Multiple Linear Regression

1 Model and Least Square Estimates

Assume the responses and the explanatory variates satisfy the following model

$$Y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \ldots + \beta_r z_{ir} + \epsilon_i, \quad i = 1, \ldots, n,$$

where $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$.

The model can be represented in the vector form:

$$Y_i = \vec{z}_i^\top \vec{\beta} + \epsilon_i, \quad i = 1, \ldots, n,$$

where $\vec{z}_i = \begin{bmatrix} 1 \\ z_{i1} \\ \vdots \\ z_{ir} \end{bmatrix}$ and $\vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix}$.

In matrix form, we have

$$\vec{Y} = \boldsymbol{Z}\vec{\beta} + \vec{\epsilon},$$

where

$$\vec{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \ \boldsymbol{Z} = \begin{bmatrix} 1 & z_{11} & z_{12} & \ldots & z_{1r} \\ 1 & z_{21} & z_{22} & \ldots & z_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & z_{n2} & \ldots & z_{nr} \end{bmatrix}, \ \vec{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Our distributional assumption implies that $\vec{\epsilon} \sim \mathcal{N}_n(\vec{0}, \boldsymbol{I}_n)$. Throughout this course, we assume that $\text{rank}(\boldsymbol{Z}) = r + 1$, which implies that $\boldsymbol{Z}^\top \boldsymbol{Z}$ is invertible.

1.1 Least square estimates

Least square estimate is to minimize

$$S(\vec{b}) = \sum_{i=1}^{n} (Y_i - b_0 - b_1 z_{i1} - \ldots - b_r z_{ir})^2 = \|\vec{Y} - \boldsymbol{Z}\vec{b}\|^2.$$

Here we need two results in multivariate calculus regarding gradients:

- For any $\vec{a}$, $\nabla_{\vec{x}}(\vec{a}^\top \vec{x}) = \vec{a}$;
- For any symmetric $\boldsymbol{S}$, $\nabla_{\vec{x}}(\vec{x}^\top \boldsymbol{S}\vec{x}) = 2\boldsymbol{S}\vec{x}$.

One can verify that

$$S(\vec{b}) = (\vec{Y} - \boldsymbol{Z}\vec{b})^\top(\vec{Y} - \boldsymbol{Z}\vec{b}) = \|\vec{Y}\|^2 - 2(\boldsymbol{Z}^\top\vec{Y})^\top\vec{b} + \vec{b}^\top(\boldsymbol{Z}^\top\boldsymbol{Z})\vec{b}.$$

Then

$$\nabla_{\vec{b}} S(\vec{b}) = -2(\boldsymbol{Z}^\top\vec{Y}) + 2(\boldsymbol{Z}^\top\boldsymbol{Z})\vec{b} = \vec{0}.$$

This gives the least square estimate

$$\hat{\vec{\beta}} = (\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\vec{Y}.$$

1.2 Hat matrix, residuals, and sum-of-squares decomposition

Define the residuals of least square fitting:

$$\hat{\epsilon}_i := Y_i - \vec{z}_i^\top\hat{\vec{\beta}}, \quad i = 1, \ldots, n,$$

or equivalently,

$$\hat{\vec{\epsilon}} = \vec{Y} - \boldsymbol{Z}\hat{\vec{\beta}} = \vec{Y} - \boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\vec{Y} = (\boldsymbol{I} - \boldsymbol{H})\vec{Y},$$

where

$$\boldsymbol{H} = \boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top$$

is referred to as the **hat matrix**.

The fitted responses are denoted as

$$\hat{\vec{Y}} = \boldsymbol{Z}\hat{\vec{\beta}} = \boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\vec{Y} = \boldsymbol{H}\vec{Y}.$$

This also gives the relationship

$$\hat{\vec{\epsilon}} = \vec{Y} - \hat{\vec{Y}}.$$

In practice, we estimate σ^2 through the residuals

$$\hat{\sigma}^2 := \frac{1}{n-r-1}\sum_{i=1}^n \hat{\epsilon}_i^2 = \frac{1}{n-r-1}\|\hat{\vec{\epsilon}}\|^2 = \frac{1}{n-r-1}\|\vec{Y} - \boldsymbol{Z}\hat{\vec{\beta}}\|^2.$$

Here we list some useful properties of the hat matrix (homework)

- Both $\boldsymbol{H}$ and $\boldsymbol{I} - \boldsymbol{H}$ are symmetric;
- $\boldsymbol{H}^2 = \boldsymbol{H}$, $(\boldsymbol{I} - \boldsymbol{H})^2 = \boldsymbol{I} - \boldsymbol{H}$, $\boldsymbol{H}(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{0}$;
- All eigenvalues of $\boldsymbol{H}$ and $\boldsymbol{I} - \boldsymbol{H}$ are either 1 or 0;
- Both $\boldsymbol{H}$ and $\boldsymbol{I} - \boldsymbol{H}$ are positive semidefinite;
- $\boldsymbol{H}\boldsymbol{Z} = \boldsymbol{Z}$ and $(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Z} = \boldsymbol{0}$.

With these properties, one can verify the following properties:

- $\boldsymbol{Z}^\top\hat{\vec{\epsilon}} = \boldsymbol{Z}^\top(\boldsymbol{I} - \boldsymbol{H})\vec{Y} = \vec{0}$;
- $\hat{\vec{Y}}^\top\hat{\vec{\epsilon}} = \vec{Y}^\top\boldsymbol{H}(\boldsymbol{I} - \boldsymbol{H})\vec{Y} = 0$.

In particular, given the first column of $\boldsymbol{Z}$ is $\vec{1}_n$, the fact $\boldsymbol{Z}^\top \hat{\vec{\epsilon}} = \vec{0}$ implies that $\vec{1}_n^\top \hat{\vec{\epsilon}} = 0$.

Denote $\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$. Then we have

$$\|\vec{Y} - \overline{Y}\vec{1}_n\|^2 = \|\hat{\vec{Y}} - \overline{Y}\vec{1}_n + \hat{\vec{\epsilon}}\|^2 = \|\hat{\vec{Y}} - \overline{Y}\vec{1}_n\|^2 + \|\hat{\vec{\epsilon}}\|^2 + 2(\hat{\vec{Y}} - \overline{Y}\vec{1}_n)^\top \hat{\vec{\epsilon}} = \|\hat{\vec{Y}} - \overline{Y}\vec{1}_n\|^2 + \|\hat{\vec{\epsilon}}\|^2.$$

This decomposition can be rewritten in the form of sum-of-squares decomposition

$$\underbrace{\sum_{i=1}^n (Y_i - \overline{Y})^2}_{\text{Total sum of squares (TSS)}} \quad = \quad \underbrace{\sum_{i=1}^n (\hat{Y}_i - \overline{Y})^2}_{\text{Explained sum of squares (ESS)}} \quad + \quad \underbrace{\sum_{i=1}^n \hat{\epsilon}_i^2}_{\text{Residual sum of squares (RSS)}} \quad .$$

With this decomposition, we can define the R^2 statistic:

$$R^2 = \frac{ESS}{TSS}.$$

2 Sampling distributions and confidence intervals

The fact $\vec{Y} \sim \mathcal{N}_n(\boldsymbol{Z}\vec{\beta}, \sigma^2 \boldsymbol{I}_n)$ implies that $\hat{\vec{\beta}}$ is also of multivariate normal distribution. Furthermore,

$$\mathbb{E}[\hat{\vec{\beta}}] = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}\boldsymbol{Z}^\top \mathbb{E}[\vec{Y}] = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}\boldsymbol{Z}^\top (\boldsymbol{Z}\vec{\beta}) = \vec{\beta}$$

and

$$\mathrm{Cov}[\hat{\vec{\beta}}] = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}\boldsymbol{Z}^\top (\sigma^2 \boldsymbol{I}_n)\boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} = \sigma^2 (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}.$$

Then

$$\hat{\vec{\beta}} \sim \mathcal{N}_{r+1}(\vec{\beta}, \sigma^2 (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}).$$

This also gives the estimated covariance of $\hat{\vec{\beta}}$:

$$\widehat{\mathrm{Cov}}(\hat{\vec{\beta}}) = \hat{\sigma}^2 (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}.$$

2.1 One-at-a-time confidence intervals

Denote $\boldsymbol{\Omega} := (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \in \mathbb{R}(r+1) \times (r+1)$ of entries ω_{jk} for $0 \le j \le r,\ 0 \le k \le r$. Then

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma^2 \omega_{jj}).$$

Replacing σ^2 with $\hat{\sigma}^2$, we get

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\sqrt{\omega_{jj}}} \sim t_{n-r-1}.$$

Again, denote the $(1 - \frac{\alpha}{2})$-th quantile of t_{n-r-1} as $t_{n-r-1}(\frac{\alpha}{2})$, we get $(1-\alpha)$ confidence interval for any fixed $j = 0, \ldots, r$:

$$\beta_j \in \left[\hat{\beta}_j - \hat{\sigma}\sqrt{\omega_{jj}}\, t_{n-r-1}(\frac{\alpha}{2}), \hat{\beta}_j + \hat{\sigma}\sqrt{\omega_{jj}}\, t_{n-r-1}(\frac{\alpha}{2})\right].$$

2.2 Confidence region based simultaneous confidence intervals

The fact $\hat{\vec{\beta}} \sim \mathcal{N}_{r+1}(\vec{\beta}, \sigma^2(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1})$ implies

$$(\hat{\vec{\beta}} - \vec{\beta})^\top \left(\sigma^2(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}\right)^{-1} (\hat{\vec{\beta}} - \vec{\beta}) \sim \chi^2_{r+1}.$$

Replacing σ^2 with $\hat{\sigma}^2$,

$$(\hat{\vec{\beta}} - \vec{\beta})^\top \left(\hat{\sigma}^2(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}\right)^{-1} (\hat{\vec{\beta}} - \vec{\beta}) \sim (r+1)F_{r+1,n-r-1}.$$

This gives $(1-\alpha)$ confidence region for $\hat{\beta}$:

$$(\hat{\vec{\beta}} - \vec{\beta})^\top \left(\hat{\sigma}^2(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}\right)^{-1} (\hat{\vec{\beta}} - \vec{\beta}) \leq (r+1)F_{r+1,n-r-1}(\alpha),$$

where $F_{r+1,n-r-1}(\alpha)$ is the $(1-\alpha)$-th quantile of $F_{r+1,n-r-1}$.

By the extended Cauchy-Schwarz inequality, for any $\vec{a} \in \mathbb{R}^{r+1}$ and any $\vec{\beta}$ in the confidence region,

$$\begin{aligned}(\vec{a}^\top(\hat{\vec{\beta}} - \vec{\beta}))^2 &\leq (\hat{\vec{\beta}} - \vec{\beta})^\top \left(\hat{\sigma}^2(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}\right)^{-1} (\hat{\vec{\beta}} - \vec{\beta}) \vec{a}^\top(\hat{\sigma}^2(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1})\vec{a} \\ &\leq (r+1)F_{r+1,n-r-1}(\alpha)\vec{a}^\top(\hat{\sigma}^2(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1})\vec{a}.\end{aligned}$$

By choosing

$$\vec{a} = [1, 0, \ldots, 0]^\top, [0, 1, 0, \ldots, 0]^\top, \ldots, [0, 0, \ldots, 0, 1]^\top,$$

we have

$$(\hat{\beta}_j - \beta_j)^2 \leq (r+1)F_{r+1,n-r-1}(\alpha)\hat{\sigma}^2\omega_{jj},$$

which implies the simultaneous confidence intervals

$$\beta_j \in \left[\hat{\beta}_j - \hat{\sigma}\sqrt{\omega_{jj}}\sqrt{(r+1)F_{r+1,n-r-1}(\alpha)}, \hat{\beta}_j + \hat{\sigma}\sqrt{\omega_{jj}}\sqrt{(r+1)F_{r+1,n-r-1}(\alpha)}\right], \quad j = 0, 1, \ldots, r.$$

2.3 Bonferroni correction based confidence intervals

For each $j = 0, 1, \ldots, r$, we have obtained

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\sqrt{\omega_{jj}}} \sim t_{n-r-1}.$$

With Bonferroni correction, this gives the $(1-\alpha)$ simultaneous confidence intervals

$$\beta_j \in \left[\hat{\beta}_j - \hat{\sigma}\sqrt{\omega_{jj}} t_{n-r-1}\left(\frac{\alpha}{2(r+1)}\right), \hat{\beta}_j + \hat{\sigma}\sqrt{\omega_{jj}} t_{n-r-1}\left(\frac{\alpha}{2(r+1)}\right)\right], \quad j = 0, 1, \ldots, r.$$

3 Model comparison and F-tests

We are interested in comparing the model

$$Y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \ldots + \beta_r z_{ir} + \epsilon_i, \quad i = 1, \ldots, n,$$

and

$$Y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \ldots + \beta_q z_{iq} + \epsilon_i, \quad i = 1, \ldots, n,$$

for some $0 \le q < r$. In other words, if we denote

$$\vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \\ \hdashline \beta_{q+1} \\ \vdots \\ \beta_r \end{bmatrix} := \begin{bmatrix} \vec{\beta}_{(1)} \\ \hdashline \vec{\beta}_{(2)} \end{bmatrix}$$

and

$$\boldsymbol{Z} = \left[\begin{array}{cccc:ccc} 1 & z_{11} & \dots & z_{1q} & z_{1,q+1} & \dots & z_{1,r} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & \dots & z_{nq} & z_{n,q+1} & \dots & z_{n,r} \end{array}\right] := [\boldsymbol{Z}_{(1)}, \boldsymbol{Z}_{(2)}],$$

we want to compare the **full model**

$$\vec{Y} = \boldsymbol{Z}\vec{\beta} + \vec{\epsilon}$$

and the reduced model

$$\vec{Y} = \boldsymbol{Z}_{(1)}\vec{\beta}_{(1)} + \vec{\epsilon}.$$

This comparison is formulated as

$$H_0 : \beta_{q+1} = \ldots = \beta_r = 0,$$

or equivalently

$$H_0 : \vec{\beta}_{(2)} = \vec{0}.$$

3.1 F-tests

Let's first consider a more general problem: Let $\boldsymbol{C} \in \mathbb{R}^{(r-q)\times(r+1)}$ with $\text{rank}(\boldsymbol{C}) = r - q$. It is of interest to test

$$H_0 : \boldsymbol{C}\vec{\beta} = \vec{0}.$$

By $\hat{\vec{\beta}} \sim \mathcal{N}_{r+1}(\vec{\beta}, \sigma^2(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1})$, we have

$$\boldsymbol{C}\hat{\vec{\beta}} \sim \mathcal{N}_{r-q}(\boldsymbol{C}\vec{\beta}, \sigma^2\boldsymbol{C}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{C}^\top),$$

which further implies

$$\frac{1}{\sigma^2}(\boldsymbol{C}\hat{\vec{\beta}} - \boldsymbol{C}\vec{\beta})^\top \left(\boldsymbol{C}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{C}^\top\right)^{-1} (\boldsymbol{C}\hat{\vec{\beta}} - \boldsymbol{C}\vec{\beta}) \sim \chi^2_{r-q}.$$

Replacing σ^2 with $\hat{\sigma}^2$,

$$\frac{1}{\hat{\sigma}^2}(\boldsymbol{C}\hat{\vec{\beta}} - \boldsymbol{C}\vec{\beta})^\top \left(\boldsymbol{C}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{C}^\top\right)^{-1} (\boldsymbol{C}\hat{\vec{\beta}} - \boldsymbol{C}\vec{\beta}) \sim (r-q)F_{r-q,n-r-1}.$$

To test

$$H_0 : \boldsymbol{C}\vec{\beta} = \vec{0},$$

it suffices to compare

$$\frac{1}{\hat{\sigma}^2}(\boldsymbol{C}\hat{\vec{\beta}})^\top \left(\boldsymbol{C}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{C}^\top\right)^{-1} (\boldsymbol{C}\hat{\vec{\beta}})$$

and

$$(r-q)F_{r-q,n-r-1}(\alpha).$$

Coming back to the problem of model comparison, if we choose

$$\boldsymbol{C} = [\mathbf{0}_{(r-q)\times(1+q)}, \boldsymbol{I}_{r-q}],$$

then $H_0 : \vec{\beta}_{(2)} = \vec{0}$ is equivalent to $H_0 : \boldsymbol{C}\vec{\beta} = \vec{0}$. We denote

$$\hat{\vec{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_q \\ \hdashline \hat{\beta}_{q+1} \\ \vdots \\ \hat{\beta}_r \end{bmatrix} := \begin{bmatrix} \hat{\vec{\beta}}_{(1)} \\ \hdashline \hat{\vec{\beta}}_{(2)} \end{bmatrix}$$

and

$$(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} = \boldsymbol{\Omega} = \left[\begin{array}{c:c} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \hdashline \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{array}\right]$$

where $\Omega_{22} \in \mathbb{R}^{(r-q)\times(r-q)}$. Then F-test is equivalent to compare

$$\hat{\vec{\beta}}_{(2)}^\top \boldsymbol{\Omega}_{22}^{-1} \hat{\vec{\beta}}_{(2)}$$

and

$$(r-q)\hat{\sigma}^2 F_{r-q,n-r-1}(\alpha).$$

3.2 Formula of F-test by comparing residuals

Here we derive another equivalent formula for the F-tests. The derivation is involved and not required to master in this course, but it shows how to derive statistically insightful results with advanced algebraic techniques.

For the full model $\vec{Y} = \boldsymbol{Z}\vec{\beta} + \vec{\epsilon}$, we have the least square estimate and the residual

$$\hat{\vec{\beta}} = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}\boldsymbol{Z}^\top \vec{Y}, \quad \hat{\vec{\epsilon}} = (\boldsymbol{I} - \boldsymbol{H})\vec{Y}.$$

The aforementioned properties of hat matrices imply that

$$\|\hat{\vec{\epsilon}}\|^2 = \hat{\vec{\epsilon}}^\top \hat{\vec{\epsilon}} = \vec{Y}^\top (\boldsymbol{I} - \boldsymbol{H})^2 \vec{Y} = \vec{Y}^\top (\boldsymbol{I} - \boldsymbol{H})\vec{Y}$$

Similarly, for the reduced model

$$\vec{Y} = \boldsymbol{Z}_{(1)}\vec{\beta}_{(1)} + \vec{\epsilon},$$

we have the hat matrix

$$\boldsymbol{H}_{(red)} = \boldsymbol{Z}_{(1)}(\boldsymbol{Z}_{(1)}^\top \boldsymbol{Z}_{(1)})^{-1}\boldsymbol{Z}_{(1)}^\top,$$

the residuals

$$\hat{\vec{\epsilon}}_{(red)} = (\boldsymbol{I} - \boldsymbol{H}_{(red)})\vec{Y},$$

and

$$\|\hat{\vec{\epsilon}}_{(red)}\|^2 = \vec{Y}^\top (\boldsymbol{I} - \boldsymbol{H}_{(red)})\vec{Y}.$$

We are interested in deriving a formula for the “difference in sum of squares of residuals”:

$$\|\hat{\vec{\epsilon}}_{(red)}\|^2 - \|\hat{\vec{\epsilon}}\|^2.$$

To derive this formula, besides the aforementioned basic properties of the hat matrices, we also need the following fact:

$$\boldsymbol{H}\boldsymbol{H}_{(red)} = \boldsymbol{H}_{(red)}\boldsymbol{H} = \boldsymbol{H}_{(red)} \quad \text{Homework}.$$

Then

$$\begin{aligned}
\|\hat{\vec{\epsilon}}_{(red)}\|^2 - \|\hat{\vec{\epsilon}}\|^2 &= \vec{Y}^\top(\boldsymbol{H} - \boldsymbol{H}_{(red)})\vec{Y} \\
&= \vec{Y}^\top \boldsymbol{H}(\boldsymbol{I} - \boldsymbol{H}_{(red)})\boldsymbol{H}\vec{Y} \\
&= \hat{\vec{\beta}}^\top \boldsymbol{Z}^\top(\boldsymbol{I} - \boldsymbol{H}_{(red)})^2\boldsymbol{Z}\hat{\vec{\beta}} \\
&= \|(\boldsymbol{I} - \boldsymbol{H}_{(red)})\boldsymbol{Z}\hat{\vec{\beta}}\|^2
\end{aligned}$$

Since

$$\boldsymbol{Z}\hat{\vec{\beta}} = [\boldsymbol{Z}_{(1)}, \boldsymbol{Z}_{(2)}]\begin{bmatrix} \hat{\vec{\beta}}_{(1)} \\ \hline \hat{\vec{\beta}}_{(2)} \end{bmatrix} = \boldsymbol{Z}_{(1)}\hat{\vec{\beta}}_{(1)} + \boldsymbol{Z}_{(2)}\hat{\vec{\beta}}_{(2)}.$$

The fact $(\boldsymbol{I} - \boldsymbol{H}_{(red)})\boldsymbol{Z}_{(1)} = \boldsymbol{0}$ implies that

$$(\boldsymbol{I} - \boldsymbol{H}_{(red)})\boldsymbol{Z}\hat{\vec{\beta}} = (\boldsymbol{I} - \boldsymbol{H}_{(red)})\boldsymbol{Z}_{(2)}\hat{\vec{\beta}}_{(2)}.$$

Then

$$\begin{aligned}
\|\hat{\vec{\epsilon}}_{(red)}\|^2 - \|\hat{\vec{\epsilon}}\|^2 &= \|(\boldsymbol{I} - \boldsymbol{H}_{(red)})\boldsymbol{Z}_{(2)}\hat{\vec{\beta}}_{(2)}\|^2 \\
&= \hat{\vec{\beta}}_{(2)}^\top \boldsymbol{Z}_{(2)}^\top(\boldsymbol{I} - \boldsymbol{H}_{(red)})^2\boldsymbol{Z}_{(2)}\hat{\vec{\beta}}_{(2)} \\
&= \hat{\vec{\beta}}_{(2)}^\top \boldsymbol{Z}_{(2)}^\top(\boldsymbol{I} - \boldsymbol{H}_{(red)})\boldsymbol{Z}_{(2)}\hat{\vec{\beta}}_{(2)} \\
&= \hat{\vec{\beta}}_{(2)}^\top \boldsymbol{Z}_{(2)}^\top(\boldsymbol{I} - \boldsymbol{Z}_{(1)}(\boldsymbol{Z}_{(1)}^\top\boldsymbol{Z}_{(1)})^{-1}\boldsymbol{Z}_{(1)}^\top)\boldsymbol{Z}_{(2)}\hat{\vec{\beta}}_{(2)} \\
&= \hat{\vec{\beta}}_{(2)}^\top \left(\boldsymbol{Z}_{(2)}^\top\boldsymbol{Z}_{(2)} - \boldsymbol{Z}_{(2)}^\top\boldsymbol{Z}_{(1)}(\boldsymbol{Z}_{(1)}^\top\boldsymbol{Z}_{(1)})^{-1}\boldsymbol{Z}_{(1)}^\top\boldsymbol{Z}_{(2)}\right)\hat{\vec{\beta}}_{(2)}
\end{aligned}$$

Notice that

$$\boldsymbol{Z}^\top\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_{(1)}^\top \\ \boldsymbol{Z}_{(2)}^\top \end{bmatrix}[\boldsymbol{Z}_{(1)}, \boldsymbol{Z}_{(2)}] = \begin{bmatrix} \boldsymbol{Z}_{(1)}^\top\boldsymbol{Z}_{(1)} & \boldsymbol{Z}_{(1)}^\top\boldsymbol{Z}_{(2)} \\ \boldsymbol{Z}_{(2)}^\top\boldsymbol{Z}_{(1)} & \boldsymbol{Z}_{(2)}^\top\boldsymbol{Z}_{(2)} \end{bmatrix}$$

and recall that

$$\boldsymbol{\Omega} = (\boldsymbol{Z}^\top\boldsymbol{Z})^{-1} = \left[\begin{array}{c|c} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \hline \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{array}\right]$$

By Schur complement,

$$\boldsymbol{\Omega}_{22} = \boldsymbol{Z}_{(2)}^\top\boldsymbol{Z}_{(2)} - \boldsymbol{Z}_{(2)}^\top\boldsymbol{Z}_{(1)}(\boldsymbol{Z}_{(1)}^\top\boldsymbol{Z}_{(1)})^{-1}\boldsymbol{Z}_{(1)}^\top\boldsymbol{Z}_{(2)}$$

Therefore,

$$\|\hat{\vec{\epsilon}}_{(red)}\|^2 - \|\hat{\vec{\epsilon}}\|^2 = \hat{\vec{\beta}}_{(2)}^\top\boldsymbol{\Omega}_{22}^{-1}\hat{\vec{\beta}}_{(2)}.$$

In other words, the F-test to test $H_0 : \vec{\beta}_{(2)} = \vec{0}$ is equivalent to comparing

$$\frac{1}{\hat{\sigma}^2}(\|\hat{\vec{\epsilon}}_{(red)}\|^2 - \|\hat{\vec{\epsilon}}\|^2)$$

and

$$(r - q)F_{r-q,n-r-1}(\alpha).$$

4 Predictive inference

After obtaining the least square estimate $\hat{\vec{\beta}}$, if we have a new observation of the explanatory variates

$$\vec{z}_0^\top = [1, z_{01}, \dots, z_{0r}],$$

how to make inference about the unknown response $Y_0 = \vec{z}_0^\top \vec{\beta} + \epsilon_0$, where $\epsilon_0 \sim \mathcal{N}(0, \sigma^2)$ is independent of the "training data"?

4.1 Inference about the regression function $\mathbb{E}[Y_0|\vec{z}_0]$

For the regression function value $\mathbb{E}[Y_0|\vec{z}_0] = \vec{z}_0^\top \vec{\beta}$, we first study the sampling distribution of $\vec{z}_0^\top \hat{\vec{\beta}}$. The fact $\hat{\vec{\beta}} \sim \mathcal{N}_{r+1}(\vec{\beta}, \sigma^2(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1})$ gives

$$\vec{z}_0^\top \hat{\vec{\beta}} \sim \mathcal{N}(\vec{z}_0^\top \vec{\beta}, \sigma^2 \vec{z}_0^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \vec{z}_0).$$

Then

$$\frac{\vec{z}_0^\top \hat{\vec{\beta}} - \vec{z}_0^\top \vec{\beta}}{\sigma \sqrt{\vec{z}_0^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \vec{z}_0}} \sim \mathcal{N}(0, 1).$$

Replacing σ^2 with $\hat{\sigma}^2$, we have

$$\frac{\vec{z}_0^\top \hat{\vec{\beta}} - \vec{z}_0^\top \vec{\beta}}{\hat{\sigma} \sqrt{\vec{z}_0^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \vec{z}_0}} \sim t_{n-r-1}.$$

Then we have a $(1-\alpha)$ confidence interval for $\vec{z}_0^\top \vec{\beta}$:

$$\vec{z}_0^\top \vec{\beta} \in \left[\vec{z}_0^\top \hat{\vec{\beta}} - \hat{\sigma} t_{n-r-1}(\frac{\alpha}{2}) \sqrt{\vec{z}_0^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \vec{z}_0},\ \vec{z}_0^\top \hat{\vec{\beta}} + \hat{\sigma} t_{n-r-1}(\frac{\alpha}{2}) \sqrt{\vec{z}_0^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \vec{z}_0} \right].$$

4.2 Prediction interval for the new response Y_0

To find a prediction interval for $Y_0 = \vec{z}_0^\top \vec{\beta} + \epsilon$, we start with the sampling distribution of

$$Y_0 - \vec{z}_0^\top \hat{\vec{\beta}} = \vec{z}_0^\top \vec{\beta} + \epsilon_0 - \vec{z}_0^\top \hat{\vec{\beta}}.$$

Notice that $\hat{\vec{\beta}}$ and ϵ_0 are independent and normally distribution, we know $Y_0 - \vec{z}_0^\top \hat{\vec{\beta}}$ is normal. Moreover,

$$\mathbb{E}[Y_0 - \vec{z}_0^\top \hat{\vec{\beta}}] = \vec{z}_0^\top \vec{\beta} + 0 - \vec{z}_0^\top \vec{\beta}$$

and

$$\mathrm{var}[Y_0 - \vec{z}_0^\top \hat{\vec{\beta}}] = \mathrm{var}(\epsilon_0) + \mathrm{var}(\vec{z}_0^\top \hat{\vec{\beta}}) = \sigma^2 + \sigma^2 \vec{z}_0^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \vec{z}_0 = \sigma^2 (1 + \vec{z}_0^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \vec{z}_0).$$

Then

$$\frac{Y_0 - \vec{z}_0^\top \hat{\vec{\beta}}}{\sigma \sqrt{1 + \vec{z}_0^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \vec{z}_0}} \sim \mathcal{N}(0, 1)$$

Replacing σ^2 with $\hat{\sigma}^2$, we have

$$\frac{Y_0 - \vec{z}_0^\top \hat{\vec{\beta}}}{\hat{\sigma} \sqrt{1 + \vec{z}_0^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \vec{z}_0}} \sim t_{n-r-1}.$$

This gives

$$\mathbb{P}\left(-t_{n-r-1}(\alpha/2) \le \frac{Y_0 - \vec{z}_0^\top \hat{\vec{\beta}}}{\hat{\sigma} \sqrt{1 + \vec{z}_0^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \vec{z}_0}} \le t_{n-r-1}(\alpha/2) \right) = 1 - \alpha,$$

which gives the $(1-\alpha)$ prediction interval

$$Y_0 \in \left[\vec{z}_0^\top \hat{\vec{\beta}} - \hat{\sigma} t_{n-r-1}(\frac{\alpha}{2})\sqrt{1+\vec{z}_0^\top(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\vec{z}_0},\ \vec{z}_0^\top \hat{\vec{\beta}} + \hat{\sigma} t_{n-r-1}(\frac{\alpha}{2})\sqrt{1+\vec{z}_0^\top(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\vec{z}_0}\right].$$