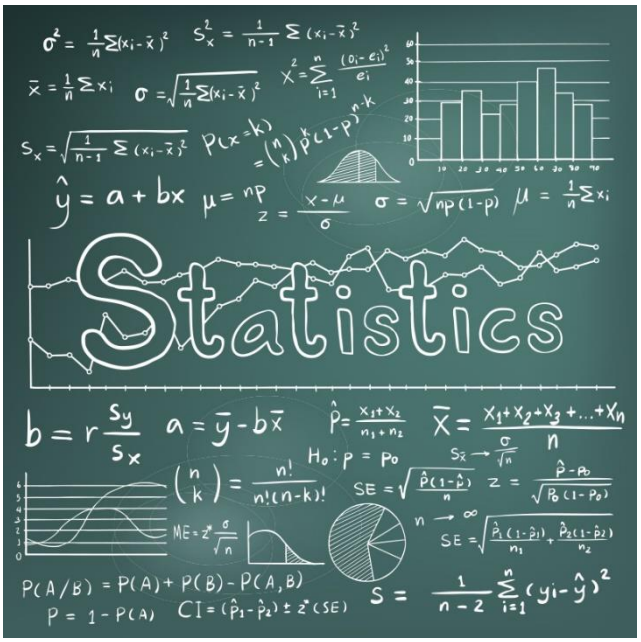
Elementary Statistics
BSH 111

Practical Manual



TAIBANGJAM LOIDANG CHANU

COLLEGE OF HORTICULTURE
CENTRAL AGRICULTURAL UNIVERSITY, IMPHAL
Bermiok, South Sikkim-737134

Pub. No.: COH/SK/BSH- 111/1

Citation: *Practical manual on Elementary Statistics*

Technical Support & Guidance

Prof. A. K. Pandey
Dean, COH, Bermiok
CAU, Imphal, Manipur

Prepared by:

Mrs. Taibangjam Loidang Chanu
Assistant Professor, CAEPHT, Ranipool

Designing and processing by:

PME Cell
College of Horticulture, Central Agricultural University, Bermiok, South Sikkim-737-134

Published by:

Dean, College of Horticulture, Central Agricultural University, Bermiok, South Sikkim-737-134

Prof. Anupam Mishra
Vice Chancellor

केंद्रीय कृषि विश्वविद्यालय
CENTRAL AGRICULTURAL UNIVERSITY
इंफाल-795004, मणिपुर
Imphal-795004, Manipur

Phone: 0385-2415933(O)
Email: vc@cau.ac.in
vcofficecau@yahoo.in

FOREWORD

The practical manual on BSH111, **Elementary Statistics** will be an excellent guide for the undergraduate students. Questions are classified as per the Fifth Dean's committee report and more than 15 questions are given to illustrate the concepts explained in the syllabus. The questions on the manual starts with a detailed step about the problems to solve. As each section has examples follow the necessary steps, the student will be highly benefited from the manual. Apart from this, it would help the undergraduate students to get familiarity with elementary statistic in solving the mathematical problem and solved the statistics analysis.

(Prof. Anupam Mishra)
Vice Chancellor
Central Agricultural University
Imphal, Manipur

Preface

Statistics has developed and grown into a full fledge inductive science dealing with the various aspects of chance phenomena. Growing at a comparatively fast pace, Statistics has now established itself as a discipline which has the widest range of applications among all science subjects. It has been common experience everywhere that theory and practical demonstration are held separately, development of theory under well define assumptions, derivation of formulas and analysis of techniques etc. for various statistical problems are discussed in theory classes, whereas practical demonstrations of the discussed theory are conducted during practical classes. The manual devotes and restricts mainly to the syllabus of our university.

The aim of this manual is specially for researchers and students of applied statistics at undergraduate levels. It is the lecture notes prepared during the past six years to teach the students of our university. This will provide a scientific approach and practical guide to the students when they conducted field experimentation. They can use statistical tools and techniques to interpret findings with confidence. In order to make the manual concise and handy I have included only a limited number of problems with their steps.

Taibangjam Loidang Chanu

SYLLABUS

Grouped data representation; Histogram, Frequency polygon; Bar and Pie Chart; Multiple components graphs and charts, Measures of Central Tendency - Mean Median and Mode; Measures of Dispersion-Standard deviation. Coefficient of Variation, Large sample test-z-test, Small sample test – t-test for single mean, Paired t-test, Chi square; Test of Goodness of Fit; Test of independence of attributes, Simple Correlation and fitting of linear regression analysis

CERTIFICATE

This is to certify that Mr./Ms ……..…….

Reg. No……. has performed Practical for the semester ……..……..

B. Sc. (Hons) Horticulture in the Course No…….…….……..

Title……..……

During the academic year…….

He/She has performed ……. practical out of ……..……

US ID:……..

Course Teacher

CONTENT

Sl. No.	Title of Exercise	Page no.
1.	Preparation of frequency distribution;	**1**
2.	Drawing of Histogram, Frequency polygon, Frequency curve, Less than ogive, greater than ogive	**2**
3.	To draw Bar diagram, pie chart, multiple components graphs and charts	**3**
4.	Calculation of mean, median, mode, standard deviation, coefficient of variation	**4-5**
5.	Test for small sample -t-test for single mean, Paired t-test, F-test	**6-8**
6.	Large sample test- Z test	**9**
7.	Test of Goodness of Fit; Test of independence of attributes	**10-11**
8.	Calculation of Simple Correlation between two variables	**12**
9.	Fitting of simple linear regression	**13**

Practical No. 1

Objective: Preparation of frequency distribution

Question: The following data relate to the grain yield of potato of 50 experimental plots in gm per plot

170, 160, 126, 181, 174, 164, 143, 165, 129, 166, 164, 154, 139, 128, 120, 168, 100, 143, 65, 162, 93, 80, 129, 166, 164, 154, 139, 128, 120, 168, 150, 118, 143, 120, 86, 170, 111, 169, 141, 164, 161, 170, 136, 195, 91, 187, 145, 189, 156, 190.

Solution: For the construction of frequency distribution table the following steps should be followed

Step I: The appropriate number of classes may be decided by Yule s formula 2.5 X $n^{\frac{1}{4}}$, where n is the total no. of observations.

Step II: Calculate range=Maximum value in the data set- minimum value in the data set

Step III: Calculate the class interval as

$$C=\frac{\text{Maximum value in the data set} - \text{minimum value in the data set}}{No.of\ classes}$$

IV: Construct the frequency distribution table

Class interval	Telly mark	Frequency
……..	……..	……
……		
……		

Practical No. 2

Objective: To drawing of Histogram, Frequency polygon, Frequency curve, Less than ogive, greater than ogive

Question: Draw histogram, frequency polygon, freq. curve, less than ogive, greater than ogive of the freq. distribution table:

Class	20-35	35-50	50-65	65-80	80-95	95-110	110-125	125-140	140-155	155-170	170-185	185-200	200-215	215-230	230-245
Freq. (f_i)	1	4	6	9	12	18	31	36	33	20	12	9	5	3	1

Solution: Construct the following frequency table;

Class interval	Frequency	Mid-point	Less than cumulative frequency	More than cumulative frequency

Practical No.3

Objective: To draw Bar diagram, pie chart, multiple components graphs and charts

Bar diagram:

Question: The production of major foodgrain crops in Sikkim in the year 2017-2018 was as follows:

Crops	Production in' 000 Tonne
Rice	17.6
Wheat	0.2
Oilseed	5.8
Pulses	5.1
Coarse cereal	71
Food grains	93.9

Draw the simple bar diagram and Pie Diagram for this data.

……
……
……
……
……
……

Pie Diagram:

Question: The production of major foodgrain crops in Sikkim in the year 2017-2018 was as follows:

Crops	Production in' 000 Tonne
Rice	17.6
Wheat	0.2
Oilseed	5.8
Pulses	5.1
Coarse cereal	71
Food grains	93.9
Total	193.6

Draw the Pie Diagram for this data.

……
……
……
……
…….

Multiple bar diagram:

Question: Draw a multiple bar diagram for the following data which represented agricultural production in Sikkim

Year	Rice(in '000 tones)	Wheat (in '000 tones)	Oilseed (in '000 tones)
2010	24.3	5.9	9.4
2011	21	2.7	7.9
2012	20.9	0.6	7.8
2013	21.3	0.4	7.1
2014	20.3	0.2	7.1

…….
……
……

Practical No. 4

Objective: To find out mean, median, mode, variance and the coefficient of variation

Question: The following observation are onion production of a village. Find out the mean, the median, the mode, variance and the coefficient of variation from the following data-

Class	20-35	35-50	50-65	65-80	80-95	95-110	110-125	125-140	140-155	155-170	170-185	185-200	200-215	215-230	230-245
Freq. (f_i)	1	4	6	9	12	18	31	36	33	20	12	9	5	3	1

Solution: Step I:

Construct the following frequency table:

Solution:

Class	Frequency(fi)	Mid-point (xi)	Cumulative freq	fixi

Step II: For a continuous frequency distribution the mean is given by-

$$\bar{X}=\frac{\sum fixi}{\sum fi}$$

Xi= mid-value

Fi= frequency of the ith class

For a continuous frequency distribution, the median is given by-

$$M_d = l + \frac{\frac{N}{2}-C}{F} \text{ X h}$$

l = lower limit of the median class

c = cumulative frequency just above the median class

f = frequency of the median class

h = height of the median class

For a continuous frequency distribution, the mode is given by-

$$\boldsymbol{M_o = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} X\ h}$$

l = lower limit of the model class

f_m = frequency of the model class

f_1 = frequency of the preceding model class

f_2 = frequency of the succeeding model class

Variance: $\sigma^2 = \frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}$

Coefficient of variation: The formula of coefficient of variation (C.V) is given by

C.V=$\frac{Standard\ deviation}{Mean}$X 100

Practical No. 5

Objective: To test significant difference between means in case of single sample, two sample (independent) and two samples (Paired) using t-test.

t-test for single mean-n

Question: From a random sample of pigs fed on a diet A the increased in weights in lb in a certain period are 10.2,6.7,16.4,17.3,13.5,12.5,8.2,14.9,15.5 &9.3

Test that the population from which the sample is drawn has a mean 14.

Solution: Step I: Null hypothesis, H_0: $\mu = 14$

Step II: Test Statistics, t= $\frac{\bar{x}-\mu}{s/\sqrt{n}}$

x_i	x_i^2

Where, $s^2 = \frac{1}{n-1}[\sum x_i^2 - n\bar{x}^2]$

Step III: Write the Conclusion

t-test for difference of mean-

Question: An anthropological survey of Jaintia people of the following data of heights of 10 males and 10 females. Test whether the Jaintia males and females are equally tall or not.

Heights of males (in cm)	:	155	160	152	158	165	150	156	161	157	154
Heights of females (in cm)	:	150	148	152	149	154	156	149	150	146	148

Solution :

Step I: Null hypothesis, H_0: Jaintia males and females are equally tall

Step II: Test Statistics ,t=$\frac{|\bar{x}-\bar{y}|}{\sqrt{s^2(\frac{1}{n1}+\frac{1}{n2})}}$

x_i	y_i	x_i^2	y_i^2

Where, $s^2 = \frac{1}{n_1+n_2-2}[\sum x_i^2 - n_1\bar{x}^2 + \sum y_i^2 - n_2\bar{y}^2]$

Step III: Write the Conclusion

Paired t-test-

Question: A certain stimulus is to be tested for its effect on blood pressure. 12 men have their blood pressure measured before and after the stimulus.

Before	:	120	124	130	118	140	128	140	135	126	130	126	127
After	:	128	131	131	127	132	127	141	137	118	132	129	135

Test whether the stimulus has any effect on blood pressure.

Solution: Null hypothesis, H_0: The stimulus has no effect on blood pressure

Step II: Test Statistics, t=$\frac{|\bar{d}|}{s_d/\sqrt{n}}$

x_i	y_i	$d_i = x_i - y_i$	d_i^2

Where, x_i =blood pressure of the ith men before the stimulus is administered
y_i = blood pressure of the ith men after the stimulus is administered
n = sample size
$d_i = x_i - y_i$
$\bar{d} = \frac{1}{n}\sum d_i$
$s_d^2 = \frac{1}{n-1}\left[\sum d_i^2 - n\bar{d}^2\right]$

Step III: Write the Conclusion

Practical No 6.

Objective: To test significant difference between means in case of single sample, two samples and test of difference between two standard deviations using Z-test.

One sample Z test

Question: A sample of 900 members has a mean 3.4cm and standard deviation 2.61 cm. Is the sample from a large population of mean 3.25 cm and standard deviation 2.61 cm?

Solution: Step I: Null hypothesis, $H_0: \mu = 3.25cm$

Step II: Test Statistics, z= $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

Step III: Write the Conclusion

Test of significance for difference of means

Question: The mean yield of wheat per hectare for 150 villages selected in a random sample from district A is 32.24 qtls. With standard deviation 21.31 qtls . A second sample of 100 villages from a district B shows the mean yield 24.58 with standard deviation19.78 qtls. Examine the significance of difference in mean yield of two districts.

Solution: Step I: Null hypothesis, $H_0: \mu_1 = \mu_2$

Step II: Test Statistics, Z= $\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma^2}_2}{n_2}}}$

Step III: Write the Conclusion

Test for difference between two standard deviations

Question: The mean nutrient contents in a sample of 100 bags of ammonium chloride is 5.4 kg per bag with standard deviation 13.7 kg. In another sample of 160 bags, the mean nutrients contents are found to be 5.4kg per bag with standard deviation 21.3 kg. Are the two results consistent?

Solution: Step I: Null hypothesis, $H_0: \sigma_1 = \sigma_2$

Step II: Test Statistics, Z= $\frac{\hat{\sigma}_1-\hat{\sigma}_2}{\sqrt{\frac{{\hat{\sigma}_1}^2}{2n_1}+\frac{{\hat{\sigma}^2}_2}{2n_2}}}$

Step III: Write the Conclusion

Practical No. 7

Objective: To test the Chi-square test of Goodness of Fit and independence of attributes

Chi-square test of goodness of fit

Question: The following are (a) the observed frequencies & (b) expected frequencies of normal variate using the mean and standard deviation of (a)

O_i :	3	21	150	335	326	135	26	4
E_i :	3	31	148	322	319	144	30	3

Test the goodness of fit by applying suitable criterion.

Solution: Step I: Null hypothesis, H_0: Normal fit is good

Step II: Test Statistics, $\chi^2 = \sum[\frac{(O_i - E_i)^2}{E_i}]$

O_i	E_i	$(O_i - E_i)^2 / E_i$

Step III: Write the Conclusion

Chi- square test of independence of attributes

Question: **Two sample polls of votes for two candidates A and B for a public office are taken, one from among the residents of rural areas. The results are given in the adjoining table. Examine whether the nature of the area is related to voting preferences in this election.**

Area	Votes for		Total
	A	B	
Rural	620	380	1000
Urban	550	450	1000
	1170	830	2000

Solution: Step I: Null hypothesis, H_0: the nature of the area is independent of the voting preference in the election

Step II: Test Statistics, $\chi^2 = \sum[\frac{(O_i - E_i)^2}{E_i}]$

Under H_0 ,we get expected frequencies as follows:

$$E_{ij} = \frac{i^{th} row\ total\ X\ j^{th}\ column\ total}{Grand\ total}$$

Step III: Write the Conclusion

Practical No 8.

Objective: To calculate correlation co-efficient between two variables

Question: Calculate the correlation coefficient for the following heights (in inches) of fathers (x) and their sons (y) :

x	:	65	66	67	67	68	69	70	72
y	:	67	68	65	68	72	72	69	71

Solution:

Step I: Construct the table-

X	**Y**	$\mathbf{X^2}$	$\mathbf{Y^2}$	**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: Number of paired observations, n =

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

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2}$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2}$$

Cov(x,y) = $\frac{1}{n}\sum_{i=1}^{n} x_i\, y_i - \bar{x}\, X\bar{y}$

Step III: Calculate the Karl Pearson correlation coefficient, $r_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y}$ =

Practical No. 9

Objective: Fitting of simple linear regression on the given data

Question: The following table gives the length of green jute plant (in cm) and weight of dry jute fibre (in gm) for 10 jute plants.

Length of green plant	Weight of dry jute fibre	Length of green plant	Weight of dry jute fibre
111	1.20	118	2.10
125	2.15	140	3.05
135	2.70	150	4.10
165	5.25	160	5.70
171	6.05	185	7.25

Calculate the linear regression of dry fibre on length of green plant. Also find the expected dry weight of a120cm long green plant.

Solution: Construct the table-

X	Y	X^2	XY
$\sum_{i=1}^{n} x_i =$	$\sum_{i=1}^{n} y_i =$	$\sum_{i=1}^{n} x_i^2 =$	$\sum_{i=1}^{n} x_i y_i =$

Step II: Number of observations, n =

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

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$b_{yx} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$

a= $\bar{y} - b_{xy}\bar{x}$

Step III: Find the regression equation, y = a+ $b_{yx}X$

Step IV: When x= 120, y =