응용통계연구 제 8권 1호
1995년 3월 pp. 119-131.

윈저화를 이용한 로버스트 분산분석[1)]

성 내 경[2)]

요 약

윈저화 자료에 기초한 분산분석법 개발의 일차 시도로 고정효과 일원 분산분석 모형에 대한 윈저화 분산분석을 제시한다. 몬테 칼로 모의실험 기법을 사용하여 각 요인 수준마다 g-g 대칭 윈저화를 적용시켰을 때 윈저화 자료에 기초한 제곱합들의 비의 경험적 분포가 통상의 F 분포로 근사됨을 보인다. 이 근사 F 분포의 자유도는 윈저화 카이제곱 통계량의 경험적 분포가 자유도 $(n-3g-1)$의 통상적인 카이제곱 분포에 만족할만하게 근사되어진다는 성내경(1994)의 연구 성과를 토대로 결정된다. 여기서 n은 표본 크기, g는 한쪽 꼬리 부분에서 윈저화가 적용되는 양이다. 산출된 분산비의 경험적 분위수의 일부를 수록하였다. 이 연구는 non-adaptive 로버스트 분산분석법을 제안하는 것으로 이상점이 존재하는 분산분석 자료에 적용하면 자료 해석이 단순화되는 실용성을 위주로 한다.

1. 서 론

고전적 분산분석 모형에서 오차항은 서로 독립이며 평균이 0, 분산은 모든 요인 수준마다 동일한 정규 확률변수로 간주된다. 그러나 실제로 나타나는 오차의 대표적인 분포는 정규분포보다 더 긴 꼬리를 대칭형이다(Tukey와 McLaughlin, 1963). Huber (1981, p.196)가 지적한대로 오차에 대한 기본 가정이 파괴되면 분산분석의 검정력은 심각한 손상을 입을 수 있다. 특히 꼬리가 아주 두터운 오차 분포이거나 이상점이 존재하는 상황에서 고전적인 F 검정은 검정력이 상당히 떨어지는 것으로 보고되고 있다.

편의상 균형된 일원 분산분석 모형을 고려하자. 요인의 수준수를 I, 전체 관측수를 N이라 가정하자. 요인제곱합을 SSB, 오차제곱합을 SSE로 표기한다. 이때 통상적인 F 통계량은 다음과 같이 주어진다.

$$\frac{\mathrm{SSB}/(I-1)}{\mathrm{SSE}/(N-I)} \tag{1}$$

Huber의 지적에 의하면, 분산분석의 기본 가정이 심각하게 손상된 경우에 식 (1)의 분자는 어떤 로버스트 추정값으로 대치해도 별 문제가 없으나, 분모는 전혀 로버스트하지 않으므로 쓸모가 없다고 단언하고 있다. 그러나 분모의 대치, 즉, 오차 분산의 로버스트 추정 방법에 대해서는 단순히 M-추정량을 사용하는 것이 한 가지 대안이고 SSE의 자유도를 다소 낮추어야 한다는 것을 제안하고 있을 뿐 분포가 어떻게 될지에 대해서는 미지라고 밝히고 있다.

물론 Huber의 지적은 상식적인 지적이지만 로버스트 분산분석을 개발할 때 고려해야할 기본적인 사항들을 간결히 요약하고 있다. 요컨대 로버스트 분산분석법의 요체는 오차분산의 적절

1) 이 연구는 1993년도 한국학술진흥재단의 자유공모과제 연구비에 의하여 수행되었음.
2) (120-750) 서울시 서대문구 대현동 11-1, 이화여자대학교 통계학과.

한 추정에 달려있다고 할 수 있다.

기실 이런 류의 문제에 대하여 정확한(exact) 분포를 구해내기는 거의 불가능하다. 대표본 이론을 적용한 고전적 분산분석에 대한 로버스트 대안의 하나로 Schrader와 Hettmansperger (1980)의 연구를 들 수 있다. 이들의 방법은 adaptive 접근법으로 간주될 수 있으며, Huber의 M-추정을 이용하여 일반 선형 가설에 대한 우도비 형식의 분산분석법을 제안하였다. 그러나 이들의 방법은 본질적으로 대표본 근사를 이용한 것이라 관측수가 적은 자료에 적용하기에 난점이 있지만, 상호작용 효과의 유무에 논란이 있던 데이터(Box & Cox, 1964)에 적용한 결과 상당히 만족할만한 결과를 얻을 수 있었다. 그러나 가정되는 오차 분포에 따라 검정 통계량의 계산이 간단하지 않은 문제가 있다.

분산분석의 일차적 관심사는 요인 수준간 차이에 의한 유의한 변동이 존재하는지 여부이다. 만일 요인 수준간에 정말로 차이가 있다고 결론을 내린다면 그다음 질문이 자연스럽게 제기된다. 고정효과 요인이라면 요인의 어느 수준들이 서로 다른가? 또, 랜덤효과 요인이라면 관심 요인에 대한 분산 요소의 추정값은 얼마인가?

고정효과 모형에서 약간의 비정규성은 중요한 문제가 못된다. 요인 수준의 평균의 점 추정량은 오차 분포에 상관없이 불편의이며, 통상의 F 검정 또한 검정력이 별로 떨어지지 않는다. 그러나 이상점이 존재하면 오차 분산이 과대 추정됨을 피할 수 없고 요인 수준 평균의 구간추정값도 더 넓게 된다. 따라서 예를 들어 Scheffe의 다중비교법에서는 훨씬 더 보수적인 결과가 나타난다. 랜덤효과 모형에서도 분산요소의 점추정량은 여전히 불편의일지라도 구간 추정값에는 상당히 악영향을 끼친다.

이 논문에서는 얻어진 자료에서 각 요인 수준마다 독립적으로 g-g 대칭 윈저화를 적용한 다음 통상적인 분산분석 절차를 그대로 진행하는 형식의 새로운 로버스트 분산분석법을 제시한다. 이 분산분석법의 아이디어는 정규분포 하에서 오차 분산을 추정할 때 오차 제곱합의 자유도 결정은 단일 표본에서 윈저화 제곱합의 경험적 분포가 자유도 $(n-3g-1)$의 카이제곱 분포에 근사될 수 있다는 성내경(1994)의 연구를 기초로 한다. 이 분산분석법을 앞으로는 윈저화 분산분석(Winsorized ANOVA)이라 부르겠다. 제 2 절에서는 오차의 분포를 대칭이라고만 가정할 때 이 방법을 한 가지의 새로운 분산분석 방법으로서 소개한다.

제 3 절에서는 특히 정규분포 하에서 몬테 칼로 모의실험을 통하여 윈저화 분산분석의 유사 F 통계량(pseudo-F statistic)의 경험적 분포를 산출하고, 이 통계량의 분포가 적당한 자유도를 갖는 통상의 F 분포로 근사됨을 제시한다.

윈저화 분산분석에서는, 따라서, 원시 자료에 대한 윈저화와 SSE의 자유도 결정이 다를 뿐 본질적으로 고전적 분산분석의 기본 형식을 그대로 답습한다.

2. 고정효과모형

여기서는 단일 요인 실험을 고려한다. 편의상 균형 자료만을 생각하고 모형을 다음과 같이 쓰기로 하자.

$$x_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i=1,\cdots,I, \quad j=1,\cdots,n, \tag{2}$$

여기서 x_{ij}는 i번째 요인 수준에서 j번째 관측이고, μ는 총평균, α_i는 i번째 요인 수준의 고정 효과, 그리고 ε_{ij}는 i번째 요인 수준에서 j번째 관측에 대한 랜덤 오차이다. 서로 독립인 ε_{ij}들은 평균이 0, 분산이 σ^2인 대칭 분포를 따른다. 또한 하나의 요인 수준 내의 관측들이 이미 크기순으로 순서화되어 있다고 가정하자. 즉, $x_{i1} \le x_{i2} \le \cdots \le x_{in}$ $\forall\, i$.

$\bar{x}_{t,i}$를 g-g (대칭) 절단 자료에서 i번째 요인 수준의 표본평균, 그리고 $\bar{x}_t$를 $\bar{x}_{t,i}$들의 평균이라 하자. 즉, 다음과 같은 표기를 약속한다.

$$\bar{x}_{t,i} = \sum_{j=g+1}^{n-g} x_{ij} / (n-2g) \qquad i=1,\cdots,I$$

$$\bar{x}_t = \sum_{i=1}^{I} \bar{x}_{t,i} / I$$

원시 자료와 g-g (대칭) 윈저화 자료를 구별하기 위하여, 윈저화 자료는 앞으로 x_{ij}대신 z_{ij}로 표기한다. 즉, g-g 윈저화 후에 요인 수준 내의 관측들은 다음과 같이 표현된다.

$$\begin{aligned} &z_{i1} = z_{i2} = \cdots = z_{ig} = x_{i,g+1} \\ &z_{i,g+k} = x_{i,g+k}, \quad k=1 \le \cdots \le h = n-2g \\ &z_{in} = z_{i,n-1} = \cdots = z_{i,n-g+1} = x_{i,n-g}, \qquad i=1,\cdots,I. \end{aligned} \tag{3}$$

$\bar{z}_{w,i}$를 g-g 윈저화 자료에서 i번째 요인 수준의 표본평균, 그리고 $\bar{z}_w$를 윈저화 자료의 총평균이라 하자. 즉, 다음과 같은 표기를 약속한다.

$$\bar{z}_{w,i} = \sum_{j=1}^{n} z_{ij} / n, \qquad i=1,\cdots,I,$$

$$\bar{z}_w = \sum_{i=1}^{I} \bar{z}_{w,i} / I.$$

본 논문에서 제안하는 윈저화 분산분석의 요체는 (i) μ와 $\mu_i=\mu+\alpha_i$를 각각 절단평균 $\bar{x}_t$와 $\bar{x}_{t,i}$로 추정하고, (ii) 요인의 유의성 검정과 오차 분산의 추정은 윈저화 자료에 대한 분산분석표를 기초로 하는데 있다. 이 절차는 Stigler (1973)가 보인 바, 단일표본에서 절단평균은 근사적으로 분산이 윈저화 분산인 정규분포를 따른다는 대표본 이론에 근거한다.

윈저화 자료에서 총편차 $(z_{ij} - \bar{z}_w)$는 평소와 같이 다음의 두 요소로 분할된다.

$$(z_{ij} - \bar{z}_w) = (\bar{z}_{w,i} - \bar{z}_w) + (z_{ij} + \bar{z}_{w,i}). \tag{4}$$

(4)를 제곱하여 더하면 다음 식을 얻게 된다.

$$\sum_i \sum_j (z_{ij} - \bar{z}_w)^2 = \sum_i \sum_j (\bar{z}_{w,i} - \bar{z}_w)^2 + \sum_i \sum_j (z_{ij} - \bar{z}_{w,i})^2 \tag{5}$$

식 (5)의 좌변을 SSTw로 표기하고 윈저화 총제곱합(Winsorized total sum of squares)이라 부른다. 비슷하게 우변의 첫번째 항은 SSBw로 표기하고 윈저화 처리 제곱합(Winsorized

between sum of squares)을 나타내며, 우변의 두번째 항은 SSE_W로 표기하고 윈저화 오차 제곱합(Winsorized error sum of squares)이다.

고전적 분산분석과 유사하게 제곱합들의 윈저화 판을 얻었기 때문에 유사 F 검정 통계량을 얻으려면 각 윈저화 제곱합에 적절한 자유도를 배당해야 한다. 오차의 분포가 정규분포라는 가정이 없었고, 윈저화 관측들은 더 이상 독립이 아니므로 고전적 분산분석의 자유도를 배정할 수는 없다.

그러나 자료의 윈저화는 각 수준별로 독립적으로 수행되었으며, 오차의 분포는 대칭이기 때문에 $\bar{z}_{w,i}$와 $\bar{z}_w$는 각각 여전히 μ_i와 μ에 대한 적법한 측도이다. 따라서 차이 $(\bar{z}_{w,i}-\bar{z}_w)$는 μ_i들의 μ로부터의 편차의 측도가 되며, I 개의 수준이 있으므로 SSB_W의 자유도로는 $(I-1)$을 배정한다. SSE_W에는 자유도로 $I(k-1)$, (단$k=n-3g$)를 배정하는 것이 적절하다. 이 부분은 성내경(1994)의 윈저화 제곱합의 경험적 분포에 대한 몬테 칼로 모의실험 결과를 적용한 것이다. 마지막으로 SST_W에는 $(Ik-1)$의 자유도를 배정한다.

따라서 처리 및 오차의 평균제곱은 아래와 같이 정의된다.

$$MSB_W = SSB_W/(I-1), \quad MSE_W = SSE_W/I(k-1).$$

여기서 MSB_W는 윈저화 처리 평균제곱(Winsorized between mean square), 그리고 MSE_W는 윈저화 오차 평균제곱(Winsorized error mean square)이다. 이때, MSE_W가 오차 분산의 추정량이다.

요인 수준 평균이 같은지 다른지에 대한 검정은 유사 F 검정 통계량인 F^*를 기준으로 한다.

$$F^* = MSB_W / MSE_W$$

그리고, F^*는 분자의 자유도가 $(I-1)$, 분모의 자유도가 $I(k-1)$인 F 분포로 근사한다.

3. 윈저화 F 통계량의 경험적 분포

이 절에서는 정규분포 하에서 윈저화 자료로부터 형성된 유사 F 통계량의 분포 행태를 탐색한다. 이 유사 F 통계량의 정확한 분포는 분석적으로 얻을 수 없기 때문에 몬테칼로 모의실험 기법으로 경험적 분포함수를 추정하고, 제 2 절에서 언급한 자유도의 F 분포로 간주하고 추론을 전개하여도 실용상 문제가 없음을 보인다.

모형 (2)에 주어진 바, i번째 수준에서 $x_{i1} \le x_{i2} \le \cdots \le x_{in}$을 정규분포 $N(\mu_i,\sigma^2)$에서 추출된 n 개 관측을 순서화한 표본이라 하자. 이 자료에 대한 g-g($g \ge 1$) 대칭 윈저화 자료의 형태는 (3)을 보라. 여기서 $n=g+h+g$ (순서화를 강조하는 방편으로 이런 표현이 $n=2g+h$보다 선호된다)이며, 양쪽 꼬리 부분에서 g 개의 관측들의 값이 수정되고 가운데의 h 개 관측들은 윈저화의 영향을 받지 않는다.

귀무가설 하에서 일반성의 상실없이 $\mu_i=0$와 $\sigma^2=1$을 가정하자. SAS 6.04 패키지에 내장된 정규 난수 생성기인 RANNOR를 이용하여 각 수준마다 표본 크기 n의 랜덤 표본 1,000 개를

10조 생성한다. 고려된 수준수는 3부터 10까지, 각 수준의 표본 크기는 5부터 15까지로 1 단위씩 증가시켰다. 각 수준마다 생성된 관측들을 순서화하고 원저화를 적용한다. 원저화의 양 g는 표본 크기에 따라 달라지지만, Gastwirth와 Cohen (1970)의 연구 결과를 따라 20%까지의 원저화를 적용하였다. 한 조의 1,000개 표본에 대하여, 각 원저화 자료마다 원저화 F 통계량을 계산하고 선택된 누적확률값들에 대하여 원저화 F 통계량의 경험적 분위수를 산출한다.

원저화 F 통계량의 경험적 분위수들을 구한 후 이 값들을 근사되는 통상의 F 분포와 비교하여, 만족할만한 근사가 이루어지는지 확인하였다. 즉, 최대 관심 영역은 임계값이 위치하는 위꼬리 부분이므로, 이 영역에서 원저화 F의 경험적 분위수와 진짜 F 분포의 분위수 간의 비를 취하여 만족스러운 근사가 되는지 확인한다. 이러한 탐색 방법은 Dixon과 Tukey (1968) 스타일의 탐색법을 원용한 것으로, 이 결과 예측한대로 만족할만한 근사가 이루어짐을 발견하였다.

표 1에 몇 가지 선택된 확률값에 대한 원저화 F 통계량의 경험적 분위수, 표준오차, 그리고 경험적 분위수와 근사되어지는 F 분포에서 도출된 분위수 간의 비가 수록되어 있다. 표 1은 완전한 분위수 표의 일부이다. 표 1을 보면 일반적으로 수준수에 상관없이 수준내 관측수 n이 5나 6일 때의 근사가 잘 되지 않는다. n=6일 때는 수준수가 증가함에 따라 근사가 향상되기는 하나 만족할만한 수준은 아니다. 이런 경향의 표출은 이표본에서 절단화 합동 t 검정을 제안한 Yuen과 Dixon (1973)의 모의실험 결과와 동일하며, n이 5나 6일 때는 경험적 분위수를 직접 사용하라는 Yuen과 Dixon의 충고를 여기서도 받아들일 수 밖에 없을 것 같다. 기타 다른 모의실험 모수 조합의 결과는 실용상 F 근사에 별다른 무리가 없어 보인다.

4. 기 타

앞서 언급하였지만 분산분석표는 원저화 자료를 기초로 만들어지지만 요인 수준 평균 μ_i의 추정량으로는 원저화 평균 대신에 절단 평균을 사용한다. 따라서 Fisher의 LSD에 대한 로버스트 대안은 다음과 같다.

$$\text{LSDT} = t_{I(k-1),\alpha/2}\sqrt{2\,\text{MSEw}/h}\,.$$

물론 원시 자료에 원저화를 적용하지 않았다면 LSDT는 통상의 LSD와 동일하다.

이 논문에서는 고정효과 일원 분산분석 모형만을 고려하였지만 랜덤효과 모형에서도 추론 방식은 일치하며, 분산 요소의 추론에서도 원저화 분산분석표의 값들을 그대로 사용함을 부기한다. 또 이원 분산분석 모형으로의 확장도 어렵지 않다.

원저화 분산분석의 또다른 대안으로 이상점이 포함된 수준들만을 원저화하는 선택적 원저화(selective Winsorization)을 고려할 수 있다. 이는 adaptive 접근 방식으로, 지금까지 고려한 원저화 방식은 선택적 원저화와 대비하여 총체적 원저화(total Winsorization)라 부를 수 있다. 선택적 원저화의 경우 분산분석표에서 검정을 할 때 오차 제곱합에 대한 자유도가 변경되어야 하나, 새로운 자유도의 공식은 자명하므로 이 부분에 대한 언급은 생략한다.

제 3 절에서는 오차가 정규분포를 따른다는 가정 하에서 경험적 분포함수를 산출하였지만 이 가정이 원저화 분산분석법의 효용을 감소시키지는 않는다. 현실적으로 분산분석 자료를 분석할 때 아무리 자료의 분포가 이상하더라도 변환을 고려하지 않는다면 자료의 분포로서 정규분포 이외에 다른 분포를 가정할만한 대안이 별로 없기 때문이다.

그리고, 제 1 절에서 Huber의 지적을 언급한 바, 자료가 정규분포보다 더 긴꼬리 분포로부터 도출되었을 때 통상의 F 통계량에서 추정에 문제가 되는 부분은 분모의 오차평균제곱항이다. 윈저화를 적용하였을 때 이 분모항 값의 감소 효과는 성내경(1994)에서 제시한 응용례에서 엿볼 수 있다. 그리고 오차평균제곱이 감소하면 다중비교시 검정력이 더 증진됨도 자명하다. 따라서 이러한 면에 대한 로버스트성의 확인례는 생략하였다.

이상에서 제안한 윈저화 분산분석은 수준당 최소한 관측수가 5 이상이 되어야 의미가 있다. 따라서 수준내 관측수를 4 이상으로 하며 동시에 근사 자유도를 더 높이면서 정밀도가 높은 근사 공식의 개발이 차후의 과제라 하겠다.

표 1. 윈저화 자료에서 산출된 F 통계량의 경험적 분위수. 수록된 값들은 몇 개의 선택된 확률값들에 대한 대칭 윈저화 분산비 분포의 경험적 분위수와 그들의 표준오차, 그리고 경험적 분위수 대 진짜 F 분포의 분위수의 비이다. l은 수준수, n은 수준 내의 표본 크기, g는 각 꼬리에서 대칭 윈저화의 양이며 df는 윈저화 자료에서 결정되는 F 분포의 자유도이다. 표 우측의 MSE는 추정된 오차평균제곱의 값이고, se는 MSE에 대한 표준오차이다. 이 표에 나와있는 값들은, $n \leq 15$과 $g \geq 1$의 모든 적절한 조합에 대하여 산출된 경험적 분위수들 중 일부이다.

l	n	g	df	.8	.9	.95	.975	.99	MSE	se
3	5	1	(2,3)	1.9009	3.2361	4.9044	6.9930	11.8699		
				0.0443	0.1023	0.1475	0.1944	0.4485		
				0.6586	0.5924	0.5134	0.4359	0.3852	1.374	0.0282
3	6	1	(2,6)	1.9084	3.0266	4.4060	5.9565	9.0236		
				0.0276	0.0458	0.0748	0.1571	0.4256		
				0.8960	0.8739	0.8567	0.8205	0.8260	1.125	0.0183
3	7	1	(2,9)	1.8826	2.8743	4.0098	5.2356	7.8145		
				0.0282	0.0536	0.0953	0.1733	0.2662		
				0.9730	0.9560	0.9421	0.9162	0.9742	1.053	0.0145
3	8	1	(2,12)	1.8782	2.8788	3.8761	5.1895	6.8079		
				0.0194	0.0370	0.0324	0.1113	0.2488		
				1.0175	1.0257	0.9976	1.0184	0.9828	1.028	0.0126
3	9	1	(2,15)	1.8674	2.7846	3.7853	4.8495	6.2881		
				0.0300	0.0414	0.0530	0.1065	0.1297		
				1.0403	1.0332	1.0280	1.0177	0.9889	0.995	0.0114
3	10	1	(2,18)	1.8209	2.6800	3.6174	4.6884	6.1803		
				0.0236	0.0261	0.0662	0.1248	0.1593		
				1.0333	1.0214	1.0177	1.0282	1.0278	0.997	0.0101
3	10	2	(2,9)	1.7974	2.7311	3.7479	4.8716	6.6691		
				0.0135	0.0451	0.0422	0.1202	0.1608		
				0.9290	0.9084	0.8805	0.8525	0.8314	1.154	0.0150
3	15	1	(2,33)	1.7494	2.5669	3.4117	4.3640	5.4466		
				0.0284	0.0419	0.0661	0.0916	0.1448		
				1.0348	1.0388	1.0386	1.0557	1.0253	0.985	0.0076

표 1. (계속됨)

l	n	g	df	.8	.9	.95	.975	.99	MSE	se
3	15	2	(2,24)	1.7877	2.6150	3.5081	4.4341	5.6500		
				0.0202	0.0346	0.0402	0.0461	0.1053		
				1.0379	1.0302	1.0309	1.0267	1.0065	1.001	0.0089
4	5	1	(3,4)	1.7077	2.6215	3.6867	4.8122	6.9952		
				0.0330	0.0546	0.0736	0.1099	0.2439		
				0.6873	0.6255	0.5593	0.4822	0.4190	1.384	0.0247
4	6	1	(3,8)	1.7783	2.6450	3.5715	4.5998	6.2094		
				0.0233	0.0437	0.0684	0.1263	0.2201		
				0.9114	0.9046	0.8784	0.8493	0.8180	1.118	0.0156
4	7	1	(3,12)	1.8067	2.5832	3.4392	4.3634	5.6050		
				0.0129	0.0159	0.0450	0.0783	0.1068		
				1.0014	0.9914	0.9853	0.9752	0.9416	1.051	0.0126
4	8	1	(3,16)	1.7568	2.4500	3.2127	4.0631	5.4155		
				0.0220	0.0374	0.0569	0.0728	0.1480		
				1.0122	0.9952	0.9919	0.9966	1.0233	1.019	0.0108
4	9	1	(3,20)	1.8132	2.5472	3.3854	4.1948	5.3657		
				0.0223	0.0305	0.0563	0.0707	0.1126		
				1.0692	1.0702	1.0926	1.0871	1.0866	1.000	0.0098
4	10	1	(3,24)	1.7353	2.4225	3.1719	3.9146	5.0369		
				0.0137	0.0299	0.0390	0.0770	0.1450		
				1.0391	1.0409	1.0542	1.0520	1.0676	0.990	0.0088
4	10	2	(3,12)	1.6571	2.3501	3.0998	3.9936	5.1635		
				0.0211	0.0282	0.0239	0.0787	0.1210		
				0.9185	0.9020	0.8881	0.8926	0.8674	1.147	0.0127
4	15	1	(3,44)	1.6984	2.3338	2.9328	3.6423	4.4215		
				0.0123	0.0176	0.0380	0.0745	0.0734		
				1.0531	1.0547	1.0413	1.0620	1.0378	0.984	0.0066
4	15	2	(3,32)	1.7120	2.3510	2.9600	3.6813	4.5639		
				0.0259	0.0415	0.0533	0.0550	0.0977		
				1.0450	1.0387	1.0203	1.0349	1.0234	1.000	0.0076

표 1. (계속됨)

l	n	g	df	.8	.9	.95	.975	.99	MSE	se
5	5	1	(4,5)	1.6021	2.3275	3.1452	4.1119	5.3336		
				0.0204	0.0495	0.0680	0.0726	0.1325		
				0.7153	0.6612	0.6058	0.5566	0.4682	1.372	0.0217
5	6	1	(4,10)	1.7266	2.4148	3.1399	4.0104	5.2309		
				0.0190	0.0336	0.0406	0.0446	0.1395		
				0.9442	0.9269	0.9028	0.8975	0.8726	1.123	0.0140
5	7	1	(4,15)	1.6776	2.3179	3.0111	3.6777	4.6373		
				0.0170	0.0249	0.0385	0.0564	0.1025		
				0.9808	0.9816	0.9855	0.9667	0.9477	1.052	0.0112
5	8	1	(4,20)	1.6887	2.3245	2.9467	3.5807	4.5530		
				0.0165	0.0303	0.0296	0.0509	0.0468		
				1.0208	1.0336	1.0281	1.0188	1.0276	1.018	0.0099
5	9	1	(4,25)	1.7244	2.3379	2.9692	3.6541	4.5899		
				0.0154	0.0218	0.0413	0.0589	0.1378		
				1.0634	1.0703	1.0763	1.0898	1.0987	0.999	0.0088
5	10	1	(4,30)	1.6700	2.2295	2.7682	3.3318	4.1331		
				0.0106	0.0239	0.0183	0.0391	0.0500		
				1.0437	1.0407	1.0292	1.0252	1.0287	0.996	0.0080
5	10	2	(4,15)	1.6158	2.1925	2.7849	3.4653	4.3388		
				0.0174	0.0235	0.0396	0.0556	0.0933		
				0.9447	0.9285	0.9114	0.9109	0.8867	1.147	0.0115
5	15	1	(4,55)	1.6315	2.1284	2.6370	3.1268	3.8028		
				0.0155	0.0237	0.0362	0.0421	0.0689		
				1.0509	1.0382	1.0383	1.0324	1.0331	0.986	0.0059
5	15	2	(4,40)	1.6969	2.2190	2.7547	3.2449	3.9358		
				0.0159	0.0222	0.0323	0.0464	0.0860		
				1.0783	1.0612	1.0571	1.0380	1.0281	0.994	0.0068
6	5	1	(5,6)	1.4894	2.0859	2.7639	3.5447	4.6106		
				0.0212	0.0315	0.0468	0.0866	0.1241		
				0.7176	0.6712	0.6300	0.5920	0.5272	1.384	0.0200

표 1. (계속됨)

l	*n*	*g*	*df*	.8	.9	.95	.975	.99	MSE	se
6	6	1	(5,12)	1.6298	2.2152	2.8550	3.4948	4.5381		
				0.0130	0.0287	0.0295	0.0654	0.1113		
				0.9365	0.9253	0.9192	0.8981	0.8961	1.126	0.0129
6	7	1	(5,18)	1.6614	2.2266	2.7733	3.4124	4.1120		
				0.0187	0.0276	0.0369	0.0606	0.1029		
				1.0124	1.0140	1.0002	1.0090	0.9680	1.048	0.0103
6	8	1	(5,24)	1.6183	2.1402	2.6281	3.1093	3.8361		
				0.0161	0.0176	0.0307	0.0419	0.0727		
				1.0157	1.0177	1.0029	0.9856	0.9848	1.024	0.0089
6	9	1	(5,30)	1.6386	2.1731	2.6471	3.1768	3.9515		
				0.0232	0.0160	0.0319	0.0456	0.0875		
				1.0468	1.0604	1.0448	1.0497	1.0683	1.001	0.0078
6	10	1	(5,36)	1.6075	2.0813	2.5731	3.0279	3.5945		
				0.0077	0.0135	0.0236	0.0367	0.0695		
				1.0391	1.0334	1.0387	1.0285	1.0056	0.994	0.0073
6	10	2	(5,18)	1.5721	2.0642	2.5627	3.0505	3.8112		
				0.0138	0.0198	0.0268	0.0339	0.0593		
				0.9580	0.9400	0.9242	0.9020	0.8972	1.142	0.0104
6	15	1	(5,66)	1.5785	2.0639	2.4874	2.9425	3.5362		
				0.0129	0.0181	0.0224	0.0340	0.0566		
				1.0482	1.0657	1.0568	1.0640	1.0690	0.984	0.0053
6	15	2	(5,48)	1.6155	2.1044	2.5572	2.9595	3.5313		
				0.0164	0.0257	0.0322	0.0392	0.0645		
				1.0599	1.0676	1.0617	1.0405	1.0310	1.001	0.0062
7	5	1	(6,7)	1.3784	1.8877	2.4482	3.0253	3.8318		
				0.0185	0.0192	0.0277	0.0305	0.0880		
				0.7042	0.6676	0.6333	0.5910	0.5328	1.395	0.0186
7	6	1	(6,14)	1.6072	2.1063	2.6282	3.1357	3.8282		
				0.0129	0.0198	0.0214	0.0390	0.0600		
				0.9603	0.9392	0.9229	0.8956	0.8591	1.116	0.0118

표 1. (계속됨)

l	*n*	*g*	*df*	.8	.9	.95	.975	.99	MSE	se
7	7	1	(6,21)	1.5934	2.0802	2.5771	3.0795	3.8585		
				0.0169	0.0231	0.0407	0.0417	0.0899		
				1.0036	1.0025	1.0017	0.9968	1.0123	1.052	0.0096
7	8	1	(6,28)	1.5990	2.0677	2.5162	2.9505	3.5886		
				0.0131	0.0147	0.0237	0.0344	0.0694		
				1.0342	1.0360	1.0290	1.0165	1.0173	1.018	0.0082
7	9	1	(6,35)	1.5667	2.0132	2.4519	2.8807	3.4089		
				0.0066	0.0110	0.0335	0.0525	0.0528		
				1.0297	1.0326	1.0338	1.0302	1.0122	1.001	0.0073
7	10	1	(6,42)	1.5954	2.0207	2.4690	2.8476	3.4086		
				0.0119	0.0205	0.0272	0.0285	0.0652		
				1.0599	1.0528	1.0624	1.0441	1.0437	0.991	0.0066
7	10	2	(6,21)	1.5278	1.9336	2.3719	2.8948	3.5642		
				0.0134	0.0183	0.0282	0.0418	0.0909		
				0.9623	0.9318	0.9220	0.9370	0.9351	1.151	0.0098
7	15	1	(6,77)	1.5239	1.9343	2.3213	2.7208	3.1875		
				0.0128	0.0180	0.0166	0.0234	0.0637		
				1.0373	1.0444	1.0462	1.0557	1.0466	0.985	0.0050
7	15	2	(6,56)	1.5898	2.0359	2.4421	2.8498	3.2500		
				0.0115	0.0179	0.0215	0.0251	0.0353		
				1.0704	1.0817	1.0779	1.0779	1.0341	0.994	0.0057

참고문헌

[1] 성내경(1994). 윈저화 χ^2의 양태에 대하여, 「응용통계연구」 제7권 2호, 1-7.

[2] Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion), *Journal of Royal Statisitcal Society,* series B, Vol. 2, 211-252.

[3] Dixon, W. J. and Tukey, J. W. (1968). Approximate behavior of the distribution of Winsorized t (Trimming/Winsorization 2), *Technometrics,* Vol. 10, 83-98.

[4] Gastwirth, J. L. and Cohen, M. L. (1970). Small sample behavior of some robust linear estimators of location, *Journal of the American Statistical Association,* Vol. 65, 946-973.

[5] Huber, P. J. (1981). *Robust statistics,* Wiley.

[6] Schrader, R. M. and Hettmansperger, T. P. (1980). Robust analysis of variance based upon a likelihood ratio criterion, *Biometrika,* Vol. 67, 93-101.

[7] Stigler, S. M. (1973). The asymptotic distribution of the trimmed mean, *Annals of Statistics,* Vol. 1, 472-477.

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

[9] Yuen, K. K. and Dixon, W. J. (1973). The approximate behaviour and performance of the two-sample trimmed t, *Biometrika,* Vol. 60, 369-374.

On a Robust Analysis of Variance Based on Winsorization[1)]

Nae Kyung Sung[2)]

Abstract

Based on Monte-Carlo simulation results we propose a robust analyis of variance procedure by utilizing trimmed mean and Winsorized variance. We deal with mainly the one-way classification case. We evaluate the empirical distribution of a pseudo-F statistic based on symmetrically Winsorized sum of squares when the population is normally distributed.

1) Research was supported by '93 Non-directed Research Fund, Korea Research Foundation.

2) Statistics Department, Ewha University, Seoul 120-750, KOREA.