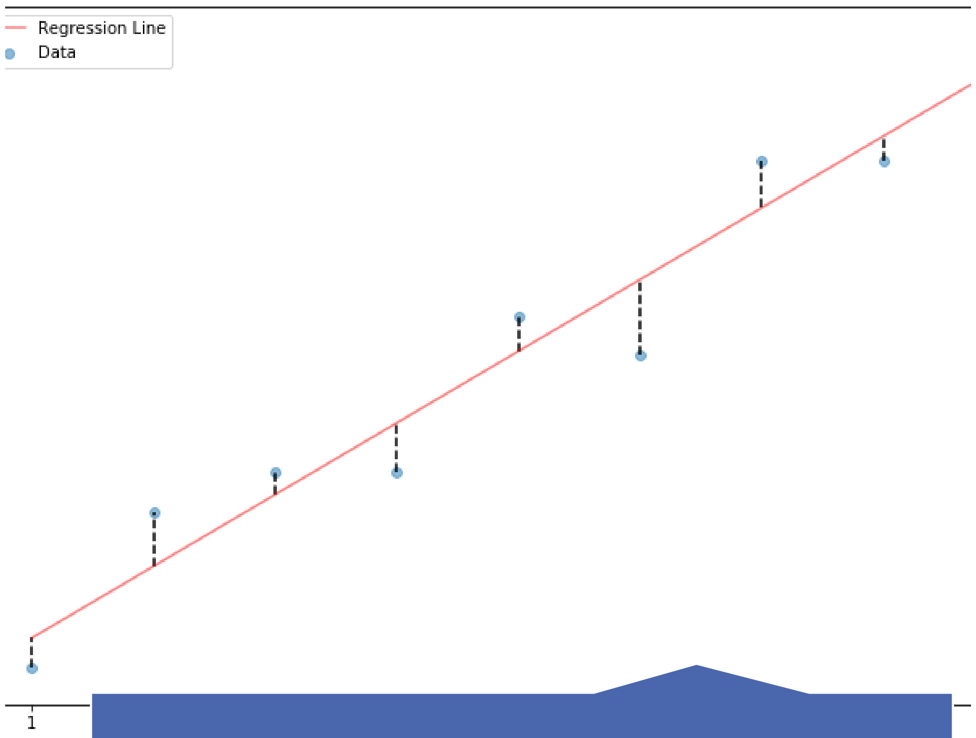
Regression Analysis – Covariance and Correlation



Covariance & Correlation

Do two random variables move together?

Covariance

Sample variance:

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

The covariance between two random variables measures the degree to which the two variables move together – it captures the linear relationship.

$$cov_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

n: sample size
X_i: ith observation on variable X
$\bar{X}$ mean of the variable X
Y_i: ith observation on variable Y
$\bar{Y}$ mean of the variable Y

Properties

- positive covariance: variables move together
- negative covariance: variables move in opposite directions
- covariance of variable with itself == variance

Covariance – Example and Pitfalls

X	1	2	3	4	5	6	7	8	9	mean
										5.00
y	2	6	7	7	11	10	15	15	18	10.11

$$cov_{XY} = \frac{(1-5.00)(2-10.11)+(2-5.00)(6-10.11)+ \ldots +(9-5.00)(18-10.11)}{9-1} = 13.75$$

Pitfalls

- actual value of covariance not meaningful
- can range from minus to plus infinity
- squared units

Correlation Coefficient

The correlation coefficient (r) measures the strength of the linear relationship (correlation) between two variables. It´s the standardized covariance and is easier to interpret as values are between -1 and +1.

$$r_{XY} = \frac{covariance\ of\ X\ and\ Y}{(sample\ standard\ deviation\ of\ X)(sample\ standard\ deviation\ of\ Y)}$$

Correlation Coefficient (r)	Interpretation
r = 1	perfect positive correlation
0 < r < 1	Positive linear relationship
r = 0	no linear relationship
-1 < r < 0	negative linear relationship
r = -1	perfect negative correlation

Correlation Coefficient - Example

X	1	2	3	4	5	6	7	8	9	sample std	cov
y	2	6	7	7	11	10	15	15	18		

	sample std	cov
X	2.74	13.75
y	5.16	

$$r_{XY} = \frac{13.75}{2.74 * 5.16} = 0.973$$

→ almost perfect positive correlation