



Universidad de Valparaíso

Departamento de Estadística

Aplicación de un Modelo Jerárquico
con
Parámetros Aleatorios.

Trabajo de titulación presentada por
Katherine Beatriz Rojas Guerrero
para optar al título de Ingeniero Estadístico.

Profesora Guía: Dra. Claudia Navarro Villarroel.

Valparaíso, 2013.

Índice general

1. Modelo Jerárquico.	5
1.1. Construcción de un modelo jerárquico.	5
1.1.1. Modelo jerárquico matricial	8
1.2. Estimación de los Parámetros	16
1.3. Residuos	23
2. Diagnóstico	24
2.1. Detección de medidas influyentes para el primer nivel.	24
2.1.1. Detección y diagnóstico de medidas influyentes para el segundo nivel.	40
3. Aplicación	44
3.1. Análisis de datos	47
3.1.1. Resultados	48
3.2. Diagnóstico.	56
3.3. Conclusión	63
Bibliografía	64

Dedicado a

*... mi familia y
a todos quienes me apoyaron durante mi carrera.*
Muchas Gracias.

Introducción

Este trabajo de titulación se desarrolla en el área de modelos lineales jerárquicos.

Este trabajo dará a conocer de forma teórica, y luego de manera práctica, en el área de la educación chilena, una técnica estadística para el análisis en un modelo jerárquico con parámetros aleatorios, llamado detección de medidas influyentes.

El modelo jerárquico con coeficientes aleatorios se basa principalmente en que los parámetros asociados a las variables regresoras tiene, a su vez, asociado un error aleatorio, lo cual implica que éstos coeficientes se transformen en coeficientes aleatorios.

Los modelos jerárquicos con coeficientes aleatorios son utilizados, en gran parte, por experimentos y estudios observacionales (Laird, et al., 1982), en estudios educacionales (Goldstein, 1995), en estudios basados en el área de la epidemiología (De la Cruz, 2008) y en el área de la psicología (Oliver, et al., 2000).

Estos modelos jerárquicos permiten solucionar dos tipos de problemas que se presentan cuando se utilizan erróneamente los modelos de un único nivel en datos que son jerárquicos (Sánchez y Ocaña, 1999):

- (i) Problemas de correlación entre los individuos, estimación de los mínimos cuadrados ordinarios ineficientes y significancias menos precisas.
- (ii) Problemas conceptuales, ya que se emplea el nivel equivocado (analizar los datos a un nivel y extraer conclusiones de otro).

Además, resuelve otros problemas como: la determinación del efecto directo de las variables explicativas individuales y de grupo, permite determinar si las variables explicativas del grupo “moderan” las relaciones a nivel individual (interacción entre niveles) y permite también, determinar qué porcentaje de la variabilidad de la variable explicada o dependiente, una vez controlada por las variables explicativas, no causa efecto al individuo y qué porcentaje afecta al grupo (Aparicio y Morera, 2007). A pesar de esto, en algunas ocasiones, la existencia de medidas influyentes puede conducir a conclusiones erróneas. Para superar, de alguna manera, esta situación existen algunas técnicas para la detección y manejo de estas medidas.

Las técnicas para diagnosticar errores de medición son comúnmente usadas en modelos de regresión lineal ordinaria. Buscan encontrar la potencial influencia en el Error Cuadrático Medio (EMC), la estimación de la variable dependiente o respuesta, R^2 (Coeficiente de determinación del modelo), entre otros. Para eso, las pruebas de diagnóstico sirven para detectar las observaciones inusuales y deben ser parte fundamental en el análisis (Galea, 1990).

Para el diagnóstico en modelos jerárquicos, Goldstein (1995), propuso un modelo de componentes de varianza simple, el cual se basa en las estimaciones de las varianzas de los errores formando un coeficiente de correlación intraclase ($\hat{\rho}$), donde basta con que $\hat{\rho} > 0,1$, para justificar el uso del modelo jerárquico. Esto significa, que se encuentran en función de las varianzas y covarianzas estimadas de los errores en las variables, y los valores de hasta 0,08 para el análisis, han sido interesantes dependiendo del área donde se este trabajando para el análisis (Ojeda, et al., 1996).

Existen algunas técnicas que son más conocidas para detectar anomalías en regresión lineal ordinaria, como por ejemplo, la influencia en el error cuadrático medio, el cual utiliza la distancia de Cook como medida de diagnóstico y está encargada de extraer el i -ésimo caso y observar su efecto, entre otras.

La implementación de análisis en modelos jerárquicos con parámetros aleatorios, implementado en algunos software estadísticos ha producido que el tiempo de ejecución del análisis sea reducido. Entre los programas se destacan SAS, S-Plus, Stata y MLwiN. El programa llamado MLwiN realizado por Rasbash, et al. en 2004 y es un programa especial que trabaja exclusivamente con modelos jerárquicos. Sin embargo, los programas estadísticos como Stata, S-Plus y SAS tienen incorporado un módulo aparte para realizar análisis basados en modelos jerárquicos.

La construcción de modelos jerárquicos con parámetros aleatorios se mostrará en el siguiente capítulo, luego una medida de diagnóstico propia en estos tipos de modelos y por último una aplicación con datos reales chilenos, obtenidos a través de la prueba SIMCE. El SIMCE es el Sistema Nacional de Evaluación de resultados de aprendizaje del Ministerio de Educación de Chile. Las pruebas SIMCE evalúan el logro de los objetivos fundamentales y contenidos mínimos obligatorios (OF-CMO) del marco curricular vigente en diferentes sectores de aprendizaje, a través de una medición que se aplica a nivel nacional, una vez al año, a los estudiantes que cursan un determinado nivel educacional. Estos datos fueron adquiridos de la página www.simce.cl, y se trabajará sólo con los colegios que posean octavo básico en todas las regiones de Chile.

Capítulo 1

Modelo Jerárquico.

Los modelos jerárquicos son una clase de modelos que se estructura a varios niveles. Este capítulo mostrará la construcción de un modelo jerárquico en dos niveles con parámetros aleatorios y sus respectivos supuestos.

1.1. Construcción de un modelo jerárquico.

Sea y la variable respuesta de interés. Esta variable respuesta puede ser explicada en dos niveles. En el primer nivel se explicará por m covariables fijas. En el segundo nivel la variable respuesta se explicará por q covariables aleatorias. Lo anterior da paso a la creación de un modelo jerárquico de dos niveles. Las covariables del primer nivel se denotarán por x_1, \dots, x_m y las covariables del segundo nivel se denotarán por w_1, \dots, w_q .

Para mayor comprensión, se realizará paso a paso la construcción del modelo jerárquico, comenzando por el primer nivel.

Modelo en el primer nivel.

Definición 1.1. El modelo para el primer nivel, con una variable explicatoria a nivel 1, x_{1ij} , tiene la forma:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + e_{ij} \quad , \quad \forall i = 1, 2, \dots, n_j \quad \text{y} \quad \forall j = 1, 2, \dots, J, \quad (1.1)$$

y se caracteriza por,

$$\begin{aligned} E(e_{ij}) &= 0 \quad \text{y} \\ \text{Var}(e_{ij}) &= \sigma_e^2 \quad . \end{aligned}$$

Modelo en el Segundo nivel.

Definición 1.2. El modelo para el segundo nivel, con una variable explicatoria a nivel 2, w_{1j} , tiene la siguiente estructura:

$$\begin{aligned}\beta_{0j} &= \beta_{00} + \beta_{01}w_{1j} + u_{0j} \quad \text{y} \\ \beta_{1j} &= \beta_{10} + \beta_{11}w_{1j} + u_{1j} \quad ;\end{aligned}\tag{1.2}$$

donde,

$$\begin{aligned}E(u_{0j}) &= 0 \quad \text{y} \\ E(u_{1j}) &= 0 \quad .\end{aligned}$$

Y además,

$$\begin{aligned}Var(u_{0j}) &= \sigma_{u0}^2 \quad , \\ Var(u_{1j}) &= \sigma_{u1}^2 \quad \text{y} \\ cov(u_{0j}, u_{1j}) &= 0 \quad .\end{aligned}$$

El modelo será combinado en la j -ésima unidad a nivel 2, del modelo nivel 1 con una variable explicatoria a nivel 1; y del modelo nivel 2 con una variable explicatoria a nivel 2, queda:

$$y_{ij} = (\beta_{00} + \beta_{01}w_{1j} + u_{0j}) + (\beta_{10}x_{1ij} + \beta_{11}w_{1j}x_{1ij} + u_{1j}x_{1ij}) + e_{ij} \quad .\tag{1.3}$$

Reestructurando los términos anteriores se tiene:

$$y_{ij} = \beta_{00} + \beta_{10}x_{1ij} + \beta_{01}w_{1j} + u_{0j} + \beta_{11}w_{1j}x_{1ij} + u_{1j}x_{1ij} + e_{ij} \quad .\tag{1.4}$$

Para incluir más variables en el primer y segundo nivel, se considerará lo siguiente:

En el primer nivel considere m variables independientes de la forma:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \cdots + \beta_{mj}x_{mij} + e_{ij} \quad , \forall i = 1, 2, \dots, n_j \forall j = 1, 2, \dots, J \quad ,\tag{1.5}$$

donde, $E(e_{ij}) = 0$ y $Var(e_{ij}) = \sigma_e^2$.

En el segundo nivel considere q variables independientes de la forma:

$$\begin{aligned}
\beta_{0j} &= \beta_{00} + \beta_{01}w_{1j} + \beta_{02}w_{2j} + \cdots + \beta_{0q}w_{qj} + u_{0j} \\
\beta_{1j} &= \beta_{10} + \beta_{11}w_{1j} + \beta_{12}w_{2j} + \cdots + \beta_{1q}w_{qj} + u_{1j} \\
&\vdots \\
\beta_{mj} &= \beta_{m0} + \beta_{m1}w_{1j} + \beta_{m2}w_{2j} + \cdots + \beta_{mq}w_{qj} + u_{mj} \quad .
\end{aligned} \tag{1.6}$$

El modelo combinado para la j -ésima unidad en el nivel 2, del modelo nivel 1 con m variables independientes en el primer nivel y q variables independientes para el segundo nivel, toma la siguiente forma:

$$\begin{aligned}
y_{ij} &= \beta_{00} + \beta_{01}w_{1j} + \beta_{02}w_{2j} + \cdots + \beta_{0q}w_{qj} + u_{0j} \\
&\quad + \beta_{10}x_{1ij} + \beta_{11}w_{1j}x_{1ij} + \cdots + \beta_{1q}w_{qj}x_{1ij} + u_{1j}x_{1ij} \\
&\quad + \cdots + \beta_{m0}x_{mij} + \cdots + \beta_{mq}w_{qj}x_{mij} + u_{mj}x_{mij} + e_{ij} \quad .
\end{aligned} \tag{1.7}$$

Al reordenar los términos de la ecuación (1.7) se obtiene el siguiente modelo:

$$\begin{aligned}
y_{ij} &= \beta_{00} + \beta_{10}x_{1ij} + \cdots + \beta_{m0}x_{mij} \\
&\quad + \beta_{01}w_{1j} + \beta_{11}w_{1j}x_{1ij} + \cdots + \beta_{m1}w_{1j}x_{mij} \\
&\quad + \cdots + \beta_{0q}w_{qj} + \beta_{1q}w_{qj}x_{1ij} + \cdots \\
&\quad + u_{0j} + u_{1j}x_{1ij} + \cdots + u_{mj}x_{mij} + e_{ij} \quad .
\end{aligned} \tag{1.8}$$

1.1.1. Modelo jerárquico matricial

Para tener una mayor simplicidad al momento de realizar cálculos, es necesario transformar la ecuación (1.8) en un modelo jerárquico de dos niveles vista desde la perspectiva matricial. Así, el modelo propuesto bajo el primer nivel dado en la ecuación (1.5), se escribirá de la siguiente manera:

Sea \mathbf{Y}_j un vector de respuestas $n_j \times 1$, \mathbf{X}_j una matriz de diseño del primer nivel de orden $n_j \times (m + 1)$, β_j un vector de parámetros de orden $(m + 1) \times 1$ y \mathbf{e}_j un vector de errores aleatorios de orden $n_j \times 1$. Los supuestos para este modelo son $E(\mathbf{e}_j) = 0$ y $Var(\mathbf{e}_j) = \sigma_e^2 \mathbf{I}_{n_j}$.

Entonces,

$$\mathbf{Y}_j = \mathbf{X}_j \beta_j + \mathbf{e}_j \quad , \text{ para } j = 1, \dots, J \quad . \quad (1.9)$$

Donde,

$$\mathbf{Y}_j = \begin{bmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{n_j j} \end{bmatrix} ; \quad \mathbf{X}_j = \begin{bmatrix} 1 & x_{11j} & x_{21j} & \dots & x_{m1j} \\ 1 & x_{12j} & x_{22j} & \dots & x_{m2j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n_j j} & x_{2n_j j} & \dots & x_{mn_j j} \end{bmatrix} ;$$

$$\mathbf{e}_j = \begin{bmatrix} e_{1j} \\ e_{2j} \\ \vdots \\ e_{n_j j} \end{bmatrix} \quad \text{y} \quad \beta_j = \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \\ \vdots \\ \beta_{mj} \end{bmatrix} .$$

La forma matricial para el segundo nivel, obtenida desde la ecuación (1.6) es la siguiente: Sea \mathbf{W}_j la matriz de diseño del segundo nivel, de orden $(m + 1) \times (q + 1)(m + 1)$, β es el vector de coeficientes fijos de orden $(q + 1)(m + 1) \times 1$ y \mathbf{u}_j el vector de errores aleatorios de nivel 2 de orden $(m + 1) \times 1$. Los supuestos para el segundo nivel son $E(\mathbf{u}_j) = 0$ y

$$Var(\mathbf{u}_j) = \Omega = \begin{bmatrix} \sigma_{u0}^2 & \sigma_{u01} & \dots & \sigma_{u0m} \\ \sigma_{u01} & \sigma_{u1}^2 & \dots & \sigma_{u1m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u0m} & \sigma_{u1m} & \dots & \sigma_{um}^2 \end{bmatrix} .$$

Así,

$$\beta_j = \mathbf{W}_j \beta + \mathbf{u}_j \quad , \quad \forall j = 1, \dots, J \quad ; \quad (1.10)$$

donde,

$$\mathbf{W}_j = \begin{bmatrix} 1 & w_{1j} & w_{2j} & \cdots & w_{qj} & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & \ddots & 0 & \ddots & 1 & w_{1j} & w_{2j} & \cdots & w_{qj} & \cdots & 0 & 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & w_{1j} & w_{2j} & \cdots & w_{qj} \end{bmatrix} \quad ,$$

$$\beta = [\beta_{00}, \beta_{01}, \cdots, \beta_{0q}, \beta_{10}, \beta_{11}, \cdots, \beta_{1q}, \cdots, \beta_{m0}, \beta_{m1}, \cdots, \beta_{mq}]^\top \quad ;$$

$$\mathbf{u}_j = [u_{0j}, u_{1j}, \cdots, u_{mj}]^\top \quad .$$

Entonces el modelo, al combinarlo por la j -ésima unidad perteneciente al segundo nivel, toma la siguiente forma:

$$\begin{aligned} \mathbf{Y}_j &= \mathbf{X}_j \mathbf{W}_j \beta + \mathbf{X}_j \mathbf{u}_j + \mathbf{e}_j \quad ; \quad j = 1, \dots, J, \\ E(\mathbf{Y}_j) &= \mathbf{X}_j \mathbf{W}_j \beta, \quad \mathbf{V}_j = \text{Var}(\mathbf{Y}_j) = \mathbf{X}_j \boldsymbol{\Omega} \mathbf{X}_j^\top + \sigma_e^2 \mathbf{I}_{n_j} \quad . \end{aligned} \quad (1.11)$$

El modelo lineal general jerárquico en la forma matricial, será expresado de la siguiente manera:

$$\mathbf{Y} = \mathbf{XW}\beta + \mathbf{Xu} + \mathbf{e} \quad , \quad (1.12)$$

donde,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_J \end{bmatrix} \quad , \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_J \end{bmatrix} \quad , \quad \mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_J \end{bmatrix} \quad , \quad \mathbf{X} = \text{diag}(\mathbf{X}_j) \quad \text{y} \quad \mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_J \end{bmatrix} \quad .$$

Además, $\text{diag}(\mathbf{A}_j)$ es una matriz diagonal en bloque (de aquí en adelante llamada matriz-bloque), con \mathbf{A}_j en el j -ésimo bloque de la diagonal.

La matriz de varianzas y covarianzas es denotada por:

$$\begin{aligned}
\mathbf{V} &= \text{Var}(\mathbf{Y}) \\
&= \mathbf{X}\text{Var}(\mathbf{u})\mathbf{X}^t + \text{Var}(\mathbf{e}) \\
&= \mathbf{X}\text{diag}(\Omega)\mathbf{X}^t + \text{diag}(\sigma_e^2\mathbf{I}_{n_j}) \\
&= \mathbf{X}\mathbf{G}\mathbf{X}^t + \mathbf{R} \quad .
\end{aligned} \tag{1.13}$$

DEMOSTRACIÓN. (Goldstein, 1995)

Considere el modelo (1.12) de la siguiente manera:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}^{(2)}\mathbf{u} + \mathbf{Z}^{(1)}\mathbf{e} \quad ; \tag{1.14}$$

donde,

- \mathbf{Y} es el vector columna de las observaciones de la i -ésima unidad del primer nivel en la j -ésima unidad del segundo nivel.
- $\mathbf{X}\beta$ es la parte fija del modelo multinivel general de dos niveles.
- $\mathbf{Z}^{(2)} = (z_{0ij}, z_{1ij}, \dots, z_{qij})^\top$ es la matriz de diseño de las $h = 1, \dots, q$ variables aleatorias para el segundo nivel.
- $\mathbf{Z}^{(1)} = (z_{0ij}, z_{1ij}, \dots, z_{mij})^\top$ es la matriz de diseño de las $h = 1, \dots, m$ variables aleatorias para el primer nivel.
- $\mathbf{u} = (u_{0j}, u_{1j}, \dots, u_{qj})^\top$ es el vector columna con $h = 0, 1, \dots, q$ de parámetros aleatorios para el segundo nivel.
- $\mathbf{e} = (e_{0ij}, e_{1ij}, \dots, e_{mij})^\top$ es el vector columna con $h = 0, 1, \dots, m$ de parámetros aleatorios para el primer nivel.

En general,

$$\mathbf{Z}_{0ij} = x_{0ij} = 1 \quad .$$

Así, la parte aleatoria del modelo supone que los coeficientes aleatorios, para los niveles 1 y 2, tienen media cero y varianza constante.

El valor esperado del modelo es,

$$\begin{aligned} E(\mathbf{Y}) &= \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij} \quad , \\ &= \mathbf{X}\beta \quad . \end{aligned} \tag{1.15}$$

La construcción de la matriz de varianzas y covarianzas, denotada por \mathbf{V} , está dada por:

$$\mathbf{V} = E \left[(\mathbf{Y} - E(\mathbf{Y}))(\mathbf{Y} - E(\mathbf{Y}))^\top \right] \quad . \tag{1.16}$$

Luego, sustituyendo las ecuaciones (1.14) y (1.15) en la expresión (1.16) se tiene que,

$$\mathbf{V} = E \left[\left(\mathbf{Z}^{(2)}\mathbf{u} + \mathbf{Z}^{(1)}\mathbf{e} \right) \left(\mathbf{Z}^{(2)}\mathbf{u} + \mathbf{Z}^{(1)}\mathbf{e} \right)^\top \right] \quad .$$

Sea $\mathbf{Z}^{(1)}\mathbf{e} = \mathbf{e}_1$ y $\mathbf{Z}^{(2)}\mathbf{e} = \mathbf{e}_2$, entonces,

$$\begin{aligned} \mathbf{V} &= E \left[\left(\mathbf{e}_1 + \mathbf{e}_2 \right) \left(\mathbf{e}_1 + \mathbf{e}_2 \right)^\top \right] \\ &= E \left[\left(\mathbf{e}_1 + \mathbf{e}_2 \right) \left(\mathbf{e}_1^\top + \mathbf{e}_2^\top \right) \right] \\ &= E \left[\left(\mathbf{e}_1\mathbf{e}_1^\top \right) + \left(\mathbf{e}_1\mathbf{e}_2^\top \right) + \left(\mathbf{e}_2\mathbf{e}_1^\top \right) + \left(\mathbf{e}_2\mathbf{e}_2^\top \right) \right] \\ &= E \left[\left(\mathbf{e}_1\mathbf{e}_1^\top \right) \right] + E \left[\left(\mathbf{e}_1\mathbf{e}_2^\top \right) \right] + E \left[\left(\mathbf{e}_2\mathbf{e}_1^\top \right) \right] + E \left[\left(\mathbf{e}_2\mathbf{e}_2^\top \right) \right] \\ &= E \left[\left(\mathbf{e}_1\mathbf{e}_1^\top \right) \right] + E \left[\left(\mathbf{e}_2\mathbf{e}_2^\top \right) \right] \quad , \end{aligned}$$

donde $E \left[\left(\mathbf{e}_1\mathbf{e}_1^\top \right) \right]$ representa la varianza para las variables aleatorias del primer nivel y $E \left[\left(\mathbf{e}_2\mathbf{e}_2^\top \right) \right]$ es la varianza de las variables aleatorias del segundo nivel.

Por otro lado,

$$\begin{aligned}
E \left[\left(\mathbf{e}_1 \mathbf{e}_1^\top \right) \right] &= E \left[\left(\mathbf{Z}^{(1)} \mathbf{e} \right) \left(\mathbf{Z}^{(1)} \mathbf{e} \right)^\top \right] \\
&= E \left[\left(\mathbf{Z}^{(1)} \mathbf{e} \right) \left(\mathbf{e}^\top \mathbf{Z}^{(1)\top} \right) \right] \\
&= E \left[\mathbf{Z}^{(1)} \mathbf{e} \mathbf{e}^\top \mathbf{Z}^{(1)\top} \right] \\
&= \mathbf{Z}^{(1)} E \left[\mathbf{e} \mathbf{e}^\top \right] \mathbf{Z}^{(1)\top} \\
&= \mathbf{V}_{(1)j} \quad ,
\end{aligned}$$

donde,

$$E \left[\mathbf{e} \mathbf{e}^\top \right] = E \left[\begin{array}{c} \left(\begin{array}{c} e_{0ij} \\ e_{1ij} \\ \vdots \\ e_{mij} \end{array} \right) \left(e_{0ij} \ , \ e_{1ij} \ , \ \dots \ , \ e_{mij} \right) \end{array} \right]_{(m+1) \times (m+1)} .$$

Considerando los supuestos del modelo jerárquico y utilizando las propiedades de esperanza matemática, se tiene:

$$E \left[\mathbf{e} \mathbf{e}^\top \right] = E \left[\begin{array}{cccccccccc} \sigma_{e011}^2 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & \\ 0 & \sigma_{e021}^2 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & \sigma_{e0k_1}^2 & \dots & 0 & 0 & \dots & 0 & \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & 0 & \dots & \sigma_{em1n}^2 & 0 & \dots & 0 & \\ 0 & 0 & \dots & 0 & \dots & 0 & \sigma_{em2n}^2 & \dots & 0 & \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & \sigma_{emk_n}^2 & \end{array} \right]_{(k_1+\dots+k_n) \times (k_1+\dots+k_n)} ,$$

$$\begin{aligned}
E \left[\mathbf{e} \mathbf{e}^\top \right] &= \mathbf{\Omega}_e \\
&= cov(\mathbf{e}_h^{(1)}) \quad .
\end{aligned}$$

La matriz $\mathbf{\Omega}_e$ es una matriz bloque-diagonal de orden $(m+1) \times (m+1)$ donde cada bloque contiene la varianza del bloque h , en el i -ésimo individuo del j -ésimo nivel.

Si,

$$\mathbf{Z}^{(1)} = \begin{bmatrix} 1 & x_{111} & x_{211} & \cdots & x_{m11} \\ 1 & x_{121} & x_{221} & \cdots & x_{m21} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1k_11} & x_{2k_11} & \cdots & x_{mk_11} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{11n} & x_{21n} & \cdots & x_{m1n} \\ 1 & x_{12n} & x_{22n} & \cdots & x_{m2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1k_n n} & x_{2k_n n} & \cdots & x_{mk_n n} \end{bmatrix}_{(k_1+\dots+k_n)\times(m+1)}$$

Se tiene que $\mathbf{V}_{(1)j} = \mathbf{Z}^{(1)}\boldsymbol{\Omega}_e\mathbf{Z}^{(1)\top}$, es una matriz de orden $(m+1) \times (m+1)$.

Análogamente,

$$\begin{aligned} E\left[\left(\mathbf{e}_2\mathbf{e}_2^\top\right)\right] &= E\left[\left(\mathbf{Z}^{(2)}\mathbf{u}\right)\left(\mathbf{Z}^{(2)}\mathbf{u}\right)^\top\right] \\ &= E\left[\left(\mathbf{Z}^{(2)}\mathbf{u}\right)\left(\mathbf{u}^\top\mathbf{Z}^{(2)\top}\right)\right] \\ &= E\left[\mathbf{Z}^{(2)}\mathbf{u}\mathbf{u}^\top\mathbf{Z}^{(2)\top}\right] \\ &= \mathbf{Z}^{(2)}E\left[\mathbf{u}\mathbf{u}^\top\right]\mathbf{Z}^{(2)\top} \\ &= \mathbf{V}_{(2)j} \end{aligned}$$

donde,

$$\begin{aligned} E\left[\mathbf{u}\mathbf{u}^\top\right] &= E\left[\begin{pmatrix} u_{0j} \\ u_{1j} \\ \vdots \\ u_{qj} \end{pmatrix} \left(u_{0j}, u_{1j}, \dots, u_{qj}\right)\right]_{(q+1)\times(q+1)} \\ &= E\begin{bmatrix} u_{0j}^2 & u_{0j}u_{1j} & \cdots & u_{0j}u_{qj} \\ u_{0j}u_{1j} & u_{1j}^2 & \cdots & u_{1j}u_{qj} \\ \vdots & \vdots & \ddots & \vdots \\ u_{0j}u_{qj} & u_{1j}u_{qj} & \cdots & u_{qj}^2 \end{bmatrix}_{(q+1)\times(q+1)} \end{aligned}$$

Luego, se obtiene

$$E[\mathbf{uu}^t] = E \begin{bmatrix} \sigma_{u0}^2 & \sigma_{u01} & \cdots & \sigma_{u0q} \\ \sigma_{u01} & \sigma_{u1}^2 & \cdots & \sigma_{u1j} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u0q} & \sigma_{u1q} & \cdots & \sigma_{uq}^2 \end{bmatrix}_{(q+1) \times (q+1)} = \Omega_u = \text{cov}(\mathbf{e}_h^{(2)}) .$$

Por otro lado,

$$\mathbf{Z}^{(2)} = \begin{bmatrix} 1 & x_{111} & x_{211} & \cdots & x_{q11} \\ 1 & x_{121} & x_{221} & \cdots & x_{q21} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1k_11} & x_{2k_11} & \cdots & x_{qk_11} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{11n} & x_{21n} & \cdots & x_{q1n} \\ 1 & x_{12n} & x_{22n} & \cdots & x_{q2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1k_n n} & x_{2k_n n} & \cdots & x_{qk_n n} \end{bmatrix}_{(k_1 + \dots + k_n) \times (q+1)} .$$

Al igual que el término anterior, si se calcula $\mathbf{V}_{(2)j} = \mathbf{Z}^{(2)} \Omega_u \mathbf{Z}^{(2)t}$, se obtiene una matriz bloque-diagonal, y ésta contiene j -ésimos bloques. El bloque j -ésimo de la matriz \mathbf{V}_2 , se puede escribir:

$$\mathbf{V}_{2j} = \mathbf{V}_{(1)j} + \mathbf{V}_{(2)j} . \quad (1.17)$$

Por ejemplo, para los datos del área de educación, considere dos estudiantes en cada $j = 1, \dots, n$ escuelas. Entonces, la matriz de $\mathbf{V}_{(2)j}$ asume la forma:

$$\mathbf{V}_{(2)j} = \left[\begin{array}{c} \left(\begin{array}{ccc} 1 & x_{11j} & x_{21j} \\ 1 & x_{12j} & x_{22j} \end{array} \right) \left(\begin{array}{ccc} \sigma_{u0}^2 & \sigma_{u01} & \sigma_{u02} \\ \sigma_{u01} & \sigma_{u1}^2 & \sigma_{u12} \\ \sigma_{u02} & \sigma_{u12} & \sigma_{u2}^2 \end{array} \right) \left(\begin{array}{cc} 1 & 1 \\ x_{11j} & x_{12j} \\ x_{21j} & x_{22j} \end{array} \right) \end{array} \right]_{j \times j} . \quad (1.18)$$

Mientras que la matriz, $\mathbf{V}_{(1)j}$ es dada por (considerando $h = 0$),

$$\mathbf{V}_{(1)j} = \begin{bmatrix} \sigma_{e01j} & 0 \\ 0 & \sigma_{e02j} \end{bmatrix} . \quad (1.19)$$

Ahora, para obtener \mathbf{V}_{2j} , es necesario sumar las matrices (1.18) y (1.19), esto es:

$$\mathbf{V}_{2j} = \left[\begin{array}{c} \left(\begin{array}{ccc} 1 & x_{11j} & x_{21j} \\ 1 & x_{12j} & x_{22j} \end{array} \right) \left(\begin{array}{ccc} \sigma_{u0}^2 & \sigma_{u01} & \sigma_{u02} \\ \sigma_{u01} & \sigma_{u1}^2 & \sigma_{u12} \\ \sigma_{u02} & \sigma_{u12} & \sigma_{u2}^2 \end{array} \right) \left(\begin{array}{cc} 1 & 1 \\ x_{11j} & x_{12j} \\ x_{21j} & x_{22j} \end{array} \right) \\ + \left[\begin{array}{cc} \sigma_{e01j} & 0 \\ 0 & \sigma_{e02j} \end{array} \right] \end{array} \right] . \quad (1.20)$$

Así, para escribir \mathbf{V}_{2j} de forma general, es necesario utilizar el operador de suma directa, la cual está en términos de matrices por bloques, de esta manera (Goldstein, 1995) :

$$\mathbf{V}_{2j} = \bigoplus_i \sigma_{eij}^2 + \mathbf{V}_{2(2)j} . \quad (1.21)$$

Finalmente, \mathbf{V}_{2j} queda expresada en la forma de la ecuación (1.13).

Parte fundamental del análisis para los modelos jerárquicos con parámetros aleatorios es la estimación de parámetros. A continuación se dará paso a la estimación de parámetros fijos y aleatorios que contiene este tipo de modelos.

1.2. Estimación de los Parámetros

Los modelos jerárquicos con coeficientes aleatorios tienen parámetros aleatorios y parámetros fijos que deben ser estimados. A través del *método de mínimos cuadrados generalizados* es posible estimar los parámetros fijos del modelo. Considere el modelo (1.14) de la siguiente forma (Goldstein, 1995):

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}, \quad \mathbf{e} = \mathbf{Z}^{(s)}e^{(s)} + \dots + \mathbf{Z}^{(1)}e^{(1)}, \quad (1.22)$$

donde,

- \mathbf{Y} es el vector de respuestas de orden $N \times 1$;
- \mathbf{X} es la matriz de diseño de orden $N \times p$;
- β es el vector de parámetros de efectos fijos;
- $e^{(k)}$ son los errores aleatorios, $k = 1, \dots, s$; y,
- $\mathbf{Z}^{(k)}$ es la matriz de diseño asociada a los errores aleatorios, $k = 1, \dots, s$.

Bajo el supuesto de normalidad sobre $e^{(k)}$, $\hat{\beta}$ y $\hat{\theta}$ corresponden a los estimadores de máxima verosimilitud de β y θ respectivamente.

Sea $\mathbf{V} = \text{cov}(\mathbf{e})$ la matriz de covarianza asociada al vector de errores aleatorios \mathbf{e} denotada por $\mathbf{V} = \mathbf{V}(\theta)$, es θ el vector de efectos aleatorios de orden $R \times 1$. Sea $\text{vec}(\mathbf{V}(\theta)) = \mathbf{Z}^*\theta$, donde \mathbf{Z}^* es una matriz sectorizada. Si θ es conocido, entonces el estimador del vector de parámetro β de efectos fijos es:

$$\hat{\beta} = (\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}\mathbf{Y}, \quad (1.23)$$

y la matriz de covarianza es dada por

$$\text{cov}(\hat{\beta}) = (\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}, \quad (1.24)$$

donde $\hat{\mathbf{V}} = \mathbf{V}(\hat{\theta})$.

Considere ahora el siguiente caso:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n_m m} \end{bmatrix} \quad \text{e} \quad \mathbf{Y} = \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n_m m} \end{bmatrix}, \quad (1.25)$$

en este caso tenemos m unidades en el segundo nivel y n_j unidades pertenecientes al primer nivel en la j -ésima unidad del segundo nivel. Cuando los errores tienen una distribución normal asociada en la ecuación (1.23) también se estima mediante el método de *verosimilitud máxima*.

Sea $\mathbf{Y}^* = \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top$, donde $\tilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\beta$. Considere:

$$\begin{aligned} E(\mathbf{Y}^*) &= E\left(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top\right) \\ &= E\left((\mathbf{Y} - \mathbf{X}\beta)(\mathbf{Y} - \mathbf{X}\beta)^\top\right) \\ &= \mathbf{V}. \end{aligned} \quad (1.26)$$

Es posible realizar un arreglo para $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top$ utilizando las propiedades de la función vec , de la siguiente manera:

$$\begin{aligned} \mathbf{Y}^{**} &= \text{vec}\left(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top\right) \\ &= \text{vec}\left((\mathbf{Y} - \mathbf{X}\beta)(\mathbf{Y} - \mathbf{X}\beta)^\top\right) \\ &= \tilde{\mathbf{Y}} \otimes \tilde{\mathbf{Y}}. \end{aligned} \quad (1.27)$$

La función matemática vec , opera como se muestra a continuación:

Sea \mathbf{T} una matriz cuadrada y simétrica de orden 2×2 ,

$$\mathbf{T} = \begin{bmatrix} a & b \\ b & c \end{bmatrix},$$

entonces $\text{vec}(\mathbf{T})$ será:

$$\text{vec}(\mathbf{T}) = \begin{bmatrix} a \\ b \\ b \\ c \end{bmatrix}.$$

Para una mayor comprensión, tome en consideración las siguientes matrices:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \\ 1 & x_{12} \\ 1 & x_{22} \end{bmatrix} \quad \text{e} \quad \mathbf{Y} = \begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{12} \\ y_{22} \end{bmatrix},$$

el vector \mathbf{Y}^* resulta ser,

$$\begin{aligned} \mathbf{Y}^* &= \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top \\ &= \begin{pmatrix} \tilde{y}_{11} \\ \tilde{y}_{21} \\ \tilde{y}_{31} \\ \tilde{y}_{12} \\ \tilde{y}_{22} \end{pmatrix} \begin{pmatrix} \tilde{y}_{11} & \tilde{y}_{21} & \tilde{y}_{31} & \tilde{y}_{12} & \tilde{y}_{22} \end{pmatrix} \\ &= \begin{pmatrix} \tilde{y}_{11}^2 & \tilde{y}_{11}\tilde{y}_{21} & \cdots & \tilde{y}_{11}\tilde{y}_{22} \\ \tilde{y}_{21}\tilde{y}_{11} & \tilde{y}_{21}^2 & \cdots & \tilde{y}_{21}\tilde{y}_{22} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{y}_{22}\tilde{y}_{11} & \tilde{y}_{22}\tilde{y}_{21} & \cdots & \tilde{y}_{22}^2 \end{pmatrix}. \end{aligned}$$

Ahora, se desarrollará \mathbf{Y}^{**} :

$$\begin{aligned}
\mathbf{Y}^{**} &= \text{vec} \left(\tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^t \right) \\
&= \text{vec} \left[\begin{array}{c} \left(\begin{array}{c} \tilde{y}_{11} \\ \tilde{y}_{21} \\ \tilde{y}_{31} \\ \tilde{y}_{12} \\ \tilde{y}_{22} \end{array} \right) \left(\begin{array}{ccccc} \tilde{y}_{11} & \tilde{y}_{21} & \tilde{y}_{31} & \tilde{y}_{12} & \tilde{y}_{22} \end{array} \right) \end{array} \right] \\
&= \text{vec} \left(\begin{array}{cccc} \tilde{y}_{11}^2 & \tilde{y}_{11} \tilde{y}_{21} & \cdots & \tilde{y}_{11} \tilde{y}_{22} \\ \tilde{y}_{21} \tilde{y}_{11} & \tilde{y}_{21}^2 & \cdots & \tilde{y}_{21} \tilde{y}_{22} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{y}_{22} \tilde{y}_{11} & \tilde{y}_{22} \tilde{y}_{21} & \cdots & \tilde{y}_{22}^2 \end{array} \right).
\end{aligned}$$

Aplicando la esperanza matemática a \mathbf{Y}^{**} , así:

$$\begin{aligned}
E(\mathbf{Y}^{**}) &= E \left[\text{vec} \left(\begin{array}{cccc} \tilde{y}_{11}^2 & \tilde{y}_{11} \tilde{y}_{21} & \cdots & \tilde{y}_{11} \tilde{y}_{22} \\ \tilde{y}_{21} \tilde{y}_{11} & \tilde{y}_{21}^2 & \cdots & \tilde{y}_{21} \tilde{y}_{22} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{y}_{22} \tilde{y}_{11} & \tilde{y}_{22} \tilde{y}_{21} & \cdots & \tilde{y}_{22}^2 \end{array} \right) \right] \\
&= E \left(\begin{array}{c} \tilde{y}_{11}^2 \\ \tilde{y}_{21} \tilde{y}_{11} \\ \vdots \\ \tilde{y}_{22}^2 \end{array} \right),
\end{aligned}$$

el desarrollo de la ecuación anterior es:

$$E(\mathbf{Y}^{**}) = \begin{pmatrix} \sigma_{u0}^2 + \sigma_{e0}^2 \\ \sigma_{u0}^2 \\ \vdots \\ \sigma_{u0}^2 \\ \sigma_{u0}^2 + \sigma_{e0}^2 \end{pmatrix}. \tag{1.28}$$

Así, se tiene que la expresión (1.28), y se puede expresar como el siguiente modelo lineal:

$$\mathbf{Y}^{**} = E(\mathbf{Y}^{**}) + \mathbf{R},$$

donde \mathbf{R} es un vector residual. En consecuencia (Goldstein, 1995),

$$\begin{aligned} \begin{pmatrix} \tilde{y}_{11}^2 \\ \tilde{y}_{21}\tilde{y}_{11} \\ \vdots \\ \tilde{y}_{22}^2 \end{pmatrix} &= \begin{pmatrix} \sigma_{u0}^2 + \sigma_{e0}^2 \\ \sigma_{u0}^2 \\ \vdots \\ \sigma_{u0}^2 + \sigma_{e0}^2 \end{pmatrix} + \mathbf{R} \\ &= \sigma_{u0}^2 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \sigma_{e0}^2 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 1 \end{pmatrix} + \mathbf{R} . \end{aligned} \quad (1.29)$$

El lado izquierdo de (1.29) contiene el vector de respuestas del modelo y el lado derecho contiene dos variables explicativas, con coeficientes σ_{u0}^2 y σ_{e0}^2 que deben ser estimados.

Sea $\mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ e $\mathbf{Y}^{**} = \text{vec}(\mathbf{e}\mathbf{e}^t)$. El modelo jerárquico se puede reescribir de la siguiente manera:

$$E(\mathbf{Y}) = \mathbf{Z}^*\boldsymbol{\theta} , \quad (1.30)$$

donde \mathbf{Z}^* es la matriz de diseño asociada a los efectos usando el método de *Mínimos Cuadrados Generalizados* se tiene que:

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{Z}^{*t} \mathbf{V}^{*-1} \mathbf{Z}^* \right)^{-1} \mathbf{Z}^* \mathbf{V}^{*-1} \mathbf{Y}^{**}, \quad \mathbf{V}^* = \mathbf{V} \otimes \mathbf{V} . \quad (1.31)$$

Consecuentemente, la matriz de covarianza de $\hat{\boldsymbol{\theta}}$ está dada por:

$$\text{cov}(\hat{\boldsymbol{\theta}}) = 2 \left(\mathbf{Z}^{*t} \mathbf{V}^{*-1} \mathbf{Z}^* \right)^{-1} . \quad (1.32)$$

DEMOSTRACIÓN. (Goldstein, 1995)

La matriz de covarianzas de $\hat{\theta}$ es:

$$\text{cov}(\hat{\theta}) = \left(\mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*t} \mathbf{V}^{*-1} \text{cov}(\mathbf{Y}^{**}) \mathbf{V}^{*-1} \mathbf{Z}^* (\mathbf{Z}^{*t} \mathbf{V}^{*-1} \mathbf{Z}^*)^{-1} ; \quad (1.33)$$

donde,

$$\mathbf{Y}^{**} = \tilde{\mathbf{Y}} \otimes \tilde{\mathbf{Y}} .$$

Usando el resultado estándar (Searle, et al., 1992), se tiene que:

$$\text{cov}(\tilde{\mathbf{Y}} \otimes \tilde{\mathbf{Y}}) = (\mathbf{V} \otimes \mathbf{V}) (\mathbf{I} + S_N) , \quad (1.34)$$

donde $(\mathbf{V} \otimes \mathbf{V}) = \mathbf{V}^*$ y S_N es matriz de permutación vec. Consecuentemente,

$$\begin{aligned} \mathbf{V}^{*-1} &= (\mathbf{V} \otimes \mathbf{V})^{-1} \\ &= \mathbf{V}^{-1} \otimes \mathbf{V}^{-1}. \end{aligned}$$

Por lo tanto,

$$\mathbf{V}^{*-1} \mathbf{Z}^* = (\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}) \text{vec}(\mathbf{A}) ,$$

donde $\text{vec}(\mathbf{A})$ corresponde al vector operador de la matriz de diseño para los parámetros aleatorios del modelo jerárquico. Luego,

$$\begin{aligned} \mathbf{V}^{*-1} \mathbf{Z}^* &= (\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}) \text{vec}(\mathbf{A}) \\ &= \text{vec}(\mathbf{V}^{-1} \mathbf{A} \mathbf{V}^{-1}). \end{aligned} \quad (1.35)$$

Como $(\mathbf{V}^{-1} \mathbf{A} \mathbf{V}^{-1})$ es matriz simétrica,

$$S_N \mathbf{V}^{*-1} \mathbf{Z}^* = \mathbf{V}^{*-1} \mathbf{Z}^* . \quad (1.36)$$

Reemplazando las expresiones (1.34) , (1.35) y (1.36) en (1.33) se tiene que la matriz de covarianza para $\hat{\theta}$, es dada por:

$$\begin{aligned}
\text{cov}(\hat{\theta}) &= (\mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^*)^{-1} \mathbf{Z}^{*\top} \mathbf{V}^{*-1} ((\mathbf{V} \otimes \mathbf{V})(\mathbf{I} + \mathbf{S}_N)) \mathbf{V}^{*-1} \mathbf{Z}^* (\mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^*)^{-1} \\
&= (\mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^*)^{-1} \mathbf{Z}^{*\top} \mathbf{V}^{*-1} (\mathbf{V}^* \mathbf{I} + \mathbf{S}_N \mathbf{V}^*) \mathbf{V}^{*-1} \mathbf{Z}^* (\mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^*)^{-1} \\
&= (\mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^*)^{-1} \mathbf{Z}^{*\top} \mathbf{V}^{*-1} \left(\mathbf{Z}^{*\top} + \mathbf{S}_N \mathbf{Z}^{*\top} \right) \mathbf{V}^{*-1} \mathbf{Z}^* (\mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^*)^{-1} \\
&= (\mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^*)^{-1} \left(\mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^* + \mathbf{S}_N \mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^* \right) (\mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^*)^{-1} \\
&= (\mathbf{I} + \mathbf{I}) (\mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^*)^{-1} \\
&= 2 \left(\mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^* \right)^{-1} .
\end{aligned} \tag{1.37}$$

Usando *minimos cuadrados generalizados iterados* (IGLS) se producen estimaciones polarizadas y en general, esto puede ser importante en muestras pequeñas (Goldstein, 1989), muestra que una simple modificación sobre la iteración general restringida de mínimos cuadrados o por *verosimilitud máxima restringida* (REML) de estimaciones son insesgadas. Si se reescribe $\mathbf{Y}^{**} = \tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^\top$ y usando las estimaciones de los parámetros fijos $\hat{\beta}$ se obtiene:

$$\begin{aligned}
E(\mathbf{Y}^*) &= \mathbf{V}_2 - \mathbf{X} \text{cov}(\hat{\beta}) \mathbf{X}^\top \\
&= \mathbf{V}_2 - \mathbf{X} \left(\mathbf{X}^\top \mathbf{V}_2^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^\top ,
\end{aligned} \tag{1.38}$$

donde,

$$\mathbf{V}_2 = \text{diag}(\Omega) + \text{diag}(\sigma_e^2 \mathbf{I}_{n_j}) .$$

Hasta ahora, se ha realizado la estimación de los parámetros de efectos fijos y aleatorios para este tipo de modelos. A continuación se definirán los residuos para este tipo de modelos.

1.3. Residuos

El cálculo de los residuos en un modelos jerárquico con parámetros aleatorios se basará en el trabajo de Goldstein, (1995).

Se considerará el modelo jerárquico de la siguiente forma:

$$\tilde{\mathbf{Y}} = \mathbf{Y} - \hat{\mathbf{Y}} \quad , \quad (1.39)$$

bajo el modelo considerado,

$$\mathbf{Y} = \beta_0 + \beta_1 x_{1ij} + (u_{0j} + u_{1ij} x_{1ij} + e_{0ij}) \quad . \quad (1.40)$$

Y además,

$$\begin{aligned} \hat{\mathbf{Y}} &= \hat{y}_{ij} \\ &= \hat{\beta}_0 + \hat{\beta}_1 x_{ij} \quad . \end{aligned}$$

Para calcular los residuos en el segundo nivel, dado el estimador $\hat{\beta}$, es necesario calcular:

$$\hat{u}_{0j} = E \left(\hat{u}_{0j} / \mathbf{Y}, \hat{\beta}, \hat{\Omega} \right) \quad . \quad (1.41)$$

Considere el siguiente estimador consistente:

$$\hat{\mathbf{u}} = \frac{n_j \sigma_{u0}^2}{n_j \sigma_{u0}^2 + \sigma_{e0}^2} \tilde{y}_j \quad . \quad (1.42)$$

Goldstein (1995), se refiere a la necesidad de realizar inferencia en torno al verdadero valor de \hat{u}_{0j} , despreciando la variabilidad muestral de los parámetros aleatorios, entonces se obtiene:

$$\hat{u}_{0j} = \frac{n_j \sigma_u^2}{n_j \sigma_u^2 + \sigma_e^2} y_i, \quad \tilde{e}_{0ij} = \tilde{y}_{ij} - \hat{u}_{0j}, \quad \tilde{y}_j = \frac{\sum_{i=1}^{n_j} \tilde{y}_{ij}}{n_j} \quad . \quad (1.43)$$

Las estimaciones residuales no son insesgadas, sino que son constantes.

A continuación se mostrará un de tipo de diagnóstico que se utilizan en modelos jerárquicos con parámetros aleatorios, el cual es llamado *detección de medidas influyentes*.

Capítulo 2

Diagnóstico

Para los modelos jerárquicos con parámetros aleatorios, una de las técnicas de diagnóstico se presenta a continuación:

2.1. Detección de medidas influyentes para el primer nivel.

En general, los investigadores discuten la dificultad para obtener estimaciones satisfactorias que permitan confiar en las estimaciones obtenidas desde el modelo propuesto y el caso de los modelos jerárquicos no es una excepción (Galea, 1990).

Si bien, existen técnicas de diagnóstico para modelos lineales ordinarios, los modelos jerárquicos también tienen técnicas de diagnóstico, pero hay que tener cuidado que exista la suficiente variabilidad para justificar el uso del modelo jerárquico (Goldstein, 1995). Goldstein, propuso el uso de las varianzas estimadas de los errores del modelo el cual al formar una relación entre la variabilidad estimada de los errores de las variables del segundo nivel y la variabilidad de las mismas más las del primer nivel, se concluye un tipo de correlación intraclase estimada de la varianza residual. También, asegura que basta con que el coeficiente de correlación intraclase sea mayor a 0,1 para justificar el uso del modelo jerárquico con error en las variables (Goldstein, 1995) . También, existe otra técnica la cual estima los coeficientes aleatorios pertenecientes al modelo jerárquico a través del método de los momentos desarrollada de forma paralela por Hengjian, et al., (2004) y por Woodhouse, et al., (1996) a través de un modelo jerárquico en particular, y a través del método de mínimos cuadrados, propuesto por Goldstein y Rasbash (1992) .

En la siguiente sección se mostrará la estimación de parámetros fijos para un modelo jerárquico basado en el método de mínimos cuadrados generalizados propuesto por Goldstein y Rabash (1992) y a través del método de mínimos cuadrados generalizados iterados la estimación de parámetros aleatorios propuesto por Goldstein (1989) .

Lei y Gemai (2008), proponen como medida de diagnóstico la eliminación de datos influyentes en el análisis utilizando como herramienta de comparación la distancia de Cook's. Usando la iteración generalizada de estimación de mínimos cuadrados (IGLS) se producen estimaciones polarizadas y en

general esto puede ser importante en muestras pequeñas Goldstein, (1989) muestra que una simple modificación sobre la iteración general restringida de mínimos cuadrados o por verosimilitud máxima restringida (REML) de estimaciones son insesgadas.

Definición 2.1. Sea $[a]$ el conjunto de datos influyentes en el modelo. Y $\widehat{\beta}_{[a]}$ y $\widehat{\theta}_{[a]}$ los estimadores IGLS de β y θ , respectivamente, sin los a casos influyentes. Entonces:

$$\begin{aligned}\widehat{\beta}_{[a]} &= (\mathbf{X}_{[a]}^\top \widehat{\mathbf{V}}_{[a]}^{-1}(\widehat{\theta}_{[a]}) \mathbf{X}_{[a]})^{-1} \mathbf{X}_{[a]}^\top(\widehat{\theta}_{[a]}) \widehat{\mathbf{V}}_{[a]}^{-1} \mathbf{Y} \quad y \\ \widehat{\theta}_{[a]} &= (\mathbf{Z}_{[a]}^{*t} \mathbf{V}_{[a]}^{*-1}(\widehat{\theta}_{[a]}) \mathbf{Z}_{[a]}^*)^{-1} \mathbf{Z}_{[a]}^{*t} \mathbf{V}_{[a]}^{*-1}(\widehat{\theta}_{[a]}) \mathbf{Y}_{[a]}^{**} \quad ,\end{aligned}\tag{2.1}$$

donde,

$$\begin{aligned}\mathbf{Y}_{[a]}^{**} &= \text{vec}(\widehat{\mathbf{e}}_{[a]}, \widehat{\mathbf{e}}_{[a]}^\top) \quad , \\ \widehat{\mathbf{e}}_{[a]} &= \mathbf{Y}_{[a]} - \mathbf{X}_{[a]} \widehat{\beta}_{[a]} \quad y \\ \mathbf{V}_{[a]}^*(\widehat{\theta}_{[a]}) &= \mathbf{V}_{[a]}(\widehat{\theta}_{[a]}) \otimes \mathbf{V}_{[a]}(\widehat{\theta}_{[a]}) \quad .\end{aligned}$$

La distancia de Cook's puede ser usada como herramienta de medida para detectar observaciones influyentes. En este caso, tal medida es definida de la siguiente manera:

$$\begin{aligned}C_{[a]}(\widehat{\beta}) &= (\widehat{\beta}_{[a]} - \widehat{\beta})^\top \mathbf{X}^\top \widehat{\mathbf{V}}^{-1} \mathbf{X} (\widehat{\beta}_{[a]} - \widehat{\beta}) \quad y \\ C_{[a]}(\widehat{\theta}) &= (\widehat{\theta}_{[a]} - \widehat{\theta})^\top \mathbf{Z}^{*t} \widehat{\mathbf{V}}^{*-1} \mathbf{Z}^* (\widehat{\theta}_{[a]} - \widehat{\theta}) / 2 \quad .\end{aligned}\tag{2.2}$$

Como $\widehat{\beta}_{[a]}$ y $\widehat{\theta}_{[a]}$ son obtenidos por medio de procesos de estimación de mínimos cuadrados iterados, generalmente se reemplaza de $\widehat{\beta}_{[a]}$ y $\widehat{\theta}_{[a]}$ en la ecuación (2.1), por $\widetilde{\beta}$ y $\widetilde{\theta}$. Y es dado por las siguientes ecuaciones:

$$\begin{aligned}\widetilde{\beta}_{[a]} &= (\mathbf{X}_{[a]}^\top \widehat{\mathbf{V}}_{[a]}^{-1} \mathbf{X}_{[a]})^{-1} \mathbf{X}_{[a]}^\top \widehat{\mathbf{V}}_{[a]}^{-1} \mathbf{Y}_{[a]} \quad \text{y} \\ \widetilde{\theta}_{[a]} &= (\mathbf{Z}_{[a]}^{*\top} \widehat{\mathbf{V}}_{[a]}^{*-1} \mathbf{Z}_{[a]}^*)^{-1} \mathbf{Z}_{[a]}^{*\top} \widehat{\mathbf{V}}_{[a]}^{*-1} \mathbf{Y}_{o[a]}^{**},\end{aligned}\tag{2.3}$$

donde,

- (1) $\mathbf{Y}_{o[a]}^{**} = \text{vec}(\widehat{\mathbf{e}}_{o[a]}, \widehat{\mathbf{e}}_{o[a]}^\top)$,
- (2) $\widehat{\mathbf{e}}_{o[a]} = \mathbf{Y}_{[a]} - \mathbf{X}_{[a]} \widehat{\beta}$,
- (3) $\widehat{\mathbf{V}}_{[a]}^{**} = \widehat{\mathbf{V}}_{[a]} \otimes \widehat{\mathbf{V}}_{[a]}$,
- (4) $\widehat{\mathbf{V}}_{[a]} = \mathbf{V}_{[a]}(\widehat{\theta})$.

Este tipo de enfoque se llamará *Aproximación de Tipo I*.

Definición 2.2. Sea $a = \{i, \dots, i_m\}$ y $\mathbf{D}_a = (d_{i_1}, \dots, d_{i_m})$ donde d_{i_k} es un vector de orden $N \times 1$ con i_k elementos iguales a uno y el resto ceros, $\forall k = 1, \dots, m$.
Sea $\mathbf{I}_{[a]}$ la matriz identidad de orden $N \times N$, entonces se verifican las siguientes propiedades:

Algunas propiedades de $\mathbf{I}_{[a]}$ y \mathbf{D}_a son:

$$\begin{aligned}\mathbf{X}_{[a]} &= \mathbf{I}_{[a]} \mathbf{X} \quad , \\ \mathbf{V}_{[a]} &= \mathbf{I}_{[a]} \mathbf{V} \mathbf{I}_{[a]}^\top \quad , \\ \mathbf{I}_{[a]}^\top \mathbf{I}_{[a]} &= \mathbf{I}_N - \mathbf{D}_a \mathbf{D}_a^\top \quad , \\ \mathbf{I}_{[a]} \mathbf{I}_{[a]}^\top &= \mathbf{I}_{N-m} \quad , \\ \mathbf{I}_{[a]} \mathbf{D}_a &= \mathbf{0} \quad , \\ \mathbf{V}_{[a]}^{-1} &= \mathbf{I}_{[a]} \left(\mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{D}_a (\mathbf{D}_a^\top \mathbf{V}^{-1} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top \mathbf{V}^{-1} \right) \mathbf{I}_{[a]}^\top \quad .\end{aligned}\tag{2.4}$$

Teorema 2.1. Sea $\hat{\beta}$ y $\hat{\theta}$ los estimadores IGLS (mínimos cuadrados generalizados iterados) de β y θ , respectivamente, $\hat{\mathbf{V}} = \hat{\mathbf{V}}(\theta)$ y $\hat{\mathbf{r}} = \hat{\mathbf{V}}^{-1}\hat{\mathbf{e}}$. Entonces

$$\begin{aligned}\tilde{\beta}_{[a]} &= \hat{\beta} - \left(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{D}_a \left(\mathbf{D}_a^\top \hat{\mathbf{Q}} \mathbf{D}_a\right)^{-1} \mathbf{D}_a^\top \hat{\mathbf{r}} \quad \text{y} \\ \tilde{\theta}_{[a]} &= \hat{\theta} - \left(\mathbf{Z}^{*\top} \hat{\mathbf{V}}^{*-1} \mathbf{Z}^*\right)^{-1} \mathbf{Z}^{*\top} \hat{\mathbf{V}}^{*-1} (\mathbf{I} - \mathbf{B}_a) \mathbf{Y}^{**} \quad ,\end{aligned}\tag{2.5}$$

donde,

$$(1) \quad \hat{\mathbf{Q}} = \hat{\mathbf{V}}^{-1} - \hat{\mathbf{V}}^{-1} \mathbf{X} \left(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^\top \hat{\mathbf{V}}^{-1} \quad ,$$

$$(2) \quad \mathbf{B}_a = \mathbf{Z}^* \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^*\right)^{-1} \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \quad ,$$

$$(3) \quad \mathbf{N}_a = \hat{\mathbf{V}}^{-1} - \hat{\mathbf{V}}^{-1} \mathbf{D}_a \left(\mathbf{D}_a^\top \hat{\mathbf{V}}^{-1} \mathbf{D}_a\right)^{-1} \mathbf{D}_a^\top \hat{\mathbf{V}}^{-1} \quad .$$

DEMOSTRACIÓN. (Lei, et al., 2008)

Sea $\hat{\mathbf{V}} = \hat{\mathbf{V}}(\hat{\theta})$ y $\hat{\mathbf{V}}_{[a]} = \mathbf{V}_{[a]}(\hat{\theta})$. Utilizando las propiedades de la ecuación (2.4) y usando $\mathbf{N}_a \mathbf{D}_a = 0$, se tiene que:

$$\begin{aligned}\mathbf{X}_{[a]}^\top \hat{\mathbf{V}}_{[a]}^{-1} \mathbf{X}_{[a]} &= \mathbf{X}^\top \mathbf{N}_a \mathbf{X} \quad , \\ \mathbf{X}_{[a]}^\top \hat{\mathbf{V}}_{[a]}^{-1} \mathbf{Y}_{[a]} &= \mathbf{X}^\top \mathbf{N}_a \mathbf{Y} \quad ,\end{aligned}\tag{2.6}$$

donde $\mathbf{N}_a = \hat{\mathbf{V}}^{-1} - \hat{\mathbf{V}}^{-1} \mathbf{D}_a \left(\mathbf{D}_a^\top \hat{\mathbf{V}}^{-1} \mathbf{D}_a\right)^{-1} \mathbf{D}_a^\top \hat{\mathbf{V}}^{-1}$. Por lo tanto,

$$\tilde{\beta}_{[a]} = \left(\mathbf{X}^\top \mathbf{N}_a \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{N}_a \mathbf{Y} \quad .\tag{2.7}$$

Note que,

$$\begin{aligned}\left(\mathbf{X}^\top \mathbf{N}_a \mathbf{X}\right)^{-1} &= \left(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X}\right)^{-1} + \left(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{D}_a \\ &\quad \times \left[\mathbf{D}_a^\top \hat{\mathbf{V}}^{-1} \mathbf{D}_a - \mathbf{D}_a^\top \hat{\mathbf{V}}^{-1} \mathbf{X} \left(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{D}_a\right]^{-1} \\ &\quad \times \mathbf{D}_a^\top \hat{\mathbf{V}}^{-1} \mathbf{X} \left(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X}\right)^{-1} \\ &= \left(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X}\right)^{-1} + \left(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{D}_a \\ &\quad \times \left(\mathbf{D}_a^\top \hat{\mathbf{Q}} \mathbf{D}_a\right)^{-1} \mathbf{D}_a^\top \hat{\mathbf{V}}^{-1} \mathbf{X} \left(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X}\right)^{-1} \quad ,\end{aligned}\tag{2.8}$$

donde $\widehat{\mathbf{Q}} = \widehat{\mathbf{V}}^{-1} - \widehat{\mathbf{V}}^{-1}\mathbf{X}(\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}$. Luego,

$$\begin{aligned}\widetilde{\beta}_{[a]} &= \widehat{\beta} - (\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}_a^\top\widehat{\mathbf{V}}^{-1}\mathbf{D}_a)^{-1}\mathbf{D}_a^\top\widehat{\mathbf{V}}^{-1}\mathbf{Y} \\ &\quad + (\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}_a^\top\widehat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}_a^\top\widehat{\mathbf{V}}^{-1}\mathbf{X}\widehat{\beta} \\ &\quad - (\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}_a^\top\widehat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}_a^\top\widehat{\mathbf{P}}\mathbf{D}_a(\mathbf{D}_a^\top\widehat{\mathbf{V}}^{-1}\mathbf{D}_a)^{-1}\mathbf{D}_a^\top\widehat{\mathbf{V}}^{-1}\mathbf{Y} \\ &= \widehat{\beta} - (\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}_a^\top\widehat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}_a^\top\widehat{\mathbf{V}}^{-1}\widehat{\mathbf{e}}.\end{aligned}$$

Donde $\widehat{\mathbf{Q}} = \widehat{\mathbf{V}}^{-1} - \widehat{\mathbf{P}}$, $\widehat{\mathbf{e}} = \mathbf{Y} - \mathbf{X}\widehat{\beta}$ y $\widehat{\beta} = (\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{Y}$.

A continuación se muestra el desarrollo de la segunda ecuación del Teorema 2.1.

Por otro lado, sea $\mathbf{Z}_{[a]}^*$ la matriz de diseño asociada a θ , cuando las $[a]$ observaciones han sido removidas. Entonces,

$$\text{vec}[\mathbf{V}_{[a]}(\theta)] = \mathbf{Z}_{[a]}^*\theta$$

Desde,

$$\begin{aligned}\text{vec}[\mathbf{V}_{[a]}(\theta)] &= \text{vec}(\mathbf{I}_{[a]}\mathbf{V}(\theta)\mathbf{I}_{[a]}^\top) \\ &= (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})\text{vec}(\mathbf{V}(\theta)) \\ &= (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})\mathbf{Z}^*\theta,\end{aligned}$$

se tiene que

$$\mathbf{Z}_{[a]}^* = (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})\mathbf{Z}^* \quad . \quad (2.9)$$

Usando las propiedades de $\mathbf{I}_{[a]}$ y \mathbf{D}_a (ecuaciones de (2.4)),

$$\begin{aligned}\widehat{\mathbf{V}}_{[a]}^{*-1} &= \widehat{\mathbf{V}}_{[a]}^{-1} \otimes \widehat{\mathbf{V}}_{[a]}^{-1} \\ &= (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})(\mathbf{N}_{[a]} \otimes \mathbf{N}_{[a]})(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})^\top.\end{aligned} \quad (2.10)$$

Desde la ecuación (2.4) y $\mathbf{N}_a \mathbf{D}_a = \mathbf{0}$, se obtiene:

$$\begin{aligned}
\mathbf{Z}_{[a]}^{*\top} \mathbf{V}_{[a]}^{*-1} (\hat{\theta}) \mathbf{Z}_{[a]}^{*\top} &= \mathbf{Z}^{*\top} (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})^\top (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]}) (\mathbf{N}_a \otimes \mathbf{N}_a) (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})^\top (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]}) \mathbf{Z}^* \\
&= \mathbf{Z}^{*\top} (\mathbf{I}_{[a]}^\top \mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]}^\top \mathbf{I}_{[a]}) (\mathbf{N}_a \otimes \mathbf{N}_a) (\mathbf{I}_{[a]}^\top \mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]}^\top \mathbf{I}_{[a]}) \mathbf{Z}^* \\
&= \mathbf{Z}^{*\top} \left((\mathbf{I}_N - \mathbf{D}_a \mathbf{D}_a^\top) \otimes (\mathbf{I}_N - \mathbf{D}_a \mathbf{D}_a^\top) \right) (\mathbf{N}_a \otimes \mathbf{N}_a) \\
&\quad \times \left((\mathbf{I}_N - \mathbf{D}_a \mathbf{D}_a^\top) \otimes (\mathbf{I}_N - \mathbf{D}_a \mathbf{D}_a^\top) \right) \mathbf{Z}^* \\
&= \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \quad .
\end{aligned} \tag{2.11}$$

Además,

$$\begin{aligned}
\mathbf{Z}_{[a]}^* \mathbf{V}_{[a]}^{*-1} (\hat{\theta}) \mathbf{Y}_{0[a]}^{**} &= \mathbf{Z}^{*\top} (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})^\top (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]}) (\mathbf{N}_a \otimes \mathbf{N}_a) (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})^\top \\
&\quad \times \text{vec} \left[(\mathbf{Y}_{[a]} - \mathbf{X}_{[a]} \hat{\beta}) (\mathbf{Y}_{[a]} - \mathbf{X}_{[a]} \hat{\beta})^\top \right] \\
&= \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})^\top (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]}) \text{vec} [(\mathbf{Y} - \mathbf{X} \hat{\beta}) (\mathbf{Y} - \mathbf{X} \hat{\beta})^\top] \\
&= \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \text{vec} [\hat{\mathbf{e}} \hat{\mathbf{e}}^\top] \quad .
\end{aligned}$$

Así,

$$\begin{aligned}
\tilde{\theta}_{[a]} &= (\mathbf{Z}_{[a]}^{*\top} \mathbf{V}_{[a]}^{*-1} (\hat{\theta}) \mathbf{Z}_{[a]}^*)^{-1} \mathbf{Z}^{*\top} \mathbf{V}_{[a]}^{*-1} \mathbf{Y}_{0[a]}^{**} \\
&= \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Y}^{**} \quad .
\end{aligned} \tag{2.12}$$

Consecuentemente,

$$\tilde{\theta}_{[a]} = \hat{\theta} - (\mathbf{Z}^{*\top} \mathbf{V}^{*-1} \mathbf{Z}^*)^{-1} \mathbf{Z}^{*\top} \hat{\mathbf{V}}^{*-1} (\mathbf{I} - \mathbf{B}_a) \mathbf{Y}^{**} \quad . \tag{2.13}$$

Definición 2.3. Sea $\widehat{\mathbf{Q}}_{aa} = \mathbf{D}_a^\top \widehat{\mathbf{Q}} \mathbf{D}_a$ y $\widehat{\mathbf{r}}_a = \mathbf{D}_a^\top \widehat{\mathbf{r}}$. Entonces la extracción de casos basados en la aproximación del tipo I queda,

$$\begin{aligned} C_\alpha^I(\widehat{\beta}) &= \widehat{\mathbf{r}}_a^\top \widehat{\mathbf{Q}}_{aa}^{-1} \widehat{\mathbf{P}}_{aa} \widehat{\mathbf{Q}}_{aa}^{-1} \widehat{\mathbf{r}}_a \quad \text{y} \\ C_\alpha^I(\widehat{\theta}) &= \frac{\mathbf{Y}^{**\top} (\mathbf{I} - \mathbf{B}_a)^\top \widehat{\mathbf{P}}^* (\mathbf{I} - \mathbf{B}_a) \mathbf{Y}^{**}}{2} \quad , \end{aligned} \quad (2.14)$$

donde,

$$\begin{aligned} \widehat{\mathbf{P}} &= \widehat{\mathbf{V}}^{-1} \mathbf{X} (\mathbf{X}^\top \widehat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \widehat{\mathbf{V}}^{-1} \quad \text{y} \\ \widehat{\mathbf{P}}^* &= \widehat{\mathbf{V}}^{*-1} \mathbf{Z}^* (\mathbf{Z}^{*\top} \widehat{\mathbf{V}}^{*-1} \mathbf{Z}^*) \mathbf{Z}^{*\top} \widehat{\mathbf{V}}^{*-1} \quad . \end{aligned}$$

Estas medidas permiten realizar el diagnóstico en modelos jerárquicos a cualquier nivel. La aproximación tipo I no tiene dificultad en implementarse, sin embargo, tiene algunas fallas, por ejemplo, las observaciones asociadas a a son muy influyentes en las estimaciones de los parámetros β y θ , entonces sustituyendo $\widehat{\beta}$ y $\widehat{\theta}$ dentro de la ecuación (2.1) para obtener la ecuación (2.3) admite un subconjunto de datos suprimidos, pero esencialmente ignora la influencia con otras variables. Para solucionar este problema, se estudia la nueva aproximación basada en la expansión de Taylor de $\mathbf{V}_{[a]}(\widehat{\theta}_{[a]})$ alrededor de $\widehat{\theta}$, esto es,

$$\mathbf{V}_{[a]}(\widehat{\theta}_{[a]}) = \mathbf{V}_{[a]}(\widehat{\theta}) + \mathbf{V}_{[a]}(\widehat{\theta})(\widehat{\theta}_{[a]} - \widehat{\theta}) + o(\|\widehat{\theta}_{[a]} - \widehat{\theta}\|^2) \quad , \quad (2.15)$$

donde,

$$\mathbf{V}_{[a]}(\widehat{\theta}) = \frac{\partial \mathbf{V}_{[a]}(\theta)}{\partial \theta} \Big|_{\widehat{\theta}} \quad ,$$

$\mathbf{V}_{[a]}(\widehat{\theta})$ es una matriz es de orden $(N - m) \times (N - m) \times R$ si los datos asociados a a tienen m índices, y $\|\cdot\|$ denota la norma euclídeana. Sustituyendo la ecuación (2.15) en la ecuación (2.1) e ignorando el término de segundo orden $o(\|\widehat{\theta}_{[a]} - \widehat{\theta}\|^2)$. Se derivan dos nuevas fórmulas, las que se denominan *Aproximación de Tipo II*.

Teorema 2.2. *Asuma que $o(\|\widehat{\theta}_{[a]} - \widehat{\theta}\|^2)$ es despreciable, entonces las fórmulas para la aproximación de tipo II son las siguientes,*

$$\begin{aligned} \bar{\beta}_{[a]} &= \widehat{\beta} - (\mathbf{X}^\top \widehat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \widehat{\mathbf{V}}^{-1} \left[\mathbf{D}_a (\mathbf{D}_a^\top \widehat{\mathbf{Q}} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top \widehat{\mathbf{r}} + (\mathbf{M}_a \widehat{\mathbf{r}} \otimes \mathbf{M}_a)^\top \mathbf{Z}^* (\bar{\theta}_{[a]} - \widehat{\theta}) \right] \\ &\text{y} \\ \bar{\theta}_{[a]} &= \widehat{\theta} + (\mathbf{Z}^{*\top} \mathbf{W}_a \mathbf{Z}^*)^{-1} \mathbf{Z}^{*\top} \text{vec}(\mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top - \mathbf{N}_a) \quad , \end{aligned} \quad (2.16)$$

donde,

$$\begin{aligned}
\mathbf{M}_a &= \mathbf{I}_N - \widehat{\mathbf{Q}}\mathbf{D}_a(\mathbf{D}_a^\top \widehat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}_a^\top \quad , \\
\mathbf{W}_a &= \mathbf{N}_a \otimes \mathbf{N}_a - \mathbf{N}_a \otimes (\mathbf{N}_a \mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}^*)\mathbf{N}_a) - (\mathbf{N}_a \mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}^*)\mathbf{N}_a) \otimes \mathbf{N}_a \\
&\quad + \mathbf{M}_a \widehat{\mathbf{Q}} \otimes (\mathbf{M}_a \widehat{\mathbf{r}}\widehat{\mathbf{r}}^\top \mathbf{M}_a^\top) + \mathbf{M}_a \widehat{\mathbf{r}}\widehat{\mathbf{r}}^\top \mathbf{M}_a^\top \otimes \mathbf{M}_a \widehat{\mathbf{Q}} \quad \text{y} \\
\widehat{\boldsymbol{\theta}}_{[a]}^* &= (\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a \mathbf{Z}^*)^{-1} \mathbf{Z}^{*\top} \text{vec}(\mathbf{M}_a \widehat{\mathbf{r}}\widehat{\mathbf{r}}^\top \mathbf{M}_a^\top)) \quad .
\end{aligned}$$

y $\widehat{\mathbf{Q}}$ y \mathbf{N}_a están descritas en el Teorema 2.1.

DEMOSTRACIÓN. (Lei, et al., 2008)

Considere la expansión de Taylor aproximado a $\mathbf{V}_{[a]}(\widehat{\boldsymbol{\theta}}_{[a]})$,

$$\mathbf{V}_{[a]}(\widehat{\boldsymbol{\theta}}_{[a]}) = \mathbf{V}_{[a]}(\widehat{\boldsymbol{\theta}}) + \mathbf{V}_{[a]}^{(1)} + o(\eta_a^2) \quad , \tag{2.17}$$

donde,

$$\begin{aligned}
\eta_a &= \|\widehat{\boldsymbol{\theta}}_a - \widehat{\boldsymbol{\theta}}\| \quad \text{y} \\
\mathbf{V}_{[a]}^{(1)} &= \mathbf{V}_{[a]}(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}}_{[a]} - \widehat{\boldsymbol{\theta}}) \quad , \\
&= o(\eta_a) \quad .
\end{aligned}$$

Considerando $o(\eta_a)$ despreciable, entonces:

$$\begin{aligned}
\mathbf{V}_{[a]}^{(1)} &= \mathbf{V}_{[a]}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}_{[a]}(\widehat{\boldsymbol{\theta}}) + o(\eta_a^2) \\
&= \mathbf{I}_{[a]}(\mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}}))\mathbf{I}_{[a]}^\top + o(\eta_a^2).
\end{aligned} \tag{2.18}$$

Por lo tanto, $\mathbf{V}_{[a]}^{(1)}$ satisface,

$$\begin{aligned}
\mathbf{I}_{[a]}^\top \mathbf{V}_{[a]}^{(1)} \mathbf{I}_{[a]} &= \mathbf{I}_{[a]}^\top \mathbf{I}_{[a]} (\mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}})) \mathbf{I}_{[a]}^\top \mathbf{I}_{[a]} + o(\eta_a^2) \\
&= (\mathbf{I}_N - \mathbf{D}_a \mathbf{D}_a^\top) (\mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}})) (\mathbf{I}_N - \mathbf{D}_a \mathbf{D}_a^\top) + o(\eta_a^2).
\end{aligned} \tag{2.19}$$

Considere el resultado:

$$[\mathbf{A} + \mathbf{B}\epsilon + o(\epsilon^2)]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\epsilon + o(\epsilon^2),$$

Asumiendo que la matriz A es definida positiva. Luego, para $\mathbf{V}_{[a]}^{(1)} = o(\eta_a)$, se tiene:

$$\begin{aligned} \mathbf{V}_{[a]}^{-1}(\hat{\theta}_{[a]}) &= (\mathbf{V}_{[a]}(\hat{\theta}) + \hat{\mathbf{V}}_{[a]}^{(1)} + o(\eta_a^2))^{-1} \\ &= \hat{\mathbf{V}}_{[a]}^{-1} - \hat{\mathbf{V}}_{[a]}^{-1}\mathbf{V}_{[a]}^{(1)}\hat{\mathbf{V}}_{[a]}^{-1} + o(\eta_a^2) \quad , \end{aligned} \quad (2.20)$$

donde $\hat{\mathbf{V}}_{[a]}^{-1} = \mathbf{V}_{[a]}^{-1}(\hat{\theta})$. Así, desde las ecuaciones (2.6), (2.19) y (2.20), se puede generar lo siguiente:

$$\begin{aligned} (\mathbf{X}_{[a]}^\top \mathbf{V}_{[a]}^{-1}(\hat{\theta}_{[a]}) \mathbf{X}_{[a]})^{-1} &= (\mathbf{X}_{[a]}^\top \hat{\mathbf{V}}_{[a]}^{-1} \mathbf{X}_{[a]} - \mathbf{X}_{[a]}^\top \hat{\mathbf{V}}_{[a]}^{-1} \mathbf{V}_{[a]}^{(1)} \hat{\mathbf{V}}_{[a]}^{-1} \mathbf{X}_{[a]} + o(\eta_a^2))^{-1} \quad , \\ &= (\mathbf{X}_{[a]}^\top \hat{\mathbf{V}}_{[a]}^{-1} \mathbf{X}_{[a]})^{-1} + (\mathbf{X}_{[a]}^\top \hat{\mathbf{V}}_{[a]}^{-1} \mathbf{X}_{[a]})^{-1} \mathbf{X}_{[a]}^\top \hat{\mathbf{V}}_{[a]}^{-1} \mathbf{V}_{[a]}^{(1)} \hat{\mathbf{V}}_{[a]}^{-1} \mathbf{X}_{[a]} \quad , \\ &\quad \times (\mathbf{X}_{[a]}^\top \hat{\mathbf{V}}_{[a]}^{-1} \mathbf{X}_{[a]})^{-1} + o(\eta_a^2) \quad , \\ &= (\mathbf{X}^\top \mathbf{N}_a \mathbf{X})^{-1} + (\mathbf{X}^\top \mathbf{N}_a \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{I}_{[a]}^\top \mathbf{I}_{[a]} \mathbf{N}_a \mathbf{I}_{[a]} \mathbf{V}_{[a]}^{(1)} \mathbf{I}_{[a]} \mathbf{N}_a \mathbf{I}_{[a]}^\top \mathbf{I}_{[a]} \\ &\quad \times \mathbf{X} (\mathbf{X}^\top \mathbf{N}_a \mathbf{X})^{-1} + o(\eta_a^2) \quad , \\ &= (\mathbf{X}^\top \mathbf{N}_a \mathbf{X})^{-1} + (\mathbf{X}^\top \mathbf{N}_a \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{N}_a (\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta})) \mathbf{N}_a \mathbf{X} (\mathbf{X}^\top \mathbf{N}_a \mathbf{X})^{-1} \quad , \\ &\quad + o(\eta_a^2) \quad . \end{aligned}$$

De manera similar,

$$\mathbf{X}_{[a]}^\top \mathbf{V}_{[a]}^{-1}(\hat{\theta}_{[a]}) \mathbf{Y}_{[a]} = \mathbf{X}^\top \mathbf{N}_a \mathbf{Y} - \mathbf{X}^\top \mathbf{N}_a (\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta})) \mathbf{N}_a \mathbf{Y} + o(\eta_a^2) \quad .$$

Sustituyendo las dos ecuaciones anteriores en la primera ecuación (2.1), se obtiene:

$$\hat{\beta}_{[a]} = \tilde{\beta}_{[a]} - (\mathbf{X}^\top \mathbf{N}_a \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{N}_a (\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta})) \mathbf{N}_a (\mathbf{Y} - \mathbf{X} \tilde{\beta}_{[a]}) + o(\eta_a^2) \quad . \quad (2.21)$$

Usando la ecuación (2.8), se obtiene:

$$(\mathbf{X}^\top \mathbf{N}_a \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{N}_a = (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{M}_a^\top ,$$

y,

$$\begin{aligned} \mathbf{N}_a(\mathbf{Y} - \mathbf{X}\tilde{\beta}_{[a]}) &= \mathbf{N}_a \left[\mathbf{I}_N + \mathbf{X}(\mathbf{X}^\top \widehat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \widehat{\mathbf{V}}^{-1} \mathbf{D}_a (\mathbf{D}_a^\top \widehat{\mathbf{Q}} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top \widehat{\mathbf{V}}^{-1} \right] \widehat{\mathbf{e}} \\ &= \left[\widehat{\mathbf{V}}^{-1} - \widehat{\mathbf{V}}^{-1} \mathbf{D}_a (\mathbf{D}_a^\top \widehat{\mathbf{V}}^{-1} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top \widehat{\mathbf{V}}^{-1} \right] \\ &\quad \times \left[\mathbf{I}_N + \mathbf{X}(\mathbf{X}^\top \widehat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \widehat{\mathbf{V}}^{-1} \mathbf{D}_a (\mathbf{D}_a^\top \widehat{\mathbf{Q}} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top \widehat{\mathbf{V}}^{-1} \right] \widehat{\mathbf{e}} \\ &= [\mathbf{I}_N + \widehat{\mathbf{P}} \mathbf{D}_a (\mathbf{D}_a^\top \widehat{\mathbf{Q}} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top - \widehat{\mathbf{V}}^{-1} \mathbf{D}_a (\mathbf{D}_a^\top \widehat{\mathbf{V}}^{-1} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top \\ &\quad - \widehat{\mathbf{V}}^{-1} \mathbf{D}_a (\mathbf{D}_a^\top \widehat{\mathbf{V}}^{-1} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top \widehat{\mathbf{P}} \mathbf{D}_a (\mathbf{D}_a^\top \widehat{\mathbf{Q}} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top] \widehat{\mathbf{r}} \\ &= [\mathbf{I}_N + \widehat{\mathbf{P}} \mathbf{D}_a (\mathbf{D}_a^\top \widehat{\mathbf{Q}} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top - \widehat{\mathbf{V}}^{-1} \mathbf{D}_a (\mathbf{D}_a^\top \widehat{\mathbf{Q}} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top] \widehat{\mathbf{r}} \\ &= [\mathbf{I}_N - \widehat{\mathbf{Q}} \mathbf{D}_a (\mathbf{D}_a^\top \widehat{\mathbf{Q}} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top] \widehat{\mathbf{r}} \\ &= \mathbf{M}_a \widehat{\mathbf{r}} , \end{aligned} \tag{2.22}$$

donde $\widehat{\mathbf{r}} = \widehat{\mathbf{V}}^{-1} \widehat{\mathbf{e}}$ y $\mathbf{M}_a = \mathbf{I}_N - \widehat{\mathbf{Q}} (\mathbf{D}_a \mathbf{D}_a^\top \widehat{\mathbf{Q}} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top$. Así,

$$\begin{aligned} \widehat{\beta}_{[a]} &= \tilde{\beta}_{[a]} - (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{M}_a^\top (\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) \mathbf{M}_a \widehat{\mathbf{r}} + o(\eta_a^2) \\ &= \tilde{\beta}_{[a]} - (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} (\mathbf{M}_a \widehat{\mathbf{r}} \otimes \mathbf{M}_a)^\top \text{vec}(\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) + o(\eta_a^2) \\ &= \widehat{\beta} - (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} [\mathbf{D}_a (\mathbf{D}_a^\top \widehat{\mathbf{Q}} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top \widehat{\mathbf{r}} + (\mathbf{M}_a \widehat{\mathbf{r}} \otimes \mathbf{M}_a)^\top \\ &\quad \times \mathbf{Z}^*(\widehat{\theta}_{[a]} - \widehat{\theta})] + o(\eta_a^2) . \end{aligned} \tag{2.23}$$

A continuación, se considera la fórmula actualizada para $\widehat{\theta}$. Se observa a partir de la ecuación (2.20) que,

$$\begin{aligned} \mathbf{V}_{[a]}^{*-1}(\widehat{\theta}_{[a]}) &= \mathbf{V}_{[a]}^{-1}(\widehat{\theta}_{[a]}) \otimes \mathbf{V}_{[a]}^{-1}(\widehat{\theta}_{[a]}) \\ &= \widehat{\mathbf{V}}_{[a]}^{-1} \otimes \widehat{\mathbf{V}}_{[a]}^{-1} - \left(\widehat{\mathbf{V}}_{[a]}^{-1} \otimes \widehat{\mathbf{V}}_{[a]}^{-1} \mathbf{V}_{[a]}^{(1)} \widehat{\mathbf{V}}_{[a]}^{-1} + \widehat{\mathbf{V}}_{[a]}^{-1} \mathbf{V}_{[a]}^{(1)} \widehat{\mathbf{V}}_{[a]}^{-1} \otimes \widehat{\mathbf{V}}_{[a]}^{-1} \right) + o(\eta_a^2) \\ &= (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]}) [\mathbf{N}_a \otimes \mathbf{N}_a - \mathbf{N}_a \otimes \mathbf{N}_a \mathbf{I}_{[a]}^\top \mathbf{V}_{[a]}^{(1)} \mathbf{I}_{[a]} \mathbf{N}_a \\ &\quad + \mathbf{N}_a \mathbf{I}_{[a]}^\top \mathbf{V}_{[a]}^{(1)} \mathbf{I}_{[a]} \mathbf{N}_a \otimes \mathbf{N}_a] \left(\mathbf{I}_{[a]}^\top \otimes \mathbf{I}_{[a]}^\top \right) + o(\eta_a^2) . \end{aligned} \tag{2.24}$$

Utilizando los resultados de (2.8), (2.9) y (2.24) se obtiene

$$\begin{aligned}
(\mathbf{Z}_{[a]}^{*-1} \mathbf{V}_{[a]}^{*-1} (\widehat{\theta}_{[a]}) \mathbf{Z}_{[a]}^*)^{-1} &= \left[\mathbf{Z}_{[a]}^{*\top} \mathbf{V}_{[a]}^{*-1} \mathbf{Z}_{[a]}^* - \mathbf{Z}_{[a]}^{*\top} (\widehat{\mathbf{V}}_{[a]}^{-1} \otimes \widehat{\mathbf{V}}_{[a]}^{-1} \mathbf{V}_{[a]}^{(1)} \widehat{\mathbf{V}}_{[a]}^{-1} \right. \\
&\quad \left. + \widehat{\mathbf{V}}_{[a]}^{-1} \mathbf{V}_{[a]}^{(1)} \widehat{\mathbf{V}}_{[a]}^{-1} \otimes \widehat{\mathbf{V}}_{[a]}^{-1} \right] \mathbf{Z}_{[a]}^* + o(\eta_a^2) \Big]^{-1} \\
&= \left[\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* - \mathbf{Z}^{*\top} (\mathbf{I}_{[a]}^\top \otimes \mathbf{I}_{[a]}^\top) (\mathbf{I}_{[a]} \mathbf{N}_a \mathbf{I}_a^\top \otimes \mathbf{I}_{[a]} \mathbf{N}_a \mathbf{I}_a^\top \mathbf{V}_{[a]}^{(1)} \mathbf{I}_{[a]} \mathbf{N}_a \mathbf{I}_a^\top \right. \\
&\quad \left. + \mathbf{I}_{[a]} \mathbf{N}_a \mathbf{I}_a^\top \mathbf{V}_{[a]}^{(1)} \mathbf{I}_{[a]} \mathbf{N}_a \mathbf{I}_a^\top \otimes \mathbf{I}_{[a]} \mathbf{N}_a \mathbf{I}_a^\top) (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]}) \mathbf{Z}^* + o(\eta_a^2) \right]^{-1} \\
&= \left[\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* - \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a \mathbf{I}_{[a]}^\top \mathbf{V}_{[a]}^{(1)} \mathbf{I}_{[a]} \mathbf{N}_a \right. \\
&\quad \left. + \mathbf{N}_a \mathbf{I}_{[a]}^\top \mathbf{V}_{[a]}^{(1)} \mathbf{I}_{[a]} \mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* + o(\eta_a^2) \right]^{-1} \\
&= \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} + \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \left\{ \mathbf{Z}^{*\top} [\mathbf{N}_a \otimes \mathbf{N}_a \mathbf{I}_{[a]}^\top \right. \\
&\quad \left. \times \mathbf{V}_{[a]}^{(1)} \mathbf{I}_{[a]} \mathbf{N}_a + \mathbf{N}_a \mathbf{I}_{[a]}^\top \mathbf{V}_{[a]}^{(1)} \mathbf{I}_{[a]} \mathbf{N}_a \otimes \mathbf{N}_a] \mathbf{Z}^* \right\} \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \\
&\quad + o(\eta_a^2) \\
&= \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} + \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^* [\mathbf{N}_a \otimes \mathbf{N}_a \\
&\quad \times (\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) \mathbf{N}_a + \mathbf{N}_a (\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) \mathbf{N}_a \otimes \mathbf{N}_a] \mathbf{Z}^{*\top} \\
&\quad \times \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} + o(\eta_a^2) \quad . \tag{2.25}
\end{aligned}$$

Entonces,

$$\mathbf{Y}_{[a]}^{**} = (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]}) \text{vec} \left((\mathbf{Y} - \mathbf{X} \widehat{\beta}_{[a]}) (\mathbf{Y} - \mathbf{X} \widehat{\beta}_{[a]})^\top \right) \quad .$$

Además,

$$\begin{aligned}
\mathbf{Z}_{[a]}^{*\top} \mathbf{V}_{[a]}^{*-1} (\widehat{\theta}_{[a]}) \mathbf{Y}_{[a]}^{**} &= \mathbf{Z}^{*\top} \left(\mathbf{I}_{[a]}^\top \otimes \mathbf{I}_{[a]}^\top \right) \left[(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]}) (\mathbf{N}_a \otimes \mathbf{N}_a - \mathbf{N}_a \otimes \mathbf{N}_a \mathbf{I}_{[a]}^\top \mathbf{V}_{[a]}^{(1)} \mathbf{I}_{[a]} \mathbf{N}_a \right. \\
&\quad \left. - \mathbf{N}_a \mathbf{I}_{[a]}^\top \mathbf{V}_{[a]}^{(1)} \mathbf{I}_{[a]} \mathbf{N}_a \otimes \mathbf{N}_a) \left(\mathbf{I}_{[a]}^\top \otimes \mathbf{I}_{[a]}^\top \right) \right] (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]}) \\
&\quad \times \text{vec} \left((\mathbf{Y} - \mathbf{X} \widehat{\beta}_{[a]}) (\mathbf{Y} - \mathbf{X} \widehat{\beta}_{[a]})^\top \right) + o(\eta_a^2) \\
&= \mathbf{Z}^{*\top} \left[\mathbf{N}_a \otimes \mathbf{N}_a - \mathbf{N}_a \otimes \mathbf{N}_a (\mathbf{V}(\widehat{\theta}_{[a]}) \right. \\
&\quad \left. - \mathbf{V}(\widehat{\theta})) \mathbf{N}_a - \mathbf{N}_a (\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) \mathbf{N}_a \otimes \mathbf{N}_a \right] \\
&\quad \times \text{vec} \left((\mathbf{Y} - \mathbf{X} \widehat{\beta}_{[a]}) (\mathbf{Y} - \mathbf{X} \widehat{\beta}_{[a]})^\top \right) + o(\eta_a^2) \quad . \tag{2.26}
\end{aligned}$$

Note que,

$$\begin{aligned}
\widehat{\beta}_{[a]} &= \widetilde{\beta}_{[a]} - (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{M}_a^\top (\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) \mathbf{M}_a \widehat{\mathbf{r}} + o(\eta_a^2) \\
&= \widetilde{\beta}_{[a]} - \widehat{\beta}_{[a]}^{(1)} + o(\eta_a^2) \quad ,
\end{aligned} \tag{2.27}$$

donde,

$$\widehat{\beta}_{[a]}^{(1)} = (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{M}_a^\top (\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) \mathbf{M}_a \widehat{\mathbf{r}} \quad .$$

Luego,

$$Y - X \widehat{\beta}_{[a]} = \widehat{e}_{[a]0} + X \widehat{\beta}_{[a]}^{(1)} \quad , \tag{2.28}$$

donde $\widehat{\mathbf{e}}_{[a]0} = \mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}_{[a]}$. Sustituyendo las ecuaciones (2.25) y (2.26) en la segunda ecuación de (2.1) y usando (2.28), se tiene que,

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_{[a]} &= \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} \left[\mathbf{N}_a \otimes \mathbf{N}_a - \mathbf{N}_a \otimes \mathbf{N}_a (\widehat{\mathbf{V}}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}})) \right. \\
&\quad \times \mathbf{N}_a - \mathbf{N}_a (\mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}})) \mathbf{N}_a \otimes \mathbf{N}_a \left. \right] \text{vec} \left((\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{[a]})(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{[a]})^\top \right) \\
&\quad + \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} \left[\mathbf{N}_a \otimes \mathbf{N}_a (\mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}})) \mathbf{N}_a \right. \\
&\quad \left. + \mathbf{N}_a (\mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}})) \mathbf{N}_a \otimes \mathbf{N}_a \right] \mathbf{Z}^* \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \\
&\quad \times \text{vec} \left((\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{[a]})(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{[a]})^\top \right) + o(\eta_a^2) \\
&= \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \text{vec} \left((\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{[a]})(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{[a]})^\top \right) \\
&\quad - \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} \left[\mathbf{N}_a \otimes \mathbf{N}_a (\widehat{\mathbf{V}}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}})) \mathbf{N}_a \right. \\
&\quad \left. + \mathbf{N}_a (\mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}})) \mathbf{N}_a \otimes \mathbf{N}_a \right] \left[\mathbf{I} - \mathbf{Z}^* \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \right] \\
&\quad \times \text{vec} \left((\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}_{[a]})(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}_{[a]})^\top \right) + o(\eta_a^2) \\
&= \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \text{vec} \left((\widehat{\mathbf{e}}_{[a]0} + \mathbf{X}\widehat{\boldsymbol{\beta}}_{[a]}^{(1)})(\widehat{\mathbf{e}}_{[a]0} + \mathbf{X}\widehat{\boldsymbol{\beta}}_{[a]}^{(1)})^\top \right) \\
&\quad - \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} \left[\mathbf{N}_a \otimes \mathbf{N}_a (\mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}})) \mathbf{N}_a \right. \\
&\quad \left. + \mathbf{N}_a (\widehat{\mathbf{V}}_{[a]} - \widehat{\mathbf{V}}) \mathbf{N}_a \otimes \mathbf{N}_a \right] \left[\text{vec}(\widehat{\mathbf{e}}_{[a]0} \widehat{\mathbf{e}}_{[a]0}^\top) - \mathbf{Z}^* \widehat{\boldsymbol{\theta}}_{[a]}^* \right] + o(\eta_a^2) \\
&= \widehat{\boldsymbol{\theta}}_{[a]}^* + \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \text{vec} \left(\widehat{\mathbf{e}}_{[a]0} \widehat{\boldsymbol{\beta}}_{[a]}^{(1)\top} \mathbf{X}^\top + \mathbf{X} \widehat{\boldsymbol{\beta}}_{[a]}^{(1)} \widehat{\mathbf{e}}_{[a]0} \right) \\
&\quad - \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} \left[\mathbf{N}_a \otimes \mathbf{N}_a (\mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}})) \mathbf{N}_a \right. \\
&\quad \left. + \mathbf{N}_a (\mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}})) \mathbf{N}_a \otimes \mathbf{N}_a \right] \left[\text{vec}(\widehat{\mathbf{e}}_{[a]0} \widehat{\mathbf{e}}_{[a]0}^\top) - \mathbf{Z}^* \widehat{\boldsymbol{\theta}}_{[a]}^* \right] + o(\eta_a^2) \quad , \tag{2.29}
\end{aligned}$$

donde $\widehat{\boldsymbol{\theta}}_{[a]}^* = \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} \text{vec} (\mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top)$. Los siguientes factores son usados sobre $\widehat{\boldsymbol{\beta}}_{[a]}^{(1)} = o(\eta_a^2)$ y $\mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}}) = o(\eta_a)$. Note que $\mathbf{N}_a \widehat{\mathbf{e}}_{[a]0} = \mathbf{N}_a (\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}_{[a]}) = \mathbf{M}_a \widehat{\mathbf{r}}$ (ver ecuación (2.22)), y además,

$$\begin{aligned}
\mathbf{N}_a \mathbf{X} \widehat{\boldsymbol{\beta}}_{[a]}^{(1)} &= \left(\mathbf{I}_N - \widehat{\mathbf{V}}^{-1} \mathbf{D}_a \left(\mathbf{D}_a^\top \widehat{\mathbf{V}}^{-1} \mathbf{D}_a \right)^{-1} \mathbf{D}_a^\top \right) \widehat{\mathbf{P}} \mathbf{M}_a^\top \left(\mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}}) \right) \mathbf{M}_a \widehat{\mathbf{r}} \\
&= \left(\mathbf{N}_a - \mathbf{M}_a \widehat{\mathbf{Q}} \right) \left(\mathbf{V}(\widehat{\boldsymbol{\theta}}_{[a]}) - \mathbf{V}(\widehat{\boldsymbol{\theta}}) \right) \mathbf{M}_a \widehat{\mathbf{r}} \quad , \tag{2.30}
\end{aligned}$$

inmediatamente se obtiene,

$$\begin{aligned}
\widehat{\theta}_{[a]} &= \widehat{\theta}_{[a]}^* + \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} \text{vec} \left[\mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top (\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) (\mathbf{N}_a - \mathbf{M}_a \widehat{\mathbf{Q}})^\top \right. \\
&\quad + (\mathbf{N}_a - \mathbf{M}_a \widehat{\mathbf{Q}}) (\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) \mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top - \mathbf{N}_a (\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) \mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top \\
&\quad - \mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top (\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) \mathbf{N}_a + \mathbf{N}_a (\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) \mathbf{N}_a \mathbf{V}(\widehat{\theta}_{[a]}^*) \mathbf{N}_a \\
&\quad \left. + \mathbf{N}_a \mathbf{V}(\widehat{\theta}_{[a]}^*) \mathbf{N}_a (\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) \mathbf{N}_a \right] + o(\eta_a^2) \\
&= \widehat{\theta}_{[a]}^* + \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} \left[\mathbf{N}_a \otimes \mathbf{N}_a \mathbf{V}(\widehat{\theta}_{[a]}^*) \mathbf{N}_a + \mathbf{N}_a \mathbf{V}(\widehat{\theta}_{[a]}^*) \mathbf{N}_a \otimes \mathbf{N}_a \right. \\
&\quad \left. - \mathbf{M}_a \widehat{\mathbf{Q}} \otimes \mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top - \mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top \otimes \mathbf{M}_a \widehat{\mathbf{Q}} \right] \text{vec}(\mathbf{V}(\widehat{\theta}_{[a]}) - \mathbf{V}(\widehat{\theta})) + o(\eta_a^2) \\
&= \widehat{\theta} + \widehat{\theta}_{[a]}^* - \widehat{\theta} + \left(\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top} \left[\mathbf{N}_a \otimes \mathbf{N}_a \mathbf{V}(\widehat{\theta}_{[a]}^*) \mathbf{N}_a + \mathbf{N}_a \mathbf{V}(\widehat{\theta}_{[a]}^*) \mathbf{N}_a \otimes \mathbf{N}_a \right. \\
&\quad \left. - \mathbf{M}_a \widehat{\mathbf{Q}} \otimes \mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top - \mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top \otimes \mathbf{M}_a \widehat{\mathbf{Q}} \right] \mathbf{Z}^* (\widehat{\theta}_{[a]} - \widehat{\theta}) + o(\eta_a^2) \quad . \quad (2.31)
\end{aligned}$$

Ahora, si se multiplica ambos lados la ecuación (2.31) por $\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^*$, se tiene:

$$\left(\mathbf{Z}^{*\top} \mathbf{W}_a \mathbf{Z}^* \right) (\widehat{\theta}_{[a]} - \widehat{\theta}) = \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* (\widehat{\theta}_{[a]}^* - \widehat{\theta}) + o(\eta_a^2), \quad (2.32)$$

donde,

$$\begin{aligned}
\mathbf{W}_a &= \mathbf{N}_a \otimes \mathbf{N}_a - \mathbf{N}_a \otimes \mathbf{N}_a \mathbf{V}(\widehat{\theta}_{[a]}^*) \mathbf{N}_a - \mathbf{N}_a \mathbf{V}(\widehat{\theta}_{[a]}^*) \mathbf{N}_a \otimes \mathbf{N}_a \\
&\quad + \mathbf{M}_a \widehat{\mathbf{Q}} \otimes \mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top + \mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top \otimes \mathbf{M}_a \widehat{\mathbf{Q}} \quad .
\end{aligned}$$

Luego,

$$\mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \widehat{\theta}_{[a]}^* = \mathbf{Z}^{*\top} \text{vec}(\mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top) \quad \text{y} \quad \mathbf{N}_a \widehat{\mathbf{V}} \mathbf{N}_a = \mathbf{N}_a \quad .$$

Desde (2.32), se tiene que

$$\begin{aligned}
(\mathbf{Z}^{*\top} \mathbf{W}_a \mathbf{Z}^*)(\hat{\theta}_{[a]} - \hat{\theta}) &= \mathbf{Z}^{*\top} \text{vec}(\mathbf{M}_a \hat{\mathbf{r}} \hat{\mathbf{r}}^\top \mathbf{M}_a^\top) - \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \mathbf{Z}^* \hat{\theta} + o(\eta_a^2) \\
&= \mathbf{Z}^{*\top} \text{vec}(\mathbf{M}_a \hat{\mathbf{r}} \hat{\mathbf{r}}^\top \mathbf{M}_a^\top) - \mathbf{Z}^{*\top} (\mathbf{N}_a \otimes \mathbf{N}_a) \text{vec}(\mathbf{V}(\hat{\theta})) + o(\eta_a^2) \\
&= \mathbf{Z}^{*\top} \left[\text{vec}(\mathbf{M}_a \hat{\mathbf{r}} \hat{\mathbf{r}}^\top \mathbf{M}_a^\top) - \text{vec}(\mathbf{N}_a) \right] + o(\eta_a^2) \\
&= \mathbf{Z}^{*\top} \text{vec}(\mathbf{M}_a \hat{\mathbf{r}} \hat{\mathbf{r}}^\top \mathbf{M}_a^\top - \mathbf{N}_a) + o(\eta_a^2) \quad .
\end{aligned} \tag{2.33}$$

Descartando $o(\eta_a^2)$ en las ecuaciones (2.23) y (2.33), y usando $\bar{\beta}_{[a]}$ y $\bar{\theta}_{[a]}$ se concluye la demostración del Teorema 2.2.

Si se observa la primera ecuación de (2.16) desde el Teorema 2.2 se tiene un término extra comparado con $\tilde{\beta}_{[a]}$ en el Teorema 2.1. Esto se debe a la estimación de θ cuando se encuentra asociado a los valores influyentes, los cuales son removidos. Así, las medidas de diagnóstico bajo la aproximación Tipo II son dadas por:

$$\begin{aligned}
C_\alpha^{II}(\hat{\beta}) &= C_\alpha^I(\hat{\beta}) + \Delta C_\alpha(\hat{\beta}) \quad \text{y} \\
C_\alpha^{II}(\hat{\theta}) &= \frac{\text{vec}^\top(\mathbf{M}_a \hat{\mathbf{r}} \hat{\mathbf{r}}^\top \mathbf{M}_a^\top - \mathbf{N}_a) \mathbf{G}_a \hat{\mathbf{V}}^{*-1} \mathbf{G}_a \text{vec}(\mathbf{M}_a \hat{\mathbf{r}} \hat{\mathbf{r}}^\top \mathbf{M}_a^\top - \mathbf{N}_a)}{2} \quad ,
\end{aligned} \tag{2.34}$$

donde $\mathbf{G}_a = \mathbf{Z}^* \left(\mathbf{Z}^{*\top} \mathbf{W}_a \mathbf{Z}^* \right)^{-1} \mathbf{Z}^{*\top}$, $C_\alpha^I(\hat{\beta})$ es dado en la ecuación (2.14), y

$$\begin{aligned}
\Delta C_\alpha(\hat{\beta}) &= (\bar{\theta}_{[a]} - \hat{\theta})^\top \mathbf{Z}^* (\mathbf{M}_a \hat{\mathbf{r}} \otimes \mathbf{M}_a)^\top \hat{\mathbf{P}} \\
&\quad \times \left[(\mathbf{M}_a \hat{\mathbf{r}} \otimes \mathbf{M}_a) \mathbf{Z}^{*\top} (\bar{\theta}_{[a]} - \hat{\theta}) + 2\mathbf{D}_a (\mathbf{D}_a^\top \hat{\mathbf{Q}} \mathbf{D}_a)^{-1} \mathbf{D}_a^\top \hat{\mathbf{r}} \right] \quad .
\end{aligned}$$

Note que $C_\alpha^{II}(\hat{\beta})$ se compone de dos partes. La primera parte es la medida reprimida de $\hat{\beta}$ cuando el parámetro aleatorio θ es fijo en $\hat{\theta}$. La segunda parte $\Delta C_\alpha(\hat{\beta})$ es la variación de la medida de la estimación para β debido a la estimación de θ cuando las observaciones se encuentran asociadas a α y estas son removidas. Si $\hat{\theta}_{[a]} \approx \hat{\theta}$, entonces $C_\alpha^{II}(\hat{\beta})$ reduce a $C_\alpha^I(\hat{\beta})$.

Al calcular las aproximaciones del Tipo I y Tipo II en medidas reprimidas, se encuentra con la complicación de lidiar con algunas matrices complicadas, tales como \mathbf{Z}^* , \mathbf{V}^{*-1} , $(\mathbf{M}_a \hat{\mathbf{r}} \otimes \mathbf{M}_a)^\top \mathbf{Z}^*$, y $\mathbf{Z}^{*\top} \mathbf{W}_a \mathbf{Z}^*$. A veces, el computador facilita los cálculos pero, en general en modelos jerárquicos, la matriz \mathbf{V} puede ser calculada como $\mathbf{V}(\theta) = \sum_{i=1}^R \mathbf{A}_i \theta_i$, donde \mathbf{A}_i tiene la misma dimensión de \mathbf{V} y es simétrica, y R tiene dimensión igual a θ . Así, $\mathbf{Z}^* = (\text{vec}(\mathbf{A}_1), \dots, \text{vec}(\mathbf{A}_R))$ y algunas de las matrices en la ecuación (2.16) pueden ser descompuestas en matrices simples. Por ejemplo, en la ecuación

(2.16), la matriz $(\mathbf{M}_a \widehat{\mathbf{r}} \otimes \mathbf{M}_a)^\top \mathbf{Z}^*$ de dimensión $\mathbf{N} \times \mathbf{R}$ y la j -ésima columna es denota así,

$$(\mathbf{M}_a \widehat{\mathbf{r}} \otimes \mathbf{M}_a)^\top \text{vec}(\mathbf{A}_j) = \mathbf{M}_a^\top \mathbf{A}_j \mathbf{M}_a \widehat{\mathbf{r}} \quad .$$

$\mathbf{Z}^{*\top} \text{vec}(\mathbf{M}_a \widehat{\mathbf{r}} \widehat{\mathbf{r}}^\top \mathbf{M}_a^\top - \mathbf{N}_a)$ en la segunda ecuación de (2.16), el cual es un vector de dimensión $\mathbf{R} \times 1$ y su j -ésimo elemento se denota de la siguiente manera:

$$\text{vec}^\top(\mathbf{A}_j) \text{vec}(\mathbf{M}_a \mathbf{r} \mathbf{r}^\top \mathbf{M}_a^\top - \mathbf{N}_a) = \text{tr}(\mathbf{M}_a \mathbf{r} \mathbf{r}^\top \mathbf{M}_a^\top \mathbf{A}_j - \mathbf{N}_a \mathbf{A}_j) \quad .$$

De manera similar, $\mathbf{Z}^{*\top} \mathbf{W}_a \mathbf{Z}^*$ es una matriz de orden $\mathbf{R} \times \mathbf{R}$, y el (i, j) -ésimo elemento es dado por:

$$\begin{aligned} \text{vec}^\top(\mathbf{A}_j) \mathbf{W}_a \text{vec}(\mathbf{A}_k) &= \text{tr} \left(\mathbf{A}_j \mathbf{N}_a \mathbf{A}_k \mathbf{N}_a - 2 \mathbf{A}_j \mathbf{V}(\tilde{\theta}_{[a]}) \mathbf{N}_a \mathbf{A}_j \mathbf{N}_a \right. \\ &\quad \left. + 2 \mathbf{A}_j \mathbf{M}_a \widehat{\mathbf{Q}} \mathbf{A}_k \mathbf{M}_a \mathbf{r} \mathbf{r}^\top \mathbf{M}_a^\top \right) \quad . \end{aligned}$$

Se ha mostrado el diagnóstico para el primer nivel del modelo jerárquico propuesto. Ahora, se mostrará el diagnóstico para el segundo nivel.

2.1.1. Detección y diagnóstico de medidas influyentes para el segundo nivel.

Se utilizará un modelo jerárquico de dos niveles para ilustrar el uso de la detección de un grupo de datos que influyen el análisis. Se considerará el caso de la extracción de datos en la aproximación tipo I.

Definición 2.4. Sean a_i los índices para las observaciones en la i -ésima unidad del segundo nivel, entonces la extracción de la i -ésima unidad en el segundo nivel se expresa de la siguiente manera:

$$\begin{aligned} C_{a_i}^I(\hat{\beta}) &= \hat{\mathbf{r}}_i^\top \hat{\mathbf{Q}}_{ii}^{-1} \hat{\mathbf{P}}_{ii} \hat{\mathbf{Q}}_{ii}^{-1} \hat{\mathbf{r}}_i \quad \text{y} \\ C_{a_i}^I(\hat{\theta}) &= \frac{\hat{\mathbf{r}}_i^{*\top} \hat{\mathbf{Q}}_{ii}^{*-1} \hat{\mathbf{P}}_{ii}^* \hat{\mathbf{Q}}_{ii}^{*-1} \hat{\mathbf{r}}_i^*}{2}, \end{aligned} \quad (2.35)$$

donde,

$$\begin{aligned} \hat{\mathbf{r}}_i &= \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{e}}_i, \\ \hat{\mathbf{e}}_i &= \mathbf{y}_i - \mathbf{X}_i \hat{\beta}, \\ \hat{\mathbf{P}}_{ii} &= \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i (\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}_i^\top \hat{\mathbf{V}}_i^{-1}, \\ \hat{\mathbf{Q}}_{ii} &= \hat{\mathbf{V}}_i^{-1} - \hat{\mathbf{P}}_{ii}, \\ \hat{\mathbf{r}}_i^* &= \hat{\mathbf{V}}_i^{*-1} \hat{\mathbf{e}}_i^*, \\ \hat{\mathbf{e}}_i^* &= \mathbf{Y}_i^{**} - \mathbf{Z}_i^* \hat{\theta}, \\ \hat{\mathbf{P}}_{ii}^* &= \hat{\mathbf{V}}_i^{*-1} \mathbf{Z}_i^* (\mathbf{Z}_i^{*\top} \hat{\mathbf{V}}_i^{*-1} \mathbf{Z}_i^*)^{-1} \mathbf{Z}_i^{*\top} \hat{\mathbf{V}}_i^{*-1}, \\ \hat{\mathbf{Q}}_{ii}^* &= \hat{\mathbf{V}}_i^{*-1} - \hat{\mathbf{P}}_{ii}^*, \\ \mathbf{V}_i^* &= \mathbf{V}_i \otimes \mathbf{V}_i, \end{aligned}$$

y \mathbf{Z}_i^* satisface la condición $\text{vec}(\mathbf{V}_i) = \mathbf{Z}_i^* \theta$.

La primera ecuación de (2.35) corresponde al caso en que se extrae el i -ésimo caso para los coeficientes estimados de la regresión en un modelo con errores correlacionados, mientras que la segunda ecuación es un nuevo caso de extracción de medidas para la estimación de parámetros aleatorios para el modelo jerárquico. Se demostrará la segunda ecuación de (2.35).

DEMOSTRACIÓN

Note que,

$$\begin{aligned} \mathbf{D}_{a_i} &\cong D_i \quad , \\ &\cong (0_{m_i \times n_1}, \dots, \mathbf{I}_{m_i}, \dots, 0_{m_j \times n_n})^\top \quad . \end{aligned}$$

es una matriz de orden $N \times m_i$, y

$$\begin{aligned} \mathbf{V} &= \text{diag}(\mathbf{V}_1, \dots, \mathbf{V}_n) \quad , \\ &= \sum_j \mathbf{D}_j \mathbf{V}_j \mathbf{D}_j^\top \quad . \end{aligned}$$

Suponga que \mathbf{Z}_j^* satisface la condición $\text{vec}(\mathbf{V}_j) = \mathbf{Z}_j^* \theta$. Entonces,

$$\begin{aligned} \mathbf{Z}^* \theta &= \text{vec}(\mathbf{V}) \quad , \\ &= \sum_j (\mathbf{D}_j \otimes \mathbf{D}_j) \text{vec}(\mathbf{V}_j) \\ &= \sum_j (\mathbf{D}_j \otimes \mathbf{D}_j) (\mathbf{Z}_j^* \theta) \quad , \end{aligned}$$

y,

$$\mathbf{Z}^* = \sum_j (\mathbf{D}_j \otimes \mathbf{D}_j) \mathbf{Z}_j^* .$$

Note que $\mathbf{N}_{a_i} = \widehat{\mathbf{V}}^{-1} - \widehat{\mathbf{V}}^{-1} \mathbf{D}_i \widehat{\mathbf{V}}_i \mathbf{D}_i^\top \widehat{\mathbf{V}}^{-1}$ y $\mathbf{D}_j^\top \mathbf{N}_{a_i} \mathbf{D}_k = \widehat{\mathbf{V}}_j^{-1}$ para $j = k \neq i$ y $j = k = 0$. Por otro lado, se tiene que,

$$\begin{aligned} \mathbf{Z}^{*\top} (\mathbf{N}_{a_i} \otimes \mathbf{N}_{a_i}) \mathbf{Z}^* &= \sum_{j \neq i} \mathbf{Z}_j^{*\top} \widehat{\mathbf{V}}_j^{*-1} \mathbf{Z}_j^* \\ &= \mathbf{Z}^{*\top} \widehat{\mathbf{V}}^{*-1} \mathbf{Z}^* - \mathbf{Z}_i^{*\top} \widehat{\mathbf{V}}_i^{*-1} \mathbf{Z}_i^* \\ &\quad \text{y} \\ \mathbf{Z}^{*\top} (\mathbf{N}_{a_i} \otimes \mathbf{N}_{a_i}) \mathbf{Y}^{**} &= \sum_{j \neq i} \mathbf{Z}_j^{*\top} \widehat{\mathbf{V}}_j^{*-1} \mathbf{Y}_j^{**} \\ &= \mathbf{Z}^{*\top} \widehat{\mathbf{V}}^{*-1} \mathbf{Y}^{**} - \mathbf{Z}_i^{*\top} \widehat{\mathbf{V}}_i^{*-1} \mathbf{Y}_i^{**} \quad . \end{aligned}$$

Entonces,

$$\begin{aligned}
\tilde{\theta}_{[a_i]} &= (\mathbf{Z}^{*\top} \widehat{\mathbf{V}}^{*-1} \mathbf{Z}^* - \mathbf{Z}_i^{*\top} \widehat{\mathbf{V}}_i^{*-1} \mathbf{Z}_i^*)^{-1} (\mathbf{Z}^{*\top} \widehat{\mathbf{V}}^{*-1} \mathbf{Y}^{**} - \mathbf{Z}_i^{*\top} \widehat{\mathbf{V}}_i^{*-1} \mathbf{Y}_i^{**}) \\
&= \left[(\mathbf{Z}^{*\top} \widehat{\mathbf{V}}^{*-1} \mathbf{Z}^*)^{-1} + (\mathbf{Z}^{*\top} \widehat{\mathbf{V}}^{*-1} \mathbf{Z}^*)^{-1} \mathbf{Z}_i^{*\top} \widehat{\mathbf{V}}_i^{*-1} \widehat{\mathbf{Q}}_{ii}^{*-1} \widehat{\mathbf{V}}_i^{*-1} \mathbf{Z}_i^* \right. \\
&\quad \left. \times (\mathbf{Z}^{*\top} \widehat{\mathbf{V}}^{*-1} \mathbf{Z}^*)^{-1} \right] (\mathbf{Z}^{*\top} \widehat{\mathbf{V}}^{*-1} \mathbf{Y}^{**} - \mathbf{Z}_i^{*\top} \widehat{\mathbf{V}}_i^{*-1} \mathbf{Y}_i^{**}) \\
&= \widehat{\theta} - (\mathbf{Z}^{*\top} \widehat{\mathbf{V}}^{*-1} \mathbf{Z}^*)^{-1} \mathbf{Z}_i^{*\top} \widehat{\mathbf{V}}_i^{*-1} \widehat{\mathbf{Q}}_{ii}^{*-1} \mathbf{V}_i^{*-1} \widehat{\mathbf{e}}_i^* , \tag{2.36}
\end{aligned}$$

donde

$$\begin{aligned}
\widehat{\mathbf{e}}_i^* &= \mathbf{Y}_i^{**} - \mathbf{Z}_i^* \widehat{\theta} , \\
\widehat{\mathbf{Q}}_{ii}^* &= \widehat{\mathbf{V}}_i^{*-1} - \widehat{\mathbf{P}}_{ii}^* , \\
\widehat{\mathbf{P}}_{ii}^* &= \widehat{\mathbf{V}}_i^{*-1} \mathbf{Z}_i^* (\mathbf{Z}^{*\top} \widehat{\mathbf{V}}^{*-1} \mathbf{Z}^*)^{-1} \mathbf{Z}_i^{*\top} \widehat{\mathbf{V}}_i^{*-1} .
\end{aligned}$$

Por lo tanto,

$$C_{\alpha_i}^I(\widehat{\theta}) = \frac{\widehat{\mathbf{r}}_i^{*\top} \widehat{\mathbf{Q}}_{ii}^{*-1} \widehat{\mathbf{P}}_{ii}^* \widehat{\mathbf{Q}}_{ii}^{*-1} \widehat{\mathbf{r}}_i^*}{2} , \tag{2.37}$$

donde $\widehat{\mathbf{r}}_i^* = \mathbf{V}_i^{*-1} \widehat{\mathbf{e}}_i^*$.

Siguiendo el enfoque estándar en el análisis de influencia, una medida adecuada de influencia de la unidad i en la estimación del parámetro fijo puede definirse como:

$$h_i^{(2)} = \frac{\sum_j \widehat{p}_{i,jj}}{\widehat{v}_i^{jj}} , \tag{2.38}$$

donde $\widehat{p}_{i,jj}$ y \widehat{v}_i^{jj} son los i -ésimos elementos ubicados en la diagonal de las matrices $\widehat{\mathbf{P}}_{ii}$ y $\widehat{\mathbf{V}}_i^{-1}$, respectivamente.

La motivación de esta definición es que $\widehat{p}_{i,jj}/\widehat{v}_i^{jj}$ representan la ponderación general de la j -ésima observación en la i -ésima unidad para los parámetros fijos. (ver 2.40))

El estudio de influencias de observaciones individuales de primer nivel, deja a $a = \{i_j\}$ como la conexión de la j -ésima observación en la i -ésima unidad, y dejan a $\mathbf{D}_a = \mathbf{d}_{i_j}$ vector de orden $\mathbf{N} \times \mathbf{1}$ con i -ésimos elementos iguales a 1 y el resto igual a cero. Entonces $\mathbf{D}_a^\top \widehat{\mathbf{Q}} \mathbf{D}_a = \widehat{v}_i^{jj} - \widehat{p}_{i,jj}$, el j -ésimo elemento de la diagonal de la matriz $\widehat{\mathbf{Q}}_{ii}$ y $\mathbf{D}_a \widehat{\mathbf{r}} = \widehat{r}_{ij}$, el elemento j -ésimo del vector $\widehat{\mathbf{r}}_i$.

Ahora,

$$\begin{aligned}
C_{ij}^I(\hat{\beta}) &= \left(\frac{\hat{r}_{ij}}{\sqrt{\hat{v}_i^{jj} - \hat{p}_{i,jj}}} \right)^2 \frac{\hat{p}_{i,jj}}{\hat{v}_i^{jj} - \hat{p}_{i,jj}} \\
&= t_{ij}^2 \frac{h_{ij}^{(1)}}{1 - h_{ij}^{(1)}},
\end{aligned} \tag{2.39}$$

donde $t_{ij} = \frac{\hat{r}_{ij}}{\sqrt{\hat{v}_i^{jj} - \hat{p}_{i,jj}}}$ es el residual standarizado y

$$h_{ij}^{(1)} = \frac{\hat{p}_{i,jj}}{\hat{v}_i^{jj}}. \tag{2.40}$$

La ecuación (2.40) se define como el valor de influencia generalizado para los parámetros fijos del primer nivel. Se observa que $h_{ij}^{(1)}$ resulta ser la misma influencia generalizada (Christensen, et. al., 1992) para modelos lineales mixtos. Una fórmula simple para las medidas eliminadas para $\hat{\theta}$ no parece alcanzable.

Para el caso de aproximación Tipo II, no hay fórmulas simples. Cuando la influencia para las unidades del segundo nivel es de interés, las matrices ligadas a a_i son \mathbf{N}_{a_i} , \mathbf{M}_{a_i} y $\mathbf{N}_{a_i} \hat{\mathbf{r}}$. En este caso, \mathbf{N}_{a_i} es una matriz bloque diagonal con el i -ésimo bloque nulo y el j -ésimo bloque $\hat{\mathbf{V}}_j^{-1}$ para $j \neq i$. El (k, j) -ésimo bloque de la matriz $\mathbf{M}_{a_i} \hat{\mathbf{r}}$, descrita como:

$$\begin{aligned}
\mathbf{M}_{a_i} \hat{\mathbf{Q}} &= \hat{\mathbf{Q}}_{kl} - \hat{\mathbf{Q}}_{ki} \hat{\mathbf{Q}}_{ii}^{-1} \hat{\mathbf{Q}}_{il}, \quad \text{con } k, l = 1, \dots, n \quad . \\
\mathbf{M}_{a_i} \hat{\mathbf{r}} &= (\tilde{\mathbf{r}}_{i1}^\top, \dots, \tilde{\mathbf{r}}_{ni}^\top)^\top,
\end{aligned}$$

donde

$$\tilde{\mathbf{r}}_{ji} = \hat{\mathbf{r}}_j - \hat{\mathbf{Q}}_{ji} \hat{\mathbf{Q}}_{ji}^{-1} \hat{\mathbf{r}}_i \quad .$$

Además,

$$\hat{\mathbf{Q}}_{ji} = \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j (\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}_i^\top \hat{\mathbf{V}}_j^{-1} \quad \text{para } i \neq j.$$

El siguiente capítulo mostrará de manera práctica la implementación de la técnica de diagnóstico expuesta en este capítulo.

Capítulo 3

Aplicación

La prueba SIMCE es un tipo de prueba que mide el conocimiento educacional adquirido durante los primeros años de educación.(SIMCE, 2012) Los puntajes utilizados en este trabajo, corresponden al rendimiento medio obtenido por cada establecimiento perteneciente a la región de Valparaíso en el año 2011. Específicamente, se consideró el puntaje obtenido por los alumnos de octavo año básico en la prueba SIMCE de matemáticas. Para este estudio, también se tomó en cuenta la localidad del establecimiento y el tipo de administración de éste.

El SIMCE es el sistema nacional de evaluación de resultados de aprendizaje del Ministerio de Educación de Chile(SIMCE, 2012). Su propósito principal es contribuir al mejoramiento de la calidad y equidad de la educación, informando sobre el desempeño de los estudiantes en diferentes áreas de aprendizaje del Currículum Nacional, y relacionando estos desempeños con el contexto escolar y social en que aprenden(Cadiz, 2006).

Las pruebas SIMCE evalúan el logro de los objetivos fundamentales y contenidos mínimos Obligatorios (OF-CMO) del Marco Curricular vigente en diferentes sectores de aprendizaje, a través de una medición que se aplica a nivel nacional, una vez al año a los estudiantes que cursan un determinado nivel educacional. Hasta el año 2005, la aplicación de las pruebas se alternó entre 4^o año básico, 8^o año básico y 2^o año Medio. A partir del año 2006, se evalúan todos los años a 4^o años básico y, se alternan 8^o año básico y 2^o Medios. Desde el año 2010 se aplica cada dos años la evaluación del sector Inglés en 3^o Medios, y todos los años una evaluación muestral del sector de Educación Física en 8^o años básico (Cadiz, 2006).

A través del tiempo se ha escuchado o leído que existe una gran brecha en el rendimiento académico de los alumnos que estudian en un establecimiento municipal versus uno particular pagado. Desde hace mucho tiempo, el Ministerio de Educación ha tratado e implantado nuevos objetivos y metas para que el sistema educacional entregue el mismo conocimiento a todos los alumnos indiferentemente al establecimiento al cual pertenezca. Sin embargo las políticas de estado no han dado los resultados esperados Cadiz, 2006.

La prueba SIMCE de matemática está formada por 45 preguntas, siendo 12 de estas relacionadas con habilidades matemáticas básicas, y las siguientes 33 relacionadas con el currículo de matemáticas de octavo año básico. La prueba SIMCE de lenguaje, por otro lado, está formada por 45 preguntas, de las cuales 12 están relacionadas con habilidades del lenguaje básico, y las siguientes 33 están relacionadas con el currículo que corresponde al lenguaje de octavo año básico.

Sobre la base de los argumentos anteriores, algunas interrogantes se han creado a través de este documento con el objetivo de examinar la aplicación de los modelos jerárquicos con coeficientes aleatorios. Estas interrogantes son las siguientes:

- Con respecto al rendimiento de los establecimientos de la región de Valparaíso, ¿Cuántos establecimientos educacionales de la región varían significativamente en su rendimiento medio en la prueba de matemáticas del SIMCE?
- ¿Es la diferencia en el rendimiento de la prueba SIMCE de matemáticas estadísticamente significativa entre las comunas de la región de Valparaíso?
- ¿Es la diferencia en el rendimiento de la prueba SIMCE de matemáticas estadísticamente significativa considerando el tipo de administración y la localidad del establecimiento?

La base de datos fue obtenida desde la página web <http://buscador.simce.cl>, de manera gratuita. Los datos se encuentran organizados de manera jerárquica, de la siguiente manera:

COMUNA: Representa la comuna donde se encuentra el i-ésimo establecimiento, las cuales son:

ID	Comuna	ID	Comuna
1	Algarrobo	20	Olmué
2	Cabildo	21	Panquehue
3	Calera	22	Papudo
4	Calle Larga	23	Petorca
5	Cartagena	24	Puchuncaví
6	Casablanca	25	Putendo
7	Catemu	26	Quillota
8	Concón	27	Quilpué
9	El Quisco	28	Quintero
10	El Tabo	29	Rinconada
11	Hijuelas	30	San Antonio
12	Isla de Pascua	31	San Esteban
13	Juan Fernandez	32	San Felipe
14	La Cruz	33	Santa María
15	La Ligua	34	Santo Domingo
16	Limache	35	Valparaíso
17	LlayLlay	36	Villa Alemana
18	Los Andes	37	Viña del Mar
19	Nogales	38	Zapallar

ESTABLECIMIENTO: Unidad de estudio en la cual se procederá a realizar el análisis a través de un modelo jerárquico con coeficiente aleatorio, de aquí en adelante llamado establecimiento.

ADM: Corresponde al tipo de administración del establecimiento educacional, el que se clasifica en tres categorías: Municipal (estos establecimientos son administrados por los municipios y son subvencionados por el gobierno), Particular Subvencionado (establecimientos administrados por empresas particulares y subvencionados por parte del gobierno) y Particular Pagado (establecimientos administrados y subvencionados en su totalidad por empresas particulares).

N°	Dependencia
MU	Municipal
PP	Particular Pagado
PS	Particular Subvencionado

RU-UR: Variable relacionada con las áreas de población en que se encuentra cada establecimiento. Su clasificación está dada por: (1) zona rural: establecimientos ubicadas en zonas rurales de la región de Valparaíso, (2) zona urbana: establecimientos ubicados en ciudades o pueblos de la región de Valparaíso y por último (0) zona no definida: establecimiento no definido dentro de las clasificaciones anteriores (1 o 2) ante el Ministerio de Educación de Chile.

N°	Ruralidad
RU	Rural
UR	Urbano
ND	No disponible

MATEMÁTICAS: Puntaje promedio de la prueba SIMCE de matemáticas rendida por los alumnos de octavo año básico en los respectivos establecimientos educacionales de la región de Valparaíso.

LENGUAJE: Puntaje promedio de la prueba SIMCE de lenguaje rendida por los alumnos de octavo año básico en los respectivos establecimientos educacionales de la región de Valparaíso.

Una vez presentadas las variables consideradas en el modelo estadístico, se procederá a revisar los datos de manera descriptiva y luego de manera inferencial en la siguiente sección.

3.1. Análisis de datos

El análisis de los efectos de las comunas en el puntaje medio de la prueba SIMCE de matemática de octavo año básico, se divide en dos tipos: el primero en un análisis descriptivo y, el segundo en un análisis inferencial. El primero permite desarrollar un análisis exploratorio de las variables en estudio y el segundo análisis, permite hacer inferencia a través de los modelos jerárquicos lineales.

Para el análisis inferencial, a su vez, se desarrollarán dos tipos de modelos jerárquicos lineales; el modelo ONE-WAY ANOVA o nulo (MN) y, el modelo con intercepto aleatorio (MJ). El primero, permite tener información acerca de la variación existente de los resultados en la prueba SIMCE de matemática que se encuentran dentro de, y entre las comunas de la región de Valparaíso. El segundo modelo propuesto, permite analizar el efecto de las comunas con respecto al logro de los establecimientos en la prueba SIMCE de matemática. En este modelo el parámetro de intersección que el primer modelo se considera como parámetro fijo, es considerado como variable aleatoria.

Este tipo de modelos jerárquicos lineales permite evitar algunos problemas metodológicos, por ejemplo, permite examinar los datos sólo a nivel de establecimiento e ignorar el hecho de que los establecimientos se encuentran anidados en las comunas de la región de Valparaíso. Así, se evita tener problemas con el error estándar estimado, pues este podría ser demasiado pequeño. A continuación se mostrarán los resultados obtenidos del análisis de los datos.

3.1.1. Resultados

Los Análisis Descriptivos

En el año 2011, sólo alumnos pertenecientes a 855 establecimientos educacionales rindieron la prueba SIMCE de matemáticas de octavo año básico, establecidos en 38 comunas de la región de Valparaíso. De éstos, sólo 653 establecimientos se les registraron los respectivos puntajes.

El análisis descriptivo, en general, reveló que el puntaje medio en matemática fue de 253,37 puntos y la desviación estándar fue de 27,37 puntos, siendo el puntaje mínimo entre todos los establecimientos de la región de Valparaíso 185 y el máximo 338 puntos.

La siguiente Tabla 3.0, muestra la comparación entre los establecimientos educacionales que rindieron la prueba SIMCE de matemáticas y los que fueron registrados por el Ministerio de Educación.

Tabla 3.0: Comparación entre el número de Establecimientos Educacionales que rindieron la prueba SIMCE de matemáticas y los Establecimientos que fueron registrados por el Ministerio de Educación.

Tipo Ruralidad	Rendidos	Registrados
Urbano	658	558
Rural	187	90
No Def.	10	5
Total	855	653

En los establecimientos municipalizados, el puntaje medio en matemáticas de octavo año básico fue 237,37 puntos y la desviación estándar fue 17,36 puntos; la mínima puntuación fue 185 y la máxima 338 puntos. De aquí, se concluye que el puntaje medio en matemática de los establecimientos municipalizados fue menor al mismo regional.

En la Tabla 3.1, se observa que los establecimientos municipalizados de la región de Valparaíso ubicadas en zonas rurales tienen mejores resultados, en media, con respecto a los establecimientos ubicados en zonas urbanas ($240,04 > 235,94$). De 286 establecimientos educacionales que registraron puntajes medios en la región de Valparaíso, 201 están en zonas urbanas, 83 establecimientos se encuentran en zonas rurales y 2 no están definidos por el Ministerio de Educación.

Tabla 3.1: Puntajes de la prueba SIMCE del año 2011 de matemáticas por establecimientos municipales.

Variabes	Categorías	Media	Desviación estándar	Mín a Max	n
Ruralidad	Rural (1)	240,036	16,119	193 a 276	83
	Urbano (2)	235,935	16,596	185 a 295	201
	No Def.(0)	271,500	78,489	216 a 327	2

Lo más llamativo de la Tabla anterior (Tabla 3.0), es que el mejor puntaje medio en la prueba SIMCE de matemáticas se encuentra entre dos establecimientos municipalizados que no tienen definido

el tipo de ruralidad en que se encuentran.

La Tabla 3.2, muestra los establecimientos educacionales particulares pagados de la región de Valparaíso ubicados en zonas rurales, tienen mayor puntaje en media con respecto a los demás, a pesar de ser sólo dos establecimientos que tienen esta característica. Existen 48 establecimientos que son particulares pagados y se encuentran en el área urbana en la región de Valparaíso, su desviación estándar es 26,56 puntos, y el puntaje oscila entre los 224 y 338 puntos.

Tabla 3.2: Puntajes del SIMCE 2011 de matemática por establecimientos particular pagado.

Variabes	Categorías	Media	Desviación estándar	Mín a Máx	n
Ruralidad	Rural (1)	312,500	13,435	303 a 322	2
	Urbano (2)	297,917	26,561	224 a 338	48
	No Def. (0)	276,000	38,184	249 a 303	2

A continuación la Tabla 3.3, muestra el detalle de los puntajes de la prueba SIMCE de matemáticas de los establecimientos particulares subvencionados de la región de Valparaíso.

Tabla 3.3: Puntajes de la prueba SIMCE 2011 de matemáticas por establecimientos particular subvencionados.

Variabes	Categorías	Media	Desviación estándar	Mín a Máx	n
Ruralidad	Rural (1)	247,800	13,917	225 a 263	5
	Urbano (2)	260,916	23,605	208 a 317	309
	No Def. (0)	221,000	-	221 a 221	1

Los establecimientos particulares subvencionados ubicados en área urbana tienen mejor rendimiento medio en matemáticas en comparación a la media general de la región de Valparaíso ($260,92 > 253,37$) siendo 309 establecimientos que tienen un puntaje entre 208 a 317 puntos en el SIMCE de matemáticas, su desviación estándar es de 23,61 puntos, el rango del puntaje medio es de 109 puntos. Mientras que los establecimientos que se encuentran en zonas rurales o desconocidas por el Ministerio de Educación, que son tan sólo 6, tienden a tener una media inferior a la media general de la región de Valparaíso.

Como resumen, se observa que en octavo año básico, el rendimiento en matemáticas varía entre las distintas comunas de la región de Valparaíso.

A continuación se realizará el análisis inferencial el cual involucra modelos jerárquicos lineales.

Análisis Inferencial: Modelos Jerárquicos Lineales.

Usando como base los análisis descriptivos de las variables independientes expuestos anteriormente, no es posible concluir que estas diferencias en el rendimiento en matemáticas por establecimiento son estadísticamente significativas. El análisis inferencial permitirá determinar si estas diferencias son significativas, en particular en relación a los establecimientos educacionales municipalizados, particulares subvencionados y particulares pagados. Para lograr esto, se utilizarán los dos modelos jerárquicos lineales propuestos y, que a continuación se presentan.

Modelo ONE-WAY ANOVA.

Este modelo, es uno de los modelos jerárquicos más básicos el cual, entre otras cualidades, permite obtener información acerca de la media general y, la partición de la varianza total en el rendimiento en la prueba SIMCE de matemáticas de los establecimientos educacionales. Es decir, permite conocer cómo cambia la varianza estimada dentro de las comunas y entre las comunas del rendimiento medio de la prueba SIMCE de matemáticas. Además, este modelo permite examinar la hipótesis de que todas los establecimientos de la región de Valparaíso tienen el mismo rendimiento medio y, con respecto a los establecimientos educacionales, ¿Cuántas comunas varían su rendimiento medio significativamente en la prueba SIMCE de matemáticas? En este caso, el modelo lineal jerárquico propuesto es:

Nivel 1 (Nivel establecimientos):

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

Y_{ij} : Resultados del rendimiento de la prueba SIMCE de matemáticas de octavo año básico de los i -ésimos establecimientos educacionales en las j -ésimas comunas de la región de Valparaíso.

β_{0j} : Intersección de los logros en la prueba SIMCE de matemáticas de cada comuna de la región de Valparaíso.

$\varepsilon_{ij} \sim N(0, \sigma^2)$, donde σ^2 es la varianza a nivel de establecimientos educacionales.

Nivel 2 (Nivel de Comunas):

$$\beta_{0j} = \beta_{00} + u_{0j}$$

β_{0j} : Rendimiento medio de la prueba SIMCE en matemáticas de cada comuna de la región de Valparaíso.

β_{00} : Media general.

$u_{0j} \sim N(0, \sigma_{u0}^2)$: Error aleatorio en el que σ_{u0}^2 es la varianza del factor denominado nivel comunal.

Nótese, que el modelo propuesto no tiene predictores ya sea a primer como a segundo nivel. La Tabla 3.4, muestra los primeros resultados del modelo propuesto.

Tabla 3.4: Resultados del modelo nulo.

Efecto Fijo	Coficiente	Error Estándar	
Gran Media (Establecimiento), β_{00}	251,091	1,596	
Efecto Aleatorio	Componentes de Varianza	Error Estándar	p valor
Media Comunal, u_{0j}	9,679	1,154	< 0,0001
Efecto Nivel 1, e_{ij}	26,971	1,117	

En la Tabla 3.4, se observa que en la estimación del rendimiento medio en la prueba SIMCE de matemáticas por establecimiento educacional, la media general es 251,091 puntos, con un error estándar de 1,596 puntos. Con respecto a las estimaciones de las componentes de la varianza, se estima que la varianza dentro de los establecimientos educacionales o a primer nivel es de 26,971 puntos, mientras que la variabilidad entre el rendimiento de la prueba SIMCE de matemáticas y las comunas es 9,679 puntos. Basados en las estimaciones anteriores, es posible determinar que gran parte de la variación está entre los establecimientos educacionales.

Es necesario determinar la proporción de la varianza entre los establecimientos educacionales, para esto, existe la correlación intraclase ρ . Los componentes de la varianza estimada son sustituidos por la estimación de sus parámetros. Así, $\rho = 9,679^2 / (9,679^2 + 26,971^2) = 0,114$ puntos. Esto explica que aproximadamente el 11,41 % de la varianza del rendimiento medio de los establecimientos educacionales, se encuentra repartido entre las comunas de la región de Valparaíso.

Si la hipótesis principal es $\sigma^2 = 0$, es decir, se supone que todas las comunas tienen el mismo rendimiento medio en la prueba SIMCE de matemáticas. Entonces, la Tabla 3.4, muestra el valor p menor a 0,0001, lo cual con un nivel del 0,001 % estadísticamente es altamente improbable que las comunas tengan el mismo rendimiento medio.

Segundo Modelo Jerárquico Propuesto:

El segundo modelo jerárquico propuesto tiene en cuenta los factores: puntaje medio del i -ésimo establecimiento educacional de la prueba SIMCE de lenguaje, el tipo de administración que tiene cada establecimiento educacional en cada comuna y, la ubicación de cada establecimiento donde los alumnos rindieron la prueba SIMCE de octavo año básico en las distintas comunas la región de Valparaíso. Utilizando un caso particular de la ecuación (1.1) descrita en el Capítulo 1, el modelo jerárquico con intercepto aleatorio propuesto es:

Nivel 1 (Nivel de establecimientos educacionales):

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

También, es necesario usar la ecuación (1.2) para realizar el segundo nivel del modelo jerárquico propuesto. Entonces:

Nivel 2 (Nivel de Comunas):

$$\beta_{0j} = \beta_{00} + \beta_{01}(URB - RUR)_{0j} + \beta_{02}(ADM)_{0j} + \beta_{03}(LENGUAJE)_{0j} + u_{0j}$$

β_{0j} : Rendimiento medio en la prueba SIMCE de matemáticas en cada comuna de la región de Valparaíso.

β_{00} : Intercepto.

β_{01} : Efecto del área donde se encuentra ubicado el establecimiento educacional en la región de Valparaíso.

β_{02} : Efecto del tipo de administración que tiene cada establecimiento educacional en la región de Valparaíso.

β_{03} : Rendimiento medio de cada establecimiento educacional en la j -ésima comuna de la región de Valparaíso.

$u_{0j} \sim N(0, \sigma_{0j}^2)$: Error aleatorio, donde σ_{0j}^2 corresponde a la varianza residual a nivel comunal.

El resultado del primer nivel del modelo MJ, es decir, sin considerar el nivel comuna de la región de Valparaíso, está dado por:

Tabla 3.5: Resultados del modelo MJ a nivel de establecimientos.

Efecto Fijo	Coefficiente	Error Estándar	p valor
Modelo para las medias comunales			
Intercepto, β_{00}	-52,683	11,055	0,001
UR-RU, β_{01}	0,167	0,048	< 0,0001
ADM, β_{02}	0,296	0,032	< 0,0001
LENGUAJE, β_{03}	0,755	0,021	< 0,0001

y su análisis de varianza correspondiente es:

Fuente de Variación	Suma Cuadrática	g.l	Cuadrado Medio
Modelo	400269,611	3	133423,204
Residuos	83332,541	648	128,599
Total	483602,152	651	133551,803

La Tabla 3.5, muestra que las variables independientes como el tipo de administración, el tipo de administración del establecimiento educacional, la ubicación comunal de los establecimientos educacionales en la región de Valparaíso y, el puntaje medio de la prueba SIMCE en lenguaje, son altamente significativos en el modelo propuesto, es decir, los factores causan algún tipo de efecto en el puntaje medio de la prueba SIMCE de matemáticas. Considerando 648 grados de libertad, la variabilidad explicada por el modelo en el primer nivel, es de un 83,7% aproximadamente ($R^2 = 0,827$).

A continuación, en el siguiente paso, se considerará la ubicación comunal de los establecimientos educacionales.

La Tabla 3.6, presenta los resultados del segundo modelo jerárquico propuesto.

Tabla 3.6: Resultados del segundo modelo jerárquico propuesto.

Efecto Fijo	Coefficiente	Error Estándar	p valor
Modelo para las medias de los establecimientos			
Intercepto, β_{00}	-48,169	16,739	0,004
UR-RU, β_{01}	0,137	0,073	0,060
ADM, β_{02}	0,274	0,032	< 0,0001
LENGUAJE, β_{03}	0,782	0,026	< 0,0001
Efecto Aleatorio	Componentes de Varianza	Error Estándar	p valor
Media Comunal, u_{0j}	3,645	0,389	< 0,0001
Efecto Nivel 1, e_{ij}	11,042	0,376	

La variabilidad residual del segundo modelo jerárquico (MJ) bajo el rendimiento medio de la prueba SIMCE en matemáticas, entre las comunas es más pequeño que el modelo jerárquico nulo (MN), así se contrasta $3,65_{MJ} < 11,23_{MN}$. Por otra parte, mediante el uso de las estimaciones de las varianzas obtenidas a través de los modelos propuestos MJ y MN, es posible observar la proporción de la varianza estimada del rendimiento medio entre las comunas de la región de Valparaíso, lo que se explica por las variables predictoras en el segundo nivel, las cuales son estadísticamente significativas. Así, la correlación intraclase bajo el MN era del 11,4% y ahora considerando las variables de administración de los establecimientos educacionales, zonas en donde se encuentran los establecimientos y rendimiento en la prueba SIMCE de lenguaje en la región de Valparaíso, es de 9,8%, esto ha marcado un descenso de la variabilidad del 0,9%. También, es posible observar que el tipo de administración de los establecimiento educacional es altamente significativa en el puntaje medio de la prueba SIMCE de matemáticas, sin embargo, a pesar de que el tipo de área donde se encuentran ubicados los establecimientos educacionales es medianamente significativo (valor $p = 0,060$).

Como último punto antes de realizar el diagnóstico, es importante para determinar cuánto rendimiento tienen los establecimientos educacionales cuando el rendimiento medio en lenguaje, la administración de los establecimientos educacionales y el tipo de ubicación de los establecimientos educacionales se encuentran considerados. La Tabla 5, muestra que su valor $p < 0,001$, lo cual es altamente significativo. Usando estos resultados, la hipótesis nula ($\sigma_{0j}^2 = 0$) es, fácilmente rechazado. Por lo tanto, es posible decir que la variación significativa en el rendimiento en matemáticas a nivel comunal ha sido explicada.

Es posible encontrar algunos datos que hacen perder información o dar información engañosa de la real situación en que se encuentra el análisis en cuestión. Estos datos pueden ser detectados y luego aislados para no ser procesados dentro de la base de datos, ya que éstos nos cambian la real situación y se puede llegar a dar una solución incorrecta a las hipótesis expuestas.

3.2. Diagnóstico.

Como ya fue realizado el MN con anterioridad, se procede a observar una de las hipótesis importantes dentro de los modelos jerárquicos, la cual es la independencia. Existen 653 establecimientos educacionales de la región de Valparaíso que el Ministerio de Educación registró con alumnos que rindieran la prueba SIMCE de matemáticas y 654 establecimientos educacionales que el Ministerio de Educación registró con alumnos que rindieran la prueba SIMCE de lenguaje, lo cual implica que asintóticamente éstas variables o factores cumplen con la normalidad y, son independientes. Pero, la Figura 3.3 muestra la normalidad residual del rendimiento medio de la prueba SIMCE de matemáticas de los establecimientos educacionales de la región de Valparaíso.

Figura 3.3: Normalidad Residual.

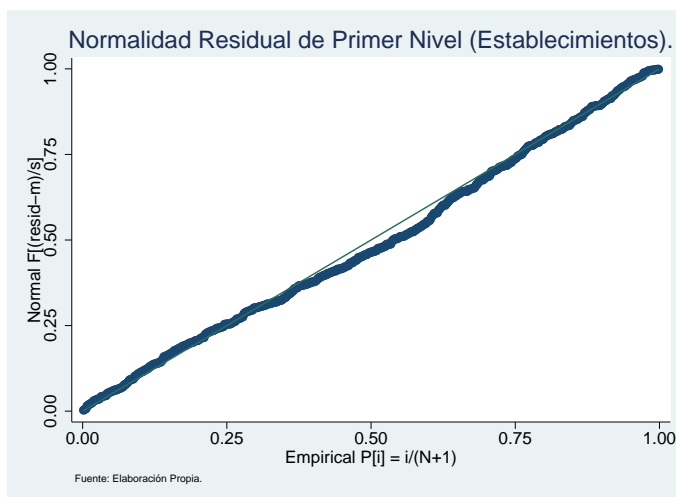
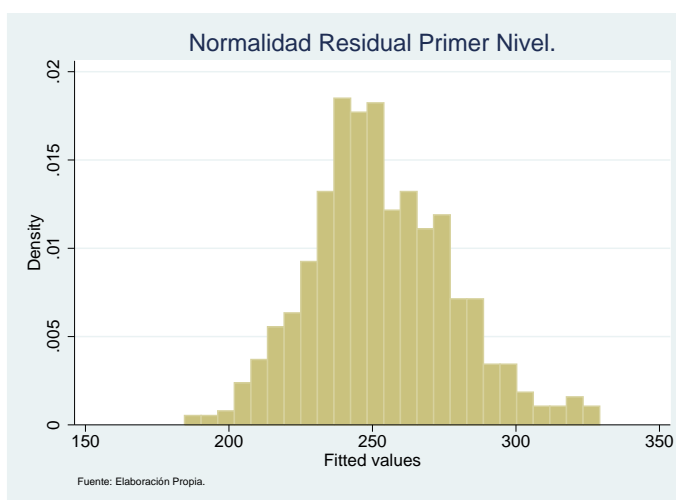


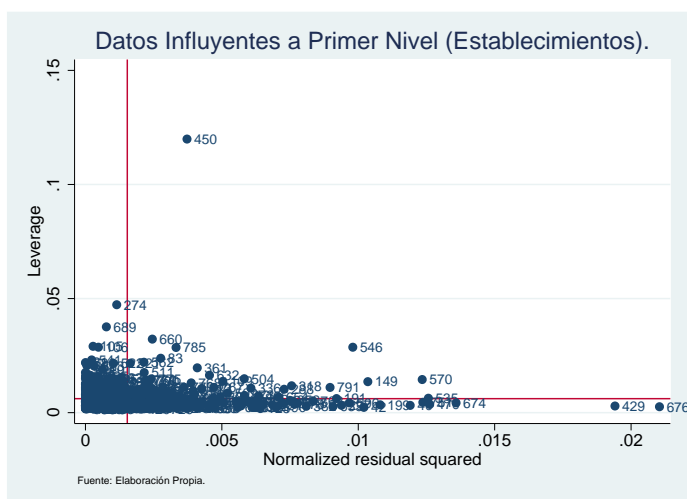
Figura 3.3.1: Histograma: Normalidad Residual del Primer Nivel.



La Figura 3.3, muestra la normalidad residual que se ajusta en gran parte a la recta diagonal de la gráfica, pero se observa un leve descenso en el centro de la gráfica, es decir, tiene forma de parábola. Sin embargo, la prueba de normalidad de Shapiro-Wilk muestra que efectivamente estadísticamente hablando, no existe evidencia suficiente para decir que existe normalidad residual en el modelo, con un valor p menor al 1% (Valor $p = 0,00058$). La Figura 3.1, muestra los residuos con una tendencia o asimetría hacia la izquierda. Esto implica que no se cumple con los supuestos del modelo, pero pueden existir datos atípico u observaciones influenciadas que modifican o alteran la normalidad residual. Para esto es necesario realizar el análisis correspondiente al diagnóstico de primer nivel al MJ, es decir, mediante gráficas y la técnica para detectar medidas influyentes de distancia de Cook's, para así obtener mejores resultados en la regresión. Las herramientas adecuadas para enfrentar este tipo de problemas se encuentran en el Capítulo 2, en la sección de “Diagnóstico para el primer nivel” utilizando las ecuaciones (2.39) y (2.40).

A continuación se presenta las medidas influyentes para el modelo MJ (parte fija o primer nivel).

Figura 3.4: Datos Influyentes Normalidad Residual Primer Nivel.



Aquí, es posible observar que los casos influyentes en una primera impresión, son los datos que se encuentra en la posición 450 (dato influyente) y los datos ubicados en la posición 429 y 676 (residuos muy altos), siendo éstos los correspondientes al rendimiento medio 327 , 313 y 295 puntos en la prueba SIMCE de matemáticas. Entonces, estos datos se extraerán y luego se aplicará el test de normalidad para los residuos y obtener conclusiones a nivel de establecimientos educativos. Entonces, la Figura 3.5 mostrará la normalidad residual sin los datos influyentes o atípicos para el modelo.

Figura 3.5: Normalidad Residual Primer Nivel, sin datos influyentes .

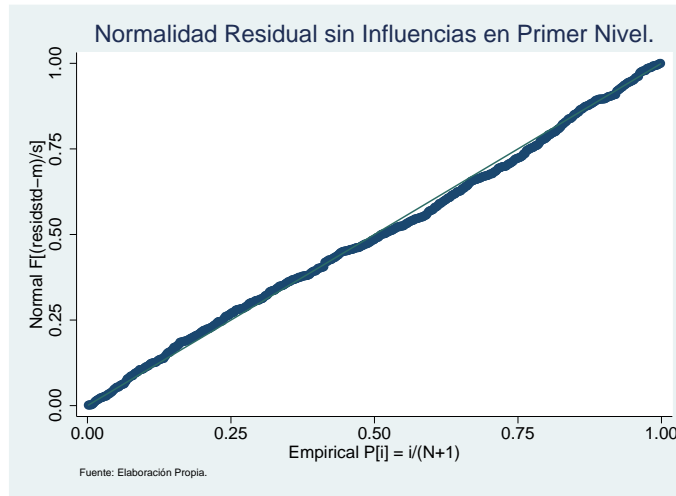
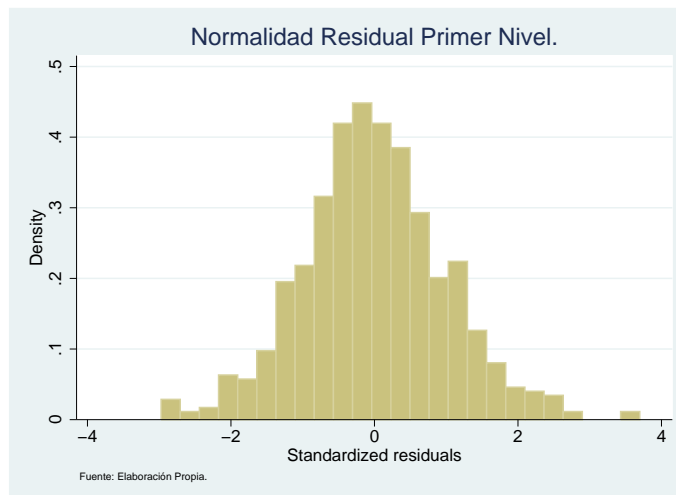


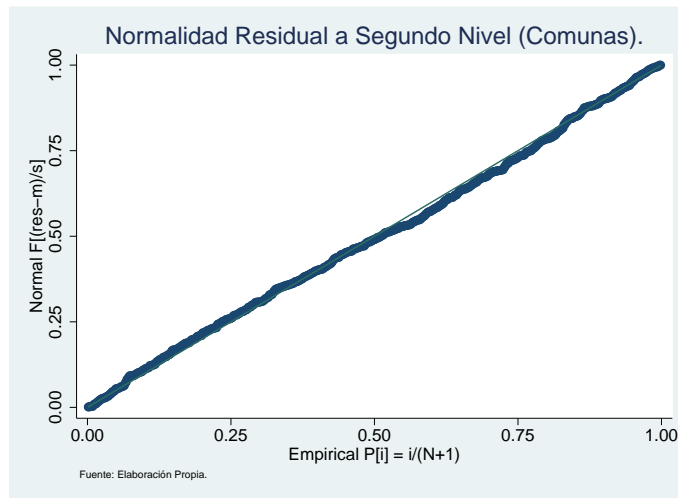
Figura 3.5.1: Histograma: Normalidad Residual sin datos influyentes .



Las Figuras 3.5, muestra que los residuos han mejorado, es decir, hay evidencia estadística suficiente para decir que existe normalidad residual en el modelo, esto lo afirma el valor p igual a 0,031 y la Figura 5.1, expone un histograma de residuos centrados en cero. Al parecer, los datos que anteriormente se extrayeron, fueron motivo suficiente para dar una respuesta errónea a las hipótesis planteadas. Los valores de los residuos, distancia de cook's correspondiente a cada dato se encuentra detallado en el anexo A.

Ahora, es posible determinar datos influyentes en el segundo nivel (comunas) del modelo. Primero, se verificó la normalidad residual a nivel de comunas, la cual es mostrada en la Figura 3.6.

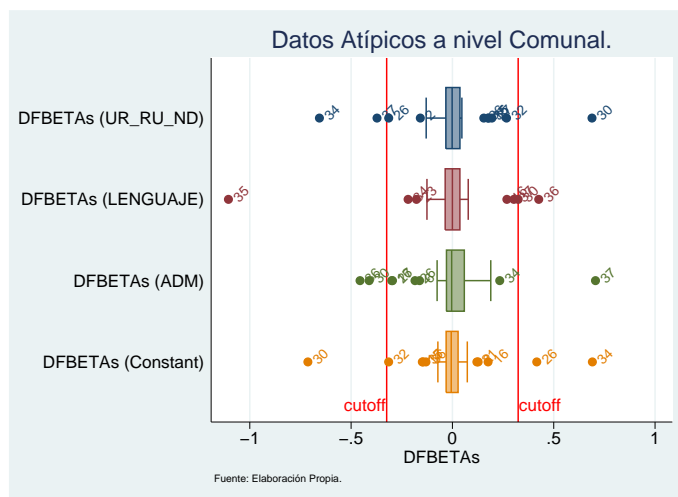
Figura 3.6: Normalidad Residual Segundo Nivel.



La Figura 3.6, muestra que los residuos en el segundo nivel del mododelo MJ cumple con la normalidad residual, y esto lo afirma el estadístico de prueba es 0,99 y un valor p igual a 0,07.

A continuación, la siguiente Figura (Figura 3.7) muestra los datos atípicos que se encuentran influenciando el nivel de las comunas, bajo el rendimiento medio de la prueba SIMCE de matemáticas a nivel comunal.

Figura 3.7: Datos Influyentes a nivel de comunas.



La Figura 3.7 muestra que si se consideran las distancias mayores a 0,5 y menores a $-0,5$ de la variable significativa para el modelo MJ la cual es área en donde se encuentra el establecimiento educacional, entonces las comunas que influyen la normalidad residual son efectivamente la números 35 (comuna de Valparaíso), 30 (comuna de San Antonio) y 34 (comuna de Santo Domingo) a pesar

de que los residuos son efectivamente normales, se considerarán las comunas correspondientes a los valores expuestos anteriormente. Si se extraen y se recalculan los residuos del modelo MJ (se muestran en la Figura 3.8) y aplicando un test de normalidad, es posible detectar que la normalidad residual se acentúa con mayor fuerza, y con evidencia estadísticamente suficiente de determina que efectivamente la

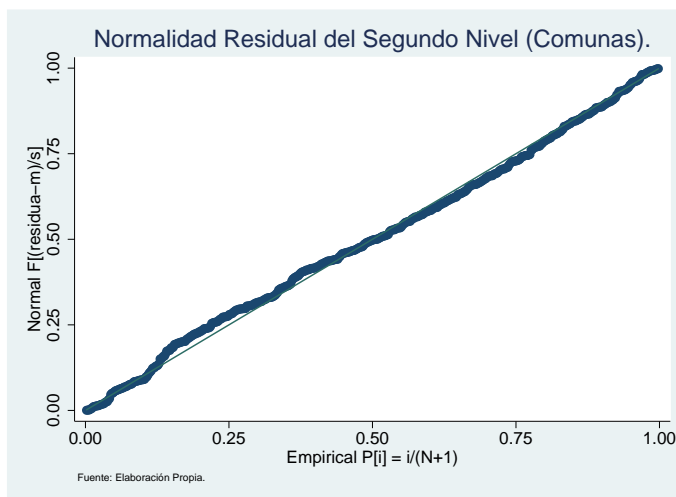
normalidad residual se cumple para el modelo MJ (valor p igual a 0,031) y la tabla (Tabla 3.6) muestra la mejora sustancial del modelo. Ya una vez aplicada la distancia de

Cook's en el primer nivel, es necesario aplicar DFBetas para el segundo nivel del modelo MJ, el cual se basa en la distancia de Cook's para el segundo nivel, éste se encuentra expuesto en el capítulo dos, sección 2.1.1, pero con el fin de ser utilizada e interpretada en el segundo nivel del modelo MJ. Así, los resultados fueron los siguientes:

Tabla 3.7: Resultados del modelo jerárquico con intercepto aleatorio sin Influencias a nivel Comunal.

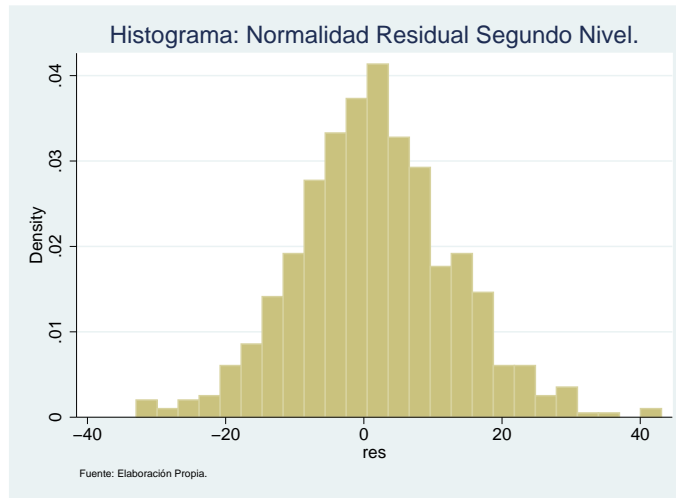
Comunas	DFB URB-RUR	DFB LENG	DFB ADM	DFB Const
35	0,179	-1,105	0,130	0,053
37	-0,372	0,303	0,706	-0,145
30	0,689	0,325	-0,411	-0,713
34	-0,656	-0,219	0,234	0,691
36	0,155	0,426	-0,456	-0,131
26	-0,314	0,078	-0,185	0,416

Figura 3.8: Normalidad Residual sin Influencias a Nivel de Comunas.



Considerando la Figura 3.8, ésta muestra que los residuos se ajustan en gran parte a la diagonal, pero, a pesar de que existe una leve curvatura ascendente en el comienzo de la gráfica, ésta no afecta la normalidad residual a nivel de las comunas de la región de Valparaíso.

Figura 3.8.1: Histograma: Normalidad Residual.



La Figura 3.8.1 muestra una centralización de los residuos alrededor del cero, formando una figura casi simétrica, confirmando lo anteriormente expuesto. A pesar de que se observa irregularidades en la normalidad residual del modelo, entonces, se mostrará a continuación los resultados del modelo jerárquico (MJ) sin influencias.

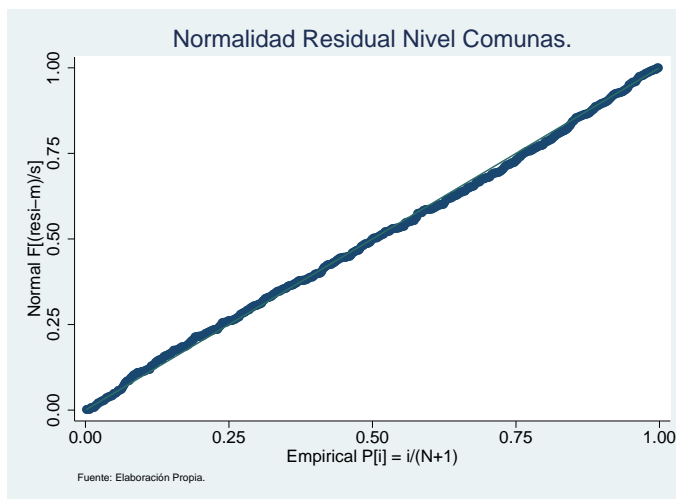
Tabla 3.9: Resultados del segundo modelo jerárquico sin Influencias a nivel Comunal.

Efecto Fijo	Coefficiente	Error Estándar	p valor
Modelo para media Establecimientos			
Intercepto, β_{00}	-48,762	13,855	< 0,0001
UR-RU, β_{01}	0,120	0,066	0,069
ADM, β_{02}	0,272	0,036	< 0,0001
LENGUAJE, β_{03}	0,804	0,019	< 0,0001
Efecto Aleatorio	Componentes de Varianza	Error Estándar	p valor
Media Comunal	3,790	0,426	< 0,0001
Efecto Nivel 1	10,660	0,439	

En la Tabla 3.8, se observa que al extraer las comunas influyentes en el MJ, la ubicación en donde se encuentra el establecimiento donde fue realizada la prueba SIMCE en la región de Valparaíso tiene relevancia en el modelo. La justificación de esto es que el nivel de influencia que tiene la variabilidad en entre los establecimientos considerando las distintas comunas en que se encuentran los establecimientos es del 11,22%. Esto implica que aproximadamente el 11,22% de la variabilidad del rendimiento de los establecimientos de la región de Valparaíso se encuentra repartido entre las comunas de la región de Valparaíso.

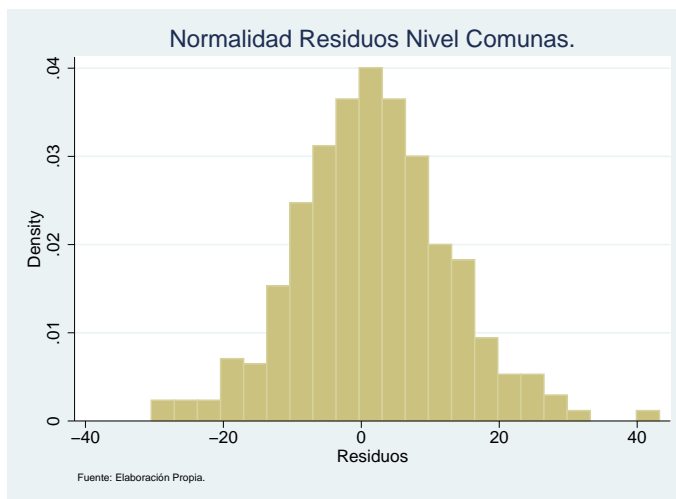
Así, siendo $\sigma^2 = 0$ la hipótesis principal, y el valor p es menor a 0,001, entonces se está en presencia de rechazar la hipótesis y decir que significativamente el rendimiento en media de los establecimientos que rindieron la prueba SIMCE de matemática, en las distintas comunas de la región de Valparaíso no son todas iguales.

Figura 3.9: Normalidad Residual sin Influencias a Nivel Comuna.



Es posible observar en la Figura 3.9, que la normalidad residual no ha sido afectada en gran parte por la extracción de algunas comunas. Sin embargo el test de bondad de ajuste arrojó un valor p igual a 0,08, lo cual en comparación a la Figura 3.8, se puede decir que ha mejorado en un 1%.

Figura 3.9.1: Histograma: Normalidad Residual sin Influencias.



La Figura 3.9.1, muestra una gran masa de datos centrada en cero, formando una casi perfecta simetría residual, lo cual confirma lo anteriormente obtenido. Para un mayor detalle, los residuos del segundo Nivel se encuentran en el cd.

3.3. Conclusión

Con base en los análisis anteriores, se observa que gran parte de la variación en el rendimiento medio en la prueba SIMCE de matemáticas está a nivel de establecimiento educacionales. Sin embargo, como variable regresora, siempre fue importante la ubicación del establecimiento donde se realizó la prueba SIMCE de matemática. El hecho de considerar factores que puedan ayudar de mejor manera la estimación del modelo, como lo fue el rendimiento medio de la prueba SIMCE en lenguaje, el tipo de administración de los establecimientos educacionales y el tipo de ubicación donde se encuentran los mismos en la región de Valparaíso, fueron útiles para mejorar la variabilidad del modelo ya que la correlación intraclase ha decrecido, se paso de un 11 % a un 9,8 %, esto provoca que las conclusiones obtenidas para las hipótesis expuestas sean las correctas y no se encuentren influenciadas por datos atípicos o infladas por la variabilidad escondida.

El modelo jerárquico con coeficientes aleatorios es muy utilizado en el área de educación y a través de la prueba SIMCE de octavo año básico fue posible comprobar que existen factores a considerar para controlar la variabilidad, que un modelo lineal general no tiene la suficiente razón para estimar de mejor manera los parámetros del modelo.

A pesar de no cumplirse en una primera instancia la normalidad residual del modelo jerárquico con coeficiente aleatorio, esto implicó que existían datos atípicos que influenciaban la normalidad residual, una vez detectados y extraídos, la normalidad residual y las conclusiones mejoraron.

Los factores considerados en el segundo modelo jerárquico son altamente significativos para el modelo, lo cual implica que efectivamente existe algún tipo de influencia en la respuesta, es más, el puntaje medio en la prueba SIMCE de matemáticas se encuentra directamente influenciada por el puntaje medio de la prueba SIMCE de lenguaje. También, el tipo de administración que tiene cada establecimiento educacional, la ubicación de cada establecimiento dentro de las comunas de la región de Valparaíso son influencias relevantes para el puntaje medio de la prueba SIMCE de matemáticas. En futuros estudios, será necesario identificar otros factores que puedan afectar el rendimiento en la prueba SIMCE de matemáticas a nivel de establecimientos y a nivel superior. También es importante el desarrollo de otros instrumentos de medición que permitan adquirir mayor información. Es decir, la prueba SIMCE necesita incluir otras fuentes de información.

Bibliografía

- [1] Aparicio, A. y Morera, M. (2007). La conveniencia del análisis multinivel para la investigación en salud: una aplicación para Costa Rica. *Rev. Electrónica Semestral, Población y Salud en Mesoamérica*. Costa Rica, 4(2).
- [2] Casal, J. y Mateu, E. (2003). Tipos de Muestreo. *Rev. Epidemiología y Medicina preventiva*, 1, 3 – 7 .
- [3] Cadiz, J. (2006). Reanalyzing fourth grade math student achievement in Chile: applying Hierarchical Linear Models (HLMs). *Electronic Journal of Educational Research, Assessment and Evaluation*, 12(1),.
- [4] Christensen, R.; Pearson, L. and Johnson W. (1992) Case - Deletion Diagnostics for Mixed Models. *American Statistical Association and American Society for Quality. Technometrics*, 34(1), 38 – 45.
- [5] De la Cruz, F. (2008). Modelos Multinivel. *Revista Peruana de Epidemiología*, 12(3).
- [6] Galea, M. (1990). Técnicas de Diagnóstico en Regresión Lineal. *Sociedad Chilena de Estadística*, 7(7), 23 – 44.
- [7] Goldstein, H. (1989) Restricted unbiased iterative generalized least squares estimation. *Biometrika*. London, 76, 45 – 51.
- [8] Goldstein, H. (1995). *Multilevel Statistical Models (2^{da} Edición)*. Edward Arnold. London: Wiley. New York.
- [9] Goldstein, H. (1999). *Multilevel Statistical Models*. London: Institute of Education, Multilevel Models Project.
- [10] Goldstein, H. y Rasbash, J. (1992). Efficient computational procedures for the estimation of parameters in multilevel models based on iterative generalized least squared. *Computational Statistics & Data Analysis*, 13(1), 63 – 71.
- [11] Goldstein, H.; Kounali, D. y Robinson, A. (2008). Modelling measurement errors and category misclassifications in multilevel models. *Statistical Modelling*, 8(3), 234 – 261.

- [12] Hengjian, C.; Kai, W. y Lixing, Z. (2004). Estimation in mixed effects model with error in variables. *Journal of Multivariate Analysis*, 91, 53 – 73.
- [13] Lair, N. y Ware, J. (1982) Random - effects models for longitudinal data. *A Journal of the International Biometric Society*, 38(4), 963 – 974.
- [14] Lei, S. y Gemai, C. (2008). Case Deletion Diagnostics in Multilevel Models. *Journal of multivariate analysis*, 99, 1860 – 1877.
- [15] Ojeda, M. y Torres, M. (1996). A Hierarchical Regression Model for a Cluster Stratified Sample. *Agrociencia*, 30(4), 587 – 596.
- [16] Oliver, J.; Rosel, J. y Jara, P. (2000). Modelos de Regresión Multinivel: aplicación en psicología escolar. *Psicothema. Universidad de Oviedo*. 12(3), 487 – 494.
- [17] Rasbash, J.; Steele, F.; Browne, W. y Goldstein, H. (2004). *A users Guide to MLwiN version 2.10*. Centre for Multilevel Modelling, University of Bristol, London.
- [18] Sánchez, E. y Ocaña, R. (1999). Los Modelos Multinivel o la importancia de la jerarquía. *Gaceta Sanitaria*. 13(5), 391 – 398.
- [19] Searle, S.; Casella, G. y McCulloch, C. (1992). *Variance Components*. New York: John Wiley & Sons Inc.
- [20] Stata. (2010) *StataCorp. Stata Statistical Software: Release 12.0. College Station: Stata Corporation*.
- [21] Velasco, F. (2006). Modelo Lineal General Jerárquico. *Ciencias Básicas UJAT*. México, 4(2), 20 – 28.
- [22] Woodhouse, G., Yang, M., Goldstein, H. y Rasbash J. (1996). Adjusting for Measurement Error in Multilevel Analysis. *Journal of the Royal Statistical Society. Series A*, 159(2), 201 – 212.