VII. Confidence Intervals: Means

Scenario. You will have sample data from a population on means and standard deviations. The observations on the sample will have been taken only once.

Data and parameters. You will be given sample data which estimate unknown population parameters.

	Parameters (unknown)	Sample estimates (known)
mean	$\boldsymbol{\mu}$	$\bar{\boldsymbol{x}}$
standard deviation	$\boldsymbol{\sigma}$	$\boldsymbol{s}$
sample size	–	$\boldsymbol{n}$

Research Objective. To place error bounds $\pm$ ERROR about the sample mean in such a way that the resulting *confidence interval*

$$\bar{\boldsymbol{x}} - \texttt{ERROR} \quad \text{to} \quad \bar{\boldsymbol{x}} + \texttt{ERROR}$$

has a high likelihood of including the true population parameter $\boldsymbol{\mu}$. The likelihood that the confidence interval

$$\bar{\boldsymbol{x}} - \texttt{ERROR} \quad \text{to} \quad \bar{\boldsymbol{x}} + \texttt{ERROR}$$

contains $\boldsymbol{\mu}$ is called the *confidence level* of the interval.

Solution Template

Step 1. Make a dictionary of the data given in the problem:

sample mean	$\bar{\boldsymbol{x}}$
sample standard deviation	$\boldsymbol{s}$
sample size	$\boldsymbol{n}$
confidence level	$\boldsymbol{C}$

$\boxed{\textit{Step 2.}}$ Find the confidence level in the last column in the table in Appendix B. Then look in the fourth column to find the confidence limits $\pm z_c$.

$\boxed{\textit{Step 3.}}$ Apply the following formula to find the value of the `ERROR` term:

$$\texttt{ERROR} = \pm z_c \frac{s}{\sqrt{n}}$$

using the dictionary from step one and the confidence limits from step 2.

$\boxed{\textit{Step 4.}}$ Write down the resulting confidence interval:

$$\bar{x} - \texttt{ERROR} \quad \text{to} \quad \bar{x} + \texttt{ERROR}$$

End of Solution Template

Interpretation. You expect that the interval

$$\bar{x} - \texttt{ERROR} \quad \text{to} \quad \bar{x} + \texttt{ERROR}$$

will contain the true value of the parameter μ. You would be surprised if the value of μ were outside of this interval. The confidence level measures the degree to which you expect the interval will contain the true parameter (or the degree to which you would be surprised if the true parameter were not within the interval).

The researcher is free to select whatever confidence level is appropriate for the problem being studied. However, it is conventional to consider only confidence levels of 90% or higher. Since you believe that the interval contains μ, the likelihood that this belief is in error is

$$100\% - \text{confidence level.}$$

Thus, for example, an 80% confidence interval results in a 20% chance of error; a chance of error this high is almost always unacceptable.

Assumptions. You must assume that the sample size is at least 30.

$\boxed{\textit{Example.}}$ A survey of 175 randomly selected apartment dwellers in Tulsa revealed that they had an average annual income of \$27,500 with a standard deviation of

\$3,200. Find a 95% confidence interval for the average annual income of apartment dwellers in Tulsa.

Solution.

Step 1. Our dictionary of values consists of

sample mean	$\bar{x}$	27,500
sample standard deviation	s	3,200
sample size	n	175
confidence level	95%	

Step 2. The confidence limits are ± 1.960 from the table in Appendix B.

Step 3. The error term is

$$\begin{aligned} \texttt{ERROR} &= \pm z_c \frac{s}{\sqrt{n}} \\ &= \pm 1.960 \frac{3200}{\sqrt{175}} \\ &= \pm 1.96 \frac{3200}{13.229} \\ &= \pm 1.96(241.89) \\ &= \pm 474.10 \end{aligned}$$

Step 4. The confidence interval is thus

$$\begin{array}{rcl} 27,500 - 474.10 & \text{to} & 27,500 + 474.10 \\ & \text{or} & \\ 27,025.90 & \text{to} & 27,974.10 \end{array} .$$

Thus, we would be surprised if the true average annual income for apartment dwellers in Tulsa were less than \$27,025.90 or were more than \$27,974.10. Note that we would not be 100% surprised, though – we would be only 95% surprised since we have found only a 95% confidence interval. ▮

Question. Suppose that, based on census data, we know that apartment dwellers in Oklahoma City have an average annual income of \$27,880. Would the data in the above experiment provide convincing evidence that apartment dwellers in Oklahoma City earn more than apartment dwellers in Tulsa? Suppose that census data shows that apartment dwellers residing in Stillwater have an average annual income of \$27,010. Do the data in the above example suggest that apartment dwellers in Stillwater earn less than those in Tulsa? (Since we have census data on both Oklahoma City and Stillwater, we are certain about the comparison of average incomes between those two cities.)

Remark on small samples. If your sample size is less than 30 the above procedure still works with only minor modification. You need to replace step two with the following:

Step 2. You will find the confidence limits by looking in the appropriate row and column in the table in Appendix C. The row and column are selected as follows:

- Select the row corresponding to

$$\text{row} = \text{sample size} - 1$$

This number is called the *degrees of freedom* for the sample.

- Select the column by finding

$$\text{column} = 100\% - \text{confidence level}$$

at the *bottom* of the table.

- The confidence limits are the plus/minus values of the entry in the row and column just selected.

Problems

1. In a sample of size 590, the sample mean was 32 and the sample standard deviation was 12. Find a 90% confidence interval for the population mean.

2. In a sample of size 100, the sample mean was 32 and the sample standard deviation was 12. Find a 90% confidence interval for the population mean.

3. In a sample of size 50, the sample mean was 68 and the sample standard deviation was 17. Find a 80% confidence interval for the population mean.

4. A researcher for an insurance company wants to estimate the average speed of male drivers on the freeways in Tulsa. The researcher randomly samples 83 male drivers on the Broken Arrow Expressway during the second week in January. The researcher finds that the average speed is 58.9 miles per hour with a standard deviation of 9.3 miles per hour.

(i) Construct a 95% confidence interval for the average speed of male drivers on the Broken Arrow Expressway.

(ii) Data from the Tulsa Police Department indicate that the average speed of all drivers on the Broken Arrow Expressway is 56.8 miles per hour. Has the researcher gathered convincing evidence that men drive faster than the prevailing speed?

(iii) The data included a car travelling at 86 miles per hour. The next fastest car was only travelling at 72 miles per hour. Should the researcher have thrown out the car going 86 miles per hour since it was going so much faster than any other vehicle and was therefore unrepresentative?

5. A researcher suspects that there may be bias against racial groups in the sentencing practices of a particular district court. In this court, the average sentence for armed robbery was 5.9 years. The researcher sampled 283 armed robbery cases adjudicated in this court involving racial minorities and found that the average sentence was 6.1 years with a standard deviation of 2.3 years.

(i) Find a 98% confidence interval for the average sentence for armed robbery given to racial minorities in this court. Has the researcher gathered convincing evidence that minorities recieve longer sentences?

(ii) What are some other variables which might influence the length of sentence? How would you advise the researchers to modify their sampling methods to control for these variables?

Video Assignment.

View the following program(s) from the series *Against All Odds*:

Program	*Title*
19	*Confidence Intervals*