

Monday 18 October 2021 – Afternoon

Paper
reference

9MA0/31

Mathematics

Advanced

PAPER 31: Statistics

1. (a) State one disadvantage of using quota sampling compared with simple random sampling. *(non-random)*

(1)

In a university 8% of students are members of the university dance club.

A random sample of 36 students is taken from the university.

The random variable X represents the number of these students who are members of the dance club.

- (b) Using a suitable model for X , find

(i) $P(X = 4)$

(ii) $P(X \geq 7)$

(3)

Only 40% of the university dance club members can dance the tango.

- (c) Find the probability that a student is a member of the university dance club and can dance the tango.

(1)

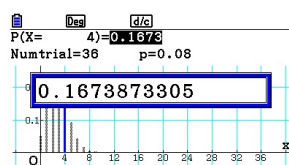
A random sample of 50 students is taken from the university.

- (d) Find the probability that fewer than 3 of these students are members of the university dance club and can dance the tango.

(2)

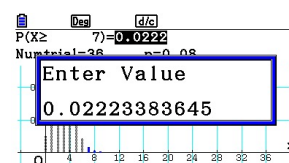
a) quota sampling is non-random,
 $\Rightarrow$ not all sampling unit has an equal chance of being selected.

b) $X \sim B(36, 0.08)$



b) $P(X=4) = 0.1674$

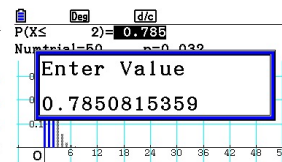
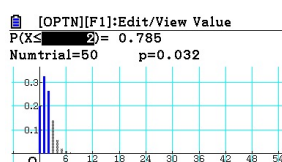
b) $P(X \geq 7) = 0.0222$



$P(X \leq 2) = 0.7851$

d) $X \sim B(50, 0.032)$

c) $0.08 \times 0.4 = 0.032$



2. Marc took a random sample of 16 students from a school and for each student recorded

- the number of letters, x , in their last name
- the number of letters, y , in their first name

His results are shown in the scatter diagram on the next page.

(a) Describe the correlation between x and y .

(1)

Marc suggests that parents with long last names tend to give their children shorter first names.

(b) Using the scatter diagram comment on Marc's suggestion, giving a reason for your answer.

(1)

The results from Marc's random sample of 16 observations are given in the table below.

x	3	6	8	7	5	3	11	3	4	5	4	9	7	10	6	6
y	7	7	4	4	6	8	5	5	8	4	7	4	5	5	6	3

(c) Use your calculator to find the product moment correlation coefficient between x and y for these data.

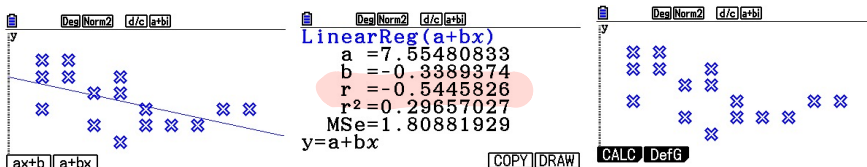
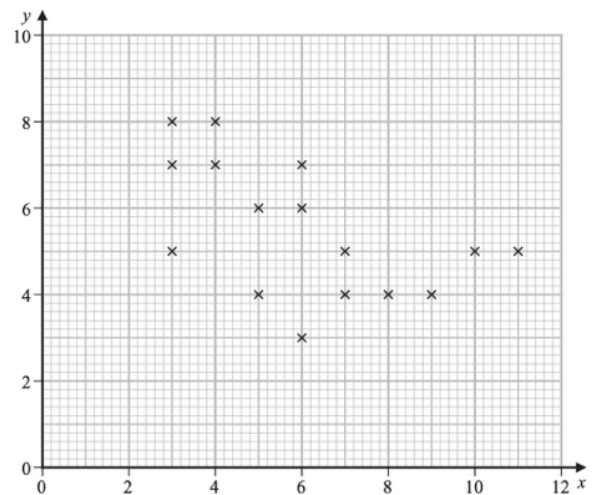
(1)

(d) Test whether or not there is evidence of a negative correlation between the number of letters in the last name and the number of letters in the first name.

You should

- state your hypotheses clearly
- use a 5% level of significance

(3)



a) weak negative correlation, as letters in last name ↑, letters in first name ↓.

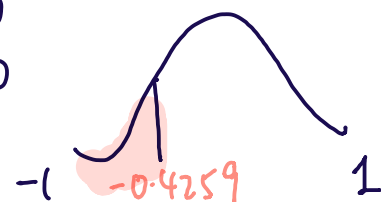
b) this is consistent to the scatter diagram, since data shows -ve correlation.

c) $p = -0.5446$

d) $n = 16$

$H_0: p = 0$

$H_1: p < 0$



$CV = p = -0.4259$

$CR: p < -0.4259$

$TS: p = -0.5446$

-0.5446 is in the CR,

Reject H_0 ,

enough evidence to conclude that there is negative correlation.

between letters in last name and letters in first name.

//

Product Moment Coefficient					Sample size, n
0.10	0.05	0.025	0.01	0.005	
0.8000	0.9000	0.9500	0.9800	0.9900	4
0.6870	0.8054	0.8783	0.9343	0.9587	5
0.6084	0.7293	0.8114	0.8822	0.9172	6
0.5509	0.6694	0.7545	0.8329	0.8745	7
0.5067	0.6215	0.7067	0.7887	0.8343	8
0.4716	0.5822	0.6664	0.7498	0.7977	9
0.4428	0.5494	0.6319	0.7155	0.7646	10
0.4187	0.5214	0.6021	0.6851	0.7348	11
0.3981	0.4973	0.5760	0.6581	0.7079	12
0.3802	0.4762	0.5529	0.6339	0.6835	13
0.3646	0.4575	0.5324	0.6120	0.6614	14
0.3507	0.4409	0.5140	0.5923	0.6411	15
0.3383	0.4259	0.4973	0.5742	0.6226	16

3. Stav is studying the large data set for September 2015

He codes the variable Daily Mean Pressure, x , using the formula $y = x - 1010$ translation

The data for all 30 days from Hurn are summarised by

$$\sum y = 214 \quad \sum y^2 = 5912$$

- (a) State the units of the variable x hPa. (1)
- (b) Find the mean Daily Mean Pressure for these 30 days. (2)
- (c) Find the standard deviation of Daily Mean Pressure for these 30 days. (3)

Stav knows that, in the UK, winds circulate

- in a **clockwise** direction around a region of **high** pressure
- in an **anticlockwise** direction around a region of **low** pressure

The table gives the Daily Mean Pressure for 3 locations from the large data set on 26/09/2015

Location	Heathrow	Hurn	Leuchars
Daily Mean Pressure	1029	1028	1028
Cardinal Wind Direction	<u>NE</u>	<u>E</u>	<u>W</u>

The Cardinal Wind Directions for these 3 locations on 26/09/2015 were, in random order,

W NE E

You may assume that these 3 locations were under a single region of pressure.

- (d) Using your knowledge of the large data set, place each of these Cardinal Wind Directions in the correct location in the table. Give a reason for your answer. (2)

b)

$$\frac{\sum y}{n} = \bar{y} = \frac{214}{30} = 7.1333$$

$$\Rightarrow \bar{x} = 1010 + 7.1333$$
$$\bar{x} = 1017.1333 \text{ hPa.}$$

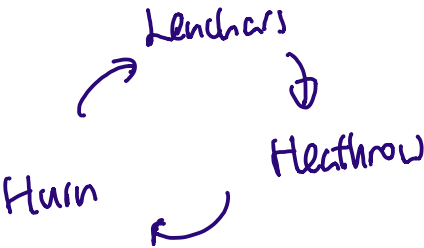
c)

$$\sigma_y = \sigma_x$$
$$\sigma_y = \sqrt{\frac{5912}{30} - \left(\frac{214}{30}\right)^2}$$
$$= 12.0906$$

$$\sigma_x = \underline{\underline{12.0906 \text{ hPa}}}$$

1000 (standard)
1028/1029 $\Rightarrow$ high

$\Rightarrow$ high pressure means clockwise



Variable	Typical value(s)
Gust (UK only)	8 – 52 nm
Rainfall	0 – 60 mm in UK, but more extreme maximums elsewhere (e.g. 102mm in Perth)
Pressure	988 – 1038 hPa
Wind Speed on Beaufort scale	Max is 'fresh' (5). Most Light or Moderate.
Sunshine (UK only)	0 – 16 hrs
Cloud Cover	0 – 8 oktas (i.e. full spread)

4. A large college produces three magazines.

One magazine is about green issues, one is about equality and one is about sports.

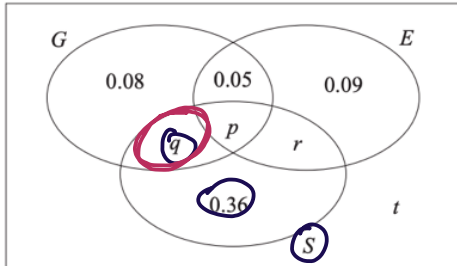
A student at the college is selected at random and the events G , E and S are defined as follows

G is the event that the student reads the magazine about green issues

E is the event that the student reads the magazine about equality

S is the event that the student reads the magazine about sports

The Venn diagram, where p , q , r and t are probabilities, gives the probability for each subset.



- (a) Find the proportion of students in the college who read exactly one of these magazines.

No students read all three magazines and $P(G) = 0.25$

- (b) Find

(i) the value of p

(ii) the value of q

Given that $P(S|E) = \frac{5}{12}$

- (c) find

(i) the value of r

(ii) the value of t

- (d) Determine whether or not the events $(S \cap E')$ and G are independent. Show your working clearly.

$$a) 0.08 + 0.09 + 0.36 = 0.53$$

$$b) 0.08 + 0.05 + p + q = 0.25$$

$$p = 0 //$$

$$\Rightarrow q = 0.12 //$$

$$(1) c) P(S|E) = \frac{5}{12}$$

$$\frac{p+r}{0.05+0.09+p+r} = \frac{5}{12}$$

(3)

$$\frac{r}{0.14+r} = \frac{5}{12}$$

(4)

$$12r = 5(0.14+r)$$

$$r = 0.1 //$$

(3)

$$\sum \text{prob} = 1$$

$$1 = 0.08 + 0.05 + 0.09 + 0.36 + t + p + q + r$$

$$t = 0.2 //$$

$$\begin{matrix} & \uparrow & \uparrow & \uparrow \\ & 0 & 0.12 & 0.1 \end{matrix}$$

$$P(G) = 0.25$$

$$P(S \cap E') = q + 0.36 = 0.48$$

$$P(G) \times P(S \cap E') = 0.12$$

$$P(G \cap (S \cap E')) = 0.12$$

$$P(G) \times P(S \cap E') = P(G \cap (S \cap E'))$$

therefore they are independent.

5. The heights of females from a country are normally distributed with

- a mean of 166.5 cm
- a standard deviation of 6.1 cm

Given that 1% of females from this country are shorter than k cm,

(a) find the value of k

(b) Find the proportion of females from this country with heights between 150 cm and 175 cm

A female, from this country, is chosen at random from those with heights between 150 cm and 175 cm

(c) Find the probability that her height is more than 160 cm

The heights of females from a different country are normally distributed with a standard deviation of 7.4 cm

Mia believes that the mean height of females from this country is less than 166.5 cm

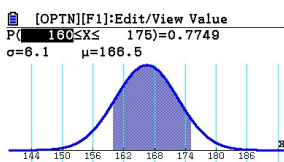
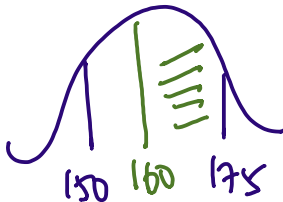
Mia takes a random sample of 50 females from this country and finds the mean of her sample is 164.6 cm

(d) Carry out a suitable test to assess Mia's belief.

You should

- state your hypotheses clearly
- use a 5% level of significance

c) $P(X > 160) \mid P(150 < X < 175)$

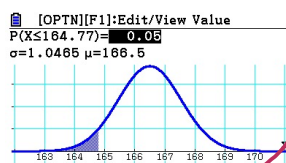


$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{0.7749}{0.9148} = \underline{\underline{0.8471}} \text{ (4dp)}$$

d) $H_0: \mu = 166.5$
 $H_1: \mu < 166.5$

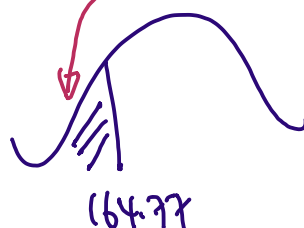
$$\bar{X} \sim N\left(166.5, \left(\frac{7.4}{\sqrt{50}}\right)^2\right)$$



TS: $\bar{X} = 164.6$

CR: $\bar{X} = 164.77$

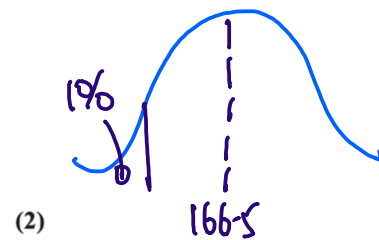
$$P(\bar{X} \leq 164.77) = 0.05$$



164.77 is in the CR.

Reject H_0

enough evidence to suggest $\mu < 166.5$.

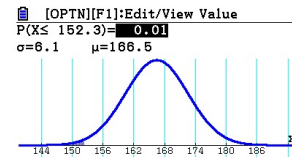


(2)

$$X \sim N(166.5, 6.1^2)$$

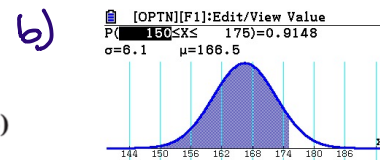
$$P(X \leq k) = 0.01$$

(4)



$$k = 152.3 //$$

(4)



$$P(150 \leq X \leq 175) = \underline{\underline{0.9148}}$$

6. The discrete random variable X has the following probability distribution

x	a	b	c
$P(X=x)$	$\log_{36} a$	$\log_{36} b$	$\log_{36} c$

where

- a, b and c are distinct integers ($a < b < c$)
- all the probabilities are greater than zero

(a) Find

- the value of a
- the value of b
- the value of c

Show your working clearly.

The independent random variables X_1 and X_2 each have the same distribution as X

(b) Find $P(X_1 = X_2)$

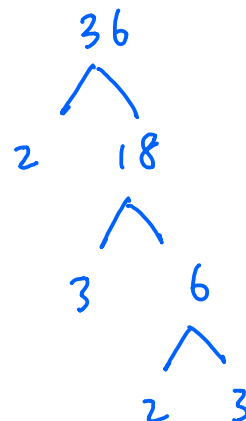
$$\log_{36} a + \log_{36} b + \log_{36} c = 1$$

$$abc = 36$$

$$\text{integer} \Rightarrow a < b < c$$

$$36 = 2^2 \times 3^2$$

(5)



(2)

b)

$$(\log_{36} 2)^2 + (\log_{36} 3)^2 + (\log_{36} 6)^2$$

$$36 = 2 \times 3 \times 6$$

$$= 2 \times 9 \times 2 \times \text{distinct}$$

$$= 1 \times 18 \times 2 \times \ln 1 = 0$$

$$= \underline{\underline{0.3814}}$$