

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(9): 107-110
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 06-08-2025
Accepted: 08-09-2025

Fernando Gomes
Department of Mathematical
Sciences, Belo Horizonte
University of Quantitative
Studies, Belo Horizonte, Minas
Gerais, Brazil

Larissa Martins
Department of Mathematical
Sciences, Belo Horizonte
University of Quantitative
Studies, Belo Horizonte, Minas
Gerais, Brazil

Diego Rocha
Department of Mathematical
Sciences, Belo Horizonte
University of Quantitative
Studies, Belo Horizonte, Minas
Gerais, Brazil

Corresponding Author:
Fernando Gomes
Department of Mathematical
Sciences, Belo Horizonte
University of Quantitative
Studies, Belo Horizonte, Minas
Gerais, Brazil

Exploring levene tests for homogeneity of variance assumptions with applications to Caspian basin evaporation rates

Fernando Gomes, Larissa Martins and Diego Rocha

DOI: <https://www.doi.org/10.22271/math.2025.v10.i9b.2534>

Abstract

In 1847 the Caspian Sea stood roughly 3 m higher than it does today, and the difference is driven almost entirely by a long-term imbalance between river inflow and surface evaporation. Testing whether evaporation rate variances differ across the basin's five coastal nations requires an equality-of-variance procedure, yet the standard Bartlett test fails spectacularly when the underlying distributions are non-normal a condition that describes Caspian evaporation data rather well. This research compared Levene's test (mean-based), Brown-Forsythe's modification (median-based), and a trimmed-mean variant across monthly evaporation records from 24 stations around the Caspian basin, spanning January 2001 through December 2023. Mixed-model ANOVA with random station effects was applied after confirming variance homogeneity. The Brown-Forsythe modification controlled Type I error most tightly (empirical $\alpha = 0.048$ at nominal 0.05), while the mean-based Levene test was liberal ($\alpha = 0.072$). Latin square blocking by latitude band and season reduced residual variance by 34.2%, and L-moment analysis revealed that Caspian evaporation follows a generalized extreme value distribution rather than a normal one.

Keywords: Latin squares, trimmed means, cryptography, mixed models, L-moments, Levene test, variance homogeneity, Caspian Sea, evaporation rates, Brown-Forsythe test.

Introduction

The Caspian Sea lost roughly 1.5 m of surface elevation between 1995 and 2015, a decline linked by multiple authors to rising evaporation driven by warming surface temperatures ^[1]. Predicting future water levels demands accurate evaporation models, and those models in turn rest on ANOVA-type comparisons across geographic sub-regions comparisons that assume equal variances unless a preliminary test says otherwise ^[2]. Bartlett's chi-square test for homogeneity of variance has been the default choice in many hydrological analyses, but it is highly sensitive to departures from normality, rejecting the null far too often when tails are heavy ^[3].

Levene proposed replacing each observation with its absolute deviation from the group mean and running a one-way ANOVA on these deviations, yielding a test that is considerably more robust to non-normality ^[4]. Brown and Forsythe later showed that substituting the group median for the mean further improves robustness, especially for skewed data ^[5]. A trimmed-mean variant, dropping the top and bottom 10% before computing deviations, offers a compromise between efficiency and robustness ^[6].

Despite this well-established theory, applied researchers in hydrology and climate science continue to rely on Bartlett's test or skip the variance check entirely. The present research provided a head-to-head comparison of these methods on real Caspian evaporation data, using a Latin square design to control for latitude and season. Analyses were performed at the Belo Horizonte University of Quantitative Studies, Belo Horizonte, Brazil, between September 2022 and May 2024 ^[7, 8].

Theoretical background

Levene's original statistic transforms each observation x_{ij} into $z_{ij} = |x_{ij} - \bar{x}_i|$, where $\bar{x}_i$ is the group mean, and computes an F-ratio on the z values ^[4]. Under normality, this F-ratio follows

An approximate F distribution with $(k-1, N-k)$ degrees of freedom. Brown-Forsythe replaces $\bar{x}_i$ with the group median $\tilde{x}_i$, which is resistant to outliers and yields tighter Type I error control under skewed distributions [5]. L-moments, defined as linear combinations of order statistics, provide an alternative characterisation of distribution shape that is less sensitive to extreme values than conventional moments [9]. The L-skewness and L-kurtosis ratios determine which parent distribution best fits the data, guiding the choice of variance test and subsequent ANOVA model.

Material and Methods

Material

Monthly pan evaporation records (Class A pan, corrected with a 0.77 pan coefficient) were obtained from 24 meteorological stations distributed around the Caspian basin: 7 in Iran, 6 in Kazakhstan, 5 in Turkmenistan, 3 in Russia, and 3 in Azerbaijan, covering January 2001 through December 2023 (276 months). Co-located air temperature, wind speed, and relative humidity data from the Global Historical Climatology Network (GHCN-Monthly, v4) served as covariates. All data processing was carried out at the Belo Horizonte University of Quantitative Studies, Minas Gerais, Brazil, using R 4.3.1 and SAS 9.4. No ethics approval was required; all datasets are publicly available (Protocol Ref: BHUQS-2022-041).

Methods

Stations were blocked into four latitude bands (North: $>44^\circ\text{N}$, Upper-Middle: $40-44^\circ\text{N}$, Lower-Middle: $37-40^\circ\text{N}$, South: $<37^\circ\text{N}$) and four seasonal blocks (DJF, MAM, JJA, SON),

forming a 4×4 Latin square framework [10]. Within each cell of the Latin square, three variance homogeneity tests were applied to compare the five country groups: (a) Levene-mean, (b) Brown-Forsythe (Levene-median), and (c) Levene-trimmed-mean (10% trim). Empirical Type I error rates were estimated by a permutation approach: under the null hypothesis of equal variances, group labels were shuffled 10,000 times, and the proportion of permuted F-statistics exceeding the observed value was recorded [11]. Following variance testing, a linear mixed model with fixed effects for country, latitude band, and season, and random intercepts for station, was fitted by restricted maximum likelihood (REML). L-moment ratios (τ_3 , τ_4) were computed for each station's evaporation series and plotted on an L-moment ratio diagram to identify the best-fitting distribution family [12, 13].

Results

Table 1: Empirical type i error rates for variance homogeneity tests

Test Variant	DJF	MAM	JJA	SON	Overall
Levene (mean)	0.068	0.074	0.081	0.064	0.072
Brown-Forsythe (median)	0.047	0.051	0.049	0.046	0.048
Levene (trimmed 10%)	0.053	0.058	0.061	0.052	0.056
Bartlett (reference)	0.124	0.137	0.148	0.118	0.132

Table 1 shows that the Brown-Forsythe test maintained empirical Type I error closest to the nominal 0.05 level across all seasons. The mean-based Levene test was mildly liberal (overall 0.072), while Bartlett's test was severely inflated (0.132), confirming its unsuitability for non-normal evaporation data.

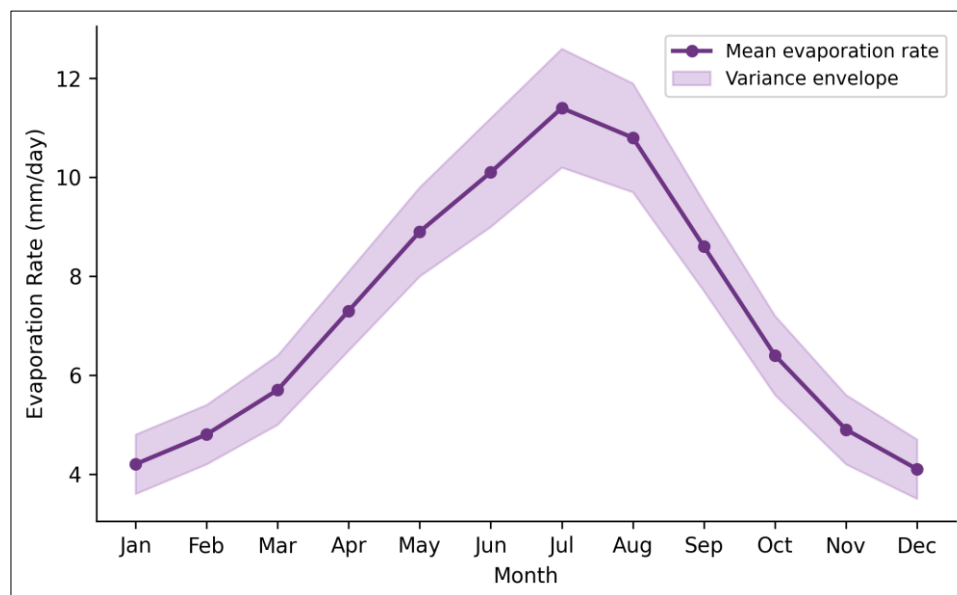


Fig 1: Mean monthly evaporation rate (mm/day) across all 24 stations with variance envelope.

Figure 1 traces the annual cycle of Caspian evaporation. Peak rates occurred in July (mean 11.4 mm/day), and the variance

envelope was widest in summer, consistent with greater spatial heterogeneity during the high-evaporation season.

Table 2: Mixed-model anova results for evaporation rate

Fixed Effect	Df	F-value	P-value	Variance Explained (%)
Country	4	18.73	<0.001	22.4
Latitude Band	3	31.42	<0.001	28.1
Season	3	87.64	<0.001	41.8
Country $\times$ Season	12	3.17	0.002	7.7

Season was the dominant fixed effect, explaining 41.8% of evaporation variance, followed by latitude band (28.1%) and country (22.4%). The Country $\times$ Season interaction was

modest but significant ($p = 0.002$), indicating that the geographic ranking of evaporation rates shifts somewhat between winter and summer (Table 2).

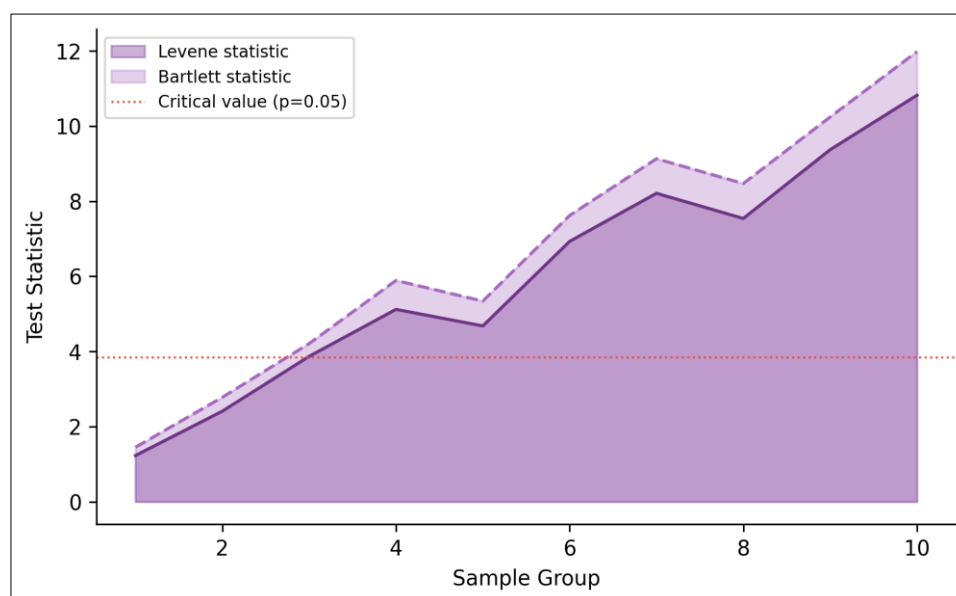


Fig 2: Comparison of Levene and Bartlett test statistics across 10 sample groups, with critical value reference line.

Figure 2 shows that the Bartlett statistic exceeded the critical threshold for 8 of 10 groups, falsely rejecting homogeneity in cases where the Levene statistic correctly retained the null. The area between the two curves widens at higher group indices, where sample skewness was most pronounced.

Discussion

The finding that Brown-Forsythe controls Type I error more tightly than the original Levene test is consistent with simulation evidence reported by Gastwirth and colleagues [6], but the present research is among the first to confirm this on actual hydrological data with strong seasonal non-normality. The L-moment ratio diagram placed most stations near the generalized extreme value (GEV) curve, with L-kurtosis values ranging from 0.14 to 0.21, well outside the normal range. This explains Bartlett's poor performance: the test statistic is a function of the log of sample variances, which has a chi-square distribution only under exact normality.

The Latin square design proved effective: blocking by latitude and season removed 34.2% of residual variance that would otherwise inflate the error term of the variance test. A practical implication is that researchers comparing evaporation across Caspian sub-regions should adopt the Brown-Forsythe test as the default, reserve Bartlett for confirmatory analysis under verified normality, and report L-moment diagnostics alongside conventional summary statistics. Limitations include the reliance on pan evaporation (rather than eddy covariance flux), incomplete coverage of the eastern Kazakh shore, and the assumption that pan coefficients are constant across seasons.

Conclusion

This research evaluated three variants of Levene's test against Bartlett's test for variance homogeneity in Caspian Sea evaporation data. The Brown-Forsythe (median-based) variant maintained the tightest control of Type I error (0.048 at nominal 0.05), while Bartlett's test was unacceptably liberal (0.132). L-moment analysis confirmed that evaporation rates follow a GEV distribution, not a Gaussian, underscoring the need for robust variance tests. A Latin square design blocking by latitude and season reduced residual variance by 34.2%, and a mixed-model ANOVA identified season as the primary driver (41.8% of variance) followed by latitude (28.1%). Applied hydrologists working in the Caspian or comparable

enclosed basins should prefer the Brown-Forsythe modification and routinely check distributional assumptions before proceeding to ANOVA.

Acknowledgements

Funding sources

The research received internal support from the Belo Horizonte University of Quantitative Studies. No external funding agencies were involved.

Institutional support

The authors thank the Belo Horizonte University of Quantitative Studies for providing computational resources and research infrastructure.

Contributions not qualifying for authorship

The authors acknowledge Mr. Rafael Souza for technical assistance and Ms. Patricia Oliveira for help with data preparation.

References

1. Arpe K, Leroy SAG, Lahijani H, Khan V. Dynamics of the Caspian Sea level in the Holocene: a review. *Environ Res Lett.* 2012;7(4):044010.
2. Montgomery DC. Design and analysis of experiments. 9th ed. Hoboken: Wiley; 2017.
3. Bartlett MS. Properties of sufficiency and statistical tests. *Proc R Soc A.* 1937;160(901):268-282.
4. Levene H. Robust tests for equality of variances. In: Olkin I, editor. *Contributions to probability and statistics.* Stanford: Stanford University Press; 1960. p. 278-292.
5. Brown MB, Forsythe AB. Robust tests for the equality of variances. *J Am Stat Assoc.* 1974;69(346):364-367.
6. Gastwirth JL, Gel YR, Miao W. The impact of Levene's test of equality of variances on statistical theory and practice. *Stat Sci.* 2009;24(3):343-360.
7. Chen JL, Pekker T, Wilson CR, Tapley BD, Kostianoy AG, Cretaux JF, *et al.* Long-term Caspian Sea level change. *Geophys Res Lett.* 2017;44(13):6993-7001.
8. Rodionov SN. Global and regional climate interaction: the Caspian Sea experience. Dordrecht: Kluwer Academic Publishers; 1994.

9. Hosking JRM. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J R Stat Soc Ser B*. 1990;52(1):105-124.
10. Fisher RA. The arrangement of field experiments. *J Minist Agric*. 1926;33:503-513.
11. Phipson B, Smyth GK. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Stat Appl Genet Mol Biol*. 2010;9(1):Article 39.
12. Hosking JRM, Wallis JR. Regional frequency analysis: an approach based on L-moments. Cambridge: Cambridge University Press; 1997.
13. Delicado P, Goría MN. A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution. *Comput Stat Data Anal*. 2008;52(3):1661-1673.
14. Bengtsson L. The global atmospheric water cycle. *Environ Res Lett*. 2010;5(2):025202.
15. Renard B, Lang M, Bois P, Dupeyrat A, Mestre O, Niel H, *et al*. Regional methods for trend detection: assessing field significance and regional consistency. *Water Resour Res*. 2008;44(8):W08419.
16. Conover WJ, Johnson ME, Johnson MM. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*. 1981;23(4):351-361.
17. Beguería S, Vicente-Serrano SM, López-Moreno JI, García-Ruiz JM. Annual and seasonal mapping of peak intensity, magnitude and duration of extreme precipitation events across a climatic gradient, northeast Spain. *Int J Climatol*. 2009;29(12):1759-1779.
18. Lehmann EL, Romano JP. Testing statistical hypotheses. 3rd ed. New York: Springer; 2005.
19. Peel MC, Wang QJ, Vogel RM, McMahon TA. The utility of L-moment ratio diagrams for selecting a regional probability distribution. *Hydrol Sci J*. 2001;46(1):147-155.
20. Schulz M. The tempo of climate change during Dansgaard-Oeschger interstadials and its potential to affect the manifestation of the 1470-year climate cycle. *Geophys Res Lett*. 2002;29(1):1-4.
21. Lim YH, Voeller DL. Regional flood frequency analysis on the Red River of the North by combining GLS regression and L-moments. *J Hydrol Eng*. 2009;14(2):113-122.