AQA SS1B Statistics 1 Statistics 22 May 2014

Question Paper and Worked Solutions

Please note, this document represents my own solutions to the questions, is entirely unofficial and is not related to the mark scheme (which I have not seen). Therefore, while it should help you see how to do the questions, it won't include every valid method or give you a break down of the mark allocation. If you spot any errors, or think you have found a better solution, please email me so I can update it.

1 The weights, in kilograms, of a random sample of 15 items of cabin luggage on an aeroplane were as follows.

4.6 3.8 3.9 4.5 4.9 3.6 3.7 5.2 4.0 5.1 4.1 3.3 4.7 5.0 4.8

For these data:

(a) find values for the median and the interquartile range;

[4 marks]

(b) find the value for the range;

[1 mark]

(c) state why the mode is **not** an appropriate measure of average.

[1 mark]

1.

a)

In order: $3.3, 3.6, 3.7, 3.8, 3.9, 4, 4.1, 4.5, 4.6, 4.7, 4.8, 4.9, 5, 5.1, 5.2$

Median: $\frac{15+1}{2}^{th}$ number: $\mathbf{4.5kg}$.

Lower quartile: $\frac{15+1}{4}^{th}$ number: $3.8kg$, Upper quartile: $\frac{3(15+1)}{4}^{th}$ number: $4.9kg \implies IQR = 4.9 - 3.8 = \mathbf{1.1kg}$

b)

$$Range = max - min = 5.2 - 3.3 = \mathbf{1.9kg}$$

c)

Each number is different, so the data set is multi-modal and every single number is a modal average.

2 (a) Tim rings the church bell in his village every Sunday morning. The time that he spends ringing the bell may be modelled by a normal distribution with mean 7.5 minutes and standard deviation 1.6 minutes.

Determine the probability that, on a particular Sunday morning, the time that Tim spends ringing the bell is:

(i) at most 10 minutes;

(ii) more than 6 minutes;

(iii) between 5 minutes and 10 minutes.

[6 marks]

(b) June rings the same church bell for weekday weddings. The time that she spends, in minutes, ringing the bell may be modelled by the distribution $\mathrm{N}(\mu, 2.4^2)$.

Given that 80 per cent of the times that she spends ringing the bell are less than 15 minutes, find the value of μ.

[4 marks]

2.

a)

i.

$$X \sim N(7.5, 1.6^2) \quad \Rightarrow \quad P(x < 10) = P\left(z < \frac{10 - 7.5}{1.6}\right) = \Phi(1.5625) = 0.94062 = \mathbf{0.941\ to\ 3\ s.f.}$$

ii.

$$P(x > 6) = P\left(z > \frac{6 - 7.5}{1.6}\right) = P(z > -0.9375) = 1 - P(z < 0.9375) = 1 - \Phi(0.9375) = \mathbf{0.933\ to\ 3\ s.f.}$$

iii.

Note that 10 and 5 are symmetrical about the mean, 7.5. Therefore $P(5 < x < 10) = 2(P(x < 10) - 0.5)$

$$2(P(x < 10) - 0.5) = 2(0.94062 - 0.5) = 0.88124 = \mathbf{0.881\ to\ 3\ s.f.}$$

b)

$$Y \sim N(\mu, 2.4^2)$$

$$P(y < 15) = 0.8 \quad \Rightarrow \quad P\left(z < \frac{15 - \mu}{2.4}\right) = 0.8 \quad \Rightarrow \quad \Phi\left(\frac{15 - \mu}{2.4}\right) = 0.8 \quad \Rightarrow \quad \frac{15 - \mu}{2.4} = 0.8416$$

$$\Rightarrow \quad 15 - \mu = 2.01984 \quad \Rightarrow \quad \mu = 12.9802 = \mathbf{13.0\ minutes\ to\ 3\ s.f.}$$

3 The table shows the body mass index (BMI), x, and the systolic blood pressure (SBP), y mmHg, for each of a random sample of 10 men, aged between 35 years and 40 years, from a particular population.

x	13	23	29	35	17	34	25	20	31	27
y	103	115	124	126	108	120	113	117	118	119

(a) Calculate the equation of the least squares regression line of y on x.

[4 marks]

(b) Use your equation to estimate the SBP of a man from this population who is aged 38 years and who has a BMI of 30.

[2 marks]

(c) State why your equation might **not** be appropriate for estimating the SBP of a man from this population:

(i) who is aged 38 years and who has a BMI of 45;

(ii) who is aged 50 years and who has a BMI of 25.

[2 marks]

(d) Find the value of the residual for the point (20, 117).

[2 marks]

(e) The mean of the vertical distances of the 10 points from the regression line calculated in part **(a)** is 2.71, correct to three significant figures.

Comment on the likely accuracy of your estimate in part **(b)**.

[1 mark]

3.

a)

$$y = a + bx \quad where \quad a = \bar{y} - b\bar{x} \quad and \quad b = \frac{S_{xy}}{S_{xx}}$$

$$S_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 29942 - \frac{254 \times 1163}{10} = 401.8 \quad and \quad S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 6924 - \frac{254^2}{10} = 472.4$$

$$b = \frac{401.8}{472.4} = 0.85055 \quad \bar{x} = \frac{\sum x}{n} = \frac{254}{10} = 25.4 \quad and \quad \bar{y} = \frac{\sum y}{n} = \frac{1163}{10} = 116.3$$

$$\Rightarrow \quad a = 116.3 - 0.85055(25.4) = 94.69603 \quad \Rightarrow \quad \boldsymbol{y = 94.69603 + 0.85055x}$$

b)

$$x = 30 \quad \Rightarrow \quad y = 94.69603 + 0.85055(30) = \mathbf{120.21253}$$

c)

i.

$x = 45$ is beyond the range of the sampled data. Extrapolation makes for less reliable predictions than interpolation.

ii.

The data set was for men of age 35 to 40. An age of 50 is beyond the range sampled.

d)

$$x = 20 \quad \Rightarrow \quad y = 94.69603 + 0.85055(20) = 111.70703 \quad \Rightarrow \quad residual = 117 - 111.70703 = \mathbf{5.29297}$$

e)

My estimate, of 120.21253, is likely to be correct to within 2.71. That is, 120.2 ± 2.71.

4 Alf and Mabel are members of a bowls club and sometimes attend the club's social events.

The probability, $\mathrm{P}(A)$, that Alf attends a social event is 0.70.
The probability, $\mathrm{P}(M)$, that Mabel attends a social event is 0.55.
The probability, $\mathrm{P}(A \cap M)$, that both Alf and Mabel attend the same social event is 0.45.

(a) Find the probability that:

(i) either Alf or Mabel or both attend a particular social event;

(ii) either Alf or Mabel but **not both** attend a particular social event.

[3 marks]

(b) Give a numerical justification for the following statement.

"Events A and M are **not** independent."

[2 marks]

(c) Ben and Nora are also members of the bowls club and sometimes attend the club's social events.

The probability, $\mathrm{P}(B)$, that Ben attends a social event is 0.85.
The probability, $\mathrm{P}(N)$, that Nora attends a social event is 0.65.
The attendance of each of Ben and Nora at a social event is independent of the attendance of all other members.

Find the probability that:

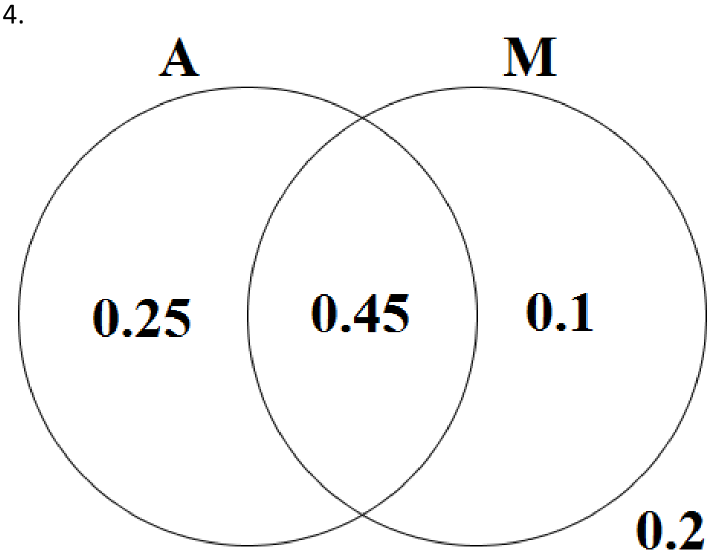
(i) all four named members attend a particular social event;

[2 marks]

(ii) none of the four named members attend a particular social event.

[3 marks]

4.



a)
i.

$$P(A \cup M) = P(A) + P(M) - P(A \cap M)$$

$$= 0.7 + 0.55 - 0.45 = \mathbf{0.8}$$

ii.

$$0.8 - 0.45 = \mathbf{0.35}$$

b)
If events A and M are independent, then:

$$P(A|M) = P(A) \quad and \quad P(M|A) = P(M)$$

But:

$$P(A|M) = \frac{0.45}{0.55} = 0.\dot{8}\dot{1} \neq P(A) = 0.7$$

$$P(M|A) = \frac{0.45}{0.7} = 0.6\dot{4}28571\dot{1} \neq P(M) = 0.55$$

Therefore not independent.

c)

$$P(A \cap M) = 0.45 \quad \Rightarrow \quad P(A \cap M \cap B \cap N) = 0.45 \times 0.85 \times 0.65 = \mathbf{0.248625}$$

d)

$$P(\neg A \cap \neg M \cap \neg B \cap \neg N) = 0.2 \times 0.15 \times 0.35 = \mathbf{0.0105}$$

5 As part of a study of charity shops in a small market town, two such shops, X and Y, were each asked to provide details of its takings on 12 randomly selected days.

The table shows, for each of the 12 days, the day's takings, £x, of charity shop X and the day's takings, £y, of charity shop Y.

Day	A	B	C	D	E	F	G	H	I	J	K	L
x	46	57	39	116	62	77	41	61	15	53	68	61
y	78	102	66	214	98	72	98	134	21	67	95	83

(a) (i) Calculate the value of the product moment correlation coefficient between x and y. **[3 marks]**

(ii) Interpret your value in the context of this question. **[2 marks]**

(b) Complete the scatter diagram shown on the opposite page. **[2 marks]**

(c) The investigator realised subsequently that one of the 12 selected days was a particularly popular town market day and another was a day on which the weather was extremely severe.

Identify **each** of these days giving a reason for each choice. **[3 marks]**

(d) Removing the two days described in part **(c)** from the data gives the following information.

$$S_{xx} = 1292.5 \qquad S_{yy} = 3850.1 \qquad S_{xy} = 407.5$$

(i) Use this information to recalculate the value of the product moment correlation coefficient between x and y.

(ii) Hence revise, as necessary, your interpretation in part **(a)(ii)**. **[3 marks]**

5.

a)

i.

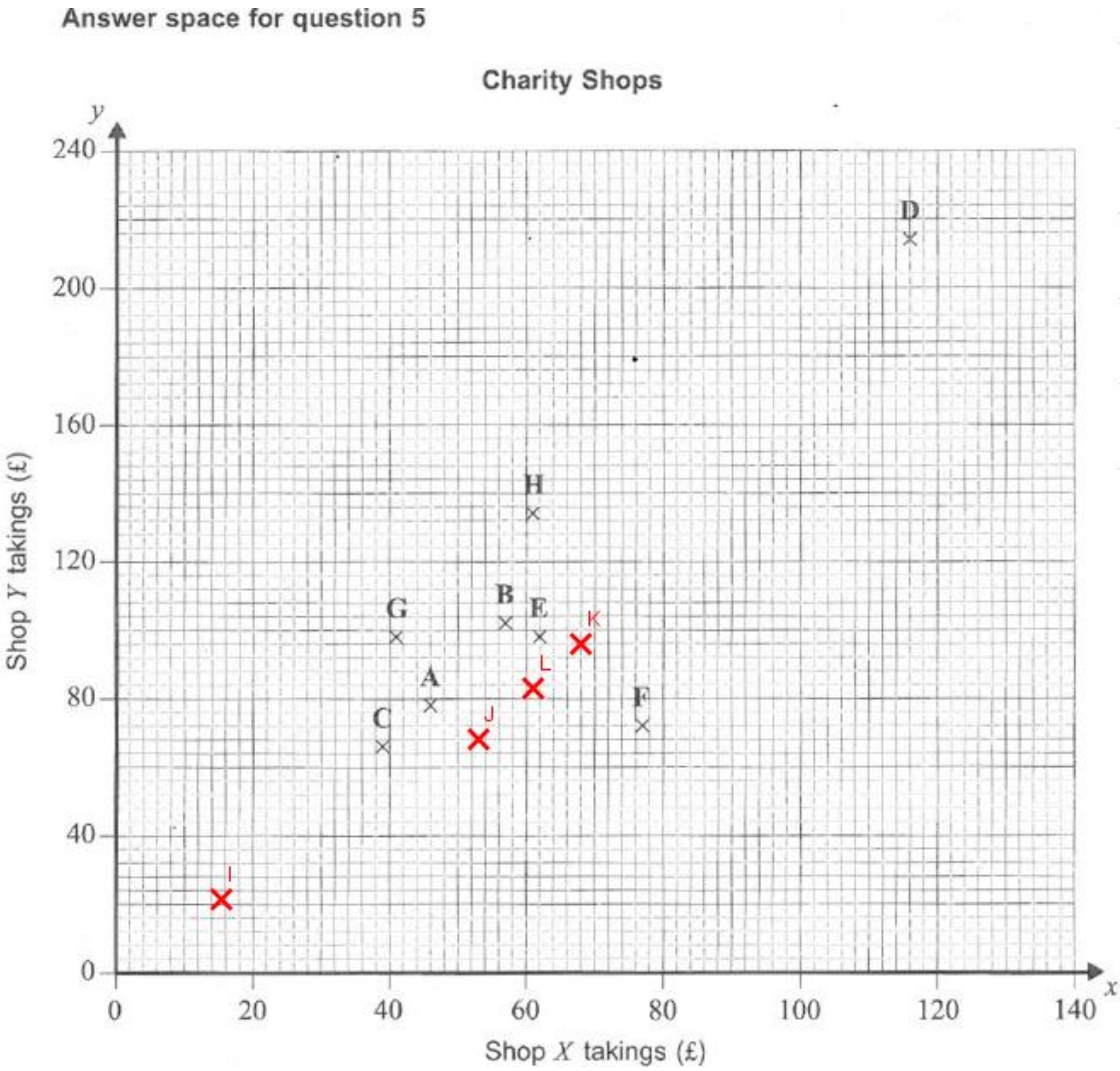
$$S_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 76001 - \frac{696 \times 1128}{12} = 10577 \qquad S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 46896 - \frac{696^2}{12} = 6528$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 129832 - \frac{1128^2}{12} = 23800 \qquad r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{10577}{\sqrt{6528 \times 23800}} = \mathbf{0.84856\ \mathit{to}\ 5\ \mathit{d.p.}}$$

ii.

There is a strong positive correlation between the takings of shop X and the takings of shop Y. Therefore, in general, the more money shop X takes in, the more money shop Y will take in.

b)



c)

D is likely to represent takings on market day, since takings were much higher for both shops.

I is likely to represent takings on the day with the severe weather, since takings were much lower for both shops.

d)

i.

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{407.5}{\sqrt{1292.5 \times 3850.1}} = \mathbf{0.18267\ to\ 5\ d.p.}$$

ii.

This number suggests a weak positive correlation, meaning that while the takings from the two shops may be related, there is no evidence to suggest a strong link between them. Takings from shop X and from shop Y are not closely correlated.

6 The probability that an online order from a supermarket chain has at least one item missing when delivered is 0.06.

Online orders are 'incomplete' if they contain substitute items and/or have at least one item missing when delivered. The probability that an order is incomplete is 0.15.

(a) Calculate the probability that exactly 2 out of a random sample of 26 online orders have at least one item missing when delivered.

[3 marks]

(b) Determine the probability that the number of incomplete orders in a random sample of 50 online orders is:

(i) fewer than 10;

(ii) more than 5;

(iii) more than 6 but fewer than 12.

[6 marks]

(c) Farokh, the manager of one of the supermarket's stores, examines 50 randomly selected online orders from each of a random sample of 100 of the store's customers. He records, for each of the 50 orders, the number, x, that were incomplete.

His summarised results, correct to three significant figures, for the 100 customers selected are

$$\bar{x} = 4.33 \quad \text{and} \quad s^2 = 3.94$$

Use this information to compare the performance of the store managed by Farokh with that of the supermarket chain as a whole.

[5 marks]

6.

a)

$$X \sim B(26, 0.06) \quad \Rightarrow \quad P(x = 2) = \binom{26}{2} \times 0.06^2 \times 0.94^{24} = \mathbf{0.26501\ \textit{to}\ 5\ \textit{d.p.}}$$

b)

i.

$$Y \sim B(50, 0.15) \quad \Rightarrow \quad P(y < 10) = P(y \le 9) = \mathbf{0.7911}$$

ii.

$$P(y > 5) = 1 - P(y \le 5) = \mathbf{0.2194}$$

iii.

$$P(6 < y < 12) = P(y \le 11) - P(y \le 6) = 0.9372 - 0.3613 = \mathbf{0.5759}$$

c)

Population: $\mu = np = 50 \times 0.15 = 7.5 \qquad \sigma^2 = np(1-p) = 6.375$

Sample: $\bar{x} = 4.33 \qquad s^2 = 3.94$. Comparing these to the population parameters tells me that this store has not only a lower incidence of incomplete orders, but also less variation than the supermarket chain as a whole.

7 For the year 2014, the table below summarises the weights, x kilograms, of a random sample of 160 women residing in a particular city who are aged between 18 years and 25 years.

Weight (x kg)	Number of women
35–40	4
40–45	9
45–50	12
50–55	16
55–60	24
60–65	28
65–70	24
70–75	17
75–80	12
80–85	7
85–90	4
90–95	2
95–100	1
Total	**160**

(a) Calculate estimates of the mean and the standard deviation of these 160 weights. **[4 marks]**

(b) (i) Construct a 98% confidence interval for the mean weight of women residing in the city who are aged between 18 years and 25 years. **[5 marks]**

(ii) Hence comment on a claim that the mean weight of women residing in the city who are aged between 18 years and 25 years has increased from that of 61.7 kg in 1965. **[2 marks]**

7.
a)
Using midpoints for each class to use as x, and using the sample standard deviation formula:

$$s^2 = \frac{\sum f}{\sum f - 1}\left(\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2\right) = \left(\frac{160}{159}\right)\left(\frac{657450}{160} - \left(\frac{10065}{160}\right)^2\right) = 152.8\ldots \quad \Rightarrow \quad \boldsymbol{s = 12.36\ to\ 2\ d.p.}$$

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{10065}{160} = \mathbf{62.90625}$$

b)
i.
By the central limit theorem, means are distributed with mean $\bar{x}$ and standard deviation $\frac{s}{\sqrt{n}} = \frac{12.36}{\sqrt{160}} \approx 0.97714$

$$P(x < a) = p\left(z < \frac{a - 62.90625}{0.97714\ldots}\right) = 0.99 \quad \Rightarrow \quad \frac{a - 62.90625}{0.97714\ldots} = 2.3263 \quad \Rightarrow \quad a = 65.179\ to\ 3\ d.p.$$

By symmetry, confidence interval: $98\%\ of\ weights\ lie\ within\ the\ interval$: $60.63 < x < 65.18$
ii.
$61.7kg$ lies within the confidence interval since $60.63 < 61.7 < 65.18$, therefore is not statistically significant for a 98% confidence interval. Our sample data is insufficient to conclude that the mean of the population has risen.

Produced by A. Clohesy; TheChalkface.net