통계학과 2020-17530 곽지호 회귀분석 및 실습 숙제 2

3.2번. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$

a. $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta$

$y = X\beta + \varepsilon$. model. $\varepsilon \sim N(0, \sigma^2 I)$

$\hat{y} = X\hat{\beta}$. how to find $\hat{\beta}$? $E(y) = X\beta$.

$(y - X\beta)^T(y - X\beta)$

$= (y^T - \beta^T X^T)(y - X\beta)$

$= y^T y - \beta^T X^T y - y^T X\beta + \beta^T X^T X\beta$

$= y^T y - 2y^T X\beta + \beta^T X^T X\beta$

미분. $-2X^T y + X^T X\beta = 0$.

$\hat{\beta} = (X^T X)^{-1} X^T y$.

$E(\hat{\beta}) = E((X^T X)^{-1} X^T y)$

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

$= \beta$.

$Var(\hat{\beta}) = Var(\underbrace{(X^T X)^{-1} X^T}_{A} y)$,

$= (X^T X)^{-1} X^T Var(A) X (X^T X)^{-1}$

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

① $\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y = Hy$.

$H = X(X^T X)^{-1} X^T$.

② $e = y - \hat{y} = y - Hy = (I - H)y$.

③ $SS_E = SS_T - SS_R$.

$= \sum_{i=1}^{n} e_i^2$

$= e^T e$

$= (y - X\hat{\beta})^T (y - X\hat{\beta}) = y^T y - y^T X\hat{\beta}$ or $\hat{\beta}^T X^T y$.

$= ((I-H)y)^T ((I-H)y) = y^T (I-H)^T (I-H) y = y^T (I-H) y$.

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

$$SST = \sum (y-\bar{y})^2$$

$$= \sum(y^2 - 2y\bar{y} + \bar{y}^2) = \sum y^2 - 2\underbrace{\bar{y}}_{\frac{1}{n}\sum y} \sum y + n\underbrace{\bar{y}^2}_{n\cdot\left(\frac{1}{n}\sum y\right)^2}$$

$$= y^T y - \frac{1}{n}\left(\sum y\right)^2$$

$$SSR = SST - SSE.$$

$$= \sum (\hat{y}_i - \bar{y})^2$$

$$= \hat{y}^T \hat{y} + ?$$

$$= \hat{\beta}^T X^T y + \frac{1}{n}\left(\sum y\right)^2$$

$$= y^T H y - \frac{1}{n} y^T J_n y. = y^T (H - \tfrac{1}{n} J_n) y.$$

$$* \; \hat{y}^T \hat{y} = (\hat{\beta}^T X^T) X \hat{\beta}.$$

$$= \hat{\beta}^T X^T y.$$

$$= (Hy)^T (Hy) = y^T H H y = y^T H y.$$

$*$ thm

$$y^T A y \sim \chi^2_{\text{rank}(A),\, \mu^T A \mu}.$$

$$\Rightarrow \frac{SSE}{\sigma^2} = \frac{y^T(I-H)y}{\sigma^2} \sim \chi^2\left(\underbrace{\text{rank}\left(\frac{I-H}{\sigma^2}\right)}_{\substack{\text{rank}(I-H) \\ \| \\ tr(I-H) = n-k-1.}},\ \underbrace{\beta^T X^T \frac{(I-H)}{\sigma^2} X\beta}_{\frac{1}{\sigma^2}\beta^T X^T \underbrace{(I-H)X}_{X - HX = 0}\beta.}\right)$$

$$* \; tr(H) = tr(X(X^T X)^{-1} X^T) = tr(I_p)$$

$$\Rightarrow \frac{SSR}{\sigma^2} = y^T \frac{(H - \frac{1}{n}J_n)}{\sigma^2} y \sim \chi^2\left(\underbrace{\text{rank}\left(\frac{H - \frac{1}{n}J_n}{\sigma^2}\right)}_{\substack{\text{rank}(H - \frac{1}{n}J_n) \\ \| \\ tr(H - \frac{1}{n}J_n) = p-1}},\ \beta^T X^T \frac{(H - \frac{1}{n}J_n)}{\sigma^2} X\beta\right)$$

Idempotent?

$$(H - \tfrac{1}{n}J_n)(H - \tfrac{1}{n}J_n)$$

$$= H - \tfrac{1}{n}J_n H - \tfrac{1}{n}HJ_n + \tfrac{1}{n^2}J_n J_n$$

$$\frac{1}{n}J_n = 1(1^T 1)^{-1} 1^T.$$

$$\frac{1}{n^2} J_n J_n = 1(\cancel{1^T 1})^{-1} \cancel{1^T 1}(1^T 1)^{-1} 1^T$$

$$= \frac{1}{n}J_n$$

$$\frac{1}{n}HJ_n = \underbrace{H \cdot 1}(1^T 1)^{-1} 1^T$$

$$= 1(1^T 1)^{-1} 1^T.$$

$$\frac{1}{n}J_n H = 1(1^T 1)^{-1} \underbrace{1^T H}$$

$$H \cdot [1 \cdots x_k] = [1 \cdots x_k]$$

$$HX = X.$$

$$X(\cancel{X^T X})^{-1} \cancel{X^T X}.$$

$$H1 = 1.$$

$$1^T H = 1^T.$$

Indep? $\mathrm{Cov}\left(\frac{y^T(I-H)y}{\sigma^2}, \frac{y^T(H-\frac{1}{n}J_n)y}{\sigma^2} \right) = 0.$

$\Leftrightarrow \quad \frac{(I-H)}{\cancel{\sigma^2}} \cdot \cancel{\sigma^2} I \cdot \frac{(H-\frac{1}{n}J_n)}{\sigma^2} = 0.$

$\frac{1}{\sigma^2}(I-H)(H-\frac{1}{n}J_n) = \frac{1}{\sigma^2}\left(H-H-\frac{1}{n}J_n+\frac{1}{n}J_n \right) = 0.$

$\Rightarrow \quad F_0 = \frac{SS_R/k}{SS_E/n-k-1} \sim F_{k,n-k-1}$ under H_0. $H_0: \beta_1 = \cdots = \beta_k = 0$

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

General Hypothesis

Full : $y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon. = X\beta + \varepsilon.$

Reduced : $y = \underline{\beta_1 X_1 + \varepsilon}.$

$\hookrightarrow \quad \hat{\beta}_1 = (X_1^T X_1)^{-1} X_1^T y.$

Ch4. Model Adequacy Checking.

Assumptions (Single Linear Reg.를 생각)

① Linearity

② Constant Variance

③ Normality

④ Independence

error에 대한 assumptions

$y = X\beta + \varepsilon \Rightarrow y = X\hat{\beta} + e$

$= \hat{y} + e$

ε와 e의 비교

②-④ 만족?

Residual의 성질

① sample mean $= 0$. $\bar{e} = \frac{\sum e_i}{n} = 0$.

$\mathbf{1}$과 e의 내적.

(e는 $X\hat{\beta}$에 수직이므로.)

$\Omega(X)$

+) $\sum e_i x_{i1} = \cdots = \sum e_i x_{ik} = 0$.

$\Rightarrow \sum e_i \hat{\mu}_i = 0$.

$\hat{\mu} = X\hat{\beta}$.

$e_1 x_{11} + \cdots + e_n x_{n1} = 0$.

$X = \begin{bmatrix} 1 & \cdots \end{bmatrix}$ $\hat{\beta} = \begin{pmatrix} 1 \\ \hat{\beta}_1 \\ \hat{\beta}_k \end{pmatrix} \to n \times 1$

$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & & \ddots & \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}$ $\rightarrow \begin{pmatrix} 1 + x_{11}\hat{\beta}_1 + \cdots + x_{1k}\hat{\beta}_k \\ \vdots \\ 1 + x_{n1}\hat{\beta}_1 + \cdots + x_{nk}\hat{\beta}_k \end{pmatrix} = \hat{\mu}$.

$\sum e_i \hat{\mu}_i$

$= e_1\hat{\mu}_1 + \cdots + e_n\hat{\mu}_n$

$= e_1(1 + x_{11}\hat{\beta}_1 + \cdots + x_{1k}\hat{\beta}_k) + \cdots + e_n(1 + x_{n1}\hat{\beta}_1 + \cdots + x_{nk}\hat{\beta}_k)$

$= \underbrace{(e_1 + \cdots + e_n)}_{0} + \hat{\beta}_1\underbrace{(e_1 x_{11} + \cdots + e_n x_{n1})}_{0} + \cdots + \hat{\beta}_k(e_1 x_{1k} + \cdots + e_n x_{nk})$

$*$ $E(e) = (I-H)E(y)$

$= (I-H)X\beta$

$= (X - HX)\beta = 0$.

$*$ $\mathrm{Cov}(\hat{\mu}, e) = \mathrm{Cov}(X\hat{\beta}, e)$

$\hat{\mu} = \hat{y}$

$= \mathrm{Cov}(Hy, (I-H)y)$

$= H\,\underbrace{\mathrm{Cov}(y,y)}_{\sigma^2 I}(I-H)^T$

$= \sigma^2 H(I-H) = 0$

(fitted value는 residual의 정보 가지지 X)

$\Rightarrow$ residual plot 그렸을 때 no pattern!

$*$ $\mathrm{Var}(e) = \mathrm{Var}((I-H)y)$ $\leftarrow$ estimated by MSE. $= \frac{SSE}{n-k-1} = \frac{\sum e_i^2}{n-k-1} = \frac{\sum(e_i - \bar{e})^2}{n-k-1}$

residual의 표본분산.

$= (I-H)\,\mathrm{Cov}(y,y)\,(I-H)^T$

$= \sigma^2(I-H)$. $= \sigma^2 \begin{bmatrix} 1-h_{11} & -h_{12} & \cdots & -h_{1n} \\ & 1-h_{12} & & \\ & & \ddots & 1-h_{nn} \end{bmatrix}$

$\Rightarrow$ $\begin{cases} \mathrm{Var}(e_i) = \sigma^2(1-h_{ii}). \\ \mathrm{Cov}(e_i, e_j) = -\sigma^2 h_{ij} \neq 0. \end{cases}$

Note $\begin{cases} \mathrm{Var}(\varepsilon_i) = \sigma^2 \\ \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0. \end{cases}$

About h_{ii} "leverage point"

Consider SLM. $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.

$H = X(X^TX)^{-1}X^T$.

$$h_{ii} = \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}.$$

$\hookrightarrow \frac{1}{n} \le h_{ii} \le 1$

pf) (첫번째 $\le$) $\frac{(x-\bar{x})^2}{S_{xx}} \ge 0$.

(두번째 $\le$) $H = H^2$.

$$\begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} \begin{bmatrix} h_1 & \cdots & h_n \end{bmatrix}$$

$$h_{ii} = \sum_{j=1}^{n} h_{ij}^2 = h_{ii}^2 + \sum_{j \ne i} h_{ij}^2$$

$$\Leftrightarrow 1 = h_{ii} + \underbrace{\frac{\sum h_{ij}^2}{h_{ii}}}_{\ge 0}.$$

Types of Residuals

① Standardized : $d_i = \frac{e_i}{\sqrt{MSE}}$ $\quad i = 1, \dots, n$ $\quad \widehat{Var}(e_i) = \hat{\sigma}^2 = MSE$.

② Studentized : $r_i = \frac{e_i}{\sqrt{MSE(1-h_{ii})}}$ $\quad i = 1, \dots, n$ $\quad \widehat{Var}(e_i) = \hat{\sigma}^2(1-h_{ii}) = MSE(1-h_{ii})$

③ PRESS : $e_{(i)} = y_i - \hat{y}_{(i)}$ $\quad \hat{y}_{(i)}$ = predicted value of the ith obs without the ith obs.

$= \frac{e_i}{1-h_{ii}}$

(why? y_i가 unusual 하면 $\hat{y}_i$ 에도 영향.
즉 e_i, d_i 도 영향 → not a good measure of assumption checking)

PRESS Residual

의미 $Var(e_{(i)}) = \left(\frac{1}{1-h_{ii}}\right)^2 Var(e_i)$

$= \left(\frac{1}{1-h_{ii}}\right)^2 \sigma^2 (1-h_{ii}) = \frac{\sigma^2}{1-h_{ii}}$ 이다.

scale

$\hookrightarrow \frac{e_{(i)}}{\sqrt{Var(e_{(i)})}} = \frac{e_i / 1-h_{ii}}{\sqrt{\hat{\sigma}^2/(1-h_{ii})}} = \frac{e_i}{\sqrt{MSE(1-h_{ii})}} = r_i$

$\hookrightarrow$ studentized residual.

$e_{(i)} = \frac{e_i}{1-h_{ii}}$ 임을 보이자.

pf) $e_{(i)} = y_i - \hat{y}_{(i)}$

$= y_i - \underline{x_i}^T \hat{\beta}_{(i)}$ $\quad * \ \underline{x_i} = (1, x_{i1} \dots x_{ik})^T$

$= y_i - \underline{x_i}^T (X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T \underline{y}_{(i)}$

$\hookrightarrow$ Note $(X^TX)^{-1} + \frac{(X^TX)^{-1} \underline{x_i}\, \underline{x_i}^T (X^TX)^{-1}}{1-h_{ii}}$.

$$= y_i - \underbrace{x_i^T (X^TX)^{-1} X_{(i)}^T y_{(i)}}_{A} - \left(\underbrace{\boxed{x_i^T (X^TX)^{-1} \underline{x_i}}}_{h_{ii}}\ \underbrace{x_i^T (X^TX)^{-1} X_{(i)}^T \underline{y}_{(i)}}_{A} \right) / (1-h_{ii})$$

$$e_{(i)} = y_i - A - \frac{h_{ii}}{1-h_{ii}} A$$

$$= \frac{1}{1-h_{ii}} \left((1-h_{ii}) y_i - A(1-\cancel{h_{ii}}) - \cancel{h_{ii} A} \right)$$

$$= \frac{1}{1-h_{ii}} \left((1-h_{ii}) y_i - \underline{x_i}^T (X^T X)^{-1} \underline{X_{(i)}}^T \underline{y_{(i)}} \right)$$

Note $X^T y = X_{(i)}^T y_{(i)} + \underline{x_i} y_i$

$\Leftrightarrow X_{(i)}^T y_{(i)} = X^T y - \underline{x_i} y_i$

$$= \frac{1}{1-h_{ii}} \left((1-h_{ii}) y_i - \underline{x_i}^T \underbrace{(X^T X)^{-1} X^T y}_{\hat\beta.} + \underbrace{x_i^T (X^T X)^{-1} x_i}_{h_{ii}} y_i \right)$$

$$= \frac{1}{1-h_{ii}} \left(y_i - \underbrace{x_i^T \hat\beta}_{\overset{\shortparallel}{\hat y_i}} \right) = \frac{e_i}{1-h_{ii}}$$

Assumption check

① Linearity : (1) plot residuals vs $\hat y_i$. $\Rightarrow$ check if residuals are <u>uniformly distributed</u> around 0

(2) plot y vs x

(3) Lack of fit

② Equal Variance : plot residuals vs $\hat y_i$. $\Rightarrow$ check if residuals are <u>uniformly distributed</u> around 0

$\hookrightarrow$ $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$.

Suppose that $\mathrm{Var}(\varepsilon_i) = \sigma_i^2$ (not equal variance).

Then $\mathrm{Var}(e) = (I-H)\,\Sigma\,(I-H)$

$$\mathrm{Var}(e_i) = (1-h_{ii})^2 \sigma_i^2 + \sum_{j \neq i} h_{ij}^2 \sigma_j^2 = E(e_i^2)$$

$(\because E(e_i) = 0)$.

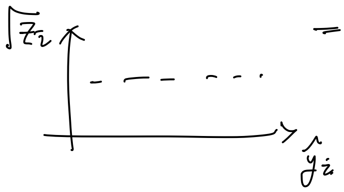
Consider the quantity $z_i = \dfrac{e_i^2}{1-h_{ii}}$

$$E(z_i) = \frac{1}{1-h_{ii}} E(e_i^2) = \frac{1}{1-h_{ii}} \left((1-h_{ii})^2 \sigma^2 + \sum_{j \neq i} h_{ij}^2 \sigma_j^2 \right)$$

If variances are all equal to σ^2, $E(z_i) = \dfrac{\sigma^2}{1-h_{ii}} \Big(\underbrace{(1-h_{ii})^2 + \textstyle\sum_j h_{ij}^2}_{\overset{\shortparallel}{1-h_{ii}}} \Big).$ $(\because (I-H)$ is idempotent$)$

$= \sigma^2$

$\Rightarrow$ z_i의 제곱값은 상수가 됨.



③ Normality : use normal probability plot.

↳ $y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$.

$$\begin{aligned} e &= (I-H)(X\beta + \varepsilon) \\ &= (I-H)X\beta + (I-H)\varepsilon \\ &= (I-H)\varepsilon. \end{aligned}$$

$e_i = \varepsilon_i - \sum_{j=1}^{n} h_{ij}\varepsilon_j$ $\Rightarrow$ $h_{ij} \simeq 0$ 이면 $e_i \simeq \varepsilon_i$.

$\Rightarrow$ e_i 로 normality 이야기해도 ok.

but h_{ij} 가 0으로부터 떨어져 있다면 ? — ε_i normal 아닌데도 e_i normal 하게 보일수도.

(주로 n이 작은 경우 발생).

Consider normal prob. plot.

* suppose we have a sample size of n $z_1, \dots, z_n$,
& wish to test that z_i's are sample from $N(\mu, \sigma^2)$

1. Order z_i $\Rightarrow$ $z_{(1)} \le \cdots \le z_{(n)}$

2. Consider normal sample of size n $u_1, \dots, u_n$ with 0 mean & unit variance.

Let $u_{(1)} \le \cdots \le u_{(n)}$ be the mean values of the order stat. of several samples

$\Rightarrow$ $u_{(i)}$'s are the expected values of normal order stat.

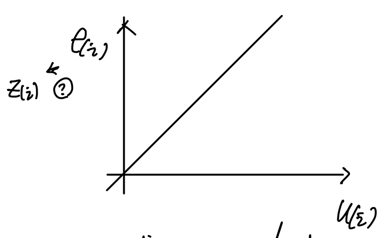
Note $u_{(i)} \sim \Phi^{-1}\left(\frac{i-\frac{3}{8}}{n-\frac{1}{4}}\right)$ (Φ : normal cdf, Φ^{-1} : quantile 값)

$P(X \le x) = \sim$.

3. If z_i's are normal,

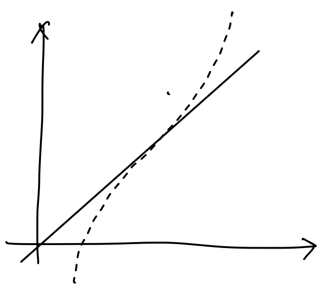
then $E(z_{(i)}) = \mu + \sigma \cdot u_{(i)}$.

$\Leftrightarrow$ $E\left(\frac{z_{(i)} - \mu}{\sigma}\right) = u_{(i)}$.

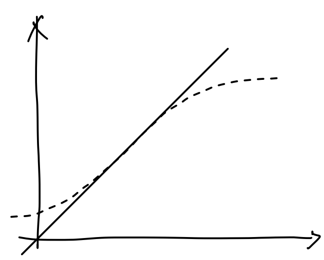


y축: normal의 quantile.

normal이 아닌 경우 :



heavy-tailed



short-tailed.

④ Independence : (1) plot residuals vs time
(2) Durbin - Watson test.

↳ $H_0 : E(\varepsilon_i \cdot \varepsilon_j) = 0.$ (uncorrelated).

$H_1 : E(\varepsilon_i \cdot \varepsilon_j) \neq 0$

$$d = \frac{\Sigma (e_t - e_{t-1})^2}{\Sigma e_t^2} \Rightarrow$$

$d < d_L$: reject H_0 = autocorr ↑.
$d > d_U$: not reject H_0 / $d_L \le d \le d_U$ → test is inconclusive.

Define $\hat{\rho}_k = \dfrac{\sum_{t=k+1}^{n} e_t e_{t-k}}{\sum_{t=1}^{k} e_t^2}$, $k = 1, 2, \dots$ ⇒ lag k 하에서 defined.

(= lag k sample autocorr. of residuals)

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = \frac{\Sigma e_t^2 + \Sigma e_{t-1}^2 - 2\Sigma e_t \cdot e_{t-1}}{\Sigma e_t^2}$$

$$\simeq 2 - 2 \cdot \frac{\Sigma e_t \cdot e_{t-1}}{\Sigma e_t^2}$$

$$= 2 - 2\hat{\rho}_k = 2(1 - \hat{\rho}_1)$$

↳ d는 $\hat{\rho}_1 \simeq 1$ 일 때 작아진다.

① Linearity : Partial Regression Plot.

↳ Consider the model $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon.$

Suppose we are interested in the relationship Y & X_1. (X_1 X_2 가 (indep.) orthogonal 이면 Y_i X_i 의 plot 만 그려도 ok)

1. Fit the reg. of Y & X_2 : $y = \theta_0 + \theta_1 X_2 + \varepsilon.$

⇒ $\hat{y}_i(X_2) = \hat{\theta}_0 + \hat{\theta}_2 X_{i2}$ $i = 1, 2, \dots n.$

& $e_i(Y|X_2) = y_i - \hat{y}_i(X_2).$

↳ X_1의 정보만 有.

2. Fit the reg. of X_1 on X_2 : $X_1 = \alpha_0 + \alpha_1 X_2 + \varepsilon.$

⇒ $\hat{X}_i(X_2) = \hat{\alpha}_0 + \hat{\alpha}_1 X_{i2}$ $i = 1, 2, \dots, n.$

& $e_i(X_1|X_2) = X_{i1} - \hat{X}_{i1}(X_2).$

3. Plot $e_i(y|X_2)$ & $e_i(X_1|X_2)$

⇒ if linear, X_1이 y를 marginal 하게 설명하는 부분 알 수 있다.

In general, $Y = X\beta + e$

$= X_{(j)}\beta_{(j)} + \beta_j \underline{X_j} + \varepsilon.$

$\underline{(I - \underset{\sim}{H_{(j)}})\underline{y}} = (I - H_{(j)}) X\beta + (I - H_{(j)})\varepsilon.$

$= \cancel{(I - H_{(j)}) X_{(j)}}\beta_{(j)} \stackrel{\to 0}{} + \beta_j (I - H_{(j)}) X_j + (I - H_{(j)})\varepsilon.$

$\Downarrow$

I의 e_i $= \beta_j \underbrace{(I - H_{(j)}) X_j}_{\Downarrow} + \underline{(I - H_{(j)})\varepsilon}$

2의 e_i 　　　 ε^*로 두면 linear reg. 한 것.

① Linearity : Lack of fit test

↳ $y = E(y|x) + \varepsilon.$

$H_0 : E(y|x) = \beta_0 + \beta_1 x$

$H_1 : E(y|x) \neq \beta_0 + \beta_1 x$ (not linear).

- Assumption : normality, independence, constant variance
- requires repeated obs at x levels.

x_i 　 $i = 1, \dots, m$ 　 ⇒ level

y_{ij} 　 $j = 1, 2, \dots, n_i$ 　 ⇒ x_i 에서의 obs.

$(n = \sum n_i)$

Testing: $y_{ij} - \hat{y}_i = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \hat{y}_i)$

※ cross-product term : $\sum\sum (\underbrace{y_{ij} - \bar{y}_i}_{0})(\bar{y}_i - \hat{y}_i)$

$\sum_{i=1}^{m}\sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{m} n_i(\underbrace{\bar{y}_i - \hat{y}_i}_{\text{가까우면}})^2$

$SS_E = SS_{PE} + SS_{LoF}.$ 　 $SS_{LoF} \downarrow$ (귀무가설 support).

SS_{PE} ↓ pure error (model indep.)

SS_{LoF} ↓ lack of fit.

df : $n - 2 = \underbrace{(n-m)}_{= \sum_{i=1}^{m}(n_i - 1)} + (m - 2)$

$SS_E(F) = \sum (y_i - \bar{y}_i)^2$

↳ $SS_{PE} = \sum\sum (y_{ij} - \bar{y}_i)^2$

Test stat: $F_0 = \dfrac{MS_{LoF}}{MS_{PE}} \underset{H_0}{\sim} F_{m-2, n-m}$

(Extra sum of squares).

$SS_E(R) = \sum (y_i - \hat{y}_i)^2$

↳ $SS_{LoF} = \sum\sum (y_{ij} - \hat{y}_i)^2$

Ch 5. Transformation & Weighting.

Transformations to Linearize

- $y = \beta_0 x^{\beta_1} \varepsilon \Rightarrow \log y = \log \beta_0 + \beta_1 \log x + \log \varepsilon$
  $y' = \beta_0' + \beta_1 x' + \varepsilon' \overset{iid}{\sim} N(0, \sigma^2)$
- $y = \beta_0 e^{\beta_1 x} \Rightarrow \log y = \log \beta_0 + \beta_1 x$
  $y' = \beta_0' + \beta_1 x$
- $y = \beta_0 + \beta_1 (\log x) \Rightarrow y = \beta_0 + \beta_1 x'$
- $y = \dfrac{x}{\beta_0 x + \beta_1} \Rightarrow \dfrac{1}{y} = \dfrac{\beta_0 x + \beta_1}{x}$
  $y' = \beta_0 + \beta_1 x'$
- $y = \beta_0 + \beta_1 \cdot \dfrac{1}{x} \Rightarrow y = \beta_0 + \beta_1 x'$

Ex) 어떤 term을 고려하지 않으면 data의 structure가 남아 있다.

real. $\underline{y} = X\underline{\beta} + W\underline{r} + \varepsilon$

fitted : $\underline{y} = X\underline{\beta} + \underline{e} \rightarrow e = (I-H)y.$

$$E(e) = (I-H)E(y)$$
$$= (I-H)(X\beta + Wr)$$
$$= (I-H)Wr \Rightarrow \text{평균이 0이 되지 X.}$$

Variance - Stabilizing Transformation

↳ equal variance 위한..

- $\sigma_i^2 \propto E(y_i) \Rightarrow y' = \sqrt{y}.$
- $\sigma_i^2 \propto \{E(y_i)\}^2 \Rightarrow y' = \log(y)$

$\Rightarrow$ 큰 값은 크게, 작은 값은 작게 커지도록.

GOAL : to find transformation $h(y)$
s.t. the variance of $h(y)$ is constant.

1. Suppose that $Var(y_i) = w(\mu_i)$ $* \mu_i = E(y_i)$
2. By Taylor expansion, $h(y) = h(\mu) + (y-\mu)h'(\mu) + \cdots$

$$\Rightarrow Var(h(y)) \simeq (h'(\mu))^2 \, Var(y) = \underset{\sim}{k}$$

constant. (우리의 목적).

$$h'(\mu) = \frac{k}{\sqrt{Var(y)}} = \frac{k}{\sqrt{w(\mu)}}$$

$$h(\mu) = \int \frac{d\mu}{\sqrt{w(\mu)}}$$

Ex. $Var(y) = Var(\varepsilon) \propto \underset{}{\overset{= w(\mu)}{\mu^2}}$ 인 경우.

$$h(\mu) = \int \frac{d\mu}{\sqrt{w(\mu)}} = \int \frac{1}{\mu} d\mu = \log(\mu)$$

· $Var(y) = Var(\varepsilon) \propto \mu$ 인 경우 (Poisson)

$$h(\mu) = \int \frac{d\mu}{\sqrt{w(\mu)}} = \int \frac{1}{\sqrt{\mu}} d\mu = \sqrt{\mu}$$

<u>Box-Cox Transformation</u> : y^λ 로 transform but $\lambda = 0$ 이면 모든 response 가 const.

$\Rightarrow$ $\frac{y^\lambda - 1}{\lambda}$ 를 대신 이용 ($\frac{y^\lambda - 1}{\lambda} \to \log y$ as $\lambda \to 0$)

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \lambda \neq 0. \\ \log y & \lambda = 0. \end{cases}$$

$y_i^{(\lambda)} = X_i^T \beta + \varepsilon_i$. → const. var & normality 성립.

$$L(\beta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\left(y^{(\lambda)} - X\beta\right)^T \left(y^{(\lambda)} - X\beta\right)\right) |J|$$

$|J|$ ↳ $\prod_{i=1}^{n} y_i^{\lambda - 1}$ ($y_i^{\lambda-1} = \frac{dy^{(\lambda)}}{dy_i}$)

LSE $\hat\beta$, $\hat\sigma^2$ 를 대입.(?)

$$L(\hat\beta, \hat\sigma^2) = (2\pi\hat\sigma^2)^{-n/2} e^{-n/2} |J|$$

↳ ($\because$) $\hat\sigma^2 = \frac{1}{n} SSE(y^{(\lambda)})$ ($\frac{1}{n}$ ↳ (MLE))

$$l(\lambda) = \log L(\hat\beta, \hat\sigma^2)$$

$$= -\frac{n}{2} \log SSE(y^{(\lambda)}) + (\lambda - 1) \sum \log y_i.$$ (constant 제외)

$$\hat\lambda = \underbrace{\arg\max\ l(\lambda)}_{\text{↳ 범위 내 } \lambda \text{에 대해...}} \Rightarrow \frac{y^{\hat\lambda} - 1}{\hat\lambda} \text{ 로 plug in.}$$

<u>Generalized (Weighted) Least Squares</u>

· $Var(\varepsilon) = \sigma^2 V$ 인 경우 $(V \neq I)$ $\begin{cases} V \text{ diagonal but unequal elements.} \\ \text{off-diagonal elements} \neq 0. \end{cases}$

→ OLS 적절하지X. ($\because$ obs' are correlated).

1. Find K (unique nonsingular symmetric) s.t. $K^T K = KK = V$.

2. Define $z = K^{-1} y$, $B = K^{-1} X$, $g = K^{-1} \varepsilon$.

3. Model $\quad z = B\beta + g$

$$\Rightarrow E(g) = E(K^{-1}\varepsilon) = K^{-1}E(\varepsilon) = 0.$$

$$\Rightarrow Var(g) = Var(K^{-1}\varepsilon) = K^{-1}Var(\varepsilon)K^{-1}$$
$$= K^{-1}\sigma^2 V K^{-1}$$
$$= \sigma^2 K^{-1}KKK^{-1} = \sigma^2 I$$

4. Find LSE. (GLSE)

$$\rightarrow \hat{\beta}_G = (B^T B)^{-1}B^T z$$
$$= (X^T K^{-1} K^{-1} X)^{-1} X^T K^{-1} K^{-1} y$$
$$= \underbrace{(X^T V^{-1} X)^{-1} X^T V^{-1}}_{A} y.$$

properties : ① $E(\hat{\beta}_G) = (X^T V^{-1} X)^{-1} X^T V^{-1} \cancel{E(y)}^{X\beta} = \beta.$

② $Var(\hat{\beta}_G) = A\, Var(y) A^T$
$$= \sigma^2 AVA^T$$
$$= \sigma^2 (X^T V^{-1} X)^{-1} X^T \cancel{V^{-1} V} V X (X^T V^{-1} X)^{-1}$$
$$= \sigma^2 (X^T V^{-1} X)^{-1}$$

* Weighted LSE 은 언제?

: V is diagonal, but unequal elements 인 경우 : $V = \begin{pmatrix} \frac{1}{w_1} & & 0 \\ & \frac{1}{w_2} & \\ 0 & & \ddots & \frac{1}{w_n} \end{pmatrix}$

Let $W = V^{-1}$ $\quad (V = KK)$

$$\hat{\beta}_W = (X^T W X)^{-1} X^T W y.$$

$$* \; B = K^{-1}X = \begin{pmatrix} \sqrt{w_1} & \sqrt{w_1}\, x_{11} & \cdots & \sqrt{w_1}\, w_{1k} \\ \vdots & \vdots & & \\ \sqrt{w_n} & \sqrt{w_n}\, x_{n1} & & \sqrt{w_n}\, w_{nk} \end{pmatrix}$$

$$z = K^{-1}y = \begin{pmatrix} \sqrt{w_1}\, y_1 \\ \vdots \\ \sqrt{w_n}\, y_n \end{pmatrix}$$

ch 6. Diagnostics for Leverage & Influence.

Leverage Point : unusual x-value.

회귀계수에는 영향 X, R^2나 $se(\hat{\beta})$에는 영향.

Influence Point : unusual x-value & y-value.

회귀계수에 영향 O.

Hat Matrix

$\boxed{\frac{1}{n} \le h_{ii} \le 1.}$

- In simple reg, $h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}$ → x_i가 $\bar{x}$로부터 떨어진 정도.

- 관찰 $\hat{E}(y|x) = \hat{y} = Hy = \hat{\mu}$.

$\hat{y}_i = h_{ii} y_i + \sum_{j \neq i}^{n} h_{ij} y_j$.

$h_{ii}=1$이라면 $H\mathbf{1} = \mathbf{1}$, 즉 $\sum_{j=1}^{n} h_{ij} = 1$이므로 나머지는 O.

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

$h_{ii} \simeq 1$이면 $Var(e_i) \simeq 0$. 즉 $y_i \simeq \hat{y}_i$. → "leverage"↑.

↳ x에서 멂 (좋은 obs 아님).

- Decision Rule : $h_{ii} > 2 \times \frac{p}{n}$, $= 2 \times \overline{h_{ii}}$

$* \sum_{i=1}^{n} h_{ii} = tr(H) = tr(I_p) = p$.

Cook's D : Measure of Influence.

$$D_i := \frac{r_i^2}{p} \times \frac{h_{ii}}{1-h_{ii}}.$$

r_i : studentized residual $\left(= \frac{e_i}{\sqrt{MSE(1-h_{ii})}} \right)$

$\frac{r_i^2}{p}$: $y \leftrightarrow \hat{y}$ $\qquad$ $\frac{h_{ii}}{1-h_{ii}}$: $x \leftrightarrow \bar{x}$.

유도

$$:= \frac{(\hat{y} - \hat{y}_{(i)})^T (\hat{y} - \hat{y}_{(i)})}{p \underbrace{\hat{\sigma}^2}_{MSE}} = \frac{(\hat{\beta} - \hat{\beta}_{(i)})^T X^T X (\hat{\beta} - \hat{\beta}_{(i)})}{p \hat{\sigma}^2}$$

① $$\hat\beta_{(i)} = (X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T y_{(i)}$$
$$= (X^TX - x_i x_i^T)^{-1}(X^Ty - x_i y_i)$$
$$= \left((X^TX)^{-1} + \frac{(X^TX)^{-1} x_i x_i^T (X^TX)^{-1}}{1 - x_i^T (X^TX)^{-1} x_i} \right)(X^Ty - x_i y_i)$$
$$= \hat\beta - (X^TX)^{-1}x_i y_i + \frac{(X^TX)^{-1}x_i x_i^T(\hat y_i - h_{ii} y_i)}{1-h_{ii}}$$
$$= \hat\beta - \frac{(X^TX)^{-1}x_i\left(y_i(1-h_{ii}) + \hat y_i - h_{ii} y_i\right)}{1-h_{ii}}$$
$$= \hat\beta - \frac{(X^TX)^{-1}x_i e_i}{1-h_{ii}}.$$

② $$\hat y - \hat y_{(i)} = X(\hat\beta - \hat\beta_{(i)})$$
$$= \frac{X(X^TX)^{-1}x_i e_i}{1-h_{ii}} = \frac{H u_i e_i}{1-h_{ii}}$$

③ $$\frac{(\hat y - \hat y_{(i)})^T(\hat y - \hat y_{(i)})}{p\, MS_E} = \frac{e_i^2\, u_i^T H u_i}{1-h_{ii}} \cdot \frac{1}{p\, MS_E} = D_i.$$

$$\text{DFFITS} := \frac{\hat y_i - \hat y_{(i)}}{\sqrt{\hat\sigma_{(i)}^2 h_{ii}}} = \frac{e_i}{\sqrt{\hat\sigma_{(i)}^2(1-h_{ii})}} \cdot \left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2}$$

$$\hat\sigma_{(i)}^2 \to \frac{\sum_{j\neq i}^{n}(y_j - \hat y_j)^2}{n-k-1}.$$

$$\hat y_i - \hat y_{(i)} = x_i^T(\hat\beta - \hat\beta_{(i)})$$
$$= \frac{h_{ii} e_i}{1-h_{ii}}$$

Ch 7. Polynomial Reg.

Removing ill-conditioning $\Rightarrow$ center the data.

$$\Rightarrow \ y = \beta_0 + \beta_1 (x - \bar{x}) + \beta_2 (x - \bar{x})^2 + \varepsilon.$$

Piecewise

$$(x-t)_+ := \begin{cases} x-t & \text{if } x-t>0. \\ 0 & \text{o.w.} \end{cases}$$

e.g. $y = \beta_0 + \beta_1 x + \beta_2 (x-t)_+ + \varepsilon$

if $x > t$: $E(y) = (\beta_0 - t\beta_2) + (\beta_1 + \beta_2)x.$

$x \le t$: $E(y) = \beta_0 + \beta_1 x + \varepsilon.$

(General Form): $\quad \tau = \{t_1, \dots, t_k\} \rightarrow$ k개의 knots.

$$y = \beta_0 + \beta_1 x + \sum_{j=1}^{k} \alpha_j (x - t_j)_+ + \varepsilon.$$

Orthogonal Poly-.

"not orthogonal" $\Rightarrow$ $\hat{\beta}_0 \dots \hat{\beta}_k$ changes if $\beta_{k+1} x^{k+1}$ is added.

$\hookrightarrow$

$$y_i = \alpha_0 P_0(x_i) + \alpha_1 P_1(x_i) + \dots + \alpha_k P_k(x_i) + \varepsilon_i = X\alpha + \varepsilon.$$

$\left(P_j(x_i) \text{ is orthogonal poly.} \Leftrightarrow \sum_{i=1}^{n} P_r(x_i) P_s(x_i) = 0, \ r \ne s. \right)$

$\& \ P_0(x_i) = 1.$

$$X = \begin{pmatrix} 1 & P_1(x_1) & \cdots & P_k(x_1) \\ \vdots & \vdots & & \vdots \\ 1 & P_1(x_n) & & P_k(x_n) \end{pmatrix}, \qquad (X^TX)^{-1} = \begin{pmatrix} \frac{1}{n} & & & 0 \\ & \frac{1}{\sum_{i=1}^{n} P_1(x_i)^2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\sum_{i=1}^{n} P_k(x_i)^2} \end{pmatrix}$$

- Least-Squares estimator $\quad \hat{\alpha} = (X^TX)^{-1}X^Ty$

$$= \begin{pmatrix} \frac{1}{n} & & & \\ & \frac{1}{\bigcirc} & & \\ & & \ddots & \\ & & & \frac{1}{\bigcirc} \end{pmatrix} \cdot \begin{pmatrix} \Sigma y_i \\ \Sigma P_1(d_i)\, y_i \\ \vdots \\ \Sigma P_k(d_i)\, y_i \end{pmatrix} = \begin{pmatrix} \frac{\Sigma y_i}{n} \\ \frac{\Sigma P_1(d_i) y_i}{\Sigma P_1(d_i)^2} \\ \vdots \\ \frac{\Sigma P_k(d_i) y_i}{\Sigma P_k(d_i)^2} \end{pmatrix}$$

$$\Leftrightarrow \quad \hat{\alpha}_j = \frac{\Sigma P_j(X_i) y_i}{\Sigma P_j^2(X_i)}$$ → X의 j^{th} col 에만 의존! $\quad \left(* \ \hat{\alpha}_0 = \frac{\Sigma y_i}{n} = \bar{y} \right)$.

- $SS_T = SS_R + SS_E$. 이고,

$\Sigma(y_i - \bar{y})^2$

$$SS_E = (y - X\hat{\alpha})^T (y - X\hat{\alpha})$$

$$= y^Ty - \hat{\alpha}^T X^T X \hat{\alpha}$$

$$\hat{\alpha}^T X^T X \hat{\alpha} \to (\hat{\alpha}_0 \cdots \hat{\alpha}_k) \begin{pmatrix} n & & \\ & \ddots & \\ & & \Sigma P_k(d_i)^2 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_0 \\ \vdots \\ \hat{\alpha}_k \end{pmatrix}$$

$$= \sum_{i=1}^n y_i^2 - \sum_{j=0}^k \hat{\alpha}_j^2 \left(\sum_{i=1}^n P_j^2(d_i) \right)$$

$$\to \hat{\alpha}_0^2 \cdot \underbrace{\sum_{i=1}^n P_0^2(d_i)}_{1} + \sum_{j=1}^n \hat{\alpha}_j^2 \left(\sum_{i=1}^n P_j^2(d_i) \right)$$

$$= \bar{y}^2 \cdot n + \Sigma \hat{\alpha}_j^2 \left(\Sigma P_j^2(x_i) \right)$$

$$= \underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{SS_T} - \underbrace{\sum_{j=1}^k \hat{\alpha}_j^2 \left(\sum_{i=1}^n P_j(X_i) \right)}_{SS_R.}$$

- $SS_R = \Sigma \hat{\alpha}_j^2 \left(\Sigma P_j^2(X_i) \right) = \Sigma \hat{\alpha}_j \left(\Sigma P_j(X_i) y_i \right)$ — 두가지 표기 방식.

$$\underset{\text{decompose}}{=} \hat{\alpha}_1^2 \, \Sigma P_1^2(X_i) + \cdots + \hat{\alpha}_k \, \Sigma P_k^2(d_i)$$

$$= \underbrace{SS_R(\alpha_1)}_{y = X_1\alpha_1 + \varepsilon.} + \cdots + SS_R(\alpha_k)$$

Testing $H_0 : \alpha_j = 0$.

- $SS_E(F) = SS_T - SS_R(F)$

$$= SS_T - \left(SS_R(\alpha_1) + \cdots + SS_R(\alpha_k) \right)$$

- $SS_E(R) = SS_T - SS_R(R)$

$$= SS_T - \left(SS_R(\alpha_1) + \cdots + SS_R(\alpha_{j-1}) + SS_R(\alpha_{j+1}) + \cdots + SS_R(\alpha_k) \right)$$

$$F_0 = \frac{SS_H / df}{SS_E / df} = \frac{SS_R(\alpha_j)/1}{SS_E(F)/n-k-1}.$$

Ch 8. Indicator Variables

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

$$x_2 = \begin{cases} 0 & \text{if type } A \\ 1 & \text{if type } B \end{cases} \Rightarrow \begin{array}{l} E(y|A) = \beta_0 + \beta_1 x_1 \\ E(y|B) = \beta_0 + \beta_1 x_1 + \beta_2 = (\beta_0 + \beta_2) + \beta_1 x_1 \end{array}$$

↳ $H_0 : \beta_2 = 0$?

⇔ two sample t-test (group # = 2) / one-way ANOVA.

t-test – $H_0 : \beta_2 = 0$

F-test – $H_0 : \beta_1 = \beta_2 = 0$.

Model Containing Interaction Effects

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

$$\Rightarrow \quad E(y/A) = \beta_0 + \beta_1 x_1$$

$$E(y|B) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) x_1$$

t-test – Common slope? – $H_0 : \beta_3 = 0$. (rpm이 주는 영향 동일한가?).

↳ ESS 혹은 coef t-test.

F-test (ESS). – $H_0 : \beta_2 = \beta_3 = 0$

↳ $SS_E(R) - SS_E(F)$.

$$= SS_R(\beta_2, \beta_3 | \beta_1, \beta_0)$$

$$= SS_R(\beta_1, \beta_2, \beta_3 | \beta_0) - SS_R(\beta_1 | \beta_0)$$

Indicator vs allocated codes

$\begin{pmatrix} A \to 1 \\ B \to 2 \\ C \to 3 \end{pmatrix} \Rightarrow$

$E(y|A) = \beta_0 + \beta_1 x_1 + \beta_2$

$E(y|B) = \beta_0 + \beta_1 x_1 + 2\beta_2$

$E(y|C) = \beta_0 + \beta_1 x_1 + 3\beta_2$

] 차이가 일치한다는 보장 X.

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

Piecewise Linear Reg.

piecewise spline을 indicator로 표현 가능.

Ex1 $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 \underbrace{(x_{i1} - 500) x_{i2}}_{} + \varepsilon_i$. $\quad (x_{i1}-500)_+$

$x_{i2} = \begin{cases} 0 & \text{if } x_{i1} \le 500 \\ 1 & \text{o.w.} \end{cases}$

$\Rightarrow E[y \mid x_1 \le 500] = \beta_0 + \beta_1 x_1$

$E(y \mid x_1 > 500) = (\beta_0 - 500\beta_2) + (\beta_1 + \beta_2) x_1$.

] → $x_{i1} = 500$ 일 때 연속!

Ex2 $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 (x_{i1} - 40) x_{i2} + \beta_3 x_{i3} + \varepsilon$.

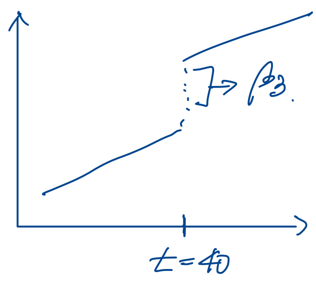
$\Rightarrow E(y \mid x \le 40) = \beta_0 + \beta_1 x_1$

$E(y \mid x > 40) = (\beta_0 - 40\beta_2 + \beta_3) + (\beta_1 + \beta_2) x_1$

↳ $H_0: \beta_2 = 0$ → slope 차이 X.

$H_0: \beta_3 = 0$ → discontinuous at $x_1 = 40$

if reject $H_0: \beta_2 = \beta_3 = 0$? → 전 구간에서 하나의 straight line으로 표현 가능.



이렇게도 표현 가능: $y = \beta_0 + \beta_1 x_1 + \beta_2 (x_1 - t)_+ + \beta_3 (x_1 - t)_+^0$.

↳ $\begin{cases} 1 & \text{if } x > t \\ 0 & \text{if } x \le t. \end{cases}$

Regression Approach to One-way ANOVA.

$y_{ij} = \mu_i + \varepsilon_{ij} \quad (i = 1, \dots, a,\ j = 1, \dots, n)$.

$\overset{iid}{\sim} N(0, \sigma^2)$.

↳ i번째 trt의 j번째 반복으로 이루어진 관측값.

목적: $\mu_1 = \cdots = \mu_a$ 인가?

$\Leftrightarrow \tau_i = 0$ for all i.

($\mu_i = \mu + \tau_i$ 로 표현하기도.
μ: y_{ij}의 전체 평균
τ_i: i^{th} treat effect)

Want to find μ & τ_i's $(i=1,\dots,a)$ $\Rightarrow$ a+1 개의 parameter.

$\hookrightarrow$ LSE : $S(\mu,\tau_i) = \sum_{i=1}^{a}\sum_{j=1}^{n} (y_{ij}-\mu-\tau_i)^2$

$\left(* \; y_{..} = \sum\sum y_{ij}, \quad y_{i.} = \sum_{j=1}^{n} y_{ij} \right)$

$\frac{\partial S}{\partial \mu}\Big|_{\hat\mu,\hat\tau_i} = -2\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\hat\mu-\hat\tau_i) = 0 \iff N\hat\mu + n\hat\tau_1 + \dots + n\hat\tau_a = y_{..} \quad \cdots ①$

$\frac{\partial S}{\partial \tau_i}\Big|_{\hat\mu,\hat\tau_i} = -2\sum_{j=1}^{n}(y_{ij}-\hat\mu-\hat\tau_i) = 0 \iff \begin{cases} n\hat\mu + n\hat\tau_1 = y_{1.} \\ \vdots \\ n\hat\mu + n\hat\tau_a = y_{a.} \end{cases}$

(n+1 equations but not indep.)

$\rightarrow$ 하나의 constraint 더 필요:

모형에서 $\mu = \frac{1}{a}\sum_{i=1}^{a}\mu_i = \frac{1}{a}\sum_{i=1}^{a}(\mu+\tau_i)$

$= \mu + \underbrace{\frac{1}{a}\sum_{i=1}^{a}\tau_i}_{=0}$. $\quad (\iff \tau_a = -\tau_1 - \dots - \tau_{a-1})$

①에서 $n\sum\hat\tau_i = 0$,

즉 $\hat\mu = \underline{y_{..}/N = \bar y_{..}}$, 아래 식들에 대입하면 $\hat\tau_i = \bar y_{i.} - \bar y_{..}$

$\rightarrow \begin{cases} \hat y_{ij} = \hat\mu + \hat\tau_i = \bar y_{i.} \\ e_{ij} = y_{ij} - \bar y_{i.} \end{cases}$

$SS_T = SS_E + SS_R$.
$\hookrightarrow SS_{TRT}$.

$\sum\sum(y_{ij}-\bar y_{..})^2 = \sum\sum(y_{ij}-\bar y_{i.})^2 + \sum\sum(\bar y_{i.}-\bar y_{..})^2$
$+ 2\sum_i \underbrace{\boxed{\sum_j (y_{ij}-\bar y_{i.})}}_{=0}(\bar y_{i.}-\bar y_{..})$.

$= \sum\sum(y_{ij}-\bar y_{i.})^2 + n\sum_{i=1}^{a}(\bar y_{i.}-\bar y_{..})^2$.

df: $a\cdot n-1 = \underbrace{a\cdot(n-1)}_{an-a.} + (a-1)$

$$F_0 = \frac{SS_{TRT}/(a-1)}{SS_E/a(n-1)} \sim F_{a-1,\, a(n-1)}$$

⇒ 모든 ANOVA를 reg. 로 표현 가능!

Ex) $a=3,\ n=2.$ $y_{ij} = \mu + \tau_i + \varepsilon_{ij} \rightarrow y = X\beta + \varepsilon.$

$$\begin{cases} y_{11} = \mu + \tau_1 + \varepsilon_{11} \\ y_{12} = \mu + \tau_1 + \varepsilon_{12} \\ y_{21} = \mu + \tau_2 + \varepsilon_{21} \\ y_{22} = \mu + \tau_2 + \varepsilon_{22} \\ y_{31} = \mu - \tau_1 - \tau_2 + \varepsilon_{31} \\ y_{32} = \mu - \tau_1 - \tau_2 + \varepsilon_{32}. \end{cases} \Rightarrow \begin{pmatrix} y_{11} \\ \vdots \\ \vdots \\ y_{32} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ \vdots & 1 & 0 \\ \vdots & 0 & 1 \\ \vdots & 0 & 1 \\ \vdots & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} + \varepsilon.$$

$$\hat{\beta} = (X^TX)^{-1}X^Ty = \begin{pmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \end{pmatrix}.$$ 추정. $(\hat{\tau}_3 = -\hat{\tau}_1 - \hat{\tau}_2)$

⇔ $y = \mu + \tau_1 x_1 + \tau_2 x_2 + \varepsilon.$

where $x_1 = \begin{cases} 1 & \text{if } 1 \\ -1 & 3 \\ 0 & \text{o.w.} \end{cases}$ $x_2 = \begin{cases} 1 & \text{if } 2 \\ -1 & 3 \\ 0 & \text{o.w.} \end{cases}$

Ex) different constraint : $\tau_1 = 0$. (잘 쓰이지는 X)

$y = \mu + \tau_2 x_1 + \tau_3 x_2 + \varepsilon,$

where $x_1 = \begin{cases} 1 & \text{if } 2 \\ 0 & \text{o.w.} \end{cases}$ $x_2 = \begin{cases} 1 & \text{if } 3 \\ 0 & \text{o.w.} \end{cases}$

+) Analysis of Covariance

Model) $y_{ij} = \mu + \underset{\text{없으면 reg.}}{\underline{\tau_i}} + \underset{\text{없으면 ANOVA.}}{\underline{\beta(x_{ij} - \bar{x}_{..})}} + \varepsilon_{ij}$ ⇒ $\sum \tau_i = 0$, $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

$E(y_{ij}) = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..})$

$Var(y_{ij}) = \sigma^2.$

→ reg: $y_j = \mu + \tau_1 z_{1j} + \tau_2 z_{2j} + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_j$

$z_1 = \begin{cases} 1 & \text{if } 1 \\ -1 & 3 \\ 0 & \text{o.w} \end{cases}$ $z_2 = \begin{cases} 1 & \text{if } 2 \\ -1 & 3 \\ 0 & \text{o.w} \end{cases}$

$\rightarrow \; y = X\beta + \varepsilon$

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{32} \end{pmatrix} = \begin{bmatrix} 1 & 1 & 0 & x_{11} - \bar{x}_{..} \\ \vdots & 1 & 0 & \vdots \\ \vdots & 0 & 1 & \vdots \\ \vdots & 0 & 1 & \vdots \\ 1 & -1 & -1 & x_{32} - \bar{x}_{..} \\ & -1 & -1 & \end{bmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \beta \end{pmatrix} + \varepsilon$$

$\hat{\beta} = (X^T X)^{-1} X^T y$.

Compute SS_E, SS_R.

$\hookrightarrow$ df $= an - a - 1$, a

① $H_0 : \tau_i = 0$. (⇔ group 간 차이 有?)

Reduced : $y_{ij} = \mu + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij}$ $\Rightarrow$ $F_0 = \dfrac{SS_E(R) - SS_E(F) \,/\, (a-1)}{SS_E(F) \,/\, an - (\underbrace{a \cancel{-1}}_{\tau\text{개수}} \underbrace{\cancel{+1}}_{\beta}) - 1}$

② $H_0 : \beta = 0$ (rpm이 y에 영향 주는가?)

Reduced : $y_{ij} = \mu + \tau + \varepsilon_{ij}$ $\rightarrow$ $F_0 = \dfrac{SS_E(R) - SS_E(F) \,/\, 1}{SS_E(F) \,/\, (an - a - 1)}$