

## Outlier Impact on the Power of Significance Test for Cronbach Alpha Reliability Coefficient

Yonghwan Um\*

\*Professor, Dept. of Industrial and Management Engineering, Sungkyul University, Anyang, Korea

### [Abstract]

In this paper, we studied the impact of outliers on the power of the significance tests for Cronbach alpha reliability coefficient. Four variables were varied: sample size, the number of items, the number of outliers and population Cronbach Alpha levels. We simulated data using multivariate normal distribution and used outliers sampled from uniform distribution. To test the significance of Cronbach Alpha Reliability, parametric approach(F statistic) and permutation method were used. Consequently, we observed that the powers of permutation test are equal to or greater than those of F test under all conditions, and also both F test and permutation test lose the power as the number of outliers increases, and that these effects of outliers on the power are enhanced for increasing population alpha levels.

▶ **Key words:** outliers, Cronbach alpha, reliability coefficient, power, permutation test

### [요 약]

본 논문은 크론바흐 알파와 신뢰계수의 유의성 검정에서 이상치가 검정력에 미치는 영향을 연구한 것이다. 표본 크기, 문항들의 수, 이상치의 수, 모집단의 크론바흐 알파 레벨의 네 개의 변수들에 변화를 주었다. 데이터 시뮬레이션을 위해 다변량 정규분포를 사용했고 균일분포로부터 이상치를 추출하여 사용했다. 크론바흐 알파와 신뢰도의 유의성 검정을 위해 모수적 검정(F 검정)과 퍼뮤테이션 검정을 사용하였다. 결과적으로 퍼뮤테이션 검정의 검정력은 F검정의 검정력 보다 크거나 같았고, 두 검정의 검정력은 모두 이상치의 수가 많아질수록 감소하였으며 이러한 이상치의 영향은 모집단의 알파 레벨이 증가할수록 크게 나타났다.

▶ **주제어:** 이상치, 크론바흐 알파, 신뢰계수, 검정력, 퍼뮤테이션 검정

- 
- First Author: Yonghwan Um, Corresponding Author: Yonghwan Um
  - Yonghwan Um (uyh@sungkyul.ac.kr), Dept. of Industrial and Management Engineering, Sungkyul University
  - Received: 2023. 04. 26, Revised: 2023. 05. 19, Accepted: 2023. 05. 22.

## I. Introduction

크론바흐 알파는 k개의 문항들로 구성된 검사에서 문항들간의 내적일관성을 측정하는 여러 방법들 중에서 가장 많이 주목을 받아온 신뢰도이다. Cronbach가 2004년에 보고한 바에 의하면 자신의 1951년 논문이 5900번 가량 인용되었고 매년 325번 정도로 SSCI에 인용된다고 하였다[1]. 크론바흐 알파 신뢰도는 문항들의 합 또는 평균과 각 문항들의 분산 그리고 문항들간의 공분산을 이용하여 쉽게 계산된다. 그러나 크론바흐 알파를 정확하게 추정하기 위해서는 문항들의 타우 동등(tau-equivalent)성, 그리고 정규성과 등분산성의 가정이 필요하다. 문항들의 타우 동등(tau-equivalent)성이란 모집단 수준에서 문항들의 공분산은 동일하나 문항들간의 분산이 동일하지 않은 것을 말하며, 분산까지 동일할 때는 평행(parallel)이라 하고 분산의 동일성과 공분산의 동일성이 모두 충족되지 않는 문항들을 동류(congeneric)라고 한다. 서로 평행한 문항들은 또한 타우 동등이라고 말할 수 있다. McDonald는 크론바흐 알파에 의해 내적 일관성을 측정하려면 문항들은 최소한 타우 동등(tau-equivalent)이어야 한다고 하였다[2]. 데이터의 정규성과 등분산성은 크론바흐 알파 추론에서 요구되는 두 번째 가정으로서 이 가정이 만족되지 않을 경우, Groggel, Wackerly와 Rao는 크론바흐 알파를 과대 추정하게 된다고 하였으며, Sheng은 시뮬레이션 연구를 통해 부정확한 크론바흐 알파 신뢰도의 정확성을 높이기 위해서는 표본의 크기를 크게 해야 한다고 하였다[3][4]. 그리고 Kristof와 van zyle et. al. 은 데이터가 정규분포를 따를 때 표본의 크론바흐 알파는 통계량으로서 가장 바람직한 성질인 불편성(unbiasedness)을 갖는다고 하였다[5][6].

본 논문은 정규분포의 데이터가 이상치(outlier)에 의해 오염(contaminated) 되었을 때 크론바흐 알파의 유의성 검정에서 이상치가 검정력(power)에 미치는 영향을 연구한 것이다. 이상치는 전체 데이터의 패턴에서 벗어난 값이므로 통계추론에 영향을 주기 때문에 이상치를 감지하고 처리하는 것은 실제 통계분석에서 매우 중요한 문제이다. 일반적으로 이상치가 존재하면 이상치를 삭제하거나 변형 또는 그대로 유지하는 방법을 사용한다. 이상치가 발생하는 이유가 분명하고 통제 가능할 경우에는 삭제하거나 변형할 수 있으나 그렇지 않을 경우에 제거하는 것은 이상치가 제공하는 유용한 정보를 상실할 수 있다. 많은 연구에서 이상치를 삭제하기 보다 포함(accomodation)하는 것이 더 신뢰할만한 연구방법이라

고 제안된 바 있다[7]. 크론바흐 알파의 추론에서 고려할 수 있는 이상치는 두가지로 존재한다. 첫째는 어느 응답자가 문항 전체 또는 일부에서 답변한 데이터가 이상치일 수 있거나(person outliers) 둘째는 각 문항에서 문항별로 서로 다른 분포를 가짐으로 발생할 수 있는 이상치(item outliers)이다. Liu와 Zumbo는 person outliers에 대해 이상치가 거리척도로 측정되는 문항들의 크론바흐 알파에 미치는 영향을 분석한 바 있으며, 이상치를 대칭적인 것과 비대칭적인 것으로 구분하여 대칭적인 이상치는 크론바흐 알파 추정에 영향을 주지 않는 반면에 비대칭적인 이상치는 크론바흐 알파를 과대 추정하고, 이러한 현상은 이상치의 비대칭성이 커지고 이상치가 많아질수록 심화된다고 보고하였다[8]. 또한 Liu, Wu와 Zumbo가 진행한 연구에서도 이상치에 의해 순서척도 문항에서의 크론바흐 알파가 과대 추정된다고 보고되었다[9]. 본 연구는 이상치가 들어있는 정규데이터에서 크론바흐 알파의 유의성검정을 기존의 모수적인 방법과 퍼뮤테이션 방법에 의해 실시하고 각각의 검정으로부터 검정력(power)을 계산한다. 이를 위해 이상치를 균일분포로부터 추출하여 정규 데이터에 삽입하고, 이상치를 특정 응답자의 모든 문항에서(person outliers) 발생하도록 하였다. 그리고 이상치의 수, 표본의 크기, 문항들의 수에 따라 이상치가 검정력에 미치는 영향을 분석함으로써 모수적 검정과 퍼뮤테이션 검정을 서로 비교한다. 퍼뮤테이션을 이용한 유의성검정은 표본의 크기가 작고 정규성 가정이 만족되지 않을 때 사용하는 대안적인 방법이며 이상치의 영향을 줄이기 위해 비모수적 통계방법과 함께 널리 사용된다. 퍼뮤테이션에 의한 유의성 검정은 실제 데이터를 이용하기 때문에 비모수적 검정에 비해 높은 검정력을 갖는 것으로 알려져 있다[10]. 퍼뮤테이션 검정에서는 데이터를 무작위로 재배치하여 얻어지는 데이터의 모든 배열을 이용하기 때문에 정확한 p값을 계산할 수 있으나 가능한 배열의 수가 많을 때는 임의로 몇 개만을 추출하는 Monte Carlo Sampling을 사용하여 검정의 p값을 구한다.

## II. Cronbach's Alpha and Hypothesis Testing

### 1. Parametric Test

크론바흐 알파는 관측값을 진점수와 오차점수의 합으로 표현한 CTT(Classic Test Theory) 모델로부터 유래되었다[11]. CTT 모델에서는 i번째 사람의 j번째 문항에

대한 관측점수를  $X_{ij}$ , ( $i = 1, 2, \dots, n; j = 1, 2, \dots, k$ )라 할 때  $X_{ij}$ 를 진점수( $t_{ij}$ )와 오차점수( $e_{ij}$ )의 선형함수로 나타낸다.

$$X_{ij} = t_{ij} + e_{ij} \quad (1)$$

여기서  $e_{ij}$ 는 무작위로 발생하는 오차이며 서로 독립이라고 가정한다. 문항별로 문항들의 진점수( $t_{ij}$ )가 동일하다고 할 때 타우 동등이라 하며 이 때 수식(1)은 수식(2)가 된다.

$$X_{ij} = t_i + e_{ij} \quad (2)$$

그리고  $k$ 개의 문항들의 값을 합산(즉,  $X = X_1 + X_2 + \dots + X_k$ )할 때  $X$ 의 분산은 진점수의 분산과 오차점수 분산의 합( $\sigma^2_x = \sigma^2_t + \sigma^2_e$ )이 되며, 타우 동등의 가정 하에서 모집단의 크론바흐 알파( $A$ )는 다음과 같이 정의된다.

$$A = \frac{k}{k-1} \left( \frac{\sigma^2_x - \sum_{i=1}^k \sigma^2_{x_i}}{\sigma^2_x} \right) \quad (3)$$

이 때 정규분포 가정하에서  $A$ 를 추정하는 최대우도 추정량(maximum likelihood estimator)  $\hat{A}$ 는 수식(4)으로 주어진다. 여기서 정규성 가정은  $t_i \sim N(\mu_t, \sigma^2_t)$ ,  $e_i \sim N(0, \sigma^2_e)$ 이다.

$$\hat{A} = \frac{k}{k-1} \left( \frac{s^2_x - \sum_{i=1}^k s^2_{x_i}}{s^2_x} \right) \quad (4)$$

크론바흐 알파의 유의성검정( $H_0 : A = A_0$ )에 대한 연구 결과는 Kristof와 Feldt에 의해 제시되었다[12]. 이들은 검정에서 사용할 검정통계량으로 자유도  $(n-1)$ 과  $(k-1)(n-1)$ 의 F분포를 따르는  $(1 - A_0)/(1 - \hat{A})$ 을 제안하였으며, 더불어 유의수준  $\alpha$ 에 대한  $100(1-\alpha)\%$  신뢰구간

$$[1 - (1 - \hat{\alpha})F_{(n-1), (k-1)(n-1), 1-\alpha/2},$$

$1 - (1 - \hat{\alpha})F_{(n-1), (k-1)(n-1), \alpha/2}]$ 과 기각역을 제시하였다. 여기서  $F_{(n-1), (k-1)(n-1), \alpha/2}$ 와

$F_{(n-1), (k-1)(n-1), 1-\alpha/2}$ 은 각각 자유도가  $n-1$ 과

$(k-1)(n-1)$ 인 F분포의  $100(\alpha/2)$  백분위수(percentile)와  $100(1-\alpha/2)$  백분위수(percentile)이다. 예를 들어  $n=31$ 이고  $k=16$ 개의 문항으로 구성된 검사에서  $\hat{A}=0.85$ 일 때, 유의수준  $\alpha=0.05$ 에서 양측검정 ( $H_0 : A = 0.8, H_1 : A \neq 0.8$ )의 결과는 검정통계량  $= (1-0.8)/(1-0.85)=1.33$ , 임계값은  $F_{30, 450, 0.025} = 0.55$ ,  $F_{30, 450, 0.975} = 1.60$  이므로 귀무가설을 기각하지 못하며 95% 신뢰구간은  $[0.76, 0.92]$ 이 된다. 마찬가지로 단측검정  $H_0 : A = 0.8, H_1 : A > 0.8$ 에서도 귀무가설을 기각하는 기각역은  $(1-0.8)/(1-\hat{A}) > F_{30, 450, 0.95} = 1.48$  이고 귀무가설을 기각하기 위해서는  $\hat{A} > 0.864$ 가 되어야 하므로 현재의 크론바흐 알파 계수  $\hat{A} = 0.85$ 은 귀무가설을 기각할만한 충분한 근거가 되지 못한다.

## 2. Permutation Test

퍼뮤테이션 검정은 모수적 검정(parametric test)에서 요구하는 모집단의 정규성 가정에 제한을 받지 않으며 데이터 값들의 독립성과 상호교환성(exchangeability)에 근거하여 데이터를 재배치하여 진행되는 방법이다 [13][14]. 퍼뮤테이션에 의한 검정은 모수적 검정처럼 검정의 임계값이 필요치 않는 장점이 있고, 데이터가 정규 분포일 때는 모수적 검정만큼의 검정력을 갖는다고 알려져 있다. 특히 이상치에 의해 오염되어 있는 데이터에 대해서도 뛰어난 검정능력을 보인 것으로 알려져 있다[15]. 퍼뮤테이션 검정을 위해 데이터를 재배치할 때 얻어지는 가능한 배열들은 매우 많을 수 있기 때문에 일부 배열만을 추출하는 Monte Carlo Sampling을 사용한다. 표본의 크기가  $n$ , 문항 수가  $k$ 인 데이터로부터 동일한 확률로 얻을 수 있는 가능한 배열의 수는  $(n!)^k$  [예를들어  $n = 20, k=4$ 일 때  $(20!)^4 \approx \infty$ ]의 매우 큰 값이므로 일부  $B$ 개를 무작위 추출하여 근사적인 검정을 실시한다.

퍼뮤테이션 검정은 다음의 단계로 진행되어 검정의  $p$  값을 결정한다.

- ① 크기가  $n$ 인 주어진 표본에서 통계량  $F_0$ 를 계산한다.
- ② 이 표본에서 각 문항별로 퍼뮤테이션을 진행하고, 재배치된 데이터(permuted data)에 대해  $F$ 를 계산하여  $F_0$ 와 비교한다.
- ③ 단계 ②를  $B$ 번 반복 시행한다.
- ④ 단측검정(upper-tail test)  $H_0 : A = A_0, H_a : A > A_0$ 에 대해서  $p$ 값을 계산한다.

$$p\text{값} = \frac{\text{(F}_0\text{보다 크거나 같은 F값들의 수)} / B}{\dots}$$

### III. Data Generation and Power of Tests

다음의 여러 조건하에서 가설  $H_0 : A = A_0$ 에 대한 단측검정(upper-tail test)을 실시하여 F검정과 퍼뮤테이션 검정의 검정력을 계산하였다.

- (1) 전체 표본의 크기 :  $n = 20, 25, 30$
- (2) 분포 형태 : 다변량 정규분포
- (3) 이상치의 분포 : 균일분포
- (4) 이상치의 수 :  $m = 1, 2, 3, 4$
- (5) 문항 수 :  $k = 4, 8, 10$

데이터를 생성하기 위해 다변량 정규분포  $N(\vec{0}, \Sigma)$ 로부터 크기가  $n - m$ 인 5000개의 임의표본들을 추출하였으며 문항들의 평행(parallel) 또는 타우동등(tau-equivalent) 조건에 맞는 공분산 행렬  $\Sigma$ 을 사용하였다. 문항 수  $k=4$  일 때  $\Sigma_p$ (평행 일 때)와  $\Sigma_t$ (타우 동등 일 때)는 다음과 같다.

$$\Sigma_p = \begin{pmatrix} 1 & \theta & \theta & \theta \\ \theta & 1 & \theta & \theta \\ \theta & \theta & 1 & \theta \\ \theta & \theta & \theta & 1 \end{pmatrix}, \quad \Sigma_t = \begin{pmatrix} 1 & \theta & \theta & \theta \\ \theta & 2 & \theta & \theta \\ \theta & \theta & 2 & \theta \\ \theta & \theta & \theta & 1 \end{pmatrix}$$

문항들이 평행이면  $\Sigma_p$ 는 대각선상의 분산성분이 1인  $k \times k$  공분산 행렬을 사용하였고, 문항들이 타우 동등이고  $k=4$ 일 때 대각선상의 각 분산성분이 (1, 2, 2, 1)이고

$k=8$ 이면 (1, 2, 3, 4, 4, 3, 2, 1),  $k=10$  일 때는 분산이 (1, 2, 3, 4, 5, 5, 4, 3, 2, 1)인  $\Sigma_t$ 를 각각 사용하였다. 여기서  $\theta$ 는 문항들 사이의 공분산이고  $\theta = 0.1, 0.4, 0.7$ 을 사용하였다. 문항들의 평행이고  $k=4$ 와 공분산  $\theta$ 에 대응되는 모집단의 크론바흐 알파 수식(3)에 의해  $A = 0.3, 0.7, 0.9$ 이 되고 문항들이 타우 동등일 때는 각각  $A = 0.2, 0.6, 0.8$ 이 된다. 이상치는 -10과 10 사이의 균일분포로부터  $m(=1, 2, 3, 4)$ 개가 추출되어 표본의 크기가  $n-m$ 인 정규데이터에 첨가되어 전체 표본의 크기는  $n$ 이 되게 하였다. 즉  $m$ 명의 응답자들의 모든 문항에서 이상치(person outliers)가 발생하도록 하였다. 예를 들어 그림 1은  $m=3$ 개의 이상치가 이변량 정규분포 데이터에 첨가된 크기  $n=20$ 인 데이터의 산점도이다.

한편 유의성 검정의 검정력을 계산하기 위해 주어진 파라미터를 갖는 분포로부터 데이터(정규데이터와 이상치)를 생성하였다. 시뮬레이션을 통해 총 5000개의 임의 표본들을 추출하여 F검정의 검정력을 구하고, 각 표본에 대해  $B = 5000$ 개의 임의 배열들을 생성하는 퍼뮤테이션을 진행함으로써 퍼뮤테이션 검정의 검정력을 계산하였다. 퍼뮤테이션 검정의 검정력은 전체 5000개의 표본들 중에서 유의수준  $\alpha=0.05$ 보다 작거나 같은 퍼뮤테이션  $p$ 값을 갖는 표본들의 비율로서 계산된다. 퍼뮤테이션 검정과 F검정을 비교하기 위해 검정력의 차이를 수식(5)로 나타내었다.

$$\begin{aligned} &\text{검정력의 차이} (\Delta (\%)) \\ &= \frac{\text{퍼뮤테이션검정의 검정력} - \text{F검정의 검정력}}{\text{F검정의 검정력}} \times 100 \end{aligned} \quad (5)$$

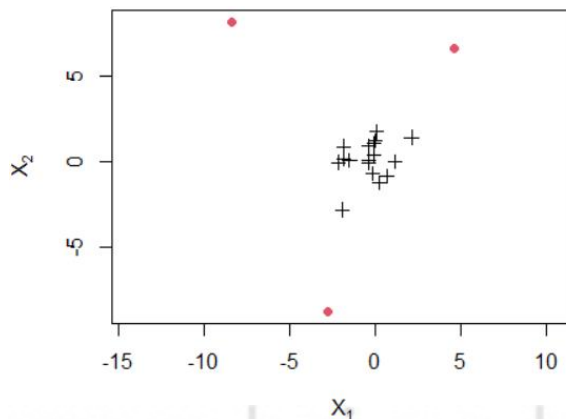


Fig. 1. Scatter plot of normal data with outliers

### IV. Numerical Results

Table 1부터 Table 6까지는 시뮬레이션 데이터를 이용해 크론바흐 알파에 대한 유의성 검정을 실시한 결과이다. 표본의 크기(n), 문항의 수(k), 이상치의 수(m), 문항들의 평행 또는 타우 동등에 따라 F검정과 퍼뮤테이션 검정의 검정력을 산출하였으며 데이터 생성, p값 계산, 검정력 계산 등의 모든 계산을 R 프로그램에 의해 진행하였다. Table 1, Table 2와 Table 3은 문항들이 평행일 때 그리고 Table 4, Table 5와 Table 6은 문항들이 타우 동등일 때의 결과로서 문항들 사이의 공분산( $\theta=0.1, 0.4, 0.7$ )에 따라 검정력을 비교하고 있으며 이 중에서 한 예로서 Table 1의 검정력을 시각적으로 볼 수 있도록 그림으로 나타내었다(Fig. 2). 또한 Fig. 3은  $\theta=0.7$ 일 때 문항수(k=4, 8, 10)별로 검정력을 비교하기 위해 플롯한 것이다.

표에서 보듯이 표본의 크기(n), 문항수(k), 문항간 공분산( $\theta$ ) (즉, 모집단의 알파계수)가 커질수록 퍼뮤테이션 검정과 F검정의 검정력은 모두 증가하였고 이런 현상은 문항들이 평행일 때와 타우 동등일 때 동일하게 나타났다. 예를 들어 문항들이 평행이고  $\theta=0.7, k=4, m=1$ 일 때 n이 20에서 30으로 변함에 따라 F검정의 검정력은 0.592  $\rightarrow$  0.711  $\rightarrow$  0.811로 증가하고 퍼뮤테이션 검정의 검정력은 0.612  $\rightarrow$  0.730  $\rightarrow$  0.830으로 증가하였으며, n=20, k=4, m=1일 때  $\theta$ 가 0.1에서 0.7로 변함에 따라 F검정의 검정력은 0.256  $\rightarrow$  0.409  $\rightarrow$  0.592로 증가하고 퍼뮤테이션 검정의 검정력은 0.261  $\rightarrow$  0.418  $\rightarrow$  0.612로 증가하였다(Table 1). 또한 문항들이 평행이고 n=20,  $\theta=0.7, m=1$ 일 때 k가 4부터 10으로 변함에 따라 F검정의 검정력은

0.592  $\rightarrow$  0.942  $\rightarrow$  0.973로 증가하였고 퍼뮤테이션 검정의 검정력은 0.612  $\rightarrow$  0.946  $\rightarrow$  0.976으로 증가하였다. 한편 이상치의 수(m)에 따른 검정력의 변화를 볼 때, 주어진 n, k,  $\theta$ 의 값에서 이상치의 수가 증가할수록 F검정과 퍼뮤테이션 검정의 검정력은 감소하였다. 예를 들어 Table 1에서 n=20, k=4,  $\theta=0.7$ 일 때 m이 1에서 4로 변함에 따라 F검정의 검정력은 0.592  $\rightarrow$  0.412  $\rightarrow$  0.347  $\rightarrow$  0.286로 감소하였으며 퍼뮤테이션 검정의 검정력은 0.612  $\rightarrow$  0.424  $\rightarrow$  0.450  $\rightarrow$  0.289로 감소하였다. 이상치의 수(m)가 증가하여 전체 표본(n) 중 이상치가 차지하는 비율(m/n)이 커짐에 따라(예를 들어 n=20 일 때, 이상치의 비율이 각각 1/20, 2/20, 3/20, 4/20) 검정력이 감소한 것이다. 그리고 이상치의 수에 따라 검정력의 감소하는 정도가  $\theta$ 가 커짐에 따라 크게 나타나는 것을 볼 수 있다. 예를 들어 n=20, k=4에서,  $\theta=0.1$ 일 때는 F검정의 검정력은 0.256  $\rightarrow$  0.191(퍼뮤테이션 검정, 0.261  $\rightarrow$  0.191),  $\theta=0.4$ 일 때 F검정의 검정력은 0.409  $\rightarrow$  0.235(퍼뮤테이션 검정, 0.418  $\rightarrow$  0.241),  $\theta=0.7$ 인 경우 F검정의 검정력은 0.592  $\rightarrow$  0.286(퍼뮤테이션 검정, 0.612  $\rightarrow$  0.289)로 감소하여 큰 값의  $\theta$ 에서 검정력의 감소폭이 크게 나타났다(Table 1, Fig. 2). 이와 같은 현상은 모든 표에서 일관되게 나타났으며, 동시에 모든 조건하에서 두 검정 사이의 검정력의 차이  $\Delta(\%) \geq 0$ 이고 최고  $\Delta(\%) = 3.49\%$ 를 보임으로써 퍼뮤테이션 검정의 검정력이 F검정의 검정력 보다 일정하게 높게 유지되고 있음을 볼 수 있다.

Table 1. Powers when data are parallel and k=4

n	m	$\theta = 0.1$			$\theta = 0.4$			$\theta = 0.7$		
		F test	perm. test	$\Delta(\%)$	F test	perm. test	$\Delta(\%)$	F test	perm. test	$\Delta(\%)$
20	1	0.256	0.261	1.95	0.409	0.418	2.20	0.592	0.612	3.38
	2	0.238	0.243	2.10	0.313	0.320	2.24	0.412	0.424	2.91
	3	0.217	0.220	1.38	0.276	0.281	1.81	0.347	0.450	0.86
	4	0.191	0.196	2.62	0.235	0.241	2.55	0.286	0.289	1.05
25	1	0.258	0.263	1.94	0.479	0.491	2.51	0.711	0.730	2.67
	2	0.251	0.255	1.59	0.378	0.388	2.65	0.527	0.538	2.09
	3	0.246	0.252	2.44	0.326	0.334	2.45	0.421	0.427	1.43
	4	0.223	0.227	1.79	0.292	0.297	1.71	0.357	0.362	1.40
30	1	0.274	0.277	1.09	0.540	0.550	1.85	0.811	0.830	2.34
	2	0.273	0.280	2.56	0.423	0.428	1.18	0.606	0.618	1.98
	3	0.254	0.259	1.97	0.360	0.370	2.78	0.484	0.493	1.86
	4	0.238	0.244	2.52	0.325	0.331	1.85	0.432	0.440	1.86

Table 2. Powers when data are parallel and k=8

n	m	$\theta = 0.1$			$\theta = 0.4$			$\theta = 0.7$		
		F test	perm. test	$\Delta(\%)$	F test	perm. test	$\Delta(\%)$	F test	perm. test	$\Delta(\%)$
20	1	0.249	0.251	0.80	0.665	0.684	2.86	0.942	0.946	0.42
	2	0.226	0.230	1.77	0.436	0.440	0.92	0.691	0.700	1.30
	3	0.205	0.207	0.98	0.352	0.356	1.14	0.549	0.554	0.91
	4	0.198	0.200	1.01	0.294	0.296	0.68	0.437	0.440	0.69
25	1	0.297	0.303	2.02	0.817	0.827	1.22	0.986	0.987	0.10
	2	0.252	0.255	1.19	0.567	0.574	1.23	0.869	0.874	0.58
	3	0.230	0.233	1.30	0.431	0.436	1.16	0.686	0.693	1.02
	4	0.229	0.231	0.87	0.391	0.392	0.26	0.591	0.595	0.68
30	1	0.320	0.324	1.25	0.916	0.921	0.55	0.996	0.997	0.10
	2	0.283	0.287	1.41	0.682	0.691	1.32	0.940	0.943	0.32
	3	0.263	0.265	0.76	0.539	0.545	1.11	0.829	0.832	0.36
	4	0.255	0.257	0.78	0.465	0.468	0.65	0.711	0.715	0.56

Table 3. Powers when data are parallel and k=10

n	m	$\theta = 0.1$			$\theta = 0.4$			$\theta = 0.7$		
		F test	perm. test	$\Delta(\%)$	F test	perm. test	$\Delta(\%)$	F test	perm. test	$\Delta(\%)$
20	1	0.275	0.281	2.18	0.794	0.800	0.76	0.973	0.976	0.31
	2	0.246	0.249	1.22	0.533	0.541	1.50	0.833	0.838	0.60
	3	0.224	0.227	1.34	0.419	0.423	0.95	0.658	0.661	0.46
	4	0.205	0.206	0.49	0.336	0.340	1.19	0.525	0.528	0.57
25	1	0.310	0.316	1.94	0.913	0.916	0.33	0.995	0.995	0.00
	2	0.268	0.272	1.49	0.674	0.678	0.59	0.940	0.941	0.11
	3	0.250	0.252	0.80	0.524	0.527	0.57	0.812	0.816	0.49
	4	0.228	0.231	1.32	0.437	0.440	0.69	0.694	0.697	0.43
30	1	0.360	0.365	1.39	0.974	0.975	0.10	0.999	0.999	0.00
	2	0.294	0.297	1.02	0.807	0.810	0.37	0.988	0.989	0.10
	3	0.283	0.285	0.71	0.677	0.680	0.44	0.939	0.942	0.21
	4	0.267	0.270	1.12	0.550	0.552	0.36	0.831	0.832	0.12

Table 4. Powers when data are tau-equivalent and k=4

n	m	$\theta = 0.1$			$\theta = 0.4$			$\theta = 0.7$		
		F test	perm. test	$\Delta(\%)$	F test	perm. test	$\Delta(\%)$	F test	perm. test	$\Delta(\%)$
20	1	0.229	0.234	2.18	0.364	0.370	1.65	0.543	0.553	1.84
	2	0.222	0.227	2.25	0.298	0.304	2.01	0.383	0.394	2.87
	3	0.205	0.208	1.46	0.259	0.264	1.93	0.328	0.333	1.52
	4	0.185	0.187	1.08	0.224	0.228	1.79	0.272	0.279	2.57
25	1	0.219	0.224	2.28	0.423	0.431	1.89	0.654	0.667	1.99
	2	0.232	0.239	3.02	0.344	0.356	3.49	0.494	0.502	1.62
	3	0.234	0.238	1.71	0.313	0.316	0.96	0.397	0.404	0.76
	4	0.208	0.213	2.40	0.278	0.281	1.08	0.346	0.348	0.58
30	1	0.244	0.248	1.64	0.483	0.492	1.86	0.751	0.766	2.00
	2	0.247	0.253	2.43	0.391	0.396	1.28	0.563	0.572	1.60
	3	0.236	0.241	2.12	0.340	0.343	0.88	0.461	0.469	1.74
	4	0.224	0.228	1.79	0.310	0.315	1.61	0.415	0.422	1.69

Table 5. Powers when data are tau-equivalent and k=8

n	m	$\theta = 0.1$			$\theta = 0.4$			$\theta = 0.7$		
		F test	perm. test	$\Delta(\%)$	F test	perm. test	$\Delta(\%)$	F test	perm. test	$\Delta(\%)$
20	1	0.186	0.188	1.08	0.501	0.507	1.20	0.825	0.828	0.36
	2	0.184	0.187	1.63	0.353	0.360	1.98	0.587	0.592	0.85
	3	0.180	0.181	0.56	0.297	0.302	1.68	0.476	0.480	0.84
	4	0.174	0.176	1.15	0.260	0.262	0.77	0.387	0.388	0.26
25	1	0.214	0.218	1.87	0.640	0.646	0.94	0.907	0.912	0.55
	2	0.201	0.201	0.00	0.457	0.460	0.66	0.753	0.760	0.93
	3	0.190	0.190	0.00	0.366	0.372	1.64	0.594	0.597	0.51
	4	0.191	0.193	1.05	0.337	0.339	0.59	0.529	0.533	0.76
30	1	0.234	0.236	0.85	0.737	0.741	0.54	0.966	0.968	0.21
	2	0.222	0.225	1.35	0.552	0.560	1.45	0.855	0.858	0.35
	3	0.219	0.224	2.28	0.446	0.452	1.35	0.733	0.740	0.95
	4	0.215	0.217	0.93	0.398	0.400	0.50	0.637	0.639	0.31

Table 6. Powers when data are tau-equivalent and k=10

n	m	$\theta = 0.1$			$\theta = 0.4$			$\theta = 0.7$		
		F test	perm. test	$\Delta(\%)$	F test	perm. test	$\Delta(\%)$	F test	perm. test	$\Delta(\%)$
20	1	0.195	0.197	1.03	0.571	0.577	1.05	0.865	0.868	0.35
	2	0.189	0.191	1.06	0.425	0.426	0.24	0.697	0.700	0.43
	3	0.175	0.176	0.57	0.339	0.341	0.59	0.563	0.566	0.53
	4	0.169	0.170	0.59	0.282	0.285	1.06	0.453	0.455	0.44
25	1	0.207	0.210	1.45	0.701	0.707	0.85	0.943	0.945	0.21
	2	0.203	0.206	1.48	0.521	0.525	0.77	0.831	0.834	0.36
	3	0.195	0.196	0.51	0.424	0.427	0.71	0.703	0.707	0.57
	4	0.186	0.187	0.54	0.362	0.363	0.28	0.594	0.595	0.17
30	1	0.234	0.241	2.99	0.807	0.811	0.50	0.977	0.978	0.10
	2	0.217	0.220	1.38	0.644	0.645	0.16	0.914	0.915	0.11
	3	0.212	0.216	1.89	0.536	0.543	1.31	0.842	0.845	0.36
	4	0.217	0.219	0.92	0.458	0.461	0.66	0.734	0.736	0.27

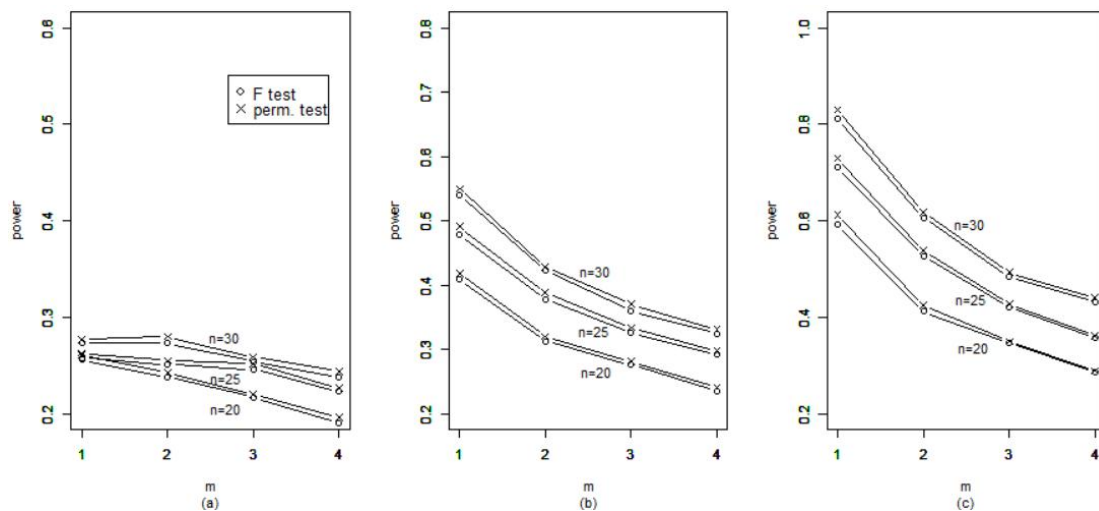


Fig. 2. Powers when data are parallel and k=4 with (a)  $\theta = 0.1$  (b)  $\theta = 0.4$  (c)  $\theta = 0.7$

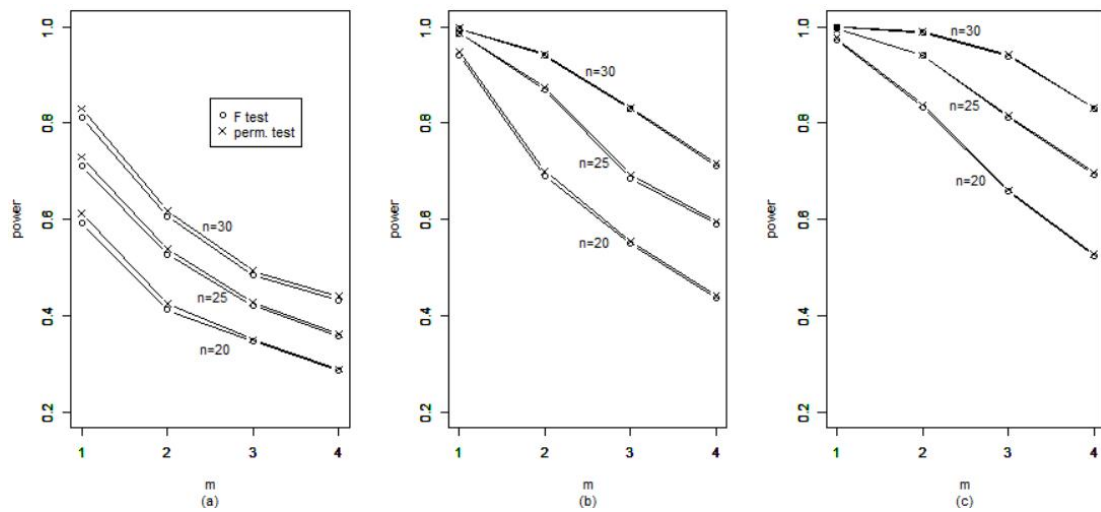


Fig. 3. Powers when data are parallel and  $\theta=0.7$  with (a)  $k=4$  (b)  $k=8$  (c)  $k=10$

## V. Conclusion

본 논문은 이상치에 의해 오염된 정규데이터에 대해 크론바흐 알파의 유의성 검정을 실시할 때 이상치가 검정력에 미치는 영향을 연구한 것이다. 이를 위해 F검정과 퍼뮤테이션 검정을 사용하였고 표본의 크기( $n$ ), 문항수( $k$ ), 문항간 공분산( $\theta$ ), 이상치의 수( $m$ )에 따른 검정력을 산출하고 비교하였다. 다변량 정규분포로부터 추출된 데이터에 균일분포로부터 추출된 이상치를 첨가하여 데이터를 생성하였으며 이렇게 생성된 5000개의 임의표본들과 각 표본 당  $B=5000$ 번의 퍼뮤테이션에 의해 검정력을 계산하였다.

결과적으로 F검정과 퍼뮤테이션 검정의 검정력은  $n$ ,  $k$ ,  $\theta$ (결과적으로 크론바흐 알파 수준)가 커짐에 따라 증가하였고, 이상치의 수( $m$ ) 즉 전체 표본 중 이상치가 차지하는 비율( $m/n$ )이 커짐에 따라 검정력은 감소하였다. 이는 이상치와 검정력에 관한 다른 연구( $t$  검정, ANOVA에서의 이상치 영향) 와도 동일하게 나타난 결과이다 [16][17]. 또한  $\theta$ 가 커짐에 따라 이상치의 수가 증가할수록 검정력은 더 급격하게 감소하였다. 더불어 퍼뮤테이션 검정의 검정력은 F검정의 검정력 보다 크거나 같게 나타남으로써 이상치에 의해 오염된 정규데이터에서 크론바흐 알파의 유의성 검정을 할 때 퍼뮤테이션 검정이 F검정 보다 우수하다고 말할 수 있다.

크론바흐 알파는 사회과학, 행동과학, 의학 등 다양한 분야에서 널리 사용되는 신뢰도 지수이므로 크론바흐 알파에 대한 통계추정과 검정은 늘 지속적인 연구과제가 되어 왔다. 이상치가 있을 때 크론바흐 알파는 과대 추정된다는 앞선 연구는 있었으나, 이상치가 유의성검정에 미치

는 영향에 대한 연구결과가 없는 상황에서 본 연구의 의미를 찾을 수 있다. 다만 본 연구결과가 거리척도의 반응(response) 데이터에 한정되어 있으므로 차후에는 순서척도(ordinal scale)의 데이터를 사용하여 문항의 수, 이상치의 비율, 이상치의 종류 등에 따른 검정력을 비교하는 추가적인 연구가 필요하다고 생각한다. 또한 본 논문의 연구방법을 크론바흐 알파 이외의 다른 신뢰계수(반분 검사 신뢰도, 재검사 신뢰도, 유사 크론바흐 알파 등)들에도 적용함으로써 신뢰계수들에 미치는 이상치의 영향 정도를 비교하는 연구도 의미 있을 것으로 생각한다.

## REFERENCES

- [1] L. J. Cronbach, My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64, 391-418, 2004. <https://doi.org/10.1177/0013164404266386>.
- [2] R. P. McDonald, *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum, 1999.
- [3] D. J. Groggel, D. D. Wackerly and P. V. Rao, Nonparametric estimation in one-way random effects models. *Communication in Statistics-Communication and Simulation*, 17, 887-903, 1988. <https://doi.org/10.1080/03610918808812702>.
- [4] Y. Sheng and Z. Sheng, Is coefficient alpha robust to non-normal Data?, *Frontiers in Psychology*, 3, Article 34, 2012. <https://doi.org/10.3389/fpsyg.2012.00034>.
- [5] W. Kristof, The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts, *Psychometrika*, 28, 221-238, 1963. <https://doi.org/10.1002/j.2333-8504.1962.tb00298.x>.

- [6] J. M. van Zyl, H. Neudecker, D. G. Nel D. G., On the distribution of the maximum likelihood estimator for Cronbach's alpha. *Psychometrika* 65, 271-280, 2000. <https://doi.org/10.1007/BF02296146>.
- [7] H. Liao, Y. Li, and G. Brooks, Outlier impact and accommodation methods: Multiple comparisons of type I error rates. *Journal of Modern Applied Statistical Methods*, 15(1), 452-471, 2016. DOI: 10.22237/jmasm/1462076520.
- [8] Y. Liu and B. D. Zumbo, The impact of outliers on cronbach's coefficient alpha estimate of reliability: Visual analog scales. *Educational and Psychological Measurement*, 67, 620-634, 2007. <https://doi.org/10.1177/0013164406296976>.
- [9] Y. Liu, A. D. Wu and B. D. Zumbo, The impact of outliers on cronbach's coefficient alpha estimate of reliability: Ordinal/rating scale item response, 70(1), 5-21, 2010. <https://doi.org/10.1177/0013164409344548>.
- [10] J. Ludbrook and H. Dudley, Why permutation tests are superior to t and F tests in biomedical research. *The American Statistician*, 52(2), 127-133, 1998. <https://doi.org/10.1080/00031305.1998.10480551>.
- [11] F. M. Lord and M. R. Novick, *Statistical Theories of Mental Test Scores*. Reading: Addison-Wesley, 1968.
- [12] L. S. Feldt, D. J. Woodruff, and F. A. Salih, Statistical inference for coefficient alpha, *Applied Psychological Measurement*, 11, 93-103, 1987. <https://doi.org/10.1177/014662168701100107>.
- [13] P. Good, *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. New York, NY: Springer-Verlag, 1994.
- [14] P. Good, Extensions of the concept of exchangeability and their applications. *Journal of Modern Applied Statistical Methods*, 1(2), 243-247, 2002. DOI: 10.22237/jmasm/103611024
- [15] E. S. Edgington, *Randomization Tests*. (3rd ed). New York, NY: Marcel Dekker, 1995.
- [16] F. Branco, A. Oliveira, T. Oliveira and C. E. Minder, The Impact of Outliers on the Power of the Randomization Test for Two Independent Groups, *Numerical Analysis and Applied Mathematics*, 1389, 1545-1548, 2011. <https://doi.org/10.1063/1.3637923>
- [17] J. H. Gleason, Comparative power of the ANOVA, approximate randomization ANOVA, and Kruskal-Wallis test, *Wane State University*, 2013.

## Authors



Yonghwan Um received the B.S. and M.S. in Chemistry from Yonsei University, Korea, in 1981, 1983, respectively, M.S. in Biostatistics from Emory University, in 1990 and Ph.D. in Statistics from University of

Florida, U.S.A. in 1995. Dr. Um joined the faculty of the Department of Computational Statistics at Sungkyul University, Anyang, Korea, in 1996. He is currently a Professor in the Department of Industrial and Management Engineering, Sungkyul University. He is interested in reliability measure, data-mining, statistical inference.