

Biometry. Lecture 17

Alexey Shipunov

Minot State University

April 7, 2014



- 1 Questions and answers
- 2 Two-dimensional statistics
 - Regression
 - Analysis of covariation



- 1 Questions and answers
- 2 Two-dimensional statistics
 - Regression
 - Analysis of covariation



Starting...

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



Previous final question: the answer

In the embedded data `USArrests`, there are numbers of murders and rapes per 100,000 for every state.

Are murders and rapes correlated? Is that correlation significant?

```
> with(USArrests, cor(Murder, Rape))  
[1] 0.5635788  
> with(USArrests, cor.test(Murder, Rape))$p.value  
[1] 0.00002030826
```



Two-dimensional statistics

Regression



Analysis of regression model

- If y is a real response, then error of model

$$E = y - m$$

- If σ^2 are dispersions of m and y , then

$$R^2 = 1 - \sigma_m^2 / \sigma_y^2,$$

- In a background, R^2 is similar to coefficient of determination



Running the example and explaining results

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



Fitting the model

```
> plot(thuesen.lm)
> ... # so what to do?
```



One more model

- We will try to understand the relation between gross state product (GDP) and rate of murders



One more model, part I

```
> gdp2010 <- read.table("
+ http://ashipunov.info/data/gdp2010.txt", h=T, sep="\t")
> mg <- data.frame(murder=USArrests$Murder,
+ gdp=gdp2010$GDP)
> mg.lm <- lm(murder ~ gdp, data=mg)
> summary(mg.lm)
> plot(mg.lm) # NOT a plot of model: 4 diagnostic plots
> plot(murder ~ gdp, data=mg) # Plot of model
> abline(mg.lm)
```

If 1st and 3rd quartiles are too far from median, residuals could be over-dispersed.

In model formula, dependent variable is always a first one, independent variable(s) are after tilde.



One more model, part II

```
> new.points <- seq(min(mg$gdp), max(mg$gdp),  
+ length.out=50)  
> new.frame <- data.frame(gdp=new.points)  
> predicted.points <- predict(mg.lm, int="c",  
+ newdata=new.frame)  
> matlines(new.points, predicted.points)  
# If you like to see names of states  
> identify(mg$gdp, mg$murder, labels=row.names(gdp2010))  
# Do not forget to click the right mouse button  
# The other way to see names of states:  
> plot(murder ~ gdp, data=mg, type="n")  
> text(y=mg$murder, x=mg$gdp, labels=row.names(gdp2010))
```



Two-dimensional statistics

Analysis of covariation



Analysis of covariation (ANCOVA)

- ANCOVA integrates several regression lines together and checks the full model
- Model formula is
$$\text{response} \sim \text{influence} * \text{factor}$$
- 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)



Grazing data

- 40 plants were treated in two groups: with grazing (in first two weeks) and without grazing
- Rootstock diameter was also measured
- At the end of season, fruit production was measured (dry weight in mg)



Visualization first

```
> ipo <- read.table(  
+ "http://ashipunov.info/data/ipomopsis.txt", h=T)  
> head(ipo)  
> with(ipo, plot(Root, Fruit,  
+ pch=as.numeric(Grazing)))  
> abline(lm(Fruit ~ Root, data=subset(ipo,  
+ Grazing=="Grazed")))  
> abline(lm(Fruit ~ Root, data=subset(ipo,  
+ Grazing=="Ungrazed")), lty=2)  
> legend("topleft", lty=1:2,  
+ legend=c("Grazed", "Ungrazed"))
```



Model output

```
> ipo.lm <- lm(Fruit ~ Root * Grazing, data=ipo)
> summary(ipo.lm)
```

Two equations:

$$\text{Fruit} = -125.174 + 23.24 * \text{Root} \text{ (for grazed)}$$

$$\text{Fruit} = (-125.174 + 30.806) + (23.24 + 0.756) * \text{Root} \text{ (for ungrazed)}$$



Finishing...

```
> savehistory("20140404.r")
```



Final question (2 points)



Final question (2 points)

What is ANCOVA?



Summary: most important commands

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



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.]

