cmtab — Tabulate chosen alternatives

Description

`cmtab` tabulates chosen alternatives, either alone in a one-way tabulation or versus another variable in a two-way tabulation.

For panel choice data, `cmtab` can display a two-way tabulation of chosen alternatives by time or a three-way tabulation of time by chosen alternative by another variable.

Quick start

Display a one-way tabulation of chosen alternatives for `cmset` data, where `depvar` is a 0/1 variable

```
cmtab, choice(depvar)
```

Tabulate chosen alternatives versus the values of variable `xvar`

```
cmtab xvar, choice(depvar)
```

Same as above, and report row percentages and Pearson's χ^2 test

```
cmtab xvar, choice(depvar) row chi2
```

Transpose rows and columns in the above tabulation

```
cmtab xvar, choice(depvar) col chi2 transpose
```

For panel choice data, display a two-way tabulation of chosen alternatives versus the time variable

```
cmtab, choice(depvar) time
```

For panel choice data, display tabulations of chosen alternatives versus `x` for each time

```
cmtab x, choice(depvar) time
```

Same as above, but display tabulations of chosen alternatives versus times for each value of `x`

```
cmtab x, choice(depvar) time timelast
```

Same as above, but create a compact display

```
cmtab x, choice(depvar) time timelast compact
```

Menu

Statistics > Choice models > Setup and utilities > Tabulate chosen alternatives

Syntax

```
cmtab [varname] [if] [in] [weight], choice(choicevar) [options]
```

options	Description
Main	
* `choice(`*choicevar*`)`	specify 0/1 variable indicating the chosen alternative
`missing`	include missing values of *varname* in tabulation
`transpose`	transpose rows and columns in tables
`time`	tabulate by time variable (only for panel CM data)
`timelast`	put time variable last in three-way tabulation; tabulate alternatives by time for each level of *varname* (only for panel CM data)
`compact`	display three-way tabulation compactly (only for panel CM data)
`altwise`	use alternativewise deletion instead of casewise deletion
Options	
tab1_options	options for one-way tables
tab2_options	options for two-way tables

*`choice()` is required.

tab1_options	Description
`sort`	display table in descending order of frequency

tab2_options	Description
`chi2`	report Pearson's χ^2
`lrchi2`	report likelihood-ratio χ^2
`column`	report column percentages
`row`	report row percentages
`cell`	report cell percentages
`rowsort`	list rows in order of observed frequency
`colsort`	list columns in order of observed frequency
`[no]key`	report or suppress cell contents key

You must `cmset` your data before using `cmtab`; see [CM] **cmset**.
`by` and `collect` are allowed; see **[U] 11.1.10 Prefix commands**.
`fweights` and `iweights` are allowed; see **[U] 11.1.6 weight**.

Options

Main

`choice(`*choicevar*`)` specifies the variable indicating the chosen alternative. *choicevar* must be coded as 0 and 1, with 0 indicating an alternative that was not chosen and 1 indicating the chosen alternative. `choice()` is required.

`missing` specifies that the missing values of *varname* are to be treated like any other value of *varname*.

`transpose` transposes rows and columns in the tabular displays.

`time` tabulates the chosen alternative versus the time variable when data are panel choice data. See [CM] **cmset**.

`timelast` puts time last in a three-way tabulation when data are panel choice data. Three-way tabulations are created when *varname* is specified as well as the option `time`. By default, the three-way tabulation is *timevar* × chosen alternative × *varname*; that is, for each value of *timevar*, a two-way table of chosen alternative versus *varname* is displayed. When `timelast` is specified, the three-way tabulation is *varname* × chosen alternative × *timevar*; that is, for each value of *varname*, a two-way table of chosen alternative versus *timevar* is displayed. To reverse the order of the two-way tabulations, you can use the option `transpose`.

`compact` creates a compact three-way tabulation when data are panel choice data.

`altwise` specifies that alternativewise deletion be used when omitting observations because of missing values in your variables. The default is to use casewise deletion; that is, the entire group of observations making up a case is omitted if any missing values are encountered. This option does not apply to observations that are excluded by the `if` or `in` qualifier or the `by` prefix; these observations are always handled alternativewise regardless of whether `altwise` is specified.

Options

`sort` puts the table in descending order of frequency in a one-way table.

`chi2` calculates and displays Pearson's χ^2 for the hypothesis that the rows and columns in a two-way table are independent. `chi2` may not be specified if `iweights` are used. `chi2` is not available when `compact` is specified.

`lrchi2` displays the likelihood-ratio χ^2 statistic for a two-way table. `lrchi2` may not be specified if `iweights` are used. `lrchi2` is not available when `compact` is specified.

`column` displays the relative frequency, as a percentage, of each cell within its column in a two-way table. `column` is not available when `compact` is specified.

`row` displays the relative frequency, as a percentage, of each cell within its row in a two-way table. `row` is not available when `compact` is specified.

`cell` displays the relative frequency, as a percentage, of each cell in a two-way table. `cell` is not available when `compact` is specified.

`rowsort` and `colsort` specify that the rows and columns, respectively, be presented in order of observed frequency in a two-way table. `rowsort` and `colsort` are not available when `compact` is specified.

[`no`]`key` displays or suppresses a key above two-way tables. The default is to display the key if more than one cell statistic is requested. `key` displays the key. `nokey` suppresses its display. [`no`]`key` is not available when `compact` is specified.

Remarks and examples

`cmtab` is a convenience command for tabulating chosen alternatives, either alone or against another variable.

The option `choice(`*choicevar*`)` is required, where *choicevar* is a 0/1 variable. *choicevar* is typically the dependent variable for choice models with 0/1 dependent variables.

For rank-ordered choice models, such as `cmroprobit`, using a dependent variable of ranks with `choice()` will give an error message. To use `cmtab` in this instance, you would have to create a 0/1 variable, such as a variable indicating the highest ranked alternative for each case.

For tabulations of choice sets, see [CM] **cmchoiceset**. For an overview of other descriptive statistics available for choice model data, see [CM] **Intro 3**.

▷ Example 1: Cross-sectional choice data

Here is an example with cross-sectional choice data. First, we `cmset` our data:

```
. use https://www.stata-press.com/data/r19/carchoice
(Car choice data)

. cmset consumerid car
note: alternatives are unbalanced across choice sets; choice sets of
      different sizes found.

     Case ID variable: consumerid
Alternatives variable: car
```

These fictitious data represent persons who purchased a car with their choices categorized by the nationality of the manufacturer, American, Japanese, European, or Korean. Second, we use `cmtab` with only the `choice()` option, which gives a one-way tabulation of the chosen alternatives.

```
. cmtab, choice(purchase)
Tabulation of chosen alternatives (purchase = 1)
```

Nationality of car	Freq.	Percent	Cum.
American	384	43.39	43.39
Japanese	326	36.84	80.23
European	135	15.25	95.48
Korean	40	4.52	100.00
Total	885	100.00	

We see that most people in this dataset purchased American cars more than any other nationality of car.

We can look at associations between chosen alternatives and other variables in the dataset. We wonder whether gender is associated with the nationality of the car purchased:

```
. cmtab gender, choice(purchase)
Tabulation for chosen alternatives (purchase = 1)
    gender is constant within case
```

Nationality of car	Gender: 0 = Female, 1 = Male: Female	Male	Total
American	96	280	376
Japanese	110	206	316
European	22	108	130
Korean	8	32	40
Total	236	626	862

We specify the option `row` to better see the percentages of gender within choices. We also specify `chi2` to get a p-value for the association of gender with the choice of car.

```
. cmtab gender, choice(purchase) row chi2
Tabulation for chosen alternatives (purchase = 1)
    gender is constant within case
```

Key
frequency *row percentage*

Nationality of car	Gender: 0 = Female, 1 = Male Female	Male	Total
American	96 25.53	280 74.47	376 100.00
Japanese	110 34.81	206 65.19	316 100.00
European	22 16.92	108 83.08	130 100.00
Korean	8 20.00	32 80.00	40 100.00
Total	236 27.38	626 72.62	862 100.00

Pearson chi2(3) = 17.6654 Pr = 0.001

There are more male car purchasers than female car purchasers in these data. Purchasers of European cars are even more overwhelmingly male. However, the percentage of Japanese cars purchased by females is greater than the percentage of American, European, or Korean cars purchased by females. The p-value from the Pearson's χ^2 test for association is 0.001.

The `transpose` option transposes rows and columns in the display:

```
. cmtab gender, choice(purchase) row chi2 nokey transpose
Tabulation for chosen alternatives (purchase = 1)
    gender is constant within case
```

Gender: 0 = Female, 1 = Male	Nationality of car American	Japanese	European	Korean	Total
Female	96 40.68	110 46.61	22 9.32	8 3.39	236 100.00
Male	280 44.73	206 32.91	108 17.25	32 5.11	626 100.00
Total	376 43.62	316 36.66	130 15.08	40 4.64	862 100.00

Pearson chi2(3) = 17.6654 Pr = 0.001

◁

▷ Example 2: Panel choice data

When you have panel choice data, cmtab is useful to see how chosen alternatives vary by time. Here is an example. First, we cmset the data:

```
. use https://www.stata-press.com/data/r19/transport, clear
(Transportation choice data)
. cmset id t alt
note: case identifier _caseid generated from id and t.
note: panel by alternatives identifier _panelaltid generated from id and alt.
                    Panel data: Panels id and time t
              Case ID variable: _caseid
         Alternatives variable: alt
Panel by alternatives variable: _panelaltid (strongly balanced)
                 Time variable: t, 1 to 3
                         Delta: 1 unit
Note: Data have been xtset.
```

Second, we specify the option time to look at chosen alternatives by time. The option column helps to see whether there is any trend with time.

```
. cmtab, choice(choice) time column chi2
Tabulation of chosen alternatives (choice = 1) by time t
```

Key
frequency *column percentage*

Alternatives	Time variable 1	2	3	Total
Car	234 46.80	359 71.80	388 77.60	981 65.40
Public	108 21.60	81 16.20	67 13.40	256 17.07
Bicycle	74 14.80	40 8.00	31 6.20	145 9.67
Walk	84 16.80	20 4.00	14 2.80	118 7.87
Total	500 100.00	500 100.00	500 100.00	1,500 100.00

Pearson chi2(6) = 148.9651 Pr = 0.000

There is a large time trend for the chosen alternatives in these data. The percentage of persons choosing cars as their mode of transportation increases from 46.8% at time 1 to 77.6% at time 3. All the other choices of modes of transportation decline over time.

Does choice of transportation vary by whether a person has a full-time or part-time job (indicated by the variable parttime)? Here is how we could look at that, aggregating across time.

```
. cmtab parttime, choice(choice) column nokey
Tabulation for chosen alternatives (choice = 1)
    parttime is constant within case
```

Alternativ	Part-time job		
es	Full-time	Part-time	Total
Car	503	478	981
	66.80	63.99	65.40
Public	132	124	256
	17.53	16.60	17.07
Bicycle	72	73	145
	9.56	9.77	9.67
Walk	46	72	118
	6.11	9.64	7.87
Total	753	747	1,500
	100.00	100.00	100.00

Because this tabulation aggregates chosen alternatives across time for the same individual, we did not calculate a Pearson χ^2. However, there does not appear to be an association between choice of transportation and whether the person is employed full time or part time.

Let's look at the choice of transportation by full-time or part-time employment for each time point. To do this, we add the option `time`. We also specify the option `transpose` to make wide tables that take up less vertical space. Because we are not aggregating counts, we also specify the `chi2` option.

```
. cmtab parttime, choice(choice) row chi2 nokey time transpose
Tabulations by chosen alternatives (choice = 1)
    parttime is constant within case
    time t = 1
```

Part-time job	Alternatives				
	Car	Public	Bicycle	Walk	Total
Full-time	119	53	34	31	237
	50.21	22.36	14.35	13.08	100.00
Part-time	115	55	40	53	263
	43.73	20.91	15.21	20.15	100.00
Total	234	108	74	84	500
	46.80	21.60	14.80	16.80	100.00

Pearson chi2(3) = 5.0154 Pr = 0.171

```
    time t = 2
```

Part-time job	Alternatives				
	Car	Public	Bicycle	Walk	Total
Full-time	186	43	18	11	258
	72.09	16.67	6.98	4.26	100.00
Part-time	173	38	22	9	242
	71.49	15.70	9.09	3.72	100.00
Total	359	81	40	20	500
	71.80	16.20	8.00	4.00	100.00

Pearson chi2(3) = 0.8683 Pr = 0.833

```
    time t = 3
```

Part-time job	Alternatives				
	Car	Public	Bicycle	Walk	Total
Full-time	198	36	20	4	258
	76.74	13.95	7.75	1.55	100.00
Part-time	190	31	11	10	242
	78.51	12.81	4.55	4.13	100.00
Total	388	67	31	14	500
	77.60	13.40	6.20	2.80	100.00

Pearson chi2(3) = 5.2158 Pr = 0.157

Is there a time trend for choice of transportation for those employed full time? For those employed part time? The tables above can be considered a three-way tabulation: time × parttime × chosen alternative. To look for time trends within parttime, we note the three-way tabulation parttime × chosen alternative × time is better. We can get this three-way tabulation by specifying the option timelast.

```
. cmtab parttime, choice(choice) column chi2 nokey time timelast
Tabulations by chosen alternatives (choice = 1)
    parttime is constant within case
    parttime = 0
```

Alternatives	Time variable 1	2	3	Total
Car	119	186	198	503
	50.21	72.09	76.74	66.80
Public	53	43	36	132
	22.36	16.67	13.95	17.53
Bicycle	34	18	20	72
	14.35	6.98	7.75	9.56
Walk	31	11	4	46
	13.08	4.26	1.55	6.11
Total	237	258	258	753
	100.00	100.00	100.00	100.00

Pearson chi2(6) = 57.2439 Pr = 0.000

parttime = 1

Alternatives	Time variable 1	2	3	Total
Car	115	173	190	478
	43.73	71.49	78.51	63.99
Public	55	38	31	124
	20.91	15.70	12.81	16.60
Bicycle	40	22	11	73
	15.21	9.09	4.55	9.77
Walk	53	9	10	72
	20.15	3.72	4.13	9.64
Total	263	242	242	747
	100.00	100.00	100.00	100.00

Pearson chi2(6) = 93.5435 Pr = 0.000

Three-way tabulations created by cmtab can be displayed more compactly using the option compact:

```
. cmtab parttime, choice(choice) time timelast compact
Tabulations by chosen alternatives (choice = 1)
    parttime is constant within case
```

Alternatives	Part-time job and Time variable: Full-time 1	Full-time 2	Full-time 3	Part-time 1	Part-time 2	Part-time 3
Car	119	186	198	115	173	190
Public	53	43	36	55	38	31
Bicycle	34	18	20	40	22	11
Walk	31	11	4	53	9	10

Stored results

cmtab stores the following in r():

Scalars

r(N)	number of observations
r(r)	number of rows
r(c)	number of columns
r(chi2)	Pearson's χ^2
r(p)	p-value for Pearson's χ^2 test
r(chi2_lr)	likelihood-ratio χ^2
r(p_lr)	p-value for likelihood-ratio test

Also see

[CM] **cmchoiceset** — Tabulate choice sets

[CM] **cmsample** — Display reasons for sample exclusion

[CM] **cmset** — Declare data to be choice model data

[CM] **cmsummarize** — Summarize variables by chosen alternatives