Multiple linear regression

Let $\underline{U} = \begin{bmatrix} Y \\ \left.\begin{matrix} X_1 \\ \vdots \\ X_p \end{matrix}\right\} \underline{X} \end{bmatrix}$ be a $p+1$ dimensional multivariate

normal distributed random vector.

Then $Y | \underline{X} = \underline{x} \sim N(\mu_y + \Sigma_{Y\underline{X}} \Sigma_{\underline{X}}^{-1}(\underline{X} - \underline{\mu}_x),\ \underline{\Sigma}_Y - \Sigma_{Y\underline{X}} \Sigma_{\underline{X}}^{-1} \Sigma_{\underline{X}Y})$

$\Rightarrow \quad E[Y | \underline{X} = \underline{x}] = \mu_y + \Sigma_{Y\underline{X}} \Sigma_{\underline{X}}^{-1} (\underline{X} - \mu_x)$

$\Rightarrow \quad \mu_y + \beta_1(X_1 - u_1) + \beta_2(X_2 - \mu_2) + \cdots + \beta_p(X_p - \mu_p)$

i.e. a multiple linear regression model.

The model for the multiple linear regression model

is: $\quad Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_k + \varepsilon$

Response-variable

regression variables
explanatory variables
predictors

error

or alternatively: $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i$ where

the normal assumptions are $E[\varepsilon_i] = 0$, $Var[\varepsilon_i] = \sigma^2$, $i = 1, 2, \ldots, m$

and $E[\varepsilon_i \varepsilon_j] = 0$, $i \neq j$.

In matrix form $\begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & & & \vdots \\ 1 & x_{m1} & \cdots & x_{mk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{bmatrix}$

$$\underline{Y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon}$$

By a linear model we mean linear in the coefficients such that X_i can be substituted by any transformation for instance $\ln(X_i)$, X_i^2 or $\sin(X_i)$ as long as the transformation is allowed.

In the model the response variables $Y_1, Y_2, \ldots, Y_n$ are random variables while the regression variables can be either deterministic or stochastic. This has some implications in the interpretation of the model.

If deterministic we have:

$$E[Y_i] = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}, \qquad i = 1, 2, \ldots, n$$

If they are stochastic

$$E[Y_i \mid X_{i1}, \ldots, X_{ik}] = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}, \quad i = 1, 2, \ldots, n$$

<u>Estimation of parameters</u>

We observe $(y_i, x_{i1}, \ldots, x_{ik})$, $i = 1, 2, \ldots, n$ which according to the model must satisfy

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \varepsilon_1^* \\
y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + \varepsilon_2^* \\
&\vdots \\
y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + \varepsilon_n^*
\end{aligned}$$

where ε_i^* is a realization of ε_i, $i = 1, 2, \ldots, n$

or
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1' \\ \varepsilon_2'' \\ \vdots \\ \varepsilon_n' \end{bmatrix}$$

which in matrix form becomes

$$\underset{\sim}{y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon}'$$

Estimates for $\underline{\beta} = (\beta_0, \beta_1, \cdots, \beta_k)^T$, $(b_0, b_1, \cdots, b_k)^T$ are obtained by minimizing

$$Q = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \cdots - b_k x_{ik})^2$$

with respect to $b_0, b_1, \cdots, b_k$. The estimated model becomes

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik}$$

Q can be shown to be a convex function in $b_0, b_1, \cdots, b_k$. Setting the partial derivatives equal to zero gives the normal equations.

$$-2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \cdots - b_k x_{ik}) = -2 \sum_{i=1}^{n} (y_i - \hat{y}_i) = 0$$

$$-2 \sum_{i=1}^{n} x_{i1} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \cdots - b_k x_{ik}) = -2 \sum_{i=1}^{n} x_{i1}(y_i - \hat{y}_i) = 0$$

$$\vdots$$

$$-2 \sum_{i=1}^{n} x_{ik} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \cdots - b_k x_{ik}) = -2 \sum_{i=1}^{n} x_{ik} (y_i - \hat{y}_i) = 0$$

which can be written as:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{m1} \\ \vdots & & & \\ x_{1k} & x_{2k} & \cdots & x_{mk} \end{bmatrix} \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_m - \hat{y}_m \end{bmatrix} = \underline{0}$$

or $\underline{X}^T(\underline{y} - \hat{\underline{y}}) = \underline{X}^T(\underline{y} - \underline{X}\underline{b}) = 0$

which gives $\underline{X}^T\underline{X}\,\underline{b} = \underline{X}^T\underline{y}$ and $b = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{y}$

if $(\underline{X}^T\underline{X})^{-1}$ exist.

The least square estimator is therefore

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

The residuals are given by $e_i = y_i - \hat{y}_i$, $i = 1, 2, \ldots, m$

and the first normal equation gurantees that

$\sum_{i=1}^{\hat{}} e_i = 0$. With no constant term in the model

there is no such guarantee.

Expectation and Covariance matrix of the estimators.

$$\hat{\underline{\beta}} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T(\underline{X}\underline{\beta} + \underline{\varepsilon})$$

$$\Rightarrow \hat{\underline{\beta}} = \underline{\beta} + (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{\varepsilon}$$

Since $E[\underline{\varepsilon}] = \underline{0}$, $E[\hat{\underline{\beta}}] = \underline{\beta}$ i.e the least square estimator is unbiased.

$$\mathrm{Cov}[\hat{\underline{\beta}}] = E[(\hat{\underline{\beta}} - \underline{\beta})(\hat{\underline{\beta}} - \underline{\beta})^T] = E[(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{\varepsilon}\,\underline{\varepsilon}^T\underline{X}(\underline{X}^T\underline{X})^{-1}]$$

Note

$$(\underline{X}^T\underline{X})^T = \underline{X}^T\underline{X} \Rightarrow \underline{X}^T\underline{X} \text{ symmetric}$$

$$\underline{A} \text{ symmetric} \Rightarrow \underline{A} = \underline{P}\underline{\Lambda}\underline{P}^T \text{ and } \underline{A}^{-1} = \underline{P}\underline{\Lambda}^{-1}\underline{P}^T$$

$$\Rightarrow (\underline{A}^{-1})^T = \underline{P}\underline{\Lambda}^{-1}\underline{P}^T = \underline{A}^{-1} \Rightarrow \underline{A}^{-1} \text{ is symmetric}$$

We get

$$\mathrm{Cov}[\hat{\underline{\beta}}] = (\underline{X}^T\underline{X})^{-1}\underline{X}^T E[\underline{\varepsilon}\,\underline{\varepsilon}^T]\underline{X}(\underline{X}^T\underline{X})^{-1}$$

$$= (\underline{X}^T\underline{X})^{-1}\underline{X}^T\sigma^2\underline{I}\,\underline{X}(\underline{X}^T\underline{X})^{-1} = \sigma^2(\underline{X}^T\underline{X})^{-1}$$

Maximum likelihood estimation

Assuming $\underline{\varepsilon} \sim N(\underline{0}, \sigma^2\underline{I})$ we have for $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$ that $\underline{Y} \sim N(\underline{X}\underline{\beta}, \sigma^2\underline{I})$ and the likelihood estimator for n indepent variables can be written

$$L(\underline{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}(\underline{Y} - \underline{X}\underline{\beta})^T(\underline{Y} - \underline{X}\underline{\beta})} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \underline{x}_i^T\underline{\beta})^2}$$

where $\underline{x}_i = \begin{bmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{ik} \end{bmatrix}$. This shows that the ML estimates for $\underline{\beta}$ are, the case of equal variances, found by minimizing $\sum_{i=1}^{n}(y_i - \underline{x}_i^T\underline{\beta})^2$ with respect to $\underline{\beta}$ and therefore equals the least squares estimates.