

Artículo de investigación

Stests: an R package to perform multivariate statistical tests

Stests: una librería de R para realizar pruebas de hipótesis multivariadas

Jean Paul Piedrahita García¹✉, Freddy Hernández Barajas²✉

¹Estadístico, Universidad Nacional de Colombia sede Medellín, Medellín, Colombia.

²Ingeniero Industrial, PhD en Estadística, Universidad Nacional de Colombia sede Medellín, Medellín, Colombia.

Recepción: 26-jun-2023 **Aceptado:** 15-abril-2024 **Publicado:** 23-jul-2024

Cómo citar: Piedrahita García, J. P., & Hernández Barajas, F. (2024). stests: librería de R para realizar pruebas de hipótesis multivariadas. Ciencia En Desarrollo, 15(2). <https://doi.org/10.19053/uptc.01217488.v15.n2.2024.16145>

Abstract

The Behrens-Fisher problem refers to a statistical challenge in comparing the mean vectors of two normally distributed p -variate populations when the covariance matrices of these populations are assumed to be unequal. This problem can be addressed using Hotelling's T^2 test that requires equality of covariance matrices. However, when the assumption of equality between the two covariance matrices is violated, the performance of this test can be affected, leading to incorrect conclusions. This article presents the implementation of 11 alternative tests proposed in the statistical literature for the Behrens-Fisher problem. These tests are hosted in the stests library in R, and any of these tests can be used through a single function. Additionally, this article conducted a Monte Carlo simulation study in which factors such as sample size, the distance between the mean vectors, and a scaling factor between the covariance matrices were examined. The results found that the rejection rate of the null hypothesis ($H_0 : \mu_1 = \mu_2$) increases when there is a greater discrepancy between the two mean vectors and when the sample size increases. The results demonstrate that all the tests developed in the stests package, which address the multivariate Behrens-Fisher problem, are plausible for comparing two mean vectors.

Keywords: Behrens-Fisher problem, multivariate statistic, statistical test.

Resumen

El problema de Behrens-Fisher se refiere a un desafío estadístico en la comparación de vectores de medias de dos poblaciones distribuidas normal p -variadas cuando se asume que las matrices de covarianzas de estas poblaciones no son iguales. Este problema se puede abordar mediante la prueba T^2 de Hotelling y que exige igualdad de matrices de covarianzas. Sin embargo, cuando el supuesto de igualdad entre las dos matrices de covarianzas se viola, el desempeño de esta prueba puede verse afectada, dando lugar a conclusiones incorrectas. En este artículo se muestra la implementación de 11 pruebas alternativas propuestas en la literatura estadística para el problema de Behrens-Fisher, estas pruebas están alojadas en la librería stests en R y por medio de una sola función se puede usar cualquiera de estas pruebas. Adicionalmente, en este artículo se hizo un estudio de simulación Monte Carlo en el cual se estudiaron factores como el tamaño de muestra, la distancia entre los vectores de medias y un factor escalar entre las matrices de covarianza. Como resultado se encontró que la tasa de rechazos de la hipótesis nula ($H_0 : \mu_1 = \mu_2$) aumenta cuando hay una mayor discrepancia entre los dos vectores de media y cuando el tamaño de muestra aumenta. Los resultados demuestran que todas las pruebas desarrolladas en el paquete stests, que abordan el problema multivariado de Behrens-Fisher, son plausibles para comparar dos vectores de medias.

Palabras Clave: Problema de Behrens-Fisher, estadística multivariada, test estadístico.

1 Introduction

Several methods have been proposed in the statistical literature to test the equality of mean vectors for two p -variate normal populations. This problem, commonly referred to as the Behrens-Fisher problem, is defined by the following set of hypotheses:

$$H_0 : \mu_1 = \mu_2 \quad \text{vs.} \quad H_1 : \mu_1 \neq \mu_2 \quad (1)$$

For example, if we are comparing two species (1 and 2) of penguins according to flipper length and body mass, the vectors μ_1 and μ_2 contain the means of these two measurements for species 1 and 2, respectively.

The Hotelling test is one approach to tackle this problem, assuming equality of the two covariance matrices. However, this test has limitations when the variance and covariance matrices are unknown and unequal. To address these deficiencies, several alternative methods have been developed with the aim of obtaining hypothesis tests that yield more consistent and efficient results.

The earliest solutions to this problem were proposed by Bennett [2] and James [6]. Bennett's solution was based on the univariate solution presented by Scheffé [15], while James extended the univariate Welch series solution. Yao [19] proposed an extension of the Welch approximate degrees of freedom and conducted a simulation study comparing it with James' solution. Johansen [7] presented an invariant solution within the context of general linear models and obtained the exact null distribution of the statistic. Nel and Van Der Merwe [12] also obtained the exact null distribution of the statistic and proposed a non-invariant solution. Kim [10] introduced another extension of the Welch approximate degrees-of-freedom solution, utilizing the geometry of confidence ellipsoids for the mean vectors. Krishnamoorthy and Yu [11] proposed a new invariant test by modifying Nel's solution. Gamage et al. [4] proposed a procedure for testing the equality of mean vectors using generalized p -values, which are functions of sufficient statistics. Yanagihara and Yuan [18] suggested the F test, Bartlett correction test, and modified Bartlett correction test. Kawasaki and Seo [9] proposed an approximate solution to the problem by adjusting the degrees of freedom of the F distribution.

In the context of conducting Hotelling's T^2 test for multivariate data assuming equality of the covariance matrices, several R packages provide useful functions. The ICSNP package offers the `HotellingsT2` function, the `rcov` package includes the `T2.test` function, the `Hotelling` package, with its `hotelling.test` function, and the `Compositional` package offers the `comp.test` function (Hotelling and James). However, those packages do not provide functions to perform the tests proposed by James, [15], [19], [7], [12], [10], [11], [4], [18] and [9] when the covariance matrices are unequal.

This paper is organized as follows: Section 2 provides an overview of the statistical tests implemented in the `stests` R package which is a programming language and software environment specifically designed for statistical computing created by the R Core Team [13]. In Section 3, an introduction to the `stests` R package, which was created to implement these statistical tests, is presented. Sections 4 and 5 detail the design and results of a simulation study conducted to compare the statistical tests. Lastly, Section 6 contains conclusions and discussion.

2 The Behrens-Fisher problem

The univariate version of the Behrens-Fisher problem was proposed initially by Behrens [1] and then reformulated by Fisher [3]. The problem concerns comparing the means of two independent normal populations without assuming the variances are equal. The multivariate version of the Behrens-Fisher problem refers to a statistical hypothesis testing dilemma concerning the comparison of two population means when the variances of the populations are unequal. The interest can be summarized as follows:

Let $X_{i1}, X_{i2}, \dots, X_{in_i}$ a random sample from a p -variate normal population $N(\mu_i, \Sigma_i)$ with $i = 1, 2$. The Behrens-Fisher problem has the next set of hypotheses:

$$H_0 : \mu_1 = \mu_2 \quad \text{vs.} \quad H_1 : \mu_1 \neq \mu_2 \quad (2)$$

The tests proposed in the literature require specific inputs, namely the sample mean vectors and sample covariance matrices, to perform the necessary calculations. The expressions to obtain these inputs are as follows:

$$\bar{X}_i = \sum_{j=1}^{n_i} X_{ij} / n_i \quad \text{with} \quad i = 1, 2,$$

and the sample variance-covariance matrices

$$S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)^\top \quad \text{with} \quad i = 1, 2.$$

Next, the statistical tests implemented in the `stests` R package to study the hypothesis given in expression (2) are listed.

- Hotelling test.
- First order James' test (1954).
- Yao's test (1965).
- Johansen's test (1980).
- NVM test (1986).
- Modified NVM test (2004).
- Gamage's test (2004).
- Yanagihara and Yuan's test (2005).
- Bartlett Correction test (2005).
- Modified Bartlett Correction test (2005).
- Second Order Procedure (S procedure) test (2015).
- Bias Correction Procedure (BC Procedure) test (2015).

All of the tests listed above have expressions for the statistics and their distributions under the null hypothesis. However, due to the length and complexity of these expressions, it is recommended that readers refer to the vignette available in [5] for detailed information. The vignette provides comprehensive explanations, equations, and examples related to the statistics and their distribution, allowing readers to delve into the specifics of each test.

3 The `stests` package

In this section, we introduce the `stests` R package, which has been developed to facilitate the implementation of multivariate statistical tests for the Behrens-Fisher problem. The package is currently available on the GitHub platform, allowing users to easily download and utilize it. To install and use the package, users can use the following code snippet:

```
install.packages("devtools")
library(devtools)
install_github(repo='fhernanb/stests',
               force=TRUE)
library(stests)
```

The main function of the `stests` package is `two_mean_vector_test()`. This function has the next structure:

```
two_mean_vector_test(xbar1, s1, n1,
                    xbar2, s2, n2,
                    delta0 = NULL,
                    method = "T2",
                    alpha = 0.05)
```

The arguments `xbar1`, `s1`, and `n1` correspond to the sample mean vector, sample covariance matrix, and size of the sample 1. In the same way, `xbar2`, `s2`, and `n2` are the information for the sample 2. The argument `delta0` is a p -dimensional numeric vector to test $H_0 : \mu_1 - \mu_2 = \delta_0$, by default, $\delta_0 = 0$.

The `method` argument can be used to indicate the name of the hypothesis test, the current values are "T2" (default), "james" (James' first order test), "yao" (Yao's test), "johansen" (Johansen's test), "nvm" (Nel and Van der Merwe test), "mnm" (modified Nel and Van der Merwe test), "gamage" (Gamage's test), "yy" (Yanagihara and Yuan test), "byy" (Bartlett Correction test), "mbyy" (modified Bartlett Correction test), "ks1" (Second Order Procedure), "ks2" (Bias Correction Procedure). Finally, the `alpha` argument is used only to obtain the critic value for the James test.

The `two_mean_vector_test()` function generates an object of class `htest`. For this type of class, we create the general S3 functions `print()` and `plot()`. In the examples below, we illustrate the utility of these functions.

3.1 Example 1

Here we revisit example 3.7 from Seber [16] on page 116, using the James first-order test to explore the null hypothesis $H_0 : \mu_1 = \mu_2$, with a significance level of $\alpha = 0.05$. We can use the function `two_mean_vector_test()` with the argument `method="james"`. The resulting output is stored in the object `a`.

```
n1 <- 16
xbar1 <- c(9.82, 15.06)
s1 <- matrix(c(120, -16.3,
              -16.3, 17.8), ncol = 2)

n2 <- 11
xbar2 <- c(13.05, 22.57)
s2 <- matrix(c(81.8, 32.1,
              32.1, 53.8), ncol = 2)

a <- two_mean_vector_test(xbar1=xbar1,
                        s1=s1,
                        n1=n1,
                        xbar2=xbar2,
                        s2=s2,
                        n2=n2,
                        method='james',
                        alpha=0.05)
```

We can write in the R console the object name `a` to obtain the next summary for the test.

```
James test for two mean vectors
```

```
data: this test uses summarized data
T2=9.4455, critical_value=7.2308, df=2
alt hyp: mu1 is not equal to mu2
```

```
sample estimates:
Sample 1 Sample 2
xbar_1      9.82  13.05
xbar_2     15.06  22.57
```

From the summary, we note that the critical value is 7.2308, and the test statistic 9.4455. For this reason there is evidence to reject $H_0 : \mu_1 = \mu_2$.

3.2 Example 2

In this second example we used the data from Yao [19] on page 141. The objective is to test $H_0 : \mu_1 = \mu_2$ versus $H_0 : \mu_1 \neq \mu_2$ using the Yao test. We can use the function `two_mean_vector_test()` with the argument `method="yao"`. The resulting output is stored in the object `b`.

```
n1 <- 16
xbar1 <- c(9.82, 15.06)
s1 <- matrix(c(120, -16.3,
              -16.3, 17.8), ncol=2)

n2 <- 11
xbar2 <- c(13.05, 22.57)
s2 <- matrix(c(81.8, 32.1,
              32.1, 53.8), ncol=2)

b <- two_mean_vector_test(xbar1=xbar1,
                        s1=s1,
                        n1=n1,
                        xbar2=xbar2,
                        s2=s2,
                        n2=n2,
                        method="yao")
```

We can write in the R console the object name `b` to obtain the next summary for the test.

```
Yao test for two mean vectors
```

```
data: this test uses summarized data
T2 = 9.4455, F = 4.3855,
df1 = 2.000, df2 = 13.001, p-value = 0.03503
alter hyp: mu1 is not equal to mu2
```

```
sample estimates:
Sample 1 Sample 2
xbar_1      9.82  13.05
xbar_2     15.06  22.57
```

From the summary, we observe that the p -value is 0.03503, which is lower than the conventional significance level of 5%. Therefore, there is evidence to reject the null hypothesis $H_0 : \mu_1 = \mu_2$.

4 Simulation study

A Monte Carlo simulation was conducted to evaluate the performance of the function developed for the Behrens-Fisher tests. In the simulation study we considered four factors:

- p : the dimension of the two normal populations, three cases were considered, $p = 2$, $p = 3$ and $p = 5$.
- n : the same sample size for both populations was equal ($n_1 = n_2 = n$) with values $n = 10, 20, \dots, 150$.
- δ : distance between the components of centroids μ_1 and μ_2 . For $p = 2$, the values for δ were 0, 10, 20, \dots , 150. For $p = 3$, the values for δ were 0, 10, 20, \dots , 150. For $p = 5$, the values of δ were 0, 0.5, 1.0, 1.5, \dots , 7.0, 7.5.
- k : a scalar factor to relate the variances and covariances matrices, that is, $\Sigma_2 = k \cdot \Sigma_1$. The values considered were

$k = 1, 2, 5$, where $k = 1$ means that both variances and covariances matrices are equal.

The combination of the aforementioned factors produces 1440 scenarios. For each scenario, 1000 samples of size n were generated from both p -variate normal populations using the `mvrnorm` function of the MASS package of [17].

When $p = 2$, the baseline mean vector and variance-covariance matrix were taken from example 4.1 by Nel and Van Der Merwe [12], located on page 3729. The baseline mean vector for population 1 was held fixed at $\mu_1 = (204.4, 556.6)$ while μ_2 was obtained by adding δ to each component of μ_1 . The baseline variance-covariance matrix Σ_1 was

$$\Sigma_1 = \begin{pmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{pmatrix}$$

When $p = 3$, the baseline mean vector and variance-covariance matrix were taken from exercise 6.18 by Johnson and Wichern [8], located on page 344. The baseline mean vector for population 1 was held fixed at $\mu_1 = (113.4, 88.3, 40.7)$ while μ_2 was obtained by adding δ to each component of μ_1 . The baseline variance-covariance matrix Σ_1 was

$$\Sigma_1 = \begin{pmatrix} 138.8 & 79.1 & 37.4 \\ 79.1 & 50.0 & 21.7 \\ 37.4 & 21.7 & 11.3 \end{pmatrix}$$

When $p = 5$, the baseline mean vector and variance-covariance matrix were taken from example 3.10.1 by Rencher and Christensen [14], located on page 79. The baseline mean vector for population 1 was held fixed at $\mu_1 = (36.09, 25.55, 34.09, 27.27, 30.73)$ while μ_2 was obtained by adding δ to each component of μ_1 . The baseline variance-covariance matrix Σ_1 was

$$\Sigma_1 = \begin{pmatrix} 65.09 & 33.65 & 47.59 & 36.77 & 25.43 \\ 33.65 & 46.07 & 28.95 & 40.34 & 28.36 \\ 47.59 & 28.95 & 60.69 & 37.37 & 41.13 \\ 36.77 & 40.34 & 37.37 & 62.82 & 31.68 \\ 25.43 & 28.36 & 41.13 & 31.68 & 58.22 \end{pmatrix}$$

The variance-covariance matrix Σ_2 was obtained as $\Sigma_2 = k \cdot \Sigma_1$. If a valid variance-covariance matrix Σ is multiplied by a positive scaling factor k , the resulting matrix retains the property of being positive definite.

The performance of each test in the simulation study was measured by the rejection rate of the null hypothesis, which is the probability of rejecting the null hypothesis when it is actually true. It is also known as the Type I error rate or significance level. The rejection rate is typically denoted by the Greek letter α (alpha).

To calculate the rejection rate, you need to specify a critical value or a range of critical values based on the desired significance level. The critical value(s) define the boundary between the region where the null hypothesis is rejected and the region where it is not rejected. The rejection rate can be expressed as:

$$\alpha = P(\text{Reject } H_0 | H_0 \text{ is true})$$

For the James test [6], we did not calculate the p -value because there is no distribution for the statistic, so the critic value was used to calculate the proportion of rejections of H_0 . In the simulation study, the significance level was constant at $\alpha = 0.05$.

5 Results

The rejection rates of the null hypothesis $H_0 : \mu_1 = \mu_2$ for scenarios with $p = 2$ are illustrated in Figures 1 to 3. When $\delta = 0$ the null hypothesis is true, the rejection rates of H_0 are approximately 0.05. Furthermore, as the value of δ increases, indicating that the null hypothesis is false, the rejection rates progressively approach 1.00. Therefore the rejection rates of the null hypothesis increase as the discrepancy between the mean vectors μ_1 and μ_2 becomes larger.

Upon observing Figures 1 to 3, it is evident that as the scaling factor k increases, the rejection rate of the null hypothesis decreases. This suggests that a larger scaling factor leads to a more conservative test, resulting in a lower probability of rejecting the null hypothesis.

Furthermore, it can be observed that as the sample size n increases, the rejection rate also increases. This implies that with larger sample sizes, the statistical test becomes more powerful in detecting differences between the mean vectors, leading to a higher probability of rejecting the null hypothesis when it is indeed false.

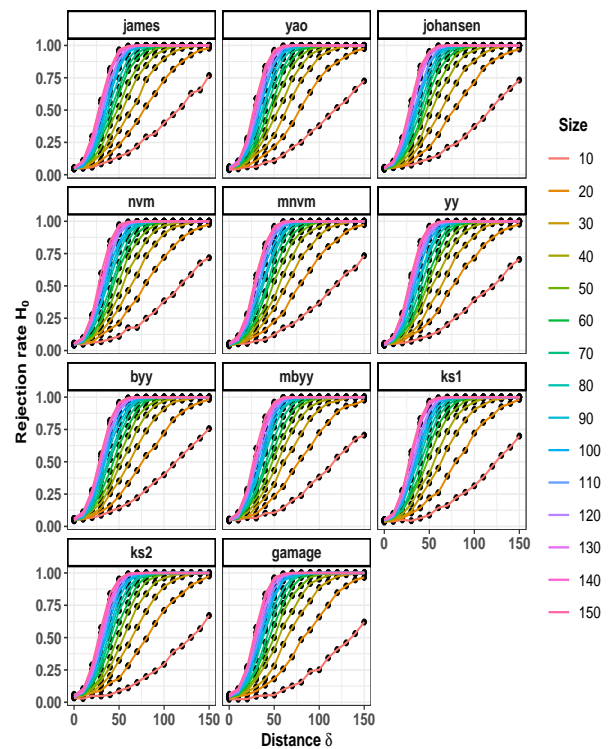


Figure 1: Rejection rate of $H_0 : \mu_1 = \mu_2$ as a function of δ and sample n when $p = 2$.

Figures 4 to 6 depict the trend of the rejection rates of the null hypothesis $H_0 : \mu_1 = \mu_2$ for scenarios with $p = 3$. The pattern observed is the same for $p = 2$. As the values of δ and n increase, the rejection rates also increase.

Figures 7 to 9 illustrate the progression of the rejection rates of the null hypothesis $H_0 : \mu_1 = \mu_2$ for scenarios with $p = 5$. Similar patterns can be observed as in $p = 2$ and $p = 3$. As the values of δ and n increase, the rejection rates also increase. This indicates that as the discrepancy between the mean vectors and the sample size grow, the statistical tests become more powerful in detecting differences, resulting in higher rejection rates of the null hypothesis when it is indeed false.

On the other hand, as the scaling factor k increases, the rejection rate decreases. This implies that a larger scaling factor leads to a more conservative test, with a lower probability of rejecting the null

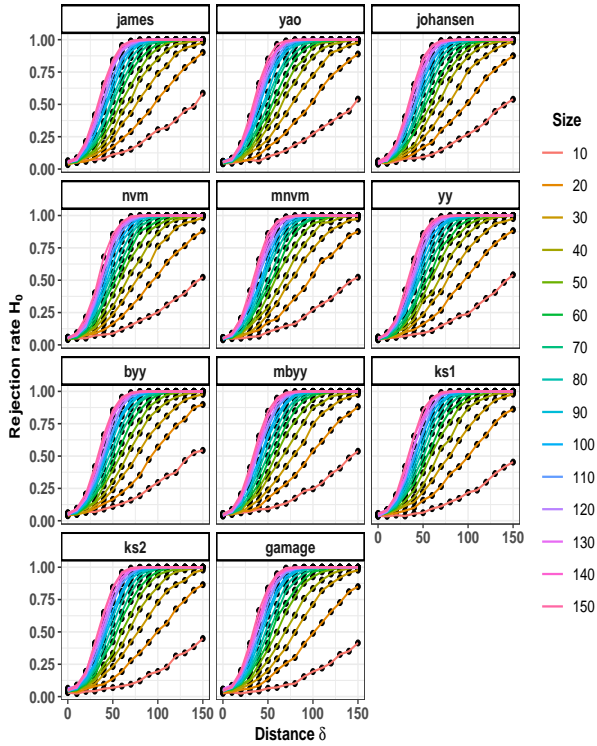


Figure 2: Rejection rate of $H_0 : \mu_1 = \mu_2$ as a function of δ and sample n when $p = 2$ and $k = 2$.

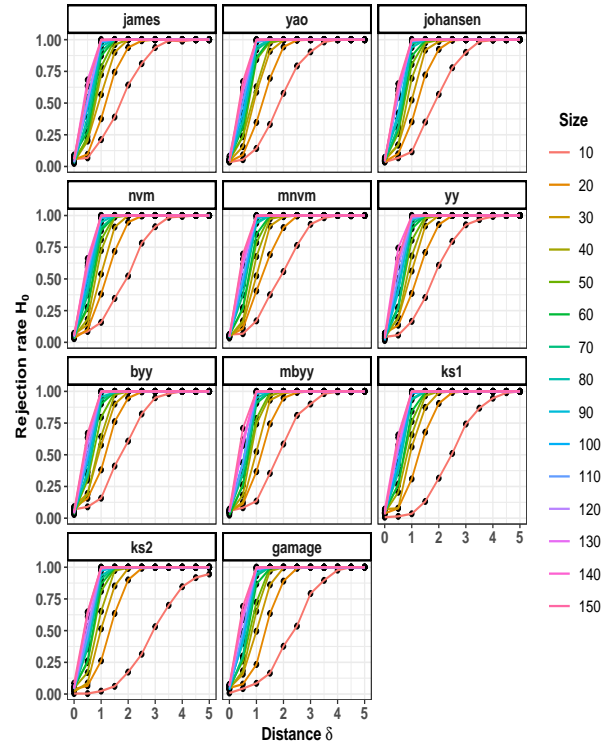


Figure 4: Rejection rate of $H_0 : \mu_1 = \mu_2$ as a function of δ and sample n when $p = 3$.

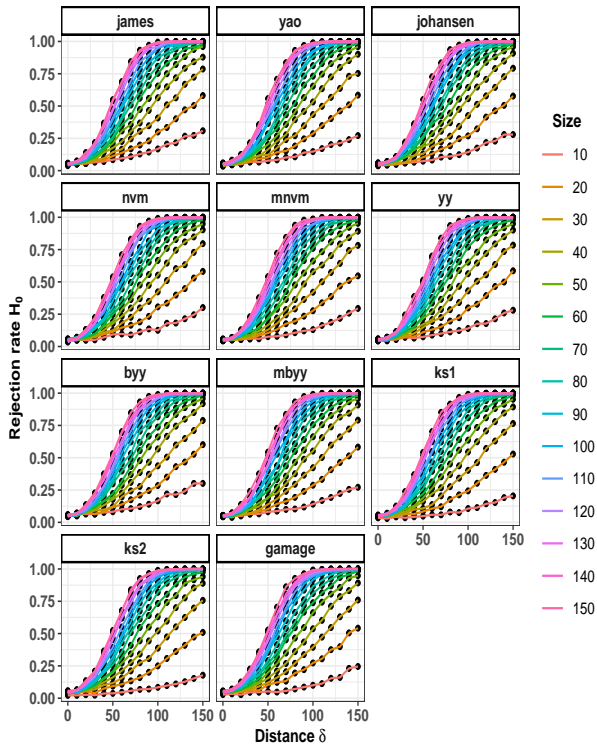


Figure 3: Rejection rate of $H_0 : \mu_1 = \mu_2$ as a function of δ and sample n when $p = 2$ and $k = 5$.

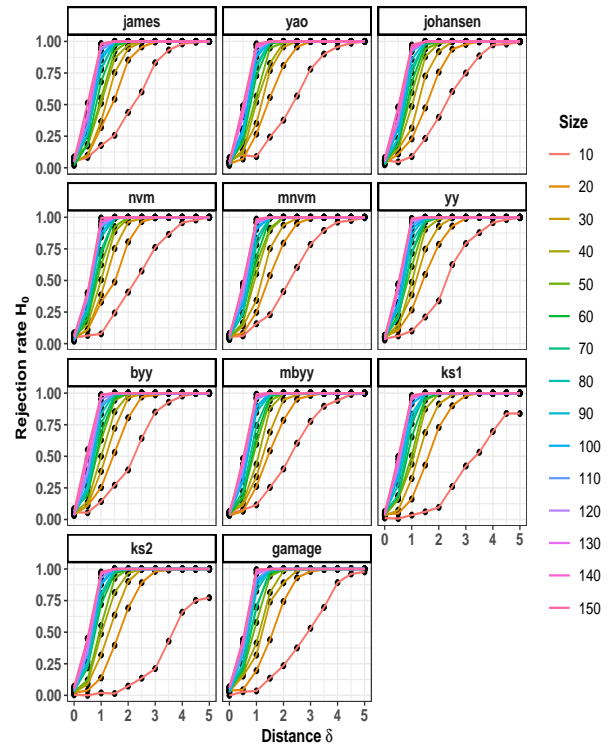


Figure 5: Rejection rate of $H_0 : \mu_1 = \mu_2$ as a function of δ and sample n when $p = 3$ and $k = 2$.

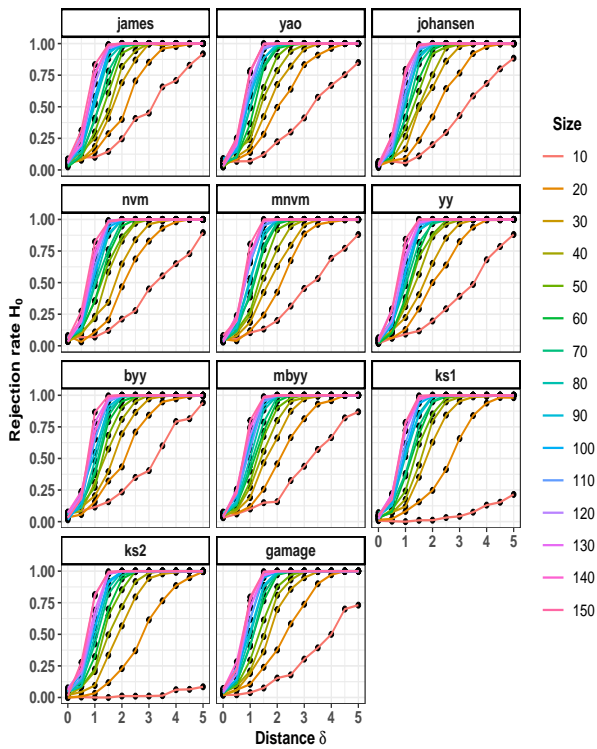


Figure 6: Rejection rate of $H_0 : \mu_1 = \mu_2$ as a function of δ and sample n when $p = 3$ and $k = 5$.

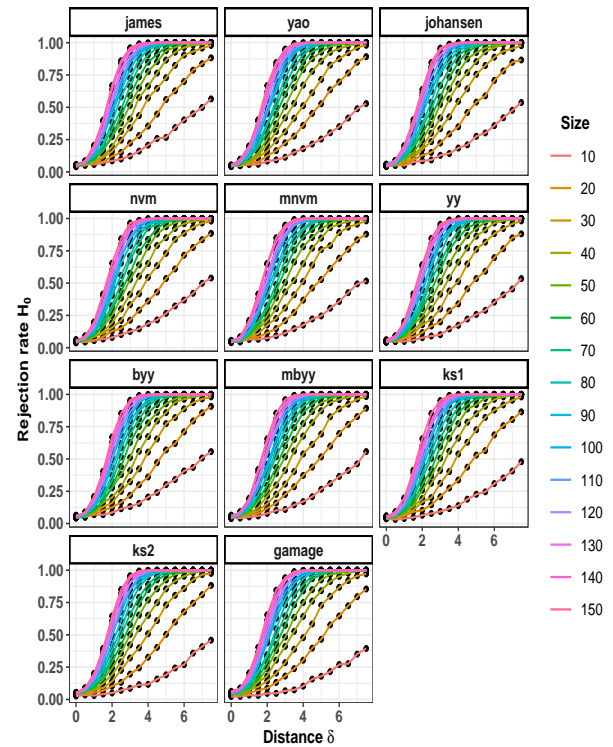


Figure 7: Rejection rate of $H_0 : \mu_1 = \mu_2$ as a function of δ and sample n when $p = 5$.

hypothesis. The consistent patterns observed in both $p = 2$ and $p = 5$ scenarios suggest a general trend across different dimensions.

Figure 10 illustrates the behavior of the rejection rate for the hypothesis $H_0 : \mu_1 = \mu_2$ as a function of sample size n and δ for Yanagihara and Yuan's test when $p = 3$ and $k = 2$. In this figure, we observe the general pattern that as the sample size n or δ increases, the rejection rate also increases. This observed pattern is consistent for each test, regardless of the dimension p or the value of k .

6 Conclusion and Discussion

In conclusion, the simulation study has provided valuable insights into the performance of the R functions developed for addressing the Behrens-Fisher problem. The analysis of various factors and their impact on the rejection rates of the null hypothesis has yielded important findings.

The results indicate that as the sample size n increases, the rejection rate of H_0 increases. This suggests that larger sample sizes enhance the power of the statistical test in detecting differences between the mean vectors. Consequently, the probability of correctly rejecting the null hypothesis when it is false is higher, reflecting the increased sensitivity of the tests.

Another interesting finding is that as the distance δ increases, the rejection rate also increases. This is an expected result because as the distance between mean vectors increases, we anticipate a rise in rejection rates. This observation confirms that the implemented tests are functioning correctly.

We observed that as the scaling factor k increases, the rejection rate of the null hypothesis decreases. This implies that a larger scaling factor leads to a more conservative test, reducing the likelihood of erroneously rejecting the null hypothesis. However, as the sample size n increases, the rejection rate of H_0 increases regardless of the value of k .

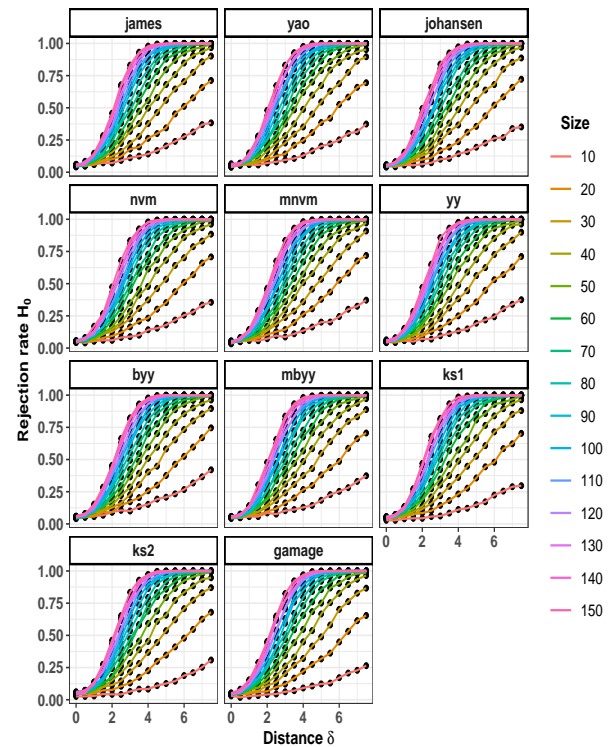


Figure 8: Rejection rate of $H_0 : \mu_1 = \mu_2$ as a function of δ and sample n when $p = 5$ and $k = 2$.

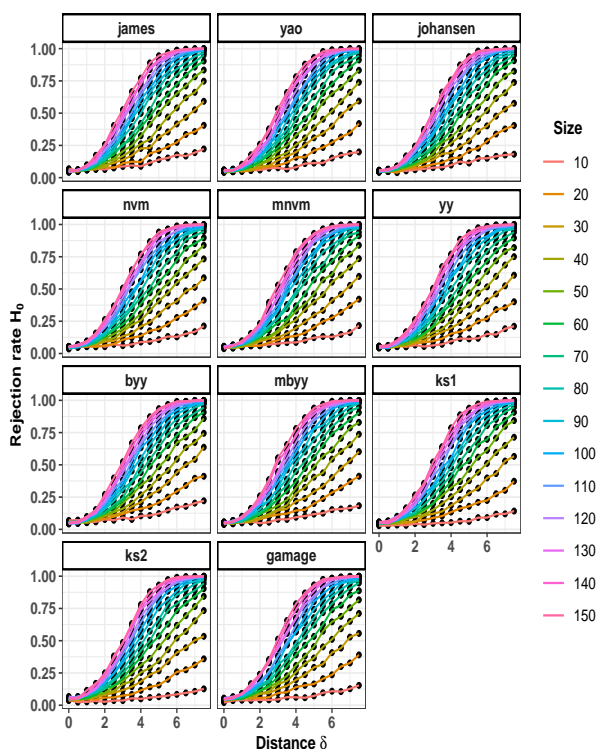


Figure 9: Rejection rate of $H_0 : \mu_1 = \mu_2$ as a function of δ and sample n when $p = 5$ and $k = 5$.

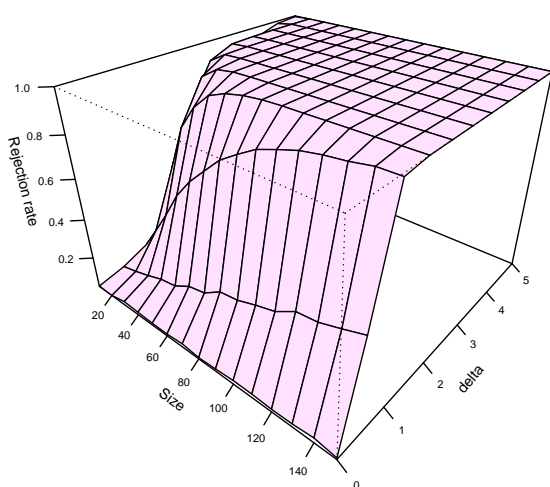


Figure 10: Rejection rate of $H_0 : \mu_1 = \mu_2$ as a function of sample n and δ for Yanagihara and Yuan's test when $p = 3$ and $k = 2$.

We observed that the trajectory lines for rejection rates of H_0 tend to decrease for larger values of k and small sample sizes, but these trajectories approach 100% as the sample size increases.

The developed R functions offer a user-friendly and accessible tool for conducting Behrens-Fisher tests. The simplicity and efficiency of these functions make them highly convenient for researchers and practitioners in the field. Users can confidently rely on the accuracy and reliability of these functions, as demonstrated by the comprehensive evaluation and validation carried out in this study.

In summary, the R functions developed for the Behrens-Fisher problem have exhibited robust performance, providing reliable and powerful tools for hypothesis testing. Researchers and practitioners can readily employ these functions to analyze their data, confident in their ability to accurately assess differences between mean vectors. The accessibility and user-friendliness of these functions further enhance their usability and contribute to the advancement of statistical analysis in multivariate data.

References

- [1] W. Behrens. A contribution to error estimation with few observations. *Journal of Agriculture Scientific Archives of the Royal Prussian State College-Economy*, 68:807–837, 1929.
- [2] B. M. Bennett. Note on a solution of the generalized behrens–fisher problem. *Ann. Inst. Statist. Math.*, 2:87–90, 1951.
- [3] R. A. Fisher. The comparison of samples with possibly unequal variances. *Annals of Eugenics*, 9(2):174–180, 1939.
- [4] J. Gamage, T. Mathew, and S. Weerahandi. Generalized p-values and generalized confidence regions for the multivariate behrens–fisher problem and manova. *Journal of Multivariate Analysis*, 88:177–189, 2004.
- [5] F. Hernández-Barajas and J. P. Piedrahita. The behrens-fisher problem. <https://fhernanb.github.io/stests/articles/Behrens-Fisher.html>, 2024. Accessed: 2024-05-24.
- [6] G. James. Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika*, 41:19–43, 1954.
- [7] S. Johansen. The welch-james approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika*, 67(1):85–92, 1980.
- [8] R. Johnson and D. Wichern. *Applied multivariate statistical analysis*. Pearson Prentice Hall, 2007.
- [9] T. Kawasaki and T. Seo. A two sample test for mean vectors with unequal covariance matrices. *Communications in Statistics - Simulation and Computations*, 44:1850–1866, 2015.
- [10] S.-J. Kim. A practical solution to the multivariate behrens-fisher problem. *Biometrika*, 79(1):171–176, 1992. ISSN 00063444. URL <http://www.jstor.org/stable/2337157>.
- [11] K. Krishnamoorthy and J. Yu. Modified nel and van der merwe test for the multivariate behrens–fisher problem. *Statistics & Probability Letters*, 66(2):161–169, 2004.
- [12] D. G. Nel and C. A. Van Der Merwe. A solution to the multivariate behrens-fisher problem. *Communications in Statistics - Theory and Methods*, 15(12):3719–3735, 1986.
- [13] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024. URL <https://www.R-project.org/>.
- [14] A. Rencher and W. Christensen. *Methods of multivariate analysis*. John Wiley & Sons, 2012.

- [15] H. Scheffé. On solutions of the behrens-fisher problem, based on the t-distribution. *The Annals of Mathematical Statistics*, 14(1): 35–44, 1943. ISSN 00034851. URL <http://www.jstor.org/stable/2236000>.
- [16] G. A. Seber. *Multivariate observations*. John Wiley & Sons, 2009.
- [17] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. URL <https://www.stats.ox.ac.uk/pub/MASS4/>. ISBN 0-387-95457-0.
- [18] H. Yanagihara and K. Yuan. Three approximate solutions to the multivariate behrens-fisher problem. *Communications in Statistics - Simulation and Computations*, 34:975–988, 2005.
- [19] Y. Yao. An approximate degrees of freedom solution to the multivariate behrens-fisher problem. *Biometrika*, 52(1/2):139–147, 1965.