

stests: paquete de R con funciones útiles para pruebas de hipótesis multivariadas^{*}

Valentina García Velásquez

Universidad Nacional de Colombia - sede Medellín

Jean Paul Piedrahita García

Universidad Nacional de Colombia - sede Medellín

Freddy Hernández Barajas

Universidad Nacional de Colombia - sede Medellín

Resumen: Actualmente R cuenta con pocas funciones para realizar pruebas de hipótesis para el vector de medias o la matriz de varianzas y covarianzas. En vista de esta situación, el paquete **stests** fue creado para implementar las diversas pruebas multivariadas en los casos de una o dos poblaciones normales que se ha reportado en la literatura estadística. Las funciones de **stests** fueron creadas de tal manera que exista una estandarización de la información, con el propósito de facilitar su uso por parte del usuario. En la ayuda de las funciones del paquete se tienen ejemplos presentados en artículos científicos y en libros, y, de esta manera, el usuario puede emplear con tranquilidad las funciones de **stests** ya que los resultados obtenidos coinciden con los presentados en la literatura.

Keywords: prueba de hipótesis, análisis multivariado.

Introducción

En la literatura estadística se pueden encontrar varias propuestas de pruebas de hipótesis para comparar vectores de medias y para comparar matrices de varianzas y covarianzas para el caso de poblaciones normales. Solo un número reducido de estas pruebas de hipótesis están disponibles en algunos paquetes de R, y desafortunadamente no hay una estandarización en la forma de ingresar la información a las funciones. Por esta razón, contar con un paquete especializado en pruebas estadísticas multivariadas sería de mucha utilidad para estudiantes, profesores y analistas.

El paquete **stests** reúne, recopila y organiza las diferentes propuestas multivariadas para realizar pruebas de hipótesis para el vector de medias y la matriz de varianzas y covarianzas. Dichas pruebas consideran tanto una como dos poblaciones normales.

Como se mencionó anteriormente, la mayoría de las funciones existentes no consideran usar datos resumidos, por lo que no serían de mucha utilidad en la comunidad académica. El paquete **stests** tiene como objetivo apoyar el trabajo de docentes y estudiantes para aplicar pruebas multivariadas.

Paquete

El paquete **stests** fue hecho en el lenguaje de programación [R Core Team \(2019\)](#) y contiene funciones útiles para aplicar pruebas estadísticas, en las cuales se destacan las propuestas multi-

^{*} Autor de contacto: valentia@unal.edu.co.

variadas para el vector de medias y la matriz de varianzas y covarianzas.

Los usuarios pueden visitar la [página web del paquete](#) para explorar tanto las viñetas (artículos) como las funciones de referencia, tanto para el caso univariado como multivariado, entre otros detalles acerca de cada función.

El paquete se encuentra alojado en GitHub y para instalar la última versión se debe utilizar el código que se muestra a continuación:

```
if (!require("devtools")) install.packages("devtools")
devtools::install_github("fhernanb/stests", force=TRUE)
```

En la siguiente figura se muestra el hexalogo que caracteriza el paquete **stests**.



Figura 1: Hexalogo del paquete **stests**

Funciones para pruebas de hipótesis multivariadas

El paquete **stests** también contiene pruebas para el caso univariado, las cuales presentan herramientas importantes para complementar las ya existentes en el paquete **stats**. En esta oportunidad se presentarán las funciones creadas para realizar pruebas de hipótesis multivariadas. Para comprender de una manera adecuada el uso de estas funciones, se expone lo siguiente:

Caso de una población normal

Para μ

La función **one_mean_vector_test** puede ser usada para probar el siguiente conjunto de hipótesis:

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$$

Esta función considera los casos donde la matriz de varianzas y covarianzas Σ es conocida o desconocida. La forma de utilizar esta función se muestra a continuación:

```
one_mean_vector_test(mu0, xbar, n, S = NULL, Sigma = NULL)
```

Para Σ

La función **one_covar_matrix_test** se usa para probar el siguiente conjunto de hipótesis:

$$H_0 : \Sigma = \Sigma_0 \quad vs \quad H_1 : \Sigma \neq \Sigma_0$$

La forma de utilizar esta función se muestra a continuación:

```
one_covar_matrix_test(Sigma0, S, n, method = c("lrt", "modlrt1", "modlrt2"))
```

Cuando se emplea `method = "lrt"` (método por defecto), la función aplica el test LRT presentado en [Mardia, Kent and Bibby \(1979\)](#), página 126, expresión 5.2.7. Cuando se emplea `method = "modlrt1"` o `method = "modlrt2"`, la función desarrolla el test LRT presentado en [Rencher and Christensen \(2012\)](#), página 260, expresiones 7.2 and 7.4.

Caso de dos poblaciones normales

Para μ_1 y μ_2

La función **two_mean_vector_test** implementa la prueba de hipótesis para dos vectores de medias que provienen de dos poblaciones normales p variadas.

$$H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 \neq \mu_2$$

La forma de utilizar esta función se muestra a continuación:

```
two_mean_vector_test(xbar1, Sigma1, n1,
                     xbar2, Sigma2, n2,
                     delta0 = NULL, alpha = 0.05,
                     method = c("T2", "james", "mvn"))
```

Cuando se emplea `method = "T2"` (método por defecto), la función desarrolla el test de Hotelling, en el cual se asume igualdad entre las dos matrices de varianzas y covarianzas. Cuando se emplea `method = "james"` o `method = "mvn"`, la función desarrolla el problema Behrens-Fisher, en el cual no se asume igualdad entre las dos matrices de varianzas y covarianzas, utilizando, respectivamente, la propuesta de [James \(1954\)](#) o de [Krishnamoorthy and Yu \(2004\)](#).

Para Σ_1 y Σ_2

La función **cov.Mtest** implementa la prueba de hipótesis desarrollada por Box para varias matrices de varianzas y covarianzas, también conocida como Prueba M de Box. En este caso, la función se empleará para probar la igualdad entre dos matrices de la siguiente manera:

$$H_0 : \Sigma_1 = \Sigma_2 \quad vs \quad H_1 : \Sigma_1 \neq \Sigma_2$$

La forma de utilizar esta función se muestra a continuación:

```
cov.Mtest(x, group, alpha = 0.05)
```

El objeto `x` es la matriz numérica de datos y `group` el vector categórico de datos que clasifica a cada observación.

Ejemplos

En esta sección se presentan algunos ejemplos para entender el uso del paquete. Luego de instalar el paquete es necesario cargarlo por medio de la siguiente instrucción:

```
library(stests)
```

Para μ_1 y μ_2

Este ejemplo corresponde al ejemplo 4.1 de Nel and Van de Merwe (1986), página 3729. A continuación se muestra la forma de ingresar los datos para probar $H_0 : \mu_1 = \mu_2$ versus $H_0 : \mu_1 \neq \mu_2$

```
n1 <- 45
xbar1 <- c(204.4, 556.6)
s1 <- matrix(c(13825.3, 23823.4, 23823.4, 73107.4), ncol = 2)

n2 <- 55
xbar2 <- c(130.0, 355.0)
s2 <- matrix(c(8632.0, 19616.7, 19616.7, 55964.5), ncol = 2)

res1 <- two_mean_vector_test(xbar1 = xbar1, Sigma1 = s1, n1 = n1,
                             xbar2 = xbar2, Sigma2 = s2, n2 = n2,
                             method = "mvn")

print(res1)
```

```
##
## Modified Nel and Van der Merwe test for two mean vectors
##
## data: this test uses summarized data
## T2 = 15.6585, F = 7.7284, df1 = 2.000, df2 = 76.592, p-value =
## 0.0008763
## alternative hypothesis: mu1 is not equal to mu2
##
## sample estimates:
##      Sample 1 Sample 2
## xbar_1      204.4      130
## xbar_2      556.6      355
```

El resultado de la prueba está almacenado en el objeto `res1` y la función genérica `S3` puede usarse para obtener un gráfico que ilustra el valor-P de la prueba. La figura 2 presenta el gráfico correspondiente.

Para Σ

Este ejemplo corresponde al ejemplo 5.3.2 de [Mardia, Kent and Bibby \(1979\)](#), página 127. A continuación se muestra la forma de ingresar los datos para probar $H_0 : \Sigma = \Sigma_0$ versus $H_1 : \Sigma \neq \Sigma_0$.

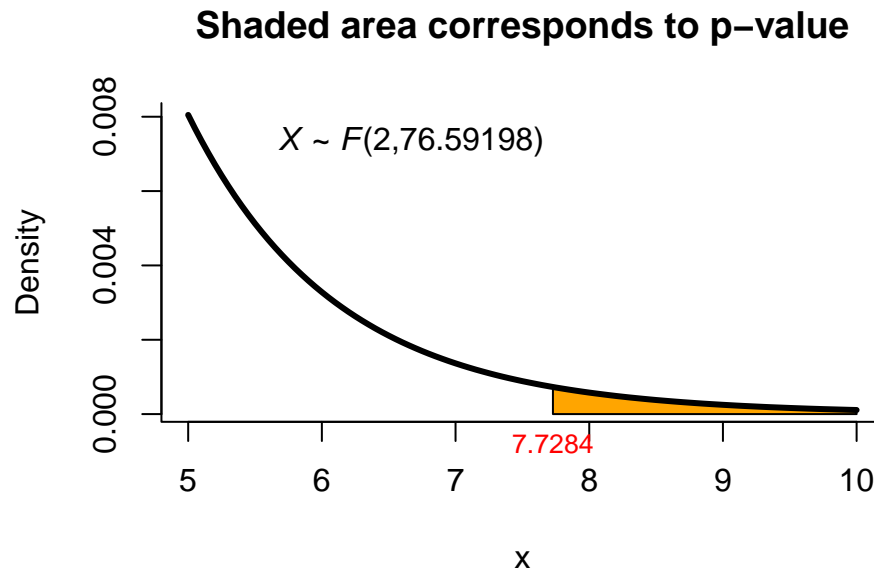


Figura 2: Ilustración del valor-P para el ejemplo 1.

```
Sigma0 <- matrix(c(100, 0, 0, 100), ncol = 2)
S <- matrix(c(91.481, 66.875, 66.875, 96.775), ncol = 2)

res2 <- one_covar_matrix_test(Sigma0 = Sigma0, S = S, n = 25, method = "lrt")

print(res2)

##
## LRT test for Sigma matrix
##
## data: this test uses summarized data
## lrt = 17.698, df = 3, p-value = 0.0005077
## alternative hypothesis: true Sigma matrix is not equal to Sigma0
##
## sample estimates:
##      xbar_1 xbar_2
## xbar_1 91.481 66.875
## xbar_2 66.875 96.775
```

El resultado de la prueba está almacenado en el objeto `res2` y la función genérica `S3` puede usarse para obtener un gráfico que ilustra el valor-P de la prueba, así como el obtenido en el **ejemplo 1**.

Referencias

- James, G. S. 1954. "Tests of Linear Hypotheses in Univariate and Multivariate Analysis when the Ratios of the Population Variances are Unknown.". ISBN 978-1498716963.
URL: <http://www.jstor.org/stable/2333003?origin=JSTOR-pdf>
- Krishnamoorthy, K. and Jianqi Yu. 2004. "Modified Nel and Van der Merwe test for the multivariate Behrens–Fisher problem.".
URL: <https://doi.org/10.1016/j.spl.2003.10.012>
- Mardia, K. V., J. T. Kent and J. M. Bibby. 1979. *Multivariate Analysis*. 10nd ed. London, United Kingdom: Academic Press, INC. ISBN 0-12-471250-9 0-12-471252-5.
- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
URL: <https://www.R-project.org/>
- Rencher, A. C. and W. F. Christensen. 2012. *Methods of Multivariate Analysis*. 3nd ed. Hoboken, New Jersey: John Wiley and Sons, INC. ISBN 978-0-470-17896-6.