

Part B

Instructions to candidates

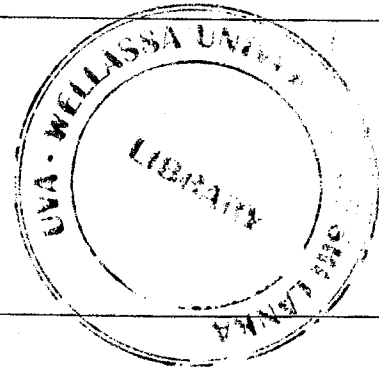
Duration: Two(02) hour and 15 minutes

Number of questions: Four(04) essay questions

Mark allocation: 70

Scientific Calculators are allowed.

Answer all questions



1.
  - a. Present arguments in favor of conducting interviews over the telephone instead of face to face. (5 mark)
  - b. Define the correct sequence in the stages of sampling. (5 mark)
  - c. The marketing director of a chain of supermarkets would like to set up an interview survey of a sample of customers. The supermarkets are located in several different regions, and are classified into three types according to size. A sample of supermarkets is to be taken, and samples of customers will be drawn from these supermarkets.
    - i. Suggest how cluster sampling could be used to choose the sample of supermarkets. Discuss any potential advantages and disadvantages of this method for the proposed survey. (5 mark)
    - ii. Suggest how stratified sampling could be used to choose the sample of supermarkets. Discuss any potential advantages and disadvantages of this method for the proposed survey. (5 mark)

2.
  - a. The table shows the average price, in dollar, of a litre of petrol in USA on 31 December each year from 2003 to 2011.

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011
Price (p)	76.02	84.17	87.35	85.47	102.10	92.79	108.70	119.20	133.40

- Draw a suitable graph for the data and briefly describe the graph. (5 mark)
- b. For these situations, state which measure of central tendency should be used. (4 mark)
    - a) The most typical case is desired
    - b) The distribution is open-ended
    - c) There is an extreme value in the set
    - d) The data are categorical

- c. A seed merchant took delivery of a sack of beans. Looking into the sack, he thought the beans appeared much more variable in size than his usual stock. He took a sample of 15 beans from the sack and measured their lengths. His results were as follows, in cm.

2.4, 2.7, 3.0, 2.5, 3.2, 2.8, 2.8, 2.3, 3.0, 2.8, 2.8, 2.4, 2.9, 2.8, 3.1

Calculate the Range, Mean and Standard deviation for these 15 beans. (6 mark)

3. Suppose that the following two data sets I and II correspond to percentage marks obtained by two groups of students of size twenty each. They are arranged in ascending order of magnitude.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
I	11	11	13	14	18	21	23	28	30	33	37	39	41	43	54	57	79	81	82	87
II	10	22	27	28	28	29	29	29	31	35	38	38	50	52	54	80	82	90	90	95

- a. Classify each of the above data sets according to the following table. (4 mark)

Class Boundaries	Frequency I	Frequency II
E: (0.05- 29.5)		
D: (29.5- 39.5)		
C: (39.5- 54.5)		
B: (54.5- 69.5)		
A: (69.5- 99.5)		

- b. Hence calculate the means using  
 (i) raw data (3 mark)  
 (ii) classified data for each of the data sets using part a. (4 mark)
- c. Compare the accuracies and comment critically on the assumption about data. (3 mark)
- d. Find the cumulative frequencies for each group of students and draw their graphs on the same graph paper. Comment on the differences. (6 mark)

4. An experiment is being conducted into the physical fitness of older people. A physical fitness test is given to each of 10 subjects (a higher score represents greater fitness). Each subject's fitness score for the test is recorded as value  $y$  and age as value  $x$  (ages are between 40 and 80 years). The following totals are given.

$$n = 10 \quad \sum_{i=1}^{10} x_i = 581 \quad \sum_{i=1}^{10} y_i = 607 \quad \sum_{i=1}^{10} x_i y_i = 33426$$

$$\sum_{i=1}^{10} x_i^2 = 37193 \quad \sum_{i=1}^{10} y_i^2 = 38795$$

Calculate:

- a. mean of  $x$  and mean of  $y$ . (2 mark)
- b. the product-moment correlation coefficient. (5 mark)

- c. Comment on both the sign and magnitude of the coefficient. (2 mark)
- d. By fitting the appropriate straight line, estimate the average fitness score of a person who is aged 45. (6 mark)

### Useful Formula

$$1. r = 1 - 6 \sum \frac{d_i^2}{n(n^2-1)}$$

$$2. r = \frac{(\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y})}{\sqrt{(\sum_{i=1}^n X_i^2 - n\bar{X}^2)(\sum_{i=1}^n Y_i^2 - n\bar{Y}^2)}}$$

$$3. \beta_1 = \frac{(\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y})}{(\sum_{i=1}^n X_i^2 - n\bar{X}^2)}$$

$$4. \beta_0 = \bar{Y} - \beta_1 \bar{X}$$

