실험계획과 분석

심송용(한림대학교 데이터과학스쿨)

http://jupiter.hallym.ac.kr

a개 그룹 비교-일원배치 ANOVA

a개의 그룹 비교와 분산분석

y_{ij}: i번째 처리의 j 번째 자료

그룹	1	2	$\cdots$	i	$\cdots$	a	전체
자료	y_{11} y_{12} $\vdots$ y_{1j} $\vdots$ y_{1n_1}	y_{21} y_{22} $\vdots$ y_{2j} $\vdots$ y_{2n_2}	$\cdots$	y_{i1} y_{i2} $\vdots$ y_{ij} $\vdots$ y_{in_i}	$\cdots$	y_{a1} y_{a2} $\vdots$ y_{aj} $\vdots$ y_{an_a}	
합	$y_{1.}$	$y_{2.}$	$\cdots$	$y_{i.}$	$\cdots$	$y_{a.}$	$y_{..}$
평균	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\cdots$	$\bar{y}_{i.}$	$\cdots$	$\bar{y}_{a.}$	$\bar{y}_{..}$

각 처리 그룹의 모평균 μ_i 들이 차이가 나는지 검정

a개 그룹 비교-일원배치 ANOVA

자료가정:

모두 독립인 $y_{ij} \sim N(\mu_i, \sigma^2)$, $i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, n_i$

정규성

독립성

등분산성

부호

n_i: i번째 그룹의 자료의 개수

$N = n_1 + n_2 + \cdots + n_a$: 전체 자료수

$y_{i.} = \sum_{j=1}^{n_i} y_{ij}$: i번째 그룹의 자료합

$\bar{y}_{i.} = \frac{y_{i.}}{n_i} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$: i번째 그룹의 평균

a개 그룹 비교-일원배치 ANOVA

$y_{\cdot\cdot} = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}$: 전체 자료합

$$\bar{y}_{\cdot\cdot} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}}{n_1 + n_2 + \cdots + n_a} = \frac{y_{\cdot\cdot}}{N}$$

가설

귀무가설: $H_0 : \mu_1 = \mu_2 = \cdots = \mu_a$

대립가설: $H_1 :$ not H_0 $\Leftrightarrow$ 적어도 하나는 나머지와 다름(모두 다름이 아님에 유의)

a개 그룹 비교-일원배치 ANOVA

모형 다시 보기

$$y_{ij} = \mu_i + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0,\sigma^2)$$

$\mu_i = \mu + \tau_i$ 라고 하면 (단, $\sum_{i=1}^{a} \tau_i = 0$)

$$\boxed{y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0,\sigma^2)}$$

로 쓸 수 있음. 이 경우

$$\tau_i = \mu_i - \mu$$

이며 각 처리군의 평균과 전체 평균의 차이를 의미

귀무가설: $H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$

대립가설: $H_1 :$ not $H_0 \Leftrightarrow$ 적어도 하나의 τ_i는 0이 아님.(모두 0이 아님에 유의)