

Nonparametric Method using Placement in an Analysis of a Covariance Model

Dong Min Hwang^a, Dongjae Kim^{1, a}

^aDepartment of Biostatistics, Catholic University of Korea

Abstract

Various methods control the influence of a covariate on a response variable. These methods are analysis of covariance(ANCOVA), RANK ANCOVA, ANOVA of (covariate-adjusted) residuals, and Kruskal-Wallis tests on residuals. Covariate-adjusted residuals are obtained from the overall regression line fit to the entire data set that ignore the treatment levels or factors. It is demonstrated that the methods on covariate-adjusted residuals are only appropriate when the regression lines are parallel and covariate means are equal for all treatments. In this paper, we proposed the new nonparametric method on the ANCOVA model, as applying joint placement in a one-way layout on residuals as described in Chung and Kim (2007). A Monte Carlo simulation study is adapted to compare the power of the proposed procedure with those of the previous procedure.

Keywords: Analysis of covariance, nonparametric, joint placement, residual.

1. 서론

세 군 이상의 처리효과의 차이를 검정하기 위해서는 분산분석(Analysis of Variance)을 실시하게 된다. 분산분석이 필요한 실험계획에서, 반응변수의 모평균에 영향을 줄 수 있는 또 다른 독립변수가 있다면 단순히 처리 간의 차이뿐만이 아니라 그 독립변수의 영향도 함께 고려해야 한다. 이러한 독립변수가 연속형일 때 이를 공변량이라고 하고, 공변량이 존재하는 경우 공분산분석(Analysis of Covariance)을 고려해야 한다. 공변량은 기저측정치, 환경요인 외에도 다양한 형태로 존재하며, 이러한 유의한 공변량이 존재할 때 이를 고려해주지 않으면 통계분석의 오차가 커지고, 이로 인해 실험의 정확도가 떨어질 수 있다. 따라서 처리간 반응변수의 모평균이 처리에 공통적으로 작용하는 어떠한 공변량에 의해 영향을 받을 때에는 공분산 분석을 선택해야 한다. 이렇게 공변량을 함께 고려하면 이를 고려하지 않은 분산분석보다 추정의 정도를 높일 수 있다 (Lee 등, 2005).

공변량이 있는 경우에 처리군간 효과를 비교하는 방법에는, 모수적인 방법으로 공분산분석이 있고, 비모수적인 방법으로는 Quade (1967)에 의해 고안된 RANK ANCOVA 방법이 있다. ANCOVA 분석은 회귀분석과 분산분석이 혼합된 형태로, 각 처리 안에서 비교하려는 변수에 유의한 영향을 주는 공변량을 회귀모형으로 제어한 후 분산분석을 하는 방법이고, RANK ANCOVA 방법은 공변량과 반응변수 각각의 순위를 구하여, 회귀분석과 분산분석을 차례로 적용시켜 처리간 효과차이를 검정하는 비모수적 방법이다. 이처럼 순위를 사용하는 비모수적 검정방법은 정규성 가정이나 등분산성 가정이 위배되는 경우 모수적인 방법에 비해 효과적일 수 있고, 모집단의 분포형태에 따라서 서로 다른 분석방법의 선택이 더 효율적임을 알 수 있다 (Conover와 Iman, 1982).

¹ Corresponding author: Professor, Department of Biostatistics, Catholic University of Korea, Banpo-Dong, Seocho-Gu, Seoul 137-701, Korea. E-mail: djkim@catholic.ac.kr

한편, 종속변수와 공변량의 관계가 선형임을 가정하고 회귀분석을 따로 실시하여 보정된 관측값을 구한 후, 기존의 관측값과의 차이인 잔차를 이용하여 분산분석과 Kruskal-Wallis 검정법을 적용한 방법이 있다. 회귀분석과 분산분석을 함께 사용한다는 점에서, 잔차를 이용한 분산분석은 공분산분석과 같은 검정으로 혼동될 수 있으나 이들은 엄연히 다른 분석이며, 자료의 분포형태에 따라 효율성이 다르게 나타난다 (Maxwell와 Delaney, 1985). 잔차를 이용한 방법들과 공분산분석을 이용한 방법의 검정력을 비교해 보면, 정규성 및 등분산(Homogeneity of Variance)이라는 가정 하에서 위의 방법들이 공분산분석과 유사한 검정력(Power)을 보인다는 것이 알려져 있다 (Ceyhan와 Goad, 2009).

Orban과 Wolfe (1982)는 두 처리간 효과의 차이를 검정하기 위해 위치(Placement)를 사용한 비모수적 검정법을 제안하였다. 이 방법은 두 처리 중 어느 한 처리에 대한 상대적 위치정보를 이용하여 처리효과를 검정하는 방법이다. 또한 Chung과 Kim (2007)은, 위치라는 비모수적 도구를 일원배치모형으로 확장하여 새로운 비모수적 방법론을 제안하였다. 모의실험의 결과로 나타나는 이 검정방법의 검정력은, 표본의 크기가 큰 처리의 처리효과가 표본의 크기가 작은 처리의 처리효과보다 클 때 Kruskal-Wallis가 제안한 검정방법보다 효율적이다.

공분산분석에도 이 위치라는 비모수적 도구를 사용하여, 특정한 상황에서 기존의 검정방법들보다 효율적인 검정방법을 제안할 수 있다. 하지만 공분산분석이 필요한 데이터에서는 종속변수에 공변량을 보정해주는 과정이 필요하기 때문에 위치를 바로 공분산분석 모형에 적용할 수 없다. 여기서 잔차를 이용함으로써 위치를 이용한 비모수적인 방법을 공분산분석이 필요한 데이터에 적용할 수 있다. 본 논문에서는 비모수적인 분석방법 중 위치를 이용한 방법을 잔차에 적용하여 공분산분석 모형의 새로운 비모수적 검정방법을 제안하였다. 귀무가설에 대한 일반적인 대립가설과 그에 대한 검정방법을 제안하고, 기존 방법들의 설명과 함께 모의실험을 통하여 공분산분석과, RANK ANCOVA, 잔차를 이용한 분산분석과 Kruskal-Wallis 검정, 그리고 본 논문에서 제안한 검정방법들의 검정력을 비교하였다. 또한 실제 자료에의 적용 사례를 들어서 그 유용성을 보였다.

2. 방법

하나의 공변량(X)이 있고 처리의 수가 k 개인 공분산분석 모형은 다음과 같다.

Treatment	X	Y	\hat{Y}	R
1	x_{11}	y_{11}	\hat{y}_{11}	$R_{11} = y_{11} - \hat{y}_{11}$
	x_{12}	y_{12}	\hat{y}_{12}	$R_{12} = y_{12} - \hat{y}_{12}$
	\vdots	\vdots	\vdots	\vdots
	x_{1n_1}	y_{1n_1}	\hat{y}_{1n_1}	$R_{1n_1} = y_{1n_1} - \hat{y}_{1n_1}$
\vdots	\vdots	\vdots	\vdots	\vdots
k	x_{k1}	y_{k1}	\hat{y}_{k1}	$R_{k1} = y_{k1} - \hat{y}_{k1}$
	x_{k2}	y_{k2}	\hat{y}_{k2}	$R_{k2} = y_{k2} - \hat{y}_{k2}$
	\vdots	\vdots	\vdots	\vdots
	x_{kn_k}	y_{kn_k}	\hat{y}_{kn_k}	$R_{kn_k} = y_{kn_k} - \hat{y}_{kn_k}$

$$Y_{ij} = \mu + \tau_i + \beta X_{ij} + \epsilon_{ij} \quad (i = 1, 2, \dots, k; j = 1, 2, \dots, n_i),$$

여기서 μ 는 미지의 전체 평균을 나타내고 τ_i 는 i 번째 처리의 효과, β 는 회귀계수, X_{ij} 는 공변량을 나타낸다. 또한 ϵ_{ij} 는 오차항으로 동일한 연속분포를 따르는 서로 독립인 확률변수로 가정한다. $Y_{i1}, Y_{i2}, \dots, Y_{in_i}; i = 1, 2, \dots, k$ 를 i 번째 처리에서 추출한 확률표본이라 하고, 전체표본의 크기는 $N = \sum_{i=1}^k n_i$ 이라 하자. 또한 각 처리 안에서 공변량의 효과는 같다고 가정한다.

각 처리의 효과가 모두 동일하다는 귀무가설과 그에 대한 일반적인 대립가설은 다음과 같다.

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_k \quad \text{vs.} \quad H_1 : \text{모든 } \tau_i \text{들이 같지는 않다.}$$

2.1. RANK ANCOVA

Quade (1967)에 의해 제안된 순위를 이용한 공분산분석 방법은, 전체표본에서 공변량(X)의 순위를 행렬 C 로 나타내고 종속변수(Y)의 순위를 행렬 U 로 나타내어, 보정된 순위행렬

$$\hat{U} = C(C'C)^{-1}C'U$$

을 구한다. 이 때, $Z = U - \hat{U} = (Z_{11}, Z_{12}, \dots, Z_{1n_1}, \dots, Z_{k1}, \dots, Z_{kn_k})$ 는 전체표본에서 구한 순위에 대한 잔차 행렬이다. 처리간에 효과차이가 없다는 귀무가설 H_0 를 검정하기 위한 검정통계량은

$$VR = \frac{(N - k) \sum_{i=1}^k \left(\sum_{j=1}^{n_i} Z_{ij} \right)^2 / n_i}{(k - 1) \left[\sum_{i=1}^k \sum_{j=1}^{n_i} Z_{ij}^2 - \sum_{i=1}^k \left(\sum_{j=1}^{n_i} Z_{ij} \right)^2 / n_i \right]}$$

이다. 검정통계량 VR 는 근사적으로 자유도 $(k - 1, N - k)$ 인 F 분포에 따른다. 이 때 기각역은 $F_\alpha(k - 1, N - k)$ 이며, $F_\alpha(k - 1, N - k)$ 은 자유도가 $(k - 1, N - k)$ 인 F -분포의 상위 100α 백분위수이다.

2.2. 잔차를 이용한 방법

하나의 자료에서 변수 X (공변량)와 Y (반응변수)의 단순선형회귀모형을 통해 예측값 $\hat{Y} = X(X'X)^{-1}X'Y$ 을 구할 수 있다. 여기서 $R = Y - \hat{Y} = (R_{11}, R_{12}, \dots, R_{1n_1}, \dots, R_{k1}, \dots, R_{kn_k})$ 로, $R_{ij}; i = 1, 2, \dots, k, j = 1, 2, \dots, n_i$ 는 i 번째 처리에서 j 번째 관측치의 잔차를 나타낸다. 이 잔차는 공변량을 보정해 준 관측값의 의미를 가지며, 그 자체를 이용하여 기존의 처리간 효과차이를 검정하는 검정법들을 적용시킬 수 있다 (Ceyhan과 Goad, 2009).

2.2.1. 잔차를 이용한 분산분석법

잔차를 이용한 분산분석은 처리간 잔차 평균의 차이가 있는지를 검정하는 것과 같다. 따라서 잔차를 종속변수로 하는 분산분석을 실시한다. 분산분석에서는 총변동을 처리간 변동과 처리내 변동으로 분해하여 두 변동의 상대적 크기를 비교함으로써 처리효과와 유의성을 검정한다. 가설을 검정하기 위해 총 제곱합(총변동)을 잔차의 전체 평균 \bar{R} 와 각 처리에서의 잔차평균 $\bar{R}_i; i = 1, 2, \dots, k$ 을 이용하여 다음과 같은 형태로 분해한다.

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (R_{ij} - \bar{R})^2 = \sum_{i=1}^k n_i (\bar{R}_i - \bar{R})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (R_{ij} - \bar{R}_i)^2,$$

여기서 $SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (R_{ij} - \bar{R})^2$ 는 총제곱합(total sum of squares), $SSA = \sum_{i=1}^k n_i (\bar{R}_i - \bar{R})^2$ 는 처리제곱합(treatment sum of squares)이고, $SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (R_{ij} - \bar{R}_i)^2$ 는 오차제곱합(error sum of squares)이다. 처리제곱합과 오차제곱합을 각각 그들의 자유도 $k - 1, N - k$ 로 나눈 평균제곱합을 각각 $MSA = SSA/(k - 1), MSE = SSE/(N - k)$ 라고 했을 때, 귀무가설 H_0 를 검정하기 위한 검정통계량은

$$F_R = \frac{MSA}{MSE}$$

이다. 분산분석법에서 F_R 은 자유도 $(k-1, N-k)$ 인 F -분포를 따른다는 것을 이용하여 검정한다. 이때 기각역은 $F_\alpha(k-1, N-k)$ 이며, $F_\alpha(k-1, N-k)$ 은 자유도가 $(k-1, N-k)$ 인 F -분포의 상위 100α 백분위수이다.

2.2.2. 잔차를 이용한 Kruskal-Wallis 검정법

K-W 검정은 Mann-Whitney U 검정의 확장 모형으로, 그룹이 2개인 경우에 두 검정은 같다. 이 검정법은 혼합표본에서 R_{ij} 의 순위인 I_{ij} 를 사용하게 된다. Kruskal-Wallis 검정법의 검정통계량 H 는 다음과 같다

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{I_i^2}{n_i} - 3(N+1).$$

이때 귀무가설 H_0 를 검정하기 위한 검정은

$$\text{Reject } H_0 \quad \text{if } H \geq h(\alpha, k, (n_1, \dots, n_k))$$

이며, $h(\alpha, k, (n_1, \dots, n_k))$ 는 K-W 통계량 분포표에서 찾을 수 있다. 또한 각 처리에서의 표본크기가 충분히 클 때, H 는 근사적으로 자유도 $k-1$ 인 χ^2 분포에 따른다. 이때, 기각역은 $\chi_{k-1, \alpha}^2$ 이며, $\chi_{k-1, \alpha}^2$ 은 자유도가 $k-1$ 인 카이제곱분포의 상위 100α 백분위수이다.

2.2.3. 잔차를 이용한 결합위치 검정법

Chung과 Kim (2007)에 의해 제안된 비모수적 도구인 결합 위치(Joint Placement) V_{ij} 를 잔차에 적용하면 다음과 같다.

$$V_{ij} = \frac{1}{N - n_i} \sum_{h=1}^k \sum_{\substack{s=1 \\ h \neq i}}^{n_k} \chi(R_{hs}, R_{ij}), \quad \text{여기서 } \chi(x, y) = \begin{cases} 1, & x \leq y, \\ 0, & \text{otherwise.} \end{cases}$$

V_{ij} 는 전체 혼합표본에서 i 번째 처리의 관측값을 제외한 후, R_{ij} 보다 작거나 같은 관측값의 개수를 이용하여 구한 확률변수이다. 또한 V_{ij} 의 처리별 평균 \bar{V}_i 와 전체평균 \bar{V} 은 다음과 같다.

$$\begin{aligned} \bar{V}_i &= \frac{\sum_{j=1}^{n_i} V_{ij}}{n_i}, \quad i = 1, 2, \dots, k \\ \bar{V} &= \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} V_{ij}}{N} = \frac{\sum_{i=1}^k n_i \bar{V}_i}{N} = \frac{\sum_{i=1}^k V_i}{N}, \end{aligned}$$

여기서 각 군에서의 \bar{V}_i 와 \bar{V} 의 차이를 이용한 검정통계량 JP 는 다음과 같다.

$$JP = \sum_{i=1}^k n_i (\bar{V}_i - \bar{V})^2 = \sum_{i=1}^k \frac{V_i^2}{n_i} - N\bar{V}^2.$$

일원배치모형에서 Y 를 변수로 처리간 효과차이를 분석하는것과는 달리, 이 모형에서는 회귀분석을 이용해 구한 R 에 대해 검정을 실시한다. 유의수준 α 에서 기각역은

$$JP \geq jp(\alpha, k, (n_1, \dots, n_k))$$

형태가 되고, 여기서 $jp(\alpha, k, (n_1, \dots, n_k))$ 는 $P_0\{JP \geq jp(\alpha, k, (n_1, \dots, n_k))\} = \alpha$ 를 만족하는 상수로써, 처리수가 3~4개이고 처리별 표본수가 5개 이하인 경우의 기각역은 Chung과 Kim (2007)에서 찾아볼 수 있다. 그리고 각 처리별 표본의 크기가 동일할 때, 결합 위치를 이용한 검정방법과 Kruskal과 Wallis가 제안한 검정법은 같은 검정방법이 된다.

3. 예제

나병환자의 치료법을 연구하기 위하여 B, C 두 항생제의 약효를 대조군 A와 함께 비교하기 위한 실험에서, 이 실험집단들에 대하여 치료 전과 일정기간 치료 후에 환자의 몸에서 leprosy bacilli(나병간균)을 측정하여 다음과 같은 자료를 얻었다. 이와 같은 실험에서 치료 후의 세균수는 치료 전의 세균수에 크게 의존할 것이므로 치료전 세균수를 공변량으로 포함시켜 분석하는 것이 바람직하다 (Song과 Jo, 2004).

Group	X(치료전)	Y(치료후)	잔차(R_{ij})	잔차 순위(I_{ij})	결합 위치	V_{ij}
A	12	20	9.46	12	7	1
	13	10	-1.37	8	6	0.86
	11	18	8.29	11	7	1
	9	5	-3.06	5	4	0.57
	21	23	5.01	10	7	1
B	14	8	-4.20	2	0	0
	19	11	-5.34	1	0	0
	11	8	-1.71	7	4	0.44
C	18	18	2.49	9	5	0.625
	19	14	-2.34	6	3	0.375
	8	4	-3.23	4	2	0.25
	15	9	-4.02	3	2	0.25

결합 위치의 그룹별 평균과 전체평균은,

$$\bar{V}_A = \frac{4.43}{5} = 0.886, \quad \bar{V}_B = \frac{0.44}{3} = 0.148, \quad \bar{V}_C = \frac{1.50}{4} = 0.375, \quad \bar{V}_{..} = \frac{6.37}{12} = 0.531$$

로 나타나며, 이를 이용하여 구한 검정통계량

$$JP = 5(0.886 - 0.531)^2 + 3(0.148 - 0.531)^2 + 4(0.375 - 0.531)^2 = 0.63 + 0.44 + 0.10 = 1.17$$

이다. 각 그룹의 종속변수 평균은 $\bar{Y}_A = 15.20, \bar{Y}_B = 9.00, \bar{Y}_C = 11.25$ 로 나타난다. 기존의 검정 방법들은 SAS 9.2를 이용하여 검정통계량과 P-value를 구하였다. 각 검정방법들의 검정통계량과 P-value를 아래에 표로 정리하였다.

검정 방법	검정 통계량	P-value
분산분석법	1.05	0.390
공분산분석법	3.26	0.092
RANK ANCOVA	3.52	0.074
잔차를 이용한 분산분석법	3.43	0.078
잔차를 이용한 K-W 검정법	5.43	0.066
잔차를 이용한 결합위치 검정법	1.17	0.046

여기서, 공변량(X)를 고려하지 않은 분산분석의 경우 p값이 0.390으로 그룹 간 차이가 있다는 귀무가설을 기각하지 못하지만, 기존의 다른 공분산분석 같은 경우 p값이 매우 낮아지는 모습을 보인다. 이는 공변량이 종속변수에 주는 영향을 고려했을 때 더욱 유의한 결과를 얻을 수 있다는 것을 말해준다.

하지만 모든 방법이 유의수준 .05에서 귀무가설을 기각하지 못하는데, 결합위치를 이용한 검정법에서 p 값이 0.046으로 유의한 결과를 나타낸다.

4. 모의실험의 계획 및 결과

공변량이 존재하는 자료의 일반적인 분석방법은 모수적 검정방법으로 공분산분석(ANCOVA)을 사용하고, 비모수적인 방법으로는 Quade (1967)가 제안한 RANK ANCOVA를 사용한다. 여기에 단순 선형회귀모형을 통해 구한 잔차에 대해 분산분석법과 Kruskal-Wallis 검정법을 사용하는 방법과 결합 위치(Joint Placement)를 이용한 검정방법의 검정력을 비교해 보았다. 실행 프로그램은 SAS를 사용하였다. 모집단의 분포로는 정규분포, 지수분포, Cauchy분포, 이중지수분포를 채택하였으며, 정규분포의 난수생성은 RANNOR 함수, 지수분포의 난수생성은 RANEXP 함수, Cauchy분포의 난수생성은 RANCAU 함수를 이용하였다. 이중지수분포는 RANUNI 함수를 이용하여 역변환 방법으로 난수를 생성하였다. 회귀계수 β 는 1로 고정하고, 공변량(X)은 한 개인 경우만 고려하였고, 연속형 변수인 것을 가정하여 RANNOR 함수를 이용하여 추출하였다.

처리의 수는 3개인 경우를 선택하였고, 각 처리별로 세 군의 표본의 크기가 모두 다를 경우와 두 표본의 크기는 같고 한 표본의 크기만 큰 경우를 고려하였다. 각 처리의 표본의 크기가 모두 같은 경우는 결합 위치를 사용한 검정법과 Kruskal-Wallis 검정법이 같은 방법이기 때문에 모의실험에서 제외하였다. 각 처리의 표본의 크기가 모두 같은 경우의 모의실험은 Ceyhan과 Goad (2009)에서 찾아볼 수 있다. 각 처리별 효과를 나타내는 τ 들은 표본의 수가 모두 다른 경우에 8가지 조합을, 나머지 경우에는 7가지 조합의 실험을 실시하였다. 이러한 조건에서 각 검정통계량들이 기각역에 포함되는지를 판단하는 과정을 10,000번 반복하는 Monte Carlo Study를 사용하였다. 표본의 수는 처리별로 3~5개의 개체가 있는 소표본인 경우와 10~20개의 개체가 존재하는 대표본인 경우를 다루었고, 소표본인 경우에는 Chung과 Kim (2007)이 제시한 기각역을 사용하고, 대표본인 경우 정렬을 통해 확률값을 계산하고 분포를 구하여 기각역을 사용하였다. 각 분포에서의 방법별 검정력의 비교결과를 소표본인 경우는 Table 1에, 대표본인 경우는 Table 2에 정리하였다.

각 처리의 효과가 모두 동일할 경우에 유의수준이 0.05를 만족하는지를 살펴보면, 공분산분석법과 잔차를 이용한 분산분석법의 유의수준은 Cauchy 분포에서 0.02에 근사한 값들을 얻었다. 또한 지수분포나 이중지수분포에서도 0.04에 근사한 값들을 얻는 것을 볼 수 있는데, 이는 정규성을 만족하지 않을 경우에 제 1종 오류를 제어하기 힘들다는 것을 보여준다. 그에 반해 비모수적인 검정법들은 모든 분포에서 제 1종 오류를 0.05근방에서 일정한 수준으로 제어할 수 있었다.

모의실험의 전체적인 결과를 살펴보면 분산분석법과 잔차를 이용한 분산분석법의 검정력은 모든 상황에서 비슷한 결과를 보인다. 그리고 표본크기가 큰 처리의 τ 값이 표본의 크기가 작은 처리의 τ 값보다 클 때 결합 위치를 사용한 검정방법의 검정력이 다른 검정방법들에 비해 높게 나타난다. 이러한 차이는 각 집단의 표본수가 모두 다르고 분포가 정규성을 만족하지 않을 때 더 크게 나타난다. 특히 지수분포와 이중지수분포에서는 그 차이가 더욱 크게 나타난다.

정규분포에서는 기존의 모수적 검정방법인 공분산분석방법이 대체로 높은 검정력을 보여주지만, 표본크기가 큰 처리의 τ 값이 가장 클 때 결합 위치를 사용한 검정방법의 검정력이 높았다. 하지만 표본수가 증가할수록 그 차이는 줄어들고, 전체 표본수가 50개가 넘어가면서는 모수적 검정방법들의 검정력이 τ 값에 상관없이 높아지는 것을 볼 수 있었다. 그리고 꼬리가 두터운 Cauchy분포와 이중지수분포, 비대칭성을 띄는 지수분포에서는 비모수적인 검정방법들의 검정력이 높았다. Cauchy 분포에서는 RANK ANCOVA 방법의 검정력이 가장 높게 나타났고, 지수분포와 이중지수분포에서는 모수적 검정방법과 결합위치를 이용한 방법의 검정력이 높았다.

Table 1: Monte Carlo power estimates: $\alpha = 0.05, k = 3, n_i \leq 5$

Dist	n_1	n_2	n_3	τ_1	τ_2	τ_3	F	VR	F_R	H	JP
Normal	3	4	5	0	0	0	.048	.049	.049	.036	.049
				0	0	1	.206	.187	.209	.163	.236
				0	1	0	.197	.175	.197	.149	.196
				1	0	0	.173	.155	.171	.122	.108
				.8	0	1.6	.369	.320	.365	.299	.408
				1	0	2	.537	.465	.522	.439	.572
				1.6	0	.8	.304	.256	.302	.230	.218
	2	0	1	.447	.374	.442	.344	.316			
	4	4	5	0	0	0	.054	.053	.053	.044	.050
				0	0	1	.235	.209	.237	.199	.246
				0	1	0	.211	.181	.212	.176	.183
				0	.8	1.6	.398	.341	.395	.341	.386
				0	1	2	.573	.492	.561	.491	.549
				0	1.6	.8	.353	.303	.354	.297	.277
0				2	1	.515	.439	.508	.432	.400	
Exponential	3	4	5	0	0	0	.044	.049	.043	.037	.049
				0	0	1	.267	.249	.269	.264	.352
				0	1	0	.254	.231	.255	.206	.269
				1	0	0	.227	.202	.229	.153	.141
				.8	0	1.6	.468	.433	.457	.442	.549
				1	0	2	.625	.572	.599	.570	.674
				1.6	0	.8	.389	.343	.382	.346	.337
	2	0	1	.536	.468	.523	.470	.448			
	4	4	5	0	0	0	.040	.043	.039	.042	.049
				0	0	1	.292	.271	.287	.299	.363
				0	1	0	.271	.243	.269	.239	.244
				0	.8	1.6	.482	.438	.470	.481	.530
				0	1	2	.639	.590	.625	.613	.660
				0	1.6	.8	.443	.398	.436	.437	.412
0				2	1	.601	.541	.584	.563	.536	
Cauchy	3	4	5	0	0	0	.025	.051	.024	.036	.048
				0	0	1	.042	.087	.039	.060	.084
				0	1	0	.042	.087	.042	.062	.081
				1	0	0	.039	.080	.038	.055	.063
				.8	0	1.6	.061	.125	.058	.085	.122
				1	0	2	.083	.166	.084	.113	.161
				1.6	0	.8	.053	.106	.052	.072	.081
	2	0	1	.070	.138	.070	.094	.100			
	4	4	5	0	0	0	.024	.055	.023	.041	.049
				0	0	1	.044	.098	.044	.073	.089
				0	1	0	.039	.088	.041	.071	.079
				0	.8	1.6	.059	.133	.060	.098	.117
				0	1	2	.083	.176	.083	.129	.148
				0	1.6	.8	.052	.121	.054	.089	.092
0				2	1	.072	.159	.076	.117	.117	
Double Exponential	3	4	5	0	0	0	.043	.048	.044	.038	.050
				0	0	1	.150	.156	.154	.134	.187
				0	1	0	.133	.141	.134	.120	.153
				1	0	0	.116	.120	.119	.104	.101
				.8	0	1.6	.247	.246	.249	.227	.307
				1	0	2	.359	.347	.355	.312	.416
				1.6	0	.8	.195	.197	.197	.173	.168
	2	0	1	.284	.282	.287	.247	.234			
	4	4	5	0	0	0	.038	.047	.038	.041	.048
				0	0	1	.143	.155	.146	.140	.174
				0	1	0	.135	.145	.139	.132	.139
				0	.8	1.6	.239	.248	.242	.224	.264
				0	1	2	.358	.353	.357	.331	.377
				0	1.6	.8	.220	.224	.220	.211	.199
0				2	1	.324	.321	.323	.303	.285	

** F = ANCOVA, VR = RANK ANCOVA, F_R = ANOVA of Residual, H = K-W of Residual, JP = Joint Placement of Residual

Table 2: Monte Carlo power estimates : $\alpha = 0.05, k = 3, n_i \geq 10$

Dist	n_1	n_2	n_3	τ_1	τ_2	τ_3	F	VR	F_R	H	JP
Normal	10	15	20	0	0	0	.048	.047	.049	.045	.048
				0	0	1	.817	.765	.815	.788	.840
				0	1	0	.771	.714	.769	.738	.757
				1	0	0	.663	.596	.664	.621	.550
				.8	0	1.6	.985	.972	.985	.977	.986
				1	0	2	.999	.999	.999	.999	.999
				1.6	0	.8	.929	.896	.929	.913	.881
	2	0	1	.993	.982	.992	.987	.979			
	15	15	20	0	0	0	.053	.054	.053	.050	.048
				0	0	1	.849	.801	.847	.826	.843
				0	1	0	.807	.755	.806	.775	.754
				0	.8	1.6	.985	.970	.984	.978	.979
				0	1	2	.999	.998	.999	.999	.999
				0	1.6	.8	.972	.954	.971	.960	.946
0				2	1	.998	.996	.998	.998	.995	
Exponential	10	15	20	0	0	0	.043	.046	.043	.045	.052
				0	0	1	.834	.872	.832	.826	.966
				0	1	0	.789	.834	.786	.935	.944
				1	0	0	.674	.724	.493	.870	.805
				.8	0	1.6	.974	.984	.973	.993	.996
				1	0	2	.996	.998	.996	.999	.999
				1.6	0	.8	.912	.940	.911	.978	.964
	2	0	1	.981	.989	.981	.997	.994			
	15	15	20	0	0	0	.046	.052	.047	.049	.046
				0	0	1	.855	.901	.854	.969	.976
				0	1	0	.803	.857	.803	.952	.943
				0	.8	1.6	.974	.987	.974	.996	.997
				0	1	2	.997	.999	.997	.999	.999
				0	1.6	.8	.957	.976	.957	.994	.989
0				2	1	.991	.997	.991	.999	.999	
Cauchy	10	15	20	0	0	0	.027	.051	.028	.048	.052
				0	0	1	.058	.244	.059	.209	.247
				0	1	0	.058	.223	.059	.185	.203
				1	0	0	.047	.181	.046	.155	.135
				.8	0	1.6	.090	.413	.090	.342	.395
				1	0	2	.130	.563	.129	.458	.514
				1.6	0	.8	.070	.319	.069	.263	.234
	2	0	1	.095	.448	.096	.360	.326			
	15	15	20	0	0	0	.019	.047	.019	.047	.045
				0	0	1	.051	.254	.052	.218	.231
				0	1	0	.044	.226	.045	.190	.178
				0	.8	1.6	.078	.421	.080	.349	.359
				0	1	2	.114	.580	.113	.471	.479
				0	1.6	.8	.069	.375	.068	.309	.279
0				2	1	.099	.528	.100	.428	.390	
Double Exponential	10	15	20	0	0	0	.046	.049	.046	.048	.053
				0	0	1	.522	.561	.522	.618	.683
				0	1	0	.488	.520	.488	.576	.604
				1	0	0	.393	.417	.393	.461	.401
				.8	0	1.6	.814	.844	.811	.878	.916
				1	0	2	.945	.956	.944	.972	.984
				1.6	0	.8	.676	.709	.677	.762	.710
	2	0	1	.859	.880	.856	.910	.876			
	15	15	20	0	0	0	.050	.052	.051	.050	.047
				0	0	1	.556	.595	.555	.660	.683
				0	1	0	.513	.546	.510	.603	.576
				0	.8	1.6	.828	.854	.827	.895	.900
				0	1	2	.950	.960	.950	.975	.977
				0	1.6	.8	.766	.798	.766	.848	.814
0				2	1	.918	.936	.916	.954	.939	

** F = ANCOVA, VR = RANK ANCOVA, F_R = ANOVA of Residual, H = K-W of Residual, JP = Joint Placement of Residual

대표본 모의실험 결과를 보면 소표본 모의실험과 같은 상황에서 전체적으로 검정력이 높게 나타난다. 전체적인 결과는 소표본 모의실험의 결과와 비슷한 모습을 보이며, 여기에서 잔차를 사용한 방법들의 유의수준이 0.05 근방으로 더욱 잘 제어되는 것을 볼 수 있다. 그리고 소표본에서와는 달리 지수분포와 이중지수 분포에서 Kruskal-Wallis 방법과 결합위치를 이용한 방법의 검정력 높게 나타난다. 여기서 Kruskal-Wallis 방법이 효율적 검정방법이 되기 위해선 다른 비모수적 방법들보다 많은 표본 수를 필요로 한다고 볼 수 있다. Cauchy분포를 제외하고는, 분포가 정규성을 만족하는지에 상관없이 표본크기가 가장 큰 처리의 값이 가장 클 때, 결합 위치를 사용한 검정방법이 효율적인 것으로 나타난다.

5. 고찰

본 논문에서는 처리간 효과차이를 비교하는 검정에서 반응변수에 영향을 주는 공변량의 효과를 제어하는 다양한 방법들을 비교해 보았다. 고려한 방법에는 공분산분석, RANK ANCOVA, 전체 데이터에 대해 회귀분석을 통해 구한 잔차에 ANOVA, Kruskal-Wallis, 그리고 결합 위치를 적용한 방법이다. 실험 설계시 분포, 표본 수, 처리별 효과크기에서 특정한 상황이 기대되는 경우에는 상황에 맞는 검정방법을 사용하는 것이 더 효율적인 실험이 된다. 특히 분포가 정규성을 만족하지 않을 것으로 예상되는 경우 비모수적인 검정방법의 사용이 실험결과에 더 큰 영향을 미치게 된다.

모의실험 결과를 보면 본 논문에서 제안한 검정방법은 작은 표본의 크기를 가지는 처리의 처리효과가 가장 작고 가장 큰 표본의 크기를 가진 처리의 처리효과가 가장 클 때 가장 효율적이라고 말할 수 있다. 이러한 경향이 나오게 된 원인은 결합위치의 처리별 평균과 전체평균의 차이가 표본의 크기에 따라 달라진다는 점에서 찾을 수 있다. 각 처리별 표본의 크기의 차이가 관측값의 결합 위치에 영향을 미치기 때문이다. 또한 꼬리가 두터운 분포나 비대칭성이 예상되는 분포에서 다른 모수적인 방법들보다 효율적이고, 상황에 따라서 다른 비모수적인 방법들보다도 효율적이다. 하지만 Chung과 Kim (2007)의 논문에서 제시된 기각역의 표본수와 처리수준이 제한적이기 때문에 실질적으로 분석에 적용하는 데는 한계가 있다. 따라서 결합위치를 사용한 검정통계량의 분포와 점근적인 성질을 밝혀낸다면 공분산분석이 필요한 모형에서 유용한 비모수적인 검정방법이 될 것이다.

References

- Ceyhan, E. and Goad, C. L. (2009). A comparison of analysis of covariate-adjusted Residuals and analysis of covariance, *Communications in Statistics - Simulation and Computation*, **38**, 2019–2038.
- Chung, T. and Kim, D. (2007). Nonparametric method using placement in one-way layout, *The Korean Communications in Statistics*, **3**, 552–560.
- Conover, W. J. and Iman, R. L. (1982). Analysis of covariance using the rank transformation, *Biometrics*, **38**, 715–724.
- Lee, J. W., Park, M. R. and Yoo, H. N. (2005). *Statistical Method for Biotechnology Research*, Freecademy, Korea.
- Maxwell, S. E. and Delaney, H. D. (1985). Anova of residuals and ancova: Correcting an illusion by using model comparisons and graphs, *Journal of Educational Statistics*, **10**, 197–209.
- Orban, J. and Wolfe, D. A. (1982). A class of distribution-free two-sample tests based on placement, *Journal of the American Statistical Association*, **77**, 666–672.
- Quade, D. (1967). Rank analysis of covariance, *Journal of the American Statistical Association*, **62**, 1187–1220.
- Song, M. S. and Jo, S. S. (2004). *Statistical Data Analysis using SAS*, Freecademy, Korea.