Lecture Notes for Oct 9, 2012
A few more examples in STATA

Using gss93_data.dta (on resources page)

Histograms and frequencies

1) Show tabulation of **attend**, followed by histogram of **attend**

2) Show tab of **age**, followed by histogram of **age**

3) Features of histogram: options **w(#)** for bucket size; **norm** for normal curve approx; **freq** option.

Getting more descriptives

1) Summarize: sum **age childs**

2) 'detail' option with sum (obtaining more information)

    a. Note that we can see the SD, skew and kurtosis.
    b. A non skew will = 0.
    c. Normal kurtosis = 3.

Using Box Plots

1) graph box **agewed**
2) can look at outliers using: marker(1, mlab(id))
3) A better way is to create a new variable if some condition is met. We know that the outliers are approximately greater than 30 or less than 1 based on the box plot.
    a. Gen problem = 1 if **agewed** > 30 | **agewed** < 1
        i. Note that the pipe symbol "|" means 'or' to STATA.
    b. Now we can display a table of the id's for the problem cases:
        i. tab **id** if problem ==1
        ii. Again, note that this is why we MUST have an id variable for all respondents, cases, etc in our dataset!