

**UNIVERSITY OF LONDON**

**279 004a ZB**

**990 004a ZB**

**996 D04a ZB**

**BSc degrees and Diplomas for Graduates in Economics, Management, Finance and the Social Sciences, the Diploma in Economics and Access Route for Students in the External Programme**

**Statistics 1 (half unit)**

Wednesday, 12 May 2010 : 10.00am to 12.00pm

Candidates should answer **THREE** of the following **FOUR** questions: **QUESTION 1** of Section A (50 marks) and **TWO** questions from Section B (25 marks each). **Candidates are strongly advised to divide their time accordingly.**

Extracts from statistical tables are given after the final question on this paper

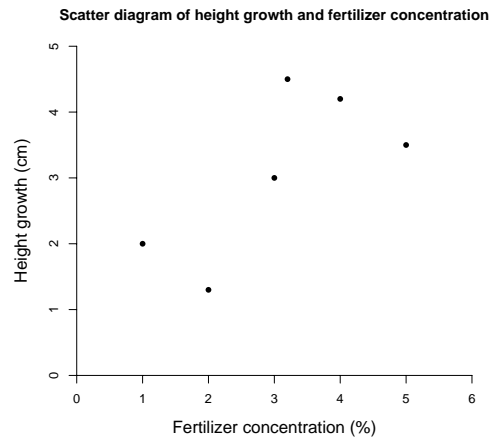
Graph paper is provided at the end of this question paper. If used, it must be detached and fastened securely inside the answer book.

A calculator may be used when answering questions on this paper and it must comply in all respects with the specification given with your Admission Notice. The make and type of machine must be clearly stated on the front cover of the answer book.

## SECTION A

Answer **all** parts of Question 1 (50 marks in total).

1. (a) The following scatter diagram shows the relationship between the height growth ( $y$ , in centimetres) of a plant and the concentration of fertilizer it receives ( $x$ , in %).



- What do you think the correlation coefficient is likely to be?
- Would you feel justified in estimating a regression line? If so, would its slope be positive or negative?
- Comment on the diagram. What does it tell you about the effect of the fertilizer?
- You want to study the effect of using a higher concentration of the fertilizer than has been tested in the study. Could you extrapolate from the data? Explain.

**(6 marks)**

- (b) Explain the difference between sampling error and sampling bias. Explain briefly which is taking place in the following situations:

- The population under study consists of all married couples in a country, but you take a sample of married couples from a list of all couples which excludes those who have moved in the last year.
- The population under study consists of all addresses in a certain city and you take a random sample from a list of addresses in the city using a table of random numbers.

**(5 marks)**

- (c) Two random samples are obtained from a population with a normal distribution. The summary statistics for the two samples are as follows:

	Sample size	Sample mean	Sample variance
$x$ data	18	4.8	0.81
$y$ data	23	5.7	1.21

(question continues on next page)

Compute the mean and the standard deviation of the combined dataset.

**(6 marks)**

- (d) We take a random sample from a population to measure a quantity with variance 400 and estimate the population mean using the sample mean. We wish to be accurate to within four units with 98% confidence. What is the smallest sample size that is necessary?

**(4 marks)**

- (e) i. Two fair dice are thrown and you are told that the sum of the upturned dice is at least nine. What is the probability that neither of the upturned faces is a four?
- ii. In a small village with an adult population of 20 only four adults watch television sports events on a pay-per-view basis. A researcher asks two adults selected at random, one after the other, whether they watch sport this way. What is the probability that the second person asked does?
- iii. Rachel is going to a party. There is a 40 percent chance Bill will go. If Bill does not go, there is an 80 percent chance she will enjoy herself. If Bill does go, there is only a 10 percent chance she will enjoy herself. What is the probability that Rachel will enjoy the party?
- iv. At the same party, suppose you know Rachel did enjoy herself. What is the probability that Bill was not present?

**(8 marks)**

- (f) A UK government agency carries out a large-scale random survey of public attitudes towards the recession. 132 of the 600 workers surveyed indicated they were worried about losing their job. Newspaper reports claim 25% of workers fear losing their job. Is such a high percentage claim justified? State and carry out an appropriate hypothesis test at two levels and explain your results.

**(8 marks)**

- (g) With  $x_1 = 2$ ,  $x_2 = 3$ ,  $x_3 = 4$ ,  $x_4 = 2$  and  $x_5 = 2$ , find:

i.  $\sum_{i=1}^{i=3} (x_i - 2)$       ii.  $\sum_{i=4}^{i=5} 3x_i$       iii.  $\sum_{i=2}^{i=4} x_i^2$

**(6 marks)**

- (h) A variable has a unimodal distribution with a mean of 50 and a median of 30. Is the distribution skewed to the left, to the right, symmetric, or is there not enough information to decide? Explain your answer briefly and give a rough sketch of the distribution.

**(3 marks)**

- (i) Which does not belong in the list in each of the following? Explain briefly. *Note no marks will be awarded without an explanation.*

- i. Standard deviation, variance, range.
- ii. Median, range, interquartile range, standard deviation.

**(4 marks)**

## SECTION B

Answer **two** questions from this section (25 marks each).

2. (a) A study is conducted of shoppers at department stores and at supermarkets to compare the proportions who regularly buy 'own brand' products in each kind of shop. The results of a random sample of shoppers are given in the following table:

	Sample size	Number buying own brand
Department stores	1,200	450
Supermarkets	1,400	600

- Carry out a hypothesis test as to whether the proportions of own brand purchasers are the same for each type of shop. Test at two appropriate levels and comment on your results.
- Calculate a 99% confidence interval for the difference between the two true population proportions of own brand purchasers.

**(13 marks)**

- (b) The merchandising group who commissioned the survey are not happy with the way the survey was carried out. They are not sure the survey was conducted in a random manner.

- Think of three distinct questions you would want to ask in order to check that they are correct in this supposition.
- You are invited to carry out a survey by telephoning loyalty card holders for the shops concerned. Explain how you might do this. Do you regard this as a random sample? Explain.

**(12 marks)**

3. (a) A leading global retailer is studying the relationship between the area available for sales in its shops ( $x$  values in metres squared) and the weekly amount of sales in each shop ( $y$  values in thousands of pounds). The figures for a random sample of 14 outlets are recorded, and are shown in the following table:

Area ( $x$ ) in m <sup>2</sup>	Sales ( $y$ ) in £1000s	Area ( $x$ ) in m <sup>2</sup>	Sales ( $y$ ) in £1000s
1.9	3.6	1.1	3.3
1.3	4.1	3.7	6.1
2.8	3.7	1.4	4.8
5.8	9.6	5.1	10.1
2.9	3.9	4.5	9.9
1.1	3.6	6.5	10.0
1.2	1.7	3.9	5.1

The summary statistics for these data are:

$$\sum x_i = 43.2 \quad \sum x_i^2 = 178.42 \quad \sum y_i = 79.5 \quad \sum y_i^2 = 563.85 \quad \sum x_i y_i = 309.39$$

- Draw and label the scatter diagram of these data carefully on the graph paper provided.
- What does your diagram show? Make at least two comments.
- Calculate and write down the least squares regression line.
- An architect has designed a series of retail areas of exactly 6 metres squared. What would you expect the sales to be in such an area?

**(12 marks)**

- (b) A researcher into the use of computational aids conducts a survey in which 195 students, as a homework assignment, compute the standard deviation of a set of data. The students are asked what computational aids, if any, they used. The researcher is particularly interested as to whether male and female students differ in this respect. The following table shows the results of this survey:

Gender	Method of computation			
	No aids	Basic calculator	Statistical function on a calculator	Computer
Male	10	45	30	10
Female	12	30	52	6

- Carry out an overall test for association between gender and method of calculation at two levels. Give the null and alternative hypotheses and comment on your results.
- The researcher is interested in whether there are any gender differences in preferred method of computation. Discuss any potential gender differences which appeared in the test for association.

**(13 marks)**

4. (a) The monthly losses ( $x$ , in \$1000s) of a random sample of thirty managed funds at the height of the 2008 ‘credit crunch’ are shown below:

53	23	30	72	55	49
65	36	40	51	48	43
35	39	42	44	56	31
50	54	64	57	53	42
52	63	49	33	47	44

- Draw and label an appropriate histogram using exactly five classes on the graph paper provided.
- Calculate the mean, median and range.
- Give the modal group.
- Comment on the data given the information you have calculated.

**(12 marks)**

- (b) A pet food supplier is studying the difference between two of its stores. It is particularly interested in the time it takes before customers receive the products they have ordered. Using standard notation, the data of delivery times from the two stores is as follows:

	Store A	Store B
$\bar{x}$	34.3 days	38.6 days
$s$	2.4 days	3.1 days
$n$	41	31

- Use an appropriate hypothesis test to see if there is a difference in the average delivery times for the two stores. Test at two appropriate levels and comment on your findings.
- Give the assumptions you have made.
- Give the 98% confidence interval for the mean delivery time for Store A.

**(13 marks)**

END OF PAPER