

Obtención de la matriz de varianzas y covarianzas a través de los productos Kronecker en modelos balanceados de dos y tres vías con aplicaciones en R

Luz Marina Moya-Moya^{1*}, Milton Januario Rueda-Varón^{2**}

¹Departamento de Matemáticas, ²Grupo de Física Matemática. Facultad de Ciencias, Pontificia Universidad Javeriana, Bogotá, D.C., Colombia

*: luz.moya@javeriana.edu.co; milton.rueda@javeriana.edu.co

Recibido: 15-06-2011; Aceptado: 08-11-2011

Resumen

Objetivo. Presentar una metodología basada en el concepto de productos Kronecker que facilita la construcción de la matriz de varianzas y covarianzas para diseños con estructura balanceada de datos a 2 y 3-vías y una aplicación realizada en R para facilitar su cálculo y aplicación en diferentes áreas. **Materiales y métodos.** Se proporciona un punto de partida para personas interesadas en utilizar R en el análisis de varianza. **Resultados.** Se utiliza una aplicación realizada en R donde se desarrolla la metodología basada en productos Kronecker mediante la cual se construye la matriz de varianzas y covarianzas cuando se trabaja en diseños con estructura balanceada de datos desarrollada por Moya 2003. De igual forma se presenta una aplicación del método con datos reales. **Conclusiones.** La metodología expuesta permite agilizar el desarrollo y solución de algunos problemas prácticos. El método propuesto puede ser aplicado a modelos mixtos con efectos fijos o aleatorios con cualquier número de factores.

Palabras clave: productos Kronecker, matriz de varianzas y covarianzas, diseños balanceados, modelos lineales, R Gui.

Abstract

Obtainment of the variance-covariance matrix through Kronecker products for balanced models of two and three ways with applications in R. Objective. To present a methodology based on the concept of Kronecker products that facilitates the construction of the variance and covariance matrix for designs with balanced data structure for 2 and 3 ways, and an application in R to facilitate its calculation and application in different areas. **Materials and methods.** We provide a starting point for people interested in using R in the analysis of variance. **Results.** We use an application made in R for a methodology based on Kronecker products through which we build the covariance matrix for working with designs with balanced data structure developed by Moya (2003). We also present an application of the method with real data. **Conclusions.** With this methodology we can accelerate the development and solution of some practical problems. The proposed methodology can be applied to mixed models with fixed or random effects with any number of factors.

Key words: Kronecker products, variance and covariance matrix, balanced designs, linear models, R Gui.

Resumo

Obtenção da matriz de variância-covariância através dos produtos Kronecker em modelos balanceados de duas e três vias com aplicações em R. Objetivo. Apresentar uma metodologia baseada no conceito de produtos de Kronecker para facilitar a construção da matriz de variância-covariância para desenhos com estrutura balanceada de dados em 2 e 3 vias e uma aplicação realizada em R para

o cálculo fácil e aplicação em diferentes áreas. **Materiais e métodos.** É fornecido um ponto de partida para as pessoas interessadas em utilizar R na análise de variância. **Resultados.** Usa-se uma aplicação realizada em R, onde é desenvolvida a metodologia baseada nos produtos de Kronecker com a qual se constrói a matriz de variância-covariância quando se trabalha em desenhos com estrutura balanceada de dados segundo Moya (2003). Da mesma forma, se apresenta uma aplicação do método com dados reais. **Conclusões.** A metodologia descrita pode acelerar o desenvolvimento e solução de alguns problemas práticos. O método proposto pode ser aplicado a modelos mistos com efeitos fixos ou aleatórios com qualquer número de fatores.

Palavras-chave: produtos de Kronecker, matriz de variância-covariância, desenhos balanceados, modelos lineares, R Gui.

Introducción

Hoy en día el avance de los procesadores y el desarrollo de software estadístico han permitido que los investigadores en diversas áreas del conocimiento tengan a su disposición una gran cantidad de herramientas computacionales que les permiten un análisis óptimo de la información. Dentro de estas herramientas, se encuentran aplicaciones de distribución libre, como es el caso del R, el software estadístico más empleado a nivel mundial en investigación. Lamentablemente, en diferentes áreas aplicadas, su uso se ha visto limitado por el desconocimiento de sus procedimientos y su lenta divulgación.

El objetivo de este artículo es proporcionar un punto de partida para personas interesadas en utilizar R en el análisis de varianza, y presentar una aplicación realizada en el mismo software, donde se desarrolla la metodología basada en productos Kronecker mediante la cual se construye la matriz de varianzas y covarianzas cuando se trabaja en diseños con estructura balanceada de datos a 2 y 3-vías desarrollada por Moya (1).

Los algoritmos presentados pueden ser empleados en modelos a 1 y 2-vías con cualquier número de factores. La matriz de varianzas y covarianzas se construye bajo los supuestos usuales para modelos finitos utilizando el método propuesto por Moya (1). Se ha escogido utilizar R, debido a la universalidad de su uso y libre distribución. R se ofrece gratuitamente bajo los términos de la *GNU General Public Licence*; su desarrollo y distribución son llevados a cabo por varios estadísticos conocidos como el *Grupo Nuclear de Desarrollo de R*. Se ha tratado de simplificar las explicaciones al máximo con el fin de hacerlas lo más comprensivas posibles para que usuarios de cualquier nivel puedan utilizar los desarrollos aquí planteados.

Para profundizar en la teoría subyacente al análisis de la varianza y sus aplicaciones se puede consultar, entre otros, Carmona (2), McCulloch (3), Sudhir y Saha (4) o los clásicos Hartley y Rao (5) y Searle (6). Si se quiere profundizar en el aprendizaje de R o desarrollar otro tipo de aplicaciones, en el link http://www.sciviews.org/_rgui/ se encuentra toda

la información necesaria. También se pueden consultar Crawley (7) y Dalgaard (8), donde se presentan algunas aplicaciones de R.

Materiales y métodos

Con el fin de realizar una adecuada descripción de los métodos de estimación de componentes de varianza, se presenta la forma más general de un modelo lineal denominado modelo mixto general. La formulación del modelo mixto general originalmente propuesto por Hartley y Rao (5) y aceptada como estándar en el campo de la estimación de las componentes de varianza esta dada por:

$$y = X\beta + ZU + e \quad [1]$$

Donde:

$y_{N \times 1}$: Vector de observaciones

$X_{N \times K}$: Matriz asociada a los efectos fijos

$\beta_{K \times 1}$: Vector de parámetros asociados a los efectos fijos

$U_{q \times 1}$: Vector asociado a los efectos aleatorios.

$Z_{N \times q}$: Usualmente es una matriz de incidencia asociada a efectos aleatorios

$e_{N \times 1}$: Vector de términos del error.

Como:

$$ZU_{N \times 1} = \begin{bmatrix} Z_1 & \cdots & Z_S \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_S \end{bmatrix} = \sum_{i=0}^S Z_i U_i \quad [2]$$

Donde: Z_i son los vectores columna de Z .

por lo tanto:

$$y = X\beta + \sum_{i=0}^S Z_i U_i \quad [3]$$

De esta forma, la varianza de y puede ser definida como:

$$y = \sum_{i=1}^S Z_i Z_i^T \sigma_i^2 + \sigma_e^2 I_N \quad [4]$$

A e se le atribuye la estructura usual de varianza-covarianza de un componente de error, dado por:

$$E(e) = 0 \quad Var(e) = \sigma_e^2 I_N \quad [5]$$

En primer lugar se presentan los algoritmos realizados en R para realizar análisis de varianza a 1 y 2-vías.

Diseño Aleatorio a 1-vía

Con el fin de ilustrar la utilización de R en el análisis de varianza, se puede definir el modelo a 1-vía como:

$$y_{i_1} = \mu + \tau_{i_1}^1 + e_{i_1}, \quad i_1 = 1, \dots, n_1 \quad [6]$$

Donde: τ^1 denota el factor 1 con $\tau^1 \sim N(0, \sigma_{\tau^1}^2)$ y el termino e corresponde al error definido anteriormente

con $e_{i_1} \sim N(0, \sigma_{e_{i_1}}^2)$.

En primer lugar se deben ingresar los datos (note que existen k niveles), para tal fin se usan las siguientes instrucciones:

```
> Data.Nivel1 <- c(y11, y12, ..., y1n1)
:
> Data.Nivelk <- c(yk1, yk2, ..., yknk)
```

Sin embargo debido a la estructura de R, existe una forma adecuada para organizar la información que se pretende analizar. Primero que todo se organizan los datos en un único vector y se agrega una variable cualitativa o factor que nos indique la población correspondiente a cada caso, así:

```
>Data <- c(Data.Nivel1, Data.Nivel2, ..., Data.Nivelk)
> Nivel <- rep(1:k, each=n1)
> Nivel <- factor(Tratamiento, labels=c("Data.Nivel
1", ..., "Data.Nivel k"))
```

Es necesario notar que en R, es imprescindible definir el vector tratamiento como un factor, ya que en caso contrario se podría confundir con un vector numérico y esto afectaría el análisis, además en este caso $n_1 = n_2 = \dots = n_k$. Existe una instrucción que implementa los dos pasos anteriores:

```
> Data <- gl(k, n1, labels=c("Data.Nivel1", "Data.
Nivel2", ..., "Data.Nivelk"))
```

Una vez se tienen los datos organizados en R, se pueden obtener algunas estadísticas descriptivas o un gráfico de medias. De igual forma, resulta útil para análisis de esta naturaleza, realizar un boxplot múltiple, mediante las instrucciones:

```
> tapply(Data, Nivel, summary)
> plotMeans(Data$data, Data$Nivel, error.bars="se")
> plot(Data~ Nivel)
```

Si se asume que la variable Data sigue una distribución normal con igual varianza en las k poblaciones, el ANOVA o la tabla de análisis de varianza, se obtiene con:

```
> Data.av <- aov(Data~ Nivel)
> summary(Data.av)
```

Si se quiere ser más específico o más formal, se puede utilizar la definición del modelo lineal y obtener los mismos resultados mediante

```
> Data.glm <- lm(Data~ Nivel)
> anova(Data.glm)
```

Además, las estimaciones de los respectivos parámetros, pueden ser calculadas con la instrucción:

```
> model.tables(Data.av)
```

Un resumen de todo el modelo con la información pertinente se obtiene con

```
>summary (Data.glm)
```

Diseño Aleatorio a 2-vías

En primer lugar, es necesario definir el modelo. Así, El modelo a 2-vías puede ser expresado como:

$$y_{i_1, i_2} = \mu + \tau_{i_1}^1 + \tau_{i_2}^2 + e_{i_1, i_2}, \quad i_1 = 1, \dots, n_1 \quad y \quad i_2 = 1, \dots, n_2 \quad [7]$$

Donde: τ^1 denota el factor 1 con $\tau^1 \sim N(0, \sigma_{\tau^1}^2)$, τ^2 denota el factor 2 con $\tau^2 \sim N(0, \sigma_{\tau^2}^2)$, y el término e_{i_1, i_2} corresponde al error definido anteriormente con $e_{i_1, i_2} \sim N(0, \sigma_{e_{i_1, i_2}}^2)$.

En este caso los datos son introducidos mediante:

```
> Data2 <- c(y111, y112, ..., y11n1, ..., ypk1, ypk2, ..., ypknk)
> Tratamiento1 <- gl(k, p)
> Tratamiento2 <- factor(rep(1:k, p))
> xtabs(Data2 ~ Tratamiento1 + Tratamiento2)
```

Usando comandos similares a los presentados en el caso a 1-vía, se puede generar un resumen de los datos y los respectivos gráficos que acompañan este análisis utilizando las instrucciones:

```
> tapply(Data2, Tratamiento, summary)
> tapply(Data2, Bloque, summary)
> stripchart(Data2 ~ Tratamiento1, method = "stack")
> stripchart(Data2 ~ Tratamiento2, method = "stack")
> interaction.plot(Tratamiento, Bloque, Data2)
> interaction.plot(Bloque, Tratamiento, Data2)
```

Y el modelo lineal y la tabla del análisis de la varianza se construye con:

```
> Data2.glm <- lm(Data2 ~ Tratamiento1 + Tratamiento2)
> anova(Data2.glm)
```

Como se puede observar en los anteriores procedimientos, se utilizan instrucciones similares a las presentadas en el caso de 1-vía, es decir al lector interesado en realizar análisis a mayor nivel, le resultará sencillo establecer un patrón de generalización de las instrucciones de acuerdo al número de componentes que requiera.

A continuación se presenta una aplicación en R que permite la construcción de la matriz de varianzas y covarianzas cuando se trabaja en diseños con estructura balanceada de datos a 1 y 2-vías basada en la metodología propuesta por Moya (2003). En este tipo de modelo es importante determinar los mejores estimadores lineales insesgados (MELI), y los

mejores predictores lineales (BLUP). Estos estimadores juegan un importante papel a la hora de realizar inferencia puntual y por intervalos sobre los parámetros. Tanto los MELIS como los BLUPS dependen del conocimiento de la estructura apropiada de la matriz de varianzas y covarianzas, así como de su inversa y es allí donde radica la importancia de la aplicación aquí presentada.

Resultados

Construcción de la matriz de varianzas y covarianzas

Un diseño factorial es aquel en el que se investigan todas las posibles combinaciones de los niveles de los factores en cada ensayo completo o réplica del experimento. Por ejemplo, si existen n_1 niveles del factor τ^1 y n_2 niveles del factor τ^2 , entonces cada réplica del experimento contiene todas

las $n_1 n_2$ combinaciones de los tratamientos. A menudo, se dice que los factores están cruzados cuando estos se arreglan en un diseño factorial. En este artículo se determina la matriz de varianzas y covarianzas (V) para diseños a dos y tres vías sin tener en cuenta el efecto de réplica. Para el cálculo de la matriz V en el caso de dos vías con efecto de réplica ver Moya (1).

Modelo Aleatorio a 2-vías

A continuación se presenta la metodología mediante la cual, se construye la matriz V de manera sencilla y directa. De igual forma, se propone una metodología para hallar la inversa de esta matriz.

En la determinación de la matriz de estructura

$$V_{i_1, i_2} = \sum_{i=0}^s Z_i Z_i^T \sigma_i^2 \quad [9]$$

Donde: $Z_i Z_i^T$ es la matriz de covarianzas y puede ser escrita en términos de productos Kronecker [\otimes] de matrices I_s (matriz identidad de tamaño $S \times S$) y J_s (matriz de unos con tamaño $S \times S$) como se presenta a continuación, y σ_i^2 es la varianza de y_{i_1, i_2} .

Sea P el número de subíndices asociados a la variable respuesta (en este caso, $P = 2$), los cuales dependen de los efectos e interacciones fijas y/o aleatorias del modelo,

entonces se tienen 2^P particiones de productos Kronecker con matrices I_s y J_s , las cuales son resumidas por la expresión:

$$\Delta_{1 \times 2^P} = [I_{n_1} \otimes I_{n_2}; I_{n_1} \otimes J_{n_2}; J_{n_1} \otimes I_{n_2}; J_{n_1} \otimes J_{n_2}] \quad [10]$$

y definiendo la función indicadora:

$$\gamma_{\text{efecto}}^2 = \begin{cases} 1 & \text{si aparece } J_s \\ 0 & \text{si aparece } I_s \end{cases} \quad [11]$$

Se puede reescribir $\Delta_{1 \times 2^P}$ para el caso $P = 2$, como:

$$\Delta_{1 \times 2^2}^* = [00; 01; 10; 11] \quad [12]$$

Caracterizando el vector $\Delta_{1 \times 2^2}^*$, se construye el vector asociado con las componentes de varianza $\theta_{1 \times 2^2} = \theta_p$, según el siguiente criterio:

$$\theta_p = \begin{cases} \sigma^2 & \text{si existe con relación entre el} \\ & \text{subíndice del efecto y la posición} \\ & \text{del cero en } \Delta_{1 \times 2^2}^* \\ 0 & \text{en caso contrario} \end{cases} \quad [13]$$

con esto

$$V = \Delta_{1 \times 2^2}^* [\theta^0]^T \quad [14]$$

Donde: $\Delta_{1 \times 2^2}^*$ es la matriz final de permutaciones y θ^0 representa una permutación de θ_p dada por

$$\theta^0 = [\sigma_e^2; \sigma_{\tau^1}^2; \sigma_{\tau^2}^2; 0] \quad [15]$$

Donde: σ_e^2 , $\sigma_{\tau^1}^2$ y $\sigma_{\tau^2}^2$ son los respectivos componentes de varianza.

Por otro lado, definiendo

$$\lambda_p = \theta_p T_p^T, \quad [16]$$

se tiene que

$$V^{-1} = \Delta_{1 \times 2^2}^* T_p^{-1} \lambda_p^{-1} \quad [17]$$

Donde:

$$T_p = \begin{bmatrix} 1 & 0 \\ 1 & n_1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 1 & n_2 \end{bmatrix} \quad [18]$$

y

$$T_p^{-1} = \frac{1}{n_1} \begin{bmatrix} n_1 & 0 \\ -1 & 1 \end{bmatrix} \otimes \frac{1}{n_2} \begin{bmatrix} n_2 & 0 \\ -1 & 1 \end{bmatrix} \quad [19]$$

A continuación se construye una función en R que realiza este cálculo de manera sencilla con el fin de facilitar su aplicación práctica.

El primer paso en la construcción de la matriz V , es la definición del vector asociado a las componentes de varianza θ^0 , como:

$$\theta^0 = [\sigma_e^2; \sigma_{\tau^1}^2; \sigma_{\tau^2}^2; 0] \quad [20]$$

Donde: σ_e^2 , $\sigma_{\tau^1}^2$ y $\sigma_{\tau^2}^2$ son los respectivos componentes de varianza.

Este vector será el insumo del algoritmo realizado en R, al igual que n_1 y n_2 que corresponden a los niveles del primer y segundo factor aleatorio (τ^1 y τ^2), respectivamente. Para calcular la matriz V , primero que todo, se debe compilar la función (V_2vias) que aparece a continuación:

```
#Construcción de V_2vias la matriz de varianzas
y covarianzas
> V_2vias <- function(Se,S1,S2,n1,n2){
+ (Se*diag(n1*n2))+(S1*kroncker ((diag(n1)),matrix
(1, ncol=n2, nrow=n2)))+
+ S2*kroncker(matrix(1, ncol=n1, nrow=n1),(diag(n2)))
+ }
```

Luego para obtener la matriz V , basta con utilizar la instrucción:

```
>V_2vias (Se,S1,S2,n1,n2)
```

Donde: Se corresponde a σ_e^2 , S1 a $\sigma_{\tau^1}^2$, S2 a $\sigma_{\tau^2}^2$, n_1 y n_2 corresponden al total de niveles de τ^1 y τ^2 , respectivamente.

Finalmente para calcular la inversa de la matriz V , primero que todo, se debe compilar la función (Vinv_2vias) que aparece a continuación:

```

> Vinv_2vias <- function(Se,S1,S2,n1,n2) + {
+ Theta <- matrix(c( Se, S1, S2, 0),nrow = 1, ncol = 4)
+ Tp <- matrix(c(1,1,1,1,0,n2,0,0,n1,n1,0,0,0,(n1
*n2)),nrow = 4, ncol = 4)
+ Lambda <- Theta %*% t(Tp)
+ Lambdai <- 1/ Lambda
+ Tpi <- solve(Tp)
+ M<- Tpi %*% t(Lamdai)
+ D1 <- diag(n1*n2)*M[1,1]
+ D2 <- (kronecker ((diag(n1)),matrix( 1, ncol=n2,
nrow=n2)))* M[2,1]
+ D3 <- (kronecker(matrix( 1, ncol=n1,
nrow=n1),(diag(n2))))* M[3,1]
+ D4 <- (matrix( 1, ncol=(n1*n2), nrow=(n1*n2)))*
M[4,1]
+ D1+D2+D3+D4
+ }
    
```

De esta forma, para obtener la matriz inversa V^{-1} , basta con utilizar la instrucción:

```
> Vinv_2vias (Se,S1,S2,n1,n2)
```

Donde: Se corresponde a σ_e^2 , S1 a $\sigma_{\tau_1}^2$, S2 a $\sigma_{\tau_2}^2$, n1 y n2 corresponden al total de niveles de τ^1 y τ^2 , respectivamente.

Modelo 3-vías

Utilizando una metodología similar a la aplicada en el modelo a 2-vías, V puede ser calculada como

$$V = \Delta_{1 \times 2^3}^0 [\theta^0]^T \tag{21}$$

Donde:

$$\theta^0 = [\sigma_e^2; 0; 0; 0; 0; \sigma_{\tau_3}^2; \sigma_{\tau_2}^2; \sigma_{\tau_1}^2; 0] \tag{22}$$

Donde: σ_e^2 , $\sigma_{\tau_1}^2$, $\sigma_{\tau_2}^2$ y $\sigma_{\tau_3}^2$ son los componentes de varianza, y

$$\Delta_{1 \times 2^3}^0 = [I_{n_1 n_2 n_3}; I_{n_1 n_2} \otimes J_{n_3}; I_{n_1} \otimes J_{n_2} \otimes I_{n_3}; J_{n_1} \otimes I_{n_2 n_3}; J_{n_1 n_2} \otimes I_{n_3}; J_{n_1} \otimes I_{n_2} \otimes J_{n_3}; I_{n_1} \otimes J_{n_2 n_3}; J_{n_1 n_2 n_3}] \tag{23}$$

Además

$$V^{-1} = \Delta^0 T_p^{-1} \lambda_p^{-1} \tag{24}$$

Donde: λ_p es definido como en [16],

y

$$T_p = \begin{bmatrix} 1 & 0 \\ 1 & n_1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 1 & n_2 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 1 & n_3 \end{bmatrix} \tag{25}$$

$$T_p^{-1} = \frac{1}{n_1} \begin{bmatrix} n_1 & 0 \\ -1 & 1 \end{bmatrix} \otimes \frac{1}{n_2} \begin{bmatrix} n_2 & 0 \\ -1 & 1 \end{bmatrix} \otimes \frac{1}{n_3} \begin{bmatrix} n_3 & 0 \\ -1 & 1 \end{bmatrix} \tag{26}$$

A continuación se construye la respectiva función en R mediante la cual se realizan los cálculos anteriormente expuestos.

Como en el caso a 2-vías, el primer paso en la construcción de la matriz de V es la definición del vector asociado a las componentes de varianza θ^0 , así:

$$\theta^0 = [\sigma_e^2; 0; 0; 0; 0; \sigma_{\tau_3}^2; \sigma_{\tau_2}^2; \sigma_{\tau_1}^2; 0] \tag{27}$$

Donde: σ_e^2 , $\sigma_{\tau_1}^2$, $\sigma_{\tau_2}^2$ y $\sigma_{\tau_3}^2$ son los componentes de varianza.

Este vector será utilizado por el algoritmo realizado en R, al igual que n_1 , n_2 y n_3 , que corresponden a los niveles del primer, segundo y tercer factor aleatorio (τ^1 , τ^2 y τ^3), respectivamente. Para calcular la matriz se varianzas y covarianzas, primero que todo, se debe compilar la función (V_{3vias}) que aparece a continuación:

```

> V_3vias <- function(Se,S1,S2,S3,n1,n2,n3){
+ (Se*diag(n1*n2*n3)) +S3* kronecker(matrix( 1,
ncol=n1*n2, nrow=n1*n2),(diag(n3)))+
+ S2* kronecker(kronecker(matrix( 1, ncol=n1,
nrow=n1),(diag(n2))), matrix( 1, ncol=n3, nrow=n3))+
+ (S1* kronecker ((diag(n1)),matrix( 1, ncol=n2*n3,
nrow=n2*n3)))
+ }
    
```

Para obtener la matriz V , se utiliza la instrucción:

```
>V_3vias (Se,S1,S2,S3,n1,n2,n3).
```

Donde: Se corresponde a σ_e^2 , S1 a $\sigma_{\tau_1}^2$, S2 a $\sigma_{\tau_2}^2$, S3 a $\sigma_{\tau_3}^2$, n1, n2 y n3 corresponden a los niveles de τ^1 , τ^2 y τ^3 , respectivamente.

Para calcular la inversa de la matriz V , primero que todo, se debe compilar la función (Vinv_3vias) que aparece a continuación:

```
> Vinv_3vias <- function(Se,S1,S2,S3,n1,n2,n3)
+ {
+ Theta <- matrix(c( Se, 0,0,0,S3,S2, S1, 0) ,nrow =
+ 1, ncol = 8)
+ Tp1 <- kronecker(matrix(c(1,1,0,n2),nrow = 2, ncol
+ = 2), matrix(c(1,1,0,n3),nrow = 2, ncol = 2))
+ Tp <- kronecker(matrix(c(1,1,0,n1),nrow = 2, ncol
+ = 2),Tp1)
+ Lambda <- Theta %*% t(Tp)
+ Lambdai <- 1/ Lambda
+ Tpi <- solve(Tp)
+ M <- Tpi %*% t(Lamdai)
+ D1 <- diag(n1*n2*n3)*M[1,1]
+ D2 <- (kronecker ((diag(n1*n2)),matrix( 1, ncol=n3,
+ nrow=n3)))* M[2,1]
+ D3 <-kronecker(diag(n1), (kronecker(matrix( 1,
+ ncol=n2, nrow=n2),(diag(n3))))) * M[3,1]
+ D4 <- kronecker(matrix( 1, ncol=n1, nrow=n1) ,
+ (matrix( 1, ncol=(n2*n3), nrow=(n2*n3)))* M[4,1]
+ D5 <- kronecker(matrix( 1, ncol=(n1*n2),
+ nrow=(n1*n2)) , (matrix( 1, ncol=n3, nrow=n3)))*
+ M[5,1]
+ D6 <-kronecker(matrix( 1, ncol=n1, nrow=n1),
+ (kronecker(diag(n2),matrix( 1, ncol=n3, nrow=n3)))*
+ M[6,1]
+ D7 <- (kronecker ((diag(n2)),matrix( 1, ncol=(n2*n3),
+ nrow=(n2*n3)))* M[7,1]
+ D8 <- matrix ( 1, ncol=(n1*n2*n3), nrow=(n1*n2*n3))*
+ M[8,1]
+ D1+D2+D3+D4+D5+D6+D7+D8
+ }
```

Así, para determinar la matriz V^{-1} , se utiliza la instrucción:

```
>Vinv_3vias (Se,S1,S2,S3,n1,n2,n3)
```

Donde: Se corresponde a σ_e^2 , S1 a $\sigma_{\tau_1}^2$, S2 a $\sigma_{\tau_2}^2$, S3 a $\sigma_{\tau_3}^2$, n1, n2 y n3 corresponden a los niveles de τ^1 , τ^2 y τ^3 , respectivamente.

Discusión

La estimación de componentes de varianza es útil en la identificación de fuentes de variación. Existen varios métodos de estimación de componentes de varianza. El problema de la estimación de los componentes de varianza se ha abordado históricamente desde dos perspectivas diferentes: estructuras balanceadas y desbalanceadas de datos. Para la primera, la estimación recae exclusivamente en el método propuesto por Henderson, mientras que para la segunda, existen varios procedimientos, entre ellos, tres extensiones distintas del método utilizado para datos balanceados. Se debe distinguir entre estas dos estructuras puesto que ambas se basan en la posibilidad que existe de encontrar formas cuadráticas no definidas positivas en una estructura desbalanceada de datos.

Después de tener un algoritmo que permite deducir la matriz de varianzas y covarianzas y la correspondiente inversa, resultan de gran utilidad las funciones realizadas en R, mediante las cuales se pueden llevar a cabo múltiples aplicaciones. La discusión entorno a este tema queda abierta y este trabajo permite explorar diferentes tipos de modelos con estructuras diferentes a los aquí expuestos.

Aplicación del modelo

González (9), evalúa el efecto del agua residual urbana con diferentes niveles de tratamiento y laminas de riego sobre las características físicas, químicas y carga bacteriológica de un suelo agrícola. Los tratamientos se establecieron a través del diseño completamente al azar con arreglo combinatorio factorial 4x3, correspondiente a 12 tratamientos. Los factores fueron calidad del agua (suelo regado con agua cruda, suelo regado con agua tratada a nivel primario, suelo regado con agua tratada a nivel secundario y suelo regado con agua blanca), y el segundo factor fue las laminas de riego 8,8 10 mm/día. El autor realizó un análisis de varianza a cada una de las variables evaluadas para determinar el efecto de los tratamientos. Para esta aplicación se tomó el análisis de varianza para el contenido de arcilla en el suelo regado con agua residual. De esta forma los componentes de varianza están dados por:

$$\theta^0 = [18.20; 28.42; 8.53; 0] \quad [28]$$

Entonces aplicando la función V_2vias (Para este caso $V_2vias(18.20, 28.4, 8.53, 4.3)$), se obtiene la matriz de varianzas y covarianzas (V) dada por:

55.15	28.42	28.42	8.53	0.00	0.00	8.53	0.00	0.00	8.53	0.00	0.00
28.42	55.15	28.42	0.00	8.53	0.00	0.00	8.53	0.00	0.00	8.53	0.00
28.42	28.42	55.15	0.00	0.00	8.53	0.00	0.00	8.53	0.00	0.00	8.53
8.53	0.00	0.00	55.15	28.42	28.42	8.53	0.00	0.00	8.53	0.00	0.00
0.00	8.53	0.00	28.42	55.15	28.42	0.00	8.53	0.00	0.00	8.53	0.00
0.00	0.00	8.53	28.42	28.42	55.15	0.00	0.00	8.53	0.00	0.00	8.53
8.53	0.00	0.00	8.53	0.00	0.00	55.15	28.42	28.42	8.53	0.00	0.00
0.00	8.53	0.00	0.00	8.53	0.00	28.42	55.15	28.42	0.00	8.53	0.00
0.00	0.00	8.53	0.00	0.00	8.53	28.42	28.42	55.15	0.00	0.00	8.53
8.53	0.00	0.00	8.53	0.00	0.00	8.53	0.00	0.00	55.15	28.42	28.42
0.00	8.53	0.00	0.00	8.53	0.00	0.00	8.53	0.00	28.42	55.15	28.42
0.00	0.00	8.53	0.00	0.00	8.53	0.00	0.00	8.53	28.42	28.42	55.15

[29]

De igual forma, aplicando la función $Vinv_2vias$ se obtiene la inversa (V^{-1}) dada por:

0,0337	-0,0123	-0,0123	-0,0062	0,0028	0,0028	-0,0062	0,0028	0,0028	-0,0062	0,0028	0,0028
-0,0123	0,0337	-0,0123	0,0028	-0,0062	0,0028	0,0028	-0,0062	0,0028	0,0028	-0,0062	0,0028
-0,0123	-0,0123	0,0337	0,0028	0,0028	-0,0062	0,0028	0,0028	-0,0062	0,0028	0,0028	-0,0062
-0,0062	0,0028	0,0028	0,0337	-0,0123	-0,0123	-0,0062	0,0028	0,0028	-0,0062	0,0028	0,0028
0,0028	-0,0062	0,0028	-0,0123	0,0337	-0,0123	0,0028	-0,0062	0,0028	0,0028	-0,0062	0,0028
0,0028	0,0028	-0,0062	-0,0123	-0,0123	0,0337	0,0028	0,0028	-0,0062	0,0028	0,0028	-0,0062
-0,0062	0,0028	0,0028	-0,0062	0,0028	0,0028	0,0337	-0,0123	-0,0123	-0,0062	0,0028	0,0028
0,0028	-0,0062	0,0028	0,0028	-0,0062	0,0028	-0,0123	0,0337	-0,0123	0,0028	-0,0062	0,0028
0,0028	0,0028	-0,0062	0,0028	0,0028	-0,0062	-0,0123	-0,0123	0,0337	0,0028	0,0028	-0,0062
0,0028	0,0028	-0,0062	0,0028	0,0028	-0,0062	0,0028	0,0028	0,0337	-0,0123	-0,0123	-0,0123
0,0028	-0,0062	0,0028	0,0028	-0,0062	0,0028	0,0028	-0,0062	0,0028	-0,0123	0,0337	-0,0123
0,0028	0,0028	-0,0062	0,0028	0,0028	-0,0062	0,0028	0,0028	-0,0062	-0,0123	-0,0123	0,0337

[30]

Conclusiones

La metodología expuesta permite agilizar el desarrollo y solución de algunos problemas prácticos mediante la estimación de la matriz de varianzas y covarianzas. El método propuesto puede ser aplicado a modelos mixtos con efectos fijos o aleatorios con cualquier número de factores y mediante la aplicación realizada en R se logran implementar los cálculos de manera óptima y eficiente. En este tipo de modelo es muy importante determinar los mejores estimadores lineales insesgados (MELI), y los mejores predictores lineales (BLUP), con el objeto de realizar inferencia puntual y por intervalos para tomar decisiones apropiadas en muchos problemas de aplicaciones. Como

los MELIS y los BLUPS dependen del conocimiento de la estructura apropiada de la matriz de varianzas y covarianzas y de su inversa, el algoritmo expuesto permite optimizar el cálculo de estos estimadores.

Financiación

Este trabajo no tuvo ningún tipo de financiación externa y fue realizado dentro de las labores académicas de los autores.

Conflicto de intereses

No existe conflicto de intereses.

Referencias

1. Moya LM. Obtención de la matriz de varianzas y covarianzas a través de los productos Kronecker para modelos balanceados. *Universitas Scientiarum* 2003; **8** (2): 39-44.
2. Carmona F. *Modelos lineales*. Volumen 3 de Electronic-university mathematical books. Editor Publicacions i Edicions, Universitat de Barcelona, Barcelona, España. 2005, 266 p.
3. MacCulloch, C.E. *Generalized mixed models*. Volume 3. NSF-CMBS Regional Conference Series in Probability and Statistics. 2003, 84 p.
4. Sudhir P, Saha K. The generalized linear model and extensions: A review and some biological and environmental applications. *Environmetrics* 2007; **18**:421-443.
5. Hartley HO, Rao JNK. Maximum likelihood estimation for the mixed analysis of variance models. *Biometrika* 1967; **54**: 93-108.
6. Searle SR. Built in restrictions on best linear unbiased predictors (BLUP) of random effects in mixed models. *The American Statistician* 1997; **51**: 19-21.
7. Crawley M. *Biodiversidad statistics: an introduction using R*. Primera edición. John Wiley & Sons. London. England. 2005, 327 p.
8. Dalgaard P. *Introductory statistics with R*. Primera edición. Springer. New York, USA. 2008, 363 p.
9. González J. Efecto de la adición de agua residual urbana sobre las características de un suelo Agrícola. **Tesis Doctoral**, área: Biotecnología. Universidad de Colima, Tecomán, México, 2003, 133 p.