Graduate Student Project: ODS – Output Delivery System

1. Given the following DATA step and data observations:

```
DATA one;
INPUT y x ;
DATALINES;
    3.1     5.5
    2.3     4.8
    3.0     4.7
    1.9     3.9
    2.5     4.5
    3.7     6.2
    3.4     6.0
    2.6     5.2
    2.8     4.7
    1.6     4.3
    2.0     4.9
    2.9     5.4
    2.3     5.0
    3.2     6.3
    1.8     4.6
    1.4     4.3
    2.0     5.0
    3.8     5.9
    2.2     4.1
    1.5     4.7
```

a.   Perform a simple linear regression analysis on the above data where y is the
        dependent variable and x is the independent variable. In your analysis include 98%
        confidence intervals for the parameter estimates and for the mean response. Recover
        the output as a pdf file. Name this file *lastname*ODS1.pdf where *lastname* is your
        surname. In this file include only the following information: a table containing the
        98% confidence intervals for the parameters and a table containing the values of the
        independent variable, dependent variable, predicted values, 98% CI for the mean
        response and the associated standard error, and the residuals.

b.   In the regression analysis of problem part a, include a ODS FitPlot of the dependent
        variable by the independent variable with the line of best fit drawn. Name the
        program that does all of problems 1a and 1b *lastname*ODS1.*sas*.

2. Obtain the SAS data set ODSHW2 from the Data Sets module in the Online Classroom. This is a permanent SAS data set that contains 2400 observations. The variables in the SAS data set are called: iteration, group, sample, response. Name the program *lastname*ODS2.*sas*.

   a. Conduct an analysis of variance using GLM for each value of iteration. Group is the *treatment* variable in the ANOVA. Sample counts the number of replications of each group. Compute the Group means for each iteration in this procedure also. Since there are 100 iterations, there will be 100 different analyses. Using ODS suppress all of the output from GLM. Using ODS recover the ANOVA tables (specifically the ANOVA F-statistics and p-values) in a temporary SAS data set.

   b. Form a SAS data set containing the sample means requested in part a for each of the groups across all 100 iterations. (Compute the means in the same GLM procedure used for parts a.) Then print the table of sample means in an HTML file, *lastname*ODS2b.HTML.

   c. **Bonus Question:** Count the number of times the ANOVA null hypothesis is rejected, that is, count the number of times the p-value is below α. Use $\alpha = 0.05$. (For those students with more statistical background, this is the empirical power, $0 < \text{power} < 1$, of this ANOVA test.) Use the FREQ procedure to count the number of times the null hypothesis is rejected. You will need to create a new variable indicatimg whether the null hypothesis was rejected or not. Do this two different ways: Use
   i. an indicator function and the FREQ procedure count the number of times the p-value is below α.
   ii. IF-THEN(-ELSE) statement(s) and the FREQ procedure count the number of times the p-value is below α.
   Print these results (2 frequency tables) in an HTML file called *lastname*ODS2c.HTML.

Submit in the Dropbox of the Online Classroom the two programs, the corresponding logs, outputs, and remarks in a single .rtf or pdf file. Submit the Problem 1 pdf file and Problem 2 html file that were requested also. You should have a total of 5 files (two SAS programs annotated with comments and titles; 6 files if you did the bonus question: 1 pdf, 1 or 2 html, 1 rtf or pdf for remarks and logs) submitted to the Dropbox. Please note the due date to the Dropbox folder in the Online Classroom.