



Una prueba de independencia completa basada en la FDR

A test for complete Independence based on FDR

Jorge Iván Vélez^a
jorgeivanvelez@gmail.com

Juan Carlos Correa^b
jccorrea@unal.edu.co

Resumen

El análisis e interpretación de datos multivariados se facilita enormemente si las variables son independientes. En la práctica, este supuesto se verifica a través de una prueba de independencia completa. Proponemos una nueva prueba de independencia completa basada en la tasa de falsos descubrimientos (FDR, en inglés), y reportamos los resultados de un estudio de simulación en el que se comparan los niveles de significancia real de esta propuesta y otras pruebas comúnmente utilizadas. Encontramos que el nivel de significancia real solo se mantiene por debajo del teórico para la prueba basada en la FDR, y que este es independiente del tamaño de muestra y el número de variables. Finalmente, ilustramos nuestra propuesta con dos ejemplos.

Palabras clave: independencia completa, tasa de falsos descubrimientos, matriz de correlación.

Abstract

Analysis and interpretation of multivariate data is largely facilitated if the variables are independent. In the practice, this supposition is verified through a test for complete independence. We propose a new test for complete independence based on the false discovery rate (FDR), and report the results of a simulation study which compares the real significance levels of this proposal and other tests generally used. We found that the real significance level only remains under the theoretical one for the test based on FDR, and that this is regardless the size of the sample and number of variables. Finally, we illustrate our proposal with two examples.

Keywords: complete independence, false discovery rate, correlation matrix.

^aTranslational Genomics Group, Genome Biology Department, John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia. Grupo de Neurociencias de Antioquia, Universidad de Antioquia, Colombia. Grupo de Investigación en Estadística, Universidad Nacional de Colombia, sede Medellín, Colombia.

^bGrupo de Investigación en Estadística, Universidad Nacional de Colombia, sede Medellín, Colombia. Profesor asociado, Escuela de Estadística, Universidad Nacional de Colombia, sede Medellín, Colombia.

1. Introducción

Desarrollos relativamente recientes en genética y procesamiento de imágenes han dado lugar a experimentos y aplicaciones cuyos resultados corresponden a grandes conjuntos de datos. En la actualidad, una de las áreas más importantes en el campo de la investigación médica es el estudio de niveles de expresión de m genes en n_1 casos y n_2 controles utilizando microarreglos (Nguyen et al. 2002, Dudoit et al. 2002).

Por lo general, en este tipo de estudios el interés se centra en determinar aquellos genes para los que sus niveles de expresión difieren significativamente entre ambos grupos, o aquellos pares de genes para los cuales su correlación es estadísticamente significativa para alguna probabilidad de error tipo I α predeterminada. En el primer caso, la detección de diferencias significativas permite determinar qué genes se encuentran alterados en los casos y no en los controles (Dudoit et al. 2002), mientras en el segundo el coeficiente de correlación corresponde a un *proxy* que cuantifica una potencial interacción entre un par de genes.

Dado un conjunto de datos con p variables numéricas, independencia completa se refiere a probar

$$H_0: \Sigma = D(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2) \quad (1)$$

donde Σ es la matriz de varianzas-covarianzas, $D(\cdot)$ corresponde a una matriz diagonal cuyas componentes son $\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$, con σ_i^2 la varianza de la i -ésima variable, $i = 1, 2, \dots, p$. Si P es la matriz de correlación, lo anterior se reduce a probar $H_0: P = I_p$, con I_p una matriz diagonal de orden p . Si se rechaza H_0 para algún nivel de significancia α , esto indica que por lo menos una de las correlaciones entre pares de variables es estadísticamente diferente de cero. En el caso de estudios con microarreglos, rechazar la hipótesis de independencia completa indicaría que existe al menos un par de genes que interactúan.

En la literatura, se encuentran disponibles la prueba basada en la razón de verosimilitud (LRT, en inglés) (Wilks 1935, Morrison 2005), y las aproximaciones según Box (1949), Bartlett (1954) y Schott (2005) para probar H_0 . El rechazo o no de H_0 en (1) puede contextualizarse de diferentes maneras, por ejemplo, en regresión lineal, es deseable que las covariables del modelo sean ortogonales, i.e., no exista multicolinealidad (no se rechaza H_0). Sin embargo, la existencia de correlación es deseable en análisis de componentes principales y análisis factorial. En el primer caso, permite reducir la dimensionalidad del problema; en el segundo, las pruebas de independencia completa se utilizan para determinar si es recomendable realizar este tipo de análisis.

En este trabajo proponemos una metodología basada en la tasa de falsos descubrimientos (FDR, en inglés) y comparamos su desempeño con los métodos mencionados previamente. Finalmente, ilustramos nuestra propuesta con dos conjuntos de datos, uno relacionado con un experimento taxonómico (Anderson 1935) y otro sobre medidas de los pies (Correa 2006).

2. Pruebas de independencia completa

Supongamos que se tiene una matriz de datos (continuos) $X_{n \times p}$ donde n corresponde al número de observaciones y p al número de variables. La matriz de correlación muestral R está dada por

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix} \quad (2)$$

donde $r_{ij} = s_{ij}(s_{ii}s_{jj})^{-1/2}$, $s_{ij} = (n-1)^{-1} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)$ y $s_{ii} = \widehat{\text{var}}(x_i)$, $i, j = 1, 2, \dots, p$.

Bajo normalidad multivariada, esto es, si $X \sim N_p(\mu, \Sigma_p)$, Wilks (1935) mostró que para probar (1) el estadístico de prueba es

$$\Lambda = |R|^{n/2} \quad (3)$$

La distribución de Λ es aún motivo de extensa investigación (Mudholkar et al. 1982).

2.1. Razón de verosimilitud (LRT)

La prueba LRT (Morrison 2005, Sección 1.9) considera el estadístico de prueba

$$G = -[n - (2p + 5)/6] \log |R| \quad (4)$$

Bajo H_0 , G tiene una distribución χ^2 con $p(p-1)/2$ grados de libertad. Schott (2005) mostró que G tiene un pobre desempeño cuando $p \rightarrow \infty$ debido a que $|R| \rightarrow 0$.

2.2. Aproximación de Box

De acuerdo con Box (1949), la distribución de Λ en (3) puede aproximarse como:

$$\begin{aligned} P[C \leq z] &\approx P\{\chi_w^2 \leq z\} \\ &+ \frac{\gamma_2}{m^2} (P\{\chi_{w+4}^2 \leq z\} - P\{\chi_w^2 \leq z\}) + O(m^{-3}) \end{aligned} \quad (5)$$

con $C = -n \log |R|$, $\gamma_2 = p(p-1)(2p^2 - 2p - 13)/288$, $m = n - (2p + 11)/6$, $w = p(p-1)/2$ y χ_w^2 una variable χ^2 con w grados de libertad.

2.3. Aproximación de Bartlett

Morrison (1976) menciona que Bartlett (1954) propuso aproximar el estadístico

$$Q = -[n - (2p + 11)/6] \log |R| \quad (6)$$

utilizando una $\chi_{p(p-1)/2}^2$. El estadístico Q , además de ser mucho más simple de calcular que la aproximación de Box, no utiliza la expansión de Taylor de tercer orden, lo cual garantiza esta converja mucho más rápido a la distribución límite. Mudholkar et al. (1982) mostraron que el estadístico Q tiene un mejor desempeño que la aproximación en (5).

2.4. Aproximación de Schott

Schott (2005) propone una prueba de independencia completa basada en probar las hipótesis

$$H_0: \rho_{i,j} = 0 \text{ vs. } H_1: \rho_{i,j} \neq 0 \quad i > j \quad (7)$$

donde $\rho_{i,j}$ es el (i, j) -ésimo elemento de la matriz de correlación P . El estadístico de prueba está dado por

$$Z = t_{n,p} / \sigma_{t_{n,p}} \quad (8)$$

donde

$$t_{n,p} = \sum_{i=2}^p \sum_{j=1}^{i-1} r_{i,j}^2 - 2p(p-1)/n,$$

$\sigma_{t_{n,p}} = n^{-2}(n+2)^{-1}p(p-1)(n-1)$ y $r_{i,j}$ es el estimador de $\rho_{i,j}$. Bajo independencia completa, $Z \sim N(0, 1)$. Para $p > n$, esta prueba mostró tener un mejor desempeño que la prueba LRT (Schott 2005).

2.5. Propuesta basada en FDR

La tasa de falsos descubrimientos (FDR, en inglés) está definida como la proporción de hipótesis nulas *verdaderas* que resultan ser *rechazadas* dentro del total de hipótesis rechazadas (Benjamini & Hochberg 1995). Para una revisión sobre este y otros métodos, ver Schaffer (1995) y Correa (2011).

Tabla 1: Posibles resultados cuando se prueban m hipótesis. T : cierto; F : falso, D : descubrimiento (rechazo de H_0); N : no descubrimiento. Fuente: modificado de Benjamini & Hochberg (1995)

| | Acepto H_0 | Rechazo H_0 | Total |
|-----------------|--------------|---------------|-------|
| H_0 Verdadera | NT | DF | m_0 |
| H_0 Falsa | NF | DT | m_1 |
| Total | N | D | m |

En la Tabla 1 presentamos los posibles resultados cuando se realizan m pruebas de hipótesis independientes. A partir de esta información, la FDR se define formalmente como (Benjamini & Hochberg 1995)

$$\text{FDR} = E \left(\frac{DF}{D} \mid D > 0 \right) P(D > 0) \quad (9)$$

y el procedimiento FDR (Benjamini & Hochberg 1995) se reduce a:

1. Probar m hipótesis independientes $H_{0,1}, H_{0,2}, \dots, H_{0,m}$ a partir de las cuales se obtienen los estadísticos de prueba $T_{0,1}, T_{0,2}, \dots, T_{0,m}$ y los valores $p_{0,1}, \dots, p_{0,m}$, respectivamente.
2. Calcular κ como

$$\hat{\kappa} = \text{máx} \left\{ i : p_{(i)} \leq \frac{i}{m} \alpha \right\} \quad (10)$$

para algún nivel de significancia $\alpha \in (0, 1)$.

3. Rechazar $H_{0,1}, H_{0,2}, \dots, H_{0,\hat{\kappa}}$. Si no existe tal $\hat{\kappa}$, ninguna hipótesis nula podrá ser rechazada.

Nuestra propuesta para realizar la prueba de independencia completa implica probar (7) para la l -ésima componente de la matriz de correlación R en (2) y aplicar el procedimiento FDR descrito anteriormente, de tal manera que para una matriz de datos (continuos) $X_{n \times p}$ deben calcularse $p(p-1)/2$ coeficientes de correlación, estadísticos de prueba y valores p . Una vez calculado el l -ésimo coeficiente de correlación muestral r_l , el estadístico de prueba es $t_l = r_l(n-2)^{1/2}(1-r_l^2)^{-1/2}$ y el valor- p puede calcularse como $p_l = P(t_l > t_{n-2})$, $l = 1, 2, \dots, p(p-1)/2$. Rechazaremos H_0 en (1) si $\hat{\kappa} \geq 1$. En el Apéndice A presentamos una implementación de nuestra propuesta en R (R Core Team 2013).

3. Estudio de simulación y resultados

3.1. Estudio de simulación

El desempeño de nuestra propuesta y las aproximaciones antes mencionadas se evaluó a través de un estudio de simulación. Para ello, implementamos un algoritmo en R que funciona de la siguiente manera¹:

1. *Generación de datos.* Defina la tripleta (n, p, ρ) y genere una muestra aleatoria de tamaño n de una distribución normal p -variada con matriz de correlación $P = (1 - \rho)I_p + \rho \mathbf{1}_p \mathbf{1}'_p$. Se utilizaron $10 \leq n \leq 200$, $p = \{2, 5, 10, 30, 50, 100\}$.
2. *Aproximaciones LRT, Box, Bartlett y Schott.* Estime la matriz de correlación muestral R y determine si se rechaza H_0 en (1) con las aproximaciones LRT, Bartlett, Box y Schott.
3. *Prueba basada en la FDR.* A partir de la matriz de correlación R , determine el valor de $\hat{\kappa}$ como se describe en la sección 2.5. Rechace H_0 en (1) si $\hat{\kappa} \geq 1$.
4. *Tasa de rechazos.* Repita los pasos 1–3, B veces. Calcule la tasa de rechazos (TdRs) para cada método como la proporción de veces que se rechaza H_0 en las B muestras.

Con el propósito de estimar el de cada una de las pruebas, se determinó su desempeño bajo H_0 , es decir, con $P = I_p$ (equivalente a $\rho = 0$). En total se evaluaron 240 escenarios de simulación. En todos ellos, la probabilidad de error tipo I fue $\alpha = 0.05$ y $B = 10000$.

3.2. Resultados

Los resultados obtenidos se presentan en la Figura 1. A diferencia de las demás pruebas para independencia completa, nuestra propuesta basada en la FDR mantiene niveles de significancia reales cercanos al 5% independiente del tamaño de muestra n y el número de variables p . Las pruebas tradicionales presentan TdRs cercanos al 5% solo para $n > 100$ y $p \leq 10$ (panel superior, Figura 1). Sin embargo, la prueba de Schott para $p = 2$ y la de Box para $p = 5$ y $p = 10$, son claras excepciones. Para $p > 2$, la prueba LRT presenta las TdRs más altas independiente del tamaño de muestra.

¹El programa en R se encuentra disponible a petición del lector.

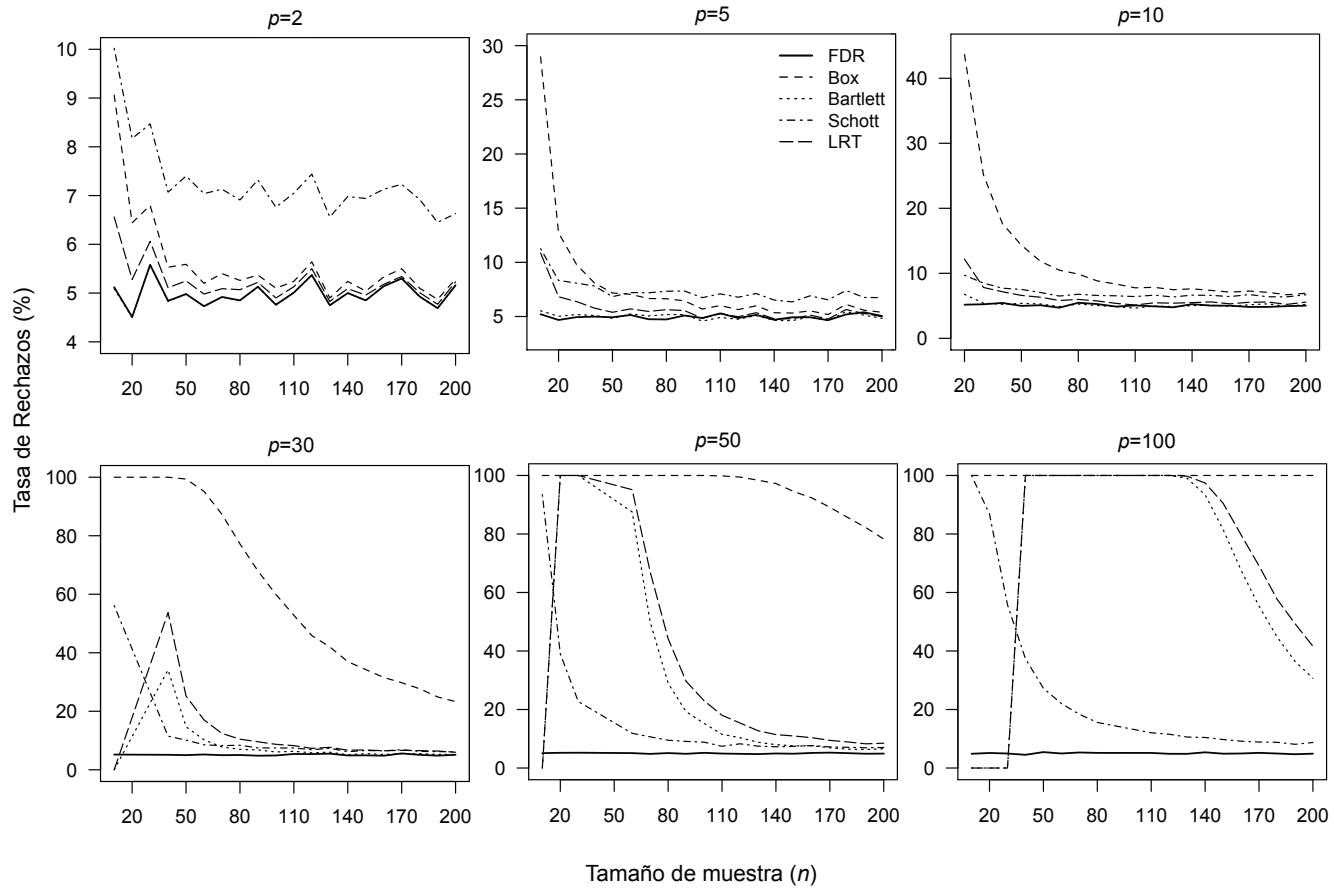


Figura 1: Nivel de significancia real de las aproximaciones LRT, Box, Bartlett, Schott y nuestra propuesta basada en la FDR como función del tamaño de muestra n y el número de variables p . Fuente: elaboración propia.

Cuando $p \geq 30$, las TdRs de las pruebas clásicas son superiores al nivel de significancia nominal del 5% en muchos órdenes de magnitud. A excepción de la prueba de Box, las TdRs son cercanas al nivel nominal cuando $p = 30$ y $n > 120$. Sin embargo, para $p = 50$ y $p = 100$ este comportamiento es menos evidente y las TdRs varían considerablemente. Para $p \geq 30$ (panel inferior, Figura 1) las únicas pruebas evaluadas con resultados consistentes, i.e., que igualan o tienden rápidamente al nivel nominal cuando n aumenta, son nuestra propuesta y la prueba de Schott.

4. Ejemplos

4.1. Datos taxonómicos

Anderson (1935) presenta un conjunto de datos, ahora clásicos, correspondiente a medidas (en centímetros) del ancho y la longitud del sépalo y los pétalos en 150 flores iris de tres especies diferentes (setosa, versicolor y virginica, Figura 2)². Los datos, disponibles en R, corresponden a mediciones de estas cuatro características en 50 flores de cada especie.



Figura 2: *Especies setosa, versicolor y virginica de la variedad de flores iris. Fuente: ver nota de pie.*

Para ilustrar el uso de nuestra propuesta se hicieron dos tipos de análisis. En el primero, se analizó el conjunto de datos sin dividir por especie; en el segundo, se realizó dicha división. En el primer caso, todas las pruebas de independencia rechazaron H_0 en (1), excepto la prueba de Box (LRT: $G = 711.77$, $p = 1.75 \times 10^{-150}$; Bartlett: $Q = 706.96$, $p = 1.92 \times 10^{-149}$; Schott: $Z = 115.79$, $p < 10^{-150}$). Resultados similares se obtienen al realizar el análisis de independencia completa por especie.

²Imágenes tomadas de www.alaska-in-pictures.com/data/media/10/wild-iris_8865.jpg, <http://goo.gl/nEGwEr> y <http://goo.gl/7PRWSf>

4.2. Medidas de pies

Correa (2006) presenta datos³ correspondientes a mediciones de la longitud (x_1) y la amplitud (x_2) máxima del pie, la amplitud máxima del talón (x_3), la longitud máxima del dedo grande (x_4) y la amplitud máxima del dedo grande (x_5) en $n = 10$ personas.

La matriz de correlación muestral (diagonal inferior) y los respectivos valores p (diagonal superior) al probar (7) son

$$D = \begin{pmatrix} & x_1 & x_2 & x_3 & x_4 & x_5 \\ x_1 & - & 0.0018 & 0.5881 & 0.1413 & 0.0107 \\ x_2 & 0.8507 & - & 0.7924 & 0.3767 & 0.1589 \\ x_3 & 0.1956 & -0.0958 & - & 0.9448 & 0.1963 \\ x_4 & 0.4998 & 0.3141 & 0.0253 & - & 0.1989 \\ x_5 & 0.7600 & 0.4814 & 0.4461 & 0.4438 & - \end{pmatrix},$$

de tal manera que $r_{x_1, x_2} = 0.8507$ y el correspondiente valor p es 0.0018.

La hipótesis de independencia completa es rechazada al utilizar nuestra propuesta basada en la FDR. Este resultado es equivalente al obtenido con cualquiera de las aproximaciones clásicas (LRT: $G = 24.45$, $p = 6.48 \times 10^{-3}$; Box: $C = 32.61$, $p = 4.48 \times 10^{-4}$; Bartlett: $Q = 21.19$, $p = 1.97 \times 10^{-2}$; Schott: $Z = 3.42$, $p = 3.10 \times 10^{-4}$).

5. Discusión

El análisis multivariado de datos se simplifica en gran medida si se asume que las p variables disponibles son independientes. La principal ventaja de tener datos con estas características radica en que cada una de las variables podría analizarse utilizando métodos univariados. Ahora, si el supuesto de independencia completa no se cumple y aún así este se asume, podrían obtenerse resultados alejados de la realidad, especialmente cuando el estudio de relaciones entre variables, como en experimentos con microarreglos, es de gran importancia.

En este artículo hemos presentado una nueva prueba de independencia completa, basada en la FDR, que consiste en realizar $p(p-1)/2$ pruebas de hipótesis independientes sobre igual número de coeficientes de correlación de una matriz de datos (continuos) $X_{n \times p}$, y que ofrece una alternativa fácil de implementar en cualquier programa de análisis estadístico (ver Apéndice A para nuestra implementación en R). A diferencia de otras pruebas de independencia completa también evaluadas, el nivel de significancia real de esta nueva alternativa es comparable con el nivel teórico nominal (en este caso del 5%) y no depende del número de variables p ni del tamaño de muestra n . En la práctica, estas tres propiedades son deseables. Por

³Disponibles bajo solicitud expresa del lector.

otro lado, la potencia de nuestra propuesta basada en la FDR es comparable o superior a las obtenidas con las demás aquí mencionadas⁴. Este resultado garantiza, en cierta medida, que nuestra propuesta representa una mejor alternativa a las ya existentes.

Tabla 2: *Tiempos de ejecución (en segundos) de la prueba de independencia completa basada en la FDR para combinaciones de n y p . Fuente: elaboración propia*

| $n p$ | 100 | 200 | 500 | 1000 | 1500 | 2000 | 3000 | 5000 |
|-------|-------|-------|-------|-------|-------|--------|--------|---------|
| 500 | 0.007 | 0.029 | 0.199 | 0.960 | 2.549 | 5.219 | 13.256 | 42.116 |
| 1000 | 0.011 | 0.043 | 0.283 | 1.239 | 3.298 | 6.572 | 16.253 | 50.630 |
| 2000 | 0.017 | 0.071 | 0.456 | 1.919 | 4.787 | 9.017 | 22.076 | 67.222 |
| 3000 | 0.024 | 0.097 | 0.643 | 2.688 | 6.321 | 11.701 | 28.206 | 84.214 |
| 5000 | 0.038 | 0.158 | 0.960 | 3.930 | 9.424 | 17.226 | 39.996 | 117.902 |

Uno de los problemas que podrían presentarse en la implementación y aplicación de esta prueba es el costo computacional (e.g., tiempo de ejecución). Por ejemplo, para $p = 5000$ e independiente del tamaño de muestra, es necesario calcular 12,497,500 coeficientes de correlación, probar igual número de hipótesis y computar los respectivos valores $-p$. Sin embargo, los tiempos de ejecución en R (ver Tabla 2) son relativamente cortos; el procedimiento tarda menos de 120 segundos⁵ para $p = n = 5000$.

Posibles direcciones de investigación podrían estar enfocadas a la evaluación de nuestra propuesta en presencia de datos faltantes. Puesto que la prueba FDR se basa en el cálculo de coeficientes de correlación y la determinación de si al menos uno es significativo, el problema se reduce a escoger diferentes métodos para el cálculo de estos (e.g., usando toda la información, solo la información completa, o solo la información completa por pares de variables) y calcular el nivel de significancia real de la prueba.

Agradecimientos

Los autores agradecen los comentarios y sugerencias de un revisor anónimo, quien ayudó a mejorar sustancialmente la versión previa de este documento. El trabajo de JIV fue financiado parcialmente por The Eccles Scholarship in Medical Sciences, The Fenner Merit Scholarship y The Australian National University (ANU) High Degree Research Scholarship. JIV agradece el apoyo del dr. Mauricio Arcos-Burgos de ANU.

Recibido: 10 de abril de 2013
Aceptado: 14 de mayo de 2013

⁴Los resultados se encuentran disponibles a petición del lector.

⁵Se utilizó R versión 3.0.0 Patched (2013-04-08 r62531) en un MacBook Pro con 8GB de RAM y procesador 2.3 GHz Intel Core i7.

Referencias

- Anderson, E. (1935), 'The irises of the Gaspe peninsula', *Bulletin of the American Iris Society* **59**, 2–5.
- Bartlett, M. (1954), 'A note on multiplying factors for various χ^2 approximations', *Journal of the Royal Statistical Society, Ser. B (Methodological)* **16**, 296–298.
- Benjamini, Y. & Hochberg, Y. (1995), 'Controlling the false discovery rate: A practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society, Series B (Methodological)* **57**(1), 389–400.
- Box, G. (1949), 'A general distribution theory for a class of likelihood criteria', *Biometrika* **36**, 317–346.
- Correa, J. C. (2006), *Control de la proporción de hipótesis rechazadas equivocadamente*, Curso de Estadística Genética, Universidad de Antioquia.
- Correa, J. C. (2011), 'Diagnósticos de regresión usando la FDR (Tasa de Descubrimientos Falsos)', *Comunicaciones en Estadística* **3**(2), 109–118.
- Dudoit, S., Yang, Y.-H., Callow, M. J. & Speed, T. P. (2002), 'Statistical methods for identifying differentially expressed genes in replicated cDNA experiments', *Statistica Sinica* **12**, 111–139.
- Morrison, D. F. (1976), *Multivariate statistical methods*, 2 edn, New York: McGraw-Hill.
- Morrison, D. F. (2005), *Multivariate statistical methods*, 4 edn, Belmont, CA: Brooks/Cole.
- Mudholkar, G. S., Trivedi, M. C. & Lin, T. (1982), 'An approximation to the distribution of the likelihood ratio statistic for testing complete independence', *Technometrics* **24**(2), 139–143.
- Nguyen, D. V., Bulak Apart, A., Wang, N. & Carrol, R. J. (2002), 'DNA microarray experiments: biological and technological aspects', *Biometrics* **58**, 701–717.
- R Core Team (2013), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*<http://www.R-project.org/>
- Schaffer, J. P. (1995), 'Multiple hypothesis testing: A review', *Annu. Rev. Psychol.* **46**, 561–84.
- Schott, J. R. (2005), 'Testing for complete independence in high dimensions', *Biometrika* **92**(4), 951–956.
- Wilks, S. S. (1935), 'On the independence of k sets of normally distributed statistical variables', *Econometrika* **3**, 309–26.

A. Apéndice. Programa en R para la prueba de independencia completa basada en la FDR

```
## Cálculo del valor p para los coef. de correlacion de una matriz X
cor.pvalue <- function(X, method = "pearson", use = "complete") {
  dfr <- nrow(X) - 2
  R <- cor(X, method = method, use = use)
  above <- row(R) < col(R)
  r2 <- R[above]^2
  Fstat <- r2 * dfr / (1 - r2)
  R[above] <- 1 - pf(Fstat, 1, dfr)
  R[above]
}

## Prueba de independencia completa basada en la FDR
fdrci <- function(x, alpha = 0.05) {
  p <- cor.pvalue(x)
  p <- p.adjust(p, method = "fdr")
  ifelse(any(p <= alpha), "Rechace H0", "No rechace H0")
}

## Ejemplo con una normal 5-variada independiente, n = 100
set.seed(1)
X <- matrix(rnorm(100 * 5), ncol = 5)
fdrci(X)

## [1] "No rechace H0"
```