

International Journal of Statistics and Applied Mathematics



ISSN: 2456-1452

NAAS Rating (2025): 4.49

Maths 2026; 11(7): 13-18

© 2025 Stats & Maths

<https://www.mathsjournal.com>

Received: 28-04-2026

Accepted: 30-05-2026

Mahassin Alnumani

Department of Statistics, North
Dakota State University, Fargo,
North Dakota, USA

Rhonda Magel

Department of Statistics, North
Dakota State University, Fargo,
North Dakota, USA

Comparison of the Friedman, modified Friedman, and Quade tests under symmetric and asymmetric distributions

Mahassin Alnumani and Rhonda Magel

DOI: <https://www.doi.org/10.22271/math.2026.v11.i7a.2604>

Abstract

This research conducts a Monte Carlo simulation to estimate significance levels and powers of 3 nonparametric tests for an RCBD design. Both 3 and 5 populations with 10 and 20 blocks are considered. Four underlying population distributions are considered: normal; t-distribution with 3 degrees of freedom; standard exponential; and chi-square with 2 degrees of freedom. The significant values are maintained for all tests. The Quade test has the highest estimated powers under the normal distribution for both 3 and 5 populations. The Friedman test has the highest estimated powers for 3 and 5 populations under all the other distributions included in the study.

Keywords: Randomized complete block design; nonparametric; 3 and 5 populations;

Introduction

Experimental designs that involve both within-subject and between-subject factors are widely used in applied research, particularly in fields such as psychology, medicine, agriculture, and industrial quality control. Classical general linear models assume normality and homogeneity of variance. Although modern linear mixed models allow flexible variance-covariance structures and generalized mixed models accommodate non-normal distributions, these approaches still require correct model specification. In practice, misspecification or complex data features such as skewness, heavy tails, or outliers can affect inference, motivating the use of nonparametric methods. In such cases, nonparametric tests provide more robust and distribution-free alternatives, making them suitable for analyzing data that deviate from normality or containing outliers (Hollander, Wolfe, & Chicken, 2014) [7]. Classic references emphasize that rank-based tests require substantially weaker assumptions than their normal-theory counterparts and often exhibit high asymptotic relative efficiency even under normality (Conover, 1999; Gibbons & Chakraborti, 2011; Lehmann, 1975) [2, 5, 9]. Lehmann (1975) [9] established the theoretical foundations for rank-based inference, demonstrating that nonparametric procedures maintain validity across a wide range of distributional forms while achieving high relative efficiency under normality.

In randomized complete block designs (RCBDs), the Friedman test (1937) has been the most used nonparametric procedure for detecting treatment effects when blocking is present. However, the Friedman test assumes that the variability among blocks is uniform, which may not hold when block variances differ substantially. To address this, the Modified Friedman test (Xu, J. et. al (2017); Jan & Shieh, 2025) [13] and the Quade test (1979) [8] were proposed as generalizations that incorporate block variability through ranking or weighting adjustments. These tests improve sensitivity to treatment differences when block effects are not homogeneous.

More recently, Jan and Shieh (2025) [8] proposed an improved nonparametric test and accompanying sample size determination procedures for RCBDs, addressing limitations of traditional Friedman-type methods in maintaining Type I error control and maximizing power under small to moderate sample conditions. Their findings emphasize the ongoing relevance

Corresponding Author:

Mahassin Alnumani

Department of Statistics, North
Dakota State University, Fargo,
North Dakota, USA

And potential for methodological innovation within the RCBD framework. Related studies have revisited the small-to-moderate sample accuracy of Friedman-type procedures, proposing F-transformations and explicit power approximations that improve Type I error control and power relative to the classical χ^2 approach (Jan & Shieh, 2025) [8].

Building upon these developments, the present research aims to systematically evaluate existing RCBD tests (Friedman, Modified Friedman, and Quade) for both $k=3$ and $k=5$ populations under both symmetric and asymmetric distributions. The general recommendation is to use the Quade test for a small number of treatment and the Friedman test if the number of treatments is 5 or more (Theodorsson-Norheim, 1987) [11].

This study aims to bridge this methodological gap and provide researchers with more reliable tools for inference when traditional assumptions of normality and homogeneity are violated. Furthermore, through Monte Carlo simulations under Normal, Exponential, Chi-Square, and t-distributions spanning symmetric, skewed, and heavy-tailed error structures components, this research will offer valuable insights into the robustness, efficiency, and power of the 3 nonparametric tests for an RCBD (Hettmansperger & McKean, 2011) [6].

The study will estimate and compare the statistical power and Type I error of these tests across different settings of the number of populations ($k = 3, 5$), the number of blocks ($b = 10, 20$). In parallel, we will evaluate the robustness of the proposed tests under four probability distributions Normal, Exponential, Chi-Square, and t-distribution ($df = 3$) using large-scale Monte Carlo simulations.

The friedman test (1937)

The Friedman test (Friedman, 1937, 1940) is a nonparametric procedure used to detect differences among treatments in a Randomized Complete Block Design (RCBD). It evaluates whether treatment effects differ significantly when blocking is used to control nuisance variability.

The hypotheses are stated as:

- $H_0: M_1 = M_2 = \dots = M_k$
- $H_1: \text{At least one median } M_j \text{ differs}$

Where M_j represents the population median of the j th treatment.

The test statistic is computed as:

$$FR = \frac{12}{bk(k+1)} \sum_{j=1}^k \left[R_j - \frac{b(k+1)}{2} \right]^2$$

Where:

b = number of blocks

k = number of treatments

R_j = sum of ranks assigned to the j th treatment within each block

Under the null hypothesis, FR approximately follows a chi-square distribution with $(k-1)$ degrees of freedom. The null hypothesis is rejected for large values of F, based on the critical value from the chi-square table.

Definitions, assumptions, and asymptotic reference distributions for the Friedman rank tests follow Gibbons and Chakraborti (2011) [5], who provided detailed derivations of their chi-square approximations and critical regions. Original development and rank-ANOVA justification appear in Friedman (1937, 1940), with practical χ^2 approximations reviewed in standard texts (Conover, 1999) [2].

The modified friedman test

The Modified Friedman test (Xu, J. et. al, 2017) [13] improves the traditional Friedman test by incorporating within-subject variability through ranking based on sample variances. Let X_{ij} be the observation for the i th subject and j th treatment, and let r_{ij} denote its rank within subject i . The sample variance of subject i is V_i , and its rank among all subjects is τ_i .

The hypotheses are stated as:

- $H_0: M_1 = M_2 = \dots = M_k$
- $H_1: \text{At least one treatment median differs}$

Where M_j represents the median of the j th treatment.

The adjusted rank is defined as

$$R_{ij} = (r_{ij} - 1) \tau_i + 1$$

The test statistic is

$$ModF = \frac{72n}{k(k+1)(n+1)(2n+1)} \sum_{j=1}^k (\bar{R}_{.j} - \bar{R}_{..})^2$$

Where:

- $\bar{R}_{.j} = \frac{1}{n} \sum_{i=1}^n R_{ij}$ is the mean adjusted rank for treatment j ,
- $\bar{R}_{..} = \frac{kn+k-n+3}{4}$ is the grand mean of adjusted ranks.

Under H_0 , the statistic $ModF$ is asymptotically chi-square distributed with $(k - 1)$ degrees of freedom.

The Quade Test (1979)

The Quade test (Quade, 1979) is a nonparametric alternative to the two-way analysis of variance (ANOVA) for data obtained from randomized complete block designs (RCBD). It is particularly useful when block variances differ, because it assigns weights to blocks according to their importance, as reflected by the ranges within each block. Blocks with larger internal variability receive larger weights, which allows the test to incorporate block heterogeneity directly into the ranking structure and improves its sensitivity under non-additivity or unequal block variances.

The Quade test is traditionally formulated with an F-distribution reference. For integration into our combined mixed-design framework, we transform the Quade F-statistic to a χ^2 -scale equivalent (see Transformation section below), maintaining consistency with the Friedman.

The hypotheses for the Quade test are stated as:

$$H_0: M_1 = M_2 = \dots = M_j,$$

H_1 : At least one treatment median differs,

Where M_j represents the median of the j th treatment.

The procedure begins by calculating the range within each block as:

$$\text{Range}_i = \max_j(X_{ij}) - \min_j(X_{ij})$$

so that blocks with a larger spread of observations are identified. Each block range is then ranked from smallest to largest, and its rank is denoted by Q_i . Next, within each block, the observations are ranked, and a weighted score is calculated as:

$$S_{ij} = Q_i[R(X_{ij}) - \frac{k+1}{2}]$$

Where $R(X_{ij})$ is the rank of the j th treatment within the i th block, and k is the number of treatments.

The Quade test statistic is then obtained by applying standard one-way ANOVA to the weighted scores S_{ij} , producing an F_{QD} -statistic that approximately follows an F-distribution with degrees of freedom:

$$F_{QD} = \frac{SSA/(k-1)}{SSE/((b-1)(k-1))}$$

$$df_1 = k - 1, df_2 = (b - 1)(k - 1)$$

Where SSA is the treatment sum of squares, SSE is the error sum of squares, and b is the number of blocks. Under H_0 , the classical rejection rule is:

$$F_{QD} > F_{\alpha; k-1, (b-1)(k-1)}$$

$$\chi_{QD}^2 = (k - 1) \times F_{QD}$$

Under H_0 , $\chi_{QD}^2 \sim \chi_{k-1}^2$.

Materials and methods

The performance the three nonparametric test statistics was examined through a comprehensive Monte Carlo simulation study conducted in SAS, where data generation followed established simulation principles (Wicklin, 2013)^[12]. The primary objective of this simulation was to evaluate and compare the behavior in a randomized complete block design (RCBD) under a wide range of experimental conditions and probability distributions differing in symmetry, skewness, and tail behavior.

Specifically, the study incorporated the following three well-established within-subject tests:

- The Friedman test (1937)
- The Modified Friedman test (1979)
- The Quade test (1979)

All datasets were generated using the SAS random number generator function RAND () with (SEED = 0) allowing the system clock to produce different random values in each run. These data-generation and computation steps were programmed in SAS using standard statistical programming practices (Bailer, 2010)^[1]. Five thousand random samples were generated for each scenario, with the number of blocks for each scenario either 10 or 20. The number of populations considered was either 3 or 5. The significance levels of each test under each of the various scenarios were estimated by counting the number of times the null hypothesis was rejected when the null hypothesis was true divided by 5000. Powers were estimated by counting the number of times each test rejected for a particular type of distribution and set of parameters and then divided by 5000. Using the maximum significance error a 95% confidence interval for each probability estimate would be $\text{est}(p) \pm 0.014$. The data were generated from two symmetric distributions being the normal and the t-distribution with 3 degrees of freedom. The data were generated from two skewed distributions being the chi-square distribution with 2 degrees of freedom and the standard exponential distribution.

The rejection regions determined using the Chi-Square and quantile corresponding to a 0.05 significance level (Conover, 1999)^[2]. Well-established references justify the large-sample χ^2 reference for the Friedman and Modified Friedman tests, and the F reference for the Quade test, with specific cautions regarding small-sample performance (Conover, 1999; Gibbons & Chakraborti, 2011; Hollander, Wolfe, & Chicken, 2014)^[2, 5, 7]. These classical sources provide the theoretical foundations for using asymptotic χ^2 and F approximations in rank-based procedures, emphasizing that accuracy improves as the number of blocks or treatments increases.

Through this simulation framework, the study aims to determine how effectively each test maintains Type I error control. If Type I error is controlled, the estimated powers are compared among the three tests.

Results and discussion

Results are given in Tables 1-8 for all distributions studied with 20 blocks. Tables 1-4 included the results for 3 populations and Tables 5-8 include the results for 5 populations. The results for 10 blocks were similar. The first line of every table gives the estimated alpha values for each test under that scenario. All tests had alpha values that were close to the stated alpha value of 0.05 with close defined to be in the interval (0.04, 0.06).

Results for the Normal distribution with 3 populations are given in Table 1. One can see that the power of the Quade test is generally the largest followed by the power of the Modified Friedman and then the Friedman. The same pattern is found for 5 populations in Table 2. The powers of the tests appear to be closer together for 5 populations than for 3 populations.

Results for the t-distribution are found in Tables 3-4. One can see that the estimated power of the Friedman test is higher than the power for the Modified Friedman or the Quade test. This is true for both 3 and 5 populations. The Quade test has the second highest power followed by the Modified Friedman. It appears that the powers between the tests are closer together for 3 populations than for 5 populations. Results for the exponential distribution are given in Tables 5-6. The same pattern holds as in the t-distribution case. The power of the Friedman test is higher than both the power of the Quade test and the power of the Modified Friedman. The Quade test has the second highest power. This is true for both 3 and 5 populations. It once again appears that the powers are closer together between the tests for 3 populations than for 5 populations.

Results for the chi-square distribution with 2 degrees of freedom are given in Tables 7-8. The same pattern holds as in both the t-distribution and the exponential distribution. The Friedman test has the highest powers. The Quade test has the second highest powers. The powers again appear to be closer together for 3 populations than for 5 populations.

Table 1: Empirical power under the Normal distribution; Variance=1; k=3

Scenario	m1	m2	m3	FR	MF	QD
b20	0	0.00	0.00	0.0582	0.0514	0.0542
	0	0.22	0.44	0.1756	0.1808	0.1914
	0	0.32	0.64	0.3154	0.3434	0.3616
	0	0.42	0.84	0.4792	0.5402	0.5550
	0	0.50	1.00	0.6414	0.7148	0.7312
	0	0.56	1.12	0.7500	0.8208	0.8308
	0	0.62	1.24	0.8286	0.8888	0.8966
	0	0.66	1.32	0.8700	0.9192	0.9250
	0	0.70	1.40	0.9136	0.9526	0.9570
	0	0.73	1.46	0.9276	0.9612	0.9652
	0	0.76	1.52	0.9438	0.9744	0.9770

Table 2: Empirical power under the Normal distribution; Variance=1; k=5

Scenario	m1	m2	m3	m4	m5	FR	MF	QD
b20	0	0.00	0.00	0.00	0.00	0.0510	0.0450	0.0490
	0	0.20	0.30	0.44	0.58	0.2356	0.2378	0.2510
	0	0.28	0.42	0.60	0.80	0.4400	0.4506	0.4680
	0	0.36	0.54	0.76	1.02	0.6504	0.6686	0.6834
	0	0.44	0.66	0.92	1.24	0.8358	0.8524	0.8598
	0	0.48	0.72	1.00	1.34	0.8924	0.9118	0.9180
	0	0.50	0.75	1.04	1.38	0.9204	0.9302	0.9326
	0	0.52	0.78	1.08	1.44	0.9312	0.9490	0.9534
	0	0.53	0.79	1.10	1.47	0.9396	0.9562	0.9598
	0	0.54	0.81	1.12	1.50	0.9594	0.9692	0.9692
	0	0.55	0.82	1.14	1.52	0.9584	0.9682	0.9724

Table 3: Empirical power under the t-distribution (df = 3); k=3

Scenario	m1	m2	m3	FR	KW	MF	QD
b20	0	0.00	0.00	0.0544	0.0430	0.0500	0.0530
	0	0.50	1.00	0.4580	0.0946	0.4212	0.4364
	0	0.60	1.20	0.5896	0.1192	0.5602	0.5754
	0	0.70	1.40	0.7306	0.1590	0.6962	0.7104
	0	0.80	1.60	0.8346	0.2050	0.7972	0.8140
	0	0.83	1.66	0.8630	0.2148	0.8278	0.8434
	0	0.86	1.72	0.8874	0.2230	0.8566	0.8654
	0	0.88	1.76	0.8978	0.2324	0.8696	0.8780
	0	0.89	1.78	0.8960	0.2464	0.8760	0.8886
	0	0.90	1.80	0.9108	0.2426	0.8848	0.8958
	0	0.91	1.82	0.9140	0.2382	0.8826	0.8880

Table 4: Empirical power under the t-distribution (df = 3); k=5

Scenario	m1	m2	m3	m4	m5	FR	MF	QD
b20	0	0.00	0.00	0.00	0.00	0.0472	0.0416	0.0450
	0	0.20	0.32	0.48	0.68	0.2104	0.1752	0.1860
	0	0.23	0.37	0.55	0.79	0.2764	0.2304	0.2390
	0	0.27	0.42	0.63	0.94	0.3856	0.3180	0.3318
	0	0.31	0.47	0.71	1.04	0.4628	0.3910	0.4056
	0	0.35	0.52	0.78	1.14	0.5372	0.4494	0.4712
	0	0.39	0.58	0.87	1.28	0.6514	0.5570	0.5756
	0	0.43	0.63	0.95	1.39	0.7246	0.6514	0.6616
	0	0.47	0.68	1.02	1.50	0.7844	0.6982	0.7098
	0	0.51	0.73	1.10	1.61	0.8560	0.7692	0.7772
	0	0.55	0.78	1.17	1.72	0.8932	0.8192	0.8276

Table 5: Empirical power under Exp (1); k=3

Scenario	m1	m2	m3	FR	MF	QD
b20	0	0.00	0.00	0.0452	0.0430	0.0482
	0	0.20	0.40	0.2880	0.2346	0.2462
	0	0.30	0.50	0.3974	0.3280	0.3474
	0	0.40	0.60	0.5554	0.4742	0.4910
	0	0.50	0.70	0.6568	0.5964	0.6184
	0	0.60	0.80	0.7452	0.6940	0.7108
	0	0.70	0.90	0.8228	0.7940	0.8094
	0	0.75	0.95	0.8516	0.8326	0.8448
	0	0.78	0.98	0.8728	0.8500	0.8626
	0	0.81	1.01	0.8910	0.8660	0.8768
	0	0.84	1.04	0.9012	0.8852	0.8948

Table 6: Empirical power under Exp (1); k=5

Scenario	m1	m2	m3	m4	m5	FR	MF	QD
b20	0	0.00	0.00	0.00	0.00	0.0524	0.0488	0.0510
	0	0.20	0.30	0.46	0.62	0.5626	0.4512	0.4696
	0	0.23	0.36	0.54	0.72	0.6994	0.5774	0.5944
	0	0.27	0.42	0.63	0.84	0.8058	0.6956	0.7114
	0	0.30	0.48	0.72	0.96	0.8948	0.8016	0.8126
	0	0.32	0.51	0.76	1.01	0.9256	0.8500	0.8574
	0	0.33	0.53	0.79	1.05	0.9458	0.8732	0.8854
	0	0.34	0.55	0.82	1.09	0.9556	0.9004	0.9082
	0	0.35	0.56	0.84	1.12	0.9602	0.9020	0.9108
	0	0.36	0.58	0.87	1.15	0.9676	0.9146	0.9218
	0	0.37	0.59	0.89	1.18	0.9770	0.9328	0.9386

Table 7: Empirical power under χ^2 (2); k=3

Scenario	m1	m2	m3	FR	MF	QD
b20	0	0.00	0.00	0.0568	0.0456	0.0504
	0	0.20	0.70	0.2520	0.1858	0.2008
	0	0.35	0.85	0.3138	0.2464	0.2616
	0	0.50	1.00	0.3902	0.3144	0.3284
	0	0.65	1.15	0.4874	0.4172	0.4336
	0	0.75	1.25	0.5550	0.4862	0.5046
	0	0.85	1.35	0.6286	0.5574	0.5774
	0	0.92	1.42	0.6570	0.5778	0.5950
	0	0.98	1.48	0.6842	0.6210	0.6416
	0	1.04	1.54	0.7130	0.6546	0.6694
	0	1.10	1.60	0.7362	0.6778	0.6950

Table 8: Empirical power under χ^2 (2); k=5

Scenario	m1	m2	m3	m4	m5	FR	MF	QD
b20	0	0.00	0.00	0.00	0.00	0.0390	0.0456	0.0514
	0	0.08	0.16	0.50	0.65	0.2496	0.1872	0.1998
	0	0.12	0.24	0.65	0.85	0.3654	0.2764	0.2894
	0	0.16	0.32	0.80	1.05	0.5002	0.3912	0.4024
	0	0.20	0.40	0.95	1.25	0.6534	0.5280	0.5358
	0	0.23	0.46	1.05	1.38	0.7276	0.5948	0.6054
	0	0.26	0.52	1.14	1.50	0.7940	0.6622	0.6716
	0	0.28	0.56	1.22	1.60	0.8546	0.7404	0.7496
	0	0.30	0.60	1.28	1.68	0.8734	0.7594	0.7674
	0	0.33	0.66	1.36	1.78	0.8986	0.7990	0.8056
	0	0.36	0.72	1.44	1.88	0.9296	0.8460	0.8484

Conclusion

The findings from this research do indicate that the Quade test has the highest powers for when the underlying distributions are normal and this was true for both 3 and 5 populations. The powers of the tests did become closer as the number of populations increased. The Quade test is recommended if one has fewer than 5 populations and thinks that the underlying distributions are symmetric with light to medium tails.

For the heavier-tailed t-distribution with 3 degrees of freedom, the Quade test had the second highest powers with the Friedman test having the highest powers. The powers were closer together for 3 populations than for 5 populations. The Friedman test is recommended for heavy-tailed symmetric distributions and especially if the number of populations becomes 5 or greater.

Results for both the exponential and chi-square with 2 degrees of freedom distributions, the Friedman test had the highest powers followed by the Quade test. The powers appear to have a greater separation for 5 populations over 3 populations. The Friedman test is recommended if one suspects the underlying distributions to be skewed. This is especially true when the number of populations exceed 3.

References

1. Bailer, A. J. (2010). *Statistical programming in SAS*. SAS Institute Inc.
2. Conover, W. J. (1999). *Practical nonparametric statistics* (3rd Ed.). Wiley.
3. Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200), 675–701.
4. Friedman, M. (1940). A comparison of alternative tests of significance for the problem of matched samples. *Annals of Mathematical Statistics*, 11(1), 86–92.
5. Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric statistical inference* (5th ed., updated version). CRC Press.
6. Hettmansperger, T. P., & McKean, J. W. (2011). *Robust nonparametric statistical methods*. CRC Press.
7. Hollander, M., Wolfe, D. A., & Chicken, E. (2014). *Nonparametric statistical methods* (3rd ed.). Wiley.
8. Jan, S. L., & Shieh, G. (2025). Improved nonparametric procedures and sample size determination for randomized block designs. *Journal of Nonparametric Statistics*. (In press).
9. Lehmann, E. L. (1975). *Nonparametrics: Statistical methods based on ranks*. Holden–Day.
10. Quade, D. (1979). Using weighted rankings in the analysis of variance. *Journal of the American Statistical Association*, 74(367), 680–683.
11. Theodorsson-Norheim, E. (1987). Friedman and Quade tests: BASIC computer program to perform nonparametric two-way analysis of variance and multiple comparisons on ranks of several related samples. *Computers in Biology and Medicine*, Vol. 17, Issue 2, pp 85-99.
12. Wicklin, R. (2013). *Simulating data with SAS*. SAS Institute Inc.
13. Xu, J. Shan, G., Amei, A., Zhao, J., Young, D., and Clark S. (2017). A Modified Friedman test for randomized complete block designs. *Communications in Statistics – Simulation and Computation*, Vol. 46. Issue 2, pp 1508-1519.