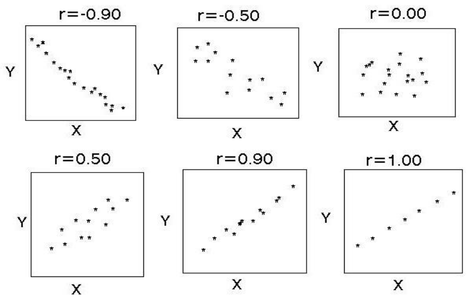
Week 1/2
Term 3
2023

HAWKER COLLEGE
Engage | Inspire | Achieve

Learning Brief

SMM4

Goals



This fortnight we are going to:

- Display, summarise, and describe relationships in bivariate data
- Identify dependent (response) and independent (explanatory) variable
- Calculate and use r and r^2 to interpret the strength of bivariate relationship between two variables. Comment on the reliability of the predicted values using r^2
- Understand that underlying theory behind least-squares to fit straight line to bivariate data
- Calculate the equation of a least-squares regression line by hand and by using the CAS calculator
- Use the equation of the line to 'predict' data values for a given value of x

Theoretical Components

Make notes on the following chapters:

Maths Quest 12 Further Maths

- 2E - Scatterplots
- 2F - Pearson's product-moment correlation coefficient
- 2G - Calculating r and the coefficient of determination
- 3C - Fitting a straight line - least-squares regression
- 3D - Interpretation, interpolation and extrapolation

Minimising Least Square error:

- https://www.youtube.com/watch?v=6OvhLPS7rj4 (Probably no need to worry about the proof)

Practical Components

Do the following questions:
Organise your solutions neatly in your exercise book.

Chapter 2 and 3 of Maths Quest 12 Further Maths (pdf – Google Classroom)

- 2E: 2, 3, 4
- 2F: 1-4
- 2G: 1, 5, 7, 8 (Use CAS for calculation of correlation coefficient)
- 3C: As many as you need
- 3D: As many as you need

Mathspace

Investigation

See next page

Other

Assignment will be issued Monday Week 2.

Fun fact: Many machine learning algorithms use linear regression modelling as a means to find the best fit linear line between the independent and dependent variables. While linear models are easy to understand and efficient to train machine learning models on, they can be prone to outliers and can be affected by noisy data.

Week 1 and 2 Investigation

Question 1: Describe the following cartoon in relation to causation and correlation.

Question 2: Copy and complete the table to estimate the parameters of linear regression (i.e. constant and the coefficient in $y = ax + c$). Then use the totals and the formulae given to work out **'a', 'b', r and r^2**. Check your answers on CAS. The first two columns give the values for age (x, in years) and systolic blood pressure (y, in mmHg) for 15 women.

X	Y	X^2	Y^2	XY
42	130	1764	16900	5460
46	115			
42	148			
71	100			
80	156			
74	162			
70	151			
80	156			
85	162			
72	158			
64	155			
81	160			
41	125			
61	150			
75	165			
Total: 984	Total: 2193			

Formulae:

$$a = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

$$b = \bar{Y} - a\bar{X}$$

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{n}\right]}}$$

Use the formula you have found to predict the systolic blood pressure for a 59 year old woman.

The coefficient of determination (r^2) provides a measure of how well the linear rule linking the two variables (x and y) predicts the value of y when we are given the value of x. Comment on the r^2 found in the above example and the predictability of the linear model found.