Section 7-1

Estimating a Population Proportion

INFERENTIAL STATISTICS

This chapter presents the beginnings of inferential statistics. The two major applications of inferential statistics are:

1. Use sample data to estimate the values of population parameters (such as a population proportion or population mean.)
2. Test hypotheses (or claims) made about population parameters.

INFERENTIAL STATISTICS (CONTINUED)

This chapter deals with the first of these.

1. We introduce methods for estimating values of these important population parameters: proportions, means, and variances.
2. We also present methods for determining sample sizes necessary to estimate those parameters.

DEFINITIONS

- An **estimator** is a formula or process for using sample data to estimate a population parameter.
- An **estimate** is a specific value or range of values used to approximate a population parameter.
- A **point estimate** is a single value (or point) used to approximate a population parameter.

ASSUMPTIONS FOR ESTIMATING A PROPORTION

We begin this chapter by estimating a population proportion. We make the following assumptions:

1. The sample is a simple random sample.
2. The conditions for the binomial distribution are satisfied. (See Section 5-2.)
3. There are at least 5 successes and 5 failures.

NOTATION FOR PROPORTIONS

$p =$ ***population*** proportion

$\hat{p} = \frac{x}{n} =$ ***sample*** proportion of x ***successes*** in a sample of size n.

$\hat{q} = 1 - \hat{p} =$ ***sample*** proportion of ***failures*** in a sample of size n.

POINT ESTIMATE

A **point estimate** is a single value (or point) used to approximate a population parameter.

The sample proportion $\hat{p}$ is the best point estimate of the population proportion p.

CONFIDENCE INTERVALS

A **confidence interval** (or **interval estimate**) is a range (or an interval) of values used to estimate the true value of a population parameter. A confidence interval is sometimes abbreviated as CI.

CONFIDENCE LEVEL

A **confidence level** is the probability $1 - \alpha$ (such as 0.95, or 95%) that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times. (The confidence level is also called the **degree of confidence**, or the **confidence coefficient**.) Some common confidence levels are:

90% or 0.90	95% or 0.95	99% or 0.99
(α = 10%)	(α = 5%)	(α = 1%)

CAUTIONS ABOUT CONFIDENCE INTERVALS

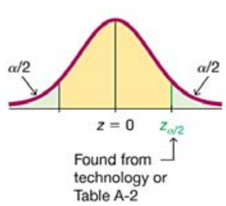
- Know the correct interpretation of a confidence interval: "We are 95% certain that the interval actually does contain the true value of the population proportion p."
- Confidence intervals can be used informally to compare different data sets, but *the overlapping of confidence intervals should not be used for making formal and final conclusions about the equality of proportions*.

CRITICAL VALUES

1. When certain requirements are met, the sampling distribution of sample proportions can be approximated by a normal distribution. (See figure on next slide.)
2. A z score associated with a sample proportion has a probability of $\alpha/2$ of falling in the right tail of Figure 7-2.
3. The z score at the boundary of the right-tail is commonly denoted by $z_{\alpha/2}$, and is referred to as a **critical value** because it is on the borderline separating z scores that are significantly high.

CRITICAL VALUE

A **critical value** is the number on the borderline separating sample statistics that are likely to occur from those that are unlikely to occur. The number $z_{\alpha/2}$ is a critical value that is a z score with the property that it separates an area of $\alpha/2$ in the right tail of the standard normal distribution.

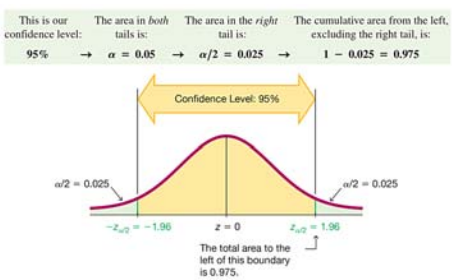


NOTATION FOR CRITICAL VALUE

The critical value $z_{\alpha/2}$ is the positive z value that is at the vertical boundary separating an area of $\alpha/2$ in the right tail of the standard normal distribution. (The value of $-z_{\alpha/2}$ is at the vertical boundary for the area of $\alpha/2$ in the left tail). The subscript $\alpha/2$ is simply a reminder that the z score separates an area of $\alpha/2$ in the right tail of the standard normal distribution.

FINDING $z_{\alpha/2}$ FOR 95% DEGREE OF CONFIDENCE

Note that when finding the critical z score for a 95% confidence level, we use a cumulative left area of 0.9750 (*not* 0.95). Think of it this way:



MARGIN OF ERROR

When data from a simple random sample are used to estimate a population proportion p, the difference between the sample proportion $\hat{p}$ and the population proportion p is an error. The maximum likely amount of that error is the **margin of error**, denoted by E. There is a probabilityof $1-\alpha$ (such as 0.95) that the difference between $\hat{p}$ and p is E or less. The margin of error E is also called the *maximum error of the estimate* and can be found using the formula on the following slide.

MARGIN OF ERROR OF THE ESTIMATE FOR *p*

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

<u>NOTE</u>: n is the size of the sample.

CONFIDENCE INTERVAL FOR THE POPULATION PROPORTION *p*

$\hat{p} - E < p < \hat{p} + E$ where $E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$

The confidence interval is often expressed in the following two equivalent formats:

$$\hat{p} \pm E$$

or

$$(\hat{p} - E, \hat{p} + E)$$

ROUND-OFF RULE FOR CONFIDENCE INTERVALS

Round the confidence interval limits to

three significant digits.

PROCEDURE FOR CONSTRUCTING A CONFIDENCE INTERVAL

1. Verify that the required assumptions are satisfied. (The sample is a simple random sample, the conditions for the binomial distribution are satisfied, and the normal distribution can be used to approximate the distribution of sample proportions because there are at least 5 successes and at least 5 failures.)
2. Use Table A-2 or technology to find the critical value $z_{\alpha/2}$ that corresponds to the desired confidence level.
3. Evaluate the margin of error $E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$

4. Using the calculated margin of error E and the value of the sample proportion $\hat{p}$, find the values of the **confidence interval limits** $\hat{p} - E$ and $\hat{p} + E$. Substitute those values in the general format for the confidence interval: $\hat{p} - E < p < \hat{p} + E$
5. Round the resulting confidence interval limits to three significant digits.

FINDING THE POINT ESTIMATE AND *E* FROM A CONFIDENCE INTERVAL

Point estimate of p:

$$\hat{p} = \frac{(\text{upper confidence limit}) + (\text{lower confidence limit})}{2}$$

Margin of error:

$$E = \frac{(\text{upper confidence limit}) - (\text{lower confidence limit})}{2}$$

FINDING A CONFIDENCE INTERVAL USING TI-83/84

1. Select **STAT**.
2. Arrow right to **TESTS**.
3. Select **A:1–PropZInt….**
4. Enter the number of *successes* as x.
5. Enter the size of the sample as n.
6. Enter the Confidence Level.
7. Arrow down to **Calculate** and press **ENTER**.

NOTE: If the sample proportion is given, you must first compute the number of successes by multiplying the sample proportion (as a decimal) by the sample size. You must round to the nearest integer.

SAMPLE SIZES FOR ESTIMATING A PROPORTION *p*

When an estimate $\hat{p}$ is known: $n = \dfrac{\left[z_{\alpha/2}\right]^2 \hat{p}\hat{q}}{E^2}$

When no estimate $\hat{p}$ is known: $n = \dfrac{\left[z_{\alpha/2}\right]^2 \cdot 0.25}{E^2}$

If a reasonable estimate of $\hat{p}$ can be made by using a previous sample, a pilot study, or someone's expert knowledge, use the first formula. If nothing is known about the value of $\hat{p}$, use the second formula.

ROUND-OFF RULE FOR DETERMINING SAMPLE SIZE

In order to ensure that the required sample size is at least as large as it should be, if the computed sample size is not a whole number,

round up to the next *higher* whole number.

CAUTION

Try to avoid these three common errors when calculating sample size.

1. Don't make the mistake of using $E = 3$ as the margin of error corresponding to "three percentage points." If the margin of error is three percentage points, use $E = 0.03$.
2. Be sure to substitute the critical z score for $z_{\alpha/2}$. For example, when working with 95% confidence, be sure to replace $z_{\alpha/2}$ with 1.96. Don't make the mistake of replacing $z_{\alpha/2}$ with 0.95 or 0.05.
3. Be sure to round up to the next higher integer; don't round using the usual rounding rules. Round 1067.11 to 1068.