

ශ්‍රී ලංකා විභාග දෙපාර්තමේන්තුව
இலங்கைப் பரீட்சைத் திணைக்களம்

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

නව නිර්දේශය/ புதிய பாடத்திட்டம்

විෂය අංකය
பாட இலக்கம்

31 - E

විෂය
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/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 required to assure the validity of data.

Disadvantages

- uage being limited
- 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) Focus Group Interview Method

Collecting data discussing with a small group of knowledgeable 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 studying the facts in great details.
- being a more appropriate method to collect qualitative 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 opinions as responses related to a particular objective.
- because of the same fact, data analysis being complicated.
- arriving at conclusions 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 - researches. There are few major techniques 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 scattered individuals

Disadvantages

- Possibility of rate of responses being poor, when the computer literacy of respondents is inadequate.
- 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 catogriz 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 : °C and °F in measuring temperature

$$0^{\circ}\text{C} = 32^{\circ}\text{F}$$

$$0^{\circ}\text{F} = -17.7746^{\circ}\text{C}$$

(iv) **Ratio Scale**

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

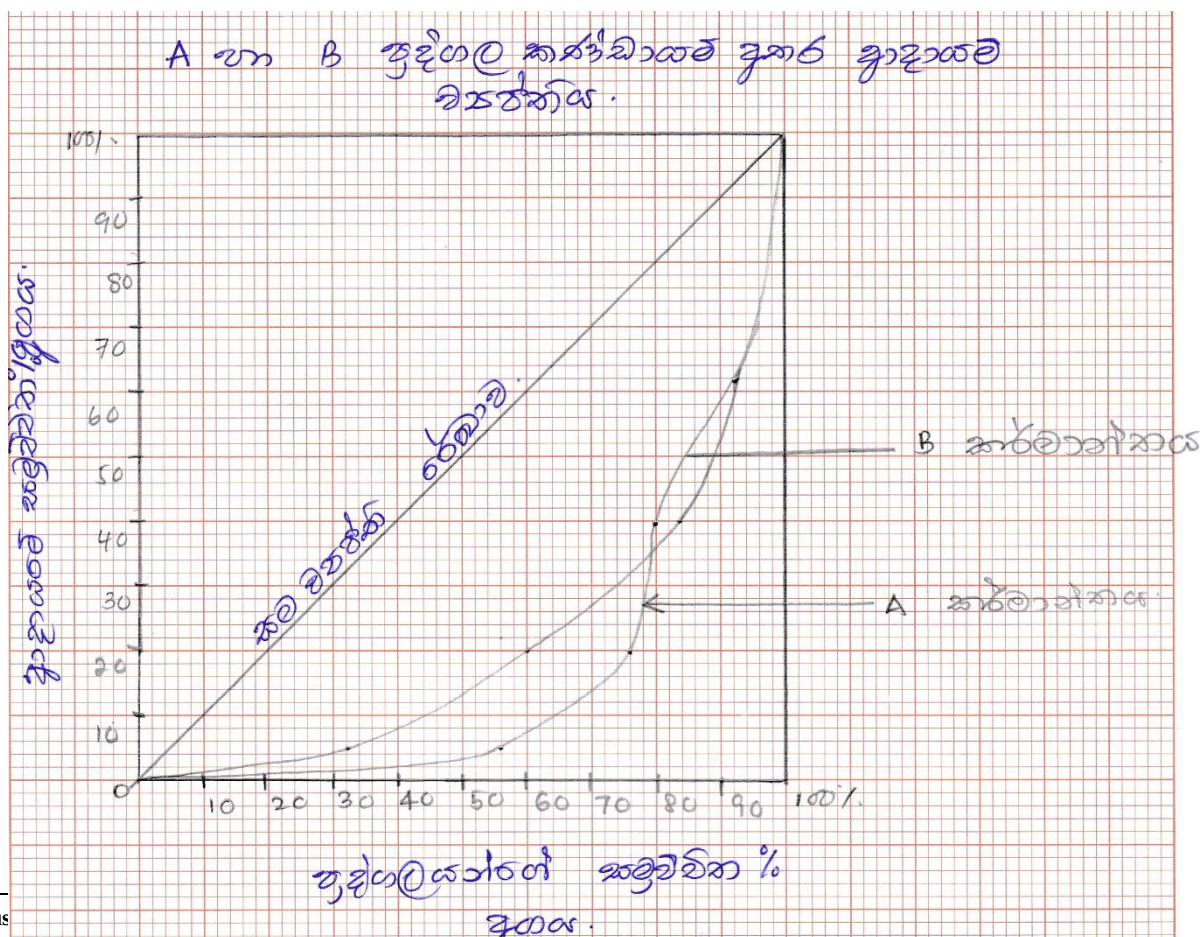
ex : Scores, age, / life time, height, mass, income, expenditure.....etc.

(04 marks)

(d) (i) **Lorenz Curve**

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

(d) (i)



- (ii) Since these two Lorenze curves are intersected by one another an accurate comparison can be made using the Gini 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 remaining 80% of the income goes to 40% of the people which is considerably a higher proportion compared to that of in group A

Accordingly a bigger variation 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

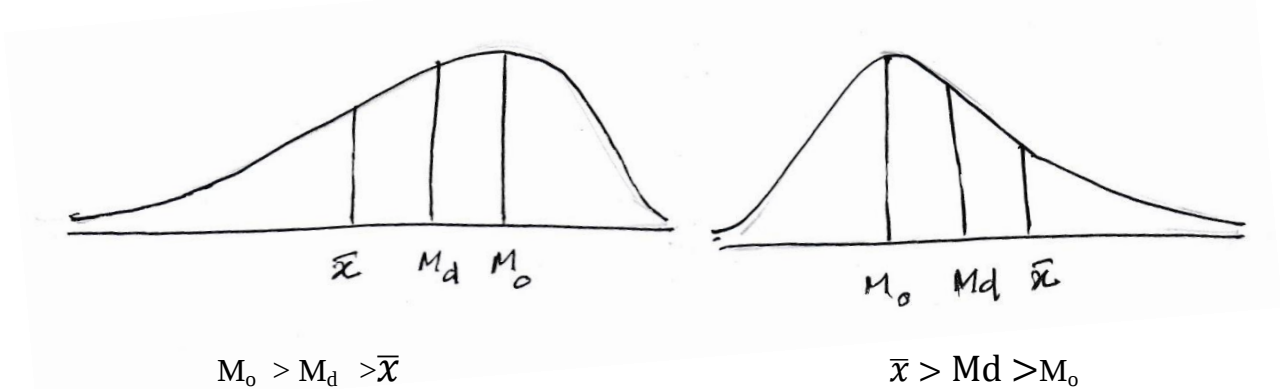
Calculate the mean, median, mode, standard deviation and Karl Pearson's coefficient of skewness and comment on the distribution.

(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 rather than towards left from the peak (maximum point) that is called a right skewed distribution and vice versa.

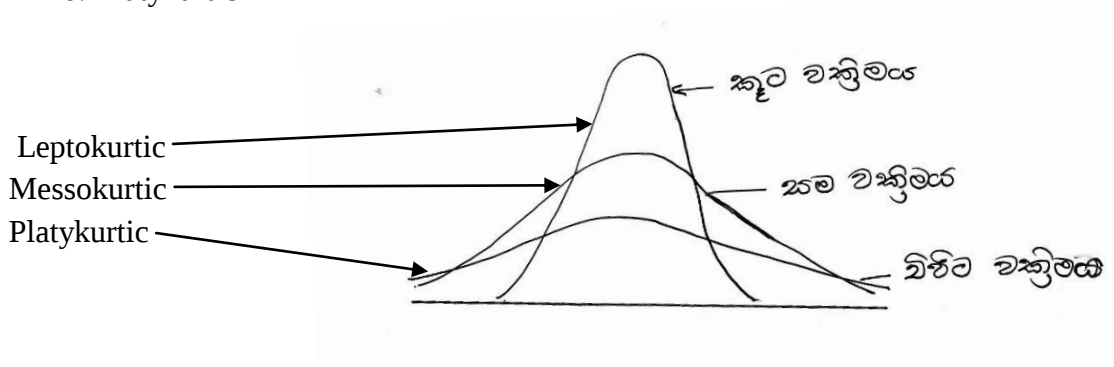


Kurtosis :

The nature of peakedness (curvature) of the frequency curve representing the distribution of a variable is known as 'Kurtosis' The type of kurtosis is explained 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

$$\begin{aligned}
 P_{10} &= L_1 + \left[\frac{\frac{10n}{100} - f_c}{f_{pm}} \right] C \\
 &= 19.5 + \left[\frac{10 - 08}{12} \right] 10 \\
 &= 19.5 + \frac{2}{12} \times 10 \\
 &= 19.5 + 1.67 \\
 &= \underline{\underline{21.17}}
 \end{aligned}$$

$$\begin{aligned}
 P_{50} &= L_1 + \left[\frac{\frac{50n}{100} - f_c}{f_{p50}} \right] C \\
 &= 39.5 + \left[\frac{50 - 40}{35} \right] 10 \\
 &= 39.5 + \left[\frac{10}{35} \right] 10 \\
 &= 39.5 + 2.857 \\
 &= \underline{\underline{42.357}}
 \end{aligned}$$

$$\begin{aligned}
 P_{90} &= L_1 + \left[\frac{\frac{90n}{100} - f_c}{f_{p90}} \right] C \\
 &= 49.5 + \left[\frac{90 - 75}{20} \right] 10 \\
 &= 49.5 + \left[\frac{15}{2} \right] \\
 &= 49.5 + 7.5 \\
 &= \underline{57.0}
 \end{aligned}$$

$$\begin{aligned}
 S_{Kp} &= \frac{P_{90} + P_{10} - 2P_{50}}{P_{90} - P_{10}} \\
 &= \frac{57.0 + 21.17 - 2 \times 42.36}{57.0 - 21.17} \\
 &= \frac{78.17 - 84.72}{35.83} \\
 &= \frac{-6.55}{35.83} \\
 &= \underline{-0.18}
 \end{aligned}$$

This is a negatively skewed distribution

(06 marks)

(b)	Male workers	Female workers
	CV = 55%	CV = 60%
	S = 22	S = 15
	$CV = \frac{S}{\bar{X}} \times 100$	$CV = \frac{S}{\bar{X}} \times 100$
	$55 = \frac{22}{\bar{X}} \times 100$	$60 = \frac{15}{\bar{X}} \times 100$
	$\bar{X} = \frac{22}{55} \times 100$	$\bar{X} = \frac{15}{60} \times 100$
	$\bar{X} = 40$	$\bar{X} = 25$
	$n_1 = 80$	$n_2 = 20$

$$\begin{aligned}
 \bar{X} &= \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2} \\
 &= \frac{80 \times 40 + 20 \times 25}{80 + 20} \\
 &= \frac{3200 + 500}{100}
 \end{aligned}$$

$$= \frac{3700}{100}$$

$$\bar{x} = 37$$

(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

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

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

$$= 63.5 + \frac{20}{30} \times 3$$

$$= \underline{\underline{65.5}}$$

Mode

$$M_0 = L_1 + \left(\frac{A_1}{A_1 + A_2} \right) C$$

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

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

$$= 63.5 + 1.5$$

$$= \underline{\underline{65}}$$

Standard Deviation

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

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

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

$$S = 3 \sqrt{1.7875}$$

$$S = \underline{\underline{4.01}}$$

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

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

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

(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.

Item	Base Year		Current Year	
	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 accordance with time or geographical location between two periods or two locations any other factor is known as an index number.

Laspar Price Index

Once 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

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

Pasche Price Index

Once 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

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

(03 marks)

(b)

Item	Base Year		Current Year		P ₀ Q ₀	P ₀ Q _n	P _n Q ₀	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

$$\begin{aligned}
 LP_{n/o} &= \frac{\sum p_n q_0}{\sum p_0 q_0} \times 100 \\
 &= \frac{1900}{1360} \times 100 \\
 &= \underline{\underline{139.7}}
 \end{aligned}$$

II. Paasche's price index

$$\begin{aligned}
 PP_{n/o} &= \frac{\sum p_n q_n}{\sum p_0 q_n} \times 100 \\
 &= \frac{2408}{1696} \times 100 \\
 &= 141.98 \\
 &= \underline{\underline{142}}
 \end{aligned}$$

III. Fisher's price index**IV. Marshal Edeworth price index**

$$\begin{aligned}
 FP_{n/o} &= \sqrt{LP_{n/o} \times PP_{n/o}} \\
 &= \sqrt{139.7 \times 141.9} \\
 &= 140.79 \\
 &= \underline{\underline{140.8}}
 \end{aligned}$$

$$\begin{aligned}
 M_E P_{n/o} &= \frac{\sum p_n (q_o + q_n)}{\sum p_o (q_o + q_n)} \times 100 \\
 &= \frac{4308}{3056} \times 100 \\
 &= 140.96 \\
 &= \underline{\underline{141}}
 \end{aligned}$$

(05 marks)

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

$$M_E P_{n/o} = \frac{\sum P_n \cdot \frac{1}{2}(q_o + q_n)}{\sum P_o \cdot \frac{1}{2}(q_o + q_n)} \times 100$$

$$\begin{aligned}
 M_E P_{n/o} \times M_E P_{o/n} &= \frac{\sum P_n (q_o + q_n)}{\sum P_o (q_o + q_n)} \times \frac{\sum P_o (q_o + q_n)}{\sum P_n (q_o + q_n)} \\
 &= \frac{4308}{3056} \times \frac{3056}{4308} \\
 &= 1
 \end{aligned}$$

The time reversal test is satisfied by Marshal Edgeworth Price index

$$\begin{aligned}
 M_E P_{n/o} \times M_E Q_{n/o} &= \frac{\sum P_n (q_o + q_n)}{\sum P_o (q_o + q_n)} \times \frac{\sum Q_n (p_o + p_n)}{\sum Q_o (p_o + p_n)} \\
 &= \frac{4308}{3056} \times \frac{4104}{3260} \\
 &= \underline{\underline{1.7746}}
 \end{aligned}$$

$$\begin{aligned}
 V_{n/o} &= \frac{\sum P_n q_n}{\sum P_o q_o} \times 100 \\
 &= \frac{2408}{1360} \times 100 \\
 &= \underline{\underline{1.7705}}
 \end{aligned}$$

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

(02 marks)

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

If $y_1, y_2, y_3 \dots y_n$ are a set of data collected for a variable denoted by Y for $t_1, t_2, t_3 \dots t_n$
Corresponding time interval $y_1, y_2, y_3 \dots y_n$ 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 :-

Oscillations found in long term trend in a time series prevailing 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 climate, regular festivals and Customs and Patterns in individual life styles etc....

(05 marks)

03. (d) (i) Annual trend equation
 $Y = 840 + 72x$ (origin 2005)
 Monthly trend equation
 $Y = \frac{840}{12} + \frac{72}{144}x$
- $Y = 70 + 0.5x$ (origin July 01th 2005)
- (ii) Sales for month of October 2011 ($X = 75.5$)
 $Y = 70 + (0.5 \times 75.5)$
 $= 70 + 37.75$
 $= 107.75$
 or
 Shift the base to January 2006
 $Y = 70 + 0.5[x + 6.5]$
 $= 70 + 0.5x + 3.25$

$$Y = 73.25 + 0.5 X \text{ (Origin January 15 2006)}$$

Sales for month of October 2011 ($X = 69$)

$$\begin{aligned} Y &= 73.25 + 0.5 \times 69 \\ &= 73.25 + 34.5 \\ &= \underline{\underline{107.75}} \end{aligned}$$

(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

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

- Calculate the correlation coefficient between test scores and sales, and state whether there is a relationship between them.
- Fit the regression line of Y on X by the method of least squares.
- Calculate the coefficient of determination and comment on your result.
- 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.

- Chance variation and Assignable variation
- 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.

Carpets 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)

$$\begin{aligned} \sum X &= 200 & \sum Y &= 400 \\ \sum X^2 &= 4120 & \sum Y^2 &= 16346 \\ \sum XY &= 8193 \end{aligned}$$

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

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

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

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

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

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

$$r = 0.9471 = 0.95$$

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

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

$$(ii) \quad \hat{\beta}_1 = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2} \quad \text{or} \quad \hat{\beta}_1 = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{n}}{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}$$

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

$$= \frac{81930 - 80000}{41200 - 40000}$$

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

$$\hat{\beta}_1 = \underline{\underline{1.608}}$$

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ &= 40 - 1.608 \times 20 \\ &= 40 - 32.16 \\ &= \underline{\underline{7.84}} \end{aligned}$$

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

(04 marks)

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

$$\text{Since } R^2 = r^2 \quad r = 0.9471$$

$$R^2 = (0.9471)^2$$

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

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)

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

$$30 = 7.8 + 1.61x$$

$$30 - 7.8 = 1.61x$$

$$22.2/1.61 = 13.78$$

$$x = 13.8$$

$$\underline{x = 14}$$

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 eliminated.

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 manufactured 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 plans.

(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 known as the 'C' chart

$$CL = \bar{C}$$

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

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

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 component of the relevant product as 'U'

$$CL = \bar{u}$$

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

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

(02 marks)

(c) $CL = \bar{C}$
 $\bar{C} = \frac{\sum C}{K}$ K = Number of samples

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

$$= 4$$

$$CL = \bar{C}$$

$$= 4$$

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

$$= 4 + 3\sqrt{4}$$

$$= 4 + 3 \times 2$$

$$UCL = 10$$

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

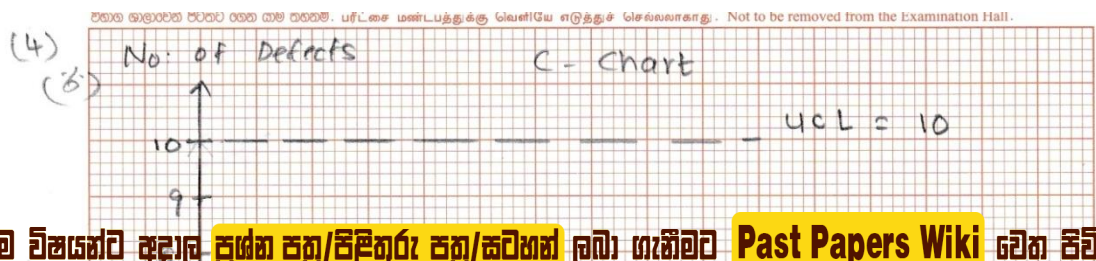
$$= 4 - 3 \times 2$$

$$= 4 - 6$$

$$= -2$$

$$= 0$$

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 exclusive 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 considered 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 frequency of occurring that event. When the number of trials is increased if this ratio approaches at a particular constant, that constant is considered 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 circumstances.
2. the probability value assigned for an event not being unique.

(04 marks)

$$(b) \quad P(A) = 1/2, \quad P(A \cup B) = 3/4, \quad P(B') = 5/8$$

$$\begin{aligned} P(B) &= 1 - P(B') \\ &= 1 - 5/8 \\ &= 3/8 \end{aligned}$$

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ 3/4 &= 1/2 + 3/8 - P(A \cap B) \\ P(A \cap B) &= 1/2 + 3/8 - 3/4 \\ &= \frac{4+3-6}{8} \\ &= 1/8 \end{aligned}$$

$$\begin{aligned} (i) \quad P(A^1 \cap B^1) &= P(A \cup B)^1 \\ &= 1 - P(A \cup B) \\ &= 1 - 3/4 \\ &= 1/4 \end{aligned}$$

$$\begin{aligned} P(A^1 \cup B^1) &= P(A \cap B)^1 \\ &= 1 - P(A \cap B) \\ &= 1 - 1/8 \\ &= 7/8 \end{aligned}$$

$$\begin{aligned} P(B \cap A^1) &= P(B) - P(A \cap B) \\ &= 3/8 - 1/8 \\ &= 2/8 \\ &= 1/4 \end{aligned}$$

(ii) If A and B events are indepent they should be related as $P(A) \cdot P(B) = P(A \cap B)$

$$P(A) \times P(B) = \frac{1}{2} \times \frac{3}{8}$$

$$= \frac{3}{16}$$

$$P(A \cap B) = \frac{1}{8}$$

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

$\therefore$ 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

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

$$= \frac{9!}{8! \cdot 1!} \times \frac{8!}{7! \cdot 1!}$$

$$\frac{17!}{15! \cdot 2!}$$

$$= \frac{9 \times 8}{17 \times 8} = \frac{72}{136}$$

$$= \frac{9}{17}$$

$$= 0.529$$

or

$$= \frac{1}{2} \left[\frac{{}^5C_1 \times {}^3C_1}{{}^8C_2} + \frac{{}^4C_1 \times {}^5C_1}{{}^9C_2} \right]$$

$$= \frac{1}{2} \left[\frac{5 \times 3}{28} + \frac{4 \times 5}{36} \right]$$

$$= \frac{1}{2} \left[\frac{15}{28} + \frac{20}{36} \right]$$

$$= \frac{275}{504}$$

$$= 0.546$$

(04 marks)

5. (d) The Law of Total Probability

Once $A_1, A_2, A_3, \dots, A_n$ are a set of mutually exclusive and collectively exhaustive events, while B is 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.

$$P(B) = P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + \dots + P(A_n) \cdot P(B/A_n)$$

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

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

Bayes' Theorem

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

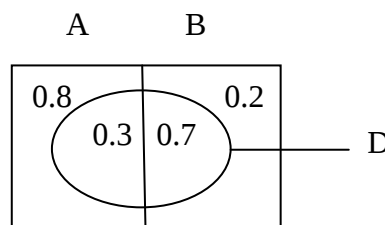
$$P(A_j / B) = \frac{P(A_j) \cdot P(B/A_j)}{\sum_{i=1}^n P(A_i) \cdot P(B/A_i)}$$

(03 marks)

A : Correct diagnosis of the disease

B : incorrect diagnosis of the disease

D : Patient being dead



$$P(A) = 0.8$$

$$P(B) = 0.2$$

$$P(D/A) = 0.3$$

$$P(D/B) = 0.7$$

$$P(D) = P(A) \cdot P(D/A) + P(B) \cdot P(D/B)$$

$$= 0.8 \times 0.3 + 0.2 \times 0.7$$

$$= 0.24 + 0.14$$

$$= 0.38$$

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

$$= \frac{0.8 \times 0.3}{0.38}$$

$$= 0.63$$

$$\begin{aligned}
 &= \frac{0.24}{0.38} \\
 &= \frac{24}{38} \\
 &= \frac{12}{19}
 \end{aligned}$$

(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.

- Find the probability that **no call is coming** when the operator is **not at the switchboard**.
- Find the probability that three or more calls are coming when the operator is **not at the switchboard**.
- 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,

- the percentage of bulbs with a life-time of at least 570 hours.
- the percentage of bulbs with life-time **between** 485 and 515 hours.
- 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 containing only two outcomes a success and failure, with probability of success (P) at each trial is a constant, the probability receiving X number of successes is given the probability mass function

$$P(X = x) = nC_x p^x q^{n-x} \quad \text{මගින් ලබාදේ.}$$

$$\text{මෙහි } x = 0, 1, 2, \dots, n$$

$$q = (1 - p)$$

Conditions

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

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

x : නමපලුර දින ගැණවසවක බිසකි

$$n = 10 \quad P = 0.2 \quad q = 0.8$$

$$x \sim B(10, 0.2)$$

$$P(X=x) = {}^n C_x p^x q^{n-x}$$

$$P(X=x) = {}^{10} C_x (0.2)^x (0.8)^{10-x} \quad x = 0, 1, 2, 3 \dots 10$$

$$\begin{aligned} \text{Probability of drawing} &= P(x=1) + P(x=2) \\ \text{the second sample} &= 0.2684 + 0.3020 \\ &= \underline{0.5704} \end{aligned}$$

(06 marks)

- (b) The probability mass function of a discrete random variable when that distributes on time or space is considered as the poisson distribution, 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 $x = 0, 1, 2, \dots$

and $e = 2.7183, \dots$

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.

$$\begin{aligned} \text{(i)} \quad \lambda &= \frac{1}{2}T \\ \lambda &= \frac{1}{2} \times 6 \\ \lambda &= 3 \end{aligned}$$

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

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

$$\text{I} \quad P(X=0) = \underline{0.0498}$$

$$\begin{aligned} \text{II} \quad P(X \geq 3) &= 1 - [P(x=0) + P(x=1) + P(x=2)] \\ &= 1 - [0.0498 + 0.1498 + 0.2240] \\ &= 1 - 0.4236 \\ &= \underline{0.5764} \end{aligned}$$

$$\begin{aligned}
 \text{III} \quad & \frac{e^{-\frac{1}{2}T} \left(\frac{1}{2}T\right)^0}{0!} = 0.9 \\
 & e^{-\frac{1}{2}T} = 0.9 \\
 & \log_{10} e^{-\frac{1}{2}T} = \log_{10} 0.9 \\
 & -\frac{1}{2}T \log_{10} e = \log_{10} 0.9 \\
 & -\frac{1}{2}T \times 0.4343 = -0.0458 \\
 & T = \frac{0.0458 \times 2}{0.4343} \\
 & = \frac{0.0916}{0.4343} \\
 & = 0.2109 \\
 \text{Number of} & = 0.2109 \times 60 \\
 \text{seconds} & = 12.654 \\
 & = \underline{\underline{12}}
 \end{aligned}$$

(06 marks)

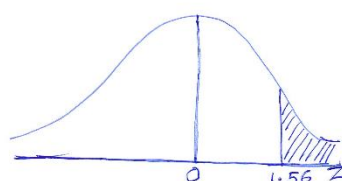
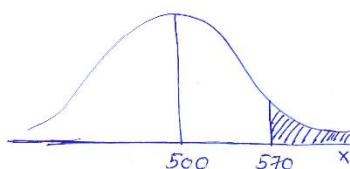
(c) **Uses of Normal distribution**

1. Because most of the continuous 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 approximation 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

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

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



I

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

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

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

$$Z = 1.56$$

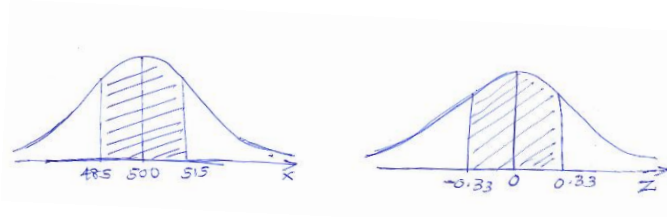
$$P(x > 570) = P(Z > 1.56)$$

$$= 0.5 - 0.4406$$

$$= 0.0594$$

Percentage of bulbs = $0.0594 \times 100\%$
 $= 5.94\%$

II



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

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

$$P(485 < x < 515) = P(-0.33 < Z < 0.33)$$

$$Z_1 = \frac{485 - 500}{45}$$

$$Z_2 = \frac{515 - 500}{45}$$

$$= 0.1293 + 0.1293$$

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

$$Z_2 = \frac{15}{45}$$

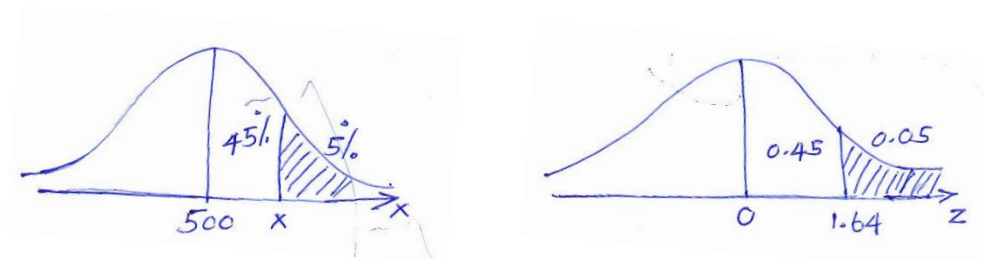
$$= 0.2586$$

$$Z_1 = -0.33$$

$$Z_2 = 0.33$$

Percentage = $0.2586 \times 100\%$
 $= 25.86\%$

III



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

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

$$x = 500 + 73.8$$

$$\underline{x = 573.8}$$

or

if $Z = 1.65$

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

$$x = 500 + 74.25$$

$$\underline{x = 574.25}$$

(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 separated in to L number of sub populations (Strata) as $N_1, N_2, N_3, \dots$ 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 representative sample from a heterogeneous population
- Ability of making estimations separately 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
- Inability to use once the strata are intersected by each other
- Dividing the population in to homogeneous strata in accordance with the characteristics of population 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 sampling 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 according to few characteristics the process of choosing pre determined number of elements from each category to the sample as desired by the invigilator / enumerator 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 administration 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 enumerator.
- 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) Systematic 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 separate 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 systematic 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.
- Possibility 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 systematic 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 systematic sampling.
- (ii) Since the population is grouped as the first 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 estimators under systematic 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 systematic sample will be poorer compared to that under simple random sampling, because of the possibility 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 population with mean μ and variance σ^2 , is approximately 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 decisions and conclusions applying the normal distribution using large samples ($n \geq 30$) when the variables of population characteristics do not fall normally, or the population distribution is unknown

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

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

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

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

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

$$= \frac{1}{5}$$

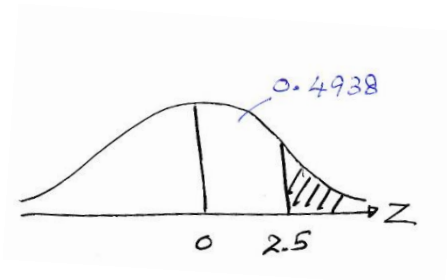
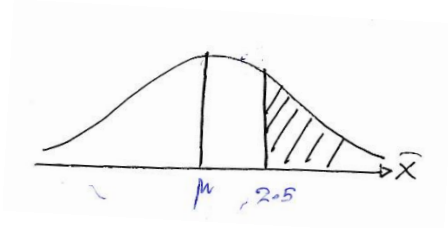
$$= \underline{0.2}$$

or

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

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

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



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

$$= \frac{2.5 - 2}{0.2}$$

$$= 2.5$$

$$P(\bar{X} > 2.5) = P(Z > 2.5)$$

$$= 0.5 - 0.4938$$

$$= \underline{\underline{0.0062}}$$

(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 particular population parameter, the estimator with the least variance is considered as the efficient estimator for that parameter.

(02 marks)

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

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

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

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

$$= \frac{1}{3} \times 3\mu$$

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

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

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

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

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

$$= \frac{1}{4} \times 4\mu$$

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

$\therefore \hat{\theta}_1$ is an unbiased estimative for μ $\therefore \hat{\theta}_2$ is an unbiased estimator for μ

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

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

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

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

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

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

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

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

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

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

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

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

(04 marks)

08. (b)

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



$$\begin{aligned}
 \mu_A - \mu_B &= (\bar{x}_A - \bar{x}_B) \pm z_{\alpha/2} \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}} \\
 &= (2015 - 2045) \pm 1.96 \sqrt{\frac{80 \times 80}{50} + \frac{60 \times 60}{70}} \\
 &= -30 \pm 1.96 \sqrt{128 + 51.43} \\
 &= -30 \pm 1.96 \sqrt{179.43} \\
 &= -30 \pm 1.96 \times 13.4 \\
 &= -30 \pm 26.26 \\
 &= \underline{\underline{(-56.26, -3.74)}}
 \end{aligned}$$

(04 marks)

(ii) $H_0 : \mu_A = \mu_B$
 $H_0 : \mu_A \neq \mu_B$

If H_0 is true $\mu_A = \mu_B$ then $\mu_A - \mu_B = 0$

H_0 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) Building up of hypothesis

$$H_0 : \mu_P = \mu_Q = \mu_R$$

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

Test statistic

$$\begin{aligned}
 T &= \sum X_P + \sum X_Q + \sum X_R \\
 &= 40 + 30 + 20 \\
 &= 90
 \end{aligned}$$

$$\begin{aligned}
 \text{Error} &= \frac{T^2}{N} \\
 &= \frac{90 \times 90}{12} \\
 &= 675
 \end{aligned}$$

$$\begin{aligned}
 \text{SST} &= \sum x_P^2 + \sum x_Q^2 + \sum x_R^2 - T^2/N \\
 &= 440 + 244 + 120 - 675 \\
 &= \underline{\underline{129}}
 \end{aligned}$$

$$\begin{aligned}
 SSC &= \frac{(\sum x_P)^2}{n} + \frac{(\sum x_Q)^2}{n} + \frac{(\sum x_R)^2}{n} - T^2/N \\
 &= \frac{40 \times 40}{4} + \frac{30 \times 30}{4} + \frac{20 \times 20}{4} - 675 \\
 &= 400 + 225 + 100 - 675 \\
 &= 725 - 675 \\
 &= 50 \\
 SSE &= SST - SSC \\
 &= 129 - 50 \\
 &= 79
 \end{aligned}$$

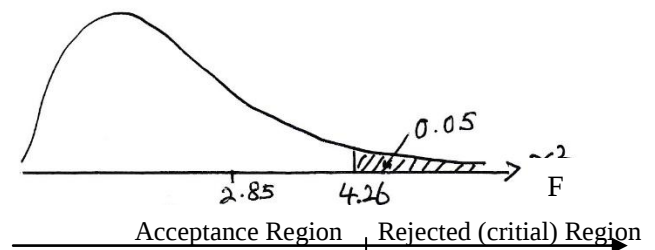
Anova Table

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



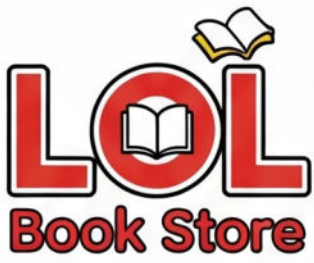
Decision :

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.

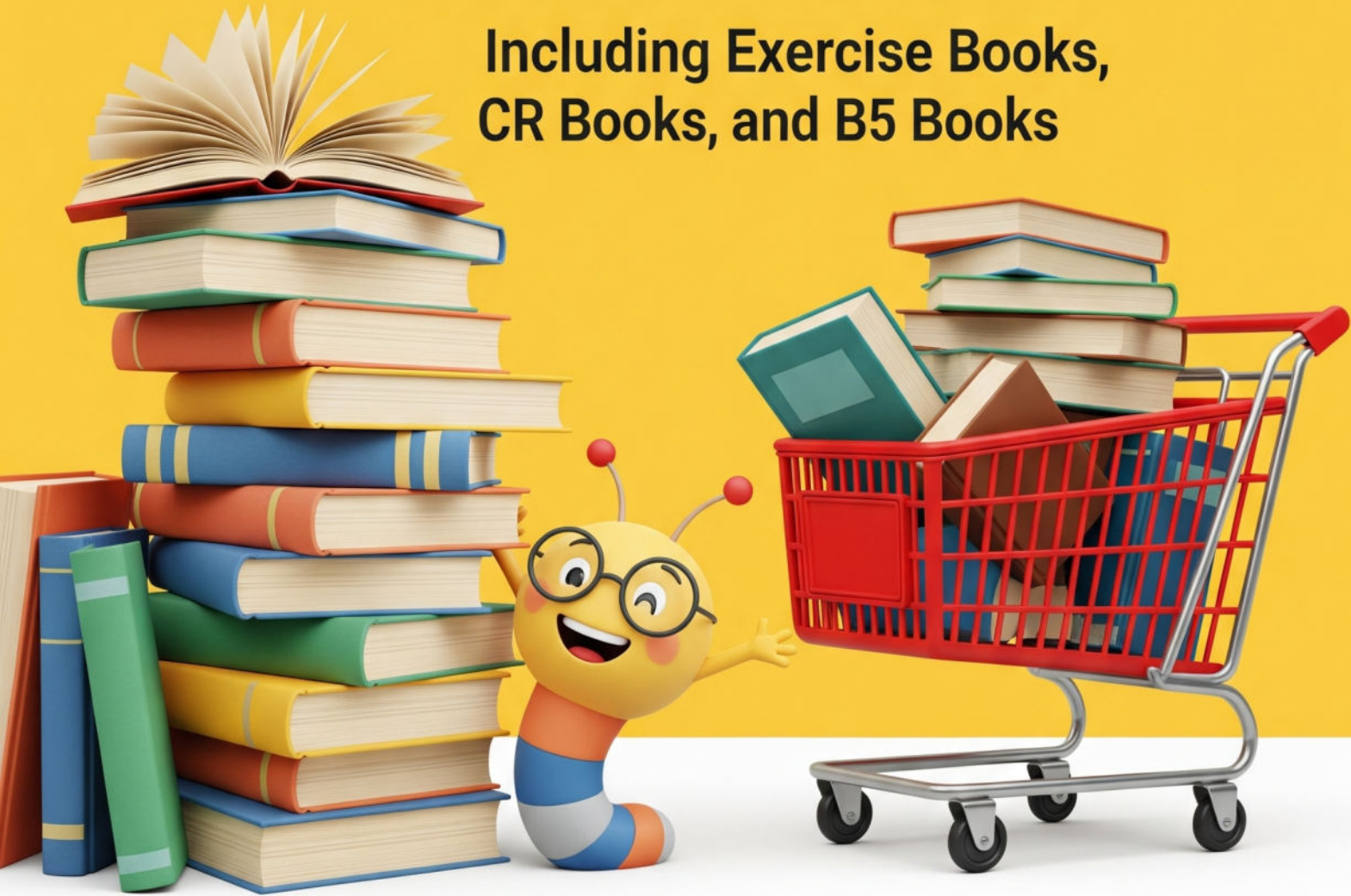
(08 marks)



වාර්තා මින කරන පොත් චිකිතාර
මිඩර් කරලා දුන් ගෙදරටම ගෙන්වා ගන්න

25% OFF ALL BOOKS

Including Exercise Books,
CR Books, and B5 Books



මිනෙම අභ්‍යාස පොතකට 25%ක විශේෂ වට්ටමක්

ORDER BOOKS FROM **LOL BOOKS STORE**



0372060110 WWW.LOL.LK



071 777 4440 (WhatsApp)

Cash On Delivery Available

ORDER A/L $a+b^2$ TERM TEST PAPERS, SHORT NOTES, WORKBOOKS & REVISION BOOKS

SINHALA, ENGLISH & TAMIL MEDIUM



LOL BOOK STORE

CASH ON DELIVERY AND KOKO PAYMENT AVAILABLE

0717774440 (WHATSAPP)

WWW.LOL.LK