Türkiye Klinikleri Biyoistatistik Dergisi
Turkiye Klinikleri Journal of Biostatistics Turkiye Klinikleri J Biostat. 2021;13(1):50-6

ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/biostatic.2020-79326

Non-parametric Randomization-based Analysis of Covariance Method in Repeated Clinical Categorical

Tekrarlanan Klinik Kategorik Veride Non-parametrik Randomizasyon Tabanlı Kovaryans Analizi Metodu

Şirin ÇETİN[a], S Kenan KÖSE[b], Erhan KARAASLAN[c], Ayşe ÜLGEN[d]

[a]Department of Biostatistics, Tokat Gaziosmanpaşa University Faculty of Medicine, Tokat, TURKEY
[b]Department of Biostatistics, Ankara University Faculty of Medicine, Ankara, TURKEY
[c]Department of Pediatrics, Tokat Gaziosmanpaşa University Faculty of Medicine, Tokat, TURKEY
[d]Girne American University, Faculty of Medicine, Kyrenia, CYPRUS

ABSTRACT Objective: Covariance analysis (ANCOVA) is a method used in biomedical and health research, but when the assumption of normality is not satisfied or the dependent variable is bivariate or on ordinal scale, various procedures are presented in the literature for covariance analysis. If assumptions are not satisfied and parametric methods are used, Type I error increases and the power of the test decreases. To overcome these issues, the researcher needs to look for an alternative approach to the analysis of covariance. For non-parametric ANCOVA, many methods are presented in the literature that can be applied to different types of data. **Material and Methods:** In our study, by analyzing various non-parametric covariance methods; the analysis of repeated categorical data obtained from the clinic is to be analyzed with the non-parametric randomization-based analysis of covariance (NPANCOVA) method. **Results:** The application of the method was performed on repeated clinical data obtained from 5 consecutive visits from pediatric patients diagnosed with Crimean-Congo hemorrhagic fever disease in the pediatric health and diseases service. **Conclusion:** As a result, we can state that the NPANCOVA method can be used in clinical analysis due to its many superior advantages in the analysis of categorically repeated non-parametric data.

Keywords: Non-parametric; non-parametric randomization-based analysis of covariance; non-parametric methods

ÖZET Amaç: Kovaryans analizi (ANCOVA), biyomedikal ve sağlık araştırmalarında kullanılan bir yöntemdir, fakat normallik varsayımı sağlanmadığında veya bağımlı değişkenin 2 değerli veya sıralı değişken olması durumunda literatürde kovaryans analizi için çeşitli prosedürler sunulmuştur. Varsayımlar sağlanmadığı hâlde, eğer parametrik yöntemler kullanılırsa Tip I hatanın artmasına ve testin gücünün azalmasına neden olmaktadır. Parametrik olmayan ANCOVA için literatürde farklı türden verilere uygulanabilecek pek çok yöntem sunulmuştur. **Gereç ve Yöntemler:** Çalışmamızda, parametrik olmayan çeşitli kovaryans analizi yöntemleri incelenerek, klinikten elde edilen tekrarlanan kategorik verilerinin analizinin non-parametrik randomizasyon tabanlı kovaryans analizi (NPANCOVA) metoduyla yapılması amaçlanmıştır. **Bulgular:** Metodun uygulaması ise çocuk sağlığı ve hastalıkları servisinde yatan Kırım-Kongo kanamalı ateşi hastalığı tanısı olan çocuk hastalardan art arda 5 vizitden elde edilen, tekrarlanan klinik veriler üzerinde yapılmıştır. **Sonuç:** Sonuç olarak, non-parametrik randomizasyon tabanlı kovaryans analizi metodunu kategorik tekrarlanan parametrik olmayan verilerin analizinde, pek çok üstün avantajları sebebiyle klinik analizlerde kullanılabileceğini ifade edebiliriz.

Anahtar kelimeler: Non-parametrik; non-parametrik randomizasyon tabanlı kovaryans analizi; parametrik olmayan yöntemler

The word non-parametric was first coined by Wofowitz in 1942 though specific non-parametric tests can be traced to medivial time. Non-parametric methods have been developed as an alternative for parametric methods. While parametric methods are based on a certain distribution assumptions and variance; non-parametric methods are not based on a specific distribution and the concept of variance. In addition, although non-parametric methods do not have any assumptions about the population, the data may have been obtained

Correspondence: Şirin ÇETİN
Department of Biostatistics, Tokat Gaziosmanpaşa University Faculty of Medicine, Tokat, TURKEY/TÜRKİYE
E-mail: cetinsirin55@gmail.com

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

Received: 30 Sep 2020 ***Received in revised form:*** 17 Dec 2020 ***Accepted:*** 17 Dec 2020 ***Available online:*** 29 Apr 2021

by using ratio, interval nominal and ordinal scales. Parametric methods are the methods that take into account many assumptions about the population. It is posible to determine whether to use parametric or non-parametric methods by examining whether the assumptions are provided. Non-parametric statistical methods have been developed as an alternative to parametric statistical methods to be used when one or more assumptions about a parametric test are not satisfied. It is problematic to use parametric statistical methods when data are not-distributed by specific known distribution such as normal distribution. In this case, two approaches have emerged. The first is to convert the data into a form close to a given known distribution such as the normal distribution (using transformations, e.g., of square root, arcsin etc). The second is to use a method that does not take into account the distribution, methods based on rank numbers (ranks). Using ranks of data instead of data, then applying parametric tests such as t, F tests to ranks, is called the rank transform approach. In addition, the whole approach of computational simulation is to obtain empirical distribution of the test statistic under the null, thus belong to the category of non-parametric methods.

Covariance analysis (ANCOVA) is a method used in biomedical and health research, but when the normality assumption is not satisfied or when the dependent variable is bivariate or ordinal scale, various procedures are presented in the literature for covariance analysis. These can be grouped under two headings

(1) Covariance analysis alternatives: Johnson-Neyman and picked-point analysis;

(2) Non-parametric methods: Non-parametric ANCOVA researchers use Johnson-Neyman and Picked-Point analyses to interpret the interaction between covariates and groups. The Johnson-Neyman technique was first described by Johnson and Neyman in 1936.[1] When regression lines are not parallel, results of ANCOVA analyses are not valid. In this case, Johnson-Neyman and Picked-Point analyses can be used.[2] Johnson-Neyman procedure is presented as an alternative to parametric ANCOVA when there is variance heterogeneity.[3] Johnson-Neyman technique can be calculated with the WILCOX program.[4] In cases where ANCOVA cannot be used in case of regression of heterogeneity; researchers can use Johnson-Neyman and Picked-Point techniques to interpret the interaction between the covariates and groups.

Many methods are presented in the literature for non-parametric ANCOVA. A few of these are the methods of Quade, Puri & Sen, Conever & Iman, Rutherford and McSweeney & Porter.[5] However, these methods do not have separate application modules in commonly used statistical package programs.

First, the rank ANCOVA proposed by Quade for non-parametric ANCOVA is presented in the literature.[6] The quade method is based on the test of equality of residues. In this method, ranks of the dependent variable and covariates are taken separately. Then the rank transformation by Conever and Iman for non-parametric ANCOVA is proposed. In this approach, ranks of the covariates and dependent variable are taken. This method gives more effective results in data with outlier observations. ANCOVA and Rank ANCOVA results are often not very different even if the assumption of normality is not satisfied in the absence of outliers. Therefore, the rank transformation approach is better when there are outlier observations.[1,4] In rank transformation, after taking ranks of the covariate and dependent variable, if there is not a linear relationship between the dependent variable and covariate, then the rank transformation method, which is not a non-parametric ANCOVA method, cannot be applied in such a case. When the assumptions of covariance analysis are not satisfied, Koch et al. (1982, 1990) proposed the rank ANCOVA Quade and Mantel Haenszel statistics for non-parametric ANCOVA.[6,7] After the covariate effects are taken into account, this method is used to make comparison between the groups by using the Mantel-Haenszel statistics. This method can easily be used to compare groups when there are one or more covariates and it can be easily implemented using the SAS package program, applied the randomization-based method when the assumptions of the covariance analysis were not satisfied.[8] This method can be used when the data are continuous, binary, ordinal or time-dependent continuous.[7-9] Later, Jaeckel proposed the rank-based estimator for linear models. Rank-based estimator is a non-parametric alternative robust method developed against the least squares estimator and likelihood estimator methods that are traditionally used. Later, McKean et al. (1978) developed the Newton Step algorithm to easily calculate rank-based estimators and presented robust ANCOVA's solution.[10] Kloke and

McKean, developed the Rfit package in the R program to calculate the Robust Linear Model. The Rfit package basically uses the rank-based estimator. Details of this package can be found in.[11]

MATERIAL AND METHODS

EXTENDED MANTEL-HAENSZEL METHOD

This method makes the comparison between the groups after correcting the effects of the covariate.

$$Q_{MH} = \frac{\left(\sum_h w_h d_h\right)^2}{\sum_h w_h^2 \, v_{h0}}$$

Where h: layer

Wh: weighted layer

$\sum_h w_h = 1$.

$w_{Mh} = \frac{(n_{h1} n_{h2})}{(n_{h1} + n_{h2})}$dh is defined as follows.

$$d_h = \left(\frac{\sum_j y_{hj1}}{n_{h1}}\right) - \left(\frac{\sum_j y_{hj2}}{n_{h2}}\right) = (\bar{y}_{h1} - \bar{y}_{h2})$$

$$v_{h0} = \frac{\left\lfloor \sum_i \sum_j \left(y_{hji} - \bar{y}_h\right)^2 \right\rfloor}{w_{Mh}(n_h - 1)}$$

Here $n_h = (n_{h1} + n_{h2})$ ve $\bar{y}_h = \left(\sum_i \sum_j y_{hji} / n_h\right)$.

Non-parametric covariance analysis randomization-based method

The randomization-based method has many advantages. The data can be continuous, binary, ordinal and time-dependent continuous. In this method, confidence limits can be calculated. In addition, p value and confidence limits can be calculated in subgroups. However, covariate effects and estimators cannot be found. In addition, estimators of subgroups cannot be found.[12] In addition, this method is preferred primarily in studies with a control group.

The basis of the randomization-based method is based on permutation testing. Permutation test is used to determine the observed difference between sample averages. This test derives random copies in a certain systematic order by changing the order of the data and the replacement of the group by replacing the averages of the two groups, creating first, second and more permutations. The results obtained are combined by analyzing the derived copies. The process is continued by changing the copies constantly. The operation is carried out on the original data or ranks.

Permutation Number= $\binom{n_A + n_B}{n_A} = \frac{(n_A + n_B)!}{n_A! n_B!}$

$(|D_i| \geq |D^*|)$number of conditions satisfying the condition $\sum_{i=1}^{\binom{n_A+n_B}{n_A}} I(|D_i| \geq |D^*|)$

$$a_p = p_{H_0}(|D_i| \geq |D^*|) = \frac{\sum_{i=1}^{\binom{n_A+n_B}{n_A}} I(|D_i| \geq |D^*|)}{\binom{n_A+n_B}{n_A}}$$

$$f_{\square} = (d_{\square}, \boldsymbol{u}_{\square}^{'})$$

$$d_{\square} = (\bar{y}_{\square 1} - \bar{y}_{\square 2})\text{ve}\boldsymbol{u}_{\square} = (\bar{x}_{\square 1} - \bar{x}_{\square 2})$$

$$n_{\square 1} n_{\square 2}/(n_{\square 1} + n_{\square 2})$$

$$f^* = \frac{\sum_{\square} w_{\square} \boldsymbol{f}_{\square}}{\sum_{\square} w_{\square}}\text{ve}S_0 = \frac{\sum_{\square} w_{\square}^2 \boldsymbol{S}_{\square 0}}{(\sum_{\square} w_{\square})^2}$$

$$S_{\square 0} = \frac{\left[\sum_i \sum_j (g_{\square ji} - \bar{g}_{\square})(g_{\square ji} - \bar{g}_{\square})^{'}\right]}{w_{M\square}(n_{\square} - 1)}$$

ile $g_{\square ji} = (y_{\square ji}, x_{\square ji}^{'})^{'}$

$$\bar{g}_{\square} = \sum_j \sum_i g_{\square ji}/n_{\square}$$

$$S_{\square A} = \sum_i \left[\frac{\sum_j (g_{\square ji} - \bar{g}_{\square i})(g_{\square ji} - \bar{g}_{\square i})^{'}}{n_{\square i}(n_{\square i} - 1)}\right]$$

$$\bar{g}_{\square i} = \left\{\sum_j (g_{\square ji}/n_{\square i})\right\} S_{\square A} = \left(\sum_{\square} w_{\square}^2 S_{\square A}\right) \Big/ \left(\sum_{\square} w_{\square}\right)^2$$

$$b = (Z^{'}S^{-1}Z)^{-1}Z^{'}S^{-1}f^*,$$

$$v_b = (Z^{'}S^{-1}Z)^{-1}.A$$

In our study, it is aimed to investigate the difference of the Severity Score Index (SSI) in ribavirin and supportive treatment after correction of the effects of covariants with non-parametric randomization-based analysis of covariance (NPANCOVA) method.

In our study, data of the patients diagnosed with Crimean-Congo hemorrhagic fever disease were obtained from the Department of Pediatrics of Tokat Gaziosmanpaşa University Hospital. A total of 275 repeatedly measured data were obtained from 55 children with the diagnosis of Crimean-Congo hemorrhagic fever in 5 consecutive days in the hospital. While ribavirin treatment was applied to 26 of pediatric patients; supportive treatment was given to 28 patients. In our study, SSI values of pediatric patients were measured on day 1, day 2, day 3, day 4, day 5. Subjects were classified with an overall SSI score index 0-13 at each of five visits. The SSI is calculated to score information (platelet count, activated partial thromboplastin time (aPTT) value, fibrinogen level, presence of signs of bleeding, and presence of somnolence). The SSI was calculated according to sum of the scores for the 5 parameters, daily (Table 1).

SSI index values were obtained from the parameter values taken from the patients for five consecutive days. The patients were classified according to their SSI on recognition. There were 275 data what had available data from all five visits. Covariates include a baseline score, contamination time, gender and age at study. NPANCOVA method was used in the analysis of categorical repeated data obtained in our study. Method using SAS V.9.2 was applied. Means were considered significant when $p<0.05$.

TABLE 1. Characteristics of SSI parameters for Crimean-Congo hemorrhagic fever.

SSI parameter	Score
Platelet count, ×10^3 platelets/mm^3	
>150	0
150-50	1
49–20	2
<20	3
aPTT	
34	0
35-45	1
46-59	2
>60	3
Fibrinogen level, mg/dL	
≥180	0
179-160	1
159-120	2
<120	3
Bleeding	
No	0
Petechia	1
Ecchymosis	2
Bleeding	3
Somnolence	
No	0
Yes	1

aPTT: Activated partial thromboplastin time; SSI: Severity scoring index.

RESULTS

Ribavirin treatment was used in 26 of the pediatric patients and supportive treatment in 28 patients. The contamination time of our patients varies between 1 and 12. SSI score index values are between 1 and 13, and were calculated as a result of values taken from patients for 5 consecutive days. Severity Score Index had the following 5 parameters: platelet count (0-3 points), activated partial thromboplastin time (aPTT, 0-3 points), fibrinogen (0-3 points), bleeding (0-3 points) and drowsiness (0 or 1 points) point). SSI was calculated daily as the sum of the scores of the 5 parameters. In our study, SSI was obtained from the daily clinical and laboratory data recorded in the files of the patients during their follow-up.

TABLE 2. Crimean-Congo hemorrhagic fever for the SSI categorical outcomes.

	Estimate	SE	OR	95% CI		p value
				Lower	Upper	
Unadjusted	0.640	0.3441	1.8981	1.2237	2.5725	p<0.001
NPANCOVA	0.612	0.2393	1.8457	1.37	2.3147	p<0.001

SE: Standard error; OR: Odds ratio; CI: Confidence interval; SSI: Severity scoring index;
NPANCOVA: Non-parametric randomization-based analysis of covariance.

The standard error for the unadjusted for the SSI outcome was 0.3441. With unadjusted, there was a 30.45% reduction in the standard error to 0.2393. While the unadjusted and NPANCOVA p values were smaller at p<0.001, all the p values for this example were statistically significant at the 0.05 level and indicated SSI score index. Ribavirin differs in the administered and unadministered patient groups (Table 2).

CONCLUSION

There are many non-parametric methods for non-parametric ancova. It is not easy to decide which of these tests to use. In such cases, the choice is usually made based on the power of the tests and the characteristics of the data. It is necessary to decide which of the many non-parametric statistical methods mentioned in the introduction section to use, and to correctly interpret it.

The appropriateness of the statistical analysis gives meaningful results when it is correctly chosen. When the normality assumption is satisfied, parametric tests are superior in terms of power and effectiveness compared to non-parametric ones. However, since the assumptions of covariance analysis are not satisfied in our application, non-parametric covariance analysis is used. It is really difficult to decide which of the many non-parametric methods presented should be used. In the literature, Sievers and McKean presented the solution of Rank ANCOVA in small samples in 1986.[13] When the literature is examined on the non-parametric covariance analysis; Fan et al. presented solutions for non-parametric Ancova by comparing parametric, semi-parametric and non-parametric methods when there is missing data in clinical trials.[14] Zhao et al., proposed solutions when there is missing data in covariance analysis in survival analysis data.[15] Hussey et al explained the use of NPANCOVA method in clinical trials.[16] Yu (2019) researched methods that can be used in repeated categorical data analysis.[17] In our study, the analysis of categorical data with repeated measurements was performed using the NPANCOVA method.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and / or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

Idea/Concept: *Şirin Çetin;* ***Design:*** *Şirin Çetin, S Kenan Köse;* ***Control/Supervision:*** *Şirin Çetin;* ***Data Collection and/or Processing:*** *Erhan Karaaslan;* ***Analysis and/or Interpretation:*** *Şirin Çetin;* ***Literature Review:*** *Şirin Çetin;* ***Writing the Article:*** *Şirin Çetin;* ***Critical Review:*** *Ayşe Ülgen.*

REFERENCES

1. Huitema BE. The Analysis of Covariance and Alternatives. 2nd ed. New York: John Wiley; 2011. [Crossref]
2. Keppel G. Design and Analysis: A Researcher's Handbook. 3rd ed. Englewood Cliffs, NJ: Prentice Hall; 1991.
3. Dorsey SG, Soeken KL. Use of the Johnson-Neyman technique as an alternative to analysis of covariance. Nurs Res. 1996;45(6):363-6. [Crossref] [PubMed]
4. Quinn GP, Keough MJ. Experimental Design and Data Analysis for Biologists. 1st ed. New York: Cambridge University Press; 2002. [Crossref]
5. Barret TJ. Computations using analysis of covariance.WIREs Computational Statistics. 2011;3(3):260-8. [Crossref]
6. Quade D. Nonparametric analysis of covariance by matching. Biometrics. 1982;38(3):597-611. [Crossref] [PubMed]
7. Koch GG, Tangen CM. Non-parametric analysis of covariance and its role in non-inferiority clinical trials. Drug Information J. 1999;33(4):1145-59. [Crossref]
8. Koch GG, Tangen CM, Jung JW, Amara IA. Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. Stat Med. 1998;15-30;17(15-16):1863-92. [Crossref] [PubMed]
9. Koch GG, Carr GJ, Amara IA, Stokes ME, Uryniak TJ. Categorical data analysis. In: DA Berry, ed. Statistical Methodology in the Pharmaceutical Sciences. 1st ed. New York: Marcel Dekker; 1990. p.389-73.
10. McKean JW, Vidmar TJ. A comparison of two rank-based methods for the analysis of linear models. American Statistician.1994;48(3):220-9. [Crossref]
11. Kloke JD, McKean JW. Rfit: Rank-based estimation for linear models. The R Journal 2012;4(2):57-64. [Crossref]
12. Zink RC, Koch GG. NParCov3: A SAS/IML macro for non-parametric randomization-based analysis of covariance. J Statistical Software. 2012;50(3):1-17. [Crossref]
13. Sievers GL, McKean JW. On the robust rank analysis of linear models with non-symmetric error distributions. J Statistical Inference Planning. 1986;13:215-30. [Link]
14. Fan C, Zhang D, Wei L, Koch G. Methods for missing data handling in randomized clinical trials with nonnormal endpoints with application to a phase III clinical trial. Statistics in Biopharmaceutical Research. 2016;8(2):179-93. [Crossref]
15. Zhao P, Tang N, Qu A, Jiang D. Semiparametric estimating equations inference with nonignorable missing data. Statistica Sinica. 2017;27(1):89-113. [Link]
16. Hussey MA, Koch GG, Preisser JS, Saville BR. Analysis of matched studies with dichotomous outcomes using nonparametric randomization-based analysis of covariance. Statistics in Biopharmaceutical Research. 2013;5(3):194-203. [Crossref]
17. YU Jing. Statistical methods for topics involving repeated measures for categorical data. 2019. [Link]