2025-겨울 통계학
Jan. 28, 2026
박성우 (sungw.ooo)

일반통계학 연습문제 풀이

- 일반통계학 (서울대학교 자연과학대학 통계학과, 2025)를 교제로 사용한 2025년 겨울학기 통계학 수업의 과제로 선정된 일부 연습문항에 대한 풀이입니다.
- 본 풀이를 참고용으로 사용하길 권장드립니다. 만약 이 풀이를 과제에 직접적으로 사용하게 된다면 표절 등의 문제에 대해 필자는 책임을 지지 않습니다.
- 풀이과정 중 궁금한 점이 있거나 오탈자를 발견했다면 블로그 글 하단 댓글 혹은 park.sw@snu.ac.kr 로 문의 바랍니다.

2장

연습문제 2.1 모집단과 표본

(a)
- 모집단: 전체 국민 (각 연령대에 속한 모든 국민)
- 표본: 실제 조사에 참여한 연령대별 100명

(b)
- 모집단: 해당 분식집에서 한 달간 판매된 모든 주문 거래
- 표본: 기록된 모든 거래 (전수조사이므로 표본 = 모집단)

(c)
- 모집단: 건강보험 가입자 전체
- 표본: 공개된 10만 명의 병원 이용 형태 데이터

(d)
- 모집단: 대한민국 모든 초등학생
- 표본: 서울대학교사범대학부속초등학교 학생 전체

연습문제 2.2 모수와 통계량

(a) 평균 연령 <u>36.7</u>세: $\boxed{\text{통계량}}$

(b) 평균 학점 <u>3.45</u>: $\boxed{\text{통계량}}$

(c)
- 공장 전체 타이어의 평균 지름 <u>60 mm</u>: $\boxed{\text{모수}}$
- 생산된 타이어 200개의 표본 평균 <u>59.8 mm</u>: $\boxed{\text{통계량}}$

연습문제 2.3 선형 변환

유한 모집단 $C = \{c_1, \ldots, c_N\}$의 모평균을 μ, 모표준편차를 σ라 할 때,

$$D = \{d_i = a\, c_i + b \mid i = 1, \ldots, N\}$$

의 모평균 μ_D와 모표준편차 σ_D는 다음과 같다:

$$\mu_D = \frac{1}{N}\sum_{i=1}^{N}(a\, c_i + b) = \frac{1}{N}\left\{a\sum_{i=1}^{N} c_i + b\right\} = a\,\mu + b,$$

$$\sigma_D = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[(a\, c_i + b) - (a\mu + b)\right]^2} = \sqrt{\frac{1}{N}\cdot a^2\sum_{i=1}^{N}(c_i - \mu)^2} = |a|\,\sigma.$$

연습문제 2.4 형태를 통한 모평균과 모분산의 비교

(a) A와 B 모두 0에 대해 대칭 꼴의 분포 형태를 가진다. 따라서 두 모평균은 모두 0이다. 또한 A의 분포가 더욱 중심에 몰려있으므로 A의 모분산이 B의 모분산보다 작다.

$$\mu_A = \mu_B = 0, \quad \sigma_A^2 < \sigma_B^2.$$

(b) 두 분포 모두 비대칭 형태이고, A가 더 왼쪽으로 쏠려 있다. 따라서 모평균은 A가 B보다 더 작다. 반대로 B의 분포는 A보다 더 퍼져있음으로, B의 모분산이 A보다 더 크다.

$$\mu_A < \mu_B, \quad \sigma_A^2 < \sigma_B^2.$$

참고: A와 B 모두 양의 왜도를 가진다.

연습문제 2.6 기말시험 성적 분석

(a)

1	0 5
2	0 0 0 0 1 5 5 9 9
3	0 3 5 7 8
4	0 0 5 5 5 7
5	0 0 3 5 5
6	0 0 0 3 5 9
7	2 5 7
8	0 2 8
9	2 3

분포 설명: 성적은 중앙값 인근(40–60점)에 다수 분포하며, 하위권(20–40점)과 상위권(60–80점)에도 분포를 보인다. 특히 20점대에 촘촘히 분포되는 것을 확인할 수 있다. 또한 극단값(10점, 93점)이 소수 존재한다.

(b) 통계량을 계산하여 일부 숫자를 반올림하면 다음과 같다.

통계량	값
평균	47.88 점
(모)표준편차	22.43 점
중앙값	45 점
제1사분위수 (Q_1)	29 점
제3사분위수 (Q_3)	63 점
사분위수범위 (IQR)	34 점
평균절대편차 (MAD)	18.75 점

참고: 문제에서 주어진 집단은 모집단이므로 표본표준편차가 아닌 모표준편차를 계산해야 한다.

(c) 다음 예시는 예시일 뿐 유일한 정답이 아닙니다. 이점 참고 바랍니다.
분류 기준 예시 (동일 폭 20점 구간):

1등급: $x \geq 80$ 점 (5명)
2등급: $60 \leq x < 80$ 점 (9명)
3등급: $40 \leq x < 60$ 점 (11명)
4등급: $20 \leq x < 40$ 점 (14명)
5등급: $x < 20$ 점 (2명)

막대그래프와 원형그래프는 Figure 1와 2 참고.

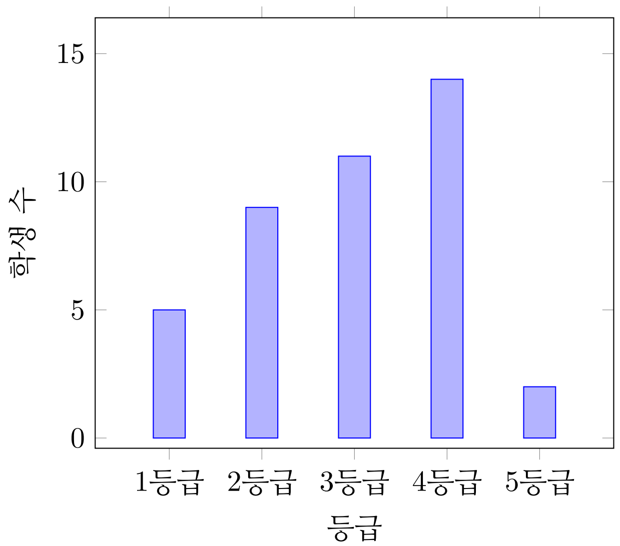

Figure 1: 2.6(c) 성적 등급별 학생 수 막대그래프

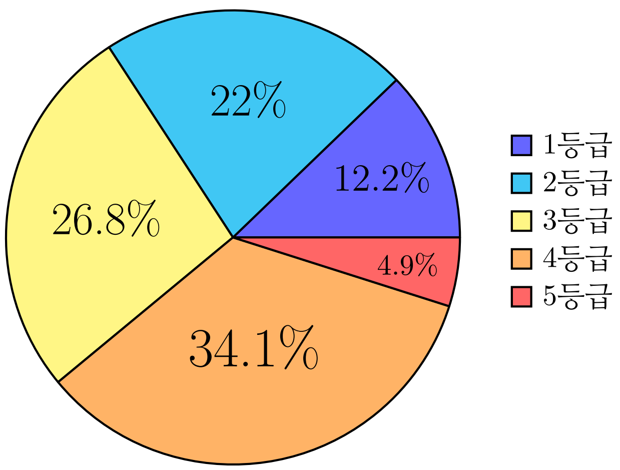

Figure 2: 2.6(c) 성적 등급별 학생 비율 원형그래프 (반올림)

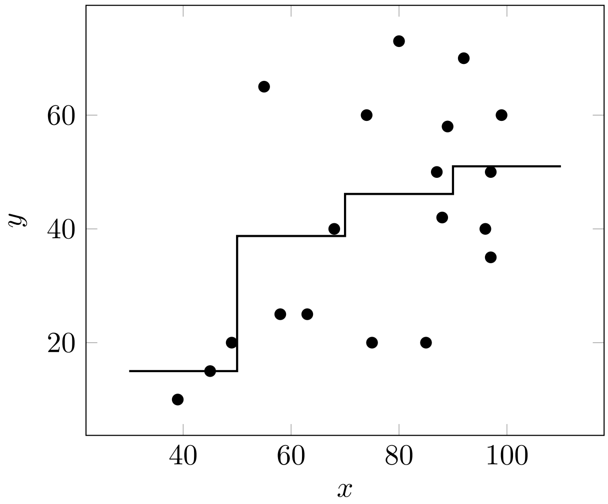

Figure 3: 2.9(a) 과제성적과 시험성적의 산점도 및 회귀도

연습문제 2.9 과제성적 vs 시험성적 분석

(a) 산점도와 회귀도는 Figure 3 참고.

참고: 회귀도는 히스토그램과 마찬가지로 구간을 어떻게 잡았느냐에 따라 다르게 그려질 수 있다. 여기서 그린 구간은 x축의 구간이 $[30, 50)$, $[50, 70)$, $[70, 90)$, $[90, 110)$로 설정되었다.

(b) 통계량을 계산하여 일부 숫자를 반올림하면 다음과 같다.

	평균	분산	표준편차
x	75.579	350.349	18.718
y	40.947	374.997	19.365

상관계수는 0.54로 양의 상관관계를 가진다.

연습문제 2.10 비선형 관계 및 상관계수

(a) Figure 4 참고. 이 자료는 $y = x^2$ 형태의 이차관계로, 선형관계가 아니므로 단순 선형상관이 약함.

(b) **통계량 및 상관계수 계산**

	평균	분산	표준편차
x	0	2	1.4142
y	2	2.8	1.6733

$$r_{xy} = \sum_i (x_i - 0)(y_i - 2) = 0$$

이는 (a)의 산점도를 보아도 자료가 선형 관계가 전혀 아니기 때문이다.

(c) Figure 5 참고. x^2와 y는 선형 상관관계에 있어, x와 y 사이에는 상관관계가 없다고 할 수는 없다.

참고: 보통 상관계수라고 말하는 피어슨 상관 계수(Pearson Correlation Coefficient)는 선형 상관관계 유무를 알려주는 척도이다. 따라서 상관계수가 0이면 선형 상관관계가 없다고 말할 수 있지만, 두 변수 사이의 상관관계가 있을 수도 있다.

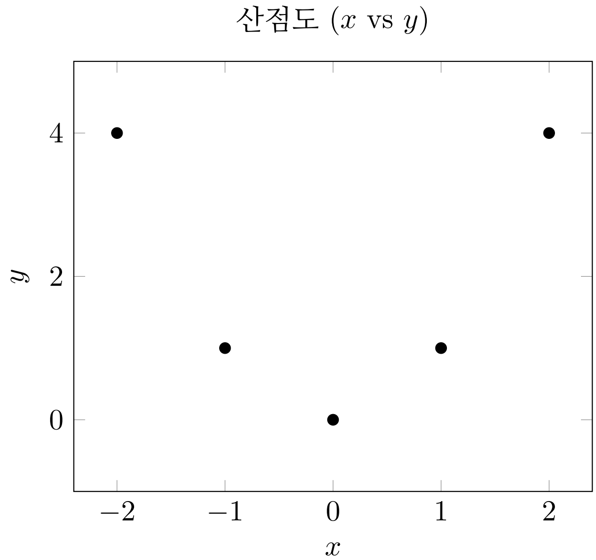

Figure 4: 2.10(a) 산점도: 포물선 형태, 이차함수 관계

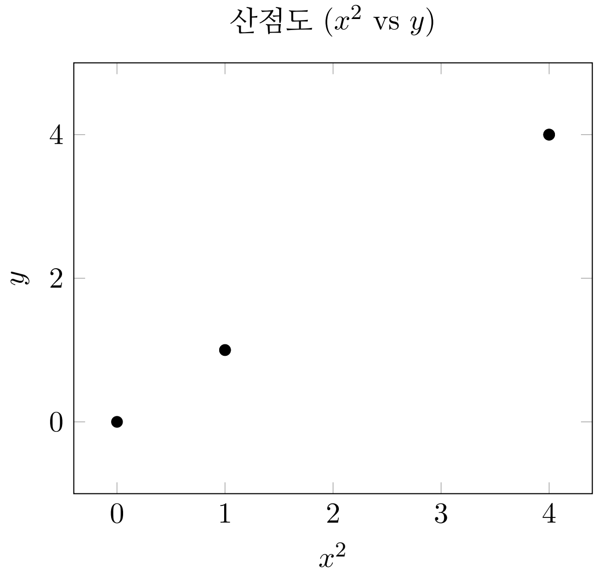

Figure 5: 2.10(c) x^2와 y의 산점도: 선형 상관관계

3장

연습문제 3.2 서버의 정상작동 확률

서버 n이 고장 나는 사건을 A_n, $(n = 1, 2, 3, 4)$이라 하자. 그렇다면 $\mathbb{P}(A_n) = 0.03, n = 1, 2, 3, 4$
각 서버가 고장 나는 사건이 독립이므로,

$$\mathbb{P}(\text{정상}) = 1 - \mathbb{P}\,(\text{모두 고장}) = 1 - \mathbb{P}\left(\bigcap_{i=1}^{4} A_i\right) = 1 - \prod_{i=1}^{4} \mathbb{P}(A_i) = 1 - (0.03)^4 = 0.99999919$$

연습문제 3.5 트럼프 카드

(a) $\mathbb{P}(A \cap B) = 1/52 > 0$ 이므로, 서로 배반이 아니다.

(b) 사건 A의 확률은 $\mathbb{P}(A) = 4/52 = 1/13$, 사건 B의 확률은 $\mathbb{P}(B) = 13/52 = 1/4$
$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \dfrac{1}{13} + \dfrac{1}{4} - \dfrac{1}{52} = \dfrac{4}{13}$

(c) $\mathbb{P}(A \cap B) = 1/52$

연습문제 3.8 이산확률분포의 기댓값과 확률

(a) $\mathbb{E}(X) = 2(0.1) + 3(0.3) + 4(0.3) + 5(0.2) + 6(0.1) = 3.9$

(b) $\mathbb{P}(X \geq 4) = 0.3 + 0.2 + 0.1 = 0.6, \quad \mathbb{P}(2 \leq X \leq 4) = 0.1 + 0.3 + 0.3 = 0.7.$

연습문제 3.9 자격시험 합격 수험생 수의 분포

(a) 각 시험 응시자의 합격 여부가 독립적임으로,

x	0	1	2	3
$\mathbb{P}(X = x)$	$(0.2)^3 = 0.008$	$3(0.8)(0.2)^2 = 0.096$	$3(0.8)^2(0.2) = 0.384$	$(0.8)^3 = 0.512$

참고 (1): X의 확률질량함수 (pmf) $p_X(x)$는 다음과 같다.

$$p_X(x) = \mathbb{P}(X = x) = \binom{3}{x}(0.8)^x(0.2)^{3-x}$$

참고 (2): 4장에서 X가 이항분포임을 확인할 수 있다. $(X \sim B(3, 0.8))$

(b)

$$\mathbb{E}(X) = 0 \times 0.008 + 1 \times 0.096 + 2 \times 0.384 + 3 \times 0.512 = 2.4$$
$$\mathbb{E}(X^2) = 0^2 \times 0.008 + 1^2 \times 0.096 + 2^2 \times 0.384 + 3^2 \times 0.512 = 6.24$$
$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 0.48$$

연습문제 3.11 연속확률분포의 기댓값과 확률

(a) 확률밀도함수의 적분값은 항상 1이여야 함으로, $\displaystyle\int_0^4 k(4 - x)\,dx = 8k = 1$
$\therefore k = \dfrac{1}{8}.$

(b) $P(1 \leq X \leq 2) = \displaystyle\int_1^2 \frac{1}{8}(4 - x)dx = 0.3125.$

(c)

$$\mathbb{E}(X) = \int_0^4 x\frac{1}{8}(4-x)dx = \frac{1}{8}\left[2x^2 - \frac{x^3}{3}\right]_0^4 = \frac{4}{3},$$
$$\mathbb{E}(X^2) = \int_0^4 x^2\frac{1}{8}(4-x)dx = \frac{8}{3},$$
$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{8}{3} - \left(\frac{4}{3}\right)^2 = \frac{8}{9}.$$

연습문제 3.14 온라인 쇼핑몰의 상품 등록/반려 분포

(a)

x	$y = 5 - x$	$\mathbb{P}(X = x, Y = y)$
0	5	$\binom{5}{0}(0.8)^0(0.2)^5 = 0.00032$
1	4	$\binom{5}{1}(0.8)^1(0.2)^4 = 0.0064$
2	3	$\binom{5}{2}(0.8)^2(0.2)^3 = 0.0512$
3	2	$\binom{5}{3}(0.8)^3(0.2)^2 = 0.2048$
4	1	$\binom{5}{4}(0.8)^4(0.2)^1 = 0.4096$
5	0	$\binom{5}{5}(0.8)^5(0.2)^0 = 0.32768$

이외 자연수 x, y값에 대해서는 확률값이 항상 0

참고 (1): 고등학교 교육과정에서 배운 조합 기호는 이후 다음과 같이 작성한다.

$$\binom{n}{r} := {}_n\mathrm{C}_r$$

참고 (2): X와 Y의 결합확률질량함수 (Joint pmf) $p_{X,Y}(x, y)$는 다음과 같다.

$$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y) = \binom{5}{x}(0.8)^x(0.2)^y I(x + y = 5)$$

참고 (3): 지시함수 (Indicator function)는 다음과 같이 정의한다.

$$I(A) = \begin{cases} 1, & \text{명제 } A\text{가 참} \\ 0, & \text{otherwise.} \end{cases}$$

(b)

x	$\mathbb{P}(X=x)$
0	$\binom{5}{0}(0.8)^0(0.2)^5 = 0.00032$
1	$\binom{5}{1}(0.8)^1(0.2)^4 = 0.0064$
2	$\binom{5}{2}(0.8)^2(0.2)^3 = 0.0512$
3	$\binom{5}{3}(0.8)^3(0.2)^2 = 0.2048$
4	$\binom{5}{4}(0.8)^4(0.2)^1 = 0.4096$
5	$\binom{5}{5}(0.8)^5(0.2)^0 = 0.32768$

y	$\mathbb{P}(Y=y)$
0	$\binom{5}{0}(0.2)^0(0.8)^5 = 0.32768$
1	$\binom{5}{1}(0.2)^1(0.8)^4 = 0.4096$
2	$\binom{5}{2}(0.2)^2(0.8)^3 = 0.2048$
3	$\binom{5}{3}(0.2)^3(0.8)^2 = 0.0512$
4	$\binom{5}{4}(0.2)^4(0.8)^1 = 0.0064$
5	$\binom{5}{5}(0.2)^5(0.8)^0 = 0.00032$

참고: 연습문제 3.9와 마찬가지로, 4장에서 X와 Y가 이항분포임을 확인할 수 있다. ($X \sim B(5, 0.8), Y \sim B(5, 0.2)$)

(c)

$$\mathbb{E}(X) = 0 \times 0.00032 + 1 \times 0.0064 + 2 \times 0.0512 + 3 \times 0.2048 + 4 \times 0.4096 + 5 \times 0.32768 = 4$$

$$\mathbb{E}(X^2) = 0^2 \times 0.00032 + 1^2 \times 0.0064 + 2^2 \times 0.0512 + 3^2 \times 0.2048 + 4^2 \times 0.4096 + 5^2 \times 0.32768 = 16.8$$

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 0.8$$

$Y = 5 - X$임으로 이를 활용하면,

$$\mathbb{E}(Y) = \mathbb{E}(5 - X) = 5 - \mathbb{E}(X) = 1, \quad \mathrm{Var}(Y) = \mathrm{Var}(5 - X) = (-1)^2 \,\mathrm{Var}(X) = 0.8$$

(d)

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}[X(5 - X)] - 4 \times 1 = 5\mathbb{E}(X) - \mathbb{E}(X^2) - 4 = -0.8$$

$$\mathrm{Corr}(X, Y) = \frac{-0.8}{\sqrt{0.8 \times 0.8}} = -1$$

참고: 모상관계수를 계산할 때 계산식을 쓰지 않고도 이미 $Y = 5 - X$이기에 음의 선형상관관계를 가지므로 -1임을 바로 구할 수도 있다.

연습문제 3.15 결합확률분포와 독립

(a)

x	1	2	3
$\mathbb{P}(X=x)$	$\frac{1}{20}+\frac{2}{20}+\frac{3}{20}=\frac{6}{20}$	$\frac{4}{20}+0+\frac{3}{20}=\frac{7}{20}$	$\frac{2}{20}+\frac{4}{20}+\frac{1}{20}=\frac{7}{20}$

y	1	5	10
$\mathbb{P}(Y=y)$	$\frac{1}{20}+\frac{4}{20}+\frac{2}{20}=\frac{7}{20}$	$\frac{2}{20}+0+\frac{4}{20}=\frac{6}{20}$	$\frac{3}{20}+\frac{3}{20}+\frac{1}{20}=\frac{7}{20}$

(b)

$$\begin{aligned}
\mathbb{E}(X) &= \frac{6+14+21}{20} = 2.05 \\
\mathbb{E}(X^2) &= \frac{6+28+63}{20} = 4.85 \\
\mathrm{Var}(X) &= 4.85 - (2.05)^2 = 0.6475 \\
\mathbb{E}(Y) &= \frac{7+30+70}{20} = 5.35 \\
\mathbb{E}(X+Y) &= E(X) + E(Y) = 7.4
\end{aligned}$$

(c) $\mathbb{P}(X=1, Y=1) = 0.05 \neq \mathbb{P}(X=1) \times \mathbb{P}(Y=1) = 0.3 \times 0.35 = 0.105$
따라서 X와 Y는 독립이 아니다.

참고(1): 두 분포가 독립이 아님을 보일 땐 반례 한가지만 보이면 되지만, 두 분포가 독립임을 보일 땐 모든 x, y에 대하여 $\mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x)\mathbb{P}(Y=y)$ 임을 보여야 한다.

참고(2): $\mathbb{E}(XY) \neq \mathbb{E}(X)\mathbb{E}(Y)$ 이므로 독립이 아님을 보이는 풀이도 맞다. 다만, X와 Y가 독립이면 $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$이지만, 그 역($\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$이면 X와 Y가 독립)은 성립하지 않는다.

4장

연습문제 4.2 공정한 주사위

① 다면체 주사위 한 면이 2일 확률을 $p = 1/6$이라 할 때, 5번 던져서 한 번도 2가 나오지 않을 확률은 $(5/6)^5$이므로,

$$\mathbb{P}(①) = 1 - (5/6)^5 \approx 0.5981$$

② 6번 던져 모든 숫자가 한 번씩 나올 확률은 각 순열 6! 가지가 모두 6^6 경우 중 유리하므로,

$$\mathbb{P}(②) = \frac{6!}{6^6} \approx 0.0154$$

따라서 ① 사건의 확률이 더 크다.

연습문제 4.7 자동차 연비 조사

광고한 모델의 연비는 평균 20 km/L, 표준편차 3 km/L인 정규분포를 따른다. 표본 36대의 $\bar{X}$는 $N\big(20, (3/\sqrt{36})^2\big) = N(20, 0.5^2)$

(a)

$$\mathbb{P}(19 \le \bar{X} \le 21) = \Phi\big(\frac{21-20}{0.5}\big) - \Phi\big(\frac{19-20}{0.5}\big) = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 \approx 0.9545$$

(b) 측정값이 18.5 km/L라면 $z = (18.5-20)/0.5 = -3$으로, 추론 논리를 활용한다면 $\mathbb{P}(\bar{X} \le 18.5) = \Phi(-3) \approx 0.0013$ 이므로 광고가 과장되었다고 판단할 수 있다.

연습문제 4.8 스마트 농장

씨앗 크기 $X \sim N(\mu, \sigma^2)$에 대해 $\mathbb{P}(X \ge 10) = 0.10,\ \mathbb{P}(X \le 3) = 0.15$ 표준화하여

$$\frac{10-\mu}{\sigma} = z_{0.90} \approx 1.2816, \quad \frac{3-\mu}{\sigma} = z_{0.15} \approx -1.0364.$$

이를 풀면

$$\sigma = \frac{10-3}{1.2816+1.0364} \approx 3.02, \quad \mu = 10 - 1.2816\,\sigma \approx 6.13.$$

연습문제 4.10 초기하분포

(a) 주어진 확률변수는 초기하분포를 따른다. $\therefore X \sim H(5,2,2)$

(b) $\mathbb{E}[X] = \frac{2}{5} \times 2 = 0.8,\ \mathrm{Var}(X) = 2 \times \frac{2}{5} \times \frac{3}{5} \times \frac{3}{4} = 0.36$

연습문제 4.13 이항분포와 정규근사

3개의 동전을 던져 앞면 2개, 뒷면 1개가 나올 확률은 $p = \frac{3}{8}$. 이를 활용하면 X의 분포는 이항분포를 따른다. $\therefore X \sim B(600, \frac{3}{8})$.

(a) $\mathbb{E}[X] = 600 \times \frac{3}{8} = 225,\ sd(X) = \sqrt{600 \times \frac{3}{8} \times \frac{5}{8}} \approx 11.85.$

(b) $np = 225 \geq 5, n(1-p) = 375 \geq 5$ 이므로 정규근사를 이용해 확률을 구할 수 있다.
$\mathbb{P}(X \leq 220) \approx \Phi\big((220.5 - 225)/11.85\big) \approx \Phi(-0.38) \approx 0.35.$

(c) $\mathbb{P}(X > 250) \approx 1 - \Phi\big((250.5 - 225)/11.85\big) \approx 1 - \Phi(2.15) \approx 0.0156.$

참고: 일부 학생분들이 연속성 수정을 언제 해야하는지에 대해 궁금해하셨습니다. 이항분포를 정규근사할 때 연속성 수정은 $\sqrt{npq}$값이 크지 않을 때 주로 사용합니다. 위 연습문제와 다음 연습문제 (4.13, 4.16)은 굳이 연속성 수정을 하지 않아도 충분히 근사할 수 있기 때문에 연속성 수정 없이 근사해야 합니다.

연습문제 4.16 전구 불량품 확률

X_1, X_2는 각각 $B(10000, 0.005)$을 따르는 확률변수이다.

(a) $np = 50 \geq 5, n(1-p) = 9995 \geq 5$ 이므로 정규근사를 이용해 확률을 구할 수 있다.
$\mu = 50, \sigma = \sqrt{10000 \times 0.005 \times 0.995} \approx 7.05.$

$$\mathbb{P}(X_1 \geq 40) = 1 - \Phi\big((39.5 - 50)/7.05\big) \approx 1 - \Phi(-1.49) \approx 0.93.$$

(b) 문제의 조건이 약간 부족하다. X_1, X_2가 서로 독립 이라는 조건이 있다면, $X_1 + X_2 \sim B(20000, 0.005)$ 이 또한 정규근사를 할 수 있다.

$$\mathbb{P}(X_1 + X_2 \geq 80) = 1 - \Phi\big((79.5 - 100)/9.99\big) \approx 1 - \Phi(-2.06) \approx 0.98.$$

연습문제 4.17 이산분포 표본평균

(a) $\mathbb{E}[X] = 0 \cdot 0.1 + 1 \cdot 0.5 + 2 \cdot 0.4 = 1.3, \quad \mathrm{Var}(X) = 2.1 - 1.3^2 = 0.41, sd(X) = \sqrt{0.41} \approx 0.64.$

(b) 표본평균 $\bar{X} = (X_1 + X_2)/2$의 분포:

x	0	0.5	1	1.5	2
$\mathbb{P}(\bar{X} = x)$	0.01	0.10	0.33	0.40	0.16

(c)

$$\mathbb{E}[\bar{X}] = 0.5 \times 0.1 + 1 \times 0.33 + 1.5 \times 0.4 + 2 \times 0.16 = 1.3$$
$$\mathbb{E}[(\bar{X})^2] = 0.25 \times 0.1 + 1 \times 0.33 + 2.25 \times 0.4 + 4 \times 0.16 = 1.895$$
$$\mathrm{Var}(\bar{X}) = 1.895 - (1.3)^2 = 0.205 \Rightarrow sd(\bar{X}) = \sqrt{0.205} \approx 0.4528$$

원분포에 비해 표준편차가 감소함.

참고: 표본수가 작아 엄밀하게 중심극한정리를 도입할 수 없지만, $\frac{\mathrm{Var}(X)}{\mathrm{Var}(\bar{X})} = n = 2$ 임을 참고할 수 있다.

연습문제 4.18 정규모집단 표본분산

모집단 $N(1, 2^2)$에서 $n = 10$. 표본분산 S^2에 대해 $(n-1)S^2/\sigma^2 \sim \chi^2(9)$

(a) $\mathbb{P}(S^2 \leq y) = 0.95$ 이므로,

$$y = \frac{\sigma^2 \chi^2_{0.05}(9)}{n-1} = \frac{4 \times 16.919}{9} \approx 7.52.$$

(b) 표본평균 $\bar{X}$와 S로 $(\bar{X}-1)/(S/\sqrt{10}) \sim t(9)$ 이며

$$P\big(\bar{X} \leq 1 + yS\big) = 0.90 \implies y = \frac{t_{0.1}(9)}{\sqrt{10}} \approx \frac{1.383}{3.162} \approx 0.44$$

연습문제 4.19 두집단 평균차와 분산비

(a) $T = \dfrac{\bar{Y} - \bar{X} - (2-1)}{S_p\sqrt{1/10 + 1/15}} \sim t(23)$

$$P\big(\bar{Y} - \bar{X} \leq 1 + yS_p\big) = 0.90 \implies y = t_{0.1}(23)\sqrt{\tfrac{1}{10} + \tfrac{1}{15}} \approx 1.319 \times 0.408 \approx 0.54.$$

(b) 분산비 $F = \dfrac{S_1^2/3^2}{S_2^2/3^2} = \dfrac{S_1^2}{S_2^2} \sim F(9, 14)$

$$P\big(S_1^2 \leq y\, S_2^2\big) = 0.90 \implies y = F_{0.90}(9, 14) \approx 2.03.$$

연습문제 4.24 두표본분산 비 확률

A표본 $n_A = 30, \sigma_A^2 = 2$, B표본 $n_B = 40, \sigma_B^2 = 3$. 또한 A와 B가 독립이므로, $U = \frac{S_A^2/2}{S_B^2/3} \sim F_{29,39}$

$$P\big(1 \leq \frac{S_A^2}{S_B^2} \leq 2\big) = P\Big(\tfrac{3}{2} \leq U \leq 3\Big) = F_{29,39}(3) - F_{29,39}(1.5)$$

5장

참고: 검정의 설계

검정을 수행하기 전에 설계 과정은 다음 네 단계를 모두 포함해야 합니다.

1. 귀무가설과 대립가설 (H_0 vs H_1)을 명확히 세우기
2. 검정통계량을 정의하기
3. 귀무가설 하의 통계량 분포를 확인하기 (어떤 분포를 따르는지, 자유도는 얼마인지)
4. 기각역을 설정하기

연습문제 5.1 표본평균의 표준편차와 신뢰구간

(a) 표본평균의 표준편차(표준오차)는 다음과 같이 계산된다

$$\mathrm{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{25}} = \frac{5}{5} = 1 \text{ (cm)}$$

(b) 모분산 σ^2를 알고 있으므로, 모평균 μ에 대한 $100(1-\alpha)\%$ 신뢰구간은 다음과 같다:

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

95% 신뢰구간의 경우 $\alpha = 0.05$이므로 $z_{\alpha/2} = z_{0.025} = 1.96$이다.
따라서 95% 신뢰구간은

$$(174 - 1.96 \times 1, 174 + 1.96 \times 1) = (172.04, 175.96) \text{ (cm)}$$

참고: 신뢰구간의 의미는 "동일한 방법으로 100번 표본을 추출하여 신뢰구간을 구하면, 그 중 약 95개의 신뢰구간이 모평균 μ를 포함한다"는 것이다.

연습문제 5.3 배터리 충전 시간의 신뢰구간과 표본크기

(a) 표본평균의 표준오차:

$$\mathrm{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{50}} = \frac{8}{7.071} \approx 1.131 \text{ (분)}$$

99% 신뢰구간:

$$\begin{aligned} &\left(\bar{x} - z_{0.005} \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + z_{0.005} \cdot \frac{\sigma}{\sqrt{n}}\right) \\ &\approx (45 - 2.576 \times 1.131, \quad 45 + 2.576 \times 1.131) \\ &\approx (45 - 2.913, 45 + 2.913) = (42.087, 47.913) \text{ (분)} \end{aligned}$$

(b) 오차한계(Margin of Error)를 d 이하로 하기 위한 표본 크기 n은 다음 조건을 만족해야 한다:

$$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq d$$

이를 n에 대해 정리하면:

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{d}\right)^2$$

95% 신뢰수준에서 $z_{0.025} = 1.96$, $\sigma = 8$분, $d = 2$분이므로:

$$n \geq \left(\frac{1.96 \times 8}{2}\right)^2 = \left(\frac{15.68}{2}\right)^2 = (7.84)^2 \approx 61.47$$

표본 크기는 정수여야 하므로 위 값을 올림하여 최소 $\boxed{n = 62}$개가 필요하다.

연습문제 5.5 귀무가설과 대립가설의 설정

(a) μ: 신제품 음료의 평균 당 함량 (g)
음료회사가 증명하고자 하는 것(대립가설)은 "평균 당 함량이 10g 미만"이다.

$$\boxed{H_0 : \mu \geq 10 \quad \text{vs.} \quad H_1 : \mu < 10}$$

또는 등호를 귀무가설에 포함하여:

$$H_0 : \mu = 10 \quad \text{vs.} \quad H_1 : \mu < 10$$

(b) μ: 아파트의 평균 면적 (m^2)
입주자들이 증명하고자 하는 것(대립가설)은 "실제 평균 면적이 85m^2보다 작다"이다.

$$\boxed{H_0 : \mu \geq 85 \quad \text{vs.} \quad H_1 : \mu < 85}$$

또는:

$$H_0 : \mu = 85 \quad \text{vs.} \quad H_1 : \mu < 85$$

(c) μ_m: 남학생의 체육 과목 선호도 (또는 체육 선호 비율)
μ_f: 여학생의 체육 과목 선호도 (또는 체육 선호 비율)
연구원이 증명하고자 하는 것(대립가설)은 "남학생이 여학생보다 체육 과목을 더 선호한다"이다.

$$\boxed{H_0 : \mu_m \leq \mu_f \quad \text{vs.} \quad H_1 : \mu_m > \mu_f}$$

또는 차이를 이용하면:

$$H_0 : \mu_m - \mu_f \leq 0 \quad \text{vs.} \quad H_1 : \mu_m - \mu_f > 0$$

참고: 답안을 작성할 시 가설을 설계할 때 사용하는 기호(위의 경우 μ, μ_m, μ_f)의 의미를 명확히 작성해야 합니다.

연습문제 5.7 DRP 점수 가설검정

(a) μ: 이 학교 초등학교 3학년 학생들의 평균 DRP 점수
교육학자가 증명하고자 하는 것(대립가설)은 "이 학교의 평균 DRP 점수가 32점보다 높다"이다.

$$\boxed{H_0 : \mu = 32 \quad \text{vs.} \quad H_1 : \mu > 32}$$

(b) 주어진 상황은 모분산을 알고 있는 경우의 한 모평균에 대한 검정 상황이다. 유의수준 $\alpha = 0.05$로 설정하여 검정을 설계하면 다음과 같다.

① 가설: $H_0 : \mu = 32$, $H_1 : \mu > 32$

② 검정통계량: $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

③ 귀무가설 하의 분포: $Z \sim N(0, 1)$ under H_0

④ 검정

- 표본수: $n = 44$

- 관측값 계산:

$$\sum_{i=1}^{44} x_i = 1544 \Rightarrow \bar{x} = \frac{1544}{44} \approx 35.0909$$

- 검정통계량의 관측값 계산:

$$z = \frac{35.0909 - 32}{11/\sqrt{44}} = \frac{3.0909}{1.659} \approx 1.86$$

- 기각역: $z \geq z_{0.05} = 1.645$
- 유의확률: p-value $= P(Z \geq 1.86) = 1 - \Phi(1.86) \approx 0.0314$

⑤ 결론: 검정통계량 값 $z = 1.86$가 기각역에 속하므로 $(1.86 > 1.645)$, 유의수준 5%에서 귀무가설 H_0를 **기각**한다. 또한 유의확률 $p = 0.0314 < \alpha = 0.05$이므로 동일한 결론을 얻는다. 따라서 이 학교의 평균 DRP 점수가 교육청 전체 평균 32점보다 높다는 교육학자의 주장을 지지하는 통계적 증거가 충분하다.

연습문제 5.10 신발 수명 가설검정

(a) μ: 새로운 공정으로 생산된 신발의 평균 수명 (마일)

주어진 상황은 모분산을 알고 있는 경우의 한 모평균에 대한 검정 상황이다. 회사가 증명하고자 하는 것(대립가설)은 “새로운 공정이 신발의 평균 수명을 늘렸다”이다. 유의수준 $\alpha = 0.05$로 설정하여 검정을 설계하면 다음과 같다.

① 가설: $H_0 : \mu = 450$, $H_1 : \mu > 450$

② 검정통계량: $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

③ 귀무가설 하의 분포: $Z \sim N(0, 1)$ under H_0

④ 검정

- 표본수: $n = 25$, 관측값: $\bar{x} = 478$마일, $\sigma = 50$마일
- 검정통계량의 관측값 계산:

$$z = \frac{478 - 450}{50/\sqrt{25}} = \frac{28}{10} = 2.8$$

- 기각역: $z \geq z_{0.05} = 1.645$

⑤ 결론: 검정통계량 값 $z = 2.8$이 기각역에 속하므로 $(2.8 > 1.645)$, 유의수준 5%에서 귀무가설 H_0를 **기각**한다. 따라서 새로운 공정이 신발의 평균 수명을 증가시켰다는 통계적 증거가 충분하다.

(b) 최소 유의수준은 유의확률(p-value)과 같다.

$$p\text{-value} = P(Z \geq 2.8) = 1 - \Phi(2.8) \approx 1 - 0.9974 = \boxed{0.0026}$$

(c) 제2종 오류(β)는 대립가설이 참일 때 귀무가설을 채택하는 오류이다.

유의수준 $\alpha = 0.05$에서 오른쪽 단측검정의 기각역은 $\bar{X} \geq c$이며, 이를 만족하는 임계값 c는

$$c = \mu_0 + z_{0.05} \cdot \frac{\sigma}{\sqrt{n}} = 450 + 1.645 \times \frac{50}{\sqrt{25}} = 450 + 1.645 \times 10 = 466.45$$

$\mu = 470$일 때 제2종 오류를 범할 확률:

$$
\begin{aligned}
\beta &= P(\bar{X} < 466.45 \mid \mu = 470) \\
&= P\left(Z < \frac{466.45 - 470}{50/\sqrt{25}}\right) \\
&= P\left(Z < \frac{-3.55}{10}\right) \\
&= P(Z < -0.355) \\
&= \Phi(-0.355) \approx \boxed{0.3613}
\end{aligned}
$$

(d) 검정력(Power) $= 1 - \beta = 0.95$이므로 $\beta = 0.05$이다.
유의수준 $\alpha = 0.01$에서 오른쪽 단측검정의 임계값은

$$c = \mu_0 + z_{0.01} \cdot \frac{\sigma}{\sqrt{n}} = 450 + 2.326 \cdot \frac{50}{\sqrt{n}}$$

$\mu_1 = 460$일 때 제2종 오류 확률 $\beta = 0.05$은

$$\beta = P(\bar{X} < c \mid \mu = 460) = 0.05$$

이는 다음을 의미한다

$$P\left(Z < \frac{c - 460}{50/\sqrt{n}}\right) = 0.05$$

$\Phi(z) = 0.05$인 $z = -z_{0.05} = -1.645$이므로

$$\frac{c - 460}{50/\sqrt{n}} = -1.645$$

두 식을 연립하면

$$
\begin{aligned}
c &= 450 + 2.326 \cdot \frac{50}{\sqrt{n}} \\
c &= 460 - 1.645 \cdot \frac{50}{\sqrt{n}}
\end{aligned}
$$

따라서 연립방정식을 풀면

$$450 + 2.326 \cdot \frac{50}{\sqrt{n}} = 460 - 1.645 \cdot \frac{50}{\sqrt{n}}$$

$$(2.326 + 1.645) \cdot \frac{50}{\sqrt{n}} = 10$$

$$3.971 \cdot \frac{50}{\sqrt{n}} = 10$$

$$\sqrt{n} = \frac{3.971 \times 50}{10} = 19.855$$

$$n = (19.855)^2 \approx 394.2$$

따라서 필요한 표본 크기는 최소 $\boxed{n = 395}$개이다.
참고: 표본 크기 공식을 직접 사용하면

$$n \geq \left(\frac{(z_\alpha + z_\beta) \cdot \sigma}{\mu_1 - \mu_0}\right)^2 = \left(\frac{(2.326 + 1.645) \times 50}{460 - 450}\right)^2 = \left(\frac{198.55}{10}\right)^2 \approx 394.2$$

6장

연습문제 6.1 기각역 구하기

문제의 정보가 부족하지만, 셋 모두 이표본에 의한 모평균의 비교 상황으로 볼 수 있다.

(a) 귀무가설 $H_0 : \mu_1 = \mu_2$, 표본 크기 $n_1 = 512, n_2 = 515$로 매우 크므로
통계량 $T = \dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$은 귀무가설 하에 표준정규분포로 근사할 수 있다.
이때 $H_1 : \mu_1 \neq \mu_2$의 유의수준 $\alpha = 0.05$의 기각역은 $|t| \geq z_{0.025}$이다.

(b) $n_1 = n_2 = 25$임으로 표본 개수가 적다. 정보가 부족하지만,
<u>두 모집단의 분포가 정규분포이고 등분산이라는 추가 조건</u>을 준다면,
통계량 $T = \dfrac{\left(\overline{X_1} - \overline{X_2}\right) - 3}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$은 귀무가설 하에 $t\,(n_1 + n_2 - 2)$를 따른다. 이때 $H_1 : \mu_1 > \mu_2 + 3$의 유의수준 $\alpha = 0.01$의 기각역은 $t \geq t_{0.01}(48)$이다.

(c) $n_1 = n_2 = 25$임으로 표본 개수가 적다. 이 문제 역시 정보가 부족하지만,
<u>두 모집단의 분포가 정규분포이고 등분산이라는 추가 조건</u>을 준다면, 이 역시
통계량 $T = \dfrac{\left(\overline{X_1} - \overline{X_2}\right) - (-9)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$은 귀무가설 하에 $t\,(n_1 + n_2 - 2)$를 따른다. 이때 $H_1 : \mu_1 < \mu_2 - 9$의 유의수준 $\alpha = 0.025$의 기각역은 $t \leq -t_{0.025}(48)$이다.

연습문제 6.2 눈의 적응시간

주어진 상황은 한 모평균에 대한 추론 상황이다. 평균 적응 시간을 μ라 하고 이전 적응 시간을 $\mu_0 = 7$이라 하면,

① 가설 : $H_0 : \mu = 7,\ H_1 : \mu \neq 7$

② 표본수 : $n = 9$

③ 검정통계량 : $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$ under H_0

④ 검정

- 관측값: $\bar{x} = 6.32, s = 1.65$
- 검정통계량의 관측값 계산 : $t = \frac{6.32 - 7}{1.65/\sqrt{9}} \approx -1.2364$
- 기각역 : $|t| \geq t_{0.05}(8) \approx 1.86$

⑤ 결론 : 검정통계량 값 -1.2364가 기각역에 속하지 않으므로 ($|-1.2364| < 1.86$), 귀무가설 H_0를 기각하지 못한다. 따라서 이 실험 결과가 기존의 연구결과와 모순된다고 말할 수 없다.

연습문제 6.5 골프 드라이버 비교

주어진 상황은 대응비교에 의한 모평균의 비교 상황이다. 어떤 드라이버가 다른 드라이버보다 비거리가 잘 나온다는 말은 차이 $d_i = A_i - B_i$가 0이 아님을 의미하기에 다음과 같이 검정을 설계할 수 있다.

① 가설: $H_0 : \mu_d = 0,\ H_1 : \mu_d \neq 0$

② 표본수 : $n = 14$

③ 검정통계량: $T = \dfrac{\bar{D}}{s_d/\sqrt{n}} \sim t(n-1)$ under H_0

④ 검정

- 관측값: $\bar{d} \approx -2.4643, s \approx 7.4742$
- 검정통계량의 관측값 계산 : $t = \frac{-2.4643}{7.4742/\sqrt{14}} \approx -1.2336$
- 기각역 : $|t| \geq t_{0.05}(13) \approx 1.771$

⑤ 결론 : 검정통계량 값 -1.2336가 기각역에 속하지 않으므로 ($|-1.2336| < 1.771$), 귀무가설 H_0를 기각하지 못한다. 따라서 특정 드라이버가 잘 나온다고 말할 통계적 근거가 없다.

연습문제 6.7 서비스 효과 비교

서비스 A, B를 사용해 달리기 시 소모되는 칼로리의 모평균을 μ_A, μ_B라 하고, $\mu_d = \mu_A - \mu_B$라 하자.

(a) 대응비교에 의한 두 모평균의 비교를 하기 위해 다음과 같은 검정을 설계할 수 있다.

① 가설: $H_0 : \mu_d = 0$, $H_1 : \mu_d \neq 0$

② 표본수 : $n = 16$

③ 검정통계량: $T_a = \dfrac{\bar{D}}{s_d/\sqrt{n}} \sim t(n-1)$ under H_0

④ 검정

- 관측값: $\bar{d} = -11.9375, s \approx 23.6607$
- 검정통계량의 관측값 계산 : $t_a = \frac{-11.9375}{23.6607/\sqrt{16}} \approx -2.0181$
- 기각역 : $|t_a| \geq t_{0.005}(15) \approx 2.947$

⑤ 결론 : 검정통계량 값 -2.0181가 기각역에 속하지 않으므로 ($|-2.0181| < 2.947$), 귀무가설 H_0를 기각하지 못한다. 따라서 특정 서비스에 따라 소모 칼로리가 다르다고 말할 통계적 근거가 없다.

(b) (a)와 같은 유의수준 $\alpha = 0.01$에서 등분산 가정 이표본에 의한 모평균의 비교 검정을 설계하면 다음과 같다.

① 가설: $H_0 : \mu_d = 0$, $H_1 : \mu_d \neq 0$

② 표본수 : $n = 16$

③ 검정통계량: $T_b = \dfrac{\bar{D}}{S_p\sqrt{\dfrac{1}{n} + \dfrac{1}{n}}} \sim t(2n-2)$ under H_0

④ 검정

- 관측값: $\bar{d} = -11.9375, S_A \approx 16.1537, S_B \approx 9.7281, S_p \approx 13.3338$
- 검정통계량의 관측값 계산 : $t = \frac{-11.9375}{13.3338\sqrt{1/16+1/16}} \approx -2.5322$
- 기각역 : $|t| \geq t_{0.005}(30) \approx 2.750$

⑤ 결론 : 검정통계량 값 -2.5322가 기각역에 속하지 않으므로 ($|-2.5322| < 2.750$), 귀무가설 H_0를 기각하지 못한다. 따라서 등분산 가정 하에 특정 서비스에 따라 소모 칼로리가 다르다고 말할 통계적 근거가 없다.

연습문제 6.9 유아의 철분공급

(a) 모유 그룹(X_1)이 유아식 그룹(X_2)보다 헤모글로빈 함유량이 높은지 검정한다.

[**Step 0**] **등분산 검정**: 우선 주어진 상황이 등분산인지 아닌지를 확인하기 위해 먼저 등분산 검정을 수행한다.

① 가설: $H_0 : \sigma_1^2 = \sigma_2^2$, $H_1 : \sigma_1^2 \neq \sigma_2^2$

② 표본수 : $n_1 = 23, n_2 = 19$

③ 검정통계량: $F = \dfrac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$ under H_0

④ 검정

- 관측값: $S_1 = 1.7, S_2 = 1.8$
- 검정통계량의 관측값 계산 : $F = \frac{1.7^2}{1.8^2} \approx 0.8920$
- 기각역 : $F \leq F_{0.975}(22, 18) \approx 0.387$ 또는 $F \geq F_{0.025}(22, 18) \approx 2.58$

⑤ 결론 : 검정통계량 값 0.8920가 기각역에 속하지 않으므로 ($0.387 < 0.8920 < 2.58$), 귀무가설 H_0를 기각하지 못한다. 따라서 두 그룹의 모분산이 다르다고 말할 통계적 근거가 없으므로, 등분산 가정 하에 검정을 진행할 수 있다.

[**본 검정**] **평균 비교**: 등분산 가정이 유효하므로 pooled t-검정을 수행한다.

① 가설: $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 > \mu_2$

② 표본수 : $n_1 = 23, n_2 = 19$

③ 검정통계량: $T = \dfrac{\overline{X_1} - \overline{X_2}}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$ under H_0

④ 검정

- 관측값: $\overline{X_1} = 13.3, \overline{X_2} = 12.4, S_1 = 1.7, S_2 = 1.8, S_p \approx 1.7457$
- 검정통계량의 관측값 계산 : $t = \frac{13.3 - 12.4}{1.7457\sqrt{1/n_1 + 1/n_2}} \approx 1.6630$
- 기각역 : $t \geq t_{0.05}(40) \approx 1.684$

⑤ 결론 : 검정통계량 값 1.6630가 기각역에 속하지 않으므로 ($1.6630 < 1.684$), 귀무가설 H_0를 기각하지 못한다. 따라서 등분산 가정 하에 모유로 자란 아기들의 평균 헤모글로빈 함유량이 더 높다고 말할 통계적 근거가 없다.

(b) 두 그룹의 평균 헤모글로빈 차이에 대한 95% 신뢰구간은 다음과 같다.

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}(n_1 + n_2 - 2)s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$t_{0.025}(40) \approx 2.021$ 이므로,

$$(13.3 - 12.4) \pm 2.021 \times 1.746\sqrt{\frac{1}{23} + \frac{1}{19}} \Rightarrow 0.9 \pm 1.094 \Rightarrow (-0.194, 1.994)$$

(c) (a)와 (b)의 질문에 답하기 위해 필요한 가정은 다음과 같다: 두 모집단은 정규분포를 따르고, 두 표본은 서로 독립이며, 두 모집단의 분산은 동일하다.

연습문제 6.11 SNS 광고의 효과

SNS 광고 그룹(X_1)이 비광고 그룹(X_2)보다 판매량이 많은지 검정한다.

[Step 0] 등분산 검정: 우선 주어진 상황이 등분산인지 아닌지를 확인하기 위해 먼저 등분산 검정을 수행한다.

① 가설: $H_0 : \sigma_1^2 = \sigma_2^2$, $H_1 : \sigma_1^2 \neq \sigma_2^2$

② 표본수 : $n_1 = 13, n_2 = 21$

③ 검정통계량: $F = \dfrac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$ under H_0

④ 검정

- 관측값: $S_1 = 7.23, S_2 = 6.98$
- 검정통계량의 관측값 계산 : $F = \frac{7.23^2}{6.98^2} \approx 1.0730$
- 기각역 : $F \leq F_{0.975}(12, 20) \approx 0.359$ 또는 $F \geq F_{0.025}(12, 20) \approx 2.68$

⑤ 결론 : 검정통계량 값 1.0730가 기각역에 속하지 않으므로 ($0.359 < 1.0730 < 2.68$), 귀무가설 H_0를 기각하지 못한다. 따라서 두 그룹의 모분산이 다르다고 말할 통계적 근거가 없으므로, 등분산 가정 하에 검정을 진행할 수 있다.

[본 검정] 평균 비교: 등분산 가정이 유효하므로 pooled t-검정을 수행한다.

① 가설: $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 > \mu_2$

② 표본수 : $n_1 = 13, n_2 = 21$

③ 검정통계량: $T = \dfrac{\overline{X_1} - \overline{X_2}}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$ under H_0

④ 검정

- 관측값: $\overline{X_1} = 58.3, \overline{X_2} = 50.6, S_1 = 7.23, S_2 = 6.98, S_p \approx 7.0748$
- 검정통계량의 관측값 계산 : $t = \frac{58.3-50.6}{7.0748\sqrt{1/n_1+1/n_2}} \approx 3.0840$
- 기각역 : $t \geq t_{0.01}(32) \approx 2.449$

⑤ 결론 : 검정통계량 값 3.0840가 기각역에 속하므로 ($3.0840 \geq 2.449$), 귀무가설 H_0를 기각할 수 있다. 따라서 등분산 가정 하에 SNS 광고의 효과가 있다고 말할 통계적 근거가 있다.

연습문제 6.13 차량의 수리비용 비교

SUV 그룹(X_1)이 승용차 그룹(X_2)보다 수리비용이 더 높은지 검정하기 위해 (a), (b)에 모두 사용되는 가설을 미리 설정하면 다음과 같다.

$$H_0 : \mu_1 = \mu_2,\ H_1 : \mu_1 > \mu_2$$

(a) **공통분산을 가정한 t-검정**:

① 표본수 : $n_1 = n_2 = 8$

② 검정통계량: $T_a = \dfrac{\overline{X_1} - \overline{X_2}}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$ under H_0

③ 검정

- 관측값: $\overline{X_1} = 17.42, \overline{X_2} = 13.11, S_1 = 7.93, S_2 = 5.04, S_p \approx 6.6440$
- 검정통계량의 관측값 계산 : $t = \frac{17.42-13.11}{6.6440\sqrt{1/n_1+1/n_2}} \approx 1.2974$
- 기각역 : $t \geq t_{0.05}(14) \approx 1.761$

④ 결론 : 검정통계량 값 1.2974가 기각역에 속하지 않음으로 ($1.2974 < 1.761$), 귀무가설 H_0를 기각할 수 없다. 따라서 등분산 가정 하에 SUV의 수리비용이 더 높다고 말할 통계적 근거가 없다.

(b) **이표본 t-검정 (Satterthwaite 근사)**:

① 표본수 : $n_1 = n_2 = 8$

② 검정통계량: $T_b = \dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \sim t(df)$ under H_0 where $df = \dfrac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1-1} + \frac{\left(S_2^2/n_2\right)^2}{n_2-1}}$

③ 검정

- 관측값: $\overline{X_1} = 17.42, \overline{X_2} = 13.11, S_1 = 7.93, S_2 = 5.04, df \approx 11.8618 \approx 12$
- 검정통계량의 관측값 계산 : $t = \frac{17.42-13.11}{\sqrt{S_1^2/n_1+S_2^2/n_2}} \approx 1.2974$
- 기각역 : $t \geq t_{0.05}(12) \approx 1.782$

④ 결론 : 검정통계량 값 1.2974가 기각역에 속하지 않음으로 ($1.2974 < 1.782$), 귀무가설 H_0를 기각할 수 없다. 따라서 SUV의 수리비용이 더 높다고 말할 통계적 근거가 없다.

(c) **등분산성 검정**:

① 가설: $H_0 : \sigma_1^2 = \sigma_2^2$, $H_1 :$ *not* $H_0 (\sigma_1^2 \neq \sigma_2^2)$

② 표본수 : $n_1 = n_2 = 8$

③ 검정통계량: $F = \dfrac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$ under H_0

④ 검정

- 관측값: $S_1 = 7.93, S_2 = 5.04$
- 검정통계량의 관측값 계산 : $F = \frac{7.93^2}{5.04^2} \approx 2.4756$
- 기각역 : $F \geq F_{0.025}(7, 7) \approx 4.99$

⑤ 결론 : 검정통계량 값 2.4756가 기각역에 속하지 않음으로 ($2.4756 < 4.99$), 귀무가설 H_0를 기각할 수 없다. 따라서 두 그룹의 (모집단) 표준편차가 다르다고 말할 통계적 근거가 없다.

연습문제 6.15 정산인과 조현병 환자의 뇌세포 대사물질 농도 비교

정상인 그룹(X_1)이 조현병 환자 그룹(X_2)과 비교하는 검정하기 위해 주어진 상황이 등분산인지 아닌지 확인하기 위해 (c) 검정을 먼저 진행해본다.

(c) **등분산성 검정**: 유의수준 $\alpha = 0.05$ 에서 다음 검정을 설계할 수 있다.

① 가설: $H_0 : \sigma_1^2 = \sigma_2^2$, $H_1 :\ \sigma_1^2 > \sigma_2^2$

② 표본수 : $n_1 = n_2 = 9$

③ 검정통계량: $F = \dfrac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$ under H_0

④ 검정

- 관측값: $S_1 = 8.16, S_2 = 6.93$
- 검정통계량의 관측값 계산 : $F = \frac{8.16^2}{6.93^2} \approx 1.3865$
- 기각역 : $F \geq F_{0.05}(8, 8) \approx 3.44$

⑤ 결론 : 검정통계량 값 1.3865가 기각역에 속하지 않음으로 ($1.3865 < 3.44$), 귀무가설 H_0를 기각할 수 없다. 따라서 두 그룹의 (모집단) 표준편차가 다르다고 말할 통계적 근거가 없다.

(d) 두 모분산의 비 σ_1^2/σ_2^2에 대한 90% 신뢰구간:

$$\left(\frac{s_1^2}{s_2^2}\frac{1}{F_{\alpha/2,n_1-1,n_2-1}}, \frac{s_1^2}{s_2^2}\frac{1}{F_{1-\alpha/2,n_1-1,n_2-1}}\right)$$

$s_1^2/s_2^2 \approx 1.386,\ F_{0.05,8,8} \approx 3.44,\ F_{0.95,8,8} = 1/F_{0.05,8,8} \approx 1/3.44 \approx 0.29.$

$$\left(\frac{1.386}{3.44}, \frac{1.386}{0.29}\right) \Rightarrow (0.403, 4.779)$$

(a) (c)에 의해 두 그룹이 등분산이라고 가정하고 검정을 설계할 수 있다.

① 가설 : $H_0 : \mu_1 = \mu_2,\ H_1 : \mu_1 > \mu_2$

② 표본수 : $n_1 = n_2 = 9$

③ 검정통계량: $T_a = \dfrac{\overline{X_1} - \overline{X_2}}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$ under H_0

④ 검정

- 관측값: $\overline{X_1} = 39.8, \overline{X_2} = 35.5, S_1 = 8.16, S_2 = 6.93, S_p \approx 7.5700$
- 검정통계량의 관측값 계산 : $t = \frac{39.8-35.5}{.57\sqrt{1/n_1+1/n_2}} \approx 1.2050$
- 기각역 : $t \geq t_{0.05}(16) \approx 1.746$

⑤ 결론 : 검정통계량 값 1.2050가 기각역에 속하지 않음으로 ($1.2050 < 1.746$), 귀무가설 H_0를 기각할 수 없다. 따라서 등분산 가정 하에 정상인보다 조현병 환자의 평균 뇌세포활동이 저조하다고 말할 통계적 근거가 없다.

(b) 두 그룹 평균 차에 대한 95% 신뢰구간 (등분산 가정):

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0.025,16}s_p\sqrt{1/9 + 1/9} \Rightarrow 4.3 \pm 2.12 \times 7.57 \times \sqrt{2/9} \Rightarrow (-3.26, 11.86)$$

7장

연습문제 7.1 신약의 치유율

신약의 치유율을 p라고 할 때, 기존 약의 치유율 70%보다 높은지 검정하는 상황이다. 판단 기준에 따라 $(np_0 = 14 \geq 5, n(1-p_0) = 6 \geq 5)$, 대표본 근사를 사용한다. 이때 유의수준은 $\alpha = 0.05$로 설정해서 검정을 설계한다면 다음과 같다.

① 가설 : $H_0 : p = p_0 := 0.7,\ H_1 : p > p_0$

② 검정통계량: $Z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)$ under H_0.

③ 검정

- 관측값: $n = 20, X = 16 \implies \hat{p} = 16/20 = 0.8$
- 검정통계량의 관측값 계산 : $z = \dfrac{0.8 - 0.7}{\sqrt{0.7(1-0.7)/20}} \approx 0.9759$
- 기각역 : $z \geq z_\alpha \approx 1.65$

④ 결론: 검정통계량 값 0.9759가 기각역에 속하지 않으므로 $(0.9759 \not\geq 1.65)$, 귀무가설 H_0를 기각할 수 없다. 따라서 새로운 약의 치유율이 기존 약보다 높다고 말할 통계적 근거가 부족하다.

연습문제 7.3 커피 선호도 조사

(a) 별다방 커피를 선호할 확률(p)이 0.5보다 작다고 할 수 있는지 검정한다. $n = 50, p_0 = 0.5$이므로 $np_0 = 25, n(1-p_0) = 25$로 모두 5보다 커 대표본 근사를 사용한다. 이때 유의수준은 $\alpha = 0.05$로 설정해서 검정을 설계한다면 다음과 같다.

① 가설 : $H_0 : p = p_0 := 0.5,\ H_1 : p < p_0$

② 검정통계량: $Z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)$ under H_0.

③ 검정

- 관측값: $n = 50, X = 19 \implies \hat{p} = 19/50 = 0.38$
- 검정통계량의 관측값 계산 : $z = \dfrac{0.38 - 0.5}{\sqrt{0.5(1-0.5)/50}} = \dfrac{-0.12}{0.0707} \approx -1.6971$
- 기각역 : $z \leq -z_\alpha \approx -1.65$

④ 결론: 검정통계량 값 -1.6971이 기각역에 속하므로 $(-1.6971 \leq -1.65)$, 귀무가설 H_0를 기각할 수 있다. 따라서 별다방 커피 선호율이 0.5보다 작다고 할 통계적 근거가 있다.

(b) 번디 커피를 선호하는 사람의 비율에 대한 90% 신뢰구간은 다음과 같다. $(z_{\alpha/2} = z_{0.05} = 1.645)$

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \implies 0.38 \pm 1.645\sqrt{\frac{0.38(0.62)}{50}}$$

$$0.38 \pm 1.645 \times 0.0686 \implies 0.38 \pm 0.1128 \implies (0.2672, 0.4928)$$

연습문제 7.5 프로야구팀 승률 비교

(a) 홈경기와 원정경기에서의 승률($\hat{p}$) 계산

$$\hat{p}_1 = \frac{37}{72} \approx 0.5139, \quad \hat{p}_2 = \frac{37}{70} \approx 0.5286$$

(b) 두 모비율 차의 표준편차는 표준오차이다. 이를 구하면 다음과 같다.

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{0.5139(0.4861)}{72} + \frac{0.5286(0.4714)}{70}} \approx 0.0838$$

(c) 두 모비율의 차에 대한 90% 신뢰구간은 ($z_{0.05} = 1.645$)

$$(\hat{p}_1 - \hat{p}_2) \pm z_{0.05} \times SE(\hat{p}_1 - \hat{p}_2) \implies (0.5139 - 0.5286) \pm 1.645 \times 0.0838$$

$$-0.0147 \pm 0.1378 \implies (-0.1525, 0.1231)$$

연습문제 7.6 프로야구팀 승률 비교 (이어서)

(a) 전체 경기에서의 승률: $\hat{p}_{\text{total}} = \frac{37+37}{72+70} = \frac{74}{142} \approx 0.5211$

(b) '홈경기 승률이 원정경기보다 높다'는 통념을 검정하기 위한 가설은 $H_0 : p_1 = p_2,\ H_1 : p_1 > p_2$ 이다. 이 또한 유의수준은 $\alpha = 0.05$로 설정해서 검정을 설계한다면 다음과 같다.

(c) ① 가설 : $H_0 : p_1 = p_2$, $H_1 : p_1 > p_2$

② 검정통계량: $H_0 : p_1 = p_2$ 하에서는 공통모비율(합동표본비율) $\hat{p} = \frac{X_1+X_2}{n_1+n_2}$을 이용한다.

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} \approx N(0,1) \text{ under } H_0$$

③ 검정

- 관측값: $\hat{p}_1 \approx 0.5139, \hat{p}_2 \approx 0.5286, \hat{p} \approx 0.5211$
- 검정통계량의 관측값 계산 : $z = \frac{0.5139-0.5286}{\sqrt{0.5211(0.4789)(\frac{1}{72}+\frac{1}{70})}} \approx -0.1752$
- 기각역 : $z \geq z_\alpha \approx 1.65$

④ 결론: 검정통계량 값 -0.1752이 기각역에 속하지 않음으로 ($-0.1752 < 1.65$), 귀무가설 H_0를 기각할 수 없다. 따라서 홈경기 승률이 원정경기보다 높다는 통념을 통계적으로 뒷받침하지 못한다.

연습문제 7.10 혈압과 심장발작

두 범주형 변수(혈압 상태, 사망 여부) 간의 연관성을 검정하므로 χ^2 독립성 검정을 실시한다.

① 가설: H_0: 혈압 상태와 사망 여부는 서로 독립이다. vs H_1: 독립이 아니다(연관되어 있다).

② 유의수준: $\alpha = 0.05$ (명시되지 않았으므로)

③ 검정통계량: $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2((r-1)(c-1))$

④ 검정:

- 관측도수표(O_{ij}) 및 기대도수표(E_{ij})

	생존	사망	합계
저혈압	2655 (2648.7)	21 (27.3)	2676
고혈압	3283 (3289.3)	55 (48.7)	3338
합계	5938	76	6014

$E_{11} = (2676 \times 5938)/6014 \approx 2648.7$

- 검정통계량의 관측값 계산 :

$$\chi^2 = \frac{(2655 - 2648.7)^2}{2648.7} + \frac{(21 - 27.3)^2}{27.3} + \frac{(3283 - 3289.3)^2}{3289.3} + \frac{(55 - 48.7)^2}{48.7}$$

$$\chi^2 \approx 0.0150 + 1.4542 + 0.0121 + 0.8152 = 2.2965$$

- 기각역: 자유도 $df = (2-1)(2-1) = 1$, 유의수준 $\alpha = 0.05$에서 $\chi^2 \geq \chi^2_{0.05}(1) = 3.841$.

⑤ 결론: 검정통계량 값 $\chi^2 = 2.2965$가 기각역에 속하지 않으므로($2.2965 < 3.841$), 귀무가설을 기각하지 못한다. 따라서 혈압상태와 심장발작 사망 여부 사이에 통계적으로 유의한 연관성이 있다고 보기 어렵다.

연습문제 7.11 강의평가 동질성 검정

두 모집단(3학년, 4학년)의 강의평가 분포가 동일한지 검정하므로 χ^2 동질성 검정을 실시한다.

① 가설: H_0: 3학년과 4학년 학생의 강의평가 분포는 동일하다. vs H_1: 분포가 동일하지 않다.

② 유의수준: $\alpha = 0.05$

③ 검정통계량: $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2((r-1)(c-1))$

④ 검정:

- 관측도수(O_{ij}) 및 기대도수(E_{ij}) (괄호 안)

	나쁘다	보통이다	좋다	뛰어나다	합계
3학년	20 (19.8)	110 (105.6)	170 (171.6)	30 (33.0)	330
4학년	10 (10.2)	50 (54.4)	90 (88.4)	20 (17.0)	170
합계	30	160	260	50	500

- **검정통계량의 관측값 계산**:

$$\chi^2 \approx 0.0020 + 0.1833 + 0.0149 + 0.2727 + 0.0039 + 0.3559 + 0.0290 + 0.5294 = 1.3911$$

- **기각역 설정**:
  - 자유도 $df = (r-1)(c-1) = (2-1)(4-1) = 3$
  - 유의수준 $\alpha = 0.05$에서 임계값은 $\chi^2_{0.05}(3) = 7.815$
  - 따라서 기각역은 $\chi^2 \geq 7.815$

⑤ 결론: 검정통계량 값 $\chi^2 \approx 1.3911$이 기각역에 속하지 않음으로 ($1.3911 < 7.815$), 귀무가설을 기각할 수 없다. 따라서 3학년과 4학년의 강의평가 분포가 다르다는 통계적 근거가 없다.

8장

연습문제 8.1 잔차에 대한 성질

단순선형회귀모형 $y_i = \alpha + \beta x_i + e_i$에서 최소제곱법으로 구한 회귀식은 $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$ 이다. 여기서 $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ 이다. 오차제곱합 $Q(\alpha, \beta) = \sum (y_i - \alpha - \beta x_i)^2$을 최소화하는 $\hat{\alpha}, \hat{\beta}$는 다음 두 식을 만족한다.

$$\left.\frac{\partial Q}{\partial \alpha}\right|_{\hat{\alpha},\hat{\beta}} = -2\sum (y_i - \hat{\alpha} - \hat{\beta} x_i) = -2\sum \hat{e}_i = 0$$
$$\left.\frac{\partial Q}{\partial \beta}\right|_{\hat{\alpha},\hat{\beta}} = -2\sum x_i(y_i - \hat{\alpha} - \hat{\beta} x_i) = -2\sum x_i \hat{e}_i = 0$$

위 두 식으로부터 $\sum \hat{e}_i = 0$ 과 $\sum x_i \hat{e}_i = 0$ 이 바로 증명된다. 이 두 제약조건 때문에 n개의 잔차 중 자유롭게 변할 수 있는 것은 $n-2$개 뿐이다. ■

연습문제 8.2 총편차의 분해식에 대한 증명

(a) 연습문제 8.1의 결과를 활용하자.

$$\begin{aligned} 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}_i) &= 2\sum_{i=1}^{n}\hat{e}_i \left\{(\hat{\beta} x_i - \epsilon_i) - (\hat{\beta}\bar{x} - \epsilon_i)\right\} = 2\sum_{i=1}^{n}\hat{e}_i \hat{\beta}(x_i - \bar{x}) \\ &= 2\hat{\beta}\left\{\sum_{i=1}^{n} x_i \hat{e}_i - \bar{x}\sum_{i=1}^{n}\hat{e}_i\right\} = 0 \end{aligned}$$

(b) $y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$ 양변에 제곱과 $\sum$을 취하고 (a)를 활용한다면,

$$\begin{aligned} \sum_{i=1}^{n}\{y_i - \bar{y}\}^2 &= \sum_{i=1}^{n}\{(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})\}^2 \\ &= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \end{aligned}$$ ■

연습문제 8.3 SSR에 대한 성질

회귀제곱합 SSR에 대해 $c_i = (x_i - \bar{x})/\sqrt{S_{xx}}$ 라 할 때, 다음이 성립함을 보여야 한다.

$$SSR \overset{(a)}{=} \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \overset{(b)}{=} \hat{\beta}^2 S_{xx} \overset{(c)}{=} \left(\sum_{i=1}^{n} c_i y_i\right)^2$$

(a) SSR의 정의로 당연하다.

(b) $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i = (\bar{y} - \hat{\beta}\bar{x}) + \hat{\beta} x_i = \bar{y} + \hat{\beta}(x_i - \bar{x})$ 이므로, $\hat{y}_i - \bar{y} = \hat{\beta}(x_i - \bar{x})$이다. 따라서,

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n}[\hat{\beta}(x_i - \bar{x})]^2 = \hat{\beta}^2\sum_{i=1}^{n}(x_i - \bar{x})^2 = \hat{\beta}^2 S_{xx}$$

(c) 먼저 $\sum c_i y_i$를 정리하면,

$$\sum_{i=1}^{n} c_i y_i = \sum_{i=1}^{n} \frac{x_i - \bar{x}}{\sqrt{S_{xx}}} y_i = \frac{1}{\sqrt{S_{xx}}} \sum_{i=1}^{n} (x_i - \bar{x}) y_i$$

$\sum(x_i - \bar{x})(y_i - \bar{y}) = \sum(x_i - \bar{x})y_i - \bar{y}\sum(x_i - \bar{x})$ 이고 $\sum(x_i - \bar{x}) = 0$ 이므로,

$$S_{xy} = \sum (x_i - \bar{x}) y_i$$

따라서,

$$\sum_{i=1}^{n} c_i y_i = \frac{S_{xy}}{\sqrt{S_{xx}}} \Rightarrow \left(\sum_{i=1}^{n} c_i y_i\right)^2 = \frac{S_{xy}^2}{S_{xx}}$$

한편, $\hat{\beta} = S_{xy}/S_{xx}$ 이므로,

$$\hat{\beta}^2 S_{xx} = \left(\frac{S_{xy}}{S_{xx}}\right)^2 S_{xx} = \frac{S_{xy}^2}{S_{xx}}$$

■

연습문제 8.4 단순선형회귀모형 에서의 성질

(a) 연습문제 8.3과 이 문제의 c_i 정의가 다름에 유의하자.

$$\sum c_i y_i = \sum \frac{x_i - \bar{x}}{S_{xx}} y_i = \frac{1}{S_{xx}} \sum (x_i - \bar{x}) y_i$$

연습문제 8.3에서 보았듯이 $\sum(x_i - \bar{x})y_i = S_{xy}$ 이므로,

$$\sum c_i y_i = \frac{S_{xy}}{S_{xx}} = \hat{\beta}$$

(b) $d_i = \dfrac{1}{n} + \dfrac{(x - \bar{x})(x_i - \bar{x})}{S_{xx}}$로 정의하자. 그러면

$$\sum d_i y_i = \sum \left(\frac{1}{n} + \frac{(x - \bar{x})(x_i - \bar{x})}{S_{xx}}\right) y_i = \frac{1}{n}\sum y_i + \frac{x - \bar{x}}{S_{xx}} \sum (x_i - \bar{x}) y_i$$

여기서 $\dfrac{1}{n}\sum y_i = \bar{y}$ 이고, $\sum(x_i - \bar{x})y_i = S_{xy}$ 이므로,

$$\sum d_i y_i = \bar{y} + \frac{x - \bar{x}}{S_{xx}} \cdot S_{xy} = \bar{y} + (x - \bar{x})\hat{\beta}$$

한편, $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ 이므로,

$$\sum d_i y_i = \bar{y} + x\hat{\beta} - \bar{x}\hat{\beta} = (\bar{y} - \hat{\beta}\bar{x}) + \hat{\beta}x = \hat{\alpha} + \hat{\beta}x$$

(c) (a)의 결과를 사용하면 $\hat{\beta}$의 기댓값은

$$E(\hat{\beta}) = E\left(\sum c_i y_i\right) = \sum c_i E(y_i)$$

모형에서 $E(y_i) = \alpha + \beta x_i$ 이므로,

$$E(\hat{\beta}) = \sum c_i(\alpha + \beta x_i) = \alpha \sum c_i + \beta \sum c_i x_i$$

여기서 $\sum c_i$ 와 $\sum c_i x_i$를 계산하면,

$$\sum c_i = \sum \frac{x_i - \bar{x}}{S_{xx}} = \frac{1}{S_{xx}} \sum (x_i - \bar{x}) = 0$$

$$\sum c_i x_i = \sum \frac{x_i - \bar{x}}{S_{xx}} x_i = \frac{1}{S_{xx}} \sum (x_i - \bar{x}) x_i = \frac{1}{S_{xx}} \sum (x_i - \bar{x})(x_i - \bar{x}) = \frac{S_{xx}}{S_{xx}} = 1$$

따라서, $E(\hat{\beta}) = \alpha(0) + \beta(1) = \beta$. 즉 $\hat{\beta}$는 β의 불편추정량이다.

(d) Y_i들이 서로 독립이고 $Var(y_i) = \sigma^2$ 이므로,

$$Var(\hat{\beta}) = Var\left(\sum c_i y_i\right) = \sum c_i^2 Var(y_i) = \sigma^2 \sum c_i^2$$

$\sum c_i^2$를 계산하면,

$$\sum c_i^2 = \sum \left(\frac{x_i - \bar{x}}{S_{xx}}\right)^2 = \frac{1}{S_{xx}^2} \sum (x_i - \bar{x})^2 = \frac{S_{xx}}{S_{xx}^2} = \frac{1}{S_{xx}}$$

따라서, $Var(\hat{\beta}) = \sigma^2 \left(\frac{1}{S_{xx}}\right) = \frac{\sigma^2}{S_{xx}}$ ■

(e) (b)의 결과를 사용하면 $\hat{\alpha} + \hat{\beta}x$의 기댓값은

$$E(\hat{\alpha} + \hat{\beta}x) = E\left(\sum d_i y_i\right) = \sum d_i E(y_i)$$

모형에서 $E(y_i) = \alpha + \beta x_i$ 이므로,

$$E(\hat{\alpha} + \hat{\beta}x) = \sum d_i(\alpha + \beta x_i) = \alpha \sum d_i + \beta \sum d_i x_i$$

여기서 $\sum d_i$ 와 $\sum d_i x_i$를 계산하면,

$$\sum d_i = \sum \left(\frac{1}{n} + \frac{(x - \bar{x})(x_i - \bar{x})}{S_{xx}}\right) = 1 + \frac{x - \bar{x}}{S_{xx}} \sum (x_i - \bar{x}) = 1 + 0 = 1$$

$$\begin{aligned}
\sum d_i x_i &= \sum \left(\frac{1}{n} + \frac{(x - \bar{x})(x_i - \bar{x})}{S_{xx}}\right) x_i = \frac{1}{n} \sum x_i + \frac{x - \bar{x}}{S_{xx}} \sum (x_i - \bar{x}) x_i \\
&= \bar{x} + \frac{x - \bar{x}}{S_{xx}} \cdot S_{xx} = \bar{x} + (x - \bar{x}) = x
\end{aligned}$$

따라서, $E(\hat{\alpha} + \hat{\beta}x) = \alpha(1) + \beta(x) = \alpha + \beta x$. 즉 $\hat{\alpha} + \hat{\beta}x$는 $\alpha + \beta x$의 불편추정량이다.

(f) y_i들이 서로 독립이고 $Var(y_i) = \sigma^2$ 이므로,

$$Var(\hat{\alpha} + \hat{\beta}x) = Var\left(\sum d_i y_i\right) = \sum d_i^2 Var(y_i) = \sigma^2 \sum d_i^2$$

$\sum d_i^2$를 계산하면,

$$\begin{aligned}
\sum d_i^2 &= \sum \left(\frac{1}{n} + \frac{(x - \bar{x})(x_i - \bar{x})}{S_{xx}}\right)^2 \\
&= \sum \left(\frac{1}{n^2} + \frac{2(x - \bar{x})(x_i - \bar{x})}{nS_{xx}} + \frac{(x - \bar{x})^2 (x_i - \bar{x})^2}{S_{xx}^2}\right) \\
&= \frac{1}{n^2} \cdot n + \frac{2(x - \bar{x})}{nS_{xx}} \sum (x_i - \bar{x}) + \frac{(x - \bar{x})^2}{S_{xx}^2} \sum (x_i - \bar{x})^2 \\
&= \frac{1}{n} + \frac{2(x - \bar{x})}{nS_{xx}} \cdot 0 + \frac{(x - \bar{x})^2}{S_{xx}^2} \cdot S_{xx} \\
&= \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}
\end{aligned}$$

따라서, $Var(\hat{\alpha} + \hat{\beta}x) = \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right)$ ■

연습문제 8.5 국민총소득과 자동차 등록대수의 상관관계

주어진 자료에 대한 기초 통계량은 다음과 같다 ($n = 10$).

- $\sum x_i = 34820 \implies \bar{x} = 3482, \ \sum y_i = 23305 \implies \bar{y} = 2330.5$
- $\sum x_i^2 = 121571546, \sum y_i^2 = 54652753, \sum x_i y_i = 81466793$
- $S_{xx} = \sum x_i^2 - (\sum x_i)^2/n = 328306$
- $S_{yy} = \sum y_i^2 - (\sum y_i)^2/n = 340450.5$
- $S_{xy} = \sum x_i y_i - (\sum x_i)(\sum y_i)/n = 318783$

(a) 산점도는 x축을 1인당 GNI, y축을 자동차 등록대수로 하여 점을 찍어 그리면 Figure 6와 같다.

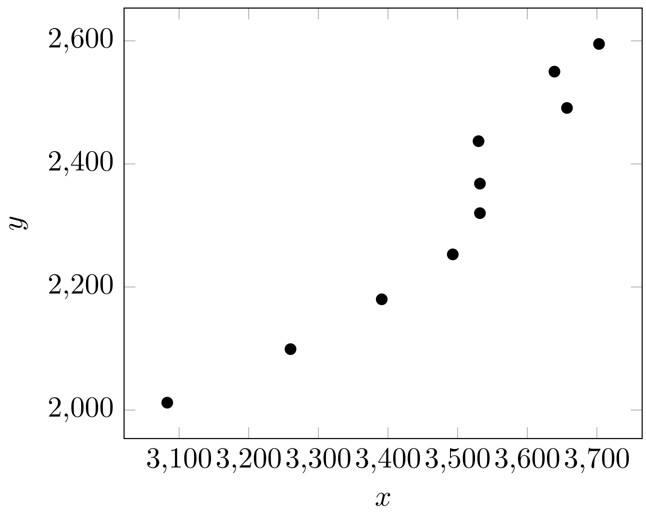

Figure 6: 국민총소득과 자동차 등록대수의 산점도

(b)

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{318783}{\sqrt{328306 \times 340450.5}} \approx 0.9535$$

(c) 모상관계수 ρ에 대한 가설검정

① 가설: $H_0 : \rho = 0$ vs $H_1 : \rho > 0$

② 검정통계량: $T = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t(n-2)$ under H_0

③ 검정 (유의수준 $\alpha = 0.05$로 가정)

- 관측값: $n = 10, r \approx 0.9535$
- 검정통계량의 관측값: $t = \frac{0.9535\sqrt{10-2}}{\sqrt{1-(0.9535)^2}} \approx 8.9481$
- 기각역: $t \geq t_{0.05}(8) \approx 1.860$

④ 결론: 검정통계량의 값 8.9481이 기각역에 속하므로 ($8.9481 \geq 1.860$), 유의수준 1%에서 귀무가설을 기각한다. 따라서 우리나라의 1인당 GNI와 자동차 등록대수 사이에는 유의한 양의 상관관계가 있다고 할 수 있다.

연습문제 8.7 국민총소득과 자동차 등록대수의 상관관계 (이어서)

(a) 분산분석표

- SST $= S_{yy} = 340450.5$
- SSR $= \frac{S_{xy}^2}{S_{xx}} = \frac{318783^2}{328306} \approx 309536.2287$
- SSE = SST - SSR $\approx 340450.5 - 309536.2287 = 30914.2713$
- DF(reg)=1, DF(err)=n-2=8, DF(total)=n-1=9
- MSR = SSR/1 ≈ 309536.2287
- MSE = SSE/8 ≈ 3864.2839
- F = MSR/MSE ≈ 80.1018

요인	제곱합	자유도	평균제곱	F 값
회귀	309536.2287	1	309536.2287	80.1018
잔차	30914.2713	8	3864.2839	
계	340450.5	9		

(b)

$$R^2 = \frac{SSR}{SST} \approx \frac{309536.2287}{340450.5} \approx 0.9092$$

참고: $r^2 \approx (0.9535)^2 \approx 0.9092 \approx R^2$

(c) $\hat{\sigma} = \sqrt{MSE} \approx \sqrt{3864.2839} \approx 62.1634$

(d) 회귀직선의 유의성검정 (유의수준 $\alpha = 0.01$)

① 가설: $H_0 : \beta = 0$ vs $H_1 : \beta \neq 0$

② 검정통계량: $F = \frac{MSR}{MSE} \sim F(1, n-2)$ under H_0

③ 검정:

- 검정통계량의 관측값: $F \approx 80.1018$ (분산분석표에서 계산됨)
- 기각역: $F \geq F_{0.01}(1, 8) \approx 11.26$

④ 결론: 검정통계량의 값 80.1018이 기각역에 속하므로 ($80.1018 \geq 11.26$), 유의수준 1%에서 귀무가설을 기각한다. 따라서 회귀직선은 통계적으로 유의하며, 1인당 GNI는 자동차 등록대수를 예측하는 데 유의한 변수라고 할 수 있다.

9장

연습문제 9.1 총편차의 분해

분해식 $(y_{ij} - \bar{y}_{..}) = (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})$의 양변을 제곱하면,

$$(y_{ij} - \bar{y}_{..})^2 = (\bar{y}_{i.} - \bar{y}_{..})^2 + 2(\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}) + (y_{ij} - \bar{y}_{i.})^2$$

양변을 $i = 1, \ldots, k$와 $j = 1, \ldots, n$에 대해 합하면,

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(\bar{y}_{i.} - \bar{y}_{..})^2 + 2\sum_{i=1}^{k}\sum_{j=1}^{n}(\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}) + \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$$

힌트에서 주어진 $\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.}) = 0$을 이용하자.
(1) 첫째 항: $(\bar{y}_{i.} - \bar{y}_{..})$는 j에 무관하므로,

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2 \cdot n = n\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2$$

(2) 둘째 항 (교차항): $(\bar{y}_{i.} - \bar{y}_{..})$는 j에 무관하므로,

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}) = \sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.}) = \sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..}) \cdot 0 = 0$$

따라서,

$$\boxed{\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = n\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2}$$

즉, **SST = SSA + SSE** (총제곱합 = 처리제곱합 + 잔차제곱합)이 성립한다. ■

연습문제 9.3 빵집 선호도 비교 (일원배치법)

일원배치법 모형: $Y_{ij} = \mu + \alpha_i + e_{ij}$, $i = 1, 2, 3$, $j = 1, \ldots, 10$, 여기서 $e_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

기초 통계량 계산 ($k = 3$, $n = 10$, $N = 30$)

- A_1: $y_{1.} = 7 + 8 + 5 + 9 + 6 + 5 + 10 + 8 + 7 + 9 = 74$, $\bar{y}_{1.} = 7.4$
- A_2: $y_{2.} = 6 + 6 + 3 + 8 + 6 + 4 + 9 + 6 + 5 + 7 = 60$, $\bar{y}_{2.} = 6.0$
- A_3: $y_{3.} = 9 + 10 + 7 + 8 + 8 + 6 + 10 + 9 + 8 + 10 = 85$, $\bar{y}_{3.} = 8.5$
- 총합: $y_{..} = 74 + 60 + 85 = 219$, 총평균: $\bar{y}_{..} = 219/30 = 7.3$

제곱합 계산

$\sum_{i,j} y_{ij}^2 = (49 + \cdots + 81) + (36 + \cdots + 49) + (81 + \cdots + 100) = 574 + 388 + 739 = 1701$

$$SST = \sum_{i,j} y_{ij}^2 - \frac{y_{..}^2}{N} = 1701 - \frac{219^2}{30} = 1701 - 1598.7 = 102.3$$

$$\begin{aligned} SSA &= \frac{1}{n}\sum_{i=1}^{k} y_{i.}^2 - \frac{y_{..}^2}{N} = \frac{1}{10}(74^2 + 60^2 + 85^2) - \frac{219^2}{30} \\ &= \frac{1}{10}(5476 + 3600 + 7225) - 1598.7 = 1630.1 - 1598.7 = 31.4 \end{aligned}$$

$$SSE = SST - SSA = 102.3 - 31.4 = 70.9$$

분산분석표

요인	제곱합	자유도	평균제곱	F 값
처리(지점)	31.4	2	15.7	5.98
잔차	70.9	27	2.626	
계	102.3	29		

검정

① 가설: $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ (모든 지점의 선호도 평균이 동일)
H_1 : 적어도 하나의 $\alpha_i \neq 0$

② 검정통계량: $F = \dfrac{MSA}{MSE} \sim F(k-1, N-k)$ under H_0

③ 검정 (유의수준 $\alpha = 0.01$)

- 검정통계량의 관측값: $F = \frac{15.7}{2.626} \approx 5.98$
- 기각역: $F \geq F_{0.01}(2, 27) \approx 5.49$

④ 결론: 검정통계량의 값 5.98이 기각역에 속하므로 ($5.98 \geq 5.49$), 유의수준 1%에서 귀무가설을 기각한다. 따라서 지점에 따른 빵 맛에 대한 선호도에 유의한 차이가 있다고 할 수 있다.

연습문제 9.5 이원배치법 분산분석표 작성 및 검정 (반복 없음)

반복이 없는 이원배치법에서 인자 A는 $a = 4$수준, 인자 B는 $b = 5$수준이고, 총 관측수는 $N = a \times b = 20$이다.

(a) 분산분석표 완성

자유도:

- $df_A = a - 1 = 4 - 1 = 3$
- $df_B = b - 1 = 5 - 1 = 4$
- $df_E = (a-1)(b-1) = 3 \times 4 = 12$
- $df_T = N - 1 = 20 - 1 = 19$

제곱합:

- $SST = 2200.77$ (주어짐)
- $SSA = 350.6$ (주어짐)
- $SSE = 30$ (주어짐)
- $SSB = SST - SSA - SSE = 2200.77 - 350.6 - 30 = 1820.17$

평균제곱:

- $MSA = SSA/df_A = 350.6/3 = 116.867$
- $MSB = SSB/df_B = 1820.17/4 = 455.04$
- $MSE = SSE/df_E = 30/12 = 2.5$

F값:

- $f_A = MSA/MSE = 116.87/2.5 = 46.75$
- $f_B = MSB/MSE = 455.04/2.5 = 182.02$

요인	제곱합	자유도	평균제곱	F 값
인자 A	350.6	3	116.87	46.75
인자 B	1820.17	4	455.04	182.02
잔차 E	30	12	2.5	
합계	2200.77	19		

(b) 가설 설정
반복이 없는 이원배치법 모형: $Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$
인자 A에 대한 검정:

- $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ (인자 A의 모든 수준의 효과가 동일)
- H_1 : 적어도 하나의 $\alpha_i \neq 0$

인자 B에 대한 검정:

- $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ (인자 B의 모든 수준의 효과가 동일)
- H_1 : 적어도 하나의 $\beta_j \neq 0$

(c) 인자 B에 대한 F 검정통계량의 분포 및 검정

① 가설: $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ vs H_1 : 적어도 하나의 $\beta_j \neq 0$

② 검정통계량: $F_B = \dfrac{MSB}{MSE} \sim F(b-1, (a-1)(b-1)) = F(4, 12)$ under H_0

③ 검정 (유의수준 $\alpha = 0.05$)

- 검정통계량의 관측값: $F_B = \frac{455.04}{2.5} = 182.02$
- 기각역: $F \geq F_{0.05}(4, 12) \approx 3.26$

④ 결론: 검정통계량의 값 182.02가 기각역에 속하므로 ($182.02 \gg 3.26$), 유의수준 5%에서 귀무가설을 기각한다. 따라서 인자 B는 반응변수에 유의한 영향을 미친다고 할 수 있다.

연습문제 9.6 패션 브랜드 매출량 분석 (반복 없는 이원배치법)

반복이 없는 이원배치법 모형: $Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$, $i = 1, \ldots, 5$, $j = 1, \ldots, 4$
기초 통계량 계산 ($a = 5$, $b = 4$, $N = 20$)

- 행합: $y_{1.} = 207$, $y_{2.} = 230$, $y_{3.} = 231$, $y_{4.} = 218$, $y_{5.} = 223$
- 열합: $y_{.1} = 583$, $y_{.2} = 297$, $y_{.3} = 73$, $y_{.4} = 156$
- 총합: $y_{..} = 1109$

제곱합 계산

$$CT = \frac{y_{..}^2}{N} = \frac{1109^2}{20} = 61494.05$$

$$\sum_{i,j} y_{ij}^2 = 100^2 + 56^2 + \cdots + 15^2 + 26^2 = 92155$$

$$SST = \sum_{i,j} y_{ij}^2 - CT = 92155 - 61494.05 = 30660.95$$

$$SSA = \frac{1}{b}\sum_{i=1}^{a} y_{i.}^2 - CT = \frac{207^2 + 230^2 + 231^2 + 218^2 + 223^2}{4} - 61494.05 = 96.70$$

$$SSB = \frac{1}{a}\sum_{j=1}^{b} y_{.j}^2 - CT = \frac{583^2 + 297^2 + 73^2 + 156^2}{5} - 61494.05 = 30058.55$$

$$SSE = SST - SSA - SSB = 30660.95 - 96.70 - 30058.55 = 505.70$$

분산분석표

요인	제곱합	자유도	평균제곱	F 값
인자 A(디자인)	96.70	4	24.175	0.574
인자 B(도시)	30058.55	3	10019.52	237.76
잔차	505.70	12	42.14	
계	30660.95	19		

인자 A(디자인)에 대한 검정

① 가설: $H_0 : \alpha_1 = \cdots = \alpha_5 = 0$ vs H_1 : 적어도 하나의 $\alpha_i \neq 0$

② 검정통계량: $F_A = \dfrac{MSA}{MSE} \sim F(4, 12)$ under H_0

③ 검정 (유의수준 $\alpha = 0.05$)

- 검정통계량의 관측값: $F_A = 24.175/42.14 \approx 0.574$
- 기각역: $F \geq F_{0.05}(4, 12) \approx 3.26$

④ 결론: $0.574 < 3.26$이므로 귀무가설을 기각하지 못한다. 디자인(인자 A)은 매출량에 유의한 영향을 미치지 않는다.

인자 B(도시)에 대한 검정

① 가설: $H_0 : \beta_1 = \cdots = \beta_4 = 0$ vs H_1 : 적어도 하나의 $\beta_j \neq 0$

② 검정통계량: $F_B = \dfrac{MSB}{MSE} \sim F(3, 12)$ under H_0

③ 검정 (유의수준 $\alpha = 0.05$)

- 검정통계량의 관측값: $F_B = 10019.52/42.14 \approx 237.76$
- 기각역: $F \geq F_{0.05}(3, 12) \approx 3.49$

④ 결론: $237.76 \gg 3.49$이므로 귀무가설을 기각한다. 도시(인자 B)는 매출량에 유의한 영향을 미친다.

연습문제 9.8 합성수지 강도 분석 (반복 있는 이원배치법)

반복이 있는 이원배치법 모형: $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$

여기서 $a = 4$ (인자 A: 성형온도), $b = 3$ (인자 B: 촉매량), $r = 2$ (반복), $N = 24$

기초 통계량 계산

셀합 $y_{ij.}$	A_1	A_2	A_3	A_4	$y_{.j.}$
B_1	25.8	25.6	26.1	26.4	103.9
B_2	25.8	25.6	25.9	26.4	103.7
B_3	24.0	24.3	25.9	28.2	102.4
$y_{i..}$	75.6	75.5	77.9	81.0	$y_{...} = 310.0$

제곱합 계산

$$CT = \frac{y_{...}^2}{N} = \frac{310^2}{24} = 4004.17$$

$$\sum_{i,j,k} y_{ijk}^2 = 12.5^2 + 13.3^2 + \cdots + 14.2^2 + 14.0^2 = 4011.56$$

$$SST = 4011.56 - 4004.17 = 7.39$$

$$SSA = \frac{1}{br}\sum_i y_{i..}^2 - CT = \frac{75.6^2 + 75.5^2 + 77.9^2 + 81.0^2}{6} - 4004.17 = 3.34$$

$$SSB = \frac{1}{ar}\sum_j y_{.j.}^2 - CT = \frac{103.9^2 + 103.7^2 + 102.4^2}{8} - 4004.17 = 0.17$$

$$SS_{cell} = \frac{1}{r}\sum_{i,j} y_{ij.}^2 - CT = \frac{25.8^2 + 25.6^2 + \cdots + 28.2^2}{2} - 4004.17 = 6.07$$

$$SS_{AB} = SS_{cell} - SSA - SSB = 6.07 - 3.34 - 0.17 = 2.57$$

$$SSE = SST - SS_{cell} = 7.39 - 6.07 = 1.32$$

분산분석표

요인	제곱합	자유도	평균제곱	F 값
인자 A(성형온도)	3.34	3	1.11	10.11
인자 B(촉매량)	0.17	2	0.08	0.75
교호작용 $A \times B$	2.57	6	0.43	3.90
잔차	1.32	12	0.11	
계	7.39	23		

반복이 있는 이원배치법에서는 **교호작용을 먼저 검정**한다.

교호작용 $A \times B$에 대한 검정

① 가설: $H_0 : (\alpha\beta)_{ij} = 0$ for all i, j vs H_1 : 적어도 하나의 $(\alpha\beta)_{ij} \neq 0$

② 검정통계량: $F_{AB} = \dfrac{MS_{AB}}{MSE} \sim F(6, 12)$ under H_0

③ 검정 (유의수준 $\alpha = 0.05$)

- 검정통계량의 관측값: $F_{AB} = 0.43/0.11 \approx 3.90$
- 기각역: $F \geq F_{0.05}(6, 12) \approx 3.00$

④ 결론: $3.90 > 3.00$이므로 귀무가설을 기각한다. 성형온도와 촉매량 사이에 **유의한 교호작용이 존재**한다.

교호작용이 유의하면 주효과에 대한 검정은 진행하지 않아도 된다.