

```
. keep logbw
```

will drop all variables except for `logbw`. Once data are dropped there is no way of getting them back other than by re-loading the whole dataset with

```
. use births, clear
```

where `clear` gives permission for the memory to be cleared before the `births` data are re-loaded.

## 1.8 Sorting data

Stata can sort the records in a file according to values (numeric or string) of a variable. The file is not physically re-arranged – instead a key is created which tells Stata commands the order in which the records should be processed. Try the following:

```
. list id matage sex sexalph in 1/10
. sort matage
. list id matage sex sexalph in 1/10
```

The records are now sorted in ascending order of maternal age. To sort on `id` within `matage` try

```
. sort matage id
. list id matage sex sexalph in 1/10
```

You will see that within each value of `matage` the records are sorted in order of `id`. To sort on a string variable try

```
. sort sexalph
. list id matage sex sexalph
```

Note that `female` comes before `male` because `f` comes before `m` in the alphabet. To restore the original sort order, try

```
. sort id
```

## 1.9 Using Stata as a calculator

The `display` command can be used to carry out simple calculations. For example, the command

```
. display 2+2
```

will display the answer 4, and

```
. display 2^3
```

will display the answer 8. The command

```
. display ln(10)
```

will display the natural logarithm of 10, which is 2.3026, and

```
. display sqrt(25)
```

will display the square root of 25. Text can also be displayed as in

```
. display "The natural logarithm of 10 is " ln(10)
```

Note that because the text contains spaces it must be surrounded by the quotes symbol " which is usually found above the number 2. The result can also be color-coded as in

```
. display as text "Square root of 25 is " as result sqrt(25)
```

The keywords `as text` and `as result` determine the colors: when the background is black `as text` displays are green and `as result` displays are yellow. Other display styles are `as input` (white) and `as error` (red).

Standard probability functions are readily available. For example, the probability below 1.96 in a standard normal (i.e. Gaussian) distribution is obtained with

```
. display norm(1.96)
```

while

```
. display 1 - norm(1.96)
```

will display the probability above 1.96. Similarly,

```
. display chi2(1,3.84)
```

will display the probability below 3.84 in a chi-squared distribution on 1 degree of freedom, and

```
. display chi2tail(1,3.84)
```

will display the probability above 3.84. Try `help probfun` for a full list of available probability functions.

## 1.10 Shortcuts

Variable names can be abbreviated, as long as the abbreviation is unique. Try

```
. list id matage hyp gestwks in 1/10  
. list id mat hy gest in 1/10  
. list i m h g in 1/10
```

Also lists of variable names can be shortened if they are consecutive

```
. list id-gestwks in 1/10
. list i-g in 1/10
```

or if they share some unique initial letters:

```
. list se* in 1/10
```

where `se*` stands for “all variables with names starting with `se`”.

Command names, as well as options within commands, can be abbreviated (with a few exceptions). Try

```
. sum matage
. l matage in 1/10
. br matage
. tab sex
```

Note that `tab` is accepted as an abbreviation for `tabulate`, *not* for `table`, which must be typed in full.

## 1.11 Stata syntax

The word syntax here refers to the rules which govern how a Stata command is put together. The heart of any Stata command takes the form

*command varlist if\_expression in\_range, options*

For example, try

```
. list bweight hyp if sex==1 in 1/10, table noobs
```

The *command* is `list`, the *varlist* is `bweight hyp`, the *if\_expression* is `if sex==1`, the *in\_range* is `in 1/10`, and the *options* are `table noobs`. Adding weights to a Stata command will be covered in Chapter 7.

## 1.12 Using the Stata help facilities

At the end of the top line on the Stata screen you will see the *Help* tab. Click on this and a small menu will appear. Click on *Stata Command* and you will see another menu in which the name of the command should be entered. For example, enter `list` and press *OK*. Alternatively just type

```
. help list
```

Each Stata command has a help file, but the amount of information can be rather overwhelming. One useful bit to look at is the line which shows the syntax:

```
list [varlist] [if exp] [in range] [, options ]
```

Parts of the syntax which are not essential are shown inside square brackets [ ]. In Stata commands the options are always entered after a comma. The syntax for `list` shows that there are several options available. If you scroll down the help screen you will see that they are described under the heading *Options*. Looking further down at the *Examples* section, you find some of the common ways in which the command is used.

Another sort of help available from the Stata screen is help on commands linked to particular operations. Cancel the current help screen, click on *Help*, and then on *Search*. Enter the keyword *tables* to search on, and you will see a long list of Stata commands relevant to making tables. These start with official commands from Stata and then go on to contributed commands which have been written by users and published in the *Stata Journal* (previously the *Stata Technical Bulletin*). We shall return to contributed commands in Chapter 20.

## Exercises

1. Load the `births` data.
2. List the variables `bweight` and `hyp` for observations 20–25 inclusive.
3. Summarize all variables.
4. Summarize `matage` in detail.
5. Use `codebook` to find out more about `sexalph`.
6. Use `count` to find out how many hypertensive women there are.
7. Summarize `matage` for hypertensive women.
8. Use `count` to find how many hypertensive women have babies with birth weight less than 2000 g.
9. Use `count` to find how many women over 30 are hypertensive.
10. Tabulate the values of `sex`.
11. Make a table of mean birth weight by `sex`.
12. Make a table of median birth weight by `sex`.
13. Generate a new variable called `bwkg` which is the birth weight in kilograms.
14. Use `display` to calculate  $\sqrt{3^2 + 4^2}$ .

15. Use `display` to find the probability above 4.3 in a chi-squared distribution on 1 degree of freedom.

Answers to these exercises can be obtained by running the program `chap1` with the command

```
. chap1
```

To get the answer for question 8 only, try

```
. chap1, q(8)
```

Chapter 2 has no answers, but answers for chapter 3 can be obtained with

```
. chap3
```

and similarly for the other chapters.