සියලු ම හිමිකම් ඇවිරිණි /முழுப் பதிப்புரிமையுடையது /All Rights Reserved]

නව නිර්දේශය/புதிய பாடத்திட்டம்/New Syllabus

NEW

ශ්‍රී ලංකා විභාග දෙපාර්තමේන්තුව
இலங்கைப் பரீட்சைத் திணைக்களம்
Department of Examinations, Sri Lanka

අධ්‍යයන පොදු සහතික පත්‍ර (උසස් පෙළ) විභාගය, 2019 අගෝස්තු
கல்விப் பொதுத் தராதரப் பத்திர (உயர் தர)ப் பரீட்சை, 2019 ஓகஸ்ற்
General Certificate of Education (Adv. Level) Examination, August 2019

15.08.2019 / 1300 – 1500

ව්‍යාපාර සංඛ්‍යානය	I
வணிகப் புள்ளிவிவரவியல்	I
Business Statistics	I

31 E I

පැය දෙකයි
இரண்டு மணித்தியாலம்
Two hours

Instructions:

- ✲ *Answer **all** questions.*
- ✲ *Write your **Index Number** in the space provided in the answer sheet.*
- ✲ *Statistical tables will be provided. Use of calculator is **not allowed**.*
- ✲ *Instructions are given on the back of the answer sheet. Follow those carefully.*
- ✲ *In each of the questions **1** to **50**, pick one of the alternatives from (1), (2), (3), (4), (5) which is **correct** or **most appropriate** and **mark your response on the answer sheet with a cross** (×) on the number of the correct option in accordance with the instructions given at the back of the answer sheet.*

1. Which of the following statements is true?

(1) The data collected through newspapers and periodicals are primary data.
(2) Making inferences about the whole population by studying only a random sample selected is a misuse of statistics.
(3) Sampling errors cannot be reduced by increasing the sample size.
(4) Statistics does not study an individual value.
(5) The purpose of a pilot survey is to test the questionnaire.

2. Consider the following statements.

A - Histogram can be constructed even for a frequency distribution with unequal class intervals.
B - The area between the 45 degree line and the Lorenze curve is called Gini co-efficient.
C - If the Lorenze curve lies exactly on 45 degree line, the value of the Gini co-efficient is zero.

Of the above statements,

(1) only *A* is true. (2) only *C* is true.
(3) only *A* and *C* are true. (4) only *B* and *C* are true.
(5) all *A*, *B* and *C* are true.

3. Consider the following statements regarding scales of measurements.

A - There is no relationship among subgroups in nominal scale of measurements.
B - Since interval scale of measurements has unit of measurements, it can be used for mathematical operations.
C - Ratio scale of measurements is the only scale which has a fixed starting point.

Of the above statements,

(1) only *A* is true. (2) only *C* is true.
(3) only *A* and *B* are true. (4) only *A* and *C* are true.
(5) all *A*, *B* and *C* are true.

4. The most suitable diagram to represent the total value with the component values is

(1) Simple bar diagram. (2) Multiple bar diagram. (3) Pictogram.
(4) Profile chart. (5) Pie diagram.

tamilguru.lk

5. The import of a commodity increased by 20% in 2008, decreased by 18% in 2009 and then increased by 30% in the following year. The increase or decrease in each year is measured relative to its previous year. Which of the following is equal to the average rate of change of imports per annum?

(1) 10% (2) 10.7% (3) 22.6%

(4) $[(0.2)(-0.18)(0.3)]^{\frac{1}{3}}$ (5) $[(100+20)(100-18)(100+30)]^{\frac{1}{3}} - 100$

6. If the mid values (X_i) of class intervals of a frequency distribution were transformed into U_i values as $U_i = \frac{X_i - A}{C}$, which of the following gives the mean $\bar{X}$ and the standard deviation σ of the distribution respectively?

(1) $\bar{X} = A + \bar{U}, \sigma_x = C\sigma_u$ (2) $\bar{X} = A + C\bar{U}, \sigma_x = C\sigma_u$

(3) $\bar{X} = A - C\bar{U}, \sigma_x = C\sigma_u$ (4) $\bar{X} = \bar{U}, \sigma_x = C\sigma_u$

(5) $\bar{X} = A + C\bar{U}, \sigma_x = \sigma_u$

7. In a moderately skewed distribution, the mode and the mean are 32 and 35 respectively. What is the median of the distribution?

(1) 32 (2) 33 (3) 34 (4) 35 (5) 36

8. For a certain distribution, Kelly's coefficient of skewness is 0.2 and P_{10} = 60 and median = 80. What is the value of P_{90} of the distribution?

(1) 100 (2) 110 (3) 130 (4) 140 (5) 160

9. Which of the following statements is **false**?

(1) Bowley's co-efficient of skewness cannot be used when a distribution has open end classes.
(2) Kelly's co-efficient of skewness covers more extreme values than Bowley's co-efficient of skewness.
(3) The distribution with a negative co-efficient of skewness has a longer tail to the right.
(4) Bowley's co-efficient of skewness is based on only the central 50% of the observations.
(5) In a distribution with a longer tail to the right, mean > median > mode.

10. The means of runs scored by the five batsmen A, B, C, D and E in a series of 10 innings are 75, 60, 50, 45 and 20 respectively. The standard deviations of their runs are 30, 25, 30, 15, 10 respectively. Who is the most consistent batsman of the five batsmen?

(1) A (2) B (3) C (4) D (5) E

11. A motor car travels 250 km with the speed of 50 km/hour, 120 km with the speed of 40 km/hour and the remaining 50 km with the speed of 25 km/hour. Which of the following is equal to average speed of the motor car for the entire trip?

(1) $38\frac{1}{3}$ km h^{-1} (2) 42 km h^{-1} (3) $63\frac{2}{3}$ km h^{-1}

(4) 140 km h^{-1} (5) $(50 \times 40 \times 25)^{\frac{1}{3}}$ km h^{-1}

12. Consider the following data set.

14, 15, 8, 10, 13, 18, 9, 11, 7, 16, 19, 22, 21

Select the correct answer which gives the first quartile, second quartile, and the third quartile of this data set respectively.

(1) 8, 9, 16 (2) 9.5, 14, 18.5 (3) 9, 14, 18 (4) 8.5, 9.5, 16.5 (5) 10, 15, 19

13. Which of the following statements is true about regression and correlation?

(1) If a constant is subtracted from the two variables X and Y, the correlation co-efficient between X and Y will also change accordingly.
(2) If correlation co-efficient between X and Y is zero, we can conclude that there is no relationship between X and Y.
(3) The correlation co-efficient is only a measure of linear relationship between X and Y.
(4) The free hand method can also be used to fit a multiple regression model.
(5) If the regression co-efficient of Y on X is b_1 and the regression co-efficient of X on Y is b_2 then the correlation co-efficient between X and Y is $b_1 b_2$.

14. Consider the following statements regarding the regression analysis.

A - If the regression co-efficient of *Y* on *X* is positive, the correlation co-efficient between *X* and *Y* is also positive.

B - The co-efficient of determination is equal to the square of the correlation co-efficient in simple linear regression.

C - A multiple regression model can have only two independent variables.

Of the above statements,

(1) only *B* is true. (2) only *A* and *B* are true.
(3) only *A* and *C* are true. (4) only *B* and *C* are true.
(5) all *A*, *B* and *C* are true.

15. If the yield increases by 12 kg when the fertilizer increases by 5 kg according to a fitted regression line, what is the regression co-efficient?

(1) 0.42 (2) 2.4 (3) 5 (4) 7 (5) 10

16. Consider the following statements about the approaches to probability.

A - For the probability of a certain event, every person gets the same answer as the correct answer under classical approach to probability.

B - If the number of all possible outcomes of an experiment is *n* and the number of outcomes favourable to the event *A* is *m*, the probability of the event *A* occurs is $P(A) = \frac{m}{n}$.

C - Under the mathematical approach to probability, it is not required that the probability of the sample space $P(S) = 1$.

Of the above statements,

(1) only *A* is true. (2) only *A* and *B* are true.
(3) only *A* and *C* are true. (4) only *B* and *C* are true.
(5) all *A*, *B* and *C* are true.

17. The sample space for a certain random experiment is $S = \{a_1, a_2, a_3, a_4\}$. The Probability function for the given sample space is

(1) $P(a_1) = \frac{1}{2}, P(a_2) = \frac{1}{2}, P(a_3) = -\frac{1}{4}, P(a_4) = \frac{1}{5}$.

(2) $P(a_1) = \frac{1}{2}, P(a_2) = \frac{1}{4}, P(a_3) = -\frac{1}{4}, P(a_4) = \frac{1}{2}$.

(3) $P(a_1) = \frac{3}{2}, P(a_2) = \frac{1}{4}, P(a_3) = \frac{1}{8}, P(a_4) = \frac{1}{8}$.

(4) $P(a_1) = \frac{1}{2}, P(a_2) = 0, P(a_3) = \frac{1}{4}, P(a_4) = \frac{1}{4}$.

(5) $P(a_1) = \frac{1}{4}, P(a_2) = \frac{1}{5}, P(a_3) = \frac{1}{5}, P(a_4) = \frac{1}{4}$.

More Past Papers at
tamilguru.lk

18. If *A* and *B* are any two events with $P(A) = P_1$, $P(B) = P_2$ and $P(A \cap B) = P_3$ then the probability of the event $A \cup (A' \cap B)$ is

(1) $P_1 + P_2 - P_3$. (2) $P_2 - P_3$. (3) $P_1 - P_3$.
(4) $1 - P_1 - P_2 + P_3$. (5) $1 - P_3$.

19. If *A* and *B* are two events with $P(A \cap B) = \frac{1}{2}$, $P(A' \cap B') = \frac{1}{3}$ and $P(A) = P(B) = k$ then the value of *k* is,

(1) $\frac{1}{3}$. (2) $\frac{1}{2}$. (3) $\frac{7}{8}$. (4) $\frac{8}{9}$. (5) $\frac{7}{12}$.

20. If *A*, *B* and *C* are any three events, which of the following expressions gives the probability that *A* or *B* occur but **not** *C* occurs?

(1) $P(A \cap B \cap C')$ (2) $P[(A \cup B) \cap C']$
(3) $P\left[\left(A' \cap C'\right) \cup \left(B' \cap C'\right)\right]$ (4) $1 - P[(A \cup B) \cap C']$
(5) $P[(A' \cup B') \cap C]$

21. The random variable X has the following probability distribution.

x	0	1	2	3	4	5
$f(x)$	0.1	K	0.2	$2K$	0.3	K

Which could be the smallest value of X for $P(X \leq x) > 0.5$?

(1) 1.0 (2) 2.0 (3) 2.5 (4) 3.0 (5) 4.0

22. If a random variable X has a poisson distribution with $P(X = 1) = P(X = 2)$ then what is the value of $P(X > 0)$?

(1) 0.1353 (2) 0.3879 (3) 0.4060 (4) 0.5940 (5) 0.8647

23. If a male birth or a female birth are equally likely, what is the probability that there are fewer girls than boys in a family of 5 children?

(1) 0.0313 (2) 0.1583 (3) 0.1876 (4) 0.5001 (5) 0.8126

24. The marks of a certain examination are in a normal distribution with mean 76 and standard deviation 15. If the best 15% of the students are awarded A passes, what is the approximately minimum mark to receive an A pass?

(1) 77 (2) 85 (3) 91 (4) 92 (5) 94

25. 2.5% of the items produced by a certain factory is defective. If a random sample of 100 items is selected from these items, the probability that at most one item is defective is

(1) 0.0821. (2) 0.2052. (3) 0.2873. (4) 0.7127. (5) 0.9179.

26. Consider the following statements about the systematic sampling.

A - If the units in the sampling frame are in a random order we can expect that precision of the systematic sampling is same as the precision of the simple random sampling.

B - Systematic sampling can be considered as a cluster sampling with the selection of one cluster from k clusters of size n.

C - In systematic sampling, $\frac{N}{n}$ is called the sampling fraction.

Of the above statements,

(1) only A is true.
(2) only A and B are true.
(3) only A and C are true.
(4) only B and C are true.
(5) all A, B and C are true.

27. Which of the following statements is true about sampling?

(1) If the sampling fraction is large, the finite population correction can be ignored.
(2) If the variations among clusters are large, cluster sampling is more efficient.
(3) Quota sampling can be considered as a non-probability stratified sampling.
(4) Cluster sampling is not used when there is no sampling frame.
(5) The method of selecting a sample giving a known probability to every unit of the population is called simple random sampling.

28. In simple random sampling without replacement, which of the following gives the probability that a certain specified unit of the population is included in the sample?

(1) $\frac{1}{N}$ (2) $\frac{n}{N}$ (3) $\frac{n-1}{N}$ (4) $\frac{1}{NC_n}$ (5) $\frac{1}{N^n}$

29. According to the central limit theorem, the sampling distribution of the sample proportion p is

(1) normal for large samples.
(2) normal if the population proportion is $\pi = 0.5$.
(3) approximately normal if the population size is large.
(4) approximately normal if the sample size is large.
(5) approximately normal only if the population is infinite.

30. Which of the following statements is true?

(1) The accuracy of an estimate is measured by its standard error.
(2) $\bar{X} - \mu$ is always a statistic since it is a function of the sample elements.
(3) The standard error of the mean of a sample from a finite population is larger than the standard error of the mean of a sample from an infinite population for the same sample size.
(4) Chi-square distribution is skewed to the left.
(5) The shape of the T-distribution depends only on the sample size.

31. It is required to estimate the population mean μ by the sample mean $\bar{X}$ of a random sample taken from the population $N(\mu, 100)$. What is the sample size 'n' required for estimating population mean μ, within the range $\mu \pm 5$ with probability 0.954?

(1) 4 (2) 11 (3) 15 (4) 16 (5) 80

32. In a random sample of size 16 from a normal population with mean μ and variance $\sigma^2 = 25$, the sample mean was $\bar{X} = 75$ and the sample variance was $s^2 = 16$. The best 95% confidence interval for μ is

(1) (73.04, 76.96) (2) (72.55, 77.45) (3) (72.33, 77.67)
(4) (72.87, 77.13) (5) (71.94, 78.06)

33. Consider the following statements about confidence intervals.

A - If the sample size is small the confidence interval for the mean μ of a normal distribution based on the *t*-distribution is wider than the confidence interval based on *z*-distribution.
B - One way of reducing the width of a confidence interval for a given confidence level is to increase the sample size.
C - The meaning of the 95% confidence interval for population mean μ is that the variable μ lies in the interval with probability 0.95.

Of the above statements,

(1) only *A* is true. (2) only *B* is true.
(3) only *A* and *B* are true. (4) only *B* and *C* are true.
(5) all *A*, *B* and *C* are true.

34. Which of the following statements is **false**?

(1) If the mean of a normal population with unknown variance is μ, $H_0 : \mu = 100$ is a composite hypothesis.
(2) If the $p-$value of a hypothesis test is high the null hypothesis is more credible.
(3) The value of a test statistic is calculated under the assumption that the null hypothesis is true.
(4) The probability that the H_1 hypothesis is accepted when H_1 is true is called the power of the test.
(5) A better hypothesis test can be performed by reducing the significance level.

35. The mean of a random sample of size 45 from distribution $N(\mu_1, 90)$ is 920 and the mean of a random sample of size 50 from $N(\mu_2, 100)$ distribution is 925. When testing hypothesis $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 < \mu_2$ at 5% significance level, the conclusion is

(1) reject H_0 : since $p-$value $= 0.0062 < 0.05$
(2) do not reject H_0 : since $p-$value $= 0.0062 < 0.05$
(3) reject H_0 : since $p-$value $= 0.0124 < 0.05$
(4) do not reject H_0 : since $p-$value $= 0.0124 < 0.05$
(5) reject H_0 : since $p-$value $= 0.0124 < 1.64$

36. The critical region for testing the hypothesis $H_0 : \mu = 62$ against $H_1 : \mu = 63$ by taking a random sample of size 30 from $N(\mu, 120)$ population is given by $\bar{X} > 64$. The probability of type I error for this hypothesis test is

(1) 0.1587. (2) 0.1915. (3) 0.3085. (4) 0.3413. (5) 0.6587.

37. A candidate in a local electoral area claims that at least 50% of the voters will vote for him. To test his claim a random sample of 100 voters was selected and 48 voters said that they would vote for him. The candidate's claim cannot be rejected at 5% significance level since,

(1) $z = -0.4 > -1.64$ (2) $z = 0.4 < 1.64$
(3) $z = -0.39 > -1.64$ (4) $z = 0.39 < 1.64$
(5) $-1.96 < z = -0.4 < 1.96$

38. The number of errors in 100 accounts selected at random from a company are given below.

Number of Errors	0	1	2	3	4	5	6
Number of Accounts	40	35	19	2	0	2	2

What is the table (critical value) value of the Chi-square distribution in testing goodness of fit at 5% level of the poisson distribution fitted for this distribution?

(1) 5.99 (2) 7.81 (3) 9.49 (4) 11.1 (5) 12.6

39. The incomplete analysis of variance table constructed to compare mean output of three machines is given below.

Analysis of variance table				
Source	SS	df	MS	F
Between samples	a	2	65	d
Within samples	96	12	c	
Total Variation	226	b		

Select the statement which gives the correct values for a, b, c, d respectively.

(1) $a=130$, $b=10$, $c=8$, $d=8.125$
(2) $a=322$, $b=14$, $c=8$, $d=8.125$
(3) $a=130$, $b=24$, $c=84$, $d=0.773$
(4) $a=130$, $b=14$, $c=8$, $d=8.125$
(5) $a=130$, $b=10$, $c=8$, $d=0.123$

40. Consider the following statements about time series analysis.

A - Method of semi-average can be used only when the trend is linear.
B - The multiplicative model of time series assumes that the components caused due to various factors affect each other.
C - In moving average method, it is assumed that the trend varies according to a line.

Of the above statements,

(1) only *A* is true. (2) only *B* is true.
(3) only *A* and *B* are true. (4) only *A* and *C* are true.
(5) all *A*, *B* and *C* are true.

41. The trend equation with the origin 2006 is given by $Y_t = 56 - 4t$. Time unit = 1 year. If the origin is shifted from 2006 to 2002 what is the new trend equation?

(1) $Y_t = 56 - t$ (2) $Y_t = 40 - 4t$ (3) $Y_t = 76 - 4t$
(4) $Y_t = 72 - 4t$ (5) $Y_t = 72 + 4t$

42. In a certain shop the seasonal index for the garment sales is 80 for the first quarter and 130 for the fourth quarter. If the value of the total sales for the first quarter is Rs. 100 000, what is the sales value of the garments that the shop should keep for the fourth quarter in order to meet the demand?

(1) Rs. 61 530 (2) Rs. 130 000 (3) Rs. 162 500 (4) Rs. 500 000 (5) Rs. 800 000

43. The moving average of order 3 of the values 15, 24, 21, 33 and 42 are given by

(1) 20, 22, 30 (2) 20, 26, 32 (3) 20, 23, 32 (4) 20, 24, 33 (5) 20, 25, 34

44. The statistical chart constructed to control the number of defects per unit of a product is
(1) nP - chart. (2) P - chart. (3) C - chart. (4) $\overline{X}$ - chart. (5) R - chart.

45. The average number of defectives in 10 samples each of the size 100 was found to be $\overline{P} = 0.20$. The Lower control limit (L.C.L.) and Upper Control Limit (U.C.L.) of the P – chart respectively are
(1) (0. 16, 0. 24). (2) (0. 18, 0. 28). (3) (0. 20, 0. 32). (4) (0. 08, 0. 32). (5) (0. 08, 0. 20).

46. Consider the following statements.
A - Rejecting a good lot is called producer's risk.
B - The maximum allowable number of defectives in the sample in acceptance sampling is called acceptance number.
C - The quality level of a bad lot is called the Acceptable Quality Level.
Of the above statements,
(1) only A is true. (2) only B is true.
(3) only A and B are true. (4) only A and C are true.
(5) all A, B and C are true.

47. For an acceptance sampling plan with N = 1200, n = 100 and C = 1, what is the probability of acceptance of a lot with fraction defective for 4% ?
(1) 0.0183 (2) 0.0733 (3) 0.0916 (4) 0.9084 (5) 0.9817

48. A worker earned Rs. 30 000 per month in year 2005. The cost of living index increased by 25% in the year 2010 compared to 2005. What should be the salary of the worker in 2010, if his standard of living is to remain the same as in 2005?
(1) Rs. 32 000 (2) Rs. 35 000 (3) Rs. 37 500 (4) Rs. 75 000 (5) Rs. 120 000

49. The price index numbers for the years 2003-2010 are given in the following table. (Base year = 1998).

2003	2004	2005	2006	2007	2008	2009	2010
140	200	210	230	250	260	280	300

Select the correct answer which gives the new index numbers for 2004 and 2010 respectively, if the base year is shifted from 1998 to 2007.
(1) 70, 110 (2) 80, 120 (3) 85, 125 (4) 90, 130 (5) 125, 83

50. In a situation where prices are increasing, the index that tends to overestimate the price increase is
(1) Laspeyre's price index. (2) Paache's price index.
(3) Marshall - Edgewerth price index. (4) Fisher's price index.
(5) Simple aggregate price index.

* * *

More Past Papers at
tamilguru.lk

ශ්‍රී ලංකා විභාග දෙපාර්තමේන්තුව
இலங்கைப் பரீட்சைத் திணைக்களம்

අ.පො.ස.(උ.පෙළ) විභාගය/ க.பொ.த. (உயர் தர)ப் பரீட்சை- 2019

නව නිර්දේශය/ புதிய பாடத்திட்டம

විෂයය අංකය
பாட இலக்கம் 31 - E

විෂයය
பாடம் **Business Statistics**

ලකුණු දීමේ පටිපාටිය/புள்ளி வழங்கும் திட்டம்

I පත්‍රය/பத்திரம் I

ප්‍රශ්න අංකය வினா இல.	පිළිතුරු අංකය விடை இல.	ප්‍රශ්න අංකය வினா இல.	පිළිතුරු අංකය விடை இல.	ප්‍රශ්න අංකය வினா இல.	පිළිතුරු අංකය விடை இல.	ප්‍රශ්න අංකය வினா இல.	පිළිතුරු අංකය விடை இல.	ප්‍රශ්න අංකය வினா இல.	පිළිතුරු අංකය விடை இல.
01.	4	11.	2	21.	4	31.	4	41.	4
02.	3	12.	2	22.	5	32.	2	42.	3
03.	4	13.	3	23.	4	33.	3	43.	2
04.	5	14.	2	24.	4	34.	5	44.	3
05.	5	15.	2	25.	3	35.	1	45.	4
06.	2	16.	1	26.	2	36.	1	46.	3
07.	3	17.	4	27.	3	37.	1	47.	3
08.	2	18.	1	28.	2	38.	1	48.	3
09.	1 හෝ 3	19.	5	29.	4	39.	4	49.	2
10.	4	20.	2	30.	5	40.	3	50.	1

✪ විශේෂ උපදෙස්/ விசேட அறிவுறுத்தல் :

එක් පිළිතුරකට/ ஒரு சரியான விடைக்கு 02 ලකුණු බැගින්/புள்ளி வீதம்

මුළු ලකුණු/மொத்தப் புள்ளிகள் 2 × 50 = 100

Part I

1. (a) State **three** ways that Statistics may be misused. (03 *marks*)

(b) Describe the following methods of data collection stating advantages and disadvantages of each method.

(i) Direct Observation Method

(ii) Focus Group Interview Method

(iii) Electronic Data Collection Method (06 *marks*)

(c) Describe the following measurement scales giving examples.

(i) Nominal Scale (ii) Ranking Scale / Ordinal Scale

(iii) Interval Scale (iv) Ratio Scale (04 *marks*)

(d) The income distributions of the two groups, **A** and **B** are given in the following table.

Income (Thousands Rs.)	Number of Persons (in thousands)	
	Group A	Group B
10	14	08
30	05	07
40	01	06
44	03	02
76	02	02

(i) Calculate the cumulative percentages for the income, for the number of persons in group **A** and for the number of persons in group **B**.

(ii) Draw the **two** Lorenze curves in the same graph and comment on the income distributions of two groups. (07 *marks*)

01. (a) 1. Missinterpretation of analysed results
2. Using inappropriate data for Comparission
3. Interpretation of Statistical outcomos in a favourable manner
4. Giving recommondations without using a Sufficient and reasonable (representative) Sample
5. Selecting the Sample in a biased manner

(03 mark)

(b) (i) **Direct Observation method**

Collecting data observing the real work situation directly in the relevant field of study by the ennumerators is known as Direct Observation Method

Advantages

- accuray being at a higher degree
- rate of responses being at a higher level
- reliability of date being greater
- no any other evidence being requirred to assure the validity of data.

Disadvantages

- uage being limitted
- time consuming and being expensive

- Possibility of collected data being subjective
- possibility of changing the results based on the shortcomings in quality of technical instruments used

(ii) Focuse Group Interview Method

Collecting data discussing with a small group of knoweldgeble people in the relevant field of study is known as Focus Group Interview Method.

Here, the group members are instructed by the invigilator regarding the data to be collected

Advantages

- ability to use in studing the facts in great details.
- being a more apppropriate method to collect qualifafive data such as attitudes, Faiths or experiences.
- being economic relative to the other methods of collecting data.
- rate of responses being at a greater degree with exchange of points of view through discussion.
- data being highly reliable.
- ability of using as an additional way of collecting descriptive data after collecting quantitative data.
- ability of expanding the sample size compared to the personal interview method, because of the possibility of interviewing many people during the same period.

Disadvantages

- receiving a variety of openions as responses related to a particcdar objective.
- because of the same Fact, data anaysis being complicated.
- achiving at conclussions may be delayed.

(iii) Electronic Data Collection Method

Collecting data applying modern electronic devices as tools (instruments) is known as Electronic Data Collection Method. This is also defined as E - reserches. There are few major techniqus of collecting data using electronic devices as Follows

- Computer Assisted Personal Interviewing (CAPI)
- Computer Assisted Self Interviewing (CASI)
- E mail Survey
- Web Surveys

Advantages

- Convenience in usage
- ability of collecting data quickly
- being economic
- Data organising being convenient
- ability of collecting data from globally scatterred individuals

Disadvantages

- Possibility of rate of rensponses being poor, when the computer literacy of respondents is inadiquate.
- Possibility of being less reliable
- inability of drawing a representative sample, in the absence of access to modern technology among some of the sample members

(06 marks)

(C) (i) **Nominal Scale**

Once the attribute related to a categorical variable is available in names; the numbers or symbols are used as codes only with the purpose of categorising (seperating) thaose attribules and the data collected in that manner are called Nominal Scale Data.

- No any mathematical operation is applicable on those coding numbers. (Those figures have no any numerical Value)

  ex :- Gender
  1 - Male
  2 - Female

(ii) **Ranking / Ordinal Scale**

The codes assigned on attributes related to categorical Variables in a convenient manner to catogrize and compare are called ranks and such data are known as Ranking scale data/ ordinal Scale data.

There isn't a constant difference among those rankings.

ex : Inquiring the preference of students for Mathematics. Mentioning the codes related to the choice.

like very much- 4
like - 3
dislike - 2
dislike at all - 1

- Grading the cricketers in accordance with their performance / cleverliness

(iii) **Interval Scale**

Data having a Zero, but not true zero, with equal diferrences followed by a relative order that can be undergone to Some (addition, Subtraetion) mathematical operations are called Interval Scale Data.

Ability of identifying the Variable stuting the magnitude as well as availability of intervals can be regarded as specific features.

ex : ℃ and ℉ in measuring temperature

$0℃ = 32\ F°$

$0F° = -17.7746\ ℃$

(iv) **Ratio Scale**

Data that having a true zero with equal differences and that can be undergone to all the methematical operations where the ratio of two numbers being meaningful are called Ratio Scale Data Categorieal variables of this scale are characterized by having a true zero, equal difference and a magnitude.

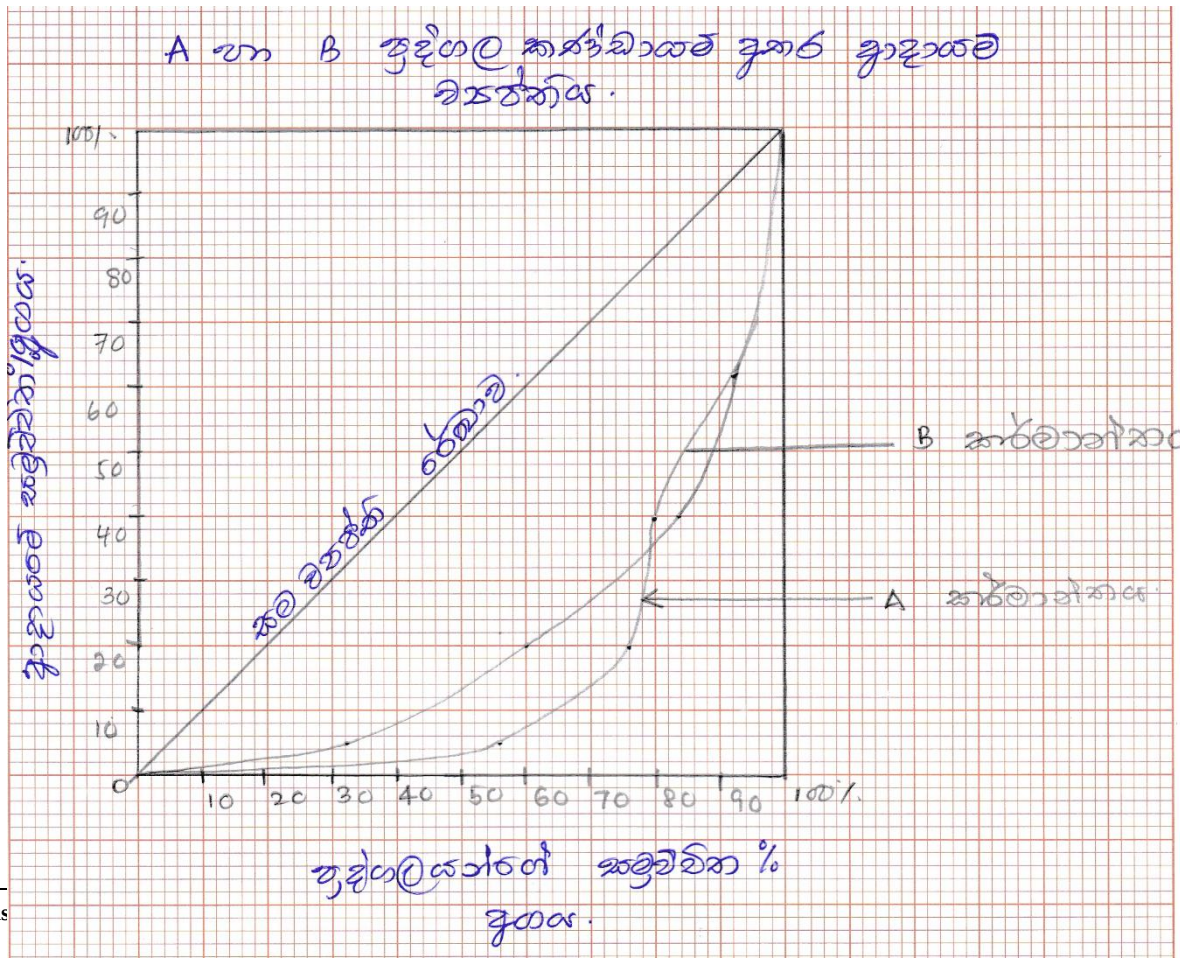
ex : Scores, age, / life time, height, mass, income, expenditure.....etc.

(04 marks)

(d) (i) Lorenz Curve

			Group A			Group B		
Income	ප්‍රතිශතය	සමුච්චිත ප්‍රතිශතය	පුද්ගලයන් ගණන	ප්‍රතිශතය	සමුච්චිත ප්‍රතිශතය	පුද්ගලයන් ගණන	ප්‍රතිශතය	සමුච්චිත ප්‍රතිශතය
10	5	5	14	56	56	08	32	32
30	15	20	05	20	76	07	28	60
40	20	40	01	04	80	06	24	84
44	22	62	03	12	92	02	08	92
76	38	100	02	08	100	02	08	100

(d) (i)



(ii) Since these two Lorenze curves are intersected by one another an accurate comparission can be made using the Ginii Coefficient,

Any way it's obvious the fact that there is a greater variation in distributing the income among the persons in group A because the area included to the curve. A is bigger than the area excluded from the curve A, through this intersection

20% of the total income in group A has distributed among 76% of the people, making a significant disparity.

But 20% of the total income in Group B has distributed among 60% of the people, While the remaing 80% of the income goes to 40% of the people which is considerably a higher proportion compared to that of in group A

Accodingly a bigger variatiion in distributing the income is depicted in group A.

(07 marks)

2. (a) Describe what is meant by skewness and kurtosis of a distribution.
The per hour wage rates of 100 workers are given in the following distribution.

Wage Rate	10-19	20-29	30-39	40-49	50-59	60-69
Number of workers	08	12	20	35	20	05

Calculate Kelly's coefficient of skewness based on percentiles and comment on the skewness of the distribution. (06 *marks*)

(b) The coefficients of variation of wages of male workers and female workers in a certain institute are 55% and 60% respectively, while the standard deviations are 22 and 15 respectively. If 80% of the workers are male, calculate the overall average wage of all workers. (04 *marks*)

(c) The heights of the students of a certain class are given in the following distribution.

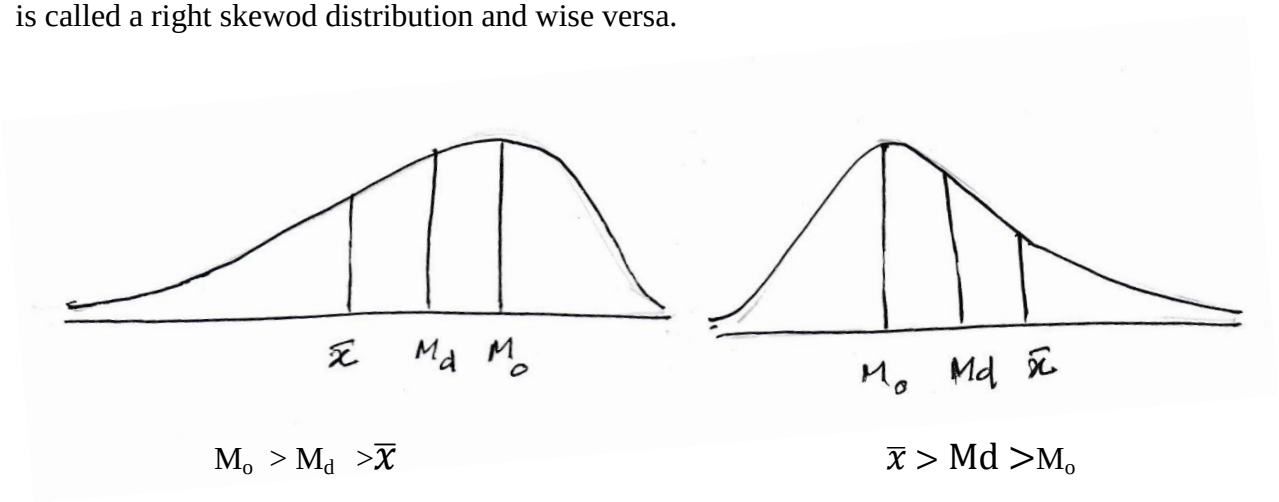
Height (inches)	58-60	61-63	64-66	67-69	70-72	73-75
Number of students	10	20	30	20	15	05

Calculate the mean, median, mode, standard deviation and Karl Pearson's coefficient of skewness and comment on the distribution. (10 *marks*)

02. (a) **Skewness**

The asymmetrical nature of a distribution or how the data are apart from the symmetrical nature is known as skewness.

If the Frequency curve representing the distribution of a variable Falls with a longer tail towards the right hand side rarther than towards left from the peak (maximum point) that is called a right skewod distribution and wise versa.



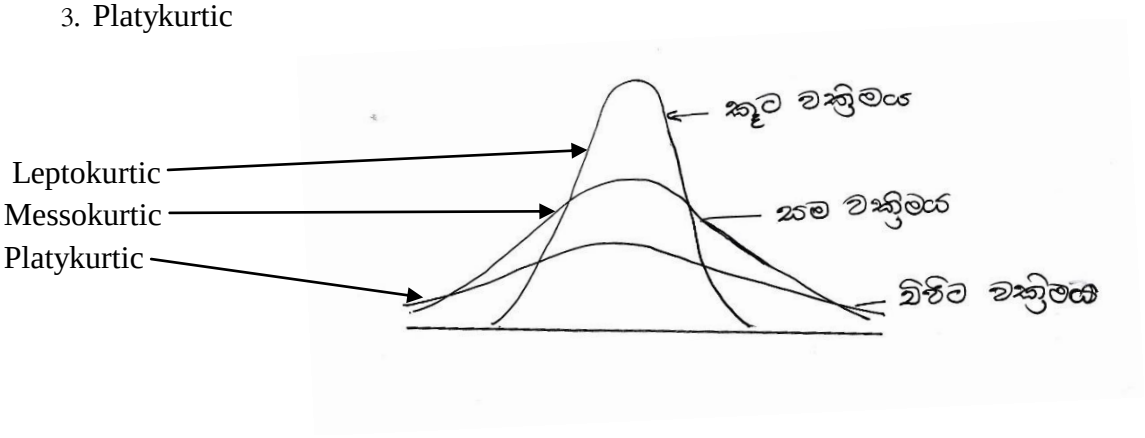
$M_o > M_d > \bar{x}$ $\bar{x} > Md > M_o$

Kurtosis :

The nature of peakedness (curvatare) of the frequncy curve representing the distribution of a variable is known as 'Kurtosis' The type of kurtosis is exp;ained relative to the normal distribution curve.

Kurtosis is in three types relative to the normal distribution

1. Leptokurtic
2. Messokurtic
3. Platykurtic



Class Interval	Frequency	Camulative Frequency
10 - 19	08	08
20 - 29	12	20
30 - 39	20	40
40 - 49	35	75
50 - 59	20	95
60 - 69	05	100

$P_{10} = L_1 + \left[\frac{\frac{10n}{100} - f_c}{f_{pm}}\right] c$

$= 19.5 + \left[\frac{10-08}{12}\right] 10$

$= 19.5 + \frac{2}{12} \times 10$

$= 19.5 + 1.67$

$= \underline{21.17}$

$P_{50} = L_1 + \left[\frac{\frac{50n}{100} - f_c}{f_{p50}}\right] c$

$= 39.5 + \left[\frac{50-40}{35}\right] 10$

$= 39.5 + \left[\frac{10}{35}\right] 10$

$= 39.5 + 2.857$

$= \underline{42.357}$

P_{90} $= L_1 + \left[\frac{\frac{90n}{100} - f_c}{f_{p90}}\right] c$

$= 49.5 + \left[\frac{90-75}{20}\right] 10$

$= 49.5 + \left[\frac{15}{2}\right]$

$= 49.5 + 7.5$

$= \underline{57.0}$

$S_{Kp} = \frac{P_{90} + P_{10} - 2P_{50}}{P_{90} - P_{10}}$

$= \frac{57.0 + 21.17 - 2 \times 42.36}{57.0 - 21.17}$

$= \frac{78.17 - 84.72}{35.83}$

$= \frac{-6.55}{35.83}$

$= \underline{-0.18}$ This is a negatively skewed distribution

(06 marks)

(b)

Male workers	Female workers
CV = 55%	CV = 60%
S = 22	S = 15
$CV = \frac{s}{\bar{x}} X \quad 100$	$CV = \frac{S}{\bar{X}} \times 100$
$55 = \frac{22}{\bar{x}} X \quad 100$	$60 = \frac{15}{\bar{x}} X \quad 100$
$\overline{x} = \frac{22}{55} X \quad 100$	$\overline{x} = \frac{15}{60} X \quad 100$
$\overline{x} = 40$	$\overline{x} = 25$
$n_1 = 80$	$n_2 = 20$

$$\overline{x} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}$$

$$= \frac{80 \times 40 + 20 \times 25}{80+20}$$

$$= \frac{3200 + 500}{100}$$

$= \frac{3700}{100}$

$\bar{\bar{x}} = 37$

(04 marks)

(c)

Height (inches)	No : of Students	Midvalue (classmarks)	u	u^2	fu	fu^2	fc
58 - 60	10	59	-2	4	-20	40	10
61 - 63	20	62	-1	1	-20	20	30
64 - 66	30	**65**	0	0	0	0	60
67 - 69	20	68	1	1	20	20	80
70 - 72	15	71	2	4	30	60	95
73 - 75	05	74	3	9	15	45	100
	100				**25**	**185**	

Mean

$\bar{x} = A + \left(\frac{\sum fu}{\sum f}\right) C$

$= 65 + \left(\frac{25}{100}\right) 3$

$\bar{x} = 65.75$

Median

$M_d = L_1 + \left(\frac{\frac{n}{2} - f_C}{fm}\right) C$

$= 63.5 + \left(\frac{\frac{100}{2} - 30}{30}\right) 3$

$= 63.5 + \frac{20}{30} \text{x } 3$

$= 65.5$

Mode

$M_o = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) C$

$= 63.5 + \left(\frac{10}{10+10}\right) 3$

$= 63.5 + \left(\frac{30}{20}\right)$

$= 63.5 + 1.5$

$= 65$

Standard Deviation

$S = C \sqrt{\left[\frac{\sum fu^2}{\sum f} - \left(\frac{\sum fu}{\sum f}\right)^2\right]}$

$S = 3 \sqrt{\left[\frac{185}{100} - \left(\frac{25}{100}\right)^2\right]}$

$S = 3 \sqrt{[1.85 - 0.0625\ ]}$

$S = 3 \sqrt{1.7875}$

$S = 4.01$

$$S_{k1} = \frac{\bar{X} - M_0}{S} \quad \text{or} \quad S_{k2} = \frac{3(\bar{X} - M_d)}{S}$$

$$= \frac{65.75 - 65}{4.01} \qquad = \frac{3(65.75 - 65.5)}{4.01}$$

$$S_{k1} = 0.187 \qquad S_{k2} = 0.187$$

(10 marks)

3. (a) What is an index number?

Explain the Laspeyre's price index and Paasche's price index in terms of the total cost of a basket of goods in the base year and total cost of a basket of goods in the given year. (03 *marks*)

(b) Consider the following table.

	Base Year		Current Year	
Item	**Price**	**Total Value**	**Price**	**Total Value**
A	6	300	10	560
B	4	240	06	360
C	2	200	02	240
D	8	320	12	960
E	10	300	12	288

Using the data in the table, calculate the following.

(i) Laspeyre's price index
(ii) Paasche's price index
(iii) Fisher's price index
(iv) Marshall-Edgeworth price index

Does the Marshall-Edgeworth price index satisfy the time reversal test and the factor reversal test? Give reasons for the answer. (07 *marks*)

(c) What is a Time Series?
Describe **three** uses of time series analysis in the business field.
Describe what is meant by cyclical variation and seasonal variation in time series analysis. (05 *marks*)

(d) The trend equation fitted by the method of least squares for the sales of garments is given below.

$Y = 840 + 72X$
Origin is 2005
Time unit = 1 year
Y = Number of units sold per year

(i) Convert this trend equation into a monthly trend equation.
(ii) Estimate the sale for the month of October in the year 2011. (05 *marks*)

03. (**a**) A statistical measure which is used to quantify the changes of a variable or a set of related variables moving in accordanco with time or geographical location between two periods or two locations any other factor is known as an index number.

Lasper Price Index

Onece the total expenditure in current year for a basket of items in base year is expressed as a percentage to the expenditure incurred for the same basket of items in the base year that index is known as Laspear price index

$$LP_{n/o} = \frac{\sum p_n q_0}{\sum p_0 q_0} \text{X}100$$

Pasche Price Index

Onece the total expenditure of the basket of items consumed in the current year is expressed as a percentage to the total expenditure of the same basket of items in base year that index is known as Pasche Price Index

$$LP_{n/o} = \frac{\sum p_n q_0}{\sum p_0 q_0} \text{X}100$$

(03 marks)

(b)

Item	Base Year		Current Year		$p_o q_o$	$p_o q_n$	$p_n q_o$	$p_n q_n$
	Price	Quantity (Units)	Price	Quantity (Units)				
A	6	50	10	56	300	336	500	560
B	4	60	6	60	240	240	360	360
C	2	100	2	120	200	240	200	240
D	8	40	12	80	320	640	480	960
E	10	30	12	24	300	240	360	288
					1360	1696	1900	2408

I. Laspere's Price index

$$LP_{n/o} = \frac{\sum p_n q_0}{\sum p_0 q_0} \text{X}100$$

$$= \frac{1900}{1360} \text{x } 100$$

$= 139.7$

II. Paasche's price index

$$PP_{n/o} = \frac{\sum p_n q_n}{\sum p_0 q_n} \text{X}100$$

$$= \frac{2408}{1696} \text{x } 100$$

$= 141.98$

$= 142$

III. Fisher's price index

IV. Marshal Edeworth price index

$FP_{n/o}$ $\sqrt{LPn/o\ X\ PPn/o}$

$= \sqrt{139.7 \text{ x } 141.9}$

$= 140.79$

$= 140.8$

$M_EP_{n/o} = \frac{\sum p_n(q_o+q_n)}{\sum p_0(q_o+q_n)} \text{ x } 100$

$= \frac{4308}{3056} \text{ x } 100$

$= 140.96$

$= 141$

(05 marks)

$P_o(q_o+q_n)$	$P_n(q_o+q_n)$
636	1060
480	720
440	440
960	1440
540	648
3056	4308

$$M_E P_{n/o} = \frac{\sum Pn.\,½(qo + qn)}{\sum Po.\,½(qo + qn)} \text{ x } 100$$

$$M_E P_{n/o} \text{ x } M_E P_{o/n} = \frac{\sum P_n (q_o + q_n)}{\sum P_o (q_o + q_n)} \text{ x } \frac{\sum P_o (q_o + q_n)}{\sum P_n (q_o + q_n)}$$

$$= \frac{4308}{3056} \quad \frac{3056}{4308}$$

$$= 1$$

The time reversal test is satisfied by Marshal Edgeworth Price index

$$M_E P_{n/o} \text{ x } M_E Q_{n/o} = \frac{\sum P_n (q_o + q_n)}{\sum P_o (q_o + q_n)} \text{ x } \frac{\sum Q_n (p_o + p_n)}{\sum Q_o (p_o + p_n)}$$

$$= \frac{4308}{3056} \text{ x } \frac{4104}{3260}$$

$$= 1.7746$$

$$V_{n/o} = \frac{\sum P_n q_n}{\sum P_o q_o} \text{ x}100$$

$$= \frac{2408}{1360} \text{ x}100$$

$$= 1.7705$$

The factor reversal test is not Satistfied by Marshal Edgeworth Price index

(02 marks)

03. (c) A sequence of data collected for a particalar variable in equal and successive time intervals is known as a time series.

If y_1, y_2, y_3 .... y_n are a set of data collected for a variable denoted by Y for t_1, t_2, t_3... t_n Corresponding time interview y_1,y_2,y_3.....yn is a time series.

Uses

1. Ability to Forecast analysing the past data.
2. Ability to use for making marketing plans and production plans.
3. Ability of identifying the time series components.
4. Ability of comparing two or more related time series.

Cyclical Variation :-

Oscilations found in long term trend in a time series prevaling for a period of more than one year are known as cyclical variations.

ex : Economic expansions (recoveries) and Recession civil wars, Political crisis,introduction of new products.

Seasonal Variations

Short term Fluctuations repeated in equal time intervals in a period less than one year are known as seasonal variations.

ex : Changes in cilimate, regular festivals and Customs and Patterns in individual life styles etc....

(05 marks)

03. (d) (i) Annual trend equation

Y = 840 + 72 x (origin 2005)

Monthly trend equation

$$Y = \frac{840}{12} + \frac{72}{144} x$$

$Y = 70 + 0.5 \times$ (origin July 01th 2005)

(ii) Sales for month of October 2011 (X = 75.5)

Y = 70 + (0.5 x 75.5)

= 70 + 37.75

= 107.75

or

Shift the base to January 2006

$Y = 70 + 0.5\,[x + 6.5]$

$= 70 + 0.5x + 3.25$

$Y = 73.25 + 0.5\,X$ (Origin January 15 2006)

Sales for month of October 2011 (X = 69)

$Y = 73.25 + 0.5$ x 69

= 73.25 + 34.5

= 107.75

(05 marks)

4. (a) A sales department of a certain company gives a training to its salesmen and then a test is held. The following table gives the test scores, and sales made by the salesmen after the training.

Test scores (*X*)	19	24	14	22	26	21	19	20	15	20
Sales (in Rs.1000) (*Y*)	36	48	31	45	50	37	39	41	33	40

$\sum X = 200$, $\sum Y = 400$, $\sum X^2 = 4120$, $\sum Y^2 = 16346$, $\sum XY = 8193$

(i) Calculate the correlation coefficient between test scores and sales, and state whether there is a relationship between them.

(ii) Fit the regression line of Y on X by the method of least squares.

(iii) Calculate the coefficient of determination and comment on your result.

(iv) The department is considering to terminate the service of some salesmen based on the test scores and sales. If the department expects a minimum sale of Rs. 30 000 from each salesman, what should be the minimum test score to consider the termination of the service of a salesman? (10 *marks*)

(b) Explain the difference between the terms of each pair given below.

(i) Chance variation and Assignable variation

(ii) Process Control and Product Control (04 *marks*)

(c) Explain the difference between **C-chart** and **U-chart**.

The number of defects in ten woollen carpets manufactured are given in the following table.

Carpet No.	1	2	3	4	5	6	7	8	9	10
Number of defects	2	3	6	5	3	3	6	4	5	3

Construct a suitable control chart for these data and state whether the quality characteristic under inspection is in control. (06 *marks*)

04. (a) (i) $\sum X = 200$ $\sum Y = 400$

$\sum X^2 = 4120$ $\sum Y^2 = 16346$

$\sum XY = 8193$

$$r = \frac{n\Sigma XY - \sum X \sum Y}{\sqrt{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2}}$$

$$r = \frac{10 \text{ x } 8193 - 200 \text{ x } 400}{\sqrt{[10 \text{ x } 4120 - -(200)^2][10 \text{ x } 16346 - (400)^2]}}$$

$$r = \frac{81930 - 80000}{\sqrt{[41200 - 40000\ \ ][163460\ - 160000\ \ ]}}$$

$$r = \frac{1930}{\sqrt{[1200\ X\ 3460]}}$$

$$r = \frac{1930}{\sqrt{4152000}}$$

$$r = \frac{1930}{2037.6}$$

$$r = 0.9471\ =\ 0.95$$

There is a strong positive correlation between test scores and sales revenue.

$$r = \frac{\sum XY - \frac{\sum XY}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

(ii) $\widehat{\beta}1 = \frac{n\sum XY - \sum X\sum Y}{n\sum X^2 - (\sum X)^2}$ or $\widehat{\beta}1 = \frac{\sum XY - \frac{\sum X\sum Y}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}}$

$$= \frac{10\ \text{x}\ 8193 - 200\ \text{x}\ 400}{10\ \text{x}\ 4120 - 200\ \text{x}\ 200}$$

$$= \frac{81930 - 800000}{41200 - 40000}$$

$$= \frac{1930}{1200}$$

$\widehat{\beta}1$ = 1.608

$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1\bar{x}$
= 40 - 1.608 x 20
= 40 - 32.16
= 7.84

$\hat{y}$ = 7.84 + 1.608x

(04 marks)

(iii) $R^2 = \hat{\beta}_1^2\left[\frac{n\Sigma X^2 - (\sum X)^2}{n\Sigma Y^2 - (\sum Y)^2}\right]$ ෙහ්

Since $R^2 = r^2$ r = 0.9471

$R^2 = (0.9471)^2$

$\underline{\underline{R^2 = 0.8969}}$

Since 89% of the variation of the dependent variable is described by independent variable through this regression model, the regression line that has been fit is appropriate.

(02 marks)

$\hat{y} = 7.8 + 1.61x$

$30 = 7.8 + 1.61x$

$30 - 7.8 = 1.61\ x$

$22.2/1.61 = 13.78$

$x = 13.8$

$\underline{\underline{x = 14}}$ Minimum test score is 14.

(01 marks)

(b) (i) Chance variation and Assignable variation

Variations occurred in random due to natural forces in a process of production which are beyond human control are called chance variations.

ex : Changes take place in liquidity ,
Changes take place in temperature,

Random causes are independent from one another and those can not be concealed or eleminated.

Variations occurred in the quality of a product due to causes that can be identified are called assignable variations

ex : Break - downs in machines, labourers becoming tired. machine not being maintained, usage of defective raw materials may cause for these variations.

(02 marks)

(ii) Process Control and Product Control

The Procedure of investigating in to the process of production to check whether the products are being manufactared to meet the pre determined quality standards is known as process control. The process control is implemented through control charts.

The procedure of investigating in to the quality of raw materials / Spare parts or Finished goods of a product to check whether they meeting the pre determined quality levels is known as product control. The product control is performed using the technique of Acceptance sampling palans.

(04 marks)

(c) C - chart

The control chart which is used to control the number of flaws (defects) contained in a unit of a product is knows as the 'C' chart

$CL = \overline{C}$

$UCL = \overline{C} + 3\sqrt{\overline{C}}$

$LCL = \overline{C} - 3\sqrt{\overline{C}}$

U - Chart

The control Chart which is used to control the number of defects contained in a unit of a product consists of few components is known as 'U' chart.

'U' chart is constructed by computing the number of defects in each cornponent of the relevant product as 'U'

$CL = \overline{u}$

$UCL = \overline{u} + 3\sqrt{\overline{u}}$

$LCL = \overline{u} - 3\sqrt{\overline{u}}$

(02 marks)

(c) $CL = \overline{C}$

$\overline{C} = \frac{\sum C}{K}$ K= Number of samples

$\overline{C} = \frac{40}{10}$

$= 4$

$CL = \overline{C}$

$= 4$

$UCL = \overline{C} + 3\sqrt{\overline{C}}$

$= 4 + 3\sqrt{4}$

$= 4 + 3 \times 2$

$\underline{UCL = 10}$

$LCL = \overline{C} - 3\sqrt{\overline{C}}$

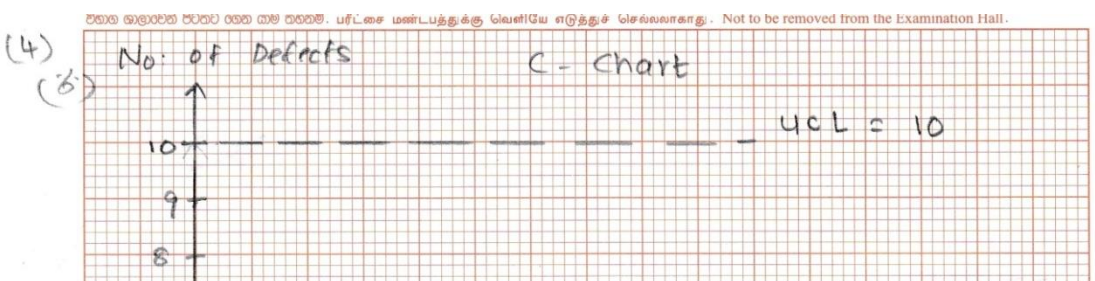
$= 4 - 3 \times 2$

$= 4 - 6$

$= -2$

$\underline{\underline{= 0}}$

d



Conslasion

Since all the points fall between the control limits of the C chart the process is under control

(04 marks)

Part II

5. (a) Describe the classical approach to probability and relative frequency approach to probability, stating **two** limitations of each. (04 *marks*)

(b) If $P(A)=\frac{1}{2}$, $P(A\cup B)=\frac{3}{4}$ and $P(B')=\frac{5}{8}$,

(i) Find $P(A'\cap B')$, $P(A'\cup B')$ and $P(B\cap A')$

(ii) State whether the events A and B are independent. (04 *marks*)

(c) In a manufacturing industrial firm, there are 5 production engineers and 3 maintenance engineers in one section and there are 4 production engineers and 5 maintenance engineers in the other section. From any of these sections, a single selection of two engineers was made. Find the probability that one of them would be a production engineer and the other person would be a maintenance engineer. (04 *marks*)

(d) State the law of total probability and Bayes' Theorem.
The probability that a doctor will diagnose a disease X correctly is 0.8. The probability that a patient with disease X will die by his treatment after correct diagnosis is 0.3. The probability that a patient with disease X will die after not diagnosing the disease correctly is 0.7. If a patient with disease X died, find the probability that the doctor had diagnosed disease X correctly. (08 *marks*)

5. (a) Classical Approach

Once all the possible outcomes of a random experiment are equally likely as well as mutually exclasive the ratio of the number of outcomes receivable in favour of a particular event defined on the sample space to the total number of outcomes to the total number of outcomes in the sample space, that ratio is considerred as the probability of cccuring that event and that approach,is called "the classical approach to probability.

Linitations

1. This approach is inapplicable when the outcomes are not equally likely,
2. Once the sample space of a random experiment is infinite, this approach is inapplicable

Relative Frequency Approach

Once a random experiment is repeated for 'n' number of times under necessarily required identical cricumstances and if the number of times that the outcomes in favour of a

particular event received is m, then $\frac{m}{n}$ is the relative frequncy of occurring that event. When the number of trials is increased if this ration approaches at a particalar constant, that constant is considerred as the probability of occurring that event and that approach is known as the Relative Frequency Approach

Limitations :

1. inability to use when an experiment can not be repeated under identical cricumstances.
2. the probability value assigned for an event not being unique.

(04 marks)

(b) $P(A) = {}^1/_2$, $P(A \cup B) = {}^3/_4$, $P(B') = {}^5/_8$

$P(B) = 1 - P(B')$
$= 1 - {}^5/_8$
$= {}^3/_8$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$¾ = ½ + {}^3/_8 - P(A \cap B)$
$P(A \cap B) = ½ + {}^3/_8 - ¾$
$= \frac{4+3-6}{8}$
$P(A \cap B) = {}^1/_8$

(i) $P(A^1 \cap B^1) = P(A \cup B)^1$
$= 1 - P(A \cup B)$
$= 1 - ¾$
$= ¼$

$P(A^1 \cup B^1) = P(A \cap B)^1$
$= 1 - P(A \cap B)$
$= 1 - {}^1/_8$
$= {}^7/_8$

$P(B \cap A^1) = P(B) - P(A \cap B)$
$= {}^3/_8 - {}^1/_8$
$= {}^2/_8$
$= ¼$

(ii) If A and B events are indepent they should be related as P(A) . P(B) = P(A∩ B)

P(A) x P(B) = ½ X $^3/_8$

$= \frac{3}{16}$

P(A∩ B) $= {}^1/_8$

$P(A \cap B) \neq P(A).P(B)$

∴ The events A and B are not independent

(04 marks)

5. (c) Section 1 **Section 2**

No : of Production Engineers	= 5	No : of production Engineers	= 4
No : of maintenance Engineers	= 3	No : of maintenance Engineers	= 5

Probability to select an production Engineer and a maintenance Engineer

$= \frac{{}^9C_1 \text{ X } {}^8C_1}{{}^{17}C_2}$

$= \frac{\frac{9!}{8!.1!} \text{X} \frac{8!}{7!.1!}}{\frac{17!}{15!.2!}}$

$= \frac{9 \text{ x } 8}{17 \text{ x } 8} = \frac{72}{136}$

$= \frac{9}{17}$

$= 0.529$

or

$= ½ \left[\frac{{}^5C_1 \text{ X } {}^3C_1}{{}^8C_2} + \frac{{}^4C_1 \text{ X } {}^5C_1}{{}^9C_2}\right]$

$= ½ \left[\frac{5 \text{ X } 3}{28} + \frac{4 \text{ X } 5}{36}\right]$

$= ½ \left[\frac{15}{28} + \frac{20}{36}\right]$

$= \frac{275}{504}$

$= 0.546$

(04 marks)

5. (d) The Law of Total Probability

Once A_1, A_2, A_3 ..... A_n are a set of mutually exclusive and collectively exhaustive events, while B is an another event defined on the same sample space common to all those events, the probability of occurring B is defined as the total probability law as follows.

$P(B) = P(A_1).\ P(B/A_1) + P(A_2).\ P(B/A_2) + \ldots\ldots\ldots + P(A_n).\ P(B/A_n)$

When this expression is summarised,the total probability of event B is given by

$$P(B) = \sum_{i=1}^{n} P(A_i)\ .\ P\ (B/A_i)$$

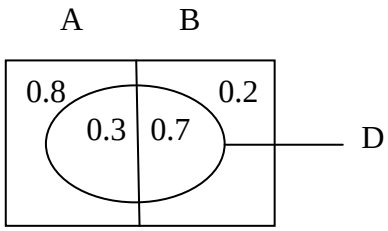
Bayes' Theorem

Given that the event B has occurred in the above sample space the conditional probability of occuring and event denoted by A_1 is the Bayes Theorm and can be expressed as

$$P(A_j\ /\ B) = \frac{P(A_j) \cdot P(B/A_j)}{\sum_{i=1}^{n} P(A_i) \cdot P(B/A_i)}$$

(03 marks)

A ⁝ Correct diagnosis of the disease
B ⁝ incorrect diagnois of the disease
D ⁝ Patient being dead



P(A) = 0.8 P(B) = 0.2
P(D/A) = 0.3 P(D/B) = 0.7

P(D) = P(A). P(D/A) + P(B) . P(D/B)
= 0.8 x 0.3 + 0.2 x 0.7
= 0.24 + 0.14
= 0.38

$$P(A/D) = \frac{P(A).P(D/A)}{P(D)}$$

$$= \frac{0.8 \text{ x } 0.3}{0.38}$$

$$= \frac{0.24}{0.38}$$

$$= \frac{24}{38}$$

$$= \frac{12}{19}$$

(05 Marks)

6. (a) State the probability function of the binomial distribution. What are the conditions that should be satisfied by a random experiment for deriving this function?

On average, 20% of the nails produced by a certain machine is defective. A batch is accepted if a random sample of 10 nails taken from that batch does not contain defective nails and the batch is rejected, if the sample contains 3 or more defectives. In other cases, a second sample is taken. Find the probability of taking a second sample. (06 *marks*)

(b) Define the Poisson distribution and state **three** examples for the application of this distribution.

The number of telephone calls received at a switchboard in any time interval of length T minutes has a Poisson distribution with mean $\frac{1}{2}T$. The telephone operator leaves the switchboard for 6 minutes.

(i) Find the probability that **no call is coming** when the operator **is not at the switchboard**.

(ii) Find the probability that three or more calls are coming when the operator is **not at the switchboard**.

(iii) Find also the maximum length of time in nearest second for which the operator could be absent with 90% probability of receiving no calls.

($\log_{10} e = 0.4343$, $\log_{10} (0.90) = -0.0458$) (06 *marks*)

(c) Explain **three** uses of the normal distribution in the field of statistics.

The life-time of a certain kind of bulbs has a normal distribution with mean life-time of 500 hours and standard deviation of 45 hours. Find,

(i) the percentage of bulbs with a life-time of at least 570 hours.

(ii) the percentage of bulbs with life-time **between** 485 and 515 hours.

(iii) the minimum life time of the best 5% of the bulbs. (08 *marks*)

6. (a) Once a random experiment is conducted for n number of independent trials containg only two outcomes an success and failure, with probability of success (P) at each trial is a constant, the probability receving X number of successes is given the probability mass function

$P(X = x) = {}^{n}C_{x}\, p^{x} q^{n-x}$ මගින් ලබාදේ.

මෙහි $x = 0, 1, 2, \ldots\ldots n$

$q = (1 - p)$

Conditions

1. The random experiment should consist of a specific number (n) of trials.
2. Each trial should consist of only two outcomes as 'success' and 'failure'
3. The probability of receiving success (p) at each trial should be equal

4. Each trial should be independent from all the other trials.

x : හම්පඉැර දෙ ාැඳේඑසඩැ බ්සකි
n = 10 P = 0.2 q = 0.8
x ~ B (10, 0.2)
$P(X=x) = {}_{n}C_{x} p^{x} q^{n-x}$
$P(X = x) = {}_{10}C_{x} (0.2)^{x} (0.8)^{10-x}$ x = 0, 1 , 2 , 3 10

Probability of drawing the second sample = P(x = 1) + P(x = 2)
= 0.2684 + 0.3020
= 0.5704

(06 marks)

(**b**) The probability mass function of a descreterandom variable when that distributes on time or space is considered as the poisson destribution,and the number of times that a particular event defined on time or space occurs is denoted by X, the probability mass function of X is as follows.

$$P(X = x) = \frac{e^{-\lambda} \lambda^{x}}{x!}$$

Where x = 0, 1 , 2 .....
and e = 2.7183.....

Examples

1. Number of calls received by a telephone exchange in an interval of 1 minute.
2. Number of customers coming to a bank counter in an hour.
3. Number of printing errors found in a page of a printed book.

(i) $\lambda = \frac{1}{2}T$

$\lambda = \frac{1}{2} x6$

$\lambda = 3$

$$P(X = x) = \frac{e^{-\lambda} \lambda^{x}}{x!}$$

$$= \frac{e^{-3} 3^{x}}{x!} \qquad x = 0, 1, 2 .....$$

I P(X = 0) = 0.0498

II $P(X \geq 3)$ = 1 - [P(x = 0) + P (x = 1) + P(x =2)]
= 1 - [0.0498 + 0.1498 + 0.2240]
= 1 - 0.4236
= 0.5764

III $\dfrac{e^{-\frac{1}{2}T}\left(\frac{1}{2}T\right)^0}{0!} = 0.9$

$e^{-\frac{1}{2}T} = 0.9$

$\log_{10} e^{-\frac{1}{2}T} = \log_{10} 0.9$

$-\frac{1}{2}T \log_{10} e = \log_{10} 0.9$

$-\frac{1}{2}T \times 0.4343 = -0.0458$

$T = \dfrac{0.0458 \times 2}{0.4343}$

$= \dfrac{0.0916}{0.4343}$

$= 0.2109$

Number of seconds $= 0.2109 \times 60$

$= 12.654$

$= \underline{\underline{12}}$

(06 marks)

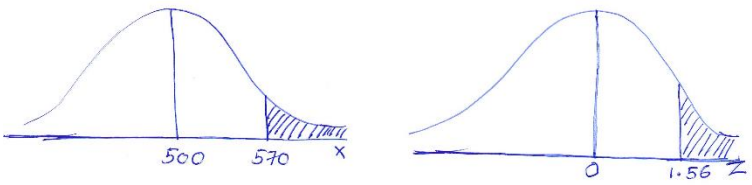
(c) **Uses of Normal distribution**

1. Because most of the continuos variables distribute normally, ability of using normal distribution to solve probability base problems .
2. Ability of using the normal distributions under required conditions,as an aproximation to other discrete distributional
3. Ability of using as a probability distribution in statistical inference, as most of the sample statistics falls or approximately falls normally.
4. Ability of using the normal distribution to derive control limits of control charts in statistical quality control

x : Life time of a bulb

$\mu = 500$ $\sigma = 45$

$x \sim N(500, 45^2)$



I $Z = \frac{x-\mu}{\sigma}$ $P(x > 570) = P(Z > 1.56)$

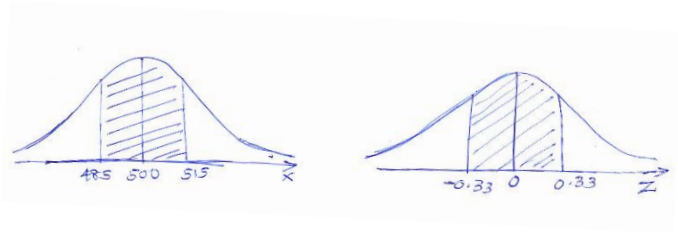
$Z = \frac{570-500}{45}$ $= 0.5 - 0.4406$

$Z = \frac{70}{45}$ $= 0.0594$

$Z = 1.56$ Percentage of bulbs $= 0.0594 \times 100\%$

$= 5.94\%$

II



$Z = \frac{x-\mu}{\sigma}$ $Z = \frac{x-\mu}{\sigma}$ $P(485 < x < 515) = P(-0.33<Z<0.33)$

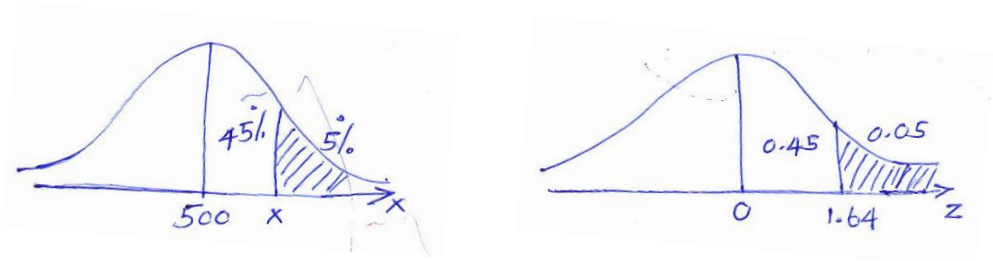
$Z_1 = \frac{485-500}{45}$ $Z_2 = \frac{515-500}{45}$ $= 0.1293 + 0.1293$

$Z_1 = \frac{-15}{45}$ $Z_2 = \frac{15}{45}$ $= 0.2586$

$Z_1 = -0.33$ $Z_2 = 0.33$ Percentage $= 0.2586 \times 100\%$

$= 25.86\%$

III



$Z = \frac{x-\mu}{\sigma}$ if $Z = 1.65$

$1.64 = \frac{x-500}{45}$ or $1.64 = \frac{x-500}{45}$

$x = 500 + 73.8$ $x = 500 + 74.25$

$\underline{x = 573.8}$ $\underline{x = 574.25}$

(08 marks)

7. (a) Describe the following methods of sampling, stating **two** advantages and **two** disadvantages of each sampling method.

(i) Stratified random sampling

(ii) Cluster sampling

(iii) Quota sampling

(iv) Systematic sampling (08 *marks*)

(b) Describe how the following population structures affect the expected precision of the systematic sampling.

(i) Population with the units in random order.

(ii) Population with linear trend.

(iii) Population with cyclic variations. (06 *marks*)

(c) (i) State the Central Limit Theorem.
Explain, why the Central Limit Theorem is considered as the most important theorem in Statistics.

(ii) A random sample of size 50 is taken from a Poisson distribution with mean $\lambda = 2$. Find the probability approximately that the sample mean will exceed 2.5. (06 *marks*)

07. **(a) (i) Stratified Random Sampling**

Once a population with N number of units is seperated in to L number of sub populations (Strata) as N_1, N_2, N_3..... the process of sampling by drawing a simple random sample, independently from each and every stratum is known as stratified random sampling

Advantages

- Population is very well represented by the sample.
- Once the Population is extremely skewed this method is more suitable for drawing a sample.
- Ability of drawing out a respresentative sample from a heterogeneous population
- Ability of making estimations seperately for each stratum, where necessary

Disadvantages

- Inability of sampling in the absence of a complete sampling frame.
- Being an expensive and time consuming sampling method
- Inbility to use once the strata are intersected by each other
- Dividing the population in to homogeneous strate in accordance with tha characteristics of populaton being inconvenient.

(ii) Cluster sampling

Once the population is naturally organized as clusters or can be identified as clusters, having selected one cluster or few clusters's sample members is known as cluster sampling method.

The population should be grouped as clusters as Variation between clusters being small while variations within the cluster (among cluster elements) being greater.

Advantages

- Being a more flexible sampling method
- Being a cost minimising method
- A sampling frame not being essential
- Being a more convenient samling method. once the population is organized naturally as clusters

Disadvantages

- Being a less accurate sampling method when compared to the other sampling methods.
- Being a more subjective sampling method (in grouping the population elements in to clusters etc....)

(iii) Quota sampling

After categorising the population accoding to few characteristics the process of choosing pare determined number of elementts from each category to the sample as desired by the invigilater / ennumerator is known as quota sampling

Advantages

- Being a cost minimising and a time saving method because data are collected from a pre determined group.
- Supervision and adminitration being convenient
- A sampling frame not being essential
- Selecting the sample being convenient unlike in probabilistic sampling methods.
- A better representative sample can be expected on the experience of the ennumerator.
- Better the sample being representative, greater the categorical aspects.

Disadvantages

- inability of estimating, because it's not followed by a probability base.
- Precision of the sample being at a lower level because of personal bias in selecting final sample elements.
- Results being less reliable
- Controlling over the field activities being inconvenient.

(iv) Systemetic Sampling

When a sample of size n should be drawn out from a population with size N, number the population elements as 1,2,3,.....N and seperate the population in to n number of intervals with width K as K = $\frac{N}{n}$, Then the first element from the first interval is drawn using simple random sampling method and since then every K^{th} element from each interval is drawn in to the sample and this method is called systemetic sampling.

Advantages

- Being a simple and easy sampling method.
- Being a cost minimising and time saving method

Disadvantages

- If the population elements are not arranged in a random order, perhaps a representative sample may not be derived.
- Posibility of the sample being biased when the population elements are in a periodic order.
- Once the sampling frame is incomplete inability of drawing the sample.

(08 marks)

(b)

(i) Once the population elements are arranged in a random order the precision of an estimator under systemetic sampling is the same as the precision of an estimator under simple random sampling method. The outcomes of simple random sampling can be applied mean under systemetic sampling.

(ii) Since the population is grouped as the fist K units, the second K units.... and an element from every K width interval is drawn in to the sample, once the population shows a linear trend, the precision of estimaters under systemetic sampling is greater than that under the simple random sampling.

(iii) Once the population shows a periodic pattern and if the sample elements fall on equal wave lengths the precision of estimators under systemetic sample will be poorer compared to that undr simple random sampling, because of the posibility of repeating the same information in the sample.

(06 marks)

(c)

(i) Central Limit Theorem

The acceptance that the sampling distribution of the sample mean $\bar{x}$ of large samples $(n \geq 30)$ drawn from any type of populaton with mean is and variance σ^2, is approximatolly distributed normally with mean μ and variance $\frac{\sigma^2}{n}$ is known as central Limit Theorem

The importance of Central Limit Theorem is that the ability of making dicisions and conclusions applying the normal distribution using large samples $(n \geq 30)$ when the variables of population cteristics do not fall normally, or the population distribution is unknown

(ii) $\lambda = 2$ n = 50

$\mu = \lambda$ $\sigma = \sqrt{\lambda}$

$\mu = 2$ $\sigma = \sqrt{2}$ n = 50

$\mu_{\bar{x}} = \mu = 2$ $\sigma_{\bar{x}} = \sigma/\sqrt{n}$

$= \frac{\sqrt{2}}{\sqrt{50}}$

$= \frac{1}{5}$

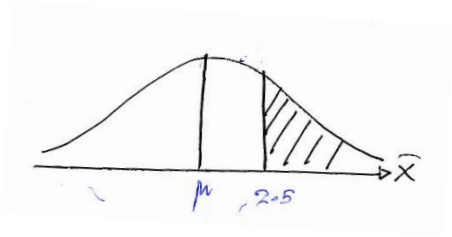
$= \underline{0.2}$

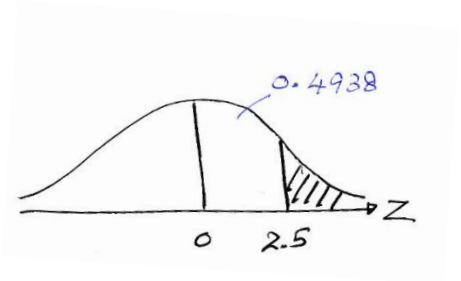
or

$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{2}{50} = \frac{1}{25}$

$\sigma_{\bar{x}} = \frac{1}{5}$

$\bar{X} \sim N [2, 1/25]$





$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

$= \frac{2.5 - 2}{0.2}$ $P(\bar{X} > 2.5) = P(Z > 2.5)$

$= 2.5$ $= 0.5 - 0.4938$

$= \underline{0.0062}$

(06 marks)

8. (a) Explain what is meant by Unbiasedness and Efficiency of a point estimator.

If $\{X_1, X_2, X_3\}$ is a random sample from a population with mean μ and variance σ^2, show that both estimators $\hat{\theta}_1 = \frac{X_1 + X_2 + X_3}{3}$ and $\hat{\theta}_2 = \frac{X_1 + 2X_2 + X_3}{4}$ are unbiased estimators for μ.

Out of these estimators, what is the most efficient estimator? (06 *marks*)

(b) Samples of two types of electric bulbs were tested for finding their life-time and the following values were observed.

Types of bulbs	No. of bulbs used	Sample mean (hours)	Standard Deviation
A	50	2015	80
B	70	2045	60

(i) Construct a 95% confidence interval for the difference of mean life-time between ***A*** and ***B***.

(ii) Using the confidence interval, test the hypothesis that mean life-time of bulbs ***A*** and ***B*** are equal. (06 *marks*)

(c) The prices of a certain commodity in three cities ***P***, ***Q*** and ***R*** are given in the table.

City		
P	***Q***	***R***
14	10	2
6	8	8
8	8	6
12	4	4

$\sum x_{ij}^2 = 804$

Test whether the average prices of the commodity in the three cities are significantly different at 5% level. (08 *marks*)

08. (a) Unbiasedness

If the expected value (mean) of a particular estimator is equal to the value of population parameter, that is known as an unbiased estimator for that parameter.

Efficiency

When there are few unbiased estimators for a particicular population parameter, the estimator with the least variance is considered as the efficient estimator for that parameter.

(02 marks)

$\hat{\theta}_1 = \frac{x_1+x_2+x_3}{3}$

$E(\hat{\theta}_1) = E\left[\frac{x_1+x_2+x_3}{3}\right]$

$E(\hat{\theta}_1) = \frac{1}{3}[E(x_1)+E(x_2)+E(x_3)]$

$= \frac{1}{3}[\mu+\mu+\mu]$

$= \frac{1}{3} x\, 3\mu$

$E(\hat{\theta}_1) = \mu$

$\therefore \hat{\theta}_1$ is an unbiased estimative for μ

$\hat{\theta}_2 = \left[\frac{x_1+2x_2+x_3}{4}\right]$

$E(\hat{\theta}_2) = E\left[\frac{x_1+2x_2+x_3}{4}\right]$

$E(\hat{\theta}_2) = \frac{1}{4}[E(x_1)+2E(x_2)+E(x_3)]$

$= \frac{1}{4}[\mu+2\mu+\mu]$

$= \frac{1}{4} x\, 4\mu$

$E(\hat{\theta}_2) = \mu$

$\therefore \hat{\theta}_2$ is an unbiased estimator for μ

$\text{Var}(\hat{\theta}_1) = Var\left[\frac{x_1+x_2+x_3}{3}\right]$

$= \frac{1}{9}[Var(x_1)+Var(x_2)+Var(x_3)]$

$= \frac{1}{9}[\sigma^2+\sigma^2+\sigma^2]$

$= \frac{3}{9}\sigma^2$

$Var(\hat{\theta}_1) = \frac{\sigma^2}{3}$

$\text{Var}(\hat{\theta}_2) = Var\left[\frac{x_1+2x_2+x_3}{4}\right]$

$\text{Var}(\hat{\theta}_2) = \frac{1}{16}[Var(x_1)+4Var(x_2)+Var(x_3)]$

$= \frac{1}{16}[\sigma^2+4\sigma^2+\sigma^2]$

$= \frac{6}{16}\sigma^2$

$\text{Var}(\hat{\theta}_2) = \frac{3}{8}\sigma^2$

$Var(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$

$\therefore \hat{\theta}_1$ is the most efficient estimator

(04 marks)

08. (b)

(i)

A		B	
n_A	= 50	n_B	= 70
$\bar{x}_A$	= 2015	$\bar{x}_B$	= 2045
S_A	= 80	S_B	= 60

$$\mu_A - \mu_B = (\bar{x}_A - \bar{x}_B) \pm z a_{/2} \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

$$=(2015 - 2045) \pm 1.96 \sqrt{\frac{80 \text{ x } 80}{50} + \frac{60 \text{ x } 60}{70}}$$

$$= - 30 \pm 1.96 \sqrt{128 + 51.43}$$

$$= - 30 \pm 1.96 \sqrt{179.43}$$

$$= - 30 \pm 1.96 \text{ x } 13.4$$

$$= - 30 \pm 26.26$$

$$= (-56.26, -3.74)$$

(04 marks)

(ii) $H_o : \mu_A = \mu_B$

$H_o : \mu_A \neq \mu_B$

If H_o is true $\mu_A = \mu_B$ then $\mu_A - \mu_B = 0$

Ho is rejected, since 0 is not contained in the above confidence interval.

There are no sufficient statistical evidence at 0.05 to accept that the lifetime of the bulbs A and B are equal

(02 marks)

(c) Buiding up of hypothesis

$H_o : \mu_P = \mu_Q = \mu_R$

H_1 Average prices of the item is significant in at least one of these cities

Test statistic

$T = \sum X_P + \sum X_Q + \sum X_R$

$= 40 + 30 + 20$

$= 90$

Error $= \frac{T^2}{N}$

$= \frac{90 \text{ X } 90}{12}$

$= 675$

SST $= \sum x_P^2 + \sum x_Q^2 + \sum x_R^2 - {T^2}/{N}$

$= 440 + 244 + 120 - 675$

$= 129$

$$SSC = \frac{(\sum x_P)^2}{n} + \frac{(\sum x_Q)^2}{n} + \frac{(\sum x_R)^2}{n} - T^2/N$$

$$= \frac{40 \text{ X } 40}{4} + \frac{30 \text{ X } 30}{4} + \frac{20 \text{ X } 20}{4} - 675$$

= 400 + 225 + 100 - 675

= 725 - 675

= 50

SSE = SST - SSC

= 129 - 50

= 79

Anova Table

Source of variation	Sum of squares	d : f	Mean sum of squares	F = Valces
Between Samples	SSC = 50	K - 1 = 3 - 1 = 2	MSC = $\frac{50}{2}$ = 25	F = $\frac{25}{8.78}$ = 2.85
Within Samples	SSE = 79	K (n-1) = 3 (4-1) = 9	MSE = $\frac{79}{9}$ = 8.78	

Test : $\alpha = 0.05$

Degrees of freedom of Numerator = k -1

= 3 - 1

= 2

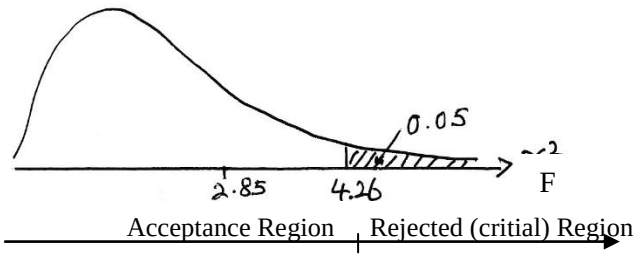
Degrees of freedom of Denomination= k(n-1)

= 3(4 - 1)

= 3 x 3

= 9

$F_{2, 9, 0.05}$ = 4.26



Decision :

Ho is not rejected since the value of test statistic $F = 2.85$ falls in acceptance region

Conclusion :

There are no sufficient statistical evidence at s% levch of significance to State that the price of this item in these three cities is significant.

(08 marks)

****