

# STAT 480 Assignment #5

(Due Aug 7th, Friday, 11:59 PM)

**Instructions:** Read lectures 8 and 9 before you start your homework. In this homework, you need to use two SAS datasets *mining.sas7bat* and *alcohol.sas7bat*, both of which can be found on Angel under folder *Data*. You need to submit a .sas file on Angel as usual. In addition, you also need submit a .word file to answer questions 1(d), 1(e) and 2(b).

**1. (Data set options and summarizing continuous data)** You need to use the SAS dataset *mining.sas7bat* for this problem. Download the dataset and store it in the following folder:

X:/STAT480/SAS library/MyLibrary2.

This dataset contains four variables: *YEAR*, *QTR*, *PRO* and *POW*.

- (a) **Rename** the variables *QTR*, *PRO* and *POW* as *QUARTER*, *PRODUCTION* and *POWER* respectively.
- (b) Print out the *YEAR* in which *PRODUCTION* of the first quarter is greater than 110.
- (c) **Group** all the observations by *YEAR* and find the mean and standard deviation of variable *PRODUCTION* and *POWER* for each year. Save the results to a new dataset and print the new dataset.
- (d) Give a **plot** of *YEAR* vs **mean** value of *PRODUCTION* of each year (*YEAR* is used as X-axis). Include this plot in the .word file.
- (e) Based on the output above, answer the following questions:
  1. How many observations are there in the dataset *mining*?
  2. In which year(s) the *PRODUCTION* of the first quarter is greater than 110?
  3. Find the mean value and standard deviation for variable *PRODUCTION* in 1990.
  4. Find the mean value and standard deviation for variable *POWER* of all observations.

**2. (Summarizing categorical data)** In this problem, you need to use the SAS dataset *alcohol.sas7bat* to answer the following questions. Download the dataset and store it in the same folder (X:/STAT480/SAS library/MyLibrary2). Dataset *alcohol.sas7bat* contains four variables *ADULTS*, *KIDS*, *INCOME* and *CONSUME*. *ADULTS* is the number of adults in a family, *KIDS* is the number of children in a family and *INCOME* is the income of a family. *CONSUME* = 1 if the family consumes alcohol, otherwise *CONSUME* = 0.

- (a) Give the two-way contingency table of variables *KIDS* and *CONSUME*.
- (b) Based on the output above, answer the following questions:
  1. How many observations are there in this dataset?
  2. What is the number of family that has at least 3 kids?
  3. What is the percentage of family that consumes alcohol?
  4. Among the family that has 4 kids, what is the percentage of family that consumes alcohol?