Mohammad Madiah

# Chapter 8

## Interval Estimation (Confidence Intervals)

**Inferential Statistics** Inferential statistics consists of methods that use sample results to help make decisions or predictions about a population.

Two types of inferential statistics:

- Estimation (chapters 7 and 8)
- Hypothesis Testing (chapter 9)

**Estimation** is a procedure by which a numerical value or values are assigned to a population ***parameter*** (population mean $\mu$) based on the information collected from a sample ***statistic*** (sample mean $\bar{x}$).

For example: A researcher may want to estimate the mean daily expenditure for BZU students, a manager may want to estimate the average time taken by new employees to learn a job.

The value(s) assigned to a population parameter based on the value of a sample statistic is called an **estimate** of the population parameter. For example, suppose the researcher takes a sample of 200 BZU students and finds that the mean daily expenditure, $\bar{x}$ is 30 ILS. If the researcher assigns this value to the population mean, then 30 ILS is called an estimate of $\mu$ .

The sample statistic used to estimate a population parameter is called an **estimator**. Thus, the sample mean $\bar{x}$ , is an estimator of the population mean $\mu$

The sample statistic used to estimate a population parameter is called an **estimator**. Thus, the sample mean $\bar{x}$ , is an estimator of the population mean $\mu$ .

**Estimate and Estimator:** The value(s) assigned to a population parameter based on the value of a sample statistic is called an **estimate.** The sample statistic used to estimate a population parameter is called an **estimator**.

The estimation procedure involves the following steps.

1. Select a sample.
2. Collect the required information from the elements of the sample.
3. Calculate the value of the sample statistic.
4. Assign value(s) to the corresponding population parameter.

Two types of Estimation

**Point Estimation**

The value of a sample statistic that is used to estimate a population parameter is called a **point estimate**. Thus, the value computed for the sample mean, $\bar{x}$ from a sample is a point estimate of the corresponding population mean, $\mu$.

**Interval Estimation**

In the case of interval estimation, instead of assigning a single value to a population parameter, an interval is constructed around the point estimate, and then a **probabilistic statement** that this interval contains the corresponding population parameter is made.

- A confidence interval is an interval (range) of values that we can be really confident contains the true unknown population parameter.
- Each interval is constructed with a **given** probability, this given probability is called the ***confidence level*** and it is **selected** by the researcher. The interval is called a **confidence interval.**
- The confidence coefficient is denoted by $(1-\alpha)$ 100%, $\alpha$ is called the significance level, which will be discussed in detail in Chapter 9
- The three most commonly used confidence levels are 90%, **95%**, and 99%.
- A 95% confidence interval means that there is a 95% chance that the confidence interval contains the population mean.
- The general form of an interval estimation is

**Point estimate ± Margin of Error**

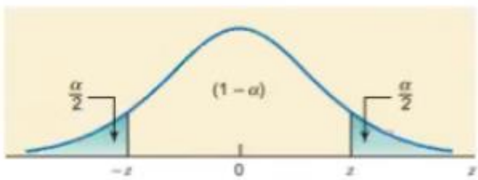
- Two cases for interval estimation
  1. $\sigma$ **known** case: The population standard deviation is known.
  2. $\sigma$ **unknown** case: The population standard deviation is unknown. The sample standard deviation $s$ will be used as a point estimate for $\sigma$

## CONFIDENCE INTERVAL F0r POPULATION MEAN σ KNOWN CASE

The $(1-\alpha)$ 100% Interval estimation for a population mean $\sigma$ known case is given by

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$z_{\frac{\alpha}{2}}$ ≡ The z-value providing an area $\frac{\alpha}{2}$ in the upper tail of the standard normal probability distribution.




For example $z_{0.1}$ is z – value providing an area 0.1 in the upper tail of the Z – table. That is, P (Z > z) = 0.1 ($z_{0.1}$ = 1.28)

$\frac{\sigma}{\sqrt{n}}$ ≡ The standard error of the mean

≡ The margin of error (maximum error of estimate).

$$E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

For example: The 95% confidence interval; $\frac{\alpha}{2} = \frac{5\%}{2} = 0.025$. So, $z_{0.025}$ is the z – value providing an area of 0.025 in the upper tail. That is, P (Z < z) = 1 - 0.025 = 0.975 ($z_{0.025}$ = 1.96)

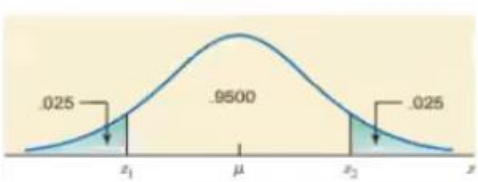



TABLE 8.1 VALUES OF $Z_{\alpha/2}$ FOR THE MOST COMMONLY USED CONFIDENCE LEVELS

| Confidence Level | α | α/2 | $z_{\alpha/2}$ |
|---|---|---|---|
| 90% | .10 | .05 | 1.645 |
| 95% | .05 | .025 | 1.960 |
| 99% | .01 | .005 | 2.576 |

**Example:**

100 randomly selected university students were asked how much money they spent on mobile over the past month. The sample mean for the 100 students was 58 ILS. Suppose that the standard deviation for the population 15 ILS. Construct and interpret a 90%, 95%, and 99% confidence intervals for the mean money spent by all university students.

**Solution**

The 90% confidence interval for this example is

$$\bar{\mathbf{x}} \pm \mathbf{z}_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{\mathbf{n}}} \Rightarrow 58 \pm 1.645 \frac{15}{\sqrt{100}} = 58 \pm 2..47$$

The interval is 55.53 to 60.47

There is a 90% chance (probability, confidence) that the interval 55.53 to 60.47 contains the mean amount of money spent by all university students.

(We are 0.90 confident that mean amount spent by all students is between 55.53 and 60.47)

The 95% confidence interval for this example is

$$\bar{\mathbf{x}} \pm \mathbf{z}_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{\mathbf{n}}} \Rightarrow 58 \pm 1.96 \frac{15}{\sqrt{100}} = 58 \pm 2.94$$

The interval is 55.06 to 60.94

There is a 95% chance that the interval 55.06 to 60.94 contains the mean amount of money spent by all university students.

The 99% confidence interval for this example is

$$58 \pm 2.575 \frac{15}{\sqrt{100}} \Rightarrow 58 \pm 3.86$$

The interval is 54.14 to 61.86

There is a 99% chance that the interval 54.14 to 61.81 contains the mean amount of money spent by all university students.

**Note that a 99% confidence interval for the population mean is the wider interval of all intervals. That is, as confidence level increases, the width of the interval increases (the margin of error increases).**

## Confidence Intervals for Unknown Mean and Unknown Standard Deviation

In most practical research, the standard deviation for the population of interest is not known. In this case, the standard deviation $\sigma$ is replaced by the estimated standard deviation *s*. The sample mean follows the ***t* distribution** with mean $\mu$ and standard deviation $\frac{s}{\sqrt{n}}$.

**The t Distribution**: The t distribution is a type of bell-shaped distribution with a lower height and a wider spread than the standard normal distribution. The *t* distribution is also described by its degrees of freedom (df). For a sample of size n, the t distribution will have n-1 degrees of freedom. As the sample size n increases, the t distribution becomes closer to the normal distribution.

For a population with unknown mean $\mu$ and unknown standard deviation, a confidence interval for the population mean, based on a simple random sample of size *n*, is

$$\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \text{ with degrees of freedom } n-1$$

, where $t_{\frac{\alpha}{2}}$ is the value for the *t* distribution with *n-1* degrees of freedom.

| | Confidence Intervals | | | | |
|---|---|---|---|---|---|
| | 80% | 90% | 95% | 98% | 99% |
| | Level of Significance for One-Tailed Test | | | | |
| *df* | 0.10 | 0.05 | 0.025 | 0.010 | 0.005 |
| | Level of Significance for Two-Tailed Test | | | | |
| | 0.20 | 0.10 | 0.05 | 0.02 | 0.01 |
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 |

## Example

An economist is interested in studying the incomes of households in Ramallah. A random sample of 25 individuals resulted in an average income of $1500 and standard deviation of $100. Construct the 95% and the 99% confidence interval for the income of all households in Ramallah?

- The 95% interval $t_{0.025} = 2.064$ with degrees of freedom df = 25 −1= 24

$$\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \Rightarrow 1500 \pm 2.064 \frac{100}{\sqrt{25}} \Rightarrow 1500 \pm 41.28$$

Interval, 1458.72 to 1541.28: There is a 95% chance that the mean of income of all households in Ramallah is between $1458.72 and $1541.28

- The 99% interval $t_{0.005} = 2.797$ with degrees of freedom df = 25 −1= 24.

$$\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \Rightarrow 1500 \pm 2.797 \frac{100}{\sqrt{25}} \Rightarrow 1500 \pm 55.94$$ Interval; 1444.06 to 1555.94 There is a 99% chance that the mean of income of all households in Ramallah is between $1444.06 and $1555.94.

## Example

A simple random sample of size 36 provided a sample mean of 416 and a sample standard deviation of 180.

a. Compute the standard error of the sample mean.

b. Compute and interpret the 99 percent confidence interval for the mean.

**c.** How large a sample is needed to estimate the population mean with a maximum error of 10?

## Determining the Sample Size for the Estimation of Population Mean μ

$$E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow n = \frac{z^2 \sigma^2}{E^2} = \left(\frac{z\sigma}{E}\right)^2$$

The confidence level is known (the researcher started with a known confidence level), the margin of error is also selected by the researcher, the unknown value of **σ** can be determined by one of the three following methods (planning value for **σ** ):

*A planning value for the population standard deviation σ must be specified before the sample size can be determined. Three methods of obtaining a planning value for σ are discussed here.*

1. Use the estimate of the population standard deviation computed from data of previous studies as the planning value for $\sigma$.
2. Use a pilot study to select a preliminary sample. The sample standard deviation from the preliminary sample can be used as the planning value for $\sigma$.
3. Use judgment or a "best guess" for the value of $\sigma$. For example, we might begin by estimating the largest and smallest data values in the population. The difference between the largest and smallest values provides an estimate of the range for the data. Finally, the range divided by 4 is often suggested as a rough approximation of the standard deviation and thus an acceptable planning value for $\sigma$.

### Example

Suppose the Palestinian Central Bureau of statistics wants to estimate the mean family size of all families in Palestine at a 99% confidence level. It is known that the standard deviation for the sizes of all families in Palestine is 0.6. How large a sample should the bureau select if it wants its estimate to be within 0.01 of the population mean?

**Solution**

$$n = \frac{z^2 \sigma^2}{E^2} = \frac{(2.575)^2 (0.6)^2}{(0.01)^2}$$
$$= 23870.25$$
$$= 23871$$

The required sample size is 23871.

[You should around your answer to the next integer.]

24. The range for a set of data is estimated to be 36.
a. What is the planning value for the population standard deviation?
b. At 95% confidence, how large a sample would provide a margin of error of 3?
c. At 95% confidence, how large a sample would provide a margin of error of 2?

**Solution:**

a. planning value for the population standard deviation is

$$\sigma = \frac{\text{Range}}{4} = \frac{36}{4} = 9$$

b. $n = \left(\frac{z\sigma}{E}\right)^2 = \left(\frac{1.96*9}{3}\right)^2 = 34.5744 = 35$

c. $n = \left(\frac{z\sigma}{E}\right)^2 = \left(\frac{1.96*9}{2}\right)^2 = 77.7924 = 78$

27. Annual starting salaries for college graduates with degrees in business administration are generally expected to be between $30,000 and $45,000. Assume that a 95% confidence interval estimate of the population mean annual starting salary is desired. What is the planning value for the population standard deviation? How large a sample should be taken if the desired margin of error is
a. $500?
b. $200?
c. $100?
d. Would you recommend trying to obtain the $100 margin of error? Explain.

$\sigma = \frac{15000}{4} = 3750$ $z_{\alpha/2} = 1.960$

E ↓ → n ↑ → Confd. ↑

a. $\left(\frac{3750 \times 1.96}{500}\right)^2$

b