# E0410 Fundamentals of Statistics for Scientific Data Using R

by Daria Sapunova, PhD student, RECETOX

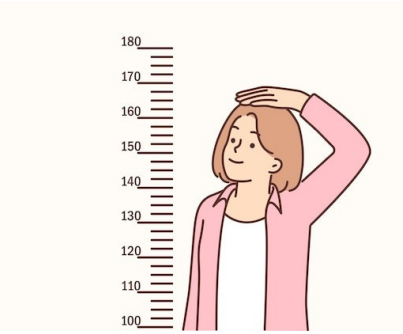
daria.sapunova@recetox.muni.cz

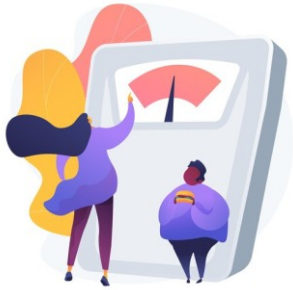
Bohunice, D29, room 123

# Repetition

# Association between multiple variables: quantification

## We have: multiple numerical and categorical variables



**Height Gender Smoker Weight**

**Salary Education Urbanization Productivity**

QUANTIFY association

QUANTIFY association

# Association between multiple variables: quantification

```
Call:
lm(formula = Weight ~ Height + Gender, data = women)

Residuals:
     Min       1Q   Median       3Q      Max
-19.8421  -3.2877   0.2537   3.6191  14.6720

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -65.4133    16.8769  -3.876 0.000193 ***
Height        0.8697     0.1116   7.792 7.39e-12 ***
GenderMale    2.1404     2.0246   1.057 7.39e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.839 on 97 degrees of freedom
Multiple R-squared:  0.599,     Adjusted R-squared:  0.5908
F-statistic: 72.46 on 2 and 97 DF,  p-value: < 2.2e-16
```

# Association between multiple variables: quantification

Height (cm) ↑ Gender ✗

**One-unit shift**

Weight (kg) ↑↓ ?

**How much will change?**

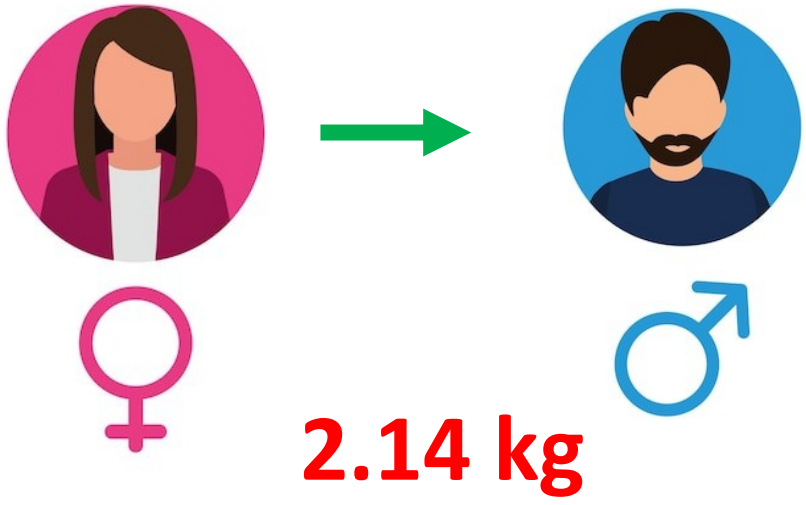

| | Estimate |
|---|---|
| (Intercept) | -65.4133 |
| Height | 0.8697 |
| GenderMale | 2.1404 |

Gender Male Female ✗ Height (cm)

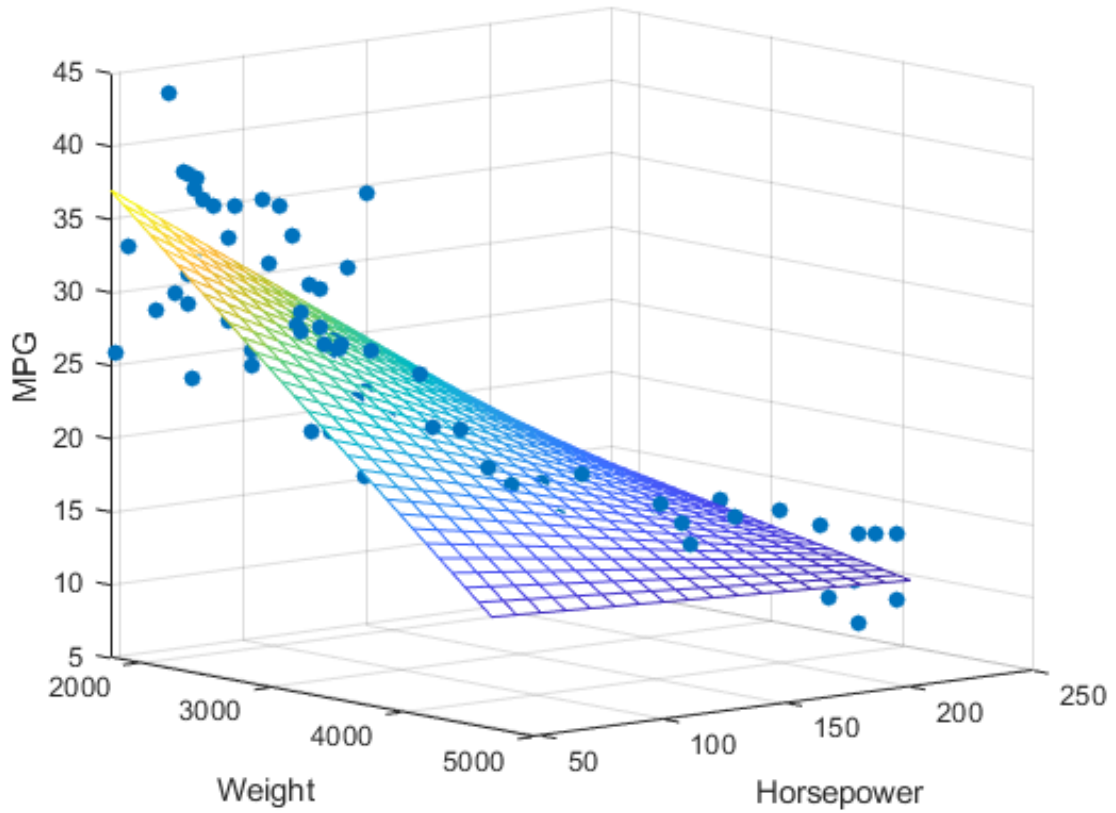
**One-unit shift**

Weight (kg) ↑↓ ?

**How much will change?**



# Multiple linear regression




The basic format of a multiple linear regression is:

$$Y = \alpha + \beta1*X1 + \beta2*X2$$

Where:

Y = outcome (i.e. dependent) variable.

X1 = first predictor (i.e. independent) variable.

X2 = second predictor (i.e. independent) variable.

α = intercept (average Y when X1=X2=0). Note: α is unit specific.

β1 = slope of the line (change in Y for a 1 unit increase in X1 when X2 is held constant). Note: β1 is unit specific.

β2 = slope of the line (change in Y for a 1 unit increase in X2 when X1 is held constant). Note: β2 is unit specific.

https://www.coursera.org/learn/linear-regression-r-public-health; https://towardsdatascience.com/graphs-and-ml-multiple-

# Association between multiple variables: quantification

**We have: multiple numerical and categorical variables**





**Fruit consumption; veggie consumption; fish consumption; meat consumption**
**Mercury level**

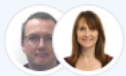
**Be aware: collinearity as an assumption in case of multiple numerical variables!**

Imperial College London

# Linear Regression in R for Public Health

This course is part of Statistical Analysis with R for Public Health Specialization

Taught in English | 22 languages available | Some content may not be translated



Instructors: Alex Bottle +1 more

Enroll for Free
Starts Apr 29

Financial aid available

**14,799** already enrolled

## Course

Gain insight into a topic and learn the fundamentals

4.8 ★ (494 reviews) | 97%

## Intermediate level

Recommended experience

## 14 hours (approximately)

## Flexible schedule

Learn at your own pace

## Prepare for a degree

Learn more

View course modules

# Practical part