

ESTIMACIÓN POR MÁXIMA VEROSIMILITUD RESTRINGIDA DE COMPONENTES DE VARIANZA Y COVARIANZA DE MÚLTIPLES CARACTERÍSTICAS BAJO LOS DISEÑOS I Y II DE CAROLINA DEL NORTE

RESTRICTED MAXIMUM LIKELIHOOD ESTIMATION OF VARIANCE AND COVARIANCE COMPONENTS OF MULTIPLE TRAITS UNDER DESIGNS I AND II OF NORTH CAROLINA

Jaime Galán Caballero¹, Efraín Jiménez Pablo¹ y Cuauhtémoc Cervantes Martínez^{2*}

¹ Departamento de Estadística, Universidad Autónoma Chapingo. Km. 38.5 Carr. México-Texcoco. C.P.56230. Chapingo, Estado de México. ² División de Ciencias Forestales. Correo electrónico: c_cervan@taurus1.chapingo.mx Tel: 01(595) 952-1500 Ext. 5492.

* Autor responsable

RESUMEN

En el presente estudio se describe el método de estimación por máxima verosimilitud restringida (REML) de componentes de varianza y covarianza con los Diseños I y II de Carolina del Norte, cuando dos o más variables son observadas en cada unidad experimental. Aunque REML es altamente demandante de recursos computacionales, los equipos modernos disponibles permiten su implementación. REML se basó en el desarrollo de los modelos mixtos para múltiples características para ambos diseños, al considerar la misma matriz diseño para todas las variables y conjuntos de datos balanceados. Junto con las ecuaciones de máxima verosimilitud restringida y las matrices de varianza-covarianza, se obtuvieron los estimadores de REML de la correlación fenotípica y genotípica de cada diseño, y sus varianzas aproximadas según el método Delta. Los modelos desarrollados en este trabajo contribuyen, como parte de la base teórica, a implementar la estimación de REML de caracteres múltiples en el mejoramiento genético vegetal.

Palabras clave: Diseños genéticos, mejoramiento genético, efectos aleatorios, método de momentos, correlación fenotípica, correlación genotípica, método delta.

SUMMARY

A method for estimating variance and covariance components by restricted maximum likelihood (REML) under Designs I and II of North Carolina when two or more variables are observed on each experimental unit, is described in this study. Even though REML requires a large amount of computational resources, it can be implemented given the current availability in modern equipments. REML is based on the mixed models development for multiple trait situations for both mating designs, considering the same matrix design for all traits and balanced data. In addition to the restricted maximum likelihood equations and the variance-covariance matrices, the REML estimators of the phenotypic and genotypic correlations, and their approximated variances by the Delta method are developed. The models described in this study are part of the theoretical basis for the multiple trait REML estimation in plant breeding.

Index words: Genetic designs, plant breeding, random effects, moments method, phenotypic correlation, genotypic correlation, delta method.

INTRODUCCIÓN

Los modelos estadísticos lineales en lo que algunos de los componentes son de efectos fijos y otros de efectos aleatorios, se conocen como modelos mixtos (Searle, 1971). Uno de los objetivos del análisis de los modelos mixtos es el de estimar las varianzas de cada uno de los efectos aleatorios, así como sus covarianzas, conocidas como componentes de varianza y covarianza (Corbeil y Searle, 1976; Harville, 1977).

La estimación de componentes de varianza y covarianza se puede realizar por los métodos de momentos (ANOVA), estimación cuadrática insesgada de mínima varianza (MIVQUE), máxima verosimilitud (ML) y máxima verosimilitud restringida (REML) (Swallow y Monahan, 1984; Meyer, 1985). Tradicionalmente, los mejoradores de plantas han estimado los componentes de varianza basados en el análisis de varianza de mínimos cuadrados (ANOVA), igualando los cuadrados medios observados a sus esperanzas matemáticas y resolviendo el sistema de ecuaciones resultantes para los diferentes componentes de varianza (Milliken y Johnson, 1992). Los estimadores basados en el método ANOVA aplicado a conjuntos de datos balanceados poseen muchas propiedades interesantes: son insesgados, de mínima varianza entre todos los estimadores insesgados que son funciones cuadráticas de las observaciones, y bajo normalidad, de mínima varianza entre todos los estimadores insesgados (Searle, 1971). Sin embargo, si el conjunto de datos es desbalanceado, como sería el caso de tener observaciones perdidas, pierden todas sus propiedades excepto la de insesgamiento (Searle *et al.*, 1992). Henderson (1953) desarrolló tres métodos análogos al ANOVA para datos desbalanceados; de éstos, el Método 3 ha sido utilizado más ampliamente, pero desafortunadamente la teoría relativa a sus propiedades estadísticas es muy poco conocida (Harville, 1977). Para realizar la

estimación por el método de MIVQUE tienen que suministrarse valores para los componentes de varianza. Los estimadores son entonces funciones de los datos y de los valores previos, y son de varianza mínima solamente cuando los valores previos son iguales a los valores verdaderos de los componentes de varianza (Swallow y Monahan, 1984).

Los estimadores de los componentes de varianza y covarianza obtenidos por los métodos de ML y REML tienen propiedades asintóticas que los hacen preferibles sobre los estimadores obtenidos con otros métodos. Cuando el tamaño de muestra es considerablemente “grande” los estimadores de ML y REML son consistentes, normalmente asintóticos, eficientes y no requieren de conjuntos de datos balanceados para mantener estas propiedades; además, la matriz asintótica de varianzas y covarianzas de los estimadores es conocida, lo que permite el establecimiento de intervalos de confianza y pruebas de hipótesis acerca de los componentes de varianza y covarianza (Searle *et al.*, 1992). Para tamaños de muestra “pequeños” los estimadores de máxima verosimilitud son generalmente sesgados, puesto que al estimar los efectos aleatorios no se toma en cuenta la pérdida en grados de libertad que resulta de la estimación de los efectos fijos (Meyer, 1985; Searle *et al.*, 1992). Para solucionar este problema, el método REML utiliza combinaciones lineales de los elementos del vector de datos, de tal manera que esas combinaciones no contienen efectos fijos. Esas combinaciones generan una función de verosimilitud que no depende de los efectos fijos (Searle *et al.*, 1992). Posteriormente la función de verosimilitud restringida se maximiza con respecto a cada uno de los componentes de varianza y covarianza. Los valores que maximizan la función son los respectivos estimadores de REML de los componentes de varianza y covarianza (Corbeil y Searle, 1976; Swallow y Monahan, 1984).

Generalmente, las ecuaciones generadas por los métodos basados en máxima verosimilitud no pueden ser resueltas analíticamente para los diferentes componentes de varianza (Corbeil y Searle, 1976), por lo que es necesario obtener las soluciones por medio de métodos iterativos que son altamente demandantes de recursos computacionales. Con el actual desarrollo de la tecnología y de nuevos algoritmos computacionales, esto ya no es un impedimento para realizar dicha estimación (Miller, 1977). Por todas las propiedades mencionadas anteriormente la estimación basada en la máxima verosimilitud está tomando nuevamente un resurgimiento en la investigación científica.

Una extensión natural del modelo lineal mixto es su aplicación en el análisis multivariado para la estimación de componentes de covarianza entre múltiples características (Meyer, 1985), que tradicionalmente se ha realizado a tra-

vés del análisis de varianza multivariado (MANOVA) (Harville, 1977; Meyer, 1985). Los estimadores de REML de covarianza entre características tienen las mismas propiedades asintóticas que los estimadores de componentes de varianza basados en máxima verosimilitud. El método de REML para estimar varianzas y covarianzas entre características múltiples, aplicado a mejoramiento genético animal, fue desarrollado por Schaeffer *et al.* (1978) y Meyer (1985).

Los Diseños I y II de Carolina del Norte juegan un papel primordial en el mejoramiento genético vegetal para estimar los componentes de varianza genética de una población de referencia (Fehr, 1991). La medición de múltiples características bajo estos diseños permite también la estimación de componentes de covarianza, y las correlaciones fenotípicas y genotípicas entre las características medidas (Hallauer y Miranda, 1988). Tal estimación se ha realizado por MANOVA, puesto que no se han desarrollado los modelos mixtos multivariados para estos diseños.

En el presente trabajo se plantea el soporte teórico de la estimación de componentes de varianza y covarianza, y las correlaciones fenotípicas y genotípicas de múltiples características bajo los Diseños I y II de Carolina del Norte, desarrollados por Comstock y Robinson (1948), a través del método de máxima verosimilitud restringida, que actualmente es factible de ser implementada con los recursos computacionales modernos.

MARCO TEÓRICO

Considérese un vector de observaciones dado por $\underline{y}' = (\underline{y}'_1, \underline{y}'_2, \dots, \underline{y}'_t)$, donde cada subvector \underline{y}'_i corresponde a las observaciones de la i -ésima variable respuesta, donde $i = 1, 2, \dots, t$, de las progenies evaluadas en varios ambientes con el Diseño I o el Diseño II de Carolina del Norte, según sea el caso, que generan un conjunto de datos totalmente balanceados con las observaciones obtenidas. Supóngase que cada variable de respuesta tiene asociado el siguiente modelo lineal aleatorio:

$$\underline{y}_i = \underline{1}_{n_i} \mu_i + Z_i \gamma_i \hat{P} \quad (i=1, 2, \dots, t) \quad (\text{Ec. 1})$$

$$\text{con } E(\underline{y}_i) = \underline{1}_{n_i} \mu_i, E(\gamma_i) = 0, \quad \text{Var}(\gamma_i) = \bigoplus_{s=1}^f \mathbf{I}_{t_s} \sigma_{ii_s}$$

$$\text{y } \text{Var}(\underline{y}_i) = \sum_{s=1}^f Z_{i_s} Z'_{i_s} \sigma_{ii_s}$$

donde $\underline{1}_{n_i}$ es un vector de unos de orden $n_i \times 1$; n_i el número de observaciones de la i -ésima característica; μ_i la media de la característica i ; $\underline{\gamma}_i$ un vector que contiene los factores aleatorios del modelo e incluyen el vector de errores; $Z_i = [Z_{i1} Z_{i2} \dots Z_{if}]$ la matriz diseño de la i -ésima característica; I_{l_s} una matriz identidad de orden l_s , con l_s como el número de niveles del factor s , $s = 1, 2, \dots, f$; σ_{ij_s} la varianza del factor s para la i -ésima característica. El símbolo \oplus denota la suma directa de matrices (Searle, 1982). Si todas las variables tienen asociado el mismo modelo lineal con igual número de observaciones, entonces $\underline{1}_{n_i} = \underline{1}_n$ y $Z_i = Z$ para toda i , y el modelo multivariado puede escribirse como:

$$\underline{y} = (I_t \otimes \underline{1}_n) \underline{\mu} + (I_t \otimes Z) \underline{\gamma} \quad (\text{Ec. 2})$$

donde I_t es una matriz identidad de orden t , $\underline{\mu}' = (\mu_1, \mu_2, \dots, \mu_t)$ y $\underline{\gamma}' = (\gamma'_1, \gamma'_2, \dots, \gamma'_t)$. El símbolo \otimes denota el producto directo de matrices (Searle, 1982).

Además de los supuestos del modelo (1), se hacen los siguientes supuestos para el modelo en (2),

$$\text{Cov}(\underline{\gamma}_i, \underline{\gamma}'_j) = \bigoplus_{s=1}^f I_{l_s} \sigma_{ij_s}, \quad E(\underline{y}) = (I_t \otimes \underline{1}_n) \underline{\mu}, \quad y$$

$$\text{Var}(\underline{y}) = \mathbf{V}_M = \left\{ m \sum_{s=1}^f Z_{i_s} Z'_{i_s} \sigma_{ij_s} \right\}$$

para $i, j = 1, 2, \dots, t$

El símbolo m denota a una matriz con las entradas especificadas y σ_{ij_s} es el componente de covarianza de los caracteres i y j para el factor s .

Los estimadores de máxima verosimilitud de los parámetros σ_{ij_s} 's, son el conjunto de valores de los parámetros que maximizan la verosimilitud de obtener el conjunto de datos. Estos estimadores son obtenidos al maximizar la función de densidad conjunta de la muestra vista como una función de los parámetros. A dicha función se le llama función de verosimilitud. La estimación por máxima verosimilitud restringida (REML) es una modificación al método

anterior para obtener estimadores con menor sesgo. El método REML primero estima los efectos fijos del modelo y luego maximiza la función de verosimilitud de los residuales de n^* contrastes ortogonales, donde n^* es el número de grados de libertad remanente después de ajustar los efectos fijos del modelo (Lynch y Walsh, 1998). El logaritmo natural de la función de verosimilitud restringida y la función de verosimilitud restringida tienen los mismos puntos máximos, pero la primera es más fácil de maximizar.

En condiciones de normalidad de los factores de efectos aleatorios, el logaritmo natural de la función de verosimilitud restringida del vector \underline{y} en Ec. 2 es:

$$l_R \log(L) = -\frac{(tn-t)}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{K}'_M \mathbf{V}_M \mathbf{K}_M| - \frac{1}{2} \underline{y}' \mathbf{K}_M (\mathbf{K}'_M \mathbf{V}_M \mathbf{K}_M)^{-1} \mathbf{K}'_M \underline{y} \quad (\text{Ec. 3})$$

donde la matriz \mathbf{K}'_M se obtiene al eliminar el n -ésimo renglón, el $2n$ -ésimo renglón, \dots , y el tn -ésimo renglón de la matriz $\mathbf{S} = I_t \otimes \left(I_n - \frac{1}{n} \mathbf{J}_n \right)$ (Corbeil y Searle, 1976).

Las derivadas parciales del logaritmo de la función de verosimilitud restringida (L) en Ec. 3, con respecto a los componentes de varianza y covarianza, son:

$$\frac{\partial l_R}{\partial \sigma_{ij_s}} = -\frac{1}{2} \text{tr} \left\{ \mathbf{P}_M (\mathbf{E}_{ij} \otimes Z_{i_s} Z'_{i_s}) \right\} + \frac{1}{2} \underline{y}' \mathbf{P}_M (\mathbf{E}_{ij} \otimes Z_{i_s} Z'_{i_s}) \mathbf{P}_M \underline{y} \quad (\text{Ec. 4})$$

y las segundas derivadas resultan en:

$$\frac{\partial^2 l_R}{\partial \sigma_{ij_s} \partial \sigma_{i'j'_s}} =$$

$$\frac{1}{2} \text{tr} \{ \mathbf{P}_M (\mathbf{E}_{ij} \otimes \mathbf{Z}_{i_s} \mathbf{Z}_{i_s}') \mathbf{P}_M (\mathbf{E}_{i'j'} \otimes \mathbf{Z}_{i'_s} \mathbf{Z}_{i'_s}') \}$$

$$- \underline{y}' \mathbf{P}_M (\mathbf{E}_{ij} \otimes \mathbf{Z}_{i_s} \mathbf{Z}_{i_s}') \mathbf{P}_M (\mathbf{E}_{i'j'} \otimes \mathbf{Z}_{i'_s} \mathbf{Z}_{i'_s}') \mathbf{P}_M \underline{y}$$

(Ec. 5)

para $s=1, 2, \dots, f$, donde la matriz $\mathbf{P}_M = \mathbf{K}_M' (\mathbf{K}_M' \mathbf{V}_M \mathbf{K}_M)^{-1} \mathbf{K}_M'$ y la matriz \mathbf{E}_{ij} es una matriz de orden $t \times t$ con el ij -ésimo elemento igual a uno y ceros en los elementos restantes (Searle *et al.*, 1992).

Los respectivos estimadores de máxima verosimilitud restringida, son los valores para las σ_{ij_s} 's que maximizan a Ec. 3, obtenidos al igualar a cero la expresión en Ec. 4 para los diferentes valores de i, j, s , y resolver las ecuaciones para los componentes de varianza y covarianza. Generalmente, las ecuaciones resultantes no pueden resolverse analíticamente, por lo que se deben emplear métodos iterativos. Estos métodos parten de un conjunto de valores iniciales de todos los parámetros, seguidos de un proceso de diferenciación como en Ec. 4 y de la estimación iterativa de parámetros que van reduciendo el valor absoluto de la primera derivada hasta obtener soluciones estables. Un ejemplo es el método de Newton-Raphson utilizado por el procedimiento MIXED de SAS (SAS Institute Inc., 1999) que se caracteriza por obtener las soluciones en pocas iteraciones, en comparación con otros métodos (Lindstrom y Bates, 1989). Si se define Θ como la matriz de parámetros a estimar, entonces el valor de la matriz en la iteración $m + 1$, según el método de Newton-Raphson, es (Searle *et al.*, 1992):

$$\Theta^{(m+1)} = \Theta^{(m)} - (\mathbf{H}^{(m)})^{-1} \frac{\partial l_R}{\partial \Theta} \bigg|_{\Theta^{(m)}}$$

donde $\Theta^{(m)}$ es el valor de la matriz de parámetros en la m -ésima iteración, $\mathbf{H}^{(m)}$ es la matriz $\mathbf{K}_M' \mathbf{V}_M \mathbf{K}_M$ con los

elementos σ_{ij_s} remplazados por $\frac{\sigma_{ij_s}^{(m)}}{\sigma_{ij_f}^{(m)}}$, para $s = 1, 2,$

\dots, f . Finalmente, $\frac{\partial l_R}{\partial \Theta} \bigg|_{\Theta^{(m)}}$ es la matriz de derivadas

evaluadas en la m -ésima iteración y cuyos elementos se muestran en la Ec. 4.

En el presente trabajo se presentará la forma de las matrices \mathbf{Z}_{i_s} 's y \mathbf{V}_M para los Diseños I y II de Carolina del Norte requeridas para resolver las ecuaciones de verosimilitud restringida.

Si ahora se considera a $\hat{\underline{\theta}}$ como el vector que contiene todos los estimadores de REML de componentes de varianza y covarianza, entonces la $\text{Var}(\hat{\underline{\theta}})$ para muestras grandes es la inversa de la matriz de información de Fisher: $\mathbf{I}_{\underline{\theta}} = E \left[- \frac{\partial^2 l_R}{\partial \underline{\theta} \partial \underline{\theta}'} \right]$, cuyas segundas derivadas se presentan en (5). Los estimadores de los componentes de la matriz $\text{Var}(\hat{\underline{\theta}})$ se obtienen substituyendo los componentes de varianza y covarianza de $\mathbf{I}_{\underline{\theta}}$ por los estimadores que maximizan la Ec. 3, es decir:

$$\hat{\text{Var}}(\hat{\underline{\theta}}) \approx 2 \left\{ \text{tr} \left[\hat{\mathbf{P}}_M (\mathbf{E}_{ij} \otimes \mathbf{Z}_{i_s} \mathbf{Z}_{i_s}') \right] \right. \\ \left. \hat{\mathbf{P}}_M (\mathbf{E}_{i'j'} \otimes \mathbf{Z}_{i'_s} \mathbf{Z}_{i'_s}') \right\}^{-1}$$

(Ec. 6)

para $i, j = 1, 2, \dots, t$ y $s = 1, 2, \dots, f$ (Searle *et al.*, 1992).

Por otra parte, si $g(\underline{\theta})$ es una función real del vector de parámetros $\underline{\theta}$, su estimador de REML puede obtenerse como $g(\hat{\underline{\theta}})$, con varianza aproximada según el método Delta:

$$\text{Var} \left[g(\hat{\underline{\theta}}) \right] \approx \left[\frac{\partial g(\underline{\theta})}{\partial \underline{\theta}} \right]' \text{Var}(\hat{\underline{\theta}}) \left[\frac{\partial g(\underline{\theta})}{\partial \underline{\theta}} \right]$$

(Ec. 7)

donde $\left[\frac{\partial g(\underline{\theta})}{\partial \underline{\theta}} \right]$ es el vector gradiente de las primeras derivadas parciales de $g(\underline{\theta})$ con respecto a $\underline{\theta}$ y evaluadas en $\hat{\underline{\theta}}$ (Meeker y Escobar, 1998). Los estimadores de REML de los coeficientes de correlación fenotípica y genotípica se presentan como funciones $g(\hat{\underline{\theta}})$ y sus varianzas aproximadas como se muestra en la Ec. 7.

RESULTADOS: ESTIMADORES DE PARÁMETROS POR EL MÉTODO REML

Diseño I de Carolina del Norte

Bajo el supuesto que “*m*” machos fueron cruzados con diferentes grupos de “*h*” hembras de una población de referencia dada y las progenies se evaluaron en un diseño experimental con “*r*” repeticiones en “*e*” ambientes, la matriz diseño $Z_i = [Z_{i1} \ Z_{i2} \ \dots \ Z_{i7}]$ del modelo lineal en (1) correspondiente al Diseño I de Carolina del Norte asociado a cada característica, tiene las siguientes submatrices: $Z_{i1} = \mathbf{I}_e \otimes \mathbf{1}_{mhr}$, que es la matriz diseño del factor ambientes; $Z_{i2} = \mathbf{1}_e \otimes (\mathbf{I}_m \otimes \mathbf{1}_{hr})$ es la matriz diseño del factor machos; $Z_{i3} = \mathbf{1}_e \otimes (\mathbf{I}_{mh} \otimes \mathbf{1}_r)$ es la matriz diseño del factor hembras anidadas en machos; $Z_{i4} = \mathbf{I}_e \otimes (\mathbf{I}_m \otimes \mathbf{1}_{hr})$ es la matriz diseño del factor de interacción entre ambientes y machos; $Z_{i5} = \mathbf{I}_e \otimes (\mathbf{I}_{mh} \otimes \mathbf{1}_r)$ es la matriz diseño del factor de interacción entre ambientes y hembras anidadas en machos; $Z_{i6} = \mathbf{I}_e \otimes (\mathbf{1}_{mh} \otimes \mathbf{I}_r)$ es la matriz diseño del factor repeticiones anidado en ambientes, y $Z_{i7} = \mathbf{I}_{emhr}$ es la matriz diseño de los errores aleatorios asociados a cada observación. El modelo multivariado para dos características ($t=2$) se puede escribir como:

$$\underline{y}_M = \begin{bmatrix} \underline{y}_i \\ \underline{y}_j \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{emhr} & 0 \\ 0 & \mathbf{1}_{emhr} \end{bmatrix} \begin{bmatrix} \underline{\mu}_i \\ \underline{\mu}_j \end{bmatrix} + \begin{bmatrix} Z_i & 0 \\ 0 & Z_j \end{bmatrix} \begin{bmatrix} \underline{\gamma}_i \\ \underline{\gamma}_j \end{bmatrix}$$

$$= \mathbf{X}_M \underline{\mu} + \mathbf{Z}_M \underline{\gamma}_M \quad (\text{Ec. 8})$$

donde las matrices Z_i y Z_j son las matrices diseño de los factores aleatorios para las características “*i*” y “*j*”; $\underline{\mu}_i$ y $\underline{\mu}_j$, y $\underline{\gamma}_i$ y $\underline{\gamma}_j$ son las medias y los vectores que contienen los factores aleatorios de las variables respuesta “*i*” y “*j*”, respectivamente.

La matriz de varianzas y covarianzas del vector de observaciones multivariado del modelo en la Ec. 8 es:

$$\text{Var}(\underline{y}) = \mathbf{V}_M = \left\{ \sum_{s=1}^7 Z_{is} Z'_{is} \sigma_{ij_s} \right\}$$

$$= \left\{ \sigma_{ij_E} (\mathbf{I}_e \otimes \mathbf{1}_{mhr} \mathbf{1}'_{mhr}) + \sigma_{ij_m} (\mathbf{1}_e \mathbf{1}'_e \otimes \mathbf{I}_m \otimes \mathbf{1}_{hr} \mathbf{1}'_{hr}) + \sigma_{ij_{hm}} (\mathbf{1}_e \mathbf{1}'_e \otimes \mathbf{I}_{mh} \otimes \mathbf{1}_r \mathbf{1}'_r) + \sigma_{ij_{Em}} (\mathbf{I}_e \otimes \mathbf{I}_m \otimes \mathbf{1}_{hr} \mathbf{1}'_{hr}) + \sigma_{ij_{Ehm}} (\mathbf{I}_e \otimes \mathbf{I}_{mh} \otimes \mathbf{1}_r \mathbf{1}'_r) + \sigma_{ij_{r/E}} (\mathbf{I}_e \otimes \mathbf{1}_{mh} \mathbf{1}'_{mh} \otimes \mathbf{I}_r) + \sigma_{ij} \mathbf{I}_{emhr} \right\}$$

para $i, j = 1, 2$, donde σ_{ii_s} es la varianza de la característica “*i*” del factor *s*, y σ_{ij_s} es la covarianza de las características “*i*” y “*j*” del factor *s*.

La matriz de varianzas y covarianzas $\mathbf{K}'_M \mathbf{V}_M \mathbf{K}_M$ de la función de verosimilitud restringida para el modelo multivariado del Diseño I de Carolina del Norte es:

$$\begin{aligned}
\mathbf{K}'_M \mathbf{V}_M \mathbf{K}_M &= \left\{ \begin{matrix} m \\ i=1 \end{matrix} \sum \mathbf{K}'_i \mathbf{Z}'_i \mathbf{Z}_i \mathbf{K}_i \sigma_{ij_s} \right\} \\
&= \left\{ \begin{matrix} m \\ i=1 \end{matrix} \sigma_{ij_E} \left[[(\mathbf{I}_{e-1} \otimes \mathbf{J}_{mhr}) \oplus \mathbf{J}_{mhr-1}] - \frac{1}{e} \mathbf{J}_{emhr-1} \right] \right. \\
&\quad + \sigma_{ij_m} \left[[(\mathbf{J}_{e-1} \otimes [(\mathbf{I}_m \otimes \mathbf{J}_{hr})]) \oplus [(\mathbf{I}_{m-1} \otimes \mathbf{J}_{hr}) \right. \\
&\quad \left. \left. \oplus \mathbf{J}_{hr-1} \right]] - \frac{1}{m} \mathbf{J}_{emhr-1} \right] + \sigma_{ij_{hm}} \left[[\mathbf{J}_{e-1} \otimes [(\mathbf{I}_{mh} \otimes \right. \\
&\quad \left. \left. \mathbf{J}_r) \right]] \oplus [(\mathbf{I}_{mh-1} \otimes \mathbf{J}_r) \oplus \mathbf{J}_{r-1}] \right] - \frac{1}{mh} \mathbf{J}_{emhr-1} \left. \right] \\
&\quad + \sigma_{ij_{Em}} \left[[(\mathbf{I}_{em-1} \otimes \mathbf{J}_{hr}) \oplus \mathbf{J}_{hr-1}] - \frac{1}{em} \mathbf{J}_{emhr-1} \right] \\
&\quad + \sigma_{ij_{Ehm}} \left[[(\mathbf{I}_{emh-1} \otimes \mathbf{J}_r) \oplus \mathbf{J}_{r-1}] - \frac{1}{emh} \mathbf{J}_{emhr-1} \right] \\
&\quad + \sigma_{ij_{r/E}} \left[[(\mathbf{I}_{e-1} \otimes [(\mathbf{J}_{mh} \otimes \mathbf{I}_r)]) \oplus \right. \\
&\quad \left. [(\mathbf{J}_{mh-1} \otimes \mathbf{I}_r) \oplus \mathbf{I}_{r-1}] \right] - \frac{1}{er} \mathbf{J}_{emhr-1} \left. \right] \\
&\quad + \sigma_{ij}^2 \left[\mathbf{I}_{emhr-1} - \frac{1}{emhr} \mathbf{J}_{emhr-1} \right] \left. \right\}, \text{ para } i, j = 1, 2.
\end{aligned}$$

(Ec. 9)

La matriz $\mathbf{K}'_M \mathbf{V}_M \mathbf{K}_M$ de la Ec. 9 es requerida para obtener las derivadas parciales de la función de verosimilitud restringida del modelo multivariado del Diseño I de Carolina del Norte, con respecto a los componentes de varianza y covarianza, como es mostrado en las ecuaciones 4 y 5.

Sea $\hat{\Sigma}$ la matriz diagonal que contiene los estimadores de REML de los componentes de varianza y covarianza para las características “i” y “j” de los siete factores:

$$\hat{\Sigma} = \text{Diag}(\hat{\theta}) = \text{Diag}\{\text{Diag}(\hat{\sigma}_{ii_s}, \hat{\sigma}_{ij_s}, \hat{\sigma}_{jj_s})\}_{21 \times 21}$$

para $s=1, 2, \dots, 7$

Si se consideran ausencia de endogamia, de epistasis, de efectos maternos y equilibrio de ligamiento se tiene que

$$\sigma_{ij_m} = \frac{1}{4} \sigma_{ij_A}, \quad \sigma_{ij_{h/m}} = \frac{1}{4} (\sigma_{ij_D} + \sigma_{ij_A}), \quad \sigma_{ij_{Em}} = \frac{1}{4} \sigma_{ij_{AE}}$$

y

$$\sigma_{ij_{Eh/m}} = \frac{1}{4} (\sigma_{ij_{DE}} + \sigma_{ij_{AE}}), \text{ donde } \sigma_{ij_{AE}} \text{ y } \sigma_{ij_{DE}}$$

son las covarianzas de las características “i” y “j” de la interacción entre los efectos aditivos (A) y de dominancia (D) con los efectos ambientales (E). Los estimadores de las varianzas fenotípicas ($\hat{\sigma}_{ii_p}$ y $\hat{\sigma}_{jj_p}$) para las variables respuesta “i” y “j” y el estimador de la covarianza fenotípica ($\hat{\sigma}_{ij_p}$) entre éstas, se obtienen en forma matricial al definir los vectores:

$$\begin{aligned}
\underline{c}'_{ii} &= [100000200000200000000] \\
\underline{c}'_{ij} &= [010000020000020000000] \\
\underline{c}'_{jj} &= [001000002000002000000]
\end{aligned}$$

y sustituyéndolos en las formas cuadráticas

$$\hat{\sigma}_{ij_p} = \underline{c}'_{ij} \hat{\Sigma} \underline{c}_{ij} = 4\hat{\sigma}_{ij_{h/m}} + 4\hat{\sigma}_{ij_{Eh/m}} + \hat{\sigma}_{E_{ij}},$$

$$\hat{\sigma}_{ii_p} = \underline{c}'_{ii} \hat{\Sigma} \underline{c}_{ii} = 4\hat{\sigma}_{ii_{h/m}} + 4\hat{\sigma}_{ii_{Eh/m}} + \hat{\sigma}_{E_{ii}} \text{ y}$$

$$\hat{\sigma}_{jj_p} = \underline{c}'_{jj} \hat{\Sigma} \underline{c}_{jj} = 4\hat{\sigma}_{jj_{h/m}} + 4\hat{\sigma}_{jj_{Eh/m}} + \hat{\sigma}_{E_{jj}},$$

de tal manera que el estimador de la correlación fenotípica entre las características “i” y “j” se obtiene como:

$$\hat{\rho}_{ij_p} = \frac{\underline{c}'_{ij} \hat{\Sigma} \underline{c}_{ij}}{\left[(\underline{c}'_{ii} \hat{\Sigma} \underline{c}_{ii}) (\underline{c}'_{jj} \hat{\Sigma} \underline{c}_{jj}) \right]^{1/2}} = \frac{\hat{\sigma}_{ij_p}}{(\hat{\sigma}_{ii_p} \hat{\sigma}_{jj_p})^{1/2}} =$$

$$\frac{4\hat{\sigma}_{ij_{h/m}} + 4\hat{\sigma}_{ij_{Eh/m}} + \hat{\sigma}_{Eij}}{\sqrt{4\hat{\sigma}_{ii_{h/m}} + 4\hat{\sigma}_{ii_{Eh/m}} + \hat{\sigma}_{Eii}} \sqrt{4\hat{\sigma}_{jj_{h/m}} + 4\hat{\sigma}_{jj_{Eh/m}} + \hat{\sigma}_{Ejj}}}$$

(Ec. 10)

Los estimadores de las varianzas genéticas ($\hat{\sigma}_{ii_g}$ y $\hat{\sigma}_{jj_g}$) para las variables respuesta “i” y “j” y del estimador de la respectiva covarianza genética ($\hat{\sigma}_{ij_g}$), se obtienen en forma matricial con los vectores:

$$\underline{c}'_{ii} = [000000020000000000000000]$$

$$\underline{c}'_{ij} = [000000002000000000000000]$$

$$\underline{c}'_{jj} = [000000000200000000000000]$$

y
sustituyéndolos en las formas cuadráticas

$$\hat{\sigma}_{ij_g} = \underline{c}'_{ij} \hat{\Sigma} \underline{c}_{ij} = 4\hat{\sigma}_{ij_{h/m}}, \quad \hat{\sigma}_{ii_g} = \underline{c}'_{ii} \hat{\Sigma} \underline{c}_{ii} = 4\hat{\sigma}_{ii_{h/m}} \text{ y}$$

$\hat{\sigma}_{jj_g} = \underline{c}'_{jj} \hat{\Sigma} \underline{c}_{jj} = 4\hat{\sigma}_{jj_{h/m}}$, de tal forma que el estimador de la correlación genética entre las características “i” y “j” se obtiene como:

$$\hat{\rho}_{ij_g} = \frac{\underline{c}'_{ij} \hat{\Sigma} \underline{c}_{ij}}{\left[(\underline{c}'_{ii} \hat{\Sigma} \underline{c}_{ii}) (\underline{c}'_{jj} \hat{\Sigma} \underline{c}_{jj}) \right]^{1/2}}$$

(Ec. 11)

$$= \frac{\hat{\sigma}_{ij_g}}{\sqrt{\hat{\sigma}_{ii_g}} \sqrt{\hat{\sigma}_{jj_g}}} = \frac{4\hat{\sigma}_{ij_{h/m}}}{\sqrt{4\hat{\sigma}_{ii_{h/m}}} \sqrt{4\hat{\sigma}_{jj_{h/m}}}}$$

donde $\hat{\sigma}_{ii_{h/m}}$ y $\hat{\sigma}_{jj_{h/m}}$ son los estimadores de las varianzas para las características “i” y “j” del factor hembras anidadas en machos, y $\hat{\sigma}_{ij_{h/m}}$ es el correspondiente estimador de la covarianza para las características “i” y “j” de este factor.

Los estimadores $\hat{\rho}_{ij_p}$ y $\hat{\rho}_{ij_g}$ continúan siendo, por la propiedad de invarianza de los estimadores de REML, estimadores de máxima verosimilitud restringida de ρ_{ij_p} y ρ_{ij_g} , respectivamente.

Sea

$$\underline{\hat{\theta}}_{21 \times 1}^t = [\hat{\sigma}_{ii_1} \quad \hat{\sigma}_{ij_1} \quad \hat{\sigma}_{jj_1} \quad \cdot \quad \cdot \quad \cdot \quad \hat{\sigma}_{ii_7} \quad \hat{\sigma}_{ij_7} \quad \hat{\sigma}_{jj_7}]^t$$

el estimador correspondiente del vector de parámetros $\underline{\theta}_{21 \times 1}$. La varianza aproximada de $\hat{\rho}_{ij_p}$ según Ec. 7, es:

$$Var(\hat{\rho}_{ij_p}) \approx (\hat{\rho}_{ij_p})^2 \left[\sum_{r=1}^{21} b_r^2 Var(\hat{\theta}_r) \right]$$

(Ec. 12)

$$+ \sum_{r=1}^{21} \sum_{\substack{l=1 \\ r \neq l}}^{21} b_r b_l Cov(\hat{\theta}_r, \hat{\theta}_l) \Bigg]$$

donde:

$$b_1 = \frac{-1}{2\sigma_{ii_p}}, \quad b_2 = \frac{1}{\sigma_{ij_p}}, \quad b_3 = \frac{-1}{2\sigma_{jj_p}},$$

$$b_7 = b_{13} = \frac{-2}{\sigma_{ii_p}}, \quad b_8 = b_{14} = \frac{4}{\sigma_{ij_p}}, \quad b_9 = b_{15} = \frac{-2}{\sigma_{jj_p}},$$

y $b_r = 0$ para $r=4, 5, 6, 10, 11, 12, 16, 17, 18, 19, 20, 21$.

La aproximación a la varianza del estimador $\hat{\rho}_{ij_g}$ se obtiene como:

$$Var(\hat{\rho}_{ij_g}) \approx (\hat{\rho}_{ij_g})^2 \left[\sum_{r=1}^{21} b_r^2 Var(\hat{\theta}_r) \right]$$

(Ec. 13)

$$+ \sum_{r=1}^{21} \sum_{\substack{l=1 \\ r \neq l}}^{21} b_r b_l Cov(\hat{\theta}_r, \hat{\theta}_l) \Bigg]$$

donde:

$$b_7 = \frac{-2}{\sigma_{ii_g}}, \quad b_8 = \frac{4}{\sigma_{ij_g}}, \quad b_9 = \frac{-2}{\sigma_{jj_g}}, \text{ y } b_r = 0 \text{ para}$$

$$r=4, 5, 6, 10, 11, 12, 16, 17, 18, 19, 20, 21.$$

Los estimadores de la varianza de la correlación fenotípica y genotípica se obtienen substituyendo a los componentes de varianza y covarianza por sus estimadores, y a la $Var(\hat{\theta})$ por $\hat{Var}(\hat{\theta})$ en (6), o bien a la $Var(\hat{\theta}_r)$ y $Cov(\hat{\theta}_r, \hat{\theta}_l)$ por la $\hat{Var}(\hat{\theta}_r)$ y $\hat{Cov}(\hat{\theta}_r, \hat{\theta}_l)$ en Ecs. 12 y 13. Los elementos de la matriz $\hat{Var}(\hat{\theta})$ son obtenidos como en la Ec. 6.

Diseño II de Carolina del Norte

Para el Diseño II de Carolina del Norte se parte del supuesto que “m” machos fueron cruzados con un grupo de “h” hembras de cierta población de referencia, y las progenies se evalúan en experimentos de campo con “r” repeticiones en “e” ambientes. En este arreglo, la matriz diseño $Z_i = [Z_{i1} \ Z_{i2} \ \dots \ Z_{i9}]$ del modelo lineal en (1) tiene las siguientes submatrices: $Z_{i1} = \mathbf{I}_e \otimes \mathbf{1}_{mhr}$, que es la matriz diseño del factor ambientes, $Z_{i2} = \mathbf{1}_e \otimes (\mathbf{I}_m \otimes \mathbf{1}_{hr})$ es la matriz diseño del factor machos; $Z_{i3} = \mathbf{1}_e \otimes [\mathbf{1}_m \otimes (\mathbf{I}_h \otimes \mathbf{1}_r)]$ es la matriz diseño del factor hembras; $Z_{i4} = \mathbf{1}_e \otimes (\mathbf{I}_{mh} \otimes \mathbf{1}_r)$ es la matriz diseño de la interacción entre los factores machos y hembras; $Z_{i5} = \mathbf{I}_e \otimes (\mathbf{I}_m \otimes \mathbf{1}_{hr})$ es la matriz diseño de la interacción entre los factores ambientes y machos; $Z_{i6} = \mathbf{I}_e \otimes [\mathbf{1}_m \otimes (\mathbf{I}_h \otimes \mathbf{1}_r)]$ es la matriz diseño de la interacción entre los factores ambientes y hembras; $Z_{i7} = \mathbf{I}_e \otimes (\mathbf{I}_{mh} \otimes \mathbf{1}_r)$ es la matriz diseño de la interacción entre los factores ambientes, machos y hembras; $Z_{i8} = \mathbf{I}_e \otimes (\mathbf{1}_{mh} \otimes \mathbf{I}_r)$ es la matriz diseño del factor repeticiones anidadas en ambientes; y $Z_{i9} = \mathbf{I}_{emhr}$ es la matriz diseño de los errores de cada observación. El modelo multivariado para dos características ($t=2$) se puede escribir como:

$$\begin{aligned} \underline{y}_M &= \begin{bmatrix} y_i \\ y_j \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{emhr} & 0 \\ 0 & \mathbf{1}_{emhr} \end{bmatrix} \begin{bmatrix} \underline{\mu}_i \\ \underline{\mu}_j \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_i & 0 \\ 0 & \mathbf{Z}_j \end{bmatrix} \begin{bmatrix} \underline{\gamma}_i \\ \underline{\gamma}_j \end{bmatrix} \\ &= \mathbf{X}_M \underline{\mu} + \mathbf{Z}_M \underline{\gamma}_M \end{aligned} \quad (\text{Ec. 14})$$

donde Z_i y Z_j son las matrices diseño de los factores aleatorios para las características “i” y “j”; $\underline{\mu}_i$ y $\underline{\mu}_j$; y $\underline{\gamma}_i$ y $\underline{\gamma}_j$ son las medias y los vectores que contienen los factores aleatorios de las variables respuesta “i” y “j”, respectivamente.

La matriz de varianzas y covarianzas del vector de observaciones multivariado del modelo en la Ec. 14, con el modelo lineal del Diseño II de Carolina del Norte asociado a cada característica es:

$$\begin{aligned} Var(\underline{y}) &= \mathbf{V}_M = \left\{ \sum_{s=1}^9 Z_{is} Z'_{is} \sigma_{ij_s} \right\} \\ &= \left\{ \sigma_{ij_E} (\mathbf{I}_e \otimes \mathbf{1}_{mhr} \mathbf{1}'_{mhr}) + \sigma_{ij_m} (\mathbf{1}_e \mathbf{1}'_e \otimes \mathbf{I}_m \otimes \mathbf{1}_{hr} \mathbf{1}'_{hr}) \right. \\ &\quad + \sigma_{ij_h} (\mathbf{1}_e \mathbf{1}'_e \otimes \mathbf{1}_m \mathbf{1}'_m \otimes \mathbf{I}_h \otimes \mathbf{1}_r \mathbf{1}'_r) \\ &\quad + \sigma_{ij_{mh}} (\mathbf{1}_e \mathbf{1}'_e \otimes \mathbf{I}_{mh} \otimes \mathbf{1}_r \mathbf{1}'_r) \\ &\quad + \sigma_{ij_{Em}} (\mathbf{I}_e \otimes \mathbf{I}_m \otimes \mathbf{1}_{hr} \mathbf{1}'_{hr}) \\ &\quad + \sigma_{ij_{Eh}} (\mathbf{I}_e \otimes \mathbf{1}_m \mathbf{1}'_m \otimes \mathbf{I}_h \otimes \mathbf{1}_r \mathbf{1}'_r) \\ &\quad + \sigma_{ij_{Emh}} (\mathbf{I}_e \otimes \mathbf{I}_{mh} \otimes \mathbf{1}_r \mathbf{1}'_r) \\ &\quad \left. + \sigma_{ij_{r/E}} (\mathbf{I}_e \otimes \mathbf{1}_{mh} \mathbf{1}'_{mh} \otimes \mathbf{I}_r) + \sigma_{ij} \mathbf{I}_{emhr} \right\} \text{ para } i, j = 1, 2. \end{aligned}$$

La matriz de varianzas y covarianzas $\mathbf{K}'_M \mathbf{V}_M \mathbf{K}_M$ de la función de verosimilitud restringida para la Ec.14 del Diseño II es:

$$\begin{aligned}
\mathbf{K}'_M \mathbf{V}_M \mathbf{K}_M &= \left\{ \begin{matrix} m \\ \sum_{i=1}^9 \mathbf{K}'_i \mathbf{Z}'_i \mathbf{Z}_i \mathbf{K}_i \sigma_{ij_s}^2 \end{matrix} \right\} \\
&= \left\{ \begin{matrix} m \\ \sigma_{ij_E}^2 \left[[(\mathbf{I}_{e-1} \otimes \mathbf{J}_{mhr}) \oplus \mathbf{J}_{mhr-1}] - \frac{1}{e} \mathbf{J}_{emhr-1} \right] \right. \\
&\quad + \sigma_{ij_m}^2 [[[\mathbf{J}_{e-1} \otimes (\mathbf{I}_m \otimes \mathbf{J}_{hr})] \oplus [(\mathbf{I}_m \otimes \mathbf{J}_{hr}) \\
&\quad \oplus \mathbf{J}_{hr-1}]] - \frac{1}{m} \mathbf{J}_{emhr-1}] + \sigma_{ij_h}^2 [[[\mathbf{J}_{em-1} \otimes (\mathbf{I}_h \otimes \mathbf{J}_r)] \\
&\quad \oplus [(\mathbf{I}_{h-1} \otimes \mathbf{J}_r) \oplus \mathbf{J}_{r-1}]] - \frac{1}{h} \mathbf{J}_{emhr-1}] \\
&\quad + \sigma_{ij_{mh}}^2 [[[\mathbf{J}_{e-1} \otimes (\mathbf{I}_{mh} \otimes \mathbf{J}_r)] \\
&\quad \oplus [(\mathbf{I}_{mh-1} \otimes \mathbf{J}_r) \oplus \mathbf{J}_{r-1}]] - \frac{1}{mh} \mathbf{J}_{emhr-1}] \\
&\quad + \sigma_{ij_{Em}}^2 \left[[(\mathbf{I}_{em-1} \otimes \mathbf{J}_{hr}) \oplus \mathbf{J}_{hr-1}] - \frac{1}{em} \mathbf{J}_{emhr-1} \right] \\
&\quad + \sigma_{ij_{Eh}}^2 [[[\mathbf{I}_{e-1} \otimes (\mathbf{J}_m \otimes [(\mathbf{I}_h \otimes \mathbf{J}_r)])] \\
&\quad \oplus [[\mathbf{J}_{m-1} \otimes [(\mathbf{I}_h \otimes \mathbf{J}_r)]]] \\
&\quad \oplus [[(\mathbf{I}_{h-1} \otimes \mathbf{J}_r) \oplus \mathbf{J}_{r-1}]] - \frac{1}{eh} \mathbf{J}_{emhr-1}] \\
&\quad + \sigma_{ij_{Emh}}^2 \left[[(\mathbf{I}_{emh-1} \otimes \mathbf{J}_r) \oplus \mathbf{J}_{r-1}] - \frac{1}{emh} \mathbf{J}_{emhr-1} \right] \\
&\quad + \sigma_{ij_{r/E}}^2 [[[\mathbf{I}_{e-1} \otimes ((\mathbf{J}_{mh} \otimes \mathbf{I}_r))]] \\
&\quad \oplus [(\mathbf{J}_{mh-1} \otimes \mathbf{I}_r) \oplus \mathbf{I}_{r-1}]] - \frac{1}{eh} \mathbf{J}_{emhr-1}] \\
&\quad + \sigma_{ij}^2 \left[\mathbf{I}_{emhr-1} - \frac{1}{emhr} \mathbf{J}_{emhr-1} \right] \left. \right\} \text{ para } i, j = 1, 2.
\end{aligned}$$

(Ec. 15)

Al igual que en el Diseño I, la matriz $\mathbf{K}'_M \mathbf{V}_M \mathbf{K}_M$ de la Ec. 15, es requerida para obtener las derivadas parciales de la función de verosimilitud restringida del modelo multivariado del Diseño II de Carolina del Norte, con respecto a los componentes de varianza y covarianza, como es mostrado en las Ecs. 4 y 5.

Sea $\hat{\Sigma}$ la matriz diagonal que contiene los estimadores de REML de los componentes de varianza y covarianza de las características “i” y “j” de los nueve factores:

$$\hat{\Sigma} = \text{Diag}(\hat{\theta}) = \text{Diag}\{\text{Diag}(\hat{\sigma}_{ii_s}, \hat{\sigma}_{ij_s}, \hat{\sigma}_{jj_s})\}_{27 \times 27},$$

para $s = 1, 2, \dots, 9$

Al considerar ausencia de endogamia, epistasis y efectos maternos, y presencia de equilibrio de ligamiento, se tiene

$$\text{que } \sigma_{ij_m} = \frac{1}{4} \sigma_{ij_A}, \sigma_{ij_h} = \frac{1}{4} \sigma_{ij_A}, \sigma_{ij_{mh}} = \frac{1}{4} \sigma_{ij_D},$$

$$\sigma_{ij_{Em}} = \frac{1}{4} \sigma_{ij_{AE}}, \sigma_{ij_{Eh}} = \frac{1}{4} \sigma_{ij_{AE}} \text{ y } \sigma_{ij_{Emh}} = \frac{1}{4} \sigma_{ij_{DE}},$$

donde $\sigma_{ij_{AE}}$ y $\sigma_{ij_{DE}}$ son las covarianzas de las aracterísticas “i” y “j” de la interacción entre los efectos aditivos y de dominancia con los efectos ambientales, respectivamente. Los estimadores de las varianzas fenotípicas

($\hat{\sigma}_{ii_p}$ y $\hat{\sigma}_{jj_p}$) para las variables respuesta “i” y “j” y el estimador de la covarianza fenotípica ($\hat{\sigma}_{ij_p}$) de éstas, se obtienen en forma matricial con los vectores:

$$\underline{c}'_{ii} = [1002000002002 \\ 00000200000000]$$

$$\underline{c}'_{ij} = [0100200000200 \\ 20000020000000]$$

$$\underline{c}'_{jj} = [0010020000020 \\ 02000002000000]$$

y sustituyéndolos en las formas cuadráticas siguientes:

$$\hat{\sigma}_{ij_p} = \underline{c}'_{ij} \hat{\Sigma} \underline{c}_{ij} = 4\hat{\sigma}_{ij_m} + 4\hat{\sigma}_{ij_{mh}} + 4\hat{\sigma}_{ij_{Em}} \\ + 4\hat{\sigma}_{ij_{Emh}} + \hat{\sigma}_{E_{ij}}$$

$$\hat{\sigma}_{ii_p} = \underline{c}'_{ii} \hat{\Sigma} \underline{c}_{ii} = 4\hat{\sigma}_{ii_m} + 4\hat{\sigma}_{ii_{mh}} + 4\hat{\sigma}_{ii_{Em}} \\ + 4\hat{\sigma}_{ii_{Emh}} + \hat{\sigma}_{E_{ii}}$$

$$\hat{\sigma}_{jj_p} = \underline{c}'_{jj} \hat{\Sigma} \underline{c}_{jj} = 4\hat{\sigma}_{jj_m} + 4\hat{\sigma}_{jj_{mh}} + 4\hat{\sigma}_{jj_{Em}} \\ + 4\hat{\sigma}_{jj_{Emh}} + \hat{\sigma}_{E_{jj}}$$

de tal manera que el estimador de la correlación fenotípica entre las características “i” y “j” se obtiene como:

$$\hat{\rho}_{ij_p} = \frac{\underline{c}'_{ij} \hat{\Sigma} \underline{c}_{ij}}{\left[(\underline{c}'_{ii} \hat{\Sigma} \underline{c}_{ii}) (\underline{c}'_{jj} \hat{\Sigma} \underline{c}_{jj}) \right]^{1/2}} = \frac{\hat{\sigma}_{ij_p}}{(\hat{\sigma}_{ii_p} \hat{\sigma}_{jj_p})^{1/2}} \\ = \frac{4\hat{\sigma}_{ij_m} + 4\hat{\sigma}_{ij_{mh}}}{\sqrt{4\hat{\sigma}_{ii_m} + 4\hat{\sigma}_{ii_{mh}} + 4\hat{\sigma}_{ii_{Em}} + 4\hat{\sigma}_{ii_{Emh}} + \hat{\sigma}_{E_{ii}}}} \\ \frac{+ 4\hat{\sigma}_{ij_{Em}} + 4\hat{\sigma}_{ij_{Emh}} + \hat{\sigma}_{E_{ij}}}{\sqrt{4\hat{\sigma}_{jj_m} + 4\hat{\sigma}_{jj_{mh}} + 4\hat{\sigma}_{jj_{Em}} + 4\hat{\sigma}_{jj_{Emh}} + \hat{\sigma}_{E_{jj}}}} \quad (\text{Ec. 16})$$

En el caso de los estimadores de las varianzas genéticas ($\hat{\sigma}_{ii_g}$ y $\hat{\sigma}_{jj_g}$) para las variables respuesta “i” y “j”, y del estimador de la respectiva covarianza genética ($\hat{\sigma}_{ij_g}$), se obtienen en forma matricial con los vectores:

$$\underline{c}'_{ii} = [0002000002000 \\ 00000000000000]$$

$$\underline{c}'_{ij} = [0002000002000 \\ 00000000000000]$$

$$\underline{c}'_{jj} = [0002000002000 \\ 00000000000000]$$

y sustituyéndolos en las formas cuadráticas

$$\hat{\sigma}_{ij_g} = \underline{c}'_{ij} \hat{\Sigma} \underline{c}_{ij} = 4\hat{\sigma}_{ij_m} + 4\hat{\sigma}_{ij_{mh}} ;$$

$$\hat{\sigma}_{ii_g} = \underline{c}'_{ii} \hat{\Sigma} \underline{c}_{ii} = 4\hat{\sigma}_{ii_m} + 4\hat{\sigma}_{ii_{mh}} ; \text{ y}$$

$\hat{\sigma}_{jj_g} = \underline{c}'_{jj} \hat{\Sigma} \underline{c}_{jj} = 4\hat{\sigma}_{jj_m} + 4\hat{\sigma}_{jj_{mh}}$, por lo que el estimador de la correlación genética de las características “i” y “j” se obtiene como:

$$\begin{aligned}\hat{\rho}_{ij_g} &= \frac{c'_{ij} \hat{\Sigma} c_{ij}}{\left[(c'_{ii} \hat{\Sigma} c_{ii}) (c'_{jj} \hat{\Sigma} c_{jj}) \right]^{1/2}} \\ &= \frac{\hat{\sigma}_{ij_g}}{(\hat{\sigma}_{ii_g} \hat{\sigma}_{jj_g})^{1/2}} \\ &= \frac{4\hat{\sigma}_{ij_m} + 4\hat{\sigma}_{ij_{mh}}}{\sqrt{4\hat{\sigma}_{ii_m} + 4\hat{\sigma}_{ii_{mh}}} \sqrt{4\hat{\sigma}_{jj_m} + 4\hat{\sigma}_{jj_{mh}}}}\end{aligned}\quad (\text{Ec. 17})$$

Los estimadores basados en el método de máxima verosimilitud tienen la propiedad de ser invariantes, lo que significa que si $\hat{\theta}$ es un estimador de máxima verosimilitud de θ y la función $g(\theta)$ es continua, entonces $g(\hat{\theta})$ también es el estimador de máxima verosimilitud de $g(\theta)$ (Freund, 1992). De acuerdo con la invarianza de los estimadores de REML de σ_{ij_g} , σ_{ii_g} y σ_{jj_g} , las Ecs. 16 y 17 corresponden a los estimadores de máxima verosimilitud restringida de ρ_{ij_p} y ρ_{ij_g} , respectivamente.

Sea

$$\hat{\theta}_{27 \times 1}^t = [\hat{\sigma}_{ii_1} \quad \hat{\sigma}_{ij_1} \quad \hat{\sigma}_{jj_1} \quad \dots \quad \hat{\sigma}_{ii_9} \quad \hat{\sigma}_{ij_9} \quad \hat{\sigma}_{jj_9}]^t$$

el estimador de los correspondientes elementos del vector de parámetros $\hat{\theta}_{27 \times 1}$. La varianza aproximada de $\hat{\rho}_{ij_p}$ según la Ec. 7, es:

$$\begin{aligned}Var(\hat{\rho}_{ij_p}) \\ \approx (\hat{\rho}_{ij_p})^2 \left[\sum_{r=1}^{27} b_r^2 Var(\hat{\theta}_r) + \sum_{r=1}^{27} \sum_{\substack{l=1 \\ r \neq l}}^{27} b_r b_l Cov(\hat{\theta}_r, \hat{\theta}_l) \right]\end{aligned}\quad (\text{Ec. 18})$$

donde:

$$b_1 = \frac{-1}{2\sigma_{ii_p}}, \quad b_2 = \frac{1}{\sigma_{ij_p}}, \quad b_3 = \frac{-1}{2\sigma_{jj_p}},$$

$$b_4 = b_{10} = b_{13} = b_{19} = \frac{-2}{\sigma_{ii_p}},$$

$$b_5 = b_{11} = b_{14} = b_{20} = \frac{4}{\sigma_{ij_p}},$$

$$b_6 = b_{12} = b_{15} = b_{21} = \frac{-2}{\sigma_{jj_p}}, \text{ y } b_r = 0 \text{ para } r=7, 8,$$

9, 16, 17, 18, 22, 23, 24, 25, 26, 27.

La aproximación a la varianza del estimador de la correlación genotípica $\hat{\rho}_{ij_g}$ se obtiene como:

$$\begin{aligned}Var(\hat{\rho}_{ij_g}) &\approx (\hat{\rho}_{ij_g}) \\ &\quad \left[\sum_{r=1}^{27} b_r^2 Var(\hat{\theta}_r) + \sum_{r=1}^{27} \sum_{\substack{l=1 \\ r \neq l}}^{27} b_r b_l Cov(\hat{\theta}_r, \hat{\theta}_l) \right]\end{aligned}\quad (\text{Ec. 19})$$

donde:

$$b_4 = b_{10} = \frac{-2}{\sigma_{ii_p}}, \quad b_5 = b_{11} = \frac{4}{\sigma_{ij_p}},$$

$$b_6 = b_{12} = \frac{-2}{\sigma_{jj_p}}, \text{ y } b_r = 0 \text{ para } r=1, 2, 3, 7, 8, 9,$$

13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27.

Los estimadores de la varianza de la correlación fenotípica y genotípica bajo el Diseño II se obtienen al sustituir a los componentes de varianza y covarianza por sus estimadores, a los elementos de la matriz $Var(\hat{\theta})$ por los de la matriz $\hat{Var}(\hat{\theta})$, como en la Ec. 6.

DISCUSIÓN

La estimación por máxima verosimilitud restringida presentada para los Diseños I y II tiene una serie de ventajas teóricas que lo hacen un método muy deseable para su aplicación en el mejoramiento genético vegetal. Para conjuntos de datos completamente balanceados y cuando no

hay estimaciones negativas de componentes de varianza con el método ANOVA, los estimadores de componentes de varianza y covarianza son insesgados y de mínima varianza, idénticos a los de REML. Sin embargo, en conjuntos de datos desbalanceados los estimadores de ANOVA sólo mantienen la propiedad de insesgamiento, mientras que los estimadores de REML conservan sus propiedades asintóticas de consistencia, eficiencia, distribución normal y matriz de dispersión conocida (Searle *et al.*, 1992). Al respecto, algunos estudios han comprobado estas propiedades teóricas para ambos tipos de estimadores. Corbeil y Searle (1976) compararon los estimadores ANOVA y REML de componentes de varianza utilizando modelos con un criterio de clasificación, y dos criterios de clasificación jerárquicos y cruzados con interacción y sin interacción; encontraron que los valores estimados con ambos métodos fueron los mismos para conjuntos de datos balanceados, y muy similares para conjuntos de datos moderadamente desbalanceados. Swallow y Monahan (1984) realizaron la comparación Monte Carlo de los estimadores ANOVA, ML y REML de componentes de varianza utilizando un modelo aleatorio con un criterio de clasificación, y encontraron que para conjuntos de datos relativamente pequeños los estimadores ML fueron más sesgados que los estimadores ANOVA y REML, y que el sesgo disminuyó conforme se incrementó el tamaño del conjunto de datos. Los estimadores ANOVA y REML fueron generalmente adecuados y muy similares en términos de sesgo y error cuadrado medio, excepto para conjuntos de datos seriamente desbalanceados, para los cuales el error cuadrado medio de los estimadores ANOVA fue considerablemente mayor que el de los estimadores REML.

En un estudio de simulación basado en los Diseños I y II de Carolina del Norte en que se utilizaron dos tamaños de conjuntos de datos balanceados y diferentes conjuntos de valores de los parámetros para realizar la simulación, Shaw (1987) encontró que la precisión de los estimadores REML de componentes de varianza aumentó conforme se incrementó el tamaño del conjunto de datos. En el análisis de conjuntos pequeños de datos, los errores estándar de los estimadores REML de componentes de varianza genética aditiva y heredabilidad fueron el doble de los correspondientes errores estándar obtenidos con conjuntos de datos relativamente grandes, y el sesgo de los estimadores REML de componentes de varianza fue debida únicamente a la propiedad de ser no negativos, el sesgo disminuyó cuando se aumentó el tamaño de los conjuntos de datos. Estas propiedades en conjunto hacen preferibles a los estimadores REML sobre los estimadores de ANOVA.

También, el uso de los modelos estadísticos mixtos en la estimación por máxima verosimilitud restringida permite el manejo de los diseños de bloques incompletos, conducen a

una estimación más precisa de los componentes de varianza y covarianza, y de la correlación fenotípica y genotípica, a través de un mejor control del error experimental. Por ejemplo, la separación de componentes de varianza para diferentes grupos de tratamientos en el mismo experimento tiene una utilidad considerable en el mejoramiento genético vegetal, especialmente cuando se incluyen genotipos provenientes de poblaciones seleccionadas y no seleccionadas. Cuando además se utiliza un diseño de bloques incompletos la estimación es todavía más complicada. Con la estimación de mínimos cuadrados (ANOVA) primero se estiman las medias poblacionales con todos los datos y después se estiman los componentes de varianza separadamente para cada grupo. Generalmente los bloques incompletos son ignorados en dicho análisis para evitar un severo desbalance de los datos y para facilitar el análisis estadístico. Aún cuando los bloques incompletos son considerados en el análisis de subgrupos, cierta información se pierde a nivel de bloques completos e incompletos, y con esto la eficiencia de los diseños incompletos en el control del error experimental.

Con el uso de los modelos mixtos es posible estimar los efectos de bloques completos e incompletos y los componentes de varianza de subgrupos con un sólo análisis, lo que aumenta la precisión de la estimación. Cervantes-Martínez *et al.* (2001) estimaron exitosamente con el procedimiento MIXED de SAS (Littell *et al.*, 1996) los componentes de varianza de cinco poblaciones simultáneas; el estudio incluyó 100 genotipos de cada población asignados al azar a conjuntos, y cada conjunto fue diseñado como un látice dentro de cada uno de los cuatro ambientes evaluados, y con cerca de 15 % de datos perdidos.

El proceso de maximizar el logaritmo natural de la matriz de verosimilitud restringida es un problema numérico complicado, puesto que la función l_R mostrada en la Ec. 3, y sus derivadas parciales dadas en Ecs. 4 y 5, incluyen la inversa de una matriz de dimensión $n^* \times n^*$, donde n^* son los grados de libertad después de ajustar los efectos fijos del modelo. Cada iteración del proceso de maximización implica el cálculo de la inversa $n^* \times n^*$, lo que demanda muchos de recursos computacionales. Sin embargo, las computadoras y paquetes estadísticos modernos han hecho que la estimación por REML de componentes de varianza sea relativamente fácil. Por ejemplo, como se mencionó anteriormente, el procedimiento MIXED de SAS proporciona un método robusto y conveniente para realizar el análisis de máxima verosimilitud restringida de varios tipos de modelos mixtos (Littell *et al.*, 1996; SAS Institute Inc., 1999), además de que la matriz de dispersión asintótica estimada puede obtenerse directamente de los archivos de salida del procedimiento.

Aún cuando el procedimiento MIXED no contempla el análisis multivariado directo de modelos mixtos, el cual es requerido para la estimación de REML de componentes de covarianza entre caracteres debida a los diferentes factores del modelo, tal análisis se puede realizar generando un conjunto de datos en formato "alto" como en la Ec. 2, es decir, que ambas variables (para $t = 2$) se manejen como una sola, pero que además se incluya una variable de clasificación adicional en el modelo que indique el carácter en cuestión, y se especifique un modelo de niveles múltiples por medio de los comandos RANDOM y REPEATED, y las opciones TYPE y SUBJECT para especificar la estructura de la matriz de varianzas y covarianzas. Con un procedimiento de este tipo se han realizado algunos análisis multivariados para modelos mixtos más simples que los descritos en el presente trabajo con el procedimiento MIXED de SAS (Wright, 1998).

Para determinar las ventajas prácticas de los estimadores de máxima verosimilitud restringida sobre los estimadores de ANOVA, es recomendable que ambos tipos de estimadores de componentes de varianza y covarianza de múltiples características, con los diseños y condiciones descritas en este trabajo, así como los estimadores de la correlación fenotípica y genotípica, sean comparados en futuras investigaciones con conjuntos de datos generados y considerando varias combinaciones de valores verdaderos de los parámetros mediante algún método de simulación, como es el de Monte Carlo, e implementando la estimación con el procedimiento MIXED y GLM de SAS (Littell *et al.*, 1996; SAS Institute Inc., 1999). Otra opción con amplias perspectivas a explorarse en el futuro, es la elaboración de programas computacionales en el lenguaje de programación IML de SAS (SAS Institute Inc., 1999) o en otro lenguaje, que sea específico para realizar la estimación de máxima verosimilitud restringida de componentes de varianza y covarianza de caracteres múltiples para los Diseños I y II, con las especificaciones de los modelos descritas en este trabajo.

CONCLUSIONES

La extensión de modelos mixtos univariados para estimar componentes de varianza y covarianza conduce a modelos multivariados que permiten estimar la máxima verosimilitud restringida de la covarianza entre características para los diferentes factores, en los modelos de los Diseños I y II de Carolina del Norte. Los estimadores de componentes de varianza y covarianza de características múltiples poseen todas las propiedades teóricas de los estimadores basados en la máxima verosimilitud, que son: consistencia, eficiencia, normalidad asintótica, y la disponibilidad de la matriz de dispersión asintótica, por lo que tales estimadores REML son preferibles sobre otros.

Una aproximación a la varianza de los estimadores de la correlación fenotípica y genotípica puede obtenerse de un modo relativamente simple con el método Delta. Esta aproximación es particularmente útil cuando la estructura de las funciones de los estimadores de componentes de varianza y covarianza incluidos en los estimadores de las diferentes correlaciones aumenta la complejidad algebraica del estimador y sus propiedades, si las primeras derivadas parciales de estos coeficientes existen y son continuas, y se cuenta con las varianzas y covarianzas muestrales de los estimadores de los componentes de varianza y covarianza.

...

BIBLIOGRAFÍA

- Cervantes-Martínez C T, K J Frey, P J White, D M Wesenberg, J B Holland (2001). Selection for greater β -glucan content in oat grain. *Crop Sci.* 41:1085-1091.
- Comstock R E, H F Robinson (1948). The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. *Biometrics* 4: 254-265.
- Corbeil R R, S R Searle (1976). Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics* 18:31-38.
- Fehr W R (1991). Principles of Cultivar Development. Volume 1. Theory and Technique. Iowa State University Press. Ames, Iowa. 536 p.
- Freund J E (1992). Mathematical Statistics. Fifth edition. Prentice Hall. New Jersey. 658 p.
- Hallauer A R, J B Miranda Fo. (1988). Quantitative Genetics in Maize Breeding. Second edition Iowa State University Press, Ames, Iowa. 468 p.
- Harville D A (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* 72:320-340.
- Henderson C R (1953). Estimation of variance and covariance components. *Biometrics* 9:226-252.
- Lindstrom M J, D M Bates (1989). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-effects models for repeated measures data. *J. Am. Stat. Assoc.* 83:1014-1022.
- Littell R C, G A Milliken, W W Stroup, R D Wolfinger (1996). SAS System for Mixed Models. SAS Inst. Inc., Cary, NC. 633 p.
- Lynch M, B Walsh (1998). Genetics and Analysis of Quantitative Traits. Sinauer Associates, Inc., Sunderland, MA. 980 p.
- Meeker W Q, L A Escobar (1998). Statistical Methods for Reliability Data. John Wiley. New York. 680 p.
- Meyer K (1985). Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics* 41:153-165.
- Miller J J (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *The Annals of Statistics* 5:746-762.
- Milliken G A, D E Johnson (1992). Analysis of Messy Data, Volume I: Designed Experiments. Chapman & Hall. New York. 473 p.
- SAS Institute Inc. (1999). SAS Online Doc, Version 8. CD-ROM. SAS Institute, Inc., Cary, N.C.
- Schaeffer L R, J W Wilton, R Thompson (1978). Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics* 34:199-208.
- Shaw R G (1987). Maximum-likelihood approaches applied to quantitative genetics of natural populations. *Evolution* 41:812-826.
- Searle S R (1971). Linear Models. New York: John Wiley. 423 p.

- Searle S R (1982).** Matrix Algebra Useful for Statistics. New York: John Wiley. 438 p.
- Searle S R, G Casella, C E McCulloch (1992).** Variance Components. New York: John Wiley. 501 p.
- Swallow W H, J F Monahan (1984).** Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimator of variance components. *Technometrics* 26:47-57.
- Wright S P (1998).** Multivariate analysis using the MIXED procedure. In: Proc. 23rd Annual SAS Users Group International Conference. March 22-25. Nashville, TN. (ed.) SAS Inst. Inc., Cary, NC. pp:1238-1242.