고급회귀분석론

이영미 교수님

학 과	통계학과
학 번	202250926
이 름	김보람
과 제 명	HW 04
제 출 일	2022.12.05.(월)

HW 4

1.(1) 각 성별 회귀분석 자료 (완전모형)

① 남자

$$X=\begin{pmatrix}1 & 151\\ 1 & 92\\ \vdots & \vdots\\ 1 & 290\\ 1 & 238\end{pmatrix} \quad y=\begin{pmatrix}17\\ 26\\ \vdots\\ 4\\ 16\end{pmatrix} \quad X^TX=\begin{pmatrix}1 & 1 & \cdots & 1 & 1\\ 151 & 92 & \cdots & 290 & 238\end{pmatrix}\begin{pmatrix}1 & 151\\ 1 & 92\\ \vdots & \vdots\\ 1 & 290\\ 1 & 238\end{pmatrix}=\begin{pmatrix}n & \Sigma x\\ \Sigma x & \Sigma x^2\end{pmatrix}=\begin{pmatrix}10 & 1688\\ 1688 & 349640\end{pmatrix}$$

$$(X^TX)^{-1}=\begin{pmatrix}0.5404 & -2.608739e^{-03}\\ -0.0026 & 1.545461e^{-05}\end{pmatrix} \quad X^Ty=\begin{pmatrix}\Sigma y\\ \Sigma xy\end{pmatrix}=\begin{pmatrix}168\\ 21897\end{pmatrix}$$

$$\hat{\beta}_{01}=(X^TX)^{-1}X^Ty=\begin{pmatrix}33.6561\\ -0.0999\end{pmatrix} \quad e=y-\hat{y}=y-X\hat{\beta}=\begin{pmatrix}-1.5775\\ 1.5309\\ \vdots\\ -0.6972\\ 6.1102\end{pmatrix}$$

$$\hat{y}_{1j}=33.6561-0.0999x_{1j}$$

$$SST=y^Ty-n(\bar{y})^2=\Sigma y^2-n(\bar{y})^2=3568-10\cdot(16.8)^2=745.6$$

$$SSE=e^Te=100.3747$$

$$SSR=SST-SSE=745.6-100.3747=645.2253$$

② 여자

$$X=\begin{pmatrix}1 & 164\\ 1 & 272\\ \vdots & \vdots\\ 1 & 124\\ 1 & 246\end{pmatrix} \quad y=\begin{pmatrix}28\\ 15\\ \vdots\\ 30\\ 14\end{pmatrix} \quad X^TX=\begin{pmatrix}10 & 1727\\ 1727 & 386463\end{pmatrix}$$

$$(X^TX)^{-1}=\begin{pmatrix}0.4361 & -1.957826e^{-03}\\ -0.0020 & 1.336571e^{-05}\end{pmatrix}$$

$$X^Ty=\begin{pmatrix}221\\ 33181\end{pmatrix} \quad y^Ty=5641$$

$$\hat{\beta}_{02}=(X^TX)^{-1}X^Ty=\begin{pmatrix}31.8611\\ -0.0565\end{pmatrix}$$

$$\hat{y}_{2j}=31.8611-0.0565x_{2j}$$

$$SST=y^Ty-n(\bar{y})^2=756.9$$

$$SSE=e^Te=475.1$$

$$SSR=SST-SSE=281.8$$

남자

$\hat{y}_{1j} = 33.6561 - 0.0999 x_{1j}$

ANOVA

요인	제곱합	자유도
회귀	$SSR_1 = 645.23$	1
잔차	$SSE_1 = 100.37$	8
계	$SST_1 = 1745.6$	9

여자

$\hat{y}_{2j} = 31.8611 - 0.0565 x_{2j}$

ANOVA

요인	제곱합	자유도
회귀	$SSR_2 = 281.8$	1
잔차	$SSE_2 = 475.1$	8
계	$SST_2 = 756.9$	9

$SSE(F) = SSE_1 + SSE_2 = 100.37 + 475.1 = 575.4747$

$df(F) = (10-2) + (10-2) = 16$

③ 축소모형

$$X = \begin{pmatrix} 1 & 151 \\ 1 & 92 \\ \vdots & \vdots \\ 1 & 124 \\ 1 & 246 \end{pmatrix} \quad y = \begin{pmatrix} 17 \\ 26 \\ \vdots \\ 30 \\ 14 \end{pmatrix} \quad X^TX = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ 151 & 92 & \cdots & 124 & 246 \end{pmatrix} \begin{pmatrix} 1 & 151 \\ 1 & 92 \\ \vdots & \vdots \\ 1 & 124 \\ 1 & 246 \end{pmatrix} = \begin{pmatrix} 20 & 3637 \\ 3637 & 776043 \end{pmatrix}$$

$$X^Ty = \begin{pmatrix} 389 \\ 58186 \end{pmatrix} \quad (X^TX)^{-1} = \begin{pmatrix} 0.2956 & -1.350493e^{-03} \\ -0.0014 & 7.426411e^{-06} \end{pmatrix}$$

$$\hat{\beta} = (X^TX)^{-1}X^Ty = \begin{pmatrix} 36.4036 \\ -0.0932 \end{pmatrix} \quad y^Ty = 9209$$

$SST = y^Ty - n(\bar{y})^2 = 1642.95$

$SSE = e^Te = 472.59$

$SSR = SST - SSE = 1170.36$

축소모형의 anova분석 자료

$\hat{y}_{ij} = 36.4036 - 0.0932 x_{ij}$

ANOVA

요인	제곱합	자유도
SSR	1170.36	1
SSE	472.59	18
SST	1642.95	19

$\therefore SSE(R) = 472.59$

$df_R = (n_1 - 1) + (n_2 - 1)$

$= 9 + 9$

$= 18$

$H_0: \beta_{01} = \beta_{02},\ \beta_{11} = \beta_{12}$ vs H_1 : not H_0

$$F_0 = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} = \frac{472.59 - 575.47}{18-16} \div \frac{575.47}{16} = -1.430258$$

$|F_0| < F_{0.05}(2,16) = 3.633723$ 이므로 H_0를 기각할 수 없다.

즉, 두 회귀계수의 기울기는 같다.

(2) $Var(\hat{\beta}_1) = \underbrace{(X^TX)^{-1}_{(1,1)}}_{\hookrightarrow 7.619135e^{-06}} \sigma^2$ $\quad s.e(\hat{\beta}_1) = \sqrt{C_{11} \cdot MSE}$

$t_{0.025}(17) = 2.109816$ $\quad = \sqrt{0.0023} = 0.0483$

95% 신뢰구간: $\hat{\beta}_1 \pm t_{0.025}(17) \cdot s.e(\hat{\beta}_1)$

$\Rightarrow -0.1009 \pm 2.109816 \times 0.0483$

$= (-0.2029,\ 0.0011)$

(3) $X = \begin{pmatrix} 1 & 151 & 0 & 0 & 1 \\ 1 & 92 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 124 & 1 & 0 & 1 \\ 1 & 246 & 1 & 0 & 0 \end{pmatrix}$ $y = \begin{pmatrix} 17 \\ 26 \\ \vdots \\ 30 \\ 14 \end{pmatrix}$ $X^Ty = \begin{pmatrix} 389 \\ 58186 \\ 221 \\ 116 \\ 138 \end{pmatrix}$

$$X^TX = \begin{pmatrix} 20 & 3637 & 10 & 4 & 6 \\ 3637 & 796043 & 1949 & 295 & 821 \\ 10 & 1949 & 10 & 1 & 3 \\ 4 & 295 & 1 & 4 & 0 \\ 6 & 821 & 3 & 0 & 6 \end{pmatrix}$$

$$\hat{\beta} = (X^TX)^{-1}X^Ty = \begin{pmatrix} 36.5227 \\ -0.1101 \\ 17.8991 \\ -1.3758 \\ -2.4036 \end{pmatrix}$$

$\hat{y} = 36.5227 - 0.1101x_1 + 7.8991x_2 - 1.3758x_3 - 2.4036x_4$

(4) $\hat{\beta}_3$: 대학원 졸업과 고등학교 졸업의 소요시간 차이

$\hat{\beta}_4$: 대학원 졸업과 대학교 졸업의 소요시간 차이

(5) $Var(\hat{\beta}_3) = (X^TX)^{-1}_{(4,4)}\,\sigma^2$

$s.e(\hat{\beta}_3) = \sqrt{C_{44} \cdot MSE} = \sqrt{1.6541 \times 10.33} = \sqrt{17.084} \doteq 4.133$

$t_{0.025}(15) = 2.1315$

$\hat{\beta}_3$의 95% 신뢰구간 : $\hat{\beta}_3 \pm t_{0.025}(15) \cdot s.e(\hat{\beta}_3)$

$\Rightarrow -1.3758 \pm 2.1315 \times 4.133 = (-10.1839,\ 7.4323)$

$Var(\hat{\beta}_4) = (X^TX)^{-1}_{(5,5)}\,\sigma^2$

$s.e(\hat{\beta}_4) = \sqrt{C_{55} \cdot MSE} = \sqrt{0.17708 \times 10.33} = \sqrt{8.169} \doteq 2.8580$

$\hat{\beta}_4$의 95% 신뢰구간 : $\hat{\beta}_4 \pm t_{0.025}(15) \cdot s.e(\hat{\beta}_4)$

$\Rightarrow -2.4036 \pm 2.1315 \times 2.8580 = (-8.4939,\ 3.6867)$

(6)

$$X = \begin{pmatrix} 1 & 151 & 0 & 0 & 1 & 0 & 0 & 151 & 0 & 0 \\ 1 & 92 & 0 & 1 & 0 & 0 & 92 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 246 & 1 & 0 & 0 & 246 & 0 & 0 & 0 & 0 \end{pmatrix} \quad y = \begin{pmatrix} 17 \\ 26 \\ \vdots \\ 14 \end{pmatrix}$$

$$\hat{\beta} = (X^TX)^{-1}X^Ty = \begin{pmatrix} 36.6397 \\ -0.1119 \\ 12.27994 \\ -3.42748 \\ -6.74634 \\ -0.01529 \\ 0.01659 \\ 0.03861 \\ 0.0294 \\ -4.10250 \end{pmatrix}$$

$\therefore \hat{y} = 36.6397 - 0.1119x_1 + 12.280x_2 - 3.4275x_3 - 6.7463x_4$
$- 0.0153x_1x_2 + 0.0166x_1x_3 + 0.0386x_1x_4$
$+ 0.0294x_2x_3 - 4.1025x_2x_4$

$H_0: \beta_5 = 0$ vs $H_1: \beta_5 \begin{matrix} >0 \\ \neq 0 \\ <0 \end{matrix}$

검정통계량 $t_0 = \dfrac{\hat{\beta}_5}{\widehat{s.e}(\hat{\beta}_5)} \sim t(0.95, 20-9-1) = 1.812$

$\widehat{s.e}(\hat{\beta}_5) = \sqrt{C_{66} \cdot MSE} = \sqrt{2.444926e^{-04} \times 12.86} = \sqrt{0.003111} = 0.05573$

$t_0 = \dfrac{-0.0153}{0.05573} \doteq -0.274$ 이고 $|t_0| < t(0.95, 10)$ 이므로 H_0를 기각못한다.

즉, $\beta_i (i = 5.6.7.8.9) = 0$ 이다.

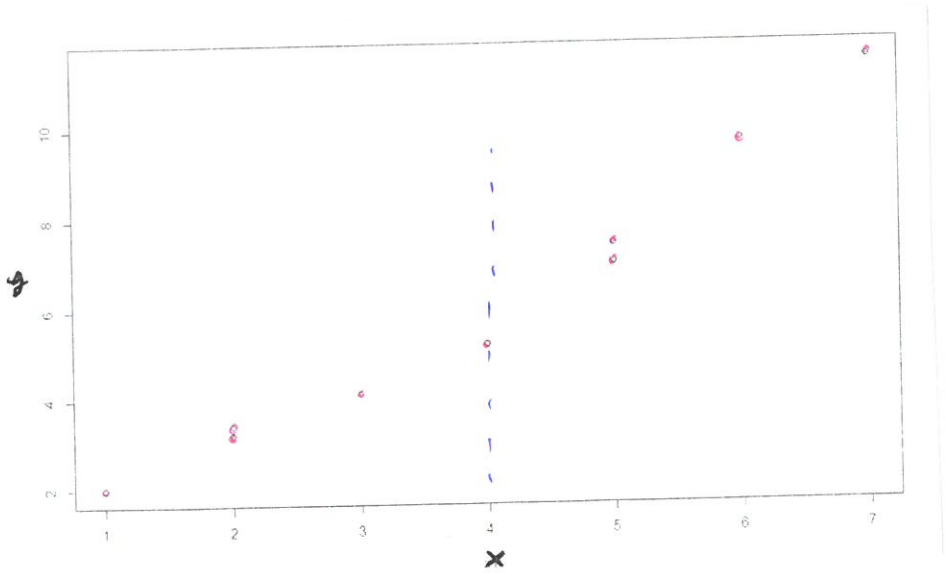
(ㄱ) $H_0: \beta_1 = \beta_2 = \cdots = \beta_9 = 0$ vs H_1: not H_0

ANOVA..

요인	제곱합	df	평균제곱	Fo
SSR	1514.345	9	168.261	13.084
SSE	128.605	10	12.860	
SST	1642.950	19		

$F(9, 10, 0.90) = 2.35 < F_0$ 이므로 귀무가설을 기각한다.

2.(1)



$x_w = 4$인 점에서 두 구간으로 나누는 것이 적당해 보인다.

(2) $x_{2i} = \begin{cases} 1 & x_{1i} > 4 \\ 0 & x_{1i} \le 4 \end{cases}$

$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 (x_{1i} - 4) x_{2i} + \varepsilon_i$

$y = \begin{pmatrix} 2.0 \\ 3.2 \\ \vdots \\ 11.5 \end{pmatrix}$ $X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \\ 1 & 4 & 0 \\ 1 & 5 & 1 \\ 1 & 5 & 1 \\ 1 & 6 & 2 \\ 1 & 7 & 3 \end{pmatrix}$ (1, x_1, $(x_1-4)x_2$)

$X^TX = \begin{pmatrix} 9 & 15 & 7 \\ 35 & 169 & 43 \\ 7 & 43 & 15 \end{pmatrix}$

$(X^TX)^{-1} = \begin{pmatrix} 1.1667 & -0.3810 & 0.5476 \\ -0.3810 & 0.1463 & -0.2415 \\ 0.5476 & -0.2415 & 0.5034 \end{pmatrix}$

$X^Ty = \begin{pmatrix} 53.5 \\ 259.0 \\ 68.3 \end{pmatrix}$ $\hat{\beta} = (X^TX)^{-1}X^Ty = \begin{pmatrix} 1.152381 \\ 1.005782 \\ 1.132313 \end{pmatrix}$

$\therefore \hat{y} = 1.152381 + 1.005782 x_1 + 1.132313 x_2$

(3) $Var(\hat{\beta}_2) = (X^TX)^{-1}_{(3,3)} \cdot \sigma^2$

$SSE = e^Te = (y - X\beta)^T(y - X\beta) = 0.2630\,3$

$MSE = SSE/(9-2-1) = SSE/6 = 0.0438$

$s.e(\hat{\beta}_2) = \sqrt{(X^TX)^{-1}_{(3,3)} \cdot MSE} = \sqrt{0.5034 \cdot 0.0438} = \sqrt{0.0221} \doteq 0.1485$

$t(0.9, 6) = 1.4398$

β_2의 90% 신뢰구간 : $\hat{\beta}_2 \pm t_{0.05}(6) \cdot s.e(\hat{\beta}_2)$

$\Rightarrow 1.1323 \pm 1.4398 \times 0.1485$

$\Rightarrow (0.9185, 1.3461)$

$x > 4$ 일때 y의 이차는 효과는 0.9185에서 1.3461까지 범위를 가진다.