Dr. Özlem İLK DAĞ

Fall 2023-2024

**STAT 545- LONGITUDINAL DATA ANALYSIS**

**HOMEWORK 1**

**Due 7 November 2023, Tuesday, 12:40**

**You should work on Part 2 on your own. Please feel free to get help from me, but not from anyone else. Let me know if my wording in the questions is not clear.**

**Your answers should be ≤ 10 pages under a standard format (normal margins, Times new roman font=12 …)**

**I will post the answers to this homework on the same day. Therefore, no late homeworks will be allowed.**

**Part 1.** Reading assignments (do not turn in anything):

* Please read Diggle et al. (2002) Chapters 1 & 3 & [13.1-13.4]
* Please read Weiss, R.E. (2005) Chapters 1 & 2 & [12.1-12.4]

**Part 2.** I simulated a data set, and post it at ODTUClass under the name sim.data.

Consider **Y0** as a baseline measurement and as **a covariate** for this dataset.

1. Is this data in long or wide form?
2. Fill in the blank: The disadvantage of such data form is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.
3. Convert it into the other form and call this new dataset sim.data2. Give the first 5 lines of sim.data2. What is the dimension of sim.data2?
4. Provide the spaghetti plot of response vs time. Interpret your Figure with just 1-2 sentences.

**e)** Provide the trellis display of response vs each of the covariates separately. Be careful/neat about the labels, layout etc. of these displays (i.e., make them look nice). Interpret your Figures with just 1-2 sentences.

**f)** Provide the Draftman’s display. Interpret your Figure with just 1-2 sentences.

**g)** Provide the covariance and correlation matrices for the response at different years. What do you see? Suggest a known variance-covariance structure to be used while modeling this data. Explain why in 1-2 sentences.

**h)** Now, delete some observations to create a dataset with missing data at random. Let the missing proportion be 10%. Provide code or steps of how you created this dataset. Explore if missingness is random. Report your findings. Apply some imputation techniques (it is up to you which ones you apply) that we learned in the class. Investigate if these methods work nicely for this situation.

**i)** Now, delete some observations to create a dataset with missing data in a non-random fashion. Let the missing proportion be 10%. Provide code or steps of how you created this dataset. Explore if missingness is random. Report your findings. Apply some imputation techniques (it is up to you which ones you apply) that we learned in the class. Investigate if these methods work nicely for this situation.