

# Biometry. Lecture 21

Alexey Shipunov

Minot State University

April 25, 2016



- 1 2D statistics
  - Logistic regression
  - ANalysis Of VAriation (ANOVA)



```
> 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)



# Finishing...

Save your commands!

`(savehistory(<today's date>.r)` or File -> Save as... on  
Mac)



# About eggs lab

Selection and droplevels ()



# 2D statistics

## Logistic regression



# Analysis of logistic regression

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



# Visualizing logistic regression

```
> new.points <- seq(min(lo$V1), max(lo$V1), length.out=14)
> predicted.points <- predict(lo.logit,
+ list(V1=new.points), type="response")
> success <- as.numeric(lo$V2) - 1
> plot(success ~ V1, data=lo)
> lines(new.points, predicted.points)
```



# Logistic regression example II: poisoning

- Caesar or tomatoes?
- High significance of both terms could be a result of coincidence (people often took these things together)
- If we construct a logistic model and then update it (taking out one of two terms), AIC will show which model is better.



# Poisoning analysis

```
> tox <- read.table("http://ashipunov.info/data/tox.txt",
+ h=TRUE)
> str(tox)
> tox.logit <- glm(formula=I(2-ILL) ~ CAESAR + TOMATO,
+ family=binomial, data=tox)
> tox.logit2 <- update(tox.logit, . ~ . - TOMATO)
> tox.logit3 <- update(tox.logit, . ~ . - CAESAR)
> tox.logit$aic
> tox.logit2$aic # lowest!
> tox.logit3$aic
> summary(tox.logit2) # highly significant!
```

## Caesar!



# 2D statistics

## ANalysis Of VAriation (ANOVA)



# Categorical influence and numerical response

- What if we want to compare all 11 types of eggs?
- There is a solution: analysis of variation (ANOVA) and its non-parametric counterpart, Kruskal-Wallis test.



# ANOVA null and alternative hypotheses and assumptions

- Null is that all groups are not different, alternative is that **at least one group is different from all others**.
- All variables **should be normally distributed**. Small deviations from normality are typically accepted but Kruskal-Wallis test is preferable for all “non-normal” data.
- Variances should be at least **similar**. This is possible to overcome with `oneway.test()`
- Actually, t-test is just an ANOVA for only two groups. Wilcoxon test, in turn, is a special case of Kruskal-Wallis test for two groups only.



# Introductory example: artificial data

See the textbook



# Summary: most important commands

- `predict()` – predicts values with model
- `glm()` – generalized linear models, including logistic
- `aov()` – classic one-way ANOVA



# For Further Reading



A. Shipunov.

*Biometry* [Electronic resource].

2012—onwards.

Mode of access:

[http://ashipunov.info/shipunov/school/biol\\_240](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](http://ashipunov.info/shipunov/school/biol_240/en/visual_statistics.pdf)

