**xttab — Tabulate xt data**

### Description

*xttab*, a generalization of *tabulate* (see [R] *tabulate oneway*), performs one-way tabulations and decomposes counts into between and within components in panel data.

*xtrans*, another generalization of *tabulate* (see [R] *tabulate oneway*), reports transition probabilities (the change in one categorical variable over time).

### Quick start

Overall, between, and within one-way tabulation of *v1* using *xtset* data

```
xtabl v1
```

Report transition probabilities for *v2*

```
xtrans v2
```

Add frequency of transitions

```
xtrans v2, freq
```

As above, but for each level of *catvar*

```
bysort catvar: xtrans v2, freq
```

### Menu

**xttab**

Statistics > Longitudinal/panel data > Setup and utilities > Tabulate xt data

**xtrans**

Statistics > Longitudinal/panel data > Setup and utilities > Report transition probabilities
Syntax

```
xttab varname [if]
```

```
xtrans varname [if] [, freq]
```

A panel variable must be specified; use `xtset`; see [XT] `xtset`. by is allowed with `xttab` and `xtrans`; see [D] `by`.

Option

```
freq, allowed with `xtrans` only, specifies that frequencies as well as transition probabilities be displayed.
```

Remarks and examples

If you have not read [XT] `xt`, please do so.

Example 1: `xtab`

Using the `nlswork` dataset described in [XT] `xt`, variable `msp` is 1 if a woman is married and her spouse resides with her, and 0 otherwise:

```
. use https://www.stata-press.com/data/r16/nlswork
   (National Longitudinal Survey. Young Women 14–26 years of age in 1968)
. xtab msp
```

<table>
<thead>
<tr>
<th>msp</th>
<th>Overall</th>
<th>Between</th>
<th>Within</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freq. Percent</td>
<td>Freq. Percent</td>
<td>Percent</td>
</tr>
<tr>
<td>0</td>
<td>11324</td>
<td>39.71</td>
<td>3113 66.08</td>
</tr>
<tr>
<td>1</td>
<td>17194</td>
<td>60.29</td>
<td>3643 77.33</td>
</tr>
<tr>
<td>Total</td>
<td>28518</td>
<td>100.00</td>
<td>6756 143.41</td>
</tr>
</tbody>
</table>

The overall part of the table summarizes results in terms of person-years. We have 11,324 person-years of data in which `msp` is 0 and 17,194 in which it is 1—in 60.3% of our data, the woman is married with her spouse present. Between repeats the breakdown, but this time in terms of women rather than person-years; 3,113 of our women ever had `msp` 0 and 3,643 ever had `msp` 1, for a grand total of 6,756 ever having either. We have in our data, however, only 4,711 women. This means that there are women who sometimes have `msp` 0 and at other times have `msp` 1.

The within percent tells us the fraction of the time a woman has the specified value of `msp`. If we take the first line, conditional on a woman ever having `msp` 0, 62.7% of her observations have `msp` 0. Similarly, conditional on a woman ever having `msp` 1, 75.8% of her observations have `msp` 1. These two numbers are a measure of the stability of the `msp` values, and, in fact, `msp` 1 is more stable among these younger women than `msp` 0, meaning that they tend to marry more than they divorce. The total within of 69.73% is the normalized between weighted average of the within percents, that is, $(3113 \times 62.69 + 3643 \times 75.75) / 6756$. It is a measure of the overall stability of the `msp` variable.
A time-invariant variable will have a tabulation with within percents of 100:

\[ \texttt{. xttab race} \]

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th></th>
<th>Between</th>
<th></th>
<th>Within</th>
</tr>
</thead>
<tbody>
<tr>
<td>race</td>
<td>Freq.</td>
<td>Percent</td>
<td>Freq.</td>
<td>Percent</td>
<td>Percent</td>
</tr>
<tr>
<td>white</td>
<td>20180</td>
<td>70.72</td>
<td>3329</td>
<td>70.66</td>
<td>100.00</td>
</tr>
<tr>
<td>black</td>
<td>8051</td>
<td>28.22</td>
<td>1325</td>
<td>28.13</td>
<td>100.00</td>
</tr>
<tr>
<td>other</td>
<td>303</td>
<td>1.06</td>
<td>57</td>
<td>1.21</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>28534</td>
<td>100.00</td>
<td>4711</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

\[ (n = 4711) \]

Example 2: \texttt{xttrans}

\texttt{xttrans} shows the transition probabilities. In cross-sectional time-series data, we can estimate the probability that \( x_{i,t+1} = v_2 \) given that \( x_{it} = v_1 \) by counting transitions. For instance

\[ \texttt{. xttrans msp} \]

<table>
<thead>
<tr>
<th>married, spouse present</th>
<th>1 if married, spouse present</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>80.49</td>
<td>19.51</td>
</tr>
<tr>
<td>1</td>
<td>7.96</td>
<td>92.04</td>
</tr>
<tr>
<td>Total</td>
<td>37.11</td>
<td>62.89</td>
</tr>
</tbody>
</table>

The rows reflect the initial values, and the columns reflect the final values. Each year, some 80% of the \texttt{msp} 0 persons in the data remained \texttt{msp} 0 in the next year; the remaining 20% became \texttt{msp} 1. Although \texttt{msp} 0 had a 20% chance of becoming \texttt{msp} 1 in each year, the \texttt{msp} 1 had only an 8% chance of becoming (or returning to) \texttt{msp} 0. The \texttt{freq} option displays the frequencies that go into the calculation:

\[ \texttt{. xttrans msp, freq} \]

<table>
<thead>
<tr>
<th>married, spouse present</th>
<th>1 if married, spouse present</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7,697</td>
<td>1,866</td>
</tr>
<tr>
<td></td>
<td>80.49</td>
<td>19.51</td>
</tr>
<tr>
<td>1</td>
<td>1,133</td>
<td>13,100</td>
</tr>
<tr>
<td></td>
<td>7.96</td>
<td>92.04</td>
</tr>
<tr>
<td>Total</td>
<td>8,830</td>
<td>14,966</td>
</tr>
<tr>
<td></td>
<td>37.11</td>
<td>62.89</td>
</tr>
</tbody>
</table>
Technical note

The transition probabilities reported by `xttrans` are not necessarily the transition probabilities in a Markov sense. `xttrans` counts transitions from each observation to the next once the observations have been put in $t$ order within $i$. It does not normalize for missing periods. `xttrans` does pay attention to missing values of the variable being tabulated, however, and does not count transitions from nonmissing to missing or from missing to nonmissing. Thus if the data are fully rectangularized, `xttrans` produces (inefficient) estimates of the Markov transition matrix. `fillin` will rectangularize datasets; see [D] `fillin`. Thus the Markov transition matrix could be estimated by typing

```
.fillin idcode year
.xttrans msp
(output omitted)
```

Stored results

`xttab` stores the following in `r()`:

Scalars
- `r(n)` number of panels

Matrices
- `r(results)` results matrix

Also see

[XT] `xtdescribe` — Describe pattern of xt data
[XT] `xtsum` — Summarize xt data