Chapter 9: Estimation from Sample Data

El Mechry, El Koudouss

Fordham University

April 21, 2020

Estimation from Sample Data

- In chapter 7, we had access to the population mean $\bar{x}$ and we made probability statements about individual x values taken from the population.
- In Chapter 8, we began with a population having a known mean μ or proportion p;
- Then we examined the sampling distribution of the corresponding sample statistic ($\bar{x}$ or $\hat{p}$) for samples of a given size, n.
- In this chapter, we'll be going in the opposite direction: based on sample data, we will be making estimates involving the (unknown) value of the population mean or proportion.

Point versus interval estimates

Definition

A ***point estimate*** of a population parameter is a single number that estimates the exact value of that parameter.

An ***interval estimate*** of a population parameter is an interval which includes a range of possible values that are likely to include the actual population parameter.

Example

Suppose the average GPA in a sample of 100 Fordham students is 3.3, what is the average GPA at Fordham?
point estimate $\mu = 3.3$
interval estimate: $\mu \in [2.8, 3.7]$

Unbiased Estimators

An estimator is ***unbiased*** if the expected value of the sample statistic is the same as the actual value of the population parameter it is intended to estimate. For example $\bar{x}$, $\hat{p}$, and s^2 are unbiased estimators of μ, p , and σ^2, respectively.

Parameter	Estimator	Formula	Expected Value
Mean: μ	$\bar{x}$	$\frac{\sum_{i=1}^{n} x_i}{n}$	$\mathbb{E}[\bar{x}] = \mu$
Proportion: p	$\hat{p}$	$\hat{p} = \frac{x}{n}$	$\mathbb{E}[\hat{p}] = p$
Variance, σ^2	s^2	$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)}$	$\mathbb{E}[\hat{s^2}] = \sigma^2$

Confidence Intervals: Definitions

- INTERVAL ESTIMATE: A range of values within which the actual value of the population parameter may fall.
- CONFIDENCE INTERVAL: An interval estimate for which there is a specified degree of certainty that the actual value of the population parameter will fall within the interval.
- CONFIDENCE LEVEL: This expresses the degree of certainty that an interval will include the actual value of the population parameter. It is usually stated as a percentage, commonly 90%, 95%, or 99%.
- LEVEL OF SIGNIFICANCE α: This expresses the the probability that an interval will NOT include the actual value of the population parameter. Note that $\alpha = 1 -$ Confidence Level

Estimating Confidence Intervals of the Mean

Suppose we take a simple random sample of size n from a population, and let $\bar{x}$ be the mean of this sample and s^2 its standard deviation. Estimating a confidence interval around the mean μ depends on whether or not the populations standard deviation σ is known:

$$\text{If } \sigma = \begin{cases} \text{is known} & \mu \in [\bar{x} \pm z * \frac{\sigma}{\sqrt{n}}] \\ \text{is unknown} & \mu \in [\bar{x} \pm t * \frac{s}{\sqrt{n}}], \text{ and } df = n - 1 \end{cases}$$

Where $z =$ the $z$score corresponding to the level of confidence desired. For example, $z = 1.96$ corresponds to the 95% confidence level.

We will come back to t latter.

Estimating Confidence Intervals of the Mean: σ is known

Suppose we take a simple random sample of size n, with $\bar{x}$, s^2. Also, suppose σ is known, then

$$\mu \in [\bar{x} \pm z * \frac{\sigma}{\sqrt{n}}]$$

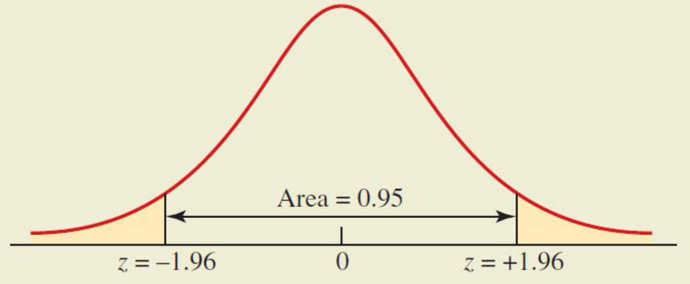
Where $z =$ the $z$score corresponding to the level of confidence desired.

Assumptions: this assumes that either (1) the underlying population is normally distributed or (2) the sample size is $n > 30$.

z Scores for Confidence Intervals

Commonly used confidence intervals and their corresponding z values:

Confidence	90%	95%	98%	99%
z	1.645	1.96	2.33	2.58



z Scores for Confidence Intervals: Example

Find the z score associated with 85% confidence level.

Answer:

z Scores for Confidence Intervals: Example

Find the z score associated with 85% confidence level.
Answer: We want z such that $P[-z < Z < z] = 0.85$. Notice, it's sufficient to find z or $-z$. So:

1. To find z, note that $P[-z < Z < z] = 0.85$ implies that $P[Z > z] = \frac{1-0.85}{2}$, due to the symmetry of the normal distribution.

z Scores for Confidence Intervals: Example

Find the z score associated with 85% confidence level.
Answer: We want z such that $P[-z < Z < z] = 0.85$. Notice, it's sufficient to find z or $-z$. So:

1. To find z, note that $P[-z < Z < z] = 0.85$ implies that $P[Z > z] = \frac{1-0.85}{2}$, due to the symmetry of the normal distribution.
   This implies that $P[Z < z] = 1 - \frac{1-0.85}{2} = 0.925$.

z Scores for Confidence Intervals: Example

Find the z score associated with 85% confidence level.
Answer: We want z such that $P[-z < Z < z] = 0.85$. Notice, it's sufficient to find z or $-z$. So:

1. To find z, note that $P[-z < Z < z] = 0.85$ implies that $P[Z > z] = \frac{1-0.85}{2}$, due to the symmetry of the normal distribution.
   This implies that $P[Z < z] = 1 - \frac{1-0.85}{2} = 0.925$.
   Using this probability with the Z table, we can find that $z = 1.44$.

z Scores for Confidence Intervals: Example

Find the z score associated with 85% confidence level.
Answer: We want z such that $P[-z < Z < z] = 0.85$. Notice, it's sufficient to find z or $-z$. So:

1. To find z, note that $P[-z < Z < z] = 0.85$ implies that $P[Z > z] = \frac{1-0.85}{2}$, due to the symmetry of the normal distribution.
   This implies that $P[Z < z] = 1 - \frac{1-0.85}{2} = 0.925$.
   Using this probability with the Z table, we can find that $z = 1.44$.
   Notice we can stop here because $-z = -1.44$

2. To find $-z$, note that $P[-z < Z < z] = 0.85$ implies, $P[-z < Z] = \frac{1-0.85}{2} = 0.075$ due to the symmetry of the normal distribution.

z Scores for Confidence Intervals: Example

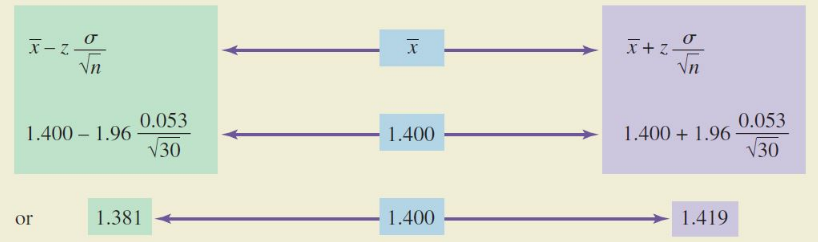
Find the z score associated with 85% confidence level.
Answer: We want z such that $P[-z < Z < z] = 0.85$. Notice, it's sufficient to find z or $-z$. So:

1. To find z, note that $P[-z < Z < z] = 0.85$ implies that $P[Z > z] = \frac{1-0.85}{2}$, due to the symmetry of the normal distribution.
   This implies that $P[Z < z] = 1 - \frac{1-0.85}{2} = 0.925$.
   Using this probability with the Z table, we can find that $z = 1.44$.
   Notice we can stop here because $-z = -1.44$

2. To find $-z$, note that $P[-z < Z < z] = 0.85$ implies, $P[-z < Z] = \frac{1-0.85}{2} = 0.075$ due to the symmetry of the normal distribution.
   Using this probability with the Z table, we can find that $-z = -1.44$.

Estimating Confidence Intervals of the Mean: σ is known, Example 1

From past experience, the population standard deviation of rod diameters produced by a machine has been found to be $\sigma = 0.053$ inches. For a simple random sample of $n = 30$ rods, the average diameter is found to be $\bar{x} = 1.4$ inches. What Is the 95% Confidence Interval for the Population Mean, μ?



Estimating Confidence Intervals of the Mean: σ is known, Example 2

The following data values are a simple random sample from a population that is normally distributed, with $\sigma = 4$: $\{8, 10, 7, 8, 5, 13, 7, 10, 4, 6\}$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n =$

Estimating Confidence Intervals of the Mean: σ is known, Example 2

The following data values are a simple random sample from a population that is normally distributed, with $\sigma = 4$: $\{8, 10, 7, 8, 5, 13, 7, 10, 4, 6\}$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 10$, $\bar{x} =$

Estimating Confidence Intervals of the Mean: σ is known, Example 2

The following data values are a simple random sample from a population that is normally distributed, with $\sigma = 4$: $\{8, 10, 7, 8, 5, 13, 7, 10, 4, 6\}$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 10$, $\bar{x} = 7.8$, and $\sigma = 4$. So depending on the confidence level:
90% CI: $\mu \in$

Estimating Confidence Intervals of the Mean: σ is known, Example 2

The following data values are a simple random sample from a population that is normally distributed, with $\sigma = 4$: $\{8, 10, 7, 8, 5, 13, 7, 10, 4, 6\}$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 10$, $\bar{x} = 7.8$, and $\sigma = 4$. So depending on the confidence level:
90% CI: $\mu \in [7.8 - 1.645\frac{4}{\sqrt{10}}\ ,\ 7.8 + 1.645\frac{4}{\sqrt{10}}] \Rightarrow \mu \in [5.7192, 9.8808]$
95% CI: $\mu \in$

Estimating Confidence Intervals of the Mean: σ is known, Example 2

The following data values are a simple random sample from a population that is normally distributed, with $\sigma = 4$: $\{8, 10, 7, 8, 5, 13, 7, 10, 4, 6\}$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 10$, $\bar{x} = 7.8$, and $\sigma = 4$. So depending on the confidence level:

90% CI: $\mu \in [7.8 - 1.645\frac{4}{\sqrt{10}}\ , 7.8 + 1.645\frac{4}{\sqrt{10}}] \Rightarrow \mu \in [5.7192, 9.8808]$

95% CI: $\mu \in [7.8 - 1.96\frac{4}{\sqrt{10}}\ , 7.8 + 1.96\frac{4}{\sqrt{10}}] \Rightarrow \mu \in [5.3208, 10.2792]$

99% CI: $\mu \in$

Estimating Confidence Intervals of the Mean: σ is known, Example 2

The following data values are a simple random sample from a population that is normally distributed, with $\sigma = 4$: $\{8, 10, 7, 8, 5, 13, 7, 10, 4, 6\}$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 10$, $\bar{x} = 7.8$, and $\sigma = 4$. So depending on the confidence level:

90% CI: $\mu \in [7.8 - 1.645\frac{4}{\sqrt{10}}\ ,\ 7.8 + 1.645\frac{4}{\sqrt{10}}] \Rightarrow \mu \in [5.7192, 9.8808]$

95% CI: $\mu \in [7.8 - 1.96\frac{4}{\sqrt{10}}\ ,\ 7.8 + 1.96\frac{4}{\sqrt{10}}] \Rightarrow \mu \in [5.3208, 10.2792]$

99% CI: $\mu \in [7.8 - 2.58\frac{4}{\sqrt{10}}\ ,\ 7.8 + 2.58\frac{4}{\sqrt{10}}] \Rightarrow \mu \in [4.5365, 11.0635]$

Estimating Confidence Intervals of the Mean: σ is known, Example 3

A simple random sample of 30 has been collected from a population for which it is known that $\sigma = 10$. The sample mean has been calculated as $\bar{x} = 240$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n =$

Estimating Confidence Intervals of the Mean: σ is known, Example 3

A simple random sample of 30 has been collected from a population for which it is known that $\sigma = 10$. The sample mean has been calculated as $\bar{x} = 240$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 30$, $\bar{x} =$

Estimating Confidence Intervals of the Mean: σ is known, Example 3

A simple random sample of 30 has been collected from a population for which it is known that $\sigma = 10$. The sample mean has been calculated as $\bar{x} = 240$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 30$, $\bar{x} = 240$ and $\sigma =$

Estimating Confidence Intervals of the Mean: σ is known, Example 3

A simple random sample of 30 has been collected from a population for which it is known that $\sigma = 10$. The sample mean has been calculated as $\bar{x} = 240$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 30$, $\bar{x} = 240$ and $\sigma = 10$. So depending on the confidence level:
90% CI: $\mu \in$

Estimating Confidence Intervals of the Mean: σ is known, Example 3

A simple random sample of 30 has been collected from a population for which it is known that $\sigma = 10$. The sample mean has been calculated as $\bar{x} = 240$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 30$, $\bar{x} = 240$ and $\sigma = 10$. So depending on the confidence level:
90% CI: $\mu \in [240 - 1.645\frac{10}{\sqrt{30}}, 240 + 1.645\frac{10}{\sqrt{30}}] \Rightarrow \mu \in [236.997, 243.003]$
95% CI: $\mu \in$

Estimating Confidence Intervals of the Mean: σ is known, Example 3

A simple random sample of 30 has been collected from a population for which it is known that $\sigma = 10$. The sample mean has been calculated as $\bar{x} = 240$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 30$, $\bar{x} = 240$ and $\sigma = 10$. So depending on the confidence level:

90% CI: $\mu \in [240 - 1.645\frac{10}{\sqrt{30}}, 240 + 1.645\frac{10}{\sqrt{30}}] \Rightarrow \mu \in [236.997, 243.003]$

95% CI: $\mu \in [240 - 1.96\frac{10}{\sqrt{30}}\ , 240 + 1.96\frac{10}{\sqrt{30}}] \Rightarrow \mu \in [236.4215, 243.5785]$

99% CI: $\mu \in$

Estimating Confidence Intervals of the Mean: σ is known, Example 3

A simple random sample of 30 has been collected from a population for which it is known that $\sigma = 10$. The sample mean has been calculated as $\bar{x} = 240$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 30$, $\bar{x} = 240$ and $\sigma = 10$. So depending on the confidence level:

90% CI: $\mu \in [240 - 1.645\frac{10}{\sqrt{30}}, 240 + 1.645\frac{10}{\sqrt{30}}] \Rightarrow \mu \in [236.997, 243.003]$

95% CI: $\mu \in [240 - 1.96\frac{10}{\sqrt{30}}, 240 + 1.96\frac{10}{\sqrt{30}}] \Rightarrow \mu \in [236.4215, 243.5785]$

99% CI: $\mu \in [240 - 2.58\frac{10}{\sqrt{30}}, 240 + 2.58\frac{10}{\sqrt{30}}] \Rightarrow \mu \in [235.2896, 244.7104]$

Estimating Confidence Intervals of the Mean: σ is known, Example 4

A simple random sample of 25 has been collected from a normally distributed population for which it is known that $\sigma = 17$. The sample mean has been calculated as 342.0, and the sample standard deviation is $s = 14.9$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n =$

Estimating Confidence Intervals of the Mean: σ is known, Example 4

A simple random sample of 25 has been collected from a normally distributed population for which it is known that $\sigma = 17$. The sample mean has been calculated as 342.0, and the sample standard deviation is $s = 14.9$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 25$, $\bar{x} =$

Estimating Confidence Intervals of the Mean: σ is known, Example 4

A simple random sample of 25 has been collected from a normally distributed population for which it is known that $\sigma = 17$. The sample mean has been calculated as 342.0, and the sample standard deviation is $s = 14.9$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 25$, $\bar{x} = 342$, and $\sigma = 17$. So depending on the confidence level:
90% CI: $\mu \in$

Estimating Confidence Intervals of the Mean: σ is known, Example 4

A simple random sample of 25 has been collected from a normally distributed population for which it is known that $\sigma = 17$. The sample mean has been calculated as 342.0, and the sample standard deviation is $s = 14.9$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 25$, $\bar{x} = 342$, and $\sigma = 17$. So depending on the confidence level:

90% CI: $\mu \in [342 - 1.645\frac{17}{\sqrt{25}}\ , 342 + 1.645\frac{17}{\sqrt{25}}] \Rightarrow \mu \in [336.407, 347.593]$

95% CI: $\mu \in$

Estimating Confidence Intervals of the Mean: σ is known, Example 4

A simple random sample of 25 has been collected from a normally distributed population for which it is known that $\sigma = 17$. The sample mean has been calculated as 342.0, and the sample standard deviation is $s = 14.9$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 25$, $\bar{x} = 342$, and $\sigma = 17$. So depending on the confidence level:

90% CI: $\mu \in [342 - 1.645\frac{17}{\sqrt{25}}\ , 342 + 1.645\frac{17}{\sqrt{25}}] \Rightarrow \mu \in [336.407, 347.593]$

95% CI: $\mu \in [342 - 1.96\frac{17}{\sqrt{25}}\ , 342 + 1.96\frac{17}{\sqrt{25}}] \Rightarrow \mu \in [335.336, 348.664]$

99% CI: $\mu \in$

Estimating Confidence Intervals of the Mean: σ is known, Example 4

A simple random sample of 25 has been collected from a normally distributed population for which it is known that $\sigma = 17$. The sample mean has been calculated as 342.0, and the sample standard deviation is $s = 14.9$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the population mean, μ.

Answer: Notice that $n = 25$, $\bar{x} = 342$, and $\sigma = 17$. So depending on the confidence level:
90% CI: $\mu \in [342 - 1.645\frac{17}{\sqrt{25}}\ , 342 + 1.645\frac{17}{\sqrt{25}}] \Rightarrow \mu \in [336.407, 347.593]$
95% CI: $\mu \in [342 - 1.96\frac{17}{\sqrt{25}}\ , 342 + 1.96\frac{17}{\sqrt{25}}] \Rightarrow \mu \in [335.336, 348.664]$
99% CI: $\mu \in [342 - 2.58\frac{17}{\sqrt{25}}\ , 342 + 2.58\frac{17}{\sqrt{25}}] \Rightarrow \mu \in [333.228, 350.772]$

Estimating Confidence Intervals of the Mean: σ is known, Example 5

You want to rent a one-bedroom apartment in NYC. You find the mean monthly rent for a random sample of 60 apartments advertised on Craig's List is \$1600. Assume a population standard deviation of \$400. Construct 90%, 95%, and 99% confidence intervals for the average monthly one-bedroom rent in NYC.

Answer: Notice that $n =$

Estimating Confidence Intervals of the Mean: σ is known, Example 5

You want to rent a one-bedroom apartment in NYC. You find the mean monthly rent for a random sample of 60 apartments advertised on Craig's List is \$1600. Assume a population standard deviation of \$400. Construct 90%, 95%, and 99% confidence intervals for the average monthly one-bedroom rent in NYC.

Answer: Notice that $n = 60$, $\bar{x} =$

Estimating Confidence Intervals of the Mean: σ is known, Example 5

You want to rent a one-bedroom apartment in NYC. You find the mean monthly rent for a random sample of 60 apartments advertised on Craig's List is \$1600. Assume a population standard deviation of \$400. Construct 90%, 95%, and 99% confidence intervals for the average monthly one-bedroom rent in NYC.

Answer: Notice that $n = 60$, $\bar{x} = 342$, and $\sigma = 17$. So depending on the confidence level:
90% CI: $\mu \in$

Estimating Confidence Intervals of the Mean: σ is known, Example 5

You want to rent a one-bedroom apartment in NYC. You find the mean monthly rent for a random sample of 60 apartments advertised on Craig's List is \$1600. Assume a population standard deviation of \$400. Construct 90%, 95%, and 99% confidence intervals for the average monthly one-bedroom rent in NYC.

Answer: Notice that $n = 60$, $\bar{x} = 342$, and $\sigma = 17$. So depending on the confidence level:

90% CI: $\mu \in [1600 - 1.645\frac{400}{\sqrt{60}}\ , 1600 + 1.645\frac{400}{\sqrt{60}}] \Rightarrow \mu \in [1515.1, 1684.9]$

95% CI: $\mu \in$

Estimating Confidence Intervals of the Mean: σ is known, Example 5

You want to rent a one-bedroom apartment in NYC. You find the mean monthly rent for a random sample of 60 apartments advertised on Craig's List is \$1600. Assume a population standard deviation of \$400. Construct 90%, 95%, and 99% confidence intervals for the average monthly one-bedroom rent in NYC.

Answer: Notice that $n = 60$, $\bar{x} = 342$, and $\sigma = 17$. So depending on the confidence level:

90% CI: $\mu \in [1600 - 1.645\frac{400}{\sqrt{60}}\ , 1600 + 1.645\frac{400}{\sqrt{60}}] \Rightarrow \mu \in [1515.1, 1684.9]$

95% CI: $\mu \in [1600 - 1.96\frac{400}{\sqrt{60}}\ , 1600 + 1.96\frac{400}{\sqrt{60}}] \Rightarrow \mu \in [1498.8, 1701.2]$

99% CI: $\mu \in$

Estimating Confidence Intervals of the Mean: σ is known, Example 5

You want to rent a one-bedroom apartment in NYC. You find the mean monthly rent for a random sample of 60 apartments advertised on Craig's List is \$1600. Assume a population standard deviation of \$400. Construct 90%, 95%, and 99% confidence intervals for the average monthly one-bedroom rent in NYC.

Answer: Notice that $n = 60$, $\bar{x} = 342$, and $\sigma = 17$. So depending on the confidence level:

90% CI: $\mu \in [1600 - 1.645\frac{400}{\sqrt{60}}\ , 1600 + 1.645\frac{400}{\sqrt{60}}] \Rightarrow \mu \in [1515.1, 1684.9]$

95% CI: $\mu \in [1600 - 1.96\frac{400}{\sqrt{60}}\ , 1600 + 1.96\frac{400}{\sqrt{60}}] \Rightarrow \mu \in [1498.8, 1701.2]$

99% CI: $\mu \in [1600 - 2.58\frac{400}{\sqrt{60}}\ , 1600 + 2.58\frac{400}{\sqrt{60}}] \Rightarrow \mu \in [1466.8, 1733.2]$

Confidence Interval Estimate for the Population Proportion

Suppose a proportion p of individuals in a population have a certain trait, and assume we don't know the value of p. Take a sample of size n, and let $\hat{p}$ be the proportion of individuals in the sample with that trait. We can construct a confidence interval for the population proportion p as follows:

$$p \in \left[\hat{p} \pm z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = \left[\hat{p} \pm z * \sqrt{\frac{\hat{p}\hat{q}}{n}}\right]$$

This assumes the normal distribution as an approximation to the binomial distribution, which holds whenever $n\hat{p} > 5$ and $n\hat{q} > 5$. The approximation becomes better for large values of n and whenever $\hat{p}$ is closer to 0.5.

Confidence Interval Estimate for the Population Proportion: Example 1

Out of a sample of 1008 adults, 22% responded YES to a survey question. What is the 95% confidence interval for the population proportion who would have answered "YES" to the same question?

Answer: $\hat{p} = 0.22$, and $z = \pm 1.96$, so

$$p \in \left[\hat{p} \pm z * \sqrt{\frac{\hat{p}\hat{q}}{n}}\right] = \left[0.22 \pm 1.96 * \sqrt{\frac{0.22 * 0.78}{1008}}\right] = [0.194\ ,\ 0.246]$$

Confidence Interval Estimate for the Population Proportion: Example 2

A pharmaceutical company found that 46% of 1000 U.S. adults surveyed knew neither their blood pressure nor their cholesterol level. Assuming the persons surveyed to be a simple random sample of U.S. adults, construct 99% confidence interval for p, the population proportion of U.S. adults who would have given the same answer if a census had been taken instead of a survey.

Answer: $\hat{p} = 0.46$, and $z = \pm 2.58$, so 99% CI:

$p \in$

Confidence Interval Estimate for the Population Proportion: Example 2

A pharmaceutical company found that 46% of 1000 U.S. adults surveyed knew neither their blood pressure nor their cholesterol level. Assuming the persons surveyed to be a simple random sample of U.S. adults, construct 99% confidence interval for p, the population proportion of U.S. adults who would have given the same answer if a census had been taken instead of a survey.

Answer: $\hat{p} = 0.46$, and $z = \pm 2.58$, so 99% CI:

$$p \in \left[\hat{p} \pm z * \sqrt{\frac{\hat{p}\hat{q}}{n}}\right] = \left[0.46 \pm 2.58 * \sqrt{\frac{0.46 * 0.54}{1000}}\right] = [0.42, 0.50]$$

Confidence Interval Estimate for the Population Proportion: Example 3

An airline has surveyed a simple random sample of air travelers to find out whether they would be interested in paying a higher fare in order to have access to e-mail during their flight. Of the 400 travelers surveyed, 80 said email access would be worth a slight extra cost. Construct 90% confidence interval for the population proportion of air travelers who are in favor of the airline's email idea.

Answer: $\hat{p} = \frac{80}{400} = 0.2$, and $z = \pm 1.645$, so 99% CI:

$p \in$

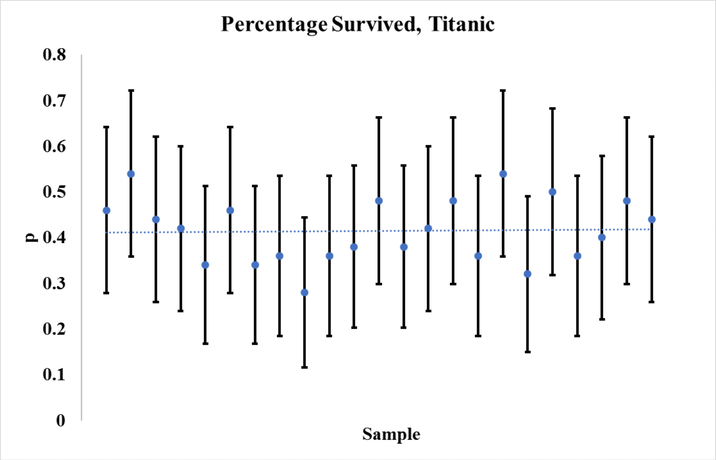
Confidence Interval Estimate for the Population Proportion: Example 3

An airline has surveyed a simple random sample of air travelers to find out whether they would be interested in paying a higher fare in order to have access to e-mail during their flight. Of the 400 travelers surveyed, 80 said email access would be worth a slight extra cost. Construct 90% confidence interval for the population proportion of air travelers who are in favor of the airline's email idea.

Answer: $\hat{p} = \frac{80}{400} = 0.2$, and $z = \pm 1.645$, so 99% CI:

$$p \in \left[\hat{p} \pm z * \sqrt{\frac{\hat{p}\hat{q}}{n}}\right] = \left[0.2 \pm 1.645 * \sqrt{\frac{0.2*0.8}{400}}\right] = [0.18, 0.22]$$

Titanic: Estimating the Proportion Survived



Estimating Confidence Intervals of the Mean: σ is unknown

Suppose we take a simple random sample of size n, with mean $\bar{x}$ and standard deviation s. Also, suppose σ is unknown, then

$$\mu \in [\bar{x} \pm t * \frac{s}{\sqrt{n}}]$$

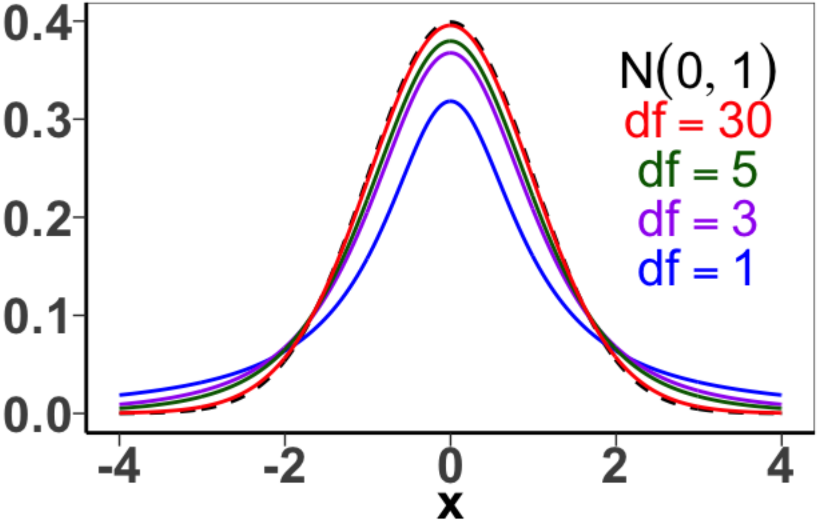
Where $t =$ the t score corresponding to the level of confidence with $n-1$ degrees of freedom, or $df = n-1$.

Assumptions: this assumes that either (1) the underlying population is normally distributed or (2) the sample size is $n > 30$.

The Student's t Distribution

- It is rare that we know the standard deviation, σ, of a population but have no knowledge about its mean, μ.
- Whenever the population standard deviation,σ, is unknown, it must be estimated by the sample standard deviation, s.
- There is a continuous distribution called the Student's t distribution that allows you to do this.
- It has a mean of zero, but its shape is determined by the number of *degrees of freedom (df)*.

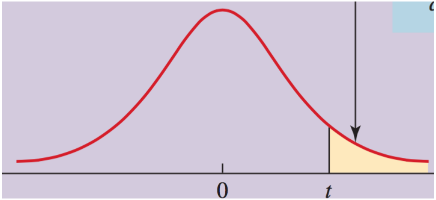
Comparing the Normal and t distributions



Using The t Distribution Table: Example 1

- Like a z score, t represents the distance in terms of standard error (standard deviation) units.
- Unlike the z Table, the t Table provides areas to the ***right*** of the t score given the number of degrees of freedom (df). This area is called α.
- Example: For a sample size of $n = 15$, what t values would correspond to an area centered at $t = 0$ and having an area beneath the curve of 95%?

$n = 15$, $df = n - 1 = 14$, and the area under the curve is 0.95. This leaves 0.05, equally distributed between the two tales. We want t to the right of which the area under the curve is 0.025. From the t Table, $\alpha = 0.025$, $df = 14$, we have: $t = \pm 2.145$. Compare this to $z = \pm 1.96$.



Using The t Distribution Table: Example 2

For a sample size of $n = 99$, what t values would correspond to an area centered at $t = 0$ and having an area beneath the curve of 90%?

$n = 99$, $df = n - 1 = 98$, and the area under the curve is 0.90. This leaves 0.1, equally distributed between the two tales. We want t to the right of which the area under the curve is 0.05. From the t Table, $\alpha = 0.05$, $df = 98$, we have: $t = \pm 1.661$. Compare this to $z = \pm 1.645$.

Should you encounter a situation in which the number of degrees of freedom exceeds the $df = 100$ limit of the t distribution table, just use the corresponding z value for the desired level of confidence.

Using The t Distribution Table: Example 3

For a sample size of $n = 31$, what t values would correspond to an area centered at $t = 0$ and having an area beneath the curve of 99%?

$n = 31$, $df = n - 1 = 30$, and the area under the curve is 0.99. This leaves 0.01, equally distributed between the two tales. We want t to the right of which the area under the curve is 0.005. From the t Table, $\alpha = 0.005$, $df = 30$, we have $t = \pm 2.750$. Compare this to $z = \pm 2.58$.

Using The t Distribution Table: Example 4

For a sample size of $n = 51$, what t values would correspond to an area centered at $t = 0$ and having an area beneath the curve of 80%?

$n = 51$, $df = n - 1 = 50$, and the area under the curve is 0.80. This leaves 0.2, equally distributed between the two tales. We want t to the right of which the area under the curve is 0.1. From the t Table, $\alpha = 0.1$, $df = 50$, we have $t = \pm 1.299$. Compare this to $z = \pm 1.2816$.

Estimating Confidence Intervals of the Mean: σ is unknown, Example 1

A simple random sample of $n = 90$ manufacturing employees has been selected. The average number of overtime hours worked last week was $\bar{x} = 8.46$ hours, with a sample standard deviation of $s = 3.61$ hours. What is the 98% confidence interval for the population mean, μ?

Answer: Here σ is unknown, so we use the sample standard deviation s instead, and this requires using the t distribution.

$$\alpha = \tfrac{1-0.98}{2} = 0.01,\ df = 89, \text{ so } t = \pm 2.369$$

$$\mu \in \left[\bar{x} \pm t * \frac{s}{\sqrt{n}}\right] = \left[8.46 \pm 2.369 * \frac{3.61}{\sqrt{90}}\right] = [7.56\ ,\ 9.36]$$

Estimating Confidence Intervals of the Mean: σ is unknown, Example 2

A simple random sample of $n = 105$ manufacturing employees has been selected. The average number of overtime hours worked last week was $\bar{x} = 8.46$ hours, with a sample standard deviation of $s = 3.61$ hours. What is the 98% confidence interval for the population mean, μ?

Answer: $\alpha = \frac{1-0.98}{2} = 0.01$, $df = 104$, so we can use the z table instead of the t table. Recall that the z score associated with 98% confidence interval is $z = t = \pm 2.33$

$$\mu \in \left[\bar{x} \pm t * \frac{s}{\sqrt{n}}\right] = \left[8.46 \pm 2.33 * \frac{3.61}{\sqrt{90}}\right] = [7.57\ ,\ 9.35]$$

Estimating Confidence Intervals of the Mean: σ is unknown, Example 3

Suppose a random sample from normally distributed student grades yields the following: $\{70, 78, 74, 98, 74, 72, 60, 100, 90, 94\}$. Construct and interpret the 90%, 95%, and 99% confidence intervals for the mean:

Answer: Here $n = 10$, and you can calculate $\bar{x}$ and s from the data to verify that $\bar{x} = 81$, and $s = 13.54$. Since the sample size is too small, we rely on the assumption that the underlying population is normally distributed. For the 90% CI:

$$\alpha = \tfrac{1-0.90}{2} = 0.05,\ df = 10 - 1 = 9, \text{ so } t = \pm 1.833$$

$$\mu \in \left[\bar{x} \pm t * \frac{s}{\sqrt{n}}\right] = \left[81 \pm 1.833 * \frac{13.54}{\sqrt{10}}\right] = [73.15\ ,\ 88.85]$$

Estimating Confidence Intervals of the Mean: σ is unknown, Example 4

In a sample of 30 current MLB pitchers, the mean age was 28 years with a standard deviation of 4.4 years. Construct a 95% confidence interval to estimate the mean age of all current MLB pitchers.

Answer:

$$\alpha = \frac{1-0.95}{2} = 0.025,\ df = 30 - 1 = 29, \text{ so } t = \pm 2.045$$

$$\mu \in \left[\bar{x} \pm t * \frac{s}{\sqrt{n}}\right] = \left[28 \pm 2.045 * \frac{4.4}{\sqrt{30}}\right] = [26.36\ ,\ 29.64]$$

Sample Size Determination

Recall that

$$\mu \in \left[\bar{x} \pm z * \frac{\sigma}{\sqrt{n}}\right] \text{ and } p \in \left[\hat{p} \pm z * \sqrt{\frac{\hat{p}\hat{q}}{n}}\right]$$

Definition

In a confidence interval, the margin of error, or maximum likely sampling error, denoted e, is defined as:

$$e = z * \frac{\sigma}{\sqrt{n}} \text{ , for the population mean } \mu \text{ , and}$$

$$e = z * \sqrt{\frac{\hat{p}\hat{q}}{n}} \text{ , for the population proportion } p$$

.

Sample Size Determination for the Mean

We can set the margin of error in advance, and choose a sample size which guarantees achieving that margin of error. If we let e be the margin of error, then:

$$
\begin{aligned}
e &= z * \frac{\sigma}{\sqrt{n}} \\
\Rightarrow e^2 &= z^2 * \frac{\sigma^2}{n} \\
\Rightarrow n &= \frac{z^2 * \sigma^2}{e^2}
\end{aligned}
$$

So, after setting e, we can use a sample size of $n = \frac{z^2 * \sigma^2}{e^2}$ to guarantee that the mean will be within the margin of error. If σ is unknown, use s instead.

Sample size determination for μ: Example

Determine the sample size required to estimate the average summer earnings among teenagers with 95% confidence level and a \$50 margin of error. Suppose $\sigma = \$400$.
Answer:

$$n = \frac{z^2 * \sigma^2}{e^2}$$
$$\Rightarrow n = \frac{1.96^2 * 400^2}{50^2} \approx 246$$

Notice that we don't know μ, nor $\bar{x}$, but we are 95% confident that $\mu \in [\bar{x} \pm \$50]$ as long as $n = 246$ and $\sigma = \$400$.
So, if you find that $\bar{x} = \$1000$ in a simple random sample of 246 teenagers from this population, you can conclude that $\mu \in [\$1000 \pm \$50]$

Sample Size Determination for the Proportion

again, we can set the margin of error in advance, and choose a sample size which guarantees achieving that margin of error. If we let e be the margin of error, then:

$$e = z * \sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$\Rightarrow e^2 = z^2 * \frac{\hat{p}\hat{q}}{n}$$
$$\Rightarrow n = \frac{z^2 * \hat{p}\hat{q}}{e^2}$$

As a conservative strategy, use $\hat{p} = 0.5$ if you have no idea about the actual value of p.
So, after setting e, we can use a sample size of $n = \frac{z^2 * \hat{p}\hat{q}}{e^2}$ to guarantee that the mean will be within the margin of error.

Sample Size Determination for the Proportion: Example

Suppose we want to estimate the proportion, p, of people who vacation in Mexico. What sample size is necessary to be 95% confident that the sample proportion will be within 0.03 (3 percentage points) of the actual population proportion?

Answer: $z = 1.96$, $e = 0.03$, and $\hat{p} = 0.5$ since we don't know p. So,

$$n = \frac{z^2 * \hat{p}\hat{q}}{e^2} = \frac{1.96^2 * 0.5 * 0.5}{0.03^2} \approx 1068$$

Confidence Intervals when Population is Finite

$$\text{If } = \begin{cases} \sigma \text{ known, infinite population} & \mu \in [\bar{x} \pm z * \frac{\sigma}{\sqrt{n}}] \\ \sigma \text{ unknown, infinite population} & \mu \in [\bar{x} \pm t * \frac{s}{\sqrt{n}}] \text{ , } df = n-1 \\ \text{infinite population} & p \in= \left[\hat{p} \pm z * \sqrt{\frac{\hat{p}\hat{q}}{n}}\right] \\ \sigma \text{ known, population finite} & \mu \in [\bar{x} \pm z * \left(\frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}}\right)] \\ \sigma \text{ unknown, population finite} & \mu \in [\bar{x} \pm t * \left(\frac{s}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}}\right)] \text{ , } df = n-1 \\ \text{population finite} & p \in \left[\hat{p} \pm z * \left(\sqrt{\frac{\hat{p}\hat{q}}{n}} * \sqrt{\frac{N-n}{N-1}}\right)\right] \end{cases}$$

Confidence Intervals when Population is Finite: Example

The population of Hinsdale County, is 838 persons. Assume a researcher has interviewed a simple random sample of 400 persons and found that their average number of years of formal education is $\bar{x} = 11.5$ years, with a standard deviation of $s = 4.3$ years. Construct a 95% CI for the population mean.

Answer: Notice that the sample size is large relative to the population, $n > 0.05 * N$. $\alpha = \frac{1-0.95}{2} = 0.025$, $df = 400 - 1 = 399$, so use the z score associated with 95% CI: $z = \pm 1.96$.

$$\mu \in \left[\bar{x} \pm z * \left(\frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}}\right)\right]$$

$$\mu \in \left[11.5 \pm 1.96 * \left(\frac{4.3}{\sqrt{400}} * \sqrt{\frac{838-400}{838-1}}\right)\right] = [11.195\ ,\ 11.805]$$