

MODELOS CON TÉRMINO MULTIPLICATIVO. APLICACIÓN EN EL ANÁLISIS DE LA INTERACCIÓN GENOTIPO AMBIENTE

M. Varela[✉] y J. G. Castillo

ABSTRACT. Residuals of the second order interaction associated to a linear model are analyzed, starting from multiplicative models for interaction (AMMI model). For its adjustment, a computational program was made in Matlab, which will be applied to the analysis of genotype environment interaction, with the objective of characterizing stable or unstable varieties, considering its interaction with the environment.

Key words: models, genotype environment interaction

RESUMEN. Se analizan los residuales de interacción de segundo orden asociados a un modelo lineal, a partir de modelos con término multiplicativo en la interacción (modelos AMMI). Para su ajuste, se elaboró un programa computacional en Matlab, el cual será aplicado en el análisis de la interacción genotipo ambiente, con el objetivo de caracterizar variedades en estables o inestables a partir de su interacción con el ambiente.

Palabras clave: modelos, interacción genotipo ambiente

INTRODUCCIÓN

Los modelos con término multiplicativo han sido muy utilizados para describir la interacción en tablas de dos vías; tienen la ventaja de permitir una representación Biplot (simultánea) de filas y columnas de la tabla, lo que facilita identificar las combinaciones de niveles causantes de la interacción. Estos modelos, a su vez, se clasifican en internos y externos (1): internos cuando estiman la interacción haciendo uso solamente de la información contenida en la tabla inicial de datos y externos cuando utilizan además información proveniente de variables externas, ya sean medidas sobre filas, columnas o ambos factores de variación.

Entre los modelos con término multiplicativo que aparecen en la literatura podemos citar: el Modelo Concurrente de Tukey (2) y los Modelos de Regresión sobre la media de Yates y Cochran, Mandel, Finlay y Wilkinson, Eberhart y Russel (3). Se caracterizan por tratar de explicar la interacción a partir de un solo término multiplicativo, lo cual en muchos casos resulta insuficiente dada su complejidad.

En los modelos AMMI (4), se mantiene la descomposición en términos multiplicativos de la interacción y esta no es forzada a tener una característica específica; en el modelo se incluyen tantos términos como sean necesarios, para explicar la variabilidad asociada a la interacción y son clasificados también como modelos internos.

Estos modelos combinan las técnicas de Análisis de Varianza y Análisis de Componentes Principales (5); tienen como objetivo explicar la interacción asociada a un ANOVA bifactorial, a partir de una representación biplot.

Existen otras formas de hacer referencia a este tipo de modelos con término multiplicativo en la interacción y efectos principales aditivos (modelos AMMI); así por ejemplo, algunos lo denominan modelos bilineales (6, 7); por otra parte, otros lo llaman modelos biaditivos (8).

Los modelos AMMI han sido aplicados fundamentalmente en experimentos de campo, más específicamente en el análisis de la interacción genotipo-ambiente, con el objetivo de clasificar genotipos en estables e inestables a partir de su interacción con el ambiente (9, 10, 11).

Es conocido que a partir del Análisis de la Varianza, podemos detectar la presencia o no de interacción mediante la F de Fisher correspondiente a esta fuente de variación. Sin embargo, una vez detectada, no la interpretamos, nos limitamos a seleccionar los genotipos con mayores valores medio, sin tener en cuenta su grado de estabilidad.

Los modelos AMMI, al permitir una representación Biplot de filas (genotipos) y columnas (ambientes), dan la posibilidad de estudiar el grado de estabilidad de los genotipos, al ser probados en diferentes ambientes.

En este trabajo se presenta un *software* elaborado en Matlab, el cual permite el ajuste de un modelo AMMI a los residuales de interacción de segundo orden asociadas a un modelo lineal. Los resultados se presentan a partir de un Biplot, es decir, mediante una representación gráfica simultánea de variedades y ambientes.

OTROS MODELOS CON TÉRMINO MULTIPLICATIVO

Como sabemos, el modelo lineal al que se ajustan los datos experimentales en un arreglo bifactorial bajo un

Dr.C. M. Varela, Investigador Agregado del Departamento de Matemática Aplicada y Ms.C. J. G. Castillo, Investigador del Departamento de Genética y Mejoramiento Vegetal, Instituto Nacional de Ciencias Agrícolas, Gaveta Postal 1, San José de las Lajas, La Habana, Cuba, CP 32 700.

✉ varela@inca.edu.cu, mvn1121@yahoo.es

Diseño Completamente Aleatorizado, es de la forma siguiente:

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \quad i=1..I \quad j=1..J \quad k=1..K$$

donde:

I: no. de niveles del primer factor

J: no. de niveles del segundo factor

K: no. de observaciones por tratamiento o combinaciones de niveles

y_{ijk} : k-ésima observación correspondiente a la combinación de niveles ij

α_i, β_j : efectos principales para filas y columnas respectivamente

$(\alpha\beta)_{ij}$: efecto interacción.

Sabemos, además, que el estimador mínimo cuadrático correspondiente a la interacción se calcula de la forma:

En la literatura se presentan modelos, en los que se trata de explicar la interacción a partir de términos multiplicativos; en dichos términos, aparecen como factores los efectos principales asociados a filas, columnas o ambas fuentes de variación. Sin embargo, en muchos casos, resulta imposible modelar la interacción a partir de estos términos, por su complejidad.

Estos modelos han sido desarrollados para el caso de una observación por celda, es decir, no tenemos repeticiones que permitan controlar la variabilidad dentro de la celda y, por tanto, no es posible estimar el error experimental.

Tukey (2) fue el primer investigador que propuso un modelo para el análisis de la interacción en experimentos de dos vías, para diseños no replicados.

El modelo propuesto es de la forma:

$$y_{ij} = \mu + \alpha_i + \beta_j + \lambda\alpha_i\beta_j + e_{ij}$$

donde:

$\alpha_i\beta_j$: es el producto de los efectos principales

: un coeficiente de regresión.

Contrastar la hipótesis sobre $\lambda=0$ será equivalente a un test de hipótesis para contrastar que el producto de los efectos no contribuye a la predicción de y_{ij} .

Luego, su finalidad es distinguir entre el modelo aditivo (solo incluye los efectos principales) y el modelo que contempla un término de interacción.

Otro tipo de modelos con término multiplicativo muy utilizado para describir la interacción en tablas de dos vías, son los Modelos de Regresión sobre la media (3), los cuales se definen de la siguiente forma:

$$y_{ij} = \mu + \alpha_i + \beta_j + \lambda\gamma_i\beta_j \quad (\hat{\alpha\beta})_{ij} = \lambda \hat{\gamma}_i \hat{\beta}_j \quad \text{regresión para filas}$$

$$y_{ij} = \mu + \alpha_i + \beta_j + \lambda\alpha_i\gamma_j \quad (\hat{\alpha\beta})_{ij} = \lambda \hat{\alpha}_i \hat{\gamma}_j \quad \text{regresión para columnas}$$

Como su nombre lo indica, se estiman los valores de γ_i a partir de las regresiones de los efectos principales en la variable dependiente.

Nótese que en los modelos anteriores, la interacción es forzada a tener características muy específicas; por tanto, puede darse el caso de que la descomposición en términos multiplicativos no sea suficiente para explicar la variabilidad asociada a la interacción.

MODELOS AMMI

En los modelos AMMI (12)-Efecto Interacción Multiplicativo y Efectos Principales Aditivos- se combinan las técnicas de Análisis de Varianza y Análisis de Componentes Principales. Se introducen al modelo tantos términos multiplicativos como sean necesarios para explicar la variabilidad de la interacción.

Gabriel (6) muestra la conexión entre el ajuste mínimo cuadrático de un modelo multiplicativo y la descomposición en valores singulares de una matriz (13).

El método consiste en hacer la descomposición en valores singulares de la matriz Z de orden IxJ, formada por los estimadores de las interacciones en el modelo anterior.

$$Z = (z_{ij}) = (\hat{\alpha\beta})_{ij} = y_{ij} - y_{i..} - y_{.j} + y_{...}$$

Al realizar la descomposición en valores singulares de Z nos queda:

$$z_{ij} = (\hat{\alpha\beta})_{ij} = \sum_{m=1}^M \lambda_m u_{mi} v_{mj}$$

siendo M el rango de Z.

Llamaremos modelo AMMI de orden M a la expresión:

$$AMMI_M : E(y_{ijk}) = \mu + \alpha_i + \beta_j + \sum_{m=1}^M \lambda_m u_{mi} v_{mj}$$

donde:

λ_m : corresponde al valor singular de orden m de Z'Z

v_{mi} : coordenada i-ésima del vector singular de ZZ' asociado a λ_m

v_{mj} : coordenada j-ésima del vector singular de Z'Z asociado a λ_m

De esta forma, podemos representar las filas (genotipos) y columnas (ambientes) en un subespacio de dimensión M, en el que las proximidades entre genotipos van a indicar que interactúan de manera similar con el ambiente. Así, los genotipos que se ubican cerca del origen de coordenadas serán los más estables, es decir, los que interactúan poco con el ambiente. Por otra parte, aquellos genotipos que se alejan del origen serán los más inestables y tendrán altos rendimientos solamente en aquellos ambientes próximos a ellos en la representación.

SELECCIÓN DEL NÚMERO DE TÉRMINOS MULTIPLICATIVOS

Nuevamente, al estar en presencia de una técnica en la que se realiza una descomposición en valores singulares, surge la problemática de cuántos ejes elegir.

Si consideramos un solo factor el modelo será:

$$AMMI_1 : E(y_{ijk}) = \mu + \alpha_i + \beta_j + \lambda_1 u_{1i} v_{1j}$$

con dos factores:

$$AMMI_2 : E(Y_{ijk}) = \mu + \alpha_i + \beta_j + \lambda_1 u_{1i} v_{1j} + \lambda_2 u_{2i} v_{2j}$$

y así sucesivamente (5).

¿Qué modelo es el más adecuado?. ¿Cuántos términos multiplicativos deben ser incluidos en el modelo?.

Para dar respuesta a estas interrogantes, se realiza la descomposición de la suma de cuadrados de la interacción (Tabla I) asociada al Análisis de Varianza para un modelo bifactorial (9), en la forma siguiente:

Tabla I. Descomposición de la suma de cuadrados de la interacción

FV	GL	SC
INT GxA	(I-1)(J-1)	$K \sum \lambda_i^2$
AMMI ₁	(I-1)+(J-1)-1	$K \lambda_1^2$
AMMI ₂	(I-1)+(J-1)-3	$K \lambda_2^2$
	-----	-----
AMMI _M	(I-1)+(J-1)-2L-1	$K \lambda_M^2$

Para conocer si un modelo con G términos multiplicativos es válido, se realiza el ANOVA, utilizando como variabilidad total la correspondiente a la interacción, se determina el residual de la interacción asociada a la parte de la variabilidad total (interacción) que no es explicada por los G términos multiplicativos y se calcula la F de Fisher correspondiente (Tabla II).

Tabla II. ANOVA para la selección del número de términos multiplicativos

FV	GL	SC	CM	F
INT GxA	(I-1)(J-1)	$K \sum \lambda_i^2$		
AMMI ₁	(I-1)+(J-1)-1	$K \lambda_1^2$		
AMMI ₂	(I-1)+(J-1)-3	$K \lambda_2^2$		
	-----	-----		
AMMI _G	(I-1)+(J-1)-2L-1	$K \lambda_G^2$	S.C/GL	CMAMMI _G /CMRes
Residual	Por diferencia	Por dif.	C.M.Res	

Si la F de Fisher es significativa, se incluye el término multiplicativo asociado a λ_G y se pasa a analizar el término G+1. En caso contrario, se elimina y no se continúa el análisis, debido a la relación de orden existente entre los valores singulares.

REPRESENTACIÓN BIPLLOT

Al ajustar un modelo AMMI de orden q (q términos multiplicativos), estamos aproximando la matriz de residuales de interacción de segundo orden a partir del siguiente modelo (con la descomposición en valores y vectores singulares de la matriz de residuales de interacción de segundo orden, estamos ajustando al siguiente modelo):

$$Z \approx U_{(q)} \Sigma_{(q)} V_{(q)}$$

$\Sigma_{(q)}$: matriz diagonal que contiene los q primeros valores propios de Z

$U_{(q)}$ y $V_{(q)}$: representan las primeras q columnas de U y V respectivamente.

Esta descomposición nos permite obtener una representación Biplot, en la que van a ser ploteados simultáneamente los genotipos y ambientes. En dicha representación tomamos como marcadores para el genotipo i, la fila i de la matriz $U_{(q)}$ y como marcador para el ambiente j la columna j de la matriz $V_{(q)}$.

En dicha representación, proximidades entre genotipos y ambientes indican una interacción positiva; en caso contrario, es decir, genotipos y ambientes distantes es sinónimo de interacción negativa. Por otra parte, genotipos próximos al origen de coordenadas nos indican estabilidad.

IMPLEMENTACIÓN COMPUTACIONAL

Se elaboró un programa en Matlab versión 5.0 para el ajuste de un modelo AMMI, el cual necesita como datos de entrada la cantidad de niveles de cada factor (variedad y ambiente) y los valores medio de cada genotipo en cada ambiente. Seguidamente se calcula la matriz de residuales de interacción y su descomposición en valores y vectores singulares.

El programa ofrece además las matrices de marcadores para genotipos y ambientes, así como la representación Biplot correspondiente (Figura 1).

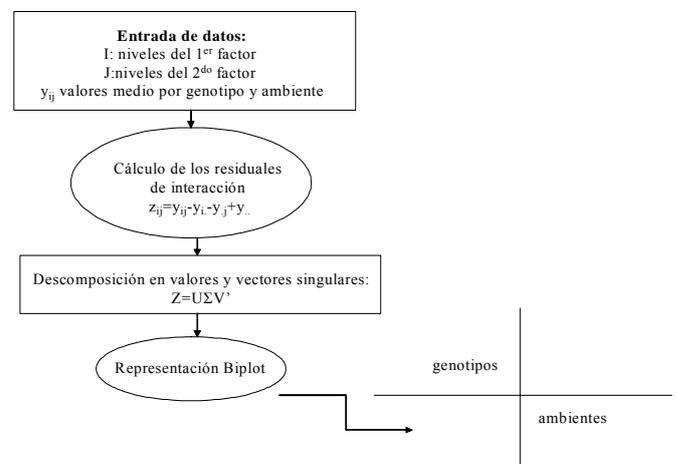


Figura 1. Etapas del programa computacional

APLICACIÓN

Se evalúa el rendimiento (t.ha⁻¹) de 10 variedades de papa, en tres localidades durante tres años (Tabla III). Las dos primeras localidades, Boyeros y San José de las Lajas, están ubicadas en la parte occidental de Cuba, mientras que la localidad de Villa Clara pertenece a la parte central del país. Son regiones de gran producción del tubérculo, con condiciones climáticas diferentes. Los datos fueron obtenidos a partir de un diseño de Bloques al Azar con tres repeticiones. Se ofrecen los valores medio del rendimiento.

Tabla III. Rendimiento (t.ha⁻¹) de las variedades en tres localidades durante tres años

Localidad Variedad/Año	Rendimiento total (t.ha ⁻¹)								
	Boyeros			San José de las Lajas			Villa Clara		
	Año 1 1993/1994	Año 2 1994/1995	Año 3 1995/1996	Año 1 1993/1994	Año 2 1994/1995	Año 3 1995/1996	Año 1 1993/1994	Año 2 1994/1995	Año 3 1995/1996
Aranka (v1)	32.07	24.10	41.77	51.37	33.43	38.04	32.04	24.00	43.03
Binella (v2)	30.18	27.33	34.66	42.56	33.12	40.02	34.36	16.89	44.80
Provento (v3)	31.91	24.79	39.33	36.83	33.49	42.03	29.36	19.44	39.50
Raja (v4)	26.64	25.95	31.55	31.84	27.72	40.07	30.26	18.62	42.30
Impala (v5)	25.20	30.36	30.44	36.12	27.18	33.60	30.97	19.56	48.90
Snowden (v6)	27.81	20.39	30.44	24.24	26.95	37.70	29.24	18.33	29.80
Granada (v7)	28.18	25.17	30.22	38.94	37.61	31.80	28.88	20.78	27.20
Desiree (v8)	28.73	24.53	38.88	34.43	29.71	35.70	26.14	20.00	43.40
Red Pont. (v9)	27.78	27.30	34.44	30.18	23.67	43.35	22.69	16.00	41.00
Baraka (v10)	32.00	28.53	36.88	27.85	34.72	40.50	22.09	20.45	32.90
Valor medio	29.05	25.84	34.86	35.43	30.76	38.28	28.60	19.40	39.28

Se consideran, por tanto, 10 genotipos y 9 ambientes formados por las combinaciones de localidades por año:

A1: Boyeros–Año 1, **A2:** Boyeros–Año 2, **A3:** Boyeros–Año 3, **A4:** San José–Año 1, **A5:** San José–Año 2, **A6:** San José–Año 3, **A7:** Villa Clara–Año 1, **A8:** Villa Clara–Año 2, **A9:** Villa Clara–Año 3.

Al realizar el Análisis de Varianza correspondiente a un arreglo bifactorial bajo un diseño de Bloques al Azar, se obtuvo una interacción genotipo ambiente significativa, lo que justifica el ajuste de un modelo AMMI a los residuales de interacción de segundo orden.

Al realizar la descomposición de la suma de cuadrados de la interacción, se pudo constatar que solamente los valores propios asociados a los dos primeros ejes del Biplot eran significativos en la explicación de la variabilidad de la interacción, acumulando entre los dos el 72.5 % de la variabilidad total. A continuación presentamos el Biplot correspondiente:

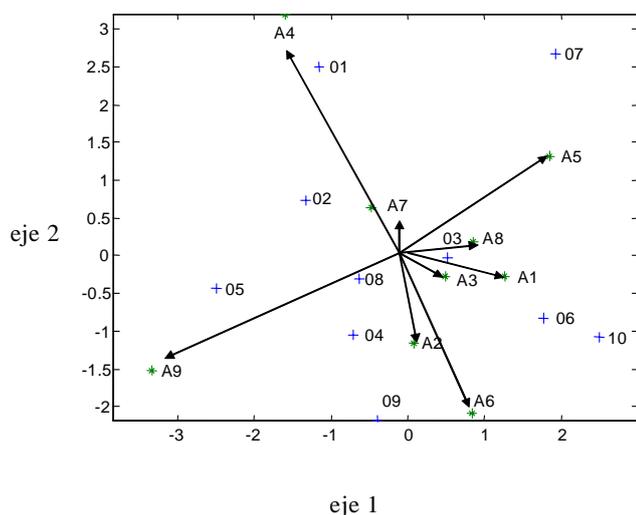


Figura 2. Representación Biplot

Como aspectos más relevantes que se destacan en el Biplot (Fig. 2), podemos decir que la variedad 1 interactúa positivamente en el ambiente 4 y negativamente en el ambiente 6, la variedad 7 interactúa positivamente

en el ambiente 5 y negativamente en los ambiente 9 y 6, las variedades 6 y 10 interactúan positivamente en los ambientes 1 y 6, y negativamente en los ambiente 9 y 4, la variedad 9 interactúa positivamente en el ambiente 6 y negativamente en el ambiente 4, la variedad 5 interactúa positivamente en el ambiente 9 y negativamente en el ambiente 5. Finalmente, podemos decir que como variedades más estables se destacan la 2, 3, 4 y 8, por estar más próximas al origen de coordenadas.

Otro estudio interesante corrobora los resultados de la aplicación de este modelo (3), al estudiar ocho clones y tres variedades de papa durante cinco años y donde un clon de los estudiados fue el más estable en todos los ambientes. Todo lo anterior evidencia la gran utilidad de estos métodos para los programas de mejoramiento de los diferentes cultivos que se desarrollan en el país.

REFERENCIAS

1. Van Eeuwijk, F. A. y Kroonenberg, P. M. Multiplicative models for interaction in three-way ANOVA, with applications to plant breeding. *Biometrics*, 1998, vol. 54, p. 1315-1333.
2. Tukey, J. W. One degree of freedom for non-additivity. *Biometrics*, 1949, vol. 5, p. 232-242.
3. González, M. E.; Estévez, A.; Castillo, J. G.; Salomón, J. L.; Varela, M.; Ortiz, U. y Ortiz, E. Análisis de la estabilidad genotípica en el cultivo de la papa (*Solanum tuberosum* L.) mediante las representaciones biplots. *Cultivos Tropicales*, 2003 vol. 24, no. 1, p 81-84.
4. Gauch, H. G. Model selection and validation for yield trials with interaction. *Biometrics*, 1988, vol. 44, p. 705-715.
5. Gauch, H. G. y Zobel, R. W. Accuracy and selection success in yield trial analyses'. *Theoretical and Applied Genetics*, 1989, vol. 77, p. 473-481.
6. Gabriel, K. R. Least squares approximation of matrices by additive and multiplicative models. *Journal of the Royal Statistical Society*, 1978, vol. 40, p. 186-196.
7. Denis, J. B. Ajustements de modèles linéaires et bilinéaires sous contraintes linéaires avec données manquantes. *Revue de Statistique Appliquée*, 1991, vol. 29, no. 2, p. 5-24.
8. Denis, J. B. y Gower, J. C. Biadditive models. *Biometrics*, 1994, vol. 50, p. 310-311.

9. Cornelius, P. L.; Crossa, J. y Seyedsadr, M. S. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. En: S. Manjit, H. G. Kang y J. R. Gauch (eds.). *Genotype by Environment Interaction*. 1996, p. 199-233.
10. Van Eeuwijk, F. A. Linear and bilinear models for the analysis of multi-environment trials: I. An inventory of models. *Euphytica*, 1995, vol. 84, p. 1-7.
11. Kang, M. S. y Gauch, H. G. *Genotype by Environment Interaction*. New York : CRC, Press. 1996.
12. Romagosa, I.; Ullrich, S. E.; Han, F. y Hayes, P. M. Use of the AMMI model in QTL mapping for adaptation in barley. *Theoretical Applied Genetics*, 1996, vol. 93, p. 30-37.
13. Gollob, H. F. A statistical model which combines features of analytic factor and analyses of variance techniques. *Psychometrika*, 1968, vol. 33, p. 73-115.

Recibido: 30 de septiembre de 2004

Aceptado: 27 de mayo de 2005

Cursos de Verano

Precio: 320 CUC

Análisis Multivariado de Datos. Aplicación a las Ciencias Agrícolas

Coordinador: Dr.C. Mario Varela Nualles

Fecha: julio

Duración: 40 horas

SOLICITAR INFORMACIÓN

Dr.C. Walfredo Torres de la Noval
Dirección de Educación, Servicios Informativos
y Relaciones Públicas
Instituto Nacional de Ciencias Agrícolas (INCA)
Gaveta Postal 1, San José de las Lajas,
La Habana, Cuba. CP 32700
Telef: (53) (64) 86-3773
Fax: (53) (64) 86-3867
E.mail: posgrado@inca.edu.cu