effect of that size with a nondirectional test and $\alpha = .05$? The correlation between conditions in this case is .94.

(f) Finally, suppose that the data in the first case shown above were actually obtained in an independent-samples design. That is, 32 subjects were tested, but were randomly assigned to two groups of 16. One group named colors carried by alcohol-related words and the other group named colors carried by neutral words. What would the $t$ ratio be in this case? Would it be significant with a nondirectional test and $\alpha = .05$? Note that the observed effect size is the same as it was in part (b). What would be the power of this study, using an independent-samples $t$ test, to detect an effect of that size?

13. A researcher is interested in determining whether ordinary human subjects can demonstrate precognition -- a type of extrasensory perception whereby an observer is shown to be influenced by future information that cannot be deduced from currently available information. In a test of this possibility, a researcher obtained a sample of 20 subjects and had each attempt to predict the outcome of a series of cards in a randomly shuffled deck of special cards. In this deck of 25 cards, five displayed a particular symbol (e.g., wavy lines), another five depicted a different symbol (e.g., a star), and so on such that there were 5 cards for each unique symbol. If a subject were simply guessing the outcome for each card, he or she would have a $5/25$ or .2 chance of being correct. On average, someone who is guessing would be expected to be correct on 5 trials out of 25. Suppose that each subject was test on the deck four times (100 cards). The expected mean performance under the null hypothesis (people are just guessing) would be 20 correct. This researcher obtained the following results for the 20 subjects who were tested: $M = 22.0$, $s = 6.0$. Using a directional alternative hypothesis and a one-sample $t$ test with $\alpha = .05$, the researcher obtained a $t$ ratio of $(22 - 20)/1.342 = 1.49$. The critical $t$ ratio for this case (directional test, $\alpha = .05$, and $df = 19$) is 1.729, so this result is not significant.

(a) Assume that precognition actually exists in the population from which these 20 subjects were sampled. What type of decision error occurred in this study?

(b) Continue to assume that precognition exists in this population. Given the outcome of this study, how might a power analysis help us to decide whether the true effect of
precognition is a large effect \((d = .8)\)? What conclusion should we reach regarding the proposal that a large effect exists, given the outcome of the present study?
(c) What does this study allow us to conclude about the proposal that precognition exists, but it is only a small effect \((d = .2)\)? Using a power analysis to guide your answer.

14. Following up on the principle behind question 12, consider an independent samples \(t\) test that fails to find a significant effect with \(\alpha = .05\) and a nondirectional alternative hypothesis. Suppose the study had 20 subjects in each condition.
(a) Use the `pwr.t.test` function in R to determine how much power this study had to find a medium effect size \((d = .5)\).
(b) Now use that function to determine how many subjects would have to be included in each condition for power to go up to .80.
(c) Sticking with the sample size obtained in part (b), use R to determine how much power a study with that sample size would have to detect an effect of \(d = .2\).
(d) Use R to find how large the sample size would have to be to have power of .80 to detect an effect size of \(d = .2\).
(e) Now repeat parts (c) and (d), but assume that the actual effect size is only \(d = .1\).