- **Estimation Defined**
- **Confidence Levels**
- **Confidence Intervals**
- **Confidence Interval Precision**
  - **Standard Error of the Mean**
  - **Sample Size**
  - **Standard Deviation**
- **Confidence Intervals for Proportions**

Estimation Defined:

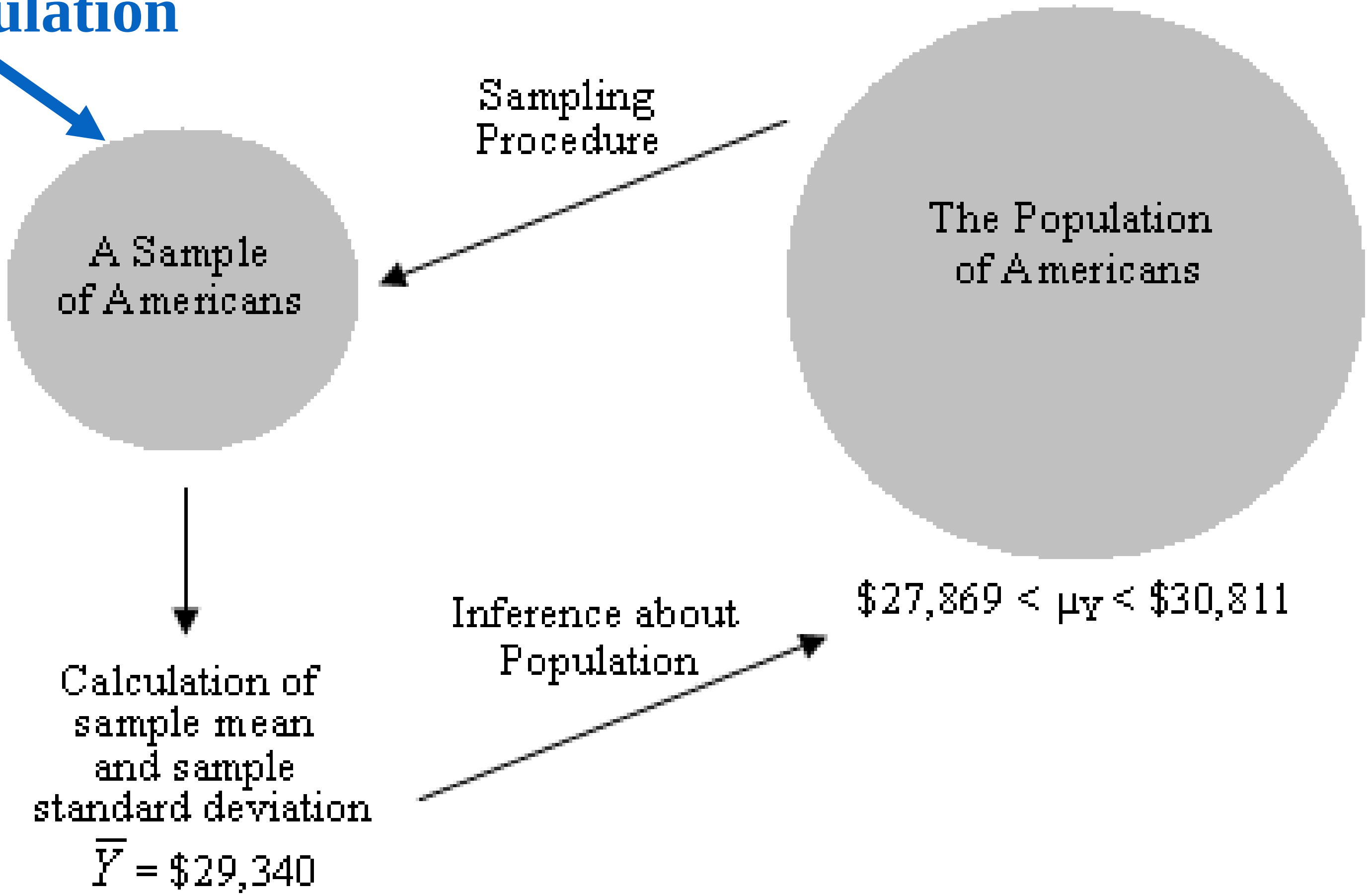
- **Estimation** – A process whereby we select a **random sample** from a population and use a **sample statistic** to **<u>estimate</u>** a **population parameter**.

Point and Interval Estimation

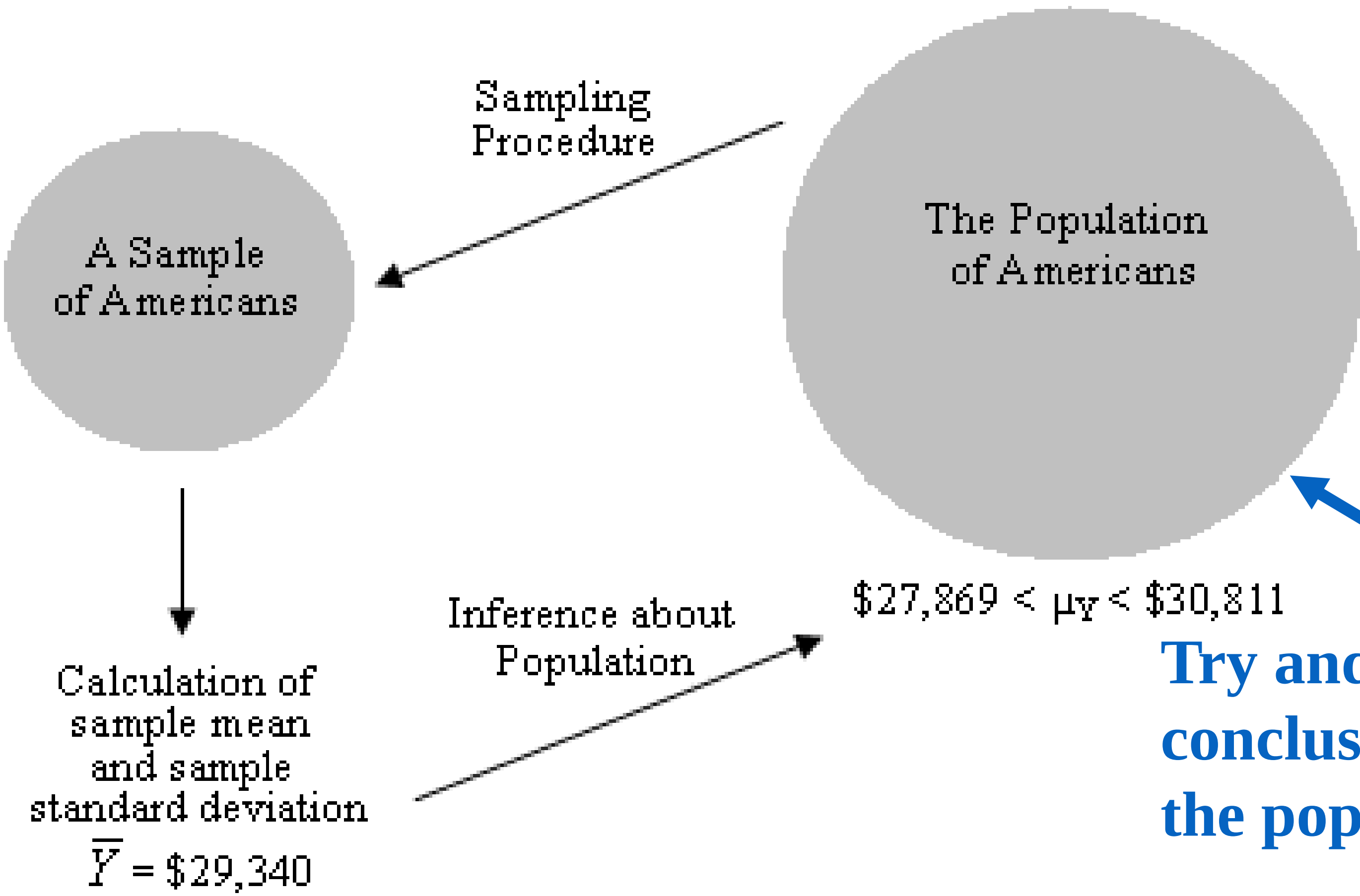
- **Point Estimate** – A sample statistic used to estimate the **exact value** of a population parameter
- **Confidence interval (*interval estimate*)** – **A range of values** defined by the confidence level within which the population parameter is **estimated** to fall.
- **Confidence Level** – The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

Estimations Lead to Inferences

Take a subset of the population



Estimations Lead to Inferences



Try and reach conclusions about the population

Inferential Statistics involves Three Distributions:

A population distribution – variation in the **larger group** that we want to know about.

A distribution of sample observations – variation in the **sample** that we **can observe**.

A sampling distribution – a normal distribution whose mean and standard deviation are **unbiased estimates** of the **parameters** and allows one to infer the parameters from the statistics.

The Central Limit Theorem Revisited

- What does this Theorem tell us:
  - Even if a population distribution is skewed, we know that **the sampling distribution** of the mean is **normally distributed**
  - As the sample size gets larger, the mean of the **sampling distribution** becomes <u>**equal**</u> to the **population mean**
  - As the **sample size gets <u>larger</u>**, the **standard error of the mean <u>decreases</u>** in size (which means that the **variability** in the sample estimates from sample to sample **decreases** as N increases).
- It is important to remember that researchers <u>**do not**</u> typically conduct repeated samples of the same population. Instead, they use the knowledge of **theoretical sampling distributions** to construct confidence intervals around estimates.

Confidence Levels:

- **Confidence Level** – The likelihood, expressed as a percentage or a probability, 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*** contain the population mean.
  - **99% confidence level** – there is 1 chance out of 100 that the interval ***DOES NOT*** contain the population mean.

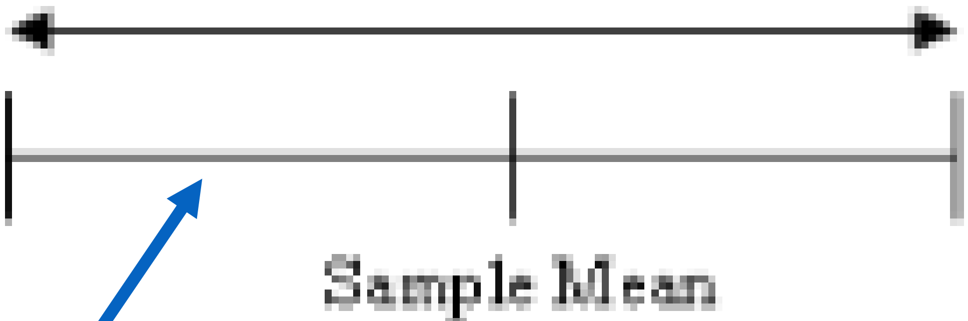
Constructing a Confidence Interval (CI)

- The **sample mean** is the point **estimate** of the **population mean**.
- The sample **standard deviation** is the point **estimate** of the **population standard deviation**.
- The **standard error** of the mean makes it possible to state the **probability** that an interval around the point estimate **contains** the actual **population mean**.

What We are Wanting to Do

We want to construct an estimate of where the population mean falls based on our sample statistics

The actual population parameter falls somewhere on this line



This is our Confidence Interval

The Standard Error

Standard error of the mean – the standard deviation of a **sampling distribution**

$$\text{Standard Error} = \sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{N}}$$

Estimating standard errors

$$\sigma_{\overline{y}} = \frac{\sigma_y}{\sqrt{N}}$$

Since the standard error is generally **not known**, we usually work with the **estimated standard error:**

$$s_{\overline{Y}} = \frac{s_Y}{\sqrt{N}}$$

Determining a Confidence Interval (CI)

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

where:

$\overline{Y}$ = sample mean (estimate of μ)

Z = Z score for one-half the acceptable error

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

Confidence Interval Width

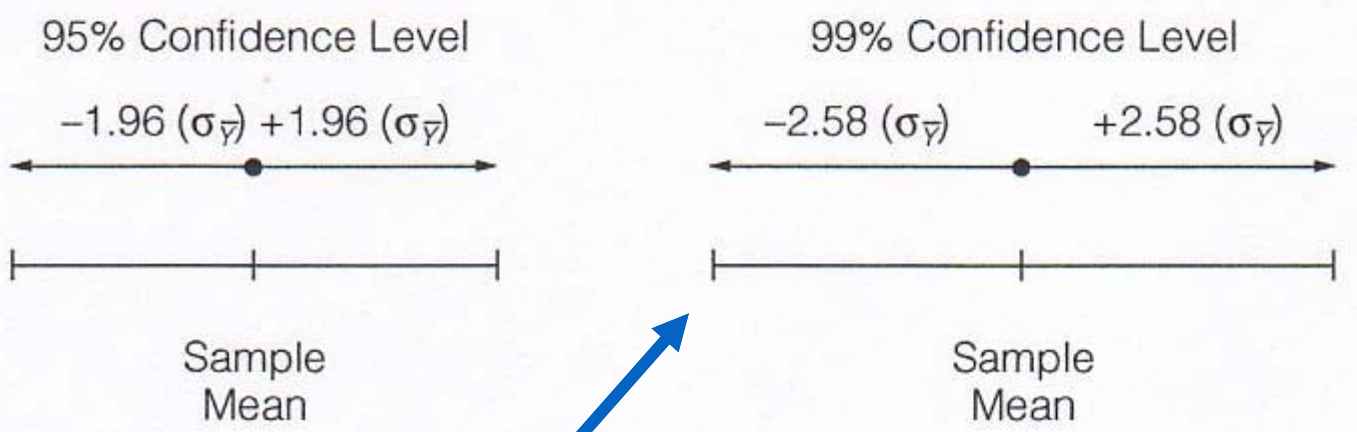
$$\bar{Y} \pm Z\left(\frac{s_Y}{\sqrt{N}}\right)$$

- **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.

  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**



More precise, less confident

More confident, less precise

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 Width

$$\overline{Y} \pm Z \left(\frac{s_Y}{\sqrt{N}}\right)$$

- **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 Width

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

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

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

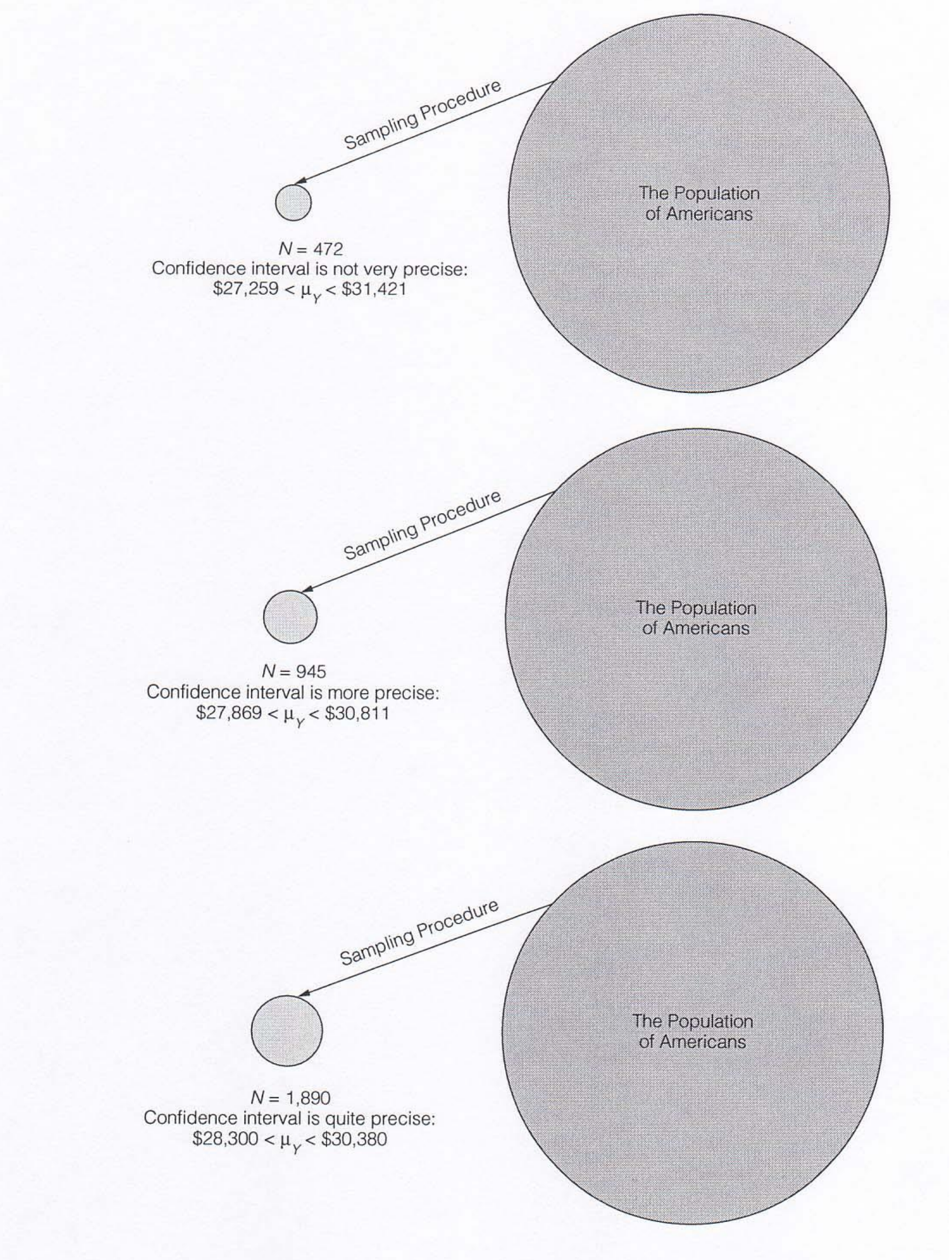
Example: Sample Size and Confidence Intervals

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: Sample Size and Confidence Intervals

Figure 12.5 **The Relationship Between Sample Size and Confidence Interval Width**



Example: 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 of \$13,342 (S_y = \$9,414), N=5726
- Puerto Ricans had an average of \$12,587 (S_y = \$8,647), N=5908

Example: Hispanic Migration and Earnings

Now, compute the 95% CI's for all three groups:

- Cubans: standard error = 3069/ $\sqrt{3895}$ **= 49.17**

$$95\%CI = 16{,}368 \pm 1.96(49.17)$$

= 16,272 to 16,464

- Mexicans: s.e. = 9414/$\sqrt{5726}$ **= 124.41**

$$95\%CI = 13{,}342 \pm 1.96(124.41)$$

= 13,098 to 13,586

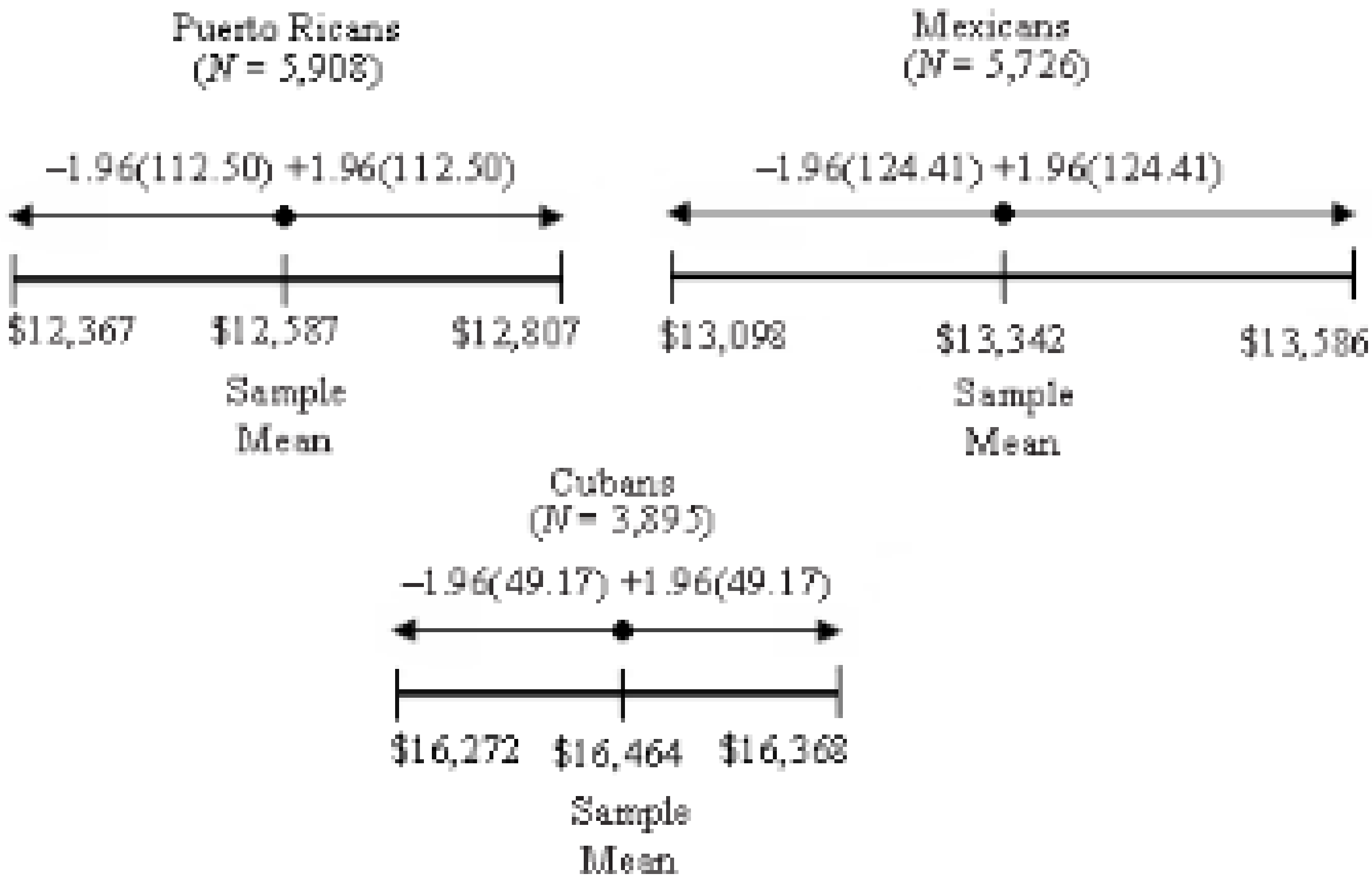
Example: Hispanic Migration and Earnings

- Puerto Ricans, s.e. = 8647/ $\sqrt{5908}$ **= 112.5**

$$95\%CI = 12{,}587 \pm 1.96(112.5)$$

= 12,367 to 12,807

Example: Hispanic Migration and Earnings



Confidence Intervals for Proportions

- **Estimating the standard error of a proportion** – based on the Central Limit Theorem, a sampling distribution of proportions is **approximately normal**, with a mean, μ_p , equal to the population proportion, π, and with a standard error of proportions equal to:

$$\sigma_p = \sqrt{\frac{(\pi)(1-\pi)}{N}}$$

Since the standard error of proportions is generally **<u>not</u> known**, we usually work with the **estimated standard error:**

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

Determining a Confidence Interval for a Proportion

$$p \pm Z(s_p)$$

where:

p = observed sample proportion (estimate of π)

Z = Z score for one-half the acceptable error

s_p = estimated standard error of the proportion

Confidence Intervals for Proportions

Protestants in favor of banning stem cell research: N = 2,188, p = .37

Calculate the estimated standard error: $S_p = \sqrt{\dfrac{(.37)(1-.37)}{2{,}188}} =$ **.10**

Determine the confidence level

Lets say we want to be **95%** confident

$= .37 + 1.96(.010)$

$= .37 \pm .020$

= **.35 to .39**

Confidence Intervals for Proportions

Catholics in favor of banning stem cell research:
N = 880, p = .32

Calculate the estimated standard error: $S_p = \sqrt{\frac{(.32)(1-.32)}{880}} =$ **.16**

Determine the confidence level

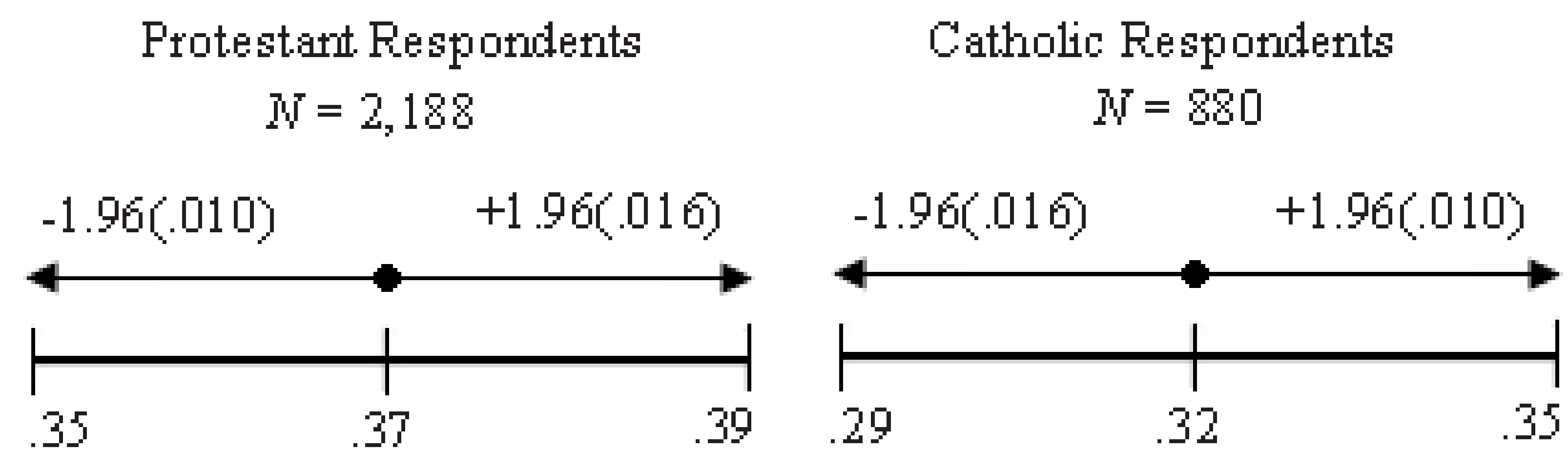
Lets say we want to be **95%** confident

= .32 + 1.96(.016)

= .32 ± .031

= **.29 to .35**

Confidence Intervals for Proportions



Interpretation:We are **95 percent confident** that the **true population proportion** supporting a ban on stem-cell research is somewhere **between .35 and .39** (or between 35.0% and 39.0%) for **Protestants**, and somewhere between .29 and .35 (or between 29.0% and 35.0%) for **Catholics.**