Course Code : BCS-040
Course Title : Statistical Techniques
Assignment Number : BCA(4)040/Assignment/2022-23
Maximum Marks : 100
Weightage : 25%
Last Date of Submission : 31st October, 2022 (for July session)
: 15th April, 2023 (for January session)

Q1. In a study on the Per capita Income for a particular year in a city, the following weekly observations were made. **(5)**

Per Capita Income (Rs.) - (1K=1000)	14K-15K	15K-16K	16K-17K	17K-18K	18K-19K	19K-20K
Number of Weeks	5	10	20	9	6	2

Draw a histogram and a frequency polygon on the same scale

Q2. Do you find any correlation between ages and playing habits of the students, whose distribution according to age groups is given in the following table **(5)**

Age of groups(Years)	15-16	16-17	17-18	18-19	19-20	20-21
Number of Students	200	270	340	360	400	300
Number of Regular players	150	152	170	180	180	120

Q3. Data are given below shows statistics viz. standard deviation & average marks secured by students, in the examination of subject A and B **(5)**

	SUBJECT A	SUBJECT B
MEAN MARKS	36	85
STANDARD DEVIATION	11	8

Assuming the Coefficient of correlation between A and B = ±0.66
Perform the following tasks:

i) Determine the two equations of regression
ii) Calculate the expected marks in A corresponding to 75 marks obtained in B.

Q4. Calculate 2-sigma and 3-sigma upper and lower control limits for means of samples 4 and prepare a control chart for a drilling machine, which bores holes with a mean deviation of 0.5230 cm and a standard deviation of 0.0032 cm. **(5)**

Q5. Construct 5- yearly moving averages from the following data **(5)**

YEAR	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
SALE	105	107	109	112	114	116	118	121	123	124	125	127	129

Q6. In 120 throws of a single dice, following distribution of faces was observed. **(5)**

FACES	1	2	3	4	5	6	TOTAL
F_0	30	25	18	10	22	15	120

From the given data, verify that the hypothesis "dice is biased" is acceptable or not.

Q7. A company wants to estimate, how its monthly costs are related to its monthly output rate. The data for a sample of nine months is tabulated below:

Out Put (Tons)	1	2	4	8	6	5	8	9	7
Cost (Lakhs)	2	3	4	7	6	5	8	8	6

Using the data given above, perform following tasks :

(a) Calculate the best linear regression, where the monthly output is the dependent variable and monthly cost is the independent variable.
(b) Use the regression line to predict the company's monthly cost, if they decide to produce 4 tons per month.

Q8. The Probability that at least one of the two Independent events occur is 0.5. Probability that 5 first event occurs but not the second is (3/25). Also the probability that the second event occurs but not the first is (8/25). Find the probability that none of the two event occurs.

Q9. Marks of six students are tabulated below : **(5)**

Name :	Raj	Anil	Amit	Om	Rita	Renu
Marks :	54	50	52	48	50	52

From the population, tabulated above, you are suppose to choose a sample of size two.
(a) Determine, how many samples of size two are possible
(b) Construct sampling distribution of means by taking samples of size 2 and organize the data.

Q10. Two new types of petrol, called premium and super, are introduced in the market, and their manufacturers claim that they give extra mileage. Following data were obtained on extra mileage which is defined as actual mileage minus 10. **(5)**

Ordinary Petrol	1	2	2	1
Premium Petrol	2	2	1	3
Super Petrol	4	1	2	3

(i) Using ANOVA, test whether premium or super gives an extra mileage.

(ii) What is your estimate for the error variance?

(iii) Assuming that the error variance is known and is equal to 1, obtain the 95 % confidence interval for the mean extra mileage of super.

Q11. Two floppies are selected at random without replacement from a box containing 7 good and 3 defective floppies. Let A be the event that the first floppy drawn is defective, and let B be the event that the second floppy drawn is defective.

(i) Find the conditional probabilities P(B/A) and P(B/AC)

(ii) Show that P(B) = P(B/A). P(A) + P(B/AC) P(AC) = P(A).

Q12. A drilling machine bores holes with a mean deviation of 0.5230 cm and a standard deviation of 0.0032 cm. Calculate 2-sigma and 3-sigma upper and lower control limits for means of samples 4 and prepare a control chart.

Q13. What are control charts briefly discuss the utility of control charts? **(5)**

Q14. Compare the following

a) Cluster sampling , Stratifies sampling and Systematic sampling
b) Parametric and Non-Parametric Tests

Q15. Explain the following with the help of an example each: **(10)**

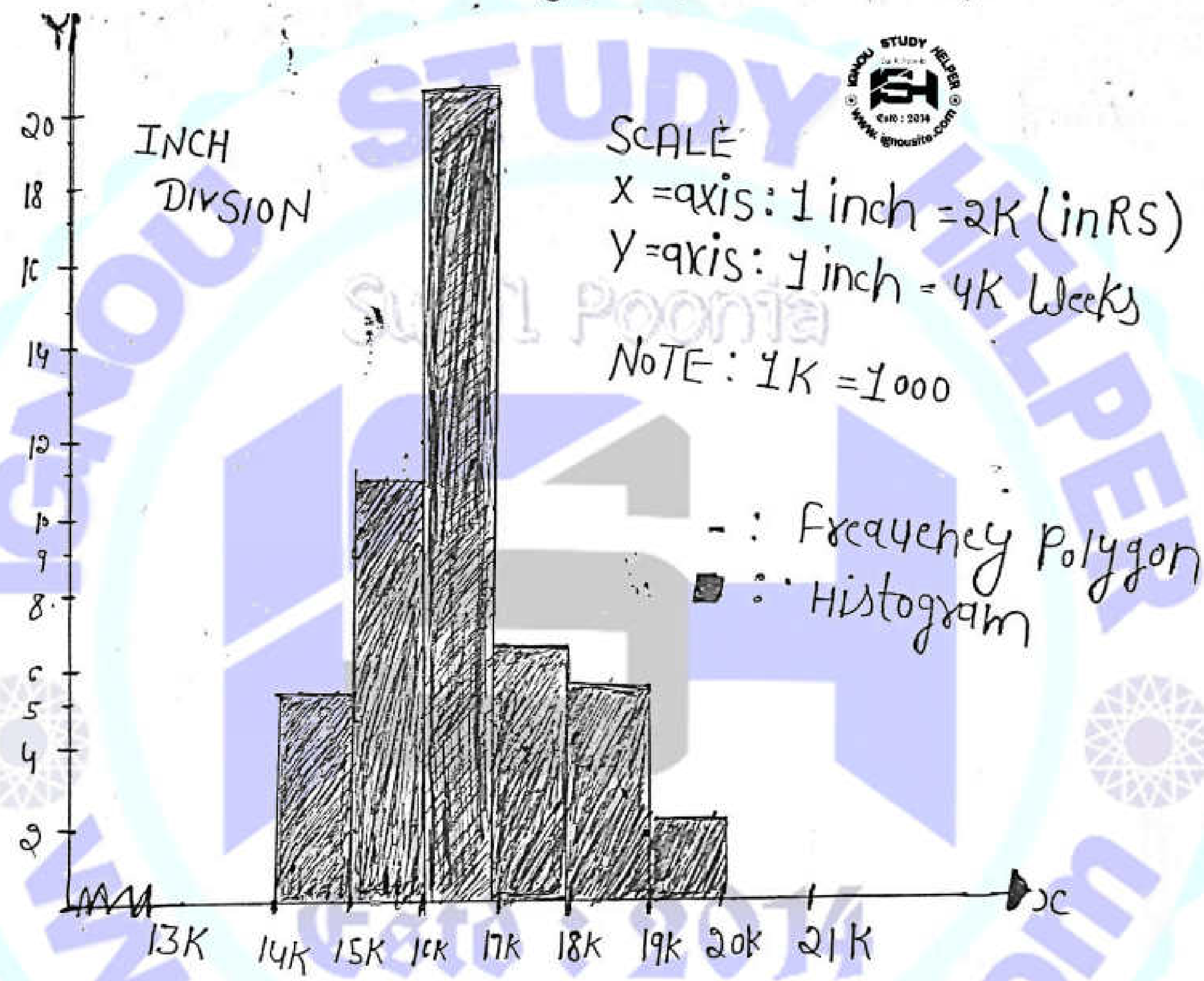
a) Goodness of fit test b) Test of Independence c) Criteria for a good estimator
d) Chi-Square Test

Ans. 1

Per Capita Income (Rs) (1k=1000)	Mid Value (x_i)	Frequency (f_i)
13k-14k	13.5k	0
14k-15k	14.5k	5
15k-16k	15.5k	10
16k-17k	16.5k	20
17k-18k	17.5k	9
18k-19k	18.5k	6
19k-20k	19.5k	2
20k-21k	20.5k	0

Answer : 1

Histogram and Frequency Polygon (Graph) :



Answer:- 2

Calculation of Percentage of Regular Players:

Number of students	Number of Regular Players	Percentage of Regular Players (y_i) in %
200	150	$\frac{150}{200} \times 100 = 75$
270	152	$\frac{152}{270} \times 100 = 56.3$
340	170	$\frac{170}{340} \times 100 = 50$
360	180	$\frac{180}{360} \times 100 = 50$
400	180	$\frac{180}{400} \times 100 = 45$
300	120	$\frac{120}{300} \times 100 = 40$

$$\text{Mean } \bar{x} = \frac{\sum x_1}{n}$$

$$= \frac{15.5 + 16.5 + 17.5 + 18.5 + 19.5 + 20.5}{6}$$

$$= \frac{180}{6}$$

$$= 18$$

$$\text{Mean } \bar{y} = \frac{\sum y_i}{n}$$

$$= \frac{75 + 56.3 + 50 + 50 + 45 + 40}{6}$$

$$= \frac{316.3}{6}$$

$$= 52.7167$$

Use assumed mean $A = \bar{x} = 18$

$\bar{y} = 52.7167$ mean is not an integer use assumed mean $B = 53$

Class-X	mid value x	y	dx=x-A=x-18	dy=y-B=y-53	dx^2	dy^2	dx-dy
15-16	15.5	75	-2.5	22	6.25	484	-55
16-17	16.5	56.3	-1.5	3.3	2.25	10.89	-4.95
17-18	17.5	50	-0.5	-3	0.25	9	1.5
18-19	18.5	50	0.5	-3	0.25	9	-1.5
19-20	19.5	45	1.5	-8	2.25	64	-12
20-21	20.5	40	2.5	-13	6.25	169	-32.5
-	-	-	-	-	-	-	-
-	108	316.3	$\sum dx = 0$	$\sum dy = 1.7$	$\sum dx^2 = 17.5$	$\sum dy^2 = 745.89$	$\sum dx\text{-}dy = -104.45$

4

correlation coeffciente :

$$r = \frac{n\sum dxdy - \sum dx \cdot \sum dy}{\sqrt{n\sum dx^2 - (\sum dx)^2} \cdot \sqrt{n\sum dy^2 - (\sum dy)^2}}$$

$$= \frac{6 \cdot -104.45 - 0 \cdot -1.7}{\sqrt{6 \cdot 17.5 - (0)^2} \cdot \sqrt{6 \cdot 745.89 - (-1.7)^2}}$$

$$= \frac{-626.7 + 0}{\sqrt{105 - 0} \cdot \sqrt{4475.34 - 2.89}}$$

$$= \frac{626.7}{\sqrt{105} \cdot \sqrt{4472.45}}$$

$$= \frac{626.7}{10.247 \cdot 66.8764}$$

$$= \frac{626.7}{685.279}$$

$$= -0.9145$$

Hence, co-efficient of correlation is $= -0.9145$ Ans

Answer :- 3

Given, $\bar{X} = 36$, $\bar{Y} = 85$ $\sigma_x = 11$, $\sigma_y = 8$ and $r = \pm 0.66$

Calculation of Regression co-efficient

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} = (\pm 0.66) \times \frac{8}{11} = \pm 0.48$$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} = (\pm 0.66) \times \frac{11}{8} = \pm 0.48 = \pm 0.9075$$

(i) Regression Equations

Regression equations of X on y :

$$X - \bar{X} = b_{xy}(y - \bar{y})$$

$$= X - 36 = \pm 0.9075(y - 85)$$

$$= X - 36 = \pm(0.9075y - 77.1375)$$

$$= X = 36 \pm (0.9075y - 77.1375)$$

$$= \text{i.e. } X = 0.9075y - 41.1375 \text{ or } X = -0.9075y + 113.1375 \text{ Ans}$$

Regression equations of y on X :

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$= y - 85 = \pm 0.48(x - 36)$$

$$= y - 85 = \pm(0.48x - 17.28)$$

$$= y = 85 \pm (0.48x - 17.28$$

$$= \text{i.e. } y = 0.85x + 67.72 \text{ or } = -0.48x + 102.28 \text{ Ans}$$

(ii) calculation of expected marks in a corresponding to 75 marks obtained in B

$x = 0.9075(75) - 41.1375$ or $X = -0.9075(75) + 113.1375$

$= x = 68.0625 - 41.1375$ or $x = -68.0625 + 113.1375$

$= x = 26.925$ or $x\ 45.075$ Ans

Answer :- 4

μ or $\bar{x} = 0.5230,\ \sigma = 0.0032,\ n = 4$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.0032}{\sqrt{4}} = 0.0016$$

2 Sigma (Internal contral) Limits

$UCL_{\bar{x}} = \mu + 2\sigma_{\bar{x}} = 0.5230 + 2 \times 0.0016 = 0.5262$ cm

$LCL_{\bar{x}} = \mu - 2\sigma_{\bar{x}} = 0.5230 - 2 \times 0.0016 = 0.5198$ cm

3 Sigma (outer contral) Limits :

$UCL_{\bar{x}} = \mu + 3\sigma_{\bar{x}} = 0.5230 + 2 \times 0.0016 = 0.5278$ cm

$LCL_{\bar{x}} = \mu - 3\sigma_{\bar{x}} = 0.5230 - 2 \times 0.0016 = 0.5182$ cm

Based on the above calculation the contral charts is the fallowling

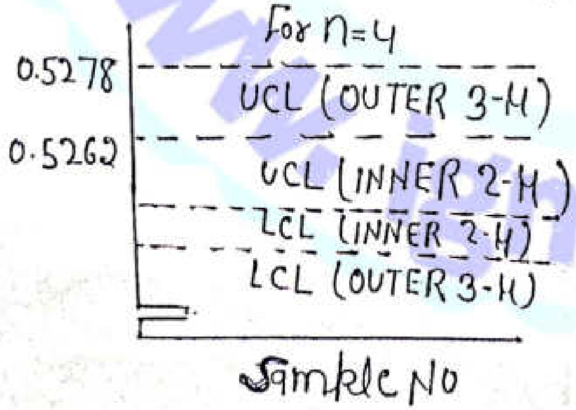


Fig contral chart

Answer :-5

Year	Sales	5-yearly totals	5 yearly moving averages
2000	105	---	---
2001	107	---	---
2002	109	547	109.4
2003	112	558	111.6
2004	114	569	113.8
2005	116	581	116.2
2006	118	592	118.4
2007	121	602	120.4
2008	123	611	122.2
2009	124	620	124.0
2010	125	628	125.6
2011	127	----	----
2012	129	----	-----

Answer :-6
on the basis of Principle of equal Probability $p = \frac{1}{6}$ the theoretical frequencies for each face is $Np = 120 \times \frac{1}{6} = 20$ Thus we have,

Faces	1	2	3	4	5	6	Total
Fo	30	25	18	10	22	15	120
Fe	20	20	20	20	20	20	120

Since on theoretical frequency is less then 10 Hence

$$x^2 = \sum \frac{(Fo - Fe)^2}{Fe}$$

$$x^2 = \frac{(30-20)^2}{20} + \frac{(25-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(10-20)^2}{20} + \frac{(22-20)^2}{20} + \frac{(15-20)^2}{20}$$

$= \frac{1}{20}[100+25+4+100+4+25]$

$= \frac{258}{20}$

$= 12.9$

Degree of freedom $V = 6-1 = 5$

$\chi^2_{0.5} = 11.070$ (from table)

$\chi^2_{cal} > \chi^2_{0.5}(5)$ hypothesis of equal probability is Rejected

Answer :- 7

(a) Let

y: monthly cost (in Lakhs)

x: monthley outputs (in Tons)

$\because \bar{x}$ = mean of monthly output

$\because \bar{x} = \frac{1+2+4+8+6+5+8+9+7}{9} \Rightarrow \frac{50}{9}$

$\because \bar{y}$ = mean of monthly cost

$\therefore \bar{y} = \frac{2+3+4+7+6+5+8+8+6}{9} = \frac{49}{9}$

x_i	x_i^2	y_i	y_i^2	$x_i y_i$
1	1	2	4	2
2	4	3	9	6
4	16	4	16	16
8	64	7	49	56
6	36	6	36	36
5	25	5	25	25
8	64	8	64	64
9	81	8	64	72
7	49	6	36	42
	$\sum x_i^2 = 340$			$\sum x_i y_i = 319$

$S_{xx} = \left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2 = 9$ (Total observations)]

$= 340 - 9 \times \frac{50}{9} \times \frac{49}{9}$

$= \frac{560}{9}$

Calculation of Sum S_{xy}

$S_{xy} = \left(\sum_{i=1}^{n} x_i y_i\right) - n\bar{x}\bar{y}$ [$n = 9$ (Total observations)]

$= 319 - 9 \times \frac{50}{9} \times \frac{49}{9}$

$= \frac{421}{9}$

Now, $b = \frac{S_{xy}}{S_{xx}} \Rightarrow \frac{\frac{421}{9}}{\frac{560}{9}} = 0.752$

Correspondingly we have

$a = \bar{y} - b\bar{x}$

$= \frac{49}{9} - 0.752 \times \frac{50}{9}$

$= 1.266$

$\therefore$ The best linear regression line is

$\Rightarrow y = a + bx$

$\therefore y = 1.266 + 0.752x$ Ans

(b) If the firm decides to produce 4 tons per month then the predicted cost is given by -

$\Rightarrow y = 1.266 + 0.752 \times 4$ [$\because$ Here $x = 4$ tons/month

$\therefore y = 4.274$ lakhs Ans

Answer :-8

Given, $P(A \cup B) = 0.5$

$P(A \cap B^c) = \frac{3}{25}$ and $P(A^c \cap B) = \frac{8}{25}$

We have $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= P(A \cup B) = P(A) + P(B) \cdot P(B)$ [∴ A and B are two independent event ∴ $P(A \cap B) = P(A).P(B)$]

$= P(A \cup B) = P(A)\{1 - P(B)\} + P(B)$

$= P(A \cup B) = P(A \cap B^c) + P(B)$

$= 0.5 = \frac{3}{25} + P(B)$

$= P(B) = 0.38$

Again, We have, $P(A \cup B) = P(B) - P(A \cap B)$

$= P(A \cup B) = P(A) + P(B) - P(A).P(B)$ [∴ A and B are two independent events ∵ $P(A \cap B) = P(A).P(B)$]

$= P(A \cup B) = P(B)\{1 - P(A)\} + P(A)$

$= P(A \cup B) = P(B \cap A^c) + P(A)$

$= P(A \cup B) = P(A^c \cap B) + P(A)$

$= 0.5 = \frac{8}{25} + P(A)$

$\therefore P(A) = 0.18$ Ans

Answer :- 9

(a) The number of sample of size 2 is given by

$^nC_2 = {}^6C_2$ [n = number of persons = 6

$= (6\times5)(2\times1) = 15$ Ans

(b) Samples of size 2 are given in the following table:

Sample	$\bar{x}_i$
54, 50	52
54, 52	53
54, 48	51
54, 50	52
54, 52	53

Sample	$\bar{x}_i$
50, 52	51
50, 48	49
50, 50	50
50, 52	51
52, 48	50

Sample	$\bar{x}_i$
52, 50	51
52, 52	52
48, 50	49
48, 52	50
50, 52	51

Answers :- 10

observation	ordinary Petrol	Premium Petrol	Super petrol
1	1	2	4
2	2	2	1
3	2	1	2
4	1	3	3

Solving using one-way ANOVA method:

ordinary petrol (A)	Premium petrol (B)	Super petrol (C)
1	2	4
2	2	1
2	1	2
1	3	3
$\sum A = 6$	$\sum B = 8$	$\sum C = 10$

A^2	B^2	C^2
1	4	16
4	4	1
4	1	4
1	9	9
$\sum A^2 = 10$	$\sum B^2 = 18$	$\sum C^2 = 30$

Date table

Group	A	B	C	Total
N	$n_1 = 4$	$n_2 = 4$	$n_3 = 4$	$n = 12$
$\sum x_i$	$T_1 = \sum x_1 = 6$	$T_2 = \sum x_2 = 8$	$T_3 = \sum x_3 = 10$	$\sum x = 24$
$\sum x_i^2$	$\sum x_1^2 = 10$	$\sum x_2^2 = 18$	$\sum x_3^2 = 30$	$\sum x^2 = 58$
Mean $\bar{x}_i$	$\bar{x}_1 = 1.5$	$\bar{x}_2 = 2$	$\bar{x}_3 = 2.5$	overall $\bar{x} = 2$
std dev S_i	$S_1 = 0.5774$	$S_2 = 0.8165$	$S_3 = 1.291$	

Let K = the number of different Samples = 3

$n = n_1 + n_2 + n_3 = 4+4+4 = 12$

Grand mean or overall

$$\bar{x} = \frac{24}{12} = 2$$

$$= \sum x = T_1 + T_2 + T_3 = 6+8+10 = 24 \quad ...(i)$$

$$= \frac{(\sum x)^2}{n} = \frac{24^2}{12} = 48 \quad ...(ii)$$

$$= \sum \frac{T_i^2}{n_i} = \left(\frac{6^2}{4} + \frac{8^2}{4} + \frac{10^2}{4}\right) = 50 \quad ...(iii)$$

$$= \sum x^2 = \sum x_1^2 + \sum x_2^2 + \sum x_3^2 = 10+18+30 = 58 \quad ...(iv)$$

Calculation of Degee of Freedom (df)

Degree of Freedom between Samples ($df_{between}$)

$df_{between} = K-1 = 3-1 = 2$ [K = no of columns]

Degree of Freedom Within Samples (df_{within})

$df_{within} = n - K = 12 - 3 = 9$

Total degree of freedom (df_{total})

$df_{total} = df_{between} + df_{within}$

$= 2+9 = 11$

ANOVA :

Step 1 : Sum of Squares between samples (SSB)

$$SSB = \left(\frac{T_i^2}{n_i}\right) - \frac{(\sum x)^2}{n} = Eq(iii) - Eq(ii)$$

$= 50 - 48 = 2$

or

$SSB = \sum n_i(\bar{x}_i - \bar{x})^2$

$= 4(1.5-2)^3 + 4(2-2)^3 + 4(2.5-2)^3 = 2$

Step 2 Sum of Squares Within Samples (SSW)

$SSW = \sum x^2 - \left(\sum \frac{T_i^2}{n_i}\right) = Eq(iv) - Eq(ii)$

$= 58 - 50$

$= 8$

Step 3 Total Sum of Squares (SST)

$SST = SSB + SSW$

$= 2 + 8$

$= 10$

Step 4 Mean Square between Samples

$MSB = \frac{SSB}{df_{between}}$

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

Step 5 Mean Square Within Samples

$MSW = \frac{SSW}{df_{within}}$

$= \frac{8}{9}$

$= 0.8889$

Step 6 Test statistic F for one way ANOVA test

$$F = \frac{MSB}{MSW}$$

$$= \frac{1}{0.8889}$$

$$= 1.125$$

ANOVA Table						
Source of variation	SS	df	MS	F	P-value	F crit
Between Groups (treatment)	2	2	1	1.125	0.366357	4.256495
Within Groups (Error)	8	9	0.888889			
Total	10	10				

H_0: There is no Significant differentiating between extra mileage Samples

H_1: There is significant differentiating between extra mileage Samples

F(2,9) at 0.05 level of Significance = 4.2565

As calculated F = 1.125 < 4.2565

(i) So, H_0 is accepted Hence there is no Significant differenting between extra mileage of Premium and Super petrol.

(ii) The error variance is 0.888889

Answer 8-11

Total number of balls $= 7 + 3 = 10$

The probability that the first floppy drawn is defective $P(A) = \frac{3}{10}$

The Probability that the first floppy drawn is good $P(A^c) = 1 - \frac{3}{10} = \frac{7}{10}$.

(i) The Probability that the second floppy drawn is defective if the first floppy drawn is defective (case of without replacement)

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{2}{9} \text{ Ans}$$

Because 1 defective floppy 10 floppies drawn

The Probability that the second floppy drawn is defective if the first floppy drawn is defective (case of without replacement)

$$P(B|A^c) = \frac{P(B \cap A^c)}{P(A)} = \frac{3}{9} \text{ Ans}$$

Because 1 good floppy from floppies was drawn.

(ii) We know that.

$= P(B \cap A) = P(A)\,P(B|A)$

$= P(B \cap A^c) = P(A)\,P(B|A^c)$

$=$ Now $P(B|A)\,P(A) + P(B|A^c)\,P(A^c)$

$= P(B \cap A) + P(B \cap A^c)$

$= P(B \cap A) + [P(B) - (B \cap A)]$ Since $P(B \cap A^c) = P(B) - P(B \cap A)$

$= P(B \cap A) + P(B) - P(B \cap A)$

$= P(B)$

Hence, $P(B) = P(B/A).P(A) + P(B/A^c)P(A^c)$

Now, on substituting the know probabilities we must have

$P(B) = P(B/A)P(A) + P(B/A^c)P(A^c)$

$= \frac{2}{9} \times \frac{3}{10} \times \frac{3}{9} \times \frac{7}{10}$

$= \frac{6+21}{90} = \frac{27}{90} = \frac{3}{10} = P(A)$

Hence, $P,(B) = P(B/A) + P(B/A^c)P(A^c) = P(A)$ (Proved)

Answer :-12

$$\mu \text{ or } \bar{x} = 0.5230, \sigma = 0.0032, n = 4$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.0032}{\sqrt{4}} = 0.0016$$

2 Sigma (Internal control) Limits

$$UCL_{\bar{x}} = \mu + 2\sigma_{\bar{x}} = 0.5230 + 2 \times 0.0016 = 0.5262 \text{cm}$$

$$LCL_{\bar{x}} = \mu - 2\sigma_{\bar{x}} = 0.5230 - 2 \times 0.0016 = 0.5198 \text{cm}$$

3 Sigma (outer control) Limits

$$UCL_{\bar{x}} = \mu + 3\sigma_{\bar{x}} = 0.5230 + 2 \times 0.0016 = 0.5278 \text{ cm}$$

$$LCL_{\bar{x}} = \mu - 3\sigma_{\bar{x}} = 0.5230 - 2 \times 1.0016 = 0.5182 \text{ cm}$$

Based on the above calculation the control charts is the Fallowing

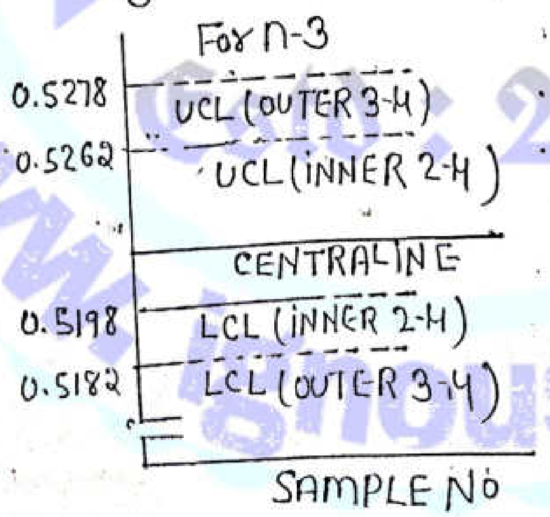
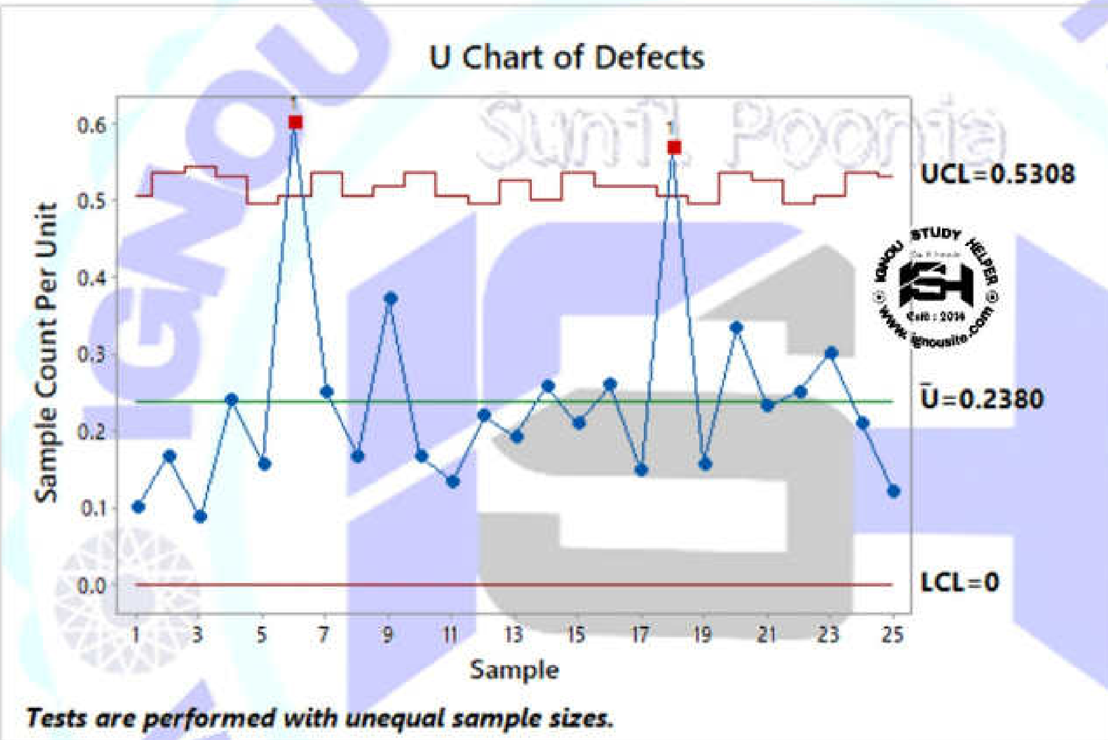


Fig control chart

Ans.13 Control charts: A control chart is a graph which displays all the process data in order sequence. It consists of a centre line, the upper limit and lower limit. Centre line of a chart represents the process average. Control limits (upper & lower) which are in a horizontal line below and above the center line depicts whether the process is in control or out of control. Control limits are based on process variation.

Use control charts:

- To examine whether the process is stable or not.
- To understand the process variation over time.
- When you need to find out any variation occurs and fixed it instantaneously.
- To find out whether the process is within the statistical control or not (Due to chance or assignable causes).



U chart used in an LCD manufacturing

Utilities of Control Charts: Control charts provide the greatest benefits in large-scale, repetitive processes. The following examples illustrate the application of control charts to some typical accounting, auditing, and tax processes. Of course, a cost/benefit analysis should be made before deciding whether to implement control charts for a given accounting process.

Internal Auditing: A function of internal auditing is to determine the extent to which company policies and procedures are being followed. Since many of the procedures are repetitive, internal auditors can use control charts to monitor various accounting processes throughout the year. Processes that can be monitored via control charts include payroll accounting, invoice payments, and revenue collection. Payroll accounting is illustrated here.

The repetitive procedures involved in payroll accounting include the proper authorization of overtime, approval of time records, and checks on the calculation of gross pay and payroll deductions. To ensure that the internal controls over these procedures are in effect throughout the year, internal auditors might sample individual employee pay records during each pay period. For each employee pay record in a sample, the internal auditors would verify documentation of the required authorizations and approvals. A missing authorization or approval constitutes an audit exception or nonconformity.

Customer Billing: Control charts may also be useful in helping a firm improve its financial position. For example, one method of improving cash position is to speed up collections from customers. Reducing the amount of time between the point that goods (services) are provided and the point that customers are billed is a key step in speeding up collections.

In using control charts to evaluate the customer billing process, a company may periodically draw random samples of invoices. For each invoice, the amount of time between the delivery date and billing date is measured.

Tax Return Preparation: Although the preceding illustrations have focused on accounting procedures within a company, public accountants can also use control chart techniques for evaluating internal controls of audit clients or for evaluating their firm's internal billing process. In addition, public accounting firms may also find unique uses for control charts, such as in tax return preparation. For example, if computers are used for preparing tax returns, the accountant takes information provided by the taxpayer and transfers it to computer input sheets. The input sheets may be sent to a tax service which generates the tax return, or the return may be generated in-house. Improper completion of an input form results in an incorrect tax return, and the process must be repeated. The end result is lower profits for the firm.

The preparation of computer input forms is a process that can be graphed on control charts. The main issue is whether or not the input forms are correct; the number of errors on the forms is not as important since even a single error results in an unusable tax return.

Travel and Entertainment Expenses: Organizations can use control charts to manage costs and expenses. Effective monitoring of expenses includes controls that ensure that expenses are properly authorized and documented, and controls that monitor the level of expenses.

Travel and entertainment is an expense that should be monitored for both cost control and tax reasons. Selection of the appropriate control chart depends upon whether the primary focus is on monitoring the level of expenses, or on monitoring whether travel and entertainment expenses have been properly authorized and documented.

Ans.14

a) Cluster sampling, Stratifies sampling and Systematic sampling

Cluster sampling: Cluster sampling when used, gives every unit/person in the population an equal and known chance of being selected in the sample group.

For this method of sampling, researchers divide the population into internally heterogeneous and externally homogeneous subpopulations known as clusters. The clusters are externally homogeneous as they appear to be grouped together by a shared characteristic/criteria but are internally heterogeneous because the subpopulations within the clusters have different compositions.

Clusters may be divided by different cities in a country, different areas in a city, different organizations, different universities, different industrial estates, etc. After these clusters have been decided, researchers select certain clusters and eliminate the rest. For example, if you're conducting a study across all cities in the United States, you can use cluster sampling to eliminate certain cities, or clusters, in order to select your final sample group.

Stratifies sampling: Stratified sampling is a method of obtaining a representative sample from a population that researchers have divided into relatively similar subpopulations (strata). Researchers use stratified sampling to ensure specific subgroups are present in their sample. It also helps them obtain precise estimates of each group's characteristics. Many surveys use this method to understand differences between subpopulations better. Stratified sampling is also known as stratified random sampling.

The stratified sampling process starts with researchers dividing a diverse population into relatively homogeneous groups called strata, the plural of stratum. Then, they draw a random sample from each group (stratum) and combine them to form their complete representative sample. Learn more about representative samples.

Strata are subpopulations whose members are relatively similar to each other compared to the broader population. Researchers can create strata based on income, gender, and race, among many other possibilities. For example, if your research question requires you to compare outcomes between income levels, you might base the strata on income. All members of the population should be in only one stratum.

Systematic sampling: Systematic sampling is a type of probability sampling method in which sample members from a larger population are selected according to a random starting point but with a fixed, periodic interval. This interval, called the sampling interval, is calculated by dividing the population size by the desired sample size. Despite the sample population being selected in advance, systematic sampling is still thought of as being random if the periodic interval is determined beforehand and the starting point is random.

When carried out correctly on a large population of a defined size, systematic sampling can help researchers, including marketing and sales professionals, obtain representative findings on a huge group of people without having to reach out to each and every one of them.

b) Parametric and Non-Parametric Tests:

Parametric Test: In Statistics, a parametric test is a kind of the hypothesis test which gives generalizations for generating records regarding the mean of the primary/original population. The t-test is carried out based on the students t-statistic, which is often used in that value.

The t-statistic test holds on the underlying hypothesis which includes the normal distribution of a variable. In this case, the mean is known, or it is considered to be known. For finding the sample from the population, population variance is identified. It is hypothesized that the variables of concern in the population are estimated on an interval scale.

Non-Parametric Test: The non-parametric test does not require any population distribution, which is meant by distinct parameters. It is also a kind of hypothesis test, which is not based on the underlying hypothesis. In the case of the non-parametric test, the test is based on the differences in the median. So, this kind of test is also called a distribution-free test. The test variables are determined on the nominal or ordinal level. If the independent variables are non-metric, the non-parametric test is usually performed.

Ans.15

a) Goodness of fit test: The goodness-of-fit test is a statistical hypothesis test to see how well sample data fit a distribution from a population with a normal distribution. Put differently, this test shows if your sample data represents the data you would expect to find in the actual population or if it is somehow skewed. Goodness-of-fit establishes the discrepancy between the observed values and those that would be expected of the model in a normal distribution case.

Goodness-of-fit tests are statistical methods often used to make inferences about observed values. These tests determine how related actual values are to the predicted values in a model, and when used in decision-making, goodness-of-fit tests can help predict future trends and patterns.

The most common goodness-of-fit test is the chi-square test, typically used for discrete distributions. The chi-square test is used exclusively for data put into classes (bins), and it requires a sufficient sample size to produce accurate results.

Example: Suppose we have bags of balls with five different colours in each bag. The given condition is that the bag should contain an equal number of balls of each colour. The idea we would like to test here is that the proportions of the five colours of balls in each bag must be exact.

b) Test of Independence: The Chi-square test of independence checks whether two variables are likely to be related or not. We have counts for two categorical or nominal variables. We also have an idea that the two variables are not related. The test gives us a way to decide if our idea is plausible or not.

Example: We have a list of movie genres; this is our first variable. Our second variable is whether or not the patrons of those genres bought snacks at the theater. Our idea (or, in statistical terms, our null hypothesis) is that the type of movie and whether or not people bought snacks are unrelated. The owner of the movie theater wants to estimate how many snacks to buy. If movie type and snack purchases are unrelated, estimating will be simpler than if the movie types impact snack sales.

c) Criteria for a good estimator: In statistics, an estimator is a function of the observable sample data that is used to estimate an unknown population parameter; an estimate is the result from the actual application of the function to a particular set of data. Many different estimators are possible for any given parameter. Some criterion is used to choose between the estimators, although it is often the case that a criterion cannot be used to clearly pick one estimator over another.

Example: If your estimates of the population mean μ are say, 10, and 11.2 from two independent samples of sizes 20 and 30 respectively, then a better estimate of the population mean μ based on both samples is [20 (10) + 30 (11.2)]/(20 + 30) = 10.75 (10.72, actually).

d) Chi-Square Test: A chi-squared test (symbolically represented as $\chi 2$) is basically a data analysis on the basis of observations of a random set of variables. Usually, it is a comparison of two statistical data sets. This test was introduced by Karl Pearson in 1900 for categorical data analysis and distribution. So it was mentioned as Pearson's chi-squared test.

The chi-square test is used to estimate how likely the observations that are made would be, by considering the assumption of the null hypothesis as true.

A hypothesis is a consideration that a given condition or statement might be true, which we can test afterwards. Chi-squared tests are usually created from a sum of squared falsities or errors over the sample variance.

Example: Let us take an example of a categorical data where there is a society of 1000 residents with four neighborhoods, P, Q, R and S. A random sample of 650 residents of the society is taken whose occupations are doctors, engineers and teachers. The null hypothesis is that each person's neighborhood of residency is independent of the person's professional division. The data are categorized as:

Categories	P	Q	R	S	Total
Doctors	90	60	104	95	349
Engineers	30	50	51	20	151
Teachers	30	40	45	35	150
Total	150	150	200	150	650

Assume the sample living in neighborhood P, 150, to estimate what proportion of the whole 1,000 people live in neighborhood P. In the same way, we take 349/650 to calculate what ratio of the 1,000 are doctors. By the supposition of independence under the hypothesis, we should "expect" the number of doctors in neighborhood P is;

$150 \times 349/650 \approx 80.54$

So by the chi-square test formula for that particular cell in the table, we get;

$(\text{Observed} - \text{Expected})^2/\text{Expected Value} = (90-80.54)^2/80.54 \approx 1.11$