실험계획과 분석

심송용(한림대학교 데이터과학스쿨)

http://jupiter.hallym.ac.kr

a개 그룹 비교-일원배치 ANOVA

모수추정

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

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

에서 모수는 σ^2, μ_i, μ, τ_i 이며 ϵ_{ij}는 정규분포를 따르는 확률변수.

추정량

- $\hat{\sigma}^2 = \mathrm{MSE} = \dfrac{\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2}{N - a}$
- $\hat{\mu}_i = \bar{y}_{i.}$
- $\hat{\mu} = \bar{y}_{..}$
- $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$

a개 그룹 비교-일원배치 ANOVA

MSE가 σ^2의 추정량

독립 이표본 t-검정에서 공통분산 σ^2의 추정량은

$$\hat{\sigma}^2 = S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{\sum_{j=1}^{n_1}(y_{1j}-\bar{y}_{1.})^2 + \sum_{j=1}^{n_1}(y_{2j}-\bar{y}_{2.})^2}{n_1+n_2-2}$$

인데 비교할 그룹의 수가 $a=2$인 경우로 보고 다시 써보면 $S_p^2 = \dfrac{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2}{N-2}$

가 되며 MSE의 공식에서 $a=2$인 경우임.

S_i^2를 i번째 처리군의 표본분산이라 하면(각각의 S_i^2은 σ^2의 추정량)

$S_i^2 = \dfrac{\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2}{n_i-1}$ 이며 $(n_i-1)S_i^2 = \sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2$인 관계에서

a개 그룹 비교-일원배치 ANOVA

$$\text{MSE} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{i.})^2}{N-a} = \frac{\sum_{j=1}^{n_1}(y_{1j}-\overline{y}_{1.})^2 + \sum_{j=1}^{n_2}(y_{2j}-\overline{y}_{2.})^2 + \cdots + \sum_{j=1}^{n_a}(y_{aj}-\overline{y}_{a.})^2}{n_1+n_2+\cdots+n_a-a}$$

$$= \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2 + \cdots + (n_a-1)S_a^2}{n_1+n_2+\cdots+n_a-a}$$

로 S_i^2 들의 가중평균임을 알 수 있다.

각각의 S_i^2은 σ^2의 추정량이므로 이들의 자료수에 비례하는 가중평균은 σ^2의 합리적인 추정량임을 짐작할 수 있다.

a개 그룹 비교-일원배치 ANOVA

μ_i에 대한 추론

$y_{ij} \sim N(\mu_i, \sigma^2)$이므로 이들의 평균은 $\bar{y}_{i.} \sim N(\mu_i, \frac{\sigma^2}{n_i})$임을 알 수 있다. 표준화하면

$\frac{\bar{y}_{i.} - \mu_i}{\sigma / \sqrt{n_i}} \sim N(0,1)$이며 σ에 대한 추정량으로 $\sqrt{MSE}$를 사용하면

$$\boxed{\frac{\bar{y}_{i.} - \mu_i}{\sqrt{MSE} / \sqrt{n_i}} \sim t_{N-a}}$$

참고: t-분포의 정의는 독립인 두 확률변수 $Z \sim N(0,1)$, χ^2_ν ~ 자유도가 ν인 카이제곱분포를 따르는 확률변수일 때 $t_\nu = \frac{Z}{\sqrt{\chi^2_\nu / \nu}}$이다. 즉, t-분포의 자유도는 나누어주는 카이제곱분포의 자유도를 그대로 가져온다. 위의 식에서 카이제곱분포는 MSE 안에 숨어 있으며 이 자유도가 $(N-a)$이다.

a개 그룹 비교-일원배치 ANOVA

μ_i의 신뢰구간

$$\frac{\bar{y}_{i.} - \mu_i}{\sqrt{MSE}/\sqrt{n_i}} \sim t_{N-a}$$ 이므로

$$1-\alpha = \Pr[-t_{N-a;\alpha/2} < \frac{\bar{y}_{i.} - \mu_i}{\sqrt{MSE}/\sqrt{n_i}} < t_{N-a;\alpha/2}]$$

$$= \Pr[-t_{N-a;\alpha/2}\frac{\sqrt{MSE}}{\sqrt{n_i}} < \bar{y}_{i.} - \mu_i < t_{N-a;\alpha/2}\frac{\sqrt{MSE}}{\sqrt{n_i}}]$$

$$= \Pr[\bar{y}_{i.} - t_{N-a;\alpha/2}\frac{\sqrt{MSE}}{\sqrt{n_i}} < \mu_i < \bar{y}_{i.} + t_{N-a;\alpha/2}\frac{\sqrt{MSE}}{\sqrt{n_i}}]$$

이므로 μ_i의 100(1-α)% 신뢰구간은

$$\boxed{\mu_i = \bar{y}_{i.} \pm t_{N-a;\alpha/2}\frac{\sqrt{MSE}}{\sqrt{n_i}}}$$

이다.

a개 그룹 비교-일원배치 ANOVA

$H_0 : \mu_i = \mu_{i0}$ **에 대한 가설검정**

귀무가설이 참이면 $t_0 = \dfrac{\bar{y}_{i.} - \mu_{i0}}{\sqrt{MSE} / \sqrt{n_i}} \sim t_{N-a}$이므로

대립가설이

- $H_1 : \mu_i > \mu_{i0}$일 때 $t_0 > t_{N-a;\alpha}$일 때 H_0 기각
- $H_1 : \mu_i < \mu_{i0}$일 때 $t_0 < -t_{N-a;\alpha}$일 때 H_0 기각
- $H_1 : \mu_i \neq \mu_{i0}$일 때 $|t_0| > t_{N-a;\alpha/2}$일 때 H_0 기각

$\Leftrightarrow$ 유의확률 P가 유의수준 α보다 작을 때 H_0 기각

대립가설이

- $H_1 : \mu_i > \mu_{i0}$일 때 $P = \Pr[t_{N-a} > t_0]$
- $H_1 : \mu_i < \mu_{i0}$일 때 $P = \Pr[t_{N-a} < t_0]$
- $H_1 : \mu_i \neq \mu_{i0}$일 때 $P = 2\Pr[t_{N-a} > |t_0|]$

a개 그룹 비교-일원배치 ANOVA

$\mu_i - \mu_j = \tau_i - \tau_j$에 대한 추론

모평균 차 $\mu_i - \mu_j$에 대한 100(1-α)% 신뢰구간

$$1-\alpha = \Pr\left[-t_{N-a;\alpha/2} < \frac{\bar{y}_{i.} - \bar{y}_{j.} - (\mu_i - \mu_j)}{\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}} < t_{N-a;\alpha/2}\right]$$

$$= \Pr\left[-t_{N-a;\alpha/2}\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}} < \bar{y}_{i.} - \bar{y}_{j.} - (\mu_i - \mu_j) < t_{N-a;\alpha/2}\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}\right]$$

=

$$\Pr\left[(\bar{y}_{i.} - \bar{y}_{j.}) - t_{N-a;\alpha/2}\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}} < \mu_i - \mu_j < (\bar{y}_{i.} - \bar{y}_{j.}) + t_{N-a;\alpha/2}\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}\right]$$

에서

$$\boxed{\mu_i - \mu_j = (\bar{y}_{i.} - \bar{y}_{j.}) \pm t_{N-a;\alpha/2}\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}}$$

a개 그룹 비교-일원배치 ANOVA

귀무가설 $H_0 : \mu_i - \mu_j = \delta$**에 대한 가설검정**

귀무가설이 참이면 $t_0 = \dfrac{\bar{y}_{i.} - \bar{y}_{j.} - \delta}{\sqrt{MSE}\sqrt{\dfrac{1}{n_i} + \dfrac{1}{n_j}}} \sim t_{N-a}$이므로

대립가설이

- $H_1 : \mu_i - \mu_j > \delta$일 때 $t_0 > t_{N-a;\alpha}$일 때 H_0 기각
- $H_1 : \mu_i - \mu_j < \delta$일 때 $t_0 < -t_{N-a;\alpha}$일 때 H_0 기각
- $H_1 : \mu_i - \mu_j \neq \delta$일 때 $|t_0| > t_{N-a;\alpha/2}$일 때 H_0 기각

$\Leftrightarrow$ 유의확률 P가 유의수준 α보다 작을 때 H_0 기각

대립가설이

- $H_1 : \mu_i - \mu_j > \delta$일 때 $P = \Pr[t_{N-a} > t_0]$
- $H_1 : \mu_i - \mu_j < \delta$일 때 $P = \Pr[t_{N-a} < t_0]$
- $H_1 : \mu_i - \mu_j \neq \delta$일 때 $P = 2\Pr[t_{N-a} > |t_0|]$