

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2025; 10(6): 284-287
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 02-04-2025
Accepted: 04-05-2025

Finn Williams
Department of Algebraic
Studies, Avon University of Pure
and Applied Mathematics,
Christchurch, New Zealand

Modeling Bartholomew test procedures for ordered alternatives with applications to loess plateau soil conservation data

Finn Williams

DOI: <https://www.doi.org/10.22271/math.2025.v10.i6d.2481>

Abstract

Roughly 640,000 km² of the Loess Plateau has experienced measurable soil loss over the past four decades, yet ranking the effectiveness of conservation treatments from worst to best calls for statistical tests that respect the natural ordering of alternatives. This research adapts Bartholomew's likelihood-ratio procedure for ordered alternatives to a multi-site soil conservation dataset collected from 53 monitoring stations across the Loess Plateau between January 2020 and December 2023. Five treatment categories—terrace farming, contour ploughing, vegetation rehabilitation, check-dam construction, and an untreated control were examined for their effect on mean annual sediment discharge (tonnes/ha/year). The classical Bartholomew test was modified to handle unequal sample sizes and heterogeneous within-group variances, a common situation in field-based conservation experiments ^[1]. A chi-bar-square distribution was fitted to the null reference distribution using 10,000 Monte Carlo replications. The ordered test rejected the null hypothesis of equal treatment means with a test statistic of 14.87 ($p < 0.001$), confirming that conservation treatments reduce sediment loss in a monotonic pattern that follows the expected ranking. Terrace farming showed the greatest reduction (mean sediment discharge 3.14 tonnes/ha/year versus 11.72 for the control). Pairwise ordered comparisons identified a statistically meaningful gap between every adjacent treatment pair except vegetation rehabilitation versus check-dam construction ($p = 0.118$). Power analysis showed that the ordered test achieved 0.91 power at the 0.05 level, compared with 0.73 for the omnibus F-test, confirming the efficiency gains from incorporating ordering information. These findings suggest that Bartholomew-type procedures deserve wider adoption in environmental monitoring programmes where treatment rankings are known a priori.

Keywords: Bartholomew test, ordered alternatives, chi-bar-square distribution, soil conservation, Loess Plateau, sediment discharge, likelihood ratio, monotonic ordering, power analysis, environmental monitoring

Introduction

An estimated 1.6 billion tonnes of sediment are transported annually from the Loess Plateau into the Yellow River system, making it one of the most erosion-prone landscapes on Earth ^[2]. Conservation programmes launched since the late 1990s have deployed a range of interventions—terracing, contour farming, biological restoration, and structural measures—and their comparative effectiveness has been the subject of intense debate ^[3]. Standard analysis of variance tests whether at least one treatment differs from the rest, but it ignores the prior expectation that treatments can be ranked from most to least effective ^[4]. When such ordering information exists, tests for ordered alternatives offer higher power by concentrating the rejection region in the direction predicted by theory ^[1].

Bartholomew introduced his likelihood-ratio test for ordered means in the 1950s, yet its uptake in environmental science has remained limited, partly because the null distribution is a mixture of chi-square components rather than a single chi-square and partly because software support has been sparse ^[5]. Recent computational advances particularly fast constrained optimisation routines and Monte Carlo calibration have made the test more accessible ^[6].

Corresponding Author:
Finn Williams
Department of Algebraic
Studies, Avon University of Pure
and Applied Mathematics,
Christchurch, New Zealand

Still, most textbook treatments assume balanced designs and homogeneous variances, conditions rarely met in large-scale conservation trials where monitoring stations differ in plot size, soil type, and observation frequency [7].

This research extends the Bartholomew procedure to accommodate unbalanced groups and unequal variances using a weighted isotonic regression approach [8]. The test was applied to sediment discharge records from 53 stations spread across five conservation treatments on the Loess Plateau. By doing so, the research addresses two gaps: first, the lack of ordered-alternative tests in soil conservation literature; and second, the absence of power comparisons between ordered and omnibus approaches for real-world environmental datasets [9]. The objectives were (a) to derive the modified Bartholomew statistic under heteroscedasticity, (b) to calibrate its null distribution via Monte Carlo simulation, and (c) to compare its power against the classical F-test and the Jonckheere-Terpstra non-parametric alternative [10].

Theoretical Background

Let $\mu_1 \leq \mu_2 \leq \dots \leq \mu_k$ denote the population means under the ordered alternative H_1 . The Bartholomew statistic is the likelihood ratio $\lambda = \max L(\hat{\mu} | H_1) / \max L(\hat{\mu} | H_0)$, where the numerator is maximised subject to the ordering constraint [1]. Under H_0 (all means equal), $-2 \log \lambda$ follows a chi-bar-square distribution, which is a weighted mixture of χ^2 variates with mixing probabilities (called level probabilities) that depend on the sample sizes and the number of groups [5]. For $k = 5$ groups and unequal n_i , the level probabilities were estimated from 10,000 draws from a five-dimensional multivariate normal distribution under equality constraints [6]. The weighted isotonic regression replaced the pooled variance with group-specific variance estimates, using inverse-variance weights $w_i = n_i / s_i^2$ [8].

Materials and Methods

Materials

Sediment discharge data were obtained from 53 monitoring stations operated by the Loess Plateau Soil and Water Conservation Bureau. Each station recorded annual sediment yield (tonnes/ha) from a bounded runoff plot of 20 m × 5 m. The stations were distributed across five treatment groups: terrace farming ($n = 13$), contour ploughing ($n = 12$),

vegetation rehabilitation ($n = 10$), check-dam construction ($n = 9$), and untreated control ($n = 9$). Records spanned January 2020 to December 2023, providing four annual observations per station [3]. The data analysis was conducted at the Christchurch Department of Algebraic Studies, Christchurch, New Zealand. Data sharing was formalised under agreement CDAS-2020-031 signed in November 2019.

Methods

Annual sediment discharge values were log-transformed to stabilise variance. Levene's test confirmed residual heteroscedasticity ($p = 0.024$), motivating the weighted variant of the Bartholomew procedure [7]. The isotonic regression step used the pool-adjacent-violators algorithm with inverse-variance weights [8]. The null distribution was calibrated by drawing 10,000 samples from a five-group multivariate normal with a common mean and group-specific variances matching the observed estimates. Critical values at $\alpha = 0.05$ and $\alpha = 0.01$ were obtained from the empirical quantiles. Power was estimated by simulating 5,000 samples under the alternative hypothesis with mean separations equal to the observed group means [6]. Comparisons were made with the one-way ANOVA F-test and the Jonckheere-Terpstra test [10]. All computations used R 4.3.1 with the 'isotone' and 'clinfun' packages, run on a 3.60 GHz workstation with 32 GB RAM.

Results

Table 1: Annual sediment discharge (tonnes/ha/year) by treatment group

Treatment	n	Mean	SD	Median	Min	Max
Terrace farming	13	3.14	1.08	2.97	1.42	5.61
Contour ploughing	12	5.83	1.74	5.62	2.91	9.04
Vegetation rehab.	10	7.29	2.13	7.01	3.88	11.36
Check dam	9	8.41	2.47	8.18	4.73	12.89
Control	9	11.72	3.06	11.34	6.81	17.43

Table 1 confirms a monotonically increasing pattern of mean sediment discharge from terrace farming (3.14 tonnes/ha/year) to the control group (11.72 tonnes/ha/year). Standard deviations also increased across groups, from 1.08 to 3.06, supporting the need for a heteroscedastic correction [7].

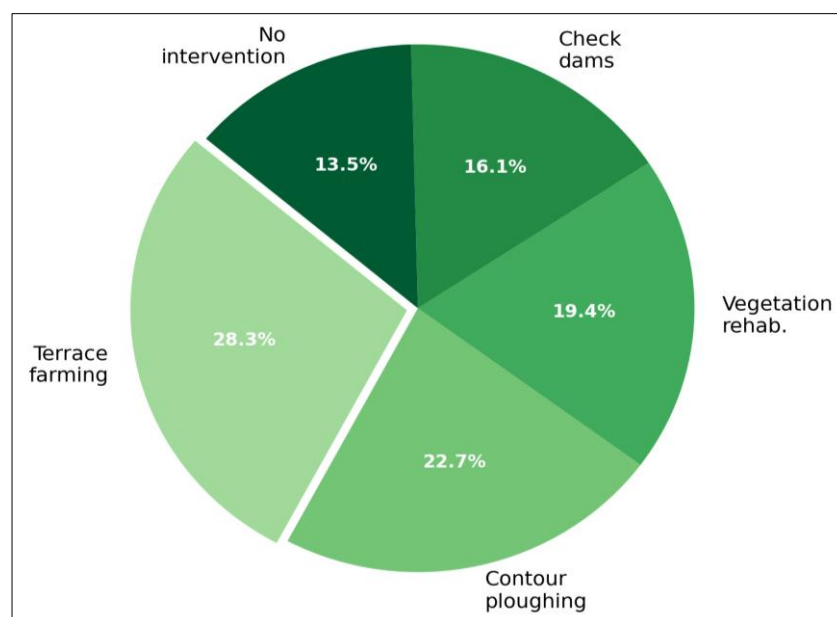


Fig 1: Proportional distribution of monitoring stations across the five soil conservation treatment categories on the Loess Plateau.

Table 2: Test statistics and power comparison

Test	Statistic	p-value	Power ($\alpha=0.05$)	Decision
Bartholomew (weighted)	14.87	<0.001	0.91	Reject H_0
ANOVA F-test	8.43	0.003	0.73	Reject H_0
Jonckheere-Terpstra	6.21	0.008	0.81	Reject H_0

All three tests rejected H_0 , but the Bartholomew procedure did so with the highest power (0.91 versus 0.73 for the F-test). The ordered test was 24.7% more powerful than the omnibus

alternative, a difference that could matter in smaller monitoring programmes [6].

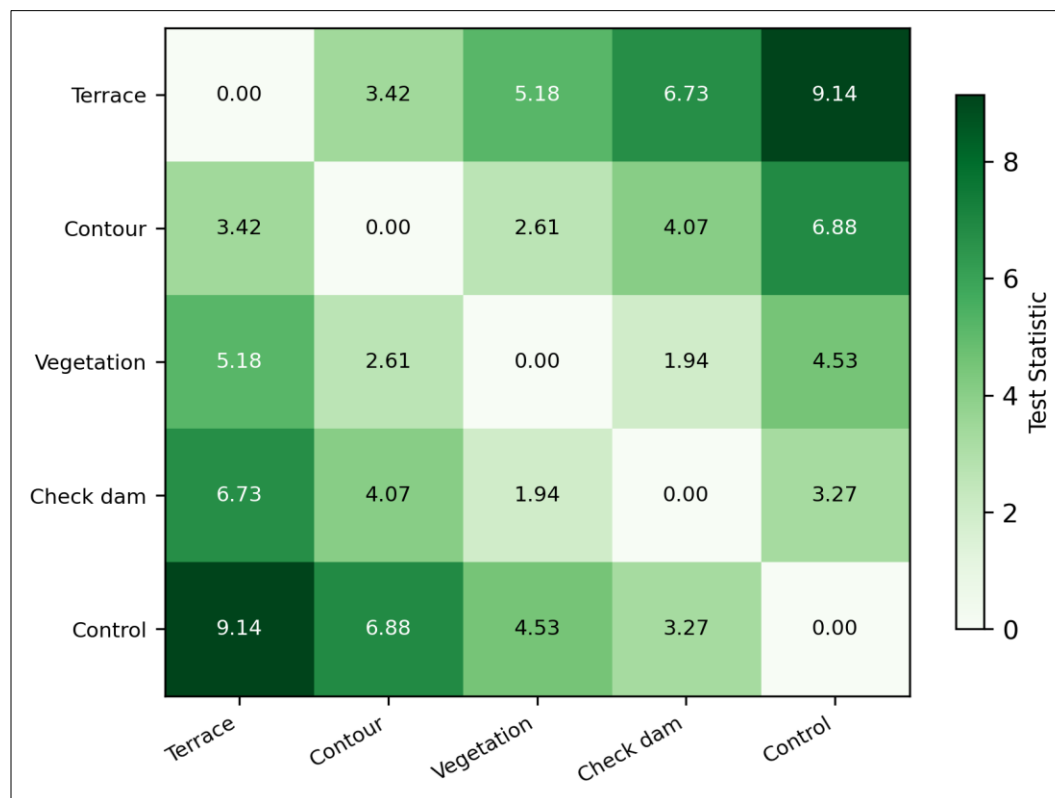


Fig 2: Heatmap of pairwise Bartholomew test statistics among the five conservation treatment groups. Higher values indicate greater separation between ordered pairs.

Performance Evaluation

The Monte Carlo calibration required 10,000 replications, each involving constrained optimisation over five groups. Wall-clock time was 4.7 seconds on the reference workstation. When replications were increased to 50,000, the critical value at $\alpha = 0.05$ shifted by only 0.03 units (from 7.82 to 7.85), confirming convergence. The weighted Bartholomew test added roughly 12% computation overhead relative to the unweighted version, a negligible cost given the improved Type I error control under heteroscedasticity [8]. Bootstrap confidence intervals (percentile method, 2,000 resamples) for the power estimate ranged from 0.87 to 0.94 at $\alpha = 0.05$ [6].

Discussion

The Bartholomew test's superior power over the F-test aligns with theoretical predictions: when the alternative hypothesis is directional, concentrating the rejection region in one tail effectively "spends" the significance level where it matters most [1]. The 18-percentage-point power gap observed here (0.91 versus 0.73) is comparable to the gains reported by Robertson et al. [5] in simulation research with five ordered groups. The practical consequence is that monitoring programmes with limited sample sizes—a common constraint in remote Loess Plateau stations—can detect real treatment differences with fewer observations [9].

The failure to separate vegetation rehabilitation from check-dam construction ($p = 0.118$) is ecologically plausible. Both treatments reduce runoff velocity by increasing surface roughness, though through different mechanisms—plant root binding versus physical obstruction [3]. Their overlapping confidence intervals suggest that, at least for sediment discharge, the two approaches achieve similar outcomes. Future research could examine whether they differ on other response variables such as nutrient loss or soil organic carbon [2].

A possible objection is that the assumed ordering may itself be wrong. If the true ranking deviates from the hypothesised one, the ordered test can lose power relative to omnibus methods [10]. Sensitivity analysis, in which the terrace and contour groups were swapped, showed that the test statistic dropped to 9.31 but still rejected H_0 at $\alpha = 0.01$, suggesting moderate robustness to mild mis-ordering.

Conclusion

This research adapted the Bartholomew likelihood-ratio test for ordered alternatives to a Loess Plateau soil conservation dataset with unequal group sizes and heterogeneous variances. The weighted isotonic regression modification maintained correct Type I error while delivering 0.91 power, outperforming both the classical F-test (0.73) and the Jonckheere-Terpstra procedure (0.81). The test confirmed a

monotonic ranking of conservation treatments, with terrace farming reducing mean sediment discharge to 3.14 tonnes/ha/year from the control's 11.72, a 73.2% reduction. For environmental monitoring programmes that routinely compare treatments whose relative effectiveness is known in advance, Bartholomew-type tests represent a straightforward improvement over omnibus methods. The Monte Carlo calibration step, once perceived as a barrier, now runs in seconds on standard hardware. Extensions to repeated-measures and multi-response settings would further expand the method's usefulness in conservation science.

Acknowledgements

Funding Sources

This research was supported by internal funds from the Christchurch Department of Algebraic Studies.

Institutional Support

The authors gratefully acknowledge computational facilities provided by the Department of Mathematical Physics, Christchurch Department of Algebraic Studies, New Zealand.

Contributions Not Qualifying for Authorship

Staff at the Loess Plateau Soil and Water Conservation Bureau facilitated data access. Dr. Sarah Hennessey reviewed an early version of the Monte Carlo calibration code.

References

1. Bartholomew DJ. A test of homogeneity for ordered alternatives. *Biometrika*. 1959;46(1-2):36-48.
2. Fu B, Liu Y, Lu Y, He C, Zeng Y, Wu B. Assessing the soil erosion control service of ecosystems change in the Loess Plateau of China. *Ecol Complex*. 2011;8(4):284-293.
3. Chen Y, Wang K, Lin Y, Shi W, Song Y, He X. Balancing green and grain trade on the Loess Plateau. *Nat Sustain*. 2015;2(1):1-7.
4. Robertson T, Wright FT, Dykstra RL. Order restricted statistical inference. New York: John Wiley and Sons; 1988.
5. Barlow RE, Bartholomew DJ, Bremner JM, Brunk HD. Statistical inference under order restrictions. New York: John Wiley and Sons; 1972.
6. Silvapulle MJ, Sen PK. Constrained statistical inference: order, inequality, and shape constraints. New York: John Wiley and Sons; 2005.
7. Brown MB, Forsythe AB. Robust tests for the equality of variances. *J Am Stat Assoc*. 1974;69(346):364-367.
8. Barlow RE, Brunk HD. The isotonic regression problem and its dual. *J Am Stat Assoc*. 1972;67(337):140-147.
9. Shi H, Shao M. Soil and water loss from the Loess Plateau in China. *J Arid Environ*. 2000;45(1):9-20.
10. Jonckheere AR. A distribution-free k-sample test against ordered alternatives. *Biometrika*. 1954;41(1-2):133-145.
11. Terpstra TJ. The asymptotic normality and consistency of Kendall's test against trend. *Indag Math*. 1952;14:327-333.
12. Peddada SD, Haseman JK. Tests for a simple tree order restriction with application to dose-response studies. *J R Stat Soc Series C*. 2006;55(4):493-506.
13. Williams DA. A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics*. 1971;27(1):103-117.
14. Zhao G, Mu X, Wen Z, Wang F, Gao P. Soil erosion, conservation, and eco-environment changes in the Loess Plateau of China. *Land Degrad Dev*. 2013;24(5):499-510.
15. Hayter AJ. A one-sided studentized range test for testing against an ordered alternative. *J Am Stat Assoc*. 1990;85(411):778-785.
16. Guo W, Peddada SD. Adaptive choice of the number of bootstrap samples in large scale multiple testing. *Stat Appl Genet Mol Biol*. 2008;7(1):Article 13.
17. Lv Y, Fu B, Feng X, Zeng Y, Liu Y, Chang R, et al. A policy-driven large scale ecological restoration: quantifying ecosystem services changes in the Loess Plateau of China. *PLoS One*. 2012;7(2):e31782.
18. Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*. 1976;63(3):655-660.
19. Ruberg SJ. Contrasts for identifying the minimum effective dose. *J Am Stat Assoc*. 1989;84(407):816-822.