PRACTICAL-II

Linear Models and Applied Regression Analysis-Practical

&

THEORY OF LINEAR ESTIMATION AND ANALYSIS OF VARIANCE and ANOVA-Practical

SEMESTER-II, PAPER - VI

(206ST24)

LESSON WRITER

Dr. S. Bhanu Prakash
Assistant Professor
Freshman Engineering Department
Godavari Global University
Rajamahendravaram - 533296.

EDITOR

Prof. G. V. S. R. Anjaneyulu
Professor of Statistics (Retd.)
Acharya Nagarjuna University

ACADEMIC ADVISOR

Prof. G. V. S. R. Anjaneyulu
Professor of Statistics (Retd.)
Acharya Nagarjuna University

DIRECTOR, I/c.

Prof. V. Venkateswarlu
M.A., M.P.S., M.S.W., M.Phil., Ph.D.

Centre for Distance Education
Acharya Nagarjuna University
Nagarjuna Nagar 522 510

Ph: 0863-2346222, 2346208
0863- 2346259 (Study Material)
Website www.anucde.info
E-mail: anucdedirector@gmail.com

M.Sc., STATISTICS: Linear Models and Applied Regression Analysis-Practical
&
Theory Of Linear Estimation And Analysis of Variance And Anova-Practical
PRACTICAL-I (205ST24)

First Edition : 2025

No. of Copies :

© Acharya Nagarjuna University

This book is exclusively prepared for the use of students of M.Sc., STATISTICS Centre for Distance Education, Acharya Nagarjuna University and this book is meant for limited circulation only.

Published by:

Prof. V. VENKATESWARLU
Director, I/c
Centre for Distance Education,
Acharya Nagarjuna University

Printed at:

FOREWORD

Since its establishment in 1976, Acharya Nagarjuna University has been forging ahead in the path of progress and dynamism, offering a variety of courses and research contributions. I am extremely happy that by gaining 'A+' grade from the NAAC in the year 2024, Acharya Nagarjuna University is offering educational opportunities at the UG, PG levels apart from research degrees to students from over 221 affiliated colleges spread over the two districts of Guntur and Prakasam.

The University has also started the Centre for Distance Education in 2003-04 with the aim of taking higher education to the door step of all the sectors of the society. The centre will be a great help to those who cannot join in colleges, those who cannot afford the exorbitant fees as regular students, and even to housewives desirous of pursuing higher studies. Acharya Nagarjuna University has started offering B.Sc., B.A., B.B.A., and B.Com courses at the Degree level and M.A., M.Com., M.Sc., M.B.A., and L.L.M., courses at the PG level from the academic year 2003-2004 onwards.

To facilitate easier understanding by students studying through the distance mode, these self-instruction materials have been prepared by eminent and experienced teachers. The lessons have been drafted with great care and expertise in the stipulated time by these teachers. Constructive ideas and scholarly suggestions are welcome from students and teachers involved respectively. Such ideas will be incorporated for the greater efficacy of this distance mode of education. For clarification of doubts and feedback, weekly classes and contact classes will be arranged at the UG and PG levels respectively.

It is my aim that students getting higher education through the Centre for Distance Education should improve their qualification, have better employment opportunities and in turn be part of country's progress. It is my fond desire that in the years to come, the Centre for Distance Education will go from strength to strength in the form of new courses and by catering to larger number of people. My congratulations to all the Directors, Academic Coordinators, Editors and Lesson-writers of the Centre who have helped in these endeavors.

Prof. K. Gangadhara Rao
M.Tech., Ph.D.,
Vice-Chancellor I/c
Acharya Nagarjuna University.

CONTENTS

LINEAR MODELS AND APPLIED REGRESSION ANALYSIS-PRACTICAL

THEORY OF LINEAR ESTIMATION AND ANALYSIS OF VARIANCE AND ANOVA-PRACTICAL

Linear Models and Applied Regression Analysis-Practical

LAB Exercise 1:

Fitting a Straight Line by Least Squares (Normal Equations)

Problem: Fit the regression of Y on X for the data:

X	1	2	3	4	5
Y	2	3	5	7	8

Model $Y = \beta_0 + \beta_1 X + \varepsilon.$

Aim: To estimate β_0 and β_1 using the least-squares normal equations.

Procedure:

First, compute $\sum X, \sum Y, \sum X^2, \sum XY, n.$

The normal equations are

$\sum Y = n\hat{\beta}_0 + \hat{\beta}_1 \sum X; \qquad \sum XY = \hat{\beta}_0 \sum X + \hat{\beta}_1 \sum X^2.$

(1)

Slove these two normal equations for $\hat{\beta}_0$ and $\hat{\beta}_1$.

State the fitted line $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X.$

Calculation:

Here $n = 5$

$$\sum X = 1+2+3+4+5 = 15$$

$$\sum Y = 2+3+5+7+8 = 25$$

$$\sum X^2 = 1+4+9+16+25 = 55$$

$$\sum XY = 2+6+15+28+40 = 91$$

Substitute the above in the normal equations (1), then we get

$$25 = 5\hat{\beta}_0 + 15\hat{\beta}_1$$

(2)

$$91 = 15\hat{\beta}_0 + 55\hat{\beta}_1$$

(3)

By solving equations (2) and (3). we get,

$$\hat{\beta}_0 = 0.2 \text{ and } \hat{\beta}_1 = 1.6.$$

Therefore, fitted line is $\hat{Y} = 0.2 + 1.6X.$

Inference:

There is a strong positive linear relation; each 1-unit increase in *X* increases the fitted *Y* by 1.6 units.

LAB EXERCISE 2:

Error Variance and Variances/Covariances of LS Estimators

Problem: For the independent dataset below, estimate the error variance and the variance-covariance of $\hat{\beta}_0$ and $\hat{\beta}_1$.

X	0 1 2 3 4
Y	1 2 2 4 3

Model $Y = \beta_0 + \beta_1 X + \varepsilon$ with $E(\varepsilon) = 0, Var(\varepsilon) = \sigma^2$.

Aim: To estimate the error variance and the variance-covariance of $\hat{\beta}_0$ and $\hat{\beta}_1$.

Procedure:

First compute $\bar{X}, \bar{Y}, S_{xx} = \Sigma(X_i - \bar{X})^2, S_{xy} = \Sigma(X_i - \bar{X})(Y_i - \bar{Y})$.

Estimate coefficients:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

Obtain fitted values $\hat{Y}_i$, residuals $e_i = Y_i - \hat{Y}_i$, and $SSE = \sum e_i^2$.

Compute $\hat{\sigma}^2 = SSE/(n-2)$.

$$Var(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{xx}}$$

$$Var(\hat{\beta}_0) = \hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\Sigma(Xi - \bar{X})^2} \right)$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\bar{X}\hat{\sigma}^2}{\Sigma(Xi - \bar{X})^2}$$

Use the formulas above to get the variances and covariance.

Calculation:

Here $n = 5$,

X	Y	X^2	XY
0	1	0	1
1	2	1	4
2	2	4	4
3	4	9	16
4	3	16	9
$\sum X = 10$	$\sum Y = 12$	$\sum X_i^2 = 30$	$\sum XY = 34$

$$S_{xx} = \sum(X_i - \bar{X})^2$$
$$= \sum X^2 - (\sum X)^2 / n$$
$$= 30 - 100/5$$
$$= 10$$
$$S_{xy} = \sum(X_i - \bar{X})(Y_i - \bar{Y}).$$

$$S_{xy} = \sum(X_i - \bar{X})(Y_i - \bar{Y})$$
$$= \sum XY - (\sum X)(\sum Y)/n$$
$$= 30 - 120/5$$
$$= 6$$

Therefore,

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{6}{10} = 0.6$$
$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 2.4 - 0.6 \times 2 = 1.2$$

Find values and residuals

X	Y	$\hat{Y} = 1.2 + 0.6X$	$e = Y - \hat{Y}$	e^2
0	1	1.2	-0.2	0.04
1	2	1.8	0.2	0.04
2	2	2.4	-0.4	0.16
3	4	3.0	1.0	1.00
4	3	3.6	-0.6	0.36

SSE = 0.04+0.04+0.16+1.00+0.36 = 1.60.

$$\hat{\sigma}^2 = SSE/(n-2) = 1.60/3 = 0.5333.$$

Variance and covariance

$$Var(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{xx}} = \frac{0.5333}{10} = 0.05333, \quad SE(\hat{\beta}_1) = \sqrt{0.05333} = 0.231.$$

$$Var(\hat{\beta}_0) = 0.5333\left(\frac{1}{5} + \frac{2^2}{10^2}\right) = 0.3200, \quad SE(\hat{\beta}_0) = \sqrt{0.3200} = 0.566.$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\bar{X}\hat{\sigma}^2}{\sum(Xi - \bar{X})^2} = -\frac{2 \times 0.5333}{10} = -0.1067.$$

Inference:

Estimated line: $\hat{Y} = 1.2 + 0.6X$.

Error Variance estimate: $\hat{\sigma}^2 = 0.5333$.

The variances and covariances of the estimates are

$$Var(\hat{\beta}_1) = 0.05333,$$

$$Var(\hat{\beta}_0) = 0.3200,$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = -0.1067.$$

LAB EXERCISE 3:

GLS (Regression with AR(1) Correlated Errors, Paris – Winsten Style)

Problem: Fit a simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ using the data:

X	1	2	3	4
Y	1.2	2.4	3.2	4.0

With AR(1) errors: $Cov(\varepsilon_i, \varepsilon_j) = \sigma^2 \rho^{|i-j|}$. Take $\rho = 0.5$. Obtain GLS estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$. Also compute *SSE*, s^2, $Var(\hat{\beta})$, $SE(\hat{\beta})$, t_{slope}.

Aim: To obtain GLS estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$. Also compute *SSE*, s^2, $Var(\hat{\beta})$, $SE(\hat{\beta})$, t_{slope}.

Procedure:

1. Compute $\sqrt{1-\rho^2}$ and apply the whitening (exact AR(1) transform:

   $$y_1^* = \sqrt{1-\rho^2}\, y_1, X_1^* = \sqrt{1-\rho^2}\,(1, x_1).$$

   For $t \geq 2; y_t^* = y_t - \rho y_{t-1}, X_t^* = (1, x_t) - \rho(1, x_{t-1}).$

2. From X^* and y^*. Compute $X^{*T} X^*$ and $X^{*T} y^*$.
3. GLS estimate: $\hat{\beta} = (X^{*T} X^*)^{-1} X^{*T} y^*$.
4. Fitted $\hat{y}^* = X^* \hat{\beta}$ and residuals $e^* = y^* - \hat{y}^*$.
5. Compute $SSE = \sum (e_i^*)^2$ and $s^2 = \dfrac{SSE}{n-p}$ where p=2.
6. Compute $\text{Var}(\hat{\beta}\ ) = s^2 (X^{*T} X^*)^{-1} . SE(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}.$
7. $t_{slope} = \hat{\beta}_1 / SE(\hat{\beta}_1), df = n - p.$.

Calculations:

Given $\rho = 0.5$

1. $\sqrt{1-\rho^2} = \sqrt{1-0.5^2} = 0.8660254.$

2. Transformed matrix:

$$y^* = \begin{bmatrix} 0.8660254 \times 1.2 \\ 2.4 - 0.5 \times 1.2 \\ 3.2 - 0.5 \times 2.4 \\ 4.0 - 0.5 \times 3.2 \end{bmatrix} = \begin{bmatrix} 1.03923048 \\ 1.8 \\ 2.0 \\ 2.4 \end{bmatrix} \text{ (from step:1 in procedure)}$$

$$X^* = \begin{bmatrix} 0.8660254 & 0.8660254 \\ 0.5 & 1.5 \\ 0.5 & 2.0 \\ 0.5 & 2.5 \end{bmatrix}$$

3. Compute $X^{*T}X^*$ and $X^{*T}y^*$.

$$X^{*T}X^* = \begin{bmatrix} 0.8660254 & 0.8660254 \\ 0.5 & 1.5 \\ 0.5 & 2.0 \\ 0.5 & 2.5 \end{bmatrix}^T \cdot \begin{bmatrix} 0.8660254 & 0.8660254 \\ 0.5 & 1.5 \\ 0.5 & 2.0 \\ 0.5 & 2.5 \end{bmatrix}$$
$$= \begin{bmatrix} 1.5 & 3.75 \\ 3.75 & 13.25 \end{bmatrix}$$

$$X^{*T}y^* = \begin{bmatrix} 0.8660254 & 0.8660254 \\ 0.5 & 1.5 \\ 0.5 & 2.0 \\ 0.5 & 2.5 \end{bmatrix}^T \cdot \begin{bmatrix} 1.03923048 \\ 1.8 \\ 2.0 \\ 2.4 \end{bmatrix}$$
$$= \begin{bmatrix} 4.00 \\ 13.60 \end{bmatrix}$$

4. GLS estimates:

$$(X^{*T}X^*)^{-1} = \begin{bmatrix} 1.5 & 3.75 \\ 3.75 & 13.25 \end{bmatrix}^{-1} = \begin{bmatrix} 0.88888889 & -0.251897 \\ -0.251897 & 0.10033557 \end{bmatrix}$$

$$\hat{\beta} = \left(X^{*T}X^*\right)^{-1} X^{*T}y^* = \begin{bmatrix} 0.88888889 & -0.251897 \\ -0.251897 & 0.10033557 \end{bmatrix} * \begin{bmatrix} 4.00 \\ 13.60 \end{bmatrix} = \begin{bmatrix} 0.34408602 \\ 0.92903226 \end{bmatrix}$$

So fitted line approximately: $\hat{y} = 0.344086 + 0.929032x.$

5. Fitted transformed values and residuals

$$\hat{y}^* = X^*\hat{\beta}$$

$$= \begin{bmatrix} 0.8660254 & 0.8660254 \\ 0.5 & 1.5 \\ 0.5 & 2.0 \\ 0.5 & 2.5 \end{bmatrix} \cdot \begin{bmatrix} 0.34408602 \\ 0.92903226 \end{bmatrix}$$ (from step:2 and

$$= \begin{bmatrix} 1.039 & 1.7809677 & 1.9806452 & 2.3311823 \end{bmatrix}^T$$

step:4)

residuals: $e^* = y^* - \hat{y}^* = [0.0002305,\ 0.0190323,\ 0.0193548,\ 0.0688172]^T$

6. Compute $SSE = \sum (e_i^*)^2$ and $s^2 = \frac{SSE}{n-p}$ where p=2

   $SSE = \sum (e_i^*)^2 = 0.0688172043010752.$

   Degrees of freedom *n-p* =4-2=2

   $s^2 = \frac{SSE}{n-p} = \frac{0.0688172043010752}{2} = 0.0344086021505376.$

7. Variance-covariance of $\hat{\beta}$:

$$\text{Var}(\hat{\beta}) = s^2 (X^{*T} X^*)^{-1}$$

$$= 0.0344086021505376 \times \begin{bmatrix} 0.88888889 & -0.251897 \\ -0.251897 & 0.10033557 \end{bmatrix}$$

$$= \begin{bmatrix} 0.07843681 & -0.02219910 \\ -0.02219910 & 0.00887964 \end{bmatrix}$$

   Standard errors:

$$SE(\hat{\beta}_0) = \sqrt{Var(\hat{\beta}_0)} = \sqrt{0.07843681} = 0.28006573$$

$$SE(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} = \sqrt{0.00887964} = 0.09423184$$

8. T-test for slope (*df=2*):

$$t_{slop} = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{0.92903226}{0.09423184} = 9.8590060.$$

9. Compare with $t_{0.975,2}$=4.303

   Therefore, we reject the null hypothesis H_0: $\beta_1 = 0$ (Slope highly significant).

Inference:

- Fitted regression line with correlated observation $\hat{y} = 0.344086 + 0.929032x$.
- computed *SSE*= 0.0688172, s^2= 0.0344086,

$$\text{Var}(\hat{\beta}) = \begin{bmatrix} 0.07843681 & -0.02219910 \\ -0.02219910 & 0.00887964 \end{bmatrix},$$

$$SE(\hat{\beta}_0) = \sqrt{Var(\hat{\beta}_0)} = \sqrt{0.07843681} = 0.28006573,$$
$$SE(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} = \sqrt{0.00887964} = 0.09423184$$

and t_{slope}=9.8590, compared with $t_{0.975,2}$=4.303

Therefore, we reject the null hypothesis H_0: $\beta_1 = 0$ (Slope highly significant).

LAB EXERCISE 4:

Restricted Least Squares (restriction: $\beta_1+\beta_2=1$)

Problem: Given that

i	1	2	3	4	5	6
x_1	1	2	3	4	5	6
x_2	2	1	4	3	5	4
y	2.8	2.3	4.6	4.2	6.1	5.8

Fit the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$.

Restriction: $\beta_0 + \beta_1 = 1$.

compute residuals, SSE$_U$, SSE$_R$, s^2, Var$(\hat{\beta})$, $SE(\hat{\beta})$ and test the restriction.

Aim: To fit the data to the given model, compute residuals, SSE$_U$, SSE$_R$, s^2, Var$(\hat{\beta})$, $SE(\hat{\beta})$ and to test the restriction.

Procedure:

1. First form $X = [1 \quad x_1 \quad x_2]$ and y.
2. Compute ***X'X*** and ***X'y***.
3. Unrestricted OLS: $\hat{\beta}_U = (X'X)^{-1} X'\text{y}$. Compute fitted values

   $\hat{y}_U = X\hat{\beta}_U$, residuals $e_U = \text{y} - \hat{\text{y}}_U$, and $SSE_U = e'_U e_U$.

   Estimate $s_U^2 = SSE_U / (n-p)$.

   Compute $\text{Var}(\hat{\beta}_U) = s_U^2 (X'X)^{-1}$ and $SE(\hat{\beta}_{U,j}) = \sqrt{diag}$.
4. Form restriction: $R\beta = r$ where R=[0 1 1], r =1.
5. Restricted estimator $\hat{\beta}_R = \hat{\beta}_U - (X'X)^{-1} R^T [R(X'X)^{-1} R^T]^{-1} (R\hat{\beta}_U - r)$.
6. Compute restricted fitted values $\hat{y}_R = X\hat{\beta}_R$, residuals $e_R = y - \hat{y}_R$, and $SSE_R = e'_R e_R$.
7. Test the restriction (q = 1):

   $$F = \frac{(SSE_R - SSE_U)/\text{q}}{SSE_U/(n-p)}.$$

   Compare with $F_{q,n-p;1-\alpha}$.
8. Variance of restricted estimates:

   $$Var(\hat{\beta}_R) = s_U^2[(X'X)^{-1} - (X'X)^{-1} R^T \{R(X'X)^{-1} R^T\}^{-1} R(X'X)^{-1}],$$

and take square roots of diagonal as standard errors.

Calculation:

First form the matrices,

$$Y=\begin{bmatrix}2.8\\2.3\\4.6\\4.2\\6.1\\5.8\end{bmatrix}, X=\begin{bmatrix}1&1&2\\1&2&1\\1&3&4\\1&4&3\\1&5&5\\1&6&4\end{bmatrix},$$

$$X'X=\begin{bmatrix}1&1&2\\1&2&1\\1&3&4\\1&4&3\\1&5&5\\1&6&4\end{bmatrix}^T\cdot\begin{bmatrix}1&1&2\\1&2&1\\1&3&4\\1&4&3\\1&5&5\\1&6&4\end{bmatrix}=\begin{bmatrix}6&21&19\\21&91&77\\19&77&71\end{bmatrix}$$

$$X'\text{y}=\begin{bmatrix}1&1&2\\1&2&1\\1&3&4\\1&4&3\\1&5&5\\1&6&4\end{bmatrix}^T\begin{bmatrix}2.8\\2.3\\4.6\\4.2\\6.1\\5.8\end{bmatrix}=\begin{bmatrix}25.8\\103.3\\92.6\end{bmatrix}$$

$$(X'X)^{-1}=\begin{bmatrix}6&21&19\\21&91&77\\19&77&71\end{bmatrix}^{-1}=\begin{bmatrix}1.1176&-0.0588&-0.2353\\-0.0588&0.1366&-0.1324\\-0.2353&-0.1324&0.2206\end{bmatrix}$$

Therefore, unrestricted OLS

$$\begin{aligned}\hat{\beta}_U&=(X'X)^{-1}X'y\\&=\begin{bmatrix}1.1176&-0.0588&-0.2353\\-0.0588&0.1366&-0.1324\\-0.2353&-0.1324&0.2206\end{bmatrix}\times\begin{bmatrix}25.8\\103.3\\92.6\end{bmatrix}\\&=\begin{bmatrix}0.9706\\0.3326\\0.6838\end{bmatrix}\end{aligned}$$

Fitted values $\hat{y}_U = X\hat{\beta}_U$, residuals $e_U = \text{y} - \hat{\text{y}}_U$ give

$SSE_U = e'_U e_U = 0.0630$

Degrees of freedom $n-p = 6 - 3 = 3$.

$s_U^2 = SSE_U / (n-p) = 0.0630/3 = 0.0210$

Variance-covariance of unrestricted estimates:

$$\text{Var}(\hat{\beta}_U) = s_U^2 (X'X)^{-1} = 0.0210\begin{bmatrix} 1.1176 & -0.0588 & -0.2353 \\ -0.0588 & 0.1366 & -0.1324 \\ -0.2353 & -0.1324 & 0.2206 \end{bmatrix}$$

$$= \begin{bmatrix} 0.023472 & -0.001235 & -0.004942 \\ -0.001235 & 0.002868 & -0.002780 \\ -0.004942 & -0.002780 & 0.004633 \end{bmatrix}$$

Standard error (unrestricted):

$$SE(\hat{\beta}_{0,\text{U}}) = \sqrt{0.023472} = 0.1532$$

$$SE(\hat{\beta}_{1,\text{U}}) = \sqrt{0.002868} = 0.0536$$

$$SE(\hat{\beta}_{2,\text{U}}) = \sqrt{0.004633} = 0.0681.$$

t – statistics for β_1 (unrestricted):

$$t_{U,\beta_1} = \frac{\hat{\beta}_{1,U}}{SE(\hat{\beta}_{1,U})} = \frac{0.3326}{0.0536} = 6.2083 \quad (\text{df} = 3)$$

Which is greater than $t_{0.975,3}$=3.182, so β_1 is significant in unrestricted fit.

Restricted Estimator:

Set $R = [0\ \ 1\ \ 1]$, $r = 1$.

Compute scalar

$$R(X'X)^{-1}R^T = [0\ 1\ 1]\begin{bmatrix} 1.1176 & -0.0588 & -0.2353 \\ -0.0588 & 0.1366 & -0.1324 \\ -0.2353 & -0.1324 & 0.2206 \end{bmatrix}[0\ 1\ 1]^T \approx 0.0924$$

$$\Rightarrow [R(X'X)^{-1}R^T]^{-1} \approx 10.8167.$$

Therefore,

$$K = (X'X)^{-1}R^T[R(X'X)^{-1}R^T]^{-1}$$

$$= \begin{bmatrix} 1.1176 & -0.0588 & -0.2353 \\ -0.0588 & 0.1366 & -0.1324 \\ -0.2353 & -0.1324 & 0.2206 \end{bmatrix}[0\ 1\ 1]^T \times 10.8167$$

$$= \begin{bmatrix} -3.1818 \\ 0.0455 \\ 0.9545 \end{bmatrix}$$

Deviation from the restriction:

$$R\hat{\beta}_U - r = (\hat{\beta}_{1,U} + \hat{\beta}_{2,U}) - 1 = 0.3326 + 0.6838 - 1 = 0.0164.$$

Restricted estimate:

$$\hat{\beta}_R = \hat{\beta}_U - K(R\hat{\beta}_U - r) = \begin{bmatrix} 1.0227 \\ 0.3318 \\ 0.6682 \end{bmatrix}$$

(check: 0.3318+0.6682=1)

Restricted fitted values and residuals (element-wise):

Using $\hat{\beta}_R$ computed above, the restricted fitted values $\hat{y}_{R,i} = \beta_{0,R} + \beta_{1,R}x_{1,R} + \beta_{2,R}x_{2i}$ and $e_{R,i} = y_i - \hat{y}_{R,i}$

i	x_1	x_2	y_i	$\hat{y}_{R,i}$	$e_{R,i}$
1	1	2	2.8000	2.6909	0.1091
2	2	1	2.3000	2.3545	-0.0545
3	3	4	4.6000	4.6909	-0.0909
4	4	3	4.2000	4.3545	-0.1545
5	5	5	6.1000	6.0227	0.0773
6	6	4	5.8000	5.6864	0.1136

$$SSE_R = \sum_{i=1}^{6} e_{R,i}^2 = 0.0659$$

We have, $SSE_U = 0.0630$

Therefore, $S_U^2 = \frac{SSE_U}{n-p} = \frac{0.0630042}{3} = 0.021$

$$Var(\hat{\beta}_R) = s_U^2[(X'X)^{-1} - (X'X)^{-1}R^T\{R(X'X)^{-1}R^T\}^{-1}R(X'X)^{-1}]$$
$$= \begin{bmatrix} 0.003818 & -0.000955 & 0.000955 \\ -0.000955 & 0.002864 & -0.002864 \\ 0.000955 & -0.002864 & 0.002864 \end{bmatrix}$$

Standard error (restricted):

$$SE(\hat{\beta}_{0,R}) = \sqrt{0.003818} = 0.0618$$
$$SE(\hat{\beta}_{1,R}) = \sqrt{0.002864} = 0.0535$$
$$SE(\hat{\beta}_{2,R}) = \sqrt{0.002864} = 0.0535.$$

F – test for the restriction:

$$F = \frac{(SSE_R - SSE_U)/q}{SSE_U/(n-p)} = \frac{0.0659091 - 0.0630042}{0.0630042/3} = 0.1383.$$

Critical $F_{1,3;0.95} \approx 10.13$. Since $0.1383 < 10.13$, **fail to reject** the restriction $\beta_1 + \beta_2 = 1$ at 5% level

— the data are consistent with the constraint.

Inference:

Fitted the data to the given model and computed the residuals,

SSE_U, SSE_R, s^2, $\text{Var}(\hat{\beta})$, $SE(\hat{\beta})$.

Since $F < F_{1,3;0.95}$, **fail to reject** the restriction $\beta_1+\beta_2=1$ at 5% level — the data are consistent with the constraint.

LAB EXERCISE 5:

Test of Hypothesis for a Single Regression Coefficient

Problem: Suppose we fit the regression model $Y = \beta_0 + \beta_1 X + \varepsilon$ using the data:

X	1 2 3 4 5
Y	2 4 5 4 5

Test the hypothesis $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ at 5% level of significance.

Aim: To test whether the slop parameter (β_1) is significantly different from zero.

Procedure:

1. Compute $\bar{X}, \bar{Y}$.
2. Compute $S_{xx} = \Sigma(Xi - \bar{X})^2$ and $S_{xy} = \Sigma(X_i - \bar{X})(Y_i - \bar{Y})$
3. Estimate $\hat{\beta}_1 = S_{xy} / S_{xx}$ and $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$.
4. Fit the simple linear regression model $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$.and residuals $e_i = Y_i - \hat{Y}_i$.
5. Compute the residual Sum of Squares (SSE) $SSE = \sum_i e_i^2$ and error variance s^2 : $s^2 = \dfrac{SSE}{n-2}$
6. Standard error: $SE(\hat{\beta}_1) = \sqrt{s^2 / S_{xx}}$.
7. Test statistic $t = \dfrac{\hat{\beta}_1}{SE(\hat{\beta}_1)}$ with $df = n - 2$.
8. Compare $|t|$ to $t_{\alpha/2, n-2}$ and conclude.

Calculation:

Given n =5,

$$\bar{X} = \frac{1+2+3+4+5}{5} = 3$$

$$\bar{Y} = \frac{2+4+5+4+5}{5} = 4$$

X_i	Y_i	$(X_i - \bar{X})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
1	2	4	4
2	4	1	0
3	5	0	0
4	4	1	0
5	5	4	2
$\sum X_i = 15$	$\sum Y_i = 20$	$\Sigma(X_i - \bar{X})^2 = 10$	$\Sigma(X_i - \bar{X})(Y_i - \bar{Y}) = 6$

Therefore, $S_{xx} = \Sigma(Xi - \bar{X})^2 = 10$ and $S_{xy} = \Sigma(X_i - \bar{X})(Y_i - \bar{Y}) = 6$

Estimates: $\hat{\beta}_1 = S_{xy}/S_{xx} = 6/10 = 0.6$ and $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 4 - 0.6(3) = 2.2$

Fitted the simple linear regression model is

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$$\Rightarrow Y_i = 2.2 + 0.6X_i$$

X_i	Y_i	$\hat{Y}_i = 2.2 + 0.6X_i$	$e_i = Y_i - \hat{Y}_i$	e_i^2
1	2	2.8	-0.8	0.64
2	4	3.4	0.6	0.36
3	5	4.0	1.0	1
4	4	4.6	-0.6	0.36
5	5	5.2	-0.2	0.04
				$\sum e_i^2 = 2.40$

The residual Sum of Squares (SSE) $SSE = \sum e_i^2 = 2.40$

and error variance s^2 : $s^2 = \dfrac{SSE}{n-2} = \dfrac{2.40}{3} = 0.80$

Standard error: $SE(\hat{\beta}_1) = \sqrt{s^2/S_{xx}} = \sqrt{0.80/10} = 0.2828$.

Test statistic $t = \dfrac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \dfrac{0.6}{0.2828} = 2.121$ with $df = n - 2 = 5-2=3$.

Therefore, $|t| = 2.121 < t_{\alpha/2,n-2} = t_{0.025,3} = 3.182$, We accept H_0.

Inference:

Therefore, we accept the null hypothesis.

That is, Slope is not significant at 5% level of significance.

LAB EXERCISE 6:

One-Way ANOVA (CRD)

Problem: Three teaching methods are compared test scores:

- Method A: 68, 72, 70
- Method B: 75, 73, 74
- Method C: 65, 67, 66

Test $H_0 : \mu_A = \mu_B = \mu_C$ at α = 0.05.

Aim: To test $H_0 : \mu_A = \mu_B = \mu_C$ using one-way ANOVA (CRD) at α = 0.05.

Procedure:

1. Compute each treatment mean $\bar{y}_i$ and grand mean $\bar{y}$.
2. Compute Sum of Squares Between (SSB): $SSB = \sum_{i=1}^{k} n(\bar{y}_i - \bar{y})^2$.
3. Compute Sum of Squares Within (SSE): $SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$.
4. Compute degrees of freedom: $df_B = k-1, df_E = N-k$.
5. Compute mean squares: $MSB = \frac{SSB}{df_B}, MSE = \frac{SSE}{df_E}$.
6. Compute $F = \frac{MSB}{MSE}$ and compare to F_{α, df_B, df_E}.
7. If $|F| \le F_{\alpha, df_B, df_E}$ then we accept H_0 otherwise we reject H_0.

Calculation:

For each treatment $n_A = n_B = n_C = 3$. Total N=9.

1. Treatment means

$$\bar{y}_A = \frac{68+72+70}{3} = \frac{210}{3} = 70$$

$$\bar{y}_B = \frac{75+73+74}{3} = \frac{232}{3} = 74$$

$$\bar{y}_C = \frac{65+67+66}{3} = \frac{198}{3} = 66$$

2. Grand Mean

$$\overline{y} = \frac{68+72+70+75+73+74+65+67+66}{9} = 70$$

3. SSB

$$SSB = \sum_{i=1}^{k} n(\overline{y}_i - \overline{y})^2$$
$$= 3(70-70)^2 + 3(74-70)^2 + 3(66-70)^2$$
$$= 96$$

4. SSE (within-treatment)

   Compute deviations within each group:

$$\text{A}: (68-70)^2 + (72-70)^2 + (70-70)^2 = 4+4+0 = 8.$$
$$\text{B}: (75-74)^2 + (73-74)^2 + (74-74)^2 = 1+1+0 = 2.$$
$$\text{C}: (65-66)^2 + (67-66)^2 + (66-66)^2 = 1+1+0 = 2.$$

$$SSE: 8+2+2 = 12.$$

5. SST (total)

$$SST = SSB + SSW = 38 + 6 = 44.$$

6. Degrees of freedom

$$df_B = k-1 = 3-1 = 2, df_W = N-k = 9-3 = 6, df_T = N-1 = 8.$$

7. Mean Squares and F

$$MSB = \frac{SSB}{df_B} = \frac{96}{2} = 48, MSE = \frac{SSE}{df_E} = \frac{12}{6} = 2.$$

$$F = \frac{MSB}{MSE} = \frac{48}{2} = 24.$$

   Compare with critical value $F_{0.05,2,6} \approx 5.14$. Since 24>5.14, reject H_0.

Inference:

Therefore, at least one reaching method procedures having different mean score.

LAB EXERCISE 7:

Simple Linear Regression

Problem: The following data represent the number of study hours (X) and the corresponding exam scores (Y) of 6 students. Fit a simple linear regression model $Y=\beta_0+\beta_1X+\varepsilon$.

Student	Hours Studied (X)	Exam Score (Y)
1	2	20
2	4	40
3	6	50
4	8	65
5	10	80
6	12	95

Aim: To fit a simple linear regression line and interpret the relationship between study hours and exam scores.

Procedure:

1. Compute $\bar{X}, \bar{Y}$.
2. Compute the slope $\hat{\beta}_1 = \dfrac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$
3. Calculate the intercept $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$.
4. Fit the simple linear regression model $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$.

Calculation:

Mean of X, $\bar{X} = \dfrac{2+4+6+8+10+12}{6} = 7$

Mean of Y, $\bar{Y} = \dfrac{20+40+50+65+80+95}{6} = 58.33$

Now,

$$\sum(X_i - \bar{X})(Y_i - \bar{Y}) = 2(-38.33) + 4(-18.33) + 6(-8.33) + 8(6.67) + 10(21.67) + 12(36.67)$$
$$= 560$$

$$\sum(X_i - \bar{X})^2 = (2-7)^2 + (4-7)^2 + (6-7)^2 + (8-7)^2 + (10-7)^2 + (12-7)^2 = 70$$

The slope $\hat{\beta}_1 = \dfrac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = \dfrac{560}{70} = 8$

The intercept $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 58.33 - 8(7) = 2.33$

Therefore, the fitted the simple linear regression model is,

$$\hat{Y} = 2.33 + 8X.$$

Inference:

The fitted the simple linear regression model is,

$$\hat{Y} = 2.33 + 8X.$$

That is, the exam score increases by approximately 8 marks for every additional hour of study.

LAB EXERCISE 8:

Polynomial Regression (Quadratic Fit)

Problem: The following data represent the years of experience (X) and the monthly salary (Y, in ₹000s) of employees in a company. Fit a quadratic polynomial regression model.

Employee	1	2	3	4	5	6
Experience (X)	1	2	3	4	5	6
Salary (Y)	15	20	28	40	55	70

Aim: To fit a quadratic model: $Y = \beta_0 + \beta_1 X + \beta_2 X^2$.

Procedure:

First, compute $\sum X, \sum Y, \sum X^2, \sum X^3, \sum X^4, \sum XY, \sum X^2Y, n.$

The normal equations of the quadratic equation are

$$\sum Y = n\beta_0 + \beta_1 \sum X + \beta_2 \sum X^2;$$
$$\sum XY = \beta_0 \sum X + \beta_1 \sum X^2 + \beta_2 \sum X^3;$$
$$\sum X^2Y = \beta_0 \sum X^2 + \beta_1 \sum X^3 + \beta_2 \sum X^4$$

Slove these three normal equations for β_0, β_1 and β_2.

The fitted line is $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X^2$.

Calculation:

Here $n = 6$

X	**Y**	**X^2**	**X^3**	**X^4**	**XY**	**X^2Y**
1	15	1	1	1	15	15
2	20	4	8	16	40	80
3	28	9	27	81	84	252
4	40	16	64	256	160	640
5	55	25	125	625	275	1375
6	70	36	216	1296	420	2520
$\sum X = 21$	$\sum Y = 228$	$\sum X^2 = 91$	$\sum X^3 = 441$	$\sum X^4 = 2275$	$\sum XY = 994$	$\sum X^2Y = 4882$

Substitute the above in the normal equations, then we get

$$228 = 6\beta_0 + 21\beta_1 + 91\beta_2 \qquad (1)$$

$$994 = 21\beta_0 + 91\beta_1 + 441\beta_2 \qquad (2)$$

$$4882 = 91\beta_0 + 441\beta_1 + 2275\beta_2 \qquad (3)$$

By solving equations (1), (2) and (3). we get,

$\hat{\beta}_0 = 10.57$, $\hat{\beta}_1 = -1.10$ and $\hat{\beta}_2 = 2.93$.

Therefore, fitted line is $Y = 10.57 - 1.10X + 2.93X^2$.

Inference:

The quadratic equation is fitted $Y = 10.57 - 1.10X + 2.93X^2$.

The salary increases with experience, but the quadratic term indicates an accelerated growth pattern (salary rises faster as experience increases).

LAB EXERCISE 9:

Multicollinearity

Problem: A company records the following data on sales (Y in ₹000), advertising expenditure (X_1 in ₹000), and sales staff salary (X_2 in ₹000) for 5 months:

Month	Sales (Y)	Advertising (X_1)	Salary (X_2)
1	25	5	6
2	30	6	7
3	38	8	9
4	45	10	11
5	50	12	13

Check whether multicollinearity exists between X_1 and X_2.

Aim: Check whether multicollinearity exists between X_1 and X_2 for the given data.

Procedure:

1. Compute correlation coefficient between X_1 and X_2.

$$r_{12} = \frac{\Sigma(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)}{\sqrt{\Sigma(X_1 - \bar{X}_1)^2 \cdot \Sigma(X_2 - \bar{X}_2)^2}}$$

If $|r_{12}| > 0.80$, strong multicollinearity is suspected.

2. Compute Variance Inflation Factor (VIF) for each predictor

$$VIF(X_j) = \frac{1}{1 - R_j^2}$$

where R_j^2 is the coefficient of determination when regressing X_j on the other predictors.

If VIF>10, serious multicollinearity exists.

Calculation:

Calculate means:

$$\bar{X}_1 = \frac{5+6+8+10+12}{5} = 8.2$$

$$\bar{X}_2 = \frac{6+7+9+11+13}{5} = 9.2$$

Now,

$$\sum(X_1-\bar{X}_1)(X_2-\bar{X}_2) = (5-8.2)(6-9.2)+(6-8.2)(7-9.2)+(8-8.2)(9-9.2) + (10-8.2)(11-9.2)+(12-8.2)(13-9.2) = 29.2$$

$$\sum(X_1-\bar{X}_1)^2 = (5-8.2)^2+(6-8.2)^2+(8-8.2)^2+(10-8.2)^2+(12-8.2)^2 = 29.2$$

$$\sum(X_2-\bar{X}_2)^2 = (6-9.2)^2+(7-9.2)^2+(9-9.2)^2+(11-9.2)^2+(13-9.2)^2 = 29.2$$

$$\therefore r_{12} = \frac{\sum(X_1-\bar{X}_1)(X_2-\bar{X}_2)}{\sqrt{\sum(X_1-\bar{X}_1)^2\cdot\sum(X_2-\bar{X}_2)^2}} = \frac{29.2}{\sqrt{29.2\times 29.2}} = 1$$

Since the correlation is perfect (1.0), $R^2 = 1$.

Variance Inflation Factor (VIF), $VIF(X) = \frac{1}{1-R^2} = \frac{1}{1-1} = \infty.$

Inference:

There is perfect multicollinearity between advertising expenditure and sales staff salary. The two predictors carry the same information, so one of them should be dropped from the regression model.

LAB EXERCISE 10:

Principal Component Regression

Problem: A researcher studies the relationship between Crop Yield (Y in quintals) and two predictors: Rainfall (X_1 in cm) and Fertilizer (X_2 in kg). The data for 4 seasons is given below:

Season	Yield (Y)	Rainfall (X_1)	Fertilizer (X_2)
1	20	10	15
2	25	12	18
3	28	14	20
4	30	16	22

Apply Principal Component Regression (PCR) and reduce collinearity among predictors (Rainfall and Fertilizer) before estimating regression coefficients.

Aim: To apply Principal Component Regression (PCR) and reduce collinearity among predictors (Rainfall and Fertilizer) before estimating regression coefficients.

Procedure:

1. Compute sample means

   $$\bar{X}_1=\frac{1}{n}\sum X_{1i}, \bar{X}_2=\frac{1}{n}\sum X_{2i}, \bar{Y}=\frac{1}{n}\sum Y_i.$$

2. Compute sample standard deviations

   $$s_1=\sqrt{\frac{\sum(X_{1i}-\bar{X}_1)^2}{n-1}}; s_2=\sqrt{\frac{\sum(X_{2i}-\bar{X}_2)^2}{n-1}}$$

3. Compute standardize predictors (optional but convenient)

   $$Z_{1i}=\frac{X_{1i}-\bar{X}_1}{s_1}, Z_{2i}=\frac{X_{2i}-\bar{X}_2}{s_2}.$$

4. Form the correlation matrix *R:*

   $$R=\begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix},\ r_{12}=\frac{\sum(X_1-\bar{X}_1)(X_2-\bar{X}_2)}{\sqrt{\sum(X_1-\bar{X}_1)^2\cdot\sum(X_2-\bar{X}_2)^2}}$$

5. Compute eigen values by solving $\det(R-\lambda I)=0$

   For the 2×2 symmetric *R* this reduces to

   $(1-\lambda)^2-r_{12}^2=0 \Rightarrow \lambda=1\pm r_{12}$.

6. Compute eigenvectors v_1, v_2 corresponding to λ_1, λ_2. Normalize them to unit length. For

r_{12} close to +1, $v_1 \propto (1, 1)^T$, $v_2 \propto (1, -1)^T$.

7. Form PC1 scores (projection of standardized predictor vector on v_1):

$$z_{1i} = v_1^T \begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix}.$$

8. Regress centred response $y_{c,i} = Y_i - \bar{Y}$ on z_{1i} :

$$\hat{\gamma} = \frac{\sum_i z_{1i} y_{c,i}}{\sum_i z_{1i}^2}, \quad y_{c,i} = \hat{\gamma} z_{1i}.$$

9. Back-transform to coefficient on original predictors: if $v_1 = (v_{11}, v_{12})^T$,

$\hat{\beta}_{(c)} = \hat{\gamma} v_1$ (coefficients for centred predictors on the standardized scale),

Coefficients on original X:

$$\hat{b}_j = \frac{\hat{\beta}_{(c),j}}{s_j}, \quad j = 1, 2.$$

Intercept:

$$\hat{\beta}_0 = \bar{Y} - \hat{b}_1 \bar{X}_1 - \hat{b}_2 \bar{X}_2.$$

Calculation:

Means:

$$\bar{X}_1 = \frac{10 + 12 + 14 + 16}{4} = 13.0$$

$$\bar{X}_2 = \frac{15 + 18 + 20 + 22}{4} = 18.75$$

$$\bar{Y} = \frac{20 + 25 + 28 + 30}{4} = 25.75.$$

sample standard deviations

$$s_1 = \sqrt{\frac{\Sigma(X_{1i} - \bar{X}_1)^2}{n-1}}$$

$$= \sqrt{\frac{(10-13)^2 + (12-13)^2 + (14-13)^2 + (16-13)^2}{4-1}}$$

$$= 2.5819889$$

$$s_2 = \sqrt{\frac{\Sigma(X_{2i} - \bar{X}_2)^2}{n-1}}$$

$$= \sqrt{\frac{(15-18.75)^2 + (18-18.75)^2 + (20-18.75)^2 + (22-18.75)^2}{4-1}}$$

$$= 2.9877719$$

standardize predictors (optional but convenient)

$$Z_{11} = \frac{X_{11} - \bar{X}_1}{s_1} = \frac{10 - 13.0}{2.5819889} = -1.151895$$

Similarly,

$Z_{12} = -0.387298$

$Z_{13} = 0.387298$

$Z_{14} = 1.161895$

$$Z_{21} = \frac{X_{21} - \bar{X}_2}{s_2} == \frac{15 - 18.75}{2.9877719} = -1.2550$$

Similarly,

$Z_{22} = -0.2510$

$Z_{23} = 0.4185$

$Z_{24} = 1.0874$

The correlation:

$$r_{12} = \frac{\Sigma(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)}{\sqrt{\Sigma(X_1 - \bar{X}_1)^2 \cdot \Sigma(X_2 - \bar{X}_2)^2}}$$

$$= \frac{(-3)(-3.75) + (-1)(-0.75) + (1)(1.25) + (3)(3.25)}{2.5819889 \times 2.9877719}$$

$$= 0.9950$$

The correlation matrix *R*:

$$R = \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.9950 \\ 0.9950 & 1 \end{pmatrix}$$

Eigen values by solving $\det(R - \lambda I) = 0$

For the 2×2 symmetric *R* this reduces to

$(1-\lambda)^2 - r_{12}^2 = 0 \Rightarrow \lambda = 1 \pm r_{12}$.

$\Rightarrow 1 - \lambda = \pm 0.9950$

$\Rightarrow \lambda_1 = 1 + 0.9950 = 1.9950$

and $\lambda_2 = 1 - 0.9950 = 0.0050$

Eigen vectors:

For $\lambda_1 = 1 + r$:

Solve $\det(R - \lambda I)v = 0$

$$\begin{pmatrix} 1-\lambda_1 & r \\ r & 1-\lambda_1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0 \Rightarrow \begin{pmatrix} -r & r \\ r & -r \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0$$

$$\Rightarrow -rv_1 + rv_2 = 0$$

$$\Rightarrow v_1 = v_2.$$

Therefore, choose unnormalized eigen vectors $[1, 1]^T$.

Then normalized to unit length:

$$v_1 = \frac{1}{\sqrt{1^2+1^2}}\begin{pmatrix}1\\1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}.$$

For $\lambda_2 = 1-r$:

Similarly, $v_2 \propto [1,\ -1]^T$; normalized:

$$v_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}.$$

Now, form PC1 scores (projection of standardized predictor vector on v_1):

$$z_{1i} = v_1^T\begin{pmatrix}Z_{1i}\\Z_{2i}\end{pmatrix} = \frac{1}{\sqrt{2}}(Z_{1i}+Z_{2i})$$

$$\frac{1}{\sqrt{2}}(Z_{11}+Z_{21}) = \frac{1}{\sqrt{2}}(-1.161895+(-1.2550)) = -1.7090$$

$$\frac{1}{\sqrt{2}}(Z_{11}+Z_{22}) = \frac{1}{\sqrt{2}}(-0.387298+(-0.2510)) = -0.4512$$

$$\frac{1}{\sqrt{2}}(Z_{11}+Z_{23}) = \frac{1}{\sqrt{2}}(0.387298+0.4185) = 0.5699$$

$$\frac{1}{\sqrt{2}}(Z_{11}+Z_{24}) = \frac{1}{\sqrt{2}}(1.161895+1.0874) = 2.249295$$

Regress centred response $y_{c,i} = Y_i - \bar{Y}$ on z_{1i}:

$$y_{c,i} = Y_i - \bar{Y}$$
$$y_{c,1} = 20-25.75 = -5.75$$
$$y_{c,2} = 25-25.75 = -0.75$$
$$y_{c,3} = 28-25.75 = 2.25$$
$$y_{c,4} = 30-25.75 = 4.25$$

OLS slope on PC1:

$$\hat{\gamma} = \frac{\sum_i z_{1i}y_{c,i}}{\sum_i z_{1i}^2}$$
$$= \frac{(-1.7090\times-5.75)+(-0.4512\times-0.75)+(0.5699\times2.25)+(1.5903\times4.25)}{(-1.7090)^2+(-0.4512)^2+(0.5699)^2+(1.5903)^2}$$
$$= 3.046$$

Therefore, $y_c = 3.046z_1$.

Back-transform to coefficient on original predictors:

$$\hat{\beta}_{(c)} = \hat{\gamma} v_1 = 3.046 \times \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2.153 \\ 2.153 \end{pmatrix}.$$

Coefficients on original X:

$$\hat{b}_j = \frac{\hat{\beta}_{(c),j}}{s_j}, \quad j = 1, 2.$$

$$\Rightarrow \hat{b}_1 = \frac{\hat{\beta}_{(c),1}}{s_1} = \frac{2.153}{2.5819889} = 0.8346$$

$$\Rightarrow \hat{b}_2 = \frac{\hat{\beta}_{(c),2}}{s_2} = \frac{2.153}{2.9877719} = 0.7209$$

Intercept:

$$\hat{\beta}_0 = \overline{Y} - \hat{b}_1 \overline{X}_1 - \hat{b}_2 \overline{X}_2$$

$$\Rightarrow \hat{\beta}_0 = 25.75 - 0.8346 \times 13 - 0.7209 \times 18.75$$

$$= 1.3832$$

Therefore, final PCR regression is, $\hat{Y} = 1.3832 + 0.8346 X_1 + 0.7209 X_2$.

Inference:

Applied Principal Component Regression and fitted regression line

$\hat{Y} = 1.3832 + 0.8346 X_1 + 0.7209 X_2$.

THEORY OF LINEAR ESTIMATION AND ANALYSIS OF VARIANCE and ANOVA-Practical

LAB Exercise 1:

Rank, Signature and Index of the quadratic form

Problem: Find a rank, signature of a index of transformed form and normal form of given quadratic form.

$x_1^2+6x_1x_2+4x_1x_3+2x_1x_4+8x_1x_5+10x_2^2+16x_2x_3+8x_2x_4+26x_2x_5+12x_3^2+8x_3x_4+20x_3x_5+2x_4^2+10x_4x_5+17x_5^2$

Aim: For the given problem find rank, signature, and index.

Procedure:

- First we convert the given quadratic form into matrix form.
- Rank = no. of identity matrix $\therefore$[Rank = r]
- Index p = Rank
- Signature s = 2p-r

Calculation:

$$\text{LET } A = \begin{bmatrix} 1 & 3 & 2 & 1 & 4 \\ 3 & 10 & 8 & 4 & 13 \\ 2 & 8 & 12 & 4 & 10 \\ 1 & 4 & 4 & 2 & 5 \\ 4 & 13 & 10 & 5 & 17 \end{bmatrix}$$

Consider A = [I_5] A [I_5]

$$\begin{bmatrix} 1 & 3 & 2 & 1 & 4 \\ 3 & 10 & 8 & 4 & 13 \\ 2 & 8 & 12 & 4 & 10 \\ 1 & 4 & 4 & 2 & 5 \\ 4 & 13 & 10 & 5 & 17 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} A \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$R_2 \rightarrow R_2 - 3R_1$

$R_3 \rightarrow R_3 - 2R_1$

$R_4 \rightarrow R_4 - R_1$

$R_5 \rightarrow R_5 - 4R_1$

$$\begin{bmatrix} 1 & 3 & 2 & 1 & 4 \\ 0 & 1 & 2 & 1 & 1 \\ 0 & 2 & 8 & 2 & 2 \\ 0 & 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -3 & 0 & 0 & 0 & 0 \\ -2 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -4 & 0 & 0 & 0 & 1 \end{bmatrix} A \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$C_2 \rightarrow C_2 - 3C_1$

$C_3 \rightarrow C_3 - 2C_1$

$C_4 \rightarrow C_4 - C_1$

$C_5 \rightarrow C_5 - 4C_1$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 & 0 \\ 2 & -1 & 1/2 & 0 & 0 \\ 2 & -1 & 0 & 1 & 0 \\ -1 & -1 & 0 & 0 & 1 \end{bmatrix} A \begin{bmatrix} 1 & -3 & 4 & 2 & -1 \\ 0 & 1 & -2 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$C_3 \rightarrow C_3/2$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 & 0 \\ 2 & -1 & 1/2 & 0 & 0 \\ 2 & -1 & 0 & 1 & 0 \\ -1 & -1 & 0 & 0 & 1 \end{bmatrix} A \begin{bmatrix} 1 & -3 & 2 & 2 & -1 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

For the given problem rank (r) $=I_3=3$

Index $=p=3$

Signature (s) $=2p-r$

$=2(3)-3$

$=6-3$

Normal form $\varnothing$ is= 3

Given by

$y_1^2+y_2^2+y_3^2+0(y_4)^2+0(y_5)^2$

i.e., $y_1^2+y_2^2+y_3^2$

TRANSFORMATION:

$X = \varnothing Y$

$$\begin{bmatrix}1&0&0&0&0\\0&1&2&1&1\\0&2&8&2&2\\0&1&2&1&1\\0&1&2&1&1\end{bmatrix}=\begin{bmatrix}1&0&0&0&0\\-3&1&0&0&0\\-2&0&1&0&0\\-1&0&0&1&0\\4&0&0&0&1\end{bmatrix}A\begin{bmatrix}1&-3&-2&-1&4\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}$$

$R_3 \to R_3 - 2R_2$ $R_2 \to R_2 \to R_2$

$R_4 \to R_4 - R_2$

$R_5 \to R_5 - R_2$

$$\begin{bmatrix}1&0&0&0&0\\0&1&2&1&1\\0&0&4&0&0\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix}=\begin{bmatrix}1&0&0&0&0\\-3&1&0&0&0\\4&-2&1&0&0\\2&-1&0&1&0\\-1&-1&0&0&1\end{bmatrix}A\begin{bmatrix}1&-3&-2&-1&4\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}$$

$C_3 \to C_3 - 2c_2$ $C_4 \to C_4 - C_2$ $C_5 \to C_5 - C_2$

$$\begin{bmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&4&0&0\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix}=\begin{bmatrix}1&0&0&0&0\\-3&1&0&0&0\\4&-2&1&0&0\\2&-1&0&1&0\\1&-1&0&0&1\end{bmatrix}A\begin{bmatrix}1&-3&4&2&1\\0&1&-2&-1&-1\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}$$

$R_3 \to R_3/2$

$$\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\end{bmatrix}=\begin{bmatrix}1&-3&2&2&-1\\0&1&-1&-1&-1\\0&0&1/2&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\\y_4\\y_5\end{bmatrix}$$

$$\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\end{bmatrix}=\begin{bmatrix}y_1-3y_2+2y_3+2y_4-y_5\\0y_1+y_2-y_3-y_4-y_5\\0y_1+0y_2+(1/2)y_3+0y_4+0y_5\\0y_1+0y_2+0y_3+y_4+0y_5\\0y_1+0y_2+0y_3+0y_4+y_5\end{bmatrix}$$

$x_1 = y_1 - 3y_2 + 2y_3 + 2y_4 - y_5$

$x_2 = y_2 - y_3 - y_4 - y_5$

$x_3 = (1/2)y_3$

$x_4 = y_4$

$x_5 = y_5$

Inference:

For the given problem find rank, index, signature

Rank= 3

Index(p)=3

Signature=3

LAB EXERCISE 2:

Inverse of the Matrix

Problem: Find inverse of the following matrix

$$A = \begin{pmatrix} 2 & 0 & 1 \\ 4 & 1 & 2 \\ 1 & 2 & 6 \end{pmatrix}$$

Aim: To find the inverse of the given matrix.

Procedure:

$$A^{-1} = \frac{1}{\det(A)} Adj\ A$$

$$A^{-1} = \frac{1}{|A|} Adj\ A$$

Calculation:

First we calculate determinant of the given matrix

$$A = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & 2 & 6 \end{pmatrix}$$

[A]= 2(6-4)-0(24-2)+1(8-1)

=2(2)-0+1(7)

=4+7

[A]=11≠0

Now we calculate adj A for the given matrix

i.e., Adj A = [cofactor matrix]^T

cofactor of 2 =(6-4)=2

cofactor of 0 =(24-2)=-22

cofactor of 1=(8-1)=7

cofactor of 4= -(0-2)=2

cofactor of 1 = (12-1)=1

cofactor of 2= (-4-0)=-4

cofactor of 1=(0-1)=-1

cofactor of 2=(4-4)=0

cofactor of 6=(12-0)=12

$$\text{COFACTOR MATRIX} = \begin{pmatrix} 2 & -22 & 7 \\ 2 & 11 & -4 \\ -1 & 0 & 12 \end{pmatrix}$$

Adj A = transpose of the cofactor matrix

$$= \begin{pmatrix} 2 & -22 & 7 \\ 2 & 11 & -4 \\ -1 & 0 & 12 \end{pmatrix}$$

$$Adj = \begin{pmatrix} 2 & 2 & -1 \\ -22 & 11 & 0 \\ 7 & -4 & 12 \end{pmatrix}$$

$$A^{-1} = \frac{1}{\det(A)} Adj\ A$$

$$A^{-1} = \frac{1}{|A|} Adj\ A$$

$$= \frac{1}{11}\begin{pmatrix} 2 & 2 & -1 \\ -22 & 11 & 0 \\ 7 & -4 & 12 \end{pmatrix}$$

$$= \begin{pmatrix} 2/11 & 2/11 & -1/11 \\ -2 & 1 & 0 \\ 7/11 & -4/11 & 2/11 \end{pmatrix}$$

Inference:

The inverse of the given matrix is $\begin{pmatrix} 2/11 & 2/11 & -1/11 \\ -2 & 1 & 0 \\ 7/11 & -4/11 & 2/11 \end{pmatrix}$

LAB EXERCISE 3:

OLS and Gauss-Markov (BLUE)

Problem: Consider the linear model

$$y = X\beta + \varepsilon, \qquad E(\varepsilon) = 0, Var(\varepsilon) = \sigma^2 I.$$

Given the data

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \end{bmatrix},$$

(so n=4, p=2), compute the OLS estimator $\hat{\beta}$, show it is unbiased, compute its estimated variance-covariance matrix, and show (sketch) why it is BLUE (Gauss–Markov).

Aim:

1. Compute $\hat{\beta} = (X^T X)^{-1} X^T y$.
2. Compute residuals, SSE and $\hat{\sigma}^2$.
3. Compute $Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$ (estimate with $\hat{\sigma}^2$).
4. Give the Gauss–Markov argument for minimal variance among linear unbiased estimators.

Procedure:

1. Compute $X^T X$ and $X^T y$.
2. Invert $X^T X$.
3. Multiply to get $\hat{\beta}$.
4. Compute fitted values $\hat{y} = X\hat{\beta}$ and residuals $\hat{e} = y - \hat{y}$.
5. Compute $SSE = \hat{e}^T \hat{e}$ and $\hat{\sigma}^2 = \dfrac{SSE}{n-p}$.
6. Compute estimated covariance matrix $\text{Var}(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1}$.
7. Sketch the Gauss–Markov variance-minimizing argument.

Calculations:

1. Compute $X^T X$ and $X^T y$

$$\mathbf{X}^{\top} X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}.$$

$$\mathbf{X}^{\top} \mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \end{bmatrix}$$

$$= \begin{bmatrix} 2+3+5+7 \\ 1\cdot 2+2\cdot 3+3\cdot 5+4\cdot 7 \end{bmatrix}$$

$$= \begin{bmatrix} 17 \\ 2+6+15+28 \end{bmatrix} = \begin{bmatrix} 17 \\ 51 \end{bmatrix}.$$

2. Invert $X^T X$

Determinant: $\det(X^T X) = 4\times 30 - 10\times 10 = 20.$

Inverse: $(X^T X)^{-1} = \dfrac{1}{20}\begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix} = \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 0.2 \end{bmatrix}$

3. Compute $\hat{\beta}$

$\hat{\beta} = \left(X^T X\right)^{-1} X^T y = \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 0.2 \end{bmatrix}\begin{bmatrix} 17 \\ 51 \end{bmatrix} = \begin{bmatrix} 0 \\ 17 \end{bmatrix}$ (intercept 0, slope 1.7)

4. Fitted values and residuals

Predicted values:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i :$$
$$x_1 = 1:\ \hat{y}_1 = 0 + 1.7(1) = 1.7 \Rightarrow \hat{e}_1 = 2 - 1.7 = 0.3$$
$$x_2 = 2:\ \hat{y}_2 = 0 + 1.7(2) = 3.4 \Rightarrow \hat{e}_2 = 3 - 3.4 = -0.4$$
$$x_3 = 3:\ \hat{y}_3 = 0 + 1.7(3) = 5.1 \Rightarrow \hat{e}_3 = 5 - 5.1 = -0.1$$
$$x_4 = 4:\ \hat{y}_4 = 0 + 1.7(4) = 6.8 \Rightarrow \hat{e}_4 = 7 - 6.8 = 0.2.$$

Check residual sum: 0.3−0.4−0.1+0.2=00.3 - 0.4 - 0.1 + 0.2 = 00.3−0.4−0.1+0.2=0 (as required).

5. Compute $SSE = \hat{e}^T \hat{e}$ and $\hat{\sigma}^2 = \dfrac{SSE}{n-p}$

$$SSE = \sum_{i=1}^{4} \hat{e}_i^2 = 0.3^2 + (-0.4)^2 + (-0.1)^2 + 0.2^2 = 0.09 + 0.16 + 0.01 + 0.04 = 0.30.$$

Degrees of freedom *n-p* =4-2=2

$$\hat{\sigma}^2 = \frac{SSE}{n-p} = \frac{0.30}{2} = 0.15$$

6. Estimated covariance matrix $\text{Var}(\hat{\beta}\ ) = \hat{\sigma}^2 (X^T X)^{-1} = 0.15 \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 0.2 \end{bmatrix}$

   Therefore,

$$SE(\hat{\beta}_0) = \sqrt{0.225} \approx 0.4743$$
$$SE(\hat{\beta}_1) = \sqrt{0.03} \approx 0.1732$$

7. Estimation of a linear parametric function

   For $a^\top \boldsymbol{\beta} = \beta_1$ (i.e. $a = [0,1]^\text{T}$), estimate:
   $\hat{\theta} = a^\top \hat{\boldsymbol{\beta}} = 1.7, \text{Var}(\hat{\theta})\ = a^\top\ \text{Var}(\hat{\boldsymbol{\beta}}) a =\ 0.03$

8. Sketch proof of Gauss-Markov (minimal among linear unbiased estimators)
   - Any linear estimator of $\boldsymbol{\beta}$ can be written $\tilde{\boldsymbol{\beta}} = \boldsymbol{Ly}$.
   - Unbiasedness requires $\mathbb{E}(\tilde{\boldsymbol{\beta}}) = \mathbf{LX}\boldsymbol{\beta} = \boldsymbol{\beta}$ for all $\mathbf{LX} = \mathbf{I}_P$
   - Write $\mathbf{L} = (\mathbf{X}^\top \mathbf{X})^{\mathbf{-1}} \mathbf{X}^\top + \mathbf{A}$ where $\mathbf{AX} = 0$. Then $Var(\tilde{\boldsymbol{\beta}}) = \sigma^2 \mathbf{LL}^\top = \sigma^2 [(\mathbf{X}^\top \mathbf{X})^{-1} + \mathbf{AA}^\top]$.

     Since $\mathbf{AA}^\top$ is positive semi-definite, $Var(\tilde{\boldsymbol{\beta}}) - Var(\hat{\boldsymbol{\beta}}) = \sigma^2 \mathbf{AA}^T \geq 0$.

     Hence $\hat{\boldsymbol{\beta}}$ has minimum variance among linear unbiased estimators (BLUE).

Inference:

- The OLS estimate is $\hat{\beta} = \begin{bmatrix} 0 \\ 17 \end{bmatrix}$.
- The model residuals give $\hat{\sigma}^2 = 0.15$ and estimated covariance matrix of $\hat{\beta}$ shown above.
- By the Gauss-Markov theorem, $\hat{\beta}$ is the BEST (minimum variance) Linear Unbiased estimator under $Var(\varepsilon) = \sigma^2 I$.

LAB EXERCISE 4:

Estimating Parameters in a Simple Linear Model

Problem: Given a simple linear regression model

$$Y = X\beta + \varepsilon$$

With

$$Y = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}, X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix},$$

Find the Best Linear Unbiased Estimator (BLUE) of $\beta = [\beta_0, \beta_1]^T$.

Aim: To compute the least squares estimates of the intercept and slope in a simple linear regression model using the Gauss–Markov theorem.

Procedure:

1. Write down ***Y*** and ***X***.
2. Compute $\boldsymbol{X^T X}$ and $\boldsymbol{X^T Y}$.
3. Find $(\boldsymbol{X^T X})^{-1}$.
4. Apply the formula $\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$

Calculation:

Given,

$$Y = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}, X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix},$$

$$\mathbf{X}^\top X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

$$\mathbf{X}^\top Y = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 10 \\ 23 \end{bmatrix}$$

$$(\mathbf{X}^\top X)^{-1} = \frac{1}{3 \times 14 - 6 \times 6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix}$$

Therefore,

$$\hat{\beta} = \frac{1}{6}\begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix}\begin{bmatrix} 10 \\ 23 \end{bmatrix} = \begin{bmatrix} 0.333 \\ 1 \\ 1.5 \end{bmatrix}$$

Inference:

The estimated regression line is: $\hat{Y} = 1.5X + 0.333$.

Which means the slope is 1.5 and the intercept is 0.333.

LAB EXERCISE 5:

One-Way ANOVA (with Fisher's LSD)

Problem: Yields (kg) from three fertilizers, 3 replicates each:

- A: 20, 22, 21
- B: 18, 19, 20
- C: 23, 25, 24

Test at α = 0.05 whether the three treatment means are equal. If ANOVA is significant, use Fisher's LSD to compare means pairwise.

Aim: To test $H_0 : \mu_A = \mu_B = \mu_C$ using one-way ANOVA; if H_0 rejected, apply Fisher's LSD to find which pairs differ.

Procedure:

1. Compute each treatment mean $\bar{y}_i$ and grand mean $\bar{y}$.
2. Compute Sum of Squares Between (SSB): $SSB = \sum_{i=1}^{k} n(\bar{y}_i - \bar{y})^2$.
3. Compute Sum of Squares Within (SSW): $SSW = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$.
4. Compute degrees of freedom: $df_B = k-1, df_W = N-k$.
5. Compute mean squares: $MSB = \frac{SSB}{df_B}, MSE = \frac{SSW}{df_W}$.
6. Compute $F = \frac{MSB}{MSE}$ and compare to F_{α, df_B, df_W} .
7. If significant, compute Fisher's LSD: $LSD = t_{\alpha/2, df_W}\sqrt{\frac{2MSE}{n}}$ (for equal n).

   Compare $\left|\bar{y}_i - \bar{y}_j\right|$ with LSD.

Calculation:

For each treatment $n_A = n_B = n_C = 3$. Total N=9.

1. Treatment means

$$\overline{y}_A = \frac{20+22+21}{3} = \frac{63}{3} = 21.0$$
$$\overline{y}_B = \frac{18+19+20}{3} = \frac{57}{3} = 19.0$$
$$\overline{y}_C = \frac{23+25+24}{3} = \frac{72}{3} = 24.0$$

2. Grand Mean

$$\overline{y} = \frac{20+22+21+18+19+20+23+25+24}{9} = 21.3$$

3. SSB

   Compute each $(\overline{y}_i - \overline{y})$:

   For A: $\overline{y}_A - \overline{y} = 21 - 64/3 = (63-64)/3 = -1/3.$

   $n(\overline{y}_A - \overline{y})^2 = 3(1/9) = 1/3.$

   For B: $\overline{y}_B - \overline{y} = 19 - 64/3 = (57-64)/3 = -7/3.$

   $n(\overline{y}_B - \overline{y})^2 = 3(49/9) = 49/3.$

   For C: $\overline{y}_C - \overline{y} = 24 - 64/3 = (72-64)/3 = 8/3.$

   $n(\overline{y}_C - \overline{y})^2 = 3(64/9) = 64/3.$

   $$SSB = \sum_{i=1}^{k} n(\overline{y}_i - \overline{y})^2 = \frac{1}{3} + \frac{49}{3} + \frac{64}{3} = 38.0$$

4. SSW (within-treatment)

   Compute deviations within each group:

   $\text{A}: (20-21)^2 + (22-21)^2 + (21-21)^2 = 1+1+0 = 2.$

   $\text{B}: (18-19)^2 + (19-19)^2 + (20-19)^2 = 1+0+1 = 2.$

   $\text{C}: (23-24)^2 + (25-24)^2 + (24-24)^2 = 1+1+0 = 2.$

   $$SSW : 2+2+2 = 6.0.$$

5. SST (total)

   $SST = SSB + SSW = 38 + 6 = 44.$

6. Degrees of freedom

   $df_B = k-1 = 3-1 = 2, df_W = N-k = 9-3 = 6, df_T = N-1 = 8.$

7. Mean Squares and F

$MSB = \frac{SSB}{df_B} = \frac{38}{2} = 19.0, MSE = \frac{SSW}{df_W} = \frac{6}{6} = 1.0.$

$F = \frac{MSB}{MSE} = \frac{19.0}{1.0} = 19.0.$

Compare with critical value $F_{0.05,2,6} \approx 5.14$. Since 19.0>5.14, reject H_0. There is a treatment effect.

8. Fisher's LSD (comparison)

   Use $t_{\alpha/2,df_W} = t_{0.025,6} \approx 2.447$

   LSD=$t_{0.025,6}\sqrt{\frac{2MSE}{n}} = 2.447\sqrt{\frac{2\times 10}{3}} \approx 1.977 \approx 2.00$

   Pairwise mean differences:

   $|\,\bar{y}_A - \bar{y}_B| = |21-19| = 2.0 \geq \text{LSD} \Rightarrow$ significant.

   $|\,\bar{y}_C - \bar{y}_A| = |24-21| = 3.0 > \text{LSD} \Rightarrow$ significant.

   $|\,\bar{y}_C - \bar{y}_B| = |24-19| = 5.0 > \text{LSD} \Rightarrow$ significant.

Inference:

- ANOVA: $F(2,6)=19.0$, $p<0.02$ – treatment means differ.
- Fisher's LSD: all three pairwise differences are significant at 5% (LSD ≈ 2.00).
- Therefore, fertilizers A, B and C all produce different mean yields in this sample.

LAB EXERCISE 6:

Two-Way ANOVA (Balanced; test main effects and interaction)

Problem:

Two factors — Factor A (2 levels: A1, A2) and Factor B (2 levels: B1, B2). Two replicates per cell (r=2). Observations:

- A1B1: 8, 9
- A1B2: 7, 6
- A2B1: 10, 11
- A2B2: 9, 8

Test main effects of A, B and interaction AB at α = 0.05.

Aim: To test main effects of A, B and interaction AB at α = 0.05

Procedure:

1. Compute cell means $\bar{y}_{ij}$, row means $\bar{y}_{i.}$, column means $\bar{y}_{.j}$, and grand mean $\bar{y}$.
2. Compute:

$$SSA = br\sum_i (\bar{y}_{i.} - \bar{y})^2$$

$$SSB = ar\sum_j (\bar{y}_{.j} - \bar{y})^2$$

3. Compute integration sum of squares:

$$SS_{AB} = r\sum_{ij} (\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} - \bar{y})^2$$

4. total sum of squares $SST = \sum (y - \bar{y})^2$, *SSE = SST-SSA-SSB-SSAB.*
5. Find *MS* values and *F* ratios.

Calculation:

Grand mean $\bar{y}$ =(8+9+7+6+10+11+9+8)/8=68/8= 8.5

Row means:

$\bar{y}_{A1}$ = (8+9+7+6)/4 = 30/4 = 7.5,

$\bar{y}_{A2}$ = (10+11+9+8)/4 = 38/4 = 9.5.

Column means:

$\bar{y}_{B1}$ = (8+9+10+11)/4 = 38/4 = 9.5,

$\bar{y}_{B2}$ = (7+6+9+8)/4 = 30/4 = 7.5.

Sum of Squares for factor A:

Here *b* = 2, *r* = 2, *br* = 4;

$$SSA = br\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y})^2 = 4[(7.5-8.5)^2 + (9.5-8.5)^2] = 8.0$$

Sum of Squares for factor B:

Here $a = 2$, $r = 2$, $ar = 4$;

$$SSB = ar\sum_{j=1}^{b}(\bar{y}_{.\mathrm{j}} - \bar{y})^2 = 4[(9.5-8.5)^2 + (7.5-8.5)^2] = 8.0$$

Interaction Sum of Squares for factor AB:

$$SS_{AB} = r\sum_{ij}(\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} - \bar{y})^2 = 2\sum(8.5-7.5-9.5+8.5)^2 = 0$$

Total sum of squares:

$$SST = \sum(y-\bar{y})^2$$

$$= (8-8.5)^2 + (0.5-8.5)^2 + (-1.5-8.5)^2 + (-2.5-8.5)^2 + (1.5-8.5)^2 + (2.5-8.5)^2 + (0.5-8.5)^2 + (-0.5-8.5)^2$$
$$= 18.0.$$

Error sum of squares: $SSE = SST - SSA - SSB - SSAB = 2.0$

Degrees of freedom:

- $df_A = a - 1 = 1$
- $df_B = b - 1 = 1$
- $df_{AB} = (a-1)(b-1) = 1$
- $df_E = ab(r-1) = 2.2.1 = 4$
- $df_T = N - 1 = 1+1+1+4 = 7$

Mean squares:

$$MSA = \frac{SSA}{df_A} = \frac{8}{1} = 8.0, MSB = \frac{SSB}{df_B} = \frac{8}{1} = 8.0,$$

$$MSAB = \frac{SSAB}{df_{AB}} = \frac{0}{1} = 0, MSE = \frac{SSE}{df_E} = \frac{2}{4} = 0.5.$$

$$F_A = \frac{MSA}{MSE} = \frac{8.0}{0.5} = 16.0, F_B = \frac{MSB}{MSE} = \frac{8.0}{0.5} = 16.0, F_{AB} = \frac{MSAB}{MSE} = \frac{0.0}{1.0} = 0.$$

For 0.05 LOS and $df = (1,4)$, $F_{0.05,1,4} \approx 7.71$

- FA=16.0>7.71 → **reject** H_{0A} : factor A has a significant effect.
- FB=16.0>7.71 → **reject** H_{0B} : factor B has a significant effect.
- FAB=0.0<7.71 → **do not reject** H_{0AB} : no evidence of interaction.

Inference:

Main effects of *A* and *B* are significant; no interaction.

LAB EXERCISE 7:

Completely Randomized Design

Problem: A set of data involving 4 tropical feed stuffs A,B,C,D tried on 20 chicks is given below. All the 20 chicks are treats a like in all aspects expected the feeding treatments is given to 5 chicks. Analysis the data weight gain of baby chains fed on different feeding materials composed on the feed stuff.

	OBSERVATIONS	TOTAL(Ti)
A	55 49 42 21 52	219
B	61 112 30 89 63	355
C	42 97 81 95 92	407
D	169 137 169 85 154	714
	Grand total	G=1695

Aim: To test whether there is any significant difference between feeding materials composed of tropical feed stuffs.

Procedure:

The linear mathematical model of completely randomized design is

$Yij = \mu + ti + \varepsilon ij$

Where,

Yij is the yield of the j^{th} experiments receiving the i^{th} treatment

μ is general mean effect

Ɛj is random error effect

Null Hypothesis:

There is no significant difference between treatments.

$\mu 0: \mu 1 = \mu 2 = \ldots\ldots = \mu t$

under the null hypothesis μ0 all the treatments are homogenous

G= grand total $= \sum_{i=1}^{t} \sum_{j=1}^{ni} y_i j$

The mathematical model is

$Yij = \mu + ti + \varepsilon ij$

$\varepsilon ij = yij - \mu - ti$

Applying the principle of least square there we use estimate the values of unknown constants or parameters.

$$\frac{\partial \varepsilon}{\partial \mu} = 0; \frac{\partial \varepsilon}{\partial ti} = 0$$

Then the estimated values are $\mu \Lambda = \bar{y}$

$Ti = y_1 - \bar{y}$, ,substituting three estimation values in the mathematical model i.e.,

$Yij = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{ij} - \bar{y}_{i.})$

$(yij - \bar{y}..) = (\bar{yi}. - \bar{y}..) + (\bar{yij} - \bar{yi})$

Squaring and summering on both sides.

$i\sum j\sum(yij - \bar{y}..)^2 = i\sum j\sum[(\bar{yi}. - \bar{y}..) + (\bar{yij} - \bar{yi}.)^2]$

$_i\sum_j\sum(yij - \bar{y}..)^2 = _i\sum_j\sum[(\bar{yi}. - \bar{y}..)^2] + _i\sum_j\sum[(\bar{yij} - \bar{yi}.)]^2 + 2_i\sum_j\sum(\bar{yi}.\bar{y}..)(\bar{yij} - \bar{yi.})$

by arithmetic mean property the sum of the algebraic deviation taken from the mean value is always equals to zero.

$_i\sum_j\sum(yij - \bar{y.})^2 = _i\sum_j\sum(\bar{yi}. - \bar{y}..)^2 + _i\sum_j\sum(\bar{yij} - \bar{yi.})^2$

TSS=SST – SSE

$Si^2 = st^2 + sE^2$

First we calculate the grand total

$G = \sum\sum yij$

Next we calculate the row sum of squares

$RSS = _i\sum_j\sum yij^2$

Correction factor $CF = G^2/N (\therefore N = tr)$

Total sum of squares due to treatments (SST)

$SST = 1/ni \; \sum Ti^2 - CF$

Where ni is the no. of observations in i^{th} sample

Sum of squares due to error (SSE)

SSE=TSS-SST

ANOVA table:

Source of variation	df	Sum of squares	Mean sum of squares
Treatments	(t_i-1)	T	$T_n=T/t-1$
Errors	*	**	E'=**/*
total	(T-1)	TSS	=

$* = (t_r - 1) = (t - 1) = (t_r - t)$

**=TSS-T

CONCLUSIONS:

If F calculated value is less than the f-t calculated value with α% level of significance at(t-a,*)degrees of freedom then we say null hypothesis otherwise we reject null hypothesis

Calculation:

	OBSERVATIONS	TP	TP^2
A	55 49 42 21 52	219	4796
B	61 112 30 89 63	355	12602
C	42 97 81 95 92	407	16564
D	169 137 169 85 154	714	5093
		G=1695	8490

Correction factor (cf)=G^2/N

Where G= grand total

N= no.of observations

=2873025/20 => 143651.25

Treatment sum of squares (T_rSS)=$\sum Ti^2/k$ .cf

=849431/5=143651.25

=26234.95

Total sum of squares =$\sum\sum xij^2$-CF

=$(55)^2+(49)^2+$……..$(85)^2+(154)^2$-CF

=$(55)^2+(49)^2+$……..$(85)^2+(154)^2$-143651.25

=181445-143651.25

=37793.75

Errors sum of squares =ESS=TSS-SST

=37793.75-26234.95

=11558.8ANOVA TABLE:

Source of variation	d.f	Sum of squares	Mean sum of squares	F-ratio
Treatments	4-1=3	26234	8744.983	12.105 ~F(3,16)
Error	16	11558.8	722.425	
total	20-1=19	37793.75	-	

Inference:

Here the F- calculated value greater than the table value of F

i.e., $F_{cal} > f_{tab}$

12.105> 8.69

We reject the null hypothesis and we conclude that all the treatments effects are not homogeneous.

LAB EXERCISE 8:

Latin Square Design

Problem: In 4*4 latin square design the following data is obtained with yield of plots.

D 29.1	B 18.9	C 29.4	A 5.7
C 16.4	A 10.2	D 21.2	B 19.1
A 12.12	D 38.8	B 24	C 37
B 24.9	C 41.7	A 9.8	D 28.9

Analysis the data and draw conclusions

Aim: To analyze the data and draw conclusion for given data.

Procedure: The linear mathematical model of latin square design is

Yij(k)= μ +ri+cj+tk+Eij(k)

Here yij(k) is the yield from the experiment unit of i^{th} row and j^{th} column receiving k^{th} treatment

μ is the general mean effect

ri is the i^{th} row addictive effect

grand total G= $_i\sum_j\sum_k\sum y\overline{ij}(k)$

correction factor(CF)= G^2/rt (∴N=rt)

row sum of squares=R= $\sum Ri^2/t$ -CF

where Ri is the i^{th} row total

column sum of squares =c= $\sum cj^2/t$ -CF

treatment sum of squares =T= $\sum T_k^2/t$ -CF

total sum of squares = $_i\sum_j\sum_k\sum y\overline{ij}(k)$-CF

ANOVA TABLE:

Source of variation	d.f.	Sum of squares	Mean sum of squares	F_{cal}	F_{tab}
Rows	(t-1)	R	R^*=R/(t-1)	R'/E'	F(t-1,*)
Columns	(t-1)	C	C'=C/(t-1)	C'/E'	F(t-1,*)
Treatments	(t-1)	T	T'=T/(t-1)	T'/E'	F(t-1,*)
Error	*	**	E'=**/*		
Total	(t^2-1)	TSS			

Calculations:

					Ri.	$Ri.^2$
	D 29.1	B 18.9	C 29.4	A 5.7	83.1	6905.6
	C 16.4	A 10.2	D 21.2	B 19.1	66.9	4475.6
	A 12.12	D 38.8	B 24	C 37	111.92	12526.08
	B 24.9	C 41.7	A 9.8	D 28.9	105.3	11088.0
CJ	82.52	109.6	84.4	90.7	G=367.22	$\sum Ri^2$=34995.39
CJ^2	6809.5304	12012.16	7123.36	8226.49	$\sum Cij^2$=34171.5604	

$G = {}_I\sum{}_J\sum{}_K\sum(YIJ(K))$

=367.22

CORRECTION FACTOR:

$CF=G^2/rt =(367.22)^2/(4)(4)$

CF=8428.1580

Row sum of squares:

$R= \sum Ri^2./r$ -CF

=34925.3964/4 -8428.1580

=8748.8491-8428.1580

R=320.6911

Column sum of squares:

$C=\sum C.j^2/t$ -CF

=3471.5604/4 =8428.1580

=8542.8901-8428.1580

C=114.7321

					Ti.	$Ti.^2$
A	5.7	10.2	12.12	9.8	37.82	1430.3524
B	18.9	19.1	24	24.9	86.9	7551.61
C	29.4	16.4	37	41.7	124.5	15500.25
D	29.1	21.2	38.8	28.9	118	13924
					367.22	38406.2124

Treatment sum of squares:

$T=\sum Ti^2./r$ - CF

=38406.2124/4 -8428.1580

=9601.5531-8428.1580 =1173.3951

Total sum of squares:

TSS= $_i\sum_j\sum_k\sum \overline{yij}^2(k)$-CF=> (1075.6044)-(8428.952)

➢ 1747.4464

ANOVA TABLE:

Source of variation	d.f	Sum of squares	Mean sum of squares	F-ratio
Rows	(r-1)(4-1)=3	R=320.6911	R'=R/3=106.89 320.6911/3	R'/E'=106.89/23.1046 =4.6203~F(3,6)
Columns	(t-1)=3	C=114.7321	C'=C/3=114.7321/3 =38.2440	C'/E'=38.2440/23.1046 =1.6632~F(3,6)
Treatments	(8-1)=3	T=1173.3951	T'=T/3=1173.395/3 =391.1317	T'/E'=391.1317/23.1046 =16.9287~F(3,6)
Errors	6	E=TSS-R-C-T	E'=E/6=138.6281 =23.1046	
Total	15	TSS=1747.4464		

Inference:

For rows and columns, we accept H_0 and for treatments we reject the null hypothesis H_0.

LAB EXERCISE 9:

Duncan's Multiple Range Test (DMRT)

Problem: Using the following data from an experiment comparing three wheat varieties (r = 3 replications), determine which variety means differ by applying **Duncan's Multiple Range Test** at significance level α=0.05.

- Variety A: 48, 52, 50
- Variety B: 45, 47, 44
- Variety C: 51, 53, 52

Aim: To separate treatment means into homogeneous groups using Duncan's multiple range procedure after confirming treatments differ (ANOVA).

Procedure:

1. Compute treatment means:

$$\bar{X}_i = \frac{1}{r}\sum_{j=1}^{r} X_{ij} (i = 1, \ldots, t)$$

2. Compute grand mean: $\bar{X} = \frac{1}{N}\sum_{i=1}^{t}\sum_{j=1}^{r} X_{ij}$
3. Compute total sum of squares (about grand mean): $SST = \sum_{i=1}^{t}\sum_{j=1}^{r}(X_{ij} - \bar{X})^2$
4. Compute treatment (between) sum of squares (balanced design): $SSA = r\sum_{i=1}^{t}(\bar{X}_i - \bar{X})^2$
5. Compute error (within) sum of squares: $SSE = \sum_{i=1}^{t}\sum_{j=1}^{r}(X_{ij} - \bar{X}_i)^2$

(Note: SST=SSA+SSE.)

6. Degrees of freedom:
   - df(treatment) $= t-1$
   - df(error) $= N-t$
   - df(total) $= N-1$
7. Mean squares: $MST = \frac{SSA}{t-1}$, $MSE = \frac{SSE}{N-t}$
8. Check ANOVA: F=MST/MSE. If F significant, proceed to DMRT.
9. DMRT specific:

- Standard error: $s_e = \sqrt{\frac{MSE}{r}}$
- For a comparison spanning ***s*** ordered means, get studentized-range $q_{s,df_{error},\alpha}$ from table.
- Critical range: $R_s = q_s \times s_e$
- Compare ordered-mean differences with $\boldsymbol{R_s}$ (use appropriate ***s***). Form groups: means that are **not** significantly different share same letter.

Calculation:

A. Treatment means $\bar{X}_i$ and grand mean $\bar{X}$

- $\bar{X}_A = \frac{48+52+50}{3} = \frac{150}{3} = 50.000$
- $\bar{X}_B = \frac{45+47+44}{3} = \frac{136}{3} = 45.3333$
- $\bar{X}_C = \frac{51+53+52}{3} = \frac{156}{3} = 52.000.$

Grand mean: $\bar{X} = \frac{150+136+156}{9} = \frac{442}{9} = 49.1111$

B. SST, SSA, SSE (formulas + arithmetic)

1. SST (total SS about grand mean)

$$SST = \sum_i \sum_j (X_{ij} - \bar{X})^2 .$$

2. SSA (between treatments)

Formula: $SSA = r\sum_{i=1}^{t}(\bar{X}_i - \bar{X})^2.$

Compute each $(\bar{X}_i - \bar{X})$ and square:

- $\bar{X}_A - \bar{X} = 50.000 - 49.1111 = 0.8888889$
  $(\bar{X}_A - \bar{X})^2 = 0.8888889^2 = 0.7901234568.$
- $\bar{X}_B - \bar{X} = 45.3333 - 49.1111 = -3.7777778$
  $(\bar{X}_B - \bar{X})^2 = 14.2716049383.$
- $\bar{X}_C - \bar{X} = 52.000 - 49.1111 = 2.8888889$
  $(\bar{X}_C - \bar{X})^2 = 8.3456790123.$

Sum of squared deviations = 0.7901234568+14.2716049383+8.3456790123

=23.4074074074.0.7901234568 + 14.2716049383

+ 8.3456790123

=23.4074074074.

Multiply by r =3:

SSA=3×23.4074074074=70.2222222222.

3. SSE (within / error SS)

Formula: $SSE = \sum_{i=1}^{t}\sum_{j=1}^{r}(X_{ij} - \bar{X}_i)^2$.

Compute within each group:

- Group A: deviations from $\bar{X}_A = 50 : (48-50)^2 + (52-50)^2 + (50-50)^2 = 4+4+0 = 8$.
- Similarly, Group B: deviations from $\bar{X}_B = 45.3 :$ ≈4.6666667.
- Group C: deviations from $\bar{X}_C = 52 :$ 2.

So

SSE=8+14/3+2=44/3≈14.66666667.

Check: SST=SSA+SSE=70.22222222+14.66666667=84.88888889.

C. Degrees of freedom and mean squares

- df(treatment) =t−1=2.
- df(error) =N−t=9−3=6.
- df(total) =N−1=8.

Mean squares:

MST=SSA/(t−1) = 70.22222222 / 2=35.1111.

MSE=SSE/(N−t) = (44/3)/6=44/18=229≈2.4444.

ANOVA F:

F=MST/MSE=35.1111 / 2.4444 ≈ 14.3636 (significant at α=0.05).

Since ANOVA is significant, proceed with DMRT.

D. DMRT calculations (use formulas)

1. Standard error for DMRT:

$$s_e = \sqrt{\frac{MSE}{r}} = \sqrt{\frac{22/9}{3}} = \sqrt{\frac{22}{27}} \approx 0.90267093.$$

2. Order means (largest → smallest): $\bar{X}_C = 52.000, \quad \bar{X}_A = 50.000, \quad \bar{X}_B = 45.3333.$
3. From q-table (Duncan / studentized range) for df_{error}=6, α=0.05 (use text table):
   - $q_{s=2} \approx 2.89$
   - $q_{s=3} \approx 3.15$

4. Critical ranges:

$$R_2 = q_2 \times s_e = 2.89 \times 0.90267093 \approx 2.6087\ (\approx 2.609)$$

$$R_3 = q_3 \times s_e = 3.15 \times 0.90267093 \approx 2.8434\ (\approx 2.843)\ .$$

5. Pairwise (ordered) differences and comparisons:

- **C vs B (range length s=3s=3s=3)**:

$$\bar{X}_C - \bar{X}_B = 52.000 - 45.333\bar{3} = 6.666\ .$$

Compare with $R_3 = 2.843$: $6.667 > 2.843$ → **significant**.

- **C vs A (s = 2)**:

$$\bar{X}_C - \bar{X}_A = 52.000 - 50.000 = 2.000.$$

Compare with $R_2 = 2.609$: $2.000 < 2.609$ → **not significant**.

- **A vs B (s = 2)**:

$$\bar{X}_A - \bar{X}_B = 50.000 - 45.3333 = 4.6666\ .$$

Compare with $R_1 = 2.609$: $4.667 > 2.609$ → **significant**.

6. Grouping (Duncan):
   - C not separated from A → C and A grouped together (same letter “a”).
   - B separated from both → B alone (letter “b”).

Write as:

- C = 52.000 — group **c**
- A = 50.000 — group **a**
- B = 45.333 — group **b**

Inference:

Duncan’s Multiple Range Test (at α=0.05, using $q_2 \approx 2.89$, $q_3 \approx 3.15$) yields **two groups**: {C,A} and {B}.

Varieties C and A are not significantly different in mean yield; **Variety B is significantly lower** than both A and C.

LAB EXERCISE 10:

Test for Equality of Variances (Bartlett's test)

Problem: Test equality of variances for three groups (each n =5)

- Group A: 12.5, 13.2, 11.8, 14.0, 13.5
- Group B: 15.0, 15.3, 14.8, 15.7, 15.1
- Group C: 11.2, 11.8, 12.0, 11.5, 12.3

Aim: To test $H_0 : \sigma_A^2 = \sigma_B^2 = \sigma_C^2$ vs H_a : not all equal, using Bartlett's test at $\alpha = 0.05$

Procedure:

1. State hypothesis.

   $H_0 : \sigma_A^2 = \sigma_B^2 = \sigma_C^2$ vs H_a : not all equal.

2. For each group compute $n_i, \bar{x}_i$, and the sum of squared deviations $SS_i = \sum(x_{ij} - \bar{x}_i)^2$. Then sample (unbiased) variance $s_i^2 = SS_i / (n_i - 1)$.

3. Compute:

   $$s_p^2 = \frac{\sum(n_i - 1)s_i^2}{N - k},$$

   Where $N = \sum n_i$ and k = number of groups.

4. Compute Bartlett statistic with correlation:

   $$\chi_B^2 = \frac{(N-k)\ln s_p^2 - \sum(n_i - 1)\ln s_i^2}{1 + \frac{1}{3(k-1)}\left(\sum \frac{1}{n_i - 1} - \frac{1}{N-k}\right)}.$$

   Under $H_0, \chi_B^2 \sim \chi_{k-1}^2$.

5. Obtain p – value and compare with α.

Calculation:

Given $k = 3, n_1 = n_2 = n_3 = 5, N - k = 12.$

Group means:

Group A mean: $\bar{x}_1 = (12.5 + 13.2 + 11.8 + 14.0 + 13.5) / 5 = 13.00.$

Group B mean: $\bar{x}_2 = (15.0 + 15.3 + 14.8 + 15.7 + 15.1) / 5 = 15.18.$

Group C mean: $\bar{x}_3 = (11.2 + 11.8 + 12.0 + 11.5 + 12.3) / 5 = 11.76.$

Sum of squared deviations SS_i and unbiased variances $s_i^2 = SS_i / (n_i - 1)$:

Group A:

$SS_1 = (12.5-13.0)^2 + (13.2-13.0)^2 + (11.8-13.0)^2 + (14.0-13.0)^2 + (13.5-13.0)^2 = 2.98$

$\Rightarrow s_1^2 = 2.98/4 = 0.7450.$

Group B:

$SS_2 = (15.0-15.18)^2 + (15.3-15.18)^2 + (14.8-15.18)^2 + (15.7-15.18)^2 + (15.1-15.18)^2 = 0.468$

$\Rightarrow s_2^2 = 0.468/4 = 0.1170.$

Group C:

$SS_3 = (11.2-11.76)^2 + (11.8-11.76)^2 + (12.0-11.76)^2 + (11.5-11.76)^2 + (12.3-11.76)^2 = 0.732$

$\Rightarrow s_3^2 = 0.732/4 = 0.1830.$

(These SS_i are the sums of squared deviations from each group mean.)

Pooled variance:

$$s_p^2 = \frac{4(0.7450)+4(0.1170)+4(0.1830)}{12} = \frac{2.98+0.468+0.732}{12} = \frac{4.18}{12} = 0.3483.{}_{A1}$$

Numerator of Bartlett expression:

$$(N-k)\ln s_p^2 = 12\times\ln(0.3483) \approx 12\times(-1.0546) = -12.6551.$$

$$\begin{aligned}\Sigma(n_i-1)\ln s_i^2 &= 4\ln(0.7450)+4\ln(0.1170)+4\ln(0.1830)\\ &\approx 4(-0.2944)+4(-2.1456)+4(-1.6983)\\ &\approx -1.1775-8.5823-6.792(\approx) = -16.5529.\end{aligned}$$

Numerator = -12.6551-(-16.5529) = 3.8977.

Correction factor CF:

$$CF = 1+\frac{1}{3(k-1)}\left(\Sigma\frac{1}{ni-1}-\frac{1}{N-k}\right) = 1+\frac{1}{6}(3\times\frac{1}{4}-\frac{1}{12}) = 1.1111$$

Bartlett statistic: $\chi_B^2 = \dfrac{3.8977}{1.1111} \approx 3.5080.$

Degrees of freedom = k-1 = 2.

p – value for χ_2^2 with value 3.5080 is approximately $p \approx 0.1731$.

Inference:

At $\alpha=0.05$: $p=0.1731>0.05$. Fail to reject H_0. There is no statistical evidence that the variance differ – the variances may be treated as homogenous.