# Chapter 8

Matt Warner

# Chapter 8 - Confidence Intervals Based on a Single Sample

**Satistical inference -** using a sample statistic (e.g. $\overline{x}$ = sample mean) to make some statement or some conclusion about a population parameter (e.g. $\mu$ = population mean) There are two types.

- **Confidence interval**
  - Uses a sample statistic to *estimate* the unknown value of a parameter.
  - We write the estimate in the form of an interval that we belive captures the actual or true parameter value.
  - The confidence interval has a specified level of confidence.
- **Hypothesis Test** (or significance test)

## Section 8.2 - A Confidence Interval for a Population Mean when $\sigma$ is Known

- Provided
  - Underlying population has a normal distribution or $n$ is large ($n \geq 30$)
  - $\sigma$ (population standard deviation) has a known value
- A $100(1-a)\%$ confidence interval to estimate $\mu$ is:

$$\overline{x} \pm \left( Z_{\frac{a}{z}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

- The **critical value** ($Z_a z$) is chosen based on the **level of confidence**

  $\pm Z_{\frac{a}{z}}$ = values from a standard normal distribution that capture the middle $100(1-a)\%$
- Simplified form
  - estimate $\pm$ margin of error
  - **margin of error** = (critical value x **standard error of $\overline{\mathbf{x}}$**)

## Question 1

**An administrator at a large university wants to estimate $\mu$, the mean GPA of all students on campus. A random sample of $n = 50$ students is selected and the GPA of each student is recorded. The resulting mean is $\overline{x} = 2.60$. Assume that the GPAs in the population are normally distributed with $\sigma = 0.75$**

**Problem 1. Calculate a 95% confidence interval**

***its asking for 95%, so our tails are***

$$\alpha = .05 \rightarrow \frac{\alpha}{2} = 0.25$$

***our z scores are then***

$$\pm 1.96$$

***So, using the formula***

$$\overline{x} \pm \left( Z_{\frac{a}{z}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

***We have***

$$2.60 \pm (1.96 \cdot \frac{.75}{\sqrt{50}})$$

***our confidence intervals are then***

$$2.60 \pm .21$$

$$= 2.39, 2.81$$

---

**Problem 2. Calculate a 90% confidence interval to estimate $\mu$**

***If its asking for 90% are tails are***

$$0.05$$

***Our z scores are then***

$$\pm 1.6499$$

***So, using the formula***

$$\overline{x} \pm \left( Z_{\frac{a}{z}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

***We have***

$$2.60 \pm (1.6449 \cdot \frac{.75}{\sqrt{50}})$$

***So, our confidence intervals are***

$$2.60 \pm .17$$

$$= 2.43, 2.77$$

---

**Problem 3. Calculate a 99% confidence interval to estimate $\mu$**

***99% means are tails are***

$$.005$$

***our z scores would then be***

$$\pm 2.5758$$

***So***

$$2.60 \pm (2.5758 \cdot \frac{.75}{\sqrt{50}})$$

***So, our confidence intervals are***

$$2.60 \pm .27$$

$$2.33, 2.87$$

## The effect of choosing the confidence level and sample size

The confidence level (90%, 95%, 99% etc. ) and the sample size ($n$) are chosen by the statistician or researcher. It is important to understand how these choices affect the overall length of the confidence interval.

- If the confidence level **increases** then the interval gets **wider** and **less** precise
- If the confidence level **decreases**, then the interval gets **narrower** and **more** precise
- If the sample size $n$ **increases** then the interval gets **narrower** and **more** precise
- If the sample size $n$ **decreases**, then the interval gets **wider** and **less** precise

## Question 2

Which confidence interval would be longer and which would be shorter?

(a) 90% confidence and $n = 50$
  95% confidence and $n = 50$

***Solution:***

95% confidence interval would be longer than 90%

(b) 95% confidence and $n = 50$
  95% confidence and $n = 100$

***Solution:***

95% confidence and $n = 50$ would be longer

## Meaning of confidence

In the previous example

- $\mu$ (= mean GPA of the entire campus) is **unknown** but it has a **fixed value**
- After getting the sample and making our calculation, our 95% confidence interval is also **fixed** with endpoints 2.39 to 2.81
- But, if we took another sample we would likely get a different $\bar{x}$ and **different** endpoints than before, and we would still state that we are confident the new interval captures $\mu$.

## Sample size calculation

**Before** the sample is picked

- we specify the desired
  - confidence level
  - bound for the margin of error (B)
- we ask: What size sample is needed?

$$n = \left( \frac{\sigma \cdot Z_{\frac{\alpha}{2}}}{B} \right)^2$$

> **Note:-**
>
> **ALways** round $n$ up to a whole number

## 8-3 A Confidence Interval for a Population Mean when $\sigma$ is Unknown

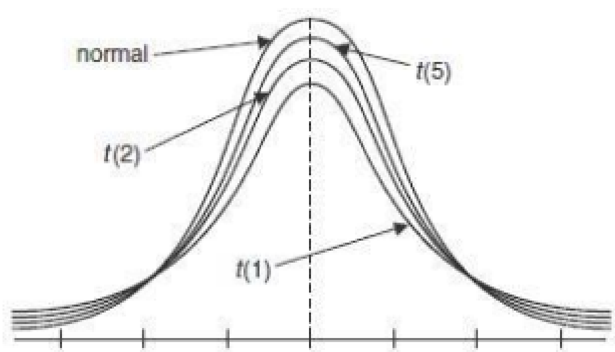
- Provided
  - Underlying population has a normal distribution
  - $\sigma$ (population standard deviation) has an **unknown value**
  - $s$ (sample standard deviation) has a **known value**
- A 100(1 - a)% confidence interval to estimate $\mu$ is

$$\bar{x} \pm \left(t_{\frac{a}{z}} \cdot \frac{s}{\sqrt{n}}\right)$$

- $\pm t_{\frac{a}{z}}$ = critical values from a **t distribution** (with **degree of freedom** $df = n-1$) that capture the middle 100(1 - $a$)%

### Facts about the t distribution

- It is described by a symmetric, bell-shaped curve that is centered at 0.
- There are many t distributions. Each is identified by giving its degree of freedom ($df$).
- It is wider and has more spread than the standard normal distribution
- As the $df$ increases the t distribution looks more like the standard normal distribution.




### Question 3

In each of the following, find the appropriate t critical value for use in constructing a condidence interval

**Problem 1.** n = 10, 90% confident

$$df = n - 1$$

$$df = 9$$

$$t_{.05,9} = 1.8331$$

**Problem 2.** n = 15, 95% confidence

$$df = 14$$

***95% confidence means our tails are both .025***

$$t_{.025,14} = 2.1448$$

**n = 21, 99% confidence**

$$df = 20$$

***our tails our .005, so looking up 20df and .005, we get***

$$t_{.005,20} = 2.8453$$

## Question 4

Oil obtained from orange blossoms through distillation is used in perfume. Suppose the oil yield is normally distributed. In a random sample of 11 distillations, the sample mean oil yield was $\bar{x} = 980.2$ g with standard deviation $s = 27.6$ g.

**Find a 95% confidence interval for the true mean oil yield per batch.**

$$\overline{x} \pm \left(T\frac{s}{\sqrt{n}}\right)$$

$$980.2 \pm \left(T\frac{27.6}{\sqrt{11}}\right)$$

$$df = 10$$

$$\text{tails} = .025$$

$$t_{.025,10} = 2.2281$$

***So,***

$$980.2 \pm \left(2.2281\frac{27.6}{\sqrt{11}}\right)$$

$$980.2 \pm 18.5$$

***We are 95% confident that $\mu$ is between 961.7, and 998.7***

## Question 5

The earth is structured in layers: crust, mantle, and core. A recent study was conducted to estimate the mean depth of the upper mantle in a specific farming region of California. Twenty-six sample sites were selected at random, and the depth of the upper mantle was measured using changes in seismic velocity and density. The sample data resulted in a mean of **127.5 km** and a standard deviation of **21.3 km**. Suppose the depth of the upper mantle is normally distributed. Find a **90%** confidence interval for the true mean depth of the upper mantle in this farming region.

***Using,***

$$\overline{x} \pm \left(T\frac{s}{\sqrt{n}}\right)$$

$$127.5 \pm \left(T\frac{21.3}{\sqrt{26}}\right)$$

$$df = 25 \qquad \text{our tails are } .05$$

***finding the value in the t-table***

$$t_{.05,25} = 1.7081$$

***So,***

$$127.5 \pm \left(1.7081\frac{21.3}{\sqrt{26}}\right)$$

$$127.5 \pm 1.7081$$

***We are 90% confident that $\mu$ is between 120.4, and 134.6***

---

**Section - 8.4**

## A large-Sample Confidence Interval for a Population Proportion

Previously

- ◦ Measurements or data values were **quantitative**
- ◦ Population parameter was $\mu$ = population mean

Now

- ◦ Measurements or data values will be **qualitative**
  - Status of each child's vision (impaired, not)
  - Quality of each manufactured part (defective, not)
  - Voting preference (for, against)
- ◦ Population parameter

  $p$ = proportion of individuals in the **population** with a specified characteristic
- ◦ Sample statistic

  $\hat{p}$ = proportion of individuals in the **sample** with the specified characteristic

## Confidence interval to estimate a population proportion $p$

- Provided
  - $n$ is large enough that both $n\hat{p} \geqslant 5$ and $n(1-\hat{p}) \geqslant 5$ are true
    (Comparing each against the values 10 or 15 are common alternatives conditions)
- A $100(1-a)\%$ confidence interval to estimate $p$ is

$$\hat{p} \pm \left(Z_{\frac{a}{z}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

- $\pm Z_{\frac{a}{2}}$ = critical values from a standard normal distribution that captures the middle $100(1-a)\%$
- Simplified form
  - estimate $\pm$ margin of error
  - margin of error = (critical value $\cdot$ standard error of $\hat{p}$)

## Question 6

A random sample of 1012 American adults was selected and 385 said that they believe in ghosts. Calculate a 95% confidence interval to estimate $p$, the true percent of all American adults who feel similarly.

***We have,***

$$n = 1012$$

$$\hat{p} = \frac{385}{1012} = .38$$

***Our z-score for a 95% confidence interval is***

$$Z = 1.96$$

***So, using the formula***

$$\hat{p} \pm \left( Z_{\frac{a}{z}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

***We have***

$$.38 \pm \left( 1.96 \cdot \sqrt{\frac{.38(1-.38)}{1012}} \right)$$

$$= .38 \pm .02991$$

$$= (0.35, 0.41)$$

---

## Sample size calculation

- Before the sample is picked we specify the desired
  - confidence level
  - bound for the margin of error (B)
- we ask: What size sample is needed?

$$n = \hat{p}(1-\hat{p}) \left( \frac{Z_{\frac{a}{2}}}{B} \right)^2$$

- Problem
  - We need $\hat{p}$ to get $n$
  - We will not have $\hat{p}$ before picking our sample

**Note:-**

Always round $n$ up to a whole number

## Question 7

A candidate wants to estimate his popularity. He wants to be 90% confident that the sample estimate $\hat{p}$ is within $\pm 3\%$ of the true $\hat{p}$ favoring him. What size sample is needed is

**a)** last month's poll estimated $p$ to be 60%?

***We have***

$$\hat{p} = .60$$

$$B = .03$$

$$Z = 1.6449$$

***So, using the formula***

$$n = \hat{p}(1-\hat{p})\left(\frac{Z_{\frac{a}{2}}}{B}\right)^2$$

***We have,***

$$.60(.40)\left(\frac{1.6449}{.03}\right)^2$$

$$= 721.518$$

**b)** No prior information about $p$ is available?

**Same method as above but use 0.50 for** $\hat{p}$

# Partial Review of Chapter 8

## Parameter

- Number or value that summarizes some aspect of an entire population.
- Examples:
  - $\mu$ = mean of an entire population (Quantitative data).
  - $\sigma$ = standard deviation of an entire population (Quantitative data).
  - $p$ = % of an entire population that has some specified trait (Qualitative data).

## Statistic

- Number calculated from the data in a random sample.
- Sample statistics are often used to estimate population parameters.
- Examples:
  - $\bar{x}$ = mean of a sample (Quantitative data).
  - $s$ = standard deviation of a sample (a.k.a. sample std. dev.) (Quantitative data).
  - $\hat{p}$ = % of a sample that has some specified trait (Qualitative data).

## Confidence Intervals (for estimating the unknown value of a parameter)

**To estimate $\mu$**

- Z-interval
  - Underlying population has a normal distribution .. or.. $n \geqslant 30$.
  - $\sigma$ = population standard deviation has a known value.
  - $\bar{x} \pm \left(Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$.
- t-interval
  - Underlying population has a normal distribution.
  - $\sigma$ = population standard deviation has an unknown value.
  - $s$ = sample standard deviation has a known value.
  - $\bar{x} \pm \left(t_{\alpha/2,df} \cdot \frac{s}{\sqrt{n}}\right)$, where $df = n - 1$.

**To estimate $p$**

- Z-interval
  - $n$ is large enough that both $n\hat{p} \geqslant 5$ and $n(1-\hat{p}) \geqslant 5$ are true. (Comparing each against the values 10 or 15 are common alternative conditions)
  - $\hat{p} \pm \left(Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$.

## Critical Values

- Confidence intervals are "two-sided."
- For a confidence level of $C\%$, go to the table and look for $\left(\frac{100-C}{2}\right)\%$.
- Examples:
  - $90\% \rightarrow$ look up 0.05.
  - $95\% \rightarrow$ look up 0.025.
  - $99\% \rightarrow$ look up 0.005.