

## Section A

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

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

- i. Interviewer bias.
- ii. Cluster sampling.
- iii. Quota sampling.

(6 marks)

(b) Which of the following is the odd one out? Explain briefly.

Variance, mean, standard deviation, range.

(2 marks)

(c) In an examination, the scores of students who attend type A schools are normally distributed about a mean of 55 with a standard deviation of 6. The scores of students who attend type B schools are normally distributed about a mean of 60 with a standard deviation of 5. Which type of school would have a higher proportion of students with marks above 70? Explain your answer.

(5 marks)

(d) What does it mean for two variables to be uncorrelated? How is it possible for two variables to be strongly related, but still uncorrelated?

(3 marks)

(e) The owner of a fish market finds that the mean weight for a catfish is 3.2 pounds with a standard deviation of 0.8 pounds. Assume that the weights of the catfish are normally distributed.

- i. What is the probability that the weight of a randomly selected catfish is greater than 4.8 pounds?
- ii. Above what weight (in pounds) do 89.8% of the weights occur?
- iii. You buy a sample of 25 catfish. What is the probability that the mean is less than 3 pounds?

(10 marks)

(f) Let  $x_1 = 5$ ,  $x_2 = 1$ ,  $x_3 = 3$  and  $x_4 = 1$ . Find:

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

(6 marks)

(g) Three fair coins are thrown (and throws are independent of each other).

- i. Find the probability that exactly one is a head.
- ii. You are told that at least one is a head. What is the probability that all are heads in this case?

(4 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. The correlation coefficient for the relationship between hours of revision and examination mark is  $-2.3$ .
  - ii. ~~Quota sampling allows us to quantify bias and standard errors.~~
  - iii. It is possible to have a chi-squared value of  $-3$ .
  - iv. If the probability that it will rain tomorrow is  $1/5$  and the probability that you will not wear a raincoat is given, then the probability that it rains and you find you have no raincoat is  $1/4$ .

**(8 marks)**

- (i) A company has two machines which produce cupboards. 75% are produced by the new machine and the remaining 25% by the older machine. In addition, the new machine produces 8% defective cupboards. The old machine produces 23% defective cupboards.
- i. What is the probability that a randomly chosen cupboard produced by the company is defective?
  - ii. Given that a randomly chosen cupboard has been chosen and found to be defective, what is the probability it was produced by the new machine?

**(6 marks)**

**Section B**

Answer **TWO** questions from Section B (25 marks each). [Goodness of Fit - Done by Gurdeep in Tutorial](#)

2. (a) Over a number of years, the demand for daily newspapers in a locality had been 20% for the Sun, 15% for the Star, 30% for the Times, 20% for the Examiner and 15% for the Independent. To determine whether demand has changed, the manager randomly records the papers purchased by 300 readers. The results are given below:

Paper	Sun	Star	Times	Examiner	Independent
No. of readers	65	40	75	90	30

Use the  $\chi^2$  test to determine whether demand for these papers has changed.

**(10 marks)**

- (b) A survey is carried out to determine whether or not there is any association between age group and being in favour of working from home on two days per week. The raw results of the survey are as follows:

Age group	18–24	25–35	36–50	Over 50
In favour	20	62	40	40
Against	32	50	46	80

- Test the claim that there is no association between age group and being in favour at the 5% significance level.
- If the null hypothesis in (i.) is rejected, identify the nature of the association.

**(15 marks)**

3. (a) Fill in the missing entries (??) in the following table:

$H_0$	??
$H_1$	$\mu > 50$
The decision rule (for $\alpha = 0.05$ ) is	??
Sample data:	$n = 36, \bar{x} = 56.5, s = 4.5$
The test statistic value is	??
Conclusion	??

**(8 marks)**

(b) Health advisory services state that 50% of heart attack patients die before reaching hospital – the chances of surviving a heart attack are dramatically increased by receiving medical help as quickly as possible. A survey of 240 heart attack victims discovered that men sought help more urgently than women.

	Number	Sample statistics of those reaching hospital after the initial attack (in hours)
Men	134	$\bar{x} = 3.40, s = 1.42$
Women	106	$\bar{x} = 8.24, s = 4.24$

- i. Calculate a 95% confidence interval for the difference between the mean times to reach hospital for men and women.
- ii. Is it possible to infer that there is a difference between the mean times taken to reach hospital from the data given above? Explain your answer.

**(12 marks)**

(c) According to the Quarterly National Household Survey, the mean weekly family income is estimated to be £1,250 with a standard deviation of £120. Assuming incomes are normally distributed, calculate the percentage of households whose weekly income is between £750 and £1,500.

**(5 marks)**

## A. Sample examination paper

4. (a) The following data give the price in euros of a ten-minute call at 11:00 on a weekday for a local call. The prices refer to August each year. Normal tariffs without special rates are used.

	1997	1998	1999	2000	2001	2002	2003	2004
Spain	0.20	0.32	0.32	0.28	0.28	0.28	0.28	0.28
Ireland	0.58	0.58	0.49	0.51	0.51	0.51	0.51	0.49

- Plot, on the same diagram, approximate lines for the price of a local call in Spain and Ireland from 1997 to 2004. Give a brief verbal comparison of the trends in prices in the two countries.
- Fit a least squares line for prices in Ireland over this period, using  $x = 0$  for 1997,  $x = 1$  for 1998, etc.

N.B. summary statistics for these data are:

$$\sum_i x_i = 28, \quad \sum_i y_i = 4.18, \quad \sum_i x_i^2 = 140, \quad \sum_i y_i^2 = 2.1934, \quad \sum_i x_i y_i = 14.17.$$

- Calculate the sample correlation coefficient.
- How appropriate do you think it is to fit a linear trend to these data?

**(15 marks)**

- (b) You run a market research company and have been asked to research the likely demand for extra sports facilities in your city following a decision by the government to promote exercise for general health.

Outline your survey design, making it clear whether it is random or not, and giving appropriate design factors. You have been given generous funding and are expected to look at all aspects of the question and all groups of the population.

**(10 marks)**

[END OF PAPER]