BS805 Fall 2022 Week 4

Be sure to follow the *Assessment Guideline 1: Writing up Homework* at the end of the syllabus in preparing the homework for submission.

Homework assignments need to be uploaded to the blackboard website by 2 PM on the due date.

In each homework report, be sure to include an introductory and a summary paragraph. Also, include the relevant parts of your SAS code where appropriate in your answer for each question.

A randomized, double blind, clinical trial of a new device for rheumatoid arthritis patients was performed. Three weekly evaluations for tender joint count were performed on all subjects before treatment was administered. Patients were then randomized to either the device or a placebo treatment. One week after the treatment was administered, another evaluation (the fourth) of tender joint count was performed. The ***outcome variable*** for a subject is to be the average of the three tender joint count evaluations done before treatment subtracted from the tender joint count after treatment.

In addition to the treatment, investigators feel that the duration of disease could be an important factor in the outcome of the trial. They feel that this effect will be captured by using 3 categories of disease duration: 1 - disease duration less than or equal to 2 years; 2 - disease duration of more than 2 years, but less than or equal to 5 years; 3 - disease duration of more than 5 years.

The data for this trial are contained in 5 Excel data files: *trial01_f22.xlsx*, *trial02_f22.xlsx*, *trial03_f22.xlsx*, *trial04_f22.xlsx*, and *trial05_f22.xlsx* in the **Course Documents/Data Sets/Arthritis Clinical Trial Data Files** section of the BS 805 web site.

The files *trial01* to *trial04* have the same structure and column names. The *trial01* file contains values from the first evaluation. The *trial02* file contains values from the second evaluation. The *trial03* file contains values from the third evaluation. The *trial04* file contains values from the fourth evaluation.

The file *trial05* contains the following information:

1. Subject ID number
2. Treatment code (0 = placebo, 1 = treatment)
3. Disease Duration code (see code above).

A) Read in the files *trial01*, *trial02*, and *trial03* successfully. Then create a new temporary SAS data set by **appending (that is, stacking vertically)** the three data sets together so that there are multiple observations per ID. Using the appended SAS data set, create another temporary SAS data set (call it, STATS1) that has one observation per subject and contains variables for the subject ID number and the mean tender joint count for each subject over the first 3 evaluations (hint: use the MEANS procedure with an output statement).

B) Read in the file *trial04* and create a temporary SAS data set. **Merge (that is, join by ID)** this data set with the data set, STATS1, created in part A, and create another temporary SAS data set and call it, TRIALS1_4. In the TRIALS1_4 SAS data set, create a variable that gives the **difference** between the

average of the first three evaluations and the fourth (after treatment) evaluation. Create this difference variable so that a **positive value denotes an increase** and a **negative value denotes a decrease**. This is the ***outcome variable*** noted above in the first paragraph. Generate a report of the mean and standard deviation for this difference in tender joint count for the data set, TRIALS1_4.

C) Read in the file *trial05* and create a temporary SAS data set. **Merge** this data set with the data set, TRIALS1_4, and call the new temporary SAS data set, TRIALS1_5. Use the TRIALS1_5 data set to make a means plot of the differences from baseline by treatment and duration. Use the MEANS procedure to generate means of the differences from baseline for the six treatment and duration combinations, save the means to a temporary SAS data set (call it DIFMNS) using the output statement, and then use the GPLOT procedure to plot the means on the vertical axis and the drug code on the horizontal axis (see the example program below). Draw in a line using the PROC GPLOT code shown between the two means for each duration of disease (the SYMBOLn statement does this, with the 'c' standing for COLOR and 'i' being short-hand for INTERPOL or "interpolate"). Be sure that this plot is identified in the output and that the lines are correctly drawn on the plot. For the report on this question, provide an interpretation of what this plot suggests about effects of treatment and duration on the mean of the outcome. You can use the SAS code provided below to create this plot. For example,

```
proc means data=differ;
     by treat durat;
     var diff;
     output out=difmns mean=diffmn;
run;

proc gplot data=difmns;
     symbol1 i=join c=black line=1;
     symbol2 i=join c=black line=2;
     symbol3 i=join c=black line=3;
     ** will plot 3 lines, one for each duration **;
     plot diffmn*treat=durat;
     title 'Part C - Plot of Mean Differences';
run;
```

D) Perform a two factor ANOVA for the ***outcome variable*** (i.e., the difference between the average of the first 3 evaluations and the fourth (after treatment) evaluation) with **treatment** and **disease duration code** as the factors (be sure to check interaction before presenting results for main effects; use an alpha of 0.05 to test the interaction and to denote statistical significance for all other relevant hypothesis tests as well). Report the ANOVA results. Be sure to present the appropriate means and measures of dispersion for the outcome for the **treatment** and ***disease duration*** factors. Adjust the significance levels for multiple comparisons if required. If it is helpful, you should refer to the plot from Part C, in explaining the ANOVA results. Did the treatment result in significantly better improvement in tender joint counts than the placebo group?