ipolate — Linearly interpolate (extrapolate) values
Description Quick start Menu Syntax Options

References

Also see

Remarks	and	examples	

Description

ipolate creates in *newvar* a linear interpolation of *yvar* on *xvar* for missing values of *yvar*.

Methods and formulas

Because interpolation requires that *yvar* be a function of *xvar*, *yvar* is also interpolated for tied values of *xvar*. When *yvar* is not missing and *xvar* is neither missing nor repeated, the value of *newvar* is just *yvar*.

Quick start

Create y2 containing a linear interpolation of y1 on x for observations with missing values of y1 or tied values of x

ipolate y1 x, generate(y2)

Same as above, but use interpolation and extrapolation

ipolate y1 x, generate(y2) epolate

Same as above, but perform calculation separately for each level of catvar

by catvar: ipolate y1 x, generate(y2) epolate

Menu

Data > Create or change data > Other variable-creation commands > Linearly interpolate/extrapolate values

Syntax

ipolate yvar xvar [if] [in], generate(newvar) [epolate]

by is allowed; see [D] by.

Options

generate (newvar) is required and specifies the name of the new variable to be created.

epolate specifies that values be both interpolated and extrapolated. Interpolation only is the default.

Remarks and examples

Example 1

We have data points on y and x, although sometimes the observations on y are missing. We believe that y is a function of x, justifying filling in the missing values by linear interpolation:

. use https://www.stata-press.com/data/r19/ipolxmpl1

. list, sep(0)

	x	у
1.	0	
2.	1	3
3.	1.5	
4.	2	6
5.	3	
6.	3.5	
7.	4	18
	(

```
. ipolate y x, gen(y1)
(1 missing value generated)
```

```
. ipolate y x, gen(y2) epolate
```

```
. list, sep(0)
```

	x	У	y1	y2
1.	0		•	0
2.	1	3	3	3
З.	1.5		4.5	4.5
4.	2	6	6	6
5.	3		12	12
6.	3.5		15	15
7.	4	18	18	18
	1			

Example 2

We have a dataset of circulations for 10 magazines from 1980 through 2003. The identity of the magazines is recorded in magazine, circulation is recorded in circ, and the year is recorded in year. In a few of the years, the circulation is not known, so we want to fill it in by linear interpolation.

```
. use https://www.stata-press.com/data/r19/ipolxmpl2, clear
```

```
. by magazine: ipolate circ year, gen(icirc)
```

When the by prefix is specified, interpolation is performed separately for each group.

4

Methods and formulas

The value y at x is found by finding the closest points (x_0, y_0) and (x_1, y_1) , such that $x_0 < x$ and $x_1 > x$ where y_0 and y_1 are observed, and calculating

$$y = \frac{y_1 - y_0}{x_1 - x_0} \left(x - x_0 \right) + y_0$$

If epolate is specified and if (x_0, y_0) and (x_1, y_1) cannot be found on both sides of x, the two closest points on the same side of x are found, and the same formula is applied.

If there are multiple observations with the same value for x_0 , then y_0 is taken as the average of the corresponding y values for those observations. (x_1, y_1) is handled in the same way.

References

Cox, N. J. 2023. Speaking Stata: Replacing missing values: The easiest problems. Stata Journal 23: 884-896.

Meijering, E. 2002. A chronology of interpolation: From ancient astronomy to modern signal and image processing. Proceedings of the IEEE 90: 319–342. https://doi.org/10.1109/5.993400.

Also see

[MI] mi impute — Impute missing values

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark 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.