



Pearson Correlation Coefficient



Pearson Correlation Coefficient

- In order to normalize values coming from two different distributions, we use:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\frac{1}{n} \sum (x - \bar{x})(y - \bar{y})}{\sqrt{\frac{\sum (x - \bar{x})^2}{n}} \sqrt{\frac{\sum (y - \bar{y})^2}{n}}} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}$$

ρ = Greek letter "rho"

cov = covariance

σ = standard deviation

\bar{x} = mean of X



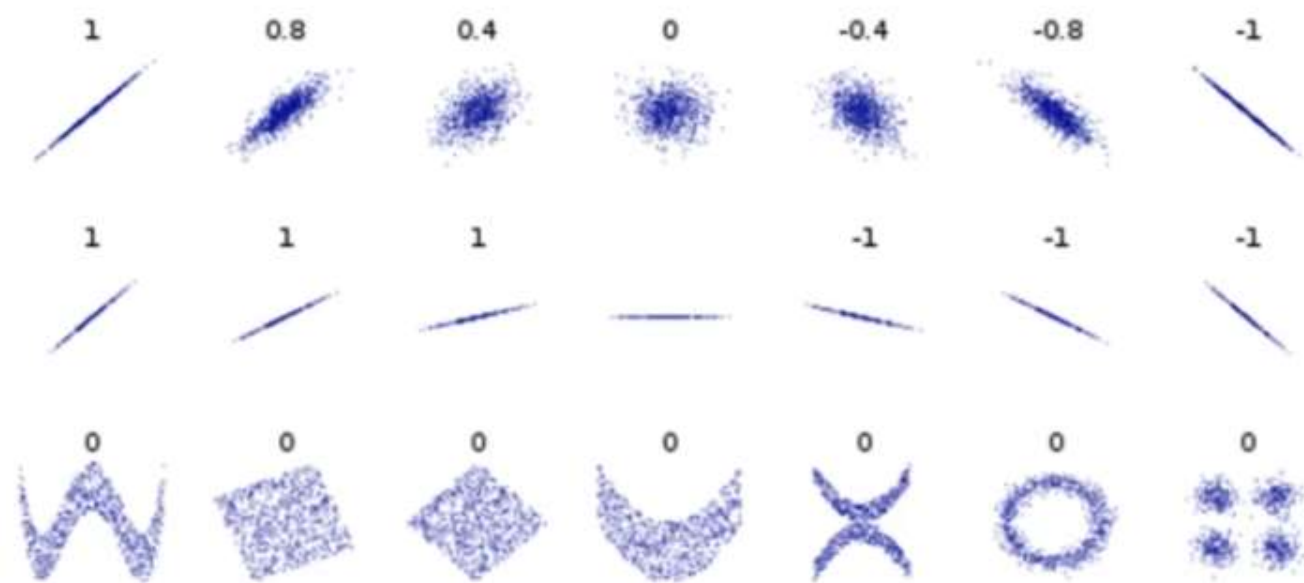
Pearson Correlation Coefficient

- Values fall between +1 and -1, where
 - 1 = total positive linear correlation
 - 0 = no linear correlation
 - 1 = total negative linear correlation



Pearson Correlation Coefficient

- Several sets of (x, y) points, with the correlation coefficient for each set:





Correlation Exercise

- A company decides to test sales of a new product in five separate markets, to determine the best price point.
- They set a different price in each market and record sales volume over the same 30 day period.

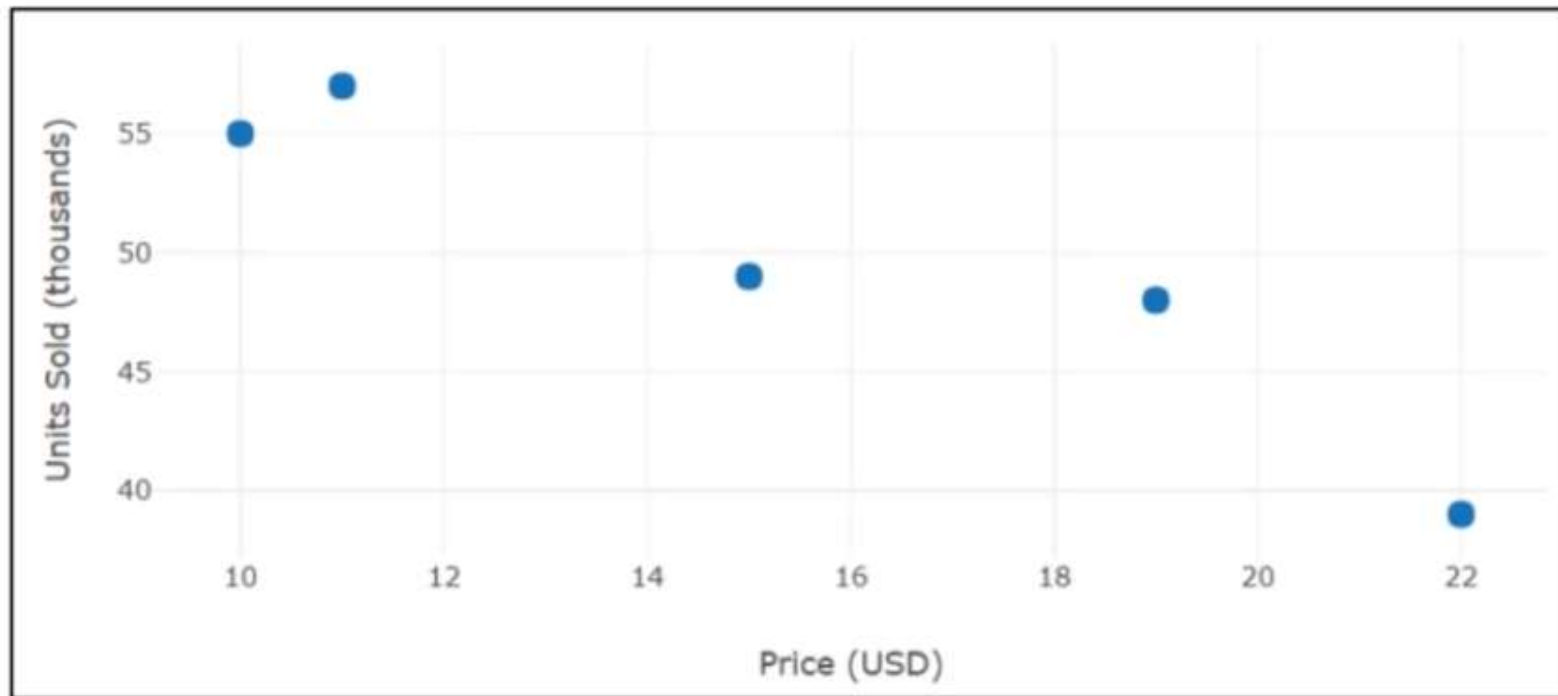




Correlation Exercise

- These are the results
- Plot the results

Price (USD)	Units Sold (thousands)
10	55
11	57
15	49
19	48
22	39

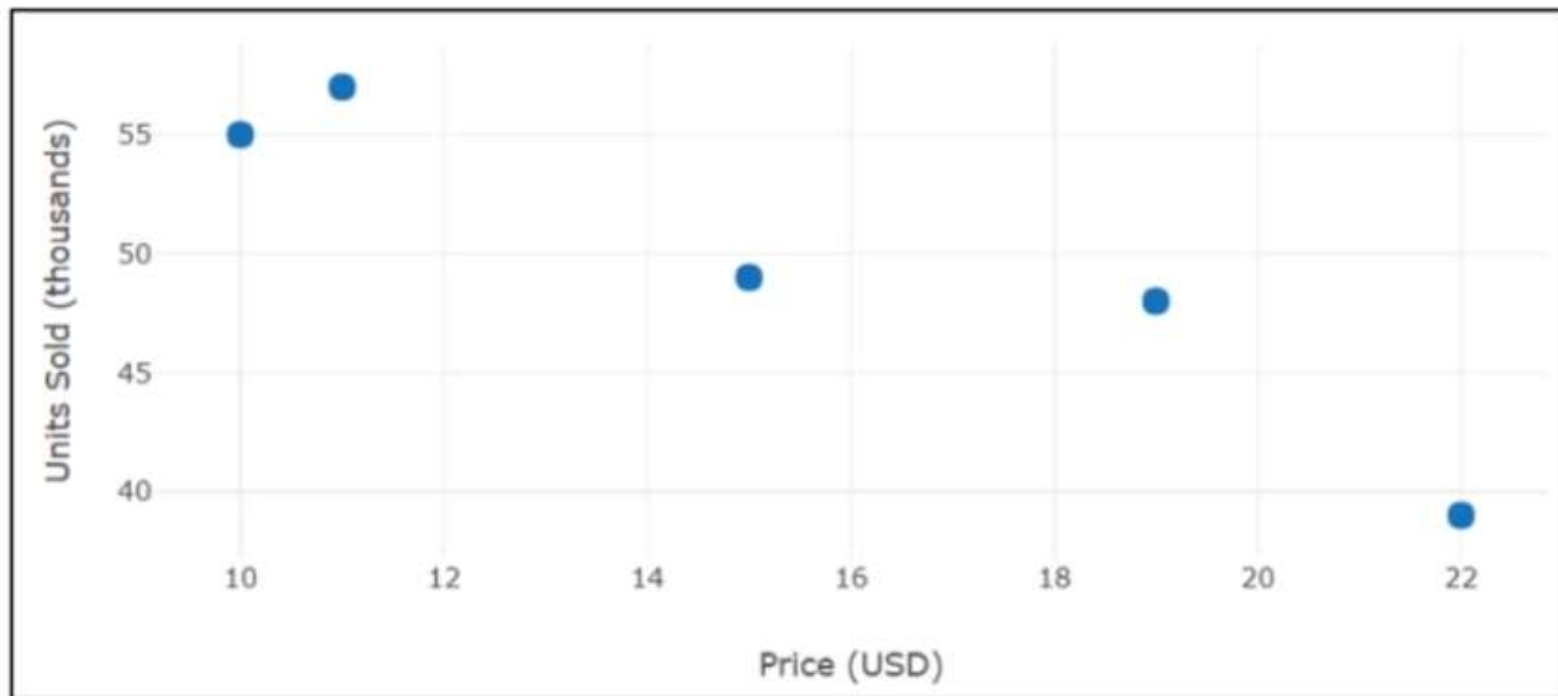




Correlation Exercise

- There appears to be a strong correlation, but how strong?

Price (USD)	Units Sold (thousands)
10	55
11	57
15	49
19	48
22	39





Correlation Exercise

1. Recall the simplified correlation formula:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}$$

Price (USD)	Units Sold (thousands)
10	55
11	57
15	49
19	48
22	39

2. Find the mean of x and y:

$$\bar{x} = \frac{10 + 11 + 15 + 19 + 22}{5} = 15.4$$

$$\bar{y} = \frac{55 + 57 + 49 + 48 + 39}{5} = 49.6$$



Correlation Exercise

$$\rho_{X,Y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}$$

$$\bar{x} = 15.4 \quad \bar{y} = 49.6$$

3. Calculate $(x - \bar{x})$ and $(y - \bar{y})$:

Price (USD)	Units Sold (thousands)	$(x - \bar{x})$	$(y - \bar{y})$
10	55	-5.4	5.4
11	57	-4.4	7.4
15	49	-0.4	-0.6
19	48	3.6	-1.6
22	39	6.6	-10.6



Correlation Exercise

$$\rho_{X,Y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}$$

$$\bar{x} = 15.4 \quad \bar{y} = 49.6$$

4. Calculate $(x - \bar{x})(y - \bar{y})$:

Price (USD)	Units Sold (thousands)	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})(y - \bar{y})$
10	55	-5.4	5.4	-29.16
11	57	-4.4	7.4	-32.56
15	49	-0.4	-0.6	0.24
19	48	3.6	-1.6	-5.76
22	39	6.6	-10.6	-69.96



Correlation Exercise

$$\rho_{X,Y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}$$

$$\bar{x} = 15.4 \quad \bar{y} = 49.6$$

5. Calculate $(x - \bar{x})^2$ and $(y - \bar{y})^2$:

Price (USD)	Units Sold (thousands)	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
10	55	-5.4	5.4	-29.16	29.16	29.16
11	57	-4.4	7.4	-32.56	19.36	54.76
15	49	-0.4	-0.6	0.24	0.16	0.36
19	48	3.6	-1.6	-5.76	12.96	2.56
22	39	6.6	-10.6	-69.96	43.56	112.36



Correlation Exercise

$$\rho_{X,Y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}$$

$$\bar{x} = 15.4 \quad \bar{y} = 49.6$$

6. Compute the sums:

Price (USD)	Units Sold (thousands)	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
10	55	-5.4	5.4	-29.16	29.16	29.16
11	57	-4.4	7.4	-32.56	19.36	54.76
15	49	-0.4	-0.6	0.24	0.16	0.36
19	48	3.6	-1.6	-5.76	12.96	2.56
22	39	6.6	-10.6	-69.96	43.56	112.36
		Σ		-137.2	105.2	199.2



Correlation Exercise

$$\rho_{X,Y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}$$

$$\bar{x} = 15.4 \quad \bar{y} = 49.6$$

7. Plug these into the original formula:

Price (USD)	Units Sold (thousands)	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
10	55	-5.4	5.4	-29.16	29.16	29.16
11	57	-4.4	7.4	-32.56	19.36	54.76
15	49	-0.4	-0.6	0.24	0.16	0.36
19	48	3.6	-1.6	-5.76	12.96	2.56
22	39	6.6	-10.6	-69.96	43.56	112.36
		Σ		-137.2	105.2	199.2



Correlation Exercise

$$\rho_{X,Y} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2} \sqrt{\sum(y - \bar{y})^2}}$$

$$\bar{x} = 15.4 \quad \bar{y} = 49.6$$

7. Plug these into the original formula:

$$\begin{aligned}\rho_{X,Y} &= \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2} \sqrt{\sum(y - \bar{y})^2}} = \frac{-137.2}{\sqrt{105.2} \sqrt{199.2}} \\ &= \frac{-137.2}{10.26 \times 14.11} = \frac{-137.2}{144.8} = -0.948\end{aligned}$$



Correlation Exercise

- $\rho_{X,Y} = -0.948$ shows a *very strong* negative correlation!

