

MODELOS MIXTOS PARA ANALIZAR EXPERIMENTOS AGROPECUARIOS DE TRES FACTORES CON UN PROGRAMA ESTADÍSTICO¹

Pedro Guerra-Martínez²

RESUMEN

El PROC GLM de SAS® es el procedimiento más utilizado en el análisis de varianza en la investigación agropecuaria, pero fue diseñado para modelos fijos. El objetivo fue capacitar a los usuarios de SAS® en el uso de modelos mixtos en el análisis de experimentos agropecuarios de tres factores. Se tomaron datos de la literatura y se analizaron por el PROC GLM (modelo fijo y mixto) y PROC MIXED. Tres factores y sus interacciones fueron: Leguminosa (V), Fósforo (P) y Control de Malezas (W). El diseño fue de Bloques Completos al Azar (BCA de tres bloques, BLK) y la variable de respuesta fue rendimiento de semilla (kg/parcela). Se presentan las rutinas de PROC GLM para modelos fijos y mixtos y PROC MIXED. Bajo los modelos mixtos, V y P fueron efectos fijos y W aleatorio. Los grados de libertad fueron ajustados por Satterthwaite. Con PROC GLM modelo fijo, BLK, V*W y P*W fueron $p < 0,01$ y con el modelo mixto solo BLK, pero los Valores-F fueron menores. Las varianzas de W y V*P*W fueron -0,59 y -0,22, respectivamente. Con PROC MIXED ninguna fuente de variación resultó $p < 0,05$ (pruebas F y Z de Wald). La varianza de W fue cero y de V*W y P*W resultaron menores. Los errores estándares (EE) de V y V*P fueron mayores con PROC MIXED y de los efectos aleatorios fueron iguales en todos los procedimientos. Se concluye que el PROC MIXED proporciona adecuada prueba de hipótesis y EE de medias, por lo que se recomienda su utilización.

Palabras clave: Análisis de varianza, efectos fijos y mixtos, errores estándares, bloques completos al azar, experimento factorial.

¹ Recepción: 21 de diciembre de 2020. Aceptación: 7 de mayo de 2021. Dirección Nacional de Recursos Genéticos y Biodiversidad. Proyecto de Investigación-Innovación en Mejoramiento Genético Aplicado en Hatos de los Sistemas Ganaderos Familiares Bovinos de Panamá. Instituto de Innovación Agropecuaria de Panamá (IDIAP).

² IDIAP, Centro de Innovación Agropecuaria de Chiriquí, Estación Experimental de Gualaca "Carlos M. Ortega". M.Sc. Mejoramiento Genético Animal. e-mail: pedroguerram16@gmail.com



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

MIXED MODELS FOR ANALIZING THREE FACTORS AGRICULTURAL EXPERIMENTS WITH A STATISTICAL SOFTWARE

ABSTRACT

PROC GLM of SAS® is the most used procedure for the analysis of variance in agricultural research, but it was designed for fixed models. The objective of this work was to capacitate SAS®'s users on the use of mixed models for the analysis of three factors agricultural experiments. Data came from literature and were analyzed by PROC GLM (fixed and random models) and PROC MIXED. Three factors and their interactions were: Legume (V), Phosphorus (P) and Weed Control (W). The design was CRB (three blocks, BLK) and the response variable was seed-yield (kg/plot). Routines of PROC GLM for fixed and mixed models and PROC MIXED were presented. Under mixed models, V and P were fixed and W was random effect. Degree of freedoms were adjusted by Satterthwaite procedure. With PROC GLM fixed model, BLK, V*W and P*W were $p < 0,01$ and with mixed model only was BLK, but Value-F's were minors. W and V*P*W variances were -0,59 and -0,22, respectively. With PROC MIXED, no one source of variation was $p < 0,05$ (F and Z Wald Test). W variance was zero while V*W and P*W variances were minors. V and V*P standard errors (SE) were higher with PROC MIXED and were equal for random effects in all procedures. It was concluded that PROC MIXED give adequate hypothesis test and SE of mean, so its utilization is recommended.

Key words: Analysis of variance, random and fixed effects, standard errors, completed random blocks.

INTRODUCCIÓN

Frecuentemente en la experimentación agropecuaria se le presta poca importancia a determinar la presencia o no de factores fijos y aleatorios³ en un experimento cruzado clasificado como el experimento factorial. Estos experimentos factoriales son analizados con PROC GLM de SAS® en la forma tradicional asumiendo un modelo de efectos fijos. Sin embargo, al tomar en cuenta la naturaleza de los factores, alguno de ellos resulta ser aleatorios y otros fijos cambiando el error experimental a ser usado en la prueba de hipótesis, lo que lleva al experimentador a expresar conclusiones erradas.

³ Efectos aleatorios: aquellos efectos que supuestamente son tomados al azar de una población de posibles niveles. Efectos fijos: aquellos que son determinados a propósito y que no representan a otros sino a ellos mismos (Correa-Morales y Salazar-Urbe, 2016).



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Generalmente, los experimentos factoriales son analizados con el procedimiento PROC GLM del paquete estadístico SAS® asumiendo un modelo de efectos fijos para facilitar el análisis. Sin embargo, Yang (2008) y Yang (2010) indica que al no tomar en cuenta esta situación, el PROC GLM brinda incorrecto denominador para la prueba-F, error estándar de los tratamientos y asume que todos los tratamientos tienen error idéntico y tales errores no están correlacionados. Sin embargo, autores como Gil (2001), Yang (2008) y Yang (2010) han reportados la situación antes mencionada y toman como ejemplo el diseño de parcelas divididas y comparan los análisis utilizando el PROC GLM *versus* PROC MIXED del paquete estadístico SAS®. Ambos autores, para modelos mixtos, recomiendan el PROC MIXED junto con técnicas que toman en cuenta el desbalance de datos y ajustes de los grados de libertad. El procedimiento recomendado proporciona los errores estándares adecuados a cada nivel de análisis, realizando las comparaciones de media en la forma correcta. Por otra parte, Yang (2008) señala que, si los efectos aleatorios verdaderos son pequeños y/o el tamaño de muestras es pequeño, los estimados de varianza negativos pueden ser obtenidos usando PROC GLM. Sin embargo, cualquier varianza por definición no debe ser menor a cero y que los estimados negativos no tienen significado. Con el REML (máxima verosimilitud restringida), el defecto de PROC MIXED establece el estimado negativo de varianza a cero independiente si los datos son balanceados o no. Esto guía a diferentes pruebas-F de los mismos efectos fijos. Agrega que dada (i) la dificultad de reportar varianzas negativas y (ii) errores estándares de la media y comparaciones de media obtenidas de PROC GLM, el uso de PROC MIXED es verdaderamente el preferido. Además, con PROC MIXED solamente se llevarían a cabo las pruebas-F para los efectos fijos para evaluar si las medias de tratamientos son diferentes, mientras que para los efectos aleatorios se utilizaría la prueba Z de Wald o la prueba Chi-Cuadrado de la Razón de Verosimilitud (Likelihood Ratio). Así con el PROC MIXED provee inferencias correctas acerca de efectos fijos y aleatorios, pero el uso de PROC GLM no (Yang, 2008; Yang, 2010).

Por otra parte, para la estimación de parámetros genéticos, Lowry (1983) ha propuesto una serie de reglas para escribir los parámetros estimados en el análisis de varianza tomando en cuenta un modelo mixto. Para este caso, Lowry (1983) indica tres reglas básicas en la estimación de los parámetros de los efectos: (1) la interacción de un efecto fijo con otro



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

efecto fijo no debe aparecer en la estimación del parámetro del efecto principal fijo (Modelo I Efectos Fijos), (2) la interacción de efectos aleatorios deben aparecer en la estimación de parámetros de cada efecto aleatorio (Modelo II Efectos Aleatorios), y (3) la interacción de un efecto fijo con un efecto aleatorio debe aparecer en la estimación del parámetro del efecto fijo, pero no debe aparecer en la estimación del parámetro del efecto aleatorio (Modelo Mixto).

Para los usuarios del SAS®, el presente trabajo tuvo como objetivo capacitar al usuario de SAS® en el uso de modelos mixtos en el análisis de experimentos factoriales (cruzados clasificados) de tres factores de datos provenientes de la experimentación agropecuaria.

MATERIALES Y MÉTODO

Datos utilizados: el conjunto de datos utilizados como ejemplo provienen de Kling (2004) del curso de Experimental Design in Agriculture of Oregon State University, USA (Cuadro 1) de un experimento factorial (tres factores) en un diseño de bloques completos al azar (BCA).

Cuadro 1. Producción de semilla (kg/parcela) de dos variedades de leguminosa forrajera (V) con tres niveles de P_2O_5 (kg/ha) (P) y con y sin control de malezas (W).

Combinación de tratamientos	Bloques			Totales
	Bloque I	Bloque II	Bloque III	
V1P1W1	61	58	55	174
V1P1W2	56	45	43	144
V1P2W1	60	59	54	173
V1P2W2	60	50	45	155
V1P3W1	59	55	51	165
V1P3W2	66	57	50	173
V2P1W1	65	61	60	186
V2P1W2	60	61	50	171
V2P2W1	60	58	56	174
V2P2W2	62	68	67	197
V2P3W1	53	53	48	154
V2P3W2	73	57	77	207
Totales	735	682	656	2073

V= Red Clover (V1) y Sweet Clover (V2); P= 0, 25 y 50 kg/ha P_2O_5 ; W= Sin control (W1) y con control (W2) de malezas. Fuente: Kling (2004).



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Análisis original de los datos experimentales: Los datos fueron analizados mediante el PROC GLM de SAS®. El diseño experimental usado por Kling (2004) fue el BCA en un arreglo factorial en donde el bloqueo consistió en el agrupamiento de las parcelas o unidades experimentales, a las cuales se aplicaron los tratamientos o efectos a estudiar en bloques para disminuir los efectos que puede causar una fuente de variación externa a los tratamientos dentro de las unidades experimentales y que distorsionen sus resultados e inferencias en caso de omitirlos. De acuerdo con Salazar-Guerrero (sf) y Steel y Torrie (1980), los bloques son grupos de unidades que se forman de tal manera que ellas dentro de los bloques sean tan homogéneas como sea posible.

Por otro lado, Steel y Torrie (1980) indican que en un experimento factorial, una interacción significativa puede causar la pérdida del interés en probar hipótesis concernientes a los efectos principales y llegar a estar más interesado en otras pruebas como las de los efectos simples y llegar a la mejor interpretación satisfactoria de los resultados. Sin embargo, en un modelo mixto, podría no ser de interés los efectos simples para un efecto fijo, ya que ellos serán medidos en niveles selectos aleatoriamente de otro factor y así serían valores de una variable aleatoria.

Modelo matemático del experimento factorial: el modelo matemático para el experimento factorial con tres factores de Kling (2004) describe los efectos de V (Variedad), P (Nivel de fósforo) y W (Control de malezas) y es el siguiente:

$$Y_{ijklm} = \mu + \rho_i + \alpha_j + \beta_k + \gamma_l + (\alpha\beta)_{jk} + (\alpha\gamma)_{jl} + (\beta\gamma)_{kl} + (\alpha\beta\gamma)_{jkl} + \varepsilon_{ijklm}$$

Donde:

Y_{ijklm} = es la m-ésima observación en el i-ésimo bloque, j-ésimo nivel del factor α , k-ésimo nivel del factor β y l-ésimo factor γ .

μ = media general,

ρ_i = efecto del i-ésimo bloque,

α_j = efecto del j-ésimo nivel del factor variedad,

β_k = efecto del k-ésimo nivel del factor fósforo,



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

γ_l = efecto del l-ésimo nivel del factor control de malezas,

$(\alpha\beta)_{jk}$ = efecto de la interacción entre el j-ésimo nivel de variedad con el k-ésimo nivel del factor fosforo,

$(\alpha\gamma)_{jl}$ = es el efecto de la interacción entre el j-ésimo nivel de variedad con el l-ésimo nivel de control de malezas,

$(\beta\gamma)_{kl}$ = es el efecto de la interacción entre el k-ésimo nivel del factor fósforo con el l-ésimo nivel del factor control de malezas,

$(\alpha\beta\gamma)_{jkl}$ = es el efecto de la triple interacción entre el j-ésimo nivel de variedad, k-ésimo nivel del factor fósforo y l-ésimo nivel del factor control de malezas,

ε_{ijklm} = es el error aleatorio ($\sim NID$, $\mu=0$, $\sigma^2=1$).

Tradicionalmente, el modelo lineal mixto generalizado se escribe en forma de matrices de la siguiente manera (Lair y Ware, 1982):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\omega} + \boldsymbol{\varepsilon}$$

Donde:

$\mathbf{y}$ = vector de observaciones o respuestas,

$\boldsymbol{\beta}$ = vector de parámetros de efectos fijos desconocidos,

$\mathbf{X}$ =matriz de incidencia que relaciona las observaciones fenotípicas con el vector $\boldsymbol{\beta}$ de efectos fijos (V y P),

$\boldsymbol{\omega}$ = vector de parámetros de efectos aleatorios desconocidos,

$\mathbf{Z}$ = matriz de incidencia que relaciona las observaciones fenotípicas con el vector $\boldsymbol{\omega}$ de efectos aleatorios (W),

$\boldsymbol{\varepsilon}$ = vector de errores aleatorios residuales, normal e independientemente distribuidos.

Este procedimiento está basado en determinaciones probabilísticas máximas o restringidas de modelos lineales estadísticos que contemplan términos fijos y aleatorios. Los errores estándar (EE) fueron derivados del diseño experimental, fuentes de variación asumidas y sus componentes de varianza (Littell et al., 2006). Los EE para PROC GLM y PROC MIXED fueron calculados con la instrucción LSMEANS/STDERR;



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Representación de las Expectativas de los Cuadrados Medios del modelo fijo y aleatorios: los promedios o valores esperados de los cuadrados medios en un experimento de tres factores fijos, en un diseño de bloques completos al azar para el modelo fijo están dados en el Cuadro 2 (Steel y Torrie, 1980). Las letras mayúsculas en la columna de Fuente de Variación se refieren a los efectos principales e interacciones y las letras minúsculas, en la columna de los grados de libertad, se refieren a los números de los niveles de los efectos designados por la correspondiente letra mayúscula. Por otra parte, las letras griegas, en la columna de las esperanzas de los cuadrados medios, se refieren a componentes de varianzas de cada efecto e interacción.

Cuadro 2. Valores esperados de los cuadrados medios (ECM) con el modelo fijo para experimentos factoriales de tres-factores.

F de V	gl	Valores Esperados de los Cuadrados Medios (CM)	
		Modelo I (Fijo)	Modelo II (aleatorio)
Bloque	r-1	$\sigma^2 + abc \sum p^2 / (r-1)$	$\sigma^2 + abc\sigma^2_p$
V	a-1	$\sigma^2 + rbc \sum \alpha^2 / (a-1)$	$\sigma^2 + r\sigma^2_{\alpha\beta\gamma} + rc\sigma^2_{\alpha\beta} + rb\sigma^2_{\alpha\gamma} +$
P	b-1	$\sigma^2 + rac \sum \beta^2 / (b-1)$	$rb\sigma^2_{\alpha\gamma}$
W	c-1	$\sigma^2 + rab \sum \gamma^2 / (c-1)$	$\sigma^2 + r\sigma^2_{\alpha\beta\gamma} + rc\sigma^2_{\alpha\beta} + ra\sigma^2_{\beta\gamma} +$
V*P	(a-1)(b-1)	$\sigma^2 + rc \sum (\alpha\beta)^2 / (a-1)(b-1)$	$rac\sigma^2_{\beta}$
V*W	(a-1)(c-1)	$\sigma^2 + rb \sum (\alpha\gamma)^2 / (a-1)(c-1)$	$\sigma^2 + r\sigma^2_{\alpha\beta\gamma} + rb\sigma^2_{\alpha\gamma} + ra\sigma^2_{\beta\gamma} +$
P*W	(b-1)(c-1)	$\sigma^2 + ra \sum (\beta\gamma)^2 / (b-1)(c-1)$	$rac\sigma^2_{\gamma}$
V*P*W	(a-1)(b-1)(c-1)	$\sigma^2 + r \sum (\alpha\beta\gamma)^2 / (a-1)(b-1)(c-1)$	$\sigma^2 + r\sigma^2_{\alpha\beta\gamma} + rc\sigma^2_{\alpha\beta}$
Error Residual	(r-1)(abc-1)	σ^2	$\sigma^2 + r\sigma^2_{\alpha\beta\gamma} + rb\sigma^2_{\alpha\gamma}$
			$\sigma^2 + r\sigma^2_{\alpha\beta\gamma} + rb\sigma^2_{\beta\gamma}$
			$\sigma^2 + r\sigma^2_{\alpha\beta\gamma}$
			σ^2

Fuente: Steel y Torrie, 1980.

Los elementos σ^2 , $\sum \alpha^2$, $\sum \beta^2$, $\sum \gamma^2$, $\sum (\alpha\beta)^2$, $\sum (\alpha\gamma)^2$, $\sum (\beta\gamma)^2$, $\sum (\alpha\beta\gamma)^2$, σ^2_{α} , σ^2_{β} , σ^2_{γ} , $\sigma^2_{\alpha\beta}$, $\sigma^2_{\alpha\gamma}$, $\sigma^2_{\beta\gamma}$, y $\sigma^2_{\alpha\beta\gamma}$ son los componentes de varianza para el modelo fijo. Estos componentes en los efectos aleatorios son importantes estimarlos para los cálculos de los parámetros genéticos como heredabilidades y correlaciones genéticas. En la mayoría de los experimentos agropecuarios se asume que los efectos son todos fijos en el modelo matemático utilizado. De ser así, en este caso (Modelo I) se observa que la varianza del error (σ^2) es un término apropiado para probar hipótesis de cualquier fuente de variación en el análisis de varianza (Steel y Torrie, 1980). Pero para el modelo aleatorio (Modelo II),



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

de acuerdo con el Cuadro 2, el Error Residual es el apropiado para probar el efecto de bloque y la triple interacción. Si $\sigma^2_{\alpha\beta\gamma}$ es real, entonces el CM(V*P*W) es el apropiado para probar las dobles interacciones. Estos autores agregan que para probar los efectos principales es necesario realizar algunos agrupamientos de CM. Por ejemplo, para probar $H_0: \sigma^2_{\gamma} = 0$, CM(W) sería el numerador y $CM(V*W) + CM(P*W) - CM(V*P*W) - 89e^{-17}(CM \text{ Error})$ sería el denominador o $CM(W) + CM(V*P*W)$ puede ser usado como numerador y $CM(V*W) + CM(P*W)$ como el denominador.

Error experimental apropiado en el modelo aleatorio: para las pruebas de hipótesis con los modelos aleatorios, la escogencia de un término de error adecuado, cuando todas las fuentes de variación son reales, es más difícil cuando las hipótesis concernientes a los efectos principales van a ser probados.

Consideraciones del modelo mixto: de acuerdo con Yang (2008) y Yang, (2010), la determinación si un efecto es fijo o aleatorio no es siempre fácil y ha sido debatido en la literatura científica. En la mayoría de los casos, los efectos de los tratamientos son considerados fijos porque los tratamientos en el experimento son los únicos sobre los cuales se realizarán inferencias y son seleccionados por el experimentador. Los efectos del bloque se consideran como efectos aleatorios porque los bloques en el experimento son sólo un pequeño subconjunto de un universo de bloques sobre los cuales se realizarán inferencias acerca de los tratamientos (Salazar-Guerrero, sf; Steel y Torrie, 1980). Así, el modelo para datos de BCA usualmente contiene efectos fijos por las contribuciones de los tratamientos y efectos aleatorios por las contribuciones de los bloques, conformándose un modelo mixto.

Análisis con el Modelo Mixto: para fines ilustrativos se tomaron los datos de Kling (2004) en donde V y P se consideraron como efectos fijos y W como un efecto aleatorio. De acuerdo con Gil (2001), este análisis con el modelo mixto está basado en estimaciones probabilísticas máximas o restringidas de modelos lineales estadísticos que contemplan términos fijos y aleatorios y es considerado la mejor opción para el análisis estadístico de este tipo de experimento (Littell et al., 1991; Littell et al., 2006; Yang, 2008).



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Representación de las Expectativas de los Cuadrados Medios del modelo mixto: el Cuadro 3 contiene las expectativas de los cuadrados medios (ECM) del modelo mixto, utilizados para calcular los componentes de varianza y considerar el error experimental apropiado para las pruebas de hipótesis.

Cuadro 3. Valores esperados de los cuadrados medios (ECM) con el modelo mixto para experimentos factoriales: el experimento de tres-factores.

Modelo mixto: V y P son fijos; W es aleatorio	
Bloque	$\sigma^2 + abc\sigma_p^2$
V	$\sigma^2 + rb \frac{a}{a-1} \sigma_{\alpha\gamma}^2 + rbc \sum \alpha^2 / (a-1)$
P	$\sigma^2 + ra \frac{b}{b-1} \sigma_{\beta\gamma}^2 + rac \sum \beta^2 / (b-1)$
W	$\sigma^2 + rab\sigma_{\gamma}^2$
V*P	$\sigma^2 + r \frac{a}{a-1} \frac{b}{b-1} \sigma_{\alpha\beta\gamma}^2 + rc \sum (\alpha\beta)^2 / (a-1)(b-1)$
V*W	$\sigma^2 + rb \frac{a}{a-1} \sigma_{\alpha\gamma}^2$
P*W	$\sigma^2 + ra \frac{b}{b-1} \sigma_{\beta\gamma}^2$
V*P*W	$\sigma^2 + r \frac{a}{a-1} \frac{b}{b-1} \sigma_{\alpha\beta\gamma}^2$
Error	σ^2

Fuente: Steel y Torrie, 1980.

Errores experimentales del Modelo Mixto: tomando en cuenta el Cuadro 3, para las pruebas de hipótesis de efectos aleatorios y de las interacciones entre uno o dos factores fijos con un factor aleatorio, la interacción es considerada como aleatoria como es el caso de W, W*P, P*W y V*P*W. Para estas interacciones, el término de error (σ^2) es usado para realizar la prueba de F. Sin embargo, para la interacción V*P, entre dos efectos fijos, el término de error adecuado es la triple interacción o V*P*W. Para probar las hipótesis de los efectos fijos principales como V y P es más complejo. Por ejemplo, para el efecto de V sería:

$$F_V = \frac{CM_V + CM_{V*P*W}}{CM_{V*P} + CM_{V*W}} = \frac{2\sigma^2 + 2r\sigma_{V*P*W}^2 + rb\sigma_{V*W}^2 + rc\sigma_{V*P}^2 + rbc\sigma_V^2}{2\sigma^2 + 2r\sigma_{V*P*W}^2 + rb\sigma_{V*W}^2 + rc\sigma_{V*P}^2}$$

Con esta fórmula se tiene el apropiado numerador y denominador para probar la hipótesis del factor V.



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Ajuste de grados de libertad: Satterthwaite (1946) y Gaylor y Hopper (1969) propusieron ajustar los grados de libertad en el modelo mixto donde CM_i representa un cuadrado medio y cualquier CM_i debe no aparecer tanto en el numerador como en el denominador. Tales relaciones son también llamados *quasi* F. El criterio de prueba es distribuido aproximadamente como F. Para este caso de un experimento factorial con tres factores: V, P y W, en un arreglo de BCA, y donde W es el factor aleatorio, la prueba de F sería:

$$F_{p,q} = \frac{CMr + \dots + CMs}{CMu + \dots + CMv}$$

Donde p y q son los grados de libertad para el numerador y denominador, respectivamente

$$p = \frac{(CMr + \dots + CMs)^2}{(CM^2r/fr + \dots + CM^2s/fs)}$$

$$(CM^2r/fr + \dots + CM^2s/fs)$$

$$q = \frac{(CMu + \dots + CMv)^2}{(CM^2u/fu + \dots + CM^2v/fv)}$$

f_i son los grados de libertad para CM_i . Aquí p y q son llamados grados de libertad “efectivos”. Además, CMr es el cuadrado medio de la variable de interés en la prueba F, en este caso es V; CMs es el cuadrado medio de la triple interacción; CMu es el cuadrado medio de la doble interacción de V*P y CMv es el cuadrado medio de la doble interacción V*W.

Procedimientos de análisis: los siguientes procedimientos o rutinas están diseñados para los usuarios de SAS®, así para:

1 - Análisis como modelo fijo: primeramente, asumiendo que los factores V, P y W son considerados como efectos fijos, el procedimiento con PROC GLM sería el siguiente:

```
data uno;
input id blk v p w rend;
cards;
1 1 1 0 1 61
```



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

```
2 1 1 0 2 56
3 1 1 25 1 60
```

Etc.

-colocar datos del cuadro 1 en columnas indicadas por el orden del input-

```
;
proc sort;
by blk v p w;
run;
proc glm;
class id blk v p w;
model rend = blk v|p|w; [automáticamente se estiman los efectos principales e
                           interacciones]
*****determinacion de contrastes ortogonales*****;
contrast 'v1 vs v2' v 1 -1; [contrastes para probar diferencias entre
                              variedades]
contrast 'v1 vs v2 con control de malezas' v 1 -1 v*p 0 1 0 -1; [prueba v1
                                                                    contra v2 con control de malezas]
*****comparacion de medias ajustadas por el modelo y error estándar*****;
lsmeans v p v*p w v*w p*w/s; [pide las medias ajustadas y errores estándar de la
                               media de los efectos principales y sus interacciones
                               y las compara todas contra todas mediante pruebas de t]
```

```
run;
quit;
```

2 - Análisis como modelo mixto con PROC GLM: primeramente, asumiendo que los factores V y P son fijos y el efecto W es aleatorio, el procedimiento con PROC GLM sería el siguiente:

```
data dos;
input id v p w blk rend;
cards;
1 1 1 0 1 61
2 1 1 0 2 56
3 1 1 25 1 60
```

Etc.

-colocar datos del Cuadro 1 en columnas indicadas por el orden del input-

```
;
proc sort;
by id blk v p w;
run;
proc glm;
class id blk v p w;
model rend = blk v|p|w;
random blk c v*w p*w v*p*w/test; [pide prueba de las variables aleatorias]
*****prueba de hipótesis*****; [prueba de efectos principales con su
                                    adecuado error experimental]

test h=v e=v*w;
test h=p e=p*w;
test h=v*p e=v*p*w;
```



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

```

***prueba de contrastes ortogonales de efectos fijos con sus
errores****;
contrast 'v1 vs v2' v 1 -1 /e=v*w;
contrast 'v1 vs v2 con control de malezas' v 1 -1 v*w 0 1 0 -1; [compara
0 contra 50 kg P2O5 con control de malezas]
*****comparaciones de medias de efectos principales*****;
lsmeans v/s e=v*w; [compara las medias ajustadas con su adecuado error
experimental y pide el error estándar]
lsmeans p/s e=p*w;
lsmeans w/s;
***comparacion de medias de las interacciones con su error*****;
lsmeans v*p/s e=v*p*w;
lsmeans v*w/s;
lsmeans p*w/s;
run;
quit;

```

3 - Análisis como modelo mixto con PROC MIXED: este procedimiento utiliza como defecto (default) la técnica de Máxima Verosimilitud Restringida (REML) y es muy preferido con datos desbalanceados sobre el método ANOVA (Yang, 2008 y Yang, 2010) y calcula correctamente efectos fijos y aleatorios. Asumiendo, nuevamente, que los factores V y P son fijos y el efecto W es aleatorio, el procedimiento con PROC MIXED sería el siguiente:

```

data tres;
input id blk v p w rend;
cards;
1 1 1 0 1 61
2 1 1 0 2 56
3 1 1 25 1 60
Etc.
-colocar datos del cuadro 1 en columnas indicadas por el orden del input-
;
proc sort;
by id blk v p w;
run;
proc mixed data=tres method=reml covtest; [covtest realiza la prueba de Wald para los
efectos aleatorios]
class id blk v p w;
model rend = v p v*p/ddfm=satterth; [ajusta los gl de acuerdo a
Satterthwaite, 1946]
random blk w v*w p*w v*p*w;
****estimación de medias y errores estándares de efectos fijos****;
estimate 'media variedad v1' intercept 1 v 1 0; [estima la media de V1 y error
estándar]
estimate 'media 50 kg-ha P2O5' intercept 1 p 0 1 0;
estimate 'variedad v1 vs v2 con 50 kg fosforo' v 1 -1 v*p 0 0 1 0 0 -1;
****comparación de medias con prueba de contrastes ortogonales****;
contrast 'v1 vs v2' v 1 -1; [contrasta V1 contra V2]
contrast 'v1 vs v2 con 50 kg-ha fosforo' v 1 -1 v*p 0 0 1 0 0 -1;
[contrasta V1 con V2 a un nivel de 50 kg-ha P2O5]

```



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

```
*****tambien estima medias ajustadas para efectos fijos*****;
lsmeans v/EST;
lsmeans p/EST;
lsmeans v*w/EST;
run;
quit;
```

Comparación de las tres rutinas de SAS®: para probar la eficiencia en el análisis de los datos con las tres rutinas de SAS® se compararon los Errores Estándares (EE) de las medias e interacciones. Para obtener los errores estándares de las medias e interacciones del modelo fijo y mixto con PROC GLM, además, de las instrucciones para el análisis, se añadieron las instrucciones (statements) para solicitar contrastes ortogonales (contrast) y comparar diferencias de medias y/o estimados (estimate) y así obtener las medias y sus errores estándares, tanto en efectos fijos como en aleatorios. También, en las dos primeras rutinas con PROC GLM se utilizó la instrucción “lsmeans” para obtener las medias ajustadas por el modelo y su error estándar, mientras que en el modelo mixto con PROC MIXED se utilizó la instrucción “estimate” para obtener la media ajustada y su error estándar de los efectos fijos. Para la diferencia de dos medias de efectos fijos se utilizó la instrucción “contrast”. Además, en el PROC MIXED se utilizó la instrucción “covtest” para realizar la prueba de Wald para efectos aleatorios (refiérase a Correa-Morales y Salazar-Urbe, 2016), la cual está basada en errores estándares que subestiman la verdadera variabilidad en la estimación de β debido a que estas no tienen en cuenta la variabilidad introducida al estimar a α . Las hipótesis serían $H_0: L\beta=0$ y $H_a: L\beta \neq 0$. L es cualquier matriz conocida y β_i es un vector q -dimensional con los coeficientes de regresión sujeto-específicos y

$$W = (\hat{\beta} - \beta)' L' [L (\sum_{i=1}^m X' i V i (\hat{\alpha}) X i)]^{-1} L' (\hat{\beta} - \beta)$$

RESULTADOS Y DISCUSIÓN

Se asumió un modelo fijo en donde el término de error se consideró como el apropiado para probar las hipótesis sobre los efectos principales y sus interacciones (Cuadro 4).



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Cuadro 4. Análisis de varianza con el PROC GLM y PROC MIXED de SAS® de la producción de semilla (kg/parcela) de dos variedades de leguminosa forrajera (V) con tres niveles de P₂O₅ (P) y con y sin control de malezas (W), usando en modelo lineal generalizado fijo y mixto.

F de V	gl	PROC GLM				PROC MIXED ¹			
		CM	FIJO ^a Valor-F	MIXTO Valor-F	σ^2 estimada	Gl Num	gl Den	Valor-F ^g Valor-Z ^h	σ^2 estimada
Bloques, BLK	2	135,08	5,93**	5,93** ^b	9,36	2	22	0,83	9,36
Variedades (V)	1	306,25	13,44**	1,08 ^c	.	1	1,74	1,64 ^g	.
Fósforo (P)	2	16,00	0,70	0,07 ^d	.	2	2,72	0,09 ^g	.
Malezas (W)	1	12,25	0,54	0,02 ^e	-0,59	.	.	.	0,00
VxP	2	9,33	0,41	0,42 ^f	.	2	1	0,41 ^g	.
VxW	1	283,36	12,44**	12,82 ^f	28,95	.	.	0,81 ^h	18,25
PxW	2	234,33	10,29**	10,60 ^f	35,26	.	.	1,01 ^h	27,28
VxPxW	2	22,11	0,97	0,97 ^b	-0,22	.	.	0,00 ^h	0,02
Error	22	22,78	.	.	22,78	.	.	3,32 ^h	22,78

**P<0,01

¹Valores-F de efectos fijos y estimados de varianza de efectos aleatorios.

^aFuente: Kling (2004).

Términos de errores (denominador) usados en la Prueba-F de PROC GLM de SAS®

^bError: CM(Error)

^cError: CM(VxW)

^dError: CM(PxW)

^eError: CM(VxW) + CM(PxW) – CM(VxPxW) – 89e⁻¹⁷(CM Error); gl error = 2,2742

^fError: CM(VxPxW)

^gValor-F para efectos fijos.

^hValor-Z para efectos aleatorios

Con PROC GLM de SAS® para el modelo fijo (Cuadro 4), se observa que los Valores-F se estimaron con el CM del error (residual) y basado en el Cuadro 2. Estos resultados muestran que los efectos BLK, V, VxW y PxW resultaron ser altamente significativos ($p < 0,01$). Las interacciones VxW y PxW toman mayor relevancia en la discusión de los resultados, principalmente se debe enfocar en analizar los efectos simples y determinar cuál de ellos fue(ron) el(los) de mayor aporte a la significancia de las interacciones.

Usando un modelo mixto con PROC GLM de SAS®, los Valores-F se calcularon de acuerdo a la estructura del Cuadro 2 y con los términos de error mostrados al pie del Cuadro 4. Los efectos de BLK y VxPxW se calcularon con el término de error residual. Las interacciones V, VxW y PxW ahora no mostraron una significancia ($p > 0,05$). También, se observa que para la Prueba-F del efecto aleatorio W (modelo aleatorio del Cuadro 2), su denominador es un término compuesto que usa PROC GLM de SAS® para modelos mixtos



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

definido por la fórmula: $\text{Error} = \text{CM}(\text{VxW}) + \text{CM}(\text{PxW}) - \text{CM}(\text{VxPxW}) - 89e^{-17}(\text{CM Error})$ con g.l. error = 2,2742 ajustado por Satterthwaite (1946). Sin embargo, basado en el Cuadro 3, el CM del término del error Residual (σ^2) sería el más apropiado para la Prueba-F y su valor debió ser 0,43.

En contraste con el modelo fijo, se observó que con el modelo mixto de PROC GLM no se detectaron diferencias significativas ($p > 0,05$) en el efecto V. Por el contrario, en el modelo fijo, el efecto V resultó altamente significativo ($p < 0,01$), llegándose a conclusiones erradas, en caso de que el modelo fuera en realidad mixto. Además, el PROC GLM de SAS® con el modelo mixto las interacciones VxW y PxW son probadas (Prueba-F) con el CM de la triple interacción (VxPxW), mientras que basado en el modelo mixto del Cuadro 3 se debiera probar adecuadamente con el término de error σ^2 . Los valores de F debieran ser 12,44 y 10,29, respectivamente.

Con el PROC MIXED de SAS®, los Valor-F de V y P, ajustado por los grados de libertad de 1,74 y 2,72, respectivamente, de acuerdo Satterthwaite, (1946), fueron ligeramente mayores que en PROC GLM, pero aún no resultaron significativos ($p > 0,05$) (Cuadro 4). Los efectos aleatorios de BLK y V, sus Valor-Z no fueron estimados, en este ejemplo, pero las interacciones VxW, PxW y VxPxW, sus Valor-Z fueron estimados, pero no resultaron significativos ($p > 0,05$).

La varianza del efecto aleatorio W fue estimada; así como las interacciones donde este efecto participa con el modelo mixto de PROC GLM de SAS® (Cuadro 4). Por otra parte, las varianzas de W y VxPxW resultaron negativas. Con el PROC MIXED las varianzas de BLK y W fueron 0,00; sin embargo, las varianzas de VxW, PxW y VxPxW fueron positivas y de VxW y PxW fueron ligeramente menores que las estimadas por PROC GLM.

Con el PROC MIXED se pueden realizar las Pruebas-F para efectos fijos y así evaluar si las medias de tratamiento son diferentes y la Prueba-Z para determinar estimados de varianzas efectos aleatorios e inferir acerca de ellas. Esta situación favorece



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

el uso de PROC MIXED para proveer inferencias correctas acerca de efectos fijos y aleatorios, pero no así para el PROC GLM.

Los Cuadros 2 y 3 guían la estimación de las varianzas de los efectos aleatorios y el cálculo de los Valor-F de los efectos fijos. Las varianzas estimadas por PROC MIXED de SAS® son ligeramente menores que las estimadas por PROC GLM de SAS® con modelo mixto; además no existen varianzas negativas, las cuales no tienen significado alguno (Yang, 2008). Para el PROC GLM, debido a que el CM de W o Control de Malezas (12,25) es menor que el del CM del Error (22,78), la σ^2_W resultó ser -0,59 [$\sigma^2_W = (12,25 - 22,78)/18 = -0,59$]. Sin embargo, el defecto (“default”) del método del análisis PROC MIXED es REML (Máxima Verosimilitud Restringida) establece que este indeseable valor negativo es estimado a ser cero. Además, establecer $\sigma^2_W = 0$ afecta el estimado REML de $\sigma^2_{V \times W}$ y $\sigma^2_{P \times W}$, el cual también afecta los Valores-F de los efectos V y P; así como los contrastes y estimados de medias o diferencias de medias (Yang, 2008).

Los Valores-F del PROC MIXED de los efectos fijos fueron calculados de los componentes de varianza. Así, para F_V se procedió así: $CM_{V \times W} = \sigma^2 + rb\sigma^2_{V \times W} = 22,78 + (3)(3)(18,25) = 187,03$. $F_V = CM_V / CM_{A \times W} = 306,25/187,03 = 1,64$. Similarmente se procede para F_P y $F_{V \times P}$. SAS® establece los valores de F tabulada con los grados de libertad ajustados por Satterthwaite (1946).

Los Valores-Z (Cuadro 4) para los efectos aleatorios del PROC MIXED indican que no existe diferencias significativas en sus efectos ($p > 0,05$). No se obtuvo Valor-Z para el efecto de W, ya que su varianza en el modelo fijo resultó ser negativa y PROC MIXED lo ubica en cero (0,00).

Los errores estándares de acuerdo con los tres procedimientos usando el programa SAS® se muestran en el Cuadro 5.

Los errores estándares (EE) con el modelo fijo de PROC GLM se estimaron todos con el CM del error residual ($\sigma^2 = 22,78$) y variaron de 1,125 con V a 2,756 con $V \times P \times W$; sin embargo, con el modelo mixto de PROC GLM, los EE de los efectos fijos de V, P y su



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

interacción VxP variaron de 1,920 en VxP a 4,419 en P. Todos estos EE fueron estimados con su correspondientes y verdaderos CM. Un aumento de 253,7%. 220,7% se registró en V y P, respectivamente y una disminución de 1,5% en V*P al compararlos con PROC GLM modelo fijo. Además, los EE del PROC MIXED para los efectos fijos, variaron de 4,249 en V a 5,449 en la interacción VxP. Los aumentos del EE en V, P y VxP fueron 7,1%, 9,1% y 183,8%, respectivamente, con respecto al modelo mixto de PROC GLM. Por otra parte, los EE de los efectos aleatorios fueron iguales en ambos modelos del PROC GLM y modelo fijo del PROC MIXED, solo que no fue mostrado en los resultados del análisis (output). En un experimento factorial de dos factores, uno fijo y otro aleatorio, Yang (2008) reportó también Errores Estándares mucho mayores con PROC MIXED que en PROC GLM para modelos fijos. Un aumento de 28,4% en los EE de los niveles del efecto fijo y 28,1% en las comparaciones de medias de los niveles del efecto fijo.

Cuadro 5. Errores estándares de las fuentes de variación de acuerdo con el tipo de modelo y analizados por SAS®.

F de V	PROC GLM	PROC GLM	PROC MIXED
	FIJO	MIXTO	
Variedad (V)	1,125	3,968	4,249
Fósforo (P)	1,378	4,419	4,819
VxP	1,949	1,920	5,449
Control de Malezas (W)	1,125	1,125	-
VxW	1,591	1,591	-
PxW	1,948	1,948	-
VxPxW	2,756	2,756	-

PROC MIXED correctamente determina los errores estándares de la media y las diferencias entre medias, mientras que PROC GLM produce valores errores incorrectos. Este procedimiento produce y acomoda más realísticamente la estructura de varianza-covarianza de efecto aleatorios. Esto se debe a que el PROC GLM considera todos los efectos e interacciones como efectos fijos y solamente el término de error es considerado como aleatorio (Yang, 2008).

Yang (2008) en su nota al editor “*Why is MIXED análisis underutilized*”, indicó lo siguiente, el PROC GLM de SAS® es el más utilizado en el análisis de varianza; sin



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

embargo, ha sido diseñado para analizar datos conteniendo solo efectos fijos. Por lo tanto, recomienda el PROC MIXED como el más adecuado para datos con efectos fijos y aleatorios. Además, agrega qué si los efectos aleatorios verdaderos son pequeños y/o los tamaños de la muestra son pequeños, estimados negativos de varianza pueden ser obtenidos usando PROC GLM. Sin embargo, cualquiera varianza por definición no debiera ser menor a cero y que los estimados negativos no tienen significado. Cuando esto ocurre, la máxima verosimilitud restringida o REML (defecto de MIXED) establece los estimados de varianza negativos a cero, sin importar si los datos están balanceados o no y esto guía a diferentes Pruebas-F de los mismos efectos fijos. Esta es una de las razones por la que PROC MIXED debe ser utilizado en modelos mixtos.

CONCLUSIONES

- La comparación entre PROC GLM y PROC MIXED en el análisis de un experimento cruzado - clasificado de tres factores con un factor aleatorio y dos factores fijos, se reafirma lo encontrado en otros estudios con otros modelos estadísticos, que PROC MIXED es capaz de producir Pruebas-F y errores estándares correctos para medias de efectos fijos, mientras que el PROC GLM no los produce.
- El uso del PROC GLM de SAS® puede estimar varianzas negativas en variables aleatorias y las interacciones como ocurrió en este estudio. Varianzas negativas no tienen ningún significado e interpretación como en esta variable de respuesta (rendimiento).
-

RECOMENDACIONES

- Con el apoyo de un biometrista, el investigador agropecuario debe estar seguro si hay o no presencia de efectos fijos y/o aleatorios.
- Utilizar siempre el PROC MIXED de presentarse un experimento en donde se combinen efectos fijos y aleatorios.



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

REFERENCIAS

- Correa-Morales, J.C., y Salazar-Uribe, J.C. (2016). Introducción a los modelos mixtos. 1ª Edición. Centro Editorial. Escuela de Estadística. Facultad de Ciencias. Universidad Nacional de Colombia. Sede Medellín, Colombia.
- Gaylor, D.W., y Hopper, F.N. (1969). Estimating the degrees of freedom for linear combinations of mean squares by Satterthwaite's formula. *Technometrics* 11, 691-706.
- Gil, J.L. (2001). Comparación de los procedimientos GLM y MIXED del SAS® para analizar parcelas divididas con bloques al azar. *Zootecnia Tropical* 19(1), 43-58.
- Kling, J. (2004). Field plot techniques and experimental design in agriculture. Oregon State University. Corvallis, OR. USA.
- Lair, N.M., y Ware, J.H. (1982). Random effects models for longitudinal data. *Biometrics*, 38, 963-974.
- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., y Schabenberger, O. (2006). SAS for mixed models. 2nd Edition. SAS Institute Inc. Cary, NC. USA.
- Littell, R.C., Freund, R.J., y Spector, P.C. (1991). SAS system for linear models. 3rd Ed. SAS Institute, Cary, NC. USA.
- Lowry, S. (1983). Rules for thumb for writing parameters estimated in the analysis of variance. Monography. In: Design of Experiment. Agriculture 802. University of Nebraska at Lincoln. NE. USA. 9p.
- Salazar-Guerrero, A. sf. Estadística superior. Clave LII. Consultado 5 sept. 2020. www.cursos.aiu.edu/Estadistica%20Superior/PDF/Tema%202.pdf



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. Biometric Bulletin 2, 110-114.

Steel, R.G.D., Torrie, J.H. (1980). Principles and procedures of statistics. A biometrical approach. 2nd Edition. McGraw-Hill, Inc. New York. USA.

Yang, R-C. (2008). ¿Why is MIXED analysis underutilized? Canadian Journal of Plant Science 88, 563-567.

Yang, R-C. (2010). Towards understanding and use of mixed-model analysis of agricultural experiments. Canadian Journal of Plant Science 90, 605-627.



Este trabajo está licenciado bajo una [licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/)