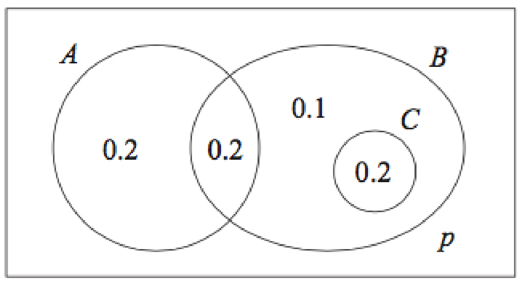
# Dr Oliver Mathematics
# Advanced Subsidiary Paper 21: Statistics
# November 2021: Calculator
# 1 hour 15 minutes

The total number of marks available is 30.
You must write down all the stages in your working.
Inexact answers should be given to three significant figures unless otherwise stated.
(It goes with Paper 22: Mechanics)

1. The Venn diagram, where $p$ is a probability, shows the 3 events $A$, $B$, and $C$ with their associated probabilities.




(a) Find the value of $p$. (1)

**Solution**

$$0.2 + 0.2 + 0.1 + 0.2 + p = 1 \Rightarrow 0.7 + p = 1$$
$$\Rightarrow \underline{\underline{p = 0.3}}.$$

(b) Write down a pair of mutually exclusive events from $A$, $B$, and $C$. (1)

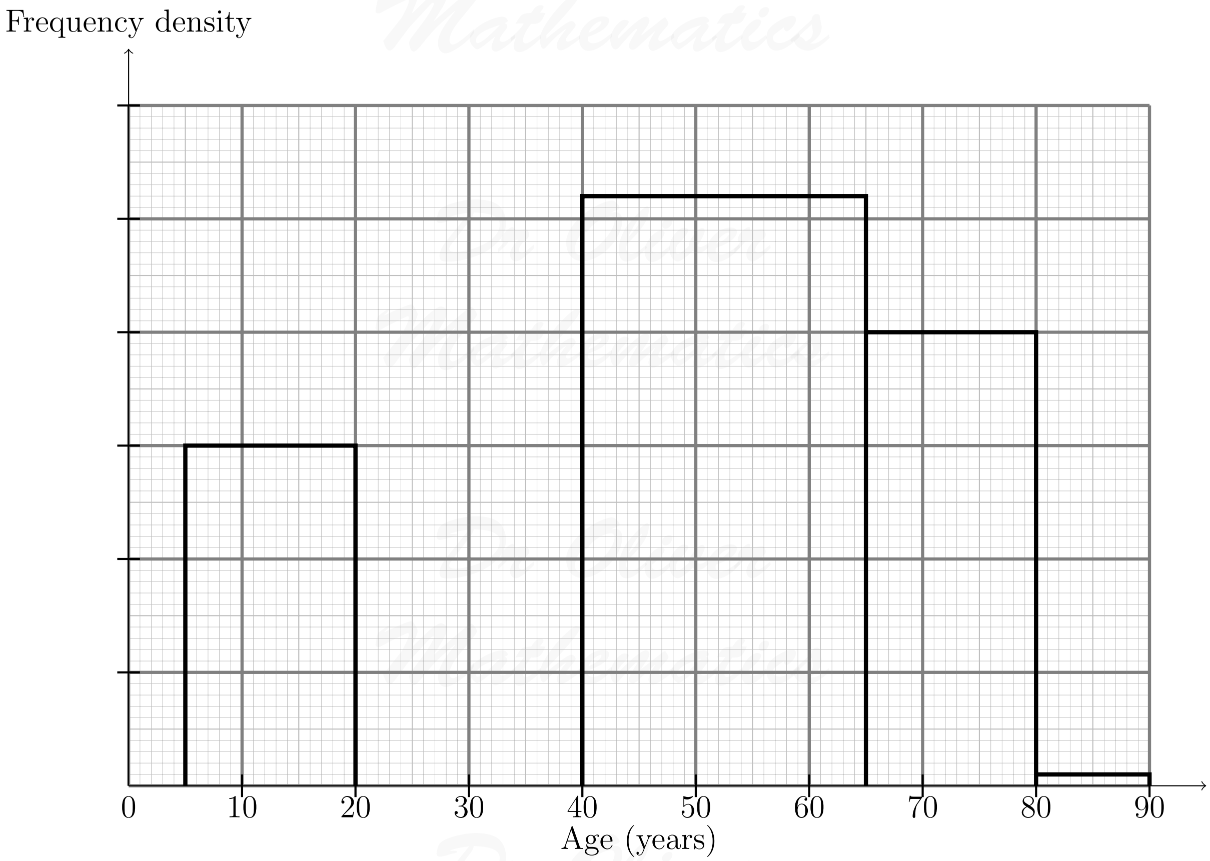
**Solution**

$\underline{\underline{A \text{ and } C}}$.

2. The partially completed table and partially completed histogram give information about the ages of passengers on an airline.

There were no passengers aged 90 or over.

| Age ($x$ years) | Frequency |
|---|---|
| $0 \leqslant x < 5$ | 5 |
| $5 \leqslant x < 20$ | 45 |
| $20 \leqslant x < 40$ | 90 |
| $40 \leqslant x < 65$ | |
| $65 \leqslant x < 80$ | |
| $80 \leqslant x < 90$ | 1 |



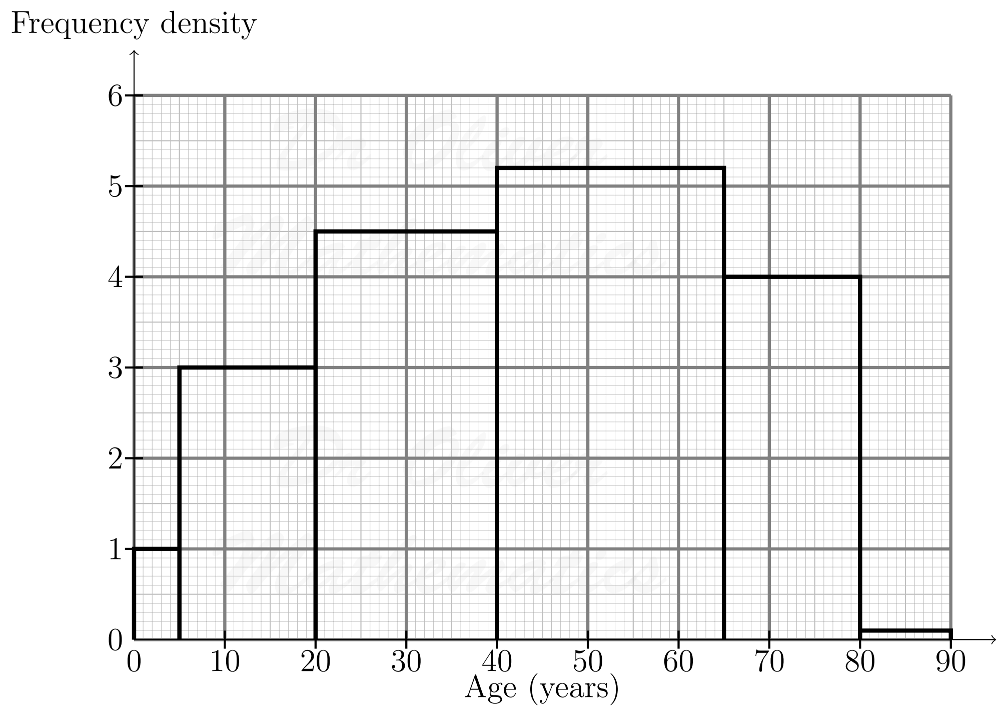

(a) Complete the histogram. (3)

**Solution**

We complete the table:

| Age ($x$ years) | Frequency | Width | Frequency Density |
|---|---|---|---|
| $0 \leqslant x < 5$ | 5 | 5 | $\frac{5}{5} = 1$ |
| $5 \leqslant x < 20$ | 45 | 15 | $\frac{45}{15} = 3$ |
| $20 \leqslant x < 40$ | 90 | 20 | $\frac{90}{20} = 4.5$ |
| $40 \leqslant x < 65$ | $25 \times 5.2 = 130$ | 25 | 5.2 |
| $65 \leqslant x < 80$ | $15 \times 4 = 60$ | 15 | 4 |
| $80 \leqslant x < 90$ | 1 | 10 | $\frac{1}{10}$ |

And now we complete the histogram:




(b) Use linear interpolation to estimate the median age. (4)

**Solution**

The total frequency is 331. The median is in the

$$\tfrac{1}{2}(331 + 1) = 166\text{th piece},$$

i.e., the 26nd of the $40 \leqslant x < 65$ interval. So, by linear interpolation,

$$\begin{aligned}\text{median} &= 40 + \tfrac{26}{130} \times 25 \\ &= \underline{\underline{45}}.\end{aligned}$$

An outlier is defined as a value greater than

$$Q_3 + 1.5 \times \text{interquartile range}.$$

Given that $Q_1 = 27.3$ and $Q_3 = 58.9$,

(c) determine, giving a reason, whether or not the oldest passenger could be considered as an outlier. (2)

**Solution**
Well,

$$\begin{aligned} Q_3 + 1.5 \times \text{interquartile range} &= 58.9 + 1.5(58.9 - 27.3) \\ &= 106.3; \end{aligned}$$

so, no, the oldest passenger cannot be considered as an outlier.

3. Helen is studying one of the qualitative variables from the large data set for Heathrow from 2015.

She started with the data from 3rd May and then took every 10th reading.

There were only 3 different outcomes with the following frequencies

| Outcome | $A$ | $B$ | $C$ |
|---|---|---|---|
| Frequency | 16 | 2 | 1 |

(a) State the sampling technique Helen used. (1)

**Solution**
Systematic sampling.

(b) From your knowledge of the large data set (2)

(i) suggest which variable was being studied,

**Solution**
Daily Mean Wind Speed.

(ii) state the name of outcome $A$.

**Solution**
Light.

George is also studying the same variable from the large data set for Heathrow from 2015.
He started with the data from 5th May and then took every 10th reading and obtained the following.

| Outcome | $A$ | $B$ | $C$ |
|---|---|---|---|
| Frequency | 16 | 1 | 1 |

Helen and George decided they should examine all of the data for this variable for Heathrow from 2015 and obtained the following.

| Outcome | $A$ | $B$ | $C$ |
|---|---|---|---|
| Frequency | 155 | 26 | 3 |

(c) State what inference Helen and George could reliably make from their original samples about the outcomes of this variable at Heathrow, for the period covered by the large data set in 2015. (1)

**Solution**
E.g., Variable $A$ occurs most (around $80-90\%$) of the time

4. A nursery has a sack containing a large number of coloured beads of which 14% are coloured red.

Aliya takes a random sample of 18 beads from the sack to make a bracelet.

(a) State a suitable binomial distribution to model the number of red beads in Aliya's bracelet. (1)

**Solution**
It is the Binomial distribution. If $R$ is the number of red beads in Aliya's bracelet, $R \sim \mathrm{B}(18, 0.14)$.

(b) Use this binomial distribution to find the probability that (3)

(i) Aliya has just 1 red bead in her bracelet,

**Solution**

$$\begin{aligned}
\text{P(1 red bead)} &= \binom{18}{1}(0.14)(0.86^{17}) \\
&= 0.194\,032\,380\,5 \text{ (FCD)} \\
&= \underline{\underline{0.194 \text{ (3 sf)}}}.
\end{aligned}$$

(ii) there are at least 4 red beads in Aliya's bracelet.

**Solution**

$$\begin{aligned}
\text{P(at least 4 red beads)} &= 1 - \text{P(at most 3 red beads)} \\
&= 1 - \left[(0.86)^{18} + \binom{18}{1}(0.14)(0.86^{17}) \right.\\
&\quad \left. + \binom{18}{12}(0.14^2)(0.86^{16}) + \binom{18}{3}(0.14^3)(0.86^{15})\right] \\
&= 0.238\,158\,851\,4 \text{ (FCD)} \\
&= \underline{\underline{0.238 \text{ (3 sf)}}}.
\end{aligned}$$

(c) Comment on the suitability of a binomial distribution to model this situation. (1)

**Solution**

Well, we require a large number of beads in the sack to ensure that removing 18 beads does not appreciably affect this probability. If that *is* the case, then it could be suitable.

After several children have used beads from the sack, the nursery teacher decides to test whether or not the proportion of red beads in the sack has changed.

She takes a random sample of 75 beads and finds 4 red beads.

(d) Stating your hypotheses clearly, use a 5% significance level to carry out a suitable test for the teacher. (4)

**Solution**

Well, let $X$ number of red beads in the sample: $X \sim \mathrm{B}(75, 0.14)$.
$H_0$: $p = 0.14$.
$H_1$: $p \neq 0.14$.
Level of significance: 5%.

1. We use our calculator (I am using fx-991EX although other calculators are available ...): $\boxed{\text{Menu}}$ and $\boxed{\text{7: Distribution}}$ (it is the normal distribution).
2. Scroll down until you see $\boxed{\text{1: Binomial CD}}$ and then $\boxed{\text{2: Variable}}$.
3. You should then see a box with x, N, and p on it: we type $\boxed{4}$ and $\boxed{=}$.
4. We type $\boxed{75}$ and $\boxed{=}$.
5. We type $\boxed{0.14}$ and $\boxed{=}$.

Now, $\text{P}(X \leqslant 4) = 0.015\,061\,720\,2$
$0.015\ldots < 0.025$ and so the result is significant: we reject $H_0$.
Hence, there is evidence at the 5% level that the proportion of red beads has changed.

(e) Find the $p$-value in this case. (1)

**Solution**
$2 \times 0.015\ldots = 0.030\,123\,440\,39$ (FCD) $= \underline{\underline{0.03 \text{ (1 sf)}}}$.

5. Two bags, **A** and **B**, each contain balls which are either red or yellow or green. (5)

Bag **A** contains 4 red, 3 yellow and $n$ green balls.
Bag **B** contains 5 red, 3 yellow and 1 green ball.

A ball is selected at random from bag **A** and placed into bag **B**.
A ball is then selected at random from bag **B** and placed into bag **A**.

The probability that bag **A** now contains an equal number of red, yellow, and green balls is $p$.

Given that $p > 0$, find the possible values of $n$ and $p$.

**Solution**
Well! The yellow must be unaffected (why?).

Case 1: We select a red ball from bag **A**. That means we have an equal number of reds and yellows and so we take the green.

| | **Start** | | | Bag **A** → Bag **B** | | | Bag **B** → Bag **A** | | |
|---|---|---|---|---|---|---|---|---|---|
| Bag **A** | 4 | 3 | $n$ | 3 | 3 | $n$ | 3 | 3 | $n+1$ |
| | **R** | **Y** | **G** | **R** | **Y** | **G** | **R** | **Y** | **G** |
| Bag **B** | 5 | 3 | 1 | 6 | 3 | 1 | 6 | 3 | 0 |

Now,

$$n + 1 = 3 \Rightarrow \underline{\underline{n = 2}}$$

and

$$\begin{aligned} \text{P (equal number)} &= \tfrac{4}{9} \times \tfrac{1}{10} \\ &= \underline{\underline{\tfrac{2}{45}}}. \end{aligned}$$

<u>Case 2</u>: We select a green ball from bag **A**. So that means we take a yellow ball for our second pick.

| | **Start** | | | Bag **A** → Bag **B** | | | Bag **B** → Bag **A** | | |
|---|---|---|---|---|---|---|---|---|---|
| Bag **A** | 4 | 3 | $n$ | 4 | 3 | $n-1$ | 4 | 4 | $n-1$ |
| | **R** | **Y** | **G** | **R** | **Y** | **G** | **R** | **Y** | **G** |
| Bag **B** | 5 | 3 | 1 | 5 | 3 | 2 | 5 | 2 | 2 |

Now,

$$n - 1 = 4 \Rightarrow \underline{\underline{n = 5}}$$

and the number of balls in bag **A** is

$$4 + 3 + 5 = 12.$$

Finally,

$$\begin{aligned} \text{P (equal number)} &= \tfrac{5}{12} \times \tfrac{3}{10} \\ &= \underline{\underline{\tfrac{1}{8}}}. \end{aligned}$$

An excellent question!