Estimation

- Estimation
- Review of Central Limit Theorum
- Confidence Levels
- Confidence Intervals
- Confidence Interval Precision
  - Standard Error of the Mean
  - Sample Size
  - Standard Deviation

Estimation Defined:

Estimation – A process whereby we select a random sample from a population and use a sample statistic to estimate a population parameter.

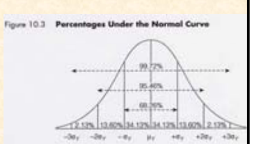
Point and Interval Estimation

- **Point Estimate** – A sample statistic used to estimate the value of a population parameter (for example, a specific mean obtained from sample and used to estimate the population mean)
- **Confidence interval (*interval estimate*)** – A range of values defined by the confidence level within which the population parameter is estimated to fall.
- **Confidence Level** – The likelihood, expressed as a percentage or a probability, that a specified interval will contain the population parameter.

Applying Properties of the Sampling Distribution

If the sample mean is $50K, N=100, and the standard deviation for the sample is $7K, we can calculate the standard error and then use it:

- We know that 95% of the sample means in a sampling distribution would fall between two standard errors of the mean:

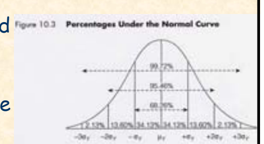


48.6K 50K 51.4K

2 Standard Errors = $1,400

Applying Properties of the Sampling Distribution

- Therefore we can say that there is a 95% chance that the true population mean falls between the two standard errors of the mean.
- Or, in other words, we are 95% confident that the population mean falls between two SEs of the sample mean.



48.6K 50K 51.4K

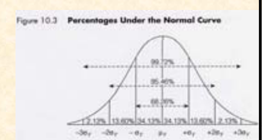
2 Standard Errors = $1,400

Applying Properties of the Sampling Distribution

Applying these assumptions to the data has the following results:

We are 95% confident that the average income for the **population** ranges between:

$48,600 and $51,400



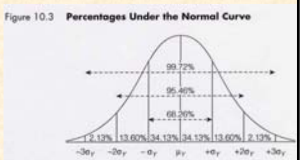
48.6K 50K 51.4K

2 Standard Errors = $1,400

Example of Applying the Properties of the Sampling Distribution

We can be 99% confident that the average income for the population ranges between:

$47,900 and $52,100.



47.9K 50K 52.1K

3 Standard Errors = $2,100

Applying Properties of the Sampling Distribution

Notice that in the example we:

1. assumed the **sampling distribution** has a normal distribution and
2. Applied the characteristics of the normal curve (i.e., 2 SEs will include 95% of the cases) to our **theoretical sampling distribution** in order to estimate the population mean.

How can we make these assumptions?

The Central Limit Theorem

The CLT tells us that, even if a **population** distribution is skewed or flat (like multiple roles of a die), we know that the **sampling distribution drawn from that population is normally distributed** (such as plotting the sample means from repeatedly rolling a die 10 times and calculating the averages).

The CLT also tells us that a single **sample of sufficient size (at least 50 cases) will often mirror the sampling distribution**.

The Central Limit Theorem

Thus, the CLT clarifies for us:

**researchers can use a single sample to construct confidence intervals around sample estimates (such as the sample mean) with specific confidence levels associated with the confidence intervals.

Confidence Levels:

Confidence Level – The likelihood, that a specified interval will contain the population parameter.

- **95% confidence level** – there is a .95 probability that a specified interval DOES contain the population mean. In other words, there are 5 chances out of 100 (or 1 chance out of 20) that the interval *DOES NOT* contains the population mean.
- **99% confidence level** – there is 1 chance out of 100 that the interval *DOES NOT* contain the population mean.

Constructing a Confidence Interval (CI)

- The **sample statistic** is the point estimate of the population parameter (such as the mean).
- The sample **standard deviation** is the point estimate of the population standard deviation.
- The **standard error** (or the standard deviation of the sampling distribution) makes it possible to state the probability that an interval around the point estimate contains the actual population parameter.

The Standard Error

is the standard deviation for a sampling distribution

Estimating standard errors

Below is the formula for estimating the standard error of the **sampling** distribution (this is also referred to as the standard deviation of the **sampling** distribution).

Standard Error of Sampling Distribution = $\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{N}}$ (standard deviation of population)

Estimating standard errors

Since the standard error of the **sampling** distribution is generally not known because we do not have the SD of the population, we usually work with the estimated standard error:

Estimated **Standard Error** Based On Large, Random Sample = $s_{\bar{Y}} = \frac{s_Y}{\sqrt{N}}$ (standard deviation of sample)

Determining a Confidence Interval (CI)

$$CI = \bar{Y} \pm Z \ (s_{\bar{Y}})$$

where:

$\bar{Y}$ = sample mean (estimate of u_y)

Z = Z score

$s_{\bar{Y}}$ = estimated standard error

Determining a Confidence Interval (CI)

$$CI = \bar{Y} \pm Z \ (s_{\bar{Y}})$$

In our previous examples we used this formula by:

(1) calculating the SE (SD/$\sqrt{N}$),
(2) multipling it by Z (two or three) and then
(3) Subtracting or adding this sum to the mean to obtain the confidence intervals.

Estimation Defined:

Estimation – A process whereby we select a random sample from a population and use a sample statistic to estimate a population parameter.

Confidence Interval and Risk Taking

$$\bar{Y} \pm Z\left(\frac{S_Y}{\sqrt{N}}\right)$$

- **Confidence Level** – Increasing our confidence level from 95% to 99% means we are less willing to draw the wrong conclusion.

  – we take a 1% risk (rather than a 5%) that the specified interval does not contain the true population mean.

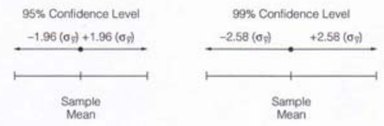
Confidence Interval and Risk Taking

$$\bar{Y} \pm Z\left(\frac{S_Y}{\sqrt{N}}\right)$$

If we reduce our risk of being wrong, then we need a wider range of values . . . ***So the interval becomes less precise.***

Confidence Interval Width

Figure 12.1 **Relationship Between Confidence Level and Z for 95 and 99 Percent Confidence Intervals**



Confidence Interval Z Values

Table 12.1 **Confidence Levels and Corresponding Z Values**

Confidence Level	Z Value
90%	1.65
95%	1.96
99%	2.58

Confidence Interval Risk

$$\bar{Y} \pm Z\left(\frac{S_Y}{\sqrt{N}}\right)$$

Sample Size – Larger samples result in smaller standard errors, and therefore, in sampling distributions that are more clustered around the population mean.

A more closely clustered sampling distribution indicates that our confidence intervals will be narrower and more precise.

Confidence Interval Risk

$$\bar{Y} \pm Z\left(\frac{S_Y}{\sqrt{N}}\right)$$

Standard Error – Smaller sample standard deviations result in smaller, more precise confidence intervals.

(Unlike sample size and confidence level, the researcher plays no role in determining the standard deviation of a sample.)

Example: Sample Size and Confidence Intervals

(remember, we are assuming a random sample)

Table 12.2 **95 Percent Confidence Interval and Width for Mean Income for Three Different Sample Sizes**

Sample Size	Confidence Interval	Interval Width	S_y	$S_{\bar{y}}$
N = 472	$27,259-$31,421	$4,162	$23,067	1061.53
N = 945	$27,869–$30,811	$2,942	$23,067	750.39
N = 1,890	$28,300–$30,380	$2,080	$23,067	530.64

Example: Determining CIs for Hispanic Migration and Earnings

From 1980 Census data:

- **Cubans** had an average income of $16,368 ($S_y$ = $3,069), N=3895
- **Mexicans** had an average income of $13,342 ($S_y$ = $9,414), N=5726
- **Puerto Ricans** had an average income of $12,587 ($S_y$ = $8,647), N=5908

Example: Determining 95% CIs for Average Cuban Earnings

Cubans

1. Standard error = $3069/\sqrt{3895}$
   = (3069/62.4) = 49.17
2. 1.96 * 49.17 = $96.37 (2 standard errors)
3. $16,368 – $96.37 to $16,368 + $96.37
   (mean-SE to mean+SE)
4. = 16,272 to 16,464 (95% confident the mean income of Cubans is within this range)

Example: Determining 95% CIs for Average Mexican Earnings

$\sqrt{N}$

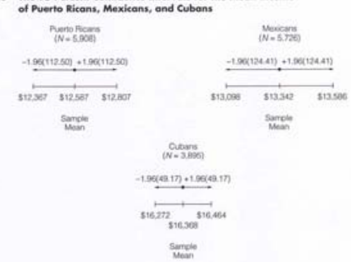
Mexicans

1. Standard error = $9414/\sqrt{5726}$
2. = (9414/75.7) = 124.41
3. 1.96 * 124.41 = 243.8
4. $13,342 – 243.8 to $13,342 + 243.8
   (mean-SE to mean+SE)

   = 13,098 to 13,586

Example: Hispanic Migration and Earnings

Figure 12.6 **The 95 Percent Confidence Intervals for the Mean Income of Puerto Ricans, Mexicans, and Cubans**



In Sum:

Now you can estimate a population's parameter (such as the mean) based on a sample.

Simply calculate the standard error, determine the level of confidence that you want, and then calculate the confidence intervals using the given formula.

Procedures for Estimating Proportions or Percentages within a Population

The procedures for determining the confidence intervals surrounding a percentage are slightly different than determining the CIs surrounding a mean.

For Example:

Based on a random sample of 1,512 adults, the percentage of our sample opposed to gay marriage was 60.

What would we estimate the CIs to be for the population if we want to be 95% confident?

Example: a random sample of 1,512 adults, the percentage opposed to gay marriage was 60.

The formula is:
$CI = p \pm Z(Sp)$

Where:
- CI = the confidence interval
- p = the observed sample percentage
- Z = the level of confidence desired
- Sp = the estimated standard error

Where the formula for the standard error of the percentage is:

Sp = the estimated standard error
or:

$$Sp = \sqrt{\frac{(p)\ (1-p)}{N}}$$

Where:
- P = sample percentage
- N = number of cases

Applying the Example: a random sample of 1,512 adults, the percentage opposed to gay marriage was 60. What are the CIs with 95% confidence?

$$Sp = \frac{(.60)\ (1 - .60)}{1{,}512} = \sqrt{.000158} = .012$$

The formula is:
$CI = p \pm Z(Sp)$
or
$CI = .60 \pm 1.96\ (.01)$
or
.58 and .62

That is, we are 95% confident that the percentage of the population opposed to gay marriage falls somewhere between 58% and 62%.