2015 NOTES



# Section 8.1 Confidence Intervals: The Basics

**Key Learning – We are making a big transition!**

- In the past, we assumed we knew the true value of a population parameter and then asked questions about the distribution of the statistic used to estimate the parameter.

- NOW we no longer pretend to know the true value of the population parameter. We start with the more realistic situation where we know only the value of the statistic and use this to estimate the value of the population parameter.

## The Idea of Confidence Intervals

ACTIVITY "The mystery mean" page 468

- We want to estimate the unknown population mean **"The mystery mean"**
- We are only given the following information
  1. **"The mystery mean"** is from a normal distribution
  2. The distribution has a **population standard deviation σ=20**
  3. We can take sampling distributions. **SRS's n=16**
  4. Ms. Groves demonstration gave a sample mean ($\bar{x}$ =_______________)
- ** **Your teams job is determine a reasonable interval to estimate the population mean μ.** (Ms. Groves will hand out calculators with **"The mystery mean"** programmed into it.)

ACTIVITY

SET UP FOR "MYSTERY MEAN"

① SETUP: • DEFINE m = 650
• PROGRAM SPARE CALCS WITH M (~4 CALCS)
STORE (m) → 650 [ENTER]
→ STO> ALPHA [M]
* check [2ND] RCL [ALPHA] M

② DEMONSTRATE:
• HAND OUT CALC'S TO STUDENTS (GROUPS OF 2 OR 3)
• WALK THROUGH STEPS TO SIMULATE SRS n=16
Command "mean (rand Norm (m, 20, 16))" ROUND TO INTEGERS
① (2ND) LIST > MATH > mean
② MATH > PROB > rand Norm >
μ = ALPHA [M]
σ = 20
trials = 16

③ ACTIVITY **
TEAMS NEED TO DEVELOP AN INTERVAL TO ESTIMATE THE TRUE MYSTERY MEAN.

# AFTER the ACTIVITY "The mystery mean"



- **Questions:**

1. How can we estimate the population mean?
   - What is the "Point Estimator"? ______________________
   - What is the "Point Estimate"? ______________________

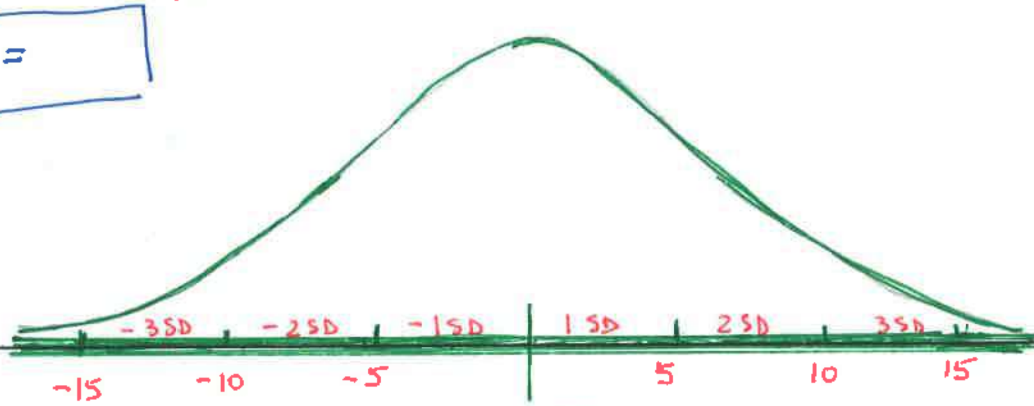
2. **Describe sampling distribution of "The mystery mean"**

- Guide students to think about (1) sampling distribution (2) The empirical rule

$\mu_{\bar{x}} = ?$   $n=16$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{16}} = 5$

$\bar{X} =$




$\mu = ?$

What is a reasonable interval that you would be confident that would include the unknown mean ($\mu$)?

MY EXAMPLE

My sampling dist. took 30 SRS's.

| -3σ | -2σ | -1σ | $\bar{x}$ | 1σ | 2σ | 3σ |
|---|---|---|---|---|---|---|
| 632 | 637 | 642 | 647 | 652 | 657 | 662 |

636, 641, 643, 645, 645, 646, 646, 647, 647, 647, 648, 648, 648, 649, 649, 649, 649, 650, 650, 650, 650, 651, 652, 652, 652, 654, 655, 656, 656, 657, 657, 660

HELP STUDENT DEFINE 95% CI

+/- 2 SD

$\bar{X} \pm 2\sigma_{\bar{x}} =$

$647 \pm 2(5)$

$647 \pm 10$ ← ME

CI: 637 - 657

$\bar{x}$ is point estimator (statistic)

647 is the point estimate (the #)

NOTICE: THIS 95% CI would miss 2 $\bar{x}$'s.
28/30 = 93.3%

# Point Estimator and Point Estimate:



**Problem:** In each of the following settings, determine the point estimator you would use and calculate the value of the point estimate.

(a) The makers of a new golf ball want to estimate the median distance the new balls will travel when hit by a mechanical driver. They select a random sample of 10 balls and measure the distance each ball travels after being hit by the mechanical driver. Here are the distances (in yards):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 285 | 286 | 284 | 285 | 282 | 284 | 287 | 290 | 288 | 285 |

(b) The golf ball manufacturer would also like to investigate the variability of the distance travelled by the golf balls by estimating the interquartile range.

(c) The math department wants to know what proportion of its students own a graphing calculator, so they take a random sample of 100 students and find that 28 own a graphing calculator.

**Definition:**

•A **point estimator** is a statistic that provides an estimate of a population parameter.

•The value of that statistic from a sample is called a **point estimate.**

•The ideal point estimate will have no bias and low variability.

**The solutions are**

**Solution:**

(a) Use the sample median as a point estimator for the true median. The sample median is 285 yards.

(b) Use the sample *IQR* as a point estimator for the true *IQR*. The sample *IQR* is $287 - 284 = 3$ yards.

(c) Use the sample proportion $\hat{p}$ as a point estimator for the true proportion $p$. The sample proportion is $\hat{p} = 0.28$.

# Finding Exact Critical Values for the 68-95-99.7 Empirical Rule

- 95% CL:


- 99% CL:


- 70% CL:


> **Summary: Confidence Interval**

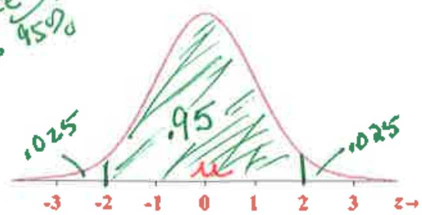
+/- 1 SD   +/- 2SD   +/- 3SD

# Finding Exact Critical Values for the 68-95-99.7 Empirical Rule

Critical Values based on N(0,1)

- 95% CL: ± 2 SD

Confidence Level is 95%

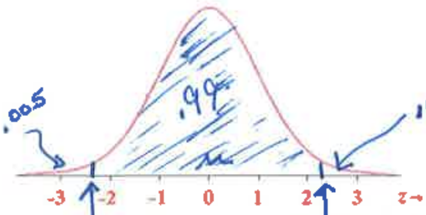



EXACT Critical Value: -C   Confidence Interval   +C

C = -1.96 — invNorm(.025,0,1)

C = 1.96 — invNorm(.975,0,1)

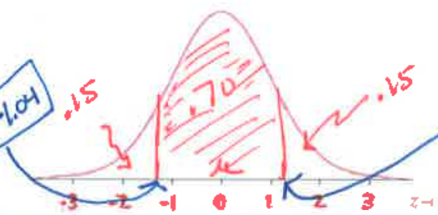
- 99% CL: ~ ± 3SD




C = -2.58 — invNorm(.005,0,1)

C = 2.58 — invnorm(.975,0,1)

- 70% CL: ~ ± 1 SD




C = -1.04

C = 1.04 — invNorm(.15,0,1)

## ➢ Summary: Confidence Interval

STATISTIC ± [CRITICAL VALUE] · [SD of statistic]

Margin of Error = [CRITICAL VALUE] · [SD of statistic]

SD of statistic = STD ERROR

**ACTIVITY** *The* **Confidence Interval** *applet*

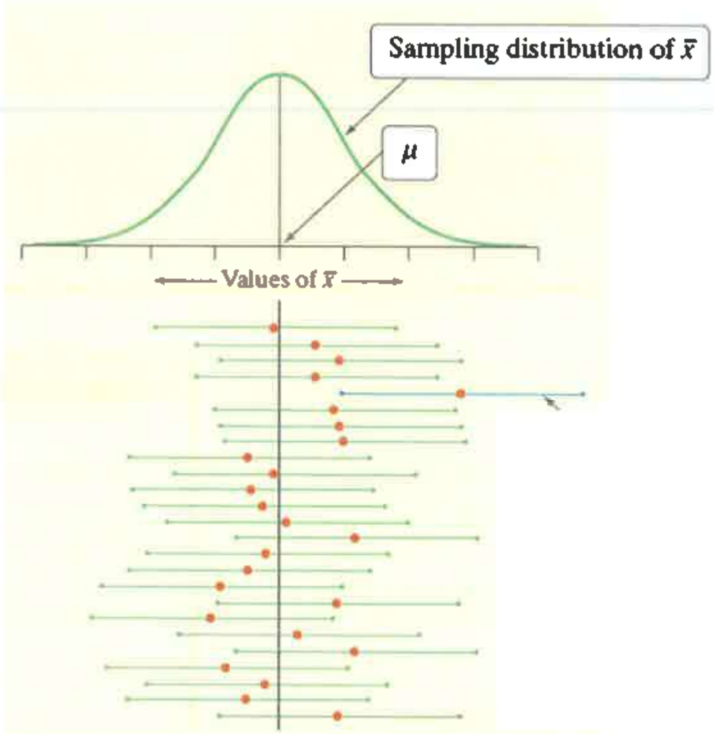
MATERIALS: Computer with Internet connection and display capability



## Interpreting Confidence Intervals:

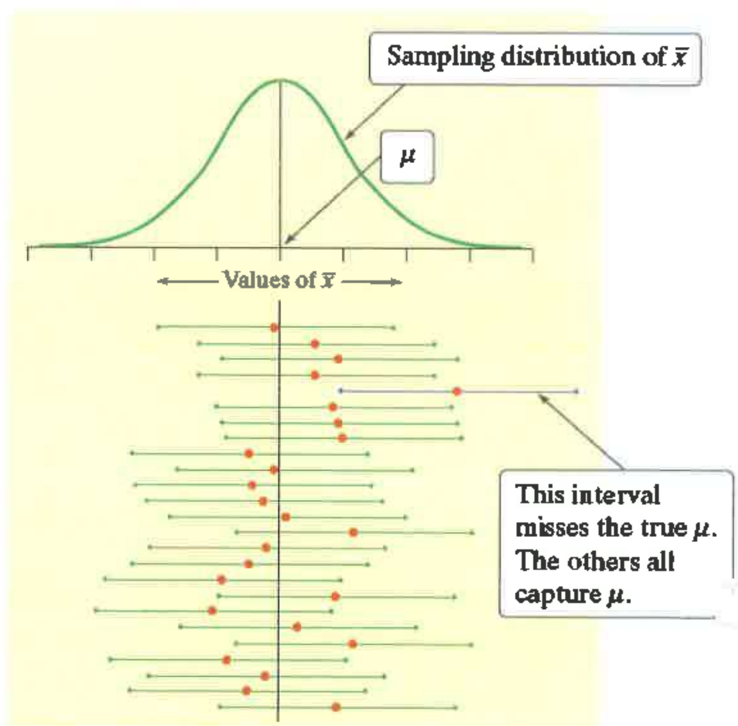
- Here is a sampling distribution for a sample mean.
- 25 samples of the same size were taken
- We choose a 95% confidence level.
- Then find a 95% confidence interval for each of the sample means.

- How many confidence intervals miss the true population mean ($\mu$)?




http://bcs.whfreeman.com/tps4e/#628644__666391__

pgroves@brunswick.k12.me.us BENKYLE1



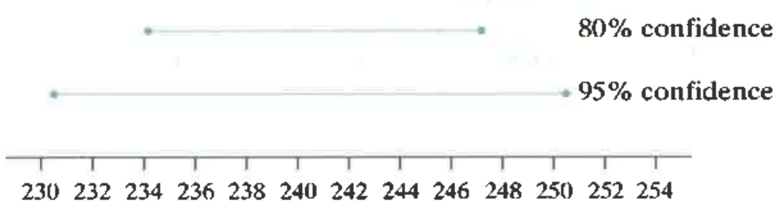
**ACTIVITY** *The* **Confidence Interval** ***applet***

MATERIALS: Computer with Internet connection and display capability



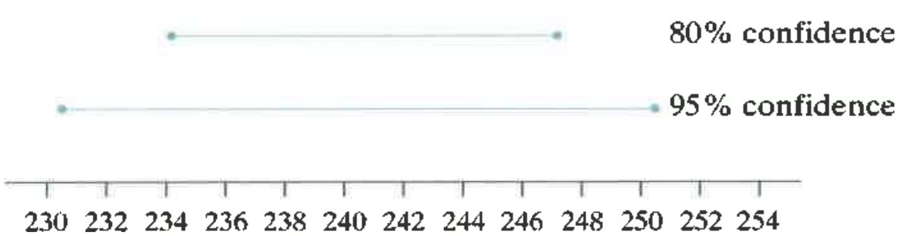
## Interpreting Confidence Levels:

Notice, the confidence interval gets larger when the confidence level increases:




Notice, the confidence interval gets larger when the confidence level increases:

- Do SRSs confidence level at 95% and what %hit the mean
- Reset. Do SRSs confidence level at 99% and what %hit the mean
- Reset. Do SRSs confidence level at 90% and what %hit the mean
- Then toggle from 90% - 95% - 99% and notice the lengths



## Interpreting Confidence Levels and intervals:

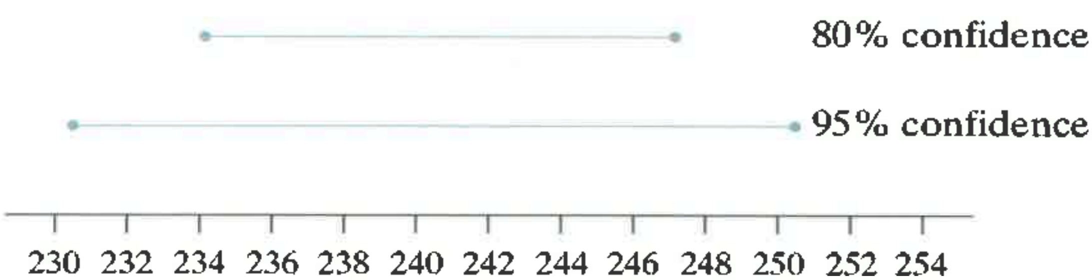


**AP Exam Common Error:**

**Use these correctly!**
**Memorize these statements!**

***Confidence level:*** To say that we are 95% *confident* is shorthand for "95% *of* all possible samples of a given size from this population will result in an interval that captures the unknown parameter."

***Confidence interval:*** To interpret a *C*% confidence interval for an unknown parameter, say, "We are *C*% confident that the interval from ______ to ______ captures the actual value of the [population parameter in context]."



# Interpreting Confidence Levels and intervals:



TPS4e: page 476

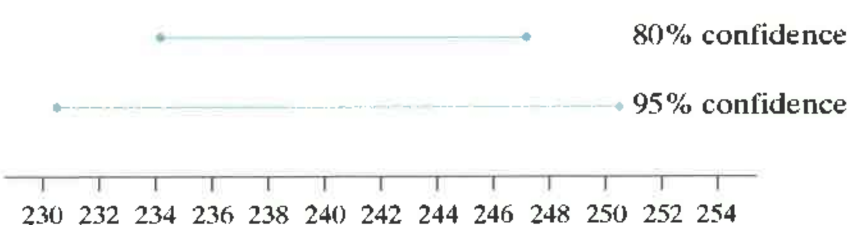
**CHECK YOUR UNDERSTANDING**

How much does the fat content of Brand X hot dogs vary? To find out, researchers measured the fat content (in grams) of a random sample of 10 Brand X hot dogs. A 95% confidence interval for the population standard deviation $\sigma$ is 2.84 to 7.55.


1. Interpret the confidence interval.
2. Interpret the confidence level.
3. True or false: The interval from 2.84 to 7.55 has a 95% chance of containing the actual population standard deviation $\sigma$. Justify your answer.

**The solutions are in the Appendix page 10**

1. CONFIDENCE INTERVAL: We are 95% confident that the interval from 2.84 to 7.55 captures the true standard deviation of the fat content of Brand X hot dogs.
2. CONFIDENCE LEVEL: In 95% of all possible samples of 10 Brand X hot dogs, the resulting confidence interval would capture the true standard deviation.
3. FALSE: The probability is either 1(if the interval contains the true standard deviation) or 0(if it does not).



# Confidence Intervals: The Basics Summary

10

- ✓ To estimate an unknown population parameter, start with a statistic that provides a reasonable guess. The chosen statistic is a **point estimator** for the parameter. The specific value of the point estimator that we use gives a **point estimate** for the parameter.
- ✓ A **confidence interval** uses sample data to estimate an unknown population parameter with an indication of how precise the estimate is and of how confident we are that the result is correct.
- ✓ Any confidence interval has two parts: an interval computed from the data and a confidence level C. The interval has the form

estimate ± margin of error •

• The margin of error tells how close the estimate tends to be to the unknown parameter in repeated random sampling.

- ✓ When calculating a confidence interval, it is common to use the form

statistic ± (critical value) · (standard deviation of statistic)

- ✓ The critical value depends on (1) the confidence level and (2) the sampling distribution of the statistic.

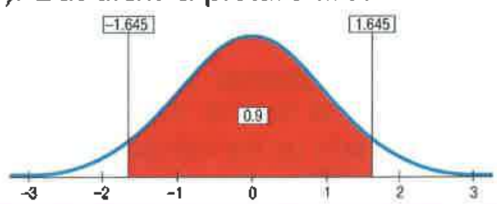
## Confidence Intervals: The Basics Summary



- ✓ The **confidence level C** is the success rate of the method that produces the interval. If you use 95% confidence intervals often, in the long run 95% of your intervals will contain the true parameter value. You don't know whether a 95% confidence interval calculated from a particular set of data actually captures the true parameter value.
  - ➢ We usually choose a confidence level of 90% or higher because we want to be quite sure of our conclusions.
  - ➢ The most common confidence level is 95%.

---

- ✓ For Normal distributions, we have used the "68-95-99.7."
  - ➢ Now we want the exact critical value and we could use Table A or the following calculator command: DISTRIB invNorm(C,0,1). But draw a picture first.
    - ➢ The critical value for the 90% CL is
    - ➢ Lower Bound: invNorm(.05,0,1)= -1.645
    - ➢ Upper Bound: invNorm(.05,0,1)= +1.645




***What are the critical values for 95%____________________ and 99% ____________________ ?***

The critical value for the 95% CL is

- Lower Bound: invNorm(.025,0,1)= -1.96
- Upper Bound: invNorm(.975,0,1)= +1.96

The critical value for the 99% CL is

- Lower Bound: invNorm(.005,0,1)= -2.57
- Upper Bound: invNorm(.995,0,1)= +2.57

# Confidence Intervals: The Basics Summary



✓ Other things being equal, the **margin of error** of a confidence interval gets smaller as the confidence level C decreases and/or the sample size *n* increases.

> ➢ **The margin of error for a confidence interval includes only chance variation, not other sources of error like nonresponse and undercoverage.**

**Before calculating a confidence interval for μ or p there are 3 important conditions to check:**

- **RANDOM**: The data should come from a well-designed random sample or randomized experiment.
- **INDEPENDENT**: Individual observations are independent. When sampling without replacement, the sample size *n* should be no more than 10% of the population size *N* (the *10% condition*) to use our formula for the standard deviation of the statistic
- **NORMAL**: The sampling distribution of the statistic is approximately Normal.
- ***Normal For Means:*** The sampling distribution is exactly Normal if the population distribution is Normal. When the population distribution is not Normal, then the central limit theorem tells us the sampling distribution will be approximately Normal if *n* is sufficiently large ($n \geq 30$).
- ***Normal For Proportions:*** We can use the Normal approximation to the sampling distribution as long as $np \geq 10$ and $n(1 - p) \geq 10$.

# SOLUTIONS

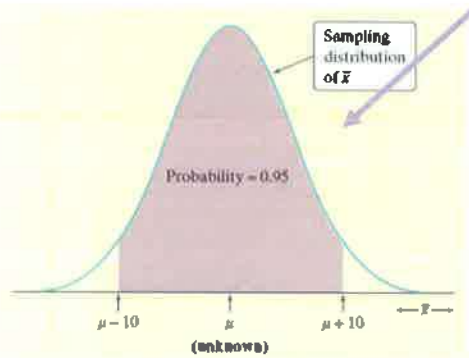
## ■ INTRODUCTION: "The Idea of a Confidence Interval"



- **We want to estimate a "mystery mean $\mu$," *was from a population with a Normal distribution and $\sigma$=20.***
- ***We take an SRS of n=16 and calculated* the sample mean= 240.79.**

First, to estimate the Mystery Mean $\mu$, we can use $\bar{x} = 240.79$ as a point estimate.
We don't expect $\mu$ to be exactly $= \bar{x}$ so we need to say how accurate we think our estimate is.

- The Mystery Mean followed a normal distribution N(240.79 , 5 ).




- Remember the "68 - 95 - 99.7 Rule"

- It tells us that in 95% of all samples, $\bar{x}$ will be within 10 (2 SD) of $\mu$.
- Therefore, the interval from $\bar{x} - 10$ to $\bar{x} + 10$ will "capture" $\mu$ in about 95% of all samples.

20/√16

CONCLUSION in CONTEXT: **If we estimate that $\mu$ lies somewhere in the interval 230.79 to 250.79, we'd be calculating an interval using a method that captures the true $\mu$ in about 95% of all possible samples of this size.**

SOLUTION

# Point Estimator and Point Estimate:



**Problem:** In each of the following settings, determine the point estimator you would use and calculate the value of the point estimate.

(a) The makers of a new golf ball want to estimate the median distance the new balls will travel when hit by a mechanical driver. They select a random sample of 10 balls and measure the distance each ball travels after being hit by the mechanical driver. Here are the distances (in yards):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 285 | 286 | 284 | 285 | 282 | 284 | 287 | 290 | 288 | 285 |

(b) The golf ball manufacturer would also like to investigate the variability of the distance travelled by the golf balls by estimating the interquartile range.

(c) The math department wants to know what proportion of its students own a graphing calculator, so they take a random sample of 100 students and find that 28 own a graphing calculator.

**Solution:**

(a) Use the sample median as a point estimator for the true median. The sample median is 285 yards.

(b) Use the sample *IQR* as a point estimator for the true *IQR*. The sample *IQR* is $287 - 284 = 3$ yards.

(c) Use the sample proportion $\hat{p}$ as a point estimator for the true proportion $p$. The sample proportion is $\hat{p} = 0.28$.

**Definition:**

•A **point estimator** is a statistic that provides an estimate of a population parameter.

•The value of that statistic from a sample is called a **point estimate**.

•The ideal point estimate will have no bias and low variability.

## HOT DOG SOLUTIONS

CHECK YOUR UNDERSTANDING (PAGE 476)

1. CONFIDENCE INTERVAL: We are 95% confident that the interval from 2.84 to 7.55 captures the true standard deviation of the fat content of Brand X hot dogs.

2. CONFIDENCE LEVEL: In 95% of all possible samples of 10 Brand X hot dogs, the resulting confidence interval would capture the true standard deviation.

3. FALSE: The probability is either 1(if the interval contains the true standard deviation) or 0(if it does not).