Leticia Hachuel
Gabriela Boggio
Guillermina Harvey
Diego Marfetán
Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística

ESTIMACIÓN DE UN MODELO LOGIT PARA DATOS CORRELACIONADOS. COMPARACIÓN DE LOS ENFOQUES CLÁSICO Y BAYESIANO

1. INTRODUCCIÓN

En muchas aplicaciones los individuos bajo estudio presentan algún tipo de agrupamiento que provoca que las observaciones provenientes de sujetos de un mismo grupo tiendan a estar correlacionadas. Tal es el caso de los estudios multicéntricos, en los cuales se presupone que las mediciones de sujetos pertenecientes al mismo centro tienden a parecerse más entre ellas que con las de otros centros o grupos. En el análisis estadístico de este tipo de datos frecuentemente se busca modelar respuestas como función de covariables, ajustando los resultados por esta potencial correlación. Un enfoque posible para ello es el que incluye en el modelo efectos no observados que varían aleatoriamente y reciben el nombre de efectos aleatorios (Goldstein, 2003; Fitzmaurice et al., 2004).

Estos modelos constituyen la clase de los denominados modelos lineales generalizados mixtos (MLGM), los cuales admiten variables respuestas no normales y permiten modelar una función de la media a través de efectos fijos, asociados a variables medidas tanto a nivel individual como grupal, y de efectos aleatorios en el predictor lineal, los que generalmente se suponen distribuidos normalmente.

El enfoque clásico de estimación de estos modelos es el de máximo-verosimilitud ya que es posible especificar la función de probabilidad paramétrica completa. Otra alternativa de estimación proviene del paradigma bayesiano, el cual hace uso de una distribución a priori que describe creencias acerca de esas cantidades desconocidas, parámetros, independiente de los datos.

Browne y Draper (2006) compararon los resultados de la aplicación de la inferencia clásica y bayesiana para este tipo de modelos, también denominados jerárquicos. Bajo el enfoque bayesiano estos modelos requieren especificar la distribución a priori del parámetro de escala o hiperparámetro. Gelman (2006) estudió el comportamiento de diferentes distribuciones a priori para dicho hiperparámetro.

En este trabajo se realiza un estudio por simulación para comparar las estimaciones de los parámetros de un modelo logit con intercepto aleatorio para datos agrupados con diferente nivel de correlación a partir del enfoque clásico y el bayesiano.

En la sección siguiente se describe brevemente el modelo estadístico utilizado, seguidamente se presenta el estudio de simulación elegido y los resultados hallados.

2. METODOLOGÍA

2.1. Modelo lineal generalizado mixto

Un MLGM se puede formalizar de la siguiente manera. Sea Y_{ij} la respuesta para el j-ésimo individuo del i-ésimo grupo, la cual puede ser continua, binaria o de conteo. Asociado con cada Y_{ij} hay un vector (fila) X_{ij} de dimensión 1x(p+1), que contiene el valor 1 asociado al intercepto y el valor de p covariables que pueden variar de grupo a grupo o bien de individuo a individuo dentro de cada grupo.

Para el caso particular en que Y_{ij} es una respuesta binaria, un posible MLGM simple es un modelo logístico con interceptos aleatorios que se especifica de la siguiente forma (Fitzmaurice et al., 2004):

1.- Condicional sobre un único efecto aleatorio, b_i, las Y_{ij} son independientes y tienen una distribución de probabilidad Bernoulli, con

$$Var(Y_{ij}/b_i)=E(Y_{ij}/b_i)\{1-E(Y_{ij}/b_i)\} \tag{1}$$

2.- La media condicional de Y_{ij} depende de efectos fijos y aleatorios a través de la siguiente expresión:

$$\log\left\{\frac{Pr(Y_{ij}=1/b_i)}{1-Pr(Y_{ij}=1/b_i)}\right\}=X_{ij}\beta+b_i \tag{2}$$

Es decir la media condicional de Y_{ij} se relaciona con el predictor lineal a través del enlace logit, siendo β el vector de (p+1) parámetros fijos.

3.- El único efecto aleatorio b_i tiene una distribución de probabilidad univariada con media cero y variancia σ_b^2 que usualmente se supone Normal.

La premisa básica de los MLGM es que la correlación entre las unidades de un mismo grupo puede pensarse que surge por el hecho de compartir un conjunto de efectos aleatorios. Condicional sobre los efectos aleatorios, las observaciones de diferentes grupos se suponen independientes y con una distribución de probabilidad perteneciente a la familia exponencial.

2.2. Métodos de estimación

Una vez postulado un MLGM, el método de estimación máximo-verosímil requiere especificar la función de verosimilitud que se considera función de los parámetros dadas las observaciones, esto es:

$$L(\beta,\sigma/y)=\int f(y/b;\beta)f(b;\sigma)db \tag{3}$$

Esta expresión generalmente es difícil de resolver y se complica aún más cuando el número de efectos aleatorios aumenta. Se pueden utilizar métodos de integración numérica para aproximar la verosimilitud. El error inducido por reemplazar la integral por una suma finita, como lo hacen los métodos de cuadratura de Gauss-Hermite, se hace cada vez más difícil de controlar a medida que la dimensión de la integral aumenta. Las aproximaciones convergen a las estimaciones máximo-verosímiles (MV) a medida que el número de puntos de cuadratura se incrementa de una manera apropiada para la integración numérica (Song, 2007).

Por otro lado, el enfoque bayesiano tiene dos partes: un modelo estadístico que describe la distribución de los datos dadas cantidades desconocidas y una distribución a priori que describe las creencias acerca de esas cantidades desconocidas independiente de los datos. En este sentido el enfoque bayesiano trata a los parámetros como variables aleatorias con una determinada distribución de probabilidad y la inferencia estadística descansa totalmente en la distribución a posteriori de los parámetros, dados los datos. Para obtener estas distribuciones a posteriori se deben elegir previamente las distribuciones a priori de los parámetros del modelo, las que se combinan luego con la función de verosimilitud de la siguiente manera, según el teorema de Bayes (Dobson, 2008; Bernardo & Smith, 1994):

$$h(\beta,\sigma/y)=\frac{f(y/\beta,\sigma)f(\beta,\sigma)}{f(y)} \tag{4}$$

El denominador de esta expresión depende sólo de los datos por lo que resulta irrelevante para la inferencia sobre los parámetros; de ahí que la distribución a priori multiplicada por la función de verosimilitud determina finalmente la distribución a posteriori.

Excepto en casos particulares, no hay una expresión cerrada para la distribución a posteriori, por lo que se usan métodos de remuestreo para aproximarla. El método más conocido es el MCMC, por el cual se diseña un proceso estocástico con forma de Cadena de Markov de manera tal que la distribución estacionaria sea la distribución a posteriori del parámetro de interés.

El proceso comienza con la selección de estimaciones iniciales y se desprecia un primer tramo de la cadena hasta que la distribución sea cercana a la estacionaria. Los restantes valores simulados se supone que proveen información acerca de la distribución a posteriori. Como las observaciones sucesivas de la cadena de Markov están correlacionadas, se poda el proceso para obtener observaciones levemente correlacionadas. Finalmente cuando se tienen las suficientes observaciones después del descarte de forma de aproximarse lo mejor posible a la distribución a posteriori, se obtienen medidas resumen de interés como la media, modo y desviación estándar, las cuales proveen información sobre los parámetros del modelo.

La metodología bayesiana necesita especificar distribuciones a priori para todos los parámetros del modelo, lo cual puede llegar a ser una complicación en el caso particular abordado ya que, además de especificar una distribución a priori para los parámetros fijos del modelo, es necesario especificar la distribución a priori del hiperparámetro relacionado con la variancia del efecto aleatorio. Gelman (2006) señala inconvenientes en el uso de la distribución inversa gamma no informativa como distribución a priori de la variancia del efecto aleatorio y sugiere utilizar la distribución uniforme para la desviación estándar como también la distribución “half Cauchy” especialmente para el caso de agrupamiento en escasa cantidad de grupos.

3. ESTUDIO DE SIMULACIÓN

Se diseña un estudio por simulación a fin de comparar el comportamiento de los estimadores de los parámetros de un MLGM particular: un modelo logit con intercepto aleatorio y una única variable explicativa, el cual se formaliza de la siguiente manera:

$$\log\left\{\frac{\Pr(Y_{ij}=1/b_i)}{1-\Pr(Y_{ij}=1/b_i)}\right\}=\eta_{ij}=\beta_0+\beta_1X_{ij}+cb_i \qquad (5)$$

donde c es una constante cuya magnitud produce diferentes niveles de correlación intragrupo. El modelo permite generar datos con una estructura de correlación intercambiable (es decir, todos los pares de observaciones dentro de cada grupo están igualmente correlacionados). Se ha comprobado en otros trabajos la relación entre la magnitud del coeficiente c y el grado de correlación entre las respuestas de los individuos de un mismo grupo para diferentes valores de c considerando que un valor de c=1 produce grupos de baja correlación y un valor de c=4, alta correlación intra-grupo.

Los estimadores a evaluar son los obtenidos por el método de máxima verosimilitud (MV) implementado en el procedimiento GLIMMIX de SAS y el método bayesiano proporcionado por el procedimiento MCMC del mismo programa computacional.

Para llevar adelante el estudio por simulación se generan datos a partir del algoritmo empleado por Evans y Hosmer (2004), que considera el modelo (5). En él, X_{ij} es el valor asumido por una variable tipo Bernoulli con probabilidad igual a 0.5 y b_i es el valor asumido por una variable aleatoria Normal (0,1). Por último el coeficiente c, con los valores 1 y 4 permiten representar situaciones con diferente nivel de correlación entre las unidades de un mismo grupo.

Para valores predeterminados de β_0, β_1 y c y valores aleatorios para b_i y X_{ij}, se determinan $\pi_{ij}=Pr(Y_{ij}=1/b_i)$, a partir de las cuales se obtienen valores binarios 0 ó 1 comparando cada probabilidad con un valor elegido al azar de una distribución Uniforme definida en el intervalo [0;1].

A los valores binarios generados por el algoritmo recién descrito se les ajusta el modelo logit mixto (5), con un coeficiente aleatorio que se supone distribuido Normal. Se obtienen las estimaciones de los coeficientes del modelo y de la variancia de los efectos aleatorios.

Se repite este proceso de generación de datos y ajuste de modelos 1000 veces. Se calcula el promedio de las estimaciones de los parámetros del modelo y el promedio de sus correspondientes desviaciones estándares a través de las 1000 muestras de acuerdo a los dos métodos de estimación presentados. Además, para el caso de la estimación MV se calcula el desvío estándar empírico de las estimaciones de los parámetros del modelo a través de las 1000 muestras. Los escenarios de análisis se definieron teniendo en cuenta diferentes valores y combinaciones de:

n: número de grupos en la muestra.

k: número de individuos dentro de cada grupo.

En este trabajo se presentan los resultados considerando en el modelo $\beta_0=1$, $\beta_1=0.8$, c=1 ó 4 y combinaciones entre valores de k=5, 30 y valores de n=10, 30. Los escenarios definidos por estos valores de k y n se eligieron en función de la falta de consenso acerca del número mínimo de grupos necesarios para garantizar los resultados del ajuste de modelos de regresión jerárquicos (Austin, 2010).

Para la estimación del modelo (5) mediante el procedimiento MCMC se eligieron distribuciones a priori no informativas para los parámetros. Para los efectos fijos se supusieron distribuciones normales con media 0 y variancia igual a 1000. Los interceptos aleatorios siempre se supusieron con distribución Normal y se consideraron diferentes distribuciones prior difusas para el hiperparámetro, a saber: distribución uniforme para el desvío estándar e inversa Gamma para la variancia y la “half Cauchy” considerada débilmente informativa por Gelman (2006). También se consideró una distribución informativa Inversa Gamma con apropiados valores de sus parámetros. Para cada parámetro se calculó la media de la distribución a posteriori respectiva. Se eligió la estadística de Geweke para comprobar la convergencia de las cadenas para los parámetros de interés.

4. RESULTADOS

La Tabla 1 muestra los resultados hallados para los escenarios planteados a través de la generación de 1000 muestras para el coeficiente de regresión del modelo bajo los dos métodos de estimación.

Tabla 1

Estimación MV y Bayesiana de β_1=0.8 a través de 1000 muestras simuladas según número de grupos (n), de individuos por grupo (k) y valores de c

c	n	k	Estimación MV			Estimación Bayesiana			
						Prior para σ U(0,100)		Prior para σ^2 IG(0.01;0.01)	
			$\hat{\beta}_1$ promedio	DS($\hat{\beta}_1$) promedio	DS empírico de $\hat{\beta}_1$	$\hat{\beta}_1$ promedio	DS($\hat{\beta}_1$) promedio	$\hat{\beta}_1$ promedio	DS($\hat{\beta}_1$) promedio
1	10	5	1.13	93.50	2.43	1.47	1.33	1.38	1.19
		30	0.80	0.31	0.31	0.83	0.32	0.83	0.31
	30	5	0.83	0.46	0.46	0.89	0.49	0.86	0.47
		30	0.80	0.18	0.18	0.80	0.18	0.80	0.18
4	10	5	1.28	28.10	3.63	2.34	3.19	1.82	2.41
		30	0.80	12.26	0.76	0.87	0.46	0.84	0.44
	30	5	0.85	0.67	0.73	0.97	0.74	0.95	0.72
		30	0.80	0.24	0.25	0.82	0.24	0.81	0.24

Cuando hay baja correlación dentro de los grupos (c=1) la estimación MV presenta un sesgo positivo cuando el número de observaciones dentro de cada grupo y la cantidad de grupos son pequeñas (k=5 y n=10). Si aumenta el grado de correlación intragrupo (c=4) se observan sesgos levemente mayores en las estimaciones puntuales y las desviaciones estándares son muy diferentes a las desviaciones estándares reales para los escenarios con n=10.

Con respecto a las estimaciones bayesianas, se observan también sesgos en la estimación del parámetro en el caso n=10 y k=5 para los dos valores de c considerados. Particularmente en los escenarios de pocas observaciones (n=10 para cualquier k) resultan muchos más precisas cualquiera sea el c que las estimaciones MV.

La Tabla 2 muestra los resultados del promedio de las estimaciones MV y bayesianas de la componente de variancia del modelo que tienen un valor teórico igual a 1 para el caso de c=1 y 16 para c=4 de acuerdo al modelo de generación de datos utilizado.

Tabla 2

Estimación MV y Bayesiana de σ_b^2 a través de 1000 muestras simuladas según número de grupos (n), de individuos por grupo (k) y valores de c

c	n	k	Estimación MV			Estimación Bayesiana			
						Prior para σ U(0,100)		Prior para σ^2 IG(0.01;0.01)	
			$\hat{\sigma}_b^2$ promedio	DS($\hat{\sigma}_b^2$) promedio	DS empírico de $\hat{\sigma}_b^2$	$\hat{\sigma}_b^2$ promedio	DS($\hat{\sigma}_b^2$) promedio	$\hat{\sigma}_b^2$ promedio	DS($\hat{\sigma}_b^2$) promedio
1	10	5	1.36	2.19	2.20	23.59	57.37	5.89	11.05
		30	0.91	0.58	0.63	2.00	1.94	1.49	1.26
	30	5	1.11	0.86	0.89	2.00	1.55	1.48	1.21
		30	0.97	0.34	0.36	1.22	0.47	1.13	0.43
4	10	5	255.95	158.13	2325.82	923.10	1002.90	1177.64	1800.47
		30	20.33	17.68	34.43	89.49	144.80	42.66	56.41
	30	5	22.99	7.11	61.35	56.16	64.59	41.56	42.28
		30	15.54	5.97	5.91	23.04	11.52	20.98	9.99

En el primer caso, cuando c=1, las estimaciones MV son aceptables aún para escenarios muestrales chicos, en cambio cuando c=4 el comportamiento es excesivamente malo tanto en la estimación puntual como en su desvío estándar llegando a estabilizarse en valores aceptables en el escenario con n=30 y k=30. Respecto a las estimaciones bayesianas, las obtenidas con la prior uniforme resultan sesgadas con sesgos mucho más pronunciados para c=4 incluso para escenarios de muchas observaciones y resultan en general muy imprecisas para este caso con c=4. Las estimaciones obtenidas bajo la distribución a priori Inversa Gamma difusa resultan sesgadas pero en mucho menor grado que en el caso de la prior Uniforme cuando c=1. Cuando c=4 las estimaciones son exageradamente sesgadas e imprecisas en el escenario n=10 y k=5. La estimación puntual toma un valor más cercano al teórico sólo para el escenario n=30 y k=30 pero con precisión menor que la MV.

Cabe aclarar que se realizaron también simulaciones para distribuciones a priori no informativas más difusas, a saber: Uniforme (0, 1000) para σ_b e Inversa gamma (0.001, 0.001) para σ_b^2 y se obtuvieron resultados análogos a los presentados.

Los resultados hallados en este estudio por simulación relacionados con la inferencia bayesiana orientaron a completar el estudio con la elección de una distribución a priori para la componente de variancia débilmente informativa y otra completamente informativa, esta última sólo a los efectos comparativos. La primera de estas distribuciones intenta restringir la distribución a posteriori a un rango de variación más apropiado para los datos. Para ello se utiliza la distribución “half Cauchy” con parámetros (0, 2) para el caso c=1 y (0, 8) cuando c=4 y una distribución Inversa Gamma con parámetros determinados de acuerdo a los niveles teóricos 1 y 16 de la σ_b^2. En esta oportunidad se eligieron sólo algunos escenarios, particularmente los que presentaron peores resultados, es decir aquellos con n=10.

En relación a la estimación del coeficiente de regresión, los resultados hallados permiten observar cómo los sesgos son levemente menores cuando se utiliza la prior “half Cauchy” que la distribución Uniforme para k=5 cualquiera sea el valor de c. Por otro lado las diferencias entre las estimaciones obtenidas con la prior informativa y la débilmente informativa no son tan evidentes (Tabla 1 y Tabla 3).

Si se comparan los resultados basados en la prior “half Cauchy” con los obtenidos a través de la estimación clásica MV, si bien la primer opción presenta sesgos un tanto mayores es notoria la mejor precisión de las estimaciones sobre todo para k=5 (Tabla 3).

Tabla 3

Estimación MV y Bayesiana de β_1=0.8 a través de 1000 muestras simuladas para muestras de 10 grupos según número de individuos por grupo (k) y valores de c

c	k	Estimación MV			Estimación Bayesiana			
					Prior para σ half-Cauchy[(a)]		Prior para σ^2 Inversa Gamma informativa[(b)]	
		$\hat{\beta}_1$ promedio	DS($\hat{\beta}_1$) promedio	DS empírico de $\hat{\beta}_1$	$\hat{\beta}_1$ promedio	DS($\hat{\beta}_1$) promedio	$\hat{\beta}_1$ promedio	DS($\hat{\beta}_1$) promedio
1	5	1.13	93.50	2.43	1.41	1.28	1.31	1.14
	30	0.80	0.31	0.31	0.83	0.31	0.81	0.31
4	5	1.28	28.10	3.63	2.06	2.72	1.07	1.35
	30	0.80	12.26	0.76	0.86	0.44	0.81	0.45

[(a)]half Cauchy (0, 2) cuando c=1 y half Cauchy (0, 8) cuando c=4.
[(b)]Inversa Gamma (3, 2) cuando c=1 y Inversa Gamma (258, 4112) cuando c=4.

Respecto a la estimación del σ_b^2 es notorio el previsto buen comportamiento de la Inversa Gamma informativa (Tabla 4). Si se utiliza la prior “half Cauchy”, para c=1, las estimaciones presentan mejor comportamiento que cuando se elige la distribución a priori Uniforme y resultados similares a los obtenidos con la Gamma Inversa no informativa. Cuando c=4, las estimaciones obtenidas con esta prior débilmente informativa presentan sesgos intermedios entre los que se observan para la distribución a priori Uniforme y la Inversa Gamma difusa salvo para el escenario n=10 y k=5 donde la estimación obtenida a través de la Inversa Gamma es la más sesgada de las tres (Tabla 2 y Tabla 4). Para los escenarios analizados no se logra la mejoría deseada, especialmente para el caso c=4 donde tanto las estimaciones bayesianas como clásicas presentan sesgos muy grandes sobre todo para k=5.

Nótese que se eligieron como estimaciones de los parámetros las medias de las distribuciones a posteriori cuando en realidad, dada la asimetría de las mismas, sería más conveniente considerar la mediana o el modo. Por ejemplo en el escenario c=4, n=10 y k=30, el promedio de las medias de la distribución a posteriori de σ_b^2 bajo la prior “half Cauchy” fue de 44.83 mientras que el promedio de las medianas se reduce a 27.33 y el modo se encuentra entre 10 y 20. Estas medidas estimarían de manera más deseable el valor teórico igual a 16.

Tabla 4

Estimación MV y Bayesiana de σ_b^2 a través de 1000 muestras simuladas para muestras de 10 grupos según número de individuos por grupo (k) y valores de c

c	k	Estimación MV			Estimación Bayesiana			
					Prior para σ half-Cauchy[(a)]		Prior para σ^2 Inversa Gamma informativa[(b)]	
		$\hat{\sigma}_b^2$ promedio	DS($\hat{\sigma}_b^2$) promedio	DS empírico de $\hat{\sigma}_b^2$	$\hat{\sigma}_b^2$ promedio	DS($\hat{\sigma}_b^2$) promedio	$\hat{\sigma}_b^2$ promedio	DS($\hat{\sigma}_b^2$) promedio
1	5	1.36	2.19	2.20	5.62	11.54	1.12	0.89
	30	0.91	0.58	0.63	1.636	1.335	1.00	0.53
4	5	255.95	158.13	2.325.80	729.40	1.790.4	16.04	1.00
	30	20.33	17.67	34.43	44.83	63.26	16.02	1.00

[(a)]half Cauchy (0, 2) cuando c=1 y half Cauchy (0, 8) cuando c=4.
[(b)]Inversa Gamma (3, 2) cuando c=1 y Inversa Gamma (258, 4112) cuando c=4.

5. CONSIDERACIONES FINALES

En este trabajo se estudia el comportamiento de los estimadores obtenidos por los métodos de estimación de máxima-verosimilitud y Bayes de los parámetros de un modelo logístico simple con intercepto aleatorio, ampliamente utilizado en el análisis de datos binarios correlacionados.

El estudio por simulación se realizó para dos niveles diferentes de correlación entre los datos: baja y alta. Con respecto a la estimación del coeficiente de regresión, las estimaciones máximo-verosímiles y las bayesianas son en general levemente sesgadas, cualquiera sea el nivel de correlación intragrupo, con sesgos más pronunciados para tamaños de muestra chicos. Respecto a la precisión de las estimaciones, es notoria la influencia del tamaño de la muestra en las estimaciones MV mientras que las bayesianas resultan siempre más precisas.

En relación a la componente de variancia, los resultados encontrados no fueron satisfactorios, en particular cuando la correlación es alta. Llamativamente la estimación bayesiana no proporciona buenos resultados bajo las distribuciones a priori señaladas en la bibliografía como las más apropiadas para el problema (Gelman, 2006; Austin, 2010). La mejora solamente se alcanza cuando se utiliza una distribución a priori sumamente informativa, la cual es difícilmente factible de obtener en situaciones reales.

Queda pendiente la utilización sistemática en un estudio de simulación de medidas diferentes a la media de la distribución a posterior de la componente de variancia como su estimación bayesiana, teniendo en cuenta la importante asimetría de esta distribución.

6. REFERENCIAS BIBLIOGRÁFICAS

Austin, P. C. (2010). “Estimating multilevel logistic regression models when the number of clusters is low: a comparison of different statistical software procedures”. *The International Journal of Biostatistics*, 6(1): 1-18.

Bernardo, J. M.; Smith, A.F.M. (1994). *Bayesian Theory*. John Wiley & sons.

Browne, W.J.; Draper, D. (2006). “A comparison of Bayesian and likelihood-based methods for fitting multilevel models”. *Bayesian Analysis*, 1(3): 473-514.

Dobson, A.; Barnett, A. (2008). “An Introduction to Generalized Linear Models”, 3rd ed. Chapman & Hall/CRC.

Gelman, A.(2006). “Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper)”. *Bayesian Analysis*, 1(3): 515-534.

Evans, S. R.; Hosmer, D. W. (2004). “Goodness of Fit Tests for Logistic GEE Models: Simulation Results”. *Communications in Statistics: Simulation and Computation*, 33(1): 247-258.

Fitzmaurice, G.; Laird, N.; Ware, J. (2004). *Applied longitudinal analysis*. John Wiley & Sons.

Goldstein, H. (2003). *Multilevel Statistical Models*, 3rd ed. Kendall’s Library of Statistics. London.

SAS Institute, Inc. (2011). SAS/STAT User’s guide, version 9.3. Cary, NC, USA.

Song, P. X. K. (2007). *Correlated data analysis: modeling, analytics and applications*. Springer, New York.