

### WORKSHEET - REGRESSION

The table below displays data on the temperature ( $^{\circ}\text{F}$ ) reached on a given day and the number of cans of soft drink sold from a particular vending machine in front of a grocery store.

temperature	70	75	80	90	93	98	72	75	75	80	90	95	98	91	98
quantity	30	31	40	52	57	59	33	38	32	45	53	56	62	51	58

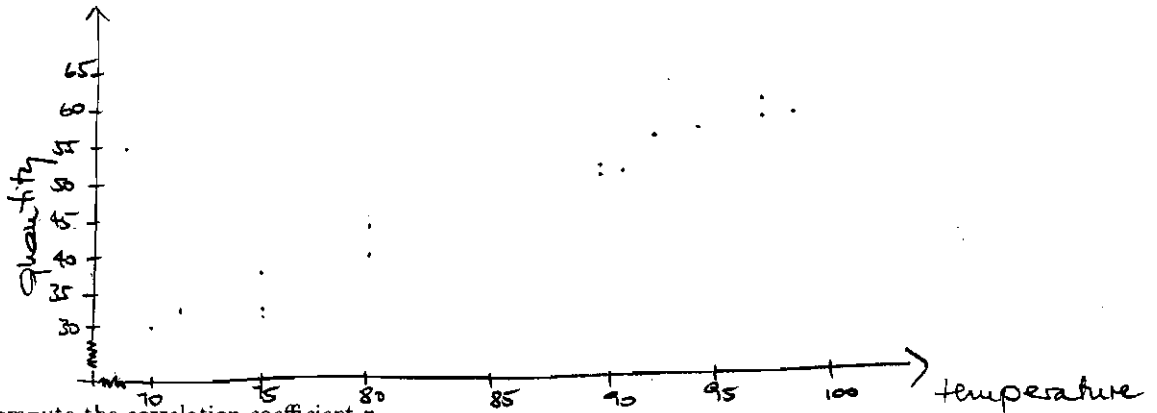
$$\bar{x} = 85.3, \quad \bar{y} = 46.46, \quad s_x^2 = 107.09, \quad s_y^2 = 131.6952, \quad \sum_{i=1}^{15} (x_i - \bar{x})(y_i - \bar{y}) = 1627.667$$

1. Identify the explanatory variable ( $x$ ) and the response variable ( $y$ ).

explanatory variable: temperature

response variable: quantity sold

2. Draw a scatterplot of the data.



3. Compute the correlation coefficient  $r$ .

$$r = \frac{\frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})}{s_x \cdot s_y} = \frac{\frac{1}{14} \cdot 1627.667}{10.35 \cdot 11.48} = \underline{\underline{0.9784}}$$

$$s_x = \sqrt{s_x^2} = \sqrt{107.09} = 10.35 \quad s_y = \sqrt{s_y^2} = \sqrt{131.6952} = 11.48$$

4. Based on the computed value of  $r$ , what can you say about the association between the temperature and the number of soft drinks sold.

there seems to be a strong, positive linear association between the temperature and the # of drinks being sold.

5. Compute the slope  $b$  for the least squares regression line and give an interpretation of the slope **within the context of the problem**.

$$b = r \cdot \frac{S_y}{S_x} = 0.9784 \cdot \frac{11.48}{10.35} = \underline{\underline{1.0852}}$$

interpret: For every additional degree (in °F) the number of cans sold increases by about 1.0852.

6. Compute the intercept  $a$  for the least squares regression line and give an interpretation of the slope **within the context of the problem**.

$$a = \bar{y} - b\bar{x} = 46.4\bar{6} - 1.0852 \cdot 85.\bar{3} = \underline{\underline{-46.13706}}$$

The intercept  $a$  does not have a meaningful interpretation in this problem

7. State the least squares regression line.

$$\hat{y} = -46.14 + 1.0852 \cdot X$$

8. For a temperature of 85°F, predict how many cans of soft drinks will be sold.

$$\hat{y} = -46.14 + 1.0852 \cdot 85 = 46.102$$

9. Can you predict the number of soft drinks being sold for a temperature of 62°F? Explain why or why not!

No, this would be extrapolation, as 62 is not in the range of  $x$ -values used to fit the line

10. Compute the residual for for the predicted values based on  $x = 72$  and  $x = 91$ .

$$\begin{aligned} \hat{y}_1 &= -46.14 + 1.0852 \cdot 72 = 31.9944 \Rightarrow \text{residual} = y - \hat{y} \\ &= 33 - 31.9944 \\ &= 1.0056 \\ \hat{y}_2 &= -46.14 + 1.0852 \cdot 91 = 52.6132 \\ \Rightarrow \text{residual} &= 51 - 52.6132 = -1.6132 \end{aligned}$$

11. Compute the coefficient of determination and give an interpretation the coefficient of determination.

$r^2 = (0.9784)^2 = 0.9573 \Rightarrow \approx 95.73\%$  of the variation in  $y$  (# of drinks sold) can be explained by the least squares regression line.