Coefficients: Correlation and Determination

DASC 512

Coefficient of Correlation (r)

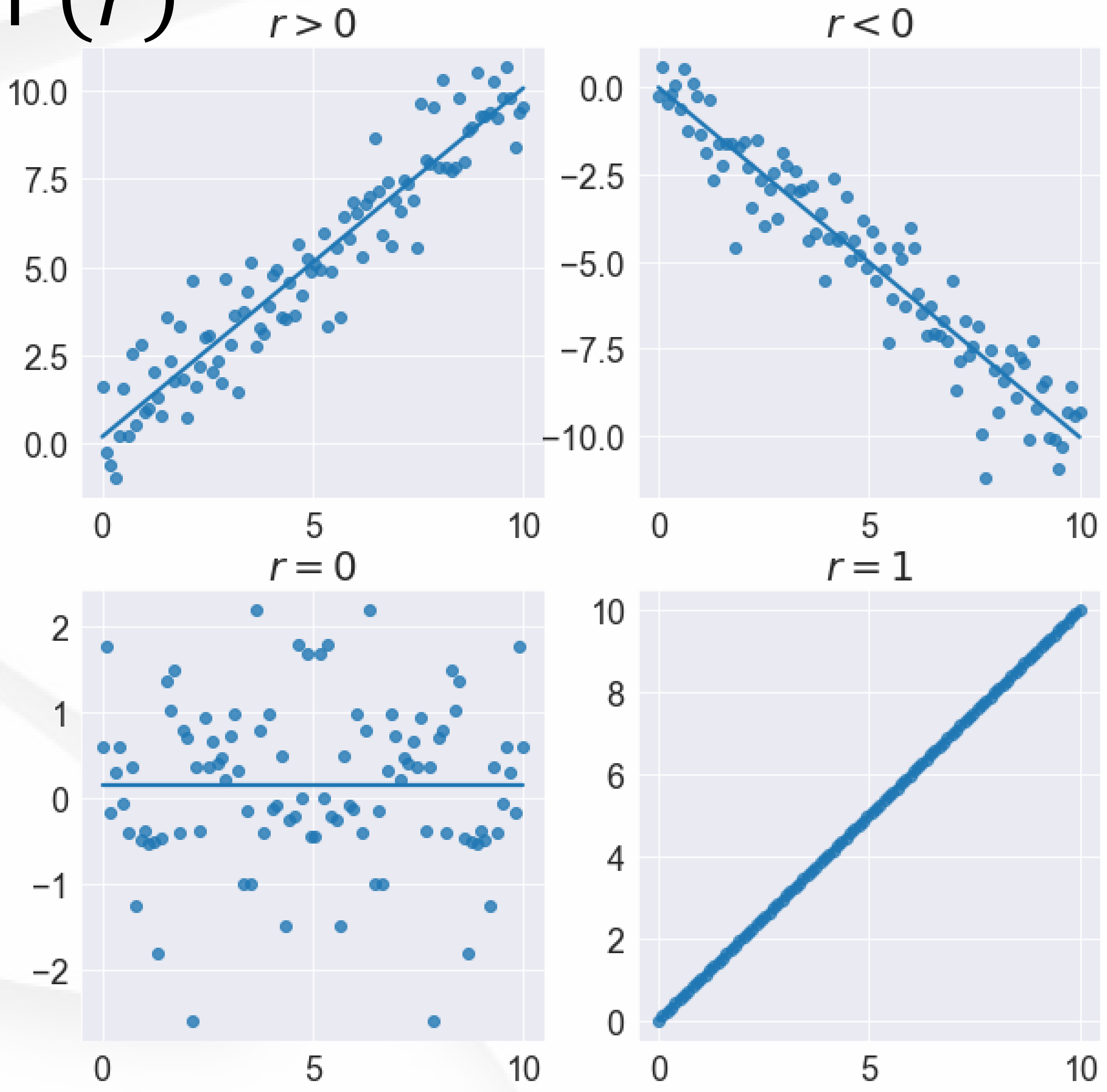
- The coefficient of correlation, r, is a measure of the strength of the **linear** relationship between two variables x and y.
- For a sample of n measurements, it is computed by:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} \times SS_{yy}}} = \sqrt{\frac{SS_{xx}}{SS_{yy}}}\,\hat{\beta}_1$$

Coefficient of Correlation (r)

$$r \in [-1,1]$$

- If $r > 0$, the model has a positive slope. The variables are positively correlated.
- If $r < 0$, the model has a negative slope. The variables are negatively correlated.
- If $r = 0$, the model has zero slope. The variables are not correlated.
- If $r = \pm 1$, the model is deterministic.



Hypothesis Test on r

We can perform a test if the population correlation coefficient is zero

$$H_0: r = 0, \qquad H_a: r < 0, r > 0, \text{or } r \neq 0$$

Our test statistic would be

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$$

Hypothesis Test on r

Our test statistic would be

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$$

This is the same test statistic and an equivalent test to last lesson's hypothesis test on whether $\beta_1 = 0$. Nothing new to learn here!

Coefficient of Determination

The <u>coefficient of determination</u>, r^2, is the proportion of the total variability that is explained by the linear relationship between x and y

$$r^2 = \frac{\text{Explained sample variability}}{\text{total sample variability}} = \frac{SS_{yy} - SSE}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}}$$

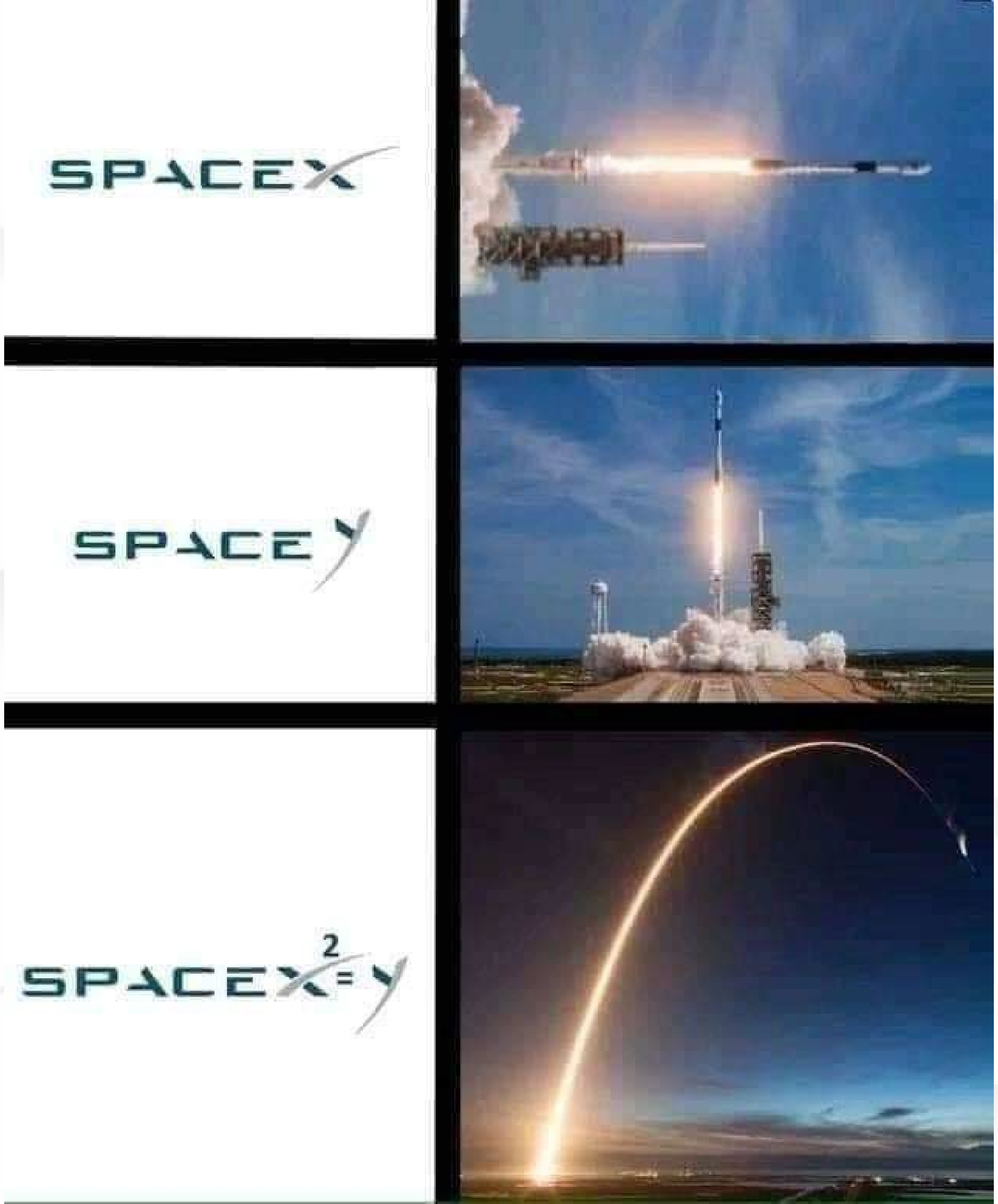
Coefficient of Determination

The coefficient of determination can be computed by squaring the coefficient of correlation

$$r^2 = r^2 = R^2$$

r takes a value between -1 and 1, so r^2 takes a value between 0 and 1

Practical interpretation: About $100 \times r^2$% of the sample variation in y can be explained by using x to predict y in the straight-line model


SPACEX
SPACEY
SPACEX2=Y


ROCKET AIN'T BEEN THE SAME
SINCE HE LOST GROOT.

Equations Recap (1 of 2)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$s^2 = MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n-2}$$

$$= \frac{SS_{yy} - \hat{\beta}_1 SS_{xy}}{n-2}$$

$$SS_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$SS_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$SS_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

Equations Recap (2 of 2)

$$s_{\hat{\beta}_1} = \frac{s}{\sqrt{SS_{xx}}}$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} \times SS_{yy}}} = \sqrt{\frac{SS_{xx}}{SS_{yy}}}\hat{\beta}_1$$

$$t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$$

$$r^2 = \frac{SS_{yy} - SSE}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}}$$

Next time…

Estimation Intervals

Prediction Intervals