Biometry. Lecture 14

Alexey Shipunov

Minot State University

February 29, 2012

Outline

Minot State UNIVERSITY

Outline

Minot State UNIVERSITY

Outline

Minot State UNIVERSITY

Starting...

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

or

“Change dir”

in menu!

Previous final question: the answer

Why we need the "`for`" construction in R?

Previous final question: the answer

Why we need the "`for`" construction in R?

- To repeat anything

One-dimensional statistics

Functions

Normality function, version 3

```
> Normality3 <- function(df, p=.05)
+ {
+ sapply(df, function(.x)
+ ifelse(shapiro.test(.x)$p.value > p,
+ "NORMAL", "NOT NORMAL"))
+ }
> Normality3(trees)
> Normality3(log(trees+1))
```

Features: shorter, faster, vectorized and applicable to modified data

One-dimensional statistics

Tests for proportions

Why we need to test proportions

- Proportions are secondary data
- The main question is: how well the proportion calculated from sample represents the population proportion?
- Null is that proportion of sample does not differ significantly from population proportion

Smokers and non-smokers example

- In hospital, among lung cancer patients, 356 from 476 are smokers ($\approx 75\%$)
- However, among all patients this proportion is lightly lower.
- How well our sample (lung cancer group) represents the whole hospital? In other words, is the deviation we see accidental?

Exact binomial test

```
> binom.test(x=356, n=476, p=0.7, alternative="two.sided")
```

`"two.sided"` means that the deviation may be to the both possible sides. It was possible to write `"greater"` instead; in this case we would test if the proportion in our sample is bigger. One-sided tests are normally more powerful but you should **newer** use two and one-sided tests together (this is not far from falsification of results)!

Proportion test

Proportion tests are more universal than binomial, but return very similar results:

```
> prop.test(x=356, n=476, p=0.7, alternative="two.sided")
```

Two-dimensional statistics

Hypotheses and tests

Hypotheses are basics of science

- The inferential science is based on hypotheses construction and calculation of their probability.
- The simplest approach is to establish null and alternative hypotheses.

Statistical errors

- Type I error is a false alarm: we accept alternative when null is true
- Type II error is a carelessness: we accept null when alternative is true

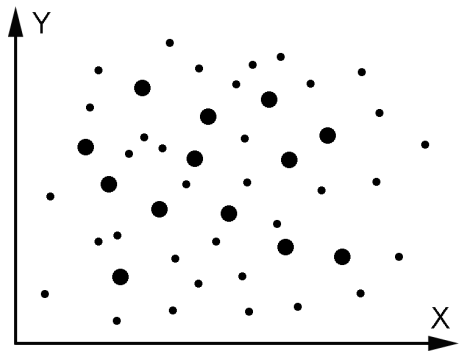
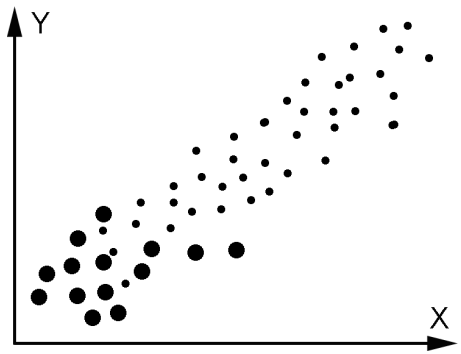
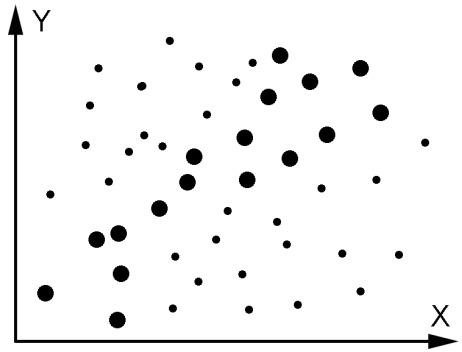
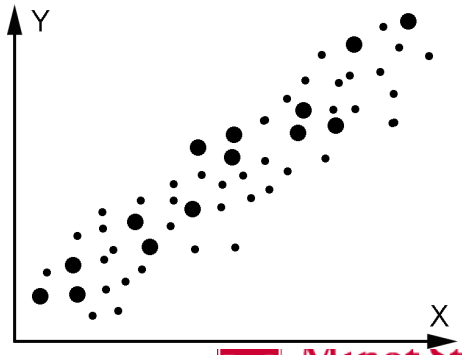
Level of significance

- The probability to make Type I error is a p-value
- We may ignore this probability if it is too low, in other words, below the level of significance
- The level of significance is a matter of experience and agreement, it could be 0.05, but sometimes also 0.1 and 0.01

Power

- Probability NOT to make a Type II error is a *power*
- The significance level for the power is normally around 0.8, tests with lesser power should be considered as weak

Type I and II errors for two variables

Sample \ Population	Null true	Alternative true
Accept null		
Accept alternative		

Finishing...

```
> savehistory("20120229.r")
```

Final question (10 points!)

Final question (10 points!)

In the exit poll, 262 persons were questioned. 136 ($\approx$ 53%) said they voted for the candidate A.

Use proportion test to check if candidate A won. Test the alternative that candidate A result was greater than 50% (0.5). Report the p-value and your conclusion.

Summary: most important commands

- !!!
- !!!

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.
Chapters 4, 5:1–2.

Minot State UNIVERSITY