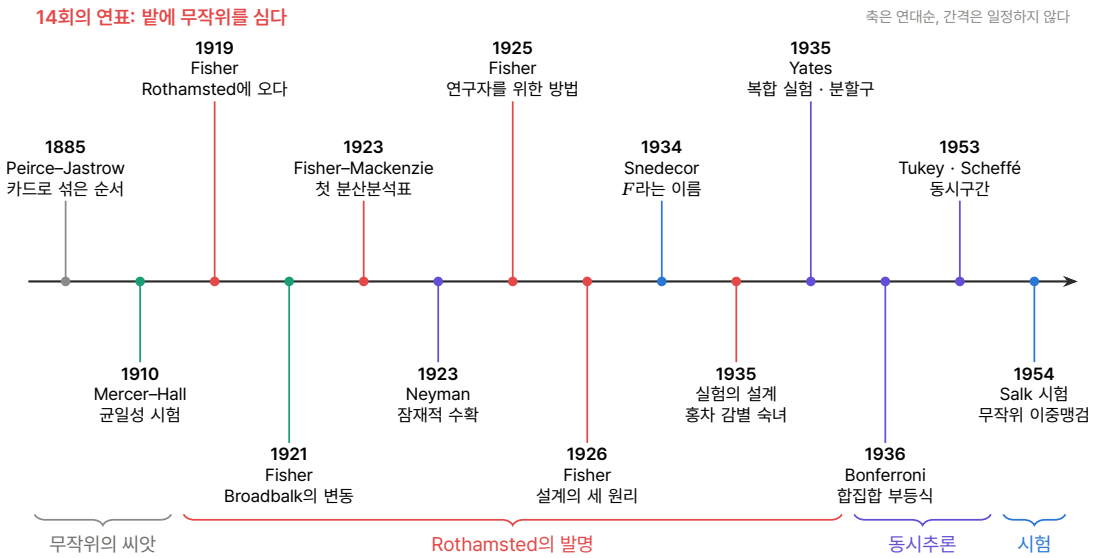
수리통계학 × 통계학의 역사

14회 강의노트

분산분석과 실험설계
밭에 무작위를 심다

1885 → 1954

비료가 얼마나 듣는지 알려면 밭의 차이와 날씨와 우연을 떼어 내야 한다. Fisher는 Rothamsted에서 그 일을 두 가지 발명으로 해냈다. 처치를 무작위로 배치하는 것, 그리고 변동을 직교하는 조각으로 나누는 분산분석이다. 이 노트는 잠재적 결과와 무작위화 검정에서 출발해, 13회의 사영으로 분산분석을 다시 세우고, 비중심 분포로 검정력과 표본 크기를, 동시추론으로 여러 비교를 다룬 뒤, 1954년 Salk 백신 시험에서 무작위화의 값을 숫자로 본다.



다루는 것	잠재적 결과와 Neyman의 분산 · Fisher의 무작위화 검정 · 일원분산분석 = 사영 · 비중심 χ^2와 F, 검정력과 표본 크기 · 블록과 이원분산분석, 분할구 · Bonferroni · Tukey · Scheffé · 국소 관리와 요인설계 · 교란과 의사반복 · CUPED
원자료	Mercer–Hall의 균일성 시험(1910) · Broadbalk 밀(1852–1925) · Yates의 귀리 분할구(1931) · npk 요인실험 · Darwin의 옥수수(1876) · Salk 백신 시험(1954)
노트북	`sessions/14_anova_design/14_anova_design.ipynb` — 절·정리 번호가 같다
워크북	`14_anova_design_workbook.pdf` (손으로 푸는 34문항) · 풀이판 `_workbook_sol.pdf`
선수 지식	1–13회 (특히 6회 $t \cdot F$, 8회 χ^2 검정, 12회 검정력, 13회 사영 · Cochran · 일반선형가설)

이 노트를 쓰는 법

이 노트는 다른 책 없이 읽도록 썼다. 쓰이는 정의와 정리는 모두 이 노트나 앞 회차의 노트에서 증명한다. 이 과정의 범위를 넘어 증명하지 않는 사실은 **증명 없이 쓰는 사실** 상자에 따로 밝힌다.

각 절은 **역사 → 정의·정리 → 증명 → 그림 → 연습문제** 순서다. 핵심 정리의 증명은 번호 붙은 단계 ①②③으로 쓰고, 단계마다 오른쪽 끝에 근거(앞 정리의 번호나 정의)를 단다. 읽을 때는 근거를 가리고 "왜 이 단계가 성립하는가"를 먼저 말해 본다. 확률이 어디서 오는지, 급수와 기댓값의 순서를 바꿔도 되는지는 **조건 확인** 단락에 따로 밝힌다. 나머지 정리는 **증명 전략**만 적고, 증명을 소문항으로 나눈 유도형 연습으로 돌린다. 이 연습은 건너뛰지 않고 손으로 쓴다.

근거에 붙은 [T1]–[T8]은 별책 **증명 도구 상자**의 도구 번호다. 이 회차에서는 [T1] 집합의 포함과 단조성, [T4] Cauchy–Schwarz, [T6] 급수와 기댓값의 교환을 쓴다.

연습문제는 소절 끝에 모여 있다. 난이도는 ★ 계산·확인, ★★ 증명·응용, ★★★ 도전이다. 모든 문제의 풀이가 부록 A에 있고, 문제 끝의 **풀이 ○○쪽**을 누르면 그 쪽으로 간다. 손으로 푸는 문제는 따로 **워크북**(계산 드릴, 빈칸 증명, 증명 문제, 퀄형)에 모았다.

상자의 뜻

> **정의** 대상을 처음 정하는 문장. 왼쪽 선이 보라다.

> **정리** 증명할 주장(정리·보조정리·따름정리·명제). 왼쪽 선이 빨강이다.

> **예제** 숫자를 넣어 끝까지 계산해 보는 곳. 왼쪽 선이 파랑이다.

> **가정 주의** 빼면 정리가 무너지는 가정. 왼쪽 선이 주황이다.

> **증명 없이 쓰는 사실** (예) 이 과정에서 증명하지 않고 쓰는 결과. 무엇이 참인지, 왜 그럴듯한지, 어디서 증명되는지를 밝힌다.

> **역사** 그 개념이 태어난 사람과 문제. 회색 선 옆의 작은 글씨다.

> **연습** 소절 끝의 연습문제. 가는 청록 선이고, 끝에 풀이가 있는 쪽 번호가 붙는다.

색의 뜻 수식과 그림에서 보라는 모수, 청록은 데이터·확률변수, 파랑은 통계량, 빨강은 핵심 결론이다. 노트북과 같은 색이다.

차례

0 들어가며 1

0.1 1–13회에서 가져오는 것 2

0.2 표기 3

1 무작위화 (1885–1935) 4

1.1 잠재적 결과 4

1.2 Fisher의 무작위화 검정 6

2 일원분산분석과 비중심 분포 9

2.1 일원분산분석은 사영이다 9

2.2 비중심 분포와 검정력 10

3 블록과 이원분산분석 15

3.1 균형 이원배치의 직교 분해 15

3.2 분할구 설계: 오차가 두 층이다 18

4 여러 비교를 한꺼번에 21

4.1 가족 단위 오류율과 Bonferroni 21

4.2 Scheffé: 모든 대비를 동시에 21

4.3 Tukey: 짝 비교 전체를 정확히 22

5 설계의 원리 — 반복, 무작위화, 국소 관리 25

5.1 국소 관리: 짝과 블록 25

5.2 요인설계: 한 번에 여러 질문을 26

6 원자료로 — 1954년 Salk 백신 시험 29

7 부숴보기 — 설계가 깨질 때 32

7.1 무작위화를 빼면: 땅심이 효과로 둔갑한다 . . . 32

7.2 오차층을 틀리면: 의사반복 . . . 32

7.3 균형을 깨면: 제곱합이 순서에 따라 달라진다 . . . 32

7.4 분산이 다르면: 작은 집단이 시끄러울 때 . . . 33

8 오늘의 실험설계 — 하이퍼파라미터와 A/B 테스트 34

8.1 Ablation과 그리드: 상호작용을 보는가 . . . 34

8.2 A/B 테스트와 CUPED . . . 34

9 정리와 역사 노트 37

9.1 한 장 요약 . . . 37

9.2 Stigler의 명명 법칙 . . . 38

9.3 7 Pillars 지도 . . . 38

9.4 다음 회 — 가정이 틀리면: 순위·로버스트·부트스트랩 (1935 → 1979) . . . 38

A 연습문제 풀이 40

0

들어가며

> **여는 질문.** 새 비료가 밀 수확을 얼마나 늘리는지 알고 싶다. 그런데 같은 밭 안에서도 땅심이 다르고, 해마다 날씨가 다르며, 같은 처치를 해도 수확은 들쭉날쭉하다. 비료의 효과를 땅과 날씨와 우연에서 어떻게 떼어 내는가? 그리고 그 판단을 무엇이 보증하는가?

1919년 29세의 Ronald Fisher는 런던 북쪽 Rothamsted 농사시험장에 통계 담당으로 왔다. 그곳에는 1843년부터 쌓인 밀 수확 기록이 있었다. 같은 밭(Broadbalk)의 여러 구획에 서로 다른 거름을 70년 넘게 주어 온 기록이다. Fisher의 첫 일은 그 숫자들에서 거름의 효과와 해마다의 날씨를 떼어 내는 것이었다. 그는 곧 문제가 분석에 있지 않고 실험을 설계하는 방식에 있음을 깨달았다. 처치를 사람이 고른 자리에 놓으면, 땅심의 차이가 처치의 효과로 둔갑한다. 무엇으로도 그 둔갑을 분석 단계에서 되돌릴 수 없다.

Fisher의 답은 두 가지였다. 하나는 **무작위화**다. 처치를 놓을 자리를 동전이나 카드로 정하면, 땅심이 처치와 상관될 길이 끊기고, 오차의 크기를 정직하게 잴 수 있다. 다른 하나는 **분산분석**이다. 수확의 전체 변동을 처치, 블록, 오차의 몫으로 나누는 계산인데, 13회의 언어로는 서로 직교하는 사영들의 피타고라스다. 1925년의 **연구자를 위한 통계적 방법**과 1935년의 **실험의 설계**가 이 둘을 세상에 퍼뜨렸다. 그 끝에 1954년 미국에서 180만 명의 아이가 참가한 Salk 소아마비 백신 시험이 있다.

이 노트는 그 순서를 따른다. 1절은 잠재적 결과로 무작위화를 정의하고, Neyman(1923)의 분산 공식과 Fisher의 무작위화 검정을 증명한다. 2절은 일원분산분석을 사영으로 세우고, 비중심 분포로 검정력과 표본 크기를 계산한다. 3절은 블록과 이원분산분석, 분할구 설계를 다룬다. 4절은 여러 비교를 동시에 할 때의 Bonferroni·Tukey·Scheffé, 5절은 Fisher의 설계 원리를 숫자로 확인한다. 6절은 Salk 시험, 7절은 무작위화와 균형이 깨질 때, 8절은 하이퍼파라미터 탐색과 A/B 테스트 속의 실험설계다.

학습 목표

1. 잠재적 결과로 처치효과를 정의하고, 무작위 배정 아래 평균 차이의 불편성과 Neyman의 분산 공식, 무작위화 검정의 정확성을 증명한다.
2. 일원·이원분산분석을 직교사영의 분해로 유도하고, 비중심 χ^2와 F로 검정력과 표본 크기를

계산한다.

3. 블록 설계와 분할구 설계에서 오차를 어느 층에서 재야 하는지 판단한다.
4. Bonferroni · Tukey · Scheffé의 동시추론을 증명하고, 가족 단위 오류율을 시뮬레이션으로 확인한다.
5. Mercer–Hall 균일성 시험과 Salk 백신 시험으로 무작위화·블록·요인설계의 값을 숫자로 보고, 같은 원리를 하이퍼파라미터 탐색과 A/B 테스트에 쓴다.

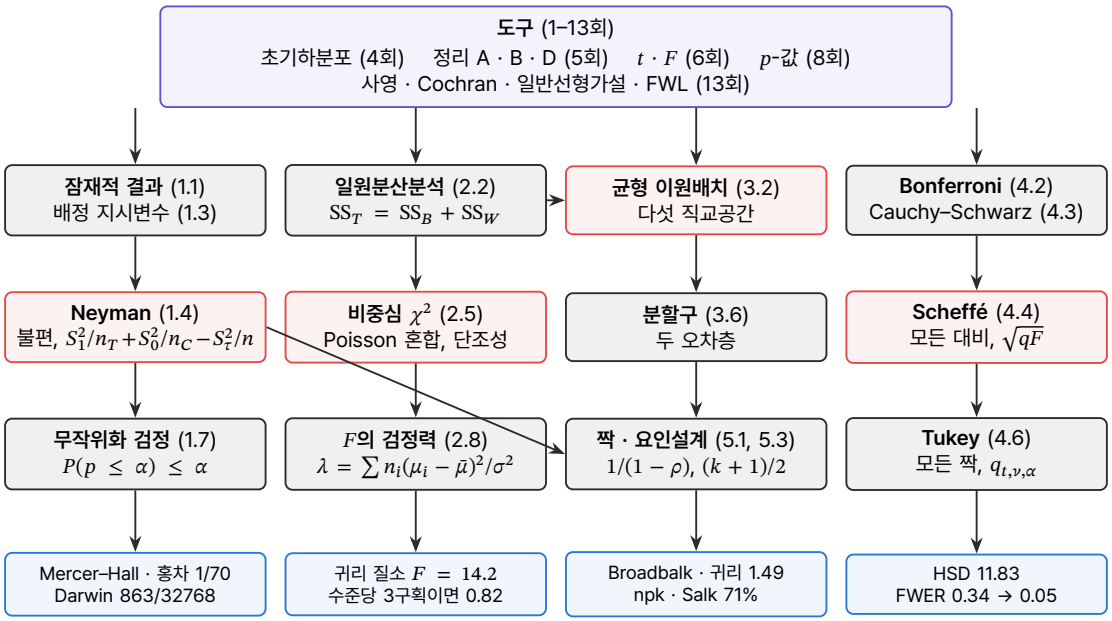


그림 0.1 이번 회차 정리들의 의존 관계. 화살표는 "증명에 쓰인다"는 뜻이다.

0.1 1–13회에서 가져오는 것

앞 회차	내용
4회 정의 1.4 · 정리 1.5	초기하분포와 그 평균·분산
5회 정리 3.8 · 3.11 · 3.17	정규의 선형변환, 무상관이면 독립, 이차형식의 χ^2
6회 정의 3.1 · 4.1	t 분포와 F 분포
8회 정의 3.2 · 정리 4.2 · 5.1	크기와 검정력, p-값의 균등성, Pearson의 χ^2
13회 보조정리 2.4 · 정리 2.5	대칭 멱등행렬과 사영, hat 행렬
13회 정리 4.1 · 4.5 · 4.9	이차형식의 χ^2 판정, Cochran의 정리, 일반선형가설의 F
13회 명제 5.6	누락변수의 치우침 (7절)

0.2 표기

처치(처리)는 t개, 처치 i를 받은 단위(밭의 구획, 환자)는 n_i개, 모두 $N = \sum n_i$개다. 관측은 y_{ij} (i번째 처치의 j번째 단위)로 쓰고, 점으로 평균을 표시한다. $\bar{y}_{i\cdot} = \frac{1}{n_i}\sum_j y_{ij}$, $\bar{y}_{\cdot\cdot} = \frac{1}{N}\sum_{i,j} y_{ij}$. 제곱합은 SS로 쓴다. 13회처럼 관측 전체를 벡터 $\mathbf{y} \in \mathbb{R}^N$로 보고, 사영행렬로 제곱합을 쓴다. $J = \mathbf{1}\mathbf{1}'/N$은 전체 평균으로의 사영이다.

1

무작위화 (1885–1935)

> **역사** 1885년 미국의 철학자 Charles S. Peirce와 제자 Joseph Jastrow는 아주 작은 무게 차이를 사람이 느낄 수 있는지 실험했다. 피험자가 기대나 습관으로 답하지 못하도록, 무게를 올리고 내리는 순서를 섞은 카드 묶음으로 정했다. 기록에 남은 최초의 의도적 무작위화다. 1923년 폴란드의 Jerzy Neyman은 농사 실험을 다룬 학위논문에서, 한 구획이 각 품종을 받았을 때 낼 수확을 "잠재적 수확"으로 적고, 무작위 배정 아래 평균 차이의 분산을 계산했다. 이 생각은 폴란드어로만 남아 있다가 1990년에 영어로 옮겨졌다. Fisher는 1926년과 1935년에 무작위화를 실험 설계의 원리로 세웠다. "무작위화만이 오차를 정당하게 추정하게 해 준다."

1.1 잠재적 결과

정의 1.1 (잠재적 결과와 처치효과) 단위 $i = 1, \dots, n$은 처치를 받으면 $y_i(1)$, 받지 않으면 $y_i(0)$을 낸다. 두 값은 실험 전에 정해진 상수이고, 실제로는 그 가운데 하나만 관측된다. $\tau_i = y_i(1) - y_i(0)$을 단위 i의 **처치효과**, $\bar\tau = \frac{1}{n}\sum \tau_i$를 **평균 처치효과**라 한다. 배정 $W_i \in \{0,1\}$ 아래 관측값은 $y_i = W_i y_i(1) + (1 - W_i) y_i(0)$이다.

이 정의에서 확률은 단위들에 있지 않다. 단위와 그 두 잠재적 결과는 고정되어 있고, 무작위인 것은 오직 배정 W다. 정규분포도, 독립인 오차도 가정하지 않는다. 실험자가 동전을 던졌다는 사실 하나가 모든 확률의 출처다.

정의 1.2 (완전무작위설계) n개 단위 가운데 n_T개를 처치에, 나머지 $n_C = n - n_T$개를 대조에 배정하는 $\binom{n}{n_T}$가지 방법이 모두 같은 확률로 뽑히는 설계를 **완전무작위설계**라 한다.

보조정리 1.3 (배정 지시변수의 적률) 완전무작위설계에서 $E W_i = n_T/n$, $\operatorname{Var} W_i = n_T n_C / n^2$이고 $i \neq j$이면 $\operatorname{Cov}(W_i, W_j) = -\dfrac{n_T n_C}{n^2 (n-1)}$이다.

증명. *전략* 한 단위가 처치에 들어가는 배정과 두 단위가 함께 들어가는 배정의 수를 세어 확률을 구한다. 공분산은 $E W_i W_j - E W_i E W_j$이다. $\sum_i W_i = n_T$가 상수라는 사실로 검산한다. 단계별 증명은 연습 1.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 41쪽에 있다.

정리 1.4 (Neyman (1923)) $S_1^2 = \frac{1}{n-1}\sum(y_i(1)-\bar{y}(1))^2$, S_0^2도 같게, $S_\tau^2 = \frac{1}{n-1}\sum(\tau_i-\bar{\tau})^2$이라 하자. 완전무작위설계에서 $\hat{\tau} = \bar{y}_T - \bar{y}_C$ (처치군과 대조군의 관측 평균의 차)는

$$E\hat{\tau} = \bar{\tau}, \qquad \operatorname{Var}\hat{\tau} = \frac{S_1^2}{n_T} + \frac{S_0^2}{n_C} - \frac{S_\tau^2}{n}$$

를 만족한다. 처치효과가 모든 단위에서 같으면($S_\tau^2 = 0$) 마지막 항이 없어진다.

증명. **전략** $\hat{\tau} - \bar{\tau}$를 중심화한 배정 지시변수 $W_i - n_T/n$의 일차결합으로 쓴다. 그러면 평균은 0이고, 분산은 보조정리 1.3의 분산과 공분산으로 계산하는 이차형식이다.

① $a_i = y_i(1)/n_T + y_i(0)/n_C$로 두면 $\hat{\tau} = \sum_i W_i y_i(1)/n_T - \sum_i (1-W_i) y_i(0)/n_C = \sum_i W_i a_i - \sum_i y_i(0)/n_C$이다. 정의 1.1의 y_i

② $\sum_i \frac{n_T}{n} a_i - \sum_i \frac{y_i(0)}{n_C} = \frac{1}{n}\sum_i y_i(1) + \left(\frac{n_T}{nn_C} - \frac{1}{n_C}\right)\sum_i y_i(0) = \bar{\tau}$이므로

$$\hat{\tau} - \bar{\tau} = \sum_i \left(W_i - \frac{n_T}{n}\right) a_i. \qquad \frac{n_T}{nn_C} - \frac{1}{n_C} = -\frac{1}{n}$$

③ $E(W_i - n_T/n) = 0$이므로 $E\hat{\tau} = \bar{\tau}$이다. 보조정리 1.3, 선형성

④ ②의 일차결합의 분산을 계산한다.

$$\operatorname{Var}\hat{\tau} = \frac{n_T n_C}{n^2}\left(\sum_i a_i^2 - \frac{1}{n-1}\sum_{i\neq j} a_i a_j\right) \qquad \text{보조정리 1.3}$$

$$= \frac{n_T n_C}{n(n-1)}\sum_i (a_i - \bar{a})^2 \qquad \sum_{i\neq j} a_i a_j = \left(\sum_i a_i\right)^2 - \sum_i a_i^2$$

⑤ S_{10}을 $y(1)$과 $y(0)$의 공분산(분모 $n-1$)이라 하면 $\frac{1}{n-1}\sum_i (a_i - \bar{a})^2 = \frac{S_1^2}{n_T^2} + \frac{S_0^2}{n_C^2} + \frac{2S_{10}}{n_T n_C}$이므로

$$\operatorname{Var}\hat{\tau} = \frac{n_C}{n}\frac{S_1^2}{n_T} + \frac{n_T}{n}\frac{S_0^2}{n_C} + \frac{2S_{10}}{n}. \qquad ④\text{, 표본공분산의 쌍선형성}$$

⑥ $2S_{10} = S_1^2 + S_0^2 - S_\tau^2$을 넣고 $\frac{n_C}{nn_T} + \frac{1}{n} = \frac{1}{n_T}$, $\frac{n_T}{nn_C} + \frac{1}{n} = \frac{1}{n_C}$를 쓰면 정리의 식이다. $\tau_i = y_i(1) - y_i(0)$

조건 확인 확률은 배정 W에서만 온다. 잠재적 결과는 상수이므로 모든 적률이 유한하고, 정규성도 표본의 독립도 쓰지 않았다. ∎

S_τ^2는 두 잠재적 결과를 함께 보아야 알 수 있으므로 자료로 추정할 수 없다. 처치군과 대조군의 표본분산 s_T^2, s_C^2로 만든 $s_T^2/n_T + s_C^2/n_C$는 $S_1^2/n_T + S_0^2/n_C$의 불편추정량이어서(연습 1.3), 참 분산을 과대추정한다. 보수적인 쪽으로 틀리는 것이다. 6회의 두 표본 t에서 쓰던 표준오차가 무작위 배정만으로 이렇게 정당화된다.

예제 1.5 (처치가 없는 밭) 1910년 Rothamsted의 Mercer와 Hall은 겉보기에 고른 1에이커의 밀밭을 $20\times 25=500$개 구획으로 나누어 모두 똑같이 기르고 구획마다 수확을 쟀다(**균일성 시험**). 아무 처치도 없으므로 모든 τ_i가 0이다. 500개 가운데 250개를 무작위로 골라 "처치"라 부르면, 정리 1.4로 $\hat{\tau}$의 분산은 $S^2\left(\frac{1}{250}+\frac{1}{250}\right)=0.001680$이다($S^2=0.2100$, 단위는 구획당 파운드의 제곱). 무작위 배정을 2만 번 되풀이하면 $\hat{\tau}$의 평균은 -0.00006, 분산은 0.001679이다.

1.2 Fisher의 무작위화 검정

정의 1.6 (날카로운 귀무가설) 모든 단위에서 $y_i(1)=y_i(0)$이라는 가설을 **날카로운 귀무가설**이라 한다. 이 가설 아래에서는 배정이 무엇이든 관측 벡터 $\mathbf{y}$가 같다.

정리 1.7 (무작위화 검정은 정확하다) 배정 W가 유한집합 $\mathcal{W}$ 위에서 균등하게 뽑히고, $T(w,\mathbf{y})$가 아무 검정통계량이라 하자. W'을 W와 독립이고 같은 분포를 갖는 배정이라 하고

$$p(W)=P_{W'}\big(T(W',\mathbf{y})\geq T(W,\mathbf{y})\big)$$

로 정의하면, 날카로운 귀무가설 아래 모든 $\alpha\in(0,1)$에 대해 $P\big(p(W)\leq\alpha\big)\leq\alpha$이다.

증명. **전략** 귀무가설 아래 $\mathbf{y}$는 배정과 무관하게 고정된다. 그러면 T는 균등 배정만으로 무작위인 이산 확률변수이고, $p(W)$는 그 생존함수 G에 T 자신을 넣은 것이다. 생존함수에 자기 자신을 넣은 값은 균등분포보다 작은 쪽으로 몰리지 않는다.

① 날카로운 귀무가설 아래 $\mathbf{y}$는 W에 의존하지 않는다. 따라서 $T=T(W,\mathbf{y})$의 무작위성은 $\mathcal{W}$ 위의 균등분포에서만 온다. — 정의 1.6

② $G(t)=P_{W'}\big(T(W',\mathbf{y})\geq t\big)$로 두면 $p(W)=G(T)$이다. — W'과 W의 분포가 같다

③ G는 감소함수다. $t\leq t'$이면 $\{T\geq t'\}\subset\{T\geq t\}$이기 때문이다. — 단조성 [T1]

④ T가 가질 수 있는 값은 유한개다. 그 가운데 $G(t)\leq\alpha$인 가장 작은 값을 t_α라 하면 $\{G(T)\leq\alpha\}=\{T\geq t_\alpha\}$이다. — ③, $\mathcal{W}$가 유한

⑤ $P\big(p(W)\leq\alpha\big)=P(T\geq t_\alpha)=G(t_\alpha)\leq\alpha$이다. 그런 값이 없으면 사건이 공집합이어서 확률은 0이다. — ②, t_α의 정의

■

이 정리는 정규분포도, 표본의 독립도, 큰 표본도 쓰지 않는다. 쓰는 것은 실험자가 실제로 무작위로 배정했다는 사실뿐이다. 무작위화하지 않은 실험에는 이 보증이 없다. 8회 정리 4.2의 p-값의

균등성을 이산판으로 다시 본 것이기도 하다.

예제 1.8 (홍차 감별 숙녀) Rothamsted의 차 시간에 조류학자 Muriel Bristol은 우유를 먼저 부은 차와 차를 먼저 부은 차를 맛으로 가릴 수 있다고 했다. Fisher는 **실험의 설계**(1935)에서 이 일을 실험으로 만들었다. 컵 8개 가운데 무작위로 고른 4개에 우유를 먼저 붓고, 숙녀에게 그 4개를 고르게 한다. 숙녀가 가리지 못한다면(날카로운 귀무가설) 맞힌 개수 K는 초기하분포를 따르고

$$P(K=k)=\frac{\binom{4}{k}\binom{4}{4-k}}{\binom{8}{4}}: \qquad \frac{1}{70},\ \frac{16}{70},\ \frac{36}{70},\ \frac{16}{70},\ \frac{1}{70} \qquad (k=0,\dots,4)$$

이다. 넷 다 맞힐 확률은 $1/70=0.014$이고, 셋 이상은 $17/70=0.243$이다. 5%에서 의미가 있는 결과는 넷 다 맞히는 것뿐이다. 이 검정이 Fisher의 정확검정이다.

예제 1.9 (Darwin의 옥수수) Darwin(1876)은 타가수정한 옥수수와 자가수정한 옥수수를 같은 화분에 짝지어 15쌍을 기르고 키(인치)를 쟀다. 짝마다 차이 d_i(타가 − 자가)의 합은 39.25다. Fisher(1935)는 각 짝 안에서 어느 쪽이 타가수정인지 동전으로 정했다고 상상했다. 날카로운 귀무가설 아래에서는 각 d_i의 부호만 동전에 따라 바뀌므로, $2^{15}=32768$가지 부호 조합이 같은 확률을 갖는다. 그 가운데 $\sum \pm|d_i|$가 39.25 이상인 것은 863개이고, 한쪽 p-값은 $863/32768=0.0263$, 양쪽은 0.0527이다. 정규분포를 가정한 대응 t 검정(6회)의 양쪽 p-값은 0.0497이다. Fisher의 결론은 t 검정이 무작위화 검정의 좋은 근사라는 것이었다(그림 1.1 오른쪽).

연습문제

연습 1.1 (컵 여섯 개 ★) 컵 6개 가운데 3개에 우유를 먼저 부으면, 숙녀가 모두 맞혀도 p-값이 1/20임을 보여라. 컵 8개 대신 우유 먼저 2개, 차 먼저 6개로 하면 모두 맞힐 확률은 얼마인가? 어느 설계가 더 좋은가? 풀이 40쪽

연습 1.2 (두 단위 ★) $n=2$, $n_T=n_C=1$이면 $\hat{\tau}$는 $y_1(1)-y_2(0)$과 $y_2(1)-y_1(0)$ 가운데 하나다. 직접 분산을 계산해 정리 1.4의 식과 같음을 보여라. 풀이 40쪽

연습 1.3 (유한 모집단의 표본분산 ★★) 완전무작위설계에서 처치군의 표본분산 $s_T^2=\frac{1}{n_T-1}\sum_{i:W_i=1}(y_i(1)-\bar{y}_T)^2$의 기댓값이 S_1^2임을 보조정리 1.3로 보여라. ($\sum_i W_i(y_i(1)-\bar{y}(1))^2$과 $(\bar{y}_T-\bar{y}(1))^2$의 기댓값을 각각 구하라.) 풀이 40쪽

연습 1.4 (정확하지만 보수적 ★★) 홍차 실험에서 "$p\le 0.05$이면 기각"하는 검정의 실제 크기가 1/70임을 보여라. 정리 1.7의 부등호가 등호가 되지 않는 이유는 무엇인가? 12회의 확률화 검정으로 크기를 정확히 0.05로 만들려면 어떻게 하는가? 풀이 40쪽

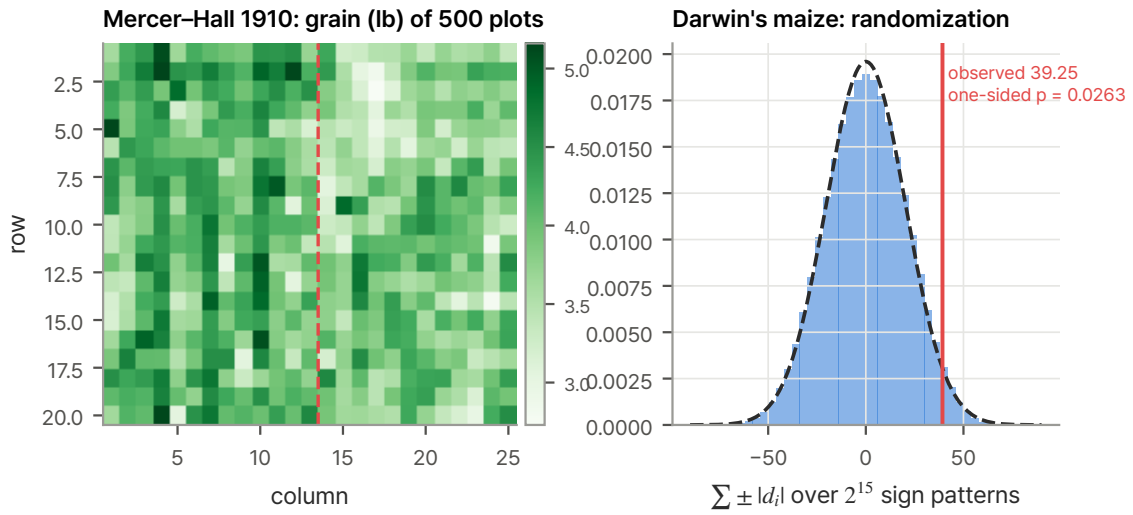


그림 1.1 왼쪽: Mercer–Hall의 균일성 시험. 같은 처치를 받은 500개 구획의 수확(파운드)이 밭의 위치에 따라 고르지 않다. 번호가 작은 열들(1–13)이 큰 열들(14–25)보다 대체로 높다. 세로 점선은 7절의 "체계적 배치"의 경계다. 오른쪽: Darwin의 15쌍에 대한 무작위화 분포. 2^{15}개 부호 조합의 $\sum \pm |d_i|$와 관측값 39.25(빨강). 점선은 같은 평균과 분산의 정규 근사다.

연습 1.5 (효과가 일정하다는 가설 ★★) 모든 단위에서 $\tau_i = \tau_0$이라는 가설 아래에서도 관측값으로부터 모든 잠재적 결과를 채울 수 있음을 보이고, 이것으로 Darwin의 자료에서 τ의 신뢰구간을 무작위화 검정을 뒤집어 만드는 방법을 설명하라(8회 정리 3.5의 쌍대성). 풀이 40쪽

연습 1.6 (배정 지시변수의 적률: 증명 ★ 유도형) 보조정리 1.3를 다음 순서로 증명하라.

(a) 단위 i가 처치에 들어가는 배정의 수가 $\binom{n-1}{n_T-1}$임을 보이고 $EW_i = P(W_i = 1) = n_T/n$을 얻어라. W_i는 0과 1만 가지므로 $\operatorname{Var} W_i = EW_i(1 - EW_i)$이다.

(b) $i \neq j$가 함께 처치에 들어가는 배정의 수가 $\binom{n-2}{n_T-2}$임을 보이고 $P(W_i = W_j = 1) = \frac{n_T(n_T-1)}{n(n-1)}$을 얻어라.

(c) $\operatorname{Cov}(W_i, W_j) = EW_iW_j - EW_iEW_j$에 (a)와 (b)를 넣어 정리의 식을 얻어라.

(d) $\sum_i W_i = n_T$가 상수이므로 $0 = \operatorname{Var} \sum_i W_i = n \operatorname{Var} W_1 + n(n-1) \operatorname{Cov}(W_1, W_2)$이다. 이것으로 (c)를 검산하라. 4회의 초기하분포의 분산과 같은 계산이다.

풀이 41쪽

2

일원분산분석과 비중심 분포

> **역사** Fisher는 1918년 유전학 논문에서 "분산"이라는 말을 처음 썼다. 1923년 Mackenzie와 함께 쓴 감자 비료 실험의 분석에서 처음으로 분산분석표가 나왔다. 전체 제곱합을 품종, 비료, 나머지의 몫으로 나누고 각 몫을 자유도로 나눈 표다. Fisher는 두 분산의 비 대신 $z = \frac{1}{2}\log(\text{비})$의 분포표를 만들었다. 1934년 아이오와의 George Snedecor가 비 자체의 표를 내면서 Fisher를 기려 그것을 F라 불렀다. Fisher는 이 이름을 달가워하지 않았다.

2.1 일원분산분석은 사영이다

정의 2.1 (일원배치 모형) 처치 $i = 1, \dots, t$의 단위 $j = 1, \dots, n_i$에서 $y_{ij} = \mu_i + \varepsilon_{ij}$, $\varepsilon_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$이다. 13회의 선형모형으로 쓰면 X는 $N \times t$의 처치 표시 행렬이고 $\boldsymbol{\beta} = (\mu_1, \dots, \mu_t)'$다.

정리 2.2 (일원분산분석) H를 $\mathrm{col}(X)$ 위로의 사영이라 하면 $(H\mathbf{y})_{ij} = \bar{y}_{i\cdot}$이고

$$\underbrace{\sum_{i,j}(y_{ij} - \bar{y}_{\cdot\cdot})^2}_{\mathrm{SS}_T = \|(I-J)\mathbf{y}\|^2} = \underbrace{\sum_{i} n_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}_{\mathrm{SS}_B = \|(H-J)\mathbf{y}\|^2} + \underbrace{\sum_{i,j}(y_{ij} - \bar{y}_{i\cdot})^2}_{\mathrm{SS}_W = \|(I-H)\mathbf{y}\|^2}$$

이다. $H - J$와 $I - H$는 계수가 $t-1$, $N-t$인 직교사영이다. $H_0 : \mu_1 = \cdots = \mu_t$ 아래

$$F = \frac{\mathrm{SS}_B/(t-1)}{\mathrm{SS}_W/(N-t)} \sim F(t-1,\ N-t)$$

이고, 이 F 검정은 H_0의 가능도비 검정이다.

증명. **전략** $\mathbb{R}^N$을 서로 직교하는 세 공간, 곧 상수 벡터, 처치 평균의 편차, 처치 안의 편차로 나눈다. 분포는 13회의 Cochran 정리가 주고, 가능도비 검정이라는 것은 일원배치의 귀무가설을 일반선형가설로 쓰면 나온다.

① $X'X = \mathrm{diag}(n_1, \dots, n_t)$이고 $X'\mathbf{y}$의 성분은 처치별 합이므로 $\hat{\boldsymbol{\beta}} = (\bar{y}_{1\cdot}, \dots, \bar{y}_{t\cdot})'$이고 $(H\mathbf{y})_{ij} = (X\hat{\boldsymbol{\beta}})_{ij} = \bar{y}_{i\cdot}$이다. 정규방정식, 13회 정리 2.2

② $\mathbf{1} = X\mathbf{1}_t \in \mathrm{col}(X)$이므로 $HJ = JH = J$이다. 따라서 $H - J$는 계수 $t-1$, $I - H$는 계수 $N-t$인 대칭 멱등행렬이다. 13회 정리 2.5의 증명

③ $I-J=(H-J)+(I-H)$이고 $(H-J)(I-H)=0$이므로 $\mathrm{SS}_T=\|(I-J)\mathbf{y}\|^2=\mathrm{SS}_B+\mathrm{SS}_W$이다. 직교하는 사영, 피타고라스

④ $I=J+(H-J)+(I-H)$이고 계수의 합이 $1+(t-1)+(N-t)=N$이다. 13회 정리 4.5(Cochran)의 조건

⑤ H_0 아래 평균 벡터는 $\mu\mathbf{1}$이고 $(H-J)\mu\mathbf{1}=(I-H)\mu\mathbf{1}=\mathbf{0}$이다. 따라서 $\mathrm{SS}_B/\sigma^2\sim\chi^2(t-1)$과 $\mathrm{SS}_W/\sigma^2\sim\chi^2(N-t)$가 독립이고 $F\sim F(t-1,N-t)$이다. ④, 13회 정리 4.5, 따름정리 4.2

⑥ H_0은 13회 정리 4.9에서 C를 대비 $\mu_i-\mu_t$ $(i<t)$로 둔 일반선형가설이다. 제약 아래의 적합은 $J\mathbf{y}$이고 $\mathrm{RSS}_0-\mathrm{RSS}=\|(H-J)\mathbf{y}\|^2=\mathrm{SS}_B$이므로 F 검정은 가능도비 검정이다. 13회 정리 4.9(d)

■

$\mathrm{SS}_W=\sum_i(n_i-1)s_i^2$이므로 $\mathrm{SS}_W/(N-t)$는 처치마다의 표본분산을 합친 것이다. $t=2$이면 F는 6회 정리 5.4의 두 표본 t의 제곱이다(연습 2.2). "분산분석"이라는 이름과 달리 이 검정이 비교하는 것은 평균이다. 처치 평균들이 흩어진 정도를 처치 안의 흩어짐과 견준다.

예제 2.3 (귀리와 질소 (Yates 1931)) 1931년 Rothamsted에서 귀리 세 품종에 질소 비료를 에이커당 0, 0.2, 0.4, 0.6 헌드레드웨이트(cwt) 주었다. 수확은 1/80에이커 구획마다 1/4파운드 단위로 쟀고, 질소 수준마다 18구획이다. 품종과 블록을 무시하고 질소만으로 일원분산분석을 하면 평균은 차례로 79.4, 98.9, 114.2, 123.4이고

$$\mathrm{SS}_B=20020.5\ (3),\qquad \mathrm{SS}_W=31965.4\ (68),\qquad F=14.20,\qquad p=2.8\times10^{-7}$$

이다(괄호는 자유도). 오차의 표준편차 추정은 $s=21.7$이다. 3절에서 블록과 품종을 넣으면 오차가 크게 줄어든다.

2.2 비중심 분포와 검정력

귀무가설이 거짓일 때 F는 어떤 분포를 따르는가? 그것을 알아야 검정력을 계산하고, 실험 전에 반복 수를 정할 수 있다.

정의 2.4 (비중심 χ^2) $\mathbf{X}\sim\mathcal{N}_k(\boldsymbol{\mu},I)$일 때 $\|\mathbf{X}\|^2$의 분포를 자유도 k, **비중심도** $\lambda=\|\boldsymbol{\mu}\|^2$의 비중심 χ^2이라 하고 $\chi^2(k,\lambda)$로 쓴다.

정리 2.5 (비중심 χ^2의 성질)

(a) 분포는 $\boldsymbol{\mu}$에 $\|\boldsymbol{\mu}\|^2$을 통해서만 의존한다.

(b) $t < \frac{1}{2}$에서 mgf는 $(1-2t)^{-k/2}\exp\!\left(\frac{\lambda t}{1-2t}\right)$이다.

(c) $J \sim \mathrm{Poi}(\lambda/2)$이고 $J = j$일 때 $\chi^2(k+2j)$인 혼합분포와 같다.

(d) 평균은 $k+\lambda$, 분산은 $2k+4\lambda$이다.

(e) 모든 c에 대해 $P\big(\chi^2(k,\lambda) > c\big)$는 λ의 증가함수다.

증명. **전략** (a)는 회전으로 평균 벡터를 한 축 위로 옮긴다. (b)는 한 성분의 mgf를 제곱 완성으로 구해 곱한다. (c)는 (b)의 mgf를 Poisson 가중치의 급수로 펼친다. (d)와 (e)는 (c)의 혼합 표현에 조건을 거는 계산이다.

(a)

① $\boldsymbol{\mu} \neq \mathbf{0}$이면 첫 행이 $\boldsymbol{\mu}/\|\boldsymbol{\mu}\|$인 직교행렬 Q가 있다. — 정규직교 기저로 확장

② $Q\mathbf{X} \sim \mathcal{N}_k\big(Q\boldsymbol{\mu}, QQ'\big) = \mathcal{N}_k\big((\|\boldsymbol{\mu}\|, 0, \dots, 0)', I\big)$이다. — 5회 정리 3.8

③ $\|Q\mathbf{X}\| = \|\mathbf{X}\|$이므로 $\|\mathbf{X}\|^2$의 분포는 $\|\boldsymbol{\mu}\|$로만 정해진다. — $Q'Q = I$

(b)

① $Z \sim \mathcal{N}(0,1)$, $t < \frac{1}{2}$이면 지수의 제곱을 완성해

$$t(z+a)^2 - \frac{z^2}{2} = -\frac{1-2t}{2}\Big(z - \frac{2ta}{1-2t}\Big)^2 + \frac{ta^2}{1-2t}.$$ 전개해 비교

② 따라서 $E e^{t(Z+a)^2} = (1-2t)^{-1/2}\exp\!\left(\frac{ta^2}{1-2t}\right)$이다. — 분산 $\frac{1}{1-2t}$인 정규밀도의 적분은 1

③ $\|\mathbf{X}\|^2 = \sum_i X_i^2$의 항들은 독립이므로 mgf를 곱하면 $(1-2t)^{-k/2}\exp\!\left(\frac{t\lambda}{1-2t}\right)$이다. — 3회 정리 3.4(d)

(c)

① $\frac{\lambda t}{1-2t} = \frac{\lambda}{2}\left(\frac{1}{1-2t} - 1\right)$이므로 지수함수를 급수로 펼치면

$$(1-2t)^{-k/2} e^{\lambda t/(1-2t)} = \sum_{j=0}^{\infty} e^{-\lambda/2}\frac{(\lambda/2)^j}{j!}(1-2t)^{-(k+2j)/2}.$$ $e^x = \sum_j x^j/j!$

② 오른쪽은 $J \sim \mathrm{Poi}(\lambda/2)$이고 $J = j$일 때 $\chi^2(k+2j)$인 혼합분포의 mgf $E\big[E(e^{tY} \mid J)\big]$이다. — 3회 정리 2.5

③ 두 mgf가 $t < \frac{1}{2}$에서 같으므로 두 분포가 같다. — mgf의 유일성, 2회 3절

(d)

① 평균은 $E\big[E(Y\mid J)\big]=E(k+2J)=k+\lambda$이다. (c), 3회 정리 2.5

② 분산은 $E\big[\mathrm{Var}(Y\mid J)\big]+\mathrm{Var}\big(E(Y\mid J)\big)=E[2(k+2J)]+\mathrm{Var}(2J)=2k+2\lambda+2\lambda$이다. 3회 정리 2.6

(e)

① $g(j)=P\big(\chi^2(k+2j)>c\big)$는 j의 증가함수다. $\chi^2(k+2j+2)$는 $\chi^2(k+2j)$에 독립인 $\chi^2(2)\ge 0$을 더한 것이기 때문이다. 단조성

② $\lambda<\lambda'$이면 $\mathrm{Poi}(\lambda'/2)$는 $J\sim\mathrm{Poi}(\lambda/2)$에 독립인 $J''\sim\mathrm{Poi}\big((\lambda'-\lambda)/2\big)$를 더한 것의 분포다. 4회 정리 2.4

③ 따라서 $P\big(\chi^2(k,\lambda')>c\big)=E\,g(J+J'')\ge E\,g(J)=P\big(\chi^2(k,\lambda)>c\big)$이다. (c), ①②

조건 확인 (c)의 급수는 모든 항이 양수이므로 기댓값과 합의 순서를 바꿀 수 있다 [T6]. 두 mgf는 0을 포함하는 구간 $t<\frac{1}{2}$에서 유한하므로 유일성을 쓸 수 있다. ■

따름정리 2.6 (사영의 비중심도) $\mathbf{y}\sim\mathcal{N}_N(\boldsymbol{\mu},\sigma^2 I)$이고 P가 계수 r의 대칭 멱등행렬이면 $\mathbf{y}'P\mathbf{y}/\sigma^2\sim\chi^2\big(r,\ \boldsymbol{\mu}'P\boldsymbol{\mu}/\sigma^2\big)$이다.

증명. **전략** P를 정규직교 열을 가진 행렬로 $P=UU'$라 쓰면 $\mathbf{y}'P\mathbf{y}$는 r차원 정규벡터의 제곱길이다. 정의 2.4에 그대로 들어간다. 단계별 증명은 연습 2.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 42쪽에 있다.

정의 2.7 (비중심 F) $Q\sim\chi^2(q,\lambda)$와 $V\sim\chi^2(m)$이 독립일 때 $\dfrac{Q/q}{V/m}$의 분포를 비중심 $F(q,m,\lambda)$라 한다.

정리 2.8 (F 검정의 검정력) 13회 정리 4.9의 일반선형가설 $C\boldsymbol{\beta}=\mathbf{d}$에서, 참 모수가 무엇이든 $F\sim F(q,\ n-p,\ \lambda)$이고

$$\lambda=\frac{(C\boldsymbol{\beta}-\mathbf{d})'\big[C(X'X)^{-1}C'\big]^{-1}(C\boldsymbol{\beta}-\mathbf{d})}{\sigma^2}.$$

일원분산분석에서는 $\lambda=\sum_i n_i(\mu_i-\bar\mu)^2/\sigma^2$, $\bar\mu=\sum n_i\mu_i/N$이다. 크기 α인 F 검정의 검정력 $P(F>F_{q,n-p,\alpha})$는 λ의 증가함수다.

증명. **전략** 분자를 q차원 정규벡터의 제곱 길이로 바꿔 비중심도를 읽는다. 분모는 평균과 무관하고 분자와 독립이다. 검정력의 단조성은 분모에 조건을 걸고 정리 2.5(e)를 쓴다. 단계별 증명은 연습 2.7(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 42쪽에 있다.

예제 2.9 (반복 수를 정하기) 질소 네 수준의 참 평균이 예제 2.3의 평균과 같고, 오차의 표준편차가 3절의 분할구 분석이 준 $\sigma = \sqrt{177.1} = 13.3$이라 하자. 수준마다 n구획이면 $\lambda = n\sum(\mu_i - \bar{\mu})^2/\sigma^2$이고, 5% 검정의 검정력은

$$n=2:\ 0.44, \qquad n=3:\ 0.82, \qquad n=4:\ 0.96$$

이다(그림 2.1). 질소의 주효과만 보려면 수준마다 세 구획이면 충분했다. Yates의 실험이 수준마다 18구획을 쓴 것은 품종의 차이와 품종 × 질소의 상호작용까지 보려 했기 때문이다.

연습문제

연습 2.1 (제곱합의 공식 ★) $\mathrm{SS}_B = \sum_i n_i \bar{y}_{i\cdot}^2 - N\bar{y}_{\cdot\cdot}^2$과 $\mathrm{SS}_W = \sum_i (n_i - 1)s_i^2$을 보여라. 풀이 41쪽

연습 2.2 (두 표본이면 ★) $t = 2$일 때 정리 2.2의 F가 6회 정리 5.4의 합동 두 표본 t의 제곱임을 보여라. 풀이 41쪽

연습 2.3 (비중심도의 계산 ★★) 일원배치에서 $(H-J)\boldsymbol{\mu}$의 (i,j) 성분이 $\mu_i - \bar{\mu}$ ($\bar{\mu} = \sum n_i\mu_i/N$)임을 보이고, 정리 2.8의 일반식에 C를 $\mu_i - \mu_t$ ($i < t$)의 대비로 넣어도 같은 λ가 나옴을 $t = 2$에서 확인하라. 풀이 41쪽

연습 2.4 (평균과 분산 ★) 정의 2.4에서 직접 $E\|\mathbf{X}\|^2 = k + \lambda$를 구하고, $\mathrm{Var}(Z+a)^2 = 2 + 4a^2$을 써서 분산 $2k + 4\lambda$를 다시 얻어라. 풀이 41쪽

연습 2.5 (오차가 두 배면 ★) 예제 2.9에서 σ가 두 배이면 같은 검정력을 얻는 데 반복 수가 대략 몇 배 필요한가? λ의 식으로 답하고 노트북으로 확인하라. 풀이 41쪽

연습 2.6 (사영의 비중심도: 증명 ★ 유도형) 따름정리 2.6를 다음 순서로 증명하라.

(a) $P = UU'$, $U'U = I_r$인 $N \times r$ 행렬 U가 있음을 13회 보조정리 2.4로 확인하라.

(b) $U'\mathbf{y}/\sigma \sim \mathcal{N}_r(U'\boldsymbol{\mu}/\sigma,\ I_r)$임을 보여라. 공분산이 $U'(\sigma^2 I)U/\sigma^2$이다.

(c) $\|U'\mathbf{y}\|^2 = \mathbf{y}'P\mathbf{y}$, $\|U'\boldsymbol{\mu}\|^2 = \boldsymbol{\mu}'P\boldsymbol{\mu}$를 보이고 정의 2.4로 결론을 내라.

풀이 42쪽

연습 2.7 (F 검정의 검정력: 증명 ★★ 유도형) 정리 2.8를 다음 순서로 증명하라. $W = C(X'X)^{-1}C'$로 둔다.

(a) $C\hat{\boldsymbol{\beta}} - \mathbf{d} \sim \mathcal{N}_q\big(C\boldsymbol{\beta} - \mathbf{d},\ \sigma^2 W\big)$임을 보이고, $W^{-1/2}(C\hat{\boldsymbol{\beta}} - \mathbf{d})/\sigma$의 분포로 분자 $(C\hat{\boldsymbol{\beta}} - \mathbf{d})'W^{-1}(C\hat{\boldsymbol{\beta}} - \mathbf{d})/\sigma^2$이 $\chi^2(q, \lambda)$임을 보여라.

(b) 분모 RSS/σ^2이 참 모수와 무관하게 $\chi^2(n-p)$이고 분자와 독립임을 13회 정리 4.7로 확인하라.

(c) 일원분산분석에서는 분자가 $\|(H-J)\mathbf{y}\|^2/\sigma^2$이다. 따름정리 2.6와 연습 2.3로 $\lambda =$

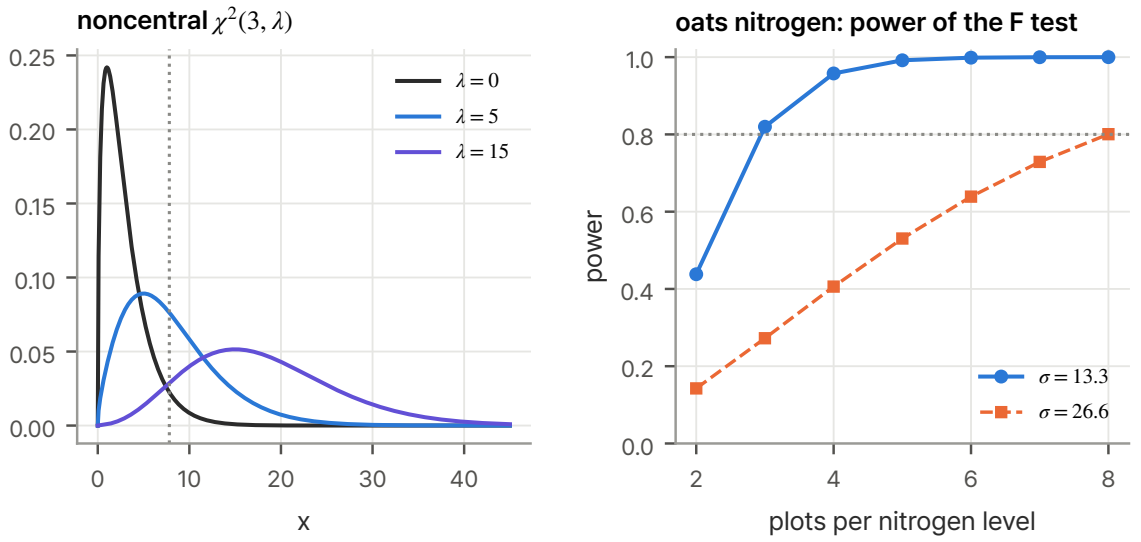


그림 2.1 왼쪽: 자유도 3인 비중심 χ^2의 밀도. 비중심도 λ가 커지면 분포가 오른쪽으로 옮겨 가고 퍼진다. 점선은 중심 $\chi^2(3)$의 95% 분위수다. 오른쪽: 귀리 질소 효과에서 수준당 구획 수 n에 따른 5% F 검정의 검정력. $\sigma = 13.3$이면(파랑) 세 구획에서 0.8을 넘고, σ가 두 배면(주황) 여덟 구획이 필요하다(연습 2.5).

$\sum_i n_i(\mu_i - \bar{\mu})^2/\sigma^2$을 얻어라.

(d) $V = \mathrm{RSS}/\sigma^2$에 조건을 걸면 $P(F > c \mid V) = P\big(Q > cqV/(n-p) \mid V\big)$이다. 이것이 λ의 증가함수임을 정리 2.5(e)로 보이고, V에 대해 평균해도 증가함수임을 보여라.

풀이 42쪽

3

블록과 이원분산분석

> **역사** Fisher의 첫 Rothamsted 논문(1921)은 Broadbalk 밀밭의 기록을 다뤘다. 1843년부터 같은 구획에 같은 거름을 주어 온 밭이다. 그는 수확의 변동을 거름(구획)의 몫, 해(날씨)의 몫, 나머지로 나누었다. 같은 생각을 설계에 쓰면 **블록**이 된다. 땅심이 비슷한 구획들을 한 블록으로 묶고, 블록 안에서 처치를 무작위로 배치하면 블록 사이의 땅심 차이가 처치의 비교에 섞이지 않는다. Fisher는 이것을 "국소 관리"라 불렀다. Frank Yates는 1935년의 **복합 실험**에서 한 구획을 다시 쪼개 두 번째 처치를 배치하는 분할구 설계를 정리했다.



그림 3.1 세 설계. 완전무작위설계는 밭 전체에서 처치를 섞고, 무작위 블록 설계는 블록마다 처치를 한 번씩 섞는다. 분할구 설계는 큰 단위(주구)에 한 요인을, 그 안의 작은 단위(세구)에 다른 요인을 섞는다. 무작위화가 일어난 단위의 크기가 오차를 재는 단위다.

3.1 균형 이원배치의 직교 분해

정의 3.1 (균형 이원배치) 요인 A의 수준 $i = 1, \dots, a$와 요인 B의 수준 $j = 1, \dots, b$의 모든 조합(칸)마다 r개씩, 모두 $N = abr$개의 관측 y_{ijk}가 있다. 모형은

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \qquad \sum_i \alpha_i = \sum_j \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$$

이고 $\varepsilon_{ijk} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$이다. $(\alpha\beta)_{ij}$를 **상호작용**이라 한다. 처치 t개를 블록 b개에 하나씩 배치한 **무작위 블록 설계**는 $r = 1$이고 상호작용이 없는(가법) 경우다.

정리 3.2 (균형 이원배치의 분해) $\mathbb{R}^N$의 다섯 부분공간

$$V_0=\{c\mathbf{1}\},\quad V_A=\{v_{ijk}=f(i):\textstyle\sum_i f(i)=0\},\quad V_B=\{v_{ijk}=g(j):\textstyle\sum_j g(j)=0\},$$
$$V_{AB}=\{v_{ijk}=h(i,j):\textstyle\sum_i h(i,j)=\sum_j h(i,j)=0\},\quad V_E=\{v:\textstyle\sum_k v_{ijk}=0\ (\text{모든 칸})\}$$

는 서로 직교하고, 차원은 $1,\ a-1,\ b-1,\ (a-1)(b-1),\ ab(r-1)$이며, 합하면 $\mathbb{R}^N$ 전체다. $\mathbf{y}$의 각 공간 위로의 사영은

$$\bar y_{\cdots},\quad \bar y_{i\cdot\cdot}-\bar y_{\cdots},\quad \bar y_{\cdot j\cdot}-\bar y_{\cdots},\quad \bar y_{ij\cdot}-\bar y_{i\cdot\cdot}-\bar y_{\cdot j\cdot}+\bar y_{\cdots},\quad y_{ijk}-\bar y_{ij\cdot}.$$

이다. 따라서 $\mathrm{SS}_T=\mathrm{SS}_A+\mathrm{SS}_B+\mathrm{SS}_{AB}+\mathrm{SS}_E$이고, 정규 오차 아래 각 제곱합은 서로 독립이며, 예컨대 $H_0:\alpha_i\equiv 0$ 아래 $F_A=\dfrac{\mathrm{SS}_A/(a-1)}{\mathrm{SS}_E/(ab(r-1))}\sim F(a-1,ab(r-1))$이다.

증명. **전략** 다섯 공간이 서로 직교하고 차원의 합이 N임을 보이면 $\mathbb{R}^N$의 직교분해가 된다. 그다음 $\mathbf{y}$를 다섯 조각의 합으로 쓰는 항등식에서 각 조각이 제 공간에 속함을 확인하면, 분해의 유일성으로 그 조각들이 사영이다.

직교

① $f\in V_A$, $g\in V_B$이면 $\sum_{ijk}f(i)g(j)=r\big(\sum_i f(i)\big)\big(\sum_j g(j)\big)=0$이다. — 균형: 칸마다 r개

② $h\in V_{AB}$이면 $\sum_{ijk}f(i)h(i,j)=r\sum_i f(i)\sum_j h(i,j)=0$이다. V_B와 V_{AB}도 같은 계산이다. — V_{AB}의 정의

③ $e\in V_E$는 칸마다 합이 0이므로, 칸 안에서 상수인 벡터(V_0,V_A,V_B,V_{AB}의 원소)와 모두 직교한다. — 칸별로 묶어 더한다

④ V_A,V_B,V_{AB},V_E의 원소는 성분의 합이 0이므로 $V_0=\{c\mathbf{1}\}$과 직교한다. — $\langle c\mathbf{1},\mathbf{v}\rangle=c\sum v$

차원과 합

⑤ 차원은 자유롭게 고를 수 있는 값의 수다. V_A는 a개 값에 합 0의 제약 하나이므로 $a-1$, V_{AB}는 $a\times b$ 표에 행과 열의 합 0의 제약이 독립으로 $a+b-1$개이므로 $(a-1)(b-1)$, V_E는 칸마다 $r-1$이다. — 제약을 센다

⑥ $1+(a-1)+(b-1)+(a-1)(b-1)+ab(r-1)=ab+ab(r-1)=N$이므로 다섯 공간의 직교합이 $\mathbb{R}^N$ 전체다. — 직교하는 공간의 차원은 더해진다

사영과 분포

⑦ y_{ijk}를 정리에 적힌 다섯 조각의 합으로 쓰면 항등식이고, 각 조각은 차례로 제 공간에 속한다. 예컨대 넷째 조각은 i로 더해도 j로 더해도 0이다. — 대입해 확인

⑧ 직교분해에서 각 성분은 하나뿐이므로 각 조각이 사영이고, $\mathrm{SS}_T = \mathrm{SS}_A + \mathrm{SS}_B + \mathrm{SS}_{AB} + \mathrm{SS}_E$ 이다. ⑥, 피타고라스

⑨ 정규 오차 아래 다섯 이차형식은 서로 독립이다. $H_0 : \alpha_i \equiv 0$ 아래 평균 벡터의 V_A 성분과 V_E 성분이 0이므로 $\mathrm{SS}_A/\sigma^2 \sim \chi^2(a-1)$, $\mathrm{SS}_E/\sigma^2 \sim \chi^2(ab(r-1))$이고 $F_A \sim F\big(a-1, ab(r-1)\big)$ 이다. 13회 정리 4.5, 따름정리 4.2

모든 곱의 합이 칸마다 r개씩 똑같이 세어진다는 것, 곧 **균형**이 ①과 ②의 열쇠다. ∎

무작위 블록 설계($r = 1$, 가법)에서는 V_E가 비어 있고, 상호작용이 없다는 가정 아래 V_{AB}가 오차의 자리를 맡는다. 오차의 자유도는 $(t-1)(b-1)$이다. 블록을 넣으면 블록 사이의 변동 SS_B가 오차에서 빠져나간다. 블록이 땅심을 잘 묶었다면 오차가 줄어 처치의 비교가 날카로워진다.

예제 3.3 (Broadbalk, 74년) Broadbalk 밀밭의 17개 구획(서로 다른 거름)에서 1852–1925년의 74년 동안 잰 수확(헥타르당 톤)을 구획 × 해의 가법 모형으로 나누면

$$\mathrm{SS}_{\text{해}} = 316.6\ (73), \qquad \mathrm{SS}_{\text{구획}} = 253.1\ (16), \qquad \mathrm{SS}_{\text{나머지}} = 144.2\ (1168)$$

이다. 전체 변동의 44%가 날씨, 35%가 거름, 20%가 나머지다. $F_{\text{구획}} = 128.1$, $F_{\text{해}} = 35.1$이다. 구획 평균은 거름이 없는 3번 0.85, 질소 없이 광물질만 준 5번 0.98, 광물질에 질소를 1·2·3배로 더한 6·7·8번 1.57·2.22·2.53, 두엄을 준 2번 2.44다. 질소의 용량 반응이 뚜렷하고, 질소가 없으면 광물질은 거의 듣지 않는다.

가법 모형은 "거름의 효과가 해마다 같다"는 가정이다. 반복이 없는 이원배치($r = 1$)에서는 상호작용 전체를 검정할 수 없지만, Tukey(1949)는 특별한 꼴의 상호작용 하나를 검정하는 방법을 냈다.

정의 3.4 (Tukey의 비가법성 1자유도) $r = 1$인 이원배치에서 $\hat{a}_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}$, $\hat{b}_j = \bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}$이라 하고

$$\mathrm{SS}_N = \frac{\left(\sum_{i,j} y_{ij}\hat{a}_i\hat{b}_j\right)^2}{\sum_i \hat{a}_i^2 \sum_j \hat{b}_j^2}, \qquad F_N = \frac{\mathrm{SS}_N}{(\mathrm{SS}_{\text{나머지}} - \mathrm{SS}_N)/\big((a-1)(b-1)-1\big)}$$

로 정의한다. SS_N은 가법 모형의 잔차를 $\hat{a}_i\hat{b}_j$에 회귀한 제곱합이다.

증명 없이 쓰는 사실 (Tukey (1949)) 가법 정규 모형 아래 $F_N \sim F\big(1, (a-1)(b-1)-1\big)$이다. $\hat{a}_i, \hat{b}_j$가 자료에서 추정된 것이어서 증명에는 그것들을 조건으로 놓는 논증이 필요하다. 이 검정은 효과가 더해지지 않고 곱해지는 경우(좋은 해에 좋은 거름의 효과가 더 큰 경우)를 잘 잡는다.

Broadbalk에서 $F_N = 73.7$ ($p = 3 \times 10^{-17}$)이다. 날씨가 좋은 해에는 거름을 많이 준 구획이 더 많이 늘었다. 로그 수확으로 바꾸면 $F_N = 17.7$로 줄지만 여전히 뚜렷하다. 효과가 곱해진다는 Fisher와 Mackenzie(1923)의 관찰이 이것이다.

3.2 분할구 설계: 오차가 두 층이다

예제 3.5 (Yates의 귀리 (1931)) 6개 블록마다 구획 3개(**주구**)에 귀리 품종 Golden Rain, Marvellous, Victory를 무작위로 배치하고, 각 주구를 다시 4개의 **세구**로 나눠 질소 네 수준을 무작위로 배치했다. 무작위화가 두 번, 두 크기의 단위에서 일어났다. 분산분석도 두 층으로 나뉜다.

층	출처	자유도	제곱합	평균제곱	F
주구	블록	5	15875.3	3175.1	
	품종	2	1786.4	893.2	1.49
	주구 오차 (블록×품종)	10	6013.3	601.3	
세구	질소	3	20020.5	6673.5	37.69
	품종×질소	6	321.7	53.6	0.30
	세구 오차	45	7968.8	177.1	

품종의 $F = 1.49$ ($p = 0.27$)는 주구 오차로, 질소와 상호작용의 F는 세구 오차로 나눈 것이다. 질소는 뚜렷하고($p < 10^{-11}$), 품종 차이와 상호작용은 우연의 크기 안이다(그림 3.2 오른쪽).

명제 3.6 (분할구의 기대평균제곱) 주구마다 독립인 효과 $\eta \sim \mathcal{N}(0, \sigma_w^2)$이 그 주구의 세구 4개에 똑같이 더해지고, 세구마다 독립인 $\varepsilon \sim \mathcal{N}(0, \sigma^2)$이 더해진다고 하자. 그러면

$$E\,\mathrm{MS}_{\text{주구 오차}} = \sigma^2 + 4\sigma_w^2, \qquad E\,\mathrm{MS}_{\text{세구 오차}} = \sigma^2$$

이다. 품종의 비교에는 η가 섞이므로 주구 오차로 나눠야 한다.

증명. **전략** 평균이 사라지는 공간 위의 제곱 길이의 기댓값은 $\mathrm{tr}(P\Sigma)$다(13회 보조정리 3.1). 오차의 공분산 Σ를 주구 사영 P_w로 쓰고, 두 오차 공간이 주구 안에서 상수인지, 주구 안에서 합이 0인지를 본다.

① $\boldsymbol{\eta}^*$를 η들을 세구 72개로 펼친 벡터라 하면 $\operatorname{Cov}\boldsymbol{\eta}^* = \sigma_w^2 K$이다. K는 같은 주구의 세구끼리 1인 블록대각 행렬이고, $K = 4P_w$이다. P_w는 주구 표시 벡터 18개가 생성하는 공간 위로의 사영이다. 주구마다 세구 4개

② $\boldsymbol{\eta}^*$와 $\boldsymbol{\varepsilon}$는 독립이므로 오차 $\boldsymbol{\eta}^* + \boldsymbol{\varepsilon}$의 공분산은 $\Sigma = 4\sigma_w^2 P_w + \sigma^2 I$이다. 독립합의 공분산

③ 주구 오차의 공간 $V_{\text{블록}\times\text{품종}}$(차원 10)은 주구 안에서 상수인 벡터들의 공간 안에 있으므로, 그 사영 P에 대해 $PP_w = P$이다. 블록 × 품종은 주구 단위의 대비

④ 평균 벡터는 P로 사라지므로

$$E\|P(\boldsymbol{\eta}^* + \boldsymbol{\varepsilon})\|^2 = \operatorname{tr}(P\Sigma) = 4\sigma_w^2 \operatorname{tr} P + \sigma^2 \operatorname{tr} P = 10(4\sigma_w^2 + \sigma^2). \quad \text{13회 보조정리 3.1, } \operatorname{tr} P = 10$$

⑤ 세구 오차의 공간(차원 45)은 주구 안에서 합이 0인 벡터들이어서 P_w와 직교한다. 그 사영 P'에 대해 $P'P_w = 0$이므로 $E\|P'(\boldsymbol{\eta}^* + \boldsymbol{\varepsilon})\|^2 = 45\sigma^2$이다. ④와 같은 계산

⑥ 각각을 자유도 10과 45로 나누면 명제의 식이다. 평균제곱의 정의

■

귀리에서 주구 오차의 평균제곱 601.3은 세구 오차 177.1의 3.4배다. 이웃한 세구들이 땅심을 함께 나누기 때문이다. 72구획을 모두 독립으로 보고 두 오차를 합친 잔차(평균제곱 254.2, 자유도 55)로 품종을 검정하면 F가 3.51 ($p = 0.037$)로 부풀어 "유의" 해진다(7절). 무작위화가 일어난 단위가 오차를 재는 단위다.

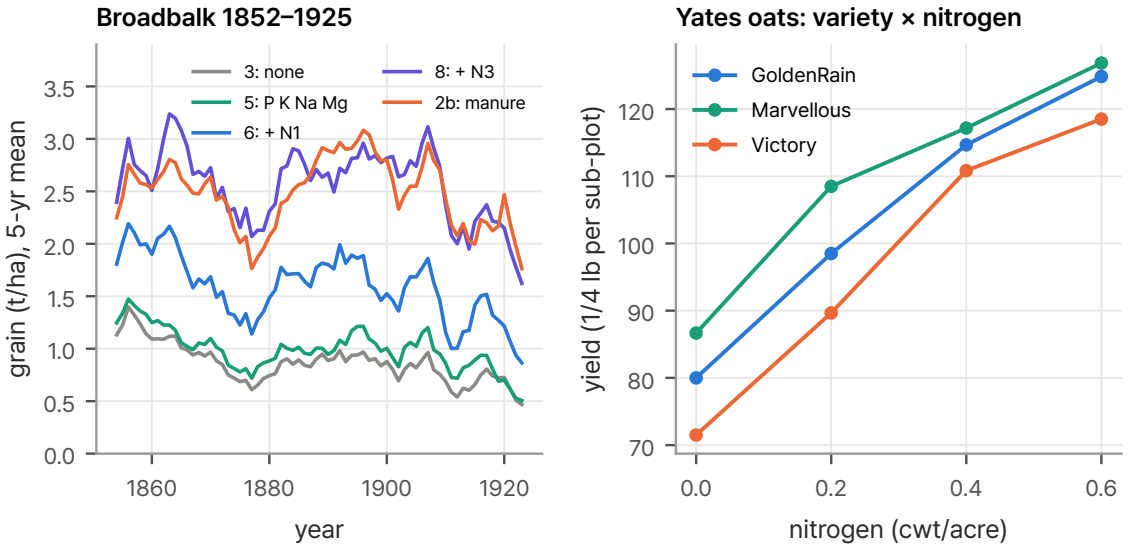


그림 3.2 왼쪽: Broadbalk의 다섯 구획의 수확(5년 이동평균). 질소 용량(6·7·8번)에 따른 차이와 해마다 함께 오르내리는 날씨의 몫이 보인다. 오른쪽: 귀리의 품종 × 질소 평균. 세 선이 거의 평행하다(상호작용 $F = 0.30$).

연습문제

연습 3.1 (무작위 블록의 표 ★) 처치 t개, 블록 b개의 무작위 블록 설계의 분산분석표(출처, 자유도, 제곱합의 식)를 쓰고, 블록을 무시한 일원분산분석과 오차 자유도가 어떻게 다른지 말하라. 풀이 42쪽

연습 3.2 (균형이 하는 일 ★★) 정리 3.2의 증명 ①에서 $\sum_{ijk} f(i)g(j) = r(\sum f)(\sum g)$가 성립하려면 칸마다 관측 수가 같아야 함을 보여라. 칸의 관측 수가 n_{ij}이면 V_A와 V_B가 직교할 조건은 무엇인가? ($n_{ij} = u_i v_j$ 꼴이면 충분하다.) 풀이 42쪽

연습 3.3 (Broadbalk의 비율 ★) 예제 3.3의 제곱합에서 세 몫의 비율과 두 F를 계산하라. 구획의 효과가 해의 효과보다 F는 크지만 제곱합은 작은 이유는? 풀이 43쪽

연습 3.4 (Tukey의 1자유도 ★★) 가법 모형의 잔차 $e_{ij} = y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}$를 $\hat{a}_i \hat{b}_j$에 (절편 없이) 회귀했을 때의 회귀제곱합이 정의 3.4의 SS_N 임을 보여라. ($\sum_{ij} e_{ij}\hat{a}_i\hat{b}_j = \sum_{ij} y_{ij}\hat{a}_i\hat{b}_j$를 먼저 보여라.) 풀이 43쪽

연습 3.5 (분할구의 품종 비교 ★★) 명제 3.6의 모형에서 한 품종의 평균(주구 6개, 세구 24개)의 분산이 $\sigma_w^2/6 + \sigma^2/24$이고, 두 품종의 차의 분산이 $(4\sigma_w^2 + \sigma^2)/12$임을 보여라. 이것이 주구 오차로 품종을 검정하는 이유다. 풀이 43쪽

4

여러 비교를 한꺼번에

> **역사** 분산분석의 F가 "처치들이 모두 같지는 않다"고 말하면, 다음 질문은 "어느 것이 어느 것과 다른가"다. 처치가 6개면 짝은 15개이고, 각각을 5%에서 검정하면 우연만으로도 어느 하나가 유의할 확률이 크다. Fisher(1935)는 F가 유의할 때만 짝 비교를 하라고 했다. 이탈리아의 Carlo Bonferroni(1936)는 확률의 합집합 부등식을 정리했다. 1953년 두 해법이 나란히 나왔다. John Tukey는 짝 비교 전체를 정확히 덮는 구간을 돌려 읽힌 원고로 내놓았고(정식 출판은 1994년), Henry Scheffé는 모든 대비를 동시에 덮는 구간을 *Biometrika*에 발표했다.

4.1 가족 단위 오류율과 Bonferroni

정의 4.1 (가족 단위 오류율) 가설 $H_{0,1}, \dots, H_{0,m}$을 함께 검정할 때, 참인 가설 가운데 하나라도 기각할 확률을 **가족 단위 오류율**(FWER)이라 한다. 신뢰구간 m개에서는 하나라도 참값을 놓칠 확률이다.

명제 4.2 (Bonferroni 부등식) 각 검정의 크기가 α/m이면 $\mathrm{FWER} \le \alpha$이다.

증명. A_k를 참인 k번째 가설을 기각하는 사건이라 하면 $P(\bigcup A_k) \le \sum P(A_k) \le m \cdot \alpha/m$이다(1회의 합집합 부등식). ∎

Bonferroni는 검정들 사이의 의존을 전혀 쓰지 않으므로 언제나 쓸 수 있지만, 검정이 많고 서로 상관되면 지나치게 보수적이다. 처치 평균들의 비교처럼 구조가 있으면 더 날카로운 방법이 있다.

4.2 Scheffé: 모든 대비를 동시에

보조정리 4.3 (Cauchy–Schwarz의 꼴) W가 양정치이면 $\sup_{\mathbf{a}\neq\mathbf{0}} \dfrac{(\mathbf{a}'\mathbf{u})^2}{\mathbf{a}'W\mathbf{a}} = \mathbf{u}'W^{-1}\mathbf{u}$이고, 최댓값은 $\mathbf{a} \propto W^{-1}\mathbf{u}$에서 이루어진다.

증명. $(\mathbf{a}'\mathbf{u})^2 = \big((W^{1/2}\mathbf{a})'(W^{-1/2}\mathbf{u})\big)^2 \le (\mathbf{a}'W\mathbf{a})(\mathbf{u}'W^{-1}\mathbf{u})$이고, 등호는 $W^{1/2}\mathbf{a} \propto W^{-1/2}\mathbf{u}$일 때다(Cauchy–Schwarz [T4]). 5회 연습 3.13과 같은 계산이다. ∎

정리 4.4 (Scheffé (1953)) 13회의 정규 선형모형에서 C를 계수 q의 $q \times p$ 행렬이라 하고 $\mathcal{L} = \{\mathbf{c} = C'\mathbf{a} : \mathbf{a} \in \mathbb{R}^q\}$라 하자. $s^2 = \mathrm{RSS}/(n-p)$이면

$$P\Big(\text{모든 } \mathbf{c} \in \mathcal{L}\text{에 대해 } |\mathbf{c}'\hat{\boldsymbol{\beta}} - \mathbf{c}'\boldsymbol{\beta}| \le \sqrt{qF_{q,n-p,\alpha}}\, s\sqrt{\mathbf{c}'(X'X)^{-1}\mathbf{c}}\Big) = 1-\alpha.$$

일원배치의 대비 $\sum c_i\mu_i$ ($\sum c_i = 0$)들에서는 $q = t-1$, $\mathbf{c}'(X'X)^{-1}\mathbf{c} = \sum c_i^2/n_i$이다.

증명. **전략** 무한히 많은 대비에 대한 부등식을 최댓값 하나에 대한 부등식으로 바꾼다. 그 최댓값은 Cauchy–Schwarz로 계산되고, 참값에서 잰 일반선형가설의 F 통계량이 된다.

① $\mathbf{u} = C(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$, $W = C(X'X)^{-1}C'$로 두면 $\mathbf{c} = C'\mathbf{a}$에 대해 $\mathbf{c}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \mathbf{a}'\mathbf{u}$, $\mathbf{c}'(X'X)^{-1}\mathbf{c} = \mathbf{a}'W\mathbf{a}$이다. 대입

② C의 계수가 q이고 $(X'X)^{-1}$이 양정치이므로 W는 양정치다. $\mathbf{a}'W\mathbf{a} = \|(X'X)^{-1/2}C'\mathbf{a}\|^2$

③ ①②에서

$$\sup_{\mathbf{c}\in\mathcal{L},\ \mathbf{c}\neq\mathbf{0}} \frac{\big(\mathbf{c}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\big)^2}{s^2\,\mathbf{c}'(X'X)^{-1}\mathbf{c}} = \frac{\mathbf{u}'W^{-1}\mathbf{u}}{s^2}.$$ 보조정리 4.3

④ $\mathbf{d} = C\boldsymbol{\beta}$로 두면 귀무가설이 참이므로, $\mathbf{u}'W^{-1}\mathbf{u}/(qs^2) \sim F(q, n-p)$이다. 13회 정리 4.9(c)의 증명

⑤ "모든 $\mathbf{c} \in \mathcal{L}$이 부등식을 만족한다"는 사건은 "③의 최댓값 $\le qF_{q,n-p,\alpha}$"와 같고, ④로 그 확률은 $1-\alpha$이다. 상한의 정의

⑥ 일원배치의 대비 공간은 $\mu_i - \mu_t$ $(i < t)$로 생성되므로 $q = t-1$이고 $(X'X)^{-1} = \mathrm{diag}(1/n_i)$이다. 정리 2.2의 증명 ①

■

Scheffé의 구간은 F 검정과 짝을 이룬다. F 검정이 H_0을 기각하는 것과 "어떤 대비의 Scheffé 구간이 0을 포함하지 않는 것"이 같은 사건이다. 대비가 자료를 본 뒤에 정해져도 보증이 깨지지 않는다.

4.3 Tukey: 짝 비교 전체를 정확히

정의 4.5 (스튜던트화 범위) $Z_1, \dots, Z_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$과 $V \sim \chi^2(\nu)$가 독립이라 하자.

$$Q = \frac{\max_i Z_i - \min_i Z_i}{\sqrt{V/\nu}}$$

의 분포를 **스튜던트화 범위** 분포라 하고, 위쪽 α 분위수를 $q_{t,\nu,\alpha}$로 쓴다.

증명 없이 쓰는 사실 (스튜던트화 범위의 분포) Q의 분포함수는 정규밀도와 χ 밀도를 포함하는 이중적분으로 쓸 수 있고, 수치적분으로 계산한다(Student 1927, Hartley 1942). 노트북은 `scipy.stats.studentized_range`를 쓴다.

정리 4.6 (Tukey (1953)) 일원배치에서 모든 $n_i = n$이고 s^2이 자유도 ν인 오차의 평균제곱이면

$$P\Big(\text{모든 } i<j \text{에 대해 } \big|(\bar{y}_{i\cdot}-\bar{y}_{j\cdot})-(\mu_i-\mu_j)\big| \le q_{t,\nu,\alpha}\frac{s}{\sqrt{n}}\Big) = 1-\alpha.$$

증명. **전략** 처치 평균을 표준화한 Z_i와 오차의 평균제곱으로 스튜던트화 범위를 만든다. 모든 짝의 차이에 대한 부등식은 최댓값 하나, 곧 범위에 대한 부등식이다. 단계별 증명은 연습 4.5(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 44쪽에 있다.

처치 평균들에 공통인 성분이 더해져도(분할구의 블록과 주구 효과처럼) 차이에서는 지워지므로, 세구 요인의 평균들에도 같은 정리가 쓰인다. 짝마다의 구간 폭 $q\,s/\sqrt{n}$을 **정직한 유의차**(HSD)라 부른다.

예제 4.7 (질소 수준의 비교) 귀리의 질소 평균은 79.4, 98.9, 114.2, 123.4이고(수준마다 18세구), 세구 오차는 $s^2 = 177.1$, $\nu = 45$다. $q_{4,45,0.05} = 3.773$이므로

$$\text{HSD} = 3.773\sqrt{177.1/18} = 11.83.$$

여섯 짝의 차 19.5, 34.8, 44.0, 15.3, 24.5, 9.2 가운데 0.4와 0.6 cwt의 차 9.2만 HSD보다 작다. 질소를 더 줄수록 수확이 늘지만 0.4를 넘으면 늘어나는 폭이 우연과 가려지지 않는다. 짝 하나의 t 통계량으로 바꾸면 기준값은 Tukey $q/\sqrt{2} = 2.668$, Bonferroni(6짝) $t_{45,0.05/12} = 2.760$, Scheffé $\sqrt{3F_{3,45,0.05}} = 2.904$로, 짝 비교만 할 때는 Tukey가 가장 좁다. 선형 추세 대비 $(-3,-1,1,3)$은 추정값 147.3, 표준오차 14.0, $t = 10.5$로 Scheffé의 기준을 훨씬 넘는다.

예제 4.8 (우연만으로 몇 번 틀리나) 처치 6개, 처치마다 5개, 모든 평균이 같은 세계를 4만 번 만들고 15쌍을 비교했다($\nu = 24$). 짝마다 5% t 검정을 하면 하나라도 "유의"한 비율이 0.34다. Bonferroni는 0.034, Tukey는 0.049, Scheffé는 0.015다(그림 4.1 오른쪽). Tukey는 정확히 5%이고, Bonferroni와 Scheffé는 짝 비교에는 보수적이다. Scheffé는 짝뿐 아니라 무한히 많은 대비를 함께 덮기 때문이다.

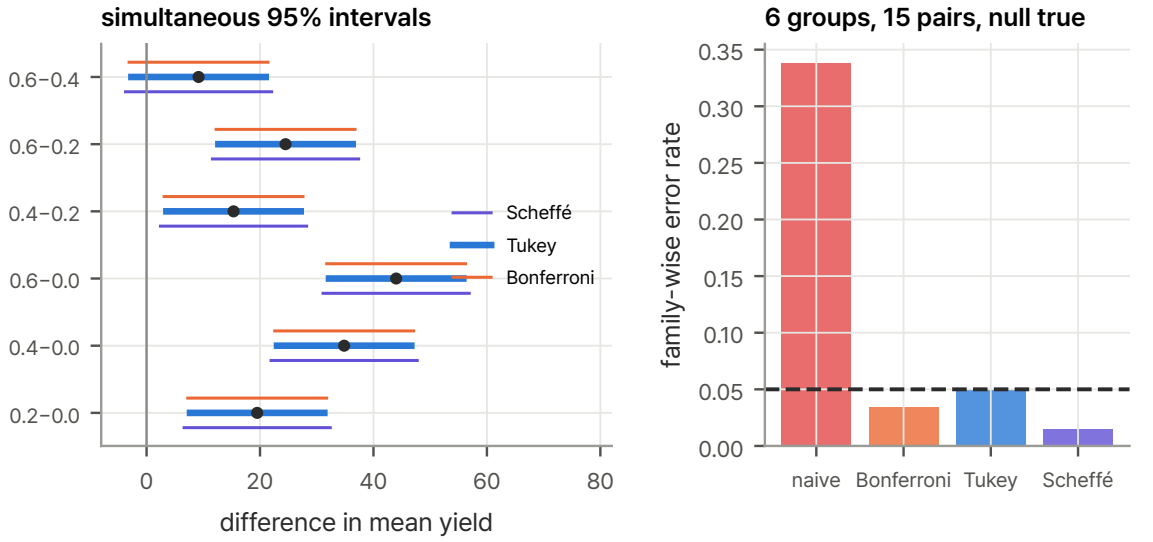


그림 4.1 왼쪽: 귀리 질소 수준의 여섯 짝 차이와 동시 95% 구간. 같은 중심에 Tukey(굵게), Bonferroni, Scheffé의 폭을 겹쳤다. 0.4–0.6 cwt의 구간만 0을 포함한다. 오른쪽: 평균이 모두 같은 6처치에서 15쌍을 비교할 때의 가족 단위 오류율(4만 번).

연습문제

연습 4.1 (독립이라면 ★) 독립인 검정 15개를 각각 5%로 하면 FWER가 $1-0.95^{15}=0.54$임을 보여라. 예제 4.8의 0.34가 이보다 작은 이유는? 풀이 43쪽

연습 4.2 (Scheffé는 t보다 넓다 ★★) 대비 하나만 볼 때의 t 구간 기준값 $t_{\nu,\alpha/2}$보다 Scheffé의 기준값 $\sqrt{qF_{q,\nu,\alpha}}$가 크거나 같음을, 정리 4.4의 사건이 대비 하나의 t 사건을 포함한다는 사실로 보여라. $q=1$이면 같다. 풀이 43쪽

연습 4.3 (HSD 계산 ★) 예제 4.7의 HSD와 여섯 짝의 판정을 다시 계산하라. 질소 0과 0.2의 차이만 관심이 있었다면 어떤 구간을 쓰는가? 풀이 43쪽

연습 4.4 (추세 대비 ★) 대비 $(-3,-1,1,3)$의 추정값과 표준오차 $s\sqrt{\sum c_i^2/n}$을 계산하라. 이 대비를 자료를 보고 골랐다면 어느 기준값을 써야 하는가? 풀이 44쪽

연습 4.5 (Tukey의 동시구간: 증명 ★ 유도형) 정리 4.6를 다음 순서로 증명하라.

(a) $Z_i=(\bar y_{i\cdot}-\mu_i)\sqrt{n}/\sigma$ $(i=1,\dots,t)$가 독립인 $\mathcal{N}(0,1)$임을 보여라.

(b) $V=\nu s^2/\sigma^2\sim\chi^2(\nu)$가 $Z_1,\dots,Z_t$와 독립임을 정리 2.2의 증명으로 확인하라. ($\bar y_{i\cdot}$은 $H\mathbf{y}$의 성분이고 s^2은 $(I-H)\mathbf{y}$의 함수다.)

(c) 모든 $i<j$에 대해 $|Z_i-Z_j|\le c$인 것과 $\max_i Z_i-\min_i Z_i\le c$인 것이 같은 사건임을 보여라.

(d) (a)–(c)와 정의 4.5로 정리의 확률이 $1-\alpha$임을 결론지어라.

풀이 44쪽

5

설계의 원리 — 반복, 무작위화, 국소 관리

> **역사** Fisher는 1926년 **밭 실험의 배치**에서 설계의 원리 셋을 내놓았다. 오차를 재기 위한 **반복**, 그 측정을 정당하게 만드는 **무작위화**, 오차를 줄이는 **국소 관리**(블록)다. 같은 글에서 그는 한 번에 한 요인만 바꾸라는 오랜 격언을 정면으로 반박했다. "자연은 논리적으로 신중하게 짠 질문지에 가장 잘 답한다. 질문을 하나만 던지면, 다른 주제를 먼저 다룰 때까지 대답을 거부하는 일이 많다." Student(Gosset)는 마지막까지 체계적 배치를 옹호했다. 처치를 번갈아 배치하면 땅의 경향이 고르게 나뉜다는 것이다. 두 사람의 논쟁은 Gosset가 죽은 1937년까지 이어졌다.

5.1 국소 관리: 짝과 블록

정리 5.1 (짝 설계의 효율) $2m$개 단위가 m개 짝 (a_i, b_i)로 나뉘어 있고, 처치효과가 모든 단위에서 τ로 같다고 하자(a_i, b_i는 대조 아래의 잠재적 결과).

(a) 완전무작위설계(m개를 처치)의 $\hat{\tau}$는 불편이고 $\operatorname{Var}\hat{\tau} = 2S^2/m$이다. S^2은 $2m$개 값의 분산(분모 $2m-1$)이다.

(b) 짝마다 동전으로 한쪽을 처치하는 설계의 $\hat{\tau}$(짝 차이의 평균)는 불편이고 $\operatorname{Var}\hat{\tau} = 2S_w^2/m$이다. $S_w^2 = \frac{1}{2m}\sum_i (a_i - b_i)^2$은 짝 안의 분산이다.

따라서 짝 설계의 상대효율은 S^2/S_w^2이고, 짝 안의 상관을 ρ라 하면 대략 $1/(1-\rho)$이다.

증명. **전략** (a)는 정리 1.4에 효과가 일정하다는 조건을 넣는다. (b)는 짝마다의 동전을 ± 1의 독립 변수로 쓰면 $\hat{\tau} - \tau$가 독립인 항의 평균이 된다. 단계별 증명은 연습 5.5(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 45쪽에 있다.

예제 5.2 (Mercer–Hall의 짝과 블록) 균일성 시험의 각 행에서 이웃한 두 열(1–2, 3–4, ⋯, 23–24)을 짝으로 묶으면 짝이 240개, 짝 안의 상관은 0.33이다. $S^2 = 0.2123$, $S_w^2 = 0.1433$이므로 짝 설계의 상대효율은 1.48이고, $1/(1-\rho) = 1.49$와 거의 같다. 짝으로 묶은 실험 하나는 무작위 배치 실험 1.48개의 값이 있다. 무작위 배정을 되풀이해 잰 분산의 비도 1.5 근처다(노트북). 이웃한 다섯 구획을 블록으로 삼아 가짜 처치 5개를 블록마다 무작위 배치하면 오차의 평균제곱이 0.210(완전무작위)에서 0.167(블록)로 줄어 상대효율이 1.26이다. 두 설계 모두 무작위화가 있으므로 5% 검정의 실제 크기는 0.05다.

블록은 공짜가 아니다. 오차의 자유도를 블록 수만큼 쓴다. 블록 안의 단위들이 서로 닮지 않았다면 자유도만 잃는다. Mercer–Hall의 밭에서는 이웃한 구획이 닮았으므로 블록이 이득이었다. 2절의 그림 1.1의 왼쪽이 그 이유를 보여 준다.

5.2 요인설계: 한 번에 여러 질문을

정리 5.3 (요인설계의 효율) 요인 k개가 각각 두 수준이고 단위가 N개, 오차의 분산이 σ^2이며 효과가 더해진다고 하자.

(a) 2^k개 조합에 $N/2^k$개씩 배치하는 **요인설계**에서, 요인 하나의 주효과(높은 수준의 평균 − 낮은 수준의 평균)의 분산은 $4\sigma^2/N$이다.

(b) 기준 조합 하나와 요인을 하나씩만 바꾼 조합 k개에 $N/(k+1)$개씩 배치하는 **한 번에 한 요인**(OFAT) 설계에서, 주효과(바꾼 조합 − 기준 조합)의 분산은 $2(k+1)\sigma^2/N$이다.

따라서 같은 N으로 요인설계의 효율은 $(k+1)/2$배이고, 요인설계만 상호작용을 추정할 수 있다.

증명. **전략** 두 추정량 모두 독립인 두 평균의 차이므로, 각 평균에 들어가는 단위의 수만 세면 분산이 나온다. 요인설계에서는 모든 단위가 높은 수준 쪽이나 낮은 수준 쪽에 들어간다. 단계별 증명은 연습 5.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 45쪽에 있다.

요인설계에서는 모든 단위가 모든 주효과의 추정에 쓰인다. Fisher가 "숨은 반복"이라 부른 것이다. 요인이 셋이면 효율이 2배, 다섯이면 3배다.

예제 5.4 (npk: 완두의 질소·인·칼륨) 교과서에 자주 나오는 완두 실험(R의 `npk` 자료)은 질소(N), 인(P), 칼륨(K)을 각각 주거나 주지 않은 2^3 요인설계를 6개 블록, 24구획에 배치했다. 블록 하나에 구획이 4개뿐이어서 8조합을 다 넣을 수 없다. 그래서 블록마다 8조합의 절반만 넣되, NPK 삼중 상호작용의 부호가 같은 4조합끼리 묶었다. 그 대가로 NPK 상호작용은

블록 차이와 구별되지 않는다(**교락**). 주효과는(구획당 파운드)

$$\text{N}: \ +5.62 \ (F = 12.26, \ p = 0.004), \qquad \text{K}: \ -3.98 \ (F = 6.17, \ p = 0.029),$$
$$\text{P}: \ -1.18 \ (F = 0.54, \ p = 0.47)$$

이다. 오차는 블록과 주효과·이중 상호작용을 뺀 자유도 12, 평균제곱 15.44다. 24구획 모두가 세 주효과의 추정에 쓰였다.

연습문제

연습 5.1 (짝 안의 분산 ★) 짝의 두 값 a, b가 같은 분산 σ^2과 상관 ρ를 가지면 $E(a-b)^2 = 2\sigma^2(1-\rho)$임을 보이고, 정리 5.1의 효율 $1/(1-\rho)$를 얻어라. $\rho < 0$이면 어떻게 되는가? 풀이 44쪽

연습 5.2 (세 요인의 요인설계 ★★) $k = 3$, $N = 24$에서 정리 5.3의 두 분산을 비교하라. 요인설계로 NP 상호작용 $\frac{1}{2}[(\text{N과 P 둘 다}) - (\text{N만}) - (\text{P만}) + (\text{둘 다 없음})]$을 추정할 때의 분산은? 풀이 44쪽

연습 5.3 (교락 ★) npk 실험에서 블록 1, 5, 6은 NPK의 부호가 -1인 4조합, 블록 2, 3, 4는 $+1$인 4조합을 받았다. NPK 상호작용의 추정량이 "블록 2·3·4의 평균 − 블록 1·5·6의 평균"과 같음을 보이고, 왜 블록 차이와 구별할 수 없는지 설명하라. 풀이 44쪽

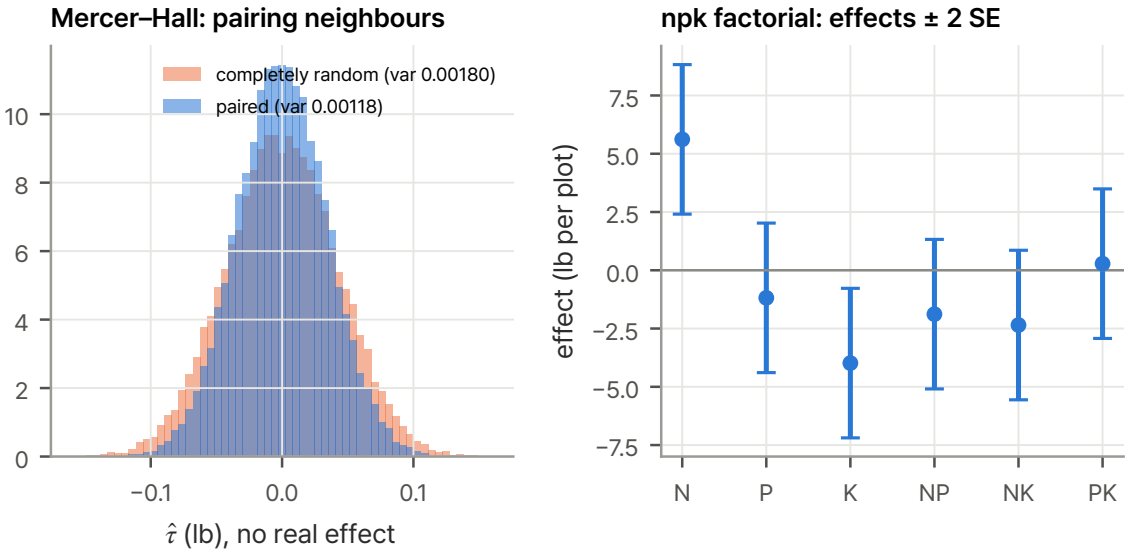


그림 5.1 왼쪽: Mercer–Hall의 480구획(240짝)에서 가짜 처치효과 $\hat{\tau}$의 무작위화 분포. 짝 안에서 무작위화한 설계(파랑)가 완전무작위설계(주황)보다 좁다(정확한 분산비 1.48). 오른쪽: npk 실험의 주효과와 이중 상호작용 추정값과 ±2 표준오차.

연습 5.4 (OFAT가 놓치는 것 ★★) 두 요인 A, B의 평균 수확이 (A 없음, B 없음) 10, (A만) 10, (B만) 10, (둘 다) 16이라 하자. 기준 조합에서 하나씩만 바꾸는 실험은 어떤 결론을 내리는가? 요인설계는 무엇을 보는가?

풀이 44쪽

연습 5.5 (짝 설계의 효율: 증명 ★★ 유도형) 정리 5.1를 다음 순서로 증명하라.

(a) 효과가 모든 단위에서 τ이면 $S_1^2 = S_0^2 = S^2$, $S_\tau^2 = 0$임을 보이고, 정리 1.4에 $n_T = n_C = m$을 넣어 (a)를 얻어라.

(b) 짝 i의 동전을 $\sigma_i = \pm 1$(처치가 a쪽이면 $+1$)로 두면, 짝 안의 관측 차이(처치 − 대조)가 $\tau + \sigma_i(a_i - b_i)$임을 보여라.

(c) σ_i들이 독립이고 평균 0, 분산 1임을 써서 $\hat{\tau} = \tau + \frac{1}{m}\sum_i \sigma_i(a_i - b_i)$의 평균과 분산 $2S_w^2/m$을 구하라.

(d) 짝의 두 값이 분산 S^2, 상관 ρ를 가지면 $S_w^2 \approx S^2(1-\rho)$임을 보이고 상대효율 $1/(1-\rho)$를 얻어라.

풀이 45쪽

연습 5.6 (요인설계의 효율: 증명 ★ 유도형) 정리 5.3을 다음 순서로 증명하라.

(a) 요인설계에서 한 요인의 높은 수준과 낮은 수준에 각각 $N/2$개의 단위가 있음을 보이고, 주효과 추정량의 분산 $4\sigma^2/N$을 구하라.

(b) 가법 모형에서 다른 요인들의 효과가 두 평균의 차에서 지워짐을 보여라. 다른 요인의 각 수준이 두 쪽에 몇 번씩 나오는지 센다.

(c) OFAT에서 주효과 추정량이 단위 $N/(k+1)$개짜리 두 평균의 차임을 보이고 분산 $2(k+1)\sigma^2/N$을 구하라.

(d) 두 분산의 비가 $(k+1)/2$임을 확인하고, OFAT로는 상호작용을 추정할 수 없는 이유를 설명하라.

풀이 45쪽

6

원자료로 — 1954년 Salk 백신 시험

> **역사** 1950년대 초 미국에서 소아마비는 해마다 수만 명의 아이를 덮쳤다. 1954년 봄, 소아마비 재단은 Jonas Salk의 사백신을 시험하기로 하고 미시간 대학의 Thomas Francis Jr.에게 평가를 맡겼다. 시험은 두 설계로 나뉘었다. 많은 지역에서는 부모가 동의한 2학년을 접종하고 접종하지 않은 1·3학년을 대조로 삼았다(**관찰 대조**). 다른 지역에서는 동의한 아이들을 무작위로 백신과 위약(소금물 주사)에 배정하고, 아이도 의사도 평가자도 누가 무엇을 맞았는지 모르게 했다(**무작위 이중맹검**). 1955년 4월 12일 Francis는 백신이 효과적이고 안전하다고 발표했다. 180만 명이 넘는 아이가 참가한, 그때까지 가장 큰 의학 실험이었다.

설계	집단	인원	소아마비 (10만 명당)	마비형 (10만 명당)
무작위 이중맹검	백신	200,745	28.4	16.4
	위약	201,229	70.6	57.1
	동의하지 않음	338,778	46.3	35.7
관찰 대조	백신 (2학년, 동의)	221,998	25.2	17.1
	대조 (1·3학년)	725,173	53.9	45.5
	동의하지 않은 2학년	123,605	43.7	34.8

자료는 Francis 등(1955)의 보고서에서 왔다. “소아마비”는 마비형과 비마비형을 합친 것이다. 진단이 확실한 마비형으로 효과를 잰다.

명제 6.1 (상대위험과 그 신뢰구간) 두 집단에서 n_1, n_2명 가운데 a, c명이 병에 걸렸을 때 상대위험 $\widehat{\mathrm{RR}} = \frac{a/n_1}{c/n_2}$의 로그는 근사적으로 정규이고

$$\mathrm{SE}\big(\log \widehat{\mathrm{RR}}\big) \approx \sqrt{\frac{1}{a} - \frac{1}{n_1} + \frac{1}{c} - \frac{1}{n_2}}$$

이다. 백신 효과는 $\mathrm{VE} = 1 - \mathrm{RR}$로 쓴다.

증명. $\hat p = a/n_1$은 근사적으로 $\mathcal{N}(p, p(1-p)/n_1)$이고, 델타 방법(7회 정리 4.4)으로 $\log \hat p$의 분산은 $\frac{p(1-p)}{n_1 p^2} = \frac{1-p}{n_1 p} \approx \frac{1}{a} - \frac{1}{n_1}$이다. 두 집단은 독립이므로 분산을 더한다. ∎

예제 6.2 (무작위 시험의 결론) 무작위 이중맹검에서 마비형의 상대위험은 $\frac{33/200745}{115/201229} = 0.288$, 95% 신뢰구간은 $[0.195,\ 0.424]$이다. 백신은 마비형 소아마비를 약 71% 줄였다. 두 비율의 차의 z 통계량은 -6.73 ($\chi^2 = 45.3$, $p = 1.7 \times 10^{-11}$)이다(8회 따름정리 5.2). 관찰 대조의 같은 계산은 상대위험 0.376, 백신 효과 62%다.

두 설계의 백신 효과가 왜 다른가? 표의 셋째 줄이 답한다.

예제 6.3 (동의는 무작위가 아니다) 무작위 시험에서 동의하지 않은 아이들의 마비형 비율(10만 명당 35.7)은 같은 지역의 위약군(57.1)보다 낮다. 상대위험 57.1/35.7 = 1.60의 95% 구간은 $[1.24,\ 2.07]$이다. 위약군과 동의하지 않은 아이들은 둘 다 백신을 맞지 않았으므로 차이는 "누가 동의하는가"에서 온다. 흔히 드는 설명(Meier 1972)은 이렇다. 동의한 가정은 소득과 교육 수준이 높았고, 위생이 좋은 환경의 아이들은 어릴 때 약한 감염으로 면역을 얻을 기회가 적어 오히려 소아마비에 더 취약했다. 관찰 대조 설계의 대조군(1·3학년 전체)에는 동의할 가정과 하지 않을 가정의 아이가 섞여 있으므로, 동의한 2학년과 비교하면 백신의 효과가 작게 나온다. 학년(나이)이 다르다는 것, 의사가 누가 접종했는지 안다는 것도 비교를 흐린다.

무작위화는 동의한 아이들 **안에서** 백신과 위약을 정했다. 그래서 위약군은 백신군과 동의 여부, 소득, 나이, 진단하는 의사의 기대까지 모든 면에서 같은 분포를 갖는다. 관찰 대조 설계는 그 가운데 어느 것도 보증하지 않는다. 13회 7절의 Yule–Simpson 역설이 분석에서 고칠 수 없는 문제였다면, 무작위화는 그것을 설계에서 없앤다(그림 6.1).

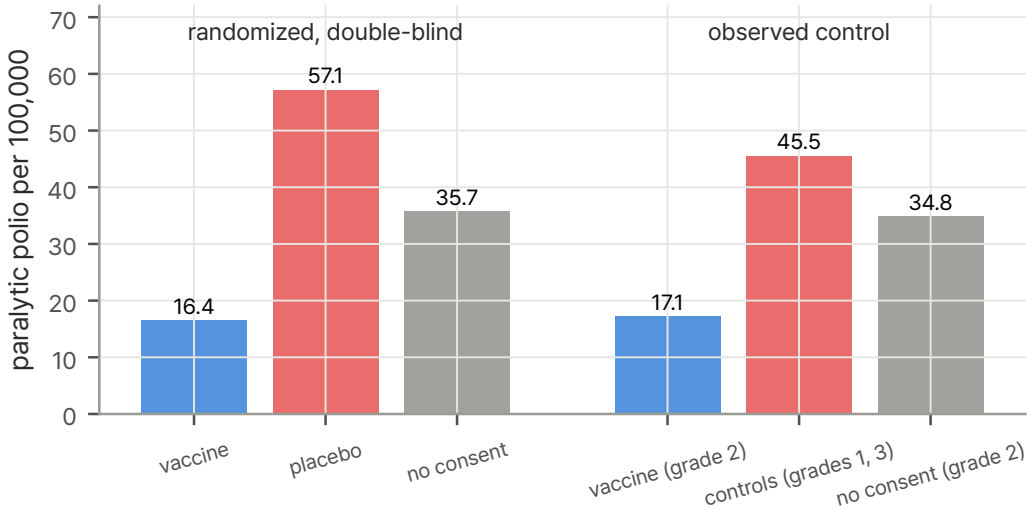


그림 6.1 Salk 시험의 마비형 소아마비 비율(10만 명당). 무작위 시험(왼쪽 묶음)에서 백신군과 위약군의 차이가 관찰 대조(오른쪽 묶음)보다 크다. 두 설계 모두 동의하지 않은 아이들의 비율이 대조군보다 낮다.

연습문제

연습 6.1 (신뢰구간 ★) 명제 6.1로 무작위 시험의 상대위험 0.288의 95% 구간 [0.195, 0.424]를 다시 계산하고, 백신 효과의 구간으로 바꿔라. 풀이 45쪽

연습 6.2 ($z^2 = \chi^2$ ★) 2×2 표에서 합동 비율로 만든 두 비율 차의 z 통계량의 제곱이 Pearson의 χ^2(연속성 수정 없음)과 같음을 보이고, 예제 6.2의 $(-6.73)^2 = 45.3$을 확인하라. 풀이 45쪽

연습 6.3 (이중맹검 ★) 마비형과 비마비형을 합친 비율로 계산하면 무작위 시험의 상대위험은 28.4/70.6이다. 비마비형만으로 계산하면 어떻게 되는가? 비마비형의 진단에 의사의 기대가 섞일 수 있다면, 이중맹검이 없는 관찰 대조 설계에서 어느 방향으로 치우칠 것인가? 풀이 45쪽

7

부숴보기 — 설계가 깨질 때

7.1 무작위화를 빼면: 땅심이 효과로 둔갑한다

예제 7.1 (체계적 배치) Mercer–Hall의 균일성 시험에는 처치가 없다. 그런데 열 14–25(12열, 240구획)를 "비료", 열 1–13(260구획)을 "대조"라 부르고 두 표본 t 검정을 하면, 차이는 -0.33파운드, $t = -8.55$, $p = 1.5 \times 10^{-16}$이다. 존재하지 않는 효과가 압도적으로 "유의"하다. 밭의 한쪽이 원래 덜 비옥했기 때문이다(그림 1.1 왼쪽). 같은 밭을 위아래로 나누면 $t = -0.77$로 아무 일도 없다. 체계적 배치가 괜찮은지는 땅심의 경향이 어느 방향인지에 달려 있고, 실험자는 그것을 모른다. 250구획을 무작위로 골라 같은 검정을 2만 번 하면 $|t| > 1.96$의 비율이 0.050이다. 정리 1.7가 약속한 대로다.

무작위화하지 않은 비교에서는 처치와 함께 움직이는 모든 것이 처치의 효과로 섞인다. 13회 명제 5.6의 누락변수 치우침이 땅심이라는 보이지 않는 변수로 나타난 것이다. 무작위 배정은 처치를 땅심과 독립으로 만들어 그 치우침의 기댓값을 0으로 만든다.

7.2 오차층을 틀리면: 의사반복

예제 7.2 (귀리의 품종을 세구로 검정하면) 예제 3.5의 품종은 주구 18개에 무작위로 배치되었다. 72세구를 모두 독립인 단위로 보고 블록, 품종, 질소, 상호작용을 뺀 잔차(평균제곱 254.2, 자유도 55)로 품종을 검정하면 $F = 3.51$, $p = 0.037$이다. 올바른 주구 오차(601.3, 자유도 10)로는 $F = 1.49$, $p = 0.27$이다. 품종의 비교에 쓸 수 있는 독립 단위는 18개 주구뿐인데, 세구를 반복으로 세면 반복 수를 네 배로 부풀리게 된다. Hurlbert(1984)는 생태학 논문의 상당수가 이런 **의사반복**을 저지른다고 보고했다.

7.3 균형을 깨면: 제곱합이 순서에 따라 달라진다

정리 3.2은 균형에 기댔다. 칸마다 관측 수가 다르면 V_A와 V_B가 직교하지 않고, "A의 제곱합"이 하나로 정해지지 않는다.

명제 7.3 (순차 제곱합) $\mathrm{col}(X_0) \subset \mathrm{col}(X_0, X_A)$ 등으로 모형을 키워 갈 때, A를 B보다 먼저 넣은 제곱합 $\mathrm{SS}(A \mid X_0) = \|(H_{0A} - H_0)\mathbf{y}\|^2$과 B 뒤에 넣은 $\mathrm{SS}(A \mid X_0, B) = \|(H_{0AB} - H_{0B})\mathbf{y}\|^2$은 일반적으로 다르다. A와 B의 공간이 X_0를 뺀 뒤 서로 직교하면 둘은 같다.

증명. 직교하면 $H_{0AB} = H_0 + (H_{0A} - H_0) + (H_{0B} - H_0)$이고 $H_{0AB} - H_{0B} = H_{0A} - H_0$이다. 직교하지 않으면 13회 정리 5.3(FWL)으로 $\mathrm{SS}(A \mid X_0, B)$는 A를 B에 대해 잔차화한 뒤의 제곱합이므로 다르다. ∎

예제 7.4 (귀리 7구획을 잃으면) 귀리 자료에서 세구 7개(2, 31, 45, 50, 54, 58, 61번 행)를 빼면 균형이 깨진다. 블록을 먼저 넣은 뒤의 품종의 제곱합은 질소보다 먼저 넣으면 1026.3, 질소 뒤에 넣으면 1742.4다. 질소는 15044.4와 15760.5다. 통계 소프트웨어의 "Type I·II·III 제곱합"은 이 순서 문제에 대한 서로 다른 관행이다. 균형 설계는 이 문제를 처음부터 없앤다.

7.4 분산이 다르면: 작은 집단이 시끄러울 때

예제 7.5 (Box (1954)의 경고) 세 집단의 크기가 5, 10, 20이고 표준편차가 3, 2, 1이며 평균은 모두 같다고 하자. 5% F 검정의 실제 크기는 0.20이다(4만 번 반복). 표준편차를 1, 2, 3으로 뒤집으면(큰 집단이 시끄러우면) 0.014가 된다. 합동 분산 $\mathrm{SS}_W/(N-t)$가 큰 집단의 분산에 끌려가기 때문이다. 집단 크기가 같으면 이 효과는 훨씬 작다. 균형 설계는 여기서도 보험이다.

연습문제

연습 7.1 (가짜 효과의 크기 ★) 예제 7.1의 $t = -8.55$를 정리 1.4의 무작위화 표준편차 $\sqrt{S^2(1/240 + 1/260)}$와 비교하라. 차이 -0.33은 무작위 배치였다면 몇 표준편차인가? 풀이 46쪽

연습 7.2 (의사반복의 자유도 ★) 품종 비교에 쓸 수 있는 독립 단위가 18개이고 블록이 6개, 품종이 3개라면 오차 자유도가 10임을 보여라. 72세구를 단위로 세면 자유도가 몇으로 부풀었는가? 풀이 46쪽

연습 7.3 (작은 불균형 ★★) 2×2 배치에서 칸의 관측 수가 $\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$이고 칸 평균이 모두 같지 않을 때, 명제 7.3의 두 제곱합이 다를 수 있음을 숫자 예로 보여라. 풀이 46쪽

연습 7.4 (합동 분산의 치우침 ★★) 예제 7.5에서 $E\,\mathrm{SS}_W/(N-t)$와 $E\,\mathrm{SS}_B/(t-1)$을 13회 보조정리 3.1로 계산하라. 둘의 비가 1보다 크면 F 검정이 너무 자주 기각한다. 풀이 46쪽

8

오늘의 실험설계 — 하이퍼파라미터와 A/B 테스트

8.1 Ablation과 그리드: 상호작용을 보는가

기계학습 논문의 **ablation** 표는 완성된 모델에서 부품을 하나씩 빼며 성능을 잰다. 하이퍼파라미터 조정도 흔히 하나씩 한다. 학습률을 고정된 배치 크기에서 고른 뒤, 그 학습률에서 배치 크기를 고르는 식이다. 둘 다 Fisher가 반박한 "한 번에 한 요인"이다. 요인 사이에 상호작용이 있으면 이 방법은 최적점을 놓친다.

> **예제 8.1** (학습률 × 배치 크기) 학습률 5단계(10^{-4}부터 10^{-2}까지)와 배치 크기 5단계(32부터 512까지)의 검증 정확도가, 최적 학습률이 배치 크기에 비례하는(선형 스케일링 규칙) 가상의 곡면을 따른다고 하자. 배치 32에서 학습률을 고르면 3×10^{-4}이고, 그 학습률에서 배치를 고르면 64다. 정확도는 0.9155다. 5×5 전체 그리드를 보면 최적은 학습률 10^{-3}, 배치 128의 0.920이다. 조합마다 씨앗 3개로 반복한 그리드를 이원분산분석하면 상호작용의 F가 97이다(그림 8.1 왼쪽). 그리드 탐색은 요인설계이고, 분산분석표는 어느 하이퍼파라미터가 얼마나, 그리고 서로 얽혀서 성능을 움직이는지 말해 준다. 씨앗을 반복하는 것은 Fisher의 반복이다.

8.2 A/B 테스트와 CUPED

온라인 서비스의 A/B 테스트는 사용자를 무작위로 두 판에 배정하는 완전무작위설계다. 1절의 모든 정리가 그대로 쓰인다. 사용자가 수백만이어도 효과가 작아서 분산을 줄이는 일이 중요하다. 블록의 연속판이 공변량 보정이다. Fisher(1932)는 이것을 **공분산분석**이라 불렀고, Deng 등(2013)은 실험 전 기간의 지표를 공변량으로 쓰는 방법을 CUPED라는 이름으로 퍼뜨렸다.

정리 8.2 (공변량 보정의 분산) 무작위 배정 전에 정해진 공변량 X(처치의 영향을 받지 않음)가 있고, 단위들이 (X, Y)의 분포에서 독립으로 뽑혔다고 하자. $\theta = \operatorname{Cov}(X,Y)/\operatorname{Var} X$, $\rho = \operatorname{Corr}(X,Y)$(대조군 기준, 처치효과는 상수)일 때

$$\hat{\tau}_\theta = (\bar{Y}_T - \theta \bar{X}_T) - (\bar{Y}_C - \theta \bar{X}_C)$$

는 불편이고 $\operatorname{Var} \hat{\tau}_\theta = (1-\rho^2)\operatorname{Var}(\bar{Y}_T - \bar{Y}_C)$이다.

증명. **전략** $\hat{\tau}_\theta$는 새 반응 $Y - \theta X$의 두 군 평균의 차다. 불편성은 무작위 배정이 두 군의 X 분포를 같게 만든다는 데서 나오고, 분산은 $\operatorname{Var}(Y-\theta X)$를 θ의 이차식으로 펼쳐 얻는다. 단계별 증명은 연습 8.4(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 47쪽에 있다.

θ는 Y를 X로 맞춘 최선의 선형예측의 기울기(13회 정리 5.1)이고, 보정은 그 잔차로 비교하는 것이다.

실제로는 θ를 자료에서 추정하지만, 표본이 크면 분산은 같다. 실험 전 지표와 실험 중 지표의 상관이 0.7이면 분산이 0.51배, 곧 같은 정밀도에 필요한 사용자가 절반이다(노트북의 2만 번 반복에서 0.52). 무작위화가 치우침을 없앤다면, 블록과 공변량 보정은 분산을 줄인다. 둘은 Fisher의 세 원리 가운데 둘이다.

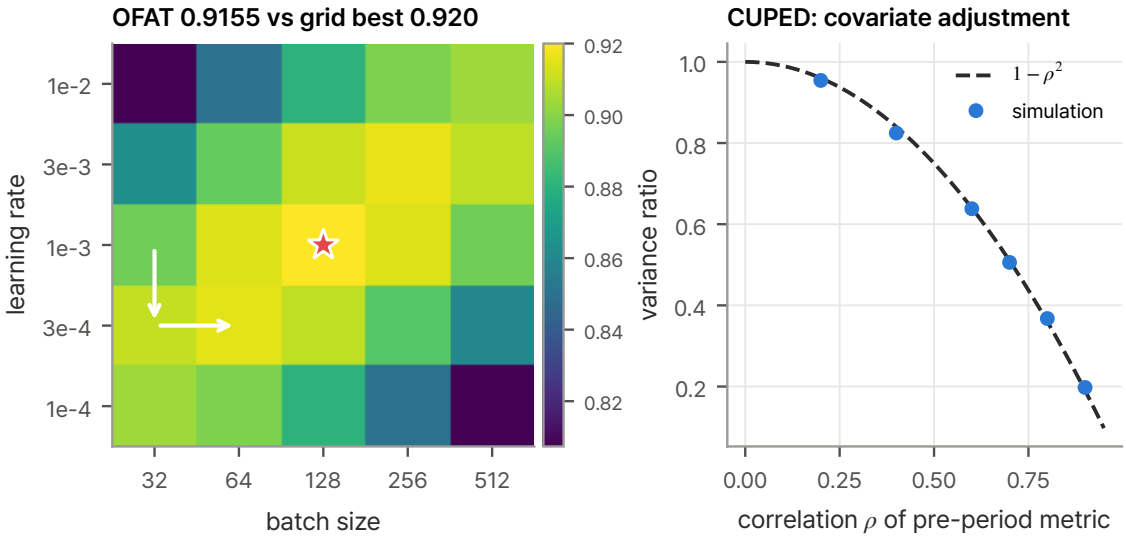


그림 8.1 왼쪽: 학습률 × 배치 크기의 가상 정확도 곡면. 한 번에 한 요인(OFAT, 흰 화살표)은 $(3\times10^{-4}, 64)$에 멈추고, 그리드의 최적은 $(10^{-3}, 128)$이다(별). 오른쪽: 공변량 보정의 분산비. 실선은 이론값 $1-\rho^2$, 점은 시뮬레이션이다.

연습문제

연습 8.1 (추정한 θ ★★) θ 대신 두 군을 합친 자료의 회귀 기울기 $\hat\theta$를 쓰면 $\hat\tau_{\hat\theta}$가 13회의 다중회귀 $Y \sim 1 + \text{처치} + X$에서 처치의 계수와 거의 같음을 FWL(13회 정리 5.3)로 설명하라.

풀이 46쪽

연습 8.2 (몇 명이 필요한가 ★) 효과 0.01, 지표의 표준편차 1인 A/B 테스트에서 양쪽 5%, 검정력 80%에 군마다 필요한 사용자 수는 대략 $2(1.96+0.84)^2/0.01^2$이다. 이 수를 계산하고, $\rho = 0.7$인 공변량 보정을 쓰면 몇 명으로 줄어드는지 구하라. 풀이 46쪽

연습 8.3 (Holm의 단계 ★★) 지표 m개의 p-값을 작은 것부터 $p_{(1)} \le \dots \le p_{(m)}$으로 놓고, $p_{(k)} \le \alpha/(m-k+1)$이 처음으로 깨지는 곳 직전까지 기각한다(Holm 1979). 참인 가설이 m_0개일 때 FWER $\le \alpha$임을 보여라. (참인 가설들 가운데 가장 작은 p-값이 α/m_0 이하일 확률을 Bonferroni로 묶어라.) 풀이 47쪽

연습 8.4 (공변량 보정의 분산: 증명 ★ 유도형) 정리 8.2를 다음 순서로 증명하라. 두 군의 크기는 고정되어 있고 처치효과는 상수 τ다.

(a) 무작위 배정 아래 $\bar X_T$와 $\bar X_C$가 같은 평균을 가짐을 보이고 $E\hat\tau_\theta = \tau$를 얻어라.

(b) $\hat\tau_\theta$가 반응 $Y - \theta X$의 두 군 평균의 차임을 보이고, 그 분산이 $\operatorname{Var}(Y-\theta X)\,(1/n_T + 1/n_C)$임을 보여라.

(c) $\operatorname{Var}(Y-\theta cX)$를 c의 이차식으로 펼쳐 $c=1$, 곧 $\theta = \operatorname{Cov}(X,Y)/\operatorname{Var}X$에서 최소이고 그 값이 $(1-\rho^2)\operatorname{Var}Y$임을 보여라.

(d) (b)와 (c)를 합쳐 정리의 식을 얻어라.

풀이 47쪽

9

정리와 역사 노트

9.1 한 장 요약

정리	내용	조건
Neyman (1.4)	$E\hat{\tau} = \bar{\tau}$, $\mathrm{Var}\,\hat{\tau} = S_1^2/n_T + S_0^2/n_C - S_\tau^2/n$	완전무작위설계
무작위화 검정 (1.7)	$P(p \le \alpha) \le \alpha$	날카로운 귀무가설
일원분산분석 (2.2)	$\mathrm{SS}_T = \mathrm{SS}_B + \mathrm{SS}_W$, $F \sim F(t-1, N-t)$, 가능도비 검정	정규 오차
비중심 분포 (2.5, 2.8)	Poisson 혼합, 평균 $k+\lambda$, 검정력은 λ의 증가함수	정규 오차
균형 이원배치 (3.2)	다섯 직교공간, 제곱합의 분해와 독립	균형
분할구 (3.6)	$E\,\mathrm{MS}_{\text{주구}} = \sigma^2 + 4\sigma_w^2$, 주구 요인은 주구 오차로	두 번의 무작위화
Bonferroni (4.2)	크기 α/m이면 FWER $\le \alpha$	—
Scheffé (4.4)	모든 대비를 $\sqrt{qF}$로 동시에	정규 선형모형
Tukey (4.6)	모든 짝을 $q_{t,\nu,\alpha}s/\sqrt{n}$으로 정확히	균형 일원배치
짝 설계 (5.1)	상대효율 $S^2/S_w^2 \approx 1/(1-\rho)$	일정한 효과
요인설계 (5.3)	주효과 분산 $4\sigma^2/N$, OFAT의 $(k+1)/2$배 효율	가법
공변량 보정 (8.2)	분산 $(1-\rho^2)$배	배정 전 공변량

이번 회차의 수학은 13회의 사영 그대로였다. 새로운 것은 확률이 어디서 오느냐였다. 1절의 정리들에서 확률의 출처는 오차분포가 아니라 실험자의 동전이다. 무작위화가 있으면 평균 차이는 불편이고(Neyman), 검정은 정확하다(Fisher). 분산분석의 F와 t는 그 정확한 무작위화 분포의 훌륭한 근사로 쓰인다. 블록, 요인, 공변량은 같은 무작위화 위에서 분산을 줄이는 장치다.

9.2 Stigler의 명명 법칙

> **이름과 발견**
>
> - **무작위화**: Fisher보다 40년 앞서 Peirce와 Jastrow(1885)가 카드로 순서를 섞었다.
> - **잠재적 결과**: Neyman(1923)의 "잠재적 수확"이 원형이다. Rubin(1974)이 관찰 연구로 넓힌 뒤 "Rubin 인과모형"이라 불리게 되었다.
> - **F 분포**: Fisher는 $z = \frac{1}{2}\log F$를 썼다. F라는 이름은 Snedecor(1934)가 붙였다.
> - **Fisher의 정확검정**: 같은 해 Irwin(1935)과 Yates(1934)도 같은 검정에 이르렀다.
> - **Bonferroni 부등식**: 합집합의 부등식 자체는 Boole(1854)까지 거슬러 올라간다.
> - **Tukey의 HSD**: 스튜던트화 범위는 Student(1927)가 처음 다뤘고, Newman(1939)과 Keuls(1952)가 그것으로 단계적 비교를 만들었다. Tukey의 원고는 41년 뒤에 출판되었다.
> - **CUPED**: Fisher(1932)의 공분산분석을 온라인 실험의 언어로 다시 쓴 것이다.

9.3 7 Pillars 지도

이번 회차는 다섯째 기둥 **Design**의 중심이다. 12회의 Wald 축차검정이 "몇 개를 관측할지"를 설계했다면, Fisher는 "어디에 무엇을 놓을지"를 설계했다. 무작위화는 추론의 근거를 모형에서 실험 절차로 옮겼다. 첫째 기둥 **Aggregation**(분산분석의 제곱합)과 둘째 기둥 **Information**(반복과 효율, $(k+1)/2$)도 함께 칠한다.

9.4 다음 회 — 가정이 틀리면: 순위 · 로버스트 · 부트스트랩 (1935 → 1979)

이번 회차의 F와 t는 정규 오차를 가정했다. 1절의 무작위화 검정은 그 가정 없이도 정확했다. 15회는 그 생각을 넓힌다. 관측값 대신 순위를 쓰는 Wilcoxon(1945)의 검정, 이상치 하나에 무너지지 않는 Huber(1964)의 로버스트 추정, 그리고 분포를 자료 자신으로 바꾸는 Efron(1979)의 부트스트랩이다. Darwin의 옥수수와 Newcomb의 빛의 속도가 다시 나온다.

참고문헌

- Peirce, C. S., Jastrow, J. On small differences of sensation. *Memoirs of the National Academy of Sciences* 3 (1885), 73–83.
- Neyman, J. On the application of probability theory to agricultural experiments (1923). 영역: *Statistical Science* 5 (1990), 465–472. · Rubin, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educational Psychology* 66 (1974), 688–701.
- Mercer, W. B., Hall, A. D. The experimental error of field trials. *J. Agricultural Science* 4 (1911), 107–132.

- Fisher, R. A. Studies in crop variation I. *J. Agricultural Science* 11 (1921), 107–135. · Fisher, R. A., Mackenzie, W. A. Studies in crop variation II. *J. Agricultural Science* 13 (1923), 311–320. · *Statistical Methods for Research Workers*. Oliver and Boyd, 1925. · The arrangement of field experiments. *J. Ministry of Agriculture* 33 (1926), 503–513. · *The Design of Experiments*. Oliver and Boyd, 1935.
- Yates, F. Complex experiments. *JRSS Suppl.* 2 (1935), 181–247. · Tukey, J. W. One degree of freedom for non-additivity. *Biometrics* 5 (1949), 232–242.
- Snedecor, G. W. *Calculation and Interpretation of Analysis of Variance and Covariance*. Collegiate Press, 1934.
- Scheffé, H. A method for judging all contrasts in the analysis of variance. *Biometrika* 40 (1953), 87–104. · Tukey, J. W. *The Problem of Multiple Comparisons* (1953). *Collected Works* VIII, Chapman & Hall, 1994. · Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* 6 (1979), 65–70.
- Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems I. *Ann. Math. Statist.* 25 (1954), 290–302. · Hurlbert, S. H. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54 (1984), 187–211.
- Francis, T. Jr., Korns, R. F., et al. An evaluation of the 1954 poliomyelitis vaccine trials. *American J. Public Health* 45 (1955), 1–63. · Meier, P. The biggest public health experiment ever. In Tanur et al. (eds.), *Statistics: A Guide to the Unknown*. Holden-Day, 1972.
- Deng, A., Xu, Y., Kohavi, R., Walker, T. Improving the sensitivity of online controlled experiments by utilizing pre-experiment data. *WSDM* (2013). · Goyal, P. et al. Accurate, large minibatch SGD. arXiv:1706.02677 (2017).
- 자료: agridat `mercer.wheat.uniformity`, `broadbalk.wheat`, `yates.oats`; datasets `npk`; HistData `ZeaMays`, `PolioTrials`.

A

연습문제 풀이

각 풀이의 머리에 있는 문제 ○○쪽을 누르면 문제로 돌아간다. 계산 문제의 수치는 노트북에서 같은 계산으로 확인했다. 유도형 연습의 풀이는 본문에서 연습으로 돌린 정리의 완성된 증명이다.

1절

연습 1.1 문제 7쪽

3개를 고르는 방법은 $\binom{6}{3} = 20$가지이므로 모두 맞힐 확률은 $1/20 = 0.05$다. 우유 먼저가 2개이면 $\binom{8}{2} = 28$가지에서 $1/28 = 0.036$이다. 4:4 설계의 $1/70 = 0.014$가 가장 작아서, 한 번의 완벽한 답이 가장 강한 증거가 된다. 2:6 설계는 "우유 먼저는 드물다"는 쪽으로 찍는 전략에 약하다.

연습 1.2 문제 7쪽

$A = y_1(1) - y_2(1)$, $B = y_1(0) - y_2(0)$이라 하면 $\hat{\tau}$의 두 값의 차는 $A + B$이고 확률이 반반이므로 $\operatorname{Var}\hat{\tau} = (A+B)^2/4$. 정리의 식은 $n = 2$에서 $S_1^2 = A^2/2$, $S_0^2 = B^2/2$, $S_\tau^2 = (A-B)^2/2$이므로 $A^2/2 + B^2/2 - (A-B)^2/4 = (A+B)^2/4$로 같다.

연습 1.3 문제 7쪽

$\mu = \bar{y}(1)$이라 하면 $\sum_{W_i=1}(y_i(1) - \bar{y}_T)^2 = \sum_i W_i(y_i(1) - \mu)^2 - n_T(\bar{y}_T - \mu)^2$. 첫 항의 기댓값은 $\frac{n_T}{n}(n-1)S_1^2$이다. 보조정리 1.3로 $\operatorname{Var}\sum W_i y_i(1) = \frac{n_T n_C}{n(n-1)}(n-1)S_1^2$이므로 $\operatorname{Var}\bar{y}_T = \frac{n_C}{n n_T}S_1^2$이고, 둘째 항의 기댓값은 $\frac{n_C}{n}S_1^2$이다. 차는 $\frac{n_T(n-1)-n_C}{n}S_1^2 = (n_T - 1)S_1^2$이다.

연습 1.4 문제 7쪽

$p \le 0.05$는 $K = 4$일 때뿐이고($p = 1/70$), 그 확률이 $1/70 = 0.014$다. p-값이 $1/70,\ 17/70, \ldots$의 이산값만 가지므로 0.05에 맞는 값이 없다. 확률화 검정(12회 예제 1.4): $K = 4$이면 기각, $K = 3$이면 확률 γ로 기각하면 크기는 $\frac{1}{70} + \gamma\frac{16}{70} = 0.05$, 곧 $\gamma = 0.156$이다.

연습 1.5 문제 8쪽

$\tau_i \equiv \tau_0$이면 관측된 쪽에서 τ_0을 더하거나 빼 다른 쪽을 채울 수 있다. Darwin의 짝에서는 $d_i - \tau_0$

이 날카로운 귀무가설 아래의 차이이므로, $d_i - \tau_0$의 부호를 뒤집는 무작위화 검정을 한다. 기각되지 않는 τ_0들의 집합이 신뢰집합이다. 검정과 구간의 쌍대성(8회 정리 3.5) 그대로다.

연습 1.6 문제 8쪽

(a) 단위 i를 처치에 넣고 나머지 $n-1$개에서 n_T-1개를 고르는 배정이 $\binom{n-1}{n_T-1}$개이므로 $P(W_i=1)=\binom{n-1}{n_T-1}/\binom{n}{n_T}=n_T/n$이다. $W_i^2=W_i$이므로 $\operatorname{Var}W_i=EW_i-(EW_i)^2=\frac{n_T}{n}\cdot\frac{n_C}{n}$이다. (b) 같은 방법으로 $P(W_i=W_j=1)=\binom{n-2}{n_T-2}/\binom{n}{n_T}=\frac{n_T(n_T-1)}{n(n-1)}$이다. (c) $\operatorname{Cov}(W_i,W_j)=\frac{n_T(n_T-1)}{n(n-1)}-\frac{n_T^2}{n^2}=\frac{n_T}{n}\cdot\frac{n(n_T-1)-(n-1)n_T}{n(n-1)}=-\frac{n_Tn_C}{n^2(n-1)}$이다. 분자에서 $n(n_T-1)-(n-1)n_T=n_T-n=-n_C$를 썼다. (d) $\sum_i W_i=n_T$이므로 분산은 0이고, $n\frac{n_Tn_C}{n^2}+n(n-1)\operatorname{Cov}(W_1,W_2)=0$에서 $\operatorname{Cov}(W_1,W_2)=-\frac{n_Tn_C}{n^2(n-1)}$이 다시 나온다. 4회 정리 1.5의 초기하분포 분산 공식이 같은 계산에서 나온다.

2절

연습 2.1 문제 13쪽

$\sum_i n_i(\bar y_{i\cdot}-\bar y)^2=\sum n_i\bar y_{i\cdot}^2-2\bar y\sum n_i\bar y_{i\cdot}+N\bar y^2$이고 $\sum n_i\bar y_{i\cdot}=N\bar y$. 처치 i 안의 제곱합은 $(n_i-1)s_i^2$이다.

연습 2.2 문제 13쪽

$t=2$이면 $\bar y=\frac{n_1\bar y_1+n_2\bar y_2}{N}$이고 $\mathrm{SS}_B=\frac{n_1n_2}{N}(\bar y_1-\bar y_2)^2$이다. $\mathrm{SS}_W/(N-2)=s_p^2$이므로 $F=\frac{(\bar y_1-\bar y_2)^2}{s_p^2(1/n_1+1/n_2)}=t^2$.

연습 2.3 문제 13쪽

$\boldsymbol\mu$의 성분이 μ_i이므로 $\boldsymbol\mu\in\operatorname{col}(X)$, $H\boldsymbol\mu=\boldsymbol\mu$, $J\boldsymbol\mu=\bar\mu\mathbf 1$이다. $t=2$, $C=(1,-1)$이면 $\lambda=\frac{(\mu_1-\mu_2)^2}{\sigma^2(1/n_1+1/n_2)}$이고, $\sum n_i(\mu_i-\bar\mu)^2=\frac{n_1n_2}{N}(\mu_1-\mu_2)^2$이므로 같다.

연습 2.4 문제 13쪽

$E(Z+a)^2=1+a^2$을 더하면 $k+\lambda$. $(Z+a)^2=Z^2+2aZ+a^2$이고 $\operatorname{Cov}(Z^2,Z)=EZ^3=0$이므로 분산은 $2+4a^2$, 더하면 $2k+4\lambda$.

연습 2.5 문제 13쪽

$\lambda=n\sum(\mu_i-\bar\mu)^2/\sigma^2$이므로 σ가 두 배면 같은 λ에 n이 네 배 필요하다. 그러나 n이 커지면 분모의 자유도도 커져 같은 λ에서 검정력이 오르므로, 실제로는 수준당 8구획(3구획의 약 2.7배)이면

검정력 0.80이다(노트북).

연습 2.6 문제 13쪽

(a) 13회 보조정리 2.4로 계수 r의 대칭 멱등행렬은 $\mathrm{col}(P)$의 정규직교 기저를 열로 하는 U에 대해 $P = UU'$, $U'U = I_r$로 쓸 수 있다. (b) $U'\mathbf{y}$는 정규벡터의 선형변환이므로 정규이고(5회 정리 3.8), 평균은 $U'\boldsymbol{\mu}$, 공분산은 $U'(\sigma^2 I)U = \sigma^2 I_r$이다. σ로 나누면 $\mathcal{N}_r(U'\boldsymbol{\mu}/\sigma, I_r)$이다. (c) $\|U'\mathbf{y}\|^2 = \mathbf{y}'UU'\mathbf{y} = \mathbf{y}'P\mathbf{y}$이고 같은 방법으로 $\|U'\boldsymbol{\mu}\|^2 = \boldsymbol{\mu}'P\boldsymbol{\mu}$이다. 정의 2.4에서 $k = r$, $\lambda = \|U'\boldsymbol{\mu}/\sigma\|^2 = \boldsymbol{\mu}'P\boldsymbol{\mu}/\sigma^2$이다.

연습 2.7 문제 13쪽

(a) $\hat{\boldsymbol{\beta}} \sim \mathcal{N}_p(\boldsymbol{\beta}, \sigma^2(X'X)^{-1})$(13회 정리 4.7)이므로 $C\hat{\boldsymbol{\beta}} - \mathbf{d} \sim \mathcal{N}_q(C\boldsymbol{\beta} - \mathbf{d}, \sigma^2 W)$이다. W는 양정치이므로 $W^{-1/2}(C\hat{\boldsymbol{\beta}} - \mathbf{d})/\sigma \sim \mathcal{N}_q\big(W^{-1/2}(C\boldsymbol{\beta} - \mathbf{d})/\sigma, I_q\big)$이고, 그 제곱 길이가 분자다. 정의 2.4로 분자는 $\chi^2(q, \lambda)$이고 $\lambda = \|W^{-1/2}(C\boldsymbol{\beta} - \mathbf{d})\|^2/\sigma^2$은 정리의 식이다. (b) $\mathrm{RSS} = \|(I - H)\mathbf{y}\|^2$이고 $(I - H)X\boldsymbol{\beta} = \mathbf{0}$이므로 평균과 무관하게 $\mathrm{RSS}/\sigma^2 \sim \chi^2(n - p)$이며, 13회 정리 4.7로 $\hat{\boldsymbol{\beta}}$과 독립이다. 분자는 $\hat{\boldsymbol{\beta}}$의 함수이므로 분모와 독립이다. (c) 일원분산분석의 분자는 $\|(H - J)\mathbf{y}\|^2/\sigma^2$이고 $H - J$는 계수 $t - 1$의 대칭 멱등이다. 따름정리 2.6로 $\lambda = \boldsymbol{\mu}'(H - J)\boldsymbol{\mu}/\sigma^2 = \|(H - J)\boldsymbol{\mu}\|^2/\sigma^2$이고, 연습 2.3로 $(H - J)\boldsymbol{\mu}$의 (i, j) 성분이 $\mu_i - \bar{\mu}$이므로 $\lambda = \sum_i n_i(\mu_i - \bar{\mu})^2/\sigma^2$이다. (d) Q를 분자, V를 분모의 χ^2이라 하면 $F = \dfrac{Q/q}{V/(n-p)}$이므로 $P(F > c \mid V = v) = P\big(Q > cqv/(n - p)\big)$이다($Q$와 V는 독립). 정리 2.5(e)로 이것은 모든 v에서 λ의 증가함수이고, V의 분포는 λ와 무관하므로 V에 대해 평균한 $P(F > c)$도 λ의 증가함수다.

3절

연습 3.1 문제 20쪽

처치 $t - 1$, 블록 $b - 1$, 오차 $(t - 1)(b - 1)$, 전체 $tb - 1$. 제곱합은 $b\sum(\bar{y}_{i\cdot} - \bar{y})^2$, $t\sum(\bar{y}_{\cdot j} - \bar{y})^2$, 나머지다. 블록을 무시하면 오차의 자유도는 $t(b - 1)$로 $b - 1$만큼 크지만, 블록 사이의 변동이 오차에 섞여 평균제곱이 커진다.

연습 3.2 문제 20쪽

칸의 수가 n_{ij}이면 $\sum_{ijk} f(i)g(j) = \sum_{ij} n_{ij} f(i)g(j)$이다. 모든 칸이 r이면 $r(\sum f)(\sum g)$로 갈라진다. 일반적으로는 $n_{ij} = n_{i\cdot}n_{\cdot j}/N$(비례 도수)이면, f와 g를 각각 무게 $n_{i\cdot}$, $n_{\cdot j}$의 가중평균이 0이 되게 중심화했을 때 $\sum n_{ij} f(i)g(j) = \frac{1}{N}(\sum n_{i\cdot} f(i))(\sum n_{\cdot j} g(j)) = 0$으로 직교한다.

연습 3.3 문제 20쪽

전체 713.9에서 316.6/713.9 = 0.44, 253.1/713.9 = 0.35, 144.2/713.9 = 0.20. 오차의 평균제곱 144.2/1168 = 0.1235로 $F_{\text{해}} = (316.6/73)/0.1235 = 35.1$, $F_{\text{구획}} = (253.1/16)/0.1235 = 128.1$. 구획은 자유도가 16뿐이어서 자유도당 변동이 해보다 훨씬 크다.

연습 3.4 문제 20쪽

$\sum_j \hat{b}_j = 0$이므로 $\sum_{ij} \bar{y}_{i\cdot}\hat{a}_i\hat{b}_j = 0$이고, 같은 이유로 $\bar{y}_{\cdot j}$와 $\bar{y}_{\cdot\cdot}$의 항도 0이어서 $\sum e_{ij}\hat{a}_i\hat{b}_j = \sum y_{ij}\hat{a}_i\hat{b}_j$이다. 원점을 지나는 회귀의 제곱합은 $(\sum e_{ij}z_{ij})^2/\sum z_{ij}^2$ ($z_{ij} = \hat{a}_i\hat{b}_j$)이고 $\sum z_{ij}^2 = \sum \hat{a}_i^2 \sum \hat{b}_j^2$이다.

연습 3.5 문제 20쪽

한 품종의 평균은 그 품종의 주구 6개의 η 평균과 세구 24개의 ε 평균을 포함하므로(블록 효과는 모든 품종에 같다) 분산이 $\sigma_w^2/6 + \sigma^2/24$이다. 두 품종은 서로 다른 주구를 쓰므로 차의 분산은 두 배, $\sigma_w^2/3 + \sigma^2/12 = (4\sigma_w^2 + \sigma^2)/12$이다. 분자에 $4\sigma_w^2 + \sigma^2$이 들어 있으므로 기댓값이 같은 주구 오차로 나눠야 한다.

4절

연습 4.1 문제 24쪽

독립이면 하나도 기각하지 않을 확률이 $0.95^{15} = 0.463$이므로 FWER는 0.537이다. 짝 비교 15개는 같은 평균 6개와 같은 s를 공유해서 양으로 상관되고, 기각 사건들이 겹치므로 합집합의 확률이 독립일 때보다 작다.

연습 4.2 문제 24쪽

대비 $\mathbf{c}$ 하나의 $T = \frac{\mathbf{c}'(\hat{\beta}-\beta)}{s\sqrt{\mathbf{c}'(X'X)^{-1}\mathbf{c}}} \sim t(\nu)$이고, Scheffé의 사건은 $|T| \le \sqrt{qF}$를 포함하므로 $P(|T| \le \sqrt{qF}) \ge 1-\alpha = P(|T| \le t_{\nu,\alpha/2})$이다. T는 연속이므로 $\sqrt{qF} \ge t_{\nu,\alpha/2}$. $q = 1$이면 $F(1,\nu)$는 $t(\nu)^2$의 분포여서 같다.

연습 4.3 문제 24쪽

HSD $= 3.773 \times \sqrt{177.1/18} = 11.83$. 짝의 차는 19.5, 34.8, 44.0, 15.3, 24.5, 9.2이고 마지막(0.4와 0.6)만 작다. 질소 0과 0.2만 미리 정해 두었다면 보통의 t 구간 $19.5 \pm t_{45,0.025}\sqrt{2 \cdot 177.1/18} = 19.5 \pm 8.9$를 쓴다.

연습 4.4 문제 24쪽

$-3(79.39) - 98.89 + 114.22 + 3(123.39) = 147.3$, 표준오차 $\sqrt{177.1 \cdot 20/18} = 14.0$. 자료를 보고 고른 대비이므로 Scheffé의 기준값 2.904를 쓴다. $t = 10.5$는 그래도 훨씬 크다.

연습 4.5 문제 24쪽

(a) $\bar{y}_{i\cdot} \sim \mathcal{N}(\mu_i, \sigma^2/n)$이고 처치가 다르면 서로 다른 관측의 평균이므로 독립이다. 표준화하면 $Z_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$이다. (b) 정리 2.2의 증명에서 $(H\mathbf{y})_{ij} = \bar{y}_{i\cdot}$이고 $s^2 = \|(I-H)\mathbf{y}\|^2/\nu$이다. $H\mathbf{y}$와 $(I-H)\mathbf{y}$는 직교하는 사영의 정규벡터이므로 독립이고(13회 따름정리 4.2), $\nu s^2/\sigma^2 = \|(I-H)\mathbf{y}\|^2/\sigma^2 \sim \chi^2(\nu)$이다. (c) $|Z_i - Z_j| \le \max Z - \min Z$이고 등호는 Z_i가 최대, Z_j가 최소일 때 이루어지므로 $\max_{i<j} |Z_i - Z_j| = \max Z - \min Z$이다. 따라서 "모든 짝에서 $\le c$"와 "범위 $\le c$"는 같은 사건이다. (d) 정리의 사건은 모든 짝에서 $|Z_i - Z_j| \le q_{t,\nu,\alpha}\, s/\sigma$, 곧 $Q = \frac{\max Z - \min Z}{\sqrt{V/\nu}} \le q_{t,\nu,\alpha}$이다. (a)(b)로 Q는 정의 4.5의 스튜던트화 범위이므로 확률은 $1-\alpha$이다.

5절

연습 5.1 문제 27쪽

평균이 같으면 $E(a-b)^2 = \mathrm{Var}(a-b) = 2\sigma^2 - 2\rho\sigma^2$이므로 $S_w^2 \approx \sigma^2(1-\rho)$이고 효율은 $1/(1-\rho)$이다. $\rho < 0$이면(짝의 두 단위가 반대로 움직이면) 효율이 1보다 작아 짝을 짓는 것이 손해다.

연습 5.2 문제 27쪽

요인설계는 $4\sigma^2/24 = \sigma^2/6$, OFAT는 $2 \cdot 4\sigma^2/24 = \sigma^2/3$로 효율이 2배다. 상호작용 추정량은 6구획씩의 평균 네 개에 $\pm\frac{1}{2}$을 곱한 것이므로 분산이 $4 \cdot \frac{1}{4} \cdot \sigma^2/6 = \sigma^2/6$으로 주효과와 같다.

연습 5.3 문제 27쪽

NPK의 추정량은 NPK 부호가 $+1$인 12구획의 평균에서 -1인 12구획의 평균을 뺀 것이다. 블록마다 한 부호만 있으므로 이것은 블록 $2\cdot3\cdot4$의 평균 − 블록 $1\cdot5\cdot6$의 평균이다. 블록 효과가 그대로 더해지므로 NPK 효과와 블록 차이를 가를 수 없다.

연습 5.4 문제 28쪽

기준(10)에서 A만 바꿔도 10, B만 바꿔도 10이므로 OFAT는 "A도 B도 효과 없음"이라 결론짓는다. 요인설계는 두 요인을 함께 바꾼 16을 보고, 주효과(각 +3)와 상호작용(곱의 효과)이 있음을 안다. 효과가 상호작용으로만 있으면 OFAT는 그것을 볼 수 없다.

연습 5.5 문제 28쪽

(a) $y_i(1) = y_i(0) + \tau$이면 두 잠재적 결과의 분산이 같아 $S_1^2 = S_0^2 = S^2$이고, τ_i가 상수이므로 $S_\tau^2 = 0$이다. 정리 1.4에 $n_T = n_C = m$을 넣으면 $\operatorname{Var}\hat{\tau} = S^2/m + S^2/m = 2S^2/m$이고 불편성도 그 정리에서 나온다. (b) $\sigma_i = +1$이면 처치는 a_i쪽이고 관측 차이는 $(a_i + \tau) - b_i$이다. $\sigma_i = -1$이면 $(b_i + \tau) - a_i$이다. 두 경우 모두 $\tau + \sigma_i(a_i - b_i)$이다. (c) $\hat{\tau} = \frac{1}{m}\sum_i\big(\tau + \sigma_i(a_i - b_i)\big) = \tau + \frac{1}{m}\sum_i \sigma_i(a_i - b_i)$이다. $E\sigma_i = 0$이므로 평균은 τ이고, σ_i들이 독립이고 분산이 1이므로 분산은 $\frac{1}{m^2}\sum_i (a_i - b_i)^2 = \frac{2S_w^2}{m}$이다. (d) a, b의 분산이 S^2, 상관이 ρ이고 평균이 같으면 $E(a-b)^2 = 2S^2(1-\rho)$이므로 $S_w^2 = \frac{1}{2m}\sum (a_i - b_i)^2 \approx S^2(1-\rho)$이다. 상대효율은 $\frac{2S^2/m}{2S_w^2/m} \approx \frac{1}{1-\rho}$이다.

연습 5.6 문제 28쪽

(a) 2^k개 조합 가운데 한 요인이 높은 수준인 것이 절반이므로 높은 수준에 $\frac{N}{2^k}\cdot 2^{k-1} = N/2$개, 낮은 수준에도 $N/2$개가 있다. 두 평균은 서로 다른 단위의 평균이어서 독립이므로 차의 분산은 $\sigma^2(2/N + 2/N) = 4\sigma^2/N$이다. (b) 높은 수준 쪽의 $N/2$개 안에서 다른 요인 B의 두 수준은 $N/4$개씩 나온다. 낮은 수준 쪽도 같다. 가법 모형에서 B의 효과 $\pm\beta$는 두 평균에 똑같이 $\frac{(N/4)\beta + (N/4)(-\beta)}{N/2} = 0$으로 들어가므로 차에서 지워진다. 모든 다른 요인이 같다. (c) OFAT의 주효과는 요인 하나만 바꾼 조합의 평균에서 기준 조합의 평균을 뺀 것이고, 두 조합에는 $N/(k+1)$개씩 있다. 분산은 $\sigma^2 \cdot 2(k+1)/N$이다. (d) $\frac{2(k+1)\sigma^2/N}{4\sigma^2/N} = \frac{k+1}{2}$이다. 상호작용 $(\alpha\beta)$를 보려면 두 요인을 함께 바꾼 조합이 있어야 하는데, OFAT의 조합은 기준에서 요인을 하나만 바꾼 것들이라 그런 조합이 없다.

6절

연습 6.1 문제 31쪽

$\log 0.288 = -1.245$, $\mathrm{SE} = \sqrt{1/33 - 1/200745 + 1/115 - 1/201229} = 0.1975$. $-1.245 \pm 1.96 \times 0.1975$를 지수로 돌리면 $[0.195,\ 0.424]$이고, 백신 효과 $1 - \mathrm{RR}$의 구간은 $[0.576,\ 0.805]$다.

연습 6.2 문제 31쪽

2×2 표의 Pearson χ^2은 $\frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$이고, 합동 비율 $\hat{p}$로 만든 $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$를 제곱해 정리하면 같은 식이다. 노트북에서 $z = -6.727$, $\chi^2 = 45.25$를 확인한다.

연습 6.3 문제 31쪽

마비형과 비마비형을 합치면 $28.4/70.6 = 0.40$. 비마비형만이면 $12.0/13.4 = 0.89$로 거의 차이가 없다. 비마비형 진단에는 소아마비가 아닌 병이 섞이기 쉽다. 이중맹검이 없으면 접종 사실을 아는 의사가 접종한 아이를 덜 진단할 수 있으므로, 관찰 대조 설계에서는 백신 효과가 부풀려지는

쪽으로 치우칠 수 있다. 무작위 이중맹검은 이 경로를 막는다.

7절

연습 7.1 문제 33쪽

$S^2 = 0.2100$이면 무작위 배치의 표준편차는 $\sqrt{0.21(1/240 + 1/260)} = 0.041$이고, -0.33은 약 -8 표준편차다. 우연으로는 일어날 수 없는 크기다. 두 표본 t가 이를 그대로 보여 준 것이다.

연습 7.2 문제 33쪽

주구 오차는 블록 × 품종이므로 $(6-1)(3-1) = 10$이다. 72세구를 단위로 세면 잔차 자유도가 55로 부풀었다.

연습 7.3 문제 33쪽

칸의 값을 A1B1: 0, 0, A1B2: 1, A2B1: 1, A2B2: 2라 하자(A와 B의 효과가 각각 1인 가법 자료). A를 먼저 넣으면 A1 평균 1/3, A2 평균 3/2, 전체 평균 0.8이므로 $\mathrm{SS}(A) = 3(1/3 - 0.8)^2 + 2(1.5 - 0.8)^2 = 1.633$. B를 먼저 넣은 모형의 잔차제곱합은 1.167이고 가법 모형은 정확히 맞으므로 $\mathrm{SS}(A \mid B) = 1.167$이다. 둘이 다르다.

연습 7.4 문제 33쪽

평균이 같으면 $E\,\mathrm{SS}_W = \sum (n_i - 1)\sigma_i^2$, $E\,\mathrm{SS}_B = \sum \sigma_i^2 (1 - n_i/N)$이다. $(n_i) = (5, 10, 20)$, $(\sigma_i) = (3, 2, 1)$이면 $E\,\mathrm{SS}_W/32 = 91/32 = 2.84$, $E\,\mathrm{SS}_B/2 = 11.0/2 = 5.5$로 비가 1.93이다. 뒤집으면 6.59와 3.79로 비가 0.57이다. 비가 1보다 크면 F가 너무 커서 크기가 0.20으로, 작으면 0.014로 간다.

8절

연습 8.1 문제 36쪽

FWL로 처치의 계수는 Y와 처치 표시를 각각 $(1, X)$로 맞춘 잔차끼리의 기울기다. 처치 표시는 무작위여서 X와 거의 무상관이므로, 그 계수는 $(\bar{Y}_T - \bar{Y}_C) - \hat{\theta}(\bar{X}_T - \bar{X}_C)$이고 $\hat{\theta}$는 군 안의 합동 기울기다. 큰 표본에서 이것은 CUPED와 같다.

연습 8.2 문제 36쪽

$2(1.96 + 0.84)^2 / 0.01^2 = 2 \times 7.84 \times 10^4 = 156{,}800$명이다. 분산이 $1 - 0.7^2 = 0.51$배가 되면 79,968

명, 곧 약 절반이다.

연습 8.3 문제 36쪽

참인 가설들의 집합을 I_0 ($|I_0| = m_0$)이라 하자. Holm이 참인 가설을 하나라도 기각했다면, 처음으로 기각된 참인 가설은 순서상 $k \le m - m_0 + 1$ 번째에 있다(그 앞에는 거짓인 가설만 있을 수 있다). 그 p-값은 $\alpha/(m-k+1) \le \alpha/m_0$ 이하다. 따라서 FWER $\le P(\min_{i\in I_0} p_i \le \alpha/m_0) \le m_0 \cdot \alpha/m_0 = \alpha$ (Bonferroni).

연습 8.4 문제 36쪽

(a) 무작위 배정은 처치 전에 정해진 X와 무관하므로 두 군의 X는 같은 분포를 갖고 $E\bar{X}_T = E\bar{X}_C$ 이다. 처치효과가 상수 τ이므로 $E(\bar{Y}_T - \bar{Y}_C) = \tau$이고, $E\hat{\tau}_\theta = \tau - \theta(E\bar{X}_T - E\bar{X}_C) = \tau$이다. (b) $\hat{\tau}_\theta = \overline{(Y-\theta X)}_T - \overline{(Y-\theta X)}_C$이다. 단위들이 독립이고 처치는 Y를 상수만큼 옮기므로 $Y - \theta X$의 분산은 두 군에서 같다. 두 평균은 독립이므로 $\operatorname{Var}\hat{\tau}_\theta = \operatorname{Var}(Y-\theta X)\left(\frac{1}{n_T} + \frac{1}{n_C}\right)$이다. (c) $\theta_0 = \operatorname{Cov}(X,Y)/\operatorname{Var}X$라 하면 $\operatorname{Var}(Y - c\theta_0 X) = \operatorname{Var}Y - 2c\theta_0 \operatorname{Cov}(X,Y) + c^2\theta_0^2 \operatorname{Var}X = \operatorname{Var}Y - (2c - c^2)\frac{\operatorname{Cov}(X,Y)^2}{\operatorname{Var}X}$이다. 이것은 $c = 1$에서 최소이고 그 값은 $\operatorname{Var}Y - \frac{\operatorname{Cov}(X,Y)^2}{\operatorname{Var}X} = (1-\rho^2)\operatorname{Var}Y$이다. (d) (b)에 (c)를 넣으면 $\operatorname{Var}\hat{\tau}_\theta = (1-\rho^2)\operatorname{Var}Y\left(\frac{1}{n_T} + \frac{1}{n_C}\right) = (1-\rho^2)\operatorname{Var}(\bar{Y}_T - \bar{Y}_C)$이다.