STAT 614 - HW 5

Due: Thursday, October 29, 2020 in Blackboard by 11:59pm.

Instructions: Please type your solutions in a separate document and upload the document in Blackboard. Include supporting work (plots, etc.) when appropriate, but do not copy all computer output. Select only relevant output. I will not be collecting syntax for this assignment.

Notes:

- For this HW you will need some concepts from chapter 5 on the ANOVA model.
- HW 6 will finish out the ANOVA section.

The effects of exposure to lead on the psychological and neurological well-being of children were studied by Landrigan et al. (1975). Complete raw data for this study are in the data set lead.sav in Blackboard. The data describe a group of children who lived near a lead smelter in El Paso, Texas. Two exposed groups of children were identified who had blood-lead levels $\geq$ 40 μg/ml in 1972 or in 1973. Because neurological and psychological tests were performed in 1973, researchers argued that it would be better to define an exposure group based on blood-lead levels in 1973 only. For this purpose, the variable `lead_typ` in the data file gives three exposure groups:

If `lead_typ` = 1, then the child had normal blood-lead levels (<40 μg/ 100 mL) in both 1972 and 1973 (control group).
If `lead_typ` = 2, then the child had elevated blood-lead levels ($\geq$40 μg/100 mL) in 1973 (the currently exposed group).
If `lead_typ` = 3, then the child had elevated blood-lead levels in 1972 and normal blood-lead levels in 1973 (the previously exposed group).

One important measure of neurological function studied was `MAXFT` = the number of finger-wrist taps in the dominant hand. Researchers are interested in whether there is evidence of differences in neurological function, as measured by `MAXFT`, on average, between the three exposure populations. They would also like to test and estimate the average difference in `MAXFT` between each pair of exposure populations, with the expectation that populations with normal blood-lead levels will have higher average `MAXFT` scores. It is unclear if previously exposed populations will have "recovered" any function as compared to a currently exposed population. Address these research questions by answering the following questions.

1. State the hypotheses of interest to be tested. Include the overall test of group differences in addition to all possible pairwise comparisons of interest.
2. Write the ANOVA model to be fit.
3. Conduct a brief exploratory analysis of the `MAXFT` variable by exposure group (`lead_typ`). Give supporting graphs, descriptive statistics, and interpret these results.
4. What are the assumptions of the model (and corresponding hypothesis tests)? Based on the exploratory analysis in (3), are the assumptions reasonably met for this data? If not, what adjustments should you make in your analysis? (You don't need to use residuals here – you'll do that on HW 6 and then for the rest of the semester!!!)
5. Conduct the appropriate analysis (i.e. incorporate any recommended adjustments from (d) if you had them). Clearly and briefly state the conclusions of your analysis. Be sure you address the researcher's questions.