Dr Oliver Mathematics
Advance Level Further Mathematics
Further Statistics 1: Calculator
1 hour 30 minutes

The total number of marks available is 75.
You must write down all the stages in your working.

1. A chocolate manufacturer places special tokens in 2% of the bars it produces so that each bar contains at most one token. Anyone who collects 3 of these tokens can claim a prize.

   Andreia buys a box of 40 bars of the chocolate.

   (a) Find the probability that Andreia can claim a prize. (2)

   Solution
   Let X be the number of prizes Andreia wins.
   Then $X \sim \text{B}(40, 0.02)$.
   Finally,

   $$\begin{aligned} \text{P}(X \geqslant 3) &= 1 - (X \leqslant 2) \\ &= 1 - 0.954\,329\,769\,6 \text{ (FCD)} \\ &= 0.045\,670\,230\,43 \text{ (FCD)} \\ &= \underline{\underline{0.045\,7 \text{ (4 dp)}}}. \end{aligned}$$

   Barney intends to buy bars of the chocolate, one at a time, until he can claim a prize.

   (b) Find the probability that Barney can claim a prize when he buys his 40th bar of chocolate. (3)

   Solution
   Let Y be the number of the bar when Barney wins.
   Then $Y \sim \text{NB}(3, 0.02)$.
   Finally,

   $$\begin{aligned} \text{P}(Y = 40) &= \left[\binom{39}{2}(0.02)^2(0.98)^{37}\right](0.02) \\ &= 2.807\,197\,676 \times 10^{-3} \text{ (FCD)} \\ &= \underline{\underline{0.002\,8 \text{ (4 dp)}}}. \end{aligned}$$

(c) Find the expected number of bars that Barney must buy to claim a prize. (1)

Solution

$$\mathrm{E}(Y) = \frac{3}{0.02} = \underline{\underline{150}}.$$

2. Indre works on reception in an office and deals with all the telephone calls that arrive. Calls arrive randomly and, in a 4-hour morning shift, there are on average 80 calls.

(a) Using a suitable model, find the probability of more than 4 calls arriving in a particular 20-minute period one morning. (3)

Solution

Well,

$$\begin{aligned} 80 \text{ calls} &\leftrightarrow 4 - \text{hour period} \\ 80 \text{ calls} &\leftrightarrow 240 - \text{minutes period} \\ \tfrac{1}{3} \text{ calls} &\leftrightarrow 1 - \text{minute period} \\ \tfrac{20}{3} \text{ calls} &\leftrightarrow 20 - \text{minute period.} \end{aligned}$$

Let X be the number of calls in a 20-minute period.
Then $X \sim \mathrm{Po}(\frac{20}{3})$.
Finally,

$$\begin{aligned} \mathrm{P}(X > 4) &= 1 - \mathrm{P}(X \leqslant 3) \\ &= 1 - 0.205\,627\,246\,7 \text{ (FCD)} \\ &= 0.794\,372\,753\,3 \text{ (FCD)} \\ &= \underline{\underline{0.794\,4 \text{ (4 dp)}}}. \end{aligned}$$

Indre is allowed 20 minutes of break time during each 4-hour morning shift, which she can take in 5-minute periods. When she takes a break, a machine records details of any call in the office that Indre has missed.

One morning Indre took her break time in 4 periods of 5 minutes each.

(b) Find the probability that in exactly 3 of these periods there were no calls. (2)

Solution

$$\begin{aligned} \tfrac{1}{3} \text{ calls} &\leftrightarrow 1 - \text{minute period} \\ \tfrac{5}{3} \text{ calls} &\leftrightarrow 5 - \text{minute period.} \end{aligned}$$

Let Y be the number of 5-minute periods with no calls.
Then $Y \sim \mathrm{B}(4, \mathrm{e}^{-\frac{5}{3}})$.
Finally,

$$\begin{aligned} \mathrm{P}(Y = 3) &= 0.021\,861\,25278\,3 \text{ (FCD)} \\ &= \underline{\underline{0.021\,9 \text{ (4 dp)}}}. \end{aligned}$$

On another occasion Indre took 1 break of 5 minutes and 1 break of 15 minutes.

(c) Find the probability that Indre missed exactly 1 call in each of these 2 breaks. (3)

Solution

$$\begin{aligned} \tfrac{5}{3} \text{ calls} &\leftrightarrow 5 - \text{minute period} \\ 5 \text{ calls} &\leftrightarrow 15 - \text{minute period}. \end{aligned}$$

Finally,

$$\begin{aligned} \mathrm{P}(\text{exactly one call in each break}) &= \left(\tfrac{5}{3} \times \mathrm{e}^{-\frac{5}{3}}\right)\left(5 \times \mathrm{e}^{-5}\right) \\ &= 0.010\,605\,281\,68 \text{ (FCD)} \\ &= \underline{\underline{0.010\,6 \text{ (4 dp)}}}. \end{aligned}$$

3. A biased spinner can land on the numbers 1, 2, 3, 4, or 5 with the following probabilities. (6)

Number on spinner	1	2	3	4	5
Probability	0.3	0.1	0.2	0.1	0.3

The spinner will be spun 80 times and the mean of the numbers it lands on will be calculated.
Find an estimate of the probability that this mean will be greater than 3.25.

Solution

$$\begin{aligned}\mathrm{E}(X) &= (1 \times 0.3) + (2 \times 0.1) + (3 \times 0.2) + (4 \times 0.1) + (5 \times 0.3) \\ &= 0.3 + 0.2 + 0.6 + 0.4 + 1.5 \\ &= 3\end{aligned}$$

and

$$\begin{aligned}\mathrm{E}(X^2) &= (1^2 \times 0.3) + (2^2 \times 0.1) + (3^2 \times 0.2) + (4^2 \times 0.1) + (5^2 \times 0.3) \\ &= 0.3 + 0.4 + 1.8 + 1.6 + 7.5 \\ &= 11.6.\end{aligned}$$

Now,

$$\begin{aligned}\sigma^2 &= \mathrm{Var}(X) \\ &= 11.6 - 3^2 \\ &= 2.6\end{aligned}$$

and, by the Central Limit Theorem, we have

$$\overline{X} \approx\sim \mathrm{N}\left(3, \frac{2.6}{80}\right).$$

Finally,

$$\begin{aligned}\mathrm{P}(\overline{X} > 3.25) &= \mathrm{P}\left(Z > \frac{3.25 - 3}{\sqrt{\frac{2.6}{80}}}\right) \\ &= \mathrm{P}\left(Z > \tfrac{5\sqrt{13}}{5}\right) \\ &= 0.082\,758\,928\,37 \text{ (FCD)} \\ &= \underline{\underline{0.082\,8 \text{ (4 dp)}}}.\end{aligned}$$

4. Liam and Simone are studying the distribution of oak trees in some woodland. They divided the woodland into 80 equal squares and recorded the number of oak trees in each square. The results are summarised in Table 1 below.

Number of oak trees in a square	0	1	2	3	4	5	6	7 or more
Frequency	1	4	21	23	13	11	7	0

Table 1: the distribution of oak trees

Liam believes that the oak trees were deliberately planted, with 6 oak trees per square and that a constant proportion p of the oak trees survived.

(a) Suggest the model Liam should use to describe the number of oak trees per square. (2)

Solution

Let X be the number of oak trees in a square.
Then $\underline{\underline{X \sim \mathrm{B}(6, p)}}$.

Liam decides to test whether or not his model is suitable and calculates the expected frequencies given in Table 2.

Number of oak trees	0 or 1	2	3	4	5	6
Expected frequency	5.53	14.89	24.26	22.24	10.87	2.21

Table 2: the expected frequencies

(b) Showing your working clearly, complete the test using a 5% level of significance. You should state your critical value and conclusion clearly. (7)

Solution

We have to merge the last two groups in the table (why?):

Number of oak trees	0 or 1	2	3	4	5 or 6
Expected frequency	5.53	14.89	24.26	22.24	13.08
$\frac{(O_i - E_i)^2}{E_i}$	0.050 8	2.507 2	0.065 4	3.838 9	1.850 4

H_0: $X \sim \mathrm{B}(6, p)$ is a good model, i.e., there is no difference between the observed and expected values.

H_1: $X \sim \mathrm{B}(6, p)$ is not a good model, i.e., there is a difference between the observed and expected values.

$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 8.312\,7$.
There are $5 - 2 = 3$ (why?) degrees of freedom.
$\chi_3^2(5\%) = 7.815$.
As we have $\chi_3^2(5\%) < X^2$, we reject H_0; $X \sim \mathrm{B}(6, p)$ is $\underline{\underline{\text{not a good model}}}$.

Simone believes that a Poisson distribution could be used to model the number of oak trees per square.
She calculates the expected frequencies given in Table 3.

Number of oak trees in a square	0 or 1	2	3	4	5	6 or more
Expected frequency	12.69	16.07	s	14.58	t	9.37

Table 3: a Poisson distribution

(c) Find the value of s and the value of t, giving your answers to 2 decimal places. (4)

Solution

Let Y be the number of oak trees in a square. Then

$$\begin{aligned} & \mathrm{E}(Y) \\ = & \frac{(0 \times 1) + (1 \times 4) + (2 \times 21) + (3 \times 23) + (4 \times 13) + (5 \times 11) + (6 \times 8)}{80} \\ = & \frac{264}{80} \\ = & 3.3. \end{aligned}$$

Then $Y \sim \mathrm{Po}(3.3)$.

$$\begin{aligned} \text{Expected frequency for 3 oaks} &= 80 \times \frac{3.3^3\,\mathrm{e}^{-3.3}}{3!} \\ &= 17.672\,938\,49 \text{ (FCD)} \\ &= \underline{\underline{17.67 \text{ (4 sf)}}} \end{aligned}$$

and

$$\begin{aligned} \text{expected frequency for 5 oaks} &= 80 \times \frac{3.3^5\,\mathrm{e}^{-3.3}}{5!} \\ &= 9.622\,915\,009 \text{ (FCD)} \\ &= \underline{\underline{9.623 \text{ (4 sf)}}}. \end{aligned}$$

(d) Write down hypotheses to test the suitability of Simone's model. (1)

Solution

H_0: $Y \sim \mathrm{Po}(3.3)$ is a good model, i.e., there is no difference between the observed and expected values.

H_1: $Y \sim \mathrm{Po}(3.3)$ is not a good model, i.e., there is a difference between the observed and expected values.

The test statistic for this test is 8.749.

(e) Complete the test. Use a 5% level of significance and state your critical value and conclusion clearly. (3)

Solution

There are $6 - 2 = 4$ (why?) degrees of freedom.
$\chi^2_4(5\%) = 9.488$.
As we have $\chi^2_3(5\%) > X^2$, we fail to reject H_0; $Y \sim \mathrm{Po}(3.3)$ is a good model.

(f) Using the results of these tests, explain whether the origin of this woodland is likely to be cultivated or wild. (2)

Solution

The Poisson model has a better fit and so it suggests that oak trees occur at random. Hence, the forest is likely to be wild.

5. Information was collected about accidents on the *Seapron* bypass. It was found that the number of accidents per month could be modelled by a Poisson distribution with mean 2.5.

Following some work on the bypass, the numbers of accidents during a series of 3-month periods were recorded. The data were used to test whether or not there was a change in the mean number of accidents per month.

(a) Stating your hypotheses clearly and using a 5% level of significance, find the critical region for this test. You should state the probability in each tail. (5)

Solution

Let X be the number of accidents in a 3-month period.

H_0: There is no difference to the rate of accidents, i.e., $X \sim \mathrm{Po}(7.5)$ is a good model.

H_1: There is a difference to the rate of accidents, i.e., $X \sim \mathrm{Po}(7.5)$ is not a good model.

Level of significance: 5%.

$\mathrm{P}(X \leqslant 2) = 0.020\,3$ and

$$\begin{aligned}\mathrm{P}(X \geqslant 14) &= 1 - \mathrm{P}(X \leqslant 13) \\ &= 1 - 0.978\,4 \\ &= 0.021\,6.\end{aligned}$$

So, the critical regions are

$$\underline{\underline{X \leqslant 2 \text{ or } X \geqslant 14}}.$$

(b) State P(Type I error) using this test. (1)

Solution

$$0.020\,3 + 0.021\,6 = \underline{\underline{0.041\,9}}.$$

Data from the series of 3-month periods are recorded for 2 years.

(c) Find the probability that at least 2 of these 3-month periods give a significant result. (3)

Solution

Let Y be the number of 3-month periods with a significant result.
Then $Y \sim \mathrm{B}(8, 0.041\,9)$.
Finally,

$$\begin{aligned}\mathrm{P}(Y \geqslant 2) &= 1 - \mathrm{P}(Y \leqslant 1) \\ &= 1 - \left[(0.958\,1)^8 + 8(0.958\,1)^7(0.041\,9)\right] \\ &= 1 - 0.958\,462\,572\,9 \text{ (FCD)} \\ &= 0.041\,537\,427\,1 \text{ (FCD)} \\ &= \underline{\underline{0.041\,5 \text{ (4 dp)}}}.\end{aligned}$$

Given that the number of accidents per month on the bypass, after the work is completed, is actually 2.1 per month,

(d) find P(Type II error) for the test in part (a). (3)

Solution

$$3 \times 2.1 = 6.3$$

and let the W be the number of accidents in a 3-month period.
Then $W \sim \text{Po}(6.3)$.
Finally,

$$\begin{aligned}
\text{P(Type II error)} &= \text{P}(3 \leqslant W \leqslant 13) \\
&= \text{P}(W \leqslant 13) - \text{P}(W \leqslant 2) \\
&= 0.994\,514\,711\,4 - 0.049\,846\,493\,17 \text{ (FCD)} \\
&= 0.944\,668\,218\,42 \text{ (FCD)} \\
&= \underline{\underline{0.944\,7 \text{ (4 dp)}}}.
\end{aligned}$$

6. The discrete random variable X has probability generating function

$$G_X(t) = k \ln\left(\frac{2}{2-t}\right),$$

where k is a constant.

(a) Find the exact value of k. (1)

Solution

$$\begin{aligned}
G_X(1) = 1 &\Rightarrow k \ln\left(\frac{2}{2-1}\right) = 1 \\
&\Rightarrow k \ln 2 = 1 \\
&\Rightarrow \underline{\underline{k = \frac{1}{\ln 2}}}.
\end{aligned}$$

(b) Find the exact value of $\text{Var}(X)$. (7)

Solution

$$G_X(t) = \frac{1}{\ln 2}\ln\left(\frac{2}{2-t}\right) \Rightarrow G_X(t) = \frac{1}{\ln 2}\cdot[\ln 2 - \ln(2-t)]$$
$$\Rightarrow G'_X(t) = \frac{1}{\ln 2}\cdot\frac{-1}{2-t}\cdot(-1)$$
$$\Rightarrow G'_X(t) = \frac{1}{\ln 2}(2-t)^{-1}$$
$$\Rightarrow G''_X(t) = \frac{1}{\ln 2}(2-t)^{-2}.$$

Now,

$$\begin{aligned}\text{Var}(X) &= G''_X(1) + G'_X(1) - [G'_X(1)]^2 \\ &= \frac{1}{\ln 2} + \frac{1}{\ln 2} - \left[\frac{1}{\ln 2}\right]^2 \\ &= \underline{\underline{\frac{2}{\ln 2} - \frac{1}{(\ln 2)^2}}}.\end{aligned}$$

(c) Find $\text{P}(X = 3)$. (4)

Solution

$$G''_X(t) = \frac{1}{\ln 2}(2-t)^{-2} \Rightarrow G'''_X(t) = \frac{2}{\ln 2}(2-t)^{-3}.$$

Now,

$$G_X(t) = G_X(0) + G'_X(0)t + \tfrac{1}{2!}G''_X(0)t^2 + \tfrac{1}{3!}G'''_X(0)t^3 + \dots$$

and hence

$$\begin{aligned}\text{P}(X = 3) &= \frac{1}{3!}G'''_X(0) \\ &= \frac{1}{6}\cdot\frac{1}{4\ln 2} \\ &= \underline{\underline{\frac{1}{24\ln 2}}}.\end{aligned}$$

7. A spinner can land on red or blue. When the spinner is spun, there is a probability of $\frac{1}{3}$ that it lands on blue. The spinner is spun repeatedly.

The random variable B represents the number of the spin when the spinner first lands on blue.

(a) Find (4)

(i) $P(B = 4)$,

Solution
$B \sim \text{Geo}(\frac{1}{3})$. Then,

$$\begin{aligned} P(B = 4) &= (\tfrac{2}{3})^3(\tfrac{1}{3}) \\ &= \underline{\underline{\tfrac{8}{81}}}. \end{aligned}$$

(ii) $P(B \leqslant 5)$.

Solution

$$\begin{aligned} & P(B \leqslant 5) \\ = \;& P(B = 1) + P(B = 2) + P(B = 3) + P(B = 4) + P(B = 5) \\ = \;& (\tfrac{1}{3}) + (\tfrac{2}{3})(\tfrac{1}{3}) + (\tfrac{2}{3})^2(\tfrac{1}{3}) + (\tfrac{2}{3})^3(\tfrac{1}{3}) + (\tfrac{2}{3})^4(\tfrac{1}{3}) \\ = \;& \tfrac{1}{3} + \tfrac{2}{9} + \tfrac{4}{27} + \tfrac{8}{81} + \tfrac{16}{243} \\ = \;& \underline{\underline{\tfrac{211}{243}}}. \end{aligned}$$

(b) Find $E(B^2)$. (3)

Solution

$$E(B) = \frac{1}{\frac{1}{3}} = 3$$

and

$$\text{Var}(B) = \frac{1 - \frac{1}{3}}{(\frac{1}{3})^2} = 6.$$

Now,

$$\begin{aligned} \text{Var}(B) = E(B^2) - E(B)^2 &\Rightarrow E(B^2) = \text{Var}(B) + E(B)^2 \\ &\Rightarrow E(B^2) = 6 + 3^2 \\ &\Rightarrow \underline{\underline{E(B^2) = 15}}. \end{aligned}$$

Steve invites Tamara to play a game with this spinner.

Tamara must choose a colour, either red or blue.

Steve will spin the spinner repeatedly until the spinner first lands on the colour Tamara has chosen. The random variable X represents the number of the spin when this occurs.

If Tamara chooses red, her score is e^X.

If Tamara chooses blue, her score is X^2.

(c) State, giving your reasons and showing any calculations you have made, which colour you would recommend that Tamara chooses. (5)

Solution

Let X be the number of the spin when it first lands on red.
$X \sim \text{Geo}(\frac{2}{3})$.
Now,

$$\begin{aligned}
\mathrm{E}(\mathrm{e}^X) &= \sum_{x=1}^{\infty} \mathrm{e}^x (\tfrac{1}{3})^{x-1} (\tfrac{2}{3}) \\
&= \tfrac{2e}{3} \sum_{x=1}^{\infty} (\tfrac{\mathrm{e}}{3})^{x-1} \\
&= \frac{2e}{3} \sum_{x=0}^{\infty} (\tfrac{\mathrm{e}}{3})^{x} \\
&= \frac{2e}{3} \cdot \frac{1}{1 - \frac{\mathrm{e}}{3}} \\
&= \frac{2\mathrm{e}}{3 - \mathrm{e}} \\
&= 19.297\,880\,67 \text{ (FCD)} \\
&> 15 \\
&= \mathrm{E}(X^2).
\end{aligned}$$

Hence, Tamara should choose red since it has the greater expected score.