Biometry. Lecture 21

Alexey Shipunov

Minot State University

April 4, 2012

Outline

Minot State UNIVERSITY

Outline

Starting...

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

or

“Change dir”

in menu!

Previous final question: the answer

What is R-squared?

Previous final question: the answer

What is R-squared?

- The strength of relation between terms in a model, it is close to the squared correlation coefficient

Two-dimensional statistics

Linear regression

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

```
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)

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:

`Fruit = -125.174 + 23.24 * Root` (for grazed)

`Fruit = (-125.174 + 30.806) + (23.24 + 0.756) * Root` (for ungrazed)

Finishing...

```
> savehistory("20120404.r")
```

Final question (3 points)

Final question (3 points)

What is ANCOVA?

Summary: most important commands

- `lm()`—estimates the linear regression model and many other models (like ANCOVA)

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.
Sections 12.7.