

Tesina

**“Modelo lineal mixto para datos  
correlacionados: evaluación de  
los supuestos”**

Alumna: Mara Lis Catalano

Directora: Mgs. Ma. del Carmen García

Carrera: Licenciatura en Estadística



Universidad Nacional de Rosario  
Facultad de Ciencias Económicas y Estadística  
Escuela de Estadística  
Noviembre 2013

## **AGRADECIMIENTOS**

Quiero agradecer a toda mi familia que siempre me acompañó, apoyó y me contuvo en este camino para que nunca bajara los brazos. También, a todos mis amigos por compartir y formar parte de mi vida. A mis compañeras de facultad, con las cuales hemos compartido muchos momentos muy lindos.

Finalmente, un agradecimiento muy especial a mi directora de tesis Mgs. María del Carmen García por su enorme dedicación y por ser tan paciente e incondicional conmigo. Le agradezco las infinitudes de cosas que me enseñó a lo largo de este tiempo.

## ÍNDICE

1. INTRODUCCIÓN .....	1
2. DATOS LONGITUDINALES .....	3
2.1 Análisis exploratorio.....	5
2.2 Modelos lineales mixtos .....	8
2.3 Modelos paramétricos para la estructura de covariancia .....	11
2.4 Estimación de los parámetros del modelo .....	13
2.4.1 Método de máxima verosimilitud.....	14
2.4.2 Método de máxima verosimilitud restringida.....	15
2.5 Inferencia sobre los parámetros del modelo .....	16
2.5.1 Test de Wald .....	17
2.5.2 Test de razón de verosimilitud .....	17
2.5.3 Criterios para seleccionar modelos .....	19
2.6 Predicción de los efectos aleatorios .....	20
2.7 Diagnóstico del modelo lineal mixto .....	21
2.7.1 Análisis de residuos .....	21
2.7.2 Residuos del modelo lineal mixto.....	21
2.7.3 Residuos estudentizados, externamente estudentizados y de Pearson.....	24
2.7.4 Residuos puros y confundidos .....	25
2.7.5 Residuos con mínima confusión .....	25
2.7.6 Técnicas gráficas para evaluar los supuestos.....	28
3. APLICACIÓN .....	31
3.1 Análisis descriptivo.....	33
3.1.1 Exploración de la estructura de la parte media .....	33
3.1.2 Exploración de la estructura de covariancia .....	36
3.2 Análisis inferencial. ....	39

3.2.1 Selección de la estructura de covariancia .....	39
3.2.2 Selección de la estructura de la parte media .....	41
3.2.3 Diagnóstico del modelo .....	44
4. CONSIDERACIONES FINALES.....	51
5. BIBLIOGRAFÍA .....	53
6. ANEXO .....	55

## 1. INTRODUCCIÓN

Un estudio longitudinal se realiza cuando el objetivo de la investigación es caracterizar el cambio de la variable en estudio a través del tiempo y los factores que influyen en ese cambio. En estos estudios, las observaciones de la variable de interés se registran en distintas ocasiones para cada unidad, lo que hace que las mediciones repetidas estén correlacionadas.

Los modelos mixtos permiten ajustar datos con estas particularidades, donde la respuesta es modelada por una parte sistemática que está formada por una combinación de características poblacionales que son compartidas por todas las unidades (efectos fijos), y una parte aleatoria que está constituida por efectos específicos de cada unidad (efectos aleatorios) y por el error aleatorio. Estos modelos permiten, además, hacer predicciones del perfil de una unidad específica.

El proceso de modelización está compuesto por reiterados ajustes de modelos y evaluaciones de los mismos, con el objetivo de identificar una forma funcional parca para describir la evolución de la respuesta en el tiempo y su relación con las covariables. Una etapa importante de este proceso lo constituye la evaluación de los supuestos del modelo.

Para que estos modelos provean un buen ajuste se tienen que cumplir los supuestos realizados sobre las partes sistemática y aleatoria (efectos aleatorios y error aleatorio). Debido a que se cuenta con dos fuentes de variación (entre e intra unidades) se definen tres tipos de residuos, conocidos como los residuos condicionales, los residuos marginales y los residuos de los efectos aleatorios (BLUP, best linear unbiased predictor que significa “mejor predictor lineal insesgado”).

En esta tesina se utilizan como una herramienta de diagnóstico los residuos. La adecuación del modelo se analiza gráficamente usando los residuos marginales, condicionales y los residuos de los efectos aleatorios.

Para la evaluación del supuesto de normalidad se introducen los residuos con mínima confusión como una alternativa a los residuos condicionales. Se realiza también la implementación computacional de los mismos.

Los residuos condicionales están “confundidos” en el sentido que no sólo son función de los errores que predicen sino también de los efectos aleatorios y podría ocurrir que si los efectos aleatorios se alejan considerablemente de la distribución normal se puede concluir que los errores no se distribuyen normalmente cuando en realidad eso no es cierto.

Para ilustrar qué supuestos se pueden diagnosticar con cada residuo y cuál es la técnica gráfica utilizada, se modelan datos suministrados por el Hospital Escuela “Eva Perón” de Granadero Baigorria correspondiente a valores de procalcitonina en pacientes con Síndrome de Respuesta Inflamatoria Sistemática medidos durante su permanencia en terapia intensiva.

## 2. DATOS LONGITUDINALES

Los datos longitudinales están conformados por mediciones repetidas de una variable realizadas a la misma unidad. Estas mediciones repetidas surgen de observar las unidades en varios momentos diferentes o bajo ciertas condiciones, denominados ocasiones.

Las ocasiones en que se miden las respuestas no necesariamente tienen que estar distribuidas equidistantes a través de la duración del estudio.

En general, las mediciones repetidas se registran en ocasiones comunes o no para todos los participantes dando lugar a los estudios balanceados y no balanceados. Muchas veces, aunque el estudio sea balanceado, por alguna razón no se cuenta con algunas observaciones, obteniendo distinto número de mediciones de la variable de interés para las unidades.

Para simplificar la escritura, de ahora en adelante se asume que los tiempos de medición son los mismos para todas las unidades y que no hay observaciones faltantes.

De esta forma, se tiene una muestra de  $N$  unidades cada una con  $n$  mediciones repetidas de la variable en estudio registradas en los tiempos  $t_1, t_2, \dots, t_n$  siendo el total de observaciones  $N^* = Nn$ . La variable respuesta  $Y_{ij}$  es la realización sobre la unidad  $i$  en la ocasión  $j$ ,  $i=1, \dots, N$ ;  $j=1, \dots, n$ .

Estas unidades en general están agrupadas de acuerdo a alguna característica que comparten (sexo, etc.) por lo que se supone que existen  $S$  grupos.

Para cada unidad se registran covariables. Existen dos tipos de covariables, aquellas cuyos valores no cambian durante el estudio y aquellas cuyos valores cambian.

El objetivo primordial de estos estudios es caracterizar el cambio en la respuesta a través de las ocasiones y los factores que influyen el cambio. Es conveniente notar que la respuesta de una unidad en una ocasión es muy probable que sea predictiva de la respuesta de la misma unidad en futuras ocasiones. En consecuencia, las observaciones de una unidad

están relacionadas. Cuando se trabaja con variables cuantitativas la dependencia entre ellas se puede medir mediante su correlación.

Estudios empíricos llevan a pensar que existen tres fuentes potenciales de variabilidad que tienen impacto sobre la correlación entre medidas repetidas; estas son:

- la heterogeneidad entre unidades. Refleja la propensión natural a responder que tiene cada unidad y si esto se debe a determinados factores (ambientales, genéticos, sociales), todas las mediciones de esa unidad van a estar afectadas por ellos. Por ejemplo, se espera que las unidades con valores altos en una ocasión tengan valores altos en la siguiente ocasión.

- la variación biológica intra unidad. Se piensa que existe algún proceso que genera las respuestas de una unidad, el cual cambia en forma continua y suave a través del tiempo produciendo que las medidas más cercanas sean más parecidas.

- error de medida. Es la fuente de variación que surge debido a los errores de medición, por ejemplo cuando se usan procedimientos de medición no demasiado confiables.

Estas fuentes de variación se clasifican como “variabilidad entre” que representa la variación entre mediciones de las distintas unidades (heterogeneidad entre unidades) y “variabilidad intra” que agrupa a las fuentes de variación que hay dentro de las unidades (variación biológica intra unidad y error de medida).

En consecuencia, la correlación entre las mediciones repetidas, tiende a ser positiva y a decrecer cuanto más separadas están las observaciones en el tiempo, raramente se aproxima a cero excepto cuando el intervalo de tiempo entre ellas es muy grande, y nunca es igual a uno aunque las medidas se registren muy cercanas en el tiempo.

Como estas fuentes de variabilidad explican la correlación existente, es necesario controlarlas y estimarlas.

Debido a la existencia de correlación entre las mediciones de una misma unidad no se pueden utilizar las pruebas estadísticas habituales que se basan en la existencia de



independencia. Surgen así técnicas especiales que consideran esa correlación cuando se quiere estudiar el cambio de la respuesta a través del tiempo permitiendo la obtención de estimaciones más precisas de los parámetros de regresión e inferencias válidas sobre los mismos.

Para modelar estos datos se debe considerar la modelación de las estructuras, media (cómo evoluciona la respuesta promedio en función del tiempo) y de covariancia (la variación entre las mediciones repetidas dentro y entre unidades). En la sección siguiente se consideran técnicas exploratorias que permiten estudiar cómo se comportan los datos para luego elegir un modelo que los ajuste.

## **2.1 Análisis exploratorio**

Para indagar sobre las estructuras de media y covariancia, además de calcular las estadísticas básicas sobre la variable de interés, se pueden utilizar las siguientes técnicas gráficas.

### **-Evaluación de la estructura media**

*Gráficos de perfiles individuales y de perfiles promedio.*

Los gráficos de perfiles individuales y de perfiles promedio se utilizan para indagar los posibles modelos para la estructura de la parte media del modelo.

En el eje de ordenadas se ubican los diferentes valores de la variable respuesta y en el eje de abscisas el tiempo.

Para el gráfico de perfiles individuales las observaciones de una misma unidad se unen mediante segmentos para identificar a que unidad pertenecen. Este gráfico permite visualizar la evolución de la variable de interés para cada unidad, la variabilidad entre las respuestas en cada ocasión y detectar posibles outliers.

En el gráfico de perfiles promedio se grafica en el eje de ordenadas el promedio de la variable respuesta en cada tiempo para cada grupo y se unen mediante segmentos. Este gráfico provee la base para seleccionar un modelo apropiado para el análisis de los cambios a través del tiempo, ya que permite observar si la tendencia promedio a través del tiempo es lineal o no.

### **-Evaluación de la estructura de covariancia**

Se han desarrollado varias técnicas gráficas que ayudan a especificar la estructura de correlación de un conjunto de unidades con medidas repetidas y a determinar un modelo inicial para la matriz de variancias y covariancias intra-unidad. Si el número de unidades que se estudian en cada grupo es considerable se puede hacer la distinción por grupo. Las técnicas diagnósticas más utilizadas son:

#### *Matriz de gráficos de dispersión.*

Consiste en un conjunto de gráficos de dispersión. Para cada par de ocasiones se grafican los valores observados de la respuesta. En general se utiliza cuando los tiempos de medición son comunes entre las unidades e igualmente espaciados.

Este gráfico permite diagnosticar estructuras de correlación, observando en cada gráfico la relación entre las mediciones repetidas.

#### *Gráfico de Draftman.*

Este gráfico es similar al anterior, pero utiliza una estandarización de la variable respuesta.

$$Z_{ij} = \frac{Y_{ij} - \bar{Y}_{.j}}{\sqrt{S^2_{.j}}}, \quad i=1, \dots, N; j=1, \dots, n; \quad (2.1.1)$$

donde,  $Y_{ij}$  variable respuesta de la unidad  $i$  en la ocasión  $j$  perteneciente al grupo  $s$ ,

$$\bar{Y}_{.j} = \frac{1}{N} \sum_{i=1}^N Y_{ij}, S_{.j}^2 = \frac{\sum_{i=1}^N (Y_{ij} - \bar{Y}_{.j})^2}{N-1}.$$

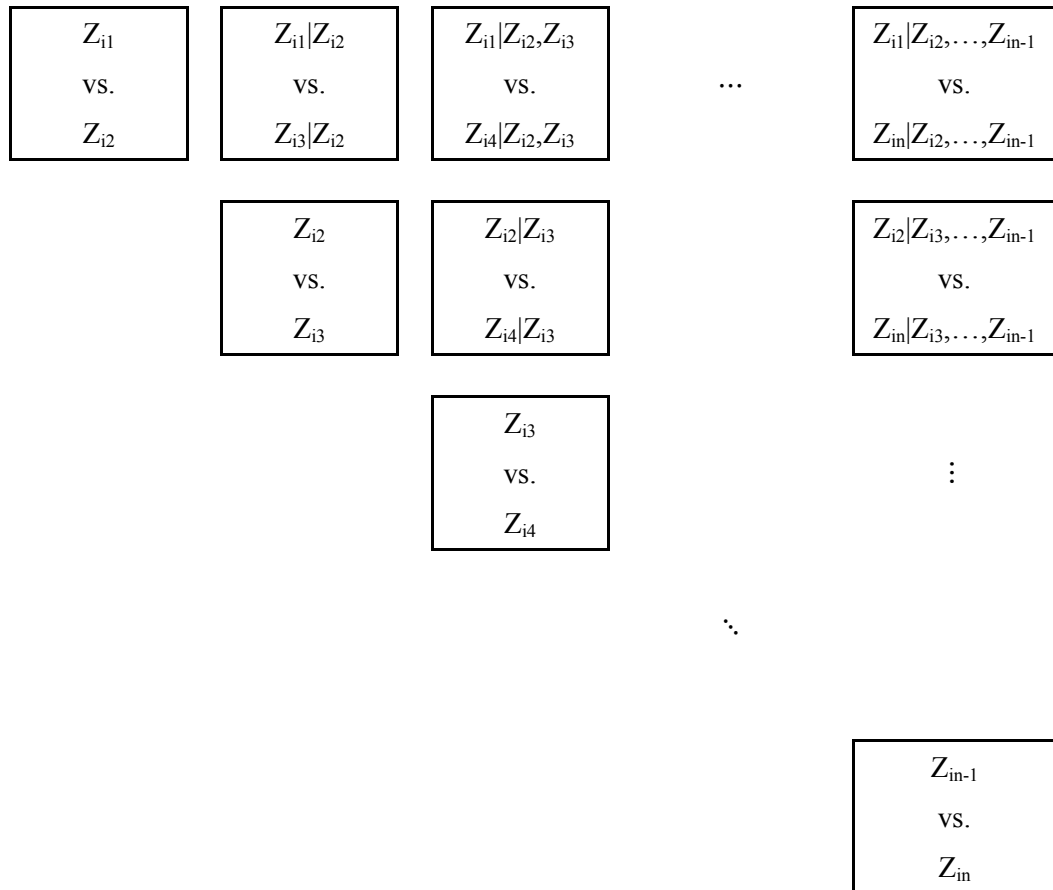
La utilización de la variable respuesta estandarizada ayuda a eliminar la variabilidad de los datos, asociada con diferencias en las medias y variancias en el tiempo, permitiendo visualizar más claramente el patrón de correlación.

*Gráfico PRISM (Partial Regression on Intervenors Scatterplot Matrix).*

Este gráfico consiste en una matriz de gráficos de dispersión utilizando la variable respuesta estandarizada. En la diagonal principal se encuentran gráficos de dispersión entre la variable respuesta en los tiempos  $t_j$  y  $t_{j+1}$ ,  $j=1, \dots, n-1$  y fuera de ella, gráficos de regresión parcial entre las variables respuestas separadas una cierta cantidad de ocasiones ajustadas por las respuestas de los tiempos intermedios. De esta forma la segunda diagonal contiene gráficos de regresión parcial de las respuestas apartadas dos tiempos, ajustada por la respuesta en el tiempo intermedio ( $Z_{i1}|Z_{i2}$  vs.  $Z_{i3}|Z_{i2}$ ,  $Z_{i2}|Z_{i3}$  vs.  $Z_{i4}|Z_{i3}$ , etc.). En la tercera diagonal se encuentran los gráficos de regresión parcial de las respuestas apartadas tres tiempos, ajustada por la respuesta dos tiempos intermedios ( $Z_{i1}|Z_{i2}, Z_{i3}$  vs.  $Z_{i4}|Z_{i2}Z_{i3}$ , etc.).

Se usa como complemento del gráfico de Draftman cuando se quiere identificar estructuras autoregresivas o de otro tipo de correlación serial.

A continuación se muestra el esquema de gráficos de regresión parcial en un PRISM (Zimmerman, 2000).



### *Correlograma.*

Consiste en graficar la correlación entre todos los pares de mediciones repetidas separadas cierta cantidad de rezagos (eje vertical) y el número de rezagos que hay entre ellas (eje horizontal).

Permite analizar cómo evoluciona la correlación a medida que aumenta el número de rezagos. La matriz de correlaciones muestrales se examina por resultar de importancia para la selección de estructuras de correlación.

## **2.2 Modelos lineales mixtos (MLM)**

En esta sección se presenta una clase de modelos para el caso particular que la distribución del vector de respuestas se aproxime a una distribución normal multivariada.

Los modelos lineales mixtos se utilizan para el análisis de datos con mediciones repetidas y, en particular, datos longitudinales. Su denominación se debe a que los parámetros ingresan a la ecuación en forma “lineal” y “mixtos” porque se consideran tanto efectos fijos (características compartidas por todas las unidades) como efectos aleatorios (efectos específicos de cada unidad). Algunas de las características más relevantes de estos modelos son:

- se aplican tanto en estudios balanceados como en no balanceados.
- permiten considerar al tiempo en forma cuantitativa.
- permiten utilizar varias covariables cuantitativas y/o cualitativas, dependientes o no del tiempo.
- admiten distintas estructuras para la matriz de variancias y covariancias de los errores.
- distinguen las fuentes de variación entre e intra unidad.

Estos modelos se plantean con el objetivo de conocer las trayectorias individuales y evaluar cómo la respuesta media cambia a través del tiempo para cada grupo, denominado con  $s, s=1, \dots, S$ .

El modelo lineal mixto para la unidad  $i$  se puede expresar en forma matricial como:

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, i=1, \dots, N; \quad (2.2.1)$$

donde,

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in} \end{bmatrix}$$

$\mathbf{Y}_i$ : vector ( $n \times 1$ ) de la variable respuesta de la unidad  $i$

$\mathbf{X}_i$ : matriz de diseño de la unidad  $i$ , que caracteriza la parte sistemática de la respuesta, de dimensión ( $n \times p$ )

$\beta$  : vector ( $p \times 1$ ) de parámetros (efectos fijos)

$Z_i$ : matriz de diseño de la unidad  $i$ , que caracteriza la parte aleatoria de la respuesta, de dimensión ( $n \times k$ )

$b_i$ : vector ( $k \times 1$ ) de efectos aleatorios de la unidad  $i$

$\epsilon_i$  : vector ( $n \times 1$ ) de los errores aleatorios de la unidad  $i$

$N$ = total de unidades

$n$ = total de observaciones de la variable respuesta para cada unidad

$p$ = total de parámetros de la parte media

$k$ = total de efectos aleatorios

Los supuestos se pueden escribir

$b_i$  y  $\epsilon_i$  son independientes

$$b_i \stackrel{iid}{\sim} N_k(\mathbf{0}, \mathbf{D}^*) \quad \mathbf{D}^* = \sigma^2 \mathbf{D} \quad (2.2.2)$$

$$\epsilon_i \stackrel{iid}{\sim} N_n(\mathbf{0}, \mathbf{R}^*) \quad \mathbf{R}^* = \sigma^2 \mathbf{R} \quad i=1, \dots, N; \quad (2.2.3)$$

Las matrices  $\mathbf{D}^*$  ( $k \times k$ ) y  $\mathbf{R}^*$  ( $n \times n$ ) son definidas positivas y contienen las variancias y covariancias de los elementos del vector  $b_i$  y  $\epsilon_i$  respectivamente. Estas matrices pueden ser diferentes para cada grupo, es decir,  $\mathbf{D}_s^*$  y  $\mathbf{R}_s^*$  con  $s=1, \dots, S$ . Se desarrolla la metodología para el caso en que las matrices anteriores son las mismas para los  $S$  grupos, nombrándolas a partir de ahora  $\mathbf{D}^*$  y  $\mathbf{R}^*$ .

Si las variancias de los errores aleatorios son iguales (hay homocedasticidad),  $\mathbf{R}$  representa a la matriz de correlación de los errores.

En los modelos mixtos se realiza la distinción entre modelo marginal y modelo condicional, siendo,

- $E(Y_i / b_i) = \mathbf{X}_i \beta + \mathbf{Z}_i b_i$  la media condicional o específica de la unidad  $i$ , (2.2.4)

- $E(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta}$  la media marginal, (2.2.5)

- $\text{Cov}(\mathbf{Y}_i / \mathbf{b}_i) = \mathbf{R}^*$  la variancia condicional, (2.2.6)

- $\text{Cov}(\mathbf{Y}_i) = \mathbf{Z}_i\mathbf{D}^*\mathbf{Z}_i' + \mathbf{R}^* = \mathbf{V}_i$  la variancia marginal. (2.2.7)

Se define  $\boldsymbol{\theta}$  como el vector que contiene los parámetros de la matriz  $\mathbf{V}$ , donde

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & 0 & \dots & 0 \\ 0 & \mathbf{V}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{V}_N \end{bmatrix}. \quad (2.2.8)$$

### 2.3 Modelos paramétricos para la estructura de covariancia

La aplicación de un modelo lineal mixto a un conjunto particular de datos longitudinales requiere especificar las estructuras para la parte media y aleatoria. Primero se elige un modelo para la covariancia. Debido a que éste depende del modelo de la parte media y para evitar que cualquier especificación incorrecta de la misma afecte la estimación de la matriz de variancias y covariancias se recomienda utilizar para representar la estructura de la parte media un modelo maximal.

Modelar adecuadamente la matriz de variancias y covariancias es importante para obtener errores estándares correctos, los cuales permiten aumentar la precisión con la que se estiman los parámetros de regresión y realizar inferencias válidas acerca de estos parámetros.

La matriz de variancias y covariancias del vector de efectos aleatorios de la unidad  $i$ ,  $\mathbf{D}^*$ , asume, en general una estructura arbitraria (no estructurada). Mientras, que la estructura de la matriz de covariancia del vector de errores puede adoptar la forma de independencia  $(\sigma^2\mathbf{I})$ , si se considera que la variación intra unidad se debe enteramente al error de medida. En cambio, si esta variación se debe a las fluctuaciones biológicas de la unidad, se espera que haya correlación entre las medidas repetidas, y  $\mathbf{R}^*$  adopte una estructura más compleja.

Se recuerda que manteniendo la misma estructura de covariancia, estas matrices puede ser diferente para cada grupo.

A continuación se mencionan algunas estructuras que puede adoptar la matriz de covariancias intra-unidad.

*Estructura arbitraria o no estructurada.*

Esta estructura considera variancias y covariancias distintas entre las mediciones repetidas. Siendo  $\sigma_{ij'} = \text{Cov}(Y_{ij}, Y_{ij'})$  se expresa como,

$$\mathbf{R}^* = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix}.$$

*Estructura de Simetría Compuesta.*

El patrón de covariancia de simetría compuesta asume que la correlación entre mediciones repetidas es la misma sin importar el número de rezagos que hay entre ellas,  $\text{Corr}(Y_{ij}, Y_{ij'}) = \rho$ . Su representación es,

$$\mathbf{R}^* = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}.$$

*Estructura de Toeplitz.*

El modelo Toeplitz para la covariancia hace el supuesto que cualquier par de respuestas igualmente espaciadas tiene la misma correlación. Notando que la  $\text{Corr}(Y_{ij}, Y_{ij+h}) = \rho_h \forall j$  y  $h$  se tiene,



$$\mathbf{R}^* = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \dots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \dots & 1 \end{bmatrix}.$$

*Estructura Autorregresiva de primer orden.*

La estructura autorregresiva asume que la correlación entre medidas repetidas disminuye a medida que el número de rezagos entre ellas aumenta, siendo esta correlación la misma para las observaciones que están separadas la misma cantidad de rezagos. Esto se sintetiza como  $\text{Corr}(Y_{ij}, Y_{ij+h}) = \rho^h \forall j \text{ y } h$ .

$$\mathbf{R}^* = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{bmatrix}.$$

*Estructura Markoviana (matriz exponencial).*

Cuando las ocasiones en las que se registran las mediciones no están igualmente espaciadas en el tiempo el modelo autorregresivo se puede generalizar asumiendo que  $\text{Corr}(Y_{ij}, Y_{ij'}) = \rho^{d_{jj'}}$ ,  $d_{jj'} = |t_j - t_{j'}| \forall j \neq j'$ , de lo que resulta la matriz,

$$\mathbf{R}^* = \sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \dots & \rho^{d_{1n}} \\ \rho^{d_{12}} & 1 & \dots & \rho^{d_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{d_{n1}} & \rho^{d_{n2}} & \dots & 1 \end{bmatrix}.$$

## 2.4 Estimación de los parámetros del modelo

Para la estimación de los parámetros del modelo (2.2.1) en la literatura están disponibles métodos bayesianos y métodos basados en la verosimilitud. Si se puede asumir

que  $\mathbf{b}_i$  y  $\boldsymbol{\varepsilon}_i$  tienen distribución normal multivariada, la mejor aproximación para la estimación de los parámetros se logra con los métodos basados en la verosimilitud (Laird and Ware, 1982) , los métodos de máxima verosimilitud y máxima verosimilitud restringida (Patterson and Thompson, 1971). La idea fundamental detrás de los estimadores obtenidos por estos métodos es usar los valores que son más probables para los datos que han sido observados.

### 2.4.1 Método de máxima verosimilitud (ML)

Considerando que el vector de mediciones de una unidad se distribuye normalmente

$$\mathbf{Y}_i \stackrel{\text{ind}}{\sim} N_n(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i) \quad i=1, \dots, N, \quad (2.4.1)$$

la función de log-verosimilitud resulta:

$$\ell = \ln f(\mathbf{Y}) = -\frac{1}{2} Nn \ln(2\pi) - \frac{1}{2} \sum_{i=1}^N \ln |\mathbf{V}_i| - \frac{1}{2} \sum_{i=1}^N [(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})' \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})], \quad (2.4.2)$$

siendo  $\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_1', \dots, \mathbf{Y}_N')'$ ,  $\mathbf{V}_i = \mathbf{Z}_i \mathbf{D}^* \mathbf{Z}_i' + \mathbf{R}^*$  y  $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\theta})$ , con el vector  $\boldsymbol{\theta}$  que agrupa los parámetros de covariancia de esa matriz.

Los estimadores máximos verosímiles de  $\boldsymbol{\beta}$  y  $\boldsymbol{\theta}$  se obtienen maximizando (2.4.2). Derivando (2.4.2) con respecto a  $\boldsymbol{\beta}$  y  $\boldsymbol{\theta}$  e igualándolas a cero. Para  $\boldsymbol{\theta}$  resulta una ecuación no lineal y requiere de algoritmos numéricos para encontrar su solución, por lo que no es posible escribir de forma simple la expresión de  $\hat{\boldsymbol{\theta}}$  (Fitzmaurice et al., 2004, pág. 91).

El estimador del vector  $\boldsymbol{\beta}$  es,

$$\hat{\boldsymbol{\beta}} = \left( \sum_{i=1}^N \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{Y}_i = (\mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{Y}, \quad (2.4.3)$$

$$\text{donde } \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix}, \quad \hat{\mathbf{V}} = \begin{bmatrix} \hat{\mathbf{V}}_1 & 0 & \dots & 0 \\ 0 & \hat{\mathbf{V}}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\mathbf{V}}_N \end{bmatrix}. \quad (2.4.4)$$

Aunque el estimador (2.4.3) sea insesgado para cualquier elección de  $\mathbf{V}_i$  es más eficiente cuando se usa el verdadero valor de  $\mathbf{V}_i$ . Como se reemplaza  $\mathbf{V}$  por  $\hat{\mathbf{V}}$  en la expresión (2.4.3) no se puede calcular exactamente la matriz de variancias y covariancias de  $\hat{\boldsymbol{\beta}}$ . Por lo tanto, no se conoce la distribución muestral exacta de este estimador pero, si el número de unidades es grande (es decir,  $N$  crece mientras el número de medidas repetidas y de parámetros del modelo permanecen fijos), se demuestra que  $\hat{\boldsymbol{\beta}}$  se distribuye aproximadamente  $N_p(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1})$ . (2.4.5)

#### 2.4.2 Método de máxima verosimilitud restringida (REML)

Si bien cuando el número de unidades de la muestra no es suficientemente grande  $\hat{\boldsymbol{\beta}}$  es un estimador insesgado de  $\boldsymbol{\beta}$ , no ocurre lo mismo con  $\hat{\mathbf{V}}_i$ , el estimador máximo verosimil de  $\mathbf{V}_i$ , el cual es sesgado. El sesgo se origina debido a que no se tiene en cuenta que  $\boldsymbol{\beta}$  es estimado a partir de los datos. Esto produce que los elementos de la diagonal de  $\mathbf{V}_i$  sean subestimados.

Para disminuir el sesgo se pueden estimar los parámetros mediante el método REML, que es una variante del método de máxima verosimilitud.

La función de log-verosimilitud restringida que se utiliza es,

$$\ell^* = -\frac{1}{2}Nn \ln(2\pi) - \frac{1}{2} \sum_{i=1}^N \ln|\mathbf{V}_i| - \frac{1}{2} \sum_{i=1}^N [(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})] - \frac{1}{2} \ln \left| \sum_{i=1}^N \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i \right| \quad (2.4.6)$$

$$\text{Observar que } -\frac{1}{2} \ln \left| \sum_{i=1}^N \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i \right| = -\frac{1}{2} \ln \left| \left( \sum_{i=1}^N \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i \right)^{-1} \right| = \ln |\text{Cov}(\hat{\boldsymbol{\beta}})|^{1/2}.$$

En consecuencia, el determinante que se agrega en (2.4.6) se llama variancia generalizada de  $\boldsymbol{\beta}$  y resume la variación en la estimación de  $\boldsymbol{\beta}$ .

Los estimadores REML de  $\beta$  y  $\theta$  se obtienen maximizando (2.4.6). Así, la expresión (2.4.3) es el estimador REML de  $\beta$  siendo  $\hat{V}_i$  el estimador REML de  $V_i$ .

En conclusión, para estimar los parámetros de covariancia se recomienda usar el estimador REML, que tiene menor que el del estimador ML. La diferencia entre los estimadores obtenidos por ambos métodos pierde importancia cuando el tamaño de muestra,  $N$ , sea considerablemente más grande que  $p$  (dimensión de  $\beta$ ). Estos dos estimadores son consistentes, normalmente asintóticos, eficientes, manteniendo estas propiedades aún para datos no balanceados.

## 2.5 Inferencia sobre los parámetros del modelo

En la mayoría de los estudios longitudinales, la inferencia de interés se centra en los parámetros de la parte media. Para plantear pruebas de hipótesis sobre estos parámetros es conveniente expresarlas mediante la denominada hipótesis lineal general. Para ello se construyen funciones lineales de los elementos de  $\beta$ , de la forma  $L'\beta$ , siendo  $L'$  una matriz de dimensión  $r \times p$  de rango completo en las filas ( $r = \text{rango}(L')$ ).

El estimador de  $L'\beta$  es  $L'\hat{\beta}$  y si el número de unidades es grande ( $N \rightarrow \infty$ ), la distribución muestral aproximada de  $L'\hat{\beta}$  es:

$$L'\hat{\beta} \text{ se distribuye aproximadamente } N_p(L'\beta, L'\text{Cov}(\hat{\beta})L), \quad (2.5.1)$$

donde  $\text{Cov}(\hat{\beta}) = (X'V^{-1}X)^{-1}$ .

Para probar hipótesis se presentan dos métodos basados en la función de verosimilitud: la prueba de Wald y la del cociente de verosimilitud. Asintóticamente las dos pruebas son equivalentes pero para muestras no demasiado grandes la prueba de Wald tiende a ser más liberal (rechaza más la hipótesis nula).

### 2.5.1 Test de Wald

Las hipótesis de la forma,

$$H_0) \mathbf{L}'\boldsymbol{\beta} = \mathbf{0} \quad H_1) \mathbf{L}'\boldsymbol{\beta} \neq \mathbf{0} ,$$

se pueden comprobar mediante la estadística,

$$W = (\mathbf{L}'\hat{\boldsymbol{\beta}})' (\mathbf{L}'\hat{\text{Cov}}(\hat{\boldsymbol{\beta}})\mathbf{L})^{-1} (\mathbf{L}'\hat{\boldsymbol{\beta}}). \quad (2.5.2)$$

Esta estadística W bajo  $H_0$  se distribuye aproximadamente  $\chi_r^2$ .

La variabilidad incorporada por reemplazar las componentes de variancia por sus estimadores no es tomada en cuenta en la construcción de la estadística del test, por lo cual sólo proveyerá inferencias válidas cuando N sea grande. Cuando N no sea demasiado grande una alternativa es reemplazar la distribución chi-cuadrado por la distribución F de Snedecor. El inconveniente surge porque no se conocen los grados de libertad del denominador. Existen algunos métodos para aproximarlos: Containment, Within-Between, Satterthwaite y Kenward y Roger, entre otros (Fitzmaurice et al., 2004, pág 98).

### 2.5.2 Test de razón de verosimilitud

Este test es una alternativa al test de Wald ya que es más confiable cuando N no es tan grande.

La prueba del cociente de verosimilitud se aplica cuando se quiere probar las denominadas hipótesis de modelos completos ( $H_0$ ) versus reducidos ( $H_1$ ) siendo,

$$H_0) \mathbf{L}'\boldsymbol{\beta} = \mathbf{0} \quad H_1) \mathbf{L}'\boldsymbol{\beta} \neq \mathbf{0} .$$

El método consiste en comparar las log-verosimilitudes maximizadas de dos modelos, uno que incorpora la restricción  $\mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$  (modelo reducido) y otro sin esta restricción (modelo completo). Estos dos modelos están anidados, en el sentido que el modelo reducido es un caso particular del modelo completo.

La función de log-verosimilitud maximizada del modelo reducido es,

$$\hat{\ell}_{\text{red}} = -\frac{1}{2} Nn \ln(2\pi) - \frac{1}{2} \sum_{i=1}^N \ln |\hat{\mathbf{V}}_i| - \frac{1}{2} \sum_{i=1}^N \left[ (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_0) \hat{\mathbf{V}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_0) \right] \quad (2.5.3)$$

$\boldsymbol{\beta}_0$ : vector de parámetros resultante de haber impuesto la restricción planteada en  $H_0$ .

La función de log-verosimilitud maximizada del modelo completo es,

$$\hat{\ell}_{\text{comp}} = -\frac{1}{2} Nn \ln(2\pi) - \frac{1}{2} \sum_{i=1}^N \ln |\hat{\mathbf{V}}_i| - \frac{1}{2} \sum_{i=1}^N \left[ (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) \hat{\mathbf{V}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) \right]. \quad (2.5.4)$$

En (2.5.3) y (2.5.4)  $\hat{\boldsymbol{\beta}}$  y  $\hat{\mathbf{V}}_i$  se obtienen mediante los métodos ML o REML.

Si el número de unidades es grande ( $N \rightarrow \infty$ ), la distribución muestral aproximada de la estadística del test es,

$$G = -2(\hat{\ell}_{\text{red}} - \hat{\ell}_{\text{comp}}) \text{ que bajo } H_0 \text{ se distribuye } \chi_{\text{GL}}^2, \quad (2.5.5)$$

donde  $\text{GL} = n^\circ$  de parámetros del modelo completo –  $n^\circ$  de parámetros del modelo reducido.

Cuando se quiere comparar la estructura media de dos modelos diferentes mediante el test de razón de verosimilitud se recomienda usar el método de ML para estimar los parámetros y usar la función de log-verosimilitud obtenida mediante este método. En cambio, si se quiere elegir un modelo entre una serie de modelos de patrones de covarianza anidados dado un modelo maximal para la media, esta elección se puede hacer comparando la log-verosimilitud restringida maximizada por los estimadores REML para cada modelo.

El test de razón de verosimilitud es válido según la naturaleza de la hipótesis nula, ya que de ésta depende la distribución que se usa para aproximar la estadística del test. Una precaución se debe tomar cuando se quiere probar una hipótesis nula que está en el límite del espacio paramétrico como sucede al hacer inferencia sobre las componentes de variancia (por ejemplo, se postula que una variancia es nula). La distribución considerada ya no es chi-cuadrado sino una mezcla de distribuciones chi-cuadrados. Si este problema se ignora, y se usa la distribución chi-cuadrado original, la probabilidad asociada resultante va a estar

sobreestimada haciendo que se seleccione un modelo para la covariancia demasiado parco. En estas situaciones se recomienda usar un nivel de significación de 0,1.

### 2.5.3 Criterios para seleccionar modelos

Cuando algún par de modelos están anidados (el modelo reducido es un caso particular del modelo completo) se puede usar el test de razón de verosimilitud que compara la log-verosimilitud maximizada del modelo reducido con la del modelo completo.

Surgen otras medidas para comparar modelos que no necesariamente tienen que estar anidados. Las mismas son funciones de la log-verosimilitud de cada modelo ajustada por el número de parámetros que tienen. Este ajuste se hace debido a que la log-verosimilitud aumenta a medida que se agregan parámetros. Por lo tanto, la idea es crear un balance entre dos cuestiones: que el modelo sea lo suficientemente complejo como para proveer un buen ajuste de los datos, pero al mismo tiempo obtener un modelo parco.

Dos de los criterios más utilizados en la práctica, son los criterios de información de Akaike (AIC) y de información bayesiana (BIC).

$$\text{Criterio de información de Akaike (AIC): } AIC = -2\hat{\ell} + 2p^*, \quad (2.5.6)$$

$$\text{Criterio de información bayesiano de Schwarz (BIC): } BIC = -2\hat{\ell} + p^* \ln(N), \quad (2.5.7)$$

siendo,

$\hat{\ell}$ : log-verosimilitud del modelo maximizada,

$p^*$ : total de parámetros del modelo.

El criterio BIC tiende a favorecer la selección de modelos más parcos, ya que penaliza de forma más severa a los modelos con muchos parámetros.

Se van a preferir los modelos que minimicen los criterios de AIC o BIC.

Para seleccionar estructuras de covariancia, para cada modelo se calculan los valores de AIC y BIC usando la log-verosimilitud restringida maximizada con los estimadores REML y el número de parámetros de covariancia como  $p^*$ .

## 2.6 Predicción de los efectos aleatorios

En algunos estudios longitudinales el interés se centra en predecir los perfiles de respuesta específicos de cada unidad. Es decir, se quiere conocer como evoluciona cada unidad por separado a través del tiempo.

Los modelos lineales mixtos son modelos específicos para la unidad en el sentido que los modelos de regresión individuales se caracterizan por tener media  $E(\mathbf{Y}_i / \mathbf{b}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$ . De manera que si se desea caracterizar el comportamiento individual se tendrían que “conocer” los valores de  $\boldsymbol{\beta}$  y  $\mathbf{b}_i$ . La estimación de  $\boldsymbol{\beta}$  ya se obtuvo en la sección 2.4 por lo que falta predecir el efecto específico de la unidad  $i$ ,  $\mathbf{b}_i$ .

Se puede demostrar que el mejor predictor lineal insesgado (BLUP) de  $\mathbf{b}_i$  es su esperanza condicional dado el vector de respuestas.

$$E(\mathbf{b}_i / \mathbf{Y}_i) = \mathbf{D}^* \mathbf{Z}_i' \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}), \quad (2.6.1)$$

donde  $\mathbf{V}_i = \text{Cov}(\mathbf{Y}_i) = \mathbf{Z}_i \mathbf{D}^* \mathbf{Z}_i' + \mathbf{R}^*$ .

Si  $\mathbf{D}^*$  y  $\mathbf{R}^*$  son desconocidas se reemplazan por sus estimadores REML y se obtiene,

$$\hat{\mathbf{b}}_i = \hat{\mathbf{D}}^* \mathbf{Z}_i' \hat{\mathbf{V}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}). \quad (2.6.2)$$

El predictor (2.6.2) es referido como mejor predictor lineal insesgado empírico (o EBLUP, sigla formada por su nombre en inglés “empirical best linear unbiased predictor”) o estimador empírico de Bayes.



Calculados  $\hat{\beta}$  y  $\hat{\mathbf{b}}_i$ , se puede obtener el perfil de respuesta predicho de la unidad  $i$  como,

$$\hat{\mathbf{Y}}_i = \mathbf{X}_i \hat{\beta} + \mathbf{Z}_i \hat{\mathbf{b}}_i. \quad (2.6.3)$$

Haciendo los reemplazos correspondientes en (2.6.3) se tiene que (ver anexo),

$$\hat{\mathbf{Y}}_i = (\hat{\mathbf{R}}^* \hat{\mathbf{V}}_i^{-1}) \mathbf{X}_i \hat{\beta} + (\mathbf{I}_n - \hat{\mathbf{R}}^* \hat{\mathbf{V}}_i^{-1}) \mathbf{Y}_i. \quad (2.6.4)$$

El perfil de respuesta predicho para la unidad  $i$  se puede expresar como un promedio ponderado de la estimación del perfil de respuesta medio para el promedio poblacional y el perfil de respuesta observado de la unidad  $i$ .

## **2.7 Diagnóstico del modelo lineal mixto**

### **2.7.1 Análisis de residuos**

Después de ajustar un modelo es importante comprobar si los supuestos distribucionales se cumplen. Este proceso involucra el uso de una técnica informal denominada examen gráfico de residuos.

Los métodos de diagnóstico para los modelos lineales clásicos están bien establecidos. En contraste, los diagnósticos para modelos mixtos son más difíciles de realizar y de interpretar, porque el modelo es más complejo debido a la presencia de efectos aleatorios y diferentes estructuras de covarianza.

En esta sección, se definen los residuos que se desprenden de un modelo lineal mixto y se comenta acerca de sus características. Además se muestran algunas técnicas gráficas muy utilizadas para este objetivo.

### **2.7.2 Residuos del modelo lineal mixto**

El modelo lineal mixto para la unidad  $i$ , presentado en la sección 2.3,

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i \quad i=1, \dots, N,$$

con  $\mathbf{b}_i$  y  $\boldsymbol{\varepsilon}_i$  independientes,  $\mathbf{b}_i \sim N_k^{\text{iid}}(\mathbf{0}, \mathbf{D}^*)$   $\mathbf{D}^* = \sigma^2 \mathbf{D}$ ,  $\boldsymbol{\varepsilon}_i \sim N_n^{\text{iid}}(\mathbf{0}, \mathbf{R}^*)$   $\mathbf{R}^* = \sigma^2 \mathbf{R}$ ,

se puede expresar más conveniente como,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}, \quad (2.7.1)$$

siendo,

$$\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_1, \dots, \mathbf{Y}'_N)' , \mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_N)' , \mathbf{b} = (\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_N)' , \boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_1, \boldsymbol{\varepsilon}'_2, \dots, \boldsymbol{\varepsilon}'_N)'$$

y  $\mathbf{Z} = \bigoplus_{i=1}^N \mathbf{Z}_i$  donde  $\oplus$  representa la suma directa.

Los supuestos se escriben en forma conjunta,

$$\begin{pmatrix} \mathbf{b} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N_{kN+N^*} \left( \begin{pmatrix} \mathbf{0}_{kN} \\ \mathbf{0}_{N^*} \end{pmatrix}, \begin{pmatrix} \sigma^2 \mathbf{G} & \mathbf{0}_{kN \times N^*} \\ \mathbf{0}_{N^* \times kN} & \sigma^2 \boldsymbol{\Sigma} \end{pmatrix} \right) \quad (2.7.2)$$

con  $N^* = Nn$ ,  $\mathbf{G} = \mathbf{I}_N \otimes \mathbf{D}$  y  $\boldsymbol{\Sigma} = \mathbf{I}_N \otimes \mathbf{R}$ ,  $\otimes$  denota el producto Kronecker e  $\mathbf{I}_N$  la matriz identidad de orden  $N$ . Bajo el modelo (2.7.1) la matriz de variancias y covariancias de  $\mathbf{Y}$ , cuyos elementos se sintetizan en un vector  $\boldsymbol{\theta}$  de parámetros de covariancia es,

$$\text{Cov}(\mathbf{Y}) = \mathbf{V} = \sigma^2 (\mathbf{Z}\mathbf{G}\mathbf{Z}' + \boldsymbol{\Sigma}). \quad (2.7.3)$$

Debido a que los modelos lineales mixtos distinguen dos fuentes de variabilidad (intra y entre individuos) se definen tres tipos de residuos.

#### Residuos marginales.

Se definen como la diferencia entre el vector de respuestas y su media marginal estimada,

$$\mathbf{r}_M = \mathbf{Y} - E(\hat{\mathbf{Y}}) = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Z}\hat{\mathbf{b}} + \hat{\boldsymbol{\varepsilon}}, \quad (2.7.4)$$

que predicen los errores marginales  $\boldsymbol{\xi} = \mathbf{Y} - E(\mathbf{Y}) = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}$ .

$$E(\mathbf{r}_M) = \mathbf{0}, \text{Cov}(\mathbf{r}_M) = \mathbf{V} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'. \quad (2.7.5)$$

### Residuos condicionales.

Se definen como la diferencia entre el vector de respuestas y su media condicional estimada,

$$\mathbf{r}_C = \mathbf{Y} - E(\hat{\mathbf{Y}} / \mathbf{b}) = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{b}}, \quad (2.7.6)$$

que predicen los errores condicionales  $\boldsymbol{\varepsilon} = \mathbf{Y} - E(\mathbf{Y} / \mathbf{b}) = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}$ .

$$E(\mathbf{r}_C) = 0, \text{ Cov}(\mathbf{r}_C) = \sigma^2 \mathbf{Q}\mathbf{\Sigma}, \quad (2.7.7)$$

siendo,

$\mathbf{Q} = \mathbf{M}(\mathbf{I} - \mathbf{X}\mathbf{T})$  matriz simétrica de dimensión  $(N^* \times N^*)$  con rango  $(N^* - p)$ ,

$\mathbf{T} = (\mathbf{X}'\mathbf{M}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}$  matriz de dimensión  $(p \times N^*)$ ,

$\mathbf{M} = \sigma^2 \mathbf{V}^{-1} = \sigma^2 [\sigma^2 (\mathbf{Z}\mathbf{G}\mathbf{Z}' + \boldsymbol{\Sigma})]^{-1} = (\mathbf{Z}\mathbf{G}\mathbf{Z}' + \boldsymbol{\Sigma})^{-1}$  matriz simétrica definida positiva de dimensión  $(N^* \times N^*)$  de rango completo.

Notar que  $\mathbf{r}_M = \mathbf{r}_C + \mathbf{Z}\hat{\mathbf{b}}$  y  $\mathbf{r}_C = \mathbf{r}_M - \mathbf{Z}\hat{\mathbf{b}}$ . Si no hubiera efectos aleatorios  $\mathbf{r}_M = \mathbf{r}_C$ .

Los residuos condicionales y marginales expresados como (2.7.6) y (2.7.7) se

consideran residuos ordinarios, siendo  $\mathbf{r}_C = \begin{bmatrix} r_{C_1} \\ r_{C_2} \\ \vdots \\ r_{C_q} \\ \vdots \\ r_{C_{N^*}} \end{bmatrix}$  y  $\mathbf{r}_M = \begin{bmatrix} r_{M_1} \\ r_{M_2} \\ \vdots \\ r_{M_q} \\ \vdots \\ r_{M_{N^*}} \end{bmatrix}$ .

### Residuos de los efectos aleatorios.

$$\mathbf{r}_{EA} = \mathbf{Z}\hat{\mathbf{b}} = E(\hat{\mathbf{Y}} / \mathbf{b}) - E(\hat{\mathbf{Y}}), \quad (2.7.8)$$

que predicen una combinación lineal de los efectos aleatorios  $\mathbf{Z}\mathbf{b} = E(\mathbf{Y} / \mathbf{b}) - E(\mathbf{Y})$ .

$$E(\mathbf{r}_{EA}) = 0, \text{ Cov}(\mathbf{r}_{EA}) = \sigma^2 \mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{Q}\mathbf{Z}\mathbf{G}\mathbf{Z}'. \quad (2.7.9)$$

En el anexo se presenta la deducción de las expresiones (2.7.5), (2.7.7) y (2.7.9).

Los residuos así calculados reciben el nombre de residuos ordinarios. En general, no es conveniente usar los residuos ordinarios para verificar los supuestos del modelo ya que éstos están relacionados y no tienen variancia constante.

### 2.7.3 Residuos estudentizados, externamente estudentizados y de Pearson

Para eliminar el efecto de variancias distintas de los residuos se proponen diferentes escalamientos de los mismos.

#### Residuos estudentizados.

Se obtienen como el cociente entre el residuo ordinario y la estimación del desvío estándar de este residuo,

$$r_q^{\text{student}} = \frac{r_q}{\sqrt{\hat{V}(r_q)}}, q = 1, \dots, N^*. \quad (2.7.10)$$

#### Residuos externamente estudentizados.

Se obtienen como el cociente entre el residuo ordinario y la estimación del desvío estándar de este residuo calculada sin la observación  $q$ ,

$$r_q^{\text{ext.student}} = \frac{r_q}{\sqrt{\hat{V}_{-q}(r_q)}}, q = 1, \dots, N^*. \quad (2.7.11)$$

#### Residuos de Pearson.

Se obtienen como el cociente entre el residuo ordinario y la estimación del desvío estándar de la variable respuesta

$$r_q^{\text{pearson}} = \frac{r_q}{\sqrt{\hat{V}(Y_q)}}, q = 1, \dots, N^*. \quad (2.7.12)$$

#### 2.7.4 Residuos puros y confundidos

Un residuo según Hilden-Minton (1995) se considera,

- Residuo puro si depende sólo de las componentes fijas y del error que predice.
- Residuo confundido si además de depender de las componentes fijas y del error que predice, están en función de otros errores.

Para poder visualizar las definiciones anteriores los vectores  $\hat{\beta}$ ,  $\hat{b}$  y los residuos se expresan de una forma alternativa,

$$\hat{\beta} = (X'V^{-1}X)^{-1} X'V^{-1}Y = TY, \quad (2.7.13)$$

$$\hat{b} = \sigma^2 GZ'V^{-1}(Y - X\hat{\beta}) = \sigma^2 GZ' \frac{M}{\sigma^2} (Y - X\hat{\beta}) = GZ'QY, \quad (2.7.14)$$

$$r_M = Y - X\hat{\beta} = M^{-1}QY = (I - T)\xi, \quad (2.7.15)$$

$$r_C = Y - X\hat{\beta} - Z\hat{b} = \Sigma QY = \Sigma QZb + \Sigma Q\epsilon, \quad (2.7.16)$$

$$r_{EA} = Z\hat{b} = ZGZ'QY = ZGZ'QZb + ZGZ'Q\epsilon. \quad (2.7.17)$$

La deducción de las expresiones (2.7.14), (2.7.15), (2.7.16) y (2.7.17) se presentan en el anexo.

Los residuos  $r_C$  y  $r_{EA}$  no sólo se expresan en función de los errores que representan sino también de otros errores, por lo cual se observa que están confundidos con  $b$  y  $\epsilon$  respectivamente.

#### 2.7.5 Residuos con mínima confusión

Supongamos que se quiere chequear la normalidad de  $\epsilon$  examinando  $r_C$ . En la sección anterior se vio que  $r_C$  está confundido con  $b$ . Entonces, si  $Zb$  se aleja notablemente de la distribución normal,  $r_C$  puede parecer no normal aún cuando  $\epsilon$  sea normal. Por otro lado, si

$\mathbf{Zb}$  es verdaderamente normal, esto puede esconder la no normalidad de  $\boldsymbol{\varepsilon}$ . No obstante, debido a que  $\mathbf{r}_C$  es un residuo confundido no es confiable para refutar el supuesto de normalidad.

Hilden-Minton (1995) comentó que la habilidad para detectar no normalidad de  $\boldsymbol{\varepsilon}$ , usando  $\mathbf{r}_C$ , decrece a medida que  $\text{Cov}(\boldsymbol{\Sigma}\mathbf{Q}\mathbf{Z}\mathbf{b}) = \sigma^2\boldsymbol{\Sigma}\mathbf{Q}\mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{Q}\boldsymbol{\Sigma}$  aumenta en relación a  $\text{Cov}(\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}$ ; esto motiva la definición de la “fracción de confusión” ( $\text{FC}_q$ ) para el residuo condicional  $q$ ,  $r_{Cq} = \mathbf{u}'_q \mathbf{r}_C$  ( $\mathbf{u}_i$  es la columna  $q$  de  $\mathbf{I}_{N^*}$ ),  $q=1, \dots, N^*$ .

$$\text{FC}_q = \frac{\text{Cov}(\mathbf{r}_C / \boldsymbol{\varepsilon})}{\text{Cov}(\mathbf{r}_C)} = \frac{\mathbf{u}'_q \boldsymbol{\Sigma}\mathbf{Q}\mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{Q}\boldsymbol{\Sigma}\mathbf{u}_q}{\mathbf{u}'_q \boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}\mathbf{u}_q} = 1 - \frac{\mathbf{u}'_q \boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}\mathbf{u}_q}{\mathbf{u}'_q \boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}\mathbf{u}_q}, 0 \leq \text{FC}_q \leq 1. \quad (2.7.18)$$

La deducción de la expresión (2.7.18) se presenta en el anexo.

Esta fracción ( $\text{FC}_q$ ) representa la proporción de la variabilidad de  $r_{Cq}$  atribuida a la confusión con los efectos aleatorios. Cuanto mayor es (2.7.18), mayor es el grado de confusión de  $r_{Cq}$ .

La propuesta de Hilden-Minton es considerar una transformación lineal,  $\mathbf{L}'\mathbf{r}_C$ , que tenga mínima confusión.

Si el vector  $\boldsymbol{\ell}_{q^*}$ , de dimensión  $N^* \times 1$ , es la columna  $q^*$  de  $\mathbf{L}$  ( $N^* \times (N^*-p)$ ),  $\boldsymbol{\ell}'_{q^*}\mathbf{r}_C$  se dice que tiene mínima confusión si  $\lambda_{q^*} = \frac{\boldsymbol{\ell}'_{q^*} \boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}\boldsymbol{\ell}_{q^*}}{\boldsymbol{\ell}'_{q^*} \boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}\boldsymbol{\ell}_{q^*}} \quad (2.7.19)$ ,  $q^*=1, \dots, (N^*-p)$  es un máximo.

Por lo tanto, los residuos con mínima confusión se obtienen maximizando (2.7.19) sujeto a la restricción que la matriz de covariancias  $\text{Cov}(\boldsymbol{\ell}'_{q^*}\mathbf{r}_C) \propto \boldsymbol{\ell}'_{q^*} \boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}\boldsymbol{\ell}_{q^*} > 0$ . Debido a que  $\boldsymbol{\Sigma}$  es una matriz definida positiva y de rango completo ( $\text{rango}(\boldsymbol{\Sigma}) = N^*$ ), sólo interesa el

espacio no nulo de  $\mathbf{Q}$  ( $\text{rango}(\mathbf{Q}) = N^* - p$ ). Para ello se utiliza la descomposición espectral de  $\Sigma^{-1/2} \mathbf{Q} \Sigma^{-1/2} = \mathbf{W} \mathbf{\Pi} \mathbf{W}'$ , donde

- $\mathbf{\Pi} = \text{diag}\{\pi_1, \dots, \pi_{N^*-p}\}$ ,  $\pi_1 \geq \dots \geq \pi_{N^*-p}$  son los autovalores positivos no nulos ordenados de  $\Sigma^{-1/2} \mathbf{Q} \Sigma^{-1/2}$ ,
- $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_{N^*-p})$ ,  $\mathbf{w}_1, \dots, \mathbf{w}_{N^*-p}$  autovectores normalizados asociados a los autovalores de  $\Sigma^{-1/2} \mathbf{Q} \Sigma^{-1/2}$  y  $\mathbf{W}' \mathbf{W} = \mathbf{I}_{(N^*-p)}$ .

Se considera  $\ell_{q^*} = \Sigma^{-1/2} \mathbf{W} \mathbf{\Pi}^{-1/2} \mathbf{v}_{q^*}$  para cualquier vector  $\mathbf{v}_{q^*}$ ,  $q^* = 1, \dots, (N^* - p)$ .

Sustituyendo en (2.7.19) resulta  $\lambda_{q^*} = \frac{\mathbf{v}_{q^*}' \mathbf{\Pi} \mathbf{v}_{q^*}}{\mathbf{v}_{q^*}' \mathbf{v}_{q^*}}$ . Lo cual implica que

$$\pi_{N^*-p} \leq \lambda_{q^*} \leq \pi_1, \text{ con } \pi_{N^*-p} \leq \dots \leq \pi_1 \leq 1.$$

Si se supone que  $\mathbf{v}_{q^*}$  es la columna  $q^*$  de  $\mathbf{I}_{(N^*-p)}$ ,

$\ell_{q^*} = \Sigma^{-1/2} \mathbf{W} \mathbf{\Pi}^{-1/2} \mathbf{v}_{q^*} = \Sigma^{-1/2} \mathbf{w}_{q^*} \pi_{q^*}^{-1/2}$ ,  $\mathbf{w}_{q^*}$  es la columna  $q^*$  de  $\mathbf{W}$ . Por lo tanto,

$$\ell_{q^*}' \mathbf{r}_C = \frac{1}{\sqrt{\pi_{q^*}}} \mathbf{w}_{q^*}' \Sigma^{-1/2} \mathbf{r}_C = \sqrt{\pi_{q^*}} \mathbf{w}_{q^*}' \Sigma^{-1/2} \mathbf{Y}, \quad (2.7.20)$$

siendo  $(1 - \pi_{q^*})$  la fracción de confusión.

La deducción de la expresión (2.7.20) se presenta en el anexo.

$$\text{Cov}(\ell_{q^*}' \mathbf{r}_C, \ell_{q^{**}}' \mathbf{r}_C) = \sigma^2 \ell_{q^*}' \Sigma \mathbf{Q} \Sigma \ell_{q^{**}} = \frac{\sigma^2}{\sqrt{\lambda_{q^{**}} \lambda_{q^*}}} \mathbf{w}_{q^*}' \mathbf{W} \mathbf{\Pi} \mathbf{W}' \mathbf{w}_{q^{**}} = \begin{cases} \sigma^2 & \text{si } q^* = q^{**} \\ 0 & \text{si } q^* \neq q^{**} \end{cases}$$

Los residuos con mínima confusión  $\ell_{q^*}' \mathbf{r}_C$  forman una secuencia de residuos no correlacionados homocedásticos con variancia  $\sigma^2$  y con mínima fracción de confusión. Así,  $\ell_{q^*}' \mathbf{r}_C / \sigma$  son residuos estandarizados no correlacionados.

### 2.7.6 Técnicas gráficas para evaluar los supuestos

Las técnicas gráficas, para evaluar el cumplimiento de los supuestos del modelo, consideran los distintos residuos que se desprenden de un modelo lineal mixto. En la literatura, muchas son las opiniones acerca de cuáles son las técnicas gráficas más apropiadas que se deben utilizar para diagnosticar cada supuesto postulado. A continuación se presentan algunas de ellas, explicitando además, qué permiten diagnosticar y cómo se tienen que analizar.

En general, se sugiere trabajar con algún escalamiento de los residuos, como los vistos en el apartado 2.7.3.

#### *Gráfico del residuo marginal vs. variable explicativa (covariable).*

Se usa para evaluar si la parte media del modelo está bien especificada. Se espera que los elementos de  $\mathbf{r}_M$  varíen aleatoriamente en torno de cero en virtud de la veracidad de esta suposición.

#### *Gráfico del residuo condicional vs. tiempo.*

Es útil para evaluar si los efectos aleatorios están correctamente especificados en el modelo o si se omitió algún término en la parte media e identificar observaciones atípicas. Si los supuestos se cumplen los elementos de  $\mathbf{r}_C$  varían aleatoriamente alrededor de cero.

#### *Gráfico del residuo condicional estudentizado vs. n° de observación.*

Permite visualizar la presencia de observaciones atípicas. Se trazan dos líneas horizontales en -3 y 3 y se comprueba si algún punto cae fuera de ellas.



*Gráfico del residuo condicional estudentizado vs. valor ajustado.*

Permite chequear la homocedasticidad de los errores condicionales. Si los puntos se distribuyen aleatoriamente alrededor del cero y no muestran ningún patrón se piensa que este supuesto se cumple.

*Gráfico probabilístico normal del residuo con mínima confusión estudentizado con bandas de confianza simuladas.*

Hilden-Minton (1995) sugiere realizar un gráfico probabilístico normal (se grafican los cuantiles muestrales versus los cuantiles poblacionales de una distribución normal estándar) con bandas de confianza (“envelopes”) simuladas (Atkinson, 1985; Flack and Flores, 1989) de los residuos con confusión mínima estandarizados  $\ell'_q \mathbf{r}_C / \sigma$ ,  $q^* = 1, \dots, (N^* - p)$ . Las bandas se agregan con el fin de quitarle subjetividad al gráfico usual de probabilidad normal.

Las bandas simuladas se obtienen del siguiente modo:

- 1) Se ajusta el modelo y se genera una muestra simulada de  $N^*$  observaciones independientes usando el modelo ajustado como si este fuera el modelo verdadero.
- 2) Se ajusta el modelo para la muestra generada, se calculan los residuos con mínima confusión estudentizados para este modelo y se ordenan en forma decreciente.
- 3) Se repite el paso 1) y 2)  $m$  veces.
- 4) Para cada uno de los  $m$  conjuntos de estadísticas de orden, se obtiene el mínimo y el máximo.

El valor mínimo y máximo de los  $m$  conjuntos de estadísticas de orden producen las bandas. Para realizar el gráfico se usan estos valores y los residuos con mínima confusión estudentizados ordenados de los datos originales.

Por lo tanto, si hay puntos que caen fuera de las bandas se va a concluir que los errores condicionales no parecen muestreados de una población normal.

*Gráfico del EBLUP vs. n° de unidad.*

Permite identificar unidades atípicas. Se trazan dos líneas horizontales en -3 y 3 y se observa si algún punto cae fuera de ellas. Si las matrices  $\mathbf{Z}_i$  son iguales para todas las unidades entonces los EBLUP ( $\hat{\mathbf{b}}$ ) son comparables y se pueden reemplazar por los residuos de los efectos aleatorios ( $\mathbf{r}_{EA} = \mathbf{Z}\hat{\mathbf{b}}$ ).

*Gráfico probabilístico normal del EBLUP con bandas de confianza simuladas.*

Evalúa la normalidad de los efectos aleatorios. Las bandas se generan de la misma manera que se expuso anteriormente. Si hay puntos que caen fuera de las bandas se va a concluir que los efectos aleatorios no parecen muestreados de una población normal.

Para resumir la utilidad de los distintos tipos de residuos se adjunta la siguiente tabla extraída del artículo presentado por Nobre y Singer (2007).

Tabla 2.7.1: Residuos de un modelo mixto para diagnosticar los supuestos.

Para diagnosticar	Tipo de residuo	Gráfico
Especificación de la parte media ( $\mathbf{X}\boldsymbol{\beta}$ )	Marginal	Residuo marginal vs. variable explicativa
Presencia de observaciones atípicas	Condicional	Residuo condicional estudentizado vs. n° de observación
Homocedasticidad del error condicional	Condicional	Residuo condicional estudentizado vs. valor ajustado
Normalidad del error condicional	Condicional	Gráfico probabilístico normal del residuo con mínima confusión estudentizado
Presencia de unidades atípicas	EBLUP	EBLUP vs. n° de unidad
Normalidad de los efectos aleatorios	EBLUP	Gráfico probabilístico normal del EBLUP

### 3. APLICACIÓN

Para la aplicación se utilizan datos suministrados por el Hospital Escuela “Eva Perón” de Granadero Baigorria. Estos datos provienen de un estudio observacional que tenía como finalidad analizar los valores de la procalcitonina (pct) para la identificación temprana y el pronóstico de sepsis, sepsis severa y shock séptico. En el estudio se consideraron pacientes ingresados a Unidad de Terapia Intensiva (UTI) del hospital en un período de 18 meses con Síndrome de Respuesta Inflamatoria Sistemática (SRIS). El SRIS está conformado por signos clínicos, hematológicos y bioquímicos. Estos signos pueden ocurrir en respuesta a estímulos tanto infecciosos como no infecciosos. Este dilema tiene connotaciones terapéuticas, dado que el diagnóstico y la administración de antibióticos en forma apropiada y precoz en pacientes con sepsis (respuesta del huésped a la infección) es esencial, pero innecesaria y aún perjudicial en el caso de SRIS por causa no infecciosa.

Múltiples variables clínicas, de laboratorio y microbiológicas están disponibles actualmente como guía para discriminar pacientes sin sepsis (SRIS por causa no infecciosa), con sepsis, sepsis severa o shock séptico (SRIS por causa infecciosa según el grado de severidad de la infección, siendo el último el diagnóstico más comprometido). Los cultivos microbiológicos son el mejor método para diagnosticar infección, pero requieren habitualmente 48 horas o más para la confirmación.

La procalcitonina ha sido propuesta como un marcador útil en el diagnóstico, pronóstico y monitoreo de la terapéutica de las enfermedades bacterianas y micóticas. Las mediciones de procalcitonina han demostrado ser más rápidas, sensibles y específicas en la discriminación de los pacientes con o sin sepsis que cualquiera de los marcadores sanguíneos establecidos.

Para aplicar la metodología expuesta se modela la evolución de la procalcitonina durante la estadía del paciente en UTI, considerando el diagnóstico presuntivo realizado al ingreso a la misma: “no sepsis”, “sepsis”, “sepsis severa” o “shock séptico”. A cada uno de

los pacientes se le mide el nivel de procalcitonina al ingreso, a las 24 y 48 horas de permanencia en UTI.

Se comienza con un análisis exploratorio utilizando gráficos adecuados para este tipo de estudio, que permiten examinar el comportamiento de los datos. Luego, a partir de la evidencia muestral se proponen modelos con distintas estructuras de variancias y covariancias para la parte aleatoria del modelo, cuyos parámetros se estiman mediante el método de máxima verosimilitud restringida. Con la ayuda de los criterios de información de Akaike y bayesiano de Schwarz y según el número de parámetros de covariancia a estimar se decide cuál es el más conveniente.

Una vez elegida la estructura de variancias y covariancias de los efectos aleatorios y los errores se evalúa la parte media del modelo. Los efectos fijos se estiman utilizando el método de máxima verosimilitud y mediante el test de razón de verosimilitud se realizan las correspondientes pruebas de interés.

Una etapa importante, cuando se trabaja con modelos, es la comprobación de los supuestos que se realizan sobre ellos. La adecuación del modelo se analiza gráficamente usando los residuos marginales, condicionales y una predicción de los efectos aleatorios (EBLUP) para evaluar los correspondientes supuestos. Se pone especial atención en el estudio del supuesto de normalidad de los errores condicionales, para el cual se calculan los residuos con mínima confusión y se efectúa un gráfico probabilístico normal (gráfico qq) con bandas (“envelopes”) simuladas.

El modelo lineal mixto se ajusta utilizando el procedimiento “MIXED” del programa estadístico SAS, del cual se obtiene; la estimación de los parámetros, los valores de AIC y BIC; las estadísticas del test de Wald y del test de razón de verosimilitud usadas para realizar pruebas de hipótesis sobre los efectos fijos, los residuos condicionales, marginales y los EBLUP. Los residuos con mínima confusión se calculan recurriendo al procedimiento “IML”,

también del programa SAS. Para realizar los gráficos probabilístico normal se ejecuta el paquete “envelope” del programa estadístico R.

Las sentencias para calcular los residuos con mínima confusión se adaptaron de lo realizado por García et al. (2012). Estas sentencias y las utilizadas para realizar los gráficos probabilístico normal se encuentran en el anexo.

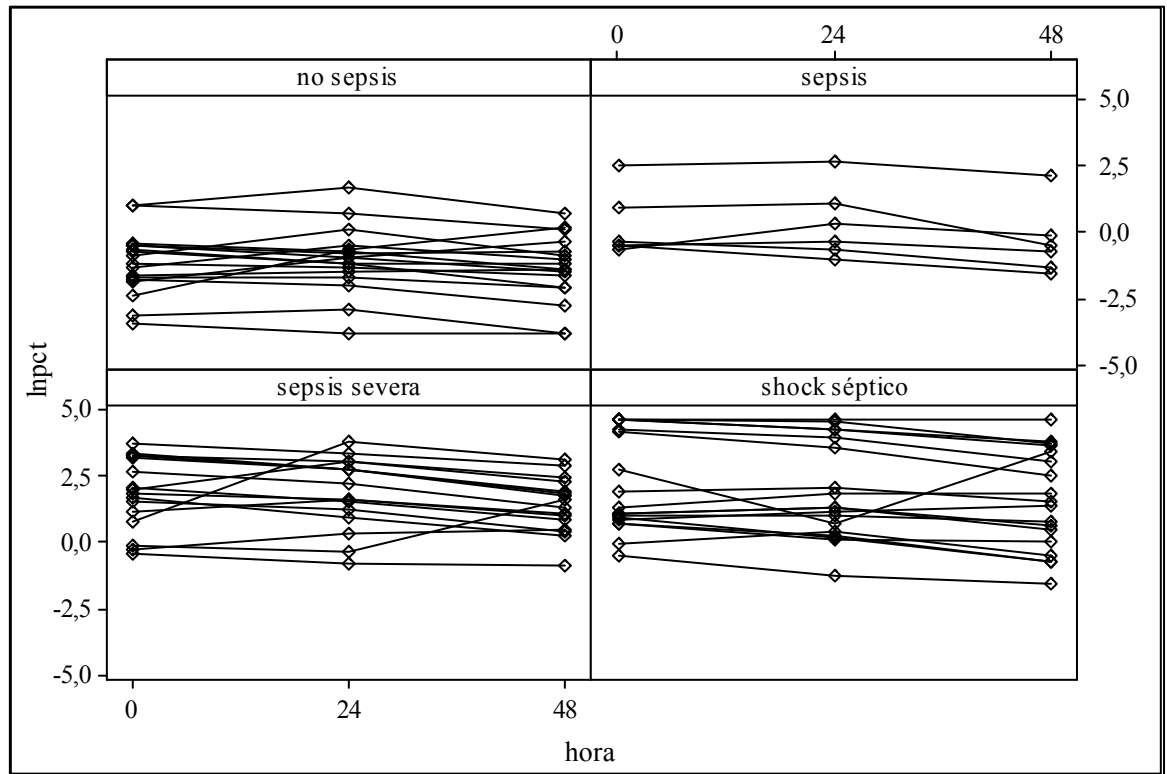
Con el fin de disminuir la variabilidad de las observaciones se aplica el logaritmo neperiano a los valores de pct (se va a llamar “lnpct”). Debido a que esta es una función monótona creciente, las interpretaciones que se realizan sobre la variable respuesta trasformada se pueden extender a la variable original. Al aplicar el logaritmo los valores de la variable respuesta se suavizan, maximizando la tendencia de la respuesta debida al paso del tiempo (la propensión que tiene la unidad a responder).

### **3.1 Análisis descriptivo**

#### **3.1.1 Exploración de la estructura de la parte media**

A continuación, se presentan gráficos que permiten observar el comportamiento de la procalcitonina (lnpct) a través del tiempo según el diagnóstico establecido al comienzo del estudio.

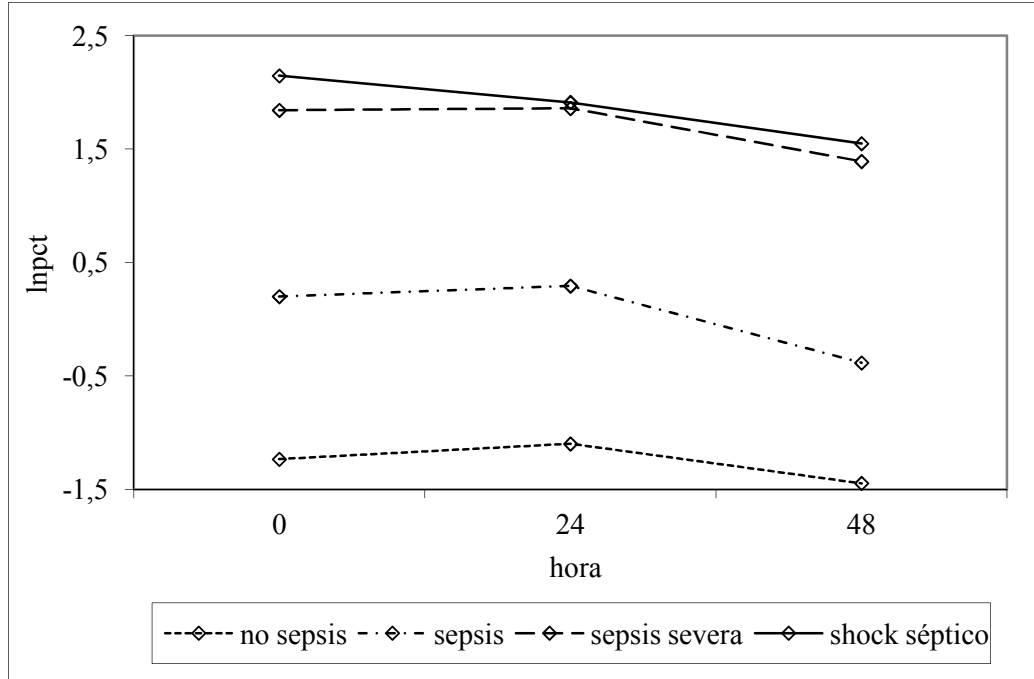
Gráfico 3.1.1: Gráficos de perfiles individuales para cada diagnóstico.



Los pacientes, para los cuales se les confirmó que el SRIS no era por causa infecciosa, presentan, en general, valores de Inpct más bajos dentro de las 48 horas que los pacientes a los que sí se les detectó infección. Se observa que los valores del Inpct, para algunos pacientes, fluctúan mucho a lo largo del estudio. En los cuatro grupos la variabilidad entre pacientes para cada ocasión parece similar en las tres ocasiones. Los perfiles parecen tener diferentes respuestas en la medición basal, pero la evolución a través del tiempo parece similar (Gráfico 3.1.1).

Para visualizar si el comportamiento del Inpct a lo largo del estudio resulta similar en los cuatro diagnósticos se construye el gráfico de perfiles promedios.

Gráfico 3.1.2: Gráfico de perfiles promedio según diagnóstico.



Los valores promedios de Inpct parecen estar en relación con la severidad del diagnóstico, por ejemplo, los promedios más altos se presentan para el diagnóstico más grave (shock séptico). La evolución del Inpct medio a lo largo del estudio es similar para los distintos diagnósticos (Gráfico 3.1.2).

Luego de realizar un estudio exploratorio se decidió elegir una función cuadrática en el tiempo para explicar el patrón que sigue el Inpct.

Además al observar los Gráficos 3.1.1 y 3.1.2 se nota que al comienzo del estudio los pacientes tienen valores de Inpct de distinta magnitud por lo cual se decide incorporar una ordenada aleatoria al modelo que ayude a explicar la heterogeneidad de la población.

El modelo utilizado es el siguiente

$$Y_{ij} = \beta_{0i} + \beta_{1i}t_j + \beta_{2i}t_j^2 + \epsilon_{ij} \text{ con } i = \overline{1,57} \text{ } j = \overline{1,3},$$

$$\beta_{0i} = \beta_{0s} + b_{0i} \text{ para el grupo } s \text{ con } s = \overline{1,4},$$

$$\beta_{1i} = \beta_{1s} \text{ para el grupo } s \text{ con } s = \overline{1,4},$$

$$\beta_{2i} = \beta_{2s} \text{ para el grupo } s \text{ con } s = \overline{1,4},$$

siendo  $t_j$  las horas de permanencia en UTI.

Se supone que  $b_{0i}$  y  $\boldsymbol{\varepsilon}_i = [\boldsymbol{\varepsilon}'_{i1} \quad \boldsymbol{\varepsilon}'_{i2} \quad \boldsymbol{\varepsilon}'_{i3}]'$  son independientes,  $b_{0i} \stackrel{\text{ind}}{\sim} N(0, D_s^*)$ ,

$$D_s^* = \sigma^2 D_s = V_s(b_{0i}); \quad \boldsymbol{\varepsilon}_i \stackrel{\text{ind}}{\sim} N_3(\mathbf{0}, \mathbf{R}_s^*), \quad \mathbf{R}_s^* = \sigma^2 \mathbf{R}_s. \quad (3.1.1)$$

### 3.1.2 Exploración de la estructura de covariancia

En estos modelos, se distinguen dos fuentes de variación, dentro y entre pacientes, por lo que se debe especificar una estructura para la matriz de variancias y covariancias de los efectos aleatorios y una para los errores aleatorios.

Como forma de guiar esta elección se utilizan las variancias y las correlaciones muestrales de cada grupo y algunas técnicas de diagnóstico gráfico (matriz de gráficos de dispersión, gráfico de Draftman, gráfico PRISM y correlograma muestral). La matriz de gráficos de dispersión y el gráfico de Draftman no se diferencian por grupo debido a que hay pocas unidades en cada uno de ellos.

Tabla 3.1.1: Medidas descriptivas de Inpct según el diagnóstico.

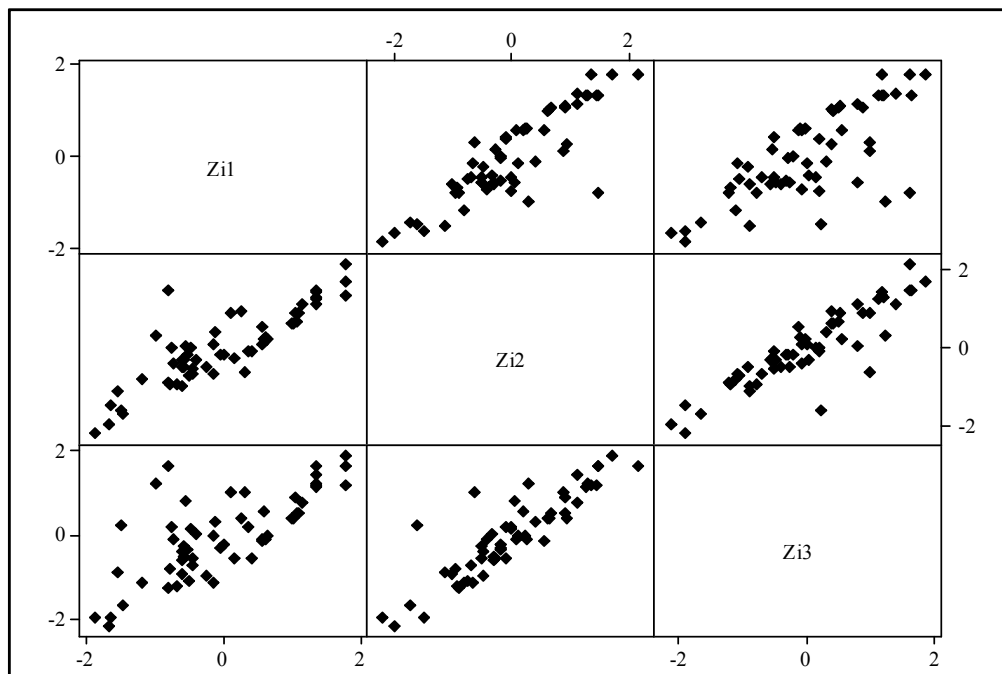
	<b>Diagnóstico</b>			
<b>Medida</b>	<b>no sepsis</b>	<b>sepsis</b>	<b>sepsis severa</b>	<b>shock séptico</b>
<b>Variancia(<math>Y_{i1}</math>)</b>	2,42	2,52	2,72	3,64
<b>Variancia(<math>Y_{i2}</math>)</b>	2,54	2,76	2,68	3,66
<b>Variancia(<math>Y_{i3}</math>)</b>	2,56	2,68	2,10	3,74
<b>Corr(<math>Y_{i1}, Y_{i2}</math>)</b>	0,86 <sup>s</sup>	0,92 <sup>s</sup>	0,75 <sup>s</sup>	0,94 <sup>s</sup>
<b>Corr(<math>Y_{i1}, Y_{i3}</math>)</b>	0,69 <sup>s</sup>	0,84 <sup>s</sup>	0,54 <sup>s</sup>	0,93 <sup>s</sup>
<b>Corr(<math>Y_{i2}, Y_{i3}</math>)</b>	0,90 <sup>s</sup>	0,94 <sup>s</sup>	0,85 <sup>s</sup>	0,89 <sup>s</sup>

**Nota:** (s) significativo al 5%.



En los cuatro grupos se puede pensar que hay homogeneidad de variancias en las tres ocasiones. Las variancias parecen aumentar levemente con el grado de severidad del diagnóstico. Con respecto a las correlaciones entre mediciones repetidas de un mismo paciente son relativamente grandes y significativas en todos los casos. Sin embargo, se aprecia que en general las correlaciones de las medidas consecutivas son más parecidas y altas que las separadas por dos rezagos (Tabla 3.1.1).

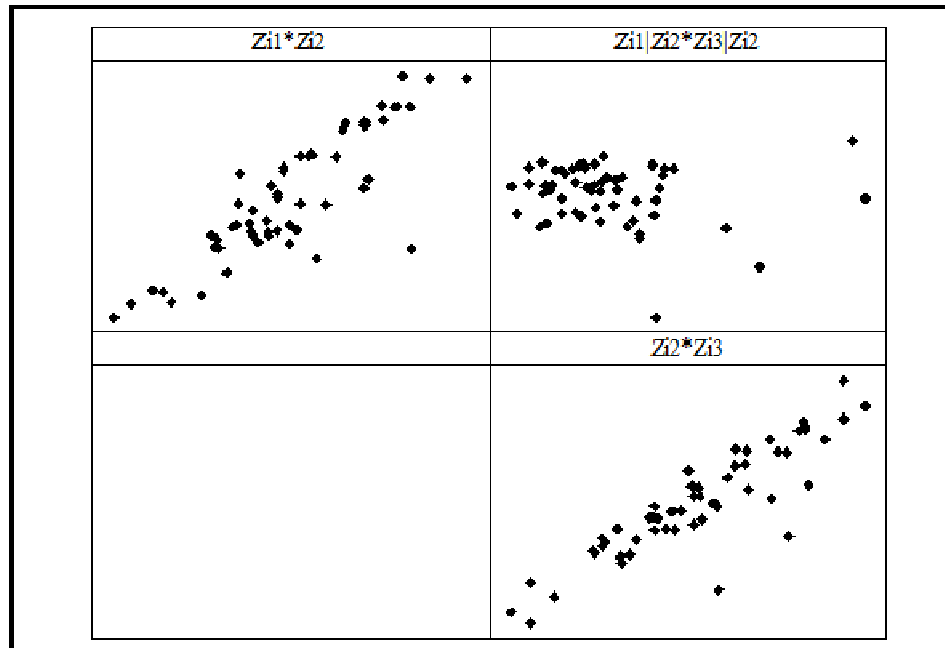
Gráfico 3.1.3: Gráfico de Draftman.



**Nota:**  $Z_{ij}$  es la variable respuesta estandarizada

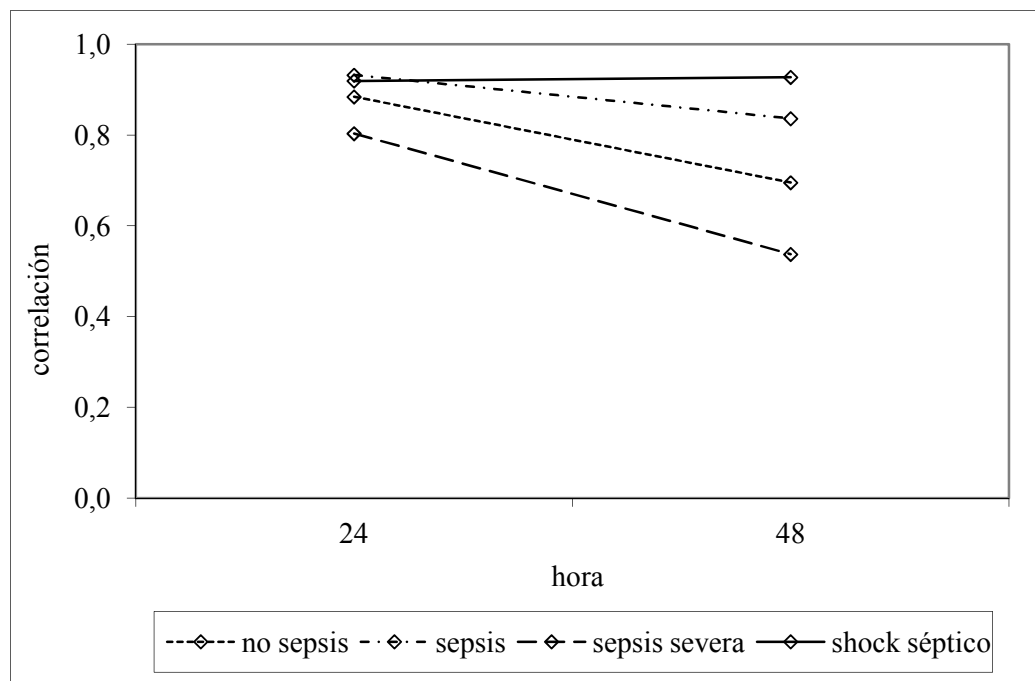
El gráfico de Draftman utiliza la variable  $\ln p_{ct}$  estandarizada lo cual permite dilucidar mejor el patrón de correlación intra-unidad. Este gráfico (Gráfico 3.1.3) muestra que las correlaciones son positivas, altas y parecidas. La correlación entre valores de  $\ln p_{ct}$  que distan 48 horas decrece con respecto a las que distan 24 horas.

Gráfico 3.1.4: Gráfico PRISM.



El gráfico PRISM refuerza lo evidenciado en los gráficos anteriores, las mediciones consecutivas de un paciente están muy relacionadas mientras que la correlación para las mediciones rezagadas dos tiempos se puede considerar inexistente.

Gráfico 3.1.5: Correlograma muestral.



La correlación entre los valores de  $\ln pect$  que distan 24 horas (medidas consecutivas) es superior a 0,8 para los cuatro diagnósticos. Si bien las observaciones que difieren 48 horas es

están menos correlacionadas, su correlación es alta. Este patrón se repite en todos los grupos, excepto para los que se les diagnosticó shock séptico para el cual parece no disminuir (Gráfico 3.1.5).

Del análisis exploratorio surge como posible estructura de covariancia intra-unidad la autorregresiva de orden uno. Esto se confirma mediante un examen analítico, comparando con otras estructuras que también se cree que son viables, como la estructura de simetría compuesta, de Toeplitz y arbitraria.

## **3.2 Análisis inferencial**

### **3.2.1 Selección de la estructura de covariancia**

Debido a que el modelo de la covariancia depende del modelo postulado para la media y para evitar que cualquier especificación incorrecta de esta última afecte la estimación de la matriz de variancias y covariancias se plantea para la misma un modelo maximal, en este caso como la cantidad de grupos es pequeña y el estudio es balanceado se utiliza el modelo (3.1.1).

La Tabla 3.2.1 presenta la log-verosimilitud restringida maximizada por los estimadores REML, los criterios de información de AIC y BIC y la cantidad de parámetros de covariancia a estimar para modelos con distintas estructuras de covariancia para los errores aleatorios. Además, se considera la posibilidad que la matriz de variancias y covariancias de los errores aleatorios y la variancia de la ordenada aleatoria sean distintas por grupo.

Con el objetivo de seleccionar el modelo que mejor ajuste los datos pero que a su vez sea más parco entre los que presentan menor AIC y BIC (Tabla 3.2.1) se realizan tests de hipótesis. Esta selección se realiza mediante tests de razón de verosimilitud ya que los modelos se encuentran anidados. Se utiliza un nivel de significación ( $\alpha$ ) de 0,05.

1º) Se comparan las estructuras (3) y (2) de la Tabla 3.2.1.

H<sub>0</sub>) La estructura (2) es adecuada

H<sub>1</sub>) La estructura (3) es adecuada

$$G_{\text{obs}} = -2(\hat{\ell}_{\text{red}} - \hat{\ell}_{\text{comp}}) = 520,7 - 510,6 = 10,1 \quad \chi^2_{6,0,05} = 12,59$$

Decisión:  $G_{\text{obs}} < \chi^2_{6,0,05}$  entonces no rechazo H<sub>0</sub>.

En base a la evidencia muestral, se puede pensar que la estructura (2) es adecuada.

2º) Se comparan las estructuras (2) y (1) de la Tabla 3.2.1.

H<sub>0</sub>) La estructura (1) es adecuada

H<sub>1</sub>) La estructura (2) es adecuada

$$G_{\text{obs}} = -2(\hat{\ell}_{\text{red}} - \hat{\ell}_{\text{comp}}) = 526,3 - 520,7 = 5,6 \quad \chi^2_{3,0,05} = 7,81$$

Decisión:  $G_{\text{obs}} < \chi^2_{3,0,05}$  entonces no rechazo H<sub>0</sub>.

En base a la evidencia muestral, se puede pensar que la estructura adecuada

$$\mathbf{D}^* = \sigma^2 \mathbf{D} = \mathbf{V}(\mathbf{b}_{0i}) \text{ y } \mathbf{R}_i^* = \sigma^2 \mathbf{R}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix} \quad i=1, \dots, 57. \quad (3.2.1)$$

Tabla 3.2.1: Criterios de log-verosimilitud penalizada.

Estructura	-2 Res Log Likelihood	AIC	BIC	N° de parámetros
<b>D* igual por grupo - R* igual por grupo</b>				
No estructurada	525,8	539,8	554,1	7
Simetría compuesta	537	543	549,1	3
Autoregresiva de orden uno (1)	526,3	532,3	538,4	3
Toeplitz	526,3	534,3	542,4	4
<b>D* igual por grupo - R<sub>s</sub>* distinta por grupo</b>				
No estructurada	497,7	547,7	598,7	25
Simetría compuesta		no converge		9
Autoregresiva de orden uno	516,3	532,3	548,6	9
Toeplitz	506,4	532,4	558,9	13
<b>D<sub>s</sub>* distinta por grupo - R* igual por grupo</b>				
No estructurada		no converge		10
Simetría compuesta	532,4	544,4	556,7	6
Autoregresiva de orden uno (2)	520,7	530,7	540,9	6
Toeplitz	520,7	534,7	549	7
<b>D<sub>s</sub>* distinta por grupo - R<sub>s</sub>* distinta por grupo</b>				
No estructurada		no converge		28
Simetría compuesta	527,6	551,6	576,1	12
Autoregresiva de orden uno (3)	510,6	528,6	547	12
Toeplitz		no converge		16

**Nota:** método de estimación REML

### 3.2.2 Selección de la estructura de la parte media

Una vez elegida la estructura de covariancias se estiman los parámetros del modelo

(3.1.1) con las estructuras (3.2.1) para **D\*** y **R\*** usando el método de ML.

Tabla 3.2.2: Estimación de los efectos fijos.

Efecto fijo	Estimación	Error estándar	p-value
$\beta_{01}$	-1,2336	0,3429	0,0005
$\beta_{02}$	0,2006	0,5771	0,7288
$\beta_{03}$	1,8430	0,3534	<0,0001
$\beta_{04}$	2,1461	0,3332	<0,0001
$\beta_{11}$	0,01570	0,01129	0,1673
$\beta_{12}$	0,01998	0,01901	0,2958
$\beta_{13}$	0,01072	0,01164	0,3591
$\beta_{14}$	-0,00712	0,01098	0,5180
$\beta_{21}$	-0,00042	0,000214	0,0530
$\beta_{22}$	-0,00067	0,000361	0,0659
$\beta_{23}$	-0,00042	0,000221	0,0603
$\beta_{24}$	-0,00011	0,000208	0,5954

**Nota:** El p-value corresponde al test de Wald para probar la significación de cada parámetro.

Para verificar hipótesis de interés para el problema se realizan pruebas usando el test de razón de verosimilitud. Se utiliza un nivel de significación ( $\alpha$ ) de 0,05.

Pruebas de interés:

1) La tasa de cambio en el tiempo es la misma para los cuatro grupos.

$$H_0) \begin{cases} \beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} \\ \beta_{21} = \beta_{22} = \beta_{23} = \beta_{24} \end{cases}$$

$H_1$ ) al menos uno es distinto

$$G_{\text{obs}} = -2(\hat{\ell}_{\text{red}} - \hat{\ell}_{\text{comp}}) = 444,3 - 442,6 = 1,7 \quad \chi^2_{6,0,05} = 12,59$$

Decisión:  $G_{\text{obs}} < \chi^2_{6,0,05}$  entonces no rechazo  $H_0$ .

En base a la evidencia muestral, se concluye que el patrón de cambio del lnpt es el mismo para los cuatro grupos.

2) Se comprueba si al ingreso en UTI los pacientes con distintos diagnósticos tienen el mismo valor promedio de Inpct.

$$H_0) \beta_{01} = \beta_{02} = \beta_{03} = \beta_{04}$$

$H_1$ ) al menos uno es distinto

$$G_{\text{obs}} = -2(\hat{\ell}_{\text{red}} - \hat{\ell}_{\text{comp}}) = 526,3 - 520,7 = 41,7 \quad \chi^2_{3,0,05} = 7,81$$

Decisión:  $G_{\text{obs}} > \chi^2_{3,0,05}$  entonces rechazo  $H_0$ .

En base a la evidencia muestral, se concluye que al ingreso en UTI el promedio de Inpct no es el mismo para los cuatro diagnósticos.

La expresión  $L'\beta$  de las hipótesis se presentan en el anexo.

Después de realizar las pruebas de hipótesis sobre los parámetros del modelo (3.1.1) se obtiene el siguiente modelo estimado

$$\hat{Y}_{ij} = (-1,1159 + \hat{b}_{0i}) + 0,0075 t_j - 0,0004 t_j^2 \quad \text{si } i \text{ pertenece al grupo con "no sepsis"}$$

$$\hat{Y}_{ij} = (0,1448 + \hat{b}_{0i}) + 0,0075 t_j - 0,0004 t_j^2 \quad \text{si } i \text{ pertenece al grupo con "sepsis"}$$

$$\hat{Y}_{ij} = (1,8411 + \hat{b}_{0i}) + 0,0075 t_j - 0,0004 t_j^2 \quad \text{si } i \text{ pertenece al grupo con "sepsis severa"}$$

$$\hat{Y}_{ij} = (2,0552 + \hat{b}_{0i}) + 0,0075 t_j - 0,0004 t_j^2 \quad \text{si } i \text{ pertenece al grupo con "shock séptico"}$$

$$i=1, \dots, 57$$

$$\hat{D}^* = \hat{V}(b_{0i}) = 0,7639 \quad \hat{R}^* = \begin{bmatrix} 1,2418 & 0,9969 & 0,8003 \\ 0,9969 & 1,2418 & 0,9969 \\ 0,8003 & 0,9969 & 1,2418 \end{bmatrix}. \quad (3.2.2)$$

Los valores predichos del efecto aleatorio  $(\hat{b}_{0i})$  para cada paciente se muestran en la tabla I del anexo.

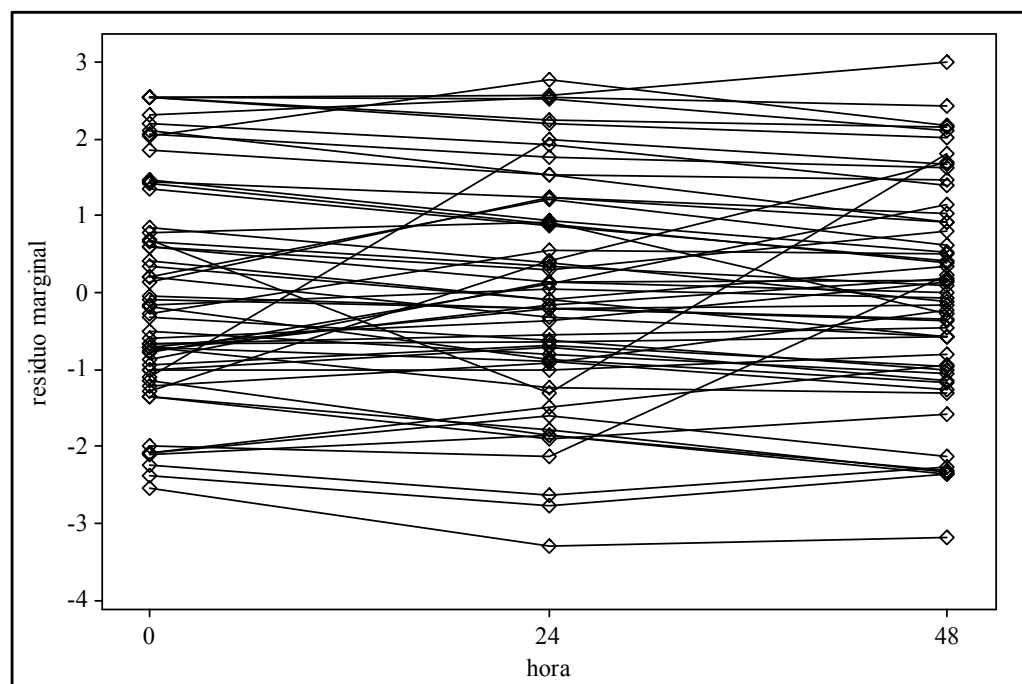
### 3.2.3 Diagnóstico del modelo

Los tres tipos de residuos que se desprenden del modelo lineal mixto (3.2.2) se utilizan en diferentes gráficos para chequear el cumplimiento de los supuestos.

En una primera instancia se evalúa si el modelo postulado para la parte media es adecuado para los datos, realizando los gráficos de los residuos marginales y de los residuos condicionales vs las horas en UTI (0-24-48).

Los valores de los residuos marginales y condicionales se encuentran en la tabla II del anexo.

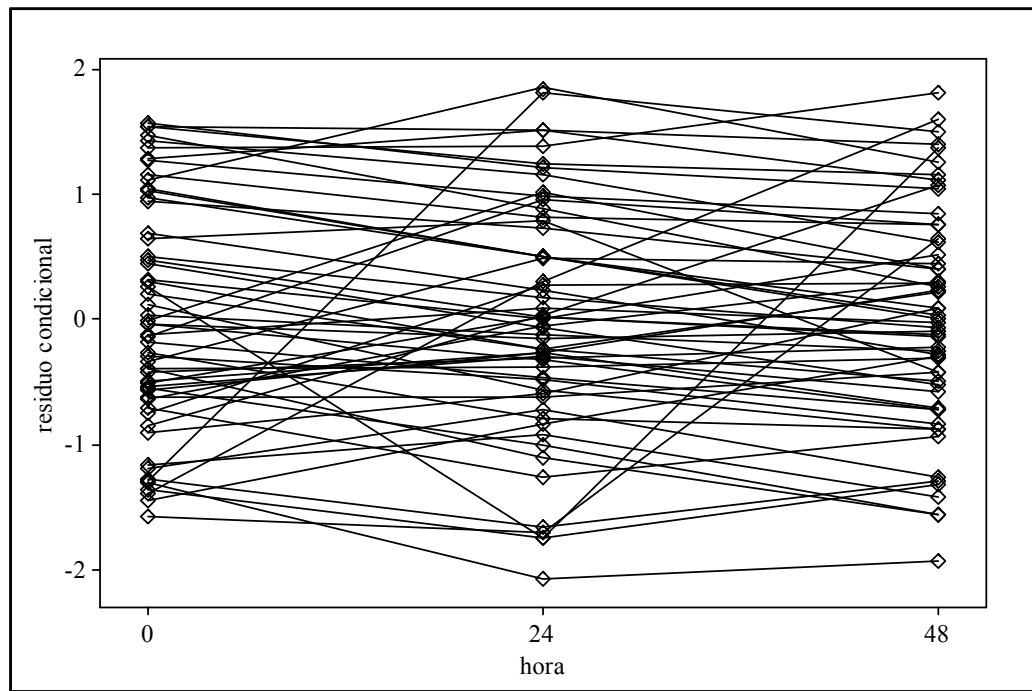
Gráfico 3.2.1: Gráfico de perfiles de los residuos marginales vs. la hora.



Los perfiles no muestran ningún patrón y varían aleatoriamente en torno de cero, indicando que no es necesario agregar términos a la parte media del modelo.



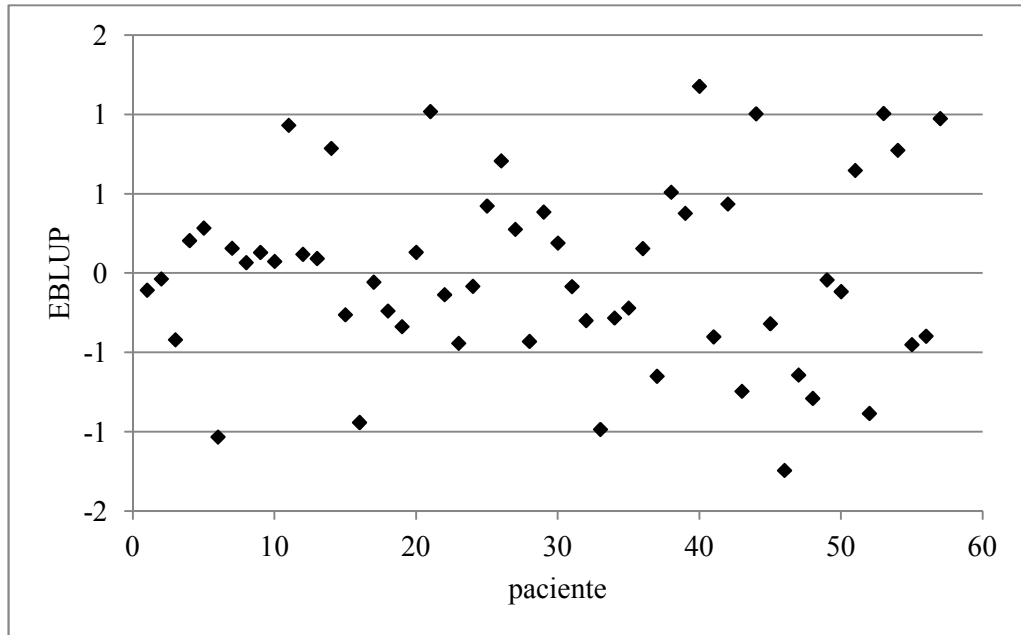
Gráfico 3.2.2: Gráfico de perfiles de los residuos condicionales vs. la hora.



El comportamiento que muestran los perfiles permite concluir que no es necesario agregar otros efectos aleatorios al modelo (Gráfico 3.2.2).

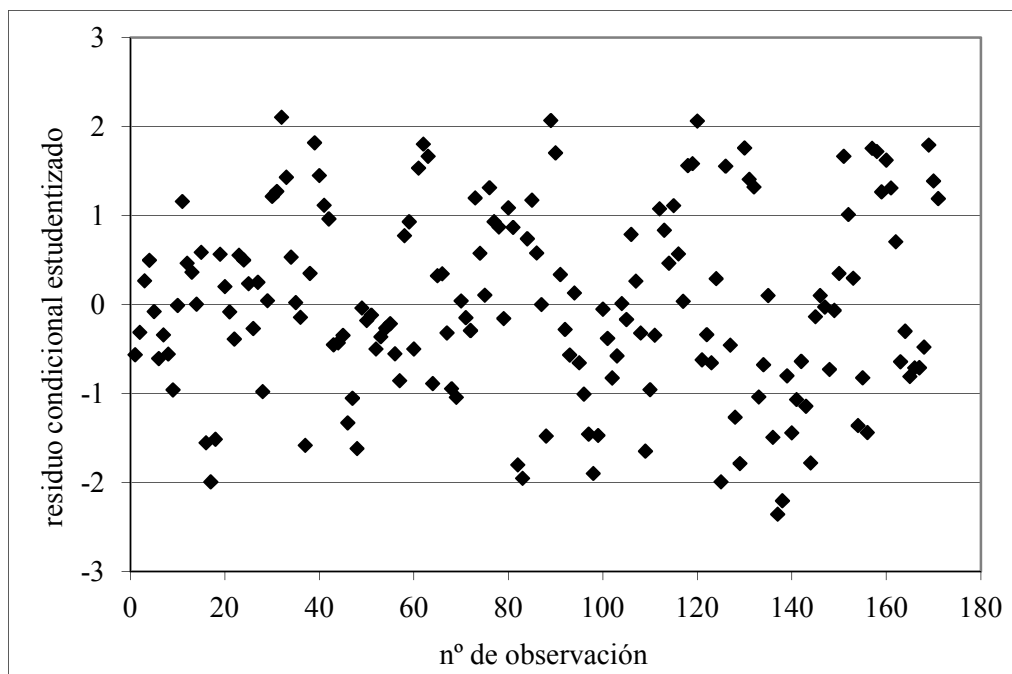
Para identificar si hay pacientes u observaciones atípicas se construye el gráfico del EBLUP vs el número de paciente, y el de los residuos condicionales estudentizados vs el número de observación respectivamente. Si en ambos gráficos los puntos se pueden encerrar entre dos líneas paralelas ubicadas en  $\pm 3$  se concluye que no hay valores alejados del resto.

Gráfico 3.2.3: Gráfico de los EBLUP vs. número de paciente



En el Gráfico 3.2.3 no se ve ningún punto fuera de las bandas horizontales  $\pm 3$ , por lo tanto ningún paciente se considera atípico.

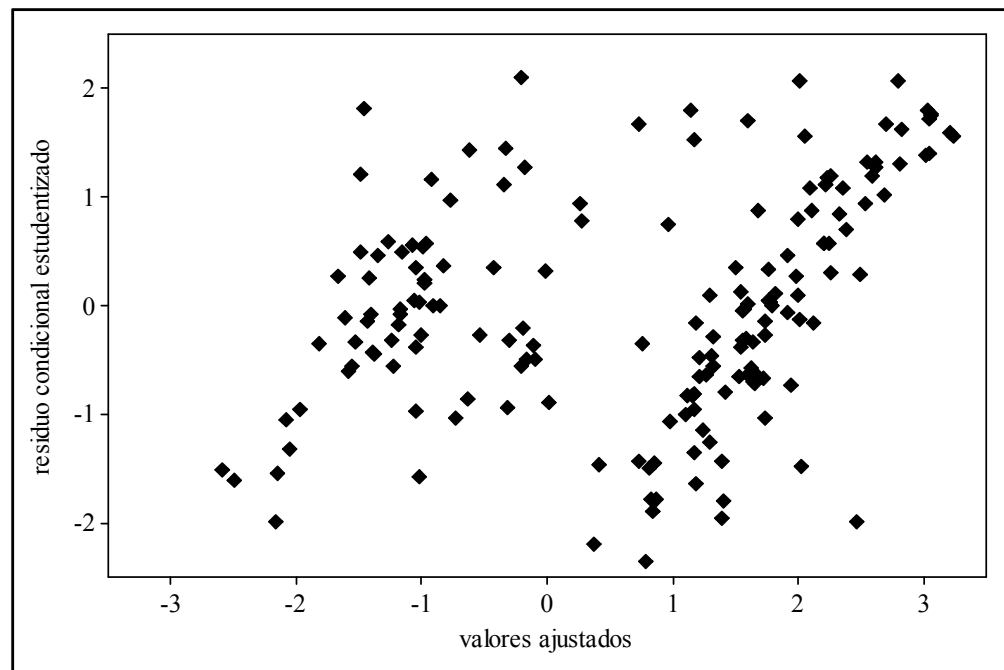
Gráfico 3.2.4: Gráfico de los residuos condicionales estudentizados vs. n° de observación.



En el Gráfico 3.2.4 no se ve ningún punto fuera de las bandas horizontales  $\pm 3$ , por lo tanto ninguna observación se considera atípica.

La homocedasticidad de los errores condicionales se comprueba usando los residuos que los predicen.

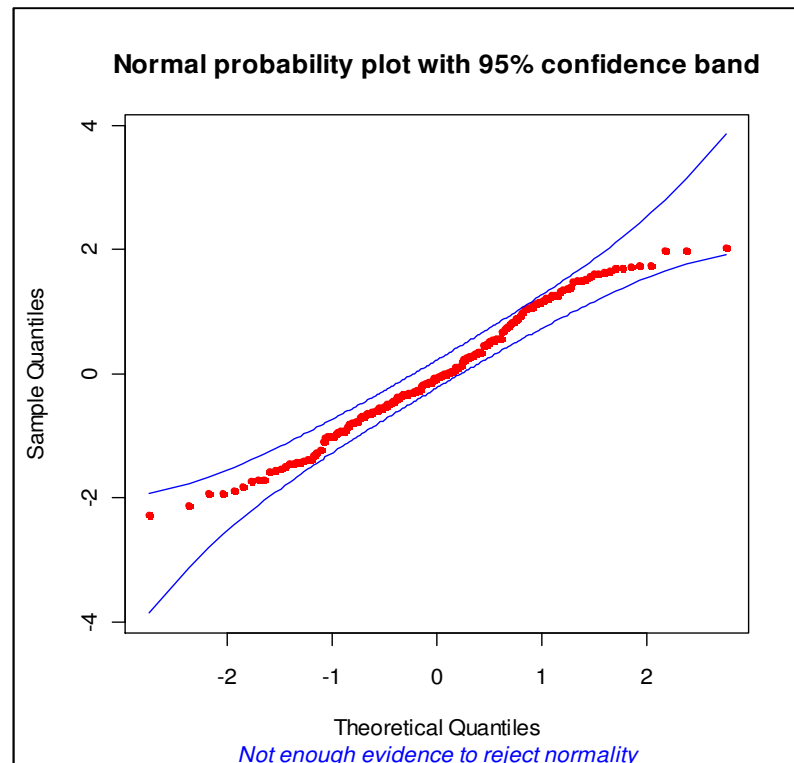
Gráfico 3.2.5: Gráfico de los residuos condicionales estudentizados vs. los valores ajustados.



En el Gráfico 3.2.5 se observa que los puntos se distribuyen aleatoriamente alrededor del cero no mostrando ningún patrón, esto lleva a creer que el supuesto de variancia constante de los errores condicionales se cumple.

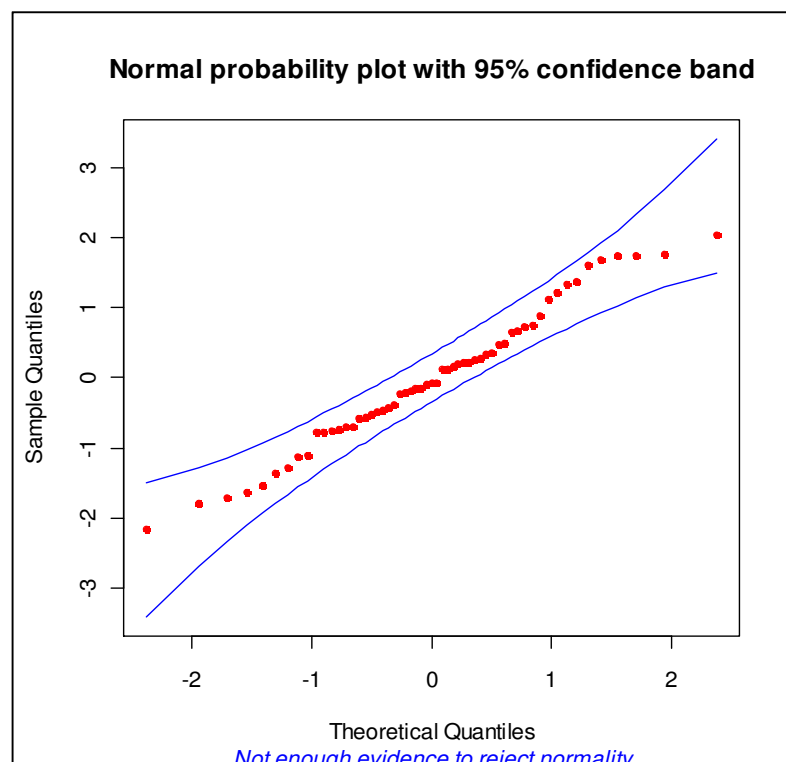
Para evaluar si la normalidad de los errores condicionales usando los residuos condicionales es errónea cuando el grado de confusión con los efectos aleatorios es grande se realizan dos gráficos probabilístico normal con bandas simuladas, uno de los residuos condicionales estudentizados y otro de los EBLUP.

Gráfico 3.2.6: Gráfico probabilístico normal de los residuos condicionales estudentizados.



Según se observa en el Gráfico 3.2.6 de los residuos condicionales no hay suficiente evidencia para pensar que los errores condicionales no se distribuyen normalmente.

Gráfico 3.2.7: Gráfico probabilístico normal de los EBLUP.

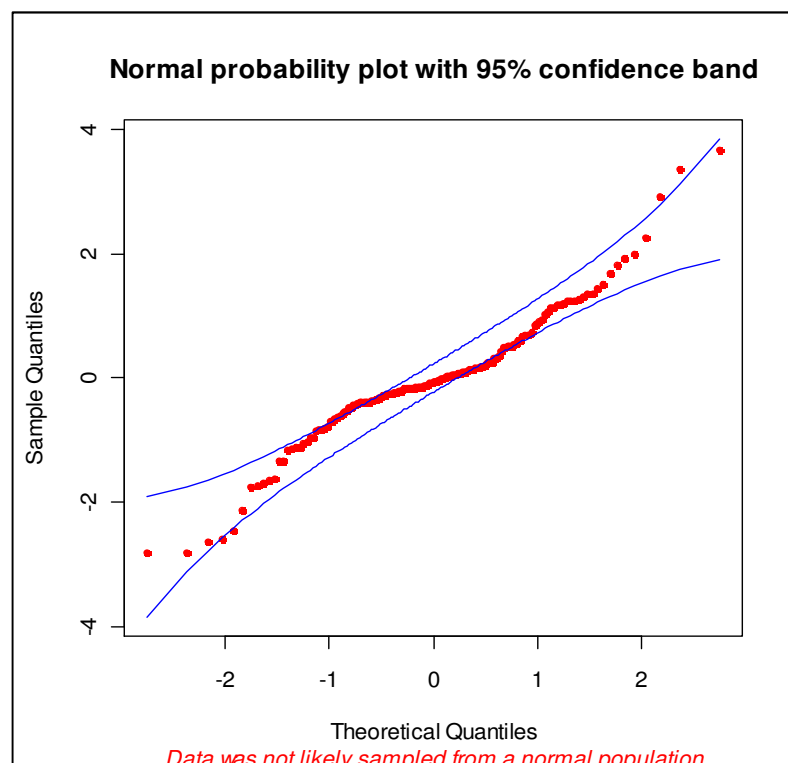


El Gráfico 3.2.7 de los EBLUP manifiesta que no hay suficiente evidencia para pensar que los efectos aleatorios no se distribuyen normalmente.

Los residuos condicionales muestran que los errores condicionales se distribuyen normalmente. Esta conclusión se puede deber a la confusión que éstos presentan con los efectos aleatorios, los cuales parecen ser muestreados de una población normal, según indican los EBLUP. Para comprobar esto se sugiere usar un gráfico probabilístico normal con bandas simuladas del residuo con mínima confusión estudentizado.

Se recuerda que los residuos con mínima confusión ayudan a evaluar la distribución de los errores condicionales, eliminando la confusión que introducen en ellos la presencia de los efectos aleatorios.

Gráfico 3.2.8: Gráfico probabilístico normal de los residuos con mínima confusión estudentizados.



Se puede observar que hay puntos fuera de las bandas, como consecuencia se concluye que los errores condicionales no parecen muestreados de una población normal (Gráfico 3.2.8).

Una explicación para esta situación podría ser que la distribución de los efectos aleatorios inciden sobre la distribución que muestran los residuos condicionales, forzando a que el error condicional se distribuya normalmente cuando realmente no lo sea, tal como se observa al utilizar los residuos con mínima confusión.

Por este motivo se recomienda utilizar para comprobar la normalidad de los errores condicionales un gráfico de probabilidad normal de los residuos con mínima confusión.

#### 4. CONSIDERACIONES FINALES

Para analizar datos provenientes de un estudio longitudinal se utiliza un modelo estadístico. Debido a la correlación existente entre las mediciones repetidas de la variable de interés en una misma unidad no se pueden usar los modelos clásicos. Cuando la variable respuesta se puede aproximar a una distribución Gaussiana se puede usar un modelo lineal mixto el cual consta de una ecuación (que relaciona la evolución de la respuesta a través de las ocasiones con covariables y explicita la variación que existe entre y dentro de las unidades por separado) y de supuestos sobre los efectos aleatorios y los errores aleatorios.

Tres tipos de residuos, marginales, condicionales y de los efectos aleatorios, se desprenden de este modelo. Estos residuos, estandarizados o no, se utilizan para evaluar los distintos supuestos del mismo. Además, algunos de ellos están sujetos a confusión debido a la presencia en los mismos de otras fuentes de variación distintas a los errores que representan. Específicamente, para evaluar la normalidad se utilizan los residuos condicionales estandarizados que presentan confusión, en el sentido que no sólo son función del error que predicen, no resultando adecuados para evaluar este supuesto. Para minimizar el efecto de la confusión se sugiere utilizar una transformación lineal de los residuos condicionados, los residuos con mínima confusión (“least confounded residuals”). En la tesis para implementar este tipo de residuos se realiza un programa usando el procedimiento “IML” de SAS. La programación de estos residuos se torna más difícil a medida que la estructura de covariancia es más compleja.

Para la aplicación se utilizó un modelo para evaluar la evolución de la procalcitonina a través del tiempo, para los pacientes clasificados en grupos según la severidad del proceso inflamatorio.

El estudio se centró en el análisis de los supuestos del modelo postulado, resultando las siguientes consideraciones:

- el uso de gráficos de residuos permitió comprobar que la estructura planteada para la parte media del modelo resulta adecuada y que no existen ni observaciones ni unidades atípicas.

- el gráfico de los residuos condicionales estudentizados y los valores ajustados permitió comprobar la homocedasticidad de los errores condicionales.

- el gráfico probabilístico normal con bandas de confianza (“envelopes”) simuladas utilizado para evaluar el supuesto de normalidad evidenció que el uso de los residuos condicionales puede no ser apropiado. En cambio, el uso de estos residuos transformados muestra que la confusión presente en los residuos condicionales se debe tomar en cuenta.

Se enfatiza que debido al incumplimiento de la normalidad este modelo necesita revisiones para poder ser usado con confianza con fines predictivos.



## 5. BIBLIOGRAFÍA

- \* Akaike, Hirotugu (1974). A new look at the statistical model identification. IEEE Transactions on automatic control, vol. 19, n° 6, p. 716-723.
- \* Atkinson, C. A. (1985). Plots, transformations and regression: an introduction to graphical methods of diagnostic regression analysis. Oxford University Press, Oxford.
- \* Bozgodan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. Psychometrika, vol. 52, p. 345-370.
- \* Davidian, M. (2001). Applied longitudinal data analysis. Department of Statistics. North Carolina State University. Lectures Notes.
- \* Dawson, K. S.; Gennings, C.; Carter, W. H. (1997). Two techniques useful in detecting correlation structure in repeated measures data. The American Statistician, vol. 51, p. 275-283.
- \* Fitzmaurice, G.; Laird, N.; Ware, J. (2004). Applied longitudinal analysis. John Wiley & Sons.
- \* Flack, Virginia F.; Flores, Rafael A. (1989). Using simulated envelopes in the evaluation of normal probability plots of regression residuals. Technometrics, vol. 31, n° 2.
- \* Garcia, M.; Rapelli, C. (2012). Estudio de la validez de los supuestos en los modelos lineales mixtos mediante un análisis de residuos. FABICIB, vol.16, p. 11-20. Universidad Nacional del Litoral, Argentina.
- \* Hilden-Minton, J. A. (1995). Multilevel diagnostics for mixed and hierarchical linear models. PhD Thesis. University of California, Los Angeles.
- \* Laird, N. M.; Ware, J. H. (1982). Random effects models for longitudinal data. Biometrics, vol. 38.
- \* Lange, N.; Ryan, L. (1989). Assessing normality in random effects models. The Annals of Statistics, vol. 17, n° 2.

- \* Littell, R.C.; Milliken, G.A; Stroup, W.W.; Wolfinger, R.D. (1996). SAS® System for mixed models. Cary, NC: SAS Institute Inc.
- \* Nobre, J. S.; Singer, J. M. (2007). Residual analysis for linear mixed models. *Biometrical Journal* vol. 49, 6, p. 863-875.
- \* Pinheiro, J. C.; Bates, D. M. (2000). *Mixed-effects in S and S-PLUS*. Springer, New York.
- \* Rolke, Wolfgang A. (2003). An extension of the normal probability plot. Departamento de matemática. Universidad de Puerto Rico. Mayaguez.
- \* SAS Institute Inc (2011). *SAS/STAT User's Guide: Version 9.2, Second Edition*.
- \* Weiss, R. E.; Lazaro, C. G. (1992). Residual plots for repeated measures. *Statistics in Medicine* 11, p. 115–124.
- \* Weiss, R. E. (1997). Exploratory data graphics for repeated measures. Department of biostatistics UCLA School of Public Health, Los Angeles.
- \* West, Brandy T.; Welch, Kathleen B.; Galecki, Andrzej T. (2007). *Linear mixed models. A practical guide using statistical software*. Chapman & Hall, New York.
- \* Zimmerman, D. L. (2000). Viewing the correlation structure of longitudinal data though a PRISM. *The American Statistician*, vol.54, nº4.

## 6. ANEXO

### Demostraciones:

#### Ecuaciones (2.6.4):

$$\begin{aligned}\hat{\mathbf{Y}}_i &= \mathbf{X}_i \hat{\boldsymbol{\beta}} + \mathbf{Z}_i \hat{\mathbf{b}}_i = \mathbf{X}_i \hat{\boldsymbol{\beta}} + \mathbf{Z}_i \hat{\mathbf{D}}^* \mathbf{Z}_i' \hat{\mathbf{V}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) = (\mathbf{I}_{n_i} - \mathbf{Z}_i \hat{\mathbf{D}}^* \mathbf{Z}_i' \hat{\mathbf{V}}_i^{-1}) \mathbf{X}_i \hat{\boldsymbol{\beta}} + \mathbf{Z}_i \hat{\mathbf{D}}^* \mathbf{Z}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{Y}_i = \\ &= (\hat{\mathbf{R}}_i^* \hat{\mathbf{V}}_i^{-1}) \mathbf{X}_i \hat{\boldsymbol{\beta}} + (\mathbf{I}_{n_i} - \hat{\mathbf{R}}_i^* \hat{\mathbf{V}}_i^{-1}) \mathbf{Y}_i\end{aligned}$$

$$\hat{\mathbf{V}}_i \hat{\mathbf{V}}_i^{-1} = \mathbf{I}_{n_i} = (\mathbf{Z}_i \hat{\mathbf{D}}^* \mathbf{Z}_i' + \hat{\mathbf{R}}_i) \hat{\mathbf{V}}_i^{-1} = \mathbf{Z}_i \hat{\mathbf{D}}^* \mathbf{Z}_i' \hat{\mathbf{V}}_i^{-1} + \hat{\mathbf{R}}_i \hat{\mathbf{V}}_i^{-1}$$

#### Ecuaciones (2.7.5):

$$\mathbf{E}(\mathbf{r}_M) = \mathbf{E}(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) = \mathbf{E}(\mathbf{Y}) - \mathbf{E}(\mathbf{X} \hat{\boldsymbol{\beta}}) = \mathbf{X} \boldsymbol{\beta} - \mathbf{X} \boldsymbol{\beta} = 0$$

$$\text{Cov}(\mathbf{r}_M) = \text{Cov}(\mathbf{M}^{-1} \mathbf{Q} \mathbf{Y}) = \mathbf{M}^{-1} \mathbf{Q} \text{Cov}(\mathbf{Y}) (\mathbf{M}^{-1} \mathbf{Q})' = \mathbf{M}^{-1} \mathbf{Q} \mathbf{V} \mathbf{Q} \mathbf{M}^{-1} = \mathbf{M}^{-1} \sigma^2 \mathbf{Q} \mathbf{M}^{-1} = \sigma^2 \mathbf{M}^{-1} \mathbf{Q} \mathbf{M}^{-1}$$

$$\begin{aligned}\mathbf{Q} \mathbf{V} \mathbf{Q} &= \mathbf{Q} \sigma^2 \mathbf{M}^{-1} \mathbf{Q} = \sigma^2 \mathbf{Q} \mathbf{M}^{-1} \mathbf{Q} = \sigma^2 \mathbf{M} (\mathbf{I} - \mathbf{T}) \mathbf{M}^{-1} \mathbf{M} (\mathbf{I} - \mathbf{T}) = \sigma^2 \mathbf{M} (\mathbf{I} - \mathbf{T}) (\mathbf{I} - \mathbf{T}) = \\ &= \sigma^2 \mathbf{M} (\mathbf{I} - 2\mathbf{T} + \mathbf{T}^2) = \sigma^2 \mathbf{M} (\mathbf{I} - 2\mathbf{T} + \mathbf{T}) = \sigma^2 \mathbf{M} (\mathbf{I} - \mathbf{T}) = \sigma^2 \mathbf{Q}\end{aligned}$$

#### Ecuaciones (2.7.7):

$$\mathbf{E}(\mathbf{r}_C) = \mathbf{E}(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} - \mathbf{Z} \hat{\mathbf{b}}) = \mathbf{E}(\mathbf{Y}) - \mathbf{E}(\mathbf{X} \hat{\boldsymbol{\beta}}) - \mathbf{E}(\mathbf{Z} \hat{\mathbf{b}}) = \mathbf{X} \boldsymbol{\beta} - \mathbf{X} \boldsymbol{\beta} - 0 = 0$$

$$\text{Cov}(\mathbf{r}_C) = \text{Cov}(\boldsymbol{\Sigma} \mathbf{Q} \mathbf{Y}) = \boldsymbol{\Sigma} \mathbf{Q} \text{Cov}(\mathbf{Y}) (\boldsymbol{\Sigma} \mathbf{Q})' = \boldsymbol{\Sigma} \mathbf{Q} \mathbf{V} \mathbf{Q} \boldsymbol{\Sigma} = \boldsymbol{\Sigma} \sigma^2 \mathbf{Q} \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Sigma} \mathbf{Q} \boldsymbol{\Sigma}$$

#### Ecuaciones (2.7.9):

$$\begin{aligned}\mathbf{E}(\hat{\mathbf{b}}) &= \mathbf{E}(\sigma^2 \mathbf{G} \mathbf{Z}' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})) = \sigma^2 \mathbf{G} \mathbf{Z}' \mathbf{V}^{-1} \mathbf{E}(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) = \sigma^2 \mathbf{G} \mathbf{Z}' \mathbf{V}^{-1} [\mathbf{E}(\mathbf{Y}) - \mathbf{E}(\mathbf{X} \hat{\boldsymbol{\beta}})] = \\ &= \sigma^2 \mathbf{G} \mathbf{Z}' \mathbf{V}^{-1} [\mathbf{X} \boldsymbol{\beta} - \mathbf{X} \boldsymbol{\beta}] = 0\end{aligned}$$

$$\text{Cov}(\hat{\mathbf{b}}) = \text{Cov}(\mathbf{G} \mathbf{Z}' \mathbf{Q} \mathbf{Y}) = \mathbf{G} \mathbf{Z}' \mathbf{Q} \text{Cov}(\mathbf{Y}) (\mathbf{G} \mathbf{Z}' \mathbf{Q})' = \mathbf{G} \mathbf{Z}' \mathbf{Q} \mathbf{V} \mathbf{Q} \mathbf{Z} \mathbf{G} = \mathbf{G} \mathbf{Z}' \sigma^2 \mathbf{Q} \mathbf{Z} \mathbf{G} = \sigma^2 \mathbf{G} \mathbf{Z}' \mathbf{Q} \mathbf{Z} \mathbf{G}$$

$$\mathbf{E}(\mathbf{r}_{EA}) = \mathbf{E}(\mathbf{Z} \hat{\mathbf{b}}) = 0$$

$$\text{Cov}(\mathbf{r}_{EA}) = \text{Cov}(\mathbf{Z} \hat{\mathbf{b}}) = \mathbf{Z} \text{Cov}(\hat{\mathbf{b}}) \mathbf{Z}' = \mathbf{Z} \sigma^2 \mathbf{G} \mathbf{Z}' \mathbf{Q} \mathbf{Z} \mathbf{G} \mathbf{Z}' = \sigma^2 \mathbf{Z} \mathbf{G} \mathbf{Z}' \mathbf{Q} \mathbf{Z} \mathbf{G} \mathbf{Z}'$$

#### Ecuación (2.7.14):

$$\begin{aligned}\Sigma QY &= \Sigma M(Y - X\hat{\beta}) = \Sigma(\Sigma^{-1} - \Sigma^{-1}ZC^{-1}Z'\Sigma^{-1})(Y - X\hat{\beta}) = (I - ZC^{-1}Z'\Sigma^{-1})(Y - X\hat{\beta}) = \\ &= (Y - X\hat{\beta}) - ZC^{-1}Z'\Sigma^{-1}(Y - X\hat{\beta}) = (Y - X\hat{\beta}) - ZGZ'M(Y - X\hat{\beta}) = Y - X\hat{\beta} - Z\hat{b}\end{aligned}$$

$$\begin{aligned}QY &= M(I - XT)Y = M(Y - XTY) = M[Y - X(X'MX)^{-1}X'MY] = \\ &= M[Y - X(X'\sigma^2V^{-1}X)^{-1}X'\sigma^2V^{-1}Y] = M[Y - X(X'V^{-1}X)^{-1}X'V^{-1}Y] = M(Y - X\hat{\beta}) \\ \therefore M^{-1}QY &= (Y - X\hat{\beta})\end{aligned}$$

$$\begin{aligned}GZ'M &= GZ'(\Sigma^{-1} - \Sigma^{-1}ZC^{-1}Z'\Sigma^{-1}) = G(Z'\Sigma^{-1} - Z'\Sigma^{-1}ZC^{-1}Z'\Sigma^{-1}) = \\ &= G(CC^{-1}Z'\Sigma^{-1} - Z'\Sigma^{-1}ZC^{-1}Z'\Sigma^{-1}) = G(C - Z'\Sigma^{-1}Z)C^{-1}Z'\Sigma^{-1} = GG^{-1}C^{-1}Z'\Sigma^{-1} = \\ &= C^{-1}Z'\Sigma^{-1}\end{aligned}$$

Ecuación (2.7.15):

$$\begin{aligned}r_M &= Y - X\hat{\beta} = M^{-1}QY = M^{-1}Q(X\beta + Zb + \varepsilon) = M^{-1}QX\beta + M^{-1}Q(Zb + \varepsilon) = \\ &= M^{-1}Q(Zb + \varepsilon) = M^{-1}Q\xi = M^{-1}M(I - T)\xi = (I - T)\xi\end{aligned}$$

$$QX\beta = [M - MX(X'MX)^{-1}X'M]X\beta = MX\beta - MX(X'MX)^{-1}X'MX\beta = MX\beta - MX\beta$$

Ecuación (2.7.16):

$$r_C = Y - X\hat{\beta} - Z\hat{b} = \Sigma QY = \Sigma Q(X\beta + Zb + \varepsilon) = \Sigma QX\beta + \Sigma Q(Zb + \varepsilon) = \Sigma QZb + \Sigma Q\varepsilon$$

Ecuación (2.7.17):

$$\begin{aligned}r_{EA} &= Z\hat{b} = ZGZ'QY = ZGZ'Q(X\beta + Zb + \varepsilon) = ZGZ'QX\beta + ZGZ'Q(Zb + \varepsilon) = \\ &= ZGZ'QZb + ZGZ'Q\varepsilon\end{aligned}$$

$$r_{EA} = r_C - r_M = \Sigma QY - (\Sigma + ZGZ')QY = \Sigma QY - \Sigma QY + ZGZ'QY = ZGZ'QY$$

Ecuación (2.7.18):

$$CF_q = \frac{u_q'\Sigma QZGZ'Q\Sigma u_q}{u_q'\Sigma Q\Sigma u_q} = \frac{u_q'\Sigma(Q - Q\Sigma Q)\Sigma u_q}{u_q'\Sigma Q\Sigma u_q} = 1 - \frac{u_q'\Sigma Q\Sigma Q\Sigma u_q}{u_q'\Sigma Q\Sigma u_q}$$

$$QVQ = Q\sigma^2(ZGZ' - \Sigma)Q = \sigma^2(QZGZ'Q - Q\Sigma Q) = \sigma^2Q$$

Ecuación (2.7.20):

$$\begin{aligned}\therefore \ell'_j \mathbf{r}_C &= \frac{1}{\sqrt{\pi_j}} \mathbf{w}'_j \Sigma^{-1/2} \mathbf{r}_C = \frac{1}{\sqrt{\pi_j}} \mathbf{w}'_j \Sigma^{-1/2} \Sigma \mathbf{QY} = \frac{1}{\sqrt{\pi_j}} \mathbf{w}'_j \Sigma^{1/2} \mathbf{QY} = \frac{1}{\sqrt{\pi_j}} \mathbf{w}'_j \mathbf{W} \Pi \mathbf{W}' \Sigma^{-1/2} \mathbf{Y} = \\ &= \frac{1}{\sqrt{\pi_j}} \mathbf{w}'_j \mathbf{w}_j \pi_j \mathbf{w}'_j \Sigma^{-1/2} \mathbf{Y} = \sqrt{\pi_j} \mathbf{w}'_j \Sigma^{-1/2} \mathbf{Y}\end{aligned}$$

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

$$\hat{\mathbf{b}}_i = E(\mathbf{b}_i / \mathbf{Y}_i) = \mathbf{D}^* \mathbf{Z}'_i \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})$$

- **Sentencias para calcular los residuos con mínima confusión.**

```
options nodate pageno=1 ls=80 ps=59 nonumber; run;
```

```
data infeccion2; set final.uti4hln;
array q(3) ln_pct0 ln_pct24 ln_pct48;
do hora= 1 to 3;
ln_pct=q(hora);
output;
end;
drop ln_pct0 ln_pct24 ln_pct48;
run;
```

```
data infeccion; set final.uti4ln;
if hora=0 then newhora=1;
if hora=24 then newhora=2;
if hora=48 then newhora=3;
run;
proc sort data=infeccion; by diagcod2 paciente; run;
proc print data=infeccion;run;
```

```
data infeccion; set final.uti4ln;
time=hora;
run;
```

```
proc mixed data=infeccion method=ml;
ods output solution R=efal;
ods output v=vmatriz;
ods output g=gmatriz;
class paciente diagcod2;
model ln_pct = diagcod2 hora hora2 / noint solution outp=residcond;
```

```

random intercept / type=un subject=paciente solution g v;
repeated / type=ar(1)subject=paciente r;
run;

proc iml;
use vmatriz;
read all var {col1 col2 col3} into vmatriz;
close vmatriz;
V=i(57)@ vmatriz;

R={1 0.8028 0.64448784,0.8028 1 0.8028,0.64448784 0.8028 1};
sig2=1.2418;
Rt=sig2*R;
verrores=i(57)@ Rt;
E=i(57)@ R;

use infeccion;
read all var {lnpct} into y;
close infeccion;

t={0,24,48};
t2={0,576,2304};
x1=j(3,1,1)||j(3,3,0)||t||t2;
x2=j(3,1,0)||j(3,1,1)||j(3,2,0)||t||t2;
x3=j(3,2,0)||j(3,1,1)||j(3,1,0)||t||t2;
x4=j(3,3,0)||j(3,1,1)||t||t2;
do i=1 to 17;
x=x//x1;
end;
do i=1 to 6;
x=x//x2;
end;
do i=1 to 16;
x=x//x3;
end;
do i=1 to 18;
x=x//x4;
end;

M=sig2*inv(V);
T=x*inv(t(x)*M*x)*t(x)*M;

Q=M*(I(3*57)-T);

Eroot=root(E);
EQE=Eroot*Q*Eroot;

call svd(left,mid,right,EQE);
reset fuzz;
zero=ssq(EQE - left*diag(mid)*t(right));

```

```

print zero;

mid2=mid[1:165,1];
delta=diag(mid2);
raiz=mid2##(1/2);
pi=diag(raiz);

k=left[1:171,1:165];

resleast=pi*t(K)*inv(Eroot)*y/sqrt(sig2);
print resleast;
create resleast from resleast;
append from resleast;
print resleast;

quit;

```

- **Sentencias para realizar los gráficos probabilístico normal con bandas de confianza simuladas.**

```

rcond<-read.table("C:/Users/mara/Desktop/rcondstudlnpct.txt",header=T,dec=",")
attach(rcond)
least<-read.table("C:/Users/mara/Desktop/rleaststudlnpct.txt",header=T,dec=",")
attach(least)
reblup<-read.table("C:/Users/mara/Desktop/ebluplnpct.txt",header=T,dec=",")
attach(reblup)

library(envelope)
library(lattice)
trellis.device(color=F)

envl.plot(rcond$rcondstud, conf = 95, plot.it = TRUE, mc = FALSE, k = 10000)
envl.plot(least$rleaststud, conf = 95, plot.it = TRUE, mc = FALSE, k = 10000)
envl.plot(reblup$eblup, conf = 95, plot.it = TRUE, mc = FALSE, k = 10000)

```

- **Matrices  $\mathbf{L}'$  y  $\boldsymbol{\beta}$  para escribir las hipótesis.**

$$H_0) \begin{cases} \beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} \\ \beta_{21} = \beta_{22} = \beta_{23} = \beta_{24} \end{cases}$$

$$\mathbf{L}' = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ \beta_{04} \\ \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{14} \\ \beta_{21} \\ \beta_{22} \\ \beta_{23} \\ \beta_{24} \end{bmatrix}$$

$$H_0) \beta_{01} = \beta_{02} = \beta_{03} = \beta_{04}$$

$$\mathbf{L}' = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ \beta_{04} \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

Escribir las hipótesis usando la descomposición  $\mathbf{L}'\boldsymbol{\beta}$  es especialmente útil cuando se quiere probarlas usando el test de Wald con el programa SAS.



Tabla I: Predicción de los efectos aleatorios de cada paciente

Paciente	Diagnóstico	Predicción
1	1	-0,1062
2	1	-0,03547
3	1	-0,4198
4	1	0,2057
5	1	0,2854
6	1	-1,0324
7	1	0,1573
8	1	0,06735
9	1	0,1309
10	1	0,07444
11	1	0,9328
12	1	0,1192
13	1	0,09247
14	1	0,7881
15	1	-0,2624
16	1	-0,9411
17	1	-0,05628
18	2	-0,2383
19	2	-0,3364
20	2	0,1325
21	2	1,0196
22	2	-0,1357
23	2	-0,4416
24	3	-0,0823
25	3	0,4243
26	3	0,7083
27	3	0,2775
28	3	-0,4303
29	3	0,3861
30	3	0,1908
31	3	-0,08349
32	3	-0,2988
33	3	-0,9841
34	3	-0,2826
35	3	-0,22
36	3	0,1561
37	3	-0,6498
38	3	0,5104
39	3	0,378
40	4	1,1792
41	4	-0,4005

42	4	0,436
43	4	-0,7443
44	4	1,0056
45	4	-0,3189
46	4	-1,2437
47	4	-0,6423
48	4	-0,7896
49	4	-0,04151
50	4	-0,1156
51	4	0,6485
52	4	-0,8843
53	4	1,007
54	4	0,7755
55	4	-0,4504
56	4	-0,3963
57	4	0,9757

Tabla II: Residuos marginales y condicionales estudentizados.

Obs	Paciente	Diagnóstico	Hora	R. marg.	R. cond.
1	1	1	0	-0,43483	-0,56253
2	1	1	24	-0,27501	-0,31119
3	1	1	48	0,09359	0,26846
4	2	1	0	0,29229	0,50015
5	2	1	24	-0,07443	-0,07654
6	2	1	48	-0,40968	-0,60375
7	3	1	0	-0,52035	-0,33901
8	3	1	24	-0,65699	-0,55388
9	3	1	48	-0,91295	-0,95641
10	4	1	0	0,14494	-0,00696
11	4	1	24	0,88692	1,15987
12	4	1	48	0,44611	0,46665
13	5	1	0	0,43933	0,36501
14	5	1	24	0,21157	0,00684
15	5	1	48	0,58158	0,5887
16	6	1	0	-1,73577	-1,55082
17	6	1	24	-2,01604	-1,99157
18	6	1	48	-1,71062	-1,51126
19	7	1	0	0,47476	0,56699
20	7	1	24	0,24458	0,20501
21	7	1	48	0,06269	-0,08102
22	8	1	0	-0,19632	-0,38562

23	8	1	24	0,40337	0,55743
24	8	1	48	0,36754	0,50108
25	9	1	0	0,24642	0,23805
26	9	1	24	-0,07443	-0,2665
27	9	1	48	0,25561	0,25251
28	10	1	0	-0,56721	-0,97698
29	10	1	24	0,08256	0,04484
30	10	1	48	0,82809	1,21724
31	11	1	0	1,48704	1,27345
32	11	1	24	2,01719	2,10715
33	11	1	48	1,58812	1,43241
34	12	1	0	0,42713	0,53554
35	12	1	24	0,10246	0,02497
36	12	1	48	-0,00336	-0,14144
37	13	1	0	-0,9381	-1,58083
38	13	1	24	0,29144	0,35274
39	13	1	48	1,22491	1,8207
40	14	1	0	1,49557	1,45208
41	14	1	24	1,2827	1,11733
42	14	1	48	1,18566	0,96472
43	15	1	0	-0,47633	-0,44941
44	15	1	24	-0,46221	-0,4272
45	15	1	48	-0,40968	-0,34459
46	16	1	0	-1,52689	-1,32663
47	16	1	24	-1,35075	-1,04963
48	16	1	48	-1,71062	-1,61555
49	17	1	0	-0,06394	-0,03629
50	17	1	24	-0,15371	-0,17747
51	17	1	48	-0,11528	-0,11703
52	18	2	0	-0,50219	-0,49732
53	18	2	24	-0,41429	-0,36057
54	18	2	48	-0,35214	-0,26387
55	19	2	0	-0,39514	-0,21386
56	19	2	24	-0,61177	-0,55089
57	19	2	48	-0,8056	-0,85245
58	20	2	0	0,6	0,77565
59	20	2	24	0,70073	0,93237
60	20	2	48	-0,21794	-0,49692
61	21	2	0	1,76831	1,53619
62	21	2	24	1,94097	1,80481
63	21	2	48	1,85347	1,66867
64	22	2	0	-0,6731	-0,88552
65	22	2	24	0,10491	0,32491
66	22	2	48	0,11949	0,34759

67	23	2	0	-0,54148	-0,31622
68	23	2	24	-0,94519	-0,94431
69	23	2	48	-1,00655	-1,03978
70	24	3	0	-0,03225	0,04341
71	24	3	24	-0,15288	-0,14623
72	24	3	48	-0,24457	-0,29038
73	25	3	0	1,07143	1,19926
74	25	3	24	0,67696	0,57914
75	25	3	48	0,3774	0,10821
76	26	3	0	1,35191	1,31546
77	26	3	24	1,10849	0,93279
78	26	3	48	1,07105	0,87394
79	27	3	0	0,10377	-0,15414
80	27	3	24	0,89393	1,08804
81	27	3	48	0,75488	0,86945
82	28	3	0	-1,45727	-1,79894
83	28	3	24	-1,55381	-1,95071
84	28	3	48	0,15886	0,74172
85	29	3	0	1,02867	1,17571
86	29	3	24	0,65003	0,58047
87	29	3	48	0,28302	0,0035
88	30	3	0	-0,79944	-1,47494
89	30	3	24	1,45658	2,07165
90	30	3	48	1,22392	1,7059
91	31	3	0	0,15495	0,33905
92	31	3	24	-0,23746	-0,27783
93	31	3	48	-0,41952	-0,56405
94	32	3	0	-0,13299	0,13257
95	32	3	24	-0,63275	-0,65307
96	32	3	48	-0,85666	-1,00508
97	33	3	0	-1,64128	-1,45502
98	33	3	24	-1,92204	-1,89638
99	33	3	48	-1,64992	-1,4686
100	34	3	0	-0,23708	-0,04961
101	34	3	24	-0,44561	-0,37744
102	34	3	48	-0,72892	-0,82283
103	35	3	0	-0,52561	-0,57479
104	35	3	24	-0,15143	0,01346
105	35	3	48	-0,26492	-0,16496
106	36	3	0	0,61587	0,78972
107	36	3	24	0,28202	0,26488
108	36	3	48	-0,08933	-0,3189
109	37	3	0	-1,51978	-1,64616
110	37	3	24	-1,08019	-0,9551

111	37	3	48	-0,69088	-0,34309
112	38	3	0	1,05727	1,07855
113	38	3	24	0,90351	0,83682
114	38	3	48	0,66803	0,46664
115	39	3	0	0,98414	1,11494
116	39	3	24	0,6379	0,57063
117	39	3	48	0,29982	0,03915
118	40	4	0	1,84859	1,56321
119	40	4	24	1,86269	1,5854
120	40	4	48	2,16765	2,06511
121	41	4	0	-0,68391	-0,61907
122	41	4	24	-0,50423	-0,33643
123	41	4	48	-0,70558	-0,65317
124	42	4	0	0,50216	0,29272
125	42	4	24	-0,94841	-1,98909
126	42	4	48	1,30562	1,55659
127	43	4	0	-0,82859	-0,45462
128	43	4	24	-1,34368	-1,26487
129	43	4	48	-1,6734	-1,78354
130	44	4	0	1,84859	1,76116
131	44	4	24	1,62454	1,40872
132	44	4	48	1,56934	1,32188
133	45	4	0	-0,88929	-1,03526
134	45	4	24	-0,66033	-0,67509
135	45	4	48	-0,16589	0,10268
136	46	4	0	-1,8483	-1,48912
137	46	4	24	-2,3987	-2,35492
138	46	4	48	-2,3023	-2,20328
139	47	4	0	-0,9731	-0,79823
140	47	4	24	-1,38086	-1,43966
141	47	4	48	-1,14246	-1,06465
142	48	4	0	-0,97666	-0,63587
143	48	4	24	-1,29689	-1,1396
144	48	4	48	-1,70299	-1,77842
145	49	4	0	-0,11535	-0,13412
146	49	4	24	0,03527	0,10281
147	49	4	48	-0,0459	-0,02487
148	50	4	0	-0,5454	-0,72606
149	50	4	24	-0,12448	-0,06393
150	50	4	48	0,14003	0,35216
151	51	4	0	1,53069	1,66834
152	51	4	24	1,11496	1,01437
153	51	4	48	0,66015	0,29893
154	52	4	0	-1,5046	-1,35839

155	52	4	24	-1,16283	-0,82076
156	52	4	48	-1,55341	-1,43516
157	53	4	0	1,84794	1,75849
158	53	4	24	1,82497	1,72237
159	53	4	48	1,53581	1,2675
160	54	4	0	1,59538	1,62522
161	54	4	24	1,3957	1,31111
162	54	4	48	1,01238	0,70814
163	55	4	0	-0,73324	-0,63984
164	55	4	24	-0,51571	-0,29764
165	55	4	48	-0,83928	-0,80664
166	56	4	0	-0,73837	-0,70956
167	56	4	24	-0,73723	-0,70777
168	56	4	48	-0,58948	-0,47535
169	57	4	0	1,84859	1,79524
170	57	4	24	1,59095	1,38996
171	57	4	48	1,46486	1,19161