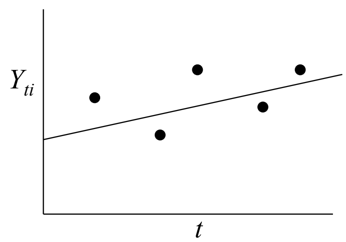
Growth Curve Models

Overview

Growth curve models investigate change over three or more time points. The statistical model is a special case of multilevel models for which multiple measurements are nested within a person. Generally of the concepts we have discussed for models for individuals nested within groups applies, because the way in which the models are estimated and the software implementation is identical to the group-nesting circumstance. There are several special considerations and interpretations for growth curve models, however.

Fundamental Concepts

Growth curve models examine change in the dependent variable as a function of time for each individual. Below is the example of a single growth curve (linear slope) for one individual.



This individual growth curve can be represented by a regression formula, with a dependent variable, Y_{ti}, measured at each time point, t, for each individual, i. The slope, β_{1i}, represents the change in the dependent variable for each unit increase in time, t, The intercept, β_{0i}, represents the value of Y_{ti} when t, (i.e., time) equals zero. If time is coded, 0, 1, 2, …, then the intercept represents the value of Y_{ti} at the beginning of the study.[1]

$$Y_{ti} = \beta_{0i} + \beta_{1i}t + R_{ti}$$

This coding for the time variable does not have to be used, and there are some instances in which the middle (centered), last time point, or other value may be of interest instead. Moreover, coding does not have to be in unit increments. For example, if a year is skipped in an otherwise annual survey, codes would likely follow suit.

Conceptually, intercepts and slopes for each individual can be used as dependent variables in second-level regression equations.

$$\beta_{0i} = \gamma_{00} + U_{0i}$$
$$\beta_{1i} = \gamma_{10} + U_{1i}$$

γ_{00} (gamma) is the average intercept (baseline score). The two levels can be combined into a single multilevel equation.

$$Y_{ti} = \gamma_{00} + \gamma_{10}t + U_{0i} + U_{1i}t + R_{ti}$$

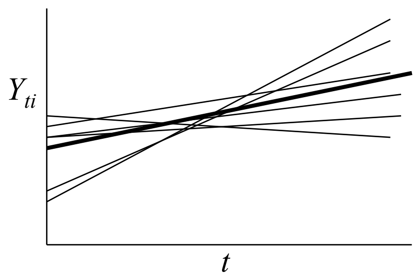
where γ_{10} now becomes the average slope.

The Snijders and Bosker (2012) text uses $(t - t_0)$ to represent the time variable to emphasize that the time variable should start with 0 (if we subtract the participant's age at baseline, for example, every case will

[1] I am following the notation from the Snijders and Bosker (2012) text, but I will discuss the notation used by Singer and Willett (2003), Raudenbush & Bryk (2002), and Hox (2010) soon.

start with 0 at baseline. This creates a more desirable baseline value for the intercept in most instances. I will discuss this coding and others in more detail later.

Random effects can be estimated as in the group-nested model, where, in this context, the random effect for the intercept represents the variance of baseline values (for the 0, 1, 2 … coding) across individuals and variance of the slopes represents individual differences in rates of change over time.



The models can be expanded in a variety of ways, including incorporating level-1 and level-2 covariates, cross-level interactions, and nonlinear effects and they have, therefore, become an invaluable tool for investigating change in longitudinal studies.

References

Hox, J. (2010). *Multilevel analysis: Techniques and applications*. Routledge.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.
Singer & Willett, (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
Snijders, T. A., & Bosker, R. (2012). Multilevel analysis: An introduction to basic and advanced multilevel modeling.