Chapter 2. Simple Linear Regression

Jiho KWAK, Department of Geography, SNU

April 23, 2025

0 Some Basic Result in Statistics

0.1 Facts about Variance and Covariance

- $\begin{cases} Cov(X,X) = Var(X) \\ Cov(aX,cY) = acCov(X,Y) \\ Cov(a+X,c+Y) = Cov(X,Y) \end{cases}$
- If $X \perp\!\!\!\perp Y$, $Cov(X,Y) = 0$
- $Var(a_1X_1 + a_2X_2) = a_1^2Var(X_1) + a_2^2Var(X_2) + 2a_1a_2Cov(X_1,X_2)$

0.2 Normal Probability Distribution and Related Distribution

- Normal $X \sim N(\mu, \sigma^2)$ if the density function of a random variable X is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$$

- Standard Normal $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
- χ^2 Let $Z_1, \ldots, Z_n \sim N(0,1)$

$$V = Z_1 + \cdots + Z_n \sim \chi^2(n) \text{ and } \begin{cases} E(V) = n \\ Var(V) = 2n \end{cases}$$

- t Let $Z \perp\!\!\!\perp V$

$$T = \frac{Z}{\sqrt{V/n}} \sim t_n \text{ and } \begin{cases} E(T) = 0 \\ Var(T) = n/(n-2) \end{cases}$$

- F Let $V \sim \chi^2_m, W \sim \chi^2_n$ and $V \perp\!\!\!\perp W$

$$F = \frac{V/m}{W/n} \sim F_{m,n}$$

0.3 Statistical Estimation

Properties of Estimators

- unbiasedness: $E(\hat{\theta}_n) = \theta$
- consistency: $\lim_{n\to\infty} \mathbb{P}(|\hat{\theta}_n - \theta| \geq \epsilon) = 0$ for any $\epsilon > 0$
- minimum variance: $Var(\hat{\theta}_n) \leq Var(\theta_n^*)$ for all θ_n^*

1 Simple Linear Regression Model

- A simple linear regression model with a single regressor x is

$$y = \beta_0 + \beta_1 x + \epsilon, \quad \epsilon \overset{\text{iid}}{\sim} (0, \sigma^2)$$

- The primary goal is to do statistical inference on $f(x) = \beta_0 + \beta_1 x$
  1. Estimate β_0 and β_1
  2. Do testing $H_0 : \beta_1 = 0$

2 Least-Squares Estimation

2.1 Least-Squares Estimation of β_0 and β_1

- Estimate β_0 and β_1 to minimize

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^{n} \epsilon_i^2$$

- Least-squares normal equations

$$\frac{\partial S}{\partial \beta_0}\Big|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \implies n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\frac{\partial S}{\partial \beta_1}\Big|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0 \implies \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i$$

- LS estimators: $\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} \end{cases}$ where $S_{xx} = \sum (x_i - \bar{x})^2$ and $S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$
- Define the fitted values and the residuals as $\hat{y}_i = \beta_0 + \beta_1 x_i$ and $e_i = y_i - \hat{y}_i$

2.2 Properties of $\hat{\beta}_0$ and $\hat{\beta}_1$

- (Unbiasedness) $E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$
- $Var(\hat{\beta}_0) = \sigma^2 (\dfrac{1}{n} + \dfrac{\bar{x}^2}{S_{xx}})$ and $Var(\hat{\beta}_1) = \dfrac{\sigma^2}{S_{xx}}$

2.3 Properties of the Least-Squares fit, $\hat{y}$

1. $\sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n} e_i = 0 \iff \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$
2. The least squares fit always passes through the point $(\bar{x}, \bar{y})$.
3. $\sum_{i=1}^{n} x_i e_i = 0$
4. $\sum_{i=1}^{n} \hat{y}_i e_i = 0$

2.4 Estimation of σ^2

- As an unbiased estimator of σ^2, use

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} = MS_E$$

- Decomposition of the SS of y_i:

$$SS_T = SS_E + SS_R$$
$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

- $SS_E = SS_T - \hat{\beta}_1 S_{xy} \quad (df = n-2)$
- $SS_R = \hat{\beta}_1 S_{xy} = \hat{\beta}_1^2 \frac{S_{xx}}{S_{xy}} \cdot S_{xy} = \hat{\beta}_1^2 S_{xx} \quad (df = 1)$
- $\hat{\sigma}^2$ is called the standard error of regression, and is a model-dependent estimate of σ^2

2.5 Alternative Form of the Original Model

- Consider an alternative model

$$\begin{aligned} y_i &= (\beta_0 + \beta_1 \bar{x}) + \beta_1(x_i - \bar{x}) + \epsilon_i \\ &= \beta_0' + \beta_1(x_i - \bar{x}) + \epsilon_i \end{aligned}$$

- $\hat{\beta}_0' = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y}$ and $\hat{\beta}_1 = S_{xy}/S_{xx}$
- $Cov(\hat{\beta}_0', \hat{\beta}_1) = Cov(\bar{y}, \hat{\beta}_1) = 0$

3 Hypothesis Testing on the Slope

- Now we need the assumption $\epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$
- Then β_1 is distributed in $N(\beta_1, \sigma^2/S_{xx})$

3.1 t-Tests

- Typically, σ^2 is unknown.
- Use t-test

$$t_0 = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \overset{H_0}{=} \frac{\hat{\beta}_1}{\sqrt{MS_E/S_{xx}}}$$

- Reject the null hypothesis if

$$|t_0| > t_{\alpha/2,n-2}$$

3.2 F-Tests (ANOVA)

- F-test와 t-test는 simple regression에서 equivalent.
- Test statistic

$$F_0 = \frac{SS_R/1}{SS_E/(n-2)} = \frac{MS_R}{MS_E}$$

- $E(F_0) = \frac{E(MS_R)}{E(MS_E)} = \frac{\sigma^2 + \beta_1^2 S_{xx}}{\sigma^2} \overset{H_0}{=} 1$

4 Interval Estimation

4.1 Confidence Intervals on β_0 and β_1

- Under $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$$\frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \text{ and } \frac{\hat{\beta}_0 - \beta_0}{se(\hat{\beta}_0)} \sim t_{n-2}$$

where

$$se(\hat{\beta}_0) = \sqrt{MS_E(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}})} \text{ and } se(\hat{\beta}_1) = \sqrt{\frac{MS_E}{S_{xx}}}$$

- Therefore, a 95% confidence interval on β_1 is

$$\hat{\beta}_1 - t_{0.025,n-2} se(\hat{\beta}_0) \leq \beta_0 \leq \hat{\beta}_0 + t_{0.975,n-2} se(\hat{\beta}_1)$$

4.2 Confidence Intervals on σ^2

- Under $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$$\frac{(n-2)MS_E}{\sigma^2} \sim \chi^2(n-2)$$

- Therefore, a 95% confidence interval on σ^2 is

$$\frac{(n-2)MS_E}{\chi^2_{0.025,n-2}} \leq \sigma^2 \leq \frac{(n-2)MS_E}{\chi^2_{0.975,n-2}}$$

4.3 Interval Estimation of the Mean Response, $E(y|x_0)$

- Use the estimates $\hat{E(y|x_0)} = \hat{\mu}_{y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$
- $E(\hat{\mu}_{y|x_0}) = E(y|x_0)$
- $Var(\hat{\mu}_{y|x_0}) = \sigma^2(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}})$
- Since

$$\frac{\hat{\mu}_{y|x_0} - E(y|x_0)}{\sqrt{MS_E(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}})}} \sim t_{n-2}$$

  a 95% confidence interval on the mean response at the point $x = x_0$ is

$$\mu_{y|x_0} - t_{0.025,n-2}A \leq E(y|x_0) \leq \mu_{y|x_0} - t_{0.975,n-2}A$$

  where $A = \sqrt{MS_E(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}})}$

4.4 Prediction of New Observation

- The variability of $\hat{y}_0 - y_0$ is

$$\begin{aligned} E(\hat{y}_0 - y_0)^2 &= E[\hat{y}_0 - E(y_0)]^2 + E[y_0 - E(y_0)]^2 \\ &= \sigma^2(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}) + \sigma^2 \end{aligned}$$

- Therefore, a 95% prediction interval on a future observation at the point $x = x_0$ is

$$\hat{y}_0 - t_{0.025,n-2}B \leq y_0 \leq \hat{y}_0 - t_{0.975,n-2}B$$

  where $B = \sqrt{MS_E(1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}})}$

4.5 Coefficient of Determination, R^2

- The coefficient of determination is defined as

$$R^2 = \frac{SS_R}{SS_T} \quad (0 \leq R^2 \leq 1)$$

- $SS_E = 0$이면 $R^2 = 1$, $SS_E \simeq SS_T$ 이면 (즉, $\hat{y}_i \simeq \bar{y}$) $R^2 \simeq 0$

Properties of R^2

1. $R^2 = \dfrac{\hat{\beta}_1 S_{xx}}{S_{yy}} = r_{xy}^2$
2. $F = \dfrac{MS_R}{MS_E} = \dfrac{SS_R}{SS_E/n-2} = \dfrac{(n-2)R^2}{1-R^2}$

4.6 Regression through the Origin

$$y_i = \beta_1 x_i + \epsilon_i \quad \epsilon \overset{iid}{\sim} N(0, \sigma^2)$$

- $S(\beta_1) = \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2$
- $\hat{\beta}_1 = \dfrac{\sum x_i y_i}{\sum x_i^2}$
- $\hat{\sigma}^2 = \dfrac{SS_E}{n-1} = \dfrac{1}{n-1} \sum_{i=1}^{n} (y_i - \hat{\beta}_1 x_i)^2$

4.7 Least-Squares와 MLE의 관계

- Under assumption that $y_i \overset{iid}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$, the likelihood function is

$$\begin{aligned} L(\beta_0, \beta_1, \sigma^2) &= f(y_1) f(y_2) \cdots f(y_n) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2) \\ &= (\frac{1}{2\pi\sigma^2})^{n/2} \exp(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2) \end{aligned}$$

- Take log: $\ln L(\beta_0, \beta_1, \sigma^2) = -\dfrac{n}{2} \ln 2\pi - \dfrac{n}{2} \ln \sigma^2 - \dfrac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$
- $$\begin{cases} \left.\dfrac{\partial \ln L}{\partial \hat{\beta}_0}\right|_{\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2} = 0 & \Longrightarrow \quad \dfrac{1}{\hat{\sigma}^2} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 0 \\ \left.\dfrac{\partial \ln L}{\partial \hat{\beta}_1}\right|_{\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2} = 0 & \Longrightarrow \quad \dfrac{1}{\hat{\sigma}^2} \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 0 \\ \left.\dfrac{\partial \ln L}{\partial \hat{\sigma}^2}\right|_{\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2} = 0 & \Longrightarrow \quad -\dfrac{n}{2\hat{\sigma}^2} + \dfrac{1}{2\hat{\sigma}^4} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 0 \end{cases}$$
- The first two equations are same as the least-squares normal equations.
- From the last equation,

$$\begin{aligned} \hat{\sigma}_{ML}^2 &= \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \\ &= \frac{1}{n} SS_E \end{aligned}$$

Therefore, $E(\hat{\sigma}_{ML}^2) = \sigma^2 \cdot \dfrac{n-2}{n}$ ($\neq \sigma^2$, biased)

Chapter 3. Multiple Linear Regression

Jiho KWAK, Department of Geography, SNU

April 23, 2025

0 Some Basic Results in Matrix Algebra

0.1 Definitions and Properties

- Properties of transposed matrix:

$$(\mathbf{A}^T)^T = \mathbf{A}, \quad (\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T \quad (\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$$

- Symmetric matrix: $\mathbf{A}$ is said to be symmetric if $\mathbf{A} = \mathbf{A}^T$
- Orthogonal matrix: $\mathbf{A}$ is said to be orthogonal if $\mathbf{A}^T\mathbf{A} = \boldsymbol{I} \iff \mathbf{A}^{-1} = \mathbf{A}^T$
- Idempotent matrix: $\mathbf{A}$ is said to be idempotent if $\mathbf{A} = \mathbf{AA}$.
  An idempotent matrix is called a *projection matrix*.
- Quadratic form: Let $\mathbf{y}$ be a $k \times 1$ vector, and let $\mathbf{A}$ be a $k \times k$ matrix. The function

$$\mathbf{y}^T\mathbf{A}\mathbf{y} = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}y_iy_j$$

  is called a quadratic form.
- Positive-definite(p.d.): A symmetric matrix $\mathbf{A}$ is said to be positive-definite if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x}, \mathbf{x} \neq 0$.

0.2 Matrix Derivatives

- $\dfrac{\partial \mathbf{a}^T\mathbf{y}}{\partial y} = \mathbf{a}, \quad \dfrac{\partial \mathbf{y}^T\mathbf{y}}{\partial \mathbf{y}} = 2\mathbf{y}$
- $\dfrac{\partial \mathbf{a}^T\mathbf{A}\mathbf{y}}{\partial \mathbf{y}} = \mathbf{A}^T\mathbf{a}, \quad \dfrac{\partial \mathbf{y}^T\mathbf{A}\mathbf{y}}{\partial y} = \mathbf{A}\mathbf{y} + \mathbf{A}^T\mathbf{y}$ $(= 2\mathbf{A}\mathbf{y}$ when $\mathbf{A}$ is symmetric)

0.3 Expectations and Variances of Matrix

- $\mathbb{E}(\mathbf{a}^T\mathbf{y}) = \mathbf{a}^T\boldsymbol{\mu}$
- $E(\mathbf{A}\mathbf{y}) = \mathbf{A}\boldsymbol{\mu}$
- $Var(\mathbf{A}\mathbf{y}) = \mathbf{a}^T\boldsymbol{V}a$
- $Var(\mathbf{A}\mathbf{y}) = \mathbf{A}\boldsymbol{V}\mathbf{A}^T$ (if $\boldsymbol{V} = \sigma^2\boldsymbol{I}$, then $Var(\mathbf{A}\mathbf{y}) = \sigma^2\mathbf{A}\mathbf{A}^T$)

- $\mathbb{E}(\mathbf{y}^T\mathbf{y}) = tr(\mathbf{A}\boldsymbol{V}) + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}$ (if $\boldsymbol{V} = \sigma^2\boldsymbol{I}$, then $\mathbb{E}(\mathbf{y}^T\mathbf{y}) = \sigma^2 tr(\mathbf{A}) + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}$
- $Cov(\mathbf{Ax}, \mathbf{By}) = \mathbf{A}Cov(\mathbf{x}, \mathbf{y})\mathbf{B}^T$

1 Multiple Linear Regression Model

- The multiple regression model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i$$

  where $i = 1, 2, \ldots, n$.

- With matrix notation, the model is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$$

  where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \ \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix},$$

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_k \end{pmatrix}, \text{ and } \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

- $\mathbf{X}$ is an $n \times p(= k + 1)$ matrix, which is also known as the design matrix of the model.
- The assumption on the error is $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I}_n)$

2 Least-Squares Estimation

2.1 Least-Squares Estimation of $\hat{\boldsymbol{\beta}}$

- Wish to find $\hat{\boldsymbol{\beta}}$ that minimize

$$\begin{aligned} S(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \epsilon_i^2 \\ &= \boldsymbol{\epsilon}^T\boldsymbol{\epsilon} \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

- $S(\boldsymbol{\beta})$ can be expressed as

$$\begin{aligned} S(\boldsymbol{\beta}) &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}^\mathbf{y} - \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}^\mathbf{y} - 2\mathbf{y}^T\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \quad (\because \mathbf{y}^T\mathbf{X}\boldsymbol{\beta} \text{ is a scalar}) \end{aligned}$$

- LS estimators must satisfy

$$\left.\frac{\partial S}{\partial \boldsymbol{\beta}}\right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}^T + 2\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = 0 \iff \mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y}$$

  It is called the **least-squares normal equations**(p개의 equations).

- The least-squares estimators of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T X)^{-1}\mathbf{X}^T\mathbf{y}$, provided that $(\mathbf{X}^T\mathbf{X})^{-1}$ exists.

- The fitted regression model is

$$\begin{aligned} \hat{\mathbf{y}} &= \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\ &= \boldsymbol{H}\mathbf{y} \end{aligned}$$

  where the $n \times n$ matrix $\boldsymbol{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is called the **hat matrix**.

- $\boldsymbol{H}$의 성질

  1. Symmetric: $\boldsymbol{H}^T = \boldsymbol{H}$
  2. Idempotent: $\boldsymbol{H}\boldsymbol{H} = \boldsymbol{H}$
  3. Orthogonal projection of $Y \in \mathbb{R}^n$ onto $\Omega(\mathbf{X})$ (column space of $\mathbf{X}$)

- The residual is

$$\begin{aligned} \boldsymbol{e} &= \mathbf{y} - \hat{\mathbf{y}} \\ &= \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{y} - \boldsymbol{H}\mathbf{y} \\ &= (\boldsymbol{I} - \boldsymbol{H})\mathbf{y} \end{aligned}$$

- $\boldsymbol{I} - \boldsymbol{H}$의 성질

  1. Symmetric: $(\boldsymbol{I} - \boldsymbol{H})^T = \boldsymbol{I} - \boldsymbol{H}$
  2. Idempotent: $(\boldsymbol{I} - \boldsymbol{H})(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{I} - \boldsymbol{H}$
  3. Orthogonal complement projection of $Y \in \mathbb{R}^n$ onto $\Omega(\mathbf{X})$: $(\boldsymbol{I} - \boldsymbol{H})\mathbf{y} \in \Omega^\perp(\mathbf{X})$

- Under the assumption $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $Var(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}_n$,

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}) \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta} \end{aligned}$$

and

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}) &= Var((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}) \\ &= Var(\mathbf{A}\mathbf{y}) = \mathbf{A}\sigma\mathbf{A}\mathbf{A}^T \\ &= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \end{aligned}$$

2.2 Estimation of σ^2

- Note that

$$\begin{aligned} SS_E &= \sum_{i=1}^{n} e_i^2 \\ &= \boldsymbol{e}^T\boldsymbol{e} \\ &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}^{\mathbf{y}} - 2\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y} + \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{y}^{\mathbf{y}} - \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y} \quad (\because \text{LS normal equations}) \end{aligned}$$

or

$$\begin{aligned} SS_E = &= \boldsymbol{e}^T\boldsymbol{e} \\ &= (y - \hat{\mathbf{y}})^T(y - \hat{\mathbf{y}}) \\ &= \mathbf{y}^T(\boldsymbol{I} - \boldsymbol{H})^T(\boldsymbol{I} - \boldsymbol{H})\mathbf{y} \quad (\because \mathbf{y} - \hat{\mathbf{y}} = (\boldsymbol{I} - \boldsymbol{H})\mathbf{y}) \\ &= \mathbf{y}^T(\boldsymbol{I} - \boldsymbol{H})\mathbf{y} \end{aligned}$$

- SS_E has $n - (k+1)$ degree of freedom.
- The estimator of σ^2 is

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-k-1}$$

- MS_E is an unbiased estimator of σ^2, that is

$$E(MS_E) = \sigma^2$$

2.3 Gauss-Markov Theorem

Theorem. Suppose that we have a model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ and $Var(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$. (No normality assumption.) Let $\tilde{\phi} = \boldsymbol{d}^T\mathbf{y}$ be any unbiased linear estimator of $\phi = \boldsymbol{c}^T\boldsymbol{\beta}$ with an arbitrary vector $\boldsymbol{c}^T$. Then,

$$Var(\tilde{\phi}) \geq Var(\hat{\phi})$$

where $\hat{\phi} = \boldsymbol{c}^T\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$.
(i.e. Each element of $\hat{\boldsymbol{\beta}}$ is a minimum variance unbiased estimator since $\boldsymbol{c}^T$ is arbitrary.)

Proof. Since $E(\boldsymbol{d}^T\mathbf{y}) = \boldsymbol{d}^T E(\mathbf{y}) = \boldsymbol{d}^T\mathbf{X}\boldsymbol{\beta} = \boldsymbol{c}^T\boldsymbol{\beta}$ ($\because$ unbiased), $\boldsymbol{c}^T = \boldsymbol{d}^T\mathbf{X}$.
Need to show that $Var(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}) \leq Var(\boldsymbol{d}^T\mathbf{y})$.
(First proof)

$$\begin{aligned} Var(\boldsymbol{d}^T\mathbf{y}) &= Var(\boldsymbol{d}^T\mathbf{y} - \boldsymbol{c}^T\hat{\boldsymbol{\beta}} + \boldsymbol{c}^T\hat{\boldsymbol{\beta}}) \\ &= Var(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}) + Var(\boldsymbol{d}^T\mathbf{y} - \boldsymbol{c}^T\hat{\boldsymbol{\beta}}) + 2Cov(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}, \boldsymbol{d}^T\mathbf{y} - \boldsymbol{c}^T\hat{\boldsymbol{\beta}}) \\ &\geq Var(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}) + 2Cov(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}, \boldsymbol{d}^T\mathbf{y} - \boldsymbol{c}^T\hat{\boldsymbol{\beta}}) \end{aligned}$$

Since $Cov(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}, \boldsymbol{d}^T\mathbf{y} - \boldsymbol{c}^T\hat{\boldsymbol{\beta}}) = 0, Var(\tilde{\phi}) \geq Var(\hat{\phi})$.
(Another proof)

$$\begin{aligned} Var(\boldsymbol{d}^T\mathbf{y}) &= \boldsymbol{d}^T Var(\mathbf{y})\boldsymbol{d} \\ &= \sigma^2\boldsymbol{d}^T\boldsymbol{d} \\ Var(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}) &= Var(\boldsymbol{d}^T\mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \boldsymbol{d}^T\mathbf{X}Var(\hat{\boldsymbol{\beta}})\mathbf{X}^T\boldsymbol{d} \\ &= \sigma^2\boldsymbol{d}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{d} \\ &= \sigma^2\boldsymbol{d}^T\boldsymbol{H}\boldsymbol{d} \end{aligned}$$

and since $\boldsymbol{d}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{d} \geq 0$...

■

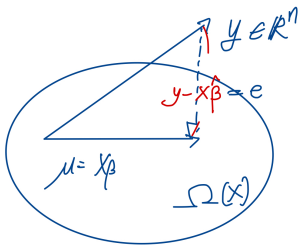
3 Hypothesis Testing

3.1 Geometric Interpretatoin of LS

- The model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = E(\mathbf{y}) + \boldsymbol{\epsilon}$ where

$$\begin{aligned} \mathbf{y} &= (y_1, \ldots, y_n)^T \in \mathbb{R}^n \\ E(\mathbf{y}) = \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} &= \text{ a linear combination of the columns of } \mathbf{X} \in \Omega(\mathbf{X}) \end{aligned}$$

- In the least-squares, we would like to find $\boldsymbol{\beta}$ such that minimizes $S(\boldsymbol{\beta})$.
  $\iff$ Find a vector $\mathbf{X}\hat{\boldsymbol{\beta}} \in \Omega(\mathbf{X})$ that is closest to $\mathbf{y}$.



- It can be obtained by making the difference $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ perpendicular to the space $\Omega(\mathbf{X})$

- Since $\mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_k$ are in the space $\Omega(\mathbf{X})$, it is required that

$$\begin{pmatrix} \mathbf{1}^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0} \\ \boldsymbol{X_1}^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0} \\ \vdots \\ \boldsymbol{X_k}^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0} \end{pmatrix} \Longrightarrow \mathbf{X}^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0} \Longleftrightarrow (\mathbf{X}^T\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y}$$

3.2 Test for Significance of Regression

- The appropriate hypotheses are

$$\begin{aligned} H_0: &\quad \beta_1 = \beta_2 = \cdots = \beta_k = 0 \\ H_1: &\quad \beta_j \neq 0 \text{ for at least one } j \end{aligned}$$

- Decompose SS_T into SS_R and SS_E

$$\begin{aligned} SS_T &= \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - 2\bar{y}\sum_{i=1}^{n} y_i + n\bar{y}^2 \\ &= \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \\ &= \mathbf{y}^T\mathbf{y} - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2 \\ SS_E &= \mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y} \\ &= \mathbf{y}^T(\boldsymbol{I} - \boldsymbol{H})\mathbf{y} \\ SS_R &= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n} \hat{y}_i - n\bar{y}^2 \\ &= \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y} - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2 \\ &= \hat{\mathbf{y}}^T\hat{\mathbf{y}} - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2 \\ &= \mathbf{y}^T\boldsymbol{H}\mathbf{y} - \frac{1}{n}\mathbf{y}^T\boldsymbol{J}_n\mathbf{y} \quad (\because \hat{\mathbf{y}} = \boldsymbol{H}\mathbf{y}) \\ &= \mathbf{y}^T(\boldsymbol{H} - \frac{1}{n}\boldsymbol{J}_n)\mathbf{y} \end{aligned}$$

Theorem. Let $\mathbf{A}$ be an $n \times n$ symmetric matrix with $rank(A) = k$, and $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is an $n \times n$ positive definite matrix. If $\mathbf{A}\boldsymbol{\Sigma}$ is idempotent,

$$\mathbf{y}^T\mathbf{A}\mathbf{y} \sim \chi^2_{k,\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}}$$

Fact. If p is idempotent, $rank(p) = tr(p)$.

- By the theorem and the fact,

$$\frac{SS_E}{\sigma^2} \sim \chi^2_{rank(\boldsymbol{I}-\boldsymbol{H}),\boldsymbol{\beta}^T\mathbf{X}^T(\boldsymbol{I}-\boldsymbol{H})\mathbf{X}\boldsymbol{\beta}} = \chi^2_{n-k-1}$$
$$\frac{SS_R}{\sigma^2} \sim \chi^2_{rank(\boldsymbol{H}-\frac{1}{n}\boldsymbol{J}_n),\boldsymbol{\beta}^T\mathbf{X}^T(\boldsymbol{H}-\frac{1}{n}\boldsymbol{J}_n)\mathbf{X}\boldsymbol{\beta}} = \chi^2_k \quad (\text{under } H_0)$$

Theorem. Let $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{A}_1, \mathbf{A}_2$ be $n \times n$ symmetric matrix. If $\mathbf{A}_1\Sigma\mathbf{A}_2^T = \mathbf{0}$,

$$\mathbf{y}^T\mathbf{A}_1\mathbf{y} \text{ and } \mathbf{y}^T\mathbf{A}_2\mathbf{y} \text{ are independent.}$$

- Let $\mathbf{A}_1 = \dfrac{\boldsymbol{I}-\boldsymbol{H}}{\sigma^2}, \mathbf{A}_2 = \dfrac{\boldsymbol{H}-\frac{1}{n}\boldsymbol{J}_n}{\sigma^2}$. Then by the theorem,

$$\mathbf{A}_1\Sigma\mathbf{A}_2^T = \frac{1}{\sigma^2}(\boldsymbol{I}-\boldsymbol{H})(\boldsymbol{H}-\frac{1}{n}\boldsymbol{J}_n) = \mathbf{0}$$

- Obtain F-statistic

$$F_0 = \frac{MS_R}{MS_E} = \frac{\frac{SS_R}{\sigma^2}/(p-1)}{\frac{SS_E}{\sigma^2}/(n-p)} \sim F_{p-1,n-p}$$

  under $H_0 : \boldsymbol{\beta} = \mathbf{0}$.

- Decision rule: reject H_0 if $F_0 > F_{\alpha,k,n-k-1}$

- ANOVA table:

Source of Variation	SS	DF	MS	F_0
Regression	$\hat{\boldsymbol{\beta}}^\top\boldsymbol{X}^\top\boldsymbol{y} - \frac{1}{n}(\sum_{i=1}^n y_i)^2$	k	MS_R	MS_R/MS_E
Error	$\boldsymbol{y}^\top\boldsymbol{y} - \hat{\boldsymbol{\beta}}^\top\boldsymbol{X}^\top\boldsymbol{y}$	$n-k-1$	MS_E	
Total	$\boldsymbol{y}^\top\boldsymbol{y} - \frac{1}{n}(\sum_{i=1}^n y_i)^2$	$n-1$		

3.3 Tests on Individual Regression Coefficients

- The appropriate hypotheses are

$$H_0 : \quad \beta_j = 0$$
$$H_1 : \quad \beta_j \neq 0$$

- Test statistic for the hypothesis is

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

  where C_{jj} is the diagonal element of $(\mathbf{X}^T\mathbf{X})^{-1}$ corresponding to $\hat{\beta}_j$ and $\hat{\sigma}^2 = MS_E$

- Decision rule: reject H_0 if $|t_0| > t_{\alpha/2,n-k-1}$

3.4 Coefficient of Multiple Determination

- The R^2 is defined as

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

- In general, R^2 always increases when a regressor is added to the model.
- Use an **adjusted** R^2 defined as

$$R^2_{\text{adj}} = 1 - \frac{n-1}{n-k-1} \cdot \frac{SS_E}{SS_R}$$

4 Confidence Intervals

4.1 C.I. on the Regression Coefficients

- The test statistic

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\sqrt{C_{jj}}}, \quad j = 0, 1, \ldots, k$$

  is distributed as t with $n-(k+1)$ df.

- $100(1-\alpha)\%$ C.I. for β_j is

$$\hat{\beta}_j - t_{\alpha/2,n-(k+1)} se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2,n-(k+1)} se(\hat{\beta}_j)$$

4.2 C.I. on the Mean Response at $\mathbf{x}_0$

- The fitted value at $\mathbf{x}_0 = (1, x_{01}, \ldots, x_{0k})^T$ is

$$\hat{y}_0 = \mathbf{x}_0^T \hat{\boldsymbol{\beta}}$$

- $E(\hat{y}_0) = \mathbf{x}_0^T \boldsymbol{\beta} = E(y|\mathbf{x}_0)$ and $Var(\hat{y}_0) = \mathbf{x}_0^T Var(\hat{\boldsymbol{\beta}})\mathbf{x}_0 = \sigma^2 \mathbf{x}_0^T (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0$
- The test statistic

$$\frac{\hat{y}_0 - E(\hat{y}|\mathbf{x}_0)}{\hat{\sigma}\sqrt{\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0}}$$

  is distributed as t with $n-(k+1)$ df.

- $100(1-\alpha)\%$ C.I. for $E(y|\mathbf{x}_0)$ is

$$\hat{y}_0 - t_{\alpha/2,n-(k+1)}\hat{\sigma}\sqrt{\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0} \leq E(y|\mathbf{x}_0) \leq \hat{y}_0 + t_{\alpha/2,n-(k+1)}\hat{\sigma}\sqrt{\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0}$$

5 Extram Sum of Squares Method

5.1 Introduction

- The goal of this section is to investigate the contribution of a subset of the regressor variable to the model.
- Let the vector of regression coefficients be partitioned as follows

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$$

where $\boldsymbol{\beta}_1$ is $(p-r) \times 1$ vector and $\boldsymbol{\beta}_2$ is $r \times 1$ vector.

- Wish to test the hypotheses

$$\begin{aligned} H_0: &\quad \boldsymbol{\beta}_2 = 0 \\ H_1: &\quad \boldsymbol{\beta}_2 \neq 0 \end{aligned}$$

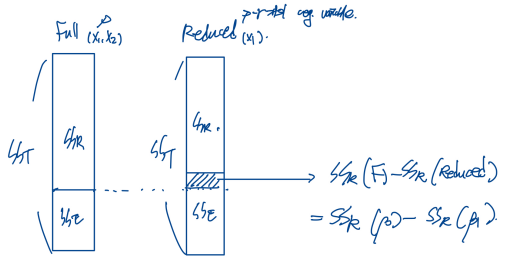
- Full model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$ under $H_0 \cup H_1$
- Reduced model: $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}$ under H_0
- In the reduced model, the (uncorrected) model sum of squares is

$$SS_R(\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}_1^T \mathbf{X}_1^T \mathbf{y}$$

- The **extra sum of squres due to** $\boldsymbol{\beta}_2$ is defined as

$$SS_R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta}_1)$$

i.e. The SS_R due to $\boldsymbol{\beta}_2$ given that $\boldsymbol{\beta}_1$ is already in the model.



- The test statistic is

$$F_0 = \frac{SS_R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/r}{SS_E(\boldsymbol{\beta})/(n-k-1)}$$

follows the distribution $F_{r,n-k-1}$ under the $H_0: \boldsymbol{\beta}_2 = 0$

5.2 Partial F-test

- Consider the full model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

- The hypothesis is

$$\begin{aligned} H_0 : &\quad \beta_2 = 0 \\ H_1 : &\quad \beta_2 \neq 0 \end{aligned}$$

- $SS_T = SS_R(\beta_1, \beta_2|\beta_0) + SS_E$
- The extra SS due to β_2 is

$$\begin{aligned} SS_R(\beta_2|\beta_1, \beta_0) &= SS_R(\beta_1, \beta_2, \beta_0) - SS_R(\beta_1, \beta_0) \text{ (uncorrected ver.)} \\ &= SS_R(\beta_1, \beta_2|\beta_0) - SS_R(\beta_1|\beta_0) \text{ (corrected ver.)} \end{aligned}$$

- The term $SS_R(\beta_1, \beta_2|\beta_0)$ can be obtained from SS_R of the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ and the term $SS_R(\beta_1|\beta_0)$ can be obtained from SS_R of the model $y = \beta_0 + \beta_1 x_1 + \epsilon$.
- The test statistic is

$$F_0 = \frac{SS_R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)/1}{SS_E(F)/(n-k-1)}$$

5.3 Testing the General Linear Hypothesis (1)

- Consider the full model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

- The hypothesis is

$$\begin{aligned} H_0 : &\quad \beta_1 = \beta_3 \\ H_0 : &\quad \boldsymbol{T\beta} = 0 \end{aligned}$$

  where $\boldsymbol{T} = (0, 1, 0, -1)$.

- The extra sum of squres due to the hypothesis is

$$SS_H = SS_{E(RM)} - SS_{E(FM)}$$

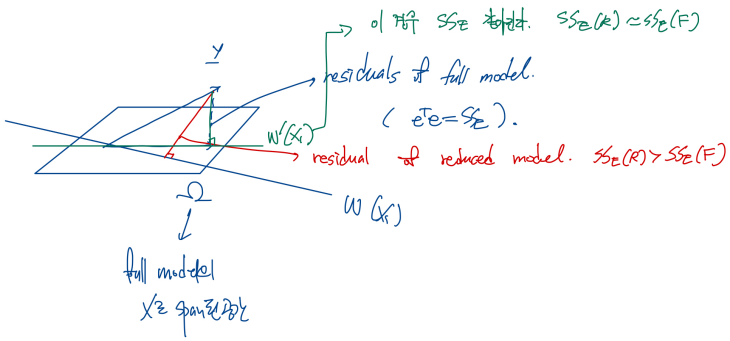
- The df is $(n-3) - (n-4) = 1$. (i.e. $\boldsymbol{T\beta}$의 independent한 row 수)

- The test statistic is

$$F_0 = \frac{SS_H/1}{SS_{E(FM)}/(n-k-1)}$$

5.4 Testing for the General Linear Hypothesis (2)

- Full model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$ under $H_0 \cup H_1$ where $\mathbf{X}_1$ is $n \times (p-r)$ matrix and $\mathbf{X}_2$ is $n \times r$ matrix.
- Reduced model: $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}$ under H_0



- $SS_E(R) \simeq SS_E(F)$ 이면 $\boldsymbol{\beta}_2$는 $\mathbf{y}$를 설명하는 데 도움이 되지 않는다.
- Let ω be a reduced model and Ω be a full model.

$$SS_E(F) = \mathbf{y}^T(\boldsymbol{I} - \boldsymbol{P}_\Omega)\mathbf{y}$$
$$SS_E(R) = \mathbf{y}^T(\boldsymbol{I} - \boldsymbol{P}_\omega)\mathbf{y}$$

  where

$$\boldsymbol{P}_\Omega = \text{projection matrix of } \Omega = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$
$$\boldsymbol{P}_\omega = \text{projection matrix of } \omega = \mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T$$

- $SS_H = SS_E(R) - SS_E(F) = \mathbf{y}^T(\boldsymbol{P}_\Omega - \boldsymbol{P}_\omega)\mathbf{y}$
- Since $(\boldsymbol{P}_\Omega - \boldsymbol{P}_\omega)$ is idempotent, under the H_0

$$\frac{SS_E(F)}{\sigma^2} \sim \chi^2_r$$

- SS_E and SS_H are independent. ($\because \frac{\boldsymbol{P}_\Omega - \boldsymbol{P}_\omega}{\sigma^2}\sigma^2\boldsymbol{I}\frac{\boldsymbol{I} - \boldsymbol{P}_\Omega}{\sigma^2} = 0$)
- Under the H_0, the test statistic

$$F_0 = \frac{SS_H/r}{SS_{E(FM)}/(n-k-1)} \sim F_{r,n-p}$$