LONGITUDINAL ANALYSIS
A Quick Guide

Huan Bui

Colby College
Physics & Mathemathical Science: Statistics
Class of 2021

February 6, 2019

Preface

Greetings,

Longitudinal Analysis, A Quick Guide to is compiled based on my SC398: Applied Longitudinal Analysis notes with professor Liam O'Brien. The sections are based on *Longitudinal Data Analysis* by Hedeker & Gibbons and professor Liam O'Brien resources.

Enjoy!

Contents

1 Introduction

Longitudinal studies are studies in which subjects are measured repeatedly through time. The defining structure of longitudinal data is repeated observations on individuals, which allows for the study of change. In longitudinal studies, the same response is measured (measurements are commensurate), even though the factor is not necessarily the same. Correlations may arise as a result of sequential measurements. However, sophisticated statistical techniques is required to correct assess these correlations. Correlations among responses must also be taken into account before making inferences.

1.1 Advantages of longitudinal studies

- Longitudinal studies require fewer subjects than cross-sectional studies. Observations from the same subject, while correlated, are not perfectly correlated. So, the net result is that repeated measurements from a single subject provide more independent information than a single measurement obtained from a single subject.

- Each subject can serve as their own control in a longitudinal study. Intra-subject variability is generally smaller than inter-subject variability. Therefore, longitudinal studies generate more efficient estimators of treatment-related effects compared to cross-sectional studies.

- Longitudinal studies allow for separating aging effects (change over time within individuals) from cohort effects (differences between subjects at baseline).

- Longitudinal data can provide information about individual change.

1.2 Challenges of longitudinal studies

- Observations are not independent. Intra-subject dependence must be accounted for.

- Missing data: a subject may be missing one of several measurement occasions. A subject can also remove themselves from the study: "drop-out" or "attrition" are some terms that describe these situations. We can do a "completer analysis" where we only work with the completed data, but we might be missing out on possible correlations between the reasons for stopping and the effects of the study.

- Time-varying covariates: estimators can be time-dependent.

- Cross-over/Carry-over effects: repeated measurements can involve different conditions that the same subjects are exposed to. Responses to one treatment may be conditional on exposure to a previous treatment.

1.3   Two general types of design

- Parallel Design: groups of subjects defined by treatment are followed over time. The objective is to compare the outcome trajectories over time.

- Crossover Design: Subjects are exposed to multiple treatments. The objective is to compare responses of the same subjects to different conditions.

1.4   General Data Structure

Generally, data is described as a matrix, where y_{ij} is the jth observation on the ith subject. There are slight differences in how parallel design and crossover design data is displayed, but the underlying principle is the same.

1.5   Basic types of Longitudinal Models

The most basic types of longitudinal models are those with two-time measures. For example:

- Paired t-test.

- Change scores

- Model response controlling for baseline (ANCOVA)

- Model change score controlling for baseline (ANCOVA)

Note: ANCOVA, or analysis of covariance, is a general linear model which blends ANOVA and regression.