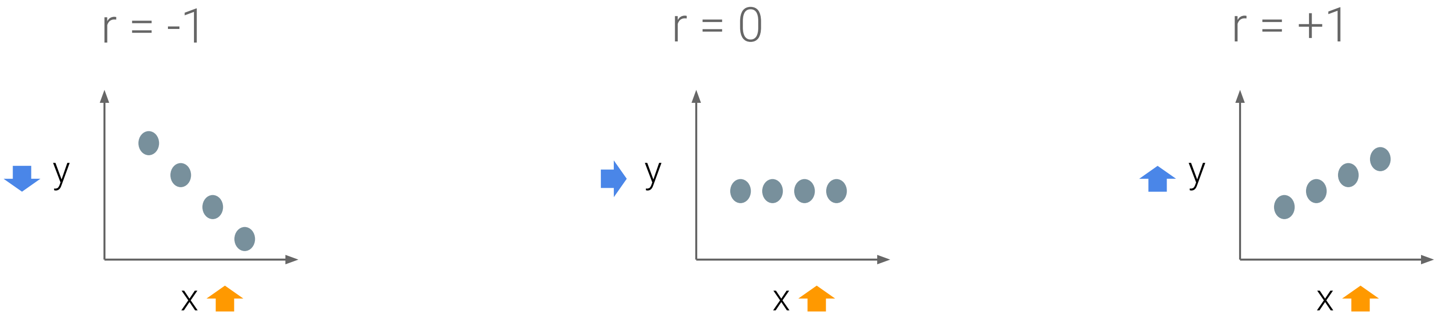
Simple linear regression

R^2: Describing the strength of a fit

Prof. Dr. Jan Kirenz

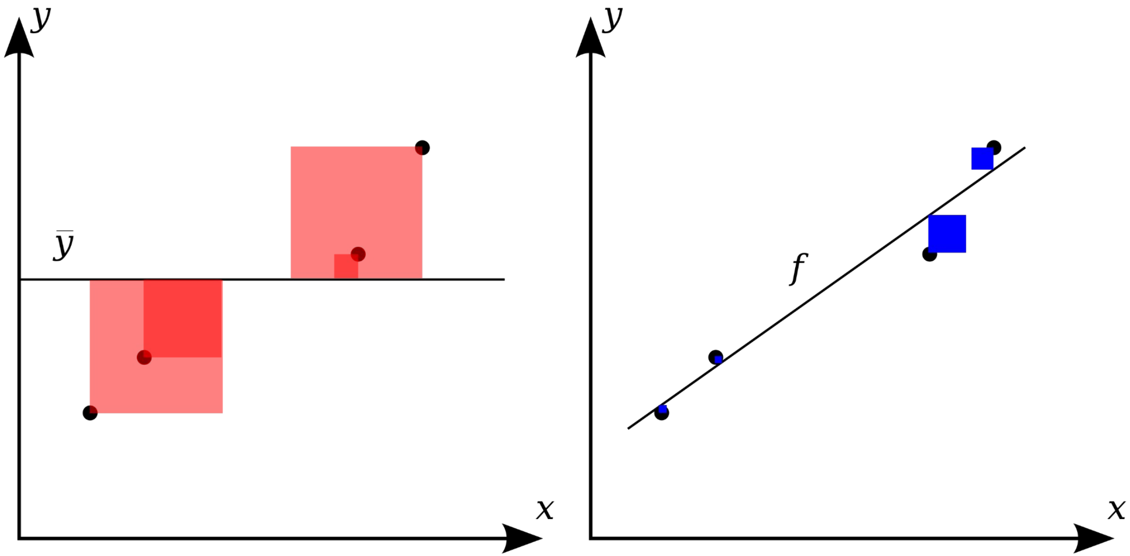
Coefficient of determination: $\mathbf{R^2}$ (R-squared)

Remember that **r** takes values between -1 and 1



- By **squaring** the value of **r** you get the proportion of variance in one variable shared by the other
- This is called the coefficient of determination or $\mathbf{R^2}$

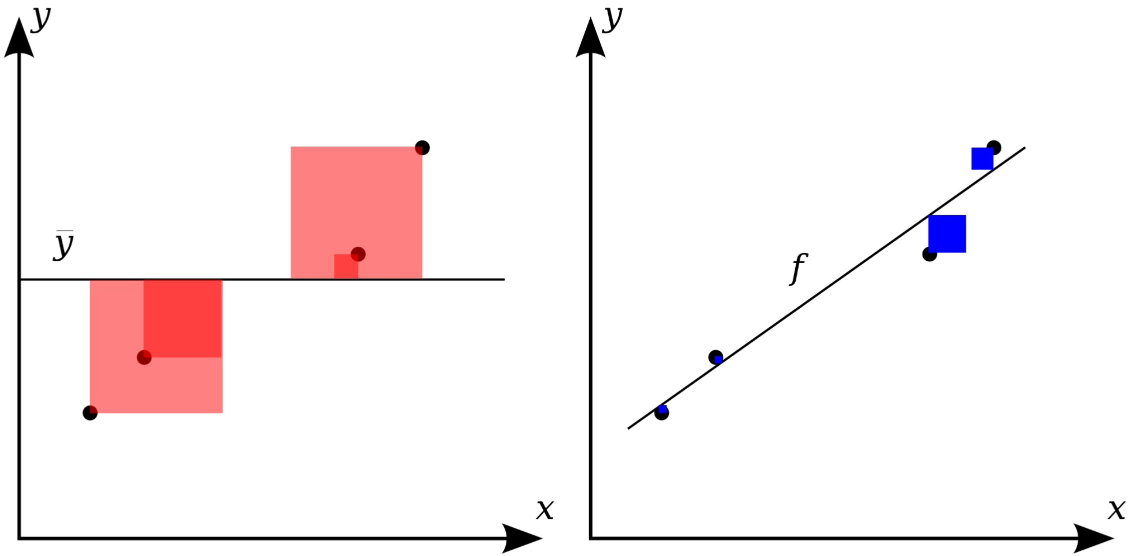
We explain the strength of a linear fit using R^2



Source: Orzetto

We explain the strength of a linear fit using R^2

$$R^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$



Source: Orzetto

We explain the strength of a linear fit using R^2

$$R^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$

SST: total sum of squares,

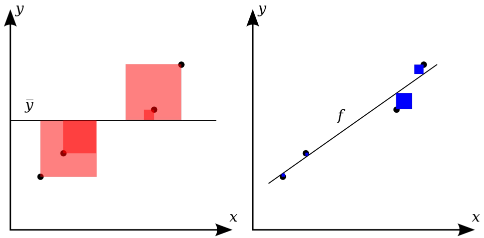
SSE: sum of squared errors

$$SST = (y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \cdots + (y_n - \bar{y})^2.$$

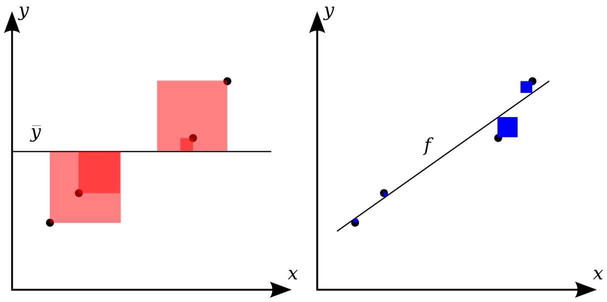
$$SSE = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \cdots + (y_n - \hat{y}_n)^2 \\ = e_1^2 + e_2^2 + \cdots + e_n^2$$

Describes the amount of variation in the outcome variable that is explained by the least squares line

R^2 will always be between 0 and 1.



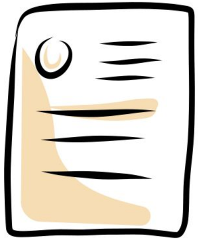
Alternative notation for R-squared



$$R^2 = 1 - \frac{\text{variability in residuals}}{\text{variability in the outcome}} = 1 - \frac{Var(e_i)}{Var(y_i)}$$

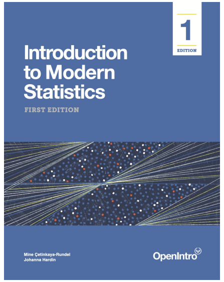
- **Var** = variance (s^2)
- $\mathbf{e_i}$ = residuals of the model for observation i
- $\mathbf{y_i}$ = outcome for observation i

To do: Calculate the R^2 in this Jupyter Notebook



Calculate the R^2 of your model

Resources



The content of this presentation is mainly based on the excellent book “Introduction to Modern Statistics” by Mine Çetinkaya-Rundel and Johanna Hardin (2021).

The online version of the book can be accessed for free:

https://openintro-ims.netlify.app/index.html