Write your name here

Surname	Other names

Pearson Edexcel
International
Advanced Level

Centre Number | | | | | |

Candidate Number | | | | |

Statistics S3

Advanced/Advanced Subsidiary

Wednesday 25 May 2016 – 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 ▶

P46669A

1/1/1/

P 4 6 6 6 9 A 0 1 2 8

PEARSON

Leave blank

1. The table below shows the distance travelled by car and the amount of commission earned by each of 8 salespersons in 2015

Salesperson	Distance travelled (in 1000's of km)	Commission earned (in $1000's)
A	20.4	17.7
B	22.2	24.1
C	29.9	20.3
D	37.8	28.3
E	25.5	34.9
F	30.2	29.3
G	35.3	23.6
H	16.5	26.8

(a) Find 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 positive correlation between the distance travelled by car and the amount of commission earned.

(4)

Leave blank

2. A researcher investigates the results of candidates who took their driving test at one of three driving test centres.

A random sample of 620 candidates gave the following results.

		Driving test centre			Total
		A	***B***	***C***	
Result	**Pass**	99	110	68	277
	Fail	108	116	119	343
Total		207	226	187	620

(a) Test, at the 5% level of significance, whether there is an association between the results of candidates' driving tests and the driving test centre. State your hypotheses and show your working clearly. You should state your expected frequencies correct to 2 decimal places.

(10)

The researcher decides to conduct a further investigation into the results of candidates' driving tests.

(b) State which driving test centre you would recommend for further investigation. Give a reason for your answer.

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

Question 2 continued

Leave blank

3. A company wants to survey its employees' attitudes to work. The company's workforce is located at three offices. The number of employees at each location is summarised in the table below.

Office location	Number of employees
Bristol	856
Dudley	429
Glasgow	1215

Each employee is located at only one office.

A personnel assistant plans to survey the first 50 employees who arrive for work at the Bristol office on a Monday morning.

(a) Give two reasons why this survey is likely to lead to a biased response.

(2)

A personnel manager has access to the company's information system that holds details of each employee including their place of work.

The manager decides to take a stratified sample of 150 employees.

(b) Describe how to choose employees for this stratified sample.

(3)

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

(1)

Leave blank

4. A random sample of 60 children and a random sample of 50 adults were taken and each person was given the same task to complete.

The table below summarises the times taken, t seconds, to complete the task.

	Mean, $\bar{t}$	**Standard deviation, s**	***n***
Children	61.2	5.9	60
Adults	59.1	5.2	50

(a) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not there is evidence that the mean time taken to complete the task by children is greater than the mean time taken by adults.

(6)

(b) Explain the relevance of the Central Limit Theorem to your calculation in part (a).

(1)

(c) State an assumption you have made to carry out the test in part (a).

(1)

Leave blank

Question 4 continued

Leave blank

5. Kylie used video technology to monitor the direction of flight, as a bearing, x degrees, for 450 honeybees that left her beehive during a particular morning. Kylie's results are summarised in the table below.

Direction of flight	Frequency
$0 \leqslant x < 72$	78
$72 \leqslant x < 140$	69
$140 \leqslant x < 190$	51
$190 \leqslant x < 260$	108
$260 \leqslant x < 360$	144

Kylie believes that a continuous uniform distribution over the interval [0, 360] is a suitable model for the direction of flight.

Stating your hypotheses clearly, use a 1% level of significance to test Kylie's belief. Show your working clearly.

(9)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

Question 5 continued

Leave blank

6. The random variable W is defined as

$$W = 3X - 4Y$$

where $X \sim N(21, 2^2)$ and $Y \sim N(8.5, \sigma^2)$ and X and Y are independent.

Given that $P(W < 44) = 0.9$

(a) find the value of σ, giving your answer to 2 decimal places.

(8)

The random variables A_1, A_2 and A_3 each have the same distribution as A, where $A \sim N(28, 5^2)$

The random variable B is defined as

$$B = 2X + \sum_{i=1}^{3} A_i$$

where X, A_1, A_2 and A_3 are independent.

(b) Find $P(B \leqslant 145 \mid B > 120)$

(7)

Leave blank

Question 6 continued

Leave blank

7. A random sample of 8 apples is taken from an orchard and the weight, in grams, of each apple is measured. The results are given below.

143 131 165 122 137 155 148 151

(a) Calculate unbiased estimates for the mean and the variance of the weights of apples.

(4)

A population has an unknown mean μ and an unknown variance σ^2

A random sample represented by $X_1, X_2, X_3, ..., X_8$ is taken from this population.

(b) Explain why $\sum_{i=1}^{8}(X_i - \mu)^2$ is not a statistic.

(1)

Given that $E(S^2) = \sigma^2$, where S^2 is an unbiased estimator of σ^2 and the statistic

$$Y = \frac{1}{8}\left(\sum_{i=1}^{8} X_i^2 - 8\bar{X}^2\right)$$

(c) find $E(Y)$ in terms of σ^2

(2)

(d) Hence find the bias, in terms of σ^2, when Y is used as an estimator of σ^2

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

Question 7 continued

Leave blank

8. A six-sided die is labelled with the numbers 1, 2, 3, 4, 5 and 6

A group of 50 students want to test whether or not the die is fair for the number six.

The 50 students each roll the die 30 times and record the number of sixes they each obtain.

Given that $\bar{X}$ denotes the mean number of sixes obtained by the 50 students, and using

$$H_0 : p = \frac{1}{6} \text{ and } H_1 : p \neq \frac{1}{6}$$

where p is the probability of rolling a 6,

(a) use the Central Limit Theorem to find an approximate distribution for $\bar{X}$, if H_0 is true.

(3)

(b) Hence find, in terms of $\bar{X}$, the critical region for this test. Use a 5% level of significance.

(4)

Leave blank

Question 8 continued