Samples and Populations

Statistics for Political Science
Levin and Fox
Chapter 6
Populations and Samples

**Population:** Set of individuals who share at least one characteristic.

**Sample:** Smaller number of individuals from the population.
Random Sampling

Random Sample:
Researcher strive to give every member of the population an equal and independent chance of being drawn into the sample.

This indicates that each member of the population must be identified before the random sample is drawn. This is done by obtaining a reliable list of the entire population.
Random Sampling

The most basic type of random sample comes from a **table of random numbers** or a **random number generator**.

A random number table or generator has **no particular pattern or order**.
Random Sample

Creating a Random Sample:

1. Obtain a list of the population.

2. Assign a unique identifying number to each member (telephone number?).

3. Draw the members of the sample from the table of random numbers.

4. Disregard numbers that come up a second time or are higher than needed.
Sampling Error

Now we have to carefully distinguish between samples and populations.

This requires that we use different symbols for each:

**Mean:**
\[ \bar{X} = \text{mean of a sample} \]
\[ \mu = (\text{mu}) \text{ mean of a population/probability} \]

**Standard Deviation:**
\[ s = \text{standard deviation of a sample} \]
\[ \sigma = (\text{sigma}) \text{ standard deviation of a population/probability} \]

**Note:** The symbols for population and probability distributions are the same.
Sampling Error

Sampling Error:
With random sampling, we can always expect some difference between a sample (random or nonrandom) and the population from which it was drawn.

This is known as sampling error. Sampling error is in every study no matter how rigorous the design or careful the researcher. As such, the sample mean (\( \bar{X} \)) will never be exactly the same as a population mean (\( \mu \)).
## Sampling Error

### Sampling Error: Table 6.1

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample A</th>
<th>Sample B</th>
<th>Sample C</th>
</tr>
</thead>
<tbody>
<tr>
<td>70 80 93</td>
<td>96</td>
<td>40</td>
<td>72</td>
</tr>
<tr>
<td>86 85 90</td>
<td>99</td>
<td>86</td>
<td>96</td>
</tr>
<tr>
<td>56 52 67</td>
<td>56</td>
<td>56</td>
<td>49</td>
</tr>
<tr>
<td>40 78 57</td>
<td>52</td>
<td>67</td>
<td>56</td>
</tr>
<tr>
<td>89 49 48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99 72 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96 94</td>
<td>$\bar{X} = 75.55$</td>
<td>$\bar{X} = 62.25$</td>
<td>$\bar{X} = 68.25$</td>
</tr>
<tr>
<td>$\mu = 71.55$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sampling Distribution of Means

Sampling Distribution of Means:
Even with sampling error, it is possible to generalize from a sample to a larger population.

What we can do is construct a sampling distribution of means. We can collect data on multiple samples and calculate their means. These means are then placed in a frequency distribution.
Sampling Distribution of Means

Sampling Distribution of Means: Example: Long-Distance Phone Calls

A researcher studying the extent of long-distance phone calls made in the US, collects a single sample of 200 households taken at random from the entire population. The researcher observers each household to determine how many minutes are spent over a one-week period on long-distance calling.

Findings:

\[ \bar{X} = 0 - 240 = 240 \text{ minutes} \]

\[ = 101.55 \text{ minutes (sample Mean)} \]

Lets pretend as though we also know the population mean (\( \mu \)):

\[ \mu = 99.75 \text{ minutes} \]
# Sampling Error

## Sampling Error: Table/Figure 6.2: Long-distance Phone Calls (Minutes)

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample …</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Long-distance calling in all US Households</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>μ = 99.75</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>200 Households</strong></td>
<td><strong>$\bar{X} = 101.55$</strong></td>
<td><strong>$\bar{X} = ?$</strong></td>
<td><strong>$\bar{X} = ?$</strong></td>
</tr>
<tr>
<td><strong>Sample 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sampling Distribution of Means, continued

Example: Long-Distance Phone Calls
Here is another way to illustrate the relationship between the population mean ($\mu$) and the sample mean ($\bar{X}$).

$$\mu = 99.75$$

$$\bar{X} = 101.5$$
Sampling Distribution of Means

Example: Long-Distance Phone Calls
What if the researchers took **100 samples of 200 households**?

Just like raw scores, we can use **sample means** to construct a frequency distribution, known as a **sampling distribution of means**.
### Sampling Distribution of Means

**Example: Long-Distance Phone Calls**

<table>
<thead>
<tr>
<th>Mean</th>
<th>$f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>109</td>
<td>1</td>
</tr>
<tr>
<td>108</td>
<td>1</td>
</tr>
<tr>
<td>107</td>
<td>2</td>
</tr>
<tr>
<td>106</td>
<td>5</td>
</tr>
<tr>
<td>105</td>
<td>4</td>
</tr>
<tr>
<td>104</td>
<td>5</td>
</tr>
<tr>
<td>103</td>
<td>7</td>
</tr>
<tr>
<td>102</td>
<td>9</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

$\mu = 100.4 \quad N = 100$

**Mean of means**
Sampling Distribution of Means

Instead of constructing a frequency distribution of raw scores, our frequency distribution is of sample means.

It is worth noting that the sampling distribution of means is a theoretical model.
Characteristics of a Sampling Distribution of Means

Consider these statements to be truths that hold for all sample distribution of means:

1. The **sampling distribution of means approximates the normal curve.** This is true regardless of the shape of the distribution of raw scores in the population from which the means are drawn, so long as the sample of means is large enough (over 30).

2. The **mean of a sample distribution of means (the mean of means)** is equal to the **true population mean.**

3. The **standard deviation** of a sample distribution of means is **smaller** than the standard deviation of the population.
Sampling Distribution of Means

Figure 6.3: Calls: Sampling distribution of means approximates the normal curve.

Population

One Sample

100 Samples

Infinite Samples

FIGURE 6.3 Population, Sample, and Sampling Distributions
Let’s take a look at Figure 6.3.

Compare **Figure A** which is the population distribution to **Figure D** which is the theoretical sampling distribution.

**Figure A** includes all of the raw scores and the range is **240**.

**Figure D** includes the means of the sample distribution and clusters around the mean in a normal distribution.

- Note how the extreme means were weeded out.
Sampling Distribution of Means as a Normal Curve (P)

Normal Curve: Probability Distribution

Remember that the normal curve can be regarded as a probability distribution.

As a result, we can find the probability of obtaining various raw scores.

Now that we are interested in sample means, we can use the normal curve to make probability statements about these sample means.
Pointers for distinguishing between:

(1) the **standard deviation** \((\sigma)\) of **raw scores** in the **population** and

(2) the **standard deviation of the sampling distribution** of **sample means**

**Notation:**

\[
\sigma_{\bar{X}} = \text{standard deviation of the sample distribution of means.}
\]

The \(\sigma\) indicates that this is an unobserved **probability** distribution.

The \(\sigma_{\bar{X}}\) indicates that this is the **standard deviation** among all possible **sample means**.
Sampling Distribution of Means as a Probability Distribution

- Mean of means
- Stand. dev of sample means

![Diagram showing the distribution of means with percentages and standard deviations]
Sampling Distribution of Means as a Normal Curve (P)

Because the sampling distribution takes the form of the normal curve, we are able to use z scores and Table A to get the probability of obtaining any sample mean.

The process is identical for a distribution of raw scores and a distribution of means. Only the names and symbols change.

So let’s calculate some z-scores!
To locate a sample mean in the sampling distribution in terms of the number of standard deviations it falls from the center, obtain the z score: subtract the sample mean from the mean of means, divided by the standard deviation of the sampling distribution of means.

\[ z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \]

Where

\[ \bar{X} = \text{sample mean in the distribution} \]
\[ \mu = \text{mean of means} \]
\[ \sigma_{\bar{X}} = \text{standard deviation of the sampling distribution of means} \]
Sampling Distribution of Means as a Normal Curve (P)

Example: Annual Income of Graduates

A university claims that the average (\( \mu \)) \textbf{(mean for a population)} income of its graduates is $25,000. Researchers investigating this claim test it on a random sample of 100 graduates. Their sample mean (\( \bar{x} \)) is only $23,500. What is the probability of getting $23,500 or less if the true population mean is $25,000?

So let’s calculate the \textit{z-score}!
Standard Deviation: Example: Annual Income of Graduates

\[ \mu = 25,000 \]

\[ X = 23,500 \]

P of 23k or Less?

mean for a population

\[ \mu = 25,000 \]

\[ X = 23,500 \]
Again, we cannot calculate our z score without knowing the **mean** (the mean of the means) and the **standard deviation**.

So, let’s say the standard deviation of the sampling distribution is $\$ 700$:

- $\bar{X} = \$23,500$ (sample mean in the distribution)
- $\mu = \$25,000$ (mean of the means (the true mean of the population))
- $\sigma_{\bar{X}} = \$700$ (standard deviation of the sampling distribution of sample means)

\[
Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}
\]

\[
Z = \frac{23,500 - 25,000}{700} = -2.142
\]
Standard Deviation: Example: Annual Income of Graduates

If 25k was the real mean, you would have a P of 2 in 100 of mean 23K in a sample.

\[
\mu = 25,000
\]

\[
P \text{ of 23k or Less} = 0.02 \ (2 \text{ in 100})
\]

\[
X = 23,500
\]

\[
\mu = 25,000
\]

mean for a population
Standard Error of the Mean

In actual practice, the researcher rarely collects data on more than one or two samples.

As a result, the researcher does not know the mean of means or the standard deviation of the sampling distribution. They have not collected several samples from which a series of means could be calculated.

**Standard error of the mean**

However, the standard deviation in a theoretical (the distribution that would exist in theory if the means of all possible samples were obtained) sampling distribution can be derived.

This derivation is known as the standard error of the mean.
The **standard error of the mean** is obtained by dividing the **population standard deviation** ($\sigma$) by the square root of the **sample size**.

$$\sigma_{X} = \frac{\sigma}{\sqrt{n}}$$

**Where**

- $\sigma = \text{population standard deviation}$
- $\sqrt{n} = \text{square root of sample size}$
Standard Error of the Mean

Standard Error of the Mean: Example: IQ
The IQ test is standardized to have a population mean ($\mu$) of 100 and a population standard deviation ($\sigma$) of 15. If we took a sample of 10 students, we could calculate the Standard Error of the Mean as follows:

\[
\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}
\]

Whereas the pop. IQ has a SD of 15, the sampling distribution of the sample mean for $N = 10$ has a SD of 4.74

\[
\sigma_{\bar{X}} = \frac{15}{\sqrt{10}} = \frac{15}{3.1623} = 4.74
\]
Standard Error of the Mean

Confidence Intervals:
With the aid of the standard error of the mean, we can find the range of mean values with which our true population mean is likely to fall.

We can estimate the probability that our population mean actually falls within that range of mean values. This is done with confidence intervals.
Confidence Intervals

Confidence Intervals and the Normal Curve:
Sampling error is the inevitable product of only taking a portion of the population. As a result, we are working with an estimate of the mean.

BUT, there are things that we can know from our use of the normal curve.

We can pull together everything we know about the nature of the normal curve and the standard error of the means to determine the margin of error and level of confidence.
Calculating and Using Confidence Intervals

1. Select a random sample from our population
2. Calculate the mean for our sample \( \bar{X} \)
3. Now, we take a look at our normal curve. We can know the percent of all random samples that fall within specific standard deviations from the true population mean.
4. Using the information that we have collected thus far, we can calculate the standard error of the mean (also known as the standard deviation from the mean).
5. Using this information, we can obtain the 68% confidence interval:

\[
\bar{X} \pm q\]

68% confidence interval (CI)
The Normal Curve: Theoretical Overview

- **Percentages**
  - 34.13%
  - 47.72%
  - 49.87%

- **X-bar/µ**
- **+1σ/σx**
- **+2σ/σx**
- **+3σ/σx**

- **Measures of mean and SD**
- **Sample or pop/prob. mean**
- **Sample or pop/prob. SD/SE**
Standard Deviation: Example: Annual Income of Graduates

One standard error of mean below

One standard error of mean above

68%
Confidence Intervals

Confidence Intervals: Example: Private School IQ

Suppose the Dean of a private school wants to estimate the mean IQ ($\mu$) of her study body, but does not want to test every student. So, she takes a sample of 25. She finds that the sample mean ($\bar{X}$) is 105. Remember the population standard deviation ($\sigma$) for IQ is 15. How confident can the Dean be that the sample mean reflects the true pop. mean?

\[
\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}
\]

\[
\sigma_{\bar{X}} = \frac{15}{\sqrt{25}} = \frac{15}{5} = 3.0
\]

\[N = 25 \quad \mu = 100 \quad \sigma = 15 \quad \bar{X} = 105\]
Confidence Intervals

Confidence Intervals: Example: Private School IQ
How confident can the Dean be that the sample mean reflects the true pop. mean?

68% confidence interval (CI) = \( \bar{X} \pm \sigma \)

68% confidence interval (CI) = 105 +/- 3

68% confidence interval (CI) = 102 - 108
Standard Deviation: Example: Annual Income of Graduates

- IQ = 102
- X = 105
- IQ = 108

68% within one standard deviation.
Confidence Intervals

**NOTE:** Confidence intervals can be constructed for any level of probability.

Most researchers are not confident enough to estimate a population mean knowing that there are only 68 out of 100 chances of being correct.

So, we generally use a wider, **less precise** confidence interval having a **better probability**.

The standard is 95% **confidence interval**, when we accept that there is a 5% chance that we could be wrong!
Confidence Intervals

So how do we find a 95% confidence interval?

1. We look at our normal distribution and know that 95.44% of the sample means lies between -$2 \sigma_X$ and $+2 \sigma_X$ from the mean of the means.

2. If we divide 95 by 2 (for both sides of the standard deviation from the means), we get 47.5.

3. Go to Table A and look in column b for 47.5 which gives us a z score of 1.96.

$$95\% \text{ confidence interval } = \overline{X} \pm 1.96 \sigma_X$$

If we apply this to our estimated mean from our sample, we would plug the values into the equation above and multiply 1.96 by the value we calculated for $\sigma_X$. 
Confidence Intervals

Confidence Intervals: Example: Private School IQ
How confident can the Dean be that the sample mean reflects the true pop. mean?

\[
\text{95\% confidence interval (CI)} = \overline{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}
\]

\[
\text{95\% confidence interval (CI)} = 105 +/- 1.96 (3.0) = 105 +/- 5.88
\]

\[
\text{95\% confidence interval (CI)} = 99.12 – 110.88
\]
Standard Deviation: Example: Annual Income of Graduates

The more “confident” you get the less precise you are pinpointing the pop. Mean.

IQ = 99.12

X = 105

IQ = 110.88

95% (slightly less than 95.44% or +/-2σx)
The t Distribution

At this point, we want to make statements about our population based on information that we have about our sample.

We know that our sample mean and sample distribution will be smaller than those for our population. That is, in any given sample, the sample variance ($s^2$) and the sample standard deviation ($s$) calculated from a sample mean will be smaller than if calculated from the population mean ($\mu$).

Specifically, sample variance and sample standard deviation tend to be smaller than population variance and the sample standard deviation.
The t Distribution

Since the sample estimates are biased (underestimation of variation), we will need to inflate them to produce more accurate estimates of the population variance and population standard deviation.

We do this by dividing by N – 1 rather than just N when calculating variance and standard deviation. By changing the denominator, the product is increased (thus making the confidence interval larger or less precise). The lack of confidence reflects the use of sample rather than population data.
Unbiased Estimates of the Population Variance and Population Standard Deviation

Population Variance:

\[ \hat{\sigma}^2 = \frac{\sum (X - \bar{X})^2}{N - 1} \]

Population Standard Deviation:

\[ \hat{\sigma} = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}} \]
Unbiased Estimates

An unbiased estimate of the standard error of the mean is:

$$s_{\bar{X}} = \frac{s}{\sqrt{N - 1}}$$

Where \( s \) is the sample standard deviation from a distribution of raw scores or from a frequency distribution.
The t Distribution

There is an complication when it comes to using an unbiased estimate of the standard error of the mean when calculating z scores in a normal distribution.

When we estimate the standard error using sampling data rather than population data, the sampling distribution of means is a bit wider (more dispersed).

That is when we use:

\[ z = \frac{\bar{X} - \mu}{S_x} \]

Instead of:

\[ z = \frac{\bar{X} - \mu}{\sigma_x} \]
So, the ratio does not allow us to follow the $z$ or normal distribution.

We will need to follow a **t distribution**, rather than a **z distribution**, which will give us a **t ratio**.

$$t = \frac{\bar{X} - \mu}{s \sqrt{\frac{1}{n}}}$$
The t Distribution

The **degrees of freedom** indicate how close the **t distribution** comes to approximating the **normal curve**. It is a means of indicating the difference between a sample and a probability distribution.

When estimating the **population mean**, the degrees of freedom are **one less than the sample size**.

\[
df = N - 1
\]

*Where*  
df = degrees of freedom  
N = sample size
The greater the degrees of freedom, the larger the sample size, the closer the \textit{t distribution} gets to the normal distribution. That is, the reliability of our estimate of the standard error of the mean increases as our sample increases, so $t$ approaches $z$.

Rather, the \textit{reliability} of our estimate of the standard error of the mean increases as our sample size increases, and so the \textit{t ratio} approaches a $z$ score.
When working with the t distribution, we use Table C instead of Table A.

Table C is calibrated for various levels of alpha (α).

Alpha is the area in the tails of the distributions and the formula for calculating alpha is

\[ \alpha = 1 - \text{level of confidence} \]

95% level of confidence, \( \alpha = .05 \)  
99% level of confidence, \( \alpha = .01 \)

To use Table C, we need two pieces of information:

1. The degrees of freedom (\( N - 1 \))
2. The alpha value (1 – level of confidence)
For example,

We want to construct a 95% confidence interval with a sample size of 20.

- Based on just this information, we know that we would have $df = 19$ and $\alpha = .05$.

We now go to Table C.

- In the df column we find 19 and in the $\alpha$ column we find .05.
- Where the two meet (at 2.093), is our t value.

Table C: Critical Values of t

<table>
<thead>
<tr>
<th>df</th>
<th>.20</th>
<th>.10</th>
<th>.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>1.328</td>
<td>1.729</td>
<td>2.093</td>
</tr>
</tbody>
</table>
The larger our sample, the larger our \( df \), and the smaller our \( t \) values.

The larger our \( df \) and the smaller the \( t \) values, the closer they approach the \( z \) or normal distribution.

SO, when the standard error of the mean is estimated, we can construct a confidence interval using the \( t \) values from Table C.

\[
CI = \overline{X} \pm ts \overline{X}
\]

**NOTE:** You *multiply* the \( t \) value by the standard error of the mean.

ALSO: \( t \) multiplied by \( s \) \( \overline{X} \) provides us with our margin of error.
So let’s pull it all together…

Suppose we have some sample data and we want to statements about the population from these data:

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>3</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

1. Find the mean of the sample.
2. Obtain the standard deviation of the sample
3. Find the estimated standard error of the mean
4. Determine the value of t from Table C
5. Find the margin of error
6. Calculate our confidence interval using t
Consider these statements to be truths that hold for all sample distribution of means:

1. The mean of a sample distribution of means (the mean of means) is equal to the true population mean.

2. The standard deviation of a sample distribution of means is smaller than the standard deviation of the population.

3. We can use the sample distribution of means to make generalizations from samples to populations.
**Sample and Populations**

- \( \bar{X} \): Mean of Sample
- \( \mu \): Mean of Population/Prob
- \( s \): Stand. Dev. of Sample
- \( \sigma \): Stand. Dev. of Population/Prob

**Sampling Error**

- Difference between Sample and Population

**Sample Distribution of Means (SDM)**

- Distribution of sample means
- Approx Normal Curve
- Means of Means = True pop mean
- SD of SDM is smaller than SD or pop

**SDM and Normal Curve**

- Can use Normal Cur. to make P statements about sample means
- Can use z scores (calculate SD of any mean within a SDM)

**Standard Error**

- Used to calculate standard dev. of SDM, without having a sampling dist. of sample means

\[ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} \]

- Pop. stand. dev divided by the square root of the sample size.

**Confidence Intervals**

- Est. range of true pop. mean.

\[ 68\% \quad \bar{X} \pm \sigma_{\bar{X}} \]

\[ 95\% \quad \bar{X} \pm 1.96\sigma_{\bar{X}} \]
Details:
- Used when we want to make statements about pop. from samples

The t distribution

Sample Variance and Stand. Dev.: biased
- Sample var. \((s^2)\) and sample stand. dev. \((s)\) is smaller than pop. var. and the pop. standard deviation.

Pop. Variance and Pop. Stand. Dev.: unbiased
- To “inflate” sample var. \((s^2)\) and sample stand. dev. \((s)\), we use \(N - 1\), rather than just \(N\).

Standard Error: Unbiased
- Used to calculate an unbiased standard error.

Formulas:
- When using sample, rather than pop. We cannot use \(z\).
- Use \(t\) scores (Table C)

\[
\hat{\sigma}^2 = \frac{\sum (X - \overline{X})^2}{N - 1}
\]

\[
\sigma^2 = \frac{\sum (X - \overline{X})^2}{\sqrt{N - 1}}
\]

\[
s_{\overline{X}} = \frac{s}{\sqrt{N - 1}}
\]

\[
t = \frac{\overline{X} - \mu}{s_{\overline{X}}}
\]

\[
df = N - 1
\]

\[
\alpha = 1 - \text{level of confidence}
\]

95/99% level of confidence, \(\alpha = .05/.01\)
Extra: Sample and Populations

Sample and Populations: Symbols
Now we have to carefully distinguish between samples and populations.

Mean:
\[ \bar{X} = \text{mean of a sample} \]
\[ \mu = (\text{mu}) \text{ mean of a population/probability} \]

Standard Deviation:
\[ s = \text{standard deviation of a sample} \]
\[ \sigma = (\text{sigma}) \text{ standard deviation of a population/probability} \]

Note: The symbols for population and probability distributions are the same.
Review: Samples and Populations

Type of Data: Standard Error (lack SDM: theoretical population data)

Population Standard Deviation: $\sigma$
Population Mean: $\mu$
Sample data: $N=?$

Formula: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$

95% Confidence Level: $= \bar{X} \pm 1.96\sigma_{\bar{X}}$
Review: Samples and Populations

Type of Data: Standard Error: Unbiased (No SDM, only sample data)
Sample Standard Deviation: $s$
Sample Mean: $\bar{X}$
Sample data: $N$

1. Formula: $s_{\bar{X}} = \frac{s}{\sqrt{N-1}}$

2. Degrees of freedom: $df = N - 1$
3. Alpha: $\alpha$: 1 – level of confidence (95%: 1 - .95 = .05)
4. Consult Table C to get t score (based on $df$ and $\alpha$)

4. 95% Confidence Level: $= \bar{X} \pm ts_{\bar{X}}$
Question Example: Standard Error: Unbiased (No SDM, only sample data)

Ch. 6: Q 16: A medical researcher interested in prenatal care interviews 35 randomly selected women and finds that they averaged 3 prenatal visits with a standard deviation of 1.

Type of Data?
Question Example: Standard Error: Unbiased (No SDM, only sample data)

N = 35 Women

Sample Mean ( ) = 3 Prenatal Visits Years of Practice

Sample Standard Deviation (s): 1 visit

Step 1: Calculate unbiased SE:

\[
\overline{S_X} = \frac{S}{\sqrt{N - 1}} = \frac{1}{5.8} = .17
\]

Step 2: Calculate t score:

df = N – 1 = 34
\[\alpha = 1 – \text{level of confidence} = .05 \ (95\%)\]
t = 2.042

Step 3: Calculate 95 % CI:

\[
\overline{X} \pm ts_{\overline{X}}
\]

\[
X +/- (2.042)(.17) = .35
\]

Answer:
\[2.65 = .35 – 3.0 + .35 = 3.35\]
Extra: Sampling Error

Sampling Error:
With random sampling, we can always expect some difference between a sample (random or nonrandom) and the population from which it was drawn.

This is known as sampling error. Sampling error is in every study no matter how rigorous the design or careful the researcher. As such, the sample mean (\(\bar{X}\)) will never be exactly the same as a population mean (\(\mu\)).