

Applied Statistical Analysis

EDUC 6050

Week 3

Finding clarity using data

Today

1. Review Statistical Terminology
2. Central Tendency
3. Variability

Reading

**Questions from
Chapter 1**

Reading

Quick, Quiet,
Qualifying Quiz

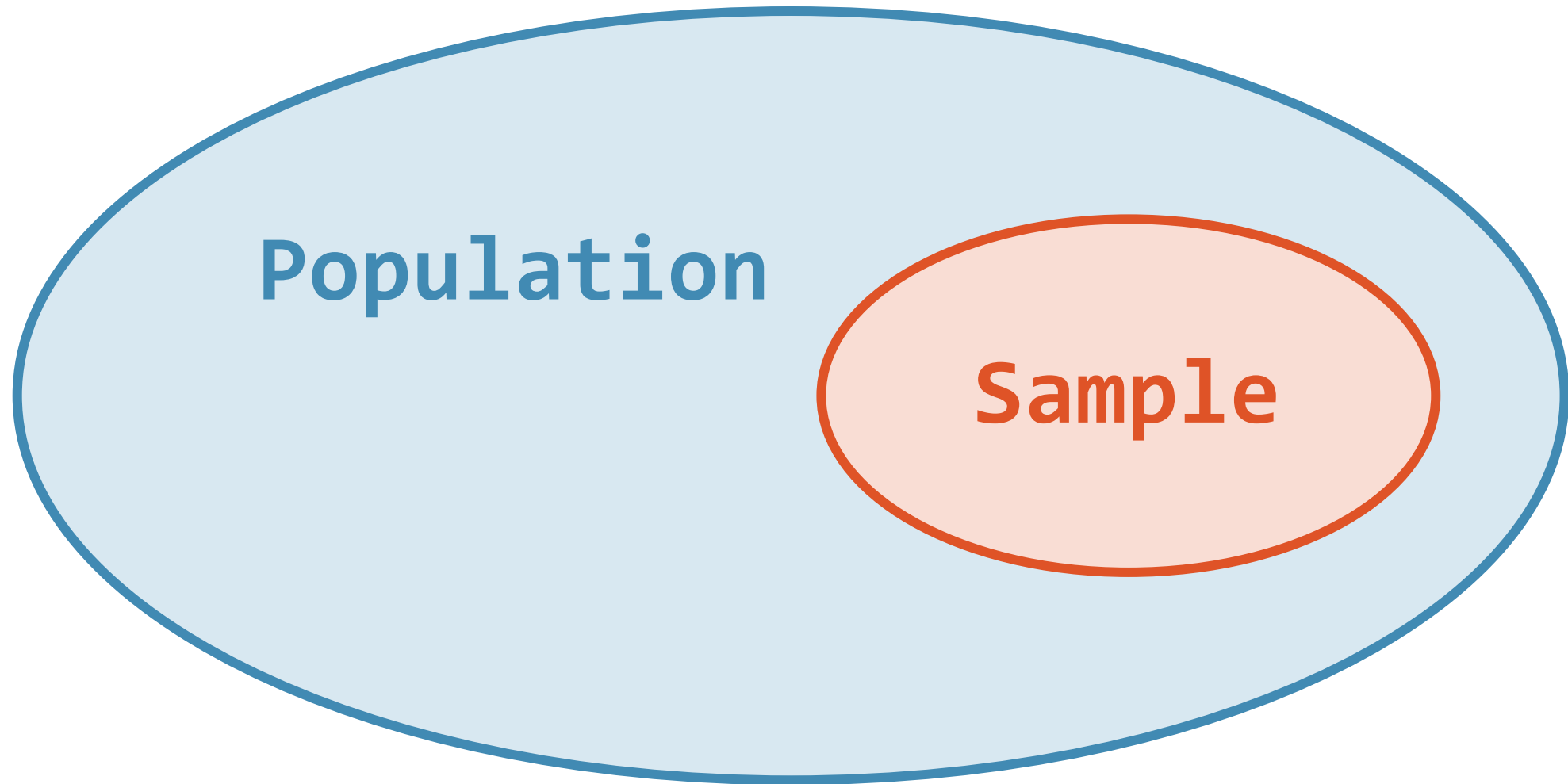
Reading

1. What is the difference between a sample and a population?
2. What are descriptive statistics?
3. True or False. Inferential statistics help us use our sample to understand the population.
4. True or False. Independent Variables are also known as outcomes.
5. Hypothesis tests inform us about the _____ of our findings.

Reading

6. True or False. Hypothesis testing informs us about the population.
7. What is the difference between qualitative and quantitative?
8. Is a nominal variable more qualitative or quantitative?
9. What information does a bar chart provide? What about a histogram?
10. How satisfied are you with your answers?

Review: Vocabulary of Statistics



Review: Vocabulary of Statistics

Descriptive Statistics

Describing the data that
you have (your sample)

Inferential Statistics

Understanding what your data
say about the population

Review: Vocabulary of Statistics

**Independent
Variables**



**Dependent
Variables**

“predictors” or “IV”

These are the variables
that we think are
causing or influencing
the outcome

“outcomes” or “DV”

These are the variables
that we think are caused
by an independent
variable

Review: Vocabulary of Statistics

Hypothesis Testing (Inferential Statistics)

“Null Hypothesis Significance Testing”

Gives us an idea about what the population may look like based on our sample (accounts for **sampling error**) = “significance”

Review: Vocabulary of Statistics

Hypothesis Testing (Inferential Statistics)

“Null Hypothesis Significance Testing”

Effect Sizes

“Magnitude of the effect”

Tells us how big the effect is = “meaningfulness”

Review: Scales of Measurement

4 General Types (see pg. 11)


<i>Scale</i>	<i>Definition</i>	<i>What the scale allows you to do</i>
Nominal	Categories based on qualitative similarity (no order to the categories)	Count the number of things in the categories
Ordinal	Like nominal, but the categories can be ranked	Count and rank the number of things in each category
Interval	Quantify how much of something	Count, rank, and quantify how much of something there is (zero does not mean there's nothing)
Ratio	Quantify how much of something (zero means there is none of that thing)	Count, rank, and quantify how much of something there is with a meaningful zero

Review: Scales of Measurement

4 General Types (see pg. 11)

<i>Scale</i>	<i>Definition</i>	<i>What the scale allows you to do</i>
Nominal	Categories based on qualitative similarity (no order to the categories)	Count the number of things in the categories
Ordinal	Like nominal, but categories can be ranked	Count and rank the number of things in each category
Interval	Quantify how much of something there is (zero does not mean there's nothing)	Count, rank, and quantify how much of something there is (zero does not mean there's nothing)
Ratio	Quantify how much of something there is with a meaningful zero	Count, rank, and quantify how much of something there is with a meaningful zero

Increasing degree of information

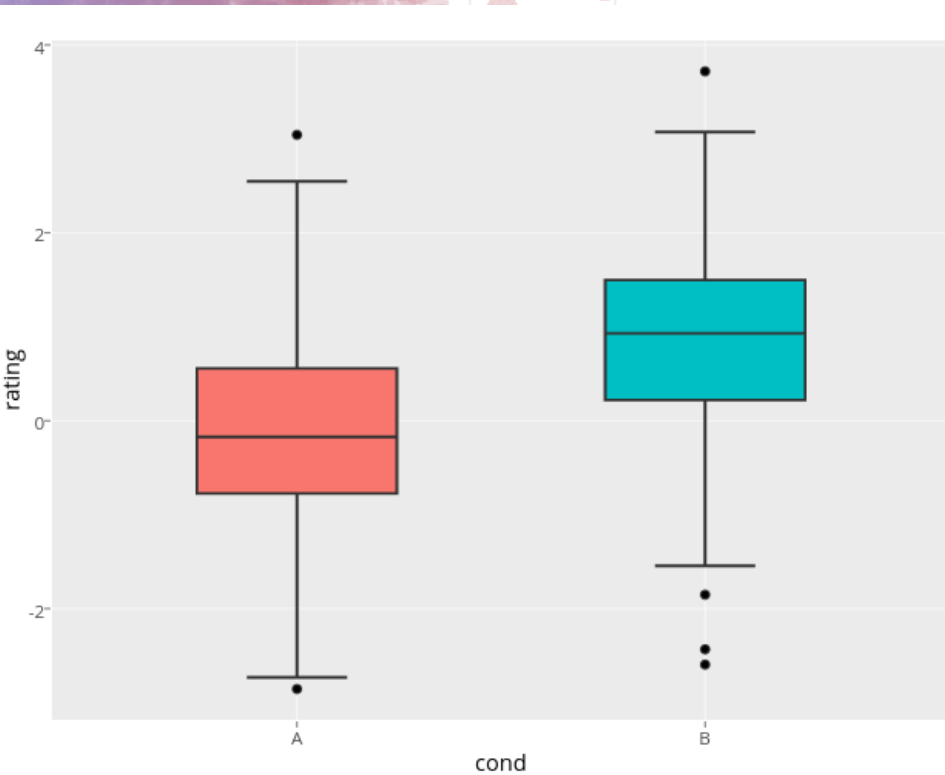
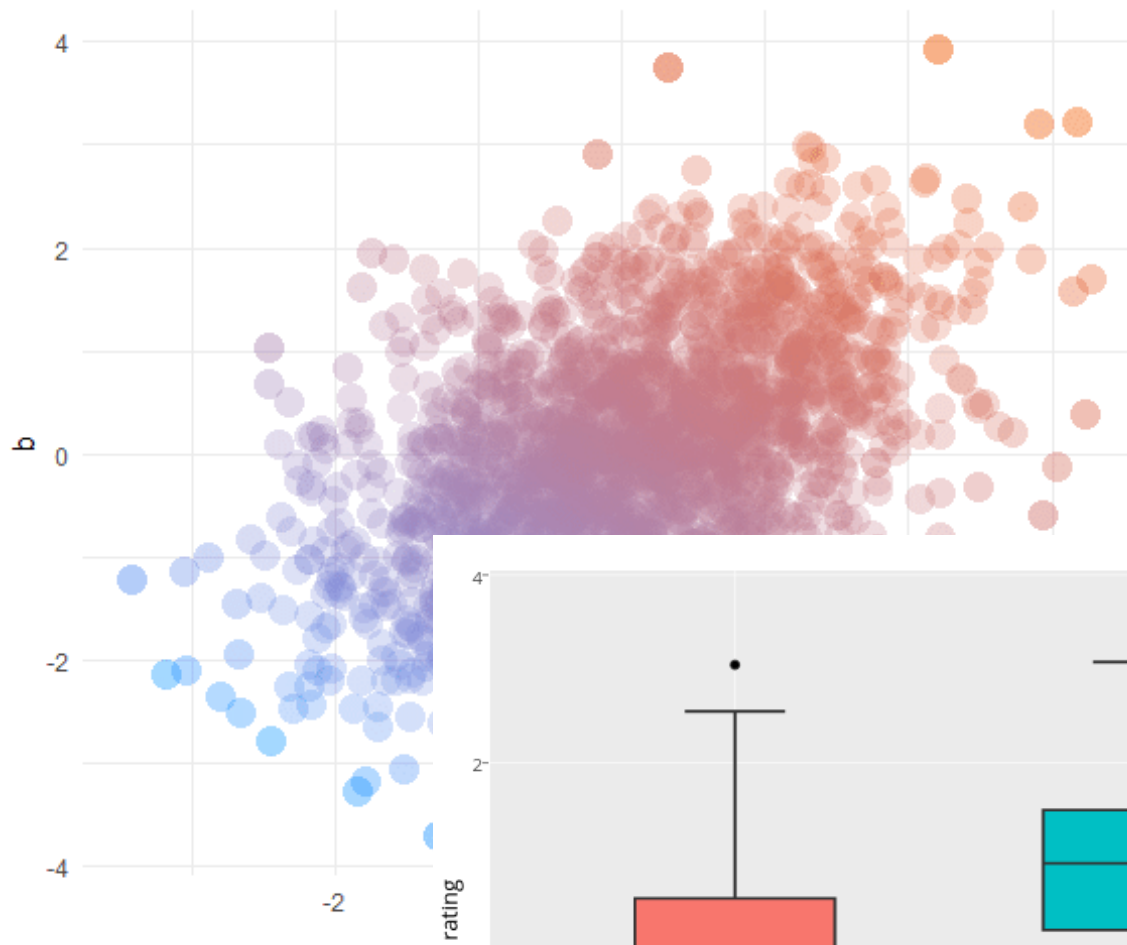
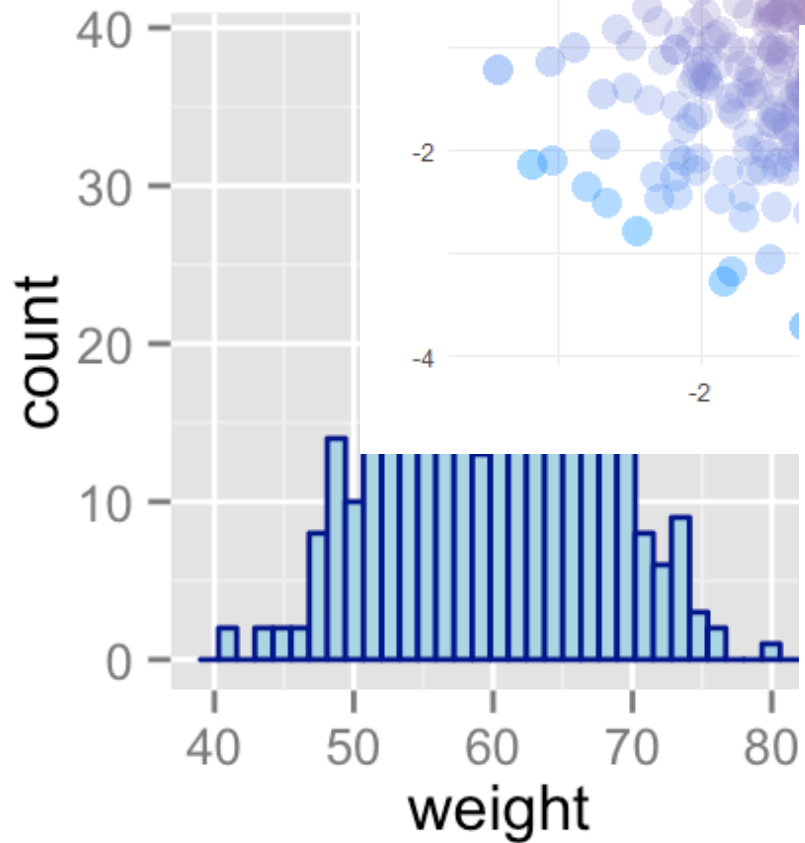
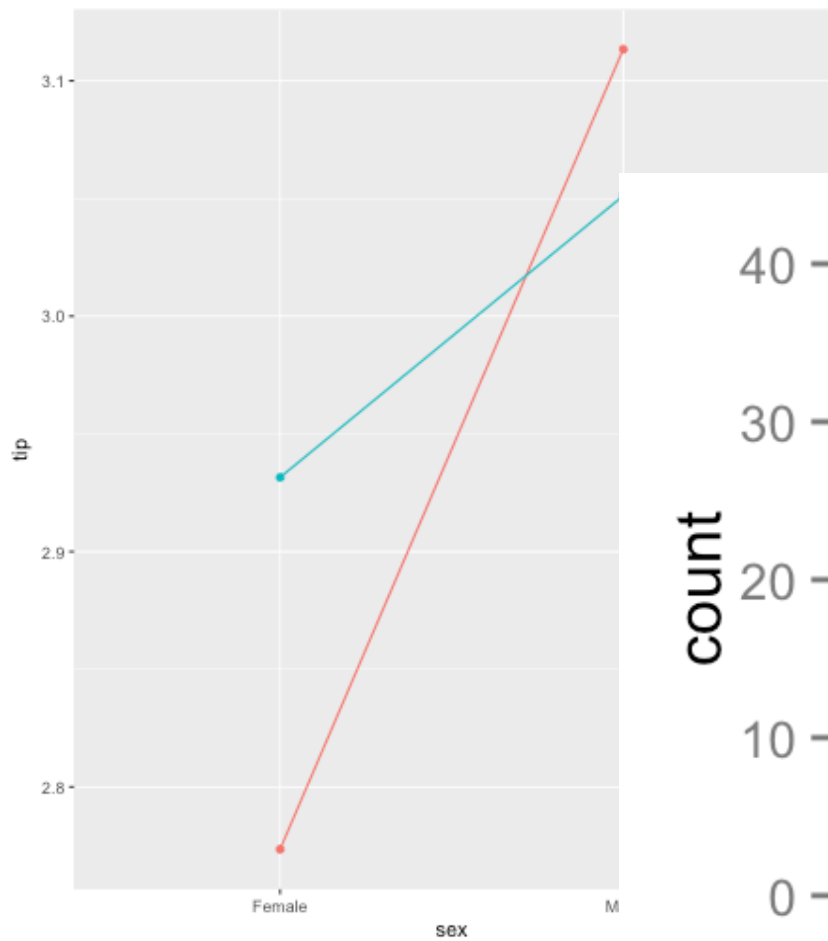


Review: Scales of Measurement

These lie on a spectrum from qualitative to quantitative



Review: Plots

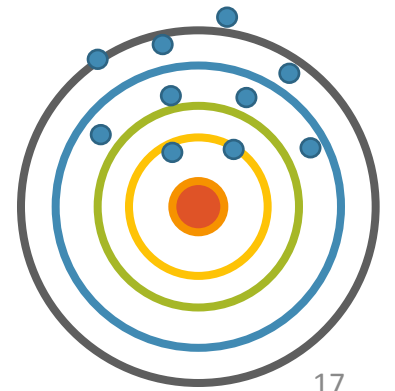
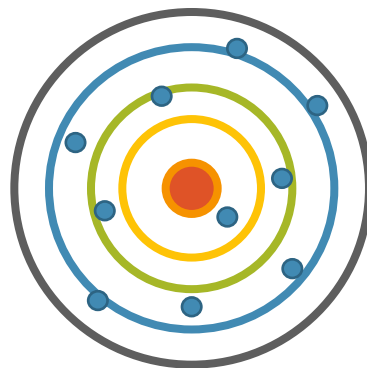
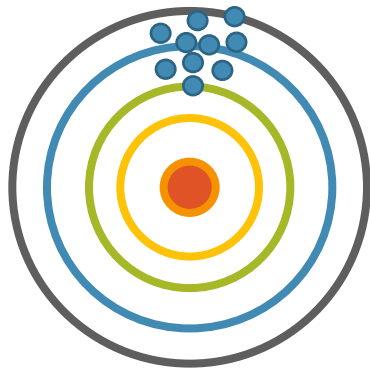


Break Time

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?

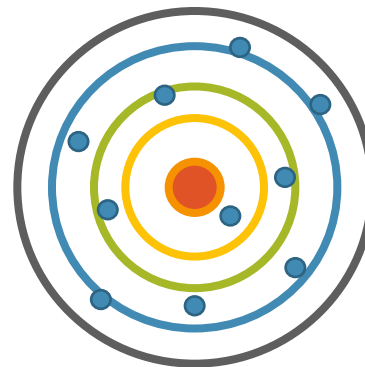
Reliable
Valid



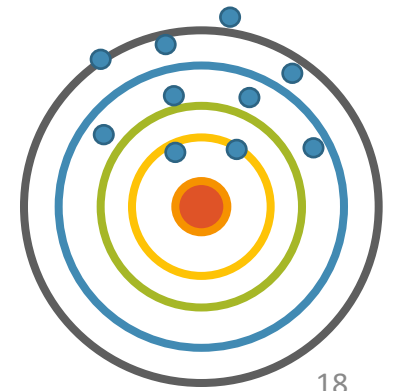
Reliable
Not Valid



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

Central Tendency

See Figure 2.2 (page 42)

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?

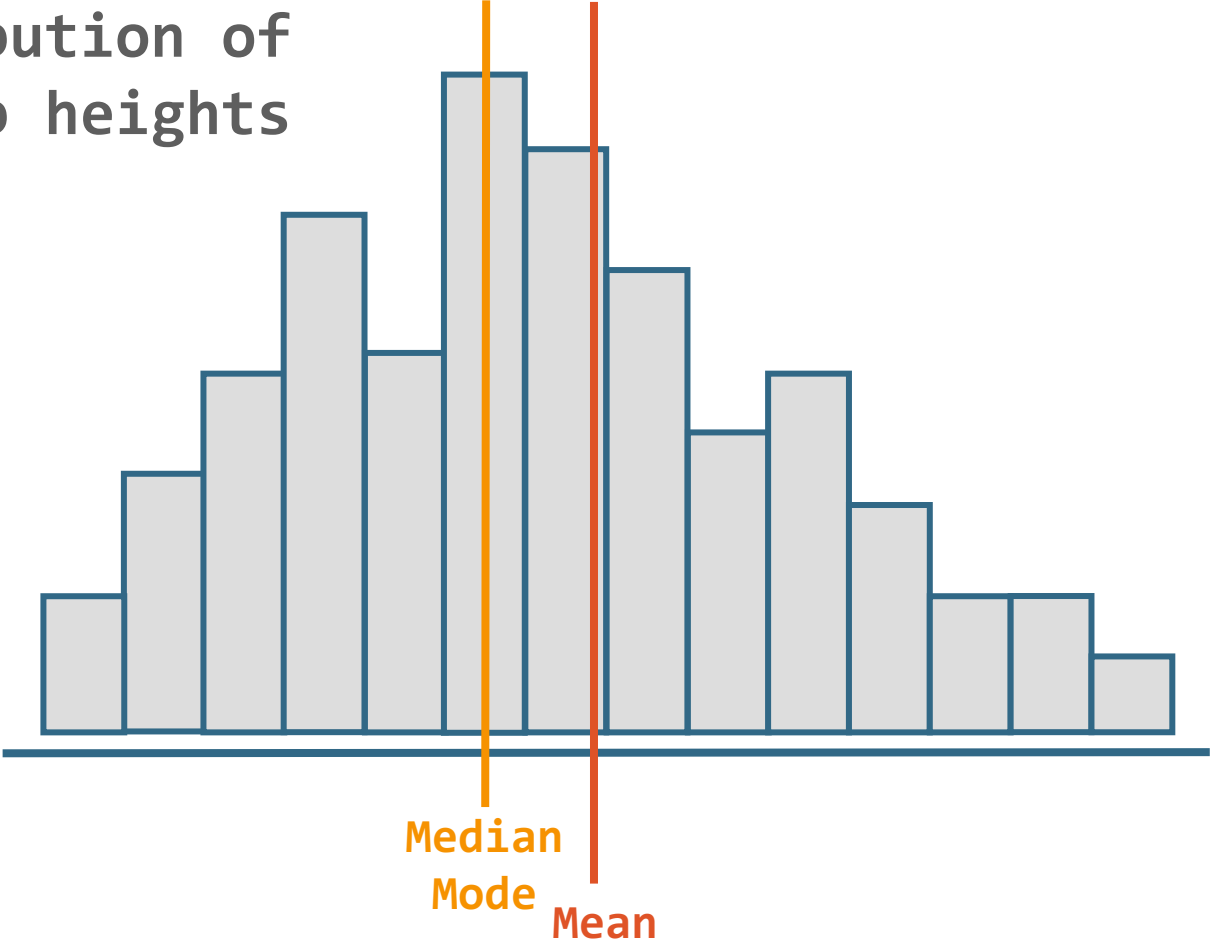
$$\text{Mean} = \dots$$

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

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

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

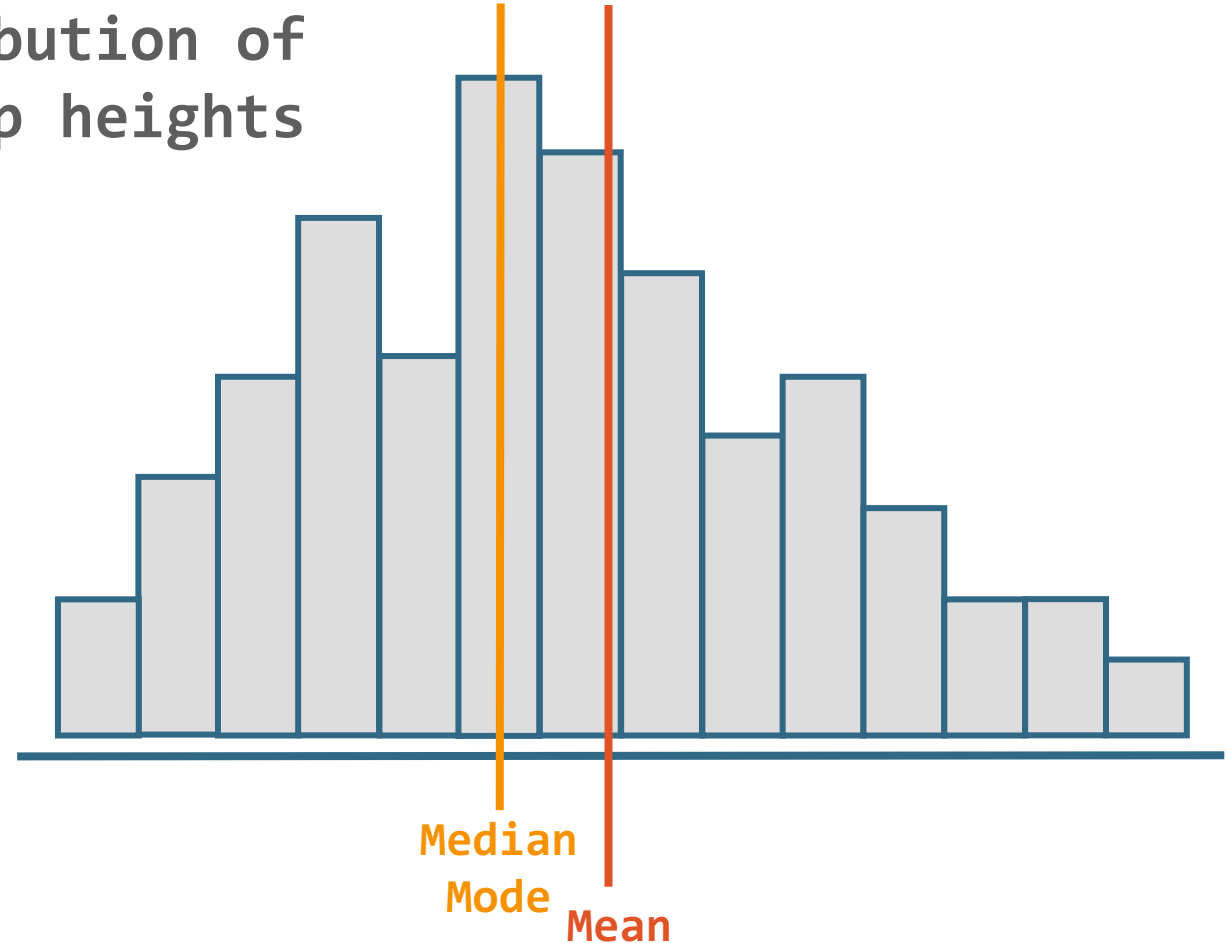
The distribution of jump heights



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 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 σ 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?

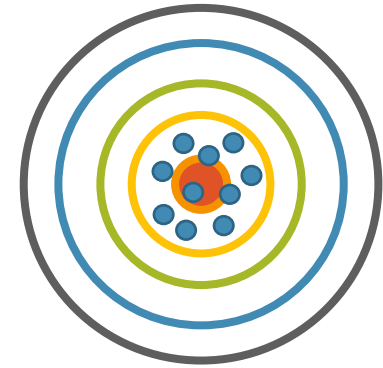
$$\begin{aligned}\text{Range} &= 34 - 21 = 13 \\ &= 21 \text{ to } 34\end{aligned}$$

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

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?

Another look at Jamovi

Plots
Central Tendency
Variability

Questions?

Next week:

1. Statistics terminology continued
(hypothesis testing, descriptive and inferential statistics, effect sizes, confidence intervals, Type I and II errors)
2. Chapters 4, 5, and 6 in Book
3. Statistical Organizer due