

Biometry. Lecture 11

Alexey Shipunov

Minot State University

March 5, 2014



- 1 Questions and answers
- 2 One-dimensional data
 - Central tendency
 - One-dimensional tests



- 1 Questions and answers
- 2 One-dimensional data
 - Central tendency
 - One-dimensional tests



Starting...

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



Previous final question: the answer

What is a main practical difference between mean and median?



Previous final question: the answer

What is a main practical difference between mean and median?

- Median is more robust



Comparative boxplot

```

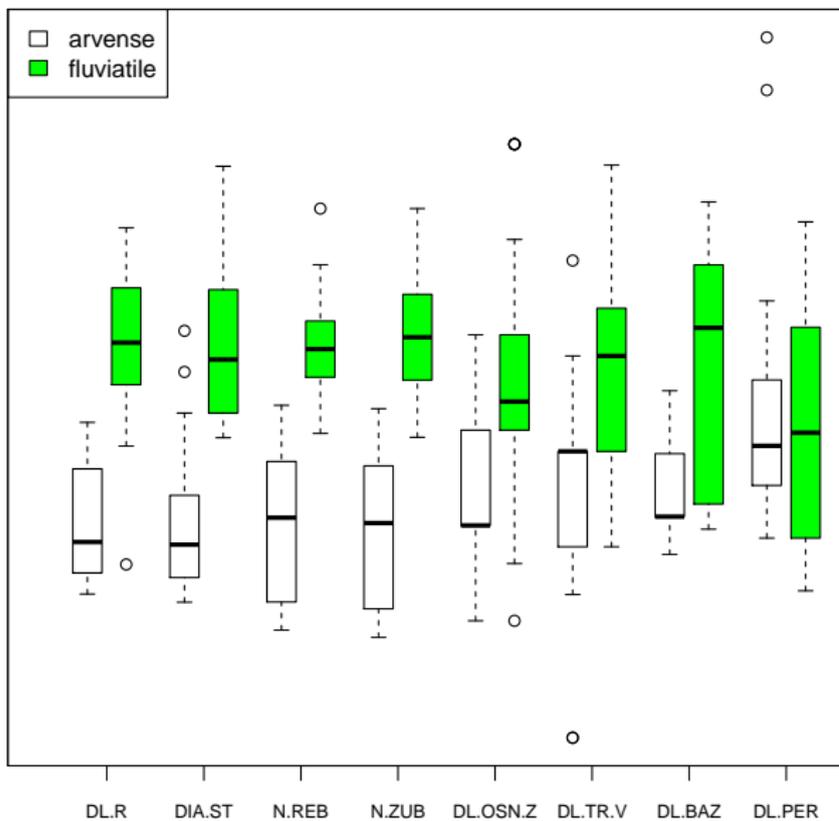
> eq <- read.table("http://ashipunov.info/data/eq.txt",
+ h=TRUE)
> seq <- scale(eq[,2:9]) # scale all except SPECIES
> boxplot(seq[eq$SPECIES=="arvensis",],
+ at=1:8-0.2, boxwex=0.3, # shift and narrow
+ xaxt="n", yaxt="n", # no labels
+ main="Horsetails, scaled and paired characters")
> boxplot(seq[eq$SPECIES=="fluviatile",],
+ at=1:8+0.2, boxwex=0.3,
+ xaxt="n", yaxt="n",
+ add=T, col="green") # overlay and colorize
> axis(1, 1:8, names(eq)[2:9], cex.axis=.75) # labels
> legend("topleft", c("arvensis","fluviatile"),
+ fill=c("white","green"))

```

This stuff is better to remember in separate *.r file



Horsetails, scaled and paired characters



One-dimensional data

Central tendency



summary()

`summary()` is a “smart” (generic) function which gives the most appropriate description of data. In many cases, it will give quantiles + mean:

```
> summary(salary)
> summary(trees)
> summary(sex)
```



summary() and “bad” data

summary() is very useful when one needs to check a reliability of data:

```
> err <- read.table("http://ashipunov.info/data/errors.txt",  
+ h=TRUE, sep="\t")  
> str(err)  
> summary(err)
```

AGE became a factor (erroneous “a”), empty name (instead of NA), and impossible minimal height.



One-dimensional data

One-dimensional tests



t-test and Wilcoxon test for one-dimensional data

- Statistical tests allow to check how well the general characteristic (central tendency or range) calculated from *sample* represents a *population*
- t-test (Student's) takes into account the normality of sample whereas Wilcoxon test do not consider the distribution, it is non-parametric
- Both give a *confidence interval*



t-test for one variable

```
> t.test(salary, mu=mean(salary))
One Sample t-test

data:  salary
t = 0, df = 6, p-value = 1
alternative hypothesis: true mean is not equal to 32.28571
95 percent confidence interval:
 3.468127 61.103302
sample estimates:
mean of x
32.28571
```



Wilcoxon test for one variable

```
> wilcox.test(salary, mu=median(salary), conf.int=TRUE)
```

```
Wilcoxon signed rank test with continuity correction
```

```
data: salary
```

```
V = 10, p-value = 0.5896
```

```
alternative hypothesis: true location is not equal to 21
```

```
80 percent confidence interval:
```

```
17.99999 63.50002
```

```
sample estimates:
```

```
(pseudo)median
```

```
24.99994
```

This will test median, not mean! Wilcoxon test is more universal but less traditional.



Understanding the test output: theory

- Alternative hypothesis (“something”) and null hypothesis (“nothing”)
- Type I error (false alarm), p-value (probability to issue the false alarm) and significance level (matter of agreement)



Understanding the test output: quick and dirty

- Which hypothesis is null?
- Does p-value less than 0.05?
 - 1 No: accept the null hypothesis—“sit and relax”
 - 2 Yes: reject the null hypothesis—“jump and call the police”



Finishing...

```
>savehistory("20140305.r")
```



Final question (3 points)



Final question (3 points)

Please explain the difference between null and alternative hypotheses.



Summary: most important commands

- `t.test()` — checks the reliability of mean (assuming that data distribution is normal)
- `wilcox.test()` — checks the reliability of median (non-parametric)



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. [Translating from Russian.]

