Biometry. Lecture 20

Alexey Shipunov

Minot State University

April 20, 2016

Outline

1. **2D statistics**
   - Anatomy of relation: regression
   - Analysis of covariation

> 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)

```
> dact <- read.table("http://ashipunov.info/data/dact2.txt", h=T)
> (d2 <- table(dact$LEAF.SP, dact$LIP.COL))
> source("http://ashipunov.info/r/pleiad.r")
> VTcoeffs(d2)
```

2D statistics
Anatomy of relation: regression

Diagnostic plots again: good Residuals vs. Fitted

```
> oldpar <- par(mfrow = c (3, 3))
> set.seed(101)
> for (i in 1:9) plot(1:50, rnorm(50), xlab="Fitted",
+ ylab = "Residuals")
```

Dispersion of residuals is homogeneous.

Diagnostic plots again: bad Residuals vs. Fitted

```
> for (i in 1:9) plot(1:50, (1:50)*rnorm(50), xlab="Fitted",
+ ylab = "Residuals")
> par(oldpar)
```

Dispersion of residuals is **not** homogeneous.

Women model again: adding squared term

```
> women.lm1 <- lm(weight ~ height, data=women)
> women.lm2 <- lm(weight ~ height + I(height^2), data=women)
> summary(women.lm2)
> plot(women.lm1, which=1)
> plot(women.lm2, which=1) # much better!
> source("http://ashipunov.info/r/asmisc.r")
> Normality(women.lm2$residuals) # normal!
> AIC(women.lm1)
> AIC(women.lm2)
```

The Akaike information criterion (AIC) reflects the entropy of the model and therefore is a measure of the model relative quality.

2D 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)

M

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)

ANCOVA model tuning

```
> ipo.lm2 <- update(ipo.lm, . ~ . - Root:Grazing)
> summary(ipo.lm2)
> ipo.lm3 <- lm(Fruit ~ Root + Grazing, data=ipo)
> summary(ipo.lm3) # same as ipo.lm2: additive model
> AIC(ipo.lm)
> AIC(ipo.lm2)
> # Let us return to women data
> AIC(women.lm1)
> AIC(women.lm2) # better!
```

" `. - something`" means "everything in the model is the same except something which has been taken out"

M

Analysis of covariation, example II

Islands of two types: islet-like and stone-like

```
> it <- read.table("http://ashipunov.info/data/it.txt",
+ h=T, sep="\t")
> str(it)
> it$SQ <- log10(it$SQ)
> plot(SP ~ SQ, data=it, type="n")
> text(it$SQ, it$SP, labels=abbreviate(it$TYPE, 1))
> abline(lm(SP ~ SQ, data=subset(it, TYPE=="islet-like")))
> abline(lm(SP ~ SQ, data=subset(it, TYPE=="stone-like")),
+ lty=2)
```

Analysis of covariation, example II

```
> it.ancova <- lm(SP ~ SQ * TYPE, data=it)
> summary(it.ancova)
> it.ancova2 <- update(it.ancova, ~ . - SQ:TYPE)
> summary(it.ancova2)
> AIC(it.ancova)
> AIC(it.ancova2) # better!
> summary(lm(SP ~ SQ + TYPE, data=it)) # like second
```

Interceptions are different but slopes are the same. In statistical language, we may say that in this case, additive model is better. Square and type are two independent terms.

Finishing...

Save your commands!

(savehistory(<todaysdate>.r) or File -> Save as... on Mac)

Summary: most important commands

- `update()` – changethe existing model
- `influence * factor` – model with interactions
- `influence + factor` – additive model

For Further Reading

A. Shipunov.
Biometry [Electronic resource].
2012—onwards.
Mode of access:
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