isid — Check for unique identifiers

Description

isid checks whether the specified variables uniquely identify the observations.

Quick start

Verify that idvar uniquely identifies observations
     isid idvar

Verify that idvar uniquely identifies observations within panels identified by pvar
     isid idvar pvar

Same as above
     isid pvar idvar

Same as above, and indicate that the data should be sorted by pvar and idvar
     isid pvar idvar, sort

Verify that idvar uniquely identifies observations in mydata.dta
     isid idvar using mydata.dta

Menu

Data > Data utilities > Check for unique identifiers

Syntax

isid *varlist* [using *filename*] [, <u>s</u>ort <u>m</u>issok]

Options

sort indicates that the dataset be sorted by *varlist*.

missok indicates that missing values are permitted in *varlist*.

Remarks and examples

▷ Example 1

Suppose that we want to check whether the mileage ratings (mpg) uniquely identify the observations in our auto dataset.

```
. use https://www.stata-press.com/data/r19/auto
(1978 automobile data)

. isid mpg
variable mpg does not uniquely identify the observations
r(459);
```

isid returns an error and reports that there are multiple observations with the same mileage rating. We can locate those observations manually:

```
. sort mpg

. by mpg: generate nobs = _N

. list make mpg if nobs >1, sepby(mpg)
```

	make	mpg
1.	Linc. Continental	12
2.	Linc. Mark V	12
	(output omitted)	
68.	Dodge Colt	30
69.	Mazda GLC	30
72.	Subaru	35
73.	Datsun 210	35

◁

▷ Example 2

isid is useful for checking a time-series panel dataset. For this type of dataset, we usually need two variables to identify the observations: one that labels the individual IDs and another that labels the periods. Before we set the data using tsset, we want to make sure that there are no duplicates with the same panel ID and time. Suppose that we have a dataset that records the yearly gross investment of 10 companies for 20 years. The panel and time variables are company and year.

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

. isid company year
```

isid reports no error, so the two variables company and year uniquely identify the observations. Therefore, we should be able to tsset the data successfully:

```
. tsset company year
Panel variable: company (strongly balanced)
 Time variable: year, 1935 to 1954
         Delta: 1 year
```

◁

❑ Technical note

The `sort` option is a convenient shortcut, especially when combined with `using`. The command

```
. isid patient_id date using newdata, sort
```

is equivalent to

```
. preserve
. use newdata, clear
. sort patient_id date
. isid patient_id date
. save, replace
. restore
```

❑

Also see

[D] **describe** — Describe data in memory or in a file

[D] **ds** — Compactly list variables with specified properties

[D] **duplicates** — Report, tag, or drop duplicate observations

[D] **lookfor** — Search for string in variable names and labels

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

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