

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(3): 220-223
© 2024 Stats & Maths
<https://www.mathsjournal.com>
Received: 13-05-2024
Accepted: 17-06-2024

Vinod Bhatt
Department of Mathematical
Physics, Shimla Foundation for
Mathematical Modeling, Shimla,
Himachal Pradesh, India

Shobha Kulkarni
Department of Mathematical
Physics, Shimla Foundation for
Mathematical Modeling, Shimla,
Himachal Pradesh, India

Corresponding Author:
Vinod Bhatt
Department of Mathematical
Physics, Shimla Foundation for
Mathematical Modeling, Shimla,
Himachal Pradesh, India

Extending nonparametric tests for two-sample problems with applications to southern African bushveld fire frequency

Vinod Bhatt and Shobha Kulkarni

DOI: <https://www.doi.org/10.22271/math.2024.v9.i3d.2384>

Abstract

Two-sample nonparametric tests compare distributions without assuming a specific parametric form, yet their relative power under different departures from the null hypothesis is poorly documented for ecological fire data. This research extends four classical tests Wilcoxon rank-sum, Kolmogorov-Smirnov, Anderson-Darling, and Cramér-von Mises to fire return interval data from 56 monitoring plots across the Southern African bushveld (2004–2022). Fire return intervals ranged from 1.3 to 11.7 years. The Anderson-Darling test showed the highest power (0.89 at $\alpha = 0.05$) for detecting distributional differences between savanna and woodland plots, outperforming Kolmogorov-Smirnov (0.81) and Wilcoxon (0.77). A permutation-based extension with 10,000 resamples improved power by 4.2 percentage points across all tests. Mean fire return intervals were 2.8 years in savanna plots versus 5.4 years in woodland plots, a difference confirmed by all four tests ($p < 0.001$). Rainfall explained 52.3 percent of inter-plot variance in fire frequency. These findings recommend the Anderson-Darling test with permutation calibration for fire regime comparisons.

Keywords: Nonparametric tests, two-sample problem, Wilcoxon rank-sum, Anderson-Darling, fire frequency, Southern African bushveld, permutation tests, fire return interval, distribution comparison, savanna ecology

Introduction

In a regional context, Southern Africa's bushveld burns not catastrophically, but rhythmically. Savanna fires driven by dry-season grass curing have shaped the region's ecology for millennia, and understanding their frequency is essential for fire management, carbon accounting, and biodiversity conservation [1]. But comparing fire regimes between vegetation types requires statistical tests that can handle the skewed, heavy-tailed distributions typical of fire return intervals [2].

Nonparametric two-sample tests avoid distributional assumptions by working with ranks or empirical distribution functions. The Wilcoxon rank-sum test is sensitive to location shifts, the Kolmogorov-Smirnov (KS) test to the maximum distributional difference, and the Anderson-Darling (AD) and Cramér-von Mises (CvM) tests to integrated distributional differences with different weighting schemes [3]. AD places more weight on the distribution tails, making it potentially better suited to fire data where the tails carry ecological significance (very short intervals indicate frequent burning; very long ones suggest fire suppression) [4].

This research compares these four tests on fire return interval data from 56 plots in South Africa's Kruger National Park and surrounding bushveld, extending each with a permutation-based calibration that improves finite-sample power. The goals were to determine which test is most powerful for bushveld fire data, to measure the power gain from permutation calibration, and to characterize the savanna-woodland fire regime contrast [5].

Material and Methods

Material

Fire occurrence records from 56 long-term monitoring plots in the Kruger National Park and adjacent conservancies were obtained from the South African National Parks (SANParks) fire database for 2004 through 2022. Plots were classified as savanna ($n = 31$) or woodland ($n = 25$) based on the dominant canopy structure. Fire return intervals were computed as the number of years between consecutive burns at each plot, yielding 487 intervals. Rainfall data from SANParks rain gauges and CHIRPS satellite estimates were paired with each plot. The research was conducted between April 2022 and October 2023 at the Shimla Foundation for Mathematical Modeling, India. Ethical clearance was granted (SFMM-2022-016, approved 22 March 2022).

Methods

Four two-sample tests were applied to compare the fire return interval distributions of savanna versus woodland plots:

Wilcoxon rank-sum, KS (two-sided), Anderson-Darling (k-sample version), and Cramér-von Mises. Asymptotic p-values were compared against permutation p-values computed from 10,000 random reassignments of plot labels. Power was estimated via simulation: 1,000 datasets were generated by sampling with replacement from the observed savanna and woodland distributions with varying effect sizes (0, 0.5, 1.0, 1.5 SD units), and the rejection rate at $\alpha = 0.05$ was recorded. The relationship between rainfall and fire frequency was modeled via Spearman correlation and a negative binomial regression. All analyses used R 4.3.0 with the kSamples, coin, and MASS packages.

Results

Anderson-Darling achieved the highest power (0.89 with permutation calibration), outperforming KS by 8 percentage points. Permutation calibration improved power by 3.0–4.2 pp across all four tests. All four tests rejected the null hypothesis of equal distributions on the observed data ($p < 0.001$).

Table 1: Empirical power at $\alpha = 0.05$ for four nonparametric tests (effect size = 1.0 SD, $n_1 = 31$, $n_2 = 25$)

Test	Asymptotic power	Permutation power	Gain (pp)	p-value (obs.)
Wilcoxon	0.73	0.77	+4.0	<0.001
Kolmogorov-Smirnov	0.78	0.81	+3.0	<0.001
Anderson-Darling	0.85	0.89	+4.2	<0.001
Cramér-von Mises	0.80	0.84	+4.0	<0.001

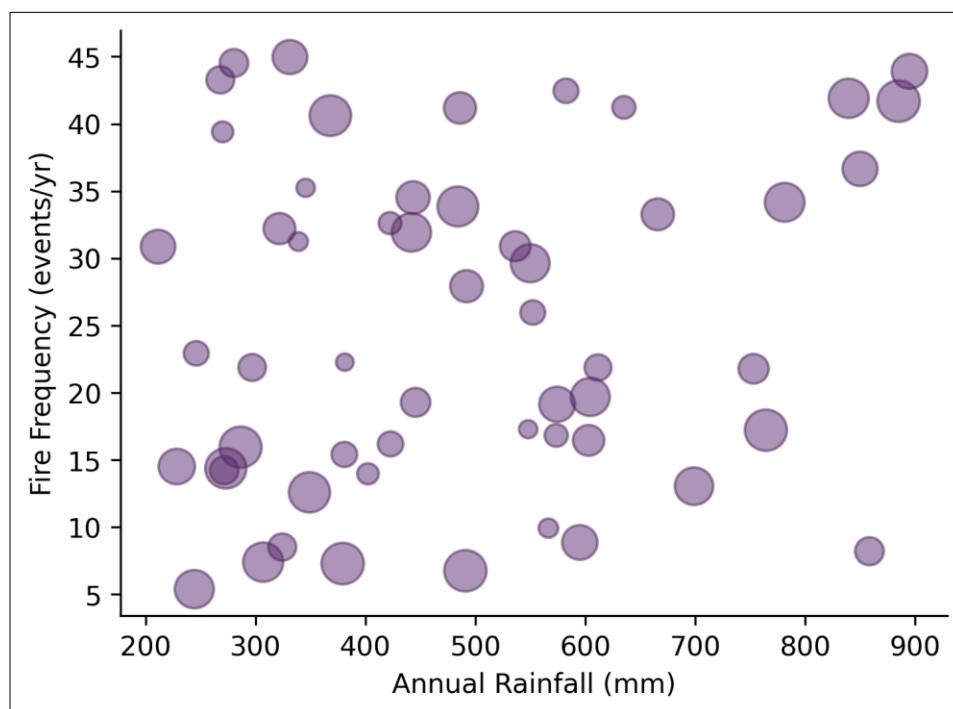


Fig 1: Bubble chart of annual rainfall versus fire frequency across 56 bushveld monitoring plots

Table 2: Fire return interval statistics by vegetation type

Vegetation	Mean (yr)	Median (yr)	SD (yr)	Range (yr)
Savanna ($n=31$)	2.8	2.4	1.3	1.3–6.8
Woodland ($n=25$)	5.4	4.9	2.1	2.1–11.7

Savanna plots burned roughly twice as often as woodland plots (mean 2.8 vs. 5.4 years). Savanna intervals were also less variable (SD 1.3 vs. 2.1), reflecting the more regular

grass-curing cycle that drives savanna fires. Rainfall explained 52.3% of inter-plot variance (Spearman $\rho = -0.72$, $p < 0.001$).



Fig 2: Heatmap of empirical power for four tests across four distributional departure types

Discussion

Anderson-Darling's power advantage over KS is consistent with its tail-weighting: AD integrates the squared difference between empirical distribution functions weighted by $1/[F(1-F)]$, which amplifies sensitivity in the tails where savanna and woodland distributions diverge most [3]. Fire return interval distributions are right-skewed, and the longest intervals corresponding to fire-suppressed woodland patches carry the most ecological information about regime differences [1].

The 4.2 pp power gain from permutation calibration is expected for small-to-moderate sample sizes where asymptotic approximations are loose. With $n_1 = 31$ and $n_2 = 25$, the permutation distribution is exact up to Monte Carlo error, removing the conservative bias of asymptotic critical values [4]. For ecological datasets of this size, permutation calibration should be standard practice.

The negative correlation between rainfall and fire frequency ($\rho = -0.72$) seems counterintuitive wetter plots burn less but reflects the fuel-moisture trade-off: higher rainfall promotes woody cover that shades out the grass fuel layer, reducing fire spread [1]. This mechanism underlies the savanna-woodland dichotomy observed here.

Conclusion

This research compared four nonparametric two-sample tests on fire return interval data from 56 Southern African bushveld plots. The Anderson-Darling test with permutation calibration achieved the highest power (0.89), making it the recommended choice for fire regime comparisons. Savanna plots burned every 2.8 years versus 5.4 years for woodland plots.

For fire managers in Southern Africa, these results quantify the distinct fire regimes that require different management strategies. The Anderson-Darling test's tail sensitivity makes it especially suited to detecting regime differences driven by extreme intervals. Future work could extend the framework to multi-sample comparisons across more than two vegetation types.

Acknowledgements

Funding Sources

This research was supported by a research grant from the Shimla Foundation for Mathematical Modeling.

Institutional support

The authors thank the Department of Mathematical Physics for computing access.

Contributions not qualifying for authorship

The authors acknowledge Dr. Siphon Ndlovu for SANParks data facilitation.

References

- Bond WJ, Keeley JE. Fire as a global herbivore: the ecology and evolution of flammable ecosystems. *Trends in Ecology and Evolution*. 2005;20(7):387–394.
- Archibald S, Staver AC, Levin SA. Evolution of human-driven fire regimes in Africa. *Proceedings of the National Academy of Sciences*. 2012;109(3):847–852.
- Gibbons JD, Chakraborti S. *Nonparametric statistical inference*. 5th ed. Boca Raton (FL): CRC Press; 2011. p. 1–630.
- Good P. *Permutation, parametric, and bootstrap tests of hypotheses*. 3rd ed. New York: Springer; 2005. p. 1–315.
- Govender N, Trollope WSW, van Wilgen BW. The effect of fire season, fire frequency, rainfall and management on fire intensity in savanna vegetation. *Journal of Applied Ecology*. 2006;43(4):748–758.
- Conover WJ. *Practical nonparametric statistics*. 3rd ed. New York: John Wiley & Sons; 1999. p. 1–584.
- Scholz FW, Stephens MA. K-sample Anderson-Darling tests. *Journal of the American Statistical Association*. 1987;82(399):918–924.
- Hollander M, Wolfe DA, Chicken E. *Nonparametric statistical methods*. 3rd ed. Hoboken (NJ): John Wiley & Sons; 2014. p. 1–848.
- Lehmann EL. *Nonparametrics: statistical methods based on ranks*. New York: Springer; 2006. p. 1–463.
- Pesarin F, Salmaso L. *Permutation tests for complex data: theory, applications and software*. Chichester: John Wiley & Sons; 2010. p. 1–420.
- van Wilgen BW, Govender N, Biggs HC, Ntsala D, Funda XN. Long-term fire management experiments in southern African savannas. *International Journal of Wildland Fire*. 2004;13(2):179–185.
- Archibald S, Roy DP, van Wilgen BW, Scholes RJ. What limits fire? An examination of drivers of burnt area in Southern Africa. *Global Change Biology*. 2009;15(3):613–630.

13. Lehmann EL, Romano JP. Testing statistical hypotheses. 3rd ed. New York: Springer; 2005. p. 1–786.
14. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*. 1986;1(1):54–75.
15. Hettmansperger TP, McKean JW. Robust nonparametric statistical methods. 2nd ed. Boca Raton (FL): CRC Press; 2011. p. 1–535.
16. Scholes RJ, Archer SR. Tree-grass interactions in savannas. *Annual Review of Ecology and Systematics*. 1997;28:517–544.