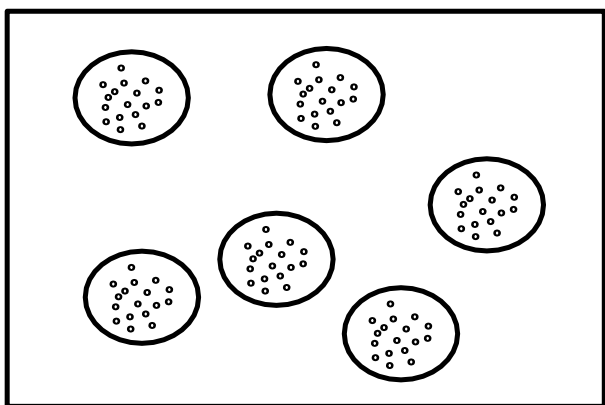


Esta página presenta brevemente los modelos lineales mixtos (MLM) como método para analizar datos no independientes, multinivel/jerárquicos, longitudinales o correlacionados. Nos centramos en los conceptos generales y la interpretación de los MLMS, dedicando menos tiempo a la teoría y los detalles técnicos.

Fondo

Los modelos lineales mixtos son una extensión de los modelos lineales simples para permitir efectos fijos y aleatorios, y se utilizan sobre todo cuando no hay independencia en los datos, como ocurre en una estructura jerárquica. Por ejemplo, los estudiantes podrían ser muestreados dentro de las aulas, o los pacientes dentro de los médicos.

Cuando hay varios niveles, como pacientes atendidos por el mismo médico, la variabilidad del resultado puede considerarse dentro de un grupo o entre grupos. Las observaciones a nivel de paciente no son independientes, ya que dentro de un mismo médico los pacientes son más similares. Las unidades muestreadas al más alto nivel (en nuestro ejemplo, los médicos) son independientes. La siguiente figura muestra una muestra en la que los puntos son pacientes dentro de médicos, los círculos más grandes.



[_ \(https://stats.idre.ucla.edu/wp-content/uploads/2016/02/lmm01.svg\)](https://stats.idre.ucla.edu/wp-content/uploads/2016/02/lmm01.svg)

Existen múltiples formas de tratar los datos jerárquicos. Un enfoque sencillo consiste en agregar. Por ejemplo, supongamos que se toman muestras de 10 pacientes de cada médico. En lugar de utilizar los datos de los pacientes individuales, que no son independientes, podríamos tomar la media de todos los pacientes de un médico. Estos datos agregados serían entonces independientes.

Aunque el análisis de datos agregados arroja estimaciones de efectos y errores estándar coherentes, en realidad no aprovecha todos los datos, porque los datos de los pacientes simplemente se promedian. Si observamos la figura anterior, en el nivel agregado sólo habría seis puntos de datos.

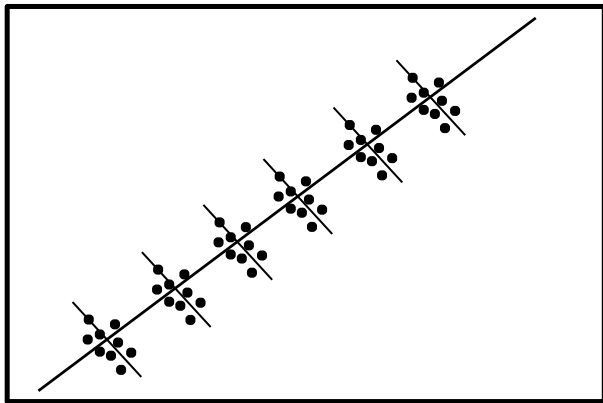
Otro enfoque de los datos jerárquicos consiste en analizar los datos de una unidad cada vez. De nuevo

en nuestro ejemplo, podríamos realizar seis regresiones lineales separadas, una para cada médico de la muestra. De nuevo, aunque esto funciona, hay muchos modelos y cada uno de ellos no aprovecha la información de los datos de otros médicos. Esto también puede hacer que los resultados sean "ruidosos" en el sentido de que las estimaciones de cada modelo son

no se basa en muchos datos

Los modelos lineales mixtos (también llamados modelos multinivel) pueden considerarse un compromiso entre estas dos alternativas. Las regresiones individuales tienen muchas estimaciones y muchos datos, pero son ruidosas. El agregado es menos ruidoso, pero puede perder diferencias importantes al promediar todas las muestras dentro de cada médico. Los MLM se sitúan en un punto intermedio.

Más allá de la mera preocupación por obtener errores estándar corregidos por la no independencia en los datos, puede haber razones importantes para explorar la diferencia entre los efectos dentro de los grupos y entre ellos. Un ejemplo de esto se muestra en la siguiente figura. Volvemos a tener pacientes de los seis médicos y observamos un gráfico de dispersión de la relación entre un predictor y el resultado. Dentro de cada médico, la relación entre el predictor y el resultado es negativa. Sin embargo, entre médicos, la relación es positiva. Los LMM nos permiten explorar y comprender estos importantes efectos.



(<https://stats.idre.ucla.edu/wp-content/uploads/2016/02/lmm02.svg>)

Efectos aleatorios

El núcleo de los modelos mixtos es que incorporan efectos fijos y aleatorios. Un efecto fijo es un parámetro que no varía. Por ejemplo, podemos suponer que existe una línea de regresión verdadera en el

población, β , y obtenemos una estimación de la misma, $\hat{\beta}$. Por el contrario, los efectos aleatorios son parámetros

variables aleatorias. Por ejemplo, podríamos decir que β se distribuye como una variable aleatoria normal con media μ y desviación típica σ , o en forma de ecuación:

$$\beta \sim N(\mu, \sigma)$$

Esto es realmente lo mismo que en la regresión lineal, donde suponemos que los datos son variables aleatorias, pero los parámetros son efectos fijos. Ahora los datos son variables aleatorias, y los parámetros son variables aleatorias (en un nivel), pero fijas en el nivel superior (por ejemplo, seguimos suponiendo unos

media de la población, μ).

Teoría de los modelos lineales mixtos

$$y = X\beta + Zu + \varepsilon$$

Donde y es un vector de columnas $N \times 1$, la variable de resultado; X es una matriz $N \times p$ de las p variables predictoras; β es un vector de columnas $p \times 1$ de los coeficientes de regresión de efectos fijos (los β s); Z es la matriz de diseño $N \times qJ$ para los q efectos aleatorios y J grupos; u es un vector $qJ \times 1$ de q efectos aleatorios.

(el complemento aleatorio del β fijo) para J grupos; y ε es un vector de columnas $N \times 1$ de los residuos, la parte de y que no explica el modelo, $X\beta + Zu$. Recapitemos:

$$y = X\beta + Zu + \varepsilon$$

$N \times p$ $p \times 1$ $N \times qJ$ $qJ \times 1$

Para concretar, veamos un ejemplo de un conjunto de datos simulado. Los médicos ($J = 407$) indexados por el subíndice j atienden cada uno a n_j pacientes. Por tanto, nuestra variable de agrupación es el médico. No todos los

El número total de pacientes es la suma de los pacientes atendidos por cada médico. El número total de pacientes es la suma de los pacientes atendidos por cada médico.

$$N = \sum_j^J n_j$$

En nuestro ejemplo, los médicos atendieron a $N = 8525$ pacientes. Nuestro resultado, y , es una variable continua, las puntuaciones de movilidad. Además, supongamos que tenemos 6 predictores de efectos fijos, Edad (en años), Casado (0 = no, 1 = sí, 1 = no, 1 = no).

sí), Sexo (0 = mujer, 1 = hombre), Recuento de glóbulos rojos (RBC) y Recuento de glóbulos blancos (WBC) más un intercepto fijo y un intercepto aleatorio ($q = 1$) para cada uno de los $J = 407$ médicos. Para simplificar, sólo vamos a considerar los interceptos aleatorios. Dejaremos que todos los demás efectos sean fijos por ahora. La dirección

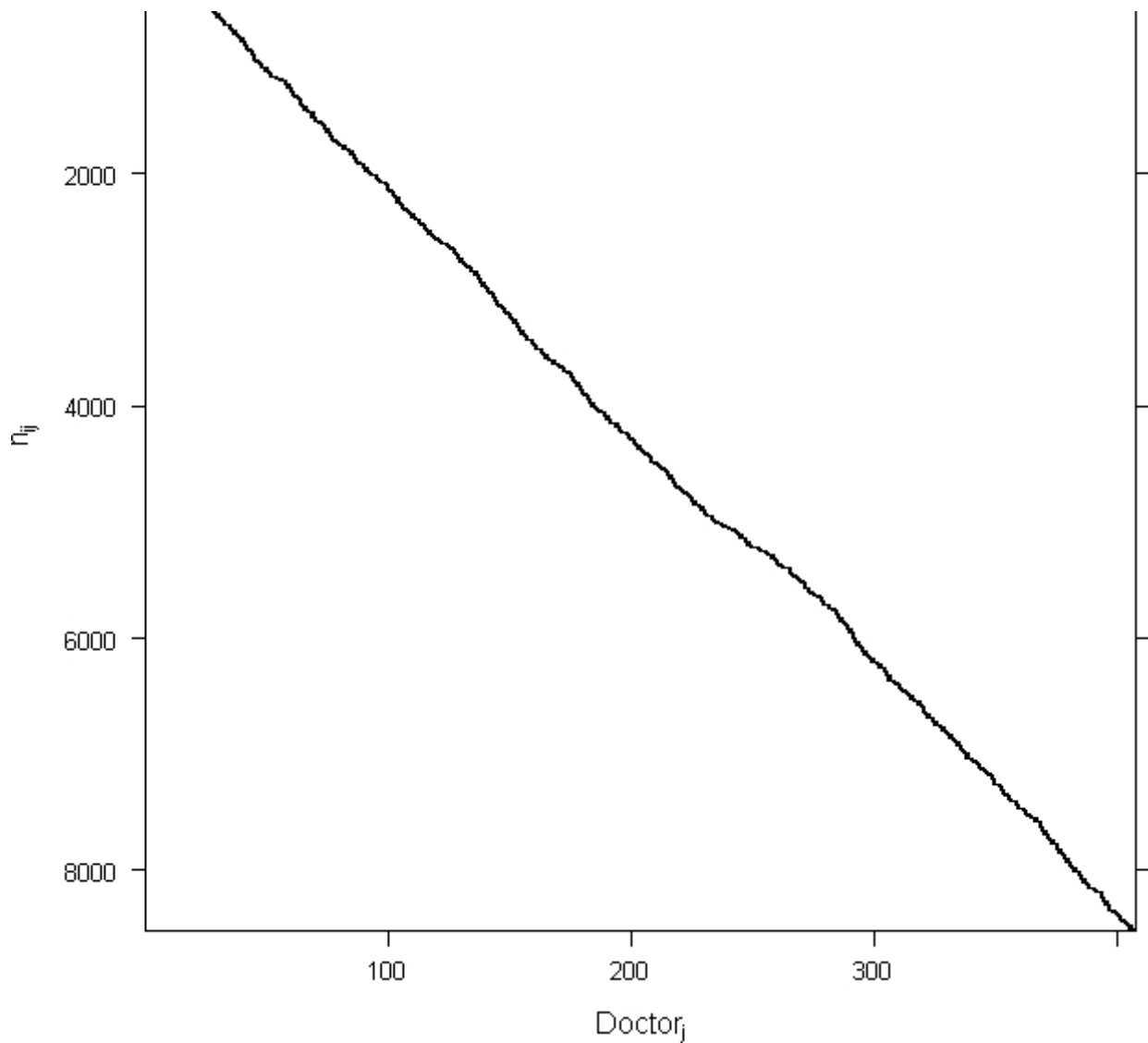
La razón por la que queremos efectos aleatorios es que esperamos que las puntuaciones de movilidad dentro de los médicos estén correlacionadas. Esto puede deberse a muchas razones. Por ejemplo, los médicos pueden tener especialidades que signifiquen que tienden a ver pacientes con cáncer de pulmón con síntomas particulares o algunos médicos pueden ver casos más avanzados, de forma que dentro de un médico, los pacientes son más homogéneos que entre médicos. Para volver a poner este ejemplo en nuestra notación matricial, para la respuesta n_j dimensional y_j para el médico j tendríamos:

$$y_j = X_j\beta + Z_ju + \varepsilon_j$$

$n_j \times 1$ $n_j \times 6$ 6×1 $n_j \times 1$ 1×1 $n_j \times 1$

y apilando las observaciones de todos los grupos, ya que $q = 1$ para el modelo de intercepto aleatorio, $qJ = (1)(407) = 407$ por lo que tenemos:

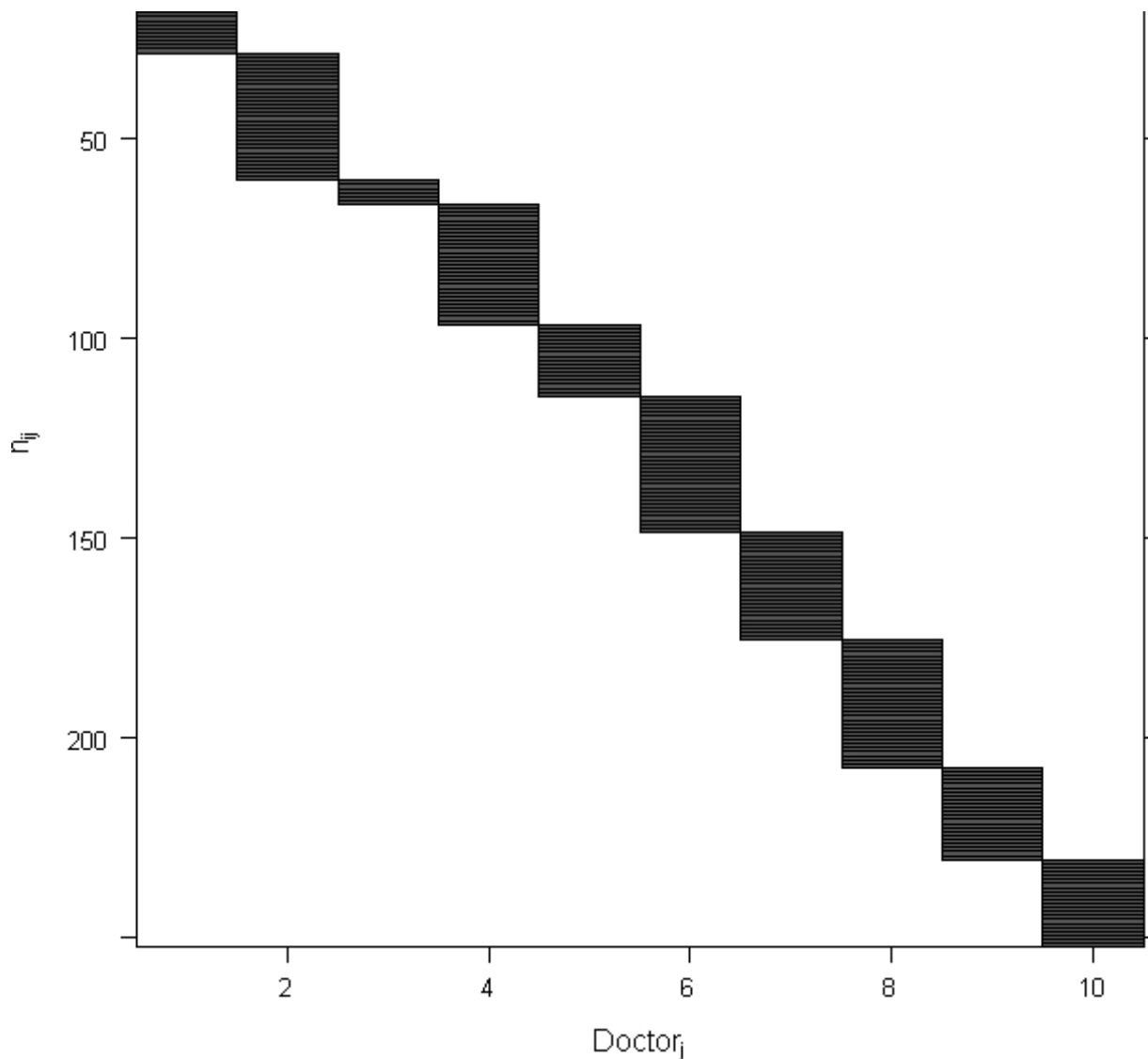
movilidad n	Intercepción	Edad	Casado	SexoWBCRBC
-------------	--------------	------	--------	------------



Para ver la estructura con más detalle, también podemos hacer zoom sólo en los 10 primeros médicos. El espacio relleno indica filas de observaciones que pertenecen al médico de esa columna, mientras que el espacio en blanco indica que no pertenecen al médico de esa columna.

First 10 Doctors Z: 252 x 10





Si la estimáramos, u sería un vector columna, similar a β . Sin embargo, en estadística clásica, en realidad no estimamos u . En su lugar, casi siempre suponemos que:

$$u \sim N(0, G)$$

Que se lee: " u se distribuye como normal con media cero y varianza G ". Donde G es la matriz de varianza-covarianza de los efectos aleatorios. Dado que estimamos directamente los efectos fijos, incluyendo el intercepto del efecto fijo, los complementos de los efectos aleatorios se modelan como desviaciones del efecto fijo, por lo que tienen media cero. Los efectos aleatorios son sólo desviaciones alrededor del valor en β , que es la media. Así que lo que queda por estimar es la varianza. Como nuestro ejemplo sólo tenía un efecto aleatorio

intercepto, G es sólo una matriz 1×1 , la varianza del intercepto aleatorio. Sin embargo, puede ser mayor. En

ejemplo, supongamos que tenemos un intercepto aleatorio y una pendiente aleatoria, entonces

$$G = \begin{bmatrix} \sigma_{\text{int}}^2 & \sigma_{\text{int,pendiente}}^2 \\ \sigma_{\text{int,pendiente}}^2 & \sigma_{\text{pendiente}}^2 \end{bmatrix}$$

Como G es una matriz de varianza-covarianza, sabemos que debe tener ciertas propiedades. En

En particular, sabemos que es cuadrada, simétrica y semidefinida positiva. También sabemos que tiene elementos redundantes. Para una matriz $q \times q$, $\frac{q(q+1)}{2}$ elementos únicos. Para simplificar

existen
mediante la eliminación de efectos redundantes y garantizar que la matriz de estimación resultante sea positiva.

definida, en lugar de modelar G directamente, estimamos θ (por ejemplo, una factorización Cholesky triangular

$G = LDL^T$). θ no siempre se parametriza de la misma manera, pero en general se puede pensar que representa los efectos aleatorios. Suele estar diseñado para contener elementos no redundantes (a diferencia del

matriz de covarianzas) y parametrizarse de forma que se obtengan estimaciones más estables que las varianzas (como tomar el logaritmo natural para garantizar que las varianzas sean positivas). Independientemente de los detalles, podemos decir que

$$G = \sigma(\theta)$$

En otras palabras, G es una función de θ . Por tanto, obtenemos una estimación de θ que llamamos

θ^\wedge . Diversas parametrizaciones y restricciones nos permiten simplificar el modelo, por ejemplo suponiendo que los efectos aleatorios son *independientes*, lo que implicaría que la verdadera estructura es

$$G = \begin{bmatrix} \sigma_{\text{int}}^2 & 0 \\ 0 & \sigma_{\text{pendiente}}^2 \end{bmatrix}$$

El elemento final de nuestro modelo es la matriz de varianza-covarianza de los residuos, ε o la matriz de varianza-covarianza de la distribución condicional de $(y|\beta; u = u)$. La estructura de covarianza residual más común es

$$R = I\sigma_\varepsilon^2$$

donde I es la matriz de identidad (matriz diagonal de 1s) y σ_ε^2 es la varianza residual. Esta estructura supone una varianza residual homogénea para todas las observaciones (condicionales) y que son

(condicionalmente) independientes. Pueden asumirse otras estructuras, como la simetría compuesta o la autorregresión. La terminología G es común en SAS, y también lleva a hablar de estructuras del lado G para la matriz de covarianza de varianza de efectos aleatorios y estructuras del lado R para la matriz de covarianza de varianza residual.

Así, los elementos fijos finales son y , X , Z y ε . Los elementos estimados finales son β^\wedge , θ^\wedge y R^\wedge . El modelo final depende de la distribución asumida, pero generalmente es de la forma

$$(y|\beta; u = u) \sim N(X\beta + Zu, R)$$

También podríamos enmarcar nuestro modelo en una ecuación de dos niveles para el i -ésimo

paciente para el j -ésimo médico. En este caso, trabajamos con variables a las que añadimos subíndices en lugar de vectores, como antes. La ecuación de nivel 1 añade subíndices a los parámetros β_S para indicar a qué médico pertenecen. Pasando a las ecuaciones de nivel 2, podemos ver que cada estimación β para un médico concreto, β_{pj} , puede representarse como

una combinación de una estimación media para ese parámetro, γ_{p0} , y un efecto aleatorio para ese médico, (u_{pj}) . En este modelo concreto, vemos que sólo se permite que el intercepto (β_{0j}) varíe entre médicos porque es la única ecuación con un término de efecto aleatorio, (u_{0j}) . Los demás β_{pj} son constantes entre médicos.

$$\begin{aligned} \text{L1 : } Y_{ij} &= \beta_{0j} + \beta_{1j} \text{ Edad}_{ij} + \beta_{2j} \text{ Casado}_{ij} + \beta_{3j} \text{ Sexo}_{ij} + \beta_{4j} \text{ WBC}_{4ij} + \beta_{5j} \text{ RBC}_{5ij} + e_{ij} \\ \text{L2 : } \beta_{0j} &= \gamma_{00} + u_{0j} \\ \text{L2 : } \beta_{1j} &= \gamma_{10} \\ \text{L2 : } \beta_{2j} &= \gamma_{20} \\ \text{L2 : } \beta_{3j} &= \gamma_{30} \\ \text{L2 : } \beta_{4j} &= \gamma_{40} \\ \text{L2 : } \beta_{5j} &= \gamma_{50} \end{aligned}$$

Si se sustituyen las ecuaciones del nivel 2 en el nivel 1, se obtiene la especificación del modelo mixto. Aquí agrupamos los parámetros de intercepción fija y aleatoria para mostrar que combinados dan la intercepción estimada para un médico concreto.

$$Y_{ij} = (\gamma_{00} + u_{0j}) + \gamma_{10} \text{ Edad}_{ij} + \gamma_{20} \text{ Casado}_{ij} + \gamma_{30} \text{ SEXO}_{ij} + \gamma_{40} \text{ WBC}_{40ij} + \gamma_{50} \text{ RBC}_{50ij} + e_{ij}$$

Referencias

- Fox, J. (2008). *Applied Regression Analysis and Generalized Linear Models*, 2ª ed. Sage Publications.
- McCullagh, P, y Nelder, J. A. (1989). *Generalized Linear Models*, 2ª ed., Chapman & Hall/CRC Press. Chapman & Hall/CRC Press.
- Snijders, T. A. B. & Bosker, R. J. (2012). *Multilevel Analysis*, 2ª ed. Sage Publications.
- Singer, J. D. & Willett, J. B. (2003). *Análisis longitudinal de datos aplicado: Modeling Change and Event Occurrence*. Oxford University Press.
- Pinheiro, J. & Bates, D. (2009). *Modelos de efectos mixtos en S y S-Plus*. 2nd printing. Springer.
- Galecki, A. & Burzykowski, T. (2013). *Linear Mixed-Effects Models Using R*. Springer.
- Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, longitudinal, and structural equation models*. Chapman & Hall/CRC Press.
- Gelman, A. y Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2003). *Bayesian Data Analysis*, 2ª ed., Chapman & Hall/CRC. Chapman & Hall/CRC.

Haga clic aquí para informar de un error en esta página o dejar un comentario

Cómo citar esta página ([https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-cite-web-
pages-and-programs-from-the-ucla-statistical-consulting-group/](https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-cite-web-pages-and-programs-from-the-ucla-statistical-consulting-group/))

© 2021 UC REGENTS (<http://www.ucla.edu/terms-of-use/>)

INICIO (/)

CONTACTO (/contacto)