Biometry. Lecture 16

Alexey Shipunov

Minot State University

April 4, 2016

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)

2D 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

Null and alternative for chi-squared

- Null: independent distribution $\approx$ equal counts in each cell $\approx$ no pattern
- Alternative: concerted distribution $\approx$ non-equal counts in each cell $\approx$ pattern

2D statistics
Exact and approximate tests

Chi-squared and Fisher exact

- Chi-squared proportion tests will **estimate** the p-value from theoretical distribution. As a consequence, it may say "*Chi-squared approximation may be incorrect*".
- Fisher exact and binomial tests will **calculate** p-value directly. That is why they are sometimes preferable.

Fisher's tea drinker

A British woman claimed to be able to distinguish whether milk or tea was added to the cup first. To test, she was given 8 cups of tea, in four of which milk was added first.

```
> tea <- matrix(c(3,1,1,3), nrow=2)
> colnames(tea) <- row.names(tea) <- c("Milk", "Tea")
> tea
> chisq.test(tea) # warning!
> fisher.test(tea) # note odds ratio
```

How to avoid the approximation with simulation

```
> chisq.test(tea, simulate.p.value=T) # no warning!
```

When some cells contain less than 5 items, `simulate.p.value=T` is recommended.

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))
```

As it was said above, here is better to use `simulate.p.value=T` for chi-squared test.

Multiple comparisons

- If we apply multiple tests to one component (e.g., test several samples against one), probability to obtain incorrect answer will grow
- To keep the error rate low, one should apply *Bonferroni* or another p-value correction, in other words—increase p-values to avoid the growing error

Toxicity with correction

```
> tox <- read.table("http://ashipunov.info/data/tox.txt",
+ h=TRUE)
> answer <- data.frame(FOOD=NA, CHISQ.P=NA)
> for (m in 2:ncol(tox))
+ {
+ tmp <- chisq.test(tox$ILL, tox[,m])
+ answer[m-1,] <- c(names(tox)[m], tmp$p.value)
+ }
> answer[,2] <- p.adjust(answer[,2])
> answer
```

2D statistics

Sign test

Sign test

- Idea is simple: to calculate differences between all pairs of values (paired test!)
- Then take only positive differences
- If two samples came from a same distribution, approximately 50% of differences should be positive—we can test with with, e.g., binomial test

Making the sign test

These are points from the first and second exam in one small class. Both exams were equivalent. Are second exam results better?

```
> first <- c(63, 72, 77, 76, 67, 56, 55, 51, 77, 64)
> second <- c(87, 86, 76, 79, 54, 60, 97, 80, 73, 97)
> dif <- second - first
> pos.dif <- dif[dif > 0]
> binom.test(length(pos.dif), length(dif))
```

2D statistics
Concordance and Cohen kappa

- Concordance is a measure of “agreement” between two expert answer sheets
- The most common application are psychological tests
- Cohen kappa test is frequently used for understanding the degree of concordance; **the null hypothesis for Cohen kappa is that two answer sheets are non-concordant**

Cohen kappa and island flora

```
> source("http://ashipunov.info/r/concordance.r")
> isl <- read.table(
+ "http://ashipunov.info/data/pokorm_03.dat",
+ h=TRUE, sep=";")
> str(isl); head(isl)
> cohen.kappa(as.matrix(isl))
```

Finishing...

Save your commands!

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

Summary: most important commands

- `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