# Simple linear regression (student version)

Dr. H.G.J. van Mil (revision by Dr. F.J. Rodenburg)

September 2021

Save this file and "Day8.Rdata" in the same location. Then, after opening this file, go to **Session** > **Set Working Directory** > **To Source File Location**.

Then run this:

```
load("Day8.Rdata")
```

## Q1

Below you find an output of a linear model:

```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5139 -0.6819  0.1292  0.9780  3.1159

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.8179     0.8709    1?   0.00459 **
x             1.9960       2?    14.031 3.92e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.694 on 18 degrees of freedom
Multiple R-squared:  0.9162,    Adjusted R-squared:  0.9116
F-statistic: 196.9 on 1 and 18 DF,  p-value: 3.919e-11
```

1. Estimate `y` for a value of $x = 3$.
2. Calculate the values of `1?` and `2?`.
3. How much of the variance in is explained by this model? Use one decimal place (xx.x%).

# Q2

It is often said (for example by the CDC) that you can approximate your maximum heart rate as:

$$\text{max. heart rate} = 220 - \text{age in years}$$

Below is some fictional data from a group of individuals, including their maximum heart rate and their age in years.

1. Fit a simple linear model to check whether we arrive at the same conclusion. The data are stored in `DF2`.
2. Does the relationship hold? What do you expect the intercept and slope to be if the rule of thumb is correct?
3. How much of the variance in maximum heart rate is explained by the age in years?
4. (hard) With `confint`, you can produce a confidence interval. Do the intercept and slope you expect from the formula differ significantly from those observed with $\alpha = 0.05$? How about at $\alpha = 0.01$ (HINT: `confint` takes on more than one argument).

# Q3

The lapse rate is the rate at which your body temperature drops as you increase elevation. A group of hikers were interested in checking empirically whether the commonly reported lapse rate of 9.8°C/km was accurate in their hike. To investigate, they grabbed their thermometers and wrist altimeters and found the following data on their hike. The data are stored in `DF3`.

1. Oh no! They didn't use SI-units. Convert the elevation from feet to kilometers and the temperature from Fahrenheit to Celsius.
2. Investigate whether the lapse rate is more or less equal to 9.8°C/km.
3. What is the difference between a confidence interval and a prediction interval?
4. (hard) Draw a plot with a confidence interval around the regression line. You will have to figure out by yourself how. How can you interpret this interval?