B.Sc. SEMESTER VI - PRACTICAL
APPLIED STATISTICS - II
R - SOFTWARE AND MS EXCEL

Dr. A. Rajini
Prof. C. Jayalakshmi

B.Sc. Semester VI – Practical Book

Applied Statistics – II

R –Software and MS - Excel

First Edition

Authors

Dr. A. Rajini

Assistant Professor,

Department of Mathematics and Statistics,

Bhavan's Vivekananda College of Science, Humanities and Commerce,

Sainikpuri, Secunderabad, Telangana.

Prof. C. Jayalakshmi (Retd)

Department of Statistics,

University College of Science,

Osmania University,

Hyderabad, Telangana.

Iterative International Publishers

Title of the Book: B.Sc. Semester VI – Practical Book, Applied Statistics – II, R –Software and MS - Excel

First Edition - 2023

Copyright 2023 © Authors

Dr. A. Rajini, Assistant Professor, Department of Mathematics and Statistics, Bhavan's Vivekananda College of Science, Humanities and Commerce, Sainikpuri, Secunderabad, Telangana, India – 500094

Prof. C. Jayalakshmi (Retd), Department of Statistics, University College of Science, Osmania University, Hyderabad, Telangana, India – 500007.

No part of this book may be reproduced or transmitted in any form by any means, electronic or mechanical, including photocopy, recording or any information storage and retrieval system, without permission in writing from the copyright owners.

Disclaimer

The authors are solely responsible for the contents published in this book. The publishers don't take any responsibility for the same in any manner. Errors, if any, are purely unintentional and readers are requested to communicate such errors to the editors or publishers to avoid discrepancies in future.

ISBN: 978-93-5747-049-0

MRP: 260/-

Publisher, Printed at & Distribution by:
Selfypage Developers Pvt Ltd.,
Pushpagiri Complex,
Beside SBI Housing Board,
K.M. Road Chikkamagaluru, Karnataka.
Tel.: +91-8861518868
E-mail:info@iiponline.org

IMPRINT: I I P Iterative International Publishers

Preface

R – Software is a freely downloadable and highly useful for Statistical analysis. It's analysis is accepted as authenticated and valid mostly. We can also do programing in R. R-Software has numerous commands for analyzing datasets. Fishing out required command in R is really a tough task for new learners.

In recent times in the curriculum of B.Sc. Statistics, a practical paper is introduced in which majority practical are to be done using R-Software. It is observed that B.Sc. students as well as teachers, who are teaching R for the first time, finding it difficult to meet the need I practical class.

In order to facilitate an easy learning to B.Sc. Statistics students as well as teachers teaching for the first time, this book is prepared keeping the practical syllabus of B.Sc., semester VI, Statistical practical in view. This book is self-explanatory.

This book would meet the needs of students as well as teachers and research scholars, who are interested to learn R on their own. This book consists of not only explanation of the commands with detailed explanation of arguments but also some problems for exercise.

Definitely, this book would be meeting the need of B.Sc. Statistics semester VI students and teachers.

A. Rajini

Contents

Practical No.1 Generation Random Samples / Numbers 1 - 5

1.0 Theoretical Procedure 1
1.1 Generation of Uniform Radom Numbers 2
1.2 Generation of Random Numbers of Binomial Distribution 3
1.3 Generation of Random Numbers of Poisson Distribution 3
1.4 Generation of Random Sample using Normal Distribution 4
1.5 Generation of Exponential Random Numbers 4

Practical No.2 Fitting of Different Types of Curves 6 - 14

2.1 R- Procedure for Fitting of a Straight Line 6
2.2 R- Procedure for Fitting of a Parabola 8
2.3 R- Procedure for Fitting of a Power Curve of the Type $y = ax^b$ 9
2.4 R- Procedure for Fitting of a Exponential Curve $y = ae^{bx}$ 11

Practical No.3 Large Sample Tests 15 - 21

3.1 Test for Single Proportion 15
3.2 Test for Difference of Proportions 16
3.3 Test for Single Mean 17
3.4 Test for Difference of Mean (Independent Samples) 18
3.5 Test for Single Standard Deviation 18
3.6 Test for Difference of Standard Deviations 19

Practical No.4 Small Sample Tests - I 22 - 26

4.1 T-Test for Specified Mean 22
4.2 T-Test for Equality of Means 23
4.3 Paired T-Test (for Dependent Samples) 24
4.4 F – Test for Equality of Two Population Variances 25

Practical No.5 Small Sample Tests - II 27 - 31

5.1 χ^2 Test for Specified Variance 27
5.2 χ^2 Test for Independence of Attributes 28
5.3 χ^2 Test for Goodness of Fit 29

Practical No.6 Nonparametric Tests - I 32 - 37

6.1 One Sample Run Test 32
6.2 Sign Test for Single Sample 33
6.3 Sign Test for Two Samples 34
6.4 Wilcoxon Signed Rank Test for Single Sample 35
6.5 Wilcoxon Signed Rank Test for Two samples 36

Practical No. 7 Nonparametric Tests - II 38 - 41

7.1 Median Test 38
7.2 Wilcoxon Mann Whitney – U Test 39
7.3 Wald – Wolfowitz's Runs Test 40

Practical No.8 Analysis of Variance **42- 51**
8.1 Analysis of Variance for CRD data using MS – Excel 42
8.2 Analysis of Variance for CRD data using MS - Excel 42
8.3 Analysis of Variance for CRD Data Using R 44
8.4 Analysis of Variance for RBD data using MS – Excel 47
8.5 Analysis of Variance for RBD Data Using R 49

Practical No.9 Construction of life Tables **52- 54**
9.1 Construction of life tables using MS-Excel 52

Practical No. 1

Generation Random Samples/Numbers

This Lesson covers the following

1.0 Theoretical Procedures to Generate Random Samples/numbers using various different distributions.
1.1 Generation of Random Numbers of Uniform distribution
1.2 Generation of Random Numbers of Binomial distribution
1.3 Generation of Random Numbers of Poisson distribution
1.4 Generation of Random Numbers of Normal distribution
1.5 Generation of Random Numbers of Exponential distribution

1.0 Theoretical Procedure

1. **Uniform Distribution:** Always first Uniform random numbers between (0, m) will be generated and then converted into Uniform random numbers between (0,1). The procedure is as follows.

   Select three 4-digit or 3-digit odd numbers arbitrarily. Call them 'a', 'x_o' and 'm'. Then, $(i+1)^{th}$ Uniform random number between (0, m) is given by

   $$\mathbf{x_{i+1} = (x_i \cdot a) \bmod m} \;\; ; \;\; i = 1, 2, 3, 4, \ldots\ldots n$$

   Divide each of the generated random numbers by 'm' to get Uniform random numbers between (0, 1).

2. **Binomial Distribution:** In a Binomial Process the numbers are generated with the help of Bernoulli random numbers. Bernoulli random numbers (b_i's) are generated using Uniform random numbers (u_i's) in the range (0, 1) and Uniform random numbers (u_i's) can be generated by the method explained in **(a)** above.

   Let Binomial (n, p) random numbers are to be generated. Consider 'n' Uniform random numbers (u_i's) in the range (0, 1).

   If $u_i \leq p$ then $b_i = 1$
   If $u_i > p$ then $b_i = 0$

   Repeat this process for 'n' Uniform random numbers (u_i's) in the range (0, 1). Then, the first Binomial random number is given by

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

   Repeat the same procedure as many times as the number of Binomial random numbers are to be generated.

3. **Poisson Distribution:** In a Poisson Process the numbers are generated with the help of Exponential random numbers. The Exponential random numbers (e_i's) are generated by the formula

$$e_i = [-\log_e u_i] / \theta, \quad i = 1, 2, 3, \ldots\ldots, m$$

Sum the generated Exponential random numbers $\left(\sum_{l=1}^{k} e_l\right)$. The highest 'k' for which $\sum_{i=1}^{k} e_i \leq 1$ is the first Poisson random number. Then find another 'k' for another summation of Exponential random numbers for which $\sum_{l=1}^{k} e_l \leq 1$ which becomes the second Poisson random number and so on.

4. **Normal Distribution:** First generate '2m' Uniform random numbers between (0, 1) to generate 'm' Normal random numbers. By inverse transformation method the distribution function of Standard Normal distribution, the ith normal random number is given by

$$x_i = -[2 \log_e u_i]^{1/2} \operatorname{Cos}(2\pi u_{i+1}), \quad i = 1, 2, 3, \ldots\ldots, m$$

Where u_i and u_{i+1} are the ith and $(i+1)^{th}$ Uniform random numbers between (0, 1). The generated random number x_i follows Standard Normal distribution. To get a random number with specified parameters (μ, σ^2), use the conversion relation $x = \mu + z \cdot \sigma$.

5. **Exponential Distribution:** First generate 'm' Uniform random numbers between (0, 1). The Distribution Function of Exponential distribution is given by

$$F(x) = 1 - e^{-\theta x}, \quad x \geq 0, \quad \theta > 0$$

And F(x) follows Uniform Distribution in the interval [0, 1]. Therefore, by inverse transformation method, the ith Exponential random number is given by

$$e_i = [-\log_e u_i] / \theta, \quad i = 1, 2, 3, \ldots\ldots, m$$

where u_i is the ith Uniform random number between (0, 1).

1.1 Generation of Uniform Random Numbers

The Procedure for generation of Uniform random numbers is explained with the help of an example.

Example 1.1: Generate 10 Uniform Random Numbers between (-1, 1)

R- Code to generate a random sample:

```
set. seed (1)
r = 10 # no. of observations we require
a = -1 # 'a' and 'b' are the parameters of Uniform distribution
b = 1
r1 = runif(r, a, b)
r1
```

Detailed Explanation of the code arguments:

'set.seed( )' - is the recommended way to specify seeds.

'runif( )' - This function generates random numbers of uniform distribution on the given interval from min to max.

Output:
-0.4689827 -0.2557522 0.1457067 0.8164156 -0.5966361
0:7967794 0.8893505 0.3215956 0.2582281 -0.8764275

1.2 Generation of Random Numbers of Binomial Distribution

The Procedure of generation of random numbers of Binomial distribution is explained with the help of the following example.

Example 1.2: Generate 10 Binomial Random Numbers with the parameters n = 7 and p = 0.8

R- Code to generate a random sample:

```
set.seed (1)
r = 10 # no. of observations we require
n = 7 # 'n' represents the no. of trials
p = 0.8 # 'p' represents the probability of success
r2 = rbinom(r, n, p)
r2
```

Detailed Explanation of the code arguments:

'set.seed( )' - is the recommended way to specify seeds.
'rbinom( )' - This function gives the generated Binomial random numbers with parameters.

Output: 6 6 6 4 7 4 4 5 5 7

1.3 Generation of Random Numbers of Poisson Distribution

Generation of Poisson random numbers is explained with the help of the following example.

Example 1.3: Generate 10 Poisson Random Numbers with the parameter 5

R- Code to generate a random sample:

```
set.seed (1)
r = 10 # no. of observations we require
lambda = 5 # 'lambda' represents the parameter of the Poisson distribution
r3 = rpois(r, lambda)
r3
```

Detailed Explanation of the code arguments:

'set.seed( )' - is the recommended way to specify seeds.
'rpois( )' - This function generates the Poisson random numbers with parameter lambda = 5.

Output: 4 4 5 8 3 8 9 6 6 2

1.4 Generation of Random Sample using Normal Distribution

Generation of Normal distribution random numbers is explained with the help of the following example.

Example 1.4: Generate 10 Normal Random Numbers with mean = 5 and variance = 25.

R- Code to generate a random sample:

```
set.seed (1)
r = 10 # no. of observations we require
m = 5 # Mean of the Normal distribution
sd = 5 # Standard deviation of the Normal distribution
r4 = rnorm(r, m, sd)
r4
```

Detailed Explanation of the code arguments:

'set.seed( )' - is the recommended way to specify seeds.
'rnorm( )' - This function generates the Normal distribution random numbers with mean 'm' and standard deviation 'sd'.

Output:

1.8677309	5.9182166	0.8218569	12.9764040	6.6475389
0.8976581	7.4371453	8.6916235	7.8789068	3.4730581

1.5 Generation of Exponential Random Numbers

Procedure to generate Exponential distribution random numbers is explained by the following example.

Example 1.5: Generate 10 Exponential Random Numbers with parameter 2.

R- Code to generate a random sample:

```
set.seed (1)
r = 10 # no. of observations we require
teeta = 2 # 'teeta' is a parameter of the exponential distribution
r5 = rexp(r, teeta)
r5
```

Detailed Explanation of the code arguments:

'set.seed( )' - is the recommended way to specify seeds.
'rexp( )' - This function generates the Exponential random numbers with 'rate' (i.e., mean = 1/rate).

Output:

0.37759092	0.59082139	0.07285336	0.06989763	0.21803431
1.44748427	0.61478103	0.26984142	0.47828375	0.07352300

Exercise for Practice

1. Generate '20' random numbers from Uniform Distribution in the range (10, 15).
2. Generate '30' random numbers from Binomial Distribution with parameters (10, 0.55).
3. Generate '15' random numbers from Poisson Distribution with parameter (λ=5).
4. Generate '20' random numbers from Normal Distribution with parameters (22, 16).
5. Generate '25' random numbers from Exponential Distribution with parameter (λ= 8).

Practical No. 2

Fitting of Different Types of Curves

This Lesson covers the following curves

2.1 Fitting of straight line
2.2 Fitting of Parabola
2.3 Fitting of Power curve of the type $y = ax^b$,
2.4 Fitting of Exponential curve $y = ae^{bx}$

2.1 R- Procedure for Fitting of a Straight Line

All the Procedures of fitting different curves is explained with the help of examples.

Example 2.1: The following data is related to Bank deposits during the years 2005 to 2013. Fit a straight line by the method of least squares for the data and construct graph. Find the estimated values of Bank Deposits and draw the graph.

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013
Bank Deposits (Rs. Crores)	27	38	44	35	51	58	50	54	62

R- Code to generate a random sample:

```
x = c(2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013)
y = c(27, 38, 44, 35, 51, 58, 50, 54, 62)

fit_st = lm(y~x)
summary(fit_st)

predict_st = predict(fit_st, data.frame(x= x))
predict_st

table_st = data.frame(x, y, predict_st)
table_st

plot(x, y, col = "blue")
lines(x, predict_st, col = "red")
```

Detailed Explanation of the code arguments:

'lm( )' – is used to fit linear models.

'summary( )' – is a generic function used to produce summaries of the results of various model fitting functions.

'predict( )' – is a generic function for predictions from the results of various model fitting Functions

'data.frame( )' – used as the fundamental data structure and data displayed in a format as a table

'plot( )' – is the generic function for plotting in R. It can be used to create basic graphs.

'lines( )' – is used to add lines of different types, colors and width to an existing plot.

Output:

Coefficients:

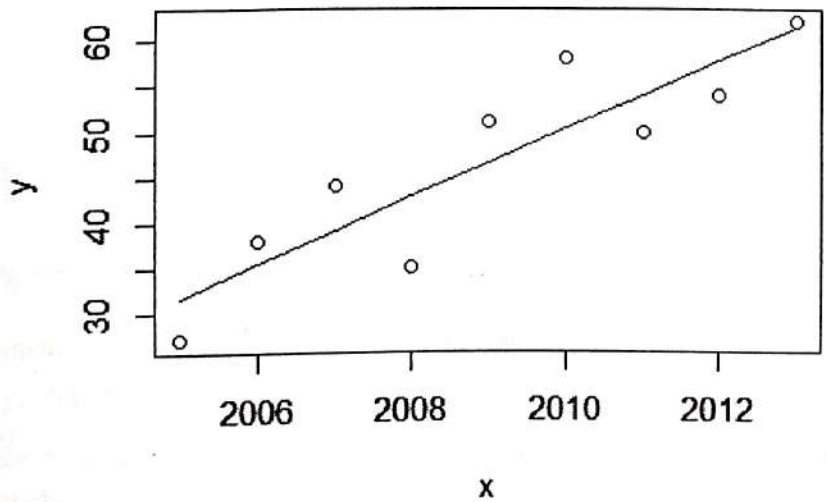
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept) -	7420.2278	1465.2193	-5.064	0.00146 **
x	3.7167	0.7293	5.096	0.00141 **

Fitted St. line equation is: y = -7420.2278 + 3.7167*x

Estimated Bank Deposits values are:

x	y	predict_st
2005	27	31.68889
2006	38	35.40556
2007	44	39.12222
2008	35	42.83889
2009	51	46.55556
2010	58	50.27222
2011	50	53.98889
2012	54	57.70556
2013	62	61.42222

Graph of Fitting of a Straight line is:



2.2 R- Procedure for Fitting of a Parabola

Example 2.2: The following data pertains to sales of mobile phones in a small city in India during 2010 to 2014. Fit a Second Degree parabola to this data and estimate the value of sales. Draw the graph.

Year	2010	2011	2012	2013	2014
Sales (Rs. Thousands)	10	12	14	10	8

R- Code to fit a parabola:

```
x = c(2010, 2011, 2012, 2013, 2014)
y = c(10, 12, 14, 10, 8)

fit_pr = lm(y~poly(x, 2, raw=TRUE))
fit_pr

predict_pr = predict(fit_pr, data.frame(x=x))
predict_pr

table_pr = data.frame(x, y, predict_pr)
table_pr

plot(x, y, col = "blue")
lines(x, predict_pr, col = "red")
```

Detailed Explanation of the code arguments:

'lm( )' – is used to fit linear models.

'poly( )' – Returns or evaluates orthogonal polynomial(s) of any specified degree in the data.

'summary( )' – is a generic function used to produce summaries of the results of various model fitting functions.

'predict( )' – is a generic function for predictions from the results of various model fitting functions

'data.frame( )' – used as the fundamental data structure and data displayed in a format as a table

'plot( )' – is the generic function for plotting in R. It can be used to create basic graphs.

'lines( )' – is used to add lines of different types, colors and width to an existing plot.

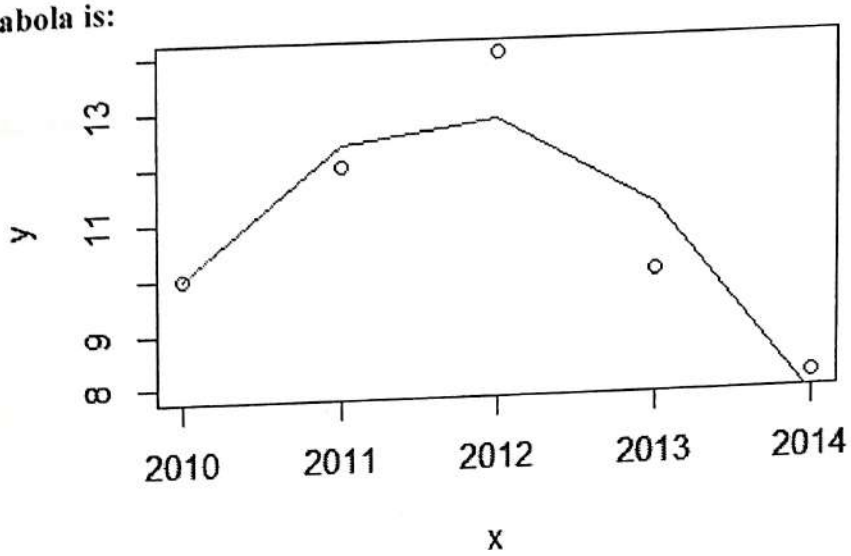
Output:

```
Coefficients: (Intercept)      poly(x, 2, raw = TRUE)1      poly(x, 2, raw = TRUE)2
                - 4046924                     4023                            -1
```

Fitted Second degree parabolic equation is: $y = -4046924 + 4023*x + (-1) * x^2$
Estimated values of Sales are:

X	y	predict_pr
2010	10	10
2011	12	12.4
2012	14	12.8
2013	10	11.2
2014	8	7.6

Graph of Parabola is:



2.3 R- Procedure for Fitting of a Power Curve of the Type $y = ax^b$

Example 2.3: Fit a Power curve equation to the following data consists of pressure of a gas and volume of the container. Find the estimated values of Volume and draw the graph.

Pressure:	0.5	0.56	0.63	0.71	0.83	0.99	1.21
Volume:	21	19	17	15	13	11	9

R- Code to fit a power curve:

```
x = c(0.5, 0.56, 0.63, 0.71, 0.83, 0.99, 1.21)
y = c(21, 19, 17, 15, 13, 11, 9)


log_y = log10(y)
log_y

log_x = log10(x)
log_x

fit_pc = lm(log_y~log_x)
summary(fit_pc)

a = 10^1.035451
```

```
a
b = -0.960417

predict_pc = a*x^b
predict_pc

table_st = data.frame(x, y, predict_pc)
table_st

plot(x, y, col = "blue")
lines(x, predict_pc, col = "red")
```

Detailed Explanation of the code arguments:

'log10' – computes common (i.e., base 10) logarithms.

'lm( )' – is used to fit linear models.

'poly( )' – Returns or evaluates orthogonal polynomial(s) of any specified degree in the data.

'summary( )' – is a generic function used to produce summaries of the results of various model fitting functions.

'predict( )' – is a generic function for predictions from the results of various model fitting functions

'data.frame( )' – used as the fundamental data structure and data displayed in a format as a table

'plot( )' – is the generic function for plotting in R. It can be used to create basic graphs.

'lines( )' – is used to add lines of different types, colors and width to an existing plot.

Output:

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.035451	0.001188	871.2	3.78e-14 ***
log_x	-0.960417	0.006559	-146.4	2.82e-10 ***

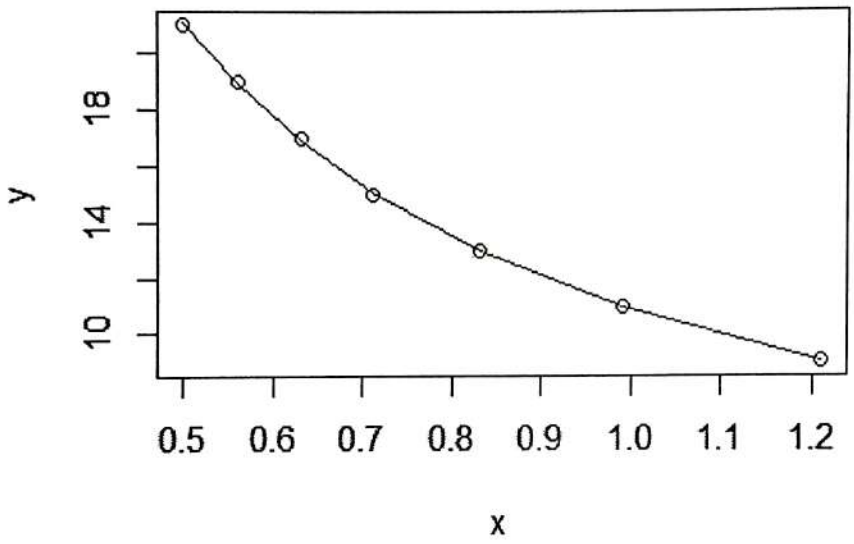
Here, A = 1.035451
b = - 0.960417
a = $antilog(A)$ = 10.85053

Fitted Power Curve equation is: y = 10.85053 * $x^{-0.960417}$

Estimated values of volume are:

X	y	predict_pc
0.5	21	21.113748
0.56	19	18.936316
0.63	17	16.91094
0.71	15	15.076655
0.83	13	12.976865
0.99	11	10.955773
1.21	9	9.035299

Graph of Power curve is:



2.4 R- Procedure for Fitting of a Exponential Curve $y = ae^{bx}$

Example 2.3: Use data of the population of the US from the table below to fit an Exponential curve. Find the estimated values of Population and draw the graph. \

Year:	1950	1960	1970	1980	1990	2000	2010	2020
Population (Millions):	62.9	76	92	105.7	122.8	131.7	151.3	179.3

R- Code to fit an Exponential curve:

```
x = c(1950, 1960, 1970, 1980, 1990, 2000, 2010, 2020)
y = c(62.9, 76, 92, 105.7, 122.8, 131.7, 151.3, 179.3)

log_y = log10(y)
log_y

fit_ec = lm(log_y~x)
summary(fit_ec)

a = 10^-10.28
a
```

```
e = exp(1)
e

b = 0.006205/log10(2.7182)
b

predict_ec = a*exp(b*x)
predict_ec

table_ec = data.frame(x, y, predict_ec)
table_ec

plot(x, y, col = "blue")
lines(x, ec, col = "red")
```

Detailed Explanation of the code arguments:

'log10' – computes common (i.e., base 10) logarithms.

'lm( )' – is used to fit linear models.

'poly( )' – Returns or evaluates orthogonal polynomial(s) of any specified degree in the data.

'summary( )' – is a generic function used to produce summaries of the results of various model fitting functions.

'exp( )' – computes the exponential function.

'predict( )' – is a generic function for predictions from the results of various model fitting functions

'data.frame( )' – used as the fundamental data structure and data displayed in a format as a table

'plot( )' – is the generic function for plotting in R. It can be used to create basic graphs.

'lines( )' – is used to add lines of different types, colors and width to an existing plot.

Output: Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.028e+01	5.391e-01	-19.07	1.35e-06 ***
x	6.205e-03	2.716e-04	22.85	4.60e-07 ***

Here, A = log a = -1.028e+01 = -10.28
B = b* log e = 6.205e-03 = 0.006205

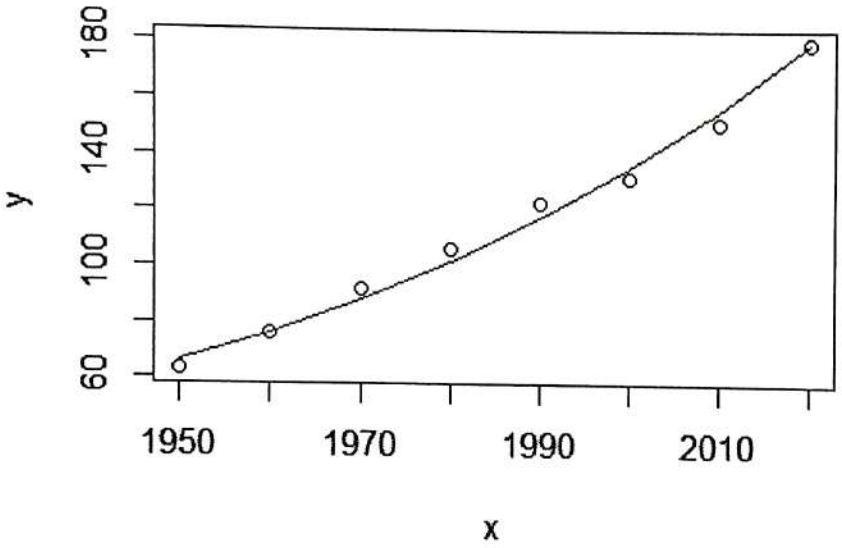
$\hat{a} = antiliog(\hat{A})$ = 5.248075e-11
$\hat{b} = \frac{\hat{B}}{\log e}$ = 0.01428797

Fitted Exponential Curve equation is: $y = (5.248075e\text{-}11)e^{0.01428797*x}$

Estimated values of Population are:

X	y	predict_ec
1950	62.9	66.08673
1960	76	76.23706
1970	92	87.94638
1980	105.7	101.45416
1990	122.8	117.03661
2000	131.7	135.01238
2010	151.3	155.74907
2020	179.3	179.67073

Graph of Exponential curve is:



Exercise Problems for Practice

1. Fit a straight line to the following data consisting of temperature (^{0}C) and Volume (cm^3). Estimate the values of Volume and draw the graph.

Temperature	10	20	25	30	35	50	80
Volume	141.4	144.8	147.8	149.7	153.1	160.5	175.4

2. The following data gives the age of mother and average number of births per 1,000 women of that age. Fit a parabolic curve, estimate the average number of births and draw the graph.

Age	16	18.5	22	27	32	37	42
Births ('000)	34	86.5	111.1	113.9	84.5	35.4	6.8

3. When a rocket is launched from the ground, its upward velocity V(t) (m/s) is measured at a certain time instant(s) t(s). Suppose that one measures the upward velocity of a rocket for time 0 to 30 and the measurements are tabulated as follows.

T(S)	0	10	15	22	25	30
V(T) (m/s)	0	250	350	655	890	910

i) Fit an exponential cure and draw the graph.
ii) Fit a power curve and draw the graph.

Practical No. 3

Large Sample Tests

This Lesson covers the following

3.1 Test for single proportion
3.2 Test for Difference of Proportions
3.3 Test for Single mean
3.4 Test for difference of means
3.5 Test for Single standard deviation
3.6 Test for difference of standard deviation

All the above Test Procedures will be explained with the help of a numerical examples.

3.1 Test for Single Proportion

Example 3.1: Two hundred people were attacked by a disease and only 180 survived. Will you accept or reject the hypothesis that the survival rate of people attacked by the disease is 85% against the hypothesis that it is more than 85%, at 5% level of significance?

R- Code to Test for single proportion:

H_0: P = 0.85; i.e., Survival rate is 85%.
H_1: P > 0.85; i.e., Survival tare is more than 85% (RTT)

 # Given,
n = 200 # Sample Size
x = 180
P = 0.85 # Population Proportion
prop.test(x = 180, n = 200, p = 0.85, alternative = "greater")

Note: If the alternative hypothesis is two tailed, then we take "two.sided" in alternative. If the test procedure is left tailed then we take "less" in alternative. If the test procedure is right tailed then we take "greater" in alternative.

Detailed Explanation of the code arguments:

'prop.test( )' – used for testing the null that the proportions (probabilities of success) in several groups are the same, or that they equal certain given values.

Output;

1-sample proportions test with continuity correction
data: 180 out of 200, null probability 0.85
X-squared = 3.5392, df = 1, **p-value = 0.02997**
alternative hypothesis: true p is greater than 0.85

95 percent confidence interval:
0.8567544 1.0000000
sample estimates:
p = 0.9

Conclusion:

Here, p-value = 0.02997
p-value < 0.05 Hence, we Reject H_0.
i.e., we conclude that Survival rate is more than 85%.

3.2 Test for Difference of Proportions

Example 3.2: A random sample of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women were in favor of the proposal. Test the hypothesis that proportion of men and women in favor of the proposal are same or not at a 5% level of significance?

R- Code to Test for difference of proportions:

 # H_0 : $P_1 = P_2 = P$; i.e., there is no significant difference between the proportion of opinions of men and women about flyover.

 # H_1 : $P_1 \neq P_2 = P$; i.e., there is a significant difference between the proportion of opinion of men and women about flyover. (Two Tailed Test)

 # Given,

$n_1 = 400$
$n_2 = 600$
$x_1 = 200$
$x_2 = 325$

```
prop.test(x = c(200, 325), n = c(400, 600), alternative = "two.sided")
```

Detailed Explanation of the code arguments:

'prop.test( )' – used for testing the null that the proportions (probabilities of success) in several groups are the same, or that they equal certain given values.

Output;

2-sample test for equality of proportions with continuity correction
data: c(200, 325) out of c(400, 600)
X-squared = 1.5079, df = 1, p-value = **0.2195**
alternative hypothesis: two.sided
95 percent confidence interval:
-0.10691965 0.02358631

Sample estimates:

Prop 1	Prop 2
0.5000000	0.5416667

Conclusion:

Here, p-value = 0.2195
p-value > 0.05 Hence, we Accept H_0.
i.e., we conclude that proportion of men and women about opinion flyover are same.

3.3 Test for Single Mean

Example 3.3: A sample of 400 individual is found to have a mean height of 67.47 inches, standard deviation is 1.3 inches. Can it be reasonable to conclude that the sample is drawn from a population with mean height more than 67.39 inches, at 1% level of significance?

R- Code to Test for single mean:

```
# H0: μ = 67.39; i.e., the sample has drawn from the population with mean height
                       67.39 inches.
# H1: μ > 67.39; i.e., the mean height of the parent population is more than 67.39
                       inches (Right Tailed Test).
# Given,
    n = 400                 # Sample Size
    s.m. = 67.47            # Sample mean
    pop.m. = 67.39          # Population mean
    s.sd. = 1.3             # Sample standard deviation
    z = (s.m. - pop.m.)/(s.sd./sqrt(n))
    cat("Z-cal = ", z)
    pvalue = pnorm(z)
    cat("p-value = ",pvalue)
```

Detailed Explanation of the code arguments:

'pnorm' – calculates the c. d. f. (cumulative distribution function) where X is a normal variate.

Output:

```
Z-cal = 1.230769
p-value = 0.8907954
```

Conclusion:

Here p-value > 0.01. Hence, we Accept H_0.
i.e., the sample has been drawn from the population whose average height = 67.39 inches.

3.4 Test for Difference of Mean (Independent Samples)

Example 3.4 : The average hourly wages of a sample of 150 workers in a plant A was Rs.2.56 with a standard deviation of Rs.1.08. The average wage of a sample of 200 workers in plant B was Rs.2.87 with the standard deviation of Rs.1.28. Can an applicant safely assume that hourly wages paid by plant B is higher than plant A? Test at 5% level of significance.

R- Code to Test for difference of mean:

 # H_0: $\mu_1 = \mu_2$; i.e., Hourly wages paid by both the plants are equal
 # H1: $\mu_1 < \mu_2$; i.e., Hourly wages paid by plant B is higher than plant A
(Left Tailed Test)

```
#Given,
        n1 = 150        # First Sample size
        n2 = 200        # Second Sample size
        fsm = 2.56      # First Sample mean
        ssm = 2.87      # Second sample mean
        s1 = 1.08       # First Sample sd
        s2 = 1.28       # Second Sample sd
        z = (fsm - ssm)/sqrt((s1^2/n1)+(s2^2/n2))
        cat("Z-cal = ", z )
        pvalue = pnorm(z)
        cat("p-value = ", pvalue)
```

Detailed Explanation of the code arguments:

'pnorm' – calculates the c. d. f. (cumulative distribution function) where X is normal.

Output:

Z-cal = -2.45322
p-value = 0.007079193

Conclusion:

Here p-value < 0.05 Hence, we reject H_0.
i.e., Hourly wages paid by plant B is higher than plant A.

3.5 Test for Single Standard Deviation

Example 3.6 : A random sample of size 50 has Standard Deviation 11.8 and has been drawn from a normal population. Can we conclude that the sample has been drawn from the population with Standard Deviation 10. Level of significance is 5%.

R- Code to Test for single standard deviation:

 # H_0: σ (psd) = 10; i.e., the sample has been drawn from the population
 # H_1: σ(psd) $\neq$ 10; i.e., the sample has not been drawn from the population.
(Two Tailed Test)

```
#Given,
        n = 50          # Sample size
        ssd = 11.8      # Sample standard deviation
        psd = 10        # Population standard deviation
        z = (ssd – psd)/(psd/sqrt(2*n))
        cat("Z-cal = ", z )
        pvalue = pnorm(z)
        cat("p-value = ", pvalue)
```

Detailed Explanation of the code arguments:

'pnorm' – calculates the c. d. f. (cumulative distribution function) where X is normal.

Output:

Z-cal = 1.8
p-value = 0.9640697

Conclusion:

Here p-value > 0.05 Hence, we Accept H_0.
i.e., we conclude that, the sample has been drawn from the population.

3.6 Test for Difference of Standard Deviations

Example 3.6: Two random samples of sizes 100 each have drawn from two normal populations with the standard deviations 2.823 and 1.548, respectively. Test for the significant difference between the sample standard deviations, if the population standard deviation is 2, at 5% level of significance.

R- Code to Test for difference of standard deviations:

H_0: $\sigma_1 = \sigma_2 = 2$; i.e., the samples are drawn from populations with equal variance.
H_1: $\sigma_1 \neq \sigma_2$; i.e., the samples are not drawn from the populations with equal variance.
(Two Tailed Test)

```
#Given,
    n1 = 100        # First Sample size
    n2 = 100        # Second Sample size
    fssd = 2.823    # First Sample standard deviation
    sssd = 1.548    # Second Sample standard deviation
    psd = 2         # Population standard deviation
    z = (fssd - sssd)/(psd*(sqrt((1/(2*n1))+(1/(2*n2)))))
    cat("Z-cal = ", z )
    pvalue = pnorm(z)           # H1 is TTT
    cat("p-value = ", pvalue)
```

Detailed Explanation of the code arguments:

'pnorm' – calculates the c. d. f. (cumulative distribution function) where X is normal.
Output:

Z-cal = 6.375
p-value = 1

Conclusion:

Here p-value > 0.05 Hence, we Accept H_0.
i.e., we conclude that, two samples were drawn from the populations with equal variance.

Exercise Problems for Practice

1. In a sample of 1,000 people in a village, 540 are educated and the rest are uneducated. Can we conclude that both educated and uneducated are equal in numbers? Test at 1% level of significance.

2. 1,000 articles from a factory A are tested and found to have 3% defects. 1,500 similar articles from factory B are found to have 4% defects. Can it be concluded that articles of first factory are inferior to second factory? Test at 5% level of significance.

3. In 2010, the market share of a two-wheeler manufacturer was 53%. In 2015, market share of the same two-wheeler manufacturer was 43%. Suppose this data is based on random sample of 100 two-wheeler customers in 2010 and 2015. Is it reasonable to conclude that there is significant change in market share of the two-wheeler manufacturer? Test at 5% level of significance.

4. A company manufacturing a new model car during 2017 claimed that the model is fuel efficient when compared to previous models. It's average mileage is 25.6 km/liter with a standard deviation of 1.5 km/liter. The first 50 cars were tested for fuel efficiency under normal conditions and found average of 23.6 km/liter. Would you conclude that the new model is fuel efficient? Test at 5% level of significance.

5. An ambulance service claim that it takes on an average 8.9 mins to reach its destination in emergency calls. To check this claim, the agency which licensed ambulance services, has put them on 40 emergency calls. They got a mean of 9.3 mins with a standard deviation of 1.6 mins to reach its destination. Test, at 5% level of significance, whether the claim is valid or not.

6. A sample of 32 Mutual Fund Companies was taken in January 2015 and the average rate of return over the past 32 days was found to be 3.23% and the sample standard deviation was 0.51%. Before one year, a sample of 38 Mutual Fund Companies showed that average rate of return was 4.36% and the sample standard deviation was 0.84%. Is it reasonable to conclude that Mutual Fund market interest rate declined during 2015? Test, at 1% level of significance.

7. A random sample of 100 students gave a mean weight of 56 kg with standard deviation of 4.36 kg. Test the Hypothesis that the standard deviation of weight in the population is 4.84 kg. Use 1% level of significance.

8. The surface hardness of a composite metal sheet is known to have variance of 0.40. From a shipment received, the sample variance from a random sample of 35 sheets found to be 0.22. Is it right to conclude that this shipment has a significant variance difference from 0.40? Test, at 5% level of significance.

9. The sum of squares of deviations based on two samples each of size 50 and 75 was found to be 2350 and 4500, respectively. Test at 5% level of significance, that the samples are drawn from the population with equal variances.

10. 30 observations fr
observations was 3
sum of squares o
second samples co

10. 30 observations from a population showed mean of 8.5 kgs and the sum of squares of observations was 3000. Observations from another sample showed a mean of 9.3 kgs and sum of squares of values was 4700. Verify at 1% level of significance, whether the second samples comes from a population with lesser variance.

Practical No. 4

Small Sample Tests - I

This Lesson covers the following

4.1 t – test for specified mean (single mean)
4.2 t – test for equality of means (two means – independent samples)
4.3 Paired t – test (two means – dependent samples)
4.4 F – test for equality of two population variances

All the test procedure based on t-distribution and F-distribution are explained with the help of appropriate numerical examples.

4.1 T-Test for Specified Mean

Example 4.1: The observed weights of 10 adult males are as given below (in kgs).

62, 64, 67, 71, 69, 68, 70, 71, 72, 66.

Test whether the average weight of the adult males is below 66 Kgs, assuming that the sample is drawn from a normal population, at 5% level of significance.

R- Code to t-test for specified mean:

 # H_0: $\mu = 66$; i.e, the average weight of the males is 66 kgs.
 # H_1: $\mu < 66$; i.e, the average weight of the males is below 66 kgs.
(One tailed test)

 # Given,

```
mu = 66  # Population mean
x = c(62, 64, 67, 71, 69, 68, 70, 71, 72, 66)
t.test(x, mu = 66, alternative = "less")
```

Detailed Explanation of the code arguments:

't.test( ) ' – Performs one and two sample t-tests on vectors of data.
'alternative' – is a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less". You can specify just the initial letter.

Output:

```
One Sample t-test
data:  x
t = 1.9365, df = 9, p-value = 0.9576
alternative hypothesis: true mean is less than 66
95 percent confidence interval: -Inf  69.89323
Sample estimates: mean of x =  68
```

Conclusion:

Here, p-value = 0.9576
p-value > 0.05 Hence, we Accept H_0.
i.e., the average weight of the males is equal to 66 kgs.

4.2 T-Test for Equality of Means

Example 4.2: The observed heights of six randomly chosen sailors are given as

63, 65, 68, 69, 71 and 72 (in inches).

Another group of 10 randomly chosen soldiers are having their heights as

61, 62, 65, 66, 69, 69, 70, 71, 72 and 73 (in inches).

Test whether sailors in first sample are, on the average, taller than soldiers in second sample, at 5% level of significance under the assumption that the samples have been drawn from normal population.

R- Code to t-test for equality of means:

```
# H0: μ1 = μ2; the height of sailors and soldiers are same.
# H1: μ1 > μ2; Sailors are taller than soldiers.
x = c(63, 65, 68, 69, 71, 72)
y = c(61, 62, 65, 66, 69, 69, 70, 71, 72, 73)
t.test(x, y, var.equal = TRUE, alternative = "greater")
```

Detailed Explanation of the code arguments:

't.test( ) ' – Performs one and two sample t-tests on vectors of data.
'alternative' – is a character string specifying the alternative hypothesis, must be one of "two. sided" (default), "greater" or "less". You can specify just the initial letter.

Output:

```
Two Sample t-test data: x and y
data:  x and y

t = 0.099154, df = 14, p-value = 0.4612
alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval: -3.352686 Inf

Sample estimates: mean of x   mean of y
                     68.0        67.8
```

Conclusion:

Here, p-value = 0.4612
p-value > 0.05 Hence, we Accept H_0.
i.e., we conclude that, the height of sailors and soldiers are same.

4.3 Paired T-Test (for Dependent Samples)

Example 4.3 : A treatment was given to 10 patients, for improving HB percentage in blood, and readings have been taken before and after applying the treatment. Assume that the samples are drawn from normal populations.

Patient Number	1	2	3	4	5	6	7	8	9	10
HB Percent before treatment	11.94	11.99	11.98	12.03	12.03	11.96	11.95	11.96	11.92	12.00
HB Percent after treatment	12.00	11.99	11.95	12.07	12.03	11.98	12.03	12.02	12.01	11.99

Can we conclude that HB percent was increased after treatment? Use 5% level of significance.

R- Code to Paired t-test:

```
# H0: μ1 = μ2;  HB percent was not increased even after treatment was given.
# H1: μ1 < μ2;  HB percent was increased after given the treatment.
x = c(11.94, 11.99, 11.98, 12.03, 12.03, 11.96, 11.95, 11.96, 11.92, 12.00)
y = c(12.00, 11.99, 11.95, 12.07, 12.03, 11.98, 12.03, 12.02, 12.01, 11.99)

t.test(x, y, paired = TRUE, alternative = "less")
```

Detailed Explanation of the code arguments:

't.test( ) ' – Performs one and two sample t-tests on vectors of data.
'alternative' – is a character string specifying the alternative hypothesis, must be one of "two. sided" (default), "greater" or "less". You can specify just the initial letter.

Output:

```
Paired t-test
Data:  x and y
t = -2.3941, df = 9, p-value = 0.02014
Alternative hypothesis: true difference in means is less than 0
95 percent confidence interval: -Inf  -0.007263724
Sample estimates: mean of the differences  =  -0.031
```

Conclusion:
Here, p-value = 0.02014
p-value < 0.05 Hence, we Reject H_0.
i.e., we conclude that, HB percent was increased after the treatment is given.

4.4 F – Test for Equality of Two Population Variances

This Procedure of F – test for equality of two normal population variances is explained with the help of the following example.

Example 4.4: Below you can find the study hours (per month) of 10 female students and 8 male students, preparing for a competitive examination. Test whether the two samples have the same population variances, at 5% level of significance. Assume that the samples have been drawn from normal populations.

Female	61	66	67	85	78	63	85	86	88	91
Male	60	65	71	74	76	82	85	87		

R- Code to F – test for equality of two population variances:

```
#H0 : Two samples have drawn from the same normal population
#H1 : Two samples have not drawn from the same population
x = c(61, 66, 67, 85, 78, 63, 85, 86, 88, 91)
y = c(60, 65, 71, 74, 76, 82, 85, 87)
var. test(x, y, alternative = "two. sided")
```

Detailed Explanation of the code arguments:

'var.test( ) ' – Performs an F test to compare the variances of two samples from normal populations.

'alternative' – is a character string specifying the alternative hypothesis, must be one of "two. sided" (default), "greater" or "less". You can specify just the initial letter.

Output:

```
F test to compare two variances
data:  x and y
F = 1.4675, num df = 9, denom df = 7, p-value = 0.6271
Alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3042586 6.1591879
Sample estimates:
Ratio of variances
        1.467505
```

Conclusion:

Here, p-value = 0.6271
p-value > 0.05 Hence, we Accept H0.
i.e., Two samples have been drawn from the same normal population.

Exercise Problems for Practice

1. The following is a random sample of 15 measurements of thickness of plastic sheet (in inches) used for making chairs

   5. 46, 7.48, 8.99, 6.44, 8.95, 6.48, 7.96, 4.96, 5.98, 2.99, 8.63, 8.24, 7.36, 4.05, 7.77.

   Test whether the sample of plastic sheet is drawn from a normal population with a mean of 6.45 inches. Test at 5% level of significance.

2. Prices of shares of a company on 10 days are found to be 65, 66, 69, 70, 69, 71, 63, 72, 64, 68. Can it be concluded that the sample of share prices has come from a normal population with an average 65? Test at 1% level of significance.

3. Two samples of sizes 14 and 21 are drawn from two different normal populations. The observed values are as follows.

Sample 1:	14	21	28	32	34	18	21	46	38	29	52	40	28	16							
Sample 2:	21	32	37	28	40	42	46	55	20	31	25	64	71	45	20	49	30	37	38	26	46

   Test at 5% level of significance, whether the two sample means differ significantly or not.

4. Recorded values of Rainfall (in cms) at two different places A and B for 10 years are as follows:

Place A	40	30	34	39	43	45	49	40	45	55
Place B	39	28	34	35	41	23	45	37	43	55

   Is there any significant difference in rainfall for two places A and B. Test at 1% level of significance. Assume that the samples were drawn from normal populations

5. In an experiment two types of food A and B are fed to the same group of 15 pigs. Below given are the gain in weights of pigs (in kgs).

Diet A	25	32	30	34	24	14	32	24	30	31	35	25	24	14	32
Diet B	44	34	22	10	47	31	40	30	32	35	18	21	35	29	22

   Can we conclude that food A is better than food B for gain in weights. Test at 1% level of significance, assuming that the samples were drawn from normal populations.

6. The following samples are drawn from normal two populations. Test whether the samples have come from two populations with equal variances. Test at 5% level of significance.

Sample 1	66.5	95.2	88.8	23.3	59.3	43.3	20.5	83.2	29.3	83.6			
Sample 2	60.5	12.1	78.8	29.2	45.7	76.4	91.5	36.7	69.7	4.7	15.1	89.5	49.1

Practical No. 5

Small Sample Tests – II

This Lesson covers the following

5.1 χ^2 test for specified variance
5.2 χ^2 test for Independence of Attributes
5.3 χ^2 test for Goodness of fit

5.1 χ^2 Test for Specified Variance

The Procedure of χ^2 test for specified variance is explained with the help of the following example.

Example 5.1 : A life time of a certain brand of bulb (in hours) produced by XYZ company is as follows:

3360, 3720, 3300, 3420, 3240, 3420, 3450, 3540, 3750, 3780

Test whether the variance is 3300 or not at 5% level of significance.

R- Code to χ^2 test for specified variance:

 # H_0: $\sigma^2 = 3300$; i.e., Variance of life time of the bulbs is equal to 3300.
 # H_1: $\sigma^2 \neq 3300$; i.e., Variance of life time of the bulbs is not equal to 3300.

```
x = c(3360, 3720, 3300, 3420, 3240, 3420, 3450, 3540, 3750, 3780)
pop.v = 33000    # Population Variance

n = length(x)  # No. of observations in the given data
v = var(x)
sv = ((n-1)/n)*v
chisquare = n*sv/pop.v
cat("chisquare = ", chisquare)
pvalue = 1-pchisq(chisquare, n-1)
cat("pvalue = ", pvalue)
```

Detailed Explanation of the code arguments:

'pchisq( )' – is used to compute cumulative chi square density for a vector of elements.

Output:

```
Pearson's Chi-squared test
Chi-square = 10.10182
p-value = 0.3423055
```

Conclusion:

Here, p –value = 0.3423055
p-value > 0.05 Hence, we Accept H_0.
i.e., Variance of life time of the bulbs is equal to 3300.

5.2 χ^2 Test for Independence of Attributes

Procedure of χ^2 test for Independence of Attributes is also explained with the help of the example given below.

Example 5.2 : Data on hair color and the eye color, of a group of people in a city, are given in the table below. Determine the association between the hair color and the eye color. Test at 5% level of significance.

		Hair color		
		Fair	**Brown**	**Black**
Eye color	**Blue**	15	20	5
	Grey	20	20	10
	Brown	25	20	15

R- Code to χ^2 test for Independence of Attributes

 # H_0 : There is no association between hair color and eye color of a person.
 # H_1 : There is an association between hair color and eye color of a person.

```
x = c(15,20,25,20,20,20,5,10,15)
mx = matrix(x, nrow = 3, ncol = 3)
chisq.test(mx)
```

Detailed Explanation of the code arguments:

'matrix( )' – Creates a matrix from the given set of values.
'chisq.test( )' – Performs chi-squared contingency table tests and goodness-of-fit tests.

Output:

Pearson's Chi-squared test data: mx
X-squared = 3.6458, df = 4, p-value = 0.4561

Conclusion:

Here, p-value = 0.4561
p-value > 0.05 Hence, we Accept H_0.
i.e., There is no association between hair color and eye color of a person.

5.3 χ^2 Test for Goodness of Fit

The χ^2 test procedure for Goodness of fit of a distribution is explained with an example as given below.

Example 5.3.1 : Records taken on the number of births in 800 families having four children in a city in India are given below:

No. of births	0	1	2	3	4
Frequency	32	178	290	236	64

Test whether the hypothesis that binomial law holds for this data with chance of male birth is equal to that of female birth. Test at 5% level of significance.

R- Code to χ^2 test for Goodness of fit:

```
# H0 : The given data follows a binomial distribution.
# H1 : The given data does not follows a binomial distribution.
x = c(0,1,2,3,4)
f= c(32,178,290,236,64)
# Fitting of binomial distribution
n = max(x)
m = sum(f*x)/sum(f)
p = m/n
px = dbinom(x,n,p)
ef = sum(f)*px
chisq.test(f, ef)
```

Detailed Explanation of the code arguments:

'dbinom( )' – calculates the probability function of the binomial distribution.
'chisq.test( )' – Performs chi-squared contingency table tests and goodness-of-fit tests.

Output:

Pearson's Chi-squared test data: f and ef
X-squared = 20, df = 16, p-value = **0.2202**

Conclusion:

Here, p-value = 0.2202
p-value > 0.05 Hence, we Accept H_0.
i.e., The given data follows Binomial distribution

Example 5.3.2 : Given below is the on number of phone calls received by a telephone exchange in a Corporate office during 10 minutes on a working day. Fit a Poisson distribution for the following data and test for the goodness of fit, at 5% level of significance?

x	0	1	2	3	4	5
f	142	156	69	27	5	1

R- Code to χ^2 test for Goodness of fit:

 # H_0 : The given data follows Poisson distribution.
 # H_1 : The given data does not follow Poisson distribution.

```
x = c(0, 1, 2, 3, 4, 5)
f= c(142, 156, 69, 27, 5, 1)
# Fitting of Poisson distribution
m = sum(f*x)/sum(f)
px = dpois(x,m)
ef = sum(f)*px
chisq.test(f, ef)
```

Detailed Explanation of the code arguments:

'dpois( )' – calculates the probability function of the binomial distribution.
'chisq.test( )' – Performs chi-squared contingency table tests and goodness-of-fit tests.

Output:

Pearson's Chi-squared test data: f and ef
X-squared = 24, df = 20, p-value = **0.2424**

Conclusion:

Here, p-value = 0.2424
p-value > 0.05 Hence, we Accept H_0.
i.e., The given data follows Poisson distribution

Exercise Problems for Practice

1. Samples of 15 observations were taken from a population with respect to the particular character of study with population standard deviation equal to 40.44. Test $\sigma^2 = (40.44)^2$ at 1% level of significance and give your conclusion.

   43.21, 44.52, 46.31, 38.26, 43.64, 40.34, 39.99, 28.46, 43.33, 37.64, 36.54, 41.54, 36.54, 44.14, 38.64

2. Weight of 10 students in kg is given as 38, 40, 45, 53, 47, 43, 55, 48, 52, 49. Can we conclude that the variance of the distribution of weights of students is equal to 20kgs. Test at 5% level of significance.

3. The following data gives the classification of college students with respect to their intelligences and economic conditions.

Economic conditions	Intelligence			
	Excellent	Good	Medium	Dull
High	24	36	42	53
Good	32	44	62	18
Bad	48	38	15	10
Worse	40	30	15	10

Test whether there is any association between intelligence and economic conditions of students at 1% level of significance.

4. At 5% level of significance, fit an appropriate distribution to the following data and test for its goodness of fit.

No. of calls received per minute	0	1	2	3	4	5	6	7
Frequency	4	15	42	60	89	94	52	80

Practical No. 6

Nonparametric Tests – I

This Lesson covers the following

6.1 One sample run test or test for randomness of the sample
6.2 Sign test for single sample
6.3 Sign test for two samples
6.4 Wilcoxon signed rank test for one sample
6.5 Wilcoxon signed rank test for two samples

All these test Procedure will be explained with the help of illustrations.

6.1 One Sample Run Test

Example 6.1: Test whether the given sample of 'Wins (W)' and 'Loses (L)' is random or not at 5% level of significance?

W, W, L, L, L, W, L, W, W, L, L, W, W, W, W, L, L, W, L, L

R- Code to One Sample Run Test:

 # H_0: The given sample is a random sample
 # H_1: The given sample is not a random sample (Two Tailed Test)

```
install.packages( ' DescTools ' )
library(DescTools)
x =  factor(c("W", "W", "L", "L", "L", "W", "L", "W", "W", "L", "L", "W", "W", "W",
                "W", "L", "L", "W", "L", "L"))
RunsTest(x, alternative = "two.sided")
```

Detailed Explanation of the code arguments:

'install.packages( )' – download and install packages from CRAN-like repositories or from local files.

'library( )' – is used to load and attach add-on packages.

'Factor( )' – is a tool for working with categorical variables.

'RunsTest( )' – is used to determine whether or not a dataset comes from a random process. (or) examines the randomness of a numeric sequence

Output:

data: x
runs = 10, m = 10, n = 10,
p-value = 0.8281
alternative hypothesis: true number of runs is not equal the expected number

Conclusion:

Here, p-value = 0.8281
p-value > 0.05, Hence, we Accept H_0.
i.e., we conclude that the given sample is a random sample.

6.2 Sign Test for Single Sample

The Procedure for Sign test for single sample is explained with the help of the following example.

Example 6.2 : A particular type of electric tubes gave the following observed life span in a sample of 8 tubes (in hours).

Life span of tubes : 680, 920, 1040, 780, 955, 1063, 895, 1105

Can we conclude that the population median is equal to 900 hours against the hypothesis that it is more than that? Test at 5% level of significance?

R- Code to Sign Test for single sample:

H_0: The life span of tube is 900 hours
H_1: The life span of tube is not 900 hours (Two Tailed Test)

```
install.packages("signmedian.test")
library(signmedian.test)

x = c(680, 920, 1040, 780, 955, 1063, 895, 1105)
md = 900
signmedian.test(x, md, alternative = "two.sided"))
```

Detailed Explanation of the code arguments:

'install.packages( )' – download and install packages from CRAN-like repositories or from local files.

'library( )' – is used to load and attach add-on packages.

'signmedian.test( )' – Perform sign test on one-sample data

Output:

```
Exact sign test
data:  x
#(x!=900) = 8, mu = 900, p-value = 0.7266
alternative hypothesis: the median of x is not equal to mu
92.96875 percent confidence interval:
 780 1063
Sample estimates: point estimator
                          937.5
```

Conclusion:

Here, p-value = 0.7266
p-value > 0.05, Hence, we Accept H_0.
i.e., we conclude that the life span of tube is 900 hours.

6.3 Sign Test for Two Samples

This Procedure is also explained with the help of the following example.

Example 6.3 : The following are the grades of 15 students in the internal and final exams in the subject Statistics. Test whether there is any improvement in their final exam grades, over internal exams grades. Test at 5% level of significance.

Internal:	66	78	75	90	63	58	75	83	73	84	85	93	70	85	90
Final:	73	91	78	86	69	87	75	80	76	89	81	96	76	82	97

R- Code to Sign Test for two samples:

```
# H0: There is no improvement in final exam grades.
# H1: There is an improvement in final exam grades. (Left Tailed Test)
x1 = c(66,78,75,90,63,58,75,83,73,84,85,93,70,85,90)
x2 = c(73,91,78,86,69,87,75,80,76,89,81,96,76,82,97)
y = x1-x2
y
signmedian.test(y, alternative = "less")
```

Detailed Explanation of the code arguments:

'signmedian.test( )' – Perform sign test on one-sample data

Output:

```
Exact sign test
data:  y
#(x<0) = 10, mu = 0, p-value = 0.08978
alternative hypothesis: the median of x is less than mu
94.2627 percent confidence interval:
 -7  3
Sample estimates: point estimator
                          -3
```

Conclusion:

Here, p-value = 0.08978
p-value > 0.05, Hence, we Accept H_0.
i.e., we conclude that there is no improvement in final exam grades when compared to internal exams.

6.4 Wilcoxon Signed Rank Test for Single Sample

The Procedure for Wilcoxon Sign Rank Test for single sample is explained with the help of the following example.

Example 6.4 : A random sample of 9 women buying new eye glasses tried on the following number of different frames.

12, 11, 14, 15, 10, 16, 9, 12, 8

Use Wilcoxon Sign Rank test to test whether a women is tried on more than 10 frames to buy new eye glasses. Use 5% level of significance.

R- Code to Wilcoxon Signed Rank Test for single sample:

```
# H0 : On an average a woman is tried 10 frames to buy new eye glasses.
# H1 : On an average a woman is tried more than 10 frames to buy new eye glasses.
mu = 10
x = c(12, 11, 14, 15, 10, 16, 9, 12, 8)
wilcox.test(x, mu = 10, alternative = "greater")
```

Detailed Explanation of the code arguments:

'wilcox.test( )' – Performs one- and two-sample Wilcoxon tests on vectors of data; the latter is also known as 'Mann-Whitney' test.

'alternative( )' – is a character string specifying the alternative hypothesis, must be one of "two. sided" (default), "greater" or "less". You can specify just the initial letter.

Output:

```
Wilcoxon signed rank test with continuity correction
data: x
V = 30.5, p-value = 0.04544
alternative hypothesis: true location is greater than 10
```

Conclusion:

Here, p-value = 0.04544
p-value < 0.05, Hence, we Reject H_0.
i.e., we conclude that on an average a woman is tried more than 10 frames to buy new eye glasses.

6.5 Wilcoxon Signed Rank Test for Two samples

This Procedure also for Wilcoxon Signed Rank Test for two samples is explained with the help of the following example.

Example 6.5 : Following data shows the weight of 10 men before and after treatment. Use Wilcoxon Signed Rank test to test the effectiveness of the drug in reducing the weight of 10 men.

Before Treatment:	70	80	88	70	85	100	110	90	70	90
After Treatment:	65	80	100	60	100	90	88	93	68	60

Test at 5% level of significance.

R- Code to Wilcoxon Signed Rank Test for two samples:

 # H_0 : There is no reduction in the weight of 10 men (or) The drug is not effective.
 # H_1 : There is a reduction in the weight of 10 men (or) The drug is effective.
(Right Tailed Test)

```
x = c(70, 80, 88, 70, 85, 100, 110, 90, 70, 90)
y = c(65, 80, 100, 60, 100, 90, 88, 93, 68, 60)

wilcox.test(x, y, paired = TRUE, alternative = "greater")
```

Detailed Explanation of the code arguments:

'wilcox.test( )' – Performs one- and two-sample Wilcoxon tests on vectors of data; the latter is also known as 'Mann-Whitney' test.
'alternative( )' – is a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less". You can specify just the initial letter.

Output:

```
Wilcoxon signed rank test with continuity correction
data:  x and y
V = 30, p-value = 0.2033
alternative hypothesis: true location shift is greater than 0
```

Conclusion:

Here, p-value = 0.2033
p-value > 0.05, Hence, we Accept H_0.
i.e., we conclude that the drug is not effective.

Exercise Problems for Practice

1. The following is a sequence of Males and Females standing in queue at a booking center.

   F, F, M, F, M, M, M, F, F, F, F, M, M, M, M, M, F, F, M, F, F, M, M

   Test whether the sequence is random or not? Test at 5% level of significance.

2. A random sample of 13 infants of one month old shows the following pulse rate.

   119, 120, 125, 127, 118, 117, 126, 114, 115, 123, 121, 120, 127

   Assuming that the distribution of pulse rate is continuous and symmetric. Is there any evidence to suggest that the median pulse rate is different from 120. Test at 5% level of significance.

3. 12 people were selected for treatment for certain diseases. The following data gives the grades of these people before and after the treatment is given. Test whether the condition of people improved after the treatment or not, at 1% level of significance.

Before:	1 3	5	6	8	2	3	6	4	8	3	4
After :	5 5	3	4	2	5	3	6	8	9	4	3

4. 14 people were selected for a special training and a test is conducted before giving training. Another test is conducted after completion of training. The marks are given below:

   Before: 2 3 3 3 5 2 1 5 6 4 5 1 2 6
   After: 4 4 5 5 3 2 5 3 1 4 5 6 2 1

   Use Wilcoxon's Signed rank test to assess the difference in their performance level, at 5% level of significance.

Practical No. 7

Nonparametric Tests – II

This Lesson covers the following

7.1 Median test
7.2 Wilcoxon Mann Whitney - U test
7.3 Wald - Wolfowitz's runs Test

7.1 Median Test

The Procedure for Median test is explained with the help of the following example.

Example 7.1 : In an experiment two different treatments were applied to a sample of 11 vegetable plants each, to improve their yield. The outputs of these plants are as follows.

Treatment 1:	34	30	51	49	48	46	42	45	32	49	50
Treatment 2:	37	35	47	50	43	61	51	57	44	40	42

Test the hypothesis that the treatment yield is same or not by using median test, at 5% level of significance.

R- Code to Median Test:

```
# H0 : The both the treatments yields are same
# H1 : The both the treatments yields are not same
x = c(34, 30, 51, 49, 48, 46, 42, 45, 32, 49, 50)
y = c(37, 35, 47, 50, 43, 61, 51, 57, 44, 40, 42)
install.packages("nonpar")
library(nonpar)
mediantest(x, y)
```

Detailed Explanation of the code arguments:
'install.packages( )' – download and install packages from CRAN-like repositories or from local files.
'library( )' – is used to load and attach add-on packages.
'mediantest( )' – to examine whether several samples came from populations having the same median.

Output:

Exact Median Test
H_0 : The 2 population medians are equal.
H_A : The 2 population medians are not equal.

Significance Level = 0.05
The p-value is 0.605144081924268
There is not enough evidence to conclude that the population medians are different at a significance level of 0.05.

Conclusion:

Here, p-value = 0.605144081924268
p-value > 0.05, Hence, we Accept H_0.
i.e., we conclude that both the treatments yields are same.

7.2 Wilcoxon Mann Whitney - U Test

The Procedure for Wilcoxon Mann Whitney - U test is explained with the help of the following example.

Example 7.2 : The following table gives the average life time (in hours) of two brands (A and B) of fluorescent tubes.

Brand A:	435	375	400	450	425	430	445	460	475	422	410	465
Brand B:	425	510	500	435	440	455	465	405	415	505	395	418

Use Wilcoxon Mann Whitney - U test to test for the significant difference between the life times of two brands of fluorescent tubes, at 1% level of significance.

R- Code to Wilcoxon Mann Whitney - U test:

```
# H0 : The life time of two brands of fluorescent tubes is same.
# H1 : The life time of two brands of fluorescent tubes is not same.
x = c( 435, 375, 400, 450, 425, 430, 445, 460, 475, 422, 410, 465)
y = c( 425, 510, 500, 435, 440, 455, 465, 405, 415, 505, 395, 418)

wilcox.test(x, y, alternative = "two.sided")
```

Detailed Explanation of the code arguments:

'wilcox.test( )' – Performs one- and two-sample Wilcoxon tests on vectors of data; the latter is also known as 'Mann-Whitney' test.

Output:

```
Wilcoxon rank sum test with continuity correction
data:  x and y
W = 61.5, p-value = 0.5634
alternative hypothesis: true location shift is not equal to 0
```

Conclusion:
Here, p-value = 0.5634
p-value > 0.05, Hence, we Accept H_0.
i.e., we conclude that the lifetime of two brands of tubes is same.

7.3 Wald - Wolfowitz's Runs Test

This Procedure for Wald-Wolgowitz's run test is explained with the help of the following example.

Example 7.3: To compare the variability of 2 brands of tires, the following mileage were observed for 8 tires of each brand.

Brand A:	32.1	20.6	17.8	28.4	19.6	21.4	19.9	30.1
Brand B:	19.8	27.6	30.8	27.6	34.1	18.7	16.9	17.9

Test for the significant difference in the average mileage of two brands of tires, at 1% level of significance.

R- Code to Wald - Wolfowitz's runs Test:

 # H_0 : There is no significant difference in the mileage of two brands of tires.
 # H_1 : There is a significant difference in the mileage of two brands of tires.
(Two Tailed Test)

```
install.packages("DescTools")
library(DescTools)
x = c(32.1, 20.6, 17.8, 28.4, 19.6, 21.4, 19.9, 30.1)
y = c(19.8, 27.6, 30.8, 27.6, 34.1, 18.7, 16.9, 17.9)

RunsTest(x, y, alternative = "two.sided")
```

Detailed Explanation of the code arguments:

'install.packages( )' – download and install packages from CRAN-like repositories or from local files.

'library( )' – is used to load and attach add-on packages.

'RunsTest( )' – Performs a test whether the elements of x are serially independent

Output:

```
Wald-Wolfowitz Runs Test
data:  x and y
runs = 11, m = 8, n = 8, p-value = 0.4289
alternative hypothesis: true number of runs is not equal the expected number
```

Conclusion:
Here, p-value = 0.4289
p-value > 0.05, Hence, we Accept H_0.
i.e., we conclude that there is no significance difference in the mileage of two brands of tires.

Exercise Problems for Practice

1. Use median test to test whether the two samples are drawn from population with equal median or not. Test at 5% level of significance.

X :	163	165	160	189	161	171	158	151	236	149	254	196	90	151
Y :	139	172	165	148	166	172	163	163	187	173	55	273	161	173

2. Given below are the weights of two groups of patients 8 in each group. Use Mann Whitney U-Test for the following data to test whether the two samples are drawn from populations with equal means or not. Test at 1% level of significance.

Group1 :	42	74	84	82	45	73	25	12
Group 2:	35	85	85	92	87	46	36	20

3. Given below are the heights of 10 soldiers in two different regiments.

Regiment 1:	163	165	160	189	161	171	158	151	169	162
Regiment 2:	163	139	172	165	148	166	172	163	187	223

Use the Wald Wolfwitz run test to test whether the two samples are drawn from populations with different variance or not. Test at 5% level of significance.

Practical No. 8

Analysis of Variance

This Lesson covers the following

8.1 Analysis of Variance for CRD data using MS - Excel.
8.2 Analysis of Variance for CRD data using R
8.3 Analysis of Variance for RBD data using MS - Excel.
8.4 Analysis of Variance for RBD data using R.

8.1 Analysis of Variance for CRD data using MS – Excel

Analysis of Variance Procedure for CRD data is explained with the help of the following example.

8.2 Analysis of Variance for CRD data using MS - Excel

Example 8.1: Three processes A, B and C are tested to see whether their outputs are equivalent. The following observations of output are made.

A :	10	12	13	11	10	14	15	13
B :	9	11	10	12	13			
C :	11	10	15	14	12	13		

Carryout the analysis of variance and state your conclusions.

Solution through MS-Excel : To start with enter the data in to the excel sheet as in the following Figure 8.1.

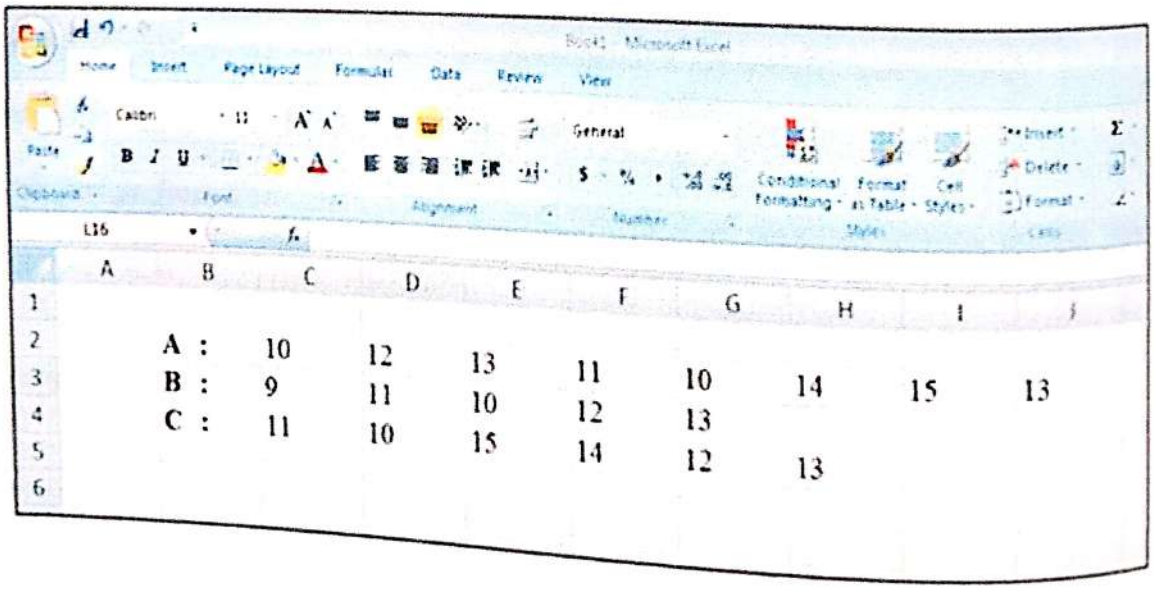


Figure 8.1

Select **"Data"** in the **Menu bar** and select **"Data Analysis"** in the drop menu as shown in Figure 8.2 below. Then select **"ANOVA : Single Factor"** in the window of Data Analysis – See Figure 8.3 below. Say **"OK"**. A dialogue box appears as shown in Figure 8.4.

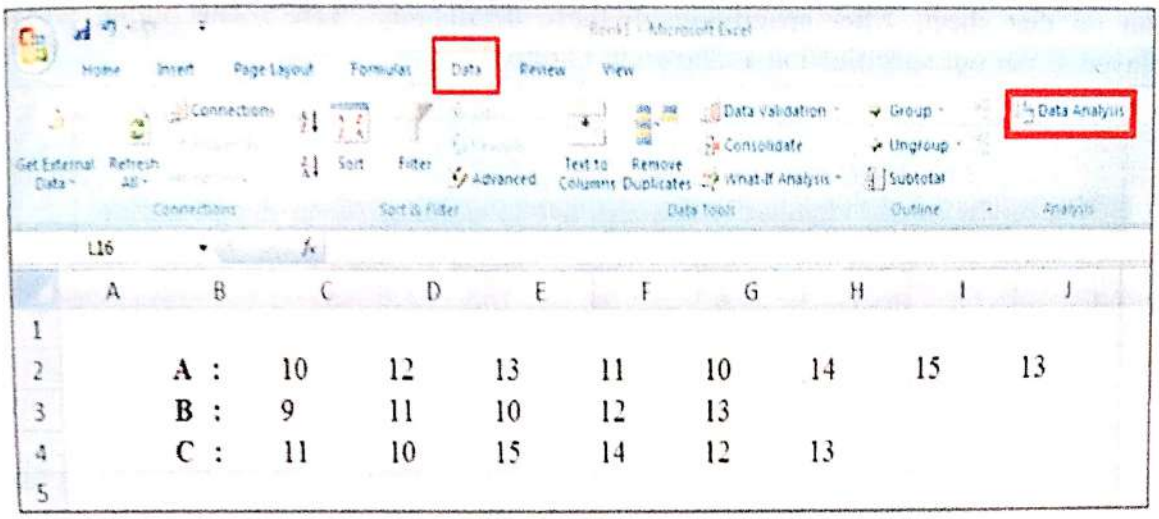


Figure 8.2

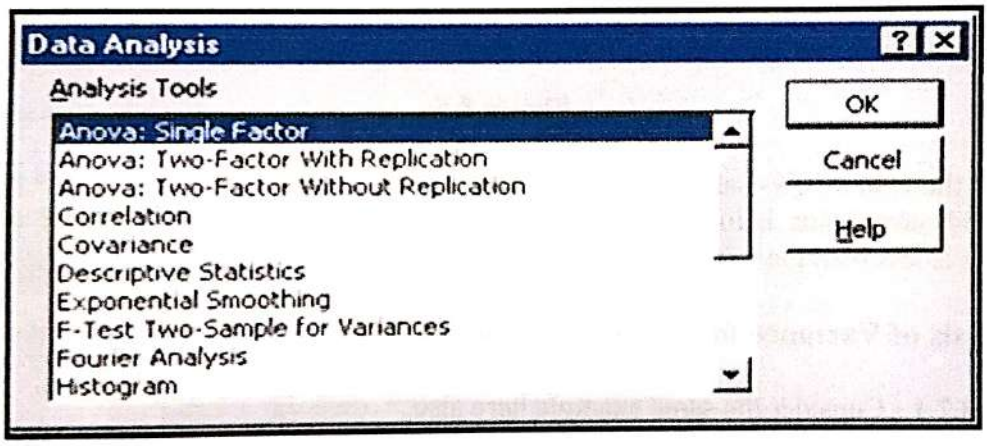


Figure 8.3

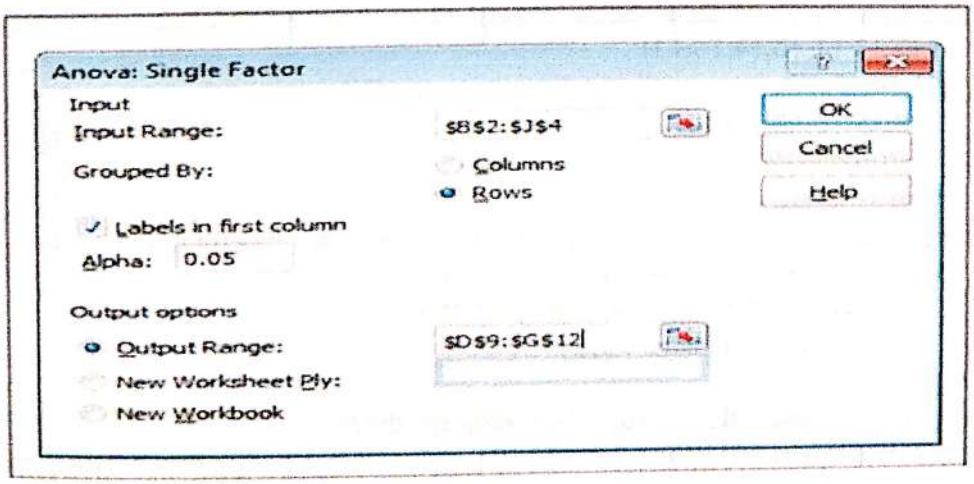


Figure 8.4

Specify the range of input data as shown in Figure 8.4. Also specify that the data is grouped by rows. The default Level of Significance is 0.05. If it is different specify that also. Here the output range also can be specified to get the output in the same sheet as specified in Figure 8.4. If no output range is specified by default it creates a new worksheet and gives the output on that sheet. After specifying all these details say **"OK"**. The output will be displayed as per our specification as shown in Figure 8.5.

Anova Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Row 1	8	98	12.25	3.3571429
Row 2	5	55	11	2.5
Row 3	6	75	12.5	3.5

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	7	2	3.5	1.0980392	0.3573859	3.633716
Within Groups	51	16	3.1875			
Total	58	18				

Figure 8.5

Based on the values of F-Calculated (in the above table 1.098) and F-Critical (as per above table 3.634) conclusion is to drawn. Here in this problem H_o is accepted as the calculated value of F is less than the critical value of F, at 5% level of significance.

8.3 Analysis of Variance for CRD Data Using R

Example 8.2.1 : Consider the same example here also.

R- Code to Analysis of Variance for CRD:

```
# In this problems no. of observations are not equal in each treatment
A = c(10,12,13,11,10,14,15,13)
B = c(9,11,10,12,13)
C = c(11,10,15,14,12,13)
max_ln = max(c(length(A), length(B), length(C) ))

combined_data = data.frame(col1 = c(A,rep(NA, max_ln - length(A))),
        col2 = c(B,rep(NA, max_ln - length(B))),
        col3 = c(C,rep(NA, max_ln - length(C))))
stacked_data = stack(combined_data)
stacked_data
oneway_anova = aov(values~ind, data = stacked_data)
summary(oneway_anova)
```

Detailed explanation of the code arguments:

'A, B, C' are vectors of the given observations in the data

'max( )' is a function it gives the maximum value among all the given values

'data.frame( )' is a function combines the different vectors into single data set

'col2 = c(B,rep(NA, max_ln - length(B))' in this "col2"compareslength of second column with the highest length of all columns of the data and if second column length is less than themaximum length, then puts NA in some cells such that second column length becomes equal to the maximum length.

'stack(combined_data)' is a stacking vector concatenates the multiple vectors into single vector along with a factor indicating where each observation originated

'aov( )' is a function fit an analysis of variance model

'values ~ind': In this "values" are the observations of the data set and "ind" represents the treatment group.

'data' is a data frame in which the variables specified in the formula will be found

'summary ( )' a function shows the ANOVA results.

Output:

```
> summary(oneway_anova)
            Df Sum Sq Mean Sq F value Pr(>F)
ind          2      7   3.500   1.098  0.357
Residuals   16     51   3.188
5 observations deleted due to missingness
```

Result: Here, p – value is 0.357 > 0. 05 therefore in this problem H_0 is accepted. Hence, it is concluded that all the three processes are homogeneous with respect to their outputs.

Example 8.2.2 : For testing the variety effect in an experiment the yield of Barley (in gms) are given in the following table. Perform analysis of variance.

Variety	Observations					
V1	320	420	353	331	358	400
V2	372	455	375	328	383	308
V3	350	417	325	378	400	275
V4	340	360	358	370	395	375
V5	398	358	334	340	320	430

R – Code to Analysis of Variance for CRD:

```
# In this problems no. of observations are equal in each treatment
v1 = c(320,420,353,331,358,400)
v2 = c(372,455,375,328,383,308)
v3 = c(350,417,325,378,400,275)
v4 = c(340,360,358,370,395,375)
v5 = c(398,358,334,340,320,430)

combined_data = data.frame(v1,v2,v3,v4,v5)
combined_data
stacked_data = stack(combined_data)
stacked_data
oneway_anova = aov(values~ind, data = stacked_data)
summary(oneway_anova)
```

Detailed explanation of the code arguments:

'v1,v2,v3,v4,v5' - are vectors of the given observations in the data

'data.frame( )' - is a function combines the different vectors into single data set

'stack(combined_data)' - is a stacking vector concatenates the multiple vectors into single vector along with a factor indicating where each observation originated

'aov( )' - is a function fit an analysis of variance model

'values ~ ind' - In this "values" are the observations of the data set and "ind" represents the treatment group.

'data' - is a data frame in which the variables specified in the formula will be found

'summary ( )' - a function shows the ANOVA results

Output:

```
> summary(oneway_anova)
            Df Sum Sq Mean Sq F value Pr(>F)
ind          4    516   129.1   0.072   0.99
Residuals   25  45000  1800.0
```

Conclusion: Here, p – value is 0.99 > 0. 05 therefore in this problem H_0 is accepted. Hence, we conclude that all the varieties are homogeneous with respect to their average yields.

<u>Note</u>: As far as analysis is concerned, ANOVA One-way classified data procedure is used to analyze Completely Randomized Design (CRD) data.

8.4 Analysis of Variance for RBD data using MS – Excel

Example 8.3: Three varieties of Coal were analyzed by Four chemists and the ash-content in the varieties was found to be as under :

	Chemists			
Varieties	I	II	III	IV
A	18	15	15	17
B	17	16	14	14
C	13	16	15	14

Do the varieties differ in their ash-content? Do the chemists differ significantly in their analysis?

Solution through MS-Excel: To start with enter the data in to the excel sheet as in the following Figure 8.6.

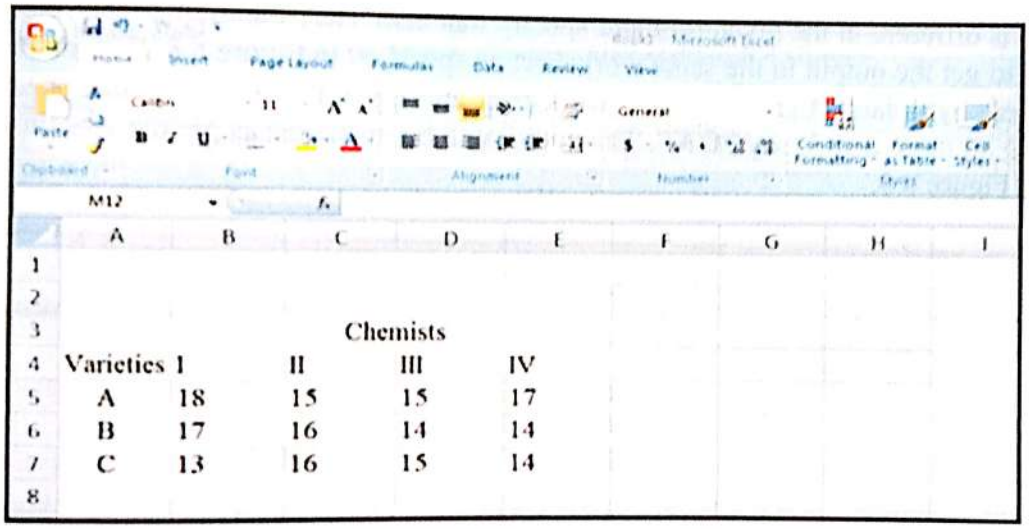


Figure 8.6

Select **"Data"** in the **Menu bar** and select **"Data Analysis"** in the drop menu as shown in Figure 8.2 above. Then select **"ANOVA: Two Factor without Replication"** in the window of Data Analysis as in Figure 8.7 below. Say **"OK"**. A dialogue box appears as shown in

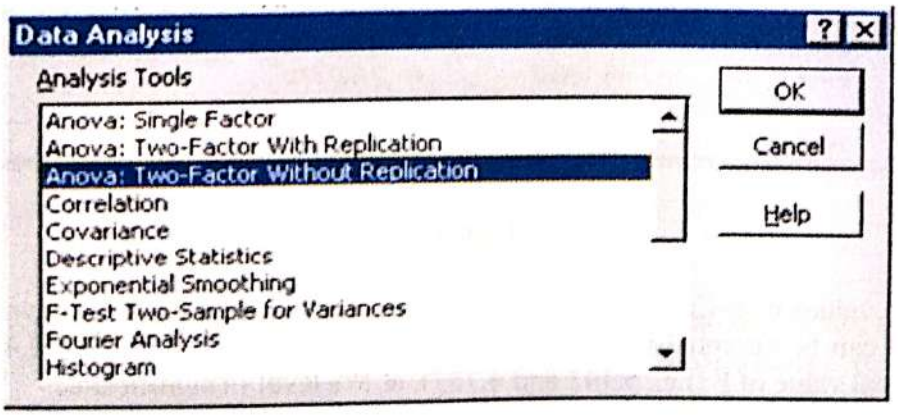


Figure 8.7

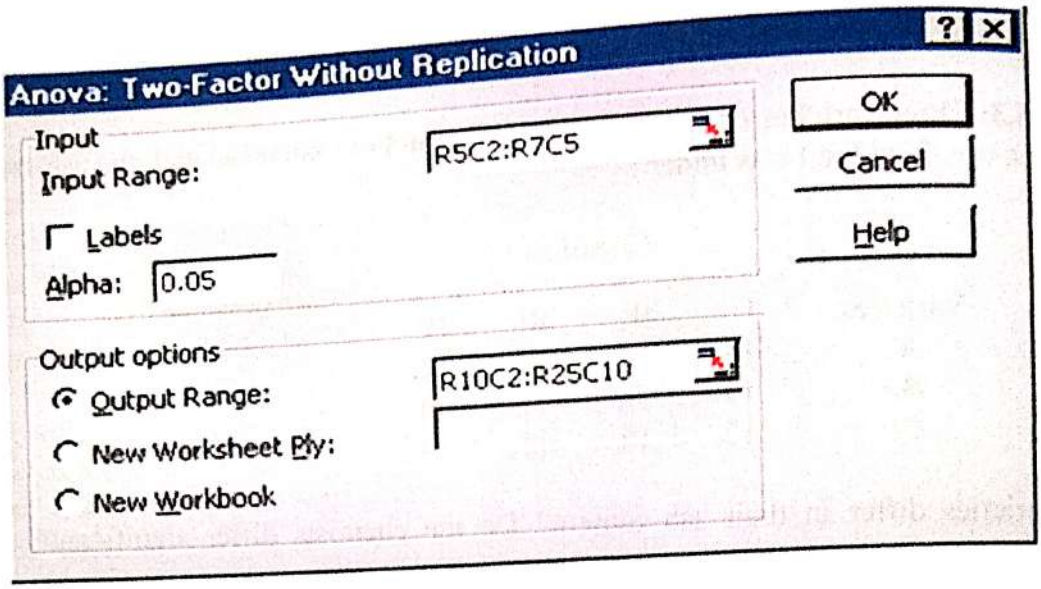


Figure 8.8:

Specify the range of input data as shown in Figure 8.8. The default Level of Significance is 0.05. If it is different in the given problem specify that also. Here the output range also can be specified to get the output in the same worksheet as specified in Figure 8.8. If no output range is specified, by default it creates a new worksheet and gives the output on that sheet. After specifying all these details say **"OK"**. The output will be displayed as per our specification as shown in Figure 8.9.

Anova: Two-Factor Without Replication

SUMMARY	Count	Sum	Average	Variance
Row 1	4	65	16.25	2.25
Row 2	4	61	15.25	2.25
Row 3	4	58	14.5	1.666667
Column 1	3	48	16	7
Column 2	3	47	15.66667	0.333333
Column 3	3	44	14.66667	0.333333
Column 4	3	45	15	3

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	6.166667	2	3.083333	1.21978	0.359331	5.143249
Columns	3.333333	3	1.111111	0.43956	0.733083	4.757055
Error	15.16667	6	2.527778			
Total	24.66667	11				

Figure 8.9

Based on the values of F-Calculated and F-Critical conclusions are to be drawn. Here in this problem H_os can be accepted as the calculated values of F (i.e., 3.083 and 0.4396) are less than the critical value of F (i.e., 5.143 and 4.757), at 5% level of significance.

8.5 Analysis of Variance for RBD Data Using R

Example 8.4 : Consider the same example here also.

R – Code to Analysis of Variance for RBD:

```
yield = c(18,15,15,17,17,16,14,14,13,16,15,14)
variety = c(rep('A',4), rep('B',4), rep('C',4))
chemist = c('I','II','III','IV','I','II','III','IV','I','II','III','IV')
combined_data = data.frame(yield,chemist,variety)
tw = aov(yield~variety+chemist, data = combined_data )
summary(tw)
```

Detailed explanation of the code arguments:

'yield' - is a vector of the given observations in the data and the values are taken in row wise

'variety' - is a vector represents the repetition of each row (here, 'A' variety repeated 4 times, 'B' variety repeated 4 times and 'C' variety also repeated 4 times)

'chemist' - is a vector represents the repetition of each column (here, first observation in the yield belongs to chemist 'I', second observation in the yield belongs to chemist 'II', third observation in the yield belongs to chemist 'III', fourth observation in the yield belongs to chemist 'IV')

'data.frame( )' - is a function combines the different vectors into single data set

'aov( )' - is a function fit an analysis of variance model

'yield~variety+chemist' - is used because here we have two factors in the data, one is variety and other is chemist. The yield is a combination of variety and chemist.

'data' - is a data frame in which the variables specified in the formula will be found

'summary ( )' - a function shows the ANOVA results

Output:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
variety	2	6.167	3.083	1.22	0.359
chemist	3	3.333	1.111	0.44	0.733
Residuals	6	15.167	2.528		

Conclusion : In this problem H_0 s can be accepted as the probability values (p-values) are (i.e., 0.359 and 0.733) are more than the 0.05, level of significance. Hence, we conclude that the varieties are homogeneous with respect to their ash content and all chemists are homogeneous in their analysis.

<u>Note</u>: As far as analysis is concerned, ANOVA Two-way classified data procedure is used to analyze Randomized Block Design (RBD) data.

Exercise Problems for Practice

1. Three different techniques namely medication, exercises and special diet are randomly assigned to (individuals diagnosed with high blood pressure) lower the blood pressure. After four weeks the reduction in each person's blood pressure is recorded. Test at 5% level, whether there is significant difference in mean reduction of blood pressure among the three techniques, using analysis variance one way classification technique.

Medication	10	12	9	15	13
Exercise	6	8	3	0	2
Diet	5	9	12	8	4

2. Three composition instructors recorded the number of spelling errors which their students made in a research paper. At 1% level of significance, test whether there is significant difference in the average number of errors in the three classes of students, using analysis variance one way classification technique.

Instructor 1	2	3	5	0	8		
Instructor 2	4	6	8	4	9	0	2
Instructor 3	5	2	3	2	3	3	

3. A reputed marketing agency in India has three different training programs for its salesmen. The three programs are Named as – A, B and C. To assess the success of the programs, 4 salesmen from each of the programs were sent to the field. Their performances in terms of sales are given in the following table.

Salesmen	Methods		
	A	B	C
1	4	6	2
2	6	10	6
3	5	7	4
4	7	5	4

Perform analysis of variance two way classifications at 5% level of significance.

4. Some illness in a city caused by food poisoning. It has happened in some restaurants due to lack of cleanliness in the restaurants. In order to investigate this issue, the director of

Food Inspection, has sent five Inspectors to inspect three restaurants and asked them to grade the three restaurants with respect to their cleanliness. The results are shown below.

Inspectors	Restaurants		
	I	II	III
1	71	55	84
2	65	57	86
3	70	65	77
4	72	69	70
5	76	64	85

Carry out analysis of variance two-way classification, at 5% level of significance, to test significant difference among Inspectors and cleanliness of restaurants.

5. There are six varieties of seeds. They are tested in three types of soils for their yields and the observations are as given below.

	Seeds					
Soil type	**A**	**B**	**C**	**D**	**E**	**F**
1	1200	1000	980	900	750	800
2	1000	1100	700	800	500	700
3	890	650	1100	900	400	350

Analyse the data and check the significance of soil and seeds with respect to their yields and state the conclusion, at 5% level of significance.

6. Consider the results of an experiment involving 6 treatments in 4 randomized blocks are given in the following table. Perform analysis of variance two-way classification, at 5% level of significance to the following data and write your conclusion.

		BLOCKS		
TREATMENTS	A	B	C	D
1	24.7	27.3	38.5	28.5
2	20.6	28.8	39.5	31
3	27.7	22.7	36.8	34.9
4	16.2	15	19.6	14.1
5	16.2	17	15.4	17.7
6	24.9	22.5	26.3	22.6

Practical No. 9

Construction of Life Tables

This Lesson covers the following

9.1 Construction of life tables using MS-Excel

The Procedure for Construction of life tables is explained with the help of the following example.

Example 1.1: Fill up the following life table using Excel.

X:	10	11	12	13	14	15	16	17	18	19	20
l_x :	79340	79110	78900	78705	78515	78300	78090	77875	77650	77420	77180

MS-Excel Procedure:

As we know a complete life table will have eight columns in total.

They are as given below.

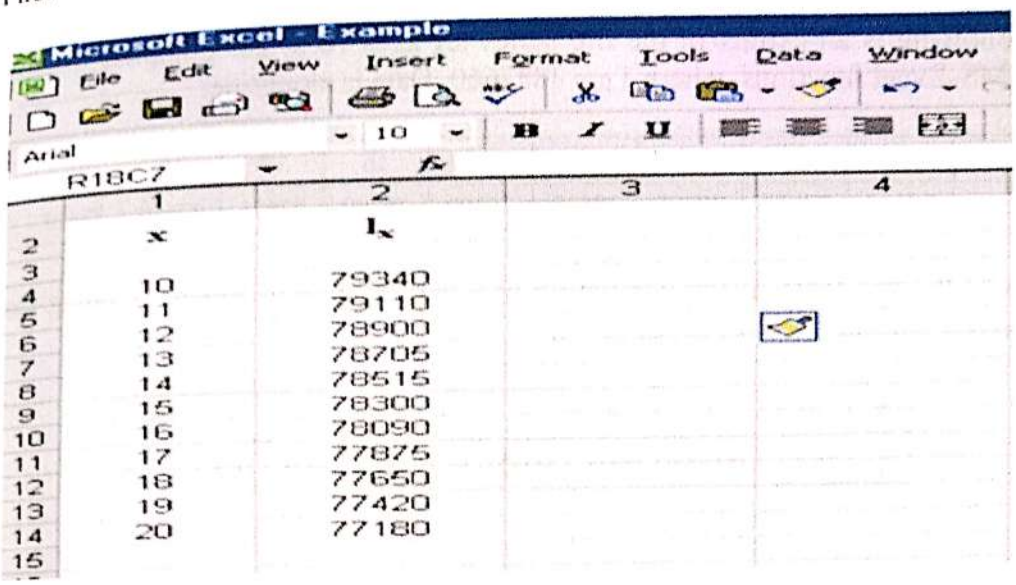
x	l_x	d_x	q_x	p_x	L_x	T_x	e_x^o

Their computational formulae are as given below.

1. x and l_x values will be given.
2. $d_x = l_x - l_{x+1}$
3. $q_x = 1 - p_x$
4. $p_x = l_{x+1} / l_x$
5. $L_x = 0.5(l_x + l_{x+1})$
6. $T_{x+1} = T_x - L_x$
7. $e_x^o = T_x / l_x$

While doing in Excel, instead of the quantities their cell addresses will be used in the formulae to fill up the life table.

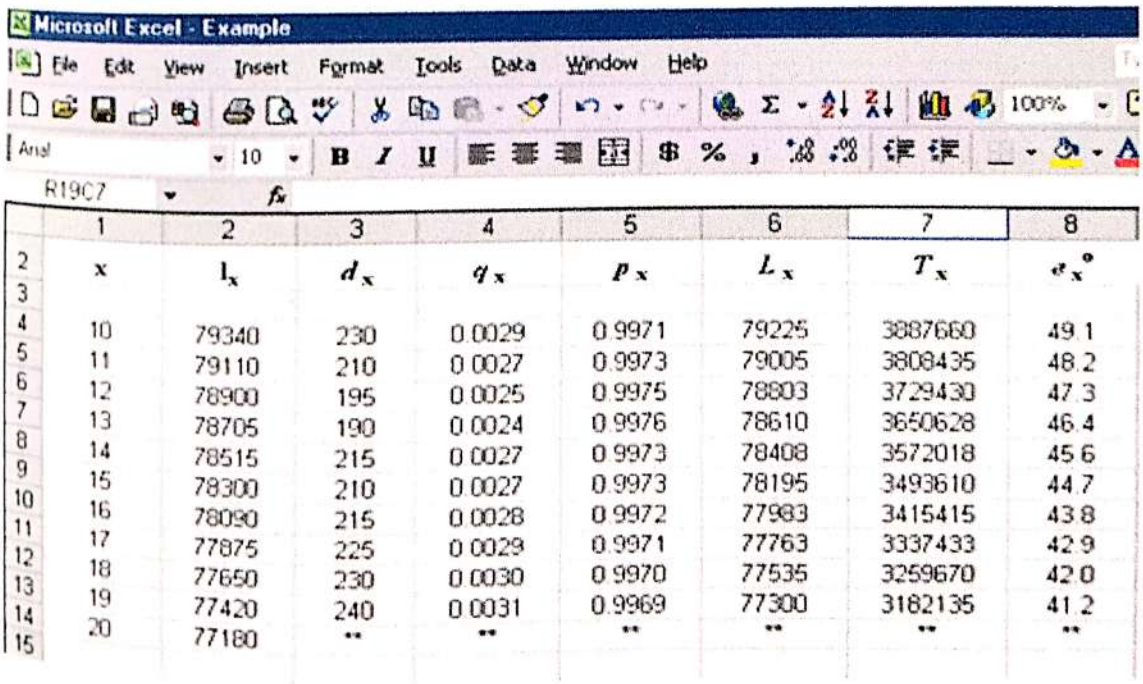
First data is entered into the worksheet as shown in the following Figure 9.1.



	1	2	3	4
2	x	l_x		
3	10	79340		
4	11	79110		
5	12	78900		
6	13	78705		
7	14	78515		
8	15	78300		
9	16	78090		
10	17	77875		
11	18	77650		
12	19	77420		
13	20	77180		
14				
15				

Figure 9.1

Then type the formulae in cells **R4C3, R4C4, R4C5** and **R4C6** as **"=R4C2-R5C2", "=1–R4C5", "=R5C2/R4C2", "=0.5*(R4C2+R5C2)"** respectively. Copy the same formulae to the remaining cells of the same columns. In cell **R4C7** first value of T_x will be given in the problem. So type the formula **"=R4C7+R4C6"** in cell **R5C7** and copy it to the remaining cells of column 7 for rest of the values of T_x. To compute the value of e_x^o type the formula **"=R4C7/R4C2"**. Copy the same to the remaining cells to get rest of the values of e_x^o. As we know the last column entries cannot be computed in any life table. All these calculations are demonstrated in the following Figure 10.2



	1	2	3	4	5	6	7	8
2	x	l_x	d_x	q_x	p_x	L_x	T_x	e_x^o
3								
4	10	79340	230	0.0029	0.9971	79225	3887660	49.1
5	11	79110	210	0.0027	0.9973	79005	3808435	48.2
6	12	78900	195	0.0025	0.9975	78803	3729430	47.3
7	13	78705	190	0.0024	0.9976	78610	3650628	46.4
8	14	78515	215	0.0027	0.9973	78408	3572018	45.6
9	15	78300	210	0.0027	0.9973	78195	3493610	44.7
10	16	78090	215	0.0028	0.9972	77983	3415415	43.8
11	17	77875	225	0.0029	0.9971	77763	3337433	42.9
12	18	77650	230	0.0030	0.9970	77535	3259670	42.0
13	19	77420	240	0.0031	0.9969	77300	3182135	41.2
14	20	77180	**	**	**	**	**	**
15								

Figure 10.2

Exercise Problems for Practice

1. The following is an extract of the life tables for ages 10 to 20. Complete the Life table using MS-Excel functions, where T_{10}= 3887660. Data is alongside.

x	10	11	12	13	14	15	16	17	18	19	20
l_x	79340	79110	78900	78705	78515	78300	78090	77875	77650	77420	77180

2. Compute the life table of the population of a certain type of insect, x being the age in days and l_x = 1000, Tx = 5134 for x = 0.

X	0	1	2	3	4	5	6	7	8
q_x	0.12	0.005	0.01	0.05	0.1	0.5	0.8	0.9	0.95

****