Biometry. Lecture 12

Alexey Shipunov

Minot State University

March 4, 2015

Outline

Starting...

`> 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)

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 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)
- In science, Type I error is more dangerous so we follow the “0.005 rule”: **if p-value is more than 0.05, we must accept the null hypothesis**

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”

How to understand which test to use? Normality.

- Normality tests will check if we can accept the normal distribution of our sample
- It is widely accepted that the strict normality testing is not generally required, it is enough, for example, to test normality graphically

Shapiro test for normality

```
> shapiro.test(salary) # What is a null hypothesis?!
> shapiro.test(rnorm(1000)) # Null is normality!
```

Quantile-quantile plot for normality

```
> qqnorm(salary); qqline(salary) # Bad!
> # Good:
> set.seed(1); qqnorm(rnorm(100)); qqline(rnorm(100))
```

`set.seed()` helps to maintain the same set of random numbers in the session.

One-dimensional data
Normality and R functions

- `shapiro.test()` is good but it is hard to apply if for data frames, and output is not very helpful.
- We will create the user function which run Shapiro-Wilks test with better output.

What is a function

```
> Sum <- function(a, b)
+ {
+ a + b
+ }
> Sum(2, 3)
```

Function is a piece of code which may run independently. All R commands are functions. Please note that every functions requires two parts: **arguments** in *round brackets* and **body** in *curly brackets*. It is too boring to enter functions line-by-line. Instead, it is better to copy function from external editor. If function contains mistake(s), one may use `fix()` command.

Let us create a simple useful function

R has no command for coefficient of variation, we will create it ourselves:

```
> CV <- function(x)
+ {
+ (sd(x) / mean(x)) * 100
+ }
> CV(trees[,3])
> CV(trees$Volume)
> sapply(trees, CV)
```

We can then run `fix(CV)` and add `round(..., 2)` function to make numbers more readable.

Finishing...

Save your commands!

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

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)
- `shapiro.test()`—Shapiro-Wilks test for normality
- `function(){}`—creates a function
- `for(){}`—the cycle
- `ifelse(check, yes, no)`—selector

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!
Ongoing translation from Russian.