Biometry. Lecture 15

Alexey Shipunov

Minot State University

March 30, 2016

Outline

1. Two-dimensional statistics

2. Effect sizes
   - Test for tables: chi-squared

Outline

1. Two-dimensional statistics
2. Effect sizes
   - Test for tables: chi-squared

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)

Quiz question (10 points!)

Quiz question (10 points!)

These are points from the first and second exam in one small class:
63, 72, 77, 76, 67, 56, 55, 51, 77, 64
and
87, 86, 76, 79, 54, 60, 97, 80, 73, 97
Both exams were equivalent. Provide a statistical support for the hypothesis that second exam went better. Report commands and all values which support your conclusion.

Effect sizes: parametric and non-parametric

- p-values do not show the extent of differences
- Means and alike are not working because data could intersect
- Consequently, effect sizes were invented. They allow to understand how big is a difference
- R library `effsize` provides parametric Cohen D and non-parametric Cliff's Delta effect sizes

Effect sizes: examples

```
> library(effsize)
> a <- 1:9
> b <- rep(5, 9)
> cliff.delta(a, b) # no effect
> cohen.d(extra ~ group, data=sleep)
```

In the `sleep` example, effect size is present even if p-value is not good enough!

Effect sizes
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
- `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 many others.
Visual statistics. Use R!
2016—onwards.
Mode of access: http://ashipunov.info/shipunov/school/biol_240/en/visual_statistics.pdf