Dr Oliver Mathematics
Advance Level Mathematics
AS Statistics: Calculator
1 hour 15 minutes

The total number of marks available is 30.
You must write down all the stages in your working.
Note: It goes with the AS Mechanics paper.

1. A sixth form college has 84 students in Year 12 and 56 students in Year 13.
The head teacher selects a stratified sample of 40 students, stratified by year group.

   (a) Describe how this sample could be taken. (3)

Solution

The head teacher could take

$$\left(\frac{84}{84+56}\right) \times 40 = 24 \text{ students}$$

from Year 12 and

$$40 - 24 = 16 \text{ students}$$

from Year 13. From Year 12, the head teacher

- could number the students 1, 2, ..., 84,
- could write all of the numbers on pieces of paper and place them all in a bag,
- and select, without replacement, twenty-four.

From Year 13, the head teacher repeat the process, ditching 57, 59, ..., 84, and select, without replacement, sixteen.

The head teacher is investigating the relationship between the amount of sleep, s hours, that each student had the night before they took an aptitude test and their performance in the test, p marks.
For the sample of 40 students, he finds the equation of the regression line of p on s to be

$$p = 26.1 + 5.60s.$$

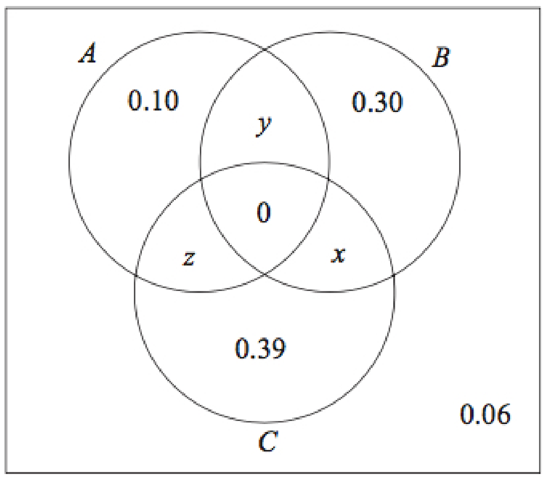
   (b) With reference to this equation, describe the effect that an extra 0.5 hours of sleep may have, on average, on a student's performance in the aptitude test. (1)

> **Solution**
> On average, for every half an hour they sleep, their performance in the aptitude test goes up by
> $$5.60 \times 0.5 = \underline{\underline{2.8 \text{ marks}}}.$$

(c) Describe one limitation of this regression model. (1)

> **Solution**
> E.g., a student cannot score below 26.1 marks on the test.

2. The Venn diagram shows three events, A, B, and C, and their associated probabilities. (5)



Events B and C are mutually exclusive.
Events A and C are independent.

Showing your working, find the value of x, the value of y, and the value of z.

> **Solution**
> $\underline{\underline{x = 0}}$ because B and C are mutually exclusive.
> Now, all the probabilities should add up to 1:
> $$0.06 + 0.10 + 0.30 + 0.39 + y + z = 1 \Rightarrow y + z = 0.15.$$

Next,

$$\begin{aligned}
\mathrm{P}(A \cap C) = \mathrm{P}(A)\,\mathrm{P}(C) &\Rightarrow z = (0.10 + 0.15)(0.39 + z) \\
&\Rightarrow z = 0.25(0.39 + z) \\
&\Rightarrow z = 0.0975 + 0.25z \\
&\Rightarrow 0.75z = 0.0975 \\
&\Rightarrow \underline{\underline{z = 0.13}} \\
&\Rightarrow \underline{\underline{y = 0.02}}.
\end{aligned}$$

3. A fair 5-sided spinner has sides numbered 1, 2, 3, 4, and 5.

The spinner is spun once and the score of the side it lands on is recorded.

(a) Write down the name of the distribution that can be used to model the score of the side it lands on. (1)

Solution
It is the discrete uniform distribution.

The spinner is spun 28 times.

The random variable X represents the number of times the spinner lands on 2.

(b) (i) Find the probability that the spinner lands on 2 at least 7 times. (5)

Solution
Use $X \sim \mathrm{B}(28, 0.2)$:

$$\begin{aligned}
\mathrm{P}(X \geqslant 7) &= 1 - \mathrm{P}(X \leqslant 6) \\
&= 1 - 0.678\,443\,503 \text{ (FCD)} \\
&= 0.321\,556\,497 \text{ (FCD)} \\
&= \underline{\underline{0.3216 \text{ (4 dp)}}}.
\end{aligned}$$

(ii) Find $\mathrm{P}(4 \leqslant X < 8)$.

Solution

$$
\begin{aligned}
\mathrm{P}(4 \leqslant X < 8) &= \mathrm{P}(X \leqslant 7) - \mathrm{P}(X \leqslant 3) \\
&= 0.818\,230\,274\,4 - 0.160\,182\,671\,1 \text{ (FCD)} \\
&= 0.658\,047\,603\,3 \text{ (FCD)} \\
&= \underline{\underline{0.658\,0 \text{ (4 dp)}}}.
\end{aligned}
$$

4. Joshua is investigating the daily total rainfall in Hurn for May to October 2015.

   Using the information from the large data set, Joshua wishes to calculate the mean of the daily total rainfall in Hurn for May to October 2015.

   (a) Using your knowledge of the large data set, explain why Joshua needs to clean the data before calculating the mean. (1)

   Solution

   'Trace' has no numerical value: the data needs to be converted to numbers before the calculation can be carried out.

   Using the information from the large data set, he produces the grouped frequency table below.

Daily total rainfall (r mm)	Frequency	Midpoint (x mm)
$0 \leqslant r < 0.5$	121	0.25
$0.5 \leqslant r < 1.0$	10	0.75
$1.0 \leqslant r < 5.0$	24	3.0
$5.0 \leqslant r < 10.0$	12	7.5
$10.0 \leqslant r < 30.0$	17	20.0

   You may use

$$\sum fx = 539.75 \text{ and } \sum fx^2 = 7\,704.187\,5.$$

   (b) Use linear interpolation to calculate an estimate for the upper quartile of the daily total rainfall. (2)

   Solution

Daily total rainfall (r mm)	Frequency	Cumulative Frequency
$0 \leqslant r < 0.5$	121	121
$0.5 \leqslant r < 1.0$	10	131
$1.0 \leqslant r < 5.0$	24	155
$5.0 \leqslant r < 10.0$	12	167
$10.0 \leqslant r < 30.0$	17	184

and so the upper quartile is at the

$$\frac{3(184+1)}{4} = 138\tfrac{3}{4}\text{th place}$$

and this is the $7\frac{3}{4}$th value in the $1.0 \leqslant r < 5.0$:

$$\begin{aligned} \text{UQ} &= 1 + \left(\frac{7\frac{3}{4}}{24} \times 4\right) \\ &= 2\tfrac{7}{24}. \end{aligned}$$

(c) Calculate an estimate for the standard deviation of the daily total rainfall in Hurn for May to October 2015. (2)

Solution

$$\begin{aligned} \sigma &= \sqrt{\frac{7\,704.187\,5}{15} - \left(\frac{539.75}{184}\right)^2} \\ &= 5.767\,634\,557 \text{ (FCD)} \\ &= 5.77 \text{ mm (3 sf)}. \end{aligned}$$

(d) (i) State the assumption involved with using class midpoints to calculate an estimate of a mean from a grouped frequency table. (3)

Solution

This assumes that the values are uniformly distributed within the classes.

(ii) Using your knowledge of the large data set, explain why this assumption does not hold in this case.

Solution

Not true: the majority of the data are 0.

(iii) State, giving a reason, whether you would expect the actual mean daily total rainfall in Hurn for May to October 2015 to be larger than, smaller than or the same as an estimate based on the grouped frequency table.

Solution

Given that 121 out of 184 values are in $0 \leqslant r < 0.5$, the actual mean is likely to be less than the estimate.

5. Past records show that 15% of customers at a shop buy chocolate. The shopkeeper believes that moving the chocolate closer to the till will increase the proportion of customers buying chocolate.

After moving the chocolate closer to the till, a random sample of 30 customers is taken and 8 of them are found to have bought chocolate.

Julie carries out a hypothesis test, at the 5% level of significance, to test the shopkeeper's belief. Julie's hypothesis test is shown below.

H_0: $p = 0.15$.
H_1: $p \geqslant 0.15$.
Let X = the number of customers who buy chocolate.
$X \sim \text{B}(30, 0.15)$
$\text{P}(X = 8) = 0.042\,0$
$0.042\,0 < 0.05$ so reject H_0.
There is sufficient evidence to suggest that the proportion of customers buying chocolate has increased.

(a) Identify the first two errors that Julie has made in her hypothesis test. (2)

Solution

First mistake: H_1: $p > 0.15$.

Second mistake: $\text{P}(X \geqslant 8) = 1 - \text{P}(X \leqslant 7) = 1 - 0.932\,2 = 0.067\,8$.

(b) Explain whether or not these errors will affect the conclusion of her hypothesis test. Give a reason for your answer. (1)

Solution

The null hypothesis should <u>not be rejected</u> as $\mathrm{P}(X \geqslant 8) = 0.067\,8 > 0.05$.

(c) Find, using a 5% level of significance, the critical region for a one-tailed test of the shopkeeper's belief. The probability in the tail should be less than 0.05. (2)

Solution

$$\begin{aligned}\mathrm{P}(X \geqslant 9) &= 1 - \mathrm{P}(X \leqslant 8) \\ &= 1 - 0.972\,2 \\ &= 0.027\,8\end{aligned}$$

and the critical region for a one-tailed test of the shopkeeper's belief in $\underline{X \geqslant 9}$.

(d) Find the actual level of significance of this test. (1)

Solution

$\underline{0.027\,8}$.