Biometry. Lecture 15

Alexey Shipunov

Minot State University

March 25, 2015

Outline

Starting...

`> 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)

Two-dimensional statistics
Hypotheses and tests

Null and alternative for two numeric samples

- Null: difference equal to 0 $\approx$ similar $\approx$ related $\approx$ samples came from same population
- Alternative: difference not equal to 0 $\approx$ different $\approx$ non-related $\approx$ samples came from different populations

p-value

- p-value is the probability to have equal or greater effect in case if null hypothesis is true
- p-value is related with the accepted level of Type I statistical error (false alarm): the bigger is the significance level of p-value, the more false alarms we accept and at the same time, the more sensitive is our research
- In biology, the most common level of significance is 0.05
- As a rule, if p-value is equal or less then 0.05, we reject the null hypothesis, if more – we stay with null hypothesis

Two main questions

- Normal?
  `t.test()` or `wilcox.test()`
- Paired?
  `(..., paired=T)` or `(...)`

Two-dimensional statistics

Tests for the independence of two variables

Leaves example (continued)

```
> leaves <- read.table(
+ "http://ashipunov.info/data/leaves.txt", h=T)
control <- leaves[,2] - leaves[,1]
exp <- leaves[,4] - leaves[,3]
t.test(control, exp)
```

“Classical” sleep data and model formula

```
> str(sleep)
> boxplot(extra ~ group, data=sleep)
> t.test(extra ~ group, data=sleep)
```

`sleep` is a data in so-called long format, `extra ~ group` is a **model formula** of `response ~ factor` form.
For t-test, “`group`” should have exactly 2 levels!

Model formula for leaves data

```
> leaves12 <- stack(leaves[,1:2])
> leaves12
> t.test(values ~ ind, data=leaves12, paired=T)
```

`stack()` converts from short to long form

Air quality data in May and August

```
> str(airquality)
> air15 <- unstack(airquality[,c(1,5)])
> Normality3(air15)
> boxplot(Ozone ~ Month, data=airquality,
+ subset=Month %in% c(5,8))
> wilcox.test(Ozone ~ Month, data=airquality,
+ subset=Month %in% c(5,8))
```

`unstack()` converts from long to short form

`%in%` is a selection operator

Two-dimensional statistics

Test for tables: chi-squared

Contingency tables

- Secondary data: counts
- May be created from any categorical variable, or from measurement variable after cutting

table() function

```
> with(airquality, table(cut(Temp, quantile(Temp)), Month))
> d <- factor(rep(c("A","B","C"), 10))
> is.na(d) <- 3:4
> table(d, exclude=NULL)
```

Graphical representation of tables

```
> Titanic # this is a multidimensional table
> ftable(Titanic, row.vars = 1:3)
> margin.table(Titanic, c(1, 4)) # make 2 dimensions
> mosaicplot(margin.table(Titanic, c(1, 4)))
```

- Chi-squared test checks the null if *variables in the table are distributed independently* (non-accordingly) between cells.
- Alternative hypothesis is that association between variables exists.

Chi-squared test

```
> HairEyeColor # multidimensional table
> margin.table(HairEyeColor, c(1, 2)) # hairs and eyes
> chisq.test(margin.table(HairEyeColor, c(1, 2)))
> margin.table(HairEyeColor, c(2, 3)) # eyes and sex
> chisq.test(margin.table(HairEyeColor, c(2, 3)))
```

Association plot

```
> assocplot(margin.table(HairEyeColor, c(1, 2)))
```

Association plots show positive and negative association between factors in the table. The key thing is the **asymmetry between the squares**.

Finishing...

Save your commands!

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

Summary: most important commands

- `response ~ factor`—if factor has exactly two levels, this is a model formula for two-sample test
- `%in%`—selection operator
- `table()`—creates contingency tables
- `chisq.test()`—test for independence of rows and columns

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.