UNIVERSITY OF LONDON

279 004a ZA
990 004a ZA
996 D04a ZA

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)

Tuesday, 16 May 2006 : 10.00am to 12.00noon

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.**

Graph paper is provided. If used, it must be fastened securely inside the answer book.

New Cambridge Statistical Tables (second edition) are provided.

A hand held calculator may be used when answering questions on this paper but it must not be pre-programmed or able to display graphics, text or algebraic equations. The make and type of machine must be stated clearly on the front cover of the answer book.

PLEASE TURN OVER

SECTION A

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

1. (a) Define each of the following briefly:

      i.   quota sampling.
      ii.   cluster sampling.
      iii. an alternative hypothesis.   **(6 marks)**

   (b) Which of the following is the odd one out?
           Mean, median, mode, range.
     Explain briefly.   **(2 marks)**

   (c) The following tables show the scores of two groups of students in a test question.

      i.

Mark	0	1	2	3	4
Frequency	10	5	8	x	3

           Determine the value of x if the median of these marks is 1.5

      ii.

Mark	0	1	2	3	4
Frequency	10	5	8	y	3

           Determine the value of y if the mean of these marks is 1.5
                                        (4 marks)

   (d) i.   Two fair dice are thrown and you are told that the sum of their upturned faces is equal to 7. What is the probability that neither face is equal to 6?

      ii.   In a class of 40 pupils, 16 have blue eyes, 12 have blonde hair, and 8 have blue eyes and blonde hair. Find each of the following probabilities:
   •   the pupil is blue-eyed or blonde or both.
           •   the pupil is neither blue-eyed nor blonde.
           •   the pupil is blue-eyed, but not blonde.   **(8 marks)**

          (question continues on next page)

(e)   A scatter plot for two variables (x and y) is shown below.

        i.   Do you consider these variables to be independent? Explain.
        ii.   State an approximate value for their correlation coefficient.
        iii.   If the data set were split into two at the centre of the x-range,
                - State an approximate value for the correlation of the smaller set.
                - State an approximate value for the correlation of the larger set.

                                                        (7 marks)
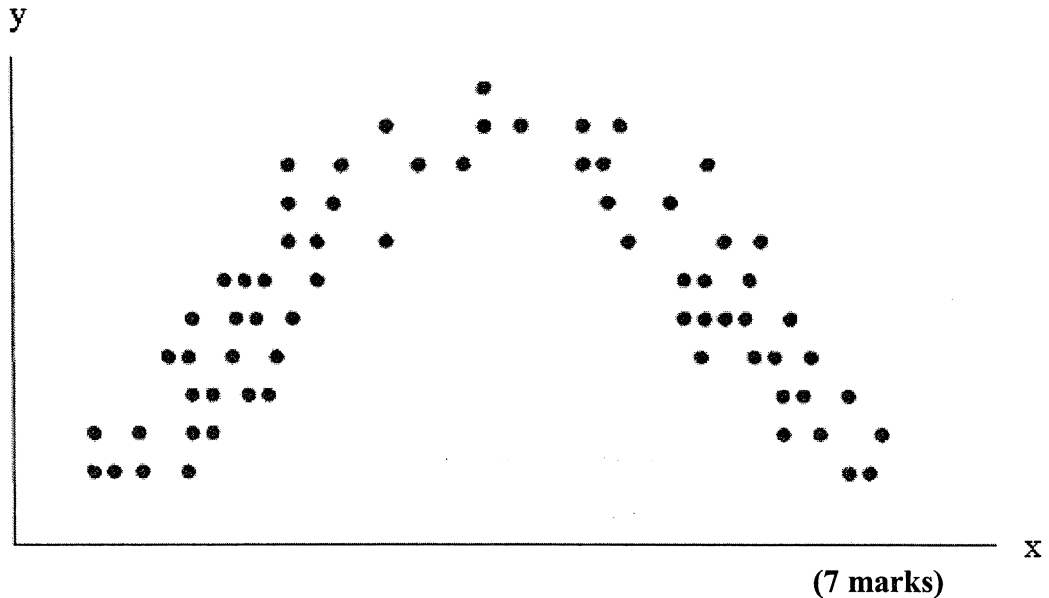
(f)   In the past, 18% of shoppers have bought a particular brand of breakfast
        cereal. After an advertising campaign for this brand, a random sample of 220
        shoppers is taken and 55 of these are found to have bought this brand of
        cereal. Carry out the appropriate significance test at the 1% level to test
        whether the advertising campaign has been successful. Comment briefly.
                                                        (9 marks)

                        (question continues on next page)

(g) $x_1 = 2,\ x_2 = 4,\ x_3 = 3,\ x_4 = 1$

Find

i. $\displaystyle\sum_{i=3}^{i=4} x_i^{\,3}$

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

iii. $\displaystyle\sum_{i=1}^{i=3} 4x_i$

(6 marks)

(h)   State whether the following are possible or not. Give a brief explanation. *(Note that no marks will be awarded for a simple possible/not possible reply)*

    i.   Only Angela or Esther may get the end of year prize and no one else. The prize must be awarded. The probability Angela gets it is ·4. That for Esther is ·5.

    ii.   The correlation coefficient for the relationship between the number of grams of pasta eaten daily and weekly weight gain in kilos is +2·5.

    iii.   The variance of the results is plus 16.

    iv.   The larger a sample, the larger the variation in its sample mean.

(8 marks)

SECTION B

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

2.   (a)   The following data show the lengths (inches) of fish caught in one day in a river.

15.2	15.6	14.6	10.5	13.7	10.6	13.9	14.7	13.9	10.6
10.3	10.6	12.1	12.8	14.9	15.2	13.8	14.5	15.0	11.7
10.9	18.4	14.1	14.9	10.2	13.9	11.6	13.9	12.5	11.1
15.1	14.9	15.0	12.6	10.9	13.7	12.9	12.1	14.0	13.5
15.3	13.8	14.9	10.6	15.1	11.6	11.0	14.8	13.8	10.3

        i.   Construct a histogram for these data (Do not use more than seven categories along the x axis). Label it carefully.

        ii.   Use this to determine the median and upper and lower data quartiles.

        iii.   Briefly describe the figures shown.   **(12 marks)**

(b)   A machine filling cans supplies, to each can, a volume X of fruit and a volume Y of juice. It is known that X has a normal distribution with mean 260 and a standard deviation of 17 whereas Y has a normal distribution with mean 150 and a standard deviation of 10. These variables may be regarded as independent.

        i.   Calculate the probability that the volume of juice supplied is greater than 290 units.

        ii.   Find the probability that the volume of fruit supplied is more than twice the amount of juice.

        iii.   If the capacity of a particular can is 400, what is the probability of the can being under-filled?   **(13 marks)**

3. (a) The following data set shows the times (y) in seconds taken by international athletes to run 100 metres together with the corresponding wind speeds (x) at the time of running. A positive wind speed indicates the wind is in the direction of running and therefore considered to be helpful whereas a negative wind speed indicates the wind is against the runner.

Wind Speed (x)	-2.94	-1.63	-0.74	-0.38	-0.17
Time for 100 metres	10.51	10.44	10.39	10.22	10.44

Wind Speed (x)	0.21	0.60	1.08	2.25	2.81
Time for 100 metres	10.18	10.30	10.15	10.20	10.21

N.B. summary statistics for these data are,

Sum of x data = 1.09   Sum of y data = 103.04

Sum of squares of x data = 26.5505   Sum of squares of y data = 1061.8770
Sum of xy products = 9.6563

  i.   Draw and label the scatter diagram of these data as described here. Label it carefully.
  ii. Comment on your diagram. How strong would you expect the connection to be between the x and y? (Do not calculate the correlation coefficient to answer this).
  iii. Calculate the regression equation and draw it.
  iv. Comment on your results.   **(15 marks)**

(b) You run a market research company and have been asked to give a tender showing how you could carry out a survey of a large international company's employees to find out about their holiday preferences in terms of their timing and length .You have been told that you should make a postal survey as the company will provide the addresses you need and are prepared to cover postage costs. The company needs reasonably accurate figures as they will be using them for a large staff deployment exercise worldwide.

Outline your design making it clear whether it is random or not and giving appropriate design factors.   **(10 marks)**

PLEASE TURN OVER

4. (a) The following table shows the classification by attitude and social class of a random group of 698 residents to a council proposal on traffic calming.

| | | Social Class | | |
		Lower	Middle	Upper
Attitude to proposal	Approved	34	45	61
	Indifferent	55	30	203
	Disapproved	28	22	220

    i.   Carry out the appropriate formal test of association at two levels of significance.

    ii.   Comment on your results - are there any particular figures you find interesting here? **(12 marks)**

(b) A simple random sample of 350 wives were asked whether they ever did 'Do it yourself' jobs in the home. Of these, 90 replied that they did. Calculate the 90% confidence interval for the proportion of housewives who do 'Do it yourself' jobs. **(4 marks)**

(c) A study was carried out to see if an extended contact with their newborn babies in the first minute of life caused the mother to have a greater attachment to her baby. An attachment index was calculated for each mother based on the mother's attitude to the baby in the week after birth. For this study, two groups (control and treatment) of mothers were used. The *control group* of 62 mothers had the standard amount of contact with their babies after birth. The *treatment group* of 73 mothers were given extended contact with their babies in the first minute after birth. Summary statistics are given below. (The higher the attachment index, the closer the mother's attachment).

	Control Group	Treatment Group
Number in group	62	73
Sample mean	6.29	8.53
Sample Standard Deviation	2.07	1.95

    i.   Carry out a test, giving the null and alternative hypotheses, to see whether the extended contact has an effect.

    ii.   Have you any reservations about these results in terms of the long term effect on family relationships? **(9 marks)**

END OF PAPER