Asia Pac. J. Math. 2026 **13**:189

AFFINE-INVARIANT ROBUST TWO-WAY MANOVA VIA WEIGHTED GENERAL LINEAR HYPOTHESES

ABDULLAH A. AMEEN[1,*], OSAMA H. ABBAS[2], MOHAMMED S. KADHIM[1], THAMER W. HUSSAIN[1]

[1]Department of Mathematics, College of Science, University of Basrah, Basrah, Iraq

[2]General Directorate of Education in Basrah, Basrah, Iraq

[*]Corresponding author: abdullah.amin@uobasrah.edu.iq

Received Aug. 30, 2026

Abstract. Classical two-way MANOVA can be affected by a few unusual observations within one factorial cell. We consider an affine-invariant Wilks-type test based on cellwise minimum covariance determinant (MCD) estimates. Robust distances are used to assign either hard or smoothly redescending weights. These weights are then applied to general linear hypothesis tests for row, column and interaction effects. The tests use coefficient contrasts, so the hypotheses do not change when weighted cell totals differ or the observed cell counts are unequal. For a fully specified common location–scatter reference law, affine equivariance removes the nuisance cell means and scatter. Exchangeability then provides a valid finite-sample Monte Carlo test. We also establish bounded direct influence of sufficiently remote observations, provided the initial MCD estimates remain stable. Simulations assess performance under clean data, cellwise contamination, masking, unequal sample sizes, increasing response dimension, alternative tuning choices, and heterogeneous covariance structures. They also compare the GLH approach with weighted-SSP and MATS tests. Applications to two morphometric datasets examine the resulting tests and the influence assigned to individual observations.

2020 Mathematics Subject Classification. Primary 62H15; Secondary 62F35, 62J10, 62F40.

Key words and phrases. two-way MANOVA; robust multivariate inference; minimum covariance determinant; general linear hypothesis; affine invariance; Monte Carlo test.

1. Introduction

Two-way multivariate analysis of variance (MANOVA) tests structured differences among vector-valued means when observations are cross-classified by two factors. Wilks' Lambda provides a well-established likelihood-ratio test with a geometric interpretation through the multivariate general linear hypothesis (GLH) and an extensive Gaussian theory [1–5]. Outliers can have an arbitrarily large effect

DOI: 10.28924/APJM/13-189

©2026 Asia Pacific Journal of Mathematics

on sample means and covariance estimates. In two-way layouts, a small cluster in one cell may be mistaken for a factorial effect while also changing the estimated residual scatter.

High-breakdown estimators provide an alternative to ordinary location and scatter estimates in this setting. We base our procedure on the minimum covariance determinant (MCD) estimator [6–12]. MCD-based methods have been considered for one-way MANOVA [13], and S- and MM-estimators provide related alternatives [14]. For balanced two-way designs, Spangl [15] used reweighted MCD estimates to construct robust factorial SSP matrices. Other studies have investigated robust Wilks statistics for one- and two-way MANOVA [16,17]. Ameen and Abbas also compared robust one-way Wilks statistics using simulations and an application [34]. Suwaidan and Ameen examined robust one-way ANOVA [35].

Other statistical studies have considered estimation and inspection design for interval-censored lifetime models [36–39], and mean-square solutions of random differential equations [40,41]. These studies address problems distinct from robust factorial MANOVA.

An important difficulty arises when case weights are introduced into a factorial decomposition. Although the original design may be balanced, the total retained weight can differ across cells. Simply substituting weighted means into a balanced-design SSP decomposition may then produce a matrix that no longer corresponds directly to the hypothesis of interest. We instead retain the saturated multivariate linear model and define every factorial effect through an explicit coefficient contrast. The resulting weighted GLH is meaningful for balanced and unequal observed cell sizes and does not require weighted orthogonality of marginal decompositions.

Calibration presents a related difficulty. Because the weights are estimated from the observations entering the determinant ratio, its null distribution is no longer given by the classical Wilks law. When the standardised reference law is specified, the weighted test has the same null distribution for cell means satisfying the null hypothesis and for any common scatter matrix. We can therefore use independent samples from that law to obtain critical values for a finite-sample Monte Carlo test. Gaussian calibration is therefore one important special case, not the only law covered by the invariance argument. If the standardized core distribution is unspecified, this conclusion does not extend to distribution-free testing.

We first center each cell robustly and estimate scatter from the pooled centered observations. Hard or smoothly redescending weights are then used in GLH tests defined by coefficient contrasts. Those contrasts remain fixed as effective cell sizes change. Affine invariance and exchangeability provide the basis for finite-sample calibration under a specified reference model. The intended setting is a common positive-definite location–scatter model with possible outliers. Exact finite-sample calibration also requires a specified standardized reference distribution. Models with heterogeneous or singular

covariance matrices, or with response dimension large relative to sample size, require other methods [18, 20–24, 26–28].

2. Relation to existing factorial MANOVA procedures

Resistance to individual outliers and validity under unequal covariance matrices are different objectives. General factorial procedures such as robust heteroscedastic tests [19], the bootstrap methods of Konietschke et al. [22], and MATS [23] allow heterogeneous group covariance matrices; MATS can additionally accommodate singular covariance matrices. Nonparametric multivariate contrasts [25] and quantile-based MANOVA [27] provide additional factorial alternatives; the latter changes the estimand from means to componentwise quantiles and thereby gains robustness through the target functional. Recent work has further broadened multivariate contrast testing in general covariate-adjusted factorial designs [28]. These approaches are important comparators, but they address inferential robustness that is not identical to high-breakdown resistance against a small group of atypical cases.

Spangl [15] proposed a closely related robust two-way MANOVA method. It uses reweighted MCD estimates from each cell to construct factorial SSP matrices for a balanced design. Ameen and Jaaze [17] retained the weighted-SSP statistic, but changed the cutoff weights and estimated its reference distribution by simulation. In our approach, each factorial hypothesis is a coefficient contrast in a saturated multivariate model. This also permits unequal cell sizes, provided the weighted fit is admissible. We give the calibration and stability results for the corresponding weighted GLH test.

The two tests can differ if their cell weights have different totals. In a balanced design with equal weighted totals, the usual factorial contrasts give identical GLH and weighted-SSP statistics. If the totals differ, the SSP margin reflects the relative retained cell weights, but the GLH continues to test the prespecified contrast. Section 7.3 compares the two formulas using identical hard weights, so their difference is attributable to the construction of the test. Table 1 summarizes the resulting methodological position relative to representative factorial MANOVA procedures.

3. Model, hypotheses, and admissibility

Let $\boldsymbol{y}_{ijk} \in \mathbb{R}^p$ denote the kth response vector in cell (i, j), with $i = 1, \ldots, r$, $j = 1, \ldots, c$, and $k = 1, \ldots, n_{ij}$. Stack all responses in the $N \times p$ matrix $\boldsymbol{Y}$, where $N = \sum_{ij} n_{ij}$. We use the saturated two-factor multivariate linear model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B}_0 + \boldsymbol{U}, \tag{1}$$

where $\boldsymbol{X}$ contains an intercept, $r-1$ row columns, $c-1$ column columns, and $(r-1)(c-1)$ interaction columns. Sum-to-zero effect contrasts are used, so row and column coefficients represent equally weighted marginal effects over the levels of the other factor. Other full-rank parameterizations give the same tests when coefficient contrasts are transformed consistently.

Table 1. Position of the proposed procedure relative to representative factorial MANOVA methods.

Procedure	Factorial scope	Covariance structure	Outlier handling	Calibration
Classical Wilks	General GLH	Common nonsingular	None	Classical Wilks law
Spangl RMCD	Balanced two-way	Common scatter	Hard MCD reweighting	Simulation-based
Ameen–Jaaze (2020)	Balanced two-way	Common scatter	Alternative finite-cutoff MCD weighting	Approximate reference law
MATS	General factorial	Heterogeneous; singular allowed	Not casewise high-breakdown	Resampling
Quantile MANOVA	General factorial	Heterogeneous	Robust estimand	Bootstrap
Proposed GLH	Balanced/unequal-cell	Common positive-definite location–scatter	Hard or redescending MCD weights	Exact MC for specified core law

Let $\boldsymbol{L}_{\mathrm{R}}$, $\boldsymbol{L}_{\mathrm{C}}$, and $\boldsymbol{L}_{\mathrm{I}}$ select the row, column, and interaction coefficient blocks. The factorial hypotheses are

$$H_{0g} : \boldsymbol{L}_g \boldsymbol{B}_0 = \boldsymbol{0}, \qquad g \in \{\mathrm{R}, \mathrm{C}, \mathrm{I}\}. \tag{2}$$

The same representation applies to balanced and unequal-cell designs, as long as all required cells are present and $\boldsymbol{X}$ has full column rank.

Let $\boldsymbol{W} = \mathrm{diag}(w_1, \ldots, w_N)$, with $w_s \geq 0$. A weighted test is admissible only when $\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}$ is nonsingular and the residual SSP matrix is positive definite. These conditions are treated as part of the statistical definition rather than as numerical inconveniences.

Proposition 1 (Admissibility in a saturated cell layout)**.** *Let $\boldsymbol{X}$ be any full-rank parameterization of a saturated two-factor model with every observed cell represented. Then $\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}$ is nonsingular if and only if every cell has strictly positive total weight. Conditional on this condition, the weighted residual SSP is positive definite if and only if the weighted residual matrix has column rank p.*

Proof. A saturated effect-coded design is related by a nonsingular column transformation to the cell-indicator design. For the latter design, $\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}$ is diagonal with entries equal to the cellwise sums of weights. Nonsingularity is therefore equivalent to a positive effective mass in every cell, and the property is preserved by nonsingular reparameterization. The residual SSP can be written as

$\boldsymbol{R}^\top \boldsymbol{W}\boldsymbol{R} = (\boldsymbol{W}^{1/2}\boldsymbol{R})^\top(\boldsymbol{W}^{1/2}\boldsymbol{R})$, which is positive definite exactly when $\boldsymbol{W}^{1/2}\boldsymbol{R}$ has column rank p. □

Accordingly, singular weighted fits are reported as inadmissible. No generalized inverse of $\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{X}$ and no artificial diagonal perturbation of a singular residual SSP is used to manufacture a statistic.

4. Cellwise robust geometry and weighted GLH statistics

4.1. **Cellwise MCD estimates and case weights.** Within cell (i, j), we calculate the MCD location $\widehat{\boldsymbol{\mu}}_{ij}^{(0)}$ from h_{ij} observations, where $h_{ij} > p$ and the subset fraction is h_{ij}/n_{ij}. The centred response in each cell is

$$\boldsymbol{z}_{ijk} = \boldsymbol{y}_{ijk} - \widehat{\boldsymbol{\mu}}_{ij}^{(0)}. \tag{3}$$

After centering, pool the response vectors and apply MCD again to estimate the common initial scatter matrix $\widehat{\boldsymbol{C}}^{(0)}$. Centering within cells keeps differences in factorial means from being absorbed into the scatter estimate. The robust location and scatter calculations use FAST-MCD with the Gaussian consistency and finite-sample corrections [8,10]. The main numerical study takes the subset fraction to be 0.75.

For each observation, calculate the squared robust distance as

$$D_{ijk}^2 = \boldsymbol{z}_{ijk}^\top \big(\widehat{\boldsymbol{C}}^{(0)}\big)^{-1} \boldsymbol{z}_{ijk}. \tag{4}$$

We consider two weighting schemes. The hard reweighted-MCD rule is

$$w_{ijk}^{\mathrm{MCD}} = \mathbb{I}\big\{D_{ijk}^2 \le \chi_{p,0.975}^2\big\}, \tag{5}$$

and the smoothly redescending rule is

$$w_{ijk}^{\mathrm{RED}} = \left(1 - \frac{D_{ijk}^2}{c_p^2}\right)^2 \mathbb{I}\big\{D_{ijk}^2 < c_p^2\big\}, \qquad c_p^2 = \chi_{p,0.9975}^2. \tag{6}$$

The initial robust estimates are held fixed when weights are assigned. Thus, case influence is reduced without further contraction of the estimated scatter matrix.

4.2. **Weighted GLH statistic.** For a specified admissible $\boldsymbol{W}$, define

$$\widehat{\boldsymbol{B}}_w = (\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{Y} \tag{7}$$

and

$$\boldsymbol{E}_w = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}_w)^\top \boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}_w). \tag{8}$$

For effect g, let

$$\boldsymbol{V}_{wg} = \boldsymbol{L}_g(\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{L}_g^\top \tag{9}$$

and

$$\boldsymbol{H}_{wg} = (\boldsymbol{L}_g\widehat{\boldsymbol{B}}_w)^\top \boldsymbol{V}_{wg}^{-1}(\boldsymbol{L}_g\widehat{\boldsymbol{B}}_w). \tag{10}$$

The determinant ratio and the reported monotone transform are

$$\Lambda_g^{(w)} = \frac{|\boldsymbol{E}_w|}{|\boldsymbol{E}_w + \boldsymbol{H}_{wg}|}, \qquad T_g^{(w)} = -\log \Lambda_g^{(w)}. \tag{11}$$

Determinants are computed using positive-definite matrix factorizations. If the admissibility conditions are not met, no value is assigned to the statistic.

Proposition 2 (Connection with the classical Wilks statistic)**.** *Setting $\boldsymbol{W} = \boldsymbol{I}_N$ in (11) gives the usual multivariate GLH Wilks statistic for H_{0g}. In a balanced two-way layout with standard factorial contrasts, this is also the classical Wilks statistic for the row, column, or interaction effect from the orthogonal SSP decomposition.*

Proof. When $\boldsymbol{W} = \boldsymbol{I}_N$, (7) reduces to ordinary least squares. Equations (8)–(10) then give the usual residual and hypothesis SSP matrices. In a balanced two-way design, they give the standard SSP matrices for the row, column and interaction effects. □

Proposition 3 (Invariance under positive weight scaling)**.** *If $\boldsymbol{W}$ is admissible and $a > 0$, multiplying every weight by a does not alter $\widehat{\boldsymbol{B}}_w$, $\Lambda_g^{(w)}$ or $T_g^{(w)}$.*

Proof. The common factor a cancels when the weighted least-squares coefficients are calculated. Scaling the weights by a multiplies $\boldsymbol{E}_w$ by a and divides $\boldsymbol{V}_{wg}$ by a; consequently, $\boldsymbol{H}_{wg}$ is also multiplied by a. Both determinants are therefore multiplied by a^p, so their ratio stays the same. □

An immediate implication is that the mean of the continuous weights does not measure efficiency, because the statistic is unchanged by a common positive rescaling. We report the scale-invariant Kish effective sample size $(\sum_s w_s)^2 / \sum_s w_s^2$, as well as the number of zero weights and selected weight quantiles.

5. Finite-sample invariance and Monte Carlo inference

We state the calibration result for a common location–scatter family, which includes but is not limited to Gaussian errors. Let the rows of $\boldsymbol{Z}$ be independent observations from a fixed, absolutely continuous, standardized distribution F_0 on $\mathbb{R}^p$. Let $\boldsymbol{A}$ be nonsingular.

Theorem 1 (Removal of nuisance parameters under a specified reference law)**.** *Suppose*

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B}_0 + \boldsymbol{Z}\boldsymbol{A}^\top, \tag{12}$$

where the rows of $\boldsymbol{Z}$ are independently and identically distributed according to F_0. Assume the MCD location and scatter selections are affine equivariant and almost surely unique under F_0. Under H_{0g}, the robust distances and the final weights are identical whether calculated from $\boldsymbol{Y}$ or $\boldsymbol{Z}$. Moreover,

$$T_g^{(w)}(\boldsymbol{Y}) = T_g^{(w)}(\boldsymbol{Z}). \tag{13}$$

For fixed design and tuning choices and a given standardized reference distribution F_0, it follows that the null law does not depend on nuisance cell means satisfying H_{0g} or on the common positive-definite scatter matrix $\boldsymbol{A}\boldsymbol{A}^\top$.

Proof. In a saturated cell model, $\boldsymbol{X}\boldsymbol{B}_0$ has the same value for every observation within a given cell. Translation equivariance of the cellwise MCD estimate removes the common mean within each cell. Affine equivariance of the pooled scatter estimator then yields $\widehat{\boldsymbol{C}}^{(0)}(\boldsymbol{Y}) = \boldsymbol{A}\widehat{\boldsymbol{C}}^{(0)}(\boldsymbol{Z})\boldsymbol{A}^\top$ after centering. Therefore, the robust squared distances and final observation weights agree in the two datasets. With the same weight matrix, $\widehat{\boldsymbol{B}}_w(\boldsymbol{Y}) = \boldsymbol{B}_0 + \widehat{\boldsymbol{B}}_w(\boldsymbol{Z})\boldsymbol{A}^\top$. Under H_{0g} the contrast removes $\boldsymbol{B}_0$, and both the residual and hypothesis SSP matrices transform by congruence with $\boldsymbol{A}$. Both determinants are multiplied by $|\boldsymbol{A}|^2$ and the factor cancels from their ratio. □

Corollary 1 (Gaussian and other fully specified reference models)**.** *The theorem applies in particular when F_0 is standard multivariate Gaussian. The same conclusion holds for another fully specified standardized location–scatter distribution, such as a standardized multivariate t law with specified parameters. The reference law itself remains part of the null model.*

Remark 1 (Randomized FAST-MCD computation)**.** *FAST-MCD uses random elemental starts to approximate the MCD optimum. Let U denote the random-start stream. Conditional on a fixed stream, the concentration steps are affine equivariant in exact arithmetic because the same observation indices are selected initially and every subsequent Mahalanobis ordering is preserved by a nonsingular response transformation. Consequently, Theorem 1 applies conditionally on U when the same stream is used to compare an observed dataset with an affine transform. In Monte Carlo calibration, the algorithm can use a fixed stream of random starts. Alternatively, independent streams $U_0, \ldots, U_B$ can be drawn from the same distribution for each dataset. In the latter case, the pairs $(\mathbf{Y}_b, U_b)$ are exchangeable. The invariance property is exact in theory. In Section 7 we check the small numerical differences caused by rounding and the random starting values.*

The test is not distribution-free. When the standardised core distribution is not known, its shape must still be accounted for. A Gaussian reference may therefore give conservative results for heavy-tailed data, even though location and common scatter have been removed.

Let T_0 denote the statistic calculated from the data. Generate $T_1, \ldots, T_B$ independently from F_0, using the same design, weighting procedure, tuning constants and admissibility rule. Define

$$p_{\mathrm{MC}} = \frac{1 + \sum_{b=1}^{B} \mathbb{I}(T_b \geq T_0)}{B+1}. \tag{14}$$

Theorem 2 (Validity of the Monte Carlo test in finite samples)**.** *Write $\mathcal{A}$ for the event that a weighted fit is admissible. Assume that the observed fit is admissible. Calculate each reference statistic from an independent null sample and keep that sample only if $\mathcal{A}$ also holds. Under the conditions of Theorem 1, and conditional on $\mathcal{A}$ for the observed sample, $T_0, T_1, \ldots, T_B$ are exchangeable under H_{0g}. If their common conditional distribution is*

continuous, then

$$\Pr_{H_{0g}}(p_{\mathrm{MC}} \le \alpha \mid \mathcal{A}) = \frac{\lfloor \alpha(B+1) \rfloor}{B+1} \le \alpha. \tag{15}$$

If ties occur, the non-randomised test in (14) *is still conservative conditional on* $\mathcal{A}$.

Proof. By Theorem 1, the observed and simulated statistics have the same standardised null distribution. The transformations used there also leave admissibility unchanged. After conditioning on $\mathcal{A}$, we retain only independent reference samples that meet this condition. The resulting $B+1$ statistics are exchangeable. If their conditional distribution is continuous, each possible rank of the observed statistic is equally likely, giving (15). The weak upper-tail rule remains conservative when there are ties [29,30]. □

Remark 2. *The argument requires common scatter across cells. If the cell covariance matrices differ, one nonsingular matrix* $\boldsymbol{A}$ *cannot standardise all groups. The nuisance-invariance result then fails, and a method for heteroscedastic factorial data is needed* [18,20–23].

6. Gross-error stability of finite-cutoff weighting

Valid calibration does not necessarily mean that a test is resistant to extreme observations. We study a sequence in which the uncontaminated data remain fixed and selected observations become increasingly distant. The result requires stable initial robust estimates and weights with a finite cutoff.

Theorem 3 (Stability under increasingly remote contamination)**.** *Consider a fixed factorial design and a specified set* $\mathcal{C}$ *of contaminated observations. Consider a sequence of samples* $\boldsymbol{Y}^{(t)}$ *in which the observations outside* $\mathcal{C}$ *remain fixed while*

$$\min_{s \in \mathcal{C}} \|\boldsymbol{y}_s^{(t)}\| \longrightarrow \infty.$$

Assume that, along this sequence, every cellwise MCD location remains bounded and the pooled MCD scatter has eigenvalues uniformly bounded away from zero and infinity. Under either rule in (5)–(6)*, all observations in* $\mathcal{C}$ *eventually receive weight zero as* t *increases. If the remaining weights converge and the limiting fit is admissible, the statistic* $T_g^{(w)}(\boldsymbol{Y}^{(t)})$ *has a finite limit for every factorial effect* g.

Proof. Write $\widehat{\boldsymbol{\mu}}_{ij}^{(0,t)}$ and $\widehat{\boldsymbol{C}}^{(0,t)}$ for the initial robust estimates. As the cell locations remain bounded, the centred norm of every contaminated observation in $\mathcal{C}$ tends to infinity. If the largest eigenvalue of $\widehat{\boldsymbol{C}}^{(0,t)}$ is uniformly bounded by $M < \infty$, then

$$D_s^2(t) \ge M^{-1} \|\boldsymbol{y}_s^{(t)} - \widehat{\boldsymbol{\mu}}_{ij}^{(0,t)}\|^2 \longrightarrow \infty.$$

Each weighting rule has a finite distance cutoff. Thus, sufficiently distant contaminated observations receive zero weight and make no direct contribution to $\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}$, $\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{Y}$ or the residual SSP matrix.

The remaining weights converge, and the limiting fit is admissible. Hence, the normal equations, coefficient estimates and SSP matrices converge. Matrix inversion and the log determinant are continuous for positive-definite matrices, so the statistic has a finite limit. □

For uncontaminated data in general position, MCD locations remain bounded and the scatter estimate remains nonsingular below the finite-sample replacement breakdown point [11, 12]. This result does not apply after that point is exceeded. Provided the initial estimates stay stable, observations that become sufficiently distant are assigned zero weight in the final fit.

7. Numerical study

7.1. **Simulation procedure.** We considered a 3×2 factorial design with $n = 30$ observations per cell and $p = 2$ response variables. The baseline model had standard Gaussian errors, and tests were carried out at the 0.05 level. We used an MCD subset fraction of 0.75. We set the cutoff at the 0.975 quantile for hard weights and at 0.9975 for redescending weights. Critical values came from 3500 accepted datasets. A different set of 2500 datasets was used to estimate the null rejection rates. Selected power estimates were based on 3000 accepted datasets. Each alternative altered only the first response mean, allowing the row, column and interaction effects to be examined separately.

A constant shift in a cell mean is removed when that cell is centred. We can therefore reuse the robust distances from each simulated error sample at different effect sizes, without approximating the test statistic. In each figure, markers identify the simulated values; shape-preserving curves connect them. We report rejection rates as binomial proportions and give their Monte Carlo standard errors to show sampling uncertainty.

7.2. **Gaussian calibration and test size.** Table 2 reports Gaussian critical values and results from an independent null experiment. The observed rejection rates of all three tests are close to 0.05 for each factorial effect. The robust tests have higher critical values on the $-\log \Lambda$ scale because their weights depend on the data. Thus, the standard Wilks cutoffs cannot be used for these tests. Figure 1 shows the observed row-effect rejection rates against the nominal levels.

7.3. **Comparison with the earlier weighted-SSP method.** To compare the construction of the tests rather than their robust estimates, we use the weighted-SSP form developed from the balanced approach of Spangl and related work [15, 17]. The example has a balanced 3×2 design with 30 observations per cell. Both statistics receive the same binary weights, while progressively more distant observations in one cell are given zero weight. Both methods use the same weights and estimation inputs; the diagnostic varies only the differences in total retained weight among cells.

With equal effective masses the two row statistics coincide numerically. As one cell loses effective mass, the weighted-SSP row margin increasingly reflects the retained-mass proportions whereas the

Table 2. Gaussian critical values and null rejection rates. The critical values are upper 5% points of $-\log \Lambda$. Rejection rates are based on 2500 independent samples.

Quantity	Method	Row	Column	Interaction
Critical value	Classical	0.0537	0.0349	0.0573
	RMCD-GLH	0.1533	0.0988	0.1556
	Redescending GLH	0.2047	0.1336	0.2034
Empirical size	Classical	0.0528	0.0488	0.0480
	RMCD-GLH	0.0460	0.0480	0.0492
	Redescending GLH	0.0480	0.0452	0.0468

GLH continues to test the equally weighted row coefficient contrast. The difference is reported in Table 3 and Figure 2. Although it is small for this particular sample, it illustrates that the procedures no longer calculate the same finite-sample statistic when weighting removes factorial orthogonality.

Table 3. Structural comparison of the two statistics with a common set of hard weights. "Zeroed" is the number of zero weights introduced in one of the six originally balanced cells; mass CV is the coefficient of variation of the six effective cell masses.

Zeroed	w_{11}	Mass CV	GLH T	SSP T	$\lvert T_{\mathrm{GLH}} - T_{\mathrm{SSP}} \rvert$
0	30	0.000	0.1780	0.1780	0.0000
5	25	0.064	0.2311	0.2238	0.0073
10	20	0.132	0.2160	0.2058	0.0101
15	15	0.203	0.2167	0.2025	0.0143

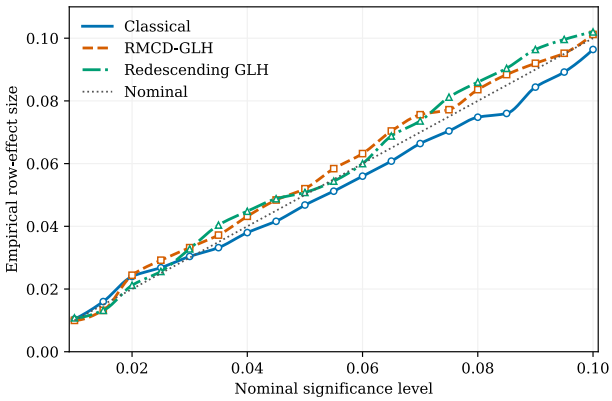

Figure 1. Row-effect size across nominal levels. Markers show evaluated levels; the diagonal denotes exact calibration.

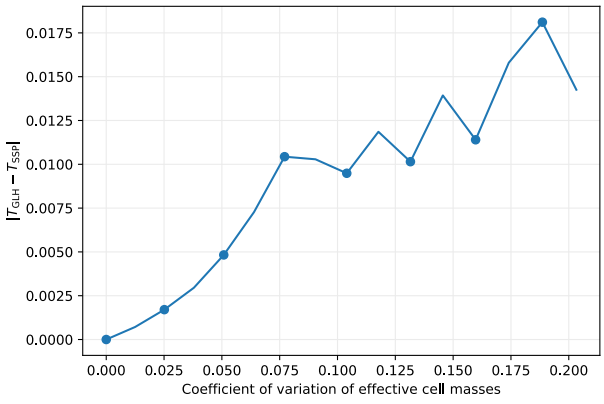

Figure 2. Difference between weighted GLH and weighted-SSP row statistics as effective cell masses become unequal under identical hard weights.

7.4. **Comparison with selected alternative tests.** We also compare the proposed row-effect tests with two benchmark procedures. Rank-Wilks applies the classical GLH statistic after componentwise ranking and is calibrated by an independent Gaussian reference experiment. MATS-PBS is the published modified ANOVA-type statistic equipped with the parametric bootstrap recommended by Friedrich and Pauly [23]: conditionally on each observed dataset, bootstrap observations are generated independently within each cell from centered multivariate normal laws having the corresponding empirical covariance matrices, and the MATS statistic is recomputed. The benchmark uses 499 inner bootstrap draws and 5000 outer Monte Carlo datasets per reported scenario.

Table 4 reports clean null size, clean power at $d = 0.75$, null rejection under a 10% localized cell shift, power when the same contamination masks a true effect in the opposite direction, and rejection under the heteroscedastic boundary model. MATS-PBS controls the heteroscedastic boundary size, as expected from its design, but does not target cellwise gross-error resistance. Rank-Wilks is comparatively stable under the localized shift, at the cost of changing the response geometry. The proposed procedures sacrifice clean Gaussian efficiency in exchange for stronger stability under localized contamination while preserving affine geometry of the original response scale.

Table 4. Targeted row-effect comparison. "Contamination size" is null rejection under 10% localized cell shift; "Masked power" combines $d = 0.75$ with an opposite-direction 10% cell shift. MATS-PBS uses 499 parametric bootstrap draws within each of 5000 outer datasets.

Method	Clean size	Clean power	Contamination size	Masked power	Heteroscedastic size
Classical Wilks	0.0528	0.928	0.140	0.316	0.051
Rank-Wilks	0.0472	0.898	0.059	0.760	0.051
MATS-PBS	0.0500	0.923	0.089	0.374	0.048
RMCD-GLH	0.0460	0.663	0.050	0.631	0.144
Redescending GLH	0.0480	0.629	0.049	0.614	0.105

7.5. **Clean-model power and efficiency.** Classical Wilks is most powerful under its uncontaminated Gaussian reference model. Figures 3–5 and Table 5 quantify the efficiency cost of robustification. The loss is most visible for intermediate effect magnitudes and for the more difficult interaction alternative. The hard and redescending procedures become similar as the signal becomes large.

Table 5. Selected clean-model powers from 3000 replications.

Effect	Method	$d = 0.25$	$d = 0.50$	$d = 0.75$	$d = 1.00$	$d = 1.25$	$d = 1.50$
Row	Classical	0.170	0.572	0.928	0.996	1.000	1.000
Row	RMCD-GLH	0.108	0.308	0.663	0.914	0.990	0.999
Row	Redescending GLH	0.103	0.304	0.629	0.898	0.985	0.999
Column	Classical	0.291	0.846	0.995	1.000	1.000	1.000
Column	RMCD-GLH	0.169	0.568	0.916	0.995	1.000	1.000
Column	Redescending GLH	0.155	0.542	0.898	0.992	1.000	1.000
Interaction	Classical	0.064	0.148	0.310	0.531	0.757	0.911
Interaction	RMCD-GLH	0.047	0.087	0.169	0.297	0.468	0.636
Interaction	Redescending GLH	0.053	0.095	0.172	0.304	0.466	0.634

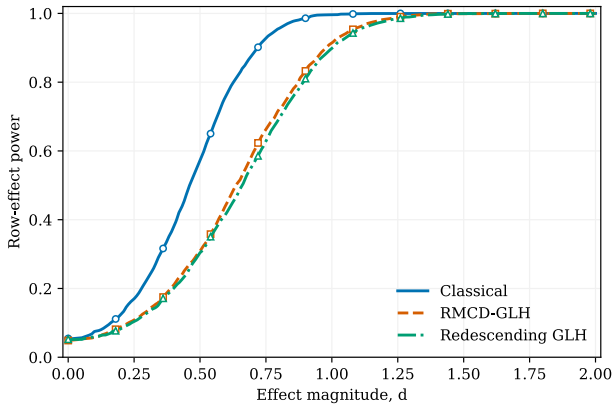

Figure 3. Row-effect power versus effect magnitude. Colored trajectories are shape-preserving connections through the Monte Carlo grid.

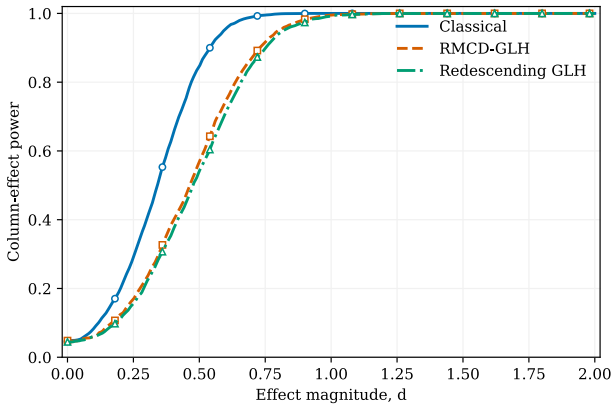

Figure 4. Column-effect power versus effect magnitude.

7.6. **Localized contamination, masking, distributional stress, and model boundaries.** The principal robustness experiment contaminates a proportion of one cell by adding a six-standard-deviation shift to the first response. Under the null this contamination can imitate a row effect. Under masking, a genuine row effect with $d = 0.75$ is combined with contamination in the opposite direction. Table 6 shows selected null stress scenarios.

With more shifted observations, the classical row test increasingly rejects a true null hypothesis. Figure 6, however, shows that both robust tests have null rejection rates near 0.05 throughout the contamination range considered. Under masking, Figure 7 shows how rejection power changes as the opposite-direction contaminated fraction increases. Under standardized t_5 and t_3 errors, Gaussian calibration makes the robust procedures conservative. Theorem 1 explains how to handle a scientifically

specified heavy-tailed core: the Monte Carlo reference law should match that standardized model rather than remain Gaussian.

The heteroscedastic row in Table 6 is intentionally included as a boundary diagnostic. The robust weighted tests are not valid heteroscedastic MANOVA procedures; their rejection rates inflate because covariance heterogeneity violates the common-scatter pivotal argument. In that setting MATS, heteroscedastic bootstrap MANOVA, or related procedures are appropriate [22,23,27].

Table 6. Null rejection rates in the selected stress scenarios, based on 1000 replications for each scenario.

Scenario	Method	Row	Column	Interaction
Gaussian	Classical	0.048	0.047	0.052
	RMCD-GLH	0.057	0.049	0.056
	Redescending GLH	0.051	0.045	0.059
t(5)	Classical	0.053	0.038	0.036
	RMCD-GLH	0.028	0.032	0.029
	Redescending GLH	0.023	0.032	0.022
t(3)	Classical	0.046	0.043	0.038
	RMCD-GLH	0.021	0.030	0.024
	Redescending GLH	0.009	0.024	0.017
5% cell shift	Classical	0.085	0.064	0.060
	RMCD-GLH	0.054	0.047	0.042
	Redescending GLH	0.054	0.044	0.046
10% cell shift	Classical	0.134	0.106	0.100
	RMCD-GLH	0.045	0.046	0.042
	Redescending GLH	0.040	0.040	0.038
15% cell shift	Classical	0.232	0.142	0.209
	RMCD-GLH	0.032	0.037	0.052
	Redescending GLH	0.025	0.036	0.042
10% cell scatter	Classical	0.057	0.064	0.051
	RMCD-GLH	0.035	0.045	0.047
	Redescending GLH	0.036	0.041	0.045
Heteroscedastic	Classical	0.051	0.041	0.043
	RMCD-GLH	0.144	0.103	0.144
	Redescending GLH	0.105	0.081	0.113

Table 7 also reports Monte Carlo uncertainty for the main row-effect results. The number of simulations, Monte Carlo standard errors, and Wilson intervals help distinguish differences in rejection rates from ordinary simulation variability.

Table 7. Precision of the main row-effect estimates. MCSE denotes the binomial Monte Carlo standard error. The final column reports the 95% Wilson confidence interval.

Metric	Method	N	Estimate	MCSE	95% Wilson CI
Clean-data null rejection	Classical Wilks	2500	0.0528	0.0045	[0.0447, 0.0623]
	RMCD-GLH	2500	0.0460	0.0042	[0.0385, 0.0549]
	Redescending GLH	2500	0.0480	0.0043	[0.0403, 0.0571]
Null rejection, 10% cell shift	Classical Wilks	800	0.1400	0.0123	[0.1177, 0.1658]
	RMCD-GLH	800	0.0500	0.0077	[0.0369, 0.0674]
	Redescending GLH	800	0.0490	0.0076	[0.0361, 0.0662]
Power with masking	Classical Wilks	800	0.3160	0.0164	[0.2847, 0.3490]
	RMCD-GLH	800	0.6310	0.0171	[0.5970, 0.6637]
	Redescending GLH	800	0.6140	0.0172	[0.5798, 0.6471]

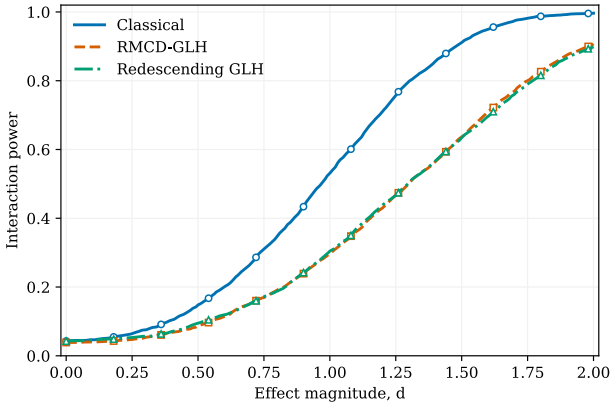

Figure 5. Power of the interaction test across the effect sizes considered.

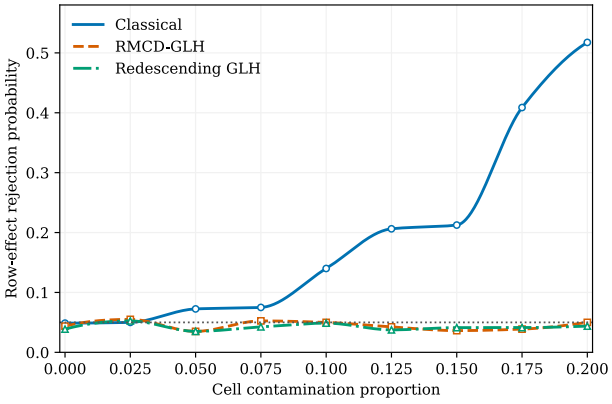

Figure 6. Row-effect null rejection with contamination in one cell. Markers indicate the contamination proportions.

7.7. **Unequal cell sizes, number of responses and tuning.** The unequal-cell design has six cells, with sizes 20, 30, 25, 35, 40 and 15. Table 8 shows that rejection rates are close to the nominal level with clean data. With a 10% shift in one cell, the robust tests remain less affected by contamination. The coefficient hypothesis $\boldsymbol{L}_g\boldsymbol{B}_0$ can be used directly despite the unequal counts, without constructing weighted balanced-design margins.

As response dimension grows, the MCD scatter estimate must be obtained from relatively fewer observations per variable. With $n = 40$ observations per cell, we vary p and report the results in Table 9 and Figure 8. Clean-model power declines as p grows, whereas the robust procedures remain less

Table 8. Rejection frequencies in the unequal-cell experiment.

Method	Clean			10% contaminated		
	Row	Column	Interaction	Row	Column	Interaction
Classical	0.0550	0.0393	0.0500	0.1336	0.0879	0.1436
RMCD-GLH	0.0529	0.0564	0.0571	0.0343	0.0400	0.0514
Redescending GLH	0.0429	0.0529	0.0536	0.0357	0.0457	0.0486

affected than classical Wilks by the localized contamination considered. The proposed approach is therefore best suited to low or moderate response dimension with sufficient data per cell, rather than to settings in which p approaches or exceeds the cell size.

Table 9. Row-effect results for different response dimensions.

p	Method	Clean size	Clean power	Contamination size
2	Classical	0.047	0.993	0.196
2	Redescending GLH	0.047	0.784	0.040
2	RMCD-GLH	0.041	0.797	0.036
5	Classical	0.049	0.930	0.146
5	Redescending GLH	0.054	0.729	0.046
5	RMCD-GLH	0.034	0.764	0.049
8	Classical	0.053	0.833	0.106
8	Redescending GLH	0.049	0.579	0.073
8	RMCD-GLH	0.071	0.656	0.061

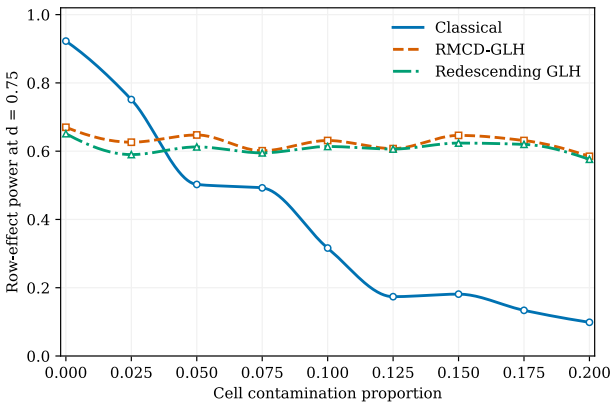

Figure 7. Power of the row effect at $d = 0.75$ with opposite-direction masking contamination.

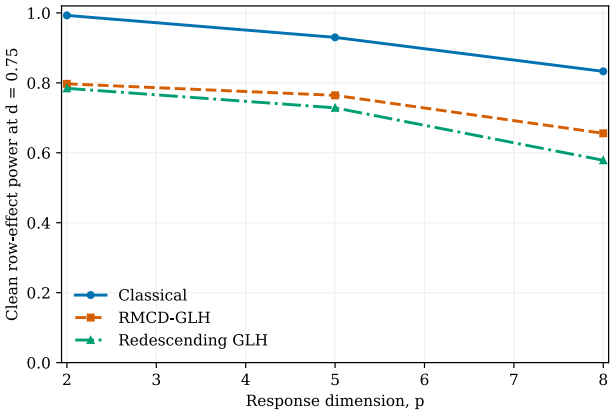

Figure 8. Clean row-effect power at $d = 0.75$ as response dimension increases.

Table 10 examines MCD subset fractions and redescending cutoffs. Clean power is also expressed relative to classical Wilks at $d = 0.75$. A larger subset improves Gaussian-model efficiency at the expense of breakdown resistance. Within the cutoffs examined, performance varies less. We use the

stated default values as a reasonable choice for these simulations, not as settings that are best for every dataset.

Table 10. Effect of tuning on the redescending row test. Relative efficiency is the ratio of clean-model power to classical Wilks power at $d = 0.75$.

Setting	h/n	Cutoff q	Clean size	Contamination size	Clean power	Relative efficiency
$h = 0.65$	0.650	0.9975	0.056	0.048	0.506	0.545
$h = 0.75$	0.750	0.9975	0.040	0.036	0.624	0.672
$h = 0.85$	0.850	0.9975	0.040	0.040	0.774	0.834
$q = 0.990$	0.750	0.9900	0.054	0.050	0.582	0.627
$q = 0.995$	0.750	0.9950	0.024	0.040	0.638	0.687
$q = 0.9975$	0.750	0.9975	0.047	0.049	0.622	0.670
$q = 0.999$	0.750	0.9990	0.056	0.047	0.689	0.742

Figure 9 compares clean-data power with null rejection under contamination for three MCD subset fractions. We use $h/n = 0.75$ as the more protective setting. When little contamination is expected, a larger subset can improve Gaussian power without much change in the contaminated-null rejection rate in this experiment. It also reduces the finite-sample replacement breakdown point, however, so the subset fraction should reflect the data and the aims of the analysis. Around 0.9975, changes in the redescending cutoff have less effect than changes in the subset fraction.

7.8. **Numerical checks of invariance.** We checked invariance using 30 randomly chosen nonsingular affine transformations. We also rescaled the weights by positive constants. Table 11 gives the largest differences. They are close to machine precision, in agreement with Theorem 1 and Proposition 3.

Table 11. Maximum differences in the affine and weight-scaling checks.

Diagnostic	Discrepancy	Interpretation
Affine transformation, 30 trials	4.83e-13	Maximum absolute change in statistic
Positive weight rescaling	$<$ 1e-12	Statistic unchanged analytically and numerically

The statistic is not defined for a singular weighted fit, so we record how often this occurs. For each setting in Table 12, we examined 100 independent datasets. In these runs, the Kish effective sample sizes were adequate for estimating the model and residual scatter. The smallest reported values apply only to the simulated samples and are not theoretical lower bounds. Inadmissible fits were also recorded in the larger experiments.

Table 12. Admissibility and Kish effective sample sizes from 100 independent samples in each setting.

	Invalid proportion		Minimum Kish ESS		Median Kish ESS	
Scenario	RMCD	RED	RMCD	RED	RMCD	RED
Gaussian	0.000	0.000	162.0	153.4	173.5	161.6
10% cell shift	0.000	0.000	163.0	151.5	171.0	159.4
Unequal cells	0.000	0.000	152.0	142.5	159.0	148.6
$p = 8$, $n = 40$ per cell	0.000	0.000	216.0	187.3	228.0	197.6

8. Morphometric applications

8.1. **Rock-crab data.** We first consider the rock-crab data reported by Campbell and Mahon [31]. The sample contains 200 crabs in four equally sized species-by-sex groups. Each crab has five body measurements. Because the test is affine invariant, standardising the responses leaves the determinant ratios and test results unchanged.

Table 13 reports the test statistics and Monte Carlo p-values under a Gaussian reference. All methods indicate effects of species, sex and their interaction. The robust tests give the same significance decisions, but also show how much influence individual crabs have on the analysis. Figure 11 gives the ordered weights. Table 14 lists the Kish effective sample sizes, weight quantiles and numbers of zero weights for both rules.

Table 13. Rock-crab morphometry. Monte Carlo results are based on 499 independent reference draws.

Method	Effect	$-\log \Lambda$	p_{MC}
Classical	Species	2.117	0.002
Classical	Sex	1.471	0.002
Classical	Species $\times$ Sex	0.259	0.002
RMCD-GLH	Species	2.245	0.002
RMCD-GLH	Sex	1.540	0.002
RMCD-GLH	Species $\times$ Sex	0.320	0.002
Redescending GLH	Species	2.410	0.002
Redescending GLH	Sex	1.823	0.002
Redescending GLH	Species $\times$ Sex	0.459	0.002

A direct construction check is also informative. Using the same hard MCD weights, the weighted-SSP and GLH statistics are close but not identical once reweighting induces unequal effective cell masses. The crab analysis thus shows the same distinction examined in Section 7.3: explicit coefficient contrasts

Table 14. Rock-crab observation-weight diagnostics. Kish ESS and quantiles are invariant to a common positive rescaling of all weights.

Method	Sum w	Zero w	Kish ESS	$Q_{0.10}$	$Q_{0.50}$	$Q_{0.90}$
Classical	200	0	200	1.000	1.000	1.000
RMCD-GLH	180.0	20	180.0	0.900	1.000	1.000
Redescending GLH	103.0	3	157.3	0.104	0.557	0.827

retain the intended factorial hypothesis even when weighting removes orthogonality from a marginal SSP decomposition.

8.2. **Sensitivity analysis using the crab measurements.** We use the observed crab data to examine the behavior described by Theorem 3 under an empirical within-cell covariance structure. We use a controlled change to the crab measurements to examine sensitivity; it is not chosen to produce a particular conclusion. Removing the observed mean from each species-by-sex cell creates an empirical null configuration while retaining the original within-cell dependence and marginal shapes. Ten observations closest to the multivariate center of one cell are then shifted together in the first standardized response. The imposed shifts are part of the sensitivity experiment and are not treated as biological effects.

Figure 10 plots each row statistic relative to its own 95% Gaussian-reference critical value as the added shift increases over a finely resolved grid. With increasing shifts, the classical statistic passes its reference threshold and stays above it. Neither finite-cutoff statistic crosses its own threshold over the range shown. Once the perturbed observations separate from the stable robust geometry, their weights change abruptly and the robust trajectories fluctuate within a bounded range. Those finite-sample switches are expected for cutoff weighting and are shown without smoothing across the evaluated grid.

8.3. **Palmer penguin data.** The second application uses Palmer Archipelago penguin morphometrics [32, 33]. Removing records with missing sex or morphometric measurements leaves 333 complete cases. Species and sex define an unequal 3×2 design. The four responses are bill length, bill depth, flipper length and body mass. Thus, this dataset also allows us to examine unequal cell sizes with a multivariate response of dimension four.

Table 15 indicates clear effects of species and sex and a smaller interaction effect. The Monte Carlo values are reported at the finite resolution implied by the reference experiment rather than as zero probabilities. Relative to the crab analysis, fewer observations receive very low weights, and both robust methods retain a high scale-invariant effective sample size (Table 16). The ordered weights appear in Figure 12.

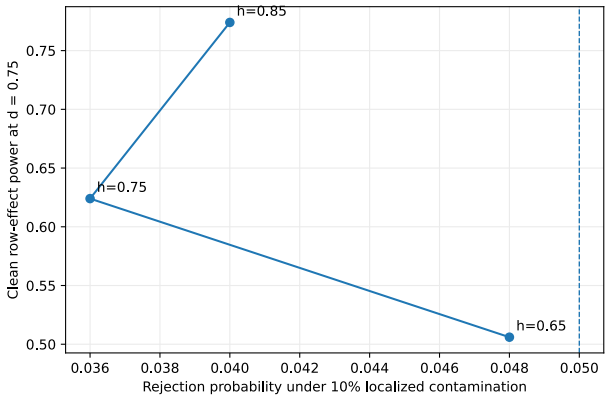

Figure 9. Robustness–efficiency frontier as the MCD subset fraction changes; the vertical reference marks nominal 5% rejection under the 10% contamination null.

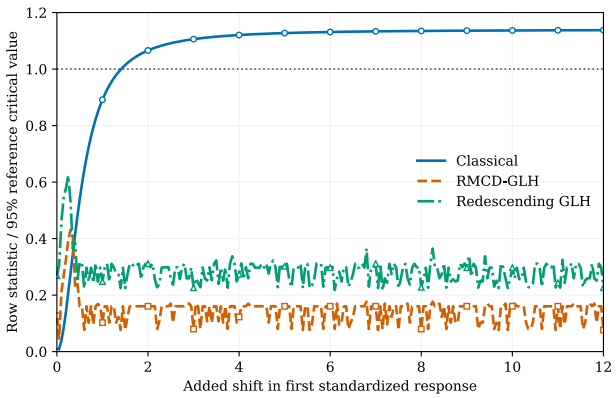

Figure 10. Controlled empirical-null crab sensitivity experiment. Curves are row statistics divided by their method-specific 95% reference critical values.

Table 15. Palmer penguin morphometric results from 199 independent Monte Carlo reference draws.

Method	Effect	$-\log \Lambda$	p_{MC}
Classical	Species	4.112	0.005
Classical	Sex	0.956	0.005
Classical	Species $\times$ Sex	0.118	0.005
RMCD-GLH	Species	4.423	0.005
RMCD-GLH	Sex	1.054	0.005
RMCD-GLH	Species $\times$ Sex	0.149	0.005
Redescending GLH	Species	4.813	0.005
Redescending GLH	Sex	1.156	0.005
Redescending GLH	Species $\times$ Sex	0.221	0.005

Table 16. Palmer penguin observation-weight diagnostics.

Method	Total w	Zero weights	Kish n_{eff}	$Q_{0.10}$	$Q_{0.50}$	$Q_{0.90}$
Classical	333	0	333	1.000	1.000	1.000
RMCD-GLH	317.0	16	317.0	1.000	1.000	1.000
Redescending GLH	196.2	6	284.2	0.241	0.633	0.889

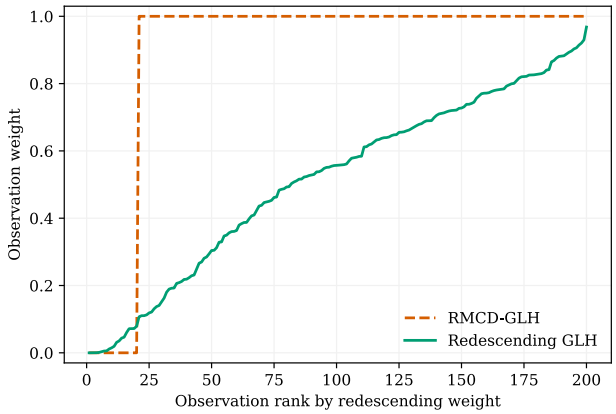

Figure 11. Rock-crab observation weights ordered by redescending weight.

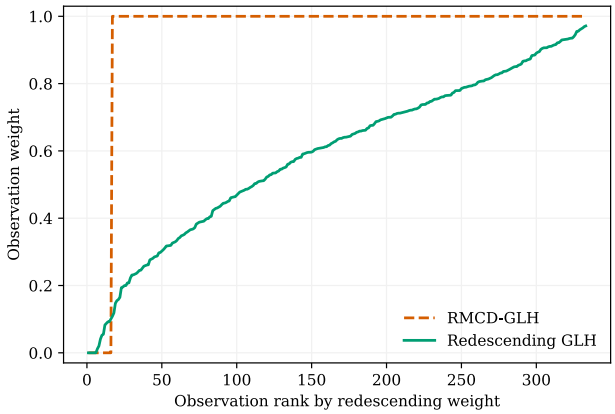

Figure 12. Palmer penguin observation weights ordered by redescending weight.

9. Discussion

The simulations show where the proposed tests are likely to be useful. They perform well when a small number of observations depart from the distribution of their own factorial cell. In such cases, classical Wilks may treat a concentrated group of outliers as a real mean difference, or it may lose power when the outliers oppose a genuine effect. We can lessen both problems by finding a robust centre for each cell before estimating scatter from the combined centred data. In the contamination and masking experiments, the two weighted GLH tests show steadier null rejection rates and greater power to detect masked effects than the classical Wilks test.

Cells with the same number of observations can still have different total weights. The GLH tests the chosen coefficient contrast, whereas the weighted-SSP margin puts more emphasis on cells with greater retained weight. As a result, the two statistics may differ despite using identical hard weights. This is due to how the tests are formed, rather than to the original MCD estimates. The procedures of Spangl [15] and Ameen and Jaaze [17] are based on weighted SSP matrices. Here, we express each factorial effect as a GLH contrast and calibrate it using a known reference law. Thus, we change the test construction, not merely the weights or the calibration of a weighted-SSP statistic.

Normality is not essential for invariance. However, the standardised reference distribution must be known. Given this law, neither changes in the cell means nor applying one nonsingular transformation to every response alters the statistic. Monte Carlo samples can be drawn from a Gaussian distribution, a fully specified multivariate t distribution or another known location–scatter model. If the core distribution is unknown, its shape is still a nuisance parameter. For heavy-tailed errors, the low rejection rates obtained with Gaussian calibration show why the choice of reference model matters.

There are limits to this method as well. Observation weights do not correct for covariance differences between cells, and the heteroscedastic simulation shows why MATS or another test designed for unequal covariance matrices is preferable in this setting [22,23,27]. MCD estimation also needs enough

observations relative to the response dimension. In addition, breakdown resistance is limited: extensive contamination or contamination resembling the majority pattern can still affect the estimates. Finally, case weights should guide scrutiny of influential observations, not serve as an automatic rule for deleting them.

The benchmark comparisons show that robustness has more than one meaning. Ranking limits the effect of unusually large marginal responses, but also changes the multivariate geometry. Parametric-bootstrap MATS controls the heteroscedastic boundary experiment because it is designed for heterogeneous and potentially singular covariance structures rather than localized casewise contamination. Quantile MANOVA changes the target from means to quantiles. The proposed method occupies a narrower niche: mean-based, affine-invariant factorial testing under a common positive-definite location–scatter family when localized multivariate outliers are the primary concern; exact finite-sample calibration further requires a specified standardized core law. The controlled crab perturbation complements the stochastic simulations by showing the bounded response predicted by Theorem 3 on an empirical within-cell geometry.

10. Conclusions

We have presented a robust two-way MANOVA test based on cellwise MCD estimates and hard or redescending weights. Row, column and interaction effects are tested through weighted general linear hypotheses. The coefficient contrasts remain fixed when the weighted cell totals change. The method also allows unequal observed cell sizes. We give the conditions under which the weighted test is defined, without adjusting singular matrices to obtain a result. The statistic is also unchanged when all weights are multiplied by a positive constant.

When the common location–scatter reference distribution is fully specified, affine equivariance removes unknown cell means and scatter. Exchangeability then gives a Monte Carlo test that is valid in finite samples. If the starting MCD estimates remain stable, finite-cutoff weights assign zero weight to sufficiently distant observations. The simulations show that the robust tests lose some power with clean Gaussian data but resist cellwise contamination and masking more effectively. The choice of tuning parameters affects this balance. The MATS comparison also shows that another test is needed if covariance matrices differ across cells. Our method is intended for mean-based factorial analysis with a common positive-definite scatter matrix. Exact Monte Carlo calibration needs a specified reference distribution. It is not intended for heteroscedastic, singular or high-dimensional MANOVA. Unlike earlier weighted-MCD Wilks tests, it retains the same factorial coefficient contrast after weighting.

Authors' Contributions. All authors contributed to the study and to the interpretation of its theoretical and numerical findings. They reviewed the manuscript, approved the final version and accept responsibility for its content.

Availability of Data and Material. The rock-crab data originate from Campbell and Mahon [31]. The Palmer penguin data originate from Gorman et al. [32] and are documented by Horst et al. [33]. Reproducibility materials and the numerical settings required to reproduce the reported experiments are supplied with the submission.

Generative AI Disclosure. During preparation of this manuscript, the authors used ChatGPT (OpenAI) for language editing, manuscript organization, and drafting assistance. The authors reviewed and verified the mathematical statements, numerical claims, references, and final text and take full responsibility for the content of the manuscript.

Conflicts of Interest. The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] S.S. Wilks, Certain Generalizations in the Analysis of Variance, Biometrika 24 (1932), 471–494. https://doi.org/10.1093/biomet/24.3-4.471.

[2] G.E.P. Box, A General Distribution Theory for a Class of Likelihood Criteria, Biometrika 36 (1949), 317–346. https://doi.org/10.1093/biomet/36.3-4.317.

[3] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, 2003. https://zbmath.org/?q=an:1039.62044.

[4] A.C. Rencher, W.F. Christensen, Methods of Multivariate Analysis, Wiley, 2012. https://doi.org/10.1002/9781118391686.

[5] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, Academic Press, 1979. https://zbmath.org/?q=an:0432.62029.

[6] P. Rousseeuw, Multivariate Estimation with High Breakdown Point, in: W. Grossmann, G.C. Pflug, I. Vincze, W. Wertz, (eds) Mathematical Statistics and Applications, Springer Netherlands, Dordrecht, pp. 283–297, (1985). https://doi.org/10.1007/978-94-009-5438-0_20.

[7] P.J. Rousseeuw, B.C. van Zomeren, Unmasking Multivariate Outliers and Leverage Points, J. Am. Stat. Assoc. 85 (1990), 633–639. https://doi.org/10.1080/01621459.1990.10474920.

[8] P.J. Rousseeuw, K. Van Driessen, A Fast Algorithm for the Minimum Covariance Determinant Estimator, Technometrics 41 (1999), 212–223. https://doi.org/10.1080/00401706.1999.10485670.

[9] C. Croux, G. Haesbroeck, Influence Function and Efficiency of the Minimum Covariance Determinant Scatter Matrix Estimator, J. Multivar. Anal. 71 (1999), 161–190. https://doi.org/10.1006/jmva.1999.1839.

[10] G. Pison, S. Van Aelst, G. Willems, Small Sample Corrections for LTS and MCD, Metrika 55 (2002), 111–123. https://doi.org/10.1007/s001840200191.

[11] M. Hubert, P.J. Rousseeuw, S. Van Aelst, High-Breakdown Robust Multivariate Methods, Stat. Sci. 23 (2008), 92–119. https://doi.org/10.1214/088342307000000087.

[12] R.A. Maronna, R.D. Martin, V.J. Yohai, M. Salibián-Barrera, Robust Statistics: Theory and Methods (with R), Wiley, 2018. https://doi.org/10.1002/9781119214656.

[13] V. Todorov, P. Filzmoser, Robust Statistic for the One-Way MANOVA, Comput. Stat. Data Anal. 54 (2010), 37–48. https://doi.org/10.1016/j.csda.2009.08.015.

[14] S. Van Aelst, G. Willems, Robust and Efficient One-Way MANOVA Tests, J. Am. Stat. Assoc. 106 (2011), 706–718. https://doi.org/10.1198/jasa.2011.tm09748.

[15] B. Spangl, Robust Test Statistics for the Two-Way MANOVA Based on the Minimum Covariance Determinant Estimator, arXiv:1806.04106, 2018. https://doi.org/10.48550/arXiv.1806.04106.

[16] A.A. Ameen, O.H. Abbas, Comparison of Some Robust Wilks' Statistics for the One-Way Multivariate Analysis of Variance (MANOVA), J. Al-Qadisiyah Comput. Sci. Math. 11 (2019), 42–58. https://doi.org/10.29304/jqcm.2019.11.2.556.

[17] A.A. Ameen, A.A. Jaaze, Robustness and Comparison of Wilks' Test Statistic for Two-Way MANOVA, J. Al-Qadisiyah Comput. Sci. Math. 12 (2020), 1–23. https://doi.org/10.29304/jqcm.2020.12.1.673.

[18] S.W. Harrar, A.C. Bathke, A Modified Two-Factor Multivariate Analysis of Variance: Asymptotics and Small Sample Approximations, Ann. Inst. Stat. Math. 64 (2012), 135–165. https://doi.org/10.1007/s10463-010-0299-0.

[19] G. Vallejo, M. Ato, Robust Tests for Multivariate Factorial Designs under Heteroscedasticity, Behav. Res. Methods 44 (2012), 471–489. https://doi.org/10.3758/s13428-011-0152-2.

[20] J.-T. Zhang, Two-Way MANOVA with Unequal Cell Sizes and Unequal Cell Covariance Matrices, Technometrics 53 (2011), 426–439. https://doi.org/10.1198/tech.2011.10128.

[21] L.-W. Xu, Parametric Bootstrap Approaches for Two-Way MANOVA with Unequal Cell Sizes and Unequal Cell Covariance Matrices, J. Multivar. Anal. 133 (2015), 291–303. https://doi.org/10.1016/j.jmva.2014.09.015.

[22] F. Konietschke, A.C. Bathke, S.W. Harrar, M. Pauly, Parametric and Nonparametric Bootstrap Methods for General MANOVA, J. Multivar. Anal. 140 (2015), 291–301. https://doi.org/10.1016/j.jmva.2015.05.001.

[23] S. Friedrich, M. Pauly, MATS: Inference for Potentially Singular and Heteroscedastic MANOVA, J. Multivar. Anal. 165 (2018), 166–179. https://doi.org/10.1016/j.jmva.2017.12.008.

[24] A.C. Bathke, S. Friedrich, M. Pauly, F. Konietschke, W. Staffen, et al., Testing Mean Differences among Groups: Multivariate and Repeated Measures Analysis with Minimal Assumptions, Multivar. Behav. Res. 53 (2018), 348–359. https://doi.org/10.1080/00273171.2018.1446320.

[25] A. Gunawardana, F. Konietschke, Nonparametric Multiple Contrast Tests for General Multivariate Factorial Designs, J. Multivar. Anal. 173 (2019), 165–180. https://doi.org/10.1016/j.jmva.2019.02.016.

[26] H. Watanabe, M. Hyodo, S. Nakagawa, Two-Way MANOVA with Unequal Cell Sizes and Unequal Cell Covariance Matrices in High-Dimensional Settings, J. Multivar. Anal. 179 (2020), 104625. https://doi.org/10.1016/j.jmva.2020.104625.

[27] M. Baumeister, M. Ditzhaus, M. Pauly, Quantile-Based MANOVA: A New Tool for Inferring Multivariate Data in Factorial Designs, J. Multivar. Anal. 199 (2024), 105246. https://doi.org/10.1016/j.jmva.2023.105246.

[28] M. Baumeister, K.E. Thiel, L. Matits, G. Zimmermann, M. Pauly, et al., Multivariate and Multiple Contrast Testing in General Covariate-Adjusted Factorial Designs, J. Multivar. Anal. 214 (2026), 105617. https://doi.org/10.1016/j.jmva.2026.105617.

[29] A.C.A. Hope, A Simplified Monte Carlo Significance Test Procedure, J. R. Stat. Soc. Ser. B Stat. Methodol. 30 (1968), 582–598. https://doi.org/10.1111/j.2517-6161.1968.tb00759.x.

[30] B. Phipson, G.K. Smyth, Permutation p-Values Should Never Be Zero: Calculating Exact p-Values When Permutations Are Randomly Drawn, Stat. Appl. Genet. Mol. Biol. 9 (2010), 39. https://doi.org/10.2202/1544-6115.1585.

[31] N.A. Campbell, R.J. Mahon, A Multivariate Study of Variation in Two Species of Rock Crab of the Genus Leptograpsus, Aust. J. Zool. 22 (1974), 417–425. https://doi.org/10.1071/zo9740417.

[32] K.B. Gorman, T.D. Williams, W.R. Fraser, Ecological Sexual Dimorphism and Environmental Variability within a Community of Antarctic Penguins (Genus Pygoscelis), PLoS One 9 (2014), e90081. https://doi.org/10.1371/journal.pone.0090081.

[33] A.M. Horst, A. Presmanes Hill, K.B. Gorman, Palmer Archipelago Penguins Data in the palmerpenguins R Package - An Alternative to Anderson's Irises, R J. 14 (2022), 244–254. https://doi.org/10.32614/rj-2022-020.

[34] A.A. Ameen, O.H. Abbas, On Compare Robust Wilks' Statistics for the One-Way MANOVA: A Simulation Study and Application, J. Stat. Appl. Probab. 11 (2022), 545–563. https://doi.org/10.18576/jsap/110215.

[35] A.A. Suwaidan, A.A. Ameen, On Robust One-Way Analysis of Variance: a Simulation Study and Application, Basrah J. Sci. 43 (2025), 420–448.

[36] R. Hasan, R.R. Al-Mosawi, A.A. Qader, Inspection Times and Optimal Censoring Scheme for Generalized Inverted Exponential Distribution with Progressive Type I Interval Censored Data, Iraqi J. Sci. 65 (2024), 2114–2131. https://doi.org/10.24996/ijs.2024.65.4.29.

[37] R. Hasan, R.R. Al-Mosawi, A.A. Qader, Inference for Generalized Inverted Exponential Distribution Under Progressive Type-I Interval Censored Data, Iraqi J. Sci. 65 (2024), 250–270. https://doi.org/10.24996/ijs.2024.65.1.22.

[38] R.R. Al-Mosawi, R. Hasan, A.A. Qader, Estimation of the Parameters of the Generalized Inverted Exponential Distribution with Progressive Type I Interval Censored Data, Electron. J. Appl. Stat. Anal. 16 (2023), 339–381. https://doi.org/10.1285/i20705948v16n2p339.

[39] A.A. Ameen, U.A. Akkash, Maximum Likelihood Estimation and Bayesian Estimation of Three-Parameter Weibull Distribution Based on Interval-Censored Data, J. Phys.: Conf. Ser. 1818 (2021), 012199. https://doi.org/10.1088/1742-6596/1818/1/012199.

[40] A.R. Khudair, A.A. Ameen, S.L. Khalaf, Mean Square Solutions of Second-Order Random Differential Equations by Using Variational Iteration Method, Appl. Math. Sci. 5 (2011), 2505–2519.

[41] A.R. Khudair, A.A. Ameen, S.L. Khalaf, Mean Square Solutions of Second-Order Random Differential Equations by Using Adomian Decomposition Method, Appl. Math. Sci. 5 (2011), 2521–2535.