

## Trabajos

---

### Uso de criterios de bondad de ajuste para seleccionar un modelo no lineal

---

RECIBIDO:  
ACEPTADO:

---

**Garcia, M.C. • Rapelli, C.**

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística. Facultad de Ciencias Económicas y Estadística. Universidad Nacional de Rosario. Bvard Oroño 1261. (2000) Rosario, Santa Fe, Argentina. Tel: 0341-4802794-int.152. Email: mgarcia@fcecon.unr.edu.ar

**RESUMEN:** Para caracterizar el proceso de crecimiento de las poblaciones se efectúan mediciones en forma repetida de la característica que es objeto del estudio y se utilizan modelos no lineales, de tipo sigmoideo, para representar la dinámica del crecimiento. Los modelos no lineales mixtos son adecuados para estudiar el crecimiento. La determinación del modelo óptimo desde el punto de vista estadístico no es una tarea sencilla, existiendo una amplia gama de criterios para la comparación de modelos. En este trabajo se utilizan algunos criterios propuestos para comprobar la bondad del ajuste de los modelos no lineales mixtos para datos longitudinales. Ellos son similares al coeficiente de determinación para regresión lineal y proveen una medida de la concordancia entre las respuestas observadas y ajustadas. Estos criterios se utilizan durante el proceso de construcción de un modelo para explicar la evolución del peso de ratones machos y hembras, mediante la función de Gompertz.

**PALABRAS CLAVE:** Datos longitudinales, curvas

de crecimiento, modelos no lineales mixtos, criterios de bondad de ajuste.

**SUMMARY:** *The use of goodness-of-fit criteria for the selection of non-linear models.*

Repeated measures on each experimental unit and non-linear models can be used to describe the growth process of a population. Non-linear mixed models constitute a powerful tool for analyzing the dynamics of growth. The selection of the appropriate model is not an easy task; model building techniques for non-linear mixed models include several goodness-of-fit criteria for comparison of different models. In this paper some of these criteria are used. They are similar to the linear regression determination coefficient and provide a concordance measure between fitted and observed responses. These criteria are used to describe and compare the weight growth in two groups of mice with different genotypes through Gompertz's function.

**KEYWORDS:** Longitudinal data, growth curve, non-linear mixed models, goodness-of-fit criteria.

## Introducción

En las investigaciones biológicas es de frecuente interés caracterizar el proceso de crecimiento de las poblaciones mediante un modelo estadístico que explique su evolución. En general, estos modelos son complejos y para la construcción del mismo se deben tener en cuenta dos aspectos. En primer lugar, la estructura de los datos bajo estudio, que consisten en medidas repetidas tomadas sobre cada una de las unidades (sujetos) en diferentes ocasiones. Estos datos longitudinales presentan algunas características que complican el análisis estadístico: las mediciones realizadas para el mismo sujeto están correlacionadas y, si bien las mediciones entre los sujetos son independientes, pueden presentar una gran variabilidad.

Por otro lado, la dinámica del crecimiento generalmente es no lineal. Los modelos de curvas de crecimiento sigmoideas permiten describir el proceso de crecimiento. El uso de modelos no lineales en los parámetros agrega complejidad al proceso de modelado, aunque aumenta la interpretación del modelo debido a la correspondencia directa entre sus parámetros y los conceptos biológicos.

Los modelos no lineales de efectos mixtos proveen una herramienta útil para analizar datos longitudinales en los cuales la relación entre las variables explicativas y la variable respuesta puede modelarse como una única función no lineal, permitiendo a los parámetros diferir entre los individuos.

Si bien los modelos no lineales mixtos para datos longitudinales se adecuan para el estudio de curvas de crecimiento, la determinación del modelo "óptimo" desde el punto vista estadístico no es una tarea sencilla, existiendo una amplia gama de criterios para la comparación de modelos (1, 2, 3).

En la etapa de construcción del modelo para verificar la bondad de ajuste del mismo, usualmente, se usan métodos basados en la verosimilitud. Estos métodos no permiten la evaluación directa del modelo que se está utilizando ya que para su aplicación se requieren ajustes repetidos de diferentes modelos anidados o no. En los últimos años se propusieron criterios (3, 4) similares al coeficiente de determinación para regresión lineal, que permiten verificar directamente la bondad del ajuste basándose sólo en el modelo que se está evaluando, lo cual presenta una ventaja sobre los anteriores.

En este trabajo se presenta el proceso de construcción de un modelo no lineal para explicar la evolución del peso de ratones y caracterizar el patrón de crecimiento de dos poblaciones (CBI y BALB). Durante la etapa de selección del modelo se utilizan algunos criterios propuestos para comprobar la bondad del ajuste, que proveen una medida de la concordancia entre las respuestas observadas y ajustadas.

## Materiales y métodos

En este trabajo se evaluaron dos genotipos de ratones CBI y BALB, los cuáles se mantienen desde hace más de 20 años en la Facultad de Ciencias Médicas de la Universidad Nacional de Rosario, inicialmente en el Centro de Biología y actualmente en el Instituto de Genética Experimental de dicha Facultad.

Ambos genotipos pertenecen a la especie *Mus musculus*, conocida vulgarmente como ratón de laboratorio. La línea BALB se utiliza extensivamente a nivel mundial en trabajos de investigación biomédica, mientras que la línea CBI, de origen local, fue generada en el antes citado Centro de Biología, de donde deriva su denominación, a partir del cruce

miento de las cepas BALB, Rockland, NIH y Suiza. En consecuencia, el acervo genético de esta población cuenta con genes del otro genotipo evaluado. Ambas líneas han sido endocriadas por limitación del número de individuos. La endocría es una estrategia de apareamiento de individuos emparentados que permite alcanzar más lentamente la uniformidad genética y sobrellevar los efectos adversos de la depresión por endogamia que se expresa sobre caracteres relacionados con la eficacia biológica tales como la fertilidad y la sobrevivencia y compromete el mantenimiento de las líneas.

Es de interés caracterizar cuantitativamente el patrón de crecimiento de estas poblaciones especialmente la cepa CBI, ya que en su conformación participó la BALB y es por eso que la primera cuenta con genes de esta última. La información sobre estos ratones (peso y longitud caudal) se encuentra registrada a través del tiempo.

En este trabajo se cuenta con 56 ratones, 28 del genotipo BALB y 28 del genotipo CBI, a los cuales se les midió el peso en 13 momentos distintos tomados en forma equiespaciada desde el nacimiento hasta los 144 días de edad.

En la Figura 1 se representa la evolución del peso de los ratones a través del tiempo.

*Métodos estadísticos*

El modelo no lineal mixto (3, 5, 6) para las observaciones del sujeto  $i, i=1, \dots, N$  se puede expresar como,

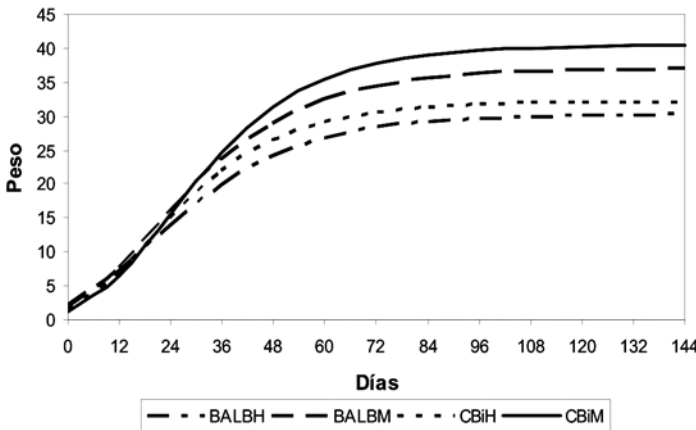
$$Y_i = f(X_i, \beta_i) + \epsilon_i, \tag{2.1}$$

donde,  $Y_i = [Y_{i1}, \dots, Y_{in}]'$  es el vector ( $n_i \times 1$ ) compuesto por las mediciones repetidas del  $i$ -ésimo individuo, e  $Y_{ij}$  la observación realizada al  $i$ -ésimo individuo en el tiempo  $t_j, j = 1, \dots, n_i,$

$f(X_i, \beta_i) = [f(x_{i1}, \beta_i), \dots, f(x_{in}, \beta_i)]'$  siendo,  $f$  una función no lineal conocida que relaciona el vector de respuestas con el tiempo y otras posibles covariables intra unidad ( $X_i$ ) y  $\beta_i$  es un vector específico del individuo que contiene los parámetros de la función no lineal.

El vector  $\beta_i$  se puede modelar, en una segunda etapa, como la suma de dos componentes, una fija o poblacional común a todos los sujetos y otra específica a cada sujeto.

$$\beta_i = A_i \beta + B_i b_i. \tag{2.2}$$



**Figura 1:** Peso de los ratones hasta los 144 días según genotipo y sexo.

Nota: BALBH= BALB hembras, BALBM= BALB machos, CBIH= CBI hembras y CBIH= CBI machos.

Los elementos del modelo no lineal mixto son, entonces,

$\mathbf{X}_i = \{\mathbf{x}_{ij}\}$ : Matriz ( $n_i \times v$ ) de diseño del  $i$ -ésimo individuo,  $j = 1, \dots, n_i$ ,

$\boldsymbol{\beta}_i$ : Vector ( $r \times 1$ ) de parámetros del sujeto  $i$ -ésimo,

$\boldsymbol{\beta}$ : Vector ( $s \times 1$ ) de efectos fijos,

$\mathbf{b}_i$ : Vector ( $q \times 1$ ) de efectos aleatorios,

$\mathbf{A}_i$ : Matriz ( $r \times s$ ) de diseño para los efectos fijos,

$\mathbf{B}_i$ : Matriz ( $r \times q$ ) de diseño para los efectos aleatorios.

Se supone que  $\mathbf{b}_i$  y  $\boldsymbol{\varepsilon}_i$  son independientes con distribución,

$$\mathbf{b}_i \stackrel{\text{i.i.d.}}{\sim} N_q(\mathbf{0}, \boldsymbol{\Psi}) \quad \boldsymbol{\varepsilon}_i \stackrel{\text{i.i.d.}}{\sim} N_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_i),$$

siendo,  $\boldsymbol{\Psi}$  la matriz de covariancias de los efectos aleatorios y  $\boldsymbol{\Sigma}_i$ , con la misma estructura para todos los individuos, la matriz de covariancias intra-individuos (3, 7).

La función no lineal que se elige para modelar el peso de los ratones (7, 8, 9) es la de Gompertz cuya expresión es la siguiente,

$$f(\mathbf{x}_{ij}, \boldsymbol{\beta}_i) = \beta_{0i} \exp(-\beta_{1i} \exp(-\beta_{2i} t_j)).$$

Los parámetros de la misma,  $\boldsymbol{\beta} = (\beta_{0i} \ \beta_{1i} \ \beta_{2i})'$  representan la asíntota de la curva ( $\beta_{0i}$ ), el inicio del crecimiento poblacional de la curva ( $\beta_{1i}$ ) y la tasa de crecimiento poblacional ( $\beta_{2i}$ ).

Uno de los objetivos del estudio es probar los efectos que tienen las covariables sexo y genotipo sobre el crecimiento. Se utilizan dos variables ficticias para identificar los niveles de los factores,

$\mathbf{S}_i$ : una variable indicadora de sexo, que vale 1 para los machos y 0 para las hembras y

$\mathbf{G}_i$ : una variable indicadora de genotipo, que vale 0 para los ratones de genotipo BALB y 1 para los ratones de genotipo CBI.

Si  $Y_{ij}$  es el peso del  $i$ -ésimo ratón en la  $j$ -ésima ocasión de medida (edad),  $i = 1, 2, \dots, 56$  y  $j = 1, 2, \dots, 13$ , y considerando que los tres parámetros del modelo dependen linealmente de las covariables, el modelo tiene la expresión,

$$Y_{ij} = \beta_{0i} \exp(-\beta_{1i} \exp(-\beta_{2i} t_j)) + \mathbf{e}_{ij}, \quad (2.3)$$

$$\beta_{0i} = \beta_0 + \beta_{01} G_i + \beta_{02} S_i + \beta_{03} G_i S_i + b_{0i}$$

$$\beta_{1i} = \beta_1 + \beta_{11} G_i + \beta_{12} S_i + \beta_{13} G_i S_i + b_{1i}$$

$$\beta_{2i} = \beta_2 + \beta_{21} G_i + \beta_{22} S_i + \beta_{23} G_i S_i + b_{2i}$$

Se supone que:

$$\mathbf{b}_i = \begin{pmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \boldsymbol{\Psi}); \quad \mathbf{e}_i = \begin{pmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{in_i} \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_i).$$

El modelo (2.3) supone que los errores son independientes y con variancia homogénea ( $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}$ ) y considera efectos aleatorios para los tres parámetros del modelo.

Los parámetros del modelo se estiman por máxima verosimilitud o máxima verosimilitud restringida (3, 5, 6), utilizando aproximaciones numéricas. Las soluciones se obtienen por métodos iterativos y la introducción de muchos efectos aleatorios (modelo sobreparametrizado) puede crear problemas de convergencia (10). Una propuesta para superar esta instancia es considerar efectos aleatorios en algunos de los parámetros.

El método de máxima verosimilitud restringida se propuso como una forma de corregir el sesgo en los parámetros de cova-

riancia que produce el procedimiento de máxima verosimilitud cuando se cuenta con muestras pequeñas.

#### *Proceso de modelado*

Debido a que una mala especificación del modelo puede tener un impacto considerable sobre las propiedades asintóticas de los estimadores y, por consiguiente, en las inferencias que se realizan sobre ellos, una etapa importante de la modelación es la determinación de los efectos aleatorios y fijos, como así también de las estructuras de covariancias entre e intra unidad.

El procedimiento, denominado "construcción de un modelo", se basa en la comparación de modelos con diferentes estructuras para la componente aleatoria y permite determinar cuáles de los parámetros del modelo contendrán una componente aleatoria y cuáles serán puramente fijos. También permite modelar la dependencia de los parámetros a covariables específicas de la unidad. Este proceso se realiza en varios pasos (10).

El primer paso a realizar consiste en la determinación de los parámetros del modelo que contendrán efectos aleatorios (primera fase) y posteriormente la elección del tipo de estructura para la matriz de covariancias de los efectos aleatorios ( $\Psi$ ) (segunda fase).

Para evaluar qué parámetros serían aleatorios y cuáles fijos se comienza postulando un modelo con todos sus efectos mixtos. Luego se plantean modelos anidados en el anterior obtenidos al eliminar los efectos aleatorios de algunos de los parámetros.

La bondad de ajuste se puede verificar usando la prueba del cociente de verosimilitud o los criterios de información de Akaike (AIC) o Bayesiano de Schwarz (BIC) (1, 2, 3). El uso de AIC o BIC consiste en el

cálculo de los mismos para varios modelos de interés y en la elección de aquel modelo que tenga el menor valor de este criterio. Para aplicar la prueba del cociente de verosimilitud, en cambio, se postulan dos modelos anidados, uno bajo la hipótesis nula y otro bajo la alternativa, y se calculan las respectivas log-verosimilitud ( $\ell_0$  y  $\ell_a$ ). La diferencia de ambas ( $U = 2\ell_0 - 2\ell_a$ ) se compara con una distribución de probabilidad adecuada para rechazar o no la hipótesis planteada.

Una vez elegido el tipo de efecto, se postulan diferentes estructuras para la matriz de covariancias de los mismos (3, 5), utilizando los criterios de información para la elección entre modelos alternativos.

De esta forma se obtiene un modelo que incluye términos con efectos aleatorios significativos y la matriz de covariancias seleccionada, mientras que, durante este paso del proceso, la covariancia intra sujeto se mantiene fija e igual a una de independencia.

La segunda fase corresponde a la elección de la matriz de covariancia de los errores (covariancia intra unidad). Para ello se plantean diferentes estructuras utilizando, nuevamente, los criterios de información para la elección de la más adecuada.

Debido a que los datos longitudinales presentan, generalmente, además de correlación, variancias heterogéneas se debe evaluar, mediante el uso de gráficos, la presencia de heterocedasticidad. La mayoría de los enfoques gráficos se basan en el uso de residuos transformados pues no están correlacionados y tienen variancia unitaria (11). Dado un conjunto de residuos transformados se pueden aplicar los diagnósticos de residuos usuales para regresión lineal estándar. En un modelo correctamente especificado este gráfico no

mostrará ningún patrón sistemático, con residuos fluctuando aleatoriamente alrededor del cero y rango constante cuando los valores predichos varían.

Durante este proceso, en las dos fases del mismo, el modelo especificado incluye todas las covariables de interés. Como el objetivo que se pretende con el uso de los modelos no sólo es caracterizar el cambio de la respuesta en el tiempo sino también los factores que influyen el cambio, se necesitan realizar pruebas de hipótesis. La significación de los parámetros se puede comprobar mediante la prueba del cociente de verosimilitud (3, 5).

#### *Criterios de bondad de ajuste*

En la etapa de construcción del modelo se usan métodos basados en la verosimilitud, los criterios de AIC y BIC o la prueba del cociente de verosimilitud, para comprobar la bondad del ajuste. Aunque son muy utilizados en la práctica tienen la desventaja de no permitir la evaluación directa del modelo que se está utilizando ya que para su aplicación se requieren ajustes repetidos de modelos.

En los últimos años, varios autores (3, 4) propusieron estadísticas de bondad de ajuste para la evaluación de la estructura media. La ventaja de estos métodos es que permiten la aplicación directa al modelo que se está utilizando, agilizando así la construcción del mismo.

Las dos medidas que se presentan son:

1.  $-R^2$ , es una modificación del coeficiente de determinación de los modelos lineales. Para calcularlo se debe plantear primero un "modelo nulo" que permita interpretar a  $R^2$  como explicando la variación residual.

Los modelos nulo y de interés se escriben como,

Modelo de interés:  $\mathbf{Y}_i = \mathbf{f}_i(\boldsymbol{\beta}_i) + \mathbf{e}_i$

$$E(\mathbf{e}_i) = \mathbf{0}$$

$$\text{Var}(\mathbf{e}_i) = \boldsymbol{\Sigma}_i,$$

Modelo nulo:  $\mathbf{Y}_i = \mathbf{f}_{i0}(\boldsymbol{\beta}_{i0}) + \mathbf{e}_{i0}$

$$E(\mathbf{e}_{i0}) = \mathbf{0}$$

$$\text{Var}(\mathbf{e}_{i0}) = \boldsymbol{\Sigma}_{i0}.$$

Además, siendo  $\hat{\mathbf{Y}}_i$  y  $d_i^2(\mathbf{V}_i) = (\mathbf{Y}_i - \hat{\mathbf{Y}}_i)' \mathbf{V}_i^{-1} (\mathbf{Y}_i - \hat{\mathbf{Y}}_i)$  e  $\hat{\mathbf{Y}}_{i0}$  y  $d_{i0}^2(\mathbf{V}_i) = (\mathbf{Y}_i - \hat{\mathbf{Y}}_{i0})' \mathbf{V}_i^{-1} (\mathbf{Y}_i - \hat{\mathbf{Y}}_{i0})$  los valores predichos y las distancias de Mahalanobis bajo los modelos a evaluar y de comparación (modelo nulo), respectivamente, la variación residual explicada de  $\hat{\mathbf{Y}}_i$  (es decir, para cada unidad) relativa al modelo nulo y a  $\mathbf{V}_i$  se define como,

$$R_i^2(\mathbf{V}_i) = 1 - d_i^2(\mathbf{V}_i) / d_{i0}^2(\mathbf{V}_i),$$

siendo,  $\mathbf{V}_i$  una matriz definida positiva cualquiera.

Por lo que, la variación explicada total (con respecto al modelo nulo y a  $\mathbf{V}_i$ ) se define como el promedio ponderado:

$$R^2(\mathbf{V}) = 1 - \frac{\sum_{i=1}^N d_i^2(\mathbf{V}_i)}{\sum_{i=1}^N d_{i0}^2(\mathbf{V}_i)} = \frac{\sum_{i=1}^N w_i R_i^2(\mathbf{V}_i)}{\sum_{i=1}^N w_i},$$

donde,  $w_i = d_{i0}^2(\mathbf{V}_i)$  y  $\mathbf{V} = [\mathbf{V}_1, \dots, \mathbf{V}_N]$ .

Para implementar esta definición de  $R^2$  se necesita especificar un modelo nulo apropiado así como también los valores de  $\mathbf{V}_i$ . Generalmente se elige el modelo nulo de forma que refleje la estructura más simple posible, pero consistente con la aplicación. Un candidato posible para utilizar como modelo nulo es el modelo con ordenada al origen aleatoria. Éste es un modelo con efectos aleatorios simple que tiene en cuenta la variabilidad intra y entre indivi-

duos. Para  $\mathbf{V}_i$  se puede optar elegir la matriz de covariancias del modelo nulo ( $\Sigma_{i0}$ ), pues sería consistente con el deseo de tener una medida de bondad de ajuste que pueda ser usada como comparación en los distintos modelos a evaluar.

**2. Coeficiente de concordancia ( $r_c$ ).**

La expresión de este coeficiente, que puede ser usado como una medida alternativa a la anterior, es,

$$r_c = \frac{\sum_{i=1}^N (\mathbf{Y}_i - \hat{\mathbf{Y}}_i)^T (\mathbf{Y}_i - \hat{\mathbf{Y}}_i)}{\sum_{i=1}^N (\mathbf{Y}_i - \bar{\mathbf{Y}} \mathbf{J}_{n_i})^T (\mathbf{Y}_i - \bar{\mathbf{Y}} \mathbf{J}_{n_i}) + \sum_{i=1}^N (\hat{\mathbf{Y}}_i - \bar{\mathbf{Y}} \mathbf{J}_{n_i})^T (\hat{\mathbf{Y}}_i - \bar{\mathbf{Y}} \mathbf{J}_{n_i}) + M (\bar{\mathbf{Y}} - \bar{\bar{\mathbf{Y}}})^2}$$

donde,  $\mathbf{J}_{n_i}$  es un vector de unos,

$$\bar{\mathbf{y}} = \sum_{i=1}^N \sum_{j=1}^{n_i} Y_{ij} / M, \quad \bar{\hat{\mathbf{y}}} = \sum_{i=1}^N \sum_{j=1}^{n_i} \hat{Y}_{ij} / M \quad \text{y} \quad M = \sum_{i=1}^N n_i$$

es el total de observaciones.

El uso de este coeficiente es más ventajoso que el del presentado anteriormente y suministra una medida de la concordancia entre los valores observados y predichos. Así  $r_c$  no requiere la especificación de un modelo nulo y se interpreta como,

- Ajuste perfecto, valores cercanos a 1.
- Falta de ajuste, valores menores o iguales a 0.

Debido a que ambos coeficientes incrementan a medida que se sobreparametizan los modelos sus valores se pueden ajustar por el número de parámetros de la siguiente manera,

$$R_a^2 = 1 - k(1 - R^2)$$

$$r_{c,a} = 1 - k(1 - r_c^2)$$

siendo,  $k = M / (M - s)$  y  $s$  el número de parámetros de efectos fijos.

La principal ventaja de usar estas medidas como bondad de ajuste es que no se requiere la especificación de la función de verosimilitud y resultan muy útiles para datos continuos.

**Resultados**

Los datos presentados en el gráfico 2.1 describen un patrón de crecimiento no lineal de tipo sigmoideo, por lo que se utiliza la función de Gompertz para modelar la evolución del peso de los ratones.

Los ratones pertenecientes a cada genotipo se dividen según el sexo resultando así cuatro grupos (CBI machos, CBI hembras, BALB machos, BALB hembras).

Cada grupo está constituido por 14 ratones y el peso de cada uno de ellos se registra en 13 momentos distintos tomados en forma equiespaciada.

El proceso de construcción del modelo comienza con el ajuste del modelo (2.3), suponiendo errores intra-individuo independientes con variancia homogénea:  $\Sigma_i = \sigma^2 \mathbf{I}$ . Pero a diferencia del modelo (2.3), para evitar los posibles problemas de convergencia mencionados anteriormente, se consideran efectos aleatorios en uno de los tres parámetros del modelo.

El ajuste de las curvas y el cálculo de las estadísticas de bondad de ajuste se realiza con la macro NLINMIX y el procedimiento IML de SAS (12).

*Especificación de los efectos aleatorios*

Se plantean diferentes modelos cada uno de los cuales posee un efecto aleatorio. Con el fin de determinar los parámetros que tienen asociada una componente aleatoria, se puede utilizar el test de razón de verosimilitud o algún criterio de información. Los valores resultantes del ajuste de los modelos planteados se presentan en la tabla 1.

Los valores de los criterios de información obtenidos indican que el modelo 1, con efecto aleatorio para el parámetro que representa el límite de crecimiento, presenta un ajuste adecuado (valor más pequeño de AIC o BIC).

**Tabla 1:** Criterio de Akaike y de Schwarz para la comparación de modelos con diferentes efectos aleatorios.

Modelo	Efectos aleatorios			AIC	BIC
	$b_0$	$b_1$	$b_2$		
Modelo 1	*			2944,8	2973,2
Modelo 2		*		3177,3	3205,6
Modelo 3			*	3129	3157,4
Modelo 4				3176,7	3202,5

El \* indica los efectos aleatorios presentes en el modelo

#### *Especificación de la estructura de variancias y covariancias de los errores intra individuos*

Mediante la Figura 2 de los residuos transformados del modelo 1 se encuentra que, aún con los efectos aleatorios, las variancias del modelo son heterogéneas.

Para solucionar este problema se vuelve a estimar el modelo 1, utilizando para la matriz de covariancias intra individuo una estructura autorregresiva de orden 1 con variancias homogéneas y otra con variancias heterogéneas. Resultan, entonces, los modelos 5 y 6 que modifican sólo la matriz de covariancias intra unidad del modelo 1.

La tabla 2 presenta los criterios que permiten evaluar diferentes estructuras de covariancias.

Si bien ambas estructuras mejoran el desempeño del modelo 1, el modelo 6, con estructura autorregresiva heterogénea de orden 1, arroja los menores valores de AIC y BIC, presentando un buen ajuste. Los gráficos de residuos para este modelo no revelan ningún patrón sistemático.

El análisis continúa con la comparación de los parámetros de los grupos.

#### *Comparación entre grupos*

Durante los dos pasos anteriores los fac-

tores genotipo, sexo y su interacción fueron incluidos en todos los parámetros como efectos fijos. Debido a que uno de los objetivos es determinar los factores que influyen en el cambio del peso de los ratones en el tiempo se prueba la significación de los efectos mencionados, mediante el test de la razón de verosimilitud.

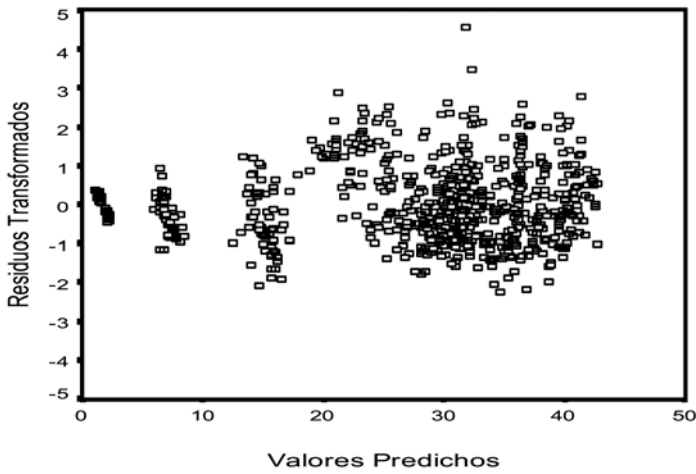
La tabla 3 presenta el modelo completo (modelo 6) y varios modelos reducidos que surgen de llevar las hipótesis planteadas al modelo. Los modelos resultantes son casos particulares del modelo general, con la misma estructura para los efectos fijos, que permiten comprobar si la reducción de los parámetros causa algún cambio significativo en el modelo.

El modelo 7 permite probar si el crecimiento límite es el mismo para los cuatro grupos. Como esa hipótesis se rechaza ( $p < 0,0001$ ) se plantean los modelos 7.1 y 7.2, concluyendo que el crecimiento límite no es el mismo para los dos genotipos ( $p < 0,0001$ ) ni para los dos sexos ( $p < 0,0001$ ).

Para comprobar si el parámetro que indica el inicio del crecimiento es el mismo para los cuatro grupos se postula el modelo 8. El rechazo de esa hipótesis ( $p < 0,0001$ ) conduce al planteo de los modelos 8.1 y 8.2 resultando que el parámetro que indica el inicio



**Figura 2:** Residuos transformados versus valores predichos para el Modelo 1.



**Tabla 2:** Criterios de Akaike y de Schwarz para la comparación de modelos con diferentes estructuras de covariancias para los errores intra individuo.

Modelo	$\Sigma_i$	AIC	BIC
Modelo 1	INDEP	2944,8	2973,2
Modelo 5	AR(1)	2779,8	2810,2
Modelo 6	ARH(1)	2316,5	2371,2

**Tabla 3:** Test de la razón de verosimilitud, criterios de Akaike y Schwarz y estadísticas de bondad de ajuste para la comparación entre modelos con diferentes efectos fijos.

Modelo	Hipótesis planteada	$-2\ell$	AIC	BIC	$R_a^2$	$r_{c,a}$	U	p
Modelo 6		2262,5	2316,5	2371,2	0,9727	0,9889		
Modelo 7	$\beta_{01} = \beta_{02} = \beta_{03} = 0$	2370,5	2418,5	2467,1	0,9736	0,9933	108	<0,0001
Modelo 7.1	$\beta_{01} = \beta_{03} = 0$	2295,4	2345,4	2396	0,9736	0,9912	32,9	<0,0001
Modelo 7.2	$\beta_{02} = \beta_{03} = 0$	2363,8	2413,8	2464,4	0,9735	0,9931	101,3	<0,0001
Modelo 8	$\beta_{11} = \beta_{12} = \beta_{13} = 0$	2300,1	2348,1	2396,7	0,9114	0,9788	37,6	<0,0001
Modelo 8.1	$\beta_{11} = \beta_{13} = 0$	2267	2317	2367,6	0,9724	0,9883	4,5	0,1054
Modelo 8.2	$\beta_{12} = \beta_{13} = 0$	2299,6	2349,6	2400,2	0,9089	0,9783	37,1	<0,0001
Modelo 9	$\beta_{21} = \beta_{22} = \beta_{23} = 0$	2274,1	2322,1	2370,7	0,9707	0,9872	11,6	0,0089
Modelo 9.1	$\beta_{21} = \beta_{23} = 0$	2265,6	2315,6	2366,2	0,9722	0,9881	3,1	0,2122
Modelo 9.2	$\beta_{22} = \beta_{23} = 0$	2274,6	2324,6	2375,2	0,9346	0,9840	12,1	0,0024
Modelo 10	$\beta_{11} = \beta_{13} = \beta_{21} = \beta_{23} = 0$	2269,2	2315,2	2361,8	0,9719	0,9874	6,7	0,1526

del crecimiento (parámetro de escala) no presenta diferencias significativas para los dos genotipos ( $p = 0,1054$ ), pero si para las dos sexos ( $p < 0,0001$ ).

El modelo 9 permite comprobar que la tasa de crecimiento no es la misma para los cuatro grupos ( $p = 0,0089$ ). Se postulan los modelos 9.1 y 9.2, concluyéndose que la tasa de crecimiento no presenta diferencias significativas entre los dos genotipos ( $p = 0,2122$ ) pero si entre los dos sexos ( $p = 0,0024$ ).

Las conclusiones anteriores, que se obtienen utilizando el test de la razón de verosimilitud, coinciden si se consideran los criterios de información. Se observa que los mismos presentan el menor valor para el modelo 10, que resulta el modelo elegido.

La tabla 3 permite ilustrar, además, el uso de las estadísticas de bondad de ajuste presentadas. Existe una disminución de los valores de los coeficientes  $R_a^2$  y  $r_{c,a}$  cuando se elimina la variable sexo y su interacción con genotipo (modelos 8.2 y 9.2), por lo que se decide incluir esta variable en el modelo.

Los coeficientes  $R_a^2$  y  $r_{c,a}$  de los modelos sin la variable genotipo (modelo 8.1 y 9.1) no difieren de los obtenidos para el modelo 6. En ambos modelos esa variable resulta ser no significativa.

Del proceso de construcción del modelo resulta el siguiente modelo estimado,

$$\hat{Y}_{ij} = \hat{\beta}_{0i} \exp(-\hat{\beta}_{1i} \exp(-\hat{\beta}_{2i} t_j)), \text{ donde,}$$

$$\hat{\beta}_{0i} = \hat{\beta}_0 + \hat{\beta}_{01} G_i + \hat{\beta}_{02} S_i + \hat{\beta}_{03} G_i S_i + \hat{b}_{0i} = 29,5429 + 2,5008 G_i + 7,1733 S_i + 0,9385 G_i S_i + \hat{b}_{0i}$$

$$\hat{\beta}_{1i} = \hat{\beta}_1 + \hat{\beta}_{12} S_i = 2,9780 + 0,2001 S_i$$

$$\hat{\beta}_{2i} = \hat{\beta}_2 + \hat{\beta}_{22} S_i = 0,0529 - 0,0033 S_i$$

La estimación de  $b_{0i}$  provee el valor del crecimiento máximo para cada uno de los ra-

tones de cada grupo. La variancia estimada de la componente aleatoria es  $\Psi = 1,8084$ , y las estimaciones de los parámetros de la matriz de covariancias intra individuo,

$$\hat{\Sigma}_i = \{\hat{\sigma}_j \hat{\sigma}_{j'} \hat{\rho}^{j-j'}\}, j, j' = 1, \dots, 13, \text{ resultan,}$$

Parámetro	Estimación
$\sigma_1^2$	0,0192
$\sigma_2^2$	0,4989
$\sigma_3^2$	4,1649
$\sigma_4^2$	26,3151
$\sigma_5^2$	3,9649
$\sigma_6^2$	1,2816
$\sigma_7^2$	1,8353
$\sigma_8^2$	1,8469
$\sigma_9^2$	1,6771
$\sigma_{10}^2$	2,0553
$\sigma_{11}^2$	2,3694
$\sigma_{21}^2$	4,4148
$\sigma_{13}^2$	8,3622
$\rho$	0,6141

Las curvas medias estimadas para los cuatro grupos resultan:

CBi Machos:

$$y_{ij} = 40,1556 \exp(-3,1781 \exp(-0,0496 t_j)) ,$$

CBi Hembras:

$$y_{ij} = 32,0437 \exp(-2,9780 \exp(-0,0529 t_j)),$$

BALB Machos:

$$y_{ij} = 36,7163 \exp(-3,1781 \exp(-0,0496 t_j))$$

BALB Hembras:

$$y_{ij} = 29,5429 \exp(-2,9780 \exp(-0,0529 t_j)) .$$

Del modelo estimado se puede concluir que:

- La tasa de crecimiento de la curva es la misma para los ratones de ambos genotipos, pero difiere entre los sexos. Los ratones hembras evolucionan más rápido que los ratones machos.
- El valor del crecimiento límite difiere entre los cuatro grupos. Los ratones del ge-

notipo CBI alcanzan tamaños mayores de crecimiento que los ratones del genotipo BALB para ambos sexos.

#### *Análisis de residuos y comprobación de supuestos*

Para completar el análisis de los datos se examinan los residuos transformados. A través de las figuras se evalúa la adecuación del modelo ajustado y se comprueba la presencia de observaciones atípicas.

En la Figura 3 se observa que los residuos están distribuidos aleatoriamente alrededor del cero y con un rango de variación constante. Esto indica que no existen alejamientos sistemáticos del modelo y que los residuos presentan variancia homogénea. Se puede concluir entonces que el modelo presenta un ajuste adecuado del patrón de cambio de la respuesta media a través del tiempo y que la estructura de covariancias es adecuada.

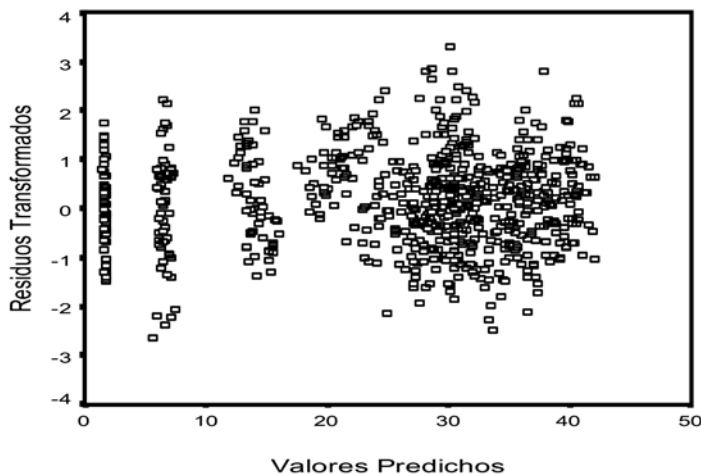
#### **Discusión**

En este trabajo se utiliza el enfoque de los modelos no lineales mixtos para modelar la evolución del peso de ratones y buscar

factores que influyen el crecimiento. Este enfoque es atractivo debido a su flexibilidad para representar múltiples fuentes de heterogeneidad y correlación y para manejar datos incompletos y no balanceados que son comunes en los estudios longitudinales. Además, se ilustra paso a paso la naturaleza interactiva del proceso de construcción del modelo. Para la selección del modelo más adecuado, generalmente, se usan criterios que requieren ajustes repetidos de varios modelos. En el trabajo se usan estadísticas de bondad de ajuste, similares al coeficiente de determinación de los modelos de regresión clásicos, que se aplican directamente al modelo que se está utilizando, agilizando así la construcción del mismo.

Del ajuste de las curvas y su comparación se puede concluir lo siguiente:

- La tasa de crecimiento de la curva es la misma para los dos genotipos, esto indica que, independientemente de cuál sea el peso máximo que alcancen los ratones evolucionan a la misma velocidad. Sin embargo, existen diferencias por sexo. Los ratones hembras evolucionan más rápido que los ratones machos.



**Figura 3:** Residuos transformados versus valores predichos para el Modelo 10.

- La característica que diferencia a los genotipos y al sexo de los ratones es la asíntota de la curva, es decir al valor del crecimiento límite, siendo los ratones machos de la línea CBI los que alcanzan tamaños de crecimiento mayores.

### Agradecimientos

Agradecemos al Dr. Ricardo J. Di Masso que generosamente aportó los datos para poder aplicar la metodología presentada en este trabajo.

### Bibliografía

1. Akaike, H. (1972) A new look look at the statistical model identification. *IEEE Trans. on Automatic Control*, **19**: 716-723.
2. Shwarz G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6**: 461-464.
3. Vonesh, E. F.; Chinchilli, V. M. (1997) *Linear and Nonlinear Models for the Repeated Measurements*. Marcel Dekker.
4. Vonesh, E.F., Chinchilli, V. M. and Pu, K. (1996) "Goodness-of-fit in generalized nonlinear mixed-effects models". *Biometrics* 52, 572-587.
5. Davidian, M.; Giltinan, D. M. (1995) *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall.
6. Lindstrom, M.J.; Bates, D.M. (1990) "Nonlinear Mixed Effects Models for Repeated Measures Data". *Biometrics* 46, 673-687.
7. García, M. del C., Col, A. (2005) "Ajuste de una curva de crecimiento utilizando la función de Gompertz". *Revista FABICIB*, 9, 121-130.
8. Kidwell, J. F.; Howard, A. and Laird, A. K. (1969) "The inheritance of growth and form in the mouse. II. The Gompertz equation". Division of Biological and Medical Sciences, Brown University.
9. Winsor, C. P. (1932) *The Gompertz Curve as a Growth curve*. National Academy of Sciences. Volume 18.
10. Pinheiro, J.C. and Bates, D.M. (1995) "Model building for nonlinear mixed-effects models". *Technical Report*, 91, Department of Biostatistics, University of Wisconsin-Madison.
11. Fitzmaurice, G. M., Laird, N. M. y J. H. Ware (2004). *"Applied Longitudinal Analysis"*. J. Wiley & Sons.
12. Littell, R.C.; Milliken, G.A.; Stroup, W.W.; Wolfinger, R.D.(1996) "SASâ System for Mixed Models". Cary, NC: SAS Institute Inc.