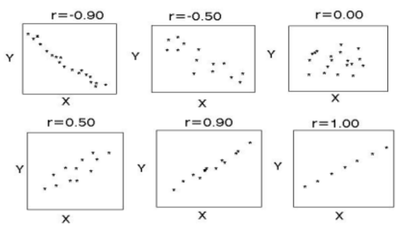
Week 1/2
Term 3
2023

HAWKER COLLEGE
Engage | Inspire | Achieve

MM4
Bivariate data
Assignment Warm - Up

Goals



By the end of this week, you should be able to:
- Display, summarise, and describe relationships in bivariate data
- Identify dependent (response) and independent (explanatory) variable
- Read and interpret information presented in back-to-back stem plots and parallel box plots
- Calculate and use $r \text{ and } r^2$ to interpret the strength of bivariate relationship between two variables.
- Use CAS to calculate $r \text{ and } r^2$ and linear regression function $y = ax + b$

Theoretical Components

Chapter 2 and 3 of Quest Further Maths 12 (pdf on Google Classrooms):

•Read through Section 2A to 2C on Dependent and Independent Variables, Back-to-Back Stem Plots, Parallel Plots. Study and make notes on Examples 1-4.
•Read through Section 2E on Scatter Plots. Study examples 8 & 9.
•Read through Section 2F on Correlation Coefficient. Study example 10.
•Read through Section 2G on Calculating $r \text{ and } r^2$. Study examples 11 & 12.
Make your notes on the following key concepts:
- Dependent/Independent Variables
- Back-to-Back Stem Plots
- Parallel Box plots
- Scatter Plots
- r and r^2

Practical Components

Do the following questions from **Chapter 2: Bivariate data** (pdf – GC). Organise your solutions neatly in your exercise book.

Ex 2A Dependent and independent variables
Qs 1, 4

Ex 2B Back-to-back stem plots
Qs 1, 4

Ex 2C Parallel boxplots
Qs 1, 2, 4, 6

Ex 2E Scatterplots
Qs 2, 3, 4

Ex 2F Pearson's correlation coefficient
Qs 1 - 4

Ex 2G Calculating $r \text{ and } r^2$.
Qs 1, 5, 7, 8 (Use CAS for calculation of correlation coefficient)

Ex 3C Fitting a straight line – least squares regression
Qs 1, 2, 3, 4, 6, 8, 9,12

Ex 3D Interpretation, interpolation, and extrapolation
Qs 1, 2, 7, 9

Investigation

See the next page.

This task is essential for understanding the calculation of the correlation co-efficient.

QFO
Quiz/Forum/Other

No Mathspace for this learning brief. 😊
Assignment will be issued Monday Week 2.

MM4 Week 1/2 INVESTIGATION

1. Complete the table below.

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 of X $\sum$ = 984	Total of Y $\sum$= 2193	Total of X^2	Total of Y^2	Total of XY

2. Use the values from the table and the formulae given to estimate the parameters of the linear regression. ie. work out 'a', 'b', r and r^2. Show all working.

Formulae:

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

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

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

Show working:

Check your answers on your CAS or the following links:
https://www.socscistatistics.com/tests/regression/default.aspx
https://www.easycalculation.com/statistics/r-squared.php

3. Write out the linear regression rule for the age (x, in years) and systolic blood pressure (y, in mmHg) for 15 women ($y = ax + b$).

4. Comment on the value of r^2.

5. Use the rule to estimate the systolic blood pressure for a 59 year old woman.

6. Is the above question an example of interpolation or extrapolation? (see notes below). Interpret the accuracy of your result.

Interpolation

Interpolation is the use of the regression line to predict values in between two values already in the data set. If the data set is highly linear (*r* is near +1 or -1) then we can be confident that our interpolated value is quite accurate. If the data are not highly linear (*r* is near 0) then our confidence is duly reduced.

Extrapolation

Extrapolation is the use of the regression line to predict values smaller than the smallest value in the data set or larger than the largest value.

Two problems may arise in attempting to extrapolate from a data set. Firstly, it may not be reasonable to extrapolate too far away from the given data values. For example, suppose there is a weather data set for 5 days. Even if it is highly linear (*r* is near +1 or -1), a regression line used to predict the same data 15 days in the future is highly risky. Weather has a habit of randomly fluctuating and patterns rarely stay stable for very long.

Secondly, the data may be highly linear in a narrow band of the given data set. For example, there may be data on stopping distances for a train at speeds of between 30 and 60 km/h. Even if they are highly linear in this range, it is unlikely that things are similar at very low spends (0-15 km/h) or high speeds (over 100 km/h).

Generally, we should feel more confident about the accuracy of a prediction derived from interpolation than on derived from extrapolation. Of course, it still depends upon the correlation coefficient (*r*). The closer to linearity the data are, the more confident our predictions in all cases.