The Korean Communications
in Statistics Vol. 5, No. 3, 1998
pp. 633-644

혼합설계의 교호작용에 대한 여러 검정법들과 절사평균을 이용하여 변형한 검정법들의 강인성 비교

김 현 철[1)]

요 약

혼합설계의 교호작용에 대한 F검정이 유효하려면 다표본 구형성(multisample sphericity) 가정과 다변량 정규분포 가정이 만족되어야 한다. F검정을 실시하기 위한 가정들이 위반된 조건하에서 혼합설계의 교호작용에 대한 검정법들의 1종오류가 비교되었다. 비교된 검정법들은 (1) F검정(F), (2)절사평균을 사용한 F검정(F_T), (3) $\tilde{\varepsilon}$-수정 F 검정($\tilde{\varepsilon}$), (4)절사평균을 사용한 $\tilde{\varepsilon}$-수정 F검정 ($\tilde{\varepsilon}_T$), (5)CIGA검정(CIGA), (6)절사평균을 사용한 CIGA검정(CIGA_T)이었다. 결과는 CIGA와 CIGA_T는 1종오류를 대체로 잘 관리하나, F검정들과 $\tilde{\varepsilon}$검정들은 일부 조건에서 아주 작은 1종오류나 아주 큰 1종오류를 갖는 것으로 나타났다.

1. 서론

수준이 J인 한 개의 표본간 인자(between-subjects factor)와 수준이 K인 한 개의 표본내 인자(within-subjects factor)가 있고, 표본간 인자 수준 j의 표본수가 $n_j(j=1,\cdot\cdot\cdot,J)$ 인 혼합설계(mixed design 또는 split plot design)는 다음과 같은 선형모형으로 표현된다.

$$X_{ijk}=\mu+\alpha_j+\beta_k+\pi_{i(j)}+\alpha\beta_{jk}+\beta\pi_{ki(j)}+\varepsilon_{ijk}$$

여기서 $i=1,\cdot\cdot\cdot,n_j$

$j=1,\cdot\cdot\cdot,J$

$k=1,\cdot\cdot\cdot,K$

X_{ijk}= j집단의 표본내 인자 k의 표본 i의 관측값

μ= 전체 평균

α_j= 집단 j의 효과

β_k= 반복작용 인자 k의 효과

$\pi_{i(j)}$= 집단 j의 표본 i의 효과

$\alpha\beta_{jk}$= 집단 j와 반복측정 인자 k의 교호작용

$\beta\pi_{ki(j)}$= 집단 j의 표본 i와 반복측정 인자 k의 교호작용

1) (110-745) 서울특별시 종로구 명륜동 3가 53번지 성균관대학교 사범대학 교육학과 조교수

ε_{ijk}= 잔차

혼합설계에서 교호작용이 없다는 가설에 대한 검정은 다음과 같은 통계량에 의해서 이루어지는데, 이 통계량은 검정을 위한 가정들이 만족되면 자유도가 $(J-1)(K-1)$과 $(N-J)(K-1)$인 F 분포를 하게 된다.

$$F=\frac{MS_{AB}}{MS_{B\cdot S/A}}=\frac{SS_{AB}/(J-1)(K-1)}{SS_{B\cdot S/A}/(N-J)(K-1)}$$

여기서 SS_{AB}와 MS_{AB}는 각각 교호작용의 제곱합과 제곱평균

$SS_{B\cdot S/A}$와 $MS_{B\cdot S/A}$는 각각 오차의 제곱합과 제곱평균

$N=n_1+n_2+\cdots+n_J$

이때 표본간 인자 수준 j의 관측값을 $(K\times 1)$차원을 갖고 서로 독립적인 벡터 $\boldsymbol{x}_j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ 라하자. 또 $\mathbf{1}_K$는 $(K\times 1)$차원의 단위벡터이고, A는 $\boldsymbol{A}'\mathbf{1}_K=0$ 이고 $\boldsymbol{A}'\boldsymbol{A}=\boldsymbol{I}_{K-1}$ 인 $[K\times(K-1)]$ 차원의 full column rank 행렬이라고 하자. Huynh과 Feldt(1970)는 이때 $\boldsymbol{A}'\boldsymbol{\Sigma}_1\boldsymbol{A}=\boldsymbol{A}'\boldsymbol{\Sigma}_2\boldsymbol{A}=\cdots=\boldsymbol{A}'\boldsymbol{\Sigma}_J\boldsymbol{A}=\sigma^2\boldsymbol{I}_{K-1}$의 조건이 혼합설계의 교호작용에 대한 검정통계량이 $(J-1)(K-1)$과 $(N-J)(K-1)$의 자유도를 갖는 F분포를 하기 위한 필요충분조건이 됨을 보였다. 이 조건은 다표본 구형성(multisample sphericity) 조건이라고 일컬어진다(Hyunh, 1978). 그런데, 대부분의 경우에 있어서 반복측정설계나 혼합설계는 구형성의 조건을 만족하지 않는다(Rogan, Keselman, 그리고 Mendoza, 1979). 예를 들면 반복측정설계나 혼합설계에서 표본내 요인이 연속된 관측인 경우에는 인접한 관측값이 인접하지 않은 관측값보다 높은 상관을 갖게 된다(Huynh과 Feldt, 1970; Rogan, Keselman, 그리고 Mendoza, 1979). Box(1954)와 Imhof(1962)는 각각 반복측정설계와 혼합설계에서 구형성의 조건이 많이 위반될수록 표본내 인자에 대한 F검정 통계량은 커진다는 것을 발견했다. 또한 구형성이 위반된 경우의 표본내 인자에 대한 검정 통계량은 자유도가 $\varepsilon(K-1)$와 $\varepsilon(N-J)(K-1)$인 F분포에 근사한다는 것을 보였다.

구형성의 조건이 위반될 때 F검정 통계량의 자유도 수정 인자인 ε를 추정하기 위한 일련의 연구들이 있었으며, Greenhouse와 Geisser(1959)의 추정치, Huynh과 Feldt(1976)의 추정치, Huynh과 Feldt(1976)의 추정치를 수정한 Lecoutre(1991)의 추정치 등이 제안되었다. Huynh(1978)은 분산-공분산 행렬의 이질성을 고려하여 각각 Greenhouse와 Geisser(1959)의 추정치와 Huynh과 Feldt의 추정치를 사용한 수정 F검정을 다시 수정한 일반근사(General Approximation, GA)검정법과 향상된 일반근사(Improved General Approximation, IGA)검정법을 제안했다. Algina(1994)는 Lecoutre(1991)의 결과를 이용하여 수정된 향상 일반근사(Corrected Improved General Approximation, CIGA)검정법을 개발했다.

혼합설계에서 교호작용 검정의 강인성에 대한 연구는 지금까지 거의 실시된 바가 없다. 김현철(1998)은 최근에 F검정, $\tilde{\varepsilon}$-수정 F검정, CIGA검정을 비교하여, 실험조건 내에서 CIGA가 월등한 결과를 나타냄을 보고하였다. 혼합설계에서 반복측정 주효과 검정의 강인성과 검정력에 대하여는

몇 편의 제한된 연구가 있었다. Algina와 Oshima(1995)는 반복측정 주효과에 대한 가설을 검정하는 검정법들을 비교하였는데, 이들의 연구결과는 $\tilde{\varepsilon}$-수정 검정법이 1종오류를 IGA검정과 CIGA검정만큼 잘 관리하는 것으로 나타났다. Kim(1997a)은 한 개의 표본간 인자와 한 개의 표본내 인자를 가진 혼합설계에서 F검정, $\tilde{\varepsilon}$-조정 F검정, 그리고 CIGA검정의 1종오류와 검정력을 조사하였는데, Algina와 Oshima(1995)와는 달리 Kim(1997a)은 $\tilde{\varepsilon}$-수정 검정법이 1종오류를 잘 관리하지 못하는 것으로 보고하였다. Kim(1997a)은 CIGA 검정법이 1종오류를 적절하게 관리하며 검정력도 가장 높은 것으로 결론지었다. Kim(1997a)의 연구에서는 선행연구보다 광범위한 실험조건에서 검정법들이 비교되었는데, 선행연구와 다른 결론이 나온 이유로는 Kim(1997a)의 실험조건에 포함된 이분산성과 집단간 표본수 비율 조건이 지적되었다.

Wilcox(1993)는 한 개의 표본내 인자를 가진 반복측정 설계에서 반복측정 주효과의 검정에 절사평균(trimmed mean)의 이용을 연구하였다. Wilcox(1993)는 반복측정에서 구형성 가정과 정규분포의 가정이 위반되었을 때, 일반적인 $\tilde{\varepsilon}$-수정 F검정($\tilde{\varepsilon}$)과, 절사평균과 윈저화분산(Winsorized variance)을 이용한 $\tilde{\varepsilon}$-수정 F검정($\tilde{\varepsilon}_T$)의 1종오류와 검정력을 비교하였다. 모의실험의 결과 두 방법 모두 1종오류를 잘 관리하였으며 $\tilde{\varepsilon}$와 $\tilde{\varepsilon}_T$는 거의 유사한 1종오류를 갖는 것으로 나타났으나, $\tilde{\varepsilon}_T$는 $\tilde{\varepsilon}$에 비하여 실질 1종오류가 명목 1종오류를 넘어서는 경향이 감소하는 것으로 보고되었다. Kim(1997b)은 절사평균을 이용하여 F검정을 변형한 검정법(F_T), 절사평균을 이용하여 $\tilde{\varepsilon}$-수정 검정을 변형한 검정법 ($\tilde{\varepsilon}_T$), CIGA검정법, 절사평균을 이용하여 CIGA검정을 변형한 검정법($CIGA_T$)의 네 가지 검정법들의 반복측정 주효과에 대한 검정의 강인성을 비교하였다. 연구의 결과는 실질 1종오류(τ)가 명목 1종오류(α)보다 커지는 현상을 절사평균을 이용한 F_T나 $\tilde{\varepsilon}_T$가 많이 완화시키는 것으로 나타났고, CIGA와 CIGA는 모두 1종오류를 잘 관리하는 것으로 나타났다.

이 연구의 목적은 혼합설계의 교호작용에 대한 세 가지 검정법 (1) F검정(F), (2) $\tilde{\varepsilon}$-수정 F검정($\tilde{\varepsilon}$), (3)CIGA 검정(CIGA)과, 절사평균을 이용하여 각각 이들을 변형한 (4)절사평균을 사용하는 F검정(F_T), (5)절사평균을 사용한 $\tilde{\varepsilon}$-수정 F검정 ($\tilde{\varepsilon}_T$), (6)절사평균을 사용한 CIGA 검정($CIGA_T$)의 여섯 가지 검정법을 검정을 위한 가정들이 위반된 상황하에서 비교하는 것이다. 김현철(1998)에 나타났던 일부 조건에서의 F검정과 $\tilde{\varepsilon}$-수정 F검정의 1종오류 상승이 절사평균의 사용에 의하여 얼마나 통제되는가를 검토하고, 이들과 CIGA, $CIGA_T$검정과의 강인성을 비교하는 것은 흥미있는 연구가 될 것이다.

2. 연구방법

이 연구에 포함된 조건들은 김현철(1998), Kim(1997a, b)와 동일하다. 모든 조건에서 $J=2$, $K=4$였으며, 실험의 조건으로는

(1) 네 종류의 분포 ($g=0$과 $h=-0.244$, $g=0$과 $h=0$, $g=0$과 $h=0.109$, 그리고 $g=0$과 $h=0.35$),

(2) 공통 분산-공분산 행렬에 대한 세 종류의 구형성 (ε = 0.96, 0.75, 그리고 0.40),

(3) 분산-공분산 행렬에 대한 세 종류의 이질성 (Σ_1: $\Sigma_2 = 1:1$, 1:2, 그리고 1:5),

(4) 두 종류의 총표본수 (N = 40과 60), 그리고

(5) 세 종류의 집단간 표본수 비율로 N = 40일 때 (n_1, n_2) = (28, 12), (20, 20), (12, 28), N = 60일 때 (42, 18), (30, 30), (18, 42)

이 고려되어 실험의 총조건수는 4 × 3 × 3 × 2 × 3 = 216가지였다. 각 실험조건에 대하여 5000번의 반복실험이 실시되었다.

g와 h를 이용한 자료의 생성은 Tukey(1977)에 의하여 제안되고, Hoaglin(1985)에 의하여 개발된 방법으로 자료의 생성 과정이 간단하고, 어떤 형태의 분포를 따르는 자료도 생성할 수 있다는 장점이 있다. g는 분포의 왜도를 나타내는 값으로 0의 값을 가질 때는 분포가 대칭이 된다. 이 연구에서는 대칭인 분포만을 고려하였으므로 g가 네 가지 분포에서 모두 0의 값을 가지고 있다. h는 분포의 첨도를 나타내는 값으로 h의 값이 음수일 때는 정규분포보다 꼬리가 짧은 분포이고, h의 값이 양수일 때는 꼬리가 긴 분포이다. 그 값이 작을수록 꼬리가 짧은 분포를 갖고, 그 값이 클수록 꼬리가 긴 분포를 갖게 된다.

고려된 네 종류의 분포에서 $g=0$이고 $h=0$인 분포는 정규분포이며, $g=0$, $h=-0.244$인 분포는 일양분포(uniform distribution), $g=0$, $h=0.109$인 분포는 이중지수분포(double exponential distribution)이다. $g=0$, $h=0.35$인 분포는 이중지수분포보다 꼬리가 더 긴 분포이다. 그러므로 이 연구에 포함된 분포는 꼬리가 짧은 분포, 정규분포, 꼬리가 긴 분포, 그리고 꼬리가 아주 긴 분포의 네 가지이다. ε = 0.96, 0.75, 그리고 0.40의 조건을 만족하는 분산-공분산 행렬은 <표1>과 같이 Keselman과 Keselman(1988, 1990), Kim(1997a, b)에 사용된 것과 동일한 행렬을 사용하였다.

모의자료를 생성하는 절차는 다음과 같다. 우선 다변량 정규분포의 자료들은 다음 순서를 따라 생성되었다.

(1) 표본간 인자 수준 j에 대하여 상호 독립적이고 정규분포를 하는 $n_j \times 4$ 차원의 행렬 Z_j가 생성되었다. 이를 위하여 SAS(SAS Institute Inc., 1989)의 NORMAL function이 사용되었다.

(2) 행렬 Z_j는 $\boldsymbol{X}_j = \boldsymbol{\mu} + d_j \boldsymbol{Z}_j \boldsymbol{U}'$로 전환되었는데, 여기서 $\boldsymbol{\mu}$는 실험조건에서 요구하는 평균값(configuration of means)을 가진 $n_j \times 4$ 차원의 평균행렬이고, d_j는 실험조건에서 요구하는 이분산성을 생성하기 위한 상수이며, $\boldsymbol{U}$는 $\boldsymbol{\Sigma}_1 = \boldsymbol{UU}'$를 만족하는 하위삼각행렬(lower triangular matrix)이다.

비정규분포의 자료는 g-와-h 분포(g-and-h distribution)를 이용하여 다음 순서를 따라 생성되었다.

(1) 표본간 인자 수준 j에 대하여 상호 독립적이고 정규분포를 하는 $n_j \times 4$ 차원의 행렬 Z_j가 생성되었다. 이를 위하여 SAS의 NORMAL function이 사용되었다.

(2) $n_j \times 4$ 차원의 행렬 $\boldsymbol{X}_j^*$가 $X_{ij}^* = Z_{ij} \cdot \exp(h\, Z_{ij}^2/2)$에 의하여 만들어졌다.

(3) $n_j \times 4$ 차원의 행렬 $\boldsymbol{X}_j^*$가 $\boldsymbol{X}_j = \boldsymbol{\mu} + d_j\ \boldsymbol{X}_j^*\ \boldsymbol{U}'$로 전환되었는데, 여기서 $\boldsymbol{\mu}$, d_j,

그리고 U 는 앞의 다변량 정규분포 자료생성의 두 번째 단계에서 정의된 것과 같다.

평균이나 분산과 같은 전통적인 통계량들은 특이값에 대하여 민감하게 반응한다. 절사와 윈저화는 표본의 극단값들을 제거하거나 대치하는 방법이다. 절사평균은 20%의 절사를 실시하여 얻어졌으며, Tukey와 McLaughlin(1963)은 윈저화 표본분산이 절사평균 표본분산의 적절한 추정치임을 발견하였다. 이 연구에서 사용된 절사평균과 윈저화분산은 다음과 같이 구해졌다.

N개의 표본으로부터의 순서통계량 $x_{(i)}$가 $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(N)}$이라고 하자. α = g/N인 α-절사평균 또는 g회 절사평균은 다음과 같이 구해진다.

$$\bar{x}_{tg} = \frac{1}{(N-2g)} \sum_{i=g+1}^{N-g} x_{(i)}$$

<표1> ε = 0.96, 0.75, 0.40의 분산-공분산 행렬

ε	분산-공분산 행렬			
0.96	12.0	6.0	5.0	5.0
		10.0	5.0	4.0
			10.0	5.0
				8.0
0.75	18.0	8.0	6.0	4.0
		8.0	5.0	4.0
			7.0	3.0
				7.0
0.40	23.8	11.9	6.4	0.9
		9.5	5.7	2.6
			3.9	2.5
				2.8

g-회 윈저화 평균(Winsorized mean)과 분산은 각각 다음과 같이 구해진다.

$$\bar{x}_{wg} = \frac{1}{N}[gx_{(g+1)} + \sum_{i=g+1}^{N-g} x_{(i)} + gx_{(N-g)}]$$

$$SSD_{wg} = g(x_{(g+1)} - \bar{x}_{wg})^2 + \sum_{i=g+1}^{N-g} (x_{(i)} - \bar{x}_{wg})^2 + g(x_{(N-g)} - \bar{x}_{wg})^2$$

$\overline{X}_j$와 S_j가 집단 j의 표본평균 벡터와 표본표준편차 행렬을 나타낸다고 하면 절사평균을 이용한 검정법들은 $\overline{X}_j$에 앞의 방식으로 얻은 $(K \times 1)$ 차원의 g-회 절사평균 벡터와 S_j에 앞의 방

식으로 얻은 $(K \times K)$ 차원의 원저화 표본 표준편차 행렬을 대체하였다. 각 검정법들의 검정통계량과 기각치는 집단별 표본수 n_j 대신 $n_j - 2g$를 대체하여 수정하였다.

3. 결과

<표2>는 실험 결과 나타난 각 검정법들의 1종오류의 분포를 보여준다. 검정법의 강인성에 대한 Bradley(1978)의 기준에 의하면, 검정법의 실질 유의수준(τ)이 $0.5\alpha \leq \tau \leq 1.5\alpha$일 때 검정법은 강인(robust)하다고 정의된다. 여기서 α는 명목 유의수준을 나타낸다. 이 기준을 적용하면 <표2>는 F, F_T, $\tilde{\varepsilon}$, 그리고 $\tilde{\varepsilon}_T$검정들이 모두 일부 조건하에서는 명목 유의수준을 상당히 초과하는 실질 1종오류를 가지는 것을 보여주고 있는 반면, CIGA와 CIGA$_T$ 검정의 1종오류는 항상 강인한 검정이 되기 위한 Bradley의 기준을 만족하고 있는 것을 알 수 있다.

추정된 1종오류의 표준편차는 $\hat{\sigma}_\tau$ = $[\tau(1-\tau)/5000]^{1/2}$로부터 구해지는데 τ = 0.05일 때의 표준편차는 0.0031이 된다. 그러므로 가설 H_0: $\alpha = .05$을 검정하기 위한 유의수준 0.05에서의 임계치는 0.055가 된다. 이 기준에 의하면 CIGA와 CIGA$_T$ 검정법 역시 몇몇 조건에서 실질 유의수준이 명목 유의수준을 유의하게 상회하고 있음을 알 수 있다.

<표2> 1종오류의 분포

Test	Min	10	25	50	75	90	Max
F	0.0034	0.0206	0.0516	0.0630	0.1028	0.1926	0.2262
F_T	0.0046	0.0194	0.0468	0.0600	0.0990	0.1856	0.2214
$\tilde{\varepsilon}$	0.0032	0.0122	0.0372	0.0498	0.0574	0.1416	0.2120
$\tilde{\varepsilon}_T$	0.0042	0.0148	0.0406	0.0510	0.0584	0.1414	0.1902
CIGA	0.0278	0.0366	0.0458	0.0498	0.0526	0.0562	0.0670
CIGA$_T$	0.0354	0.0404	0.0450	0.0488	0.0528	0.0578	0.0778

참조. F = F검정; F_T = 절사자료 F검정; $\tilde{\varepsilon}$ = $\tilde{\varepsilon}$-조정 F검정; $\tilde{\varepsilon}_T$ = 절사자료 $\tilde{\varepsilon}$-조정 F검정; CIGA = CIGA검정; CIGA$_T$ = 절사자료 CIGA검정.

1종오류의 분석을 위하여 검정법이 반복측정 인자인 4(분포) × 3(구형성) × 3(분산-공분산 동일성) × 3(집단간 표본크기 비율) × 2(총표본수) × 3(검정법)의 분산분석이 실시되었다. 이 연구에서는 1종오류에 영향을 미치는 많은 인자들을 고려하였기 때문에 분산분석의 결과는 많은 수의 유효한 효과를 포함할 것으로 보인다. 그러므로 효과의 상대적인 크기를 비교하기 위하여 각 효과의 제곱평균성분(mean square component)이 다음과 같이 계산되었다.

$$MSC_{effect} = \frac{(MS_{effect} - MS_{error})}{T}$$

여기서 T = 효과에 포함되지 않은 인자들의 수준수의 곱

총분산을 제곱평균성분과 두 개의 오차분산의 합으로 정의하고, 이로부터 총분산 중 각 효과의 비율($\hat{\omega}^2$)이 계산되었다 (Myers, 1979).

<표3>은 총분산의 1% 이상을 설명하는 효과들을 보여준다. $\hat{\omega}^2$가 5%보다 큰 효과들만이 해석되었는데 $\hat{\omega}^2$가 5%보다 큰 효과들은 (1)검정× n_1/n_2 교호작용(0.2911), (2)검정× n_1/n_2× $\Sigma_1:\Sigma_2$ 교호작용(0.2089), (3)검정의 주효과(0.1851), (4) n_1/n_2의 주효과(0.1046), (5) n_1/n_2× $\Sigma_1:\Sigma_2$ 교호작용(0.737), 그리고 (6)검정×구형성(0.0587)인 것으로 나타났다.

총분산의 5% 이상을 설명하는 효과 6개 중 앞의 다섯 효과들에 대하여는 검정법과, n_1/n_2, 그리고 $\Sigma_1:\Sigma_2$의 교호작용이 가장 높은 차수를 가지므로 이의 설명을 위하여 세 인자의 각 수준별 평균 1종오류를 <표4>에 수록하였고, 여섯 번째 효과인 검정과 구형성의 교호작용에 대한 각 수준별 평균 1종오류는 <표5>에 수록되었다.

<표3> 분산의 1%이상을 설명하는 요인들

요인	$\hat{\omega}^2$
T × n_1/n_2	0.2911
T × n_1/n_2 × $\Sigma_1:\Sigma_2$	0.2089
T	0.1851
n_1/n_2	0.1046
n_1/n_2 × $\Sigma_1:\Sigma_2$	0.0737
T × ε	0.0587
T × $\Sigma_1:\Sigma_2$	0.0322
$\Sigma_1:\Sigma_2$	0.0105

참조. T= test; n_1/n_2 = 집단간 표본크기 비율;
$\Sigma_1:\Sigma_2$ = 집단간 이분산성; ε = 구형성.

<표4>에 나타난 결과는 집단간 표본의 수가 같을 경우($n_1 = n_2$)에는 분산-공분산의 이질성이 실질 1종오류의 추정치($\hat{\tau}$)에 거의 영향을 미치지 않는 것으로 나타났다. 집단간 표본의 수가 같은 경우에는 이분산성이 커짐에 따라 두 개의 F검정들(F, F_T)과 $\tilde{\varepsilon}_T$, CIGA$_T$에서 아주 작은 1종오류의 증가가 관측되었으나, $\tilde{\varepsilon}$와 CIGA는 이분산성에 영향을 받지 않았다. F검정들에서는 0.05 유의수준에서 실질 1종오류의 추정치($\hat{\tau}$)가 명목 1종오류(α)보다 유의하게 큰 값을 가졌으나 Bradley의 강인성 범위를 벗어나지는 않았으며, $\tilde{\varepsilon}$검정들과 CIGA검정들은 1종오류를 잘 관리하는 것으로 나타났다.

두 집단의 분산-공분산이 동일한 경우($\Sigma_1:\Sigma_2$ = 1:1)에는 집단간 표본수의 비율이 여섯 가지 검

정법 모두에 거의 영향을 미치지 않는 것으로 나타났다. 그러나 여섯 가지 검정법 모두 집단간 표본 크기가 동일한 경우의 1종오류가 집단간 표본 크기가 동일하지 않은 경우의 1종오류보다 낮았다. 두 집단의 분산-공분산이 동일한 경우에 $\tilde{\varepsilon}$와 CIGA는 $\hat{\tau}$가 α보다 낮은 수준을 유지하였으며, $\tilde{\varepsilon}_T$와 $CIGA_T$는 $\hat{\tau}$와 α와 거의 같거나 오차범위 내에서 α보다 약간 큰 $\hat{\tau}$수준을 유지하였다. F검정들(F, F_T)은 유의수준 0.05에서 $\hat{\tau}$가 α보다 유의하게 큰 값을 가졌으나, Bradley의 강인성 내의 값을 가졌다.

분산-공분산이 작은 집단의 표본수가 적고, 분산-공분산이 큰 집단의 표본수가 많은 경우 ($\Sigma_1 : \Sigma_2 \neq 1:1$, $n_1 < n_2$)에는, F와 F_T, $\tilde{\varepsilon}$와 $\tilde{\varepsilon}_T$의 1종오류가 심하게 축소되었다. 축소의 정도는 F검정들보다 $\tilde{\varepsilon}$검정들에서 극단적으로 나타났다. F_T와 $\tilde{\varepsilon}_T$에서 축소의 정도가 완화되기는 하였지만, 네 검정법에서 모두 이분산성이 큰 경우($\Sigma_1 : \Sigma_2$ = 1:5, $n_1 < n_2$)에는 Bradley의 강인성 범위를 벗어나는 아주 작은 값의 $\hat{\tau}$를 가졌다. 반면에 CIGA와 $CIGA_T$는 1종오류의 완만한 감소가 관측되었으며, 감소의 정도는 오차한계 내였다.

<표4> 집단간 이분산성과 집단간 표본크기의 비율에 대한 1종오류의 추정치

$\Sigma_1 : \Sigma_2$	검정	$n_1 < n_2$	$n_1 = n_2$	$n_1 > n_2$
1 : 1	F	0.0721	0.0695	0.0714
	F_T	0.0682	0.0635	0.0681
	$\tilde{\varepsilon}$	0.0501	0.0483	0.0498
	$\tilde{\varepsilon}_T$	0.0514	0.0485	0.0510
	CIGA	0.0497	0.0472	0.0491
	$CIGA_T$	0.0511	0.0470	0.0498
1 : 2	F	0.0367	0.0698	0.1265
	F_T	0.0349	0.0651	0.1205
	$\tilde{\varepsilon}$	0.0225	0.0480	0.0958
	$\tilde{\varepsilon}_T$	0.0242	0.0494	0.0975
	CIGA	0.0483	0.0467	0.0507
	$CIGA_T$	0.0477	0.0476	0.0521
1 : 5	F	0.0159	0.0713	0.2108
	F_T	0.0162	0.0689	0.1980
	$\tilde{\varepsilon}$	0.0078	0.0492	0.1689
	$\tilde{\varepsilon}_T$	0.0097	0.0523	0.1665
	CIGA	0.0471	0.0468	0.0504
	$CIGA_T$	0.0470	0.0488	0.0527

분산-공분산이 작은 집단의 표본수가 적고, 분산-공분산이 큰 집단의 표본수가 많은 경우($\Sigma_1:\Sigma_2 \neq 1:1$, $n_1 > n_2$)에는, F와 F_T, $\tilde{\varepsilon}$와 $\tilde{\varepsilon}_T$의 1종오류가 심하게 확대되었다. 확대의 정도는 $\tilde{\varepsilon}$ 검정들보다 F검정들에서 극단적으로 나타났다. F_T와 $\tilde{\varepsilon}_T$에서 축소의 정도가 완화되기는 하였지만, 네 검정법에서 모두 이분산성이 극단적이지 않은 경우($\Sigma_1:\Sigma_2$ = 1:2, $n_1 < n_2$)에도 Bradley의 강인성 범위를 벗어나는 아주 큰 값의 $\hat{\tau}$를 가졌다. 반면에 CIGA와 $\mathrm{CIGA_T}$의 1종오류는 오차한계 내에서 완만한 증가가 관측되었으며, CIGA의 증가가 $\mathrm{CIGA_T}$의 증가보다 적게 나타났다.

반복측정에 대한 구형성의 가정이 위반됨에 따라 F검정들(F, F_T)과 CIGA검정들(CIGA, $\mathrm{CIGA_T}$)에서는 1종오류의 증가가 관측되었다. 그러나 $\tilde{\varepsilon}$검정들($\tilde{\varepsilon}$, $\tilde{\varepsilon}_T$)은 구형성의 조건에 거의 영향을 받지 않는 것으로 나타났다. 구형성의 조건이 거의 만족되는 조건(ε=0.96)하에서는 CIGA 검정들이 1종오류를 잘 관리하였으며, F검정들과 $\tilde{\varepsilon}$검정들은 유의수준 0.05에서 α보다 유의하게 큰 $\hat{\tau}$을 가졌다.

<표5> 검정법과 구형성에 대한 1종오류의 추정치

	e=.96	e=.75	e=.40
F	0.0675	0.0756	0.1049
F_T	0.0618	0.0718	0.1008
$\tilde{\varepsilon}$	0.0625	0.0597	0.0580
$\tilde{\varepsilon}_T$	0.0599	0.0623	0.0613
CIGA	0.0467	0.0474	0.0513
$\mathrm{CIGA_T}$	0.0446	0.0491	0.0542

4. 결론

연구의 결과는 집단간에 분산-공분산의 크기가 다르고, 표본의 수도 다른 경우에는 전통적인 F검정(F)이나 절사평균을 이용하여 변형된 F검정(F_T), $\tilde{\varepsilon}$-조정 F검정($\tilde{\varepsilon}$), 절사평균을 이용하여 변형된 $\tilde{\varepsilon}$-조정 F검정($\tilde{\varepsilon}_T$)은 모두 사용이 적절하지 않은 것으로 나타났다. F와 F_T는 분산-공분산이 동일한 경우에도 두 집단간에 표본의 수가 다르면 사용이 적절하지 않은 것으로 나타났다. 즉, F 검정들은 두 집단의 분산-공분산이 같고 표본의 크기도 같을 경우에만 실질 1종오류가 명목 1종오류와 근사한 값을 가지며, 그 외의 경우에는 명목 1종오류와 실질 1종오류간에 커다란 차이가 있는 것으로 나타나 이들을 사용할 때는 검정이 적용되는 자료의 상태에 대한 점검의 필요성이 지적되었다. 반면에 CIGA검정들(CIGA, $\mathrm{CIGA_T}$)은 이 연구의 실험조건 내에서 어느 경우에도 1종오류를 잘 관리하고 있는 것으로 나타나, 앞의 두 가지 검정법들보다 월등히 우수한 것으로 보고되었다. 이 결과는 혼합설계에서 반복측정 주효과에 대한 검정법들을 비교한

Kim(1997a)의 결과와 거의 유사하여 CIGA와 $CIGA_T$검정은 혼합설계에서 반복측정의 주효과 검정이나 집단과 반복측정간의 교호작용 검정에 F검정이나 F_T검정, $\tilde{\varepsilon}$검정이나 $\tilde{\varepsilon}_T$검정보다 우수한 것으로 나타났다.

절사평균과 윈저화분산을 사용한 검정법들은 평균과 분산을 사용한 검정법들에 비하여 실질 1종오류가 명목 1종오류보다 너무 작아지거나 너무 커지는 경향을 약간 완화시켰으나, 원래 F검정과 $\tilde{\varepsilon}$-조정 F검정이 가지고 있던 문제점을 크게 개선하지는 못하였다. 자료의 분포가 꼬리가 길거나(h = 0.109) 아주 길 때는(h = 0.35) 절사평균의 사용으로 인하여 원래의 F검정과 $\tilde{\varepsilon}$-조정 F검정이 가지고 있던 문제점이 상당히 완화될 것으로 기대되었으나, 검정과 분포의 교호작용이 전체 분산의 1%도 설명하지 못하는 것으로 나타났다. 전체 분산에 대한 설명력이 낮아서 결과 단원에서 그 효과가 해석되지는 않았으나 효과 자체에 대한 분산분석의 결과는 검정과 분포의 교호작용이 유의함을 보여주었다.

구형성 가정(ε = 1)의 위반 효과는 F검정들(F, F_T)과 CIGA검정들(CIGA, $CIGA_T$)에서 큰 것으로 나타났고, $\tilde{\varepsilon}$검정들($\tilde{\varepsilon}$, $\tilde{\varepsilon}_T$)에서는 그 효과가 거의 없는 것으로 관측되었다. 이는 $\tilde{\varepsilon}$검정들($\tilde{\varepsilon}$, $\tilde{\varepsilon}_T$)을 분산-공분산에 이질성이 있을 경우에 대하여 사용할 수 있도록 일반화시킨 CIGA검정들(CIGA, $CIGA_T$)이 본래의 $\tilde{\varepsilon}$검정들($\tilde{\varepsilon}$, $\tilde{\varepsilon}_T$)의 장점을 잃은 결과로 그 이유와 이의 개선에 대한 이론적 연구가 필요하다고 보여진다.

여기서 제시된 결과들은 이 연구에서 고려된 실험조건들의 (1)분포, (2)집단간 표본비율의 범위, (3)집단간 이분산성의 범위에 의하여 일반화가 제한되고 있다. 이후의 연구에서는 다음 사항들을 고려하는 실험이 필요하다고 생각된다. 우선 첫째로 이 연구에서는 첨도가 서로 다르고 대칭인 분포들만을 고려하였는데, 대칭이 아닌 분포에서는 어떤 결과를 보이는가에 대한 이후의 연구가 필요하다고 보여진다. 둘째로는 이 연구나 선행연구들에서 사용된 형태의 분산-공분산 행렬과는 다른 형태의 분산-공분산 행렬을 고려하여 검정법들을 비교할 수 있을 것으로 생각된다. 예를 들면, 1차나 2차의 자기상관(Autocorrelation, AR)모형에 의한 분산-공분산의 생성은 반복측정설계나 혼합설계에서 적절한 방법이 될 것이다. 셋째로, 이 연구에서는 반복측정이 완성된 경우만을 고려하였는데 자료의 누락(missing data) 효과에 대한 연구 역시 흥미있는 주제가 될 것이다. 마지막으로, 비정규성은 많은 자료들이 공유하고 있는 일반적인 성질이기 때문에 비모수통계를 이용한 검정법들에 대한 연구 결과와 여기서 검토된 방법들의 결과를 비교하는 것도 중요한 과제라고 할 수 있겠다.

참고문헌

[1] 김현철(1998). 혼합설계의 교호작용에 대한 검정의 강인성 비교, 게재 심사 중.

[2] Algina, J. (1994). Some alternative approximate tests for a split plot design. *Multivariate Behavioral Research*, Vol. 29, 365-384.

[3] Algina, J., & Oshima, T. C. (1995). An improved general approximation test for the main effect in a split-plot design. *British Journal of Mathematical and Statistical Psychology*, Vol. 48, 149-160.

[4] Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics*, Vol. 25, 290-302.

[5] Bradley, J. C. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, Vol. 31, 144-152.

[6] Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, Vol. 24, 95-112.

[7] Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distributions. In D.Hoaglin, F. Mosteller, & J. Tukey (Eds.), *Exploring data tables, trends, and shapes*. New York: Wiley.

[8] Huynh, H. (1978). Some approximate tests for repeated measurement designs, *Psychometrika*, Vol. 43(2), 161-175.

[9] Huynh, H., & Feldt, L. S. (1970). Conditions under which mean squared ratios in repeated measurements designs have exact F distributions. *Journal of the American Statistical Association*, Vol. 65, 1582-1589.

[10] Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs, *Journal of Educational Statistics*, Vol. 1(1), 69-82.

[11] Imhof, J. P. (1962). Testing the hypothesis of fixed main-effects in Scheffe's mixed model, *Annals of Mathematical Statistics*, Vol. 33, 1085-1095.

[12] Keselman, J. C., & Keselman, H. J. (1988). Repeated measures multiple comparison procedures: Effects of violating multisample sphericity in unbalanced designs, *Journal of Educational Statistics*, Vol. 13, 215-226.

[13] Keselman, J. C., & Keselman, H. J. (1990). Analysing unbalanced repeated measures designs. *British Journal of Mathematical and Statistical Psychology*, Vol. 43, 265-282.

[14] Kim, H. (1997a). Type I error rates and power for omnibus tests of repeated measures means in the split plot design: F test, $\tilde{\varepsilon}$-adjusted F test, and CIGA test, 한국통계학회논문집, Vol. 4(1), 139-149.

[15] Kim, H. (1997b). Robustness for omnibus tests using trimmed means under violated assumptions, 한국통계학회논문집, Vol. 4(2), 581-588.

[16] Lecoutre, B. (1991). A correction for the ε approximate test in repeated measures designs with two or more independent groups. *Journal of Educational Statistics*, Vol. 16, 371-372.

[17] Myers, J. L. (1979). *Fundamentals of experimental design* (3rd ed.). Boston: Allyn and Bacon.

[18] Rogan, J. C., Keselman, H. J., & Mendoza, J. L. (1979). Analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology*, Vol. 32, 269-286.

[19] SAS Institute Inc. (1989). *SAS/IML Software: Usage and Reference, Version 6*, 1st ed. Cary, NC: Author.

[20] Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA, Addison-Wesley.

[21] Tukey, J. W., & McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization 1. *Sankyha A*, Vol. 25, 331-352.

[22] Wilcox, R. R. (1993). Analyzing repeated measures or randomized block designs using trimmed means. *British Journal of Mathematical and Statistical Psychology*, Vol. 46. 63-76.