STAT 5110/6110: SAS Programming and Applications

8-A. SAS Graphics

Peng Zeng

Department of Mathematics and Statistics

Auburn University

Outline

Three Generations of SAS Graphics

Three generations of SAS graphics

- proc plot
- proc gplot
- proc sgplot

ODS Graphics

- high-quality graphs with easier syntax
- use sgplot for single-celled graphs and sgpanel for multi-celled graphs.

Proc SGPLOT

We can use proc sgplot to draw the following plots.

X-Y plot	band, bubble, ellipse, high-low, loess, needle, penalized B-spline, regression, scatter, step, and vector
continuous variable	box, density, histogram
categorical variable	dot, bar, line

Specify a location to save graphs.

```
ods listing gpath = "/home/zengpen";
```

Bar Chart

A bar chart is used to display the frequencies of different levels of a categorical variable.

```
proc sgplot data = SAS-data-set;
   vbar variable / more-options;
run;
```

You may use the following options to further control the appearance.

- group = specifies a variable used to group the data
- groupdisplay = how to display grouped bars, either stack or cluster

We can also use title statement to add a title and use format and label statements to change the texts for variables/levels.

vbar can be changed to hbar for horizontal bar chart.

Histogram and Density

A histogram or density plot is used to display the distribution of a continuous variable.

```
proc sgplot data = SAS-data-set;
   histogram variable / more-options;
   density variable / more-options;
run;
```

- type = species the type of density curve, either normal (default) or kernel.
- The histogram and density statements can be used together, but not with other types of graphs.

Box Plot

A boxplot is used to display the five summary statistics of a continuous variable.

```
proc sgplot data = SAS-data-set;
   vbox variable / more-options;
run;
```

- category = specifies a categorical variable.
- group = specifies a second categorical variable.
- vbox can be changed to hbox for horizontal boxplot.

Scatter Plot

A scatter plot is used to show the relationship between two continuous variables.

```
proc sgplot data = SAS-data-set;
   scatter x = variable y = variable / more-options;
run;
```

- group = specifies a categorical variable.
- datalabel = specifies a label for each point.

Dot Plot

A dot plot is used to display some summary statistics of a continuous variable at different levels of a categorical variable.

```
proc sgplot data = SAS-data-set;
   dot categorical-variable / more-options;
run;
```

The options can be

- response = specifies a numeric response variable for the plot
- stat = specifies the statistic for the horizontal axis (freq, mean, median, percent, sum)
- limitstat = specifies the statistic for the limit lines (clm, stddev, stderr)

Series Plot

Series plot is also called line plot. It makes sense when data must be displayed in a particular order.

```
proc sgplot data = SAS-data-set;
   series x = variable y = variable / more-options;
run;
```

- group = specifies a categorical variable.
- markers adds a marker for each data point.

Fitted Curves

Show fitted curves by linear regression, loess, or splines.

```
proc sgplot data = SAS-data-set;
   statment-name x = variable y = variable / more-options;
run;
```

The *statement-name* can be

- reg for regression line
- loess for loess curve
- pbsline for penalized B-spline curves.
- group = specifies a categorical variable.
- datalabel = specifies a label for each point.

Controlling Axes and Reference Lines

```
proc sgplot data = data-set;
   xaxis options;
   yaxis options;
   refline value / options;
run;
```

options for xaxis and yaxis include

- grid creates a line at each tick mark on the axis
- label = 'text' specifies axis label

options for refline include

- axis = x or axis = y specify the location
- label = specify labels
- transparency = specifies the degree of transparency

Legends and Insets

keylegend / *options*;

Options include

- title = adds a title for the legend
- across = specifies number of columns in the legend
- down = specifies number of rows in the legend
- location = specifies location of legend (inside or outside)
- position = location of legend (top, topleft, topright, bottom, bottomleft, bottomright, left, right)

inset '*text-string*' / *options*;

Options include

- border adds a border
- position = specify the location

Graph Attribute

Use options after a slash at the end of a basic plot statement

statement / fillattrs = (attribute = value);

- fillattrs can be changed to labelattrs, lineattrs, markerattrs, valueattrs
- attribute can be
  - color = use name such as red or #FF0000
  - pattern = solid, dash, dot, ...
  - size = numbers with units cm, in, px, ...
  - style = italic or normal
  - symbol = circle, diamond, plus, square, ...
  - thickness = numbers with units cm, in, px, ...
  - weight = bold or normal