Scatter Plot

Used to compare 2 different pieces of data - related.

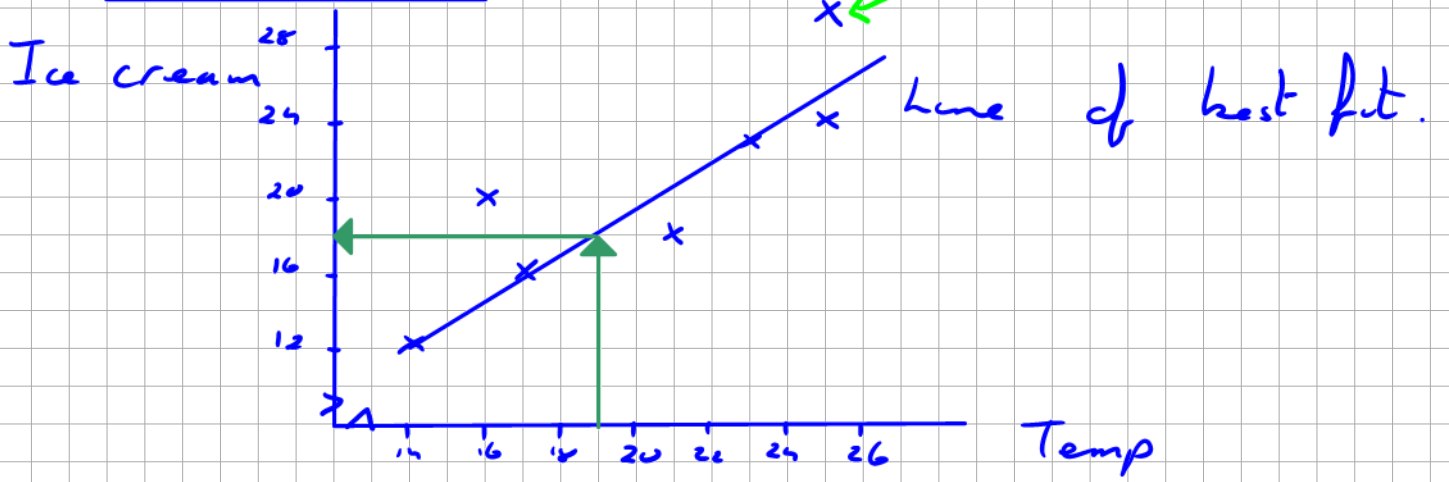
Temp	21°	23°	16°	17°	25°	14°	25°
Ice Cream	18	23	20	16	30	12	24

Bottom

Top

Top

Bottom



As temperature rises the number of ice cream sales rises. This is called a correlation. This is a positive correlation.

Positive as one variable increase the other increases.

Line of best fit ?

$(14, 12)$ $(23, 23)$

x_1 y_1 x_2 y_2

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{23 - 12}{25 - 14} = \frac{11}{9}$$

$$y - y_1 = m(x - x_1)$$

$$y - 12 = \frac{11}{9}(x - 14)$$

How many ice creams would you expect to sell at 19°C.

Expect to sell 18 ice creams from graph.

$$y - 12 = \frac{11}{9}(x - 14) \quad \text{sub } x = 19$$

$$y - 12 = \frac{11}{9}(19 - 14)$$

$$y = 18.1 \quad \text{Ans } y = 19$$

round up to nearest ice cream.

Eq of Line

$$y = a + bx \quad \text{from calculator}$$

$a = -2.33 \quad b = 1.13$

$$y = -2.33 + 1.13x$$

$x = 19 \quad y = -2.33 + 1.13(19)$

Correleration

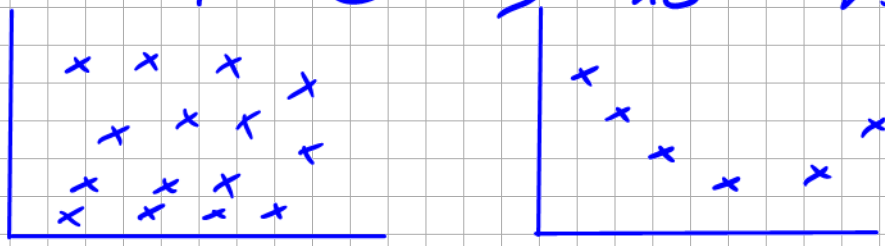
Is there a relationship between 2 variables = LINEAR.

Variable used is r.

$$-1 \leq r \leq 1$$

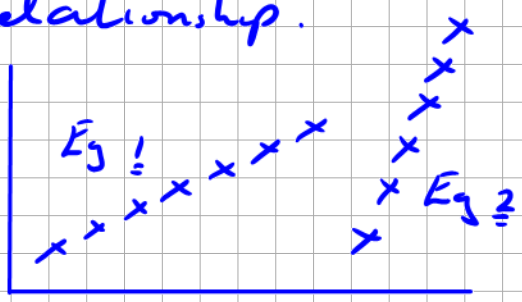
This shows the strength of the relationship (correleration)

$r = 0 \Rightarrow$ no relationship

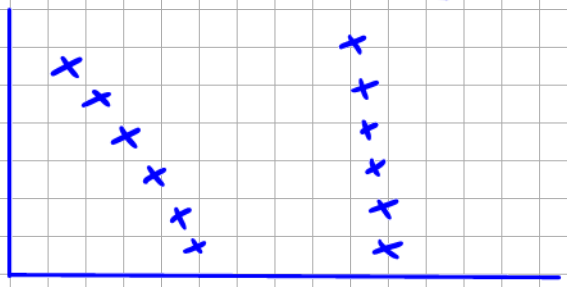


$r = 1 \Rightarrow$ perfect positive relationship.

It is not telling the ratio (slope) of relationship.



$r = -1$ perfect negative relationship.



Closer to 1 or -1 the stronger the relationship.

Causality

Smoking correlation cancer = causality

Heat correlation and ice cream = not causality.

2. A teacher asked a random sample of eight students to record their study times. She then made a table for total hours studied, t, over four weeks and class exam score, y, at the end of four weeks. The results are given in the table below.

Study time (t)	25	35	29	45	21	37	33	47
Exam score (y)	92	81	84	74	85	80	84	80

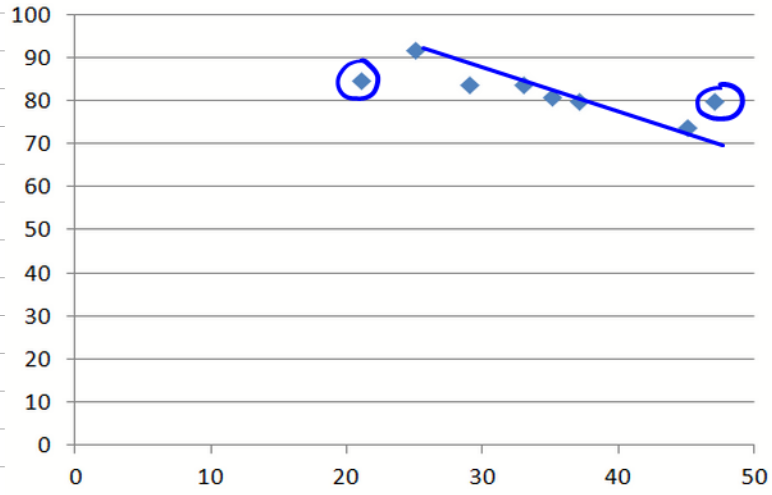
(i) Draw a scatter plot of the data.

(ii) Describe the apparent relationship between study time and test score. Does it surprise you?

(iii) Identify outliers and potential influential variables.

(iv) Compute the linear correlation coefficient, r.

Exam Dependent ≈



Time = Independent ≈

(ii) More time put the worse the scored. Surprised ⇒ yes

(iii) 21, 85 is outlier ⇒ might be not favourite topic

47, 80 ⇒ like subject.

(iv) $r = -0.796$

Note. r is not the slope.

r is affected by outliers.

Causality ? ⇒ Yes ⇒ pressure is telling