

Chapter

21

Linear modelling

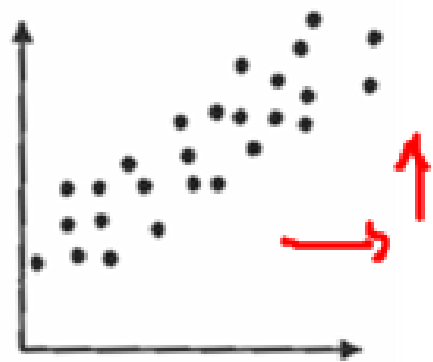
Ch 21 A – Correlation

Correlation - A mutual relationship or connection between two or more things.

There are several characteristics that describe correlation:

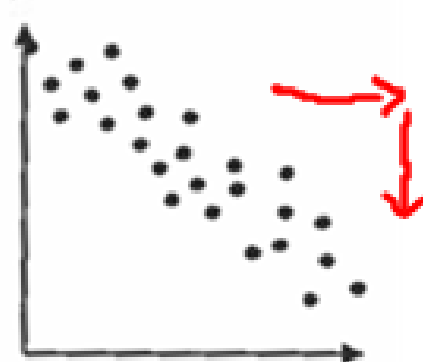
Direction:

Upward trend



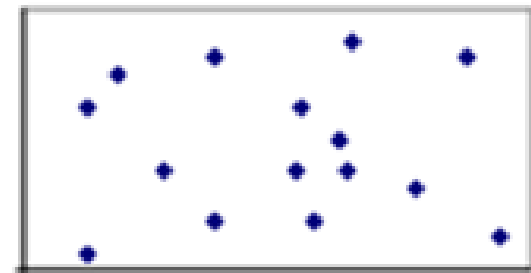
positive slope

Downward trend



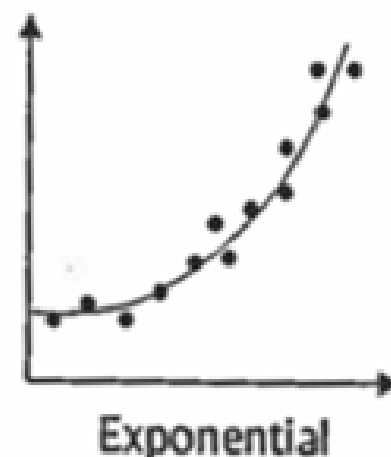
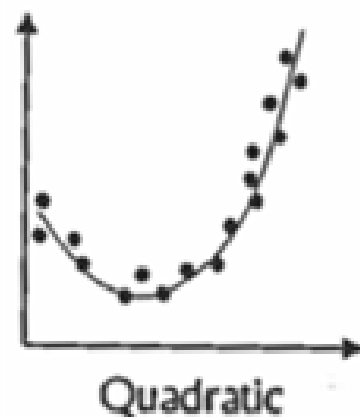
negative slope

Randomly scattered

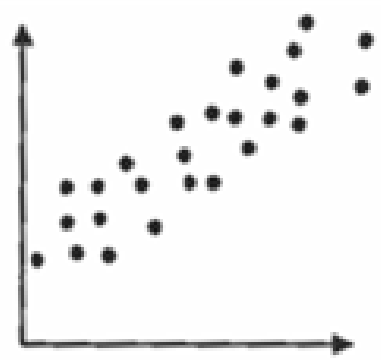


Linearity:

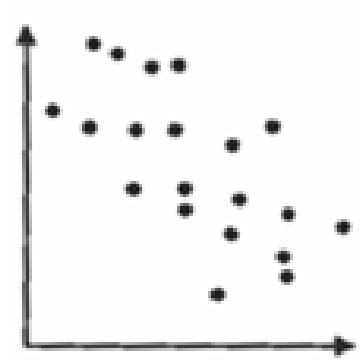
Scatter plots can be used to model a variety of situations. The above examples demonstrated linear scatter plots; however, scatter plots can be modelled by a variety of functions such as quadratic or exponential.



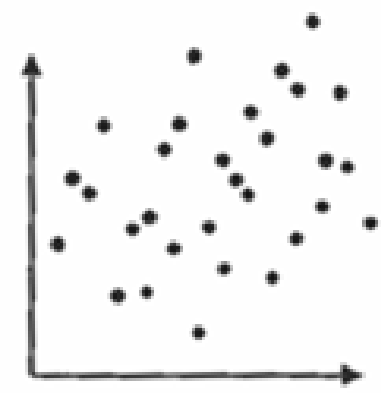
Strength:



strong

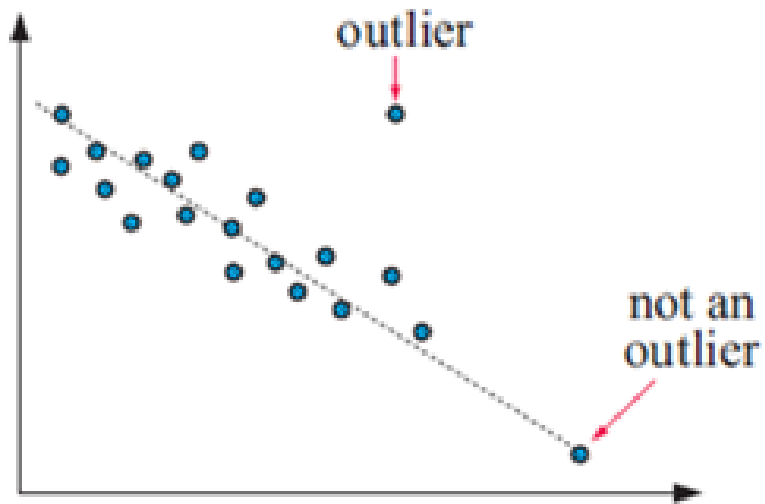


moderate



weak

Outliers: Points that do not follow the trend of the data



Ch 21 B – Pearson's Correlation Coefficient

Pearson's Correlation Coefficient is a measurement of the strength of linear correlation between two variables.

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 (y - \bar{y})^2}}$$

$x \rightarrow x$ value of the data

$\bar{x} \rightarrow$ mean of all x values of the data

$y \rightarrow y$ value of the data

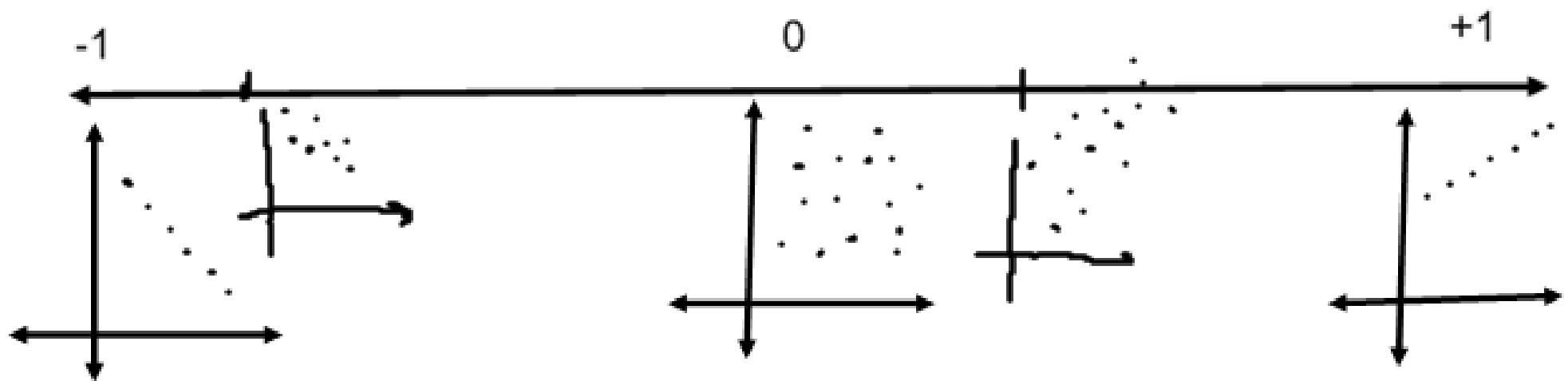
$\bar{y} \rightarrow$ mean of all y values of the data

The values of r range from -1 to +1

Perfect
negative
correlation

NO
correlation

Perfect
positive
correlation



Example 2**Self Tutor**

A group of adults is weighed, and their maximum speed when sprinting is measured:

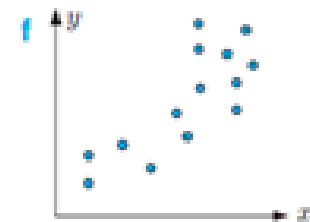
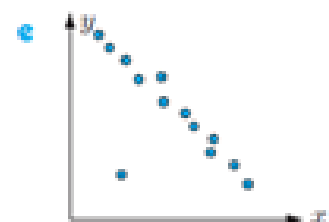
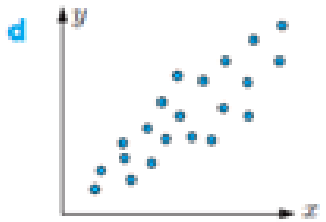
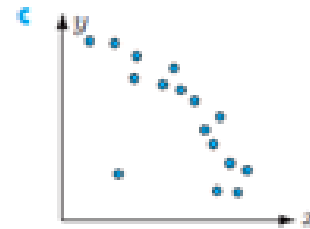
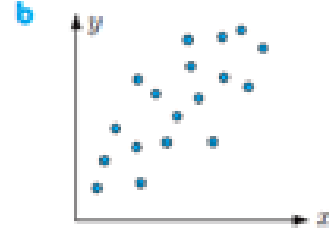
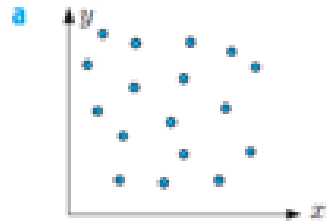
<i>Weight</i> x (kg)	85	60	78	100	83	67	79	62	88	68
<i>Maximum speed</i> y (km h ⁻¹)	26	29	24	17	22	30	25	24	19	27

- a Use technology to find r for the data.
- b Describe the correlation between *weight* and *maximum speed*.

Homework:

EXERCISE 21A

- 1 Describe what is meant by:
 - a a scatter diagram
 - b correlation
 - c positive correlation
 - d negative correlation
 - e an outlier.
- 2 For the following scatter diagrams, comment on:
 - i the existence of any *pattern* (positive, negative or no correlation)
 - ii the relationship *strength* (zero, weak, moderate or strong)
 - iii whether the relationship is linear
 - iv whether there are any outliers.



- 3 Ten students participated in a typing contest, where the students were given one minute to type as many words as possible. The table below shows how many words each student typed, and how many errors they made:

<i>Student</i>	A	B	C	D	E	F	G	H	I	J
<i>Number of words x</i>	40	53	20	65	35	60	85	49	35	76
<i>Number of errors y</i>	11	15	2	20	4	22	30	16	27	25

- a Draw a scatter diagram for this data.
- b Name the student who is best described as:
- slow but accurate
 - fast but inaccurate
 - an outlier.
- c Describe the direction and strength of correlation between these variables.
- d Is the data linear?

You can use technology to construct scatter diagrams.



GRAPHICS
CALCULATOR
INSTRUCTIONS



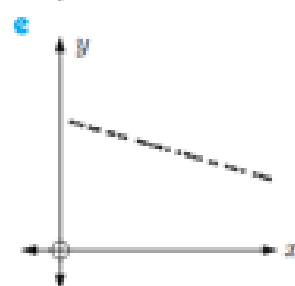
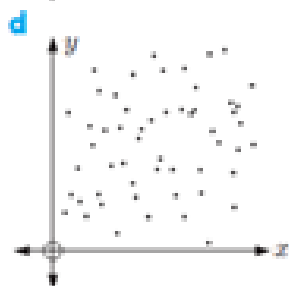
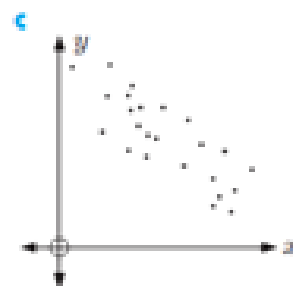
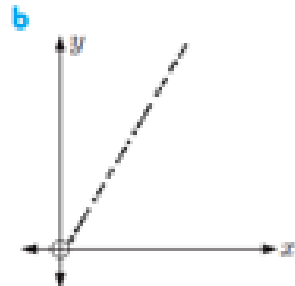
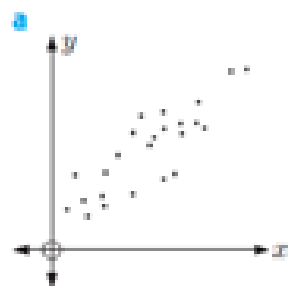
- 4 The scores awarded by two judges at an ice skating competition are shown in the table.

<i>Competitor</i>	P	Q	R	S	T	U	V	W	X	Y
<i>Judge A</i>	5	6.5	8	9	4	2.5	7	5	6	3
<i>Judge B</i>	6	7	8.5	9	5	4	7.5	5	7	4.5

- a Construct a scatter diagram for this data with Judge A's scores on the horizontal axis and Judge B's scores on the vertical axis.
- b Copy and complete the following comments about the scatter diagram:
There appears to be, correlation between Judge A's scores and Judge B's scores. This means that as Judge A's scores increase, Judge B's scores
- c Would it be reasonable to conclude that an increase in Judge A's scores *causes* an increase in Judge B's scores?

EXERCISE 21B

1 Match each scatter diagram with the correct value of r .



A $r = 1$

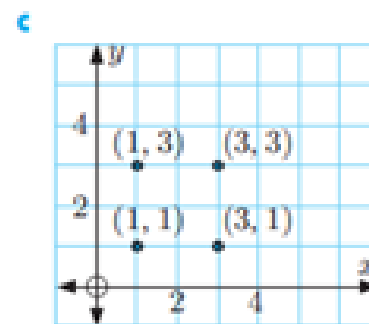
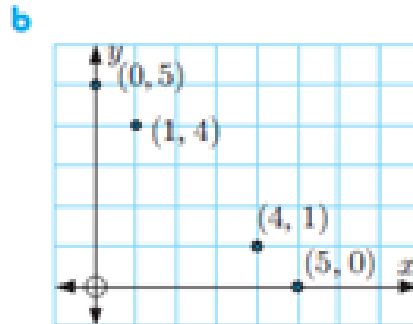
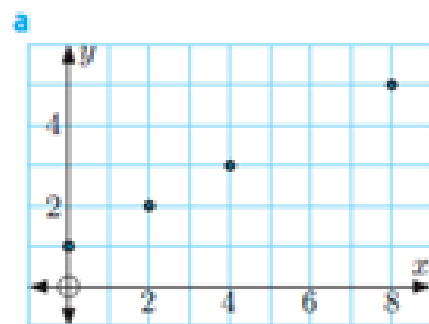
B $r = 0.6$

C $r = 0$

D $r = -0.7$

E $r = -1$

2 Use the formula $r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$ to determine the correlation coefficient r in the following:



Check your answers using a calculator.

- 3 The table alongside shows the ages of five children, and the number of times they visited the doctor in the last year:

<i>Age</i>	2	5	7	5	8
<i>Number of doctor visits</i>	10	6	5	4	3

- Draw a scatter diagram of the data.
 - Calculate the correlation coefficient by hand. *Check your answer using technology.*
 - Describe the correlation between *age* and *number of doctor visits*.
- 4 Jill hangs her clothes out to dry every Saturday, and notices that the clothes dry more quickly some days than others. She investigates the relationship between the temperature and the time her clothes take to dry:

<i>Temperature x ($^{\circ}\text{C}$)</i>	25	32	27	39	35	24	30	36	29	35
<i>Drying time y (min)</i>	100	70	95	25	38	105	70	35	75	40

- Draw a scatter diagram for this data.
 - Calculate r .
 - Describe the correlation between *temperature* and *drying time*.
- 5 The table below shows the ticket and beverage sales for each day of a 12 day music festival:

<i>Ticket sales ($\\$x \times 1000$)</i>	25	22	15	19	12	17	24	20	18	23	29	26
<i>Beverage sales ($\\$y \times 1000$)</i>	9	7	4	8	3	4	8	10	7	7	9	8

- Draw a scatter diagram for this data.
 - Calculate r .
 - Describe the correlation between *ticket sales* and *beverage sales*.
- 6 A local council collected data from a number of parks in the area, recording the size of the parks and the number of trees each contained:

<i>Size (hectares)</i>	2.8	6.9	7.4	4.3	8.5	2.3	9.4	5.2	8.0	4.9	6.2	3.3	4.5
<i>Number of trees</i>	18	31	33	24	13	17	40	32	37	30	32	25	28

- Draw a scatter diagram for this data.
- Would you expect r to be positive or negative? Explain your answer.
- Calculate r .
- Are there any outliers?
- Remove the outlier, and re-calculate r .