

Biometry. Lecture 21

Alexey Shipunov

Minot State University

April 28, 2014



- 1 Two-dimensional statistics
 - Exact and approximate tests
 - Bootstrapping in R
 - Logistic regression



```
> setwd("<working folder>")  
or  
"Change dir"  
in menu!
```



Advanced scripting: “Nanodrop” Lab

```
> download.file("http://ashipunov.info/data/nanodrop.r",  
+ "nanodrop.r")  
> file.show("nanodrop.r")
```



Two-dimensional 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.
- Somehow similar difference exists between t-test and Wilcoxon test. The later sometimes says “*Cannot compute exact p-values with ties*”. Pearson (default) and Spearman correlations are also different this way.



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)
```



How to avoid the approximation with simulation

```
> eq <- read.table("http://ashipunov.info/data/eq.txt", h=T)
> table(eq$N.REB, eq$N.ZUB) # less than 5 in cells
> chisq.test(eq$N.REB, eq$N.ZUB)
> chisq.test(tea, simulate.p.value=T) # no warning!
> chisq.test(eq$N.REB, eq$N.ZUB, simulate.p.value=T)
```

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



Two-dimensional statistics

Bootstrapping in R



Bootstrap and jackknife

- **Bootstrap** is the resampling method which replaces part of sample with the subsample of its own
- **Jackknife** is similar but in this case observations will be taking out without replacement

These methods were developed in 1980s (B. Efron, 1979–1982) but only recently started to be widely used



How to bootstrap the median

```
> library(boot)
> med <- function (data, indices)
> {
> d <- data[indices]
> return(median(d))
> }
> result <- boot(data=trees$Height, statistic=med, R=100)
> result
> plot(result)
> boot.ci(result, type="bca")
```



How to jackknife the median

```
> library(bootstrap)
> theta <- function (x)
> {
> median(x)
> }
> (results <- jackknife(x=trees$Height, theta))
```



Two-dimensional statistics

Logistic regression



Numeric influence but categorical response

- What if response is binary?
- It is possible to convert success/failure to the **probability of success** and then apply a **generalized linear model**



Analysis of logistic regression

```
> lo <- read.table("http://ashipunov.info/data/logit.txt")
> head(lo); str(lo)
> lo.logit <- glm(formula=V2 ~ V1, family=binomial,
+ data=lo)
> summary(lo.logit)
```



Finishing...

```
> savehistory("20140428.r")
```



Final question (3 points)



Final question (3 points)

What is a logistic regression?



Summary

- **Exact tests** will not need approximations but they are normally less powerful than “regular” tests
- **Bootstrap** and **jackknife** are powerful ways to know about your data avoiding any tests
- `glm()` —estimates the logistic regression model and many others



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!

DMK Press, 2012. [Under translation from Russian.]

