**SAS Enterprise Guide**
**9th Annual SAS® Summer Institute**
by the University of Iowa SAS® User Group
August 14-15, 2017
Yelena Perkhounkova
elena-perkhounkova@uiowa.edu

**What is SAS Enterprise Guide?**

A point-and-click graphical interface to SAS that provides

- easy access to data sources/export capabilities

- ready-to-use *tasks* to make reporting & analytics more available

- programming interface

- organizational structure to **projects** that includes

    - SAS programs

    - References to data and associated tasks

    - Results in various formats

    - Logs

    - Relationships among the items above

SAS EG project shows references to data, tasks, and programs in the project tree.

The program, log, output data, and results are available via tabs in the main pane.

**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 to do a lot of different analyses in SAS EG.

**Enterprise Guide Program Editor provides convenient tools for writing code**

- Autocomplete [syntax]

- dynamic syntax tooltips [point cursor at SAS keyword]

- formatting programs to provide consistent spacing [Control-I]

- analyzing program flow
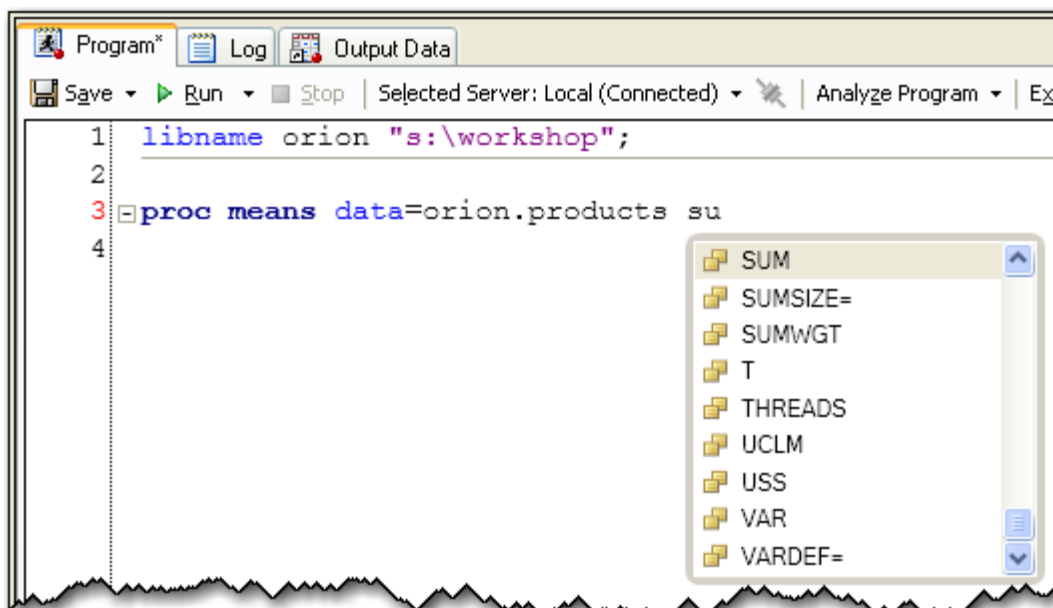
**How you can access SAS EG**

- Virtual Desktop (7.1)

- SAS installed on your PC (7.1)



**Additional information**

- Getting Started with SAS Enterprise Guide Tutorial

    http://support.sas.com/documentation/onlinedoc/guide/tut71/en/

    http://support.sas.com/eguide

**Free SAS e-Learning Courses**

- http://helpdesk.its.uiowa.edu/software/

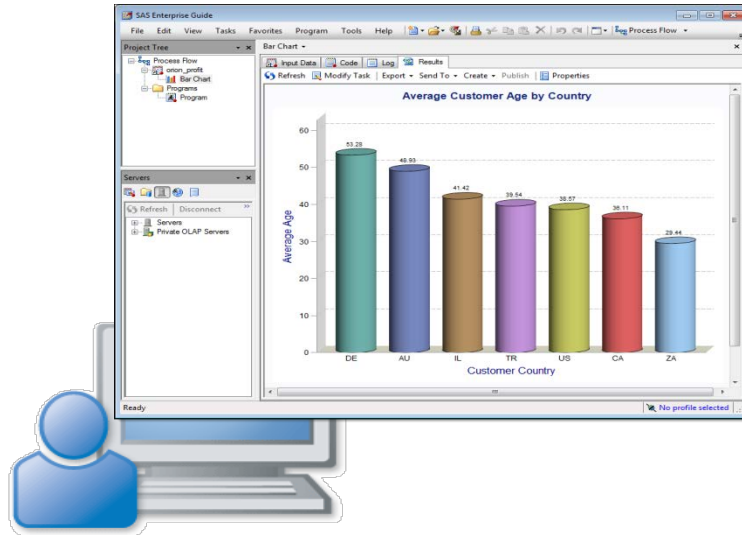- Click on **SAS Self-Paced e-Learning** for instructions (information about available courses is somewhat outdated)

- **To access SAS Self-Paced e-Learning:**

    – Go to http://support.sas.com/myelearn and log into your profile. If you do not have a profile, create one using the link on this page.

    – Enter the activation code **XXXXXXXXX** into the activation code box, then click Submit.

    – Review the license agreement and accept it.

    – Select a course title to start your learning.

**Today I will show you how to use SAS EG to**

- Create and save projects
- Use SAS EG tasks to
    - Add SAS [or Excel] data to the project /export data outside SAS
    - Create subsets of data
    - Summarize data
    - Run simple statistical analyses
- Add programs to the projects
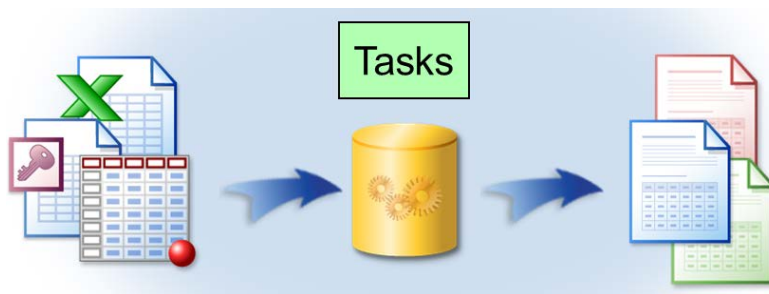


Start by creating a convenient folder structure on C:\SASClass\ [or H-drive to access after class]



Copy SAS datasets (*newheart* and *hr*) to DATA folder.

Open SAS EG 7.1.

Start a new project in EG.

**To open an existing SAS file into the project**

- Select **File → Open → Data**

- Navigate to DATA folder, select & open file NEWHEART

**To explore file properties, right-click on the file in project tree (or click on "properties" tab)**

**To select only the people who were alive at discharge, use the "Filter and Sort" task**

- Select the "Filter and Sort" tab

- Click on the double arrow to select all variables

- Choose filter (**dstat** equal to "0")

- Click "OK"
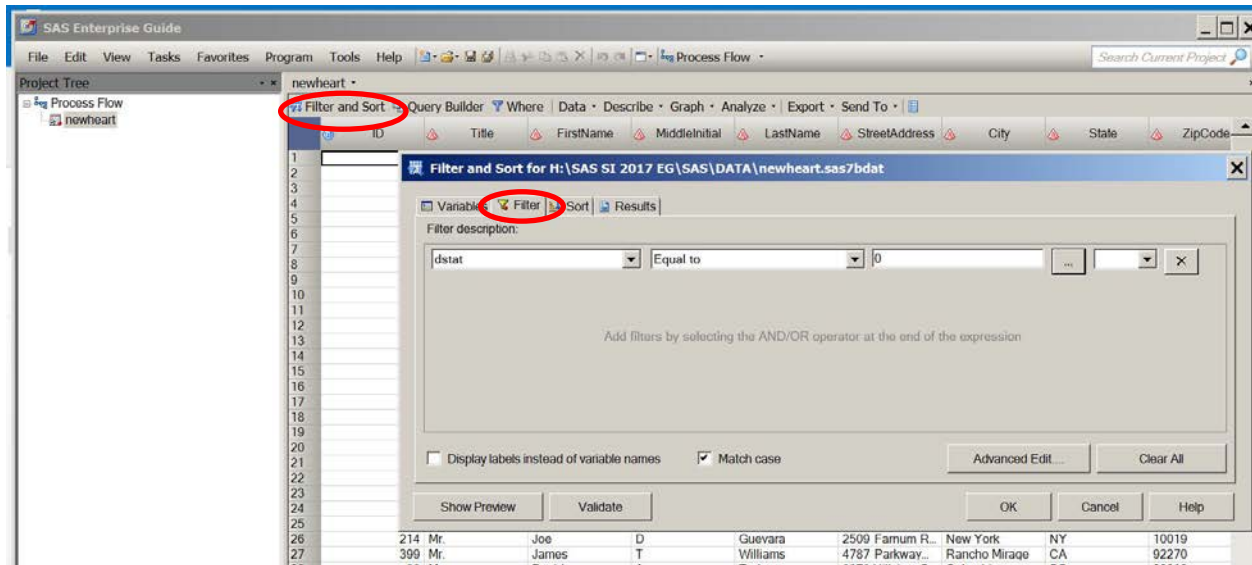


Name the new "Filter and Sort" task "alive at discharge"

**To edit the "Filter and Sort" task, select it**

- Select "Modify Task"

- Sort by ID

- Click "OK"

**To save new data to a permanent location, select the "export" tab**

- Choose "Export 'filename'"

- Navigate to the DATA folder

- Save as HEARTALIVE
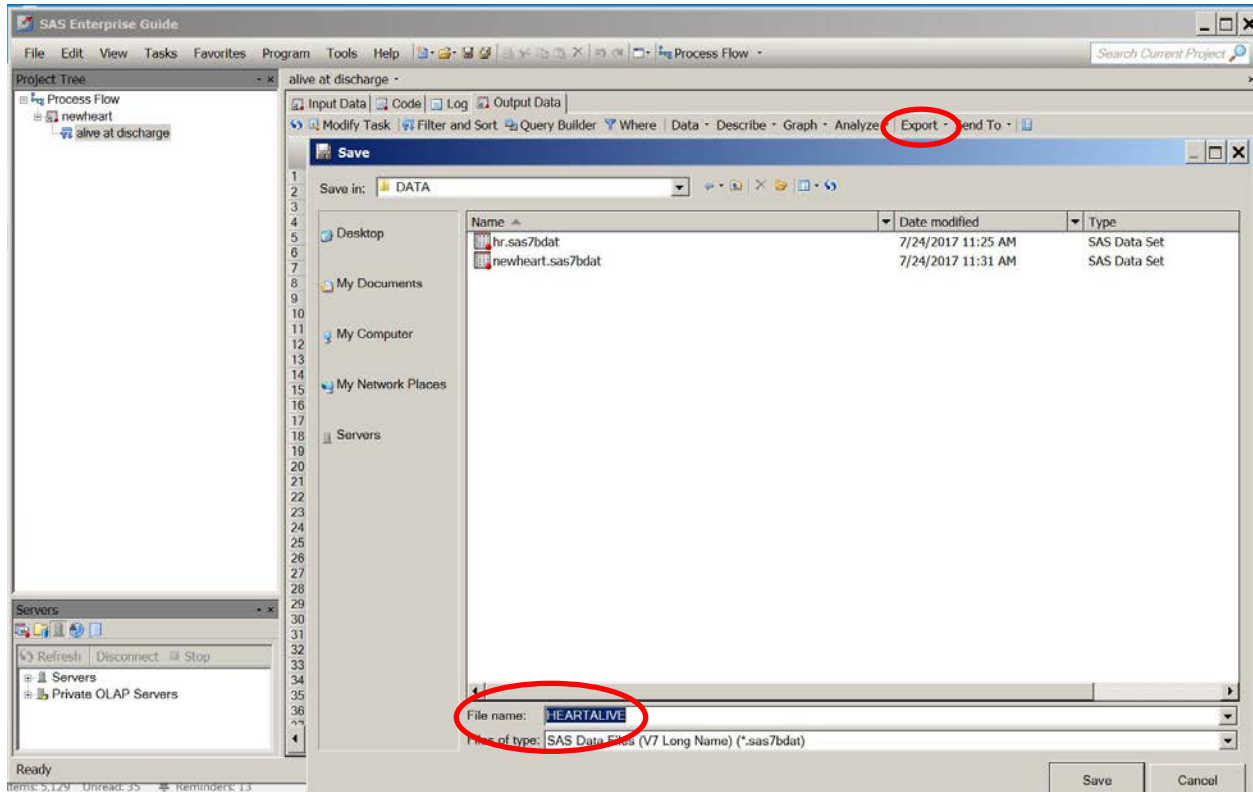


Save your project to \SAS\PROGRAMS

Name the project "SASEG17"

Close the project

Reopen your project.

Open HEARTALIVE data into the project

To select people with BMI ≥ 25 with high heart rate status, create a new "Filter and Sort" task

- Select the "Filter and Sort" tab

- Select all variables

- Choose filters (**bmi** greater than or equal to 25 AND **hr_status** equal to "High")

- Sort by **bmi**

- Name task "highrisk" in "Results" tab

- Click "OK"
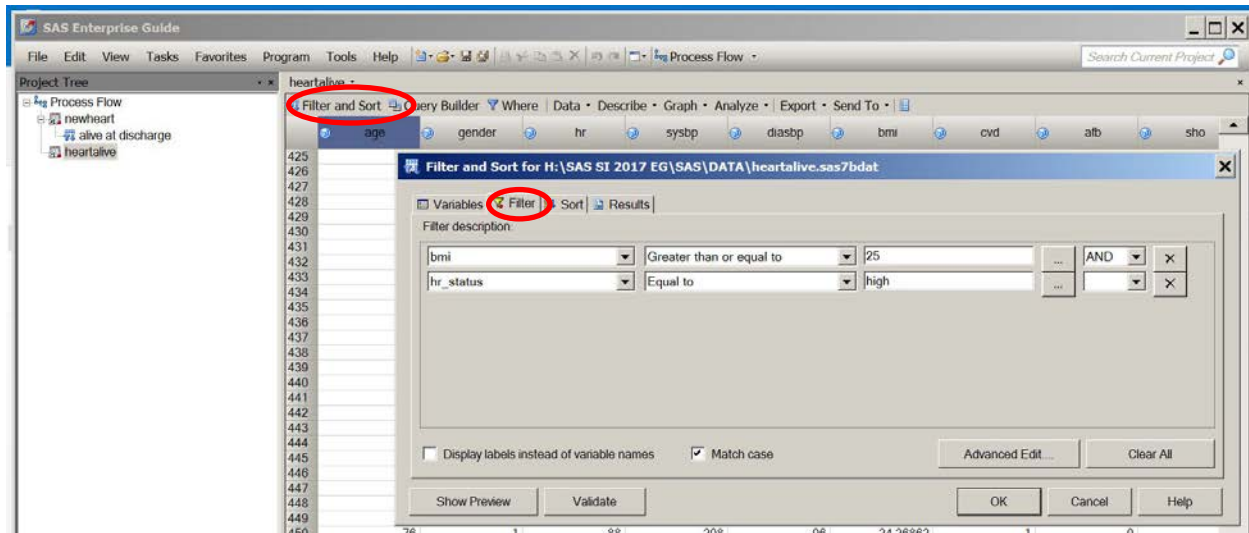
## To save data in Excel format, use the "Export" task

- Select  the "Export" tab

- Choose "Export 'filename'"

- Choose Excel file type

- Name file as HEARTHIGHRISK

- Click "Save"
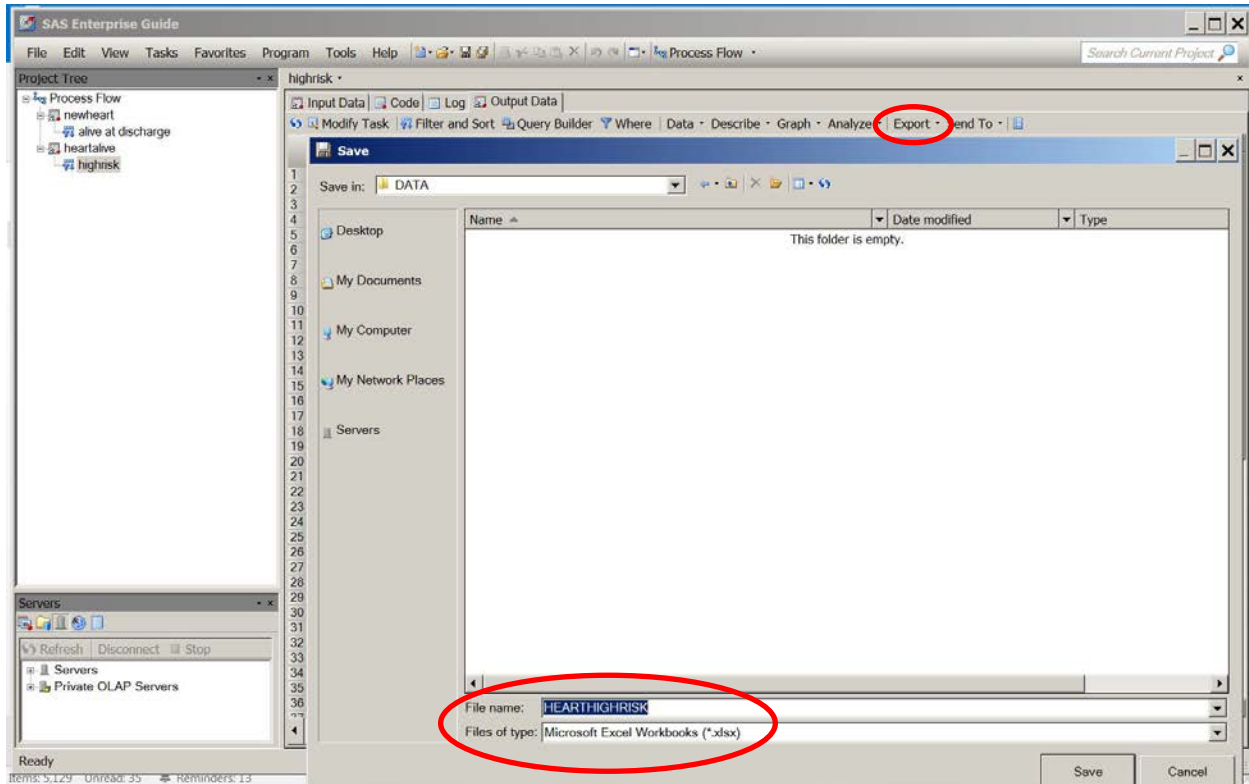


Open the file in Excel

## To open non-SAS data (e.g. Excel) into the project

- Select **File→Import data**

Browse to the DATA folder, select HEARTHIGHRISK.xlsx file, open, follow these 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

Export the new SAS data to your DATA folder with the name "HEARTHIGHRISKfromExcel"

Open "HEARTHIGHRISKfromExcel" data into your project

Compare properties of HEARTALIVE and "HEARTHIGHRISKfromExcel" files

Double-click on HEARTALIVE data to make it active (we will use it from now on)

To calculate summary statistics for continuous variables, use the "Summary statistics" task (PROC MEANS)

- Select **"Describe" tab→Summary Statistics**
    - Data

        Analysis variables: **age, hr, bmi, los**

        Classification variable: **gender**
    - Statistics

        Basic: select mean, standard deviation, min, max, number of observations

        Percentiles: select median

        Additional: select confidence limits of the mean
    - Plots: Select histogram and box-and-whisker plot
    - Titles: "Summary Statistics for Age, Heart Rate, BMI, and Length of Stay"
    - Click "Run"
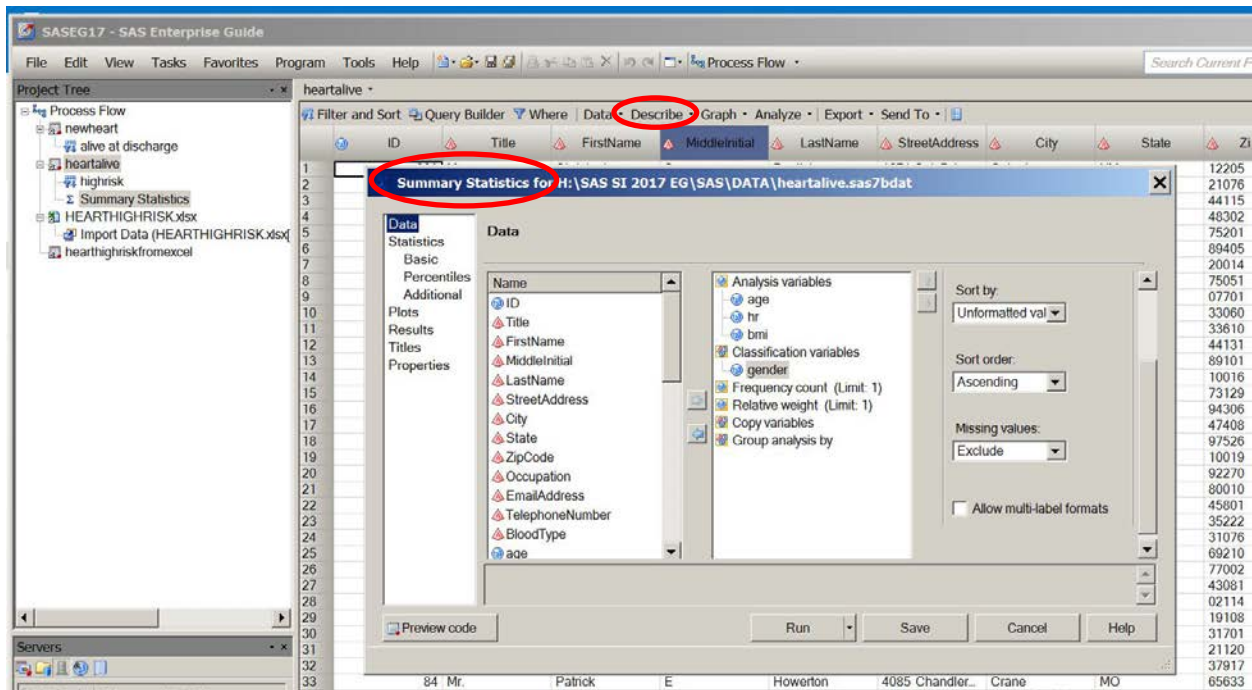


Examine output and code

To edit the task select "Modify Task"

        Statistics: For "Maximum decimal", select 2

Name task "age hr bmi los"

**To examine distributions of continuous variables, use the "Distribution Analysis" task (PROC UNIVARIATE)**

- Double-click on HEARTALIVE data
- Select **"Describe" tab→Distribution Analysis**
  - Data

    Analysis variables**: sysbp, diasbp, los**
  - 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"

Examine distributions

**To create a new SAS program in the project**

- Select **File → New → Program**
- To create new variables, type (see hints below)

```
DATA tempheart;
  SET 'path to your project\SAS\DATA\heartalive';
  los_sqrt=sqrt(los);
  los_ln=log(los+1);
RUN;
```

- Click on tab "Run"

Hints:

- Go to the DATA folder and select and copy the path to the folder, then paste it in place of "*path to your project*"
- Fix the quotes around the path if needed
- Adding 1 to avoid errors related to log transforming zeros
- New variables will be at the end of the <u>temporary</u> SAS data "TEMPHEART"

## To create a permanent SAS data set, edit the program

- Add to your program a LIBNAME statement with reference to your library

    ```
    LIBNAME data 'path to your project\SAS\DATA';
    ```

- Edit DATA statement to

    ```
    DATA data.heartalive;
    ```

- Click on tab "Run"

Double-click on HEARTALIVE in the project tree to make sure the new variables (**los_sqrt** and **los_ln**) are there

Add **los_sqrt** and **los_ln** to the "Distribution Analysis" task (in the project tree) and compare distributions to the original **los**. How do they differ?

To create a user-defined format, add code to the program:

```
PROC FORMAT;
   VALUE gender 0='Male' 1='Female';
RUN;
```

To assign a format to a variable, use the FORMAT statement:

```
   FORMAT gender gender.;
```

To assign labels to variables, use the LABEL statement:

```
   LABEL los='Length of stay';
```

Run the program

Check HEARTALIVE properties

Refresh "age hr bmi los" task to see how the **gender** format is used in the output

Refresh "Distribution Analysis" task to see how **los** label is used in the output

## To create output in other formats, modify SAS EG options

- Select **Tools→Options**

    – Results: Select PDF and RTF

    – Can choose output styles for each output format (e.g. Journal for PDF)

- Click "OK"
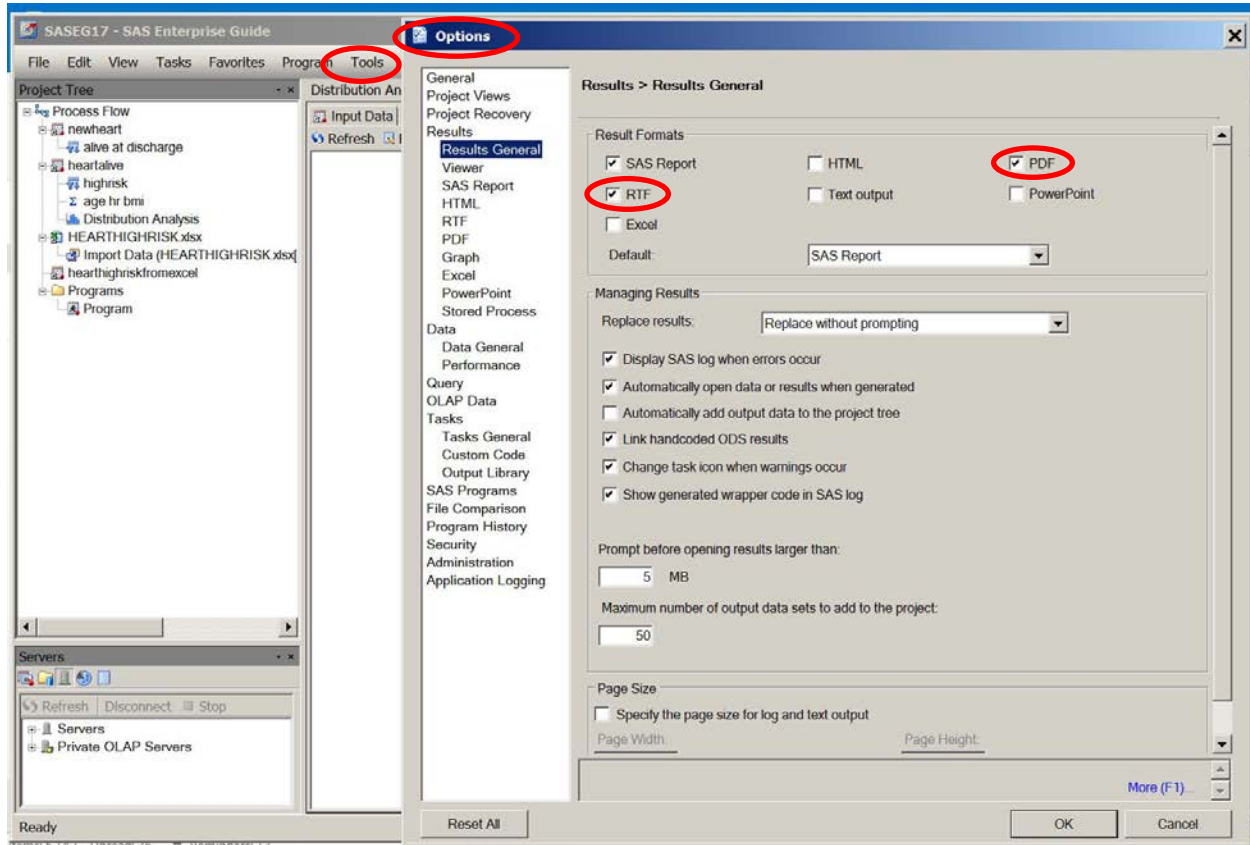


To add PDF and RTF outputs for "age hr bmi los" task to the project, select "Refresh"

Select "Results-RTF" tab, then click on "View"

Save the output to your OUTPUTS folder \SAS\OUTPUTS

**To summarize categorical variables, use the "One-Way Frequencies" task (PROC FREQ)**

- Double-click on HEARTALIVE data to make it active

- Select **"Describe" tab→One-Way Frequencies**

  - Data

    Analysis variables: **State, gender, cvd, year, hr_status**

- Results: select "Create data with frequencies and percentages"

- Click "Run"

Examine output

Check code

**To examine bivariate relationships between continuous variables, use the "Correlations" task (PROC CORR)**

- Select **"Analyze" tab → Multivariate → Correlations**

  - Data

    Analysis variables: **bmi**

    Correlate with: **hr, sysbp, diasbp**

  - Results: Select "Create a scatter plot for each correlation pair"

- Click "Run"

- Use "Modify Task" to run analysis by **gender**

**To compare two independent samples on a continuous normal variable, use the "t test" task (PROC TTEST)**

- Select **"Analyze" tab → ANOVA → t Test**

  - t Test type: two sample

  - Data

    Classification variable: **gender**

    Analysis variables: **hr, bmi, sysbp**

  - Plots: select summary plot

- Click "Run"

Change classification variable to **cvd** (history of cardiovascular disease) and rerun

To compare several independent groups on a continuous normal variable, use the "One-Way ANOVA" task (PROC ANOVA)

Select **"Analyze" tab → ANOVA → One-Way ANOVA**

- Data
    - o Dependent variable: **sysbp, bmi**
    - o Independent variable: **hr_status**
- Means
    - o Comparison: select Bonferroni t test
    - o Breakdown: select Mean, Standard deviation
- Click "Run"

To compare two or more independent samples on a continuous non-normal variable, use the "Nonparametric One-Way ANOVA" task (PROC NPAR1WAY)

- Select **"Analyze" tab → ANOVA → Nonparametric One-Way ANOVA**
    - – Data

      Dependent variables: **los**

      Independent variable: **gender**

- Analysis: uncheck all, but Wilcoxon
- Click "Run"

When more than 2 groups (e.g., **hr_status**), use Kruskal-Wallis test results.

To investigate a bivariate relationship between two categorical variables, use the "Table analysis" task (PROC FREQ)

- Select **"Describe" tab→ Table Analysis**
    - – Data

      Table variables: **gender, cvd, hr_status**

    - – Tables: define 3 tables to be generated (**gender** by **cvd**, **gender** by **hr_status**, **hr_status** by **cvd**)

    - – Cell Statistics: Row percentages, column percentages, cell frequencies

    - – Table Statistics

      Association: check Chi-square tests
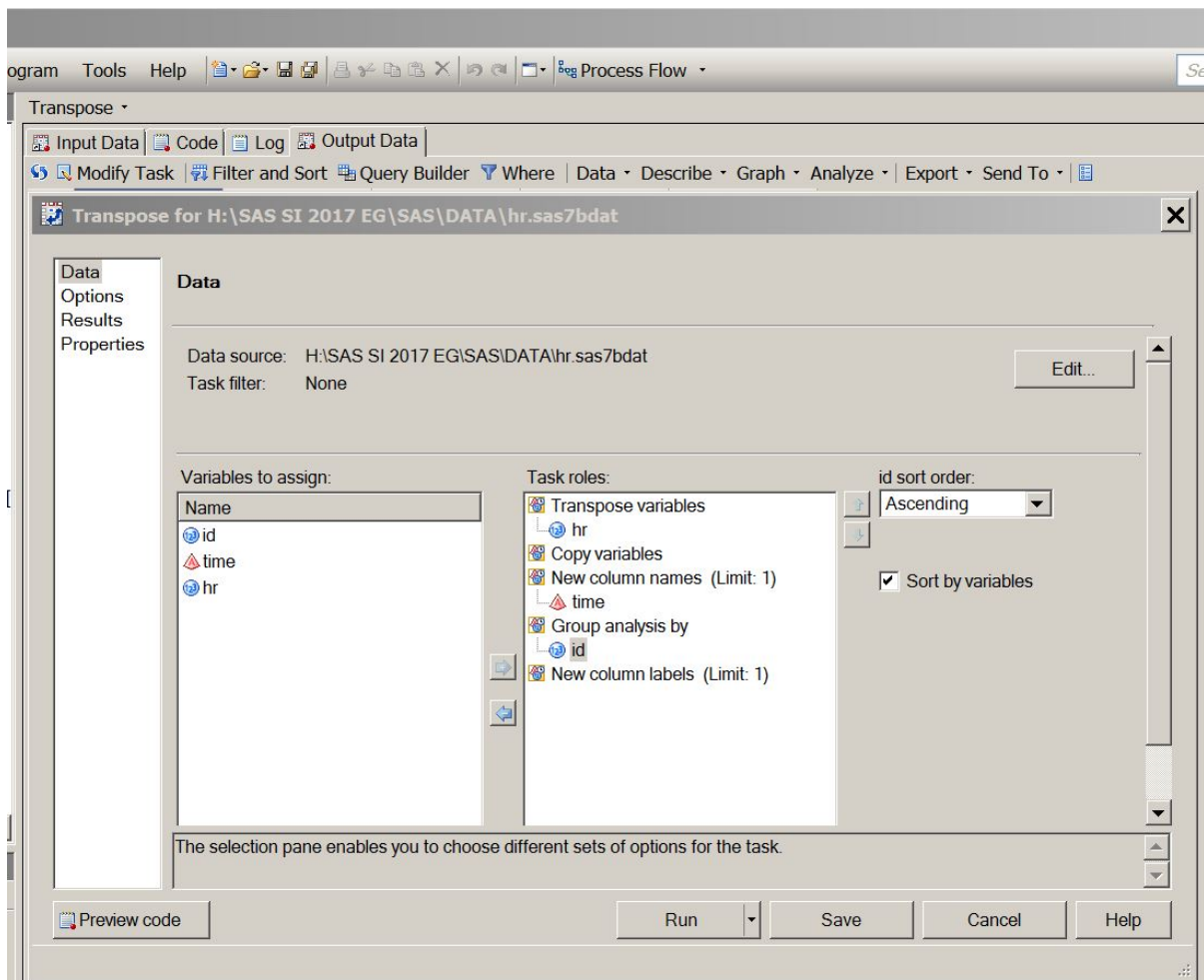
- Click "Run"

Check code

Need data for repeated measures analyses, such as paired samples t-test.

Open SAS dataset HR into the project.

Need to transpose data to compare pre- to post-treatment **hr** using the paired samples t-test.

To transpose data, use the "Transpose" task

- Select **"DATA" tab → Transpose**

  – Data

     Transpose variables: **hr**

     New column names: **time**

     Group analysis by: **id**

  – Options

     Column name prefix: uncheck "Use prefix"



- Click "Run"

Examine output data

Export to your DATA library as hr_trans

Open into the project

To compare two related samples on a continuous normal variable use the "t Test" task.

Select **"Analyze" tab → ANOVA → t Test**

- t Test type: Paired

- Data

  Paired variables: Pre-Treatment, Post-Treatment

- Plots: select "Summary plot"

- Click "Run"


Save your project


SAS How To Tutorials

http://video.sas.com/#category/videos/how-to-tutorials