Lab Assignment 3

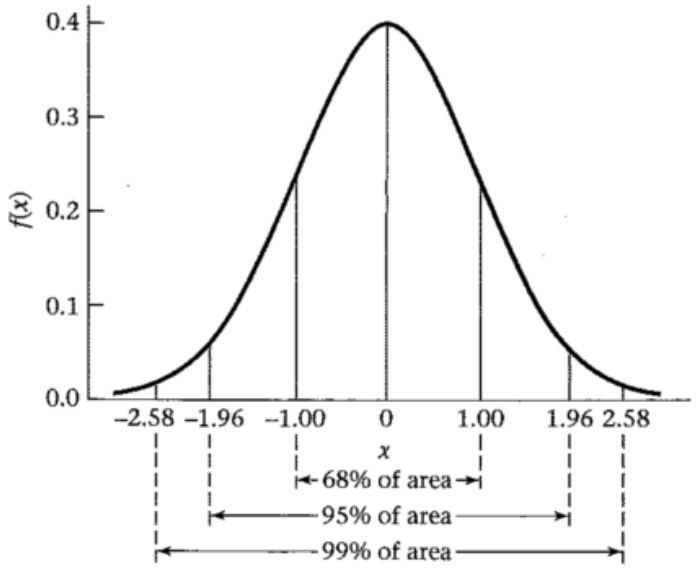
Normal Distributions

This week we analyze normal distributions. Use the new file Valid_ch5.sav that you can download from the lab web page. As before answer questions in the body of the email.

1. Consider the nutrients saturated fat, total fat, alcohol and total calories. Plot the distribution of each nutrient for both the diet record (DR) and the food-frequency questionnaire (FFQ).
   *Do steps a. through c. for **sfat_dr** during the lab demonstration. Do total_ fat and calories on your own.*

   Your assignment will be to do it for the other nutrients. You will try to answer this question: **Do you think a normal distribution is appropriate for these nutrients? For which nutrients? How else are the non-normal data distributed? Do you think the log-transformation makes the data more "normal"?**

   a. Go to Analyze\Descriptive Statistics\Q-Q and check "Estimate from data" and make sure "Normal distribution" is selected in the "Test Distribution" selection box. Create Q-Q plots for **sfat_dr** and later for two other different nutrients including alcoh_dr.
   b. Go to Graphs\Legacy Dialogs\Histogram and check "Display Normal Curve" to create the histograms for **sfat_dr** and later for the different nutrients.
   c. The histograms show the mean and standard deviation. Compute the observed proportion of people for **sfat_dr** who fall within 1.0 standard deviation of the mean. If the data is normally distributed the proportion should be 68%.
      i. You will need to find the values that fall between lowerbound=mean-1 standard deviation and upperbound=mean+1.



Use transform\compute to create a new field onestdev_ sfat_dr with value 1 only for those values of **sfat_dr** which fall between the bounds mentioned previously.

ii. Go to Analyze\Descriptive Statistics\Frequencies and select onestdev_ sfat_dr. Read the Statistics table and calculate Valid number divided by total. If this value is close to 68% then you can say it is normally distributed. Although you would still have to test this for 2 standard deviations (95.45%) and 2.5(98.76%).

iii. Create a new field LN_sfat_dr that will use the natural logarithm transformation for the sfat_dr values. Select LN from the list of functions. The numeric expression should be "LN(sfat_dr)." Make sure "If" includes all cases.

iv. Create a histogram with normal curve. Does this look more normally distributed than the sfat_dr values? Then let's repeat the test from i-iii, in which we compare the observed proportions with the expected proportions based on the assumption of normality.

v. Create a Q-Q plot for the logarithm transformed values. Describe how this plot compares with the untransformed values.

2. Let's take a quick look at alcohol.
    a. Is it normally distributed? How can you find out? (Use the two graphical methods discussed above)
    b. If it is not normally distributed how is it distributed? How do you prove it, using a visual approach?