수리통계학 × 통계학의 역사

13회 워크북

최소제곱
손으로 푸는 34문항

강의노트 13회를 읽은 뒤 손으로 푼다. 최소제곱의 계산은 작은 자료로 직접 해 보고, Gauss–Markov, Cochran, Frisch–Waugh–Lovell의 증명은 각 단계의 근거를 빈 종이에서 다시 찾는다. 문제 아래의 줄 친 공간에 직접 쓴다.

부	하는 일	문항	예상 시간
A 계산 드릴	작은 자료의 회귀, 사영, 지렛대, t와 F를 손으로 계산한다	14	1시간 51분
B 빈칸 증명	노트의 단계형 증명을 책을 덮고 다시 쓴다	7	2시간 55분
C 증명 문제	변형, 일반화, 반례 만들기	9	2시간 45분
D 퀄형	여러 절을 잇는 종합 문제. 제한 시간 안에	4	2시간 25분
	합계	34	약 10시간

노트 `13_least_squares_notes.pdf` — 문항의 "정리 3.3" 같은 번호는 노트의 번호다

도구 `00_toolkit_notes.pdf` — [T1] 원소 추적, [T2] 귀류·반례, [T4] 부등식, [T6] 적률생성함수

검산 노트북 `13_least_squares.ipynb`의 "워크북 검산" 절: `check("A1a", "3/5")`

이 워크북을 쓰는 법

1. **노트를 덮고 푼다.** 문제의 머리 오른쪽에 쓸 도구, 관련 정리 번호, 예상 시간이 있다. 예상 시간의 두 배가 지나도 진척이 없으면 그때 노트를 편다.
2. **막히면 3단으로 본다.** 뒤의 **힌트와 최종답**에서 힌트를 먼저 본다. 다 푼 뒤에는 최종답과 맞춰 본다. 그래도 어긋나거나 증명이 막히면 풀이판을 본다.
3. **B부는 순서대로 푼다.** 같은 정리를 여러 번 만난다. 처음에는 근거만 빈칸이고(①), 다음에는 단계의 이름만 있고(②), 마지막에는 전략만 있다(③). 앞 단계의 답을 보면서 다음 단계를 쓰지 않는다.
4. **계산은 검산한다.** 최종답이 수나 식인 문항은 노트북의 `check()`에 넣으면 맞는지만 알려 준다. 정답은 보여 주지 않는다. σ^2이 곱해진 분산은 σ^2의 계수만 넣는다. 분포표에서 읽는 값은 소수 넷째 자리까지 맞으면 맞다고 본다.
5. **증명은 근거를 단다.** 등호와 부등호마다 어느 정의, 정리, 앞 회차의 결과를 썼는지 오른쪽에 적는다. 노트의 단계형 증명과 같은 모양이다.

표시 ★ 정의를 바로 쓰는 문제, ★★ 두세 단계를 잇는 문제, ★★★ 퀄에 나오는 길이의 문제. [T1]–[T8]은 증명 도구 상자의 도구 번호다.

A

계산 드릴

작은 자료로 최소제곱을 손으로 돌린다. 답이 수로 떨어지므로 다 풀면 뒤의 최종답이나 노트북의 `check()`로 맞춰 본다. 계산기는 마지막 나눗셈과 분포의 꼬리에만 쓴다.

A1 다섯 점의 단순회귀 ★

정리 2.2, 명제 2.6 · 10분

자료 $x = (1, 2, 3, 4, 5)$, $y = (2, 4, 5, 4, 5)$에 $y_i = b_0 + b_1 x_i + \varepsilon_i$를 최소제곱으로 맞춘다. (a) $\hat{b}_1$, (b) $\hat{b}_0$, (c) RSS, (d) R^2을 구하라. 잔차의 합이 0인지도 확인하라.

A2 단순회귀의 지렛대 ★

명제 2.7 · 6분

A1의 설계에서 (a) h_{55}와 (b) h_{33}을 구하라. 다섯 지렛대의 합이 p와 같은지 확인하라.

A3 세 점의 사영 ★

정리 2.2, 정리 2.5 · 10분

$X = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{smallmatrix}\right)$, $\mathbf{y} = (1, 2, 6)'$ 이다. 정규방정식을 풀어 (a) 기울기 $\hat\beta_1$과 (b) RSS를 구하라. 잔차가 X의 두 열과 직교하는지 확인하라.

A4 손으로 하는 QR ★★

알고리즘 2.8 · 12분

X의 두 열이 $\mathbf{1} = (1, 1, 1, 1)'$과 $\mathbf{x} = (0, 1, 2, 3)'$이다. (a) Gram–Schmidt로 $X = QR$을 만들고 r_{22}를 구하라. (b)(c) $\mathbf{y} = (1, 3, 2, 4)'$에 대해 $\mathbf{z} = Q'\mathbf{y}$를 구하고 $R\hat{\boldsymbol\beta} = \mathbf{z}$를 거꾸로 대입해 $\hat\beta_1$과 $\hat\beta_0$을 구하라.

A5 이차형식의 기댓값 ★

보조정리 3.1 · 6분

$\mathbf{u}$의 평균이 $\boldsymbol\mu = (1, 2)'$, 공분산이 $\Sigma = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 3 \end{smallmatrix}\right)$이고 $A = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 2 \end{smallmatrix}\right)$이다. $E\,\mathbf{u}'A\mathbf{u}$를 구하라.

A6 잔차제곱합의 평균과 분산 ★

정리 3.2(c), 정리 4.7(b) · 6분

정규 선형모형에서 $n = 20$, $p = 4$, $\sigma^2 = 2$이다. (a) $E\,\mathrm{RSS}$와 (b) $\mathrm{Var}\,\mathrm{RSS}$를 구하라.

A7 계수의 공분산 ★

정리 3.2(a) · 8분

$x = (0, 1, 2, 3, 4, 5)$, $\sigma^2 = 1$인 단순회귀에서 (a) $\mathrm{Var}\,\hat{b}_1$, (b) $\mathrm{Var}\,\hat{b}_0$, (c) $\mathrm{Cov}(\hat{b}_0, \hat{b}_1)$을 $(X'X)^{-1}$로 구하라.

A8 두 끝점의 기울기 ★

명제 1.4, 정리 3.3 · 8분

$x = (-2, -1, 0, 1, 2)$인 단순회귀에서 기울기를 두 끝점으로 $\tilde{b}_1 = (y_5 - y_1)/4$로 추정한다. $\tilde{b}_1$이 선형불편추정량임을 보이고, $\mathrm{Var}\,\tilde{b}_1 / \mathrm{Var}\,\hat{b}_1$을 구하라.

A9 계수의 t와 구간 ★

따름정리 4.8 · 8분

$n = 24$, $p = 4$인 정규 선형모형에서 $\hat{\beta}_j = 1.8$, $s = 2$, $[(X'X)^{-1}]_{jj} = 0.09$이다. (a) $\mathrm{SE}(\hat{\beta}_j)$, (b) $H_0 : \beta_j = 0$의 t 통계량, (c) 95% 신뢰구간의 반폭을 구하라. ($t_{20,0.025} = 2.086$)

A10 일반선형가설의 F ★

정리 4.9 · 8분

$n = 30$, $p = 5$인 정규 선형모형에서 $q = 2$개의 제약을 건 모형의 $\mathrm{RSS}_0 = 150$이고 전체 모형의 $\mathrm{RSS} = 120$이다. (a) F 통계량과 (b) 유의확률을 구하라. 분자 자유도가 2인 $F(2, m)$의 꼬리확률은 $P(F > f) = (1 + 2f/m)^{-m/2}$이다.

A11 R^2에서 F로 ★

예제 4.10 · 5분

상수항이 있는 모형에서 $n = 33$, $p = 3$, $R^2 = 0.4$이다. "상수항 말고 모든 계수가 0"을 검정하는 F를 구하라.

A12 분산팽창계수 ★★

따름정리 5.4 · 6분

(a) 두 설명변수의 상관이 0.9이면 각 계수의 VIF는? (b) 세 설명변수의 상관이 모두 0.5이면 x_1을 x_2, x_3으로 맞춘 R_1^2과 VIF_1은?

A13 한 점을 빼고 맞춘 잔차 ★★

명제 7.3 · 10분

A1의 자료에서 (a) 다섯째 점을 빼고 맞춘 직선으로 그 점을 예측한 잔차 $y_5 - \mathbf{x}_5'\hat{\boldsymbol{\beta}}_{(5)}$와 (b) 다섯 점 모두의 PRESS $\sum_i \big(e_i/(1-h_{ii})\big)^2$을 구하라. 다시 맞추지 말고 A1, A2의 결과만 써라.

A14 Ridge 한 번 ★

정의 8.4, 명제 8.5 · 8분

$X'X = \mathrm{diag}(4, 1)$, $X'\mathbf{y} = (8, 3)'$ 이고 $\lambda = 1$ 이다. Ridge 추정량의 (a) 첫 성분과 (b) 둘째 성분을 구하고, 최소제곱 해와 비교해 각 성분이 몇 배로 줄었는지 써라.

B

빈칸 증명

노트의 단계형 증명을 책을 덮고 다시 쓴다. 같은 정리를 여러 번 만나며 도움이 줄어든다. ① 단계는 주어지고 근거만 빈칸이다. ② 단계의 이름만 주어진다. ③ 전략만 주어진다. 근거에는 정의나 정리의 번호, 앞 회차의 결과, 도구 번호를 쓴다.

B1 Gauss–Markov, 근거 채우기 ★

① · 정리 3.3 · 15분

$\tilde{\boldsymbol{\beta}} = A\mathbf{y}$가 선형불편추정량이고 $A_0 = (X'X)^{-1}X'$, $D = A - A_0$이다. $\operatorname{Cov}\tilde{\boldsymbol{\beta}} - \operatorname{Cov}\hat{\boldsymbol{\beta}} = \sigma^2 DD' \succeq 0$의 증명이다. 근거를 써라.

① $AX = I$이고 $A_0X = I$이므로 $DX = 0$이다. 왜? _____

② $DA_0' = DX(X'X)^{-1} = 0$이고 $A_0D' = 0$이다. 왜? _____

③ 공분산을 전개한다.

$$\operatorname{Cov}\tilde{\boldsymbol{\beta}} = \sigma^2(A_0 + D)(A_0 + D)'$$ 왜? _____

$$= \sigma^2\left(A_0A_0' + DD'\right)$$ 왜? _____

$$= \operatorname{Cov}\hat{\boldsymbol{\beta}} + \sigma^2 DD'$$ 왜? _____

④ $\mathbf{c}'DD'\mathbf{c} = \|D'\mathbf{c}\|^2 \geq 0$이므로 $\sigma^2 DD' \succeq 0$이다. 왜? _____

B2 Cochran의 정리, 근거 채우기 ★★

① · 정리 4.5 · 25분

$I_n = A_1 + \cdots + A_k$, A_i 대칭, $\operatorname{rank} A_i = r_i$, $\sum_i r_i = n$이면 $\mathbf{Z}'A_i\mathbf{Z}$들이 서로 독립인 $\chi^2(r_i)$임을 보이는 증명이다. 근거를 써라.

① $\operatorname{rank}(I - A_i) \le \sum_{j \neq i} r_j = n - r_i$이다. 왜? _____

② $\operatorname{rank}(I - A_i) \ge n - r_i$이다. 왜? _____

③ A_i의 고유값 1의 고유공간의 차원은 r_i이다. 왜? _____

④ $A_i^2 = A_i$이다. 왜? _____

⑤ $j \neq i$이면 $A_j A_i = 0$이다. 왜? _____

⑥ $A_i = U_i U_i'$로 쓰면 $U = [U_1, \dots, U_k]$는 직교행렬이다. 왜? _____

⑦ $\mathbf{W} = U'\mathbf{Z}$의 블록의 제곱합 $\|U_i'\mathbf{Z}\|^2$이 서로 독립인 $\chi^2(r_i)$이다.

왜? _____

B3 Frisch–Waugh–Lovell, 근거 채우기 ★

① · 정리 5.3 · 15분

$X = [X_1, X_2]$가 완전 계수이고 $M_1 = I - H_1$이다. $\hat{\boldsymbol{\beta}}_2 = (X_2'M_1X_2)^{-1}X_2'M_1\mathbf{y}$의 증명이다. 근거를 써라.

① $\mathbf{y} = X_1\hat{\boldsymbol{\beta}}_1 + X_2\hat{\boldsymbol{\beta}}_2 + \mathbf{e}$이고 $X_1'\mathbf{e} = \mathbf{0}$, $X_2'\mathbf{e} = \mathbf{0}$이다. 왜? _____

② $M_1X_1 = 0$이고 $M_1\mathbf{e} = \mathbf{e}$이다. 왜? _____

③ $M_1\mathbf{y} = M_1X_2\hat{\boldsymbol{\beta}}_2 + \mathbf{e}$이다. 왜? _____

④ $(M_1X_2)'\mathbf{e} = \mathbf{0}$이다. 왜? _____

⑤ $\hat{\boldsymbol{\beta}}_2$는 $M_1\mathbf{y}$를 M_1X_2로 맞추는 정규방정식을 푼다. 왜? _____

⑥ 그 해는 하나뿐이다. 왜? _____

B4 Gauss–Markov, 단계의 이름만 ★★

② · 정리 3.3 · 25분

정리 3.3을 증명하라. 단계의 이름만 주어진다. 각 단계 아래에 식과 근거를 써라.

① **계수행렬을 $A = A_0 + D$로 나누고 불편성을 D의 조건으로 바꾼다**

② D와 A_0가 "직교"함을 보인다

③ $\mathrm{Cov}\,\tilde{\boldsymbol{\beta}}$을 전개해 교차항을 지운다

④ $\sigma^2 DD'$가 양의 준정치임을 보이고 결론을 쓴다

B5 Cochran의 정리, 단계의 이름만 ★★★

② · 정리 4.5 · 35분

정리 4.5의 $\Leftarrow$ 방향을 증명하라.

① $I - A_i$의 계수를 위에서 누른다

② $I - A_i$의 계수를 아래에서 누른다

③ A_i의 고유값 1의 고유공간의 차원을 구하고 A_i가 멱등임을 보인다

④ 서로 다른 A_i, A_j의 곱이 0임을 보인다

⑤ **기저를 모아 직교행렬을 만든다**

⑥ **$\mathbf{Z}$를 회전해 독립인 χ^2을 얻는다**

B6 Gauss–Markov, 전략만 ★★

③ · 정리 3.3, 따름정리 3.4 · 30분

정리 3.3을 증명하고, 따름정리 3.4(b) "$\mathrm{Var}(\mathbf{c}'\tilde{\boldsymbol{\beta}}) = \mathrm{Var}(\mathbf{c}'\hat{\boldsymbol{\beta}})$이면 $\mathbf{c}'\tilde{\boldsymbol{\beta}} = \mathbf{c}'\hat{\boldsymbol{\beta}}$"도 증명하라.

전략 아무 선형불편추정량의 계수행렬을 최소제곱의 계수와 그 차로 나눈다. 불편성이 차와 최소제곱 계수의 직교성을 주고, 공분산이 두 조각의 합으로 갈라진다.

B7 Frisch–Waugh–Lovell, 전략만 ★★

③ · 정리 5.3 · 30분

정리 5.3을 증명하라. 그리고 $X_1 = \mathbf{1}$인 경우에 결론이 "단순회귀의 기울기는 중심화한 자료의 기울기" 임을 확인하라.

전략 전체 회귀의 식에 M_1을 곱해 X_1을 지우고, 남은 식이 새 회귀의 정규방정식임을 확인한다.

C

증명 문제

노트의 정리를 변형하고, 일반화하고, 반례로 경계를 확인한다. 증명은 노트의 단계형 모양으로 쓴다. 한 줄에 한 단계, 줄마다 근거.

C1 원점을 지나는 직선의 BLUE ★

[T4] · 명제 1.4, 정리 3.3 · 15분

$y_i = \beta x_i + \varepsilon_i$ $(i = 1, \dots, n)$이고 x_i가 모두 0은 아니며 오차는 Gauss–Markov 가정을 만족한다. 행렬을 쓰지 말고 다음을 보여라.

(a) $\sum_i w_i y_i$가 불편 $\iff$ $\sum_i w_i x_i = 1$.

(b) 그런 w에 대해 $\sum_i w_i^2 \geq 1/\sum_i x_i^2$이고, 등호는 $w_i = x_i / \sum_j x_j^2$일 때뿐이다.

(c) β의 BLUE는 $\sum_i x_i y_i / \sum_i x_i^2$이다.

C2 포함된 두 사영 ★★

[T1] · 보조정리 2.4 · 15분

P와 Q가 $n \times n$ 대칭 멱등행렬이고 $\mathrm{col}(Q) \subseteq \mathrm{col}(P)$이다.

(a) $PQ = Q$와 $QP = Q$를 보여라.

(b) $P - Q$가 대칭 멱등이고 $\mathrm{rank}(P - Q) = \mathrm{rank}\, P - \mathrm{rank}\, Q$임을 보여라.

(c) $\mathrm{col}(P - Q) = \mathrm{col}(P) \cap \mathrm{col}(Q)^{\perp}$ 임을 보여라.

C3 상수항이 없으면 ★

[T2] · 따름정리 2.3, 명제 2.6 · 15분

$x = (1, 2, 3)$, $y = (3, 2, 1)$에 원점을 지나는 직선 $y = bx$를 최소제곱으로 맞춘다. (a) 잔차의 합을 구하라. (b) $1 - \mathrm{RSS}/\mathrm{TSS}$ ($\mathrm{TSS} = \sum (y_i - \bar{y})^2$)를 구하라. (c) 따름정리 2.3과 명제 2.6의 어느 가정이 깨졌는지 밝혀라.

C4 포함된 두 모형의 F 검정 ★★

정리 4.5, 4.9 · 25분

$\mathrm{col}(X_0) \subseteq \mathrm{col}(X)$이고 $\mathrm{rank}\, X_0 = p_0 < p = \mathrm{rank}\, X$이다. 두 모형의 hat 행렬을 H_0, H라 하자. $\mathbf{y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \sigma^2 I)$이다.

(a) $\mathrm{RSS}_0 - \mathrm{RSS} = \mathbf{y}'(H - H_0)\mathbf{y}$를 보여라.

(b) $\boldsymbol{\mu} \in \mathrm{col}(X_0)$이면 $\mathbf{Z} = (\mathbf{y} - \boldsymbol{\mu})/\sigma$와 $I = H_0 + (H - H_0) + (I - H)$에 Cochran의 정리를 써서

$$F = \frac{(\mathrm{RSS}_0 - \mathrm{RSS})/(p - p_0)}{\mathrm{RSS}/(n - p)} \sim F(p - p_0,\ n - p)$$

임을 보여라.

C5 무게를 버리면 잃는 것 ★★

[T4] · 따름정리 3.7 · 20분

$y_i = \mu + \varepsilon_i$이고 ε_i는 서로 무상관, $\mathrm{Var}\,\varepsilon_i = \sigma^2/w_i$ ($w_i > 0$ 알려짐)이다.

(a) 따름정리 3.7로 μ의 BLUE가 $\hat{\mu}_w = \sum w_i y_i / \sum w_i$이고 $\mathrm{Var}\,\hat{\mu}_w = \sigma^2 / \sum w_i$임을 보여라.

(b) 무게를 무시한 $\bar{y}$의 분산을 구하라.

(c) $\mathrm{Var}\,\bar{y} \geq \mathrm{Var}\,\hat{\mu}_w$를 Cauchy–Schwarz로 직접 보이고 등호 조건을 밝혀라.

C6 모집단의 분산 분해 ★★

정리 5.1 · 20분

정리 5.1의 가정에서 $\mathbf{x}$의 첫 성분이 상수 1이라 하고 $u = Y - \mathbf{x}'\boldsymbol{\beta}^*$이다.

(a) $Eu = 0$이고 $\mathbf{x}$의 모든 성분과 u의 공분산이 0임을 보여라.

(b) $\operatorname{Var} Y = \operatorname{Var}(\mathbf{x}'\boldsymbol{\beta}^*) + \operatorname{Var} u$를 보여라.

(c) $\mathbf{x} = (1, X)'$이면 $\operatorname{Var}(\mathbf{x}'\boldsymbol{\beta}^*)/\operatorname{Var} Y = \operatorname{Corr}(X, Y)^2$임을 보여라.

C7 지렛대가 1이면 ★★

[T2] · 명제 2.7, 7.3 · 15분

$\operatorname{rank} X = p$이고 어떤 i에서 $h_{ii} = 1$이다. $\mathbf{u}_i$를 i번째 단위벡터라 하자.

(a) $j \neq i$이면 $h_{ij} = 0$임을 보이고, 모든 $\mathbf{y}$에서 $e_i = 0$임을 보여라.

(b) $H\mathbf{u}_i = \mathbf{u}_i$를 보이고, $X\mathbf{v} = \mathbf{u}_i$인 $\mathbf{v} \neq \mathbf{0}$가 있음을 보여라.

(c) 관측 i를 뺀 설계 $X_{(i)}$의 계수가 p보다 작음을 보여라. 명제 7.3이 $h_{ii} < 1$을 가정하는 이유는 무엇인가?

C8 Ridge는 모든 방향에서 분산을 줄인다 ★★

[T4] · 명제 8.5, 정리 8.6 · 20분

정의 1.1의 모형에서 $\hat{\boldsymbol{\beta}}_\lambda = (X'X + \lambda I)^{-1} X'\mathbf{y}$ $(\lambda > 0)$이다.

(a) $E\hat{\boldsymbol{\beta}}_\lambda = (X'X + \lambda I)^{-1} X'X\boldsymbol{\beta}$를 보이고, $\boldsymbol{\beta} \neq \mathbf{0}$이면 치우쳤음을 보여라.

(b) $\operatorname{Cov} \hat{\boldsymbol{\beta}}_\lambda = \sigma^2 (X'X + \lambda I)^{-1} X'X (X'X + \lambda I)^{-1}$을 보여라.

(c) 특이값 분해 $X = UDV'$로 $\operatorname{Cov} \hat{\boldsymbol{\beta}} - \operatorname{Cov} \hat{\boldsymbol{\beta}}_\lambda \geq 0$을 보여라.

C9 공분산 0인 두 이차형식 ★★

[T2] · 보조정리 4.3, 정리 4.4 · 20분

$\mathbf{Z} \sim \mathcal{N}_2(\mathbf{0}, I)$이고 $Q_1 = Z_1^2 - Z_2^2$, $Q_2 = Z_1^2 + Z_2^2$이다.

(a) 보조정리 4.3으로 $\mathrm{Cov}(Q_1, Q_2) = 0$을 보여라.

(b) $P(|Q_1| > 1) > 0$, $P(Q_2 < 1) > 0$인데 $P(|Q_1| > 1,\ Q_2 < 1) = 0$임을 보이고, Q_1과 Q_2가 독립이 아님을 결론지어라.

(c) 정리 4.4(c)의 어느 가정이 깨졌는가? AB를 계산하라.

D

퀄형

여러 절의 도구를 잇는 종합 문제다. 머리의 제한 시간을 재고, 노트를 덮고 푼다. 모형을 행렬로 적는 것부터 답안의 일부다.

D1 정규 단순회귀의 모든 것 ★★★ 정리 3.6, 4.4, 따름정리 4.8 · 제한 40분

$y_i = b_0 + b_1 x_i + \varepsilon_i$, $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$이다.

(a) (b_0, b_1, σ^2)의 MLE를 구하라.

(b) $\hat{b}_1$과 $\bar{y}$가 독립임을 보여라.

(c) $(\hat{b}_1 - b_1) \big/ \left(s/\sqrt{S_{xx}}\right) \sim t(n-2)$임을 보여라. $s^2 = \text{RSS}/(n-2)$이다.

(d) A1의 자료에서 $x_0 = 6$의 적합값 $\hat{y}_0$를 구하라.

(e) 같은 점에서 새 관측 y_0의 95% 예측구간의 반폭을 구하라. ($t_{3,0.025} = 3.182$)

D2 절편이 0인가 ★★★

정리 4.9, 따름정리 4.8 · 제한 35분

A1의 자료에서 $H_0 : b_0 = 0$을 두 방법으로 검정한다.

(a) 원점을 지나는 직선으로 맞춘 RSS_0을 구하라.

(b) 정리 4.9의 F를 구하라.

(c) 유의확률을 구하라. 자유도 3인 t에 대해 $P(|T| > c) = 1 - \frac{2}{\pi}\left[\arctan a + \frac{a}{1+a^2}\right]$, $a = c/\sqrt{3}$이다.

(d) $\hat{b}_0$의 t 통계량을 따름정리 4.8로 따로 구해 $F = t^2$임을 확인하라. 이것이 일반적으로 성립하는 이유를 정리 4.9(b)로 설명하라.

D3 원점을 지나는 직선의 세 추정량 ★★★

[T4] · 정리 3.3 · 제한 35분

$y_i = \beta x_i + \varepsilon_i$, $x_i = i$ $(i = 1, \dots, 10)$이고 오차는 Gauss–Markov 가정을 만족한다. 세 추정량을 비교한다.

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}, \qquad \tilde{\beta}_R = \frac{\sum y_i}{\sum x_i}, \qquad \tilde{\beta}_M = \frac{1}{n}\sum_i \frac{y_i}{x_i}$$

세 추정량이 모두 선형불편임을 보이고, (a)–(c)의 분산은 σ^2의 계수로 답하라.

(a) $\operatorname{Var}\hat{\beta}$ (b) $\operatorname{Var}\tilde{\beta}_R$ (c) $\operatorname{Var}\tilde{\beta}_M$ (소수로)

(d) $\operatorname{Var}\tilde{\beta}_R / \operatorname{Var}\hat{\beta}$를 구하라.

(e) 모든 양수 $x_1, \dots, x_n$에서 $\operatorname{Var}\hat{\beta} \le \operatorname{Var}\tilde{\beta}_R \le \operatorname{Var}\tilde{\beta}_M$ 임을 보여라.

D4 두 표본 비교를 선형모형으로 ★★★

정리 4.9, 6회 정리 5.4 · 제한 35분

집단 1의 관측이 $(3, 5, 7)$, 집단 2의 관측이 $(6, 8, 10, 12)$이고 $y_{jk} = \mu_j + \varepsilon_{jk}$, $\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$이다.

(a) 두 집단 표시를 열로 하는 7×2 설계 X를 쓰고 $\hat{\boldsymbol{\beta}}$과 RSS를 구하라.

(b) $H_0 : \mu_1 = \mu_2$를 $C = (1, -1)$로 쓰고 정리 4.9(b)로 $\text{RSS}_0 - \text{RSS} = (\bar{y}_1 - \bar{y}_2)^2 / (\frac{1}{n_1} + \frac{1}{n_2})$을 보여라.

(c) F를 구하고, 이것이 합동 t 통계량의 제곱임을 확인하라.

(d) 유의확률을 구하라. ($P(|t_5| > 2.213)$을 계산기로)

힌트와 최종답

막히면 먼저 힌트를 보고, 다 푼 뒤에 최종답과 맞춰 본다. 그래도 어긋나면 풀이판을 본다. 최종답은 노트북의 `check()`로도 확인할 수 있다.

A1 힌트 $\bar{x}=3$, $\bar{y}=4$에서 S_{xx}와 S_{xy}를 먼저 구한다. $\hat{b}_1 = S_{xy}/S_{xx}$, $\hat{b}_0 = \bar{y} - \hat{b}_1\bar{x}$.

답 (a) $\hat{b}_1 = 0.6$ (b) $\hat{b}_0 = 2.2$ (c) RSS $= 2.4$ (d) $R^2 = 0.6$

A2 힌트 단순회귀에서 $h_{ii} = \frac{1}{n} + (x_i-\bar{x})^2/S_{xx}$이다(노트 연습 2.3).

답 (a) $h_{55} = 0.6$ (b) $h_{33} = 0.2$

A3 힌트 $X'X = \left(\begin{smallmatrix}3 & 3\\ 3 & 5\end{smallmatrix}\right)$, $X'\mathbf{y} = (9, 14)'$이다. 2×2 역행렬은 행렬식으로 쓴다.

답 (a) $\hat{\beta}_1 = 2.5$ (b) RSS $= 1.5$

A4 힌트 $\mathbf{q}_1 = \mathbf{1}/2$, $r_{12} = \mathbf{q}_1'\mathbf{x}$이다. $\mathbf{x} - r_{12}\mathbf{q}_1$은 중심화한 $\mathbf{x}$이다.

답 (a) $r_{22} = \sqrt{5}$ (b) $\hat{\beta}_1 = 0.8$ (c) $\hat{\beta}_0 = 1.3$

A5 힌트 $E\,\mathbf{u}'A\mathbf{u} = \operatorname{tr}(A\Sigma) + \boldsymbol{\mu}'A\boldsymbol{\mu}$.

답 23

A6 힌트 $\mathrm{RSS}/\sigma^2 \sim \chi^2(n-p)$이고 $\chi^2(k)$의 분산은 $2k$이다.

답 (a) 32 (b) 128

A7 힌트 $X'X = \left(\begin{smallmatrix}n & \sum x_i\\ \sum x_i & \sum x_i^2\end{smallmatrix}\right) = \left(\begin{smallmatrix}6 & 15\\ 15 & 55\end{smallmatrix}\right)$.

답 (a) $\frac{2}{35}$ (b) $\frac{11}{21}$ (c) $-\frac{1}{7}$

A8 힌트 $\tilde{b}_1 = \mathbf{a}'\mathbf{y}$, $\mathbf{a} = (-1,0,0,0,1)/4$이다. 불편이려면 $\mathbf{a}'\mathbf{1}=0$, $\mathbf{a}'\mathbf{x}=1$이어야 한다.

답 $\frac{5}{4}$

A9 힌트 $\mathrm{SE}(\hat{\beta}_j) = s\sqrt{[(X'X)^{-1}]_{jj}}$이고 자유도는 $n-p$이다.

답 (a) 0.6 (b) 3 (c) ≈ 1.252

A10 힌트 $F = \dfrac{(\mathrm{RSS}_0 - \mathrm{RSS})/q}{\mathrm{RSS}/(n-p)}$이고 자유도는 $(q,\ n-p)$이다.

답 (a) $F = 3.125$ (b) ≈ 0.061

A11 힌트 $F = \dfrac{R^2/(p-1)}{(1-R^2)/(n-p)}$ (예제 4.10(ii)).

답 10

A12 힌트 표준화한 변수에서 $R_1^2 = \mathbf{r}' R_{22}^{-1}\mathbf{r}$이다. $\mathbf{r}$은 x_1과 나머지의 상관, R_{22}는 나머지끼리의 상관행렬이다.

답 (a) $\dfrac{100}{19} \approx 5.26$ (b) $\dfrac{3}{2}$

A13 힌트 빼고 맞춘 잔차는 $e_i/(1-h_{ii})$이다. A1의 잔차와 A2의 지렛대를 짝지어 나눈다.

답 (a) -0.5 (b) $\dfrac{5709}{784} \approx 7.282$

A14 힌트 $\hat{\boldsymbol{\beta}}_\lambda = (X'X + \lambda I)^{-1}X'\mathbf{y}$. 줄어드는 비율은 $d_j^2/(d_j^2+\lambda)$이다.

답 (a) $\dfrac{8}{5}$ (b) $\dfrac{3}{2}$

B1 힌트 ②의 근거에는 A_0'가 무엇인지가 들어간다. $(X'X)^{-1}$이 대칭이라는 것도 쓴다.

B2 힌트 ⑤의 근거는 A_j가 대칭 멱등이면 $\mathbf{v}'A_j\mathbf{v} = \|A_j\mathbf{v}\|^2$이라는 것과 $\sum_j A_j = I$이다.

B3 힌트 ②의 둘째 식은 "$\mathbf{e}$가 X_1의 열공간과 직교하면 $H_1\mathbf{e} = \mathbf{0}$"에서 나온다.

B4 힌트 B1에서 채운 근거를 다시 보지 말고 쓴다. 넷째 단계에서는 아무 $\mathbf{c}$에 대해 $\mathrm{Var}(\mathbf{c}'\tilde{\boldsymbol{\beta}}) - \mathrm{Var}(\mathbf{c}'\hat{\boldsymbol{\beta}})$을 쓴다.

B5 힌트 셋째 단계: 대칭행렬의 계수는 0이 아닌 고유값의 수(중복 포함)다. 넷째 단계: $\mathbf{v} \in \mathrm{col}(A_i)$를 잡고 $\|\mathbf{v}\|^2 = \mathbf{v}'I\mathbf{v} = \sum_j \mathbf{v}'A_j\mathbf{v}$를 쓴다.

B6 힌트 따름정리 3.4(b)는 B4의 넷째 단계에서 등호가 성립하는 경우다. $\|D'\mathbf{c}\| = 0$이 무엇을 뜻하는지 쓴다.

B7 힌트 $X_1 = \mathbf{1}$이면 $H_1 = \mathbf{1}\mathbf{1}'/n$이고 $M_1\mathbf{x} = \mathbf{x} - \bar{x}\mathbf{1}$이다.

C1 힌트 (b) Cauchy–Schwarz $\left(\sum w_i x_i\right)^2 \le \sum w_i^2 \sum x_i^2$. 등호 조건은 w와 x가 평행할 때다.

C2 힌트 (a) Q의 각 열은 $\mathrm{col}(P)$에 있고, P는 $\mathrm{col}(P)$의 원소를 그대로 둔다. $QP = (PQ)'$. (b) 계수는 대각합이다(보조정리 2.4(a)).

C3 힌트 $\hat{b} = \sum x_i y_i / \sum x_i^2$ (C1). 잔차가 직교하는 것은 X의 열, 곧 $\mathbf{x}$뿐이다.

답 (a) $\dfrac{12}{7}$ (b) $-\dfrac{17}{7}$

C4 힌트 C2로 $H - H_0$은 계수 $p - p_0$의 사영이다. $\boldsymbol{\mu} \in \mathrm{col}(X_0)$이면 $(H-H_0)\boldsymbol{\mu} = \mathbf{0}$, $(I-H)\boldsymbol{\mu} = \mathbf{0}$

이다.

C5 힌트 (c) $n^2 = \left(\sum \sqrt{w_i}\cdot \frac{1}{\sqrt{w_i}}\right)^2 \le \sum w_i \sum \frac{1}{w_i}$.

C6 힌트 (a) $E\mathbf{x}u = \mathbf{0}$의 첫 성분이 $Eu = 0$이다. (c) 정리 5.1(c)의 $\beta_1^* = \operatorname{Cov}(X,Y)/\operatorname{Var}X$를 쓴다.

C7 힌트 (a) 명제 2.7(a)의 증명에서 $h_{ii} = h_{ii}^2 + \sum_{j\neq i} h_{ij}^2$. (b) $H\mathbf{u}_i$는 H의 i번째 열이다. (c) (b)의 $\mathbf{v}$에 $X_{(i)}$를 곱한다.

C8 힌트 (c) 두 공분산이 모두 $V(\cdots)V'$ 꼴이고 가운데가 대각행렬이다. $\frac{1}{d^2} - \frac{d^2}{(d^2+\lambda)^2} \ge 0$을 확인한다.

C9 힌트 (b) $|Q_1| = |Z_1^2 - Z_2^2| \le Z_1^2 + Z_2^2 = Q_2$. 독립이면 교집합의 확률은 곱이다.

D1 힌트 (b) $\hat b_1 = \sum c_i y_i$, $c_i = (x_i - \bar x)/S_{xx}$이고 $\sum c_i = 0$이다. 결합정규인 두 선형결합은 공분산이 0이면 독립이다. (e) 반폭은 $t\,s\sqrt{1 + \frac{1}{n} + (x_0-\bar x)^2/S_{xx}}$이다.

답 (d) $\hat y_0 = 5.8$ (e) ≈ 4.125

D2 힌트 (a) C1의 공식 $\hat b = \sum x_i y_i / \sum x_i^2$. (d) $C = (1,0)$이면 $W = [(X'X)^{-1}]_{11} = \frac{1}{n} + \bar x^2/S_{xx}$이다.

답 (a) $\mathrm{RSS}_0 = 6.8$ (b) $F = 5.5$ (c) ≈ 0.101

D3 힌트 선형불편의 조건은 $\sum w_i x_i = 1$이다(C1). (e) 첫 부등식은 정리 3.3. 둘째는 $\frac{1}{n}\sum \frac{1}{x_i^2} \ge \left(\frac{1}{n}\sum \frac{1}{x_i}\right)^2$과 산술–조화 평균 부등식 $\frac{1}{n}\sum \frac{1}{x_i} \ge n/\sum x_i$를 잇는다.

답 (a) $\frac{\sigma^2}{385}$ (b) $\frac{2\sigma^2}{605}$ (c) $\approx 0.01550\,\sigma^2$ (d) $\frac{14}{11}$

D4 힌트 $X'X = \operatorname{diag}(n_1, n_2)$이므로 $\hat{\boldsymbol\beta} = (\bar y_1, \bar y_2)$이고 $W = C(X'X)^{-1}C' = \frac{1}{n_1} + \frac{1}{n_2}$이다.

답 (c) $F = \frac{240}{49} \approx 4.90$ (d) ≈ 0.078