Write your name here

Surname	Other names

Pearson Edexcel
International Advanced Level

Centre Number | Candidate Number

Statistics S3

Advanced/Advanced Subsidiary

Thursday 22 May 2014 – Morning
Time: 1 hour 30 minutes

Paper Reference
WST03/01

You must have:
Mathematical Formulae and Statistical Tables (Blue)

Total Marks

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.

Turn over

P44515A

©2014 Pearson Education Ltd.

5/1/1/1/

P 4 4 5 1 5 A 0 1 2 4

PEARSON

Leave blank

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)

Leave blank

Question 1 continued

Q1

(Total 8 marks)

Leave blank

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)

P 4 4 5 1 5 A 0 4 2 4

Leave blank

Question 2 continued

Q2

(Total 7 marks)

Leave blank

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)

Leave blank

Question 3 continued

Q3

(Total 7 marks)

Leave blank

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)

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Q4

(Total 12 marks)

Leave blank

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)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Q5

(Total 12 marks)

Leave blank

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
Expected 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)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

(Total 14 marks)

Q6

Leave blank

7. The random variable X is defined as

$$X = 4Y - 3W$$

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

(a) Find $\mathrm{P}(X > 25)$

(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 $C \sim \mathrm{N}(115, \sigma^2)$

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

(b) find the variance of C.

(8)

P 4 4 5 1 5 A 0 2 0 2 4

Leave blank

Question 7 continued

Leave blank

Question 7 continued

Leave blank

Question 7 continued

Leave blank

Question 7 continued

Q7

(Total 15 marks)

TOTAL FOR PAPER: 75 MARKS

END

P 4 4 5 1 5 A 0 2 4 2 4