CHAPTER 9

DISTRIBUTION OF MEANS AND DIFFERENCES BETWEEN MEANS

In the last chapter we talked about the probability that a *single* observation or score "fits" into a normal distribution with an established mean and standard deviation. However, we are seldom interested in single observations, since we usually collect our data from *groups* of *S*s. When we have collected data from our group and computed the mean for the group, we need to make a judgment about whether that mean is exceptional—very high or very low—as compared with other mean scores. To make this judgment we must first identify our research as a Case I or Case II study.

In Case I studies we want to know whether our group characteristics are the same or different from those of the population at large. Case II studies are concerned with the comparison of two sample means (usually control and experimental groups). Case II comparisons are made to decide whether the two means are truly different. In this chapter we will once again use the z score formula to help us test these differences. However, for Case II studies with *small* sample size, we will use the t-test instead. That statistical procedure will be discussed in the next chapter.

CASE I STUDIES

When we run an experiment in our field, we are usually interested in whether some special treatment influences our dependent variable—for example, whether new techniques for teaching vocabulary will make a difference in vocabulary retention. We give our treatment and then we measure vocabulary retention on some kind of test, and we get our results which we display as a mean and standard deviation. We can then use inferential statistics to tell us how important our finding really is. Imagine that the same vocabulary test was given to learners in other programs. We had 30 *S*s; so we draw samples of 30 *S*s from many, many schools. Once we get the $\overline{X}$ scores from each of these samples, we can again turn them into a frequency distribution, a sampling distribution of means.

SAMPLING DISTRIBUTION OF MEANS

In Case I studies, our distribution will be made up not of individual scores but of the $\bar{X}$ scores found for each of the groups. If we have $\bar{X}$ scores from at least 30 schools, we know that once we start plotting this sampling distribution of means, it will begin to approach the bell-shaped curve of a normal distribution. *This normal distribution of means is referred to as the sampling distribution of means.* We can, then, take the mean from our school which was taught vocabulary using the special techniques and see where it falls in that normal distribution, the sampling distribution of means.

When we have given a special instructional treatment, it is our hope that our sample mean falls far to the right of the distribution so that we can say that our group is so much better that it does not "fit" the curve of the sampling distribution of means. That is, we think that if we had given every single school this treatment for vocabulary retention, we would have had a distribution of means which would have had a much higher central point.

Suppose, however, that the $\bar{X}$ for our group fell right in the middle of the sampling distribution of means. What would we have to say then? Obviously the techniques for teaching vocabulary retention didn't make that much difference in the long run, for the $\bar{X}$ of our group was quite typical of sample means drawn from many schools which did not have the benefit of the special instructional program.

Finally, imagine that the $\bar{X}$ for our group fell way down at the bottom of the distribution, at the left tail. Then we'd have to say that it looks as if our group doesn't really belong in the normal distribution of means either; it is significantly lower than the other means in the distribution. Maybe we should disband our treatment.

Sometimes we really do want to get results which place our group under the curve at the left side of the distribution. Consider the case where you want to show that some special treatment makes it possible for your *S*s to do a task much more rapidly than other groups. Then you would hope that your mean score for time would be significantly *lower* than that of other groups.

Now let's consider how we go about locating the $\bar{X}$ of our group on the sampling distribution of means. If we gave our teacher-trainees the applied linguistics examination that we mentioned in the preceding chapter, we could easily compute the $\bar{X}$ and the s for our group. We don't really expect to go out and administer the test to teacher-trainees in programs at other universities. However, if we did, one of the things that we would notice immediately is how much more similar the means are to each other than we might have guessed they'd be. But thinking about it for a minute, we'd probably predict this would be the case. We know that individual differences in scores seem to average out when we compute the $\bar{X}$ for a group; the high and the low scores disappear and the $\bar{X}$ is a more central figure. So, since our new distribution is made up of these means, it will be much more compact. The $\bar{X}$ scores will not spread out as far as might be expected. So, the standard deviation when the observations are mean

Figure 9-1

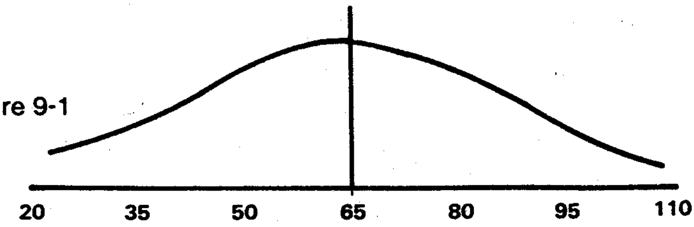


Figure 9-2

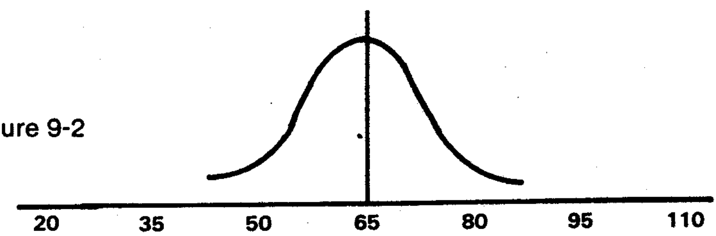
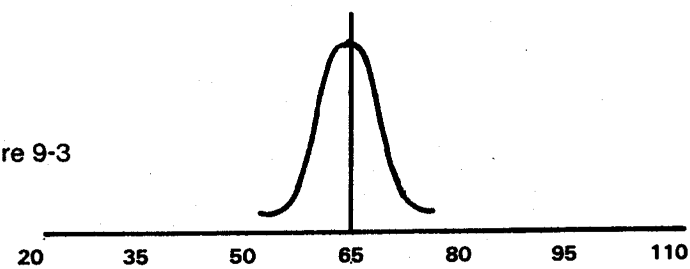


Figure 9-3



scores is always much smaller than the standard deviation of the original individual scores around the mean.

The size of the samples used in obtaining each mean will influence how much spread we are likely to find among the means. If the sample size (the number of *S*s in each of the samples) is large, the means will resemble each other very closely. So the *s* will be very small when there are lots of *S*s in each of the samples.

Perhaps a picture will make this difference, which is due to sample size, clearer; see Figure 9-1. Say we gave the applied linguistics test to a group of 36 teacher-trainees. Let's assume for a moment that the mean for our group was 65. The spread of individual scores is likely to be fairly wide; not everyone will score 65. The *s* might be 15. If we collected means from 36 *S*s in 30 other schools across the country and plotted these means, the distribution of the means for the groups would be much closer, as in Figure 9-2. If we collected data from, say, 100 *S*s at each school, the distribution would be even closer, as in Figure 9-3. The larger the size of the sample groups and the larger the number of sample groups, the closer the distribution of means will be to the central point of the distribution.

This change in the dispersion of scores is really the only thing that is different or new from the material presented in the preceding chapter. We have a distribution just as we did before. However, we call it a sampling distribution of means. We have a point of central tendency, a balance point for all the means. This is assumed to be the population mean; it is symbolized by μ (the Greek letter mu). The reason it is called the population mean is that we have drawn a large enough number of sample means from the population (and we have selected them at random from representative groups) that we have a normal distribution which allows us to make the assumption that the central point does equal the population mean.

The sampling distribution of means has three basic characteristics:

1. For 30 or more samples (with 30 or more *S*s per sample), it is normally distributed.
2. Its mean is equal to the mean of the population.
3. Its standard deviation, called the *standard error of means*, is equal to the standard deviation of the population divided by the square root of the sample size.

The third characteristic is the hard part. Just as we need a measure of central tendency for the population, we also need a measure of the dispersion of the sample means around the population mean. How can we find that? Unfortunately, all we have is the standard deviation for our sample. We can't use that, you'll remember, because individual scores are always spread out much more in a raw score distribution than means are spread out in a sampling distribution of means.

Since we know the standard deviation for our sample and since we know that the dispersion of means will be less than that of our sample, we can use our sample standard deviation to estimate the dispersion of scores for the population. Our sample standard deviation is a *statistic*, and we can use statistics to estimate those of the population. The estimate that we get for the population from this statistic will be called a *parameter*. *Statistic* refers to the *sample* and *parameter* refers to the *population*. A sample statistic is used to estimate a population parameter. The following diagram is simpler and perhaps clearer than this explanation:

Sample ↓ *Statistic* — *Estimator* ——→ *Parameter* ↑ *Population*

What we want to do is to use our sample standard deviation (statistic) to estimate the standard deviation for the means (parameter). To do that we have to make it sensitive to the size of the sample (not the number of samples but the size of the sample). This estimated standard deviation for the means which has been made sensitive to sample size is called the *standard error of means*. The symbol of the standard error of means as a parameter is the small sigma σ with the mean symbol slightly below it. The formula for the standard error of means is:

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{N}}$$

However, since we are dealing with sample data to estimate the parameter, we use our sample statistics for the formula:

$$s_{\overline{X}} = \frac{s_X}{\sqrt{N}}$$

The formula says that we can find the standard deviation of the means by dividing standard deviation of our sample by the square root of the sample size. From the first notation, above, you can see that we are talking about a population parameter while the second is our formula for estimating the parameter from our sample statistics. Sample statistics are used as the best unbiased estimate of the population parameters.

The standard error of means becomes a "ruler" for measuring the distance of the sample mean from the population mean in the same way that standard deviation was a "ruler" for measuring the distance of one score from the mean. The standard error "ruler," however, will always be very short in comparison with the standard deviation "ruler." The reason for this, you'll remember, is that a collection of means will always be similar to the population means—there's not much spread—while the spread of individual scores will make standard deviation "rulers" much larger.

Now that we have the measure of central tendency for a sample distribution of means and our "ruler" for dispersion of the means from the central point, the standard error of means, we can plug this information into our old z score formula. Remember that the z score formula is *distance of score from the mean ÷ standard deviation.* We can easily change this to deal with the sampling distribution of means. It becomes the z score of means formula: *distance of our mean from the population mean ÷ standard error of means.* To get the distance of our mean score from the population mean we just subtract. Using our new symbols, that's $\overline{X} - \mu$. And then divide it by the standard error of means.

Let's try part of this. Going back to the data from our applied linguistics exam, we gave the test to 36 of our teacher-trainees and let's say that our group got a mean score of 80. We know (because the test publishers told us so) that the population mean (μ) for the test is 65. So we subtract 80 − 65 and find that 15 is the distance of our mean from the population mean. We found that in our group of teacher-trainees the s for the distribution of scores was 30. So we can estimate the standard error of means by dividing the s by the square root of our sample size. Now that we have the μ, the s_X, and our sample $\overline{X}$, we can complete the computation. To do this we use the same z score formula as before:

$$z_{\bar{X}} = \frac{\bar{X} - \mu}{s_X / \sqrt{N}}$$

Let's put in our values:

$$z_{\bar{X}} = \frac{80 - 65}{30 / \sqrt{36}}$$

$$= \frac{15}{30 \div 6}$$

$$= \frac{15}{5}$$

$$= 3.0$$

The z score for our mean is 3. If you remember the critical value of z from the last chapter, you don't even have to look it up in the table to know that it is a very unusual score, that it doesn't really "fit" into the normal distribution of mean scores from all the other schools. We wouldn't expect any group to attain this high a mean score just by chance (our program is really good in preparing people for this test).

The symbols and numbers may be getting in the way, but if you stop and think about it for a moment, you will see that we are really doing exactly the same procedure as in the last chapter. We find the mean for the group; we look to see where it falls in the distribution (either above, at, or below μ, the mean of the population). We calculate the distance from μ and divide it by an amended measure of standard deviation. That gives us the "ruler," which we use in finding out how far the group mean is from μ, and allows us to see what proportion of the scores fall between it and μ. If it is far enough above μ, we know that this is not a probable score in this distribution, that it is substantially "better" than the population. If it is far enough below μ, we again know it doesn't really "fit" into that distribution.

The basic concept is exactly the same. The z score formula to find the value of the difference between an *individual* score and the group mean was

$$z = \frac{\text{difference between score and mean}}{\text{standard deviation}} \qquad z = \frac{X - \bar{X}}{s}$$

The z score formula to find the value of the difference between a *single sample mean* and the population mean is

$$z_{\bar{X}} = \frac{\text{difference between sample mean and population mean}}{\text{standard error of means}} \qquad z_{\bar{X}} = \frac{\bar{X} - \mu}{s_{\bar{X}}}$$

In using the z score formula we try to show that a mean which we have obtained does not truly "fit" that of the population. Our procedure has been to test the hypothesis that the mean has the same value as the population mean and reject that hypothesis when we found that the mean had a value sufficiently higher or lower than the population mean. Let's go through this procedure one more time to be sure that it is clear.

Imagine that you have hypothesized that foreign students entering American universities all score about the same on tests of reading speed. You could test a sample of 50 *S*s and another sample of 50 *S*s and another and so on. After you had collected approximately 30 such 50-member samples, you could take the 30 means and draw your sampling distribution of means. Suppose you found that the population mean was 350 WPM. You could draw another random sample of 50, and it is likely that you would find that it, too, fit nicely into this distribution of means, that the WPM would be around 340 to 360 WPM. Now suppose that in making your hypothesis what you really had in mind was that some particular group would differ from the population. You believe that *S*s with first languages which use different types of alphabets are not likely to receive scores similar to those of the population. Not only are you speculating that there will be a difference, you are also predicting the direction of that difference, that it will be lower. You remember the steps in hypothesis testing; so you first state the hypotheses:

Null hypothesis: The scores of *S*s with first languages which use a different alphabet system will be the same as the population mean.

Negative directional hypothesis: The scores of *S*s with first languages which use a different alphabet system will be lower than the population mean.

Then you select your probability level. Since you want to make the chances of error as small as possible, you pick the .01 level of probability. Then you go out and randomly select *S*s whose first language has a different alphabet (Arabic, Chinese, etc.). Your group size will be 50. You test them and compute their $\overline{X}$ for reading speed. The $\overline{X}$ is 325 WPM, 25 points less than the established population mean of 350 WPM, and the s is 100. So you put the values into the z score formula to determine whether the difference in means is a significant one:

$$z_{\overline{X}} = \frac{\overline{X} - \mu}{s_X / \sqrt{N}}$$

$$= \frac{325 - 350}{100 / \sqrt{50}}$$

$$= \frac{-25}{100 \div 7.07}$$

$$= \frac{-25}{14.14}$$

$$= -1.77$$

At this point, having done all the computations, you are going to be a bit upset when you check the probability level of a z score of 1.77. Since you selected the .01 level, you will not be able to reject the null hypothesis. To do that you would have needed a z value of 2.33 or more. If you had chosen the .05 level, you would have just barely made it. Since you chose the .01 level to be conservative in your claims, you must stay with that and say that the null hypothesis is correct: the scores of your sample group are not different from those of the population.

While it is possible that you may often need to compare sample data with the population mean, it is more likely that you will need to compare two sample means. Such studies are called Case II studies.

CASE II STUDIES

We can once again use the z score formula to compare means of two samples *provided* the size of the two samples is *large*. For small size samples we must use the t-test, which is discussed in the next chapter.

Imagine that we have gathered test data on reading scores from two matched groups of second grade bilingual children. They have been matched for degree of bilingualism, SES (socioeconomic status), IQ, and whatever other variables we decided were important. Then one S from each matched pair was randomly assigned to either treatment or control group. Both groups received reading instruction in both English and Spanish, but one class had reading instruction in each language on alternate days while the other group had reading in each language every day. (The total time devoted to reading was the same for both groups.) At the end of the year, we gathered test data on reading comprehension. One group's scores yield a mean of 41.6 and the other 38.9. Are these two means simply what you might get if you started testing bilingual children all over the state, or do the means reflect differences that can be associated with the difference in instructional format? In this example, we have no information about the reading scores on the test for the population of bilingual second-graders, nor do we have any information about the standard deviation for the population.

Once again, we will have to use our sample statistics to estimate the population parameters. Using our statistics, we must estimate the differences we'd get if we went out and tested another two classes and another two classes and another two classes until we felt that we had tested the population. We would compute the difference between our two sample means and use it to estimate the difference in means that we'd get for all the pairs of means in the population. We then could compare the differences we found between our two sample means and the estimated differences between means for the population. That would let us decide whether our difference in means is significant or just what one would expect from the population estimate. It is likely that in this case we would use the null hypothesis because we have no reason to believe that either format will work better than the other. That is, we don't believe we can predict that our differences for the two groups will be either larger or smaller than that of the population nor do we see any reason for proposing that it will be different. We decide on a probability level of .05 and then are ready to test the hypothesis. Since we are now comparing the difference between two sample means (rather than comparing one sample mean with the population mean), we will have to establish a new type of distribution, one which samples *differences between two sample means.*

SAMPLING DISTRIBUTION OF DIFFERENCES BETWEEN MEANS

Using our bilingual reading example, you know that if we collect reading scores from bilingual second grade classes, those scores are likely to be similar, but it is still *un*likely that any two samples we collect will be exactly the same. There will be some differences between the two means. We need to visualize collecting two samples over and over again from the total population and looking at the difference between the two means each time. We will then have an infinite number of difference scores between two means. If we construct a frequency distribution of all these differences, we will then have a *sampling distribution of differences between means.* This distribution (since we would have pulled two samples and found the difference between them many times over) would have the following properties according to central limit theorem:

1. It is normally distributed (bell-shaped).
2. It has a mean of zero.
3. It has a standard deviation called the *standard error of differences between means* (the formula will follow).

Since our distribution for the differences between means is a normal distribution, any value we find for differences between our two means can be converted into a z score and tested for significance. The concepts, you see, are once again the same. The z score to find the value of the difference between an *individual* score and a group mean was

$$z = \frac{\text{difference between score and mean}}{\text{standard deviation}} \qquad z = \frac{X - \overline{X}}{s}$$

The z score for the difference between a *sample mean* and the population mean was

$$z = \frac{\text{difference between sample mean and population mean}}{\text{standard error of means}}$$

$$z_{\overline{X}} = \frac{\overline{X} - \mu}{s_{\overline{X}}}$$

Now our formula for differences between *two sample means* will be

$$z = \frac{\text{difference between 2 means minus difference between 2 population means}}{\text{standard error of differences between means}}$$

Since we believe that the difference between our two population means is zero (they are from the same population), we can immediately simplify our formula by deleting the second part of the numerator. The formula becomes

$$z = \frac{\text{difference between 2 sample means}}{\text{standard error of differences between means}}$$

You can find the top half of the formula very easily. All you have to do is subtract one mean from the other. From our work so far, you probably have already guessed that we will use the sample standard deviation statistics to estimate the population parameter, standard error of differences between means. And you can guess, since it says "standard error," that we will have to make that estimate sensitive to the number of *S*s (or observations) in the two sample groups. Perhaps you can already predict what the formula will be, but let's stop for a moment and review by trying to do the first part of the process as we work through a new example.

Make believe that the Ministry of Education of Kenya has hired you to find out whether *S*s who received all their schooling in English do better on the school-leaving exam for 7th grade than those taught in vernacular languages in grades 1 to 3 and in English from grades 4 to 7. The exams are important since they determine which *S*s will continue their education in which schools. The school-leaving exams are in English. While waiting for the thousands of test figures to arrive, you decide to look at 300 scores from 20 schools which arrived early. 150 scores were for *S*s who attended vernacular-medium schools and 150 were from *S*s who attended English-medium schools instead. You computed the means for each group. Group 1, the vernacular-medium school *S*s, have a mean of 70.93 and a standard deviation of 18.27. Group 2, the English-medium school *S*s, have a mean of 72.07 and a standard deviation of 17.57.

Now that we have our means and standard deviations for the two groups we want to compare, let's think about the *z* score formula once again for comparing two means:

$$z = \frac{\text{difference between two sample means}}{\text{standard error of differences between two means}}$$

It's a simple matter to subtract our one sample mean from the other, but how do we find the standard error of differences between means? The formula using population parameters is

$$\sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{(\sigma_{\bar{X}_1})^2 + (\sigma_{\bar{X}_2})^2}$$

From Case I studies in this chapter, we know that

$$\sigma_{\bar{X}_1} = \sigma_{X_1} / \sqrt{n_1} \qquad \sigma_{\bar{X}_2} = \sigma_{X_2} / \sqrt{n_2}$$

Substituting these in the formula above, we have

$$\sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\left(\frac{\sigma_{X_1}}{\sqrt{n_1}}\right)^2 + \left(\frac{\sigma_{X_2}}{\sqrt{n_2}}\right)^2}$$

We will, however, use our sample statistics to estimate these population parameters. To estimate the population variance σ^2, we will use the sample variance s^2, but we will make that estimate sensitive to sample size by dividing it by the square root of sample size. The formula for estimating the standard error of differences between means of the population is

$$s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{(s_1 / \sqrt{n_1})^2 + (s_2 / \sqrt{n_2})^2} \quad \text{or, more simply,} \quad \sqrt{\left(\frac{s_1}{\sqrt{n_1}}\right)^2 + \left(\frac{s_2}{\sqrt{n_2}}\right)^2}$$

or, more simply

$$\sqrt{(s_1^2 / n_1) + (s_2^2 / n_2)}$$

From this formula for the standard error of differences between means we can estimate the variation we would find if we collected many, many two-group samples from the population. You may see both the population formula and the sample statistic formula in statistics books, but we will use the sample statistics to estimate the parameters.

To find the standard error of differences between our means, all we need is to plug our data into the formula:

$$s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$

$$= \sqrt{\frac{(18.27)^2}{150} + \frac{(17.57)^2}{150}}$$

$$= \sqrt{\frac{333.79}{150} + \frac{308.7}{150}}$$

$$= \sqrt{2.23 + 2.05}$$

$$= \sqrt{4.28}$$

$$= 2.07$$

And now the grand finale: we put our standard error of differences between means into our z score formula:

$$z_{(\bar{X}_1 - \bar{X}_2)} = \frac{\bar{X}_1 - \bar{X}_2}{s_{(\bar{X}_1 - \bar{X}_2)}}$$

$$= \frac{70.93 - 72.07}{2.07}$$

$$= -.55$$

The z score value from these preliminary data won't look very encouraging to anyone doing such an evaluation who hoped to find that it's better to start English early. It looks as if they are both from the same population at this point.

That is, the difference found so far between the two groups isn't very impressive. Of course, these are preliminary data; they may be from *non*typical schools, and so you may want to hold off making any judgments until all the data are in.

In this chapter, we have compared a number of means in example problems:

Applied Linguistics Test	$\overline{X} = 80$	and $\mu = 65$
Reading speed	$\overline{X} = 325$	and $\mu = 350$
Kenya schools	$\overline{X}_1 = 70.93$	and $\overline{X}_2 = 72.07$

If we just look at the two means each time, can we judge whether the differences between them are important or not? Sometimes the differences are so great that we are sure they are "real." But can we be certain? To judge the probability of finding these differences we cannot just look at the means themselves; we must subject the differences to formal analysis.

But, even when all the data are available and you have completed all the analyses, there is always the chance that you might make a mistake and reject the null hypothesis when you shouldn't or not reject it when you should. When we claim that a result is significant at the .05 level, it means that there are still 5 chances in 100 that we might be wrong. If we claim significance at the .01 level, there is still 1 chance in 100 we might be wrong. If we claim significance at the .001 level (and in social sciences that's almost bragging about how sure we are!), there is still 1 chance in 1,000 we could be wrong. While there are still chances that we are wrong, by doing the analysis we have drastically reduced the possibilities of making a mistake. When important decisions are to be made on the basis of our research, it is mandatory that we feel confident that our claims are correct.

So far, we have used *z* scores to test the importance of differences between group data and the population or between sets of means. The *z* score distribution, however, will not always be appropriate for our research. Unless you work in a state agency or a ministry of education, it is unlikely that you will have available large groups of *S*s for your research. The *z* score distribution is based on a normal distribution; so the larger the sample size, the better. When we have very few *S*s, our chance of getting a normal distribution is not great enough to make us feel comfortable about using the *z* score formulas. Fortunately, mathematicians have worked out a distribution that will take care of the problem of small sample size. When working with small numbers of *S*s, we will use the *t*-distribution instead of the *z*-distribution. The *t*-test will be presented in the next chapter.

ACTIVITIES

1. A report published by the Antarctic Academy of Neurolinguistics indicates the left-handed people may be better language learners than right-handed people. To test this hypothesis, you did the following experiments:

   In an ESL center, 250 *S*s were randomly selected and then assigned to two groups (left-handed vs. right-handed). After equal amounts of instruction, you administered a battery of language tests to all *S*s. The information you obtained was:

$\bar{X}$ left-handed	55	$s=10$
$\bar{X}$ right-handed	45	$s=10$

   If the mean on the test for the population of ESL students is 50:
   a. Test whether left-handed people are better than the population.
   b. Test whether right-handed people are better than the population.
   c. Test whether left-handed people are better than right-handed people.

$$z_{left} = \frac{\bar{X}-\mu}{s_{\bar{X}}} = \frac{55-50}{10/\sqrt{125}} \qquad z_{right} = \frac{\bar{X}-\mu}{s_{\bar{X}}} = \frac{45-50}{s_{\bar{X}}}$$

   Set α at .05 for a two-tailed hypothesis.

$$z_{(l-r)} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(s_1^2/n_1)+(s_2^2/n_2)}} = \frac{55-45}{\sqrt{(10)^2/125+(10)^2/125}}$$

2. A large population of *S*s were tested for language aptitude. The population scores are normally distributed with a $\bar{X}$ of 152.68 and s of 40.22. If a sample of 50 scores is selected, what is the probability that the sample is over 160?
3. The population mean for second language learners on a timed vocabulary test is 200 with a σ of 40. Would a set of 140 scores, randomly selected from the population with a mean of 195 fit in the population distribution?
4. State the difference between a one-tailed and a two-tailed test and clarify when and under what circumstances a researcher would select one over the other.
5. State the difference between null and alternative hypotheses.
6. The population mean on the GRE (Graduate Record Examination) is reported to be 500. Two samples with an N of 100 are selected and given the test. The means were: $\bar{X}_1 = 560, s_1 = 80$ and $\bar{X}_2 = 520, s_2 = 100$. We want to see whether the groups are different from the population and from one another.
   a. State the null hypothesis.
   b. Test for the difference between $\bar{X}_1$ and the population, $\bar{X}_2$ and the population, and $\bar{X}_1$ vs. $\bar{X}_2$ at the .01 level.
   c. If the sample sizes were 30, what differences would have occurred?
   d. If α were set at .05, one-tailed, what difference would have occurred?

Suggested further reading for this chapter: Johnson, Slakter.