G.C.E. (A/L)

Business statistics - I

Channel NIE

Marking scheme of Rehearsal Test for G.C.E. Advance Level Examination

1) – **(4)**	11) – **(1)**	21)– **(3)**	31) – **(5)**	41) – **(2)**
2) **– (2)**	12) – **(3)**	22)– **(2)**	32) – **(2)**	42) – **(1)**
3) **– (4)**	13) – **(4)**	23) – **(1)**	33) – **(5)**	43) – **(4)**
4) **– (3)**	14) – **(5)**	24) – **(3)**	34) – **(3)**	44) – **(3)**
5) – **(3)**	15) – **(4)**	25) – **(1)**	35) – **(2)**	45) – **(1)**
6) – (**1)**	16) – **(2)**	26) – (**1)**	36) – (2)	46) – **(1)**
7) – **(3)**	17) – **(1)**	27) – **(5)**	37) – **(5)**	47) – (4)
8) – (3)	18) – **(4)**	28) – **(5)**	38) – **(3)**	48) – **(2)**
9) – **(5)**	19) – **(3)**	29) – **(4)**	39) – **(2)**	49) – **(4)**
10) –**(2)**	20) – **(5)**	30) – (5)	40) – **(5)**	50) – **(2)**

Marking Scheme

1) (a)

(i) **Secondary data**: These data can be easily collected from the annual report of the Central bank of Sri Lanka.Such a study can not be done by an individual or a single firm using primary data.

(ii) **Primary data**: This is current issue faced by farmers in agriculture field.sufficient secondary data regarding that issue may not be available.Furthe discussing with farmers individually would be more convenient to understand the reality of the issues faced by them and to propose more appropriate possible solutions.

(iii)**Secondary data:** Collecting primary data is impossible in a situation where an epidemic is being rapidly spread out.It would be more convenient to collect secondary data,because all the detail related to on-line marketing are computerized from A to Z(from order to the sale).

(b(

Leading (bias) questions

If the question itself focuses the respondents towards a particular answer, such questions are called leading questions.

Ex: You usually prefer to eat bread for the breakfast, don’t you?

Ambiguious questions

If two or more meanings are given by a particular question is an ambigious question.

Ex: Where are you?

When this question is heard the respondent may answer supposing the place where he is at the moment is questioned or supposing that his residence/village is being questioned or related to a certain activity consists of a long process,supposing that the level he/she has already approached are being planned.

)c(

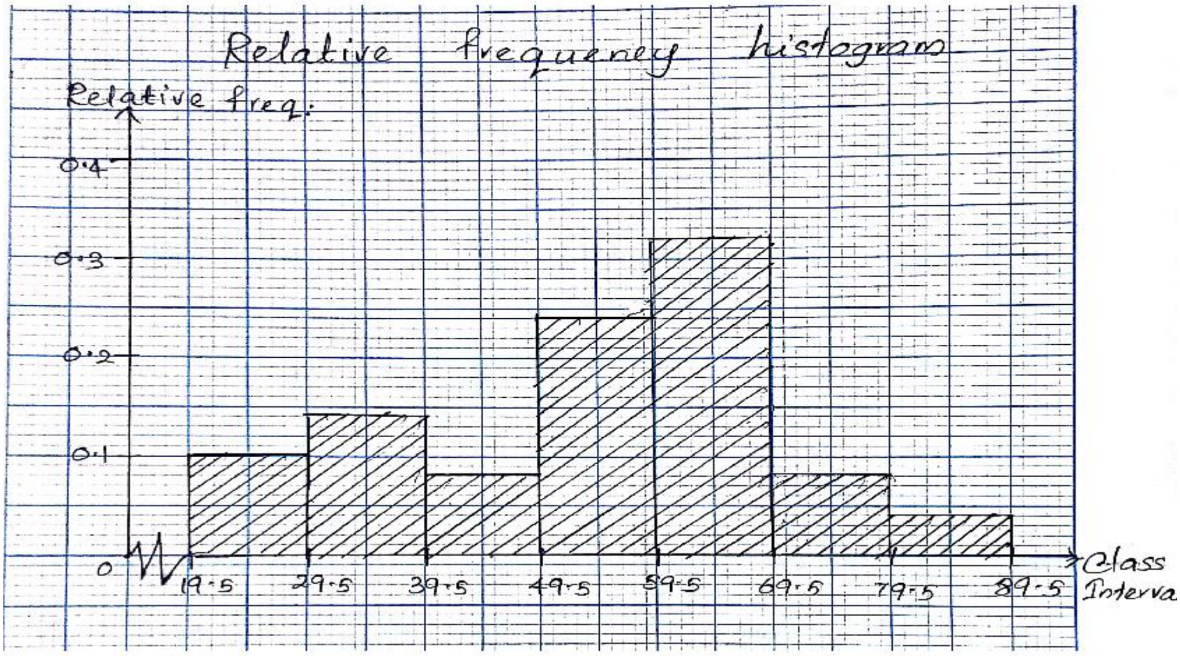
(i) **False**.Editing is making adjustments for completeness, accuracy, consistency and relevancy of the responses found in completed questionnaire copies by respondents' before the data being analysed.

(ii) **True**. This is an example for a possible missuse of statistical information through making conclusions with out using a sufficient sample.

(iii)**False.** These data belongs to rank scale.A true zero is available in ratio scale.

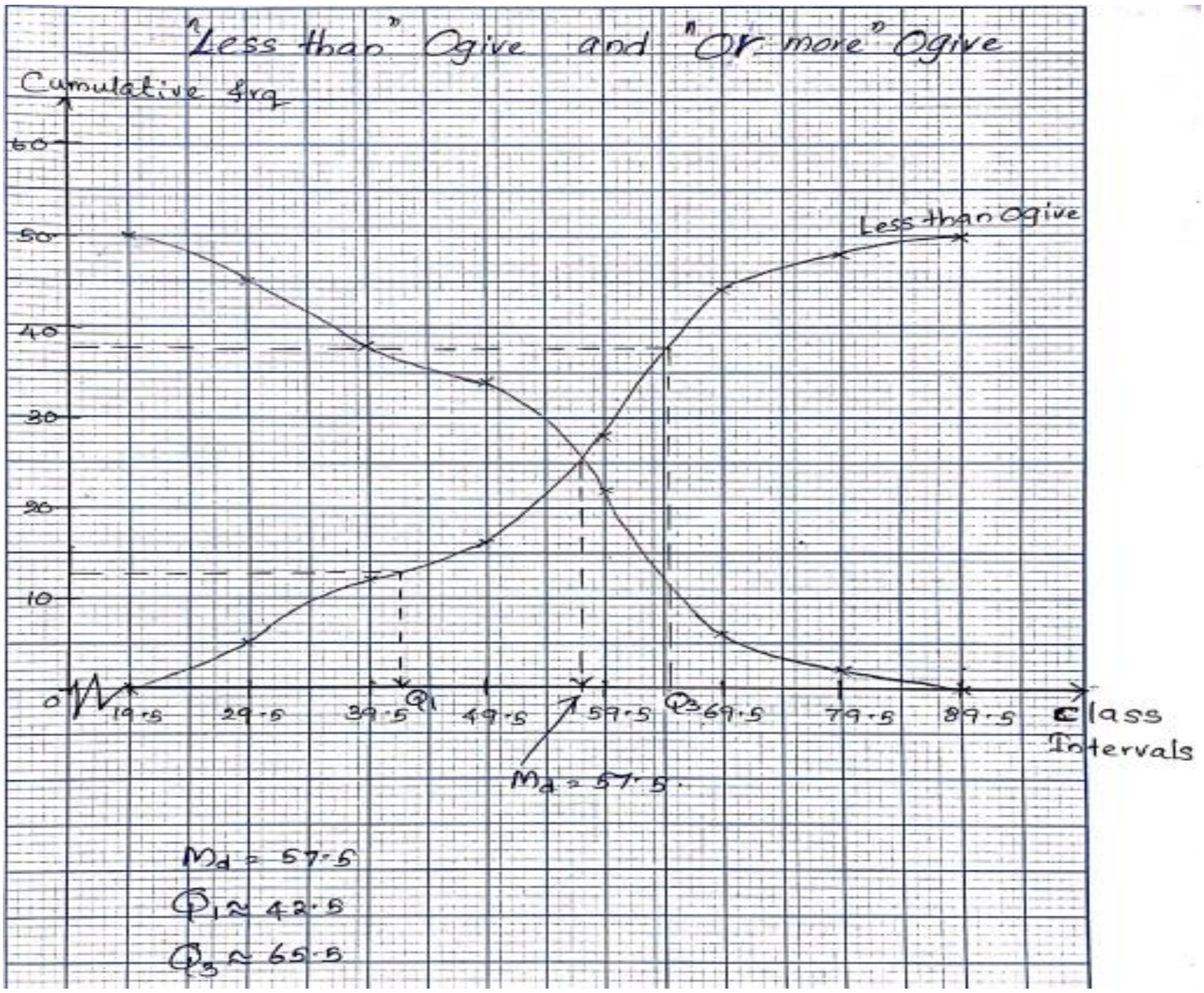
)d(

Class Intervals	**Tally marks**	**Frequency f_i**	**Relative Frequency $f_i \div 50$**
19.5 – 29.5	~~////~~	05	0.10
29.5 – 39.5	~~////~~ //	07	0.14
39.5 - 49.5	////	04	0.08
49.5 – 59.5	~~////~~ ~~////~~ //	12	0.24
59.5 – 69.5	~~////~~ ~~////~~ ~~////~~ /	16	0.32
69.5 – 79.5	////	04	0.08
79.5 – 89.5	//	02	0.04
Total		50	1.00



This seems to be a negatively skewed distribution.

"Less than" Cumulative frquency distribution		"Or more "Cumulative frquency distribution	
Upper class boundary	Less than cumulative frquency	Lower class boundary	Or more cumulative frquency
Less than 19.5	00	19.5 or more	50
Less than 29.5	05	29.5 or more	45
Less than 39.5	12	39.5 or more	38
Less than 49.5	16	49.5 or more	34
Less than 59.5	28	59.5 or more	22
Less than 69.5	44	69.5 or more	06
Less than 79.5	48	79.5 or more	02
Less than 89.5	50	89.5 or more	00



2)

)a(As expressed by A ,it's true that the simple Arithmetic Mean satisfies algebraic property.Since $\bar{x} = \frac{\sum x}{n}$,it can be derived as $\sum x = n\bar{x}$ and $n = \frac{\sum x}{\bar{x}}$. Further there can be one and only one mean available for any data set. (01 mark(

Further it's true that mode and median do not satisfy the algebraic property, because sum of the data or number of data can not be found using those two measures.Any distribution may have one mode or more or may not have any mode.Therefore it's also true that mode is not an identical measure, but median is an identical measure, because one and only one median can be calculated for any distrinbution .So that the cause that median is not identical is false (01 mark(

Statement of B is not true. Algebric formulae are used to calculate mode and median only for grouped data and those formulae also have been developed only on various assumptions.Using mode or median sum of the total observations can not be derived. Therefore median and mode do notsatisfy the algebraic property, any way B's opinion that both median and mean are identical is acceptable.

)b(

(i)

Class intervals	Mid value x_i	Frequency	u_i	$f_i u_i$
36-40	38	08	-4	-32
41-45	43	18	-3	-54
46-50	48	30	-2	-60
51-55	53	40	-1	-40
56-60	58	32	0	0
61-65	63	26	1	26
66-70	68	20	2	40
71-75	73	16	3	48
76-80	78	10	4	40
				-186 +154 $\sum fi\, ui = -32$

Mean salary of an employee on proposal **A**

$$\bar{x} = A + \left(\frac{\sum fi\, ui}{\sum fi}\right) C$$

$$= 58 + \left(\frac{-32}{200}\right) 5$$

$= 58 - 0.8$

$= 57.2 \times 1000$

= <u>Rs. 57 200</u>

$$\left.\begin{array}{l}\text{Additional allowance to be added to the} \\ \text{Basic salary of employee on proposal } \mathbf{A}\end{array}\right\} = \text{Rs } (57\,000/\,4)$$

$$= \underline{\text{Rs. } 14\,300}$$

Median monthly salary of an employee on proposal **B**

Meadian containing class interval is the interval that contains the first cumulative frequency value exceeding half of the total number of observations that is 100.accordingly median is included in

56-60 class interval.

$$M_d = L_1 + \left[\frac{\frac{n}{2} - Fc}{fm}\right]C$$

$$= 55.5 + \left[\frac{\frac{200}{2} - 96}{32}\right]5$$

$$= 55.5 + \left[\frac{100 - 96}{32}\right]5$$

$$= 55.5 + \left[\frac{4}{32}\right]5$$

$$= 55.5 + 0.625$$

$$= 56.125 \times 1000 = \underline{\text{Rs . } 56\,125}$$

$$\left.\begin{array}{l}\text{Additional amount to be added to the monthly basic salary} \\ \text{Of an employee, on proposal } \mathbf{B}\end{array}\right\} = 56\,125 \times \frac{20}{100}$$

$$= \underline{\text{Rs .} 11\,225}$$

Since the amount of Rs 11 225,which is 20% of the median salary of an employee is less than the amount of Rs 14 300/=,which is 25% of the mean salary of an employee,implementation of the proposal **B** is more appropriate by means of controlling the salary expenses of the firm.

(ii) Being an independent trade union activist a special attention should be given to the profitability of the firm as well as the employee welfare, If the maximum employee satisfaction to be assured, the proposal **A** should be recommended. Any way, after deciding of the real salary of the employee is evaluated using a suitable price index,it would be better to recommend a salary increment that will be sufficient to assure the existed living standard for the employees,or to enhance the living standard up to a certain extent from the existed level.

(iii)A constant amount is added to the existing monthly salary of each and every employee through implementation of either proposal, so that the existing variance will not be changed.

(iv)Q_1 containing class $= \frac{1}{4} \times 200^{th}$ value containing class

$= 50^{th}$ value containing class

$= 46 - 50$

Q_3 containing class $= \frac{3}{4} \times 200^{th}$ value containing class

$= 150^{th}$ value containing class

$= 61 - 65$

$$Q_1 = L_1 + \left[\frac{\frac{n}{4} - Fc}{fq1}\right]C$$

$$= 45.5 + \left[\frac{50-26}{30}\right]5$$

$$= 45.5 + \left[\frac{24 \times 5}{30}\right]$$

$$= 45.5 + 4$$

$$= 49.5$$

$$Q_3 = L_1 + \left[\frac{\frac{3n}{4} - Fc}{fq3}\right]C$$

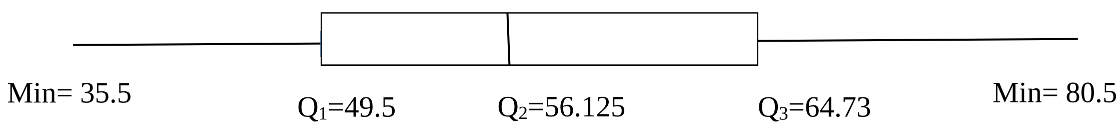
$$= 60.5 + \left[\frac{150-128}{26}\right]5$$

$$= 60.5 + \left[\frac{22 \times 5}{30}\right]$$

$$= 60.5 + 4.23$$

$$= 64.73$$

$Q_2 = M_d = 56.125$



)c(

(i) Coefficient of variance is a measure of relative disperson which is free from units, since the absolute dispersion pof the distribution is comparatively assesses relative to the central tendency of that distribution.Therefore coefficient of variation is an appropriate measure to be applied for comparission of variation in different measuring units as well as the variation of distribution with same measuring unit, but the means are significant.

(ii) Let's compute the coefficient of variation of each data set separately.

$$CV_a = \frac{\text{Rs. } 5\,400}{\text{Rs. } 54\,000} \times 100\% = 10\ \%$$

$$CV_b = \frac{3.5 \text{ km}}{18.5 \text{ km}} \times 100\% = 18.19\ \%$$

$$CV_c = \frac{8\ l}{40.8\ l} \times 100\% = 19.6\ \%$$

Hence, fuel consumption of the firm is highly deviated, while employee salary distribution is more consistent(less variated)

3))a(

(i) Normally wage ,which is paid for employees after assessing the task performed(done) in terms of the existing current monetary unit (rupees in Sri lanka) is known as money salary.Purchasing power of money salary is the real salary.In other words, the value received through deflating the money salary by an appropriate price index..

Ex : The money salary of an employee is Rs 60 000 and the price index for the period is 120. Then,

His money salary = Rs. 60 000

$$\text{Real salary} = \frac{\text{Rs } 60\,000}{120} \times 100$$

$$= \text{Rs } 50\,000$$

(ii) Although a constant amount of Rs 5 000 is added to the monthly salary of every employee, their real salary will drop down, if the price level of all the goods and services increases with a significant rate.

(iii)Salary is the cost of labour, which is one of the main production factors.the production cost of goods and services increases, through increasing of the labour cost. Hence, the general price level of all goods and services sold in the market will increae, so that the inflation may be uncontrollable.As a result the living condition of the people will further drop down.Therefore the most intelligent measure to be taken in an inflationary situation is to offer concessions for (production subsidies) producers and paving the way to decrease the prices of goods and services.

)b(Computing the percentage expenditure of each group within the total expenditure in January 2022

Expenditure Element	Percentages of expenses
Food and beverages	$\frac{8\,000}{40\,000} \times 100 = 20\%$
Clothing	$\frac{2\,000}{40\,000} \times 100 = 5\%$
Travelling	$\frac{10\,000}{40\,000} \times 100 = 25\%$
Health care	$\frac{3\,000}{40\,000} \times 100 = 7.5\%$
Education	$\frac{8\,000}{40\,000} \times 100 = 20\%$
Social affairs	$\frac{5\,000}{40\,000} \times 100 = 12.5\%$
Entertainment	$\frac{4\,000}{40\,000} \times 100 = 10\%$
	100%

Compute the Price Relatives for each commodity = $\frac{Pn}{P0} \times 100$

Lunch packet 1 = $\frac{280}{160} \times 100 = 175.0\%$

Cash of Soap = $\frac{120}{80} \times 100 = 150.0\%$

Minimum bus fare = $\frac{32}{14} \times 100 = 228.6\%$

A single doctor visit = $\frac{600}{400} \times 100 = 150.0\%$

A pack of stationary items = $\frac{900}{500} \times 100 = 180.0\%$

Cost of wedding feast = $\frac{12\,000}{8\,000} \times 100 = 150.0\%$

Admission fee to a theatre = $\frac{280}{150} \times 100 = 186.7\%$

Multiplying each Price Relative by the real weight.

If the percentage of expenses incurred in January 2022 for each item is considered as the weight of that item representing the relative importance, the contribution of each item to the cost of living index can be computed as the product of those weights and the simple price relatives

Group of items	Weight)%(	Simple price relatives	Contribution to cost of living index
Food and beverages	20.0	175.0	3 500.0
Clothing	5.0	150.0	750.0
Travelling	25.0	228.6	5 715.0
Health care	7.5	150.0	1 125.0
Education	20.0	180.0	3 600.0
Social affairs	12.5	150.0	1 875.0
Entertainment	10.0	186.7	1 867.0
Total amount			**18 432.0**

Cost of living index = 18 432.0 ÷ 100

= 184.32

(iv) Basic salary in January 2022 = Rs. 40 000

Cost of living index June/Jan 2022 = 184.32

$$\text{Real salary June 2022} = \frac{\text{Money salary June 2022}}{\text{Cost of living index}} \times 100$$

$$\text{Rs } 40\,000 = \frac{\text{Money salary June 2022}}{184.32} \times 100$$

$$\text{Money salary to be paid in June 2022} = \frac{40\,000 \times 184.32}{100}$$

= Rs 73 728.00

The amount to be added to the basic salary in June 2022 } = Rs 73 728.00 – 40 000.00

= Rs 33 728.00

)d((i) Computing the seasonal indices for the purposes of separating the influence of seasonal component,among the components on which the total value of a time series variable is determined is known as estimation of seasonal indices.

- Though estimation of seasonal indices appropriates sales estimates can be prepared for each season and it leads to maintain a proper store keeping mechanism that minimising the storage cost aas well as the wastage.

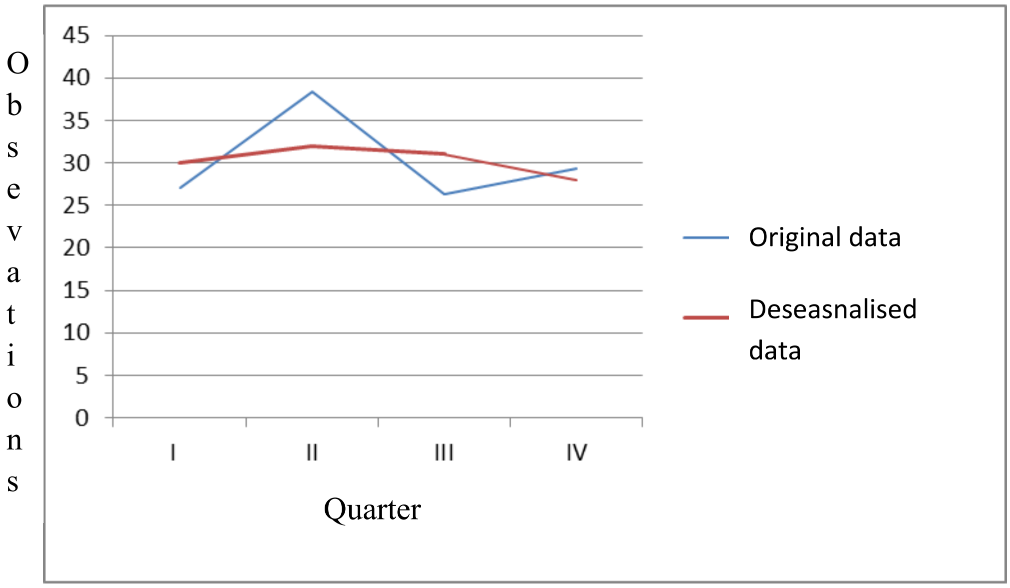
- Goodwill of the firm also can be maintained by providing with the consumer demand with no any shortage.
- At the end of certain seasons, unnecessary costs of advertising can be ruled out by tolerating the decreasing demand as usual.

(ii)

$$\text{Deseasonalised data} = \frac{\text{Original data}}{\text{Seasonal Index}} \times 100$$

$$\text{Seasonal index} = \frac{\text{Original data}}{\text{Dseasonalised data}} \times 100$$

Quarters	Seasonal index
I	$\frac{27.00}{30} \times 100 = 90$
II	$\frac{38.40}{32} \times 100 = 120$
III	$\frac{26.35}{31} \times 100 = 85$
IV	$\frac{29.40}{28} \times 100 = 105$



)d(

Year	X	Y	XY	X^2
2012	-9	14	-126	81
2013	-7	16	-112	49
2014	-5	20	-100	25
2015	-3	19	-57	9
2016	-1	22	-22	1
2017	1 0	25	25	1
2018	3	24	72	9
2019	5	29	145	25
2020	7	32	224	49
2021	9	30	270	81
		231	736-417=319	330

(i) Trend equation on least squre method $\hat{Y} = \hat{a} + \hat{b}X$

$$\hat{b} = \frac{\sum XY}{\sum X^2} = \frac{319}{330} = 0.97$$

$$\hat{a} = \bar{Y} = \frac{231}{10} = 23.1$$

$$\underline{\underline{\hat{Y} = 23.1 + 0.97X}}$$

(ii) Trend value for the year 2014

$$\hat{Y} = 23.1 + 0.97 \times (-5)$$

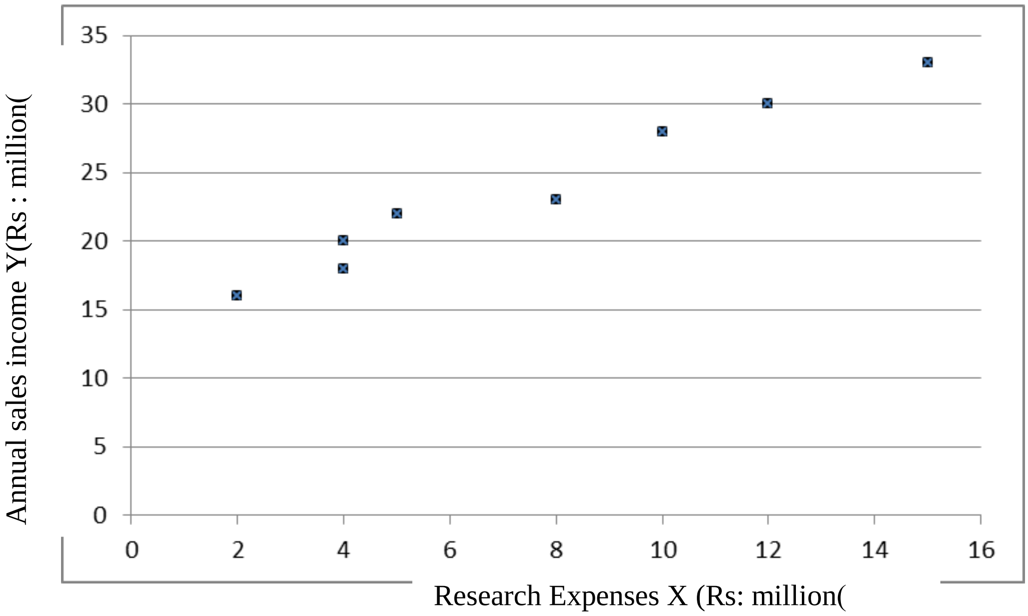
$$= 23.1 - 4.85 = 18.25$$

= Rs. <u>18.25 Million</u>

The actual sales income in 2014 is Rs 20 million is the sales income expected to have earned in the year 2014, according to the estimated trend line.The actual sales income in 2014 being Rs 1.75 million greater than the expected trend,it can be concluded that the sales in 2014 have been successfully organized.

4))a(

(i)



(ii) $\hat{Y} = 13.925 + 1.31 X$

The intercept is 13.925.It expresses that the average estimate of the sales income is Rs: 13.925 millions, once no any cent is incurred for reseach work.

The gradient is 1.31.It can be pointed out that when the expenses incurred for research work is increased by Rs:1 million ,the average estimate of the annual sales income will also increased by Rs:1.31 million.

(iii) Since $r = 0.98,$

$$R^2 = 0.98^2$$
$$= 0.9604$$

Since 96.04% of the total variation of the dependent variable is explained by the independent variable through this regression model of $\hat{Y} = 13.925 + 1.31 X$ is a very well fitted regression line.

(iv) For this purpose the regression line of X on Y should be derived

$$\hat{X} = \hat{a} + \hat{b} Y$$

$$\hat{b}_{x/y} = \frac{n \sum XY - \sum X \sum Y}{n \sum Y^2 - (\sum Y)^2}$$

$$= \frac{8 \times 1\,613 - 60 \times 190}{8 \times 4\,760 - (190)^2}$$

$$= \frac{12\,904 - 11\,400}{38\,128 - 36\,100}$$

$$= \frac{1\,504}{2\,028}$$

$$\hat{b}_{x/y} = 0.742$$

Accodingly ,the average estimate of expenditure incurred by this company for research works will increase by Rs:742 000 ,once the annual sales income gos up by Rs: 1 million.

)0.742 × 1 000 000 = Rs 742 000(

A sales income of Rs: 45 billion is a very extreme value beyond the given range of observations.Estimating the sales income, in accordance with such an extreme value is up to extrapolation, but extrapolation is not allowed under simple linear regression analysis.

(b(A disagreement is found in rankings of the judges A and b,while the rankings of a and c judges are stongly agreed,.Ranking of A and B have been disagreed,may be because of either the lack of awareness of judge B about the bowlers or him being biased in ranking.

)c(

(i) - Ability of giving an assurance to the consumer, that they can receive a high quality product.
- Ability to enhance the goodwill of the firm.
- Ability to secure and expand the existing market quota of the firm.
- Ability to enhance the efficiency of employees.
- Ability to minimise the wastage and missuse of resources.

(ii)

Sample No	$\sum x$	$\bar{x}$	R
1	246	49.2	6
2	251	50.2	4
3	257	51.4	5
4	264	52.8	8
5	243	48.6	7
6	252	50.4	5
7	238	47.6	8
8	262	52.4	6
9	258	51.6	6
10	271	54.2	10
		508.4	65

Central line $CL_{\bar{x}} = \bar{\bar{x}} = \frac{\sum \bar{x}}{k} = \frac{508.4}{10} = 50.84$ $\quad \bar{R} = \frac{65}{10} = 6.5$

Upper control limit UCL $= \bar{\bar{x}} + A_2 \bar{R}$

$= 50.84 + 0.577 \times 6.5$

$= 50.84 + 3.7505$

$= 54.5905$

$= \underline{\underline{54.59}}$

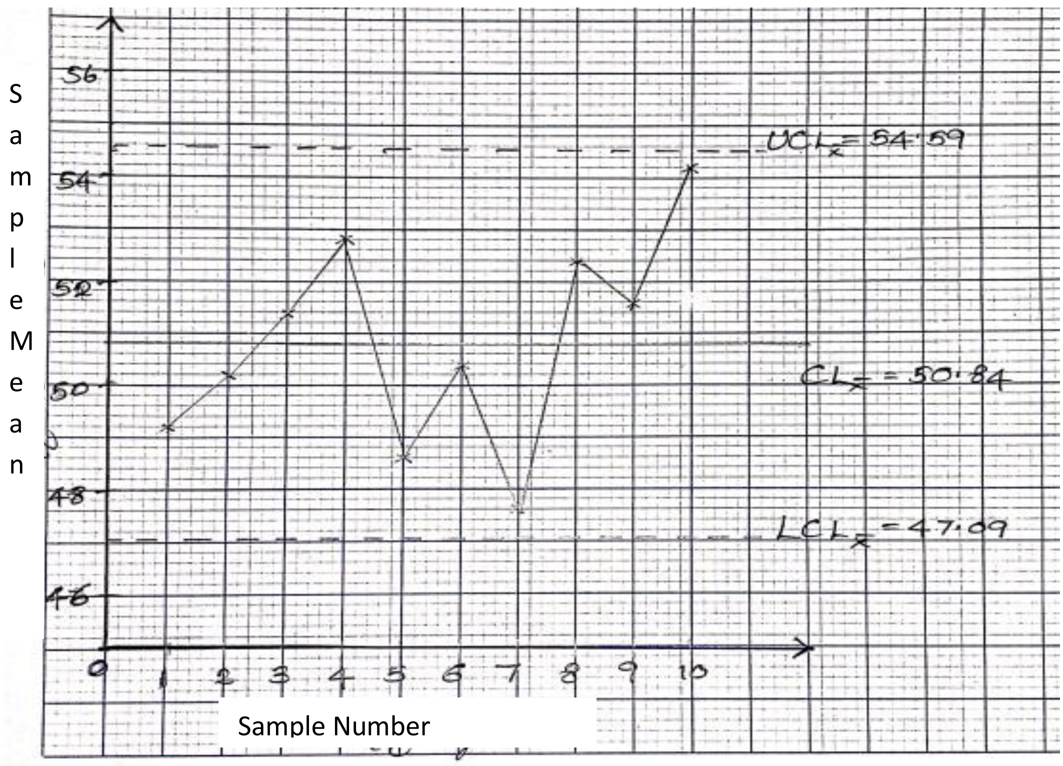
Lower control limit LCL $= \bar{\bar{x}} - A_2 \bar{R}$

$= 50.84 - 0.577 \times 6.5$

$= 50.84 - 3.7505$

$= 47.0895$

$= \underline{\underline{47.09}}$



)d(n= 80 C = 2

Since the probability accepting a lot at AQL = 0.01

Hence, P = 0.01 and n = 80

No: of defectives $x \approx Po(\lambda = np)$

$\lambda = 80 \times 0.010$

$= 0.8$

$P(x \geq 2(\quad = P(x = 0\ ) + P(x = 1) + P(x = 2\ )$

$= 0.4493 + 0.3595 + 0.1438$

$= 0.9526$

Producers Risk = 1- 0.9526 = 0.0474

Consumers risk at LTPD = 0.0625

$\lambda = 80 \times 0.0625$

$= 5$

$P(x \geq 2\ ) = P(x = 0\ ) + P(x = 1) + P(x = 2\ )$

$= 0.0067 + 0.0337 + 0.0842$

$= 0.1246$

(ii) In order to reduce the consumer's risk ,

1. Decreasing the acceptance number C, keeping n at constant.
2. Increasing LTPD.
3. Increasing the sample size.

5))a(Can be agreed.

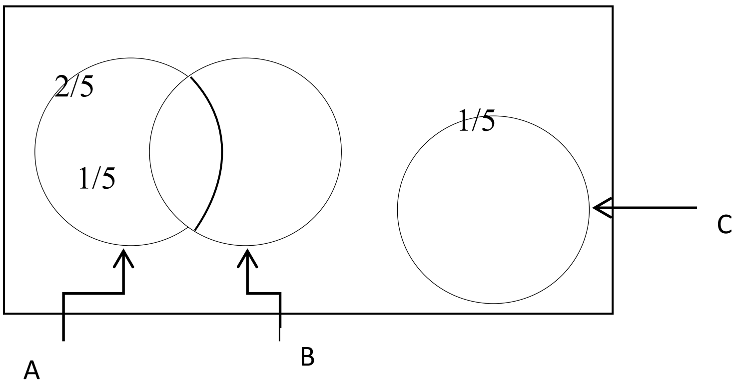
The events that exclude the occurance of other event, when one event occurs are called mutually exclusive eents.Since the occurance of one event is strongly affected by another event; mutually exclusive events can not be independent.

But if the probability of the occurrence of one of two independent events is zero, then the simultaneous occurrence of the two events or the intersection is zero. Then the events are mutually exclusive. If the probability of one of two events occurring is zero then those events are independent but also mutually exclusive.

If $P(A) = 0$, then $P(A \cap B) = P(A) \cdot P(B($

Hence, $P(A \cap B) = 0 \times P(B) = 0$

)b(



i) $P(A \cap B) = P(A) - P(A \cap B')$

$= \frac{2}{5} - \frac{1}{5} = \frac{1}{5}$

ii) P(A) . P(B(= $P(A \cap B)$,) hence A and B are independent events(

$\frac{2}{5} \times$ P(B($= \frac{1}{5}$

P(B($= \frac{1}{5} \times \frac{2}{5} = \frac{1}{2}$

iii) $P(B \cap C) = 0$

iv) The occurrence of only B can be defined as $P(B \cap A' \cap C')$.

If so, $P(B \cap A' \cap C') = P(B) - (P(A \cap B) + P(B \cap C))$

$= \frac{1}{2} - (\frac{1}{5} + 0)$

$= \frac{1}{2} - \frac{1}{5}$

$= \frac{5}{10} - \frac{2}{10}$

$= \frac{3}{10}$

v) $P(A \cup B \cup C)' = 1 - P(A \cup B \cup C)$

$= 1 - \{[P(A) + P(B) + P(C)] - [P(A \cap B) + P(B \cap C)]\}$

$= 1 - \left\{\left[\frac{2}{5} + \frac{1}{2} + \frac{1}{5}\right] - \left[\frac{1}{5} + 0\right]\right\}$

$= 1 - \left\{\left[\frac{4+5+2}{10}\right] - \left[\frac{2+0}{10}\right]\right\}$

$= 1 - \left\{\left[\frac{11}{10}\right] - \left[\frac{2}{10}\right]\right\}$

$= \frac{10}{10} - \frac{11}{10} + \frac{2}{10} = \frac{1}{10}$

)c(

Preference	Rural(A)		Urban(B)		**Total**
	Employeed(E)	Unemployeed(F)	Employeed(E)	Unemployeed(F)	
Like(C)	120	30	140	40	**330**
Dislike(D)	30	70	40	30	**170**
T	**150**	**100**	**180**	**70**	**500**

Let's identify events defined on this sample space.

Rural -A Urban - B Like -C Dislike -D

Employeed -E Unemployeed- F

i) $P(C) = \frac{n(C)}{n(S)} = \frac{330}{500}$

ii) $P(B) = \frac{n(B)}{n(S)} = \frac{250}{500}$

iii) $P(F) = \frac{n(F)}{n(S)} = \frac{170}{500}$

iv) $P(C \cap B \cap E) = \frac{n(C \cap B \cap E)}{n(S)} = \frac{140}{500}$

v) $\frac{P(A \cap E)}{P(D)} = \frac{P(A \cap E \cap D)}{P(D)} = \frac{30}{500} \div \frac{170}{500}$

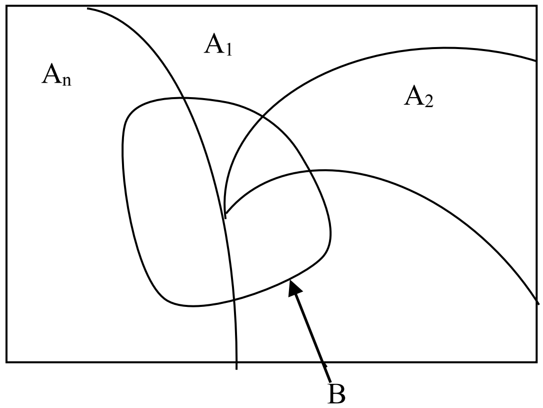
$$= \frac{30}{500} \times \frac{500}{170} = \frac{30}{170}$$

vi) If "C" is for favoring the proposed constitutional reforms and "A" for being a resident of a rural area,
For C and A to be independent,
$P(C).\, P(A) = P(C \cap A)$

$$\frac{330}{500} \times \frac{250}{500} \neq \frac{150}{500}$$

Therefore, it cannot be said that giving consent for the proposed constitutional reform is independent of the person's area of residence.

)d(



Given that an event B which is common to A_1 , A_2 , A_3 ……. A_n set of mutually exclusive and collectively exhaustive events has already occurred, the probability of occurring an event indicated by A_i in that sample space is defined by Bayes' theorem. It can be computed as follows.

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)}$$

$$P(A_i/B) = \frac{P(Ai).\ P(B/Ai)}{\sum P(Ai)\, P(B/Ai)}$$

If,

being a marine insurance policy holder is A_1,

being a motor vehicle insurance holder is A_2,

being a life insurance holder is A_3,

being an intellectual property insurance holder is A_4, and

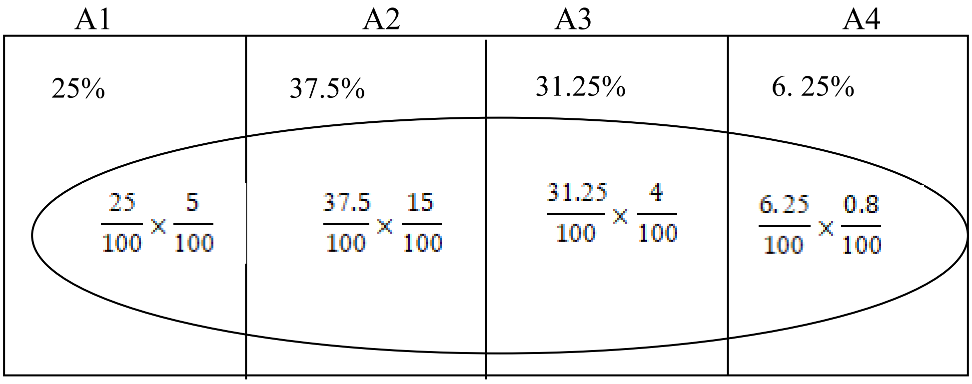
applying for insurance compensation within the first 3 years is B,

$$P(A_i/B) = \frac{P(Ai).\ P(B/Ai)}{\sum P(Ai)\, P(B/Ai)}$$

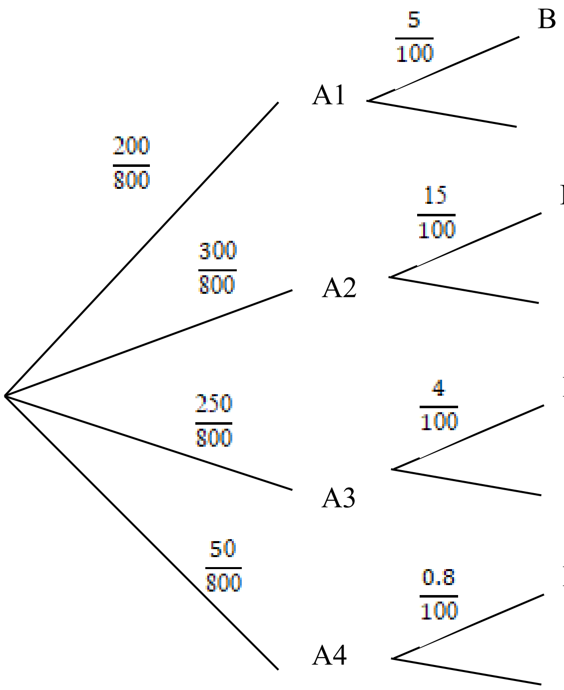
$$= \frac{\left(\frac{200}{800} \times \frac{5}{100}\right)}{\left(\frac{200}{800} \times \frac{5}{100}\right)\left(\frac{300}{800} \times \frac{15}{100}\right)\left(\frac{250}{800} \times \frac{4}{100}\right)\left(\frac{50}{800} \times \frac{0.8}{100}\right)}$$

$$= \frac{\left(\frac{1000}{800 \times 100}\right)}{\left(\frac{1000+4500+1000+40}{800 \times 100}\right)} = \left(\frac{1000}{6540}\right) = 0.1529$$

This problem can be solved by analyzing the total probability theorem very clearly by dividing the sample space as mentioned below.



This problem can also be solved by using tree diagrams.



$$\frac{200}{800} \times \frac{5}{100} = \frac{1\,000}{80\,000}$$

$$\frac{300}{800} \times \frac{15}{100} = \frac{4\,500}{80\,000}$$

$$\frac{250}{800} \times \frac{4}{100} = \frac{1\,000}{80\,000}$$

$$\frac{50}{800} \times \frac{0.8}{100} = \frac{40}{80\,000}$$

$$P(B) = \frac{6\,540}{80\,000}$$

$$P\,(A_i/B) = \frac{P(Ai).\ P\,(B/Ai)}{\sum P(Ai)\,P\,(B/Ai)}$$

$$= \left(\frac{1000}{80\,000}\right) \times \left(\frac{80\,000}{6540}\right) = \left(\frac{1000}{6540}\right) = 0.1529$$

6))a(

(i)

x	0	1	2	3	4
P(x)	$\frac{256}{625}$	$\frac{256}{625}$	$\frac{96}{625}$	$\frac{16}{625}$	$\frac{1}{625}$

- Having a certain n number of trials.

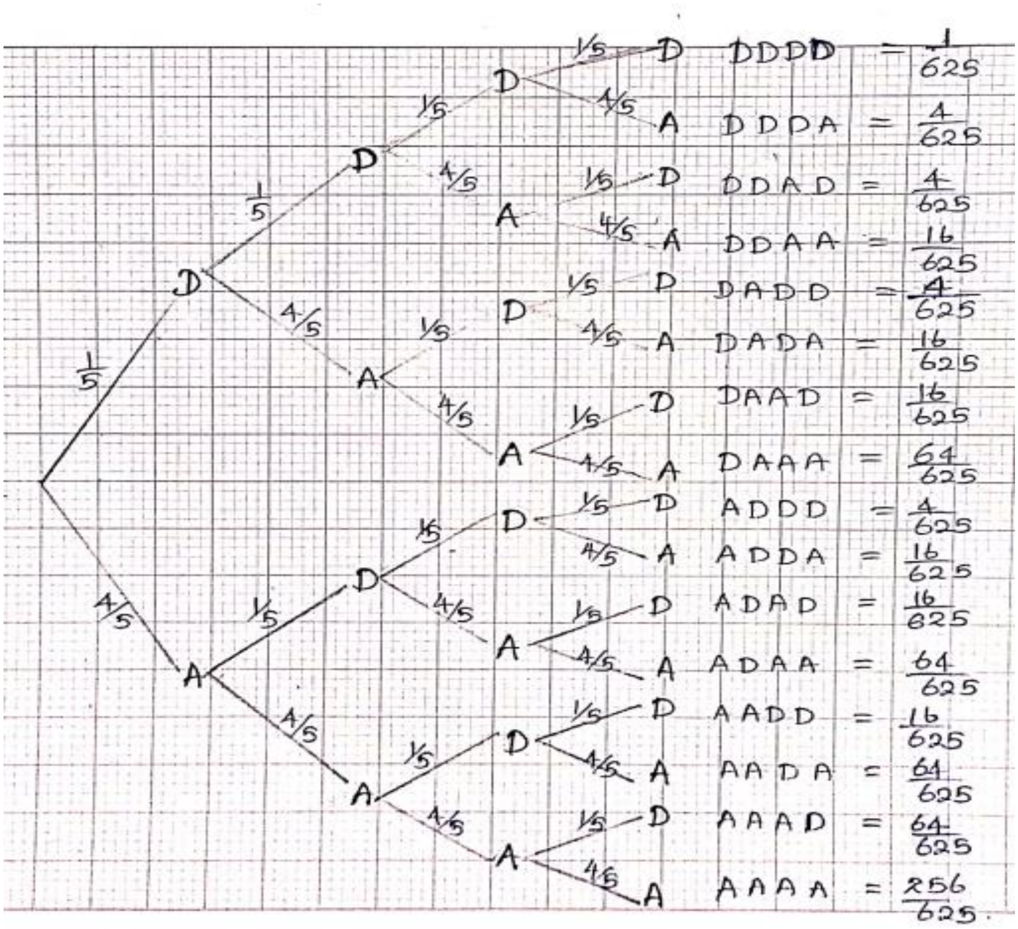
  There are four investment sources.(n=4)
- Having only 2 outcomes, success and failure.

  Here are unfavorable benefits and favorable benefits.
- The probability of success at each time is a constant value.

  Here the probability of receiving a return in each case is 1/5=0.2.
- Each trial is independent of other trials.

  Here, receiving unfavorable results from each investment source does not affect receiving unfavorable returns from the other investment source.

It is appropriate to use a diagram as follows to construct this probability distribution.



From the probability distribution you constructed, $P(x \leq 1) = P(x=0) + P(x=1)$

$= 256/625 + 256/625$

$= 512/625$

$= 0.8192$

or

x	0	1	2	3	4
P(x)	$\frac{256}{625}$	$\frac{256}{625}$	$\frac{96}{625}$	$\frac{16}{625}$	$\frac{1}{625}$

(ii) $P(x \geq 1($

(a(By using Binomial distribution, $P(X=x) = {}^{n}C_{x}\ p^{x} q^{n-x}$

$P(x \geq 1(= {}^{4}C_{0} \times 0.2^{0} \times 0.8^{4} + {}^{4}C_{1} \times 0.2^{1} \times 0.8^{3}$

$=)\ 1 \times 1 \times 0.4096(+)\ 4 \times 0.2 \times 0.512($

$= 0.4096 + 0.4096$

$= \underline{\underline{0.8192}}$

)b(

(i) Mean in 5 days = 1

$\lambda = 1/5 = 0.2$

$x \sim Po(0.2)$

- The mean number of accidents occurring in a certain time period is proportional to the length of the time period.
- The probability of two events occurring simultaneously in a very small time interval is negligible.
- The occurrence of accidents in a certain time period is independent of the occurrence of accidents in other non-overlapping time periods.

(ii) Large n. $(n > 50)$

Small p $(p < 0.1)$

)When these 2 conditions are combined, the mean of the binomial distribution can also be stated as $\mu = np < 5($

(iii) n= 20 000 p = 1/100 000 = 0.00001

$\lambda = np$

$\lambda = 20\,000 \times 0.00001 = 0.2$

$P(x \geq k) = 0.98$

$\lambda = 0.2$ Reading the Poisson distribution table, the cumulative probability value from x = 0 is 0.98 (approximate) is the value of n.

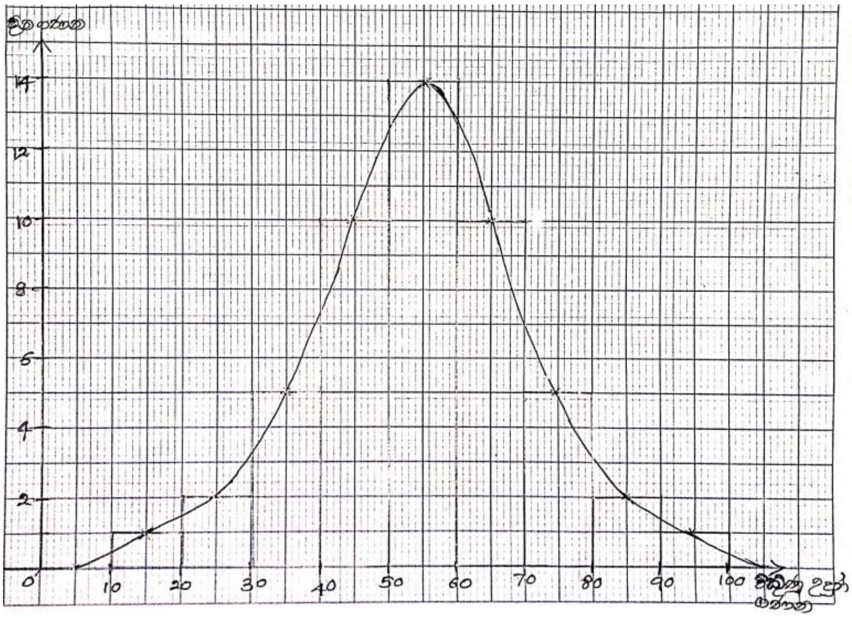
$P(x \leq 1) = P(x=0) + P(x=1)$

$= 0.8187 + 0.1637$

$= 0.9824$

So that, k = 1

)c(

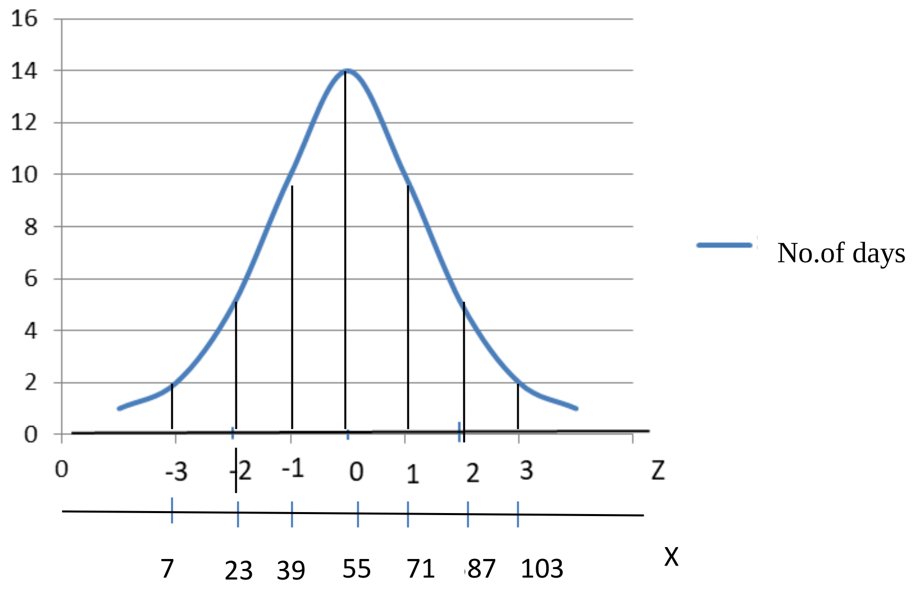


Class Intervals	f	u	fu	fu^2
10-20	1	-4	-4	16
20-30	2	-3	-6	18
30-40	5	-2	-10	20
40-50	10	-1	-10	10
50-60	14	0	0	0
60-70	10	1	10	10
70-80	5	2	10	20
80-90	2	3	6	18
90-100	1	4	4	16
			0	128

$$\bar{x} = A + \left(\frac{\Sigma fu}{\Sigma f}\right) C$$
$$= 55 + \left(\frac{0}{50}\right) 10$$
$$= 55$$

$$S^2 = \left[\frac{\Sigma fu2}{\Sigma f} - \left(\frac{\Sigma fu}{\Sigma f}\right)^2\right] C^2$$
$$= \left[\frac{128}{50} - \left(\frac{0}{50}\right)^2\right] 10^2$$
$$= (2.56 - 0) \times 100$$
$$= 256$$

$$S = \sqrt{\frac{\Sigma fu2}{\Sigma f} - \left(\frac{\Sigma fu}{\Sigma f}\right)^2} \times C$$
$$= \sqrt{\frac{128}{50} - \left(\frac{0}{50}\right)^2} \times 10$$
$$= \sqrt{2.56 - 0} \times 10$$
$$= 1.6 \times 10$$
$$= 16$$



P $(39 \leq X \leq 71)$

$$= P\left(\frac{39-55}{16} \leq Z \leq \frac{71-55}{16}\right)$$
$$= P(-1 \leq Z \leq 1)$$
$$= 0.3413 \times 2$$
$$= 0.6826$$

P $(23 \leq X \leq 87)$

$$= P\left(\frac{23-55}{16} \leq Z \leq \frac{87-55}{16}\right)$$
$$= P(-2 \leq Z \leq 2)$$
$$= 0.4772 \times 2$$
$$= 0.9544$$

P $(7 \leq X \leq 103)$

$$= P\left(\frac{7-55}{16} \leq Z \leq \frac{103-55}{16}\right)$$
$$= P(-3 \leq Z \leq 3)$$
$$= 0.0.4987 \times 2$$
$$= 0.9974$$

Thus, most of the characteristics that can be seen in a normal distribution can be seen in this frequency distribution. 100% of the observations in this distribution are located in the range of $-3 \leq Z \leq +3$ from the mean.

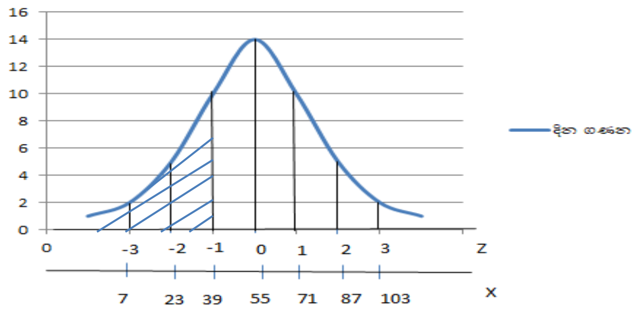
But 0.13 of the observations of a normal distribution extend to infinity on either side.

(iv) $P(X < 39) = P(Z > \frac{39-55}{16})$

$= P(Z > -1)$

$= 0.5 - 0.3413$

$= 0.1587$

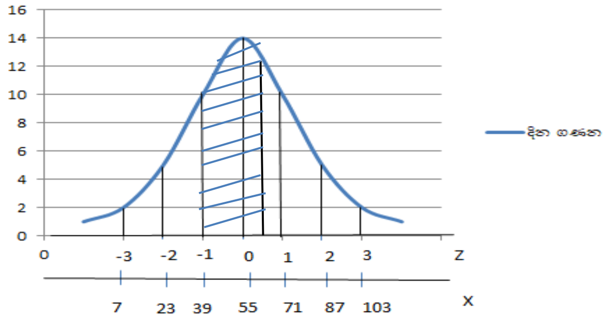


$P(39 \le X \le 64) = P(\frac{39-55}{16} \le Z \le \frac{64-55}{16})$

$= P(-1 \le Z \le 0.5625)$

$= 0.3413 + 0.2123$

$= 0.5536$



7) (a(

(i) Purposive Sampling

This is a problem that needs to be solved with the active contribution of epidemiologists

(ii) Quota Sampling

It is sufficient to select the sample according to the discretion of the investigator to represent different categories of passengers who travel from Colombo to Badulla railway stations like rural, urban, female, male, office workers, school students etc.

(iii) Judgement Sampling

In order to study financial irregularities, it is advisable to select a sample of the necessary documents and persons under the guidance of auditors with exceptional experience in the matter.

)b(

(i) Parameter

A numerical measure that is defined in relation to a variable is called a parameter. A parameter is an unknown constant.

Ex: Population means μ

Population Variance σ^2

Statistics

Numerical measures calculated with respect to the sample elements are called the sample statistics. The sample statistics is always a random variable.

Ex: Sample Mean $\bar{x}$

Sample propotion p

(ii) Estimator

A function of sample statistics used to estimate the value of a particular population parameter is called an estimator for that population parameter.

Ex : Since the sample mean $\bar{x} = \frac{\sum x}{n}$ is used to estimate the population mean μ, $\bar{x}$ is a predictor of μ.

Estimate

The numerical value calculated from the sample statistics corresponding to a given estimator is called the estimate.

Ex: The average daily wage of a sample of 25 small businessmen is Rs. 88,500.

)c(

(i) The sampling distribution of the difference of two sample means,

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$$

If two populations are infinite,

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

If two populations are finite,

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1}\left[\frac{N_1 - n_1}{N_1 - 1}\right] + \frac{\sigma_2^2}{n_2}\left[\frac{N_2 - n_2}{N_2 - 1}\right]$$

(ii) Sampling distribution of sample proportion,

$$\mu_p = \pi$$

$$\sigma_p^2 = \frac{\pi(1-\pi)}{n}$$

If the population is finite,

$$\sigma_p^2 = \frac{\pi(1-\pi)}{n}\left[\frac{N-n}{N-1}\right]$$

(d) x_1 – Sales of charcoal cooker
x_2 - Sale of biogas cooker

$x_1 \sim N(24.5, 4.2^2)$

$x_2 \sim N(23.2, 5.6^2)$

$(\bar{x}_1 - \bar{x}_2) \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}($

$(\bar{x}_1 - \bar{x}_2) \sim N(24.5 - 23.2, \frac{4.2^2}{25} - \frac{5.6^2}{25}($

(i) $P(\bar{x}_1 > \bar{x}_2) = P(\bar{x}_1 - \bar{x}_2 > 0)$

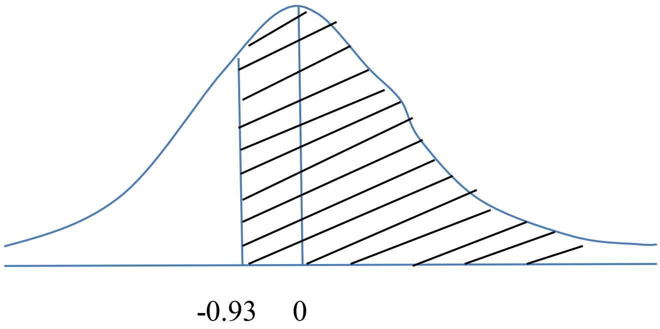
$= P\left(z > \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right.$

$= P\left(z > \frac{0 - 1.3}{\sqrt{\frac{4.2^2}{25} - \frac{5.6^2}{25}}}\right($

$= P(z > -0.93)$

$= 0.5 + 0.3238$

$= 0.8238$



(ii) $P(\bar{x}_1 - \bar{x}_2 > 2.5) = P\left(z > \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right($

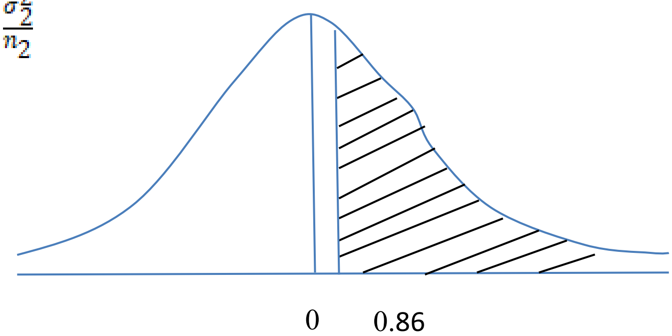
$= P\left(z > \frac{2.5 - 1.3}{\sqrt{\frac{4.2^2}{25} - \frac{5.6^2}{25}}}\right($

$= P\left(z > \frac{1.2}{1.6}\right($

$= P(z > 0.86($

$= 0.5 - 0.3051$

$= 0.1949$



8)

(a) 1. Unbiasedness

2. Efficiency

3. Consistency

4. Sufficiency

1. Since the expected value of sampling distribution of the sample means is equal to the population mean (E ($\bar{x}$) = μ), the sample mean $\bar{x}$ satisfies the property of unbiasness.

2. Var ($\bar{x}$) = $\frac{\sigma^2}{n}$, Var (Median) = $\frac{\pi\sigma^2}{2n}$

Mean or Median or Mode of the sample should be used as an estimate for population mean μ.Mode is not an identical measure,so that we can't keep hope on mode to be a good estimator.Median is greater than the variance of mean.

$\frac{\pi\sigma^2}{2n} > \frac{\sigma^2}{n}$

Accordingly the sample mean is the efficient estimator for population mean μ, as the estimator with the least variance.

3. Sample size and the variance of the sample mean is inversely related ,and therefore (Var ($\bar{x}$) = $\frac{\sigma^2}{n}$).Accordingly the sample mean ($\bar{x}$) is also a consistant estimator for population mean μ,because of the fact that,the variance of the sample mean gradually approaches to zero,once the sample gradually increasing.

4. The sample mean is the only measure of central tendency that all the observations of a data set are taken in to consideration. Therefore all the characteristics of the observation can be expected to be summarised for the population mean by the sample mean, so that the sample mean also is a sufficient estimator.Thus the sample mean $\bar{x}$ satisfies all the properties expected from a good point estimator.

(b)

(i) The confidence level and the width of the confidence interval are directly proportionate, so that it's true the fact that "A narrower confidence level can be estimated with a less confedent level".

The width of the confidence level is not influenced only by the sample size and confidence level.It's also directly proportionate with the standard deviation.Therefore a narrower confidence interval with a greater confidence level can be estimated with a samaller standard deviation .The second part of this statement is false.

ii. Hence It's given that,

$n_1 = 10$ $n_2 = 10$

$\bar{x}_1 = 1.8$ $\bar{x}_2 = 0.1$

$S_1 = 2.95$ $S_2 = 0.2$

$X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$

It seems to be σ_1^2 and σ_2^2 are unknown.Since both the samples are small, the t distribution with $n_1 + n_2 - 2$ degrees of freedom can be used to estimate the confidence intervals for $(\mu_1 - \mu_2)$ assuming that unknown population variances are equal ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) ;

Therefore the pooled variance of sample variances can be used as a good estimator for common population variance σ^2 as follows,

$$S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

$$= \frac{9\times 1.8^2 + 9\times 0.1^2}{10+10-2}$$

$$= \frac{9\times 3.24 + 9\times 0.01}{18}$$

$$= \frac{9(3.24+0.01)}{18}$$

$$= \frac{3.25}{2}$$

$$S_p^2 = 1.625$$

$\therefore$ $S_p = \sqrt{1.625} = \underline{\underline{1.275}}$

$\therefore$ 98% confidence intervals for $\mu_1 - \mu_2$ can be estimated as follows,

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm t_{n_1+n_2-2, \propto/2} . S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= (1.8 - 2.95) \pm t_{18, 0.01.} 1.275 \sqrt{\frac{1}{10} + \frac{1}{10}}$$

$$= -1.15 \pm 2.55 \times 1.275 \sqrt{0.2}$$

$$= -1.15 \pm 1.454$$

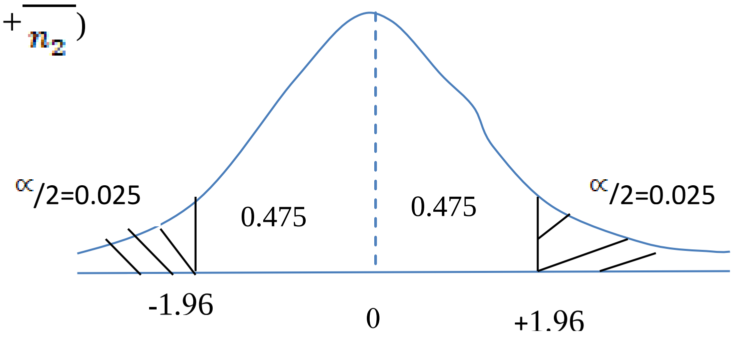
$\therefore$ $-2.604 \leq \mu_1 - \mu_2 \leq 0.304$

It’s 98% confident that the average time taken by a trainee to provide servces to a customer can be 2.604 minutes more than a trained officer from one end and same times a trainee may be quicker with an average of 0.304 minutes than a trained officer at the other end.

(c) Both the population distributions are unknown.

$n_1 = 50$ $n_2 = 50$

$$(\bar{x}_1 - \bar{x}_2) = N\left(\mu_1 - \mu_2, \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)$$



Test statistic

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu 1 - \mu 2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(1920 - 1840) - (0)}{\sqrt{\frac{80^2}{50} + \frac{120^2}{50}}}$$

$$= \frac{80\sqrt{50}}{\sqrt{6400 + 14\,400}}$$

$$= \frac{80\sqrt{50}}{\sqrt{20\,800}}$$

$Z = 3.92$

Decision: H_0 is rejected,

Since the test statistic falls in the critical region.(Since Z_{cal} ,3.92 > Z_{table} 1.96)

Conclusion:

It's highly evident that the mean harvests of these two paddy species are significant at 5% leel of significance.

(d) $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1: Means of at least two populations are significant

$n_1 = 5$ $n_2 = 4$ $n_3 = 6$ $n_4 = 3$

$N = n_1 + n_2 + n_3 + n_4$

$= 5 + 4 + 6 + 3$

$\therefore N = \underline{\underline{18}}$

$\sum X_1 = 36$ $\sum X_2 = 52$ $\sum X_3 = 56$ $\sum X_4 = 42$

∴ Error term = $\frac{T^2}{N} = \frac{(36+52+56+42)^2}{18} = \frac{(186)^2}{18}$ = 1 922

SST = $(\sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2) - \frac{T^2}{N}$

= (280 +696+554+596) – 1 922

= 2 126 – 1 922

= 204

SSC = $\left(\frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \frac{(\sum X_3)^2}{n_3} + \frac{(\sum X_4)^2}{n_4}\right) - \frac{T^2}{N}$

$= \left(\frac{(36)^2}{5} + \frac{(52)^2}{4} + \frac{(56)^2}{6} + \frac{(42)^2}{3}\right) - 1\ 922$

$= \frac{1\ 296}{5} + \frac{2\ 704}{4} + \frac{3\ 136}{6} + \frac{1\ 764}{3}$ -1 922

= 259.2 + 676 + 522.7 +588 – 1 922

=2 045.9 – 1 922

= 123.9

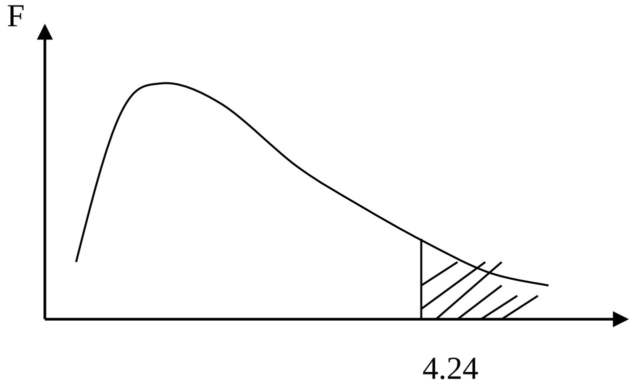
∴SSE = SST – SSC

= 204.0 – 123.9

= 80.1

ANOVA Table

Source of Variance	Square sum of Errors	df	Mean Sum of squres	F- Value
Between the samples	SSC= 123.9	K-1 = 4-1 = 3	MSC =SSC/(k-1) = 123.9/3	F= $\frac{MSC}{MSE}$ = 41.3 /5.72 = 7.22
Within the samples	SSE = 80.1	N-K = 18-4 =14	MSE = SSE / (N-K) = 80.1/14	
Total	SST = 204			



$F_{\frac{K-1}{N-K}}$, 0.025

$F_{\frac{3}{14}}$, 0.025 = 4.24

Decision: H_0 is rejected, since the test statistic 7.22 is greater than the critical value 4.24

Conclution: Average sales of 4 cities can not be expected to be equal at 2.5% level of significance.