Filter Methods

31 May 2023 08:36 AM

pearson ⟶ 'r' value

-1 to +1

Good predictor 0.7 to 1 and -1 to -0.7

Bad predictor -0.3 to +0.3

$$r = \frac{cov(x,y)}{std(x).std(y)}$$

pearson correlation coeff.

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

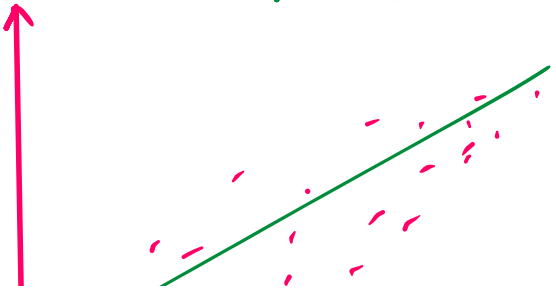
$$\text{spearman corr. coeff.} = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

n = no. of rows

d^2 = diff betn rank.

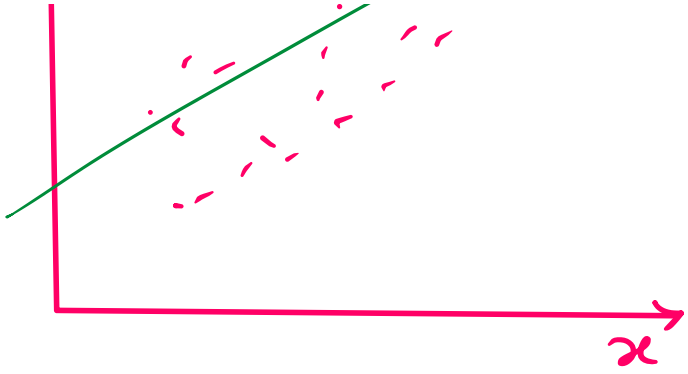
when to use spearman and pearson

case-1 — Linear Data

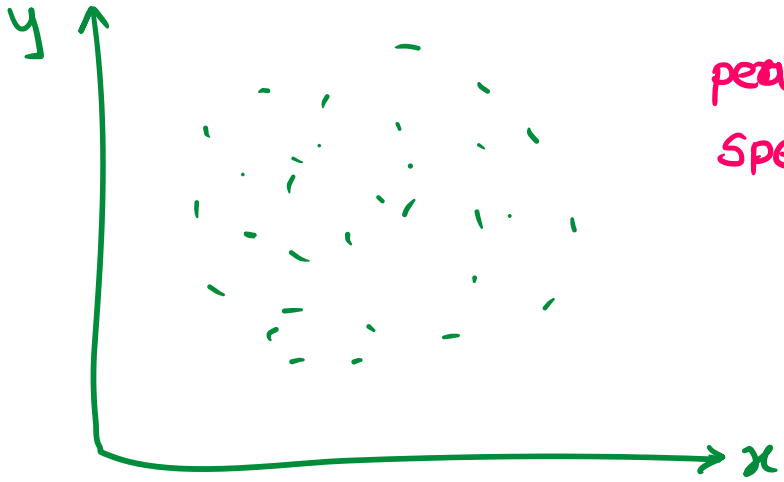


✓ pearson ≥ 0.9

✓ Spearman > 0.9



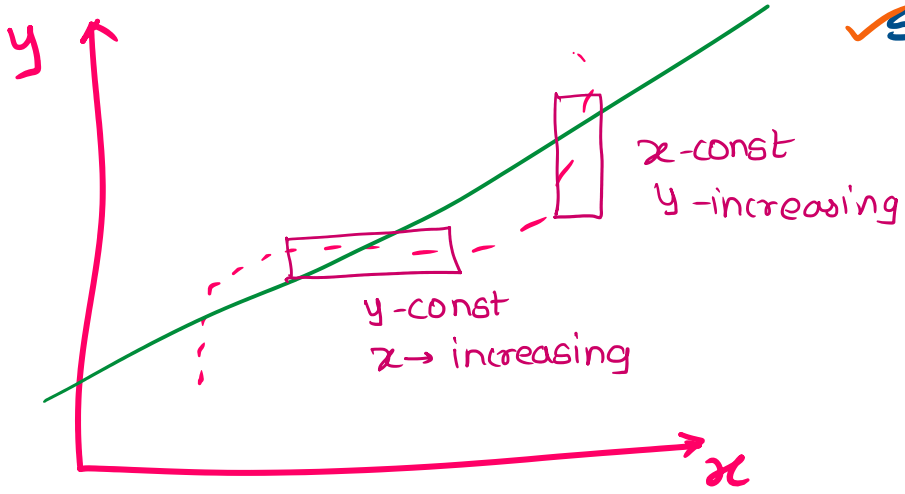
Case-2 Non Linear Data



pearson ≤ 0.3

Spearman ≤ 0.3

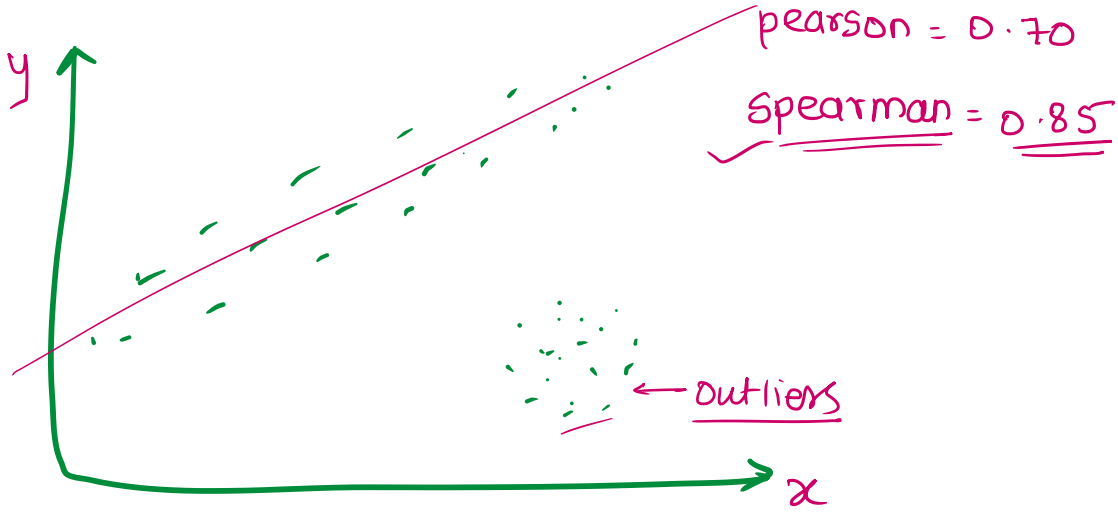
Case-3 Sine wave



pearson = 0.70

✓Spearman = 0.90 (Good)

Case-4 - When Outliers in Data


pearson = 0.70
Spearman = 0.85
y
outliers
x