S1 MAY 2013 (R)

1. Sammy is studying the number of units of gas, g, and the number of units of electricity, e, used in her house each week. A random sample of 10 weeks use was recorded and the data for each week were coded so that $x = \frac{g-60}{4}$ and $y = \frac{e}{10}$. The results for the coded data are summarised below

$$\sum x = 48.0 \quad \sum y = 58.0 \quad S_{xx} = 312.1 \quad S_{yy} = 2.10 \quad S_{xy} = 18.35$$

(a) Find the equation of the regression line of y on x in the form $y = a + bx$.

Give the values of a and b correct to 3 significant figures.

(4)

(b) Hence find the equation of the regression line of e on g in the form $e = c + dg$.

Give the values of c and d correct to 2 significant figures.

(4)

(c) Use your regression equation to estimate the number of units of electricity used in a week when 100 units of gas were used.

(2)

a) $b = \frac{Sxy}{Sxx} = 0.0589 \qquad a = \bar{y} - b\bar{x} = 5.52$

$\Rightarrow y = 5.52 + 0.0589x$

b) $\frac{e}{10} = 5.52 + 0.0589\left(\frac{g-60}{4}\right) = \cancel{-19.48}\ 4.64 + 0.0147g$

$\therefore e = \cancel{-19.48 +}\ 0.147g + 46.4$

c) $g = 100 \Rightarrow e = 61.1$

2. The discrete random variable X takes the values 1, 2 and 3 and has cumulative distribution function $F(x)$ given by

x	1	2	3
$F(x)$	0.4	0.65	1

(a) Find the probability distribution of X.

(3)

(b) Write down the value of $F(1.8)$.

(1)

a)

x	1	2	3
$P(x)$	0.4	0.25	0.35

b) $F(1.8) = 0.4$

3. An agriculturalist is studying the yields, y kg, from tomato plants. The data from a random sample of 70 tomato plants are summarised below.

Yield (y kg)	Frequency (f)		Yield midpoint (x kg)
$0 \leqslant y < 5$	16	0	2.5
$5 \leqslant y < 10$	24	16	7.5
$10 \leqslant y < 15$	14	40	12.5
$15 \leqslant y < 25$	12	54	20
$25 \leqslant y < 35$	4	66	30
		70	

(You may use $\sum fx = 755$ and $\sum fx^2 = 12037.5$)

A histogram has been drawn to represent these data.

The bar representing the yield $5 \leqslant y < 10$ has a width of 1.5 cm and a height of 8 cm.

(a) Calculate the width and the height of the bar representing the yield $15 \leqslant y < 25$

(3)

(b) Use linear interpolation to estimate the median yield of the tomato plants.

(2)

(c) Estimate the mean and the standard deviation of the yields of the tomato plants.

(4)

(d) Describe, giving a reason, the skewness of the data.

(2)

(e) Estimate the number of tomato plants in the sample that have a yield of more than 1 standard deviation above the mean.

(2)

a) width = 3cm height = 2cm (Area × 2 = freq)

b) $\frac{1}{2}n = 35$

5 Q_2 10

16 35 40

$\frac{Q_2 - 5}{5} = \frac{19}{24} \qquad Q_2 = 8.96$

c) $\bar{x} = \frac{755}{70} = 10.8 \qquad Sxx = 3894.2\ldots \quad \div n \ \sqrt{\ }$

$\sigma_x = 7.46$.

d) mean (10.8) > median (8.96) $\Rightarrow$ Positive skew

e) P(yield > 18.2) 32% of 15-25 below 18.2 ≃ 4

∴ 8 more than 18.2 ∴ 12 plants

4. The time, in minutes, taken to fly from London to Malaga has a normal distribution with mean 150 minutes and standard deviation 10 minutes.

(a) Find the probability that the next flight from London to Malaga takes less than 145 minutes.

(3)

The time taken to fly from London to Berlin has a normal distribution with mean 100 minutes and standard deviation d minutes.

Given that 15% of the flights from London to Berlin take longer than 115 minutes,

(b) find the value of the standard deviation d.

(4)

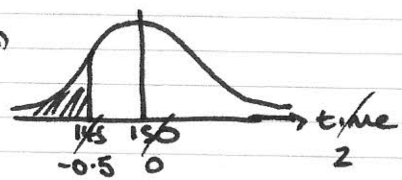
The time, X minutes, taken to fly from London to another city has a normal distribution with mean μ minutes.

Given that $P(X < \mu - 15) = 0.35$

(c) find $P(X > \mu + 15 \mid X > \mu - 15)$.

(3)

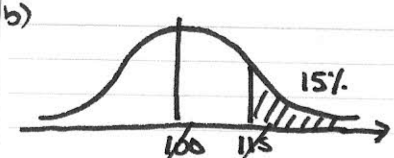
a)



$\phi(0.5) = 0.6915$

$\therefore P(t<145) = \underline{0.3085}$

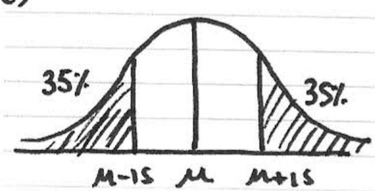
b)



$\frac{15}{d} = 1.0364$

$\therefore \underline{d = 14.5}$

c)



$P(X>\mu-15) = 0.65$

$\frac{P(X>\mu+15 \cap X>\mu-15)}{P(X>\mu-15)}$

$= \frac{0.35}{0.65} = \frac{7}{13}$

5. A researcher believes that parents with a short family name tended to give their children a long first name. A random sample of 10 children was selected and the number of letters in their family name, x, and the number of letters in their first name, y, were recorded.

The data are summarised as:

$$\sum x = 60, \ \sum y = 61, \ \sum y^2 = 393, \ \sum xy = 382, \ S_{xx} = 28$$

(a) Find S_{yy} and S_{xy}

(3)

(b) Calculate the product moment correlation coefficient, r, between x and y.

(2)

(c) State, giving a reason, whether or not these data support the researcher's belief.

(2)

The researcher decides to add a child with family name "Turner" to the sample.

(d) Using the definition $S_{xx} = \sum(x - \bar{x})^2$, state the new value of S_{xx} giving a reason for your answer.

(2)

Given that the addition of the child with family name "Turner" to the sample leads to an increase in S_{yy}

(e) use the definition $S_{xy} = \sum(x - \bar{x})(y - \bar{y})$ to determine whether or not the value of r will increase, decrease or stay the same. Give a reason for your answer.

(2)

a) $S_{yy} = 20.9$ $\quad S_{xy} = 16$

b) $r = \frac{16}{\sqrt{20.9 \times 28}} = 0.661$

c) Some evidence to suggest positive correlation, which does not support the researcher's belief

d) $x = 6 \Rightarrow x - \bar{x} = 0$, so no impact.

e) Increase in $S_{yy} \Rightarrow y >$ average y (6.1)

S_{xy}, unchanged $\therefore r$ will decrease.

6.

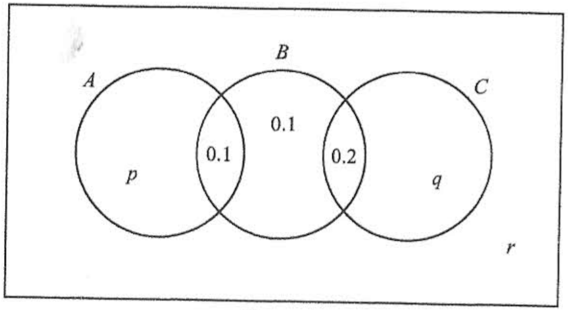


Figure 1

The Venn diagram in Figure 1 shows three events A, B and C and the probabilities associated with each region of B. The constants p, q and r each represent probabilities associated with the three separate regions outside B.

The events A and B are independent.

(a) Find the value of p. **(3)**

Given that $P(B|C) = \frac{5}{11}$

(b) find the value of q and the value of r. **(4)**

(c) Find $P(A \cup C | B)$. **(2)**

a) A, B independent $\Rightarrow P(A \cap B) = P(A) \times P(B)$

$P(A) = p + 0.1$, $P(B) = 0.4$, $P(A \cap B) = 0.1$

$\Rightarrow 0.1 = 0.4(p + 0.1) \Rightarrow 0.1 = 0.4p + 0.04 \quad \therefore p = 0.15$

b) $P(B|C) = \frac{P(B \cap C)}{P(C)} \Rightarrow \frac{5}{11} = \frac{0.2}{P(C)} \Rightarrow P(C) = 0.44$

$\Rightarrow q = 0.24 \quad r = 0.21$

c) $\frac{0.3}{0.4} = \frac{3}{4}$

7. The score S when a spinner is spun has the following probability distribution.

s	0	1	2	4	5
$P(S = s)$	0.2	0.2	0.1	0.3	0.2

(a) Find $E(S)$. **(2)**

(b) Show that $E(S^2) = 10.4$ **(2)**

(c) Hence find $Var(S)$. **(2)**

(d) Find

(i) $E(5S - 3)$,

(ii) $Var(5S - 3)$. **(4)**

(e) Find $P(5S - 3 > S + 3)$ **(3)**

The spinner is spun twice.

The score from the first spin is S_1 and the score from the second spin is S_2

The random variables S_1 and S_2 are independent and the random variable $X = S_1 \times S_2$

(f) Show that $P(\{S_1 = 1\} \cap X < 5) = 0.16$ **(2)**

(g) Find $P(X < 5)$. **(3)**

a) $E(S) = 0 + 0.2 + 0.2 + 1.2 + 1 = 2.6$

b) $E(S^2) = 0 + 0.2 + 0.4 + 4.8 + 5 = 10.4$ #

c) $V(S) = E(S^2) - E(S)^2 = 3.64$

d) i) $5E(S) - 3 = 10$ ii) $25V(S) = 91$

e)

$5S-3$	-3	2	7	17	22
$S+3$	3	4	5	7	8
p	0.2	0.2	0.1	0.3	0.2

$= \underline{0.6}$

f)

S_2 \ X	0	1	2	4	5
0	0 (0.04)	0 (0.04)	0 (0.02)	0 (0.06)	0 (0.04)
1	0 (0.04)	1 (0.04)	2 (0.02)	4 (0.06)	5 (0.04)
2	0 (0.02)	2 (0.02)	4 (0.01)	8 (0.03)	10 (0.02)
4	0 (0.06)	4 (0.06)	8 (0.03)	16 (0.09)	20 (0.06)
5	0 (0.04)	5 (0.04)	10 (0.02)	20 (0.06)	25 (0.04)

$$P((S=1) \cap X<5) = 0.04+0.04+0.02+0.06 = 0.16 \ \#$$

g) $$P(X<S) = 1-P(X \geq S) = 1-\begin{bmatrix} 0.04+0.02+0.06+0.04 \\ +0.03+0.09+0.06 \\ +0.03+0.02+0.04 \end{bmatrix}$$

$$= 1-0.43 = 0.57$$