WORK BOOK

MT-330 Applied Probability & Statistics

Department of Mathematics

NED University of Engineering & Technology

WORK BOOK

MT-330 Applied Probability & Statistics

Name : ____________________

Roll No : ____________________

Enrollment No : ____________________

Section : ____________________

Group : ____________________

Batch : ____________________

Certificate

Certified that Mr. / Ms. ____

Bearing Roll No.________________ of Class ___________ Year, Section ______________

Batch _______________ Department of ____

has completed the laboratory work for the course Applied Probability & Statistics (MT-330) as prescribed by the NED University of Engineering & Technology for the Academic Session __________________

Date: _____________

INDEX

Name: ____ **Roll No:** ____________________

S.NO.	DATE	PRACTICAL	GRADE / REMARKS			SIGNATURE
			A	B	C	
01		Introduction to Minitab				
02		Ungrouped Data, Graphs, Measures of Central Tendency & Dispersion				
03		Grouped Data, Graphs, Measures of Central Tendency & Dispersion				
04		Introduction to Global-Macro in Minitab				
05		Skewness & Kurtosis				
06		Binomial Probability Distribution				
07		Poisson Probability Distribution				
08		Normal Probability Distribution				
09		Linear Regression and Correlation				
10		Multiple and Non Linear Regression				
11		One Mean Hypothesis				
12		Two Mean Hypothesis				

PREFACE

PURPOSE:

The basic purpose of this practical workbook is to provide the necessary computational data analysis approach in statistics using Minitab for the students of engineering. Since practical covers wide variety of different and interesting statistical applications therefore it is hope that this workbook would be of great value for the engineering graduates.

PRE-REQUISITE:

Student should know the fundamental principles of statistics and computers.

OBJECTIVE:

To provide reader with a working knowledge of statistics, This workbook is designed for students having numerous applications in engineering. The material in this workbook consists of sketch of functions, plotting of variables, basic statistics, regression analysis, correlation analysis, probability distributions & hypothesis testing. This workbook is designed to be a clear, readable and even enjoyable introduction to the statistical concept that have become an important part of every engineering problems. Moreover statistical measures such as psychological and educational testing and increasingly important application regarding probability and statistics are also discussed and streamlined with engineering disciplines.

INTRODUCTION TO STATISTICAL SOFTWARE

MINITAB

WHAT MINITAB WILL DO FOR YOU

Before the widespread availability of powerful computers and prepackaged statistical software, tedious manual computations were routine in statistics courses. Today, computers have revolutionized data analysis, which is a fundamental task of statistics. Packages such as Minitab allow the computer to automate calculations and graphs. Minitab can perform a wide variety of tasks, from the construction of graphical and numerical summaries for a set of data to the more complicated statistical procedures and tests. Minitab will free you from mathematical calculations and allow you to concentrate more on the analysis of data and the interpretation of the results.

MINITAB WINDOWS

Minitab has six types of windows which can all be open at the same time: *Session, Data, Help, Info, History* and *Graph*. A window can be made active by selecting it. For example, to make the *Session* window active, place the mouse pointer on the window and just click it, if it is visible on your screen. The *Session* window and *Data* window are used most frequently.

Session Window

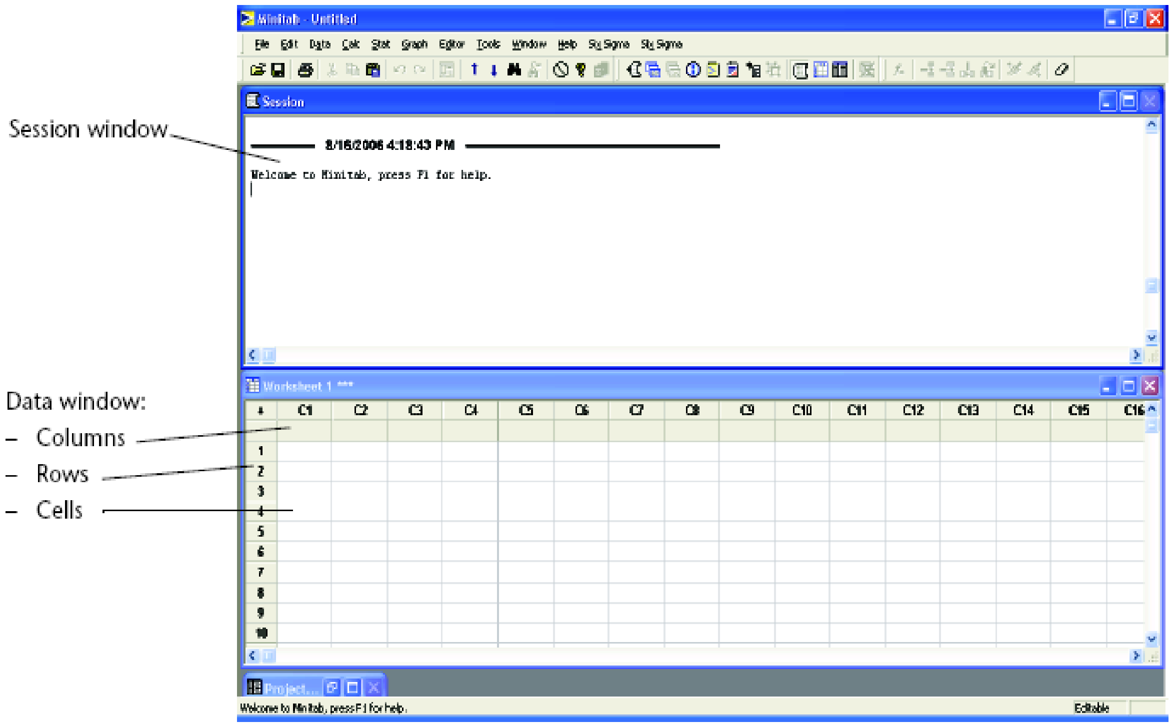
The *Session* window is used to enter Minitab commands and to display output. It has a menu bar on top and the usual Minitab prompt MTB> on the left. To enter a command. If MTB> prompt is not displayed in the session window, choose Editor ► Enable Command Language.

The *Session* window scrolls as more output goes into it. You can scroll up and down to see various parts of your output.

Data Window

The *Data* window, sometimes called the *worksheet*, displays active data entered by you or produced by the computer. You can enter, edit and view the data. The current cell is highlighted;

others can be selected by clicking the mouse on them or using the arrow keys. The scroll bars are used to view different parts of the worksheet.



MINITAB COMMANDS

Commands tell Minitab what to do. You can issue commands in Minitab by choosing commands from the menus or by typing session commands directly into the Session window. Session commands are used throughout this workbook.

MINITAB SESSION COMMANDS

Session Commands and the Session Window

Session commands allow you to provide specific instructions, through a command language. Most session commands are simple, easy to remember words like SET, READ, PLOT, SAVE, or SORT etc. The Session window is primarily used for displaying the results of commands, as text.

However, you can also type session commands in the Session window by turning on the MTB> command prompt.

Rules for Entering Session Commands

- A session command consists of one main command and may have one or more subcommands. Arguments and symbols may also be included in the command. Subcommands, which further define how the main command should be carried out, are usually optional. Arguments specify data characteristics.
- To execute a command, type the main command followed by any arguments. If the command has subcommands, end the command line with a semicolon. Type subcommands at the SUBC> prompt. Put a semicolon (;) after each subcommand. Put a period (.) after the last subcommand. Press <Enter> to execute a command.

- Commands and column names are not case-sensitive; you can type them in lowercase, uppercase, or any combination. You can abbreviate any session command or subcommand by using the first four letters.

- Arguments specify data characteristics, such as location or titles. They can be variables (columns or constants) as well as text strings or numbers. Enclose variable names in single quotation marks (for example, HISTOGRAM ‘Salary’). In arguments, variable names and variable numbers can be used interchangeably. For example, DESCRIBE C1 C2 and DESCRIBE ‘Sales’ C2 do the same thing if C1 is named ‘Sales.’

- You can abbreviate a consecutive range of columns, stored constants, or matrices with a dash. For example, PRINT C2-C5 is equivalent to PRINT C2 C3 C4 C5. You can use a stored constant (such as K20) in place of any constant. You can even use stored constants to form a range such as K20:15, which represents all integers from the value of K20 to 15.

Command Prompts

The prompts that appear in the Session window help you know what kind of input Minitab expects. There are four different prompts:

MTB>	Command prompt; type the session commands here and press Enter.
SUBC>	Subcommand prompt; type the subcommands here or type ABORT to cancel the entire command.
DATA>	Data prompt; enter data here. To finish entering data and return to the MTB> prompt, type END and press Enter.
CONT>	Continuation prompt; if the command from your previous line ends with the continuation symbol &, Minitab displays CONT> on the next line so you can enter the rest of the command or data.

S.NO.	PRACTICAL
01	**Introduction to Minitab** Introduction to Minitab environment, entering, editing and manipulating data, arithmetic operations.
02	**Ungrouped Data** **Graphical Representation, Measures of Central Tendency & Dispersion** Graphical representation, simple bar diagram, pie chart, stem-leaf display, box-plot, A.M, G.M, H.M, median, mode, range, variance and standard deviation.
03	**Grouped Data** **Graphical Representation, Measures of Central Tendency & Dispersion** Construction of frequency distribution, graphical representation, histogram, frequency curve, cumulative frequency curve, A.M, G.M, H.M, median, Mode, range, variance and standard deviation.
04	**Introduction to Global-Macro** Calculation of various statistical measures, first four moments about origin, first four moments about mean, measure of skewness and kurtosis using global Macro.
05	**Skewness & Kurtosis** Calculation of Moments about origin and mean, skewness and kurtosis.
06	**Binomial Probability Distributions** Binomial distribution with applications
07	**Poisson Probability Distribution** Poisson distribution with applications

08	**Normal Probability Distribution** Normal distribution, central limit theorem and applications
09	**Linear Regression Line and Correlation Analysis** Scatter diagram, correlation coefficient, linear regression line and prediction
10	**Multiple Regression Analysis and Non-Linear Models** Multiple regression model, quadratic and exponential curves
11	**One Mean Hypothesis** Testing a Hypothesis about one Mean. Z test and t-test
12	**Two Mean Hypothesis** Testing a Hypothesis about two Mean. Z test and t-test

PRACTICAL NO. 01

Introduction to Minitab

Note: Show all necessary session commands, results and interpretations (where required)

1. The following table gives some facts on 10 major countries.

Country	**Population**	**Area**
Pakistan	133500000	307374
China	1217600000	3691500
United States	265200000	3615278
Taiwan	21400000	13900
Netherlands	15500000	16133
Egypt	63700000	386661
Russia	147700000	6592850
Sweden	8800000	170250
Bangladesh	119800000	55598
Indonesia	949600000	1222244

i. Enter, Save and Print data.

ii. Sort data according to alphabetical order and print.

iii. Insert following information in dictionary order and print.

Country	**Population**	**Area**
Poland	38600000	120728

2. Given $x_1 = 1$ $x_2 = 2$ $x_3 = 3$ $x_4 = 4$ $x_5 = 5$

$y_1 = 6$ $y_2 = 7$ $y_3 = 8$ $y_4 = 9$ $y_5 = 10$

Evaluate the following expressions using Minitab commands and store in constants K1, K2…K7 and print.

i. $\sum_{i=1}^{5} x_i \sum_{i=1}^{5} y_i$

ii. $\sum_{i=1}^{5} x_i\, y_i$

iii. $\sum_{i=1}^{5} \log_{10}(x_i)$

iv. $\sqrt{\sum_{i=1}^{5}(5x_i + y_i)}$

v. $\sum_{i=1}^{5} |2x_i - 5y_i|$

vi. $\sum_{i=1}^{5}(x_i + y_i)^3$

vii. $\exp\left(\sum_{i=1}^{5} x_i\right)$

PRACTICAL NO. 02

Ungrouped Data, Graphical Representations, Measures of Central Tendency & Dispersion

1. The following table gives some facts on 10 major countries.

Country	**Population**	**Area**
Bangladesh	119800000	55598
China	1217600000	3691500
Egypt	63700000	386661
Indonesia	949600000	1222244
Netherlands	15500000	16133
Pakistan	133500000	307374
Russia	147700000	6592850
Sweden	8800000	170250
Taiwan	21400000	13900
United States	265200000	3615278

i. Construct Pie chart for country and area.
ii. Construct Simple Bar diagram for country and population.

2. Suppose a reference librarian selected a random sample of 36 customers before starting a new information retrieval system. The retrieval times in minutes are given below. Represent data by Stem-Leaf display.

10	15	15	10	20	15	10	20	30	15	25	05
45	22	10	27	25	22	10	30	15	05	08	15
05	36	23	17	15	10	35	48	50	25	03	31

i. Represent data by Stem-Leaf display.
ii. Calculate Arithmetic mean, Geometric mean, Harmonic mean, Median, Range, Variance and Standard deviation.

PRACTICAL NO. 03

Grouped Data, Graphical Representation, Measures of Central Tendency & Dispersion

1. Following data shows the length of time (in minutes) that customers had to wait before receiving the information they requested. The company selected 44 customers at random.

09	16	19	21	11	16	19	22	12	17	19
22	13	17	20	22	13	17	20	24	14	17
20	24	14	18	21	25	15	18	21	27	15
18	21	29	16	19	21	38	12	17	30	32

i. Construct stem and leaf plot and then make Frequency distribution by the help of stem and leaf and then construct Relative and Cumulative Frequency distribution

ii. Using Frequency distribution calculate Arithmetic mean, Geometric mean, Harmonic mean, Variance, Standard deviation and Mean absolute deviation.

PRACTICAL NO. 04

Introduction to Global-Macro

1. Write Global macro for a weekly sales (in $) given in the table of four different brands to calculate Arithmetic mean, Geometric mean, Harmonic mean, Range, Variance, Standard deviation, Mean absolute deviation and Coefficient of variation for each brand and determine which brand gives more consistent sales.

Brand A	Brand B	Brand C	Brand D
114	116	116	110
110	118	115	117
111	114	108	121
113	115	114	123
20	113	100	150

PRACTICAL NO. 05

Skewness and Kurtosis

1. From the following frequency distribution calculate first four moments about origin, first four moments about mean , $\sqrt{\beta_1}$ $(measure of\ skewness)$ and , $\beta_2\,(measure of\ kurtosis)$ also comments on the shape of distribution.

Class Interval	07 — 09	10 — 12	13 — 15	16 — 18	19 — 21	**Total**
Frequency	02	08	14	09	01	**34**

Hint: Formulae for above problem given in formula list at the end of workbook

PRACTICAL NO. 06

Binomial Probability Distribution

1. Obtain Binomial probability distributions for $p = 0.1$, $p = 0.5$ and $p = 0.9$ with $n = 10$. Also graph these probability distributions on same graph.

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}; \quad x = 0,1,2,\ldots,n$$

2. Suppose that 60% of all consumers favor Internet shopping. Five consumers are randomly sampled and the number who favors internet shopping is recorded. Let 'X' be the no. of consumers who favor this concept.

   i. Obtain and describe probability distribution of X.
   ii. Using properties of Expectation find mean and standard deviation of X and compare with theoretical mean and standard deviation.
   iii. What is the probability that 3 or more favor Internet shopping?

PRACTICAL NO. 07

Poisson Probability Distribution

1. Obtain and plot the Poisson probability and cumulative probability distributions for $\mu = 2.5$

$$P(X = x) = \frac{e^{-\mu}\,\mu^{x}}{x!} \quad ; \quad x = 0,1,2\ldots,\infty$$

2. A QC inspector at an automobile assembly plant has found that the number of paint defects on a car has a Poisson distribution with a mean of five defects per car.

   i. Calculate mean and standard deviation of 'X', the no. of paint defects per car.
   ii. Graph the probability distribution.
   iii. What is the probability that there are no paint defects? Five or more defects? Less than 8 defects?

PRACTICAL NO. 08

Normal Probability Distribution

1. Plot the following Normal distributions

   i. Normal random variable with $\mu = 100$ *and* $\sigma = 10$. Use values of x form approximately three standard deviations below the mean and three standard deviations above the mean.

   ii. Normal distribution with $\mu = 0$ *and* $\sigma = 1.0, 1.5, 2.0$

   iii. Standard Normal distribution with $\mu = 0$ *and* $\sigma = 1$

2. The average amount of time it takes a student to complete a certain task is Approximately normally distributed with $\mu = 50 \min$ *and* $\sigma = 5 \min$.

   i. Obtain a graph of Normal distribution.

   ii. What proportion of all students take between 40 and 55 minutes?

   iii. What proportion of all students take longer than 1 hour?

   iv. 75% of all students take less than what time to complete the task?

PRACTICAL NO. 09

Linear Regression Line and Correlation Analysis

1. The selling prices y and living areas in square feet x of 20 homes are given in the following table. Write a macro to display the results of following tasks.

Home	**Price ($)**	**Area**	**Home**	**Price ($)**	**Area**
1	86000	870	11	118900	2052
2	86600	840	12	125000	1590
3	92000	1032	13	130600	1600
4	92500	1168	14	139875	2044
5	93500	1100	15	144400	1916
6	94000	1430	16	148000	2024
7	104000	1520	17	156000	1840
8	104500	1468	18	151900	1684
9	109900	1160	19	159500	1760
10	111900	1800	20	163000	2260

(a) What is the average square footage and selling price of the homes in the sample?

(b) Construct Scatter plot.

(c) Calculate and interpret correlation coefficient (r)?

(d) Determine least square regression line. $\hat{y} = a + bx$ By performing arithmetic operations.

(e) Draw regression line. $\hat{y} = a + bx$ And scatter plot on the same graph.

(f) Calculate SSE, SS Regression and SS Total?

(g) Calculate Coefficient of determination and comments?

PRACTICAL NO. 10

Multiple Regression Analysis and Non-Linear Models

1. The quality y of particular finished product is dependent on temperature x_1 and pounds per square inch of pressure x_2. The results of experiment are as follows.

Temp	Pressure	Quality	Temp	Pressure	Quality	Temp	Pressure	Quality
80	50	50.8	90	50	63.4	100	50	46.6
80	50	50.7	90	50	61.6	100	50	49.1
80	50	49.4	90	50	63.4	100	50	46.4
80	55	93.7	90	55	93.8	100	55	69.8
80	55	90.9	90	55	92.1	100	55	72.5
80	55	90.9	90	55	97.4	100	55	73.2
80	60	74.5	90	60	70.9	100	60	38.7
80	60	73.0	90	60	68.8	100	60	42.5
80	60	71.2	90	60	71.3	100	60	41.4

(a) Fit multiple regression model $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$

(b) Write a global macro using matrices technique to estimate model.

2. The following table gives the size x of the home in square feet and the number of Kilowatt hours y of electrical usage for each of ten homes during a particular month.

Size	1290	1350	1470	1600	1710	1840	1980	2230	2400	2930
Usage	1182	1172	1264	1493	1571	1711	1804	1840	1956	1954

Write a global macro for the following tasks.

(a) Construct Scatter plot between size of home and electrical usage.

(b) Fit Quadratic model of the form $\hat{y} = b_0 + b_1 x + b_2 x^2$

(c) Predict the monthly electrical usage of a 1600 square foot home.

PRACTICAL NO. 11

One Mean Hypothesis

1. A random sample of 64 bags of white Cheddar popcorn weighed, on average, 5.23 ounces with a standard deviation of 0.24 ounces. Test the hypothesis that $\mu = 5.5$ ounces against the alternative hypothesis, $\mu < 5.5$ ounces at the 0.05 level of significance.

2. According to a dietary study, a high sodium intake may be related to ulcers, stomach cancer, and migraine headaches. The human requirement for salt is only 220 milligrams per day, which is surpassed in most single servings of ready-to-eat cereals. If a random sample of 20 similar servings of certain cereal has a mean sodium content of 244 milligrams and a standard deviation of 24.5 milligrams, does this suggest at the 0.05 level of significance that the average sodium content for a single serving of such cereal is greater than 220 milligrams? Assume the distribution of sodium contents to be normal.

3. A machine produces metal rods used in an automobile suspension system. A random sample of 15 rods is selected, and the diameter is measured. The resulting data (in millimeters) are as follows:

8.24	8.25	8.20	8.23	8.24
8.21	8.26	8.26	8.20	8.25
8.23	8.23	8.19	8.28	8.24

   Test the hypothesis that population mean of diameter is at least 8.25? Use 5% level of significance.

4. A machine is producing metal pieces that are cylindrical in shape. A sample of pieces is taken & the diameter is 1.07, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01 & 1.03cms. Test the hypothesis at 1% level of significance that populations mean of the diameter of pieces from this machine is greater than 1 cm. assuming an approximately normally distributed?

PRACTICAL NO. 12

Two Mean Hypothesis

1. In a study conducted by the Department of Mechanical Engineering and analyzed by the Statistics Consulting Center at Virginia Tech, steel rods supplied by two different companies were compared. Ten sample springs were made out of the steel rods supplied by each company, and the “bounciness” was studied. The data are as follows:

   Company A:
   9.3 8.8 6.8 8.7 8.5 6.7 8.0 6.5 9.2 7.0

   Company B:
   11.0 9.8 9.9 10.2 10.1 9.7 11.0 11.1 10.2 9.6

   Can you conclude that there is virtually no difference in means between the steel rods supplied by the two companies? Use a *P*-value to reach your conclusion.

2. Two chemical processes for manufacturing the same product are being compared under the same conditions. Yield from Process A gives an average value of 96.2 from 35 runs, and the estimated standard deviation of yield is 2.75. Yield from Process B gives an average value of 93.3 from 40 runs, and the estimated standard deviation is 3.35. Yields follow a normal distribution. Is the difference between the mean yields statistically significant? Use 5% level of significance, and show rejection regions for the difference of mean yields on a sketch.

3. According to *Chemical Engineering* an important property of fiber is its water absorbency. The average percent absorbency of 25 randomly selected pieces of cotton fiber was found to be 20 with a standard deviation of 1.5. A random sample of 25 pieces of acetate yielded an average percent of 12 with a standard deviation of 1.25. Is there strong evidence that the population mean percent absorbency for cotton fiber is significantly higher than the mean for acetate? Assume that the percent absorbency is approximately normally distributed and that the population variances in percent absorbency for the two fibers are the same. Use a significance level of 0.05.

Symbol Table

x_i	Random variable or mid points of grouped data
n	No. of observations in a sample
N	No. of observations in a population
Σ	Capital sigma; summation
Σx	Sum of the random variable
Σx^2	Sum of the square of observation
$(\Sigma x)^2$	Square of the sum of the observations
Σxy	Sum of the products of each *x* variable multiplied by corresponding *y* variable
C.I	Class interval
C.B	Class boundary
f	Frequency
A.M	Arithmetic mean
G.M	Geometric mean
H.M	Harmonic mean
C.f	Cumulative frequency
$Q_1, Q2, Q3$	Quartiles
$D_1, D_2, \ldots, D_9$	Deciles
$P_1, P_2, \ldots, P_{99}$	Percentiles
S.D	Standard deviation
$\bar{x}$	Sample mean
s^2	Sample variance
μ	Population mean
σ^2	Population variance
$\hat{\sigma}^2$	Estimated variance
c.o.v	Coefficient of variation
β_1	Coefficient of Skewness
β_2	Coefficient of Kurtosis

$P(A)$	Probability of an event A
$P(A/B)$	Probability of event A, assuming event B has occurred
$n!$	n factorial
$\begin{pmatrix} n \\ c \\ x \end{pmatrix}$	No. of the combinations of n items selected x at a time
$\begin{pmatrix} n \\ p \\ x \end{pmatrix}$	No. of the permutations of n items selected x at a time
p	Probability of an event or the population proportion
q	Probability or proportion equal to $1-p$
$E(x)$	Expected value of x or mean of x
$E(x^2)$	Expected value of x^2
$V(x)$	Variance of x
k	No. of samples or population or categories
$f(x)$	Function of x
$\hat{y}$	Expected value from estimated regression line
a	Estimate of y-intercept of the regression line
b	Estimate of the slope of the regression line
r	Simple correlation coefficient
r^2	Coefficient of determination
z	Standard normal variable/score
$z_{\alpha/2}$	Critical value of z
χ^2	Chi square distribution
$\chi^2_{\alpha/2}$	Critical value of chi square
$d.f$	Degree of freedom
α	Level of significance
H_0	Null hypothesis
H_1	Alternative Hypothesis

Formulae List

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

$$A.M = \bar{x} = \frac{\sum_{i=1}^{n} fix_i}{\sum_{i=1}^{n} fi}$$

$$G.M = Anti\log\left[\frac{\sum_{i=1}^{n} \log x_i}{n}\right]$$

$$G.M = Anti\log\left[\frac{\sum_{i=1}^{n} f_i \log x_i}{\sum_{i=1}^{n} f_i}\right]$$

$$H.M = \frac{n}{\sum_{i=1}^{n}\left[\frac{1}{x_i}\right]}$$

$$H.M = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n}\left[\frac{f_i}{x_i}\right]}$$

$$Median = \left(\frac{n+1}{2}\right) th\, value$$

$$Median = l + \frac{h}{f}\left[\frac{\sum_{i=1}^{n} f_i}{2} - C.F\right]$$

$$Mode = most\ frequent\, value$$

$$Mode = l + h\left[\frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)}\right]$$

$$Q_i = i\left(\frac{n+1}{4}\right) th\,value; \qquad i = 1,2,3$$

$$Q_i = l + \frac{h}{f}\sum \left[i\left(\frac{\sum_{i=1}^{n} f_i}{4}\right) - CF \right]; \; i = 1,2,3$$

$$D_i = i\left(\frac{n+1}{10}\right) th\,value; \qquad i = 1,2,\ldots,9$$

$$D_i = l + \frac{h}{f}\left[i\left(\frac{\sum_{i=1}^{n} f_i}{10}\right) - CF \right]; \qquad i = 1,2,\ldots,9$$

$$P_i = i\left(\frac{n+1}{100}\right) th\,value; \qquad i = 1,2,\ldots,99$$

$$P_i = l + \frac{h}{f}\left[i\left(\frac{\sum_{i=1}^{n} f_i}{100}\right) - CF \right]; \qquad i = 1,2,\ldots,99$$

$$M.D = \frac{\sum_{i=1}^{n} \left| x_i - mean \right|}{n}$$

$$M.D = \frac{\sum_{i=1}^{n} f_i \left| x_i - mean \right|}{\sum_{i=1}^{n} f_i}$$

$$Variance = \frac{\sum_{i=1}^{n} f_i \left(x_i - mean\right)^2}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i x_i^2}{\sum_{i=1}^{n} f_i} - \left(mean\right)^2$$

$$Variance = \frac{\sum_{i=1}^{n}(x_i - mean)^2}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - (mean)^2$$

$$c.o.v = \frac{S.D}{mean} \times 100\%$$

$Moments\ about\ origin\,(ungrouped\ data)$

$$\mu_k{}' = \frac{\sum_{i=1}^{n} x_i{}^k}{n}, \quad k = 1,2,3,4...$$

$Moments\ about\ origin\,(grouped\ data)$

$$\mu_k{}' = \frac{\sum_{i=1}^{n} f_i x_i{}^k}{\sum_{i=1}^{n} f_i}, k = 1,2,3,4...$$

$Moments\ about\ mean\,(ungrouped\ data)$

$$\mu_k = \frac{\sum_{i=1}^{n}(x_i - mean)^k}{n}, \quad k = 1,2,3,4...$$

$Moments\ about\ mean\,(grouped\ data)$

$$\mu_k = \frac{\sum_{i=1}^{n} f_i (x_i - mean)^k}{\sum_{i=1}^{n} f_i}, k = 1,2,3,4...$$

$$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2{}^{\frac{3}{2}}} \qquad \beta_2 = \frac{\mu_4}{\mu_2{}^2}$$

$$z = \frac{[x - mean]}{S.D} = \frac{[x - E(x)]}{\sqrt{V(x)}}$$

$$E(\bar{x}) = \mu \qquad V(\bar{x}) = \frac{\sigma^2}{n}$$

$$E\left(\sum_{i=1}^{n} x_i\right) = n\mu \qquad V\left(\sum_{i=1}^{n} x_i\right) = n\sigma^2$$

$$\mu = E(x) = \begin{bmatrix} \sum xf(x & \rightarrow discrete\,distribution \\ \int_{-\infty}^{\infty} xf(x) & \rightarrow continuous\ distribution \end{bmatrix}$$

$$E(x^2) = \begin{bmatrix} \sum x^2 f(x & \rightarrow discrete\,distribution \\ \int_{-\infty}^{\infty} x^2 f(x) & \rightarrow continuous\ distribution \end{bmatrix}$$

$$V(x) = E(x^2) - [E(x)]^2$$

$$\hat{y} = a + bx \qquad \rightarrow Linear\ \mathrm{Re}\,gression\,equation$$

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

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}}$$