

Description

`xtdescribe` describes the participation pattern of cross-sectional time-series (xt) data.

Quick start

Describe the 9 most common participation patterns of `xtset` data

```
xtdescribe
```

Describe up to 15 of the most common participation patterns

```
xtdescribe, patterns(15)
```

Same as above, but list all participation patterns

```
xtdescribe, patterns(1000)
```

Describe patterns only for study subjects, denoted by binary variable `insample = 1`

```
xtdescribe if insample
```

Menu

Statistics > Longitudinal/panel data > Setup and utilities > Describe pattern of xt data

Syntax

```
xtdescribe [if] [in] [, options]
```

options

Description

Main

`patterns(#)`

maximum participation patterns; default is `patterns(9)`

`width(#)`

display # width of participation patterns; default is `width(100)`

A panel variable and a time variable must be specified; use `xtset`; see [\[XT\] xtset](#).
`by` is allowed; see [\[D\] by](#).

Options

Main

`patterns(#)` specifies the maximum number of participation patterns to be reported; `patterns(9)` is the default. Specifying `patterns(50)` would list up to 50 patterns. Specifying `patterns(1000)` is taken to mean `patterns(∞)`: all the patterns will be listed.

width(#) specifies the desired width of the participation patterns to be displayed; width(100) is the default. If the number of times is greater than width(), then each column in the participation pattern represents multiple periods as indicated in a footnote at the bottom of the table. The actual width may differ slightly from the requested width depending on the span of the time variable and the number of periods.

Remarks and examples

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

xtdescribe describes the cross-sectional and time-series aspects of the data in memory.

► Example 1

In [XT] xt, we introduced data based on a subsample of the NLSY data on young women aged 14–24 years in 1968. Here is a description of the data used in many of the [XT] xt examples:

```
. use https://www.stata-press.com/data/r19/nlswork
(National Longitudinal Survey of Young Women, 14-24 years old in 1968)
. xtdescribe
idcode: 1, 2, ..., 5159          n =          4711
year:   68, 69, ..., 88          T =           15
Delta(year) = 1 unit
Span(year) = 21 periods
(idcode*year uniquely identifies each observation)
Distribution of T_i:  min    5%    25%    50%    75%    95%    max
                   1      1      3      5      9     13     15
Freq.  Percent  Cum. | Pattern
-----|-----
  136    2.89   2.89 | 1.....
  114    2.42   5.31 | .....1
   89    1.89   7.20 | .....1.11
   87    1.85   9.04 | .....11
   86    1.83  10.87 | 111111.1.11.1.11.1.11
   61    1.29  12.16 | .....11.1.11
   56    1.19  13.35 | 11.....
   54    1.15  14.50 | .....1.1.11
   54    1.15  15.64 | .....1.11.1.11.1.11
 3974   84.36 100.00 | (other patterns)
-----|-----
 4711  100.00          | XXXXXX.X.XX.X.XX.X.XX
```

xtdescribe tells us that we have 4,711 women in our data and that the idcode that identifies each ranges from 1 to 5,159. We are also told that the maximum number of individual years over which we observe any woman is 15, though the year variable spans 21 years. The delta or periodicity of year is one unit, meaning that in principle we could observe each woman yearly. We are reassured that idcode and year, taken together, uniquely identify each observation in our data. We are also shown the distribution of T_i ; 50% of our women are observed 5 years or less. Only 5% of our women are observed for 13 years or more.

Finally, we are shown the participation pattern. A 1 in the pattern means one observation that year; a dot means no observation. The largest fraction of our women (still only 2.89%) was observed in the single year 1968 and not thereafter; the next largest fraction was observed in 1988 but not before; and the next largest fraction was observed in 1985, 1987, and 1988.

At the bottom is the sum of the participation patterns, including the patterns that were not shown. We can see that none of the women were observed in six of the years (there are six dots). (The survey was not administered in those six years.)

We could see more of the patterns by specifying the `patterns()` option, or we could see all the patterns by specifying `patterns(1000)`.



▷ Example 2

The strange participation patterns shown above have to do with our subsampling of the data, not with the administrators of the survey. Here are the data from which we drew the sample used in [XT] `xt`:

```
. xtdescribe
idcode: 1, 2, ..., 5159          n =          5159
year:   68, 69, ..., 88          T =           15
Delta(year) = 1; (88-68)+1 = 21
(idcode*year does not uniquely identify observations)
Distribution of T_i:  min      5%      25%      50%      75%      95%      max
                   1         2         11         15         16         19         30

   Freq.  Percent   Cum. | Pattern
-----|-----
  1034    20.04   20.04 | 111111.1.11.1.11.1.11
   153     2.97   23.01 | 1.....
   147     2.85   25.86 | 112111.1.11.1.11.1.11
   130     2.52   28.38 | 111112.1.11.1.11.1.11
   122     2.36   30.74 | 111211.1.11.1.11.1.11
   113     2.19   32.93 | 11.....
    84     1.63   34.56 | 111111.1.11.1.11.1.12
    79     1.53   36.09 | 111111.1.12.1.11.1.11
    67     1.30   37.39 | 111111.1.11.1.11.1.1.
  3230    62.61  100.00 | (other patterns)
-----|-----
  5159   100.00          | XXXXXX.X.XX.X.XX.X.XX
```

We have multiple observations per year. In the pattern, 2 indicates that a woman appears twice in the year, 3 indicates 3 times, and so on—X indicates 10 or more, should that be necessary.

In fact, this is a dataset that was itself extracted from the NLSY, in which *t* is not time but job number. To simplify exposition, we made a simpler dataset by selecting the last job in each year.



▷ Example 3

When the number of periods is greater than the width of the participation pattern, each column will represent more than one period.

```
. use https://www.stata-press.com/data/r19/xtdesxmpl
. xtdescribe, width(16)
patient: 1, 2, ..., 30          n =          30
time:    09mar2007 16:00:00, 09mar2007 17:00:00, ...,      T =          32
        10mar2007 23:00:00
        Delta(time) = 1 hour
        Span(time)  = 32 periods
        (patient*time uniquely identifies each observation)

Distribution of T_i:  min      5%      25%      50%      75%      95%      max
                   30      30      31      32      32      32      32

   Freq.  Percent   Cum. | Pattern*
-----+-----
    21     70.00   70.00 | 2222222222222222
     4     13.33   83.33 | 1222222222222222
     3     10.00   93.33 | 22222222222222.
     2      6.67  100.00 | .222222222222222

    30     100.00          | XXXXXXXXXXXXXXXXX
```

*Each column represents 2 periods.

We have data for 30 patients who were observed hourly between 4:00 PM on March 9, 2007, and 11:00 PM on March 10, a span of 32 hours. We have complete records for 21 of the patients. The footnote indicates that each column in the pattern represents two periods, so for four patients we have an observation taken at either 4:00 PM or 5:00 PM on March 9, but we do not have observations for both times. There are three patients for whom we are missing both the 10:00 PM and 11:00 PM observations on March 10, and there are two patients for whom we are missing the 4:00 PM and 5:00 PM observations for March 9.



Reference

Cox, N. J. 2007. [Speaking Stata: Counting groups, especially panels](#). *Stata Journal* 7: 571–581.

Also see

- [XT] [xtsum](#) — Summarize xt data
- [XT] [xttab](#) — Tabulate xt data

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow and NetCourseNow are trademarks of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.



For suggested citations, see the FAQ on [citing Stata documentation](#).