응용통계연구 제 5권 2호
1992년 pp 243-254

불균형 자료 분석과 가설 검정에 관한 연구[1)]

장석환[2)], 송규문[2)], 김장한[2)]

1. 서 론

불균형 자료 분석에 대해서는 일찌기 Brown(1932)과 Yates(1934)의 연구 이 Finney(1948), Stevens(1948), Henderson(1953), Kramer(1955)등 많은 사람들이 관 을 가지고 연구하였고 Searle(1971, 1977, 1981)은 R( )표기법으로 모형식의 상수적합에 의한 변동을 나타내었으며 Hocking과 Speed(1975), Speed와 Hocking (1976)이 사용한 제한들을 Σ-, W-, O-restrictions라고 하였다. 또한 Speed등(1978)은 비가중평균법(method of unweighted means), 평균의 가중제곱법(method of weighted squares of means), 상수적합법 (method of fitting constants), Overall- Spiegel법, Henderson 방법 등을 비교설명하고 Burdick 등(1974)은 각 변동을 기하학적으로 해석하려 하였다. 백(1987a,1987b)은 SAS팩키지에 의한 변동을 설명하였고 장(1988)도 Searle (1977, 1981)의 방법을 이용하여 가설검정과 변동을 검토한바 있다.

본 연구에서는 여러가지 모형에 대하여 $n_{ij} > 0$, 또는 $n_{ij} \geq 0$ 인 경우에 변동계산과 W-, Σ-, O-제한 조건하에서의 변동과 가설를 재조명해 보고저 한다.

2. 자료 및 모형

두인자 A 와 B의 수준수를 각각 a, b라하고 이들로 만들어지는 각 처리조합의 자료수를 n_{ij} 라고 하면 일반적인 선형모형식은

$$y_{ijk} = \mu_{ij} + e_{ijk}, \quad i = 1,2,\cdots,a \quad j = 1,2,\cdots,b \quad k = 1,2,\cdots,n_{ij} \tag{1}$$

로 나타낼수 있으며 y_{ijk} 는 (i,j)조합의 k번째 관측치, μ_{ij}는 모평균, 그리고 e_{ijk} 는 오차항으로서 $N(0, \sigma^2)$인 확률변수라고 가정한다. 이원분류자료를 분석할때 일반적으로 식(1)의 모수를 재구성하므로서 다음의 초과모수모형을 생각하게된다.

$$y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \tag{2}$$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \tag{3}$$

---

1) 이 논문은 1990년도 문교부지원 한국학술진흥재단의 대학 부설 연구소지원 학술 연구 조성비에 의하여 연구되었슴.

2) (700-701) 대구직할시 달서구 신당동 1000, 계명대학교 통계학과.

여기서 μ 는 전체평균, α_i 는 인자 A의 주효과, β_j 는 인자 B의 주효과, 그리고 γ_{ij} 는 교호작용을 나타낸다. 특히 식(3)을 행렬로

$$y = Xb + e \tag{4}$$

와 같이 나타낼 수 있고 정규방정식

$$X'Xb^o = X'y \tag{5}$$

에서 $X'X$ 는 $n_{ij} > 0$ 인 경우에 다음과 같다.

$$X'X = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} \tag{6}$$

단

$$X_1'X_1 = \left| \begin{array}{ccccccccc} n & n_{1.} & n_{2.} & \cdots & n_{a.} & n_{.1} & n_{.2} & \cdots & n_{.b} \\ n_{1.} & n_{1.} & 0 & \cdots & 0 & n_{11} & n_{12} & \cdots & n_{1b} \\ n_{2.} & 0 & n_{2.} & \cdots & 0 & n_{21} & n_{22} & \cdots & n_{2b} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ n_{a.} & 0 & 0 & & n_{a.} & n_{a1} & n_{a2} & \cdots & n_{ab} \\ n_{.1} & n_{11} & n_{12} & \cdots & n_{a1} & n_{.1} & 0 & \cdots & 0 \\ n_{.2} & n_{12} & n_{22} & \cdots & n_{a2} & 0 & n_{.2} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ n_{.b} & n_{1b} & n_{2b} & \cdots & n_{ab} & 0 & 0 & \cdots & n_{.b} \end{array} \right|$$

$$X_1'X_2 = \left| \begin{array}{cccccccccccc} n_{11} & n_{12} & \cdots & n_{1b} & n_{21} & \cdots & n_{2b} & \cdots & n_{a1} & n_{a2} & \cdots & n_{ab} \\ n_{11} & n_{12} & \cdots & n_{1b} & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & n_{21} & \cdots & n_{2b} & \cdots & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & n_{a1} & n_{a2} & \cdots & n_{ab} \\ n_{11} & 0 & \cdots & 0 & n_{21} & \cdots & 0 & \cdots & n_{a1} & 0 & \cdots & 0 \\ 0 & n12 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & n_{a2} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & n_{1b} & 0 & \cdots & n_{2b} & \cdots & 0 & 0 & \cdots & n_{ab} \end{array} \right|$$

$$X_2'X_2 = diag\ \{ n_{11}, n_{12}, \cdots, n_{1b}, n_{21}, \cdots, n_{2b}, \cdots, \cdots, n_{a1}, \cdots, n_{ab} \}$$

본 연구에서는 $n_{ij} > 0$ 인 경우 Searle(1981)의 자료와 $n_{ij} \geq 0$ 인 경우 표 2 의 자료를 이용하기로 한다.

표 2. $n_{ij} \geq 0$ 인 경우

	A1	A2	A3	합계$(y_{.j})$
B1	8	9, 13		$30(n_{.1}=3)$
B2	10, 12		14	$36(n_{.2}=3)$
B3		16, 14		$30(n_{.3}=3)$
B4	6		22	$28(n_{.4}=2)$
합계 $(y_{i.})$	36 $(n_{1.}=4)$	52 $(n_{2.}=4)$	36 $(n_{3.}=2)$	124 $(n=10)$

표 3. 총변동의 분할

S.V.	D.F.	S.S.
$R(\mu)$	1	1537.6
$R(\alpha\|\mu)$	2	110.4
$R(\beta\|\mu\ \alpha)$	3	17.7143
$R(\beta\|\mu)$	3	36.4
$R(\alpha\|\mu\ \beta)$	2	91.7143
$R(\gamma\|\mu\ \alpha\ \beta)$	1	48.2857
Error	3	12
Total	10	1726

3. 모수적합과 변동

3.1 분할행렬에 의한 방법

식(1)에서 $\hat{\mu}_{ij}$ 의 불편추정량은 $\mu_{ij} = \bar{y}_{ij\cdot}$, i = 1,2,…,a, j = 1,2,…,b, 이고 따라서 $\hat{b}' = (\bar{y}_{ij.})'$ 로 놓으면 설명되는 변동은 식(7)과 같다.

$$R(b) = b'X_2'y = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij\cdot}^2/n_{ij} \tag{7}$$

교호작용이 없는 경우, 초과모수모형식(2)에서 인자 A 와 인자 B 의 수정되지 않은 변동은 잘 알려진바와 같이 다음과 같다. 즉

$$\mathrm{SSA} = \sum_{i=1}^{a} n_{i.}(\bar{y}_{i..} - \bar{y}_{...})^2\ ,\quad \mathrm{SSB} = \sum_{j=1}^{b} n_{.j}(\bar{y}_{.j.} - \bar{y}_{...})^2 \tag{8}$$

또 모형식 (2)에 대한 정규방정식은 식(6)의 $X_1'X_1$ 을 이용하면

$$X_1'X_1 b^o = X'y \tag{9}$$

와 같고 이의 해 중 한가지는

$$b^o = (X_1'X_1)^- X_1'y \tag{10}$$

이며 $(X_1'X_1)^-$ 는 $X_1'X_1$ 의 일반역행렬로서 $X_1(X_1'X_1)^-X_1'$ 이 불변이므로 어느 일반역행렬을 사용해도 설명되는 변동은 잘 알려진 바와같이 동일하다. X_1 에서 μ, α에 해당하는열을 Z_1, β에 해당하는 열을 Z_2 로 놓고, X_1 과 모수벡터 b^o 를 각각

$$X_1 = [Z_1\ \ Z_2]\ ,\quad b^{o\prime} = [\mu\ \ \alpha' \,\vdots\, \beta'] = [b_1^{o\prime}\ \ b_2^{o\prime}] \tag{11}$$

로 분할하면 식(9)는

$$y = X_1 b^o + e = Z_1 b_1^o + Z_2 b_2^o + e \tag{12}$$

로 쓸 수 있다. 따라서 인자 B만을 생각하면

$$y = Z_2 b_2^o + e \tag{13}$$

이므로 식(9) 또는 식(12)에 의하여 설명되는 변동은

$$R(b^o) = R(\mu,\ \alpha' \,\vdots\, \beta') = R(b_1^o,\ b_2^o) \tag{14}$$

이다. 식(13)에 의하여 설명되는 변동 $R(b_2^o)$ 에서 주의할것은 μ 에대한 1을 Z_2 에 포함시키고 b_2^o 에도 μ 를 포함시켜야 하지만 실제로 $R(b_2^o)$ 에는 영향을 주지않으므로 $R(b_2^o)$ = $R(\mu,\beta)$와 같다. 그러므로 인자 B를 적합한 후에 인자 A에 의하여 추가적으로 설명(β에 대하여 수정된)되는 변동은

$$R(b_1^o | b_2^o) = R(b_1^o,\ b_2^o) - R(b_2^o) = R(\alpha | \mu,\ \beta) \tag{15}$$

이다. X_1 = $[\ Z_1 \quad Z_2\ ]$을 이용하여 $X_1' X_1$ 에서 $W_1 = Z_1' Z_1 - Z_1' Z_2 (Z_2' Z_2)^{-1} Z_2' Z_1$으로 정의하면 W_1 을 이용한 $(X_1' X_1)^-$ 는

$$(X_1' X_1)^- = \begin{pmatrix} W_1^- & -W_1^- Z_1' Z_2 (Z_2' Z_2)^{-1} \\ -(Z_2' Z_2)^{-1} Z_2' Z_1 W_1^- & (Z_2' Z_2)^{-1} \{ I + Z_2' Z_1 W_1^- Z_1' Z_2 (Z_2' Z_2)^{-1} \} \end{pmatrix} \tag{16}$$

이며 식(16)을 이용한 식(10)의 해 는 다음과 같다.

$$b^o = \begin{pmatrix} b_1^o \\ b_2^o \end{pmatrix} = \begin{pmatrix} W_1^- Z_1' \{ I - Z_2 (Z_2' Z_2)^{-1} Z_2' \} y \\ (Z_2' Z_2)^{-1} Z_2' (y - Z_1 b_1^o) \end{pmatrix} \tag{17}$$

따라서 식(17)의 결과를 이용하면 식(15)는

$$R(b_1^o | b_2^o) = b_1^{o\prime} W_1 b_1^o \tag{18}$$

임을 알 수 있고, 또 $X_1' X_1$ 에서 $W_2 = Z_2' Z_2 - Z_2' Z_1 (Z_1' Z_1)^- Z_1' Z_2$ 로 정의하면 $(X_1' X_1)^-$ 를 다음과 같이 나타낼 수 있다. 즉

$$(X_1' X_1)^- = \begin{pmatrix} (Z_1' Z_1)^- + (Z_1' Z_1)^- Z_1' Z_2 W_2^- Z_2' Z_1 (Z_1' Z_1)^- & -(Z_1' Z_1)^- Z_2' Z_1 W_2' \\ -W_2^- Z_2' Z_1 (Z_1' Z_1)^- & W_2' \end{pmatrix}$$

$$b^o = \begin{pmatrix} b_1^o \\ b_2^o \end{pmatrix} = \begin{pmatrix} (Z_1' Z_1)^- Z_1' \{ y - Z_2' b_2^o \} \\ b_2^o \end{pmatrix}$$

로 구해지며 여기서 $b_2^o = W_2^- Z_2' [I - Z_1 (Z_1' Z_1)^- Z_1'] y$ 이다. 따라서

$$R(b_2^o|b_1^o)=b_2^{o\prime}W_2b_2^o=R(\beta|\mu,\alpha) \tag{19}$$

Searle등(1981)은 R(β| μ, α) 를 구하기 위하여 식(11)의 $b^{o\prime}$를 $b^{o\prime}=[\beta' \vdots \mu \ \alpha']=[b_2^{o\prime} \ \ b_1^{o\prime}]$ 로 모수의 순서를 바꾸고, X_1도 이와 상응하게 변형 분할하여 모수를 적합하고 변동을 구하였으나 W_2를 이용하면 모수의 순서를 바꾸지 않고도 R(β | μ, α)를 구할 수 있다.

모형식(3)의 적합을 위해서는 식(6)에서 가장 쉽게 구할 수 있는 $(X'X)^-$는

$$(X'X)^-=\begin{bmatrix}0 & 0\\ 0 & (X_2'X_2)^{-1}\end{bmatrix} \tag{20}$$

이므로 식(5)의 b^o 는

$$b^{o\prime}=[0' \ \ \hat{b}']=[0' \ \ \bar{y}_{11} \ \cdots \ \bar{y}_{1b} \ \cdots \ \bar{y}_{a1} \ \cdots \ \bar{y}_{ab}] \tag{21}$$

로 구해지며 변동은

$$R(b^o)=\sum\sum y_{ij}^2/n_{ij}=R(\mu,\alpha,\beta,\gamma) \tag{22}$$

로 식(7)의 결과와 같다. 따라서 식(14)와 식(22)에서 교호작용에 의하여 추가적으로 설명되는 변동은 잘 알려진바와 같이

$$R(\gamma|\mu,\alpha,\beta)=R(\mu,\alpha,\beta,\gamma)-R(\mu,\alpha,\beta) \tag{23}$$

로 구해진다. 표 2의 자료에서 식(8), 식(15), 식(19) 및 식(23) 에 의한 변동은 표3과 같다. 식(6)에서 $W_1^*=X_1'X_1-X_1'X_2(X_2'X_2)^{-1}X_2'X_1=0$ 이므로 이 결과를 식(17)의 W_1^- 과 대치하면 식(21)과 같은 결과를 얻는다.

3.2 제한 조건에 의한 방법

(1) Σ - 제한

식(5) 와 식(9)의 해를 구하기 위하여 모수들 간에

$$\sum_{i=1}^{a}\alpha_i=\sum_{j=1}^{b}\beta_j=0 .$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\gamma_{ij}=0 \ \ , \sum_{i=1}^{a}\gamma_{ij}=0, \ j=1,2,\cdots,b \tag{24}$$

$$\sum_{j=1}^{b}\gamma_{ij}=0, \quad i=1,2,\cdots,a$$

의 제한을 생각하게 되는데 모형식(2)와 (3)에 이들 제한조건을 적용할때 모형행렬을 각각 X^*_Σ , X_Σ 로 나타내면 X^*_Σ 는 X_Σ 에서 γ_{ij} 에 해당하는 열을 제거시킨행렬과 같다. 따라서 정규방정식은 각각 $(X^{*\prime}_\Sigma X^*_\Sigma)b^*_\Sigma = X^{*\prime}_\Sigma y$, $(X^\prime_\Sigma X_\Sigma)b_\Sigma = X^\prime_\Sigma y$ 이며 $X^{*\prime}_\Sigma X^*_\Sigma$ 와 $X^\prime_\Sigma X_\Sigma$ 모두 정칙행렬이므로 $\hat{b}^{*\prime}_\Sigma = [\ \hat{\mu}^\prime, \hat{\alpha}^\prime, \hat{\beta}^\prime\ ]$과 $\hat{b}^\prime_\Sigma = [\ \hat{\mu}, \hat{\alpha}^\prime, \hat{\beta}^\prime, \hat{\gamma}^\prime\ ]$ 을 추정할 수 있다. 이들 결과로 부터 모형식 (2) 와 (3)에 의하여 설명되는 변동 $R(\mu,\alpha,\beta)_\Sigma$, $R(\mu,\alpha,\beta,\gamma)_\Sigma$ 을 구하여 표 3의 변동을 구할 수 있다. 또 X_Σ 에서 α, β, γ 를 각각 제거한 후 적합할때 설명되는 y의 변동을 각각 $R(\mu,\underline{\alpha},\beta,\gamma)_\Sigma$,$R(\mu,\alpha,\underline{\beta},\gamma)_\Sigma$,$R(\mu,\alpha,\beta,\underline{\gamma})_\Sigma$라 하고 α, β, γ에 의하여 추가적으로 설명되는 변동은 각각 $R(\alpha|\mu,\beta,\gamma)_\Sigma$,$R(\beta|\mu,\alpha,\gamma)_\Sigma$,$R(\gamma|\mu,\alpha,\beta)_\Sigma$ 로 나타낼 때 빈칸이 없는 자료에서는 Searle(1977), Searle 등(1981)은 평균의 가중제곱분석(Yates, 1934)에서 인자 A의 변동(SSA_W)과 인자 B의 변동(SSB_W)사이에 다음의 관계가 성립됨을 밝혔다. 즉

$$R(\alpha|\mu,\beta,\gamma)_\Sigma = R(\mu,\alpha,\beta,\gamma)_\Sigma - R(\mu,\underline{\alpha},\beta,\gamma)_\Sigma = SSA_W$$
$$R(\beta|\mu,\alpha,\gamma)_\Sigma = R(\mu,\alpha,\beta,\gamma)_\Sigma - R(\mu,\alpha,\underline{\beta},\gamma)_\Sigma = SSB_W \tag{25}$$

한편 교호작용에 있어서는 변함이 없다. 즉

$$R(\gamma|\mu,\alpha,\beta)_\Sigma = R(\mu,\alpha,\beta,\gamma)_\Sigma - R(\mu,\alpha,\beta,\underline{\gamma})_\Sigma = R(\gamma|\mu,\alpha,\beta).$$

그러나 표 2 와 같이 빈칸이 있는 자료에서는 식(25)의 관계가 성립되지 않는다.

(2) W - 제한

W - 제한 조건은 식(2), 식(3)에서 각 처리조합의 표본 수(n_{ij}) 또는 이들의 주변합($n_{i.}$, $n_{.j}$)을 가중치로 하는 모수의 가중합이 0 인 조건이다. 즉

$$\sum_{i=1}^{a} n_{i.}\alpha_i = 0, \quad \sum_{j=1}^{b} n_{ij}(\alpha_i + \gamma_{ij}) = 0, \quad j = 1,2,\cdots,b$$
$$\sum_{j=1}^{b} n_{.j}\beta_j = 0, \quad \sum_{i=1}^{a} n_{ij}(\beta_j + \gamma_{ij}) = 0, \quad i = 1,2,\cdots,a \tag{26}$$

Σ - 제한에서와 같이 식(26)의 제한 조건하에서 식(2)의 모형행렬을 X^*_W, 모수벡터를 $b^{*\prime}_W = [\mu,\alpha_1,\beta_1,\beta_2]$ 로 놓으면 $\hat{b}^{*\prime}_W = (X^{*\prime}_W X^*_W)^{-1} X^{*\prime}_W y$ 로 추정되며 W-제한 조건하에서 설명되는 변동 $R(b^*_W)$는

$$R(b^*_W) = R(\mu,\alpha,\beta)_W = R(\mu,\alpha,\beta) \tag{27}$$

이다. 또한 X_W^*로 부터 차례로 α 와 β를 제거한 축소모형에서 설명되는 y의 변동을 각각 $R(\mu,\underline{\alpha},\beta)_W$ 와 $R(\mu,\alpha,\underline{\beta})_W$ 로 나타내면

$$R(\alpha|\mu,\beta)_W = R(\mu,\alpha,\beta)_W - R(\mu,\underline{\alpha},\beta)_W = R(\alpha|\mu,\beta),$$
$$R(\beta|\mu,\alpha)_W = R(\mu,\alpha,\beta)_W - R(\mu,\alpha,\underline{\beta})_W = R(\beta|\mu,\alpha).$$

와 같다.

또 식 (3)에 W-제한 조건을 적용할때 모형행렬을 X_W, 모수벡터를 $b_W' = [\mu,\alpha',\beta',\gamma']$ 라 하면 $\hat{b}_W = (X_W'X_W)^{-1}X_W'y$로 추정되고 설명되는 변동 $R(\mu,\alpha,\beta,\gamma)_W$ 는 $R(\mu,\alpha,\beta,\gamma)_\Sigma$ 와 같다.

모형식 (3)에서 W - 제한조건하에서는 Σ - 제한조건에서 와는 달리

$$R(\alpha|\mu,\beta,\gamma)_W = R(\mu,\alpha,\beta,\gamma)_W - R(\mu,\underline{\alpha},\beta,\gamma)_W = R(\alpha|\mu),$$
$$R(\beta|\mu,\alpha,\gamma)_W = R(\mu,\alpha,\beta,\gamma)_W - R(\mu,\alpha,\underline{\beta},\gamma)_W = R(\beta|\mu)$$

가 성립되나 교호작용에 의한 변동 $R(\gamma|\mu,\alpha,\beta)_W$ 는 $R(\gamma|\mu,\alpha,\beta)_W \neq R(\gamma|\mu,\alpha,\beta)_\Sigma$ 를 보이고 있다.

(3) 0-제한

Searle(1977)과 Searle 등(1981)이 말하는 0_{st} - 제한은

$$\alpha_s = \beta_t = \gamma_{sj} = \gamma_{it} = 0, \quad i=1,2,\cdots,a, \quad j=1,2,\cdots,b \tag{28}$$

로 나타내며 s,t는 각각 인자 A 와 B의 수준이다. 식(28)은 각 인자 또는 교호작용의 효과가 없다는 제한이라기 보다는 $X_1'X_1$ 또는 $X'X$ 에서 s행과 t열을 제거시키므로서 정칙행렬을 유도하는 체계적인 방법이라고 할 수 있다. 이들 정칙행렬의 역행렬을 이용하여 $R(\alpha|\mu,\beta)$, $R(\beta|\mu,\alpha)$, $R(\gamma|\mu,\alpha,\beta)$ 를 구할 수 있다.

4. 가설과 변동

일반선형모형의 모수벡터를 식(5)에서와 같이 b^o 로 나타내면 일반적으로 검정되는 가설은

$$H_0 \ : \ K'b^o = 0 \tag{29}$$

이고 이에 의한 변동은 다음과 같다.

$$Q(H_0) = (K'b^o)'[K'(X'X)^- K]^{-1}(K'b^o) \tag{30}$$

식(20)에서 보는바와 같이 $\hat{b}$ 에 이용되는 $(\mathbf{X}'\mathbf{X})^-$ 는 실제로 $(\mathrm{X}_2'\mathrm{X}_2)^{-1}$ 이고 또 $\mathbf{K}'$ 는 $\bar{y}_{ij}$, $i=1,2,\cdots,a$ $j=1,2,\cdots,b$ 의 독립적인 대비의 행렬이므로 식(30)에서 $Q(H_o) = (K'\ \hat{b})'[K'(X_2'X_2)^{-1}K]^{-1}(K'\ \hat{b}) = R(\alpha,\beta,\gamma|\mu)$ 임을 알 수 있다.

모형식(2)의 경우에 흥미로운 가설은 인자의 수준평균을 비교하는 것으로 $H_o : \mu_{i\cdot} = \mu_{i'\cdot}$ 또는 $H_o : \mu_{\cdot j} = \mu_{\cdot j'}$ 이고, 이는 실제로

$$H_o \ : \ \alpha_i + \frac{1}{n_{i\cdot}} \sum_{j=1}^{b} n_{ij}\beta_j \ = \ \alpha_{i'} \ + \frac{1}{n_{i'j}} \sum_{j=1}^{b} n_{i'j}\beta_j \tag{31}$$

과

$$H_o \ : \ \beta_j + \frac{1}{n_j} \sum_{i=1}^{a} n_{ij}\alpha_i = \beta_{j'} + \frac{1}{n_{j'}} \sum_{i=1}^{a} n_{ij'}\alpha_i \tag{32}$$

를 검정하며 이들에 대한 변동은 각각 SSA = $R(\alpha|\mu)$, SSB = $R(\beta|\mu)$ 이다. 또 흔히 $H_o : \alpha_i = \alpha_i'$ 과 $H_o : \beta_j = \beta_j'$ 을 검정하게 되는데 이때 Searle의 자료에서는 a = 2 이므로 $H_o : \alpha_1 = \alpha_2$ 의 가설을 검정하게 되고 표 2에서는 $H_o : \alpha_1 = \alpha_2$ 와 $H_o : \alpha_1 = \alpha_3$ 를 동시에 검정하게 된다. 따라서 $H_o : \alpha_i = \alpha_i'$ 과 $H_o : \beta_j = \beta_j'$ 의 변동은 식(30)에 의하여 $R(\alpha|\mu,\beta)$, $R(\beta|\mu,\alpha)$임은 잘알려진 사실이다. 모형식(3)에 있어서 인자 A의 수준평균은 $\mu_{i.} = \mu + \alpha_i + \frac{1}{n_{i.}} \sum_{j=1}^{b} n_{ij}(\beta_j + \gamma_{ij})$, i = 1,2,⋯,a 이므로

$H_o : \mu_{i\cdot} = \mu_{i'\cdot}$, $i \neq i'$, 는 실제로

$$H_o \ : \ \alpha_i + \frac{1}{n_{i\cdot}} \sum_{j=1}^{b} n_{ij}(\beta_j + \gamma_{ij}) = \alpha_{i'} + \frac{1}{n_{i'\cdot}} \sum_{j=1}^{b} n_{ij'}(\beta_j + \gamma_{ij'}) \tag{33}$$

이며 표 2에서 $\mathbf{K}'$는

$$K' = \left(\begin{array}{cccccccc:ccccccc} 0 & 1 & -1 & 0 & \frac{-1}{4} & \frac{1}{2} & \frac{-1}{2} & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \frac{-1}{2} & \frac{-1}{2} & 0 & 0 \\ 0 & 1 & 0 & -1 & \frac{1}{4} & 0 & 0 & \frac{-1}{4} & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & \frac{-1}{2} & \frac{-1}{2} \end{array} \right)$$
$$= [\ L' \ \vdots \ M'\ ]$$

로 구해진다. 따라서 식 (33)의 가설에 대한 변동은 식(30)에 의하여 표 3의 $R(\alpha|\mu)$ 임을 알 수 있고 $H_o : \beta_j = \beta_{j'}$ 는

$$H_o : \beta_j + \frac{1}{n_{\cdot j}} \sum_{i=1}^{a} n_{ij}(\alpha_i + \gamma_{ij}) = \beta_{j'} + \frac{1}{n_{\cdot j'}} \sum_{i=1}^{a} n_{ij'}(\alpha_i + \gamma_{ij'})$$

이므로 $R(\beta|\mu)$ 에 의하여 검정된다. 또 $K' = [\ L' \quad M'\ ]$에서 L'는 식(31)의 대비 행렬로서 식(33)의 $\{\alpha_i + \sum_j n_{ij}\beta_j/n_{i.}\}$ 부분에 해당되고 M'는 $\{\sum_j n_{ij}\gamma_{ij}/n_{i.}\}$ 부분에 해당되는 대비행렬이다. 따라서 식(31)의 변동은

$$Q(H_o) = (L'b^o)'[L'(X_1'X_1)^{-1}L]^{-1}(L'b^o)$$

이고 식(33)의 변동은 $K' = [\ L' \quad M'\ ]$을 이용하면 식(30)에서

$$(K'b_o)'[K'(X_1'X_1)^{-}K]^{-1}(K'b^o) = (M'\hat{b})'[M'(X_2'X_2)^{-}M]^{-1}(M'\hat{b})$$

이므로 $(L'b^o)'[L'(X_1'X_1)^{-}L]^{-1}(L'b^o) = (M'\hat{b})'[M'(X_2'X_2)^{-1}M]^{-1}(M'\hat{b})$임을 알 수 있다. 결국 표 2의 자료를 이용한 식(31) 과 식(33)의 가설은 $R(\alpha|\mu)$에 의하여 검정됨을 알 수 있다.

또 Searle(1971), Hocking-Speed(1975), Speed-Hocking-Hackney(1978)등 이 지적한 바와 같이 모형식 (3)에서 $R(\alpha|\mu,\beta)$ 와 $R(\beta|\mu,\alpha)$ 는 각각

$$H_o : \left(n_{i.} - \sum_{j=1}^{b} \frac{n_{ij}^2}{n_{.j}}\right)\alpha_i + \sum_{j=1}^{b}\left(n_{ij} - \frac{n_{ij}^2}{n_{.j}}\right)\gamma_{ij} = \sum_{i\neq i'}^{a}\left(\sum_{j=1}^{b}\frac{n_{ij}n_{i'j}}{n_{.j}}\right)\alpha_{i'} + \sum_{i\neq i'}^{a}\sum_{j=1}^{b}\frac{n_{ij}n_{i'j}}{n_{.j}}\gamma_{i'j} \tag{34}$$

와

$$H_o : \left(n_{.j} - \sum_{i=1}^{a} \frac{n_{ij}^2}{n_{i.}}\right)\beta_j + \sum_{i=1}^{a}\left(n_{ij} - \frac{n_{ij}^2}{n_{i.}}\right)\gamma_{ij} = \sum_{j'\neq j}^{a}\left(\sum_{i=1}^{a}\frac{n_{ij}n_{ij'}}{n_{i.}}\right)\beta_{j'} + \sum_{j'\neq j}^{b}\sum_{i=1}^{a}\frac{n_{ij}n_{ij'}}{n_{i.}}\gamma_{ij'} \tag{35}$$

을 검정한다는 것은 Searle의 자료 및 표 2 를 이용하여 보여줄 수 있다.

교호작용에 대한 가설은 일반적으로 $H_o: \gamma_{ij} = \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu = 0$ 로 표현되며 비록 $R(\gamma|\mu,\alpha,\beta) = R(\gamma|\mu,\alpha,\beta)_W \neq \sum(\gamma|\mu,\alpha,\beta)_\Sigma$ 이더라도 교호작용에 대한 가설은 W-제한하에서는 $R(\gamma|\mu,\alpha,\beta)_W$ 에 의하여 검정되며, Σ-제한하에서는 $R(\gamma|\mu,\alpha,\beta)_\Sigma$ 에 의하여 검정됨을 알 수 있다.

<참 고 문 헌>

[1] 백 운봉(1987a). "반복수가 같지않은 이원표의 분석," *응용통계 제 2권 1호*, 7 - 28, 고려대학교 통계연구소.

[2] ———.(1987b). *SAS 일반선형 모형분석*, 47 - 61, 자유아카데미.

[3] 장 석환(1988). "표본수가 다른 이원표에 있어서 가설검정," *계명대학교 수리과학논집 제 8집*, 87 - 103.

[4] Brown, B. (1932), "A Sampling Test of the Technique of Analyzing Variance in a 2×n Table With Disproportionate Frequencies," ***Proceedings of the Iowa Academy of Science***, 38, 205.

[5] Burdik, D. S., Herr, D. G., O'Fallon, W. M., and O'Neill, B. V.(1974), "Exact Methods in the Unbalanced, Two-Way Analysis of Variance – A Geometric View," ***Communications in Statistics***, 3(6), 581 -595.

[6] Finney, D. J. (1948), "Main Effects and Interactions." ***Journal of American Statistical Association***. Vol. 43, 566 - 571.

[7] Henderson, C. R.(1953), "Estimation of Variance and Covariance Component s." ***Biometrics***, Vol. 9, 226 - 252.

[8] Hocking, R. R., and Speed, F. M. (1975), "A Full Rank Analysis of Some Linear Model Problems." ***Journal of American Statistical Association***, Vol. 70, 706 - 712.

[9] Kramer, C. R. (1955), "On the Analysis of Variance of a Two-way Classification with Unequal Sub-class Numbers." ***Biometrics***, Vol. 11, 441 - 452.

[10] Searle, S. R. (1971), ***Linear Models***, New York, John Wiley.

[11] ————.(1977) "Illustrative Calculations of Sums of Squares in the 2-way Crossed Classification, Unbalanced Data, All Cell Filled." ***Paper No. BU 608-M in the Biometrics Unit Mimeo Series***, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14853.

[12] ————, Speed, F. M., and Henderson, H. V. (1981),"Some Computational and Model Equivalences in Analysis of Variance of Unequal-Subclass -Numbers Data." ***The American Statistician***, Vol. 35, No.1, 10 - 33.

[13] Speed, F. M. and Hocking, R. R. (1976), "The Use of the R( )-Notation with Unbalanced Data." ***The American Statistician***, Vol. 30, No.1 30 -33.

[14] ———— , ———— , and Hackney, O. P. (1978), "Methods of Analysis

of Linear Models with Unbalanced Data." *Journal of the American Statistical Association*, Vol. 73, 105 - 112.

[15] Stevens, W. L. (1948),"Statistical Analysis of a Non-orthogonal Trifactorial Experiment." *Biometrics*, Vol 35, 346 - 367.

[16] Yates, F. (1934), "The Analysis of Mutiple Classifications with Unequal Numbers in the Different Classes." *Journal of the American Statistical Association*, Vol. 29, 51 - 66.

Study on the Analysis of Disproportionate Data and Hypothesis Testing[1)]

Suk-Hwan Chang[2)], Gyu-Moon Song[2)], Jang-Han Kim[2)].

In the present study two sets of unbalanced two-way cross- classification data with and without empty cell(s) were used to evaluate empirically the various sums of squares in the analysis of variance table. Searle(1977) and Searle et.al.(1981) developed a method of computing $R(\alpha|\mu, \beta)$ and $R(\beta|\mu, \alpha)$ by the use of partitioned matrix of $X'X$ for the model of no interaction, interchanging the columns of X in order of α, μ, β and accordingly the elements in b. An alternative way of computing $R(\alpha|\mu, \beta)$, $R(\beta|\mu, \alpha)$ and $R(\gamma|\mu, \alpha, \beta)$ without interchanging the columns of X has been found by means of$(X'X)^-$ derived, using $W_2 = Z_2'Z_2 - Z_2'Z_1(Z_1'Z_1)^-Z_1'Z_2$.

It is true that $R(\alpha|\mu,\beta,\gamma)_\Sigma = SSA_W$ and $R(\beta|\mu,\alpha,\gamma)_\Sigma = SSB_W$ where SSA_W and SSB_W are sums of squares for the factors A and B in the weighted squares of means analysis and $R(\gamma|\mu,\alpha,\beta) = R(\gamma|\mu,\alpha,\beta)_\Sigma$ for the data without empty cell, but not for the data with empty cell(s). It is also noticed that for the data with empty cells under W - restrictions $R(\alpha|\mu,\beta,\gamma)_W = R(\mu,\alpha,\beta,\gamma)_W - R(\mu,\underline{\alpha},\beta,\gamma)_W = R(\alpha|\mu)$ and $R(\beta|\mu,\alpha,\gamma)_W = R(\mu,\alpha,\beta,\gamma)_W - R(\mu,\alpha,\underline{\beta},\gamma)_W = R(\beta|\mu)$ but $R(\gamma|\mu,\alpha,\beta)_W = R(\mu,\alpha,\beta,\gamma)_W - R(\mu,\alpha,\beta,\underline{\gamma})_W \neq R(\gamma|\mu,\alpha,\beta)$. The hypotheses $H_o : K'b = 0$ commonly tested were examined in the relation with the corresponding sums of squares for $R(\alpha|\mu)$, $R(\beta|\mu)$, $R(\alpha|\mu,\beta)$, $R(\beta|\mu,\alpha)$, $R(\alpha|\mu,\beta,\gamma)$, $R(\beta|\mu,\alpha,\gamma)$, and $R(\gamma|\mu,\alpha,\beta)$ under the restrictions.

---

1) Research was supported by the Korean Research Foundation for University Institutes 1990

2) Department of Statistics, Keimyung University, Shindangdong Dalsogu, Taegu, 704-701, Korea