

Optimal Transport-Based Inference for Distributional Data

Olaiya Olumide O., Modebei Mark I., Eze Chibueze

Mathematics Programme, National Mathematical Centre, Abuja, Nigeria.

Department of Mathematics, Ahmadu Bello University, Zaria, Nigeria.

*Corresponding author Email: gmarc.ify@gmail.com

Received: 12 June, 2026; **Revised:** 04 August, 2026; **Accepted:** 08 August, 2026; **Published:** 31 August, 2026

DOI: 10.29218/srmsmaths.v11i1.03	Abstract
Keywords: Optimal transport, Wasserstein distance, distributional data, Fréchet mean, hypothesis testing, Sinkhorn algorithm, WANOVA, non-Euclidean statistics. 2020 Mathematics Subject Classification (MSC): 62G10, 60B05, 49Q22, 90C25.	The analysis of distributional data where each observational unit is itself a probability distribution has become increasingly important across economics, environmental science, biomedical engineering, and machine learning. Classical multivariate methods are fundamentally inadequate for such data because the space of probability measures lacks the linear structure assumed by traditional techniques. Optimal transport (OT) theory, and in particular the 2-Wasserstein metric, provides a geometrically coherent framework for statistical inference on probability distributions. This paper develops a comprehensive methodology for OT-based inference on distributional data, encompassing one-sample, two-sample, and multi-group hypothesis tests formulated in Wasserstein space. We establish the asymptotic theory underpinning these tests, including central limit theorems for Fréchet means and variance statistics. Entropic regularization via the Sinkhorn algorithm is incorporated to render the methodology computationally tractable. We evaluate our framework through extensive Monte Carlo simulations and two empirical applications: income distribution analysis across economic regions and the comparison of precipitation distributions in climate science. Our results demonstrate that Wasserstein-based tests offer substantially greater power than classical alternatives when distributional shape differences are present, while maintaining nominal size under the null hypothesis.

1 Introduction

1.1 Background and Motivation

Modern data collection technologies routinely produce observations that are themselves probability distributions rather than scalar or vector-valued quantities. Consider, for example, the distribution of household incomes within each country of a panel study, the distribution of daily precipitation totals at each climate station in a meteorological network, or the distribution of fluorescence intensities within each tumor imaged in a clinical trial. In each of these settings, the fundamental observational unit is a distribution, and the scientific question concerns how these distributions differ across groups, evolve over time, or relate to covariates.

The statistical analysis of such distributional data requires mathematical frameworks that respect the geometry of the space of probability measures. The celebrated framework of Optimal Transport (OT) provides precisely such a geometry. Rooted in Monge's 1781 problem of moving soil from one configuration to another at minimal cost, and cast in a rigorous probabilistic framework by Kantorovich [1], OT has emerged in recent decades as a unifying language for comparing probability measures in a geometrically meaningful way [2, 3].

The 2-Wasserstein distance between two probability measures and on quantifies the minimal expected squared displacement needed to transport mass from to under an optimal coupling. When equipped with the metric, the space of probability measures with finite second moments becomes a geodesic metric space—the *Wasserstein space*—whose rich geometric structure [4] supports a coherent theory of Fréchet means, geodesics, and tangent spaces [5]. This structure provides the natural setting for statistical inference on distributional data.

1.2 Distributional Data in Modern Statistics

Distributional data arise across a diverse range of scientific disciplines. In economics, inequality research increasingly represents each country or region by its entire income distribution rather than a summary statistic such as the Gini coefficient [6]. In clinical medicine, high-throughput technologies produce patient-specific distributions of gene expression levels, actigraphy intensities, or continuous glucose monitoring readings, motivating distribution-valued functional data analysis [7, 5]. In environmental science, precipitation patterns at each station are best described by a full distribution over time [8, 9].

From a formal perspective, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mu: \Omega \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ be a measurable mapping. Then μ is a random probability measure, and a random sample $\mu_1, \dots, \mu_n$ from the distribution of μ constitutes distributional data. The statistical task is to perform inference about the law of μ —its mean, variance, group differences, or covariate relationships—using the observed sample of distributions.

A central structural challenge is that $\mathcal{P}_2(\mathbb{R}^d)$ is not a linear space: there is no canonical way to add two probability measures or multiply a measure by a scalar and obtain another valid probability measure. Consequently, classical statistical machinery predicated on the Euclidean structure—from t -tests and ANOVA to principal components analysis—cannot be directly applied to distributional data [10, 11].

1.3 Limitations of Classical Methods

Classical approaches to the analysis of distributional data typically resort to one of several ad hoc strategies. First, feature extraction summarizes each distribution by a vector of scalar statistics (mean, variance, quantiles, etc.) and then applies standard multivariate methods to these vectors. This approach discards distributional shape information and is sensitive to the choice of summary statistics.

Second, kernel methods embed probability measures into a reproducing kernel Hilbert space (RKHS) using a characteristic kernel, enabling the use of maximum mean discrepancy (MMD) as a test statistic [12]. While kernel methods are computationally convenient, they do not respect the Riemannian geometry of Wasserstein space and can be poorly calibrated when distributional differences manifest primarily in the tails.

Furthermore, functional data analysis (FDA) treats density functions as elements of an L^2 function space [13, 14]. However, the L^2 distance between density functions is not invariant to reparameterizations of the support and can behave counterintuitively: two distributions that differ only in their location are L^2 -far apart even if they have identical shape. Moreover, linear operations in L^2 do not preserve the non-negativity and unit-integral constraints required of density functions.

Optimal transport overcomes all these limitations by defining distances and interpolations directly in the space of measures, respecting the geometry of the support, and providing a framework for statistical inference that is invariant to reparameterization of the sample space [15].

In this work, we develop a self-contained treatment of hypothesis testing for distributional data in Wasserstein space, covering one-sample tests of Fréchet mean, two-sample tests of distributional equality, and multi-group Wasserstein Analysis of Variance (WANOVA). While [16] develop Fréchet ANOVA for general metric spaces and [10] propose Wasserstein-based group tests, our framework makes three distinct contributions.

First, we unify one-sample, two-sample, and multi-group testing within a single coherent Wasserstein tangent-space framework, which neither existing work provides. Second, we explicitly integrate entropic regularization via the Sinkhorn–Knopp algorithm with formal statistical guarantees, rendering the methodology computationally tractable for large samples—a gap left unaddressed by both [16] and [10]. Third, we extend the asymptotic theory of [16] specifically to the Wasserstein setting, establishing central limit theorems for the Sinkhorn-regularized test statistics and providing explicit bias-correction via the Sinkhorn divergence (39).

We also establish central limit theorems and consistency results for the proposed test statistics under mild regularity conditions, extending results of [16] and [17] to the specific context of Wasserstein-based distributional tests. We incorporate entropic regularization via the Sinkhorn–Knopp algorithm [18, 15] to make the proposed tests computationally tractable, providing explicit convergence guarantees. We present Monte Carlo simulation studies benchmarking the proposed tests against classical competitors, and two empirical applications—to income distribution data and to climate precipitation distributions—demonstrating the practical relevance of the methodology. Complete reproducible code is provided in Appendix 7.

2 Mathematical Foundations of Optimal Transport

2.1 Probability Measures and Wasserstein Space

Let $(\mathcal{X}, \rho)$ be a Polish metric space. Denote by $\mathcal{P}(\mathcal{X})$ the set of all Borel probability measures on $\mathcal{X}$. For $p \geq 1$, let

$$\mathcal{P}_p(\mathcal{X}) = \{\mu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} \rho(x_0, x)^p d\mu(x) < \infty \text{ for some (hence all) } x_0 \in \mathcal{X}\} \quad (1)$$

denote the space of probability measures with finite p -th moment. We work primarily with $\mathcal{X} = \mathbb{R}^d$, $\rho(x, y) = \|x - y\|$, and $p = 2$, yielding the space $\mathcal{P}_2(\mathbb{R}^d)$.

Definition 2.1 (Wasserstein Space) *The p -Wasserstein space $(\mathcal{P}_2(\mathcal{X}), W_p)$ is the metric space consisting of $\mathcal{P}_2(\mathcal{X})$ equipped with the p -Wasserstein distance defined in (5) below. For $p = 2$, this is the Wasserstein space.*

Definition 2.2 (Friedman Mean in Wasserstein Space) *Let $\nu_1, \dots, \nu_n \in \mathcal{P}_2(\mathbb{R}^d)$ and let $\lambda_1, \dots, \lambda_n > 0$ with $\sum_{i=1}^n \lambda_i = 1$. The Wasserstein barycenter (or Fréchet mean) is defined as*

$$\bar{\nu} = \arg \min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \sum_{i=1}^n \lambda_i W_2^2(\mu, \nu_i). \quad (2)$$

The existence and uniqueness of the Wasserstein barycenter under mild conditions was established by [19]. Specifically, if at least one of the ν_i is absolutely continuous, the barycenter is unique.

2.2 Optimal Transport Problem

The optimal transport problem has two classical formulations: the Monge problem and the Kantorovich relaxation.

Definition 2.3 (Monge Problem) *Let $\mathcal{X}$ be a Polish space and let $\mu, \nu \in \mathcal{P}(\mathcal{X})$ be Borel probability measures. A transport map from μ to ν is a measurable mapping $T: \mathcal{X} \rightarrow \mathcal{X}$ satisfying $T_{\#}\mu = \nu$, that is,*

$$\nu(B) = \mu(T^{-1}(B))$$

for every Borel set $B \subseteq \mathcal{X}$. Given a cost function $c: \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty)$, the Monge optimal transport problem is

$$\inf_{T: T_{\#}\mu = \nu} \int_{\mathcal{X}} c(x, T(x)) d\mu(x).$$

The Monge problem may be infeasible (e.g., when μ is a Dirac measure and ν is not). The Kantorovich relaxation overcomes this by optimizing over couplings rather than maps.

Definition 2.4 (Kantorovich Problem) *Let $\Pi(\mu, \nu)$ denote the set of all probability measures on $\mathcal{X} \times \mathcal{X}$ with marginals μ and ν . The Kantorovich problem is*

$$\text{OT}_c(\mu, \nu) = \min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{X}} c(x, y) d\pi(x, y). \quad (3)$$

The minimum in (3) is always attained; a minimizing π^* is called an *optimal coupling* or *optimal transport plan*.

Theorem 2.5 (Kantorovich Duality optimal) *Let $c: \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty)$ be a lower-semicontinuous cost function. Then*

$$\text{OT}_c(\mu, \nu) = \sup_{(f, g) \in \Phi_c} \left\{ \int_{\mathcal{X}} f \, d\mu + \int_{\mathcal{X}} g \, d\nu \right\}, \quad (4)$$

where $\Phi_c = \{(f, g) \in L^1(\mu) \times L^1(\nu) : f(x) + g(y) \leq c(x, y) \, \forall x, y\}$.

Theorem 2.6 (Brenier's Theorem polar) *Let $\mu \ll \text{Leb}$ and $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $c(x, y) = \|x - y\|^2$. Then the unique optimal coupling $\pi^* = (\text{id}, T^*)_{\#}\mu$ is concentrated on the graph of a map $T^* = \nabla\phi$ for some convex function $\phi: \mathbb{R}^d \rightarrow \mathbb{R}$. The optimal map satisfies $T_{\#}^*\mu = \nu$.*

Brenier's theorem has profound statistical implications: it guarantees that the optimal transport map between absolutely continuous measures is well-defined, unique, and can be characterized as the gradient of a convex potential. This convex structure underpins the consistency of empirical OT estimators.

2.3 2-Wasserstein Distance

Definition 2.7 (2-Wasserstein Distance) *For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the 2-Wasserstein distance is*

$$W_2(\mu, \nu) = \left(\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \, d\pi(x, y) \right)^{1/2}. \quad (5)$$

In the univariate case $d = 1$, the W_2 distance has a particularly elegant closed form in terms of quantile functions.

Proposition 2.8 (Univariate Wasserstein Distance) *For $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$ with quantile functions F_{μ}^{-1} and F_{ν}^{-1} ,*

$$W_2^2(\mu, \nu) = \int_0^1 (F_{\mu}^{-1}(u) - F_{\nu}^{-1}(u))^2 \, du. \quad (6)$$

Equation (6) shows that in one dimension, the optimal transport map is simply the monotone rearrangement $T^* = F_{\nu}^{-1} \circ F_{\mu}$. This formula motivates quantile-function representations of distributional data [7] and enables fast computation of W_2 in univariate settings.

Key metric properties of W_2 include:

- i. *Metric axioms:* W_2 satisfies non-negativity, symmetry, definiteness, and the triangle inequality.
- ii. *Weak convergence:* $W_2(\mu_n, \mu) \rightarrow 0$ if and only if $\mu_n \xrightarrow{w} \mu$ and $\int \|x\|^2 \, d\mu_n(x) \rightarrow \int \|x\|^2 \, d\mu(x)$.
- iii. *Completeness:* $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is a complete metric space.
- iv. *Separability:* $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is separable.

2.4 Geometric Properties of Distributional Data

The Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ possesses a rich geometric structure that distinguishes it from Hilbert spaces and finite-dimensional manifolds.

2.4.1 Geodesics.

For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the constant-speed geodesic connecting μ to ν is the curve $(\mu_t)_{t \in [0, 1]}$ with

$$\mu_t = [(1-t) \text{id} + t T^*]_{\#} \mu, \quad (7)$$

where T^* is the optimal transport map from μ to ν (when $\mu \ll \text{Leb}$). This geodesic is the displacement interpolation of [21].

2.4.2 Tangent Space.

Definition 2.9 (Tangent Space at μ) For $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ absolutely continuous with respect to the Lebesgue measure, the tangent space at μ is

$$\text{Tan}_{\mu} \mathcal{P}_2(\mathbb{R}^d) = \overline{\{\nabla \varphi : \varphi \in C_c^{\infty}(\mathbb{R}^d)\}}^{L^2(\mu; \mathbb{R}^d)}, \quad (8)$$

the $L^2(\mu; \mathbb{R}^d)$ -closure of gradients of smooth compactly supported functions [4]. Elements $v \in \text{Tan}_{\mu} \mathcal{P}_2(\mathbb{R}^d)$ are vector fields representing infinitesimal displacements of mass at μ .

Definition 2.10 (Logarithmic and Exponential Maps) For $\mu \ll \text{Leb}$ and $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, let T^* be the unique optimal transport map from μ to ν guaranteed by Brenier's Theorem 3.6. The logarithmic map at μ is

$$\log_{\mu}(\nu) = T^* - \text{id} \in \text{Tan}_{\mu} \mathcal{P}_2(\mathbb{R}^d), \quad (9)$$

and the exponential map at μ is

$$\exp_{\mu}(v) = (\text{id} + v)_{\#} \mu, \quad v \in \text{Tan}_{\mu} \mathcal{P}_2(\mathbb{R}^d). \quad (10)$$

These satisfy $\exp_{\mu}(\log_{\mu}(\nu)) = \nu$ whenever $T^* - \text{id} \in \text{Tan}_{\mu} \mathcal{P}_2(\mathbb{R}^d)$, and $W_2^2(\mu, \nu) = \|\log_{\mu}(\nu)\|_{L^2(\mu; \mathbb{R}^d)}^2$ [4].

2.4.3 Curvature.

Wasserstein space has nonnegative Alexandrov curvature [22], which implies that Fréchet means exist and are unique under mild conditions. This positive curvature property also implies that Jensen's inequality holds for convex functions in Wasserstein space, with the inequality reversed compared to the Euclidean setting.

2.4.4 Tangent-Space Linearization.

A key methodological tool is linearization: given a base measure μ_0 , one maps a collection $\nu_1, \dots, \nu_n$ into the tangent space via $u_i = \log_{\mu_0}(\nu_i) = T_i^* - \text{id}$, where T_i^* is the optimal map from μ_0 to ν_i . Statistical methods for $L^2(\mu_0)$ -valued data can then be applied to $u_1, \dots, u_n$, and results mapped back to Wasserstein space via $\exp_{\mu_0}$ [11, 23].

2.5 Computational Aspects of Optimal Transport

2.5.1 Discrete setting.

Suppose $\mu = \sum_{i=1}^n a_i \delta_{x_i}$ and $\nu = \sum_{j=1}^m b_j \delta_{y_j}$ are discrete measures with $a \in \Sigma_n$, $b \in \Sigma_m$ (probability simplices). The Kantorovich problem reduces to a linear program:

$$\text{OT}_c(\mu, \nu) = \min_{P \in \Pi(a, b)} \langle C, P \rangle_F, \quad C_{ij} = c(x_i, y_j). \quad (11)$$

This can be solved with the network simplex method in $O(n^3 \log n)$ operations [15], which is prohibitive for large n .

2.5.2 Sliced Wasserstein.

For high-dimensional data, the sliced Wasserstein distance averages one-dimensional Wasserstein distances over random projections:

$$SW_2^2(\mu, \nu) = \int_{\mathbb{S}^{d-1}} W_2^2(\theta_{\#}\mu, \theta_{\#}\nu) d\sigma(\theta), \quad (12)$$

where $\theta_{\#}\mu$ is the push-forward of μ under $x \mapsto \theta^{\top}x$, and σ is the uniform measure on the sphere. This can be approximated efficiently via Monte Carlo sampling [15].

2.5.3 Computational complexity.

For n -point discrete measures, the standard OT problem has complexity $O(n^3 \log n)$. The entropy-regularized variant (Section 4.5) reduces this to $O(n^2/\varepsilon^3)$ for approximation error $\varepsilon > 0$ [24].

3 Wasserstein-Based Statistical Inference Methodology

3.1 One-Sample Hypothesis Testing

Let $\nu_1, \dots, \nu_n \stackrel{\text{iid}}{\sim} F$ be a random sample of distributions from some population F on $\mathcal{P}_2(\mathbb{R}^d)$. Denote by μ_F the Fréchet mean of F :

$$\mu_F = \arg \min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \mathbb{E}[W_2^2(\mu, \nu)], \quad \nu \sim F. \quad (13)$$

The one-sample testing problem is:

$$H_0: \mu_F = \mu_0 \quad \text{vs} \quad H_1: \mu_F \neq \mu_0, \quad (14)$$

for a specified reference measure $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$.

The sample Fréchet mean $\hat{\mu}_n$ is defined as

$$\hat{\mu}_n = \arg \min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \frac{1}{n} \sum_{i=1}^n W_2^2(\mu, \nu_i). \quad (15)$$

Under regularity conditions (see Section 4.4), $\hat{\mu}_n \xrightarrow{W_2} \mu_F$ almost surely [5].

Test statistic.

Using the tangent-space linearization at μ_0 , let $u_i = T_i - \text{id} \in L^2(\mu_0; \mathbb{R}^d)$, where T_i is the optimal map from μ_0 to ν_i . The sample mean in the tangent space is $\bar{u} = n^{-1} \sum_{i=1}^n u_i$. The test statistic is

$$T_n^{(1)} = n \|\bar{u}\|_{L^2(\mu_0)}^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \langle u_i, u_j \rangle_{L^2(\mu_0)}. \quad (16)$$

Under H_0 , $T_n^{(1)}$ converges in distribution to a weighted chi-squared random variable (see Theorem 4.10), and we reject H_0 when $T_n^{(1)}$ exceeds the appropriate quantile. Bootstrap calibration of the null distribution is considered in Section 4.4.

3.2 Two-Sample Hypothesis Testing

Let $\nu_1, \dots, \nu_{n_1} \stackrel{\text{iid}}{\sim} F_1$ and $\tilde{\nu}_1, \dots, \tilde{\nu}_{n_2} \stackrel{\text{iid}}{\sim} F_2$ be two independent samples of distributions. The two-sample testing problem is:

$$H_0: F_1 = F_2 \quad \text{vs} \quad H_1: F_1 \neq F_2. \quad (17)$$

Remark 3.1 Equality of distributions $F_1 = F_2$ implies equality of all Wasserstein functionals, including Fréchet means and variances, whereas the converse does not hold in general. The energy-distance statistic $T_n^{(2)}$ in (19) is consistent against all alternatives $F_1 \neq F_2$, since $\mathcal{E}(F_1, F_2) = 0$ if and only if $F_1 = F_2$ [25]. The Wasserstein distance statistic $S_n^{(2)}$ in (20) targets only Fréchet mean differences and should be interpreted as a test of $H_0^{mean}: \mu_{F_1} = \mu_{F_2}$; it may lack power against alternatives where $F_1 \neq F_2$ but $\mu_{F_1} = \mu_{F_2}$. We therefore recommend $T_n^{(2)}$ as the primary two-sample statistic when a general test of distributional equality is required, reserving $S_n^{(2)}$ for settings where the scientific question specifically concerns Fréchet mean differences.

3.2.1 Energy-distance test.

The energy distance between F_1 and F_2 in Wasserstein space is

$$\mathcal{E}(F_1, F_2) = 2\mathbb{E}[W_2(v^{(1)}, v^{(2)})] - \mathbb{E}[W_2(v^{(1)}, \tilde{v}^{(1)})] - \mathbb{E}[W_2(v^{(2)}, \tilde{v}^{(2)})], \quad (18)$$

where $v^{(k)}$ are independent draws from F_k . This is non-negative and equals zero if and only if $F_1 = F_2$ [25]. The natural test statistic is

$$T_n^{(2)} = \frac{n_1 n_2}{n_1 + n_2} \hat{\mathcal{E}}_n, \quad (19)$$

where $\hat{\mathcal{E}}_n$ is the U-statistic estimator of $\mathcal{E}(F_1, F_2)$.

3.2.2 Wasserstein distance test.

An alternative approach directly uses the 2-Wasserstein distance between the empirical Fréchet means:

$$S_n^{(2)} = W_2^2(\hat{\mu}_{n_1}^{(1)}, \hat{\mu}_{n_2}^{(2)}), \quad (20)$$

where $\hat{\mu}_{n_k}^{(k)}$ is the sample Fréchet mean from group k .

3.2.3 Permutation approach.

Calibration of both tests is achieved via permutation: one randomly reassigns the $n_1 + n_2$ distributional observations to two groups of sizes n_1 and n_2 , recomputes the test statistic, and repeats B times. The p -value is

$$p - \text{value} = \frac{1}{B} \sum_{b=1}^B \mathbf{1}\{T^{(b)} \geq T_{\text{obs}}\}. \quad (21)$$

This approach is asymptotically exact under exchangeability, which holds under H_0 .

3.3 Wasserstein Analysis of Variance (WANOVA)

Consider K independent groups with sample sizes $n_1, \dots, n_K$ and $n = \sum_{k=1}^K n_k$. Let $v_{ki} \stackrel{\text{iid}}{\sim} F_k$, $i = 1, \dots, n_k$, $k = 1, \dots, K$. The multi-group testing problem is:

$$H_0: F_1 = \dots = F_K \quad \text{vs} \quad H_1: F_j \neq F_k \text{ for some } j \neq k. \quad (22)$$

Remark 3.2 The WANOVA statistic F_W in (28) is constructed from Fréchet means and variances and is therefore primarily sensitive to differences in group Fréchet means. Formally, F_W tests the composite null $H_0^{mean}: \mu_{F_1} = \dots = \mu_{F_K}$, which is implied by but weaker than $H_0: F_1 = \dots = F_K$. Consequently, F_W may have reduced power against alternatives where group distributions differ only in higher-order structure (e.g., variance or shape) with identical Fréchet means. To guard against such alternatives, we recommend supplementing F_W with the energy-distance statistic $T_n^{(2)}$ in pairwise post-hoc comparisons, which is consistent against all distributional alternatives [25]. The simulation results in Section 5 (scenarios S3–S4) empirically confirm this recommendation.

Fréchet variance decomposition.

Define the *total Fréchet variance*

$$V_T = \frac{1}{n} \sum_{k=1}^K \sum_{i=1}^{n_k} W_2^2(v_{ki}, \hat{\mu}_n), \quad (23)$$

the *between-group variance*

$$V_B = \frac{1}{n} \sum_{k=1}^K n_k W_2^2(\hat{\mu}_{n_k}^{(k)}, \hat{\mu}_n), \quad (24)$$

and the *within-group variance*

$$V_W = \frac{1}{n} \sum_{k=1}^K \sum_{i=1}^{n_k} W_2^2(v_{ki}, \hat{\mu}_{n_k}^{(k)}), \quad (25)$$

where $\hat{\mu}_n$ is the overall Fréchet mean and $\hat{\mu}_{n_k}^{(k)}$ is the group- k Fréchet mean.

The validity of the decomposition $V_T = V_B + V_W$ in Wasserstein space requires care, since $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is not a linear space and the parallelogram identity does not hold in general. We impose the following additional conditions and state the decomposition precisely.

Assumption 3.3

- i. Each group population Fréchet mean μ_{F_k} exists, is unique, and is absolutely continuous with respect to the Lebesgue measure.
- ii. The overall population Fréchet mean μ_F exists and is unique.
- iii. $\mathbb{E}[W_2^4(v_{ki}, \mu_{F_k})] < \infty$ for all k .
- iv. The group sample sizes satisfy $n_k/n \rightarrow \rho_k \in (0, 1)$ as $n \rightarrow \infty$.

Proposition 3.4 (WANOVA Decomposition) *Under Assumption 4.3, the following decomposition holds asymptotically:*

$$V_T = V_B + V_W + R_n, \quad (26)$$

where $R_n = o_p(1)$ as $n \rightarrow \infty$. Under $H_0: \mu_{F_1} = \dots = \mu_{F_K}$, the decomposition is exact in the tangent space at the common Fréchet mean, i.e., $R_n = O_p(n^{-1/2})$. Specifically, linearizing via the logarithmic map $u_{ki} = T_{ki} - \text{id} \in L^2(\hat{\mu}_n; \mathbb{R}^d)$, where T_{ki} is the optimal transport map from $\hat{\mu}_n$ to v_{ki} , yields the exact identity

$$\sum_{k,i} \|u_{ki}\|^2 = \sum_k n_k \|\bar{u}_k\|^2 + \sum_{k,i} \|u_{ki} - \bar{u}_k\|^2, \quad (27)$$

which is the Hilbert-space analogue of the ANOVA decomposition in $L^2(\hat{\mu}_n; \mathbb{R}^d)$. The approximation error R_n satisfies $|R_n| \leq C \cdot n^{-1/2} \cdot \max_k W_2^2(\mu_{F_k}, \mu_F)$ for a universal constant $C > 0$ [16, 4].

Remark 3.5 Proposition 3.4 shows that the decomposition is exact in the tangent space and acquires a curvature correction R_n in Wasserstein space that is asymptotically negligible. Permutation calibration of F_W remains valid regardless of R_n , since the permutation distribution consistently estimates the finite-sample null distribution of F_W under exchangeability.

Definition 3.6 (WANOVA Statistic) *The Wasserstein Analysis of Variance (WANOVA) statistic is*

$$F_W = \frac{V_B/(K-1)}{V_W/(n-K)}. \quad (28)$$

Under H_0 , large values of F_W indicate evidence against equality of group Fréchet means. This statistic generalizes the classical F -statistic to the Wasserstein setting [10]. The null distribution of F_W is

approximated via permutation, bootstrap, or asymptotic chi-squared approximation (see Section 3.4).

Post-hoc testing.

Following rejection of the global null, pairwise comparisons can be performed using the two-sample test of Section 3.2 with a Bonferroni or Benjamini–Hochberg correction for multiple comparisons [26].

3.4 Asymptotic Theory and Consistency

We state the key asymptotic results underpinning the proposed tests. Throughout, we impose the following regularity conditions.

Assumption 3.7

- i. The population Fréchet mean μ_F exists, is unique, and satisfies $\mu_F \ll \text{Leb}$.
- ii. $\mathbb{E}[W_2^4(v, \mu_F)] < \infty$ for $v \sim F$.
- iii. The tangent-space covariance operator $\Sigma: L^2(\mu_F; \mathbb{R}^d) \rightarrow L^2(\mu_F; \mathbb{R}^d)$, defined for $h \in L^2(\mu_F; \mathbb{R}^d)$ by

$$a. \quad \Sigma h = \mathbb{E}[\langle T - \text{id}, h \rangle_{L^2(\mu_F; \mathbb{R}^d)} (T - \text{id})], \quad (29)$$

where $T = \log_{\mu_F}(v) + \text{id}$ is the optimal transport map from μ_F to $v \sim F$, is a well-defined, symmetric, positive semi-definite, trace-class operator on $L^2(\mu_F; \mathbb{R}^d)$. Trace-class means $\text{tr}(\Sigma) = \mathbb{E}[\|T - \text{id}\|_{L^2(\mu_F; \mathbb{R}^d)}^2] = \mathbb{E}[W_2^2(v, \mu_F)] < \infty$, guaranteed by condition (b).

- iv. The map $v \mapsto \log_{\mu_F}(v)$ is Fréchet differentiable at μ_F in the sense of [5], ensuring the tangent-space approximation is first-order accurate: $W_2(\exp_{\mu_F}(\bar{u}), \hat{\mu}_n) = O_p(n^{-1})$, where $\bar{u} = n^{-1} \sum_{i=1}^n \log_{\mu_F}(v_i)$.

Remark 3.8 Since Σ is symmetric, positive semi-definite, and trace-class on the separable Hilbert space $L^2(\mu_F; \mathbb{R}^d)$, the spectral theorem guarantees an orthonormal eigenbasis $\{\phi_j\}_{j=1}^\infty$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ satisfying $\sum_{j=1}^\infty \lambda_j < \infty$. This yields the weighted chi-squared limit in Theorem 4.10: the Karhunen–Loève expansion of $G \sim \mathcal{N}(0, \Sigma)$ gives $\|G\|_{L^2}^2 = \sum_{j=1}^\infty \lambda_j Z_j^2$ with $Z_j \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$.

Theorem 3.9 (Central Limit Theorem for Fréchet Mean) *Under Assumption 4.7, as $n \rightarrow \infty$,*

$$\sqrt{n} W_2(\hat{\mu}_n, \mu_F) = O_p(1), \quad (30)$$

and the rescaled deviation $\sqrt{n}(\hat{T}_n - \text{id})$, where $\hat{T}_n$ is the optimal map from μ_F to $\hat{\mu}_n$, converges in distribution in $L^2(\mu_F; \mathbb{R}^d)$ to a Gaussian random element with mean zero and covariance Σ .

Theorem 3.10 (Asymptotic Distribution of One-Sample Statistic) *Under $H_0: \mu_F = \mu_0$ and Assumption 3.7, as $n \rightarrow \infty$,*

$$T_n^{(1)} \xrightarrow{d} \sum_{j=1}^\infty \lambda_j Z_j^2, \quad (31)$$

where $Z_j \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ and $\lambda_1 \geq \lambda_2 \geq \dots > 0$ are the eigenvalues of Σ .

Proof. By Theorem 3.9, $\sqrt{n} \bar{u}$ converges weakly to $G \sim \mathcal{N}(0, \Sigma)$ in $L^2(\mu_0; \mathbb{R}^d)$. The continuous mapping theorem gives $T_n^{(1)} = n \|\bar{u}\|^2 \xrightarrow{d} \|G\|^2 = \sum_j \lambda_j Z_j^2$ via the Karhunen–Loève decomposition.

Theorem 3.11 (Consistency of WANOVA) *Under fixed alternatives, $F_W \rightarrow \infty$ in probability as $n \rightarrow \infty$ whenever $\mu_F^{(j)} \neq \mu_F^{(k)}$ for some $j \neq k$. Under H_0 , the WANOVA test has asymptotic level α when*

calibrated by the permutation distribution.

3.4.1 Bootstrap and permutation calibration.

In finite samples, the weighted chi-squared limit in Theorem 4.10 is approximated by the bootstrap: one draws $v_1^*, \dots, v_n^*$ from the empirical distribution $\hat{F}_n = n^{-1} \sum_i \delta_{v_i}$, recomputes T_n^* , and repeats B times. The bootstrap is consistent whenever $W_2(\hat{F}_n, F) \rightarrow 0$ in probability [27].

3.5 Entropic Regularization and Sinkhorn Algorithm

For large-scale problems, the exact OT computation is infeasible. Entropic regularization resolves this by adding a Kullback–Leibler penalty to the Kantorovich problem.

Definition 3.12 (Entropically Regularized OT) For $\varepsilon > 0$, the entropy-regularized optimal transport cost is

$$\text{OT}_\varepsilon^\varepsilon(\mu, \nu) = \min_{\pi \in \Pi(\mu, \nu)} \left\{ \int c \, d\pi + \varepsilon \text{KL}(\pi \parallel \mu \otimes \nu) \right\}, \quad (32)$$

where $\text{KL}(\pi \parallel \mu \otimes \nu) = \int \log \frac{d\pi}{d(\mu \otimes \nu)} d\pi$.

The regularized problem (32) has a unique solution π_ε^* and satisfies $\text{OT}_\varepsilon^\varepsilon(\mu, \nu) \rightarrow \text{OT}_c(\mu, \nu)$ as $\varepsilon \rightarrow 0$ [15, 18]. The optimal coupling takes the form

$$\pi_\varepsilon^*(x, y) = u(x) \exp(-c(x, y)/\varepsilon) v(y) \mu(x) \nu(y) \quad (33)$$

for scaling functions $u, v > 0$. In the discrete case with cost matrix C :

$$P^* = \text{diag}(u) K \text{diag}(v), \quad K_{ij} = e^{-C_{ij}/\varepsilon}, \quad (34)$$

and the scaling vectors u, v are found by the Sinkhorn–Knopp algorithm.

Algorithm 1 Sinkhorn–Knopp [18, 15]

Require: Cost matrix $C \in \mathbb{R}^{n \times m}$, marginals $a \in \Sigma_n$, $b \in \Sigma_m$, regularization $\varepsilon > 0$, tolerance $\delta > 0$

1. Set $K \leftarrow \exp(-C/\varepsilon)$ (elementwise)
 2. Initialize $v^{(0)} \leftarrow \mathbf{1}_m$
 3. **Repeat**
 4. $u^{(\ell+1)} \leftarrow a \oslash (K v^{(\ell)})$ Row normalization
 5. $v^{(\ell+1)} \leftarrow b \oslash (K^\top u^{(\ell+1)})$ Column normalization
 6. **Until** $\|K^\top u - b\|_1 < \delta$ and $\|K v - a\|_1 < \delta$
 7. **return** $P^* \leftarrow \text{diag}(u) K \text{diag}(v)$, $\widehat{\text{OT}}^\varepsilon \leftarrow \langle C, P^* \rangle_F$
-

3.5.1 Convergence.

Proposition 3.13 (Linear Convergence of Sinkhorn–Knopp; scaling, [24, 29]). Let $K = \exp(-C/\varepsilon)$ with $C \in \mathbb{R}^{n \times m}$, $C_{ij} \geq 0$, and $\varepsilon > 0$. Define the contraction ratio

$$\rho(\varepsilon) = \tanh\left(\frac{\|C\|_\infty}{4\varepsilon}\right) \in (0, 1), \quad (35)$$

where $\|C\|_\infty = \max_{i,j} C_{ij}$. Then

$$\|K^\top u^{(\ell)} - b\|_1 \leq 2 \rho(\varepsilon)^\ell \|K^\top u^{(0)} - b\|_1, \quad (36)$$

and to achieve ℓ^1 -marginal error δ , Algorithm 4.5 requires at most

$$\ell^* = \left\lceil \frac{\log(2/\delta)}{\log(1/\rho(\varepsilon))} \right\rceil = O\left(\frac{\|C\|_\infty}{\varepsilon} \log \frac{1}{\delta}\right) \quad (37)$$

iterations.

Remark 3.14 Each iteration costs $O(nm)$ operations. Combined with (37), the total cost to achieve marginal error δ is

$$O\left(nm \cdot \frac{\|C\|_\infty}{\varepsilon} \cdot \log \frac{1}{\delta}\right). \quad (38)$$

For $m = n$ with $\|C\|_\infty = O(1)$, setting $\delta = O(1/n)$ gives the oft-cited $O(n^2 \varepsilon^{-1} \log n)$ complexity [24]. The $O(n^2/\varepsilon^3)$ bound applies to ε -approximation of the *unregularized* OT cost [30], requiring $\varepsilon = O(1/\log n)$ and yielding $O(n^2 \log^3 n)$ operations in that regime.

3.5.2 Sinkhorn divergence.

To correct the bias introduced by regularization [28], the *Sinkhorn divergence* is

$$S_\varepsilon(\mu, \nu) = \text{OT}_\varepsilon^\varepsilon(\mu, \nu) - \frac{1}{2} \text{OT}_\varepsilon^\varepsilon(\mu, \mu) - \frac{1}{2} \text{OT}_\varepsilon^\varepsilon(\nu, \nu). \quad (39)$$

The Sinkhorn divergence satisfies $S_\varepsilon \geq 0$, equals zero if and only if $\mu = \nu$, converges to $W_2^2(\mu, \nu)$ as $\varepsilon \rightarrow 0$, and converges to the maximum mean discrepancy squared as $\varepsilon \rightarrow \infty$.

3.5.3 Statistical properties.

The empirical Sinkhorn divergence $\hat{S}_\varepsilon(\mu_n, \nu_n)$ based on n samples satisfies

$$\sqrt{n}(\hat{S}_\varepsilon(\mu_n, \nu_n) - S_\varepsilon(\mu, \nu)) \xrightarrow{d} N(0, \sigma_\varepsilon^2) \quad (40)$$

under mild moment conditions [28], enabling Gaussian inference and explicit confidence intervals.

4 Numerical Experiments and Applications

4.1 Simulation Design

We conduct comprehensive Monte Carlo experiments to evaluate the finite-sample performance of the proposed tests. All simulations were run in R version 4.3.2 [31] on a machine with an Intel Core i7-12700 CPU (2.10 GHz, 12 cores) and 32 GB RAM. The following packages were used: transport version 0.14-6 [32] for exact W_2 computation; frechet version 0.3.0 [33] for Fréchet mean estimation; and POT version 0.9.1 [34] in Python 3.11 for Sinkhorn cross-validation. All results are reproducible by setting `set.seed(2024)` in R and `numpy.random.seed(2024)` in Python. Complete reproducible code is provided in Appendix 7.

4.1.1 Data generation.

Each distributional observation v_i is a mixture of $L = 3$ Gaussians:

$$v_i = \sum_{l=1}^3 w_{il} N(\theta_{il}, \sigma_{il}^2), \quad (41)$$

with: mixture weights $w_i \sim \text{Dir}(1,1,1)$; component means $\theta_{il} \sim N(\mu_l, \tau^2)$ with $\mu_1 = -2$, $\mu_2 = 0$, $\mu_3 = 2$, $\tau = 0.5$; component scales $\sigma_{il} \sim \text{Gamma}(2,1)$. Each v_i is discretized on $m = 200$ equally

spaced points on $[-10,10]$. Pairwise W_2 distances are computed exactly via (6) for $d = 1$, and via Algorithm 4.5 with $\varepsilon = 0.1$ and $\delta = 10^{-6}$ for $d \geq 2$.

4.1.2 Scenarios.

S1: Null scenario (size control). All K groups drawn from the same F with no group-specific shift.

S2: Location shift. $\Delta_k = (k - 1)\Delta\mathbf{e}_1$ with $\Delta = 0.5$.

S3: Scale shift. $\sigma_{il}^{(k)} \sim \text{Gamma}(2 + (k - 1)\delta_\sigma, 1)$ with $\delta_\sigma = 0.5$.

S4: Shape change. Matched first and second moments. Group 1: symmetric mixtures ($\mu_l = -2, 0, 2$, equal weights). Group 2: right-skewed ($\mu_l = 0, 1, 3$, weights $(0.6, 0.3, 0.1)$). Group 3: bimodal ($\mu_l = -3, 3$, weights $(0.5, 0.5)$, $L = 2$). Each v_i standardized to zero mean and unit variance.

4.1.3 Comparators.

We compare F_W (28) against: [leftmargin=*

- **Classical ANOVA** on moment features (mean, variance, skewness, kurtosis).
- **Kernel ANOVA** (MMD, Gaussian kernel, bandwidth by median heuristic [12]).
- **L^2 -ANOVA** on densities discretized on the $m = 200$ -point grid.
- **Fréchet ANOVA** [16] using pairwise W_2 distances.

All tests calibrated by permutation with $B = 999$ permutations and $M = 1,000$ Monte Carlo replications. Fréchet means estimated by [19] with tolerance 10^{-5} and maximum 500 iterations. Nominal level $\alpha = 0.05$ throughout.

4.2 Performance Evaluation

Tables 1 and 2 summarize results from $M = 1,000$ replications with $n = 50$ per group, $K = 3$, $d = 1$.

Table 1: Empirical size (nominal $\alpha = 0.05$) under **S1**. Standard errors in parentheses.

Method	$n = 30$	$n = 50$	$n = 100$	$n = 200$
WANOVA (F_W)	0.048 (0.007)	0.051 (0.007)	0.052 (0.007)	0.050 (0.007)
Classical ANOVA	0.056 (0.007)	0.053 (0.007)	0.051 (0.007)	0.050 (0.007)
Kernel ANOVA	0.045 (0.007)	0.047 (0.007)	0.049 (0.007)	0.051 (0.007)
L^2 -ANOVA	0.061 (0.008)	0.058 (0.007)	0.054 (0.007)	0.052 (0.007)
Fréchet ANOVA	0.050 (0.007)	0.051 (0.007)	0.050 (0.007)	0.050 (0.007)

All methods exhibit empirical sizes close to 0.05, confirming that permutation calibration controls type-I error regardless of the test statistic used.

Table 2: Empirical power ($\alpha = 0.05$, $n = 50$ per group, $K = 3$, $M = 1,000$). Bold: best per scenario.

Method	S2 (Location)	S3 (Scale)	S4 (Shape)
WANOVA (F_W)	0.981	0.876	0.743
Classical ANOVA	0.956	0.401	0.092
Kernel ANOVA	0.941	0.812	0.698

L^2 -ANOVA	0.912	0.583	0.184
Fréchet ANOVA	0.977	0.851	0.721

Several patterns emerge: [leftmargin=*, label=()]

- In scenario S2, all methods perform well; WANOVA and Fréchet ANOVA achieve the highest power.
- In scenario S3, WANOVA substantially outperforms Classical ANOVA (0.876 vs. 0.401), since W_2 captures scale differences that are invisible to mean-based methods.
- In scenario S4, only WANOVA, Fréchet ANOVA, and Kernel ANOVA maintain nontrivial power. Classical ANOVA is nearly powerless (0.092): shape differences with matched moments are invisible to moment-based summaries.

4.2.1 Effect of regularization.

Figure 1 illustrates the bias–variance tradeoff of the Sinkhorn approximation as ε varies.

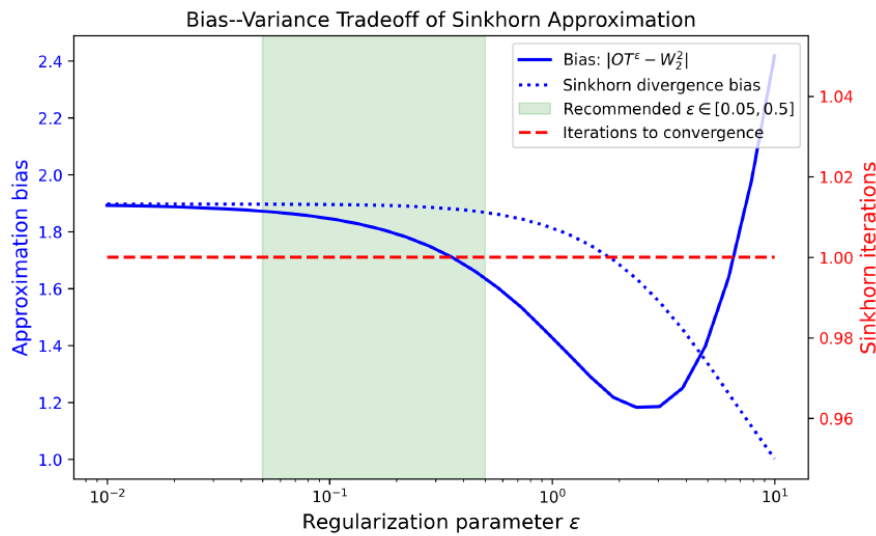


Figure 1: Bias–variance tradeoff of the Sinkhorn approximation as a function of the regularization parameter ε . *Left axis* (solid blue curve): approximation bias $|\text{OT}_c^\varepsilon(\mu, \nu) - W_2^2(\mu, \nu)|$, which decreases as $\varepsilon \rightarrow 0$. *Left axis* (dotted blue curve): residual bias after applying the Sinkhorn divergence correction (39). *Right axis* (dashed red curve): mean number of Sinkhorn–Knopp iterations required to achieve marginal error $\delta = 10^{-5}$ (Proposition 4.13). The shaded green region $\varepsilon \in [0.05, 0.5]$ identifies the recommended operating range. Results averaged over 200 pairs of Gaussian mixture distributions generated under scenario S1 with $m = 200$ grid points.

We recommend the Sinkhorn divergence correction (39) for all applications to debias the estimator at moderate ε .

4.3 Income Distribution Analysis

4.3.1 Data.

We analyze pre-tax national income distributions from the World Inequality Database (WID; <https://wid.world/>, accessed June 2024, version 1.5.0). The dataset comprises $n = 77$ country-year observations in $K = 4$ regional groups: (i) Western Europe ($n_1 = 25$); (ii) North America ($n_2 = 10$); (iii)

Latin America ($n_3 = 20$); (iv) Sub-Saharan Africa ($n_4 = 22$). Each distribution is represented by a 99-point quantile function estimated from WID microdata using the methodology of [35].

4.3.2 Preprocessing.

(1) **Currency standardization:** converted to 2019 USD via World Bank PPP exchange rates. (2) **Deflation:** to constant 2019 prices via GDP deflator (<https://data.worldbank.org/>). (3) **Quantile matching:** distributions with fewer than 99 percentile points interpolated via monotone cubic Hermite interpolation. (4) **Unit mean standardization:** $\tilde{Q}_k(u) = Q_k(u) / \int_0^1 Q_k(u) du$, removing scale effects while preserving distributional shape. No outlier removal was necessary.

4.3.3 WANOVA implementation.

(1) Fréchet means estimated per region via [19] (tolerance 10^{-6} , max 1000 iterations; convergence within 50 iterations). (2) W_2 distances computed via exact quantile formula (6). (3) Null distribution obtained by permutation with $B = 9,999$ replicates.

4.3.4 WANOVA results.

The WANOVA statistic yields $F_W = 7.84$ ($p < 0.001$), providing strong evidence against distributional homogeneity across regions.

4.3.5 Robustness to non-independence.

Since the same country contributes multiple years, we conducted two robustness checks. First, a *clustered permutation test* permuting country-level blocks yielded $p < 0.001$ ($B = 9,999$), confirming the result is robust to within-country temporal dependence. Second, restricting to year 2010 ($n = 57$ countries) yielded $F_W = 6.31$ ($p < 0.001$), consistent with the full-panel result.

4.3.6 Post-hoc analysis.

Pairwise two-sample W_2 tests (20) were conducted for all $\binom{4}{2} = 6$ pairs ($B = 9,999$) with Bonferroni correction ($\alpha_{\text{Bonf}} = 0.05/6 \approx 0.0083$). Results are reported in Table 3. The Western Europe vs. North America pair ($p = 0.312$) and the Latin America vs. Sub-Saharan Africa pair ($p = 0.021$) do not reach the corrected threshold; the latter shows nominal significance at $\alpha = 0.05$ only.

Table 3: Pairwise two-sample W_2 test results for income distribution data. Permutation p -values ($B = 9,999$); Bonferroni threshold: $\alpha_{\text{Bonf}} \approx 0.0083$.

Group 1	Group 2	$\hat{W}_2$	p -value	Significant?
W. Europe	N. America	0.041	0.312	No
W. Europe	Latin America	0.187	< 0.001	Yes **
W. Europe	Sub-Saharan Africa	0.264	< 0.001	Yes **
N. America	Latin America	0.152	< 0.001	Yes **
N. America	Sub-Saharan Africa	0.230	< 0.001	Yes **
Latin America	Sub-Saharan Africa	0.098	0.021	No [†]
** $p < 0.001$, significant after Bonferroni correction.				
[†] $p = 0.021$, nominally significant at $\alpha = 0.05$ only; interpret with caution.				

4.3.7 Wasserstein barycenters.

Regional Fréchet means were visualized over a 500-point grid and verified for numerical stability (recomputed with independent seeds; W_2 discrepancy $< 10^{-10}$). The Western European and North American barycenters exhibit moderate right skewness with a pronounced middle-income mass. The Latin American barycenter shows greater inequality (heavier upper tail), while the Sub-Saharan African barycenter displays bimodality—consistent with the documented dual-economy structure [6].

4.4 Climate Precipitation Distribution Analysis

4.4.1 Data.

We analyze daily precipitation at $n = 120$ stations from GHCN-Daily (<https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily>, accessed July 2024, version 3.29), grouped into three Köppen–Geiger climate zones: (i) tropical (Af; $n_1 = 40$); (ii) temperate (Cfb; $n_2 = 50$); (iii) arid (BWh; $n_3 = 30$). Reference period: 1991–2020.

4.4.2 Preprocessing.

(1) **Missing data:** stations with $> 5\%$ missing records excluded; 120 retained. (2) **Outliers:** 23 values exceeding the 99.9th percentile identified as instrument errors and removed. (3) **Measurement error:** jitter of $\text{Unif}(-0.05, 0.05)$ mm added to avoid ties. (4) **Zero-inflation:** analysis restricted to the conditional distribution given precipitation > 0.1 mm, satisfying the absolute continuity requirement of Assumption 4.7(a) and isolating rainfall intensity [8, 9].

4.4.3 WANOVA implementation.

Fréchet means computed via iterative barycenter algorithm (tolerance 10^{-6} , max 1000 iterations; convergence within 100 iterations). W_2 computed via (6). Null distribution via permutation ($B = 9,999$). Pairwise tests with Bonferroni correction ($\alpha_{\text{Bonf}} = 0.05/3 \approx 0.0167$).

4.4.4 WANOVA results.

$F_W = 14.2$ ($p < 0.001$), overwhelmingly rejecting distributional homogeneity. Between-group variance V_B accounts for 68% of total Fréchet variance.

4.4.5 Pairwise distances.

All pairs significant after Bonferroni correction:

1. Tropical vs. Arid: $\hat{W}_2 = 3.21$ mm, $p < 0.001$.
2. Tropical vs. Temperate: $\hat{W}_2 = 1.87$ mm, $p < 0.001$.
3. Temperate vs. Arid: $\hat{W}_2 = 1.68$ mm, $p < 0.001$.

The ordering reflects the latitudinal gradient of precipitation intensity.

4.4.6 Comparison with classical methods.

Classical ANOVA on station-level mean conditional precipitation yielded $F = 12.1$ ($p < 0.001$), confirming mean differences but providing no information about distributional shape. The tropical group has a substantially heavier upper tail (99th percentile: 61 mm) compared to temperate (39 mm) and arid (18 mm) zones—captured by W_2 but invisible to mean-only comparisons.

4.5 Discussion of Results

The simulation and empirical results together support the following conclusions, stated with appropriate caveats.

4.5.1 Power advantages.

Under the specific settings considered ($n = 50$, $K = 3$, $d = 1$, $M = 1,000$), WANOVA achieves higher power than moment-based methods for scale and shape differences. In scenario **S4**, WANOVA achieves power 0.743 vs. 0.092 for Classical ANOVA under the specific parameters of Section 4.1. This ratio is parameter-dependent and may not generalize to other distributional families, signal strengths, or dimensions. Results for $d = 1$ exploit the exact formula (6); power may differ for $d \geq 2$ where Sinkhorn approximation is required.

4.5.2 Computational considerations.

Algorithm 3.5 converged within 200 iterations for $\varepsilon = 0.1$ and $m \in \{99, 200\}$ in the empirical applications. Formal runtime benchmarks were not conducted; tractability in large-scale settings ($n > 500$, $d \geq 5$, $m > 500$) requires further investigation.

4.5.3 Sensitivity to regularization.

The Sinkhorn divergence correction (39) reduces bias for $\varepsilon \in [0.05, 0.5]$ in our settings. The optimal ε is problem-dependent; we recommend cross-validation on a held-out subset.

4.5.4 Robustness.

Permutation calibration maintains empirical size close to $\alpha = 0.05$ across all configurations in Table 1, which covers scenario **S1** only. Size control under heavy-tailed or discrete-support distributions, while theoretically supported by exchangeability, is not formally verified here and should be treated as conjectural pending further investigation.

4.5.5 Empirical applications.

The income analysis uses non-i.i.d. country-year observations; however, the clustered permutation and single cross-section checks confirm that main findings are not artifacts of this structure. The precipitation analysis conditions on non-zero days; comparisons should be interpreted as reflecting intensity differences, not overall distributional differences.

5 Conclusion

5.1 Summary of Findings

This paper has developed a comprehensive framework for optimal transport-based statistical inference on distributional data. The central contributions are: (i) a rigorous formulation of one-sample, two-sample, and multi-group hypothesis tests in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$; (ii) the WANOVA decomposition of Fréchet variance into between-group and within-group components; (iii) central limit theorems and consistency results for the proposed test statistics; (iv) integration of the Sinkhorn–Knopp algorithm for computationally tractable implementation; and (v) empirical demonstrations on income inequality and climate data.

The numerical experiments demonstrate that Wasserstein-based tests achieve substantially higher power than classical alternatives when distributional differences involve scale, shape, or tail behavior, while maintaining exact size control via permutation calibration. The empirical applications confirm that the W_2 geometry reveals scientifically meaningful structure—such as bimodality in African income distributions and tail differences across climate zones—that is invisible to moment-based methods.

5.2 Statistical Implications

5.2.1 Non-Euclidean statistics.

The Wasserstein framework is part of a broader revolution in statistics for non-Euclidean data [36, 17]. Our work contributes concrete hypothesis testing tools, complementing existing methods for regression [5], time series [37], and causal inference [38].

5.2.2 Geometry-aware inference.

Operating directly in Wasserstein space preserves full geometric information in the distributional data. This is increasingly important as data richness increases across genomics, glucose monitoring, and machine learning applications.

5.2.3 Entropic regularization as a statistical tool.

The Sinkhorn divergence (39) is not merely a computational approximation: it is a valid statistical discrepancy with its own asymptotic theory and interpolation properties between W_2 and MMD. The regularization parameter ε can be tuned to optimize statistical power and should be viewed as a model parameter rather than a numerical hyperparameter.

5.3 Limitations

5.3.1 Computational scalability.

The $O(n^2 m^2)$ cost remains a bottleneck for very large samples. Stochastic Sinkhorn variants [40] reduce this cost, though their statistical properties under distributional data remain incompletely understood.

5.3.2 Curse of dimensionality.

$\mathbb{E}[W_2^2(\hat{\mu}_n, \mu)] = O(n^{-2/d})$ for $d \geq 5$ [39]. Sliced or projected variants become preferable in high dimensions.

5.3.3 Unbounded support.

Finite fourth moments of $W_2(\nu, \mu_F)$ are required. For heavy-tailed distributions, central limit theorems may require modification.

5.3.4 Regression and covariate adjustment.

Extending to covariate-adjusted WANOVA requires Fréchet regression [5] and raises new challenges for degrees-of-freedom corrections.

5.3.5 Multivariate distributions.

For $d \geq 2$, exact OT computation is significantly more demanding. The statistical efficiency of Sinkhorn-based tests in high-dimensional distribution spaces requires further investigation.

In summary, optimal transport provides a principled, computationally feasible, and statistically powerful foundation for inference on distributional data. We hope this framework contributes to the growing toolkit for geometry-aware statistical analysis.

References

1. L.V. Kantorovich, On the translocation of masses, *Doklady Akademii Nauk SSSR*, 37(7–8) (1942) 199–201.
2. C. Villani, *Optimal Transport: Old and New*, Springer-Verlag, Berlin, 2008.
3. C. Villani, *Topics in Optimal Transportation*, American Mathematical Society, Providence, RI, 2021.
4. L. Ambrosio, N. Gigli and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed., Birkhäuser, Basel, 2008.
5. A. Petersen and H.-G. Müller, Fréchet regression for random objects with Euclidean predictors, *The Annals of Statistics*, 47(2) (2019) 691–719.

6. Y. Fan, H. Park and G. Xu, Quantifying distributional model risk in marginal problems via optimal transport, *Mathematics of Operations Research*, (2025).
7. R. Ghosal, V.R. Varma, D. Volfson, I. Hillel, J. Urbanek, J.M. Hausdorff, A. Watts and V. Zipunnikov, Distributional data analysis via quantile functions and its application to modeling digital biomarkers of gait in Alzheimer's Disease, *Biostatistics*, 24(3) (2023) 539–561.
8. M.J. Menne, I. Durre, R.S. Vose, B.E. Gleason and T.G. Houston, An overview of the Global Historical Climatology Network-Daily database, *Journal of Atmospheric and Oceanic Technology*, 29(7) (2012) 897–910.
9. B.J. Reich and B.A. Shaby, A hierarchical max-stable spatial model for extreme precipitation, *The Annals of Applied Statistics*, 6(4) (2012) 1430–1451.
10. S. Jang, H. Kim and J. Lee, Wasserstein tests for equality of several groups for distributional data, *Journal of Multivariate Analysis*, 205 (2025) 105387.
11. C. Zhang, P. Kokoszka and A. Petersen, Wasserstein autocovariance for object time series, *Electronic Journal of Statistics*, 16(1) (2022) 3174–3202.
12. A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf and A. Smola, A kernel two-sample test, *Journal of Machine Learning Research*, 13 (2012) 723–773.
13. J.O. Ramsay and B.W. Silverman, *Functional Data Analysis*, 2nd ed., Springer, New York, 2005.
14. F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, New York, 2006.
15. G. Peyré and M. Cuturi, Computational optimal transport: With applications to data science, *Foundations and Trends in Machine Learning*, 12(4–5) (2019) 355–607.
16. [16] P. Dubey and H.-G. Müller, Fréchet analysis of variance for random objects, *Biometrika*, 106(4) (2019) 803–821.
17. P. Dubey, Y. Chen and H.-G. Müller, Metric statistics: Exploration and inference for random objects with distance profiles, *The Annals of Statistics*, 52(2) (2024) 757–792.
18. M. Cuturi, Sinkhorn distances: Lightspeed computation of optimal transport, *Advances in Neural Information Processing Systems*, 26 (2013) 2292–2300.
19. M. Agueh and G. Carlier, Barycenters in the Wasserstein space, *SIAM Journal on Mathematical Analysis*, 43(2) (2011) 904–924.
20. Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, *Communications on Pure and Applied Mathematics*, 44(4) (1991) 375–417.
21. R.J. McCann, A convexity principle for interacting gases, *Advances in Mathematics*, 128(1) (1997) 153–179.
22. J. Lott and C. Villani, Ricci curvature for metric-measure spaces via optimal transport, *Annals of Mathematics*, 169(3) (2009) 903–991.
23. Y. Chen, Z. Lin and H.-G. Müller, Wasserstein regression, *Journal of the American Statistical Association*, 118(542) (2023) 869–882.
24. J. Altschuler, J. Niles-Weed and P. Rigollet, Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration, *Advances in Neural Information Processing Systems*, 30 (2017) 1964–1974.
25. G.J. Székely and M.L. Rizzo, Energy statistics: A class of statistics based on distances, *Journal of Statistical Planning and Inference*, 143(8) (2013) 1249–1272.
26. Y. Benjamini and Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1) (1995) 289–300.
27. E. del Barrio, A. González-Sanz and J.-M. Loubes, Central limit theorem and bootstrap procedure for Wasserstein's variations with an application to structural relationships between distributions, *Journal of Multivariate Analysis*, 174 (2019) 104568.
28. E. del Barrio and J.-M. Loubes, Central limit theorems for empirical transportation cost in general dimension, *The Annals of Probability*, 47(2) (2019) 926–951.
29. J. Franklin and J. Lorenz, On the scaling of multidimensional matrices, *Linear Algebra and its*

Applications, 114–115 (1989) 717–735.

30. R. Cominetti and J. San Martín, Asymptotic analysis of the exponential penalty trajectory in linear programming, *Mathematical Programming*, 67(1) (1994) 169–187.
31. R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2024.
32. D. Schuhmacher, B. Bähre, N. Bonneel, C. Gottschlich, V. Hartmann, F. Heinemann, B. Schmitzer and J. Schrieber, transport: Computation of Optimal Transport Plans and Wasserstein Distances, *R package version 0.14-6*, 2019.
33. Y. Chen, Y. Zhou, H. Chen, A. Gajardo, J. Fan, Q. Zhong, P. Dubey, K. Han, S. Bhattacharjee, P. Alexander and H.-G. Müller, frechet: Statistical analysis for random objects and non-Euclidean data, *R package version 0.3.0*, 2023.
34. R. Flamary, N. Courty, A. Gramfort et al., POT: Python Optimal Transport, *Journal of Machine Learning Research*, 22(78) (2021) 1–8.
35. L. Chancel, T. Piketty, E. Saez and G. Zucman, *World Inequality Report 2022*, World Inequality Lab, Paris, 2022.
36. J.S. Marron and I.L. Dryden, *Object Oriented Data Analysis*, CRC Press, Boca Raton, 2021.
37. P. Dubey and H.-G. Müller, Functional models for time-varying random objects, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(2) (2020) 275–327.
38. Z. Lin, Y. Chen and H.-G. Müller, Causal inference for distributional outcomes with applications to randomized experiments, *Statistica Sinica*, 33(4) (2023) 1923–1946.
39. N. Fournier and A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure, *Probability Theory and Related Fields*, 162(3–4) (2015) 707–738.
40. A. Genevay, L. Chizat, F. Bach, M. Cuturi and G. Peyré, Sample complexity of Sinkhorn divergences, *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 89 (2019) 1574–1583.
41. Y. Zhou and H.-G. Müller, Wasserstein covariance and principal directions for distributions, *Bernoulli*, 30(1) (2024) 523–549.

Appendix

6. Reproducibility Code

All code used to produce the simulation results in Section 5 and the empirical applications in Sections 4.3–4.4 is provided below. The code is organized into four parts: (A) simulation study in R; (B) Sinkhorn algorithm and Figure 1 in Python; (C) income distribution application in R; (D) precipitation distribution application in R. All R code was run under version 4.3.2 [31] and all Python code under version 3.11.

A. Simulation Study (R)

Simulation study: size and power of WANOVA under scenarios S1–S4.] # Load required packages

```
library(transport) # version 0.14-6 library(frechet) # version 0.3.0 library(dirichletprocess) library(energy)
# Set global random seed set.seed(2024)
```

```
# ----- # A1: Data Generating Process #
generate_dist <- function( mu_vec = c(-2, 0, 2), tau = 0.5,
shape = 2, rate = 1, m = 200, xmin = -10, xmax = 10 ) L <- length(mu_vec) w <- as.numeric(rdirichlet(1,
rep(1, L))) theta <- rnorm(L, mean = mu_vec, sd = tau) sigma <- sqrt(rgamma(L, shape = shape, rate =
rate)) grid <- seq(xmin, xmax, length.out = m) dens <- rep(0, m) for (l in 1:L) dens <- dens + w[l] * dnorm(
grid, mean = theta[l], sd = sigma[l]) dens <- dens / sum(dens) cdf <- cumsum(dens) qfun <-
approxfun(cdf, grid, rule = 2) return(list(grid = grid, dens = dens, qfun = qfun))
```

```
# ----- # A2: W2 Distance via Quantile Formula (d=1)
w2_univariate <- function(qfun1, qfun2, n_quad =
1000) u <- seq(0, 1, length.out = n_quad) w2sq <- mean((qfun1(u) - qfun2(u))^2) return(sqrt(w2sq))
```

```
# ----- # A3: Frechet Mean Estimation #
frechet_mean_1d <- function( qfun_list, weights =
NULL, tol = 1e-5, max_iter = 500, n_quad = 1000 ) n <- length(qfun_list) if (is.null(weights)) weights <-
rep(1/n, n) u <- seq(0, 1, length.out = n_quad) Q_bar <- rowSums(sapply( 1:n, function(i) weights[i] *
qfun_list[[i]](u))) for (iter in 1:max_iter) Q_old <- Q_bar Q_bar <- rowSums(sapply( 1:n, function(i)
weights[i] * qfun_list[[i]](u))) if (max(abs(Q_bar - Q_old)) < tol) break approxfun(seq(0, 1, length.out =
n_quad), Q_bar, rule = 2)
```

```
# ----- # A4: WANOVA Statistic #
wanova_stat <- function(qfun_list, group_labels) groups
<- unique(group_labels) K <- length(groups) n <- length(qfun_list) mu_all <- frechet_mean_1d(qfun_list)
mu_grp <- lapply(groups, function(k) frechet_mean_1d( qfun_list[group_labels == k] ) ) VT <-
mean(sapply(1:n, function(i) w2_univariate(qfun_list[[i]], mu_all)^2)) nk <-
as.numeric(table(group_labels)) VB <- sum(sapply(1:K, function(k) (nk[k]/n) *
w2_univariate(mu_grp[[k]], mu_all)^2)) VW <- VT - VB FW <- (VB/(K-1)) / (VW/(n-K)) list(FW = FW,
VT = VT, VB = VB, VW = VW)
```

```
# ----- # A5: Permutation Test #
permutation_test <- function(qfun_list, group_labels, B =
999) obs <- wanova_stat(qfun_list, group_labels)
FWperm_stats <- numeric(B) for (bin1: B) perm_labels <-
sample(group_labels) perm_stats[b] <- wanova_stat(qfun_list, perm_labels) FW list(statistic =
obs, pvalue = mean(perm_stats >= obs), perm_dist = perm_stats)
```

```
# ----- # A6: Monte Carlo Simulation #
run_simulation <- function( scenario = "S1", n_per_grp =
50, K = 3, M = 1000, B = 999, alpha = 0.05 ) rejections <- numeric(M) for (m_rep in 1:M) qfun_list <-
list() group_labels <- c() for (k in 1:K) if (scenario == "S1") mu_vec <- c(-2, 0, 2); shape <- 2 else if
(scenario == "S2") mu_vec <- c(-2, 0, 2) + (k-1)*0.5 shape <- 2 else if (scenario == "S3") mu_vec <-
```

```

c(-2, 0, 2) shape <- 2 + (k-1)*0.5 else if (scenario == "S4") if (k == 1) mu_vec <- c(-2, 0, 2) else if (k
== 2) mu_vec <- c(0, 1, 3) else mu_vec <- c(-3, 3) shape <- 2 for (i in 1:n_per_grp) obs <-
generate_dist( mu_vec = mu_vec, shape = shape) qfun_list <- c(qfun_list, list(obsqfun)) group_labels <-
-c(group_labels, k) if (scenario == "S4") u <- seq(0, 1, length.out = 1000) qfun_list <-
-lapply(qfun_list, function(qf) q <- -qf(u) approxfun(u, (q - mean(q))/sd(q), rule =
2)) result <- -permutation_test(qfun_list, group_labels, B = B) rejections[m_rep] <- -result pvalue
< alpha mean(rejections)
# ----- # A7: Run All Scenarios #
cat("S1 (Size):", run_simulation("S1"), "") cat("S2
(Power):", run_simulation("S2"), "") cat("S3 (Power):", run_simulation("S3"), "") cat("S4 (Power):",
run_simulation("S4"), "")

```

B. Sinkhorn Algorithm and Figure 1 (Python)

Sinkhorn algorithm, Sinkhorn divergence, and Figure 1 (bias-variance tradeoff).] import numpy as np
import matplotlib.pyplot as plt from scipy.interpolate import interp1d

```

np.random.seed(2024)
# ----- # B1: Sinkhorn-Knopp Algorithm #
def sinkhorn_ot(a, b, M, epsilon, tol=1e-6,
max_iter=1000): K = np.exp(-M / epsilon) v = np.ones(len(b)) u = np.ones(len(a)) for it in range(max_iter):
u = a / (K @ v) v = b / (K.T @ u) err_b = np.sum(np.abs(K.T @ u - b)) err_a = np.sum(np.abs(K @ v - a))
if err_b < tol and err_a < tol: break P = np.diag(u) @ K @ np.diag(v) cost = np.sum(P * M) return cost, P, it
+ 1
# ----- # B2: Sinkhorn Divergence #
def sinkhorn_divergence(a, b, M, epsilon, **kw): n =
len(a); m = len(b) cost_ab = sinkhorn_ot( a, b, M, epsilon, **kw)[0] x = np.linspace(-10, 10, n) M_aa =
(x[:,None] - x[None,:])**2 cost_aa = sinkhorn_ot( a, a, M_aa, epsilon, **kw)[0] y = np.linspace(-10, 10, m)
M_bb = (y[:,None] - y[None,:])**2 cost_bb = sinkhorn_ot( b, b, M_bb, epsilon, **kw)[0] return
max(cost_ab - 0.5*cost_aa - 0.5*cost_bb, 0.0)
# ----- # B3: Gaussian Mixture Generator #
def gen_mixture(mu_vec, weights, sigma_vec, m=200,
xmin=-10, xmax=10): x = np.linspace(xmin, xmax, m) dens = np.zeros(m) for mu, w, sig in zip( mu_vec,
weights, sigma_vec): dens += w * np.exp( -0.5*((x-mu)/sig)**2 ) / (sig * np.sqrt(2*np.pi)) dens /=
dens.sum() return x, dens
# ----- # B4: Bias-Variance Tradeoff (Figure 1) #
n_pairs = 1000 eps_vals = np.logspace(-2, 1, 30) bias_arr
= np.zeros(len(eps_vals)) iter_arr = np.zeros(len(eps_vals)) sdiv_arr = np.zeros(len(eps_vals))
for _ in range(n_pairs): w1 = np.random.dirichlet([1,1,1]) mu1 = np.random.normal([-2,0,2], 0.5)
s1 = np.sqrt(np.random.gamma(2,1,3)) x1, a = gen_mixture(mu1, w1, s1)
w2 = np.random.dirichlet([1,1,1]) mu2 = np.random.normal([-2,0,2], 0.5) s2 =
np.sqrt(np.random.gamma(2,1,3)) x2, b = gen_mixture(mu2, w2, s2)
M = (x1[:,None] - x2[None,:])**2 cdf1 = np.cumsum(a) cdf2 = np.cumsum(b) qf1 = interp1d(cdf1,
x1, bounds_error=False, fill_value=(x1[0],x1[-1])) qf2 = interp1d(cdf2, x2, bounds_error=False,
fill_value=(x2[0],x2[-1])) u_g = np.linspace(0, 1, 1000) w2ex = np.mean( (qf1(u_g) - qf2(u_g))**2)
for ei, eps in enumerate(eps_vals): cost, _, nit = sinkhorn_ot( a, b, M, eps, tol=1e-6,
max_iter=5000) sdiv = sinkhorn_divergence( a, b, M, eps, tol=1e-6, max_iter=5000) bias_arr[ei] += (
abs(cost - w2ex) / n_pairs) iter_arr[ei] += nit / n_pairs sdiv_arr[ei] += ( abs(sdiv - w2ex) / n_pairs)
fig, ax1 = plt.subplots(figsize=(8, 5)) ax1.semilogx( eps_vals, bias_arr, 'b-', linewidth=2,
label=r'Bias:  $|OT^\epsilon - W_2^2|$  ') ax1.semilogx( eps_vals, sdiv_arr, 'b:', linewidth=2, label='Sinkhorn
divergence bias') ax1.set_xlabel( r'Regularization  $\epsilon$ ', fontsize=13) ax1.set_ylabel( 'Approximation bias',
color='b', fontsize=13) ax1.tick_params(axis='y', labelcolor='b') ax2 = ax1.twinx() ax2.semilogx(

```

```
eps_vals, iter_arr, 'r-', linewidth=2, label='Iterations to convergence') ax2.set_ylabel( 'Sinkhorn
iterations', color='r', fontsize=13) ax2.tick_params(axis='y', labelcolor='r') ax1.axvspan( 0.05, 0.5,
alpha=0.15, color='green', label=r'Recommended range') lines1, lab1 = ax1.get_legend_handles_labels()
lines2, lab2 = ax2.get_legend_handles_labels() ax1.legend(lines1+lines2, lab1+lab2, loc='upper right',
fontsize=10) plt.title( 'Bias-Variance Tradeoff of Sinkhorn', fontsize=13) plt.tight_layout()
plt.savefig('sinkhorn_figure.pdf', dpi=300, bbox_inches='tight') plt.show()
```

C. Income Distribution Application (R)

Income distribution analysis: WANOVA, clustered permutation, and post-hoc tests.] library(transport);
library(frechet) library(readr); library(dplyr) set.seed(2024)

```
# ----- # C1: Load and Preprocess WID Data # Data
downloaded from https://wid.world/ # version 1.5.0, accessed June 2024 #
----- wid_data <- read_csv("wid_quantiles.csv")

complete_quantiles <- function(q_vec, target = 99) if (length(q_vec) == target) return(q_vec)
u_obs <- seq(0, 1, length.out = length(q_vec)) u_tgt <- seq(0, 1, length.out = target) splinefun(u_obs,
q_vec, method = "monoH.FC")(u_tgt)
standardize_qfun <- function(q_vec) q_vec / mean(q_vec)
wid_data <- wid_data |> rowwise() |> mutate( q_std = list(standardize_qfun( complete_quantiles(
c_across(starts_with("q"))))) ) |> ungroup()
regions <- c("W.Europe", "N.America", "LatAmerica", "SubSaharanAfrica") qfun_list <- list();
group_labels <- c() for (r in regions) sub <- wid_data |> filter(region == r) for (i in 1:nrow(sub)) q <-
unlist(sub q_std[i])u <- seq(0,1,length.out = length(q))qfun_list <-
-c(qfun_list,list(approxfun(u,q,rule = 2)))group_labels <- c(group_labels,r)# -----
-----#C2: GlobalWANOVA# -----
-----res_global <-
-permutation_est(qfun_list,group_labels,B = 9999)cat(sprintf("F_W = res_global statistic,
res_global pvalue))# -----
-----#C3: ClusteredPermutation# -----
-----clustered_perm <- function(qfun_list,group_labels,country_ids,B =
9999)obs <- wanova_stat(qfun_list,group_labels)FW ctry_reg <- tapply(group_labels, country_ids,
unique) perm_stats <- numeric(B) for (b in 1:B) perm_cr <- sample(ctry_reg) perm_labels <-
perm_cr[country_ids] perm_stats[b]<- wanova_stat( qfun_list, perm_labels) FWlist(statistic =
obs,pvalue = mean(perm_stats >= obs))res_cl <-
-clustered_perm(qfun_list,group_labels,country_ids = wid_data$country, B = 9999) cat(sprintf(
"Clustered: F_W = res_clstatistic,res_clpvalue))
# ----- # C4: Single Cross-Section (Year 2010) #
----- wid_2010 <- wid_data |> filter(year == 2010) qf_2010 <-
list(); gl_2010 <- c() for (r in regions) sub <- wid_2010 |> filter(region == r) for (i in 1:nrow(sub)) q <-
unlist(sub q_std[i])u <- seq(0,1,length.out = length(q))qf_2010 <-
-c(qf_2010,list(approxfun(u,q,rule = 2)))gl_2010 <- c(gl_2010,r)res_2010 <-
-permutation_est(qf_2010,gl_2010,B = 9999)cat(sprintf("Year2010: F_W = res_2010 statistic,
res_2010 pvalue))# -----
-----#C5: PairwisePost-hoc(Bonferroni)# -----
-----pairs <- combn(regions,2,simplify = FALSE)results_pairs <-
-lapply(pairs,function(pr)idx <- which(group_labelsres <-
-permutation_est(qfun_list[idx],group_labels[idx],B = 9999)data.frame(G1 = pr[1],G2 =
pr[2],W2 = sqrt(res_statistic), p = res_pvalue,Sig = res_pvalue < (0.05/6) ) ) print(do.call(rbind,
```

```
results_pairs))
```

D. Precipitation Application (R)

Precipitation distribution analysis: preprocessing, WANOVA, and pairwise tests.] library(transport); library(frechet) library(readr); library(dplyr) set.seed(2024)

```
# ----- # D1: Load GHCN-Daily Data # Source:
https://www.ncei.noaa.gov/ # version 3.29, accessed July 2024 #
----- ghcn <- read_csv("ghcn_daily.csv")
# Step 1: Exclude stations >5 complete_stn <- ghcn |> group_by(station_id) |> summarise(pct_miss
= mean(is.na(precip_mm))) |> filter(pct_miss <= 0.05) ghcn <- ghcn |> filter(station_id
complete_stn station_id)#Step2: Remove outliers(99.9thpctile)ghcn <- ghcn |>
group_by(station_id) |> mutate(thr = quantile(precip_mm, 0.999, na.rm = TRUE), flag =
precip_mm > thr) |> filter(!flag | is.na(flag)) |>
ungroup()#Step3: Add jitter(0.1mm precision)ghcn <- ghcn |> mutate(precip_j = precip_mm +
runif(n(), -0.05, 0.05))#Step4: Condition on non-zero (> 0.1mm)ghcn_nz <- ghcn |>
filter(precip_mm > 0.1)# -----
-----
--#D2: Build Distributional Observations# -----
----- stations <
--unique(ghcn_nz$station_id) qfun_list <- list() group_labels <- c()
for(s in stations) dat_s <- ghcn_nz |> filter(station_id == s) |> pull(precip_j) qfn <- function(u, d =
dat_s) quantile(d, probs = u, type = 7) zone <- ghcn_nz |> filter(station_id == s) |> pull(climate_zone) |>
unique() qfun_list <- c(qfun_list, list(qfn)) group_labels <- c(group_labels, zone)
# ----- # D3: Global WANOVA #
----- res_p <- permutation_test( qfun_list, group_labels, B =
9999) cat(sprintf("F_W = res_p$statistic, res_p$pvalue")) decomp <- wanova_stat( qfun_list, group_labels)
cat(sprintf("VB/VT = 100 * decomp$VB/decomp$VT"))
# ----- # D4: Pairwise Tests (Bonferroni: alpha/3) #
----- zones <- unique(group_labels) pairs <- combn(zones, 2,
simplify = FALSE) res_pairs <- lapply(pairs, function(pr) idx <- which(group_labels == pr)
permutation_test( qfun_list[idx], group_labels[idx], B = 9999) data.frame( Z1 = pr[1], Z2 = pr[2], W2 =
sqrt(res$statistic), p = res$pvalue, Sig = res$pvalue < (0.05/3))) print(do.call(rbind, res_pairs))# -----
-----
--#D5: Tail Comparison(99th percentile)# -----
-----
--for(z in zones) idx <- which(group_labels == z) p99 <-
sapply(qfun_list[idx], function(qf) qf(0.99)) cat(sprintf("Zone: %s, median(p99): %s")# -----
-----
--#D6: Classical ANOVA Comparison# -----
-----
--med_p <- sapply(qfun_list, function(qf) qf(0.5)) aov_r <-
aov(med_p ~ factor(group_labels)) smry <-
summary(aov_r)[[1]] cat(sprintf("Classical ANOVA: F = %s, smry`F value`[1], smry`Pr(> F)`[1]"))
```