

[Description](#)[Remarks and examples](#)[Quick start](#)[Stored results](#)[Menu](#)[Also see](#)[Syntax](#)

Description

`xtsum`, a generalization of `summarize` (see [\[R\] summarize](#)), reports means and standard deviations for panel data; it differs from `summarize` in that it decomposes the standard deviation into between and within components.

Quick start

Report means and overall, between, and within standard deviations for all numeric variables in `xtset` data

```
xtsum
```

Same as above, but restrict to `v1`, `v2`, and `v3`

```
xtsum v1 v2 v3
```

Same as above, but calculate statistics separately for each level of `catvar`

```
bysort catvar: xtsum v1 v2 v3
```

Menu

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

Syntax

`xtsum [varlist] [if]`

A panel variable must be specified; use `xtset`; see [XT] `xtset`.

`varlist` may contain time-series operators; see [U] 11.4.4 Time-series varlists.

`by` and `collect` are allowed; see [U] 11.1.10 Prefix commands.

Remarks and examples

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

`xtsum` provides an alternative to `summarize`. For instance, in the `nlswork` dataset described in [XT] `xt`, `hours` contains the usual hours worked:

```
. use https://www.stata-press.com/data/r19/nlswork
(National Longitudinal Survey of Young Women, 14-24 years old in 1968)
. summarize hours
      Variable |       Obs        Mean    Std. dev.      Min       Max
  hours | 28,467  36.55956  9.869623         1       168
. xtsum hours
      Variable |       Mean    Std. dev.      Min       Max |   Observations
  hours   overall | 36.55956  9.869623         1       168 |   N = 28467
           between |          7.846585  1          83.5 |   n = 4710
           within  |          7.520712 -2.154726  130.0596 | T-bar = 6.04395
```

`xtsum` provides the same information as `summarize` and more. It decomposes the variable x_{it} into a between (\bar{x}_i) and within ($x_{it} - \bar{x}_i + \bar{x}$, the global mean \bar{x} being added back in make results comparable). The overall and within are calculated over 28,467 person-years of data. The between is calculated over 4,710 persons, and the average number of years a person was observed in the `hours` data is 6.

`xtsum` also reports minimums and maximums. Hours worked last week varied between 1 and (unbelievably) 168. Average hours worked last week for each woman varied between 1 and 83.5. “Hours worked within” varied between -2.15 and 130.1, which is not to say that any woman actually worked negative hours. The within number refers to the deviation from each individual’s average, and naturally, some of those deviations must be negative. Then the negative value is not disturbing but the positive value is. Did some woman really deviate from her average by +130.1 hours? No. In our definition of within, we add back in the global average of 36.6 hours. Some woman did deviate from her average by $130.1 - 36.6 = 93.5$ hours, which is still large.

The reported standard deviations tell us something that may surprise you. They say that the variation in hours worked last week across women is nearly equal to that observed within a woman over time. That is, if you were to draw two women randomly from our data, the difference in hours worked is expected to be nearly equal to the difference for the same woman in two randomly selected years.

If a variable does not vary over time, its within standard deviation will be zero:

```
. xtsum birth_yr
      Variable |       Mean    Std. dev.      Min       Max |   Observations
  birth_yr  overall | 48.08509  3.012837        41       54 |   N = 28534
           between |          3.051795  41          54 |   n = 4711
           within  |          0  48.08509  48.08509 | T-bar = 6.05689
```

Stored results

xtsum stores the following in `r()`:

Scalars

<code>r(N)</code>	number of observations
<code>r(n)</code>	number of panels
<code>r(Tbar)</code>	average number of years under observation
<code>r(mean)</code>	mean
<code>r(sd)</code>	overall standard deviation
<code>r(min)</code>	overall minimum
<code>r(max)</code>	overall maximum
<code>r(sd_b)</code>	between standard deviation
<code>r(min_b)</code>	between minimum
<code>r(max_b)</code>	between maximum
<code>r(sd_w)</code>	within standard deviation
<code>r(min_w)</code>	within minimum
<code>r(max_w)</code>	within maximum

Also see

[XT] **xtdescribe** — Describe pattern of xt data

[XT] **xtab** — Tabulate xt data

