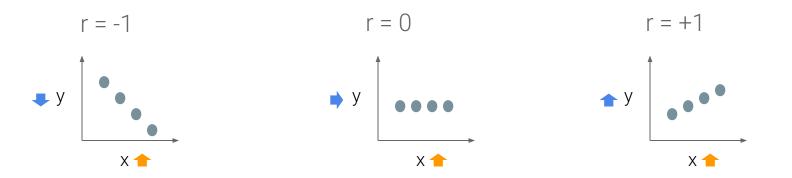
Simple linear regression

R²: Describing the strength of a fit

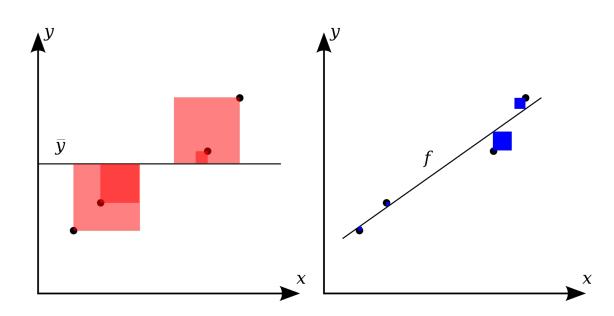
Coefficient of determination: R² (R-squared)

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 R²

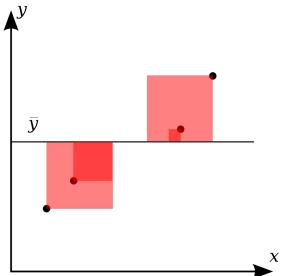
We explain the strength of a linear fit using R²

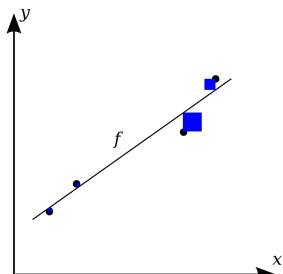


Source: Orzetto

We explain the strength of a linear fit using R²

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





Source: Orzetto

We explain the strength of a linear fit using R²

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

SST: total sum of squares,

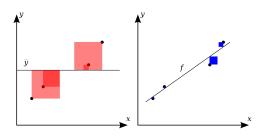
SSE: sum of squared errors

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

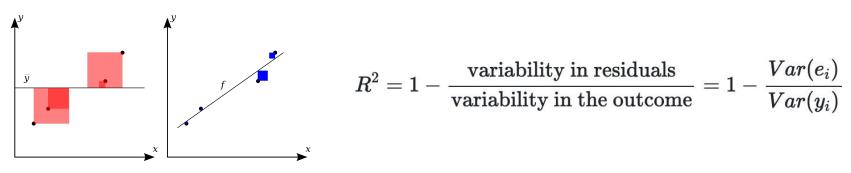
$$egin{align} SSE &= (y_1 - {\hat y}_1)^2 + (y_2 - {\hat y}_2)^2 + \dots + (y_n - {\hat y}_n)^2 \ &= e_1^2 + e_2^2 + \dots + e_n^2 \ \end{cases}$$

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



- **Var** = variance (s^2)
- **e**_i = residuals of the model for observation i
- **y**_i = outcome for observation i



To do: Calculate the R2 in this Jupyter Notebook





Calculate the R² 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