Simple Regression Analysis

AW

2022/7/10

## Key terms

* Dependent vaiable
* Expanatory variable, independent variable, regressor
* Disturbance term
* Residual
* Residual sum of squares, **RSS**
* Ordinary least squuares, **OLS**
* Normal equations

## 1. The simple linear model

**The model**:

where is usually described as the **dependent variable**, and as the **explanatory variable** or **independent variable** or the **regressor**. In statistical analysis, one generally acknowledgesthe fact that the relationship is not exact by explicitly including in it a random factor known as the **disturbance term**, .

**The fitted model**:

The difference between the actual value of and the fitted value $, in observation is known as the **residual** in the observation . It will be denoted :

## 2. Derivation of the regression coefficients

We use the **least squares** criterion to choose amd so as to minimize **RSS**, the **residual sum of squares** (sum of the squares of the residuals).

This is usually referred to as **ordinary least squares** and abbreviated **OLS**.  
 The square of the residual in observaiont interms of and $, and the data on and :

(For formatting of the above equation array, see the website [Using R Markdown for Class Reports](https://www.stat.cmu.edu/~cshalizi/rmarkdown/) by Cosma Shalizi for more information.)

Summing over all the observations, we can write *RSS* as

We find the particular values of and by minimize *RSS*. The partial differentials of *RSS* with respect to and are:

The values of and that minimize *RSS* must satisfy the first-order conditions

Hence

These equations are known as the **normal equations** for the regression coefficients.

(Please see the website <https://tex.stackexchange.com/questions/288222/two-equations-in-one-line> for more information on how to format two equations on one line.)

## 3. An example

# Example in page 98 of the text  
eawe21 <- read.csv('./Data/EAWE21.csv', header=T) # Import data  
linear\_model <- lm(EARNINGS ~ S, data = eawe21) # Run the model  
summary(linear\_model) # The result

##   
## Call:  
## lm(formula = EARNINGS ~ S, data = eawe21)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -20.079 -6.726 -2.203 3.451 79.037   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.7647 2.8038 0.273 0.785   
## S 1.2657 0.1855 6.824 2.58e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11.36 on 498 degrees of freedom  
## Multiple R-squared: 0.08551, Adjusted R-squared: 0.08368   
## F-statistic: 46.57 on 1 and 498 DF, p-value: 2.579e-11

summary(eawe21$S)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 8.00 12.00 15.00 14.87 17.00 20.00

# Create the demeaned variable SDEV  
library(tidyverse) # Load the required package  
  
eawe21\_dm <- eawe21 %>%  
 mutate(SDEV = S - mean(S)) %>% # Create the demeaned variable  
 select(EARNINGS, S, SDEV) # Select the required columns  
  
# EARNINGS regress on the demeaned variable SDEV  
linear\_model\_dm <- lm(EARNINGS ~ SDEV, data = eawe21\_dm) # Run the model  
summary(linear\_model\_dm) # The result

##   
## Call:  
## lm(formula = EARNINGS ~ SDEV, data = eawe21\_dm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -20.079 -6.726 -2.203 3.451 79.037   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 19.5808 0.5082 38.528 < 2e-16 \*\*\*  
## SDEV 1.2657 0.1855 6.824 2.58e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11.36 on 498 degrees of freedom  
## Multiple R-squared: 0.08551, Adjusted R-squared: 0.08368   
## F-statistic: 46.57 on 1 and 498 DF, p-value: 2.579e-11