```sas
/******************************************************
   Week 6: High-Quality Statistical Plots - SGPLOT
     Author:  Cheng Peng
     Date:    02/27/2021
  Topics:    1 - Histograms
             2 - Barcharts
             3 - Scatter Plots
             4 - Box-plots
             5 - Simple Linear Regression Lines
             6 - Density Curves
*******************************************************/


DM "CLEAR OUT";
DM "CLEAR LOG";


/***  Permanent library  ***/
LIBNAME sgplot "C:\STA311\w06";   /* permanemt library for saving SAS data
related to PROC SGPLOT  */


/** Global options for ouputs **/

OPTIONS PS = 76 LS = 76 NONUMBER NODATE;


/***********************************
***    Explore car.sas7bdat
***********************************/
TITLE "SAS Built-in Data Set: cars.sas7bdat";   /* global option. Keep in
mind that this title should be
                                                   updated appropriately
according to the new output*/
PROC CONTENTS DATA = sashelp.cars;      /* sashelp is the library reference to
the SAS built-in permanent library */
RUN;                                        /* IMPORTANT: If you want to load a
SAS data set to SAS from a folder, you
                                                should always create a SAS
permanent library pointing to the folder, then
                                                use the library reference and SAS
data or procedure step to access the SAS
                                                format data.
*/


/***********************
** Topic 1: Histogram
***********************/

* Example 1.0 - basic histogram;
PROC SGPLOT DATA = Sashelp.cars;
 HISTOGRAM MSRP;    /* the default histogram is based on the relative
frequency table */
 TITLE "The Simplest Histogram of MSRP";
RUN;

/********************************************************************
Example 1.1. Histogram with more controls by specifying options.
                The syntax is: HISTOGRAM MSRP/<options>
                Available options can be found from
```

```
https://documentation.sas.com/?docsetId=grstatproc&docsetTarget=n17xrpcduau1f
8n1c1nhe477pv18.htm&docsetVersion=9.4&locale=en
******************************************************************/

ODS GRAPHICS ON / WIDTH = 4.5IN HEIGHT = 3.5IN;   /** control the size of the
SAS graphic  **/
PROC SGPLOT DATA = Sashelp.cars;
 HISTOGRAM MSRP/ NBINS = 6          /* 6 vertical bars are requested. */
                 SCALE = count      /* Default scale is percent (relative
frequency) */
                         Y2AXIS              /* draw two vertical axes */
                         DATASKIN = GLOSS  /* effects of vertical bars */
                         ;
 TITLE "Histogram of MSRP with more controls";
RUN;
ODS GRAPHICS OFF;

* Example 1.3 - histogram with a density curve: i.e., stack one curve on the
other;
ODS GRAPHICS ON / WIDTH = 4.5IN HEIGHT = 3.5IN;   /** control the size of the
SAS graphic  **/
PROC SGPLOT DATA = Sashelp.cars;
 HISTOGRAM MSRP;   /* use the default number of vertical bars */
 DENSITY MSRP;     /* by default, a default density curve is normal: mu =
sample mean, sd = sample std
                      you can also choose KERNEL option --> data driven
density estimator.          */
 TITLE "Histogram of MSRP with Density Curve";
RUN;
ODS GRAPHICS OFF;

/***********************
** Topic 2: Bar Charts
************************/

* Example 2.1 Basic Bar Charts;
ODS GRAPHICS ON / WIDTH = 4.5IN HEIGHT = 3.5IN;
PROC SGPLOT DATA = Sashelp.cars;
 VBAR Type;     /* name of character variable to be used to plot the bar
chart */
 TITLE 'Barchart of Type of Vehicles';
RUN;
ODS GRAPHICS OFF;

* Example 2.2 - barchart by the origin ;
ODS GRAPHICS ON / WIDTH = 4.5IN HEIGHT = 3.5IN;
PROC SGPLOT DATA = Sashelp.cars;
 VBAR Type / GROUP = Origin;
 TITLE 'Barchart of Type by the origin';
RUN;
ODS GRAPHICS OFF;

*Example 2.3 - Response option  ;
DATA temp_cars;
SET Sashelp.cars;
counter = 1;   /* add a new variable to the SAS data set with a contant value
1 */
```

```sas
RUN;

/* we use counter as a response variable   */
ODS GRAPHICS ON / WIDTH = 4.5IN HEIGHT = 3.5IN;
PROC SGPLOT DATA = temp_cars;
 VBAR Type / RESPONSE = counter;/* RESPOSE option adds up all values of
variable
                           counter in each category of the categprical
variable Type.
                           The resulting bar chart is the same as the regular
bar-chart */
 TITLE 'Barchart - bar';
RUN;
ODS GRAPHICS OFF;

/***************************************
**  Topic 3:  Scatter plot
***************************************/
* Example 3.1 - scattere plot - basic;
ODS GRAPHICS ON / WIDTH = 4.5IN HEIGHT = 3.5IN;
PROC SGPLOT DATA = Sashelp.cars;
 SCATTER X = MSRP Y = invoice;  /* it desn't matter which variable is X and
which is Y. */
 TITLE 'Scatter plot of MSRP vs Invoices';
RUN;
ODS GRAPHICS OFF;

* Example 3.2 - scatter plot - by groups
  PROC SGPanel - allows making each for each category of the categorical
               variable specified in the PANELBY statement        ;

ODS GRAPHICS ON / WIDTH = 4.5IN HEIGHT = 3.5IN;
PROC SGPANEL DATA = Sashelp.cars;
 PANELBY Type;       /* The  */
 SCATTER X = MPG_Highway Y = MPG_City;
 TITLE 'Line plot of MSRP vs Invoices';
RUN;
ODS GRAPHICS OFF;

* plot matrix: plot    ;
TITLE "Scatter Plot Matrix Several Continuous Variables in CARS";
ODS GRAPHICS ON / WIDTH = 4.5IN HEIGHT = 3.5IN;
PROC SGSCATTER DATA =Sashelp.cars;
MATRIX MSRP INVOICE HORSEPOWER MPG_CITY MPG_HIGHWAY /
                                TRANSPARENCY = 0.6  /* degree of
transparency of the plot */
                                ELLIPSE             /* add a ellipse to
the data points to visualize

the strength of the association between the two

variables.                              */
                                MARKERATTRS =(symbol=circlefilled)
                                DIAGONAL=(histogram normal kernel);
RUN;
ODS GRAPHICS OFF;
```

```
/*************************************
**    Topic 4: Box-plots
*************************************/

/** Example 4.1 Simple box-plot **/
TITLE 'Distribution of Mileage';
ODS GRAPHICS ON / WIDTH =4.5IN HEIGHT=3.5IN;
PROC SGPLOT DATA =sashelp.cars noautolegend;
HBOX mpg_city ;      /* Horizontal box-plot */
YAXIS GRID;  /* DISPLAY is not specified. The default value of DISPLAY =
(all):
                ticks, label and values are all displayed
*/
XAXIS DISPLAY = (nolabel); /* only the x-label is not displayed
*/
RUN;
ODS GRAPHICS OFF;


/** Example 4.2: Box-plot - by type  **/
ODS GRAPHICS ON / WIDTH =4.5IN HEIGHT=3.5IN;
TITLE 'Distribution of Mileage by Type';
PROC SGPLOT DATA =sashelp.cars noautolegend;
HBOX mpg_city / CATEGORY =type;
YAXIS GRID;
XAXIS DISPLAY = (nolabel);
RUN;
ODS GRAPHICS OFF;

/** 4.3. Box-plot with a line plot connecting means  **/
TITLE 'Mileage by Type';
ODS GRAPHICS ON / WIDTH =4.5IN HEIGHT=3.5IN;   /** control the size of the
SAS graphic  **/
PROC SGPLOT DATA =sashelp.cars;
VBOX mpg_city /                 /* vertical box-plot, opposed to horizontal
box-plot with HBOX option */
            CATEGORY =type  /* The categorical to be used to  */
            CONNECT = mean  /* connect the soecified quntities of each
category with a line segment */
            DATALABEL;      /* use observation ID of the data value if
the value is an outlier      */
XAXIS GRID DISPLAY = (noticks nolabel); /* whether ticks, label, values
associated with axis are displayed  */
YAXIS GRID DISPLAY = (novalues nolabel);
RUN;
ODS GRAPHICS OFF;

/** Example 4.4. multiple box-plots on the same plot   **/
PROC SGPLOT DATA =sashelp.cars;
 VBOX enginesize / CATEGORY=type   /* Make a box-plot for each category of
the cahracter variable*/
                BOXWIDTH=0.25   /* box width of the box-plot */
                DISCRETEOFFSET=-0.15; /* the deviation of box-plot
associated with variable enginesize
                                        to the left of the center
of the category                     */
```

```sas
  VBOX horsepower / CATEGORY=type
                    BOXWIDTH=0.25
                    DISCRETEOFFSET=0.15  /* the deviation of box-plot
associated with variable enginesize
                                                  to the right of the center
of the category.
                                                  The distance between the
two box-plots is 0.15-(-0.15) = 0.3 */
  Y2AXIS;              /* make two vertical axes since two numerical variables
are at different scales!   */
RUN;


/*************************************
**   Topic 5: Regression Line
*************************************/

/* Example 5.1. Scatter plot with a regression line  */
TITLE "Scatter Plot with Regression Line";
PROC SGPLOT DATA = sashelp.cars;
REG y = Horsepower x=Weight; /* y = response variable, x = horizontal
variable */
RUN;


/** Example 5.2. Regression with confidence limits **/
TITLE1 "Scatter Plot with Regression Line";
TITLE2 "with Confidence Lmits/Band";
PROC SGPLOT DATA = sashelp.cars;
REG y = Horsepower x=Weight/ CLI   /* confidence limits of individual
predicted values */
                            CLM   /* confidence limit of regression line -
confidence band */
                            alpha=0.1;  /* confidence level = 1 - 0.1 = 90%,
default ALPHA = 0.05  */
RUN;

/** Example 5.3. Regression with confidence limits **/
TITLE1 "Scatter Plot with Regression Line";
TITLE2 "with Confidence Lmits/Band, line/marker attrbutes";
PROC SGPLOT DATA = sashelp.cars;
REG y = Horsepower x=Weight/ CLI       /* confidence limits of individual
predicted values */
                            CLM        /* confidence limit of regression
line - confidence band */
                            alpha=0.1  /* confidence level = 1 - 0.1 = 90%,
default ALPHA = 0.05  */
                            lineattrs=(color=red thickness=5)
                            markerattrs=(color=blue size=10
                            symbol=squarefilled);
RUN;


/*************************************
** Topic 6: Density function
*************************************/
```

```sas
/* Example 6.1. calculate the mean and standard deviation and then use
   the mean and standard deviation to fit a normal distribution.
   The density curve will be placed on the histogram*/

TITLE 'Normal Density for Horsepower';
ODS GRAPHICS ON / WIDTH = 4.5IN HEIGHT = 3.5IN;
PROC SGPLOT DATA=sashelp.cars NOAUTOLEGEND;  /* disables automatic legends
from being generated.  */
HISTOGRAM HORSEPOWER;
DENSITY HORSEPOWER;
YAXIS GRID;
XAXIS DISPLAY = (nolabel);  /* X-axis label is dsabled. */
RUN;
ODS GRAPHICS OFF;

/**Example 6.2: use a different variable: MSRP  **/
TITLE 'Normal Density for MSRP';
ODS GRAPHICS ON / WIDTH = 4.5IN HEIGHT = 3.5IN;
PROC SGPLOT DATA=sashelp.cars NOAUTOLEGEND;  /*  By default, legends are
created automatically for some plots,
                                             depending on their content.
This option has no effect
                                             if you specify a KEYLEGEND
statement. */
HISTOGRAM MSRP;
DENSITY MSRP;
YAXIS GRID;
XAXIS DISPLAY = (nolabel);   /* X-axis label is dsabled. */
RUN;
ODS GRAPHICS OFF;

/** Example 6.3: kernel ensity curve **/
TITLE 'Normal Density for MSRP';
ODS GRAPHICS ON / WIDTH = 4.5IN HEIGHT = 3.5IN;
PROC SGPLOT DATA=sashelp.cars; *NOAUTOLEGEND;
HISTOGRAM MSRP / SCALE = percent;
DENSITY MSRP / SCALE = percent
               TYPE = Kernel;
YAXIS GRID;
XAXIS DISPLAY = (nolabel);
RUN;
ODS GRAPHICS OFF;

/** Example 6.4. Kernel and normal density  **/
TITLE 'Normal/Kernel Density for MSRP';
ODS GRAPHICS ON / WIDTH = 4.5IN HEIGHT = 3.5IN;
PROC SGPLOT DATA=sashelp.cars;
HISTOGRAM MSRP / SCALE = percent;   /* precent = relative frequency, this the
defualt scale in SAS */
DENSITY MSRP;
DENSITY MSRP / TYPE = Kernel;
KEYLEGEND / LOCATION =inside
            POSITION = topright
            ACROSS = 1;
YAXIS GRID DISPLAY = (nolabel);
XAXIS GRID DISPLAY = (nolabel);
RUN;
```

```
ODS GRAPHICS OFF;
```