

Estimación en modelos lineales mixtos con datos continuos usando transformaciones y distribuciones no normales.

Marisol Valencia Cárdenas

Trabajo presentado para optar al título de Magister en Estadística.

Director: Juan Carlos Salazar Uribe,

Ph.D University of Kentucky, Profesor Asociado, Universidad Nacional de Colombia

Escuela de Estadística

Facultad de Ciencias

Universidad Nacional de Colombia

Medellín

Año 2010



Agradecimientos

Al profesor Juan Carlos Salazar Uribe por su valiosa y acertada dirección.

Al profesor Juan Carlos Correa Morales, por su asesoría y valiosa ayuda.

A todos quienes fueron apoyo para realizar este trabajo.

Contenido

Resumen	9
Abstract	10
1. Introducción	1
1.1. Planteamiento del Problema	2
1.2. Aportes del trabajo	4
2. Marco Teórico	6
2.1. Estado del arte	6
2.2. Efectos del modelo lineal mixto	8
2.2.1. Modelo lineal mixto	8
2.2.2. Varianza en el modelo lineal mixto, considerando la estructura de simetría compuesta	12
2.2.3. Estimación de los efectos β y b bajo normalidad	14
2.2.4. Prueba de normalidad para residuales	16
2.2.5. Transformación de variables en el modelo lineal mixto	17
2.2.6. Familia de distribuciones asimétricas	18
2.2.7. Aspectos importantes de la familia de distribuciones asimétricas	19
2.2.8. Estimación del modelo lineal mixto con distribución normal asimétrica en la componente aleatoria	21
2.2.9. Sesgo relativo	22

3. Metodología propuesta	23
3.1. Evaluación de la normalidad de los residuales del modelo	25
3.2. Transformación	28
3.3. Ejemplos de efecto aleatorio con características sesgadas	28
4. Simulación con error normal	31
4.1. Efecto sobre la estimación de parámetros del modelo	33
4.1.1. Error y efecto con distribución normal	33
4.1.2. Distribución normal para el error, variación en la distribución del efecto aleatorio	34
5. Combinación de distribuciones para el error y el efecto aleatorio	40
5.1. Efecto sobre la estimación de los parámetros del modelo lineal mixto	40
5.2. Diagnósticos de los residuales del modelo lineal mixto	48
5.2.1. Normalidad de error total	48
5.3. Normalidad en efecto aleatorio y el error crudo	53
5.3.1. Resultados de las simulaciones	56
5.4. Problemas detectados	58
6. Alternativas de estimación en el modelo lineal mixto con respuesta continua	59
6.1. Uso de transformaciones	60
6.2. Estimación del sesgo relativo con el uso de algoritmo EM	63
6.3. Propuesta de una prueba de normalidad sesgada	64
6.3.1. Aplicación	69
7. Conclusiones y Recomendaciones	73
ANEXOS	78

Índice de Tablas

3.1. Coeficientes fijos del modelo lineal mixto para la variable respuesta Densidad Celular.	30
3.2. Asimetría para los residuales del modelo mixto sobre Densidad Celular.	30
4.1. Parámetros controlados en el modelo.	32
4.2. Comportamiento del sesgo relativo con b_i y $\varepsilon \sim N(0, \sigma^2)$	34
4.3. Comportamiento del sesgo relativo cuando $b_i \sim SN(0, \sigma_b^2 = 50, \lambda)$	35
4.4. Comportamiento del sesgo relativo cuando $b_i \sim SN(0, \sigma_b^2, \lambda)$ variando σ_b^2	36
4.5. Modelo de regresión para la respuesta sesgo de β_0 con error normal.	37
5.1. Comportamiento de sesgo de β_0 para diferentes varianzas y distribuciones.	43
5.2. Comportamiento de sesgo de β_0 cuando $b_i \sim Normal$	44
5.3. Comportamiento de sesgo de β_0 cuando $\varepsilon \sim N(0, \sigma^2)$	46
5.4. Modelo de regresión para explicar el comportamiento del sesgo de β_0	48
5.5. Diagnósticos de normalidad para residuales totales del modelo mixto.	51
5.6. Diagnósticos de normalidad para residuales totales del modelo mixto.	52
5.7. Diagnósticos de normalidad, caso $b_i \sim Normal, \varepsilon \sim SN$	54
5.8. Diagnósticos de normalidad, $b_i \sim SN, \varepsilon \sim Normal$	55
5.9. Diagnósticos de normalidad, $b_i \sim SN, \varepsilon \sim SN$	56
5.10. Diagnósticos de normalidad para residuales totales del modelo mixto.	57
6.1. Resultados al realizar la transformación sobre la variable respuesta (1).	60
6.2. Resultados al realizar la transformación sobre la variable respuesta (2).	61
6.3. Resultados al realizar la estimación con el algoritmo EM.	64

6.4. Prueba de diferencia de proporciones.	67
6.5. Diferentes escenarios para la prueba de diferencia de proporciones.	68
6.6. Cambio en los coeficientes del modelo lineal mixto para la respuesta Densidad Celular. . .	71
6.7. Tres estimaciones de los parámetros del modelo lineal mixto para la Densidad Celular. . .	72

Índice de Figuras

2.1.	Sesgo de la función normal asimétrica, caso univariado.	20
2.2.	Sesgo de la función normal asimétrica, caso bivariado.	21
4.1.	Sesgo de β_0 cuando es igual a 10 y a 100 vs σ_b^2	36
4.2.	Sesgo de β_0 vs varianza de b_i y por parámetro λ	38
5.1.	Sesgo de β_0 vs varianza de b_i cuando $\beta_{0i} = 10$	41
5.2.	Sesgo de β_0 vs varianza de b_i cuando $\beta_{0i} = 100$	42
5.3.	Sesgo de β_0 vs σ_b^2 y distribución del efecto. 1). $b_i \sim N$, 2). $b_i \sim SN$, 3). $b_i \sim ST$	45
5.4.	Sesgo de β_0 vs σ_b^2 y distribución del error.	46
5.5.	Normalidad del error total vs σ_b^2 cuando $b_i \sim N(0, \sigma_b^2)$, 1) $\varepsilon \sim SN$, 2) $\varepsilon \sim ST$	49
5.6.	Normalidad del error total vs σ_b^2 cuando $b_i \sim ST(0, \sigma_b^2, \lambda, \nu)$, 1) $\varepsilon \sim N$, 2) $\varepsilon \sim SN$, 2) $\varepsilon \sim ST$	50
5.7.	QQplot de error y efectos aleatorios, $b_i \sim Normal$, $\varepsilon \sim SN$	53
5.8.	QQplot de error y efectos aleatorios, $b_i \sim SN$ y $\varepsilon \sim Normal$	54
5.9.	QQplot de error y efectos aleatorios, caso SN.	55
6.1.	Comportamiento del error vs parámetro de sesgo λ	62
6.2.	Comportamiento normalidad del error total vs distribución del efecto.	63
6.3.	Gráfico de probabilidad con bandas de la normal sesgada para el efecto aleatorio.	65
6.4.	Gráfico de probabilidad con bandas de la normal sesgada para el error crudo.	66
6.5.	Gráfico de probabilidad con bandas de la normal para el error total.	66
6.6.	Gráfico de probabilidad con bandas de la normal sesgada para el error total.	67
6.7.	Gráfico de probabilidad normal y normal sesgada para residuales.	70

6.8. Gráfico de probabilidad normal y normal sesgada para residuales. 71

Resumen

Los modelos lineales mixtos tienen una amplia aplicación para la estimación de efectos fijos en estudios que involucran datos correlacionados (Laird N.M. y Ware J., 1982, Zhang y Davidian, 2001). Usualmente, estos modelos se basan en el supuesto de que los efectos aleatorios y los errores son independientes y se distribuyen normalmente; sin embargo, este último supuesto no siempre se satisface en situaciones prácticas. Varios autores han analizado los supuestos de normalidad de los efectos aleatorios. Lange y Ryan (1989), proponen un método gráfico para la detección de normalidad. Otro trabajo que ilustra los problemas de este supuesto es el de Verbeke y Lesaffre (1996), que describe cómo los efectos aleatorios son estimados deficientemente cuando provienen de mezclas de distribuciones gaussianas. Las técnicas para mejorar u obtener normalidad de los efectos aleatorios y el error puro han recibido poca atención. Al respecto, la técnica de Box-Cox para transformar la variable respuesta ha sido explorada por Gurka y otros (2006), aportando mejoras al modelo pero con algunos inconvenientes en las componentes de varianza. En este trabajo se propone estudiar metodologías alternativas para mejorar el proceso de estimación de parámetros dentro de un modelo lineal mixto cuando el supuesto de normalidad para el vector de efectos aleatorios es cuestionable. Específicamente, se pretende estudiar el comportamiento de los estimadores de efectos fijos y los residuales del modelo lineal mixto asumiendo que los efectos y errores aleatorios siguen una distribución no normal, tal como la normal sesgada expuesta en Azzalini y Capitanio (1999), y la t-sesgada. Vía simulación, se estudiarán alternativas para la estimación de los efectos fijos del modelo lineal mixto: usando transformaciones de normalidad (Gurka y otros, 2007), y adaptando el algoritmo EM realizado por Arellano y otros (2005), cuando el efecto y el error aleatorio siguen distribuciones sesgadas. También se exponen propuestas de diagnósticos del modelo, incluyendo un gráfico de probabilidad con bandas bajo la distribución normal sesgada y una prueba de diferencia de proporciones que es útil en la aplicación de datos longitudinales mostrada.

Palabras clave: Modelos lineales mixtos, Distribución Normal Sesgada, Transformaciones de Normalidad, Datos Longitudinales

Abstract

Linear mixed models are useful to estimate fixed effects in studies involving correlated data (Laird N.M. y Ware J., 1982, Zhang y Davidian, 2001). Usually, these models are based on the assumption that both the random effects and the errors are independent and normally distributed, nevertheless, this last assumption is not always satisfied in practice. Some authors have studied the normality assumptions of the random effects, one of these Lange and Ryan (1989), propose a graphical method for normality detection. Other studies show the problems that arise with this assumption of normality of the random effects. The work of Verbeke y Lesaffre (1996), describes how random effects are poorly estimated when they come from Gaussian distribution mixtures. Techniques for improving or obtaining normality of the random effects and the pure error have received few attention, in this sense, the Box-Cox technique used to transform the response variable has been explored by Gurka y otros (2006); this study contributes to improve the model but it presents some pitfalls in the variance components. In this work we propose to study alternative methodologies to improve the parameters estimation process in a linear mixed model when the normality assumption for the random effect and residuals component is questionable. Specifically using simulation, we're going to study alternatives to estimate the fixed effects of the linear mixed models: using transformations (Gurka y otros, 2007), and adapting the EM algorithm developed by Arellano et all (2005), assuming that the random effects follow a skew distribution (Azzalini y Capitanio, 1999). It is also exposed diagnostic tools to study the behavior of normality in residuals, including a probability plot with intervals that follow the skew normal distribution and a proportion test useful in the application of longitudinal data showed.

Key words: Linear Mixed Models, Skew Normal Distribution, Normality Transformations, Longitudinal Data

Capítulo 1

Introducción

En la práctica es frecuente encontrar fenómenos cuya distribución de probabilidad puede reflejar con facilidad un comportamiento normal, especialmente cuando los tamaños de muestra aumentan. Sin embargo, y también con mucha frecuencia, se presentan otras situaciones que reflejan no normalidad, por ejemplo, en diversos casos en el análisis de los residuales de un modelo lineal. Este fenómeno puede ocurrir cuando se presentan alejamientos no normales en la distribución y en algunas ocasiones el uso de transformaciones en la variable respuesta podrían mejorar esta situación, verificando los supuestos originales.

Cuando se presentan variables de respuesta cuya distribución tiene un comportamiento sesgado, que pretenden ser explicadas mediante un modelo lineal para datos correlacionados temporal o espacialmente, los residuales pueden mostrar el tipo de alejamientos que presenta la variable respuesta, haciendo que los supuestos originales del modelo no se cumplan, lo que lleva a búsquedas de alternativas para mejorar las inferencias, como el uso de transformaciones en la variable respuesta, o en las covariables del diseño, situaciones en las cuales puede suceder que algunas covariables de interés modifiquen su verdadero efecto, pero además, que no se encuentre normalidad en la distribución de los errores, aspectos que no han sido ampliamente explorados.

Específicamente, el problema que se quiere estudiar en este trabajo es el de la estimación de

efectos fijos del modelo lineal mixto cuando el supuesto de normalidad de la componente aleatoria no es plausible. Este problema ha sido estudiado desde diversas ópticas, por ejemplo, Zhou y He (2007), proponen un proceso de estimación de tres pasos para encontrar los efectos de un modelo lineal mixto bajo el supuesto de distribución t-sesgada de los errores y el efecto aleatorio, y Verbeke y Lesaffre (1996), analizan las estimaciones cuando los efectos aleatorios provienen de mezclas de distribuciones Gaussianas .

Gurka y otros (2006), utilizan la transformación de Box-Cox, obteniendo dentro de sus resultados, que el supuesto de normalidad es más razonable que si no se realizara transformación alguna, sin embargo, muestran que la estimación bajo transformación de la variable respuesta lleva a estimaciones sesgadas de las componentes de varianza. Así mismo, Gurka y otros (2007), estudian pruebas de hipótesis sobre el parámetro de la transformación λ de Box-Cox (1964).

1.1. Planteamiento del Problema

Es necesario entonces establecer estimaciones más robustas en el modelo lineal mixto con un tipo de distribución diferente a la normal, por ejemplo, una con característica sesgada de la componente aleatoria. Una familia que reúne estas características es la normal sesgada (Azzalini, A. y Capitanio, A., 1999). Por ello, es importante determinar los efectos que puede generar la procedencia de distribuciones diferentes de la normal, para el error y el efecto aleatorio en las estimaciones de este modelo lineal mixto.

Estudiar el comportamiento distribucional del efecto aleatorio es importante ya que, además de explicar una variabilidad extra que no es tenida en cuenta por un modelo de sólo efectos fijos, afecta la distribución del error y además permite la "individualización" del modelo, aspecto que no puede estudiarse en el modelo lineal que tiene sólo efectos fijos.

Como objetivos de este trabajo se encuentran: evaluar la calidad de las estimaciones de los efec-

tos fijos de un modelo lineal mixto cuando el error y el efecto aleatorio, provienen de distribuciones no normales, determinando el comportamiento distribucional de los residuales, y determinar una posible alternativa que mejore la estimación del efecto fijo del modelo lineal mixto.

Entre los objetivos específicos se pretende comparar, vía simulación, el efecto en las estimaciones de los efectos fijos de un modelo lineal mixto y distribuciones sesgadas en el escenario de diseños balanceados, evaluar cuando es posible el uso de transformaciones de normalidad de la respuesta, o qué alternativa puede generar mejores resultados, y finalmente ilustrar las metodologías usando datos reales.

En cada uno de los siguientes capítulos se ilustrará el procedimiento utilizado, de manera que en el capítulo 2 se hace referencia al marco teórico y estado del arte. En el capítulo 3 se explica la metodología usada en las simulaciones, además de la descripción de casos reales donde ocurre el sesgo en este tipo de modelos lineales mixtos. En el capítulo 4 se describen los resultados cuando se fija la distribución normal para el error del modelo. En el capítulo 5 se combinan las distribuciones normal, normal sesgada y t-sesgada para ambos: error y efecto aleatorio del modelo lineal mixto. En este último fue estimado un modelo de regresión lineal, donde se encontró una asociación lineal significativa entre la variable respuesta "sesgo relativo del intercepto estimado" con las covariables: varianza del efecto aleatorio, la distribución de la cual proviene tanto el error como el efecto aleatorio, así como el parámetro de forma o sesgo y el valor de éste, sin embargo, no fue alta la incidencia del número de sujetos. En el capítulo 6 se describen los resultados cuando es posible utilizar una transformación de Box Cox, además, se incorpora la prueba de comportamiento normal sesgado realizada en el trabajo y se ilustra el resultado de la estimación con el algoritmo EM (Maximización de la Esperanza) propuesta por Arellano y otros (2005) (presentado en los anexos), quienes desarrollan la estimación del modelo lineal mixto asumiendo que los errores y efecto aleatorio provienen de una distribución sesgada. Adicional a los resultados mostrados, se usa una prueba de normalidad sesgada para los residuales y el efecto aleatorio mostrando los escenarios y aplicación donde se vé su correcta funcionalidad para mostrar un comportamiento sesgado.

Las transformaciones son útiles en limitados escenarios, y además, dada la mezcla de diferentes distribuciones, se ve la necesidad de realizar otro tipo de metodologías para la estimación, como el algoritmo EM, adaptado para el modelo estudiado en este trabajo.

1.2. Aportes del trabajo

Los principales aportes de este trabajo son:

- En un modelo lineal mixto para medidas repetidas con un intercepto aleatorio, los residuales totales están conformados por dos componentes: el error crudo y el efecto aleatorio, luego, dichos residuales son una mezcla de diferentes variables que pueden generar una con comportamiento no determinado. Por esta razón, se considera que la metodología propuesta para realizar el diagnóstico de normalidad de forma separada para el error crudo y el efecto aleatorio, además del error total constituye un aporte de este trabajo.

Al realizar el diagnóstico del error total, se esperaba que no tuviera un comportamiento normal cuando en la mezcla de variables una de ellas o ambas no tienen un comportamiento normal, a pesar de esto, se encontró un buen diagnóstico en la normalidad de los residuales del modelo lineal mixto al fijar la distribución normal para el efecto aleatorio y la distribución normal sesgada para el error, caso en el cual se afectó el sesgo del intercepto cuando aumentó la varianza.

Contrario a lo anterior, se identificó la violación en el supuesto de normalidad del modelo lineal mixto en todos sus componentes al fijar las distribuciones del error y efecto aleatorio en combinaciones donde está presente la distribución t-sesgada, y con un valor alto de la varianza del efecto aleatorio, además ocurre con menor frecuencia cuando el efecto tiene la distribución normal sesgada y el error es normal con alta varianza del modelo.

- La aplicación del gráfico de probabilidad con bandas normales sesgadas, y la prueba de diferencia de proporciones, en la cual se detecta el comportamiento normal sesgado de los residuales totales.

- La estimación robusta de los parámetros fijos del modelo lineal mixto vía el algoritmo EM, aplicando el trabajo de Arellano y otros (2005), dejando ver una adecuada forma de resolver el problema del sesgo en los parámetros fijos del modelo lineal mixto.

Capítulo 2

Marco Teórico

2.1. Estado del arte

El problema de considerar distribuciones alternativas a la normal para el efecto aleatorio en un modelo lineal mixto, así como el problema de violación de normalidad, son aspectos abordados por diferentes autores tal y como se evidencia en la literatura. A continuación se presentan apartes de algunos de estos trabajos.

- Lange y Ryan (1989) analizan modelos lineales mixtos desbalanceados. Describen un método para chequear los supuestos de normalidad en los errores y en particular, del efecto aleatorio en el modelo y describen la detección de desviaciones de la normalidad comparando las funciones de distribución acumuladas empíricas y teóricas de combinaciones lineales de residuales estimados para cada individuo.
- Verbeke y Lesaffre (1996) investigan el impacto del supuesto de normalidad en las estimaciones de los efectos aleatorios, comparando con el modelo gaussiano cuando los efectos provienen de mezclas de distribuciones normales o gaussianas, y muestran que los estimadores de máxima verosimilitud de los parámetros de efectos fijos son consistentes y asintóticamente normales sin importar cual sea la distribución de los efectos aleatorios.

- Fernández y Steel (1999) analizan la distribución multivariable con errores comunes al usar métodos bayesianos basados en la distribuciónn T. Además, hacen énfasis en dificultades encontradas en el modelo de regresión cuando los errores se asumen con una distribución T con grados de libertad no específicos.
- Piepho y McCulloch (2004) evalúan un tratamiento en una aplicación para el análisis de riesgo en un ensayo multiambiental, utilizando modelos lineales mixtos asumiendo que los ambientes son una muestra aleatoria en la cual los términos aleatorios del modelo están normalmente distribuidos. En el artículo se sugiere una transformación basada en el Sistema de Johnson. Vía simulación, mostraron que el sistema de Johnson se ajusta adecuadamente para modelar problemas de no normalidad.
- Arellano y otros (2005) presentan una metodología de estimación del modelo lineal mixto cuando el efecto y error aleatorios siguen distribuciones diferentes a la normal, como la normal sesgada o t sesgada, utilizando el algoritmo EM (Maximización de la Esperanza).
- Gurka et al. (2006), proporcionan métodos de exploración sobre transformaciones en el modelo lineal mixto para datos longitudinales. Utilizan la transformación de Box-Cox, realizando simulaciones para la estimación del parámetro λ usado en ésta (Box-Cox., 1964), y determinan que el supuesto de normalidad del modelo luego de aplicarla es más razonable que si no se realizara transformación alguna. Además muestran que la estimación bajo transformación de un modelo lineal mixto lleva a sesgos en las componentes de varianza, indicando que debe realizarse un mejoramiento en las estimaciones relativas al efecto aleatorio. También analizan la influencia de outliers, a partir de la construcción de diagnósticos apropiados.
- Gurka et al. (2007), estudian pruebas de hipótesis sobre el parámetro de la transformación y el efecto de los outliers en la inferencia sobre este parámetro. Los autores usan gráficos cuantil-cuantil para comparar los cambios en los residuales del modelo de acuerdo a la transformación.
- Zhou y He (2007), proponen un método de estimación basado en tres pasos para el modelo lineal mixto con el supuesto de distribución t-asimétrica de los errores y el efecto aleatorio.

- Jara et al (2008), realizan una aproximación bayesiana con distribuciones elípticas asimétricas bajo las distribuciones: normal, normal asimétrica y T de student y las aplican a un ejemplo de datos longitudinales.
- Banerjee y Frees (2009), proponen diagnósticos para efectos fijos y aleatorios en modelos longitudinales, basados en estadísticos parciales influenciales.

2.2. Efectos del modelo lineal mixto

Dentro de los aspectos del proceso de estimación de un modelo lineal mixto se contemplan los efectos fijos y aleatorios. Estos y otros conceptos, serán definidos de forma general en esta sección.

2.2.1. Modelo lineal mixto

Previamente a la definición del modelo matemático, se darán unos conceptos generales sobre componentes del modelo lineal mixto.

Efecto aleatorio: Puede verse como aquel que se relaciona con situaciones donde una unidad es tomada al azar de una población caracterizada por varios niveles. Existen diferentes razones para incorporar este efecto dentro de un modelo lineal (Correa J., 2008):

- Cuando no se está interesado en estimar su efecto localizado sino su componente de varianza.
- Cuando se desea marginalizar de manera que las unidades compartan dicho efecto.

Ejemplos: Humedad relativa, cuya medición tiene el efecto aleatorio de un punto a otro en la atmósfera. Otro es la altura que se mide a un sujeto en diferentes años, cuyo un efecto aleatorio es variante de un año al otro.

Efecto fijo: Algunas veces la función de la respuesta media involucra efectos de factores que son tratados como constantes conocidas o controladas y que representan los efectos fijos, además los niveles comprendidos en el estudio son todos los posibles niveles del factor, la inferencia se realiza

sobre todos los niveles y por ello, este efecto queda completamente determinado. Además, para el efecto fijo, es de interés determinar su magnitud, y algunas veces se condicionan en la presencia de otros factores no fijos.

Por ejemplo, se desea determinar el efecto de 5 drogas existentes para la presión sanguínea. Por ello, las conclusiones se restringen a éstas. Otro ejemplo: cuando se desea estimar el efecto de incorporar una mezcla, dada la temperatura. En este caso la temperatura es el efecto fijo.

Teniendo en cuenta estos conceptos, se puede presentar la forma general del modelo lineal mixto (Lange y Ryan, 1989):

$$\begin{array}{ccccccc}
 y & = & \mathbf{X} & \beta & + & \mathbf{Z} & b & + & \varepsilon \\
 T \times 1 & & T \times p & p \times 1 & & T \times Nr & Nr \times 1 & & T \times 1
 \end{array}$$

- El vector y tiene como componentes a y_{ij} con $i = 1, \dots, N$, $j = 1, \dots, n_i$, donde n_i es la cantidad de medidas por sujeto y N es el total de sujetos. Además $T = \sum_{i=1}^N n_i$. [p es el número de covariables de la matriz de diseño \mathbf{X} más 1, r es el número de covariables de la matriz de diseño \mathbf{Z} más 1 (intercepto aleatorio), asociadas al componente aleatorio.
- \mathbf{X} es la matriz de covariables de diseño que relaciona el vector de parámetros β con la variable respuesta y .
- \mathbf{Z} es la matriz de diseño de los efectos aleatorios que relaciona al vector b con la variable respuesta y , más precisamente,

$$\mathbf{Z} = \text{diag}(Z_1, Z_2, \dots, Z_N) \quad y \quad b = (b'_1, b'_2, \dots, b'_N)', \quad \dim(b_i) = r \times 1$$

$$\varepsilon \sim N(0, \mathbf{R})$$

$$b \sim N(0, \mathbf{B})$$

- Los b_i son independientes entre sí e independientes de los ε_i e idénticamente distribuidos.
- $\mathbf{R} = \sigma_e^2 I$ matriz diagonal de varianza del error.

- \mathbf{B} es matriz diagonal de varianza del efecto aleatorio.

El objetivo al implementar un análisis basado en modelos lineales mixtos es la estimación de los β 's y la predicción de los valores de los b_i 's, además de estimar las componentes de varianza.

El modelo dado por la expresión: $Y = \mathbf{X}\beta + \mathbf{Z}b + \varepsilon$, puede expresarse de manera más detallada, así, para el sujeto i :

$$\begin{bmatrix} y_{i1} \\ \dots \\ y_{ij} \\ \dots \\ y_{in_i} \end{bmatrix} = \begin{bmatrix} 1 & x_{i11} & \dots & x_{i1k} \\ 1 & x_{i21} & \dots & x_{i2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{ij1} & \dots & x_{ijk} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \dots & x_{in_i k} & \dots \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} 1 & z_{i11} & \dots & z_{i1r-1} \\ 1 & z_{i21} & \dots & z_{i2r-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & z_{ij1} & \dots & z_{ijr-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \dots & z_{in_i r-1} & \dots \end{bmatrix} \begin{bmatrix} b_{i0} \\ b_{i1} \\ \vdots \\ b_{ij} \\ \vdots \\ b_{ir-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{ij} \\ \vdots \\ \varepsilon_{in_i} \end{bmatrix}$$

Donde el i ésimo sujeto de y tiene n_i observaciones: $(y_{i1}, y_{i2}, \dots, y_{in_i})$ con: $i = 1, \dots, N$, N es el número de sujetos.

La matriz \mathbf{Z} puede expresarse como una matriz de bloques conformada por las submatrices \mathbf{Z}_i , con $i = 1, \dots, N$.

$$\mathbf{Z}_{T \times Nr} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_N \end{bmatrix}$$

Aquí, cada \mathbf{Z}_i tiene dimensión $n_i \times r$. Si el modelo es balanceado definiendo, $n_i = k$, se tiene que $T=Nk$, así el orden de \mathbf{Z} será $Nk \times Nr$.

Las matrices diagonales de varianza-covarianza \mathbf{R} y \mathbf{B} , del error y del efecto aleatorio respectivamente, pueden expresarse así:

$$\mathbf{R}_{T \times T} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_N \end{bmatrix} \quad \mathbf{B}_{Nr \times Nr} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_N \end{bmatrix}$$

Donde R_i es la matriz de orden $n_i \times n_i$, y si el diseño es balanceado podemos decir que: $n_i = k$, luego, la matriz R_i sería de orden $k \times k$. La matriz D_i es de orden $r \times r$. Usualmente la estructura de varianza-covarianza se define a partir de la matriz D_i .

Un modelo lineal mixto que sólo considera intercepto aleatorio, y que se asume con un componente de varianza de simetría compuesta, puede ser expresado de esta forma: $y_{ij} = \beta_0 + x_{ij}\beta_1 + b_i + \varepsilon_{ij}$
Con:

- y_{ij} es la respuesta del sujeto i en la medida j , donde $i=1, \dots, N$, N es el número de sujetos, $j = 1, \dots, n_i$, n_i representa la cantidad de medidas por sujeto.
- x_{ij} representa la covariable de la matriz de diseño \mathbf{X} de los efectos fijos para el sujeto i en la medida j , con $i = 1, \dots, N$ y $j = 1, \dots, n_i$.
- β_0 y β_1 son los efectos fijos del intercepto y la covariable x_{ij} respectivamente.
- b_i representa el intercepto aleatorio en el modelo.
- ε_{ij} es el término de error para el sujeto i en la medida j .

Cuando se asume que la componente de varianza es de simetría compuesta, dos parámetros de varianza σ_e^2 y σ_b^2 son necesarios para obtener una estimación del coeficiente de correlación intraclase ρ (que se define más abajo).

- Componentes de varianza de simetría compuesta

$$\begin{aligned} \sigma_b^2 &: \text{Varianza entre medidas del mismo sujeto,} \\ \sigma_e^2 &: \text{Covarianza entre sujetos,} \\ \sigma^2 = \sigma_e^2 + \sigma_b^2 &: \text{Componente de covarianza.} \end{aligned}$$

- El coeficiente de correlación intraclase (CCI) en la estructura de simetría compuesta está dado por la expresión:

$$\rho = \frac{\sigma_b^2}{\sigma_e^2 + \sigma_b^2}$$

El CCI representa la correlación entre 2 medidas repetidas para el mismo sujeto.

Con respecto a esta componente, Verbeke y Molenberghs (2001), menciona que cuando la componente de varianza en un modelo lineal mixto es negativa, tiene interpretaciones significativas en el modelo marginal, lo cual indica que el supuesto de correlación constante positiva no es válido para los datos. Es decir, no siempre es válido considerar este supuesto en los datos.

La forma de estructurar la varianza del modelo mixto, y su relación con la simetría compuesta que será utilizada en este trabajo, se describe a continuación.

2.2.2. Varianza en el modelo lineal mixto, considerando la estructura de simetría compuesta

En esta sección se presentarán los elementos de la descomposición de la varianza total del modelo lineal mixto, describiendo la parte relativa al efecto aleatorio y el error del modelo, cuando se asume un modelo balanceado y $r = 1$, esto es, se asumirá una sola componente aleatoria que en este caso es el intercepto aleatorio. Además $n_i = k$ representa el número de repeticiones de cada clase o sujeto, N es la cantidad de sujetos, y b es el efecto aleatorio.

Ahora, se inicia con la formulación matemática de la varianza del modelo lineal mixto \mathbf{V} (ver McCulloch C.E. and Searle S.R. (2000)).

$$\mathbf{V} = V(E[y|b]) + E[V(y|b)].$$

Sabiendo además que $E[y|b] = \mathbf{X}\beta + \mathbf{Z}b$, y que $\mathbf{V} = V(\mathbf{X}\beta + \mathbf{Z}b + \varepsilon) = \mathbf{R}$,

$$\mathbf{V} = V(\mathbf{X}\beta + \mathbf{Z}b) + E[\mathbf{R}].$$

$$\mathbf{V} = V(\mathbf{Z}b) + \mathbf{R}.$$

Como $V(b) = \mathbf{D}$, entonces

$$\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R}.$$

La matriz Z_i asociada al sujeto i , para el modelo con intercepto aleatorio, se reduce a un vector de 1's denotado: $\mathbf{1}_k$. Luego, como lo establece McCulloch (2000), la matriz \mathbf{Z} puede expresarse como el producto Kronecker:

$$\mathbf{Z} = \mathbf{I}_N \otimes \mathbf{1}_k,$$

donde k es el número de repeticiones por clase (balanceado), y N es el número de clases (o sujetos). Luego, la matriz \mathbf{V} puede expresarse como el producto Kronecker:

$$\mathbf{V} = (\mathbf{I}_N \otimes \mathbf{1}_k)(\mathbf{D} \otimes \mathbf{1}_k)(\mathbf{I}_N \otimes \mathbf{1}'_k) + \mathbf{R}.$$

Denotando por $\mathbf{J}_k = (\mathbf{I}_N \otimes \mathbf{1}_k)(\mathbf{I}_N \otimes \mathbf{1}'_k)$, y así la matriz \mathbf{V} , puede expresarse:

$$\mathbf{V} = \mathbf{D} \otimes \mathbf{J}_k + \mathbf{R}.$$

Además,

$$\mathbf{D} = \sigma_b^2[(1 - \rho_b)\mathbf{I}_N + \rho_b\mathbf{J}_N].$$

De esta forma,

$$\mathbf{V} = \sigma_b^2[(1 - \rho_b)\mathbf{I}_N + \rho_b\mathbf{J}_N] \otimes \mathbf{J}_k + \mathbf{R}.$$

Expresando la matriz \mathbf{R} como $\mathbf{R} = \mathbf{I}_N \otimes \mathbf{R}_o$, $\mathbf{R}_o = \sigma^2\mathbf{I}_k$, se obtiene:

$$\mathbf{V} = \sigma_b^2[(1 - \rho_b)\mathbf{I}_N + \rho_b\mathbf{J}_N] \otimes \mathbf{J}_k + \mathbf{I}_N \otimes \sigma^2\mathbf{I}_k.$$

Por la estructura de simetría compuesta, $\rho_b = 0$, ya que sólo es admisible la correlación intra sujetos y sacando factor común I_N se tiene:

$$\begin{aligned}\mathbf{V} &= \sigma_b^2[\mathbf{I}_N] \otimes \mathbf{J}_k + \sigma^2\mathbf{I}_N \otimes \mathbf{I}_k, \\ \mathbf{V} &= \mathbf{I}_N \otimes [\sigma_b^2\mathbf{J}_k + \sigma^2\mathbf{I}_k].\end{aligned}$$

El orden de la matriz \mathbf{V} es $kN \times kN$. Luego, para un sujeto, $\mathbf{V}_i = \sigma_b^2\mathbf{J}_k + \sigma_e^2\mathbf{I}_k$, de orden $k \times k$. Así, es posible reescribir la correlación intraclase en el modelo con simetría compuesta de esta manera:

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}$$

Al expresar cada \mathbf{V}_i de manera matricial (ver Verbeke, 1996) (pág 117), se tiene:

$$\mathbf{V}_i = \begin{pmatrix} \sigma_b^2 + \sigma^2 & \sigma_b^2 & \sigma_b^2 \\ \vdots & \ddots & \vdots \\ \sigma_b^2 & \cdots & \sigma_b^2 + \sigma^2 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{k \times 1} \sigma_b^2 \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix}_{1 \times k} + \sigma^2 \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}_{k \times k}$$

Como $r=1$, el orden de la matriz D_i es 1×1 , es decir, $D_i = \sigma_b^2$ y la matriz \mathbf{R}_i sería de orden $k \times k$.

2.2.3. Estimación de los efectos β y b bajo normalidad

Bajo el supuesto de normalidad e independencia del error y del efecto aleatorio, b y ε , tienen una distribución conjunta normal multivariada:

$$f(b, \varepsilon) = \frac{1}{(2\pi)^{(T+nk)/2}} \begin{vmatrix} \mathbf{B} & 0 \\ 0 & \mathbf{R} \end{vmatrix}^{-n/2} \exp \left[-\frac{1}{2} \begin{bmatrix} b \\ Y - \mathbf{X}\beta - \mathbf{Z}b \end{bmatrix}' \begin{bmatrix} \mathbf{B} & 0 \\ 0 & \mathbf{R} \end{bmatrix}^{-1} \begin{bmatrix} b \\ Y - \mathbf{X}\beta - \mathbf{Z}b \end{bmatrix} \right]$$

Donde $T = \sum_{i=1}^N n_i$. El proceso de estimación de máxima verosimilitud se basa en la minimización del exponente de esta expresión. El proceso usual para llevar a cabo la minimización del exponente, es como sigue. Sea:

$$Q = \begin{bmatrix} b \\ Y - \mathbf{X}\beta - \mathbf{Z}b \end{bmatrix}' \begin{bmatrix} \mathbf{B} & 0 \\ 0 & \mathbf{R} \end{bmatrix}^{-1} \begin{bmatrix} b \\ Y - \mathbf{X}\beta - \mathbf{Z}b \end{bmatrix}$$

La matriz de varianzas en el término de la mitad, tiene submatrices en la diagonal invertibles, por lo tanto, la inversa existe, y Q se convierte en:

$$Q = (b\mathbf{B}^{-1})'b + (Y - \mathbf{X}\beta - \mathbf{Z}b)' \mathbf{R}^{-1} (Y - \mathbf{X}\beta - \mathbf{Z}b)$$

$$Q = (b\mathbf{B}^{-1})'b + ((Y - \mathbf{X}\beta)' - (\mathbf{Z}b)') \mathbf{R}^{-1} ((Y - \mathbf{X}\beta) - \mathbf{Z}b).$$

Ahora Q se deriva con respecto a b y β :

$$\frac{\partial Q}{\partial b}, \frac{\partial Q}{\partial \beta}.$$

$$\frac{\partial Q}{\partial \beta} = (-2\mathbf{X})' \mathbf{R}^{-1} (Y - \mathbf{X}\beta) + 2\mathbf{X}' \mathbf{R}^{-1} \mathbf{Z}b,$$

luego, igualando a cero,

$$\mathbf{X}' \mathbf{R}^{-1} \mathbf{X} \hat{\beta} + \mathbf{X}' \mathbf{R}^{-1} \mathbf{Z} \hat{b} = \mathbf{X}' \mathbf{R}^{-1} Y.$$

Y de:

$$\frac{\partial Q}{\partial b} = 2\mathbf{B}^{-1}b + 2\mathbf{Z}' \mathbf{R}^{-1} \mathbf{Z}b - 2\mathbf{Z}' \mathbf{R}^{-1} (Y - \mathbf{X}\beta),$$

Al igualar a cero se encuentra que:

$$\mathbf{Z}' \mathbf{R}^{-1} \mathbf{X} \hat{\beta} + (\mathbf{Z}' \mathbf{R}^{-1} \mathbf{Z} + \mathbf{B}^{-1}) \hat{b} = \mathbf{Z}' \mathbf{R}^{-1} Y.$$

En forma matricial:

$$\begin{bmatrix} \mathbf{X}' \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}' \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}' \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}' \mathbf{R}^{-1} \mathbf{Z} + \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \mathbf{R}^{-1} Y \\ \mathbf{Z}' \mathbf{R}^{-1} Y \end{bmatrix}$$

Posteriormente, es posible y luego de otros cálculos, es posible encontrar los valores de $\hat{\beta}$ y \hat{b} :

$$\hat{b} = \mathbf{BZ}'\mathbf{V}^{-1}(Y - \mathbf{X}\hat{\beta})$$

$$\hat{\beta} = [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1} \mathbf{X}'\mathbf{V}^{-1}Y$$

Nota: Este estimador del efecto aleatorio b (\hat{b}) es el mejor predictor lineal insesgado (BLUP) para b (ver Correa, 2008, pág. 33.). Adicionalmente, McCulloch (2000) menciona que es posible usar un predictor para b , uno de ellos, el mejor predictor BP es $BP(b) = E[b|y]$, “Válido para todas las formas de distribución de probabilidad de b y y ” (pág. 92). Esto señala que para la variable aleatoria dada por el efecto aleatorio, puede estimarse un valor esperado condicionado en y , estimación que se expresa como \hat{b} .

2.2.4. Prueba de normalidad para residuales

Es importante determinar si el ajuste adecuado o no de los residuales totales a la distribución normal, puede provenir del error crudo o del efecto aleatorios, para lo cual, se utiliza la prueba de normalidad de Shapiro Wilks, para variables no correlacionadas. Por dicha razón debe eliminarse la dependencia o correlación presente entre éstos.

Para ello, se obtienen los residuales del modelo estimado con la función `lme` del paquete estadístico R^1 , usando la opción `resid(modelo,type='normalized')`, la cual usa la descomposición de Cholesky que elimina la dependencia entre los residuales dada la estructura de correlación.

Esto es, si $y \sim N(\mu, \Sigma)$, la descomposición de este tipo para Σ es $\Sigma^{1/2} * \Sigma^{1/2}$, por ello, al usar la estandarización $(y - \mu) * \Sigma^{-1/2}$, se obtiene $y \sim N(0, \mathbf{I})$.

La prueba para verificar normalidad de Shapiro Wilks es entonces aplicada sobre las variables estandarizadas no correlacionadas que resultan de dicha descomposición, resultando adecuada debido a que en el modelo aquí estudiado la dimensión del componente aleatorio es univariado (1×1), por tratarse de un modelo de intercepto aleatorio.

¹R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. [Vienna, Austria]. Disponible en Web: <<http://www.R-project.org>>

Una vez realizada la prueba, puede ocurrir que no se encuentre normalidad en los residuales del modelo estimado. Una alternativa propuesta para mejorar las inferencias, es la realización de una transformación sobre la respuesta y , cuando es posible realizarla. En este trabajo se describirá la transformación de Box Cox, como se describe en la siguiente sección .

2.2.5. Transformación de variables en el modelo lineal mixto

Desde un punto de vista práctico, los métodos para remediar la no normalidad consisten en realizar transformaciones sobre las variables, y en algunos trabajos se han mostrado beneficios al utilizarlas (Box Cox, 1964, Piepho H. P., McCulloch C.E., 2004). Sin embargo, se presentan varios problemas con las transformaciones en el modelo lineal mixto, según lo afirman Gurka y otros (2006):

(i) Las transformaciones se realizan usualmente en cada componente por separado esperando obtener normalidad conjunta.

(ii) Las variables transformadas son más difíciles de interpretar, especialmente cuando se usan diferentes transformaciones para cada variable.

(iii) Una transformación en el contexto del modelo lineal mixto puede llevar a la estimación sesgada de los componentes de varianza.

Transformación de Box Cox

Box-Cox (1964) introduce una transformación para lograr la normalidad de una variable, pero además, describe el uso de otra cuando hay presencia de datos correlacionados. La transformación de Box-Cox es citada de nuevo en Gurka (2006), quienes la usan para lograr normalidad para el modelo mixto con la forma:

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^\lambda - 1}{\lambda(\tilde{y})^{\lambda-1}}, & \text{if } \lambda \neq 0 \\ \tilde{y} \log y_i, & \text{if } \lambda = 0 \end{cases}$$

Donde \tilde{y} es la media geométrica.

2.2.6. Familia de distribuciones asimétricas

La siguiente proposición introduce el concepto de una función escalar y su relación con una función densidad de probabilidad.

Proposición (Azzalini y Capitanio, 2003, pág.372). Denote por $f(y)$ la función densidad de una variable aleatoria continua d-dimensional que es simétrica alrededor de 0, y por G una función de distribución escalar tal que $G(-x) = 1 - G(x)$ para cualquier real x , si $w(y)$ es una función de \mathbb{R}^d a \mathbb{R} tal que $w(-y) = -w(y)$, para todo real $y \in \mathbb{R}^d$, entonces:

$$2f(y)G\{w(y)\} \tag{2.1}$$

es una función densidad.

Luego, la forma de la función densidad normal sesgada (SN), que fue formulada en principio por Azzalini (1985), es la siguiente:

$$f(x; \lambda) = 2g(x)G(\lambda x), \quad \lambda \in \mathbb{R} \tag{2.2}$$

Siendo g una función de densidad normal de una variable aleatoria continua d-dimensional centrada en 0, y G es función de distribución escalar normal estándar. Lo anterior resulta en una distribución sesgada, donde λ es el parámetro que controla el sesgo o la asimetría, es decir, asume el papel del parámetro de forma. En forma análoga se definen las distribuciones t asimétricas (Skew-T o ST).

En este trabajo se utiliza el término Skew Normal o SN para hacer referencia a la distribución normal sesgada y se usa Skew T o ST para la distribución t sesgada.

2.2.7. Aspectos importantes de la familia de distribuciones asimétricas

- Cuando en la ecuación (2.2), el parámetro de forma λ es cero, la función corresponde a la densidad $N(0, 1)$. (Azzallini y Dalla Valle, 1996, pág. 715.)
- Por otro lado, una variable aleatoria d-dimensional Y tiene una distribución normal asimétrica SN si es continua y tiene una función densidad del tipo (Azzalini, A. y Capitanio, A., 1999, pág. 584.):

$$2\phi_d(y - \xi; \Omega)\Phi\{\lambda^T \omega^{-1}(y - \xi)\}, \quad (2.3)$$

Donde $\Omega = \omega\Omega_c\omega$ es la matriz de varianza covarianza. Luego, $Y \sim SN_d(\xi, \Omega, \lambda)$ donde ξ, Ω, λ , son los parámetros de localización, dispersión y forma o asimetría, respectivamente.

- Sea una variable aleatoria Z , definida de la siguiente forma:

$$Z = \begin{cases} Y & \text{Si } X < w(Y) \\ -Y & \text{Si } X > w(Y) \end{cases},$$

donde Y tiene función densidad f , y X es una variable independiente con función de distribución G , independiente de Y , entonces Z tiene función densidad de la forma (2.3).

A manera de ilustración, a continuación se presentarán algunas gráficas de la distribución SN para el caso univariado y bivariado cuya programación se puede encontrar en el paquete `sn` que funciona en R^2 . Estas representaciones gráficas fueron tomadas del sitio Web:

²Disponible en www.cran.r-project.org/

www.tri.org.au/azzalini.html.

- **Gráficas de las funciones densidad normales asimétricas**

- Gráfica de la densidad normal asimétrica en el caso univariado asociada a los parámetros: Media $\mu = 0$ (localización), escala $\sigma^2 = 1$, y forma o sesgo $\lambda = 5$.

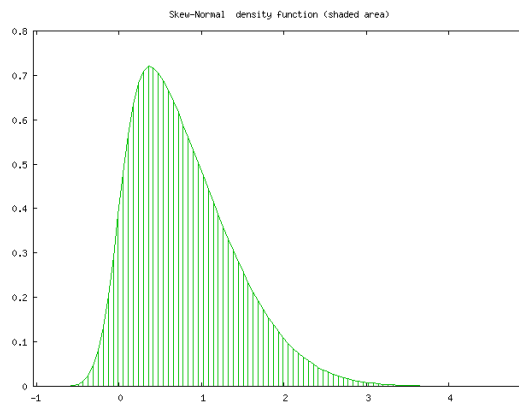


Figura 2.1: Sesgo de la función normal asimétrica, caso univariado.

En el caso univariado, la gráfica muestra un sesgo a la derecha, o alejamiento de la media de forma sesgada, lo cual indica una perturbación en la original simetría de la distribución normal.

- Gráfica de la distribución normal asimétrica bivariada asociada a los parámetros: Media $\mu = (0, 0)$, escala=(1,1), forma1 $\lambda_1 = -5$, forma2 $\lambda_2 = 2$, asociación $cov = 0,75$.

En el caso multivariado (bivariado), la campana muestra un sesgo o alejamiento del punto central, que también se refleja en la forma elíptica del contorno graficado en el plano, que muestra más lejanía en dirección del sesgo.

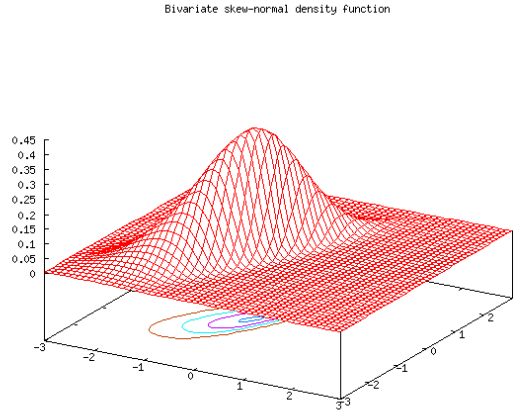


Figura 2.2: Sesgo de la función normal asimétrica, caso bivariado.

2.2.8. Estimación del modelo lineal mixto con distribución normal asimétrica en la componente aleatoria

Para la estimación por máxima verosimilitud de los efectos fijos del modelo lineal mixto, se parte de la distribución marginal del vector $(y_{i1}, \dots, y_{in_i})$, asumiendo que las distribuciones $f(y_{ij}|b_i)$ y $f(b_i)$ son funciones densidad de $y_{ij}|b_i$ y b_i respectivamente. La función de densidad marginal se puede expresar como:

$$f(y_{i1}, \dots, y_{in_i}) = \int f(y_{i1}, \dots, y_{in_i} | b_i) f(b_i) db_i,$$

y por la independencia de los y_{ij} , ésta se convierte en,

$$f(y_{i1}, \dots, y_{in_i}) = \int \left\{ \prod_{j=1}^{n_i} f(y_{ij} | b_i) \right\} f(b_i) db_i.$$

La función densidad normal asimétrica $SN_1(\xi, \sigma^2, \lambda)$, tiene la siguiente forma:

$$2\varphi(y - \xi; \sigma^2)\phi(\lambda^{-1}\sigma^{-1}(y - \xi)); \quad y, \xi, \lambda \in R,$$

donde, ξ representa el valor esperado $E[y]$ en el modelo lineal mixto, $\xi = \beta_0 + x_{ij}\beta + b_{ij}$. Y, para el efecto aleatorio, este valor esperado es igual a 0, $\xi = 0$.

2.2.9. Sesgo relativo

Cassella (2001), pg 330, plantea la siguiente definición: El error cuadrático medio (MSE) de un estimador W para un parámetro θ es una función dada por:

$$MSE = E(W - \theta)^2 = \text{var}_\theta(W) + (\text{sesgo}_\theta W)^2,$$
$$(\text{sesgo}_\theta W)^2 = E(W - \theta)^2 - \text{var}_\theta(W).$$

Y el sesgo relativo se obtendría así: $\frac{|\text{sesgo}_\theta W|}{\theta}$

En el trabajo de simulación se considerará el sesgo relativo, que se reduce a la división por el parámetro teórico fijado en la simulación, y representaría una proporción sobre éste.

Capítulo 3

Metodología propuesta

En este capítulo se describe la metodología utilizada para realizar el análisis del problema vía simulación.

Antes de realizar las simulaciones, se controlaron diferentes parámetros iniciales del modelo mixto, como son:

- La distribución del efecto aleatorio y del error aleatorio.
- El tamaño de muestra.
- El coeficiente de correlación intraclase.
- El vector de parámetros fijos del modelo: (β_0 y β_1).

Como distribuciones de probabilidad para generar tanto el efecto como el error aleatorio se usaron:

- La distribución normal.
- La distribución normal sesgada (SN).
- La distribución T sesgada (ST).

Para cada distribución se controlaron otros parámetros: el sesgo en el caso de las distribuciones sesgadas y los grados de libertad para las relacionadas con la distribución t de Student.

Adicionalmente, los cambios realizados en el coeficiente de correlación intraclase obedecen a variaciones en sus componentes:

- La varianza del efecto σ_b^2 .
- La varianza del error aleatorio σ_e^2 .

De manera que su variación hace que cambie la magnitud de la correlación intraclase.

Los comportamientos que se pretenden analizar son:

1. El sesgo relativo de los parámetros fijos resultante de la estimación del modelo.
2. El efecto causado sobre el supuesto de normalidad en las componentes del error y el efecto aleatorio del modelo en los escenarios controlados.

La cuantificación del sesgo relativo del coeficiente de correlación intraclase será observada a pesar de no ser uno de los principales objetivos de este trabajo.

La presentación de los resultados se hará en los siguientes tres capítulos, como se describe a continuación:

En el capítulo 4 se presentarán los resultados en los sesgos relativos de los parámetros obtenidos en la estimación al fijar la distribución normal en el término del error aleatorio del modelo y controlando el efecto aleatorio con: la distribución normal, la normal sesgada (SN) y la t sesgada (ST).

En el capítulo 5 se muestran los resultados del sesgo relativo de los parámetros estimados cuando se combinan las distribuciones normal, SN y ST para el error y el efecto aleatorio, así como el comportamiento de la normalidad de los residuales del modelo estimado. A partir de

los resultados obtenidos con la medición del sesgo relativo en los parámetros y la normalidad de los componentes de error y efecto aleatorio, se responderá a la hipótesis sobre los efectos e impactos que producen la variación de las distribuciones y los demás factores mencionados, en la estimación de un modelo lineal mixto, que se realiza bajo el supuesto de normalidad de los residuales.

Al realizar los diagnósticos de normalidad puede ocurrir que en los casos donde no se da, sea causada por el efecto, o por el error aleatorios, ésto es lo que se pretende mostrar al separar el efecto del error, con la metodología propuesta. En el capítulo 6 se muestran los impactos sobre los sesgos relativos en los parámetros del modelo lineal mixto al usar una transformación de Box Cox sobre la variable respuesta, y también, probando la estimación en los parámetros fijos del modelo lineal mixto con el algoritmo EM propuesto por Arellano y otros (2005).

3.1. Evaluación de la normalidad de los residuales del modelo

Como fue mencionado en el capítulo 1, se propone usar la prueba de normalidad de Shapiro Wilks para determinar el comportamiento de los residuales normalizados totales.

Los residuales del modelo mixto tienen dos componentes: el error crudo y el efecto aleatorio. La separación de los componentes del error total es mencionado por Gurka y otros (2006), donde el residual total para el sujeto i es expresado como: $\varepsilon_i = Z_i b_i + e_i$, siendo las componentes b_i y e_i independientes.

La normalidad en los residuales totales ocurre si ambos componentes también provienen de la distribución normal, si éstos son independientes (Gurka y otros, 2006), sin embargo, cuando ambos no provienen de la misma distribución, qué ocurre con el residual total?. Para ello, se propone cuantificar el residual crudo e y el efecto aleatorio b de manera separada y probar en

cada uno la normalidad, y de qué manera puede relacionarse con el comportamiento normal o no normal del error total.

Para responder a los diferentes interrogantes, se usaron 3000 corridas por simulación, cada una de las cuales estima un modelo lineal mixto, y luego de éste, la media de cada parámetro como el estimador del valor esperado, es decir, la media muestral dada por: $\hat{\beta}_i = \frac{\sum_{j=1}^{nsim} \hat{\beta}_{ij}}{nsim}$, $i=0,1$.

Antes de mencionar los pasos para extraer el residual total, el error crudo y el efecto, deben aclararse los siguientes aspectos relativos a los valores esperados condicionales.

El valor esperado de y condicionado en b es dado por: $E[y|b] = \mathbf{X}\beta + \mathbf{Z}b$ (McCulloch, 2000, pág. 95), y como el efecto y error aleatorio son variables aleatorias, entonces se podría predecir el valor de y condicionado en el efecto y error aleatorio estimando el valor esperado: $E[y|b, \varepsilon] = \mathbf{X}\beta + \mathbf{Z}b + \varepsilon$, por ello la predicción del residual parte de estimar el siguiente valor esperado: $\varepsilon = E[y|b, \varepsilon] - \mathbf{X}\beta - \mathbf{Z}b$, de aquí se encuentra: $\mathbf{Z}b = E[y|b, \varepsilon] - \mathbf{X}\beta - \varepsilon$.

Por otro lado, el mejor predictor del efecto aleatorio ($BP[b]$), según McCulloch (2000), puede obtenerse a partir de los datos y su estimación se asocia con su predicción, que para el efecto y error aleatorios, se denotan así: $BP\hat{[b]} = \hat{b}$, $Predictor[\varepsilon] = \hat{\varepsilon}$, respectivamente.

Así, el método propuesto para cuantificar el valor esperado del efecto aleatorio y el del error crudo, es el siguiente:

1. Para realizar este procedimiento, primero se calculan los residuales totales del modelo lineal mixto, $\hat{\varepsilon} = y_i - \hat{y}_i$
2. Seguido a esto, entonces el predictor del efecto aleatorio puede extraerse así:

La estimación del efecto aleatorio se realiza por medio de dos procedimientos, el primero es un estimador de Bayes empírico (Verbeke and Lesaffre, 1996), y el segundo, a partir del error del modelo, que consta de ambos componentes: el error y el efecto aleatorio.

- a) Estimador de Bayes empírico. Partiendo de la propuesta del estimador propuesto por Verbeke and Lesaffre (1996).

$$b_i = \mathbf{D}^* \mathbf{Z}'_i W_i (y_i - X_i \hat{\beta}),$$

Donde \mathbf{D}^* es la matriz de varianza covarianza del efecto aleatorio b para un sujeto, y los parámetros son reemplazados por sus estimaciones de máxima verosimilitud. Lo anterior entonces se expresa de forma general, así:

$$\hat{b} = \mathbf{BZ}'\mathbf{V}^{-1}(Y - \mathbf{X}\hat{\beta}).$$

Donde las estimaciones se encuentran vía simulación. El orden de este vector es $Nr \times 1$, y tomando en cuenta que la matriz \mathbf{Z} es de orden $T \times Nr$, entonces:

$$Nr \times 1 = (Nr \times Nr)(Nr \times T)(T \times T)(T \times 1 - (Nk \times p)(p \times 1))$$

b) A partir del modelo

$$\mathbf{Z}\hat{b} = \hat{y} - \mathbf{X}\hat{\beta} - \hat{\varepsilon}$$

$$(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{Z}\hat{b} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'(\hat{y} - \mathbf{X}\hat{\beta} - \hat{\varepsilon}),$$

Y así, se obtiene:

$$\hat{b} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'(\hat{y} - \mathbf{X}\hat{\beta} - \hat{\varepsilon}).$$

3. Una vez se encuentra el predictor del efecto aleatorio, por una de las dos vías, se calcula el error crudo: $\hat{e} = \hat{\varepsilon} - \hat{b}_{bayes}$ o $\hat{e} = \hat{\varepsilon} - \hat{b}_{modelo}$.
4. Luego de estos procedimientos, se analiza la normalidad que puede tener el error total, el residual crudo, y el efecto aleatorio, con la prueba de Shapiro Wilks.

De esta manera se desea analizar los resultados de simulaciones realizadas bajo diferentes situaciones controladas, descritas previamente, y establecer de dónde puede provenir la no

normalidad.

3.2. Transformación

La transformación usada es la siguiente: (Box y Cox (1964) y retomada en Gurka (2007)):

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^\lambda - 1}{\lambda(\tilde{y})^{\lambda-1}}, & \text{si } \lambda \neq 0 \\ \tilde{y} \log y_i, & \text{si } \lambda = 0 \end{cases}$$

Donde \tilde{y} es la media geométrica.

Partiendo de la siguiente expresión para la media geométrica:

$$\tilde{y} = \left(\prod_{i=1}^n y_i^{\alpha_i} \right)^{\frac{1}{\sum_i \alpha_i}} = (y_1^{\alpha_1} y_2^{\alpha_2} \dots y_n^{\alpha_n})^{\frac{1}{\alpha_1 + \dots + \alpha_n}}.$$

En este trabajo se propone usar una media geométrica donde los valores α_i se podría usar el inverso de la varianza de la variable respuesta, quedando una expresión más simple:

$$\tilde{y} = \prod_{i=1}^n y_i^{1/n}$$

3.3. Ejemplos de efecto aleatorio con características sesgadas

En un ejemplo dado por Zhou (2007), se describe un estudio sobre la ingesta de alimentos en humanos (deglution apnea), donde encuentran influencia del volumen de bolo (bolus) y tipo de alimentación, y se reitera que la distribución del error del modelo mixto y la del efecto

aleatorio son no normales, aun después de transformar la variable respuesta con el logaritmo de ésta.

Así mismo, en Arellano y otros (2005), se muestra un ejemplo donde la variable colesterol tiene un claro comportamiento asimétrico. Además, desarrollan el algoritmo EM para estimar el modelo mixto cuando el error y el efecto aleatorio se comportan como una normal sesgada, incluso, usando simulación bajo diferentes escenarios entre la normal y la normal sesgada, entre otras distribuciones con sesgo.

En la Tesis de grado de Muñoz (2004), realizan un diseño experimental para determinar el comportamiento de unas especies de microalgas donde extraen diferentes variables respuesta, una de ellas fue la Densidad Celular medida como el promedio del número de células en unidades dadas por: $\times 10^6/ml$, midiendo su valor por 7 días. Encontraron que la densidad se encuentra directamente relacionado con la biomasa algal y por ello fue de gran interés su medición.

En este trabajo se utiliza un conjunto de datos tomados de esa Tesis de grado, pero el modelo lineal mixto de interés, considerará el efecto aleatorio que causan las medidas repetidas en el tiempo de la Densidad Celular. Se estima entonces un modelo lineal mixto donde la variable respuesta es la Densidad Celular, la covariable x_1 será es el tiempo en el crecimiento algal, siendo un modelo con intercepto aleatorio, donde se toman 5 medidas repetidas por sujeto (tomando para 36 plantas del 1 al 5o día). Sus resultados se muestran a continuación.

- Estimación de un modelo lineal mixto donde los residuales se comportan de forma no normal.

Al estimar un modelo lineal mixto con 5 medidas repetidas (5 días) donde la variable respuesta es la Densidad Celular y la variable explicativa es el día, se encuentra que su valor $p < 0,0001$, lo cual indica su alta significancia o efecto sobre el crecimiento celular, sin embargo, no se encuentra normalidad ni para esta variable, ni al estar combinada

con el factor medio (tabla 3.1).

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-1.856222	1.3948201	143	-1.330797	0.1854
x1	5.390444	0.3006689	143	17.928175	0.0000

Tabla 3.1: Coeficientes fijos del modelo lineal mixto para la variable respuesta Densidad Celular.

Los parámetros estimados toman valores: $\hat{\beta}_0 = -1.85$, $vp = 0.185$, no significativo, y $\hat{\beta}_1 = 5.39$ con $vp < 0,0001$, significativo al nivel 0.05 (tabla 3.1).

El aparente sesgo en los residuales puede visualizarse al medir los indicadores de asimetría de Fisher, usando la función skewness del paquete fBasics de R. Para cada día se encuentran las medidas de asimetría para los residuales vistas en la tabla 3.2 Los valores positivos indican que puede existir una asimetría a derecha para los residuales, evaluados por día.

También se estimó el valor p de la prueba de normalidad de Shapiro Wilks, indicando con un $4,6 \times 10^{-9}$, la no aceptación de normalidad, con un nivel de significancia de 0.05. En este ejemplo, puede apreciarse cómo el tiempo de crecimiento puede generar un efecto significativo que incorpora sesgo en el comportamiento de los residuales, lo que puede tener un sesgo en la estimación de los parámetros, aspectos que serán entonces estudiados vía simulación en los siguientes capítulos.

Día 1:	1.077633
Día 2:	1.061791
Día 3:	1.370898
Día 4:	0.998884
Día 5:	0.880008

Tabla 3.2: Asimetría para los residuales del modelo mixto sobre Densidad Celular.

Capítulo 4

Simulación con error normal

En el esquema de la simulación de este capítulo, se fija la distribución normal del error aleatorio durante todo el proceso, pero la distribución del efecto aleatorio variará junto con los demás parámetros descritos en la metodología. Sin embargo, se estima el modelo lineal mixto bajo el supuesto de normalidad e independencia entre ε_i y b_i , determinando con estas variaciones los efectos producidos sobre sus estimaciones.

Para la simulación, además, se tomaron en cuenta los aspectos descritos en el marco teórico sobre el modelo lineal mixto con estructura de correlación de simetría compuesta, con intercepto aleatorio y medidas repetidas.

La siguiente es la notación relacionada con los parámetros y componentes del modelo lineal mixto utilizados en este trabajo.

1. **X**: Matriz de diseño, de $kN \times p$. Donde k es el número de medidas repetidas del modelo, N es el número de individuos, p es el número de parámetros fijos del modelo lineal mixto.
2. **Z**: Matriz de diseño asociada con el efecto aleatorio $kN \times rN$. Donde r es el número de

efectos aleatorios en el modelo más un intercepto aleatorio y como se asume $r=1$ para todo el proceso, el orden de \mathbf{Z} es $kN \times N$ y el efecto aleatorio es de orden $N \times 1$, luego la matriz \mathbf{Zb} es de orden $Nk \times 1$.

3. β : Vector de coeficientes de los efectos fijos: β_0 y β_1 .
4. \mathbf{R} : Matriz de varianza del error $kN \times kN$, donde σ_e^2 es valor de la varianza del error en este modelo (sig en la simulación).
5. Nsim: Número de simulaciones para estimación de modelos.
6. σ_b^2 : Varianza del efecto aleatorio del i -ésimo sujeto (D_i en simulación).
7. dis: Distribución del efecto aleatorio: 0 si es normal, 1 si es normal sesgada, 3 si es t sesgada, con parámetro de sesgo $a=\lambda$ para las respectivas distribuciones sesgadas.
8. distr: Distribución del error aleatorio: 0 si es normal, 1 si es normal sesgada, 2 si es t sesgada.
9. rho: ρ , Valor del coeficiente de correlación presente en la matriz de simetría compuesta.

En la tabla 4.1. se presentan los valores que fueron fijados en el trabajo sobre los parámetros descritos.

β_0	10, 20, 30, 50, 90, 100
β_1	10, 20, 30, 50, 200
λ	10, 12, 30, 50, 150, 200, 300, 600, 1200
σ_b^2	0.5, 1, 2.5, 5, 7.5, 9, 9.5, 25, 50, 75, 90, 95, 100, 250, 500, 750, 950
σ_e^2	0.5, 1, 2.5, 5, 7.5, 9, 9.5, 25, 50, 75, 90, 95, 100, 250, 500, 750, 950
$\sigma_{total}^2 = \sigma_b^2 + \sigma_e^2$	10, 100, 1000
N	30, 50, 60, 100, 150, 160, 320, 640

Tabla 4.1: Parámetros controlados en el modelo.

El número de medidas repetidas se deja fijo en $k=3$ para todas las simulaciones, éste no se cambiará en todo el proceso, como tampoco la matriz de diseño X , pues se desea explorar los efectos de las otras fuentes de variación mencionadas. Se dejan fijos el parámetro asociado a los grados de libertad de la distribución t sesgada en $\nu = 2$, ya que si se aleja mucho, se aproxima mucho a la normal. Además, en todas las corridas, el número de simulaciones $Nsim = 3000$. El software utilizado fue el R¹.

Como estimación de los parámetros se tomó la media de los valores encontrados en las 3000 simulaciones realizadas, y se denotarán en este trabajo así: $\hat{\beta}_0 = \frac{\sum_{i=1}^{nsim} \hat{\beta}_{0i}}{nsim}$ y $\hat{\beta}_1 = \frac{\sum_{i=1}^{nsim} \hat{\beta}_{1i}}{nsim}$, y los sesgos relativos se obtuvieron dividiendo la raíz del MSE por el respectivo valor inicial, como se describió en el capítulo 1.

En total se realizaron 155 corridas, que obedecen al control y variación en los diferentes parámetros del modelo, en las cuales la distribución del error aleatorio se conservó con la distribución normal.

4.1. Efecto sobre la estimación de parámetros del modelo

4.1.1. Error y efecto con distribución normal

En este caso, como era de esperarse cuando la distribución de ambos: error y efecto es normal, los parámetros estimados son insesgados, generando poco o ningún sesgo relativo importante. Por ejemplo, en la tabla 4.2 se muestran simulaciones cuando $\beta_0 = 20$ y $\beta_1 = 20$ para $N=50$, y $N=30$.

En esta tabla, se ve que para ambos tamaños muestrales, el sesgo es casi imperceptible (aproximadamente 0) para los dos parámetros fijos del modelo.

¹R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. [Vienna, Austria]. Disponible en Web: <<http://www.R-project.org>>

En cuanto a diagnósticos de normalidad, se logró en todos los escenarios mostrados en la tabla 4.2. por encima del 90 %, con un promedio de 92,4 % ($s = 0,03$) de las veces simuladas (3000) una aceptación de normalidad, viéndose un poco menor cuando la varianza del efecto aleatorio es alta. Razones por las cuales no fue necesaria mayor exploración de este escenario. Esto ocurre también para diferentes tamaños muestrales.

N	ρ_i	σ_e^2	σ_b^2	$\hat{\beta}_0$ (sd)	$\hat{\beta}_1$ (sd)	$\hat{\rho}$ (sd)	sesgobo	sesgobl	sesgorho
50	0.9	10	90	20 (1.53)	20.01 (0.32)	0.67 (0.01)	0	0	0.3
	0.75	25	75	19.92 (1.64)	20.03 (0.53)	0.57 (0.12)	0	0	0.2
	0.5	50	50	20.05 (1.81)	20 (0.7)	0.41 (0.09)	0	0	0.2
	0.25	75	25	20.03 (2.12)	19.97 (0.9)	0.14 (0.05)	0	0	0.5
	0.1	90	10	20.08 (2.1)	19.96 (0.94)	0 (0.05)	0	0	1
	0.05	95	5	20.07 (2.14)	19.97 (0.97)	-0.04 (0.05)	0	0	1.8
	0.01	99	1	19.94 (2.21)	20.03 (1.03)	-0.07 (0.05)	0	0	7.7
30	0.9	10	90	20.04 (1.95)	19.99 (0.43)	0.69 (0.02)	0	0	0.24
	0.75	25	75	19.92 (1.84)	20.02 (0.6)	0.57 (0.13)	0	0	0.24
	0.5	50	50	20.1 (2.18)	19.95 (0.83)	0.42 (0.1)	0.01	0	0.16
	0.25	75	25	20.01 (2.53)	19.98 (1.09)	0.13 (0.07)	0.01	0	0.49
	0.1	90	10	19.79 (2.62)	20.06 (1.19)	-0.02 (0.07)	0.01	0	1.16
	0.05	95	5	19.62 (2.65)	20.13 (1.19)	-0.04 (0.06)	0.02	0.01	1.89
	0.01	99	1	19.94 (2.73)	20.05 (1.22)	-0.07 (0.06)	0.01	0	8.07

Tabla 4.2: Comportamiento del sesgo relativo con b_i y $\varepsilon \sim N(0, \sigma^2)$.

4.1.2. Distribución normal para el error, variación en la distribución del efecto aleatorio

Inicialmente se presentarán los cambios observados sobre el intercepto y la pendiente del modelo lineal mixto con sus valores estimados, y se analizarán las posibles causas de éstos.

Cuando la distribución del efecto es normal sesgada, y dejando fija la varianza del efecto aleatorio ($\sigma_b^2=50$), así como los valores teóricos de los parámetros y con diferentes tamaños muestrales, se aprecia que el tamaño muestral no parece producir cambios notorios sobre el sesgo de β_0 , por ejemplo cuando el valor teórico del intercepto es 10, el sesgo para los tamaños de muestra considerados está entre 56 % – 57 %, y cuando es 100, el valor del sesgo es alrededor del 6 % (esto se resume en la tabla 4.3). De esta manera puede verse que al modificar el valor teórico del parámetro β_0 , el sesgo relativo cambia.

Es importante además, determinar los posibles efectos de la varianza del efecto aleatorio causados sobre las estimaciones, como se verá en los siguientes resultados.

β_{0i}	β_{1i}	n	λ	ρ_i	σ_e^2	$\hat{\beta}_0(sd)$	$\hat{\beta}_1(sd)$	$\hat{\rho}(sd)$	sesgobo	sesgobl	sesgorho
10	20	30	50	0.5	50	15.64 (2.19)	20.03 (0.9)	0.18 (0.14)	0.56	0	0.64
		60	300	0.5	50	15.67 (1.48)	19.99 (0.64)	0.13 (0.13)	0.57	0	0.37
		60	600	0.5	50	15.62 (1.48)	20.01 (0.63)	0.13 (0.12)	0.56	0	0.37
		320	600	0.5	50	15.65 (0.65)	20 (0.28)	0.13 (0.06)	0.57	0	0.74
		640	600	0.5	50	15.64 (0.45)	20 (0.2)	0.14 (0.05)	0.56	0	0.36
		320	50	0.05	950	15.6 (2.7)	20.02 (1.24)	-0.05 (0.02)	0.56	0	1.97
		320	150	0.05	950	15.62 (2.58)	20.01 (1.19)	-0.05 (0.02)	0.56	0	1.97
100	200	160	300	0.5	50	105.7 (0.91)	200 (0.39)	0.12 (0.1)	0.06	0	0.77
		320	300	0.5	50	105.7 (0.64)	200 (0.27)	0.13 (0.07)	0.06	0	0.74
		320	1200	0.05	950	105.72 (2.66)	200 (1.24)	-0.05 (0.02)	0.06	0	1.96
		320	800	0.05	950	105.66 (2.7)	200 (1.24)	-0.05 (0.02)	0.06	0	1.98
		320	600	0.05	950	105.7 (2.66)	200 (1.23)	-0.05 (0.02)	0.06	0	1.97
		320	650	0.05	950	105.67 (2.65)	200 (1.21)	-0.05 (0.02)	0.06	0	1.97

Tabla 4.3: Comportamiento del sesgo relativo cuando $b_i \sim SN(0, \sigma_b^2 = 50, \lambda)$.

En la tabla 4.4, se ve variar el sesgo en un rango entre 8 % y 25 %, para valores de la varianza del efecto aleatorio σ_b^2 entre 1 y 9, pero con igual tamaño de muestra y parámetro de sesgo λ . A partir de ésto se observa un aparente cambio en el sesgo del intercepto para diferentes valores de esta varianza.

β_{0i}	β_{1i}	n	λ	rhoi	σ_b^2	$\hat{\beta}_0(sd)$	$\hat{\beta}_1(sd)$	$\hat{\rho}(sd)$	sesgobo	sesgob1	sesgorho
				0.95	9.5	12.46 (0.22)	10.00 (0.05)	0.86 (0.02)	0.25	0	0.1
				0.9	9	12.38 (0.23)	10.00 (0.07)	0.7 (0.04)	0.24	0	0.23
10	10	100	12	0.8	7.5	12.18 (0.29)	9.996 (0.11)	0.32 (0.02)	0.22	0	0.58
				0.3	2.5	11.25 (0.43)	10.01 (0.2)	0.02 (0.05)	0.13	0	0.91
				0.1	1	10.77 (0.47)	10.014 (0.22)	-0.03 (0.04)	0.08	0.001	1.3

Tabla 4.4: Comportamiento del sesgo relativo cuando $b_i \sim SN(0, \sigma_b^2, \lambda)$ variando σ_b^2 .

Esto también puede verse por medio de una exploración gráfica. Por ejemplo, en la figura 4.1, se explora el comportamiento del sesgo relativo de β_0 para 2 valores teóricos: en el izquierdo $\beta_0=10$ y en el derecho $\beta_0=100$, vs la varianza del efecto aleatorio. Se observan diferentes escalas de variación del sesgo relativo para ambos valores de los parámetros, y es apreciable el aumento el sesgo a medida que aumenta la varianza del efecto aleatorio σ_b^2 .

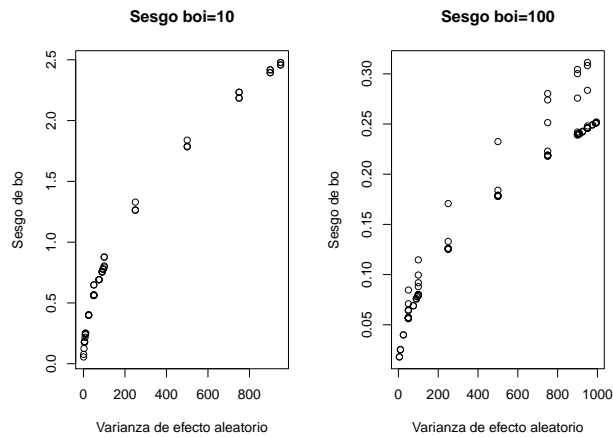


Figura 4.1: Sesgo de β_0 cuando es igual a 10 y a 100 vs σ_b^2 .

Por otro lado, en simulaciones adicionales, se observa que cuando la distribución para b_i es ST, los sesgos relativos son también altos, en algunos casos de altas varianzas, similares al comportamiento normal sesgado, bajo un parámetro de sesgo λ similar. El sesgo del parámetro β_1 no cambia, es cero en casi todos los casos, éste no se ve afectado por el cambio de la

distribución del efecto aleatorio. Estas observaciones serán ampliadas en el siguiente capítulo. También puede verse que sobre el sesgo del intercepto puede también ser influyente el parámetro de sesgo de la distribución. Para ello es útil otro tipo de exploración como la de un modelo de regresión, que será descrito a continuación.

Exploración con un modelo lineal

Al estimar un modelo de regresión lineal donde la variable respuesta el sesgo relativo de β_0 , y usando 97 datos de los 155 que habían, con el fin de lograr un balance de tamaños muestrales y eliminando valores atípicos, se encuentra un modelo que logra explicar significativamente dicha respuesta. Los resultados de este modelo se muestran en la tabla 4.5.

Se encontró que, con un nivel de significancia de 0,05, y un valor $p = 0$, las variables varianza del efecto aleatorio, varianza del error aleatorio y parámetro de sesgo λ producen un aumento significativo sobre el sesgo relativo de β_0 , por otro lado, el valor teórico de β_0 y su interacción con las varianzas, así como la interacción entre la varianza del efecto aleatorio y λ , generan una disminución sobre este sesgo relativo.

Esto permite inferir que sí son ciertas las afirmaciones realizadas en la sección anterior.

Coefficients:	Estimate	Std.Error	t	Valor p
(Intercept)	1.49e-01	1e-02	14.47	2e-16
λ	1.17e-04	1.95e-05	5.98	4.51e-08
σ_b^2	7.33e-03	1.71e-04	42.77	2e-16
σ_e^2	7.56e-05	1.8e-05	4.14	7.91e-05
β_{oi}	-1.64e-03	1.5e-04	-10.59	2e-16
$\sigma_b^2;\lambda$	-6.44e-07	1.5e-07	-4.23	5.71e-05
$\sigma_e^2;\beta_{oi}$	-4.83e-07	2.4e-07	-2.0	0.0486
$\sigma_b^2;\beta_{oi}$	-6.7e-05	1.8e-06	-37.16	2e-16

Tabla 4.5: Modelo de regresión para la respuesta sesgo de β_0 con error normal.

Este modelo tiene un comportamiento adecuado de normalidad pues en la prueba de Shapiro Wilks se obtuvo un estadístico $W = 0,986$, $valor p = 0,3962$, indicando aceptación de nor-

malidad para los residuales. Además, un valor de $R^2 = 98,4\%$ indica una adecuada explicación de variabilidad.

Sesgo sobre el coeficiente de correlación intraclase

Este sesgo estimado presenta alta variación y tendencias que cambian según algunos factores modificados, uno de ellos la varianza del efecto aleatorio, como se ve en la figura 4.2.

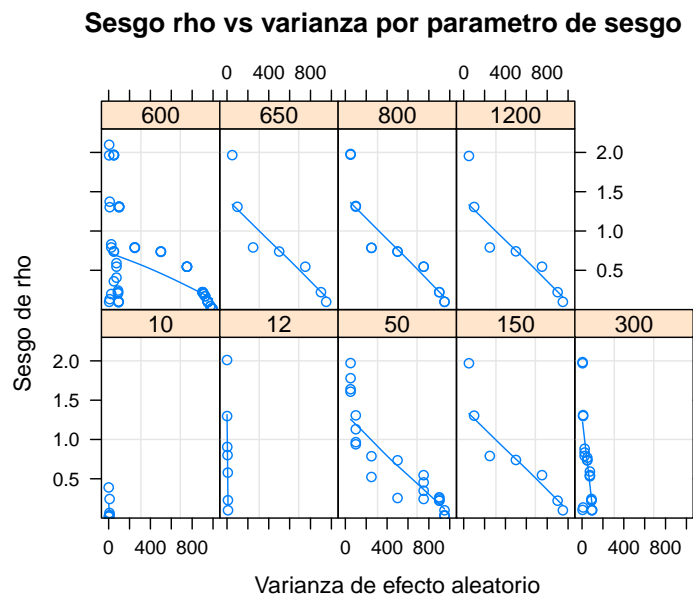


Figura 4.2: Sesgo de β_0 vs varianza de b_i y por parámetro λ .

Una de las posibles causas de este cambio es la varianza del efecto aleatorio, pues se aprecia un patrón de posible linealidad entre la correlación estimada y el valor σ_b^2 y otra causa posible es el parámetro λ de la distribución. Sin embargo, no se aprecian cambios aparentes con respecto al tamaño muestral o distribución.

Hasta el momento se ha encontrado un problema en el comportamiento del sesgo relativo del intercepto y de la correlación intraclase, cuando el efecto aleatorio se fija con una distribución que no es normal y el error sí lo es, en las estimaciones simuladas de un modelo lineal mixto con intercepto aleatorio. Ahora, en el siguiente capítulo se mostrarán los efectos sobre los sesgos relativos que tiene realizar diferentes combinaciones entre distribuciones para el efecto y el error aleatorios, además de los diagnósticos de normalidad.

Capítulo 5

Combinación de distribuciones para el error y el efecto aleatorio

En este capítulo se describe, vía simulación, el comportamiento del sesgo sobre las estimaciones de los parámetros β_0 , β_1 y ρ usando diferentes combinaciones de distribuciones entre la normal, normal sesgada y t sesgada, tanto para el error como para el efecto aleatorio. Para ello se estimaron 157 corridas en las cuales se combinaron las distribuciones y las diferentes variables de entrada del modelo lineal mixto con intercepto aleatorio.

5.1. Efecto sobre la estimación de los parámetros del modelo lineal mixto

En el capítulo anterior se encontró que el sesgo relativo de β_0 es afectado significativamente por el parámetro del sesgo de la distribución, así como la varianza del error y del efecto aleatorio. En este capítulo se analizará si ocurre lo mismo con el comportamiento del sesgo relativo de dicho coeficiente y del parámetro β_1 , cuando se dan combinaciones distribucionales diferentes entre el error y el efecto aleatorio.

Al realizar una exploración preliminar del sesgo relativo de β_0 , se encuentra una variación interesante, por ejemplo, en la figura 5.1, se muestra el comportamiento del sesgo con respecto a la varianza σ_b^2 cuando el parámetro teórico es $\beta_{0i} = 10$, notando que hay un aparente crecimiento del sesgo del parámetro β_0 , a medida que aumenta dicha varianza, comportamiento que es similar para las diferentes distribuciones del error ε .

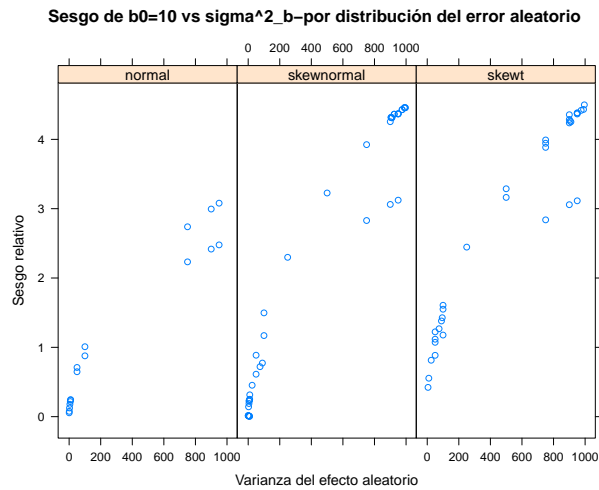


Figura 5.1: Sesgo de β_0 vs varianza de b_i cuando $\beta_{0i} = 10$.

En la figura 5.2, se muestra que para $\beta_{0i} = 100$, ocurre algo semejante al caso anterior, un crecimiento del sesgo relativo a medida que aumenta la varianza σ_b^2 . Sin embargo, la escala del sesgo relativo es más baja.

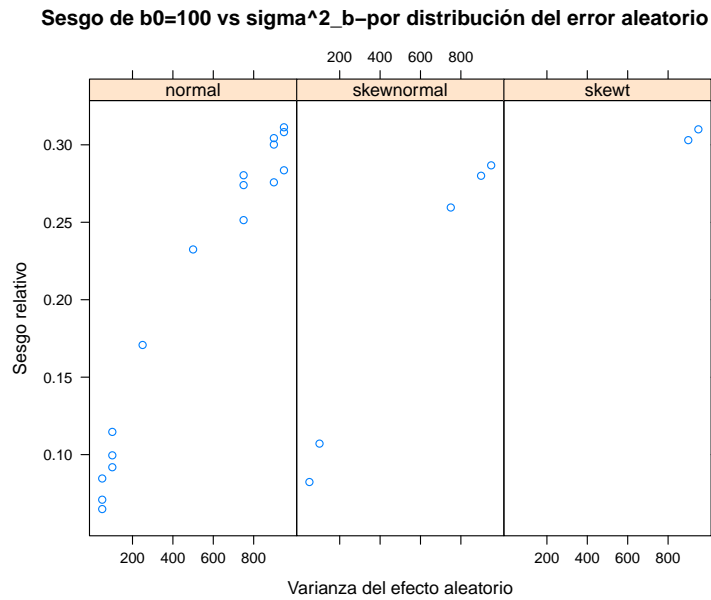


Figura 5.2: Sesgo de β_0 vs varianza de b_i cuando $\beta_{0i} = 100$.

En todas las corridas realizadas, el sesgo relativo β_0 , toma valores que en porcentajes oscilan de 0,33 % a 449,8 % y es importante explorar las posibles causas de su variación.

En la tabla 5.1 se muestran los valores de los sesgos estimados para tres valores de la varianza del efecto: $\sigma_b^2 = 50, 100, 900, \lambda = 600$. Se ve cómo el sesgo relativo del intercepto β_0 toma valores más altos que los del sesgo del parámetro β_1 .

En dicha tabla se aprecia que el sesgo del parámetro β_0 parece tener variaciones según los diferentes factores controlados. Por ejemplo, en la primera franja, los sesgos de β_0 , están entre 242 % y 436 %, cuando $\sigma_b^2 = 900$, fijando $b_i \sim ST$, o SN y $\varepsilon \sim Normal, SN$ o ST , además $\beta_{0i} = 10$. En la franja 3, se ve un sesgo menor, entre 28 % y 30 %, cuando $b_i \sim ST, \sigma_b^2 = 900$, y $\beta_{0i} = 100$. Pero cuando $\beta_{0i} = 10$ y $\sigma_b^2 = 50$, están entre 65 % y 112 % (franja 7).

					$\sigma_e^2=100$	$\sigma_b^2=900$	$\rho_i = 0,9$			
β_{0i}	β_{1i}	dis. b_i	dis. ε	n	$\hat{\beta}_0(sd)$	$\hat{\beta}_1(sd)$	$\hat{\rho}$	sesgob0	sesgob1	sesgorho
10	10	SN	normal	320	34.17 (1.34)	10.01 (0.4)	0.7 (0.02)	2.42	0	0.22
		ST	SN	320	52.55 (5.66)	10 (0.4)	0.68 (0.08)	4.26	0	0.24
		ST	ST	320	52.87 (8.53)	9.99 (2.52)	0.65 (0.11)	4.29	0	0.28
		ST	ST	100	53.56 (21.08)	9.85 (7.35)	0.62 (0.14)	4.36	0.01	0.31
		ST	ST	100	40.58 (3.68)	9.99 (0.99)	0.68 (0.06)	3.06	0	0.24
50	50	normal	SN	30	50.72 (5.98)	50.01 (1.26)	0.69 (0.02)	0.01	0	0.23
		normal	ST	30	51.75 (10.09)	49.85 (3.83)	0.53 (0.16)	0.03	0	0.41
		ST	normal	160	92.47 (7.33)	50 (0.57)	0.7 (0.09)	0.85	0	0.23
100	200	ST	normal	320	127.6 (1.61)	200 (0.39)	0.7 (0.02)	0.28	0	0.22
		ST	SN	100	128 (3.06)	200 (0.87)	0.69 (0.05)	0.28	0	0.23
		ST	ST	320	130.3 (2.09)	200 (0.55)	0.7 (0.03)	0.3	0	0.23
					$\sigma_e^2=900$	$\sigma_b^2=100$	$\rho_i = 0,1$			
β_{0i}	β_{1i}	dis. b_i	dis. ε	n	$\hat{\beta}_0(sd)$	$\hat{\beta}_1(sd)$	$\hat{\rho}$	sesgob0	sesgob1	sesgorho
10	10	SN	normal	320	18.77 (2.62)	9.98 (1.21)	-0.03 (0.02)	0.88	0	1.31
		ST	normal	320	20.09 (2.64)	9.96 (1.19)	0 (0.02)	1.01	0	0.95
		ST	SN	100	21.7 (6.59)	10 (2.97)	-0.02 (0.04)	1.17	0.01	1.23
		ST	SN	320	24.97 (3.14)	9.97 (1.18)	0.1 (0.03)	1.5	0	0.03
		ST	ST	100	21.77 (6.55)	10 (2.97)	-0.02 (0.04)	1.18	0.01	1.22
50	50	normal	SN	30	52.89 (8.55)	49.75 (3.89)	0 (0.06)	0.06	0	1.03
		normal	ST	30	54.6 (24.84)	49.79 (11.46)	-0.05 (0.06)	0.09	0.01	1.48
100	200	ST	normal	320	109.2 (2.64)	200 (1.19)	-0.01 (0.02)	0.09	0	1.13
		ST	normal	100	111.47 (4.7)	200 (2.13)	0 (0.04)	0.11	0	0.97
		ST	SN	100	110.7 (5.62)	200 (2.57)	-0.03 (0.04)	0.11	0	1.26
					$\sigma_e^2=950$	$\sigma_b^2=50$	$\rho_i = 0,05$			
β_{0i}	β_{1i}	dis. b_i	dis. ε	n	$\hat{\beta}_0(sd)$	$\hat{\beta}_1(sd)$	$\hat{\rho}$	sesgob0	sesgob1	sesgorho
10	10	SN	normal	320	16.48 (2.59)	9.99 (1.21)	-0.05 (0.02)	0.65	0	1.97
		ST	normal	320	17.1 (2.66)	9.98 (1.22)	-0.03 (0.02)	0.71	0	1.59
		ST	SN	100	18.87 (6.79)	9.98 (3.1)	-0.05 (0.04)	0.89	0.01	1.92
		ST	ST	320	21.18 (10.37)	10.07 (4.61)	-0.03 (0.03)	1.12	0	1.68
		ST	ST	100	18.85 (6.69)	10.01 (3.1)	-0.05 (0.04)	0.88	0.01	1.91
50	50	normal	SN	30	52.52 (8.73)	49.9 (3.98)	-0.04 (0.06)	0.05	0	1.89
		normal	ST	30	55.12 (27.96)	49.87 (12.26)	-0.06 (0.06)	0.1	0.01	2.3
100	200	ST	normal	320	106.5 (2.63)	200 (1.21)	-0.04 (0.02)	0.06	0	1.78
		ST	normal	100	108.46 (4.79)	200 (2.18)	-0.03 (0.04)	0.08	0	1.64
		ST	SN	100	108.23 (5.8)	200 (2.67)	-0.05 (0.04)	0.08	0	1.96

Tabla 5.1: Comportamiento de sesgo de β_0 para diferentes varianzas y distribuciones.

Por las razones expuestas, se sospecha la influencia de la varianza del efecto aleatorio σ_b^2 o la varianza total, la distribución, e incluso el valor teórico del parámetro β_0 , sobre el sesgo relativo del parámetro β_0 .

Explorando más a fondo una de las apreciaciones encontradas, cuando la distribución del efecto aleatorio es normal, y la del error no, el problema del sesgo parece no ser tan grave, esta situación se muestra en la tabla 5.2, en la cual el sesgo de β_0 no es tan alto cuando $\lambda = 200$.

		<i>bi ~ Normal</i>										
		$\lambda = 200$										
β_{0i}	β_{1i}	distr. ε	n	ρ_i	σ_ε^2	σ_b^2	$\hat{\beta}_0(sd)$	$\hat{\beta}_1(sd)$	$\hat{\rho}(sd)$	sesgob0	sesgob1	sesgorho
10	10	SN	100	0.9	1	9	10 (0.309)	10 (0.065)	0.65 (0.009)	0.002	0	0.278
				0.5	5	5	10.107 (0.408)	10.003 (0.15)	0.409 (0.054)	0.01	0.001	0.182
				0.1	9	1	10.123 (0.448)	9.993 (0.2)	0.006 (0.039)	0.012	0.001	0.941
30	30	SN	100	0.9	1	9	30.03 (0.338)	30 (0.068)	0.649 (0.009)	0	0	0.279
				0.5	5	5	30.11 (0.42)	30 (0.157)	0.404 (0.052)	0.004	0	0.192
				0.1	9	1	30.21 (0.485)	29.97 (0.214)	0.006 (0.037)	0.007	0.001	0.94
30	30	ST	100	0.9	1	9	30.11 (0.586)	30 (0.225)	0.427 (0.185)	0.004	0.001	0.525
				0.5	5	5	30.09 (1.254)	30.06 (0.538)	-0.04 (0.069)	0.001	0.001	1.081
				0.1	9	1	30.23 (1.573)	29.99 (0.681)	-0.045 (0.035)	0.007	0.002	1.447
10	10	SN	100	0.9	100	900	10.47 (3.3)	10.03 (0.68)	0.65 (0.01)	0.04	0	0.28
				0.5	500	500	11.17 (4.41)	10.01 (1.57)	0.41 (0.05)	0.11	0.01	0.19
				0.1	900	100	11.68 (4.99)	9.87 (2.27)	0.01 (0.04)	0.16	0.01	0.93
30	30	ST	30	0.9	1	9	30.24 (0.99)	29.96 (0.42)	0.52 (0.16)	0.01	0	0.42
				0.5	5	5	30.16 (3.65)	29.99 (0.89)	0.07 (0.13)	0.01	0	0.85
				0.1	9	1	30.61 (2.81)	29.9 (1.32)	-0.05 (0.07)	0.02	0	1.54
100	200	ST	30	0.9	10	90	100.32 (3.04)	200.1 (1.2)	0.55 (0.16)	0	0	0.39
				0.5	50	50	100.99 (5.53)	200 (2.54)	0.09 (0.12)	0.01	0	0.83
				0.1	90	10	101.23 (8.7)	200.1 (4.25)	-0.04 (0.06)	0.01	0	1.43
		<i>bi ~ Normal</i>										
		$\lambda = 600$										
10	10	ST	30	0.9	100	900	12.657 (10.378)	9.526 (4.097)	0.516 (0.171)	0.255	0.038	0.427
				0.5	500	500	13.953 (20.679)	9.284 (11.581)	0.084 (0.121)	0.367	0.04	0.831
				0.1	900	100	16.008 (28.29)	9.282 (12.015)	-0.047 (0.067)	0.567	0.045	1.471

Tabla 5.2: Comportamiento de sesgo de β_0 cuando $bi \sim Normal$.

En esta última tabla, se ve que el sesgo relativo toma un valor máximo de 1.2 %, cuando $b_i \sim Normal$, $\lambda = 200$, y la varianza total está entre 10 y 100, pero es máximo de 16 % cuando la varianza total es 1000. Sin embargo, el sesgo relativo llega a un máximo de 57 % cuando $\lambda = 600$. Esto también sugiere la existencia un efecto del parámetro de sesgo λ , de la varianza, la distribución, e incluso el valor teórico del parámetro β_0 , sobre el sesgo relativo

del parámetro β_0 .

Estos posibles efectos serán explorados en las siguientes figuras sobre el comportamiento del sesgo relativo del parámetro β_0 para diferentes distribuciones del efecto aleatorio.

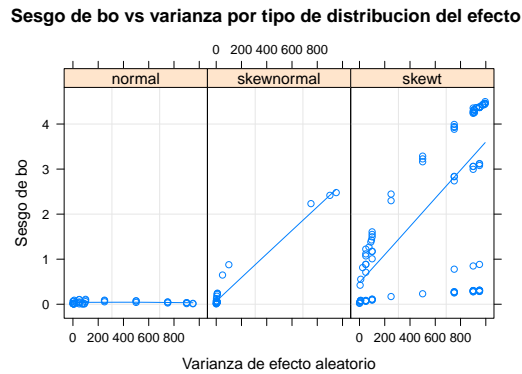


Figura 5.3: Sesgo de β_0 vs σ_b^2 y distribución del efecto. 1). $b_i \sim N$, 2). $b_i \sim SN$, 3). $b_i \sim ST$.

En la figura 5.3 se muestra en cada panel el comportamiento del sesgo relativo vs la varianza σ_b^2 con una distribución diferente para el efecto aleatorio, así que en el primer escenario, cuando $b_i \sim Normal$, no se aprecian cambios o efectos ascendentes que indiquen problemas de estimación, mientras que si $b_i \sim SN$, a medida que aumenta la varianza, el sesgo también aumenta y si $b_i \sim ST$, esto es más evidente. Este comportamiento puede detallarse más para un valor teórico de β_0 .

Por ejemplo, el ver en la tabla 5.3, cuando se dejan fijos: la distribución para el error $\varepsilon \sim N$, y el valor del parámetro teórico $\beta_{0i} = 10$, pero cambiando β_{1i} y las distribuciones SN y ST para el efecto aleatorio, los valores de sesgo de β_0 son altos cuando la varianza total es 100, pero disminuye cuando es menor la varianza σ_b^2 (última franja).

$\beta_{0i} = 10$				$\varepsilon \sim Normal$						
β_{1i}	n	λ	dis. b_i	σ_ε^2	σ_b^2	$\hat{\beta}_0(sd)$	$\hat{\beta}_1(sd)$	$\hat{\rho}(sd)$	sesgob0	sesgorho
20	30	50	ST	10	90	22.98 (4.13)	19.96 (0.41)	0.77 (0.07)	1.3	0.15
				50	50	20.13 (4.85)	20.04 (0.92)	0.46 (0.1)	1.01	0.08
				90	10	14.48 (2.87)	20.01 (1.2)	0.04 (0.06)	0.45	0.64
20	30	50	SN	10	90	17.67 (1.43)	20 (0.41)	0.66 (0.05)	0.77	0.26
				50	50	15.64 (2.19)	20.03 (0.9)	0.18 (0.14)	0.56	0.64
				90	10	12.07 (2.37)	20.16 (0.99)	-0.04 (0.07)	0.21	1.38
10	50	600	SN	0.5	9.5	12.47 (0.31)	10 (0.07)	0.85 (0.04)	0.25	0.11
				5	5	11.76 (0.56)	9.99 (0.25)	0.13 (0.12)	0.18	0.74
				9	1	10.78 (0.63)	10 (0.28)	-0.04 (0.06)	0.08	1.41

Tabla 5.3: Comportamiento de sesgo de β_0 cuando $\varepsilon \sim N(0, \sigma^2)$.

Sesgo de β_0 vs varianza por tipo de distribución del error

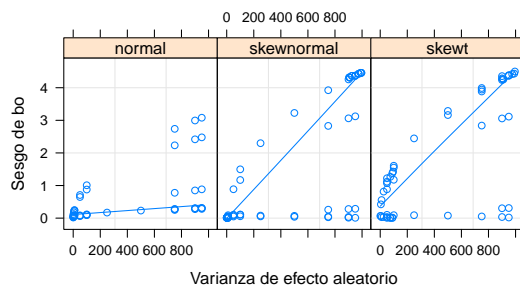


Figura 5.4: Sesgo de β_0 vs σ_b^2 y distribución del error.

Cuando se observa el sesgo de β_0 por tipo de distribución del error en la figura 5.4, se evidencia un patrón de variabilidad en el primer panel, donde $\varepsilon \sim N$, pero es aún más alto y con mayor valor del sesgo cuando $\varepsilon \sim SN$ o con $\varepsilon \sim ST$.

Los hallazgos sobre los efectos causados pueden verse reflejados con mayor seguridad en la exploración del comportamiento del sesgo con un modelo lineal, como se verá a continuación.

- Exploración con un modelo lineal

Es importante estimar un modelo de regresión lineal con el fin de analizar la significancia de los diferentes factores que alteran el comportamiento del sesgo relativo del parámetro que se ha observado más afectado por la estimación con error y efecto no normal: β_0 , en la exploración preliminar.

Al estimar los posibles efectos sobre el sesgo del intercepto β_0 por medio de un modelo lineal con 157 datos de las simulaciones, el resultado es que con un nivel de significancia de 0.01, las variables significativas con un $p \sim 0$ para explicar el comportamiento del sesgo relativo son: la varianza del efecto aleatorio, el parámetro λ de la distribución sesgada, ambos tipos de distribuciones (efecto y error aleatorios) cuyos efectos indican que a medida que aumentan, éstos hacen que el sesgo aumente, contrario el valor del parámetro teórico, pues su aumento disminuye el sesgo. Sin embargo, no hay normalidad para el mencionado modelo.

Se buscaron datos influenciales, y se eliminaron datos cuyo tamaño de muestra era de 320 sujetos, quedando un total de 130 corridas, para verificar su posible incidencia. Se encontró con una base más balanceada por tamaño muestral, una mejora en la normalidad del modelo lineal estimado ($Vp = 0,0226$ Shapiro Wilks) al nivel de significancia de 0.01, mostrando los efectos vistos en la tabla 5.4. Lo anterior muestra que sólo decrece el sesgo de β_0 con el aumento del parámetro teórico, de lo contrario, los demás factores ejercen un efecto de incremento significativo: varianza de b_i , λ , las distribuciones SN o ST del error o del efecto aleatorio, según la estimación del modelo de regresión lineal para explicar el comportamiento de dicho sesgo relativo.

Por otro lado, el sesgo de la correlación ρ sólo aparenta estar influenciado por la varianza. A medida que ésta aumenta, el sesgo parece disminuir. Sin embargo puede tener otras causas de variación que ameritan mayor exploración en otros trabajos.

Coefficients:	Estimate	Std. Error	t value	$Pr(> t)$
(Intercept)	-0.381927	0.204973	-1.863	0,0647
Varianza de efecto aleatorio	0.0015804	0.0001564	10.106	$< 2e - 16$
λ	0.0009062	0.0002234	4.057	$8,51E - 05$
$b_i \sim SN$	0.7790465	0.1880204	4.143	$6,12E - 05$
$b_i \sim ST$	1.5062649	0.1546026	9.743	$< 2e - 16$
$\varepsilon \sim SN$	0.4104984	0.1692389	2.426	0,0167
$\varepsilon \sim ST$	0.3414764	0.1724251	1.98	0,0498
β_{0i}	-0.019804	0.002355	-8.409	$6,36E - 14$

Tabla 5.4: Modelo de regresión para explicar el comportamiento del sesgo de β_0 .

Una razón por la cual la estimación del parámetro β_0 tienen más sesgo relativo que la estimación de β_1 , cuando b_i no tiene distribución normal, puede deberse a que el modelo lineal mixto considerado, tiene estructura con intercepto aleatorio, así que el comportamiento de $\hat{\beta}_0$ está alterado por el comportamiento no normal de esta componente aleatoria del modelo lineal mixto. Por ello también se observa un aumento de $\hat{\beta}_0$ al aumentar la varianza del efecto aleatorio σ_b^2 . Esto también puede generar alteraciones en los comportamientos no normales en el error total, como se verá al explorar la proporción de no rechazo de normalidad del error total y sus componentes.

5.2. Diagnósticos de los residuales del modelo lineal mixto

Los diagnósticos presentados, se basan en revisar el cumplimiento de los supuestos de normalidad en todos los componentes de los residuales del modelo lineal mixto: el residual total del modelo, el componente del efecto aleatorio y el error crudo. En esta sección se mostrarán los posibles efectos que tienen los factores controlados en este trabajo sobre dichos supuestos.

5.2.1. Normalidad de error total

Se usó la prueba de normalidad de Shapiro Wilks para todas las componentes, incluido el error total, después de aplicar la normalización de Cholesky, reiterando que se logra incorrelación

por el uso de la descomposición sobre la matriz de varianza Σ .

La exploración de normalidad inicia observando el comportamiento de la proporción de veces que se encuentra normalidad para todas las corridas por simulación por variable: error total, error crudo.

En la figura 5.5 se observa que si la distribución del efecto aleatorio es normal entonces la proporción de no rechazo de normalidad del error total es mayor cuando el error crudo tiene una distribución normal sesgada, que si es t sesgada. Esto indica que hay una mayor frecuencia en la violación del supuesto si la distribución del error aleatorio es t sesgada que cuando es normal sesgada, cuando para ambos casos, el efecto tiene distribución normal.

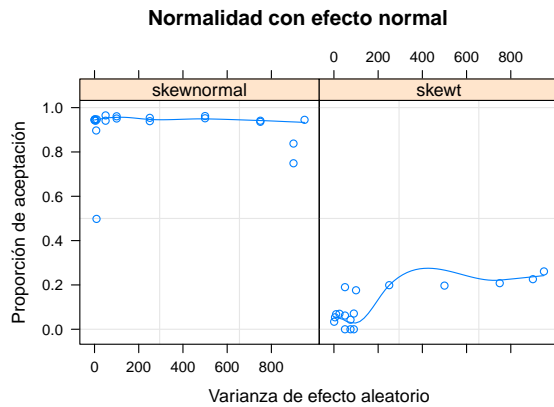


Figura 5.5: Normalidad del error total vs σ_b^2 cuando $b_i \sim N(0, \sigma_b^2)$, 1) $\varepsilon \sim SN$, 2) $\varepsilon \sim ST$.

Algo similar ocurre si la distribución del efecto aleatorio es t sesgada, pues la proporción de no rechazo de normalidad es menor cuando se combina con la t sesgada para el error, tal como se aprecia en la figura 5.6.

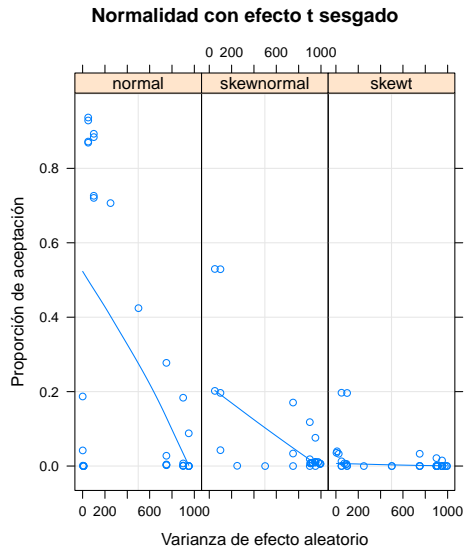


Figura 5.6: Normalidad del error total vs σ_b^2 cuando $b_i \sim ST(0, \sigma_b^2, \lambda, \nu)$, $\varepsilon \sim N$, $\varepsilon \sim SN$, $\varepsilon \sim ST$.

Es importante notar que bajo la distribución t sesgada para el efecto aleatorio b_i , con los grados de libertad usados ($\nu = 2$), se rechaza normalidad con alta frecuencia, si adicionalmente se aumenta la varianza del modelo, es más difícil alcanzar la normalidad en los errores.

En la tabla 5.5 se pueden resumir los resultados descritos sobre la normalidad del error. Aquí, $P1$ es la proporción de veces que se acepta normalidad para el error total, de las 3000 simulaciones realizadas.

En esta tabla es notorio el cambio del comportamiento de la proporción de no rechazo de normalidad en los errores según el valor teórico del vector de parámetros β , y al fijar la distribución t sesgada en el error aleatorio, y en algunos casos, cuando el parámetro λ toma el valor 300.

dis. b_i	distr. ε	β_{0i}	β_{1i}	n	λ	σ_e^2	σ_b^2	ρ_i	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\rho}$	sesgob0	P1
Normal	SN	10	10	100	300	1	9	0.9	10.05	10	0.65	0.01	0.5
						5	5	0.5	10.08	10	0.41	0.01	0.94
						7.5	2.5	0.25	10.11	10	0.14	0.01	0.95
SN	Normal	10	10	100	300	1	9	0.9	12.42	10	0.7	0.24	0.64
						5	5	0.5	11.8	9.99	0.08	0.18	0.87
						9	1	0.1	10.81	10	-0.03	0.08	0.94
Normal	ST	10	10	100	300	1	9	0.9	10.09	9.99	0.43	0.01	0.01
						5	5	0.5	10.29	9.95	-0.04	0.03	0
						7.5	2.5	0.25	10.25	10	-0.01	0.02	0
SN	ST	10	10	100	300	1	9	0.9	12.55	9.97	0.23	0.26	0
						5	5	0.5	12.11	9.94	-0.05	0.21	0
						7.5	2.5	0.25	11.55	9.96	-0.03	0.15	0
Normal	SN	50	50	30	100	50	950	0.95	50.56	50	0.93	0.01	0.95
						250	750	0.75	51.4	50	0.58	0.03	0.94
						900	100	0.1	52.89	49.8	0	0.06	0.95
SN	Normal	50	50	30	100	50	950	0.95	74.65	49.95	0.85	0.49	0.87
						100	900	0.9	74.32	49.98	0.66	0.49	0.88
						900	100	0.1	58.47	49.85	-0.05	0.17	0.96
Normal	ST	50	50	30	100	50	950	0.95	50.43	50.26	0.73	0	0.24
						100	900	0.9	51.58	49.79	0.54	0.03	0.25
						900	100	0.1	56.34	49.05	-0.05	0.12	0.21
SN	ST	50	50	30	100	50	950	0.95	75.27	49.89	0.5	0.51	0.21
						100	900	0.9	75.98	49.64	0.34	0.52	0.16
						900	100	0.1	63.44	49.74	-0.06	0.27	0.21
SN	SN	20	20	50	300	10	90	0.9	27.79	20.02	0.67	0.39	0.87
						50	50	0.5	26.1	19.99	0.17	0.31	0.91
						90	10	0.1	23.03	20.05	-0.04	0.15	0.94
SN	Normal	20	20	50	300	10	90	0.9	27.49	19.99	0.67	0.37	0.85
						50	50	0.5	25.57	20.02	0.14	0.28	0.94
						90	10	0.1	22.36	20.07	-0.04	0.12	0.96
Normal	ST	20	20	50	300	10	90	0.9	20.34	20.1	0.5	0.02	0.07
						25	75	0.75	20.51	20.1	0.25	0.02	0.04
						50	50	0.5	20.7	20	0.04	0.03	0.06
SN	SN	10	10	50	10	10	90	0.9	17.87	10	0.68	0.79	0.85
						50	50	0.5	16.15	10	0.16	0.62	0.95
						90	10	0.1	13.23	9.96	-0.04	0.32	0.95
Normal	SN	10	10	50	10	10	90	0.9	10.3	10	0.67	0.03	0.71
						50	50	0.5	10.47	9.98	0.41	0.05	0.96
						90	10	0.1	10.51	10.07	0	0.05	0.95
SN	ST	10	10	50	10	10	90	0.9	17.74	10.05	0.32	0.77	0.04
						50	50	0.5	16.07	10.06	-0.02	0.61	0.04
						90	10	0.1	14.48	9.61	-0.06	0.44	0.05

Tabla 5.5: Diagnósticos de normalidad para residuales totales del modelo mixto.

dis. b_i	distr. ε	β_{0i}	β_{1i}	n	λ	σ_e^2	σ_b^2	ρ_i	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\rho}$	sesgob0	P1
Normal	SN	10	10	100	100	1	9	0.9	10.06	10	0.65	0.01	0.44
						5	5	0.5	10.12	10.01	0.41	0.01	0.95
						7.5	2.5	0.25	10.13	10.01	0.13	0.01	0.95
SN	Normal	10	10	100	100	1	9	0.9	12.38	10.01	0.69	0.24	0.66
						5	5	0.5	11.77	10	0.09	0.18	0.91
						7.5	2.5	0.25	11.31	9.98	0.02	0.13	0.95
Normal	ST	10	10	100	100	1	9	0.9	10.04	10.02	0.44	0	0.01
						5	5	0.5	10.18	9.98	-0.04	0.02	0
						7.5	2.5	0.25	10.36	9.95	-0.01	0.04	0.01
SN	ST	10	10	100	100	1	9	0.9	12.45	10.01	0.21	0.24	0
						5	5	0.5	12.09	9.97	-0.05	0.21	0
						7.5	2.5	0.25	11.48	9.98	-0.04	0.15	0
Normal	SN	10	10	30	100	1	9	0.9	10.11	10	0.69	0.01	0.9
						5	5	0.5	10.22	9.99	0.41	0.02	0.94
						7.5	2.5	0.25	10.23	9.99	0.14	0.02	0.94
SN	Normal	10	10	30	100	1	9	0.9	12.43	9.99	0.67	0.24	0.88
						5	5	0.5	11.75	10.01	0.18	0.17	0.95
						7.5	2.5	0.25	11.29	10	0.05	0.13	0.96
Normal	ST	10	10	30	100	1	9	0.9	10.15	10.02	0.54	0.01	0.24
						5	5	0.5	10.19	10.04	0.08	0.01	0.24
						7.5	2.5	0.25	10.47	9.97	0.01	0.04	0.2
SN	ST	10	10	30	100	1	9	0.9	12.55	9.99	0.33	0.26	0.21
						5	5	0.5	12.23	9.98	0.01	0.22	0.18
						7.5	2.5	0.25	11.87	9.9	-0.03	0.19	0.21

Tabla 5.6: Diagnósticos de normalidad para residuales totales del modelo mixto.

La franja No. 1 de la tabla 5.5 puede compararse con los escenarios mostrados en la tabla 5.6, donde β_0 y β_1 y todos los demás parámetros se dejan fijos excepto el sesgo λ que se disminuye a 100, se aprecia que se conservan similares las proporciones de aceptación de normalidad del error total por combinación distribucional, comparadas con la tabla 5.5; situación que cambia cuando se disminuye el tamaño muestral a 30, donde las proporciones de aceptación de normalidad son más altas cuando la distribución ST está presente, sin embargo, el comportamiento de los sesgos relativos de β_0 es similar al caso respectivo de la tabla 5.5.

Es de notar en las dos tablas anteriores, que el sesgo del parámetro β_0 , se ve afectado por el sesgo λ , el valor teórico y el tipo de distribución, como anteriormente se mencionó, pero su comportamiento no cambia de igual manera que el de la proporción de no rechazo de normalidad en el error. Seguido a lo anterior se extraen los componentes del error crudo y efecto aleatorios, lo que permitirá explorar si la distribución del error total puede deberse al comportamiento de uno de sus dos componentes, como se muestra a continuación.

5.3. Normalidad en efecto aleatorio y el error crudo

A continuación se muestran tres casos donde se grafican cuantiles normales realizados sobre los errores del modelo lineal mixto con la descomposición de Choleski, y de los componentes de error y efecto aleatorio por separado. Adicionalmente, se muestra el resultado de la prueba de normalidad de Shapiro Wilks.

- Caso: $bi \sim N(0, 9,5)$, $\varepsilon \sim SN(0, 0,5, 12)$, $n=50$

La figura 5.7 muestra los cuantiles normales para las variables escaladas de los errores totales, los crudos y los efectos aleatorios estimados con ambos modelos, al aplicar la descomposición de Choleski a los residuales de un modelo lineal mixto. En ésta se ve que el comportamiento de los errores totales es un poco similar al del efecto obtenido por el estimador de Bayes empírico, en sus colas.

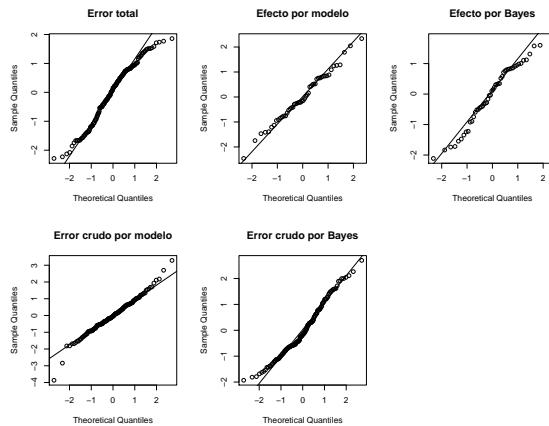


Figura 5.7: QQplot de error y efectos aleatorios, $bi \sim Normal$, $\varepsilon \sim SN$.

Variable	Valor p
Error total, normalizado	0.36
Efecto aleatorio, extraído del modelo estimado	0.97
Efecto aleatorio, estimador de Bayes Empírico	0.19
Error crudo, extraído del modelo estimado	0.077
Error crudo, extraído con el estimador de Bayes Empírico	0.008

Tabla 5.7: Diagnósticos de normalidad, caso $b_i \sim Normal$, $\varepsilon \sim SN$.

Se detecta normalidad en los componentes del efecto aleatorio, pero no en el error estimado por Bayes, con un nivel de significancia de 0.05, como se ve en la tabla 5.7. Así que en este caso, los componentes para los que se detectó un correcto ajuste a la distribución normal reflejan el escenario fijado previo a la estimación, indicando así una correcta identificabilidad del comportamiento distribucional.

- Caso: $b_i \sim SN(0, 9,5, 600)$, $\varepsilon \sim N(0, 0,5)$, $n=50$.

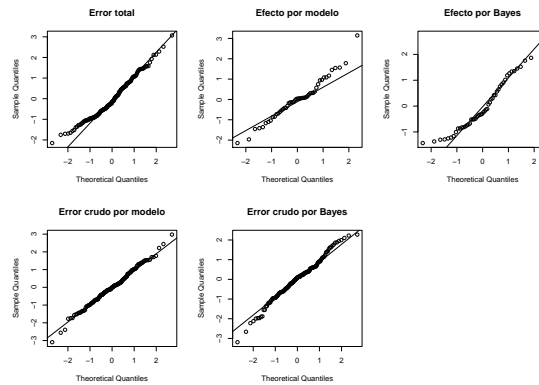


Figura 5.8: QQplot de error y efectos aleatorios, $b_i \sim SN$ y $\varepsilon \sim Normal$.

Los errores totales tienen aparente normalidad, y su comportamiento, según esta gráfica es similar al efecto calculado con el estimador de Bayes empírico en la cola izquierda (figura 5.8), sin embargo, los puntos de los errores totales se encuentran más alineados, pareciéndose un poco al error crudo.

Variable	Valor p
Error total, normalizado	0.59
Efecto aleatorio, extraído del modelo estimado	0.35
Efecto aleatorio, estimador de Bayes Empírico	0.026
Error crudo, extraído del modelo estimado	0.87
Error crudo, extraído con el estimador de Bayes Empírico	0.23

Tabla 5.8: Diagnósticos de normalidad, $b_i \sim SN$, $\varepsilon \sim Normal$.

En este caso la normalidad se presenta para los residuales totales, y el error en ambos casos pero no para el efecto encontrado con el estimador de Bayes empírico (ver tabla 5.8), hechos que también reflejan el comportamiento distribucional fijado antes de la estimación del modelo, exceptuando el efecto estimado por el modelo, mostrando además cómo es posible que se presente normalidad cuando se combina la distribución normal para el error crudo ε , con la SN para el efecto b_i .

- Caso: $b_i \sim SN(0, 9,5, 12)$, $\varepsilon \sim SN(0, 0,5, 12)$, $n=50$.

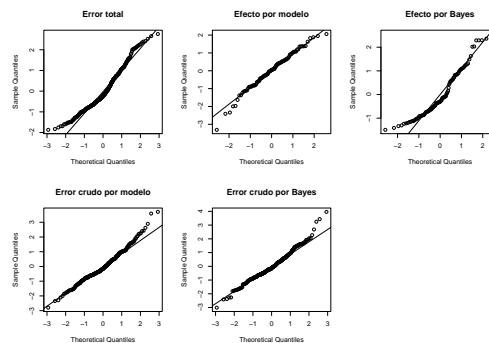


Figura 5.9: QQplot de error y efectos aleatorios, caso SN.

Variable	Valor
Error total, normalizado	0.00104
Efecto aleatorio, extraído del modelo estimado	0.285
Efecto aleatorio, estimador de Bayes Empírico	4.9e-05
Error crudo, extraído del modelo estimado	0.00068
Error crudo, extraído con el estimador de Bayes Empírico	0.00069

Tabla 5.9: Diagnósticos de normalidad, $b_i \sim SN$, $\varepsilon \sim SN$.

La figura 5.9 muestra un comportamiento de los residuales totales aparentemente no normal, lo que se ve reflejado en las colas y puede verse reflejado en el efecto aleatorio. Para cada variable, se aplicó la prueba de Shapiro Wilks, los valores p muestran los resultados vistos en la tabla 5.9.

La no normalidad se da en los residuales totales, en el efecto estimado por Bayes, y también en el error crudo en ambos casos, pero según la gráfica, las colas del error total son similares a las del efecto obtenido por Bayes. Lo anterior también refleja el control previo de las distribuciones para el error y el efecto aleatorio.

Cuando los escenarios se simulan 3000 veces, los resultados se ven en la tabla 5.10. En esta tabla es posible ver que la probabilidad de no rechazo de normalidad de los errores se ve similar al del efecto aleatorio en algunos escenarios simulados, como se describe a continuación.

5.3.1. Resultados de las simulaciones

En la tabla 5.10, $P1$ es la proporción de aceptación de normalidad $(1-\alpha)$ para el error completo, $Perm$ para el error crudo obtenido del modelo estimado, $Perb$ para el error obtenido con el estimador de bayes empírico, $Pefm$ para el efecto obtenido con el modelo $Pefb$ para el efecto obtenido con el estimador de bayes empírico.

$b_i \sim Normal$														
β_{0i}	β_{1i}	distr. ε	n	λ	σ^2	σ_b^2	$\hat{\beta}_0(sd)$	$\hat{\beta}_1(sd)$	sesgob0	P1	Perm	Perb	Pefm	Pefb
10	10	SN	100	200	1	9	10 (0.309)	10 (0.065)	0.002	0.51	0.72	0.71	1	0.945
					5	5	10.107 (0.408)	10.003 (0.15)	0.01	0.96	0.62	0.37	0.9	0.955
					9	1	10.123 (0.448)	9.993 (0.2)	0.012	0.95	0.695	0.84	1	0.95
30	30	SN	100	200	1	9	30.03 (0.338)	30 (0.068)	0	0.55	0.73	0.79	1	0.965
					5	5	30.11 (0.42)	30 (0.157)	0.004	0.93	0.65	0.3	0.9	0.945
					9	1	30.21 (0.485)	29.97 (0.214)	0.007	0.96	0.725	0.85	0.9	0.945
30	30	ST	100	200	1	9	30.11 (0.586)	30 (0.225)	0.004	0	0.01	0.26	0.5	0.7
					5	5	30.09 (1.254)	30.06 (0.538)	0.001	0.01	0.005	0	0.3	0.295
					9	1	30.23 (1.573)	29.99 (0.681)	0.007	0	0	0	0	0.055
10	10	SN	100	200	100	900	10.47 (3.3)	10.03 (0.68)	0.04	0.47	0.22	0.21	1	0.95
					500	500	11.17 (4.41)	10.01 (1.57)	0.11	0.97	0.505	0.21	0.9	0.95
					900	100	11.68 (4.99)	9.87 (2.27)	0.16	0.95	0.575	0.59	1	0.955
30	30	ST	30	200	1	9	30.24 (0.99)	29.96 (0.42)	0.01	0.23	0.145	0.55	0.6	0.855
					5	5	30.16 (3.65)	29.99 (0.89)	0.01	0.22	0.13	0.14	0.5	0.63
					9	1	30.61 (2.81)	29.9 (1.32)	0.02	0.19	0.12	0.17	0.4	0.47
100	200	ST	30	200	10	90	100.32 (3.04)	200.1 (1.2)	0	0.24	0.345	0.45	0.7	0.865
					50	50	100.99 (5.53)	200 (2.54)	0.01	0.23	0.235	0.25	0.6	0.66
					90	10	101.23 (8.7)	200.1 (4.25)	0.01	0.19	0.055	0.16	0.3	0.38
10	10	ST	30	600	100	900	12.657 (10.378)	9.526 (4.097)	0.255	0.19	0.345	0.22	0.8	0.86
					500	500	13.953 (20.679)	9.284 (11.581)	0.367	0.21	0.29	0.22	0.6	0.615
					900	100	16.008 (28.29)	9.282 (12.015)	0.567	0.2	0.215	0.19	0.5	0.48

Tabla 5.10: Diagnósticos de normalidad para residuales totales del modelo mixto.

En los casos en los que se fija el comportamiento $b_i \sim Normal$, $\varepsilon \sim SN$, la proporción de no rechazo de normalidad del error total es similar a la del efecto aleatorio, cuando $\sigma_b^2 = 5, 1$, sin embargo, cuando $b_i \sim Normal$, $\varepsilon \sim ST$, la proporción de no rechazo de normalidad en el error total es muy baja, siendo similar a la del error crudo. Es decir, en este caso, el comportamiento de la proporción de no rechazos del error total cuando $\varepsilon \sim SN$, pueden deberse al efecto, pero cuando $\varepsilon \sim ST$, podría ser similar al error crudo.

Gráficamente puede entonces explorarse el comportamiento de estas variables al extraerse del modelo estimado, y asociar al comportamiento del error total, una similitud con alguna de sus componentes, error crudo o efecto aleatorio, lo que abre la posibilidad a realizar algún tipo de correctivo para mejorar la estimación.

5.4. Problemas detectados

Los siguientes aspectos son los más importantes a señalar como problemas observados en el estudio de simulación realizado.

1. La estimación del intercepto β_0 en el modelo, se encuentra afectado tanto por la varianza del efecto aleatorio, como también por la distribución de la cual proviene el error o el efecto aleatorio. Además el parámetro λ de las distribuciones sesgadas, la varianza e incluso el valor del parámetro teórico del modelo, afectan también esta estimación. Sin embargo, el tamaño de muestra no tiene un efecto significativo.
2. El supuesto de normalidad del error y efecto aleatorio del modelo mixto es violado para los siguientes escenarios: distribución t sesgada para el error y el efecto aleatorio, normal sesgada para el efecto aleatorio y t sesgada para el error, y empeora cuando la varianza del modelo es alta. Sin embargo, este hecho no es tan grave cuando el efecto tiene una distribución normal sesgada y el error es normal, cuando se presenta alta varianza del modelo.
3. Cuando el efecto aleatorio proviene de una distribución normal, pero el error proviene de la distribución normal sesgada, el problema de violación de supuestos no es grave, hay adecuada normalidad, que se ve reflejado en una alta proporción de no rechazos de normalidad, sólo se ve mas afectado el sesgo de β_0 cuando aumenta la varianza.
4. Es posible observar la existencia de un problema en el sesgo relativo para la correlación estimada del modelo, donde una de las posibles causas es σ_b^2 , la cual al aumentar, genera una disminución de su magnitud, sin embargo, este problema no tiene otras claras fuentes de variación, por lo que podría ser explorado en otros trabajos posteriores.
5. Al realizar un diagnóstico en el comportamiento distribucional del error total podría mostrarse que su comportamiento es similar al efecto aleatorio o al error crudo en algunos escenarios, lo cual puede abrir una posibilidad para elaborar metodologías correctivas previas a la estimación bajo el supuesto de normalidad.

Capítulo 6

Alternativas de estimación en el modelo lineal mixto con respuesta continua

Como se ha encontrado hasta ahora, se ve la necesidad de realizar un mejoramiento en el sesgo del intercepto β_0 para el modelo lineal mixto propuesto, cuando el comportamiento del efecto y error aleatorio provienen de las distribuciones sesgadas utilizadas: SN, ST o su combinación con la normal. Para ello se ha mencionado que se proponen dos alternativas: una basada en una transformación sobre la variable respuesta y , y la otra basada en el uso del algoritmo EM basada en el trabajo de Arellano y otros (2005) (Ver Anexos).

En primer lugar se presentan los resultados de las transformaciones utilizadas (capítulo 3), sobre la estimación de los parámetros fijos, en segundo lugar, se presentan los resultados de utilizar el algoritmo EM para encontrar los parámetros fijos, la estimación de varianza y la estimación del sesgo. Posteriormente, se muestra una aplicación de dicho algoritmo para el conjunto de datos usado previamente.

6.1. Uso de transformaciones

En el primer escenario de la tabla 6.1 se ha fijado para ε y b_i un comportamiento normal, y se muestra la estimación de parámetros cuando $\beta_0 = 10$ y $\beta_1 = 10$, con la respuesta y transformada, arrojando un valor medio de: $\hat{\beta}_{0n}^* = 15,7$ y $\hat{\beta}_{1n}^* = 9,97$. De esta manera, el error relativo se obtuvo con respecto a los valores bajo normalidad, para otras combinaciones distribucionales, así: $Error_{b_0} = |\hat{\beta}_0^* - \hat{\beta}_{0n}^*| / \hat{\beta}_{0n}^*$ y $Error_{b_1} = |\hat{\beta}_1^* - \hat{\beta}_{1n}^*| / \hat{\beta}_{1n}^*$.

Aquí, $\hat{\beta}_{in}^*$ ($i = 0$ ó 1), es el parámetro estimado luego de transformar a y , pero fijando el caso Normal-Normal, y $\hat{\beta}_i^*$, es la estimación, luego de transformar a y , cuando se hace otra combinación distribucional.

n=100				$\lambda = 12$									
dis. ε	dis. b_i	β_{0i}	β_{1i}	ρ_i	σ_ε^2	$\sigma_{b_i}^2$	$\hat{\lambda}^*$	$\hat{\beta}_0^*$	$\hat{\beta}_1^*$	$\hat{\rho}^*$	$Error_{b_0}$	$Error_{b_1}$	P1
Normal	Normal	10	10	0.95	0.5	9.5	0.77	15.42 (0.44)	9.97 (0.75)	0.92 (0.01)	-	-	0.94
				0.9	1	9	0.77	15.51 (0.48)	9.97 (0.83)	0.66 (0.01)	-	-	0.94
				0.5	5	5	0.77	15.78 (0.68)	9.97 (1.3)	0.4 (0.05)	-	-	0.95
				0.25	7.5	2.5	0.76	15.97 (0.79)	9.97 (1.48)	0.13 (0.05)	-	-	0.95
				0.1	9	1	0.77	15.74 (0.85)	9.96 (1.66)	0 (0.04)	-	-	0.95
				0.05	9.5	0.5	0.77	15.73 (0.83)	9.96 (1.68)	-0.03 (0.03)	-	-	0.93
Normal	SN	10	10	0.95	0.5	9.5	0.66	23.66 (0.53)	9.96 (0.52)	0.82 (0.03)	0.5	0	0.43
				0.9	1	9	0.66	23.23 (0.57)	9.96 (0.58)	0.67 (0.07)	0.5	0	0.58
				0.5	5	5	0.72	19.86 (0.92)	9.96 (1.23)	0.08 (0.14)	0.26	0	0.87
				0.25	7.5	2.5	0.75	18.16 (0.97)	9.96 (1.51)	0.02 (0.05)	0.16	0	0.92
				0.1	9	1	0.77	16.99 (0.95)	9.98 (1.63)	-0.03 (0.04)	0.08	0	0.95
				0.05	9.5	0.5	0.77	16.61 (0.95)	9.97 (1.68)	-0.05 (0.04)	0.06	0	0.95
SN	SN	10	10	0.95	0.5	9.5	0.66	23.49 (0.52)	9.97 (0.5)	0.82 (0.03)	0.52	0	0.41
				0.9	1	9	0.66	23.28 (0.6)	9.96 (0.62)	0.67 (0.07)	0.5	0	0.58
				0.5	5	5	0.72	20.31 (0.92)	9.96 (1.18)	0.09 (0.14)	0.29	0	0.85
				0.25	7.5	2.5	0.75	18.44 (1.02)	9.96 (1.54)	0.02 (0.04)	0.16	0	0.94
				0.1	9	1	0.75	17.65 (0.96)	9.96 (1.54)	-0.03 (0.04)	0.12	0	0.95
				0.05	9.5	0.5	0.77	16.75 (1)	9.97 (1.74)	-0.05 (0.04)	0.07	0	0.94
Normal	ST	10	10	0.95	0.5	9.5	0.09	104.82 (1.92)	9.41 (0.73)	0.87 (0.05)	5.8	0.06	0.01
				0.9	1	9	0.07	112.37 (2.01)	9.31 (0.79)	0.67 (0.02)	6.25	0.07	0.02
				0.5	5	5	0.36	57.34 (2.17)	9.72 (1.34)	0.51 (0.05)	2.63	0.03	0.06
				0.25	7.5	2.5	0.53	39.03 (2.04)	9.91 (1.69)	0.22 (0.04)	1.44	0.01	0.23
				0.1	9	1	0.67	22.87 (1.7)	9.86 (1.92)	0.07 (0.04)	0.45	0.01	0.45
				0.05	9.5	0.5	0.71	20.24 (1.31)	9.96 (1.69)	0.01 (0.03)	0.29	0	0.59

Tabla 6.1: Resultados al realizar la transformación sobre la variable respuesta (1).

En la primera franja de la tabla 6.1 no se estima error relativo, ya que éste es precisamente el valor de referencia, es decir, los valores que deben ser obtenidos bajo normalidad con los cuales serán comparados los demás casos.

En la franja 2 de esta tabla se presentan los resultados cuando las variables b_i y ε tienen un comportamiento SN y normal, respectivamente, y bajo los mismos valores teóricos previos, se ve que el uso de la transformación de Box Cox mostró ser funcional cuando la varianza del efecto aleatorio es pequeña donde se reduce el error relativo de β_0 entre 6 % y 8 %. Sin embargo, en las franjas 3 y 4 se ve que el error relativo es mayor, por ejemplo, en la combinación $\varepsilon \sim ST$ y $b \sim Normal$, el error relativo sólo disminuye a 29 %.

		n=50		$\lambda = 200$									
$dis.\varepsilon$	$dis.b_i$	β_{0i}	β_{1i}	ρ_i	σ_e^2	σ_b^2	$\hat{\lambda}^*$	$\hat{\beta}_0^*$	$\hat{\beta}_1^*$	$\hat{\rho}^*$	$Error_{b0}$	$Error_{b1}$	P1
Normal	Normal	30	30	0.95	0.5	9.5	0.6	75.99 (0.63)	29.85 (0.57)	0.79 (0.04)	-	-	0.74
				0.9	1	9	0.61	75.11 (0.9)	29.86 (0.82)	0.69 (0.03)	-	-	0.89
				0.5	5	5	0.62	73.51 (1.59)	29.85 (1.41)	0.33 (0.09)	-	-	0.94
				0.25	7.5	2.5	0.62	73.72 (1.89)	29.84 (1.67)	0.11 (0.06)	-	-	0.94
				0.1	9	1	0.62	73.96 (1.8)	29.86 (1.62)	0 (0.06)	-	-	0.94
				0.05	9.5	0.5	0.62	73.86 (1.86)	29.86 (1.66)	-0.1 (0.05)	-	-	0.95
SN	ST	30	30	0.95	0.5	9.5	0.5	159.13 (3.31)	29.53 (1.56)	0.87 (0.05)	1.09	0.01	0.05
				0.9	1	9	0.51	160.52 (3.23)	29.72 (1.56)	0.67 (0.02)	1.14	0	0.18
				0.5	5	5	0.57	118.15 (2.89)	29.83 (1.82)	0.46 (0.1)	0.61	0	0.22
				0.25	7.5	2.5	0.59	96.1 (2.7)	29.75 (1.84)	0.19 (0.06)	0.3	0	0.45
				0.1	9	1	0.61	93.16 (2.48)	29.87 (1.87)	0.04 (0.06)	0.26	0	0.61
				0.05	9.5	0.5	0.61	80.48 (2.29)	29.83 (1.83)	0 (0.05)	0.09	0	0.73
SN	SN	30	30	0.95	0.5	9.5	0.6	80.79(0.06)	29.86(0.01)	0.59(0.07)	0.06	0	0.08
				0.9	1	9	0.6	80.59(0.36)	29.87(0.29)	0.43(0.07)	0.07	0	0.49
				0.5	5	5	0.61	78.69(1.55)	29.86(1.29)	0.04(0.09)	0.07	0	0.89
				0.25	7.5	2.5	0.62	76.32(1.88)	29.86(1.56)	0(0.07)	0.04	0	0.93
				0.1	9	1	0.62	75.87(1.92)	29.88(1.61)	-0.07(0.06)	0.03	0	0.95
				0.05	9.5	0.5	0.62	75.39(1.9)	29.86(1.59)	-0.08(0.06)	0.02	0	0.96
Normal	ST	30	30	0.95	0.5	9.5	0.52	154.96(2.94)	29.86(1.45)	0.87(0.06)	1.04	0	0.02
				0.9	1	9	0.51	178.26(3.17)	29.82(1.52)	0.67(0.03)	1.37	0	0.18
				0.5	5	5	0.55	164.3(3.12)	29.67(1.78)	0.47(0.09)	1.23	0.01	0.25
				0.25	7.5	2.5	0.58	116.02(3.02)	29.77(1.95)	0.19(0.06)	0.57	0	0.43
				0.1	9	1	0.61	83.62(2.24)	29.88(1.84)	0.04(0.06)	0.13	0	0.66
				0.05	9.5	0.5	0.61	80.29(2.28)	29.86(1.86)	-0.02(0.05)	0.09	0	0.71

Tabla 6.2: Resultados al realizar la transformación sobre la variable respuesta (2).

En la tabla 6.2 se combina el caso Normal-Normal, cuando los valores teóricos de los parámetros son 30, y en la franja 2, el error es SN con efecto ST, el mejoramiento del error relativo de β_0 se ve sólo cuando la varianza del efecto aleatorio es la menor (error relativo de 9 %). Para las mismas condiciones de las franjas 3 y 4, si se combina el error ST con efecto SN, se generan valores negativos en la respuesta, razones por las cuales no se puede realizar dicha transformación.

El error relativo es mejor en el caso SN-SN, pero no en el caso $\varepsilon \sim ST$ y $b \sim Normal$, el error

relativo sólo disminuye a 9 %.

Adicional a estos resultados, se aprecia que la proporción de no rechazo de normalidad se ve baja cuando una de las distribuciones es la ST, y cuando es mayor la varianza σ_b^2 . Lo anterior indica una no adecuada mejoría en la estimación en todos los escenarios donde la variable respuesta es positiva.

- Normalidad

La proporción de no rechazo de normalidad para el error total no se aprecia mejorar al usar la transformación. La figura 6.1 muestra el comportamiento de algunos de los resultados para 50 corridas, donde la línea continua, con valor 1, es el resultado posterior a la transformación y con 0, antes de esta, cada par con iguales características, aquí se aprecia que para los valores inferiores del parámetro de sesgo λ es un poco mayor la proporción de no rechazo de normalidad para el efecto aleatorio cuando los datos son transformados.

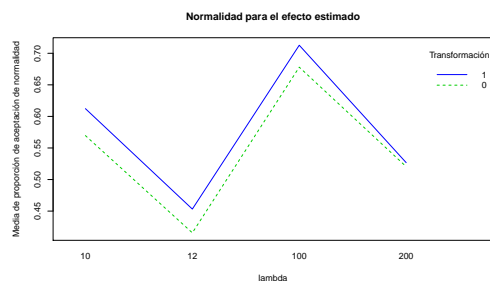


Figura 6.1: Comportamiento del error vs parámetro de sesgo λ .

En la figura 6.2 se muestra que el mejoramiento en la normalidad del residual total es un poco más notable cuando $b_i \sim N$, en los otros casos no se aprecia.

En general, se encuentra que para la distribución t sesgada no hay adecuada mejoría aparente del no rechazo de normalidad de los errores por medio de esta transformación.



Figura 6.2: Comportamiento normalidad del error total vs distribución del efecto.

En los resultados vistos en general, es claro que es casi nula la mejoría de la normalidad, lo cual muestra una necesidad de cambiar la metodología de la estimación del modelo usando otras técnicas cuando este tipo de distribuciones se presentan.

6.2. Estimación del sesgo relativo con el uso de algoritmo EM

En esta sección se realizaron 1000 simulaciones de los escenarios fijando la distribución SN para ambos ε y b_i ó ST para ambos, notando una muy buena estimación de los parámetros del modelo. Las primeras dos franjas de la tabla 6.3 resumen estas estimaciones calculadas en el primer caso SN-SN y en la tercera, se dan las estimaciones para el caso ST-ST.

Puede notarse que el sesgo relativo de β_0 varía en el rango de 0 a 2 %, en el caso SN-SN entre las varianzas del error que van desde 0.5 a 9, para ambos valores teóricos: 10 y 30, o para el efecto aleatorio de 1 a 9.5. El sesgo relativo sólo aumenta a 4 % en el caso ST-ST con la mayor varianza del error, mientras el sesgo relativo de β_1 va de 0 a 1 % en todos los escenarios. Lo anterior indica una mejoría notable en la estimación de ambos coeficientes.

		$b_i, \varepsilon \sim SN \quad \lambda = 12$						
β_{0i}	β_{1i}	$\hat{\beta}_0$	$\hat{\beta}_1$	\hat{count}	$\hat{\sigma}_\varepsilon^2$	sesgobo	sesgob1	σ_i^2
		10.171	10.006	1389.4	0.4715	0.0171	0.0006	0.5
10	10	10.217	10.098	4765.3	4.7395	0.0217	0.0098	5
		10.126	10.113	11169.5	8.26	0.0126	0.011	9
		$b_i, \varepsilon \sim SN \quad \lambda = 200$						
30	30	30.15	30.02	1135.5	0.484	0.0049	0.0	0.5
		30.65	29.99	3972.2	4.92	0.0218	0.0003	5
		$b_i, \varepsilon \sim ST \quad \lambda = 200$						
20	20	19.907	19.84	1919	2.76	0.0046	0.0079	2.5
		19.21	20.1	3725.5	9.2	0.040	0.0048	9.5

Tabla 6.3: Resultados al realizar la estimación con el algoritmo EM.

6.3. Propuesta de una prueba de normalidad sesgada

En esta sección se muestra una propuesta de prueba de normalidad, basada en bandas de confianza en el gráfico de probabilidad de los errores.

La prueba se construye por medio del siguiente proceso:

1. Obtenga bandas de confianza bajo la distribución normal sesgada y otras bandas bajo la distribución normal estándar.
2. Obtenga cuantiles reales de los residuales bajo la distribución normal sesgada y otros bajo la distribución normal estándar.
3. Calcule la proporción de puntos que caen por fuera de las bandas normales y normales sesgadas.
4. Realice una prueba de hipótesis estadística de diferencia de proporciones entre los puntos que caen por fuera de las bandas normales y los puntos que caen por fuera de las bandas normales sesgadas.

La prueba se construye antes y después de la transformación, para la distribución normal sesgada y la normal, con parámetros fijos.

El proceso para una sola corrida, puede arrojar los resultados puntuales siguientes:

En una transformación realizada con el uso de la media geométrica, bajo la distribución normal sesgada con un sesgo de 10 para el efecto aleatorio y normal para el error, con 100 sujetos, con vector teórico $\underline{\beta} = (50, 50)$ con varianza del efecto aleatorio 0.5 y una correlación intraclass de 0.05, se acepta la normalidad del modelo en todas las variables del efecto y el error estimados, además, se encuentra que la proporción de puntos por fuera de las bandas normales (0.023) es menor que la de los puntos por fuera de las bandas normales sesgadas (0.1467), con un $valor p = 1,36E - 07$.

La figura 6.3 muestra las bandas estimadas con la distribución normal sesgada para el efecto aleatorio, obtenido con el estimador de Bayes empírico.

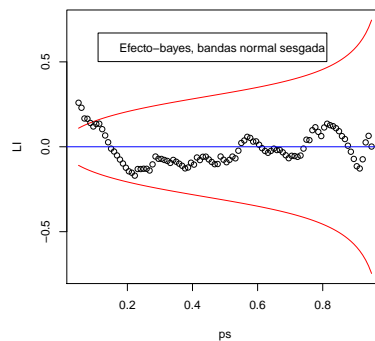


Figura 6.3: Gráfico de probabilidad con bandas de la normal sesgada para el efecto aleatorio.

Son pocos los puntos por fuera de las bandas, lo cual podría indicar buen ajuste a la distribución normal sesgada.

La figura 6.4 muestra las bandas estimadas con la distribución normal sesgada para los cuantiles del error crudo estimado por el estimador de Bayes empírico, en la cual tampoco son muchos los puntos por fuera de éstas.

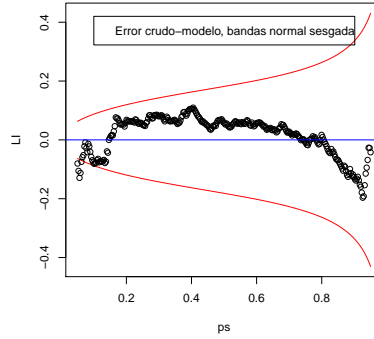


Figura 6.4: Gráfico de probabilidad con bandas de la normal sesgada para el error crudo.

La figura 6.5 muestra las bandas estimadas con la distribución normal para el error total estimado, la cual muestra también una cobertura aparentemente buena de los puntos.

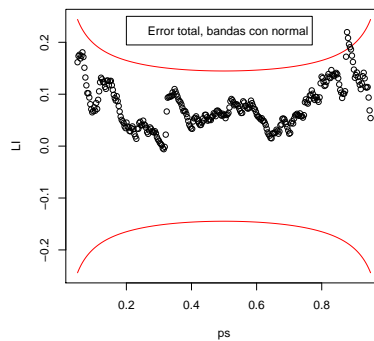


Figura 6.5: Gráfico de probabilidad con bandas de la normal para el error total.

La figura 6.6 muestra las bandas de los puntos para el error total estimado, cuando las bandas graficadas provienen de la distribución normal sesgada con parámetro de sesgo 10.

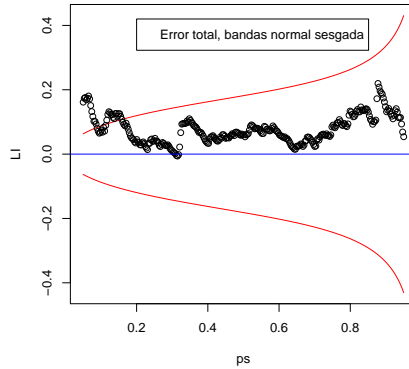


Figura 6.6: Gráfico de probabilidad con bandas de la normal sesgada para el error total.

De las anteriores dos gráficas se aprecia que en la que muestra las bandas bajo normalidad es menor la proporción de puntos por fuera de las bandas, que los que están por fuera de las bandas bajo la normal sesgada. Cuando este mismo escenario es repetido 3000 veces, los resultados numéricos muestran la proporción de veces en que los puntos por fuera de las bandas normales es menor que la proporción de puntos fuera de las bandas provenientes de la distribución normal sesgada (tabla 6.4), con $n=100$, $\beta_0, \beta_1 = 50$.

		$\beta_0 = 50$	$\lambda = 100$	$n=100$									
$disb_i$	$dis.\varepsilon$	ρ_i	σ_e^2	σ_b^2	$\hat{\beta}_0$ (sd)	sesgob0	P1	Perm	Perb	Pfm	Pfb	PropSN	PropN
		0.05	9.5	0.5	49.66 (0.03)	0.01	0.9	0.94	0.95	0.94	0.94	0.02	0.82
SN	Normal	0.1	9	1	49.66 (0.03)	0.01	0.89	0.95	0.95	0.96	0.96	0.02	0.81
		0.75	2.5	7.5	49.68 (0.02)	0.01	0.16	0.31	0.58	0.36	0.36	0.59	0.06

Tabla 6.4: Prueba de diferencia de proporciones.

En esta tabla, las columnas correspondientes a las proporciones de normalidad son: $P1$ es la frecuencia de aceptación de normalidad para el error total, $Perm$ y $Perb$, para el error crudo estimado con modelo o cuando se usa bayes, Pfm y Pfb para el efecto estimado con modelo o con bayes y las dos últimas columnas son la proporción de casos en que hay más puntos dentro de las bandas de confianza normal sesgada: $PropSN : PSN < PN$, esto es,

más puntos por fuera de las bandas normales, y en el segundo, más dentro de las bandas de confianza normales: $PropN : PSN > PN$. En la primera línea, cuando el valor de la varianza del efecto aleatorio es 0.5 ó 1, la proporción de aceptación de normalidad obtenida con Shapiro Wilks, es adecuada ($P1$ a Pfb) y los valores de $PropSN$ muestran que en un 82 % y 81 % de los casos hay más puntos contenidos en las bandas normales, lo que coincide con el resultado de la prueba de Shapiro Wilks, contrario a lo que ocurre con el resultado de no normalidad cuando la varianza del efecto aleatorio es 7.5, donde se detecta mayor la proporción de casos con distribución normal sesgada. Incorporando más resultados, puede apreciarse otro comportamiento de esta prueba.

		$\beta_0 = 50$	$n=100$	$\lambda = 100$										
dis. b_i	dis. ε	ρ_i	σ_e^2	σ_b^2	$\hat{\beta}_0$ (sd)	sesgob0	$P1$	$Perm$	$Perb$	Pfm	Pfb	$PropSN$	$PropN$	
		0.06	8.5	0.5	51.2 (0.49)	0.02	0.6	0.34	0.42	0.61	0.39	0.29	0.19	
ST	SN	0.11	8.5	1	51.43 (0.53)	0.03	0.48	0.18	0.19	0.51	0.21	0.4	0.21	
		0.47	8.5	7.5	54 (0.95)	0.08	0.1	0.02	0	0.05	0	0.69	0.07	
		0.05	9	0.5	51.11 (0.53)	0.02	0.49	0.44	0.6	0.67	0.42	0.21	0.18	
ST	SN	0.1	9	1	51.52 (0.52)	0.03	0.43	0.17	0.2	0.37	0.12	0.47	0.19	
		0.45	9	7.5	53.98 (0.72)	0.08	0.09	0.02	0	0.04	0	0.74	0.05	
		0.05	9.5	0.5	51.21 (0.63)	0.02	0.55	0.3	0.5	0.59	0.37	0.24	0.21	
ST	SN	0.1	9.5	1	51.58 (0.55)	0.03	0.4	0.16	0.22	0.33	0.16	0.42	0.16	
		0.44	9.5	7.5	54.08 (1.02)	0.08	0.13	0.02	0.01	0.05	0.01	0.75	0.09	
		0.94	0.5	8.5	54.05 (0.69)	0.08	0.02	0.01	0	0.02	0	0.89	0.01	
ST	SN	0.89	1	8.5	54.09 (0.72)	0.08	0.01	0	0	0	0	0.96	0	
		0.53	7.5	8.5	54.21 (0.99)	0.08	0.03	0	0	0.02	0	0.81	0.05	
		0.95	0.5	9	54.14 (0.63)	0.08	0.01	0	0	0.01	0	0.91	0.01	
ST	SN	0.9	1	9	54.28 (0.73)	0.09	0	0	0	0	0	0.94	0	
		0.55	7.5	9	54.38 (0.92)	0.09	0.03	0	0	0	0	0.73	0.08	
		0.95	0.5	9.5	54.25 (0.69)	0.08	0	0.02	0	0.01	0	0.95	0	
ST	SN	0.9	1	9.5	54.42 (0.85)	0.09	0	0	0	0	0	0.98	0	
		0.56	7.5	9.5	54.58 (1.05)	0.09	0.09	0.01	0	0.04	0	0.72	0.1	
		$\beta_0 = 50$	$n=100$	$\lambda = 10$										
dis. b_i	dis. ε	ρ_i	σ_e^2	σ_b^2	$\hat{\beta}_0$ (sd)	sesgob0	$P1$	$Perm$	$Perb$	Pfm	Pfb	$PropSN$	$PropN$	
		0.94	0.5	8.5	54.09 (0.82)	0.08	0.012	0.014	0	0.02	0	0.91	0.01	
ST	SN	0.9	1	9	54.27 (1.13)	0.09	0	0	0	0	0	0.97	0	
		0.56	7.5	9.5	54.5 (1.17)	0.09	0.08	0	0	0.02	0	0.73	0.11	

Tabla 6.5: Diferentes escenarios para la prueba de diferencia de proporciones.

En la tabla 6.5 se extienden las simulaciones en otros casos, en sus primeras 3 franjas se fijan valores para σ_e^2 : 9.5,9 y 8.5 y se mantienen fijos todos los demás parámetros y el escenario ST-SN, variando σ_b^2 así: 0.5,1 y 7.5. En las siguientes 3, se fijan de manera inversa, σ_b^2 : 9.5,9 y 8.5, variando σ_e^2 así: 0.5,1 y 7.5.

Se detectan proporciones altas de casos donde el error se distribuye normal sesgado (*PropSN* es mayor), que parece que aumentan a medida que la correlación intraclase fijada es mayor. Situación similar ocurre en la última franja, dejando ver una clara tendencia de la mezcla de variables con estas distribuciones hacia una distribución normal sesgada.

Lo anterior indica aciertos en la detección de normalidad o normalidad sesgada con la prueba propuesta, generando una alternativa para mostrar el comportamiento del residual total, pero también puede extenderse en cada uno de sus componentes. En la siguiente subsección se presenta esta prueba usando datos reales.

6.3.1. Aplicación

A continuación se muestra una aplicación de la prueba de normalidad sesgada propuesta al caso mostrado en el capítulo 3 usando los datos de la Tesis de grado de Muñoz (2004), que presenta un diseño experimental para determinar el comportamiento de unas especies de microalgas donde extraen diferentes variables respuesta en sus diseños, una de ellas fue la Densidad Celular: Promedio del número de células $\times 10^6/ml$.

En el capítulo 3 se mostraron los resultados al estimar un modelo lineal mixto con 5 medidas repetidas (5 días) donde la variable respuesta es Densidad Celular y la variable explicativa es el día, cuyo valor $P < 0,0001$ que indica su significancia sobre el crecimiento celular, sin embargo, no se encuentra normalidad ni para esta variable, ni al incorporar el factor 'medio'.

Para el modelo estimado se realizó un análisis de los residuales totales por medio del gráfico de las bandas, donde se aprecia cómo es posible una semejanza del comportamiento de éstos a una distribución normal sesgada, utilizando un parámetro de sesgo alto ($\lambda = 200$), donde

la cola derecha es la que parece encontrarse más cubierta por las bandas de confianza normal sesgada que la de la normal, como se observa en la figura 6.7.

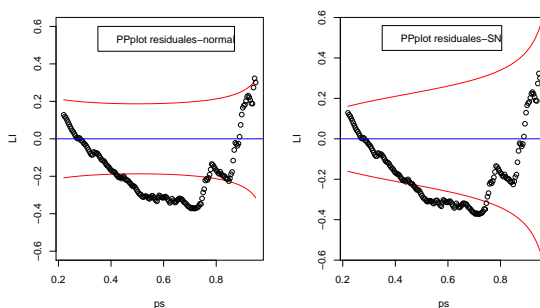


Figura 6.7: Gráfico de probabilidad normal y normal sesgada para residuales.

La prueba de diferencia de proporciones arroja un IC al 90 % para la diferencia: $(P_{sn} - P_n) = (-0,38, -0,015)$, lo cual indica que la proporción de puntos por fuera de las bandas normales es mayor que la de normales sesgadas.

Un poco diferente para el efecto aleatorio, en el cual, todos los puntos quedan dentro de las bandas bajo la distribución normal sesgada.

■ Estimación con el uso de transformación para el problema de la Densidad Celular

Al usar la transformación de Box Cox con la desviación estándar, se llega a un modelo donde los parámetros estimados cambian un poco, pero mejora la normalidad.

El *valorp* de la prueba de Shapiro Wilks, cambia a 0,3898. Y los parámetros fijos estimados, luego de retransformarlos (ver anexo 4), cambiaron según se aprecia en la tabla 6.6, donde la mayor magnitud del cambio se refleja en el intercepto estimado, indicando que el sesgo aparente en los parámetros fue modificado con la transformación.

Coefficiente	Antes de transformar	Después de transformar
β_0	-1.856	-2.91
β_1	5.39	5.328

Tabla 6.6: Cambio en los coeficientes del modelo lineal mixto para la respuesta Densidad Celular.

En los gráficos de las bandas (figura 6.8) se observan más puntos dentro de las bandas normales, y al hacer la prueba de diferencia de proporciones, éstas dan iguales con un nivel de confianza del 95 %: $(-0,02, 0,01)$ el IC incluye el 0, lo cual sugiere que el supuesto de normalidad puede ser viable ahora con transformación, ya que las proporciones de puntos por fuera de las bandas son de 0 % para la SN y 0,006 % para la normal.

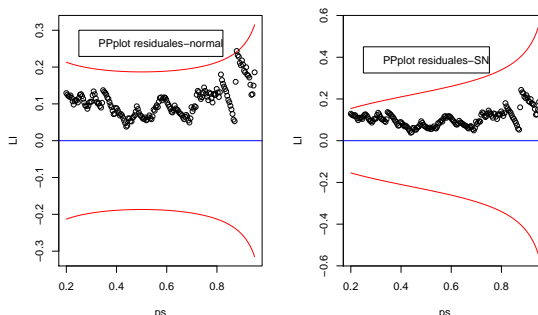


Figura 6.8: Gráfico de probabilidad normal y normal sesgada para residuales.

■ Estimación alternativa con el algoritmo EM para el problema de la Densidad Celular

Al realizar la estimación de los parámetros del modelo se consideró que el parámetro $\lambda = 200$, por el comportamiento de los residuales en las bandas SN, y los valores de σ_e^2 y σ_b^2 , son 64.1 y 65, a partir de estos valores, el parámetro que cambia es el intercepto, por medio del algoritmo EM: $\hat{\beta}_0 = -8,732$, el otro permanece igual: $\hat{\beta}_1 = 5,39$, las iteraciones que realiza van entre 3200 y 4581, para encontrar estos valores (tabla 6.7).

Coficiente	Estimación inicial	Después de transformar	Estimación con algoritmo EM
β_0	-1.856	-2.91	-8.7323
β_1	5.39	5.328	5.39

Tabla 6.7: Tres estimaciones de los parámetros del modelo lineal mixto para la Densidad Celular.

Para los modelos estimados, se aprecia el cambio en el parámetro β_0 principalmente, lo cual indica que el sesgo estimado de la distribución que fue detectado por medio del método de las bandas sirvió para realizar una mejor estimación de los efectos fijos del modelo lineal mixto considerado para estos datos.

La retransformación en el segundo caso, se realizó usando el proceso visto en el anexo 4.

Capítulo 7

Conclusiones y Recomendaciones

1. Se ha encontrado un problema en el sesgo relativo del intercepto β_0 cuando las distribuciones del efecto y error aleatorios del modelo lineal mixto provienen de combinaciones distribucionales diferentes al caso Normal-Normal: SN-Normal, SN-SN, SN-ST, ST-Normal, ST-ST. Cuando la varianza del efecto aleatorio es mayor, el sesgo de relativo de β_0 también lo es. También existe un efecto del tipo de distribución del efecto b_i o error ε y su parámetro de sesgo cuando pertenece a una familia sesgada, cuando λ es mayor, el sesgo relativo también, mas no tiene efecto el tamaño muestral (capítulos 4 y 5). Sólo decrece el sesgo de β_0 con el aumento del parámetro teórico, de lo contrario, los demás factores ejercen un efecto de incremento significativo: σ_b^2 , λ , las distribuciones *SN* o *ST* del error o del efecto aleatorio. En resumen, según los resultados encontrados puede afirmarse que la estimación de efectos fijos del modelo lineal mixto con intercepto aleatorio y con estructura de simetría compuesta, se ve alterada por comportamientos distribucionales diferentes al caso Normal.
2. El supuesto de normalidad del error y efecto aleatorio del modelo mixto es violado para los siguientes escenarios: distribución t sesgada para el error y el efecto aleatorio, normal sesgada para el efecto aleatorio y t sesgada para el error, y empeora cuando la varianza del modelo es alta. Sin embargo, este hecho no es tan grave cuando el efecto tiene una

distribución normal sesgada y el error es normal, cuando se presenta alta varianza del modelo.

3. Cuando el efecto aleatorio proviene de una distribución normal, pero el error proviene de la distribución normal sesgada, el problema de violación de supuestos no es grave, hay adecuada normalidad, que se ve reflejado en una alta proporción de no rechazos de normalidad, sólo se ve mas afectado el sesgo de β_0 cuando aumenta la varianza.
4. La transformación usada sólo mejora un poco el sesgo relativo del intercepto cuando la varianza del efecto aleatorio no es muy grande y cuando las distribuciones se combinan como normal sesgada para ambos efecto y error, o normal sesgada con normal. Sin embargo, se encontró una mejor alternativa de solución, que utiliza el algoritmo EM para encontrar las estimaciones del modelo lineal mixto. Este algoritmo puede ser mejorado para que sea aplicado en cualquier tipo de modelo lineal mixto, de manera que permita su fácil utilización.
5. Si bien, no es el propósito de este trabajo, fue observado un problema de sesgo en la estimación de la correlación intraclase, problema que puede ser mejorado en futuros trabajos.
6. En la aplicación de la prueba de sesgo, fue adecuada su funcionalidad. Esto permite tener una alternativa para su utilización en la evaluación del comportamiento de residuales del modelo lineal mixto.

Bibliografía

1. Arellano, R.B. y otros (2005). Skew-normal Linear Mixed Models. *Journal of Data Science*. 3. pp. 415-438.
2. Azzalini, A. (1985). A class of distribution which includes the normal ones. *Scand. J. Statist.* 12. pp. 171-178
3. Azzalini A. y Dalla Valle A., (1996). The multivariate skew-normal distribution. *Biometrika*. Vol 83, No. 4. pp. 715 - 726.
4. Azzalini A. y Capitanio A., (1999). Statistical applications of the multivariate skew-normal distribution. *J.Roy.Statist.Soc, series B*. Vol 61. pp. 579 - 602.
5. Azzalini A. y Capitanio A., (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *J. Roy. Statist. Soc. Ser. B*.65. pp. 367 389.
6. Banerjee M. y Frees E., (2009). Diagnostics for Linear Longitudinal Models. *Journal of the American Statistical Association*. Vol 92, No. 439. pp. 999-1005
7. Box G. y Cox D., (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 26, No. 2. pp. 211-252.
8. Cassella G., Berger R., (2001). Statistical Inference. *Duxbury Advanced Series*.2nd Ed. Thomson.
9. Correa J.C. (2008). Introducción a los modelos mixtos. [Notas de clase]. Universidad Nacional de Colombia, Sede Medellín.

10. Fernández C., Steel, M. F. J., (1999). Multivariate Student-regression models: pitfalls and inference. *Biometrika*.86. pp. 153 - 168.
11. Gurka, M. y otros., (2006). Extending the Box-Cox transformation to the linear mixed model. *J. R. Statistical Soc. A*. pp.273-288.
12. Gurka, M. y otros., (2007). Testing transformations for the linear mixed model. *Computational Statistics and Data Analysis*. No. 51. pp. 4297-4307.
13. Laird N.M., Ware J., (1982). Random effects models for longitudinal data. *Biometrics*. Vol 38. pp. 963 - 974.
14. Lange N. y Ryan L., (1989). Assessing normality in random effects models. *Ann. Statist.* 17. pp. 624 - 642.
15. Jara A. et al.,(2008). Linear mixed models with Skew-T elliptical distributions: A bayesian approach. *Computational Statistics and Data Analysis*, 52. pp. 5033 - 5045.
16. McCulloch C.E., Searle S.R., (2000). Generalized, Linear and Mixed Models. *Wiley Series in Probability and Statistics*. pg. 156.
17. Muñoz H., (2004). Estudio *In Vitro* de los efectos e interacciones ambientales en el crecimiento y la producción de ácidos grasos poliinsaturados de las microalgas marinas. *Tesis de Grado*. Universidad Nacional de Colombia.
18. Piepho H. P., McCulloch C.E., (2004). Transformations in Mixed Models: Application to Risk Analysis for a Multienviroment Trial. *American Statistical Association and the International Biometric Society*. Vol 9. No 2. pp 123-137.
19. Valencia M., Salazar J.C., (2009). Evaluación del impacto de un programa de bienestar sobre una comunidad en Colombia usando un modelo para datos correlacionados. [En revisión, FSP].
20. Verbeke G., Lesaffre E., (1996). A linear mixed effects model with heterogeneity in the random effects population. *J. Amer. Statist. Assoc.* 91. pp. 217-221.
21. Verbeke G., Molenberghs, G. (2001) *Linear Mixed Models for Longitudinal Data*. Springer, NY.

22. Zeger S.L., Liang K.Y., Albert, P.S., (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 44. pp. 1049 - 1060.
23. Zhang D., Davidian M., (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*. 57. pp. 795 - 802.
24. Zhou T., He X., (2007). Three-step estimation in linear mixed models with skew-t distributions. *Journal of Statistical Planning and Inference*. 138. pp. 1542 - 1555.

ANEXOS

1. Funciones de estimación del modelo lineal mixto.

```
#####  
#Valores de entrada para ejecutar funciones #####  
#Estructura de Simetria compuesta:cs, rhoi=Di/(Di+sig)#####  
#sig (varianza de error), Di (varianza de efecto aleatorio)#####  
#Función de ejecución de las funciones de estimación: ejectran####  
#Función para la estimacion del efecto aleatorio: efectosujeto####  
#z: Creación de la matriz Z de diseño del efecto aleatorio#####  
#Simulación de parámetros del modelo lineal mixto, bajo diferentes distribuciones del efecto aleatorio\\  
#Error bajo 3 distribuciones: normal, t y normal sesgada\\  
require(nlme)           #Paquete para el modelo lineal mixto  
require(mnormt)        #Paquete para generar normalidad multivariada  
require(MASS)          #Paquete para generar normalidad multivariada  
require(sn)            #Paquete para la distribución skew normal  
require(car)  
  
#Función para estimar el efecto aleatorio b por individuo según la matriz de covarianza  
#Argumentos de entrada  
#Zi: matriz para el iesimo sujeto, orden kxr(#repeticiones x #efectos aleatorios),  
#r=1 1 efecto aleatorio  
#N: numero de individuos  
#Di: Varianza del ef aleatorio del i-esimo sujeto  
#dis: Distribución que tendrá en efecto aleatorio: 0 si es normal, 1 si es normal sesgada, 2 si es t  
  
#Inicio  
efectosujeto=function(i,Di,dis,a,v){  
  r=1  
  if(dis==0){  
    efesujeto<-matrix(rmnorm(1,0,Di),nrow=1,ncol=1)  
  }  
  if(dis==1){  
    efesujeto=matrix(rmsn(1,xi=0,Omega=Di,alpha=a),ncol=1,nrow=1) #Skew normal  
  }  
  if(dis==2){  
    Si=(v-2)*Di/v #La matriz S dentro de la función rmt es matriz escala, la varianza es S(v/(v-2)), Di=5, se despejar S  
    efesujeto=matrix(rmt(n =1, mean =0, S=Si, df=v),ncol=1,nrow=1) #distribución t  
  }  
  if(dis==3){  
    efesujeto=matrix(rmst(1,xi=0,Omega=Di,alpha=a,df=v),ncol=1,nrow=1) #Skew t  
  }  
  return(efesujeto)  
}#Fin  
  
#Función para generar matriz de diseño X  
MX=function(k,n){  
  if(k==2){  
    X=matrix(c(rep(1,k*n),rep(c(1,0),n)),k*n,2)  
  }  
  if(k==3){  
    X=matrix(c(rep(1,k*n),rep(c(1,2,3),n)),k*n,2)  
  }  
  return(X)  
}#Fin  
  
#Funcion principal #Simula y, entrega coeficientes, argumentos de entrada:  
#X: matriz de diseño, de nk x p (num parámetros del modelo)  
#Z: matriz asociada con el efecto aleatorio kn x rn (donde r es num de efectos aleatorios, k repeticiones, n individuos)
```

```

#Zbe: producto de Z por el efecto aleatorio
#bet: Vector de coeficientes de los efectos fijos
#R: matriz de varianza del error nk x nk
#Nsim: Numero de simulaciones
#repetido: vector de entrada de medidas repetidas segun sujeto

parametro=function(X,bet,n,repetido,persona,temp,Zbe,sig,k,er,rhoi){
bet=matrix(bet,ncol=1)
#Vector simulado de valores de la variable respuesta
y=X%*%bet+Zbe+er
x1=X[,2]
#Ahora la estimacion del modelo mixto
tall1<-data.frame(y=y,x1=x1) #Data frame para lectura de funcion lme
tall2<-groupedData(y~x1|repetido,
data = as.data.frame(tall1)) #Archivo para lectura en funcion lme
model1<-lme(y~x1, data=tall2,random=~1,corCompSymm(rhoi)) #Estimacion de modelo con intercepto aleatorio, accion en indicadoras
list(modelo=model1,y=y)
} #Fin de la funcion parámetro

#Funcion para generar error aleatorio er
#Inicio
error=function(sig,distr,n,a,v){
R=sig*diag(k*n)
if(distr==0){
er=matrix(rmnorm(k*n,c(rep(0,k*n)),R),ncol=1,nrow=k*n)
}
if(distr==1){
er=matrix(rmsn(k*n,xi=c(rep(0,k*n)),Omega=R,alpha=c(rep(a,k*n))),ncol=1,nrow=k*n)
}
if(distr==2){
er=matrix(rmst(k*n,xi=c(rep(0,k*n)),Omega=R,alpha=c(rep(a,k*n)),df=v),ncol=1,nrow=k*n)
}
return(er)
}
#Fin

#Función para generar Zbe knxr para el modelo (r es num de efectos aleatorios)
#Zi: matriz para un sujeto de orden kx1
#Di: matriz o valor de varianza de 1 efecto aleatorio de orden rxr
paraz=function(Di,dis,k,a,v){
#Zi=t(c(rep(1,k)))
Zb=c(0)
for(i in 1:n){
efectob=rep(efectosujeto(matrix(ncol=1,nrow=3),Di,dis,a,v),3)
Zb=t(c(Zb,efectob))
}
Zbe=matrix(t(Zb[-1]),nrow=k*n,ncol=1)
return(Zbe)
}#fin

#FINALMENTE SE EJECUTAN LAS FUNCIONES

```

2. Programa para estimación del modelo con generación de gráficas de bandas

```
#Gráfico y estimación de prueba con bandas normales sesgadas, con parámetro de sesgo  $\lambda = 10$  #####  
#Programador: Marisol Valencia Cárdenas, basado en programa para bandas normales de Correa, J.C. ####  
#Inicio#####
```

```
grafico2.QQ<-function(x,distri='SN'){  
  if(distri=='SN'){  
    y<-(x) #x previamente transformado.  
    ps<-seq(0.05,0.95,length=length(y))  
    xi.obs<-quantile(y,probs=ps)  
    sd.xis<-sqrt(ps*(1-ps)/(length(x)*dsn(qsn(ps,location=0,scale=1,shape=10))^2))  
    dif<-xi.obs-qsn(ps)  
    LI<- -2*sd.xis  
    LS<- -LI  
    fi=length(which(dif>LS))  
    fs=length(which(dif<LI))  
    plot(ps,LI,type='l',ylim=c(min(LI),max(LS)),col='red')  
    points(ps,LS,type='l',col='red')  
    points(ps,dif)  
    con=cbind(fi,fs)  
    return(con)  
  }  
}
```

3. Fórmulas para la construcción de gráfico de probabilidad de bandas normales o normales sesgadas¹

Sea p : proporción para hallar los cuantiles de la variable de interés: $x_i = Q_p$.

Q : Cuantiles a graficar

$$Z_n = F^{-1}(p)$$

$$X_i = Q_p$$

$$Q = X_i - Z_n$$

$$Li = -2 \sqrt{\frac{p(1-p)}{nF(Z_n)}}$$

$$Ls = -Li$$

4. Estimación del parámetro real.

La propuesta realizada por Gurka et al (2006) de retransformación del valor real, cuando se aplica la transformación de Box Cox, no funciona adecuadamente en este modelo

¹Basado en gráfico de probabilidad normal inicialmente propuesto por Correa, J. C. (2009), en revisión. Universidad Nacional de Colombia.

lineal mixto. Para la aplicación de datos reales se propuso el siguiente proceso de re-transformación, pero no para las simulaciones:

Se parte de la siguiente estimación del modelo cuando $\lambda \neq 0$:

$$\frac{(\mathbf{X}\hat{\underline{\beta}} + \mathbf{Z}\hat{b} + \hat{\varepsilon})^\lambda - 1}{\lambda \tilde{y}^{\lambda-1}} = \mathbf{X}\hat{\underline{\beta}}^*$$

De aquí se despeja el vector real de parámetros $\underline{\beta}$, y se utiliza la matriz $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ para encontrar:

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \hat{\underline{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T [(\lambda \tilde{y}^{\lambda-1} \mathbf{X} \hat{\underline{\beta}}^* + 1)^{1/\lambda} - \mathbf{Z}\hat{b} - \hat{\varepsilon}]$$

$$\hat{\underline{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \left[(\lambda \tilde{y}^{\lambda-1} \mathbf{X} \hat{\underline{\beta}}^* + 1)^{1/\lambda} - \mathbf{Z}\hat{b} - \hat{\varepsilon} \right]$$

5. Algoritmo EM para estimar un modelo lineal mixto

```
matrix.sqrt <- function(A)
{
  sva <- svd(A)
  if (min(sva$d)>=0)
    Asqrt <- sva$u%*%diag(sqrt(sva$d))%*%t(sva$v) # svd e decomposição espectral
  else
    stop("Matrix square root is not defined")
  return(Asqrt)
}

#####
#Information Matrix
#####

matriznu<-function(nj,y,x,z,betal,sigmae,Dl,lambda,nu,Ind) {

# MLE of theta: beta(vector),sigmae(random errors),Dl(matrix),lambda(vector) and nu(vector or scalar).
# nj: vector with size of clusters (vector )
# y: vector of Responses (size=sum(nj) x 1)
# x: Design matrix of fixed effects (size=sum(nj) x p)
# z: Design matrix of random effects (size=sum(nj) x q)
# Ind: if 1- skew-normal; 2- skew- t; 3- skew-slash; and 4- skew-contaminated normal
# mont: Number Monte Carlo to find the integration of the Observed Information Matrix

mont<-1000
m<-n #Num de sujetos
```

```

N<-sum(nj) #Total de datos en y
p<-dim(X)[2]
q1<-dim(z)[2]
sigmae2<-as.numeric(sig)
lambda<-as.vector(lambda)
beta1<-as.vector(beta)
D1<-as.matrix(D)

#F1<- matrix.sqrt(D1)
F1<- D1

if (Ind==1){
  nu<-as.numeric(nu)
  RN<-matrix(1,ncol=mont,nrow=m)
  jj=1
}
if (Ind==2){
  nu<-as.numeric(nu)
  digama1<-psigamma(nu/2)
  Raux1<-matrix(runif(mont*m),nrow=mont,ncol=m)
  jj<-1
  Raux<-Raux1/(1-Raux1)
  jacob<-(Raux+1)^2
}

if (Ind==3){
  nu<-as.numeric(nu)
  Raux<-matrix(runif(mont*m),nrow=mont,ncol=m)
  jj<-1
}

if (Ind==4){
  nu<-as.vector(nu)
  Raux<-matrix(runif(mont*m),mont,m)
  Raux2<-matrix(1,mont,m)
  id = (Raux < nu[1])
  Raux2[id] <-nu[2]
  RN<-Raux2
  jj<-2
}

sumal<-0
for (j in 1:m){
  y1=y[(sum(nj[1:j-1])+1) : (sum(nj[1:j])) ]
  x1=matrix(x[(sum(nj[1:j-1])+1) : (sum(nj[1:j]))],ncol=p)
  z1=matrix(z[(sum(nj[1:j-1])+1) : (sum(nj[1:j])) ],ncol=q1)
  Psi<-sigmae2*diag(nj[j])+z1%*t(z1)*(F1^2)[1]
  psi1<-solve(Psi)
  H<-(solve(F1))%*solve(F1)+(t(z1)*(sigmae2^(-1)))%*diag(nj[j])%*z1
  omega<-solve(H)
  a<-as.numeric((1+t(lambda))%*solve(F1)%*omega%*solve(F1)%*lambda)^(1/2))
  Aj<-as.numeric((t(lambda))%*F1%*t(z1)%*psi1%*(y1-x1%*beta1))*solve(a)
  dj<-as.numeric(t(y1-x1%*beta1)%*psi1%*(y1-x1%*beta1))

  #derivadas de Aj

  derAjbeta<--(t(x1)%*psi1%*z1%*F1%*lambda)/a

  derAjlambda<-as.vector(t(y1-x1%*beta1)%*psi1%*z1%*F1/a - ((a)^(-2))*Aj*lambda%*solve(F1)%*omega%*solve(F1))

```

```

derH<-sigmae2^(-2)*t(z1)%*diag(nj[j])%*z1
deromega<-solve(H)%*derH%*solve(H)
derasigmae<-as.numeric((1/2)*a^(-1)*t(lambda)%*solve(F1)%*deromega%*solve(F1)%*lambda)
derinvasigmae<-(a)^(-2) * derasigmae
derpsi<-diag(nj[j])
der<-psi1%*derpsi%*psi1
derAjsigmae<-as.numeric(lambda%*F1%*t(z1)%*der%*(y1-x1%*beta1)/a+t(lambda)%*F1%*t(z1)%*psi1%*(y1-x1%*beta1)%*derinvasigmae)

#trocar quando as dimensões forem diferentes
# B<-matrix(c(1, 0, 0, 0,0,0),ncol=3,nrow=3)
# Caa<-matrix(c(0,1,1,1,0,1,1,1,0),nrow=3,ncol=3)
# Daa<-matrix(c(0,0,0,0,0,0,0,1),nrow=3,ncol=3)
B<-c(1)
Caa<-c(1)
Daa<-c(1)

derinvF1<-solve(F1)%*B%*solve(F1)
derHalfa1<-derinvF1%*solve(F1) + solve(F1)%*derinvF1
deromegaalfa1<-solve(H)%*derHalfa1%*solve(H)
deraalfal<-as.numeric((1/2)*a^(-1)*t(lambda)%*derinvF1%*omega%*solve(F1)%*lambda+t(lambda)%*solve(F1)%*omega%*derinvF1%*lambda
+t(lambda)%*solve(F1)%*deromegaalfa1%*solve(F1)%*lambda)
derinvaalfal<-(a)^(-2) * deraalfal
deralfal<-z1%*(B%*F1+F1%*B)%*t(z1)
derpsialfal<-psi1%*deralfal%*psi1
derAjalfa1<-as.numeric((t(lambda)%*B%*t(z1)%*psi1%*(y1-x1%*beta1))/a +t(lambda)%*F1%*t(z1)%*derpsialfal%*(y1-x1%*beta1))/a
+ t(lambda)%*F1%*t(z1)%*psi1%*(y1-x1%*beta1)%*derinvaalfal)

derinvF2<-solve(F1)%*Caa%*solve(F1)
derHalfa2<-derinvF2%*solve(F1) + solve(F1)%*derinvF2
deromegaalfa2<-solve(H)%*derHalfa2%*solve(H)
deraalfal2<-as.numeric((1/2)*a^(-1)*t(lambda)%*derinvF2%*omega%*solve(F1)%*lambda+t(lambda)%*solve(F1)%*omega%*derinvF2%*lambda
+t(lambda)%*solve(F1)%*deromegaalfa2%*solve(F1)%*lambda)
derinvaalfal2<-(a)^(-2) * deraalfal2
deralfal2<-z1%*(Caa%*F1+F1%*Caa)%*t(z1)
derpsialfal2<-psi1%*deralfal2%*psi1
derAjalfa2<-as.numeric((t(lambda)%*Caa%*t(z1)%*psi1%*(y1-x1%*beta1))/a +t(lambda)%*F1%*t(z1)%*derpsialfal2%*(y1-x1%*beta1))/a
+ t(lambda)%*F1%*t(z1)%*psi1%*(y1-x1%*beta1)%*derinvaalfal2)

derinvF3<-solve(F1)%*Daa%*solve(F1)
derHalfa3<-derinvF3%*solve(F1) + solve(F1)%*derinvF3
deromegaalfa3<-solve(H)%*derHalfa3%*solve(H)
deraalfal3<-((1/2)*a^(-1)*t(lambda)%*derinvF3%*omega%*solve(F1)%*lambda+t(lambda)%*solve(F1)%*omega%*derinvF3%*lambda
+t(lambda)%*solve(F1)%*deromegaalfa3%*solve(F1)%*lambda)
derinvaalfal3<-(a)^(-2) * deraalfal3
deralfal3<-z1%*(Daa%*F1+F1%*Daa)%*t(z1)
derpsialfal3<-psi1%*deralfal3%*psi1
derAjalfa3<-as.numeric((t(lambda)%*Daa%*t(z1)%*psi1%*(y1-x1%*beta1))/a +t(lambda)%*F1%*t(z1)%*derpsialfal3%*(y1-x1%*beta1))/a
+ t(lambda)%*F1%*t(z1)%*psi1%*(y1-x1%*beta1)%*derinvaalfal3)
dim1<-q1*(q1+1)/2
derAjalfa<-matrix(c(derAjalfa1, derAjalfa2, derAjalfa3),ncol=1,nrow=dim1)

dim2<-p+1+q1+dim1
derAj<-matrix(c(t(derAjbeta), derAjsigmae, t(derAjalfa),derAjlambda),nrow=dim2,ncol=1)

#derivadas de dj
derdjbeta<-2*t(x1)%*psi1%*(y1-x1%*beta1)

```

```

derdjlamba<-as.vector(matrix(0,q1,1))

derdjsigmae<-as.numeric(t(y1-x1%%beta1)%%der%%(y1-x1%%beta1))
deralfa1<-z1%%(B%%F1+F1%%B)%%t(z1)
deralfa2=z1%%(Caa%%F1+F1%%Caa)%%t(z1)
deralfa3=z1%%(Daa%%F1+F1%%Daa)%%t(z1)

derpsialfa1<--psi1%%deralfa1%%psi1
derpsialfa2<--psi1%%deralfa2%%psi1
derpsialfa3<--psi1%%deralfa3%%psi1

derdjalfa1<-as.numeric(t(y1-x1%%beta1)%%derpsialfa1%%(y1-x1%%beta1))
derdjalfa2<-as.numeric(t(y1-x1%%beta1)%%derpsialfa2%%(y1-x1%%beta1))
derdjalfa3<-as.numeric(t(y1-x1%%beta1)%%derpsialfa3%%(y1-x1%%beta1))

derdjalfa<-matrix(c(derdjalfa1, derdjalfa2, derdjalfa3),nrow=dim1,ncol=1)

derdj<-matrix(c(t(derdjbeta), derdjsigmae, t(derdjalfa), t(derdjlamba)),nrow=dim2,ncol=1)

#derivatives of psi

derpsibeta<-as.vector(matrix(0,p,1))
derpsilambda<-as.vector(matrix(0,q1,1))
derpsisigmae<-sum(diag(psi1%%diag(nj[j]))))
derpsialfa1<-sum(diag(psi1%%z1%%(F1%%B+B%%F1)%%t(z1)))
derpsialfa2<-sum(diag(psi1%%z1%%(F1%%Caa+Caa%%F1)%%t(z1)))
derpsialfa3<-sum(diag(psi1%%z1%%(F1%%Daa+Daa%%F1)%%t(z1)))
derpsialfa<-matrix(c(derpsialfa1, derpsialfa2, derpsialfa3),ncol=1,nrow=dim1)

if (Ind==1){
  R<-RN[,j]
  derKiTheta<-0.5*mean(R^(nj[j]/2+1)*exp(-R*dj/2)*(pnorm(R^(1/2)*Aj,0,1)))*derdj+mean(R^(nj[j]/2+1/2)*exp(-R*dj/2)*dnorm(R^(1/2)*Aj,0,1))*derAj
  derlogpsi<-matrix(c(t(derpsibeta), derpsisigmae, t(derpsialfa), t(derpsilambda)),ncol=1,nrow=dim2)
  ValKi<-mean(R^(nj[j]/2)*exp(-0.5*R*dj)*pnorm(R^(1/2)*Aj,0,1))
  parE<-ValKi^(-1)*derKiTheta
}

if (Ind==2){
  derlogpsi<-matrix(c(t(derpsibeta), derpsisigmae, t(derpsialfa), t(derpsilambda), as.vector(matrix(0,1,j))),nrow=(dim2+1),ncol=1)
  w<-(nj[j]+1)/2
  w1<-(nj[j]+2)/2
  w2<-nj[j]/2
  derKiTheta<-((2^w)*nu^(nu/2)*gamma((nu+2*w)/2))/(sqrt(2*pi)*gamma(nu/2)*(dj+Aj^2+nu)^((nu+2*w)/2))*derAj
  -0.5*(((2^w1)*nu^(nu/2)*gamma((nu+2*w1)/2))/(gamma(nu/2)*(nu+dj)^((nu+2*w1)/2)))*pt(sqrt((nu+2*w1)/(dj+nu))*Aj,nu+2*w1))*derdj;
  ValKi<-((2^w2)*nu^(nu/2)*gamma(w2+nu/2)*pt(sqrt((nu+2*w2)/(dj+nu))*Aj,nu+2*w2))/(gamma(nu/2)*(nu+dj)^(w2+nu/2))
  parte1e4<-0.5*(log(nu/2)+1-digama1)*ValKi
  parte3<-0.5*((2^w1)*nu^(nu/2)*gamma(w1+nu/2)*pt(sqrt((nu+2*w1)/(dj+nu))*Aj,nu+2*w1))/(gamma(nu/2)*(nu+dj)^(w1+nu/2))
  hden<-(((nu/2)^(nu/2))*((Raux[,j])^(nu/2-1))*exp(-0.5*Raux[,j]*nu))/gamma(nu/2)
  parte2<-mean(0.5*Raux[,j]^(nj[j]/2)*exp(-0.5*Raux[,j]*dj)*pnorm(Raux[,j]^(1/2)*Aj,0,1)*log(Raux[,j])*hden*jacob[,j])
  derKinu<-parte1e4+parte3+parte2
  parE<-ValKi^(-1)*matrix(c(t(derKiTheta), t(derKinu)),nrow=(dim2+1),ncol=1)
}

if (Ind==3){
  w<-(nj[j]+1)/2
  w1<-(nj[j]+2)/2
  w2<-nj[j]/2
  w3<-(nj[j]/2)+nu-1

```

```

U <- runif(10000)
V <- pgamma(1,nu+w1, dj/2)*U
S <- qgamma(V,nu+w1, dj/2)

V1 <- pgamma(1,nu+w2, dj/2)*U
S1 <- qgamma(V1,nu+w2, dj/2)

V3 <- pgamma(1,nu+w3, dj/2)*U
S3 <- qgamma(V3,nu+w3, dj/2)

derKiTheta<-((nu*2^(nu+w)*gamma(nu+w)*pgamma(1,nu+w,(dj+Aj^2)/2))/(sqrt(2*pi)*(dj+Aj^2)^(nu+w))*derAj
- 0.5* (((2^(2+w1)*gamma(nu+w1)*pgamma(1,nu+w1,dj/2))/dj^(nu+w1))*mean(pnorm(S^(1/2)*Aj)))*derdj

derlogpsi<-matrix(c(t(derpsibeta), derpsisigmae, t(derpsialfa), t(derpsilambda), as.vector(matrix(0,1,jj))),nrow=(dim2+1),ncol=1)

derKinu<-mean(Raux[,j]^(nj[j]/2)*exp(-Raux[,j]*dj/2)*(pnorm(Raux[,j]^(1/2)*Aj,0,1))^(nu* Raux[,j]^(nu-1)*log(Raux[,j])) )
+ ((2^(2+w3)*gamma(nu+w3)*pgamma(1,nu+w3,dj/2))/dj^(nu+w3))*mean(pnorm(S3^(1/2)*Aj))

ValKi<-(((2^(2+w2)*gamma(nu+w2)*pgamma(1,nu+w2,dj/2))/dj^(nu+w2))*mean(pnorm(S1^(1/2)*Aj)))
parE<-ValKi^(-1)*matrix(c(t(derKiTheta), t(derKinu)),nrow=(dim2+1),ncol=1)
}
if (Ind==4){
R<-RN[,j]
derKiTheta<-0.5*mean(R^(nj[j]/2+1)*exp(-R*dj/2)*(pnorm(R^(1/2)*Aj,0,1))*derdj+mean(R^(nj[j]/2+1/2)
*exp(-R*dj/2)*dnorm(R^(1/2)*Aj,0,1))*derAj
derlogpsi<-matrix(c(t(derpsibeta), derpsisigmae, t(derpsialfa), t(derpsilambda), as.vector(matrix(0,1,jj))),nrow=(dim2+2),ncol=1)
derkinu1<-nu[2]^(nj[j]/2)*exp(-0.5*nu[2]*dj)*pnorm(nu[2]^(1/2)*Aj,0,1)-exp(-0.5*dj)*pnorm(Aj,0,1)
derkinu2<-nu[1]*(nj[j]/2)*nu[2]^(nj[j]/2-1)* exp(-0.5*nu[2]*dj)*pnorm(nu[2]^(1/2)*Aj,0,1)
+ nu[2]^(nj[j]/2)*exp(-0.5*nu[2]*dj)*(-0.5*dj)
* pnorm(nu[2]^(1/2)*Aj,0,1) +
nu[2]^(nj[j]/2)*exp(-0.5*nu[2]*dj)*pnorm(nu[2]^(1/2)*Aj,0,1)*0.5*Aj*nu[2]^(-1/2))
derKinu<-matrix(c(derkinu1,derkinu2),nrow=2,ncol=1)
ValKi<-mean(R^(nj[j]/2)*exp(-0.5*R*dj)*pnorm(R^(1/2)*Aj,0,1)
parE<-ValKi^(-1)*matrix(c(t(derKiTheta), t(derKinu)),nrow=(dim2+2),ncol=1)
}

derlog<-0.5*derlogpsi+parE
suma1<-suma1+derlog%*t(derlog)

}
desvio<- sqrt(diag(solve(suma1)))
return(desvio)
}
#####
#Log-likelihood: need the mvtnorm package
#####

graf4<-function(y1,med,Psi,Ajj,nu,pp) {
f <- function(u) 2*nu*u^(nu - 1)*((2*pi)^(-pp/2))*(u^(pp/2))*((det(Psi))^(-1/2))*exp(-0.5*u*t(y1-med)%*solve(Psi)%*(y1
-med))*pnorm(u^(1/2)*Ajj)
resp <- integrate(f,0,1)$value
return(resp)
}

veroST1<-function(Psi,Ajj,dj,nu,pp) {
aux1<-gamma((pp+nu)/2)*nu^(-pp/2)*(det(Psi))^(-1/2)*(1+dj/nu)^(-pp+nu/2)/(gamma(nu/2)*pi^(pp/2))
aux2<-sqrt(nu+pp)*Ajj/sqrt(dj+nu)
g1<-2*aux1*pt(aux2,nu+pp)

```



```

return(g1)
}

logvero<-function(nj,y,x,z,beta1,sigmae,D1,lambda,nu,Ind){
# MLE of theta: beta(vector),sigmae(random errors),D1(matrix),lambda(vector) and nu(vector or scalar).
# nj: vector with size of clusters (vector )
# y: vector of Responses      (size=sum(nj) x 1)
# x: Design matrix of fixed effects      (size=sum(nj) x p)
# z: Design matrix of random effects      (size=sum(nj) x q)
# Ind: if 1- skew-normal; 2- skew- t; 3- skew-slash; and 4- skew-contaminated normal
# mont: Number Monte Carlo to find the integration of the Observed Information Matrix

m<-length(nj)
N<-sum(nj)
p<-dim(x)[2]
q1<-dim(z)[2]

delta<-lambda/as.numeric(sqrt(1+t(lambda)%%lambda));
Deltab<-D1%%delta
Gammab<-D1-Deltab%%t(Deltab)
sumal<-0

if (Ind==1){
require(mvtnorm)
for (j in 1:m){
y1=y[(sum(nj[1:j-1])+1) : (sum(nj[1:j])) ]
x1=matrix(x[(sum(nj[1:j-1])+1) : (sum(nj[1:j]))], , ncol=p)
z1=matrix(z[(sum(nj[1:j-1])+1) : (sum(nj[1:j]))] , ncol=q1)
med<-x1%%beta1
Psi<-(z1)%%(Gammab+Deltab%%t(Deltab))%%t(z1)+sigmae*diag(nj[j])
dj<-as.numeric(t(y1-med)%%solve(Psi)%%(y1-med))
Mtj2<-(1+t(Deltab))%%t(z1)%%solve(sigmae*diag(nj[j])+z1%%Gammab%%t(z1))%%z1%%Deltab)^(-1)
mutj<-Mtj2*t(Deltab)%%t(z1)%%solve(sigmae*diag(nj[j])+z1%%Gammab%%t(z1))%(y1-med)
Ajj<-mutj/sqrt(abs(Mtj2))
sumal<-sumal+(log(2*dmvnorm(y1,med,Psi)*pnorm(Ajj,0,1)))
}
}

if (Ind==2){
for (j in 1:m){
y1=y[(sum(nj[1:j-1])+1) : (sum(nj[1:j])) ]
x1=matrix(x[(sum(nj[1:j-1])+1) : (sum(nj[1:j]))], , ncol=p)
z1=matrix(z[(sum(nj[1:j-1])+1) : (sum(nj[1:j]))] , , ncol=q1)
med<-x1%%beta1
Psi<-(z1)%%(Gammab+Deltab%%t(Deltab))%%t(z1)+sigmae*diag(nj[j])
dj<-as.numeric(t(y1-med)%%solve(Psi)%%(y1-med))
Mtj2<-(1+t(Deltab))%%t(z1)%%solve(sigmae*diag(nj[j])+z1%%Gammab%%t(z1))%%z1%%Deltab)^(-1)
mutj<-Mtj2*t(Deltab)%%t(z1)%%solve(sigmae*diag(nj[j])+z1%%Gammab%%t(z1))%(y1-med)
Ajj<-mutj/sqrt(abs(Mtj2))
sumal<-sumal+log(veroST1(Psi,Ajj,dj,nu,pp=nj[j]))
}
}

if (Ind==3){
for (j in 1:m){
y1=y[(sum(nj[1:j-1])+1) : (sum(nj[1:j])) ]
x1=matrix(x[(sum(nj[1:j-1])+1) : (sum(nj[1:j]))], , ncol=p)

```

```

z1=matrix(z[(sum(nj[1:j-1])+1) : (sum(nj[1:j]))] , , ncol=q1)
med<-x1%*beta1
Psi<-(z1)%*(Gammab+Deltab%*t(Deltab))%*t(z1)+sigmae*diag(nj[j])
Mtj2<-(1+t(Deltab))%*t(z1)%*solve(sigmae*diag(nj[j])+z1%*Gammab%*t(z1))%*z1%*Deltab)^(-1)
mutj<-Mtj2*t(Deltab)%*t(z1)%*solve(sigmae*diag(nj[j])+z1%*Gammab%*t(z1))%*(y1-med)
Ajj<-mutj/sqrt(abs(Mtj2))

pp<-nj[j]
sumal<-sumal + (log(graf4(y1,med,Psi,as.numeric(Ajj),nu,pp)))
}
}

}
return(sumal)
}

#####
#Algoritmo EM
#####
#carregar
library(mnormt)
require(mvtnorm)

EM.Skew<- function(nj,y,x,z,betas,beta1,sigmae,D1,lambda,nu,Ind,lb,lu,precisao,loglik=T,informa=T,calcbi=T){
# beta1,sigmae,D1,lambda: Initial values of : beta(vetor),sigmae(número),D1(matriz),lambda(vetor) e nu(número ou vetor, depende
#da distribuição considerada).
# nj: vector with size of clusters (vetor )
# y: vector of Responses (size=sum(nj) x 1)
# x: Design matrix of fixed effects (size=sum(nj) x p)
# z: Design matrix of random effects (size=sum(nj) x q)
# Ind: if 1- skew-normal; 2- skew- t; 3- skew-slash; and 4- skew-contaminated normal

# lb: lower limit to nu
# lu: upper limit for nu
# precisao: número de casas decimais dadas aos desvios
# loglik: TRUE or FALSE - If you need to calculate the log-likelihood
# informa: TRUE or FALSE - If you need to calculate the standard error
# print count: each iteration
# calcbi: True ou False - f you need to calculate the random effect bi

m<-length(nj)
N<-sum(nj)
p<-dim(x)[2]
q1<-dim(z)[2]

delta<-lambda/sqrt(1+t(lambda)%*lambda);
Deltab<-(D1)%*delta
Gammab<-D1-Deltab%*t(Deltab)
dim1<-q1*(q1+1)/2
neta<-(D1)%*lambda

teta <- c(beta1,sigmae,Gammab[upper.tri(Gammab, diag = T)],Deltab,nu)

criterio<-1
count<-0

if(calcbi==TRUE){
bi=matrix(0,nrow=m,ncol=q1)
}

```

```

precisão=0.00001

while(criterio > precisão){
  uj <- c(0)
  tj <- c(0)
  t2j <- c(0)
  sum1<- matrix(0,nrow=p,ncol=p)
  sum2<- matrix(0,nrow=p,ncol=1)
  sum3<-0
  sum4<-matrix(0,nrow=q1,ncol=q1)
  sum5<-matrix(0,nrow=q1,ncol=1)
  b<-matrix(0,nrow=q1,ncol=m)
  btj<-matrix(0,nrow=q1,ncol=m)
  Omega<-matrix(0,nrow=m*q1,ncol=q1)
  count <- count + 1
  for (j in 1:m ){

y1=y[(sum(nj[1:j-1])+1) : (sum(nj[1:j])) ]
x1=matrix(x[(sum(nj[1:j-1])+1) : (sum(nj[1:j]))], ,ncol=p)
z1=matrix(z[(sum(nj[1:j-1])+1) : (sum(nj[1:j]))] , ,ncol=q1)

med<-x1%*%beta1

Psi<-(z1)%*(Gammab+Deltab%*%t(Deltab))%*%t(z1)+sigmae*diag(nj[j])
dj<-as.numeric(t(y1-med)%*%solve(Psi)%*(y1-med))
Mtj2<-(1+t(Deltab)%*%t(z1)%*%solve(sigmae*diag(nj[j])+z1%*%Gammab%*%t(z1))%*%z1%*%Deltab)^(-1)
mutj=abs(Mtj2)*t(Deltab)%*%t(z1)%*%solve(sigmae*diag(nj[j])+z1%*%Gammab%*%t(z1))%*(y1-med)
Ajj<-mutj/sqrt(abs(Mtj2))
if(calcbi==TRUE){
mediab<-D1%*%t(z1)%*%solve(Psi)%*(y1-med)
Lambda<-solve(solve(D1)+t(z1)%*%z1/sigmae)}

if (Ind==1){
  uj[j]<-1
  esper<-as.numeric(dnorm(Ajj,0,1)/pnorm(Ajj,0,1))
  if(calcbi==TRUE){
bi[j,]<-mediab+Lambda%*%neta/as.numeric(sqrt(1+t(neta)%*%Lambda%*%neta))*as.numeric(dnorm(Ajj,0,1))/as.numeric(pnorm(Ajj,0,1))
}
}
if (Ind==2){
esper <-as.numeric(2*(nu/2)^(nu/2)*gamma((nj[j]+nu+1)/2)/(gamma(nu/2+1)*sqrt(2*pi)^(nj[j]+1)*det(Psi)^(1/2)*((dj+nu+Ajj^2)/2)^(nj[j]
+1+nu)/2)*veroST1(Psi,Ajj,dj,nu,nj[j]))
uj[j]<-2*(nu/2)^(nu/2)*gamma((nj[j]+nu+2)/2)*pt(sqrt((nj[j]+nu+2)/(dj+nu))*Ajj,nj[j]+nu+2)/(gamma(nu/2
+1)*sqrt(2*pi)^(nj[j])*det(Psi)^(1/2)*veroST1(Psi,Ajj,dj,nu,nj[j]))*((dj+nu)/2)^(nj[j]+2+nu)/2))
if(calcbi==TRUE){
esper2<-as.numeric(2*(nu/2)^(nu/2)*gamma((nj[j]+nu-1)/2)/(gamma(nu/2+1)*sqrt(2*pi)^(nj[j]+1)*det(Psi)^(1/2)*((dj
+nu+Ajj^2)/2)^(nj[j]-1+nu)/2)*veroST1(Psi,Ajj,dj,nu,nj[j]))
bi[j,]<-mediab+Lambda%*%neta/as.numeric(sqrt(1+t(neta)%*%Lambda%*%neta))*esper2
}
}

Tbj<-solve(solve(Gammab)+t(z1)%*%z1/sigmae)
r<-Tbj%*%t(z1)%*(y1-x1%*%beta1)/sigmae
s1<-(diag(q1)-Tbj%*%t(z1)%*%z1/sigmae)%*%Deltab
tj[j]<-uj[j]*mutj+(sqrt(abs(Mtj2)))*esper
t2j[j]<-uj[j]*mutj^2+Mtj2+sqrt(abs(Mtj2))*mutj*esper
b[,j]<-uj[j]*r+s1*tj[j]
btj[,j]<- r*tj[j]+s1*t2j[j]
Omega[(j-1)*q1+1:(j*q1),]<-Tbj+uj[j]*r%*%t(r)+s1%*%t(r)*tj[j]+r%*%t(s1)*tj[j]+s1%*%t(s1)*t2j[j]

```

```

sum1<-sum1+uj[j]*t(x1)%%x1
sum2=sum2+(t(x1)%%(uj[j]*y1-z1)%%b[,j])
sum3<-sum3+uj[j]*t(y1-x1)%%beta1)%%(y1-x1)%%beta1)-t(y1-x1)%%beta1)%%z1)%%b[,j]-t(b[,j])%*(z1)%%(y1-x1)%%beta1)+sum(diag(Omega[(j-1)*q1+1):(j*q1),])%*(z1)%%z1))
sum4<-sum4+Omega[(j-1)*q1+1):(j*q1),]-btj[,j])%*(Deltab)-Deltab%*(btj[,j])+2j[j]*Deltab%*(Deltab)
sum5<- sum5 + btj[,j]
}

beta1<-solve(sum1)%%sum2
sigmae<-as.numeric(sum3)/N
Gammab<-sum4/m
Deltab<-sum5/sum(t2j)
D1<-Gammab+Deltab%*(Deltab)
lambda<-solve(D1)%%Deltab/as.numeric(sqrt(1-t(Deltab)%%solve(D1)%%Deltab))

logvero1<-function(nj,y,x,z,beta1,sigmae,D1,lambda,nu,Ind){logvero(nj,y,x,z,beta1,sigmae,D1,lambda,nu,Ind)}

if (Ind==1){ nu<-0} else
{nu <- optim(nu,logvero1(nj,y,x,z,beta1,sigmae,D1,lambda,nu,Ind=2),gr = NULL,method = "L-BFGS-B", lower =lb,
upper = lu,control = list(fnscale=-1))$par
}
param <- teta
teta <- c(beta1,sigmae,Gammab[upper.tri(Gammab, diag = T)],Deltab,mu)
criterio <- sqrt((teta-param)%%(teta-param))

dd<-D1[upper.tri(D1, diag = T)]
count
beta1
sigmae
D1
lambda
nu
}
vect=beta1
vecto=cbind(betas,vect,c(count,sigmae),c(lambda,nu),c(sigmae,sigmae))
vect=vecto
return(vect)

if(calcbi==TRUE){
write.table(bi,"bi.txt")
}

if (informa==TRUE) {
desvios<-matriznu(nj,y,x,z,beta1,sigmae,D1,lambda,nu,Ind)

obj.out <- list(theta=c(beta1,sigmae,dd,lambda,nu), iter = count, desvios = desvios)
}
else
{obj.out <- list(theta=c(beta1,sigmae,dd,lambda,nu), iter = count)}

if (loglik==TRUE){
saida<-as.numeric(logvero(nj,y,x,z,beta1,sigmae,D1,lambda,nu,Ind))
return(saida)
}
class(obj.out) <- "EM.Skew"
return(obj.out)
}

```

```

#Ejecución
n=50      #Número de sujetos
k=3      #k: numero de repeticiones por sujeto (intercepto aleatorio)
r=1
Ind=1     #r: Num de efectos aleatorios: intercepto, y m variables aleatorias
dis=1     #0 para normal, 1 para sn, Dis=3 para skewt, para el ef aleatorio
distr=1   #Distr=0 para normal, 1 skew normal, 2 skew t, para el error
nj=rep(k,n)

a=200
lambda=a
sig=5
Di=10-sig
rhoi=Di/(Di+sig)
sigmae2=sig
sigmae=sig

v=2
persona=matrix(c(seq(1:n)),n,1)
repetido=c(rep(persona,each=k))
X=MX(k,n)      #X matrix de variables de efectos fijos
x=X
bet=c(-30,-30)
betini=c(-29.9,-29.9)
beta=betini
beta1=betini
dise=matrix(0,n*k,n)
for (i in 1:n){
  persona=rep(1:n,each=k)
  pos=which(persona==i)
  dise[pos,i]=1
}
zt=dise
z=matrix(rep(1,n*k),n*k,1)
nsim=1000
resultados5=matrix(c(0),1,5)
for(i in 1:nsim){
  Zbe=paraz(Di,dis,k,a,v)
  er=error(sig,distr,n,a,v)
  y=X%*%bet+Zbe+er
D=Di
D1=Di
ques=EM.Skew(nj,y,x,z,betas=bet,beta1=betini,sigmae=sigmae2,D1,lambda,nu=2,Ind=1,lb,lu,precisão,loglik=T,informa=T,calcbi=T)
vector=rbind(resultados5,ques)
resultados5=vector
}

```