Estructuras factoriales aumentadas en ensayos de investigación: una aplicación en palma de aceite*

Enhanced Factor Structures in Research Trials: An Application in Oil Palm

Citación: Mesa-Fuquen, E., Molina, D. L., Rincón-Numpaque, A., Ruiz, E., Fontanilla, C. A., & Fernández, C. A. (2019). Estructuras factoriales aumentadas en ensayos de investigación: una aplicación en palma de aceite. *Palmas*, *40*(1), 36-45.

Palabras clave: contrastes ortogonales, comparaciones múltiples, diseño de tratamientos, nutrimentos.

Keywords: Orthogonal contrasts, multiple comparisons, treatment design, nutrients.

Recibido: noviembre de 2018.

Aprobado: diciembre de 2018.

* Artículo de investigación científica y tecnológica.

Eloina Mesa Fuquen
Investigadora Asociada. División de Validación, Cenipalma

Diego Luis Molina López
Investigador Asociado. Programa de Agronomía, Cenipalma

Álvaro Rincón Numpaque
Asistente de Investigación. Unidad de Investigación e Innovación, Cenipalma

Elizabeth Ruiz Álvarez
Asistente de Investigación. Economía Agrícola, Cenipalma

Carlos Andrés Fontanilla Díaz
Investigador Asociado II de Cenipalma (cargo que desempeñaba a la fecha de su retiro: 17/08/2016)

Carlos Andrés Fernández Padilla
Auxiliar de Investigación II (cargo que desempeñaba a la fecha de su retiro: 22/04/2016)

Resumen

En investigación agrícola, y en particular en cultivos perennes como la palma de aceite, frecuentemente se realizan ensayos con estructuras factoriales completas, en las cuales todos los niveles de cada factor analizado se combinan entre sí para conformar la estructura de tratamientos. Sin embargo, algunas veces es necesario considerar tratamientos adicionales, generalmente testigos, diseñando ensayos con estructuras factoriales aumentadas, es decir, la estructura básica del experimento más los tratamientos adicionales, los cuales generalmente corresponden a testigos absolutos o relativos. En el primer caso, se considera la no aplicación de tratamiento, mientras que en el segundo se incorpora un tratamiento local de la zona de estudio. Para este tipo de ensayos surge el interrogante de cómo analizar los datos de tal manera que no solo se evalúen

los tratamientos en general como una única fuente de variación, sino que, por el contrario, se consideren y evalúen los efectos principales de cada factor y sus posibles interacciones. Por ello, el objetivo de este trabajo es presentar una alternativa de análisis de datos que consiste en la realización de una descomposición de las sumas de cuadrados de tratamientos en comparaciones de interés para el investigador, a partir de contrastes ortogonales que involucren comparaciones de efectos principales, efectos de interacción y comparaciones que involucren los tratamientos adicionales aplicados al cultivo de palma de aceite.

Abstract

Trials with complete factor structures, where all the levels of each factor being studied are combined with each other to create the treatment structure, are frequently carried out in agricultural research, particularly on perennial crops, such as oil palm. However, one or more additional treatments must be considered sometimes. These are generally control subjects, with whom trials with enhanced factor structures are formed: the basic structure of the experiment + the additional processing, which usually corresponds to absolute or relative control subjects. In the first case, we consider the non-application of the treatment, and in the second, a local treatment of the area under study. These trials pose the question of how to analyze the data so that the treatments in general are not only evaluated as a single source of variation but, on the contrary, to consider and evaluate the main effects of every factor and their possible interactions. The objective of this paper is to present a data analysis alternative, which consists in factorizing the sum of squares for treatments in comparisons that are relevant to the researcher based on orthogonal contrasts which involved comparisons for main effects, interaction effects, and comparisons involving the additional treatments in oil palm research studies.

Introducción

En investigación en palma de aceite suele ser de interés conocer simultáneamente el efecto de algunos factores de estudio sobre variables de respuesta, como, por ejemplo, el efecto que tiene la fertilización nitrogenada cuando se aplica a diferentes cultivares de palma de aceite en términos de producción: número de racimos, peso medio de racimo y potencial de aceite. En estos casos, se plantean experimentos con estructuras factoriales, en los cuales los tratamientos corresponden a las combinaciones de los distintos niveles de cada factor. Para un caso de fertilización nitrogenada, por ejemplo, si se consideran cuatro cultivares de palma de aceite (Brasil x Djongo Mix, Coari x La Mé, Taisha x Avros y Manaos x Compacta) y tres niveles de aplicación de fertilización (0,5, 1,0 y 1,5 kg/palma/año), entonces se tienen dos factores: cultivar de palma de aceite y fertilización con cuatro y tres niveles respectivamente, lo cual genera una estructura factorial 4x3 = 12 tratamientos. Una vez definida esta estructura, también conocida como estructura de tratamientos, se define el diseño experimental, el cual está asociado a la estructura de las unidades experimentales. Vale la pena aclarar que la estructura de tratamientos es diferente a la estructura de las unidades experimentales; la primera hace referencia a la forma como se conforman los tratamientos y la segunda, a la forma como se asignan los tratamientos a las unidades experimentales, es decir, el diseño experimental (Melo, López, & Melo, 2007).

Los experimentos factoriales permiten evaluar efectos principales de cada factor y el efecto de interacción entre estos, esto es, si hay un cambio significativo en la variable de respuesta generada por la aplicación de las combinaciones de los distintos niveles de cada factor (tratamientos). Por ende, existe un efecto de interacción entre dos factores si el efecto sobre la variable de respuesta del nivel de un factor se da en forma distinta para los niveles del otro factor con el que se combine (Kuehl, 2000).

Para el ejemplo de fertilización, el efecto principal cuantifica las diferencias en promedio del peso de racimo para las tres aplicaciones de nitrógeno (0,5, 1,0 y 1,5 kg/palma/año), cuando se promedia para los cuatro cultivares; el efecto de interacción se presenta cuando la respuesta del cultivar depende de la cantidad de fertilizante que se está aplicando; en otras palabras, si la respuesta de los cultivares no es la misma para los tres niveles de aplicación del nutrimento.

A manera de ejemplo, en la Figura 1 se muestra el comportamiento de dos factores: cultivar y fertilización, donde se observan posibles diferencias entre los cultivares para los distintos niveles de N, pero no un efecto de interacción entre los dos factores. Por su parte, la Figura 2 muestra el efecto de interacción entre los dos factores, señalando que el comportamiento del cultivar 3 a través de los tres niveles de fertilización nitrogenada es distinto a los otros cultivares, puesto que mientras la respuesta para el cultivar 3 muestra una tendencia creciente en todo el rango de niveles de N, los otros tres cultivares muestran un comportamiento creciente hasta el segundo nivel de N, a partir del cual el peso medio de racimo incrementa a una tasa mucho menor. La pregunta ahora es si dicha variación observada en el cultivar 3 genera o no un peso medio de racimo estadísticamente diferente al de otros cultivares.

Figura 1. No efecto de interacción entre los dos factores.

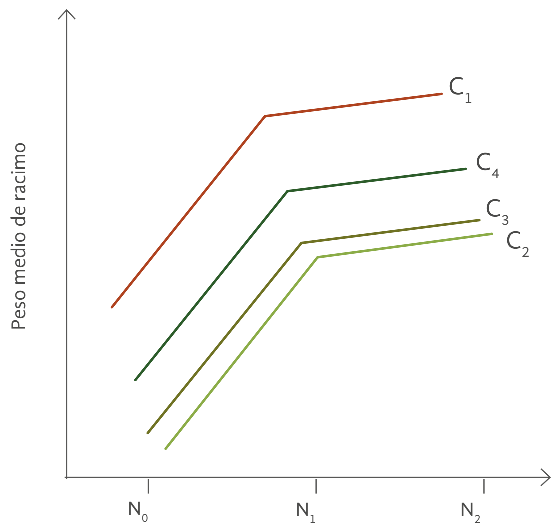
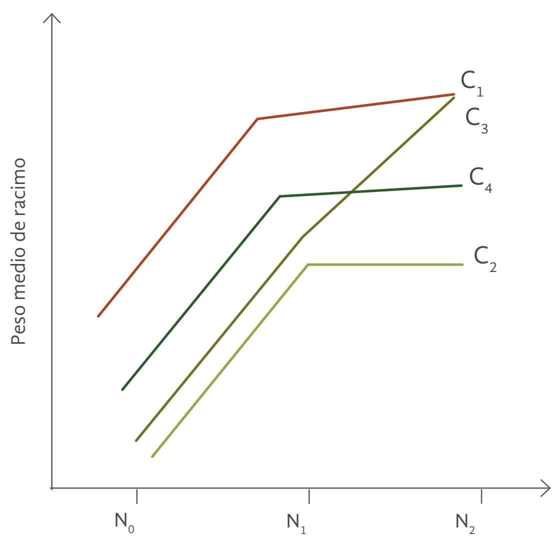


Figura 2. Efecto de interacción entre los factores.



De acuerdo con lo reportado por Melo *et al.* (2007), las principales ventajas de este tipo de experimentos factoriales son: *i)* la ampliación del rango de inferencia dada por los niveles de cada factor bajo estudio y *ii)* la posibilidad de realizar una estimación del efecto de interacción de los factores. En este tipo de experimentos factoriales y, en particular, en ensayos de fertilización, en varias ocasiones se considera uno o más tratamientos testigos, adicionales a la estructura factorial básica (Piepho, Williams, & Fleck, 2003; Schaarsschmidt & Vaas, 2009). En el análisis de los datos se considera el efecto de los tratamientos de forma general, y no se proporcionan soluciones para los efectos principales de los factores bajo estudio ni de sus posibles interacciones. El objetivo de este trabajo es mostrar cómo a través de contrastes ortogonales se puede llegar a estimar el efecto principal de cada factor, el efecto de interacción entre estos, las posibles comparaciones entre la estructura factorial básica y los tratamientos testigos adicionales.

Contrastes ortogonales

Un contraste es una combinación lineal de las medias de tratamientos de la forma presentada en la ecuación 1 (Kuehl, 2001):

$$C = \sum_{j=1}^{t} C_j \, \mu_j = C_1\mu_1 + \cdots + C_t\mu_t \quad \text{(Ec. 1)}$$

Donde t indica el número de tratamientos, las ponderaciones C_j se generan a partir de las comparaciones de interés y deben cumplir la restricción: $\sum_{j=1}^{t} C_j = 0$. La prueba estadística para cada contraste se hace a partir de la prueba t de Student así (Ec. 2):

$$t_c = \frac{\sum_{j=1}^{t} c_j \hat{\mu}_j}{\hat{\sigma}\sqrt{\sum_{j=1}^{t} \frac{c_j^2}{r}}} \quad \text{(Ec. 2)}$$

Donde $\hat{\mu}_j$ y $\hat{\sigma}^2$ corresponden a la media estimada del tratamiento j y al cuadrado medio del error experimental, simultáneamente. Los grados de libertad para la prueba corresponden a los grados asociados al cuadrado medio del error experimental. Para que dos contrastes C_1 y C_2 sean ortogonales entre sí, debe cumplirse la siguiente restricción (Ec. 3):

$$\sum_{j=1}^{t} C_{1j} \, C_{2j} = 0 \quad \text{(Ec. 3)}$$

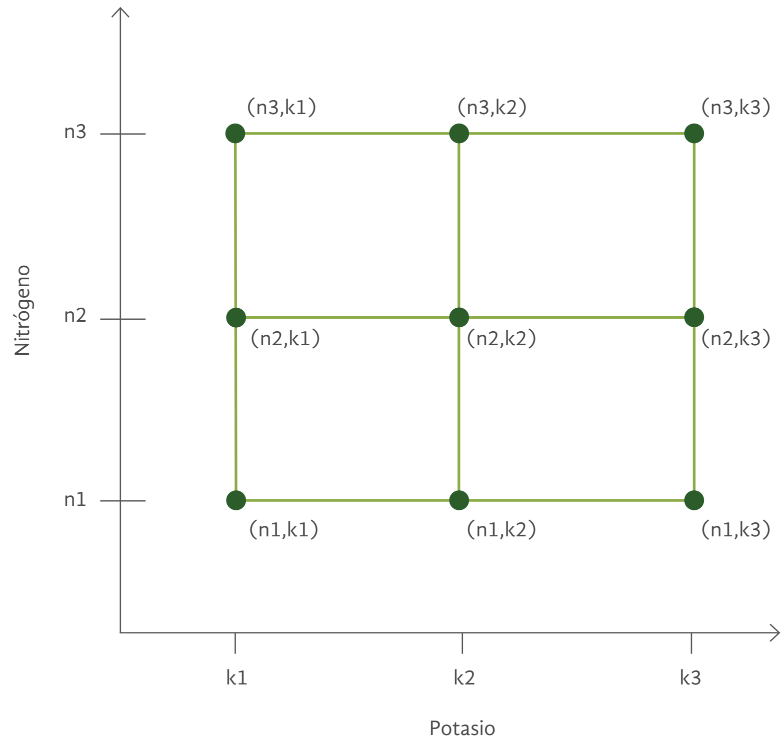
Análisis de varianza

Es una técnica estadística que permite descomponer la variación total presentada por la variable de respuesta (suma de cuadrados total) en la variación debida a las distintas componentes del modelo o fuentes de variación; para el caso de un diseño de bloques completamente aleatorizados, la variación ocasionada por los bloques, los tratamientos y el error experimental (Steel & Torrie, 1960). Esta variación cuantificada para cada fuente de variación debe totalizar la suma de cuadrados total; similarmente, se hace con los grados de libertad, cada fuente de variación tiene asociado un número de grados de libertad y la suma de estos para las tres fuentes de variación deben sumar los grados de libertad totales.

Materiales y métodos

Con el propósito de identificar las dosis de nitrógeno (N) y potasio (K) que se deben aplicar al híbrido Coari x La Mé para alcanzar mayores rendimientos, se llevó a cabo un ensayo en una plantación de la Zona Suroccidental palmera de Colombia, en la que se probaron tres niveles de aplicación de N (0,50, 0,75 y 1,0 kg/palma/año) y tres niveles de aplicación de P (0,96, 1,44 y 1,92 kg/palma/año). En el ensayo se consideraron, además de los nueve tratamientos generados por la estructura factorial básica, dos tratamientos adicionales: un tratamiento testigo, que corresponde a la no aplicación de los nutrimentos en prueba, y un testigo de referencia, que corresponde a la aplicación que realiza la plantación donde se realizó el ensayo. Así, se generó una estructura factorial aumentada con 11 tratamientos. Con el propósito de controlar la variabilidad dada por la condición del terreno, se utilizó un diseño de bloques completamente aleatorizados con cinco repeticiones; la unidad experimental correspondió a una parcela con 32 palmas de las cuales se evaluaron las 12 centrales. El diseño básico de tratamientos se muestra en la Figura 3.

Figura 3. Puntos del diseño básico de tratamientos.



El modelo estadístico para este diseño se escribe de la forma:

$$Y_{ij} = \mu + \beta_i + \alpha_j + \varepsilon_{ij}$$

Donde Y_{ij} corresponde al valor de la variable respuesta en el bloque i y el tratamiento j, μ es la media general de la variable respuesta en el ensayo, β_i es el efecto del bloque i, α_j es el efecto del j-ésimo tratamiento y ε_{ij} la variable aleatoria no observable con distribución $NI(0, \sigma^2)$.

En este ensayo, el efecto de los tratamientos contiene el efecto de los factores principales N y K, el efecto de interacción NxK y el efecto de los tratamientos adicionales. En la Tabla 1 se muestra la descripción de los once tratamientos, así como las cantidades aplicadas, expresadas en kilos/palma/año de N y K. Además, la Tabla 2 presenta el análisis de varianza para este experimento considerando sumas de cuadrados y grados de libertad para cada fuente de variabilidad como se hace tradicionalmente.

Una vez se realiza el análisis de los datos, se revisa la última columna Pr > F para la línea de tratamientos, con el propósito de verificar la hipótesis de interés en el estudio: $H_0: \mu_1 = \mu_2 = \cdots = \mu_{11}$, es decir, verificar si todas la medias de los tratamientos son estadísticamente iguales, o no, en términos de la variable de respuesta; si este valor es mayor de 0,05, no se rechaza la hipótesis H_0, lo que indica que en las condiciones bajo las cuales se realizó el experimento y con un margen de error del 5 % no se encuentran diferencias estadísticas entre los tratamientos; luego la aplicación de cualquiera de los tratamientos genera los mismos resultados en términos de la variable de interés. La recomendación de uno o más tratamientos en este caso depende de la facilidad de la consecución y aplicación de estos, del efecto ambiental que puedan generar y de su costo, entre otros aspectos. Ahora, si, por el contrario, el valor Pr > F en la línea de tratamientos es menor de 0,05, entonces se rechaza la hipótesis H_0, lo que indica que al menos uno de los tratamientos es distinto a los demás. Una alternativa de análisis frecuentemente utilizada es realizar una prueba de comparación múltiple, para indicar cuáles tratamientos son los que generan la diferencia y seleccionar aquellos que ofrecen la mejor respuesta.

¿Cómo proceder si no se tiene información relacionada con el efecto de cada uno de los factores o del efecto de interacción entre los mismos? Por ejemplo, el investigador puede estar interesado en responder preguntas como: ¿Cuál es el efecto de N y K sobre la variable de respuesta? ¿Es lineal? ¿Es cuadrático? ¿Existe efecto de interacción entre N y K? ¿Es mejor aplicar K o no hacerlo? Las comparaciones múltiples tradicionalmente utilizadas no permiten contestar puntualmente estas preguntas, pero es posible hacerlo a través de contrastes ortogonales, que es el objeto de este documento: hacer una descomposición de la suma de cuadrados de los tratamientos en sumas de cuadrados de contraste asociadas a preguntas de interés. Aquí surge un interrogante: ¿Cuántas preguntas se pueden formular? La respuesta es que se pueden formular tantas preguntas como grados de libertad se tengan para los tratamientos.

Para este ensayo se plantearon las siguientes hipótesis de trabajo: *i)* existe un efecto lineal o cuadrático para N; *ii)* existe un efecto lineal o cuadrático para K; *iii)* existe un efecto de interacción entre los dos nutrimentos; *iv)* algunos de los tratamientos propuestos son superiores a los testigos; y *v)* es necesario hacer aplicación de al menos uno de los dos nutrimentos. A manera de ejemplo, a continuación se muestra cómo resolver la hipótesis de efecto lineal de los dos nutrimentos y de su efecto de interacción a partir de contrastes ortogonales (Tabla 3).

Tabla 1. Descripción de los tratamientos considerados en el ensayo.

Tratamiento	N	K	Notación
1	0,50	0,96	$n_1 k_1$
2	0,50	1,44	$n_1 k_2$
3	0,50	1,92	$n_1 k_3$
4	0,75	0,96	$n_2 k_1$
5	0,75	1,44	$n_2 k_2$
6	0,75	1,92	$n_2 k_3$
7	1,00	0,96	$n_3 k_1$
8	1,00	1,44	$n_3 k_2$
9	1,00	1,92	$n_3 k_3$
10	0,00	0.00	Testigo absoluto
11	n_p	k_p	Testigo referente

Tabla 2. Estructura de análisis de varianza para el ensayo 3x3 más dos testigos de la forma tradicional.

Fuente de variación	Grados de libertad	Cuadrados medios	Pr > F
Bloques	B -1 = 4	-	-
Tratamientos	t – 1 = 10	$CM_{Tratamientos}$	Valor
Error experimental	(b-1) (t-1) = 40	CM_{Error}	-
Total	Bt -1 = 54	SC_{Total}	-

Tabla 3. Hipótesis de efecto lineal de N y K a partir de contrastes ortogonales.

Contraste	n_0k_0	n_0k_1	n_0k_2	n_1k_0	n_1k_1	n_1k_2	n_2k_0	n_2k_1	n_2k_2	TAbs	TRef	Suma
N_L	1	1	1	0	0	0	-1	-1	-1	0	0	0
K_L	1	0	-1	1	0	-1	1	0	-1	0	0	0
$N_L * K_L$	1	0	-1	0	0	0	-1	0	1	0	0	0

Los coeficientes para estimar el efecto lineal de cada nutrimiento están en las primeras dos líneas (Steel & Torrie, 1960). Para verificar que los dos contrastes propuestos son ortogonales, se multiplica uno a uno los coeficientes de cada columna y luego se suman. En la última columna de la tabla se verifica que la suma de cada línea es cero, para las dos primeras líneas se indica que se tienen dos contrastes y en la tercera línea se verifica la ortogonalidad de estos. Para ilustrar la metodología, se utilizarán los datos generados para la variable potencial de aceite. El análisis de varianza con la descomposición de sumas de cuadrados y grados de libertad se muestra en la Tabla 4.

En la Tabla 4 se observa que los 10 grados de libertad para tratamientos se descomponen inicialmente en cuatro líneas (color azul), en las cuales se prueba el efecto de los dos nutrimentos (N y K), se considera el efecto de la interacción entre estos, se hace una comparación entre los tratamientos de la estructura factorial completa y los dos testigos, para finalmente comparar los testigos. Los dos grados de libertad asociados al efecto del N y del K, a su vez se pueden descomponer así: un grado de libertad para verificar si el efecto de cada nutrimento es lineal, y otro para verificar si el efecto es cuadrático; de igual manera, cada línea se puede descomponer más dependiendo de cuantas preguntas o hipótesis se tengan por parte del investigador, recordando que en total se pueden hacer tantas preguntas como grados de libertad se tengan para los tratamientos.

Resultados

Una vez calculadas las sumas de cuadrados y grados de libertad para construir la tabla de análisis de varianza, se estiman los residuales y se hace la verificación de los supuestos. Por último, se procede a la verificación de las hipótesis planteadas. Los resultados del análisis de varianza, la correspondiente descomposición de las sumas de cuadrados y grados de libertad, así como la verificación de las hipótesis planteadas, se muestran en la Tabla 5.

El análisis muestra que hay un efecto de los tratamientos ($Pr < 0{,}0462$), razón para hacer la descomposición de las sumas de cuadrados y grados de libertad de los tratamientos. De acuerdo con la descomposición presentada en la Tabla 5, no se observa efecto de interacción entre los dos nutrimentos. Seguidamente se hace el análisis para cada factor independientemente, y para los dos nutrimentos se observa un efecto cuadrático ($Pr < 0{,}0072$, para N, y $Pr < 0{,}0296$, para K). No se observan diferencias entre los tratamientos de la estructura básica y los tratamientos adicionales ($Pr < 0{,}1192$); igualmente, se observa este comportamiento al comparar los dos tratamientos adicionales entre sí. En la Tabla 6 se presentan las medias de potencial de aceite para los tres niveles de cada nutrimento y en la Figura 4, el comportamiento de cada nutrimiento en sus tres niveles donde se muestra el efecto cuadrático de los dos nutrimentos.

Tabla 4. Descomposición de sumas de cuadrados y grados de libertad del análisis de varianza para el ensayo 3x3 más dos testigos a partir de contrastes ortogonales.

Fuente de variación	GL	Sumas de cuadrados	Cuadrados medios	F_c
Bloques	b-1 = 4	$SC_{Bloques}$	-	-
Tratamientos	t – 1 = 10	$SC_{Tratamientos}$	$CM_{Tratamientos}$	Fc_{Trat}
Efecto del N	n-1 = 2	SC_N	CM_N	Fc_N
Efecto del K	k-1 = 2	SC_K	CM_K	Fc_K
Efecto de la interacción N*K	(n-1)(k-1) = 4	SC_{NK}	CM_{NK}	Fc_{NK}
Factorial vs testigos	Diferencia= 2	$SC_{Fact\ vs\ Test}$	$CM_{Fact\ vs\ Test}$	$Fc_{Fact\ vs\ Test}$
Error experimental	(b-1) (t-1) = 40	SC_{Error}	CM_{Error}	-
Total	bt-1 = 54	SC_{Total}	-	-

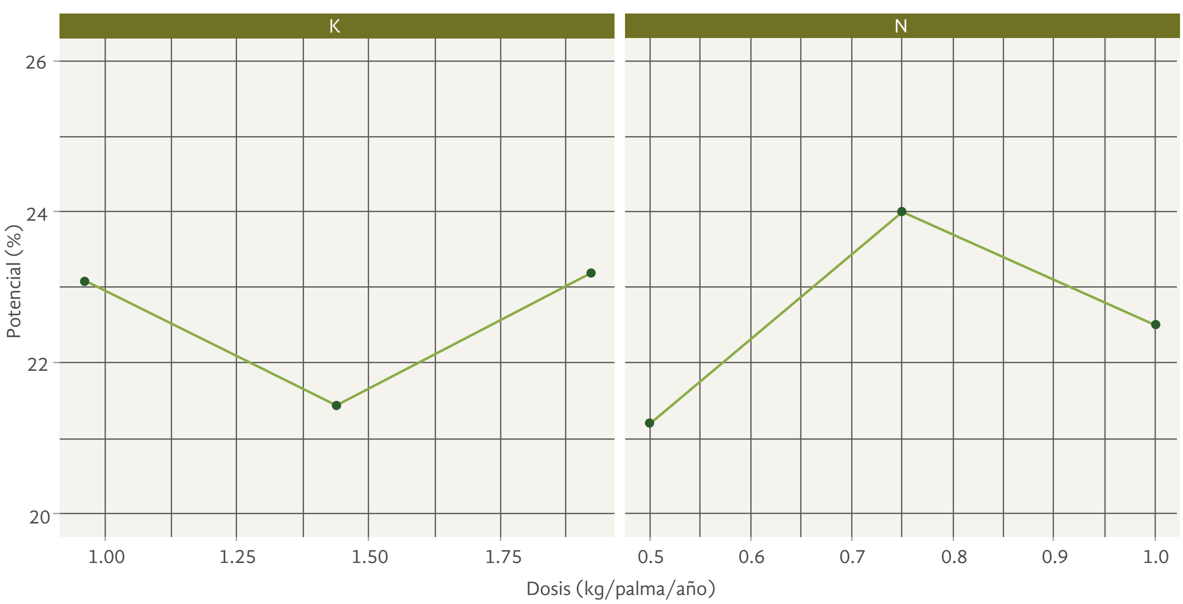
Tabla 5. Resultados del análisis de varianza para la variable potencial de aceite.

Fuente	DF	Tipo III SS	Cuadrados medios	F Valor	Pr > F
Rep	4	61,4570155	15,3642539	2,67	0,0461
Trat	10	121,7253727	12,1725373	2,11	0,0462
Efecto lineal de N	1	12,19856333	12,19856333	2,12	0,1535
Efecto cuadrático de N	1	46,31104000	46,31104000	8,04	0,0072
Efecto lineal de K	1	0,10920333	0,10920333	0,02	0,8912
Efecto cuadrático de K	1	29,34369000	29,34369000	5,09	0,0296
N lineal * K lineal	1	1,88498000	1,88498000	0,33	0,5706
N cuad * K cuad	1	14,78080667	14,78080667	2,57	0,1171
N cuad * K lineal	1	0,00032667	0,00032667	0,00	0,9940
N cuad * K cuad	1	0,33282000	0,33282000	0,06	0,8113
Factorial frente a testigos	1	14,61098273	14,61098273	2,54	0,1192
Entre testigos	1	2,15296000	2,15296000	0,37	0,5445
Error	40	230,4883545	5,7622089		
Total corregido	54	413,6707427			

Tabla 6. Promedios de potencial de aceite para cada nivel de nutrimento.

N	Obs	Potencial	K	Obs	Potencial
0,50	30	21,2133	0,96	30	23,0790
0,75	30	24,0030	1,44	30	21,4263
1,00	30	22,4887	1,92	30	23,1997

Figura 4. Comportamiento del potencial de aceite para cada nutrimento.



Conclusiones y reflexiones

Los contrastes ortogonales son una herramienta estadística importante para descomponer las sumas de cuadrados de tratamientos de acuerdo con las hipótesis de trabajo de interés en una investigación.

Cuando se tienen factores de estudio cuantitativos, la suma de cuadrados para cada factor puede descomponerse en efectos lineales, cuadráticos y más, dependiendo del número de niveles de cada factor. Cuando los factores de estudio son cualitativos, también es posible hacer la descomposición a través de contrastes ortogonales, según las hipótesis de interés.

De los resultados se concluye que, aunque se presentaron efectos cuadráticos del N y el K sobre el contenido de aceite de las palmas analizadas, solo el N está influenciando positivamente la variable.

Agradecimientos

Los autores agradecen a Colciencias por los recursos aportados para financiar la investigación realizada en el marco del proyecto Referenciación Competitiva a las Empresas Productoras de Palma de Aceite Híbrido OxG Alto Oleico y al Fondo de Fomento Palmero, administrado por Fedepalma, por cofinanciar la investigación. Igualmente, expresan un agradecimiento especial a la plantación Oleaginosas Salamanca S. A., en la cual se llevó a cabo el ensayo planteado en este documento.

Referencias

Kuehl, R. O. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis*. 2nd ed. Brooks/Cole.

Little, R., Freund, R., & Spector P. (1993). *SAS System for linear models*. Third edition. Cary, NC, USA: SAS Institute Inc.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. Boca Raton, FL, USA: Chapman & Hall/CRC.

Manly, F. J. (2009). *Statistics for Environmental Science and Management*. 2nd ed. Cheyenne, WY, USA: Chapman & Hall.

Marini, R. P. (2003). Approaches to Analyzing Experiments with Factorial Arrangements of Treatments Plus Other Treatments. *HortScience*, *38*(1), 117-120.

Melo, O., López, L. A., & Melo, S. E. (2007). *Diseño de Experimentos: métodos y aplicaciones*. Bogotá: Pro-Offset.

Milliken, G. A., & Johnson, D. E. (1992). *Analysis of Messy Data*. *Vol. I: Designed Experiments*. Boca Raton, FL, USA: Chapman & Hall.

Montgomery, D. C. (2005). *Design and Analysis of Experiments*. 6th Ed. John Wiley & Sons.

Piepho, H. P., Williams, E. R., & Fleck, M. (2006). A note on the analysis of designed experiments with complex treatment structure. *HortScience*, *41*(2), 446-452.

SAS Institute (2017). *SAS/STAT. User guide, version 9.4.* Cary, N.C.: SAS Institute.

Schaarschmidt, F., & Vaas, L. (2009). Analysis of trials with complex treatment structure using multiple contrast tests. *HortScience*, *44*(1), 188-195.

Steel, R., & Torrie, J. (1960). *Principles and Procedures of Statistics*, 1st ed. New York: McGraw-Hill.