Week 9 Lecture - Multivariate Correlation

Undergraduate Research Methods in Psychology

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

Department of Psychology

Table of Contents

1 Objectives and Overview

1.1 Learning Objective(s)

- State why simple bivariate correlations are not sufficient for establishing causation.
- Explain how longitudinal correlational designs can establish temporal precedence.
- Explain how multiple-regression analyses can rule out some (but not all) third variables.
- Describe the value of pattern and parsimony, in which a variety of research results support a single, parsimonious causal theory.
- Explain the function of a mediating variable.
- Understand the relative strengths and weaknesses of a multivariate design, compared to bivariate designs
- Be able to describe interpreting longitudinal and multiple regression designs, and interpret statistical values from these designs

1.2 Chapter Overview

- _____ $\rightarrow$ multiple (more than two) variables
  - We will still be using a lot of correlation-like techniques
- *Example:* In a _____ design, I might have just depression and anxiety that I am looking at (two variables). In a multivariate version, may I will have depression, anxiety, well-being, and academic success.
- This correlational research is *still* comprised purely of _____ variables
  - We still don't have manipulation, yet
- The techniques and designs discussed today will get us **closer** to a causal claim, but we are still just short of reaching the necessary criteria, and still in range of _____ claim

2 Review of the Casual Criteria

2.1 Overview

- **Covariance**: are _____ (or more) variables varying in some consistent pattern?

- **Temporal Precedence**: does one variable come _____ another in time?
- _____ Validity: is the relationship between our variables not better explained by a third variable?

? Which of these is already met by a bivariate design?

- (●) A) Covariance
- ( ) B) Temporal precedence
- ( ) C) Internal validity
- ( ) D) None of them

Explanation:

3 Longitudinal Designs and Temporal Precedence

3.1 Overview

- **Longitudinal Research:**
  - Research conducted over multiple _____ points (usually over an extended period of time)
  - Normally follows the _____ group of people taking *same* multiple measures at the many time points
- *Example:* I study people's rating of their own physical and mental health health over 5 years, taking measurements every half-year.
- The "multivariate" nature of this comes from the 2 or more measurements happening at _____ time points
  - In essence, the *same* measures are treated as _____ variables because they fall at different time points
- With this design, comes multiples types of _____, we will discuss each type separately in a moment

3.2 Cross-sectional Correlation

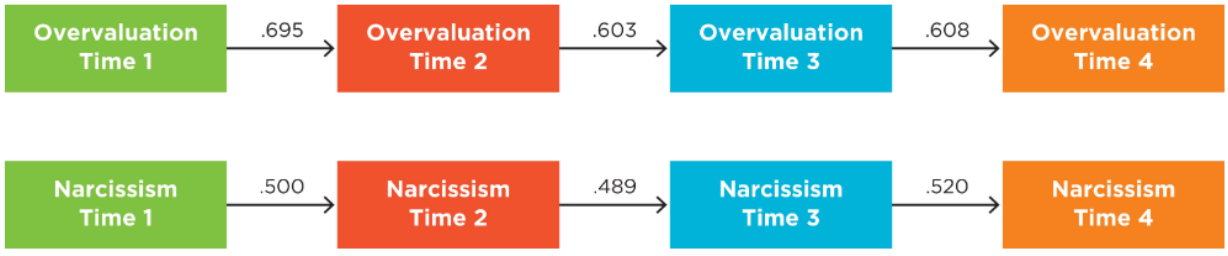
- Are the variables at the same time point correlating with one _____?
- The _____ refer to the separate measures

? Which of the following r values suggest at least weak relationship?

- (•) A) r = .007
- ( ) B) r = .070
- ( ) C) r = .138
- ( ) D) r = .099

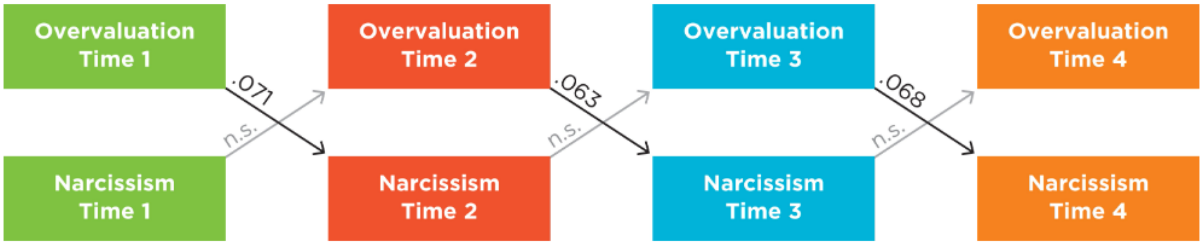
Explanation:

3.3 Autocorrelations



- Are the variables correlated with _____ at the other time points?
- “Auto-” as a root prefix means “self”, so this is a measure correlated with itself

3.4 Cross-lag Correlation



- Are the variables correlated with each other _____ time points?
- This is sufficient for establishing _____
- This is also the _____ outcome we are looking at for longitudinal designs
- The “cross” here should remind you of the diagonal direction of the arrow in this diagram

Discuss: Normally these are interval/ratio variables, but what statistic could we use if they are ordinal?

3.5 Longitudinal and Causation Criteria

? What is the one causal criteria we are still missing in a longitudinal design?

- (•) A) Internal validity
- ( ) B) Covariance
- ( ) C) Temporal precedence
- ( ) D) All of them are met

Explanation:

- However, there needs to be an _____ effort in order to rule out third variables
  - This can be done by including _____ variables to track which may have an impact on one or both of the variables
  - *Example:* If I am worried that energy level may be affecting the relationship between exercise amount and self-worth, I should measure that as well.
- But remember, that even if we get pretty good evidence for all the _____ criteria - we are not fully there yet.

3.6 Why Not Experiment

- You can't always easily _____ variables that may be causing something

> Discuss: What are examples of variables that are difficult to manipulate for practical reasons?

- You can't always _____ manipulate a variable

> Discuss: What are examples of variables that are difficult to manipulate for ethical reasons?

- Sometimes we can do a _____ study that would be unethical to over a longer period of time. Then we can combine those experimental designs with longitudinal designs.
  - *Example:* We can sleep deprive people for one night, but can't do it ethically for an entire month.

4 Multiple Regression and Third Variables

4.1 Overview

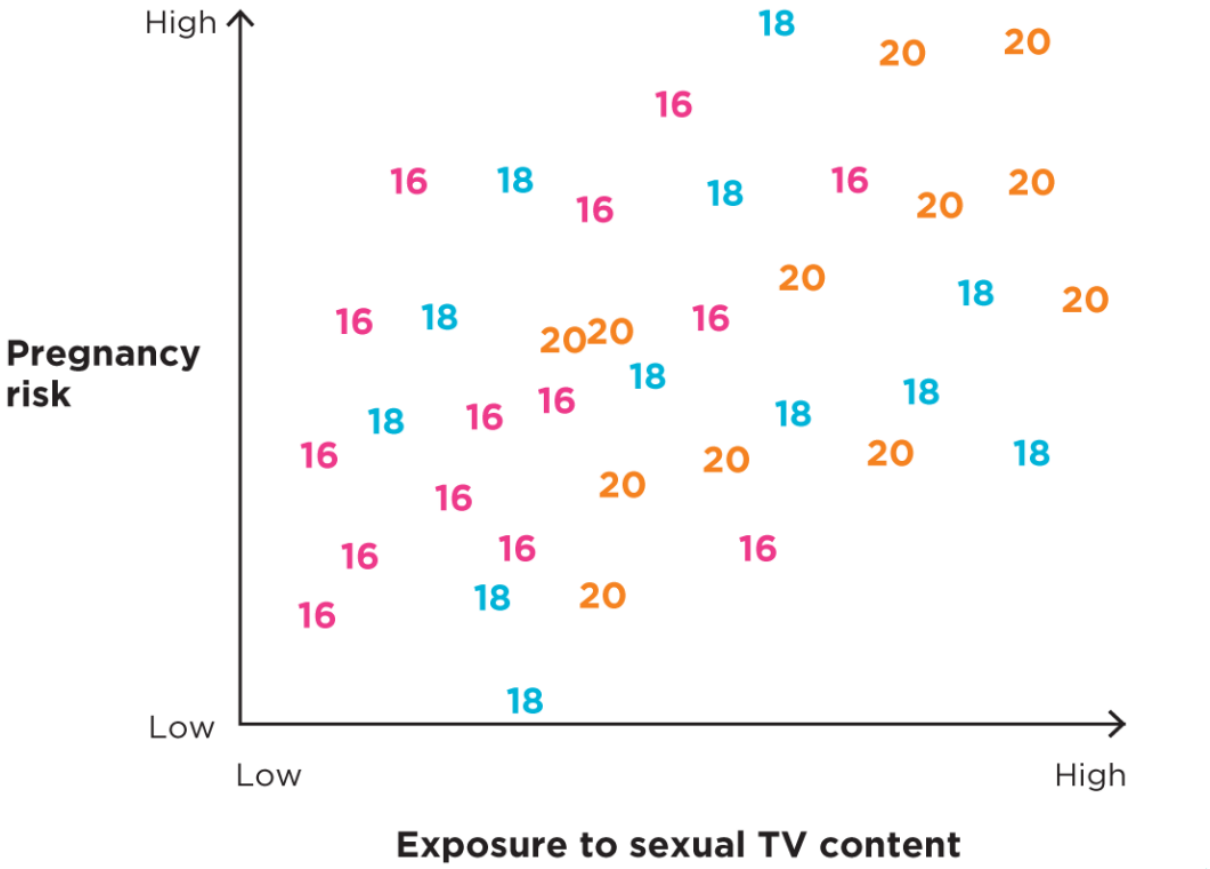
- Multiple _____ regression is when we predict one continuous variable by way of multiple _____ variables
  - Stats sidebar: some books temporarily refer to this as "multivariate regression" - don't call it that, because that is a separate, more complicated technique
- Technically, we can predict with _____ or continuous variables, but our **single** outcome should be continuous
  - Stats sidebar: outcomes of multiple regression *can* be binary/categorical, but this becomes much more complicated and won't be our focus in this class

4.2 Criterion and Predictor Variables

- When we use multiple regression, we have:
  - _____ the "dependent" variable we are trying to predict
  - Predictor variables: Those that we are using to _____ the criterion (duh)
- In summary: multiple predictor variables $\rightarrow$ single criterion variable

4.3 More Than Two Variables

- The benefit to this is we can _____ control for the effects of other variables and _____ the effects of the variables we care about the most



4.4 Statistics for Third Variables

- When we talk about *controlling* for some variable in a model, we are talking about "holding it still" or _____ the effect it has on the criterion or _____
- That way, we can see the individual _____ of each predictor variable on the criterion
- This is *not* the same as making a variable a _____ in the design itself - this is purely a statistical procedure

4.5 Beta coefficients

- Beta, represented as β, is a _____ effect that one unit on the predictor variable changes on the outcome variable

- For example, if our age β is +1.25 on pregnancy risk - that means that for every 1 year of age, _____ pregnancy risks raises by 1.25.
- A beta further from 0 signifies a _____ effect, and we can look for statistical _____ in β just like in other stats.

Discuss: If I have a beta of 4.00 on predictor X for criterion Y, provide an interpretation of the relationship between the two variables.

4.6 Interpretation Example

Multiple-Regression Results from a Study Predicting Pregnancy from Sexual Content on TV and Age

CRITERION (DEPENDENT) VARIABLE: PREGNANCY RISK	BETA	95% CI FOR BETA	Statistical Significance
Predictor (independent) variables:			
Exposure to sex on TV	0.25	[.14, .36]	*
Age	0.33	[.20, .46]	*

Note: Data are fabricated, based on imagined results if the researchers had used only two predictor variables.

*$p < .05$, meaning the result is statistically significant and the 95% CI does not include zero.

4.7 More Predictors, The Merrier?

- There is no natural limit to however many predictors to put into a model, with some caveats:
  - A good rule of thumb is that sample should be at least _____ times the number of predictor variables
  - Too many predictors can _____ effects if the predictors correlate with one another!

Discuss: AI is sometimes described as an especially complicated multiple regression model - why do you think this is?

4.8 In Popular Media

- "Controlling for", "Adjusting for", "Considering" are all phrases that often indicate use of a _____ model. Be on the lookout for these.

4.9 Still Not Causal

- Despite the fact that longitudinal and multiple regression studies are _____ at statistically control of third variable and establishing temporal precedence…
- They don't quite reach the _____ gold standard of experiments

5 Pattern and Parsimony

5.1 Overview

- Sometimes the _____ of evidence from substantial correlation studies appears to be sufficient for establishing causality… is it?

5.2 Meaning

- _____ is the principle that says we should seek the simplest possible (and still accurate) description of a phenomenon or relationship
- The "pattern" we are speaking of here is the pattern of _____ coming from numerous studies

? What type of source is going to give a nice quantitative overview of the weight of evidence in an area?

- (●) A) Original empirical journal article
- ( ) B) Meta-analysis journal article
- ( ) C) Literature review journal article
- ( ) D) Scientific journalism

Explanation:

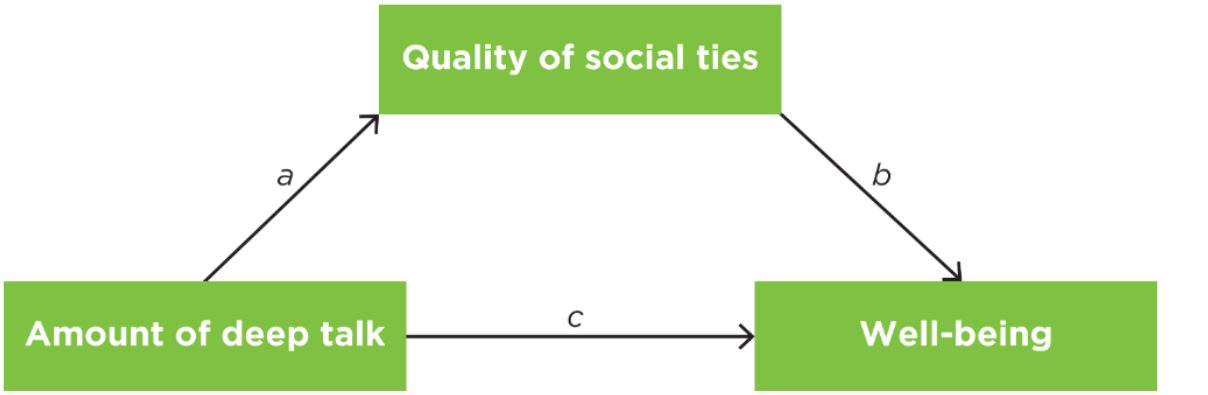
5.3 In Popular Media

- Popular media tends to like big _____ papers that seem to point out a prominent pattern in research
- However, remember the _____ associated with putting too much trust in scientific journalism

6 Mediation

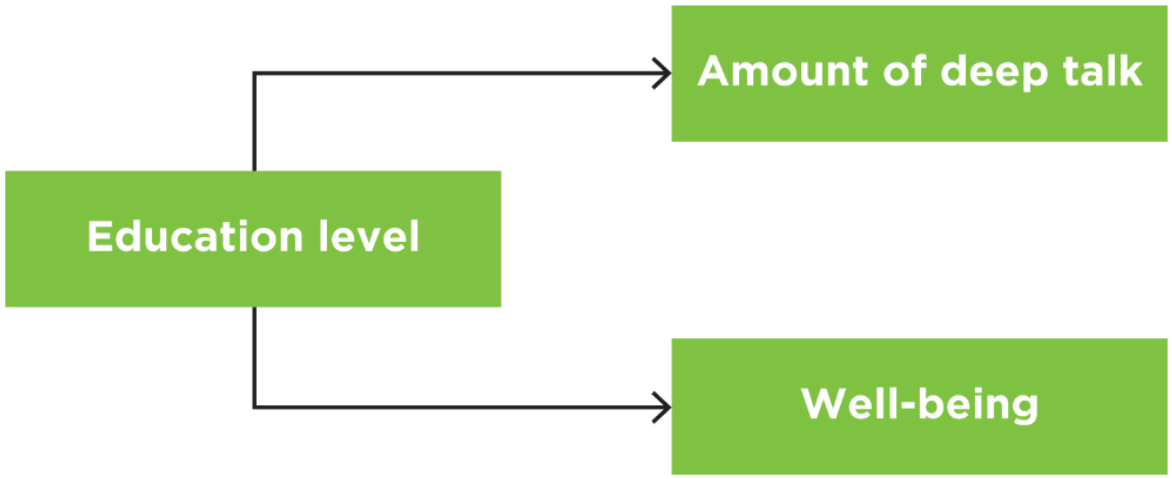
6.1 Overview

- **Mediation:** a claim that one variable's relationship with another is better _____ by another variable, i.e., some mechanism through which an effect occurs - the "why"
- Technically this is a _____ claim, but is often hinted at via correlational designs.



6.2 Compared to "Third Variable"

- A "third variable" more so refers to a variable unrelated to a mechanism that just _____ to be related to both variables of interest



▶ More Information

6.3 Compared to "Moderators"

- A _____ is best described as a variable that's state changes the relationship between two others
- E.g., The link between depression and anxiety are stronger when trauma is high

📢 Discuss: How common do you suspect mediators and moderators are in psychological phenomena?

7 Analysis with the Four Validities

7.1 Overview

- How do we investigate these designs? Largely the same as the bivariate designs!

7.1.1 External Validity

- How was the sample _____? To a lesser extent, how many people are in the sample, and what are their demographics?

7.1.2 Statistical Validity

- What is our _____ (effect size)?
- What is our precision (_____ intervals)?
- Are our statistics _____?
- Do these results _____ in other studies?
- Do we have a restriction of _____ or outliers?

❓ Which of the following scenarios seems most indicative of poor precision?

- (●) A) We have an r-squared of 0.25
- ( ) B) Our p-value is 0.10
- ( ) C) Our effect isn't shown in other studies
- ( ) D) Our confidence intervals for r are [0.20, 0.80]

Explanation:

7.1.3 Construct Validity

- Are our measures consistent and _____?
- Are our measures accurate and _____ to our intended construct?