1. (a) $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 < \mu_2$ (assuming Relaxation group is group 1). This is a directional test because the researcher claims that relaxation training can reduce (not just change) phobic reactions. Keep in mind that if the researcher wanted to apply a more conservative test, he or she could have opted to use a nondirectional test with $H_1: \mu_1 \neq \mu_2$.

(b) The correct $t$ test to use here is the independent samples $t$ test because two independent groups of subjects are involved. The two variances are similar to each other ($3.72^2$ and $4.58^2$ equal 13.84 and 20.98, and the larger variance value is not at least 4 times larger than the smaller one), so a pooled estimate of variance can be used. The pooled variance estimate is obtained by computing the $SS$ value from each standard deviation. First convert each $s$ value to a variance value by squaring, then multiply each variance value by its $df$. For the relaxation group, we have $s^2 = 3.72^2 = 13.84$, $SS = 13.84(11–1) = 138.40$. For the control group, we have $s^2 = 4.58^2 = 20.98$, $SS = 20.98(15–1) = 293.72$. The pooled variance estimate $(138.40 + 293.72)/(10 + 14) = 18.00$. The estimate for the standard error of difference is $\frac{s_{M_1-M_2}}{\sqrt{\frac{18.00}{11} + \frac{18.00}{15}}} = 1.68$ so $t = (8.30–12.40)/1.68 = –2.44$.

With $\alpha = .05$ and a directional test with $df = 24$, the critical $t = –1.711$ (+1.711 if the Control group is group 1). The observed $t$ ratio is more extreme than the critical value, so $H_0$ is rejected.

(c) Relaxation training reduces the number of phobic reactions. Thanks to the random assignment of subjects to conditions, a cause-and-effect conclusion is valid here.

2. (e) about half (this distribution is symmetrical and because $H_0$ is true, the mean is zero).

3. (a) The distribution of differences would be normal with mean = 0 (see diagram in (c) below).

(b) Start by computing the variance of the distributions of sample means, then combine them to compute the variability of the distribution of differences. The sample variances are very similar, so a pooled estimate of sample variances can be used. Sample sizes are equal, so one way of computing pooled variance is simply to take the mean of the two sample variances: $4.2^2 + 4.0^2)/2 = 16.82$. Alternatively, one can use the general formula: $[(19)(4.2)^2 + 19(4.0)^2]/(19+19) = 16.82$. So the standard error of the difference (i.e., the estimated standard deviation of the distribution of differences between sample means) is $s_{M_1-M_2} = \sqrt{\frac{16.82}{20} + \frac{16.82}{20}} = 1.30$.

(c) The observed difference between means is 14.5–11.5 = 3.0; this difference is positioned in the distribution of differences below; its position is determined by the fact that one standard deviation in this distribution is estimated to be 1.30 raw score points (this was
the answer to (b)); a difference of 3 raw-score points is roughly $3/1.3 = 2.3$ standard deviations from the mean; hash marks on the $X$ axis of the diagram are spaced at intervals of 0.5 standard deviations.

(d) The observed difference between means of 3.0 is more than two standard errors away from the mean of the distributions of differences, indicating that the observed outcome is an unlikely one if the difference between the two population means is zero. Yes, you could have computed a $t$ test to come to this conclusion, but the purpose of this question is to reveal to you the inner workings of the $t$ test.

4. The 2-tailed probability associated with the observed $t$ ratio is .078, which means that, under the null hypothesis, the probability that the study will produce a difference between means that is as extreme or more extreme than the actually obtained difference is .078. This probability is comprised by an equal-sized area in each of the two tails of the distribution of differences (see the diagram below). The observed $p$ value is greater than .05, so the obtained difference between means is not among the .05 most extreme outcomes and therefore the null hypothesis is not rejected. Clearly, the obtained result is not among the .01 most extreme outcomes either, so the null is not rejected with that value of $\alpha$ either. If a directional test is used, then we are interested only in the area in the tail of the distribution occupied by the obtained difference (assuming the difference is in the direction expected by $H_1$). The two-tailed probability is .078, which means that there is .078/2 or .039 in each tail. For a directional test, the obtained difference cuts off the most extreme .039 of one tail and therefore, by a directional test, is among the most extreme .05 outcomes, so the null hypothesis is rejected with $\alpha = .05$. If $\alpha = .01$, however, the null would not be rejected because the obtained difference is not that extreme. The diagram below depicts a hypothetical distribution of differences between means and shows the approximate location of the obtained difference. The hash marks on the $X$ axis of the diagram are spaced at intervals of 0.5 standard deviations.
The two samples have very different variances ($2.3^2 = 5.29$ and $4.8^2 = 23.04$), which differ by more than a ratio of 4 to 1, so the variance estimates should not be pooled. Working with separate variance estimates for each population, we estimate the standard error of difference as follows:

$$s_{M_1-M_2} = \sqrt{\frac{5.29}{12} + \frac{23.04}{24}} = 1.18$$

yielding $t = (22.0 - 19.5)/1.18 = 2.12$. Because the variances are deemed to be heterogeneous (unequal), the critical $t$ ratio will be determined by using the $df$ value from the smaller of the two samples, with is the control group with $n_1 = 12$, $df_1 = 12 - 1 = 11$. So with a nondirectional test and $\alpha = .05$, we have critical $t = 2.201$. The observed $t$ ratio is not as extreme as this, so the researcher cannot reject the null hypothesis and he or she has no evidence that the treatment is effective. Note that two factors were working against a positive outcome here. First, a nondirectional test was used and therefore a more extreme outcome was required for significance than would have been the case with a directional test. A directional test would have been justifiable in this case. The other factor working against finding a significant effect was the inequality of the two sample variances, leading to the use of a more conservative critical $t$ ratio (one based on fewer $df$). The researcher has some control over the first factor, but not the second factor. A possible cause of unequal variances is that the subjects in the treatment condition may have varied in their reactions to the treatment, with some subjects showing remarkable improvement but others showing no improvement or perhaps even getting worse. This variation in reaction to the treatment would push the subjects' scores further apart than they would have been without treatment, resulting in an elevated variance.

Your values for the two samples will be determined by a random process, so there is no single correct answer. But note that the true value of the standard error of the difference between means (based on $\sigma = 3$ and $n_1 = n_2 = 5$) is

$$\sigma_{M_1-M_2} = \sqrt{\frac{9}{5} + \frac{9}{5}} = 1.897$$

More often than not, sample variance will be somewhat smaller than population variance (recall that the distribution of sample variances is positively skewed), so you are likely to wind up with estimates of the standard error of the difference (which is itself determined by sample variance) that is a bit smaller than 1.897. That is, when you compute $s_{M_1-M_2}$ from your samples, you are likely to find values somewhat less than 1.897 (but these are random samples, so this may not happen in your case). If you do, notice how that fact can lead to your $t$ ratios being somewhat larger than your $z$ scores. Also notice that because the estimated standard error of difference is probably different for your two sets of samples, the two resulting $t$ ratios probably are more different from each other than the corresponding $z$ scores.

Within R, start by reading in the data file and loading the psych library so that you can use the describe function to compute the descriptive statistics. Remember to use some unique name as the place to assign the data file in R.
You can see that the sample mean is greater for the second condition (NoRel), so the $t$ ratio is going to be a negative value. That doesn't mean much since we have a nondirectional alternative hypothesis. And now for the $t$ test. We are assuming that we have equal variance in the two populations, so the last parameter used with the `t.test` command specifies that assumption; otherwise the default assumption of unequal variance would be used by `t.test`. This function by default uses a nondirectional test and assumes the two groups are independent samples, which is just what we want in this case.

```
> t.test(data$Rel, data$NoRel, var.equal=T)

Two Sample t-test

data:  data$Rel and data$NoRel
  t = -2.7318, df = 124, p-value = 0.007218
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.169362 -1.304997
sample estimates:
  mean of x mean of y
  45.34615  50.08333
```

The $p$ value for the observed $t$ ratio is very small (.007), and certainly less than .05, so the null hypothesis can be rejected. Thus, relationship status is related to the number of Facebook friends. The means show that those who are not in a romantic relationship have more Facebook friends. Because subjects were not randomly assigned to conditions, we cannot make a causal inference here; we can say only that there is an association between being in a romantic relationship and number of Facebook friends.

8. Assume that the data file has been read into an R variable called `data`. The two columns of data (conditions) are labeled `Treat` and `Control`. Here is the summary of the data:

```
> describe(data)

 vars  n  mean    sd median trimmed  mad  min  max  range  skew kurtosis   se
Treat 1  78 45.35 8.58  45.0  44.92  8.15  29  68   39 0.43  -0.16 0.97
NoRel 2  48 50.08 10.73 50.5  50.52 11.86  25  69   44 -0.34  -0.64 1.55
```

Notice that the two sample means differ by about 5 points, but is this enough to conclude that the treatment had an effect? The required command to run a $t$ test of the null and alternative hypotheses related to these two conditions is the following. The `t.test` function
by default uses a nondirectional test and assumes the two groups are independent samples, which is just what we want in this case. We are assuming that the two populations have equal variance, so we have to specify that:

\[
> \texttt{t.test(data$Treat, data$Control, var.equal=T)}
\]

Two Sample t-test

data: data$Treat and data$Control
t = 2.1302, df = 78, p-value = 0.03631
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  0.325442 9.624558
sample estimates:
  mean of x mean of y
  48.825    43.850

Notice that the \( p \) value for these data is .036, which is less than .05. This means that the null hypothesis can be rejected. If \( \alpha = .01 \), then these data would \textit{not} allow the null hypothesis to be rejected because the observed difference between means is no among the 1% least likely outcomes. With \( \alpha = .05 \), we can reject the null hypothesis and conclude that the treatment condition has a higher mean than the control condition. With \( \alpha = .01 \), the null hypothesis is not rejected, so we conclude that we have no evidence that the treatment and control conditions are different.

9. For Outcome 1 we have:

\[
s_p^2 = \frac{2.5^2 + 2.1^2}{2} = 5.33 \quad s_{M_1-M_2} = \sqrt{\frac{5.33}{35} + \frac{5.33}{35}} = .55 \quad t = \frac{9.0 - 7.5}{0.55} = 2.73
\]

In this case, \( df = 35+35-2 = 68 \), but there is no entry for \( df = 68 \) in the \( t \) table. Taking the next lowest \( df \) value, we have \( df = 68 \), which gives 50 a critical \( t \) ratio of 2.009 for a two-tailed test and \( \alpha = .05 \). The observed \( t \) ratio is more extreme than this critical value, so we can reject the null hypothesis and conclude that performance was better when notetaking was done with a tablet.

For Outcome 2, we have:

\[
s_p^2 = \frac{3.8^2 + 4.2^2}{2} = 16.04 \quad s_{M_1-M_2} = \sqrt{\frac{16.04}{35} + \frac{16.04}{35}} = .96 \quad t = \frac{9.0 - 7.5}{0.96} = 1.56
\]

In this case, we cannot reject the null hypothesis. The reason this happened is that scores are more variable in this case (variance of raw scores is higher than in the first case). That higher variability means that the sample means are more variable, and in turn, the
distribution of differences between means is more variable. That means that larger differences between means are to be expected, even though the null hypothesis may be true. So a more extreme difference between means is needed to reject the null hypothesis.