

# Data Analysis in the Social Sciences

A Student's t-test in R

*Alla Tamboutseva*

First, load data on Titanic passengers we discussed before:

```
df <- read.csv("https://raw.githubusercontent.com/allatambov/Py-programming-3/master/28-05/Titanic.csv")
```

And delete rows with missing values:

```
df <- na.omit(df)
```

Now let us compare the average age of survived male and female passengers. To do so we need to select rows that correspond to survived passengers:

```
surv <- df[df$Survived == 1, ]
```

Now let's look at sample means by groups:

```
# select males and females
males <- surv[surv$Sex == "male", ]
females <- surv[surv$Sex == "female", ]
```

```
# calculate means
mean(males$Age)
```

```
## [1] 27.27602
```

```
mean(females$Age)
```

```
## [1] 28.84772
```

Judging by sample means we cannot decide whether the average age of males and females that survived is significantly different (since we look at samples, not populations). So, we have to test it formally. Let us proceed to a two sample Student's t-test.

$$H_0 : \mu_{female} = \mu_{male}$$

$$H_1 : \mu_{female} \neq \mu_{male}$$

```
# t-test for two means
# after ~ goes a grouping variable
t.test(surv$Age ~ surv$Sex)
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: surv$Age by surv$Sex
```

```
## t = 0.7909, df = 158.22, p-value = 0.4302
```

```
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## -2.353227 5.496616
```

```
## sample estimates:
```

```
## mean in group female mean in group male
```

```
## 28.84772 27.27602
```

Judging by this output we can conclude that the null hypothesis about the equality of two population means should not be rejected at the 5% level of significance. The mean age of survived males and survived females is not different.

We can consider a test with a one-sided alternative:

$$H_0 : \mu_{female} = \mu_{male}$$

$$H_1 : \mu_{female} > \mu_{male}$$

```
t.test(surv$Age ~ surv$Sex, alternative = "greater")
```

```
##  
## Welch Two Sample t-test  
##  
## data:  surv$Age by surv$Sex  
## t = 0.7909, df = 158.22, p-value = 0.4302  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -2.353227  5.496616  
## sample estimates:  
## mean in group female    mean in group male  
##           28.84772           27.27602
```

However, in our case, no difference in results is detected.