시험볼때 계산기 지참하기
4/24(월) 18시 시작
1시간 내외 → 75분을 초과하지 않을 예정

2장 Review

■ 기본개념

X : 설명변수
Y : 반응변수

$$E(Y|X=x) = \beta_0 + \beta_1 x$$

$$Var(Y|X=x) = \sigma^2$$

– 단순선형회귀모형

오차 : 기대값까지의 수직거리

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \cdots, n$$

$$E(\epsilon_i) = 0$$
$$Var(\epsilon_i) = \sigma^2$$
$$Cov(\epsilon_i, \epsilon_j) = 0, \quad i \neq j$$

– 대체모형 절편이 달라짐

$$y_i = \alpha + \beta_1 \underbrace{(x_i - \bar{x})}_{\text{합은 } 0} + \epsilon_i$$

1

2장 Review

★ ■ $\boxed{\text{최소제곱법}}$

특정 선이 존재함, 그 선까지의 수직거리

잔차의 제곱합

– $Q(\beta_0, \beta_1) = \Sigma(y_i - \beta_0 - \beta_1 x_i)^2$를 최소화 하는 β_0, β_1

– 정규방정식 ($\partial Q / \partial \beta_0 = 0$, $\partial Q / \partial \beta_1 = 0$) → 미분해서 0으로 두고 두식을 연립

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad \hat{\beta}_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

→ β_0, β_1의 추정

– 추정회귀식, 적합값, 잔차

$$\boxed{\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x} \Rightarrow \begin{cases} \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \\ e_i = y_i - \hat{y}_i \end{cases}$$

→ 각각의 x_i를 대입한것
→ 자료값과 적합값의 차이 : 잔차

– σ^2의 추정량

$$\hat{\sigma}^2 = \frac{\Sigma(y_i - \hat{y}_i)^2}{n-2} = \frac{SSE}{n-2} = MSE$$

비편향추정량이 되게 하기 위한 자유도
(n에서 모수의 수를 뺀것)

2

산점도를 그려보는것이 좋음!
한눈에 알수있고, 잔차와 적합값에
문제가 있는지도 확인가능

2장 Review

■ 최소제곱추정량의 특성

1) 최소제곱추정량 $\hat{\beta}_0$와 $\hat{\beta}_1$은 비편향추정량(unbiased estimator)

$$E(\hat{\beta}_0) = \beta_0 \qquad E(\hat{\beta}_1) = \beta_1$$

2) 최소제곱추정량 $\hat{\beta}_0$와 $\hat{\beta}_1$은 y_i의 선형결합

$$\hat{\beta}_1 = \Sigma k_i y_i \qquad k_i = (x_i - \bar{x})/S_{xx}$$

$$\hat{\beta}_0 = \Sigma m_i y_i \qquad m_i = 1/n - \bar{x} k_i$$

✱ 알아두는게 좋음

3) 최소제곱추정량 $\hat{\beta}_0$와 $\hat{\beta}_1$은 ϵ_i의 선형결합

$$\hat{\beta}_1 = \beta_1 + \Sigma k_i \epsilon_i \qquad \hat{\beta}_0 = \beta_0 + \Sigma m_i \epsilon_i$$

4) $\hat{\beta}_1$와 $\hat{\beta}_0$의 분산

$$Var(\hat{\beta}_1) = \sigma^2 / S_{xx} \qquad Var(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) \qquad Cov(\hat{\beta}_0, \hat{\beta}_1) = -\bar{x}\frac{\sigma^2}{S_{xx}}$$

$\bar{x}=0$ 인경우를제외하면 $\hat{\beta}_0$, $\hat{\beta}_1$은 서로 영향을 주고받음
대체모형의경우 $\bar{x}=0$이므로 $\hat{\beta}_0$, $\hat{\beta}_1$은 서로독립!
(일반적으로 $\bar{x} \neq 0$)

5) 최소제곱법에서의 $\Sigma e_i = 0$, Σe_i^2 은 다른 추정량의 잔차제곱합보다 작음

↳ 증명 시험에안나옴!

3

2장 Review

6) $\boxed{\text{가우스-마코브(Gauss-Markoff)정리}}$

✱ 오차의 독립성, 등분산성의 가정하에서 ✱
최량선형비편향추정량(best linear unbiased estimator: BLUE)

가장좋다!
(가장 분산이작음)

$$\tilde{\beta}_1 = \Sigma c_i y_i \quad E(\tilde{\beta}_1) = \beta_1 \qquad Var(\hat{\beta}_1) \le Var(\tilde{\beta}_1)$$

$$\tilde{\beta}_0 = \Sigma d_i y_i \quad E(\tilde{\beta}_0) = \beta_0 \qquad Var(\hat{\beta}_0) \le Var(\tilde{\beta}_0)$$

7) 적합값 $\hat{y}_i$의 합과 관측값 y_i의 합은 같음: $\Sigma y_i = \Sigma \hat{y}_i$

8) $x_i / \hat{y}_i$을 가중값으로 하는 잔차의 가중합은 0: $\Sigma x_i e_i = 0$, $\Sigma \hat{y}_i e_i = 0$

9) 추정회귀선 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$은 반드시 $(\bar{x}, \bar{y})$를 통과

10) 잔차제곱합은 $SSE = SS_{yy} - \hat{\beta}_1 S_{xy}$

4

2장 Review

오차항이 서로독립적이고 분산이 같다는 가정하에

■ 가중최소제곱법(method of weighted least squares)

- 등분산가정이 만족되지 못하는 경우
  : 최소제곱법에 의한 추정량은 최량선형비편향추정량이 아님

- 가중오차제곱 Q를 최소화하는 회귀계수를 추정하는 방법

$$Q(\beta_0,\beta_1) = \sum \omega_i (y_i - \beta_0 - \beta_1 x_i)^2$$

알아두면 좋음 ☆

- 가중값 ω_i는 반응변수의 분산의 역수에 비례하도록 설정 : 컴퓨터가 정하지 못하므로 기준을 제시한것!
  - ▫ n_i개 자료의 평균, $Var(y_i) = \sigma^2/n_i \propto 1/n_i$: $\omega_i = n_i$
  - ▫ n_i개 자료의 합, $Var(y_i) = n_i\sigma^2 \propto n_i$: $\omega_i = 1/n_i$
  - ▫ 분산이 x_i에 비례, $Var(y_i) \propto x_i$: $\omega_i = 1/x_i$

5

2장 Review

■ 최대가능도방법(method of maximum likelihood))

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ \ i = 1,\ldots n$$

$$E(\epsilon_i) = 0,\ \ Var(\epsilon_i)\sigma^2,\ \ Cov(\epsilon_i, \epsilon_j) = 0, i \neq j$$

- 분포에 대한 가정 $\epsilon_i \sim N(0,\sigma^2)$ → 오차항이 정규분포를 따른다!

- 가능도함수 (수리통계학에서 자세히 배움)

$$L(\beta_0,\beta_1,\sigma^2|y_1,\cdots,y_n) = f(y_1,\cdots,y_n|\beta_0,\beta_1,\sigma^2) = \prod_{i=1}^{n} f(y_i)$$

- 최대가능도 추정량: 가능도함수를 최대화하는 β_0, β_1, σ^2

$$\hat{\beta}_{0_{MLE}} = \hat{\beta}_0,\qquad \hat{\beta}_{1_{MLE}} = \hat{\beta}_1,\qquad \hat{\sigma}^2_{MLE} = \frac{SSE}{n}$$

↳ 최소제곱법을 사용할때와 같다

→ 얘만 다름! n-2가 아닌 n으로 나눠준다

6

3장 Review

- 분포에 대한 가정

$$\begin{cases} y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n \\ \epsilon_i \sim N(0, \sigma^2), \quad Cov(\epsilon_i, \epsilon_j), \ i \neq j \\ y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2), \quad Cov(y_i, y_j), \ i \neq j \end{cases}$$

어느정도 신뢰할수 있는것인지!

- 추정회귀식의 타당성: ① 모형의 비교 / ② β_1에 대한 추론 — 따지고보면 같은방법이다

- 모형의 비교

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad SSE = \Sigma(y_i - \hat{y}_i)^2 = S_{yy} - \hat{\beta}_1 S_{xy}$$

$$y_i = \beta_0 + \epsilon_i \qquad SSE = \Sigma(y_i - \bar{y})^2 = S_{yy}$$

↑ $\beta_1 = 0$ → y는 x의 영향을 받지 않음

반응변수가 같으면 SSE를 비교해도 됨!
(ex. 하나의 단위가 cm, 다른 하나가 m인 경우 안됨)

7

3장 Review

- $SSR = \underbrace{\hat{\beta}_1 S_{xy}}_{\text{항상양수}} = \Sigma(\hat{y}_i - \bar{y})^2$: 크기에 따라 회귀식의 타당성 여부 결정

x가 존재해서 생기는 차이

$$(y_i - \bar{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

$$\Sigma(y_i - \bar{y})^2 = \Sigma(y_i - \hat{y}_i)^2 + \Sigma(\hat{y}_i - \bar{y})^2$$

→ x로 설명되어지는 부분

x와 관계없는 y자체의 변동

$$\boxed{S_{yy} = SSE + SSR}$$

- $S_{yy}/\sigma^2 \sim \chi^2(n-1)$

서로독립

$$\begin{cases} SSR/\sigma^2 \sim \chi^2(1) & : \beta_1 = 0\text{인 경우} \\ SSE/\sigma^2 \sim \chi^2(n-2) & : SSR/\sigma^2\text{과 } SSE/\sigma^2\text{는 서로 독립} \end{cases}$$

서로 독립인 2개의 χ^2 분포
→ F분포

- $$F^* = \frac{\frac{SSR}{\sigma^2}/1}{\frac{SSE}{\sigma^2}/(n-2)} = \frac{SSR}{SSE/(n-2)} = \frac{MSR}{MSE} \sim F(1, n-2) \ (\beta_1 = 0\text{인 경우})$$

8

3장 Review

- 분산분석과 F-검정 : F^*(표준화한 SSR)의 크기로 타당성 결정
  - 가설검정의 관점

$$\begin{cases} H_0:\ y_i = \beta_0 + \epsilon_i & \rightarrow H_0:\ \beta_1 = 0 \\ H_1:\ y_i = \beta_0 + \beta_1 x_i + \epsilon_i & \rightarrow H_1:\ \beta_1 \neq 0 \end{cases}$$

  - $F^* \geq F(\alpha, 1, n-2)$이면 H_0 기각
  - p-값 $= P(F(1, n-2) > F^*)$

- 결정계수 R^2 (선형 관계)

$$R^2 = \frac{SSR}{S_{yy}} = 1 - \frac{SSE}{S_{yy}}, \qquad R^2 = r_{xy}^2$$

✓ $0 \leq R^2 \leq 1$

↳ 'y자체에 대한 변동'과 '회귀로 설명되어지는 부분'의 비율

선형

$\because S_{yy} = SSR + SSE$

3장 Review

- $\hat{\beta}_0$와 $\hat{\beta}_1$는 y_i의 선형결합

$$\begin{cases} \hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right) \\ \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right) \end{cases}$$

  - 표본의 크기가 큰 경우 (중심극한정리) → MLE 사용

$$\hat{\beta}_0 \approx N\left(\beta_0, \hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right)$$

$$\hat{\beta}_1 \approx N\left(\beta_1, \frac{\hat{\sigma}^2}{S_{xx}}\right)$$

  - 표본의 크기가 작으면 중심극한정리 적용은 부적절

3장 Review

- $\hat{\beta}_0$, $\hat{\beta}_1$: 정규분포를 따르는 비편향추정량 θ에대한설명으로 교재에있음

$$\frac{\hat{\beta}_0-\beta_0}{s.e.(\hat{\beta}_0)} \sim t(n-2) \qquad \frac{\hat{\beta}_1-\beta_1}{s.e.(\hat{\beta}_1)} \sim t(n-2)$$

*기억해두기!

- 표준오차(standard error: s.e.): 추정량의 분산의 제곱근

$$s.e.(\hat{\beta}_1)=\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \quad \text{← MSE사용} \qquad s.e.(\hat{\beta}_0)=\hat{\sigma}\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}}$$

- 구간추정

$$\left(\hat{\beta}_1-t\left(\frac{\alpha}{2},n-2\right)\frac{\hat{\sigma}}{\sqrt{S_{xx}}},\ \hat{\beta}_1+t\left(\frac{\alpha}{2},n-2\right)\frac{\hat{\sigma}}{\sqrt{S_{xx}}}\right)$$

$$\left(\hat{\beta}_0-t\left(\frac{\alpha}{2},n-2\right)\hat{\sigma}\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}},\ \hat{\beta}_0+t\left(\frac{\alpha}{2},n-2\right)\hat{\sigma}\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}}\right)$$

11

3장 Review

- 가설검정
  - 양측검정

$$\begin{cases} H_0: \beta_0=\beta_0^* \\ H_1: \beta_0 \neq \beta_0^* \end{cases} \qquad \begin{cases} H_0: \beta_1=\beta_1^* \\ H_1: \beta_1 \neq \beta_1^* \end{cases}$$

$$t^*=\frac{\hat{\beta}_0-\beta_0^*}{s.e.(\hat{\beta}_0)} \qquad t^*=\frac{\hat{\beta}_1-\beta_1^*}{s.e.(\hat{\beta}_1)}$$

→ *가 있음!

아거 의미가아직 이해가잘안됨...
다시꼭보기

- H_0이 사실이면 $t^* \sim t(n-2)$
- $|t^*| \geq t(\alpha/2, n-2)$이면 H_0 기각
- p-값$=2P(t(n-2)>|t^*|)$

12

3장 Review

– 단측검정

$$\begin{cases} H_0: \beta_0 = \beta_0^* \\ H_1: \beta_0 > \beta_0^* \end{cases} \quad \begin{cases} H_0: \beta_1 = \beta_1^* \\ H_1: \beta_1 > \beta_1^* \end{cases} \qquad \begin{cases} H_0: \beta_0 = \beta_0^* \\ H_1: \beta_0 < \beta_0^* \end{cases} \quad \begin{cases} H_0: \beta_1 = \beta_1^* \\ H_1: \beta_1 < \beta_1^* \end{cases}$$

- $t^* \geq t(\alpha, n-2)$면 H_0 기각 　　 $t^* \leq -t(\alpha, n-2)$면 H_0 기각
- p-값$= P(t(n-2) > t^*)$ 　　 p-값$= P(t(n-2) < t^*)$

회귀분석에서 가장 많이 사용되는 가설검정

■ 가설 $\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$ 검정통계값의 제곱은 분산분석에서의 F^*값과 같음

$$t^{*2} = \left(\frac{\hat{\beta}_1}{s.e.(\hat{\beta}_1)} \right)^2 = \frac{SSR}{MSE} = F^*$$

t-분포를 제곱하면 F-분포니까 $(\sim (1, n-2))$

3장 Review

- 기대값과 새로운 관측값

1) $X = x_0$일때 반응변수의 기댓값: $E(Y|X = x_0)$

2) $X = x_0$일때 반응변수의 새로운 관측(예측)값: $y_0 = (y|X = x_0)$

3) $X = x_0$일때 반응변수 m개의 평균: $\bar{y}_0 = (\bar{y}|X = x_0)$

1) $E(Y|X = x_0)$

$$\hat{E}(Y|X = x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0 = \hat{y}_0$$

기억하기!!

$$Var\left[\hat{E}(Y|X = x_0)\right] = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)$$

σ^2을 $MSE(= \hat{\sigma}^2)$로 대체하여 추정

$$\frac{\hat{E}(Y|X = x_0) - E(Y|X = x_0)}{s.e[\hat{E}(Y|X = x_0)]} \sim t(n-2)$$

교재에는 안나와있는 내용

3장 Review

추정방법 자체에 큰 차이는 없지만
분산이 달라짐을 주의!

$$\left(\hat{y}_0 - t(\alpha/2, n-2)\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}},\ \hat{y}_0 + t(\alpha/2, n-2)\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}}\right)$$

2) $y_0 = (y|X = x_0)$

$$(\hat{y}|X = x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0 = \hat{y}_0$$

$$\left(\hat{y}_0 - t(\alpha/2, n-2)\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}},\ \hat{y}_0 + t(\alpha/2, n-2)\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}}\right)$$

3) $\bar{y}_0 = (\bar{y}|X = x_0)$

$$(\hat{\bar{y}}|X = x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0 = \hat{y}_0$$

$$\left(\hat{y}_0 - t(\alpha/2, n-2)\hat{\sigma}\sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}},\ \hat{y}_0 + t(\alpha/2, n-2)\hat{\sigma}\sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}}\right)$$

15

3장 Review

- 가설검정

✓
$$\begin{cases} H_0: E(Y|X = x_0) = y^* \\ H_1: E(Y|X = x_0) \neq y^* \quad (> y^*) \quad (< y^*) \end{cases}$$

$$\frac{\hat{E}(Y|X = x_0) - E(Y|X = x_0)}{s.e[\hat{E}(Y|X = x_0)]} \sim t(n-2), \qquad t^* = \frac{\hat{y}_0 - y^*}{s.e[\hat{E}(Y|X = x_0)]}$$

✓
$$\begin{cases} H_0: (y|X = x_0) = y^* \\ H_1: (y|X = x_0) \neq y^* \quad (> y^*) \quad (< y^*) \end{cases}$$

✓
$$\begin{cases} H_0: (\bar{y}|X = x_0) = y^* \\ H_1: (\bar{y}|X = x_0) \neq y^* \quad (> y^*) \quad (< y^*) \end{cases}$$

16

3장 Review

- 분산에 대한 추론

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

$$\frac{(n-2)MSE}{\sigma^2} = \frac{SSE}{\sigma^2} \sim \chi^2(n-2)$$ → 이미 알고 있는 내용!

  - σ^2에 대한 $(1-\alpha)\times 100\%$의 신뢰구간

$$\left(\frac{(n-2)MSE}{\chi^2(\alpha/2, n-2)}, \frac{(n-2)MSE}{\chi^2(1-\alpha/2, n-2)} \right)$$

χ^2 분포는 좌우값을 모두 구해야함!
(대칭이 아님)

17

3장 Review

- 원점을 통과하는 모형 X 시험에 안나옴

$$y_i = \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim N(0.\sigma^2), \quad Cov(\epsilon_i, \epsilon_j) = 0, \ i \neq j$$

  - β_1의 추정: $Q(\beta_1) = \Sigma(y_i - \beta_1 x_i)^2$를 최소화

$$\hat{\beta}_1 = \Sigma y_i x_i / \Sigma x_i^2 = S_{xy}^* / S_{xx}^*$$

$$S_{xy}^* = \Sigma(x_i - 0)y_i = \Sigma x_i y, \quad S_{xx}^* = \Sigma(x_i - 0)^2 = \Sigma x_i^2$$

$E(\hat{\beta}_1) = \beta_1$: 비편향추정량

  - $\hat{\beta}_1 \sim N(\beta_1, \sigma^2 / S_{xx}^*)$: 추론에서 $t(n-1)$ 분포의 사용

18

3장 Review

- $\hat{\beta}_1$, $\hat{E}(y|X=x_0)$, $(\hat{y}|X=x_0)$의 분산

$$Var(\hat{\beta}_1) = \sigma^2/S^*_{xx}$$

$$Var[\hat{E}(Y|X=x_0)] = \sigma^2 x_0^2/S^*_{xx}$$

$$Var(\hat{y}_0|X=x_0) = \sigma^2(1+\sigma^2 x_0^2/S^*_{xx})$$

- σ^2의 추정

$$\hat{\sigma}^2 = MSE = (S^*_{yy} - \hat{\beta}_1 S^*_{xy})/(n-1)$$

- 결정계수 R^2 (개념과 해석에 주의)

$$R_0^2 = \Sigma\hat{y}_i^2/\Sigma y_i^2$$

  - 이론적으로 $\beta_0 = 0$이라도 $H_0 : \beta_0 = 0$의 검정 후 모형 선택

19

3장 Review

- 적합결여검정(lack of fit test)

  H_0: 주어진 모형이 적절함 H_0: $E(Y) = \beta_0 + \beta_1 X$ → 설명변수가 1개인 선형모형

  H_1: 주어진 모형이 적절하지 않음 H_1: $E(Y) \neq \beta_0 + \beta_1 X$ → ① 직선이 아닌 경우 ② 설명변수가 더 필요한 경우

- σ^2의 값이 알려진 경우

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

  - H_0이 사실이면 $\chi^2 \sim \chi(n-2)$
  - $\chi^2 \geq \chi^2(\alpha, n-2)$이면 H_0 기각

20

3장 Review

- σ^2의 값을 모르는 경우 (MSE는 이미 모형을 이용한 것이므로 이때 사용이 무의미하다.)

모형에 의존하지 않는 σ^2의 추정방법이 필요

– y_{ij} : i번째 그룹의 j번째 반응변수의 값

$$i = 1, \cdots, m, \quad \sum_{i=1}^{m} n_i = n$$

$$SD_i^2 = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \,/\, (n_i - 1), \quad i = 1, \cdots, m$$

- 순수제곱합 SS_{PE}

$$SS_{PE} = \sum_{i=1}^{m}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^{m} (n_i - 1) SD_i^2$$

$$d.f = \sum_{i=i}^{m} (n_i - 1) = \sum_{i=i}^{m} n_i - \sum_{i=i}^{m} 1 = n - m$$

(관측값의 수 - 그룹의 수)

21

3장 Review

설명변수로 설명되지 않는 경우

- $SSE = SS_{PE} + SS_{LOF}$ ($S_{yy} = SSE + SSR$과 유사)

모형 자체의 문제

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

	SSE	$=$	SS_{PE}	$+$	SS_{LOF}
df:	$(n - p')$	$=$	$(n - m)$	$+$	$(m - p')$

p'는 대부분 2

- 적합결여에 따른 제곱합

$$SS_{LOF} = SSE - SS_{PE} = \sum_{i=1}^{m}\sum_{j=1}^{n_i} (\bar{y}_i - \hat{y}_i)^2 = \sum_{i=1}^{m} n_i (\bar{y}_i - \hat{y}_i)^2$$

- 적합결여검정 : SS_{LOF}의 크기에 따라 H_0의 기각 여부 결정

$$F^* = \frac{SS_{LOF} / (m - p')}{SS_{PE} / (n - m)} \sim F(m - p', n - m)$$

22