Multiple linear regression

July 24, 2021

1 Multiple linear regression

1.1 Matrix expression of the multiple linear regression

The multiple linear regression model assumes that the statistical relationship between the response variable Y_i $(i = 1, 2, \cdots, n)$ and the explanatory variables x_{ij} $(i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, k)$ is of the form

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \\ \epsilon_i &\sim N(0, \sigma^2) \end{aligned}$$

where β_j $(j = 0, 1, 2, \cdots, k)$ are the regression parameters and ϵ_i denote the independent normal random variables with zero mean and common variance σ^2.

The matrix representation for this regression model is denoted by

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

where

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad , \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \quad , \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

The matrix $\mathbf{X}$ is the $n \times (k + 1)$ matrix and called the *design matrix*.

The fitted values $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots \hat{\beta}_k x_{ik}$ and the residual vector $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\beta}$ can also be represented using matrix notation as follows;

$$\begin{aligned} \hat{\mathbf{Y}} &= \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 + \hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{12} + \cdots + \hat{\beta}_k x_{1k} \\ \hat{\beta}_0 + \hat{\beta}_1 x_{21} + \hat{\beta}_2 x_{22} + \cdots + \hat{\beta}_k x_{2k} \\ \vdots \\ \hat{\beta}_0 + \hat{\beta}_1 x_{n1} + \hat{\beta}_2 x_{n2} + \cdots + \hat{\beta}_k x_{nk} \end{pmatrix} = \mathbf{X}\hat{\beta} \\ \mathbf{e} &= \begin{pmatrix} Y_1 - \hat{Y}_1 \\ Y_2 - \hat{Y}_2 \\ \vdots \\ Y_n - \hat{Y}_n \end{pmatrix} = \mathbf{Y} - \mathbf{X}\hat{\beta} \end{aligned}$$

where

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}$$

are the estimated parameters.

The residual vector is orthgonal to the columns of the design matrix.

$$\mathbf{X}^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) = 0$$

Then the parameter estimates $\hat{\beta}$ are derived form the normal equations,

$$\mathbf{X}^T\mathbf{X}\hat{\beta} = \mathbf{X}^T\mathbf{Y}$$

When the inverse matrix $(\mathbf{X}^T\mathbf{X})^{-1}$ is exists, the solutions to the normal equations, the vector of fitted values, and the residual vector in matrix form are given by

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} && (1)\\
\hat{\mathbf{Y}} &= \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{H}\mathbf{Y} && (2)\\
\mathbf{e} &= \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{H}\mathbf{Y} = (\mathbf{I} - \mathbf{H})\mathbf{Y} && (3)
\end{aligned}$$

where

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \qquad (4)$$

The matrix $\mathbf{H}$ is called the *hat matrix*.

1.2 Estimated standard errors of the estimated regression parameters

Suppose the random vector $\mathbf{W}$ is obtained by multiplying the random vector $\mathbf{Y}$ by matrix $\mathbf{A}$; that is, $\mathbf{W} = \mathbf{A}\mathbf{Y}$. Then

$$\mathrm{Cov}(\mathbf{W}) = \mathbf{A}\mathrm{Cov}(\mathbf{Y})\mathbf{A}^T \qquad (5)$$

It follows from Eq.(5) with

$$\begin{aligned}
\mathbf{A} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\\
\mathbf{A}^T &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\\
\mathrm{Cov}(\mathbf{Y}) &= \sigma^2\mathbf{I}
\end{aligned}$$

that

$$\mathrm{Cov}(\hat{\beta}) = \mathbf{A}\sigma^2\mathbf{I}\mathbf{A}^T = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \qquad (6)$$

When we substitute the unbiased estimator

$$s^2 = \frac{\mathbf{e}^T\mathbf{e}}{n - k - 1}$$

for the variance σ^2 in Eq.(6), we obtained the *estimated variance-covariance matrix*:

$$s^2(\hat{\beta}) = s^2(\mathbf{X}^T\mathbf{X})^{-1}$$

and the estimated standard errors $s(\hat{\beta})$ for the parameter estimates $\hat{\beta}$.

1.3 Coefficients of multiple determination

The SST (sum of total squares), the SSR (sum of the squared regression) and the SSE (sum of the squared errors) are defined as follows respectively;

$$\begin{aligned}
\mathrm{SST} &= \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \mathbf{Y}^T\mathbf{Y} - n\bar{Y}^2\\
\mathrm{SSR} &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \hat{\mathbf{Y}}^T\hat{\mathbf{Y}} - n\bar{Y}^2\\
\mathrm{SSE} &= \sum_{i=1}^{n}e_i^2 = \mathbf{e}^T\mathbf{e}
\end{aligned}$$

These satisfies the following relation,

$$\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}.$$

The coefficients of multiple determination, denoted R^2, is defined as follows:

$$R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}}$$

The adjusted R^2 is defined as follows:

$$R_a^2 = 1 - \left(\frac{n-1}{n-k-1}\right)(1-R^2)$$

1.4 Confidence intervals for the regression parameters

The confidence intervals for the regression parameters can derived from

$$\frac{\hat{\beta}_i - \beta_i}{s(\hat{\beta}_i)} \sim t_{n-k-1}$$

where $s(\hat{\beta}_i)$ is the estimated standard error of the paramter estimate $\hat{\beta}_i$ and t_{n-k-1} is the Student's t distribution with $(n-k-1)$ degrees of freedom. For example, a 100(1-α)% confidence interval for the parameter estimate $\hat{\beta}_i$ is

$$\hat{\beta}_i \pm s(\hat{\beta}_i) t_{n-k-1}\left(\frac{\alpha}{2}\right)$$

1.5 Studentized residual

The residual vector $\mathbf{e}$ can be written in tersm of the hat matrix Eq.(4) as follows;

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

It follows from Eq.(5) with

$$\begin{aligned} \mathbf{H}^T &= \mathbf{H}, \\ (\mathbf{I} - \mathbf{H})^2 &= (\mathbf{I} - \mathbf{H}) \end{aligned}$$

that

$$\mathrm{Cov}(\mathbf{e}) = (\mathbf{I} - \mathbf{H})\sigma^2 \mathbf{I}(\mathbf{I} - \mathbf{H})^T = \sigma^2(\mathbf{I} - \mathbf{H})^2 = \sigma^2(\mathbf{I} - \mathbf{H})$$

It can be shown that the variance of the ith residual e_i are given by

$$V(e_i) = \sigma^2(1 - h_{ii})$$

where h_{ii} is the ith diagonal element of the hat matrix.

We obtain the *estimated standard error for the ith residual*, denoted $s(e_i)$, by replaceing σ^2 with its estimate s^2. Thus,

$$s(e_i) = s\sqrt{1 - h_{ii}}$$

The *Studentized residual* is difined, denoted e_i^*, as follows:

$$e_i^* = \frac{e_i}{s(e_i)}$$

1.6 Confidence intervals and prediction intervals in multiple linear regression

The vector

$$\mathbf{x}_0^T = (1, x_{01}, x_{02}, \cdots, x_{0k})$$

denotes the values of the explanatory variables; that is,

$$X_1 = x_{01}, X_2 = x_{02}, \cdots, X_k = x_{0k}$$

The estimated mean response, denoted $\hat{Y}(\mathbf{x}_0)$, can be written as the matrix product

$$\hat{Y}(\mathbf{x}_0) = \mathbf{x}_0^T \hat{\beta}$$

It follows from Eq.(2),

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

that

$$\mathbf{x}_0^T\hat{\beta} = \mathbf{A}\mathbf{Y}$$

where

$$\begin{aligned} \mathbf{A} &= \mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\ \mathbf{A}^T &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0 \end{aligned}$$

It follows from Eq.(5), with $\mathrm{Cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$ that

$$\mathrm{Cov}(\mathbf{x}_0^T\hat{\beta}) = \mathbf{A}\sigma^2\mathbf{I}\mathbf{A}^T = \sigma^2\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0 = \sigma^2\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0$$

The variance of $\hat{Y}(\mathbf{x}_0)$ is given by

$$\mathrm{V}(\hat{Y}(\mathbf{x}_0)) = \sigma^2\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0$$

We obtain the estimated *standard error of prediction*, denoted $s(\hat{Y}(\mathbf{x}_0))$, by replaceing σ^2 with its estimate $s^2 = \mathrm{SSE}/(n-k-1)$; thus,

$$s(\hat{Y}(\mathbf{x}_0)) = s\sqrt{\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0}$$

A $100(1-\alpha)\%$ confidence interval for the mean response at the values $\mathbf{x}_0$ is given by

$$\mathbf{x}_0^T\hat{\beta} \pm t_{n-k-1}\left(\frac{\alpha}{2}\right)s\sqrt{\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0}$$

A future response is given by $Y(\mathbf{x}_0) = \beta_0 + \beta_1 x_{01} + \cdots + \beta_k x_{0k} + \epsilon_0$ and the predicted future response is given by $\hat{Y}(\mathbf{x}_0) = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \cdots + \hat{\beta}_k x_{0k}$ at the values $\mathbf{x}_0$ of the explanatory variables, where ϵ_0 denote the independent normal random variable with zero mean and variance σ^2 that is independent of $\hat{Y}(\mathbf{x}_0)$.

Cosequently the variance of the difference between the future response $Y(\mathbf{x}_0)$ and the predicted future response $\hat{Y}(\mathbf{x}_0)$, denoted $V(Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0))$, is given by

$$\mathrm{V}(Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0)) = \mathrm{V}(Y(\mathbf{x}_0)) + \mathrm{V}(\hat{Y}(\mathbf{x}_0)) = \sigma^2(1 + \mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0)$$

Similarly, the estimated standard error of the difference between the future response $Y(\mathbf{x}_0)$ and the predicted future response $\hat{Y}(\mathbf{x}_0)$ is given by

$$s(Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0)) = s\sqrt{1 + \mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0}$$

The coresponding $100(1-\alpha)\%$ prediction interval for a future response $Y(\mathbf{x}_0)$ is given by

$$\mathbf{x}_0^T\hat{\beta} \pm t_{n-k-1}\left(\frac{\alpha}{2}\right)s\sqrt{1 + \mathbf{x}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}}$$

References

[1] W.A.Rosenkrantz, "Introduction to Probability and Statistics for Scientists and Engineers", McGraw-Hill、1997