Exercise 4 – Detailed Solution

Data (given)

Measurements every five years between 1960 and 1995:

Year 1995	1960	1965	1970	1975	1980	1985	1990
X (Rainy days) 162	198	196	199	164	170	163	149
Y (Rainfall mm) 670	739	880	631	658	690	501	501

Number of observations: $N = 8$.

Step 1 — Scatter plot

A scatter plot with the least-squares line is shown below. Each point (X_i, Y_i) represents one year, and the red dashed line is the fitted regression line.

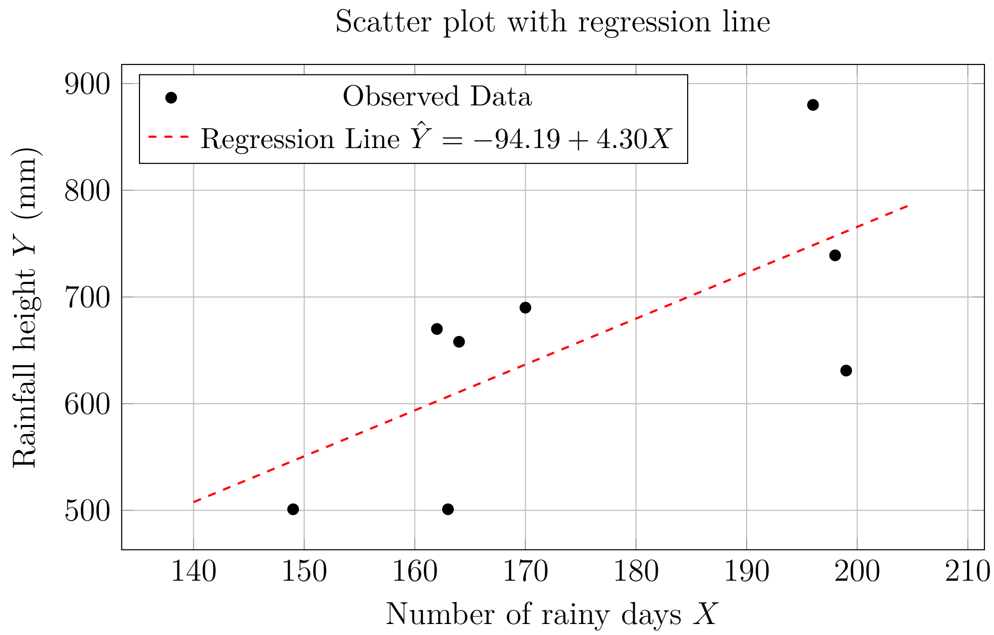

Figure 1: Scatter plot of rainfall (Y) versus rainy days (X).

Step 2 — Notation and formulas (population version)

We use the following (population) formulas, dividing by N:

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i, \quad \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i.$$

$$E[X^2] = \frac{1}{N}\sum_{i=1}^{N} X_i^2, \quad E[Y^2] = \frac{1}{N}\sum_{i=1}^{N} Y_i^2.$$

$$\mathrm{Var}(X) = E[X^2] - \bar{X}^2, \quad \mathrm{Var}(Y) = E[Y^2] - \bar{Y}^2.$$

$$s_X = \sqrt{\mathrm{Var}(X)}, \quad s_Y = \sqrt{\mathrm{Var}(Y)}.$$

$$E[XY] = \frac{1}{N}\sum_{i=1}^{N} X_i Y_i, \quad \mathrm{Cov}(X,Y) = E[XY] - \bar{X}\bar{Y}.$$

$$r = \frac{\mathrm{Cov}(X,Y)}{s_X s_Y}, \quad b = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}, \quad a = \bar{Y} - b\bar{X}.$$

Step 3 — Compute sums, means and second moments (digit-by-digit)

$$\sum_{i=1}^{8} X_i = 198 + 196 + 199 + 164 + 170 + 163 + 149 + 162 = 1401.$$

$$\sum_{i=1}^{8} Y_i = 739 + 880 + 631 + 658 + 690 + 501 + 501 + 670 = 5270.$$

Therefore:

$$\bar{X} = \frac{1401}{8} = 175.125, \quad \bar{Y} = \frac{5270}{8} = 658.75.$$

$$\sum X_i^2 = 198^2 + 196^2 + \cdots + 162^2 = 248031, \quad \sum Y_i^2 = 739^2 + 880^2 + \cdots + 670^2 = 3578648.$$

$$E[X^2] = \frac{248031}{8} = 31003.875, \quad E[Y^2] = \frac{3578648}{8} = 447331.0.$$

Step 4 — Variances and standard deviations

$$\mathrm{Var}(X) = E[X^2] - \bar{X}^2 = 31003.875 - (175.125)^2.$$

$$(175.125)^2 = 30668.765625 \Rightarrow \mathrm{Var}(X) = 335.109375.$$

$$s_X = \sqrt{335.109375} = 18.305992871188387.$$

Similarly for Y:

$$(\bar{Y})^2 = (658.75)^2 = 434951.5625,$$

$$\mathrm{Var}(Y) = 447331.0 - 434951.5625 = 13379.4375, \quad s_Y = \sqrt{13379.4375} = 115.66951845667899.$$

Step 5 — Covariance and Pearson correlation

$$\sum X_i Y_i = 198 \cdot 739 + 196 \cdot 880 + \cdots + 162 \cdot 670 = 934435,$$

$$E[XY] = \frac{934435}{8} = 116804.375.$$

$$\mathrm{Cov}(X,Y) = E[XY] - \bar{X}\bar{Y} = 116804.375 - (175.125)(658.75).$$

$$(175.125)(658.75) = 115363.59375 \Rightarrow \mathrm{Cov}(X,Y) = 1440.78125.$$

$$r = \frac{1440.78125}{(18.30599287)(115.66951846)} \approx 0.6804337261.$$

Interpretation: $r \approx 0.6804$ indicates a moderately strong positive linear association between the number of rainy days X and total rainfall Y: when X increases, Y tends to increase.

Step 6 — Least-squares regression of Y on X

$$b = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} = \frac{1440.78125}{335.109375} \approx 4.2994358185.$$

$$a = \bar{Y} - b\bar{X} = 658.75 - 4.2994358185 \times 175.125 = -94.18869772.$$

$$\boxed{\hat{Y} = -94.18869772 + 4.2994358185\, X.}$$

Interpretation: Each additional rainy day is associated with an increase of about 4.30 mm in total rainfall. The intercept -94.19 has no physical meaning (it represents the extrapolated value at $X = 0$).

Step 7 — Relationship between X and Y

Yes, there is a positive linear association between X and Y. Pearson $r \approx 0.680$ (moderately strong), and the regression slope $b \approx 4.299$ is positive. Thus, years with more rainy days tend to have greater total rainfall.

Numeric summary

Statistic	Symbol	Value
Mean of X	$\bar{X}$	175.125
Mean of Y	$\bar{Y}$	658.75
Variance of X	$\mathrm{Var}(X)$	335.1094
Variance of Y	$\mathrm{Var}(Y)$	13379.4375
Std. deviation of X	s_X	18.306
Std. deviation of Y	s_Y	115.670
Covariance	$\mathrm{Cov}(X, Y)$	1440.7813
Correlation	r	0.6804
Slope	b	4.2994
Intercept	a	-94.1887

Conclusion

There exists a statistically meaningful and positive linear relationship between the number of rainy days and the total rainfall height. The fitted regression model is:

$$\boxed{\hat{Y} = -94.1887 + 4.2994\, X.}$$

This indicates that an increase in rainy days corresponds to higher total rainfall on average, with correlation coefficient $r = 0.68$ — a moderately strong positive association.