1. The probability decreases. As sample size increases, sample means will become less variable, so it will become less likely that a sample mean very different from \( \mu \) will be obtained. With \( \sigma = 10 \), the standard error of the mean \( (\sigma_M) \), which is the standard deviation of the distribution of sample means would be \( 10/\sqrt{20} = 10/4.472 = 2.236 \). A sample mean of 54 corresponds to a \( z \) score of \( (54 - 50)/2.236 = 1.79 \). The area corresponding to values greater than or equal to a \( z \) value of 1.79 is the smaller portion. From the \( z \) table, that portion is .0367, so the probability of drawing a sample with \( M \geq 54 \) is .0367.

2. The sampling distribution of the mean will be normal because the population of raw scores is normal. The mean of this distribution, \( \mu_M \) will be equal to \( \mu \) and its standard deviation will be \( \sigma/\sqrt{n} = 4/\sqrt{10} = 1.26 \). In a normal distribution with \( \mu_M = 20 \) and \( \sigma_M = 1.26 \), a sample mean of 22 would correspond to a \( z \) score of \( (22-20)/1.26 = 1.59 \). Using the standard normal distribution (Table E.10 in Howell's book), the smaller portion of the normal distribution corresponding to a \( z \) value of 1.59 is .0559. We use the smaller portion because we are interested in sample means of 22 or larger and means in that range occupy the small portion of the distribution when it is cut along a \( z \) value of 1.59.

With a sample size of 30, the only thing that changes is the standard error of the mean: \( \sigma_M = 4/\sqrt{30} = 0.73 \). For \( M = 22 \), \( z = (22-20)/0.73 = 2.74 \); area in smaller portion = .0031. Notice how increasing the sample size makes it less likely to get a sample mean that is larger than \( \mu \) by 2 or more points.

3. A rejection criterion is the least extreme observed outcome that will allow rejection of the null hypothesis. A region of rejection represents the part(s) of a distribution of outcomes that will allow rejection of the null hypothesis (i.e., outcomes that are equal to, or more extreme than, the rejection criterion).

(a) The rejection criterion is the value of the sample mean that corresponds to a \( z \) score that cuts off the lower .05 of the standard normal distribution. From the standard normal distribution (\( z \) table, the closest to .05 we can get (without exceeding .05) when examining the smaller portion column is 1.65. Because we are working with the lower end of the distribution (the alternative hypothesis claims the true value of \( \mu \) is less than 50), the relevant \( z \) score is -1.65. That is the cutoff for rejecting \( H_0 \). The value of the sample mean corresponding to that cutoff is \( M = \mu + \sigma_M(z) \). \( \sigma_M = \sigma/\sqrt{n} = 10/\sqrt{30} = 1.83 \), so we have \( M = 50 + (1.83)(-1.65) = 46.98 \). This is the rejection criterion; the least extreme value of \( M \) that will allow rejection of \( H_0 \). The region of rejection is any sample mean with a value equal to or less than (more extreme than) 46.98.

(b) With a nondirectional test, there are two rejection criteria and two corresponding regions of rejection, one at each end of the distribution. The relevant \( z \) values cutting off the lower and upper .025 of the standard normal distribution are \( z = \pm 1.96 \). Combined, these two regions account for the most extreme .05 of the normal distribution. The rejection criteria are as follows. At the lower end, \( M = 50 + \)}
(1.83)(−1.96) = 46.41. At the upper end, \( M = 50 + (1.83)(1.96) = 53.59 \). The regions of rejection are values of \( M \) that are equal to or less than 46.41 and values that are equal to or greater than 53.59.

4. Copying the data into R as suggested would produce:

```r
> data = c(48, 45, 51, 41, 46, 48, 41, 58, 52, 51, 51, 53, 54, 55, 48, 57, 45, 47, 43, 48, 56, 52, 45, 51, 48, 32, 41, 50, 48, 51)
```

The `t.test` function with the correct parameters would be set up as follows. The first parameter is the variable containing the scores, the second parameter specifies a directional alternative hypothesis with the true value of \( \mu \) expected to be less than the null value, and the third parameter specifies the null hypothesis value for \( \mu \).

```r
> t.test(data, alternative = "less", mu = 50)
```

```
One Sample t-test

data:  data
t = -1.4522, df = 29, p-value = 0.0786
alternative hypothesis: true mean is less than 50
95 percent confidence interval:
   -Inf 50.24943
sample estimates:
mean of x
48.53333
```

The obtained \( t \) ratio is reported as -1.4522, with the negative sign implying that the sample mean is less than 50, which at least is in the direction expected by the alternative hypothesis. The last line of the output shows the sample mean (48.53). But is it small enough to reject the null hypothesis? The reported \( p \)-value of .0786 indicates that the obtained \( t \) ratio is located at a point below which the lower .0786 of the \( t \) distribution falls (see the diagram below). This means that the obtained \( t \) ratio is not among the .05 most extreme outcomes, and therefore we cannot reject the null hypothesis.

5. (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.

(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 \( (3.72^2 \text{ and } 4.58^2 \text{ equal } 13.84 \text{ and } 20.98, \text{ and the larger variance value is not at least 4 times larger than the smaller one}) \), 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

\[
S_{M_1–M_2} = \sqrt{\frac{18.00}{11} + \frac{18.00}{15}} = 1.68
\]

so \( t = (8.30–12.40)/1.68 = -2.44 \).

With \( \alpha = .05 \), one-tailed test with \( df = 24 \), 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.

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

7. (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 (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.

8. 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 H₁). 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 α = .05. If α = .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.

9. The two samples have very different variances (2.3² = 5.29 and 4.8² = 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 \( t = \frac{(22.0 - 19.5)}{1.18} = 2.12 \). 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.

10. 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 \( \sigma = 10 \) and \( n_1 = n_2 = 5 \)) is

\[ \sigma_{M_1-M_2} = \sqrt{\frac{100}{5} + \frac{100}{5}} = 6.325 \]
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
6.325. That is, when you compute $s_{M1-M2}$ from your samples, you are likely to find values somewhat less
than 6.325 (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.

11. 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 set your working directory to the
folder than contains the data file social.txt, which you will have retrieved from the course web site.

```r
> data = read.table("social.txt", header = T)
> library(psych)
> describe(data$Rel)
vars n mean sd median trimmed mad min max range skew kurtosis se
1  1 78 45.35  8.58 45 44.92  8.15 29 68 39 0.43 -0.16  0.97
> describe(data$NoRel)
vars n mean sd median trimmed mad min max range skew kurtosis se
1  1 48 50.08 10.73 50.5 50.52 11.86 25 69 44 -0.34 -0.64 1.55
```

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 nondirectional alternative hypothesis.

```r
> t.test(data$Rel, data$NoRel, alternative = "two.sided", paired = F, 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.