compare — Compare two variables

Description

compare reports the differences and similarities between *varname*$_1$ and *varname*$_2$.

Quick start

Describe differences in missing and defined values of v1 and v2

    compare v1 v2

Same as above, but only for observations where catvar is equal to 3

    compare v1 v2 if catvar==3

Same as above, but for each level of catvar

    by catvar: compare v1 v2

Menu

Data > Data utilities > Compare two variables

Syntax

    compare *varname*$_1$ *varname*$_2$ [*if*] [*in*]

by is allowed; see [D] **by**.

Remarks and examples

▷ Example 1

One of the more useful accountings made by compare is the pattern of missing values:

    . use https://www.stata-press.com/data/r19/fullauto
    (Automobile models)
    . compare rep77 rep78

	Count	Minimum	Difference Average	Maximum
rep77<rep78	16	-3	-1.3125	-1
rep77=rep78	43			
rep77>rep78	7	1	1	1
Jointly defined	66	-3	-.2121212	1
rep77 missing only	3			
Jointly missing	5			
Total	74			

We see that both `rep77` and `rep78` are missing in 5 observations and that `rep77` is also missing in 3 more observations.

◁

❑ Technical note

`compare` may be used with numeric variables, string variables, or both. When used with string variables, the summary of the differences (minimum, average, maximum) is not reported. When used with string and numeric variables, the breakdown by <, =, and > is also suppressed.

❑

Also see

[D] **cf** — Compare two datasets

[D] **codebook** — Describe data contents

[D] **inspect** — Display simple summary of data's attributes