Research Article GU J Sci, Part A, 9(4): 367-377 (2022) https://doi.org/10.54287/gujsa.1130039

Gazi University
Journal of Science
PART A: ENGINEERING AND INNOVATION
http://dergipark.org.tr/gujsa

Sample Size Estimation of Nonparametric Tests with Ordered Alternatives for Longitudinal Data in Randomized Complete Block Designs

Melike BAHCECITAPAR[1*] Hatice Tul Kubra AKDUR[2]

[1]Hacettepe University, Faculty of Science, Department of Statistics, Ankara, Türkiye
[2]Gazi University, Faculty of Science, Department of Statistics, Ankara, Türkiye

Keywords
Nonparametric Tests
Autocorrelated Longitudinal Data
Randomized Complete Block Designs

Abstract
Longitudinal studies involve repeated measurements from the same subjects or blocks over short or an extended periods of time. In longitudinal studies, usually the most important step is to decide how many experimental units to use. There are no closed form equations for determining sample size in many complex designs. Monte Carlo simulation method is an effective tool in complex designs to estimate power or sample size. This paper introduces estimating sample size for the number of blocks or experimental units based on a fixed number of treatment/time in randomized complete block designs with correlated longitudinal responses analyzed by nonparametric tests against ordered alternatives. The sample size of subjects is estimated for each test statistics by taking into account the autocorrelation structure of the error terms which form either a stationary first-order moving average or autoregressive with non-normally distributed white noise terms. An extensive sample size/power comparison among the recently proposed Modification of S test and the other two well-known nonparametric tests such as the Page test and the generalized Jonckheere test against ordered alternatives in randomized complete block designs is carried out under stationary first-order autoregressive and moving average error structures with white noise terms distributed with either Laplace or Weibull distributions. Simulation study indicates that the distribution of white noise and the error structure have an important role on sample size estimation for each nonparametric test. The Modification of S test requires large sample size in contrast to other tests for longitudinal data in the specified simulation setting.

Cite
Bahcecitapar, M., & Akdur, H. T. K. (2022). Sample Size Estimation of Nonparametric Tests with Ordered Alternatives for Longitudinal Data in Randomized Complete Block Designs. *GU J Sci, Part A*, 9(4), 367-377

Author ID (ORCID Number)
M. Bahcecitapar, 0000-0002-5443-6278
H. T. K. Akdur, 0000-0003-2144-0518

Article Process

Submission Date	13.06.2022
Revision Date	03.10.2022
Accepted Date	07.10.2022
Published Date	31.12.2022

1. INTRODUCTION

Longitudinal data taken from same subjects over time can occur within the randomized complete block designs (RCBDs). In RCBDs with longitudinal data, homogeneous subjects are blocks and a complete set of treatments/time points are randomly assigned to each block. Testing the trend/direction change of responses over time can be assessed by nonparametric tests for assessing the efficiency of treatment in a RCBD which is often expected to increase or decrease with the ordering of treatments. Many studies have examined the behavior of nonparametric tests in independent situations. Although longitudinal studies are quite popular today, the use of nonparametric tests in these studies is quite rare and nonparametric trend tests' behavior has almost never been investigated. In some clinical trials, it is reasonable to observe that the efficiency of treatment may increase or decrease based on the dose value of a drug throughout the study (Dmitrienko et al., 2007). The case that responses would incline or decline over time leads to the statistical testing problem with ordered alternatives and many nonparametric tests have been developed for this problem in RCBDs (Hollander, 1967; Shan et al., 2014; Best & Rayner, 2015). The Page (*P*) test is a well-known distribution-free tests for ordered alternatives on treatments in RCBDs (Page, 1963). The generalized Jonckheere (*GJ*) test was proposed for repeated measures in RCBDs (Zhang & Cabilio, 2012). The Modification of S (*MS*) test is a new

*Corresponding Author, e-mail: mlk@hacettepe.edu.tr

nonparametric test based on the rank differences within each block to detect monotonic trend of treatments in RCBDs (Akdur et al., 2019). All of these nonparametric tests assume that the within-block responses are independent and error terms follow an identical continuous distribution with known parameters. However, in some situations, such as when the blocks/subjects have repeated measures, the existence of autocorrelated errors and distribution of responses should be taken into account commonly. Recently, Akdur (2020) has investigated power performances of some nonparametric trend tests in RCBDs when randomized blocks contain dependent observations by using the circular block bootstrap method. The defining feature of longitudinal data is that repeated measures are likely to be autocorrelated over time. It is unreasonable to assume that the within-block responses or repeated measurements are independent over time in the analysis of this kind of popular statistical experimental designs. In RCBDs with longitudinal data, let Y_{ij} be the measurement of the ith subject at jth time point, where $i = 1,2,\dots,b$ and $j = 1,2,\dots,n$. A model for RCBDs with balanced and equally spaced longitudinal data can be written as

$$Y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}, i = 1,\dots,b; j = 1,\dots,n; \tag{1}$$

where μ is a common mean, β_i are block (subject) effects, τ_j are time (treatment) effects, b is the number of blocks (subjects) and n is the number of time points (treatments) at which subjects (blocks) are measured. The error terms ε_{ij} are here assumed to follow popular stationary time series processes, a stationary first-order moving average (MA1) or a stationary first-order autoregressive (AR1) model. The null hypothesis on the treatment effect is as

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_n = 0 \tag{2}$$

against the increasing ordered alternative

$$H_0: \tau_1 \leq \tau_2 \leq \cdots \leq \tau_n \tag{3}$$

or the decreasing ordered alternative

$$H_0: \tau_n \leq \tau_{n-1} \leq \cdots \leq \tau_1 \tag{4}$$

with at least one inequality. When the ordering of treatments is already known, the ordered alternative is a natural choice as in Equation (3) or Equation (4). Nonparametric tests are useful for statistical testing problem with the ordered alternative, when longitudinal data are not distributed normally or quite small to satisfy assumptions of traditional ANOVA models in RCBDs. Zhang and Cabilio (2012) proposed the *GJ* test against ordered alternatives for longitudinal data in RCBDs which are assumed to be autocorrelated in AR1 and second-order autoregressive (AR2) model to test the direction of change over treatments/time, when the errors follow $N(0,1)$ or Student-t distribution. Akdur et al. (2019) modified the *S* test for ordered alternative hypothesis in RCBDs for two-way ANOVA layout under the assumption of independent errors. However, they do not investigate the most general case in which the responses are assumed to be autocorrelated and non-normally distributed in RCBDs. Before starting clinical studies, it is an indispensable prerequisite to determine a suitable sample size according to the structure of the data. For the intended power value of the statistical test of interest, in recent years, many statistical studies have been carried out especially for the parametric tests with various effect sizes under independence assumption (Aslan et al., 2021; Serdar et al., 2021; Unalan, 2021). This paper proposes a new idea for calculating sample sizes (i.e. number of blocks) of *P*, *GJ* and *MS* test statistics for a target power against ordered alternatives for longitudinal data in RCBDs, when the error terms form either a stationary MA1 or AR1 model under Laplace and Weibull distributions. A broad set of Monte Carlo simulations is performed to model autocorrelated non-normally distributed responses and to compare these three nonparametric tests in terms of sample sizes and power values under a wide range of conditions. The remainder of this paper is organized as follows. The *P*, *GJ* and *MS* test statistics are briefly introduced in Section 2. In Section 3, the performance of *MS* test is compared with other commonly used nonparametric tests with regard to sample size under various conditions. In Section 4, a real example from a clinical trial is given to illustrate sample size calculations for the *MS*, *P* and *GJ* tests to obtain required 80% power. Section 5 is given to discussion and further works.

2. MATERIAL AND METHOD

This paper focuses on the *P* and *GJ* tests, well-known rank-based nonparametric tests for ordered alternatives in RCBDs. The *MS* test is a recently proposed nonparametric test which is an alternative to the *GJ* test for the ordered alternative problem in RCBDs.

2.1. Page's Test Statistics

For an ordered alternative hypothesis, Page's test statistics has been proposed as a nonparametric test based on within-block ranks, in which responses are ranked within each block (subject) (Page, 1963). The Page (*P*) test statistics is given as $P = \sum_{j=1}^{n} j r_j$, where $r_j = \sum_{j=1}^{b} R_{ij}$ is the sum of ranks of the j th treatment (time point) for block (subject) i and R_{ij} is the rank of response within block i at treatment j. Under the null hypothesis in Equation (2), the *P* test asymptotically follows a normal distribution with the mean $E(P) = \frac{bn(n+1)^2}{4}$ and the variance $V(P) = \frac{b(n-1)n^2(n+1)^2}{144}$ (Page, 1963). The null hypothesis would be rejected for a large value of the *P* test. Thas et al. (2012) provided a new version of Page test $PL = \sqrt{c}\sum_{j=1}^{n} \frac{l_j \bar{R_J}}{d}$ statistic using orthogonal trend contrast for tied and untied data where $\bar{R_J}$ is the mean of the ranks for treatment j, l_j are the linear trend coefficients, $d^2 = \sum_{j=1}^{n} l_j^2$. Here, $c = b(n-1)/(nV)$ for tied $V = \left\{\sum_{i,j} R_{ij}^2 / (bn)\right\} - (n+1)^2/4$ whereas for untied data, $V = (n^2 - 1)/12$. Additionally, Best and Rayner (2015) suggested a new test statistic which developed upon the idea of orthogonal trend analysis used in ANOVA and provided as $PT = \sqrt{b}\sum_{j=1}^{n} l_j \bar{R}_j / dS$ where S^2 is the mean square error of a randomized block ANOVA of the R_{ij}.

2.2. Generalized Jonckheere Test Statistics

The *GJ* test statistics is an alternative nonparametric test to the *P* test for the problem of testing the ordered alternative hypothesis in RCBDs (Zhang & Cabilio, 2012). It is based on the Kendall's Tau correlation that measures the association between responses within each block and the alternative ordering where each block is ranked within itself over treatment or time (Kendall, 1938). The *J* test statistics was generalized for RCBDs as given $GJ = \frac{1}{b}\sum_{i=1}^{b} T_K(i)$, where $T_K(i) = \binom{n}{2}^{-1} A_K(i)$ is a Mann-Kendall statistics for testing the direction of change in responses (repeated measures over time for subject i and $A_K(i) = \sum_{l<m}^{n} (R_{im} - R_{il})$ is the non-standardized Kendall's tau correlation where $sgn(R_{im} - R_{il})$ is either 1 or -1, depending on whether $R_{im} > R_{il}$ or $R_{im} < R_{il}$ (Skillings & Wolfe, 1978; Zhang & Cabilio, 2012).

2.3. Modification of S Test Statistics

The Modification of *S* (*MS*) test statistics is the recently proposed nonparametric test based on the *S* test for RCBDs. Rank differences within each block are calculated in the *MS* test to compose an overall test statistic (Shan et al., 2014; Akdur et al., 2019). The *MS* test statistic is defined as $MS = \frac{1}{b}\sum_{i=1}^{b} T_K(i)$, where $T_K(i) = \binom{n}{2}^{-1} S_K(i)$ and $S_K(i) = \sum_{l<m}^{n} (R_{im} - R_{il})\ I(R_{im} > R_{il})$ for subject i. $I((R_{im} > R_{il})$ is the indicator function which is either 1 or 0, depending on whether $R_{im} > R_{il}$ or $R_{im} < R_{il}$.

3. SIMULATION-BASED SAMPLE SIZE ESTIMATION

In this section, Monte Carlo simulations are conducted to compare the performances of the three nonparametric tests against ordered alternatives in RCBDs when observations are autocorrelated longitudinal data: 1) the *P* test; 2) the *GJ* test; and 3) the *MS* test. These tests are compared with each other in terms of sample sizes required to obtain 80% empirical power. Laplace and Weibull distributions are taken into account for longitudinal data. The nominal value of α is set to be as 0.05 for all simulation scenarios. Total 10000 iterations are utilized to obtain the 95% cutpoint for all tests. Giving the number of repeated measures over time, sample sizes are computed from simulated data based on the 95% cutpoint and 80% power. These are estimated for both AR1 and MA1 error models with parameter values ρ or $\theta = 0.20$, 0.40, 0.60 and 0.80 under various approximations to the distribution of longitudinal data and the whole is carried out for $n = 4$, 5, 7, 10. Additionally, sample sizes are also estimated for independent errors (ρ or θ = 0) over time/treatment. For random block and error distributions, Laplace distribution with location parameter 0 and dispersion parameter 1 (L(0,1)) and Weibull distribution with shape parameter 1 and scale parameter 1 (W(1,1)) are considered. In order to compare sample size calculations for the *MS*, *P* and *J* tests, the model with a linear trend alternative with individual block (subject) effects is considered here is written as

$$Y_{ij} = \beta_i + \phi j + \varepsilon_{ij}, i = 1, \ldots, b; j = 1, \ldots, n \quad (5)$$

assuming that for each subject, ε_{ij} are time series and they represent AR1 or MA1 errors. In Equation (5), ϕ is the slope term to generate an increasing order of treatments. For the linear trend model in Equation (5), ϕ is set to be as 0.2, 0.3 and 0.4 to assess the effect of different trends of longitudinal data on nonparametric tests. In such a model as in Equation (5), the repeated measures are considered to be from a linear model with errors following a AR1 or MA1 process over time and non-normal white noise terms. It is assumed that the error terms form either a stationary AR1 model

$$\varepsilon_{ij} = \rho\varepsilon_{ij-1} + e_{ij}, \quad (6)$$

or a stationary MA1 model

$$\varepsilon_{ij} = e_{ij} - \theta e_{ij-1}, \quad (7)$$

where ρ and θ are autocorrelation coefficients for AR1 and MA1 process, respectively and e_{ij} are independent and non-normally distributed white noise terms. In AR1 error model (Equation 6), it is assumed that ρ >0 and the magnitude of correlation among responses declines as they become farther apart. In MA1 error model (Equation 7), the correlation is the same for any two consecutive responses. The results of 10000 Monte Carlo iterations generated by the model in Equation (5) for each combination of n, ρ, θ and error distributions are summarized in Table 1-3. All simulations are performed for the *P*, *J* and *MS* tests against increasing ordered alternatives for repeated measures over time in RCBDs. Simulation studies are conducted in Cran R 3.4.3. The steps of this algorithm are inspired by the bi-section method and a sample size can be estimated based on a Monte Carlo simulation for the model in Equation (5) through the following steps:

(1) Obtain the pre-specified parameters through either historical data or previous clinical trials.
(2) Specify a desired statistical power (i.e. 80%) and a type-1 error rate (i.e. 5%).
(3) Simulate correlated responses within blocks for a fixed sample size *n* of treatment/time within each block under null hypothesis to obtain critical values of each trend test and record them.
(4) Simulate correlated responses within blocks for a fixed sample size *n* of treatment/time within each block under alternative hypothesis. From the obtained data set, estimate each nonparametric test statistics value based on the simulated data set.
(5) Repeat steps 3 to 4, M (i.e. M=10 000) times under various conditions explained through of this section.
(6) Then, using critical values record if a *p*-value is smaller than 0.05. Estimate the empirical power of the model based on the fraction of *p*-values that are smaller than 0.05.
(7) If the power calculated is less than the targeted power, divide the sample size into half and add it to itself and estimate the new power with this new sample size. If the power obtained is greater than the targeted power, divide the sample size in half and subtract it from itself and estimate the new power with this new sample size. Go to first step and repeat them through Step 7.
(8) Stop the simulation if the desired statistical power is obtained with a small tolerance value, i.e. 0.009.

Table 1 displays sample size calculations to obtain 80% empirical power of tests for longitudinal data generated by the model in Equation (5) with uncorrelated error terms within blocks for *n* = 4, 5, 7, 10. It can be seen that sample sizes estimated for Laplace distribution are greater than those for Weibull distribution. For each test, under all combinations for n and ϕ, the smallest sample size is obtained for Weibull distribution. Under uncorrelated error terms over time, sample sizes for 80% power are same for the *MS* and *P* tests for Weibull distribution. On the other hand, under both distributions, for each *n*, sample sizes of the *GJ* test are smaller than the others. As *n* increases, sample size decreases for all values of ϕ. Additionally, as shown in Figure 1, for fixed *n*, as ϕ increases, sample size estimated to obtain 80% empirical power of tests decreases.

Table 1. *Simulated sample size and power study based on uncorrelated errors following Laplace or Weibull distribution (α =0.05)*

ϕ	*n*	Laplace distribution			Weibull distribution		
		MS	*P*	*GJ*	*MS*	*P*	*GJ*
0.2	4	59 (0.799)	53 (0.808)	54 (0.809)	59 (0.807)	53 (0.793)	41 (0.794)
	5	41 (0.808)	41 (0.794)	36 (0.807)	41 (0.797)	41 (0.793)	32 (0.799)
	7	32 (0.804)	32 (0.800)	27 (0.793)	27 (0.802)	27 (0.797)	21 (0.807)
	10	24 (0.799)	24 (0.796)	20 (0.799)	20 (0.800)	21 (0.804)	17 (0.806)
0.3	4	24 (0.795)	24 (0.797)	20 (0.792)	21 (0.799)	21 (0.806)	17 (0.802)
	5	18 (0.795)	18 (0.799)	17 (0.808)	17 (0.806)	17 (0.794)	14 (0.797)
	7	14 (0.795)	14 (0.801)	11 (0.797)	11 (0.794)	11 (0.796)	9 (0.801)
	10	11 (0.802)	11 (0.796)	9 (0.799)	9 (0.805)	9 (0.793)	8 (0.806)
0.4	4	14 (0.799)	14 (0.792)	11 (0.791)	11 (0.798)	11 (0.792)	9 (0.808)
	5	11 (0.802)	11 (0.804)	9 (0.794)	9 (0.795)	10 (0.800)	8 (0.799)
	7	8 (0.794)	8 (0.796)	6 (0.799)	6 (0.793)	6 (0.795)	5 (0.796)
	10	6 (0.800)	6 (0.880)	5 (0.795)	4 (0.793)	4 (0.799)	3 (0.801)

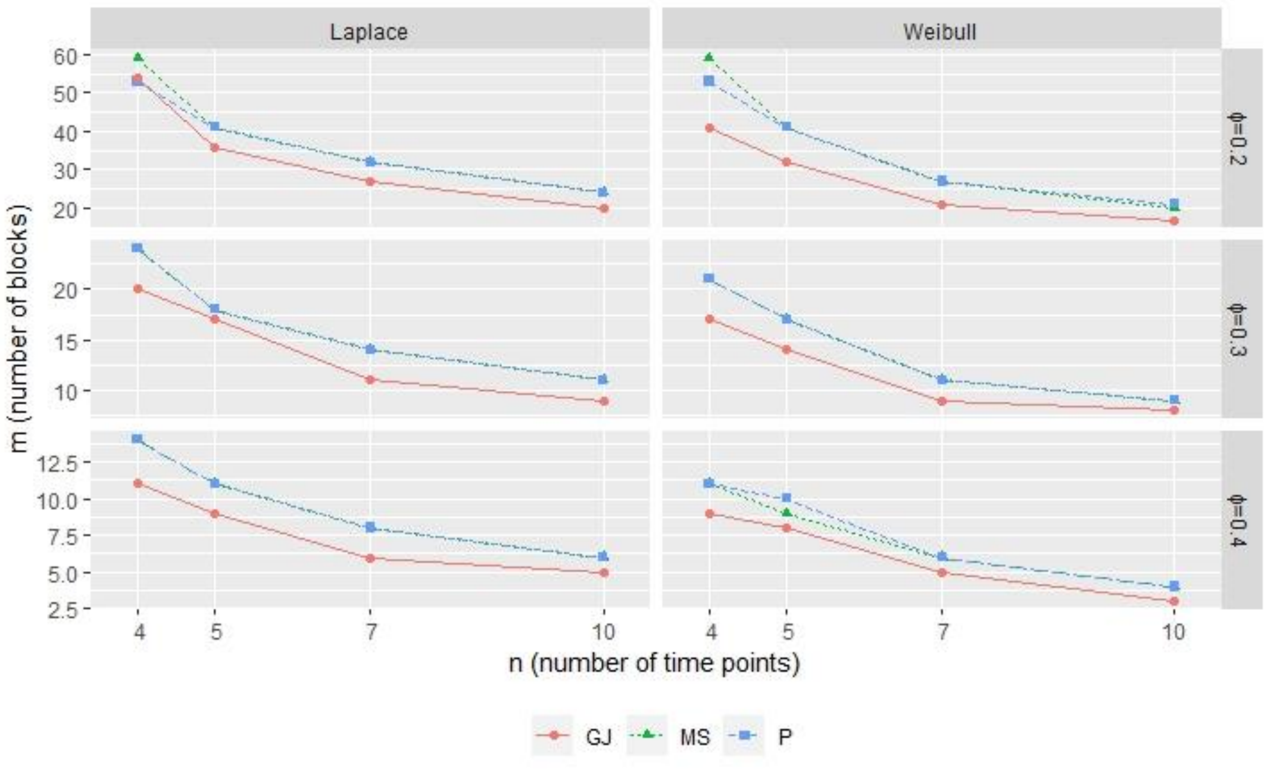

Figure 1. *Sample size (number of blocks) required to obtain 80% power curves of the MS, P and GJ tests for longitudinal data with uncorrelated errors following L(0,1) and W(1,1) distribution (α = 0.05)*

Table 2. *Simulated sample size and power study based on MA1 and AR1 error models following L(0,1) for* α*=0.05*

			L(0,1)					
			MA1			AR1		
ϕ	ρ or θ	*n*	*MS*	*P*	*GJ*	*MS*	*P*	*GJ*
0.2	0.20	4	47 (0.809)	47 (0.796)	41 (0.793)	47 (0.806)	47 (0.808)	41 (0.806)
		5	35 (0.796)	36 (0.796)	32 (0.807)	36 (0.801)	35 (0.798)	32 (0.808)
		7	27 (0.791)	27 (0.804)	24 (0.807)	27 (0.802)	27 (0.808)	24 (0.799)
		10	21 (0.791)	21 (0.805)	17 (0.806)	21 (0.795)	21 (0.798)	18 (0.797)
	0.40	4	47 (0.792)	47 (0.798)	41 (0.807)	41 (0.801)	41 (0.803)	36 (0.806)
		5	32 (0.796)	35 (0.802)	32 (0.807)	32 (0.808)	35 (0.808)	27 (0.802)
		7	27 (0.804)	27 (0.799)	21 (0.806)	24 (0.793)	24 (0.797)	20 (0.802)
		10	23 (0.808)	21 (0.794)	18 (0.799)	21 (0.797)	21 (0.808)	17 (0.804)
	0.60	4	47 (0.802)	47 (0.795)	41 (0.799)	35 (0.793)	36 (0.804)	35 (0.807)
		5	35 (0.808)	32 (0.803)	32 (0.803)	32 (0.803)	27 (0.792)	24 (0.797)
		7	27 (0.795)	27 (0.795)	21 (0.792)	21 (0.795)	21 (0.793)	20 (0.798)
		10	21 (0.795)	24 (0.807)	18 (0.806)	18 (0.806)	18 (0.801)	14 (0.799)
	0.80	4	51 (0.797)	51 (0.793)	47 (0.803)	36 (0.793)	41 (0.806)	35 (0.803)
		5	41 (0.808)	41 (0.801)	32 (0.794)	27 (0.800)	27 (0.796)	24 (0.803)
		7	27 (0.794)	32 (0.808)	24 (0.807)	21 (0.799)	21 (0.808)	17 (0.796)
		10	24 (0.802)	24 (0.796)	20 (0.804)	18 (0.808)	18 (0.805)	14 (0.795)
0.3	0.20	4	21 (0.792)	21 (0.794)	18 (0.806)	21 (0.805)	21 (0.799)	17 (0.803)
		5	17 (0.803)	17 (0.808)	14 (0.799)	17 (0.805)	18 (0.808)	14 (0.806)
		7	11 (0.796)	14 (0.807)	11 (0.806)	11 (0.805)	14 (0.809)	9 (0.796)
		10	11 (0.808)	11 (0.808)	8 (0.793)	11 (0.805)	11 (0.809)	8 (0.793)
	0.40	4	20 (0.804)	21 (0.795)	18 (0.802)	18 (0.792)	20 (0.806)	18 (0.807)
		5	14 (0.799)	17 (0.808)	14 (0.802)	14 (0.805)	17 (0.807)	14 (0.796)
		7	11 (0.801)	11 (0.795)	9 (0.795)	11 (0.800)	11 (0.793)	9 (0.796)
		10	9 (0.794)	9 (0.791)	8 (0.792)	9 (0.803)	9 (0.792)	8 (0.799)
	0.60	4	23 (0.805)	24 (0.805)	20 (0.793)	18 (0.803)	18 (0.792)	17 (0.806)
		5	17 (0.808)	17 (0.800)	14 (0.807)	14 (0.805)	14 (0.804)	11 (0.804)
		7	11 (0.794)	11 (0.793)	11 (0.806)	11 (0.799)	11 (0.806)	9 (0.798)
		10	9 (0.800)	11 (0.808)	8 (0.807)	9 (0.800)	9 (0.803)	8 (0.800)
	0.80	4	24 (0.801)	27 (0.803)	21 (0.792)	18 (0.792)	18 (0.797)	17 (0.799)
		5	17 (0.794)	18 (0.802)	14 (0.808)	14 (0.801)	14 (0.802)	11 (0.793)
		7	14 (0.805)	14 (0.798)	11 (0.797)	11 (0.808)	11 (0.808)	9 (0.809)
		10	11 (0.793)	11 (0.801)	9 (0.796)	9 (0.806)	9 (0.807)	6 (0.794)
0.4	0.20	4	14 (0.805)	14 (0.807)	11 (0.800)	14 (0.806)	14 (0.803)	11 (0.808)
		5	9 (0.791)	9 (0.798)	8 (0.794)	9 (0.794)	9 (0.795)	8 (0.792)
		7	8 (0.799)	8 (0.808)	6 (0.805)	8 (0.805)	8 (0.805)	6 (0.894)
		10	6 (0.797)	6 (0.808)	5 (0.796)	6 (0.806)	6 (0.806)	5 (0.793)
	0.40	4	14 (0.806)	14 (0.808)	11 (0.806)	11 (0.793)	11 (0.798)	11 (0.801)
		5	9 (0.808)	9 (0.803)	8 (0.799)	9 (0.795)	9 (0.802)	8 (0.794)
		7	6 (0.791)	6 (0.793)	6 (0.802)	6 (0.801)	6 (0.793)	6 (0.801)
		10	6 (0.803)	6 (0.800)	5 (0.800)	5 (0.798)	5 (0.795)	5 (0.800)
	0.60	4	14 (0.805)	14 (0.803)	12 (0.798)	11 (0.808)	11 (0.801)	9 (0.793)
		5	9 (0.793)	9 (0.799)	8 (0.800)	8 (0.799)	8 (0.805)	8 (0.804)
		7	8 (0.809)	8 (0.797)	6 (0.796)	6 (0.796)	6 (0.794)	5 (0.795)
		10	6 (0.802)	6 (0.807)	5 (0.805)	5 (0.801)	4 (0.794)	5 (0.806)
	0.80	4	17 (0.805)	14 (0.793)	14 (0.794)	11 (0.804)	11 (0.797)	9 (0.792)
		5	11 (0.806)	11 (0.797)	9 (0.803)	8 (0.805)	9 (0.802)	8 (0.809)
		7	8 (0.797)	8 (0.808)	6 (0.792)	6 (0.793)	6 (0.800)	5 (0.794)
		10	6 (0.796)	6 (0.801)	5 (0.802)	5 (0.792)	5 (0.793)	5 (0.806)

Table 3. *Simulated sample size and power study based on MA1 and AR1 error models following W(1,1) distribution for α=0.05*

			W(1,1)					
			MA1			AR1		
ϕ	ρ or θ	*n*	*MS*	*P*	*GJ*	*MS*	*P*	*GJ*
0.2	0.20	4	195 (0.803)	173 (0.793)	80 (0.796)	135 (0.799)	131 (0.801)	59 (0.798)
		5	135 (0.806)	131 (0.799)	68 (0.794)	153 (0.805)	153 (0.798)	68 (0.798)
		7	80 (0.791)	87 (0.803)	41 (0.793)	116 (0.806)	116 (0.802)	53 (0.793)
		10	53 (0.792)	47 (0.799)	32 (0.804)	60 (0.797)	60 (0.792)	35 (0.801)
	0.40	4	456 (0.793)	456 (0.802)	131 (0.806)	99 (0.794)	102 (0.808)	47 (0.802)
		5	405 (0.795)	386 (0.802)	102 (0.791)	401 (0.801)	405 (0.805)	87 (0.803)
		7	281 (0.799)	270 (0.795)	80 (0.795)	597 (0.794)	597 (0.793)	168 (0.802)
		10	132 (0.805)	135 (0.799)	51 (0.808)	339 (0.803)	339 (0.801)	120 (0.802)
	0.60	4	867 (0.803)	827 (0.807)	162 (0.796)	32 (0.797)	32 (0.807)	18 (0.803)
		5	591 (0.794)	608 (0.798)	113 (0.794)	87 (0.798)	87 (0.806)	36 (0.802)
		7	584 (0.802)	566 (0.796)	106 (0.792)	196 (0.799)	196 (0.796)	93 (0.799)
		10	249 (0.806)	249 (0.791)	77 (0.808)	138 (0.798)	138 (0.801)	63 (0.800)
	0.80	4	2228 (0.793)	2228 (0.792)	356 (0.806)	11 (0.807)	11 (0.797)	9 (0.798)
		5	734 (0.802)	699 (0.801)	131 (0.799)	14 (0.801)	14 (0.806)	9 (0.798)
		7	734 (0.793)	684 (0.796)	128 (0.804)	32 (0.804)	32 (0.793)	17 (0.800)
		10	239 (0.805)	230 (0.793)	77 (0.803)	17 (0.802)	17 (0.807)	8 (0.793)
0.3	0.20	4	41 (0.804)	36 (0.792)	24 (0.804)	27 (0.804)	27 (0.802)	17 (0.799)
		5	27 (0.803)	32 (0.806)	20 (0.804)	14 (0.801)	14 (0.804)	11 (0.803)
		7	20 (0.791)	21 (0.809)	14 (0.803)	21 (0.796)	24 (0.799)	14 (0.794)
		10	14 (0.808)	14 (0.809)	9 (0.807)	17 (0.808)	14 (0.794)	11 (0.800)
	0.40	4	53 (0.800)	51 (0.797)	32 (0.804)	27(0.802)	27 (0.807)	18 (0.800)
		5	41 (0.807)	41 (0.799)	21 (0.796)	41 (0.806)	36 (0.796)	21 (0.798)
		7	21 (0.802)	32 (0.793)	17 (0.795)	47 (0.803)	47 (0.797)	21 (0.804)
		10	14 (0.803)	21 (0.796)	14 (0.806)	36 (0.793)	36 (0.799)	18 (0.806)
	0.60	4	68 (0.800)	77 (0.804)	36 (0.804)	14 (0.805)	14 (0.801)	11 (0.806)
		5	47 (0.807)	47 (0.806)	27 (0.800)	20 (0.804)	21 (0.808)	11 (0.805)
		7	41 (0.804)	41 (0.792)	21 (0.796)	53 (0.8069	53 (0.796)	20 (0.798)
		10	27 (0.796)	27 (0.803)	14 (0.797)	135 (0.801)	135 (0.803)	27 (0.798)
	0.80	4	111 (0.802)	116 (0.806)	53 (0.791)	8 (0.793)	8 (0.804)	5 (0.799)
		5	60 (0.793)	60 (0.796)	32 (0.804)	8 (0.804)	8 (0.804)	5 (0.801)
		7	47 (0.809)	47 (0.792)	27 (0.808)	9 (0.800)	9 (0.804)	6 (0.792)
		10	32 (0.802)	32 (0.808)	17 (0.799)	27 (0.800)	27 (0.797)	11 (0.797)
0.4	0.20	4	18 (0.791)	18 (0.799)	14 (0.809)	17 (0.806)	17 (0.806)	11 (0.809)
		5	14 (0.807)	14 (0.804)	9 (0.793)	14 (0.808)	14 (0.808)	9 (0.794)
		7	9 (0.791)	9 (0.800)	6 (0.800)	11 (0.799)	11 (0.808)	6 (0.793)
		10	6 (0.800)	6 (0.791)	5 (0.804)	8 (0.806)	8 (0.809)	5 (0.796)
	0.40	4	21 (0.804)	21 (0.800)	14 (0.794)	14 (0.808)	14 (0.798)	9 (0.791)
		5	17 (0.807)	17 (0.800)	11 (0.798)	14 (0.801)	14 (0.798)	9 (0.795)
		7	11 (0.792)	11 (0.794)	8 (0.808)	14 (0.797)	14 (0.794)	8 (0.808)
		10	9 (0.804)	8 (0.792)	6 (0.806)	11 (0.798)	11 (0.792)	6 (0.798)
	0.60	4	27 (0.798)	27 (0.805)	15 (0.805)	8 (0.801)	8 (0.794)	5 (0.793)
		5	18 (0.791)	18 (0.799)	11 (0.798)	9 (0.805)	9 (0.808)	6 (0.806)
		7	14 (0.808)	14 (0.797)	9 (0.803)	14 (0.796)	14 (0.802)	8 (0.804)
		10	11 (0.809)	11 (0.806)	6 (0.792)	17 (0.797)	18 (0.796)	8 (0.807)
	0.80	4	36 (0.792)	36 (0.800)	24 (0.808)	6 (0.802)	6 (0.803)	5 (0.794)
		5	24 (0.805)	24 (0.800)	14 (0.807)	6 (0.800)	6 (0.798)	5 (0.799)
		7	17 (0.799)	17 (0.799)	11 (0.808)	5 (0.803)	5 (0.802)	3 (0.792)
		10	11 (0.808)	11 (0.792)	8 (0.804)	9 (0.799)	9 (0.803)	5 (0.808)

As can be seen in Table 2, the *GJ* test provides the smallest sample size even with small *n*. Additionally, for each ϕ and fixed correlation coefficient of either AR1 or MA1 structure, when *n* increases, sample sizes required to obtain 80% decreases for each test. It can be also seen that for each ϕ, sample size of each test under AR1 structure is smaller than sample size estimated under MA1 structure, when ρ or θ is greater than 0.40 for fixed values of *n*. In some simulation combinations, sample sizes are found to be same for the *MS* and *P* tests. When error terms have AR1 or MA1 structure, as *n* increases, sample size required to obtain at least 80% power of each test decreases for fixed ρ or θ. Additionally, for fixed *n* and ρ or θ, sample sizes estimated by AR1 error model are much smaller. For all combinations, the sample size of the *GJ* test is smaller than that of other tests. Figure 2 displays the plots of sample sizes required to obtain 80% power of the tests, when error terms have AR1 or MA1 structure and follows L(0,1) distribution.

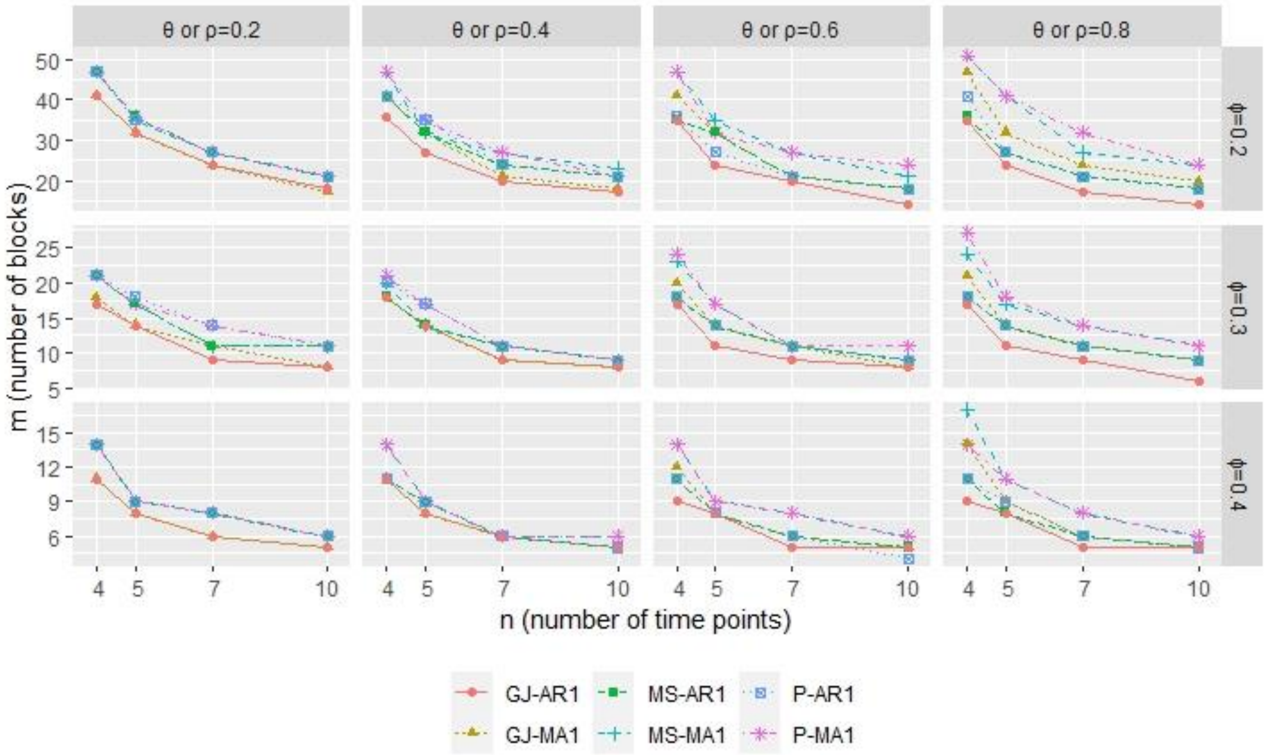

Figure 2. *Sample size (number of blocks) required to obtain 80% power curves of the MS, P and GJ tests for longitudinal data, when errors are autocorrelated in MA1 or AR1 structures and distributed in L(0,1) (α =0.05)*

Table 3 presents sample sizes for the *P*, *GJ* and *MS* tests required to obtain 80% power under the MA1 and AR1 models with white noise terms following Weibull distribution. The results of simulations under Weibull distribution are not completely similar with those under the Laplace distribution. For ρ from 0.4 to 0.8, as *n* increases from 4 to 7, sample sizes of all tests increase. Under AR1 structure, for all ρ, sample sizes for $n = 10$ are smaller than those for $n = 7$. On the other hand, when errors terms have MA1 structure, for all θ, as *n* increases from 4 to 10, sample size decreases. Also, sample sizes of the all tests under MA1 error model are larger than those provided under AR1 error model for Weibull distribution. The greatest sample size in this simulation study is obtained for $\theta = 0.80$, $n = 4$ and $\phi = 0.2$, when errors terms have MA1 model with white noise terms following W(1,1) distribution. The sample size of the *GJ* test has been observed to be the smallest in comparison to the other two tests. The smallest sample sizes for each combination of *n*, ϕ and ρ or θ are obtained by the *GJ* test. Usually, sample sizes of the *MS* and *P* tests are same. Figure 3 represents estimated sample sizes to obtain 80% power for the tests in RCBDs with longitudinal data, when error terms have either MA1 or AR1 model with W(1,1) distributed white noise terms.

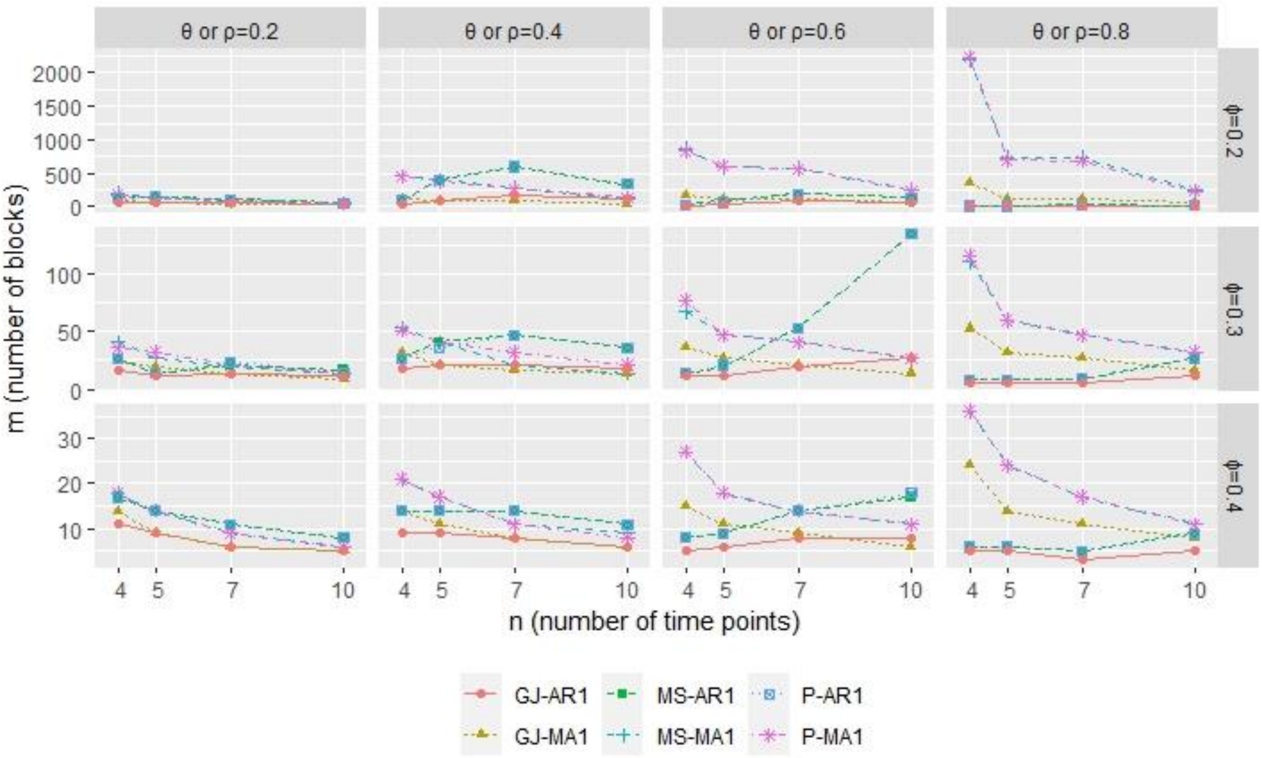

Figure 3. *Sample size (number of blocks) required to obtain 80% power curves of the MS, P and GJ tests for longitudinal data, when errors are autocorrelated in MA1 or AR1 structures and distributed in W(1,1) (α=0.05)*

4. DATA EXAMPLE

Data set in (Zhang & Cabilio, 2012) is taken into consideration to estimate sample sizes required to obtain 80% power of the *MS*, *P* and *GJ* tests for analyzing longitudinal data in RCBDs. This data set from a self-report survey was containing the total scores (0 (worst)-100 (best)) of pain, stiffness, and disability scores for each knee using the Western Ontario and McMasters Osteoarthritis Index (WOMAC). The longitudinal data set analyzed in (Zhang & Cabilio, 2012) includes 88 participants (b = 88) who have no missing pain records at the baseline and three 12-month interval visits (n = 4). They were concerned with changes of the total score over time in order to illustrate the implementation of the *GJ* test. The total WOMAC scores are appeared to be L-shaped with a long tail to the right and a declining trend in terms of medians for the left knee. Zhang and Cabilio (2012) analyzed these scores assuming AR(1) error terms with ρ = 0.5 and standard normal white noise. Here, the data are investigated for the *MS, P* and *GJ* tests for the AR(1) error model with white noise terms following W(1,1) and ρ = 0.5. At α = 0.05, sample sizes required for 80% power of the *MS, P* and *GJ* test statistics based on 10 000 MC simulations are estimated as 60 (0.806), 59 (0.7983) and 36 (0.8053), respectively.

5. DISCUSSION AND CONCLUSION

In the case of independent error terms, the distribution of the *GJ* statistics developed for RCBDs only depends on n and b and has an asymptotic normal distribution (Zhang & Cabilio, 2012). The expected values and variances of nonparametric tests depend on the error model parameters and distribution. Even if the parameters and distributions of the error terms are known, it is quite difficult to determine the asymptotic distributions of nonparametric tests, especially in the case of a dependent response variable. Therefore, it is almost impossible to obtain the exact formulas for the power and sample size of these tests. In this case, it will be inevitable for practitioners to estimate the power and sample size with simulation-based methods. This study is the first study in the literature that estimates the number of experimental units, that is, the number of blocks, which should be included in a research in an increasing or decreasing order for nonparametric tests to reach 0.80 power for testing the ordered alternatives hypothesis in the case of longitudinal data in two-way ANOVA form of

RCBDs. This study shows that, in the autocorrelated error models for RCBDs with longitudinal data, the number of blocks, i.e. sample size of subjects depends on the number of time points, the magnitude of autocorrelation coefficient, the slope term of the linear model and the distribution of errors. The fewer experimental units involved in clinical trials, the less costly it is, so choosing the trend test that requires a small sample size is very important. Monte Carlo simulations are conducted in this paper, because asymptotic distributions of tests are not known due to non-normal distributions and auto correlated error, even when the error term model is known. The sample size performance of nonparametric tests under various non-normal distributions is examined through an extensive simulation study in this paper. It is assumed that the error terms form either a stationary AR1 or MA1 model with independent white noise terms from Laplace or Weibull distribution. However, all simulations based on the fact that blocks are independent. The main purpose of this study is to investigate how sample size is affected by the distributions of errors and blocks, autocorrelation structure and the magnitude of trend in RCBDs with longitudinal data. The sample sizes for the *P*, *GJ* and *MS* tests under each non-normal distribution assuming that each subject has the same autocorrelation structure over time/ treatment were estimated by Monte Carlo simulations. When the error distribution follows laplace or weibull distributions under AR1 or MA1 correlation structures, the required sample size of GJ test is smaller than the required sample size of MS and Page tests in this study. For fixed ϕ and the coefficient parameter θ in the MA1 error model, as the number of time points n increases, sample size appears to be decreasing for each test under both non-normal error distributions. However, under Weibull distribution, when errors are autocorrelated in AR1 model, as n increases, sample sizes of all tests are increasing. For fixed n and ϕ, when the coefficient parameter ρ in the AR1 model approaches to 1, even for large n, sample size required for 80% power values of the *P*, *GJ* and *MS* tests appears to be decreasing under both distributions. However, for fixed n and ϕ, when the coefficient parameter θ in the MA1 model approaches to 1, sample sizes of the *P*, *GJ* and *MS* tests appears to be increasing under both distributions. In the light of simulation results, it can be seen that for RCBDs with longitudinal data, sample size of subjects is closely associated with the underlying error model, the values of its parameters, the distribution of data and the direction of time, all of which are unknown in practice. In comparison of sample sizes in terms of error distributions, for all combinations of correlation coefficient, number of time points and direction of time, the sample sizes estimated under Weibull distribution are much greater than those under Laplace distribution. Under Laplace distribution of errors, the more time points involved in the study and the larger correlation coefficient, the smaller sample size tends to be. The *MS*, *P* and *GJ* tests depend on the error distribution, the error correlation model and the direction of time/treatments. As a future study, the performance of alternative Page tests can be investigated in terms of sample size requirements for a target power in a Monte Carlo simulation study. In conclusion, we provide a simulation-based tool for situations in which the response variable is correlated within/ blocks and for which the accuracy of normal approximations is not sufficient in RCBDs.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

Akdur, H. T. K., Ozonur, D., Gul, H. H., & Bayrak, H. (2019). Comparison of rank-based tests for ordered alternative hypotheses in randomized complete block designs. *Gazi University Journal of Science*, *32*(2), 705-716.

Akdur, H. T. K. (2020). Comparison of non-parametric tests of ordered alternatives for repeated measures in randomized blocks. *Communications in Statistics - Simulation and Computation*, *51*(7), 4146-4158. doi:10.1080/03610918.2020.1740262

Aslan, E., Koskan, O., & Altay, Y. (2021). Determination of the Sample Size on Different Independent K Group Comparisons by Power Analysis. *Turkish Journal of Agricultural Research*, *8*(1), 34-41. doi:10.19159/tutad.792694

Best, D. J., & Rayner, J. C. W. (2015). An alternative to page's test permitting both tied and missing data. *Journal of Statistical Theory and Practice*, *9*(3), 524-536. doi:10.1080/15598608.2014.940098

Dmitrienko, A., Chuang-Stein, C., & D'Agostino Sr, R. B. (2007). *Pharmaceutical statistics using SAS: a practical guide*. SAS Institute.

Hollander, M. (1967). Rank tests for randomized blocks when the alternatives have an a priori ordering. *Annals of the Institute of Statistical Mathematics, 38*(3), 867-877. doi:10.1214/aoms/1177698880

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1/2), 81-93. doi:10.2307/2332226

Page, E. B. (1963). Ordered hypotheses for multiple treatments: a significance test for linear ranks. *Journal of the American Statistical Association*, *58*(301), 216-230. doi:10.1080/01621459.1963.10500843

Shan, G., Young, D., & Kang, L. (2014). A new powerful nonparametric rank test for ordered alternative problem. *PloS one*, *9*(11), e112924. doi:10.1371/journal.pone.0112924

Serdar, C. C., Cihan, M., Yücel, D., & Serdar, M. A. (2021). Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochemia Medica*, *31*(1), 27-53. doi:10.11613/bm.2021.010502

Skillings, J. H., & Wolfe, D. A. (1978). Distribution-free tests for ordered alternatives in a randomized block design. *Journal of the American Statistical Association*, *73*(362), 427-431. doi:10.1080/01621459.1978.10481595

Thas, O., Best, D. J., & Rayner, J. C. W. (2012). Using orthogonal trend contrasts for testing ranked data with ordered alternatives. *Statistica Neerlandica*, *66*(4), 452-471. doi:10.1111/j.1467-9574.2012.00525.x

Unalan, A. (2021). Sample Size in Clinical Researches: Power of the Test and Effect Size. *Black Sea Journal of Health Science*, *4*(3), 221-227. doi:10.19127/bshealthscience.866556

Zhang, Y., & Cabilio, P. (2012). A generalized Jonckheere test against ordered alternatives for repeated measures in randomized blocks. *Statistics in Medicine*, *32*(10), *Special Issue*, 1635-1645. doi:10.1002/sim.5606