Multiple Linear Regression

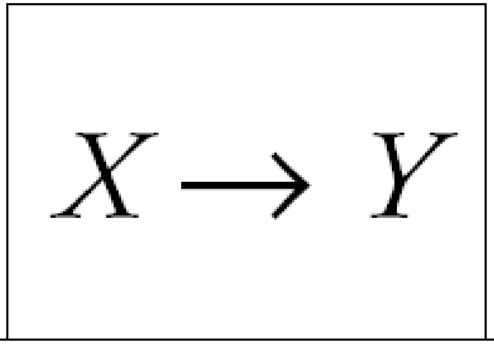
we have seen the concept of simple linear regression where a single predictor variable X was used to model the response variable Y .

In many applications, there is more than one factor that influences the response. Multiple regression models thus describe how a single response variable Y depends linearly on a number of predictor variables.

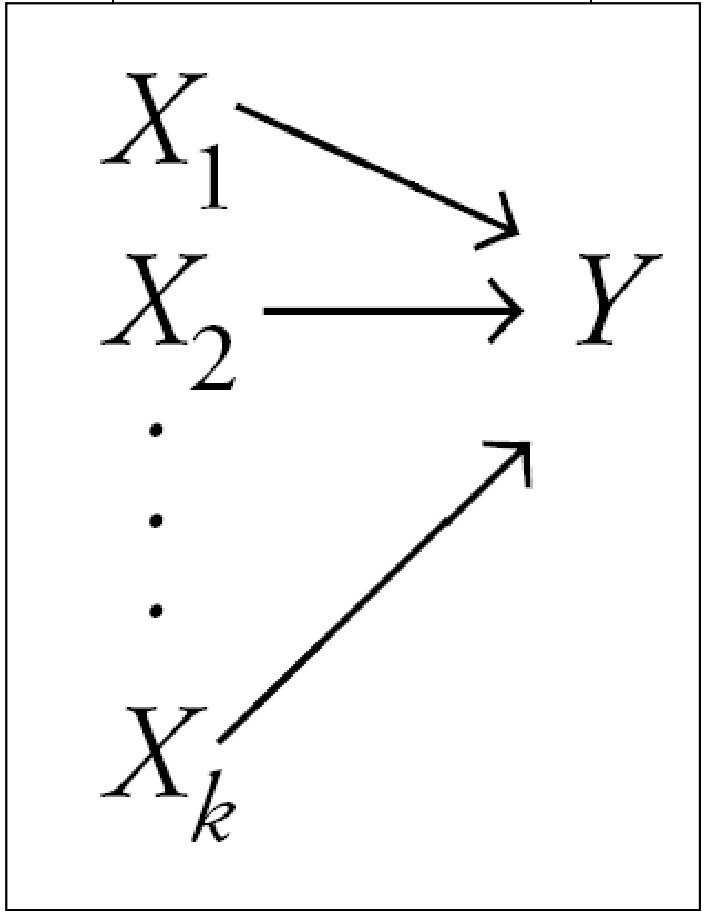
Examples:

- The selling price of a house can depend on the desirability of the location, the number of bedrooms, the number of bathrooms, the year the house was built, the square footage of the lot and a number of other factors.
- The height of a child can depend on the height of the mother, the height of the father, nutrition, and environmental factors.

Simple regression considers the relation between a single explanatory variable and response variable

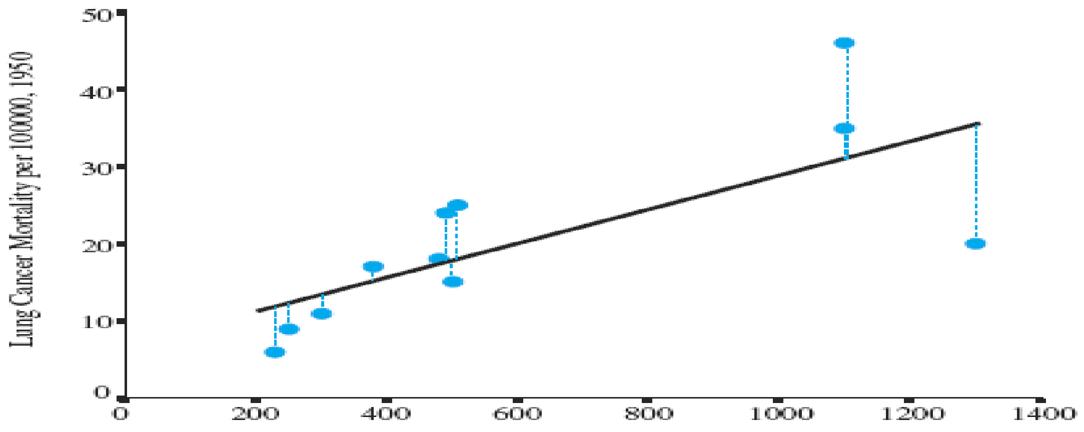


Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the [linear relationship]() between the explanatory (independent) variables and response (dependent) variable



Regression Modeling

- A simple regression model (one independent variable) fits a regression *line* in 2-dimensional space
- A multiple regression model with two explanatory variables fits a regression plane in *3*-dimensional space



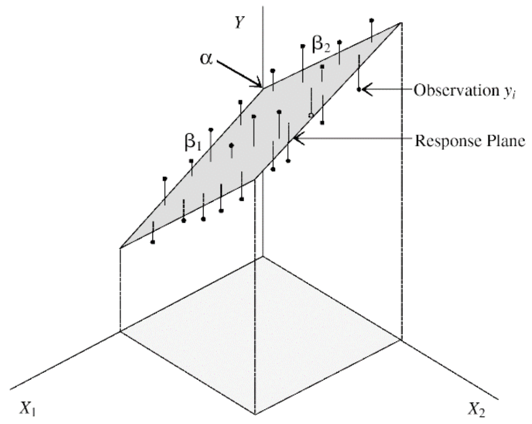
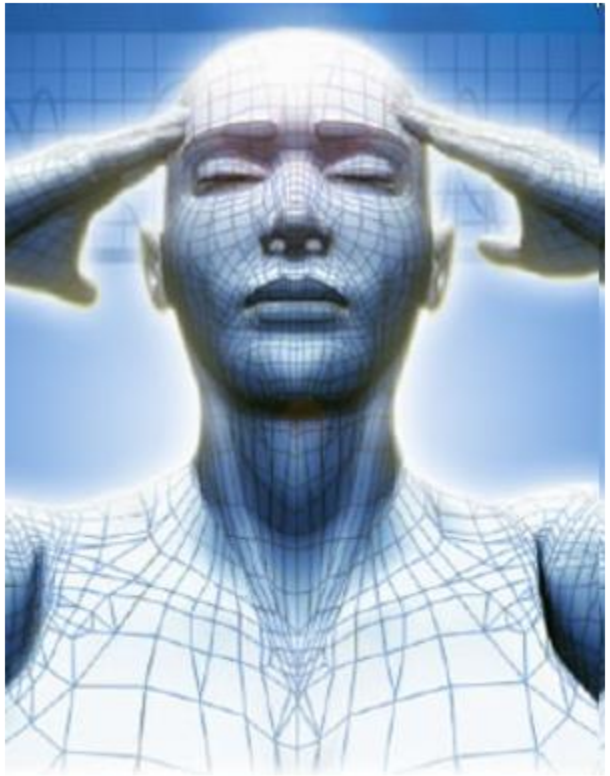


FIGURE 15.1 Three-dimensional response plane.

Multiple Regression Model

A multiple regression model with *k* independent variables fits a regression "surface" in k + 1 dimensional space (cannot be visualized)



Earlier simple regression example

Sr No	Height (cms)	Weight (Kg)
1	135	57
2	165	70
3	155	63
4	160	65
5	150	62….

Multiple Linear Regression

- More than one input /explanatory variables (features): say $x_1, x_2, \ldots, x_n$
- Number of training examples : say m
- One dependent (output)/response variable : y

Sr No	Height in cms $\boldsymbol{x_1}$	Age in years $\boldsymbol{x_2}$	Activity profile (Calorie burned) in Kcal $\boldsymbol{x_3}$	Food (K Calories) $\boldsymbol{x_4}$	Weight in kg y
1	158	20	2.5	2.1	52
2	150	39	2.1	2.5	70
3	170	45	1.8	2.0	60
4	165	56	1.7	2.8	80

Example

$m = 4; n = 4$

x_{ij} denotes value of input variable (feature) x_i for jth training example.

First suffix is for index of input variable and second suffix is for the index of training example

x_{13} means value of first input variable (feature) for third training example , in this example height of person number 3

x_{31} means value of third input variable (feature) for first training example , in this example activity profile of person number 1

Matrix form

Age of person in fourth training example is represented as $x_{??}$

What is the value of x_{21} ? Describe x_{21} in words.

Value of -------- variable in -------- training example.

What is vector 2nd training example ?

What does x_{42}stand for? Describe in words pertaining to this example

Examples: $m = 4.$

x_0	Size (feet²) x_1	Number of bedrooms x_2	Number of floors x_3	Age of home (years) x_4	Price ($1000) y
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178

$$\underbrace{X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix}}_{m \times (n+1)} \qquad \underbrace{y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}}_{m\text{-dimensional vector}}$$

$$\boxed{\theta = (X^T X)^{-1} X^T y}$$

Vector $[x_{1i}, x_{2i},\ldots, x_{ni}]$ represents i^{th} training example

Vector $[x_{j1}, x_{j2},\ldots, x_{jm}]'$ represents vector of values of j^{th} input variable

$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x}_3$	....	$\boldsymbol{x_n}$	$\boldsymbol{y}$
x_{11}	x_{21}	x_{31}	...	x_{n1}	y_1
x_{12}	x_{22}	x_{32}	...	x_{n2}	y_2
...	...	...	...	...	...
x_{1i}	x_{2i}	x_{3i}	...	x_{ni}	y_i
...	...	...	...	...	...
x_{1m}	x_{2m}	x_{3m}	...	x_{nm}	y_m

Hyper plane to be fitted

- $y(x_1, x_2, \dots, x_n) = \; b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$ using least square technique
- $y = \; b_0 + \sum_1^n b_j x_j$
- $y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \ldots + b_n x_{ni}$for $i = 1,2, \dots, m$
- $y = \; b_0 + \sum_1^n b_j x_j$

Matrix form to be fitted

- For convenience of writing the expressions introduce zeroth variable x_0; which always takes the value 1
- Hypothesis function is $y = y(x_1, x_2, \dots, x_n) = b_0x_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$;
- $\boldsymbol{b} = \begin{pmatrix} b_0 \\ b_1 \\ \cdots \\ b_n \end{pmatrix} \epsilon\ \mathcal{R}^{n+1}$, $\boldsymbol{x} = \begin{pmatrix} x_0 \\ x_1 \\ \cdots \\ x_n \end{pmatrix} \epsilon\ \mathcal{R}^{n+1}$; $\boldsymbol{b^T} = (b_0, b_1, \dots, b_n)$ is 1 x (n+1) vector and $\boldsymbol{x}$ is (n+1)x1 vector
- $\boldsymbol{b^T x}$ is 1x1 vector : $= (b_0, b_1, \dots, b_n) \cdot \begin{pmatrix} x_0 \\ x_1 \\ \cdots \\ x_n \end{pmatrix}$
- = $b_0x_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$
- y = $\boldsymbol{b^T}\ \boldsymbol{x}$ is the hypothesis function

$\boldsymbol{x_0}$	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$	. . . .	$\boldsymbol{x_n}$	$\boldsymbol{y}$
1	x_{11}	x_{21}	x_{31}	. . .	x_{n1}	y_1
1	x_{12}	x_{22}	x_{32}	. . .	x_{n2}	y_2
1	. . .	. . .	. . .	. . .	. . .	. . .
1	x_{1i}	x_{2i}	x_{3i}	. . .	x_{ni}	y_i
1	. . .	. . .	. . .	. . .	. . .	. . .
1	x_{1m}	x_{2m}	x_{3m}	. . .	x_{nm}	y_m

Predicted Values

- $b_0 x_{01} + b_1 x_{11} + b_2 x_{21} + \ldots + b_n x_{n1} = \widehat{y_1}$
- $b_0 x_{02} + b_1 x_{12} + b_2 x_{22} + \ldots + b_n x_{n2} = \widehat{y_2}$
- . . . .
- $b_0 x_{0i} + b_1 x_{1i} + b_2 x_{2i} + \ldots + b_n x_{ni} = \widehat{y_i}$
- ………..
- $b_0 x_{0m} + b_1 x_{1m} + b_2 x_{2m} + \ldots + b_n x_{nm} = \widehat{y_m}$

- $b_0 x_{0i} + b_1 x_{1i} + b_2 x_{2i} + \ldots + b_n x_{ni} = \widehat{y_i}\ for i = 1,2, \ldots, m$
- $\sum_j b_j\, x_{ji} = \widehat{y_i}\ for i = 1,2, \ldots, m; j = 0,1, \ldots, n$

Least Square method

- $e_i = y_i - \widehat{y_i} \ = y_i - \sum_j b_j \, x_{ji} \ \ for \ i = 1,2, \dots, m$
- ${e_i}^2 = (y_i - \widehat{y_i})^2 = (y_i - \sum_j b_j \, x_{ji})^2 \ for \ i = 1,2, \dots, m$
- SSE $= \sum_i {e_i}^2 = \sum_i (y_i - \sum_j b_j \, x_{ji})^2$
- Observations in n+1 dimensional space
- For least square fit, need to find b_j for $j = 0,1, \dots, n$ so that SSE is minimum. Denote SSE as S.
- Partial derivatives of S with respect to each of the parameters must be zero.

$\frac{\partial S}{\partial b_j} = 0$; for $j = 0,1, \dots, n$

Normal Equations

- $\frac{\partial S}{\partial b_j} = 0$; for $j = 0,1, \dots, n$
- $S = \sum_i (y_i - \sum_j b_j x_{ji})^2$;
- $\frac{\partial S}{\partial b_j} = \frac{\partial}{\partial b_j} \sum_i (y_i - \sum_j b_j x_{ji})^2$
- $\frac{\partial S}{\partial b_j} = \sum_i \frac{\partial}{\partial b_j} (y_i - \sum_j b_j x_{ji})^2$
- $\frac{\partial S}{\partial b_j} = 2 \sum_i \left(y_i - \sum_j b_j x_{ji}\right) \left(-x_{ji}\right) = 0$ for $j = 0,1, \dots, n$
- **$\sum_i \left(y_i - \sum_j b_j x_{ji}\right) \left(x_{ji}\right) = 0$ for $j = 0, 1, \dots, n$ are n+1 simultaneous linear equations in b_j's for $j = 0, 1, \dots, n$ called normal equations**
- For example for $j = 0$:
- $\sum_i y_i = b_0 \sum_i x_{0i} + b_1 \sum_i x_{1i} + b_2 \sum_i x_{2i} + \dots + b_n \sum_i x_{ni}$
-

Normal Equations

- For example for $j = 0$:
- $\sum_i y_i = b_0 \sum_i x_{0i} + b_1 \sum_i x_{1i} + b_2 \sum_i x_{2i} + \ldots + b_n \sum_i x_{ni}$
- For $j = 1$
- $\sum_i x_{1i} y_i = b_0 \sum_i x_{1i} x_{0i} + b_1 \sum_i {x_{1i}}^2 + b_2 \sum_i x_{1i}\, x_{2i} + \ldots + b_n \sum_i x_{1i} x_{ni}$
- …………
- For $j = n$
- $\sum_i x_{ni} y_i = b_0 \sum_i x_{ni} x_{0i} + b_1 \sum_i x_{ni} x_{1i} + b_2 \sum_i x_{ni}\, x_{2i} + \ldots + b_n \sum_i {x_{ni}}^2$
- Solution of above equations

Normal Equations for two input variables

- $\sum_i y_i = b_0 \sum_i x_{0i} + b_1 \sum_i x_{1i} + b_2 \sum_i x_{2i}$
- $\sum_i x_{1i} y_i = b_0 \sum_i x_{1i} x_{0i} + b_1 \sum_i {x_{1i}}^2 + b_2 \sum_i x_{1i}\, x_{2i}$
- $\sum_i x_{2i} y_i = b_0 \sum_i x_{2i} x_{0i} + b_1 \sum_i x_{2i} x_{1i} + b_2 \sum_i {x_{2i}}^2$

i	x_0	x_1	x_2	${x_1}^2$	$x_1\, x_2$	${x_2}^2$	y	$x_1 y$	$x_2 y$
1	1								
2	1								
3	1								
……	1								
Total	m								

Illustration: (Home Work)

- Suppose that a random sample of five families yielded the following data (income in thousand dollars)
- Estimate the multiple regression equation of S on I and W

Solution: Savings dependent variable y

Input variables income x_1 and Assets x_2

Family	Saving S	Income I	Assets W
A	0.6	8	12
B	1.2	11	6
C	1.0	9	6
D	0.7	6	3
E	0.3	6	18