Biometry. Lecture 10

Alexey Shipunov

Minot State University

February 25, 2015

Outline

Outline

> setwd("<working folder>")

or

"Change dir"

in menu!

On Mac, be sure that startup option is working: getwd()
(getwd() checks if R is in working folder, dir() checks the folder content)

M

R and data
Lists

Indexing of vectors, matrices and lists

```
> m[3] # third element of vector m
> ma[2, 1] # second row, first column
> l[1] # lists may be indexed like vectors
> str(l[1]) # it's a list!
> l[[1]] # not the same as l[1]!
> str(l[[1]]) # it's a vector!
```

Names

In R, elements of vectors and lists, columns and rows of matrices may have *names*:

```
> names(fred)
> fred$wife # this is a selection by name
> w <- 60:66
> names(w) <- c("Rick", "Amanda", "Peter", "Alex",
+ "Kathy", "Ben", "George")
> w
> w["Rick"]
> rownames(ma) <- c("a1", "a2")
> colnames(ma) <- c("b1", "b2")
> ma
```

R and data
Data frames (data tables)

More important than any other object

- This is a most important type of object; most of data are represented by data frames
- *Date frame is a list of vectors of same length*

How to create a data frame

```
> x <- 171:177
> sex.f <- c("m","m","f","f","f","m","f")
> m.o <- c("L","XL","S","M","S","M","XL")
> d <- data.frame(weight=w, height=x, size=m.o, sex=sex.f)
> d
> str(d)
```

Selection from data frames

```
> d$weight # by name
> d[[1]] # by number, as list
> d[,1] # by number of column, as matrix
> d[,"weight"] # by name of column
> d[,2:4] # columns 2, 3, 4
> d[,-1] # all columns except first
> d[-1,] # all rows except first
```

Conversions

```
> b <- 1:8 # vector
> dim(b) <- c(4,2) # two columns, four rows
> b <- data.frame(b) # convert to data frame
> b[2, 2] <- "string" # replace one number with characters
> b[!is.na(as.numeric(b[,2])),] # remove all rows with chars
```

Selection by condition

```
> d[d$sex=="f",] # will select only women
> d[d$sex!="f",] # will select all other genders ;)
```

== is "equal?", & "and", | "or" and ! is "not"

Sorting and ordering

```
> sort(x) # ascending
> rev(sort(x)) # descending
> d[order(d$sex, d$height), ] # sort by sex then by height
```

One-dimensional data
Central tendency

Mean and median

- These are two most frequently used characteristics of the central tendency.
- Median is more robust than mean.

Mean and median

```
> salary <- c(21, 19, 27, 11, 102, 25, 21)
> mean(salary); median(salary)
> median(1:3); median(1:4)
```

When number of elements is odd, median is a central value; if even—median is the average between two centrals.

Median is the third quartile

Quartiles take out 0% (minimum, `min()`), 25% (lower hinge) , 50%, 75% (upper hinge) and 100% (maximum, `max()`) of ordered data. Median is simply a 50% (third) quartile.

```
> fivenum(salary)
```

Mode

Mode is the most frequent value:

```
> sex <- c("m", "f", "m", "m", "f", "m", "m")
> t.sex <- table(sex)
> mode <- t.sex[which.max(t.sex)]
> mode
```

Finishing...

Save your commands!

(savehistory(<todaysdate>.r) or File -> Save as... on Mac)

Summary: most important commands

- [—selects an element, row or column
- $—selects by name from list or data frame
- median()—returns a median value

For Further Reading

A. Shipunov.
Biometry [Electronic resource].
2012—onwards.
Mode of access:
http://ashipunov.info/shipunov/school/biol_240

A. Shipunov, and others.
Visual statistics. Use R!
Ongoing translation from Russian.