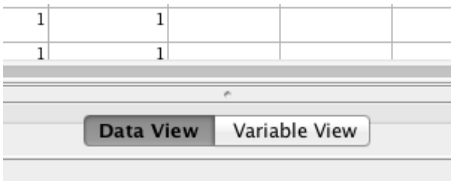
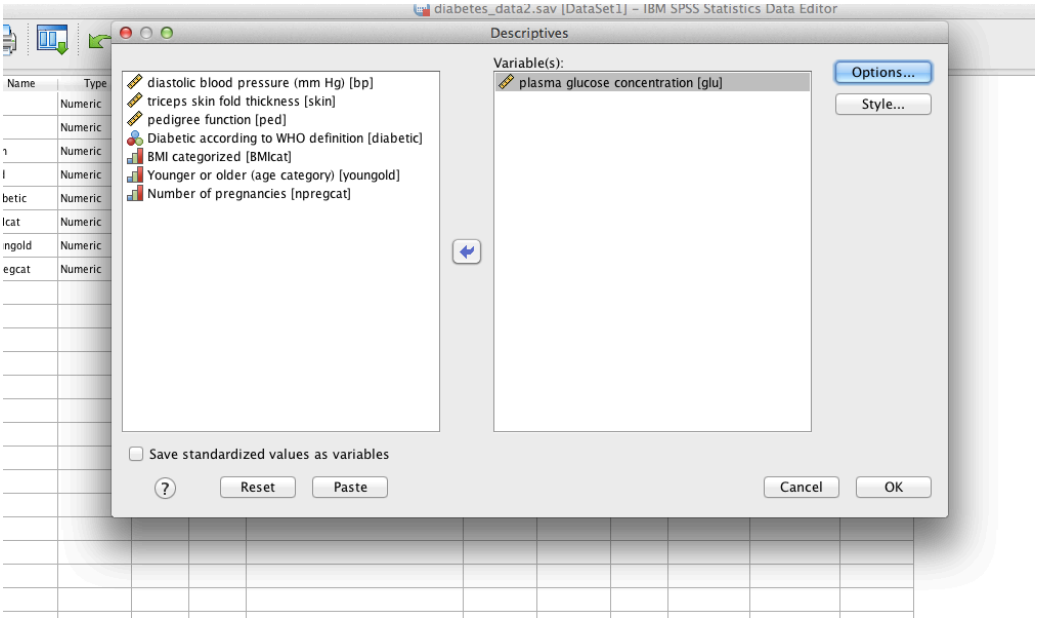
SPSS Tutorial 1: Intro to statistics

1. The dataset for this tutorial is **diabetes_data2.sav** (a *.sav file is an SPSS dataset, if you want to use other software, you can use **diabetes_data2.csv** instead, though some information will be lost (variable descriptions, variable level definitions, etc.))

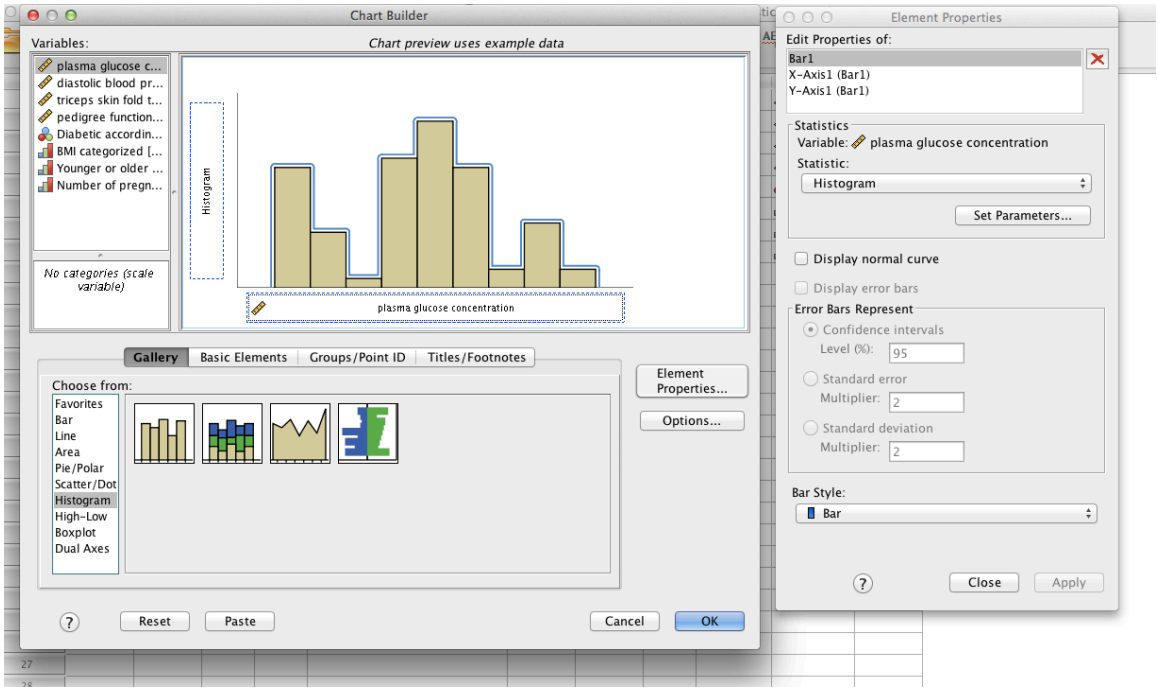
2. At the bottom of the SPSS data window you can see two icons to switch between the data view and variable view. Click on the variable view to get an idea of the kinds of variables available in this dataset.



3. Look at the descriptive statistics for plasma glucose concentration (glu). The function can be found under the menu Analyze->Descriptive Statistics->Descriptives. Move the plasma glucose concentration variable to the box on the right side. Under the options menu, make sure mean, standard deviation, variance, minimum, and maximum are selected. Press "Continue" and "Okay", and see the descriptive statistics in the output window.



4. Plot a histogram of plasma glucose concentration (glu). This can be done through Graphs->Chart Builder. Select the histogram option on the bottom of the window, and drag the first histogram icon to the chart preview area. Then drag plasma glucose concentration to the horizontal axis. Press "Okay" to plot the histogram.



5. Plot a scatterplot with plasma glucose concentration (plu) on the horizontal axis and triceps skinfold thickness (skin) on the vertical axis. To do this, go back to the chart builder and drag the first scatterplot icon to the chart preview area. Drag plasma glucose and triceps skinfold to the appropriate axes and click "Okay" to view the scatterplot.

6. Plot a boxplot of triceps skin fold tickness (skin) with respect to categorized BMI (BMIcat). Again, this can be done using the chart builder, selecting the first boxplot icon, and dragging the appropriate variables to the appropriate axes.

7. Look at the descriptive statistics and create some more plots using one (or more if you want) of the other variables in the dataset.