Description

`cf` compares `varlist` of the dataset in memory (the master dataset) with the corresponding variables in `filename` (the using dataset). `cf` returns 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. Only the variable values are compared. Variable labels, value labels, notes, characteristics, etc., are not compared.

Quick start

Compare values of `v1` and `v2` from `mydata1.dta` in memory to `mydata2.dta`
```
cf v1 v2 using mydata2
```
As above, but give a detailed listing of the differences
```
cf v1 v2 using mydata2, verbose
```
As above, but for all variables in memory
```
cf _all using mydata2, verbose
```

Menu

Data > Data utilities > Compare two datasets
Syntax

```
cf varlist using filename [, all verbose]
```

Options

- `all` displays the result of the comparison for each variable in `varlist`. Unless `all` is specified, only the results of the variables that differ are displayed.
- `verbose` gives a detailed listing, by variable, of each observation that differs.

Remarks and examples

```
cf produces messages having the following form:
    varname: does not exist in using
    varname: __ in master but __ in using
    varname: __ mismatches
    varname: match
```

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

Example 1

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

```
. cf _all using mydata
```

All the variables in the master dataset are in `mydata.dta`, and these variables are the same in both datasets. We might see instead

```
. 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
```

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:

```
. 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:

```
. 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`.

```
. 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

#### Acknowledgment

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

#### Also see

[D] `compare` — Compare two variables