Part 4. Inferential statistics: t-test

Natalia Levshina © 2015

Part of the course taught at the University of Mainz, Germany
26-28 May 2015

Research questions for t-test

- Do language learners who are taught by an innovative method show better results than those who are taught traditionally?
- Do speakers of one language variety speak faster than speakers of another variety?
- Do people of one gender use more hedging constructions than people of another?

Only for variables on the ratio and interval scale!!!

Low- vs high-frequency nouns

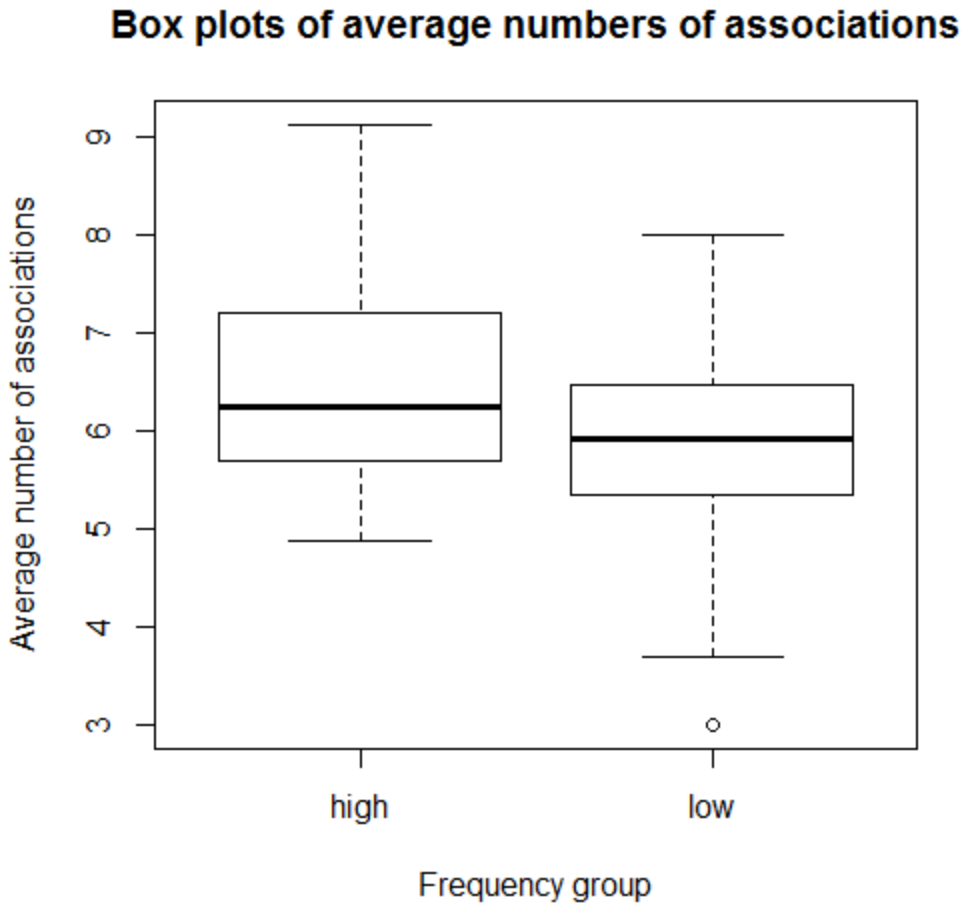
```
> data(pym_high) # high-frequency nouns
> data(pym_low) # low-frequency nouns
> head(pym_high) # first 6 observations
      syl let imag conc assoc
time    1   4 4.13 2.47  7.00
life    1   4 4.07 2.96  6.78
home    1   4 6.50 6.25  6.88
church  1   6 6.63 6.59  7.52
mind    1   4 3.03 2.60  5.88
door    1   4 6.60 7.00  7.96
```

Variable *assoc*

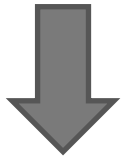
- Represents the average number of associations produced in 30 seconds, which are triggered by a noun
- Which nouns will trigger more associations, the ones with low or high frequencies?

Boxplot with two boxes

```
> boxplot(pym_high$assoc, pym_low$assoc, names = c("high", "low"), main = "Box plots of average numbers of associations", xlab = "Frequency group", ylab = "Average number of associations")
```


Box plots of average numbers of associations
Average number of associations
3
4
5
6
7
8
9
high
low
Frequency group

It seems that the high-frequency nouns trigger on average more associations, but is this difference statistically significant? That is, if we take another sample of high- and low-frequency nouns, will we observe the difference?



We need inferential statistics: t-test

A one-tailed test

```
> t.test(pym_high$assoc, pym_low$assoc, alternative = "greater")  # or alternative = “less”
Welch Two Sample t-test data:
pym_high$assoc and pym_low$assoc
t = 2.6717, df = 98.281, p-value = 0.004417
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
0.1977777       Inf
sample estimates:
mean of x mean of y
6.380000  5.857451
```

One-tailed vs. two-tailed tests

- One-tailed tests are used when your alternative hypothesis is directional:

  a) mean of group 1 is GREATER than mean of group 2)

  t.test(group1, group2, alternative = “greater”)

  b) mean of group 1 is LESS than mean of group 2

  t.test(group1, group2, alternative = “less”)

- Two-tailed tests are used if the alternative hypothesis is non-directional, e.g. mean of group 1 is DIFFERENT from mean of group 2

  t.test(group1, group2)

Interpretation of *p*-value

A *p*-value less than 0.05 means that the null hypothesis of no differences between the groups can be rejected. In other words, we believe that the difference between the high- and low-frequency nouns is not due to chance alone.

A *p*-value greater than 0.05 means that the null hypothesis of no differences between the groups CANNOT be rejected. In other words, we don't have enough evidence in favour of our alternaitive hypothesis.

In our case…

- *p*-value = 0.004417, which is much smaller than 0.05
- This means that the difference is statistically significant!

Exercise

Is there difference in the imagery scores between high- and low-frequency nouns? Compare the imagery scores of the high- and low-frequency nouns, available as the *imag* variable in the data frames pym_high and pym_low. Will you use a one- or two-tailed test?