



---

# **Week 9 Lecture - Multivariate Correlation**

Undergraduate Research Methods in Psychology

---

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

Department of Psychology

## Table of Contents

<b>1</b>	<b>Objectives and Overview</b>	<b>2</b>
1.1	Learning Objective(s)	2
1.2	Chapter Overview	2
<b>2</b>	<b>Review of the Casual Criteria</b>	<b>2</b>
2.1	Overview	2
<b>3</b>	<b>Longitudinal Designs and Temporal Precedence</b>	<b>3</b>
3.1	Overview	3
3.2	Cross-sectional Correlation	4
3.3	Autocorrelations	4
3.4	Cross-lag Correlation	5
3.5	Longitudinal and Causation Criteria	5
3.6	Why Not Experiment	6
<b>4</b>	<b>Multiple Regression and Third Variables</b>	<b>7</b>
4.1	Overview	7
4.2	Criterion and Predictor Variables	7
4.3	More Than Two Variables	7
4.4	Statistics for Third Variables	8
4.5	Beta coefficients	8
4.6	Interpretation Example	9
4.7	More Predictors, The Merrier?	9
4.8	In Popular Media	10
4.9	Still Not Causal	10
<b>5</b>	<b>Pattern and Parsimony</b>	<b>10</b>
5.1	Overview	10
5.2	Meaning	10
5.3	In Popular Media	11
<b>6</b>	<b>Mediation</b>	<b>11</b>
6.1	Overview	11
6.2	Compared to “Third Variable”	12
6.3	Compared to “Moderators”	12
<b>7</b>	<b>Analysis with the Four Validities</b>	<b>13</b>
7.1	Overview	13
7.1.1	External Validity	13
7.1.2	Statistical Validity	13
7.1.3	Construct Validity	14

# 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

- \_\_\_\_\_ → 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 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



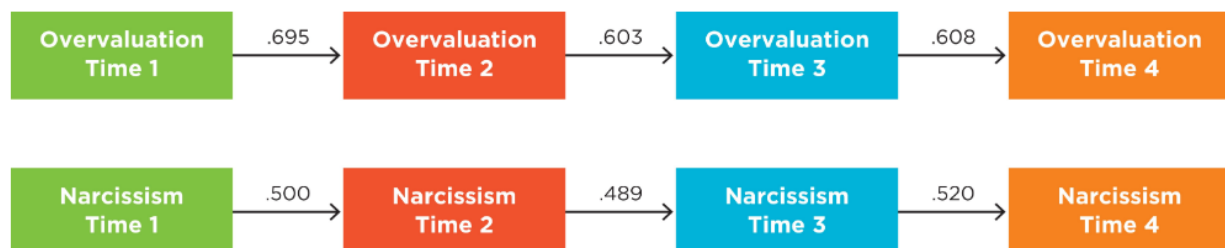
- Are the variables at the same time point correlating with one \_\_\_\_\_?
- The \_\_\_\_\_ refer to the separate measures

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

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

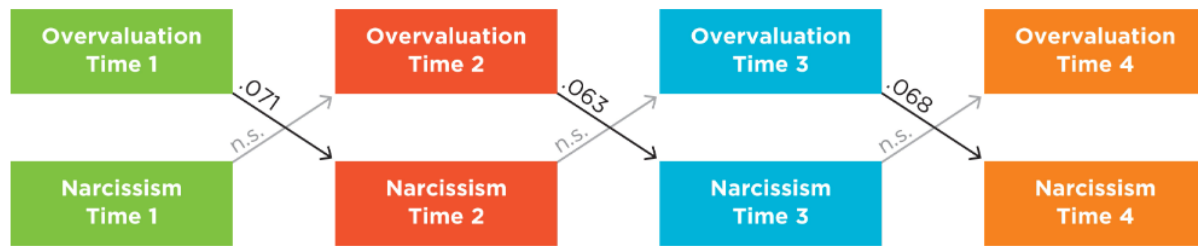
Explanation:

### 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

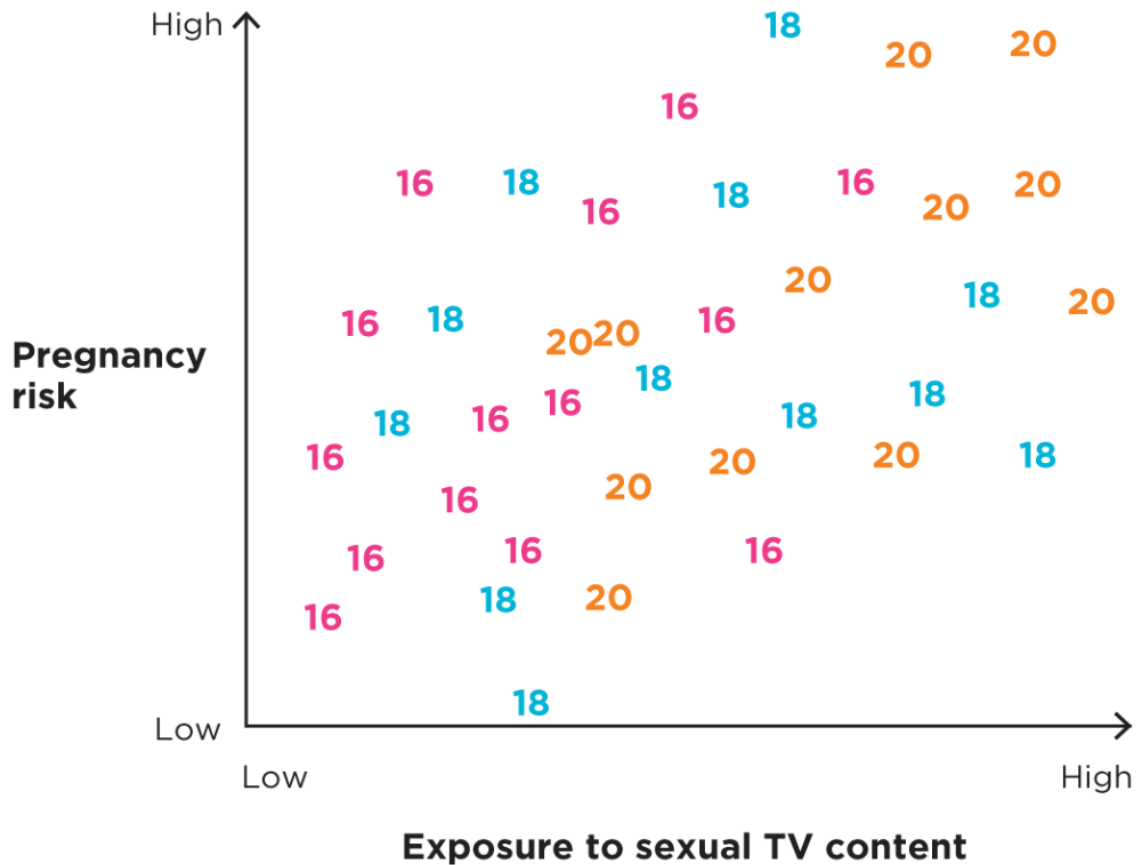
### 4.2 Criterion and Predictor Variables

- 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 → 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  $\beta$ , is a \_\_\_\_\_ effect that one unit on the predictor variable changes on the outcome variable

- For example, if our age  $\beta$  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  $\beta$  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
<b>Predictor (independent) variables:</b>			
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:

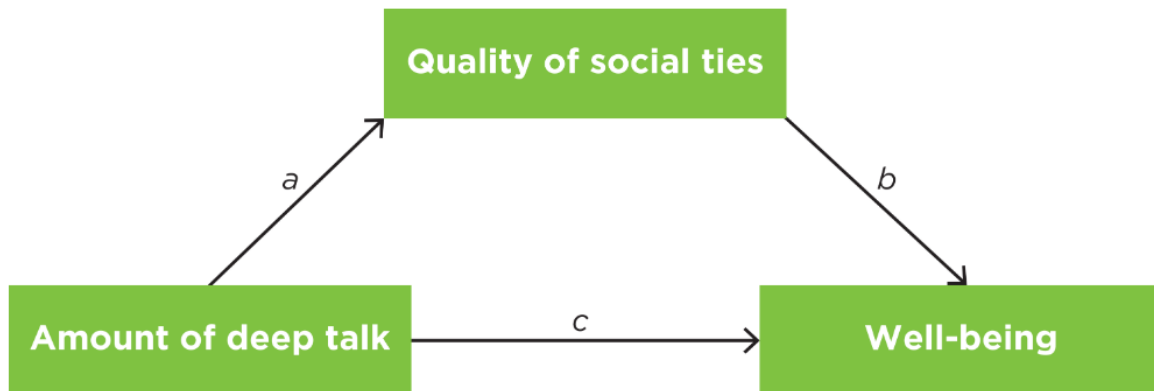
### 5.3 In Popular Media

- 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

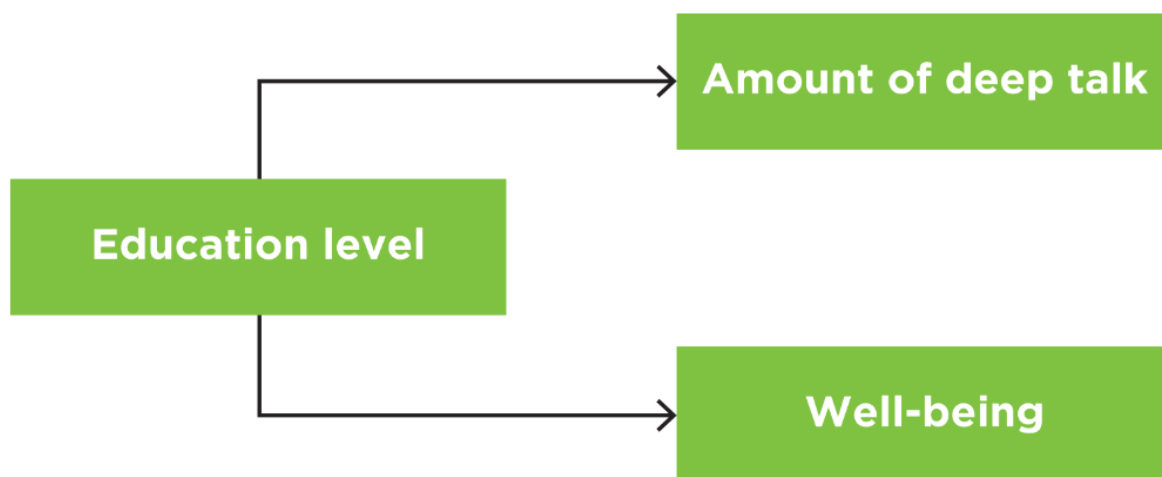
### 6.1 Overview

- **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?