MULTIVARIATE LINEAR REGRESSION MODEL

MULTIVARIATE REGRESSION

* Standard multiple Regression with emphasis on detection of collinearity
* outliers
* Non-normality and autocorrelation
* Validation of Model assumptions
* Assumptions of Multivariate Regression model
* Parameter Estimation

Multivariate Regression is a method used to measure the degree at which more than one independent variable (Predictors) and more that one dependent variable (response) are linearly related. The method is broadly used to predict the behaviour of the dependent variables associated to changes in the independent variables once a desired degree of relation has been established.

Question 1: Can a Supermarket owner maintain stock of water, icecream, frozen foods, canned foods, meat as a function of temperature, tornado chance and gas price during tornado

Obvious Assumptions: If it is too hot, icecream sales increases; If it ~~so~~ a tornado hits water and canned foods sales increase while ice cream sales will decrease.

4 A mathematical model based on multivariate linear regression analysis will address this.

Multiple Regression Model

The multiple Regression model relates more than one predictor and response

Let y be $n \times 1$ dependent vector

X be an $n \times (q+1)$ matrix such that all the entries of the first columns are 1's and q predictors

The model is

$$\Rightarrow \boxed{\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1q} \\ 1 & x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nq} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}}$$

(ie) $y = X\beta + \epsilon$

ϵ is $n \times 1$ vector, $\epsilon_i \sim N(0, \sigma^2)$

β is $(q+1) \times 1$ vector

Multivariate Regression

The multivariate Regression model relates more than one predictor and more than one response.

Let $Y \rightarrow$ be $n \times p$ response matrix

$X \rightarrow$ be $n \times (q+1)$ matrix such that all entries of the first column are 1's and q predictions

$B \rightarrow$ be an $(q+1) \times p$ matrix of fixed parameters

$\epsilon \rightarrow$ $n \times p$ matrix such that $\epsilon \sim N(0, \Sigma)$ (Multivariate normally distributed with covariance matrix Σ).

The model is as follows

$$\boxed{Y = X\beta + \epsilon}$$

$$
\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & & & \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1q} \\ 1 & x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & & & \\ 1 & x_{n1} & x_{n2} & & x_{nq} \end{pmatrix} \times
$$

$$
\begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0p} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1p} \\ \vdots & & & \\ \beta_{q1} & \beta_{q2} & \cdots & \beta_{qp} \end{pmatrix} + \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1p} \\ \epsilon_{21} & \epsilon_{22} & \cdots & \epsilon_{2p} \\ \vdots & & & \\ \epsilon_{n1} & \epsilon_{n2} & \cdots & \epsilon_{np} \end{pmatrix}.
$$

Hence

Multiple linear Regression model.

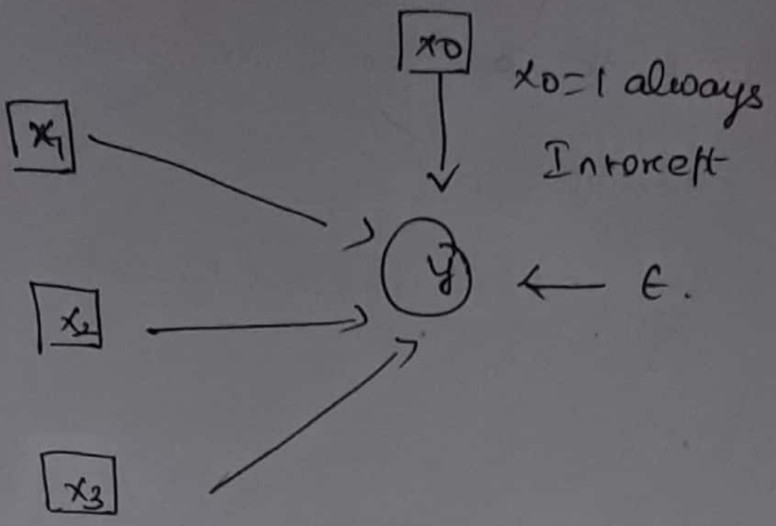
$$
\boxed{Y = X_1\beta_1 + X_2\beta_2 + \cdots + X_K\beta_K + \epsilon}
$$

The parameters $\beta_1, \beta_2 \cdots \beta_K$ are the regression coefficients associated with $X_1 X_2 \cdots X_K$ respectively and ϵ is the random error component reflecting the difference between the observed and fitted linear relation relationship.

Note: j^{th} regression coefficient

$$
\beta_j = \frac{\partial E(y)}{\partial X_j}
$$

Linear Model Representation



$x_1 =$ % absenteeism.

$x_2 \rightarrow$ Breakdown hrs

$x_3 \rightarrow$ M-ratio

$$Y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + + \beta_3 x_3 + \epsilon .$$

$$x = \begin{bmatrix} x_0 \\ \hline x_1 \\ \vdots \\ x_p \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon .$$

Lineal model

A model is said to be lineal when it is lineal in Paramerers.

In such cases $\frac{\partial y}{\partial \beta_j}$ $\left(\frac{\partial E(y)}{\partial \beta_j}\right)$ should not depend on any β's.

1) $y = \beta_0 + \beta_1 x$ → lineal model

2) $y = \beta_0 x^{\beta_1}$ → lineal model

[Since $\log y = \log \beta_0 + \beta_1 \log x$

$y^* = \beta_0^* \beta_1 x^*$]

3) $y = \beta_0 + \beta_1 x + \beta_2 x^2$ lineal in parameters but nonlineal in variable x

→ Lineal model.

4) $y = \beta_0 + \frac{\beta_1}{x - \beta_2}$ → Nonlineal in Paramerees and variable

∴ Non lineal model.

5) $y = \beta_0 + \beta_1 x^{\beta_2}$ → non lineal in Parameters and variable

∴ Non lineal model.

vi) $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$.

$= \beta_0 + \beta_1 x + \beta_2 x_2 + \beta_3 x_3$.

Model set up:

Let an experiment be conducted n times and the data is obtained as follows:

Observation no.	Response	Explanatory variables x_1 x_2 ... x_k.
1	y_1	x_{11} x_{12} ... x_{1k}
2	y_2	x_{21} x_{22} ... x_{2k}
$\vdots$	$\vdots$	$\vdots$ $\vdots$ $\ddots$ $\vdots$
n	y_n	x_{n1} x_{n2} ... x_{nk}.

$\therefore y = \beta_0 + \beta_1 x + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon$

Thus the n-tuples of observations are also assumed to follow the same model.

Thus $y_1 = \beta_0 + \beta_1 x_{11} + \dots + \beta_k x_{1k} + \epsilon_1$

$y_2 = \beta_0 + \beta_1 x_{21} + \dots + \beta_k x_{2k} + \epsilon_2$

$\vdots$

$y_n = \beta_0 + \beta_1 x_{n1} + \dots + \beta_k x_{nk} + \epsilon_n$.

Scanned by CamScanner

or $\boxed{\underset{n\times 1}{y} = \underset{n\times(p+1)}{X}\underset{(p+1)\times 1}{\beta} + \epsilon}$ (Regression coefficient.

If intercept term is present, take first column of X to be $(1, 1, \ldots 1)$

Assumptions in Multiple Linear Regression Model.

(1) $E(\epsilon) = 0$

(2) $E(\epsilon\epsilon') = \sigma^2 I_n$

(3) Rank $(X) = k$

(4) X is a non-stochastic matrix

(5) $\epsilon \sim N(0, \sigma^2 I_n)$.

Estimation of Parameters:

A general Procedure for the estimation of regression coeff. vector is to minimize.

$$\sum_{i=1}^{n} M(\epsilon_i) = \sum_{i=1}^{n} M(y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - \cdots x_{ik}\beta_k)$$

for a suitably chosen function M.

Some examples of choice of M are

$$M(x) = |x|$$

$$M(x) = x^2$$

$$M(x) = |x|^p \text{ in general}$$

We consider the principle of least square which is related to $M(x) = x^2$ and method of ML estimation for the estimation of parameters.

model befor

Ex Let n data points are to be collected.

$\boxed{DV}$ $\boxed{IV's}$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad X_{n\times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & \cdots & & x_{2p} \\ & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}.$$

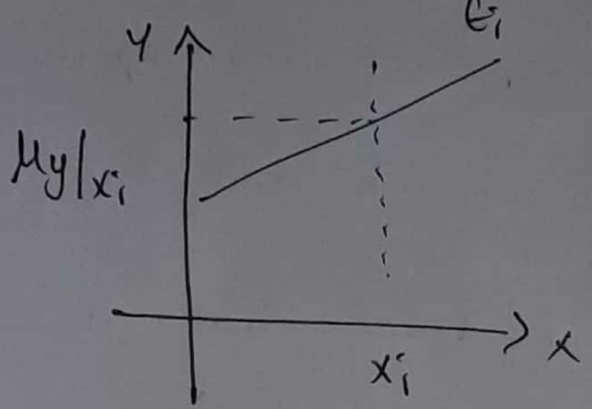
$$y_i = \underbrace{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}}_{\text{Variates.}} + \epsilon_i$$

$$y_i = E[y_i \mid x_i = x_{i1}, x_{i2} \cdots x_{ip}) + \epsilon_i$$

$$\downarrow$$

$$\hat{y}_i + \epsilon_i$$

$$\therefore \boxed{\epsilon_i = y_i - \hat{y}_i}$$



$$\boxed{Y = X\beta + \epsilon}$$

Y: $n \times 1$

X: $n \times (p+1)$ Data matrix / Design matrix

β: $(p+1) \times 1$ Regression coefficient

ϵ: $n \times 1$ error terms

Assumptions:

1. Linearity
2. Equal y variance across the value of x
3. Uncorrelated error terms
4. Normality of the error.

Scanned by CamScanner

Outliers:

Outliers are points that fall away from the cloud of points.

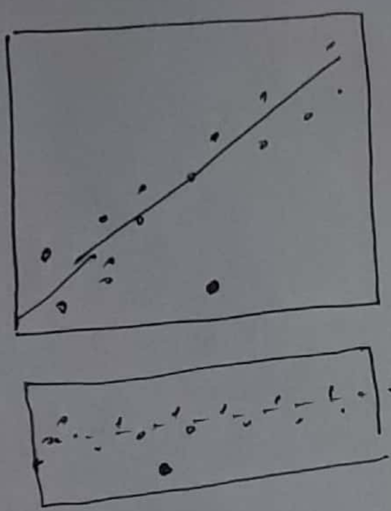
* outliers that fall horizontally away from the centre of the cloud are called the leverage points.

* High leverage points that actually influence the slope of the regression line are called influential points.

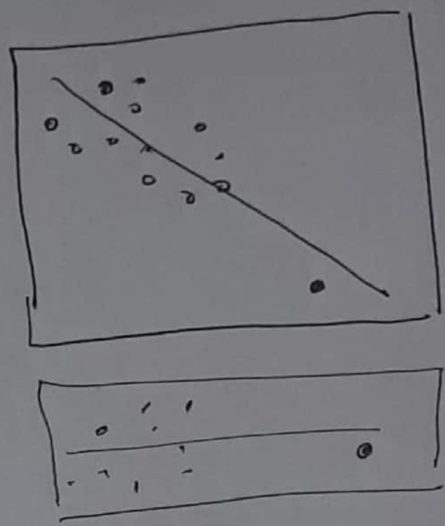
Note: In order to determine if a point is influential, visualize the regression line with and without that points. If the slope of the line change considerably the point is influential, if not then it is not.

Q1 : There are six plots shown in Figure along with the least squares lines and residual plots. For each scatter plot and residual plot pair, identify the outliers and note how may influence the least squares line

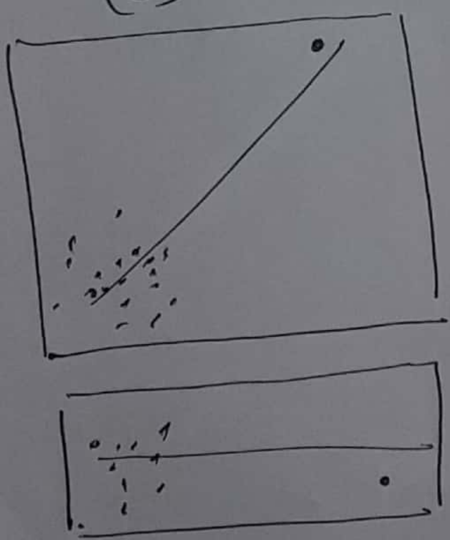
(1)



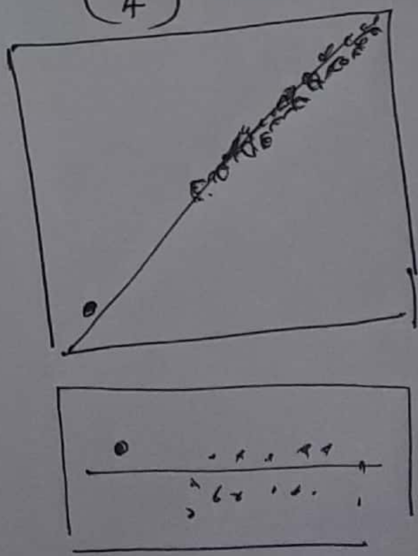
(2)



(3)



(4)



In most of the situations the efficiency of the method goes down, If multi collinearity is present, the estimated values of the coefficients are very sensitive to slight changes in the data and those coefficients will have large sampling errors which may affect the prediction or Estimation. The condition of nonorthogonality is known as the problem of collinear data or multi collinearity.

Indication of multi collinearity that appears as instability in the estimated coefficients are

- large changes in the estimated coefficients when a variable is added or deleted
- large changes in the coefficients when the data points are altered or dropped.

The coefficients of the variables tend to have large standard error.

The indications of multicollinearity can be obtained using std. regression computation.

Corrolation coefficient

Let the Population correlation coefficient matrix be the $p \times p$ symmetric matrix

$$\rho = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} \\ \vdots & & & \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & & \rho_{2p} \\ \cdots & \cdot & & \cdot \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}$$

Q: suppose $\Sigma = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix}$

Obtain correlation coefficient

Soln:

Given $\Sigma = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix}$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}$$

$$V^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & 0 \\ 0 & \sqrt{\sigma_{22}} & 0 \\ 0 & 0 & \sqrt{\sigma_{33}} \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

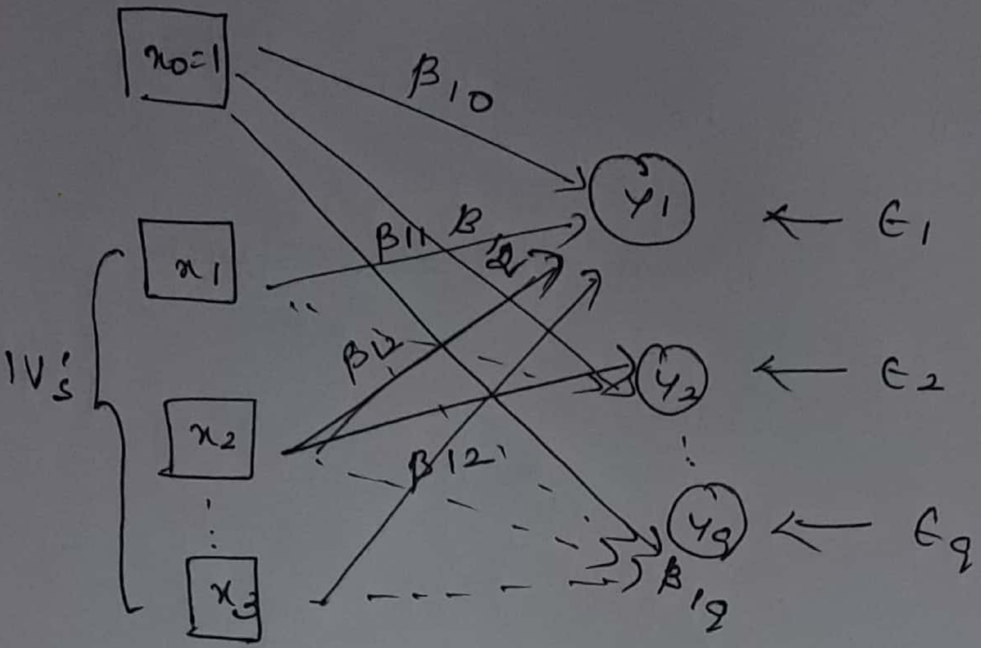
$$(V^{1/2})^{-1} = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/5 \end{bmatrix}$$

$$P = (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1}$$

$$= \begin{bmatrix} 1 & 1/6 & 1/5 \\ 1/6 & 1 & -1/5 \\ 1/5 & -1/5 & 1 \end{bmatrix}$$

Multivariate Linear Regreression:

Intercept



$$Y_1 = \beta_{11} + \beta_{11} X_1 + \beta_{12} X_2 + \cdots + \beta_{1p} X_p + \epsilon_1 \quad \text{---(1)}$$

$$Y_2 = \beta_{21} + \beta_{22} X_1 + \beta_{23} X_2 + \cdots + \beta_{2p} X_p + \epsilon_2 \quad \text{---(2)}$$

MLR

DV → 1

IV's = P

MVLR

DV → 1 , DV = q

IV's = P.

MVLR $\quad P = \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1p} \end{bmatrix}_{(P+1)\times 1} \qquad \beta = \begin{bmatrix} \beta_{10} & \beta_{20} \\ \beta_{11} & \beta_{21} \\ \vdots & \vdots \\ \beta_{1p} & \beta_{2q} \end{bmatrix}_{(P+1)\times q}$

⋮

$$Y_q = \beta_{q0} + \beta_{q1} X_1 + \beta_{q2} X_2 + \cdots + \beta_{qp} X_p + \epsilon_p \quad \text{---(3)}$$

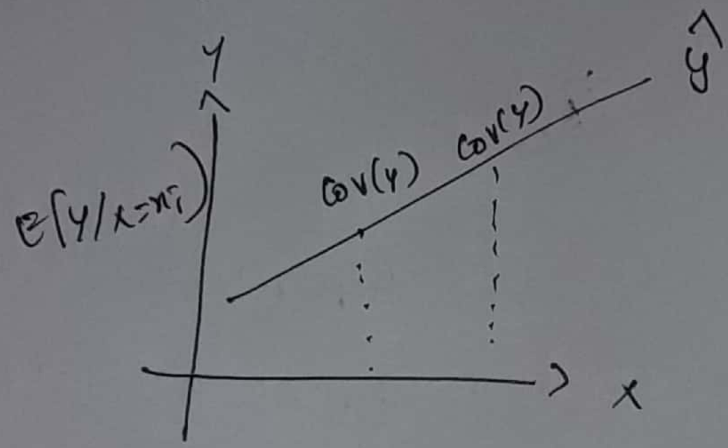
Assumptions

1) Errors ($E_{n \times q}$) are multivariate normal

2) Error variances are equal (homogeneous)

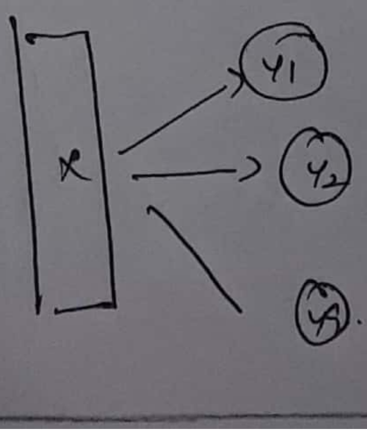
3) Errors have common covariance structure

4) Independent observations.



$$\text{Cov } Y = \Sigma_{q \times q}$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1q} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2q} \\ \vdots & & & \\ \sigma_{1q} & \sigma_{2q} & \cdots & \sigma_q^2 \end{bmatrix}$$



$$\text{Cov}(E) = \Sigma$$

Multivariate Linear Regression (MVLR)

Estimation of Parameters

$$Y = X\beta + \epsilon$$

Y: $n \times q$, X: $n \times (p+1)$, β: $(p \times 1) \times q$, ϵ: $n \times q$

$$\hat{\beta}_1, \hat{\beta}_2 \dots \hat{\beta}_q \longrightarrow \hat{\beta} = (X^T X)^{-1} X^T Y$$

Example consider $Y_{3\times 2} = \begin{bmatrix} 10 & 100 \\ 12 & 110 \\ 11 & 105 \end{bmatrix}$

IVs: $p = 2$ $\quad X = \begin{bmatrix} 9 & 62 \\ 8 & 58 \\ 7 & 64 \end{bmatrix}$

$$X_{3 \times p+1} = \begin{bmatrix} 1 & 9 & 62 \\ 1 & 8 & 58 \\ 1 & 7 & 64 \end{bmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

Step 1: compute $X^T X$

$$\begin{bmatrix} 1 & 1 & 1 \\ 9 & 8 & 7 \\ 62 & 58 & 64 \end{bmatrix} \begin{bmatrix} 1 & 9 & 62 \\ 1 & 8 & 58 \\ 1 & 7 & 64 \end{bmatrix}$$

$$x^T x = \begin{bmatrix} 3 & 24 & 184 \\ 24 & 194 & 1470 \\ 184 & 1470 & 11304 \end{bmatrix}$$

Step 2:

$$(x^T x)^{-1} = \frac{1}{|x^T x|} \operatorname{Adj}(x^T x)$$

$$= \begin{bmatrix} 320.76 & -8.16 & -4.16 \\ -8.16 & 0.56 & 0.06 \\ -4.16 & 0.06 & 0.06 \end{bmatrix}$$

Step 3:

$$x^T y = \begin{bmatrix} 1 & 1 & 1 \\ 9 & 8 & 7 \\ 62 & 58 & 64 \end{bmatrix} \begin{bmatrix} 10 & 100 \\ 12 & 110 \\ 11 & 105 \end{bmatrix}$$

$$= \begin{bmatrix} 33 & 315 \\ 263 & 2515 \\ 2020 & 19300 \end{bmatrix}$$

Step 4:

$$\hat{\beta} = (x^T x)^{-1} x^T y$$

$$= \begin{bmatrix} 35.80 & 229 \\ -0.80 & -4.0 \\ -0.30 & -1.50 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 : \hat{\beta}_2 \end{bmatrix}$$

$$-x^T y + x^T x \beta = 0$$

$$x^T x \beta = x^T y$$

$$(x^T x)^{-1} (x^T x)\beta = (x^T x)^{-1} x^T y$$

$$\boxed{\hat{\beta} = (x^T x)^{-1} x^T y}$$

formula for estimation of parameters)

Prob:

If $y = \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix}_{5\times 1}$ $x = \begin{bmatrix} 1 & 5 \\ 1 & 7 \\ 1 & 10 \\ 1 & 12 \\ 1 & 20 \end{bmatrix}_{5\times 2}$

Soln $\hat{\beta} = (x^T x)^{-1} x^T y$

Step 1 $x^T x = \underset{2\times 5}{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 7 & 10 & 12 & 120 \end{bmatrix}} \begin{bmatrix} 1 & 5 \\ 1 & 7 \\ 1 & 10 \\ 1 & 12 \\ 1 & 20 \end{bmatrix}$

$$x^T x = \begin{bmatrix} 5 & 54 \\ 54 & 718 \end{bmatrix}$$

Step 2: $(x^T x)^{-1} = \frac{1}{|x^T x|} \operatorname{adj}(x^T x)$

$$(X^T X)^{-1} = \frac{1}{874}\begin{bmatrix} 718 & -54 \\ -54 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} 1.07 & -0.0801 \\ -0.0801 & 0.007 \end{bmatrix}$$

Step 3: compute $(X^T Y)$

$$X^T Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 7 & 10 & 12 & 20 \end{bmatrix}\begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix}$$

$$= \begin{bmatrix} 150 \\ 1970 \end{bmatrix}$$

Step 4: $\hat{\beta} = (X^T X)^{-1} X^T Y$

$$= \begin{bmatrix} 1.07 & -0.0801 \\ -0.0801 & 0.007 \end{bmatrix}\begin{bmatrix} 150 \\ 1970 \end{bmatrix}$$

$$= \begin{bmatrix} 2.70 \\ 1.775 \end{bmatrix}_{2\times 1}.$$

$$y = \beta_0 + \beta_1 x_1 + \epsilon$$

$$\boxed{y = 2.70 + 1.775\, x_1 + \epsilon.}$$

$$\boxed{\hat{y} = 2.70 + 1.775 x_1}$$

$\hat{\epsilon} = y - \hat{y}$

Note

$y = f(x)$

$= \beta_0 + \beta_1 x_1 + \epsilon_0$

→ Simple Regression

$\beta + 1 \geq 3$ → Multiple regression.

Scanned by CamScanner

$$= \begin{bmatrix} \hat{\beta}_{10} & \hat{\beta}_{20} \\ \hat{\beta}_{11} & \hat{\beta}_{21} \\ \hat{\beta}_{12} & \hat{\beta}_{22} \end{bmatrix}$$

$$Y = X\beta + \epsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_{0} & \hat{\beta}_{20} \\ \hat{\beta}_{11} & \hat{\beta}_{21} \\ \hat{\beta}_{12} & \hat{\beta}_{22} \end{bmatrix} + \epsilon$$

$$y_1 = \hat{\beta}_{10} + \hat{\beta}_{11} x_1 + \hat{\beta}_{12} x_2 + \epsilon_1$$

$$y_2 = \hat{\beta}_{20} + \hat{\beta}_{21} x_1 + \hat{\beta}_{22} x_2 + \epsilon_2$$

βH

$$\boxed{\begin{aligned} y_1 &= 35.8 - 0.80x_1 - 0.30x_2 + \epsilon_1 \\ y_2 &= 229 - 4.0x_1 - 1.50x_2 + \epsilon_2 \end{aligned}}$$

βH

$y_1 \rightarrow$ Profit , $y_2 \rightarrow$ sales.

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

$$\hat{y} = X\hat{\beta} + \hat{\epsilon}$$

Fitte values.

$$\hat{e} = y - \hat{y}$$

$$= \begin{bmatrix} 10 & 100 \\ 12 & 110 \\ 11 & 105 \end{bmatrix} - \hat{y}$$

$$\hat{y} = X\hat{\beta} = \begin{bmatrix} 1 & 9 & 62 \\ 1 & 8 & 58 \\ 1 & 7 & 64 \end{bmatrix} \begin{bmatrix} 52.8 & 229 \\ -0.8 & -4.0 \\ -0.3 & -1.50 \end{bmatrix}$$

$$= \begin{bmatrix} 10 & 100 \\ 12 & 110 \\ 12 & 105 \end{bmatrix}$$

$$\therefore \hat{e} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Note : sample distribution

$$\boxed{s_{k}^{2} = (n-p-1)\,\hat{\sigma}_{k}^{2}}$$

$$= (n-p-1) \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12}^2 & \cdots & \hat{\sigma}_{1q}^2 \\ \vdots & & & \\ \hat{\sigma}_{1q} & \cdots & & \hat{\sigma}_k^2 \end{bmatrix}$$

$Cov(\epsilon_x, \epsilon_j) = 0$

$Cov(\epsilon_x, \epsilon_x) = \sigma_x^2$

Estimation of Parameters.

$Y = X\beta + \epsilon$ observed values

$\hat{Y} = X\hat{\beta}$

fitted values

Estimated β.

$\hat{\beta}$ MLR

$(p+1) \times q$ $\beta (p+1) \times 1$

$Y = [Y_1 : Y_2 \cdots : Y_q]$

$= X[\beta_1 : \beta_2 \cdots \beta_q] + [\epsilon_1 : \epsilon_2 : \cdots : \epsilon_q]$

$Y_1 = X\beta_1 + \epsilon_1 \rightarrow$ MLR-1

$Y_2 = X\beta_2 + \epsilon_2 \rightarrow$ MLR-2

$\vdots$

$Y_q = X\beta_q + \epsilon_q \leftarrow$ MLR-q.

$\boxed{MLR} \rightarrow y_1 = x\beta_1 + \epsilon_1$

$$\epsilon_1 = y_1 - x\beta_1$$

$$\epsilon = y - x\beta.$$

$$SSE = \epsilon^T\epsilon = (y - x\beta)^T(y - x\beta)$$

$$\frac{\partial SSE}{\partial \beta} = 0$$

$$\hat{\beta}_1 = (x^T x)^{-1} x^T y_1$$

$\boxed{MVLR}$

$$\underset{n\times q}{y} = \underset{n\times(p+1)}{x}\,\underset{(p+1)\times q}{\beta} + \underset{n\times q}{\epsilon}$$

$$\epsilon = y - x\beta$$

$$\underbrace{\underset{q\times n}{\epsilon^T} \cdot \underset{n\times q}{\epsilon}}_{q\times q} = \begin{bmatrix} \sum_{i=1}^{n}\epsilon_{i1}^2 & \sum \epsilon_{i1}\epsilon_{i2} & \cdots & \\ & \sum_{i=1}^{n}\epsilon_{i2}^2 & & \\ & & \ddots & \\ & & & \sum_{i=1}^{n}\epsilon_{ik}^2 \end{bmatrix}$$

$$\underbrace{\text{trace}}_{\text{minimize.}}(SSCP) = tr(\epsilon^T\epsilon).$$

$$\hat{\beta} = (X^T X)^{-1} X^T [y_1 : y_2 : \cdots y_q]$$

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

<u>MVLR</u>

$$\hat{\beta} = [\hat{\beta}_1 : \hat{\beta}_2 \cdots : \hat{\beta}_q]$$

$$= \begin{bmatrix} \hat{\beta}_{10} & \hat{\beta}_{20} & \cdots & \hat{\beta}_{q0} \\ \hat{\beta}_{11} & \hat{\beta}_{21} & & \beta_{q1} \\ \hat{\beta}_{1p} & \hat{\beta}_{2p} & & \hat{\beta}_{qp} \end{bmatrix}$$

$$= (X^T X)^{-1} X^T \begin{bmatrix} y_1 : y_2 \cdots : y_q \end{bmatrix}$$