Biometry. Lecture 24

Alexey Shipunov

Minot State University

April 18, 2012

Outline

Minot State UNIVERSITY

Outline

Minot State UNIVERSITY

Starting...

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

or

“Change dir”

in menu!

Previous final question: the answer

What is a logistic regression?

Previous final question: the answer

What is a logistic regression?

- Analysis of categorical (binary) response from numerical input

Lab 10, Part I

```
> download.file("http://ashipunov.info/data/eq.txt",
+ "eq.txt")
> eq <- read.table("eq.txt", h=T)
> str(eq); head(eq)
> download.file("http://ashipunov.info/data/normality3.r",
+ "normality3.r")
> source("normality3.r")
> Normality3(eq[,-1])
> wilcox.test(DIA.ST ~ SPECIES, data=eq)
> chisq.test(eq$N.REB, eq$N.ZUB)
```

Lab 10, Part II

```
> cor(eq[,-c(1, 4:5)], method="spearman")
> cor.test(eq$DL.R, eq$DIA.ST, method="spearman")
> eq.lm <- lm(DIA.ST ~ DL.R, data=eq)
> summary(eq.lm)
```

Lab 10, Part III

```
> eq.ancova <- lm(DIA.ST ~ DL.R * SPECIES, data=eq)
> summary(eq.ancova)
# Below are extra commands:
> eq.ancova2 <- lm(DL.R ~ DIA.ST * SPECIES, data=eq)
> summary(eq.ancova2)
> eq.ancova3 <- update(eq.ancova2, . ~ . - DIA.ST:SPECIES,
+  data=eq)
> summary(eq.ancova3)
> AIC(eq.ancova2)
> AIC(eq.ancova3)
```

Minot State UNIVERSITY

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

```
> table(eq$N.REB, eq$N.ZUB) # less than 5 in cells
> chisq.test(eq$N.REB, eq$N.ZUB) # 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
Logistic regression

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("20120418.r")
```

Final question (3 points)

Final question (3 points)

What is AIC? How to use it?

Summary: most important commands

- `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_299`

P. Dalgaard
Introductory Statistics with R. 2nd edition.
Springer, 2008.
Chapters 5, 6, 8, 12, 13.

Minot State UNIVERSITY