수리통계학 × 통계학의 역사

6회 워크북

Student의 t 손으로 푸는 34문항

강의노트 6회를 읽은 뒤 손으로 푼다. Helmert 회전, 변수변환, 축량은 눈으로 따라간 것과 빈 종이에 다시 쓰는 것이 다르다. 문제 아래의 줄 친 공간에 직접 쓴다. 분위수가 필요한 문항은 문제에 그 값을 적어 두었다.

부	하는 일	문항	예상 시간
A 계산 드릴	표본분산, χ^2, t, F, 구간을 손으로 계산한다	14	1시간 52분
B 빈칸 증명	노트의 단계형 증명을 책을 덮고 다시 쓴다	7	2시간 50분
C 증명 문제	변형, 일반화, 반례 만들기	9	2시간 50분
D 퀄형	여러 절을 잇는 종합 문제. 제한 시간 안에	4	2시간 20분
	합계	34	약 10시간

노트 `06_student_t_notes.pdf` — 문항의 "정리 2.2" 같은 번호는 노트의 번호다

도구 `00_toolkit_notes.pdf` — [T2] 귀류·반례, [T4] 부등식, [T6] 극한의 교환, [T7] 변수변환, [T8] O 표기

검산 노트북 `06_student_t.ipynb`의 "워크북 검산" 절: `check("A3", "5*exp(-4)")`

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. 분위수표의 값으로 계산한 답은 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.

예: 빈칸과 풀이 공간

① $Z \perp V$이면 $Z^2 \perp V$이다. 왜? _____

A

계산 드릴

표본분산, χ^2, t, F와 구간을 손으로 계산한다. 분위수가 필요한 문항은 문제에 값을 적어 두었다. 답이 수나 식으로 떨어지므로 다 풀면 뒤의 최종답이나 노트북의 `check()`로 맞춰 본다.

A1 표본분산을 세 가지로 ★

정리 1.1, 연습 1.1 · 6분

자료 3, 7, 5, 9, 6의 표본분산 s^2를 (i) 편차제곱합, (ii) $\sum x_i^2 - n\bar{x}^2$, (iii) $\frac{1}{2n}\sum_{i,j}(x_i - x_j)^2$의 세 가지로 계산해 같은 값이 나오는지 확인하라.

A2 $n = 4$의 Helmert 회전 ★

보조정리 2.1, 정리 2.2 · 8분

$n = 4$의 Helmert 행렬 H의 네 번째 행을 쓰고, $\mathbf{z} = (1, 2, 3, 6)'$에 대해 $\mathbf{Y} = H\mathbf{z}$의 성분 $Y_1, \dots, Y_4$를 구하라. (a) Y_4는 얼마인가? (b) $Y_2^2 + Y_3^2 + Y_4^2$을 구하고 $\sum_i (z_i - \bar{z})^2$과 같음을 확인하라.

A3 표본분산이 두 배 넘게 나올 확률 ★

정리 2.2(c) · 8분

$X_1, \dots, X_5 \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 4)$이다. $P(S^2 > 8)$을 정확히 구하라. (힌트: $\chi^2(4)$는 모양모수 2인 감마분포다.)

A4 표본분산의 흔들림 ★

따름정리 2.3 · 6분

$X_1, \dots, X_9 \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 2)$이다. (a) $\operatorname{Var} S^2$를 구하라. (b) S^2의 상대 표준편차 $\sqrt{\operatorname{Var} S^2}/ES^2$를 구하라.

A5 세 관측의 표본표준편차 ★

연습 2.4 · 8분

$n = 3$인 정규 표본에서 $ES = c_4(3)\,\sigma$의 $c_4(3)$을 감마함수로 구하고 소수로 나타내라.

A6 $t(3)$의 봉우리 ★

정리 3.2 · 6분

$t(3)$의 밀도의 0에서의 값 $f_3(0)$을 정확히 구하고, 표준정규의 $\phi(0) = \frac{1}{\sqrt{2\pi}} \approx 0.399$와 비교하라.

A7 자유도 1의 t ★

정리 3.3(c) · 8분

$T \sim t(1)$이다. (a) $P(T > \sqrt{3})$, (b) $P(|T| < 1)$을 정확히 구하라.

A8 t와 F의 적률 ★

정리 3.3(b), 4.2(b)(c) · 6분

(a) $T \sim t(6)$의 분산을 구하라. (b) $F \sim F(3, 8)$의 평균을 구하라. (c) T^2의 분포를 쓰고 (a)와 ET^2이 맞는지 확인하라.

A9 다섯 관측의 t-구간 ★

정리 5.2 · 10분

정규 모집단에서 10.2, 9.6, 10.8, 10.1, 9.8을 얻었다. $t_{4,\,0.025} = 2.776$을 써서 μ의 95% 신뢰구간의 (a) 아래 끝과 (b) 위 끝을 구하라.

A10 대응 t 통계량 ★

정리 3.4, 예제 5.3 · 10분

환자 6명의 처리 전후 차이가 $2,\ 4,\ -1,\ 3,\ 5,\ 1$이다. (a) 대응 t 통계량 $t = \bar{d}/(s_d/\sqrt{n})$을 정확한 수로 구하라. (b) 귀무가설 $\mu_D = 0$ 아래의 자유도는?

A11 합동분산의 두 표본 t ★

정리 5.4 · 10분

X: $5, 7, 6$과 Y: $3, 4, 2, 3$이 등분산 정규 모집단의 독립 표본이다. (a) $\mu_1 = \mu_2$ 아래의 통계량 T를 정확한 수로 구하라. (b) 자유도는?

A12 분산비의 구간 ★

정리 4.3 · 10분

$m = 7$, $n = 9$인 독립 정규 표본에서 $s_X^2 = 6$, $s_Y^2 = 2$를 얻었다. $F_{6,8,\,0.025} = 4.652$, $F_{8,6,\,0.025} = 5.600$을 써서 σ_1^2/σ_2^2의 95% 신뢰구간의 (a) 아래 끝과 (b) 위 끝을 구하라.

A13 다음 한 관측의 구간 ★

정리 8.1 · 8분

A9의 자료로 다음 관측 하나에 대한 95% 예측구간의 (a) 아래 끝과 (b) 위 끝을 구하라. A9의 신뢰구간보다 몇 배 넓은가?

A14 세 관측에 1.96을 쓰면 ★★

정리 3.4, 노트 연습 3.3 · 8분

$n = 3$인 정규 표본에서 $\bar{X} \pm 1.96\,S/\sqrt{3}$의 실제 포함확률을 정확한 식으로 구하라. $t(2)$의 분포함수 $F_2(t) = \frac{1}{2}\left(1 + \frac{t}{\sqrt{2+t^2}}\right)$를 써도 된다.

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 정리 번호, 앞 회차의 정리, 도구 번호를 쓴다.

B1 Student의 정리, 근거 채우기 ★★

① · 정리 2.2 · 15분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$이면 $\bar{X} \perp S^2$, $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$이다. 각 단계의 근거를 써라.

① $Z_i = (X_i - \mu)/\sigma$로 두면 $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, I)$이고 $\mathbf{Z}$에 대해 보이면 된다. 표준화

② $\mathbf{Y} = H\mathbf{Z}$이면 $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{0}, I)$이다. 왜? _____

③ $Y_1, \dots, Y_n$은 독립인 표준정규다. 왜? _____

④ $Y_1 = \sqrt{n}\,\bar{Z}$이다. 왜? _____

⑤ $\sum_{i \geq 2} Y_i^2 = \sum_i (Z_i - \bar{Z})^2$이다. 왜? _____

⑥ $\bar{Z} \perp \sum_i (Z_i - \bar{Z})^2$이다. 왜? _____

⑦ $\bar{Z} \sim \mathcal{N}(0, \frac{1}{n})$이다. 왜? _____

⑧ $\sum_i (Z_i - \bar{Z})^2 \sim \chi^2(n-1)$이다. 왜? _____

B2 t의 밀도, 근거 채우기 ★★

① · 정리 3.2 · [T7] · 20분

$Z \sim \mathcal{N}(0,1)$과 $V \sim \chi^2(r)$이 독립이면 $T = Z/\sqrt{V/r}$의 밀도는 다음과 같다. 근거를 채워라.

$$f_r(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\,\Gamma\left(\frac{r}{2}\right)}\left(1+\frac{t^2}{r}\right)^{-\frac{r+1}{2}}$$

① (Z, V)의 결합밀도는 $\phi(z) \cdot \frac{v^{r/2-1}e^{-v/2}}{\Gamma(r/2)2^{r/2}}$이다. 왜? _____

② 보조 변수 $W = V$를 두면 역변환은 $z = t\sqrt{w/r}$, $v = w$이다.

왜? _____

③ 야코비안은 $\sqrt{w/r}$이다. 왜? _____

④ $f_{T,W}(t,w) = \phi\left(t\sqrt{w/r}\right) \cdot \frac{w^{r/2-1}e^{-w/2}}{\Gamma(r/2)2^{r/2}} \cdot \sqrt{w/r}$이다. 왜? _____

⑤ $f_T(t) = \int_0^\infty f_{T,W}(t,w)\,dw$이다. 왜? _____

⑥ $\int_0^\infty w^{\frac{r+1}{2}-1}e^{-w\frac{1+t^2/r}{2}}dw = \Gamma\left(\frac{r+1}{2}\right)\left(\frac{2}{1+t^2/r}\right)^{\frac{r+1}{2}}$이다. 왜? _____

B3 두 표본 t, 근거 채우기 ★★

① · 정리 5.4 · 15분

등분산 정규 표본 $X_1, \dots, X_m$과 $Y_1, \dots, Y_n$이 독립이면 $T = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_p\sqrt{1/m+1/n}} \sim t(m+n-2)$이다. 근거를 채워라.

① $Z = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{1/m+1/n}} \sim \mathcal{N}(0,1)$이다. 왜? _____

② $\frac{(m-1)S_X^2}{\sigma^2}$와 $\frac{(n-1)S_Y^2}{\sigma^2}$는 독립인 $\chi^2(m-1)$, $\chi^2(n-1)$이다. 왜? _____

③ $V = \frac{(m+n-2)S_p^2}{\sigma^2} \sim \chi^2(m+n-2)$이다. 왜? _____

④ $Z \perp V$이다. 왜? _____

⑤ $T = Z/\sqrt{V/(m+n-2)} \sim t(m+n-2)$이다. 왜? _____

B4 Student의 정리, 단계의 이름만 ★★

② · 정리 2.2 · 30분

정리 2.2의 (a)(b)(c)를 증명하라. 단계의 이름만 주어진다. 각 단계 아래에 식과 근거를 써라.

① **표준화해 $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, I)$로 줄인다**

② **Helmert 행렬로 돌린다**

③ **돌린 좌표가 독립인 표준정규임을 보인다**

④ **첫 성분이 평균임을 보인다**

⑤ **나머지 성분의 제곱합이 편차제곱합임을 보인다**

⑥ **(b): 독립을 결론짓는다**

⑦ **(a)와 (c)를 결론짓는다**

B5 t의 밀도, 단계의 이름만 ★★

② · 정리 3.2 · [T7] · 25분

정리 3.2를 증명하라. 단계의 이름만 주어진다.

① **결합밀도를 쓴다**

② **보조 변수를 두고 역변환을 구한다**

③ **야코비안을 구한다**

④ **(T, W)의 결합밀도를 쓴다**

⑤ **보조 변수를 적분한다(감마 적분)**

⑥ **상수를 모은다**

B6 Student의 정리, 전략만 ★★

③ · 정리 2.2 · 35분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$이면 $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, $\bar{X} \perp S^2$, $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$임을 증명하라.

전략 표본을 표준화하고 직교행렬로 돌린다. 돌린 좌표에서 첫 성분은 평균이고, 나머지 성분의 제곱합은 편차제곱합이며, 좌표들은 서로 독립인 표준정규다.

B7 두 표본 t, 전략만 ★★ ③ · 정리 5.4 · 30분

정리 5.4를 증명하라.

전략 표준정규를, 그와 독립인 χ^2를 자유도로 나눈 것의 제곱근으로 나눈 꼴을 만든다. 두 표본에 Student의 정리를 따로 쓰고 독립을 합친다.

C

증명 문제

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

C1 Helmert가 아니어도 된다 ★★

보조정리 2.1, 정리 2.2 · 20분

H가 첫 행이 $\mathbf{1}'/\sqrt{n}$인 임의의 $n \times n$ 직교행렬이고 $\mathbf{Y} = H\mathbf{Z}$이다.

(a) H의 둘째 행부터 마지막 행까지는 모두 $\mathbf{1}$과 직교함을 보여라.

(b) $\sum_{i \geq 2} Y_i^2 = \sum_i (Z_i - \bar{Z})^2$임을 보여라.

(c) 정리 2.2의 증명이 이 H로도 그대로 됨을 밝혀라.

(d) $n = 3$에서 Helmert 행렬과 다른 예를 하나 만들어라.

C2 μ를 알 때의 잘못된 t ★★

[T2], [T4] · 정리 2.2, 3.4 · 20분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$이고 μ를 안다. σ^2을 $\hat{\sigma}_0^2 = \frac{1}{n}\sum_i (X_i - \mu)^2$로 어림해 $W = \frac{\sqrt{n}(\bar{X}-\mu)}{\hat{\sigma}_0}$를 만들었다.

(a) Cauchy–Schwarz로 $|W| \le \sqrt{n}$임을 보이고, W가 $t(n)$을 따를 수 없음을 보여라.

(b) 정리 3.4의 증명의 어느 단계가 W에서는 성립하지 않는가?

(c) $\frac{W^2}{n} \sim \text{Beta}\left(\frac{1}{2}, \frac{n-1}{2}\right)$임을 보여라.

C3 분산의 비를 알 때의 두 표본 t ★★

정리 5.4 · 15분

$X_1, \dots, X_m \overset{\text{iid}}{\sim} \mathcal{N}(\mu_1, \sigma^2)$, $Y_1, \dots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu_2, \lambda\sigma^2)$가 독립이고 $\lambda > 0$은 안다. $t(m+n-2)$를 따르는 $\mu_1 - \mu_2$의 축량을 만들고, 95% 신뢰구간을 써라. $\lambda = 1$이면 정리 5.4가 됨을 확인하라.

C4 t-구간이 z-구간보다 짧을 확률 ★★

정리 2.2(c) · 20분

$n=4$인 정규 표본에서 t-구간 $\bar{X}\pm t_{3,0.025}S/2$와 ($\sigma$를 알 때의) z-구간 $\bar{X}\pm 1.96\,\sigma/2$를 비교한다. $t_{3,0.025}=3.182$이다.

(a) t-구간이 더 짧은 사건이 $\{\chi^2(3) < c\}$ 꼴임을 보이고 c를 구하라.

(b) $\chi^2(3)$의 분포함수가 $2\Phi(\sqrt{x})-1-\sqrt{2x/\pi}\,e^{-x/2}$ 임을 보이고, (a)의 확률을 구하라.

C5 무상관이지만 독립이 아니다 ★★

[T2] · 명제 7.1 · 15분

$X_1, X_2 \overset{\text{iid}}{\sim} \mathrm{U}(-1,1)$이고 $\bar{X}=\frac{X_1+X_2}{2}$, $S^2=\frac{(X_1-X_2)^2}{2}$이다.

(a) 명제 7.1로 $\mathrm{Cov}(\bar{X}, S^2)=0$임을 보여라.

(b) $P(|\bar{X}|>\tfrac{1}{2})$와 $P(S^2>\tfrac{1}{2})$를 구하라.

(c) $P(|\bar{X}|>\tfrac{1}{2},\ S^2>\tfrac{1}{2})=0$임을 보이고, $\bar{X}$와 S^2가 독립이 아님을 결론지어라.

C6 k개 집단의 합동분산 ★★★

정리 2.2, 5.4 · 25분

k개의 독립 집단에서 $Y_{j1}, \dots, Y_{jn_j} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_j, \sigma^2)$ $(j = 1, \dots, k)$이고 $N = \sum_j n_j$, $S_p^2 = \frac{1}{N-k}\sum_j (n_j - 1)S_j^2$이다.

(a) $\frac{(N-k)S_p^2}{\sigma^2} \sim \chi^2(N-k)$이고 S_p^2가 $(\bar{Y}_1, \dots, \bar{Y}_k)$와 독립임을 보여라.

(b) 상수 $c_1, \dots, c_k$에 대해 $\theta = \sum_j c_j \mu_j$, $\hat{\theta} = \sum_j c_j \bar{Y}_j$이면 $T = \frac{\hat{\theta} - \theta}{S_p \sqrt{\sum_j c_j^2 / n_j}} \sim t(N-k)$임을 보여라.

(c) $k = 2$, $c = (1, -1)$이면 정리 5.4가 됨을 확인하라.

C7 단위를 바꾸어도 같은 구간 ★

정리 5.2 · 10분

$Y_i = a + bX_i$ $(b \neq 0)$이다. $\bar{Y} = a + b\bar{X}$, $S_Y = |b| S_X$ 임을 보이고, μ_Y 의 t-구간이 μ_X 의 t-구간을 같은 아핀 변환으로 옮긴 것임을 보여라($b < 0$이면 양 끝이 바뀐다). 섭씨를 화씨로 바꿔도 결론이 같다는 뜻이다.

C8 t의 분위수는 정규의 분위수로 간다 ★★★

[T6] · 정리 3.3(d) · 30분

(a) $r \geq 2$이면 $\left(1 + \frac{t^2}{r}\right)^r \geq \left(1 + \frac{t^2}{2}\right)^2$ 임을 이항전개의 처음 세 항으로 보여라. 이것으로 $r \geq 2$에서 $\left(1 + \frac{t^2}{r}\right)^{-(r+1)/2} \leq \left(1 + \frac{t^2}{2}\right)^{-1}$ 을 얻어라.

(b) 정리 3.3(d)와 지배수렴 정리로, 모든 x에서 $F_r(x) = \int_{-\infty}^{x} f_r \to \Phi(x)$ 임을 보여라. 상수 c_r이 유계인 것도 확인하라.

(c) (b)로 $t_{r,p} \to z_p$ $(r \to \infty)$를 보여라.

C9 지수분포의 축량 ★★

정의 5.1 · 15분

$X_1, \dots, X_n \overset{\text{iid}}{\sim}$ 지수분포(평균 θ)이다.

(a) $Q = \frac{2}{\theta}\sum_i X_i \sim \chi^2(2n)$임을 보이고, Q가 축량임을 밝혀라.

(b) θ의 95% 신뢰구간을 쓰고 포함확률이 정확히 0.95임을 보여라.

(c) $n = 5$, $\sum x_i = 10$이고 $\chi^2_{10,0.025} = 20.483$, $\chi^2_{10,0.975} = 3.247$이면 구간의 아래 끝은?

(d) 같은 자료에서 위 끝은?

D

퀄형

여러 절과 앞 회차의 도구를 잇는 종합 문제다. 머리의 제한 시간을 재고, 노트를 덮고 푼다. 분포를 유도할 때는 어느 정리를 쓰는지 밝히는 것까지 답안이다.

D1 σ의 불편추정량과 평균제곱오차 ★★★

정리 2.2(c), 연습 2.4 · 제한 35분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$이고 $c_4(n) = \sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}$로 $ES = c_4(n)\sigma$이다(노트 연습 2.4).

(a) $k_n S$가 σ의 불편추정량이 되는 k_n을 쓰고, $n = 5$에서 정확한 값을 구하라.

(b) $n = 5$에서 $E(k_5 S - \sigma)^2/\sigma^2$을 구하라.

(c) $n = 5$에서 $E(S - \sigma)^2/\sigma^2$을 구하고 (b)와 비교하라.

(d) $E(aS - \sigma)^2$을 가장 작게 하는 a를 $c_4(n)$으로 쓰고, $n = 5$의 값을 구하라.

(e) (d)의 최솟값 $E(a^* S - \sigma)^2/\sigma^2$을 $n = 5$에서 구하라.

D2 대응을 무시하면 ★★★

정리 5.4, 예제 5.3, 3회 정리 4.2 · 제한 40분

(X_i, Y_i), $i = 1, \dots, n$은 독립인 이변량 정규 쌍이고, $\mathrm{Var}\, X_i = \mathrm{Var}\, Y_i = \sigma^2$, $\mathrm{Corr}(X_i, Y_i) = \rho > 0$, $\Delta = EX_i - EY_i$이다.

(a) $D_i = X_i - Y_i \overset{\text{iid}}{\sim} \mathcal{N}\big(\Delta,\ 2\sigma^2(1-\rho)\big)$임을 보이고, 대응 t 통계량이 $\Delta = 0$에서 $t(n-1)$을 따름을 보여라.

(b) $\mathrm{Var}(\bar X - \bar Y) = \dfrac{2\sigma^2(1-\rho)}{n}$임을 보여라. 대응을 무시하고 두 표본을 독립으로 보면 이 분산을 $\dfrac{2\sigma^2}{n}$으로 잘못 어림한다.

(c) σ와 ρ를 안다고 하자. $n = 10$, $\rho = 0.8$, $\Delta = 0.5\sigma$일 때 올바른 z-검정 $|\bar D| / \sqrt{2\sigma^2(1-\rho)/n} > 1.96$의 검정력을 구하라.

(d) 같은 조건에서 대응을 무시한 검정 $|\bar X - \bar Y| / \sqrt{2\sigma^2/n} > 1.96$의 검정력을 구하라.

(e) (d)의 검정의 실제 유의수준($\Delta = 0$에서 기각할 확률)을 구하라.

D3 균등분포의 로그와 χ^2, F ★★★

2회 정리 4.3, 4회 정리 3.7, 정의 4.1 · 제한 35분

$U_1, U_2, \dots$ 와 $V_1, V_2, \dots$ 가 모두 독립이고 $\mathrm{U}(0,1)$을 따른다.

(a) $-2\log U_1 \sim \chi^2(2)$임을 보여라.

(b) $-2\sum_{i=1}^{k} \log U_i \sim \chi^2(2k)$임을 보여라.

(c) $\dfrac{\frac{1}{m}\sum_{i=1}^{m} \log U_i}{\frac{1}{n}\sum_{j=1}^{n} \log V_j} \sim F(2m, 2n)$임을 보여라.

(d) $F(2,2)$의 분포함수가 $\frac{x}{1+x}$ $(x>0)$임을 보이고, $P\big(\log U_1 / \log V_1 > 3\big)$을 구하라.

(e) 독립인 두 검정의 유의확률이 $p_1 = 0.04$, $p_2 = 0.10$이다. 귀무가설 아래에서 유의확률은 $\mathrm{U}(0,1)$을 따른다. 결합 통계량 $X = -2(\log p_1 + \log p_2)$의 유의확률 $P(\chi^2(4) \ge X)$를 $q = p_1 p_2$로 나타내고 값을 구하라.

D4 관측 두 개의 t ★★★

정리 2.2, 3.3(c), 3회 정리 5.8 · 제한 30분

$X_1, X_2 \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$이고 $T = \frac{\bar{X}-\mu}{S/\sqrt{2}}$이다.

(a) $T = \frac{X_1+X_2-2\mu}{|X_1-X_2|}$임을 보여라.

(b) $Y_1 = \frac{Z_1+Z_2}{\sqrt{2}}$, $Y_2 = \frac{Z_1-Z_2}{\sqrt{2}}$ ($Z_i = (X_i-\mu)/\sigma$)로 $T = Y_1/|Y_2|$임을 보여라. 정리 3.3(c)를 쓰지 말고 3회 정리 5.8로 T가 Cauchy를 따름을 보여라.

(c) $t_{1,\,0.025}$를 정확한 식으로 쓰고 값을 구하라.

(d) $\bar{X} \pm 1.96\,S/\sqrt{2}$의 실제 포함확률을 정확한 식으로 쓰고 값을 구하라.

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 (iii)에서 $i<j$인 쌍 10개의 차의 제곱을 더하면 $\frac{1}{2}\sum_{i,j}$이다.

답 $s^2 = 5$

A2 힌트 k번째 행은 $\frac{1}{\sqrt{k(k-1)}}(1,\dots,1,-(k-1),0,\dots,0)$이다. $Y_1 = \sqrt{n}\,\bar{z}$.

답 (a) $-2\sqrt{3} \approx -3.464$ (b) 14

A3 힌트 $\frac{4S^2}{4} = S^2 \sim \chi^2(4)$이다. $\chi^2(4)$의 밀도 $\frac{1}{4}xe^{-x/2}$를 부분적분한다.

답 $5e^{-4} \approx 0.0916$

A4 힌트 따름정리 2.3의 $\operatorname{Var} S^2 = \frac{2\sigma^4}{n-1}$.

답 (a) 1 (b) $\frac{1}{2}$

A5 힌트 $c_4(n) = \sqrt{\frac{2}{n-1}}\,\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}$, $\Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$.

답 $\frac{\sqrt{\pi}}{2} \approx 0.886$

A6 힌트 $f_r(0) = \frac{\Gamma((r+1)/2)}{\sqrt{r\pi}\,\Gamma(r/2)}$이고 $\Gamma(2) = 1$.

답 $\frac{2}{\sqrt{3}\,\pi} \approx 0.368$

A7 힌트 $t(1)$은 Cauchy이고 분포함수는 $\frac{1}{2} + \frac{1}{\pi}\arctan t$이다.

답 (a) $\frac{1}{6}$ (b) $\frac{1}{2}$

A8 힌트 $\operatorname{Var} T = \frac{r}{r-2}$, $EF = \frac{r_2}{r_2-2}$, $T^2 \sim F(1,r)$.

답 (a) $\frac{3}{2}$ (b) $\frac{4}{3}$

A9 힌트 $\bar{x} = 10.1$이다. 편차제곱합을 먼저 구한다.

답 (a) 9.531 (b) 10.669

A10 힌트 $\bar{d} = \frac{7}{3}$, $\sum d_i^2 = 56$. 분수로 계산하면 근호 하나가 남는다.

답 (a) $\sqrt{7} \approx 2.646$ (b) 5

A11 힌트 두 편차제곱합이 모두 2다. $S_p^2 = \frac{2+2}{5}$.

답 (a) $3\sqrt{15/7} \approx 4.392$ (b) 5

A12 힌트 $F_{6,8,0.975} = 1/F_{8,6,0.025}$ (노트 연습 4.2).

답 (a) 0.645 (b) 16.80

A13 힌트 반폭이 $t_{4,0.025}\, s\sqrt{1+1/5}$로 바뀐다.

답 (a) 8.706 (b) 11.494

A14 힌트 포함되는 사건은 $|T_2| \le 1.96$이다.

답 $\frac{1.96}{\sqrt{2+1.96^2}} \approx 0.811$

B1 힌트 근거는 짧게 쓴다. 예: "5회 정리 3.8", "4회 정리 3.9". 독립을 결론짓는 단계는 "어느 변수들의 함수인가"를 쓴다.

B2 힌트 변수변환의 근거는 3회 정리 5.1, 감마 적분의 근거는 감마함수의 정의다. 상수를 모으는 계산도 아래 줄에 해 본다.

B3 힌트 등분산 가정이 쓰이는 단계를 찾아 표시해 둔다. 셋째 단계다.

B4 힌트 B1을 다시 보지 않고 쓴다. 다섯째 단계에는 직교행렬이 길이를 보존한다는 사실과 정리 1.1의 항등식이 필요하다.

B5 힌트 B2와 같은 순서다. 마지막 단계에서 $2^{(r+1)/2}/2^{r/2} = \sqrt{2}$이고 $\sqrt{2\pi}$와 만나 $\sqrt{\pi}$만 남는다.

B6 힌트 단계는 여덟이다. 다 쓴 뒤, 정규성을 쓴 곳이 "직교변환해도 $\mathcal{N}_n(\mathbf{0}, I)$"와 "무상관이면 독립" 두 곳뿐인지 확인한다.

B7 힌트 다섯 단계다. 분자의 분산 $\sigma^2(1/m + 1/n)$, 분모의 χ^2 자유도 $m+n-2$, 그리고 둘의 독립을 차례로 보인다.

C1 힌트 (a) 직교행렬의 행들은 정규직교다. (d) Helmert의 둘째, 셋째 행을 평면 안에서 돌린다.

C2 힌트 (c) $\sum_i (X_i-\mu)^2/\sigma^2 = Y_1^2 + \sum_{i\ge 2} Y_i^2$, $Y_1 = \sqrt{n}(\bar{X}-\mu)/\sigma$. 4회 정리 4.2.

C3 힌트 Y를 $\sqrt{\lambda}$로 나눈 표본의 분산이 σ^2이다. 분모는 $\frac{(m-1)S_X^2}{\sigma^2} + \frac{(n-1)S_Y^2}{\lambda\sigma^2}$로 만든다.

C4 힌트 (b) $\chi^2(3)$은 $Z_1^2+Z_2^2+Z_3^2$이다. 밀도 $\frac{\sqrt{x}e^{-x/2}}{\sqrt{2\pi}}$를 부분적분하거나, 오른쪽을 미분해 밀도와

비교한다.

답 (a) $c \approx 1.138$ (b) 0.232

C5 힌트 (b) 정사각형 $[-1,1]^2$에서 $|x_1+x_2|>1$인 부분과 $|x_1-x_2|>1$인 부분은 각각 모서리의 삼각형 두 개다. (c) 삼각부등식.

답 (b) 둘 다 $\frac{1}{4}$

C6 힌트 집단마다 정리 2.2를 쓰고, 집단들이 독립이라는 것으로 $2k$개의 통계량 $\bar{Y}_j, S_j^2$이 상호독립임을 먼저 보인다.

C7 힌트 편차 $Y_i - \bar{Y} = b(X_i - \bar{X})$이다.

C8 힌트 (a) $u=t^2$, $(1+u/r)^r \ge 1+u+\binom{r}{2}\frac{u^2}{r^2}$이고 $r\ge 2$이면 $\frac{r-1}{2r}\ge\frac{1}{4}$이다. (c) $\varepsilon>0$을 잡고 $F_r(z_p\pm\varepsilon)$의 극한과 $1-p$를 비교한다.

C9 힌트 $\frac{2}{\theta}X_i$는 평균 2인 지수, 곧 $\Gamma(1,2)=\chi^2(2)$이다. 독립인 감마의 합(4회 정리 3.7).

답 (c) 0.976 (d) 6.160

D1 힌트 $ES^2=\sigma^2$이므로 $E(aS-\sigma)^2=\sigma^2(a^2-2ac_4+1)$이다. $c_4(5)=\sqrt{\frac{1}{2}}\,\frac{\Gamma(5/2)}{\Gamma(2)}$, $\Gamma(\frac{5}{2})=\frac{3\sqrt{\pi}}{4}$.

답 (a) $k_5=\frac{4\sqrt{2}}{3\sqrt{\pi}}\approx 1.064$ (b) $\frac{32}{9\pi}-1\approx 0.1318$ (c) $2-\frac{3\sqrt{\pi}}{2\sqrt{2}}\approx 0.1200$ (d) $a^*=c_4(5)\approx 0.940$ (e) $1-\frac{9\pi}{32}\approx 0.1164$

D2 힌트 (c) 통계량은 $\mathcal{N}(\delta,1)$, $\delta=\Delta/\sqrt{2\sigma^2(1-\rho)/n}=2.5$이다. (d) 통계량은 $\mathcal{N}\!\left(\frac{0.5}{\sqrt{0.2}},\,1-\rho\right)$이다. 표준정규표를 쓴다.

답 (c) ≈ 0.705 (d) ≈ 0.030 (e) $\approx 1.2\times10^{-5}$

D3 힌트 (a) 단조변환. $P(-2\log U\le x)=P(U\ge e^{-x/2})$. (d) 정리 4.2(a)의 밀도에 $r_1=r_2=2$를 넣는다. (e) $\chi^2(4)$의 꼬리 $e^{-x/2}(1+\frac{x}{2})$에서 $e^{-X/2}=q$이다.

답 (d) $\frac{1}{4}$ (e) $q(1-\log q)\approx 0.0261$

D4 힌트 (b) Y_1, Y_2는 독립인 표준정규이고 $Y_1/|Y_2|$와 Y_1/Y_2는 분포가 같다(Y_1의 대칭과 독립). (c) Cauchy의 분포함수 $\frac{1}{2}+\frac{1}{\pi}\arctan t$.

답 (c) $\tan(0.475\pi)\approx 12.71$ (d) $\frac{2}{\pi}\arctan 1.96\approx 0.700$