Title

> **tabulate, summarize()** — One- and two-way tables of summary statistics

Syntax   Menu   Description   Options
Remarks and examples   Also see

Syntax

> <u>ta</u>bulate *varname₁* [*varname₂*] [*if*] [*in*] [*weight*] [, *options*]

options	Description
Main	
<u>summarize</u>(*varname₃*)	report summary statistics for *varname₃*
[<u>no</u>]means	include or suppress means
[<u>no</u>]standard	include or suppress standard deviations
[<u>no</u>]freq	include or suppress frequencies
[<u>no</u>]obs	include or suppress number of observations
<u>nol</u>abel	show numeric codes, not labels
wrap	do not break wide tables
missing	treat missing values of *varname₁* and *varname₂* as categories

by is allowed; see [D] **by**.

aweights and fweights are allowed; see [U] **11.1.6 weight**.

Menu

Statistics > Summaries, tables, and tests > Other tables > Table of means, std. dev., and frequencies

Description

tabulate, summarize() produces one- and two-way tables (breakdowns) of means and standard deviations. See [R] **tabulate oneway** and [R] **tabulate twoway** for one- and two-way frequency tables. See [R] **table** for a more flexible command that produces one-, two-, and n-way tables of frequencies and a wide variety of summary statistics. table is better, but tabulate, summarize() is faster. Also see [R] **tabstat** for yet another alternative.

Options

> **Main**

summarize(*varname₃*) identifies the name of the variable for which summary statistics are to be reported. If you do not specify this option, a table of frequencies is produced; see [R] **tabulate oneway** and [R] **tabulate twoway**. The description here concerns tabulate when this option is specified.

1

[no]means includes or suppresses only the means from the table.

> The summarize() table normally includes the mean, standard deviation, frequency, and, if the data are weighted, number of observations. Individual elements of the table may be included or suppressed by the [no]means, [no]standard, [no]freq, and [no]obs options. For example, typing
>
> > . tabulate category, summarize(myvar) means standard
>
> produces a summary table by category containing only the means and standard deviations of myvar. You could also achieve the same result by typing
>
> > . tabulate category, summarize(myvar) nofreq

[no]standard includes or suppresses only the standard deviations from the table; see [no]means option above.

[no]freq includes or suppresses only the frequencies from the table; see [no]means option above.

[no]obs includes or suppresses only the reported number of observations from the table. If the data are not weighted, the number of observations is identical to the frequency, and by default only the frequency is reported. If the data are weighted, the frequency refers to the sum of the weights. See [no]means option above.

nolabel causes the numeric codes to be displayed rather than the label values.

wrap requests that no action be taken on wide tables to make them readable. Unless wrap is specified, wide tables are broken into pieces to enhance readability.

missing requests that missing values of *varname*$_1$ and *varname*$_2$ be treated as categories rather than as observations to be omitted from the analysis.

Remarks and examples

tabulate with the summarize() option produces one- and two-way tables of summary statistics. When combined with the by prefix, it can produce n-way tables as well.

Remarks are presented under the following headings:

> *One-way tables*
> *Two-way tables*

One-way tables

▷ Example 1

We have data on 74 automobiles. Included in our dataset are the variables foreign, which marks domestic and foreign cars, and mpg, the car's mileage rating. Typing tabulate foreign displays a breakdown of the number of observations we have by the values of the foreign variable.

```
. use http://www.stata-press.com/data/r13/auto
(1978 Automobile Data)

. tabulate foreign
```

Car type	Freq.	Percent	Cum.
Domestic	52	70.27	70.27
Foreign	22	29.73	100.00
Total	74	100.00	

We discover that we have 52 domestic cars and 22 foreign cars in our dataset. If we add the summarize(*varname*) option, however, tabulate produces a table of summary statistics for *varname*:

```
. tabulate foreign, summarize(mpg)
              Summary of Mileage (mpg)
   Car type        Mean   Std. Dev.        Freq.

   Domestic   19.826923   4.7432972           52
    Foreign   24.772727   6.6111869           22

      Total   21.297297   5.7855032           74
```

We also discover that the average gas mileage for domestic cars is about 20 mpg and the average foreign is almost 25 mpg. Overall, the average is 21 mpg in our dataset.

◁

❑ Technical note

We might now wonder if the difference in gas mileage between foreign and domestic cars is statistically significant. We can use the oneway command to find out; see [R] **oneway**. To obtain an analysis-of-variance table of mpg on foreign, we type

```
. oneway mpg foreign
                     Analysis of Variance
    Source            SS       df       MS          F      Prob > F

Between groups    378.153515     1   378.153515    13.18     0.0005
 Within groups    2065.30594    72   28.6848048

     Total        2443.45946    73   33.4720474

Bartlett's test for equal variances:  chi2(1) =   3.4818  Prob>chi2 = 0.062
```

The F statistic is 13.18, and the difference between foreign and domestic cars' mileage ratings is significant at the 0.05% level.

There are several ways that we could have statistically compared mileage ratings — see, for instance, [R] **anova**, [R] **oneway**, [R] **regress**, and [R] **ttest** — but oneway seemed the most convenient.

❑

Two-way tables

▷ Example 2

tabulate, summarize can be used to obtain two-way as well as one-way breakdowns. For instance, we obtained summary statistics on mpg decomposed by foreign by typing tabulate foreign, summarize(mpg). We can specify up to two variables before the comma:

```
. generate wgtcat = autocode(weight,4,1760,4840)
. tabulate wgtcat foreign, summarize(mpg)
        Means, Standard Deviations and Frequencies of Mileage (mpg)
                     Car type
  wgtcat │   Domestic    Foreign │     Total
─────────┼────────────────────────┼──────────
    2530 │  28.285714    27.0625 │ 27.434783
         │  3.0937725  5.9829619 │ 5.2295149
         │          7         16 │        23
─────────┼────────────────────────┼──────────
    3300 │      21.75       19.6 │ 21.238095
         │  2.4083189  3.4351128 │ 2.7550819
         │         16          5 │        21
─────────┼────────────────────────┼──────────
    4070 │   17.26087         14 │    17.125
         │  1.8639497          0 │ 1.9406969
         │         23          1 │        24
─────────┼────────────────────────┼──────────
    4840 │  14.666667          . │ 14.666667
         │    3.32666          . │   3.32666
         │          6          0 │         6
─────────┼────────────────────────┼──────────
   Total │  19.826923  24.772727 │ 21.297297
         │  4.7432972  6.6111869 │ 5.7855032
         │         52         22 │        74
```

In addition to the means, standard deviations, and frequencies for each weight–mileage cell, also reported are the summary statistics by weight, by mileage, and overall. For instance, the last row of the table reveals that the average mileage of domestic cars is 19.83 and that of foreign cars is 24.77—domestic cars yield poorer mileage than foreign cars. But we now see that domestic cars yield better gas mileage within weight class—the reason domestic cars yield poorer gas mileage is because they are, on average, heavier.

◁

▷ Example 3

If we do not specify the statistics to be included in a table, tabulate reports the mean, standard deviation, and frequency. We can specify the statistics that we want to see using the means, standard, and freq options:

```
. tabulate wgtcat foreign, summarize(mpg) means
               Means of Mileage (mpg)
                     Car type
  wgtcat │   Domestic    Foreign │     Total
─────────┼────────────────────────┼──────────
    2530 │  28.285714    27.0625 │ 27.434783
    3300 │      21.75       19.6 │ 21.238095
    4070 │   17.26087         14 │    17.125
    4840 │  14.666667          . │ 14.666667
─────────┼────────────────────────┼──────────
   Total │  19.826923  24.772727 │ 21.297297
```

When we specify one or more of the means, standard, and freq options, only those statistics are displayed. Thus we could obtain a table containing just the means and standard deviations by typing means standard after the summarize(mpg) option. We can also suppress selected statistics by placing no in front of the option name. Another way of obtaining only the means and standard deviations is to add the nofreq option:

```
. tabulate wgtcat foreign, summarize(mpg) nofreq
            Means and Standard Deviations of Mileage (mpg)
                       Car type
   wgtcat │   Domestic     Foreign │      Total
──────────┼──────────────────────┼────────────
     2530 │  28.285714    27.0625 │  27.434783
          │  3.0937725  5.9829619 │  5.2295149
──────────┼──────────────────────┼────────────
     3300 │      21.75       19.6 │  21.238095
          │  2.4083189  3.4351128 │  2.7550819
──────────┼──────────────────────┼────────────
     4070 │   17.26087         14 │     17.125
          │  1.8639497          0 │  1.9406969
──────────┼──────────────────────┼────────────
     4840 │  14.666667          . │  14.666667
          │    3.32666          . │    3.32666
──────────┼──────────────────────┼────────────
    Total │  19.826923  24.772727 │  21.297297
          │  4.7432972  6.6111869 │  5.7855032
```

◁

Also see

[R] **table** — Flexible table of summary statistics

[R] **tabstat** — Compact table of summary statistics

[R] **tabulate oneway** — One-way table of frequencies

[R] **tabulate twoway** — Two-way table of frequencies

[D] **collapse** — Make dataset of summary statistics

[SVY] **svy: tabulate oneway** — One-way tables for survey data

[SVY] **svy: tabulate twoway** — Two-way tables for survey data

[U] **12.6 Dataset, variable, and value labels**

[U] **25 Working with categorical data and factor variables**