Stata tip 35: Detecting whether data have changed

William Gould
StataCorp
College Station, TX
wgould@stata.com

Included in the 17 May 2006 update is a new command that you may find useful, 
\texttt{datasignature}. If you have installed the update, type \texttt{help datasignature}. If you have not updated, or are unsure, type \texttt{update query} to find out, and type \texttt{update all} to install.

Here is the result of running \texttt{datasignature} on \texttt{auto.dta}:

\begin{verbatim}
. sysuse auto
   (1978 automobile data)
. datasig
74:12(71728):3831085005:1395876116
\end{verbatim}

The output is \texttt{auto.dta}'s data signature. If you change the data, even just a little bit, the last two numbers will change:

\begin{verbatim}
. replace mpg = mpg + 1 in 2
   (1 real change made)
. datasig
74:12(71728):1616229321:1400086868
\end{verbatim}

If you change the name of a variable, the last two numbers stay the same and the number inside the parentheses changes:

\begin{verbatim}
. rename mpg miles_per_gallon
. datasig
74:12(57876):1616229321:1400086868
\end{verbatim}

\texttt{datasignature} is designed to help those who use data maintained by others and those who worry that they might themselves have accidentally changed their data.

In the latter case, you could save the signature in the dataset,

\begin{verbatim}
. datasig
74:12(57876):1616229321:1400086868
. note: `r(datasignature)'
\end{verbatim}

and check it later,

\begin{verbatim}
. notes
_dta:
1. from Consumer Reports with permission
2. 74:12(57876):1616229321:1400086868
\end{verbatim}
Another idea is to include `datasetsignature` at the beginning of logs:

```plaintext
begin myfile.do
log using myfile, replace
sysuse auto, clear
datasig
...
log close
end myfile.do
```

You can specify a varlist and `if` and `in`, so if you have a large dataset and want to check just part of it, you can write:

```plaintext
begin myfile.do
log using myfile, replace
sysuse auto, clear
datasig mpg weight price if foreign
...
log close
end myfile.do
```