Practical Manual

STATISTICAL METHODS

Dr. Shailendra Kumar

ABB 254 2(1+1)

2020

College of Agriculture
Rani Lakshmi Bai Central Agricultural University,
Jhansi-284003

Syllabus ABB 254 3(2+1): Graphical Representation of Data. Measures of Central Tendency (Ungrouped data) with Calculation of Quartiles, Deciles & Percentiles. Measures of Central Tendency (Grouped data) with Calculation of Quartiles, Deciles & Percentiles. Measures of Dispersion (Ungrouped Data). Measures of Dispersion (Grouped Data). Moments, Measures of Skewness & Kurtosis (Ungrouped Data). Moments, Measures of Skewness & Kurtosis (Grouped Data). Correlation & Regression Analysis. Application of One Sample t-test. Application of Two Sample Fisher's t-test. Chi-Square test of Goodness of Fit. Chi-Square test of Independence of Attributes for 2 ׀׀2 contingency table. Analysis of Variance One Way Classification. Analysis of Variance Two Way Classification. Selection of random sample using Simple Random Sampling.

Name of Students……………………………………………………………………………

Roll No………………………………………………….Batch……………………………….

Session…………………………………………………………………Semester…………..

Course Name: ……………………………………………………………………………

Course No…………………………………………………………………………………….

Credit…………………………………………………………………………………………….

No. of Copies:

Price: Rs.

@ RLBCAU, Jhansi

CERTIFICATE

This is to certified that Shri./Km. …………………………………………………. ID. No………..………

has completed the practical of course ……………………………………..Course No……………………..

as per the syllabus of B.Sc. (Hons.) Agriculture…… semester in the year…………………………….in the respective lab of College.

Date: Course Teacher

Contents

S. N.	Practical Name	Page No.
1	To draw Histogram, Polygon, Frequency Curve and Ogive	
2	To calculate Mean, Median, Mode, Geometric mean and Harmonic mean of ungrouped data	
3	To calculate Quartiles, Deciles and Percentiles of ungrouped data.	
4	To calculate Mean, Median, Mode, Geometric mean and Harmonic mean of grouped data.	
5	To calculate Quartiles, Deciles and Percentiles of grouped data	
6	To calculate range, quartile deviation, mean deviation and standard deviation and coefficient of variance (CV) for ungrouped data.	
7	To calculate range, quartile deviation, mean deviation, standard deviation and coefficient of variance (CV) for grouped data	
8	To calculate first four central moments, coefficient of Skewness (β_1) and Kurtosis (β_2) for ungrouped data. Comment on nature of the data.	
9	To calculate first four central moments, coefficient of Skewness (β_1) and Kurtosis (β_2) for grouped data. Comment on nature of the data	
10	To calculate correlation co-efficient between two variables	
11	To fit linear regression equation on the given data	
12	To test significant difference between sample mean and population mean using t-test	
13	To test significant difference between two sample means for independent samples and paired samples using t-test	
14	To test goodness fit of the distribution and association between two attributes using chi Square test	
15	To test goodness fit of the distribution and association between two attributes using chi Square test.	
16	To compare several sample means using one-way ANOVA	
17	To test equality of several treatment means using two-way ANOVA	
18	To draw a simple random sample and estimate population mean and population variance.	

Practical No. 1

Objective: To draw Histogram, Polygon, Frequency Curve and Ogive

Problem: The following data is related to milk production in liter of 225 farmers of a village.

Milk in liter(x_i)	0 - 4	4 - 8	8 - 12	12-16	16 - 20	20-24	24-28
No. of farmer (f_i):	30	90	60	12	8	4	1

Draw histogram, polygon, frequency curve, and ogive of the given data.

Solution:

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

Practical No. 2

Objective: To calculate Mean, Median, Mode, Geometric mean and Harmonic mean of ungrouped data.

Problem: The following data is related to crop yield in quintals of last 7 years of a farmer. Find the arithmetic mean of crop yield?

112, 124, 104, 140, 136, 132, 144.

Solution: Step –I: count number of observations, n =

Step-II: calculate the total of all of observation (or values) = $\sum_{i=1}^{n} x_i$

= ....

Step-III: Finally, we calculate the arithmetic mean of the following formula,

$\overline{x}$ = $\frac{\sum_{i=1}^{n} x_i}{n}$ =

....

Problem: The following data is related to plant height in cm. Find the median height of the plants?

12, 14, 18, 15, 21, 22, 24, 25, 3

Solution:

Step-I: arrange the data in ascending or descending order

....

Step –II: count number of observation (or values) = n =...

Step –III: Since n is odd number so we find median by the following formula:

Median = $(\frac{n+1}{2})^{th}$ $item$ =..

....

Problem: The following data is the marks of 12 students out of 30 in Mid-term examination. To find the median marks of the students?

19, 20, 22, 12, 14, 18, 15, 21, 22, 24, 25, 2

Solution: Step-I: arrange the data in ascending order

....

Step –II: count number of observation (or values) = n =...

Step –III: Since n is even number so we find median by the following formula:

Median = $\frac{(\frac{n}{2})^{th}\ item + (\frac{n}{2}+1)^{th}\ item}{2}$ =...

....

Problem: Suppose that 15 student's shoes size number are following:

8, 7, 6, 7, 10, 7, 8, 9, 9, 7, 7, 7, 8, 7, 7

Find the mode of size of shoes?

Solution: Mode =...

Problem: The following data are related to Cow population in a village every fifth year.

No. of fifth year	First	Second	Third	Fourth	Fifth	Six
Population	80	120	185	300	448	680

Find the geometric mean population of cow?

Solution: Step-I

x_i	log x_i
	$\sum_{i=1}^{n} logx_i$=

Step-II: Geometric mean, G=Antilog [$\frac{\sum_{i=1}^{n} logx_i}{n}$] =...

....

....

Problem: Find Harmonic mean of the given data 1/3, 1/4, 1/5, 1/6, 1/7

Solution: Step- I:

x_i	$\frac{1}{x_i}$
	$\sum_{i=1}^{n} \frac{1}{x_i}$=

Step-II: Harmonic mean, H = $\frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$ =..

....

Practical No-3

Objective: To calculate Quartiles, Deciles and Percentiles of ungrouped data.

Problem: The following data is related to plant height in cm. Compute the first quartile (Q_1) height of the plants?

20, 22, 12, 14, 18, 15, 21, 22, 24, 25, 3, 10, 26, 28, 32, 33, 39

Solution:

Step-I: arrange the data in ascending order

....

Step –II: count number of observation (or values), n =..

Step –III: First Quartile (Q1) = $[(\frac{n+1}{4})]^{th}\ item$ =..

....

....

Problem: The following data is height children in cm. compute the 5^{th}Deciles height of the children?

23, 31, 32, 20, 22, 12, 14, 18, 15, 21, 22, 24, 25, 3, 10, 26, 28, 32, 33, 39, 28, 29

Solution: Step-I: arrange the data in ascending order

....

Step –II: count number of observation (or values), n =..

Step –III: k^{th} Deciles = $[k(\frac{n+1}{10})]^{th}\ item$ =..

5^{th}Deciles = $[5(\frac{n+1}{10})]^{th}\ item$ =..

....

....

Problem: The following data is related to plant height in cm. Compute the 20^{th} percentile height of the plants?

20, 22, 12, 14, 18, 15, 21, 22, 24, 25, 3, 10, 26, 28, 32, 33, 39

Solution: Step-I: arrange the data in ascending order

....

Step –II: count number of observation (or values), n =..

Step –III: k^{th} percentile = $[k(\frac{n+1}{100})]^{th}\ item$ =..

20^{th} percentile = $[20(\frac{n+1}{100})]^{th}\ item$ =..

....

Practical No. 4

Objective: To calculate Mean, Median, Mode, Geometric mean and Harmonic mean of grouped data.

Problem: The following data is related to yearly income of 40 farmers of a village which are selected randomly. To find the Arithmetic Mean income of farmers of the given village?

Income in Lakh (x_i):	2 8	3	4	5	6	7
No. of farmer (f_i):	1 3	3	6	12	8	7

Solution: Step-I to construct the following table:

Variable (x_i)	Frequency (f_i)	$f_i \times x_i$
Total	$\sum_{i=1}^{n} f_i$=	$\sum_{i=1}^{n} f_i x_i$=

Step-II: To calculate the arithmetic mean, $\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$ =..

Problem: The following data is related to milk production in liter of 240 farmers of a village. Calculate arithmetic mean milk production of the given village?

Milk in liter(x_i)	0 - 4	4 - 8	8 - 12	12-16	16 - 20
No. of farmer (f_i):	30	90	60	40	20

Solution: Step-I: to construct the following table:

Class Interval ($x_i - x_{i+1}$)	**Frequency (f_i)**	**Mid Value of Class Interval** $m_i=\frac{x_i+x_{i+1}}{2}$	$f_i \times m_i$
Total	$N=\sum_{i=1}^{n} f_i$=		$\sum_{i=1}^{n} f_i m_i$=

Step-II: To calculate the arithmetic of the following formula, $\overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$=..

……….….

……………………………………………………………………………………………………….…...

Problem: The following data is related to yearly income of 50 farmers of a village which are selected randomly. To find the median income of farmers of the given village?

Income in Lakh (x_i):	2	3	4	5	6	7	8
No. of farmer (f_i):	1	3	6	12	18	7	3

Solution: Step-I: To construct the following table

Variable (x_i)	Frequency(f_i)	Cumulative Frequency
Total	N=$\sum_{i=1}^{n} f_i$=	

Step-II: To calculate $\frac{N}{2}$=...

Step-III: To see the value just greater than $\frac{N}{2}$ in cumulative frequency column =......................................

Step-IV: In first column, see the value of x correspond to the value getting in step-III,

Median =...

Problem: The following data is related to height of tree in feet of 250 trees of a forest which are selected randomly. To calculate median height of tree of the given forest?

Height of tree (x_i)	35-40	40-45	45-50	50-55	55-60
No. of tree (f_i):	30	70	90	40	20

Solution: Step-I to construct the following table:

Class Interval	**Frequency (f_i)**	**Cumulative frequency**

Total	N=$\sum_{i=1}^{n} f_i$=	

Step-II: To calculate $\frac{N}{2}$=...

Step-III: To see the value just greater than $\frac{N}{2}$ in cumulative frequency column =........................

Step-IV: corresponding class of step-III is called middle class

Step-V: Median, M_d = L + [h ($\frac{\frac{N}{2}-C}{f}$)] Where L is lower limit of middle class, h is magnitude value of class interval, f is frequency of middle class and C is cumulative frequency of preceding middle class.

Median, M_d=...

....

Problem: Following table gives the category of 250 trees in a forest.

Tree's Name	A	B	C	D	E
No. of trees	12	150	50	20	18

Find the measure of central tendency (mode) category of trees.

Solution:

Mode =..

Problem: The following data is related to milk production in liter of 225 farmers of a village. Calculate mode milk production of the given village?

Milk in liter(x_i)	0 - 4	4 - 8	8 - 12	12-16	16 - 20	20-24	24-28
No. of farmer (f_i):	30	90	60	12	8	4	1

Solution: Step-I: See the maximum frequency. Correspond class to maximum frequency is called model class.,f_1=................................

Step-II: Mode = L + [$\frac{h(f_1-f_0)}{2f_1-f_0-f_2}$] where f_1 is frequency of model class, f_0 is frequency of preceding model class, f_2 is frequency of succeeding model class, L is lower limit of model class and h is magnitude value of class interval.

Mode=...

...

....

....

Problem: Find the Geometric mean of the given data:

x_i	2	4	8	16	32
f_i	4	5	7	6	3

Solution: Step-I:

x_i	f_i	log x_i	f_i log x_i
	N=$\sum_{i=1}^{n} f_i$=		$\sum_{i=1}^{n} f_i log x_i$=

Step-II:

Geometric mean, G= Antilog $[\frac{\sum_{i=1}^{n} f_i log x_i}{N}]$ =...

....

....

Problem: Find the Geometric mean of the given data:

x_i	0 - 4	4 – 8	8 - 12	12-16	16 - 20	20-24	24-28
f_i	5	7	10	15	11	8	4

Solution:

Step-I: To construct table:

Class Interval (x_i – x_{i+1})	frequency (f_i)	Mid-Value m_i=$\frac{x_i+x_{i+1}}{2}$	log m_i	f_i log m_i
	N=$\sum_{i=1}^{n} f_i$=			$\sum_{i=1}^{n} f_i log m_i$=

Step-II: Geometric mean, G= Antilog $[\frac{\sum_{i=1}^{n} f_i log m_i}{N}]$ =...

....

....

Problem: A person went from City-A to City-B by different transport mode which speed and cover distance are given below:

transport mode	by foot	Taxi	train	Airplane	taxi
Speeds in km/h(x_i)	5	30	70	800	40
distances in km (f_i)	1	10	430	1200	50

Find the average speed (Harmonic Mean) of the complete journey.

Solution: Step-I:

x_i	f_i	$\frac{1}{x_i}$	$\frac{f_i}{x_i}$
	$N=\sum_{i=1}^{n} f_i$=		$\sum_{i=1}^{n} \frac{f_i}{x_i}$=

Step-II: Harmonic mean, H = $\frac{\sum_{i=1}^{n} \frac{f_i}{x_i}}{N}$ =..

....

Problem: Find the Harmonic mean of the given data:

x_i	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f_i	2	3	5	8	6	4	2

Solution:

Step-I:

Class Interval (x_i – x_{i+1})	frequency (f_i)	Mid-Value $m_i=\frac{x_i+x_{i+1}}{2}$	$\frac{1}{m_i}$	$\frac{f_i}{m_i}$
	$N=\sum_{i=1}^{n} f_i$=			$\sum_{i=1}^{n} \frac{f_i}{x_i}$ =

Step-II: Harmonic mean, H = $\frac{\sum_{i=1}^{n} \frac{f_i}{m_i}}{N}$ =...

....

....

Practical No. 5

Objective: To calculate Quartiles, Deciles and Percentiles of grouped data.

Problem: The following data is related to yearly income of 80 farmers of a village. Compute the Third Quartile (Q_3) income of farmers of the given village?

Income in Lakh (x_i):	2	3	4	5	6	7	8	9	10
No. of farmer (f_i):	1	3	6	12	18	1 7	13	8	2

Solution: Step-I: To construct the following table

Variable (x_i)	frequency(f_i)	cumulative frequency
Total	N=$\sum_{i=1}^{n} f_i$=	

Step-II: To calculate $\frac{3N}{4}$ =..

Step-III: To see the value just greater than $\frac{3N}{4}$ in cumulative frequency column =............................

Step-IV: In first column, see the value of x correspond to the value getting in step-III,

Third Quartile (Q_3) =..

Problem: The following data is related to height of tree in feet of 270 trees of a forest. Calculate Third Quartile (Q_3) height of tree of the given forest?

Height of tree (x_i)	35-40	40-45	45-50	50-55	55-60	60-65	65-70
No. of tree (f_i):	30	70	90	40	20	12	8

Solution: Step-I: To construct the following table

Variable (x_i)	frequency(f_i)	cumulative frequency

Total	$N=\sum_{i=1}^{n} f_i=$	

Step-II: To calculate $\frac{3N}{4}$ =...

Step-III: To see the value just greater than $\frac{3N}{4}$ in cumulative frequency column =..........................

Step-IV: Corresponding class is called Third Quartile class

Step-V: Third Quartile (Q_3) = L +[$\frac{\frac{3N}{4}-C}{f}$]$\times h$ =……………………………............................

Problem: The following data is related to yearly income of 80 farmers of a village. Compute the 7th Deciles income of farmers of the given village?

Income in Lakh (x_i):	2	3	4	5	6	7	8	9	10
No. of farmer (f_i):	1	3	6	12	18	1 7	13	8	2

Solution: Step-I: To construct the following table

Variable (x_i)	frequency(f_i)	cumulative frequency
Total	$N=\sum_{i=1}^{n} f_i=$	

Step-II: To calculate $\frac{7N}{10}$ =..............

Step-III: To see the value just greater than $\frac{7N}{10}$ in cumulative frequency column =..........

Step-IV: In first column, see the value of x correspond to the value getting in step-III,

7th Deciles =...

Problem: The following data is related to height of tree in feet of 270 trees of a forest. To calculate 6th Deciles height of tree of the given forest?

Height of tree (x_i)	35-40	40-45	45-50	50-55	55-60	60-65	65-70
No. of tree (f_i):	30	70	90	40	20	12	8

Solution: Step-I to construct the following table:

Class Interval	**Frequency (f_i)**	**cumulative frequency**
Total	N=$\sum_{i=1}^{n} f_i$=	

Step-II: To calculate $\frac{6N}{10}$ =..

Step-III: To see the value just greater than $\frac{6.N}{10}$ in cumulative frequency column =........................

Step-IV: corresponding class of step-III is called 6th Decilesclass

Step-V: 6th Deciles= L + [h $(\frac{\frac{6N}{10} - C}{f})$] Where L is lower limit of 6th Deciles class, h is magnitude value of class interval, f is frequency of 6th Decilesclass and C is cumulative frequency of preceding 6th Decilesclass.

6th Deciles=L + [h $(\frac{\frac{30.N}{10} - C}{f})$] =...

....

Problem: The following data is related to yearly income of 180 farmers of a village which are selected randomly. Compute the 30th percentile income of farmers of the given village?

Income in Lakh (x_i):	2	3	4	5	6	7	8	9	10
No. of farmer (f_i):	11	13	26	32	48	27	13	8	2

Solution: Step-I: To construct the following table

Variable (x_i)	frequency(f_i)	cumulative frequency

Total	N=$\sum_{i=1}^{n} f_i$=	

Step-II: To calculate $\frac{k.N}{100}$ =..............

Step-III: To see the value just greater than $\frac{k.N}{100}$ in cumulative frequency column =..........

Step-IV: In first column, see the value of x correspond to the value getting in step-III,

k^{th} percentile =

30^{th} percentile =

Problem: The following data is related to height of tree in feet of 270 trees of a forest which are selected randomly. To calculate 30^{th} percentile height of tree of the given forest?

Height of tree (x_i)	35-40	40-45	45-50	50-55	55-60	60-65	65-70
No. of tree (f_i):	30	70	90	40	20	12	8

Solution: Step-I to construct the following table:

Class Interval	**Frequency (f_i)**	**cumulative frequency**
Total	N=$\sum_{i=1}^{n} f_i$=	

Step-II: To calculate $\frac{k.N}{100}$ =..............

Step-III: To see the value just greater than $\frac{k.N}{100}$ in cumulative frequency column =..........

Step-IV: corresponding class of step-III is called k^{th} percentile class

Step-V: k^{th} percentile = L + [h ($\frac{\frac{k.N}{100} - C}{f}$)] Where L is lower limit of kth percentile class, h is magnitude value of class interval, f is frequency of k^{th} percentile class and C is cumulative frequency of preceding k^{th} percentile class.

30^{th} percentile =L + [h ($\frac{\frac{30.N}{100} - C}{f}$)] =...

....

....

Practical No. 6

Objective: To calculate range, quartile deviation, mean deviation and standard deviation and coefficient of variance (CV) for ungrouped data.

Problem: Find the range and its coefficient of the given data set:

8, 12, 11, 34, 22, 5, 10, 35, 28

Solution: Range, R = L – S where L is largest value and S is smallest value of the data

Range, R =...

Coefficient of Range = $\frac{L-S}{L+S}$ =...

Problem: Find the Quartile Deviation of the given data: 4, 8, 12, 11, 34, 22, 5, 10, 35, 28, 30, 35, 26, 27 also find coefficient of quartile deviation.

Solution: Quartile Deviation, Q = $\frac{Q_3-Q_1}{2}$

Coefficient of Quartile Deviation = $\frac{Q_3-Q_1}{Q_3+Q_1}$

....

....

....

....

....

....

....

....

....

....

....

Problem: The height of 10 mango trees are given below. Find the mean deviation about mean.

22.0, 22.5, 22.9, 23.0, 23.2, 23.1, 23.4, 23.0, 24.1, 23.6 in meter.

Solution:

Step-I: To calculate Average A (Mean)

Step-II: Construct a table:

x_i	$\lvert x_i - A \rvert$
	$\sum_{i=1}^{n} \lvert xi - A \rvert$ =

Step-III: Mean Deviation, η = $\frac{\sum_{i=1}^{n} |x_i - A|}{n}$ =...

Coefficient of Mean Deviation = $\frac{\eta}{A}$ =...

Problem: The yield of wheat of 10 plots are given below. Find the standard deviation in yield data: 22.0, 22.5, 22.9, 23.0, 23.2, 23.1, 23.4, 23.0, 24.1 and 23.6 in quintal. Also calculate CV.

Solution:

Step-I: To calculate Arithmetic Mean, $\overline{x}$

Step-II: Construct a table:

x_i	$x_i - \overline{x}$	$(xi - \overline{x})^2$

		$\sum_{i=1}^{n}(\mathrm{xi}-\overline{x})^2 =$

Step-III: Standard Deviation, σ = $\sqrt{\frac{\sum_{i=1}^{n}(\mathrm{xi}-\overline{x})^2}{n}}$ =…………………………………………………………………………

………..

………..

………..

Step-IV: Coefficient of Variance (CV) = $\frac{\sigma}{\overline{x}} \times 100$ =…………………………………………………………………………

………..

………..

………..

………..

………..

………..

Practical No. 7

Objective: To calculate range, quartile deviation, mean deviation, standard deviation and coefficient of variance (CV) for grouped data.

Problem: Find the range and its coefficient of the given data set:

x	10	15	20	25	30	35	40	45	50
f	4	8	12	10	21	17	19	7	3

Solution: Range, R = L – S where L is largest value and S is smallest value of the data

Range, R =...

Coefficient of Range = $\frac{L-S}{L+S}$ =...

Problem: Find the range and its coefficient of the given data set:

x	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
f	2	5	6	9	13	14	6	4	2

Solution: Range, R = L – S where L is largest value and S is smallest value of the data

Range, R =...

Coefficient of Range = $\frac{L-S}{L+S}$ =...

Problem: The following data is related to yearly income of 70 farmers of a village which are selected randomly. Measure the deviation in farmer's incomes using quartile deviation.

Income in Lakh (x_i):	2	3	4	5	6	7	8	9	10	11
No. of farmer (f_i):	1	3	6	12	18	11	7	6	4	2

Solution:...

....

....

....

....

....

....

....

....

....

....

....

....

....

....

Quartile Deviation, Q = $\frac{Q_3 - Q_1}{2}$ =...

Coefficient of Quartile Deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$ =..

Problem: Find the Quartile Deviation of the following frequency distribution:

x	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
f	2	5	6	9	13	14	6	4	2

Solution:..

....

....

....

....

....

....

....

....

....

....

....

....

....

....

Quartile Deviation, Q = $\frac{Q_3 - Q_1}{2}$ =...

Problem: The following data is related to yield of wheat of 60 plots in quintal. To find the mean deviation about median and coefficient of mean deviation in yield.

yield	8	9	10	11	12	13	14	15
frequency	2	5	6	20	10	10	5	2

Solution:

Step-I: To calculate Average A (Median)

....

....

....

....

....

....

....

....

....

....

....

....

....

....

Step-II: Construct table

Variable (x_i)	frequency(f_i)	$\|x_i - A\|$	$f_i\|x_i - A\|$
	$N = \sum_{i=1}^{n} f_i$=		$\sum_{i=1}^{n} f_i \|x_i - A\|$=

Step-III: Mean Deviation, $\eta = \frac{\sum_{i=1}^{n} f_i|x_i - A|}{N}$ =..

....

Problem: The following data is related to height of 200 plants in cm. To calculate mean deviation about mode of the given data.

Height of tree (x_i)	30-32	32-34	34-36	36-38	38-40
No. of tree (f_i):	3	70	90	35	2

Solution:

Step-I: To calculate Average A (Mode)...

....

....

....

....

....

....

....

Step-II: Construct table

Class Interval	frequency(f_i)	Mid value (m_i)	$\|m_i - A\|$	$f_i\|m_i - A\|$
	$N = \sum_{i=1}^{n} f_i$=			$\sum_{i=1}^{n} f_i \|m_i - A\|$=

Step-III: Mean Deviation, η = $\frac{\sum_{i=1}^{n} f_i|m_i-A|}{N}$=...

....

Problem: The following data is related to income of farmer. To find the standard deviation and CV of the data.

Income ('00000') Rs.	8	9	10	11	12	13	14	15
No. of Farmers	2	5	16	20	10	10	5	2

Solution:

Step-I: To calculate Arithmetic Mean, $\overline{x}$ =...

....

Step-II: Construct a table:

x_i	f_i	$x_i - \overline{x}$	$(\text{xi} - \overline{x})^2$	$f_i(\text{xi} - \overline{x})^2$
				$\sum_{i=1}^{n} f_i(\text{xi} - \overline{x})^2 =$

Step-III: Standard Deviation, σ =$\sqrt{\frac{\sum_{i=1}^{n} f_i(\text{xi}-\overline{x})^2}{N}}$ ....

....

Step-IV: Coefficient of Variance (CV) = $\frac{\sigma}{\overline{x}} \times 100$ =..

....

Problem: Find the standard deviation, coefficient of standard deviation, variance and CV of the given data.

Class Interval	10-20	20-30	30-40	40-50	50-60
frequency	8	15	45	20	12

Solution:

Step-I: To calculate Arithmetic Mean, $\overline{x}$ =..

....

Step-II: Construct a table:

Class Interval	f_i	Mid-Value (m_i)	$m_i - \overline{x}$	$(m_i - \overline{x})^2$	$f_i(m_i - \overline{x})^2$
					$\sum_{i=1}^{n} f_i(m_i - \overline{x})^2 =$

Step-III: Standard Deviation, $\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i(m_i-\overline{x})^2}{N}}$..

....

Step-IV: Coefficient of Variance (CV) = $\frac{\sigma}{\overline{x}} \times 100$ =...

....

Practical No. 8

Objective: To calculate first four central moments, coefficient of Skewness (β_1) and Kurtosis (β_2) for ungrouped data. Comment on nature of the data.

Problem: Compute first four central moments and the Karl Pearson's coefficient of skewness (β_1) and Kurtosis (β_2) from the following data: 8, 10, 11, 15, 16, 18, 21, 25, 28, 32, 35, 39, 41 and 42

Solution:

x	(x-$\bar{x}$)	(x-$\bar{x}$)^2	(x-$\bar{x}$)^3	(x-$\bar{x}$)^4
$\sum x =$.....		$\sum$(x-$\bar{x}$)^2=......	$\sum$(x-$\bar{x}$)^3=.......	$\sum$(x-$\bar{x}$)^4=.....

Arithmetic Mean, $\bar{x} = \frac{\sum x}{n}$ =...

First central moment, $\mu_1 = \frac{\sum(x-\bar{x})}{n} = 0$

Second central moment, $\mu_2 = \frac{\sum(x-\bar{x})^2}{n}$ =..

Third central moment, $\mu_3 = \frac{\sum(x-\bar{x})^3}{n}$ =..

Fourth central moment, $\mu_4 = \frac{\sum(x-\bar{x})^4}{N}$ =...

Coefficient of Skewness $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

=...Coefficient of Kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2}$

=...

Practical No. 9

Objective: To calculate first four central moments, coefficient of Skewness ($\boldsymbol{\beta_1}$) and Kurtosis ($\boldsymbol{\beta_2}$) for grouped data. Comment on nature of the data.

Problem: Compute first four central moments and the Karl Pearson's coefficient of skewness (β_1) and Kurtosis (β_2) from the following data:

Height of tree (x)	32	33	34	35	36	37	38	39	40
No of trees (f)	3	5	8	11	19	23	40	28	14

Solution:

x	f	f. x	(x-$\overline{x}$)	f. (x-$\overline{x}$)	f. (x-$\overline{x}$)^2	f. (x-$\overline{x}$)^3	f. (x-$\overline{x}$)^4
Total	$N = \sum f$ =........	$\sum f.x = \cdots$		$\sum$f. (x-$\overline{x}$)=0	$\sum$f. (x-$\overline{x}$)^2 =.........	$\sum$f. (x-$\overline{x}$)^3 =........	$\sum$f. (x-$\overline{x}$)^4 =.......

Arithmetic Mean, $\overline{x} = \frac{\sum f.x}{N}$ =..

First central moment, $\mu_1 = \frac{\sum f.(x-\overline{x})}{N} = 0$

Second central moment, $\mu_2 = \frac{\sum f.(x-\overline{x})^2}{N}$ =..

Third central moment, $\mu_3 = \frac{\sum f.(x-\overline{x})^3}{N}$ =..

Fourth central moment, $\mu_4 = \frac{\sum f.(x-\overline{x})^4}{N}$ =..

Coefficient of Skewness $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ =...

Coefficient of Kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2}$ =..

Practical No. 10

Objective: To calculate correlation co-efficient between two variables.

Problem: Find the correlation coefficient between height and weight of yield of the plants. Data are given below:

Height in cm	6	7	8	9	10
weight in gm	20	23	24	26	26

Solution:

Step-I: To construct table:

x	**y**	**x^2**	**y^2**	**xy**
$\sum_{i=1}^{n} x_i$=	$\sum_{i=1}^{n} y_i$=	$\sum_{i=1}^{n} x_i^2 =$	$\sum_{i=1}^{n} y_i^2 =$	$\sum_{i=1}^{n} x_i y_i =$

Step-II: (a) n= number of paired observations =

(b) $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ =...

(c) $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ =...

Step-III: (a) $\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2}$ =..

(b) $\sigma_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \overline{y}^2}$ =..

(c) Cov(x, y) = $\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x} \times \overline{y}$ =...

Step-IV: Karl Pearson correlation coefficient, $r_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y}$=..

....

Problem: Find the rank correlation coefficient between height and biomass of the plants. Data are given below:

Rank in Height	1	2	3	4	5	6	7	8
Rank in biomass	1	2	3	5	6	4	7	8

Solution:

Step-I: Count number of paired observations, n =

Step-II: To construct table

R_x	R_y	$d_i = R_y - R_x$	d_i^2
			$\sum_{i=1}^{n} d_i^2$ =

Step-III: Rank correlation coefficient, $\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$ =...

....

Practical No. 11

Objective: To fit linear regression equation on the given data.

Problem: Fit the regression equation of y (yield in kg) on x (number of root fibers) of turmeric crop from the following data.

No. of roots	8	7	5	10	11	9	12	14	13
Yield (in kg)	1.2	1.1	0.7	1.3	1.3	1.0	1.4	1.3	1.4

Solution:

Step-I: To construct table:

x	y	x^2	y^2	xy
$\sum_{i=1}^{n} x_i$=	$\sum_{i=1}^{n} y_i$=	$\sum_{i=1}^{n} x_i^2 =$	$\sum_{i=1}^{n} y_i^2 =$	$\sum_{i=1}^{n} x_i y_i =$

Step-II: (a) n= number of paired observations =..

(b) $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ =...

(c) $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ =...

Step-III: (a) $\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2}$ =..

(b) $\sigma_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \overline{y}^2}$ =...

(c) Cov(x, y) = $\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x} \times \overline{y}$ =...

Step-IV: Karl Pearson correlation coefficient, $r_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y}$=..

Step-V: regression coefficient, $b_{yx} = \frac{\sigma_y}{\sigma_x} \times r_{xy}$=...

Step-VI: regression equation, $y - \overline{y} = b_{yx}(x - \overline{x})$

....

....

Practical No. 12

Objective: To test significant difference between sample mean and population mean using t-test

Problem: A sample of 10 trees which height are 10.5, 10.4, 10.8, 11.3, 12.5, 12.7, 11.5, 11.8, 12.1 and 11.5 meter. Test whether the sample comes from a forest which trees mean height is 11.

Solution: Step-I (a): Null Hypothesis H_0: Here we set up the null hypothesis that there is no significant difference between sample mean and population mean.

(b): Alternative hypothesis H_1: we set up the alternative hypothesis that there is significant difference between sample mean and population mean.

Step-II: to calculate test statistic t

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
$\sum_{i=1}^{n_1} x_i$=		$\sum_{i=1}^{n_1} (x_i - \overline{x})^2$=

Sample mean, $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ =..

S = $\sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$ =..

t = $\frac{\overline{x} - \mu}{S/\sqrt{n}}$ =..

Step-III: Conclusion: Since calculated value of |t| is .... tabulated value of t at α % level of significance then null hypothesis is..

Practical No. 13

Objective: To test significant difference between two sample means for independent samples and paired samples using t-test

Problem: A sample of 10 trees from forest A which height are 11.5, 11.4, 11.8, 12.3, 12.5, 13.3, 12.5, 12.8, 13.1 and 11.5 meter and second sample of 8 trees from forest B which height are 12.8, 12.3, 12.5, 13.8, 12.5, 12.8, 13.1 and 14.1 meter. Test whether forest A and B have same average height trees or not?

Solution: Step-I:(a) Null Hypothesis H_0: Here we set up the null hypothesis that there is no significant difference between two sample means.

(b) Alternative hypothesis H_1: we set up the alternative hypothesis that there is significant difference between two sample means.

Step-II: to calculate test statistic t

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
$\sum_{i=1}^{n_1} x_i$=		$\sum_{i=1}^{n_1}(x_i - \overline{x})^2$=

Sample mean, $\overline{x} = \frac{\sum_{i=1}^{n_1} x_i}{n_1}$ =..

y_i	$y_i - \overline{y}$	$(y_i - \overline{y})^2$

$\sum_{i=1}^{n_2} y_i$=		$\sum_{i=1}^{n_2}(y_i - \overline{y})^2$=

Sample mean, $\overline{y} = \frac{\sum_{i=1}^{n_2} y_i}{n_2}$ =...

S = $\sqrt{\frac{\sum_{i=1}^{n_1}(x_i - \overline{x})^2 + \sum_{i=1}^{n_2}(y_i - \overline{y})^2}{n_1 + n_2 - 2}}$ =...

t = $\frac{\overline{x} - \overline{y}}{S\sqrt{[\frac{1}{n_1} + \frac{1}{n_2}]}}$ =...

....

Step-III: Conclusion: Since calculated value of |t| is .... tabulated value of t at α % level of significance then null hypothesis is...

....

Problem: The following given yield data is related to before and after applying a soil treatment

Before treatment	7.9	8.5	7.3	9.7	10.3	10.2	11.1	8.9	8.5
After treatment	9.6	10.1	8.9	10.8	12.1	11.5	11.0	10.1	8.9

Test whether there is any significant effect of the soil treatment on the yield or not?

Solution: Step-I: (a) Null Hypothesis H_0: Here we set up the null hypothesis that there is no significant difference between two sample means.

(b) Alternative hypothesis H_1: we set up the alternative hypothesis that there is significant difference between two sample means.

Step-II: to calculate test statistic t

x_i	y_i	$d_i = y_i - x_i$	$d_i - \overline{d}$	$(d_i - \overline{d})^2$
		$\sum_{i=1}^{n} d_i$=		$\sum_{i=1}^{n}(d_i - \overline{d})^2$=

Sample mean, $\overline{d} = \frac{\sum_{i=1}^{n} d_i}{n}$ =...

S = $\sqrt{\frac{\sum_{i=1}^{n}(d_i - \overline{d})^2}{n-1}}$ =..

t = $\frac{\overline{d}}{S/\sqrt{n}}$ =...

Step-III: Conclusion: Since calculated value of |t| is .... tabulated value of t at α % level of significance then null hypothesis is..

....

Practical No. 14

Objective: To test goodness fit of the distribution and association between two attributes using chi Square test.

Problem: The no. of deaths due to covid-19 over the days of week is following:

Day's	Sun	Mon	Tue	Wed	Fri	Sat	Sun
No. of death	14	11	13	10	12	9	15

Test whether death due to covid-19 is distributed uniformly over the days of week or not?

Solution: Step-I: (a) Null Hypothesis H_0: Here we set up the null hypothesis that there is no significant difference between observed frequencies and expected frequencies.

(b) Alternative hypothesis H_1: we set up the alternative hypothesis that there is significant difference between observed frequencies and expected frequencies.

Step-II: Calculate test statistic, χ^2

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
				$\chi^2 = \sum_{i=1}^{n} \frac{(O_i-E_i)^2}{E_i}$= ....

Step-III: Conclusion: Since calculated value of $\chi^2$is ....than tabulated value of χ^2at 5 % level of significance then null hypothesis is...

....

Practical No. 15

Objective: To test goodness fit of the distribution and association between two attributes using chi Square test.

Problem: The following data is related to eye colour father and their son.

		Father's eye colour		
		$Blue$	Black	Total
Sons eye colour	$Blue$	70	30	100
	Black	20	8o	100
	Total	90	110	200

Test whether there is any association between father's eye colour and son's eye colour or not?

Solution:

Step-I:(a) Null Hypothesis H_0: Here we set up the null hypothesis that there is no association between two attributes.

(b) Alternative hypothesis H_1: we set up the alternative hypothesis that there is any association between two attributes.

Step-II: calculate expected frequencies:

		Attribute A		
		α	A	Total
Attribute B	β	$(\alpha\beta)$	$(A\beta)$	(β)
	B	(αB)	(AB)	(B)
	Total	(α)	(A)	N

Expected frequencies,

E$(\alpha\beta)$ = $\frac{(\beta)(\alpha)}{N}$ =...

E$(A\beta)$ = $\frac{(\beta)(A)}{N}$ =...

E(αB) = $\frac{(B)(\alpha)}{N}$ =..

E(AB) = $\frac{(B)(A)}{N}$ =...

Step-III: to calculate test statistic, χ^2

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
				$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$=

Step-IV: Conclusion: Since calculated value of $\chi^2$is ....than tabulated value of χ^2at 5 % level of significance then null hypothesis is...

Practical No. 16

Objective: To compare several sample means using one-way ANOVA

Problem: The data relate to the five varieties of fertilizers using CRD conducted in a field with four plots per variety.

Varieties	Seed yield of sesame (gm/plot)			
V1	6	7	7	6
V2	8	9	10	8
V3	5	6	6	5
V4	6	6	7	4
V5	5	7	6	5

Test whether the varieties are differs significantly or not?

Solution: Null Hypothesis: Here we set up the null hypothesis that the treatments are not differ significantly.

Alternative Hypothesis: At least one of the treatments differs significantly.

Varieties	Seed yield of sesame (gm/plot)				Total	Mean
V1	6	7	7	6	……	……
V2	8	9	10	8	……	……
V3	5	6	6	5	……	……
V4	6	6	7	4	……	…….
V5	5	7	6	5	…...	……
					G=…….	

1. No. of treatment = k =……………………………………………………………………….
2. No. of observation, N = $n_1+n_2+n_3+n_4+n_5$ =………………………………………………
3. Grand total, G=$\sum_1^k \sum_1^{n_i} x_{ij}$ =………………………………………………………………
4. Correction Factor, C.F.=$\frac{G^2}{N}$ =………………………………………………………………
5. Raw of Sum Squares, RSS =………………………………………………………………
   …………………………………………………………………………………………...
   …………………………………………………………………………………………..
   ………………………………………………………………………………………….
6. Total Sum of Squares, TSS= RSS – C. F. =………………………………………………
7. Sum of squares due to treatment, SST= $\frac{T_1^2}{n_1}+\frac{T_2^2}{n_2}+\frac{T_3^2}{n_3}+\frac{T_4^2}{n_4}+\frac{T_5^2}{n_5}$ =…………………………
   ………………………………………………………………………………………………………
8. Sum of Squares due to Error, SSE= TSS – SST =……………………………………..

Table: Analysis of Variance

Source of variation	Degree of freedom	Sum of Square	Mean sum of square	Variance Ratio	Tabulated F
Treatment	k-1=……	SST=……	MSST=………..	F =………..	$F_{.05}$ =…….
Error	N-k=…….	SSE=…….	MSSE=……….		
Total	N-1=…….				

Since calculated value of F is …… than tabulated of $F_{.05}$ then the null hypothesis is …………………………at 5% level of significance.

Since the null hypothesis is rejected so we calculate Critical difference (C.D.)

Standard error of difference between two treatments = $\sqrt{\left(\frac{1}{n_i}+\frac{1}{n_2}\right)*MSSE}$ =…………………

……………………………………………………………………………………………………

Critical Difference =CD= $(S.E.)_{diff} \times t_{.05}$ (error d.f.) =……………………………………………

……………………………………………………………………………………………………

Practical No. 17

Objective: To test equality of several treatment means using two-way ANOVA

Problem: Four varieties of Onion were compared as regard yields within four block in RBD. The data pertaining to yield in kg per plot are given below:

varieties	Block			
	I	II	III	IV
A B C D	4 6 4 6	5 8 6 7	3 4 5 6	4 6 5 7

Analyze the data and give conclusion?

Solution: Null hypothesis H_0: There is no significant difference between treatments as well as blocks as regard yield.

Alternative hypothesis H_1: At least two treatments as well as blocks are differ significantly as regard yield.

varieties	Block				Total	Mean
	I	II	III	IV		
A B C D	4 6 4 6	5 8 7 9	3 4 5 8	4 6 5 7	……… ……… ……… ……….	……. ……. ……. …….
Total	…..	…..	…..	…..	G=……	
Mean	…..	…..	……	…..		

1. Number of treatments = k =…………………………………………………………………
2. Number of replication = r =………………………………………………………………..
3. Total number of observation = r.k =…………………………………………………………..
4. Grand Total (G) = $\sum_{i=1}^{k}\sum_{j=1}^{r} y_{ij}$ =…………………………………………………………………
5. Correction factor = $\frac{G^2}{rk}$ =…………………………………………………………………………
6. Total sum of squares (TSS) = ($y_{11}^2 + y_{12}^2 + y_{13}^2 + …. + y_{kr}^2$) – C. F =……………………

…………………………………………………………………………………………………..

…………………………………………………………………………………………………..

7. Sum of Squares due to treatment (SST) = $\frac{T_1^2 + T_2^2 + \cdots + T_k^2}{r}$ – C.F =…………………………………

…………………………………………………………………………………………………..

8. Sum of Squares due to Block (SSB) = $\frac{B_1^2 + B_2^2 + \cdots + B_k^2}{k}$ – CF =……………………………………

………………………………………………………………………………………………….

9. Error sum of square (SSE) = TSS – SST – SSB =………………………………………….

ANOVA TABLE

Sources of variation	D.F	S.S	M.S.S	Variance Ratio	Tabulated F
Treatments	…………	……….	…………	…………	……….
Blocks	…………	……….	…………	…………..	………..
Error	…………	……….	…………		
Total	………..	………..			

Since calculated value of F is .... than tabulated of F at 5% level of significance so null hypothesis is ………………………………………………………………………………

……
……
……
……………………………………………

Now we calculate Critical difference (C.D.)

Standard Error of difference between two treatment means = $\sqrt{\frac{2\,MSSE}{r}}$ =…………………………

Standard Error of difference between two Block means = $\sqrt{\frac{2\,MSSE}{k}}$ =……………………………..

Critical Difference (C.D) = $(S.E.)_{diff} \times t_{.05}$ (error d.f) =……………………………………………

……

……

……

………..

Practical No. 18

Objective: To draw a simple random sample and estimate population mean and population variance.

Problem: Draw a sample of size n=3 using simple random sampling with –out replacement (SRSWOR) from a population unit 1, 2, 3, 4 and 5. Show the sample mean and sample mean square is an unbiased estimate of population mean and population mean square and to determine its variances and S.E.

Solution: Step-I: Construct the following table:

S. N.	Possible Samples	Sample mean $\overline{y_n}$	Sample Mean Square (s^2)	$(\overline{y_n} - \overline{y_N})$	$(\overline{y_n} - \overline{y_N})^2$
Total					

Step-II: calculate:

Number of all possible samples of size n from N = $\binom{N}{n}$

=…………………………………………………………………..

Sample Mean, $\overline{y_n} = \frac{\sum y_i}{n}$

=………

Sample Mean Square, $s^2 = \frac{\sum(y_i - \overline{y_n})^2}{n-1}$

=…………………………………………………………………………………………

Population Mean, $\overline{y_N} = \frac{\sum y_i}{N}$

=………..

Population Mean Square, $S^2 = \frac{\sum(y_i - \overline{y_N})^2}{N-1}$

=……………………………………………………………………………………….

E($\overline{y_n}$) = $\frac{\sum \overline{y_n}}{\binom{N}{n}}$

=……

E(s^2) = $\frac{\sum s_i^2}{\binom{N}{n}}$

=……

Var($\overline{y_n}$) = $\frac{N-n}{Nn} S^2$

=…….

Standard Error = $\sqrt{\text{Var}(\overline{y_n})}$ = $\frac{\sum(\overline{y_n} - \overline{y_N})^2}{\binom{N}{n}}$ =

………………………………………………………………………………………..

Step-III: Now we have to check whether

E($\overline{y_n}$) = $\overline{y_N}$

=……

E(s^2) = S^2

=……