Causal Inference in Jamovi

# Welcome to Causal Inference!

You've built an impressive toolkit of statistical skills over the past several chapters. Now, in Chapter 6, you'll learn techniques to **strengthen causal claims** even when you can't run a true experiment.

## The Two Big Problems with Correlation

| **Problem** | **What It Means** |
| --- | --- |
| **Third-Variable Problem** | A hidden variable might cause BOTH variables you're studying (e.g., age causes both marriage and lower depression) |
| **Directionality Problem** | You can't tell which variable causes the other (does anxiety cause depression, or depression cause anxiety?) |

**🔄 Building on What You Know:** This chapter uses the same clinical dataset from Chapters 4-5. You'll use correlations and t-tests (Ch. 5), but now add covariates to control for third variables!

## What You'll Learn

* **ANCOVA** — Compare groups while controlling for a covariate
* **Multiple Regression** — Examine correlations while controlling for third variables
* **Partial Correlations** — See how relationships change after controlling for other factors
* **Cross-Lagged Analysis** — Establish temporal precedence with longitudinal data

# Getting Started: The Chapter 6 Data Files

Chapter 6 uses two data files from the OSF page (https://osf.io/a8kev/):

* **RITC\_DATA\_CH06\_ClinicalStudy** — For Activities 6.1-6.3 (same clinical data, but may include marital status)
* **RITC\_DATA\_CH06\_Longitudinal** — For Activity 6.4 (anxiety & depression measured at two time points)

## Filtering Data in Jamovi

Some analyses require filtering to include only certain participants (e.g., only Single and Married, excluding Divorced). As a reminder, to filter data In Jamovi:

1. **Click on the Data tab**
2. **Click "Filters"**
3. **In the filter row, type a formula** like:

MaritalStatus == 1 or MaritalStatus == 6

*(This keeps only Single [1] and Married [6] participants)*

1. **Filtered-out rows will be grayed out** and excluded from analyses

# Activity 6.1: Controlling for Third Variables with ANCOVA

The textbook shows that married people report less depression than single people. But married people are also **older**, and older people tend to be less depressed. Is marriage really the cause, or is age the hidden explanation?

**ANCOVA** (Analysis of Covariance) lets us compare groups while **statistically holding age constant** — as if everyone were the same age.

## Part A: First, Run the Basic T-Test (No Control)

Before controlling for age, let's see the original difference:

1. **Filter the data** to include only Single and Married participants. The result should look like the image below.

**\*Note**: the relationship status variable is labeled ‘mar’ in this dataset.

A screenshot of a computer

AI-generated content may be incorrect.

1. **Go to Analyses → T-Tests → Independent Samples T-Test**
2. **Move Depression** to Dependent Variables
3. **Move MaritalStatus** to Grouping Variable
4. **Check "Descriptives"** to see the means for each group. The result should look like the image below.

A screenshot of a test

AI-generated content may be incorrect.

**💡 Expected Result:** Single people (M ≈ 7.4) have significantly higher depression than married people (M ≈ 5.8), t ≈ 2.5, p < .05

## Part B: Now Run the ANCOVA (Controlling for Age)

Now let's see what happens when we control for age:

1. **Go to Analyses → ANOVA → ANCOVA**
2. **Move Depression** to the "Dependent Variable" box
3. **Move Marital Status (mar)** to the "Fixed Factors" box
4. **Move Age** to the "Covariates" box
5. **Under "Estimated Marginal Means":**

• Check the box for Marginal means tables. This will give you the mean for each group

• Move mar to the "Term 1" box

• Check "Marginal means plots" to visualize the adjusted means

**🎯 What to Look For:** Look at the ANCOVA table for MaritalStatus (mar). Is the p-value still < .05? Look at the "Estimated Marginal Means" — these are the adjusted means as if everyone were the same age.

**💡 Expected Result:** After controlling for age, the difference shrinks! Married (M ≈ 6.10) vs Single (M ≈ 6.97) — and the difference is NO LONGER significant (p > .05). This suggests age, not marriage, explains the original difference.

### A screenshot of a computer AI-generated content may be incorrect.

A screenshot of a graph

AI-generated content may be incorrect.

### Interpreting Your Results

When the effect **disappears** after controlling for a third variable, it suggests the third variable was driving the original relationship. In this case: married people aren't less depressed *because* they're married. Instead, it seems, they're less depressed because they're **older**.

# Activity 6.2: Multiple Regression — Controlling for Trauma

In Chapter 5, we found a strong correlation between anxiety and depression (r = .82). But could a third variable like **trauma** explain this relationship? Maybe people who've experienced trauma develop BOTH anxiety and depression.

**Multiple regression** lets us examine the anxiety-depression correlation while controlling for trauma. The result is called a **partial correlation**.

## Step-by-Step: Running Multiple Regression in Jamovi

1. **Go to Analyses → Regression → Linear Regression**
2. **Move Depression** to the "Dependent Variable" box
3. **Move BOTH variables to "Covariates":**

• Anxiety (your main predictor)

• Trauma (the third variable you're controlling for)

1. **Expand "Model Coefficients" and check:**

• "Standardized estimate" — this gives you the partial correlation. The result should look like below.

A screenshot of a computer

AI-generated content may be incorrect.

### Understanding the Output

In the "Model Coefficients" table, look at:

* **Anxiety row:** The "Stand. Estimate" (β) is essentially the partial correlation
* **p-value:** Is the anxiety-depression relationship still significant after controlling for trauma? The p value show tells you whether the correlation is significant.

**💡 Expected Result:** Original r = .82 → Partial r ≈ .76 after controlling for trauma. The relationship is slightly weaker but STILL strong and significant! This tells us trauma explains SOME of the relationship, but anxiety and depression have a robust connection independent of trauma.

# Activity 6.3: Controlling for MULTIPLE Third Variables

What if multiple factors explain the anxiety-depression relationship? Let's control for trauma, sleep, income, AND education all at once!

## Step-by-Step

1. **Go to Analyses → Regression → Linear Regression**
2. **Move Depression** to Dependent Variable
3. **Move ALL these variables to Covariates:**

• Anxiety (main predictor)

• Trauma (control #1)

• Sleep (control #2)

• Income (control #3)

• Education (control #4)

1. **Check "Standardized estimate"** under Model Coefficients. The result should look like the image below.

A screenshot of a computer

AI-generated content may be incorrect.

**💡 Expected Result:** The partial correlation drops further to about .60, but remains highly significant. Even when controlling for trauma, sleep, income, AND education, anxiety still strongly predicts depression.

### What This Tells Us

* Original r = .82
* Controlling for trauma: r = .76
* Controlling for trauma + sleep + income + education: r = .60

Each control variable explains SOME variance, but the anxiety-depression relationship remains robust. This builds confidence that there's something meaningful connecting these two conditions.

# Activity 6.4: Cross-Lagged Correlations (Longitudinal Data)

Controlling for third variables addresses one problem. But what about directionality? Does anxiety cause depression, or does depression cause anxiety?

**Longitudinal data** helps establish which variable comes first. A **cross-lagged correlation** examines whether one variable at Time 1 predicts another variable at Time 2.

**📁 Use the Longitudinal Dataset:** Download RITC\_DATA\_CH06\_Longitudinal from OSF. This contains anxiety and depression from 500 Connect participants measured one year apart!

## Understanding the Variables

* **Anxiety\_T1** — Anxiety at Time 1 (baseline)
* **Depression\_T1** — Depression at Time 1 (baseline)
* **Anxiety\_T2** — Anxiety at Time 2 (one year later)
* **Depression\_T2** — Depression at Time 2 (one year later)

## The Question

Does Depression at Time 1 predict Anxiety at Time 2, **even after controlling for Anxiety at Time 1**? If yes, it suggests depression *precedes* the development of anxiety.

## Step-by-Step: Cross-Lagged Regression

* **Go to Analyses → Regression → Linear Regression**
* **Move Anxiety\_T2** to Dependent Variable (what you're predicting)
* **Move BOTH to Covariates:**

• Depression\_T1 (predictor from Time 1)

• Anxiety\_T1 (control — baseline anxiety)

* **Check "Standardized estimate."** The result should look like the image below.

**A screenshot of a computer

AI-generated content may be incorrect.**

**💡 Expected Result:** Depression\_T1 significantly predicts Anxiety\_T2 even after controlling for Anxiety\_T1. This establishes temporal precedence — depression comes first and predicts INCREASES in anxiety over time.

### Try the Reverse!

Run another regression with Depression\_T2 as the outcome, and Anxiety\_T1 + Depression\_T1 as predictors. Does anxiety also predict future depression? If both directions are significant, there may be bidirectional causality.

# Quick Reference: Which Analysis Should I Use?

| **Your Question** | **Use This** | **Jamovi Location** |
| --- | --- | --- |
| Compare groups, control for covariate | **ANCOVA** | ANOVA → ANCOVA |
| Correlation controlling for third variable(s) | **Linear Regression** | Regression → Linear Regression |
| Does T1 predict T2? (longitudinal) | **Cross-Lagged Regression** | Regression → Linear Regression |

# How to Report Your Results

## Reporting an ANCOVA

*"I conducted an ANCOVA to examine whether marital status predicted depression while controlling for age. After controlling for age, the difference between married (M = 6.10) and single (M = 6.97) participants was no longer statistically significant, F(1, 297) = 1.92, p = .17. This suggests that age, rather than marital status itself, may explain why married people report lower depression."*

## Reporting Multiple Regression with Controls

*"I conducted a multiple regression analysis to examine the relationship between anxiety and depression while controlling for trauma. The zero-order correlation was r = .82. After controlling for trauma, the partial correlation remained significant, β = .76, p < .001, suggesting that anxiety and depression share a robust relationship independent of trauma."*

## Reporting Cross-Lagged Regression

*"I conducted a cross-lagged regression to examine whether Time 1 depression predicted Time 2 anxiety, controlling for baseline anxiety. Results indicated that depression at Time 1 significantly predicted anxiety one year later, β = .12, p < .05, even after accounting for initial anxiety levels. This establishes temporal precedence for depression in the development of anxiety symptoms."*

# Summary

Congratulations! You've learned powerful techniques for strengthening causal claims:

* **ANCOVA** — Compare groups while controlling for continuous covariates
* **Multiple Regression** — Examine correlations while controlling for third variables
* **Partial Correlations** — See the "pure" relationship between two variables
* **Cross-Lagged Analysis** — Establish which variable comes first over time

**⚠️ Remember:** Even with these techniques, correlational research cannot PROVE causation. There might always be unmeasured third variables. But these methods help build a stronger case!

**📚 Looking Ahead:** In Chapter 7, you'll learn about TRUE experiments — where researchers actually manipulate variables and randomly assign participants. This is the gold standard for establishing causation!