

A Novel Test for Comparing Two Populations Based on Principal Points

Furqan Al-Hossein

alhussen12346@gmail.com

Abstract

In this paper, we present a novel approach for comparing two statistical populations based on their principal points. Principal points are a set of locations that minimize the average distance to the data and, as a generalization of the mean, provide richer information about the structure and dispersion of the data [14]. We review classical and modified definitions and properties of principal points, and introduce a new test based on the distance between the principal points of two populations to assess the equality of their distributions. The proposed test is implemented using the bootstrap method, and its performance is evaluated under various scenarios, including equality and inequality of means, variances, and in the presence of multimodality. Simulation results demonstrate that the principal point-based test offers higher power than the classical t-test in detecting structural differences between two populations, especially in situations where the t-test fails (such as differences in variance or multimodality). Furthermore, empirical evidence using real datasets shows that the principal point test can reveal hidden differences in data structure. These features make principal points a powerful and flexible tool for comparing statistical populations, particularly in the analysis of complex and non-normal data.

1. Introduction

Comparing two populations is a fundamental problem in statistics, often reduced to testing the equality of means:

$$H_0: \mu_1 = \mu_2 \quad \text{vs.} \quad H_1: \mu_1 \neq \mu_2,$$

where μ_1 and μ_2 are the population means. The classical two-sample t-test assumes normality and equal variances, with the statistic [15, 17]:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{1/n_1 + 1/n_2}},$$

10.48047/jocaaa.2025.34.08.13

where \bar{X}_1, \bar{X}_2 are sample means, n_1, n_2 sample sizes, and S_p^2 the pooled variance. Under the assumptions, T follows a t-distribution. However, violations of normality or equal variances can lead to incorrect inference, and the test ignores differences in variance, shape, or multimodality.

To address these limitations, we propose a test based on principal points, which generalize the mean by capturing distributional structure [4, 14]. Principal points are k points minimizing the expected squared distance to the random variable, providing information on location, dispersion, and shape. We estimate principal points from samples and use the bootstrap method [3, 6] for inference, as exact distributions are intractable. The bootstrap resamples data to approximate sampling distributions, enabling variance estimation and hypothesis testing without strong parametric assumptions.

Simulation studies show our test has higher power than the t-test for detecting non-mean differences, while maintaining Type I error. Application to the Boston housing dataset reveals hidden structures missed by traditional tests. The paper is organized as follows: Section 2 details methods, including principal points and bootstrap; Section 3 presents simulations; Section 4 the real data analysis; Section 5 concludes.

2. Methods

2.1. Principal Points: Definition, Estimation, and Properties

Let X be a random variable with distribution function F . The k principal points $\xi = (\xi_1, \dots, \xi_k)^\top$, with $\xi_1 \leq \dots \leq \xi_k$, minimize the expected mean squared error [4, 9]:

$$\xi = \operatorname{argmin}_y \mathbb{E} \min_{j=1, \dots, k} (X - y_j)^2.$$

Equivalently, they are the means of optimal partitions of the support into k intervals, satisfying self-consistency: each ξ_j is the conditional expectation of X in its Voronoi cell.

For univariate symmetric distributions, [4] and [12] showed that for $k = 2$, $\xi_{1,2} = \pm \mathbb{E}|X|$ if $f(0)\mathbb{E}|X| < 0.5$, where f is the density. For the standard normal, the two principal points are approximately ± 0.612 .

Estimation:

Given a sample X_1, \dots, X_n i. i. d. F , the sample principal points $\hat{\xi}$ are the cluster centers from the k-means algorithm applied to the data, minimizing [11]:

$$\sum_{i=1}^n \min_{j=1, \dots, k} (X_i - y_j)^2.$$

The k-means algorithm initializes centers, assigns points to nearest centers, updates centers as group means, and iterates until convergence.

Theoretical properties:

Under mild conditions (e.g., finite second moment, continuous F), the sample principal points are strongly consistent: $\hat{\xi} \rightarrow \xi$ almost surely as $n \rightarrow \infty$ [10, 1]. For asymptotic normality, under additional regularity (e.g., twice differentiable density, unique optimal partition), $\sqrt{n}(\hat{\xi} - \xi) \rightarrow_d N(\mathbf{0}, \Sigma)$, where Σ depends on the distribution and partition boundaries [14, 13]. However, due to potential non-uniqueness or boundary issues, we use bootstrap for variance estimation.

The covariance matrix Σ can be complex; for example, in normal distributions, explicit forms exist, but generally, bootstrap provides a non-parametric estimate.

2.2. Bootstrap Inference and the Proposed Test

The bootstrap approximates the distribution of estimators by resampling. For a statistic $T(\mathbf{X})$, draw B resamples \mathbf{X}^{*b} with replacement, compute $T^{*b} = T(\mathbf{X}^{*b})$, and use the empirical distribution of $\{T^{*b}\}$ for inference [3, 6].

For principal points, to estimate $\text{Cov}(\hat{\xi})$, compute $\hat{\xi}^{*b}$ from each resample and take the sample covariance of $\{\hat{\xi}^{*b}\}$.

The proposed test compares two populations via their principal points. Let $\hat{\xi}_1, \hat{\xi}_2$ be estimates from samples of sizes n_1, n_2 . The test statistic is the Mahalanobis distance

$$D = (\hat{\xi}_1 - \hat{\xi}_2)^\top \hat{\Sigma}^{-1} (\hat{\xi}_1 - \hat{\xi}_2),$$

where $\hat{\Sigma}$ is the pooled covariance from bootstrap on each sample.

10.48047/jocaaa.2025.34.08.13

Under H_0 : equal distributions, pool samples, resample two groups of sizes n_1, n_2 , compute D^{*b} , and obtain p-value as proportion of $D^{*b} > D$.

This bootstrap ensures validity without parametric assumptions, leveraging the asymptotic properties of principal points estimators.

3. Simulation Studies

We assess the performance of the principal points (PP) test, leveraging its bootstrap-based estimation and ability to capture distributional features beyond means, through simulations that compare it to the t-test under controlled scenarios. The evaluation begins with normal populations to highlight PP's advantages in detecting mean and variance differences, followed by non-normal distributions to demonstrate its sensitivity to shape variations.

3.1. Comparison of t-Test and Principal Points for Normal Populations

To evaluate the performance of the bootstrap t-test and the PP test in comparing two normal populations, simulations were conducted with $\mu_1 = 0$, $\sigma_1^2 = 1$, and varying $\mu_2 \in \{0,1\}$, $\sigma_2^2 \in \{0.25,1,4\}$, using sample sizes $n_1, n_2 \in \{30,50,100\}$ [7, 2]. The tests employed $B = 1000$ bootstrap samples and $M = 3000$ repetitions to estimate rejection probabilities at $\alpha = 0.05$. Results are presented in Table 1. For the t-test under the null hypothesis ($\mu_2 = 0$), Type I error rates are generally conservative (0.000–0.034) when variances are smaller ($\sigma_2^2 = 0.25$) and closer to the nominal level (0.023–0.057) when equal ($\sigma_2^2 = 1$), but inflate slightly (0.058–0.081) with larger variances ($\sigma_2^2 = 4$), indicating sensitivity to heteroscedasticity. Under the alternative ($\mu_2 = 1$), the t-test shows high power (0.629–1), though it drops significantly with unequal variances (e.g., 0.629 for $\sigma_2^2 = 4$, $n_1 = n_2 = 100$), reflecting its reliance on equal variance assumptions. In contrast, the PP test with $k = 2$ and $k = 3$ demonstrates robust Type I error control when variances are equal (0.021–0.093 for $k = 2$, 0.020–0.136 for $k = 3$), though it inflates with unequal sample sizes [1, 14], suggesting sensitivity to balance. Remarkably, under $\mu_2 = 0$ with unequal variances, PP rejection rates rise substantially (e.g., 0.765–0.998 for $k = 2$, $\sigma_2^2 = 0.25$; 0.656–0.995 for $\sigma_2^2 = 4$), highlighting its ability to detect variance differences where the t-test fails. For $\mu_2 = 1$, PP power is competitive (0.458–0.982 for $k = 2$, 0.288–0.98 for $k = 3$), improving with sample size and k , and remains effective across variance scenarios, often

10.48047/jocaaa.2025.34.08.13

exceeding the t-test when variances differ (e.g., 0.947 vs. 0.629 for $k = 2$, $\sigma_2^2 = 4$, $n_1 = n_2 = 100$). These results underscore the PP method’s superiority, capturing both mean and variance differences with greater flexibility, making it a powerful alternative to the t-test in normal settings.

Table 1. Type I Error and Power of Bootstrap t-Test and Principal Points Test for Comparing Two Normal Populations

Bootstrap t-Test										
μ_2	n_1/n_2	$\sigma_2 = 0.5$			$\sigma_2 = 1$			$\sigma_2 = 2$		
		30	50	100	30	50	100	30	50	100
0	30	0.004	0.017	0.034	0.023	0.042	0.057	0.058	0.07	0.074
	50	0	0.002	0.014	0.008	0.022	0.047	0.052	0.061	0.081
	100	0	0	0.002	0.001	0.006	0.017	0.029	0.041	0.071
1	30	0.942	0.989	0.999	0.918	0.986	0.998	0.702	0.876	0.968
	50	0.965	0.997	1	0.934	0.995	1	0.7	0.895	0.99
	100	0.947	1	1	0.894	0.996	1	0.629	0.898	0.997
Principal Points Test ($k = 2$)										
μ_2	n_1/n_2	$\sigma_2 = 0.5$			$\sigma_2 = 1$			$\sigma_2 = 2$		
		30	50	100	30	50	100	30	50	100
0	30	0.545	0.629	0.765	0.021	0.023	0.035	0.52	0.558	0.656
	50	0.748	0.832	0.919	0.035	0.023	0.039	0.806	0.847	0.924
	100	0.938	0.992	0.998	0.093	0.055	0.024	0.967	0.989	0.995
1	30	0.607	0.681	0.651	0.458	0.539	0.611	0.487	0.541	0.698
	50	0.671	0.843	0.898	0.626	0.749	0.848	0.595	0.673	0.864
	100	0.712	0.896	0.982	0.872	0.906	0.971	0.846	0.857	0.947
Principal Points Test ($k = 3$)										
μ_2	n_1/n_2	$\sigma_2 = 0.5$			$\sigma_2 = 1$			$\sigma_2 = 2$		
		30	50	100	30	50	100	30	50	100
0	30	0.208	0.292	0.361	0.03	0.025	0.036	0.193	0.204	0.256
	50	0.405	0.437	0.542	0.052	0.02	0.025	0.382	0.4	0.509
	100	0.688	0.778	0.886	0.136	0.057	0.026	0.804	0.785	0.869
1	30	0.561	0.59	0.801	0.313	0.288	0.328	0.384	0.387	0.486
	50	0.649	0.772	0.906	0.445	0.42	0.455	0.603	0.651	0.768
	100	0.826	0.917	0.98	0.65	0.642	0.734	0.867	0.914	0.972

3.2. Principal Points for Detecting Shape Differences in Non-Normal Populations

Building on the superior performance of the PP test over the t-test in detecting mean and variance differences within normal populations, we now explore its ability to identify shape disparities in non-normal distributions [8], a domain where traditional tests like the t-test are inherently limited. Simulations compared gamma and log-normal populations with equal means but varying variances, as outlined in Table 2. The parameter settings ensure $\mathbb{E}(X) \approx \mathbb{E}(Y)$ across cases (e.g., 2 for cases 1 and 3, 1.5 for cases 2 and 4), with variances either closely matched (cases 1 and 2) or moderately unequal (cases 3 and 4), as depicted in the density plots of Figure 1, which highlight distinct skewness and tail behaviors despite mean equality. For $n_1 = n_2 = 100$ and $M = 100$ tests, rejection rates at $\alpha = 0.05$ are reported in Table 3. The t-test, designed to assess mean equality, shows negligible rejection rates (3%–7%), failing to detect any differences since means are identical, regardless of variance or shape. Conversely, the PP test, with $k = 2$ and $k = 5$, rejects the null hypothesis at significantly higher rates—27% to 46% for $k = 2$ and 64% to 92% for $k = 5$ —even in cases of equal variances (e.g., 68% for case 1 with $k = 5$). Rejection rates increase with unequal variances (e.g., 87% and 92% for cases 3 and 4 with $k = 5$) and higher k , reflecting heightened sensitivity to skewness, kurtosis, and overall distributional form. This capability allows the PP test to uncover structural differences that the t-test misses, such as those arising from differing family memberships (gamma vs. log-normal), making it a robust tool for analyzing complex, non-normal data where shape variations are critical.

Table 2. Parameter values of Gamma and Log-Normal distributions in four scenarios with equal means and two cases of equal and unequal variances

case	α	λ	μ	σ	$\mathbb{E}(X)$	$\mathbb{V}(X)$	$\mathbb{E}(Y)$	$\mathbb{V}(Y)$	Variances
1	1	0.5	0.35	0.69	2	4	2.004	3.990	equal
2	3	2	0.26	0.29	1.5	0.75	1.499	0.756	
3	1	0.5	0.44	0.51	2	4	2.004	2.671	not equal
4	3	2	0.20	0.41	1.5	0.75	1.499	1.139	

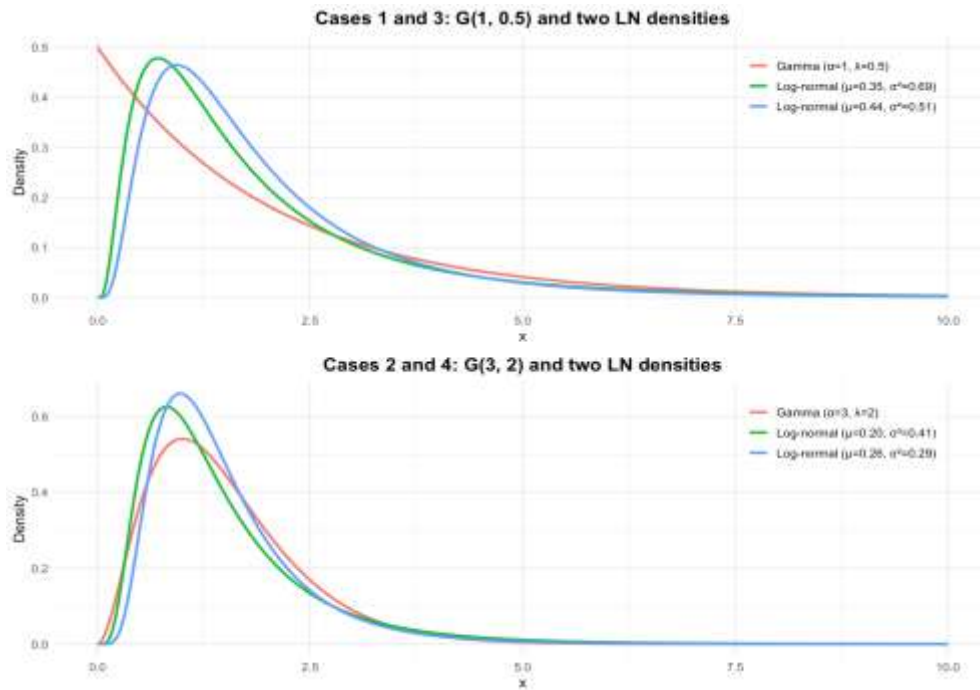


Figure 1. Density plots of Gamma distribution with parameters specified in Table 2 alongside corresponding Log-Normal densities

Table 3. Percentage of null hypothesis rejection in mean equality and principal points tests for $n = 100$

Case	t-Test	Principal Points ($k = 2$)	Principal Points ($k = 5$)
1	3%	27%	68%
2	4%	31%	64%
3	7%	41%	87%
4	5%	46%	92%

4. Real Data Application

Following the simulation results demonstrating the PP test’s superior ability to detect mean, variance, and shape differences compared to the t-test and F-test, we apply this method to the Boston dataset from the MASS package in R [5]. This dataset includes variables *indus* (proportion of non-retail business acres), *nox* (nitric oxides concentration), *rm* (average number of rooms), and *lstat* (percentage of lower status population), split by *chas* (proximity to the *Charles River*,

10.48047/jocaaa.2025.34.08.13

coded as 0 for far and 1 for near). The analysis explores whether these groups exhibit distributional differences beyond mere means or variances.

Descriptive statistics in Table 4 reveal variations in skewness (e.g., 0.93 for *lstat* far from the river vs. 0.48 near) and kurtosis (e.g., 4.97 for *rm* far vs. 3.08 near), suggesting potential non-normality and shape differences [16]. Violin and density plots in Figures 2 and 3 visually support these observations, showing distinct distributional forms. P-values in Table 5 quantify the tests' performances: the t-test (p-values 0.104–0.205) and F-test (p-values 0.029–0.638) largely fail to reject the null of equal means or variances, except for *nox* (F-test $p=0.029$), indicating limited sensitivity to structural differences. In contrast, the PP test yields highly significant p-values (0.000–0.034), detecting differences across all variables. For instance, *indus* and *nox* show near-zero p-values despite moderate t-test and F-test results, suggesting PP captures variance or shape disparities (e.g., *indus*'s differing spreads and *nox*'s skewness). Similarly, *lstat*'s $p=0.005$ reflects shape differences (higher skewness far from the river), while *rm*'s $p=0.034$ aligns with kurtosis variations. These findings echo the simulation results, where PP outperformed the t-test in detecting variance differences in normal populations and shape differences in non-normal cases, confirming its practical utility in uncovering hidden distributional structures that traditional tests overlook.

Table 4. Descriptive statistics of the selected variables from the Boston dataset

Variable	Near River	Mean	Std. Dev.	Skewness	Kurtosis	Principal Points	
<i>indus</i>	No	11.02	6.91	0.33	1.79	5.83	18.77
	Yes	12.72	5.96	-0.20	1.53	7.16	17.97
<i>nox</i>	No	0.55	0.11	0.68	2.86	0.47	0.67
	Yes	0.59	0.14	0.75	2.25	0.51	0.80
<i>rm</i>	No	6.27	0.69	0.31	4.97	5.93	7.02
	Yes	6.52	0.88	0.73	3.08	5.91	7.33
<i>lstat</i>	No	12.76	7.17	0.93	3.50	8.69	21.22
	Yes	11.24	6.69	0.48	2.42	6.17	17.26

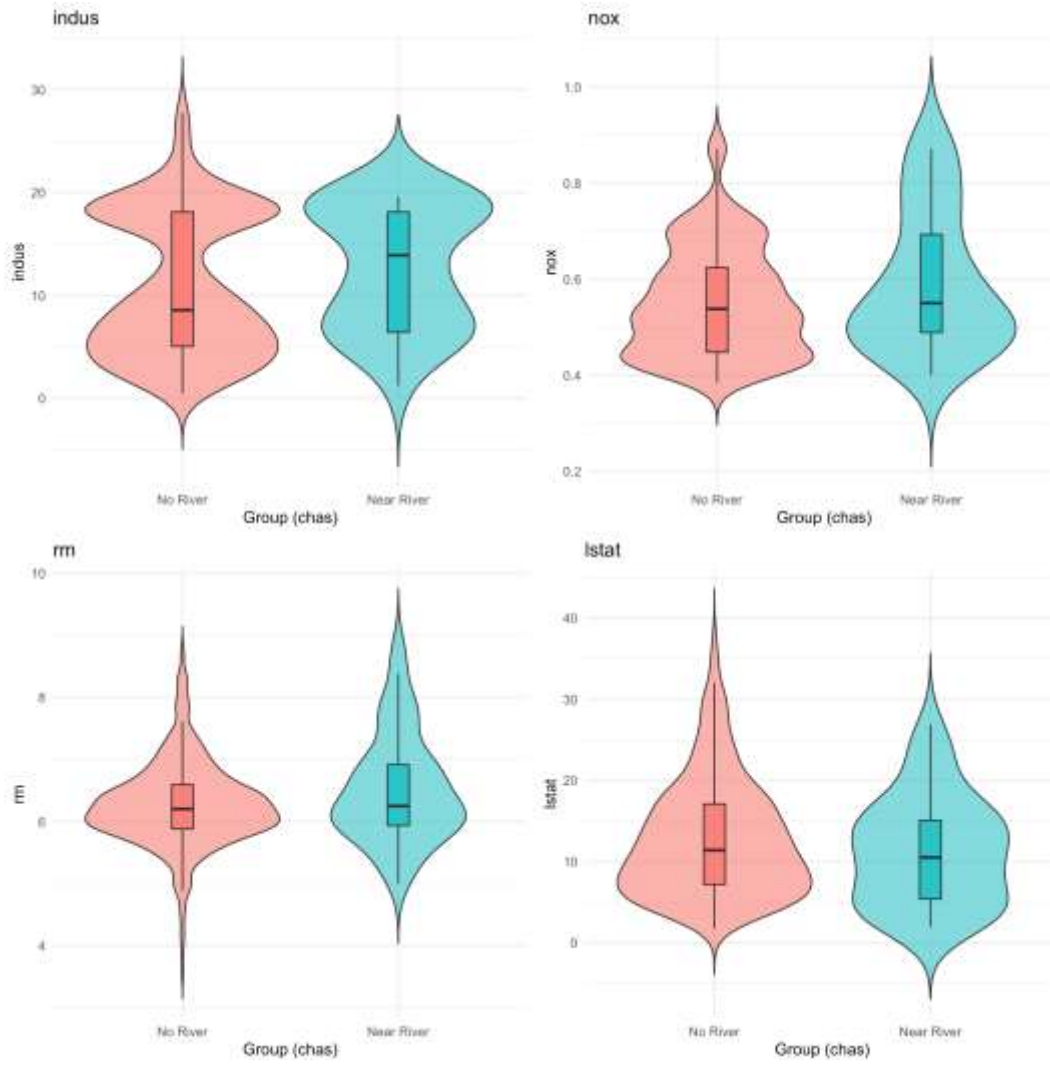


Figure 2. Violin plots of the four variables from the Boston dataset by groups near and far from the river

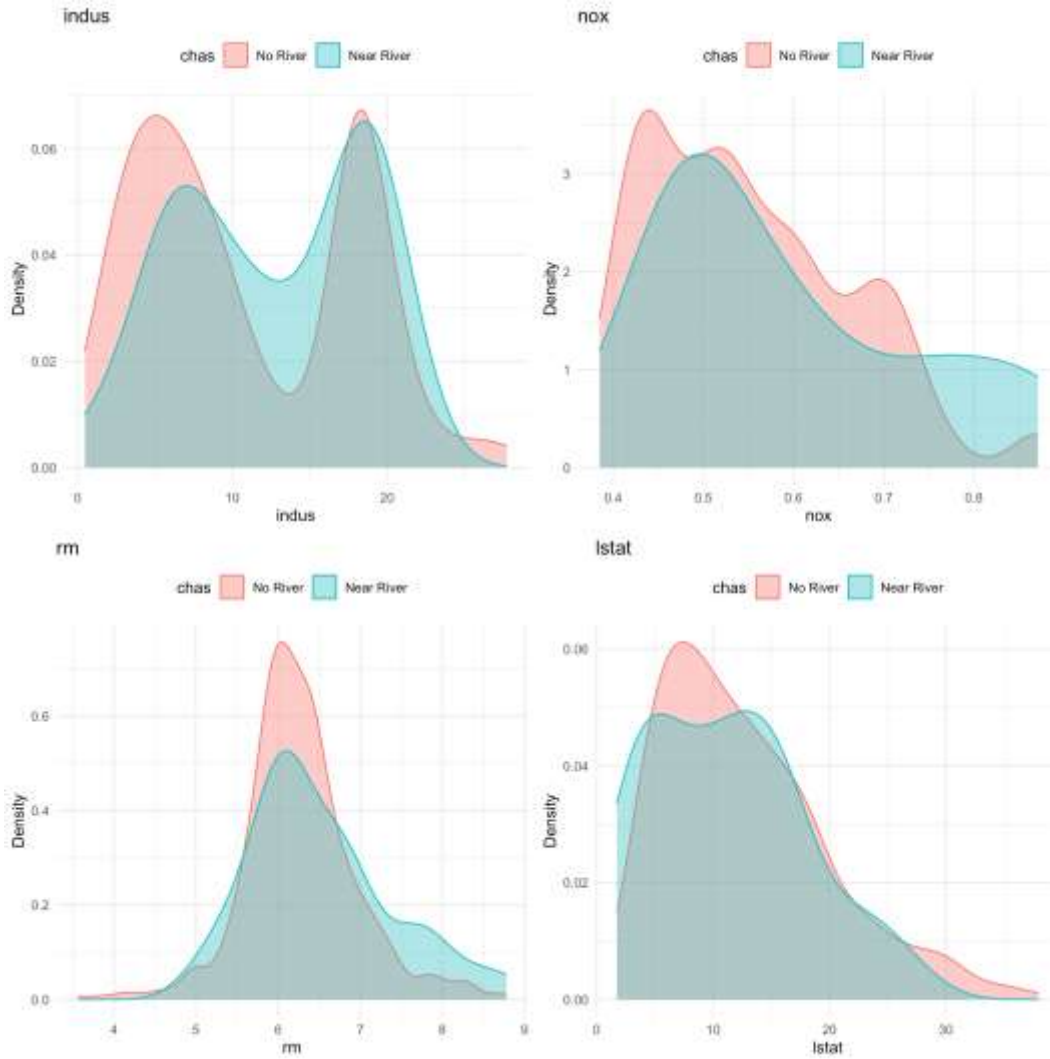


Figure 3. Density plots of the four variables from the Boston dataset by groups near and far from the river

Table 5. P-values of t-test, F-test, and principal points test for four variables from the Boston dataset

Variable	T-Test	F-Test	Principal Points
<i>indus</i>	0.115	0.289	0.000
<i>nox</i>	0.105	0.029	0.000
<i>rm</i>	0.104	0.030	0.034
<i>lstat</i>	0.205	0.638	0.005

5. Discussion and Conclusion

The principal points (PP) test demonstrates clear superiority over the traditional t-test in detecting distributional differences, fulfilling the article's aim to showcase its power and usefulness. Simulation results reveal that the PP test effectively identifies mean differences in normal populations, maintains Type I error control under equal variances, and excels in detecting variance disparities [14, 9] (e.g., rejection rates up to 0.998 for $\sigma_2^2 = 0.25$ with $k = 2$), where the t-test's performance falters (e.g., conservative errors as low as 0.000). Furthermore, the PP test's ability to capture shape differences in non-normal distributions, such as between gamma and log-normal populations with equal means [8] (rejection rates of 64%–92% with $k = 5$), addresses a critical limitation of the t-test, which remains insensitive to such structural variations (rejection rates of 3%–7%). The real data application to the Boston dataset corroborates these findings, with the PP test uncovering significant differences in variables like *indus* ($p=0.000$) and *lstat* ($p=0.005$) that the t-test ($p=0.115$ – 0.205) and F-test ($p=0.029$ – 0.638) largely miss, likely due to variance and shape effects reflected in skewness and kurtosis. While the PP test shows promise, its Type I error inflation with unequal sample sizes suggests a need for further refinement. Future research could extend the method to multivariate settings, where principal points could summarize multidimensional distributions [17], or explore higher k values to enhance sensitivity to complex structures, potentially with adaptive selection criteria. Additionally, investigating computational efficiency for large datasets and integrating PP into practical statistical software could broaden its applicability. This study establishes the PP test as a robust tool for uncovering hidden distributional differences, offering a valuable complement to classical methods in both theoretical and applied contexts.

References

- [1] Chakraborty, S. and Das, K. and Li, J. (2020). On Properties and Estimation of Principal Points for Univariate Distributions. *Statistical Methodology* **47**, 1–18.
- [2] Cochran, W. G. and Cox, G. M. (1947). *Experimental Designs*. Wiley.

10.48047/jocaaa.2025.34.08.13

- [3] Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* **7**(1), 1–26.
- [4] Flury, B. (1990). Principal Points. *Biometrika* **77**(1), 33–41.
- [5] Harrell, F. E. (2015). *Regression Modeling Strategies*. Springer Series in Statistics.
- [6] Horowitz, J. L. and Krishnamurthy, A. (2018). A bootstrap method for constructing pointwise and uniform confidence bands for conditional quantile functions. *Statistica Sinica* **28**(4), 2609–2632.
- [7] James, G. S. (1951). The Comparison of Several Groups of Observations When the Ratios of the Population Variances are Unknown. *Biometrika* **38**, 324–329.
- [8] Li, L. and Flury, B. (1995). Uniqueness of Principal Points for Univariate Distributions. *Statistics & Probability Letters* **23**, 361–364.
- [9] Matsuura, S. and Kurata, H. and Tarpey, T. (2015). Optimal estimators of principal points for minimizing expected mean squared distance. *Journal of Statistical Planning and Inference* **167**, 102–122.
- [10] Pollard, D. (1981). Strong Consistency of K -Means Clustering. *The Annals of Statistics* **9**(1), 135–140.
- [11] Stampfer, H. and Stadlober, E. (1997). Estimation of Principal Points. *Journal of Statistical Planning and Inference* **64**, 101–120.
- [12] Tarpey, T. (1994). Two Principal Points of Symmetric, Strongly Unimodal Distributions. *Statistics & Probability Letters* **20**, 253–257.
- [13] Tarpey, T. and Li, J. and Flury, B. (1994). Principal Points and Self-Consistent Points of Elliptical Distributions. *The Annals of Statistics* **22**, 1501–1521.
- [14] Tarpey, T. (1997). Estimating Principal Points of Univariate Distributions. *Computational Statistics & Data Analysis* **25**, 321–333.
- [15] Welch, B. L. (1947). The Generalization of Student's Problem When Several Different Population Variances are Involved. *Biometrika* **34**, 28–35.

10.48047/jocaaa.2025.34.08.13

- [16] Westfall, P. H. and Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*. John Wiley & Sons.
- [17] Yuen, K. (1974). The Two-Sample Trimmed t for Unequal Population Variances. *Biometrika* **61**, 165–170.