(AR)

Q16) The following information has been gathered from a random Sample of apartment renters in a city. We have information of rent in (000 Rs per month) based on the size of apartment, (number of rooms) and the distance from downtown (in KM).

Rent (000 Rs)	16	20	25	22	20	25
Number of rooms	4	6	3	4	5	3
Distance from downtown	8	10	4	6	2	1

a. Obtain the multiple regression models that best relate these variables

b. Interpret the obtained regression coefficients.

c. If someone is looking for a two bed apartment 8 km from down town, what rent should he expect to pay?

d. Obtain residuals

e. Calculate standard error of ~~dist~~ estimate

f. Test the significance of regression coefficient at 5% level of significance

g. Test overall significance of regression equation at 5% level of significance.

⇒

Working Steps:

1) Find out dependent and independent variable.
Here, Rent is dependent (say y), Number of rooms and Distance from downtown (say x_1 and x_2) respectively independent

2) Regression Equation of y on x_1 and x_2
To fit, $y = b_0 + b_1 x_1 + b_2 x_2$ Where, b_0, b_1 & b_2 are parameters of the three variable

- $\Sigma y = n b_0 + b_1 \Sigma x_1 + b_2 \Sigma x_2$
- $\Sigma y x_1 = b_0 \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2$
- $\Sigma y x_2 = b_0 \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2$

3) Use Cramer's rule for calculating the value of parameters a_0, a_1 & a_2

4) Residual Calculation

R_i = Observed rent − predicted rent

5) Standard Error Calculation,

$$S_e = \sqrt{\frac{SSE}{n-k-1}}$$

Where, $SSE = \Sigma(x_1 - \bar{x}_1)^2$

$= \Sigma x_1^2 - a\,\Sigma x_1 - b_2 \Sigma x_1 x_2 - b_3 \Sigma x_1 x_3$.

6) Test statistic calculation,

$$t = \frac{b_i}{S_{b_i}}$$

Where, b_i = Sample regression coefficient

S_{b_i} = Standard error of regression Coefficient

7) Test of Significance for Regression Coefficients

- Check the t-statistics and p-value for each Coefficient
- if p value > α level of significance then the Coefficient is significant.

9) Test of Overall significance of the Regression Coefficients

8) Calculate, test statistic

$$F = \frac{MSR}{MSE}$$

Where, MSR = mean sum of square due to regression

MSE = Mean Sum of square due to error

- Check f-Statistic and p value if p value $\geq$ α level of significance then regression model is significant.

// Output

Residuals :

1	2	3	4	5	6
-4.3168	2.3942	2.0852	0.8549	-1.8603	0.8428

Coefficients :

	Estimate	std. Error	t value	Pr(>|t|)	
Intercepts	27.3954	5.6128	4.881	0.0164	*
noofrooms	-0.9414	1.5725	-0.599	0.5916	
distance	-0.4141	0.5270	-0.786	0.4893	

Residual standard error : 3.348 on 3 degree of freedom
Multiple R-squared : 0.4334, Adjusted R-squared : 0.05569
F-statistic : 1.147 on 2 and 3 Df, p-value : 0.4265

// predicted rent
22.1998

Conclusion :

a. The multiple regression model obtained from the data is :

Rent = 27.3954 - 0.9414 No. of rooms - 0.4141 Distance

$$y = 27.3954 - 0.9414 x_1 - 0.4141 x_2$$

b. Interpretation of Coefficients :

- The intercept of 27.3954 represents the expected rent when the number of rooms and distance from downtown are both zero.
- For each additional room, the rent decreases by approximately 0.9414 RS with constant distance from downtown.
- For each additional kilometer away from downtown, the rent decreases by approximately 0.4141 RS with constant number of rooms.

c. If someone is looking for a two-bedroom apartment 8 km from downtown, the predicted rent would be approximately 22.1998 RS.

d. Residuals

1	2	3	4	5	6
−4.3168	2.3942	2.0852	0.8549	−1.8603	0.8428

e. The standard error of estimate is approximately 3.348 RS.

f. Testing the significance of regression coefficients:

- The coefficients b_1 and b_2 are insignificant at the 5% level, as indicated by their p-values being greater than 0.05. And Intercept (b_0) is significant.

g. Testing the overall significance of the regression equation:

- The p-value for the F-statistic is 0.4265, indicating that the regression equation as a whole is statistically insignificant at the 5% level. (P-value $= 0.4265 > \alpha = 0.05$).

Q17) A developer of food for pig would like to determine what relationship exists among the age of a pig when it starts receiving a newly developed food supplement, the initial weight of the pig and the amount of weight it gains in a week period with the food supplement. The following information is the result of study of eight piglets.

Initial Weight (Pounds)	39	52	49	46	61	36	28	57
Initial age (weeks)	8	7	6	11	8	7	9	5
Weight gain	8	7	6	9	10	6	4	5

a. Determine multiple correlation coefficient and partial correlation coefficients of dependent variable with independent variables.

b. Determine multiple coefficient of determination and interpret

c. Determine adjusted multiple coefficient of determination

⇒ <u>Working Steps</u>

1. Multiple Correlation Coefficient
   - It is lies between 0 and 1
     i.e. $0 < R_{1.23} \leq 1$, $0 \leq R_{2.13} \leq 1$, $0 \leq R_{3.12} \leq 1$
   - It is not less than simple correlation
     $R_{1.23} \geq r_{12}, r_{13}, r_{23}$

     Where, $R_{1.23} = \sqrt{\dfrac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}$
   - If $R_{1.23} = 0$ then $r_{12} = 0$, $r_{13} = 0$
   - $R_{1.23} = R_{1.32}$

2. Partial Correlation Coefficients
   Let, three variables X_1, X_2 & X_3 then, partial correlation coefficients betn X_1 and X_2 keeping X_3 constant is dentoted by $r_{12.3}$.

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2} \cdot \sqrt{1 - r_{23}^2}}$$

Same way,

$$r_{13.2} = \frac{r_{13} - r_{12}r_{32}}{\sqrt{1 - r_{13}^2} \cdot \sqrt{1 - r_{32}^2}} , \quad r_{23.1} = \frac{r_{23} - r_{21}r_{31}}{\sqrt{1 - r_{23}^2} \cdot \sqrt{1 - r_{31}^2}}$$

Calculate,

3. Multiple Coefficient of determination

$$R^2_{1.23} = \left(\sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}\right)^2$$

$$= \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}$$

in the same way $R^2_{2.13}$ and $R^2_{3.12}$.

4. Calculate adjusted multiple Coefficien of determinatio

$$R^2_{adjusted}(\bar{R}_2) = 1 - \frac{n-1}{(n-k-1)}[1 - R^2]$$

Where, n = no. of pair of observations
k = no. of independent Variables

//output

Coefficients :

	Estimate std.	Error	t-value	Pr(>|t|)
(Intercept)	-5.50561	3.77392	-1.459	0.2044
initialweight	0.13775	0.04926	2.797	0.0381
initialage	0.79269	0.29630	2.675	0.0441

Multiple R-squared : 0.6865 , Adjusted R-squared : 0.561

//sqrt(0.6865)

0.828553

$ estimate

	weightgain	initialweight	initialage
Weightgain	1.0000000	0.7810314	0.7672784
initialweight	0.7810314	1.0000000	-0.7470585
initialage	0.7672784	-0.7470585	1.0000000

a. Multiple Correlation Coefficient and Partial Correlation Coefficients:

- Multiple Correlation Coefficient
  - The multiple correlation coefficient R is 0.828553 indicating a strong positive linear relationship between the dependent variable (weight gain) and the independent variables (initial weight and initial age).

- Partial Correlation Coefficient
  - The partial correlation coefficients between weight gain and initial weight is 0.7810314, and between weight gain and initial age is 0.7672784. These values indicate the strength and direction of the relationships betn weight gain & each independent variable, controlling for the effect of the other independent variable.

b. Multiple Coefficient of Determination (R^2) & interpretation:

- Multiple coefficient of determination (R^2): 0.6865, which means that approximately 68.65 % of the variance in weight gain can be explained by the linear relationship with initial weight & initial age.

c. Adjusted Multiple Coefficient of determination

- The adjusted $R^2_{adj} = 0.561$

  This indicates that approximately 56.1 % of the variance in weight gain can be explained by the linear relationship with initial weight and initial age, adjusting for the number of predictors and the sample size.

Q18) let A, H, D and L represents Acer, HP, Dell and Lenovo laptop and following information represents their operating time in hours before Charge is required.

A	H	D	H	D	L
5.2	3.8	4.6	5.2	3.6	4.4
L	A	H	L	L	A
5.6	3.9	4.6	6.2	4.8	3.5
H	D	L	D	A	D
4.4	3.6	5.2	4.8	4.2	5.4
A	L	A	H	D	H
6.1	4.7	3.2	5.3	4.8	3.9

carryout analysis of the design at 1% level of significance.

→

1. Find out which design of experiment and calculate
 - It is CRD design of experiment
 So, calculate treatment $(t) = 4$ and Replication $(r) = 6$
 (A, H, D, L)

2. Mathematical Model

$$y_{ij} = \mu + T_i + e_{ij}$$

Where, $i = 1, 2, \dots, t$; $j = 1, 2, \dots, r$

y_{ij} = Observation or Yield due to ith treatment and jth replication
μ = General mean
T_i = Effect due to ith treatment
e_{ij} = Error due to ith treatment and jth replication

3. Calculate,

$$TSS = SST + SSE$$

Where, $TSS = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$

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

$SSE = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$

OR,
$TSS = \sum_i \sum_j y_{ij}^2 - \frac{G^2}{N}$
$SST = \sum_i \frac{T_{i.}^2}{r} - \frac{G^2}{N}$, $SSE = TSS - SST$

4. ANOVA Table

SV	df	SS	MS	Fcal	Ftab
Treatment	$t-1$	SST	$MST = \frac{SST}{t-1}$	$F_T = \frac{MST}{MSE}$	$F_\alpha(t-1, t(r-1))$
Error	$t(r-1)$	SSE	$MSE = \frac{SSE}{t(r-1)}$		
Total	$rt-1$	TSS			

5. Check F-value with a corresponding p-value, if p-value > α level of significance then Accept H_0. Else Reject.

// Output

	Df	Sum Sq	Mean Sq	F value	Pr(>F).
Laptop	3	2.308	0.7694	1.197	0.336
Residuals	20	12.857	0.6428		

Here, from the above result,

- The design used in the analysis is a Completely Randomized Design (CRD), as laptops of different brands were randomly assigned to the experimental unit without any specific blocking or grouping criteria.

- The value of F is 1.197 with a corresponding p-value is 0.336. Since, the p-value > $\alpha = 0.01$.
  So, Accept H_0 at $\alpha = 0.01$.

Problem to Test

H_0: Design is insignificant

H_1: Design is significant

Conclusion:

The design is insignificant.

Q19) Let A, H, D and L represents Acer, HP, Dell and Lenovo laptop and following information represents their operating time in hours before charge is required.

A 5.0	H 3.6	D 4.8	A 4.2	D 3.8	L 4.6
L 5.4	A 4.9	H 4.3	L 5.2	L 5.8	A 5.5
H 4.8	D 4.6	L 5.5	D 4.6	A 5.2	D 5.0
D 6.0	L 4.5	A 3.9	H 5.1	H 4.9	H 4.9

Carryout analysis of the design at 1% level of significance.

→

Working Steps

1. Find out that which design of experiment is it
   - If design is RBD then calculate treatment $(t) = 4$ and Block$(b) = 6$
     (A, H, D, L)

2. Mathematical Model

$$y_{ij} = \mu + T_i + B_j + e_{ij}$$

Where, $i = 1, 2, 3, \ldots, t$; $j = 1, 2, 3, \ldots, r$

y_{ij} = Yield or observation due to ith treatment and jth block.

μ = General Mean

T_i = Effect due to ith treatment

B_j = Effect due to jth block

e_{ij} = Error due to ith treatment and jth block

3. Calculate,

$$TSS = SST + SSB + SSE$$

Where, $TSS = \sum_i \sum_j y_{ij}^2 - \frac{G^2}{N}$

$$SST = \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_i \frac{T_{i.}^2}{r} - \frac{G^2}{N}$$

$$SSB = \sum_j \frac{T_{.j}^2}{t} - \frac{G^2}{N}$$

$$SSE = TSS - SST - SSB.$$

4. ANOVA Table

SV	df	SS	MS	Fcal	Ftab
Treatment	$t-1$	SST	$MST = \frac{SST}{t-1}$	$F_T = \frac{MST}{MSE}$	$F_\alpha(t-1, (t-1)(r-1))$
Block	$r-1$	SSB	$MSB = \frac{SSB}{r-1}$	$F_B = \frac{MSB}{MSE}$	
Error	$(r-1)(t-1)$	SSE	$MSE = \frac{SSE}{(t-1)(r-1)}$		
Total	$rt-1$	TSS			

5. Check F_T value and F_B value with a corresponding p-value if p-value $> \alpha$ level of significance then Accept H_0T and H_0B. Otherwise Reject.

// output

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Laptop	3	1.015	0.3382	0.945	0.44
Block	5	1.954	0.3908	1.092	0.405
Residuals	15	5.368	0.3579		

→ Problem to Test

H_0T : Treatment of design is insignificant
H_1T : Treatment of design is significant

H_0B : Block of design is insignificant
H_1B : Block of design is significant

- The F value of treatment is 0.945 with a corresponding p-value is 0.44. Since, p-value $= 0.44 > \alpha = 0.01$ Accept H_0T at $\alpha = 0.01$.
- The F value of Block is 1.092 with a corresponding p-value is 0.405. Since, p-value $= 0.405 > \alpha = 0.01$ Accept H_0B at $\alpha = 0.01$

Hence,
Treatment of design is insignificant.
Block of design is insignificant.

(The design used in this analysis is a Randomized Block Design (RBD), where laptops are grouped into blocks and treatments (laptop brands) are randomly assigned within each block.)

Q20) Let A, H, D and L represents Acer, HP, Dell and Lenovo laptop and following information represents their operating time in hours before charge is required.

A 4.2	H 4.8	D 4.2	L 6.2
L 4.6	A 5.9	H 4.8	D 5.2
H 5.4	D 5.6	L 5.6	A 4.8
D 4.1	L 5.7	A 4.2	H 4.3

Carry out analysis of the design at 5% level of significance.

→

Working steps.

1. Find out which design of experiment is it.
   - If design is LSD then calculate, treatment (t) = (4), row = 4, column = 4
     $\Downarrow$ (A, H, D, L)

2. Mathematical Model

$$y_{ijk} = \mu + r_i + c_j + T_k + e_{ijk}$$

Where, $i = 1, 2, 3, ----, m$; $j = 1, 2, 3, ----, m$; $k = 1, 2, 3, ---, m$

y_{ijk} = Observation due to i^{th} row, j^{th} column and k^{th} treatment

μ = General Mean

r_i = Effect due to ith row

c_j = Effect due to jth column

T_k = Effect due to k^{th} treatment

e_{ijk} = Error due to ith row, j^{th} column & k^{th} treatment

3. Calculate,

$$TSS = SSR + SSC + SST + SSE$$

Where,

$$TSS = \sum_{i,j,k} y_{ijk}^2 - \frac{G^2}{N}$$

$$SSR = \sum_i \frac{T_{i..}^2}{m} - \frac{G^2}{N}$$

$$SSC = \sum_j \frac{T_{.j.}^2}{m} - \frac{G^2}{N}$$

$$SST = \sum_k \frac{T_{..k}^2}{m} - \frac{G^2}{N}$$

$$SSE = TSS - SSR - SSC - SST$$

4. ANOVA Table

SV	df	SS	MS	Fcal
Row	$m-1$	SSR	$MSR = SSR/m-1$	$FR = MSR/MSE$
Column	$m-1$	SSC	$MSC = SSC/m-1$	$FC = MSC/MSE$
Treatment	$m-1$	SST	$MST = SST/m-1$	$F_T = MST/MSE$
Error	$(m-1)(m-2)$	SSE	$MSE = SSE/(m-1)(m-2)$	
Total	m^2-1	TSS		

Ftab

$F_\alpha[m-1, (m-1)(m-2)]$

5. Check F_T value, F_r value, F_C value with a corresponding p-value, if p-value > α-level of significance then Accept H_0T, H_0r & H_0C. Reject otherwise.

// Output

	Df	Sum sq	Mean sq	F-value	pr(>F)
laptop	3	1.620	0.5400	1.588	0.288
row	3	1.355	0.4517	1.328	0.350
Column	3	2.135	0.7117	2.093	0.203
Residuals	6	2.040	0.3400		

→ Problem To Test

H_0T: Treatment of design is insignificant
H_1T: Treatment of design is significant
H_0R: row of design is insignificant
H_1R: row of design is significant
H_0C: Column of design is insignificant
H_1C: Column of design is Significant

- The f value of treatment is 1.588 with a corresponding p-value is 0.288. Since, p-value = 0.288 > α = 0.05. Accept H_0T at $\alpha = 0.05$ level of significance.

- The F-value of Row is 1.328 with a Corresponding p-value is 0.350. Since, p-value > $\alpha = 0.05$. Accept H_{0R} at $\alpha = 0.05$ level of significance.

- The F-value of Column is 2.093 with a Corresponding p-value is 0.203. Since, p-value > $\alpha = 0.05$. Accept H_{0C} at $\alpha = 0.05$ level of significance.

Hence,

Treatment of design is insignificant.
Row of design is insignificant.
Column of design is insignificant.

[The design used in this analysis is an ~~Inco~~ Latin Square Design (LSD), with laptops grouped into blocks (row and column), and treatments (laptop brands) assigned within each block.].