

The objective of this assignment is to run a regression analysis. We have not covered SAS regression output yet, so we will fit an initial program today and embellish this program next week. Therefore, make sure you save a copy of this program.

1) First, modify a previous program or create a new program to do the following.

a) Input a dataset (ex0723.csv). This dataset is measured intervals between, and duration of, eruptions of the Old Faithful Geyser. There are 3 variables (**DATE INTERVAL DURATION**). The data was taken over 8 days and the “DATE” is simply day number 1 through 8. We will not use this variable, but include it in your dataset.

b) Arrange for the usual comments and title, and include an HTML output dataset to the C:\temp\ directory.

c) Do a scatter plot of the data using statements similar to the following:

```
proc plot data=geyser;  
    plot duration * interval;  
run;
```

You may want to put a statement line “**options ps=51 ls=111;**” before the PROC PLOT to keep the plot from getting too large.

d) Run a regression analysis using statements like the following:

```
Title2 'Regression with labels and simple output';  
proc reg data=geyser;  
    model duration = interval;  
run;
```

When this file is complete email me a copy and make sure to save a copy. Next week we will add a number of elements to this program and we will consider some of the conclusions that can be drawn from this type of analysis.