



---

# **Week 8 Lecture - Bivariate Correlation**

Undergraduate Research Methods in Psychology

---

Quinton Quagliano, M.S., C.S.P

Department of Psychology

## Table of Contents

<b>1</b>	<b>Chapter Overview</b>	<b>2</b>
1.1	Learning Objectives . . . . .	2
1.2	Chapter Overview . . . . .	2
<b>2</b>	<b>Bivariate Correlation</b>	<b>3</b>
2.1	Overview . . . . .	3
2.2	Association Between Two Continuous Variables . . . . .	3
2.3	Association Between Categorical and Continuous Variables . . . . .	4
2.4	Fully Correlation Design . . . . .	6
<b>3</b>	<b>Critical Association Claims</b>	<b>7</b>
3.1	Construct Validity . . . . .	7
3.2	Statistical Validity . . . . .	7
3.2.1	Effect Size . . . . .	7
3.2.2	Precision . . . . .	8
3.2.3	Replication . . . . .	8
3.2.4	Outliers . . . . .	9
3.2.5	Restriction of Range . . . . .	9
3.2.6	Curvilinear Relationships . . . . .	10
3.3	Internal Validity . . . . .	11
3.4	External Validity . . . . .	13
<b>4</b>	<b>Key Points</b>	<b>13</b>
4.1	Key Points . . . . .	13

# 1 Chapter Overview

## 1.1 Learning Objectives

- Explain that measured variables - not any particular statistic - make a study correlational.
- Interrogate the construct validity and statistical validity (and, of lower priority, external validity) of an association claim.
- Explain why a correlational study can support an association claim, but not a causal claim.
- Be able to understand the defining characteristics of a correlational, bivariate research design
- Be able to use appropriate vocabulary and logic used in describing and assessing bivariate designs

## 1.2 Chapter Overview

- Moving forward, much of the content will be about applying chapters 1 - 7 to different research \_\_\_\_\_. Make sure you review vocab from the first half of class, as it will often come back in these upcoming lectures.
- “Bivariate” → means \_\_\_\_\_ variables
- Two, **measured** variables → likely an \_\_\_\_\_ claim being made
- *Review:* Remember the vocabulary we should expect with associative claims between  
– Example: \_\_\_\_\_, “associated with”, “correlated with”, etc.



Discuss: Any other examples of vocab normally used with associative claims?


## 2 Bivariate Correlation

### 2.1 Overview

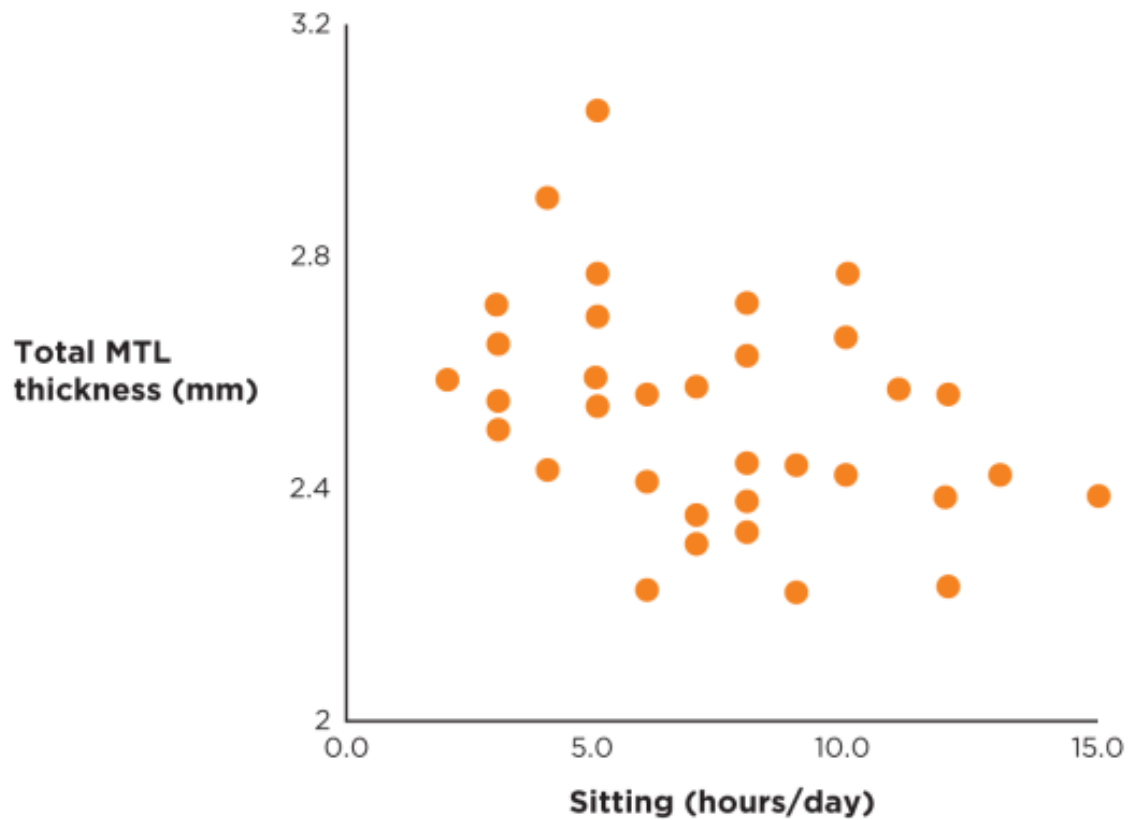
- *Review:* Recall that there are three possible association / \_\_\_\_\_ descriptions between two variables:
  - Positive
  - Negative
  - \_\_\_\_\_ /no relationship
- Also recall that these are descriptors for \_\_\_\_\_ or straight-line relationships - we'll discuss how curvilinear relationships are a bit more complicated later today

### 2.2 Association Between Two Continuous Variables

- Example:
  - Two self-report measurements continuous outcome are strongly, positively correlated with one another
  - *Review:* what type of validity is this possibly associated with?

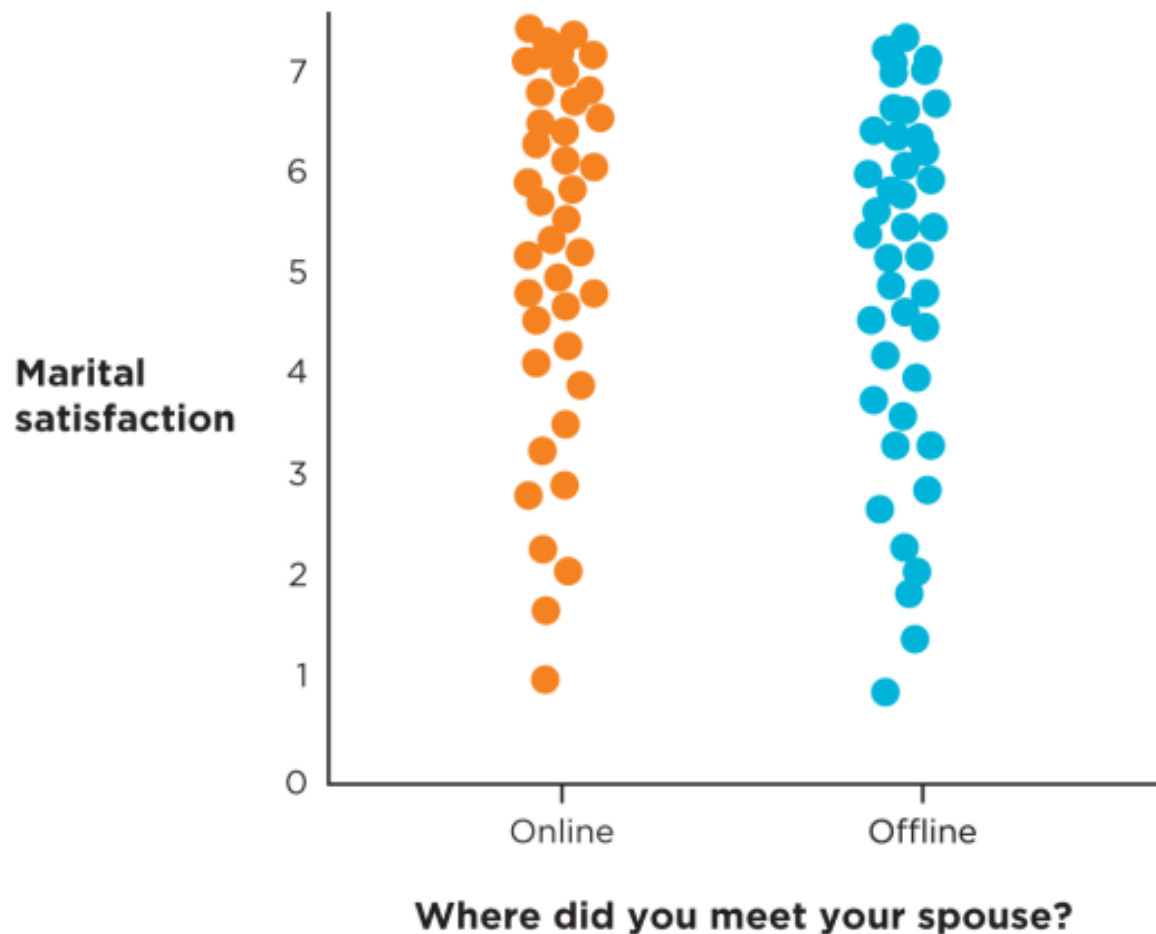
 Discuss: What statistic do we use for association between two continuous variables, and what value would indicate a strong, positive relationship?


- *Review:* When using correlation between two continuous variables, \_\_\_\_\_ are the most appropriate graphing method



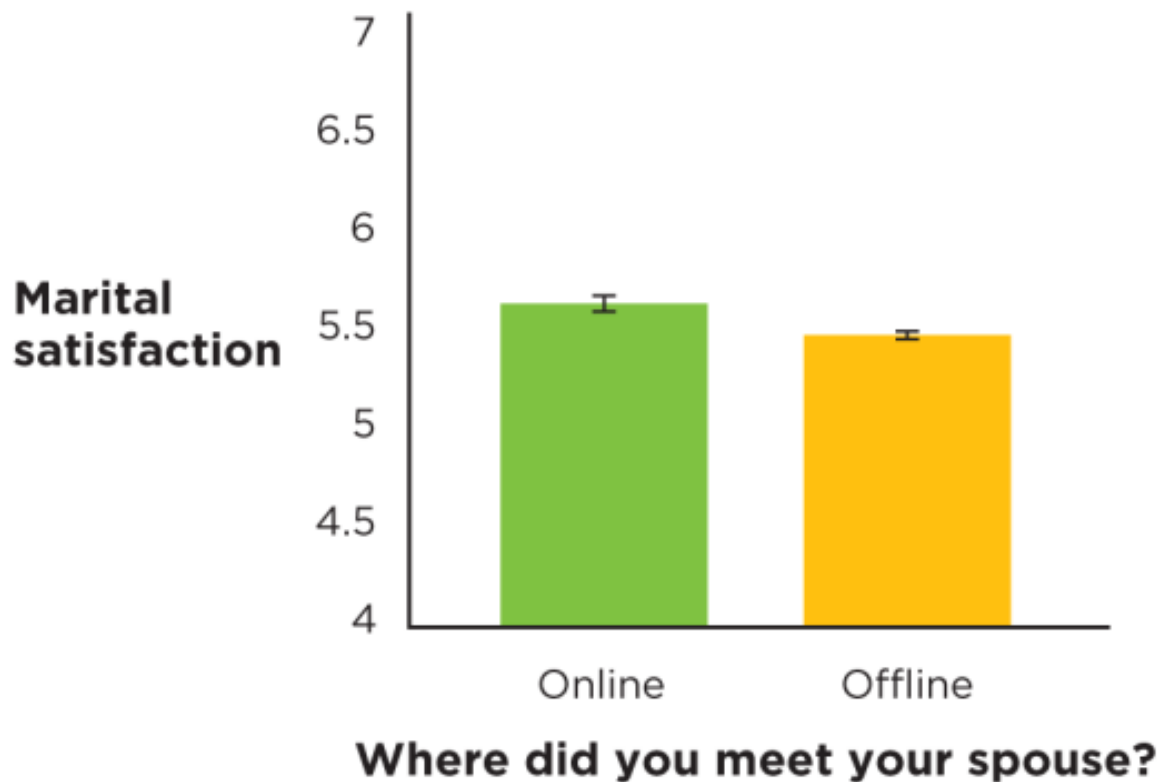
### 2.3 Association Between Categorical and Continuous Variables


- The prior example shows how to work with continuous variables, what about when we have one \_\_\_\_\_ /discrete and one continuous variable?
- One option: A sort of \_\_\_\_\_ scatterplot by the categorical variable



 Discuss: What do you think is the direction and strength of relationship here?

- In most cases, we will probably prefer other \_\_\_\_\_ to graphical display information when we have categorical and continuous variables, as the split scatterplot can be somewhat confusing or difficult to interpret.
- Another option: A \_\_\_\_\_ graph



 Discuss: What do you all think those small, black bars mean? What sorts of helpful information could they tell us about differences between the groups?


## 2.4 Fully Correlation Design

- A \_\_\_\_\_ design is one that results from all relevant variables being measured, *not* based on what scale of measurement was used, statistic, or graph.
- The designs that we discuss are largely defined by whether variables are measured or manipulated *and* the \_\_\_\_\_ we hypothesize between the variables (i.e., what claim we are attempting to make)

## 3 Critical Association Claims

### 3.1 Construct Validity

- Review: construct validity is all about the \_\_\_\_\_ reliability and validity of the tools we use to measure a phenomenon.
  - 3 reliabilities: test-retest, \_\_\_\_\_, and internal
  - 5 validities: \_\_\_\_\_, content, criterion, convergent, divergent/discriminant
- Use the terms and suggestions from chapters 5 and 6 to \_\_\_\_\_ this!
- Example: One of the items on one of my measures may not be measuring the same constructs as the others due to questionable use of the word “not” in the item.

 Discuss: How would we identify an item not well correlated with other items of the same construct (hint: what statistics and what measurement validity type)? What vocabulary do we use to describe this problem presented by 'not' in the question?

### 3.2 Statistical Validity

- Statistical validity has multiple components to be looking for:

#### 3.2.1 Effect Size

- Effect size is all about the *magnitude* of the results we found, it is not the same as \_\_\_\_\_.
- Assessing  $r$  correlation \_\_\_\_\_ effect
  - General guideline for behavioral sciences: 0.10 (weak), 0.20 (moderate), 0.30 (powerful), >0.40 (very powerful)
- We may also use \_\_\_\_\_
  - This refers to the amount of \_\_\_\_\_ that one variable accounts for in another




- Low effect size → possible \_\_\_\_\_ or non-meaningful effect
- High effect size → possible meaningful/ \_\_\_\_\_ effect
- Example: My continuous variables x and y have an  $r = 0.50$  and  $r^2 = 0.25$ 
  - This would be considered a very strong or powerful effect

### 3.2.2 Precision


- *Review from stats class:* What do we call the intervals in which 95% of estimates will fall? \_\_\_\_\_ intervals
- Common \_\_\_\_\_ : CI 95%:[Low-end Estimate, High-end Estimate]
- Larger sample size → usually \_\_\_\_\_ confidence intervals
- More precision = “better”, more \_\_\_\_\_ estimate
- Example: For an r statistic of 0.30, I have CI 95%:[0.28, 0.32], this shows a precise estimate with \_\_\_\_\_ confidence intervals

### 3.2.3 Replication

- A question we should ask: has this association been shown before in the \_\_\_\_\_ ?
- We may \_\_\_\_\_ the study and measurements again on a new group and see if we get a similar value
- *Review:* This is a good use case for literature \_\_\_\_\_ and meta-analysis articles, as they can compile different studies investigating the same association. Remember, we are looking for the \_\_\_\_\_ of evidence in favor of a claim.

 Discuss: What have you all heard about replication in science in your other classes?

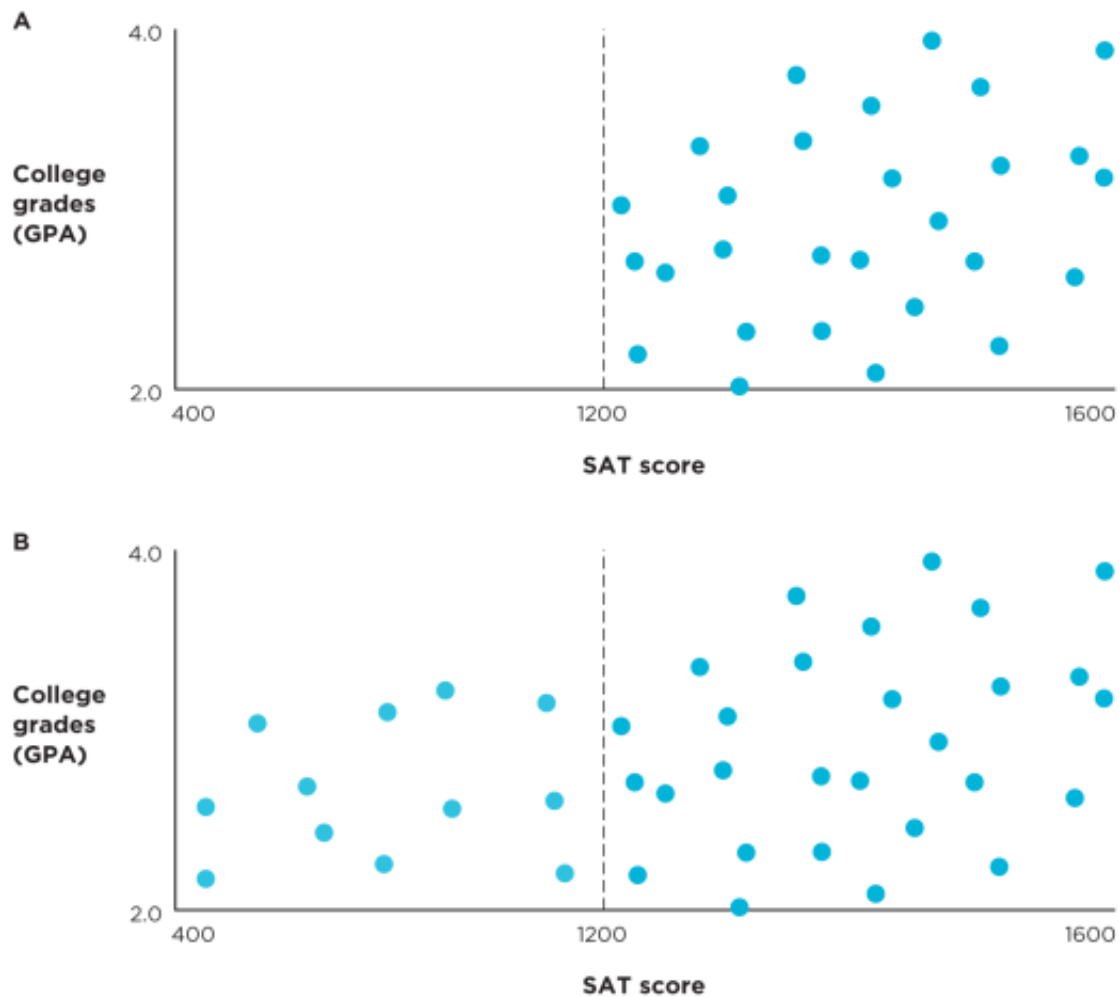
### 3.2.4 Outliers

 Discuss: Based on what you all know from previous classes, what is the definition of an outlier?

- An outlier, especially one \_\_\_\_\_ on both variables, can greatly sway the relationship of two variables.
- However, we should not simply \_\_\_\_\_ out outliers because they are inconvenient findings.
  - Instead, we may consider **robust/non-parametric** \_\_\_\_\_ that can work around them as is
- Depending on who you ask, any value 3 or 4 standard deviations away from the mean would be considered an outlier.

### 3.2.5 Restriction of Range

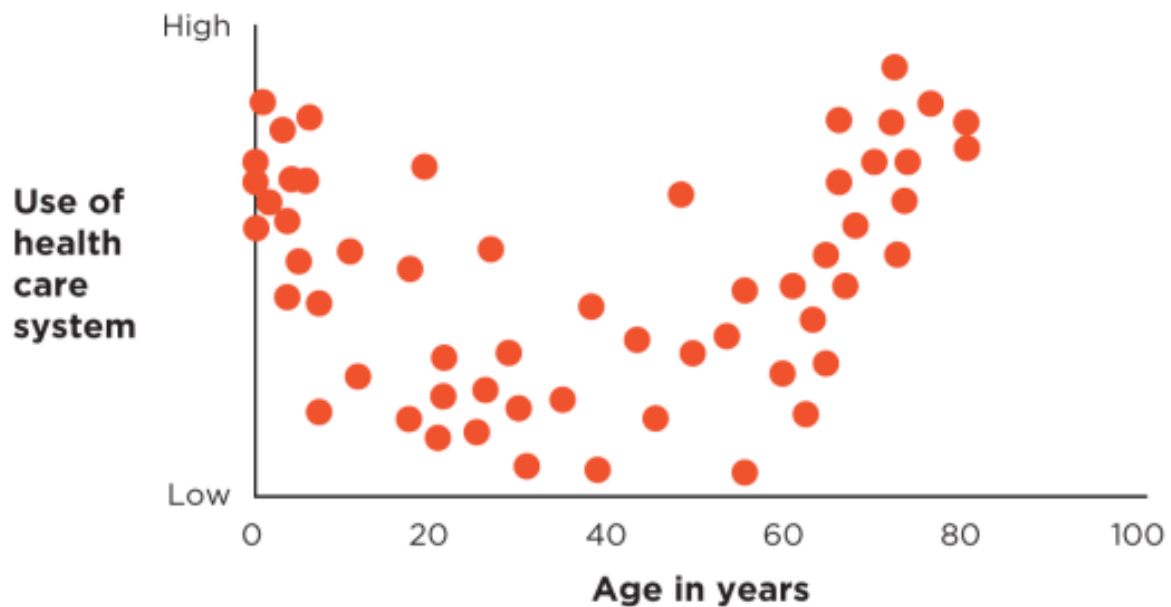
- This is when we do not have points across the full number of \_\_\_\_\_ levels on one or both of our variables



- Example: the possible scores on the test are 0 - 100, but all students only got between 70 and 85 → this is restriction of range problem.


### 3.2.6 Curvilinear Relationships

- $r$  assumes a linear relationship between the two variables, but this is not true of all relationships in reality



### 3.3 Internal Validity

- For association claims, we do not need to be concerned with meeting standards of internal validity, because a correlative design cannot explore \_\_\_\_\_ relationships.

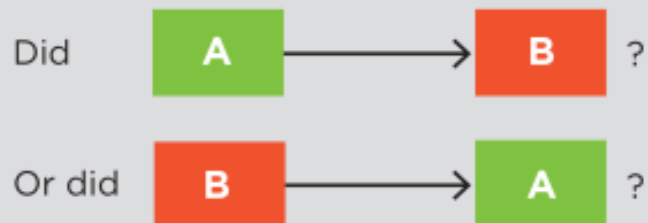
 Discuss: What research design can investigate a causal relationship

- *Review:* What is the common phrase we use to describe the relationship between correlation and causation?
  - Correlation does not \_\_\_\_\_ causation
  - Put simply, association alone does not tell us one “caused” the other!
- *Review:* What were the 3 causation criteria to establish a causal claim?
  - Covariance
  - \_\_\_\_\_ Precedence - **directionality problem**
  - Internal validity - **third-variable problem**

1. *Covariance*: Do the results show that the variables are correlated?

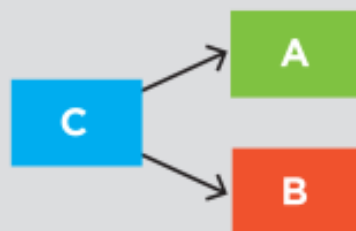


2. *Temporal precedence* (directionality problem): Does the method establish which variable came first in time?



(If we cannot tell which came first, we cannot infer causation.)

3. *Internal validity* (third-variable problem): Is there a C variable that is associated with both A and B, independently?




(If there is a plausible third variable, we cannot infer causation.)

- Example: I am measuring the relationship between exercise and feeling of self-worth, both \_\_\_\_\_ as continuous
  - High, positive correlation between the two - can I say exercise causes higher self-worth? \_\_\_\_\_ !
  - Say energy level is correlated with both exercise and feeling of self-worth, then what really is the cause?

- What is the so-called *third-variable* here? \_\_\_\_\_ level

### 3.4 External Validity

 Discuss: What matters more for 'good', representative sampling: size of sample or methodology?

- Bad external validity does not necessarily \_\_\_\_\_ a claim, but just limits its generalizability
- Example: I use quota sampling to gather my participants, and find an  $r = 0.60$  relationship between the variables
  - What direction and strength of relationship is this? \_\_\_\_\_ and \_\_\_\_\_.
  - Is this a biased or unbiased sample? \_\_\_\_\_ ! As it is non-probabilistic
- \_\_\_\_\_ are some third variable that plays a partial role in the relationship between two variable
  - These may result in a broader, more complicated theory around certain constructs
  - But most psychological constructs are complex!

## 4 Key Points

### 4.1 Key Points

- Bivariate, correlational designs are those that focus on the relationship between two measured variables, i.e., investigating an association claim
  - Assessing a study using this design is a function of (mostly) external, statistical, and construct validity. Internal validity doesn't really apply, as we are missing the critical criteria necessary for a causal claim.
-

- We have several ways to graphically and statistically represent the relationship between two variables, but we also have several considerations such as effect size and precision of our estimate.
- We began to discuss some of the nuance in assessing internal validity, we'll return to this as we creep closer to experiments!