Multivariate Analysis
Chapter 7: Regression

Shaobo Jin

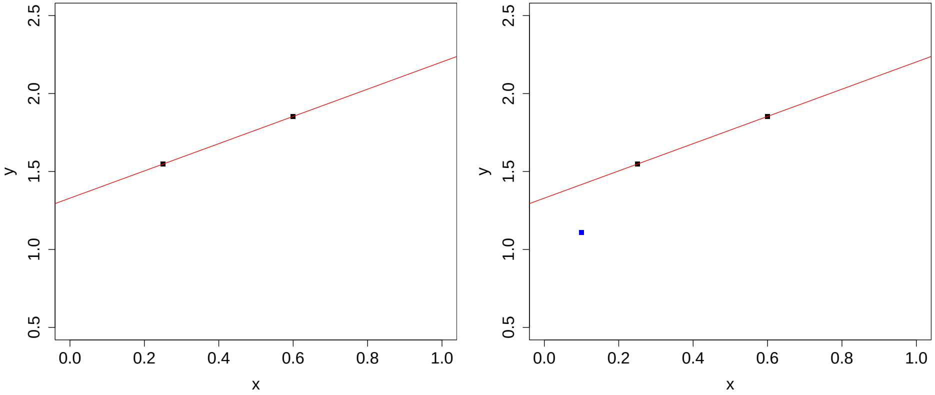
Department of Mathematics

Intended Learning Outcome

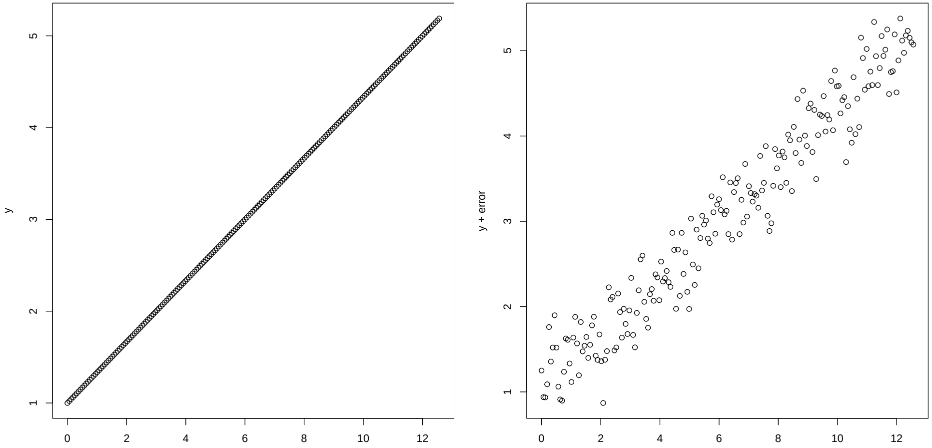
Through this chapter, you should be able to

1. Fit classic linear regression models,
2. Fit multivariate linear regression models,
3. Test regression coefficients,
4. Construct confidence regions/intervals for regression coefficients and regression functions,
5. Construct prediction regions/intervals.

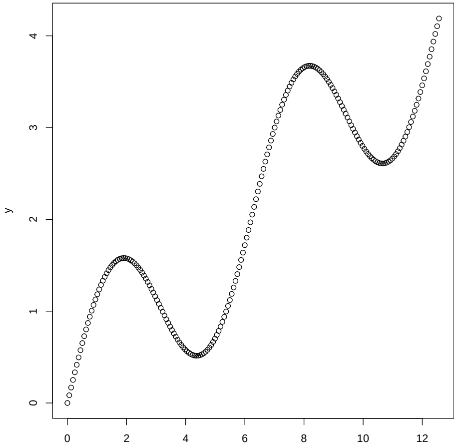
Two Points Determine A Line

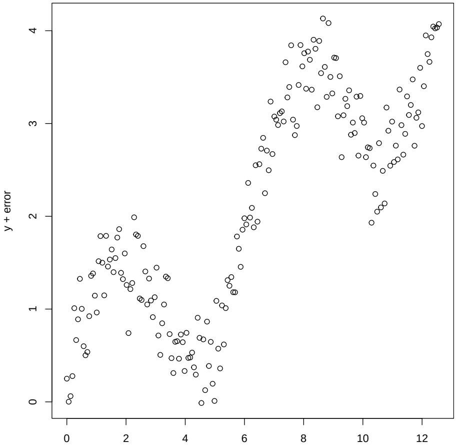


Draw A Line/Curve



Draw A Line/Curve





Classic Linear Regression

Let $\boldsymbol{Z}$ be a vector of covariates (treated as fixed in our course). The linear regression model is

$$\underset{\text{response}}{Y} = \underset{\text{conditional mean}}{\boldsymbol{Z}^T\boldsymbol{\beta}} + \underset{\text{error}}{e},$$

where $\mathbb{E}(e) = 0$ and $\text{var}(e) = \sigma^2$. In this model

$$\mathbb{E}(Y) = \boldsymbol{Z}^T\boldsymbol{\beta} = \sum_{k=1}^{r} Z_k\beta_k.$$

Some examples are:

1. Y is apartment price, $\boldsymbol{Z}$ includes crime rate, number of rooms, size of the apartment, year of construction, etc.
2. Y is waste water flow rate, $\boldsymbol{Z}$ includes temperature, precipitation, date of the year, time, etc.
3. Y is test score, $\boldsymbol{Z}$ includes school, age, gender, nationality, parents education level, etc.

Matrix Notation

Suppose that we have n observations of Y as

$$Y_j = \boldsymbol{Z}_j^T \boldsymbol{\beta} + e_j, \quad j = 1, 2, ..., n.$$

The assumptions are

1. $\mathbb{E}(e) = 0$,
2. $\text{var}(e) = \sigma^2$ (homoscedasticity),
3. observations are independent, i.e., Y_j is independent of Y_k for $j \neq k$.

In matrix notation,

$$\boldsymbol{Y}_{n\times 1} = \boldsymbol{Z}_{n\times r}\boldsymbol{\beta}_{r\times 1} + \boldsymbol{e}_{n\times 1}$$

where $\mathbb{E}(\boldsymbol{e}) = \boldsymbol{0}$, $\text{cov}(\boldsymbol{e}) = \sigma^2\boldsymbol{I}$. $\boldsymbol{Z}$ is the design matrix.

ANOVA Is Regression

The ANOVA model (MANOVA with $p = 1$) is

$$Y_{\ell j} \quad = \quad \mu + \tau_\ell + e_{\ell j}.$$

It is equivalent to

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 & \cdots \\ 1 & 0 & 1 & 0 & \cdots \\ 1 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{1n_1} \\ e_{21} \\ \vdots \\ e_{2n_2} \\ \vdots \end{bmatrix}.$$

Example: ANOVA With $g = 2$

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{24} \end{bmatrix}.$$

Instead of the restriction $\sum_{\ell=1}^{g} \tau_\ell = 0$, we can simply let $\tau_1 = 0$.

Example: ANOVA With $g = 2$ and $b = 2$

$$
\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{113} \\ Y_{121} \\ Y_{122} \\ Y_{123} \\ Y_{211} \\ Y_{212} \\ Y_{213} \\ Y_{221} \\ Y_{222} \\ Y_{223} \end{bmatrix}
=
\begin{bmatrix}
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \beta_1 \\ \beta_2 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \end{bmatrix}
+
\begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{24} \end{bmatrix}.
$$

Instead of the restriction $\sum_{\ell=1}^{g} \gamma_{\ell k} = \sum_{k=1}^{b} \gamma_{\ell k} = 0$, we can simply let $\gamma_{\ell 1} = \gamma_{1k} = 0$ for all ℓ and k.

Ordinary Least Squares

The method of ordinary least squares (OLS) is often used to estimate $\boldsymbol{\beta}$. The OLS estimator minimizes the sum of squares

$$\sum_{j=1}^{n} \left(y_j - \boldsymbol{z}_j^T \boldsymbol{\beta}\right)^2 = \left(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\beta}\right)^T \left(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\beta}\right).$$

Lemma

Consider the quadratic form $S\left(\boldsymbol{\beta}\right) = \left(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\beta}\right)^T \left(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\beta}\right)$. Its gradient and Hessian are

$$\begin{aligned} \frac{\partial S\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}} &= -2\boldsymbol{Z}^T \left(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\beta}\right), \\ \frac{\partial^2 S\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= 2\boldsymbol{Z}^T \boldsymbol{Z}, \end{aligned}$$

respectively.

OLS Estimator

Result 7.1

Let $\boldsymbol{Z}$ have full rank $r \leq n$. The OLS estimate of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{Z}^T \boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^T \boldsymbol{y}.$$

Let $\hat{\boldsymbol{y}} = \boldsymbol{Z}\hat{\boldsymbol{\beta}} = \boldsymbol{H}\boldsymbol{y}$ denote the fitted values of $\boldsymbol{y}$, where

$$\boldsymbol{H} = \boldsymbol{Z}\left(\boldsymbol{Z}^T \boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^T$$

is called the hat matrix. Then the residuals

$$\hat{\boldsymbol{e}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{H})\, \boldsymbol{y}$$

satisfy $\boldsymbol{Z}^T \hat{\boldsymbol{e}} = \mathbf{0}$ and $\hat{\boldsymbol{y}}^T \hat{\boldsymbol{e}} = 0$. The residual sum of squares is $\boldsymbol{y}^T (\boldsymbol{I} - \boldsymbol{H})\, \boldsymbol{y}$.

Illustration (2D): Projection

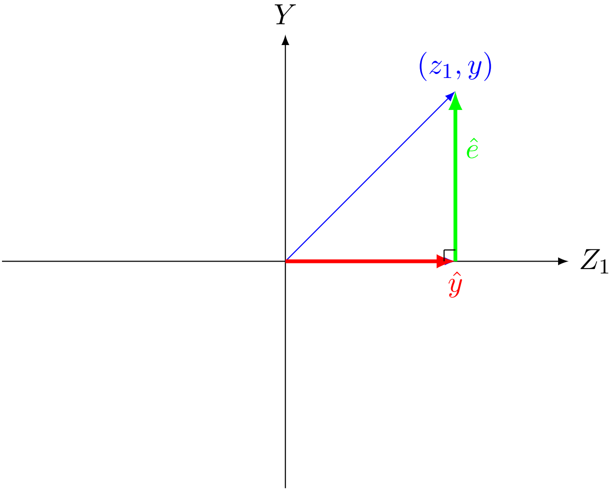
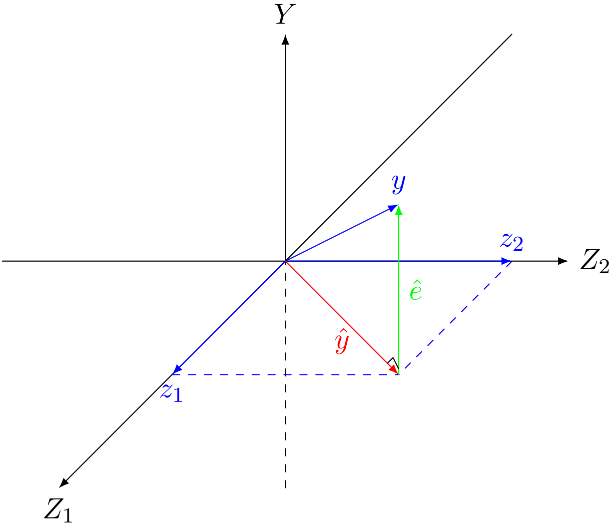


Illustration (3D): Projection



Sampling Properties of OLS Estimators

Result 7.2

Under the classic linear regression model, the OLS estimator $\hat{\boldsymbol{\beta}}$ has

$$\mathbb{E}\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}, \qquad \operatorname{cov}\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2 \left(\boldsymbol{Z}^T \boldsymbol{Z}\right)^{-1}.$$

The residuals have the properties

$$\mathbb{E}\left(\hat{\boldsymbol{e}}\right) = \mathbf{0}, \qquad \operatorname{cov}\left(\hat{\boldsymbol{e}}\right) = \sigma^2 \left(\boldsymbol{I} - \boldsymbol{H}\right).$$

Further, $\mathbb{E}\left(\hat{\boldsymbol{e}}^T \hat{\boldsymbol{e}}\right) = (n-r)\,\sigma^2$ and an unbiased estimator of σ^2 is

$$S^2 = \frac{\hat{\boldsymbol{e}}^T \hat{\boldsymbol{e}}}{n-r}.$$

Moreover, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{e}}$ are uncorrelated.

Maximum Likelihood Estimator

We add one assumption $\boldsymbol{e} \sim N_r\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$ to the classic linear regression model. Then, $\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{e} \sim N_n\left(\boldsymbol{Z}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right)$. The log-likelihood is

$$\ell\left(\boldsymbol{\beta}, \sigma^2\right) = -\frac{n}{2}\log\left(2\pi\right) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\beta}\right)^T\left(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\beta}\right).$$

The MLE is

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{y}, \\ \hat{\sigma}^2 &= \frac{1}{n}\left(\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}\right) \\ &= \frac{1}{n}\hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}}. \end{aligned}$$

Same $\hat{\boldsymbol{\beta}}$ as the OLS estimator!

Distribution of Regression Coefficients

Result 7.4

Let $\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{e}$, where $\boldsymbol{Z}$ has full rank r and $\boldsymbol{e} \sim N_n\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}\right)$. Then the maximum likelihood estimator of $\boldsymbol{\beta}$ is the same as the least squares estimator $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{Y}$. Moreover,

$$\hat{\boldsymbol{\beta}} \;\sim\; N_r\left(\boldsymbol{\beta}, \sigma^2\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\right),$$

and is distributed independent of the residuals $\hat{\boldsymbol{e}}$. Further,

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \frac{\hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}}}{\sigma^2} \;\sim\; \chi^2_{n-r},$$

where $\hat{\sigma}^2$ is the MLE of σ^2.

Confidence Region

By Result 7.4,

$$\sigma^{-2}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)^T \boldsymbol{Z}^T\boldsymbol{Z}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \sim \chi^2_r, \qquad \sigma^{-2}\hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}} \sim \chi^2_{n-r},$$

and they are independent. Hence,

$$\frac{\sigma^{-2}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)^T \boldsymbol{Z}^T\boldsymbol{Z}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)/r}{\sigma^{-2}\hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}}/\left(n-r\right)} \sim F_{r,n-r}.$$

Result 7.5

Let $\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{e}$, where $\boldsymbol{Z}$ has full rank r and $\boldsymbol{e} \sim N_n\left(\mathbf{0}, \sigma^2\boldsymbol{I}\right)$. Then a $1-\alpha$ confidence region for $\boldsymbol{\beta}$ is given by

$$\frac{\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)^T \boldsymbol{Z}^T\boldsymbol{Z}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)/r}{\hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}}/\left(n-r\right)} \le F_{r,n-r}\left(\alpha\right).$$

Confidence Interval

Since $\hat{\beta}_i \sim N\left(\beta_i, \sigma^2 \left[\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\right]_{ii}\right)$ is independent of $(n-r)\,S^2/\sigma^2 = \hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}}/\sigma^2 \sim \chi^2_{n-r}$,

$$\frac{\left(\hat{\beta}_i - \beta_i\right)/\sqrt{\sigma^2\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\right]_{ii}}}{\sqrt{\sigma^{-2}\hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}}/\left(n-r\right)}} \quad\sim\quad t_{n-r}.$$

A $1-\alpha$ individual confidence interval for β_i is

$$\hat{\beta}_i \pm t_{n-r}\left(\frac{\alpha}{2}\right)\sqrt{\hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}}\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\right]_{ii}}.$$

You can also construct simultanous confidence intervals and Bonferroni confidence intervals.

More Than One Responses

Suppose that each subject has m responses

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} = \begin{bmatrix} \beta_{01} + \beta_{11} z_1 + \cdots + \beta_{r-1,1} z_{r-1} \\ \beta_{02} + \beta_{12} z_1 + \cdots + \beta_{r-1,2} z_{r-1} \\ \vdots \\ \beta_{0m} + \beta_{1m} z_1 + \cdots + \beta_{r-1,m} z_{r-1} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}.$$

In other words, for the ith response, $i = 1, 2, ..., m$, we have a classic regression model

$$Y_i = \underbrace{\begin{bmatrix} \beta_{0i} & \beta_{1i} & \cdots & \beta_{r-1,i} \end{bmatrix}}_{\boldsymbol{\beta}_{(i)}^T: \quad 1 \times r} \boldsymbol{Z} + e_i.$$

Then, for each subject,

$$\underset{m \times 1}{\boldsymbol{Y}} = \underset{m \times r}{\boldsymbol{\beta}^T} \ \underset{r \times 1}{\boldsymbol{Z}} + \underset{m \times 1}{\boldsymbol{E}}.$$

Matrix Notation

Suppose that we have n observations. Then,

$$\begin{bmatrix} \boldsymbol{Y}_1^T \\ \boldsymbol{Y}_2^T \\ \vdots \\ \boldsymbol{Y}_n^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{Z}_1^T \\ \boldsymbol{Z}_2^T \\ \vdots \\ \boldsymbol{Z}_n^T \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{E}_1^T \\ \boldsymbol{E}_2^T \\ \vdots \\ \boldsymbol{E}_n^T \end{bmatrix}.$$

$$\underset{n \times m}{\boldsymbol{Y}} = \underset{n \times r}{\boldsymbol{Z}} \; \underset{r \times m}{\boldsymbol{\beta}} + \underset{n \times m}{\boldsymbol{E}}$$

$$\begin{bmatrix} Y_{11} & Y_{12} & \cdots \\ Y_{21} & Y_{22} & \cdots \\ \vdots & \vdots & \ddots \\ Y_{n1} & Y_{n2} & \cdots \end{bmatrix} = \boldsymbol{Z} \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots \\ \beta_{11} & \beta_{12} & \cdots \\ \vdots & \vdots & \ddots \\ \beta_{r-1,1} & \beta_{r-1,2} & \cdots \end{bmatrix} + \begin{bmatrix} E_{11} & E_{12} & \cdots \\ E_{21} & E_{22} & \cdots \\ \vdots & \vdots & \ddots \\ E_{n1} & E_{n2} & \cdots \end{bmatrix}$$

Assumptions

The model is

$$\underset{n\times m}{\boldsymbol{Y}} = \underset{n\times r}{\boldsymbol{Z}} \underset{r\times m}{\boldsymbol{\beta}} + \underset{n\times m}{\boldsymbol{e}}.$$

Equivalently, the ith response $\boldsymbol{Y}_{(i)}$ follows the classic linear regression

$$\boldsymbol{Y}_{(i)} = \boldsymbol{Z}\boldsymbol{\beta}_{(i)} + \boldsymbol{e}_{(i)}.$$

The assumptions are

1. $\mathbb{E}\left(\boldsymbol{e}_{(i)}\right) = \mathbf{0}$,
2. $\operatorname{cov}\left(\boldsymbol{e}_{(i)}\right) = \sigma_{ii}\boldsymbol{I}$ and $\operatorname{cov}\left(\boldsymbol{e}_{(i)}, \boldsymbol{e}_{(k)}\right) = \sigma_{ik}\boldsymbol{I}$, for $i, k = 1, 2, ..., m$. That is, the m observations on the jth subject have covariance matrix $\boldsymbol{\Sigma} = \{\sigma_{ik}\}$, but observations from different subjects are uncorrelated.

Least Squares

The error sum of squares and cross products matrix $(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})^T (\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})$ is

$$\begin{bmatrix} \left(\boldsymbol{Y}_{(1)} - \boldsymbol{Z}\boldsymbol{\beta}_{(1)}\right)^T \left(\boldsymbol{Y}_{(1)} - \boldsymbol{Z}\boldsymbol{\beta}_{(1)}\right) & \cdots & \left(\boldsymbol{Y}_{(1)} - \boldsymbol{Z}\boldsymbol{\beta}_{(1)}\right)^T \left(\boldsymbol{Y}_{(m)} - \boldsymbol{Z}\boldsymbol{\beta}_{(m)}\right) \\ \left(\boldsymbol{Y}_{(2)} - \boldsymbol{Z}\boldsymbol{\beta}_{(2)}\right)^T \left(\boldsymbol{Y}_{(1)} - \boldsymbol{Z}\boldsymbol{\beta}_{(1)}\right) & \cdots & \left(\boldsymbol{Y}_{(2)} - \boldsymbol{Z}\boldsymbol{\beta}_{(2)}\right)^T \left(\boldsymbol{Y}_{(m)} - \boldsymbol{Z}\boldsymbol{\beta}_{(m)}\right) \\ \vdots & \ddots & \vdots \\ \left(\boldsymbol{Y}_{(m)} - \boldsymbol{Z}\boldsymbol{\beta}_{(m)}\right)^T \left(\boldsymbol{Y}_{(1)} - \boldsymbol{Z}\boldsymbol{\beta}_{(1)}\right) & \cdots & \left(\boldsymbol{Y}_{(m)} - \boldsymbol{Z}\boldsymbol{\beta}_{(m)}\right)^T \left(\boldsymbol{Y}_{(m)} - \boldsymbol{Z}\boldsymbol{\beta}_{(m)}\right) \end{bmatrix}.$$

We want $(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})^T (\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})$ to be "small". For example, we want to minimize

$$\operatorname{tr}\left\{(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})^T (\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})\right\} = \sum_{i=1}^{m} \left(\boldsymbol{Y}_{(i)} - \boldsymbol{Z}\boldsymbol{\beta}_{(i)}\right)^T \left(\boldsymbol{Y}_{(i)} - \boldsymbol{Z}\boldsymbol{\beta}_{(i)}\right).$$

Least Squares Estimator of $\boldsymbol{\beta}$

The sum of squares $\left(\boldsymbol{Y}_{(i)} - \boldsymbol{Z}\boldsymbol{\beta}_{(i)}\right)^T \left(\boldsymbol{Y}_{(i)} - \boldsymbol{Z}\boldsymbol{\beta}_{(i)}\right)$ is minimized by the OLS estimator

$$\hat{\boldsymbol{\beta}}_{(i)} = \left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^T\boldsymbol{Y}_{(i)}.$$

Hence,

$$\operatorname{tr}\left\{(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})^T (\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})\right\} = \sum_{i=1}^{m} \left(\boldsymbol{Y}_{(i)} - \boldsymbol{Z}\boldsymbol{\beta}_{(i)}\right)^T \left(\boldsymbol{Y}_{(i)} - \boldsymbol{Z}\boldsymbol{\beta}_{(i)}\right)$$

is minimized by

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \begin{bmatrix} \hat{\boldsymbol{\beta}}_{(1)} & \hat{\boldsymbol{\beta}}_{(2)} & \cdots & \hat{\boldsymbol{\beta}}_{(m)} \end{bmatrix} \\ &= \left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^T \begin{bmatrix} \boldsymbol{Y}_{(1)} & \boldsymbol{Y}_{(2)} & \cdots & \boldsymbol{Y}_{(m)} \end{bmatrix} \\ &= \left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^T\boldsymbol{Y}. \end{aligned}$$

Sampling Property of Estimator

Result 7.9

Suppose that $\boldsymbol{Z}$ has full column rank. The least squares estimator $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$:

$$\mathbb{E}\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}.$$

The residuals $\hat{\boldsymbol{E}} = \boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}$ satisfy

$$\mathbb{E}\left(\hat{\boldsymbol{E}}\right) = \mathbf{0}, \qquad \mathbb{E}\left(\hat{\boldsymbol{E}}^T\hat{\boldsymbol{E}}\right) = (n-r)\,\boldsymbol{\Sigma}.$$

Moreover, $\hat{\boldsymbol{E}}$ and $\hat{\boldsymbol{\beta}}$ are uncorrelated.

With Normality Assumption

Suppose that we further assume $\boldsymbol{E}_j \sim N_m(\mathbf{0}, \boldsymbol{\Sigma})$ for any subject j. Then,

$$\boldsymbol{Y}_j = \boldsymbol{\beta}^T \boldsymbol{Z}_j + \boldsymbol{E}_j \sim N_m\left(\boldsymbol{\beta}^T \boldsymbol{Z}_j, \boldsymbol{\Sigma}\right).$$

The maximum likelihood estimator is

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{(i)} &= \left(\boldsymbol{Z}^T \boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^T \boldsymbol{Y}_{(i)} \sim N_r\left(\boldsymbol{\beta}_{(i)}, \sigma_{ii}\left(\boldsymbol{Z}^T \boldsymbol{Z}\right)^{-1}\right), \\ \hat{\boldsymbol{\Sigma}} &= \frac{1}{n} \hat{\boldsymbol{E}}^T \hat{\boldsymbol{E}}, \end{aligned}$$

where $n\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{E}}^T \hat{\boldsymbol{E}} \sim W_m(\boldsymbol{\Sigma}, n-r)$. Further, $\hat{\boldsymbol{\Sigma}}$ is independent of $\hat{\boldsymbol{\beta}}$.

Regression Coefficients With Zero Constraints

Suppose that we want to test

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0} \text{ where } \boldsymbol{\beta}_{r\times m} = \begin{bmatrix} \boldsymbol{\beta}_1 & (q \times m) \\ \boldsymbol{\beta}_2 & ((r-q)\times m) \end{bmatrix}.$$

Under H_0, our model reduces to

$$\begin{aligned} \boldsymbol{Y}_{n\times m} &= \boldsymbol{Z}_1\boldsymbol{\beta}_1 + \boldsymbol{Z}_2\boldsymbol{\beta}_2 + \boldsymbol{E} \\ &= \boldsymbol{Z}_1\boldsymbol{\beta}_1 + \boldsymbol{E}. \end{aligned}$$

The MLE under H_0 is

$$\begin{aligned} \hat{\boldsymbol{\beta}}_1 &= \left(\boldsymbol{Z}_1^T\boldsymbol{Z}_1\right)^{-1}\boldsymbol{Z}_1^T\boldsymbol{Y}, \\ \hat{\boldsymbol{\Sigma}}_1 &= \frac{1}{n}\hat{\boldsymbol{E}}_1^T\hat{\boldsymbol{E}}_1, \end{aligned}$$

where $n\hat{\boldsymbol{\Sigma}}_1 \sim W_m\left(\boldsymbol{\Sigma}, n-q\right)$.

Likelihood Ratio

The likelihood ratio is

$$
\begin{aligned}
\Lambda &= \frac{L\left(\hat{\boldsymbol{\beta}}_1, \mathbf{0}, \hat{\boldsymbol{\Sigma}}_1\right)}{L\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}}\right)} \\
&= \frac{\det^{-n/2}\left(\hat{\boldsymbol{\Sigma}}_1\right) \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\left(\boldsymbol{Y}_j - \hat{\boldsymbol{\beta}}_1^T \boldsymbol{Z}_{1j}\right)^T \hat{\boldsymbol{\Sigma}}_1^{-1}\left(\boldsymbol{Y}_j - \hat{\boldsymbol{\beta}}_1^T \boldsymbol{Z}_{1j}\right)\right\}}{\det^{-n/2}\left(\hat{\boldsymbol{\Sigma}}\right) \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\left(\boldsymbol{Y}_j - \hat{\boldsymbol{\beta}}^T \boldsymbol{Z}_{j}\right)^T \hat{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{Y}_j - \hat{\boldsymbol{\beta}}^T \boldsymbol{Z}_{j}\right)\right\}} \\
&= \left(\frac{\det\left(\hat{\boldsymbol{\Sigma}}\right)}{\det\left(\hat{\boldsymbol{\Sigma}}_1\right)}\right)^{n/2},
\end{aligned}
$$

since $\sum_{j=1}^{n}\left(\boldsymbol{Y}_j - \hat{\boldsymbol{\beta}}^T \boldsymbol{Z}_{j}\right)^T \hat{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{Y}_j - \hat{\boldsymbol{\beta}}^T \boldsymbol{Z}_{j}\right) = nm$ and $\sum_{j=1}^{n}\left(\boldsymbol{Y}_j - \hat{\boldsymbol{\beta}}_1^T \boldsymbol{Z}_{1j}\right)^T \hat{\boldsymbol{\Sigma}}_1^{-1}\left(\boldsymbol{Y}_j - \hat{\boldsymbol{\beta}}_1^T \boldsymbol{Z}_{1j}\right) = nm$. Here $\Lambda^{2/n}$ is the Wilks' lambda.

LRT

Result 7.11

Let the multivariate multiple regression model hold where $\boldsymbol{Z}$ has full column rank, and the errors $\boldsymbol{E}$ have a normal distribution. Consider testing $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$. The likelihood ratio test rejects H_0 if

$$-2 \log \Lambda \quad = \quad -n \log \left[\frac{\det\left(\hat{\boldsymbol{\Sigma}}\right)}{\det\left(\hat{\boldsymbol{\Sigma}}_1\right)} \right] \text{ is too large.}$$

For a sufficiently large n, with Bartlett correction,

$$-\left[n - r - \frac{1}{2}\left(m - r + q + 2\right)\right] \log \left[\frac{\det\left(\hat{\boldsymbol{\Sigma}}\right)}{\det\left(\hat{\boldsymbol{\Sigma}}_1\right)} \right]$$

is approximately a chi-square distribution with $m\left(r - q\right)$ degrees of freedom.

Special Case: $m = 1$

If $m = 1$, the unrestricted MLE is

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{Z}^T \boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^T \boldsymbol{Y}, \qquad \hat{\sigma}^2 = \frac{1}{n}\left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}\right)^T \left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}\right),$$

and the MLE under H_0 is

$$\hat{\boldsymbol{\beta}}_1 = \left(\boldsymbol{Z}_1^T \boldsymbol{Z}_1\right)^{-1} \boldsymbol{Z}_1^T \boldsymbol{Y}, \qquad \hat{\sigma}_1^2 = \frac{1}{n}\left(\boldsymbol{Y} - \boldsymbol{Z}_1\hat{\boldsymbol{\beta}}_1\right)^T \left(\boldsymbol{Y} - \boldsymbol{Z}_1\hat{\boldsymbol{\beta}}_1\right).$$

Then,

$$\Lambda = \left(\frac{\det\left(\hat{\boldsymbol{\Sigma}}\right)}{\det\left(\hat{\boldsymbol{\Sigma}}_1\right)}\right)^{n/2} = \left(\frac{\left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}\right)^T \left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}\right)}{\left(\boldsymbol{Y} - \boldsymbol{Z}_1\hat{\boldsymbol{\beta}}_1\right)^T \left(\boldsymbol{Y} - \boldsymbol{Z}_1\hat{\boldsymbol{\beta}}_1\right)}\right)^{n/2}.$$

LRT When $m = 1$

The LRT rejects H_0 if

$$\Lambda = \left(\frac{\det\left(\hat{\boldsymbol{\Sigma}}\right)}{\det\left(\hat{\boldsymbol{\Sigma}}_1\right)}\right)^{n/2} = \left(\frac{\left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}\right)}{\left(\boldsymbol{Y} - \boldsymbol{Z}_1\hat{\boldsymbol{\beta}}_1\right)^T\left(\boldsymbol{Y} - \boldsymbol{Z}_1\hat{\boldsymbol{\beta}}_1\right)}\right)^{n/2}$$

is too small.

Result 7.6: F test

Let $\boldsymbol{Z}$ have full column rank and $\boldsymbol{e} \sim N_n\left(\boldsymbol{0}, \sigma^2\boldsymbol{I}\right)$. The LRT is equivalent to a test that rejects H_0 if

$$\frac{\left[\left(\boldsymbol{Y} - \boldsymbol{Z}_1\hat{\boldsymbol{\beta}}_1\right)^T\left(\boldsymbol{Y} - \boldsymbol{Z}_1\hat{\boldsymbol{\beta}}_1\right) - \left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}\right)\right] / (r-q)}{\left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}\right) / (n-r)}$$

is greater than $F_{r-q,n-r}(\alpha)$.

Prediction of Regression Function

Suppose that a new subject has the covariate value $\boldsymbol{z}_0$ and we want to predict the mean response.

- The predicted mean response is $\hat{\boldsymbol{\beta}}^T \boldsymbol{z}_0$.
- Under the normality assumption,

$$\hat{\boldsymbol{\beta}}^T \boldsymbol{z}_0 \sim N_m\left(\boldsymbol{\beta}^T \boldsymbol{z}_0, \boldsymbol{z}_0^T \left(\boldsymbol{Z}^T \boldsymbol{Z}\right)^{-1} \boldsymbol{z}_0 \boldsymbol{\Sigma}\right).$$

- For each $\boldsymbol{\beta}_{(i)}$,

$$\frac{\hat{\boldsymbol{\beta}}_{(i)}^T \boldsymbol{z}_0 - \boldsymbol{\beta}_{(i)}^T \boldsymbol{z}_0}{\sqrt{\boldsymbol{z}_0^T \left(\boldsymbol{Z}^T \boldsymbol{Z}\right)^{-1} \boldsymbol{z}_0 \sigma_{ii}}} \sim N_m\left(0, 1\right), \qquad \frac{n\hat{\sigma}_{ii}}{\sigma_{ii}} \sim \chi_{n-r}^2,$$

  where $\hat{\sigma}_{ii}$ is the ith diagonal element of $\hat{\boldsymbol{\Sigma}}$.

Confidence Interval For Regression Function

Hence,

$$\frac{\frac{\hat{\boldsymbol{\beta}}_{(i)}^T \boldsymbol{z}_0 - \boldsymbol{\beta}_{(i)}^T \boldsymbol{z}_0}{\sqrt{\boldsymbol{z}_0^T (\boldsymbol{Z}^T \boldsymbol{Z})^{-1} \boldsymbol{z}_0 \sigma_{ii}}}}{\sqrt{\frac{n\hat{\sigma}_{ii}}{\sigma_{ii}} / (n-r)}} = \frac{\hat{\boldsymbol{\beta}}_{(i)}^T \boldsymbol{z}_0 - \boldsymbol{\beta}_{(i)}^T \boldsymbol{z}_0}{\sqrt{\boldsymbol{z}_0^T \left(\boldsymbol{Z}^T \boldsymbol{Z}\right)^{-1} \boldsymbol{z}_0 n \hat{\sigma}_{ii} / (n-r)}} \quad \sim \quad t_{n-r}.$$

A $1-\alpha$ confidence interval for $\boldsymbol{\beta}_{(i)}^T \boldsymbol{z}_0$ is

$$\hat{\boldsymbol{\beta}}_{(i)}^T \boldsymbol{z}_0 \pm t_{n-r} \left(\frac{\alpha}{2}\right) \sqrt{\boldsymbol{z}_0^T \left(\boldsymbol{Z}^T \boldsymbol{Z}\right)^{-1} \boldsymbol{z}_0 \frac{n}{n-r} \hat{\sigma}_{ii}}.$$

You can also construct simultanous confidence intervals and Bonferroni confidence intervals.

Forecast New Response

Now we want to forecast the new response $\boldsymbol{Y}_0$ using $\boldsymbol{z}_0$.

- Under the independence and normality assumption,

$$\boldsymbol{Y}_0 - \hat{\boldsymbol{\beta}}^T \boldsymbol{z}_0 \quad \sim \quad N_m\left(\mathbf{0}, \left(1 + \boldsymbol{z}_0^T \left(\boldsymbol{Z}^T \boldsymbol{Z}\right)^{-1} \boldsymbol{z}_0\right) \boldsymbol{\Sigma}\right),$$

  which is independent of $n\hat{\boldsymbol{\Sigma}} \sim W_m\left(\boldsymbol{\Sigma}, n-r\right)$.
- A $1-\alpha$ prediction interval for Y_{0i} is

$$\boldsymbol{z}_0^T \hat{\boldsymbol{\beta}}_{(i)} \pm t_{n-r}\left(\frac{\alpha}{2}\right) \sqrt{\frac{n}{n-r} \hat{\sigma}_{ii} \left[1 + \boldsymbol{z}_0^T \left(\boldsymbol{Z}^T \boldsymbol{Z}\right)^{-1} \boldsymbol{z}_0\right]}.$$

- You can also construct simultanous confidence intervals and Bonferroni confidence intervals.