

Chapter 4 - Correlation

4.1 - Recap on Correlation - Pg. 2

4.2 - Linear Regression - Pg. 3 - 4



Personal notes:

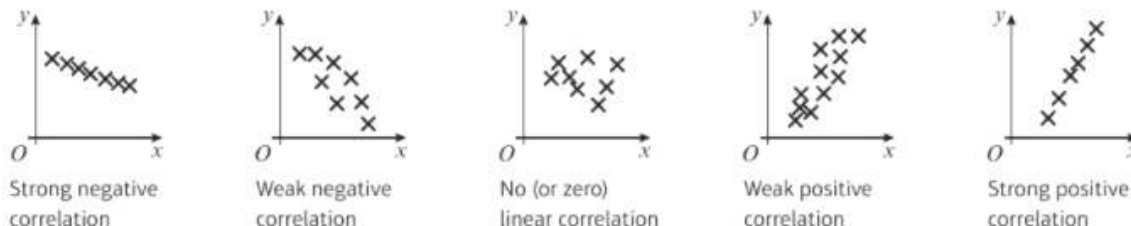


4.1 - Recap on Correlation

Notes

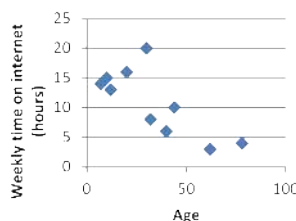
- Bivariate data is data which has pairs of values of two variables.
- Common types of bivariate analysis include
 - Scatter diagram
 - Regression analysis
 - Correlation coefficient

■ Correlation describes the nature of the linear relationship between two variables.

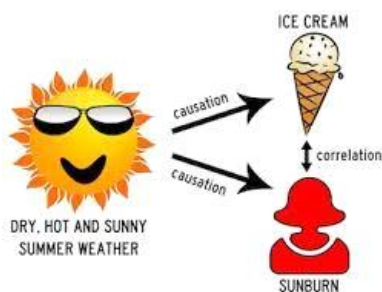


Practice

- State the correlation shown.
- Describe/interpret the relationship between age and weekly time on the internet.

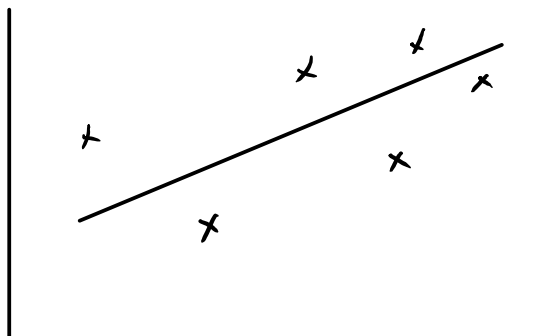


- Two variables have a **causal relationship** if a change in one variable directly causes a change in the other. Just because two variables show correlation it does not necessarily mean that they have a causal relationship.
- When two variables are correlated, use common sense to determine whether they have a causal relationship.



4.2 - Linear Regression

Notes

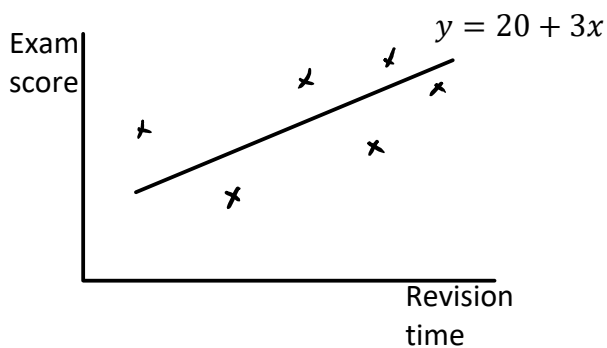


- One type of line of best fit is
- It is a straight line that
- The regression line

- It is written in the form

Example

Interpret the meaning of the value 20 and 3 respectively.



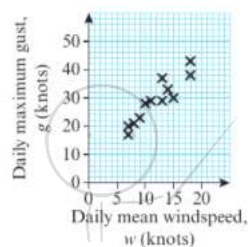
Example

From the large data set, the daily mean windspeed, w knots, and the daily maximum gust, g knots, were recorded for the first 15 days in May in Camborne in 2015.

w	14	13	13	9	18	18	7	15	10	14	11	9	8	10	7
g	33	37	29	23	43	38	17	30	28	29	29	23	21	28	20

© Crown Copyright Met Office

The data was plotted on a scatter diagram:



- a** Describe the correlation between daily mean windspeed and daily maximum gust.

The equation of the regression line of g on w for these 15 days is $g = 7.23 + 1.82w$.

- b** Give an interpretation of the value of the gradient of this regression line.
- c** Justify the use of a linear regression line in this instance.

BF MATHS

4.2 - Linear Regression

- You should only use the regression line to make predictions for values of the dependent variable (to be measured) that are within the range of given data.
- Estimating a value **inside** the data range is known as
- Estimating a value **outside** the data range is known as



"WE ONLY TEST OUR DRUGS ON PLANTS, SO WE HAVE TO DO LOTS AND LOTS OF EXTRAPOLATING."



"TELL ME ABOUT YOURSELF. I WORK MAINLY BY EXTRAPOLATION."

Example

The head circumference, y cm, and gestation period, x weeks, for a random sample of eight newborn babies at a clinic were recorded.

Gestation period (x weeks)	36	40	33	37	40	39	35	38
Head circumference (y cm)	30.0	35.0	29.8	32.5	33.2	32.1	30.9	33.6

The scatter graph shows the results.

The equation of the regression line of y on x is $y = 8.91 + 0.624x$.

The regression equation is used to estimate the head circumference of a baby born at 39 weeks and a baby born at 30 weeks.

a Comment on the reliability of these estimates.

A nurse wants to estimate the gestation period for a baby born with a head circumference of 31.6 cm.

b Explain why the regression equation given above is not suitable for this estimate.

