

Obtaining Tables of Descriptive Statistics, Separately for Groups

This set of notes shows how to use Stata to obtain reports that display descriptive statistics (mean, standard deviation, median, etc.) of a quantitative variable for cases/observations in different groups within a data set. This can be accomplished in several ways, but we will focus here on the “tabstat” command. This handout 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 a table of means for different groups, one issues the “tabstat” command, which has this syntax:

```
tabstat <variable 1>, by(<variable 2>)
```

This calculates (by default) the mean of “variable 1” within groups defined by “variable 2”. For example, to look at differences in the number of siblings (variable “sibs”) for people who hold different highest educational degrees (variable “degree”), one could type

```
tabstat sibs, by(degree)
```

This produces a report like this:

```
Summary for variables: sibs
by categories of: degree (r's highest degree)

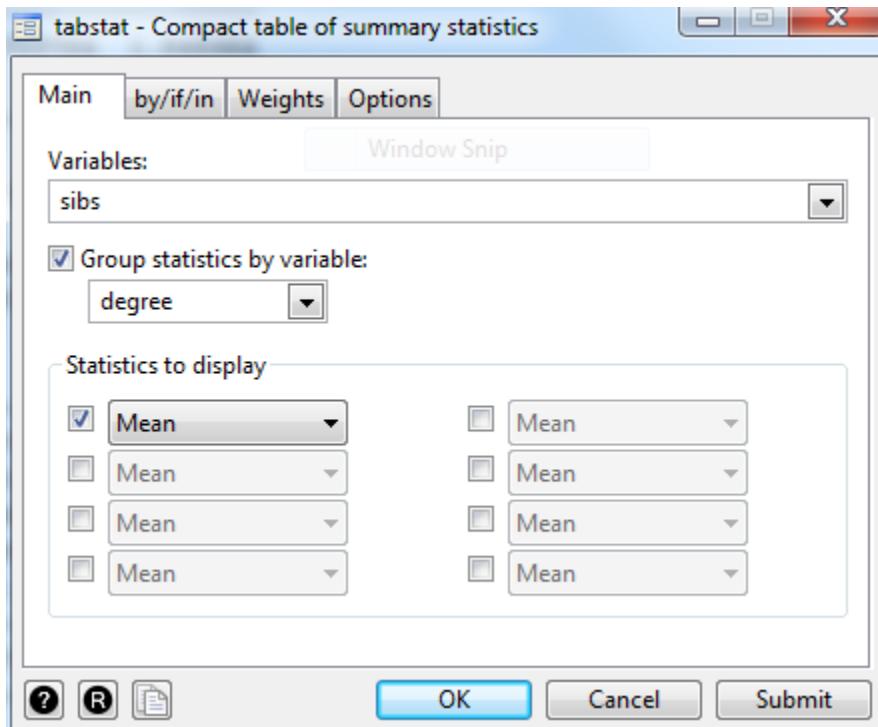
      degree |      mean
-----+-----
  lt high school |  5.789634
    high school |   3.8122
junior college |  3.804651
      bachelor |  2.895522
     graduate |   2.41195
-----+-----
          Total |  3.710784
-----
```

We see that those who have earned higher educational degrees tend to have fewer siblings.

To obtain this report using the Stata menus, proceed as follows:

- Click on “Statistics”
- Click on “Summaries, tables and tests”
- Click on “Other tables”
- Click on “Compact table of summary statistics”

A window like the one shown on the next page will open up:



Enter the quantitative variable for which you want descriptive statistics under “Variables:”, check the “Group statistics by variable:” box, and fill in the name of your grouping variable in the box beneath “Group statistics by variable:”. Check the first box beside “mean” under “Statistics to display”. Then click on “OK” or “Submit” to obtain the report displayed above.

You can obtain means for more than one quantitative variable by listing more than one after the “tabstat” command, or entering more than one in the “Variables:” box in the above menu.

You can obtain a report of statistics other than the mean by listing them in a “statistics” option in the command line, or by checking additional boxes under “Statistics to display” in the above menu and then selecting from among the available statistics in the drop-down list that appears afterwards.

For example, this command

```
tabstat sibs childs, statistics( mean median ) by(degree)
```

displays the means and medians for the numbers of siblings (“sibs”) and children ever born (“child”), for groups of people who have earned different highest degrees. It produces the report displayed on the following page:

Summary statistics: mean, p50
by categories of: degree (r's highest degree)

	degree	sibs	child
lt high school	5.789634	2.806748	
		5	3
high school	3.8122	1.864291	
		3	2
junior college	3.804651	1.767442	
		3	2
bachelor	2.895522	1.453358	
		2	1
graduate	2.41195	1.518868	
		2	2
Total	3.710784	1.848984	
		3	2

Each box shows descriptive statistics for the variable in the column, calculated within the group in the row; the mean is above the median.

The menu needed to obtain this latter report would look like this:

