Regression

# Intoduction

This is about regression, we are about to perform regression analysis using marketing dta of the datarium package in r.

# Data

The marketing data has 200 observations and 4 variables namely: youtube, facebook, newspaper, and sales.

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.5   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## Loading required package: lattice  
##   
##   
## Attaching package: 'caret'  
##   
##   
## The following object is masked from 'package:purrr':  
##   
## lift

# Split the data into training and test set  
set.seed(123)  
training.samples <- marketing$sales %>%  
 createDataPartition(p = 0.8, list = FALSE)  
train.data <- marketing[training.samples, ]  
test.data <- marketing[-training.samples, ]

##   
## Call:  
## lm(formula = sales ~ ., data = train.data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.7142 -0.9939 0.3684 1.4494 3.3619   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.594142 0.420815 8.541 1.05e-14 \*\*\*  
## youtube 0.044636 0.001552 28.758 < 2e-16 \*\*\*  
## facebook 0.188823 0.009529 19.816 < 2e-16 \*\*\*  
## newspaper 0.002840 0.006442 0.441 0.66   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.043 on 158 degrees of freedom  
## Multiple R-squared: 0.8955, Adjusted R-squared: 0.8935   
## F-statistic: 451.2 on 3 and 158 DF, p-value: < 2.2e-16

## [1] 1.965508

## [1] 0.9049049

# Predicting with new data   
# using the model  
youtube <- 0  
facebook <- 90  
newspaper <- 0  
a <- data.frame(youtube, facebook, newspaper)  
p <- model %>%predict(a)  
p

## 1   
## 20.58823

Our model is as follows

where $is sales, and id youtube, is facebook, and is the error term.

model <- lm(sales ~ youtube, data = train.data)  
summary(model)$coef

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 8.58914961 0.616044182 13.94242 1.987874e-29  
## youtube 0.04671639 0.003003398 15.55451 8.019035e-34

```