Mutiple Linear Regression and Estimation

Multiple Linear Regression analysis can be looked upon as an extension of simple linear regression analysis to the situation in which more than one independent variables must be considered.

The general model with response Y and regressors X_1, X_2,... X_p:

$$E(Y|X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

Suppose we observe data for *n* subjects with *p* variables. The data might be presented in a matrix or table like:

$$\begin{array}{cccc} y_1 & x_{11} & \cdots & x_{1p} \\ y_2 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ y_n & x_{n1} & \cdots & x_{np} \end{array}$$

We could then write the model as:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_p x_{2p} + \varepsilon_2 \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} + \varepsilon_n \end{aligned}$$

We can think of Y and E as (n x 1) vectors when transposed, and β as a ((p + 1) x 1) vector. p +1 is number of predictors + the intercept. Thus X would be:

$$\boldsymbol{X}_{n\times(p+1)} = \begin{pmatrix} x_1{}' \\ \vdots \\ x_n{}' \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

The general linear regression may be written as:

$$\boldsymbol{Y}_{n\times 1} = \boldsymbol{X}_{n\times(p+1)}\boldsymbol{\beta}_{(p+1)\times 1} + \boldsymbol{\varepsilon}_{n\times 1}$$

Or **$y_i = x_i^{|} * \beta + \in_i$**

The model is represented as the systematic structure plus the random variation with n dimensions = (p + 1) + { n - (p + 1) }

Ordinal Least Squares Estimators

The least squares estimate β_hat of β is chosen by minimizing the residual sum of squares function:

$$RSS(\boldsymbol{\beta}) = \sum (y_i - \boldsymbol{x}_i{}'\boldsymbol{\beta})^2 = (\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})$$

By differentiating with respect to β_i and solve by setting equal to 0:

$$\frac{\partial RSS}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}'\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{X\beta}$$

The least squares estimate of β_hat of β is given by:

$$(\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$$

and if the inverse exists:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

Fitted Values and Residuals

The fitted values are represetned by Y_hat = X*β_hat

$$\begin{aligned} \boldsymbol{e} &= \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \\ &= (\mathbf{I} - \mathbf{H})\mathbf{Y}, \end{aligned}$$

where the **hat matrix** is defined as $H = X(X^{|}X)^{-1}X^{|}$

The residual sum of squares (RSS):

$$\boldsymbol{e'e = Y'(I-H)'(I-H)Y = Y'(I-H)Y}$$

Gauss-Markov Conditions

In order for estimates of β to have some desirable statistical properties, we need a set of assumptions referred to as the Gauss-Markov conditions; for all i, j = 1... n

1. $E[\in_i] = 0$
2. $E[\in_i^2] =$ ^{2}
3. $E[\in_i \in_j] = 0$, where i != j

Or we can write these in matrix notation as: E[∈] = 0, $E[\in\in^{|}] =$ ^{2}*I

The GM conditions imply that E[Y] = Xβ and cov(Y) = $E[(Y\text{-}X\beta)(Y\text{-}X\beta)^{|}] = E[\in\in^{|}] = \in$

Under the GM assumptions, the LSE are the Best Linear Unbiased Estimators (BLUE). In this expression, "best" means minimum variance and linear indicates that the estimators are linear functions of y.

The LSE is a good choice but it does require the errors are uncorrelated and have equal variance. Even if the errors behave, then nonlinear or biased estimates may work better.

Estimating Variance

By definition:

$$\sigma^2 = E\big[y_i - E[y_i]\big]^2 = E[(y_i - x_i'\beta)^2]$$

We can estimate variance by an average from the sample:

$$s^2 = \frac{1}{n-p-1}\sum_{i=1}^{n}(y_i - x_i'\beta)^2$$

$$= \frac{1}{n-p-1}(\mathbf{Y}\text{-}\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y}\text{-}\mathbf{X}\hat{\boldsymbol{\beta}}) = \frac{RSS}{n-p-1}$$

Under GM conditions, s^2 is an **unbiased estimate** of variance.

Total Sum of Squares

Total sum of squares is Syy = SSreg + RSS

The corrected total sum of squares with n-1 degrees of freedom:

$$SYY = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

$$= Y'\,Y - \frac{1}{n}Y'JY = Y'\left[I - \frac{1}{n}J\right]Y$$

where J is an n x n matrix of 1s and H = X(X$^{|}$X)$^{-1}$X$^{|}$

Regression and Residual Sum of Sqaures

Regression sum of squares represent the number of X variables:

$$SSreg = \Sigma_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \Sigma_{i=1}^{n}\widehat{y_i^2} - n\bar{y}^2$$

$$= \hat{\beta}'X'Y - \frac{1}{n}Y'JY = Y'\left[H - \frac{1}{n}J\right]Y$$

Residual Sum of Squares with n - (p + 1) degrees of freedom:

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \mathrm{e}'\mathrm{e} = \left(\mathrm{Y} - \mathrm{X}\hat{\beta}\right)'\left(\mathrm{Y} - \mathrm{X}\hat{\beta}\right)$$

$$= \mathrm{Y}'\mathrm{Y} - \hat{\beta}'X'Y = Y'(I - H)Y$$

F Test for Regression Relation

To test whether there is a regression relation between Y and a set of variables X, use the hypothesis test:

□□0 : □□1 = □□2 = □□3 = ⋯ = □□□□ = 0 v.s. □□1 : not all □□□□ = 0, □□ = 1, … , □□

We use the test statistic:

$$F^{\,*} = \frac{MSReg}{MSE}$$

The chances of a Type I error is alpha, our degrees of freedom is n - p -1

The Coefficient of Determination

Recall this measures the model of fit by the proportionate reduction of total variation in Y associated with the use of the set of X variables.

R^2 = SSreg / Syy = 1 - RSS / Syy

In a multi-linear regression we must adjust the coefficient by the associated degrees of freedom:

$$R_a^2 = 1 - \frac{RSS/(n-p-1)}{SYY/(n-1)} = \frac{(n-1)R^2-p}{n-p-1}$$

Add more independent variables to the model can only increase R^2, but R^2_{alpha} may become smaller when more independent variables are introduced, as the decrease in RSS may be offset by the loss of degrees of freedom.

T Tests and Intervals

Tests for β_k are pretty standard:

$$t^* = \frac{\widehat{\beta}_k}{SE(\widehat{\beta}_k)}$$

With the rejection rule of □ >= t(1 − □/2; □ − □ − 1)

And likewise confidence limits for β_k and 1 - alpha confidence

$$\hat{\beta}_k \pm t(1-\alpha/2;\ \text{n-p-1})^* SE\left(\hat{\beta}_k\right)$$

Revision #6
Created 15 September 2022 22:02:12 by Elkip
Updated 15 September 2022 23:58:18 by Elkip