실험계획과 분석

심송용(한림대학교 데이터과학스쿨)

http://jupiter.hallym.ac.kr

내포설계(Nested Design)

내포설계(Nested Design)

➡ 교차요인(crossed factor): 한 요인의 수준이 다른 요인의 수준에 영향을 미치지 않는 요인.

➡ 내포요인(nested factor): 한 요인의 수준에 따라 다른 요인의 수준이 영향을 받는 경우. 요인 A의 수준에 따라 요인 B의 수준 속성이 달라지는 경우 요인 B는 요인 A에 내포되었다고 하고, A에 내포된 것을 표시하기 위해 요인 B는 B(A)로 표시하기도 한다.

B in A

교차요인

	B1	B2
A1	A1B1	A1B2
A2	A2B1	A2B2

내포요인

A	A1		A2	
B	B1(A1)	B2(A1)	B1(A2)	B2(A2)

내포설계(Nested Design)

예: 돼지를 사육하는데 두 종류의 사료(요인 A)에 따른 무게 증가가 차이가 있는지 알아보기 위해 사료 1(수준 A1)을 사용하여 두 곳의 민간 농장에서 위탁 사육하고 사료 2(수준 A2)를 사용하여 또 다른 두 곳의 농장에서 사육을 한다고 하자. 사료1을 사용한 두 곳은 농장과 사료 2를 사용한 두 곳의 농장을 편의상 B1, B2라고 할 수 있으나 사료 1을 사용한 B1과 사료 2를 사용한 B1은 다른 농장임. 사료 2도 마찬가지. A의 수준에 따라 B의 수준이 변하므로 내포설계

4개 다 다름

예: 두 종류의 비료와 두 종류의 농약에 따른 농작물의 수확량이 달라지는지 알아보기 위해 실험. 비료 1일 때의 농약 1, 2가 비료 2일 때의 농약 1, 2가 같은 종류이며 차이가 없음. 따라서 교차설계

예: 학교(요인 A)와 교사(요인 B)에 따른 학업성취도를 알아보기 위해 학교 2곳을 선택하고 각 학교에서 2명의 교사를 선택한 실험. 학교 1의 교사1과 학교 2의 교사 2는 다른 사람임. 학교 2도 마찬가지. 교사(요인 B)는 요인 A(학교)에 내포 되어 있음.

∴ 내포설계

B in A ← 다른 사람

B가 A에 내포

A_1		A_2	
B_1	B_2	B_1	B_2

내포설계(Nested Design)

요인 B가 요인 A에 내포된 경우 요인 A의 i번째 수준, 요인 B의 j번째 수준에서의 k번째 관측치를 y_{ijk}라고 하면

$$y_{ijk} = \mu + \overset{A}{\alpha_i} + \overset{B}{\underset{j \text{ in } i}{\beta_{j(i)}}} + \epsilon_{ijk},$$

$$i = 1,2,\ldots,a;\ j = 1,2,\ldots,b;\ k = 1,2,\ldots,n$$

$\sum_{i=1}^{a} \alpha_i = 0$ 또는 $\alpha_i \sim N(0, \sigma_\alpha^2)$ 이고 독립

$\sum_{j=1}^{b} \beta_{j(i)} = 0$ 또는 $\beta_{j(i)} \sim N(0, \sigma_\beta^2)$ 이고 독립

$\epsilon_{ijk} \sim N(0, \sigma^2)$ 이고 모두 독립

ϵ_{ijk}, α_i, $\beta_{j(i)}$ 도 모두 독립

- α_i: 요인 A에 의한 차이
- $\beta_{j(i)}$: 요인 B에 의한 차이. 이 경우 B라고 하더라도 정확히는 B(A)임.

제곱합의 분해와 분산분석-Nested Design

제곱합의 분해

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

에서 양변을 제곱하여 합하면

$$\sum_{i,j,k}(y_{ijk} - \bar{y}_{...})^2 = \sum_{i,j,k}(\bar{y}_{i..} - \bar{y}_{...})^2 + \sum_{i,j,k}(\bar{y}_{ij.} - \bar{y}_{i..})^2 + \sum_{i,j,k}(y_{ijk} - \bar{y}_{ij.})^2 + (\text{교차항들})$$

이며 교차항들은 일원배치법, 이원배치법들에서 본 것과 같이 0이므로

0

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

을 얻음

전체합^2

제곱합의 분해와 분산분석-Nested Design

- $CT = \frac{y_{...}^2}{abn} = abn \cdot \bar{y}_{...}^2$ 라 하면
- $\sum_{i,j,k} (y_{ijk} - \bar{y}_{...})^2 = \sum_{i,j,k} y_{ijk}^2 - abn \times \bar{y}_{...}^2 = \boxed{\sum_{i,j,k} y_{ijk}^2} - CT$
- $\sum_{i,j,k} (\bar{y}_{i..} - \bar{y}_{...})^2 = bn \sum_{i=1}^{a} (\bar{y}_{i..} - \bar{y}_{...})^2 = bn \sum_{i=1}^{a} \bar{y}_{i..}^2 - \frac{y_{...}^2}{abn} = bn \sum_{i=1}^{a} \bar{y}_{i..}^2 - CT$
- $\sum_{i,j,k} (\bar{y}_{ij.} - \bar{y}_{i..})^2 = n \sum_{i,j} (\bar{y}_{ij.} - \bar{y}_{i..})^2 = n \sum_{i,j}^{b} \bar{y}_{ij.}^2 - bn \sum_{i=1}^{a} \bar{y}_{i..}^2$

각 제곱합은 다음과 같음.

- 전체 제곱합 SST: $\sum_{i,j,k} (y_{ijk} - \bar{y}_{...})^2 = \sum_{i,j,k} y_{ijk}^2 - CT$
- 요인 A에 의한 제곱합 SSA: $\sum_{i,j,k} (\bar{y}_{i..} - \bar{y}_{...})^2 = bn \sum_{i=1}^{a} \bar{y}_{i..}^2 - CT$

제곱합의 분해와 분산분석-Nested Design

- 요인 B(A) 에 의한 제곱합 SSB(A) : $\sum_{i,j,k}(\bar{y}_{ij.}-\bar{y}_{i..})^2 \equiv n\sum_{i,j}^{b}\bar{y}_{ij.}^2 - bn\sum_{i=1}^{a}\bar{y}_{i..}^2$

B in A

- 오차에 의한 제곱합 SSE: $\sum_{i,j,k}(y_{ijk}-\bar{y}_{ij.})^2$

SST = SSA + SSB(A) + SSE

실제 계산에서

SSE = SST - SSA - SSB(A) 로 SSE 계산

제곱합의 분해와 분산분석-Nested Design

이원배치법의 제곱합의 분해

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

S T A B AB E

내포설계의 제곱합의 분해

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

SST SSA SSB(A) SSE

를 비교하면

$$\sum_{i,j,k}(\bar{y}_{ij.}-\bar{y}_{i..})^2 = \sum_{i,j,k}(\bar{y}_{.j.}-\bar{y}_{...})^2 + \sum_{i,j,k}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2$$ 임. 즉

SSB(A) = SSB + SSAB