Dr Oliver Mathematics
Advance Level Mathematics
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. The discrete random variable X has the following probability distribution

x	2	4	7	10
$\text{P}(X = x)$	a	b	0.1	c

where a, b, and c are probabilities.

The cumulative distribution function of X is $\text{F}(x)$ and $\text{F}(3) = 0.2$ and $\text{F}(6) = 0.8$.

(a) Find the value of a, the value of b, and the value of c. (3)

Solution

Remember, $\text{F}(x) = \text{P}(X \leqslant x)$ and so

$$\text{F}(3) = 0.2 \Rightarrow \text{P}(X \leqslant x) = 0.2 \Rightarrow \text{P}(X = 2) = \underline{\underline{0.2}},$$

$$\text{F}(6) = 0.8 \Rightarrow \text{P}(X = 2) + \text{P}(X = 4) = 0.8 \Rightarrow \text{P}(X = 6) = \underline{\underline{0.6}},$$

and

$$0.2 + 0.6 + 0.1 + c = 1 \Rightarrow c = \underline{\underline{0.1}}.$$

Hence,

x	2	4	7	10
$\text{P}(X = x)$	$\underline{\underline{0.2}}$	$\underline{\underline{0.6}}$	0.1	$\underline{\underline{0.1}}$

(b) Write down the value of $\text{F}(7)$. (1)

Solution

$$\text{F}(7) = \text{P}(X \leqslant 7) = \underline{\underline{0.9}}.$$

2. The following grouped frequency distribution summarises the number of minutes, to the nearest minute, that a random sample of 100 motorists were delayed by roadworks on a stretch of motorway one Monday.

Delay (minutes)	Number of motorists (f)	Delay midpoint (x)
$3-6$	38	4.5
$7-8$	25	7.5
$9-10$	18	9.5
$11-15$	12	13
$16-20$	7	18

You may use

$$\sum fx^2 = 8\,096.25.$$

A histogram has been drawn to represent these data.

The bar representing a delay of $(3-6)$ minutes has a width of 2 cm and a height of 9.5 cm.

(a) Calculate the width and the height of the bar representing a delay of $(11-15)$ minutes. (3)

Solution

Well, the $2.5-6.5$ covers 2 cm,

$$\text{area of } (3-6) \text{ bar} = 2 \times 9.5 = 19 \text{ cm}^2,$$

and

$$\frac{38}{19} = 2 \text{ persons per cm}^2.$$

On other hand, the $(11-15)$ goes from $10.5-15.5$, i.e., 2.5 cm, and we have

$$\text{area of } (11-15) \text{ bar} = \frac{12}{2 \times 2.5} = \underline{\underline{2\tfrac{2}{5} \text{ cm.}}}$$

(b) Use linear interpolation to estimate the median delay. (2)

Solution

Delay (minutes)	Number of motorists (f)	Cumulative Frequency
$3-6$	38	38
$7-8$	25	63
$9-10$	18	81
$11-15$	12	93
$16-20$	7	100

The median is

$$\frac{100+1}{2} = 50.5\text{th}$$

piece of data: it is the 12.5th piece of data in $7-8$ interval. Finally,

$$\text{median delay} = 6.5 + \frac{12.5}{25} \times 2 = \underline{\underline{7.5 \text{ minutes}}}.$$

(c) Calculate an estimate of the mean delay. (2)

Solution

$$\begin{aligned}\text{Mean delay} &\approx \frac{(4.5 \times 38) + (7.5 \times 25) + (9.5 \times 18) + (13 \times 12) + (18 \times 7)}{100} \\ &= \frac{811.5}{100} \\ &= \underline{\underline{8.115 \text{ minutes}}}.\end{aligned}$$

(d) Calculate an estimate of the standard deviation of the delays. (2)

Solution

$$\begin{aligned}\text{Standard deviation} &\approx \sqrt{\frac{8\,096.25}{100} - 8.115^2} \\ &= 3.887\,065\,088 \\ &= \underline{\underline{3.89 \text{ minutes (3 sf)}}}.\end{aligned}$$

One coefficient of skewness is given by

$$\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}.$$

(e) Evaluate this coefficient for the above data, giving your answer to 2 significant figures. (1)

Solution

$$\begin{aligned}\text{Skew} &= \frac{3(8.115 - 7.5)}{3.887\ldots} \\ &= 0.474\,651\,172\,1 \text{ (FCD)} \\ &= \underline{\underline{0.47 \text{ (2 sf)}}}.\end{aligned}$$

On the following Friday, the coefficient of skewness for the delays on this stretch of motorway was -0.22.

(f) State, giving a reason, how the delays on this stretch of motorway on Friday are different from the delays on Monday. (2)

Solution

The skewness for Monday and Friday are different (one is $+0.47$ and the other is -0.22) and suggests more longer delays on Fridays.

3. The random variable Y has a normal distribution with mean μ and standard deviation σ. $\mathrm{P}(Y > 17) = 0.4$.

Find

(a) $\mathrm{P}(\mu < Y < 17)$, (1)

Solution

$$\mathrm{P}(Y < 17) = 1 - \mathrm{P}(Y > 17) = 0.6$$

and

$$\begin{aligned}\mathrm{P}(\mu < Y < 17) &= \mathrm{P}(Y < 17) - \mathrm{P}(\mu < Y) \\ &= 0.6 - 0.5 \\ &= \underline{\underline{0.1}}.\end{aligned}$$

(b) $\mathrm{P}(\mu - \sigma < Y < 17)$ (4)

Solution

$$\begin{aligned}\mathrm{P}(Y > \mu - \sigma) &= \mathrm{P}\left(Z > \frac{(\mu - \sigma) - \mu}{\sigma}\right) \\ &= \mathrm{P}\left(Z > -1\right) \\ &= \mathrm{P}\left(Z < 1\right) \\ &= 0.8413\end{aligned}$$

and

$$\mathrm{P}(\mu - \sigma < Y < 17) = 0.8413 - 0.4 = \underline{\underline{0.4413}}.$$

4. A bag contains 64 coloured beads. There are r red beads, y yellow beads, 1 green bead and

$$r + y + 1 = 64.$$

Two beads are selected at random, one at a time without replacement.

(a) Find the probability that the green bead is one of the beads selected. (4)

Solution

$$\begin{aligned}\mathrm{P}(\text{neither is green}) &= \tfrac{63}{64} \times \tfrac{62}{63} \\ &= \tfrac{31}{32}\end{aligned}$$

and

$$\mathrm{P}(\text{at least one is green}) = 1 - \tfrac{31}{32} = \underline{\underline{\tfrac{1}{32}}}.$$

The probability that both of the beads are red is $\frac{5}{84}$.

(b) Show that r satisfies the equation (3)

$$r^2 - r - 240 = 0.$$

Solution

$$\begin{aligned} \mathrm{P}(RR) = \tfrac{5}{84} &\Rightarrow \frac{r}{64} \times \frac{r-1}{63} = \frac{5}{84} \\ &\Rightarrow r(r-1) = 240 \\ &\Rightarrow r^2 - r = 240 \\ &\Rightarrow \underline{\underline{r^2 - r - 240 = 0}}, \end{aligned}$$

as required.

(c) Hence show that the only possible value of r is 16. (2)

Solution

$$\left.\begin{array}{ll} \text{add to:} & -1 \\ \text{multiply to:} & -240 \end{array}\right\} -16,\ +15$$

$$\begin{aligned} r^2 - r - 240 = 0 &\Rightarrow r^2 - 16r + 15r - 240 = 0 \\ &\Rightarrow r(r-16) + 15(r-16) = 0 \\ &\Rightarrow (r+15)(r-16) = 0 \\ &\Rightarrow r = -15 \text{ or } r = 16; \end{aligned}$$

hence, $\underline{\underline{r = 16}}$.

(d) Given that at least one of the beads is red, find the probability that they are both red. (4)

Solution

$$\begin{aligned} \mathrm{P}(RR) &= \tfrac{16}{64} \times \tfrac{15}{63} = \tfrac{5}{84}, \\ \mathrm{P}(\text{neither is red}) &= \tfrac{48}{64} \times \tfrac{47}{63} = \tfrac{47}{84}, \text{ and} \\ \mathrm{P}(\text{at least one is red}) &= 1 - \tfrac{47}{84} = \tfrac{37}{84}. \end{aligned}$$

Finally, that at least one of the beads is red, the probability that they are both red

$$\frac{\frac{5}{84}}{\frac{37}{84}} = \underline{\underline{\tfrac{5}{37}}}.$$

5. The score when a spinner is spun is given by the discrete random variable X with the following probability distribution, where a and b are probabilities.

x	-1	0	2	4	5
$\mathrm{P}(X=x)$	b	a	a	a	b

(a) Explain why $\mathrm{E}(X) = 2$. (1)

Solution

It is symmetric about 2: 0 and 4 are equally spaced out (and their probability is a) and -1 and 5 are equally spaced out (and their probability is b).

(b) Find a linear equation in a and b. (1)

Solution

$$\begin{aligned} & [(-1)\times b] + (0\times a) + (2\times a) + (4\times a) + (5\times b) = 2 \\ \Rightarrow\ & -b + 0 + 2a + 4a + 5b = 2 \\ \Rightarrow\ & 6a + 4b = 2 \\ \Rightarrow\ & \underline{\underline{3a + 2b = 1 \quad (1)}}. \end{aligned}$$

Given that $\mathrm{Var}(X) = 7.1$,

(c) find a second equation in a and b and simplify your answer. (3)

Solution

$$\begin{aligned} & \mathrm{Var}(X) = 7.1 \\ \Rightarrow\ & [(-1)^2\times b] + (0^2\times a) + (2^2\times a) + (4^2\times a) + (5^2\times b) - 2^2 = 7.1 \\ \Rightarrow\ & b + 0 + 4a + 16a + 25b - 4 = 7.1 \\ \Rightarrow\ & \underline{\underline{20a + 26b = 11.1 \quad (2)}}. \end{aligned}$$

(d) Solve your two equations to find the value of a and the value of b. (3)

Solution

$13 \times (1) - (2)$:

$$19a = 1.9 \Rightarrow \underline{\underline{a = 0.1}}$$
$$\Rightarrow 0.3 + 2b = 1$$
$$\Rightarrow 2b = 0.7$$
$$\Rightarrow \underline{\underline{b = 0.35}}.$$

The discrete random variable $Y = 10 - 3X$.

(e) Find (3)

(i) $\mathrm{E}(Y)$,

Solution

$$\mathrm{E}(Y) = 10 - 3\,\mathrm{E}(X) = 10 - 3 \times 2 = \underline{\underline{4}}.$$

(ii) $\mathrm{Var}(Y)$.

Solution

$$\mathrm{Var}(Y) = 3^2\,\mathrm{Var}(X) = 9 \times 7.1 = \underline{\underline{63.9}}.$$

The spinner is spun once.

(f) Find $\mathrm{P}(Y > X)$. (3)

Solution

x	-1	0	2	4	5
$\mathrm{P}(X = x)$	0.35	0.1	0.1	0.1	0.35

$$Y > X \Rightarrow 10 - 3X > X$$
$$\Rightarrow 10 > 4X$$
$$\Rightarrow X < 2.5;$$

hence,

$$\mathrm{P}(Y > X) = 0.35 + 0.1 + 0.1 = \underline{\underline{0.55}}.$$

6. A group of climbers collected information about the height above sea level, h metres, and the air temperature, t°, at the same time at 8 different points on the same mountain. The data are summarised by

$$\sum h = 6\,370, \sum t = 61, \sum th = 31\,070, \text{ and } \sum t^2 = 693.$$

(a) Show that $S_{th} = -17\,501.25$ and $S_{tt} = 227.875$. (3)

Solution

$$S_{th} = 31\,070 - \frac{61 \times 6\,370}{8} = \underline{\underline{-17\,501.25}}$$

and

$$S_{tt} = 693 - \frac{61^2}{8} = \underline{\underline{227.875}}.$$

The product moment correlation coefficient for these data is -0.985.

(b) State, giving a reason, whether or not this value supports the use of a regression equation to predict the air temperature at different heights on this mountain. (1)

Solution

The product moment correlation coefficient is close to -1; <u>it does support the linear model</u>.

(c) Find the equation of the regression line of t on h, giving your answer in the form $t = a + bh$. Give the value of your coefficients to 3 significant figures. (7)

Solution

We need S_{hh}:

$$r = \frac{S_{th}}{\sqrt{S_{hh}S_{tt}}} \Rightarrow r^2 = \frac{S_{th}^2}{S_{hh}S_{tt}}$$
$$\Rightarrow S_{hh} = \frac{S_{th}^2}{r^2 S_{tt}}$$
$$\Rightarrow S_{hh} = \frac{(-17\,501.25)^2}{(-.985)^2 \times 227.875}$$
$$\Rightarrow S_{hh} = 1\,385\,380.258 \text{ (FCD)}$$

Now,

$$b = \frac{S_{th}}{S_{hh}} = -0.012\,632\,813\,19 \text{ (FCD)}$$

and

$$a = \frac{61}{8} - (-0.012\ldots) \times \frac{6\,370}{8} = 17.683\,877\,5 \text{ (FCD)};$$

hence, the equation of the regression line is

$$\underline{\underline{t = 17.7 - 0.012\,6h}}.$$

(d) Give an interpretation of your value of a. (1)

Solution

$\underline{\underline{17.7^\circ}}$ is an estimate of the $\underline{\underline{\text{temperature at sea level}}}$.

One of the climbers has just stopped for a short break before climbing the next 150 metres.

(e) Estimate the drop in temperature over this 150 metre climb. (2)

Solution

$$150 \times (-0.012\,6\ldots) = -1.894\,921\,979 \text{ (FCD)};$$

hence, it is $\underline{\underline{1.89^\circ \text{ (3 sf)}}}$.

7. Farmer Adam grows potatoes. The weights of potatoes, in grams, grown by Adam are normally distributed with a mean of 140 g and a standard deviation of 40 g.

Adam cannot sell potatoes with a weight of less than 92 g.

(a) Find the percentage of potatoes that Adam grows but cannot sell. (3)

Solution

$$\begin{aligned}
P(W < 92) &= P\left(Z < \frac{92 - 140}{40}\right) \\
&= P(Z < -1.2) \\
&= P(Z > 1.2) \\
&= 1 - P(Z < 1.2) \\
&= 1 - 0.8849 \text{ (from tables)} \\
&= 0.1151
\end{aligned}$$

and so percentage is 11.51%.

The upper quartile of the weight of potatoes **sold** by Adam is q_3.

(b) Find the probability that the weight of a randomly selected potato **grown** by Adam is more than q_3. (2)

Solution

$$\begin{aligned}
P(W > q_3) &= P(W > 92) \times P(W > q_3 | W > 92) \\
&= P(W > 92) \times P(W > q_3) \\
&= 0.8849 \times 0.25 \\
&= 0.221\,225.
\end{aligned}$$

(c) Find the lower quartile, q_1, of the weight of potatoes **sold** by Adam. (5)

Solution

$$\begin{aligned}
P(W > q_1) &= P(W > 92) \times P(W > q_1 | W > 92) \\
&= P(W > 92) \times P(W > q_1) \\
&= 0.8849 \times 0.75 \\
&= 0.663\,675
\end{aligned}$$

and

$$z = \Phi(0.663\,675) = 0.42.$$

Finally, (note the '−' sign!)

$$\begin{aligned}
\frac{q_1 - 140}{40} = -0.42 &\Rightarrow q_1 - 140 = -16.8 \\
&\Rightarrow q_1 = 123.2 \text{ g}.
\end{aligned}$$

Betty selects a random sample of 3 potatoes **sold** by Adam.

(d) Find the probability that one weighs less than q_1, one weighs more than q_3, and one has a weight between q_1 and q_3. (3)

Solution

$$\begin{aligned}\text{Probability} &= 3! \times \tfrac{1}{4} \times \tfrac{1}{2} \times \tfrac{1}{4} \\ &= \underline{\underline{\tfrac{3}{16}}}.\end{aligned}$$