

<http://revistas.ucr.ac.cr/index.php/ingenieria>  
[www.ucr.ac.cr](http://www.ucr.ac.cr) / ISSN 2215-2652

# Ingeniería

Revista de la Universidad de Costa Rica  
ENERO / JUNIO 2021 - VOLUMEN 31 (1)



# Uso de modelos mixtos en el análisis de estudios de homogeneidad para ensayos de aptitud: un ejemplo de aplicación para la medición de elementos en agua potable

## Use of mixed models in the analysis of homogeneity studies for proficiency testing: an application example for the measurement of elements in drinking water

*Gabriel Molina-Castro,*

*Laboratorio Costarricense de Metrología, Costa Rica.*

*Email: gmolina@lcm.go.cr*

ORCID: 0000-0002-4051-7229

*Jimmy Venegas-Padilla,*

*Laboratorio Costarricense de Metrología, Costa Rica.*

*Email: jvenegas@lcm.go.cr*

ORCID: 0000-0001-8662-4348

*Bryan Calderón-Jiménez,*

*Laboratorio Costarricense de Metrología, Costa Rica.*

*Email: bcalderon@lcm.go.cr*

ORCID: 0000-0003-3117-7709

**Recibido:** 30 de octubre 2020

**Aceptado:** 18 de diciembre 2020

### Resumen

Los ensayos de aptitud han demostrado ser herramientas de calidad sumamente poderosas para evaluar la calidad de las mediciones de diversos laboratorios. El presente trabajo evalúa las ventajas del uso de modelos mixtos lineales frente a modelos tradicionales de efectos aleatorios en estudios de homogeneidad aplicados en ensayos de aptitud, y describe su aplicación con un ejemplo en la medición de elementos de agua potable, correspondiente al ensayo de aptitud DMQ-001-2018 del Laboratorio Costarricense de Metrología (LCM). Ambos modelos fueron ajustados y evaluados para los datos de medición de calcio (Ca) y magnesio (Mg) utilizando el software estadístico R. Se evidenció la presencia de una tendencia por medición para Mg ( $p = 0.0005$ ), pero no para Ca ( $p = 0.4265$ ). Un análisis de los componentes de incertidumbre por falta de homogeneidad en las unidades ( $u_{hom}$ ) y por repetibilidad del método ( $u_r$ ) demostró la similitud entre los modelos cuando no existe una tendencia aparente por medición (modelo mixto lineal:  $u_{hom} = 0.45$  mg/L y  $u_r = 0.46$  mg/L, modelo aleatorio:  $u_{hom} = 0.47$  mg/L y  $u_r = 0.45$  mg/L). Sin embargo, cuando existen tendencias aparentes por medición (modelo mixto lineal:  $u_{hom} = 0.00$  mg/L y  $u_r = 0.08$  mg/L, modelo aleatorio:  $u_{hom} =$  no estimable y  $u_r = 0.23$  mg/L) se observó diferencias significativas. Para este último caso, se evidenció grandes ventajas para el modelo mixto lineal en comparación con el modelo tradicional, resaltando la posibilidad de



eliminar indefiniciones en la estimación de  $u_{hom}$  y la mitigación de posibles sobreestimaciones de  $u_r$ . Finalmente, el presente estudio brinda un código programado en R para procesar datos de estudios de homogeneidad basado en el ajuste de un modelo mixto lineal.

*Palabras clave:*

Agua potable, ensayo de aptitud, estudio de homogeneidad, incertidumbre, metrología, modelo mixto.

**Abstract**

Proficiency testing has proven to be an extremely powerful tool to evaluate the quality of measurements from various laboratories. This paper evaluates the advantages of using linear mixed models over traditional random-effects models in the analysis of homogeneity studies in proficiency testing and describes their application with a study case of elements measurement in drinking water (proficiency testing DMQ-001-2018 by LCM). Both models were adjusted and evaluated for calcium (Ca) and magnesium (Mg) measurement data using R software. A trend by measurement was evidenced for Mg ( $p = 0.0005$ ) but not for Ca ( $p = 0.4265$ ). An analysis of uncertainty components due to lack of homogeneity between units ( $u_{hom}$ ) and method repeatability ( $u_r$ ) showed the similarity in the components obtained for the models in the case without trends (linear mixed model:  $u_{hom} = 0.45$  mg/L and  $u_r = 0.46$  mg/L, random-effects model:  $u_{hom} = 0.47$  mg/L and  $u_r = 0.45$  mg/L). However, significant differences were observed in the case with trends (linear mixed model:  $u_{hom} = 0.00$  mg/L and  $u_r = 0.08$  mg/L, random-effects model:  $u_{hom} = \text{undefined}$  and  $u_r = 0.23$  mg/L). For the latter, great advantages were evidenced for the mixed linear model compared to the traditional model, highlighting the possibility of eliminating mathematical undefinities in the estimation of  $u_{hom}$  and the mitigation of possible overestimations of  $u_r$ . Finally, a R-code is provided to process data from homogeneity studies based on the fit of a linear mixed model.

*Keywords:*

Drinking water, homogeneity study, linear mixed model, metrology, proficiency testing, uncertainty.

## 1. INTRODUCCIÓN

Los ensayos de aptitud corresponden a herramientas de evaluación de la conformidad no solo altamente poderosas, si no también ampliamente utilizadas para múltiples propósitos. Dentro de sus aplicaciones más comunes en ensayos químicos se encuentran la evaluación del desempeño de diversos tipos de laboratorios, la identificación de interferentes o efecto matriz en ensayos, la identificación de procedimientos de ensayo inadecuados, la comparación y evaluación de distintos métodos de ensayo, la validación de métodos y de las incertidumbres declaradas en sus alcances, entre muchas otras [1]. Específicamente, en la medición de elementos en agua potable, este tipo de ejercicios toma una gran importancia al asegurar y evaluar las mediciones realizadas por diversos organismos y laboratorios interesados en determinar la potabilidad del agua. Así, los ensayos de aptitud tienen un impacto crucial para asegurar la salud y bienestar de la población en general.

Debido a su gran importancia, los ensayos de aptitud típicamente incluyen, entre muchas otras etapas, estudios de homogeneidad para caracterizar las unidades del material a utilizar [1]. Como su nombre lo indica, los estudios buscan evaluar si las diferencias entre las unidades del ítem analizado (falta de homogeneidad) pueden llegar a afectar el propósito de los ensayos de aptitud; es decir, evaluar objetivamente el desempeño de los participantes [2].

Diversos enfoques han sido planteados y utilizados para analizar los datos provenientes de un estudio de homogeneidad aplicado en ensayos de aptitud; destacan los propuestos por la norma internacional ISO 13528 [3] y el Protocolo Armonizado desarrollado por la Unión Internacional de Química Pura y Aplicada (IUPAC, por sus siglas en inglés) [4]. De forma homologada, el análisis de datos en ambos documentos se aborda mediante el ajuste de un modelo de efectos aleatorios, con un descarte previo de la presencia de tendencias en los datos, y evaluado posteriormente mediante un análisis de varianza (ANOVA, por sus siglas en inglés) de una vía o factor, en este caso, la unidad o ítem de ensayo. Algunos valores particulares obtenidos, que representan la variación cuantificada por las diferencias promedio entre las unidades, son extraídos para estimar una componente de incertidumbre por falta de homogeneidad, cuyo concepto es acorde con la Guía para la Expresión de la Incertidumbre de Medida (GUM, por sus siglas en inglés) [5]. Este componente finalmente es comparado contra criterios de aceptación para determinar si dichas diferencias pueden impactar significativamente el objetivo del ensayo de aptitud. Sin embargo, no es claro cómo proceder, cuáles acciones tomar o cómo interpretar los resultados obtenidos en los estudios de homogeneidad ante la presencia de tendencias en los datos. Van der Veen et al. [6] advierten de la imposibilidad de interpretar correctamente los resultados de un ANOVA en presencia de tendencias.

Distintas aplicaciones se han propuesto para abordar esta problemática, incluyendo la corrección de la tendencia (pendiente e intercepto) en los datos antes del análisis tradicional y el uso de modelos mixtos lineales. El segundo representa un enfoque más práctico y directo, dado que el primero requiere un ajuste y substracción inicial del modelo lineal a los datos, y una reducción posterior en los grados de libertad del ANOVA [7]. Este último aspecto puede presentar el inconveniente práctico de tener que aplicarse manualmente si el software utilizado no ajusta los grados de libertad

automáticamente, y el inconveniente teórico de reducir la fiabilidad en el ajuste del ANOVA. Según Hox [8], los modelos mixtos se caracterizan por la estructura jerárquica de sus datos, donde piezas de información de niveles inferiores (individuos o réplicas) se encuentran anidadas en niveles superiores (conglomerados o unidades replicadas). A diferencia de los modelos clásicos de regresión, estos permiten la estimación simultánea de variables en distintos niveles jerárquicos, por lo que pueden ser utilizados en experimentos de medidas repetidas [9, 10, 11]. Algunas de sus aplicaciones en el campo de la metrología incluyen los trabajos desarrollados por Ciarlini et al. [12], Pavese *et al.* [13] y Cui et al [14], entre otros. Recientemente, la nueva versión de la ISO Guía 35 [7] incluye un breve ejemplo de un estudio de homogeneidad de un material de referencia, en donde se aborda la presencia de tendencias en el marco del análisis de datos requerido para un proceso de certificación de este tipo de materiales. Pese a lo anterior, actualmente poco o nada se ha estudiado sobre la presencia de tendencias en los estudios de homogeneidad desarrollados para ensayos de aptitud haciendo imperativo su estudio en este tipo de aplicación metrológica.

Debido a lo anterior, el presente artículo planteó como objetivos evaluar las ventajas del uso de modelos mixtos lineales frente a los modelos de efectos aleatorios utilizados tradicionalmente en el análisis de datos de un estudio de homogeneidad en ensayos de aptitud; así como describir la metodología de uso del primer tipo de modelos en un ejemplo de aplicación real correspondiente a un ensayo de aptitud en la medición de elementos de agua potable. Como se mostrará más adelante, estos modelos cuentan con el potencial de cambiar las metodologías normalizadas que han sido implementadas de manera generalizada en el análisis de datos en ensayos de aptitud.

## 2. METODOLOGÍA

### 2.1 Metodología experimental

El presente análisis proviene de los estudios experimentales de homogeneidad realizados para el ensayo de aptitud DMQ-001-2018 y su correspondiente estudio exploratorio DMQ-001-2018-EE01, desarrollado por el Departamento de Metrología Química del Laboratorio Costarricense de Metrología (LCM) durante el 2018, y que incluyó la medición de los elementos calcio (Ca) y magnesio (Mg) en agua potable [15].

Para la evaluación experimental, 7 unidades de ítems de ensayo (almacenados en botellas de polipropileno de 125 ml) fueron seleccionadas mediante muestreo aleatorio simple de un lote de 25 unidades. Cada una de las unidades muestreadas fue medida por duplicado, en una sola corrida experimental, siguiendo un orden sistemático predefinido para evitar la presencia de correlaciones fuertes con la secuencia de preparación o el agrupamiento de las réplicas de la misma unidad [3, 7]. Las mediciones de Ca y Mg se realizaron por medio de la técnica de Espectrometría de Absorción Atómica en Llama (FAAS, por sus siglas en inglés). El instrumento utilizado fue un espectrofotómetro modelo PinnAAcle 900T (PerkinElmer, USA), acoplado a un automuestreador modelo S10 (PerkinElmer, USA). Por su parte, las disoluciones calibradoras de Ca y Mg fueron preparadas a partir de los materiales de referencia certificados SRM 3109a y SRM 3131a, ambos

producidos por el Instituto Nacional de Estándares y Tecnología de Estados Unidos (NIST, por sus siglas en inglés) y trazables al Sistema Internacional de Unidades (SI). La exactitud, el efecto memoria y la presencia de deriva o tendencias en las mediciones fueron evaluadas utilizando el material de referencia certificado SLRS-6 para ambos elementos (producido por el Consejo Nacional de Investigación de Canadá NRC, por sus siglas en inglés) como control de calidad externo. En la Fig. 1 se presenta un diagrama de la evaluación experimental utilizada para la obtención de los datos.

La exactitud, el efecto memoria y la presencia de deriva o tendencias en las mediciones fueron evaluadas utilizando el material de referencia certificado SLRS-6 para ambos elementos (producido por el Consejo Nacional de Investigación de Canadá NRC, por sus siglas en inglés) como control de calidad externo. En la Fig. 1 se presenta un diagrama de la evaluación experimental utilizada para la obtención de los datos.

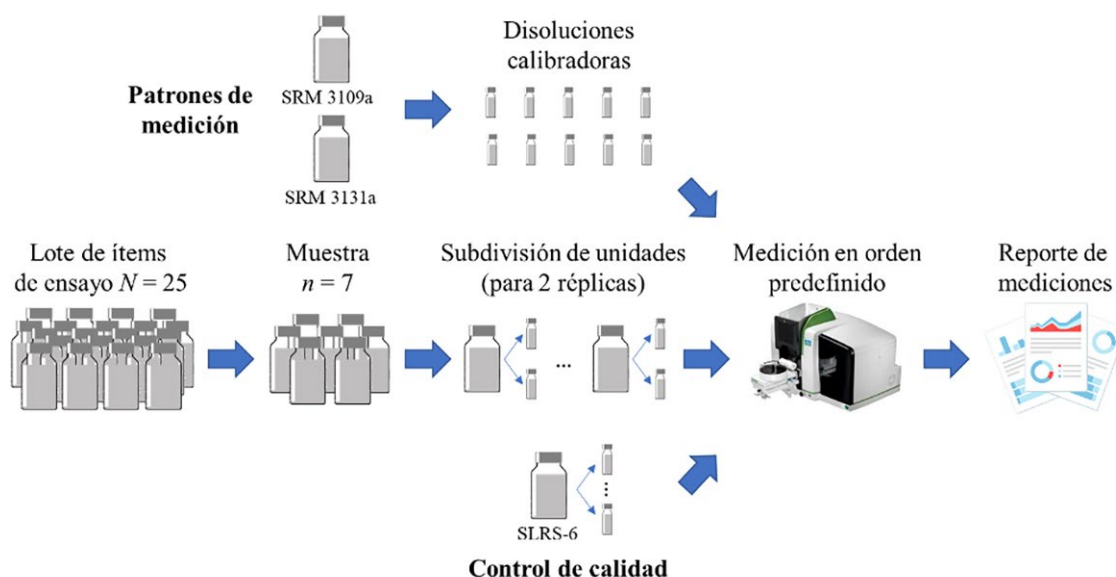


Figura 1. Esquema de evaluación experimental utilizada para la obtención de los datos (imagen confeccionada por los autores a partir de [7] y [28])

## 2.2 Base de datos y modelos estadísticos

### 2.2.1 Base de datos

De los resultados de los estudios de homogeneidad, se generó una base de datos con las siguientes variables:

- *elem*: variable categórica que únicamente identifica el elemento evaluado (Ca o Mg);
- *posic*: variable independiente discreta, en escala de 1 a 14, correspondiente a la posición de cada una de las réplicas, de cada una de las unidades ensayadas, en la secuencia de medición seguida;

- *unidad*: variable independiente categórica que contiene el número de identificación de cada una de las unidades ensayadas;
- *conc*: variable dependiente continua que corresponde a la concentración (en mg/L) de cada elemento analizado, obtenida como resultado de la medición realizada a cada réplica de cada unidad.

La base de datos utilizada se incluye dentro del material suplementario de la presente investigación.

### 2.2.2 Análisis preliminar

Inicialmente, se estimó estadísticos descriptivos de las variables. Luego, de forma exploratoria, se graficó los valores medidos de concentración, para cada uno de los elementos, en función de la posición de las réplicas de las unidades (secuencia de medición). Por último, se ajustó una regresión lineal por ajuste de mínimos cuadrados ordinarios (OLS, por sus siglas en inglés) para valorar visualmente la presencia de posibles tendencias por el método de medición, la cual fue corroborada con un ajuste similar aplicado a los resultados de los controles de calidad externos.

### 2.2.3 Modelo mixto lineal (modelo 1)

Dado que los casos de estudio corresponden a experimentos de medidas repetidas (sobre unidades), la información disponible se estructura en dos niveles jerárquicos de análisis: un primer nivel inferior correspondiente a las réplicas ( $i$ ) y un segundo nivel superior correspondiente a las unidades ( $j$ ). Se ajustó así el modelo mixto lineal mostrado en la ecuación (1).

$$conc_{ij} = \beta_0 + \beta_1 posic_{ij} + v_{0j} + \varepsilon_{ij} \quad (1)$$

El término de perturbación  $v_{0j}$  modela la variabilidad existente entre los promedios de las unidades (efecto aleatorio), mientras que los términos  $\beta_0$  y  $\beta_1$  corresponden a los coeficientes de regresión del modelo (componentes fijos) y  $\varepsilon_{ij}$  corresponde al error residual no explicado por el modelo ajustado (componente aleatorio). El ajuste del modelo se realizó considerando un criterio de máxima verosimilitud restringida (REML, por sus siglas en inglés) dado que este criterio logra minimizar el sesgo de estimación para los efectos aleatorios [8, 10, 11].

Respecto a las componentes de incertidumbre de medida atribuibles a las diferencias entre las unidades (incertidumbre por falta de homogeneidad entre unidades,  $u_{hom}$ ) y a las diferencias entre las réplicas de una misma unidad (incertidumbre por repetibilidad del método,  $u_r$ ), estas se estiman a través de los términos de variabilidad asociados con  $v_{0j}$  y  $\varepsilon_{ij}$  respectivamente [7], siguiendo las ecuaciones (2) y (3). Cabe señalar que los términos  $\sigma_{v_0}$  (desviación asociada a  $v_{0j}$ ) y  $\sigma_\varepsilon$  (desviación asociada a  $\varepsilon_{ij}$ ) surgen de un proceso de iteración matemática de minimización de la función objetivo de máxima verosimilitud restringida, cuyo detalle puede ser consultado en Galecki et al. [9] y West et al. [10].

$$u_{hom} = \sigma_{v_0} \quad (2)$$

$$u_r = \sigma_\varepsilon \quad (3)$$

### 2.2.4 Modelo de efectos aleatorios (modelo 2)

El modelo de efectos aleatorios, tradicionalmente utilizado para analizar los resultados de los estudios de homogeneidad en ensayos de aptitud (denominado también modelo ANOVA de efectos aleatorios), se encuentra descrito en la norma ISO 13528 [3] y se muestra en la ecuación (4). El mismo fue ajustado para el análisis de datos realizado en el ensayo de aptitud DMQ-001-2018 y el estudio exploratorio DMQ-001-2018-EE01 [15].

$$conc_{ij} = \mu + v_j + \varepsilon_{ij} \quad (4)$$

Dicho modelo sustituye los términos asociados con los efectos fijos por una media global común ( $\mu$ ). Su ajuste se realizó tras considerar un criterio de minimización del error cuadrático medio (ajuste por OLS), correspondiente a la metodología más común para este tipo de modelos [16].

Al modelo mostrado en la ecuación (4) se le aplica posteriormente un ANOVA de una vía, permitiendo la estimación de los correspondientes promedios cuadrados “entre muestras” ( $MS_{between}$ ) y “dentro de muestras” ( $MS_{within}$ ) [3, 6]. Estos términos se utilizan para estimar las componentes de incertidumbre  $u_{hom}$  y  $u_r$ , siguiendo las ecuaciones (5) y (6) respectivamente (el término  $n_0$  corresponde a una ponderación de la cantidad de réplicas realizadas a cada unidad).

$$u_{hom} = \sqrt{\frac{MS_{between} - MS_{within}}{n_0}} \quad (5)$$

$$u_r = \sqrt{MS_{within}} \quad (6)$$

## 2.3 Software utilizado

Para el procesamiento de los datos y el ajuste de los modelos se utilizó el ambiente de programación libre R, versión 3.6.1 [17]. Las librerías utilizadas se describen en la Tabla I. Cabe señalar que las funciones del paquete *nlme* [22] permiten la estimación directa de las variables  $\sigma_{v_0}$  y  $\sigma_\varepsilon$  como las desviaciones estándar (*StdDev*) asociadas a los efectos aleatorios (*Random effects*) resultantes de la aplicación de la función de regresión *lme*. Esto facilita la implementación práctica del algoritmo iterativo mencionado anteriormente.



Tabla 1. Librerías de r utilizadas en el presente estudio

Librerías	Uso	Referencias
<i>readr</i> , <i>dplyr</i> , <i>tidyr</i>	Lectura y manipulación de la base de datos	[18], [19] y [20]
<i>skimr</i>	Estadísticos descriptivos de las variables	[21]
<i>nlme</i>	Ajuste de los modelos lineales mixtos	[22]
<i>ggplot2</i> , <i>ggpubr</i> , <i>ggpmisc</i> , <i>ggthemes</i>	Generación de gráficos	[23], [24], [25] y [26]

El código de programación utilizado en la presente investigación puede ser consultado en el material suplementario.

### 3. RESULTADOS Y DISCUSIÓN

La Tabla II muestra un resumen estadístico para cada una de las variables utilizadas, donde se evidencia que la base de datos no posee valores faltantes. Además, corrobora que los valores de concentración en ambos elementos poseen distribuciones simétricas, aunque la dispersión de los valores de Ca es mayor.

Tabla 2. Resumen y descriptores estadísticos para las variables utilizadas

Elemento	Ca				Mg			
	Casos únicos		Total de casos		Casos únicos		Total de casos	
V. categórica	<i>unidad</i>		7		14			
	7		14		7		14	
V. continua	<i>prom</i>	<i>sd</i>	<i>med</i>	<i>m</i>	<i>prom</i>	<i>sd</i>	<i>med</i>	<i>m</i>
posic (1)	7.5	4.2	7.5	14	7.5	4.2	7.5	14
conc (mg/L)	66.4	0.6	66.4	14	30.1	0.2	30.1	14

V.: variable, *prom*: promedio, *sd*: desviación estándar, *med*: mediana, *m*: cantidad de datos.

Relacionado con el análisis preliminar, se evaluó visualmente la presencia de tendencias en el proceso de medición de las muestras (Fig. 2). Estas fueron verificadas al realizar el mismo proceso para los valores de los controles de calidad externos (Fig. A1).

Para ambos elementos, es claro que las concentraciones tienden a aumentar con el avance del proceso de medición, siendo esta tendencia más apreciable para el caso de Mg. Este comportamiento es acorde con lo esperado, ya que la técnica analítica de FAAS puede presentar distorsiones en las señales de respuesta debido a la deriva instrumental. Cabe señalar que, aunque ambas pendientes presentan valores numéricos similares, el nivel de concentración de Mg es, aproximadamente, la mitad de la concentración de Ca, por lo que los cambios por deriva instrumental afectan en mayor

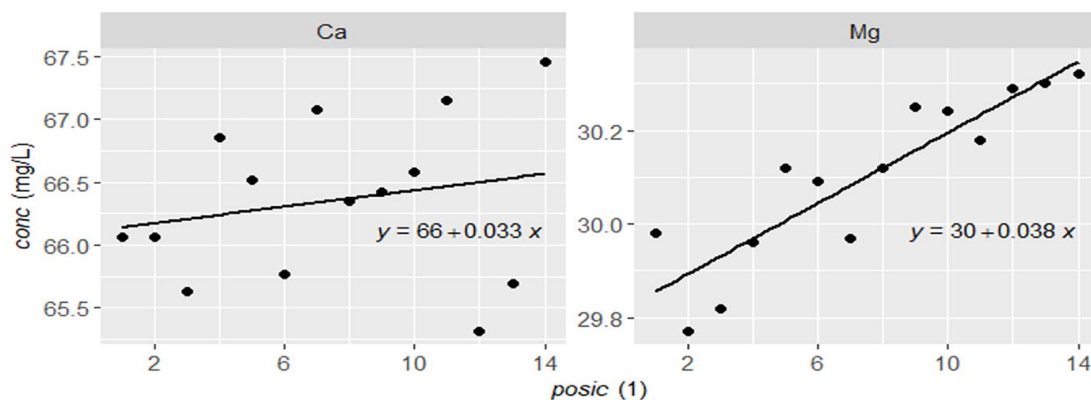


Figura 2. Ajuste exploratorio de modelos lineales ordinarios para evaluar la posible presencia de tendencias por medición en los datos provenientes de los estudios de homogeneidad para Ca y Mg del ensayo de aptitud DMQ-001-2018

proporción las mediciones de Mg. Esta diferencia en la relevancia de las tendencias puede definir el análisis de datos posterior, ya que para el caso de Ca se podría aplicar la metodología tradicional sin acciones adicionales [3], mientras que, para abordar esta tendencia en el caso de Mg, se deberían tomar acciones adicionales. Por ejemplo, eliminar la tendencia mediante una corrección numérica en los datos, descartar datos, repetir el experimento o utilizar técnicas alternativas para el procesamiento de datos. De esta manera, la opción propuesta de análisis corresponde al uso del modelo 1 el cual considera la tendencia observada como un efecto fijo presente en los datos y deja el componente de variabilidad por la falta de homogeneidad como un efecto aleatorio. Dicho modelo fue ajustado para ambos elementos, de forma que se pueda valorar su uso de manera generalizada y no solo en un caso específico. Los resultados respectivos se muestran en la Tabla 3. La comprobación de sus respectivos supuestos se muestran en la Fig. A2, todos con resultados positivos en cuanto a su cumplimiento.

Tabla 3. Resultados obtenidos con el ajuste del modelo 1 para ambos elementos

Elemento	Ca			Mg		
Coeficiente	<i>prom</i>	<i>sd</i>	<i>Prob. p</i>	<i>prom</i>	<i>sd</i>	<i>Prob. p</i>
$\beta_0$	66.1512	0.3168	0.0000	29.8164	0.0475	0.0000
$\beta_1$	0.0268	0.0314	0.4265	0.0379	0.0056	0.0005
$\sigma_{v0}$	0.4536			$3.047 \cdot 10^{-6}$		
$\sigma_\varepsilon$	0.4643			0.0842		

*prom*: promedio (mg/L), *sd*: desviación estándar (mg/L), *Prob. p*: probabilidad o valor-p (1).

En la Tabla 3 puede observarse que los coeficientes  $\beta_1$  para ambos elementos mantienen el comportamiento observado en el análisis preliminar al ajustar el modelo mixto lineal. Considerando un porcentaje de significancia del 5 %, la tendencia para el caso del Ca no presenta significancia estadística ( $\beta_1 = 0.0268$  mg/L,  $p = 0.4265$ ), mientras que se observa el caso contrario para el Mg ( $\beta_1 = 0.0379$  mg/L,  $p = 0.0005$ ). Estos resultados son de gran importancia, pues el análisis preliminar ajusta el modelo considerando únicamente el efecto fijo de la tendencia por medición. Por tanto, la inclusión del efecto aleatorio no altera de manera importante las conclusiones resultantes del análisis exploratorio.

En relación con los efectos aleatorios, se destaca en primera instancia los valores de  $\sigma_{v0}$ , los cuales corresponden a la variabilidad observada en el intercepto  $\beta_0$  del modelo (expresada como una desviación estándar) debido a la diferencia en las unidades. Es decir, que por el simple hecho que las unidades son distintas entre sí, el modelo cuantifica una componente de variabilidad igual a  $\sigma_{v0}$ . Esto es precisamente lo que justifica su uso como componente de incertidumbre atribuible a la falta de homogeneidad (diferencias) en las unidades, como se definió en la ecuación (2). Asimismo, destaca el hecho que, al incluirse la tendencia de medición como un efecto fijo en el modelo de forma independiente, el mismo no tiene influencia en la variabilidad cuantificada mediante  $\sigma_{v0}$ , convirtiéndola en una estimación “pura”. Para el caso del Ca  $\sigma_{v0} = 0.4536$  mg/L = 0.45 mg/L), es claro que la diferencia entre las unidades causa una variación más evidente en  $\beta_0$  que para el caso del Mg, donde las diferencias respectivas prácticamente no provocan un impacto sobre dicho estimador  $\sigma_{v0} = 3.047 \cdot 10^{-6}$  mg/L = 0.00 mg/L).

Según la ecuación (2), estos valores corresponderían a los respectivos componentes de incertidumbre  $u_{hom}$  para cada estudio de homogeneidad de cada elemento. Posteriormente, se señalan los valores de  $\sigma_\varepsilon$ , los cuales cuantifican la variabilidad aleatoria residual del modelo ajustado después de considerar la tendencia por medición y la diferencia entre unidades como variables explicativas. Por ello, este componente incluye el efecto de la replicación de las unidades del experimento; es decir, cualquier diferencia presente entre las porciones de ensayo de una misma unidad y la repetibilidad misma de aplicar el método de medición, lo que justifica su uso como componente de incertidumbre por repetibilidad del método. Nuevamente, para el caso del Ca ( $\sigma_\varepsilon = 0.4643$  mg/L = 0.46 mg/L), es claro que las diferencias de aplicar el método en réplicas de una misma unidad parecen ser mayores que para el caso del Mg, donde las diferencias respectivas provocan una menor dispersión residual ( $\sigma_\varepsilon = 0.0842$  mg/L = 0.08 mg/L). Según la ecuación (3), estos valores corresponderían a las respectivas componentes de incertidumbre  $u_r$  para cada elemento.

Lo más relevante del presente análisis surge al comparar los resultados del modelo 1 con los obtenidos al ajustar el modelo 2, correspondiente al propuesto por la norma ISO 13528 [3] para el análisis de resultados en estudios de homogeneidad de ensayos de aptitud y que considera únicamente efectos aleatorios para explicar el comportamiento de los datos. Los resultados del ajuste del modelo 2 para ambos elementos, seguido de la aplicación de un ANOVA de una vía, se muestran en la Tabla 4. Por su parte, la comprobación de los respectivos supuestos de los modelos se muestran en la Fig. A2, todos con resultados positivos en cuanto a su cumplimiento.

Tabla 4. Resultados obtenidos del ANOVA aplicado con el ajuste del modelo 2 para ambos elementos

Elemento: Ca	SS	df	MS	Prob. p
Entre grupos ( <i>between</i> )	3.8680	6	0.6447	0.0773
Dentro de los grupos ( <i>within</i> )	1.4138	7	0.2020	

Elemento: Mg	SS	df	MS	Prob. p
Entre grupos ( <i>between</i> )	0.0276	6	0.0046	0.9962
Dentro de los grupos ( <i>within</i> )	0.3844	7	0.0549	

SS: Suma de cuadrados ( $\text{mg}^2/\text{L}^2$ ), df: grados de libertad (1), MS: promedio cuadrado ( $\text{mg}^2/\text{L}^2$ ), Prob. p: probabilidad o valor-p (1).

Respecto a la Tabla 4, se puede observar una estructura típica resultante de un ANOVA de una vía, en la que se podría evaluar estadísticamente la igualdad entre los distintos niveles del factor (en este caso, las unidades ensayadas) a partir de las probabilidades estimadas. Considerando un nivel de significancia de 5 %, podría concluirse tanto que no existen diferencias significativas entre las unidades, para Ca ( $p = 0.0773$ ) y para Mg ( $p = 0.9962$ ), como que hay suficiente homogeneidad entre las unidades sin la necesidad de estimar  $u_{hom}$ . Esta es una práctica señalada como inadecuada por Van der Veen *et al.* [2], aunque común en el análisis de datos para ensayos de aptitud. No obstante, no fue aplicada en la presente investigación. En primer lugar, un valor de la probabilidad es muy cercano al umbral de decisión (0.05), por lo que no existe una evidencia contundente de que las diferencias sean realmente despreciables.

En segundo lugar, y de forma consistente con lo expresado en la norma ISO 13528 [3], los estudios de homogeneidad en ensayos de aptitud deben considerar la magnitud de las diferencias y evaluar si ésta puede llegar a afectar la evaluación de los participantes que medirán unidades similares a las estudiadas. Dicho aspecto técnico no es evaluado directamente por la prueba estadística, por lo que podrían obviarse diferencias significativas al observar solo los valores de probabilidad. Por estas razones es que, a partir de los resultados expuestos en la Tabla 4, se calculan las componentes de incertidumbre respectivas siguiendo las ecuaciones (5) y (6) en lugar de concluir con base en las probabilidades estimadas. Dado que no existen valores faltantes en la base de datos, para ambos elementos se utiliza  $n_0 = 2$  réplicas por unidad.

$$u_{hom}(\text{Ca}) = \sqrt{\frac{0.6447 - 0.2020}{2}} = 0.4705 \text{ mg/L} = 0.47 \text{ mg/L}$$

$$u_r(\text{Ca}) = \sqrt{0.2020} = 0.4494 \text{ mg/L} = 0.45 \text{ mg/L}$$

$$u_{hom}(\text{Mg}) = \sqrt{\frac{0.0046 - 0.0549}{2}} = \text{NaN}$$

$$u_r(\text{Mg}) = \sqrt{0.0549} = 0.2343 \text{ mg/L} = 0.23 \text{ mg/L}$$

De estos resultados surgen algunos aspectos muy relevantes. En primer lugar, la componente  $u_{hom}$  para el Mg no puede ser estimada dentro del ámbito de los números racionales (*NaN*), resultando en la raíz cuadrada de un número negativo. Aunque pueda parecer un error de metodología, esta situación no es extraña para la aplicación y ha sido tratada con anterioridad por Linsinger *et al.* [27], donde se señala que el problema se fundamenta en la dependencia de la estimación de la variabilidad entre unidades con la variabilidad asociada con el método de medición. De hecho, diversos documentos técnicos en la materia [3, 7] consideran la posibilidad de que se presente este problema, y recomienda el uso de la estimación  $u_{hom} = 0$  en caso de obtener resultados indefinidos (*NaN*). Esta aproximación es consistente con lo señalado por Van der Veen *et al.* [2], aunque estos últimos lo razonan cuando existe justificación técnica de que, por su naturaleza de preparación o por experiencia previa, las unidades sean sumamente homogéneas. Seguidamente, es importante señalar que el componente  $u_r$  para el Mg, que corresponde a una desviación estándar estimada a partir de la segregación de la varianza global de los datos, presenta un valor superior a la propia desviación estándar global calculada para los datos (0.2 mg/L). Esto corresponde a una clara posibilidad de sobreestimación de incertidumbre por el propio método de análisis de datos, y junto con la situación discutida anteriormente, debería levantar una señal de advertencia para detener el análisis y profundizar su interpretación y sus causas.

En el caso de Ca, las estimaciones de  $u_{hom}$  y  $u_r$  no parecen presentar mayores problemas de estimación ni anomalías en sus valores resultantes, lo que lleva a considerar que el problema anterior podría presentarse debido a la presencia de una tendencia en el proceso de medición. Esta, al no ser considerada durante el planteamiento del análisis de los datos de Mg utilizando el modelo 2, genera las inconsistencias en la estimación de las respectivas componentes de incertidumbre. De hecho, esta conclusión resulta aún más clara al comparar los resultados obtenidos por ambos modelos, y que se muestran en la Fig. 3.

En la Fig. 3, para el caso de Ca, se nota que ambos modelos arrojan componentes de incertidumbre muy similares, y cuyas diferencias, técnicamente despreciables, son atribuidas al uso de distintas metodologías para el ajuste de los modelos (REML y mínimos cuadrados). Sin embargo, se evidencia diferencias claras entre ambos modelos para el caso de Mg, donde el modelo 1 no presentó los problemas de estimación señalados anteriormente para el modelo 2. En su lugar, el componente  $u_{hom}$  presentó un valor de prácticamente cero, lo cual es consistente con la recomendación dada en la literatura cuando se presentan errores con el modelo 2.

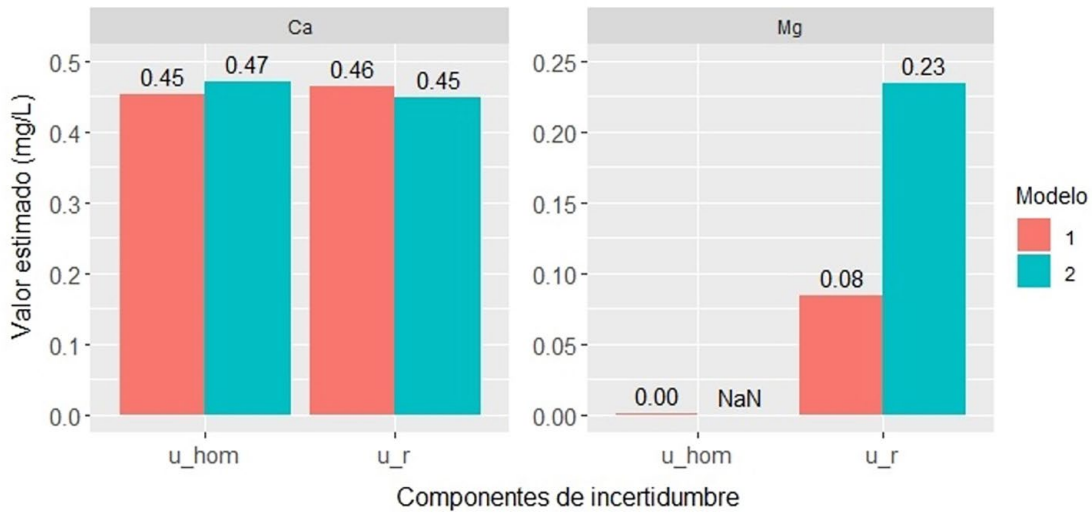


Figura 3. Comparación de las componentes estimadas de incertidumbre para el modelo 1 (mixto lineal) y modelo 2 (aleatorio) ajustados a los datos de Mg y Ca (NaN: No estimable, u\_hom: u\_hom y u\_r: u\_r).

Por su parte, el componente  $u_r$  estimado con el modelo 1, presentó un valor próximo a un tercio del estimado con el modelo 2, el cual es consistente con la desviación estándar global de los datos. Este último aspecto relacionado con el valor de  $u_r$  toma especial relevancia si, en lugar de estimar  $u_{hom} = 0$ , se reconoce que la repetibilidad del método puede influir en su estimación y se procede a calcular un valor máximo esperado de  $u_{hom}$  a partir de  $u_r$ , tal y como recomienda Van der Veen *et al.* [6]. Así, el lograr un posible valor no sobreestimado de  $u_r$  puede generar una mejor estimación en este proceso alternativo de análisis, que puede incluso brindar un valor más conservador para la posterior consideración de su impacto en la evaluación del desempeño de los participantes.

Finalmente, y a raíz de lo evidenciado anteriormente, la inclusión de un diagrama similar al presentado en la Fig. 4 se hace esencial para cualquier procedimiento de análisis de datos aplicado para estudios de homogeneidad en ensayos de aptitud, de manera que se incluyan posibles alternativas para el abordaje del análisis de resultados en caso de presentarse este tipo de tendencias en los datos.

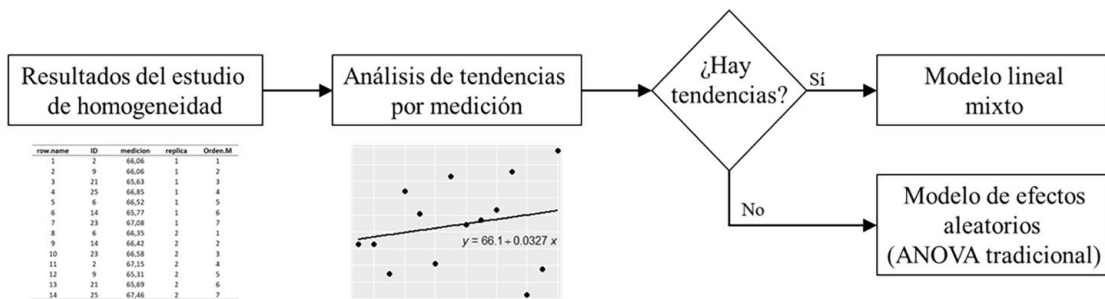


Figura 4. Esquema recomendado para abordar el análisis de resultados de un estudio de homogeneidad

## CONCLUSIONES

A partir del presente estudio, la aplicabilidad del modelo mixto lineal para abordar estudios de homogeneidad en ensayos de aptitud fue evidente. También, resultó apreciable la comparabilidad de los resultados obtenidos con modelos mixtos lineales y aquellos obtenidos con las metodologías tradicionales en el análisis de casos sin tendencias por medición, presentándose diferencias despreciables en las estimaciones para las incertidumbres relevantes en un estudio de homogeneidad.

Además, se demostraron las ventajas del uso de modelos mixtos lineales en el análisis de casos con presencia de tendencias por medición; esto al lograr mejores estimaciones para las mismas incertidumbres consideradas. Se hace especial énfasis en estos últimos casos, en los que los modelos tradicionales abordados con un ANOVA de una vía no presentan buenos resultados y su aplicación puede hacer incurrir en errores a los proveedores de ensayos de aptitud menos experimentados en el análisis estadístico de sus estudios de homogeneidad. Por lo tanto, se recomienda incluir una disposición de no utilizar dichos modelos en presencia de tendencias por medición en los respectivos procedimientos de análisis de datos o normativa de referencia pertinente en general, e incluir un diagrama similar al presentado con anterioridad.

Por último, los resultados mostrados en este estudio dejan patente la necesidad de explorar nuevas metodologías, como los modelos mixtos, durante la ejecución del análisis de datos en ensayos de aptitud y no limitarse a simplemente aplicar aquellas recomendadas en la normativa de referencia, las cuales pueden presentar limitaciones importantes en casos particulares y resultar en estimaciones inadecuadas si no se analizan con el criterio técnico correspondiente.

## 5. AGRADECIMIENTOS

Los autores extienden un agradecimiento a Eiliana Montero Rojas, PhD, docente de la Escuela de Estadística e investigadora del Instituto de Investigaciones Psicológicas de la UCR, por el apoyo y la revisión estadística del contenido expuesto en el presente documento.

**Conflictos de interés:** Los autores declaran no tener conflictos de interés en el desarrollo de esta investigación.

## REFERENCIAS

- [1] *Conformity assessment – General requirements for proficiency testing*, ISO/IEC Standard 17043, 2010.
- [2] A. Van der Veen, T. Linsinger y J. Pauwels, “Uncertainty calculations in the certification of reference materials. 2. Homogeneity study”, *Accreditation and Quality Assurance*, vol. 6, no. 1, pp. 26-30, 2001. doi: 10.1007/s007690000238.
- [3] *Statistical methods for use in proficiency testing by interlaboratory comparison*, ISO Standard 13528, 2015.
- [4] M. Thompson, S. Ellison y R. Wood, “The international harmonized protocol for the proficiency

- testing of analytical chemistry laboratories (IUPAC Technical Report)”, *Pure and Applied Chemistry*, vol. 78, no. 1, pp. 145-196, 2006. doi: 10.1351/pac200678010145.
- [5] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP y OIML, “Evaluación de datos de medición - Guía para la expresión de la incertidumbre de medida (JCGM 100 - GUM 1995 con ligeras correcciones)”, 2008. [En línea]. Obtenido de <http://www.cem.es/sites/default/files/gum20digital1202010.pdf>
- [6] A. Van der Veen y J. Pauwels, “Uncertainty calculations in the certification of reference materials. 1. Principles of analysis of variance”, *Accreditation and Quality Assurance*, vol. 5, no. 12, pp. 464-469, 2000. doi: 10.1007/s007690000237
- [7] *Reference materials — Guidance for characterization and assessment of homogeneity and stability*, ISO Guide 35, 2007.
- [8] J. Hox, *Multilevel Analysis*, 2nd ed, New York, USA: Routledge, 2010.
- [9] A. Galecki y T. Burzykowski, *Linear mixed-effects models using R: a step-by-step approach*, New York, USA: Springer, 2013.
- [10] B. West, K. Welch y A. Galecki, *Linear mixed models: A practical guide using statistical software*, Florida, USA: Chapman & Hall/CRC, 2007.
- [11] E. Montero, “El potencial de los modelos mixtos de efectos fijos y aleatorios para el análisis de datos en la investigación social”, *IV Encuentro Latinoamericano de Metodología de las Ciencias Sociales (ELMeCS)*, Heredia, Costa Rica, Agosto, 2014.
- [12] P. Ciarlini, M. Cox, F. Pavese y D. Richter, “Robust Alternatives to Least Squares” en *Advanced Mathematical Tools In Metrology III*. Berlin, Germany: World Scientific Publishing Company, 1997, pp. 118-133.
- [13] F. Pavese, P. Ciarlini, y G. Regoliosi, “On the use of mixed models in metrology”, *AMCTM Conference*, Turin, Italy, 2003.
- [14] H. Cui, K. Ng, y L. Zhu, “Estimation in mixed effects model with errors in variables”, *Journal of Multivariate Analysis*, vol. 91, no. 1, pp. 53-73, 2004. doi: 10.1016/j.jmva.2004.04.014.
- [15] LCM, “Informe final de resultados - Ensayo de aptitud DMQ-001-2018: Medición de metales en disolución acuosa (LACOMET-10781318)”, 2018. [En línea]. Obtenido de: <http://lcm.go.cr/index.php/documentacion-historica/analisis-fisicoquimicos-historico>.
- [16] R. Walpole, R. Myers, S. Myers y K. Ye, *Probabilidad y estadística para ingeniería y ciencias*, 8va ed, Ciudad de México, México: Pearson Educación, 2007.
- [17] *R: A Language and Environment for Statistical Computing*. (2019). R Core Team: R Foundation for Statistical Computing. Obtenido de <https://www.R-project.org/>
- [18] H. Wickham, J. Hester y R. François. *readr: Read Rectangular Text Data*. (2018). [En línea]. Obtenido de <https://CRAN.R-project.org/package=readr>.
- [19] H. Wickham, R. François, L. Henry y K. Müller. *dplyr: A Grammar of Data Manipulation*. (2019). [En línea]. Obtenido de <https://CRAN.R-project.org/package=dplyr>.
- [20] H. Wickham y L. Henry. *tidyr: Tidy Messy Data*. (2019). [En línea]. Obtenido de <https://CRAN.R-project.org/package=tidyr>.
- [21] E. Waring *et al.* *skimr: Compact and Flexible Summaries of Data*. (2020). [En línea]. Obtenido de <https://CRAN.R-project.org/package=skimr>.
- [22] J. Pinheiro *et al.* *nlme: Linear and Nonlinear Mixed Effects Models*. (2019). [En línea]. Obtenido de <https://CRAN.R-project.org/package=nlme>.



- [23] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. (2016). [En línea]. Obtenido de <https://ggplot2.tidyverse.org>.
- [24] A. Kassambara. *ggpubr: 'ggplot2' Based Publication Ready Plots*. (2019). [En línea]. Obtenido de <https://CRAN.R-project.org/package=ggpubr>.
- [25] P. Aphalo. *ggpmisc: Miscellaneous Extensions to 'ggplot2'*. (2019). [En línea]. Obtenido de <https://CRAN.R-project.org/package=ggpmisc>.
- [26] J. Arnold. *ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'*. (2019). [En línea]. Obtenido de <https://CRAN.R-project.org/package=ggthemes>.
- [27] T. Linsinger *et al.*, "Homogeneity and stability of reference materials", *Accreditation and Quality Assurance*, vol. 6, no. 1, pp. 20-25, 2001. doi: 10.1007/s007690000261.
- [28] SelectScience, "PinAAcle 900 AA Spectrometers by PerkinElmer Inc.", *de Products & Reviews*: <https://www.selectscience.net/products/pinaacle-900-aa-spectrometers/?prodID=107799> (accesado 10 Feb. 2020).

## ANEXOS:

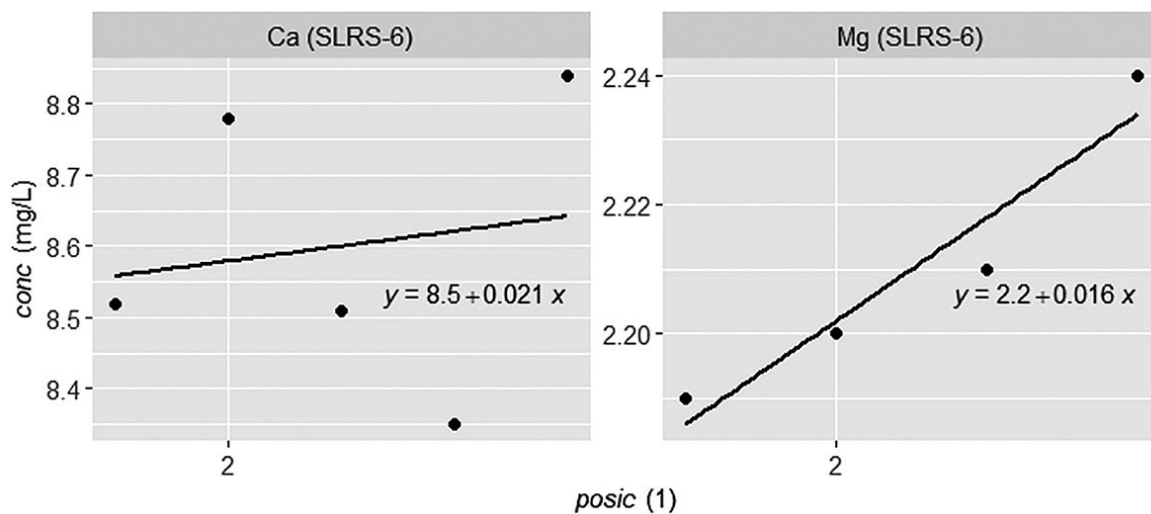


Figura. A1. Ajuste exploratorio de modelos lineales ordinarios para confirmar la posible presencia de tendencias por medición en los datos provenientes de los controles de calidad externos.

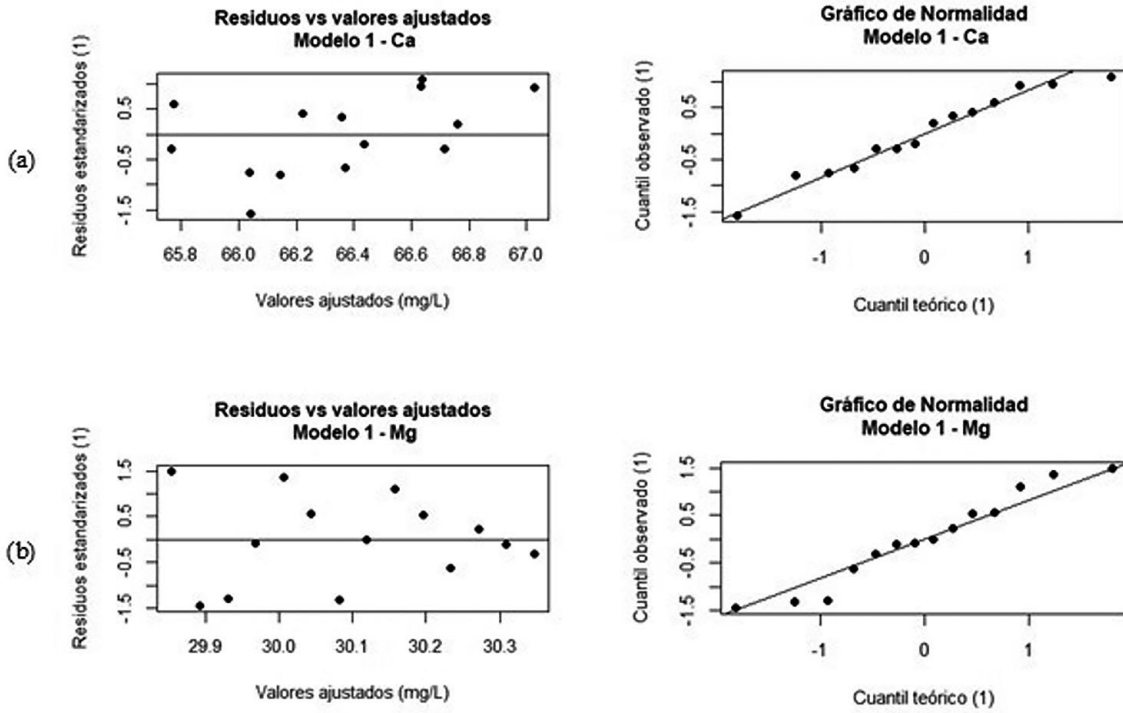


Figura A2. Comprobación gráfica de supuestos del modelo 1 (mixto lineal) para (a) Ca y (b) Mg.

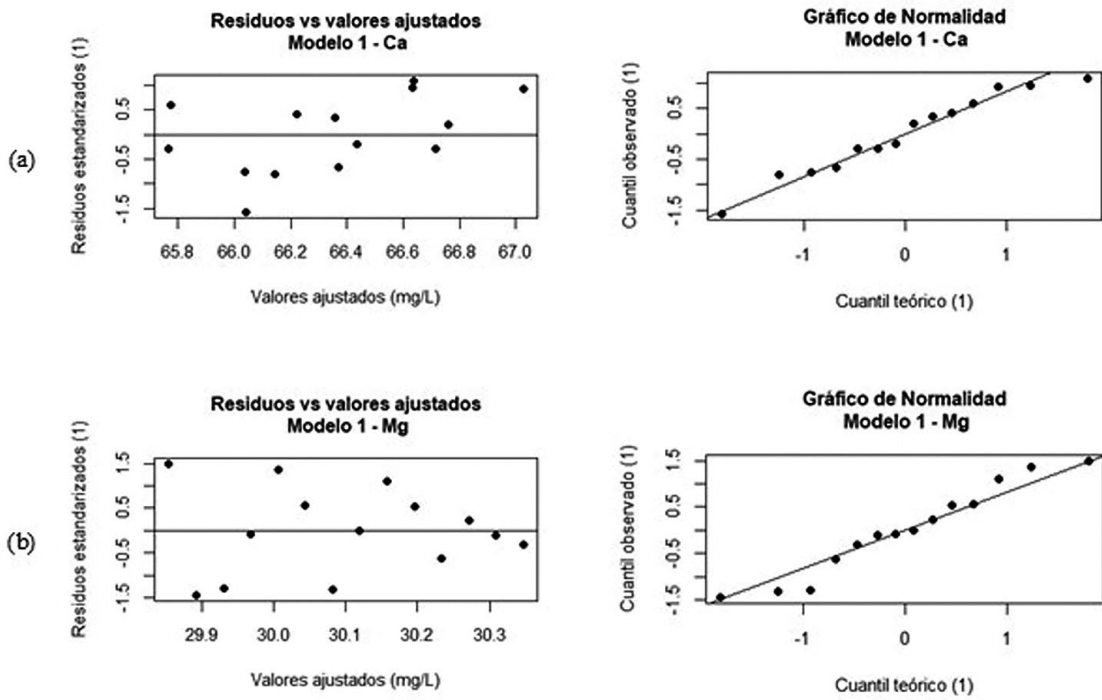


Figura A3. Comprobación gráfica de supuestos del modelo 2 (aleatorio) para (a) Ca y (b) Mg.