1. (a) State one disadvantage of using quota sampling compared with simple random sampling.
(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) $P(X \geqslant 7)$
(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)

Solution 1a

Quota sampling is less random and therefore more likely to be biased.

Solution 1bi

$X \sim B(36, 0.08)$

no. of students sampled = 36
probability of being in dance club = 0.08

calculator

$P(X=4) = 0.167387...$

Solution 1bii

$P(X \geqslant 7) = 1 - P(X \leqslant 6) = 0.22233$

Solution 1c

P(Student is member of dance club and can dance tango)

$= 0.4 \times 0.08 = 0.032$

Solution 1d Y = members of club who can dance tango

$T \sim B(50, 0.032)$

calculator

$P(T<3) = 0.7850815$

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)

Solution 2a

Negative correlation

Solution 2b

Since the scatter diagram has a negative correlation, it seems compatible with Marc's claim.

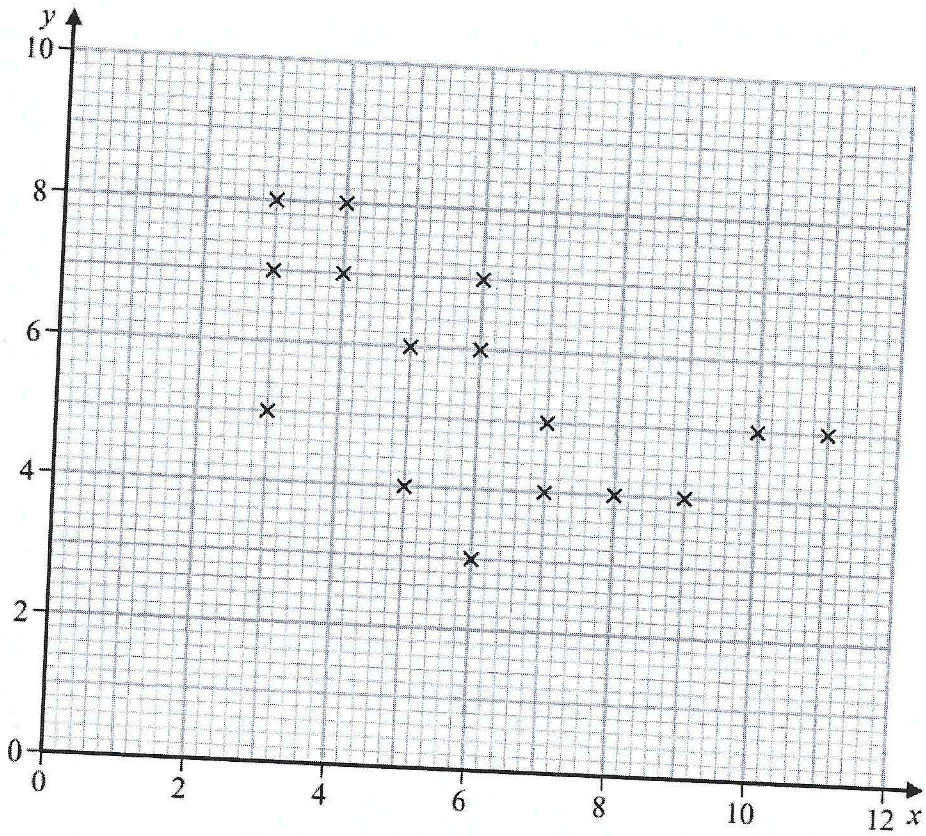
Solution 2c

$\Sigma x = 97$, $\Sigma y = 88$, $\Sigma x^2 = 681$, $\Sigma y^2 = 520$, $\Sigma xy = 502$

$n = 16$.

$$r = \frac{502 - (97 \times 88)/16}{\sqrt{\left(681 - \frac{97^2}{16}\right)\left(520 - \frac{88^2}{16}\right)}} = -0.54458266\ldots$$

Question 2 continued.



Solution 2d

$H_0: \rho = 0$

$H_1: \rho < 0$

$\alpha = 0.05$

The 5% significance 1 tail test critical value is -0.4259

$r = -0.544\ldots < -0.4259 \Rightarrow$ There is evidence to reject H_0.

Hence there is evidence of negative correlation between length of first names and length of second names

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

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

The data for all 30 days from Hurn are summarised by

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

(a) State the units of the variable x

(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	NE	E	W

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)

Solution 3a

Hectopascal (hPa)

Solution 3b

$\bar{x} = \bar{y} + 1010$, $\bar{y} = \frac{214}{30}$

$\bar{x} = 1017.1\dot{3}$

Solution 3c

Since $y = x - 1010$, $\sigma_x = \sigma_y$, $\sigma_y = \sqrt{\frac{5912}{30} - \left(\frac{214}{30}\right)^2} = 12.0905... = \sigma_x$

Solution 3d

As all of these pressures are nearly a standard deviation from the mean, we can consider them high pressures. So wind circulating clockwise.

Locations from North to South: Leuchars, Heathrow, Hurn

Solution 3d continued

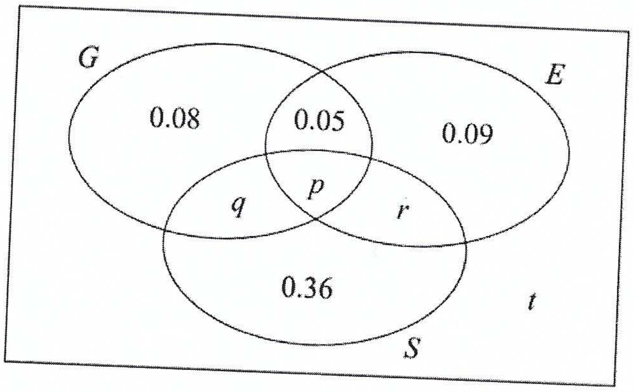
Locations from North to South are: Leuchars, Heathrow, Hurn.

So the directions of their wind is as written on the table.

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.

(1)

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

(b) Find

(i) the value of p

(ii) the value of q

(3)

Given that $P(S \mid E) = \dfrac{5}{12}$

(c) find

(i) the value of r

(ii) the value of t

(4)

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

(3)

Solution 4a

P(Proportion of students who read exactly one of these magazines)

$= 0.08 + 0.05 + 0.09 + 0.36$

$= 0.53$

Solution 4bi

$P(G \cap E \cap S) = p = 0$

Solution 4bii

$P(G) = 0.25 = 0.08 + 0.05 + p + q$

$\Rightarrow \quad 0.25 = 0.13 + 0 + q$

$\Rightarrow \quad q = 0.12$

Solution 4ci

$P(S \mid E) = p + r = \frac{5}{12}$

But $p = 0$.

So $r = \frac{5}{12}$

Solution 4cii

$P(S \mid E) = \frac{P(S \cap E)}{P(E)} = \frac{p + r}{p + r + 0.05 + 0.09}$

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

$\Rightarrow \quad 12r = 5r + 5 \times 0.14$

$\Rightarrow \quad 7r = 0.7$

$\Rightarrow \quad r = 0.1$

Solution 4d

$P(S \cap E') = 0.36 + q = 0.36 + 0.12$
$= 0.48$

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

$P(G) = 0.25$

$P(S \cap E') \times P(G) = 0.48 \times 0.25 = 0.12$

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

Hence 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

(2)

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

(1)

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

(4)

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

(4)

Solution 5a

Let $F \sim N(\boxed{166.5}, \boxed{6.1}^2)$

1% of females shorter than k cm

$\Rightarrow P(F<k) = 0.01$

$\Rightarrow \dfrac{k-\boxed{166.5}}{\boxed{6.1}} = -2.3263$

$\Rightarrow k = 152.3\text{cm}$

From tables
Percentage Points of the Normal distribution

p	z
0.0100	2.3263

As we are considering $P(Z<z)$, we must use $z=-2.3263$

Solution 5b

calculator

$P(150 < F < 175) = 0.91484$

Solution 5c

$F \sim N(166.5, 6.1)$

$P(F > 160 \mid 150 < F < 175)$

$= \dfrac{P((F>160) \cap (150 < F < 175))}{P(150 < F < 175)}$

$= \dfrac{P(160 < F < 175)}{P(150 < F < 175)} = \dfrac{0.7749487\ldots}{0.914840\ldots}$ calculator

$= 0.84708$

Solution 5d

Null hypothesis $H_0 : \mu = 166.5$

Mia's hypothesis $H_1 : \mu < 166.5$

Let X = height of females from this country

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

Central Limit Theorem applied to:
$X \sim N(166.5, 7.4^2)$
so $X \sim N(\mu, \sigma^2)$
$\Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$P(\bar{X} < 164.6) = 0.03472$
< 0.05

So we can reject H_0 and conlude that there is sufficient evidence to support Mia's belief that the mean height of females from this country is less than 166.5cm.

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

(i) the value of a

(ii) the value of b

(iii) the value of c

Show your working clearly.

(5)

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

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

(2)

Solution 6ai, ii, iii

$\sum P(X=x)=1$

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

$\Rightarrow \log_{36} abc = 1$

$\Rightarrow abc = 36$

As all probabilites are greater than 0,

$1 < a, b, c$

Prime factors of $36 = 2 \times 2 \times 3 \times 3$

Since we require 3 distinct integers that multiply to 36, we get

$36 = 2 \times 3 \times 6$

$\Rightarrow a = 2,\ b = 3,\ c = 6$ (as $a < b < c$)

Solution 6b

Three ways they can be equal:

$P(X_1 = X_2 = a) = (\log_{36} a)^2$, $[P(X_1=a) \times P(X_2=a)]$

$P(X_1 = X_2 = b) = (\log_{36} b)^2$, $[P(X_1=b) \times P(X_2=b)]$

$P(X_1 = X_2 = c) = (\log_{36} c)^2$, $[P(X_1=c) \times P(X_2=c)]$

So

$$P(X_1 = X_2) = (\log_{36} a)^2 + (\log_{36} b)^2 + (\log_{36} c)^2$$
$$= 0.381401\ldots$$