**Week 1 Assignment: ITMD 529**

**Variable Selection:**

|  |  |  |
| --- | --- | --- |
| **Predictor(X)** | **Variable** | **Type** |
| **votes** | **continuous** |
| **fraction\_votes** | **continuous** |
| **Response(Y)** | **party** | **categorical** |

**The SAS procedures we are planning to use are**

**Proc means**

**Proc freq**

**Proc univariate**

**Proc sgplot,…**

**SAS Code**

**Sample Data selection from the dataset:**

/\*Select sample data from dataset\*/

/\*Sample data\*/

data election.data\_set;

set election.data;

keep state party candidate votes fraction\_votes county;

where votes>=10000;

run;

**Exploratory Data Analysis**

/\*Frequency Plot to understand data distribution\*/

/\*Exploration of all variables that are available for analysis.\*/

/\*%let statements define macro variables containing lists of continuous variables\*/

%let tfilename=election.data\_set;

%let interval= votes fraction\_votes;

/\*UNIVARIATE proc is used to plot histogram, probability graph and to display basic statistics\*/

proc univariate data=&tfilename;

var &interval;

histogram &interval / normal kernel;

inset n mean median std skewness kurtosis / position=ne;

probplot &interval / normal (mu=est sigma=est);

inset skewness kurtosis;

title 'Descriptive Statistics Using PROC UNIVARIATE';

run;

**Exploration of all categorical variables that are available for analysis**

/\*Exploration of all categorical variables that are available for analysis.\*/

%let tfilename=election.data\_set;

%let categorical= party candidate state county;

proc means data=&tfilename maxdec=2 fw=10 printalltypes n mean median std var

q1 q3;

class &categorical;

var fraction\_votes;/\* U can check fr votes also\*/

output out=means mean=votes;

title 'Selected Descriptive Statistics for number of votes';

run;

title;

/\*proc FREQ displays frequency graph of categorical variables\*/

proc freq data=&tfilename;

tables &categorical / plots=freqplot;

title "Categorical Variable Frequency Analysis";

run;

title;

**Association between categorical response and continuous predictores using SGPLOT (VBOX**

/\* Association between categorical response and continuous predictors using SGPLOT (VBOX) \*/

proc sgplot data=election.data\_set;

vbox fraction\_votes/ category=party connect=mean;

run;

proc sgplot data=election.data\_set;

vbox votes/ category=party connect=mean;

run;