SAS Enterprise Guide

6th Annual SAS® Summer Training Institute
by the University of Iowa SAS® User Group
August 20, 2014

Yelena Perkhounkova
elena-perkhounkova@uiowa.edu

1

–

2

What is SAS Enterprise Guide?

- A point-and-click graphical interface to SAS
- Provides easy access to data sources
- Makes reporting & analytics more available by providing ready-to-use ***tasks***
- Helps organize work in **projects** that include
  - SAS programs
  - References to data
  - Results in various formats
  - Logs
  - Relationships among the items above

3

Why SAS Enterprise Guide?

- Important: you **can write code** in SAS EG just like in any SAS environment!
- However, if your data are already clean and properly set up, you don't need to write a single line of code in SAS EG to do a lot of different analyses.

4

Accessing SAS EG

- Virtual Desktop
  - http://helpdesk.its.uiowa.edu/software/
  - Click on "Software Available Online through Virtual Desktop"
  - Select SAS, then SAS Enterprise Guide
- SAS installed on your PC
  - Use Start menu
  - Select SAS, then SAS Enterprise Guide

5

Using SAS Enterprise Guide

I will show today how to use SAS EG to

- Create and save projects
- Add SAS data to the project/export data outside SAS
- Manipulate data
- Summarize data
- Plot data
- Run simple statistical analysis

6

Starting with EG

- Open EG
- Start new project
- Under Server List, go to Servers→Local→Libraries→SASHELP
- Open HEART dataset
- To explore file properties
  - Right-click on file in project tree

7

Manipulating Data I

- Click on “Filter and Sort” tab
  - Select all variables
  - Choose filter (**status** equal to “Alive”)
  - Click “OK”
- Save new data to H:\your project\SAS\DATA
  - Click on the “export” tab
  - Save as HEARTALIVE
- Save project to H:\your project\SAS\PGM
- Close the project

8

Opening Data into SAS EG

- Reopen your project
- To open existing SAS file into project
  File→Open → Data→
  - Browse, select & open file HEARTALIVE
- Check file properties

9

Manipulating Data II

- Double-click on HEARTALIVE data
- Click on "Filter and Sort" tab
  - Select all variables
  - Choose filters (**weight** > 200 AND **chol_status** equal to "High")
  - Click "OK"
- Click on "Modify Task" tab
  - Sort by **weight**
  - Click "OK"

10

Saving Data in Excel Format

- Click on the “export” tab
- Choose file type “Excel”
- Name file as HEARTHIGHRISK
- Save

11

Importing Data

- To open non-SAS data file, e.g. Excel File→Import data→
  - Browse, select file, open, follow instructions:
  1. “Specify the data”: click next
  2. “Select Data Source”: choose worksheet, select “rename columns to comply with SAS naming”, click next
  3. “Define field attributes”: click next
  4. “Advanced options”: select “remove characters that can cause transmission errors from text-based data files”, click finish

12

Summary Statistics

- Double-click on HEARTALIVE data
- Click on "Describe" tab→Summary Statistics
  - Data: Choose **AgeatStart**, **Height**, and **Weight** for analysis variables and **Sex** for classification variable
  - Statistics
    - Basic: select mean, standard deviation, min, max
    - Percentiles: select median
    - Additional: select confidence limits of the mean
  - Plots: Select histogram and box and whisker plot
  - Titles: Change title
- Click on "Run"
- Examine output

13

One-Way Frequencies

- Double-click on HEARTALIVE data
- Click on "Describe" tab→One-Way Frequencies
  - Data: Choose **Chol_status, BP_Status, Weight_Status,** and **Smoking_Status** for analysis variables
  - Plots: Select bar chart
- Click on "Run"
- Examine output
- Check code

14

Saving Output

- Choose other output formats

Tools→Options

  - Results: Select HTML and RTF

- Click on "Refresh" tab
- Click on "Results-RTF" tab
- Export the output to

  H:\your project \SAS\OUT

15

Distribution Analysis

- Double-click on HEARTALIVE data
- Click on "Describe" tab→Distribution Analysis
  - Data: Choose **Diastolic** and **Systolic** for analysis variables
  - Distributions
    - Normal: select "Normal", "Suppress distribution tables"
  - Plots
    - Appearance: select histogram plot and probability plots
  - Tables: select basic measures, extreme rows, moments, tests for normality
- Click on "Run"

16

Plotting Data I

- Double-click on HEARTALIVE data
- Click on "Graph" tab → Line Plot
  - Data: Choose **BP_status** for horizontal and **Weight** for vertical variable
  - Titles: Specify
- Click on "Run"
- Examine output

17

Plotting Data II

- Click on "Modify Task" tab
- Select "Multiple line plots by group column"
  - Data
    - For **weight** select "summarize for each distinct horizontal value", function "average"
    - Choose **Sex** for group variable
  - Appearance
    - Plots: add symbols
  - Axes
    - Horizontal axis: select reverse Axis
- Click on "Run"

18

Correlation

- Double-click on HEARTALIVE data
- Click on "Analyze" tab → Multivariate → Correlations
  - Data: Choose **Diastolic** and **Systolic** for analysis variables and **Weight** for correlate with variable
  - Results: Select "create a scatter plot for each correlation pair"
- Click on "Run"
- Use "Modify Task" to group analysis by **Sex**

19

Linear Regression

- Double-click on HEARTALIVE data
- Click on "Analyze" tab → Regression → Linear Regression
  - Data: Choose **Cholesterol** for dependent variable, **Weight**, **Diastolic**, and **Systolic** for explanatory variables
  - Statistics: select "Standardized regression coefficients" and "Confidence limits for parameter estimates"
  - Plots: select custom list of plots (histogram plot of residuals)
- Click on "Run"
- Modify Task to do stepwise selection

20

t-test

- Double-click on HEARTALIVE data
- Click on "Analyze" tab → ANOVA → t Test
  - t Test type: two sample
  - Data: Choose **Sex** for classification variable and **Weight** for analysis variable
  - Plots: select summary plot
- Click on "Run"

21

Nonparametric

- Double-click on HEARTALIVE data
- Check descriptive stats and distribution for **Smoking** variable
- Click on "Analyze" tab → ANOVA → Nonparametric One-Way ANOVA
  - Data: Choose **Smoking** for dependent variable and **BP_status** for independent variable
  - Analysis: Wilcoxon
- Click on "Run"

22

Manipulating Data III

- To create a program in the project
  File → New → Program
- To create new variables, type

```
Data tempheart;
  set 'H:\your project\SAS\DATA\heartalive';
  sqrtsystolic=sqrt(systolic);
  lnsystolic=log(systolic);
  lnsystolic70=log(systolic-70);
run;
```

- Click on tab "Run"
- Examine distributions for new variables

23

Manipulating Data IV

- To clean data in SAS EG directly

Tools→Options

  - Data General: Select "Use data in unprotected mode"
- Create a copy of your original data (export with a new name into DATA folder)
- Open this copy into your project
- Make changes to data
- WARNING – remember to unselect "Use data in unprotected mode" after you are done cleaning.

24

Linear Models

- Double-click on HEARTALIVE data
- Click on "Analyze" tab → ANOVA → Linear Models
  - Data: Choose **Cholesterol** for dependent, **Weight, Diastolic, and Systolic** for quantitative variables
  - Model: Specify model
- Click on "Run"

25

Linear Models (cont.)

- Click on "Modify Task" tab
  - Data: Remove **Weight** from quantitative, and add **Weight_Status** to classification variables
  - Model: Specify model
- Click on "Run"

26

SAS Enterprise Guide Tutorial

- Getting Started with SAS Enterprise Guide
- http://support.sas.com/documentation/onlinedoc/guide/tut51/en/
- http://support.sas.com/eguide

27