S1 UK June 16 Kprime2

Leave blank

1. A biologist is studying the behaviour of bees in a hive. Once a bee has located a source of food, it returns to the hive and performs a dance to indicate to the other bees how far away the source of the food is. The dance consists of a series of wiggles. The biologist records the distance, d metres, of the food source from the hive and the average number of wiggles, w, in the dance.

Distance, d m	30	50	80	100	150	400	500	650
Average number of wiggles, w	0.725	1.210	1.775	2.250	3.518	6.382	8.185	9.555

[You may use $\sum w = 33.6$ $\sum dw = 13\,833$ $S_{dd} = 394\,600$ $S_{ww} = 80.481$ (to 3 decimal places)]

(a) Show that $S_{dw} = 5601$

(2)

(b) State, giving a reason, which is the response variable.

(1)

(c) Calculate the product moment correlation coefficient for these data.

(2)

(d) Calculate the equation of the regression line of w on d, giving your answer in the form $w = a + bd$

(4)

A new source of food is located 350 m from the hive.

(e) (i) Use your regression equation to estimate the average number of wiggles in the corresponding dance.

(ii) Comment, giving a reason, on the reliability of your estimate.

(2)

1(a) $S_{dw} = \sum dw - \dfrac{\sum d \sum w}{n}$

$\sum d = 1960$, $n = 8$

$\therefore S_{dw} = 13\,833 - \dfrac{1960 \times 33.6}{8} = 5601$

$\therefore S_{dw} = 5601$ as required.

blank

(b) Average number of wiggles, w, is the response variable, it is what is being measured.

(c) $r = \dfrac{S_{dw}}{\sqrt{S_{dd} S_{ww}}}$

$\therefore r = \dfrac{5601}{\sqrt{394600 \times 80.481}}$

$\therefore r = 0.994$ (3dp)

(d) $b = \dfrac{S_{dw}}{S_{dd}} = \dfrac{5601}{394600} = 0.0142$ (3sf)

$\bar{w} = \dfrac{\sum w}{n} = \dfrac{33.6\cancel{003}}{8} = 4.2\cancel{0 (3sf)}$

$\bar{d} = \dfrac{1960}{8} = 245$

$\therefore a = 4.2 - \dfrac{5601}{394600}(245)$

Question 1 continued

$d = 0.722$ (3sf)

$\therefore \; W = 0.722 + 0.0142d$

(e)(i) $W \approx 0.7224\ldots + 0.01419\ldots(350)$
$= 5.6903\ldots$

$W \approx 5.69$ (3sf)

(ii) The estimate is fairly reliable since $d = 350$ falls within the data range $[30, 650]$.

Leave blank

2. The discrete random variable X has the following probability distribution, where p and q are constants.

x	-2	-1	$\frac{1}{2}$	$\frac{3}{2}$	2
$P(X = x)$	p	q	0.2	0.3	p

(a) Write down an equation in p and q

(1)

Given that $E(X) = 0.4$

(b) find the value of q

(3)

(c) hence find the value of p

(2)

Given also that $E(X^2) = 2.275$

(d) find $\mathrm{Var}(X)$

(2)

Sarah and Rebecca play a game.

A computer selects a single value of X using the probability distribution above.

Sarah's score is given by the random variable $S = X$ and Rebecca's score is given by the random variable $R = \dfrac{1}{X}$

(e) Find $E(R)$

(3)

Sarah and Rebecca work out their scores and the person with the higher score is the winner. If the scores are the same, the game is a draw.

(f) Find the probability that

(i) Sarah is the winner,

(ii) Rebecca is the winner.

(4)

~~2 (a) $-2p - q + \frac{1}{2}(0.2) + \frac{3}{2}(0.3) + 2p$~~

~~$= -q + \frac{11}{20} = 1$~~

(a) $p + q + 0.2 + 0.3 + p = 1$

2 ~~(a) $p + 0$~~

$2p + q = \frac{1}{2}$

P 4 6 6 7 3 A 0 6 2 4

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Leave blank

Question 2 continued

(b) $-\cancel{2p}$ $E(X) = -2p - q + \frac{1}{2}(0.2) + \frac{3}{2}(0.3) + 2p$

$\therefore E(X) = \frac{11}{20} - q = 0.4$

$\Rightarrow q = \frac{3}{20} = 0.15$

(c) $2p + \frac{3}{20} = \frac{1}{2} \Rightarrow p = \frac{7}{40} = 0.175$

(d) $V(X) = E(X^2) - [E(X)]^2$

$\therefore Var(X) = 2.115$

(e) $E(R) = -\frac{1}{2}\left(\frac{7}{40}\right) - \frac{3}{20} + 2(0.2) + \frac{2}{3}(0.3) + \frac{1}{2}\left(\frac{7}{40}\right)$

$\therefore E(R) = 0.45$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

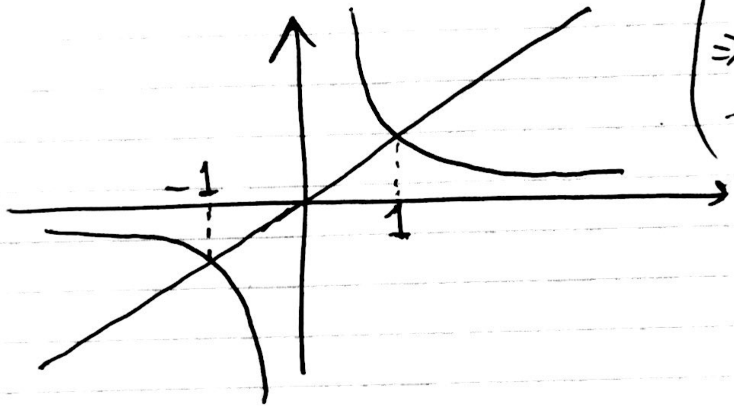
DO NOT WRITE IN THIS AREA

Question 2 continued

(f) Sarah wins if $S > R$

$S > R \Rightarrow X > \frac{1}{X}$

$\left(X = \frac{1}{X} \Rightarrow X^2 = 1 \therefore X = \pm 1 \right)$



$\therefore X > \frac{1}{X}$ for $X > 1$

$-1 < X < 0$

(i) $P(S>R) = P(X>1) + P(-1<X<0)$

$= 0.3 + 0.175 = 0.475$

$P(\text{Sarah wins}) = \underline{0.475}$

(ii) $P(R>S) = P\left(\frac{1}{X} > X\right) = P(0<X<1) + P(X<-1)$

$\therefore P(\text{Rebecca wins}) = 0.2 + 0.175 = \underline{0.375}$

(Total 15 marks)

Leave blank

Q2

Leave blank

3. Before going on holiday to *Seapron*, Tania records the weekly rainfall (x mm) at *Seapron* for 8 weeks during the summer. Her results are summarised as

$$\sum x = 86.8 \qquad \sum x^2 = 985.88$$

(a) Find the standard deviation, σ_x, for these data.

(3)

Tania also records the number of hours of sunshine (y hours) per week at *Seapron* for these 8 weeks and obtains the following

$$\bar{y} = 58 \qquad \sigma_y = 9.461 \text{ (correct to 4 significant figures)} \qquad \sum xy = 4900.5$$

(b) Show that $S_{yy} = 716$ (correct to 3 significant figures)

(1)

(c) Find S_{xy}

(2)

(d) Calculate the product moment correlation coefficient, r, for these data.

(2)

During Tania's week-long holiday at *Seapron* there are 14 mm of rain and 70 hours of sunshine.

(e) State, giving a reason, what the effect of adding this information to the above data would be on the value of the product moment correlation coefficient.

(2)

3.

(a) $\sigma_x = \sqrt{\frac{985.88}{8} - \left(\frac{86.8}{8}\right)^2}$

$\therefore \sigma_x = 2.35$ (3sf)

(b) $\sigma_y = \sqrt{\frac{\sum y^2}{8} - 58^2} = 9.461$

$\therefore 9.461^2 = \frac{\sum y^2}{8} - 3364$

$\therefore \sum y^2 = 27628.08...$

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Question 3 continued

$\bar{y} = 58 = \dfrac{\sum y}{8}$

$\therefore \sum y = 58 \times 8 = 464$

$\therefore S_{yy} = \sum y^2 - \dfrac{(\sum y)^2}{n}$

$\therefore S_{yy} = 27628.08\ldots - \dfrac{464^2}{8}$

$= 716.0841\ldots$

$\Rightarrow S_{yy} = 716$ (3sf) as required.

(c) $S_{xy} = \sum xy - \dfrac{\sum x \sum y}{n}$

$\sum xy = 4900.5$

$\sum x = 86.8$

$\sum y = 464$

$n = 8$

$S_{xy} = -\dfrac{1339}{10}$

$S_{xy} = -133.9$

Leave blank

Question 3 continued

(d) $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$ | $S_{xx} = \sum x^2 - \dfrac{(\sum x)^2}{n}$ $\therefore$ $S_{xx} = 44.1$

$\therefore r = \dfrac{-133.9}{\sqrt{44.1 \times 716}}$

$r = -0.754$ (3sf)

(e) ~~$x = 14$~~ $\bar{x} = 10.85$

$(x = 14) > (\bar{x} = 10.85)$

$S_{xx} = \sum (x - \bar{x})^2$ will increase

$S_{yy} = \sum (y - \bar{y})^2$ will increase, $y = 70 > \bar{y}$

$S_{xy} = \sum (x - \bar{x})(y - \bar{y})$ increases.

$\therefore$ correlation coefficient r increases

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

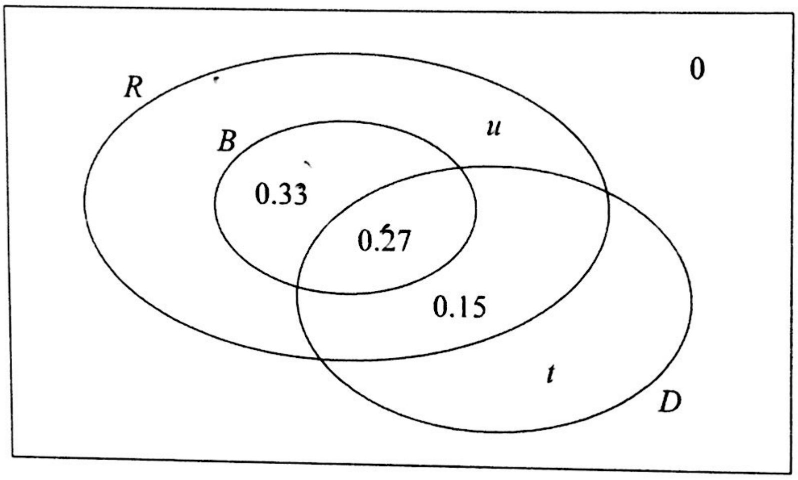
4. The Venn diagram shows the probabilities of customer bookings at Harry's hotel.

R is the event that a customer books a room

B is the event that a customer books breakfast

D is the event that a customer books dinner

u and t are probabilities.



(a) Write down the probability that a customer books breakfast but does not book a room.

(1)

Given that the events B and D are independent

(b) find the value of t

(4)

(c) hence find the value of u

(2)

(d) Find

(i) $P(D|R \cap B)$

(ii) $P(D|R \cap B')$

(4)

A coach load of 77 customers arrive at Harry's hotel.

Of these 77 customers

40 have booked a room and breakfast

37 have booked a room without breakfast

(e) Estimate how many of these 77 customers will book dinner.

(2)

Question 4 continued

4(a). 0

(B is inside R !!)

(b) $P(B \cap D) = P(B) \times P(D)$ | $P(B) = 0.33 + 0.27 = 0.6$

$\therefore 0.27 = 0.6 \times (0.15 + 0.27 + t)$

$0.27 = 0.6 \times (0.42 + t)$

$t = 0.03$

(c) $u = 1 - 0.33 - 0.27 - 0.15 - 0.03$

$\therefore u = 0.22$

(d)(i) $P(R \cap B) = P(B)$

B and D are independent

$\therefore P(D \mid R \cap B) = P(D \mid B) = P(D) =$

$0.27 + 0.15 + 0.03$

$= 0.45$

Question 4 continued

(ii) $P(R \cap B') = 0.22 + 0.15 = 0.37$

$$P(D \mid R \cap B') = \frac{P(D \cap R \cap B')}{P(D)}$$

$$= \frac{P(D \cap R)}{P(D)} = \frac{0.15}{0.37} = \frac{15}{37}$$

$$\therefore \; P(D \mid R \cap B') = \frac{15}{37}$$

~~(e)~~

~~Dinners booked = $0.18 \times 40 + 37 \times \frac{15}{37}$~~

=

(e) $P(D \mid R \cap B) \times 40 + P(D \mid R \cap B') \times 37$

$$= 40 \times 0.45 + 37 \times \frac{15}{37}$$

$= 33$ have dinner.

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

5. A midwife records the weights, in kg, of a sample of 50 babies born at a hospital. Her results are given in the table below.

Weight (w kg)	Frequency (f)	Weight midpoint (x)
$0 \leqslant w < 2$	1	1
$2 \leqslant w < 3$	8	2.5
$3 \leqslant w < 3.5$	17	3.25
$3.5 \leqslant w < 4$	17	3.75
$4 \leqslant w < 5$	7	4.5

1
9
26

[You may use $\sum fx^2 = 611.375$]

A histogram has been drawn to represent these data.

The bar representing the weight $2 \leqslant w < 3$ has a width of 1 cm and a height of 4 cm.

(a) Calculate the width and height of the bar representing a weight of $3 \leqslant w < 3.5$
(3)

(b) Use linear interpolation to estimate the median weight of these babies.
(2)

(c) (i) Show that an estimate of the mean weight of these babies is 3.43 kg.

(ii) Find an estimate of the standard deviation of the weights of these babies.
(3)

Shyam decides to model the weights of babies born at the hospital, by the random variable W, where $W \sim N(3.43, 0.65^2)$

(d) Find $P(W < 3)$
(3)

(e) With reference to your answers to (b), (c)(i) and (d) comment on Shyam's decision.
(3)

A newborn baby weighing 3.43 kg is born at the hospital.

(f) Without carrying out any further calculations, state, giving a reason, what effect the addition of this newborn baby to the sample would have on your estimate of the

(i) mean,

(ii) standard deviation.
(3)

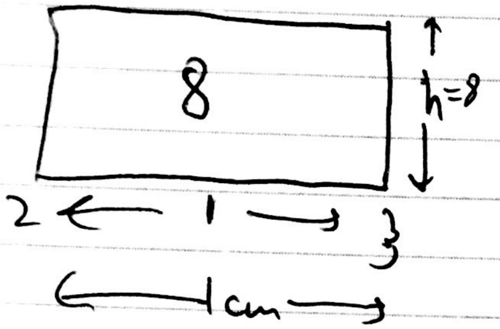
blank

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

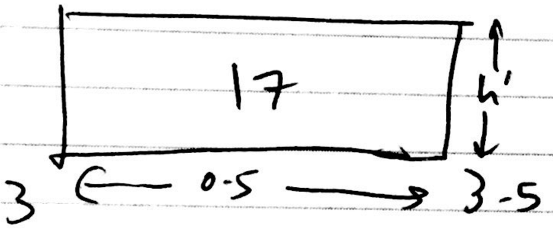
DO NOT WRITE IN THIS AREA

5(a)



$h \times 1 = 8 \quad \therefore h = 8$

~~Halved by s.f $\frac{1}{2}$~~



$0.5h' = 17$

$h' = 34$

$\therefore$ Height $= 17$ cm

Width $= 0.5$ cm

(b) $\Sigma f = 50$ median $= 25^{th}$ value

$\Rightarrow \dfrac{3.5 - m}{3.5 - 3} = \dfrac{26 - 25}{26 - 9}$

$\Rightarrow m = 3.47$ (3 s.f)

Leave blank

Question 5 continued

c(i) mean $= \dfrac{\sum fx}{\sum f} = \dfrac{171.5}{50} = 3.43\text{kg}$

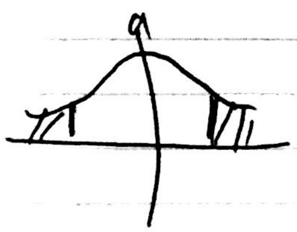
$\therefore$ mean $= 3.43$ kg
as required

(ii) $\sigma = \sqrt{\dfrac{\sum fx^2}{\sum f} - (3.43)^2}$

$\sigma = \sqrt{\dfrac{611.375}{50} - 3.43^2}$

$\sigma \approx 0.68$ (2dp)

(d) $P(W<3) = P\left(Z < \dfrac{3-3.43}{0.65}\right) = P(Z<-0.66)$



$= 1 - P(Z<0.66)$

$= 1 - 0.7454$

$= 0.2546$

Question 5 continued

(e) mean < median (3.43 < 3.47)

∴ Data is negatively skewed.

∴ Shyam's decision to use the normal distribution model is not sensible because the data is not symmetrical.

(f) (i) Mean stays the same because the weight of the newborn added baby is equal to the mean itself (=3.43 kg). Therefore, the mean will remain unchanged.

· Simple example:

Consider: 1, 2, 3, 4,

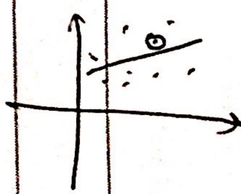
mean = $\frac{1+2+3+4}{4}$ = 2.5

Consider: 1, 2, 3, 4, 2.5

mean = $\frac{1+2+3+4+2.5}{5}$ = 2.5

mean does not change!

(ii) S.D decreases since the variance will decrease the closer the new added value is to the mean. New added value is the mean! Therefore the variance will decrease, hence lowering the s.d.



Q5

(Total 17 marks)

P46673A02124

6. The time, in minutes, taken by men to run a marathon is modelled by a normal distribution with mean 240 minutes and standard deviation 40 minutes.

(a) Find the proportion of men that take longer than 300 minutes to run a marathon. **(3)**

Nathaniel is preparing to run a marathon. He aims to finish in the first 20% of male runners.

(b) Using the above model estimate the longest time that Nathaniel can take to run the marathon and achieve his aim. **(3)**

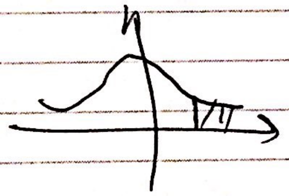
The time, W minutes, taken by women to run a marathon is modelled by a normal distribution with mean μ minutes.

Given that $P(W < \mu + 30) = 0.82$

(c) find $P(W < \mu - 30 \mid W < \mu)$ **(3)**

6(a) $T \sim N(240, 40^2)$

$P(T > 300) = P\left(Z > \frac{300-240}{40}\right) = P(Z > 1.5)$

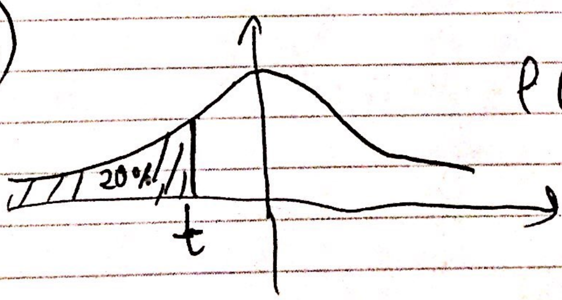


$= 1 - \Phi(1.5)$

$= 1 - 0.9332$

$= 0.0668$

(b)

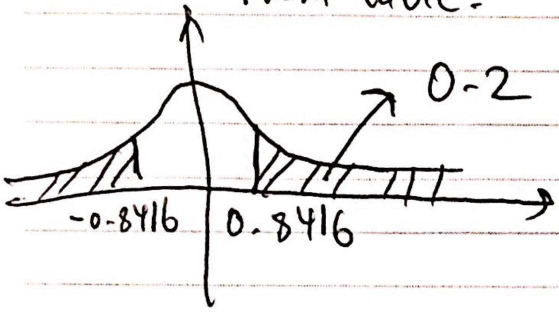


$P(T \leq t) = 0.2$

$\therefore P\left(Z \leq \frac{t-240}{40}\right) = 0.2$

Question 6 continued

From table:

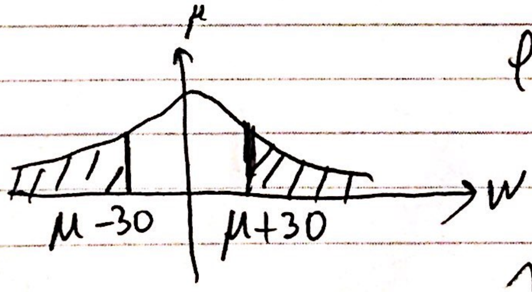


$$\therefore \frac{t-240}{40} = -0.8416$$

$$\Rightarrow t = \frac{25792}{125}$$

$$t = 206.336$$

(c) $W \sim N(\mu, \sigma^2)$



$$P(W < \mu-30) = 1 - P(W < \mu+30)$$

$$= 1 - 0.82 = 0.18$$

$$P(W < \mu - 30) = 0.18$$

$$P(W < \mu) = 0.5$$

$$P(W < \mu-30 \mid W < \mu) = \frac{P(W < \mu - 30)}{P(W < \mu)} = \frac{0.18}{0.5} = 0.36$$