평균분석을 이용한 교호작용의 균형 검정 방법
-불균형 관측치의 경우-

김 병천*　　정 강모**

요　약

관측치가 불균형을 이룰 때, Nelson(1988)에 의한 균형 관측치에 관한 이론이 확장될 수 있음을 보여 주고 있다. 이 방법을 이용하여 이원 요인 계획법에서의 교호작용항에 대한 통계량을 구하고, 그래프를 이용한 방법을 제시 하였다.

1. 서　　론

평균분석을 이용한 모형분석은 처음으로 Ott(1967)에 의하여 제안되었다고 볼 수 있다. 이 분석방법은 Halperin(1955) 등의 다중 유의성 검정에 기초를 한 것으로, 독립적인 K개의 평균을 비교하는 방법이었으나, Ott(1967)의 방법은 이원 요인 계획법에서 하나의 요인에 대한 수준이 2일 때로 제한 되고 있다. Schilling(1973)은 요인수준의 수에 제한을 두지 않고 교호작용을 검정할 수 있는 방법(처리효과의 평균 분석(Anaysis of Means for Treatment of Effects)이라고 부른다)을 제시하였으나 이 Schilling의 방법은 실제적으로 의미가 없는 것으로 알려지고 있다. 특히 관측치가 균형을 이룰 때, Nelson(1988)은 상관계수의 관계를 이용하여 교호작용을 검정할 수 있는 새로운 통계량을 개발하였다. 일반적으로 분산분석에 의해 통계적 유의성이 클 때, 평균분석방법은 분산분석방법과 거의 대등한 검정력을 갖는 것으로 알려져 있으며, 요인의 수가 클 때 평균분석방법이 좀 더 검정력이 큰 것으로 알려져 있다(Nelson, 1983). 평균분석방법의 가장 큰 장점은 요인에 대한 효과를 비교하는데 있어 그래프를 이용할 수 있는 장점이 있다.

본 논문의 2절에서 모형설정에 관하여 설명하며, 3절에서는 관측치가 불균형을 이룰 때 평균 분석방법에 대하여 연구 하였으며, 4절에서는 예제를 들어 설명한다.

* 한국과학기술원 수학과
** 삼성종합기술원

2. 모형과 기호

본 논문에서 다루고자 하는 모형은 이원 요인 모형이다. 이원요인 모형을 식으로 나타내면 다음과 같다.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \tag{2. 1}$$
$$i = 1, \cdots, p\,;\ j=1, \cdots, q\,;\ k=1, \cdots, n_{ij}\,;\ n_{ij}>0$$

여기서 μ, α_i, β_j 와 γ_{ij}는 고정효과이고, $\varepsilon_{ijk} \sim N(0, \sigma^2)$이다. 이 모형을 간단히 나타내면 다음과 같다.

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}\text{이며, } \mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}. \tag{2. 2}$$

다음은 교호작용에 대한 검정을 하기 위하여 사용되는 기호들에 대하여 설명한다.

$$z_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk}, \qquad z_{i\cdot} = \frac{1}{q} \sum_{j=1}^{q} z_{ij}$$

$$z_{\cdot j} = \frac{1}{p} \sum_{j=1}^{p} z_{ij}, \qquad z_{\cdot\cdot} = \frac{1}{pq} \sum_{i=1}^{p} \sum_{j=1}^{q} z_{ij}$$

$$N = \sum_{i=1}^{p} \sum_{j=1}^{q} n_{ij}$$

여기서 검정하고자 하는 가설 H는 교호작용에 관한 가설로써 다음과 같다.

$$H : \text{모든 } i, j\text{에 대하여 } \gamma_{ij}=0\text{이다.} \tag{2. 3}$$

그러나, γ_{ij}는 추정 불가능하므로, Graybill(1976)에 의한 추정가능한 교호효과를 이용하면

$$H : \text{모든 } i, j\text{에 대하여 } \gamma^*_{ij}=0\text{이다.} \tag{2. 4}$$

여기서 $\gamma^*_{ij}=\gamma_{ij}-\gamma_{i\cdot}-\gamma_{\cdot j}+\gamma_{\cdot\cdot}$이며, γ^*_{ij}의 추정값은 $\hat{\gamma}^*_{ij} = z_{ij}-z_{i\cdot}-z_{\cdot j}+z_{\cdot\cdot}$가 됨을 쉽게 알 수 있다. 이 가설은 추정 가능하며, γ^*_{ij}는 이원요인모형의 교호작용의 효과에 대하여 검정을 하는 데 사용되며, 이 가설에 대해서는 Graybill(1976, pp. 575-576)이 논하였으며 이러한 가설을 균형 가설 검정이라고도 부른다. 통계페키지(SPSS^X, SAS, BMDP)들이 다루고 있는 추정가능한 가설에 대하여 Hemmerle(1979)와 Speed, Hocking, Hackney(1978)들이 자세히 논하였다.

3. 불균형 관측치 경우의 평균분석 방법

모형 (2. 1)에 대하여 γ^*_{ij}의 추정치 $\hat{\gamma}^*_{ij}$의 분산을 구하면 다음의 결과를 얻을 수 있다.

$$\mathrm{Var}(\hat{\gamma}^*_{ij}) = \frac{1}{p^2q^2}\left\{(p-1)^2(q-1)^2\frac{1}{n_{ij}} + (p-1)\sum_{j'\neq j}^{q}\frac{1}{n_{ij'}} + (q-1)^2\sum_{i'\neq i}^{p}\frac{1}{n_{i'j}} + \sum_{i'\neq i}^{p}\sum_{j'\neq j}^{q}\frac{1}{n_{i'j'}}\right\}\sigma^2 \tag{3. 1}$$

이 분산의 계산과정은 부록에 나타나 있는 정리에 의하여 쉽게 얻을 수 있음과 동시에 n 요인 계획법에서의 모든 요인의 교호작용에 대한 분산을 수학적 귀납법을 이용하여 구할 수 있다. 또한 (3. 1)의 결과를 이용하여 다음의 공분산을 구할 수 있다.

$$\mathrm{Cov}(\hat{\gamma}^*_{ij}, \hat{\gamma}^*_{i'j'})$$

$$= \frac{1}{p^2q^2}\left[pq(2-q)\left(\frac{1}{n_{ij}} + \frac{1}{n_{i'j}}\right) + q(q-2)\sum_{i}^{p}\frac{1}{n_{ij}} - p\left\{\sum_{j}^{q}\left(\frac{1}{n_{ij}} + \frac{1}{n_{i'j}}\right)\right\} + \sum_{i}^{p}\sum_{j}\frac{q}{n_{ij}}\right]\sigma^2 \quad ; i'\neq i,\ j'=j.$$

$$= \frac{1}{p^2q^2}\left[pq(2-q)\left(\frac{1}{n_{ij}} \quad \frac{1}{n_{ij'}}\right) + p(p-2)\sum_{j}^{q}\frac{1}{n_{ij}} - q\left\{\sum_{j}^{q}\left(\frac{1}{n_{ij}} + \frac{1}{n_{ij'}}\right)\right\} + \sum_{i}^{p}\sum_{j}^{q}\frac{1}{n_{ij}}\right]\sigma^2 \quad ; i'=i,\ j'\neq j.$$

$$= \frac{1}{p^2q^2}\left[pq\left(\frac{1}{n_{ij}} + \frac{1}{n_{i'j}} + \frac{1}{n_{ij'}} + \frac{1}{n_{i'j'}}\right) - p\sum_{j}^{q}\left(\frac{1}{n_{ij}} + \frac{1}{n_{i'j}}\right) - q\sum_{i}^{p}\left(\frac{1}{n_{ij}} + \frac{1}{n_{ij'}}\right) + \sum_{i}^{p}\sum_{j}^{q}\frac{1}{n_{ij}}\right]\sigma^2 \quad ; i'\neq i,\ j'\neq j. \tag{3. 2}$$

3. 1 각 요인의 수준이 2인 경우

먼저 각 요인의 수준이 2인 경우를 생각해 보면, 즉 $p=q=2$, (3. 1)과 (3. 2)의 결과에 의해 다음과 같은 상관 계수를 구할 수 있다. 그 결과

$$\begin{aligned} \mathrm{Corr}(\hat{\gamma}^*_{ij}, \hat{\gamma}^*_{i'j'}) &= -1, \quad i'\neq i,\ j'=j \\ &= -1, \quad i'=i,\ j'\neq j \\ &= \ \ 1, \quad i'\neq i,\ j'\neq j \end{aligned} \tag{3. 3}$$

임을 보일 수 있다. 이 상관 계수가 모두 1아니면 -1이므로, (i, j)조합 중에서 어느 하나의 교호작용에 대하여 검정을 하면 충분하다. 그러나, 이론을 일반적으로 확장시키기 위하여 $\hat{\gamma}^*_{2j}$, $\hat{\gamma}^*_{2j}$모두 고려해야 한다. 만약 그래프를 이용한 교호작용을 보면 그림 1과 같다.

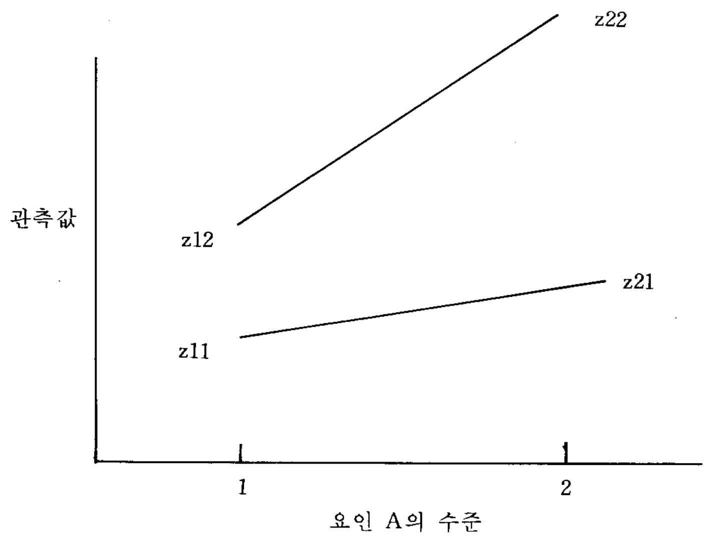

그림 1. 이원 모형의 교호작용에 관한 그래프

그림 1에서 보는 바와 같이 두 직선이 서로 평행이면 교호작용이 없고, 평행이 아니면 교호작용이 있다. 그러므로 교호작용을 검정하는 데 두 직선의 기울기를 비교하는 것이 자연스러운 방법이다. $j=1,\ 2$에 대하여 $X_j = z_{2j} - z_{1j}$는 두직선의 기울기를 나타내는데, 평균 분석을 이용한 두 직선의 기울기에 대한 검정의 통계량으로 다음을 이용할 수 있다.

$$T_j = (X_j - X.)/\left\{\frac{1}{2}\, s\, \sqrt{\sum_i \sum_j \left(\frac{1}{n_{ij}}\right)}\right\}, \qquad j=1,\ 2$$

단, $X.$는 X_j들의 평균이고, s^2은 분산분석에서 구한 평균제곱오차이며, s^2은 자유도 υ에서,

$$s^2 = \left(\sum_i \sum_j \sum_k y^2_{ijk} - \sum_i \sum_j n_{ij} z^2_{ij}\right)/\upsilon, \qquad \upsilon = N - pq$$

이다. 또

$$\mathrm{Var}(X_j - X.) = \frac{1}{4}\left(\sum_i \sum_j \frac{1}{n_{ij}}\, \sigma^2\right)$$

이다. 그러므로 그래프를 이용한 분석에서는

$$X. \pm \frac{1}{2}\, t(\alpha/2\,;\, \upsilon) s\ \sqrt{\sum_i \sum_j \frac{1}{n_{ij}}}$$

을 계산하여 X_1 또는 X_2가 이 구간 밖에 있으면 가설 H'을 기각한다. 이와 같은 관계는

$$\hat{\gamma}^*_{2j} = z_{2j} - z_{2.} - z_{.j} + z_{..}$$
$$= z_{2j} - \frac{z_{21} + z_{22}}{2} - \frac{z_{1j} + z_{2j}}{2} + \frac{z_{11} + z_{12} + z_{21} + z_{22}}{4}$$
$$= \frac{z_{2j} - z_{1j}}{2} - \frac{(z_{21} - z_{11}) + (z_{22} - z_{12})}{4}$$
$$= \frac{1}{2}(X_j - X.)$$

이므로, T_j는 $\hat{\gamma}^*_{2j}$를 스튜던트화하여 계산한 것이다.

3. 2 한 요인의 수준이 2인 경우(p=2, q>2)

p=2, q>2인 경우 (3. 1)과 (3. 2)식으로부터 다음의 관계식을 얻는다.

$$\mathrm{corr}(\hat{\gamma}^*_{ij}, \hat{\gamma}^*_{i'j'}) = -1, \qquad i' \neq i,\ j' \neq j.$$

이 결과로부터 $\hat{\gamma}^*_{1j}$와 $\hat{\gamma}^*_{2j}$는 완전하게 음으로 상관되어 있으므로 교호 작용을 검정함에 있어 $\hat{\gamma}^*_{2j}$에 대하여 검정을 하면 충분하다. j=1, 2, …, q에 대하여

$$X_j = Z_{2j} - Z_{1j},$$
$$\delta_j = \frac{1}{q^2}\left\{q(q-2)\sum_i \frac{1}{n_{ij}} + \sum_i \sum_j \frac{1}{n_{ij}}\right\},$$
$$X_j = X_j - X.$$

라 할 때 $\hat{\gamma}^*_{2j}$는 다음의 관계식을 얻을 수 있다.

$$\hat{\gamma}^*_{2j} = Z_{2j} - Z_{2.} - Z_{.j} + Z_{..}$$
$$= Z_{2j} - \frac{1}{q}\sum_{j'=1}^{q} Z_{2j'} - \frac{1}{2}(Z_{1j} + Z_{2j}) + \frac{1}{2q}\sum_{j'=1}^{q}(Z_{1j'} + Z_{2j'})$$
$$= \frac{1}{2}(Z_{2j} - Z_{1j}) - \frac{1}{2q}\sum_{j'=1}^{q}(Z_{2j'} - Z_{1j'})$$
$$= \frac{1}{2}(X_j - X.). \qquad (j=1, 2, \cdots, q)$$

3.1절에서와 마찬가지로 T_j는 $\hat{\gamma}^*_{2j}$를 스튜던트화 한 것이 된다. 또 $X_j - X.$의 분산은 다음과 같은 식으로 나타난다.

$$\mathrm{Var}(X_j - X.) = 4\mathrm{var}(\hat{\gamma}^*_{2j})$$
$$= \frac{1}{q^2}\left\{(q-1)^2\frac{1}{n_{2j}} + \sum_{j' \neq j}\frac{1}{n_{2j'}} + (q-1)^2\frac{1}{n_{1j}} + \sum_{j' \neq j}\frac{1}{n_{1j'}}\right\}\sigma^2$$

$$= \frac{1}{q^2}\left\{q(q-2)\sum_i \frac{1}{n_{ij}} + \sum_i\sum_j \frac{1}{n_{ij}}\right\}\sigma^2$$
$$= \delta_j\sigma^2, \qquad j=1,\ 2,\ \cdots,\ q.$$

이 결과들을 이용하여 T_j의 통계량에 적절한 기각값을 구하여야 한다. 이는,

$$A_j = \{\ |T_j| < C_j \quad ;\ j=1,\ 2,\ \cdots,\ q \quad ;\ C_j>0\}$$

라 하면 주이진 α에 대하여 다음을 만족하는 C_j를 구한다.

$$P(\bigcap_{j=1}^{q} A_j) = 1-\alpha \tag{3. 4}$$

그러나 관측치가 불균형을 이루고 있는 경우, 관측치가 균형을 이루고 있을 때와는 달리 T_j는 동일 상관값을 갖고 있기 때문에 일반적인 다중 t분포에 대하여 정확한 기각값을 구한다는 것이 쉬운 일은 아니다. T_j들에 대한 상관 계수 행렬의 비대각원소를 구하면 다음과 같다.

$$\mathrm{Corr}(T_j,\ T_{j'})$$
$$= \frac{-q\left(\sum_i \frac{1}{n_{ij}} + \sum_i \frac{1}{n_{ij'}}\right) + \sum_i\sum_j \frac{1}{n_{ij}}}{\sqrt{q(q-2)\sum \frac{1}{n_{ij}} + \sum_i\sum_j \frac{1}{n_{ij}}}\ \sqrt{q(q-2)\sum_i \frac{1}{n_{ij'}} + \sum_i\sum_j \frac{1}{n_{ij}}}}$$

다중 t분포를 푸는데 시간이 너무 많이 걸리므로 다음과 같은 Sidak(1967)의 부등식을 이용하여 기각값을 구한다.

$$\prod_{j=1}^{q} P(A_j) \leq P(\bigcap_{j=1}^{q} A_j). \tag{3. 5}$$

위의 식에서 좌변을 $1-\alpha$라 하고 해를 구하면

$$\{P(\ |T_j| < C(\alpha)\ \}^q = 1-\alpha$$

$$\leftrightarrow \int_{C_{\alpha^*/2}}^{\infty} P(t)dt = \frac{\alpha^*}{2} \qquad ;\ \alpha^* = 1-(1-\alpha)^{\frac{1}{q}}$$

여기서 $P(t)$는 t의 분포 밀도함수이고 $C(\alpha)=C_1=C_2=\cdots=C_q$이다.

이상의 관계를 이용하여

$$\max_{1\leq j\leq q}\ |T_j| > C_{\alpha^*/2}$$

이면 교호작용에 관한 가설 H'을 기각한다. 또는 X_j중 어느 하나에 대하여

$$X_j \in (X. \pm C_{\alpha^*/2} s\sqrt{\delta_j})$$

이면 H′을 기각한다. 여기서 사용되는 $C_{\alpha^*/2}$는 Games(1977)가 작성한 표를 이용한다.

3. 3 요인의 수준이 모두 2보다 큰 경우(p>2, q>2)

p>2, q>2인 경우에 두 교호작용 효과에 대한 상관계수는 앞절의 결과와 같은 완전한 상관관계를 갖고 있지 않으며 상관계수 구조를 주요인 효과로 얻을 수 없다. 본 논문은 Nelson (1988)이 균형 관측치에 대하여 구한 것과 마찬가지 방법으로 불균형 관측치에 대하여 적용하기로 한다. 이 방법은 B요인을 A요인의 2개의 수준과 비교하여 교호작용 효과를 분석하는 것이다. 이때 AB의 교호작용 항이 p(p−2)2개가 나오므로 $2\alpha/p(p-2)$의 유효 수준을 이용하여 기각값을 구하게 된다. 고정된 i, i′(i≠i′)에 대하여

$$X(i, i'; j) = Z_{ij} - Z_{i'j};$$

$$X. = \frac{1}{q}\sum_{j=1}^{q} X(i, i'; j);$$

$$\delta(i, i'; j) = \sqrt{q(q-2)\left(\frac{1}{n_{ij}} + \frac{1}{n_{i'j}}\right) + \sum_j \left(\frac{1}{n_{ij}} + \frac{1}{n_{i'j}}\right)\Big\} / q^2};$$

$$T(i, i'; j) = \{X(i, i'; j) - X.\}/S\sqrt{\delta(i, i'; j)} \qquad (3.\ 6)$$

라 놓으면 3. 2절에서와 마찬가지로 다음을 만족시키는 기각값을 찾을 수 있다.

$$g(\alpha; (p, q), v) = C(\{2\alpha/p(p-2)\}^*; q, v),$$

$$\int_g^\infty p(t)dt = \{1-(1-2\alpha/p(p-1))\}^{\frac{1}{q}}/2$$

여기서 P(t)는 t-분포이다. 그러므로 $\max_{i, i' j} | T(i, i'; j) \geq g(\alpha; (p, q), v)$이면 H′를 기각하며 또 X(i, i′; j)중 어느 하나가 다음의 구간밖에 있으면 같은 결론을 내릴 수 있다.

$$(X. - gs\sqrt{\delta}, \quad X. + gs\sqrt{\delta})$$

g의 값은 부록에 나타나 있다. g의 값을 구하기 위해 IMSL의 t분포 밀도함수 역함수를 이용하였으며 계산의 정밀도는 Double precision을 사용하였다. 부록의 표 2에는 α=0.01과 0.05에 대하여 p≦q인 경우에만 기각값 g가 나타나 있다. 이와 같은 결과를 그래프를 이용한 예제를 다음절에서 설명한다.

4. 그래프를 이용한 예제

여기서는 불균형 관측치 경우의 평균분석 방법을 이용한 교호작용 검정을 수행하는 예를 들기로 한다. 표 3. 1의 데이터는 3≦p≦q≦5인 경우에 kshirsager(1983)가 보여준 데이타이다.

표 3. 1 Kshirsagar(1983)의 데이타

Smoking history(A)		Stress test(B) 1 Bicycle	2 Treamill	3 Step
1	None	12.8 13.5 11.2	16.2 17.8	22.6 19.3 18.9
2	Moderate	10.9	15.5 13.8 16.2	21.0 15.9
3	Heavy	9.2 7.5	13.2 8.1	16.2 16.1 17.8

$$SSE = \sum_i \sum_j \sum_k y_{ijk} - \sum_i \sum_j n_{ij} x^2_{ij}$$

$$s^2 = 1.928$$

그리고 다음과 같은 값을 얻는다.

$$T(1,\ 2,\ 1) = -0.088 \quad T(1,\ 2,\ 2) = 0.055 \quad T(1,\ 2,\ 3) = 0.044$$
$$T(1,\ 3,\ 1) = -0.375 \quad T(1,\ 3,\ 2) = 1.102 \quad T(1,\ 3,\ 3) = 0.823$$
$$T(2,\ 3,\ 1) = 0.218 \quad T(2,\ 3,\ 2) = 1.031 \quad T(2,\ 3,\ 3) = 0.777$$

또한 모든 i, i', j에 대하여 $\mathrm{Max}\,|\,T(i,\ i',\ j)\,| = 1.102$이다.
부록에서 주어진 α에 대해 다음과 같은 $g(\alpha\ ;\ (3,\ 3)\ ;\ 12)$의 값을 얻는다.

$$g(0.05\ ;\ (3,\ 3)\ ;\ 12) = 3.3684,$$
$$g(0.01\ ;\ (3,\ 3)\ ;\ 12) = 4.2575$$

이 값들은 1.102보다 크므로 stress test와 smoking history사이의 교호효과가 유의하지 않음을 알 수 있다. 같은 결과가 Kshirsaga(1983, pp 290-307)의 방법으로 얻어지며, 또한 3. 3절 $X(i,\ i'\ ;j)-X.$ 들이 그림 3. 2와 같으므로 같은 결론을 내릴 수 있다. 그림 3. 2에서 가로축은 $X(i,\ i'\ ;j)-X.$이고 가로축은 식 (3. 6)에서 $(i,\ i'\ ;j)$들의 조합으로, 그 순서는

$$(1,\ 2\ ;\ 1),\ (1,\ 2\ ;\ 2),\ (1,\ 2\ ;\ 3),\ \cdots,\ (2,\ 3\ ;\ 3)$$

이다. 그리고 계단으로 나타난 선분들은 결정선으로

$$g(0.05\ ;\ \ (p,\ q)\ ;\ \ n)s\sqrt{\delta(i,\ i'\ ;\ j)}$$

이다. 그래프에서 보듯이 모든 X(i, i' ; j)-X. 들이 결정선안에 있으므로 교호효과가 유의하지 않음을 결론지을 수 있다.

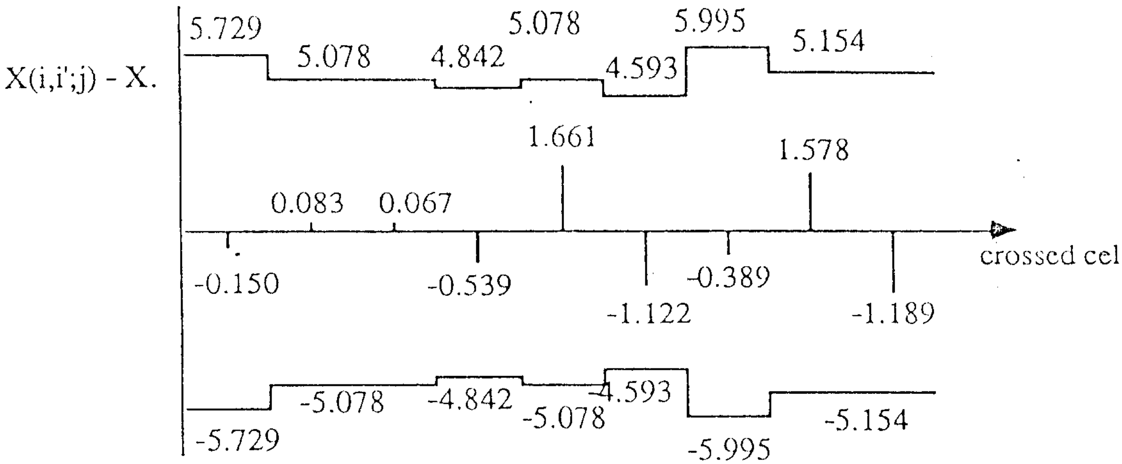

그림 3. 2 그래프를 이용한 평균 분석

참 고 문 헌

1) Games, P.A.(1977), "An Improved t Table for Simultaneous Control on g Contrasts," Journal of the American Statistical Association, 72, pp 531-534.

2) Graybill, F. A.(1976), Theory and Application of the Linear Model, North Scituate : Duxbury.

3) Halperin, M., Greenhouse, S. W., Cornfield, J., and Zalokar, J.(1955), "Tables of Percentage Points for the Studentized maximum Absolute Deviate in Normal Samples," Journal of the American Statistical Association, 50, pp 185-195.

4) Hermmerle, W. J.(1979) "Balanced Hypotheses and Unbalanced Data," Journal of the American Statistical Association, 74, pp 794-798.

5) Kshirsagar, A. M.(1983), A Course in Linear Models, New York : Dekker.

6) Nelson, P. R.(1983), "A comparison of sample sizes for the Analysis of Means and the Analysis of Variance," Jorunal of Quality Technology, 15, pp 33-39.

7) ________ (1988), "Testing for Interactions using the Analysis of Means," Technometrics, 30, pp 53-61.

8) Ott, E. R.(1967), "Analysis of Means-A Graphical procedure," Industrial Quality Control, 24, pp 101-109.

9) Schilling, E. G.(1973), "A Systematic Approach to the Analysis of Means," Journal of Quality Technology, 5, pp 92-108, pp 147-159.

10) Sidak, Z. (1967), "Rectanglar Confidence Regions for the Means of Multivariate Normal Distributions," Journal of the American Statistical Association, 62, pp 626-633.

11) Speed, F. M., Hocking, R. P., and Hackney, O. P., (1978) "Methods of Analysis of Linear Models with Unbalanced Data," Journal of the American Statistical Association, 73, pp 105-112.

부록 1 : 식 (3. 1)에 대한 증명

정리 : n개의 요인을 갖고 있는 실험계획에서 교호작용항을 $\hat{\gamma}_{i_1 \cdots i_n}^{\bullet F_1 \cdots F_n}$이라고 가정할 때, i_1은 요인 F_1의 수준이며, i_n을 요인 F_n의 수준이라고 할 때, $k=1, 2, \cdots, n$에 대하여 $i_k=1, 2, \cdots, c_k$가 되면 c_k는 요인 F_k의 수준 갯수가 된다. 그리고 $n_{i_1 \cdots i_n}$은 각 cell($i_1, i_2, \cdots, i_n$)의 관측치의 갯수이며, $y_{i_1, \cdots, i_n}$ 는 관측치이며, 동일의 분산 σ^2을 가질 때,

$$
\begin{aligned}
&VAR(\hat{\gamma}_{i_1 \cdots i_n}^{\bullet F_1 \cdots F_n}) \\
&= \prod_{k=1}^{n} \frac{\sigma^2}{c_k^2} \Bigg[\prod_{k=1}^{n} \frac{(c_k-1)}{n_{i_1, i_2, \cdots, i_n}} + \sum_{\alpha=1}^{n} \prod_{k \neq \alpha}^{n} (c_k-1)^2 \sum_{i'_\alpha = i_\alpha}^{c_\alpha} \frac{1}{n_{i_1, \cdots, i'_\alpha, \cdots, i_n}} \\
&\quad + \sum_{\alpha=1}^{n} \sum_{(\beta=1, \alpha \neq \beta)}^{n} \prod_{(k \neq \alpha, k \neq \beta)}^{n} (c_k-1)^2 \sum_{(i'_{\alpha} \neq i_{\alpha}, i'_{\beta} \neq i_{\beta})} \frac{1}{n_{i_1, \cdots, i'_\alpha, \cdots, i'_\beta, \cdots, i_n}} + \cdots + G_h + \cdots \\
&\quad + \sum_{i'_1 \neq i_1} \sum_{i'_2 \neq i_2} \cdots \sum_{i'_n \neq i_n} \frac{1}{n_{i'_1, \cdots, i'_n}} \Bigg]
\end{aligned}
$$

여기서,

$$
G_h = \sum_{\alpha_1=1}^{n} \sum_{\alpha_2=1}^{n} \cdots \sum_{\alpha_h=1}^{n} \prod_{k=1}^{n} (c_k-1)^2 \sum_{i'_{\alpha_1} \neq i_{\alpha_1}} \sum_{i'_{\alpha_2} \neq i_{\alpha_2}} \cdots \sum_{i'_{\alpha_h} \neq i_{\alpha_h}} \frac{1}{n_{i_1, \cdots, i'_{\alpha_1}, \cdots, i'_{\alpha_h}, \cdots, i_n}} \text{ 이며,}
$$

$\alpha_1, \alpha_2, \cdots, \alpha_h$는 서로 다른 것이며, k도 $\alpha_1, \alpha_2, \cdots, \alpha_h$와 다르며, $1<h<n$이다.

[증명] 수학적 귀납법을 사용하여 증명한다. 먼저 $n=2$인 경우,

$$
\hat{\gamma}_{i_1 \cdot i_2}^{\bullet F_1 \cdot F_2} = \hat{\gamma}_{i_1 i_2}^{\bullet F_1} - \hat{\gamma}_{i_1 \cdot}^{\bullet F_1}
$$

$$\mathrm{Var}(\hat{\gamma}_{i_1 \cdot i_2}^{\bullet F_1 \cdot F_2}) = \mathrm{Var}(\hat{\gamma}_{i_1 \cdot i_2}^{\bullet F_1} - \frac{1}{c_2}\sum_{c_2=1}^{l_2}\hat{\gamma}_{i_1 \cdot i_2}^{\bullet F_1})$$

$$= \frac{(c_2-1)^2}{c^2_2}\mathrm{Var}(\hat{\gamma}_{i_1 \cdot i_2}^{\bullet F_1}) + \frac{1}{l_2^2}\sum_{i'_2 \neq i_2}\mathrm{Var}(\hat{\gamma}_{i_1 \cdot i'_2}^{\bullet F_1})$$

$$= \frac{1}{c_1^2 c_2^2}\Big\{(c_2-1)^2(c_1-1)^2\frac{1}{n_{i_1 \cdot i_2}} + (c_2-_1)^2\sum_{i_1 \neq i_1}\frac{1}{n_{i'_1 \cdot i_1}} + (c_1-1)^2\sum_{i'_2 \neq i_2}\frac{1}{n_{i_1 \cdot i_2}}$$

$$+ \sum_{i'_1 \neq i_1}\sum_{i'_2 \neq i_2}\frac{1}{n_{i'_1 \cdot i'_2}}\Big\}\sigma^2$$

따라서 n=2인 경우에 식이 성립한다. 이 식이 n-요인 모형에서 성립한다고 가정하자. 그러면,

$$\mathrm{Var}(\hat{\gamma}_{i_1, \cdots, i_{n+1}}^{\bullet F_1, \cdots, F_{n+1}}) = \mathrm{Var}(\gamma_{i_1, \cdots, i_{n+1}}^{\bullet F_1, \cdots, F_n} - \hat{\gamma}_{i_1, \cdots, i_n, \cdot}^{\bullet F_1, \cdots, F_n})$$

$$= \frac{(c_{n+1}-1)^2}{c^2_{n+1}}\mathrm{Var}(\hat{\gamma}_{i_1, \cdots, i_{n+1}}^{\bullet F_1, \cdots, F_n}) + \frac{1}{c^2_{n+1}}\sum_{i'_{n+1} \neq i_{n+1}}\mathrm{Var}(\hat{\gamma}_{i_1, \cdots, i_n, i'_{n+1}}^{\bullet F_1, \cdots, F_n})$$

$$= \prod_{k=1}^{n+1}\left(\frac{1}{c_k}\right)^2\Big[\prod_{k=1}^{n=1}(c_k-1)^2\frac{1}{n_{i_1, \cdots, i_{n+1}}} + \sum_{a=1}^{n}\prod_{k \neq \alpha}^{n+1}(l_k-1)^2\sum_{i'_\alpha = i_\alpha}\frac{1}{n_{i_1, \cdots, i_\alpha, \cdots, i_{n+1}}}$$

$$+ \sum_{\alpha \neq \beta}^{n+1}\prod_{k \neq \alpha}^{n+1}(c_k-1)^2\sum_{i'_\alpha \neq i_\alpha}\frac{1}{n_{i_1, \cdots, i_\alpha, \cdots, i_\beta, \cdots, i_n}} + \cdots + M_1 + \cdots$$

$$+ \prod_{k=1}^{n+1}(c_k-1)^2\sum_{i'_1 \neq i_1}\cdots\sum_{i'_n \neq i_n}\frac{1}{n_{i_1, \cdots, i_\alpha, \cdots, i_\beta, \cdots, i_n}}\Big]$$

$$+ \prod_{k=1}^{n+1}\left(\frac{1}{c_k}\right)^2\Big[\prod_{k=1}^{n}(c_k-1)^2\sum_{i'_{n=1} \neq i_{n+1}}\frac{1}{n_{i_1, \cdots, i_{n+1}}}$$

$$+ \sum_{\alpha=1}^{n}\prod_{k \neq \alpha}^{n}(c_k-1)^2\sum_{i'_{n+1} \neq i_{n+1}}\sum_{i'_\alpha \neq i_\alpha}\frac{1}{n_{i_1, \cdots, i'_\alpha, \cdots, i'_{n+1}}}$$

$$+ \sum_{\alpha \neq \beta}^{n}\sum_{k \neq \alpha}^{n}(c_k-1)^2\sum_{i'_\alpha \neq i_\alpha}\sum_{i'_\beta \neq i_\beta}\sum_{i'_{n+1} \neq i_{n+1}}\frac{1}{n_{i_1, \cdots, i'_\alpha, \cdots, i'_\beta, \cdots, i'_n}} + \cdots + M_2 + \cdots$$

$$+ \sum_{i'_1 \neq i_1}\sum_{i'_2 \neq i_2}\cdots\sum_{i'_{n+1} \neq i_{n+1}}\frac{1}{n_{i'_1, \cdots, i'_{n+1}}}\Big]$$

여기서 서로 다른 $\alpha_1, \cdots, \alpha_h$들과 $\alpha_1, \cdots, \alpha_h$이 아닌 k에 대하여,

$$M_1 = \sum_{\alpha_1=1}^{n}\cdots\sum_{\alpha_h=1}^{n}\prod_{k=1}^{n+1}(c_k-1)^2\sum_{i'_{\alpha_1} \neq i_{\alpha_1}}\cdots\sum_{i'_{\alpha_h} \neq i_{\alpha_h}}\frac{1}{n_{i_1, \cdots, i'_{\alpha_1}, \cdots, i'_{\alpha_h}, \cdots, i_{n+1}}}$$

$$M_2 = \sum_{\alpha_1=1}^{n}\cdots\sum_{\alpha_h=1}^{n}\prod_{k=1}^{n}(c_k-1)^2\sum_{i'_{\alpha_1} \neq i_{\alpha_1}}\cdots\sum_{i'_{\alpha_h} \neq i_{\alpha_h}}\sum_{i'_{n+1} \neq i_{n+1}}\frac{1}{n_{i_1, \cdots, i'_{\alpha_1}, \cdots, i'_{\alpha_h}, \cdots, i'_{n+1}}}$$

그러면 다음과 같은 식이 계산되므로 증명이 이루어진다.

$$\prod_{k=1}^{n+1}\left(\frac{1}{c_k}\right)^2\left[\prod_{k=1}^{n+1}(c_k-1)^2\frac{1}{n_{i_1,\ldots,i_{n+1}}}+\sum_{\alpha=1}^{n+1}\prod_{k\neq\alpha}^{n+1}(c_k-1)^2\sum_{i'_\alpha=i_\alpha}\frac{1}{n_{i_1,\ldots,i_\alpha\ldots,i_{n+1}}}\right.$$

$$+\cdots+\sum_{\alpha_1=1}^{n+1}\cdots\sum_{\alpha_h=1}^{n+1}\prod_{k=1}^{n+1}(C_k-1)^2\sum_{i'_{\alpha_1}\neq i_{\alpha_1}}\cdots\sum_{i'_{\alpha_h}\neq i_{\alpha_h}}\frac{1}{n_{i_1,\ldots,i_{\alpha_1},\ldots,i_{\alpha_h'}\ldots,i_{n+1}}}+\cdots$$

$$\left.+\sum_{i_1\neq i_1}\sum_{i_2\neq i_2}\cdots\sum_{i'_{n+1}\neq i_{n+1}}\frac{1}{n_{i'_1,\ldots,i'_{n+1}}}\right]$$

부록 2. 1 g(0.01 ; (p, q) ; ν)의 기각값

ν \ (p, q)	(3. 3)	(3. 4)	(3. 5)	(4. 4)	(4. 5)	(5. 5)
9	3.6183					
10	3.5148					
11	3.4337					
12	3.3684	3.5239				
13	3.3148	3.4641				
14	3.2699	3.4142				
15	3.2319	3.3719	3.4804			
16	3.1992	3.3356	3.4411	3.6653		
17	3.1709	3.3041	3.4071	3.6255		
18	3.1460	3.2765	3.3773	3.5907		
19	3.1241	3.2522	3.3511	3.5601		
20	3.1046	3.2306	3.3277	3.5328	3.6289	
21	3.0871	3.2112	3.3068	3.5085	3.6028	
22	3.0713	3.1938	3.2881	3.3867	3.5794	
23	3.0571	3.1780	3.2711	3.4669	3.5583	
24	3.0441	3.1637	3.2556	3.4490	3.5391	
25	3.0323	3.1506	3.2415	3.4326	3.5216	3.7244
26	3.0214	3.1386	3.2286	3.4176	3.5056	3.7058
27	3.0114	3.1276	3.2167	3.4038	3.4909	3.6888
28	3.0022	3.1174	3.2058	3.3911	3.4773	3.6731
29	2.9937	3.1080	3.1956	3.3794	3.4647	3.6587
30	2.9857	3.0992	3.1862	3.3684	3.4531	3.6452
35	2.9533	3.0634	3.1477	3.3289	3.4055	3.5804
40	2.9293	3.0370	3.1193	3.2911	3.3706	3.5503
45	2.9109	3.0618	3.0976	3.2661	3.3439	3.5196
50	2.8964	3.0008	3.0804	3.2462	3.3228	3.4953
60	2.8748	2.9770	3.0549	3.2169	3.2169	3.4596
80	2.8482	2.9478	3.0236	3.1809	3.2532	3.4157
100	2.8325	2.9305	3.0051	3.1597	3.2306	3.3899

부록 2. 2 g(0.05 ; (p, q) ; ν)의 기각값

ν \ (p, q)	(3. 3)	(3. 4)	(3. 5)	(4. 4)	(4. 5)	(5. 5)
9	4.7050					
10	4.5177					
11	4.3728					
12	4.2575	4.4208				
13	4.1637	4.3184				
14	4.0859	4.2335				
15	4.0203	4.1621	4.2727			
16	3.9644	4.1012	4.2078	4.4338		
17	3.9160	4.0487	4.1518	4.3703		
18	3.8739	4.0029	4.1031	4.3151		
19	3.8368	3.9627	4.0603	4.2666		
20	3.8039	3.9270	4.0224	4.2237	4.3192	
21	3.7746	3.8952	3.9887	4.1856	4.2788	
22	3.7244	3.8409	3.9310	4.1205	4.2101	
23	3.7244	3.8409	3.9310	4.1205	4.2101	
24	3.7029	3.8176	3.9062	4.0925	4.1806	
25	3.6832	3.7963	3.8837	4.0671	4.1537	4.3515
26	3.6652	3.7768	3.8630	4.0439	4.1292	4.3239
27	3.6487	3.7590	3.8441	4.0226	4.1067	4.2985
28	3.6334	3.7425	3.8266	4.0029	4.0860	4.2753
29	3.6193	3.7273	3.8105	3.9848	4.0668	4.2537
30	3.6063	3.7132	3.7956	3.9680	4.0491	4.2338
35	3.5530	3.6557	3.7347	3.8997	3.9771	4.1530
40	3.5139	3.6136	3.6901	3.8497	3.9245	4.0941
45	3.4841	3.5814	3.6561	3.8116	3.8844	4.0492
50	3.4605	3.5560	3.6293	3.7816	3.8528	4.0139
60	3.4257	3.5158	3.5896	3.7374	3.8063	3.9020
80	3.3830	3.4726	3.5412	3.6833	3.7495	3.8986
100	3.3578	3.4456	3.5127	3.6515	3.7161	3.8615

Balanced Hypothesis Testing for Interactions of Unbalanced Case Using Analysis of Means

Kim, Byung Chun* and Jung, Kang Mo**

Abstract

An analysis of mean procedure is presented for testing the significance of two factor interactions for unbalanced case. When at least one factor if at only two levels, the technique is the same as that of Ott(1967) and Nelson(1988) except that we have used the critical value by using the Sidak's multiplicative inequality. The technique can be extended to the case in which both factors more than two levels. Tables of the critical values $g(\alpha\,;\,(p,\ q)\,;\,\nu)$ are given for $\alpha=0.05$ and 0.01 ; (p, q) combinations satisfying $3\leqq p\leqq q\leqq 5$; and various degrees of freedom ν. We present and prove the formula of the variance of the interactions for the n-way model.

* KAIST, P.O. Box 150, cheongryang, Seoul.

** Samsung Advanced Institute of Technology, Nongseo-ri, kihung-eup, Yougin-gun, Kyungki-do.