

1. A young family were looking for a new 3 bedroom semi-detached house. A local survey recorded the price x , in £1000, and the distance y , in miles, from the station of such houses.

The following summary statistics were provided

$$S_{xx} = 113573, S_{yy} = 8.657, S_{xy} = -808.917$$

- (a) Use these values to calculate the product moment correlation coefficient. (2)
- (b) Give an interpretation of your answer to part (a). (1)

Another family asked for the distances to be measured in km rather than miles.

- (c) State the value of the product moment correlation coefficient in this case. (1)

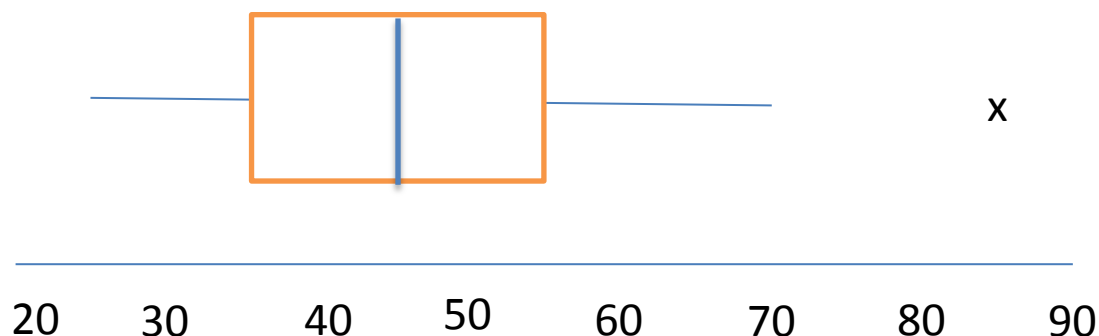
1 a)

$$r = \frac{S_{xy}}{S_x S_y} = \frac{S_{xy}}{\sqrt{S_{xx}} \sqrt{S_{yy}}} \quad r = \frac{-808.917}{\sqrt{113573} \times \sqrt{8.657}} = -0.816 \text{ (3. d. p)}$$

b) For these questions always answer specifically to the data i.e. yes is a strong negative correlation but the answer requires you to say that houses get cheaper the further from the station.

c) The same $= -0.816$ as it is a ratio and won't change.

2. The box plot in Figure 1 shows a summary of the weights of the luggage, in kg, for each musician in an orchestra on an overseas tour.



The airline's recommended weight limit for each musician's luggage was 45 kg.

Given that none of the musicians' luggage weighed exactly 45 kg,

(a) state the proportion of the musicians whose luggage was below the recommended weight limit.

(1)

A quarter of the musicians had to pay a charge for taking heavy luggage.

(b) State the smallest weight for which the charge was made.

(c) Explain what you understand by the + on the box plot in Figure 1, and suggest an instrument that the owner of this luggage might play.

(d) Describe the skewness of this distribution. Give a reason for your answer.

(2)

One musician of the orchestra suggests that the weights of luggage, in kg, can be modelled by a normal distribution with quartiles as given in Figure 1.

(e) Find the standard deviation of this normal distribution.

(4)

2 a) 45g is the median line so 50%.

b) The charge will be made for anything above $Q_3 = 54kg$.

c) The x represents an outlier i.e. a very heavy instrument. Possibly a double bass.

d) Carry out $Q_3 - Q_2$ and $Q_3 - Q_2 = 54 - 45 = 9$

$Q_2 - Q_1$ $Q_2 - Q_1 = 45 - 36 = 9$

As they are equal there is no skew and the distribution is symmetric.

$$e) P\left(M < \frac{54-45}{\sigma}\right) = 0.75 \quad \frac{54 - 45}{\sigma} = 0.67 \text{ (from tables)}$$

$$\sigma = 13.43 \text{ (2. d. p)}$$

3. A student is investigating the relationship between the price (y pence) of 100g of chocolate and the percentage ($x\%$) of cocoa solids in the chocolate.

The following data is obtained

Chocolate brand	A	B	C	D	E	F	G	H
x (% cocoa)	10	20	30	35	40	50	60	70
y (pence)	35	55	40	100	60	90	110	130

(You may use: $\sum x = 315$, $\sum x^2 = 15225$, $\sum y = 620$, $\sum y^2 = 56550$, $\sum xy = 28750$)

(a) On the graph paper on page 9 draw a scatter diagram to represent these data. (2)

(b) Show that $S_{xy} = 4337.5$ and find S_{xx} . (3)

The student believes that a linear relationship of the form $y = a + bx$ could be used to describe these data.

(c) Use linear regression to find the value of a and the value of b , giving your answers to 1 decimal place. (4)

(d) Draw the regression line on your scatter diagram. (2)

The student believes that one brand of chocolate is overpriced.

(e) Use the scatter diagram to
 (i) state which brand is overpriced,
 (ii) suggest a fair price for this brand.
 Give reasons for both your answers. (4)

3 a) Plot on graph

b) Using

$$S_{xy} = \sum xy - \frac{\sum x \sum y}{n}$$

$$S_{xx} = \sum x^2 - \frac{\sum x \sum x}{n}$$

$$S_{xy} = 28750 - \frac{315 \times 620}{8} = 4337.5$$

$$S_{xx} = 15225 - \frac{315 \times 315}{8} = 2821.875$$

c) Find b first and then a.

Using

$$b = \frac{S_{xy}}{S_{xx}}$$

$$b = \frac{4337.5}{2821.9} = 1.537084943$$

Using

$$a = \bar{y} - b\bar{x}$$

$$a = \frac{620}{8} - 1.537084943 \times \frac{315}{8} = 16.9777$$

Therefore

$$y = 17.0 + 1.5x \text{ (1.d.p.)}$$

d) When $x=0$ then $y=17.0$ and when $y=0$ $x=-17/1.5=11.333$ and draw line between these points.

e) Brand D as it is a long way above the line on the scatter graph. Take x to be 35 and put into the line of regression yields $y = 17.0 + (1.5 \times 35) = 69.5$

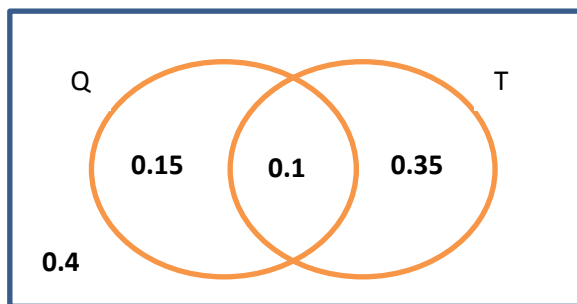
4. A survey of the reading habits of some students revealed that, on a regular basis, 25% read quality newspapers, 45% read tabloid newspapers and 40% do not read newspapers at all.

(a) Find the proportion of students who read both quality and tabloid newspapers. (3)

(b) In the space on page 13 draw a Venn diagram to represent this information. (3)

A student is selected at random. Given that this student reads newspapers on a regular basis,

(c) find the probability that this student only reads quality newspapers. (3)



4 a) As in total $45+40+25=110$ then 10% must read both Quality and Tabloid.

b) See diagram

c) Given they read a newspaper is 0.6 and ONLY quality is 0.15 therefore it is $\frac{0.15}{0.6} = \frac{1}{4} = 25\%$

5. Figure 2 shows a histogram for the variable t which represents the time taken, in minutes, by a group of people to swim 500m.

(a) Complete the frequency table for t .

T	5-10	10-14	14-18	18-25	25-40	Totals
Frequency	10	16	24	$7 \times 5 = 35$	$1 \times 15 = 15$	100
T(Midpts)	7.5	12	16	21.5	32.5	
Ft	75	192	384	752.5	487.5	1891
T ²	56.25	144	256	462.25	1056.25	
F T ²	562.5	2304	6144	16178.75	15843.75	41033

(2)

(b) Estimate the number of people who took longer than 20 minutes to swim 500m.

(2)

(c) Find an estimate of the mean time taken. (4)

(d) Find an estimate for the standard deviation of t . (3)

(e) Find the median and quartiles for t . (4)

One measure of skewness is found using

$$\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

(f) Evaluate this measure and describe the skewness of these data. (2)

a) The frequency is the area of the bar on the histogram. See calculations in table.

b) Use the histogram to work out frequency which is $(25 - 20) \times 5 + 15 = 40$

c) Mean is $\mu = \frac{\sum ft}{\sum f}$ $\frac{\sum ft}{\sum f} = \frac{1891}{100} = 18.91$

d) Using $\sigma^2 = \frac{\sum f t^2}{n} - \mu^2$ $\sigma^2 = \frac{41033}{100} - 18.91^2 = 52.7419$

$\sigma = 7.262$ (3. d. p)

e) Using

$$Q_2 = \text{lower class boundary} + \frac{\frac{n}{2} - \sum f(\text{before median class})}{f \text{ of median class}} \times \text{class width}$$

$Q_2 = 18$ (exactly between two classes)

For $Q_1 = 25\text{th value}$

$Q_1 = 10 + \frac{25 - 10}{16} \times 4 = 13.75$

For $Q_3 = 75\text{th value}$

$Q_3 = 18 + \frac{75 - 50}{35} \times 7 = 23$

f) $\frac{3(\text{mean} - \text{median})}{\text{standard deviation}} = \frac{3(18.91 - 18)}{7.262} = 0.376$ (3. d. p)

The data is positively skewed.

6. The random variable X has a normal distribution with mean 20 and standard deviation

4.

(a) Find $P(X > 25)$. (3)

(b) Find the value of d such that $P(20 < X < d) = 0.4641$ (4)

6 a)
$$P(x > 25) = 1 - P(X < 25)$$

$$= 1 - P\left(z < \frac{25 - 20}{4}\right) = 1 - P(z < 1.25)$$

$$= 1 - P(z < 1.25) = 1 - 0.8944 = 0.1056$$

b)
$$P(20 < X < d) = P(X < d) - P(X < 20) = 0.4641$$

 But 20 is the mean so
$$P(X < d) = 0.4641 + 0.5 = 0.9641$$

$$P(X < 20) = 0.5$$

Find z from tables.
$$P\left(z < \frac{d - 20}{4}\right) = 0.9641$$

$$\frac{d - 20}{4} = 1.8 \quad d = 27.2$$

7. The random variable X has probability distribution

x	1	3	5	7	9
P(X=x)	0.2	p	0.2	q	0.15

(a) Given that $E(X) = 4.5$, write down two equations involving p and q .

(3)

Find

(b) the value of p and the value of q ,

(3)

(c) $P(4 < X \leq 7)$.

(2)

Given that $E(X^2) = 27.4$, find

(d) $\text{Var}(X)$,

(2)

(e) $E(19 - 4X)$,

(1)

(f) $\text{Var}(19 - 4X)$.

(2)

7 a)
$$\sum P(X = x) = 1 \quad 0.2 + p + 0.2 + q + 0.15 = 1$$

$$p + q = 0.45 \quad (1)$$

$$E(X) = \sum xP(X = x) \quad 1 \times 0.2 + 3 \times p + 5 \times 0.2 + 7 \times q + 9 \times 0.15 = 4.5$$

$$0.2 + 3p + 1 + 7q + 1.35 = 4.5$$

	$3p + 7q = 4.5 - 0.2 - 1 - 1.35 = 1.95 \quad (2)$
b) Multiple (1) by 3	$3p + 3q = 1.35$
(2)-(1)	$4q = 0.6 \quad q = 0.15$
From (1)	$p + 0.15 = 0.45 \quad p = 0.3$
c) $P(4 < X \leq 7)$	$P(4 < X \leq 7) = P(5) + P(7)$
	$= 0.2 + q = 0.2 + 0.15 = 0.35$
d) Using	$Var(X) = 27.4 - (4.5)^2 = 7.15$
$Var(X) = E(X^2) - (E(X))^2$	
e) Using	$E(19 - 4X) = -4E(X) + 19 = 1$
$E(aX + b) = aE(X) + b$	
f) Using	$Var(19 - 4X) = 16Var(X) = 16 \times 7.15 = 114.4$
$Var(aX + b) = a^2Var(X)$	