Modelos No Lineales

En los modelos lineales los parámetros (valores poblacionales desconocidos) entran linealmente en la ecuación. No importa que la(s) variable(s) independiente(s) estén en la ecuación en forma no lineal: recordemos que son fijas, ya sea por el diseño o por observación. Lo importante es que cada parámetro aparezca multiplicado por una cantidad conocida y después sumado. Esto permite encontrar soluciones usando métodos para resolver ecuaciones múltiples. Ejemplos clásicos de modelos lineales de regresión son: 1) el modelo de regresión lineal simple

$$\mu_Y = \beta_0 + \beta_1 x\,,$$

2) el modelo de regresión múltiple

$$\mu_Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

y 3) el modelo de regresión polinomial

$$\mu_Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k$$

Por el contrario, los modelos no lineales son modelos de regresión en los cuales los parámetros aparecen en forma no lineal en la ecuación. Por ejemplo,

$$\mu_Y = \frac{1}{\beta_0 + \beta_1 e^{-\beta_2 x}}, \quad \mu_Y = \frac{1}{\left(\beta_0 + \beta_1 x\right)^{\beta_2}}, \quad \mu_Y = \frac{1}{\beta_0 + \beta_1 x + \beta_2 x^2}$$

$$\mu_Y = \beta_0 \exp\left(-\exp\left(-\beta_0 + \beta_1 x\right)\right), \quad \mu_Y = \frac{\beta_0}{1 + \beta_1 e^{-\beta_2 x}}, \quad \mu_Y = \frac{\beta_0}{\left[1 + e^{-\left(\beta_1 + \beta_2 x\right)}\right]^{\beta_3}}$$

Particularmente, estas últimas tres ecuaciones son comúnmente usadas para describir crecimiento, y se llaman, respectivamente, el modelo de Gompertz, el modelo Logístico y el modelo de Richards.

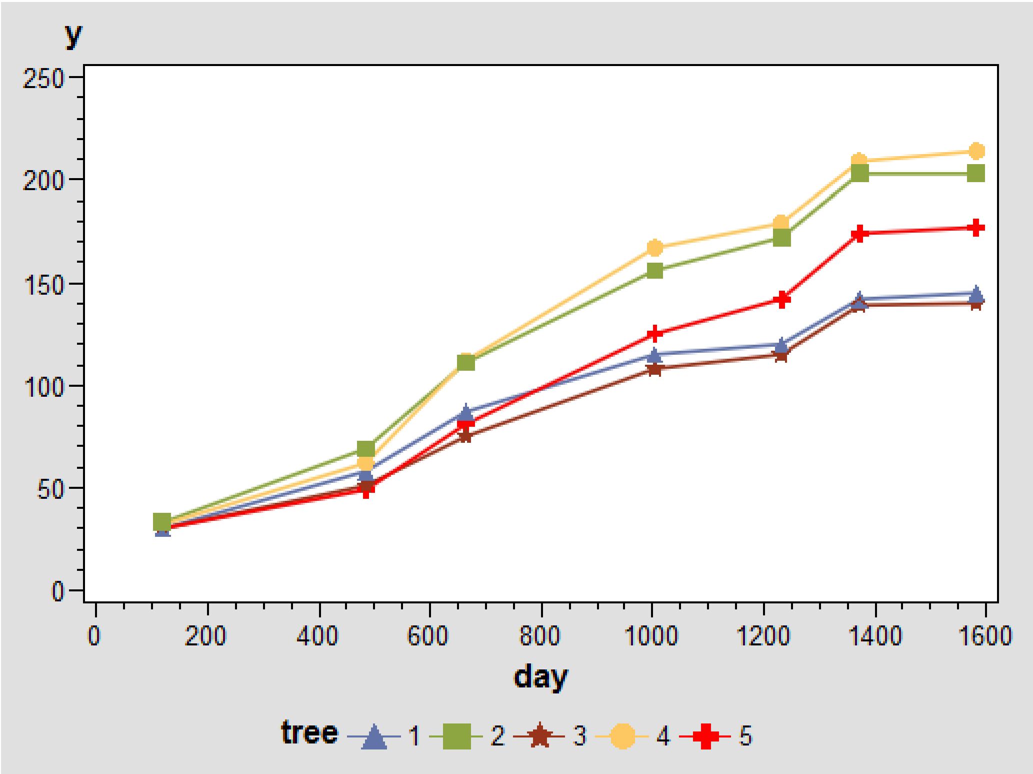
Cuando la respuesta crece (o decrece) monotónicamente, pero la magnitud de la tasa de crecimiento (decrecimiento) se hace cada vez más pequeña, y la variable dependiente se aproxima a una constante (la “asíntota”), la siguiente función exponencial suele usarse:

$$\mu_Y = \beta_0 + \beta_1 e^{-\beta_2 x}$$

En todos estos modelos estamos presentando la media (esperanza) de la respuesta Y como una función no lineal de una (o más variables) independientes. Para completar la especificación del modelo, debemos indicar además la distribución de Y y la independencia (o dependencia) entre los valores de Y.

Al igual que lo estuvimos haciendo en modelos lineales mixtos, existen muchas situaciones en las que el modelo para la media incluye también componentes aleatorias. Por ejemplo, es posible que algunos de los parámetros de una curva (no lineal) de crecimiento varíen de individuo a individuo, y podemos reflejar esta variabilidad mediante términos aleatorios en el modelo. Lo que cambia ahora es que la función no lineal que estamos modelando (y sobre cuyos coeficientes nos interesa hacer inferencias) expresan la esperanza (media) condicional de la variable de respuesta dado(s) el(los) efecto(s) aleatorio(s).

Consideremos mediciones de circunferencia (en mm) de 5 árboles de naranja, con 7 mediciones por árbol. La siguiente figura muestra los perfiles individuales de los 5 árboles:



Un modelo que podemos usar es el Logístico:

$$\mu_Y = \frac{\beta_0}{1 + \beta_1 e^{-x/\beta_2}}$$

En este modelo β_0 representa la asíntota, $\beta_o \big/ (1+\beta_1)$ representa el intercepto y β_2 está relacionado a la pendiente (velocidad con que alcanza la asíntota).

Cuando observamos las curvas individuales, podemos pensar que todas las curvas comienzan más o menos en el mismo diámetro, pero cada una alcanza un diámetro máximo diferente. Supongamos que asumimos que la asíntota β_0 tiene un efecto aleatorio, que le llamaremos u_i. La especificación del modelo condicional sería:

$$E\left(Y_{ij} \mid u_i\right) = \frac{\beta_0 + u_i}{1+\beta_1 e^{-x_{ij}/\beta_2}}$$

Si además suponemos que cada observación, dado el efecto aleatorio del individuo, es independiente de las otras y tiene una distribución normal con varianza constante, y si suponemos que los u_i son a su vez variables normales con cierto valor medio y cierta varianza, tenemos completamente especificado nuestro modelo. Como hemos visto en modelos lineales, la presencia de efectos aleatorios tiene el efecto que las observaciones provenientes de la misma unidad están correlacionadas, ya que comparten el mismo valor observado de u_i. El problema que esta correlación no siempre es simple de estimar, y varía dependiendo de los valores de x involucrados.

Esta formulación del modelo permite considerar varios efectos aleatorios, y no está limitada a datos con distribución normal. La principal dificultad de realizar inferencias con este modelo es el aspecto computacional. No existen fórmulas explícitas para el cálculo de los estimadores máximo verosímiles, y las rutinas de optimización usadas deben combinarse con rutinas de integración numérica, que son particularmente difíciles de usar cuando tenemos más de 3 o 4 efectos aleatorios en un modelo. Un modelo alternativo al presentado anteriormente podría formularse como uno que tenga no solamente un efecto aleatorio en la asíntota sino un efecto aleatorio en β_1:

$$Y_{ij} \mid u_i, v_i \sim N\left(\frac{\beta_0 + u_i}{1+(\beta_1+v_i)e^{-x_{ij}/\beta_2}}, \sigma_\varepsilon^2\right)$$

$$[u_i, v_i] \sim N\left([0,0], \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}\right)$$

Otra dificultad que no surge en modelos lineales con distribución normal es la interpretación de los parámetros. Si el modelo no incluye coeficientes aleatorios el modelo estudia la relación entre la esperanza de la respuesta y la(s) variable(s) independiente(s). Esto significa que podemos interpretar, por ejemplo, el intercepto en el ejemplo de crecimiento de troncos, $\beta_o / (1+\beta_1)$, como un intercepto promedio para la población de árboles de la cual obtuvimos la muestra estudiada (inferencia para el promedio poblacional). En cambio, en el modelo con coeficientes aleatorios no lineales (como el segundo), la relación modelada es la de la esperanza condicional de la respuesta con la(s) variable(s) independiente(s). Entonces $\beta_o / (1+\beta_1)$ ahora se interpreta como el intercepto de un árbol "típico" (típico en el sentido que el valor realizado de los efectos aleatorios es su promedio: [0,0]). Este tipo de interpretación se denomina "inferencia específica para sujetos".

Es decir, que en el modelo no lineal con efectos aleatorios modelamos relaciones condicionales (para un valor dado del efecto aleatorio), mientras que en un modelo no lineal sin efectos aleatorios modelamos relaciones marginales. Más aún, excepto en casos especiales, si en el modelo no lineal mixto se cumple la relación entre la Y y la x con la función logística (como la indicada en el segundo modelo), entonces la esperanza marginal no va a tener la misma relación. Esto se debe a que para obtener la esperanza marginal de la Y a partir de su esperanza condicional debemos "promediar" la esperanza condicional para cada valor posible del efecto aleatorio. En el caso de efectos aleatorios con distribución normal, este proceso implica integrar respecto a la distribución normal bivariada de $[u_i, v_i]$. Este proceso no siempre mantiene la misma relación entre la Y y la x como en el caso de los modelos lineales mixtos. Para ver un ejemplo gráfico de este proceso, consideremos el efecto que tendría promediar pendientes en una regresión lineal y en una regresión no lineal:

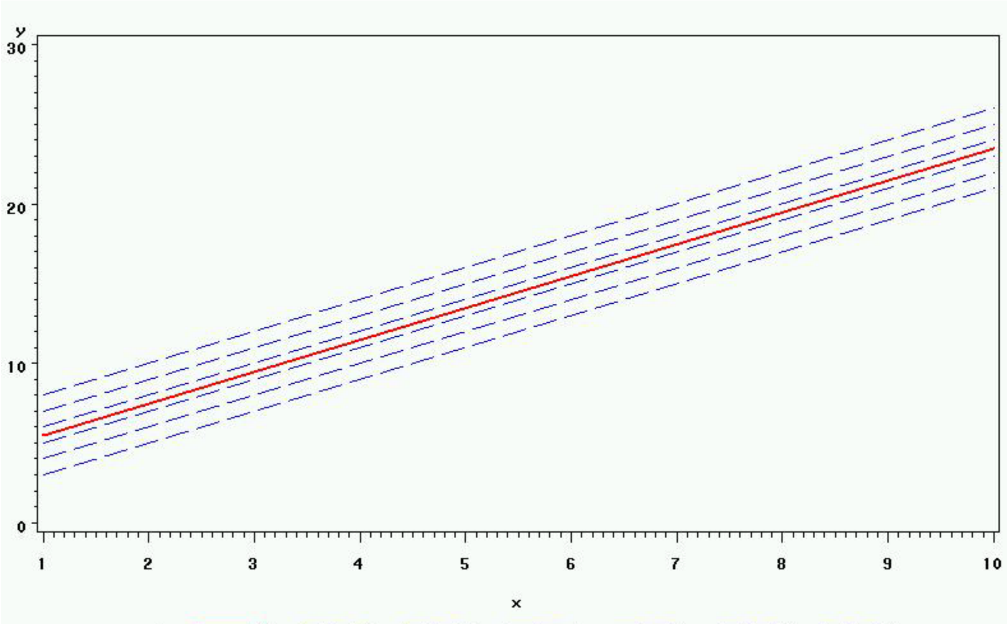


Figura 1. Promedio de modelos lineales.

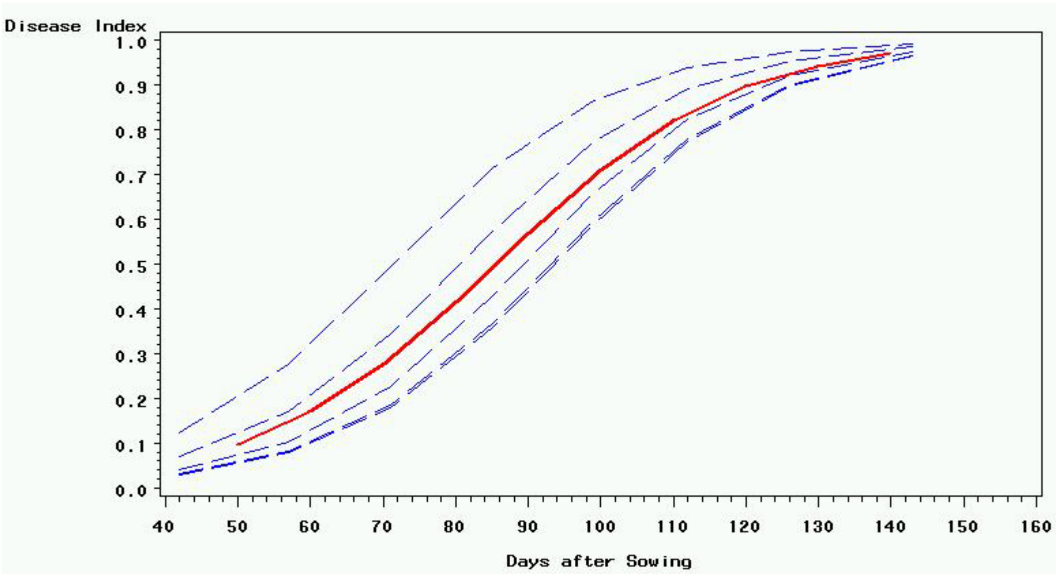


Figura 2. Promedios de modelos no-lineales

Podemos apreciar en las Figuras 1 y 2 que cuando promediamos rectas con pendientes iguales en una línea recta la pendiente promedio es la misma, pero cuando promediamos curvas logísticas con pendientes iguales la pendiente promedio es menor.

¿Cuál de las dos interpretaciones (promedio poblacional o específica de sujetos) es de mayor interés? No hay un consenso general sobre esto, y en alguna aplicaciones (por ejemplo, curvas de crecimiento, los modelos formulados con la esperanza condicional (inferencia específica para sujetos) se consideran más útiles, ya que "controlan" el efecto del sujeto en vez de ignorarlo. Por otra parte, si lo que se desea es interpretar un efecto para el promedio de la población (por ejemplo el efecto general de aplicar cierto tratamiento de descontaminación a predios contaminados) la idea de usar modelos formulados con la esperanza marginal (es decir, sin incluir efectos aleatorios) parece preferible.

Para ajustar los modelos se usa el método de máxima verosimilitud. Recordemos que en este caso para obtener la función de verosimilitud se debe integrar a través de la distribución de los efectos aleatorios, por lo que los algoritmos computacionales incluyen los aspectos de integración y maximización. Para ajustar este tipo de modelos No-Lineales en SAS se usa Proc NLMIXED. Es necesario especificar valores iniciales de los parámetros $\beta_0, \beta_1, \beta_2, \sigma_u^2, \sigma_\varepsilon^2$

```
proc nlmixed data=tree;
   parms b0=190 b1=5 b2=500 su=35 se=8;
   num = b0+u;
   ex = b1*exp(-day/b2);
   den = 1 + ex;
   model y ~ normal(num/den,se*se);
   random u ~ normal(0,su*su) subject=tree;
run;
```

Salida SAS . Datos naranjo

Specifications	
Data Set	WORK.TREE
Dependent Variable	Y
Distribution for Dependent Variable	Normal
Random Effects	U
Distribution for Random Effects	Normal
Subject Variable	Tree
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Dimensions	
Observations Used	35
Observations Not Used	0
Total Observations	35
Subjects	5
Max Obs Per Subject	7
Parameters	5
Quadrature Points	1

Parameters

b0	b1	b2	su	se	NegLogLike
190	5	500	35	8	148.929439

Iteration History

Iter		Calls	NegLogLike	Diff	MaxGrad	Slope
1		3	143.259161	5.670278	1.646156	-57.2423
2		5	141.982244	1.276917	0.515834	-1.31485
3		7	141.869988	0.112256	0.18302	-0.08575
4		10	140.987721	0.882267	1.700026	-0.04074
…		…	…	…	…	…
27		55	131.571885	9.999E-6	0.000781	-0.00002
28		57	131.571885	3.509E-7	0.000068	-5.02E-7

NOTE: GCONV convergence criterion satisfied.

Fit Statistics	
-2 Log Likelihood	263.1
AIC (smaller is better)	273.1
AICC (smaller is better)	275.2
BIC (smaller is better)	271.2

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
b0	192.05	15.6575	4	12.27	0.0003	0.05	148.58	235.52	-5.77E-6
b1	8.0950	0.8567	4	9.45	0.0007	0.05	5.7165	10.4736	8.521E-6
b2	348.07	27.0798	4	12.85	0.0002	0.05	272.89	423.26	-6.47E-7
su	31.6459	10.2612	4	3.08	0.0368	0.05	3.1563	60.1356	-2.85E-6
se	7.8431	1.0126	4	7.75	0.0015	0.05	5.0318	10.6544	0.000068