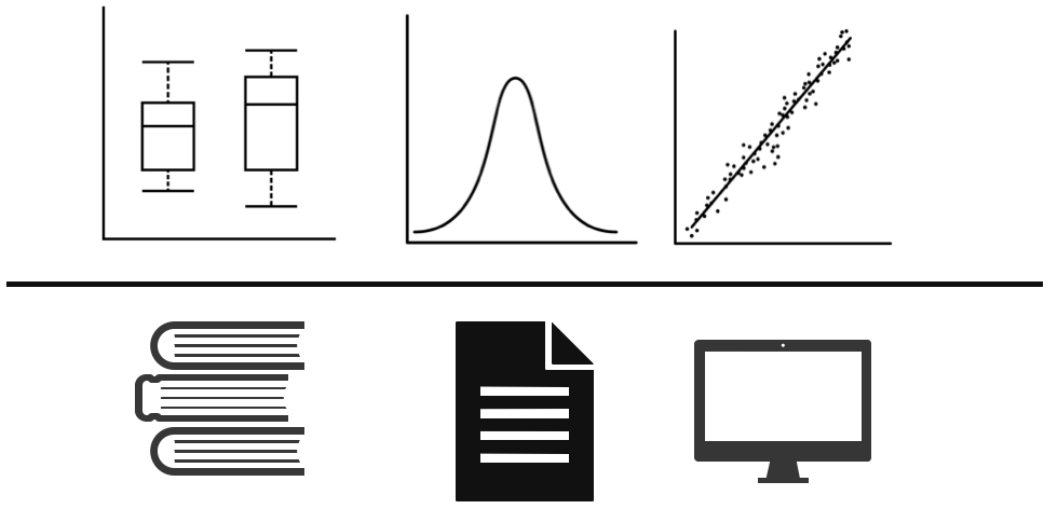
PSYC 2300

Introduction to Statistics



Lecture 04: Correlations, Reliability, & Validity

Outline for today

- **Defining relationships in the world**
  - Correlation coefficient
- **Reliability**
  - Test-retest reliablity
  - Parallel forms reliability
  - Internal consistency
  - Interrater reliability
- **Validity**
  - Content validity
  - Criterion validity
  - Construct validity

Defining Relationships in the World

Correlations

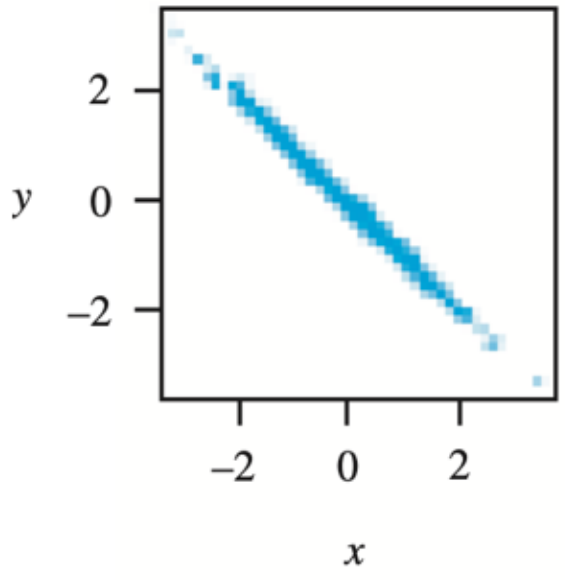
In psychological science, **correlational designs** examine the extent to which two variables are associated

Examples

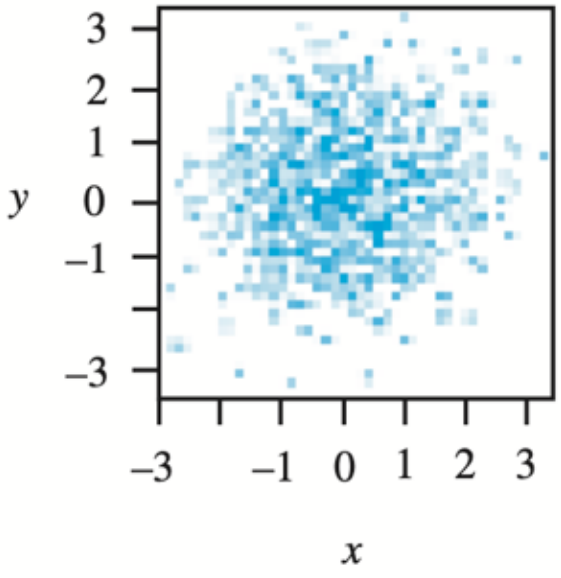
- Do people who are more creative do better in school?
- Is there a relationship between hours studied and exam scores? Class attendance and final grades?
- Are people who have higher marital satisfaction better parents?

Correlations

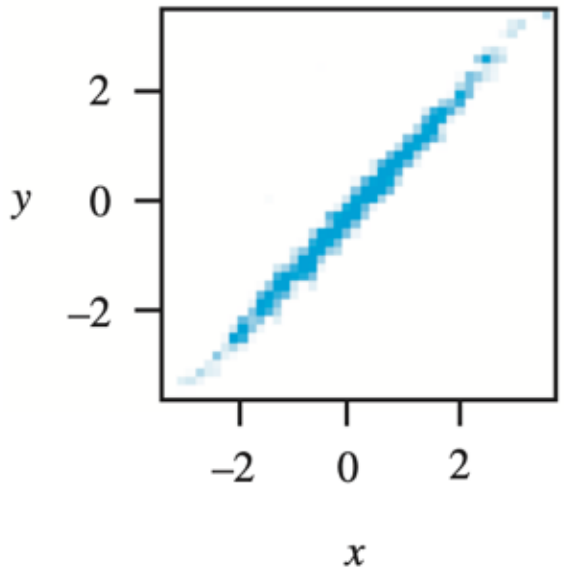
Correlations are either *positive*, *negative*, or *zero*



Negative



Zero



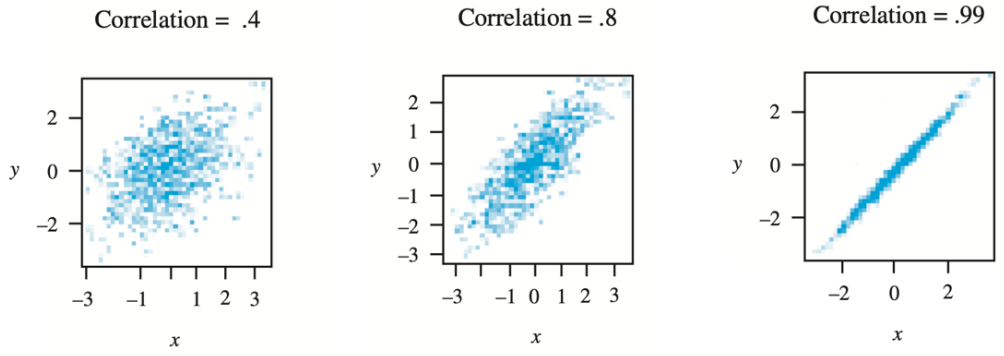
Positive

Correlations

Positive correlation: as one variable changes, the other variable changes in the *same direction*

Examples

- Relationship between hours studied and exam grade
  - More hours studied, better exam grade
  - Fewer hours studied, worse exam grade



Examples of positive correlations?

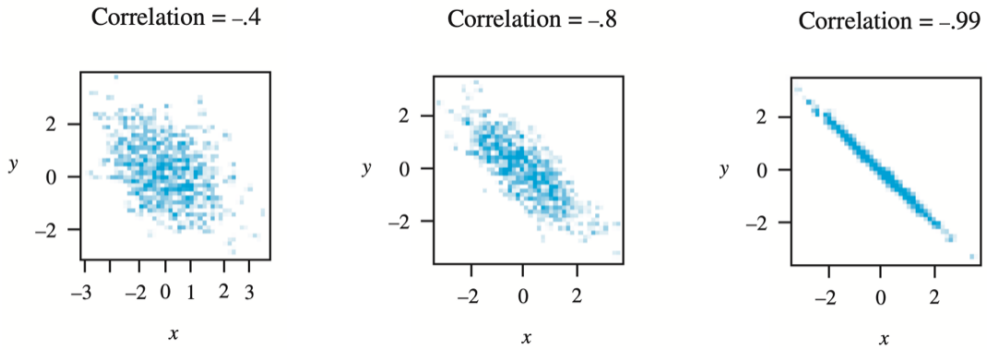
-

Correlations

Negative correlation: as one variable changes, the other variable changes in the *opposite direction*

Examples

- Relationship between running speed and time to finish a race
  - Faster running = Less time to finish
  - Slower running = More time to finish



Examples of negative correlations?

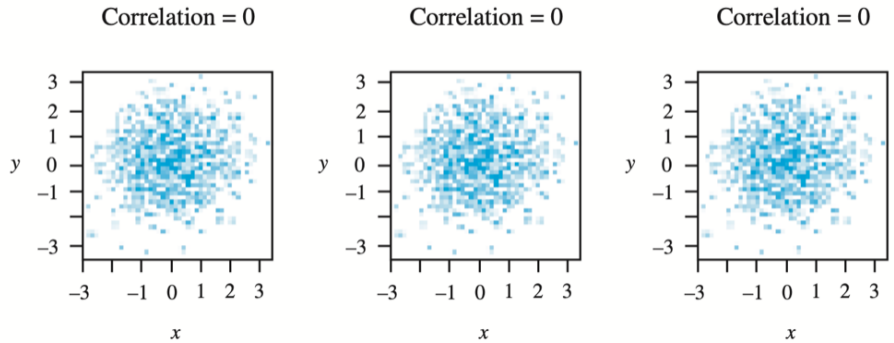
-

Correlations

- **Zero correlation**: there is *no* relationship between the two variables

Examples

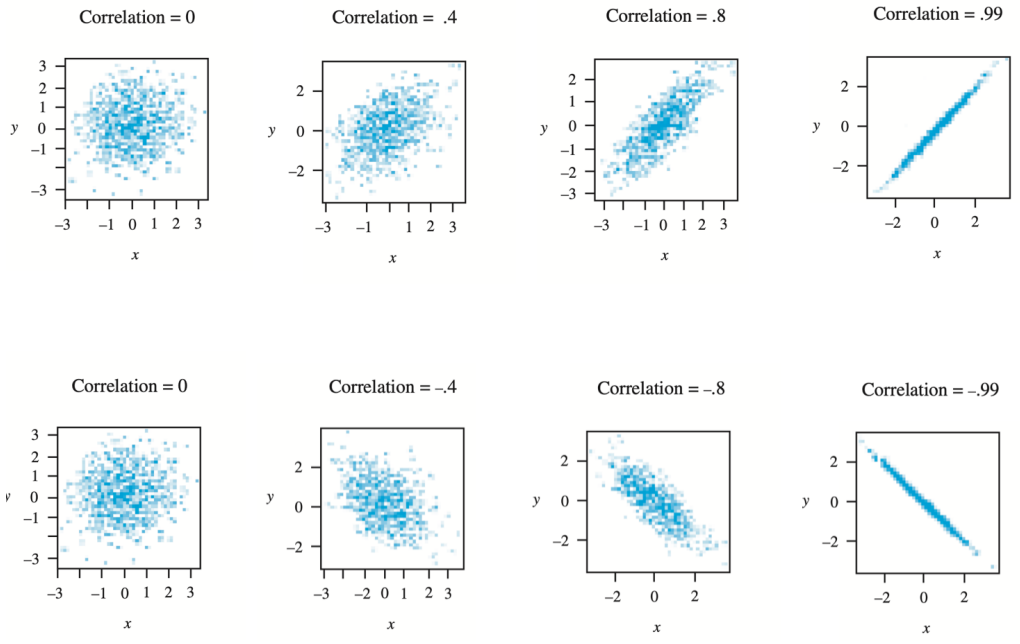
- Relationship between math ability and dancing ability
    - Being better at one tells us nothing about how good you'll be at the other



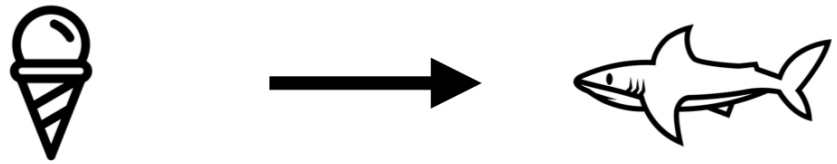
Correlations

Correlations are *always* between -1.0 and +1.0

- This tells you the **magnitude** of relationship
  - Correlations closer to absolute value of 1 are *stronger*
  - Correlations closer to a value of 0 are *weaker*



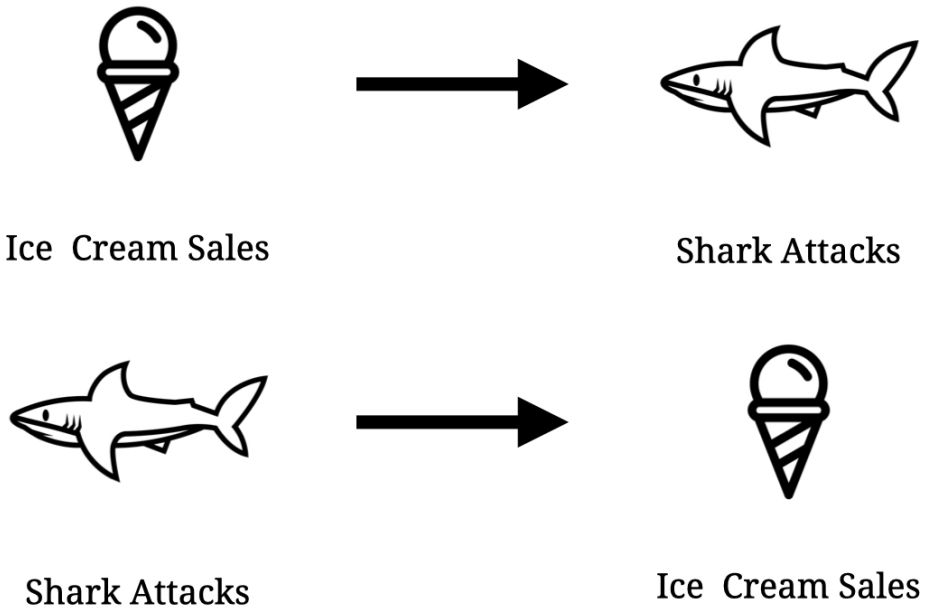
Correlation ≠ Causation



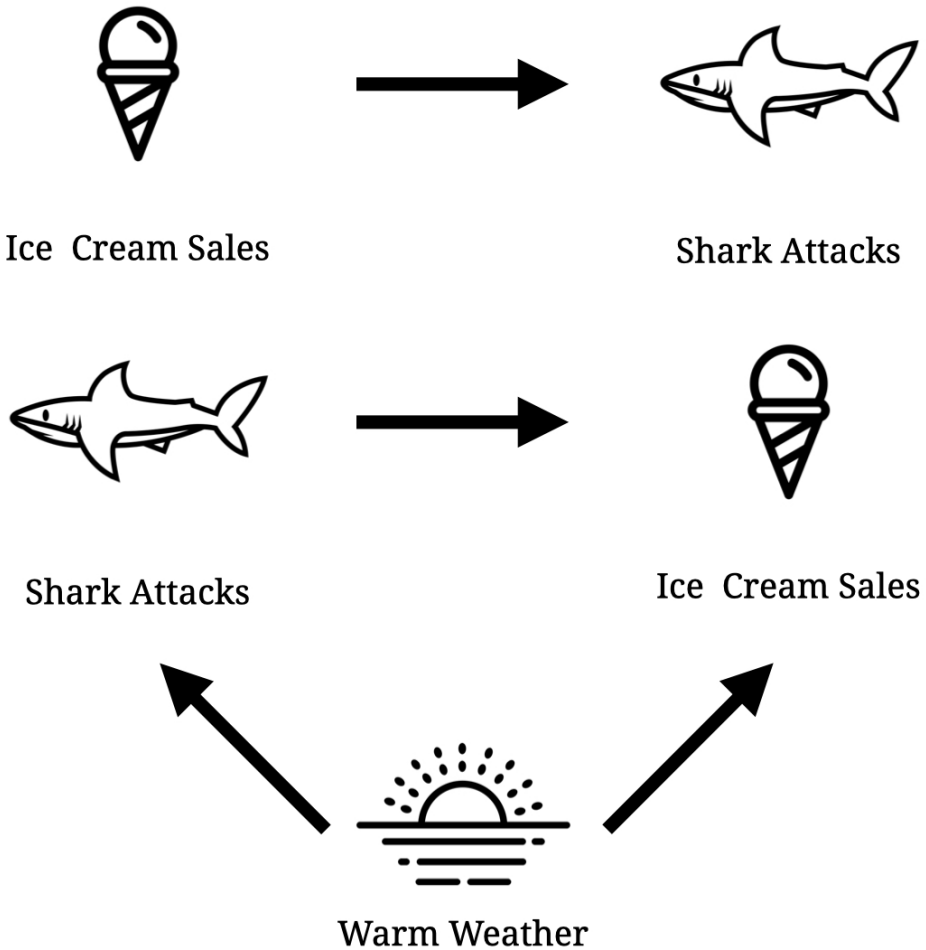
Ice Cream Sales

Shark Attacks

Correlation ≠ Causation



Correlation ≠ Causation



A *third variable* could be the cause of the two events

Correlation: Calculation

Pearson's *r*

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

We just need six values:

$\sum X$
$\sum Y$
$\sum XY$
$\sum X^2$
$\sum Y^2$
n

Correlation: Calculation

Set up our table with the x and y values

x	y	x^2	y^2	xy
5	6			
9	11			
10	6			
3	4			
5	6			
7	9			

Correlation: Calculation

Square our x values

x	y	x^2	y^2	xy
5	6	25		
9	11	81		
10	6	100		
3	4	9		
5	6	25		
7	9	49		

Correlation: Calculation

Square our y values

x	y	x^2	y^2	xy
5	6	25	36	
9	11	81	121	
10	6	100	36	
3	4	9	16	
5	6	25	36	
7	9	49	81	

Correlation: Calculation

Multiply x and y together row-wise

x	y	x^2	y^2	xy
5	6	25	36	30
9	11	81	121	99
10	6	100	36	60
3	4	9	16	12
5	6	25	36	30
7	9	49	81	63

Correlation: Calculation

x	y	x^2	y^2	xy
5	6	25	36	30
9	11	81	121	99
10	6	100	36	60
3	4	9	16	12
5	6	25	36	30
7	9	49	81	63

Sum each of the columns

$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$
39	42	289	326	294

Correlation: Calculation

Pearson's *r*

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$
39	42	289	326	294

$$r_{xy} = \frac{(6)(294) - (39)(42)}{\sqrt{[(6)(289) - (39)^2][(6)(326) - (42)^2]}}$$

Correlation: Calculation

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$$r_{xy} = \frac{(6)(294) - (39)(42)}{\sqrt{[(6)(289) - (39)^2][(6)(326) - (42)^2]}}$$

$$r_{xy} = \frac{126}{\sqrt{(213)(192)}} = 0.62306 = 0.62$$

*In psychology, round to two decimals except *p*-values

Reliability and Validity

A few preliminaries

Terminology

Indendent variable: The variable that is hypothesized to have an effect on some outcome of interest

Dependent variable: The outcome of interest that the independent variable might have an effect on

	Hypothesized to influence	**Hypothesized to be influenced**
Experiment	Independent variable	Dependent variable
Correlational study	Predictor variable	Outcome variable

Terminology

Reliability: Does it work accurately every time?

Validity: Is it actually measuring what it was designed to measure?

Why does this matter?

Since we're heading toward *inferential statistics*, it's very important that our data are what we think they are

You'll learn more about this topic in Research Methods

Validity & Reliability: Example

- Are manipulations and measures reliable and valid?
- **Valid**: measuring/manipulating what you intend to measure?
- **Reliable**: consistency of measurement/manipulation?

Participant watches **sad** video

Webcam shows participant display prototypical **sadness**

Participant self-reports **sadness**

Classical Test Theory

True score: The true value that we are trying to measure

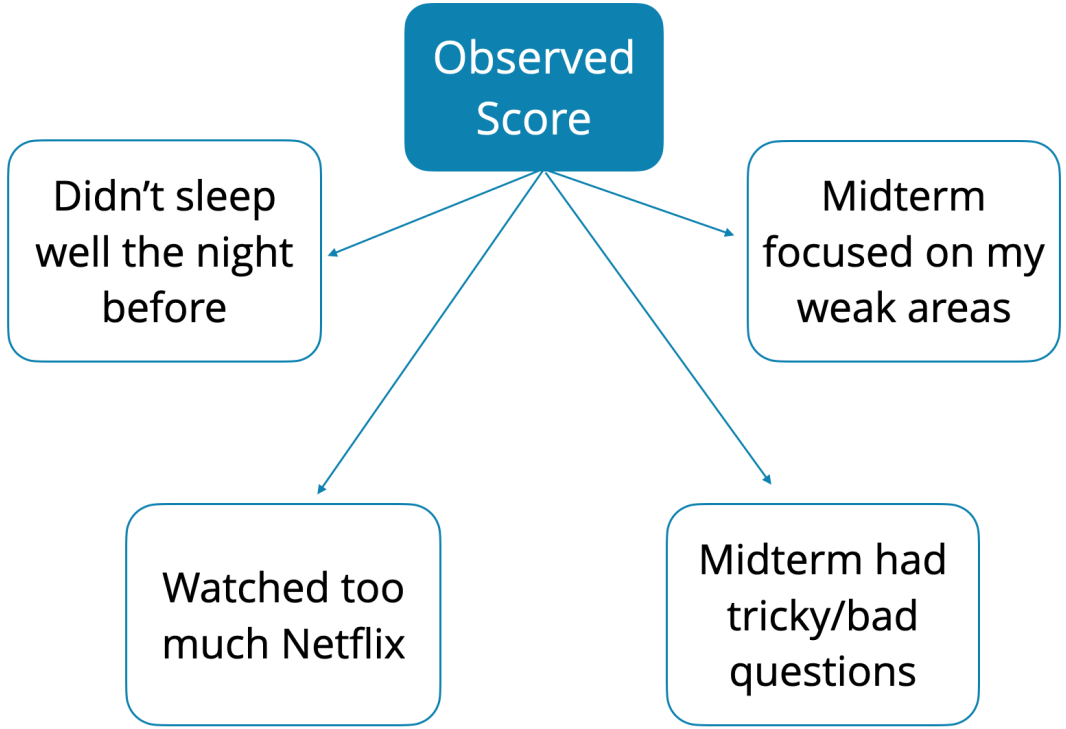
You understand 95% of the concepts and computations

In an ideal world, you would receive a 95% on the midterm

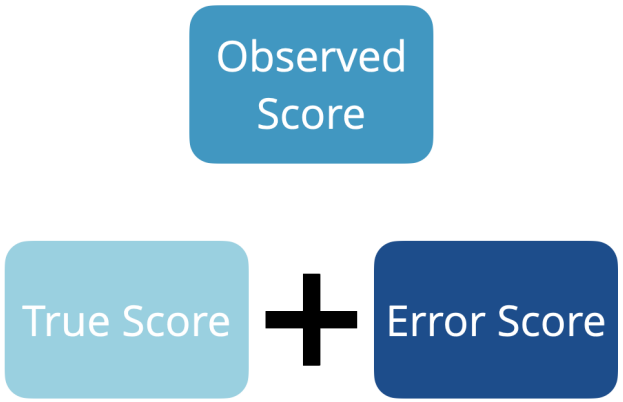
Observed Score: The value that we actually measure

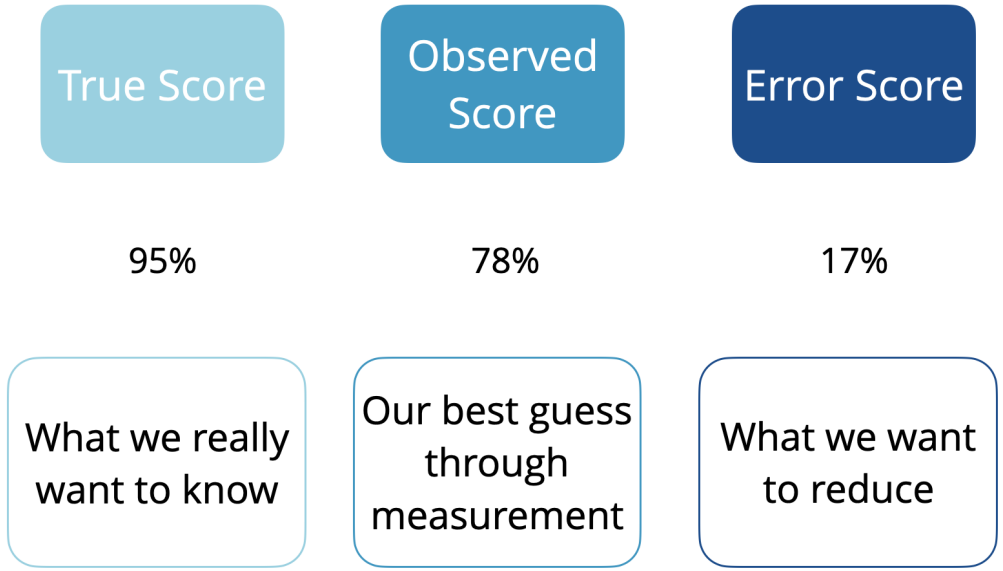
You receive a 78% on the midterm

Classical Test Theory



Classical Test Theory





Estimating and evaluating reliability helps us to reduce the error score and more closely align the true score with the observed score

Reliability

Reliability

- **Test-Retest Reliability**: Used to determine whether a test (or scale) is reliable over time
- **Parallel Forms Reliability**: Used to examine the equivalence or similarity between two forms of the same test (or scale)
- **Internal consistency**: Used to determine whether the items on a test (or scale) are consistent with each other

Test-Retest Reliability

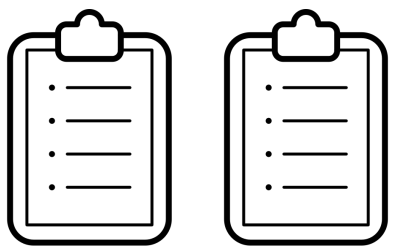
Test-Retest Reliability: Measured as the correlation r_{xy} between scores on a measure at Time 1 and the same measure at Time 2

- Example: Developing a new personality measure

Parallel Forms Reliability

Parallel Forms Reliability: Measured as the correlation r_{xy} between scores for the same individuals on Form A and Form B

- Example: Measuring your ability to calculate standard deviations on two different days
    - We would use two different forms because giving you identical problems would be trivial



Internal Consistency

Internal consistency: Used to determine whether the items on a test (or scale) are consistent with each other, typically measured with Cronbach's α

Anxiety Scale

Rate your agreement to the following statements

1. I often have worrying thoughts
2. I have trouble getting out of bed in the morning
3. I often feel nervous
4. I no longer take pleasure in things I used to enjoy
5. My heart often beats fast as fears enter in
6. I often feel sluggish and tired

Internal Consistency

Anxiety Scale

1. I often have worrying thoughts
2. I have trouble getting out of bed in the morning
3. I often feel nervous
4. I no longer take pleasure in things I used to enjoy
5. My heart often beats fast as fears enter in
6. I often feel sluggish and tired

Issues with this scale

- This scale has anxiety and depression items
- This would result in *poor* internal consistency

Internal Consistency

Internal consistency: Used to determine whether the items on a test (or scale) are consistent with each other, typically measured with Cronbach's α

- The more consistently individual items vary with the total score on the test, the higher the value of Cronbach's α
- Higher values means more internal consistency

Cronbach's Alpha

Cronbach's Alpha

$$\alpha = (\frac{k}{k-1})(\frac{s_y^2 - \sum s_i^2}{s_y^2})$$

k the number of items

s_y^2 the variance associated with the observed score

$\sum s_i^2$ the sum of all the variances for each individual item

Cronbach's Alpha

Item_1	Item_2	Item_3
6	6	8
5	5	6
9	8	6
3	2	4
2	3	2
1	1	2
5	4	6

1. I often have worrying thoughts
2. I often feel nervous
3. My heart often beats fast as fears enter in

Cronbach's Alpha

Item_1	Item_2	Item_3	Total
6	6	8	20
5	5	6	16
9	8	6	23
3	2	4	9
2	3	2	7
1	1	2	4
5	4	6	15

Sum the values of Item 1, Item 2, and Item 3 row-wise to create a "Total" column

Cronbach's Alpha

Item_1	Item_2	Item_3	Total
6	6	8	20
5	5	6	16
9	8	6	23
3	2	4	9
2	3	2	7
1	1	2	4
5	4	6	15

Calculate the **sample variance** for each item and 'total'

Item_1	Item_2	Item_3	Total
7.29	5.81	5.14	48.95

$$\sum s_i^2 = s_1^2 + s_2^2 + s_3^2$$

$$\sum s_i^2 = 7.29 + 5.81 + 5.14 = 18.24$$

Cronbach's Alpha

Item_1	Item_2	Item_3	Total
6	6	8	20
5	5	6	16
9	8	6	23
3	2	4	9
2	3	2	7
1	1	2	4
5	4	6	15

Calculate the **sample variance** for each item and 'total'

Item_1	Item_2	Item_3	Total
7.29	5.81	5.14	48.95

$k = 3$ items in our scale

$s_y^2 = 48.95$ is the 'total' column variance

Cronbach's Alpha

Cronbach's Alpha

$$\alpha = \left(\frac{k}{k-1}\right)\left(\frac{s_y^2 - \sum s_i^2}{s_y^2}\right)$$

$$\alpha = \left(\frac{3}{3-1}\right)\left(\frac{48.95 - 18.24}{48.95}\right)$$

$$\alpha = .941$$

$k = 3$

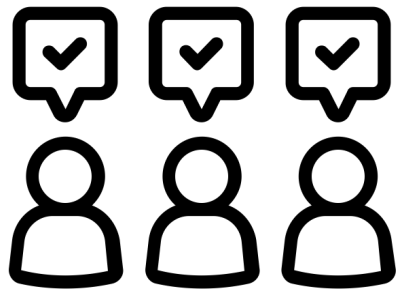
$\sum s_i^2 = 18.24$

$s_y^2 = 48.95$

Interrater Reliability

Interrater Reliability: the degree of agreement among independent observers who rate, code, or agree on their judgments of an outcome of interest

$$Interrater\ Reliability = \frac{Number\ of\ agreements}{Number\ of\ possible\ agreements}$$



Validity

Content Validity

Content validity: the extent to which the items on a test are fairly representative of the entire domain the test seeks to measure

"I want to design a test to measure my students' ability in statistics."

1. Calculate the mean of sample A.
2. Calculate the mean of sample B.
3. Calculate the mean of sample C.
4. Calculate the mean of sample D.
5. Calculate the mean of sample E.
6. Calculate the mean of sample F.

1. Calculate the mean.
2. Calculate the standard deviation.
3. Calculate the correlation.
4. Calculate a *t*-test.
5. Calculate an ANOVA.
6. Calculate a regression.

Criterion Validity

Criterion validity: measures how well one measure predicts an outcome for another measure

Here and now

Concurrent Criterion Validity: Does this test accurately assess my students' current level of ability?

In the future

Predictive Validity: Does this test accurately assess how my students will do on the final exam?

To find out, correlate scores on your test with some well-established measure of ability

Construct Validity

Construct Validity: the degree to which a test, scale, or assessment measures the construct it claims to measure

A *construct* is a group of interrelated variables that you care about

Example: When I measure extraversion as a self-report measure, does this correlate with behavioral extraversion?

Next time

Lecture

- Using hypotheses to test questions

Reading

- Chapter Five
- Chapter Six

Quiz 1

- Due Wednesday 01/19/2022 at 11:59pm MT
    - Lecture 1-4, Ch.1-5