Biometry. Lecture 15

Alexey Shipunov

Minot State University

March 7, 2012

Outline

Outline

Starting...

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

or

“Change dir”

in menu!

Exam 2: key points, problems and solutions

Optimized Exam 2 script (without saving)

```
> leaves <- read.table(
+ "http://ashipunov.info/data/leaves.txt", h=T)
> sapply(leaves, median)
> sapply(leaves, IQR)
> boxplot(leaves)
> plot(leaves$CONTROL.2, leaves$CONTROL.1)
> lines(loess.smooth(leaves$CONTROL.2, leaves$CONTROL.1))
> plot(leaves$EXP.2, leaves$EXP.1)
> lines(loess.smooth(leaves$EXP.2, leaves$EXP.1))
```

Previous final question: the answer

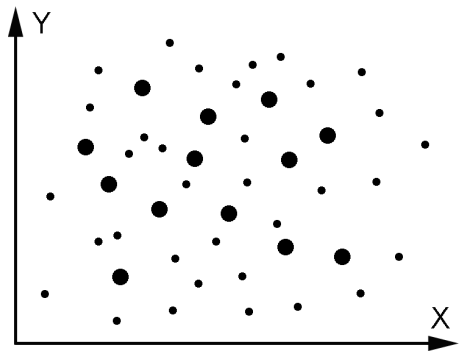
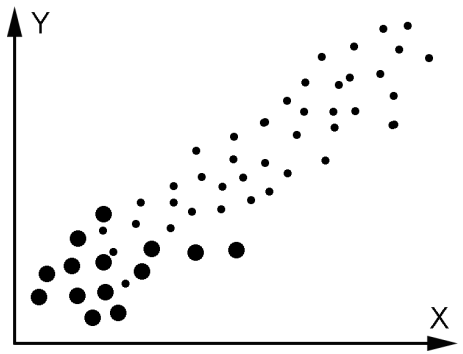
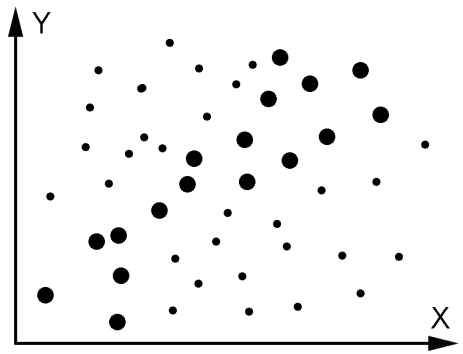
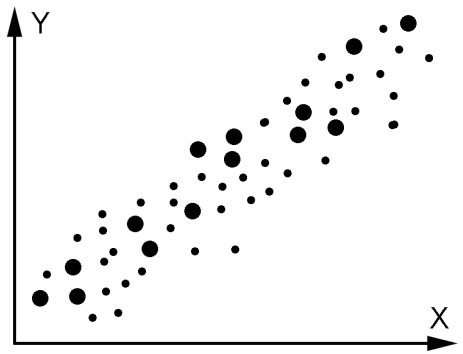
In the exit poll, 262 persons were questioned. 136 ($\approx$ 53%) said they voted for the candidate A. Use proportion test to check if candidate A won. Test the alternative that candidate A result was greater than 50% (0.5). Report the p-value and your conclusion.

```
> prop.test(x=136, n=262, p=.5, alt="greater")

1-sample proportions test with continuity correction

data:  136 out of 262, null probability 0.5
X-squared = 0.3092, df = 1, p-value = 0.2891
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.4664802 1.0000000
sample estimates:
       p
0.519084
```

Two-dimensional statistics

Tests for the independence of two variables

Type I and II errors for two variables

Sample \ Population	Null true	Alternative true
Accept null		
Accept alternative		

What is tested?

- Null: difference equal to 0 $\approx$ similar $\approx$ related $\approx$ samples came from same population
- Alternative: difference not equal to 0 $\approx$ different $\approx$ non-related $\approx$ samples came from different populations

Tests are based on central values

```
> a <- 51:59
> b <- 1:9
> x <- rep(5, 9)
> t.test(a, b)
> t.test(b, x)
```

Homoscedasticity, similarity of variance (like in a and b but not like in b an x) is an important assumption of all two variable tests.

Paired and non-paired

- Paired: came from one set of objects (e.g., measurements done at different time)
- Non-paired: do not belong to one set of objects

Tests are based on central values

```
> set.seed(1);  t.test(a, (a+rnorm(9)), paired=T)
```

We introduced here a random noise (`rnorm()` function)

Parametric and non-parametric

- Parametric: Student's, or t-test (in R, with Welch correction)
- Non-parametric: Wilcoxon tests

Leaves example

```
> leaves <- read.table(
+ "http://ashipunov.info/data/leaves.txt", h=T)
> Normality3 <- function(df, p=.05)
+ {
+ sapply(df, function(.x)
+ ifelse(shapiro.test(.x)$p.value > p,
+ "NORMAL", "NOT NORMAL"))
+ }
> Normality3(leaves) # all normal!
> t.test(leaves[,1], leaves[,2], paired=T)
> wilcox.test(leaves[,1], leaves[,2], paired=T)
> t.test(leaves[,1], leaves[,3])
> wilcox.test(leaves[,1], leaves[,3])
```

Minot State UNIVERSITY

Finishing...

```
> savehistory("20120307.r")
```

Final question (3 points)

Final question (3 points)

How to decide, which two-sample test is preferable, t-test or Wilcoxon test?

Summary: most important commands

- `t.test()`—paired and non-paired two-sample parametric test
- `wilcox.test()`—paired and non-paired two-sample non-parametric test

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 5.

Minot State UNIVERSITY