

ESTRATEGIA MULTINIVEL PARA DETECTAR COMPORTAMIENTOS DE ABANDONOS EN EL ANÁLISIS DE DATOS LONGITUDINALES

MULTILEVEL STRATEGY FOR DETECTING BEHAVIOURS OF DROPOUT CASES IN THE ANALYSIS OF LONGITUDINAL DATA

WILMER FERMÍN¹, PURIFICACIÓN GALINDO², JAVIER MARTÍN²

¹Departamento de Estadística, Universidad de Oriente, Núcleo de Nueva Esparta, E-mail: wfermin@ne.udo.edu.ve

²Departamento de Estadística, Universidad de Salamanca-España, E-mail: pgalindo@usal.es

RESUMEN

El abandono es un problema frecuente en estudios longitudinales. Cuando el abandono es no-aleatorio (MNAR) se requiere considerar conjuntamente la respuesta principal y el mecanismo de abandono. Los modelos de patrones mezclados de coeficientes aleatorios son un enfoque de modelización para tratar abandonos MNAR. Estos modelos consideran el tiempo de abandono como una covariable que interacciona con las demás en el estudio. En este trabajo se propone un método en el cual los grupos de abandono se consideran como unidades de un tercer nivel de jerarquía y el análisis se realiza mediante un modelo multinivel. Adicionalmente, se describe una estrategia para constatar el posible sesgo causado por los abandonos cuando el análisis de los datos se realiza mediante supresión listwise, inclusión pairwise o sólo-abandonos contra el modelo propuesto. El método y la estrategia se ilustran mediante un conjunto de datos referidos al contenido de proteína en leche de vacas sometidas a tres diferentes dietas.

PALABRAS CLAVE: Análisis multinivel, datos longitudinales, abandonos, tiempo de abandono.

ABSTRACT

Dropout is a frequent problem in longitudinal studies. When dropout is not at random (MNAR), it is required to consider jointly the response and the dropout mechanism. The mixed pattern models of random coefficients have been proposed for longitudinal data when dropout is thought to be MNAR. These models consider the time of dropout as a covariate that interacts with others in the study. In this paper we propose a method in which the groups of dropout are considered as units of a third level of hierarchy and the analysis is carried out by means of a multilevel model. Additionally, a strategy is described to verify the possible bias caused by the dropouts when the data analysis is carried out through listwise suppression, pairwise inclusion or dropout-only cases, versus the proposed model. The method and the strategy are illustrated by means of a set of data referred to the protein content in milk from cows assigned to three different

KEY WORDS: Multilevel analysis, longitudinal data, dropouts, time of drop-outs.

INTRODUCCIÓN

Considérese un estudio longitudinal en el cual se examina una respuesta, Y , sometida a la acción de “ P ” variables explicativas, (X_1, X_2, \dots, X_p) , a “ m ” sujetos, en cada uno de “ n ” puntos u ocasiones de medición. Sea Y_{ij} la respuesta, observable del j -ésimo sujeto en la t -ésima ocasión y $\mathbf{Y}_j = (Y_{1j}, Y_{2j}, \dots, Y_{mj})$ su correspondiente vector de resultados, posiblemente incompleto a causa de datos faltantes. Cuando algunos \mathbf{Y}_j 's contienen códigos: “no se conoce”, “*rehúso*”, “*ausencia*”, etc., o simplemente hay vacíos, se dice que hay datos faltantes, ya que todas las mediciones planificadas no fueron registradas, están ausentes a causa de que algunas unidades de investigación no fueron examinadas.

En investigaciones longitudinales podemos distinguir dos patrones de datos faltantes, según la manera en la cual aparecen en la serie de medición. *Datos faltantes intermitentes*, que se refiere a la situación en la cual una o varias unidades de investigación participan en algunas, pero no en todas las ocasiones del estudio. El otro tipo de dato faltante, y del que se ocupa este trabajo, es el *abandono*, el cual se refiere a la situación cuando una o varias unidades dejan el estudio prematuramente, después de haber participado en una o más ocasiones consecutivas.

Los abandonos son considerados mediante una variable indicadora, R , que hace referencia a la ausencia y se le llama *mecanismo de abandono*. Para el j -ésimo

sujeto, R_j , toma el valor cero cuando el sujeto completa el estudio o r , con $2 \leq r \leq n$, si el sujeto es visto por última vez en la $(r-1)$ -ésima ocasión de medición. Diggle y Kenward (1994) clasificaron el abandono en estudios longitudinales, siguiendo la terminología de Little y Rubin (1987), como sigue: (1) abandono completamente aleatorio (MCAR), en el cual el mecanismo de abandono y el proceso de medición son independientes; es decir, la ausencia es independiente de cualquier covariable, medida o no medida relacionada al proceso de medición, (2) abandono aleatorio (MAR), también llamado ignorable o no-informativo, en el cual el mecanismo de datos faltantes depende de covariables y al menos de una de las respuestas anteriores a la ausencia; cuando la ausencia depende exclusivamente de las covariables consideradas y no de las respuestas anteriores a la ausencia, Little (1995), acuñó el término “abandono dependiente de covariable” y, (3) abandono no-aleatorio (MNAR), también llamado abandono informativo o no-ignorable, en este caso la ausencia está relacionada con las respuestas no observadas, es decir, esas que pudieron haber sido observadas si el sujeto no se ausenta.

Estadísticamente el tipo de abandono más problemático es el no-aleatorio puesto que no se puede ni excluirlos del análisis ni imputarles algún valor sin introducir sesgos. Hay varios métodos para tratar los abandonos Hogan *et al.* (2004) y dos de ellos ampliamente estudiados para modelar datos longitudinales con abandonos potencialmente MNAR son los modelos de selección y los modelos de patrones mezclados. En los modelos de selección se especifica una distribución para los datos completos y luego se modela el mecanismo de abandono condicionado sobre los hipotéticos datos completos. Muchos investigadores han señalado que los resultados de estos modelos descansan sobre las suposiciones asumidas respecto a la forma de la distribución del mecanismo de abandono y de la poblacional; cambios menores pueden alterar significativamente las conclusiones (Kenward, 1998; Schafer y Graham, 2002). En consecuencia, estos modelos se consideran demasiado inestables para aplicaciones científicas y más útiles para erigir preguntas que para generar respuestas (Little, 1995; Demirtas y Schafer, 2003).

Los modelos de patrones mezclados son una alternativa a los de selección (Little 1993). En estos modelos, los sujetos están clasificados por el tiempo de abandono y luego se describen los datos observados dentro de cada grupo de ausencia, incluso en el grupo de los que completan el estudio, para extrapolar aspectos de sus comportamientos a la porción no vista (ausentes) de los datos. Una ventaja inmediata de los modelos de

patrones mezclados es que no requieren amplias teorías de distribución o la especificación precisa de un modelo de abandono; estos modelos son percibidos, por muchos investigadores, como descriptivos más que teóricos. Los modelos de patrones mezclados de coeficientes aleatorios (MPMCA) son una versión ampliada de los anteriores. En éstos se usa el tiempo de abandono como una covariable que interacciona con el tiempo de medición y otras covariables vinculadas en el estudio (Hedeker y Gibbons, 1997; Hogan y Laird, 1997). Entre las limitaciones que presentan los modelos de patrones mezclados, se tiene: (1) impone restricciones en algunos grupos de abandono para identificar parámetros de la distribución asumida para la respuesta principal, y (2) no se pueden obtener directamente las estimaciones de los parámetros de interés, es decir, los de la distribución de la respuesta principal. Se han planteado varias soluciones en ambos casos (Molenberghs *et al.* 1998; Daniels y Hogan, 2000).

Demirtas y Schafer, (2003) plantean una crítica a la consideración del tiempo de abandono como una covariable, ya que ésta no es como una covariable ordinaria; si una covariable ordinaria resulta significativa y no se incluye en el modelo, ocasionará sesgo de predicción; si el coeficiente del tiempo de abandono resulta significativo en el modelo, sin embargo, esto no hace necesario que si se continúa añadiendo potencias del tiempo de abandono mejorará la habilidad para sacar inferencias respecto a la población de interés, representada por los sujetos de todos los grupos de abandono combinados.

En este trabajo se propone una nueva alternativa de análisis para datos longitudinales con abandonos MNAR, tratando los grupos de abandonos como unidades de un tercer nivel de jerarquía y modelizar la respuesta principal mediante un modelo multinivel. El resto de este trabajo está organizado de la siguiente manera. Un apartado Materiales y métodos, donde se describe la metodología propuesta de análisis de datos longitudinales con abandonos potencialmente no-aleatorios y se realiza una aplicación; una sección de resultados de aplicación y comparación con otros métodos y, se finaliza con el apartado de discusión.

MATERIALES Y MÉTODOS

Para ilustrar la propuesta se utilizaron datos, tomados de Verbyla y Cullis (1990) quienes describen los objetivos y diseño del experimento. Los datos se refieren al contenido de proteína en leche de vacas sometidas a diferentes dietas; se determinó el contenido de proteína

en muestras de leche, semanalmente, durante 19 semanas, en cada una de las 79 vacas estudiadas. Las vacas fueron aleatoriamente asignadas a una de tres dietas: cebada, mezcla de cebada-altramuz y altramuz y el número de vacas por dieta fue de 25, 27 y 27 respectivamente. El contenido de proteína se evaluó, a través del tiempo, con el propósito de comparar las tres dietas, pero desde la semana 15 en adelante algunas vacas (38) fueron quedando fuera del estudio por dejar de producir leche. En

la Tabla 1 se presenta el número de animales por dieta y grupo de abandono, el contenido promedio de proteína por dieta y grupo de abandono (en []), total de animales por grupo de abandono, el número de mediciones registradas en cada animal por grupos de abandono y, el promedio y la desviación estándar por grupo de abandono. Diggle, y Kenward, (1994) también analizaron estos datos mediante un modelo de selección considerando un modelo de regresión logístico para el mecanismo de abandono.

Tabla 1. Información de contenido de proteína en la leche por dieta y grupo de abandono

		Patrón de abandono				
		0: no abandono (pt0)	1: abandono en la semana 15 (pt 1)	2: abandono en la semana 16 (pt 2)	3: abandono en la semana 17 (pt 3)	4: abandono en la semana 19 (pt 4)
dieta	Cebada	13 [3,5]	6 [3,17]	2 [3,42]	2 [3,37]	2 [3,37]
	Ceb-altr	14 [3,38]	7 [3,46]	3[3,38]	1[3,44]	2[3,42]
	Altramuz	14 [3,21]	7 [3,31]	4[3,35]	1 [3,36]	1 [3,35]
Total vacas		41	20	9	4	5
Medición por animal		19	14	15	16	18
Promedio		3,7	3,79	3,5	3,59	3,79
Desviación estándar (sd)		0,357	0,177	0,236	0,169	0,109

[3,5]: Cantidad de proteína promedio por dieta y grupo de abandonado

Análisis multinivel de datos longitudinales.

El análisis multinivel o análisis de regresión jerárquico es un enfoque para analizar datos multinivel. Estos datos se caracterizan porque las unidades de observación están agrupadas dentro de otras en un nivel superior y porque exhiben patrones complejos de variabilidad (Bryk y Raudenbush, 1992). La metodología multinivel establece una ecuación para modelar los datos tomando en cuenta la estructura jerárquica y la posible incorporación de variables de todos los niveles de la jerarquía. La ecuación o modelo multinivel es una combinación de submodelos, uno para cada nivel de la estructura jerárquica, quedando especificado cómo las variables en un nivel dado afectan las relaciones en otros niveles. Aplicaciones del análisis multinivel en diversas disciplinas han progresado

aceleradamente en los últimos años Goldstein, (1995); Fermín y Gallindo (2003) debido a la aceptación de la noción de datos multinivel, el desarrollo de algoritmos y métodos de estimación y la implementación de estos algoritmos en programas que permiten el uso práctico de los modelos multinivel. Algunos programas para multinivel son: MLn y MLwiN Rasbash *et al.* (2001); MIXREG (Hedeker y Gibbons, 1996). Para una mayor revisión de los software que permiten ajustar modelos multinivel se puede consultar a (Heo *et al.* 2003).

Los datos longitudinales pueden ser considerados como multinivel con estructura jerárquica, como mínimo de dos niveles: ocasiones de medición (nivel uno) anidadas en los sujetos (nivel dos). Cuando los sujetos, a su vez, están anidados en grupos entonces los datos tienen una

estructura jerárquica de tres niveles, siendo los grupos las unidades del nivel tres. Varias ventajas se hacen presentes al analizar datos longitudinales usando modelos multinivel Hox (2002), entre ellas: (1) Se obtienen curvas de crecimiento para cada unidad del nivel dos (y niveles superiores) que caracterizan el crecimiento en términos de sus tasas promedios de cambio, aceleración, etc., y que se pueden usar como diagnóstico para detectar esos con patrones anormales de crecimiento, y (2) con la inclusión de variables explicativas, de los distintos niveles es posible analizar cómo cambian las estimaciones de los parámetros de crecimiento a través de los grupos y cuánto de las diferencias entre grupos y dentro de los sujetos puede ser explicada por otros factores. Se han señalado estas ventajas puesto que son importantes para comprender la filosofía de la propuesta de considerar los grupos de abandono como unidades de un tercer nivel de jerarquía para tratar los abandonos potencialmente no-ignorables.

Los grupos de abandono como nivel superior de jerarquía

Dada la limitación planteada por Demirtas y Schafer (2003) con el tiempo de abandono como una covariable, se propone como alternativa al análisis de datos longitudinales con abandonos no-aleatorios, que los grupos de abandonos sean considerados como unidades de un tercer nivel de jerarquía; es decir, ocasiones agrupadas en sujetos y sujetos anidados en grupos de abandono. Mediante un modelo multinivel de tres niveles se puede caracterizar el comportamiento de la respuesta sobre todos los grupos de abandono y, además, en cada grupo particular; de esta manera se proporciona una mayor comprensión del mecanismo de abandono dentro de los grupos.

El modelo multinivel para los datos longitudinales con abandonos, estructurados en la jerarquía de tres niveles (ocasiones agrupadas en sujetos agrupados en tiempos de abandono), es la combinación de los submodelos de cada nivel. Explícitamente, supóngase que Y_{tjk} es la respuesta observada en la t -ésima ocasión del j -ésimo sujeto dentro del k -ésimo grupo de abandono y que $\mathbf{X}_{tjk} = (X_{1tjk}, X_{2tjk}, \dots, X_{ptjk})$ y $\mathbf{Z}_{jk} = (Z_{1jk}, Z_{2jk}, \dots, Z_{Qjk})$ son variables explicativas definidas en el nivel uno (N1) y nivel dos (N2) respectivamente; el tiempo de medición y potencias de éste generalmente están incluidas en \mathbf{X}_{tjk} . En el nivel de grupos de abandono o nivel tres (N3) no se consideran variables explicativas ya que, en éste se quiere caracterizar la magnitud de la variabilidad entre los grupos y cómo difieren grupo a grupo los \mathbf{X}_{tjk} y \mathbf{Z}_{jk} . Para dar mayor generalidad al modelo, supóngase que de las P covariables

definidas en N1, p_1 admiten coeficientes fijos; p_2 admiten coeficientes aleatorios en N1, N2 y N3, y P_3 ($p_3 = P - p_1 - p_2$) variables explicativas que admiten coeficientes aleatorios en N2 y N3. Definir otros coeficientes aleatorios en N1, después del intercepto, permite modelar una estructura de covarianza compleja en dicho nivel Goldstein *et al.* (1994); los coeficientes aleatorios de N2 serán funciones de \mathbf{Z}_{jk} ; los coeficientes aleatorios de N3 serán de variación simple. El modelo multinivel de tres niveles, con las especificaciones señaladas, viene dado por:

$$Y_{tjk} = \sum_{h=0}^p \gamma_h X_{htjk} + \sum_{h=0}^{p_2+p_3} V_{h0k} X_{htjk} + \sum_{h=0}^{p_2+p_3} \sum_{s=0}^Q \gamma_{h0s} Z_{sjk} X_{htjk} + \sum_{h=0}^{p_2+p_3} \sum_{s=0}^Q V_{hsk} Z_{sjk} X_{htjk} + \sum_{h=0}^{p_2+p_3} U_{htjk} X_{htjk} + \sum_{h=0}^{p_2} e_{htjk} X_{htjk} \tag{1}$$

Las componentes e_{htjk} , U_{htjk} y V_{hsk} , representan las fluctuaciones o residuos al nivel de ocasiones, sujetos y clases de abandono respectivamente. Se supone independencia entre estas componentes nivel a nivel, pero pueden estar relacionadas las componentes en un mismo nivel. En forma matricial, $\mathbf{\Omega}_1$ es la matriz $p_2 \times p_2$ que contiene las varianzas y covarianzas de las componentes aleatorias, e_{htj} 's, del nivel uno; el caso más simple de modelización de estas componentes es cuando únicamente el intercepto es aleatorio al nivel uno; en este caso $\mathbf{\Omega}_1 = \sigma_e^2$. $\mathbf{\Omega}_2$ es la matriz $(p_2 + p_3) \times (p_2 + p_3)$ que contiene las varianzas y covarianzas de las componentes aleatorias, U_{htj} 's, y $\mathbf{\Omega}_3$ es la matriz que contiene las varianzas y covarianzas de N3, generalmente $\mathbf{\Omega}_3 = \sigma_v^2$ la cual representa la varianza del intercepto al nivel tres.

El aspecto clave para diferenciar los grupos de abandonos, es la matriz de varianza-covarianza al nivel tres. Si la variabilidad no es significativa al nivel tres, las clases de abandono son homogéneas con igual comportamiento al de los casos completos; si la variabilidad es significativa se debe examinar cómo cambia esta varianza cuando se incluyen, en el modelo, las diferentes covariables tanto de N1 como de N2, cómo cambian los valores medios de los parámetros a través de las clases, y que parte de las diferencias entre las clases de abandono y de la variación dentro de los sujetos puede ser explicada por las covariables del estudio.

Estrategia para destacar comportamientos de abandono

Se pueden aprovechar las ventajas del análisis multinivel de datos longitudinales para explorar las características del comportamiento de esos sujetos que

abandonan, individualmente y en grupos de abandono. La estrategia propuesta para comparar analítica y gráficamente los modelos ajustados en los casos completos, en los casos disponibles y sólo los que abandonan frente al modelo de tres niveles con el tiempo de abandono como tercer nivel de jerarquía está basada en:

- 1) Ajustar un modelo de dos niveles en la respuesta principal para:
 - (a) Los casos disponibles (inclusión pairwise) y analizar los patrones de cambio.
 - (b) Los casos completos (supresión listwise) y analizar los patrones de cambio.
 - (c) El conjunto datos de los sujetos que abandonaron (sólo-abandonos) y analizar los patrones de cambio.
- 2) Ajustar un modelo de tres niveles en la respuesta principal bajo la estructura de tres niveles: ocasiones anidadas en sujetos anidados en grupos de abandono.
- 3) Valorar y contrastar gráfica y analíticamente los modelos de (a), (b), (c) y (2) para detectar posibles diferencias y la magnitud del sesgo.

Ventajas

Analizar datos longitudinales con abandonos potencialmente no-aleatorios requiere considerar conjuntamente la respuesta principal y el mecanismo de abandono. Con la propuesta que se está presentando se considera simultáneamente ambos aspectos. Dentro de las ventajas están, en primer lugar, no se modela explícitamente el mecanismo de abandono tal como en el caso de los modelos de selección, situación que puede ser difícil e inútil. Una segunda ventaja es que el tiempo de abandono no se considera una covariable en el modelo como es el caso de los modelos de patrones mezclados de coeficientes aleatorios. Por otra parte, mediante la modelización multinivel, considerando los grupos de abandono como unidades de jerarquía, se obtiene una caracterización promedio sobre todos los grupos y una para cada grupo. Además, existen muchos software de análisis multinivel que pueden ser usados para la metodología propuesta.

Aplicación

Se caracterizó el comportamiento del contenido de

proteína (respuesta principal) en función del factor dieta (de N2) y el tiempo de medición, en semanas (de N1); éste último se consideró en el modelo con un término lineal y uno cuadrático. Se ajustó un modelo de dos niveles en: (1) los casos disponibles, (2) los casos completos y (3) en el conjunto de datos de todas las vacas que fueron retiradas (sólo-abandonos). Por otra parte, los grupos de abandono obtenidos mediante la estratificación por tiempo de abandono, se consideraron unidades del nivel tres; a esta estructura se ajustó un modelo de tres niveles. Las estimaciones se obtuvieron con el software MLwiN 1.10 Rasbash *et. al.* (2001), bajo el método mínimo cuadrado generalizado iterativo (IGLS). Las figuras 1, 2, 3 y 4 también fueron obtenidas mediante este software.

RESULTADOS

La Tabla 2 presenta los resultados de cada modelo estimado. En los modelos estimados para todos los casos, excepto para el caso sólo-abandonos, se obtuvo diferencias significativas en el contenido promedio de proteína en la leche para las tres dietas. En el modelo estimado desde los casos disponibles, el mayor contenido de proteína se presentó en la dieta de cebada ($\gamma_0 = 3,7402$) seguida por la dieta cebada+altramuz ($\gamma_1 = 3,6421$) y, última altramuz ($\gamma_2 = 3,5475$); éstos resultados son similares al del modelo de tres niveles; con el modelo estimado para los casos completos se aprecia mejor la diferencia entre las dietas.

Con el modelo estimado desde los datos del caso sólo-abandonos no es posible apreciar diferencias entre las dietas; este resultado contiene información respecto a posibles sesgos en los parámetros estimados en el modelo para los casos disponibles y casos completos; además, puede ocultar diferencias reales entre las dietas o en la caracterización del perfil promedio. Otra valoración importante se refiere al comportamiento cuadrático, con una ligera subida, tanto para los casos disponibles como para los casos completos. Sin embargo, el modelo estimado con el de abandono como jerarquía decrece linealmente, puesto que el coeficiente del tiempo (γ_3) es significativo, mientras que el cuadrático (γ_4) no lo es. Este resultado que se obtiene con el modelo de tres niveles está en armonía con el de Diggle y Kenward (1994) quienes se basaron en un modelo de selección con el mecanismo de abandono caracterizado mediante un modelo logístico. En relación con los efectos aleatorios, en el nivel uno la menor variación la sostiene el modelo para los casos disponibles ($\sigma_E^2 = 0,054$) mientras que la más grande es para el modelo estimado desde los datos del caso sólo-abandonos, esta última varianza es seguida por la del modelo de tres niveles con los grupos de abandono como

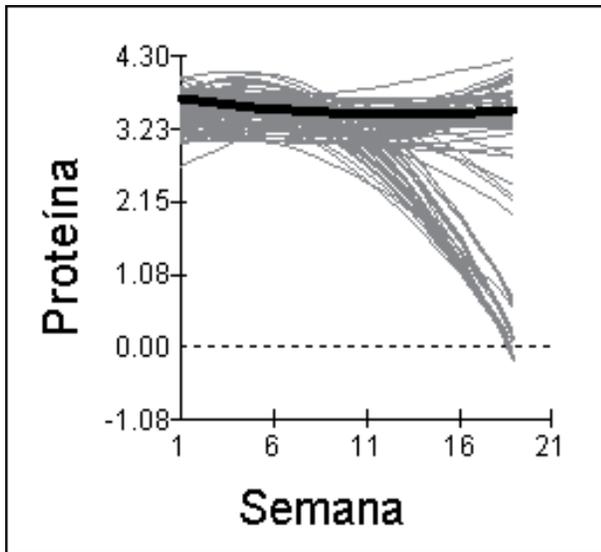
unidades del nivel tres. Las componentes de varianza del nivel dos ($\sigma_{\mu 0}^2$, $\sigma_{\mu 3}^2$ y $\sigma_{\mu 4}^2$) y las interacciones de las componentes aleatorias en ese nivel ($\sigma_{\mu 03}$, $\sigma_{\mu 04}$ y $\sigma_{\mu 34}$) resultaron significativas en los diferentes modelos pero algunas resultaron verdaderamente pequeñas. Las componentes de varianza del nivel tres ($\sigma_{v 0}^2=0,0118$ y $\sigma_{v 4}^2$

$=0,0000$) resultaron significativas, aunque desde un punto de vista práctico, la componente aleatoria para el tiempo al cuadrado parece insignificante; con esto se puede concluir que existe diferencia entre el comportamiento de la respuesta en los grupos de abandono; esta diferencia está fundamentalmente con los casos completos.

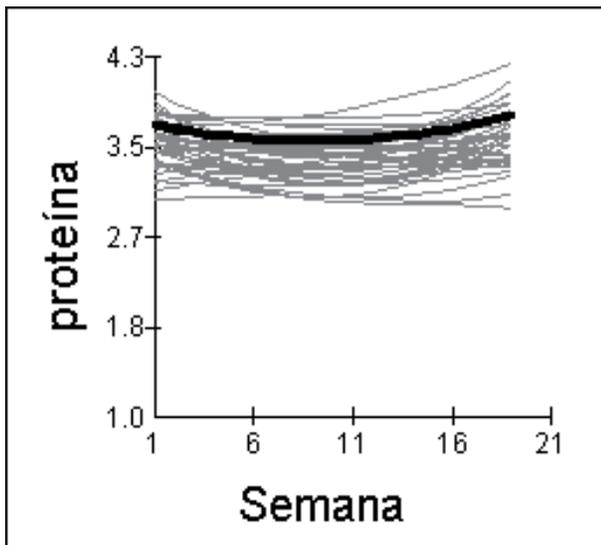
Tabla 2. Resultados de los modelos ajustados a los datos, de las pruebas del contenido de proteína en la leche, clasificados en: casos disponibles, casos completos, abandonos solamente y por patrón de abandono.

Resultados de los diferentes modelos				
Parámetros	Casos disponible	Casos completos	Solo-abandonados	Grupo de abandono (tres niveles)
Efectos fijos				
γ_0 (cebada)	3,7402 (0,0516)	3,7228 (0,0632)	3,6907 (0,0006)	3,7163 (0,0592)
γ_1 (ceb-altr)	-0,0981(0,0479)	-0.1762 (0,0632)	0,0010 (0,0742)	-0,1063 (0,037)
γ_2 (altramuz)	-0,1927 (0,048)	-0,3341 (0,058)	-0,0155 (0,0743)	-0,1727 (0,037)
γ_3 (tiempo)	-0,0475 (0,0091)	-0,0435 (0,0115)	-0,0277 (0,0109)	-0,0372 (0,0074)
γ_4 (tiempo al cuadrado)	0,0018 (0,0005)	0,0024 (0,0005)	-0,0006 (0,0006)	0,0008 (0,0006)
Efecto aleatorio				
σ_E^2	0,0539 (0,0023)	0,0572 (0,0032)	0,0649 (0,0042)	0,0608 (0,0025)
$\sigma_{\mu 0}^2$	0,1028(0,022)	0,0809 (0,0254)	0,0614 (0,0186)	-
$\sigma_{\mu 3}^2$	0,0042 (0,001)	0,0037 (0,0012)	0,0006 (0,0002)	0,0018 (0,0004)
$\sigma_{\mu 4}^2$	0,0000 (0,0000)	0,0000 (0,0000)	-	0,0000 (0,0000)
$\sigma_{\mu 03}$	-0,0161 (0,0043)	-0,0145 (0,0052)	-0,0045 (0,0017)	-
$\sigma_{\mu 04}$	0,0006 (0,0002)	0,0006 (0,0002)	-	-
$\sigma_{\mu 34}$	-0,0002 (0,0000)	-0,0002 (0,0000)	-	-0,0001 (0,0000)
$\sigma_{v 0}^2$				0,0118 (0,0087)
$\sigma_{v 4}^2$				0,0000 (0,0000)
$\sigma_{v 04}$				-0,0001(0,0000)
-2logverosimilit	273,03	152,77	186,17	295,86

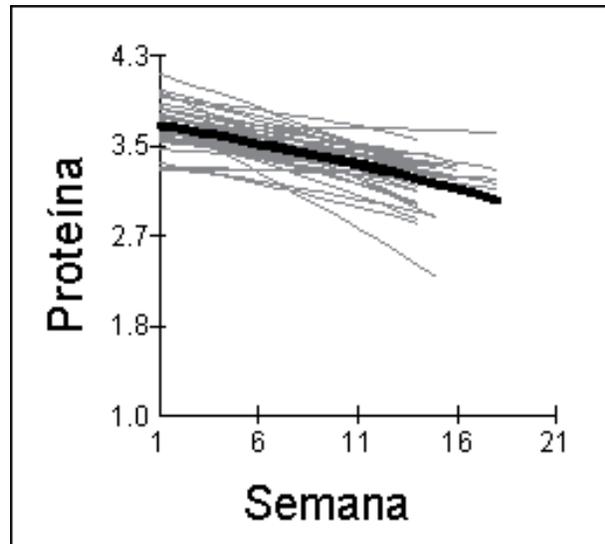
_ No se incluye en el modelo



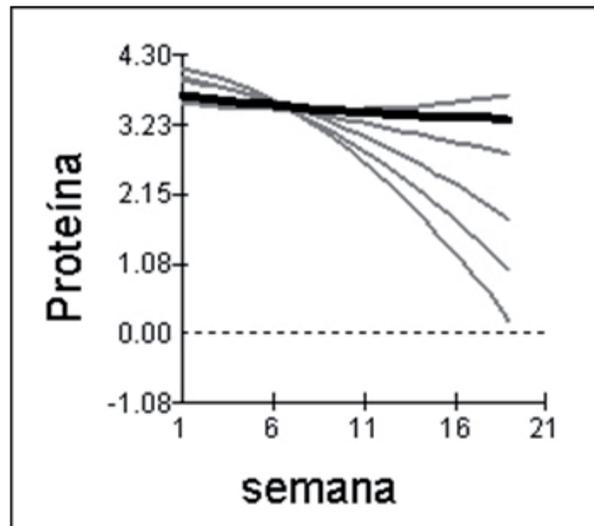
La Figura 1, obtenida mediante el software MLwiN 1.10, muestra cómo la curva de regresión promedio, para los casos disponibles, tiene una ligera subida a través del tiempo mientras que, examinando el comportamiento individual de cada animal, en general, se forman cuatro grupos bien definidos, dos de los cuales tienen un declive acentuado.



La Figura 2 presenta la curva de regresión promedio ajustada desde los casos completos y la curva para cada animal; se puede apreciar que son esas las que corresponden al grupo que está más arriba de los cuatro que se aprecian en la Figura 1.



En la Figura 3 se aprecia un comportamiento opuesto al descrito en las Figuras 1 y 2; Se puede ver que hay un acentuado declive del perfil del contenido de proteína, evidenciándose diferencias entre comportamientos de los casos completos y los abandonos.



La Figura 4 presenta la curva de regresión promedio para cada grupo de abandono, incluso la de los casos completos (pt0) puesto que éste también se incluye como grupo y, la curva promedio sobre todos los grupos (la más gruesa y en negritas); los abandonos en la semana 15 (pt1) están acentuadamente separados de ésta curva promedio. En esta última se puede apreciar que es casi constante a través del tiempo.

DISCUSIÓN

En este trabajo se ha propuesto un método para tratar los abandonos potencialmente no-aleatorios y una estrategia para destacar el comportamiento de los casos de abandonos. El método consiste en tratar los grupos que se forman, mediante la estratificación por el tiempo de abandono, como unidades de un tercer nivel de jerarquía y analizar los datos mediante un modelo de tres niveles. Esto simplifica el análisis sin necesidad de hacer restricciones de identificación ni de promediar sobre cada grupo de abandono para obtener los parámetros de la respuesta principal (Little, 1993). Además, no se da la connotación al tiempo de abandono como covariable que interacciona con las demás; por el contrario, se proporciona una mejor comprensión del mecanismo de abandono a través de la caracterización del comportamiento global, por grupo de abandono y por unidad experimental. En cuando a su comparación con los modelos de selección para tratar abandonos no aleatorios, nuestra propuesta no requiere modelizar explícitamente ni hacer suposiciones sobre el mecanismo de abandono, sin embargo se puede descubrir el comportamiento de cada grupo de abandono usando la información que los propios datos proveen.

Con respecto a la estrategia, ésta permite destacar el comportamiento de los casos completos, casos disponibles, sólo-abandonos y los diferentes grupos de abandono mediante representaciones gráficas de los modelos correspondientes. La estrategia no representa una prueba de hipótesis sobre la relación del proceso de medición y el mecanismo de abandonos (para decidir si el mecanismo de abandono es MCAR, MAR o MNAR) pero es verdaderamente útil para destacar el comportamiento de cualquier sujeto del estudio; en particular, de esas unidades que abandonan.

En el análisis de los datos, mediante la alternativa propuesta y los demás modelos, se puede decir que el perfil de la respuesta promedio de los casos completos, fue muy parecido al de los casos disponibles; además, en ambos se detectaron diferencias significativas entre las dietas respecto al contenido de proteína, descenso lineal y un ligero crecimiento cuadrático al final del experimento. En el modelo de tres niveles no se captura ese ligero crecimiento cuadrático (γ_4 0,0008 con error estándar de 0,0006). La reconciliación en estos resultados subyace en el mecanismo de abandono; éste está influyendo en la interpretación del perfil promedio de la respuesta principal; en los casos completos y casos disponibles el proceso de abandono no es considerado mientras que con el modelo de tres niveles si lo está y por tanto las

estimaciones en este último están amoldadas al proceso de abandono, originando así, una interpretación más real del comportamiento del perfil promedio de la respuesta. Los resultados del modelo de tres niveles no difieren significativamente de los de Diggle y Kenward (1994), quienes usan un modelo de selección, el cual es mucho más complicado ya que requiere modelar explícitamente el mecanismo de abandono en función de las dietas, el tiempo y las observaciones anteriores al abandono y, además, se requiere software específicos que aún no están disponibles.

AGRADECIMIENTO

Agradecemos al Dr. José Francisco Vallejo por sus recomendaciones en la parte gramatical y discursiva del texto.

REFERENCIAS BIBLIOGRÁFICAS

- BRYK, A. S.; RAUDENBUSH, S. W. 1992. Hierarchical linear models, applications and data analysis methods. Newbury Park, CA: Sage Publications.
- DANIELS, M.J.; HOGAN, J. W. 2000. Reparameterizing the pattern mixture model for sensitivity analyses under informative dropout. *Biometrics*, 56,1241-1248.
- DEMIRTAS, H.; SCHAFER, J. L. 2003. On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine*, 22, 2553-2575.
- DIGGLE, P. J.; KENWARD, M.G. 1994. Informative dropout in longitudinal data analysis (with Discussion). *Applied Statistics*, 43, 49-93.
- FERMÍN, W.; GALINDO, P. 2003. Modelos multinivel para analizar datos del estado ecológico de ecosistemas acuáticos subdivididos a causa de su heterogeneidad. EMA'03.
- GOLDSTEIN, H., M. J. R.; RASBASH, J. 1994. Multilevel time series models with applications to repeated measures data. *Statistics in Medicine*, 13, 1643-1655.
- GOLDSTEIN, H. 1995. *Multilevel Statistical Model*. Second Edition. London: Edward Arnold.
- HEDEKER, D.; GIBBONS, R. D. 1996. MIXREG: A computer program for mixed-effects regression analysis with autocorrelated errors. *Computer Methods and*

- Programs in Biomedicine, 49, 229-252.
- HEDEKER, D.; GIBBONS, R. D. 1997. Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, 2, 64-78.
- HEO, M.; FAITH, M. S.; MOTT, J. W.; GORMAN, B. S.; REDDEN, D. T.; ALLISON D. B. 2003. Tutorial in Biostatistics. Hierarchical linear models for the development of growth curves: an example with body mass index in overweight/obese adults. *Statistics in Medicine*, 22, 1911-1942.
- HOGAN, J. W.; LAIRD, N. M. 1997. Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, 16, 239-257.
- HOGAN, J. W.; JASON R.; KORKONTZELOU, C. 2004. Tutorial in biostatistics. Handling drop-out in longitudinal studies. *Statistics in Medicine*, 23, 1455-1497.
- HOX, J.J. 2002. *Multilevel analysis: Techniques and applications*. Lawrence Erlbaum Avenue, New Jersey.
- KENWARD, M.G. 1998. Selection models for repeated measurements with nonrandom dropout: an illustration of sensitivity. *Statistical in Medicine*, 17, 2723-2732.
- LITTLE, R. J. A. 1993. Pattern-Mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88, 125-134.
- LITTLE, R. J. A. 1995. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90, 1112-1121.
- LITTLE, R. J. A.; RUBIN, D. B. 1987. *Statistical analysis with missing data*. New York: John Wiley.
- MOLENBERGHS, G.; MICHIELS, B.; KENWARD, M. G.; DIGGLE P. J. 1998. Missing dat mechanisms and patterns mixture models. *Statistica Neerlandica*, 52, 153-161.
- RASBASH, J.; BROWNE, W.; GOLDSTEIN, H.; YANG, M.; PLEWIS, I.; HEALY, M.; WOODHOUSE, G.; DRAPER, D.; LANGFORD, I.; LEWIS T. 2001. *A user's guide to MLwiN*. Centre for Multilevel Modelling, Unviversity of London. London- England.
- SCHAFFER, J.L.; GRAHAM, J.W. 2002. Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147-177.
- VERBYLA, A.P.; CULLIS, B. R. 1990. Modelling in repeated measures experiments. *Applied Statistics*, 39, 341-356.