# Description

`tabstat` displays summary statistics for a series of numeric variables in one table. It allows you to specify the list of statistics to be displayed. Statistics can be calculated (conditioned on) another variable. `tabstat` allows substantial flexibility in terms of the statistics presented and the format of the table.

## Quick start

- Mean of v1 displayed using v1’s display format
  ``` stata
  tabstat v1, format
  ```

- As above, but use format with 2 significant digits and a comma
  ``` stata
  tabstat v1, format(%9.2fc)
  ```

- Nonmissing observations, mean, standard error, and coefficient of variation for v1
  ``` stata
  tabstat v1, statistics(n mean semean cv)
  ```

- Quartiles and interquartile range of v1 and v2
  ``` stata
  tabstat v1 v2, statistics(q iqr)
  ```

- As above, but report statistics separately for each level of catvar
  ``` stata
  tabstat v1 v2, by(catvar) statistics(q iqr)
  ```

- As above, but display a separate column for each statistic
  ``` stata
  tabstat v1 v2, by(catvar) statistics(q iqr) columns(statistics)
  ```

## Menu

Statistics > Summaries, tables, and tests > Other tables > Compact table of summary statistics
Syntax

```
tabstat varlist [ if ] [ in ] [ weight ] [ , options ]
```

**options**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>group statistics by variable</td>
</tr>
<tr>
<td>report specified statistics</td>
</tr>
</tbody>
</table>

**Main**

- `by(varname)`
  - group statistics by variable
- `statistics(statname[...])`
  - report specified statistics

**Options**

- `labelwidth(#)`
  - width for by() variable labels; default is labelwidth(16)
- `varwidth(#)`
  - variable width; default is varwidth(12)
- `columns(variables)`
  - display variables in table columns; the default
- `columns(statistics)`
  - display statistics in table columns
- `format[ (%fmt) ]`
  - display format for statistics; default format is %9.0g
- `casewise`
  - perform casewise deletion of observations
- `nototal`
  - do not report overall statistics; use with by()
- `missing`
  - report statistics for missing values of by() variable
- `noseparator`
  - do not use separator line between by() categories
- `longstub`
  - make left table stub wider
- `save`
  - store summary statistics in r()

**Options**

- `by(varname)` specifies that the statistics be displayed separately for each unique value of `varname`; `varname` may be numeric or string. For instance, `tabstat height` would present the overall mean of height. `tabstat height, by(sex)` would present the mean height of males, and of females, and the overall mean height. Do not confuse the by() option with the by prefix (see [D] by); both may be specified.

- `statistics(statname[...])` specifies the statistics to be displayed; the default is equivalent to specifying `statistics(mean)`. (stats() is a synonym for statistics().) Multiple statistics may be specified and are separated by white space, such as `statistics(mean sd)`. Available statistics are

- `labelwidth(#)` width for by() variable labels; default is labelwidth(16)
- `varwidth(#)` variable width; default is varwidth(12)
- `columns(variables)` display variables in table columns; the default
- `columns(statistics)` display statistics in table columns
- `format[ (%fmt) ]` display format for statistics; default format is %9.0g
- `casewise` perform casewise deletion of observations
- `nototal` do not report overall statistics; use with by()
- `missing` report statistics for missing values of by() variable
- `noseparator` do not use separator line between by() categories
- `longstub` make left table stub wider
- `save` store summary statistics in r()
### tabstat — Compact table of summary statistics

<table>
<thead>
<tr>
<th>statname</th>
<th>Definition</th>
<th>statname</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>mean</td>
<td>p1</td>
<td>1st percentile</td>
</tr>
<tr>
<td>count</td>
<td>count of nonmissing</td>
<td>p5</td>
<td>5th percentile</td>
</tr>
<tr>
<td>n</td>
<td>same as count</td>
<td>p10</td>
<td>10th percentile</td>
</tr>
<tr>
<td>sum</td>
<td>sum</td>
<td>p25</td>
<td>25th percentile</td>
</tr>
<tr>
<td>max</td>
<td>maximum</td>
<td>median</td>
<td>median (same as p50)</td>
</tr>
<tr>
<td>min</td>
<td>minimum</td>
<td>p50</td>
<td>50th percentile (same as median)</td>
</tr>
<tr>
<td>range</td>
<td>range = max − min</td>
<td>p75</td>
<td>75th percentile</td>
</tr>
<tr>
<td>sd</td>
<td>standard deviation</td>
<td>p90</td>
<td>90th percentile</td>
</tr>
<tr>
<td>variance</td>
<td>variance</td>
<td>p95</td>
<td>95th percentile</td>
</tr>
<tr>
<td>cv</td>
<td>coefficient of variation (sd/mean)</td>
<td>p99</td>
<td>99th percentile</td>
</tr>
<tr>
<td>semean</td>
<td>standard error of mean (sd/√n)</td>
<td>iqr</td>
<td>interquartile range = p75 − p25</td>
</tr>
<tr>
<td>skewness</td>
<td>skewness</td>
<td>q</td>
<td>equivalent to specifying p25 p50 p75</td>
</tr>
<tr>
<td>kurtosis</td>
<td>kurtosis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Options**

- `labelwidth(#)` specifies the maximum width to be used within the stub to display the labels of the `by()` variable. The default is `labelwidth(16)`. 8 ≤ # ≤ 32.

- `varwidth(#)` specifies the maximum width to be used within the stub to display the names of the variables. The default is `varwidth(12)`. `varwidth()` is effective only with `columns(statistics)`. Setting `varwidth()` implies `longstub`. 8 ≤ # ≤ 32.

- `columns(variables | statistics)` specifies whether to display variables or statistics in the columns of the table. `columns(variables)` is the default when more than one variable is specified.

- `format` and `format(%,fmt)` specify how the statistics are to be formatted. The default is to use a `%9.0g` format.

  - `format` specifies that each variable’s statistics be formatted with the variable’s display format; see [D] `format`.
  
  - `format(%,fmt)` specifies the format to be used for all statistics.

  The column width is the maximum width of these formats. The minimum column width is nine display characters.

- `casewise` specifies casewise deletion of observations. Statistics are to be computed for the sample that is not missing for any of the variables in `varlist`. The default is to use all the nonmissing values for each variable.

- `nototal` is for use with `by()`; it specifies that the overall statistics not be reported.

- `missing` specifies that missing values of the `by()` variable be treated just like any other value and that statistics should be displayed for them. The default is not to report the statistics for the `by()==missing` group. If the `by()` variable is a string variable, `by()==""` is considered to mean missing.

- `noseparator` specifies that a separator line between the `by()` categories not be displayed.

- `longstub` specifies that the left stub of the table be made wider so that it can include names of the statistics or variables in addition to the categories of `by(varname)`. The default is to describe the statistics or variables in a header. `longstub` is ignored if `by(varname)` is not specified.

- `save` specifies that the summary statistics be returned in `r()`. The overall (unconditional) statistics are returned in matrix `r(StatTotal)` (rows are statistics, columns are variables). The conditional statistics are returned in the matrices `r(Stat1), r(Stat2), ...,` and the names of the corresponding variables are returned in the macros `r(name1), r(name2), ...`
Remarks and examples

This command is probably most easily understood by going through a series of examples.

> Example 1

We have data on the price, weight, mileage rating, and repair record of 22 foreign and 52 domestic 1978 automobiles. We want to summarize these variables for the different origins of the automobiles.

```
. use https://www.stata-press.com/data/r16/auto
    (1978 Automobile Data)
. tabstat price weight mpg rep78, by(foreign)
  Summary statistics: mean
  by categories of: foreign (Car type)
foreign |   price    weight    mpg    rep78
---------|-----------|-----------|-------|--------
  Domestic | 6072.423  3317.115  19.82692  3.020833
  Foreign   | 6384.682  2315.909  24.77273  4.285714
  Total     | 6165.257  3019.459  21.2973   3.405797
```

More summary statistics can be requested via the `statistics()` option. The group totals can be suppressed with the `nototal` option.

```
. tabstat price weight mpg rep78, by(foreign) stat(mean sd min max) nototal
  Summary statistics: mean, sd, min, max
  by categories of: foreign (Car type)
foreign |   price    weight    mpg    rep78    min    max
---------|-----------|-----------|-------|--------|-------|------
  Domestic | 6072.423  3317.115  19.82692  3.020833 3097.104  695.3637  84.743297  .837666
            | 3291      1800      12        115906    4840    34      5
  Foreign   | 6384.682  2315.909  24.77273  4.285714 2621.915  433.0035  6.611187  .7171372
            | 3748      1760      14        3       12990    3420    41      5
```

Although the header of the table describes the statistics running vertically in the “cells”, the table may become hard to read, especially with many variables or statistics. The `longstub` option specifies that a column be added describing the contents of the cells. The `format` option can be issued to specify that `tabstat` display the statistics by using the display format of the variables rather than the overall default `%9.0g`. 
We can specify a layout of the table in which the statistics run horizontally and the variables run vertically by specifying the `col(statistics)` option.

```
. tabstat price weight mpg rep78, by(foreign) stat(min mean max) col(stat) long

<table>
<thead>
<tr>
<th>variable</th>
<th>price</th>
<th>weight</th>
<th>mpg</th>
<th>rep78</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>price</td>
<td>3291</td>
<td>6072.423</td>
<td>15906</td>
<td></td>
</tr>
<tr>
<td>weight</td>
<td>1800</td>
<td>3317.115</td>
<td>4840</td>
<td></td>
</tr>
<tr>
<td>mpg</td>
<td>12</td>
<td>19.82692</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>rep78</td>
<td>1</td>
<td>3.020833</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Foreign</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>price</td>
<td>3748</td>
<td>6384.682</td>
<td>12990</td>
<td></td>
</tr>
<tr>
<td>weight</td>
<td>1760</td>
<td>2315.909</td>
<td>3420</td>
<td></td>
</tr>
<tr>
<td>mpg</td>
<td>14</td>
<td>24.77273</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>rep78</td>
<td>3</td>
<td>4.285714</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>price</td>
<td>3291</td>
<td>6165.257</td>
<td>15906</td>
<td></td>
</tr>
<tr>
<td>weight</td>
<td>1760</td>
<td>3019.459</td>
<td>4840</td>
<td></td>
</tr>
<tr>
<td>mpg</td>
<td>12</td>
<td>21.2973</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>rep78</td>
<td>1</td>
<td>3.405797</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
```

Finally, `tabstat` can also be used to enhance summarize so we can specify the statistics to be displayed. For instance, we can display the number of observations, the mean, the coefficient of variation, and the 25%, 50%, and 75% quantiles for a list of variables.

```
. tabstat price weight mpg rep78, stat(n mean cv q) col(stat) long

<table>
<thead>
<tr>
<th>variable</th>
<th>N</th>
<th>mean</th>
<th>cv</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
</tr>
</thead>
<tbody>
<tr>
<td>price</td>
<td>74</td>
<td>6165.257</td>
<td>.478406</td>
<td>4195</td>
<td>5006.5</td>
<td>6342</td>
</tr>
<tr>
<td>weight</td>
<td>74</td>
<td>3019.459</td>
<td>.2573949</td>
<td>2240</td>
<td>3190</td>
<td>3600</td>
</tr>
<tr>
<td>mpg</td>
<td>74</td>
<td>21.2973</td>
<td>.2716543</td>
<td>18</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>rep78</td>
<td>69</td>
<td>3.405797</td>
<td>.290661</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
```

Because we did not specify the `by()` option, these statistics were not displayed for the subgroups of the data formed by the categories of the `by()` variable.
Video example

Descriptive statistics in Stata

Acknowledgments

The `tabstat` command was written by Jeroen Weesie and Vincent Buskens both of the Department of Sociology at Utrecht University, The Netherlands.

Reference


Also see

[R] `summarize` — Summary statistics
[R] `table` — Flexible table of summary statistics
[R] `tabulate, summarize()` — One- and two-way tables of summary statistics
[D] `collapse` — Make dataset of summary statistics