B Multiple Linear Regression

Question A statistician wants to predict the income of resturants, using two independent variables: the number of resturant's employees and resurant floor area. Caclulate the estimated multiple linear regression equation for the following data collected by statistician. Also find Variance of Multiple regression, Multiple Coefficient of Determination, Multipla Correlation Coefficient.

Rest #	Income (000) Y	Floor area (000 sq.ft) X_1	Number of employees X_2	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	Y^2	$\hat{Y}$	$Y-\hat{Y}$	$(Y-\hat{Y})^2$
A	30	10	15	100	225	150	300	450	900	26.75	3.25	10.53
B	22	5	8	25	64	40	110	176	484	13.54	8.46	71.61
C	16	10	12	100	144	120	160	192	256	21.90	-5.90	34.79
D	7	3	7	9	49	21	21	49	49	11.16	-4.16	17.35
E	14	2	10	4	100	20	28	140	196	15.64	-1.64	2.71
	$\sum Y$	$\sum X_1$	$\sum X_2$	$\sum X_1^2$	$\sum X_2^2$	$\sum X_1X_2$	$\sum X_1Y$	$\sum X_2Y$	$\sum Y^2$			$\sum(Y-\hat{Y})^2$
	89	30	52	238	582	351	619	1007	1885		0.0000	136.9804

$\bar{Y}=$ 17.8 $\bar{X}_1=$ 6 $\bar{X}_2=$ 10.4

$$b = \frac{(n\sum xy - \sum x \sum y)/n}{[n\sum x^2 - (\sum x)^2]/n}$$

163.8195717 300.8

Multiple Linear Regression Equation

$\hat{Y} = a + b_1X_1 + \cancel{b_2X_2}$

$$b = \frac{\sum xy - \sum x\sum y/n}{\sum x^2 - (\sum x)^2/n}$$

$$b = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$$

$x-\bar{x}=x$

$y-\bar{y}=y$

$$b_1 = \frac{\sum x_1y\sum x_2^2 - \sum x_2y\sum x_1x_2}{\sum x_1^2\sum x_2^2 - (\sum x_1x_2)^2} = \frac{327.4}{868.6} = \boxed{0.38}$$

$$b_2 = \frac{\sum x_2y\sum x_1^2 - \sum x_1y\sum x_1x_2}{\sum x_1^2\sum x_2^2 - (\sum x_1x_2)^2} = \frac{1406.2}{868.6} = \boxed{1.62}$$

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2 = \boxed{-1.30}$$

$$\hat{Y} = -1.30 + 0.38X_1 + 1.62X_2$$

$$\sum x_1^2 = \sum X_1^2 - \frac{(\sum X_1)^2}{n} = 58$$

$$\sum x_2^2 = \sum X_2^2 - \frac{(\sum X_2)^2}{n} = 41.2$$

$$\sum x_1x_2 = \sum X_1X_2 - \frac{(\sum X_1)(\sum X_2)}{n} = 39$$

$$\sum x_1y = \sum X_1Y - \frac{(\sum X_1)(\sum Y)}{n} = 85$$

$$\sum x_2y = \sum X_2Y - \frac{(\sum X_2)(\sum Y)}{n} = 81.4$$

$$\sum (x_2-\bar{x}_2)^2 = \sum x_2^2 = \sum X_2^2 - \frac{(\sum x_2)^2}{n}$$

Dr. Shabbir Ahmad

Assistant Professor

Department of Mathematics

COMSATS University Islamabad, Wah Campus

Variance and Standard Deviation of Multiple Linear Regression

$$b = \frac{\sum xy}{\sum x^2}$$

$x=X-\bar{X}$

$y=Y-\bar{Y}$

$$S^2_{y.x_1x_2} = \frac{\sum(Y-\hat{Y})^2}{n-3} = \frac{136.9804}{2} = \boxed{68.49}$$

$$S^2_{y.x_1x_2} = \frac{\sum Y^2 - a\sum Y - b_1\sum X_1Y - b_2\sum X_2Y}{n-3} = \frac{136.9804}{2} = \boxed{68.49}$$

$$S_{y.x_1x_2} = \boxed{8.28}$$

Multiple Coefficient of Determination

$$R^2_{y.x_1x_2} = \frac{a\sum Y + b_1\sum X_1Y + b_2\sum X_2Y - \frac{(\sum Y)^2}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}} = \frac{163.82}{300.80} \quad \boxed{0.54}$$

About 54% of total variation in Income (Y) is explained by its multiple linear relationship Floor Area (X1) and No of Employes (X2).

Multiple Coefficient of Correlation is a measure of the strength of the association among the independent (explanatory) variables and the one dependent (prediction) variable.

$$r_{yx_1x_2} = \frac{\sqrt{r^2_{yx_1} + r^2_{yx_2} - 2\, r_{yx_1}\, r_{yx_2}\, r_{x_1x_2}}}{\sqrt{1 - r^2_{x_1x_2}}} = \frac{0.4449296}{0.6029028} \quad \boxed{0.74}$$

$r_{yx_1} =$ 0.64353

$r_{yx_2} =$ 0.73120

$r_{x_1x_2} =$ 0.79781

Partial Coefficient of Correlation It measures the degree of linear relationship between any two variables in a multivariable problem, under the condition that any common relationship of influence with all other variables (or some of them) has been removed or assumed to be as constant.

X_1	X_2	X_3
7	4	1
12	7	2
14	8	4
17	9	5
20	12	8

$r_{12} =$ 0.98747

$r_{23} =$ 0.97065

$r_{13} =$ 0.95902

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r^2_{13})(1 - r^2_{23})}} = \frac{0.05659}{0.06814} \quad \boxed{0.83}$$

$$r_{23.1} = \frac{r_{23} - r_{21}r_{31}}{\sqrt{(1 - r^2_{21})(1 - r^2_{31})}} = \frac{0.02364}{0.04472} \quad \boxed{0.53}$$

$$r_{31.2} = \frac{r_{31} - r_{32}r_{12}}{\sqrt{(1 - r^2_{32})(1 - r^2_{12})}} = \frac{0.00054}{0.03796} \quad \boxed{0.01}$$

Dr. Shabbir Ahmad
Assistant Professor
Department of Mathematics
COMSATS University Islamabad, Wah Campus