SAS / INSIGHT

ShortCourse Handout

February 2005

Copyright © 2005 Heide Mansouri, Technology Support, Texas Tech University. ALL RIGHTS RESERVED. Members of Texas Tech University or Texas Tech Health Sciences Center may print and use this material for their personal use only. No part of this material may be reproduced in any form without written permission from Heide Mansouri, the <u>author</u>.

Introduction

SAS/INSIGHT is a module in the SAS system that allows you to display and to analyze your data interactively. It provides a higher level of interactivity than other modules in the SAS system. SAS/INSIGHT software is used to:

- identify observations in plots and to highlight observations in linked scatter plots, histograms, box plots, line plots, and three-dimensional rotating plots.
- search, sort, and edit data, and to color observations based on the value of a variable.
- compute descriptive statistics, create plots, and perform hypotheses test.
- analyze relationships between a response variable and a set of explanatory variables by fitting curves, and adding confidence bands for mean and predicted values.

The main point of SAS/INSIGHT is its interactivity with the user. This ShortCourse is designed to help you get started using SAS/INSIGHT. It is assumed that you have completed the Base SAS ShortCourse or already know how to create SAS data sets. This ShortCourse also assumes a familiarity with elementary statistics. We will be using two sample data sets: BASEBALL and GPA, included with SAS/INSIGHT, throughout this ShortCourse.

Course Objectives

After completing this ShortCourse, you should be able to use SAS/INSIGHT to:

- Enter values in the data window,
- Identify observations in plots,
- Color observations in plots, histograms, box-plots, and 3-dimensional rotating plot,
- Analyze your data interactively,
- Create Bar charts, Box-plots, and Scatter plots,

- Perform Distribution Analysis,
- Create a "Summary of Fit" table,
- Create an ANOVA table, and
- Save data to a SAS data set.

Invoking SAS/INSIGHT

- Start the SAS program.
- Choose Solutions > Analysis > Interactive Data Analysis (or type insight in the Command line by the check mark).
- Click the New button to open a new data window in which you can enter data.
- You may also open an existing SAS data set with which to work.

Entering Values

SAS/INSIGHT software displays SAS data sets in tabular form (SAS spreadsheet). By default, the first value in a new data window is selected and is displayed with a frame around it. This *active value* marks your current location in the data window. To enter data, begin typing, and then press the Enter key. The count of variables and observations is shown in the upper left of the data window. You can rename a column by double-clicking on the box above it. You can specify the variable type as Interval (**Int**), Nominal (**Nom**), Group, Label, Freq, or Weight.

Using the Data Table

- The upper-left corner of this window shows the number of variables in the top right and the number of observations in the bottom left.
- Click and hold down the mouse button on the small triangle in the upper-left corner of this window. A pop-up menu will open that lets you Find, Sort, and Rearrange your observations; add variables or observations; set variable characteristics; and set other options for the display of the data set.
- Double-click on an observation number listed down the left side to view all the values for the observation.
- Click on the **black square** next to each observation to specify if you want that observation included in calculations and graphs and labeled in plots. If you have specified any color for an observation, this square will be colored instead of black.

- Double-click on one of the variable names listed across the top to define that variable.
- Click the left box above a variable name to specify if a variable is to be used for grouping or labeling observations in your plots. The right box above a variable name tells you whether that variable is numeric (Int) or character (Nom).

🛞 WORK.A							×
▶ 1	Nom						^
10	A						
1	male						
■ 2							
a 3							
■ 4							
5							
6							
■ 7							
8							
9							
10							*
<						>	

Objects in SAS/INSIGHT

- Variables
- Observations
- Values
- Graphs
- Curves, and
- Tables

You can select multiple objects by dragging the mouse across the objects. You can also use the Shift (or Ctrl) key to extend your selection for objects that are far apart.

Adding Variables and Observations

- Click in the upper-left corner of the data window.
- Choose New Observations.
- Enter the number of observations, and then click **OK**.
- Select all variables.
- Click the variable count in the upper-left corner of the data window.

• Enter data, using Tab and Back Tab to navigate within the selected variables.

Defining Variables

- Double-click on the measurement level indicator for variables (variable name) to display a pop-up menu. The radio mark on the measurement level indicates the current measurement level (Int or Nom).
- Click in the upper-left corner of the data window to display the data pop-up menu, and then
- Choose **Define Variables** from the pop-up menu to assign its variable storage type, measurement level, default roles, name, and label.
- Enter a name for your variable, then
- Click the **Apply** button.

Closing the Data Window

You can close any window by choosing **File> End**. When you close the data window, you close all windows using that data set. When you close all your data windows, you exit the SAS/INSIGHT software.

Creating Sample Data Sets

- In SAS/INSIGHT, click Help.
- Click on Create Samples.
- Click OK.

Using Baseball Sample Data set

- Select File > Open ...
- Select **Sasuser** from the list of Libraries.
- Select **Baseball** from the list of Data Set.
- Click Open.

Sorting Observations

- Select a variable.
- Click on the menu button in the upper-left corner.
- Click on **Sort**.

Finding Observations

- Choose Edit > Observations > Find.
- Select the variables.
- Select the values.
- Click the **Apply** button to find the data.
- Click OK.

Creating Charts

- Click the Analyze menu on the main SAS window to make a
 - o Histogram/Bar Chart (Y)
 - Box Plot/Mosaic Plot (Y)
 - Line Plot (Y X)
 - Scatter Plot (Y X)
 - Contour Plot (Z Y X)
 - Rotating Plot (Z Y X)
- Right-clicking anywhere on a chart will bring up a menu.

Options for Charts

You can use the **Edit** menu to alter your charts.

- **Renew:** allows you to alter the input you specified for a chart.
- Animate: allows you to put animation into your charts.
- Tools: allows you to color or mark observations.
- Fonts: allows you to change the font of text in your charts.
- Variables: Allows you to perform various transformations on variables.

Statistical Analysis

From the **Analyze** menu you can use **Distribution**, **Fit**, and **Multivariate** to examine a variable's distribution, perform regressions of one variable on others, and perform correlation analyses.

Creating Bar Charts

- Select **SALARY** in the data window.
- Choose Histogram/Bar Chart (Y) from the Analyze menu.
- Click on any bar, and then click on the menu button in the bottom-left corner of the chart.

- Select LEAGUE in the data window. Note that LEAGUE is a *nominal* variable. Nominal variables contain a discrete set of values. For example, LEAGUE contains only two values, American and National, for the American and National leagues.
- Choose Histogram/Bar Chart (Y) from the Analyze menu.
- Select **Values** from the bar chart pop-up menu in the new bar chart.
- Arrange the windows so that you can see both bar charts (click Window > Tile vertically).
- Click on the bar that represents the American League. This selects all observations for players in the American League.
- Click on the bar that represents the National League.

Creating Box Plots

- Click anywhere inside the data window.
- Choose Analyze > Box Plot/Mosaic Plot (Y).
- Select **SALARY** in the list at the left, and then click the **Y** button.
- Select **LEAGUE** in the list at the left, and then click the **X** button.
- Select **NAME** in the list at the left, and then click the **Label** button. This assigns the **Label** role to **NAME**. The label variable is used to identify extreme values in the box plot.
- Click OK.
- Close the Baseball data window.



Creating Scatter Plots

- In SAS, click Solutions > Analysis > Interactive Data Analysis.
- Choose **Sasuser > GPA data set**, and click Open.
- Select both the SATM and SATV variables.
- Choose Analyze > Scatter Plot (Y X).
- Click on a marker. This selects the observation in the data window because windows are linked to their data. Any change to the data is automatically reflected in all windows.
- Double-click on a marker. This selects the marker and displays the Examine Observation dialog.

Coloring Observations

- Click on **SATM** and **SATV** variables in the data window.
- Click Edit > Windows > Tools.
- Click on the large, multiple colors button in the tools window.
- Choose Analyze > Scatter Plot (Y X).



Using SAS/INSIGHT Tools

- Click on Edit > Window > Tools. The sets of icons appear
- In this window are included:
 - Pointer, Hand and Magnifying glass
 - Color pallet



- o Markers
- Use the **pointer** to highlight specific points, either in the actual spreadsheet or on a plot.
- Use the **Hand** tool to manipulate the constructed plot.
- Use the Magnifying glass to look at pictures in more detail.
- Highlight an observation either on the plot or on the spreadsheet
- Click on the desired marker type to give one observation a different Marker type.
- For any interval (categorical) variable, you can assign each possible value to have a different marker.

Performing Distribution Analysis

- Select the variable **SATM** by clicking on its name in the data window.
- Choose Analyze > Distribution (Y).
- Identify the extreme observations by clicking on them.

Editing the Histogram

- Choose Edit > Windows > Tools.
- Click on the hand in the Tools window.
- Click on the histogram.
- Drag the hand horizontally and vertically in the histogram until you find a histogram that captures the foremost shape of the distribution.
- Click on the arrow in the tool.



Moments and Quantiles Tables

- **N** is the number of nonmissing observations.
- **Mean** is the arithmetic mean.
- **Std Dev** is the standard deviation.
- Variance is the variance.
- Skewness and Kurtosis are both measures of the shape of the distribution.

Creating Fit Analysis (Regression Analysis)

- Open the **GPA** data set.
- Choose Analyze > Fit (Y X).
- Select the variable **GPA** in the list on the left, and then click the **Y** button.
- Select the variables HSM, HSS, and HSE, and then click the X button.
- Click the **Apply** button.

Summary of Fit Table

- The **Root MSE** value is **0.6998** and is the square root of the mean square error given in the **Analysis of Variance** table.
- The **R-Square** value is **0.2046**, which means that 20% of the variation in **GPA** scores is explained by the fitted model.

Saving and Printing Graphics

- Click on the edge of the graphic to select it.
- From the Edit menu, click Copy to copy it to the Windows Clipboard.
- Save it to a text file, and then print it.

Printing Graphics using Windows Printing Facilities

- Select the graphic by clicking on the edge of the graphic.
- From the File menu, select Print Preview.
- Make your print selections.
- Click Print.

Saving & Printing Tables

- 1. Select the tables by clicking on their edges.
- From the File menu, choose Save > Tables (the selected tables will be sent to the Output window).

3. Send the contents of the Output window to a **file** or to a **printer**.

<u>Note</u>: You can also print graphics or tables using the **Screen-dumping** utilities.

Online References

- <u>http://www.sas.com/technologies/analytics/statistics/insight</u>
- <u>http://support.sas.com/rnd/app/da/insight.html</u>
- <u>http://support.sas.com/rnd/app/doc.html</u>
- <u>http://www.ens.gu.edu.au/stats/sashelp/inhelp.htm</u>
- <u>http://gsbwww.uchicago.edu/computing/research/SASManual/insight</u>

Where to Get Help

- If you need professional statistical advice, you may contact Statistical Consulting Lab (SCL) @ <u>www.math.ttu.edu/~scl</u> or call 742-2580 ext. 232.
- If you need help from your instructor, you may e-mail <u>heide.mansouri@ttu.edu</u>.

Exercise: Data Transformations (Taken from SAS Documentation)

A data *transformation* generates a new variable from existing variables according to a formula you specify. The most common transformations are available in the **Edit > Variables** menu. For example, you can use the **log transformation** to **linearize relationships**, **stabilize variances**, or reduce **skewness**.

Using the **BASEBALL** Sample data set included in **Sasuser** *library* (in SAS), complete the following exercises:

- 1. Open the **BASEBALL** data set.
- 2. Create a Fit Analysis of SALARY versus CR_HOME.
 - Choose Analyze > Fit (Y X).
 - > Select the variable **SALARY** in the list on the left, and then click the **Y** button.
 - > Select the variable **CR_HOME**, and then click the **X** button.
 - Click the Apply button. This window shows the results of a regression analysis of CR_HOME on SALARY. <u>Notice</u> that the majority of the observations appear in the lower-left corner of the scatter plot and that the regression line does not fit the data well. To make the relationship clearer, apply a logarithmic transformation to both variables.

- 3. Select both variables in the scatter plot.
- 4. Choose Edit > Variables > log(Y). This performs a log transformation on both SALARY and CR_HOME and transforms the scatter plot to a log-log plot. Now the regression fit is improved, and the relationship between salary and home run production is clearer. <u>Notice</u> that the degrees of freedom (DF) is reduced from 261 to 258. This is due to missing values resulting from the log transformation, described in the following step. Also <u>notice</u> that in addition to residual and predicted values from the regression, the log transformations create two new variables: L_SALARY and L_CR_HOM. The log transformation is useful in many cases. However, the result of log(Y) is undefined where Y is less than or equal to 0. In such cases, SAS/INSIGHT software cannot transform the value, so a missing value (.) is generated. To see this, sort the data in the data window.
- Select L_CR_HOM in the data window, using the menu on the upper-left corner of the window (right-click on the triangle), and choose Sort. The log transformation has removed data from the fit analysis.
- 6. Select **CR_HOME** in the data window.
- Choose Edit > Variables > Other... In the dialog you can see that the variable CR_HOME is already assigned as the Y variable.
- 8. Scroll down the transformation window, and select **log(Y + a)**.
- 9. In the field for a, enter the value 1, and then press the Enter key. <u>Notice</u> that the Label value changes from log(CR_HOME) to log(CR_HOME + 1) to reflect the new value of a. Setting a to 1 avoids the problem of generating missing values because (CR_HOME + 1) is greater than or equal to zero in all cases for this data.
- **10**. Click **OK** to perform the transformation.
- 11. Scroll all the way to the right to see the new variable, L_CR_H_1. <u>Notice</u> that the new variable contains no missing values.
- 12. Select L_SALARY and L_CR_H_1 (using the Shift or Ctrl Keys); then choose Analyze > Fit (Y X). At the lower, left corner of the scatter plot, you can see observations that were not used in the previous fit analysis. Also note that the degrees of freedom (DF) is back to 261.

Please e-mail your comments or suggestions to: <u>heide.mansouri@ttu.edu</u>