

# Applied Statistical Analysis

EDUC 6050

Week 3

Finding clarity using data

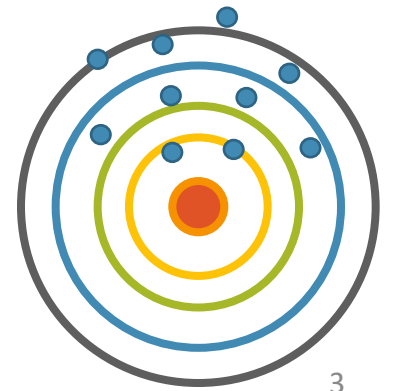
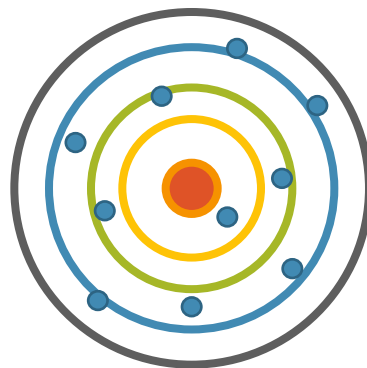
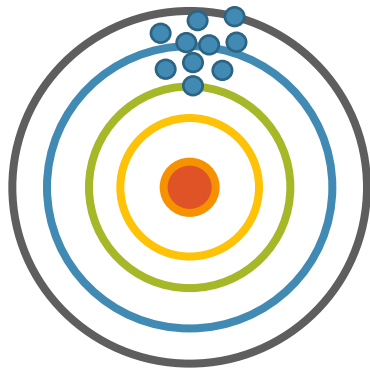
# Today

1. Review Statistical Terminology
2. Central Tendency
3. Variability
4. Confidence Interval

# Reliability and Validity

**Reliability**: the consistency of the measure

**Validity**: does it measure what we think it measures?



# Reliability and Validity

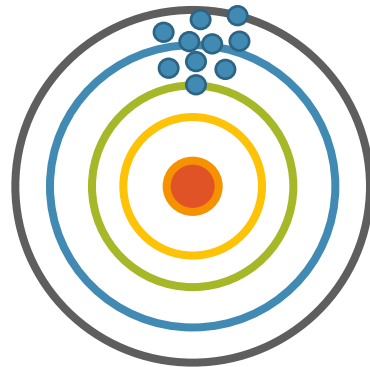
**Reliability**: the consistency of the measure

**Validity**: does it measure what we think it measures?

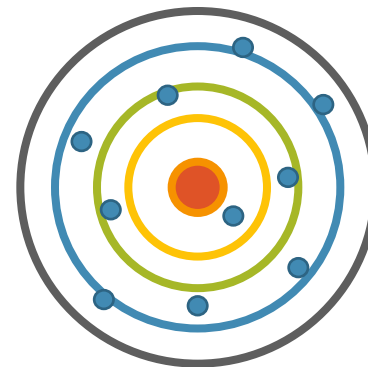
Reliable  
Valid



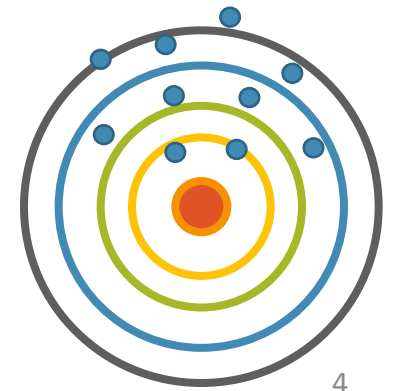
Reliable  
Not Valid



Not Reliable  
Valid



Not Reliable  
Not Valid



# Reliability and Validity

## Reliability

Compare with factor analyses (not covered in the class)

## Validity

Compare with correlations with things that should correlate or shouldn't

Often based on theory

# Correlation and Experimentation

## Correlation

observational, no  
treatment/intervention

## Experimentation

treatment/intervention (best  
if groups are randomized)

What are the pro's and  
con's of each?

# Correlation and Experimentation

Depends on the field how often each are used

Possible, but difficult, to convince of causation with correlational (observational) data

**Correlation does not imply causation**

**AND**

**Correlation does not imply it isn't causal**

# Central Tendency

What does this mean?

Mean

“arithmetic average”  
Sum of scores  
divided by number of  
scores

Median

“the middle score”  
The number where  
half of the scores  
are above and half  
are below

Mode

“most common score”  
The most common  
score



# Central Tendency

Measure	When to use it
Mean	With interval/ratio data that are ~normally distributed
Median	With ordinal data With interval/ratio data that are skewed or have outliers
Mode	With nominal data

**Outliers** = points far from the other points

# Computing the Mean

Sum of scores  
Number of scores

$$M = \frac{\sum X}{N}$$

# Computing the Median

1. Order the values from lowest to highest
2. Find the middle value
3. If two are in the middle, take the average of those two

# Computing the Mode

Find the value that is  
the most common

Mean, Median, and Mode for Age?

Mean, Median, and Mode for Degree?

Age	Degree
21	MS
25	MEd
34	PhD
21	PhD
22	MEd
28	MS
33	MS
29	MS

## Mean, Median, and Mode for Age?

$$\text{Mean} = 213/8 = 26.6$$

$$\text{Median} = 21 \ 21 \ 22 \ \boxed{25 \ 28} \ 29 \ 33 \ 34 = 26.5$$

$$\text{Mode} = 21$$

## Mean, Median, and Mode for Degree?

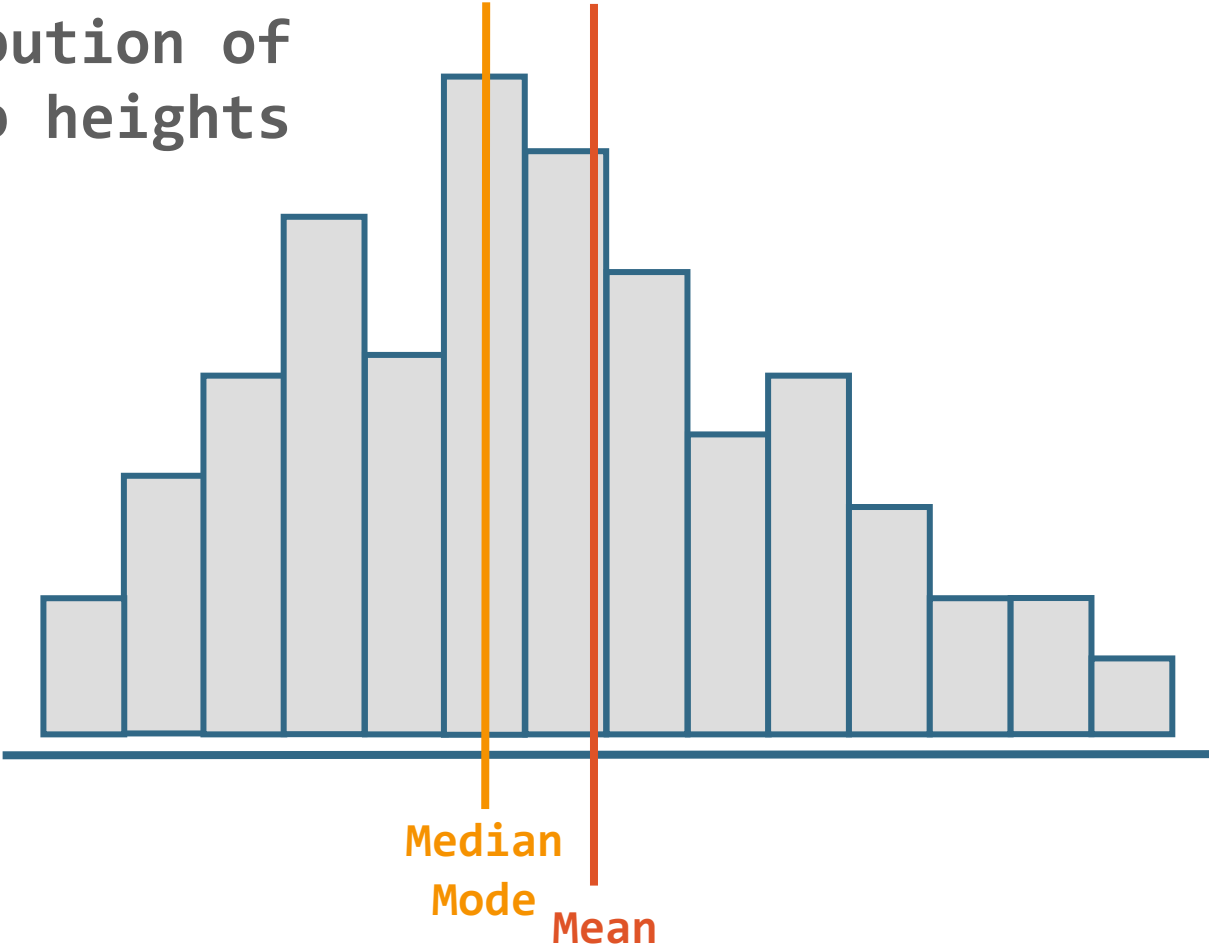
$$\text{Mean} = \dots$$

$$\text{Median} = \dots$$

$$\text{Mode} = \text{MS}$$

Age	Degree
21	MS
25	MEd
34	PhD
21	PhD
22	MEd
28	MS
33	MS
29	MS

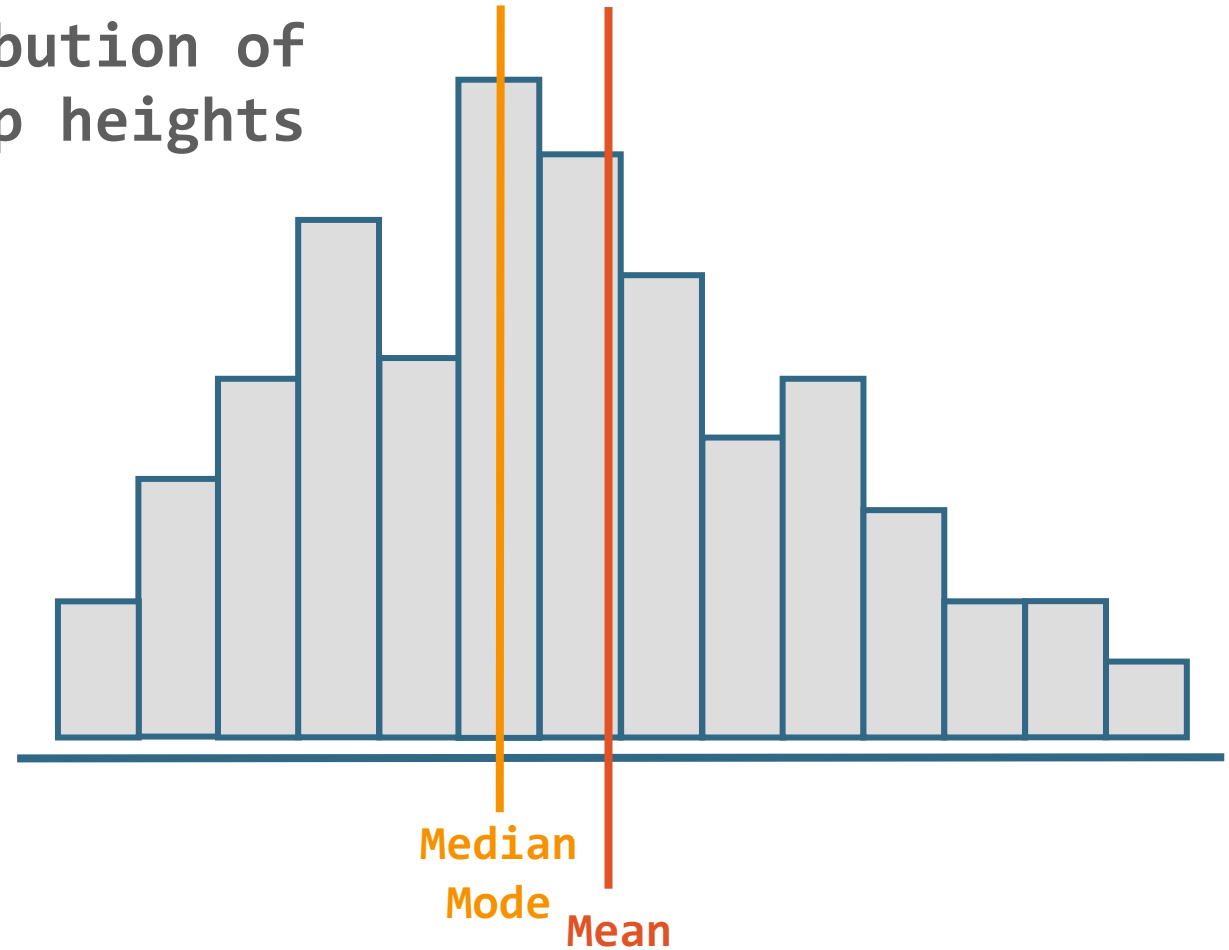
# The distribution of jump heights



What about the spread of the data?

The distribution of jump heights

This is what **variability** is all about



What about the spread of the data?



# Variability

## “spread”

Measure	What is It
Range	Max - Min
Standard Deviation	The typical (or standard) distance each score is from the mean

# Variability

Measure	When to Use	Possible Values
Range	Ordinal, Interval, Ratio	$\theta+$
Standard Deviation	Interval, Ratio	$\theta+$

# Computing **Range**

Two approaches:

1. **Max - Min**

2. **“[Min] to [Max]”**

# Computing Standard Deviation

Essentially it is the average deviation from the mean  $(X - M)$

$$SD = \sqrt{\frac{\sum (X - M)^2}{N - 1}}$$

# Other Stuff about **Standard Deviation**

There is a population standard deviation denoted  $\sigma$  but is usually unknown

**Our SD is an estimate of the population standard deviation**

# Variability

What is the range of Age?

What is the range of Grade?

Age	Grade
21	A
25	B
34	A
21	B
22	C
28	B
33	B
29	A

# Variability

What is the range of Age?

$$\begin{aligned}\text{Range} &= 34 - 21 = 13 \\ &= 21 \text{ to } 34\end{aligned}$$

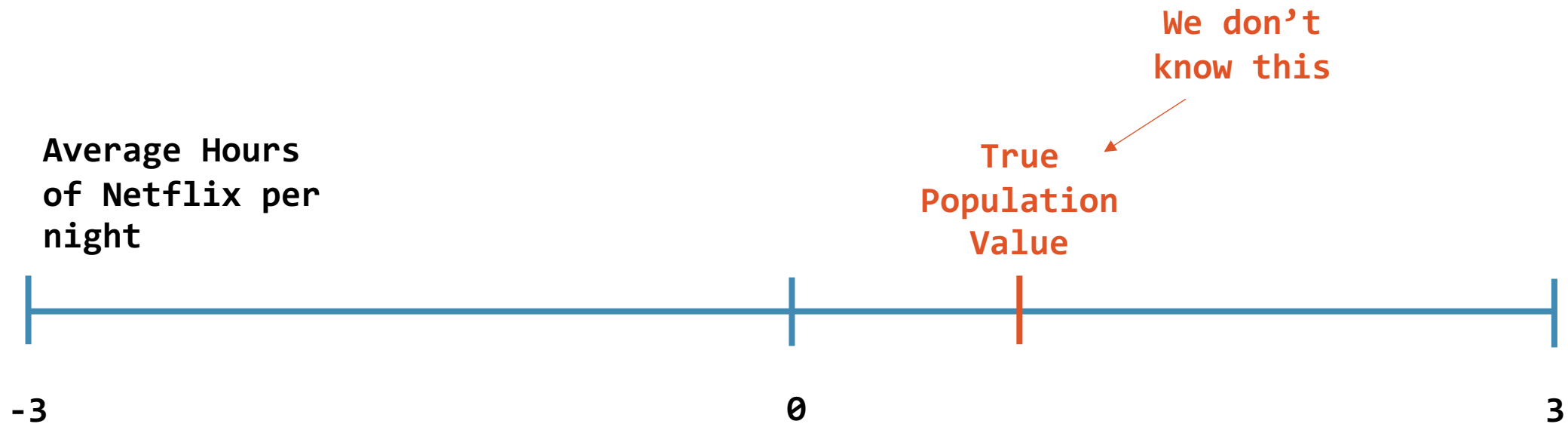
What is the range of Grade?

$$\text{Range} = \text{A to C}$$

Age	Grade
21	A
25	B
34	A
21	B
22	C
28	B
33	B
29	A

# Confidence Interval

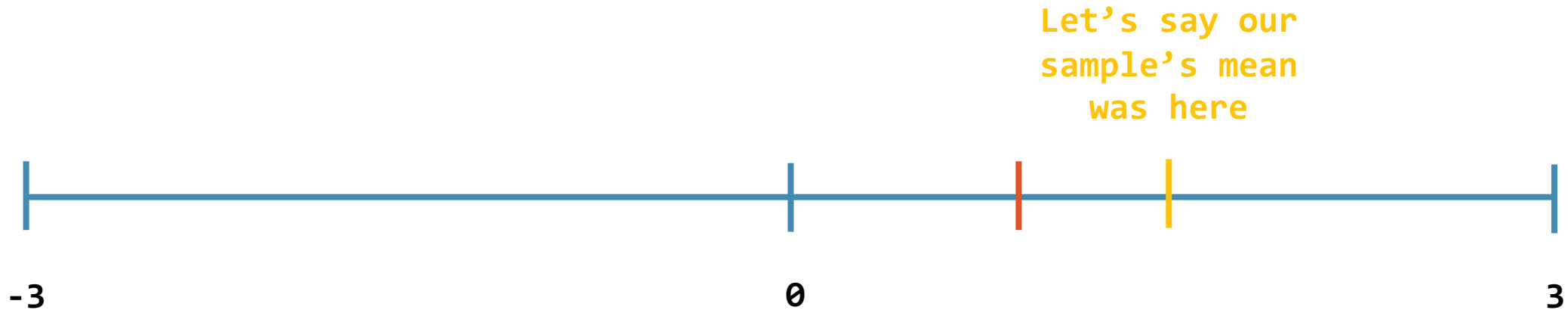
An interval that helps us understand the uncertainty in an estimate





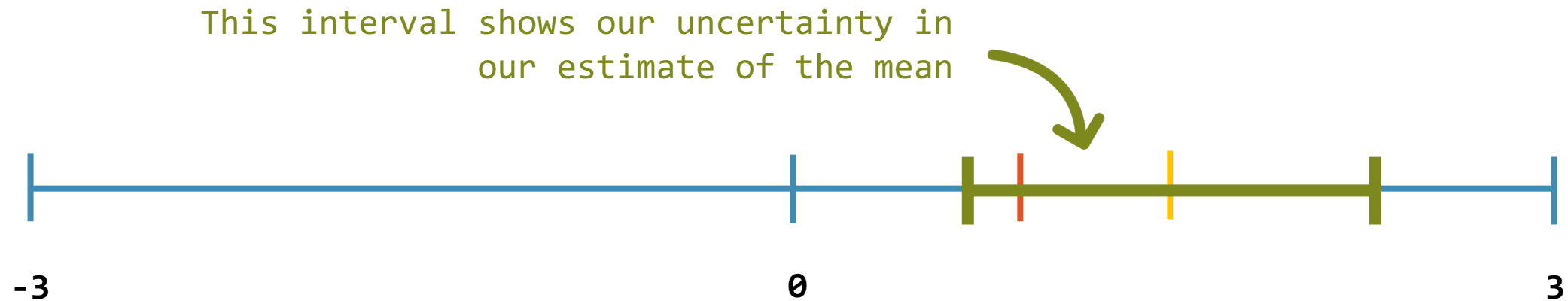
# Confidence Interval

An interval that helps us understand the uncertainty in an estimate



# Confidence Interval

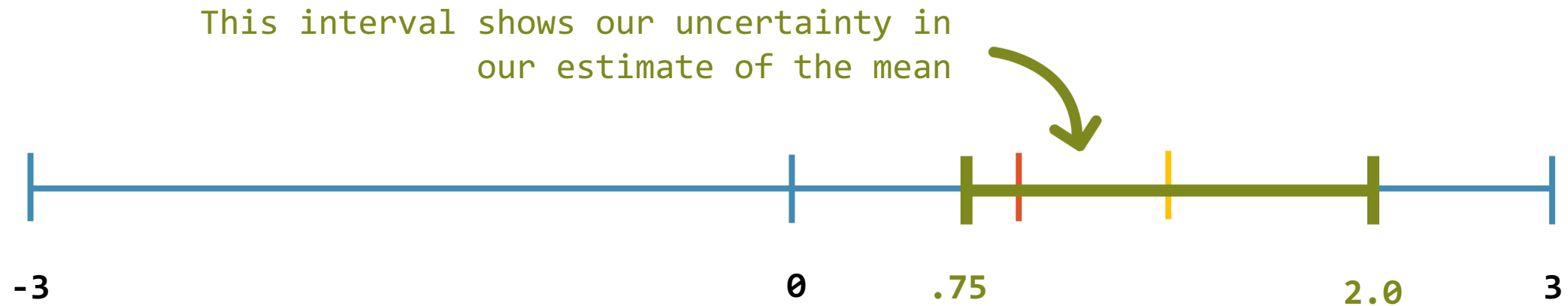
An interval that helps us understand the uncertainty in an estimate



95% Confidence Interval will contain the true population parameter 95% of the time

# Confidence Interval

An interval that helps us understand the uncertainty in an estimate

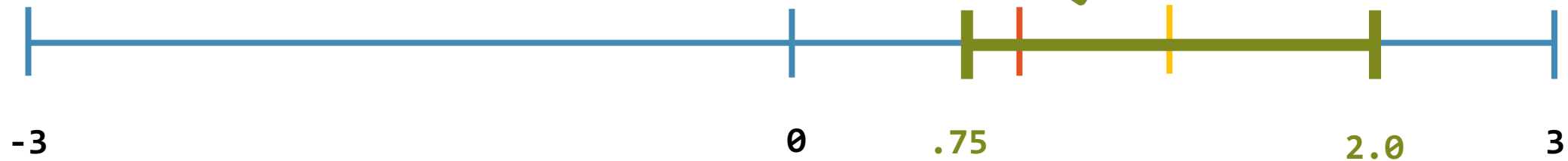


“We are 95% confident that the true population mean is between .75 and 2.0.”

# Confidence Interval

An interval that helps us understand the uncertainty in an estimate

This interval shows our uncertainty in our estimate of the mean



$$\text{Interval} = \text{Estimate} \pm \text{Uncertainty}$$

Depends on the SD and the confidence level

Bigger SD,  
bigger interval



# Confidence Interval

An interval that helps us understand the uncertainty in an estimate

We will use this throughout the semester, not just with means but with other estimates too

Basically always have the same interpretation

# Questions?

Please post them to the  
discussion board before  
class starts

End of Pre-Recorded Lecture Slides

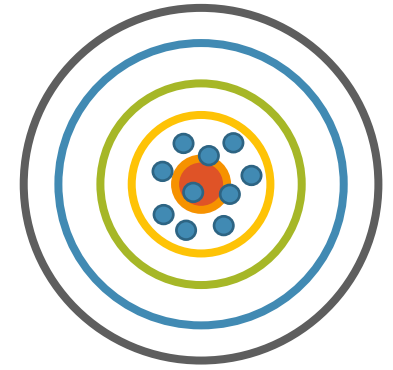
# In-class discussion slides



# Quick, Quiet, Qualifying Quiz



# Review



1. The figure to the right is reliable/unreliable and valid/invalid.
2. When should you use the mean? What about the median?
3. What does the standard deviation tell us?
4. Can we obtain a standard deviation with nominal data?

# Reading

5. What is the difference between a sample and a population?
6. What are descriptive statistics?
7. True or False. Inferential statistics help us use our sample to understand the population.
8. True or False. Independent Variables are also known as outcomes.
9. Hypothesis tests inform us about the \_\_\_\_\_ of our findings.

# Reading

10. True or False. Hypothesis testing informs us about the population.
11. Is a nominal variable more qualitative or quantitative?
12. What information does a bar chart provide? What about a histogram?
13. What is an estimate?
14. What does a confidence interval tell us?

# Application

Example Using the Class Data &  
The Office/Parks and Rec Data Set

Visualizations and Descriptives  
in Jamovi