

PRELIMINARY EXAM

PROGRAMME(S) : University of London Degree and Diploma Programmes
(Lead College: London School of Economics & Political Science)

SUBJECT : **04a STATISTICS 1**

DATE : 27 February 2010

DURATION : 2 hours

INSTRUCTIONS :-

**DO NOT TURN OVER THIS QUESTION PAPER UNTIL YOU
ARE TOLD TO DO SO.**

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).

Graph paper is provided. When used, it should be fastened securely **inside** the answer book.

Statistical tables are attached.

A handheld 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 the machine must be clearly stated on the front cover of the answer book.

Candidates are strongly advised to divide their time accordingly.

Total number of pages: 6, including this page.

SECTION A

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

1. (a) The manufacturer of a new family car claims that petrol consumption is 20.0 km/litre. The mean petrol consumption for a random sample of 25 cars was found to be 18.8 km/litre with a sample standard deviation of 5.25 km/litre.
- Construct a 95% confidence interval for the true mean consumption.
 - Describe your result verbally as you would to a non-statistician.
 - Is the manufacturer's claim supported?
- (8 marks)**
- (b) State whether the following are true or false and give brief explanations.
(Note that no marks will be given for a simple True/False answer without a satisfactory explanation.)
- The area under the curve between $z = -1$ and $z = +1$ is 0.9475.
 - When the correlation coefficient between two variables is negative, it could be that the slope of the least squares line is positive.
 - If the value of a chi-squared test statistic is 12.00, then there is no relationship between the variables in the contingency table.
 - The median value of a set of data is always greater than that of the third quartile.
- (8 marks)**
- (c) A spot check is carried out in the Accident and Emergency department in a local hospital. The waiting time, in minutes, for routine blood tests, for ten randomly selected patients is given below:
- 5 25 62 25 15 28 15 45 19 74
- Calculate the mean, median, mode and range. Explain why the three measures of location differ.
- (6 marks)**
- (d) Two key components in the office multi-coffee machine are the percolator and the steamer. During a normal working day, the probability that the percolator breaks down is 0.92, while the probability that the steamer breaks down is 0.67. Assuming that the percolator and steamer function independently, calculate the probability that the machine is not working due to:
- the percolator only
 - the steamer only
 - either the percolator or the steamer
 - at least one of them.
- (6 marks)**

- (e) A total of 2,000 students attend evening classes in a Singapore university. Students are classified according to the course they attend: professional, continuous professional development (CPD), and leisure. The table below gives the figures:

| Professional | | CPD | | Leisure | |
|--------------|--------|------|--------|---------|--------|
| Male | Female | Male | Female | Male | Female |
| 450 | 296 | 591 | 261 | 60 | 342 |

Calculate the number that should be selected from each group for a stratified random sample of 250. Describe how you would select the sample.

(8 marks)

- (f) A survey of household income reveals that expenditure on holidays drops when family income drops. The least squares line and correlation coefficient for a random sample of 18 families were calculated as:

$$y = 5 + 0.25x \quad \text{and} \quad r = +0.65$$

respectively, where both x and y are given in pounds sterling.

- Use the graph paper provided to plot the line for values of x between 10 and 100. Label your diagram carefully.
- Describe the line and intercept verbally.
- Calculate the likely drop in holiday expenditure if family income drops by £3,500.
- What additional information might you require to use the above equation to predict holiday expenditure?

(10 marks)

- (g) Two balls are selected without replacement from a box with 30 balls: 18 white, 8 yellow and 4 black. Calculate the probability that:

- Both are the same colour.
- Neither is white.

(4 marks)

SECTION B

Attempt **two** of the three questions in Section B (25 marks each).

2. (a) A construction company has three suppliers for four types of under-floor insulation material. The number of units of each type of insulation material supplied by the three suppliers is summarised in the following contingency table.

| | A | B | C |
|---------|----|----|----|
| Poly E | 22 | 17 | 20 |
| Poly S | 35 | 29 | 42 |
| M board | 13 | 14 | 33 |
| Seal F | 10 | 20 | 45 |

A chi-squared test is carried out to test whether there is an association between suppliers and the type of insulation material supplied.

- State the null and alternative hypotheses.
- The value of the test statistic calculated from the data was 16.97. Carry out the test yourself at two sensible levels. Be careful about the order in which you do this!
- State your conclusions carefully.

(12 marks)

- (b) A randomly selected group of 35 A&E patients was asked “how many hours have you been waiting for medical attention?” The data are given below.

| | | | | |
|-----|-----|-----|-----|-----|
| 6.2 | 2.5 | 6.6 | 2.6 | 1.8 |
| 1.5 | 1.4 | 5.3 | 1.0 | 2.4 |
| 4.4 | 3.2 | 2.7 | 5.1 | 5.2 |
| 7.6 | 4.2 | 2.1 | 5.4 | 2.4 |
| 6.0 | 2.1 | 4.6 | 3.4 | 1.8 |
| 4.4 | 2.9 | 1.6 | 2.4 | 2.6 |
| 6.0 | 3.0 | 7.1 | 3.7 | 3.8 |

- Sort the data by a stem-and-leaf plot and label it carefully.
- Calculate the mean, median, quartiles and quartile deviation.
- Write a brief report, describing the waiting time for this sample of patients.

(13 marks)

3. (a) Explain why each of the following is important in statistics:

least squares; correlation coefficient.

Construction companies contend that the total amount paid in bonuses to executives is related to total sales. To investigate the relationship between executive bonus as a function of sales, data were collected from 8 companies, shown below:

| | | | | | | | | |
|------------------|-----|-----|-----|-----|------|------|----|-----|
| y: Bonus (€000s) | 45 | 32 | 50 | 68 | 115 | 126 | 25 | 42 |
| x: Sales (€m) | 485 | 435 | 840 | 565 | 1244 | 1563 | 84 | 104 |

- Plot and label the scatter diagram for bonus against sales using the graph paper provided. Comment on the nature of the relationship.
- Calculate the least squares line for bonus against sales. Describe the results verbally.
- Calculate the total amount paid in bonuses when sales are:

a. €75m b. €1,000m.

How reliable are these results?

(13 marks)

- (b) Anecdotal evidence suggests that the hourly rate of pay is higher for skilled operatives compared to unskilled. The results from a survey of 160 operatives in the private sector are given below:

| | Sample size | Sample mean | Sample standard deviation |
|------------|-------------|-------------|---------------------------|
| Operatives | n | \bar{x} | s |
| Skilled | 58 | 20.54 | 0.36 |
| Unskilled | 92 | 17.46 | 0.94 |

Calculate the 95% confidence interval for the difference in the mean hourly rate for skilled and unskilled operatives.

Write a brief report on this confidence interval. Does it support (i.) anecdotal evidence, and (ii.) a claim that skilled operatives earn €3 per hour more than unskilled operatives?

(12 marks)

4. (a) Health and Safety regulations require manufacturers of lifts to state the maximum number of people allowed. If the average weight of an adult is 71 kg and the standard deviation is 2.25 kg, calculate:
- the percentage of adults whose weight exceeds 75 kg
 - the percentage of adults who weigh between 63 kg and 75 kg
 - the weight above which 99% of all adults weigh.
 - If a random sample of 36 adults is selected, calculate the probability that the mean weight exceeds 72 kg.
- (10 marks)**
- (b) i. List the advantages and disadvantages of using each of the following when surveying individuals:
- Interviewer-administered questionnaires face-to-face
 - Interviewer-administered questionnaires using the telephone
 - Questionnaires delivered by post for respondents to fill in themselves
 - Questionnaires delivered by email for respondents to fill in themselves.
- ii. Taking your list into account, explain how you might administer each of the following:
- A survey of teachers on behalf of their professional association about pay and conditions at work.
 - A survey of teenagers about their participation in leisure activities.
 - A survey on behalf of a government inquiry about the national diet and households' food consumption.

Be careful to use at least three of the points you have made in (i.) to justify your decision.

(15 marks)

[END OF PAPER]