Biometry. Lecture 17

Alexey Shipunov

Minot State University

March 21, 2012

Outline

1. Questions and answers
2. Two-dimensional statistics
   - Test for tables: chi-squared

Outline

Starting...

```
> setwd("<working folder>")
```

or

“`Change dir`”

in menu!

Previous final question: the answer

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.

```
> first <- c(63, 72, 77, 76, 67, 56, 55, 51, 77, 64)
> second <- c(87, 86, 76, 79, 54, 60, 97, 80, 73, 97)
```

```
> wilcox.test(first, second, alt="greater", paired=TRUE)

Wilcoxon signed rank test with continuity correction

data:  first and second
V = 9.5, p-value = 0.9704
alternative hypothesis: true location shift is greater
than 0

> wilcox.test(first, second, alt="greater")

Wilcoxon rank sum test with continuity correction

data:  first and second
W = 21.5, p-value = 0.9859
alternative hypothesis: true location shift is greater
than 0
```

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 of squares.

Food intoxication example

- The poisoning took place on the party of Epidemiology Statistics association
- 13 food choices and 45 persons
- Data file `tox.txt`: `ILL` $= 1$ (poisoned), $= 2$ (not poisoned)

```
> tox <- read.table("http://ashipunov.info/data/tox.txt",
+ h=TRUE)
> str(tox)
> head(tox)
> for (m in 2:ncol(tox))
+ {
+ tmp <- chisq.test(tox$ILL, tox[,m])
+ print(paste(names(tox)[m], tmp$p.value))
+ }
> assocplot(table(ILL=tox$ILL, CAESAR=tox$CAESAR))
> assocplot(table(ILL=tox$ILL, TOMATO=tox$TOMATO))
```

Final question (2 points)

Final question (2 points)

What is a null hypothesis for the chi-squared test?

Summary: most important commands

- `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_299

P. Dalgaard
Introductory Statistics with R. 2nd edition.
Springer, 2008.
Chapters 5 and 8.