**AQA MS1B Statistics 1 Mathematics 06 June 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** Henrietta lives on a small farm where she keeps some hens.

For a period of 35 weeks during the hens' first laying season, she records, **each week**, the total number of eggs laid by the hens.

Her records are shown in the table.

| **Total number of eggs laid in a week ($x$)** | **Number of weeks ($f$)** |
|---|---|
| 66 | 1 |
| 67 | 2 |
| 68 | 3 |
| 69 | 5 |
| 70 | 7 |
| 71 | 8 |
| 72 | 4 |
| 73 | 2 |
| 74 | 2 |
| 75 | 1 |
| **Total** | **35** |

**(a)** For these data:

**(i)** state values for the mode and the range;

**[2 marks]**

**(ii)** find values for the median and the interquartile range;

**[3 marks]**

**(iii)** calculate values for the mean and the standard deviation.

**[4 marks]**

**(b)** Each week, for the 35 weeks, Henrietta sells 60 eggs to a local shop, keeping the remainder for her own use.

State values for the mean and the standard deviation of the number of eggs that she keeps.

**[2 marks]**

1.

a)

i.

Mode: **71**

Range: $75 - 66 = \mathbf{9}$

ii.

| x | f | cumulative freq |
|---|---|---|
| 66 | 1 | **1** |
| 67 | 2 | **3** |
| 68 | 3 | **6** |
| 69 | 5 | **11** |
| 70 | 7 | **18** |
| 71 | 8 | **26** |
| 72 | 4 | **30** |
| 73 | 2 | **32** |
| 74 | 2 | **34** |
| 75 | 1 | **35** |

Median: $\frac{35+1}{2}^{th}$ value: **70**.

Lower quartile: $\frac{35+1}{4}^{th}$ value: 69 Upper quartile: $\frac{3(35+1)}{4}^{th}$ value: 72 $\Longrightarrow$ $IQ\ range = 72 - 69 = \mathbf{3}$

iii.

| x | f | *cumulative freq* | xf |
|---|---|---|---|
| 66 | 1 | *1* | **66** |
| 67 | 2 | *3* | **134** |
| 68 | 3 | *6* | **204** |
| 69 | 5 | *11* | **345** |
| 70 | 7 | *18* | **490** |
| 71 | 8 | *26* | **568** |
| 72 | 4 | *30* | **288** |
| 73 | 2 | *32* | **146** |
| 74 | 2 | *34* | **148** |
| 75 | 1 | *35* | **75** |
| Totals: | **35** | | **2464** |

$$\mu = \frac{\sum fx}{\sum f} = \frac{2464}{35} = \mathbf{70.4}$$

$$\sigma^2 = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2 = \frac{173610}{35} - (70.4)^2 = 4.1257\ldots \quad \Longrightarrow \quad \boldsymbol{\sigma = 2.03\ to\ 2\ d.p.}$$

b)

$$Mean = \mu - 60 = \mathbf{10.4} \qquad Standard\ deviation = \sigma = \mathbf{2.03}$$

**2** A garden centre sells bamboo canes of nominal length $1.8$ metres. The length, $X$ metres, of the canes can be modelled by a normal distribution with mean $1.86$ and standard deviation $\sigma$.

**(a)** Assuming that $\sigma = 0.04$, determine:

**(i)** $\mathrm{P}(X < 1.90)$;

**(ii)** $\mathrm{P}(X > 1.80)$;

**(iii)** $\mathrm{P}(1.80 < X < 1.90)$;

**(iv)** $\mathrm{P}(X \neq 1.86)$.

**[7 marks]**

**(b)** It is subsequently found that $\mathrm{P}(X > 1.80) = 0.98$.

Determine the value of $\sigma$.

**[3 marks]**

2.

a)

i.

$$X \sim N(1.86, 0.04^2) \quad \Rightarrow \quad P(X < 1.90) = P\left(Z < \frac{1.90 - 1.86}{0.04}\right) = \Phi(1) = \mathbf{0.84134}$$

ii.

$$P(X > 1.80) = P\left(Z > \frac{1.80 - 1.86}{0.04}\right) = P(Z > -1.5) = P(Z < 1.5) = \Phi(1.5) = \mathbf{0.93319}$$

iii.

$$P(1.80 < X < 1.90) = P(X < 1.90) - P(X < 1.80) = 0.84134 - (1 - 0.93319) = \mathbf{0.77453}$$

iv.

$$P(X \neq 1.86) = 1 - P(X = 1.86) = 1 - 0 = \mathbf{1}$$

b)

$$P(X > 1.80) = 0.98 \quad \Rightarrow \quad P\left(Z < -\frac{1.80 - 1.86}{\sigma}\right) = 0.98 \quad \Rightarrow \quad \frac{0.06}{\sigma} = 2.0537 \quad \Rightarrow \quad \boldsymbol{\sigma = 0.0292 \text{ to } 3\ s.f.}$$

**3** The table shows the colour of hair and the colour of eyes of a sample of 750 people from a particular population.

| | | Colour of hair | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Black** | **Dark** | **Medium** | **Fair** | **Auburn** | **Total** |
| **Colour of eyes** | **Blue** | 6 | 51 | 68 | 66 | 24 | 215 |
| | **Brown** | 14 | 92 | 97 | 90 | 47 | 340 |
| | **Green** | 0 | 37 | 55 | 64 | 39 | 195 |
| | **Total** | 20 | 180 | 220 | 220 | 110 | 750 |

**(a)** Calculate, **to three decimal places**, the probability that a person, selected at random from this sample, has:

**(i)** fair hair;

**(ii)** auburn hair and blue eyes;

**(iii)** either auburn hair or blue eyes but not both;

**(iv)** green eyes, given that the person has fair hair;

**(v)** fair hair, given that the person has green eyes.

**[8 marks]**

**(b)** Three people are selected at random from the sample.

Calculate, **to three significant figures**, the probability that two of them have dark hair and brown eyes and the other has medium hair and green eyes.

**[4 marks]**

3.

a)

i.

$$\frac{220}{750} = \mathbf{0.293\ to\ 3\ d.p.}$$

ii.

$$\frac{24}{750} = \mathbf{0.032\ to\ 3\ d.p.}$$

iii.

$$\frac{110}{750} + \frac{215}{750} - \frac{24}{750} = \frac{301}{750} = \mathbf{0.401\ to\ 3\ d.p.}$$

iv.

$$\frac{64}{220} = \mathbf{0.291\ to\ 3\ d.p.}$$

v.

$$\frac{64}{195} = \mathbf{0.328\ to\ 3\ d.p.}$$

b)

$$P(Dark\ hair\ and\ brown\ eyes) = \frac{92}{750} \qquad P(Medium\ hair\ and\ green\ eyes) = \frac{55}{750}$$

$$\binom{3}{1}\left(\frac{92}{750}\right)^2 \times \frac{55}{750} = \frac{1396560}{421875000} = \mathbf{0.00331\ to\ 3\ s.f.}$$

**4** Every year, usually during early June, the Isle of Man hosts motorbike races. Each race consists of three consecutive laps of the island's course. To compete in a race, a rider must first complete at least one qualifying lap.

The data refer to the lightweight motorbike class in 2012 and show, for each of a random sample of 10 riders, values of

$$u = x - 100 \quad \text{and} \quad v = y - 100$$

where

$x$ denotes the average speed, in mph, for the rider's fastest qualifying lap
and
$y$ denotes the average speed, in mph, for the rider's three laps of the race.

| | Rider | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I | J |
| $u$ | 7.88 | 13.02 | 4.29 | 2.88 | 6.26 | 7.03 | 3.60 | 11.78 | 13.15 | 11.69 |
| $v$ | 6.63 | 10.16 | 3.63 | 0.47 | 5.70 | 8.01 | 3.30 | 7.31 | 13.08 | 11.82 |

**(a) (i)** Calculate the value of $r_{uv}$, the product moment correlation coefficient between $u$ and $v$. **[3 marks]**

**(ii)** Hence state the value of $r_{xy}$, giving a reason for your answer. **[2 marks]**

**(b)** Interpret your value of $r_{xy}$ in the context of this question. **[2 marks]**

4.
a)
i.

| | A | B | C | D | E | F | G | H | I | J | totals |
|---|---|---|---|---|---|---|---|---|---|---|---|
| u | 7.88 | 13.02 | 4.29 | 2.88 | 6.26 | 7.03 | 3.6 | 11.78 | 13.15 | 11.69 | **81.58** |
| v | 6.63 | 10.16 | 3.63 | 0.47 | 5.7 | 8.01 | 3.3 | 7.31 | 13.08 | 11.82 | **70.11** |
| u² | 62.0944 | 169.5204 | 18.4041 | 8.2944 | 39.1876 | 49.4209 | 12.96 | 138.7684 | 172.9225 | 136.6561 | **808.2288** |
| v² | 43.9569 | 103.2256 | 13.1769 | 0.2209 | 32.49 | 64.1601 | 10.89 | 53.4361 | 171.0864 | 139.7124 | **632.3553** |
| *uv* | *52.2444* | *132.2832* | *15.5727* | *1.3536* | *35.682* | *56.3103* | *11.88* | *86.1118* | *172.002* | *138.1758* | ***701.6158*** |

$$S_{uv} = \sum uv - \frac{\sum u \sum v}{n} = 701.6158 - \frac{81.58 \times 70.11}{10} = 129.65842$$
$$S_{uu} = \sum u^2 - \frac{(\sum u)^2}{n} = 808.2288 - \frac{81.58^2}{10} = 142.69916$$
$$S_{vv} = \sum v^2 - \frac{(\sum v)^2}{n} = 632.3553 - \frac{70.11^2}{10} = 140.81409$$
$$r_{uv} = \frac{S_{uv}}{\sqrt{S_{uu}S_{vv}}} = \frac{129.65842}{\sqrt{142.69916 \times 140.81409}} = \mathbf{0.91468\ to\ 5\ d.p.}$$

ii.

$$r_{xy} = r_{uv} = \mathbf{0.91468\ to\ 5\ d.p.}$$

$u$ and $v$ are generated from $x$ and $y$ by shifting the points down 100 and left 100. This is a linear transformation, and therefore the correlation is unchanged.

b)

There is a strong positive correlation between the average speed of the fastest qualifying lap and the average speed of the three race laps. That is, the faster a rider is on the qualifying lap, the faster he is likely to be for the race.

**5** An analysis of the number of vehicles registered by each household within a city resulted in the following information.

| Number of vehicles registered | 0 | 1 | 2 | $\geqslant 3$ |
|---|---|---|---|---|
| Percentage of households | 18 | 47 | 25 | 10 |

**(a)** A random sample of 30 households within the city is selected.

Use a binomial distribution with $n = 30$, together with relevant information from the table in each case, to find the probability that the sample contains:

**(i)** exactly 3 households with **no** registered vehicles;

**[3 marks]**

**(ii)** at most 5 households with **three or more** registered vehicles;

**[2 marks]**

**(iii)** more than 10 households with **at least two** registered vehicles;

**[3 marks]**

**(iv)** more than 5 households but fewer than 10 households with **exactly two** registered vehicles.

**[3 marks]**

**(b)** If a random sample of **150** households within the city were to be selected, estimate the mean and the variance for the number of households in the sample that would have **either one or two** registered vehicles.

**[2 marks]**

5.

a)

i.

Using the first column:

$$X \sim B(30, 0.18) \quad \Rightarrow \quad P(X = 3) = \binom{30}{3} 0.18^3 (1 - 0.18)^{27} = \mathbf{0.1115\ to\ 4\ d.p.}$$

ii.

$$X_1 \sim B(30, 0.1) \quad \Rightarrow \quad P(X_1 \leq 5) = \mathbf{0.9268}$$

iii.

$$X_2 \sim B(30, 0.35) \quad \Rightarrow \quad P(X_2 > 10) = 1 - P(X_2 \leq 10) = 1 - 0.5078 = \mathbf{0.4922}$$

iv.

$$X_3 \sim B(30, 0.25) \quad \Rightarrow \quad P(5 < X_3 < 10) = P(X_3 \leq 9) - P(X_3 \leq 5) = 0.1298 - 0.1047 = \mathbf{0.0251}$$

b)

$$Y \sim B(150, 0.72) \quad \Rightarrow \quad mean = np = 150 \times 0.72 = \mathbf{108}$$

$$variance = np(1 - p) = 150 \times 0.72 \times 0.28 = \mathbf{30.24}$$

**6** A rubber seal is fitted to the bottom of a flood barrier. When no pressure is applied, the depth of the seal is $15$ cm. When pressure is applied, a watertight seal is created between the flood barrier and the ground.

The table shows the pressure, $x$ kilopascals (kPa), applied to the seal and the resultant depth, $y$ centimetres, of the seal.

| $x$ | 25 | 50 | 75 | 100 | 125 | 150 | 175 | 200 | 250 | 300 |
|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 14.7 | 13.4 | 12.8 | 11.9 | 11.0 | 10.3 | 9.7 | 9.0 | 7.5 | 6.7 |

**(a) (i)** State the value that you would **expect** for $a$ in the equation of the least squares regression line, $y = a + bx$.

**[1 mark]**

**(ii)** Calculate the equation of the least squares regression line, $y = a + bx$.

**[4 marks]**

**(iii)** Interpret, in context, your value for $b$.

**[2 marks]**

**(b)** Calculate an estimate of the depth of the seal when it is subjected to a pressure of $225$ kPa.

**[2 marks]**

**(c) (i)** Give a statistical reason as to why your equation is unlikely to give a realistic estimate of the depth of the seal if it were to be subjected to a pressure of $400$ kPa.

**(ii)** Give a reason based on the context of this question as to why your equation will not give a realistic estimate of the depth of the seal if it were to be subjected to a pressure of $525$ kPa.

**[3 marks]**

6.
a)
i.
$a = 15$ since when no pressure is applied, the depth is $15cm$.
ii.

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

$$S_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 13490 - \frac{1450 \times 107}{10} = -2025 \qquad S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 280000 - \frac{1450^2}{10} = 69750$$

$$b = \frac{-2025}{69750} = -0.02903 \quad \bar{x} = \frac{\sum x}{n} = \frac{1450}{10} = 145 \quad and \quad \bar{y} = \frac{\sum y}{n} = \frac{107}{10} = 10.7$$

$$\Rightarrow \quad a = 10.7 - (-0.02903)(145) = 14.90968 \quad \Rightarrow \quad \boldsymbol{y = 14.90968 - 0.02903x}$$

iii.
The seal reduces in depth by $0.02903cm$ for every $1KPa$ of pressure applied.
b)

$$y = 14.90968 - 0.02903 \times 225 = \mathbf{8.377419cm}$$

c)
i.
$400KPa$ is outside the range of the observed data. Extrapolation is unlikely to be as reliable as interpolation as a predictor of data.
ii.
Substituting $y = 525$ into our regression line gives $x = -0.33226$ which would be a negative depth (not possible).

**7** The volume of water, $V$, used by a guest in an en suite shower room at a small guest house may be modelled by a random variable with mean $\mu$ litres and standard deviation 65 litres.

A random sample of 80 guests using this shower room showed a mean usage of 118 litres of water.

**(a) (i)** Give a numerical justification as to why $V$ is unlikely to be normally distributed.

**[2 marks]**

**(ii)** Explain why $\overline{V}$, the mean of a random sample of 80 observations of $V$, may be assumed to be approximately normally distributed.

**[2 marks]**

**(b) (i)** Construct a 98% confidence interval for $\mu$.

**[4 marks]**

**(ii)** Hence comment on a claim that $\mu$ is 140.

**[2 marks]**

7.
a)
i.

$$V \sim N(118, 65^2) \quad \Rightarrow \quad P(V < 0) = P\left(Z < \frac{0-118}{65}\right) = 1 - \mathrm{P}\left(Z < \frac{118}{65}\right) = 1 - \Phi(1.8154) = 0.03438$$

This suggests that there is a 3.4% chance that a guest will use a negative quantity of water, which is impossible.

ii.

By the central limit theorem, the mean of a sufficiently large sample taken from any distribution may be assumed to be normally distributed with mean $\mu$ and standard deviation $\frac{\sigma}{\sqrt{n}}$. The negative quantity issue will not be a problem because the chance of being negative with this distribution is vanishingly small:

$$\bar{V} \sim N\left(118, \frac{65^2}{80}\right) \quad \Rightarrow \quad P(\bar{V} < 0) = P\left(Z < \frac{0-118}{\frac{65}{\sqrt{80}}}\right) = 1 - P\left(Z < \frac{118}{\frac{65}{\sqrt{80}}}\right) = 1 - \Phi(16.24) \approx 0$$

b)
i.

$$P(\bar{V} < U) = 0.99 \quad \Rightarrow \quad P\left(Z < \frac{U-118}{\frac{65}{\sqrt{80}}}\right) = 0.99 \quad \Rightarrow \quad \frac{U-118}{\frac{65}{\sqrt{80}}} = 2.3263 \quad \Rightarrow \quad U = 134.91$$

$$L = 118 - (134.91 - 118) = 101.09 \quad \Rightarrow \quad \mathbf{101.09 < \bar{V} < 134.91}$$

ii.

The chance of the population mean being above 134.91 is less than 1%, so it is very unlikely to be 140 $litres$.