EPSY 887: Computation Statistics

Working with Data

Jason Bryer

http://github.com/jbryer/CompStats
jason@bryer.org

Week 2
February 4, 2013

Agenda

1. Reading Data
2. Viewing and Subsetting Data
3. Descriptive Statistics

Outline

Reading Data

read.table	Reads in a table where each line is a record. Lots of options to define the structure of the file.
read.csv	Comma deliminited files.
read.spss	In the `foreign` package, reads SPSS files.
read.xls	In the `gdata` package, reads Excel files.
RODBC	This package has functions to read data from most ODBC databases.
RMySQL	Package for reading data from MySQL databases.
RPoostgreSQL	Package for reading data from PostgreSQL databases.
load	Read in R data object files saved using the `save`. This is very useful for saving intermediate data files.

Outline

Subsetting Lists and Vectors

```
> mylist <- list(letters=letters, numbers=1:10)
> class(mylist)

[1] "list"

> str(mylist)

List of 2
 $ letters: chr [1:26] "a" "b" "c" "d" ...
 $ numbers: int [1:10] 1 2 3 4 5 6 7 8 9 10

> length(mylist)

[1] 2
```

Subsetting Lists and Vectors

```
> mylist[1]

$letters
 [1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j" "k"
[12] "l" "m" "n" "o" "p" "q" "r" "s" "t" "u" "v"
[23] "w" "x" "y" "z"

> mylist[[1]]

 [1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j" "k"
[12] "l" "m" "n" "o" "p" "q" "r" "s" "t" "u" "v"
[23] "w" "x" "y" "z"

> mylist$letters

 [1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j" "k"
[12] "l" "m" "n" "o" "p" "q" "r" "s" "t" "u" "v"
[23] "w" "x" "y" "z"

> mylist$numbers

 [1]  1  2  3  4  5  6  7  8  9 10
```

Subsetting Data Frames and Matrices

```
> data(mtcars)
> head(mtcars)
```

	mpg	cyl	disp	hp	drat	wt	qsec
Mazda RX4	21	6	160	110	3.9	2.6	16
Mazda RX4 Wag	21	6	160	110	3.9	2.9	17
Datsun 710	23	4	108	93	3.9	2.3	19
Hornet 4 Drive	21	6	258	110	3.1	3.2	19
Hornet Sportabout	19	8	360	175	3.1	3.4	17
Valiant	18	6	225	105	2.8	3.5	20

	vs	am	gear	carb
Mazda RX4	0	1	4	4
Mazda RX4 Wag	0	1	4	4
Datsun 710	1	1	4	1
Hornet 4 Drive	1	0	3	1
Hornet Sportabout	0	0	3	2
Valiant	1	0	3	1

Subsetting Data Frames and Matrices

```
> tail(mtcars)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs
Porsche 914-2	26	4	120	91	4.4	2.1	17	0
Lotus Europa	30	4	95	113	3.8	1.5	17	1
Ford Pantera L	16	8	351	264	4.2	3.2	14	0
Ferrari Dino	20	6	145	175	3.6	2.8	16	0
Maserati Bora	15	8	301	335	3.5	3.6	15	0
Volvo 142E	21	4	121	109	4.1	2.8	19	1

	am	gear	carb
Porsche 914-2	1	5	2
Lotus Europa	1	5	2
Ford Pantera L	1	5	4
Ferrari Dino	1	5	6
Maserati Bora	1	5	8
Volvo 142E	1	4	2

Subsetting Data Frames and Matrices

```
> mtcars[1,]
           mpg cyl disp  hp drat  wt qsec vs am
Mazda RX4   21   6  160 110  3.9 2.6   16  0  1
           gear carb
Mazda RX4     4    4
> mtcars[,2]
 [1] 6 6 4 6 8 6 8 4 4 6 6 8 8 8 8 8 8 4 4 4 4 8 8
[24] 8 8 4 4 4 8 6 8 4
> mtcars$cyl
 [1] 6 6 4 6 8 6 8 4 4 6 6 8 8 8 8 8 8 4 4 4 4 8 8
[24] 8 8 4 4 4 8 6 8 4
> mtcars[,c("cyl","disp")]
                     cyl disp
Mazda RX4              6  160
Mazda RX4 Wag          6  160
Datsun 710             4  108
Hornet 4 Drive         6  258
Hornet Sportabout      8  360
```

Outline

Descriptive Statistics

table	Crosstabs.
summary	Provides summary information relevant to the type.
describe	In the `psych`, provides many of the most common descriptives statistics (e.g. mean, median, standard deviation, range, etc.)
describeBy	Same as `describe` but will provide descriptive stats based upon grouping varaible(s).
fivenum	Returns Tukey's five number summary (minimum, lower-hinge, median, upper-hinge, maximum)
mean	Mean
median	Median
sd	Standard deviation
var	Variance