Syntax

\[ \text{cf} \ varlist \ using \ filename \ [\ , \ all \ verbose] \]

Menu

Data > Data utilities > Compare two datasets

Description

\text{cf} \ compares \ varlist \ of \ the \ dataset \ in \ memory \ (the \ master \ dataset) \ with \ the \ corresponding \ variables \ in \ filename \ (the \ using \ dataset). \ \text{cf} \ \text{returns} \ \text{nothing} \ (that \ is, \ a \ return \ code \ of \ 0) \ if \ the \ specified \ variables \ are \ identical \ and \ a \ return \ code \ of \ 9 \ if \ there \ are \ any \ differences. \ \text{Only} \ \text{the variable values} \ \text{are} \ \text{compared}. \ \text{Variable labels, value labels, notes, characteristics, etc., are not} \ \text{compared}.

Options

\text{all} \ \text{displays} \ \text{the result of the comparison for each variable} \ \text{in} \ varlist. \ \text{Unless} \ all, \ \text{only} \ \text{the results of the variables that differ are displayed}.

\text{verbose} \ \text{gives} \ \text{a} \ \text{detailed} \ \text{listing, by variable, of each observation that differs}.

Remarks and examples

\text{cf} \ \text{produces} \ \text{messages} \ \text{having} \ \text{the following} \ \text{form}:

\begin{verbatim}

 varname: does not exist in using
 varname: ___ in master but ___ in using
 varname: ___ mismatches
 varname: match

\end{verbatim}

An example of the second message is “str4 in master but float in using”. \text{Unless} all, \text{the fourth message} \ \text{does not appear—silence} \ \text{indicates matches}.

Example 1

\text{We think the dataset in memory is identical to mydata.dta, but we are unsure. We want to understand any differences before continuing}:

\begin{verbatim}

 . \ cf \ _all \ using \ mydata
 . _

\end{verbatim}
All the variables in the master dataset are in mydata.dta, and these variables are the same in both datasets. We might see instead

```stata
  . cf _all using mydata
      mpg: 2 mismatches
    headroom: does not exist in using
   displacement: does not exist in using
  gear_ratio: does not exist in using
  r(9);
```

Two changes were made to the mpg variable, and the headroom, displacement, and gear_ratio variables do not exist in mydata.dta.

To see the result of each comparison, we could append the all option to our command:

```stata
  . cf _all using mydata, all
      make: match
     price: match
      mpg: 2 mismatches
    rep78: match
  headroom: does not exist in using
    trunk: match
      weight: match
     length: match
   turn: match
   displacement: does not exist in using
  gear_ratio: does not exist in using
     foreign: match
  r(9);
```

For more details on the mismatches, we can use the verbose option:

```stata
  . cf _all using mydata, verbose
      mpg: 2 mismatches
      obs 1. 22 in master; 33 in using
      obs 2. 17 in master; 33 in using
    headroom: does not exist in using
   displacement: does not exist in using
  gear_ratio: does not exist in using
  r(9);
```

This example shows us exactly which two observations for mpg differ, as well as the value stored in each dataset.

Example 2

We want to compare a group of variables in the dataset in memory against the same group of variables in mydata.dta.

```stata
  . cf mpg headroom using mydata
      mpg: 2 mismatches
    headroom: does not exist in using
  r(9);
```
Stored results

cf stores the following in r():

Macros

r(Nsum)  number of differences

Methods and formulas

If you are using Small Stata, you may get the error “too many variables” when you stipulate _all and have many variables in your dataset. (This will not happen if you are using Stata/MP, Stata/SE, or Stata/IC.) If this happens, you will have to perform the comparison with groups of variables. See example 2 for details about how to do this.

Acknowledgment

Speed improvements in cf were based on code written by David Kantor.

Reference


Also see

[D] compare — Compare two variables