알찬대학 따뜻한 동행

The Best Glocal University

고급회귀분석론

이영미 교수님

학 과	통계학과
학 번	202250926
이 름	김보람
과 제 명	HW 02
제 출 일	2022.10.23.(일)

전북대학교

JEONBUK NATIONAL UNIVERSITY

1. $SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$= (y_i - \hat{y}_i)^T (y_i - \hat{y}_i)$

$= (y - Xb)^T (y - Xb)$

$= y^T y - 2b^T X^T y + b^T X^T X b$

$* \ b = (X^T X)^{-1} X^T y$ 대입

$= y^T y - y^T X (X^T X)^{-1} X^T y$

$= y^T \underbrace{(I - X(X^T X)^{-1} X^T)}_{B} y$

$= y^T B y$

$E(SSE) = E(y^T [I - X(X^T X)^{-1} X^T] y)$

$= E(y^T B y) = \sigma^2 tr(B) + \beta^T X^T B X \beta$

$tr(B) = tr(I - X(X^T X)^{-1} X^T)$

$= tr(I) - tr(X(X^T X)^{-1} X^T)$

$= n - (k+1) = n-k-1$ 이고

$X^T B X = X^T [I - X(X^T X)^{-1} X^T] X = 0$ 이므로

$E(SSE) = (n-k-1)\sigma^2$.

2. 오차제곱합 $S = (y_1 - \beta_0)^2 + (y_2 - 2\beta_0 + \beta_1)^2 + (y_3 - \beta_0 - 2\beta_1)^2$

$$\frac{\partial S}{\partial \beta_0} = -\cancel{2}_{1}(y_1 - \beta_0) - \cancel{4}_{2}(y_2 - 2\beta_0 + \beta_1) - \cancel{2}_{1}(y_3 - \beta_0 - 2\beta_1) = 0$$

$$= y_1 - \beta_0 + 2y_2 - 4\beta_0 + 2\beta_1 + y_3 - \beta_0 - 2\beta_1 = 0$$

$$\therefore \beta_0 = \frac{y_1 + 2y_2 + y_3}{6}$$

$$\frac{\partial S}{\partial \beta_1} = 0 + \cancel{2}_{1}(y_2 - 2\beta_0 + \beta_1) - \cancel{4}_{2}(y_3 - \beta_0 - 2\beta_1) = 0$$

$$\therefore \beta_1 = \frac{y_2 - 2y_3}{5}$$

잔차제곱합은 S에 β_0, β_1 대입.

$$\text{잔차제곱합} = \left(y_1 - \frac{y_1 + 2y_2 + y_3}{6}\right)^2 + \left(y_2 - \frac{y_1 + 2y_2 + y_3}{3} + \frac{y_2 - 2y_3}{5}\right)^2$$

$$+ \left(y_3 - \frac{y_1 + 2y_2 + y_3}{6} - \frac{2}{5} \times (y_2 - 2y_3)\right)^2$$

$$= \left(\frac{5y_1 - 2y_2 - y_3}{6}\right)^2 + \left(\frac{-5y_1 + 8y_2 + 11y_3}{15}\right)^2 + \left(\frac{-5y_1 - 22y_2 + 49y_3}{30}\right)^2$$

3. 단순회귀모형 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$ IId 일때

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ 이며, $\hat{\beta}_1 = \dfrac{S(xy)}{S(xx)} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$

$S_{y \cdot x}^2 = \dfrac{\sum (y_i - \hat{y}_i)^2}{n-2}$, $R^2 = \dfrac{SSR}{SST} = \dfrac{(S(xy))^2}{S(xx)S(yy)}$, t검정 $T = \dfrac{\hat{\beta}_1 - \beta_1^0}{\hat{\sigma} / \sqrt{S(xx)}}$

x_i가 cx_i로 대체된다면,

$$\hat{\beta}_1 = \frac{\sum (cx_i - \overline{cx_i})(y_i - \bar{y}_i)}{\sum (cx_i - \overline{cx_i})^2} = \frac{c\, S(xy)}{c^2 S(xx)} = \frac{1}{c} \frac{S(xy)}{S(xx)}$$ 이다.

$\hat{\beta}_0 = \bar{y} - \frac{1}{c}\frac{S(xy)}{S(xx)} \cdot \overline{cx} = \bar{y} - \hat{\beta}_1 \bar{x}$ 이므로 변화가 없다.

$S_{y \cdot x}^2$ 도 변화없음.

$$R^2 = \frac{SSR}{SST} = \frac{(S(xy))^2}{S(xx)S(xy)} = \frac{[\sum (cx_i - \overline{cx_i})(y_i - \bar{y}_i)]^2}{\sum (cx_i - c\bar{x}_i)^2 \cdot \sum (y_i - \bar{y}_i)^2}$$

$$= \frac{S(xy)}{S(xx)S(yy)}$$ 이므로, R^2도 변화없음.

t검정결과값은 분자 $\frac{1}{c}$배, 분모 $\frac{1}{\sqrt{c^2}}$배이므로, 변화없음....

4. (1) $H_0: y=\beta_0+\beta_1 x+\varepsilon$ vs $H_1: y \neq \beta_0+\beta_1 x+\varepsilon$

$\sum x_i = 0+0+3+\cdots+12=60$

$\sum x_i^2 = 0^2+0^2+3^2+\cdots+12^2=540$

$\sum x_i y_i = 0\times 8.5+\cdots+12\times 6.7=431.1$

$\sum y_i = 8.5+8.4+\cdots=76.1$

$\bar{x}=6.\ \bar{y}=7.511$

$$\hat{\beta}_1 = \frac{S(xy)}{S(xx)} = \frac{\sum(x_i-\bar{x}_i)(y_i-\bar{y}_i)}{\sum(x_i-\bar{x}_i)^2} = \frac{\sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}$$

$$= \frac{431.1-\frac{1}{10}\times 60\times 76.1}{540-\frac{1}{10}\times 60^2} = -0.1417$$

$$\hat{\beta}_0 = \bar{y}-\bar{x}\hat{\beta}_1 = \frac{1}{10}\times 76.1 + 0.1417\times 60\times\frac{1}{10} = 8.4584$$

$$\hat{y} = 8.46-0.1417x$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\hat{y}_i)^2 = (8.5-0.04)^2+(8.4+0.06)^2+\cdots+(6.7+0.0596)^2 \doteqdot 0.1165$$

$$SSPE = \sum\sum(y_{ij}-\bar{y}_i)^2 = (8.5-8.45)^2+(8.4-8.45)^2+\cdots+(6.7-6.75)^2 \doteqdot 0.095$$

$$SSLF = SSE-SSPE = 0.1165-0.095 = 0.0215$$

$$F_0 = \frac{\frac{SSLF}{df_L}}{\frac{SSPE}{df_P}} = \frac{\frac{0.0215}{3}}{\frac{0.095}{5}} \doteqdot 0.377 < 5.41 = F(3,5;0.95)$$ 이므로

H_0을 채택한다. 회귀직선은 타당하다.

즉, A사 B사

$\hat{y}_{1j} = 6.756 + 0.2811x_{1j}$

$\hat{y}_{2j} = 12.85714 + 0.145x_{2j}$

ANOVA (A사)

요인	제곱합	자유도
SSR	221.203	1
SSE	5.56	5
SST	226.76	6

ANOVA (B사)

요인	제곱합	자유도
SSR	58.87	1
SSE	8.5071	5
SST	67.37714	6

$SSE(F) = 5.56 + 8.5071 = 14.0671$

$df_F = 5 + 5 = 10$

축소모형 회귀계수를 구하면,

$\Sigma x_i = 560.$ $\Sigma x_i^2 = 28000$

$\Sigma y_i = 256.6$ $\Sigma x_i y_i = 11457$

$$\hat{\beta}_1 = \frac{11457 - 560 \times 256.6 \times \frac{1}{14}}{28000 - 560^2 \times \frac{1}{14}} \doteqdot 0.213036$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \doteqdot 18.32857 - 0.21306 \times 40 \doteqdot 9.80713$$

$\therefore \hat{y}_{ij} = 9.807 + 0.213\hat{x}_{ij}$

$SST = 4998.76 - \frac{1}{14} \times 256.6^2 \doteqdot 295.6486$

$SSR = \frac{1193^2}{5600} \doteqdot 254.1516$

$SSE = SST - SSR \doteqdot 41.49696$

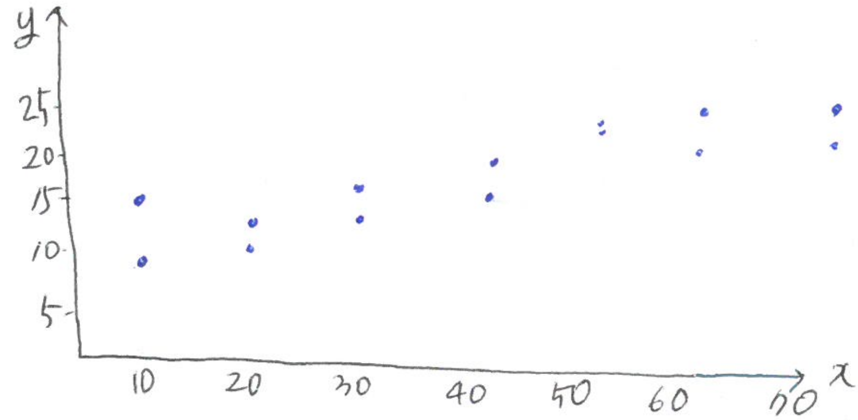
ANOVA

요인	제곱합	자유도
SSR	254.1516	1
SSE	41.49696	12
SST	295.6486	13

$$F_0 = \frac{41.49696 - 14.0671}{12 - 10} \div \frac{14.0671}{10} \doteqdot 9.74965 > F(2, 10; 0.95) = 4.10$$

이므로, H_0를 기각한다. 즉 두 회귀직선은 동일하지 않다.

5. (1)



(2) $H_0: \beta_{01} = \beta_{02}$ and, $\beta_{11} = \beta_{12}$ vs $H_0: \beta_{01} \neq \beta_{02}$ or, $\beta_{11} \neq \beta_{12}$

$\sum x_i = 280$ $\quad \sum y_{1j} = 126$ $\quad \sum x_i y_{1j} = 5827$

$\sum x_i^2 = 14000$ $\quad \sum y_{2j} = 130.6$ $\quad \sum x_i y_{2j} = 5630$

A사 회귀직선 $\hat{y}_{1j} = 6.756 + 0.2811 x_{1j}$

$$\hat{\beta}_{11} = \frac{S(xy)}{S(xx)} = \frac{5827 - \frac{1}{7} \times 126 \times 280}{14000 - 280^2 \times \frac{1}{7}} \doteqdot 0.2811$$

$$\hat{\beta}_{01} = \bar{y}_1 - \hat{\beta}_{11} \bar{x} = 18 - 0.2811 \times 40 \doteqdot 6.756$$

B사 회귀직선 $\hat{y}_{2j} = 12.857 + 0.145 x_{2j}$

$$\hat{\beta}_{12} = \frac{S(xy)}{S(xx)} = \frac{5630 - \frac{1}{7} \times 130.6 \times 280}{14000 - 280^2 \times \frac{1}{7}} \doteqdot 0.145$$

$$\hat{\beta}_{02} = \bar{y}_2 - \hat{\beta}_{12} \bar{x} = 18.65714 - 0.145 \times 40 \doteqdot 12.85714$$

A사 $SST = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{1}{n} (\sum y_i)^2$

$\doteqdot 2494.76 - \frac{1}{7} (126)^2 \doteqdot 226.76$

A사 $SSR = \sum (\bar{y} - \hat{y}_i)^2 = \frac{(S(xy))^2}{S(xx)} = \frac{787^2}{2800} = 221.203$

A사 $SSE = SST - SSR = 5.56$

B사 $SST = \sum (y_i - \bar{y})^2 = 2504 - \frac{1}{7} (130.6)^2 = 67.37714$

B사 $SSR = \sum (\bar{y} - \hat{y}_i)^2 = \frac{(S(xy))^2}{S(xx)} = \frac{406^2}{2800} \doteqdot 58.87$

B사 $SSE = SST - SSR = 67.37714 - 58.87 \doteqdot 8.5071$

5.(3) $H_0 : \beta_{11} = \beta_{12}$ vs $H_1 : \beta_{11} \neq \beta_{12}$

검정통계량

$$t_0 = \frac{\hat{\beta}_{11} - \hat{\beta}_{12}}{\sqrt{\widehat{var}(\hat{\beta}_{11} - \hat{\beta}_{12})}} \sim_{H_0} t((n_1-2)+(n_2-2))$$

$\hat{\beta}_{11} = 0.2811$, $\hat{\beta}_{12} = 0.145$

$SSE(F) = 14.0671$, $MSE(F) = \frac{SSE(F)}{df_F} = \frac{14.0671}{10} = 1.40671$

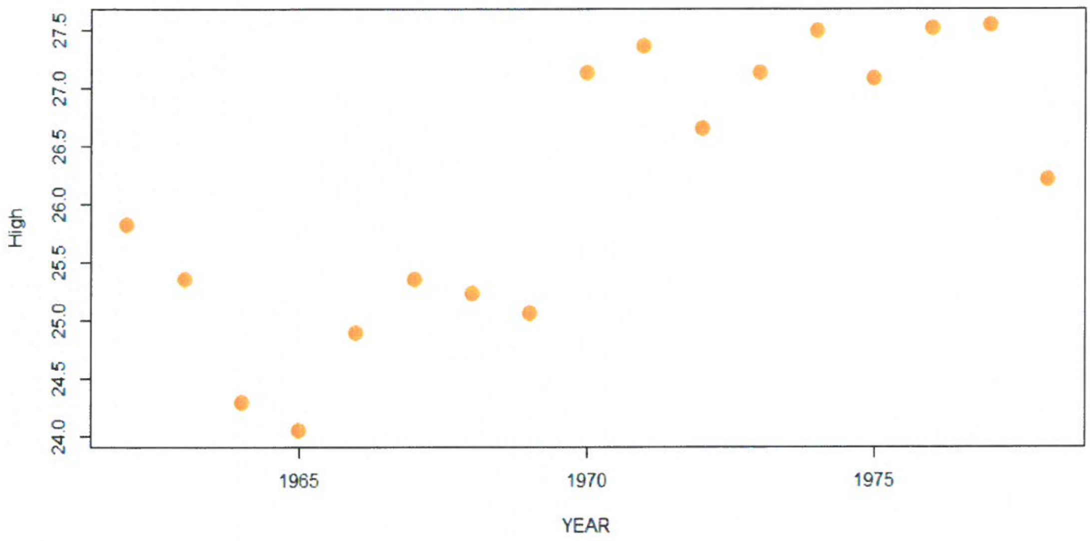
$$\widehat{Var}(\hat{\beta}_{11} - \hat{\beta}_{12}) = 1.40671 \left[\frac{1}{2800} + \frac{1}{2800} \right] = 0.001005$$

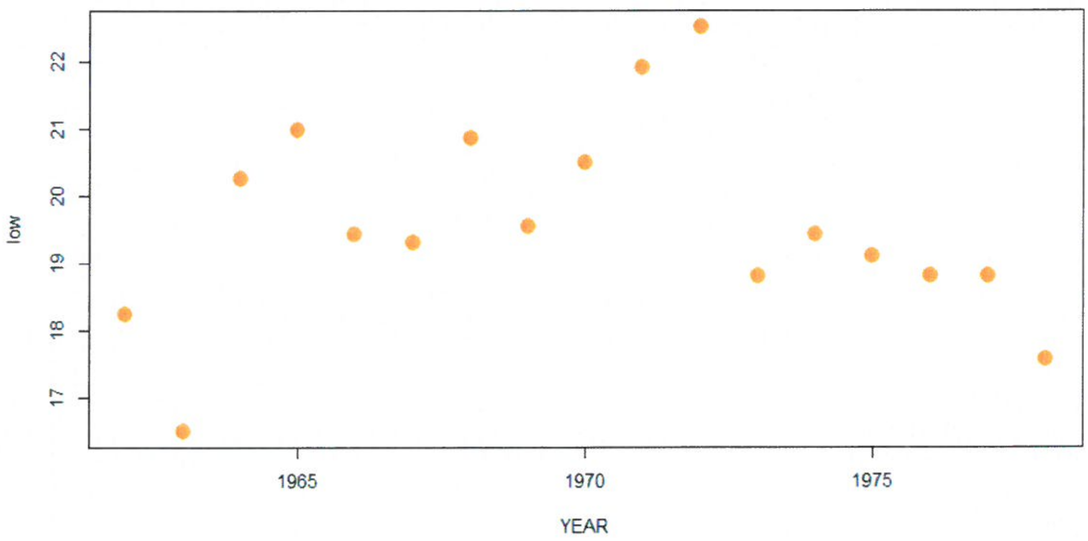
$$t_0 = \frac{0.2811 - 0.145}{\sqrt{0.001005}} \doteqdot 4.293583$$

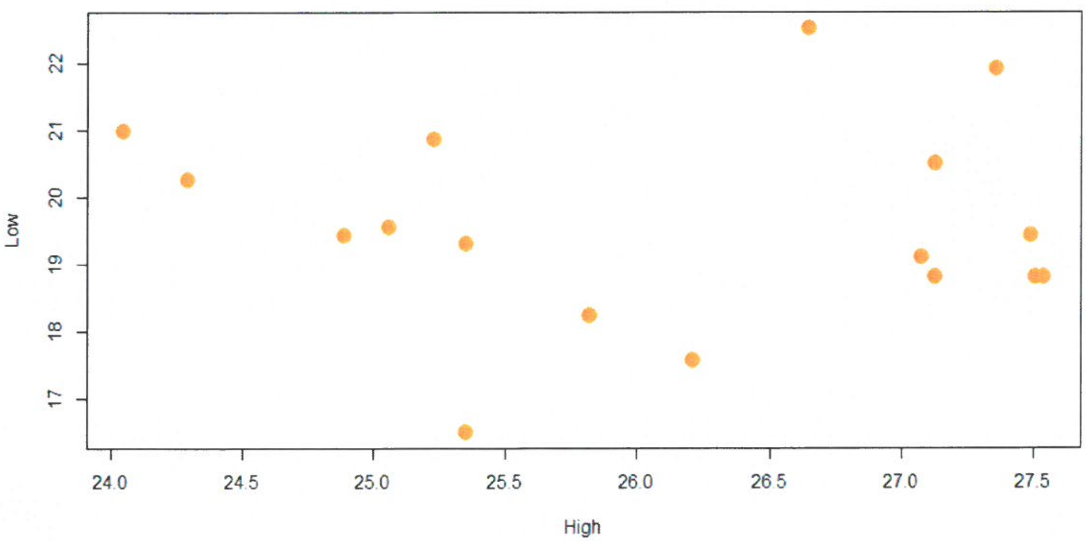
$|t_0| > t_{0.025}(10) = 2.228$ 이므로 H_0를 기각한다.

즉, 기울기는 같지 않다.

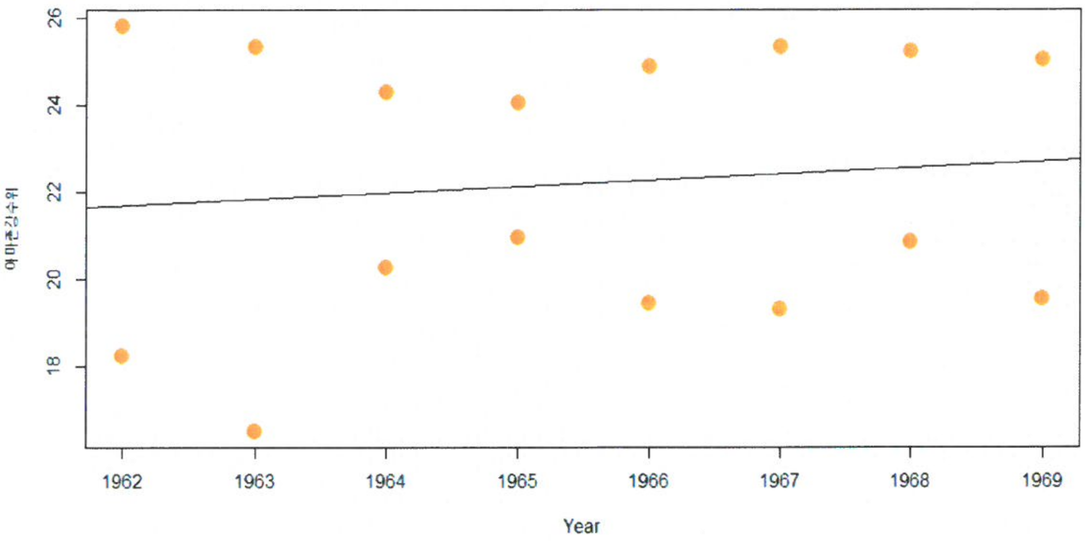
6. (1)

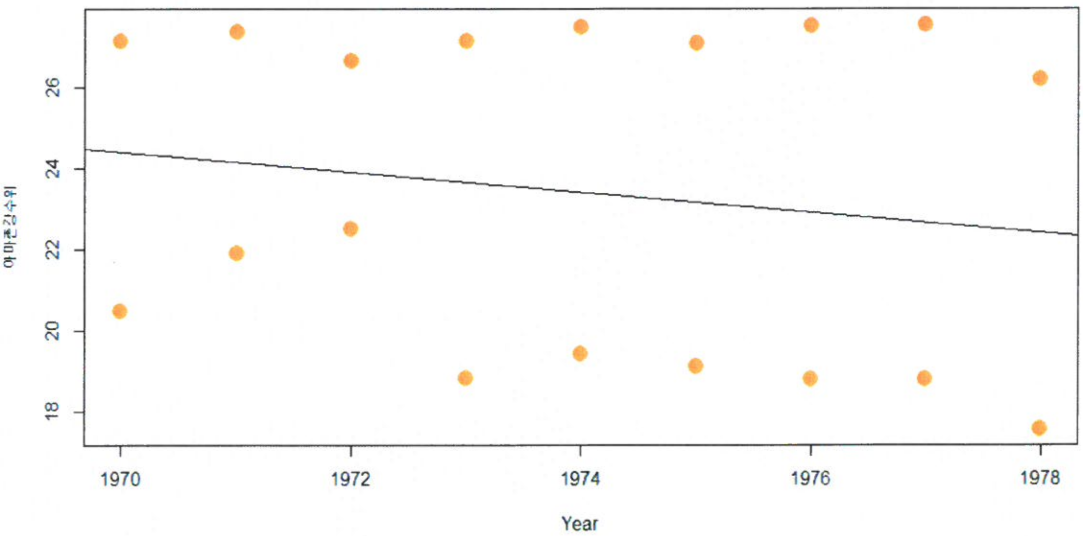

High
24.0
24.5
25.0
25.5
26.0
26.5
27.0
27.5
1965
1970
1975
YEAR


low
17
18
19
20
21
22
1965
1970
1975
YEAR


Low
17
18
19
20
21
22
24.0
24.5
25.0
25.5
26.0
26.5
27.0
27.5
High

6. (4)


아마존강수위
Year
26
24
22
20
18
1962
1963
1964
1965
1966
1967
1968
1969


아마존강수위
Year
26
24
22
20
18
1970
1972
1974
1976
1978

6. (2) Year에 대한 High의 회귀모형 ; Year가 1만큼 증가하면 물의수위(High)는 0.180882만큼 증가한다.

$$High = -330.2124 + 0.180882 \cdot Year$$

Year에 대한 low의 회귀모형 ; Year가 1만큼 증가하면 물의수위(Low)는 0.007392만큼 감소한다.

$$low = 35.106961 - 0.007392 \cdot Year$$

Low에 대한 High의 회귀모형

$$High = 26.40088 - 0.01406 \cdot Low$$

(3) 아마존 강 수위의 변화를 일으킨다고 볼수있다. → 선형관계가 있다.

(5) $H_0: \beta_{01} = \beta_{02},\ \beta_{11} = \beta_{12},\quad vs\quad H_1: \beta_{01} \neq \beta_{02}\ or\ \beta_{11} \neq \beta_{12}$

1960년대 회귀작선 $\hat{y}_{1j} = -265.1411 + 0.1462\, \hat{x}_{1j}$

ANOVA

요인	제곱합	자유도
SSR	1.795	1
SSE	141.912	14
SST	143.7076	15

1970년대 회귀작선 $\hat{y}_{2j} = 510.5017 - 0.2467\, \hat{x}_{2j}$

ANOVA

요인	제곱합	자유도
SSR	7.306	1
SSE	262.144	16
SST	269.45	17

Step 1 : $SSE(F) = SSE_1 + SSE_2 = 141.912 + 262.144 = 403.336$

$df_F = 14 + 16 = 30$

축소모형의 회귀직선 $\hat{y} = -147.5527 + 0.0865\hat{x}$

ANOVA

요인	제곱합	자유도
SSR	6.10	1
SSE	419.68	32
SST	425.78	33

Step 2 : $SSE(R) = 419.68$

$df_R = 32$

Step 3 : 검정통계량 $F_0 = \frac{419.68 - 403.336}{32-30} \div \frac{403.336}{30} \doteq 0.607830$

$F_0 < F(2, 32; 0.01) = 5.336$ 이므로 H_0 채택

즉, 두 회귀모형은 동일하다.

(6) $H_0: \beta_{11} = \beta_{12}$ vs $H_1: \beta_{11} \neq \beta_{12}$

$\hat{\beta}_{11} = 0.1462$, $\hat{\beta}_{12} = -0.2467$

$SSE(F) = 403.336$, $MSE(F) = \frac{403.336}{30} = 13.44453$

$\widehat{Var}(\hat{\beta}_{11} - \hat{\beta}_{12}) = 13.4445\left(\frac{1}{84} + \frac{1}{120}\right) \doteq 0.272092$

검정통계량 : $t_0 = \frac{0.1462 + 0.2467}{\sqrt{0.272092}} \doteq 0.753224$

$|t_0| < t_{0.005}(30) = 2.750$ 이므로 유의수준 1%에서

귀무가설 기각하지 못한다.