Homework # 3

Biostatistics 210

DUE October 12, 2021

This assignment will give you practice comparing methods for repeated measures data. This is a toy dataset with 1,000 observation two predictors. It can illustrate the effect on a real dataset of making different assumptions about the data covariance structure and the effect this has on estimating covariate effects.

1. DOWNLOAD

Download the dataset hw3 long.dta and hw3 wide.sta from the course website. The variables are

•id (a participant id number — participants are our level of clustering)

• y (a continuous outcome)

• u (a continuous cluster-level outcome)

• w (a covariate varying within cluster)

The long and wide datasets are identical but some analyses are easier with one than the other

2. ANALYSIS

Q1: Read in the wide dataset. Using the pwcorr command looking at the correlation of the outcome (y) variables

```
use "hw3 wide.dta", clear
pwcorr y1 y2 y3 y4 y5
```

These are the values at times 1-5. Are they independent? Does the correlation look exchangeable? What do you note about the correlation structure?

Q2 Read in the long dataset and fit the linear regression model for y in u and w. What is the major problem with a linear regression for this data? What effect will it have on your conclusions?

```
use "hw3 long.dta", clear
reg y u w
```

Q3 Suppose we a linear mixed effects model with a single random effect at the level of each person. What type of correlation structure does such a model assume? Does that agree with the results of pwcorr?

Q4 Fit the linear effects model in Q3. How to the results differ from the linear effects model? Specifically, how have the coefficients changed? How have the standard errors changed? Have they change in the same way for both coefficients?

```
xtmixed y u w || id:,
```

Q5 Type

```
browse id w u if id < 5
```

```
This will give a pop up window allowing you to see the data?
```

What do you notice about the pattern of the covariates within the data? Speculate as to why a repeated measures model has the effects it does for u and w.

Q6 Fit a GEE model to the data.

```
xtgee y u w, i(id) link(identity) vce(robust) corr(indep)
```

How do the results differ? What are the advantages and disadvantages for this data?