UNIT - II

Multiple Regression

A regression models that involves more than one regressor variable (independent variable) is called a multiple regression model.

A multiple regression model is

$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon_i \longrightarrow (1)$. This is a multiple linear regression model with two independent variables. The term linear is used because there is a linear function of the unknown parameters β_0, β_1 and β_2

Where * β_0 is the intercept of the regression plain * β_1 is the expected change in y per unit change in x_1 ~~and~~ when x_2 is held constant and * β_2 is the expected change in y per unit change in x_2 when x_1 is held constant. * ϵ_i is the error term.

~~The~~ General, the response variable y may be related to, ~~In gen~~ k regressor variables. the model is

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon_i \longrightarrow (2)$$

is called a multiple linear regression model with k regressor variables. The parameters β_j, $j = 0, 1, 2, \dots k$ are called the regressor co-efficients. This model describes a hyper plain in the k dimentional space of the regressor variable x_j. The parameter β_j represents the expected change in the response y

22/02/20

Per unit changing x_j when all of the remaining regressor variables x_i $(i \neq j)$ are constant. For these reason the parameters β_j $(j=1,2,\ldots k)$ are often called partial regression co-efficients. For example, consider the regression model $y = 50 + 10x_1 + 7x_2 + 5x_3 + \varepsilon_i$ → (3)
in this model the variable x_1 is highly influence the reponse variable. In this situation we will no bother about the sign of the co-efficients.

Model that of more complex in structure in equ (2) may often still analyzed by multiple linear regression techniques for example, consider the cubic polynomial model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon_i \quad \text{——} \quad (4)$$

If we assume $x_1 = x$, $x_2 = x^2$ and $x_3 = x^3$. then equ (4) can be rewritten as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon_i \quad \text{——} \quad (5)$$

which is a multiple regression model with 3 regressor variables.

IIIly, models that include interaction effects may also we analyze by multiple linear regression models. For example, Suppose that the model is

$$y_0 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon_i \quad \text{——} \quad (6)$$

If we assume $x_3 = x_1 x_2$ and $\beta_3 = \beta_{12}$ then the equ (6) can be rewritten as

$$y_0 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon_i \quad \text{——} \quad (*)$$

Aaigo which is multiple linear regression model with 8 regressor variable.

Finally any regression model is linear in the parameters the β's is the linear regression model regardless of the shape of the surface that it generalized

The Gauss - Markov Theorem, Var (ε) . V

The Gauss - Markov theorem establishes that the generalized least - squares (GLS) estimator of β. $\hat{\beta} = (x'V^{-1}x)^{-1} x'V^{-1}y$, is BLUE (Best Linear Unbiased estimator). Once again by best, we mean $\hat{\beta}$ minimizes the variance for any linear combination of the estimated co-efficients, $\ell'\hat{\beta}$.

We note that if the model is correct

$$E\left[(x'V^{-1}x)^{-1} x'V^{-1}y\right] = (x'V^{-1}x)^{-1} x'V^{-1} E(y)$$

$$= (x'V^{-1}x)^{-1} x'V^{-1} x\beta = \beta.$$

Thus $(x'V)^{-1} x'V^{-1} y$ is an unbiased estimator of β. The Variance of this estimator is,

$$V\left[(x'V^{-1}x)^{-1} x'V^{-1}y\right] = \left[(x'V^{-1}x)^{-1} x'V^{-1}\right] V(y), (x'V^{-1}x)^{-1} x'V^{-1}]$$

$$= \left[(x'V^{-1}x)^{-1}\right] V \left[(x'V^{-1}x)^{-1} x'V^{-1}\right]'$$

$$= \left[x'V^{-1}x'V^{-1}\right] V \left[V^{-1}x(x'V^{-1}x)^{-1}\right]$$

$$= (x'V^{-1}x)^{-1}$$

Thus, $Var(\ell'\hat{\beta}) = \ell' \, var(\hat{\beta}) \, \ell$

$$= \ell \left[(x'V^{-1}x)^{-1}\right] \ell$$

Let $\tilde{\beta}$ be another unbiased estimator of β that is a linear combination of the data. Our goal, then is to show that $\text{Var}(L'\tilde{\beta}) \geq L'(X'V^{-1}X)^{-1}L$ with at least one L such that $\text{Var}(L'\tilde{\beta}) \geq L'(X'V^{-1}X)^{-1}L$

We first note that we can write any other estimator of β that is a linear combination of the data as,

$$\tilde{\beta} = [(X'V^{-1}X)^{-1}X'V^{-1} + B]y + b_0$$

Where, B is an $p \times n$ matrix and b_0 is a $p \times 1$ vector of constant that appropriately adjusts the GLS estimator to from the alternative estimate. We next note that if the model is correct, then

$$E(\tilde{\beta}) = E\{[(X'V^{-1}X)^{-1}X'V^{-1} + B]y + b_0\}$$

$$= [(X'V^{-1}X)^{-1}X'V^{-1} + B]E(y) + b_0$$

$$= [(X'V^{-1}X)^{-1}X'V^{-1} + B]X\beta + b_0$$

$$= [X'V^{-1}X]^{-1}X'V^{-1}X\beta + BX\beta + b_0$$

$$= \beta + BX\beta + b_0$$

consequently, $\tilde{\beta}$ is unbiased if and only if both $b_0 = 0$ and $BX = 0$ the $\text{Var}(\tilde{\beta})$ is

$$V(\tilde{\beta}) = V\{[(X'V^{-1}X)^{-1}X'V^{-1} + B]y\}$$

$$= [(X'V^{-1}X)^{-1}X'V^{-1} + B]V(y)[(X'V^{-1}X)^{-1}X'V^{-1} + B]'$$

$$= [(X'V^{-1}X)^{-1}X'V^{-1} + B]V[(X'V^{-1}X)^{-1}X'V^{-1} + B]'$$

Properties of least square estimators:

The statistical properties of the least square estimator $\hat{\beta}$ may be easily explained

1. $E(\hat{\beta}) = E\left[(x'x)^{-1} x'y\right] = E\left[(x'x)^{-1} x'(x\beta + \varepsilon_i)\right]$

$= E\left[(x'x)^{-1} x'x\beta + (x'x)^{-1} x'\varepsilon_i\right] = \beta$

Since, $E(\varepsilon_i) = 0$, and $(x'x)^{-1} x'x = I$.

Thus $\hat{\beta}$ is an unbaised estimator of β

2. The variance property of $\hat{\beta}$ is expressed by the co-variance matrix

$$\operatorname{Cov}(\hat{\beta}) = E\left\{\left[\hat{\beta} - E(\hat{\beta})\right]\left[\hat{\beta} - E(\hat{\beta})\right]'\right\}$$

which is ~~the~~ a $p \times p$ symmetric matrix whose jth diagonal element is the variance of $\hat{\beta}_j$ ($\operatorname{Var}(\hat{\beta}_j)$) and whose ijth off-diagonal element is the co-variance between $\hat{\beta}_i$ and $\hat{\beta}_j$. Therefore the covariance matrix of $\hat{\beta}$ is

$$\operatorname{Cov}(\hat{\beta}) = \sigma^2 (x'x)^{-1}$$

If we assume $C = (x'x)^{-1}$ then

$$V(\hat{\beta}_j) = \sigma^2 C_{jj} \quad \text{and}$$

$$\operatorname{Cov}(\hat{\beta}_j) = \sigma^2 C_{ij}.$$

3. According to Gauss-Markov theorem the least square estimator $\hat{\beta}$ is the BLUE of β.

4. If we further assume that the errors ε_i are normally distributed then the least square

estimator $\hat{\beta}$ is also the Maximum likelihood estimator (MLE) of β.

5. The MLE of $\hat{\beta}$ is the minimum variance unbiased estimator of β.

Confidence Interval on regression co-efficients.

Confidence Intervals on individual regression co-efficient and confidence intervals on the mean response given specific levels of the regressor variables play the same important important role in multiple regression that they do in simple linear regression.

To construct confidence interval estimates for the regression co-efficients β_j; $j = 0, 1, 2, \cdots, k$, we will continue to assume that the errors ε_i are normally and idependently distributed with mean zero and variance σ^2.

Therefore the observations y_i are normally and independly distributed with mean $\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$ and variance σ^2. Since, the least square estimator $\hat{\beta}$ is the linear combination of the observations, and its follows that $\hat{\beta}$ is normally distributed with mean vector β and the co-variance matrix $cov(\hat{\beta}) = \sigma^2 (X'X)^{-1}$.

Finally, the test statistic is

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \quad \text{—— (1)}$$

is distributed with $(n-p)$ d.f at t-distribution

Where, $\hat{\sigma}^2$ is the estimates of the error variance

Based on the result given in equ ① we may define a $100(1-\alpha)\%$ confidence interval for the regression co-efficient β_j; $j = 0, 1, 2, \ldots, k$ as

$$\boxed{\hat{\beta}_j - t_{\alpha/2, n-p}\sqrt{\hat{\sigma}^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2, n-p}\sqrt{\hat{\sigma}^2 C_{jj}}}$$

→ ②.

that, we call that quantity $SE(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}}$ → ③

i.e., the S.E of regression co-efficient $\hat{\beta}_j$.

26/02/2020.

PREDICTION OF NEW OBSERVATION:

The regression model can be used to predict future observation on Y corresponding to particular values of the regressor variable, for example, $x_{01}, x_{02}, \ldots, x_{0k}$.

If $x_0' = [1, x_{01}, x_{02}, \ldots, x_{0k}]$, then the point estimate of the future observation Y_0 at the point $x_{01}, x_{02}, \ldots, x_{0k}$ is $\hat{Y}_0 = x_0'\hat{\beta}$

A $100(1-\alpha)$ percent prediction interval for this future observation is

$$\hat{y}_0 - t_{\alpha/2, n-p}\,\hat{\sigma}^2\sqrt{(1+x_0'(x'x)^{-1}x_0)}$$
$$\leq y_0 \leq \hat{Y}_0 + t_{\alpha/2, n-p}\sqrt{\hat{\sigma}^2(1+x_0'(x'x)^{-1}x_0)}$$

This a generalization of the prediction interval for a future observation in

Observation Number	Delivery Time (Minutes) y	No. of cases x_1	Distance (feet) x_2.
1	16.68	7	560
2	11.50	3	220
3	12.03	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8.00	2	110
8.	17.83	7	210
9.	79.24	30	1460
10.	21.50	5	605
11	40.33	16	688
12.	21.00	10	215
13.	13.50	4	255
14	19.75	6	462
15.	24.00	9	448
16.	29.00	10	776
17	15.35	6	200
18.	19.00	7	132
19.	9.50	3	36
20.	35.10	17	770
21.	17.90	10	140
22.	52.32	26	810
23.	18.75	9	450
24.	19.83	8	635
25.	10.75	4	150.

Standardized regression co-efficients

It is usually difficult to directly compare regression co-efficients because the magnitude of $\hat{\beta}_j$ reflects the units of measurement of the regressor x_j. For example, suppose that the regression model is

$$\hat{y} = 5 + x_1 + 1000 x_2.$$

and y is measure in liter's, x_1 is measured in milliters, and x_2 is measured in liters. Note that although $\hat{\beta}_2$ is considerabley larger than $\hat{\beta}_1$, the effect of both regressor on $\hat{y}$ is identical. Since a 1-liter change in either x_1 or x_2. When the other variable is held constant produces the same change in $\hat{y}$. (Generally the units of the regression co-efficient β_j are units of y/ units of x_j. For this reason, it is sometimes helpful to work with scaled regressor and response variables that produce dimension less regression co-efficients. This diamensionless co-efficients are usually called standardized regression co-efficients.) We now show how they are computed using two popular scaling techniques.

Unit normal scaling:

The 1st approach employs unit normal scaling for the regressors and the response variable, i.e,

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} \; ; \quad \left.\begin{array}{l} i = 1, 2, \dots n \\ j = 1, 2, \dots k \end{array}\right\} \text{ — } ①$$

and.

$$y_i^* = \frac{y_i - \bar{y}}{S_y} \; ; \quad i = 1, 2, \dots n \text{ — } ②$$

where, $S_j^2 = \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{n-1}$

is the Sample variance of regressor x_j and,

$S_y^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 / n-1$ is the Sample Variance of the regression. Note that Similarity to Standardizing a normal random variables. All of the scaled regressors ~~and the~~ response have sample mean equal to zero and the sample variance equal to 1.

Using these new variables, the regression model becomes,

$$y_i^* = b_1 z_{i1} + b_2 z_{i2} + \dots + b_k z_{ik} + \varepsilon_i \; ; \; i = 1, 2, \dots n \text{ — } ③$$

Centering the regresson and response variables by subtracting $\bar{x}_j$ and $\bar{y}$ removes the intercept from the model (actually the least squares estimate of b_0 is $\hat{b} = \bar{y}^* = 0$).

The least square estimator of b is,

$$\hat{b} = (z'z)^{-1} z' y^* \text{ — } ④$$

Unit length scaling:

The 2nd popular scaling is unit length scaling;

$$w_{ij} = \frac{x_{ij} - \bar{x}_j}{S_{jj}^{1/2}} \; , \quad \begin{array}{l} i = 1, 2, \dots n \\ j = 1, 2, \dots k. \end{array}$$