WORK BOOK

MS-330 Applied Probability & Statistics

Sohail Ahmed

Lecturer

Department of Mathematics & Basic Sciences

NED University of Engineering & Technology, Karachi

Department of Mathematics & Basic Sciences

NED University of Engineering & Technology, Karachi-75270, Pakistan

WORK BOOK

MS-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 Probability & Statistics (MS-331) as prescribed by the NED University of Engineering & Technology, Karachi for the Academic Session _____

Sohail Ahmed

Lecturer,

Department of Mathematics & Basic Sciences

NED University of Engineering & Technology, Karachi

Date: _____

INDEX

Name: _____ **Roll No:** _____

S.NO.	DATE	PRACTICAL	GRADE			SIGN.
			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				
05		Probability Experiments				
06		Invoking Sub-Macro within a Macro using Control Statements				
07		Linear Regression Line and Correlation Analysis				
08		Multiple Regression Analysis and Non-Linear Models				
09		Discrete Probability Distributions				
10		Continuous Probability Distribution				
11		Sampling & Sampling Distribution (Discrete)				
12		Sampling & Sampling Distribution (Continuous) & Goodness of Fit Test				

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 I hope this workbook would be of great value for the engineering graduates.

PRE-REQUISITE:

Student should know the fundamental principles of statistics.

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 our every day lives. 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 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

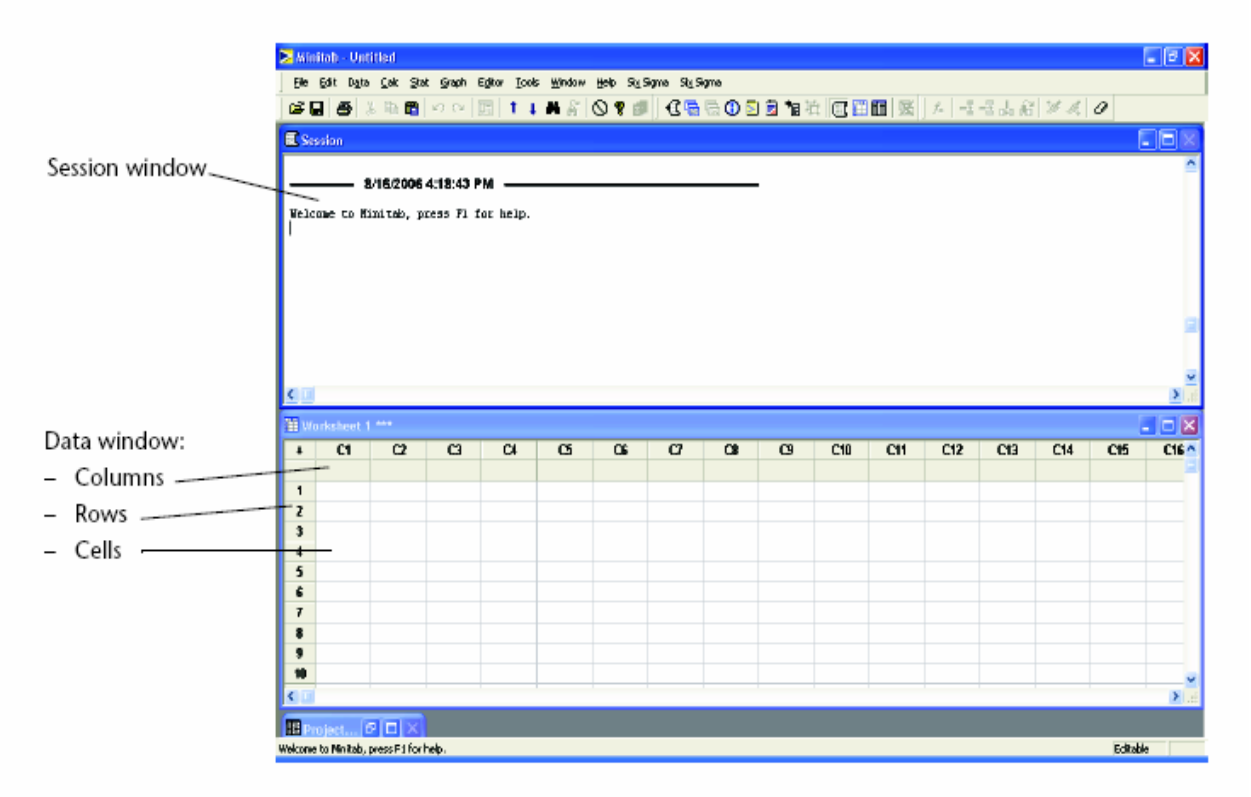
Minitab has six types of windows which can all be open at the same time: *Session, Data, Help, Info, History* and *Graph*. A window can be made active by selecting it. For example, to make the *Session* window active, place the mouse pointer on the window and just click it, if it is visible on your screen.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

LIST OF MINITAB SESSION COMMANDS

General Information

HELP command
INFO [C...C]
STOP

Managing Data

SET data into C
READ data into C…C
INSERT data [between rows K and K] of C...C
END
NAME E = 'name' ... E = 'name'
PRINT the data in E...E
SAVE [in file in "filename" or K]
RETRIEVE [file in "filename" or K]

Editing and Manipulating Data

CODE (K...K) to K ... (K...K) to K for C...C, put in C...C
DELETE rows K...K of C...C
ERASE E...E
INSERT data [between rows K and K] of C...C
LET C(K) = K
SORT C [carry along C...C] put into C [and C...C]
STACK (E...E) on ... on (E...E), put in (C...C)
UNSTACK (C...C) into (E...E) ... (E...E)

Arithmetic

LET E = expression
ADD E to E...E, put into E
SUBTRACT E from E, put into E
MULTIPLY E by E...E, put into E

DIVIDE E by E, put into E
RAISE E to the power E put into E
ABSOLUTE value of E put into E
SQRT of E put into E
LOGE of E put into E
LOGTEN of E put into E
EXPONENTIATE E put into E
ANTILOG of E put into E
ROUND E put into E
CENTER the data in C...C put into C...C COUNT the number of values in C [put into K] NMISS (number of missing values in) C [put into K] SUM of the values in C [put into K]
MEAN of the values in C [put into K]
STDEV of the values in C [put into K]
MEDIAN of the values in C [put into K]
MINIMUM of the values in C [put into K]
MAXIMUM of the values in C [put into K]
RCOUNT of E...E put into C
RN of E...E put into C RNMISS of E...E put into C RSUM of E...E put into C RMEAN of E...E put into C RSTDEV of E...E put into C RMEDIAN of E...E put into C RMINIMUM of E...E put into C RMAXMUM of E...E put into C

Distributions and Random Data

RANDOM K observations into C...C
PDF for values in E...E [put results in E...E]
CDF for values in E...E [put results in E...E]
INVCDF for values in E [put into E]
SAMPLE K rows from C...C put into C...C

Graphics

BOXPLOT of C...C CHART
C...C HISTOGRAM of
C...C STEM-AND-LEAF of
C...C PIECHART C...C
PLOT C vs C
TSPLOT [period = K] of C

Basic Statistics

CORRELATION C...C
DESCRIBE variables in C...C
ONET C...C
ONEZ C...C
PAIR C C
PONE C...C or K K...K
POWER
PTWO C C or K K K K
TWOSAMPLE test and CI [K% confidence] samples in C C
TWOT test with [K% confidence] data in C, groups in C
CORRELATION between C...C

Regression

REGRESS C on K predictors C...C
FITLINE y in C, predictor in C

Analysis of Variance

ANOVONEWAY for samples in C...C
ONEWAY data in C, levels in C

Non-parametric

KRUSKAL-WALLIS test for data in C, levels in C
MANN-WHITNEY two-sample rank test with [K% confidence] on C C
WTEST one-sample rank test [of median = K] on C...C

Tables

TALLY the data in C...C TABLE the data
classified C...C CHISQUARE test on table
stored in C...C

Quality Control

ICHART for C
PCHART number of nonconformities are in C...C, sample size = E
SCHART
XBARCHART
XSCHART

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, component 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	**Probability Experiments** Coin and die tossing experiments
06	**Invoking Sub-Macro within a Macro using Control Statements** Simulation of coin, die tossing experiments and evaluating probabilities. Construction of frequency distribution for given set(s) of measurements.

07	**Linear Regression Line and Correlation Analysis** Scatter diagram, correlation coefficient, linear regression line and prediction
08	**Multiple Regression Analysis and Non-Linear Models** Multiple regression model, quadratic and exponential curves
09	**Discrete Probability Distributions** Binomial distribution and Poisson distribution with applications
10	**Continuous Probability Distribution** Normal distribution, central limit theorem and applications
11	**Sampling & Sampling Distribution (Discrete)** Discrete sampling experiments, simulation of random samples from discrete distributions and verifies properties of sampling distribution.
12	**Sampling & Sampling Distribution (Continuous) & Goodness of Fit Test** Simulation random samples from continuous distributions and fitting of a Binomial / Poisson distribution.

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. Plot the given four relative functions for $0 \leq x \leq 64$ and step size h = 0.5, on the same Graph.

$$y = x \qquad y = x\log_2 x$$

$$y = x\left(\log_2 x\right)^2 \qquad y = x^2$$

Hint: $\log_a x = \dfrac{\log_{10} x}{\log_{10} a}$

3. 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…K10 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. $\sum_{i=1}^{5} x_i^{\,2}\, y_i$

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

v. $\sum_{i=1}^{5} \ln(y_i)$

vi. $\sum_{i=1}^{5} (2x_i + y_i - 3)$

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

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

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

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

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. Construct Multiple Bar diagram for the following table gives the results of students of a college for last three years.

Year	**Division**		
	First	**Second**	**Third**
2007	28	18	08
2008	32	20	10
2009	48	32	12

3. Construct Component Bar diagram for the following table gives the no. of students studying in a University.

Class (Year)	No. of Students		
	Boys	**Girls**	**Total**
First	250	50	300
Second	220	30	250
Third	80	20	100
Fourth	70	10	80

4. 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 & Box plot.

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

5. A large international organization rents homes for employees in Phoenix, Arizona. The following prices (in $) of 36 homes were randomly selected.

925	1095	665	895	1100	900	900	875	870	1150	1175	1050
925	895	1350	850	1065	1075	1035	1420	1250	1465	925	1275
1025	975	1020	800	1100	1285	935	875	875	1475	950	1325

i. Calculate Arithmetic mean, Geometric mean, Harmonic mean, Median, Range, Variance and Standard deviation.

ii. Compare above measures using default Minitab commands (if exist)

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 Frequency distribution, Relative and Cumulative Frequency distribution.

ii. Plot Histogram, Frequency curve and Cumulative frequency curve.

iii. Using Frequency distribution calculate Arithmetic mean, Geometric mean, Harmonic mean, Median, Mode, Range, Variance and Standard 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

2. Write a global macro for the following frequency distribution to calculate first four moments about origin, first four moments about mean and $\sqrt{\beta_1}$ $(measure of\ skewness)$, β_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. 05

Probability Experiments

1. Consider an experiment in which two fair coins are tossed 50 times and the total number of Heads recorded.
   i. Find proportion of heads in fifty tosses of a coins.
   ii. Plot the proportion of heads versus the number of tosses i.e. from n =1 to 50.

2. Determine different behavior for the proportion of heads (Problem No.1) using global macro.

3. Three coins are tossed simultaneously and no. of Heads recorded, experiment is repeated 100 times. If "X" represents the no. of heads i.e. X = 0, 1, 2, 3.
   Construct probability distribution of X.

4. In an experiment two dice are rolled 100 times and sum of dice is observed.

   i. What is probability that the sum is greater than 10 ?
   ii. What is probability that the sum is less than 7 ?
   iii. What is probability that the sum is equal to 8 ?
   iv. What are the theoretical probabilities of these events (i to iii)?
   v. Compare empirical and theoretical probabilities.

PRACTICAL NO. 06

Invoking Sub-Macro within a Global Macro using Control Statements

1. Write a global macro using control statements to construct frequency distribution of "N" measurements.

2. Write a global macro using control statements to construct probability distribution for following experiments

    i. Sum or Product of dice rolling problems
    ii. Coin tossing experiments.

PRACTICAL NO. 07

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

i. What is the average square footage and selling price of the homes in the sample?

ii. Construct Scatter plot.

iii. Calculate and interpret correlation coefficient (*r*)?

iv. Determine least square regression line. $\hat{y} = a + bx$ by performing arithmetic operations.

v. Draw regression line. $\hat{y} = a + bx$ and scatter plot on the same graph.

vi. Calculate sum of squared errors?

vii. Compare results of part (iii) & (iv) by default Minitab commands.

PRACTICAL NO. 08

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

i. Fit multiple regression model $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$

ii. 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.

i. Construct Scatter plot between size of home and electrical usage.

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

iii. Predict the monthly electrical usage of a 1600 square foot home.

PRACTICAL NO. 09

Discrete Probability Distributions

1. Write a global macro to 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?

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

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

4. 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. 10

Continuous Probability Distribution

1. Write a global macro to 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 $\mu = 50\,\text{min}$ *and* $\sigma = 5\,\text{min}$.

   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?

3. Suppose that the length of time between arrivals of customers at a franchise follows Exponential distribution, with mean time equal to 10 minutes.

   i. Find mean and standard deviation of *x*, the time between arrivals. Graph the probability distribution.

   ii. What is the probability that the inter arrival time is more than 15 minutes?

   iii. Find probability that the length of time between arrivals is within two standard deviations of the mean.

PRACTICAL NO. 11

Sampling & Sampling Distribution (Discrete)

1. Consider the selling prices of homes. Select 3 random samples of size five from 20 selling prices. Calculate mean and variance of each sample.

Home	Price ($)	Home	Price ($)
1	86000	11	118900
2	86600	12	125000
3	92000	13	130600
4	92500	14	139875
5	93500	15	144400
6	94000	16	148000
7	104000	17	156000
8	104500	18	151900
9	109900	19	159500
10	111900	20	163000

2. Write a global macro simulate 50 Random samples of size m = 5 from Binomial Distribution with 'n = 10' & 'p = 0.20'.

   i. Find out mean and variance of each sample.

   ii. Verify the following relations

   $E(\bar{x}) = \mu$, $E(s^2) = \sigma^2$ and $V(\bar{x}) = \frac{\sigma^2}{m}$

   Hint: $\mu = np$ *and* $\sigma^2 = np(1-p)$

PRACTICAL NO. 12

Sampling & Sampling Distribution (Continuous) and Goodness of Fit Test

1. Write a global macro to simulate 100 random samples of size n = 10 from Normal distribution with μ = 10 and σ = 4. using Global macro also verify the following relations.

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

2. Write a global macro for an experiment in which seven coins are tossed and number of heads recorded. If the experiment is repeated 128 times with the following distribution.

No. of Heads	0	1	2	3	4	5	6	7	**Total**
Frequency	7	6	19	35	30	23	7	1	**128**

Fit Binomial distribution and plot both observed and expected frequency on the same Graph if,

i. The coin is unbiased
ii. The nature of coin is unknown.

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
$\binom{n}{c}$ or $\binom{n}{x}$	No. of the combinations of n items selected x at a time
$\binom{n}{p}$ or $\binom{n}{x}$	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

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

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

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

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

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

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

$$Median = \left[\frac{n+1}{2}\right]th\ value$$

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

$$Mode = most\ frequent\ value$$

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

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

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

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

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

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

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

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

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

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

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

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

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)

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

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

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

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

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

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

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

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

$$\hat{y} = a + bx \qquad \rightarrow \textit{Linear Regression equation}$$

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

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

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

$$\chi^2_{calculated} = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i}$$

$$\chi^2_{tabulated} = \chi^2_{\alpha,\nu}$$

$d.f = \nu = n - k - 1$

where $n = no.\ of$ expected *frequency after adjust*

$k = no.of$ *parameter we estimate*

$e_i = N\,p(x_i)$

Critical Region = Reject H_0 *if* $\chi^2_{calculated} > \chi^2_{tabulated}$

$otherwise accept H_0$