PSY 5939: Longitudinal Data Analysis

Two observations

Two observations

Basic models for 2 observations per unit

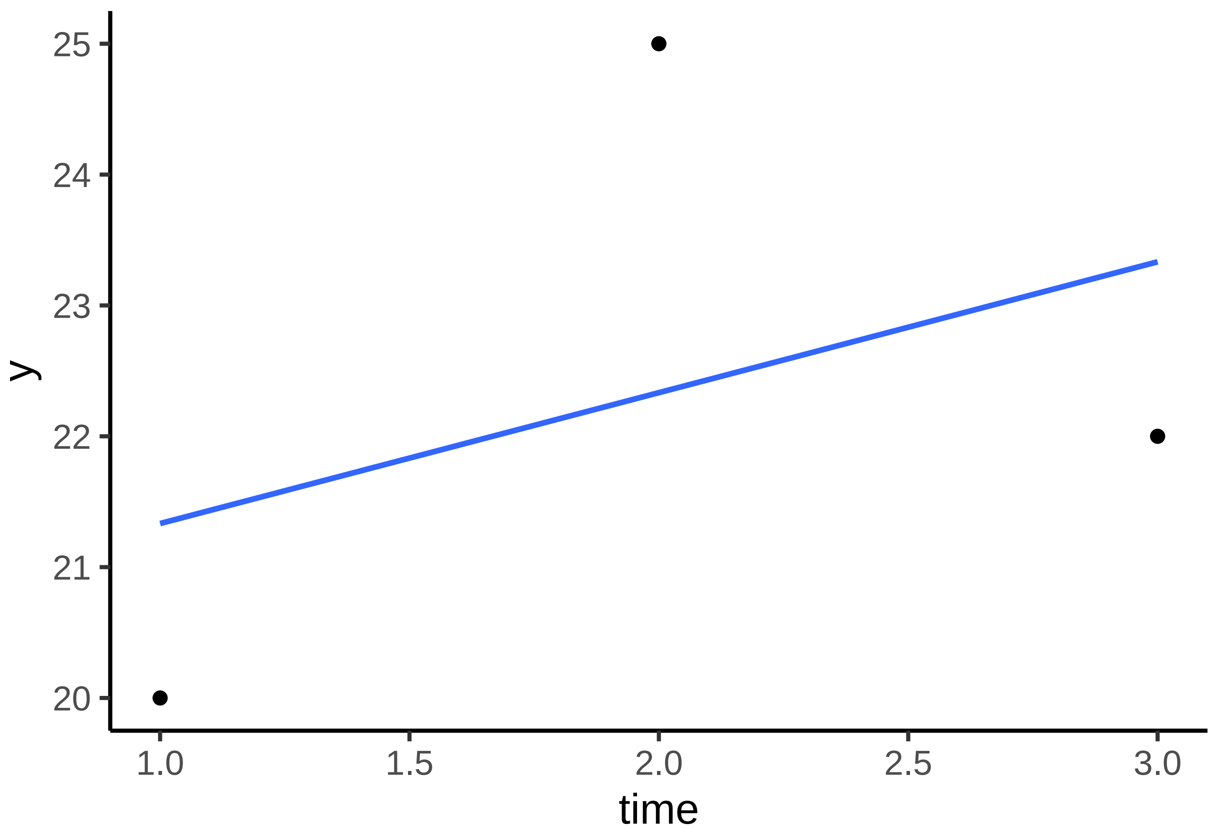
True longitudinal data has 3 or more observations per unit
You need at least 3 observations per unit for both mixed models and latent growth models

But sometimes you just don't have 3 observations

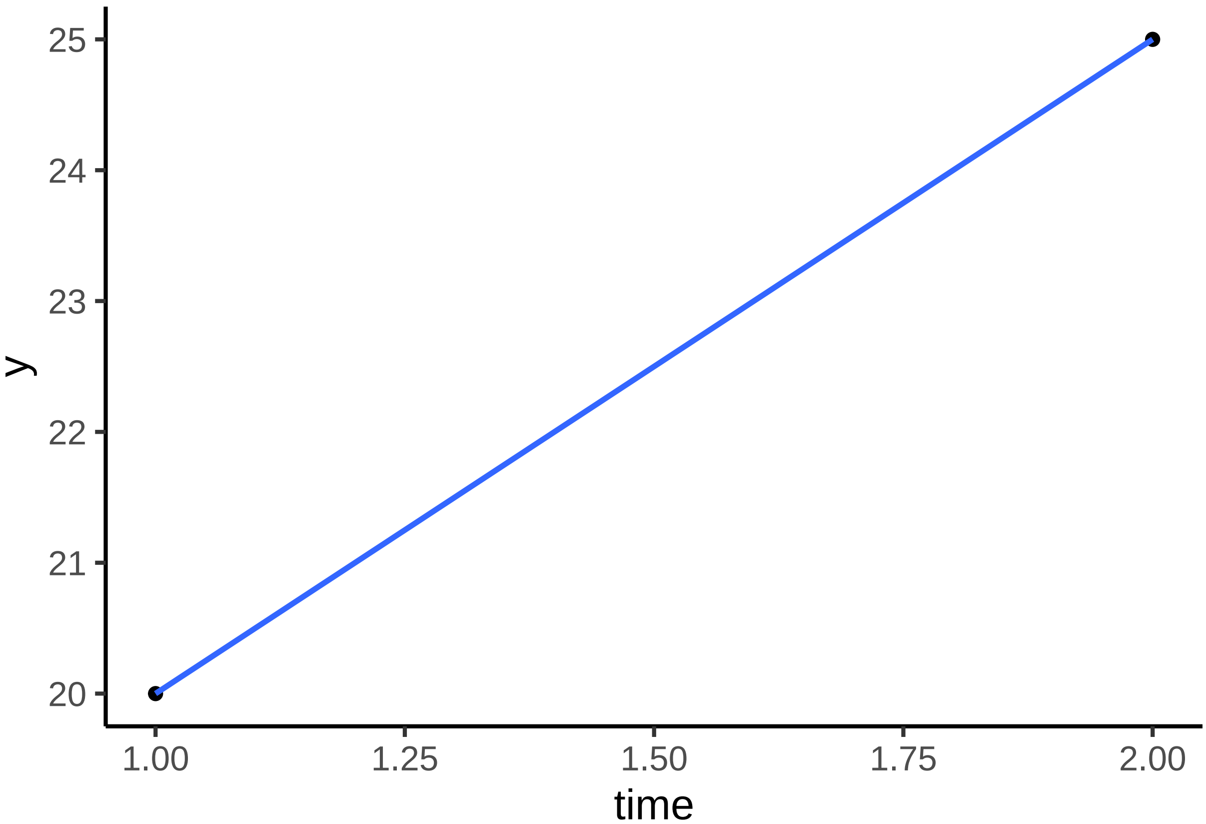
- Someone else's data (your advisor's, public data, etc.)
- Poor planning on your part :(

There **are** some basic models based on ANOVA and regression that can be used when you have just 2 observations per unit

3 observations per unit



2 observations per unit



Types of analysis

- Difference scores (also called "change scores")
- Partial change scores (also called "residualized change scores")
- Analysis of covariance

These are **different analyses**, so they **may give different results**

- Shocking!

The analyses themselves are quite simple

We'll focus on how each method handles change (or does not) and how the nuanced differences between methods lead to different results

Difference scores

Difference scores

Difference scores

For longitudinal data, a difference score is the **difference** between scores on a variable at two time points

$Y_{2-1} = Y_2 - Y_1$

Typical to subtract time 1 from time 2

- **Positive** values indicate an **increase** over time

Difference scores should not be standardized in any way

Difference score models

$$X_1 \xrightarrow{b_1} Y_{2-1}$$

How does X (at time 1) predict change in Y?

$\hat{Y}_{2-1} = b_0 + b_1 X_1$

Difference score models

$$X_1 \xrightarrow{b_1} Y_{2-1}$$

b_0 is the intercept

- Expected value of Y_{2-1} when X_1 equals 0
- Expected **change** in Y when X_1 equals 0

b_1 is the slope

- Increase in Y_{2-1} for a 1 unit increase in X_1
- Expected change in **change** in Y for a 1 unit increase in X_1

Fixed regression analysis of change

Fixed regression analysis of change

An expansion on the basic difference score model – Allison (2005)

Use the difference score for the predictor too

$\hat{Y}_{2-1} = b_0 + b_1 X_{2-1}$

Does **change** in X predict **change** in Y?

Later in the semester, we'll talk about "time varying predictors"

- This is the simplest version of time varying predictors

Fixed regression analysis of change

$$X_{2-1} \xrightarrow{b_1} Y_{2-1}$$

How does **change** in X from time 1 to 2 predict **change** in Y from time 1 to 2?

$\hat{Y}_{2-1} = b_0 + b_1 X_{2-1}$

Fixed regression analysis of change

$$X_{2-1} \xrightarrow{b_1} Y_{2-1}$$

b_0 is the intercept

- Expected value of Y_{2-1} when X_{2-1} equals 0
- Expected change in Y for someone who doesn't change on X

b_1 is the slope

- Increase in Y_{2-1} for a 1 unit change in X_{2-1}
- Expected change in **change** in Y for a 1 unit change in **change** in X

Summary

Summary of difference score models

Difference score regression

Difference score as outcome

$\hat{Y}_{2-1} = b_0 + b_1 X_1$

Fixed regression analysis of change

Difference score as outcome

Difference score as predictor

$\hat{Y}_{2-1} = b_0 + b_1 X_{2-1}$

Partial change scores

Correlation and partial correlation

Correlations with difference scores

The idea behind difference scores is that you remove the influence of the individual timepoints and examine only change

The problem: Y_1 is always correlated with Y_{2-1}

- Difference score doesn't remove all the influence of time 1

The question: How to get a difference score that **removes all the influence of time 1**?

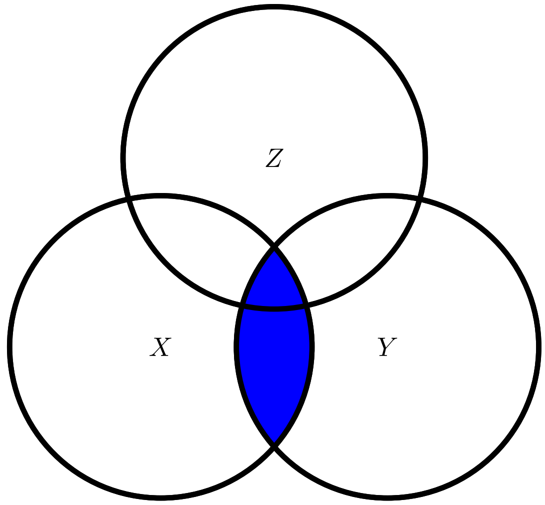
The answer: Look more carefully at **correlations**, specifically **partial correlations**

Correlation

Degree of association between two variables

Proportion of overlap

The correlation between X and Y

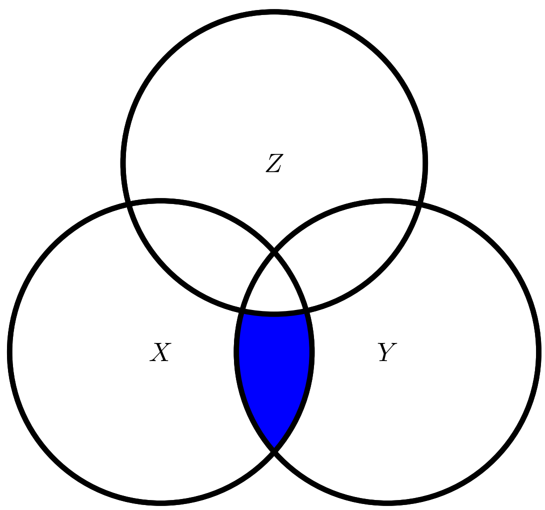


Partial correlation

Degree of association between two variables, with the effect of other variables **removed**

Shaded section is the relationship between X and Y, partially out the effect of Z

(Like a partial regression coefficient)



Partial change score

Partial change score

Use partial correlations to create a difference score that's uncorrelated with time 1

- Specifically, we will use regressions to get partial regression coefficients

Also called "regressed change score"

Cohen, Cohen, West, & Aiken (2003), page 570

3 steps

1. How much does Y_1 relate to Y_2?
2. Remove that part from Y_2

3. Use that new variable as the outcome

Step 1: Regress Y_2 on Y_1

$\hat{Y}_2 = b_0 + b_{Y_2Y_1} Y_1$

$b_{Y_2Y_1}$ is the relationship between Y_1 and Y_2

- Typically, this is **positive** and **relatively large**

Step 2: Remove the part of Y_1 that's related to Y_2 from Y_2

$b_{Y_2Y_1}$ reflects the portion of Y_1 that's related to Y_2

Subtract $b_{Y_2Y_1} \times Y_1$ from Y_2

New quantity:

$Y_2 - (b_{Y_2Y_1} \times Y_1)$

- Y_2 with all time 1 influence removed

[Note: Imagine that we used 1 instead of $b_{Y_2Y_1}$ here. That would be the same as the difference score. More on that later.]

Step 3: Use $Y_2 - (b_{Y_2Y_1} \times Y_1)$ as your outcome

The new outcome variable is $Y_2 - (b_{Y_2Y_1} \times Y_1)$

- I'll call it $newY$

Regress $newY$ on predictor(s):

$ne\hat{w}Y = b_0 + b_1 X_1$

b_1 is the effect of X **on the part of Y_2 that's unrelated to Y_1**

- The part of Y that has **changed** from time 1 to time 2
- b_1 is the relationship between X and the **change** in Y

Summary of partial change scores

Literally remove the part of Y_2 that is related to Y_1

- Then predict what's left over

The part of Y_2 that isn't related to Y_1 is the part of Y that **changes**

- You're predicting change in Y

Analysis of covariance

ANCOVA and covariates

Analysis of covariance (ANCOVA)

Strictly speaking, ANCOVA refers to:

- A relationship between two variables
- A third variable is included as a **covariate**
- Covariate is related to the outcome but not the predictor(s)

There are a number of assumptions of the ANCOVA model

(Note that I'm going to talk about this as a *regression model* rather than the traditional ANOVA approach, but they're the same thing because ANOVA is just a special type of regression)

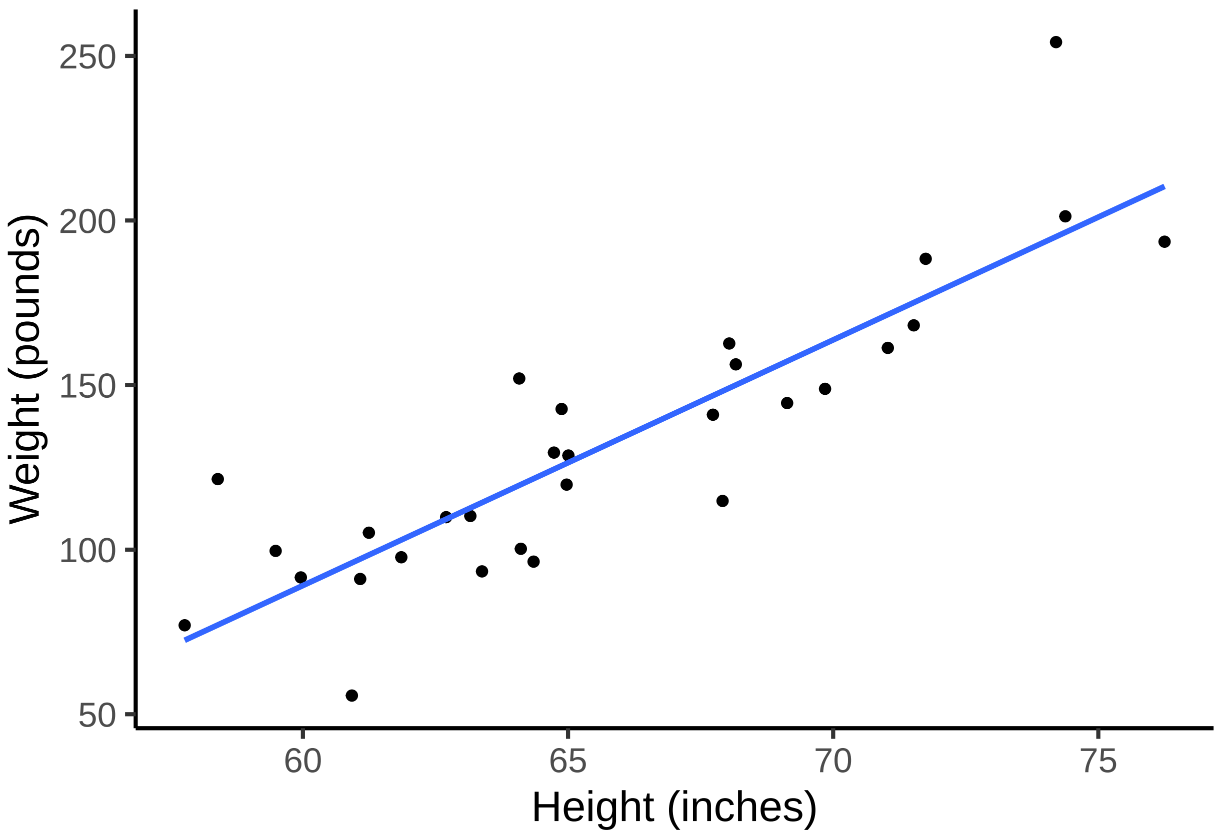
ANCOVA model

$\hat{Y} = b_0 + b_1 X + b_2 C$

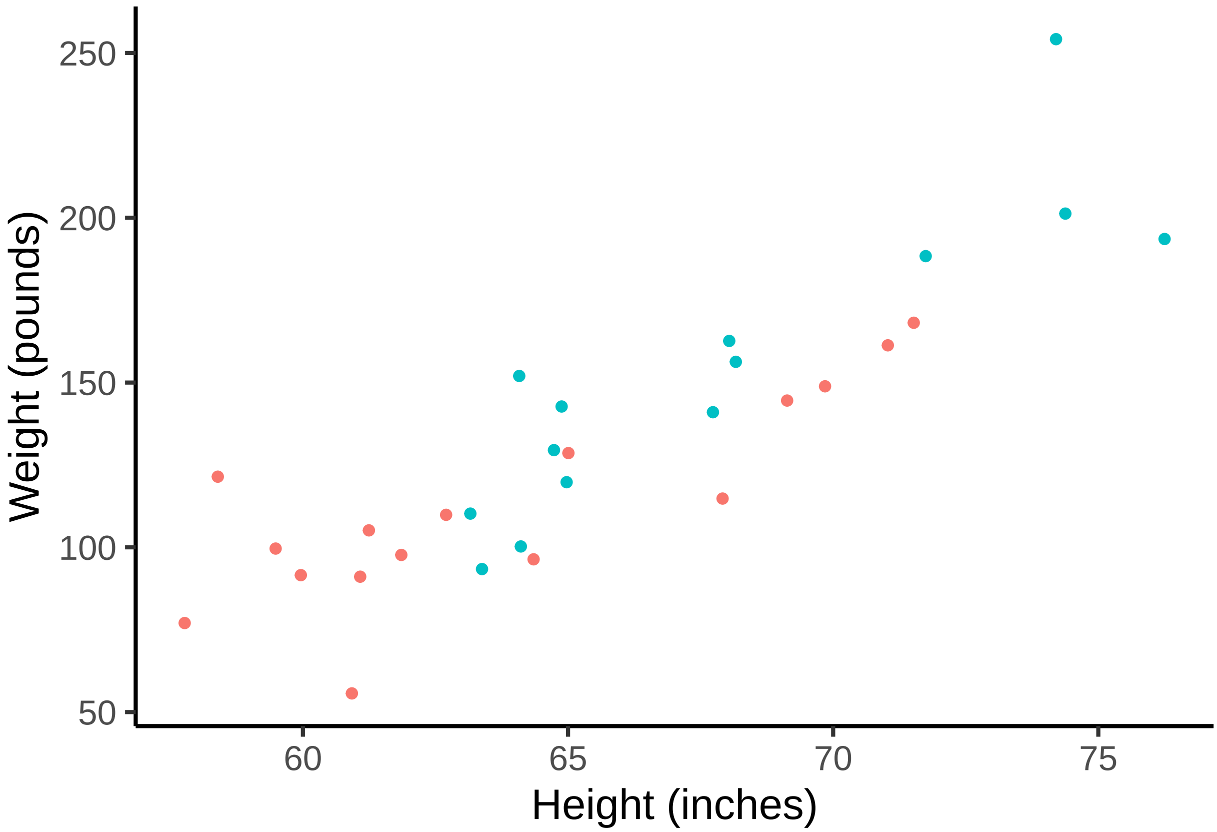
We are interested in the relationship between X and Y

But we are concerned that the relationship between C and Y could distort the $X - Y$ relationship

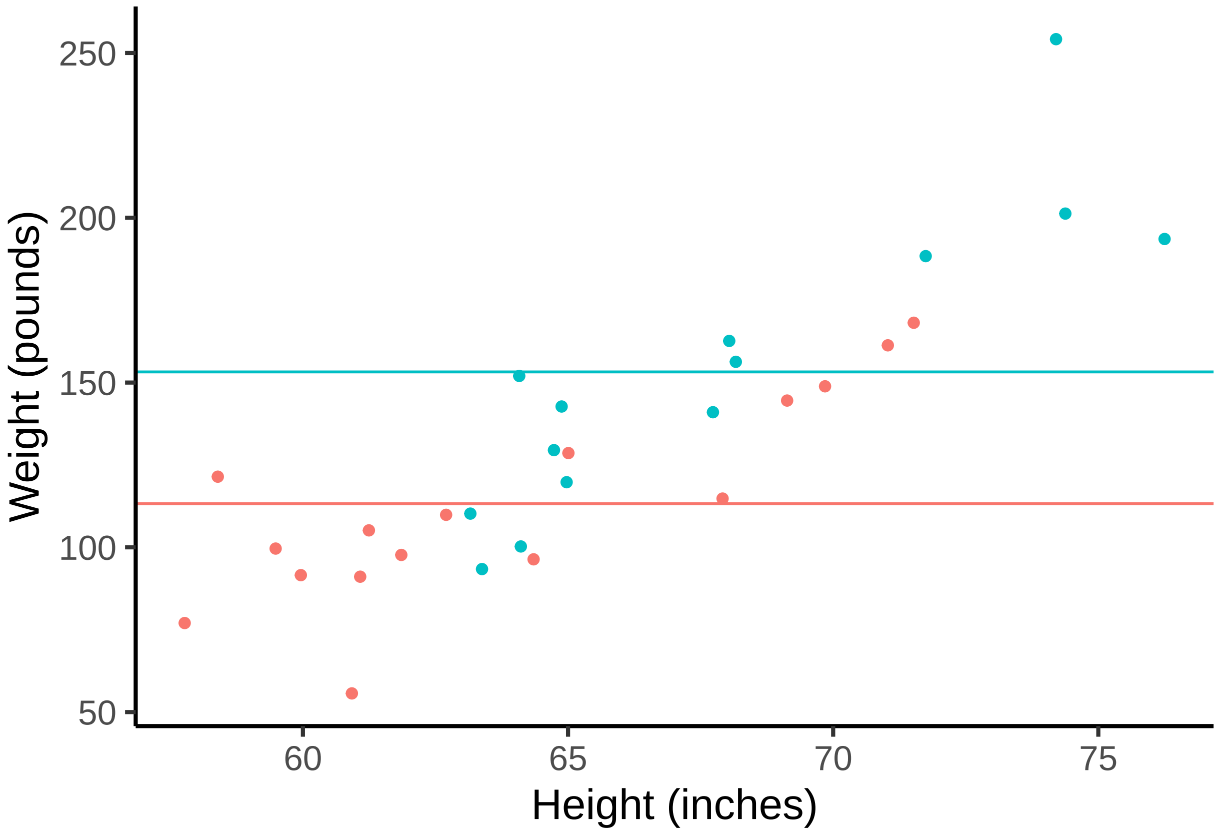
What ANCOVA looks like



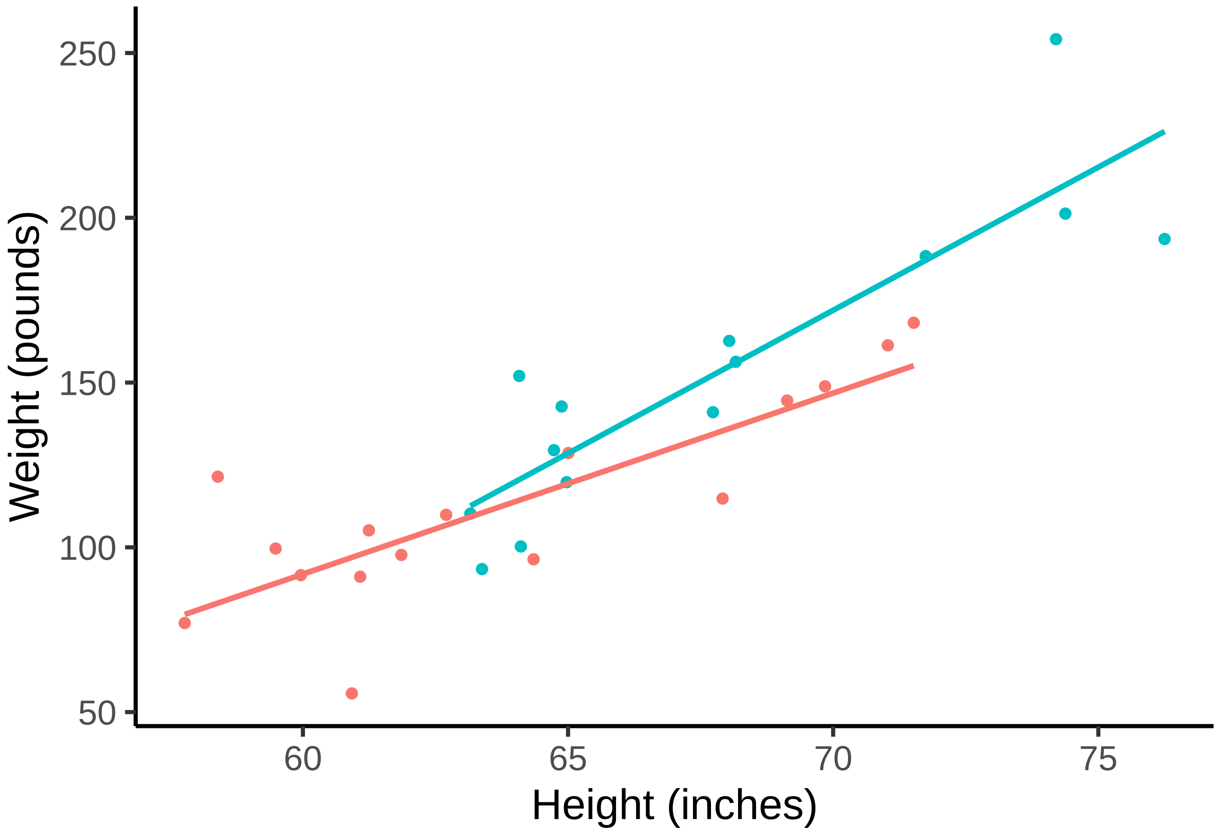
What ANCOVA looks like



What ANCOVA looks like



What ANCOVA looks like



Assumptions

Assumptions of ANCOVA

Under classic ANCOVA assumptions include:

- Homogeneity of variance
  - Same residual variance for each group
- Homogeneity of slopes
  - No interaction between predictor and covariate
- Covariate related to outcome but not predictor

In this example, the slopes for males and females were not exactly the same and the covariate (gender) was related to both height and weight (not just the outcome)

ANCOVA for longitudinal data

Analysis of covariance, more generally

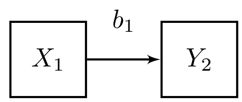
When used with multiple waves of data, these more general ANCOVA-type modelse are also called lagged regression, autoregression, baseline adjusted regression

The covariate is a **baseline or time 1** measure of the outcome

When you include a covariate, you are saying that the effects of other variables are "holding the covariate equal"

- In other words, you're equating people on the baseline measure
- This model says, if everyone were the same at time 1, what is the relationship between X and Y?

Prospective model

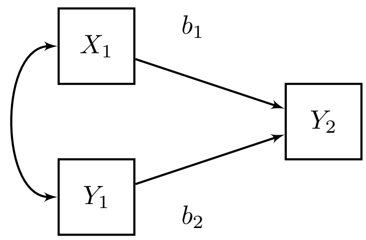


$\hat{Y}_2 = b_0 + b_1 X_1$

How does X at time 1 predict Y at time 2?

No *change* in this model since each variable is only measured once

ANCOVA for two waves



$\hat{Y}_2 = b_0 + b_1 X_1 + b_2 Y_1$

How does X at time 1 predict Y at time 2, controlling for Y at time 1?

ANCOVA for two waves

$\hat{Y}_2 = b_0 + b_1 X_1 + b_2 Y_1$

Two slightly different but **equally valid** ways to think about it

ANCOVA way:

b_1 coefficient reflects the effect of X at time 1 on Y at time 2, assuming that everyone is the same on Y at time 1

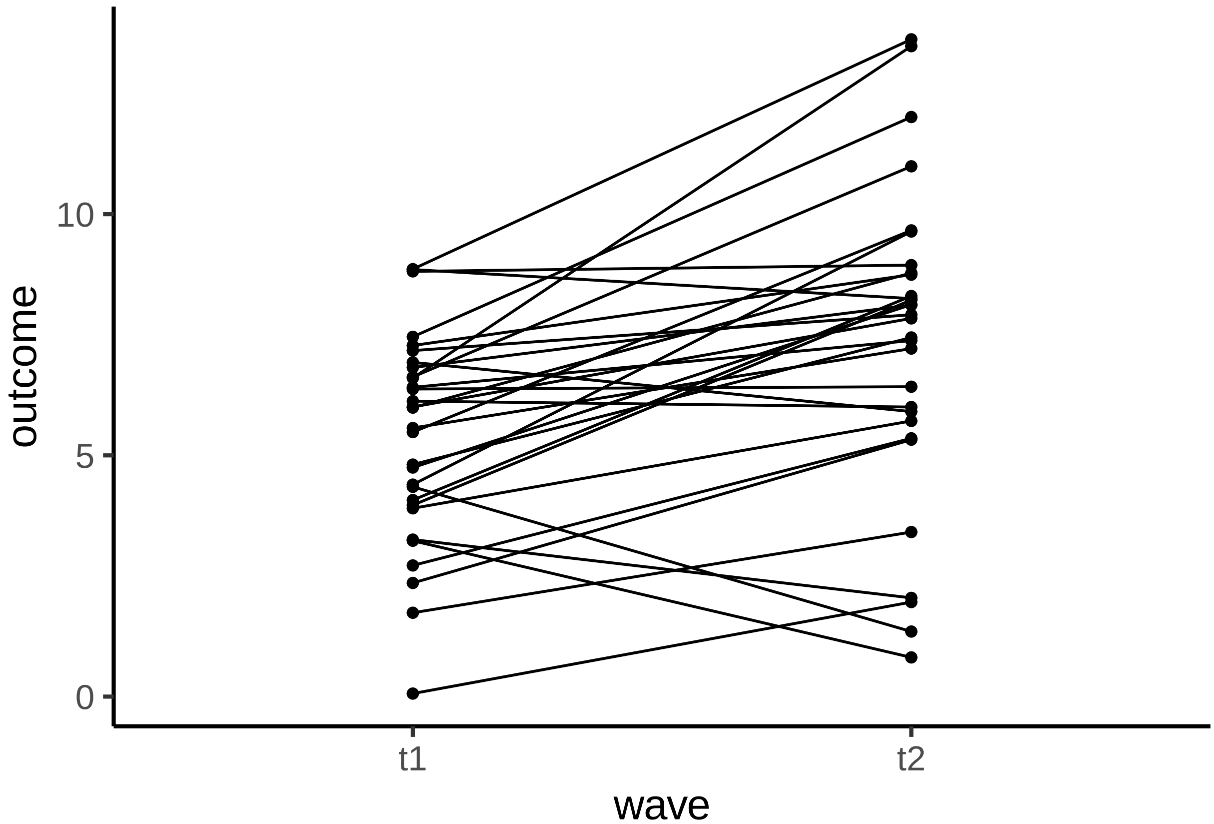
Partial regression coefficient way:

b_2 coefficient reflects the stability of the Y construct over time

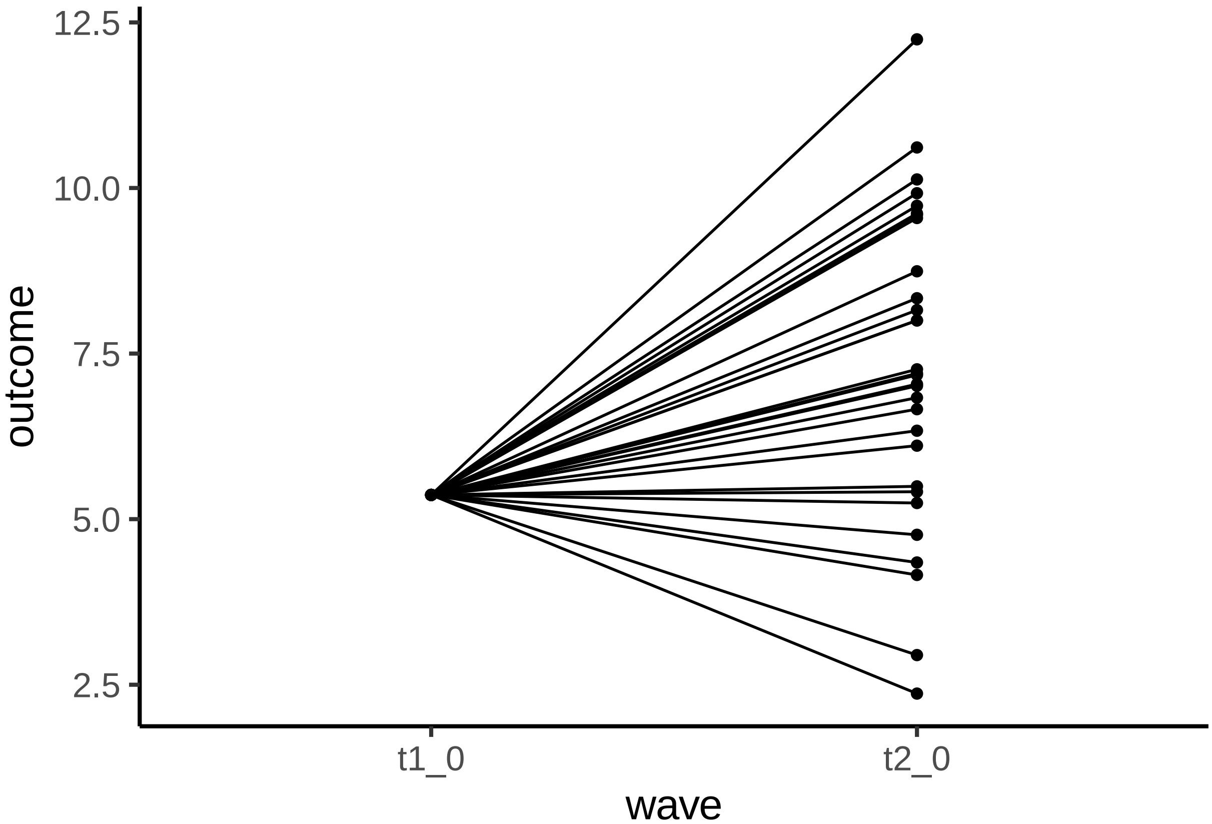
b_1 coefficient reflects what X_1 is able to predict about Y_2, after accounting for that stability

- b_1 coefficient reflects what X can predict about **change** in Y

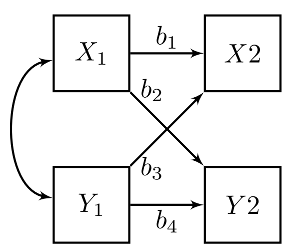
ANCOVA figure



ANCOVA figure



Assessing stability on multiple variables



Comparisons and shortcomings

Does it matter?

Difference, partial change, or ANCOVA?

You have 2 waves of data and want to analyze them

1. Which method should you use?

- Difference score, partial change score, or ANCOVA?

2. Does it matter which one you use?

- Are the conclusions the same or not?

It matters - sometimes

In certain **common** situations, difference score, partial change score, and ANCOVA methods produce different results (i.e., different conclusions)

Lord's Paradox: Lord (1967, 1969, 1973)
Pretest-posttest designs with *non-randomized* groups
Difference score does not show significant change
ANCOVA does show significant change

"... there simply is no logical or statistical procedure that can be counted on to make proper allowances for uncontrolled pre-existing differences between groups."

Really a paradox?

Not really

These methods are asking slightly different questions

- Makes sense that they could produce different results

See Lord (1967), Pike (2004), Tu et al. (2008) for more discussion

Difference and partial change score (UN)reliability

Unreliability of a difference?

Reliability of a variable is a quantitative measure of consistency

Usually, we want reliability of at least 0.8

There is *some evidence* that difference scores have poor reliability (Lord, 1956), even when the scores that go into the difference have good reliability

Situations where this happens are very **common** in practice

Reliability of a difference?

However, there are situations in which difference scores show good reliability (Rogosa et al., 1982 and Rogosa & Willett, 1983)

Situations where this happens are very **uncommon** in practice

What are these methods doing?

Definition of change

The definition of "change" is different for the models

Difference score (and partial change score)

- Absolute amount of change in Y
- i.e., 5 unit change is 5 unit change, regardless of where you started

ANCOVA

- Change in rank ordering
- Specifically, are you higher or lower at time 2, assuming everyone is the same at time 1?

Implied relationship between Y_1 and Y_2

Difference score

- Implies a **perfect** relationship between Y_1 and Y_2
- Not perfect: model overadjusts, leading to NS results

Partial change score

- Estimates the relationship between Y_1 and Y_2
- And **removes** it

ANCOVA

- Estimates the relationship between Y_1 and Y_2
- And **partials** it out

Individual versus relative change

Difference score (and partial change score)

- Individual change
- Each subject has their own "change score"

ANCOVA

- Relative rank
- Relative change, assuming everyone is the same at pre-test

When should you use each method?

What is your research question?

Do you want to rule out the possibility of **pre-test differences** producing your effect?

- Use ANCOVA
  - Difference scores don't account for baseline / time 1 differences
- But keep in mind that ANCOVA is equating everyone at time 1
  - Is that realistic? Does that make sense in your situation?

Is equating realistic?

Often not realistic:

- Non-randomized groups: gender, race, self-selected treatments
  - Especially when the groups do not, by nature, overlap or have similar means on the outcome

Could be realistic:

- Randomized groups that have different pre-test means (by chance)
- Non-randomized groups that have similar means or have been matched

What is your research question?

Are you interested in the **specific, individual amount of change**?

- Use difference scores
  - ANCOVA tells you about relative change
- Partial change scores are a good choice too
  - Individual change but also *remove* time 1 differences

Conclusions

Difference score models work well if there are **few pre-test differences** and you are interested in **individual level of change** from one timepoint to another

ANCOVA models work well if you have **pre-test differences** that make sense to equate across groups
They focus on **relative changes** in rank from one timepoint to another

Partial change scores are kind of in the middle and can help you answer questions that don't quite fit difference score or ANCOVA approaches