ODS graphics or template-based graphics
- ODS Graphics was the default for producing graphs in most SAS statistical procedures.
- Thus, ODS Graphics produced by various procedures appear automatically in the output.
- In SAS 9.2 under the SAS Windowing environment, the ODS GRAPHICS ON statement was required to enable ODS Graphics.
- With ODS Graphics, styles and templates control the appearance of tables and graphics.
- SAS 9.3 uses the default style called HTMLBlue available for the HTML destination.
- When statements that produce graphics are included in a proc step, ODS Graphics produced will be output in HTML by default.
- For example, the HISTOGRAM statement used in a proc univariate step will produce ODS Graphics output in HTML.
- As usual, graphical and other output (such as tables produced by the procedure) may be sent to a destination such as a PDF or an RTF file.

**SAS Example C4**

```sas
libname libc "U:\Documents\Stat479\";
data new;
set libc.biology;
BMI=703*Weight/Height**2;
run;
proc ttest data=new cochran ci=equal;
class Sex;
var BMI;
title "Biology class: Two Sample T-test";
run;
```

Figure 1: SAS Example C4: Output from PROC TTEST (Part 1)

Figure 2: SAS Example C4: Output from PROC TTEST (Part 2)

Figure 3: SAS Example C4: Output from PROC TTEST (Part 3)
What are SAS ODS Procedures?

- Base SAS now includes several procedures for creating plots required for statistical analysis.
- These create single-cell or multi-cell plots or panels of plots using simple syntax.
- Several of these procedures will be illustrated in this section via examples.
- A subset of the statements and options required for each procedure are presented.
- The three main procs are SGPLOT, SGPANEL, and SGSCATTER.

PROC SGPLOT < option(s)>;

Some ELLIPSE Statement Options

- alpha= specifies the confidence level for the ellipse.
- clip= specifies whether the ellipse will be clipped because the axes are determined without the ellipse.
- legendlabel="text-string" specifies a label that identifies the ellipse in the legend.
- lineattrs= specifies appearance of plotted lines (as a style-element or by using suboptions color=pattern=thickness=).
- name= specifies a name for the plot.
- outline=nouline specifies whether the outlines for the bars are displayed.
- type= specifies the type of ellipse. mean specifies a confidence ellipse for the population mean. predicted specifies a prediction ellipse for a new observation. Default is predicted.

The Ellipse statement

- The ellipse statement may be used along with the scatter statement to create bivariate confidence or prediction regions for a specified level $100(1-\alpha)%$.
- These are calculated under the assumption that pairs of data values $(x,y)$ have a bivariate normal distribution.
- A 95% prediction ellipse is generated by default. This implies that the probability of a point falling within the region is 0.95.
- On the other hand, one would have 95% confidence that the bivariate mean of the distribution lies within a 95% confidence ellipse.
- It is possible to draw multiple ellipses by including more than one ellipse statement specifying different types or alpha levels.
- Default legends for each ellipse, such as “95% Prediction Ellipse” are automatically generated.
- The user may specify a legend label using the legendlabel= option.
- The line attributes for the outline of the region may be specified using a style-element or by using suboptions.

The SGPLOT Procedure

- A few statements available in proc sgplot are illustrated via examples.
- The first statement discussed is the scatter statement.
- The user specifies an X variable and a Y variable to be plotted on the horizontal and the vertical axes, respectively.
- This statement produces scatter plots, ordered pairs $(x,y)$ plotted as points.
- A scatter plot visually displays the relationship between the two variables such as trends in the data or the occurrence of clusters of points.
- Other than simple interpretations such as whether the relationship is linear or not, scatter plots can be used for examining more complex statistical properties.
- They can be used check whether the spread of the data values of a response variable change as the X values change indicating increasing variance.

Some SCATTER Statement Options

datalabel= uses the Y values to label the points.
datalabel= uses the values of a variable to label the points.
group= specifies a variable that is used to group the data.
name= specifies a name for the plot.
markerattrs= specifies appearance of markers in the plot (as a style-element or using suboptions color=shape=pattern=).
markerchar= specifies a variable whose values replace the marker symbols in the plot.
markercharattrs= specifies the appearance of markers in the plot when the markerchar= option is used.

### Table 1: List of Marker Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ssqx</td>
<td>X</td>
</tr>
<tr>
<td>Ssqy</td>
<td>Y</td>
</tr>
<tr>
<td>Ssqxy</td>
<td>X</td>
</tr>
<tr>
<td>Ssqyx</td>
<td>Y</td>
</tr>
<tr>
<td>Ssqz</td>
<td>Z</td>
</tr>
<tr>
<td>Ssqxyz</td>
<td>X, Y, Z</td>
</tr>
</tbody>
</table>

### Table 2: Units of Measurement

<table>
<thead>
<tr>
<th>Unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>in</td>
<td>inches</td>
</tr>
<tr>
<td>mm</td>
<td>millimeters</td>
</tr>
<tr>
<td>pt</td>
<td>point size, calculated at 100 dpi</td>
</tr>
<tr>
<td>ps</td>
<td>pixels</td>
</tr>
</tbody>
</table>

### Table 3: Line Patterns

- Table of line patterns with various styles and patterns.
SAS Example C5
libname libc "U:\Documents\Stat479\";

title "Plot of Height vs. Weight with Prediction Ellipse";
proc sgplot data=libc.biology;
    scatter x=Height y=Weight;
    ellipse x=Height y=Weight/lineattrs=(Color=red);
run;

title "Plot of Height vs. Weight with Age as Data Labels";
proc sgplot data=libc.biology;
    scatter x=Height y=Weight/datalabel=Age
        markerattrs= (Color=magenta Size=2 mm Symbol=Asterisk);
run;

title "Height vs. Weight grouped by Gender";
proc sgplot data=libc.biology;
    scatter x=Height y=Weight/group=Sex
        keylegend / location=inside position=bottomright;
run;

Some HISTOGRAM Statement Options
- binstart= specifies the X coordinate of the first bin
- binwidth= specifies the bin width
- boundary=lower|upper specifies how boundary values are assigned to bins (default=upper)
- fill|nofill specifies whether the area fill is visible
- fillattrs= specifies appearance of the area fill (as a style-element or using suboptions)
- legendlabel="text-string" specifies a label that identifies the histogram in the legend
- name="text-string" specifies a name for the plot
- nbins= specifies the number of bins
- outline|nooutline specifies whether the outlines of the ellipse are visible
- scale=count|percent|proportion specifies the scale of the vertical axis. Default is percent

The HISTOGRAM statement
- The histogram statement produces a histogram of a continuous variable.
- Number of bins (classes) and binwidths (class interval widths) are determined automatically
- Options available to specify the start of the first bin, number of bins, and the binwidth
- The vertical axis represents the frequency in each of the bins as a count, or as a percentage or proportion
- Percentage frequency is plotted on the vertical axis by default

The DENSITY statement
- The density statement allows the user to overlay a density plot fitted to the data
- One type of a density plot is obtained by fitting a normal distribution to the data
- The user may give values for the two parameters \(\mu\) and/or \(\sigma\) using suboptions mu= and sigma= specify
- If values for either parameter is not specified they will be estimated from the data
- The other type of density plot available is a nonparametric kernel density estimate.
- The user may specify a standardized bandwidth \(c\) and a kernel function weight= specify
- The weight= suboption accepts three possible kernel functions normal, quadratic or triangular as its value, with normal being the default.

Notes on Bandwidth Selection
- The standardized bandwidth is a value between 0 and 100 and controls the level of smoothing
- Too small a value will show little smoothing showing spikes by attempting to fit every detail
- A too large a value will perform oversmoothing hiding most of the structure in the data
- An optimal bandwidth is one that results in a density estimate that is close to the true density
SAS Example C6

data chicks;
  input Wtgain @@;
  label Wtgain = 'Weight gain (in gms) after 8-weeks';
datalines;
 3.7 4.2 4.4 4.4 4.3 4.2 4.4 4.8 4.9 4.4
 4.2 3.8 4.2 4.4 4.6 3.9 4.1 4.5 4.8 3.9
 4.7 4.2 4.2 4.8 4.5 4.3 3.9 4.2
 4.0 4.2 4.0 4.5 4.4 4.1 4.0 4.0 3.8 4.6
 4.9 3.8 4.3 4.3 3.9 3.8 4.7 3.9 4.0 4.2
 4.3 4.7 4.1 4.0 4.6 4.4 4.6 4.4 4.9 4.4
 4.0 3.9 4.5 4.3 3.8 4.1 4.3 4.2 4.5 4.4
 4.2 4.7 3.8 4.5 4.0 4.2 4.1 4.0 4.7 4.1
 4.7 4.1 4.8 4.1 4.3 4.7 4.2 4.1 4.4 4.8
 4.1 4.9 4.3 4.4 4.4 4.3 4.6 4.5 4.6 4.0

  proc sgplot data=chicks;
    title 'Weight Gain Distribution';
    histogram Wtgain;
    density Wtgain / type=kernel;
    keylegend / location=inside position=topright;
  run;

Notes on the Construction of the Histogram

- Analysis of the data suggests class intervals of width 0.1 units beginning with a midpoint at the smallest data value of 3.6.
- However, we used the histogram statement (with no options)
- The first density statement specifies that a normal density curve be superimposed on the histogram
- By default, the procedure fits a normal density to the data using the sample mean and sample standard deviation estimated from the data.
- The second density statement will superimpose a kernel density plot using default values for the bandwidth and kernel (weight) function

SAS Example C7

data emissions;
  input period @;
  label period = 'Year of Manufacture';
datalines;
 1 2351 1293 541 1058 411 570 800 630 905 347 -1
 2 620 940 350 700 1150 2000 823 1058 423 270 900 405 780 -1
 3 1088 388 111 558 294 211 460 353 71 241 2999 199 188 353 117 -1
 4 141 359 940 882 494 306 200 100 300 190 140 880 200 223 188 435 940 24
 5 140 160 20 223 60 20 95 360 70 220 400 217 58 235 1880 200 175 85 -1

  proc format;
    do until (hc<0);
      hc=1;
      input hc @;
    output;
  end;
  format period pp.;
  vbox hc / category=period datalabel;
  title "Hydrocarbon Emissions Distribution by Period";
run;

Figure 7: Output from SAS Example C6

Some VBOX Statement Options

- boxwidth= specifies the width of the box, as a value between 0.0 (0% of the available width) and 1.0 (100% of the available width). Default is 0.4.
- category= specifies the category variable for the plot. A box plot is created for each distinct value of the category variable.
- connect= specifies appearance of the connecting lines (as a style-element or using suboptions color=pattern=thickness=).
- datalabel<=variable> adds data labels for the outlier markers. If you specified a variable, then the values for that variable are used for the data labels. If you did not specify a variable, then the values of the analysis variable are used.
- datalabelattrs= specifies appearance of labels (as a style-element or using suboptions color=, family= "font-family", size=, style=italic/normal, weight=bold/normal).
Figure 8: Output from SAS Example C7

Table 4: Mice survival time data

<table>
<thead>
<tr>
<th>Poison</th>
<th>Drug</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>0.31</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>II</td>
<td>0.46</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>0.63</td>
<td>0.62</td>
</tr>
<tr>
<td>III</td>
<td>0.22</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.22</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Table 5: Data arranged for input to proc sgplot

SAS Example C8

```sas
data survival;
input Poison 1. @;
do Drug=1 to 4;
    input Time 3.2 @;
    output;
end;
datalines;
1 3 18 24 34 5
1 45 110 45 71
1 4 68 86 36 6
1 4 37 27 66 2
2 3 69 24 45 6
2 29 61 35 102
2 4 93 17 1
2 23 124 40 38
3 2 23 02 33 0
3 2 13 72 53 6
3 1 83 82 43 1
3 2 32 92 23 3
;
title1 "Analysis of Survival Time data";
title2 "Interaction Plot of Cell Means";
proc sgplot data=survival;
vline Poison/response=Time stat=mean group=Drug markers;
run;
```
PROC SGPANEL < option(s)>;
    PANELBY variable(s) </option(s)>;
    COLAXIS <option(s)>;
    DENSITY response-variable </option(s)>;
    DOT category-variable </option(s)>;
    HBAR category-variable </option(s)>
    HBOX response-variable </option(s)>
    HISTOGRAM response-variable </option(s)>
    HLINE category-variable </option(s)>
    KEYLEGEND <"name(s)"> </option(s)>
    HLINE variable </option(s)>
    LOESS X= numeric-variable Y= numeric-variable </option(s)>
    NEEDLE X= variable Y= numeric-variable </option(s)>
    PBSPLINE X= numeric-variable Y= numeric-variable </option(s)>
    RFLINE value(s) </option(s)>
    REG X= numeric-variable Y= numeric-variable </option(s)>
    ROWAXIS <option(s)>
    SCATTER X= variable Y= variable </option(s)>
    SERIES X= variable Y= variable </option(s)>
    STEP X= variable Y= variable </option(s)>
    VBAR category-variable </option(s)>
    VBOX response-variable </option(s)>
    VLINE category-variable </option(s)>

SAS Example C8
libname mylib 'U:\Documents\Stat479\';
proc format;
    value ing 1 = 'Low Income'
          2 = 'Middle Income'
          3 = 'High Income';
run;

title 'Horizontal Bar Chart of Miles of Primary Highways';
proc sgplot data=mylib.fueldat;
    hbar Incomgrp/response=Roads stat=mean group=Taxgrp;
    keylegend / title="Fuel Tax" location=outside position=bottom;
    format Incomgrp ing.;
run;

title 'Vertical Bar Chart Fuel Use by Income Group for each Fuel Tax Group';
proc sgpanel data=mylib.fueldat;
    panelby Taxgrp;
    vbar Incomgrp/response=Fuel stat=mean;
    format Incomgrp ing.;
run;

Some PANELBY Statement Options:
- border/noborder specifies whether borders are to be drawn around each cell in the panel display.
- colheaderpos=top/bottom/both specifies the location of the column headings in the panel.
- columns=n specifies the number of columns in the panel.
- layout=lattice/panel/columnlattice/rowlattice specifies the type of layout that is used for the panel.
- missing a missing value is taken as a valid category and creates a cell for it.
- novarname removes the variable names from the cell headings of a panel layout, or from the row and column headings of a lattice layout.
- onepanel places the entire panel in a single output image.
- rowheaderpos=left/right/both specifies the location of the row headings in the panel.
- rows=n specifies the number of rows in the panel.
- spacing=n specifies the number of pixels between the rows and columns in the panel.
- start=topleft/bottomleft specifies whether the first cell in the panel is placed at the upper left or the lower left corner.
- uniscale=column/row/all scales the shared axes in the panel to be identical.
**SAS Example C10**

```sas
data world;
infile "U:\Documents\Stat479\demogr.txt";
if LevTech<13 then TechGrp=1;
else if 13<= LevTech<25 then TechGrp=2;
else if 25<= LevTech<60 then TechGrp=3;
else if LevTech>= 60 then TechGrp=4;
label LifeExp='Life Expectancy in yrs.'
TechGrp='Level of Technology'
run;
proc format;
  value tg 1='Low'
      2='Moderate'
      3='High'
      4='Advanced' ;
run;
proc sgscatter data=world;
  title "Scatterplot Matrix of Demographic Variables";
  matrix PopUrban PercGnp InfMort LifeExp
    /group=TechGrp;
  format TechGrp tg.;
run;
```

**Notes on the Dot plot**

- Values of a quantitative variable needs to be plotted by values of a nominal variable or a category variable.
- The nominal variable has a large range of values or category variable has a large number of levels.
- In this situation dot plots more compact than, say horizontal bar charts.
- They are easily modified if the category variable is divided into groups into clusters.
- Or when the quantitative variable is itself classified into several groups.
- They can be grouped or clustered in separate panels.

**SAS Example C11**

```sas
libname mylib "U:\Documents\Stat479\";
proc sgscatter data=mylib.fueldat;
  title "Scatterplot Matrix of Fuel Use by Fuel Tax";
  panelby Taxgrp/layout=rowlattice;
  dot State/response=Fuel categoryorder=respasc;
run;
```

**SAS Example C12**

```sas
data world;
infile "U:\Documents\Stat479\demogr.txt";
if LevTech<13 then TechGrp=1;
else if 13<= LevTech<25 then TechGrp=2;
else if 25<= LevTech<60 then TechGrp=3;
else if LevTech>= 60 then TechGrp=4;
label TechGrp='Level of Technology';
run;
proc format;
  value tg 1='Low'
      2='Moderate'
      3='High'
      4='Very High' ;
run;
proc sgscatter data=world;
  title "Scatterplot Matrix of Demographic Variables";
  matrix PopUrban PercGnp InfMort LifeExp
    /group=TechGrp;
  format TechGrp tg.;
run;
```
Figure 14: Output from SAS Example C12: Scatter Plot Matrix of world demographic data.