In Chapter 7, we dealt with hypothesis testing in the one-sample case. In that situation, we were concerned with the significance of the difference between a sample statistic and a population parameter. In this chapter, we consider a new research situation where we will be concerned with the significance of the difference between two separate populations. For example, do men and women in the United States vary in their support for gun control? Obviously, we cannot ask every male and female for their opinions on this issue. Instead, we must draw random samples of both groups and then use the information gathered from these samples to infer population patterns.

The question in hypothesis testing in the two-sample case is: Is the difference between the samples large enough to allow us to conclude (with a known probability of error) that the populations represented by the samples are different? Thus, if we find a large enough difference in support of gun control between random samples of men and women, we can argue that the difference between the samples did not occur by random chance but, rather, represents a real difference between men and women in the population.

In this chapter, we consider tests for the significance of the difference between sample means and sample proportions. In both tests, the five-step model will be the framework for our decision making. The hypothesis-testing process is very similar to that of the one-sample case, but we also need to consider some important differences.

The One-Sample Case versus the Two-Sample Case

There are three important differences between the one-sample case considered in Chapter 7 and the two-sample case covered in this chapter. First, the one-sample case requires that the sample be selected following the principle of EPSEM (each case in the population must have an equal chance of being selected for the sample). The two-sample case requires that the samples be selected independently as well as randomly. This means that the procedure for selecting cases for one of the samples cannot affect the probability that any particular case will be selected for the other sample. In our example concerning gender differences in support of gun control, this would mean that the selection of a specific male for the sample would have no effect on the probability of selecting any particular female. This new requirement will be stated as **independent random sampling** in step 1.

The requirement of independent random sampling can be satisfied by drawing EPSEM samples from separate lists (for example, one for females and one for males). However, it is usually more convenient to draw a single EPSEM sample from a single list of the population and then to subdivide the cases into separate groups (males and females, for example). As long as the original sample is selected randomly, any subsamples created by the researcher will meet the assumption of independent random samples.

The second important difference is in the form of the null hypothesis stated in step 2 of the five-step model. The null is still a statement of "no difference." Now, however, instead of saying that the population from which the sample is drawn has a certain characteristic, it will say that the two populations are the same. ("There is no significant difference between men and women in their support of gun control.") If the test statistic falls in the critical region, the null hypothesis of no difference between the populations can be rejected and the argument that the populations are different will be supported.

A third important new element concerns the sampling distribution or the distribution of all possible sample outcomes. In Chapter 7, the sample outcome was either a mean or a proportion. Now we are dealing with two samples (e.g., samples of men and women), and the sample outcome is the *difference between* the sample statistics. In terms of our example, the sampling distribution would include all possible differences in sample means for support of gun control between men and women. If the null hypothesis is true and men and women do *not* have different views about gun control, the difference between the population means would be zero, the mean of the sampling distribution would be zero, and the huge majority of differences between sample means would be zero or very close to zero. The greater the differences between the sample means, the further the sample outcome (the *difference* between the two sample means) will be from the mean of the sampling distribution (zero) and the more likely that the difference reflects a real difference between the populations represented by the samples.

Hypothesis Testing With Sample Means (Large Samples)

To illustrate the procedure for testing sample means, assume that a researcher has access to a nationally representative random sample and that the individuals in the sample have answered a survey that measures attitudes toward gun control. The sample is divided by sex, and sample statistics are computed for males and females. Assuming that the survey yields interval-ratio-level data, a test for the significance of the difference in sample means can be conducted.

As long as the sample size is large (that is, as long as the combined number of cases in the two samples exceeds 100), the sampling distribution of the differences in sample means will be normal and the normal curve (Appendix A) can be used to establish the critical regions. The test statistic—*Z*(obtained)—will be computed by the usual formula: sample outcome (the difference between the sample means) minus the mean of the sampling distribution divided by the standard deviation of the sampling distribution. The formula is presented as Formula 8.1. Note that numerical subscripts are used to identify the samples and populations. The subscript attached to $\sigma\,(\sigma_{\bar{X}-\bar{X}})$ indicates that we are dealing with the sampling distribution of the *differences* in sample means:

FORMULA 8.1

$$Z\,(\text{obtained}) = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}-\bar{X}}}$$

where: $(\bar{X}_1 - \bar{X}_2)$ = the difference between the sample means
$(\mu_1 - \mu_2)$ = the difference between the population means
$\sigma_{\bar{X}-\bar{X}}$ = the standard deviation of the sampling distribution of sample means

Recall that tests of significance are always based on the assumption that the null hypothesis is true. If the means of the two populations are equal, then the term ($\mu_1 - \mu_2$) will be zero and can be dropped from the equation. In effect, then, the formula we will actually use to compute the test statistic in step 4 will be:

FORMULA 8.2

$$Z(\text{obtained}) = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}-\bar{X}}}$$

For large samples, the standard deviation of the sampling distribution of the difference in sample means is defined as:

FORMULA 8.3

$$\sigma_{\bar{X}-\bar{X}} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

Because we will rarely, if ever, know the values of the population standard deviations (σ_1 and σ_2), we must use the sample standard deviations, corrected for bias, to estimate them. Formula 8.4 displays the equation used to estimate the standard deviation of the sampling distribution in this situation. This is called a **pooled estimate** because it combines information from both samples:

FORMULA 8.4

$$\sigma_{\bar{X}-\bar{X}} = \sqrt{\frac{s_1^2}{N_1 - 1} + \frac{s_2^2}{N_2 - 1}}$$

The sample outcomes for support of gun control are

Sample 1 (Men)	Sample 2 (Women)
$\bar{X}_1 = 6.2$	$\bar{X}_2 = 6.5$
$s_1 = 1.3$	$s_2 = 1.4$
$N_1 = 324$	$N_2 = 317$

We see from the sample statistics that men have a lower average score and are less supportive of gun control. The test of hypothesis will tell us if this difference is large enough to conclude that it did not occur by random chance alone but, rather, reflects an actual difference between the populations of men and women on this issue.

Step 1. Making Assumptions and Meeting Test Requirements. We now assume that the random samples are independent but the rest of the model is the same as in the one-sample case.

Model: Independent random samples
Level of measurement is interval-ratio
Sampling distribution is normal

Step 2. Stating the Null Hypothesis. The null hypothesis states that the *populations* represented by the samples are not different on this variable. No direction for the difference has been predicted, so a two-tailed test is called for:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Step 3. Selecting the Sampling Distribution and Establishing the Critical Region. For large samples, the Z distribution can be used to find areas under the sampling distribution and establish the critical region. Alpha will be set at 0.05:

$$\text{Sampling distribution} = Z \text{ distribution}$$
$$\text{Alpha} = 0.05$$
$$Z(\text{critical}) = \pm 1.96$$

Step 4. Computing the Test Statistic. The population standard deviations are unknown, so Formula 8.4 will be used to estimate the standard deviation of the sampling distribution. This value will then be substituted into Formula 8.2 and Z(obtained) will be computed:

$$\sigma_{\bar{X}-\bar{X}} = \sqrt{\frac{s_1^2}{N_1 - 1} + \frac{s_2^2}{N_2 - 1}}$$
$$\sigma_{\bar{X}-\bar{X}} = \sqrt{\frac{(1.3)^2}{324 - 1} + \frac{(1.4)^2}{317 - 1}}$$
$$\sigma_{\bar{X}-\bar{X}} = \sqrt{(0.0052) + (0.0062)}$$
$$\sigma_{\bar{X}-\bar{X}} = \sqrt{0.0114}$$
$$\sigma_{\bar{X}-\bar{X}} = 0.107$$
$$Z(\text{obtained}) = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}-\bar{X}}}$$
$$Z(\text{obtained}) = \frac{6.2 - 6.5}{0.107}$$
$$Z(\text{obtained}) = \frac{-0.300}{0.107}$$
$$Z(\text{obtained}) = -2.80$$

Step 5. Making a Decision and Interpreting the Results of the Test. Comparing the test statistic with the critical region:

$$Z(\text{obtained}) = -2.80$$
$$Z(\text{critical}) = \pm 1.96$$

We see that the Z score falls into the critical region, which means that a difference as large as -0.30 $(6.2 - 6.5)$ between the sample means is unlikely if the null hypothesis is true. The null hypothesis of no difference can be rejected, and the notion that men and women are different in terms of their support of gun control is supported. The decision to reject the null hypothesis has only a 0.05 probability (the alpha level) of being incorrect.

Note that the value for Z(obtained) is negative, indicating that men have significantly lower scores than women for support for gun control. The sign of the test statistics reflects our arbitrary decision to label men sample 1 and women sample 2. If we had reversed the labels and called women sample 1 and men sample 2, the sign of the Z(obtained) would have been positive, but its value (2.80) would have been exactly the same, as would our decision in step 5. *(For practice in testing the significance of the difference between sample means for large samples, see problems 8.1 to 8.6, 8.9, 8.15d to f, and 8.16d to f.)*

ONE STEP AT A TIME — Completing Step 4 of the Five-Step Model: Computing Z(obtained)

Use these procedures when samples are large.
Solve Formula 8.4 first and then solve Formula 8.2.

Step	***Operation***
To Solve Formula 8.4	
1.	Subtract 1 from N_1.
2.	Square the value of the standard deviation for the first sample (s_1^2).
3.	Divide the quantity you found in step 2 by the quantity you found in step 1.
4.	Subtract 1 from N_2.
5.	Square the value of the standard deviation for the second sample (s_2^2).
6.	Divide the quantity you found in step 5 by the quantity you found in step 4.
7.	Add the quantity you found in step 6 to the quantity you found in step 3.
8.	Take the square root of the quantity you found in step 7.
To Solve Formula 8.2	
1.	Subtract $\bar{X}_2$ from $\bar{X}_1$.
2.	Divide the value you found in step 1 by the quantity you found in step 8 of "To Solve Formula 8.4." This is Z(obtained).

ONE STEP AT A TIME — Completing Step 5 of the Five-Step Model: Making a Decision and Interpreting Results

Step	***Operation***
1.	Compare the Z(obtained) to the Z(critical). If Z(obtained) is *in* the critical region, *reject* the null hypothesis. If Z(obtained) is *not in* the critical region, *fail to reject* the null hypothesis.
2.	Interpret the decision to reject or fail to reject the null hypothesis in terms of the original question. For example, our conclusion for the example problem was "There is a significant difference between men and women in their support for gun control."

Applying Statistics 8.1 A Test of Significance for Sample Means

A scale measuring satisfaction with family life has been administered to a sample of married respondents. On this scale, higher scores indicate greater satisfaction. The sample has been divided into respondents with no children and respondents with at least one child, and means and standard deviations have been computed for both groups. Is there a significant difference in satisfaction with family life between these two groups?

The sample information is:

Sample 1 (No Children)	Sample 2 (At Least One Child)
$\bar{X}_1 = 11.3$	$\bar{X}_2 = 10.8$
$s_1 = 0.6$	$s_2 = 0.5$
$N_1 = 78$	$N_2 = 93$

Applying Statistics 8.1 *(continued)*

We can see from the sample results that respondents with no children are more satisfied. Is this difference significant?

Step 1. Making Assumptions and Meeting Test Requirements.

Model: Independent random samples
Level of measurement is interval-ratio
Sampling distribution is normal

Step 2. Stating the Null Hypothesis.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Step 3. Selecting the Sampling Distribution and Establishing the Critical Region.

Sampling distribution = Z distribution
Alpha = 0.05, two-tailed
$Z(\text{critical}) = \pm 1.96$

Step 4. Computing the Test Statistic.

$$\sigma_{\bar{X}-\bar{X}} = \sqrt{\frac{s_1^2}{N_1 - 1} + \frac{s_2^2}{N_2 - 1}}$$

$$\sigma_{\bar{X}-\bar{X}} = \sqrt{\frac{(0.6)^2}{78 - 1} + \frac{(0.5)^2}{93 - 1}}$$

$$\sigma_{\bar{X}-\bar{X}} = \sqrt{0.008}$$

$$\sigma_{\bar{X}-\bar{X}} = 0.09$$

$$Z(\text{obtained}) = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{X}-\bar{X}}}$$

$$Z(\text{obtained}) = \frac{11.3 - 10.8}{0.09}$$

$$Z(\text{obtained}) = \frac{0.50}{0.09}$$

$$Z(\text{obtained}) = 5.56$$

Step 5. Making a Decision and Interpreting the Results of the Test. Comparing the test statistic with the critical region,

$$Z(\text{obtained}) = 5.56$$
$$Z(\text{critical}) = \pm 1.96$$

we reject the null hypothesis. Parents and childless couples are significantly different in their satisfaction with family life. Given the direction of the difference, we can also note that childless couples are significantly happier.

STATISTICS IN EVERYDAY LIFE

The reading scores of American children have not improved significantly over recent years. Tests are administered every other year to very large, representative samples of about 180,000 fourth-graders and scores on the test range from 0 to 500. The average score in 2009 was 221, which was significantly higher than the 2005 average (219) but exactly the same as the 2007 average (also 221).* The tests were conducted with alpha set at 0.05. A different report# concludes that girls, on average, have caught up with boys in math skills (that is, there is no significant difference between the average math scores for girls and boys) and remain well ahead of boys in verbal and reading skills. What sociological factors might explain these patterns?

*This report is available at http://nationsreportcard.gov/reading_2009/reading_2009_report.

#This report is available at http://www.cep-dc.org/document/docWindow.cfm?fuseaction=document.viewDocument&documentid=304&documentFormatId=4644.

Hypothesis Testing With Sample Means (Small Samples)

When the population standard deviation is unknown and the sample size is small (combined *N*s of less than 100), the *Z* distribution cannot be used to find areas under the sampling distribution. Instead, we will use the *t* distribution to find the critical region and identify unlikely sample outcomes. To do this, we need to perform one additional calculation and make one additional assumption. The calculation is for degrees of freedom, which we need to use the *t* table (Appendix B). In the two-sample case, degrees of freedom are equal to $N_1 + N_2 - 2$.

The additional assumption is a more complex matter. When samples are small, we must assume that the population variances are equal in order to justify the assumption of a normal sampling distribution and to form a pooled estimate of the standard deviation of the sampling distribution. The assumption of equal variance in the population can be tested by a statistical technique known as the *analysis of variance*, or ANOVA (see Chapter 9). However, for our purposes here, we will simply assume equal population variances without formal testing. This assumption is safe as long as sample sizes are approximately equal.

The Five-Step Model and the *t* Distribution To illustrate this procedure, assume that a researcher believes that center-city families have significantly more children than suburban families. Random samples from both areas are gathered and the following sample statistics computed:

Sample 1 (Suburban)	Sample 2 (Center-City)
$\bar{X}_1 = 2.37$	$\bar{X}_2 = 2.78$
$s_1 = 0.63$	$s_2 = 0.95$
$N_1 = 42$	$N_2 = 37$

The sample data show a difference in the predicted direction. The significance of this observed difference can be tested with the five-step model.

Step 1. Making Assumptions and Meeting Test Requirements. The sample size is small and the population standard deviation is unknown. Hence, we must assume equal population variances in the model.

Model: Independent random samples
Level of measurement is interval-ratio
Population variances are equal
Sampling distribution is normal

Step 2. Stating the Null Hypothesis. Because a direction has been predicted, a one-tailed test will be used. The research hypothesis is stated accordingly:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 < \mu_2$$

Step 3. Selecting the Sampling Distribution and Establishing the Critical Region. With small samples, the *t* distribution is used to establish the critical region. Alpha will be set at 0.05 and a one-tailed test will be used:

$$\begin{aligned} \text{Sampling distribution} &= t \text{ distribution} \\ \text{Alpha} &= 0.05, \text{ one-tailed} \\ \text{Degrees of freedom} &= N_1 + N_2 - 2 = 42 + 37 - 2 = 77 \\ t(\text{critical}) &= -1.671 \end{aligned}$$

Note that the critical region is placed in the lower tail of the sampling distribution in accordance with the direction specified in H_1.

Step 4. Computing the Test Statistic. With small samples, a different formula (Formula 8.5) is used for the pooled estimate of the standard deviation of the sampling distribution. This value is then substituted directly into the denominator of the formula for *t*(obtained) in Formula 8.6:

FORMULA 8.5

$$\sigma_{\bar{X}-\bar{X}} = \sqrt{\frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2 - 2}}\sqrt{\frac{N_1 + N_2}{N_1 N_2}}$$

$$\sigma_{\bar{X}-\bar{X}} = \sqrt{\frac{(42)(.63)^2 + (37)(.95)^2}{42 + 37 - 2}}\sqrt{\frac{42 + 37}{(42)(37)}}$$

$$\sigma_{\bar{X}-\bar{X}} = \sqrt{\frac{50.06}{77}}\sqrt{\frac{79}{1{,}554}}$$

$$\sigma_{\bar{X}-\bar{X}} = (.81)(.23)$$

$$\sigma_{\bar{X}-\bar{X}} = 0.19$$

FORMULA 8.6

$$t(\text{obtained}) = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}-\bar{X}}}$$

$$t(\text{obtained}) = \frac{2.37 - 2.78}{0.19}$$

$$t(\text{obtained}) = \frac{-0.41}{0.19} = -2.16$$

$$t(\text{obtained}) = -2.16$$

Step 5. Making a Decision and Interpreting the Results of the Test. The test statistic falls in the critical region:

$$\begin{aligned} t(\text{obtained}) &= -2.16 \\ t(\text{critical}) &= -1.671 \end{aligned}$$

ONE STEP AT A TIME **Step 4 of the Five-Step Model: Computing *t*(obtained)**

Solve Formula 8.5 first and then solve Formula 8.6 to compute the test statistic.

Step	***Operation***
To Solve Formula 8.5	
1.	Add N_1 and N_2 and then subtract 2 from this total.
2.	Square the standard deviation for the first sample (s_1^2) and then multiply the result by N_1.
3.	Square the standard deviation for the second sample (s_2^2) and then multiply the result by N_2.
4.	Add the quantities you found in steps 2 and 3.
5.	Divide the quantity you found in step 4 by the quantity you found in step 1 and then take the square root of the result.
6.	Multiply N_1 by N_2.
7.	Add N_1 and N_2.
8.	Divide the quantity you found in step 7 by the quantity you found in step 6 and then take the square root of the result.
9.	Multiply the quantity you found in step 8 by the quantity you found in step 5.
To Solve Formula 8.6	
1.	Subtract $\bar{X}_2$ from $\bar{X}_1$.
2.	Divide the difference between the sample means—the quantity you found in step 1—by the quantity you found in step 9 of "To Solve Formula 8.5." This is *t*(obtained).

If the null hypothesis ($\mu_1 = \mu_2$) were true, this would be a very unlikely outcome, so the null hypothesis can be rejected. There is a statistically significant difference (a difference so large that it is unlikely to be due to random chance) in the sizes of center-city and suburban families. Furthermore, center-city families are significantly larger in size. The test statistic and sampling distribution are depicted in Figure 8.1. *(For practice in testing the significance of the difference between sample means for small samples, see problems 8.7 and 8.8.)*

FIGURE 8.1 The Sampling Distribution With Critical Region and Test Statistic Displayed