**Missing topics – Large Data Set questions, hypothesis tests for the PMCC and hypothesis test for the mean of a Normal distribution**

**2018**

**S1**

1. The discrete random variable *X* has the following probability distribution

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *x* | 2 | 4 | 7 | 10 |
| P(*X* = *x*) | *a* | *b* | 0.1 | *c* |

where *a*, *b* and *c* are probabilities.

The cumulative distribution function of *X* is F(*x*) and F(3) = 0.2 and F(6) = 0.8

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

**(3)**

(*b*)Write down the value of F(7).

**(1)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**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  = 8096.25)

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)**

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

**(2)**

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

**(2)**

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

**(2)**

|  |  |
| --- | --- |
| One coefficient of skewness is given by | 3(mean – median) |
| standard deviation |

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

**(1)**

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)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**3** The random variable *Y* has a normal distribution with mean *μ* and standard deviation *σ*

The P(*Y* > 17) = 0.4

Find

(*a*)P(*μ* < *Y* < 17)

**(1)**

(*b*)P(*μ* – *σ* < *Y* < 17)

**(4)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**4** A bag contains 64 coloured beads. There are *r* red beads, *y* yellow beads and 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)**

The probability that both of the beads are red is 

(*b*)Show that *r* satisfies the equation *r*2 – *r* – 240 = 0

**(3)**

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

**(2)**

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

**(4)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_6** A group of climbers collected information about the height above sea level, *h* metres, and

the air temperature, *t* °C, at the same time at 8 different points on the same mountain.

The data are summarised by

|  |  |  |  |
| --- | --- | --- | --- |
| = 6370 | = 61 | = 31070 | = 693 |

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)**

*t* = 17.7 – 0.0126*h*

(*d*)Give an interpretation of the value 17.7.

**(1)**

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)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**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)**

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)**

(*c*)Find the lower quartile, *q*1, of the weight of potatoes **sold** by Adam.

**(5)**

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)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**S2**

**2** A fair coin is spun 6 times and the random variable *T* represents the number of tails obtained.

(*a*)Give two reasons why a binomial model would be a suitable distribution for

modelling *T*.

**(2)**

(*b*)Find P(*T* = 5)

**(2)**

(*c*)Find the probability of obtaining more tails than heads.

**(2)**

A second coin is biased such that the probability of obtaining a head is 

This second coin is spun 6 times.

(*d*)Find the probability that, for the second coin, the number of heads obtained is greater

than or equal to the number of tails obtained.

**(3)**

**5** Past records show that the proportion of customers buying organic vegetables from *Tesson*

supermarket is 0.35.

During a particular day, a random sample of 40 customers from *Tesson* supermarket was

taken and 18 of them bought organic vegetables.

(*a*)Test, at the 5% level of significance, whether or not this provides evidence

that the proportion of customers who bought organic vegetables has increased.

State your hypotheses clearly.

**(5)**

The manager of *Tesson* supermarket claims that the proportion of customers buying

organic eggs is different from the proportion of those buying organic vegetables. To test

this claim the manager decides to take a random sample of 50 customers.

(*b*)Using a 5% level of significance, find the critical region to enable the *Tesson*

supermarket manager to test her claim. The probability for each tail of the region

should be as close as possible to 2.5%

**(3)**

During a particular day, a random sample of 50 customers from *Tesson* supermarket is

taken and 8 of them bought organic eggs.

(*c*)Using your answer to part (*b*), state whether or not this sample supports the manager’s

claim. Use a 5% level of significance.

**(1)**

(*d*)State the actual significance level of this test.

**(1)**

The proportion of customers who buy organic fruit from *Tesson* supermarket is 0.2.

During a particular day, a random sample of 200 customers from *Tesson* supermarket is

taken. Using a suitable approximation, the probability that fewer than *n* of these customers

bought organic fruit is 0.0465 correct to 4 decimal places.

(*e*)Find the value of *n*.

**(6)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**S3**

**1** Phil measures the concentration of a radioactive element, *c*, and the amount of dissolved

solids, *a*, of 8 random samples of groundwater. His results are shown in the table below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* |
| *c* | 625 | 700 | 650 | 645 | 720 | 600 | 825 | 665 |
| *a* | 1.28 | 1.30 | 1.00 | 1.20 | 1.55 | 1.15 | 1.40 | 1.45 |

Given r = 0.549

(*b*)Use your value of the product moment correlation coefficient to test whether or not

there is evidence of a positive correlation between the concentration of this radioactive

element and the amount of dissolved solids in groundwater. Use a 5% significance

level. State your hypotheses clearly.

**(3)**

**2** Merchandise is sold at concerts. The manager of a concert claims that the mean value of

merchandise sold to premium ticket holders is more than £6 greater than the mean value

of merchandise sold to standard ticket holders.

(*a*)Given that all the tickets for the next concert have been sold, describe how a stratified

sample should be taken at the concert.

**(3)**

**(Total 82)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**2017**

**S1**

**1.** A clothes shop manager records the weekly sales figures, £ *s*, and the average weekly

temperature, *t* °C, for 6 weeks during the summer. The sales figures were coded so that

*w* =**

The manager of the clothes shop believes that a linear regression model may be appropriate

to describe these data.

The product moment correlation coefficient r = -0.801.

(*d*)State, giving a reason, whether or not your value of the correlation coefficient supports

the manager’s belief.

**(1)**

The equation of the regression line of *w* on *t* is *w* = 20.0 – 0.655*t*

(*f*)Hence find the equation of the regression line of *s* on *t*, giving your answer

in the form *s* = *c* + *dt*, where *c* and *d* are correct to 3 significant figures.

**(2)**

(*g*)Using your equation in part (*f*), interpret the effect of a 1°C increase in average

weekly temperature on weekly sales during the summer.

**(1)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**2.** An estate agent is studying the cost of office space in London. He takes a random sample

of 90 offices and calculates the cost, £*x* per square foot. His results are given in the table

below.

|  |  |  |
| --- | --- | --- |
| **Cost (£*x*)** | **Frequency (f)** | **Midpoint (£*y*)** |
| 20 ⩽ *x* < 40 | 12 | 30 |
| 40 ⩽ *x* < 45 | 13 | 42.5 |
| 45 ⩽ *x* < 50 | 25 | 47.5 |
| 50 ⩽ *x* < 60 | 32 | 55 |
| 60 ⩽ *x* < 80 | 8 | 70 |

(You may use  *y*2 = 226 687.5)

A histogram is drawn for these data and the bar representing 50 ⩽ *x* < 60 is 2 cm wide and

8 cm high.

(*a*)Calculate the width and height of the bar representing 20 ⩽ *x* < 40

**(3)**

(*b*)Use linear interpolation to estimate the median cost.

**(2)**

(*c*)Estimate the mean cost of office space for these data.

**(2)**

(*d*)Estimate the standard deviation for these data.

**(2)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**3.** The Venn diagram shows three events *A*, *B* and *C*, where *p*, *q*, *r*, *s* and *t* are probabilities.



P(*A*)= 0.5, P(*B*)= 0.6 and P(*C*)= 0.25 and the events *B* and *C* are independent.

(*a*)Find the value of *p* and the value of *q*.

**(2)**

(*b*)Find the value of *r*.

**(2)**

(*c*)Hence write down the value of *s* and the value of *t*.

**(2)**

(*d*)State, giving a reason, whether or not the events *A* and *B* are independent.

**(2)**

(*e*)Find P(*B* | *A*  *C*).

**(3)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**4.** The discrete random variable *X* has probability distribution

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *x* | –1 | 0 | 1 | 2 |
| P(*X* = *x*) | *a* | *b* | *b* | *c* |

The cumulative distribution function of *X* is given by

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *x* | –1 | 0 | 1 | 2 |
| F(*x*) |  | *d* |  | *e* |

(*a*)Find the values of *a*, *b*, *c*, *d* and *e*.

**(5)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**5.** Yuto works in the quality control department of a large company. The time, *T* minutes, it

takes Yuto to analyse a sample is normally distributed with mean 18 minutes and standard

deviation 5 minutes.

(*a*)Find the probability that Yuto takes longer than 20 minutes to analyse the next sample.

**(3)**

The company has a large store of samples analysed by Yuto with the time taken for

each analysis recorded. Serena is investigating the samples that took Yuto longer than

15 minutes to analyse.

She selects, at random, one of the samples that took Yuto longer than 15 minutes to analyse.

(*b*)Find the probability that this sample took Yuto more than 20 minutes to analyse.

**(4)**

Serena can identify, in advance, the samples that Yuto can analyse in under 15 minutes and

in future she will assign these to someone else.

(*c*)Estimate the median time taken by Yuto to analyse samples in future.

**(5)**

**S2**

**1.** A potter believes that 20% of pots break whilst being fired in a kiln. Pots are fired in

batches of 25.

(*a*)Let *X* denote the number of broken pots in a batch. A batch is selected at random.

Using a 10% significance level, find the critical region for a two tailed test of the

potter’s belief. You should state the probability in each tail of your critical region.

**(4)**

The potter aims to reduce the proportion of pots which break in the kiln by increasing the

size of the batch fired. He now fires pots in batches of 50. He then chooses a batch at

random and discovers there are 6 pots which broke whilst being fired in the kiln.

(*b*)Test, at the 5% level of significance, whether or not there is evidence that increasing

the number of pots in a batch has reduced the percentage of pots that break whilst

being fired in the kiln. State your hypotheses clearly.

**(5)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**5.** The time taken for a randomly selected person to complete a test is *M* minutes, where

*M* ~ N (14, *σ* 2)

Given that 10% of people take less than 12 minutes to complete the test,

(*a*)find the value of *σ*

**(3)**

Graham selects 15 people at random.

(*b*)Find the probability that fewer than 2 of these people will take less than 12 minutes

to complete the test.

**(3)**

Jovanna takes a random sample of *n* people.

Using a normal approximation, the probability that fewer than 9 of these *n* people will take

less than 12 minutes to complete the test is 0.3085 to 4 decimal places.

(*c*)Find the value of *n*.

**(8)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**S3**

**1.** A company director decides to survey staff about changes to the company calendar. The

company has staff in 4 different job roles

72 managers, 108 drivers, 180 administrators and 360 warehouse staff.

The director decides to take a stratified sample.

(*a*)Write down one advantage of using a stratified sample rather than a simple random

sample for this survey.

**(1)**

(*b*)Find the number of staff in each job role that will be included in a stratified sample

of 40 staff.

**(3)**

(*c*)Describe how to choose managers for the stratified sample.

**(2)**

**(Total 70)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**2016**

**S1**

**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 |

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

**(1)**

The equation of the regression line of *w* on *d*, is w = 0.722 + 0.0142d

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

(*e*)(i) Use the 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)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *x* | –2 | –1 |  |  | 2 |
| P(*X* = *x*) | *p* | *q* | 0.2 | 0.3 | *p* |

(*a*)Write down an equation in *p* and *q*.

**(1)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**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

= 86.8  = 985.88

(*a*)Find the standard deviation, *σx*, 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

** = 58 *σy =* 9.461 (correct to 4 significant figures) r = - 0.753

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)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**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* ∩ *B*),

(ii) P(*D*|*R* ∩ *Bʹ* ).

**(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)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**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 ≤ *w <* 2 | 1 | 1 |
| 2 ≤ *w* < 3 | 8 | 2.5 |
| 3 ≤ *w* < 3.5 | 17 | 3.25 |
| 3.5 ≤ *w* < 4 | 17 | 3.75 |
| 4 ≤ *w* < 5 | 7 | 4.5 |

[You may use  = 611.375]

A histogram has been drawn to represent these data.

The bar representing the weight 2 ≤ *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 ≤ *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* ~ N(3.43, 0.652).

(*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)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**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* < *μ* + 30) = 0.82,

(*c*)find P(*W* < *μ* – 30 | *W* < *μ*).

**(3)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**S2**

**2.** In a region of the UK, 5% of people have red hair. In a random sample of size *n*, taken from this region, the expected number of people with red hair is 3.

(*a*)Calculate the value of *n*.

**(2)**

A random sample of 20 people is taken from this region.

Find the probability that

(*b*)(i) exactly 4 of these people have red hair,

(ii) at least 4 of these people have red hair.

**(5)**

Patrick claims that *Reddman* people have a probability greater than 5% of having red hair. In a random sample of 50 *Reddman* people, 4 of them have red hair.

(*c*)Stating your hypotheses clearly, test Patrick’s claim. Use a 1% level of significance.

**(5)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**5.** In a large school, 20% of students own a touch screen laptop. A random sample of *n* students is chosen from the school. Using a normal approximation, the probability that more than 55 of these *n* students own a touch screen laptop is 0.0401 correct to 3 significant figures.

Find the value of *n*.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**S3**

1. (*a*)State two reasons why stratified sampling might be a more suitable sampling method than simple random sampling.

**(2)**

(*b*)State two reasons why stratified sampling might be a more suitable sampling method than quota sampling.

**(2)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**(Total 72)**

**2015**

**S1**

**1.** Each of 60 students was asked to draw a 20° angle without using a protractor. The size of

each angle drawn was measured. The results are summarised in the box plot below.



size of angle

(*a*) Find the range for these data.

**(1)**

(*b*) Find the interquartile range for these data.

**(1)**

The students were then asked to draw a 70° angle.

The results are summarised in the table below.

|  |  |
| --- | --- |
| **Angle, *a*, (degrees)** | **Number of students** |
| 55 ≤ *a* < 60 | 6 |
| 60 ≤ *a* < 65 | 15 |
| 65 ≤ *a* < 70 | 13 |
| 70 ≤ *a* < 75 | 11 |
| 75 ≤ *a* < 80 | 8 |
| 80 ≤ *a* < 85 | 7 |

(*c*) Use linear interpolation to estimate the size of the median angle drawn. Give your answer to 1 decimal place.

**(2)**

(*d*) Show that the lower quartile is 63°.

**(2)**

For these data, the upper quartile is 75°, the minimum is 55° and the maximum is 84°.

An outlier is an observation that falls either

more than 1.5 × (interquartile range) above the upper quartile or

more than 1.5 × (interquartile range) below the lower quartile.

(*e*) (i) Show that there are no outliers for these data.

(ii) On graph paper, draw a box plot for these data.

**(5)**

(*f*) State which angle the students were more accurate at drawing. Give reasons for your answer.

**(3)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**2.** An estate agent recorded the price per square metre, *p* £/m2, for 7 two-bedroom houses.

He then coded the data using the coding *q* = , where *a* and *b* are positive constants.

His results are shown in the table below.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *p* | 1840 | 1848 | 1830 | 1824 | 1819 | 1834 | 1850 |
| *q* | 4.0 | 4.8 | 3.0 | 2.4 | 1.9 | 3.4 | 5.0 |

(*a*) Find the value of *a* and the value of *b*.

**(2)**

The estate agent also recorded the distance, *d* km, of each house from the nearest train station and calculated the product moment correlation coefficient between *d* and *q* to be  **0.749**.

(*c*) Write down the value of the product moment correlation coefficient between *d* and *p*.

**(1)**

The estate agent records the price and size of 2 additional two-bedroom houses, *H* and *J*.

|  |  |  |
| --- | --- | --- |
| House | Price (£) | Size (m2) |
| *H* | 156 400 | 85 |
| *J* | 172 900 | 95 |

(*d*) Suggest which house is most likely to be closer to a train station. Justify your answer.

**(3)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**3.** A college has 80 students in Year 12.

20 students study Biology.

28 students study Chemistry.

30 students study Physics.

7 students study both Biology and Chemistry.

11 students study both Chemistry and Physics.

5 students study both Physics and Biology.

3 students study all 3 of these subjects.

(*a*) Draw a Venn diagram to represent this information.

**(5)**

A Year 12 student at the college is selected at random.

(*b*) Find the probability that the student studies Chemistry but not Biology or Physics.

**(1)**

(*c*) Find the probability that the student studies Chemistry or Physics or both.

**(2)**

Given that the student studies Chemistry or Physics or both,

(*d*) find the probability that the student does not study Biology.

**(2)**

(*e*) Determine whether studying Biology and studying Chemistry are statistically independent.

**(3)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**4.** Statistical models can provide a cheap and quick way to describe a real world situation.

(*a*) Give two other reasons why statistical models are used.

**(2)**

A scientist wants to develop a model to describe the relationship between the average daily temperature, *x* °C, and her household’s daily energy consumption, *y* kWh, in winter.

A random sample of the average daily temperature and her household’s daily energy consumption are taken from 10 winter days and shown in the table.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *x* | –0.4 | –0.2 | 0.3 | 0.8 | 1.1 | 1.4 | 1.8 | 2.1 | 2.5 | 2.6 |
| *y* | 28 | 30 | 26 | 25 | 26 | 27 | 26 | 24 | 22 | 21 |

The equation of the regression line of *y* on *x* is .

(*d*) Give an interpretation of the value of *a*.

**(1)**

(*e*) Estimate her household’s daily energy consumption when the average daily temperature is 2°C.

**(2)**

The scientist wants to use the linear regression model to predict her household’s energy consumption in the summer.

(*f*) Discuss the reliability of using this model to predict her household’s energy consumption in the summer.

**(2)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**5.** In a quiz, a team gains 10 points for every question it answers correctly and loses 5 points for every question it does not answer correctly. The probability of answering a question correctly is 0.6 for each question. One round of the quiz consists of 3 questions.

The discrete random variable *X* represents the total number of points scored in one round.

The table shows the incomplete probability distribution of *X*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *x* | 30 | 15 | 0 | –15 |
| P(*X* = *x*) | 0.216 |  |  | 0.064 |

(*a*) Show that the probability of scoring 15 points in a round is 0.432.

**(2)**

(*b*) Find the probability of scoring 0 points in a round.

**(1)**

(*c*) Find the probability of scoring a total of 30 points in 2 rounds.

**(3)**

**6.** The random variable *Z* ~ N(0, 1).

*A* is the event *Z* > 1.1

*B* is the event *Z >* –1.9

*C* is the event –1.5 < *Z* < 1.5

(*a*) Find

(i) P(*A*),

(ii) P(*B*),

(iii) P(*C*),

(iv) P(*A* ∪ *C*).

**(6)**

The random variable *X* has a normal distribution with mean 21 and standard deviation 5.

(*b*) Find the value of *w* such that P(*X* > *w* | *X* > 28) = 0.625.

**(6)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**S2**

**2.** The proportion of houses in Radville which are unable to receive digital radio is 25%. In a survey of a random sample of 30 houses taken from Radville, the number, *X*,of houses which are unable to receive digital radio is recorded.

(*a*) Find P(5 ≤ *X* < 11).

**(3)**

A radio company claims that a new transmitter set up in Radville will reduce the proportion of houses which are unable to receive digital radio. After the new transmitter has been set up, a random sample of 15 houses is taken, of which 1 house is unable to receive digital radio.

(*b*) Test, at the 10% level of significance, the radio company’s claim. State your hypotheses clearly.

**(5)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**S3**

**1.** A mobile library has 160 books for children on its records. The librarian believes that books with fewer pages are borrowed more often. He takes a random sample of 10 books for children.

(*a*) Explain how the librarian should select this random sample.

**(2)**

**3.** A nursery has 16 staff and 40 children on its records. In preparation for an outing the manager needs an estimate of the mean weight of the people on its records and decides to take a stratified sample of size 14.

(*a*) Describe how this stratified sample should be taken.

**(3) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**(Total 71)**

**2014(R)**

**S1**

**1.** The discrete random variable *X* has probability distribution

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *x* | –4 | –2 | 1 | 3 | 5 |
| P(*X* = *x*) | 0.4 | *p* | 0.05 | 0.15 | *p* |

(*a*) Show that *p* = 0.2.

**(2)**

(*c*) F(0)

**(1)**

(*d*) P(3*X* + 2 > 5)

**(2)**

**2.** The discrete random variable *X* has probability distribution

 *x* = 1, 2, 3, … 10

(*a*) Write down the name given to this distribution.

**(1)**

(*b*) Write down the value of

(i) P(*X* = 10)

(ii) P(*X* < 10)

**(2)**

The continuous random variable *Y* has the normal distribution N(10, 22).

(*c*) Write down the value of

(i) P(*Y* = 10)

(ii) P(*Y* < 10)

**(2)**

**3.** A large company is analysing how much money it spends on paper in its offices every year. The number of employees, *x*, and the amount of money spent on paper, *p* (£ hundreds), in   
8 randomly selected offices are given in the table below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *x* | 8 | 9 | 12 | 14 | 7 | 3 | 16 | 19 |
| *p* (£ hundreds) | 40.5 | 36.1 | 30.4 | 39.4 | 32.6 | 31.1 | 43.4 | 45.7 |

(You may use Σ*x*2 = 1160 Σ*p* = 299.2 Σ*p*2 = 11 422 Σ*xp* = 3449.5)

The equation of the regression line of *p* on *x* is given in the form *p* = 28.3 *+ 0.824x*.

(*d*) Estimate the amount of money spent on paper in an office with 10 employees.

**(2)**

(*e*) Explain the effect each additional employee has on the amount of money spent on paper.

**(1)**

Later the company realised it had made a mistake in adding up its costs, *p*. The true costs were actually half of the values recorded. The product moment correlation coefficient and the equation of the linear regression line are recalculated using this information.

(*f*) Write down the new value of

(i) the product moment correlation coefficient,

(ii) the gradient of the regression line.

**(2)**

**4.** *A* and *B* are two events such that

P(*B*) =  P(*A* | *B*) =  

(*a*) Find .

**(2)**

(*b*) Draw a Venn diagram to show the events *A*, *B* and all the associated probabilities.

**(3)**

Find

(*c*) P(*A*)

**(1)**

(*d*) P(*B* | *A*)

**(2)**

(*e*) 

**(1)**

**5*.***The table shows the time, to the nearest minute, spent waiting for a taxi by each of 80 people one Sunday afternoon.

|  |  |
| --- | --- |
| **Waiting time (in minutes)** | **Frequency** |
| 2–4 | 15 |
| 5–6 | 9 |
| 7 | 6 |
| 8 | 24 |
| 9–10 | 14 |
| 11–15 | 12 |

(*a*) Write down the upper class boundary for the 2–4 minute interval.

**(1)**

A histogram is drawn to represent these data. The height of the tallest bar is 6 cm.

(*b*) Calculate the height of the second tallest bar.

**(3)**

(*c*) Estimate the number of people with a waiting time between 3.5 minutes and 7 minutes.

**(2)**

(*d*) Use linear interpolation to estimate the median, the lower quartile and the upper quartile of the waiting times.

**(4)**

**6.** The time taken, in minutes, by children to complete a mathematical puzzle is assumed to be normally distributed with mean *μ* and standard deviation *σ*. The puzzle can be completed in less than 24 minutes by 80% of the children. For 5% of the children it takes more than   
28 minutes to complete the puzzle.

(*a*) Show this information on a Normal curve.

**(2)**

(*b*)Write down the percentage of children who take between 24 minutes and 28 minutes to complete the puzzle.

**(1)**

(*c*) (i) Find two equations in *μ* and *σ*.

(ii) Hence find, to 3 significant figures, the value of *μ* and the value of *σ*.

**(7)**

A child is selected at random.

(*d*) Find the probability that the child takes less than 12 minutes to complete the puzzle.

**(3)**

**7.** In a large company,

78% of employees are car owners,

30% of these car owners are also bike owners,

85% of those who are not car owners are bike owners.

(*a*) Draw a tree diagram to represent this information.

**(3)**

An employee is selected at random.

(*b*) Find the probability that the employee is a car owner or a bike owner but not both.

**(2)**

Another employee is selected at random.

Given that this employee is a bike owner,

(*c*) find the probability that the employee is a car owner.

**(3)**

Two employees are selected at random.

(*d*) Find the probability that only one of them is a bike owner.

**(3) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**S2**

**1.** Before Roger will use a tennis ball he checks it using a “bounce” test. The probability that a ball from Roger’s usual supplier fails the bounce test is 0.2. A new supplier claims that the probability of one of their balls failing the bounce test is less than 0.2. Roger checks a random sample of 40 balls from the new supplier and finds that 3 balls fail the bounce test.

Stating your hypotheses clearly, use a 5% level of significance to test the new supplier’s claim.

**(5)**

**(Total 63)**

**2014**

**S1**

**1.** A random sample of 35 homeowners was taken from each of the villages Greenslax and Penville and their ages were recorded. The results are summarised in the back-to-back stem and leaf diagram below.



Some of the quartiles for these two distributions are given in the table below.

|  |  |  |
| --- | --- | --- |
|  | **Greenslax** | **Penville** |
| Lower quartile, *Q*1 | *a* | 31 |
| Median, *Q*2 | 64 | 39 |
| Upper quartile, *Q*3 | *b* | 55 |

(*a*) Find the value of *a* and the value of *b*.

**(2)**

An outlier is a value that falls either

more than 1.5 × (*Q*3 – *Q*1) above *Q*3

or more than 1.5 × (*Q*3 – *Q*1) below *Q*1

(*b*) On the graph paper draw a box plot to represent the data from Penville.   
Show clearly any outliers.

**(4)**

**2.** The mark, *x*, scored by each student who sat a statistics examination is coded using

*y* = 1.4*x* – 20

The coded marks have mean 60.8 and standard deviation 6.60.

Find the mean and the standard deviation of *x*.

**(4)**

**3.** The table shows data on the number of visitors to the UK in a month, *v* (1000s), and the amount of money they spent, *m* (£ millions), for each of 8 months.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Number of visitors  *v* (1000s) | 2450 | 2480 | 2540 | 2420 | 2350 | 2290 | 2400 | 2460 |
| Amount of money spent  *m* (£ millions) | 1370 | 1350 | 1400 | 1330 | 1270 | 1210 | 1330 | 1350 |

The product moment correlation coefficient between *m* and *v* is found to be**0.962**.

(*b*) Give a reason to support fitting a regression model of the form *m* = *a* + *bv* to these data.

**(1)**

The equation of the regression line of *m* on *v* is ***m* = – 467+ 0.74*ν***

(*e*) Interpret your value 0.74.

**(2)**

(*f*) Use the equation to estimate the amount of money spent when the number of visitors to the UK in a month is 2 500 000.

**(2)**

(*g*) Comment on the reliability of your estimate in part (*f*). Give a reason for your answer.

**(2)**

**4.** In a factory, three machines, *J*, *K* and *L*, are used to make biscuits.

Machine *J* makes 25% of the biscuits.

Machine *K* makes 45% of the biscuits.

The rest of the biscuits are made by machine *L*.

It is known that 2% of the biscuits made by machine *J* are broken*,* 3% of the biscuits made by machine *K* are broken and 5% of the biscuits made by machine *L* are broken.

(*a*) Draw a tree diagram to illustrate all the possible outcomes and associated probabilities.

**(2)**

A biscuit is selected at random.

(*b*) Calculate the probability that the biscuit is made by machine *J* and is not broken.

**(2)**

(*c*) Calculate the probability that the biscuit is broken.

**(2)**

(*d*) Given that the biscuit is broken, find the probability that it was not made by machine *K*.

**(3)**

**5.** The discrete random variable *X* has the probability function



where *k* is a constant.

(*a*) Show that *k* = .

**(2)**

(*b*) Find the exact value of F(5).

**(1)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**6.** The times, in seconds, spent in a queue at a supermarket by 85 randomly selected customers, are summarised in the table below.

|  |  |
| --- | --- |
| Time (seconds) | Number of customers, *f* |
| 0 – 30 | 2 |
| 30 – 60 | 10 |
| 60 – 70 | 17 |
| 70 – 80 | 25 |
| 80 – 100 | 25 |
| 100 – 150 | 6 |

A histogram was drawn to represent these data. The 30 – 60 group was represented by a bar of width 1.5 cm and height 1 cm.

(*a*) Find the width and the height of the 70 – 80 group.

**(3)**

(*b*) Use linear interpolation to estimate the median of this distribution.

**(2)**

Given that *x* denotes the midpoint of each group in the table and

Σ*fx* =6460 Σ*fx*2 = 529 400

(*c*) calculate an estimate for

(i) the mean,

(ii) the standard deviation,

for the above data.

**(3)**

**7.** The heights of adult females are normally distributed with mean 160 cm and standard deviation 8 cm.

(*a*) Find the probability that a randomly selected adult female has a height greater than   
170 cm.

**(3)**

Any adult female whose height is greater than 170 cm is defined as tall.

An adult female is chosen at random. Given that she is tall,

(*b*) find the probability that she has a height greater than 180 cm.

**(4)**

Half of tall adult females have a height greater than *h* cm.

(*c*) Find the value of *h*.

**(5)**

**8.** For the events *A* and *B*,

 and 

(*a*) Find P(*A*).

**(1)**

(*b*) Find .

**(1)**

Given that P(*A* | *B*) = 0.6,

(*c*) find .

**(3)**

(*d*) Determine whether or not *A* and *B* are independent.

**(2)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**S2**

**4.** A cadet fires shots at a target at distances ranging from 25 m to 90 m. The probability of hitting the target with a single shot is *p*. When firing from a distance *d* m, .

Each shot is fired independently.

The cadet fires 10 shots from a distance of 40 m.

(*a*) (i) Find the probability that exactly 6 shots hit the target.

(ii) Find the probability that at least 8 shots hit the target.

**(5)**

The cadet fires 20 shots from a distance of *x* m.

(*b*) Find, to the nearest integer, the value of *x* if the cadet has an 80% chance of hitting the target at least once.

**(4)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**5.** (*a*) State the conditions under which the normal distribution may be used as an approximation to the binomial distribution.

**(2)**

A company sells seeds and claims that 55% of its pea seeds germinate.

(*b*) Write down a reason why the company should not justify their claim by testing all the pea seeds they produce.

**(1)**

To test the company’s claim, a random sample of 220 pea seeds was planted.

(*c*) State the hypotheses for a two-tailed test of the company’s claim.

**(1)**

Given that 135 of the 220 pea seeds germinated,

(*d*) use a normal approximation to test, at the 5% level of significance, whether or not the company’s claim is justified.

**(7) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**S3**

**1.** (*a*) Explain what you understand by a random sample from a finite population.

**(1)**

(*b*) Give an example of a situation when it is not possible to take a random sample.

**(1)**

A college lecturer specialising in shoe design wants to change the way in which she organises practical work.

She decides to gather ideas from her 75 students.

She plans to give a questionnaire to a random sample of 8 of these students.

(*c*) (i) Describe the sampling frame that she should use.

(ii) Explain in detail how she should use a table of random numbers to obtain her sample.

**(3)**

**(Total 81)**