Biometry. Lecture 20

Alexey Shipunov

Minot State University

March 30, 2012

Minot State
UNIVERSITY

Outline

1 Questions and answers

2 Two-dimensional statistics
  - Regression
  - Analysis of covariation

Minot State
UNIVERSITY

Outline

1. Questions and answers

2. Two-dimensional statistics
   - Regression
   - Analysis of covariation

Minot State
UNIVERSITY

Starting...

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

Minot State
UNIVERSITY

Previous final question: the answer

What is a difference between regression and correlation?

Minot State
UNIVERSITY

Previous final question: the answer

What is a difference between regression and correlation?

- Correlation will show only the strength of relation (association) whereas regression will also show form of it (and may predict values)

Minot State
UNIVERSITY

Two-dimensional statistics
Regression

Thuesen data example

- 24 rows and 2 columns data for observations of ventricular velocity with different levels of blood glucose
- Data was taken from patients with diabetes type I.

Running the example and explaining results

```
> install.packages("ISwR")
> library(ISwR)
> str(thuesen); head(thuesen)
> thuesen <- na.omit(thuesen)
> thuesen.lm <- lm(short.velocity ~ blood.glucose,
+ data=thuesen)
> thuesen.lm
> summary(thuesen.lm)
```

Minot State
UNIVERSITY

Scatterplot with regression line

```
> plot(short.velocity ~ blood.glucose, data=thuesen)
> abline(thuesen.lm)
```

Minot State
UNIVERSITY

Visualizing residuals

```
> with(thuesen, segments(blood.glucose,
+ fitted(thuesen.lm), blood.glucose, short.velocity))
```

Confidence intervals for regression

```
> pred.frame <- data.frame(blood.glucose=4:20)
> pc <- predict(thuesen.lm, int="c", newdata=pred.frame)
> plot(short.velocity ~ blood.glucose, data=thuesen)
> pred.gluc <- pred.frame$blood.glucose
> matlines(pred.gluc, pc, lty=c(1,2,2), col="black")
```

Minot State
UNIVERSITY

Diagnostic plots for regression

- "Residuals vs. Fitted": checks outliers
- "Normal Q-Q": checks residuals for normal distribution, if they are not normal then our regression is not linear
- "Scale-Location": checks the trend in dispersion
- "Residuals vs. Leverage & Cook's distance": checks the most influential observations

Minot State
UNIVERSITY

Regression diagnostics

```
> plot(thuesen.lm)
> plot(lm(height ~ weight, data=women))
```

Minot State
UNIVERSITY

Two-dimensional statistics
Analysis of covariation

Minot State University

Analysis of covariation (ANCOVA)

- ANCOVA integrates several regression lines together and checks the full model
- Model formula is
  ```
  response ~ influence * factor
  ```
- The ANCOVA will check if there is any difference between intersection and slope of the first line and intersections and slopes of all other lines (each line corresponds with one factor level)

Minot State
UNIVERSITY

Final question (3 points)

Final question (3 points)

What is R-squared?

Finishing...

```
> savehistory("20120330.r")
```

Summary: most important commands

- `lm()`—estimate the linear regression
- `predict()`—predict values with model

Minot State
UNIVERSITY

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.
Chapter 6.

Minot State
UNIVERSITY