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

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.

Quick start

Tabulation of `v1`, reporting means and standard deviations of `x` and frequencies

```
tabulate v1, summarize(x)
```

Same as above, but report summary statistics for the two-way tabulation of `v1` and `v2`

```
tabulate v1 v2, summarize(x)
```

Weighted summary statistics using frequency weight `wvar`

```
tabulate v1 v2 [fweight=wvar], summarize(x)
```

Report only the mean of `x` for each group

```
tabulate v1 v2, summarize(x) means
```

Do not report standard deviations

```
tabulate v1 v2, summarize(x) nostandard
```

Show numeric values of `v1` and `v2` rather than value labels

```
tabulate v1 v2, summarize(x) nolabel
```

Menu

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

Syntax

```
tabulate varname1 [varname2] [if] [in] [weight] [, options]
```

options	Description
Main	
summarize(*varname*$_3$)	report summary statistics for *varname*$_3$
[no]means	include or suppress means
[no]standard	include or suppress standard deviations
[no]freq	include or suppress frequencies
[no]obs	include or suppress number of observations
nolabel	show numeric codes, not labels
wrap	do not break wide tables
missing	treat missing values of *varname*$_1$ and *varname*$_2$ as categories
Collect	
collect	post results to collection Tabulate
collect([*cname*][, *collect_options*])	post results to a named collection

collect_options	Description
append	append results to an existing collection
replace	replace results of an existing collection
label(*filename*)	specify the collection labels
style(*filename*[, override])	specify the collection style

by is allowed; see **[U] 11.1.10 Prefix commands**.

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

Options

Main

summarize(*varname*$_3$) 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.

[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.

Collect

collect and collect([*cname*][, *collect_options*]) specify that results be posted to a collection. This collection produces a table that you can customize and publish to Microsoft Word, Microsoft Excel, PDF, HTML, LaTeX, SMCL, or Markdown. Output does not change when these options are specified. Use collect preview to see the customizeable table.

collect is a shortcut for collect(Tabulate).

cname specifies that a collection named *cname* be associated with the collected results. The default is Tabulate.

append specifies that results be appended to collection *cname*.

replace permits tabulate to overwrite an existing collection. This option is implied for collection Tabulate when append is not specified.

label(*filename*) specifies the *filename* containing the collection labels to use for your table. Labels in *filename* will be loaded into the collection, and any labels not specified in *filename* will be taken from the labels defined in c(collect_label). The default is to use only the collection labels set in c(collect_label); see [TABLES] **set collect_label**.

style(*filename*[, override]) specifies the *filename* containing the collection styles to use for your table. The default collection styles will be discarded, and only the collection styles in *filename* will be applied.

If you prefer the default collection styles but also want to apply any styles in *filename*, specify override. If there are conflicts between the default collection styles and those in *filename*, the ones in *filename* will take precedence.

The default is to use only the collection styles set in c(tabulate_style); see [TABLES] **set tabulate_style**.

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

▷ 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 https://www.stata-press.com/data/r19/auto
(1978 automobile data)
. tabulate foreign
```

Car origin	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 origin	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 equal-variances test: 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 origin		
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 origin
    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 origin
    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
```

◁

Publish your tables

With the `collect` option, `tabulate` posts the tabulated values to a collection named `Tabulate` and sets it as the current collection. With collections, you can customize the look of your table, then publish it to HTML, Word, LaTeX, PDF, Excel, or another format appropriate for your report.

With the by prefix, `tabulate` appends the resulting tabulations into a single collection, and the default layout produces a separate table for each by group.

If you are not familiar with collections, see [TABLES] **Intro**. The predefined styles for `tabulate` are documented in [TABLES] **Predefined styles**.

▷ Example 4

Recall our table from the previous example. Let's add the `collect` option in the call to `tabulate` to produce a collection with its tabulated values.

```
. tabulate wgtcat foreign, summarize(mpg) collect
        Means, Standard Deviations and Frequencies of Mileage (mpg)
```

wgtcat	Car origin Domestic	Foreign	Total
2530	28.285714 3.0937725 7	27.0625 5.9829619 16	27.434783 5.2295149 23
3300	21.75 2.4083189 16	19.6 3.4351128 5	21.238095 2.7550819 21
4070	17.26087 1.8639497 23	14 0 1	17.125 1.9406969 24
4840	14.666667 3.32666 6	. . 0	14.666667 3.32666 6
Total	19.826923 4.7432972 52	24.772727 6.6111869 22	21.297297 5.7855032 74

The output does not change; however, we can use the `collect dir` command to see that `tabulate` created a collection named `Tabulate`.

```
. collect dir
Collections in memory
Current: Tabulate
──────────────────────
    Name   No. items
──────────────────────
   Table   51
Tabulate   45
──────────────────────
```

In this collection, means are tagged with `result[mean]`, standard deviations with `result[sd]`, and frequencies with `result[frequency]`. When weights are specified, observation counts are tagged with `result[count]`. Here we use `collect label list` to show the levels and labels of the `result` dimension.

```
. collect label list result
  Collection: Tabulate
   Dimension: result
       Label: Result
Level labels:
   frequency  Frequency
        mean  Mean
          sd  Std. dev.
```

The tabulated variables (that is, `wgtcat` and `foreign`) are added to the collection as dimensions and are used to tag the collected results. In addition to the name, label, level values, and value labels of the tabulated variables, these dimensions each have the `__margCode__` level with the `Total` label for tagging the marginal summary statistics. Here we use `collect label list` to show the levels and labels of each tabulated variable dimension. For `wgtcat`, we add the `all` option to show the levels without labels.

```
. collect label list wgtcat, all
  Collection: Tabulate
   Dimension: wgtcat
       Label:
Level labels:
        2530
        3300
        4070
        4840
__margCode__  Total
. collect label list foreign
  Collection: Tabulate
   Dimension: foreign
       Label: Car origin
Level labels:
           0  Domestic
           1  Foreign
__margCode__  Total
```

`tabulate` constructs a default layout, so you can view your customizable table with the `collect preview` command. Here we use the `collect layout` command to report the default layout specification and corresponding table.

```
. collect layout

Collection: Tabulate
      Rows: wgtcat#result
   Columns: var#foreign
    Tables: cmdset
   Table 1: 21 x 3
```

	Mileage (mpg)		
	Car origin		
	Domestic	Foreign	Total
wgtcat			
2530			
Mean	28.3	27.1	27.4
Std. dev.	3.1	6.0	5.2
Frequency	7	16	23
3300			
Mean	21.8	19.6	21.2
Std. dev.	2.4	3.4	2.8
Frequency	16	5	21
4070			
Mean	17.3	14.0	17.1
Std. dev.	1.9	0.0	1.9
Frequency	23	1	24
4840			
Mean	14.7	.	14.7
Std. dev.	3.3	.	3.3
Frequency	6	0	6
Total			
Mean	19.8	24.8	21.3
Std. dev.	4.7	6.6	5.8
Frequency	52	22	74

We now change some cell formats, remove the summary statistic labels from the row header, and add a note suggesting the table's cell contents.

```
. collect style cell result[mean], nformat(%9.2f)
. collect style cell result[sd], nformat(%9.2f) sformat("(%s)")
. collect style header result, level(hide)
. collect note "Mean (SD) N"
. collect preview
```

	Mileage (mpg)		
	Car origin		
	Domestic	Foreign	Total
wgtcat			
2530	28.29	27.06	27.43
	(3.09)	(5.98)	(5.23)
	7	16	23
3300	21.75	19.60	21.24
	(2.41)	(3.44)	(2.76)
	16	5	21
4070	17.26	14.00	17.12
	(1.86)	(0.00)	(1.94)
	23	1	24
4840	14.67	.	14.67
	(3.33)	(.)	(3.33)
	6	0	6
Total	19.83	24.77	21.30
	(4.74)	(6.61)	(5.79)
	52	22	74

Mean (SD) N

We can make further changes to the table with the `collect` suite of commands. But we are happy with this layout and ready to publish the table to a LaTeX file with `collect export`. We simply specify the filename to which we want to export it.

```
. collect export tab2.tex
(collection Tabulate exported to file tab2.tex)
```

With `collect export`, you can publish the table to several formats, such as HTML, PDF, and LaTeX files, by specifying the appropriate file extension.

◁

Also see

[R] **table** — Table of frequencies, summaries, and command results

[R] **table summary** — 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

[TABLES] **Intro** — Introduction

[U] 12.6 Dataset, variable, and value labels

[U] 26 Working with categorical data and factor variables