

## Clinical Trials Homework Set 4

1. (60 points) This project relates to the analysis of the data in the file GWU-A1C-Graphs.csv. This is a longitudinal trial that observed levels of HbA1c from patients with type 1 diabetes at various visits including a baseline visit (Visit.number=1). The objective of the trial is to demonstrate that the treatment would lower patients' HbA1c more than the control.

We will evaluate the treatment efficacy based on visits 1-9, so you will need to delete data collected at visit 10 and later visits. Labels of the column variables are given below.

Visit.number: This gives the number for each visit the patient paid to the clinic. 1 is the baseline visit prior to the start of intervention, thus 2 is the first visit after intervention.

HbA1C.level: This is the primary endpoint of the trial. HbA1c is a biomarker for diabetes which is interpreted as the three-month average of blood sugar level in the body. This column gives the level of HbA1c at each visit.

PATID: This gives the unique identification number of each patient. The same number is repeated because of multiple visits.

Control.treatment: Tells you which arm the patients were assigned to.

Age: Age of patients at enrollment to the study.

(1). Randomly select 15 patients from each treatment arm. For each arm, make a spaghetti plot of the HbA1c levels of the 15 patients against their visits. Put the two plots in the same figure with the same scales, like the one on page 18 of the lecture note. Use red color for treatment arm and blue for control arm. What can you comment on the comparison of the two arms?

(2). Again using the HbA1c levels from (1), compute the sample means of HbA1c at each visit for each treatment arm. Make the mean plots by treatment group, like the one on page 19 of the lecture note. Use red color for treatment arm and blue for control arm. What can you comment on the comparison of the two arms?

(3). Repeat the analysis (1) and (2), but this time using all patients. Comment on your findings. Are the findings similar to those from (1) and (2)?

(Hint: You can google search for "How to make spaghetti plot in R or SAS" to find codes to make the figures. Here is one I found in R: <http://www.medicalnerds.com/producing-spaghetti-plots-using-r/>)

2. (40 points) The sample size of 400 patients was based on comparison of the rate of change in HbA1c level between treatment and control. A compound symmetry correlation among repeated HbA1c was assumed with a common correlation coefficient of 0.5. A common standard deviation of 1.2. Suppose everything else remains the same, what would the sample size be, respectively, if the

correlation coefficient is 0.4, 0.45, 0.55, 0.60, 0.65, 0.70 and the standard deviation is 1, 1.1, 1.4, 1.5? Describe the relationship between the correlation and the sample size. (Hint: Use the formula on page 31 of the lecture note).