Dr Oliver Mathematics
Advanced Level Paper 31: Statistics
November 2021: 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. (a) State one disadvantage of using quota sampling compared with simple random sampling. (1)

Solution
E.g., quota sampling is not random.

In a university 8% of students are members of the university dance club.

A random sample of 36 students is taken from the university.

The random variable X represents the number of these students who are members of the dance club.

(b) Using a suitable model for X, find (3)

(i) $\text{P}(X = 4)$,

Solution
Well, $X \sim \text{B}(36, 0.08)$ and so

$$\begin{aligned}\text{P}(X = 4) &= \binom{36}{4}(0.08)^4(0.92)^{32} \\ &= 0.167\,387\,330\,5 \text{ (FCD)} \\ &= \underline{\underline{0.167 \text{ (3 sf)}}}.\end{aligned}$$

(ii) $\text{P}(X \geqslant 7)$.

Solution

$$
\begin{aligned}
\mathrm{P}(X \geqslant 7) &= 1 - \mathrm{P}(X \leqslant 6) \\
&= 1 - 0.977\,766\,163\,35 \text{ (FCD)} \\
&= 0.022\,233\,836\,46 \text{ (FCD)} \\
&= \underline{\underline{0.022 \text{ (3 sf)}}}.
\end{aligned}
$$

Only 40% of the university dance club members can dance the tango.

(c) Find the probability that a student is a member of the university dance club and can dance the tango. (1)

Solution

$$
\begin{aligned}
\mathrm{P}(\text{dance club } \textbf{and} \text{ dance the tango}) &= 0.4 \times 0.08 \\
&= \underline{\underline{0.032}}.
\end{aligned}
$$

A random sample of 50 students is taken from the university.

(d) Find the probability that fewer than 3 of these students are members of the university dance club and can dance the tango. (2)

Solution

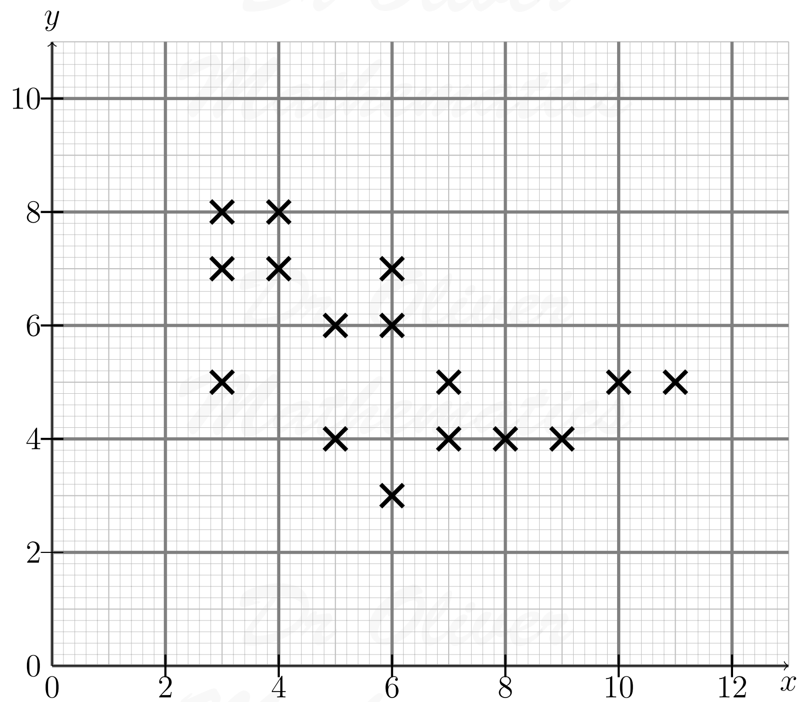
Let T be those who can dance the tango. Now, $T \sim \mathrm{B}(50, 0.032)$ and

$$
\begin{aligned}
\mathrm{P}(T < 3) &= (0.968)^{50} + \binom{50}{1}(0.032)(0.968)^{49} + \binom{50}{2}(0.032)^2(0.968)^{48} \\
&= 0.196\,683\,471\,7 + 0.325\,096\,647\,4 + 0.263\,301\,416\,9 \text{ (FCD)} \\
&= 0.785\,081\,535\,9 \text{ (FCD)} \\
&= \underline{\underline{0.785 \text{ (3 sf)}}}.
\end{aligned}
$$

2. Marc took a random sample of 16 students from a school and for each student recorded

- the number of letters, x, in their last name and
- the number of letters, y, in their first name.

His results are shown in the scatter diagram.



(a) Describe the correlation between x and y. (1)

> **Solution**
> Negative correlation.

Marc suggests that parents with long last names tend to give their children shorter first names.

(b) Using the scatter diagram comment on Marc's suggestion, giving a reason for your answer. (1)

> **Solution**
> E.g., Marc's suggestion is compatible because it is negative correlation.

The results from Marc's random sample of 16 observations are given in the table below.

x	3	6	8	7	5	3	11	3	4	5	4	9	7	10	6	6
y	7	7	4	4	6	8	5	5	8	4	7	4	5	5	6	3

(c) Use your calculator to find the product moment correlation coefficient between x and y for these data. (1)

Solution
Well,

$$\sum x = 97, \sum y = 88, \sum x^2 = 681, \sum y^2 = 520, \text{ and } \sum xy = 502$$

which leads to

$$r = -0.544\,582\,622 \text{ (FCD)} = \underline{\underline{-0.545 \text{ (3 sf)}}}.$$

(d) Test whether or not there is evidence of a negative correlation between the number of letters in the last name and the number of letters in the first name. (3)

You should

- state your hypotheses clearly and
- use a 5% level of significance.

Solution
$H_0 : \rho = 0$.
$H_1 : \rho < 0$.
Level of significance: 5%
Using $n = 16$, we get the PMCC $-0.429\,5$.
Now, $-0.544\ldots < -0.429\,5$: we reject H_0.
There is evidence of negative correlation between the number of letters first and last names.

3. Stav is studying the large data set for September 2015.

He codes the variable Daily Mean Pressure, x, using the formula

$$y = x - 1\,010.$$

The data for all 30 days from Hurn are summarised by

$$\sum y = 214 \text{ and } \sum y^2 = 5\,912.$$

(a) State the units of the variable x. (1)

Solution
Hectopascal.

(b) Find the mean Daily Mean Pressure for these 30 days. (2)

Solution

Well,

$$\begin{aligned}\bar{y} = \bar{x} - 1\,010 &\Rightarrow \frac{214}{30} = \bar{x} - 1\,010\\ &\Rightarrow \bar{x} = \frac{214}{30} + 1\,010\\ &\Rightarrow \bar{x} = 1\,017\tfrac{2}{15}\\ &\Rightarrow \underline{\underline{\bar{x} = 1\,017 \text{ (4 sf)}}}.\end{aligned}$$

(c) Find the standard deviation of Daily Mean Pressure for these 30 days. (3)

Solution

Now, the standard deviation is **not** affected by this coding (why?) so

$$\begin{aligned}\text{standard deviation} &= \sqrt{\frac{5\,912}{30} - (7\tfrac{2}{15})^2}\\ &= 12.090\,584\,03 \text{ (FCD)}\\ &= \underline{\underline{12.1 \text{ (3 sf)}}}.\end{aligned}$$

Stav knows that, in the UK, winds circulate

- in a **clockwise** direction around a region of **high** pressure and
- in an **anticlockwise** direction around a region of **low** pressure.

The table gives the Daily Mean Pressure for 3 locations from the large data set on 26/09/2015.

Location	Heathrow	Hurn	Leuchars
Daily Mean Pressure	1 029	1 028	1 028
Cardinal Wind Direction			

The Cardinal Wind Directions for these 3 locations on 26/09/2015 were, in random order,

W NE E.

You may assume that these 3 locations were under a single region of pressure.

(d) Using your knowledge of the large data set, place each of these Cardinal Wind Directions in the correct location in the table. (2)
Give a reason for your answer.

Solution

Well, the locations are, from North to South, Leuchars, Heathrow, and Hurn. Now,

$$1\,029 = 1\,017\tfrac{2}{15} + 12.090\ldots$$

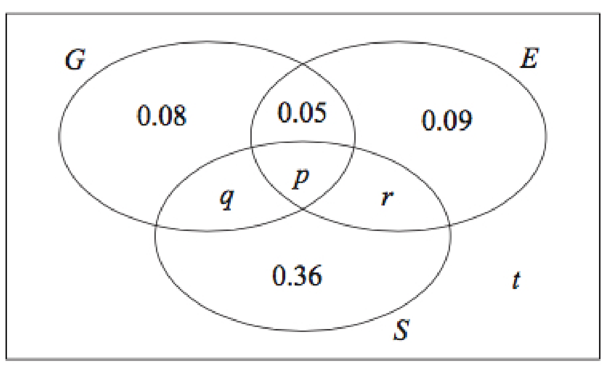
so we have high pressure which means that the wind circulates in a clockwise direction. Finally, the 'wind direction' is the direction wind blows **from**:

Location	Heathrow	Hurn	Leuchars
Daily Mean Pressure	1 029	1 028	1 028
Cardinal Wind Direction	NE	E	W

4. A large college produces three magazines.
One magazine is about green issues, one is about equality, and one is about sports.
A student at the college is selected at random and the events G, E and S are defined as follows

- G is the event that the student reads the magazine about green issues,
- E is the event that the student reads the magazine about equality, and
- S is the event that the student reads the magazine about sports

The Venn diagram, where p, q, r, and t are probabilities, gives the probability for each subset.



(a) Find the proportion of students in the college who read exactly one of these magazines. (1)

Solution

$$0.08 + 0.09 + 0.36 = \underline{\underline{0.53}}.$$

No students read all three magazines and $\mathrm{P}(G) = 0.25$.

(b) Find (3)

(i) the value of p,

Solution

If no students read all three magazines, then $\underline{\underline{p = 0}}$.

(ii) the value of q.

Solution

$$\begin{aligned} 0.08 + 0.05 + 0 + q = 0.25 &\Rightarrow 0.13 + q = 0.25 \\ &\Rightarrow \underline{\underline{q = 0.12}}. \end{aligned}$$

Given that $\mathrm{P}(S|E) = \frac{5}{12}$,

(c) find (4)

(i) the value of r,

Solution

$$\begin{aligned} \mathrm{P}(S|E) = \tfrac{5}{12} &\Rightarrow \frac{\mathrm{P}(S \cap E)}{\mathrm{P}(E)} = \tfrac{5}{12} \\ &\Rightarrow \frac{0 + r}{0.05 + 0.09 + r + 0} = \tfrac{5}{12} \\ &\Rightarrow \frac{r}{0.14 + r} = \tfrac{5}{12} \\ &\Rightarrow 12r = 5(0.14 + r) \\ &\Rightarrow 12r = 0.7 + 5r \\ &\Rightarrow 7r = 0.7 \\ &\Rightarrow \underline{\underline{r = 0.1}}. \end{aligned}$$

(ii) the value of t.

Solution

$$0.08 + 0.05 + 0.09 + 0.12 + 0 + 0.1 + 0.36 + t = 1 \Rightarrow 0.8 + t = 1$$
$$\Rightarrow \underline{\underline{t = 0.2}}.$$

(d) Determine whether or not the events $(S \cap E')$ and G are independent. Show your working clearly. (3)

Solution

Well,

$$P(S \cap E') = 0.12 + 0.36 = 0.48,$$
$$P(G) = 0.08 + 0.05 + 0 + 0.12 = 0.25,$$

and so

$$P(S \cap E') \times P(G) = 0.48 \times 0.25 = 0.12.$$

But

$$P((S \cap E') \cap G) = 0.12.$$

Yes, they are $\underline{\underline{\text{independent}}}$.

5. The heights of females from a country are normally distributed with

- a mean of 166.5 cm and
- a standard deviation of 6.1 cm.

Given that 1% of females from this country are shorter than k cm,

(a) find the value of k. (2)

Solution

Let F be the heights of females. Then $F \sim N(166.5, 6.1^2)$. Now,

$$P(F < k) = 0.01 \Rightarrow \frac{k - 166.5}{6.1} = -2.326\,3$$
$$\Rightarrow k - 166.5 = -14.190\,43$$
$$\Rightarrow k = 152.309\,57$$
$$\Rightarrow \underline{\underline{k = 152 \text{ (3 sf)}}}.$$

(b) Find the proportion of females from this country with heights between 150 cm and 175 cm. (1)

Solution

$$\begin{aligned} \mathrm{P}(150 < F < 175) &= 0.914\,840\,945\,5 \text{ (FCD)} \\ &= \underline{\underline{0.915 \text{ (3 sf)}}}. \end{aligned}$$

A female, from this country, is chosen at random from those with heights between 150 cm and 175 cm.

(c) Find the probability that her height is more than 160 cm. (4)

Solution

Well,

$$\begin{aligned} \mathrm{P}(F > 160 | 150 < F < 175) &= \frac{\mathrm{P}(160 < F < 175)}{\mathrm{P}(150 < F < 175)} \\ &= \frac{0.774\,948\ldots}{0.914\,840\ldots} \text{ (FCD)} \\ &= 0.847\,085\,868\,9 \text{ (FCD)} \\ &= \underline{\underline{0.84 \text{ (3 sf)}}}. \end{aligned}$$

The heights of females from a different country are normally distributed with a standard deviation of 7.4 cm.

Mia believes that the mean height of females from this country is less than 166.5 cm.

Mia takes a random sample of 50 females from this country and finds the mean of her sample is 164.6 cm.

(d) Carry out a suitable test to assess Mia's belief. (4)
You should

- state your hypotheses clearly and
- use a 5% level of significance.

Solution

Let G be the heights of female from the second country. Then

$$\bar{G} \sim \mathrm{N}\left(166.5, \left(\frac{7.4}{\sqrt{50}}\right)^2\right).$$

> H_0: $\mu = 166.5$.
> H_1: $\mu < 166.5$.
> Level of significance: 5%
> $\text{P}(\bar{G} < 164.6) = 0.034\,720\,141\,69$ (FCD)
> $0.034\ldots < 0.05$ and so we reject H_0: there is evidence of to support Mia's belief.

6. The discrete random variable X has the following probability distribution:

x	a	b	c
$\text{P}(X = x)$	$\log_{36} a$	$\log_{36} b$	$\log_{36} c$

where

- a, b, and c are distinct integers ($a < b < c$) and
- all the probabilities are greater than zero.

(a) Find, show your working clearly, (5)

(i) the value of a,

> **Solution**
> Well,
>
> $$\log_{36} a + \log_{36} b + \log_{36} c = 1 \Rightarrow \log_{36}(abc) = 1$$
> $$\Rightarrow abc = 36.$$
>
> Now,
>
> | | |
> |---|---|
> | | 36 |
> | 2 | 18 |
> | 2 | 9 |
> | 3 | 3 |
> | 3 | 1 |
>
> and
>
> $$36 = 2^2 \times 3^2.$$
>
> As we know that a, b, and c are distinct integers, $a = 2$.

(ii) the value of b, and

Solution
$\underline{\underline{b = 3}}$.

(iii) the value of c.

Solution
$\underline{\underline{c = 6}}$.

The independent random variables X_1 and X_2 each have the same distribution as X.

(b) Find $\mathrm{P}(X_1 = X_2)$. (2)

Solution

$$\begin{aligned}
\mathrm{P}(X_1 = X_2) &= (\log_{36} 2)^2 + (\log_{36} 3)^2 + (\log_{36} 6)^2 \\
&= 0.037\,413\,773\,62 + 0.093\,987\,37 + 0.25 \text{ (FCD)} \\
&= 0.381\,401\,143\,6 \text{ (FCD)} \\
&= \underline{\underline{0.381 \text{ (3 sf)}}}.
\end{aligned}$$