다중비교

정의

- 개요

1. 여러 개의 모수들을 따르는 확률변수들의 선형결합에 대한 신뢰구간을 구하는 것

1) $[\mu_1, \cdots, \mu_n]$, 공통분산 σ^2을 가지는 서로 독립인 확률변수 $X_1, \cdots, X_n$에서 추출한 를 가정하자.

(1) 이 때, 모수 $\mu_1, \cdots, \mu_k$가 선형결합한 일차함수 $\sum k_j \mu_j$(단 k_1, ⋯, k_j는 어떤 알려진 실수)의 신뢰구간을 구한다고 하자.

(2) 이 때, 각각의 확률변수에서 추출한 확률표본을 X_{ij}라고 하자.

$$\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1j} \\ X_{21} & X_{22} & \cdots & X_{2j} \\ \vdots & \vdots & \cdots & \vdots \\ X_{i1} & X_{i2} & \cdots & X_{ij} \end{bmatrix}$$

(3) 이 때, 각 열들의 평균과 x^2을 정의하면

- $\overline{x}_j = \frac{\sum x_{ij}}{a}$ (j = 1,2, ⋯,b) 라고 한다면, 이 통계량은 $\mathrm{N}(\mu_j, \frac{\sigma^2}{a})$을 따른다.
- $\frac{\sum (x_{ij} - \overline{x}_j)^2}{\sigma^2}$(j=1, ⋯,b)는 $x^2(\mathrm{a}-1)$을 따른다.

(4) 이 때, 위치불변통계량 $\frac{\sum (x_{ij} - \overline{x}_j)^2}{\sigma^2}$은 완비충분통계량의 함수 $\frac{\sum x_{ij}}{a}$와 확률적으로 독립이다.

정의

- 개요

2) 한편, x^2(a – 1)들의 결합분포인 $\sum_{j=1}^{b}\frac{\sum_{i=1}^{a}(x_{ij}-\overline{x}_j)^2}{\sigma^2}$는 x^2(b[a – 1])을 따르고 이 또한 완비충분통계량의 함수 $\frac{\sum x_{ij}}{a}$와 확률적으로 독립이다.

3) 이 때, 관심있는 확률변수 Z = $\sum k_j\overline{x}_j$ 는 다변량 정규분포의 성질에 따라 N($\sum k_j\overline{x}_j$, $\frac{\sum {k_j}^2\sigma^2}{a}$) 을 따른다.

(1) 이 때, $\sum k_j \frac{\sum x_{ij}}{a}$는 또한 다변량 정규분포의 완비충분통계량의 함수이고, 따라서

(2) $\frac{\sum\sum(x_{ij}-\overline{x}_j)^2}{b(a-1)}$ ~x^2(b[a – 1])는 위와 확률적으로 독립이다.

4) 여기까지 전개하여 종합하였을 때, $\frac{\sum k_j\overline{x}_j-\sum k_j\mu_j \Big/ \sqrt{\frac{\sum {k_j}^2\sigma^2}{a}}}{\sqrt{\frac{\sum\sum(x_{ij}-\overline{x}_j)^2 / b(a-1)}{\sigma^2}}}$ = $\frac{\sum k_j\overline{x}_j-\sum k_j\mu_j}{\sqrt{V/a\sum {k_j}^2}}$ = $\frac{N}{\sqrt{V/n}}$은 T[b(a-1)]을 따른다.

정의

- 개요

2. 위에서 도출한 통계량을 이용하면

1) $\mathrm{P}\left[-\mathrm{c} \le \sum k_j \bar{x}_j \le c\right]$ 에서

2) $\mathrm{P}\left[-\mathrm{c} \le \sum \frac{\sum k_j \bar{x}_j - \sum k_j \mu_j}{\sqrt{V/a \sum {k_j}^2}} \le c\right] = P\left[\sum k_j \bar{x}_j - \mathrm{c}\sqrt{\frac{V}{a \sum {k_j}^2}} \le \sum k_j \mu_j \le \sum k_j \bar{x}_j + c\sqrt{\frac{V}{a \sum {k_j}^2}}\right]$

3. 이제, 이 결합 모수에 대한 신뢰구간을 통해 다양한 일차 함수의 신뢰구간을 구할 수 있다.

1) 예를 들어, $\mu_3 - \frac{\mu_1 + \mu_2}{2}$, $\mu_2 + \mu_1$ 등의 1차 선형함수를 고려해볼 수 있다.

2) 이 때 첫번 째 예의 경우 k = $\begin{bmatrix} 1 \\ -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$ 이고, 두 번째 예의 경우 k = $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ 이고

(1) 벡터의 각 원소들의 선형결합 $\sum {k_j}^2$을 통해 신뢰구간을 정의할 수 있다.

정의

- 셰페의 다중 비교

1. 위에서 예시로 든 것과 같은 선형함수들이 동시에 포함된 경우의 신뢰구간을 계산
2. $\mathrm{P}[F \geq d]$에서, 이 F분포를 풀어서 $\mathrm{P}[\sum k_j\mu_j \geq K]$꼴의 신뢰구간을 정의한다.

1) 이 때, 어떤 상수 d = $\mathrm{F}[\alpha, b, b(a-1)]$, 즉 유의수준 α일때의 F 통계량 값이다.

3. 이를 유도하면

1) $\frac{\sum_{j=1}^{b}\left(\bar{x}_j-\mu_j\right)^2}{\sigma^2/_a}$ ~ x^2(b) 이고, 이는 $\bar{x}_j$에 대한 통계량이다. 한편

(1) V = $\frac{\sum_{j=1}^{b}\sum_{j=1}^{a}(x_{ij}-\bar{x}_j)^2}{b(a-1)}$는

$\bar{x}_j$에 대한 함수인 함수 $\frac{\sum_{j=1}^{b}\left(\bar{x}_j-\mu_j\right)^2}{\sigma^2/_a}$와 확률적으로 독립이다.

(2) 따라서 $\frac{a\sum_{j=1}^{b}\left(\bar{x}_j-\mu_j\right)^2/_b}{V}$ 는 F[b,b(a-1)] 를 따른다.

정의

- 셰페의 다중 비교

2) $\mathrm{P}[\mu \leq d] = 1 - a$ 에서

(1) $\mathrm{P}\left[\sum_{j=1}^{b}(\overline{x}_j - \mu_j)^2 \leq \frac{bd}{a}V\right] = 1 - a$ 이고,

이제 목표는 $\sum_{j=1}^{b}(\overline{x}_j - \mu_j)^2 \Rightarrow [\sum_{j=1}^{b}(\overline{x}_j - \mu_j)]^2$로, 시그마를 제곱항 안으로 집어넣는 것이다.

(2) 이를 실현하기 위해, 기하학적인 방법론을 빌려오면

- $\overline{x}_j$와 μ_j를 포함하는 초평면의 방정식은

$k_1(\overline{x}_1 - \mu_1) + k_2(\overline{x}_2 - \mu_2) + \cdots + k_b(\overline{x}_b - \mu_b) = 0$ 이고,

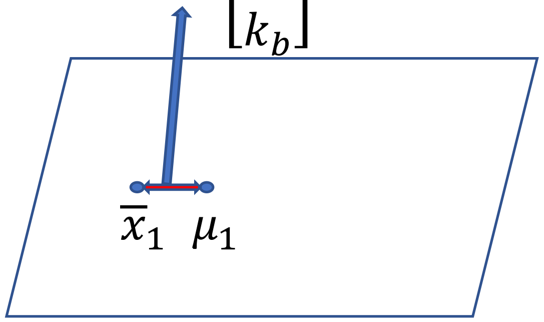
- 어떤 점과 이 초평면의 최단거리는

$\frac{\mathrm{d} - k_1(\overline{x}_1 - \mu_1) - \cdots - k_b(\overline{x}_b - \mu_b)}{\sqrt{k_1^2 + k_2^2 + \cdots + k_n^2}}$ 로 나타낼 수 있다.

- 이 때, d = 0 이기 때문에, 이를 다시 정리하면

$$\frac{-k_1(\overline{x}_1 - \mu_1) - \cdots - k_b(\overline{x}_b - \mu_b)}{\sqrt{\sum k_j^2}}$$

$\vec{K} = \begin{bmatrix} k_1 \\ \vdots \\ k_b \end{bmatrix}$



정의

- 셰페의 다중 비교

(3) $\frac{-k_1(\overline{x}_1 -\ \mu_1)-\cdots-k_b(\overline{x}_b -\ \mu_b)}{\sqrt{\sum {k_j}^2}}$ 을 제곱하면

- $\frac{[k_1(\overline{x}_1 -\ \mu_1)+\cdots+k_b(\overline{x}_b -\ \mu_b)]^2}{\sum {k_j}^2} = \frac{\left[\sum_{i=1}^{b} k_i(\overline{x}_i -\ \mu_i)\right]^2}{\sum {k_j}^2}$ 이것이 원하는 결과다.

3) $\frac{\left[\sum_{i=1}^{b} k_i(\overline{x}_i -\ \mu_i)\right]^2}{\sum {k_j}^2} = \frac{\left[\sum_{i=1}^{b} k_i\overline{x}_i - \sum_{i=1}^{b} k_i\mu_i)\right]^2}{\sum {k_j}^2}$ 이고

(1) 다시, $\mathrm{P}\left[\sum_{j=1}^{b}\left(\overline{x}_j - \mu_j\right)^2 \le \frac{bd}{a}v\right] = \mathrm{P}\left[\frac{\left[\sum_{i=1}^{b} k_i\overline{x}_i - \sum_{i=1}^{b} k_i\mu_i)\right]^2}{\sum {k_j}^2} \le \frac{bd}{a}v\right]$

$= \mathrm{P}\left[\left[\sum_{i=1}^{b} k_i\overline{x}_i - \sum_{i=1}^{b} k_i\mu_i)\right]^2 \le \sum {k_j}^2 \frac{bd}{a}v\right]$

(2) 위 부등식에 제곱근을 씌워도, 그 정보는 그대로 보존된다. 따라서

$\mathrm{P}\left[|\sum_{i=1}^{b} k_i\overline{x}_i - \sum_{i=1}^{b} k_i\mu_i| \le \sqrt{\sum {k_j}^2 \frac{bd}{a}v}\right]$

정의

- 셰페의 다중 비교

4. 이를 이용해 신뢰구간을 정의하면

(1) $P\left[\sum_{i=1}^{b} k_i\bar{x}_i - \sqrt{\sum k_j^{\,2}\frac{bd}{a}V} \le \sum_{i=1}^{b} k_i\mu_i \le \sum_{i=1}^{b} k_i\bar{x}_i + \sqrt{\sum k_j^{\,2}\frac{bd}{a}V}\right] = 1 - a$

(2) 다시 정리하면, 이 때

- d = $F[\alpha, b, b(a-1)]$ 인 F통계량 값이고
- V = $\frac{\sum_{j=1}^{b}\sum_{j=1}^{a}(x_{ij}-\bar{x}_j)^2}{b(a-1)}$ ~ x^2[b(a-1)]
- a,b는 행, 열의 차원인 상수
- k_i는 선형결합을 나타내는 벡터의 원소들이다.

정의

- 본페로니 다중비교

1. 확률변수 $T_1 = \frac{\sum k_j \bar{x}_j - \sum k_j \mu_j}{\sqrt{\sum {k_j}^2 \frac{V}{a}}}$인 T통계량을 정의하자.

1) 사건 ${A_i}^c$이 $-c_i \le T_i \le c_i$ 로 주어졌다고 할 때,

(1) $-c_i \le \frac{\sum k_j \bar{x}_j - \sum k_j \mu_j}{\sqrt{\sum {k_j}^2 \frac{V}{a}}} \le c_i$ 에서

(2) $-c_i\sqrt{\sum {k_j}^2 \frac{V}{a}} \le \sum k_j \bar{x}_j - \sum k_j \mu_j \le c_i\sqrt{\sum {k_j}^2 \frac{V}{a}} \Rightarrow \sum k_j \bar{x}_j - c_i\sqrt{\sum {k_j}^2 \frac{V}{a}} \le \sum k_j \mu_j \le \sum k_j \bar{x}_j + c_i\sqrt{\sum {k_j}^2 \frac{V}{a}}$

2) $U_i = \sum k_j \bar{x}_j - c_i\sqrt{\sum {k_j}^2 \frac{V}{a}}$, $W_i = \sum k_j \bar{x}_j + c_i\sqrt{\sum {k_j}^2 \frac{V}{a}}$ 라고 정의하자.

(1) 이 때, $c_i = t_{\frac{a}{2m}, b(a-1)}$라고 한다면, P$({A_i}^c) = 1 - \frac{a}{m}$, P$(A_i) = \frac{a}{m}$ 이다.

예제

- 본페로니 다중비교

1. 어떤 선형함수들의 집합 $\{\mu_1 - \mu_2,\ \mu_2 - \mu_3,\ \mu_3 - \mu_4,\ \mu_4 - \frac{\mu_5+\mu_6}{2},\ \frac{\mu_1+\cdots+\mu_6}{6}\}$의 결합 신뢰구간을 구하라.

1) 위 선형결합을 행렬 K로 표현하면 다음과 같다.

	식1	식2	식3	식4	식5
k_1	1	0	0	0	$\frac{1}{6}$
k_2	-1	1	0	0	$\frac{1}{6}$
k_3	0	-1	1	0	$\frac{1}{6}$
k_4	0	0	-1	$\frac{1}{2}$	$\frac{1}{6}$
k_5	0	0	0	$\frac{1}{2}$	$\frac{1}{6}$
k_6	0	0	0	0	$\frac{1}{6}$

예제

- 본페로니 다중비교

1. 어떤 선형함수들의 집합 $\{\mu_1 - \mu_2, \mu_2 - \mu_3, \mu_3 - \mu_4, \mu_4 - \frac{\mu_5+\mu_6}{2}, \frac{\mu_1+\cdots+\mu_6}{6}\}$의 결합 신뢰구간을 구하라.
   (단, b=6, a = 3, $\alpha = 0.05$)

2) 따라서,

(1) $\sum {k_j}^2$ = 1+1+1+1+1+1+ $\frac{1}{4}$+ $\frac{1}{4}$+$\frac{1}{6}$ = $\frac{46}{6}$ 이고

(2) V~ x^2[3(6-1)] = 7.261

(3) $c_i = t_{\frac{a}{2m}, b(a-1)} = t_{0.005,15}$=3.732

3) 이를 통해 결합 신뢰구간을 정의할 수 있고,
   이를 벗어나는 어떤 선형결합은 그 값이 유의미하게 다름을 입증할 수 있다.