Homework 3

Problem 1. The file Utility.dat contains a monthly record of telephone, electricity, and fuel costs for several years. Naming the 5 columns as Month ,Year, Telephone, Electricity and Fuel, respectively.

(a) Import data set into SAS using the INFILE statement. Add the title 'Descriptive Statistics for Utilities' to the dataset and print out the **first ten observation** of the dataset with **the title**.

(b) Report **only** the three descriptive statistics mean, s.d. and t-test for being zero for variable Telephone and Fuel. Use PROC Univariate. Please provide **only relevant information** from your SAS output, rather than copying everything from your output window without screening.

Problem 2. The file China.dat contains export and import information (in dollars) by year. Naming four columns as Years Total, Exports and Imports.

(a)Import data set into SAS using the INFILE statement. Create a new variable Trade Balance, where Trade_Balance = Exports - Imports. Print the first ten observation of the whole dataset, including the new variable.

(b) Show these three histograms for Imports, Exports, and Trade_Balance, using PROC UNIVARIATE.

midpoints=0:40 by 4

(c) Histogram, by definition, is a type of bar plot that group whole data into bins. In the histogram of Exports from (b), there are 6 bins by default, and each bin with the same size of roughly **8 units**, in terms of range of Exports. You may check the fact from (b). Now we try to produce a finer description of the data, in another word, more bins are need in the histogram of Export. Please draw a new histogram of Export, each bin with size of 4 units. (hint: histogram … midpoints=…by.)

Problem 3. The file Handinj.dat contains the costs (in Irish pounds) and lost work days due to hand injuries for workers in Dublin, Ireland. Naming the four columns as ID,Type, Days and Cost, respectively;

(a) Import the dataset using the INFILE statement.

(b) Create a scatter plot of Days and Cost using PROC PLOT, with Days on y-axis and Cost on x-axis. Then please label each point in the scatter plot by values of Type, i.e. given an observation of Days and Cost, a point in the 2-dimentional scatter plot is fixed. Then assign 'w' to the point if the observation of Type is 'work' and 's' if 'sport'. Based on the scatter plot, is it possible to seperate these two classes(Type = 'work' and Type = 'sport') by a linear function on the plane. If true, draw the line on the plot, otherwise, venture a guess what function could separate them? Here, a line seperates two class of points is defined as one class of points sitting on one side of a line and the other class on the other side.

Problme 4. The file Debate.dat contains the survey responses from debate students at various schools. Naming these variables as

SurveyNumber, School, Gender, Comparison, Argumentation, Research, Reasoning, and Speaking.

(a) Write a SAS program to read the above data set into SAS using the INFILE statement.

(b) Generate a two way table (with PROC FREQ) for Gender and Research when School=7. And compute the chi-square statistic to test independence of Gender and Research. What is your conclusion?