

Paper Reference(s)

6685**Edexcel GCE****Statistics S3****Advanced/Advanced Subsidiary****Tuesday 28 May 2002 – Afternoon****Time: 1 hour 30 minutes****Materials required for examination**

Answer Book (AB16)

Graph Paper (ASG2)

Mathematical Formulae (Lilac)

Items included with question papers

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6685), your surname, other name and signature.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

This paper has seven questions. Pages 6, 7 and 8 are blank.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. A hotel has 160 rooms of which 20 are classified as De-luxe, 40 Premier and 100 as Standard. The manager wants to obtain information about room usage in the hotel by taking a 10% sample of the rooms.

(a) Suggest a suitable sampling method.

(1)

(b) Explain in detail how the manager should obtain the sample.

(4)

2. A random sample of 100 classical CDs produced by a record company had a mean playing time of 70.6 minutes and a standard deviation of 9.1 minutes. An independent random sample of 120 CDs produced by a different company had a mean playing time of 67.2 minutes with a standard deviation of 8.4 minutes.

(a) Using a 1% level of significance, test whether or not there is a difference in the mean playing times of the CDs produced by these two companies. State your hypotheses clearly.

(8)

(b) State an assumption you made in carrying out the test in part (a).

(1)

3. The weights of a group of males are normally distributed with mean 80 kg and standard deviation 2.6 kg. A random sample of 10 of these males is selected.

(a) Write down the distribution of \bar{M} , the mean weight, in kg, of this sample.

(2)

(b) Find $P(\bar{M} < 78.5)$.

(3)

The weights of a group of females are normally distributed with mean 59 kg and standard deviation 1.9 kg. A random sample of 6 of the males and 4 of the females enters a lift that can carry a maximum load of 730 kg.

(c) Find the probability that the maximum load will be exceeded when these 10 people enter the lift.

(5)

4. At the end of a season an athletics coach graded a random sample of ten athletes according to their performances throughout the season and their dedication to training. The results, expressed as percentages, are shown in the table below.

Athlete	Performance	Dedication
<i>A</i>	86	72
<i>B</i>	60	69
<i>C</i>	78	59
<i>D</i>	56	68
<i>E</i>	80	80
<i>F</i>	66	84
<i>G</i>	31	65
<i>H</i>	59	55
<i>I</i>	73	79
<i>J</i>	49	53

(a) Calculate the Spearman rank correlation coefficient between performance and dedication.

(5)

(b) Stating clearly your hypotheses and using a 10% level of significance, interpret your rank correlation coefficient.

(5)

(c) Give a reason to support the use of the rank correlation coefficient rather than the product moment correlation coefficient with these data.

(1)

5. The manager of a leisure centre collected data on the usage of the facilities in the centre by its members. A random sample from her records is summarised below.

Facility	Male	Female
Pool	40	68
Jacuzzi	26	33
Gym	52	31

Making your method clear, test whether or not there is any evidence of an association between gender and use of the club facilities. State your hypotheses clearly and use a 5% level of significance.

(11)

6. Data were collected on the number of female puppies born in 200 litters of size 8. It was decided to test whether or not a binomial model with parameters $n = 8$ and $p = 0.5$ is a suitable model for these data. The following table shows the observed frequencies and the expected frequencies, to 2 decimal places, obtained in order to carry out this test.

Number of females	Observed number of litters	Expected number of litters
0	1	0.78
1	9	6.25
2	27	21.88
3	46	<i>R</i>
4	49	<i>S</i>
5	35	<i>T</i>
6	26	21.88
7	5	6.25
8	2	0.78

(a) Find the values of *R*, *S* and *T*.

(4)

(b) Carry out the test to determine whether or not this binomial model is a suitable one. State your hypotheses clearly and use a 5% level of significance.

(7)

An alternative test might have involved estimating p rather than assuming $p = 0.5$.

(c) Explain how this would have affected the test.

(1)

7. The weights of tubs of margarine are known to be normally distributed. A random sample of 10 tubs of margarine were weighed, to the nearest gram, and the results were as follows.

498 502 500 496 509 504 511 497 506 499

- (a) Find unbiased estimates of the mean and the variance of the population from which this sample was taken.

(5)

Given that the population standard deviation is 5.0 g,

- (b) estimate limits, to 2 decimal places, between which 90% of the weights of the tubs lie,

(2)

- (c) find a 95% confidence interval for the mean weight of the tubs.

(5)

A second random sample of 15 tubs was found to have a mean weight of 501.9 g.

- (d) Stating your hypotheses clearly and using a 1% level of significance, test whether or not the mean weight of these tubs is greater than 500 g.

(5)

END

Paper Reference(s)

6685

Edexcel GCE

Statistics S3

Advanced/Advanced Subsidiary

Thursday 5 June 2003 – Morning

Time: 1 hour 30 minutes

Materials required for examination

Answer Book (AB16)

Graph Paper (ASG2)

Mathematical Formulae (Lilac)

Items included with question papers

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6685), your surname, other name and signature.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

This paper has seven questions.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. Explain how to obtain a sample from a population using
- (a) stratified sampling, (2)
- (b) quota sampling. (2)
- Give one advantage and one disadvantage of each sampling method. (4)
-

2. A random sample of 30 apples was taken from a batch. The mean weight of the sample was 124 g with standard deviation 20 g.
- (a) Find a 99% confidence interval for the mean weight μ grams of the population of apples. Write down any assumptions you made in your calculations. (6)
- Given that the actual value of μ is 140,
- (b) state, with a reason, what you can conclude about the sample of 30 apples. (2)
-

3. Given the random variables $X \sim N(20, 5)$ and $Y \sim N(10, 4)$ where X and Y are independent, find
- (a) $E(X - Y)$, (2)
- (b) $\text{Var}(X - Y)$, (2)
- (c) $P(13 < X - Y < 16)$. (5)
-

4. A new drug to treat the common cold was used with a randomly selected group of 100 volunteers. Each was given the drug and their health was monitored to see if they caught a cold. A randomly selected control group of 100 volunteers was treated with a dummy pill. The results are shown in the table below.

	Cold	No cold
Drug	34	66
Dummy pill	45	55

Using a 5% significance level, test whether or not the chance of catching a cold is affected by taking the new drug. State your hypotheses clearly. (11)

5. A scientist monitored the levels of river pollution near a factory. Before the factory was closed down she took 100 random samples of water from different parts of the river and found an average weight of pollutants of 10 mg l^{-1} with a standard deviation of 2.64 mg l^{-1} . After the factory was closed down the scientist collected a further 120 random samples and found that they contained 8 mg l^{-1} of pollutants on average with a standard deviation of 1.94 mg l^{-1} .
- Test, at the 5% level of significance, whether or not the mean river pollution fell after the factory closed down. (11)
-

6. Two judges ranked 8 ice skaters in a competition according to the table below.

Skater \ Judge	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
A	2	5	3	7	8	1	4	6
B	3	2	6	5	7	4	1	8

- (a) Evaluate Spearman's rank correlation coefficient between the ranks of the two judges. (6)
- (b) Use a suitable test, at the 5% level of significance, to interpret this result. (5)
-

7. A bag contains a large number of coins of which 30% are 50p coins, 20% are 10p coins and the rest are 2p coins.

(a) Find the mean μ and the variance σ^2 of this population of coins. (4)

A random sample of 2 coins is drawn from the bag one after the other.

(b) List all possible samples that could be drawn. (2)

(c) Find the sampling distribution of \bar{X} , the mean of the coins drawn. (4)

(d) Find $P(2 \leq \bar{X} < 7)$. (2)

(e) Use the sampling distribution of \bar{X} to verify $E(\bar{X}) = \mu$ and $\text{Var}(\bar{X}) = \frac{1}{2} \sigma^2$. (5)

END

Paper Reference(s)

6685

Edexcel GCE

Statistics S3

Advanced/Advanced Subsidiary

Friday 11 June 2004 – Morning

Time: 1 hour 30 minutes

Materials required for examination

Answer Book (AB16)

Graph Paper (ASG2)

Mathematical Formulae (Lilac)

Items included with question papers

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6685), your surname, other name and signature.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

This paper has seven questions.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. There are 64 girls and 56 boys in a school.

Explain briefly how you could take a random sample of 15 pupils using

(a) a simple random sample, (3)

(b) a stratified sample. (3)

2. A random sample of 8 students sat examinations in Geography and Statistics. The product moment correlation coefficient between their results was 0.572 and the Spearman rank correlation coefficient was 0.655.

(a) Test both of these values for positive correlation. Use a 5% level of significance. (6)

(b) Comment on your results. (2)

3. It is known from past evidence that the weight of coffee dispensed into jars by machine A is normally distributed with mean μ_A and standard deviation 2.5 g. Machine B is known to dispense the same nominal weight of coffee into jars with mean μ_B and standard deviation 2.3 g. A random sample of 10 jars filled by machine A contained a mean weight of 249 g of coffee. A random sample of 15 jars filled by machine B contained a mean weight of 251 g.

(a) Test, at the 5% level of significance, whether or not there is evidence that the population mean weight dispensed by machine B is greater than that of machine A . (7)

(b) Write down an assumption needed to carry out this test. (1)

4. Kylie regularly travels from home to visit a friend. On 10 randomly selected occasions the journey time x minutes was recorded. The results are summarised as follows.

$$\Sigma x = 753, \quad \Sigma x^2 = 57\,455.$$

(a) Calculate unbiased estimates of the mean and the variance of the population of journey times. (3)

After many journeys, a random sample of 100 journeys gave a mean of 74.8 minutes and a variance of 84.6 minutes².

(b) Calculate a 95% confidence interval for the mean of the population of journey times. (5)

(c) Write down two assumptions you made in part (b). (2)

5. A random sample of 500 adults completed a questionnaire on how often they took part in some form of exercise. They gave a response of 'never', 'sometimes' or 'regularly'. Of those asked, 52% were females of whom 10% never exercised and 35% exercised regularly. Of the males, 12.5% never exercised and 55% sometimes exercised.

Test, at the 5% level of significance, whether or not there is any association between gender and the amount of exercise. State your hypotheses clearly. (12)

6. Three six-sided dice, which were assumed to be fair, were rolled 250 times. On each occasion the number X of sixes was recorded. The results were as follows.

Number of sixes	0	1	2	3
Frequency	125	109	13	3

(a) Write down a suitable model for X . (2)

(b) Test, at the 1% level of significance, the suitability of your model for these data. (11)

(c) Explain how the test would have been modified if it had not been assumed that the dice were fair. (2)

7. The random variable D is defined as

$$D = A - 3B + 4C$$

where $A \sim N(5, 2^2)$, $B \sim N(7, 3^2)$ and $C \sim N(9, 4^2)$, and A , B and C are independent.

(a) Find $P(D < 44)$. (9)

The random variables B_1 , B_2 and B_3 are independent and each has the same distribution as B . The random variable X is defined as

$$X = A - \sum_{i=1}^3 B_i + 4C.$$

(b) Find $P(X > 0)$. (7)

END

Paper Reference(s)

6685/01 6691/01

Edexcel GCE**Statistics S3****Advanced Subsidiary****Thursday 9 June 2005 – Morning****Time: 1 hour 30 minutes****Materials required for examination**Mathematical Formulae (Lilac)
Graph Paper (ASG2)**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6685), your surname, other name and signature.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

This paper has seven questions.

The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

N21146A

This publication may only be reproduced in accordance with London Qualifications Limited copyright policy.
©2005 London Qualifications Limited

1. (a) State two reasons why stratified sampling might be chosen as a method of sampling when carrying out a statistical survey. (2)

- (b) State one advantage and one disadvantage of quota sampling. (2)

(Total 4 marks)

2. A sample of size 5 is taken from a population that is normally distributed with mean 10 and standard deviation 3. Find the probability that the sample mean lies between 7 and 10.

(Total 6 marks)

3. A researcher carried out a survey of three treatments for a fruit tree disease. The contingency table below shows the results of a survey of a random sample of 60 diseased trees.

	No action	Remove diseased branches	Spray with chemicals
Tree died within 1 year	10	5	6
Tree survived for 1–4 years	5	9	7
Tree survived beyond 4 years	5	6	7

Test, at the 5% level of significance, whether or not there is any association between the treatment of the trees and their survival. State your hypotheses and conclusion clearly.

(Total 11 marks)

4. Over a period of time, researchers took 10 blood samples from one patient with a blood disease. For each sample, they measured the levels of serum magnesium, s mg/dl, in the blood and the corresponding level of the disease protein, d mg/dl. The results are shown in the table.

s	1.2	1.9	3.2	3.9	2.5	4.5	5.7	4.0	1.1	5.9
d	3.8	7.0	11.0	12.0	9.0	12.0	13.5	12.2	2.0	13.9

[Use $\sum s^2 = 141.51$, $\sum d^2 = 1081.74$ and $\sum sd = 386.32$]

- (a) Draw a scatter diagram to represent these data. (3)
- (b) State what is measured by the product moment correlation coefficient. (1)
- (c) Calculate S_{xx} , S_{dd} and S_{sd} . (3)
- (d) Calculate the value of the product moment correlation coefficient r between s and d . (2)
- (e) Stating your hypotheses clearly, test, at the 1% significance level, whether or not the correlation coefficient is greater than zero. (3)
- (f) With reference to your scatter diagram, comment on your result in part (e). (1)
- (Total 13 marks)**

5. The number of times per day a computer fails and has to be restarted is recorded for 200 days. The results are summarised in the table.

Number of restarts	Frequency
0	99
1	65
2	22
3	12
4	2

Test whether or not a Poisson model is suitable to represent the number of restarts per day. Use a 5% level of significance and state your hypothesis clearly.

(Total 12 marks)

6. A computer company repairs large numbers of PCs and wants to estimate the mean time to repair a particular fault. Five repairs are chosen at random from the company's records and the times taken, in seconds, are

205 310 405 195 320.

- (a) Calculate unbiased estimates of the mean and the variance of the population of repair times from which this sample has been taken. (4)

It is known from previous results that the standard deviation of the repair time for this fault is 100 seconds. The company manager wants to ensure that there is a probability of at least 0.95 that the estimate of the population mean lies within 20 seconds of its true value.

- (b) Find the minimum sample size required. (6)
- (Total 10 marks)**

7. A manufacturer produces two flavours of soft drink, cola and lemonade. The weights, C and L , in grams, of randomly selected cola and lemonade cans are such that $C \sim N(350, 8)$ and $L \sim N(345, 17)$.

- (a) Find the probability that the weights of two randomly selected cans of cola will differ by more than 6 g. (6)

One can of each flavour is selected at random.

- (b) Find the probability that the can of cola weighs more than the can of lemonade. (6)

Cans are delivered to shops in boxes of 24 cans. The weights of empty boxes are normally distributed with mean 100 g and standard deviation 2 g.

- (c) Find the probability that a full box of cola cans weighs between 8.51 kg and 8.52 kg. (6)

- (d) State an assumption you made in your calculation in part (c). (1)

(Total 19 marks)

END

TOTAL FOR PAPER: 75 MARKS

Paper Reference(s)

6685/01 6691/01

Edexcel GCE

Statistics S3

Advanced/Advanced Subsidiary

Friday 27 January 2006 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination

Mathematical Formulae (Lilac)

Items included with question papers

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6685 or 6691), your surname, other name and signature.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet ‘Mathematical Formulae and Statistical Tables’ is provided.

Full marks may be obtained for answers to ALL questions.

The marks for individual questions and the parts of questions are shown in round brackets: e.g. (2).

There are 7 questions on this paper. The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

N22339A

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.
©2006 Edexcel Limited.

1. A school has 15 classes and a sixth form. In each class there are 30 students. In the sixth form there are 150 students. There are equal numbers of boys and girls in each class. There are equal numbers of boys and girls in the sixth form. The head teacher wishes to obtain the opinions of the students about school uniforms.

Explain how the head teacher would take a stratified sample of size 40.

(7)

2. A workshop makes two types of electrical resistor.

The resistance, X ohms, of resistors of Type A is such that $X \sim N(20, 4)$.

The resistance, Y ohms, of resistors of Type B is such that $Y \sim N(10, 0.84)$.

When a resistor of each type is connected into a circuit, the resistance R ohms of the circuit is given by $R = X + Y$ where X and Y are independent.

Find

(a) $E(R)$, (1)

(b) $\text{Var}(R)$, (2)

(c) $P(28.9 < R < 32.64)$ (6)

3. The drying times of paint can be assumed to be normally distributed. A paint manufacturer paints 10 test areas with a new paint. The following drying times, to the nearest minute, were recorded.

82, 98, 140, 110, 90, 125, 150, 130, 70, 110.

- (a) Calculate unbiased estimates for the mean and the variance of the population of drying times of this paint. (5)

Given that the population standard deviation is 25,

- (b) find a 95% confidence interval for the mean drying time of this paint. (5)

Fifteen similar sets of tests are done and the 95% confidence interval is determined for each set.

- (c) Estimate the expected number of these 15 intervals that will enclose the true value of the population mean μ . (2)

4. People over the age of 65 are offered an annual flu injection. A health official took a random sample from a list of patients who were over 65. She recorded their gender and whether or not the offer of an annual flu injection was accepted or rejected. The results are summarised below.

Gender	Accepted	Rejected
Male	170	110
Female	280	140

Using a 5% significance level, test whether or not there is an association between gender and acceptance or rejection of an annual flu injection. State your hypotheses clearly.

(9)

5. Upon entering a school, a random sample of eight girls and an independent random sample of eighty boys were given the same examination in mathematics. The girls and boys were then taught in separate classes. After one year, they were all given another common examination in mathematics.

The means and standard deviations of the boys' and the girls' marks are shown in the table.

Examination marks				
	Upon entry		After 1 year	
	Mean	Standard deviation	Mean	Standard deviation
Boys	50	12	59	6
Girls	53	12	62	6

You may assume that the test results are normally distributed.

- (a) Test, at the 5% level of significance, whether or not the difference between the means of the boys' and girls' results was significant when they entered school.
- (b) Test, at the 5% level of significance, whether or not the mean mark of the boys is significantly less than the mean mark of the girls in the 'After 1 year' examination.
- (c) Interpret the results found in part (a) and part (b).

(7)

(5)

(1)

6. An area of grass was sampled by placing a $1\text{ m} \times 1\text{ m}$ square randomly in 100 places. The numbers of daisies in each of the squares were counted. It was decided that the resulting data could be modelled by a Poisson distribution with mean 2. The expected frequencies were calculated using the model.

The following table shows the observed and expected frequencies.

Number of daisies	Observed frequency	Expected frequency
0	8	13.53
1	32	27.07
2	27	r
3	18	s
4	10	9.02
5	3	3.61
6	1	1.20
7	0	0.34
≥ 8	1	t

- (a) Find values for r , s and t .

(4)

- (b) Using a 5% significance level, test whether or not this Poisson model is suitable. State your hypotheses clearly.

(7)

An alternative test might have been to estimate the population mean by using the data given.

- (c) Explain how this would have affected the test.

(2)

7. The numbers of deaths from pneumoconiosis and lung cancer in a developing country are given in the table.

Age group (years)	20–29	30–39	40–49	50–59	60–69	70 and over
Deaths from pneumoconiosis (1000s)	12.5	5.9	18.5	19.4	31.2	31.0
Deaths from lung cancer (1000s)	3.7	9.0	10.2	19.0	13.0	18.0

The correlation between the number of deaths in the different age groups for each disease is to be investigated.

- (a) Give **one** reason why Spearman's rank correlation coefficient should be used. (1)
- (b) Calculate Spearman's rank correlation coefficient for these data. (6)
- (c) Use a suitable test, at the 5% significance level, to interpret your result. State your hypotheses clearly. (5)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6691/01

Edexcel GCE

Statistics S3

Advanced Level

Monday 19 June 2006 – Morning

Time: 1 hour 30 minutes

Materials required for examination

Mathematical Formulae (Green)

Items included with question papers

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6685), your surname, other name and signature.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

This paper has 8 questions.

The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

N22341A

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.
©2006 Edexcel Limited

1. Describe one advantage and one disadvantage of

(a) quota sampling, (2)

(b) simple random sampling. (2)

2. A report on the health and nutrition of a population stated that the mean height of three-year old children is 90 cm and the standard deviation is 5 cm. A sample of 100 three-year old children was chosen from the population.

(a) Write down the approximate distribution of the sample mean height. Give a reason for your answer. (3)

(b) Hence find the probability that the sample mean height is at least 91 cm. (3)

3. A biologist investigated whether or not the diet of chickens influenced the amount of cholesterol in their eggs. The cholesterol content of 70 eggs selected at random from chickens fed diet *A* had a mean value of 198 mg and a standard deviation of 47 mg. A random sample of 90 eggs from chickens fed diet *B* had a mean cholesterol content of 201 mg and a standard deviation of 23 mg.

(a) Stating your hypotheses clearly and using a 5% level of significance, test whether or not there is a difference between the mean cholesterol content of eggs laid by chickens fed on these two diets. (7)

(b) State, in the context of this question, an assumption you have made in carrying out the test in part (a). (2)

4. The table below shows the price of an ice cream and the distance of the shop where it was purchased from a particular tourist attraction.

Shop	Distance from tourist attraction (m)	Price (£)
<i>A</i>	50	1.75
<i>B</i>	175	1.20
<i>C</i>	270	2.00
<i>D</i>	375	1.05
<i>E</i>	425	0.95
<i>F</i>	580	1.25
<i>G</i>	710	0.80
<i>H</i>	790	0.75
<i>I</i>	890	1.00
<i>J</i>	980	0.85

(a) Find, to 3 decimal places, the Spearman rank correlation coefficient between the distance of the shop from the tourist attraction and the price of an ice cream. (5)

(b) Stating your hypotheses clearly and using a 5% one-tailed test, interpret your rank correlation coefficient. (4)

5. The workers in a large office block use a lift that can carry a maximum load of 1090 kg. The weights of the male workers are normally distributed with mean 78.5 kg and standard deviation 12.6 kg. The weights of the female workers are normally distributed with mean 62.0 kg and standard deviation 9.8 kg.

Random samples of 7 males and 8 females can enter the lift.

(a) Find the mean and variance of the total weight of the 15 people that enter the lift. (4)

(b) Comment on any relationship you have assumed in part (a) between the two samples. (1)

(c) Find the probability that the maximum load of the lift will be exceeded by the total weight of the 15 people. (4)

6. A research worker studying colour preference and the age of a random sample of 50 children obtained the results shown below.

Age in years	Red	Blue	Totals
4	12	6	18
8	10	7	17
12	6	9	15
Totals	28	22	50

Using a 5% significance level, carry out a test to decide whether or not there is an association between age and colour preference. State your hypotheses clearly.

(11)

7. A machine produces metal containers. The weights of the containers are normally distributed. A random sample of 10 containers from the production line was weighed, to the nearest 0.1 kg, and gave the following results

49.7, 50.3, 51.0, 49.5, 49.9

50.1, 50.2, 50.0, 49.6, 49.7.

- (a) Find unbiased estimates of the mean and variance of the weights of the population of metal containers.

(5)

The machine is set to produce metal containers whose weights have a population standard deviation of 0.5 kg.

- (b) Estimate the limits between which 95% of the weights of metal containers lie.

(4)

- (c) Determine the 99% confidence interval for the mean weight of metal containers.

(5)

8. Five coins were tossed 100 times and the number of heads recorded. The results are shown in the table below.

Number of heads	0	1	2	3	4	5
Frequency	6	18	29	34	10	3

- (a) Suggest a suitable distribution to model the number of heads when five unbiased coins are tossed.

(2)

- (b) Test, at the 10% level of significance, whether or not the five coins are unbiased. State your hypotheses clearly.

(11)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6685/01**Edexcel GCE****Statistics S3****Advanced Level****Wednesday 20 June 2007 – Afternoon****Time: 1 hour 30 minutes****Materials required for examination**

Mathematical Formulae (Green)

Items included with question papers

Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6685), your surname, other name and signature.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

This paper has 7 questions.

The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

N29602A

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.
©2007 Edexcel Limited

1. During a village show, two judges, P and Q , had to award a mark out of 30 to some flower displays. The marks they awarded to a random sample of 8 displays were as follows:

Display	A	B	C	D	E	F	G	H
Judge P	25	19	21	23	28	17	16	20
Judge Q	20	9	21	13	17	14	11	15

- (a) Calculate Spearman's rank correlation coefficient for the marks awarded by the two judges. **(6)**

After the show, one competitor complained about the judges. She claimed that there was no positive correlation between their marks.

- (b) Stating your hypotheses clearly, test whether or not this sample provides support for the competitor's claim. Use a 5% level of significance. **(4)**

2. The Director of Studies at a large college believed that students' grades in Mathematics were independent of their grades in English. She examined the results of a random group of candidates who had studied both subjects and she recorded the number of candidates in each of the 6 categories shown.

	Maths grade A or B	Maths grade C or D	Maths grade E or U
English grade A or B	25	25	10
English grade C to U	15	30	15

- (a) Stating your hypotheses clearly, test the Director's belief using a 10% level of significance. You must show each step of your working. **(9)**

The Head of English suggested that the Director was losing accuracy by combining the English grades C to U in one row. He suggested that the Director should split the English grades into two rows, grades C or D and grades E or U as for Mathematics.

- (b) State why this might lead to problems in performing the test. **(1)**

N29602A

2

3. The time, in minutes, it takes Robert to complete the puzzle in his morning newspaper each day is normally distributed with mean 18 and standard deviation 3. After taking a holiday, Robert records the times taken to complete a random sample of 15 puzzles and he finds that the mean time is 16.5 minutes. You may assume that the holiday has not changed the standard deviation of times taken to complete the puzzle.

Stating your hypotheses clearly test, at the 5% level of significance, whether or not there has been a reduction in the mean time Robert takes to complete the puzzle.

(7)

4. A quality control manager regularly samples 20 items from a production line and records the number of defective items x . The results of 100 such samples are given in Table 1 below.

x	0	1	2	3	4	5	6	7 or more
Frequency	17	31	19	14	9	7	3	0

Table 1

- (a) Estimate the proportion of defective items from the production line.

(2)

The manager claimed that the number of defective items in a sample of 20 can be modelled by a binomial distribution. He used the answer in part (a) to calculate the expected frequencies given in Table 2.

x	0	1	2	3	4	5	6	7 or more
Expected frequency	12.2	27.0	r	19.0	s	3.2	0.9	0.2

Table 2

- (b) Find the value of r and the value of s giving your answers to 1 decimal place.
- (c) Stating your hypotheses clearly, use a 5% level of significance to test the manager's claim.
- (d) Explain what the analysis in part (c) tells the manager about the occurrence of defective items from this production line.

(3)

(7)

(1)

5. In a trial of diet A a random sample of 80 participants were asked to record their weight loss, x kg, after their first week of using the diet. The results are summarised by

$$\sum x = 361.6 \quad \text{and} \quad \sum x^2 = 1753.95.$$

- (a) Find unbiased estimates for the mean and variance of weight lost after the first week of using diet A.

(5)

The designers of diet A believe it can achieve a greater mean weight loss after the first week than a standard diet B. A random sample of 60 people used diet B. After the first week they had achieved a mean weight loss of 4.06 kg, with an unbiased estimate of variance of weight loss of 2.50 kg².

- (b) Test, at the 5% level of significance, whether or not the mean weight loss after the first week using diet A is greater than that using diet B. State your hypotheses clearly.

(7)

- (c) Explain the significance of the central limit theorem to the test in part (b).

(1)

- (d) State an assumption you have made in carrying out the test in part (b).

(1)

6. A random sample of the daily sales (in £s) of a small company is taken and, using tables of the normal distribution, a 99% confidence interval for the mean daily sales is found to be

$$(123.5, 154.7).$$

Find a 95% confidence interval for the mean daily sales of the company.

(6)

7. A set of scaffolding poles come in two sizes, long and short. The length L of a long pole has the normal distribution $N(19.7, 0.5^2)$. The length S of a short pole has the normal distribution $N(4.9, 0.2^2)$. The random variables L and S are independent. A long pole and a short pole are selected at random.

- (a) Find the probability that the length of the long pole is more than 4 times the length of the short pole.

(7)

Four short poles are selected at random and placed end to end in a row. The random variable T represents the length of the row.

- (b) Find the distribution of T .

(3)

- (c) Find $P(|L - T| < 0.1)$.

(5)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6691/01

Edexcel GCE

Statistics S3

Advanced Level

Monday 16 June 2008 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination

Mathematical Formulae (Green)

Items included with question papers

Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6691), your surname, other name and signature.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

This paper has 7 questions.

The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may gain no credit.

1. Some biologists were studying a large group of wading birds. A random sample of 36 were measured and the wing length, x mm, of each wading bird was recorded. The results are summarised as follows.

$$\sum x = 6046, \quad \sum x^2 = 1016338.$$

- (a) Calculate unbiased estimates of the mean and the variance of the wing lengths of these birds. **(3)**

Given that the standard deviation of the wing lengths of this particular type of bird is actually 5.1 mm,

- (b) find a 99 % confidence interval for the mean wing length of the birds from this group. **(5)**

2. Students in a mixed sixth form college are classified as taking courses in either Arts, Science or Humanities. A random sample of students from the college gave the following results.

		Course		
		Arts	Science	Humanities
Gender	Boy	30	50	35
	Girl	40	20	42

Showing your working clearly, test, at the 1 % level of significance, whether or not there is an association between gender and the type of course taken. State your hypotheses clearly.

(11)

3. The product moment correlation coefficient is denoted by r and Spearman's rank correlation coefficient is denoted by r_s .

- (a) Sketch separate scatter diagrams, with five points on each diagram, to show

- (i) $r = 1$,
 (ii) $r_s = -1$ but $r > -1$.

(3)

Two judges rank seven collie dogs in a competition. The collie dogs are labelled A to G and the rankings are as follows.

Rank	1	2	3	4	5	6	7
Judge 1	A	C	D	B	E	F	G
Judge 2	A	B	D	C	E	G	F

- (b) (i) Calculate Spearman's rank correlation coefficient for these data.

(6)

- (ii) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the judges are generally in agreement.

(5)

4. The weights of adult men are normally distributed with a mean of 84 kg and a standard deviation of 11 kg.

- (a) Find the probability that the total weight of 4 randomly chosen adult men is less than 350 kg. **(5)**

The weights of adult women are normally distributed with a mean of 62 kg and a standard deviation of 10 kg.

- (b) Find the probability that the weight of a randomly chosen adult man is less than one and a half times the weight of a randomly chosen adult woman.

(6)

5. A researcher is hired by a cleaning company to survey the opinions of employees on a proposed pension scheme. The company employs 55 managers and 495 cleaners.

To collect data the researcher decides to give a questionnaire to the first 50 cleaners to leave at the end of the day.

(a) Give 2 reasons why this method is likely to produce biased results. (2)

(b) Explain briefly how the researcher could select a sample of 50 employees using

- (i) a systematic sample,
(ii) a stratified sample.

(6)

Using the random number tables in the formulae book, and starting with the top left hand corner (8) and working across, 50 random numbers between 1 and 550 inclusive were selected. The first two suitable numbers are 384 and 100.

(c) Find the next two suitable numbers.

(2)

6. Ten cuttings were taken from each of 100 randomly selected garden plants. The numbers of cuttings that did not grow were recorded.

The results are as follows.

No. of cuttings which did not grow	0	1	2	3	4	5	6	7	8, 9 or 10
Frequency	11	21	30	20	12	3	2	1	0

(a) Show that the probability of a randomly selected cutting, from this sample, not growing is 0.223.

(2)

A gardener believes that a binomial distribution might provide a good model for the number of cuttings, out of 10, that do not grow.

He uses a binomial distribution, with the probability 0.2 of a cutting not growing. The calculated expected frequencies are as follows.

No. of cuttings which did not grow	0	1	2	3	4	5 or more
Expected frequency	r	26.84	s	20.13	8.81	t

(b) Find the values of r , s and t .

(4)

(c) State clearly the hypotheses required to test whether or not this binomial distribution is a suitable model for these data.

(2)

The test statistic for the test is 4.17 and the number of degrees of freedom used is 4.

(d) Explain fully why there are 4 degrees of freedom.

(2)

(e) Stating clearly the critical value used, carry out the test using a 5% level of significance.

(3)

7. A sociologist is studying how much junk food teenagers eat. A random sample of 100 female teenagers and an independent random sample of 200 male teenagers were asked to estimate what their weekly expenditure on junk food was. The results are summarised below.

	n	mean	s.d.
Female teenagers	100	£5.48	£3.62
Male teenagers	200	£6.86	£4.51

- (a) Using a 5% significance level, test whether or not there is a difference in the mean amounts spent on junk food by male teenagers and female teenagers. State your hypotheses clearly. (7)
- (b) Explain briefly the importance of the central limit theorem in this problem. (1)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6691/01

Edexcel GCE

Statistics S3

Advanced Level

Wednesday 17 June 2009 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination

Mathematical Formulae (Orange or Green)

Items included with question papers

Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6691), your surname, other name and signature.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

This paper has 8 questions.

The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

M34278A

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.
©2009 Edexcel Limited

1. A telephone directory contains 50 000 names. A researcher wishes to select a systematic sample of 100 names from the directory.

(a) Explain in detail how the researcher should obtain such a sample. (2)

(b) Give one advantage and one disadvantage of
 (i) quota sampling,
 (ii) systematic sampling. (4)

2. The heights of a random sample of 10 imported orchids are measured. The mean height of the sample is found to be 20.1 cm. The heights of the orchids are normally distributed.

Given that the population standard deviation is 0.5 cm,

(a) estimate limits between which 95% of the heights of the orchids lie, (3)

(b) find a 98% confidence interval for the mean height of the orchids. (4)

A grower claims that the mean height of this type of orchid is 19.5 cm.

(c) Comment on the grower's claim. Give a reason for your answer. (2)

3. A doctor is interested in the relationship between a person's Body Mass Index (BMI) and their level of fitness. She believes that a lower BMI leads to a greater level of fitness. She randomly selects 10 female 18 year-olds and calculates each individual's BMI. The females then run a race and the doctor records their finishing positions. The results are shown in the table.

Individual	A	B	C	D	E	F	G	H	I	J
BMI	17.4	21.4	18.9	24.4	19.4	20.1	22.6	18.4	25.8	28.1
Finishing position	3	5	1	9	6	4	10	2	7	8

(a) Calculate Spearman's rank correlation coefficient for these data. (5)

(b) Stating your hypotheses clearly and using a one tailed test with a 5% level of significance, interpret your rank correlation coefficient. (5)

(c) Give a reason to support the use of the rank correlation coefficient rather than the product moment correlation coefficient with these data. (1)

4. A sample of size 8 is to be taken from a population that is normally distributed with mean 55 and standard deviation 3. Find the probability that the sample mean will be greater than 57. (5)

5. The number of goals scored by a football team is recorded for 100 games. The results are summarised in Table 1 below.

Number of goals	Frequency
0	40
1	33
2	14
3	8
4	5

Table 1

- (a) Calculate the mean number of goals scored per game.

(2)

The manager claimed that the number of goals scored per match follows a Poisson distribution. He used the answer in part (a) to calculate the expected frequencies given in Table 2.

Number of goals	Expected Frequency
0	34.994
1	r
2	s
3	6.752
≥ 4	2.221

Table 2

- (b) Find the value of r and the value of s giving your answers to 3 decimal places.

(3)

- (c) Stating your hypotheses clearly, use a 5% level of significance to test the manager's claim.

(7)

6. The lengths of a random sample of 120 limpets taken from the upper shore of a beach had a mean of 4.97 cm and a standard deviation of 0.42 cm. The lengths of a second random sample of 150 limpets taken from the lower shore of the same beach had a mean of 5.05 cm and a standard deviation of 0.67 cm.

- (a) Test, using a 5% level of significance, whether or not the mean length of limpets from the upper shore is less than the mean length of limpets from the lower shore. State your hypotheses clearly.

(8)

- (b) State two assumptions you made in carrying out the test in part (a).

(2)

7. A company produces climbing ropes. The lengths of the climbing ropes are normally distributed. A random sample of 5 ropes is taken and the length, in metres, of each rope is measured. The results are given below.

120.3 120.1 120.4 120.2 119.9

- (a) Calculate unbiased estimates for the mean and the variance of the lengths of the climbing ropes produced by the company.

(5)

The lengths of climbing rope are known to have a standard deviation of 0.2 m. The company wants to make sure that there is a probability of at least 0.90 that the estimate of the population mean, based on a random sample size of n , lies within 0.05 m of its true value.

- (b) Find the minimum sample size required.

(6)

8. The random variable A is defined as

$$A = 4X - 3Y$$

where $X \sim N(30, 3^2)$, $Y \sim N(20, 2^2)$ and X and Y are independent.

Find

(a) $E(A)$, (2)

(b) $\text{Var}(A)$. (3)

The random variables Y_1, Y_2, Y_3 and Y_4 are independent and each has the same distribution as Y . The random variable B is defined as

$$B = \sum_{i=1}^4 Y_i.$$

(c) Find $P(B > A)$. (6)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6691/01

Edexcel GCE

Statistics S3

Advanced Level

Friday 18 June 2010 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination

Mathematical Formulae (Pink)

Items included with question papers

Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6691), your surname, other name and signature.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

This paper has 7 questions.

The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

H35397A

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.
©2010 Edexcel Limited

1. A report states that employees spend, on average, 80 minutes every working day on personal use of the Internet. A company takes a random sample of 100 employees and finds their mean personal Internet use is 83 minutes with a standard deviation of 15 minutes. The company's managing director claims that his employees spend more time on average on personal use of the Internet than the report states.

Test, at the 5% level of significance, the managing director's claim. State your hypotheses clearly.

(7)

2. Philip and James are racing car drivers. Philip's lap times, in seconds, are normally distributed with mean 90 and variance 9. James' lap times, in seconds, are normally distributed with mean 91 and variance 12. The lap times of Philip and James are independent. Before a race, they each take a qualifying lap.

(a) Find the probability that James' time for the qualifying lap is less than Philip's.

(4)

The race is made up of 60 laps. Assuming that they both start from the same starting line and lap times are independent,

(b) find the probability that Philip beats James in the race by more than 2 minutes.

(5)

3. A woodwork teacher measures the width, w mm, of a board. The measured width, X mm, is normally distributed with mean w mm and standard deviation 0.5 mm.

(a) Find the probability that X is within 0.6 mm of w .

(2)

The same board is measured 16 times and the results are recorded.

(b) Find the probability that the mean of these results is within 0.3 mm of w .

(4)

Given that the mean of these 16 measurements is 35.6 mm,

(c) find a 98% confidence interval for w .

(4)

4. A researcher claims that, at a river bend, the water gradually gets deeper as the distance from the inner bank increases. He measures the distance from the inner bank, b cm, and the depth of a river, s cm, at seven positions. The results are shown in the table below.

Position	A	B	C	D	E	F	G
Distance from inner bank b cm	100	200	300	400	500	600	700
Depth s cm	60	75	85	76	110	120	104

(a) Calculate Spearman's rank correlation coefficient between b and s .

(6)

(b) Stating your hypotheses clearly, test whether or not the data provides support for the researcher's claim. Use a 1% level of significance.

(4)

5. A random sample of 100 people were asked if their finances were worse, the same or better than this time last year. The sample was split according to their annual income and the results are shown in the table below.

Finances	Worse	Same	Better
Annual income			
Under £15 000	14	11	9
£15 000 and above	17	20	29

Test, at the 5% level of significance, whether or not the relative state of their finances is independent of their income range. State your hypotheses and show your working clearly.

(10)

6. A total of 228 items are collected from an archaeological site. The distance from the centre of the site is recorded for each item. The results are summarised in the table below.

Distance from the centre of the site (m)	0–1	1–2	2–4	4–6	6–9	9–12
Number of items	22	15	44	37	52	58

Test, at the 5% level of significance, whether or not the data can be modelled by a continuous uniform distribution. State your hypotheses clearly.

(12)

7. A large company surveyed its staff to investigate the awareness of company policy. The company employs 6000 full-time staff and 4000 part-time staff.

- (a) Describe how a stratified sample of 200 staff could be taken. (3)
- (b) Explain an advantage of using a stratified sample rather than a simple random sample. (1)

A random sample of 80 full-time staff and an independent random sample of 80 part-time staff were given a test of policy awareness. The results are summarised in the table below.

	Mean score (\bar{x})	Variance of scores (s^2)
Full-time staff	52	21
Part-time staff	50	19

- (c) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the mean policy awareness scores for full-time and part-time staff are different. (7)
- (d) Explain the significance of the Central Limit Theorem to the test in part (c). (2)
- (e) State an assumption you have made in carrying out the test in part (c). (1)

After all the staff had completed a training course the 80 full time staff and the 80 part-time staff were given another test of policy awareness. The value of the test statistic z was 2.53.

- (f) Comment on the awareness of company policy for the full-time and part-time staff in light of this result. Use a 1% level of significance. (2)
- (g) Interpret your answers to part (c) and part (f). (1)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6691/01

Edexcel GCE

Statistics S3

Advanced Level

Thursday 20 June 2011 – Morning

Time: 1 hour 30 minutes

Materials required for examination

Mathematical Formulae (Pink)

Items included with question papers

Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6691), your surname, other name and signature.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

This paper has 7 questions.

The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

P38359A

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.
©2011 Edexcel Limited

1. Explain what you understand by the Central Limit Theorem. (3)
-

2. A county councillor is investigating the level of hardship, h , of a town and the number of calls per 100 people to the emergency services, c . He collects data for 7 randomly selected towns in the county. The results are shown in the table below.

Town	A	B	C	D	E	F	G
h	14	20	16	18	37	19	24
c	52	45	43	42	61	82	55

- (a) Calculate the Spearman's rank correlation coefficient between h and c . (6)

After collecting the data, the councillor thinks there is no correlation between hardship and the number of calls to the emergency services.

- (b) Test, at the 5% level of significance, the councillor's claim. State your hypotheses clearly. (4)
-

3. A factory manufactures batches of an electronic component. Each component is manufactured in one of three shifts. A component may have one of two types of defect, D_1 or D_2 , at the end of the manufacturing process. A production manager believes that the type of defect is dependent upon the shift that manufactured the component. He examines 200 randomly selected defective components and classifies them by defect type and shift.

The results are shown in the table below.

Shift	Defect type	
	D_1	D_2
First shift	45	18
Second shift	55	20
Third shift	50	12

Stating your hypotheses, test, at the 10% level of significance, whether or not there is evidence to support the manager's belief. Show your working clearly.

(10)

4. A shop manager wants to find out if customers spend more money when music is playing in the shop. The amount of money spent by a customer in the shop is $\pounds x$. A random sample of 80 customers, who were shopping without music playing, and an independent random sample of 60 customers, who were shopping with music playing, were surveyed. The results of both samples are summarised in the table below.

	$\sum x$	$\sum x^2$	Unbiased estimate of mean	Unbiased estimate of variance
Customers shopping without music	5 320	392 000	x	s^2
Customers shopping with music	4 140	312 000	69.0	446.44

- (a) Find the values of \bar{x} and s^2 . (5)

- (b) Test, at the 5% level of significance, whether or not the mean money spent is greater when music is playing in the shop. State your hypotheses clearly. (8)
-

5. The number of hurricanes per year in a particular region was recorded over 80 years. The results are summarised in Table 1 below.

No of hurricanes, h	0	1	2	3	4	5	6	7
Frequency	0	2	5	17	20	12	12	12

Table 1

- (a) Write down two assumptions that will support modelling the number of hurricanes per year by a Poisson distribution. (2)
- (b) Show that the mean number of hurricanes per year from Table 1 is 4.4875. (2)
- (c) Use the answer in part (b) to calculate the expected frequencies r and s given in Table 2 below to 2 decimal places. (3)

h	0	1	2	3	4	5	6	7 or more
Expected frequency	0.90	4.04	r	13.55	s	13.65	10.21	13.39

Table 2

- (d) Test, at the 5% level of significance, whether or not the data can be modelled by a Poisson distribution. State your hypotheses clearly. (6)

6. The lifetimes of batteries from manufacturer A are normally distributed with mean 20 hours and standard deviation 5 hours when used in a camera.

- (a) Find the mean and standard deviation of the total lifetime of a pack of 6 batteries from manufacturer A . (2)

Judy uses a camera that takes one battery at a time. She takes a pack of 6 batteries from manufacturer A to use in her camera on holiday.

- (b) Find the probability that the batteries will last for more than 110 hours on her holiday. (2)

The lifetimes of batteries from manufacturer B are normally distributed with mean 35 hours and standard deviation 8 hours when used in a camera.

- (c) Find the probability that the total lifetime of a pack of 6 batteries from manufacturer A is more than 4 times the lifetime of a single battery from manufacturer B when used in a camera. (6)

7. Roastie's Coffee is sold in packets with a stated weight of 250 g. A supermarket manager claims that the mean weight of the packets is less than the stated weight. She weighs a random sample of 90 packets from their stock and finds that their weights have a mean of 248 g and a standard deviation of 5.4 g.

- (a) Using a 5% level of significance, test whether or not the manager's claim is justified. State your hypotheses clearly. (5)
- (b) Find the 98% confidence interval for the mean weight of a packet of coffee in the supermarket's stock. (4)
- (c) State, with a reason, the action you would recommend the manager to take over the weight of a packet of Roastie's Coffee. (2)

Roastie's Coffee company increase the mean weight of their packets to μ g and reduce the standard deviation to 3 g. The manager takes a sample of size n from these new packets. She uses the sample mean \bar{X} as an estimator of μ .

- (d) Find the minimum value of n such that $P(|\bar{X} - \mu| < 1) \geq 0.98$. (5)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6691/01**Edexcel GCE****Statistics S3****Advanced Level****Thursday 21 June 2012 – Afternoon****Time: 1 hour 30 minutes****Materials required for examination**

Mathematical Formulae (Pink)

Items included with question papers

Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6691), your surname, other name and signature.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

This paper has 7 questions.

The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

P40472A

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.
©2012 Edexcel Limited

1. Interviews for a job are carried out by two managers. Candidates are given a score by each manager and the results for a random sample of 8 candidates are shown in the table below.

Candidate	A	B	C	D	E	F	G	H
Manager X	62	56	87	54	65	15	12	10
Manager Y	54	47	71	50	49	25	30	44

- (a) Calculate Spearman's rank correlation coefficient for these data. (5)
- (b) Test, at the 5% level of significance, whether there is agreement between the rankings awarded by each manager. State your hypotheses clearly. (5)

Manager Y later discovered he had miscopied his score for candidate D and it should be 54.

- (c) Without carrying out any further calculations, explain how you would calculate Spearman rank correlation in this case. (2)

2. A lake contains 3 species of fish. There are estimated to be 1400 trout, 600 bass and 450 pike in the lake. A survey of the health of the fish in the lake is carried out and a sample of 30 fish is chosen.

- (a) Give a reason why stratified random sampling cannot be used. (1)
- (b) State an appropriate sampling method for the survey. (1)
- (c) Give one advantage and one disadvantage of this sampling method. (2)
- (d) Explain how this sampling method could be used to select the sample of 30 fish. You must show your working. (4)

3. (a) Explain what you understand by the Central Limit Theorem.

A garage services hire cars on behalf of a hire company. The garage knows that the lifetime of the brake pads has a standard deviation of 5000 miles. The garage records the lifetimes, x miles, of the brake pads it has replaced. The garage takes a random sample of 100 brake pads and finds that $\sum x = 1\,740\,000$.

- (b) Find a 95% confidence interval for the mean lifetime of a brake pad. (5)

- (c) Explain the relevance of the Central Limit Theorem in part (b). (2)

Brake pads are made to be changed every 20 000 miles on average. The hire car company complain that the garage is changing the brake pads too soon.

- (d) Comment on the hire company's complaint. Give a reason for your answer. (2)

4. Two breeds of chicken are surveyed to measure their egg yield. The results are shown in the table below.

Breed	Egg yield		
	Low	Medium	High
Leghorn	22	52	26
Cornish	14	32	4

Showing each stage of your working clearly, test, at the 5% level of significance, whether or not there is an association between egg yield and breed of chicken. State your hypotheses clearly. (10)

5. Mr Allan and Ms Burns are two mathematics teachers teaching mixed ability groups of students in a large college. At the end of the college year all students took the same examination. A random sample of 29 of Mr Allan's students and a random sample of 26 of Ms Burns' students are chosen. The results are summarised in the table below.

	Sample Size, n	Mean, \bar{x}	Standard Deviation, s
Mr Allan	29	80	10
Ms Burns	26	74	15

- (a) Stating your hypothesis clearly, test, at the 10% level of significance, whether there is evidence that there is a difference in the means scores of their students. (6)

Ms Burns thinks the comparison was unfair as the examination was set by Mr Allan. She looks up a different set of examination marks for these students and, although Mr Allan's sample has a higher mean, she calculates the test statistic for this new set of results to be 1.6.

However, Mr Allan now claims that the mean marks of his students are higher than the mean marks of Ms Burns' students.

- (b) Test Mr Allan's claim, stating the hypothesis and critical values you would use. Use a 10% level of significance. (3)

6. A total of 100 random samples of 6 items are selected from a production line in a factory and the number of defective items in each sample is recorded. The results are summarised in the table below.

Number of defective items	0	1	2	3	4	5	6
Number of samples	6	16	20	23	17	10	8

- (a) Show that the mean number of defective items per sample is 2.91. (2)

A factory manager suggests that the data can be modelled by a binomial distribution with $n = 6$. He uses the mean from the sample above and calculates expected frequencies as shown in the table below.

Number of defective items	0	1	2	3	4	5	6
Expected frequency	1.87	10.54	24.82	a	22.01	8.29	b

- (b) Calculate the value of a and the value of b , giving your answers to 2 decimal places. (4)
- (c) Test, at the 5% level, whether or not the binomial distribution is a suitable model for the number of defective items in samples of 6 items. State your hypotheses clearly. (8)

7. The heights, in cm, of the male employees in a large company follow a normal distribution with mean 177 and standard deviation 5.

The heights, in cm, of the female employees follow a normal distribution with mean 163 and standard deviation 4.

A male employee and a female employee are chosen at random.

- (a) Find the probability that the male employee is taller than the female employee. (5)

Six male employees and four female employees are chosen at random.

- (b) Find the probability that their total height is less than 17 m. (6)

END **TOTAL FOR PAPER: 75 MARKS**

Paper Reference(s)

6691/01R Edexcel GCE

Statistics S3 (R)

Advanced/Advanced Subsidiary

Thursday 13 June 2013 – Morning

Time: 1 hour 30 minutes

Materials required for examination
Mathematical Formulae (Pink)

Items included with question papers
Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

This paper is strictly for students outside the UK.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature. Check that you have the correct question paper. Answer ALL the questions. You must write your answer for each question in the space following the question. Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided. Full marks may be obtained for answers to ALL questions. The marks for the parts of questions are shown in round brackets, e.g. (2). There are 8 questions in this question paper. The total mark for this paper is 75. There are 24 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled. You must show sufficient working to make your methods clear to the Examiner. Answers without working may not gain full credit.

1. A gym club has 400 members of which 300 are males.

Explain clearly how a stratified sample of size 60 could be taken.

(3)

2. A random sample of size n is to be taken from a population that is normally distributed with mean 40 and standard deviation 3. Find the minimum sample size such that the probability of the sample mean being greater than 42 is less than 5%.

(5)

3. The table below shows the population and the number of council employees for different towns and villages.

Town or village	Population	Number of council employees
<i>A</i>	211	10
<i>B</i>	356	2
<i>C</i>	1047	12
<i>D</i>	2463	21
<i>E</i>	4892	16
<i>F</i>	6479	25
<i>G</i>	6571	67
<i>H</i>	6573	45
<i>I</i>	9845	48
<i>J</i>	14 784	34

- (a) Find, to 3 decimal places, Spearman's rank correlation coefficient between the population and the number of council employees.

(5)

- (b) Use your value of Spearman's rank correlation coefficient to test for evidence of a positive correlation between the population and the number of council employees. Use a 2.5% significance level. State your hypotheses clearly.

(4)

It is suggested that a product moment correlation coefficient would be a more suitable calculation in this case. The product moment correlation coefficient for these data is 0.627 to 3 decimal places.

- (c) Use the value of the product moment correlation coefficient to test for evidence of a positive correlation between the population and the number of council employees. Use a 2.5% significance level.

(2)

- (d) Interpret and comment on your results from part (b) and part (c).

(2)

4. John thinks that a person's eye colour is related to their hair colour. He takes a random sample of 600 people and records their eye and hair colours. The results are shown in Table 1.

		Hair colour				Total
		Black	Brown	Red	Blonde	
Eye colour	Brown	45	125	15	58	243
	Blue	34	90	10	58	192
	Hazel	20	38	16	26	100
	Green	6	29	7	23	65
	Total	105	282	48	165	600

Table 1

John carries out a χ^2 test in order to test whether eye colour and hair colour are related. He calculates the expected frequencies shown in Table 2.

		Hair colour			
		Black	Brown	Red	Blonde
Eye colour	Brown	42.5	114.2	19.4	66.8
	Blue	33.6	90.2	15.4	52.8
	Hazel	17.5	47	8	27.5
	Green	11.4	30.6	5.2	17.9

Table 2

- (a) Show how the value 47 in Table 2 has been calculated. (1)
- (b) Write down the number of degrees of freedom John should use in this χ^2 test. (1)

Given that the value of the χ^2 statistic is 20.6, to 3 significant figures,

- (c) find the smallest value of α for which the null hypothesis will be rejected at the $\alpha\%$ level of significance. (1)
- (d) Use the data from Table 1 to test at the 5% level of significance whether or not the proportions of people in the population with black, brown, red and blonde hair are in the ratio 2:6:1:3. State your hypotheses clearly. (9)

5. A manufacturer produces circular discs with diameter D mm, such that $D \sim N(\mu, \sigma^2)$. A random sample of discs is taken and, using tables of the normal distribution, a 90% confidence interval for μ is found to be

$$(118.8, 121.2)$$

- (a) Find a 98% confidence interval for μ . (6)
- (b) Hence write down a 98% confidence interval for the circumference of the discs. (1)

Using three different random samples, three 98% confidence intervals for μ are to be found.

- (c) Calculate the probability that all the intervals will contain μ . (2)

6. The continuous random variable X is uniformly distributed over the interval

$$[a - 1, a + 5]$$

where a is a constant.

Fifty observations of X are taken, giving a sample mean of 17.2.

- (a) Use the Central Limit Theorem to find an approximate distribution for \bar{X} . (3)
- (b) Hence find a 95% confidence interval for a . (4)

7. A farmer monitored the amount of lead in soil in a field next to a factory. He took 100 samples of soil, randomly selected from different parts of the field, and found the mean weight of lead to be 67 mg/kg with standard deviation 25 mg/kg. After the factory closed, the farmer took 150 samples of soil, randomly selected from different parts of the field, and found the mean weight of lead to be 60 mg/kg with standard deviation 10 mg/kg.
- (a) Test at the 5% level of significance whether or not the mean weight of lead in the soil decreased after the factory closed. State your hypotheses clearly. (7)
- (b) Explain the significance of the Central Limit Theorem to the test in part (a). (1)
- (c) State an assumption you have made to carry out this test. (1)

8. A farmer supplies both duck eggs and chicken eggs. The weights of duck eggs, D grams, and chicken eggs, C grams, are such that

$$D \sim N(54, 1.2^2) \text{ and } C \sim N(44, 0.8^2).$$

- (a) Find the probability that the weights of 2 randomly selected duck eggs will differ by more than 3 g. (6)
- (b) Find the probability that the weight of a randomly selected chicken egg is less than $\frac{4}{5}$ of the weight of a randomly selected duck egg. (5)

Eggs are packed in boxes which contain either 6 randomly selected duck eggs or 6 randomly selected chicken eggs. The weight of an empty box has distribution $N(28, \sqrt{5}^2)$.

- (c) Find the probability that a full box of duck eggs weighs at least 50 g more than a full box of chicken eggs. (6)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6691/01

Edexcel GCE

Statistics S3

Advanced/Advanced Subsidiary

Thursday 13 June 2013 – Morning

Time: 1 hour 30 minutes

Materials required for examination

Mathematical Formulae (Pink)

Items included with question papers

Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature. Check that you have the correct question paper. Answer ALL the questions.

You must write your answer for each question in the space following the question.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

The marks for the parts of questions are shown in round brackets, e.g. (2).

There are 7 questions in this question paper. The total mark for this paper is 75.

There are 20 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

P41817A

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.
©2013 Edexcel Limited.

1. A doctor takes a random sample of 100 patients and measures their intake of saturated fats in their food and the level of cholesterol in their blood. The results are summarised in the table below.

Cholesterol level \ Intake of saturated fats		High	Low
		High	8
High	12	8	
Low	26	54	

Using a 5% level of significance, test whether or not there is an association between cholesterol level and intake of saturated fats. State your hypotheses and show your working clearly.

(10)

2. The table below shows the number of students per member of staff and the student satisfaction scores for 7 universities.

University	A	B	C	D	E	F	G
Number of students per member of staff	14.2	13.1	13.3	11.7	10.5	15.9	10.8
Student satisfaction score	4.1	4.2	3.8	4.0	3.9	4.3	3.7

- (a) Calculate Spearman's rank correlation coefficient for these data.

(5)

- (b) Stating your hypotheses clearly test, at the 5% level of significance, whether or not there is evidence of a correlation between the number of students per member of staff and the student satisfaction score.

(3)

3. A college manager wants to survey students' opinions of enrichment activities. She decides to survey the students on the courses summarised in the table below.

Course	Number of students enrolled
Leisure and Sport	420
Information Technology	337
Health and Social Care	200
Media Studies	43

Each student takes only one course.

The manager has access to the college's information system that holds full details of each of the enrolled students including name, address, telephone number and their course of study. She wants to compare the opinions of students on each course and has a generous budget to pay for the cost of the survey.

- (a) Give one advantage and one disadvantage of carrying out this survey using

(i) quota sampling,

(ii) stratified sampling.

(2)

The manager decides to take a stratified sample of 100 students.

- (b) Calculate the number of students to be sampled from each course.

(3)

- (c) Describe how to choose students for the stratified sample.

(2)

4. Customers at a post office are timed to see how long they wait until being served at the counter. A random sample of 50 customers is chosen and their waiting times, x minutes, are summarised in Table 1.

Waiting time in minutes (x)	Frequency
0–3	8
3–5	12
5–6	13
6–8	9
8–12	8

Table 1

- (a) Show that an estimate of $\bar{x} = 5.49$ and an estimate of $s_x^2 = 6.88$. (3)

The post office manager believes that the customers' waiting times can be modelled by a normal distribution.

Assuming the data is normally distributed, she calculates the expected frequencies for these data and some of these frequencies are shown in Table 2.

Waiting Time	$x < 3$	3–5	5–6	6–8	$x > 8$
Expected Frequency	8.56	12.73	7.56	a	b

Table 2

- (b) Find the value of a and the value of b . (3)
- (c) Test, at the 5% level of significance, the manager's belief. State your hypotheses clearly. (8)
-

5. Blumen is a perfume sold in bottles. The amount of perfume in each bottle is normally distributed. The amount of perfume in a large bottle has mean 50ml and standard deviation 5ml. The amount of perfume in a small bottle has mean 15ml and standard deviation 3ml.

One large and 3 small bottles of Blumen are chosen at random.

- (a) Find the probability that the amount in the large bottle is less than the total amount in the 3 small bottles. (6)

A large bottle and a small bottle of Blumen are chosen at random.

- (b) Find the probability that the large bottle contains more than 3 times the amount in the small bottle. (6)

6. Fruit-n-Veg4U Market Gardens grow tomatoes. They want to improve their yield of tomatoes by at least 1 kg per plant by buying a new variety. The variance of the yield of the old variety of plant is 0.5 kg^2 and the variance of the yield for the new variety of plant is 0.75 kg^2 . A random sample of 60 plants of the old variety has a mean yield of 5.5 kg. A random sample of 70 of the new variety has a mean yield of 7 kg.

- (a) Stating your hypotheses clearly test, at the 5% level of significance, whether or not there is evidence that the mean yield of the new variety is more than 1 kg greater than the mean yield of the old variety. (9)
- (b) Explain the relevance of the Central Limit Theorem to the test in part (a). (2)

7. Lambs are born in a shed on Mill Farm. The birth weights, x kg, of a random sample of 8 newborn lambs are given below.

4.12 5.12 4.84 4.65 3.55 3.65 3.96 3.40

- (a) Calculate unbiased estimates of the mean and variance of the birth weight of lambs born on Mill Farm.

(3)

A further random sample of 32 lambs is chosen and the unbiased estimates of the mean and variance of the birth weight of lambs from this sample are 4.55 and 0.25 respectively.

- (b) Treating the combined sample of 40 lambs as a single sample, estimate the standard error of the mean.

(7)

The owner of Mill Farm researches the breed of lamb and discovers that the population of birth weights is normally distributed with standard deviation 0.67 kg.

- (c) Calculate a 95% confidence interval for the mean birth weight of this breed of lamb using your combined sample mean.

(5)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6691/01R

Edexcel GCE

Statistics S3 (R)

Advanced Subsidiary

Thursday 22 May 2014 – Morning

Time: 1 hour 30 minutes

Materials required for examination

Mathematical Formulae (Pink)

Items included with question papers

Nil

Calculators may NOT be used in this examination.

This paper is strictly for students outside the UK.

Instructions to Candidates

Write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6691R), your surname, initials and signature.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for the parts of questions are shown in round brackets, e.g. (2).
There are 7 questions in this question paper. The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You must show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

1. A journalist is investigating factors which influence people when they buy a new car. One possible factor is fuel efficiency. The journalist randomly selects 8 car models. Each model's annual sales and fuel efficiency, in km/litre, are shown in the table below.

Car model	A	B	C	D	E	F	G	H
Annual sales	1800	5400	18100	7100	9300	4800	12200	10700
Fuel efficiency	5.2	18.6	14.8	13.2	18.3	11.9	16.5	17.7

- (a) Calculate Spearman's rank correlation coefficient for these data. (5)

The journalist believes that car models with higher fuel efficiency will achieve higher sales.

- (b) Stating your hypotheses clearly, test whether or not the data support the journalist's belief. Use a 5% level of significance. (4)

- (c) State the assumption necessary for a product moment correlation coefficient to be valid in this case. (1)

- (d) The mean and median fuel efficiencies of the car models in the random sample are 14.5 km/litre and 15.65 km/litre respectively. Considering these statistics, as well as the distribution of the fuel efficiency data, state whether or not the data suggest that the assumption in part (c) might be true in this case. Give a reason for your answer. (No further calculations are required.) (1)

2. A survey asked a random sample of 200 people their age and the main use of their mobile phone.

The results are shown in Table 1 below.

		Main use of their mobile phone		
		Internet	Texts	Phone calls
Age	Under 20	27	14	9
	From 20 to 40	32	34	29
	Over 40	15	19	21

Table 1

The data are to be used to test whether or not age and main use of their mobile phone are independent.

Table 2 shows the expected frequencies for each group, assuming people's age and main use of their mobile phone are independent.

		Main use of their mobile phone		
		Internet	Texts	Phone calls
Age	Under 20	18.5	16.75	14.75
	From 20 to 40	35.15	31.825	28.025
	Over 40	20.35	18.425	16.225

Table 2

- (a) For users under 20 choosing the Internet as the main use of their mobile phone,
- (i) verify that the expected frequency is 18.5.
 - (ii) show that the contribution to the χ^2 test statistic is 3.91 to 3 significant figures. (2)
- (b) Given that the χ^2 test statistic for the data is 9.893 to 3 decimal places, test at the 5% level of significance whether or not age and main use of their mobile phone are independent. State your hypotheses clearly. (5)

3. A company produces two types of milk powder, 'Semi-Skimmed' and 'Full Cream'. In tests, each type of milk powder is used to make a large number of cups of coffee. The mass, S grams, of 'Semi-Skimmed' milk powder used in one cup of coffee is modelled by $S \sim N(4.9, 0.8^2)$. The mass, C grams, of 'Full Cream' milk powder used in one cup of coffee is modelled by $C \sim N(2.5, 0.4^2)$.

(a) Two cups of coffee, one with each type of milk powder, are to be selected at random. Find the probability that the mass of 'Semi-Skimmed' milk powder used will be at least double that of the 'Full Cream' milk powder used.

(6)

(b) 'Semi-Skimmed' milk powder is sold in 500g packs. Find the probability that one pack will be sufficient for 100 cups of coffee.

(5)

4. A manufacturing company produces solar panels. The output of each solar panel is normally distributed with standard deviation 6 watts. It is thought that the mean output, μ , is 160 watts.

A researcher believes that the mean output of the solar panels is greater than 160 watts. He writes down the output values of 5 randomly selected solar panels. He uses the data to carry out a hypothesis test at the 5% level of significance.

He tests $H_0: \mu = 160$ against $H_1: \mu > 160$.

On reporting to his manager, the researcher can only find 4 of the output values. These are shown below

168.2 157.4 173.3 161.1

Given that the result of the hypothesis test is that there is significant evidence to reject H_0 at the 5% level of significance, calculate the minimum possible missing output value, μ . Give your answer correct to 1 decimal place.

(6)

5. A student believes that there is a difference in the mean lengths of English and French films. He goes to the university video library and randomly selects a sample of 120 English films and a sample of 70 French films. He notes the length, x minutes, of each of the films in his samples. His data are summarised in the table below.

	Σx	Σx^2	s^2	n
English films	10650	956909	98.5	120
French films	6510	615849	151	70

(a) Verify that the unbiased estimate of the variance, s^2 , of the lengths of English films is 98.5 minutes².

(2)

(b) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the mean lengths of English and French films are different.

(7)

(c) Explain the significance of the Central Limit Theorem to the test in part (b).

(1)

(d) The university video library contained 724 English films and 473 French films. Explain how the student could have taken a stratified sample of 190 of these films.

(3)

6. Bags of £1 coins are paid into a bank. Each bag contains 20 coins.

The bank manager believes that 5% of the £1 coins paid into the bank are fakes. He decides to use the distribution $X \sim B(20, 0.05)$ to model the random variable X , the number of fake £1 coins in each bag.

- (a) State the assumptions necessary for the binomial distribution to be an appropriate model in this case. (2)

The bank manager checks a random sample of 150 bags of £1 coins and records the number of fake coins found in each bag. His results are summarised in Table 1.

Number of fake coins in each bag	0	1	2	3	4 or more
Observed frequency	43	62	26	13	6
Expected frequency	53.8	56.6	r	8.9	s

Table 1

- (b) Calculate the values of r and s , giving your answers to 1 decimal place. (3)
- (c) Carry out a hypothesis test, at the 5% significance level, to see if the data supports the bank manager's statistical model. State your hypotheses clearly. (7)

The assistant manager thinks that a binomial distribution is a good model but suggests that the proportion of fake coins is higher than 5%. She calculates the actual proportion of fake coins in the sample and uses this value to carry out a new hypothesis test on the data. Her expected frequencies are shown in Table 2.

Number of fake coins in each bag	0	1	2	3	4 or more
Observed frequency	43	62	26	13	6
Expected frequency	44.5	55.7	33.2	12.5	4.1

Table 2

- (d) Explain why there are 2 degrees of freedom in this case. (2)
- (e) Given that she obtains a χ^2 test statistic of 2.67, test the assistant manager's hypothesis that the binomial distribution is a good model for the number of fake coins in each bag. Use a 5% level of significance and state your hypotheses clearly. (3)

7. A petrol pump is tested regularly to check that the reading on its gauge is accurate. The random variable X , in litres, is the quantity of petrol actually dispensed when the gauge reads 10.00 litres. X is known to have distribution $X \sim N(\mu, 0.08^2)$.

- (a) Eight random tests gave the following values of x

10.01 9.97 9.93 9.99 9.90 9.95 10.13 9.94

- (i) Find a 95% confidence interval for μ to 2 decimal places. (5)
- (ii) Use your result to comment on the accuracy of the petrol gauge. (5)
- (b) A sample mean of 9.96 litres was obtained from a random sample of n tests. A 90% confidence interval for μ gave an upper limit of less than 10.00 litres. Find the minimum value of n . (5)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6691/01**Edexcel GCE****Statistics S3****Advanced Subsidiary****Thursday 22 May 2014 – Morning****Time: 1 hour 30 minutes****Materials required for examination**

Mathematical Formulae (Pink)

Items included with question papers

Nil

Calculators may NOT be used in this examination.**This paper is strictly for students outside the UK.****Instructions to Candidates**

Write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6691), your surname, initials and signature. Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided. Full marks may be obtained for answers to ALL questions. The marks for the parts of questions are shown in round brackets, e.g. (2). There are 8 questions in this question paper. The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled. You must show sufficient working to make your methods clear to the Examiner. Answers without working may not gain full credit.

P43156A

This publication may only be reproduced in accordance with Pearson Education Limited copyright policy.
©2014 Pearson Education Limited.

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)

2. The weights of pears in an orchard are assumed to have unknown mean μ and unknown standard deviation σ .

A random sample of 20 pears is taken and their weights recorded.

The sample is represented by X_1, X_2, \dots, X_{20} . State whether or not the following are statistics. Give reasons for your answers.

- (a) (i) $\frac{X_1 + 3X_{20}}{2}$
(ii) $\sum_{i=1}^{20} (X_i - \mu)$
(iii) $\sum_{i=1}^{20} \left(\frac{X_i - \mu}{\sigma} \right)$ (4)

- (b) Find the mean and variance of $\frac{3X_1 - X_{20}}{2}$. (3)

3. A number of males and females were asked to rate their happiness under the headings “not happy”, “fairly happy” and “very happy”.

The results are shown in the table below

		Happiness			Total
		Not happy	Fairly happy	Very happy	
Gender	Female	9	43	34	86
	Male	13	25	16	54
Total		22	68	50	140

Stating your hypotheses, test at the 5% level of significance, whether or not there is evidence of an association between happiness and gender. Show your working clearly.

(10)

4. The random variable A is defined as

$$A = B + 4C - 3D$$

where B , C and D are independent random variables with

$$B \sim N(6, 2^2) \quad C \sim N(7, 3^2) \quad D \sim N(4, 1.5^2)$$

Find $P(A < 45)$.

(6)

5. A research station is doing some work on the germination of a new variety of genetically modified wheat.

They planted 120 rows containing 7 seeds in each row.

The number of seeds germinating in each row was recorded. The results are as follows

Number of seeds germinating in each row	0	1	2	3	4	5	6	7
Observed number of rows	2	6	11	19	25	32	16	9

- (a) Write down two reasons why a binomial distribution may be a suitable model.

(2)

- (b) Show that the probability of a randomly selected seed from this sample germinating is 0.6.

(2)

The research station used a binomial distribution with probability 0.6 of a seed germinating. The expected frequencies were calculated to 2 decimal places. The results are as follows:

Number of seeds germinating in each row	0	1	2	3	4	5	6	7
Observed number of rows	0.20	2.06	s	23.22	t	31.35	15.68	3.36

- (c) Find the value of s and the value of t .

(2)

- (d) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the data can be modelled by a binomial distribution.

(7)

6. A random sample X_1, X_2, \dots, X_n is taken from a population with mean μ .

(a) Show that $\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ is an unbiased estimator of the population mean μ . (1)

A company produces small jars of coffee.

Five jars of coffee were taken at random and weighed.

The weights, in grams, were as follows

197 203 205 201 195

(b) Calculate unbiased estimates of the population mean and variance of the weights of the jars produced by the company. (3)

It is known from previous results that the weights are normally distributed with standard deviation 4.8 g.

The manager is going to take a second random sample. He wishes to ensure that there is at least a 95% probability that the estimate of the population mean is within 1.25 g of its true value.

(c) Find the minimum sample size required. (4)

7. A machine fills packets with X grams of powder where X is normally distributed with mean μ . Each packet is supposed to contain 1 kg of powder.

To comply with regulations, the weight of powder in a randomly selected packet should be such that $P(X < \mu - 30) = 0.0005$.

(a) Show that this requires the standard deviation to be 9.117 g to 3 decimal places. (3)

A random sample of 10 packets is selected from the machine. The weight, in grams, of powder in each packet is as follows:

999.8 991.6 1000.3 1006.1 1008.2 997.0 993.2 1000.0 997.1 1002.1

(b) Assuming that the standard deviation of the population is 9.117 g, test, at the 1% significance level, whether or not the machine is delivering packets with mean weight of less than 1 kg. State your hypotheses clearly. (7)

8. The heights, in metres, and weights, in kilograms, of a random sample of 9 men are shown in the table below.

Man	A	B	C	D	E	F	G	H	I
Height (x)	1.68	1.74	1.75	1.76	1.78	1.82	1.84	1.88	1.98
Weight (y)	75	76	100	77	90	95	110	96	120

(a) Given that $S_{xx} = 0.0632$, $S_{yy} = 1957.5556$ and $S_{xy} = 9.3433$ calculate, to 3 decimal places, the product moment correlation coefficient between height and weight for these men. (2)

(b) Use your value of the product moment correlation coefficient to test whether or not there is evidence of a positive correlation between the height and weight of men. Use a 5% significance level. State your hypotheses clearly. (4)

Peter does not know the heights or weights of the 9 men. He is given photographs of them and asked to put them in order of increasing weight. He puts them in the order

A C E B G D I F H

(c) Find, to 3 decimal places, Spearman's rank correlation coefficient between Peter's order and the actual order. (6)

(d) Use your value of Spearman's rank correlation coefficient to test for evidence of Peter's ability to correctly order men, by their weight, from their photographs. Use a 5% significance level and state your hypotheses clearly. (4)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

WST03/01

Pearson Edexcel International Advanced Level

Statistics S3

Advanced/Advanced Subsidiary

Thursday 22 May 2014 – Morning

Time: 1 hour 30 minutes

Materials required for examination

Mathematical Formulae (Blue)

Items included with question papers

Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.

Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B). Coloured pencils and highlighter pens must not be used.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.
- Answer the questions in the spaces provided – *there may be more space than you need*.
- You should show sufficient working to make your methods clear. Answers without working may not gain full credit.
- Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information

- The total mark for this paper is 75.
- The marks for **each** question are shown in brackets – *use this as a guide as to how much time to spend on each question*.

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

P44515A

This publication may only be reproduced in accordance with Pearson Education Limited copyright policy.
©2014 Pearson Education Limited.

1. A tennis club's committee wishes to select a sample of 50 members to fill in a questionnaire about the club's facilities. The 300 members, of whom 180 are males, are listed in alphabetical order and numbered 1 – 300 in the club's membership book.

The club's committee decides to use a random number table to obtain its sample. The first three lines of the random number table used are given below.

319	952	241	343	278	811	394	165	008	413	063	179	749
722	962	334	461	267	114	806	992	414	837	837	657	339
470	684	554	127	067	459	142	920	144	575	311	605	412

Starting with the top left-hand corner (319) and working across, the committee selects 50 random numbers. The first 2 suitable numbers are 241 and 278. Numbers greater than 300 are ignored.

(a) Find the next two suitable numbers.

(1)

When the club's committee looks at the members corresponding to their random numbers they find that only 1 female has been selected.

The committee does not want to be accused of being biased towards males so considers using a systematic sample instead.

(b) (i) Explain clearly how the committee could take a systematic sample.

(ii) Explain why a systematic sample may not give a sample that represents the proportion of males and females in the club.

(3)

The committee decides to use a stratified sample instead.

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

(3)

(d) Explain an advantage of using a stratified sample rather than a quota sample.

(1)

2. The random variable X follows a continuous uniform distribution over the interval $[\alpha - 3, 2\alpha + 3]$ where α is a constant.

The mean of a random sample of size n is denoted by \bar{X} .

- (a) Show that \bar{X} is a biased estimator of α , and state the bias. (3)

Given that $Y = k\bar{X}$ is an unbiased estimator for α ,

- (b) find the value of k . (1)

A random sample of 10 values of X is taken and the results are as follows

3 5 8 12 4 13 10 8 5 12

- (c) Hence estimate the maximum value of X . (3)

3. A grocer believes that the average weight of a grapefruit from farm A is greater than the average weight of a grapefruit from farm B . The weights, in grams, of 80 grapefruit selected at random from farm A have a mean value of 532 g and a standard deviation, s_A , of 35 g. A random sample of 100 grapefruit from farm B have a mean weight of 520 g and a standard deviation, s_B , of 28 g.

Stating your hypotheses clearly and using a 1% level of significance, test whether or not the grocer's belief is supported by the data.

(7)

4. In a survey 10 randomly selected men had their systolic blood pressure, x , and weight, w , measured. Their results are as follows:

Man	A	B	C	D	E	F	G	H	I	J
x	123	128	137	143	149	153	154	159	162	168
w	78	93	85	83	75	98	88	87	95	99

- (a) Calculate the value of Spearman's rank correlation coefficient between x and w . (5)

- (b) Stating your hypotheses clearly, test at the 5% level of significance, whether or not there is evidence of a positive correlation between systolic blood pressure and weight. (4)

The product moment correlation coefficient for these data is 0.5114.

- (c) Use the value of the product moment correlation coefficient to test, at the 5% level of significance, whether or not there is evidence of a positive correlation between systolic blood pressure and weight. (2)

- (d) Using your conclusions to part (b) and part (c), describe the relationship between systolic blood pressure and weight. (1)

5. A random sample of 200 people were asked which hot drink they preferred from tea, coffee and hot chocolate. The results are given below.

		Type of drink preferred			Total
		Tea	Coffee	Hot Chocolate	
Gender	Males	57	26	11	94
	Females	42	47	17	106
Total		99	73	28	200

- (a) Test, at the 5% significance level, whether or not there is an association between type of drink preferred and gender. State your hypotheses and show your working clearly. You should state your expected frequencies to 2 decimal places. (10)

- (b) State what difference using a 0.5% significance level would make to your conclusion. Give a reason for your answer. (2)

6. Eight tasks were given to each of 125 randomly selected job applicants. The number of tasks failed by each applicant is recorded.

The results are as follows:

Number of tasks failed by an applicant	0	1	2	3	4	5	6 or more
Frequency	2	21	45	42	12	3	0

- (a) Show that the probability of a randomly selected task, from this sample, being failed is 0.3. (2)

An employer believes that a binomial distribution might provide a good model for the number of tasks, out of 8, that an applicant fails.

He uses a binomial distribution, with the estimated probability 0.3 of a task being failed. The calculated expected frequencies are as follows

Number of tasks failed by an applicant	0	1	2	3	4	5	6 or more
Frequency	7.21	24.71	37.06	r	17.02	5.83	s

- (b) Find the value of r and the value of s giving your answers to 2 decimal places. (3)
- (c) Test, at the 5% level of significance, whether or not a binomial distribution is a suitable model for these data. State your hypotheses and show your working clearly. (8)

The employer believes that all applicants have the same probability of failing each task.

- (d) Use your result from part (c) to comment on this belief. (1)

7. The random variable X is defined as

$$X = 4Y - 3W$$

where $Y \sim N(40, 3^2)$, $W \sim N(50, 2^2)$ and Y and W are independent.

- (a) Find $P(X > 25)$. (7)

The random variables Y_1 , Y_2 and Y_3 are independent and each has the same distribution as Y . The random variable A is defined as

$$A = \sum_{i=1}^3 Y_i$$

The random variable C is such that $C \sim N(115, \sigma^2)$.

Given that $P(A - C < 0) = 0.2$ and that A and C are independent,

- (b) find the variance of C . (8)

TOTAL FOR PAPER: 75 MARKS

END