Biometry. Lecture 13

Alexey Shipunov

Minot State University

March 21, 2016

Outline

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* which is even more important because their hypotheses-related output is typically not interesting.

t-test for one variable

```
> salary <- c(21, 19, 27, 11, 102, 25, 21)
> 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
```

As you see, p-value is high which is the strong argument to stay with a **null hypothesis**: **population mean is equal to sample mean** which is 32.28571.
Alternative hypothesis (**population mean is not equal to sample mean**) is not accepted.

Wilcoxon test for one variable

(At the moment, please ignore warnings)

```
> 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! Again, because p-value is much greater than 0.05, we must stay with our null hypothesis: population median is equal to sample median which is 21.
Alternative hypothesis (population median is not equal to sample median) is not accepted.

Understanding the test output: theory

This will be explained later, but it is probably better to look on it beforehand.

- Alternative hypothesis ("something") and null hypothesis ("nothing")
- Type I error (false alarm) is when nothing happens but you think that something happens. In other words, you accept alternative when null is true.
- Probability of the Type I error (probability to issue the false alarm), p-value, and significance level (matter of agreement) are related.
- More strictly, the **probability to have greater or equal effect (deviation) when null hypothesis is true (no deviation)** is a p-value.
- Type II error (carelessness) is when we accept null but alternative is true. Non-parametric tests are more careless (less powerful) then parametric so we always must check if parametric tests are applicable.
- In science, it is **agreed** that Type I error is more dangerous so we follow the "0.05 rule": **if p-value is more than 0.05, we must stay with the null hypothesis**
- But look here: `http://www.nature.com/news/statisticians-issue-warning-over-misuse-of-p-values-1.19503`. There is a current movement to replace p-values with other measures like: (1) confidence intervals, (2) bootstrap support, (3) effect sizes and (4) likelihood.

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 do something”

5% (0.05) criteria is accepted in science because “jumping” is much more costly so it is better to stay with null hypothesis until we 95% and more sure that we see something important.

How to understand which test to use? Normality.

- Normality tests will check if we can accept the normal distribution of our sample. If yes, we use parametric tests (like t-test), if not, we use non-parametric (like Wilcoxon test).
- It is widely accepted that the strict normality testing is not generally required, it is enough, for example, to test normality graphically

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.

Shapiro test for normality

```
> shapiro.test(salary) # What is a null hypothesis?!
> set.seed(1); shapiro.test(rnorm(1000)) # Null is normality!
```

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.

Normality() function for human use

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

One-dimensional data
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 ($\approx$ 70%, or 0.7).
- 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)!

Since the null hypothesis was that “true probability of success is equal to 0.7” and p-value was less than 0.05, we can reject it in favour to alternative hypothesis, “true probability of success is not equal to 0.7”. Consequently, proportion of smokers in our group looks different from their proportion in hospital.

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")
```

Voters example

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

```
> 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
Hypotheses and tests

Hypotheses are cornerstones of science

- The inferential science is based on hypotheses construction and calculation of their probability.
- The simplest approach is to establish null hypothesis and reject it if needed.
- More complicated approach is to consider null and alternative hypotheses together.

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 have greater or equal effect when null hypothesis is true is a p-value
- We may ignore this probability if it is too low, in other words, below the level of significance (“treshold”)
- The level of significance is a matter of experience and agreement, it could be 0.05, but sometimes also 0.1 and 0.01
- p-value is related with the probability of the Type I error

Two-dimensional statistics

Tests for the independence of two variables

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

Parametric and non-parametric

- Parametric: Student's, or t-test (in R, with Welch correction for **non-homogeneity of variance**)
- Non-parametric: Wilcoxon tests

Two sample tests for Keller's data

```
> ph <- read.table(
+ "http://ashipunov.info/data/phaseolus.txt", h=T)
> sapply(ph, Normality) # all normal!
> with(ph, t.test(EXPER.2 - EXPER.1, CONTROL.2 - CONTROL.1,
+ alt="less"))
```

We use "less" because we **already** have the preliminary hypothesis was that experimental leaves grow **slower**.

Two main questions

- Normal?
- Paired?

Final question (3 points)

Final question (3 points)

Please explain the difference between null and alternative hypotheses.

Finishing...

Save your commands!

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

Summary: most important commands

- `binom.test()` and `prop.test()`—tests for the equality of proportions
- `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_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