

# Growth Curve Modeling Introduction

*Mike Bader*

*June 24, 2016*

Welcome!

This document provides a companion to what I will be discussing in class. You should follow along, but know that not everything will be contained in this document. In addition, this will be my first time using this document, so there might be errors that need correcting in it.

## Course Objectives

As you will note from the [course description][epic-description], I hope that by the end of our time together you will be able to:

- Identify the proper modeling technique for analytical questions regarding change
- Describe how latent growth, latent growth trajectory analysis, and growth mixture models are related
- Interpret results from growth models
- Clearly articulate the benefits, assumptions, and shortcomings of growth trajectory analysis

I hope to cover a lot in five hours, with the primary focus being that you will be able to understand how the family of growth curve models works and to be able to understand the differences between all of the different siblings within the family. I also think it's important to identify what *can't* be covered in five hours. We will not have time to cover model diagnostics in depth, nor will we have any time to talk about estimation procedures. I will try, however, to flag places where common problems arise.

## Structure of Topics

Throughout the class, I will (generally) follow a series of steps as we go through different types of models. We will start off very simple, so simple that you might have covered the basics of what we start with in middle school. Doing this accomplishes several goals. Most importantly, all regression typically breaks down to understanding means and standard deviations. The *structure* of those means and standard deviations becomes more complicated, but ultimately nearly all regression used in practice hopes to find the average effect and the amount of variability around that average.

More practically, starting very simple allows us to have a common language and notation with which to discuss the elements of models. Language sometimes varies from discipline to discipline; in general, I learned mostly through sociology so much of the language I use derives from how sociologists talk about models. The notation that I use generally comes from Raudenbush & Bryk (2002), though I sometimes deviate.

### Step 1: Be Our Own “Data Generating Process”

In the first phase of every problem, we will “play god” by creating data to follow some known process. Doing so will help understand what data following a particular model look like *when all of the assumptions of the model are met*. This helps us discuss the underlying models without worrying about the messiness of real data.

- **Write the model.** We will write out a process that we think describes what would occur. Don't worry if the equation does not immediately make sense; it just provides a formal way to think about what the structure of the data should look like. We will always come back to the equation later.

- **Create data following the model.** We will then create our own data. Yup, that’s right, make it up. The formal statistical term would be to “simulate” the data, but it’s more fun – for the few times in our careers when it is not verboten to do so – to make it up. This step allows us to not get bogged down in all of the messiness of real data with all of its problems. This made-up data represents the cleanest form of data that we can use.
- **Analyze the data with the model.** We will then run the model on our made-up dataset so that we can see how the model recovers the known process that we created in the data. In this phase, we will interpret the results of the model *as if we didn’t know what the model looked like ahead of time*.

## Step 2: Introduce Real Data to Analyze Problems

The second phase introduces real data to analyze so that you can see what data analysis will look like. These data *are not* completely analyzed or cleaned. They simply provide a way to look at the process of analyzing data in which society creates the “data generating processes”.

- **Gather data.** We will “gather” data; by which I mean that we will use data that already exist. In this example, I intentionally use data from Zillow® because it is freely available *and* we don’t get bogged down in debates about the epidemiologic principles of the underlying processes.
- **Analyze data.** We will then structure the dataset to conform to how we need it for analyses and conduct the same types of analyses that we ran on the data that we generated.
- **Interpret the results.** We will then interpret the results. Hopefully in doing so, the model’s equation that we started with will make a little bit more sense.

Without further ado, let’s move on to our first section!

## References

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications.