Revised July 2018

Correlation and Covariance

This set of notes shows how to use Stata to obtain correlations and covariances. It assumes that you have set Stata up on your computer (see the "Getting Started with Stata" handout), and that you have read in the set of data that you want to analyze (see the "Reading in Stata Format (.dta) Data Files" handout).

In Stata, most tasks can be performed either by issuing commands within the "Stata command" window, **or** by using the menus. These notes illustrate both approaches, using the data file "GSS2016.DTA" (this data file is posted here: https://canvas.harvard.edu/courses/53958).

To obtain Pearson correlations among two or more variables, issue the following command:

```
corr <varlist>
```

where you fill in the names of the variables you want to correlate in place of "<varlist>". You may use several names if you wish, but your "varlist" must always include at least two variables.

For example, to obtain the correlation of the variables educational attainment ("educ") and vocabulary test score ("wordsum"), the command would be (order of the variables doesn't matter):

```
corr wordsum educ
```

with the following results:

```
. corr wordsum educ
(obs=1,861)

             |  wordsum     educ
-------------+------------------
     wordsum |   1.0000
        educ |   0.4433   1.0000
```

Variables always have a perfect (1.0) positive correlation with themselves (diagonal values). The (moderate) education-vocabulary score correlation of 0.44 is at the lower left of the table.

If you want covariances, add the "cov" option after the "varlist"; for this example, the command would be

```
corr wordsum educ,cov
```

with the following results:

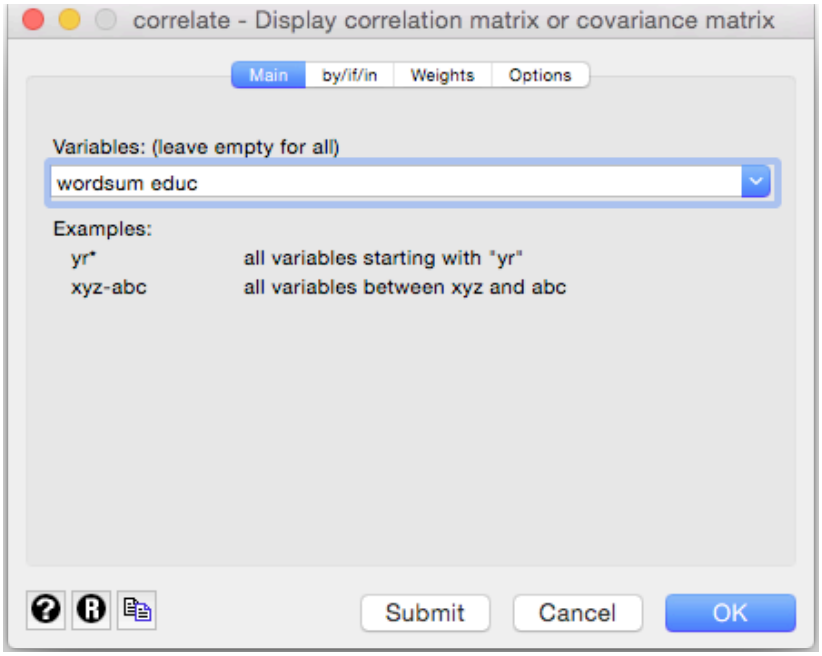
```
. corr wordsum educ, cov
(obs=1,861)

             |  wordsum     educ
-------------+------------------
     wordsum |  3.69042
        educ |  2.43968  8.20695
```

Variances of the variables now appear along the diagonal (recall that a variance is the covariance of a variable with itself); the covariance of education and vocabulary score is at the lower left.

Using the Stata menus, you can obtain correlations and covariances as follows:

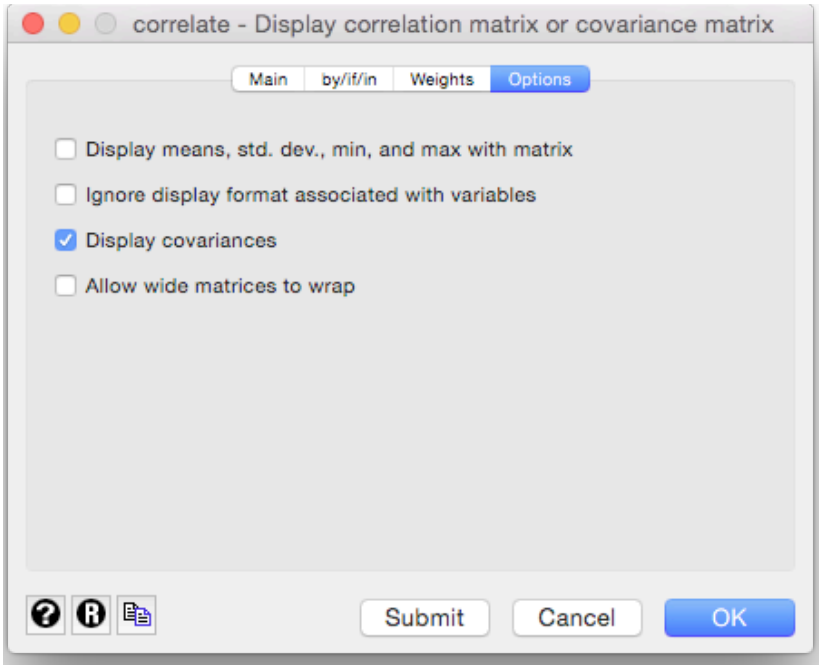
click on "Statistics"
click on "Summaries, tables, & tests"
click on "Summary and descriptive statistics"
click on "Correlations & covariances"

A window like this will open up:



Fill in the names of the variables you want to correlate in the "Variables:" box; this list may be as long as you like. If you click "OK" now, you will see the output for correlations shown on the previous page.

By default, Stata calculates Pearson correlations. If you want it to calculate covariances instead, click on the "Options" tab after you fill in your "varlist" on the above screen. The screen for the "Options" tab looks like this:



Check the "Display covariances" box and click "OK." You will then see the output shown for covariances shown on the second page of these notes.