**Analysis of Bike Sharing Data**

**SAS Code**

Firstly, source file day.csv is downloaded from [Bike-Sharing-Dataset.zip](https://archive.ics.uci.edu/ml/machine-learning-databases/00275/Bike-Sharing-Dataset.zip) and imported in SAS with Bike\_Sharing\_Data name.

**Sample data selection :**

%let path=/folders/myfolders/Bike\_Sharing;

libname BIKES "&path";

/\*Sample data selection year wise\*/

data BIKES.Bike\_Sharing\_2011;

set BIKES.Bike\_Sharing\_Data(where=(Yr=0));

keep Season Yr holiday workingday weathersit temp hum windspeed cnt;

run;

data BIKES.Bike\_Sharing\_2012;

set BIKES.Bike\_Sharing\_Data(where=(Yr=1));

keep Season Yr holiday workingday weathersit temp hum windspeed cnt;

run;

**Exploratory Data Analysis :**

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

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

%let tfilename=BIKES.Bike\_Sharing\_2011;

%let interval= temp hum windspeed;

/\*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.\*/

%let categorical= weathersit holiday Season;

%let tfilename=BIKES.Bike\_Sharing\_2011;

proc means data=&tfilename maxdec=2 fw=10 printalltypes

n mean median std var q1 q3;

class &categorical;

var cnt;

output out=means mean=cnt;

title 'Selected Descriptive Statistics for bike rental count';

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 predictors and continuous response using SGPLOT (VBOX) :**

/\*PROC SGPLOT is used here with the VBOX statement to produce vertical bar charts\*/

/\*PROC SGPLOT can only produce one plot at a time and so the macro is written to\*/

/\*produce one plot for each member in the list of the &categorical macro variable.\*/

%let categorical= weathersit holiday Season;

%let tfilename=BIKES.Bike\_Sharing\_2011;

ods graphics on/width= 800;

%macro box(dsn = ,

response = ,

Charvar = );

%let i = 1 ;

%do %while(%scan(&charvar,&i,%str( )) ^= %str()) ;

%let var = %scan(&charvar,&i,%str( ));

proc sgplot data=&dsn;

\*panelby season;

vbox &response / category=&var

grouporder=ascending

connect=mean;

title "&response across Levels of &var";

run;

%let i = %eval(&i + 1 ) ;

%end ;

%mend box;

%box(dsn = &tfilename,

response = cnt, /\*temp\*/

charvar = &categorical);

title;

options label;

**Association between continuous predicotrs and response using scatter plot and correlation matrix :**

%let tfilename=BIKES.Bike\_Sharing\_2012;

%let interval= temp hum windspeed;

/\*SGSCATTER displays scatter plot with regression line for continuous predictors and response variable\*/

proc sgscatter data=&tfilename;

plot cnt\*(&interval) / reg;

title "Associations of Bike rental count with predictors";

run;

/\*Correlation between predictors and response \*/

ods graphics / reset=all imagemap;

proc corr data=&tfilename rank

plots(only)=scatter(nvar=all ellipse=none);

var &interval;

with cnt;

title "Correlations and Scatter Plots with count of bike rented";

run;

title;

**Regression modeling of continuous predictors and response:**

%let tfilename=BIKES.Bike\_Sharing\_2011;

/\*Simple regression analysis of temperature vs bike rental count\*/

proc reg data=&tfilename;

model cnt=temp; \* / clm cli;

title "Simple Regression with temperature";

run;

quit;

/\*Simple regression analysis of wind speed vs bike rental count\*/

proc reg data=&tfilename;

model cnt= windspeed; \* / clm cli;

title "Simple Regression with wind speed";

run;

quit;

**Regression modeling of categorical predictors and continuous response:**

%let categorical=Season holiday weathersit;

%let tfilename=BIKES.Bike\_Sharing\_2011;

/\*Individual categorical predictors' analysis using one-way ANOVA\*/

/\*HOVTEST is used to check equality of variances\*/

/\*SOLUTION option is used to display parameter estimates for categorical variables\*/

/\* Analysis of impact of season on bike rental count\*/

ods graphics on / width=700;

proc glm data=&tfilename plots(only)=diagnostics(unpack);

class season ;

model cnt=Season /solution ;

means Season / hovtest;

title "ANOVA with PROC GLM";

run;

quit;

/\* Analysis of impact of holiday on bike rental count\*/

proc glm data=&tfilename plots(only)=diagnostics(unpack);

class holiday ;

model cnt=holiday /solution ;

means holiday / hovtest;

title " ANOVA with PROC GLM";

run;

quit;

/\* Analysis of impact of weather on bike rental count\*/

proc glm data=&tfilename plots(only)=diagnostics(unpack);

class weathersit ;

model cnt=weathersit /solution ;

means weathersit / hovtest;

title "ANOVA with PROC GLM";

run;

quit;

title;

**Regression modeling of categorical/continuous predictors and continuous response:**

%let tfilename=BIKES.Bike\_Sharing\_2011;

/\*Categorical predictors' analysis using n-way ANCOVA with interaction\*/

/\*Analysis of effects of season, temperature and weather together on bike rental count\*/

ods graphics on / width=800;

proc glm data=&tfilename plots(only)=intplot;

class season weathersit;

model cnt=season|weathersit|temp/solution;

title "Analyze the Effects of season, weather and temperature";

title2 "Including Interaction";

run;

quit;

title;