**CONFIDENCE INTERVALS: INTRODUCTION**

Susan Dean  
Barbara Illowsky, Ph.D.

This work is produced by The Connexions Project and licensed under the  
Creative Commons Attribution License †

1 Student Learning Objectives

By the end of this chapter, the student should be able to:

- Calculate and interpret confidence intervals for one population average and one population proportion.
- Interpret the student-t probability distribution as the sample size changes.
- Discriminate between problems applying the normal and the student-t distributions.

2 Introduction

Suppose you are trying to determine the average rent of a two-bedroom apartment in your town. You might look in the classified section of the newspaper, write down several rents listed, and average them together. You would have obtained a point estimate of the true mean. If you are trying to determine the percent of times you make a basket when shooting a basketball, you might count the number of shots you make and divide that by the number of shots you attempted. In this case, you would have obtained a point estimate for the true proportion.

We use sample data to make generalizations about an unknown population. This part of statistics is called **inferential statistics**. The sample data help us to make an estimate of a population parameter. We realize that the point estimate is most likely not the exact value of the population parameter, but close to it. After calculating point estimates, we construct confidence intervals in which we believe the parameter lies.

In this chapter, you will learn to construct and interpret confidence intervals. You will also learn a new distribution, the Student-t, and how it is used with these intervals.

If you worked in the marketing department of an entertainment company, you might be interested in the average number of compact discs (CD’s) a consumer buys per month. If so, you could conduct a survey and calculate the sample average, \( \bar{x} \), and the sample standard deviation, \( s \). You would use \( \bar{x} \) to estimate the population mean and \( s \) to estimate the population standard deviation. The sample mean, \( \bar{x} \), is the **point estimate** for the population mean, \( \mu \). The sample standard deviation, \( s \), is the point estimate for the population standard deviation, \( \sigma \).

Each of \( \bar{x} \) and \( s \) is also called a statistic.

A **confidence interval** is another type of estimate but, instead of being just one number, it is an interval of numbers. The interval of numbers is an estimated range of values calculated from a given set of sample data. The confidence interval is likely to include an unknown population parameter.
Suppose for the CD example we do not know the population mean $\mu$ but we do know that the population standard deviation is $\sigma = 1$ and our sample size is 100. Then by the Central Limit Theorem, the standard deviation for the sample mean is 

$$\frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{100}} = 0.1.$$ 

The Empirical Rule, which applies to bell-shaped distributions, says that in approximately 95% of the samples, the sample mean, $\bar{x}$, will be within two standard deviations of the population mean $\mu$. For our CD example, two standard deviations is $(2)(0.1) = 0.2$. The sample mean $\bar{x}$ is within 0.2 units of $\mu$.

Because $\bar{x}$ is within 0.2 units of $\mu$, which is unknown, then $\mu$ is within 0.2 units of $\bar{x}$ in 95% of the samples. The population mean $\mu$ is contained in an interval whose lower number is calculated by taking the sample mean and subtracting two standard deviations ($2 \cdot 0.1$) and whose upper number is calculated by taking the sample mean and adding two standard deviations. In other words, $\mu$ is between $\bar{x} - 0.2$ and $\bar{x} + 0.2$ in 95% of all the samples.

For the CD example, suppose that a sample produced a sample mean $\bar{x} = 2$. Then the unknown population mean $\mu$ is between 

$$\bar{x} - 0.2 = 2 - 0.2 = 1.8 \text{ and } \bar{x} + 0.2 = 2 + 0.2 = 2.2.$$ 

We say that we are 95% confident that the unknown population mean number of CDs is between 1.8 and 2.2. The 95% confidence interval is $(1.8, 2.2)$.

The 95% confidence interval implies two possibilities. Either the interval $(1.8, 2.2)$ contains the true mean $\mu$ or our sample produced an $\bar{x}$ that is not within 0.2 units of the true mean $\mu$. The second possibility happens for only 5% of all the samples (100% - 95%).

Remember that a confidence interval is created for an unknown population parameter like the population mean, $\mu$. A confidence interval has the form 

$$(\text{point estimate} - \text{margin of error}, \text{point estimate} + \text{margin of error})$$

The margin of error depends on the confidence level or percentage of confidence.

3 Optional Collaborative Classroom Activity

Have your instructor record the number of meals each student in your class eats out in a week. Assume that the standard deviation is known to be 3 meals. Construct an approximate 95% confidence interval for the true average number of meals students eat out each week.

1. Calculate the sample mean.
2. $\sigma = 3$ and $n =$ the number of students surveyed.
3. Construct the interval $\left( \bar{x} - 2 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + 2 \cdot \frac{\sigma}{\sqrt{n}} \right)$

We say we are approximately 95% confident that the true average number of meals that students eat out in a week is between __________ and __________.

Glossary

**Definition 1: Confidence Interval (CI)**

An interval estimate for an unknown population parameter. This depends on:

- The desired confidence level.
- Information that is known about the distribution (for example, known standard deviation).
- The sample and its size.

**Definition 2: Inferential Statistics**

Also called statistical inference or inductive statistics. This facet of statistics deals with estimating a population parameter based on a sample statistic. For example, if 4 out of the 100 calculators sampled are defective we might infer that 4 percent of the production is defective.
**Definition 3: Parameter**  
A numerical characteristic of the population.

**Definition 4: Point Estimate**  
A single number computed from a sample and used to estimate a population parameter.