

## Modelos de transición en presencia de pérdidas intermitentes: implementación y ejemplo de aplicación a un ensayo clínico cubano

Rolando Uranga\*

Centro Nacional Coordinador de Ensayos Clínicos. Calle 23, esq. 200. Reparto Atabey. Playa. Ciudad de La Habana, Cuba.

email: rolando@cencec.sld.cu

El presente trabajo brinda una herramienta útil para enfrentar el ajuste de los modelos de transición bajo el problemático escenario de patrones de pérdidas intermitentes en datos recolectados de tipo longitudinal. Mediante las facilidades implícitas en la instrucción NLMIXED del software estadístico SAS (Statistical Analysis System) para Windows, versión 9.1.3, se da solución a esta problemática. Se ofrece, además, una aplicación práctica, donde se ajusta un modelo de transición a los datos de laboratorio clínico de un ensayo que evaluó, entre otros objetivos, la seguridad de una vacuna cubana contra el melanoma cutáneo metastásico.

**Palabras clave:** Datos longitudinales, modelo de transición, patrones de pérdidas, mecanismo de pérdidas, ensayo clínico.

### Introducción

Cuando se observa una determinada característica de manera repetida en el tiempo, se habla de datos de tipo longitudinal. En la agricultura se colecta, por ejemplo, la cosecha anual de azúcar de remolacha y su precio por tonelada; las secciones de negocios de los periódicos reportan precios diarios de provisiones, tasas mensuales de desempleo y producciones anuales; la Meteorología recopila velocidades de vientos hora a hora, temperaturas diarias máximas y mínimas y precipitaciones anuales; la Geofísica observa constantemente la agitación o el temblor de la Tierra, con el objetivo de predecir posibles terremotos inminentes.

Un electroencefalograma sondea ondas cerebrales reportadas por un electroencefalógrafo con el objetivo de detectar una enfermedad cerebral y un electrocardiograma sondea ondas cardíacas; las ciencias sociales examinan tasas anuales de muertes y nacimientos, el número de accidentes domésticos y diversas formas de actividad criminal.

Los parámetros en un proceso de fabricación se monitorean permanentemente durante la ejecución de inspecciones de mantenimiento de la calidad. Resulta también importante la evaluación periódica de la respuesta inmune inducida por vacunas preventivas o terapéuticas, así como su seguridad y reactogenicidad.

Existen, obviamente, innumerables razones para recolectar y analizar datos de tipo longitudinal. Entre estas está el deseo de obtener un mejor entendimiento del mecanismo generador de datos, la predicción de valores futuros o el control óptimo de un sistema. La propiedad característica de los datos longitudinales es el hecho de que no se generan

independientemente. Los procedimientos estadísticos que asumen datos independientes e idénticamente distribuidos quedan, por tanto, relegados, y se precisan métodos específicos de análisis (1).

Los modelos condicionales y en particular los modelos de transición que asumen una dependencia lineal de la observación actual en la historia reciente, constituyen útiles herramientas disponibles para enfrentar el escenario descrito. Sin embargo, un aspecto que complica el contexto es la posible presencia de datos faltantes en la secuencia de datos, que puede contener patrones monótonos o bien, en general, patrones intermitentes de pérdidas. Molenberghs y Verbeke (2) describen detalladamente el modo de implementar modelos de transición mediante un código elaborado sobre el paquete estadístico Statistical Analysis System (SAS), en presencia de patrones de pérdidas monótonos, pero no se hallan herramientas disponibles en un software estándar que permitan el ajuste de modelos de transición en el caso general de la presencia de patrones de pérdidas intermitentes.

El objetivo primario del presente trabajo es proporcionar un código elaborado en SAS, capaz de ajustar modelos condicionales en presencia de patrones de pérdidas intermitentes. Los resultados se ilustran mediante el análisis de los datos de un ensayo clínico que evaluó, entre otros objetivos, la seguridad de una vacuna cubana contra el melanoma cutáneo metastásico.

La seguridad de la vacuna se estudió mediante variables de laboratorio clínico como hemoglobina, leucocitos, basófilos, linfocitos, monocitos, entre otros. A los efectos de nuestro ejemplo se considerará solo la hemoglobina.

\* Máster en Ciencias Matemáticas del Centro Nacional Coordinador de Ensayos Clínicos. Calle 23, esq. 200. Reparto Atabey. Playa. Ciudad de La Habana, Cuba.

**Métodos**

Cuando los valores de una respuesta repetida en el tiempo se observan en su totalidad, o bien se observan parcialmente pero de manera tal que a partir del primer valor no observado tampoco se observan los valores sucesivos, estamos en presencia de un patrón de pérdidas de tipo monótono. Si por el contrario, la secuencia de valores faltantes es arbitraria, entonces el patrón de pérdidas se dice de tipo intermitente.

En este trabajo se reformula una variante del modelo de transición de orden n clásico, de manera que puedan asimilarse datos faltantes intermitentes que sustituyan la formulación condicional usual por una formulación marginal. Con la variante clásica la implementación es posible solo en presencia de patrones de pérdidas monótonos del vector de respuestas, mientras que la formulación marginal permite una implementación computacional, en el caso general de patrones de pérdidas intermitentes.

**Modelo clásico de transición de orden n**

Dado un escenario hipotético, donde se colectan datos de tipo longitudinal, denotemos por  $Y_j$  el valor de la variable de respuesta de un individuo genérico en la ocasión  $j$ . La siguiente relación caracteriza el llamado modelo de transición de orden  $n$  (2, 3):

$$Y_j = X_j\beta + \alpha_1 Y_{j-1} + \alpha_2 Y_{j-2} + \dots + \alpha_n Y_{j-n} + \varepsilon_j$$

El modelo introduce de este modo una dependencia lineal del valor actual de la respuesta  $Y_p$  en términos de valores previos. La matriz de diseño  $X$ , conformada por las filas  $X_p$ , recoge valores de covariables de interés, en tanto que el vector  $\beta$  contiene parámetros desconocidos. Se asume que el término de error  $\varepsilon_j$  es una variable aleatoria normal de media cero.

**Variante del modelo clásico de transición de orden n a implementar**

Es conveniente, sin embargo, introducir el modelo de transición de orden n mediante la siguiente formulación equivalente (3):

$$\begin{aligned} Y_j &= X_j\beta + \delta_j \\ \delta_j &= \alpha_1 \delta_{j-1} + \alpha_2 \delta_{j-2} + \dots + \alpha_n \delta_{j-n} + \varepsilon_j \end{aligned} \tag{1}$$

La formulación (1) tiene la ventaja de que en ella el vector de parámetros desconocidos  $\beta$  tiene una interpretación marginal, a diferencia de la formulación clásica donde es difícil de interpretar. La relación de transición se expresa ahora mediante los términos auxiliares  $\delta_p$  que representan la "respuesta centrada", como sugiere la relación.

$$\delta_j = Y_j - X_j\beta$$

Puede verificarse, en efecto, que el modelo (1) expresa la respuesta actual como función lineal de valores previos, si se sustituyen los  $\delta$ 's y se reagrupan términos, como se muestra a continuación:

$$\begin{aligned} \delta_j &= \alpha_1 \delta_{j-1} + \alpha_2 \delta_{j-2} + \dots + \alpha_n \delta_{j-n} + \varepsilon_j \\ &\Rightarrow \\ Y_j - X_j\beta &= \alpha_1 (Y_{j-1} - X_{j-1}\beta) + \alpha_2 (Y_{j-2} - X_{j-2}\beta) + \dots + \alpha_n (Y_{j-n} - X_{j-n}\beta) + \varepsilon_j \\ &\Rightarrow \\ Y_j &= (X_j - \alpha_1 X_{j-1} - \alpha_2 X_{j-2} - \dots - \alpha_n X_{j-n})\beta + \alpha_1 Y_{j-1} + \alpha_2 Y_{j-2} + \dots \\ &\quad + \alpha_n Y_{j-n} + \varepsilon_j \\ &\Rightarrow \\ Y_j &= X_j\beta + \alpha_1 Y_{j-1} + \alpha_2 Y_{j-2} + \dots + \alpha_n Y_{j-n} + \varepsilon_j \end{aligned}$$

La última expresión es exactamente la representativa de un modelo de transición de orden n en su formulación clásica, con matriz de diseño  $X'$  conformada por las filas  $X_j'$  (donde  $X_j' = X_j - \alpha_1 X_{j-1} - \alpha_2 X_{j-2} - \dots - \alpha_n X_{j-n}$ ).

**Supuestos y método de ajuste**

El ajuste del modelo (1) se basa en el supuesto de mecanismo de pérdidas aleatorio o completamente aleatorio (Missing at Random, MAR; Missing Completely at Random, MCAR) y asume la validez de la llamada condición de separabilidad (4, 5). Los parámetros se estiman por el método de Máxima Verosimilitud (6). Se asume, además, una distribución normal multivariada del vector de respuestas.

Para la obtención del modelo final ajustado a los datos de la hemoglobina se parte de un modelo que contiene el mayor número de covariables posible. Mediante el método de selección *backward* -hacia atrás- (7) se obtiene un modelo simplificado o reducido.

**Herramientas utilizadas en la implementación**

Se utiliza la instrucción NLMIXED incluida en el paquete estadístico SAS para Windows, versión 9.1.3 (8). Gracias a las facilidades que esta instrucción brinda se obtiene, mediante un trabajo algebraico, la expresión explícita de la función, objetivo que automáticamente se maximiza. Se crean macros (funciones) auxiliares que posibilitan el proceso de inversión de matrices.

**Descripción de los datos del ensayo clínico**

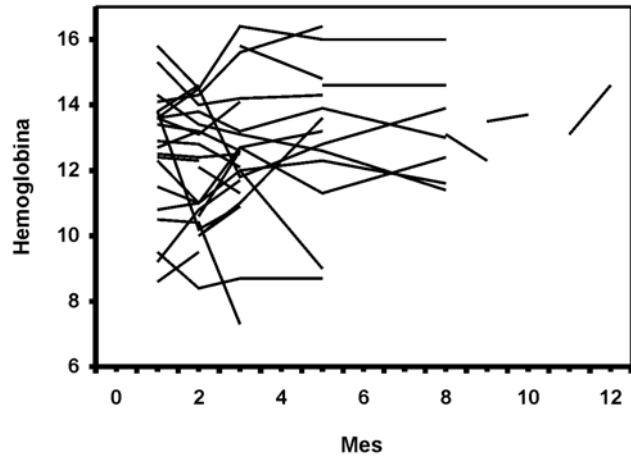
En la Tabla 1 se muestran los momentos de tomas de mediciones de la variable hemoglobina en el ensayo clínico mencionado en la sección introductoria. Los momentos mostrados representan ocasiones en que se realizó toma de mediciones para al menos un paciente. El total de pacientes estudiados fue 34. En la Tabla 2 se representan los patrones de pérdidas, que obviamente están lejos de ser monótonos. Por ejemplo, la secuencia "O N N O N N N N" significa

que se dispone del valor de la hemoglobina en los meses 1 y 5 solamente. Note que todos los pacientes presentan al menos una evaluación incompleta.

Este es un modelo de transición de cuarto orden que expresa el valor actual de la hemoglobina en función de valores previos alcanzados: uno, dos, tres y cuatro meses antes.

**Tabla 1.** Esquema de toma de muestra de hemoglobina.

Día de extracción	Mes
28	1
56	2
84	3
140	5
224	8
252	9
280	10
308	11
336	12



**Figura 1.** Perfiles observados de hemoglobina.

**Resultados**

**Exploración y selección del modelo**

Se procede seguidamente a ajustar un modelo de transición a la variable hemoglobina, considerando como covariables de interés el momento de toma de medición (*mes*) y el nivel de dosis de la vacunación (*nivdosis*), que adquiere uno de los siguientes valores: 150 µg, 300 µg, 600 µg, 900 µg, 1200 µg, 1500 µg.

Los seis valores de nivel de dosis se sustituyen por valores proporcionales mediante división por 150, de modo que las nuevas unidades son: 1, 2, 4, 6, 8 y 10. El modelo inicial que se ajusta es el siguiente:

$$Y_j = \beta_1 + \beta_2 \text{nivdosis} + \beta_3 \text{mes} + \delta_j$$

$$\delta_j = \alpha_1 \delta_{j-1} + \alpha_2 \delta_{j-2} + \alpha_3 \delta_{j-3} + \alpha_4 \delta_{j-4} + \epsilon_j \quad (II)$$

La elección del orden cuarto es empírica y se basa en que la inclusión de un quinto orden no es significativa. Como se verá, el parámetro recurrente que caracteriza el orden cuarto,  $\alpha_4$ , tampoco lo es.

En la Figura 1 se muestran los perfiles observados de la variable hemoglobina del ensayo, en gramos por decilitro, donde nuevamente se aprecia la elevada incidencia de datos faltantes. La aplicación del método de selección *backward* al modelo (II) produce como resultado el siguiente modelo reducido (de orden 3):

$$Y_j = \beta_1 + \beta_2 \text{nivdosis} + \delta_j$$

$$\delta_j = \alpha_1 \delta_{j-1} + \alpha_3 \delta_{j-3} + \epsilon_j \quad (III)$$

**Tabla 2.** Patrones de pérdidas de hemoglobina.

Patrones de pérdidas	Frecuencia	%	Frec. Acum	% Acum
ff				
N N N N N N N N N N	3	8,82	3	8,82
N O N N N N N N N N	1	2,94	4	11,76
N O O N N N N N N N	3	8,82	7	20,59
N O O N N N N N N N	1	2,94	8	23,53
O N N N N N N N N N	1	2,94	9	26,47
O N N O N N N N N N	1	2,94	10	29,41
O N O O N N O N N N	1	2,94	11	32,35
O N O O N O N N N N	1	2,94	12	35,29
O N O O O N O N N N	1	2,94	13	38,24
O O N O N N N N N N	2	5,88	15	44,12
O O N O N O O N N N	1	2,94	16	47,06
O O N O O N O N O N	1	2,94	17	50,00
O O O N N N N N N N	5	14,71	22	64,71
O O O N O N N N N N	1	2,94	23	67,65
O O O N O O N N O N	1	2,94	24	70,59
O O O O N N N N N N	5	14,71	29	85,29
O O O O O N O N N N	2	5,88	31	91,18
O O O O O N O N O N	3	8,82	34	100,00

O = Observado, N = No observado

## Implementación del modelo reducido

A continuación se expone el código que permite efectuar el ajuste del modelo reducido (III) a los datos del ensayo clínico. Basta ejecutarlo para obtener la salida reportada en la Tabla 3.

```
data labclinnlm;input incl niveldosis y1 y2 y3 y4 y5 y6 y7 y8 y9 y10 y11 y12 @@;
```

```
datalines;
```

```
1 150 8.6 9.5 . . 9.3 . . . . . 2 150 . . . . . . . . . . 3 150 . 10.6 12.6 . . (Se expone el código de creación de la base de datos íntegramente en anexo);
```

```
data labclinnlm;set labclinnlm;nivdosis=niveldosis/150;run;
```

```
/*Macros*/
```

```
%macro incionlm(n,ntime);n=&n;ntime=&ntime;array alfa(&n);
array xmat(&n,&n);array ymat(&n,&n);array ixmat(&n,&n);array iymat(&n,&n);
array alfamat(&n,&n);array y(&ntime);array rho(&ntime);array indexobs(&ntime);
array c(&ntime,&ntime);array q(&ntime,&ntime);array v(&ntime,&ntime);
array vobs(&ntime,&ntime);do i=1 to ntime;rho[i]=0;end;
do i=1 to ntime;do j=1 to ntime;v[i,j]=0;vobs[i,j]=0;c[i,j]=0;q[i,j]=0;end;end;
do i=1 to n;do j=1 to n;xmat[i,j]=0;yamat[i,j]=0;ixmat[i,j]=0;iymat[i,j]=0;end;end;%mend;
%macro cholesky(n,matrix,cholesky);do i=1 to &n;do j=1 to i;sumaij=0;sumajj=0;
if j eq 1 then &cholesky[i,j]=&matrix[i,j]/&matrix[j,j]**0.5;else do;do k=1 to j-1;
sumaij=sumaij+&cholesky[i,k]*&cholesky[j,k];sumajj=sumajj+&cholesky[j,k]**2;end;
&cholesky[i,j]=(&matrix[i,j]-sumaij)/(&matrix[j,j]-sumajj)**0.5;end;end;end;%mend;
%macro ccholesky(n,matrix,cholesky);do i=1 to &n;do j=1 to i;sumaij=0;sumajj=0;
if j eq 1 then &cholesky[i,j]=&matrix[i,j]/&matrix[j,j]**0.5;else do;do k=1 to j-1;
sumaij=sumaij+&cholesky[i,k]*&cholesky[j,k];sumajj=sumajj+&cholesky[j,k]**2;end;
if (i eq j) and (j gt 2) then &matrix[i,j]=tau2+sumajj;
else if i eq j+1 then &matrix[i,j]=alfa1*&matrix[i-1,j]+alfa2*&matrix[j,i-2];
else if i ge j+2 then &matrix[i,j]=alfa1*&matrix[i-1,j]+alfa2*&matrix[i-2,j];
&cholesky[i,j]=(&matrix[i,j]-sumaij)/(&matrix[j,j]-sumajj)**0.5;end;end;end;%mend;
%macro descomposicion_triangular(n,matrix,xmat,yamat);
do j=1 to &n;do i=1 to &n;if i=j then &yamat[i,j]=1;
if i<j then do;suma=0;do k=1 to i-1;suma=suma+&xmat[i,k]*&yamat[j,k];end;
&yamat[j,i]=(&matrix[i,j]-suma)/&xmat[i,i];end;
if i>=j then do;suma=0;do k=1 to j-1;suma=suma+&xmat[i,k]*&yamat[j,k];end;
&xmat[i,j]=&matrix[i,j]-suma;end;end;end;%mend;
%macro triangular_inversa(n,matrix,inversa);
do i=1 to &n;do j=1 to i;if i eq j then &inversa[i,j]=1/&matrix[i,j];
else do;suma=0;do k=j to i-1;suma=suma+&matrix[i,k]*&inversa[k,j];
end;&inversa[i,j]=-suma/&matrix[i,i];end;end;end;%mend;
/*Ajuste */
```

```
proc nlmixed data=labclinnlm cov maxiter=10000 maxfunc=100000;
```

```
parms beta1=11 beta2=0 alfa1=0.5 alfa3=0.3 lambda2=3.9;alfa2=0;%incionlm(3,12);
```

```
do i=1 to n;do j=1 to n;alfamat[i,j]=(i=j);if i+j<=n then alfamat[i,j]=alfamat[i,j]-alfa[i+j];if i>j then alfamat[i,j]=alfamat[i,j]-alfa[i-j];end;end;
```

```
%descomposicion_triangular(n,alfamat,xmat,yamat);%triangular_inversa(n,xmat,ixmat);
```

```
%triangular_inversa(n,yamat,iymat);do i=1 to n;do j=1 to n;suma=0;do k=max(i,j) to n;
```

```
suma=suma+iymat[k,i]*ixmat[k,j];end;rho[i]=rho[i]+suma*alfa[j];end;end;
```

```
do i=n+1 to ntime;do j=1 to n;rho[i]=rho[i]+alfa[j]*rho[i-j];end;end;
```

```
alfarho=0;do i=1 to n;alfarho=alfarho+alfa[i]*rho[i];end;tau2=lambda2*(1-alfarho);
```

```
do i=1 to ntime;do j=1 to i;if i=j then v[i,j]=lambda2;else v[i,j]=lambda2*rho[i-j];end;end;dimyobs=0;i=0;do j=1 to ntime;if y[j] ne . then do;i=i+1;indexobs[i]=j;
```

```
dimyobs=i;end;end;do i=1 to dimyobs;do j=1 to i;vobs[i,j]=v[indexobs[i],indexobs[j]];
```

```
end;end;%cholesky(dimyobs,vobs,c);%triangular_inversa(dimyobs,c,q);suma=0;sqrtdet=1;
```

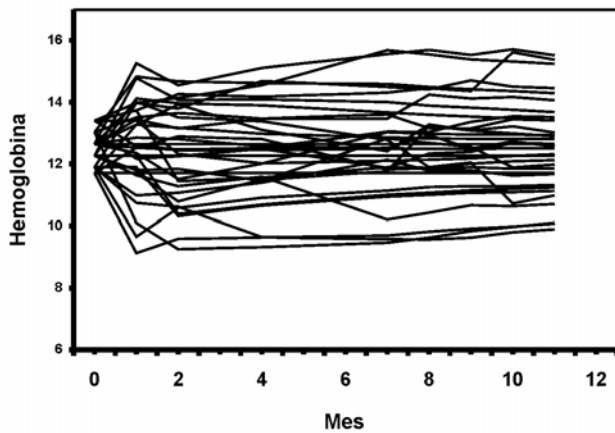
```
do j=1 to dimyobs;sqrtdet=sqrtdet*c[j,j];u=0;do k=1 to j;u=u+q[j,k]*(y[indexobs[k]]-beta1-beta2*nivdosis);end;suma=suma+u**2;end;loglike=-0.5*dimyobs*log(2*constant("pi"))-log(sqrtdet)-0.5*suma;model incl ~ general(loglike);estimate "cor_1" rho[1];run;
```

**Tabla 3.** Estimadores de los parámetros del modelo (III).

Efecto	Parámetro	Estimado	Error Estándar	p Valor
Intercepto	$\beta_1$	11,5157	0,6412	<0,0001
	$\beta_2$	0,1884	0,1050	0,0815
Parámetros Recurrentes	$\alpha_1$	0,5605	0,1021	<0,0001
	$\alpha_2$	0,3546	0,1121	0,0033
Var( $\delta_i$ )	$\lambda^2$	3,8306	0,8050	<0,0001

### Resultados del ajuste

Los resultados de las estimaciones de los parámetros del modelo (III) se presentan en la Tabla 3. Se aprecia una estrecha dependencia del valor actual de la hemoglobina en valores alcanzados un mes y tres meses antes. Esto puede ser reflejo de una regularidad trimestral en la evolución del paciente, debido quizás a la administración periódica de transfusiones de sangre para estabilizar la hemoglobina. La Figura 2 muestra los perfiles ajustados. Observe cómo el modelo es capaz de predecir valores donde antes había datos faltantes.



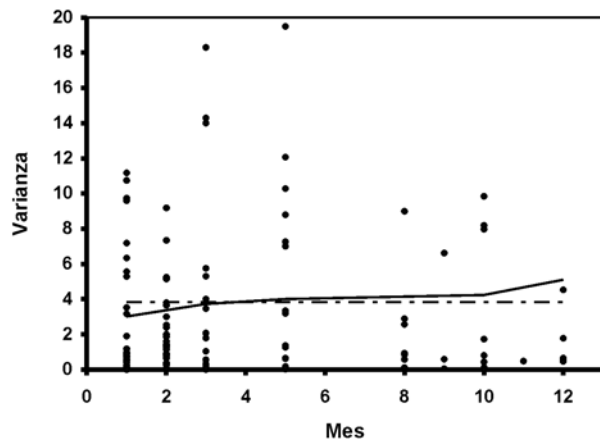
**Figura 2.** Perfiles ajustados de hemoglobina.

Se obtiene, además, un indicio de incremento de la hemoglobina con respecto al nivel de la dosis de vacunación, aunque no se alcanza significación estadística en este caso ( $p = 0,08$ ). La correlación mensual estimada por el modelo entre las variables que identifican a la hemoglobina durante las nueve tomas de mediciones es alta, de un 83%. Esto corrobora el resultado de dependencia del valor actual en el valor alcanzado al mes anterior. Un supuesto implícito en el modelo de transición (I) es la constancia de la varianza a través del tiempo. En este sentido, la Figura 3 grafica la función de la varianza estimada a partir de los datos observados – línea sólida– y la predicha por el modelo (III) –línea punteada–, confirmando la consistencia de los datos con este supuesto.

### Discusión

Se expusieron los resultados del ajuste de un modelo de transición a datos de tipo longitudinal en presencia de patrones de pérdidas intermitentes. Este es un escenario que usualmente no se halla contemplado en el marco del ajuste de modelos condicionales, debido probablemente a la no trivialidad de la implementación (2). Sin embargo, como se ha mostrado, es posible efectuar el ajuste y el código presentado puede emplearse con moderadas modificaciones a situaciones prácticas similares.

Morariu y colaboradores (9) reportan una aplicación de modelación autorregresiva a longitudes de secuencias de codificación del genoma bacteriano. Ding-Fei y cols (10) aplican la modelación autorregresiva a datos de electrocardiogramas. Otros campos de aplicación de los modelos de transición han sido mencionados en la sección introductoria. En general, en datos longitudinales se recomienda considerar siempre la opción de una modelación e interpretación condicional para obtener mejor entendimiento del mecanismo generador de los datos o predecir valores futuros de la respuesta. El supuesto de dependencia lineal permite simplificar y, ocasionalmente, hacer posible un análisis teórico (11). Se recomienda incorporar los resultados obtenidos dentro del arsenal de



**Figura 3.** Función de varianza observada (línea sólida) y ajustada (línea punteada).

técnicas estadísticas para el tratamiento de modelos condicionales en datos de tipo longitudinal y aplicar el código propuesto a situaciones reales, con el objetivo de acumular experiencia práctica y evaluar el desempeño de la herramienta aportada.

### Agradecimiento

El autor agradece al Departamento Estadísticas y Biometría del Instituto Finlay, en especial a la MSc Mayelín Mirabal Sosa, por su apoyo incondicional en la preparación del presente trabajo.

### Referencias

1. Michael Falk, Frank Marohn, René Michel, Daniel Hofmann, Maria Macke, editors. A First Course on Time Series Analysis. Examples with SAS. Würzburg, Germany: Chair of Statistics, University of Würzburg; 2006. Disponible en: [http://www.statistik-mathematik.uni-wuerzburg.de/fileadmin/10040800/user\\_upload/time\\_series/the\\_book/2006-February-01-times.pdf](http://www.statistik-mathematik.uni-wuerzburg.de/fileadmin/10040800/user_upload/time_series/the_book/2006-February-01-times.pdf) Consultado: 1ro de febrero de 2011.

2. Molenberghs G, Verbeke G. Models for discrete longitudinal data. New York: Springer-Verlag; 2005.
3. Diggle PJ, Liang KY, Zeger SL. Analysis of Longitudinal Data. Oxford Science Publications. Oxford: Clarendon Press; 1994.
4. Rubin DB. Inference and missing data. Biometrika 1976;63:581-92.
5. Little RJA, Rubin DB. Statistical Analysis with Missing Data. New York: John Wiley & Sons; 1987.
6. Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. New York: Springer-Verlag; 2000.
7. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W Applied Linear Statistical Models. 4th ed. Chicago: IRWIN; 1996.
8. SAS Institute Inc. SAS OnlineDoc® 9.1.3. Cary, NC: SAS Institute Inc; 2004.
9. Morariu VV, Buimaga-Iarinca L. Autoregressive modeling of coding sequence lengths in bacterial genome, 2009. Disponible en: <http://arxiv.org/ftp/arxiv/papers/0907/0907.1159.pdf> . Consultado: 15 de octubre de 2010.
10. Ding-Fei GE, Bei-Ping H, Xin-Jian X. Study of Feature Extraction Based on Autoregressive Modeling in ECG Automatic Diagnosis. Acta Automática Sinica 2007; 33(5): 462-6.
11. Feller W. An Introduction to Probability Theory and Its Applications. 3rd ed. New York: John Wiley; 1968.

### Anexo

#### Código íntegro para la creación de la base de datos de hemoglobina

```
data labclinnlm;input incl niveldosis y1 y2 y3 y4 y5 y6 y7 y8 y9 y10 y11 y12 @@;
```

```
datalines;
```

```
1 150 8.69.5 . 9.3 . . . . . 2 150 . . . . . 3 150 . 10.6 12.6 . . . . 13.8 . 13.1 14.6 4 150 12.7 13.2 . . 11.6 . . . . . 5 150 13.6 14.5
11.8 . 12.8 . . 13.9 . 12.6 . . 6 300 . 10 11 . . . . . 7 300 10.8 11 12 . 12.3 . . 11.6 . 11.6 . 11.3 8 300 10.5 10.4 7.3 . . . . . 9 300 9.2 10.8
11.7 . . . . . 10 300 14.4 . 15.8 . 14.8 . . 15 . . 11 600 11.7 . 11 . 13.6 . . . . 11.5 . . 12 600 11.5 11 12 . 9 . . . . . 13 600 . . . . .
14 600 15.3 14 14.2 . 14.3 . . . . . 15 600 . 10 10.9 . . . . . 16 600 14.5 . . . . . 17 900 12.3 11 12.7 . 13.2 . . . . . 18 900 13.4
13.2 12.6 . . . . .
19 900 . 12.4 . . . . . 20 900 13.6 13.1 14.1 . . . . 14.6 . . . . 21 900 12.9 12.8 12.1 . . . . . 22 900 . . . . . 23 900 15.8 14.5 16.4
. 16 . . 16 . . 24 900 12.5 12.4 12.5 . . . . 13.1 12.3 . . 12.5 25 900 14.3 13.4 13.1 . 12.6 . . 11.4 . 10 . 6.9 26 1200 9.5 8.4 8.7 . 8.7 . . . . .
27 1200 13.6 13.8 13.2 . 13.9 . . 13 . 13.8 . 14.4 28 1200 15.2 . . 14.9 . . . . . 29 1500 . 12.1 11.3 . . . . . 30 1200 12 . 12.7 . 11.3 . . 12.4
. 12.8 . . 31 1500 12.4 12.3 . . 13.1 . . . 13.5 13.7 . . 32 1500 13.8 14.6 . . 14.6 . . 14.6 . 16.6 . 14.6 33 1500 13.8 10.2 10.9 . . . . . 34 1500
14.1 14.3 15.6 . 16.4 . . . . .
```

```
data labclinnlm;set labclinnlm;nivdosis=niveldosis/150;run;
```

---

## Transition models with intermittent missing data: implementation and application to a Cuban clinical trial

### Abstract

The present work offers a useful tool for fitting transition models under the problematic scenario of intermittent missing data patterns in collected longitudinal data. A solution to this problem is given using the implicit easiness in the NL MIXED procedures from the statistical software SAS (Statistical Analysis System) for Windows, version 9.1.3. In addition, a practical application is provided, where a transition model matches the clinical laboratory data of a trial that evaluated, among other objectives, the safety of a skin cancer Cuban vaccine.

**Keywords:** Longitudinal data, transition model, missing data patterns, missing data mechanism, clinical assay.

---