통계학의 이해 II

이원배치 분산분석: 반복이 없는 경우
-고정효과모형

이원배치 분산분석: 반복이 없는 경우

- 요인이 두 개이고 각 처리에 하나의 관측값이 있는 경우, 각 요인의 처리효과를 확인하기 위해 어떻게 모형을 설정하는지 알아본다.
- 고정효과 모형 하에서의 통계적 추론을 알아본다.

이원배치 분산분석

실험설계

- 요인 A의 수준 수는 p, 요인 B의 수준 수는 q
- $p \times q$ 처리를 완전 확률화 하여 실험을 진행

자료구조

요인 A	요인 B			
	1	2	$\cdots$	q
1	Y_{11}	Y_{12}	$\cdots$	Y_{1q}
2	Y_{21}	Y_{22}	$\cdots$	Y_{2q}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
p	Y_{p1}	Y_{p2}	$\cdots$	Y_{pq}

- 요인의 수준 선택
  - 두 요인 모두 실험자가 결정 ⇒ 고정효과모형
  - 두 요인 모두 무작위 선택 ⇒ 변량효과모형
  - 하나는 실험자가 결정, 다른 하나는 무작위 선택

    ⇒ 혼합효과모형(mixed effect models)

- 동일 개체를 반복측정하는 경우 ⇒ 상관관계가 존재할 수 있음
  - 변량효과로 처리: ICC

고정효과모형

고정효과모형식

- 1-요인 설계: $Y_{ij} = \mu + (\mu_i - \mu) + \varepsilon_{ij} = \mu + \alpha_i + \varepsilon_{ij}$
- 2-요인 설계: $Y_{ij} = \mu + (\mu_{i+} - \mu) + \left(\mu_{+j} - \mu\right) + \varepsilon_{ij}$

  $= \mu + \alpha_i + \beta_j + \varepsilon_{ij}$

  - μ : 전체평균, $\varepsilon_{ij} \sim iid\ N\left(0, \sigma^2\right)$
  - μ_{i+}: 요인 A의 i번째 수준 평균, μ_{+j}: 요인 B의 j번째 수준 평균
  - $\alpha_i = \mu_{i+} - \mu$: 요인 A의 i번째 처리 효과, $\sum \alpha_i = 0$
  - $\beta_j = \mu_{+j} - \mu$: 요인 B의 j번째 처리 효과, $\sum \beta_j = 0$

변동분해

- 모형식: $Y_{ij} = \mu + (\mu_{i+} - \mu) + (\mu_{+j} - \mu) + \varepsilon_{ij}$
- $Y_{ij} - \bar{Y} = \bar{Y}_{i+} - \bar{Y} + \bar{Y}_{+j} - \bar{Y} + Y_{ij} - \bar{Y}_{i+} - \bar{Y}_{+j} + \bar{Y}$
- TSS: $\sum_{i=1}^{p}\sum_{j=1}^{q}(Y_{ij} - \bar{Y})^2$, 자유도 $= N - 1$
- SSA: $\sum_{i=1}^{p} q(\bar{Y}_{i+} - \bar{Y})^2$, 자유도 $= p - 1$
- SSB: $\sum_{j=1}^{q} p(\bar{Y}_{+j} - \bar{Y})^2$, 자유도 $= q - 1$
- SSE: $\sum_{i=1}^{p}\sum_{j=1}^{q}(Y_{ij} - \bar{Y}_{i+} - \bar{Y}_{+j} + \bar{Y})^2$, 자유도 $= (p-1)(q-1)$

가설검정

- 요인 A의 처리효과 유무: $H_{A0}: \alpha_1 = \ \cdots\ = \alpha_p = 0$
- 요인 B의 처리효과 유무: $H_{B0}: \beta_1 = \ \cdots\ = \beta_q = 0$

분산분석표

변인	자유도	SS	MS	F
처리 A	$p-1$	SSA	MSA	MSA/MSE
처리 B	$q-1$	SSB	MSB	MSB/MSE
오차	$(p-1)(q-1)$	SSE	MSE	
전체	$N-1$	TSS		

- 유의하지 않는 요인의 처리효과는 오차에 흡수시켜 다시 분석

◇ 처리 평균 추정

- $\mu(A_i)$의 구간추정: $\bar{Y}_{i+} \pm t_{\frac{\alpha}{2},(p-1)(q-1)}\sqrt{MSE/q}$
  - ○ 요인B의 처리효과가 없는 경우: $\bar{Y}_{i+} \pm t_{\frac{\alpha}{2},(p-1)q}\sqrt{MSE_A/q}$
- $\mu(B_j)$의 구간추정: $\bar{Y}_{+j} \pm t_{\frac{\alpha}{2},(p-1)(q-1)}\sqrt{MSE/p}$
  - ○ 요인A의 처리효과가 없는 경우: $\bar{Y}_{i+} \pm t_{\frac{\alpha}{2},p(q-1)}\sqrt{MSE_B/p}$

◇ 각 요인의 수준에 대한 다중비교 가능

- 두 요인의 처리효과가 있는 경우
  - ○ $|\bar{Y}_{i+} - \bar{Y}_{j+}| > A\sqrt{MSE}\sqrt{2/q} \Rightarrow$ 처리 A_i와 A_j 간 유의한 차이
  - ○ $|\bar{Y}_{+i} - \bar{Y}_{+j}| > A\sqrt{MSE}\sqrt{2/p} \Rightarrow$ 처리 B_i와 B_j 간 유의한 차이

◇ 예제】 원료와 반응온도에 따른 제품의 생산량 비교

- 원료(A): 미국 M사, 일본 Q사, 한국 P사
- 반응온도(B): 180, 190, 200, 210
  - 온도에 따라 차이가 있는지를 확인하는 것이 목적

요인 A	요인 B			
	180	190	200	210
M	97.6	98.6	99.0	98.0
Q	97.3	98.2	98.0	97.7
P	96.7	96.9	97.9	96.5

분산분석표

변인	자유도	SS	MS	F	p-값
처리 A	2	3.44	1.72	18.43	0.0027
처리 B	3	2.22	0.74	7.93	0.0165
오차	6	0.56	0.093		
전체	11	6.22			

- 두 요인 모두 처리효과가 있음

  $\Rightarrow$ 원료사와 반응온도에 따라 생산량에 차이가 있음

구간추정

- $\mu(A_1)$의 95% 신뢰구간 = $97.2 \pm 2.447\sqrt{0.093/3} = (96.77, 97.63)$
- $\mu(B_1)$의 95% 신뢰구간 = $98.3 \pm 2.447\sqrt{0.093/4} = (97.93, 98.67)$

이원배치 분산분석: 반복이 없는 경우

- 요인이 두 개이고 각 처리에 하나의 관측값이 있는 경우, 각 요인의 처리효과를 확인하기 위해 어떻게 모형을 설정하는지 알아본다.
  - 2-요인 설계: $Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$
    - α_i: 요인 A의 i번째 처리 효과, $\sum \alpha_i = 0$
    - β_j: 요인 B의 j번째 처리 효과, $\sum \beta_j = 0$
- 고정효과 모형 하에서의 통계적 추론을 알아본다.
  - 요인 A의 처리효과 유무: $H_{A0}: \alpha_1 = \cdots = \alpha_p = 0$
  - 요인 B의 처리효과 유무: $H_{B0}: \beta_1 = \cdots = \beta_q = 0$