CHAPTER 5
STATISTICAL GRAPHICS
ON SAS
YI-TING HWANG
DEPARTMENT OF STATISTICS
NATIONAL TAIPEI UNIVERSITY
TYPE OF VARIABLES

• Discrete variable:
  – Pie chart (圓形比例圖)
  – Bar chart (長條圖)

• Continuous variable:
  – Histogram (直方圖)
  – Box plot (盒形圖)
  – Scatter plot (散佈圖)
  – Stem-leaf plot (莖葉圖)
PROC CHART is a useful tool to visualize data quickly,
The GCHART procedure in SAS/GRAPH software produces the same types of charts as PROC CHART does.

– produce presentation-quality graphics that include color and various fonts.
SAS SYNTAX

PROC CHART < option(s)>;
BLOCK variable(s)< / option(s)>;
BY<variable(s)>;
HBAR variable(s) < / option(s)>
PIE variable(s)< / option(s)>;
VBAR variable(s)< / option(s)>;
SAS SYNTAX

PROC GCHART <DATA=input-data-set> ;
BLOCK chart-variable(s) </ option(s)>;
HBAR | HBAR3D | VBAR | VBAR3D
    chart-variable(s) </ option(s)>;
PIE | PIE3D | DONUT chart-variable(s) </ option(s)>;
OPTIONS

• Statistics
  – TYPE
    • CFREQ (CPERCENT)
    • MEAN
    • PERCENT
    • SUM
  – SUMVAR=summary-variable

• Midpoint
  – GROUP=group variable
  – SUBGROUP=subgroup variable
EXAMPLE 1

Consider a study on whether a patient has surgery with general anesthesia (麻醉) experienced a sore throat on waking. Variables includes:

- D = duration of the surgery (in minutes)
- T = type of device used to secure the airway
  - 0 = laryngeal (喉頭治療用的) mask airway
  - 1 = tracheal (氣管的) tube
- Y = sore throat (0 = no, 1 = yes)
PIE CHART

SUM of AERO by SEX

F
701.1

M
759.4
EXAMPLE 2

data speed;
  input rated actual @@;
  label rated='Rated Speed'
        actual='Actual Speed';
cards;
  75  85  110  112  75  81  105  108
112  115  75  77  90  89  70  73
118  121  103  100  118  121
run;
SCATTER PLOT

Rated Speed Versus Actual Speed
PLOT

The PLOT procedure plots the values of two variables for each observation in an input SAS data set. The coordinates of each point on the plot correspond to the two variables' values in one or more observations of the input data set.
SAS SYNTAX

PROC PLOT <option(s)>;
BY <DESCENDING> variable-1
   <...<DESCENDING> variable-n>
   <NOTSORTED>;
PLOT plot-request(s) <l option(s)>;
SAS SYNTAX

PROC GPLOT <DATA=input-data-set>;
   PLOT plot-request(s) </option(s)>;
   PLOT2 plot-request(s) </option(s)>;
SYMBOL

The SYMBOL statement defines the characteristics of symbols that display the data plotted by a PLOT statement used by PROC GBARLINE, PROC GCONTOUR, and PROC GPLOT.

Syntax
SYMBOL<1...255> <COLOR=symbol-color>
   <appearance-option(s)> 
   <interpolation-option> ;
APPEARANCE OPTIONS

BWIDTH=box-width
CI=line-color
CO=color
CV=value-color
FONT=font
HEIGHT=symbol-height<units>
LINE=line-type
POINTLABEL<=(label-description(s)) | NONE>
VALUE=special-symbol | text-string | NONE
WIDTH=thickness-factor
## INTERPOLATION OPTION I

<table>
<thead>
<tr>
<th>Method</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>General methods</td>
<td>INTERPOL=JOIN</td>
</tr>
<tr>
<td></td>
<td>• JOIN connects data points with straight lines.</td>
</tr>
<tr>
<td></td>
<td>• STEP specifies that the data are plotted with a step function.</td>
</tr>
<tr>
<td></td>
<td>• Specify one of these values for placement: L/R/C</td>
</tr>
<tr>
<td></td>
<td>• J produces steps joined with a vertical line.</td>
</tr>
<tr>
<td></td>
<td>• S sorts unordered data by the independent variable before plotting.</td>
</tr>
<tr>
<td>Regression interpolation methods</td>
<td>INTERPOL=R&lt;type&gt;&lt;0&gt;&lt;CLM</td>
</tr>
<tr>
<td></td>
<td>• Type specify type of regression (L: linear Q: quadratic C: cubic)</td>
</tr>
<tr>
<td></td>
<td>• 0 eliminates intercept from the regression equation.</td>
</tr>
<tr>
<td></td>
<td>• CLM displays confidence limits for mean predicted values</td>
</tr>
<tr>
<td></td>
<td>• CLI displays confidence limits for individual predicted values.</td>
</tr>
</tbody>
</table>
### INTERPOLATION OPTION II

<table>
<thead>
<tr>
<th>Method</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box plot</td>
<td>INTERPOL=BOX&lt;option(s)&gt;&lt;00...25&gt; produces box and whisker plots.</td>
</tr>
<tr>
<td></td>
<td>- The bottom and top edges of the box are located at the sample 25th and 75th percentiles.</td>
</tr>
<tr>
<td></td>
<td>- The center horizontal line is drawn at the 50th percentile (median).</td>
</tr>
<tr>
<td></td>
<td>- Any value more extreme than this is marked with a plot symbol.</td>
</tr>
<tr>
<td></td>
<td>Values for option(s) are one or more of these:</td>
</tr>
<tr>
<td></td>
<td>F =fills the box with the color specified by CV= and outlines the box</td>
</tr>
<tr>
<td></td>
<td>J =joins the median points of the boxes with a line</td>
</tr>
<tr>
<td></td>
<td>T=draws tops and bottoms on the whiskers.</td>
</tr>
</tbody>
</table>
EXAMPLE 3

GRADES contains codes to identify each class section, and the grades scored by students in each section. Generate the box plot.