# Dr Oliver Mathematics
# Advanced Level Paper 31: Statistics
# June 2022: Calculator
# 2 hours

The total number of marks available is 50.
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 32: Mechanics)

1. George throws a ball at a target 15 times.
   Each time George throws the ball, the probability of the ball hitting the target is 0.48.

   The random variable $X$ represents the number of times George hits the target in 15 throws.

   (a) Find (3)

   (i) $\mathrm{P}(X = 3)$,

   **Solution**
   Let $X \sim \mathrm{B}(15, 0.48)$. Then

   $$\begin{aligned} \mathrm{P}(X = 3) &= 0.019\,668\,680\,81 \text{ (FCD)} \\ &= \underline{\underline{0.019\,7 \text{ (3 sf)}}}. \end{aligned}$$

   (ii) $\mathrm{P}(X \geqslant 5)$.

   **Solution**
   Well,

   $$\begin{aligned} \mathrm{P}(X \geqslant 5) &= 1 - \mathrm{P}(X \leqslant 4) \\ &= 1 - 0.079\,868\,918\,14 \text{ (FCD)} \\ &= 0.920\,131\,081\,9 \text{ (FCD)} \\ &= \underline{\underline{0.920 \text{ (3 sf)}}}. \end{aligned}$$

   George now throws the ball at the target 250 times.

   (b) Use a normal approximation to calculate the probability that he will hit the target more than 110 times. (3)

**Solution**

Well, $X \sim \mathrm{B}(250, 0.48)$ so we will use the Normal distribution:

$$n = 250 \times 0.48 = 120$$
$$\sigma^2 = 250 \times 0.48 \times 0.52 = 62.4;$$

in other words, $Y \approx\sim \mathrm{N}(120, 62.4)$. Now,

$$\begin{aligned} \mathrm{P}(X > 110) &\approx \mathrm{P}(Y > 110.5) \\ &= \mathrm{P}\left(Z > \frac{110.5 - 120}{\sqrt{62.4}}\right) \\ &= \mathrm{P}\left(Z > -1.202\right) \\ &= \mathrm{P}\left(Z < 1.202\right) \\ &= 0.885\,439\,846 \text{ (FCD)} \\ &= \underline{\underline{0.885 \text{ (3 sf)}}}. \end{aligned}$$

2. A manufacturer uses a machine to make metal rods.

The length of a metal rod, $L$ cm, is normally distributed with

- a mean of 8 cm and
- a standard deviation of $x$ cm.

Given that the proportion of metal rods less than 7.902 cm in length is 2.5%,

(a) show that $x = 0.05$ to 2 decimal places. (2)

**Solution**

Well,

$$\begin{aligned} \mathrm{P}(L < 7.902) = 0.025 &\Rightarrow \frac{7.902 - 8}{x} = -1.96 \\ &\Rightarrow x = \frac{-0.098}{-1.96} \\ &\Rightarrow \underline{\underline{x = 0.05}}. \end{aligned}$$

(b) Calculate the proportion of metal rods that are between 7.94 cm and 8.09 cm in length. (1)

**Solution**

$$\begin{aligned}\mathrm{P}(7.94 < L < 8.09) &= 0.084\,900\,001\,06 \text{ (FCD)} \\ &= \underline{\underline{0.084\,9 \text{ (3 sf)}}}.\end{aligned}$$

The cost of producing a single metal rod is 20 p.

A metal rod

- where $L < 7.94$ is **sold** for scrap for 5 p,
- where $7.94 \leqslant L \leqslant 8.09$ is **sold** for 50 p, and
- where $L > 8.09$ is shortened for an extra **cost** of 10 p and then **sold** for 50 p.

(c) Calculate the expected profit per 500 of the metal rods. (5)
Give your answer to the nearest pound.

**Solution**

Now,

$$\mathrm{P}(L < 7.94) = 0.115\,069\,702 \text{ (FCD)}$$

and

$$\begin{aligned}\mathrm{P}(L > 8.09) &= 1 - \mathrm{P}(L < 8.09) \\ &= 1 - 0.964\,069\,809 \text{ (FCD)} \\ &= 0.035\,930\,191\,4 \text{ (FCD)}.\end{aligned}$$

Finally,

$$\begin{aligned}&\text{expected profit} \\ =\ &\text{retail} - \text{wholesale} \\ =\ &500\,[(0.05 \times 0.115\ldots) + (0.5 \times 0.084\ldots) + (0.4 \times 0.035\ldots)] - (500 \times 0.2) \\ =\ &222.312\,808\,2 - 100 \\ =\ &\underline{\underline{\text{£}122 \text{ (nearest pound)}}}.\end{aligned}$$

The same manufacturer makes metal hinges in large batches.

The hinges each have a probability of 0.015 of having a fault.

A random sample of 200 hinges is taken from each batch and the batch is accepted if fewer than 6 hinges are faulty.

The manufacturer's aim is for 95% of batches to be accepted.

(d) Explain whether the manufacturer is likely to achieve its aim. (4)

**Solution**

Let $X \sim \mathrm{B}(200, 0.015)$. Then

$$\mathrm{P}(X \leqslant 5) = 0.917\,608\,302 \text{ (FCD)}.$$

So, it is $\underline{\underline{\text{unlikely}}}$ that the manufacturer will achieve their aim as $0.917\ldots < 0.95$.

3. Dian uses the large data set to investigate the Daily Total Rainfall, $r$ mm, for Camborne.

(a) Write down how a value of $0 < r \leqslant 0.05$ is recorded in the large data set. (1)

**Solution**

$\underline{\underline{\text{tr}}}$.

Dian uses the data for the 31 days of August 2015 for Camborne and calculates the following statistics:

$$n = 31, \sum r = 174.9, \text{ and } \sum r^2 = 3\,523.283.$$

(b) Use these statistics to calculate (3)

(i) the mean of the Daily Total Rainfall in Camborne for August 2015,

**Solution**

$$\begin{aligned} \mu &= \frac{174.9}{31} \\ &= 5.641\,935\,484 \text{ (FCD)} \\ &= \underline{\underline{5.64 \text{ (3 sf)}}}. \end{aligned}$$

(ii) the standard deviation of the Daily Total Rainfall in Camborne for August 2015.

**Solution**

$$\begin{aligned}\sigma &= \sqrt{\frac{3\,523.283}{31} - 5.641\ldots^2} \\ &= 9.045\,598\,616 \text{ (FCD)} \\ &= \underline{\underline{9.05 \text{ (3 sf)}}}.\end{aligned}$$

Dian believes that the mean Daily Total Rainfall in August is less in the South of the UK than in the North of the UK.

The mean Daily Total Rainfall in Leuchars for August 2015 is 1.72 mm to 2 decimal places.

(c) State, giving a reason, whether this provides evidence to support Dian's belief. (2)

**Solution**

Well, Leuchars is in the *North* and Camborne is in the *South*.
The mean is smaller for Leuchars than Camborne.
Hence, there is no evidence that Dian's belief is true

Dian uses the large data set to estimate the proportion of days with no rain in Camborne for 1987 to be 0.27 to 2 decimal places.

(d) Explain why the distribution $B(14, 0.27)$ might not be a reasonable model for the number of days without rain for a 14-day summer event. (1)

**Solution**

E.g., consecutive 14 days unlikely to be independent, probability of rain not being constant.

4. A dentist knows from past records that 10% of customers arrive late for their appointment.

A new manager believes that there has been a change in the proportion of customers who arrive late for their appointment.

A random sample of 50 of the dentist's customers is taken.

(a) Write down (1)

- a null hypothesis corresponding to no change in the proportion of customers who arrive late and

- an alternative hypothesis corresponding to the manager's belief.

**Solution**
$\underline{\underline{H_0\colon p = 0.1}}$
$\underline{\underline{H_1\colon p \neq 0.1}}$.

(b) Using a 5% level of significance, find the critical region for a two-tailed test of the null hypothesis in (a). (3)
You should state the probability of rejection in each tail, which should be less than 0.025.

**Solution**
Well, we will use $X \sim \mathrm{B}(50, 0.1)$. The critical regions are

$$\underline{\underline{X = 0 \quad (0.005\,2)}}$$

and

$$\underline{\underline{X \geqslant 10 \quad (1 - 0.975\,5 = 0.024\,5)}}.$$

(c) Find the actual level of significance of the test based on your critical region from part (b). (1)

**Solution**

$$0.005\,2 + 0.024\,5 = \underline{\underline{0.029\,7}}.$$

The manager observes that 15 of the 50 customers arrived late for their appointment.

(d) With reference to part (b), comment on the manager's belief. (1)

**Solution**
E.g., 15 is in the $\underline{\underline{\text{critical region}}}$.
Hence, there is $\underline{\underline{\text{evidence}}}$ to support the manager's belief.

5. A company has 1 825 employees.
The employees are classified as professional, skilled, or elementary.

The following table shows

- the number of employees in each classification and

- the two areas, $A$ or $B$, where the employees live

| | $A$ | $B$ |
|---|---|---|
| Professional | 740 | 380 |
| Skilled | 275 | 90 |
| Elementary | 260 | 80 |

An employee is chosen at random.

Find the probability that this employee

(a) is skilled, (1)

**Solution**

$$\frac{275 + 90}{1\,825} = \underline{\underline{0.2}}.$$

(b) lives in area $B$ and is not a professional. (1)

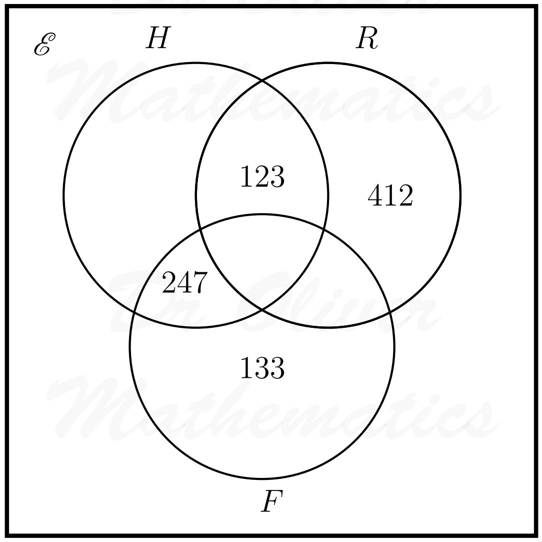
**Solution**

$$\frac{90 + 80}{1\,825} = \tfrac{34}{365} = \underline{\underline{0.093\,2 \text{ (3 sf)}}}.$$

Some classifications of employees are more likely to work from home.

- 65% of professional employees in both area $A$ and area $B$ work from home,
- 40% of skilled employees in both area $A$ and area $B$ work from home,
- 5% of elementary employees in both area $A$ and area $B$ work from home.
- Event $F$ is that the employee is a professional,
- Event $H$ is that the employee works from home, and
- Event $R$ is that the employee is from area $A$.

(c) Using this information, complete the Venn diagram. (4)




**Solution**
Well,

the number in all three:

$$740 \times 0.65 = 481,$$

completing $F$:

$$\begin{aligned}(740 + 380) - (133 + 247 + 481) &= 1\,120 - 861 \\ &= 259,\end{aligned}$$

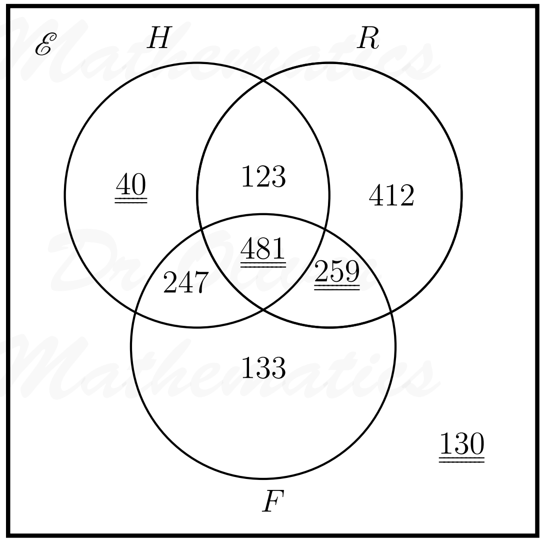
$H$ only:

$$\begin{aligned}(90 \times 0.4) + (80 \times 0.05) &= 36 + 4 \\ &= 40,\end{aligned}$$

the rest of them:

$$\begin{aligned}1\,825 - (40 + 123 + 412 + 247 + 259 + 481 + 133) &= 1\,825 - 1\,695 \\ &= 130.\end{aligned}$$

So, we complete the Venn diagram:




(d) Find $\mathrm{P}(R' \cap F)$. (1)

**Solution**

$$\begin{aligned}\mathrm{P}(R' \cap F) &= \frac{247 + 133}{1\,825} \\ &= \tfrac{380}{1\,825} \\ &= \underline{\underline{0.208 \text{ (3 sf)}}}.\end{aligned}$$

(e) Find $\mathrm{P}([H \cup R]')$. (1)

**Solution**

$$\begin{aligned}\mathrm{P}([H \cup R]') &= \frac{133 + 130}{1\,825} \\ &= \tfrac{263}{1\,825} \\ &= \underline{\underline{0.144 \text{ (3 sf)}}}.\end{aligned}$$

(f) Find $\mathrm{P}(F|H)$. (2)

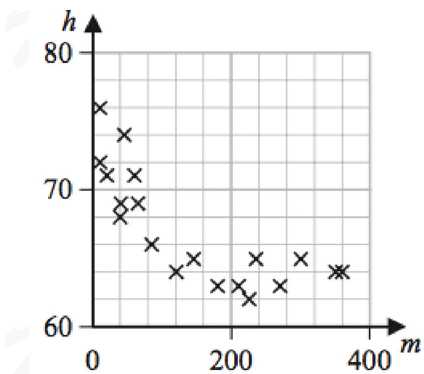
**Solution**

$$
\begin{aligned}
\mathrm{P}(F|H) &= \frac{\mathrm{P}(F \cap H)}{\mathrm{P}(H)} \\
&= \frac{247 + 481}{40 + 123 + 247 + 481} \\
&= \tfrac{728}{891} \\
&= \underline{\underline{0.817 \text{ (3 sf)}}}.
\end{aligned}
$$

6. Anna is investigating the relationship between exercise and resting heart rate.
   She takes a random sample of 19 people in her year at school and records for each person

   - their resting heart rate, $h$ beats per minute and
   - the number of minutes, $m$, spent exercising each week.

   Her results are shown on the scatter diagram.




   (a) Interpret the nature of the relationship between $h$ and $m$. (1)

   **Solution**
   E.g., as the number of minutes exercise ($m$), increases the resting heart rate ($h$) decreases.

   Anna codes the data using the formulae

$$
\begin{aligned}
x &= \log_{10} m, \\
y &= \log_{10} h.
\end{aligned}
$$

   The product moment correlation coefficient between $x$ and $y$ is $-0.897$.

   (b) Test whether or not there is significant evidence of a negative correlation between $x$ and $y$. (3)

You should

- state your hypotheses clearly,
- use a 5% level of significance, and
- state the critical value used.

**Solution**

$H_0$: $\rho = 0$
$H_1$: $\rho < 0$
Level of significance: 5%
Critical vale: $-0.388\,7$
There is evidence that the PMCC is less than 0.
Hence, there is a negative correlation.

The equation of the line of best fit of $y$ on $x$ is

$$y = -0.05x + 1.92.$$

(c) Use the equation of the line of best fit of $y$ on $x$ to find a model for $h$ on $m$ in the form (5)

$$h = am^k,$$

where $a$ and $k$ are constants to be found.

**Solution**

$$\begin{aligned} y = -0.05x + 1.92 &\Rightarrow \log_{10} h = -0.05 \log_{10} m + 1.92 \\ &\Rightarrow \log_{10} h + 0.05 \log_{10} m = 1.92 \\ &\Rightarrow \log_{10} h + \log_{10} m^{0.05} = 1.92 \\ &\Rightarrow \log_{10}(hm^{0.05}) = 1.92 \\ &\Rightarrow hm^{0.05} = 10^{1.92} \\ &\Rightarrow h = 10^{1.92} m^{-0.05} \\ &\Rightarrow h = 83.176\,377\,11 m^{-0.05} \text{ (FCD)}; \end{aligned}$$

hence, $h = 83.2$ (3 sf) and $k = -0.05$.