제3장 일원배치법

3.1 일원배치법이란?

한 인자에서 처리의 수가 2개인 경우 → 독립표본을 이용한 두 모평균의 비교

한 인자에서 처리의 수가 3개 이상인 경우 → 일원배치법

일원배치법은 특성값에 영향을 미치는 다양한 요인 중에서 특히 알고 싶은 특정 요인의 영향을 조사하고자 할 경우 흔히 활용됨

일원배치법의 가정

- 처리를 받는 대상 또는 실험환경인 실험단위(experiment unit)가 모두 동질적

일원배치법에서는 처리를 실험단위에 배치하는 순서 또는 실험실시의 순서를 랜덤하게 결정하므로 일원배치법을 완전확률화계획이라고도 한다.

요인수준(처리수준)에는 두 가지 방법이 있다.

- 고정요인(모수인자, 고정인자): 실험자가 스스로 특정한 실험수준을 선택한 경우
  · 고정요인의 각 수준은 기술적인 의미를 갖는다
  · 고정요인의 경우 특정 수준에서의 효과와 최적수준을 구하는데 관심이 있다.
- 랜덤요인(변량인자, 랜덤인자): 특정 범위 내에서 임의로 선택된 경우
  · 특성한 수준에서의 반응치(특성치)가 기술적인 의미를 갖지 못한다.
  · 랜덤요인의 경우 수준 간의 산포의 크기인 분산성분에 관심이 있다.

〈표 3-1〉 일원배치법의 구조모형

		실험의 반복	합계	평균
인자의 수준	A_1	$x_{11}\ x_{12} \cdots x_{1r}$	$T_{1.}$	$\bar{x}_{1.}$
	A_2	$x_{21}\ x_{22} \cdots x_{2r}$	$T_{2.}$	$\bar{x}_{2.}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_a	$x_{a1}\ x_{a2} \cdots x_{ar}$	$T_{a.}$	$\bar{x}_{a.}$
			T	$\bar{\bar{x}}$

일원배치 데이터의 구조모형

$$x_{ij} = \mu_i + \epsilon_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

- 모수모형의 경우 $\sum \alpha_i = 0$ 이라고 가정함

오차 ε_{ij}에 대한 가정

· 정규성: 오차의 분포는 정규분포를 따른다.

· 독립성: 오차들은 서로 독립이다.

· 불편성: 오차의 기대값은 0으로 치우침이 없다.

· 등분산성: 오차의 분산은 인자의 수준과 실험의 반복에 관계없이 일정하다.

3.2 분산분석

귀무가설

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$

- 수준 효과 간 차이가 없다

대립가설

$H_1: \alpha_i$ 모두 0인 것은 아니다.

- 어떤 수준 효과 간 차이가 있다

변동 및 자유도의 분해

$$x_{ij} - \bar{\bar{x}} = (\bar{x}_{i.} - \bar{\bar{x}}) + (x_{ij} - \bar{x}_{i.})$$

총 변동 / α_i에 기인하는 변동 / 잔차변동

$$\sum_i \sum_j (x_{ij} - \bar{\bar{x}})^2 = \sum_i r(\bar{x}_{i.} - \bar{\bar{x}})^2 + \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2$$

$$SS_T = SS_A + SS_E$$

자유도 : $ar - 1 = a - 1 + a(r-1)$

변동의 간단한 표현

$$SS_T = \sum_i \sum_j x_{ij}^2 - CT, \quad CT = \frac{T^2}{ar}$$

$$SS_A = \sum_i \frac{T_{i.}^2}{r} - CT$$

$$SS_E = SS_T - SS_A$$

분산분석표

요인	제곱합	자유도	평균제곱	F_0	$F(\alpha)$
A	SS_A	$a-1$	MS_A	$\frac{MS_A}{MS_E}$	$F(a-1, a(r-1); \alpha)$
E	SS_E	$a(r-1)$	MS_E		
T	SS_T	$ar-1$			

가설의 검정

- 검정통계량 '$F_0 > F(a-1, a(r-1); \alpha)$' 이면 유의수준 α에서 귀무가설 기각

- 유의확률을 구하여 '유의확률 p값 < 유의수준 α' 이면 대립가설 채택

3.3 분산분석 후의 추정

각 수준에서 모평균 μ_i의 추정

일원배치법의 경우 각 수준에서 μ_i 신뢰구간 : $\bar{x}_{i.} \pm t(\phi_E; \frac{\alpha}{2})\sqrt{\frac{MS_E}{r}}$

최소유의차 검정

- 일원배치에서 두 수준 A_i 와 A_j 에서 모평균의 차이인 $\mu_i - \mu_j$ 의 $100(1-\alpha)\%$ 신뢰구간 :

$$(\bar{x}_{i.} - \bar{x}_{j.}) \pm t(\phi_E; \frac{\alpha}{2})\sqrt{MS_E \frac{2}{r}}$$

$t(\phi_E; \frac{\alpha}{2})\sqrt{MS_E \frac{2}{r}}$ 를 최소유의차(least significant differense, LSD)라고 부름

3.4 반복수가 같지 않은 실험

모형

$$x_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$
$$i = 1, \ldots, a \qquad j = 1, \ldots, r_i$$
$$SS_A = \sum_i \frac{{T_{i.}}^2}{r_i} - CT$$

분산분석표

요 인	제곱합	자유도	평균제곱	F_0	$F(\alpha)$
A	SS_A	$a-1$	MS_A	$\frac{MS_A}{MS_E}$	$F(a-1, N-a : \alpha)$
E	SS_E	$N-a$	MS_E		
T	SS_T	$N-1$			

3.5 랜덤모형

실험일, 원료로트, 블록, 원자재의 배치(batch) 등과 같이 요인(인자)의 수준이 고정되지 않고 랜덤하게 뽑아 시험하는 랜덤요인의 구조요형

$$x_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad \cdots\cdots\cdots\cdots \quad (3.25)$$

$\alpha_i \sim N(0, \sigma^2{}_A)$이고, 서로 독립

$\epsilon_i \sim N(0, \sigma^2{}_E)$이고, 서로 독립

α_i와 ϵ_{ij} : 서로 독립

랜덤모형

* 가설

$H_0 : \sigma^2{}_A = 0$ (랜덤요인이 동질적이다)

$H_1 : \sigma^2{}_A > 0$ (랜덤요인이 이질적이다)

$$^{*}\ Var(x_{ij}) = Var(\alpha_i) + Var(\varepsilon_{ij})$$
$$= \sigma^2{}_A + \sigma^2{}_E \quad \cdots\cdots\cdots\cdots \quad (3.31)$$

기여율 $\varrho = \dfrac{\sigma^2{}_A}{\sigma^2{}_A + \sigma^2{}_E} \times 100\%$

$$E(MS_A) = \sigma^2{}_E + r\sigma^2{}_A$$
$$E(MS_E) = \sigma^2{}_E$$

$$\Rightarrow \quad \widehat{\sigma^2_E} = MS_E$$
$$\widehat{\sigma^2_A} = \frac{1}{r}(MS_A - MS_E)$$

총분산의 추정치 $\hat{V}ar(x_{ij}) = \hat{\sigma}_A^2 + \hat{\sigma}_E^2$

기여율 ϱ의 추정치 : $\hat{\varrho} = \dfrac{\hat{\sigma}_A^2}{\hat{V}ar(x_{ij})} \times 100\%$