Multiple Regression (Muti-linear regression)

In multiple regression, we deal with data consisting of $n(r+1)$-tuples $(x_{i1}, x_{i2}, \dots x_{ir}, y_i)$ where the x_i's are assumed to be known while the y_i's are values of the random variable.

For any given set of values $x_1, x_2 \dots x_r$ for the r independent variables, the mean of the distribution of y is given by

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_r x_r .$$

For any two independent variables, the problem of fitting a plane to n points with co-ordinates are (x_{i1}, x_{i2}, y_i).

Apply the method of least squares to obtain estimates of the co-efficients $\beta_0, \beta_1, \beta_2$. We minimize the sum of squares of the vertical distances from the observations y_i to the plane.

we minimize

$$q = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}))^2$$

Diff. the above equation w.r. to $\beta_0, \beta_1, \beta_2$ we get the normal equations.

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

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

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

Problem:- The following are data on the number of twists required to break a certain kind of forged alloy bar and the percentages of two alloying elements present in the metal.

No. of twists: (y)	41	49	69	65	40	50	58	57	31	36	44	57	19	31	33	4
Percentage of elt A (x_1):	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Percentage of elt B (x_2):	5	5	5	5	10	10	10	10	15	15	15	15	20	20	20	20

Fit a least square regression plane and use its equation to estimate the no. of twists required to break one of bars where $x_1 = 2.5$ and $x_2 = 12$.

Soln:- Let the required equation of plane be $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

ie $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

y	x_1	x_2	$x_1 y$	$x_2 y$	x_1^2	x_2^2	$x_1 x_2$
41	1	5	41	205	1	25	5
49	2	5	98	245	4	25	10
69	3	5	207	345	9	25	15
65	4	5	260	325	16	25	20
40	1	10	40	400	1	100	10
50	2	10	100	500	4	100	20
58	3	10	174	580	9	100	30
57	4	10	228	570	16	100	40
31	1	15	31	465	1	225	15
36	2	15	72	540	4	225	30
44	3	15	142	660	9	225	45
57	4	15	228	855	16	225	60
19	1	20	19	380	1	400	20
31	2	20	62	620	4	400	40
33	3	20	99	660	9	400	60
43	4	20	172	860	16	400	80
$\Sigma y = 723$	$\Sigma x_1 = 40$	$\Sigma x_2 = 200$	1963	8210	120	3000	500

The normal equations are

$$\Sigma y = nb_0 + b_1 \Sigma x_1 + b_2 \Sigma x_2$$

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

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

Substituting the values we have

$$723 = 16 b_0 + 40 b_1 + 200 b_2$$

$$1963 = 40 b_0 + 120 b_1 + 500 b_2$$

$$8210 = 200 b_0 + 500 b_1 + 3000 b_2.$$

Solving we have

$$b_0 = 46.4, \quad b_1 = 7.78, \quad b_2 = -1.65$$

The equation of the estimated regression plane is $y = 46.4 + 7.78 x_1 - 1.65 x_2$.

Substituting $x_1 = 2.5$, $x_2 = 12$, we have

$$y = 46.4 + 7.78 \times 2.5 - 1.65 \times 12$$

$$\underline{\underline{y = 46.0}}$$

Problem ②:

A set of experimental runs was made to determine a way of predicting cooking time y at various levels of ovenwidth x_1 and flue temperature x_2. The coded data were recorded as follows.

y:	6.40	15.05	18.75	30.25	44.85	48.94	51.55	61.50	100.44
x_1:	1.32	2.69	3.56	4.41	5.35	6.20	7.12	8.87	9.80
x_2:	1.12	3.40	4.10	8.75	14.82	15.15	15.32	18.18	35.19

Estimate the multilinear regression plane $y = b_0 + b_1 x_1 + b_2 x_2$.

Scanned by CamScanner