

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

MODELIZACIÓN MULTINIVEL CON ESTIMACIÓN BAYESIANA
MEDIANTE CADENAS DE MÁRKOV

PROYECTO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO MATEMÁTICO

MARIO FERNANDO COBA CISNEROS
mariofercoba@gmail.com

Director: Dr. Luis Horna
luis.horna@epn.edu.ec

2014

DECLARACIÓN

Yo MARIO FERNANDO COBA CISNEROS, declaro bajo juramento que el trabajo aquí escrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Mario Fernando Coba Cisneros

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por MARIO FERNANDO COBA CISNEROS, bajo mi supervisión

Dr. Luis Horna
DIRECTOR

AGRADECIMIENTOS

De manera especial a mis padres que han sido pilar fundamental para la culminación de este ciclo; a mis hermanos que siempre me han apoyado y a mi compañera de vida Gabby.

Mario Fernando Coba Cisneros

Para mi hijo Emilio.

Índice de Contenido

Lista de Figuras	iv
Lista de Tablas	vii
Resumen	1
Abstract	2
1 Modelos jerárquicos de dos niveles	3
1.1 Introducción	3
1.2 Tratamiento estadístico y agrupación de datos	5
1.2.1 Correlación intraclase	5
1.2.2 Varianza intra grupos y entre los grupos	7
1.3 Modelos de componentes de la varianza	9
1.3.1 Modelo con ordenada aleatoria, sin variable explicativa	9
1.3.2 Modelo con ordenada aleatoria, con una variable explicativa	12
1.3.3 Modelo con ordenada aleatoria y varias variables explicativas	14
1.4 Modelos con pendientes aleatorias	16
1.4.1 Modelo con ordenada y pendiente aleatorias	16
1.4.2 Modelo con ordenada aleatoria, pendiente aleatoria y una variable de nivel 2	19
1.4.3 Modelo lineal jerárquico de nivel 2	20
1.5 Especificación de los modelos	22
1.5.1 Modelización de variabilidad dentro de los grupos	24
1.5.2 Modelización de variabilidad entre grupos	25
1.6 Heteroscedasticidad	26
1.6.1 Heteroscedasticidad a nivel 1	26
1.6.2 Heteroscedasticidad a nivel 2	27
1.7 Estimación	28
1.7.1 Residuos	28
1.7.2 Método de Mínimos Cuadrados Generalizados Iterativo (MCGI)	28

1.8	Pruebas de Hipótesis	31
1.8.1	Pruebas para los parámetros fijos	31
1.8.2	Prueba de desviación	33
1.8.3	Pruebas de Heteroscedasticidad	34
1.8.4	Estudio de residuos	36
2	Modelos de tres niveles y estructuras jerárquicas más complejas	37
2.1	Modelo de nivel 3	37
2.1.1	Introducción	37
2.1.2	Modelo de nivel 3 con pendiente aleatoria	38
2.2	Modelos de ecuaciones estructurales multinivel	39
2.2.1	Introducción	39
2.2.2	Análisis de senderos	40
2.2.3	Análisis factorial	41
2.2.4	Modelos de ecuaciones estructurales	43
2.2.5	Análisis Factorial Confirmatorio (AFC)	50
3	Estimación Bayesiana usando Cadenas de Márkov Monte Carlo	52
3.1	Introducción	52
3.2	Distribuciones a priori por defecto	55
3.3	Muestreador de Gibbs	55
3.4	Muestreador de Metropolis - Hastings (MH)	59
4	Aplicación de los modelos	62
4.1	Introducción	62
4.2	Modelos de nivel 2	63
4.2.1	Modelo con ordenada aleatoria, sin variable explicativa	64
4.2.2	Modelo con ordenada aleatoria, con una variable explicativa	79
4.2.3	Modelo con ordenada y pendiente aleatorias	88
4.2.4	Modelo lineal jerárquico de nivel 2	99
4.3	Modelos de tres niveles y estructuras jerárquicas más complejas	108
4.3.1	Modelo lineal jerárquico de nivel 3	108
4.3.2	Modelo de ecuaciones estructurales multinivel	114
5	Conclusiones y Recomendaciones	118
5.1	Conclusiones	118
5.2	Recomendaciones	119
	Referencias	120

Anexos	122
A Demostración que la varianza observada entre grupos es insesgada	123
B Certificación de la Dirección Provincial de Educación de Imbabura	125
C Pruebas aplicadas	127
D Instituciones Educativas	139
E Modelos multinivel descartados	140

Lista de Figuras

1.1	Modelo con ordenada aleatoria.	13
1.2	Diagrama de especificación del modelo.	23
1.3	Diagrama de selección de efectos en el modelo.	25
1.4	Gráfico de la distribución t de Student.	32
1.5	Gráfico de la distribución χ^2	33
2.1	Diagrama de Senderos.	40
2.2	Modelo Estructural	45
2.3	Ejemplo de Análisis Factorial Confirmatorio (AFC.)	50
4.1	Resultados del modelo con ordenada aleatoria, sin variable explicativa con estimación de MCGI.	64
4.2	Resultados del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.	66
4.3	Resultados de la primera convergencia del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.	66
4.4	Resultados del primer diagnóstico de $\hat{\beta}_0$ del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.	67
4.5	Resultados de la segunda convergencia del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.	68
4.6	Resultados del segundo diagnóstico de $\hat{\beta}_0$ del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.	69
4.7	Resultados de la tercera convergencia del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.	69
4.8	Resultados del tercer diagnóstico de $\hat{\beta}_0$ del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.	70
4.9	Resultados del diagnóstico de $(\hat{\sigma}_{u_0}^2)$ del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.	70
4.10	Resultados del modelo con ordenada aleatoria, sin variable explicativa y sin término u_{0j}	73

4.11 Residuos de nivel 1 del modelo con ordenada aleatoria, sin variable explicativa.	74
4.12 Residuos de nivel 1 por colegio del modelo con ordenada aleatoria, sin variable explicativa.	75
4.13 Gráfico de probabilidad normal de los residuos estandarizados de nivel 1 del modelo con ordenada aleatoria, sin variable explicativa. . .	76
4.14 Residuos de nivel 2 del modelo con ordenada aleatoria, sin variable explicativa.	77
4.15 Residuos de nivel 2 en función del número de colegios del modelo con ordenada aleatoria, sin variable explicativa.	78
4.16 Gráfico de probabilidad normal de los residuos estandarizados de nivel 2 del modelo con ordenada aleatoria, sin variable explicativa. . .	78
4.17 Prueba de normalidad de K-S de los residuos de nivel 2 del modelo con ordenada aleatoria, sin variable explicativa.	79
4.18 Variable Prueba mat vs. Variable Prueba raz.	80
4.19 Resultados del modelo con ordenada aleatoria, con una variable explicativa con estimación de MCGI.	80
4.20 Regresiones de 12 colegios según el modelo con ordenada aleatoria, con una variable explicativa.	82
4.21 Resultados del modelo con ordenada aleatoria, con una variable explicativa sin término u_{0j} con estimación de MCGI.	84
4.22 Residuos de nivel 1 del modelo con ordenada aleatoria, con una variable explicativa.	86
4.23 Gráfico de probabilidad normal de los residuos estandarizados de nivel 1 del modelo con ordenada aleatoria, con una variable explicativa	86
4.24 Gráfico de probabilidad normal de los residuos estandarizados de nivel 2 del modelo con ordenada aleatoria, con una variable explicativa.	87
4.25 Prueba de normalidad de (K-S) de los residuos de nivel 2 del modelo con ordenada aleatoria, con una variable explicativa.	87
4.26 Modelo con ordenada y pendiente aleatorias.	88
4.27 Resultados del modelo con ordenada y pendiente aleatorias con estimación de MCGI.	89
4.28 Regresiones de 12 colegios según el modelo con ordenada y pendiente aleatorias.	91
4.29 Modelo con ordenada aleatoria y pendiente fija.	92
4.30 Modelo con ordenada y pendiente aleatorias.	92
4.31 Función de la varianza del modelo con ordenada y pendiente aleatorias.	93
4.32 Residuos de nivel 1 del modelo con ordenada y pendiente aleatorias .	96

4.33 Gráfico de normalidad de los residuos de nivel 1 del modelo con ordenada y pendiente aleatorias.	97
4.34 Gráfico de normalidad de los residuos de nivel 2 del modelo con ordenada y pendiente aleatorias.	97
4.35 Prueba de normalidad de (K-S) de los residuos u_{0j} de nivel 2 del modelo con ordenada y pendiente aleatorias.	98
4.36 Prueba de normalidad de (K-S) de los residuos \hat{u}_{1j} de nivel 2 del modelo con ordenada y pendiente aleatorias.	98
4.37 Modelo lineal jerárquico de nivel 2.	99
4.38 Resultados del modelo lineal jerárquico de nivel 2 con estimación de MCGI.	100
4.39 Función de la varianza del modelo lineal jerárquico de nivel 2.	103
4.40 Resultados del modelo lineal jerárquico de nivel 2, sin el término e_{3ij}	105
4.41 Gráfico de normalidad de los residuos de nivel 1 del modelo lineal jerárquico de nivel 2.	106
4.42 Gráfico de normalidad de los residuos de nivel 2 del modelo lineal jerárquico de nivel 2.	107
4.43 Prueba de normalidad de (K-S) de los residuos \hat{u}_{0j} de nivel 2 del modelo lineal jerárquico de nivel 2.	107
4.44 Prueba de normalidad de (K-S) de los residuos \hat{u}_{0j} de nivel 2 del modelo lineal jerárquico de nivel 2.	108
4.45 Modelo lineal jerárquico de nivel 3.	109
4.46 Resultados del modelo lineal jerárquico de nivel 3 con estimación de MCGI.	110
4.47 Análisis Factorial Confirmatorio a nivel estudiante-colegio.	114
4.48 Resultados del AFC a nivel estudiante	116
4.49 Resultados del AFC a nivel colegio	116

Lista de Tablas

4.1	Descripción de las variables.	63
4.2	Comparación de estimaciones CMMC-Gibbs con dos distribuciones a priori y MCGI del modelo con ordenada aleatoria, sin variable explicativa.	71
4.3	Comparación de estimaciones CMMC-Gibbs con dos distribuciones a priori, CMMC-MH con dos distribuciones a priori y MCGI del modelo con pendiente aleatoria, sin variable explicativa.	72
4.4	Comparación de estimaciones CMMC-Gibbs con dos distribuciones a priori, CMMC-MH con dos distribuciones a priori y MCGI del modelo con ordenada aleatoria, con una variable explicativa.	83
4.5	Comparación de estimaciones CMMC-Gibbs con dos distribuciones a priori, CMMC-MH con dos distribuciones a priori y MCGI del modelo con ordena y pendiente aleatorias.	94
4.6	Cuadro de predicciones del modelo lineal jerárquico de nivel 2.	101
4.7	Comparación de estimaciones CMMC-Gibbs con dos distribuciones a priori y MCGI del modelo lineal jerárquico de nivel 2.	104
4.8	Cuadro de predicciones del modelo lineal jerárquico de nivel 3.	111
4.9	Comparación de estimaciones CMMC-Gibbs con dos distribuciones a priori, CMMC-MH con dos distribuciones a priori y MCGI del modelo lineal jerárquico de nivel 3.	112

Resumen

Los modelos multinivel son herramientas completas y precisas en las cuales se puede basar decisiones, ya que permiten estudiar las relaciones dentro de los grupos, entre los grupos y estimar la variación de cada una de dichas relaciones. En este trabajo, los modelos multinivel son desarrollados desde un modelo sencillo sin variables explicativas (modelo vacío) hasta un modelo de estructura jerárquica de nivel 3 con variables explicativas.

La estimación de los parámetros es sumamente importante para el ajuste del modelo. Los modelos bayesianos para datos multinivel son tratados en este trabajo utilizando métodos de Cadenas de Markov Monte Carlo (MCMC) con el muestreador de Gibbs y el muestreador de Metropolis-Hastings.

Los modelos de ecuaciones estructurales son una poderosa herramienta de análisis que se utiliza para determinar los fenómenos existentes, por medio de una relación de causa y efecto, de variables latentes y no observadas. Estos modelos se fusionan con los modelos multinivel resultando los modelos de ecuaciones estructurales multinivel, presentados de una manera general debido a la profundidad y complejidad del tema. Los modelos de ecuaciones estructurales multinivel son utilizados cuando las unidades de observación forman una jerarquía de grupos anidados y algunas variables de interés deben ser medidas por una variable no observada o por un conjunto de variables no observadas.

Palabras clave: modelos multinivel, modelo vacío, métodos de Cadenas de Markov Monte Carlo, muestreadores de Gibbs y Metropolis-Hastings, modelos de ecuaciones estructurales, relación de causa y efecto, variables latentes y no observadas, modelos de ecuaciones estructurales multinivel.

Abstract

Multilevel models are complete and precise tools in which one can base decisions, due to the fact that they allow to study relations within groups, between groups, and estimate the variation of each one of those relations. In this work, multilevel models are developed from a simple model without any explanatory variable (empty model), to a level 3 hierarchical structure model with explanatory variables.

The parameter estimation is extremely important for the model fit. Bayesian models for multilevel data are treated in this work using Markov chain Monte Carlo (MCMC) methods, with the Gibbs sampling and the Metropolis-Hastings sampling.

Structural equations models are a powerful analysis tool that is used in order to determine the existing phenomena, through a relation of cause and effect, of unobserved and latent variables. These models are fused with multilevel models resulting multilevel structural equations models, presented in a general way because of the depth and complexity of the subject. Multilevel structural equations models are used when the observation units form a hierarchy of nested groups and some variables of interest must be measured by one unobserved variable or by a set of unobserved variables.

Key words: Multilevel models, empty model, Markov chain Monte Carlo methods, Gibbs sampling and Metropolis-Hastings algorithm, structural equations model, relation of cause and effect, unobserved and latent variables, multilevel structural equations model.

CAPÍTULO 1

Modelos jerárquicos de dos niveles

1.1 Introducción

Existen muchos tipos de datos que se encuentran estructurados jerárquicamente, como los del ser humano con la herencia es una jerarquía natural, los abuelos, padres e hijos, tienen una estructura jerárquica, anidados o agrupados. Al referirnos a una jerarquía, hablamos de un conjunto de unidades agrupadas en diferentes niveles. De esta manera los hijos pueden ser las unidades del nivel 1 y el nivel 2 los padres, en una estructura de 2 niveles. Otro ejemplo son los estudiantes que pueden ser las unidades del nivel 1 dentro de las escuelas que estarían en el nivel de 2.

Los estudios en animales y humanos de acuerdo con la herencia es una jerarquía natural en el que crías se agrupan dentro de las familias. Descendientes de los mismos padres tienden a ser más parecidos en sus características físicas y mentales de los individuos elegidos al azar de la población en general. Como son los niños de la misma familia que tienden a ser pequeños, tal vez porque sus padres son pequeños o debido a un ambiente no favorable.

Las principales aplicaciones de estructuras jerárquicas se encuentran en las ciencias médicas y sociales, pero las técnicas son aplicables a nivel general. A medida que se desarrolle la teoría y las técnicas con ejemplos, vamos a ver cómo un reconocimiento adecuado de estas jerarquías nos permite obtener más respuestas satisfactorias a preguntas importantes.

La existencia de las jerarquías no es accidental ni ignorable. Por ejemplo, estudiantes con motivaciones similares o aptitudes se agrupan en las escuelas o universidades altamente selectivas. En otros casos, los grupos pueden surgir por razones asociadas a las características de los individuos, tales como la asignación de los

niños a las escuelas primarias, o la asignación de pacientes a diferentes clínicas. Una vez que los grupos se han establecido, incluso si su establecimiento es efectivamente al azar, a menudo tenderán a diferenciarse. Esta diferenciación implica que el grupo y sus miembros influyen y son influidos por la pertenencia al grupo.

Un estudio conocido e influyente de los estilos de enseñanza utilizados con niños de educación primaria se llevó a cabo en la década de los setenta por Bennett [9], afirmó que los niños expuestos a estilos formales de enseñanza de la lectura, mostraron más progresos que los que no lo eran. Los datos fueron analizados mediante técnicas tradicionales de regresión múltiple que sólo reconocen a los niños como unidades individuales de análisis e ignoró sus agrupaciones dentro de los maestros y en las clases. Los resultados mostraron diferencias estadísticamente significativas. Posteriormente, Aitkin [9] demostró que cuando en el análisis se toma en cuenta la agrupación de los niños en clases, las diferencias significativas desaparecen.

Este es un ejemplo importante de análisis multinivel con datos de ciencias sociales. El problema en el estudio fue que los niños de cualquier clase, debido a que aprenden en conjunto, tienden a ser similares en su rendimiento. Es por esto que proporcionan menos información de lo que hubiera sido el caso, si el mismo número de estudiantes hubiesen sido formados por separado por los diferentes profesores, es decir, la unidad objetivo de estudio para fines de comparación debió haber sido el maestro, no el estudiante, ya que los estudiantes pueden proporcionar a cada maestro una estimación de la efectividad del maestro.

Modelos estructurados en los cuales, estudiantes se agrupan dentro de las escuelas tienen varias ventajas. En primer lugar, permite obtener estimaciones estadísticamente eficientes de los coeficientes de regresión. En segundo lugar, con la información de la agrupación permite corregir los errores estándar, intervalos de confianza y pruebas de significación, que generalmente serán más conservadores que estimaciones tradicionales que se obtienen ignorando la presencia de la agrupación.

Los datos multinivel representan de manera simultánea, relaciones dentro de los grupos y entre los grupos. Un modelo multinivel es el modelo lineal jerárquico que contiene coeficientes aleatorios, para datos estructurados jerárquicamente.

Se denominarán como *individuos* a las unidades de nivel 1 y como *grupos* a las

unidades de nivel 2. El número de individuos dentro del j-ésimo grupo viene dado por n_j . El número de grupos es N ; y el número total del tamaño de la muestra es $M = \sum_{j=1}^N n_j$.

El objetivo de los modelos lineales jerárquicos es construir un modelo que exprese a la variable dependiente en función de las variables explicativas, con la particularidad que la variable dependiente debe ser una variable del nivel 1 y las variables explicativas pueden estar definidas en cualquier nivel.

Para el individuo i del grupo j se utilizará la notación de y_{ij} para la variable dependiente, x_{1ij}, \dots, x_{pij} las variables explicativas a nivel 1, z_{1j}, \dots, z_{qj} las variables explicativas a nivel 2.

1.2 Tratamiento estadístico y agrupación de datos

Antes de comenzar a explicar detalladamente la modelización estadística de los datos con la estructura multinivel, vamos a ilustrar lo que sucede cuando se ignora la estructura multinivel de los datos y los distintos tipos de errores que se pueden cometer.

1.2.1 Correlación intraclase

El *coeficiente de correlación intraclase* [20] es el grado de relación que se tiene entre los individuos y sus grupos. El término utilizado de *clase* es referente a los grupos. Existen varias definiciones utilizadas para este coeficiente que serán explicadas más adelante.

Uno de los modelos iniciales multinivel, es el llamado *modelo vacío*, debido a que únicamente tiene los efectos de la varianza del modelo. Donde y_{ij} es el valor observado del individuo i en el grupo j , que puede ser expresado de la siguiente manera:

$$y_{ij} = \beta_0 + u_{0j} + e_{0ij} \quad (1.1)$$

Donde β_0 es la gran media poblacional, u_{0j} es el efecto específico del grupo j (nivel 2) y e_{0ij} es el efecto residual de los individuos (nivel 1) dentro de los grupos. Se asume que todas las variables son independientes; el grupo de efectos u_{0j} tienen media cero y varianza $\sigma_{u_0}^2$ (varianza entre grupos); y también que los residuos e_{0ij}

tienen media cero y varianza σ_{e0}^2 (varianza intra grupos). Es así que la varianza total de y_{ij} es igual a la suma de las dos varianzas¹,

$$\text{var}(y_{ij}) = \sigma_{u0}^2 + \sigma_{e0}^2$$

El coeficiente de correlación intraclase ρ_I se define de la siguiente manera:

$$\begin{aligned} \rho_I &= \frac{\text{varianza poblacional entre grupos}}{\text{varianza total}} \\ &= \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_{e0}^2} \end{aligned}$$

Este coeficiente es la proporción de varianza que es representada por cada nivel. También es conocido como coeficiente de correlación, debido a que es igual a la correlación que existe entre dos individuos extraídos aleatoriamente de un grupo cualquiera.

1.2.1.1 Agregación de los datos

La agregación de los datos consiste en la agrupación de los datos de un nivel individuos a un nivel de grupo, pero pueden existir grandes errores al momento de realizar esto.

El primer error potencial que puede existir al añadir una variable de nivel grupo, que solamente se refiere a las unidades de grupo y no directamente a los individuos, es decir, hay que tener cuidado con las variables que se desee incluir en el modelo y qué es lo que en verdad se quiere medir. Por ejemplo, el desempeño promedio de los empleados de una empresa en el trabajo, es usado como una variable para medir el clima organizacional. Esto no sería correcto, debido a que la variable se refiere a la empresa y no directamente a los empleados.

El segundo error que se puede cometer es, que si existe correlación entre las variables de nivel de grupo, uno no puede hacer aseveraciones sobre las relaciones que pudieran existir al nivel de individuos. Por ejemplo, el porcentaje de habitantes negros en el vecindario puede ser relativo a las opiniones políticas promedio en el vecindario. Esto de ninguna manera nos da un indicador sobre la relación a nivel raza - convicción política.

¹Será explicado más adelante en el desarrollo de *Modelos de componentes de la varianza*.

Estos son algunos de los errores que se pueden realizar al momento de la modelización.

1.2.2 Varianza intra grupos y entre los grupos

A continuación se va a explicar la varianza entre grupos y la varianza dentro de los grupos o intra grupos. Para ello se va a considerar *varianza observada entre los grupos* y la *varianza observada intra grupos*, que se irá definiendo más adelante. Primeramente, se define la media del grupo j :

$$\bar{y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$$

y la media general es:

$$\begin{aligned} \bar{y}_{..} &= \frac{1}{M} \sum_{j=1}^N \sum_{i=1}^{n_j} y_{ij} \\ &= \frac{1}{M} \sum_{j=1}^N n_j \bar{y}_{.j} \end{aligned}$$

La varianza observada dentro del grupo j es:

$$S_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{.j})^2$$

De esta manera se tendría un valor para cada grupo, pero lo ideal sería tener un parámetro que exprese la variabilidad dentro de los grupos para todos los grupos al mismo tiempo. Es así que se tiene un promedio ponderado de la varianza dentro de varios grupos, definido de la siguiente manera:

$$\begin{aligned} S_{intra}^2 &= \frac{1}{M - N} \sum_{j=1}^N \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{.j})^2 \\ &= \frac{1}{M - N} \sum_{j=1}^N (n_j - 1) S_j^2 \end{aligned}$$

El valor esperado de la varianza observada intra grupos es igual a la varianza poblacional intra grupos, es decir:

$$E(S_{intra}^2) = \sigma_{e0}^2$$

El caso para la varianza entre grupos es más complicado. Para grupos con igual tamaño n_j , la varianza observada entre grupos se define:

$$S_{entre}^2 = \frac{1}{(N-1)} \sum_{j=1}^N (\bar{y}_{.j} - \bar{y}_{..})^2$$

Para grupos con diferentes tamaños, la contribución de los diferentes grupos debe ser ponderada. A continuación se presenta una fórmula que usa pesos para poder estimar la población de individuos para la varianza entre grupos [19].

$$S_{entre}^2 = \frac{1}{w(N-1)} \sum_{j=1}^N n_j (\bar{y}_{.j} - \bar{y}_{..})^2$$

donde w es:

$$w = \frac{1}{N-1} \left(M - \frac{\sum_j n_j^2}{M} \right)$$

En el Anexo A se demuestra cómo se llega a definir w . La varianza total observada viene por la combinación de las varianzas intra grupos y entre grupos, expresado de la siguiente manera:

$$\begin{aligned} \text{varianza total observada} &= \frac{1}{M-1} \sum_{j=1}^N \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{..})^2 \\ &= \frac{M-N}{M-1} S_{intra}^2 + \frac{w(N-1)}{M-1} S_{entre}^2 \end{aligned}$$

La complicación que existe con la varianza entre grupos es que los residuos de nivel individuos también contribuyen en la varianza, aunque en menor grado.

$$E(S_{entre}^2) = \sigma_{u0}^2 + \frac{\sigma_{e0}^2}{w}$$

El segundo término de esta fórmula viene a ser pequeño cuando w es grande. Para tamaños de grupos grandes la varianza esperada observada entre grupos es prácticamente igual que la varianza entre grupos real. Para tamaños de grupos pequeños, tiende a ser más grande que la varianza entre grupos real. En el Anexo A se demuestra que la varianza observada entre grupos es insesgada.

En la práctica no se conocen los valores de las varianzas entre grupos e intra grupos, estos deben ser estimados de los datos.

1.3 Modelos de componentes de la varianza

Se considera el modelo lineal jerárquico más sencillo, el cual no considera los efectos aleatorios [20]. La variable dependiente puede ser expresada como la suma de una parte determinística y una parte aleatoria.

$$y_{ij} = \underbrace{\beta_0 + \beta_1 x_{ij} + \beta_2 z_j}_{\text{parte determinística}} + \underbrace{e_{ij}}_{\text{parte aleatoria}} \quad (1.2)$$

Además β_0 es la ordenada en el origen, β_1 es el coeficiente de la variable x definida a nivel 1, β_2 es el coeficiente de la variable z definida a nivel 2 y los errores e_{ij} son las variables aleatorias independientes con distribución normal de media cero y varianza constante al interior de cada grupo.

El modelo (1.2) no considera la interacción entre las variables a distinto nivel, es decir la variable x de nivel 1 con la variable z de nivel 2.

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_j + \beta_3 x_{ij} z_j + e_{ij}$$

Este modelo no toma en cuenta la variabilidad entre grupos. Esto se lo hace haciendo que los coeficientes β_0 y β_1 dependan del grupo j , denotados por β_{0j} , β_{1j} .

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \beta_2 z_j + e_{ij} \quad (1.3)$$

1.3.1 Modelo con ordenada aleatoria, sin variable explicativa

En el modelo con ordenada aleatoria no se incluyen las variables x y z , y se lo conoce también como *modelo vacío* [20], representado de la siguiente manera:

$$\begin{aligned} y_{ij} &= \beta_{0j} + e_{0ij} \\ \beta_{0j} &= \beta_0 + u_{0j} \end{aligned}$$

de manera simplificada:

$$y_{ij} = \beta_0 + u_{0j} + e_{0ij} \quad (1.4)$$

Donde β_0 representa la media general, u_{0j} el efecto aleatorio a nivel 2 y e_{0ij} el efecto aleatorio a nivel 1.

Las hipótesis de este modelo son:

- Las variables aleatorias $e_{0ij} \sim N(0, \sigma_{e0}^2)$ y son independientes.
- Las variables aleatorias $u_{0j} \sim N(0, \sigma_{u0}^2)$ y son independientes.
- Las variables e_{0ij} y u_{0j} son mutuamente independientes.

La varianza total del modelo viene dada por:

$$\begin{aligned}
 Var(y_{ij}) &= Var(\beta_{0j} + e_{0ij}) \\
 &= Var(\beta_{0j}) + Var(e_{0ij}) + 2Cov(\beta_{0j}, e_{0ij}) \\
 &= Var(\beta_0) + Var(u_{0j}) + 2Cov(\beta_0, u_{0j}) + Var(e_{0ij}) + 2[E(\beta_{0j}e_{0ij}) - E(\beta_{0j})E(e_{0ij})] \\
 &= Var(u_{0j}) + 2[E(\beta_0 u_{0j}) - E(\beta_0)E(u_{0j})] + Var(e_{0ij}) \\
 &= \sigma_{u0}^2 + \sigma_{e0}^2
 \end{aligned}$$

Esto es la suma de la varianza del nivel 1 con la varianza del nivel 2. Considerando dos variables aleatorias y_{ij} y y_{kj} con ($i \neq k$) definidas en el mismo grupo j se tiene,

$$\begin{aligned}
 Cov(y_{ij}, y_{kj}) &= Cov(\beta_0 + u_{0j} + e_{0ij}, \beta_0 + u_{0j} + e_{0ik}) \\
 &= Cov(u_{0j}, u_{0j}) \\
 &= Var(u_{0j}) \\
 &= \sigma_{u0}^2
 \end{aligned}$$

y la *correlación intraclass* está definida por:

$$Corr(y_{ij}, y_{kj}) = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_{e0}^2}$$

Ésta puede ser interpretada como la razón entre la variabilidad de los grupos y la variabilidad total, o como la correlación que existe entre dos individuos extraídos aleatoriamente de un grupo escogido al azar.

En un modelo con tres niveles, por ejemplo con las escuelas, aulas y alumnos, vamos a tener dos correlaciones, la correlación intra-escuela, que es también la proporción de la variación que hay entre las escuelas y la correlación intra-aulas que es entre aulas. En modelos más complejos con coeficientes aleatorios de la correlación intra unidades no es equivalente a la proporción de la variación en el

nivel superior, esto lo analizaremos mas adelante.

Escritura matricial: Se consideran un número arbitrario de grupos, digamos m grupos y cada uno con n_1, \dots, n_m individuos. Matricialmente el modelo se lo expresa de la siguiente manera:

$$Y = X\beta + U + E \quad (1.5)$$

donde

$$Y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{n_1 1} \\ \vdots \\ y_{1m} \\ \vdots \\ y_{n_m m} \end{pmatrix}, \quad X = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \beta = \beta_0$$

$$U = \begin{pmatrix} u_{01} \\ \vdots \\ u_{01} \\ \vdots \\ u_{0m} \\ \vdots \\ u_{0m} \end{pmatrix}, \quad E = \begin{pmatrix} e_{011} \\ \vdots \\ e_{0n_1 1} \\ \vdots \\ e_{01m} \\ \vdots \\ e_{0n_m m} \end{pmatrix}$$

Como se describió anteriormente $U \sim N(0, \Omega_U)$ y $E \sim N(0, \Omega_E)$. Considerando dos grupos y cada uno con n_1 y n_2 individuos se tiene:

$$\Omega_U = \sigma_{u0}^2 \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & \cdots & 1 \end{pmatrix} = \sigma_{u0}^2 A$$

Esta estructura diagonal por bloques, refleja el hecho de que cualquier covarian-

za entre individuos de diferentes grupos es igual a cero.

$$\Omega_E = \sigma_{e0}^2 \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix} = \sigma_{e0}^2 B$$

Con lo que se tiene la matriz de varianza global igual a:

$$V = \begin{pmatrix} \sigma_{u0}^2 + \sigma_{e0}^2 & \cdots & \sigma_{u0}^2 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{u0}^2 & \cdots & \sigma_{u0}^2 + \sigma_{e0}^2 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \sigma_{u0}^2 + \sigma_{e0}^2 & \cdots & \sigma_{u0}^2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma_{u0}^2 & \cdots & \sigma_{u0}^2 + \sigma_{e0}^2 \end{pmatrix} = \sigma_{u0}^2 A + \sigma_{e0}^2 B$$

Para el caso general de m grupos, la matriz V mantiene la misma estructura de bloques y es de dimensión N x N.

1.3.2 Modelo con ordenada aleatoria, con una variable explicativa

En este modelo la variable dependiente se encuentra expresada en función de alguna variable explicativa. El modelo viene formulado de la siguiente manera:

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + e_{0ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

de manera simplificada:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{0ij} \quad (1.6)$$

Realizando una gráfica de la ecuación (1.6), cada grupo se encuentra representado por una recta con pendiente β_1 y ordenada en el origen $\beta_0 + u_{0j}$, donde el término β_0 es la media de las ordenadas en el origen de todos los grupos y el coeficiente β_1 es el coeficiente de regresión común.

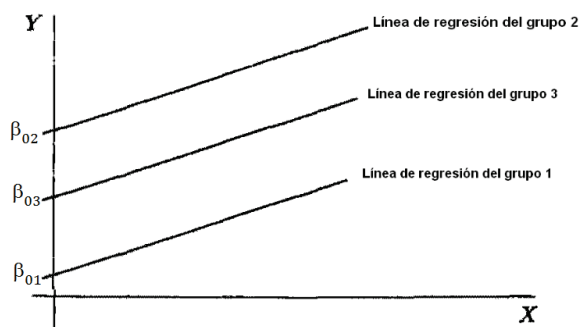


Figura 1.1: Modelo con ordenada aleatoria.

Debido a que el parámetro β_{0j} varía aleatoriamente de un grupo a otro, se lo llama *modelo con ordenada aleatoria* [20]. Las variables u_{0j} son los residuos a nivel de grupo y los e_{oij} son los residuos a nivel de individuos.

Los efectos aleatorios u_{0j} no son parámetros, sino son *variables latentes*²; es decir, variables no observables directamente.

La varianza residual, es decir, la varianza condicional de x , viene dada por:

$$\begin{aligned} \text{Var}(y_{ij} | x_{ij}) &= \text{Var}(u_{0j}) + \text{Var}(e_{oij}) \\ &= \sigma_{u0}^2 + \sigma_{e0}^2 \end{aligned}$$

La covarianza entre dos individuos, y_{ij} y y_{kj} con ($i \neq k$), de un mismo grupo j , es:

$$\text{Cov}(y_{ij}, y_{kj} | x_{ij}, x_{kj}) = \text{Var}(u_{0j}) = \sigma_{u0}^2$$

De la covarianza descrita anteriormente entre dos individuos de un mismo grupo j , una parte es explicada por los valores de x y otra parte no es explicada. Esta parte no explicada viene dada por el *coeficiente de correlación residual* [20].

$$\text{Corr}(y_{ij}, y_{kj} | x_{ij}, x_{kj}) = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_{e0}^2}$$

Esto es análogo al coeficiente de correlación intra clase, pero ahora los parámetros de la varianza σ_{u0}^2 y σ_{e0}^2 descritos en el modelo (1.6), incluyen los efectos de la variable x .

El modelo de la ecuación (1.6) es válido, si el coeficiente de correlación intracla-

²Será explicado más adelante en el capítulo 2

se, o equivalentemente $\sigma_{u_0}^2$, es positivo, entonces el modelo lineal jerárquico es el mejor método de análisis, debido a que indicaría que hay variabilidad entre grupos, lo cual no se tiene en la regresión clásica.

Escritura matricial: Se consideran m grupos y cada uno con n_1, \dots, n_m individuos. Matricialmente el modelo se lo expresa de la siguiente manera:

$$Y = X\beta + U + E \quad (1.7)$$

donde,

$$Y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{n_11} \\ \vdots \\ y_{1m} \\ \vdots \\ y_{n_mm} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} \\ \vdots & \vdots \\ 1 & x_{n_11} \\ \vdots & \vdots \\ 1 & x_{1m} \\ \vdots & \vdots \\ 1 & x_{n_mm} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$U = \begin{pmatrix} u_{01} \\ \vdots \\ u_{01} \\ \vdots \\ u_{0m} \\ \vdots \\ u_{0m} \end{pmatrix}, \quad E = \begin{pmatrix} e_{011} \\ \vdots \\ e_{0n_11} \\ \vdots \\ e_{01m} \\ \vdots \\ e_{0n_mm} \end{pmatrix}$$

Los vectores U y E tienen las siguientes distribuciones:

$$U \sim N(0, \Omega_U) \quad \text{y} \quad E \sim N(0, \Omega_E). \\ \Omega_U = \sigma_{u_0}^2 A \quad \text{y} \quad \Omega_E = \sigma_{e_0}^2 B$$

La matriz de varianzas global es:

$$V = \sigma_{u_0}^2 A + \sigma_{e_0}^2 B$$

1.3.3 Modelo con ordenada aleatoria y varias variables explicativas

En un modelo con ordenada aleatoria más de una variable explicativa puede intervenir en el modelo [20]. Es así que las variables explicativas de nivel 1 son:

X_1, X_2, \dots, X_p y las variables explicativas de nivel 2 son las siguientes: $X_{p+1}, X_{p+2}, \dots, X_{p+q}$. El modelo se lo formula de la siguiente manera:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} x_{(p+1)j} + \dots + \beta_{p+q} x_{(p+q)j} + u_{0j} + e_{0ij} \quad (1.8)$$

Los parámetros de la regresión son: β_k ($k = 1, \dots, p$) para las variables explicativas de nivel 1 y β_ℓ ($\ell = p + 1, \dots, p + q$) para las variables explicativas de nivel 2 [2].

La ecuación (1.8) está compuesta por la *parte fija* del modelo que viene dada por:

$$\beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \beta_{p+1} x_{(p+1)j} + \dots + \beta_{p+q} x_{(p+q)j}$$

y la parte restante,

$$u_{0j} + e_{0ij}$$

llamada *parte aleatoria* del modelo. Los supuestos para la parte aleatoria del modelo son los siguientes:

- Las variables aleatorias $e_{0ij} \sim N(0, \sigma_{e0}^2)$ y son independientes.
- Las variables aleatorias $u_{0j} \sim N(0, \sigma_{u0}^2)$ y son independientes.
- Las variables e_{0ij} y u_{0j} son mutuamente independientes.

Escritura matricial:

$$Y = X\beta + U + E$$

donde,

$$Y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{n1} \\ \vdots \\ y_{1m} \\ \vdots \\ y_{nmm} \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & x_{111} & \cdots & x_{p11} & x_{(p+1)1} & \cdots & x_{(p+q)1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n_11} & \cdots & x_{pn_11} & x_{(p+1)1} & \cdots & x_{(p+q)1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{11m} & \cdots & x_{p1m} & x_{(p+1)m} & \cdots & x_{(p+q)m} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n_m m} & \cdots & x_{pn_m m} & x_{(p+1)m} & \cdots & x_{(p+q)m} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p+q} \end{pmatrix}, \quad U = \begin{pmatrix} u_{01} \\ \vdots \\ u_{01} \\ \vdots \\ u_{0m} \\ \vdots \\ u_{0m} \end{pmatrix}, \quad E = \begin{pmatrix} e_{11} \\ \vdots \\ e_{n_11} \\ \vdots \\ e_{1m} \\ \vdots \\ e_{n_m m} \end{pmatrix}$$

Los vectores U y E tienen distribuciones de probabilidad:

$$U \sim N(0, \Omega_U) \quad \text{y} \quad E \sim N(0, \Omega_E)$$

$$\Omega_U = \sigma_{u_0}^2 A \quad \text{y} \quad \Omega_E = \sigma_{e_0}^2 B$$

La matriz de varianza global es:

$$V = \sigma_{u_0}^2 A + \sigma_{e_0}^2 B$$

1.4 Modelos con pendientes aleatorias

Hasta ahora se han analizado modelos en los cuales la ordenada es aleatoria y el resto de coeficientes y variables explicativas son parámetros fijos. A continuación se presentan modelos en los cuales las pendientes también son aleatorias.

1.4.1 Modelo con ordenada y pendiente aleatorias

El modelo con ordenada aleatoria:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{0ij}$$

Si se tiene que β_1 depende del grupo j, se tiene el siguiente modelo:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{0ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

de manera simplificada:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + e_{0ij} \quad (1.9)$$

Este modelo tiene como *parte fija* $\beta_0 + \beta_1 x_{ij}$ y como *parte aleatoria* $u_{0j} + u_{1j} x_{ij} + e_{0ij}$. Los residuos u_{0j} y u_{1j} de nivel 2, como los residuos e_{0ij} de nivel 1 son independientes. El término $u_{1j} x_{ij}$ es una interacción aleatoria entre el grupo y la variable x ; por esto se lo llama pendiente aleatoria [20]. Las hipótesis de este modelo son las siguientes:

- Las variables residuales $e_{0ij} \sim N(0, \sigma_{e0}^2)$.
- Las variables residuales $u_{0j} \sim N(0, \sigma_{u0}^2)$.
- Las variables residuales $u_{1j} \sim N(0, \sigma_{u1}^2)$.
- Las variables e_{0ij} y $[u_{0j}, u_{1j}]$ son mutuamente independientes.

además:

$$Cov(u_{0j}, u_{1j}) = \sigma_{u01}$$

$$Cov(u_{0j}, u_{0k}) = 0, \quad \text{con } j \neq k$$

$$Cov(u_{1j}, u_{1k}) = 0, \quad \text{con } j \neq k$$

La matriz de varianza-covarianza de la ordenada aleatoria y la pendiente aleatoria del nivel 2; es decir, los residuos de nivel 2 es la siguiente:

$$\Omega_U = \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix}$$

La varianza de residuos de nivel 1 es:

$$\Omega_E = \sigma_{e0}^2$$

Heteroscedasticidad

El modelo de la ecuación (1.9) no implica solamente que los individuos dentro de un mismo grupo tienen valores de variable respuesta correlacionados, sino que la

varianza de la variable respuesta depende de los valores x . Esto se llama *heteroscedasticidad* en la literatura estadística [20].

$$\begin{aligned}
Var(y_{ij}|x_{ij}) &= Var(\beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + e_{0ij}) \\
&= \beta_1^2 Var(x_{ij}) + Var(u_{0j}) + Var(u_{1j} x_{ij}) + Var(e_{0ij}) \\
&\quad + 2\beta_1 Cov(x_{ij}, u_{0j}) + 2\beta_1 Cov(x_{ij}, u_{1j} x_{ij}) \\
&\quad + 2\beta_1 Cov(x_{ij}, e_{0ij}) + 2Cov(u_{0j}, u_{1j} x_{ij}) \\
&\quad + 2Cov(u_{0j}, e_{0ij}) + 2Cov(u_{1j} x_{ij}, e_{0ij}) \\
&= 0 + \sigma_{u0}^2 + x_{ij}^2 \sigma_{u1}^2 + \sigma_{e0}^2 + 0 + 0 + 0 + 2x_{ij} \sigma_{u01} + 0 + 0 \\
&= \sigma_{u0}^2 + 2x_{ij} \sigma_{u01} + x_{ij}^2 \sigma_{u1}^2 + \sigma_{e0}^2
\end{aligned} \tag{1.10}$$

Si se consideran dos diferentes individuos y_{ij} y y_{kj} ($i \neq k$), en el mismo grupo j se tiene que:

$$Cov(y_{ij}, y_{kj}) = \sigma_{u0}^2 + \sigma_{u01}(x_{ij} + x_{kj}) + \sigma_{u1} x_{ij} x_{kj} \tag{1.11}$$

La expresión (1.10) de la varianza de y , es un polinomio cuadrático con respecto a x_{ij} y alcanza su mínimo en $x_{ij}^* = -\frac{\sigma_{u01}}{\sigma_{u1}^2}$. Cuando x_{ij}^* está dentro del rango de los valores posibles de x , la varianza residual primero decrece y luego crece; si x_{ij}^* es menor que todos los valores de x , entonces la varianza residual es una función creciente de x , si x_{ij}^* es mayor que todos los valores de x , entonces la varianza residual es decreciente.

El modelo de la ecuación (1.9) implica que el coeficiente de regresión, o pendiente del grupo j es $\beta_1 + u_{1j}$. Este término es una variable aleatoria normalmente distribuida con media β_1 y varianza σ_{u1}^2 .

Escritura matricial:

$$Y = X\beta + W + E$$

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{n_1 1} \\ \vdots \\ y_{1m} \\ \vdots \\ y_{n_m m} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} \\ \vdots & \vdots \\ 1 & x_{n_1 1} \\ \vdots & \vdots \\ 1 & x_{1m} \\ \vdots & \vdots \\ 1 & x_{n_m m} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} u_{01} + u_{11}x_{11} \\ \vdots \\ u_{01} + u_{11}x_{n_1 1} \\ \vdots \\ u_{0m} + u_{1m}x_{1m} \\ \vdots \\ u_{0m} + u_{1m}x_{n_m m} \end{pmatrix} + \begin{pmatrix} e_{11} \\ \vdots \\ e_{n_1 1} \\ \vdots \\ e_{1m} \\ \vdots \\ e_{n_m m} \end{pmatrix}$$

La matriz W se la construye de la siguiente manera:

$$W = \begin{pmatrix} u_{01} \\ \vdots \\ u_{01} \\ \vdots \\ u_{0m} \\ \vdots \\ u_{0m} \end{pmatrix} + \text{diag}(x_{11}, \dots, x_{n_1 1}, \dots, x_{1m}, \dots, x_{n_m m}) \begin{pmatrix} u_{11} \\ \vdots \\ u_{11} \\ \vdots \\ u_{1m} \\ \vdots \\ u_{1m} \end{pmatrix}$$

$$= U_0 + \text{diag}(x_{11}, \dots, x_{n_1 1}, \dots, x_{1m}, \dots, x_{n_m m}) U_1$$

Los vectores U_0 , U_1 y E siguen las siguientes distribuciones de probabilidad.

$$U_0 \sim N(0, \Omega_{U_0}), \quad U_1 \sim N(0, \Omega_{U_1}), \quad E \sim (0, \Omega_E)$$

$$\Omega_{U_0} = \sigma_{u_0}^2 I, \quad \Omega_{U_1} = \sigma_{u_1}^2 I, \quad \Omega_E = \sigma_{e_0}^2 I$$

La varianza global viene dada por:

$$V = \Omega_{U_0} + \Omega_{U_1} D + \Omega_E$$

donde $D = \text{diag}(x_{11}^2, \dots, x_{n_1 1}^2, \dots, x_{1m}^2, \dots, x_{n_m m}^2)$.

1.4.2 Modelo con ordenada aleatoria, pendiente aleatoria y una variable de nivel 2

Este modelo incorpora una variable z de nivel 2, al modelo visto anteriormente (1.9) y se lo formula de la siguiente manera:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{0ij}$$

$$\beta_{0j} = \beta_0 + \beta_{01}z_j + u_{0j}$$

$$\beta_{1j} = \beta_1 + \beta_{11}z_j + u_{1j}$$

sustituyendo las expresiones de β_{0j} y β_{1j} , se tiene:

$$\begin{aligned} y_{ij} &= (\beta_0 + \beta_{01}z_j + u_{0j}) + (\beta_1 + \beta_{11}z_j + u_{1j})x_{ij} + e_{0ij} \\ &= \beta_0 + \beta_{01}z_j + \beta_1x_{ij} + \beta_{11}z_jx_{ij} + u_{0j} + u_{1j}x_{ij} + e_{0ij} \end{aligned} \quad (1.12)$$

Comparando con la ecuación del modelo (1.9), la ecuación (1.12) presenta una parte fija diferente del modelo, pero la parte aleatoria del modelo no cambia que es $u_{0j} + u_{1j}x_{ij} + e_{0ij}$. Se espera que las varianzas $\sigma_{u_0}^2$ y $\sigma_{u_1}^2$ sean menores que las del modelo (1.9), ya que parte de la variabilidad ahora será explicada por la variable z de segundo nivel.

En la ecuación (1.12) se observa que la ordenada β_{0j} por una variable z de segundo nivel, produce un efecto principal de z que es $\beta_{01}z_j$. A diferencia el coeficiente β_{1j} de la variable x por una variable z de nivel 2, produce un efecto de interacción entre x y z , que es $\beta_{11}z_jx_{ij}$. La interacción entre las variables de nivel 1 y nivel 2 es llamado *interacción entre niveles* [2].

1.4.3 Modelo lineal jerárquico de nivel 2

Los modelos precedentes pueden ser extendidos por inclusión de más variables que tengan efectos aleatorios y variables que expliquen estos efectos aleatorios. Supongamos que se tiene p variables explicativas de nivel 1, x_1, x_2, \dots, x_p y q variables explicativas de nivel 2, z_1, z_2, \dots, z_q , con lo que se formula el modelo:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \dots + \beta_{pj}x_{pij} + e_{0ij} \quad (1.13)$$

donde

$$\beta_{0j} = \beta_{00} + \beta_{01}z_{1j} + \dots + \beta_{0q}z_{qj} + u_{0j}$$

$$\vdots$$

$$\beta_{pj} = \beta_{p0} + \beta_{p1}z_{1j} + \dots + \beta_{pq}z_{qj} + u_{pj}$$

Este modelo considera todas las pendientes de las variables X aleatorias y se expresan como combinaciones lineales de las variables Z .

Los residuos de nivel 2, u_{0j}, \dots, u_{pj} , indican que los parámetros $\beta_{0j}, \dots, \beta_{pj}$ varían en cada grupo. Como por ejemplo el término β_{00} es la media general del conjunto de ordenadas, mientras que u_{0j} mide la desviación respecto de la media general en cada grupo.

Reemplazando los términos de la ecuación (1.13) se tiene,

$$y_{ij} = (\beta_{00} + \beta_{01}z_{1j} + \dots + \beta_{0q}z_{qj} + u_{0j}) + \dots + (\beta_{p0} + \beta_{p1}z_{1j} + \dots + \beta_{pq}z_{qj} + u_{pj})x_{pij} + e_{0ij}$$

agrupando términos se llega a:

$$y_{ij} = \beta_{00} + \sum_{h=1}^p \beta_{h0}x_{hij} + \sum_{k=1}^q \beta_{0k}z_{kj} + \sum_{h=1}^p \sum_{k=1}^q \beta_{hk}z_{kj}x_{hij} + u_{0j} + \sum_{h=1}^p u_{hj}x_{hij} + e_{0ij} \quad (1.14)$$

Esta ecuación indica los efectos principales de las variables X y Z, como el producto de interacciones entre niveles [20].

Los grupos se caracterizan por $p + 1$ coeficientes aleatorios, u_{0j}, \dots, u_{pj} . Estos coeficientes aleatorios son independientes entre los grupos, pero pueden estar correlacionados dentro de los grupos. Se supone que el vector (u_{0j}, \dots, u_{pj}) es independiente de los residuos de nivel 1, e_{0ij} y que todos los residuos están distribuidos con media cero. Los residuos de nivel 1, e_{0ij} tienen distribución normal con varianza constante σ_{e0}^2 y que el vector (u_{0j}, \dots, u_{pj}) tiene distribución normal multivariante con matriz de covarianza constante, con componentes:

$$Var(u_{hj}) = \sigma_{uh}^2, \quad (h = 0, 1, \dots, p)$$

$$Cov(u_{hj}, u_{kj}) = \sigma_{uhk}, \quad (h, k = 0, 1, \dots, p)$$

A menos que p y q sean bastante pequeños, el modelo de la ecuación (1.14), tiene un número de parámetros muy grande. Para simplificar el modelo usualmente se usan los siguientes criterios [20]:

- No todas las variables X consideradas que tengan pendientes aleatorias. Esto se justifica, porque al estar las pendientes aleatorias explicadas por las variables Z, pueden aparecer coeficientes residuales no significativos.
- Dado que los coeficientes β_{hj} , de una variable x_h varía a través de los grupos, no es necesario usar todas las variables Z para explicar esa variabilidad. El número de términos de interacción entre niveles pueden ser restringidos al elegir un conveniente subconjunto de variables Z.

Una formulación general del modelo (1.14), con el fin de unificar la notación de las variables, ya que se utilizan dos tipos de variables, las X y Z, es unificar las variables de nivel 1, las variables de nivel 2 y las interacciones entre niveles utilizando la notación de X. De esta manera facilita el manejo de los modelos jerárquicos multinivel.

Si se tienen r variables explicativas, de manera que las primeras p tengan coeficientes fijos y aleatorios, mientras que las $r - p$ restantes tengan coeficientes fijos. Con esta notación se tiene el siguiente modelo:

$$y_{ij} = \beta_0 + \sum_{h=1}^r \beta_h x_{hij} + u_{0j} + \sum_{h=1}^p u_{hj} x_{hij} + e_{0ij} \quad (1.15)$$

en dos términos, $\beta_0 + \sum_{h=1}^r \beta_h x_{hij}$ y $u_{0j} + \sum_{h=1}^p u_{hj} x_{hij} + e_{0ij}$ son la parte fija y aleatoria del modelo, respectivamente.

Escritura matricial: El modelo de la ecuación (1.15) se formula matricialmente [2] de la siguiente manera:

$$Y = X\beta + W\alpha + E$$

donde la variable respuesta Y es un vector ($n \times 1$), X es una matriz de dimensión [$n \times (r + 1)$], β es un vector [$(r + 1) \times 1$] respecto a los parámetros de la parte fija del modelo, W es una matriz de dimensión [$n \times (p + 1)$], α es un vector [$(p + 1) \times 1$] respecto a los parámetros de la parte aleatoria del modelo y E es un vector de errores de dimensión ($n \times 1$). Además se asume que,

$$E(E) = 0 \quad Var(E) = \Omega_e = \sigma_{e0}^2 I$$

$$E(\alpha) = 0 \quad Var(\alpha) = \Omega_\alpha = diag(\sigma_1^2, \dots, \sigma_{p+1}^2) \quad Cov(\alpha, E) = 0$$

1.5 Especificación de los modelos

El objetivo de la especificación del modelo es obtener una descripción adecuada de los datos observados, a partir de los modelos de pendientes aleatorias; el investigador tiene algunas opciones de modelos para sus datos. Se deben seleccionar variables explicativas que sean relevantes y sus interacciones, para la parte fija y aleatoria del modelo.

Una buena guía para escoger entre una pendiente fija y una aleatoria para una variable en concreto se debería encontrar en los fundamentos teóricos del problema

que se investiga. Si la teoría no proporciona una pista para poder decidir, entonces se recomienda asignar pendiente aleatoria a la variable en estudio y decidir la permanencia a través de las pruebas de hipótesis [20].

La naturaleza del modelo lineal jerárquico es complicada, ya que no existen reglas a seguir para obtener un modelo óptimo, pero se pueden seguir considerando los siguientes principios para obtener un modelo satisfactorio observando el siguiente diagrama:

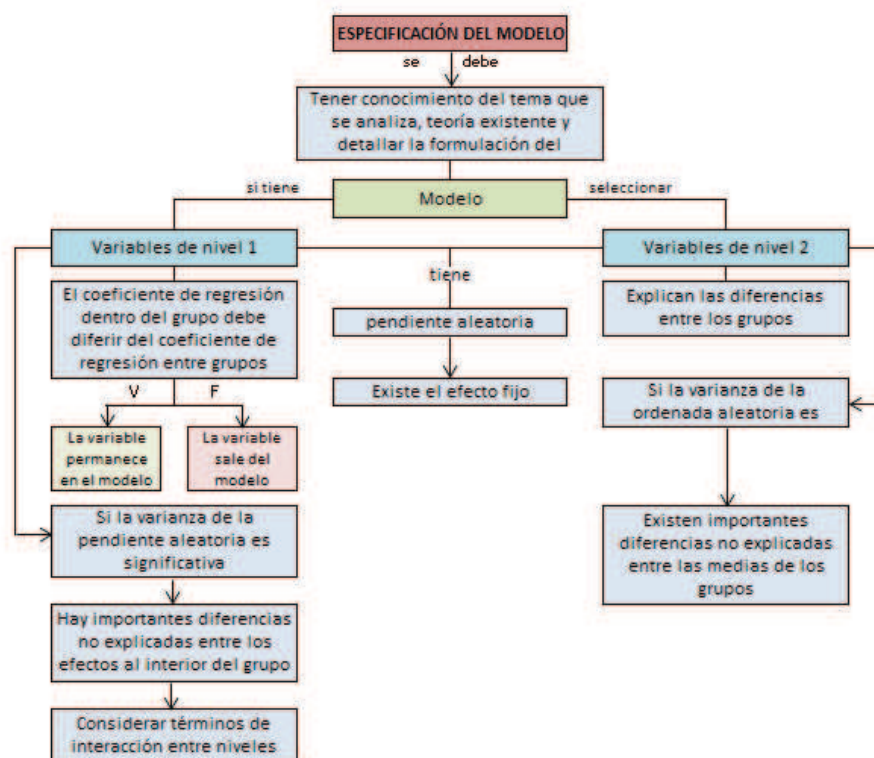


Figura 1.2: Diagrama de especificación del modelo.

Para los modelos multinivel es aconsejable comenzar el ajuste a partir del modelo vacío, el cual da las varianzas dentro y entre los grupos, a partir de las cuales se calcula la correlación intraclase.

El proceso de especificación de modelos contempla la selección de efectos adicionales (fijos o aleatorios), probar su significación y decidir si se les incluye o no en el modelo. Para realizar este proceso se realiza la modelización de la variabilidad dentro de los grupos y luego la modelización de la variabilidad entre los grupos.

1.5.1 Modelización de variabilidad dentro de los grupos

Una vez que se disponen de los datos, se seleccionan p variables de nivel 1, consideradas importantes, para explicar la variable y . Considerando la ecuación (1.13):

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \dots + \beta_{pj}x_{pij} + e_{0ij}$$

Esta ecuación representa los efectos dentro del grupo, de las variables X sobre y. La variabilidad entre grupos, es primero modelada por la variabilidad aleatoria. Esto se representa por la descomposición de los coeficientes de regresión β_{hj} , en un coeficiente β_{h0} (media general) y una desviación u_{hj} , con $h = 0, \dots, p$.

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

$$\vdots$$

$$\beta_{pj} = \beta_{p0} + u_{pj}$$

sustituyendo las β_{hj} se tiene:

$$y_{ij} = \beta_{00} + \sum_{h=1}^p \beta_{h0}x_{hij} + u_{0j} + \sum_{h=1}^p u_{hj}x_{hij} + e_{0ij}$$

Este modelo tiene p variables de nivel 1 con efectos fijos y aleatorios, pero no se deben incluir todos los efectos aleatorios necesariamente. Para esto se debe considerar las siguientes recomendaciones [20]:

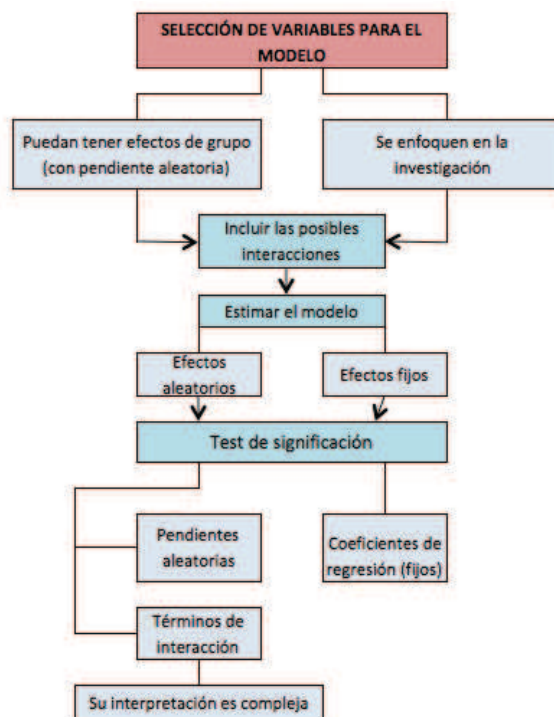


Figura 1.3: Diagrama de selección de efectos en el modelo.

En general los datos no contienen mucha información sobre efectos aleatorios; es así que si el modelo presenta muchas pendientes aleatorias el proceso de estimación puede volverse muy largo e incluso puede no converger, es por esto que solo se debe especificar un pequeño número de pendientes aleatorias.

Luego de realizar este proceso se tendrá un modelo con variables, efectos fijos y aleatorios. Es posible que la ordenada en el origen sea el único término aleatorio.

1.5.2 Modelización de variabilidad entre grupos

La modelización de variabilidad entre grupos va a ser explicada por los efectos aleatorios de variables de nivel 2. La varianza del intercepto aleatorio puede ser explicada por variables de nivel 2 y las pendientes aleatorias pueden ser explicadas por la interacciones de las variables de nivel 1 con las variables del nivel 2.

Para seleccionar los efectos de las variables de nivel 2 y las interacciones entre niveles a ser incluidas, es recomendable seleccionar aquellas cuyos efectos se consideren importantes y realizar las respectivas pruebas de significación de tal manera que justifiquen su inclusión.

Las interacciones entre niveles; es decir, una interacción entre una variable x de nivel 1 y una variable z de nivel 2, se tiene solamente si x tiene pendiente aleatoria significativa. Existen razones teóricas para considerar la interacción entre los dos niveles (xz); esto se puede probar incluso si x no tiene pendiente aleatoria significativa. De hecho la prueba de interacción entre niveles es más potente que la prueba de pendiente aleatoria.

1.6 Heteroscedasticidad

El modelo lineal jerárquico es un modelo muy flexible; este tiene algunas características que permiten representar la estructura de datos jerárquicos. Una de estas características es la posibilidad de representar los modelos multinivel de regresión de nivel 1 donde la varianza residual no es constante [9].

Al momento que se asume que una pendiente es aleatoria, implica proponer una hipótesis de *heteroscedasticidad* [20] en el modelo y la presencia de correlación entre las unidades de nivel 1, que pertenecen a un mismo grupo. Se debe probar la naturaleza aleatoria de las pendientes de todas las variables de interés y no únicamente de las que se esperaría tengan una pendiente aleatoria.

En el modelo lineal general, se supone que las varianzas residual a nivel 1 y la varianza de la ordenada en el origen a nivel 2 son constantes. En este caso de heteroscedasticidad la varianza de y depende en forma lineal o cuadrática de las variables explicativas.

1.6.1 Heteroscedasticidad a nivel 1

En el modelo lineal jerárquico existe la posibilidad de que la varianza residual a nivel 1 dependa de una de las variables explicativas.

Función de varianza lineal

Si la varianza depende linealmente de alguna variable x_1 , puede ser expresada por:

$$\text{varianza nivel 1} = \alpha + \delta_1 x_{1ij}$$

donde α y δ_1 son los parámetros a ser estimados.

Si la varianza depende linealmente de p variables explicativas, puede ser expresado de la siguiente manera:

$$\text{varianza nivel 1} = \alpha + \sum_{k=1}^p \delta_1 x_{kij}$$

Función de varianza cuadrática

Si la dependencia es cuadrática con respecto a x_1 , la varianza se la expresa de la siguiente manera:

$$\text{varianza nivel 1} = \alpha + \delta_1 x_{1ij} + \delta_2^2 x_{1ij}^2$$

De la misma manera si la dependencia es cuadrática pero con respecto a varias variables, es decir, p variables explicativas, se tiene:

$$\text{varianza nivel 1} = \alpha + \sum_{k=1}^p \delta_k x_{kij}^2 + \sum_{h=0}^{p-1} \sum_{k=h+1}^p \delta_{hk} x_{hij} x_{kij}$$

1.6.2 Heteroscedasticidad a nivel 2

En el modelo lineal jerárquico general se supone que existe homoscedasticidad a nivel 2, pero puede ser que no se verifique dicha suposición; de ser el caso, se puede suponer que la varianza existente depende de algunas variables de nivel 2.

Función de varianza lineal

Si la varianza depende linealmente de alguna variable z_1 , puede ser expresada por:

$$\text{varianza nivel 2} = \gamma + \theta_1 z_{1j}$$

donde γ y θ_1 son los parámetros a ser estimados.

Función de varianza cuadrática

La función de varianza cuadrática es respecto a z_1 y puede ser representada por una dependencia curvilínea de la varianza de la variable dependiente.

$$\text{varianza nivel 2} = \gamma + \theta_1 z_{1j} + \theta_2^2 z_{1j}^2$$

Para generalizar para p variables de nivel 2 o la dependencia es cuadrática, se

tiene que:

$$\text{varianza nivel 2} = \gamma + \sum_{k=1}^p \theta_k z_{kj}^2 + \sum_{h=0}^{p-1} \sum_{k=h+1}^p \theta_{hk} z_{hj} z_{kj}$$

1.7 Estimación

1.7.1 Residuos

Un modelo multinivel generalmente tiene residuos en diferentes niveles [11]. Los verdaderos valores de los residuos son desconocidos, pero se puede predecirlos con los datos observados y los parámetros estimados del modelo.

Supóngase que y_{ij} es el valor observado del i -ésimo estudiante en el j -ésimo colegio y que \hat{y}_{ij} es el valor predicho de la regresión. Entonces el *residuo bruto* para un estudiante es $r_{ij} = y_{ij} - \hat{y}_{ij}$. El residuo bruto para el j -ésimo colegio es la media de r_{ij} para los estudiantes en el colegio, escrito como r_{+j} . Entonces el residuo estimado de nivel 2 se obtiene multiplicando r_{+j} por el siguiente factor:

$$\hat{u}_{0j} = \left(\frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_{e0}^2/n_j} \right) r_{+j}$$

donde n_j es el número de estudiantes en el colegio j .

El factor que multiplica en la fórmula descrita anteriormente es siempre menor o igual a 1; por tanto, el residuo estimado es generalmente menor en magnitud que el residuo bruto. Al factor que multiplica el residuo bruto es llamado *factor de contracción* y el residuo estimado es llamado *residuo reducido*, en adelante lo llamaremos únicamente *residuo*. El factor de contracción será notablemente menor que 1 cuando σ_{e0}^2 es grande comparado con σ_{u0}^2 o cuando n_j es pequeño (o ambos).

La estimación de los residuos de nivel 1 viene dada por:

$$\hat{e}_{ij} = r_{ij} - \hat{u}_{0j}$$

1.7.2 Método de Mínimos Cuadrados Generalizados Iterativo (MCGI)

Consideremos el modelo de dos niveles descrito en la ecuación (1.6):

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{0ij}$$

Con las siguientes hipótesis:

$$\begin{aligned} u_{0j} &\sim N(0, \Omega_U) & \mathbf{y} & \quad e_{0ij} \sim N(0, \Omega_E), \\ \text{Var}(u_{0j}) &= \sigma_{u0}^2 & \mathbf{y} & \quad \text{Var}(e_{0ij}) = \sigma_{e0}^2. \end{aligned}$$

En este modelo los parámetros a estimar son β_0 , β_1 , σ_{u0}^2 y σ_{e0}^2 , mediante el método de *Mínimos Cuadrados Generalizados Iterativo* (MCGI) [9].

Primeramente se realiza una estimación inicial de $\hat{\beta}^{(0)}$, que viene dada por:

$$\hat{\beta}^{(0)} = (X^T X)^{-1} X^T Y$$

Luego se procede a calcular el vector de residuos $\tilde{Y} = Y - \hat{Y}$, es decir:

$$\tilde{y}_{ij} = y_{ij} - (\hat{\beta}_0 + \hat{\beta}_1 x_{ij})$$

Una vez obtenida la matriz \tilde{Y} , la matriz resultante del producto $\tilde{Y}\tilde{Y}^t$ se la ordena como vector de la siguiente manera:

$$\text{vector}(\tilde{Y}\tilde{Y}^t) = \begin{pmatrix} \tilde{y}_{11}^2 \\ \tilde{y}_{12}\tilde{y}_{11} \\ \vdots \\ \tilde{y}_{n_m m}^2 \end{pmatrix}$$

Como la $E(\tilde{Y}\tilde{Y}^t) = \text{Var}(Y)$; se tiene lo siguiente:

$$\begin{aligned} \begin{pmatrix} \tilde{y}_{11}^2 \\ \tilde{y}_{12}\tilde{y}_{11} \\ \vdots \\ \tilde{y}_{n_m m}^2 \end{pmatrix} &= \begin{pmatrix} \hat{\sigma}_{u0}^2 + \hat{\sigma}_{e0}^2 \\ \hat{\sigma}_{u0}^2 \\ \vdots \\ \hat{\sigma}_{u0}^2 + \hat{\sigma}_{e0}^2 \end{pmatrix} + R \\ &= \sigma_{u0}^2 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \sigma_{e0}^2 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 1 \end{pmatrix} + R \end{aligned}$$

donde R es un vector de residuos. A partir de esto se estiman σ_{u0}^2 y σ_{e0}^2 mediante Mínimos Cuadrados Generalizados, y se obtiene:

$$\hat{V} = \hat{\sigma}_{u0}^2 A + \hat{\sigma}_{e0}^2 B$$

Luego, se generan nuevas estimaciones $\hat{\beta}^{(1)}$, utilizando \hat{V} , mediante:

$$\hat{\beta}^{(1)} = (X^T \hat{V}^{-1} X)^{-1} X^T V^{-1} Y$$

Se repite este procedimiento hasta que el algoritmo converga, es decir, se repite el algoritmo el número de veces que sea necesario hasta que los valores obtenidos de los parámetros sean invariantes. Cuando se obtengan las estimaciones definitivas de los parámetros se procede a estimar los residuos u_{0j} y e_{0ij} de la siguiente manera:

Denotamos por $\tilde{y}_{ij} = \frac{\sum_{i=1}^{n_j} \tilde{y}_{ij}}{n_j}$, entonces se tiene:

$$\hat{u}_{0j} = \frac{n_j \hat{\sigma}_{u_0}^2}{n_j \hat{\sigma}_{u_0}^2 + \hat{\sigma}_{e_0}^2} \tilde{y}_j$$

$$\tilde{e}_{0ij} = \tilde{y}_{ij} - \hat{u}_{0j}$$

Ahora, consideremos el modelo lineal jerárquico de dos niveles descrito en la ecuación (1.15), que se lo puede representar matricialmente de la siguiente manera:

$$Y = X\beta + W\alpha + E$$

donde β es un vector de los parámetros que corresponde a la parte fija, $W = U_0 + \text{diag}(x_1 1, \dots, x_{n_1 1}, x_{1m}, \dots, x_{n_m m})$, α es un vector de parámetros que corresponde a la parte aleatoria y E es el vector de errores, que tiene una distribución normal con $E(E) = 0$ y $\text{Var}(E) = \Omega_e = \sigma_{e_0}^2 I$. También se asume que $E(\alpha) = 0$, $\text{Var}(\alpha) = \Omega_\alpha = \text{diag}(\sigma_1^2, \dots, \sigma_{p+1}^2)$ y $\text{Cov}(\alpha, E) = 0$

Para la estimación de los parámetros mediante el método de Mínimos Cuadrados Generalizados Iterativo (MCGI), procedemos al cálculo de la matriz V^{-1} , de la siguiente manera:

$$V^{-1} = \Omega_e^{-1} - \Omega_e^{-1} W C W^t \Omega_e^{-1}$$

donde $C = \Omega_\alpha^{-1} + W^t \Omega_e W$. Con la obtención de esta matriz, se encuentra una estimación de β

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

Siguiendo el mismo procedimiento descrito en la estimación del modelo con ordenada aleatoria y una variable explicativa, se obtiene de manera iterativa las estimaciones de α y β .

Otra forma de realizar las estimaciones de los parámetros β y α , es mediante la resolución del siguiente sistema [2]:

$$\begin{pmatrix} X^t \Omega_e^{-1} X & X^t \Omega_e^{-1} W \\ W^t \Omega_e^{-1} X & \Omega_\alpha^{-1} + W^t \Omega_e^{-1} W \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{\alpha} \end{pmatrix} = \begin{pmatrix} X^t \Omega_e^{-1} Y \\ W^t \Omega_e^{-1} Y \end{pmatrix}$$

1.8 Pruebas de Hipótesis

Las pruebas de hipótesis son parte importante de la interpretación. No queremos saber solamente la magnitud de los parámetros fijos y las varianzas de cada nivel, sino que también queremos saber si esos parámetros fijos son significativos, y si existe una varianza significativa en el segundo nivel.

Es por esta razón, que se deben realizar pruebas de hipótesis para determinar la significación de los coeficientes fijos y aleatorios, con el objetivo de llegar a un modelo que contenga las variables mas importantes para la explicación del problema.

1.8.1 Pruebas para los parámetros fijos

Considerando el modelo (1.9), de manera general se tiene:

$$y_{ij} = \beta_0 + \sum_{h=1}^r \beta_h x_{hij} + u_{0j} + \sum_{h=1}^p u_{hj} x_{hij} + e_{0ij}$$

donde $\beta_0 + \sum_{h=1}^r \beta_h x_{hij}$ es la parte fija del modelo y $u_{0j} + \sum_{h=1}^p u_{hj} x_{hij} + e_{0ij}$ es la parte aleatoria del modelo.

Para la parte fija del modelo, las pruebas de hipótesis consisten en comprobar si los coeficientes son estadísticamente significativos (distintos de cero). Para lo cual se plantea:

$$\begin{cases} H_0 : \beta_i = 0, & i = 0, \dots, r \\ H_1 : \beta_i \neq 0 \end{cases}$$

Para tomar una decisión se define el test estadístico ³, de la razón t:

$$t_i = \frac{\hat{\beta}_i}{ee(\hat{\beta}_i)}$$

³Esta prueba es llamada el *test de Wald*, gracias al estadístico Abraham Wald (1902 - 1950)

Donde $\hat{\beta}_i$ es el parámetro estimado, y $ee(\hat{\beta}_i)$ es el error estándar del parámetro. Si $|t_i|$ es mayor que el cuantil de orden $1 - \alpha$, entonces β_i es significativo al nivel α . Son comunes los niveles de significancia alpha (α) del 0,05, 0,01 y 0,1. A continuación se presenta el gráfico de la distribución t de Student.

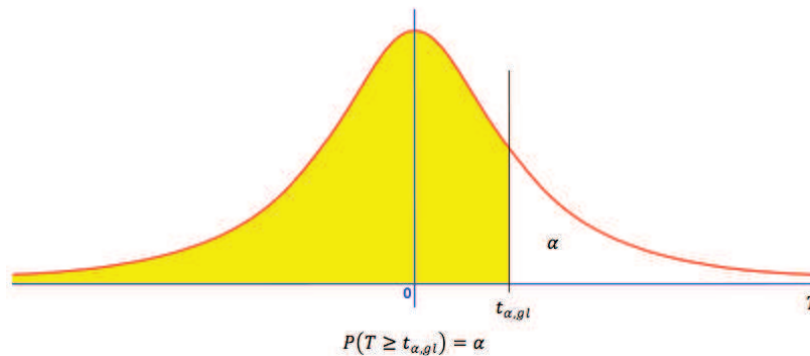


Figura 1.4: Gráfico de la distribución t de Student.

Bajo la hipótesis nula, t_i sigue aproximadamente una distribución t de Student, pero el cálculo de los grados de libertad (g.l.) es algo más complicado que en la regresión lineal múltiple, debido a que en este caso se tiene la presencia de dos niveles.

Si el número total de unidades de nivel 1 es M y el número total de variables explicativas de nivel 1 es r , los grados de libertad se calculan de la siguiente manera:

$$g.l. = M - r - 1$$

Para la prueba de coeficiente de variables de nivel 2, si N es el número de grupos y q es el número de variables explicativas de nivel 2.

$$g.l. = N - q - 1$$

Si se tiene que el número de grados de libertad es mayor que 40, la distribución t se aproxima a la distribución normal estándar.

1.8.1.1 Pruebas para multi-parámetros fijos

Para probar la significación de un grupo de parámetros, existe dos tipos de pruebas que son muy usadas para este propósito: la prueba de Wald multivariante y la prueba de la razón de verosimilitud o también llamada prueba de desviación [20].

Prueba multivariante de Wald: Supongamos que se considera un vector β con q parámetros de regresión es decir, $\beta = (\beta_1, \beta_2, \dots, \beta_q)^t$ y se desea probar la hipótesis nula:

$$\begin{cases} H_0 : \beta = 0 \\ H_1 : \beta \neq 0 \end{cases}$$

La decisión se toma mediante el estadístico de Wald:

$$W = \hat{\beta}^t \hat{\Sigma}_\beta^{-1} \hat{\beta}$$

donde $\hat{\Sigma}_\beta$ es la matriz de varianza-covarianza del vector de estimadores $\hat{\beta}$.

Bajo la hipótesis nula H_0 , la distribución de este estadístico puede ser aproximado por una distribución ji-cuadrado con q grados de libertad. Se rechaza la hipótesis H_0 , a un nivel de significación α , si el valor de W es mayor que el cuantil de orden $1 - \alpha$ de la ley ji-cuadrado con q grados de libertad.

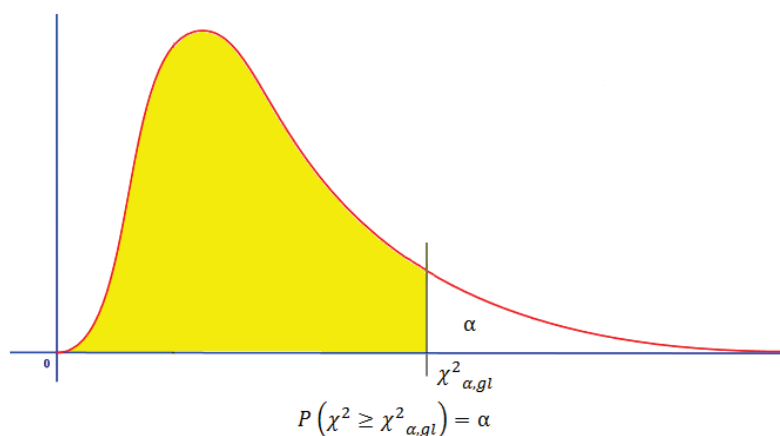


Figura 1.5: Gráfico de la distribución χ^2 .

1.8.2 Prueba de desviación

También conocida como prueba de razón de verosimilitud. En aplicaciones de modelos lineales jerárquicos es una prueba principalmente usada para pruebas de multi-parámetros y pruebas de parte aleatoria del modelo. Cuando los parámetros del modelo estadístico son estimados, se obtiene la razón de verosimilitud, que se puede transformar en la denominada *Desviación* (D) [20] que se define de la siguiente manera:

$$D(y) = -2(\log(y | \hat{\theta}_0) - \log(y | \hat{\theta}_s))$$

donde $(y | \hat{\theta}_0)$ son los valores del modelo M_1 y $(y | \hat{\theta}_s)$ son los valores de un modelo completo o más grande M_2 . La desviación sigue una ley ji-cuadrado con q grados de libertad, es decir, $\chi^2(q)$. La desviación es una medida de la falta de ajuste entre el modelo y los datos. Los valores obtenidos de esta medida sirven para comparar el ajuste de dos modelos al mismo conjunto de datos.

Sea el modelo M_1 con m_1 parámetros y el modelo más grande, M_2 con m_2 parámetros, con la adición de $m_2 - m_1$ parámetros. El modelo M_2 puede ser considerado como una extensión del modelo M_1 . Denotando por D_1 y D_2 las desviación respectivas en cada caso, la diferencia de $D_1 - D_2$ puede ser utilizado como un estadístico que sigue una distribución ji-cuadrado con $m_2 - m_1$ grados de libertad, es decir, $\chi^2(m_2 - m_1)$. Se desea probar las siguientes hipótesis:

$$\begin{cases} H_0 : \text{Los datos se ajustan al modelo } M_1 \\ H_1 : \text{Los datos se ajustan al modelo } M_2 \end{cases}$$

Se rechaza H_0 al nivel α , si el valor observado de $D_1 - D_2$ es mayor que el cuantil de orden $1 - \alpha$ de la distribución $\chi^2(m_2 - m_1)$. Esta prueba puede ser aplicada a parámetros fijos como a la parte aleatoria del modelo.

Spiegelhalter [3] usó la prueba de la *Desviación* con CMMC, que derivó en un diagnóstico llamado *Criterio de Información de Desviación (DIC)* por sus siglas en inglés. El diagnóstico de DIC es simple de calcular cuando se realiza una estimación con CMMC, ya que calcula la *Desviación* en cada iteración y el valor esperado de la desviación de los parámetros ($D(\bar{\theta})$). Además se puede calcular el número efectivo de parámetros (p_D), substrayendo $D(\bar{\theta})$ de la desviación promedio del número de iteraciones (\bar{D}). El diagnóstico de DIC puede ser usado para comparar modelos ya que se compone de dos términos que miden el ajuste y la complejidad de un modelo en particular.

$$DIC = \bar{D} + p_D = D(\bar{\theta}) + 2p_D = 2\bar{D} - D(\bar{\theta})$$

1.8.3 Pruebas de Heteroscedasticidad

Como fue expresado en la ecuación de la varianza total del modelo con ordenada y pendiente aleatorias:

$$Var(y_{ij|x_{ij}}) = \sigma_{u0}^2 + 2x_{ij}\sigma_{u01} + x_{ij}^2\sigma_{u1}^2 + \sigma_{e0}^2$$

pero los residuos e_{0ij} y u_{hj} se asume que son homoscedásticos.

Como fue explicado, se pueden representar modelos en los cuales, los residuos de nivel 1 pueden depender lineal o cuadráticamente de una variable explicativa. Probar la heteroscedasticidad de una variable es equivalente a contrastar si el coeficiente de esta variable es aleatoria.

Un método diferente [20], que puede ser usado para detectar la heteroscedasticidad es mediante el estudio de la diferencia en la varianza residual entre grupos, sin necesidad de una conexión específica con alguna variable explicativa. Este método se basa en las estimaciones de los residuos de Mínimos Cuadrados Ordinarios (MCO) dentro de cada grupo. Este método solo es aplicable si el número de grupos considerados (m) es mucho mayor que el número de variables explicativas de nivel 1. Lo siguiente se aplica a los grupos que:

$$n_j - r - 1 \geq 10, \quad j = 1, \dots, m$$

donde r es el número de variables explicativas de nivel 1 y n_j el número de individuos de cada grupo.

Para cada uno de los grupos se realiza una regresión ordinaria, donde las variables de nivel 1 son las variables explicativas. Con los resultados obtenidos en cada una de las regresiones, se denota por s_j^2 la varianza residual estimada del grupo j y por $gl_j = n_j - r - 1$ el número de grados de libertad. A continuación de esto se calcula el promedio ponderado de:

$$l_s = \frac{\sum_j gl_j \ln(s_j^2)}{\sum_j gl_j}$$

Luego se procede al cálculo de la medida de dispersión residual estandarizada, que viene dado por:

$$d_j = \sqrt{\frac{gl_j}{2} [\ln(s_j^2) - l_s]}$$

Si el modelo de nivel 1 está bien especificado y la varianza residual poblacional de nivel 1 es la misma en todos los grupos, entonces la distribución de d_j es una normal estándar. Para contrastar la hipótesis de que las varianzas residuales de nivel 1 son constantes se usa el siguiente estadístico:

$$H = \sum_j d_j^2$$

donde su distribución es una $\chi^2(g - 1)$, g es el número de grupos incluidos en la suma.

Si el resultado de la prueba es significativa, los valores individuales de d_j se los pueden analizar más detenidamente para encontrar ciertos patrones de heteroscedasticidad. También se pueden rastrear estos patrones mediante gráficos como lo son los gráficos de dispersión, gráficos de residuos estandarizados, entre otros. Una vez que se haya detectado la variable causante de la heteroscedasticidad se proceda averiguar qué tipo de relación presenta la variable con la varianza residual.

1.8.4 Estudio de residuos

El estudio de los residuos se utiliza generalmente para detectar casos atípicos que pueden influir en el resultado del modelo [20].

Cuando se realiza un análisis de residuos en modelos multinivel hay que considerar que las observaciones dependen conjuntamente de los residuos de nivel 1 y nivel 2, pero al momento de tratamiento de los residuos, se lo debe hacer de manera independiente.

El análisis de los residuos de nivel 1 se lo puede realizar con regresiones de Mínimos Cuadrados Ordinarios (MCO), calculadas dentro de cada grupo independientemente. Así se examinan los residuos de nivel 1, sin mezclar con los residuos de nivel 2.

La varianza de los residuos por MCO depende de las variables explicativas y estos residuos pueden ser estandarizados dividiéndolos para su desviación estándar. Estos residuos se los analiza de la siguiente manera:

1. Graficar los residuos no estandarizados contra las variables explicativas de nivel 1, con el fin de encontrar de posibles efectos no lineales.
2. Realizar el gráfico de probabilidad normal de los residuos estandarizados para poder corroborar las hipótesis de normalidad de los residuos y verificar la existencia de valores atípicos.

CAPÍTULO 2

Modelos de tres niveles y estructuras jerárquicas más complejas

2.1 Modelo de nivel 3

2.1.1 Introducción

En este capítulo trataremos modelos de nivel 3, que viene a ser una extensión de los modelos vistos en el capítulo anterior.

Un ejemplo de una estructura jerárquica de nivel 3 en la educación se puede presentar de la siguiente manera: estudiantes en un primer nivel, las clases en un segundo nivel y por último las escuelas en tercer nivel. De esta manera se tiene un modelo jerárquico de nivel tres, donde los estudiantes se encuentran en distintas clases y su a vez estas clases se las imparte en diferentes unidades educativas.

A la variable dependiente ahora se la notará como y_{ijk} , donde i es la unidad de nivel 1, j la unidad de nivel 2 y k la unidad de nivel 3.

A continuación se va a tratar uno de los modelos vistos en el capítulo 1, que fue el modelo con pendiente aleatoria, pero en tres niveles. La extensión para cada uno de los distintos modelos viene a ser de la misma forma, es por esto que solamente se explicará uno de ellos para no repetir nuevamente toda la explicación ya antes vista.

2.1.2 Modelo de nivel 3 con pendiente aleatoria

El modelo con pendiente aleatoria de nivel tres tiene la siguiente formulación :

$$y_{ijk} = \beta_{0jk} + \beta_1 x_{ijk} + e_{ijk}$$

$$\beta_{0jk} = \beta_{00k} + u_{0jk}$$

$$\beta_{00k} = \beta_{000} + v_{00k}$$

de forma compacta,

$$y_{ijk} = \beta_{000} + \beta_1 x_{ijk} + v_{00k} + u_{0jk} + e_{ijk}$$

donde β_{0jk} es el parámetro que varía aleatoriamente de grupo a grupo (unidades de nivel 2) en las distintas unidades de nivel 3, β_{00k} es la pendiente promedio en las distintas unidades k de nivel 3 y el efecto aleatorio a nivel 3, v_{00k} .

En la parte aleatoria se supone que:

- Las variables aleatorias e_{ijk} son independientes y tienen distribución de media cero y varianza $Var(e_{ijk}) = \sigma_{e0}^2$.
- Las variables aleatorias u_{0jk} son independientes y tienen distribución de media cero y varianza $Var(u_{0jk}) = \sigma_{u0}^2$
- Las variables aleatorias v_{00k} son independientes y tienen distribución de media cero y varianza $Var(v_{00k}) = \sigma_{v0}^2$

La varianza residual, es la siguiente:

$$Var(y_{ijk}) = \sigma_{e0}^2 + \sigma_{u0}^2 + \sigma_{v0}^2$$

En el capítulo 4, de la aplicación de los modelos se plantea un modelo lineal jerárquico de nivel 3 con el fin de explicar el mejoramiento del ajuste del modelo con la incorporación de un nuevo nivel en su estructura y presentar un modelo más general incorporando ordenada aleatoria, pendientes aleatorias y variables categóricas.

2.2 Modelos de ecuaciones estructurales multinivel

2.2.1 Introducción

La popularidad de los modelos multinivel y modelos de ecuaciones estructurales (MES) es una característica notable de la investigación cuantitativa en las ciencias médicas y sociales. Aunque desarrollados por separado y con diferentes propósitos, MES y la modelización multinivel tienen comunalidades importantes, ya que ambos enfoques son las variables latentes o efectos aleatorios para inducir, y por lo tanto explicar las correlaciones entre las respuestas.

Los modelos multinivel se utilizan cuando la estructura de datos es jerárquica, con unidades elementales en el nivel 1 anidados en grupos en el nivel 2, que a su vez pueden estar anidados en grupos de en el nivel 3, y así sucesivamente. Las variables latentes, o efectos aleatorios, se interpretan como la heterogeneidad no observada en los diferentes niveles que inducen dependencia entre todas las unidades de nivel inferior pertenecientes a una unidad de nivel superior. Las pendientes aleatorias representan heterogeneidad entre los grupos en la respuesta global y coeficientes aleatorios representan heterogeneidad en la relación entre la respuesta y las variables explicativas.

En cambio los modelos de ecuaciones estructurales son utilizados cuando las variables de interés no se puede medir perfectamente, y las variables latentes o factores son las que permiten explicar la variabilidad entre las variables observadas en términos de un número menor de variables no observadas.

Es importante destacar que los modelos de ecuaciones estructurales multinivel, son una síntesis de multinivel y modelización de ecuaciones estructurales. Estos modelos son utilizados cuando las unidades de observación forman una jerarquía de grupos anidados y algunas variables de interés deben ser medidas por una o conjunto de variables no observadas.

A continuación se va a desarrollar únicamente de manera general cómo se conforman las ecuaciones estructurales, debido a que es un tema complejo y profundo de sobrellevar. El objetivo es tener primeramente una idea clara de las ecuaciones estructurales, para después poder conectar las ideas de una mejor manera en la fusión con los modelos multinivel.

2.2.2 Análisis de senderos

El análisis de senderos estudia los efectos directos e indirectos en un conjunto de variables observables. Este análisis contempla la relación de causa-efecto [13] (efecto causal) de una variable sobre otra, entre las variables observadas.

El objetivo principal del análisis de senderos es explicar por qué las variables observadas están correlacionadas.

El análisis de senderos está compuesto por:

- Diagrama de senderos
- Modelo de senderos y Ecuaciones estructurales
- Coeficientes de Wright

Diagrama de senderos

El diagrama de senderos se encuentra representado por un gráfico, donde se encuentran representadas las relaciones de causa-efecto que se supone existen en un conjunto de variables [16].

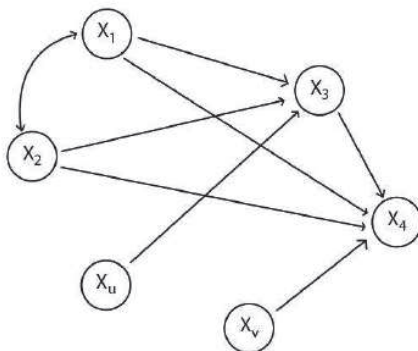


Figura 2.1: Diagrama de Senderos.

En la figura 2.1 se pueden observar las variables representadas en círculos, que se encuentran relacionadas mediante flechas en una misma dirección. La flecha va desde la variable independiente (influyente) hasta la variable dependiente (influida). También se tienen las flechas con línea curva y doble punta, que representan la supuesta correlación que existe entre las variables. Se identifican tres tipos de variables: endógenas o dependientes x_3 y x_4 , exógenas o independientes x_1 y x_2 , y

residuales x_u y x_v , que influyen en las variables x_3 y x_4 y representan a los factores no observados.

Modelo de senderos y ecuaciones estructurales

El diagrama de senderos que se explicó anteriormente puede ser planteado como un conjunto de ecuaciones estructurales, que se denomina modelo de senderos [13]. Existen dos tipos de modelos de senderos: recursivos y no recursivos. Los modelos recursivos se caracterizan porque sus residuos están no correlacionados, y sus efectos causales son unidireccionales. Los modelos no recursivos tienen lazos de retroalimentación o residuos correlacionados.

Se puede expresar a la variable dependiente en el diagrama de senderos como una función lineal de las variables independientes. El modelo de senderos de la figura 2.1 viene dado por:

$$\begin{aligned}x_3 &= \rho_{31}x_1 + \rho_{32}x_2 + \rho_{3u}x_u \\x_4 &= \rho_{41}x_1 + \rho_{42}x_2 + \rho_{43}x_3 + \rho_{4v}x_v\end{aligned}$$

donde, las x_i representan funciones de variables y los parámetros ρ_{ij} se denominan coeficientes de Wright.

Coefficientes de Wright

Como se vio anteriormente en el modelo de senderos y ecuaciones estructurales, los parámetros ρ_{ij} de las ecuaciones estructurales reciben el nombre de coeficientes de Wright [16]. El subíndice i representa la variable dependiente, y el subíndice j representa la variable independiente. Los coeficientes de Wright son las incógnitas; y su valor se determina resolviendo el sistema de ecuaciones estructurales del modelo en función de los coeficientes de correlación.

2.2.3 Análisis factorial

El análisis factorial es una técnica estadística de reducción de datos, utilizada para explicar las correlaciones entre las variables observadas en términos de un número menor de variables no observadas llamadas factores [12].

Sea un vector aleatorio observable $X^t = (X_1, X_2, \dots, X_p)$, con p variables, vector de medias μ y matriz de varianzas-covarianzas Σ . En el modelo del análisis factorial,

X son combinaciones lineales de las variables aleatorias no observadas F_1, F_2, \dots, F_m llamadas factores comunes, las variables e_1, e_2, \dots, e_p llamados errores o factores específicos y los ℓ_{ij} se denominan cargas factoriales o saturaciones factoriales. La notación matricial vendría dada por:

$$X - \mu = LF + \varepsilon$$

donde ℓ_{ij} es la carga de la i -ésima variable del factor j -ésimo, con lo que se tiene a L como la matriz de cargas factoriales. Las dimensiones son las siguientes: $(X - \mu)$ es un vector $(p \times 1)$, L es una matriz $(p \times m)$, F es un vector $(m \times 1)$ y ε es un vector $(p \times 1)$.

Sobre los factores comunes F y los factores específicos ε se presentan las siguientes hipótesis:

- F y ε son independientes
- $E(F) = 0, Cov(F) = E(FF^t) = \Phi = I_{m \times m}$
- $E(\varepsilon) = 0, Cov(\varepsilon) = E(\varepsilon\varepsilon^t) = \Psi$, donde $\Psi = [\psi_1, \psi_2, \dots, \psi_p]$ es una matriz diagonal.

Con las anteriores hipótesis se tiene que la matriz de varianzas-covarianzas Σ viene dada por:

$$\Sigma = L\Phi L^t + \Psi$$

y que,

$$Cov(X, F) = L\Phi$$

Si $\Phi = I_{m \times m}$, los factores comunes F , no están correlacionados entre sí, y este es el **Modelo Factorial Ortogonal**.

$$\begin{aligned} Var(X_i) &= \sum_{h=1}^m \ell_{ih}^2 + \psi_i, & \forall i \\ &= h_i^2 + \psi_i \end{aligned}$$

Los h_i^2 se denominan comunalidades y los ψ_i se denominan especificidades. Las comunalidades son los elementos de la diagonal principal de LL^t .

La comunalidad es la parte de la varianza de X_i explicada por los factores comunes, es decir, es la contribución de los factores comunes a la varianza de X_i . La contribución restante a la varianza de X_i proviene del factor específico ε_i y está

dada por ψ_i [12].

En el análisis factorial multinivel, asumimos que el vector de respuestas de los estudiantes puede ser descompuesto en la suma de componentes intra y entre grupos. Esto es:

$$y_{ij} = v + y_{Bj} + y_{Wj} \quad (2.1)$$

Por ejemplo tendríamos, y_{ij} es el vector de respuestas del i -ésimo estudiante en el j -ésimo colegio, v es la media total y y_{Wj} representa la contribución intra colegial (nivel estudiante) al vector de respuestas individuales. De aquí las varianzas y covarianzas de y_{ij} pueden ser descompuestas en:

$$\Sigma_T = \Sigma_B + \Sigma_W \quad (2.2)$$

Donde Σ_T es la matriz de covarianza poblacional, Σ_B es la matriz de covarianza entre colegios (nivel colegio) y Σ_W es la matriz de covarianza intra colegios (nivel estudiante).

Con lo que se tiene:

$$\begin{aligned} \Sigma_W &= \Lambda_W \Phi_W \Lambda_W^t + \Psi_W \\ \Sigma_B &= \Lambda_B \Phi_B \Lambda_B^t + \Psi_B \end{aligned}$$

donde Φ_W y Φ_B son las matrices de covarianza de los factores intra y entre colegios respectivamente, Λ_W , Λ_B son las cargas factoriales intra colegios y entre colegios, Ψ_W y Ψ_B son matrices diagonales de varianzas intra y entre colegios respectivamente.

2.2.4 Modelos de ecuaciones estructurales

Los modelos de ecuaciones estructurales es una técnica estadística multivariante, utilizada para especificar fenómenos en función de sus variables causales (causa-efecto). Dichos modelos es la incorporación de dos metodologías descritas anteriormente, el análisis de senderos y en el análisis factorial [15].

Estos modelos presentan variables no observables directamente, llamadas variables latentes o constructos, que son medidas a través de otras variables directamente observables .

Los modelos de ecuaciones estructurales como en muchos análisis presentan errores de medición, llamados errores de medida, que se dan por no medir correctamente las variables latentes mediante las variables observadas, sea esto debido a un error por parte del investigador o de los encuestados.

Las variables latentes vienen a tener el rol de los factores en el análisis factorial y además son variables medibles.

Un modelo de ecuaciones estructurales se lo puede plantear de diferentes maneras: mediante un diagrama, matricialmente o escribiendo un sistema de ecuaciones simultáneas. A continuación se presentan las variables que describen un modelo causal de ecuaciones estructurales:

- Variables latentes: endógenas η_i y exógenas ξ_j .
- Variables observadas: endógenas Y_k y exógenas X_l .
- Errores de medida: de las variables observadas endógenas ε_k y de variables observadas exógenas δ_l .
- Coeficientes de correlación θ_ε y θ_δ que correlacionan a los errores de medida.
- Términos de perturbación: ζ_m que presentan los efectos de las variables omitidas, los errores de medida y la aleatoriedad del proceso. La variación del término de perturbación viene dado por ψ y la covariación entre los términos de perturbación del i -ésimo y j -ésimo se representa por ψ_{ij} .
- Coeficientes de regresión: λ_x y λ_y que son los que relacionan a las variables latentes con las observadas.
- Coeficientes de regresión: γ_{ij} y β_{ji} que relacionan las variables latentes y las variables observadas entre sí respectivamente.

El modelo estructural viene dado por una serie de relaciones que se presentan como un conjunto de ecuaciones lineales relacionando variables latentes endógenas con variables latentes exógenas. Cada variable latente endógena es la variable dependiente de la ecuación y el resto de variables latentes endógenas y exógenas relacionadas son las variables independientes.

El modelo de ecuaciones estructurales, está compuesto por un *modelo estructural* y un *modelo de medida*.

Modelo Estructural

Este modelo se lo representa de la siguiente manera en forma de ecuación:

$$\eta = B\eta + \Gamma\xi + \zeta \quad (2.3)$$

donde: η vector ($m \times 1$) de variables latentes endógenas, B matriz ($m \times m$) de coeficientes de variables endógenas, Γ matriz ($m \times k$) de coeficientes de variables exógenas, ξ vector ($k \times 1$) de variables latentes exógenas, ζ vector ($m \times 1$) de términos de perturbación aleatoria [15].

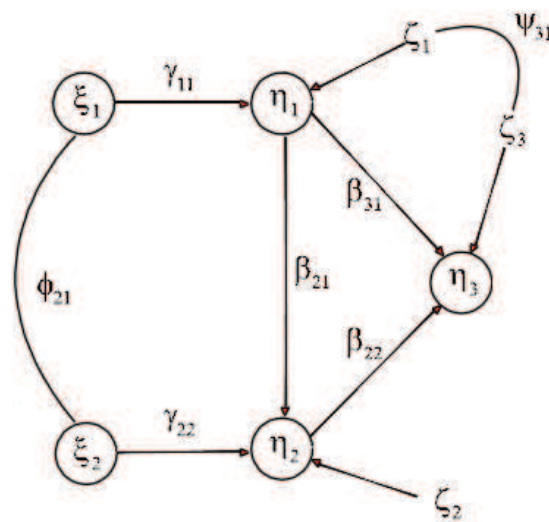


Figura 2.2: Modelo Estructural

La representación en forma de ecuación para el modelo de la figura anterior es:

$$\eta_1 = \gamma_{11}\xi_1 + \zeta_1$$

$$\eta_2 = \gamma_{22}\xi_2 + \beta_{21}\eta_1 + \zeta_2$$

$$\eta_3 = \beta_{31}\eta_1 + \beta_{32}\eta_2 + \zeta_3$$

En forma matricial, el modelo estructural viene dado por:

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} + \begin{pmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix}$$

Modelo de Medida

Especifica las variables observadas, que serán utilizadas para medir las hipótesis planteadas en el modelo estructural. El modelo de medida se lo representa en ecuaciones de la siguiente manera, separando lo exógeno (x) de lo endógeno (y):

$$y = \Lambda_y \eta + \varepsilon \quad (2.4)$$

$$x = \Lambda_x \xi + \delta \quad (2.5)$$

donde η es el vector de tamaño $(m \times 1)$ de variables latentes endógenas, ξ es el vector $(k \times 1)$ de variables latentes exógenas, Λ_x es la matriz $(q \times k)$ de coeficientes de variables exógenas, Λ_y es la matriz $(p \times m)$ de coeficientes de variables endógenas, δ es el vector $(q \times 1)$ de errores de medición para los indicadores exógenos, ε vector $(p \times 1)$ de errores de medición para los indicadores endógenos [14].

Los modelos de ecuaciones estructurales son un caso especial del modelo de estructura de covarianza general. La matriz de covarianza para y , se obtiene sustituyendo la ecuación (2.3) en la ecuación (2.4), como se indica a continuación.

$$y = \Lambda_y (B + \Gamma \xi + \zeta) + \varepsilon$$

$$y = \Lambda_y B + \Lambda_y \Gamma \xi + \Lambda_y \zeta + \varepsilon$$

Haciendo que las variables endógenas estén en un lado de la ecuación y las variables exógenas en el otro lado, se tiene:

$$(I - B)y = \Lambda_y \Gamma \xi + \Lambda_y \zeta + \varepsilon$$

con I de dimensión $(m \times m)$. Suponiendo que $(I-B)$ es no singular, es decir su inversa existe, la ecuación anterior puede reescribirse como:

$$y = \Lambda_y (I - B)^{-1} \Gamma \xi + \Lambda_y (I - B)^{-1} \zeta + (I - B)^{-1} \varepsilon$$

Sea Ω un vector que contiene los parámetros estructurales del modelo, en este caso $\Omega = (B, \Gamma, \Psi, \Phi)$.

Además se tiene que,

$$cov(x) = E(xx^t) = \Phi$$

$$E(\zeta) = 0$$

$$\begin{aligned}
\Sigma_{xx} &= cov(xx^t) \\
&= E[(\Lambda_x \xi + \delta)(\xi^t \Lambda_x^t + \delta^t)] \\
&= \Lambda_x E(\xi \xi^t) \Lambda_x^t + E(\delta \delta^t) \\
&= \Lambda_x \Phi \Lambda_x^t + \Theta_\delta
\end{aligned}$$

De esta manera se tiene:

$$\begin{aligned}
\Sigma &= \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix} \\
&= \begin{pmatrix} \Lambda_y (I - B)^{-1} (\Gamma \Phi \Gamma^t + \Psi) (I - B)^{t-1} \Lambda_y + \Theta_\varepsilon & \Lambda_y (I - B)^{-1} \Gamma \Phi \Lambda_x \\ \Lambda_x \Phi (I - B)^{t-1} \Lambda_y^t & \Lambda_x \Phi \Lambda_x^t + \Theta_\delta \end{pmatrix}
\end{aligned}$$

donde: Φ matriz de covarianzas ($k \times k$) de variables latentes exógenas, Ψ matriz de covarianzas ($m \times m$) de los términos de perturbación, Θ_ε matriz de covarianzas de los errores de medición ε y Θ_δ matriz de covarianzas de los errores de medición δ [5].

Identificación del modelo

De igual manera que un sistema de ecuaciones lineales, es posible que se tengan más ecuaciones que incógnitas o también tener más incógnitas que ecuaciones.

Se dice que un modelo se encuentra *identificado* si sus grados de libertad son 0, si se cumple esta condición es posible realizar la estimación de los parámetros [10].

La diferencia que existe entre el número de correlaciones y el número de coeficientes del modelo se denomina grados de libertad. Para realizar el cálculo de los grados de libertad en un modelo de ecuaciones estructurales se lo realiza de la siguiente manera:

$$g.l. = \frac{1}{2}[(p + q)(p + q + 1)] - t$$

donde p es el número de indicadores endógenos, q es el número de indicadores exógenos y t es el número de coeficientes estimados. Según los grados de libertad

se tiene los siguientes casos:

- Si los g.l. son iguales a cero el modelo está *identificado*, es decir se puede realizar la estimación de los parámetros.
- Si los g.l. son mayores a cero el modelo está *sobreidentificado*, es decir existe más información en la matriz de datos que parámetros a estimar.
- Si los g.l. son menores a cero el modelo está *subidentificado*, es decir los parámetros de este modelo no pueden ser estimados hasta que se fije o se restrinja algún o algunos parámetros en el modelo.

Estimación

En la estimación de los parámetros del modelo, la selección de los datos de entrada es importante. La matriz de varianzas-covarianzas o la matriz de correlaciones de las variables observadas pueden ser utilizadas como datos iniciales para los modelos de ecuaciones estructurales.

Se consideran, las ecuaciones de los submodelos estructural y de medida que conforman un modelo de ecuaciones estructurales:

$$\begin{aligned}\eta &= B\eta + \Gamma\xi + \zeta \\ y &= \Lambda_y\eta + \varepsilon \\ x &= \Lambda_x\xi + \delta\end{aligned}$$

Los parámetros del modelo son:

- Las varianzas y covarianzas de las variables exógenas de Φ (matriz de covarianza $(k \times k)$ de variables latentes exógenas).
- Las varianzas y covarianzas de los términos de perturbación de Ψ (matriz de covarianzas $(m \times m)$ de los términos de perturbación).
- Los coeficientes de regresión de B (matriz $(m \times m)$ de coeficientes de las variables endógenas) y Γ (matriz $(m \times k)$ de coeficientes de las variables exógenas).

Además se debe considerar al vector de parámetros, denotado como Ω , que contiene los parámetros antes descritos juntos:

$$\Omega = (\Lambda_y, \Lambda_x, \Theta_\varepsilon, \Theta_\delta, \Phi, B, \Gamma, \Psi)$$

La finalidad es obtener las estimaciones del vector de parámetros $\hat{\Omega}$, de tal manera que minimice la función $F(S, \hat{\Sigma})$, donde $\hat{\Sigma} = \Sigma(\hat{\Omega})$ es la matriz de varianza-covarianza de las estimaciones del modelo. Esta función, $F(S, \hat{\Sigma})$, mide la distancia entre la matriz de varianza-covarianza muestral S y la matriz de varianza-covarianza ajustada $\hat{\Sigma}$.

La función es la siguiente:

$$F = [S - \Sigma(\Omega)]^t W^{-1} [S - \Sigma(\Omega)]$$

Donde: S es la matriz de varianzas-covarianzas de la muestra, $\Sigma(\Omega)$ es la matriz de varianzas-covarianzas estimada, W es una matriz de ponderaciones.

Además si se propone como hipótesis que la distribución muestral de las variables es normal multivariante, entonces se tiene:

$$F = \frac{1}{2} \text{Traza}[(S - \Sigma(\Omega))W_2]^2$$

Donde S es la matriz de varianzas-covarianzas de la muestra, $\Sigma(\Omega)$ es la matriz de varianzas-covarianzas, W_2 es una matriz que puede tener diferentes formas dependiendo del tipo de método de estimación que se escoja.

Método de Mínimos Cuadrados Generalizados

Con el método de MCG se obtiene estimadores no sesgados, aunque cuando se tiene muestras pequeñas dichos estimadores tiene un sesgo próximo a cero.

Asumiendo que los datos tienen una distribución normal multivariante. De manera general la función de ajuste es la siguiente:

$$F_{MCG} = [S - \Sigma(\Omega)]^t W^{-1} [S - \Sigma(\Omega)]$$

donde W^{-1} es una matriz de ponderaciones, que pondera las desviaciones $S - \Sigma(\Omega)$ en términos de sus varianzas y covarianzas con otros elementos.

Con $W^{-1} = S^{-1}$, la función de ajuste de los MCG se puede escribir:

$$\begin{aligned} F &= \frac{1}{2} \text{Traza}[S^{-1}(S - \Sigma(\Omega))]^2 \\ &= \frac{1}{2} \text{Traza}[I - S^{-1}\Sigma(\Omega)]^2 \end{aligned}$$

2.2.5 Análisis Factorial Confirmatorio (AFC)

Es una extensión de las ecuaciones estructurales y de igual manera establece las relaciones entre variables observadas y las variables latentes establecidas inicialmente por el investigador.

En este análisis el investigador tiene una hipótesis a priori de la estructura que tendría el diagrama de senderos. A continuación se presenta un ejemplo.

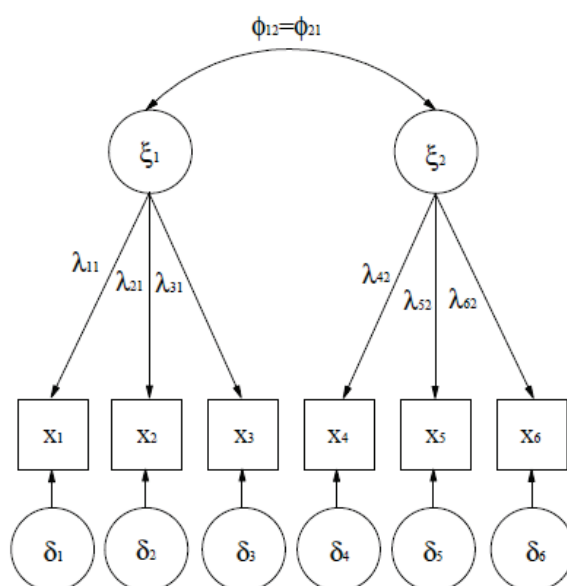


Figura 2.3: Ejemplo de Análisis Factorial Confirmatorio (AFC).

Donde las variables observadas x_i se encuentran representadas en cuadrados, las variables latentes exógenas o no observables están representadas por círculos ξ_j y se los denomina *factores comunes* debido a que comparten sus efectos con más de una variable observada, los λ_{ij} que aparecen en cada relación causal (causa-efecto) denominas cargas factoriales y los errores de medición de las variables observadas exógenas, denominadas factores específicos δ_i . Por último la flecha curva de doble punta que une a los factores comunes indica la correlación que existe entre dichos factores, ϕ [1].

La estructura presentada estaría representada de la siguiente manera:

$$x_1 = \lambda_{11}\xi_1 + \delta_1$$

$$x_2 = \lambda_{21}\xi_1 + \delta_2$$

$$x_3 = \lambda_{31}\xi_1 + \delta_3$$

$$x_4 = \lambda_{42}\xi_2 + \delta_4$$

$$x_5 = \lambda_{52}\xi_2 + \delta_5$$

$$x_6 = \lambda_{62}\xi_2 + \delta_6$$

en forma matricial se tendría:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{pmatrix}$$

Es decir, el AFC satisface la siguiente ecuación:

$$x = \Lambda_x \xi + \delta$$

donde x es un vector de dimensión $(q \times 1)$ con q variables observadas, ξ es un vector $(k \times 1)$ con s factores comunes, Λ_x es una matriz de tamaño $(q \times k)$ que contiene las cargas factoriales y δ es un vector de tamaño $(q \times 1)$ de errores o factores específicos.

Como se explicó en *Modelos de ecuaciones estructurales* se asume que ξ tiene una distribución $N(0, I)$, δ tiene una distribución $N(0, \Theta_\delta)$ y δ no está correlacionada con ξ .

El vector x tiene una distribución $N(0, \Sigma)$,

$$\Sigma = \Lambda_x \Phi \Lambda_x^t + \Theta_\delta$$

Donde Φ es la matriz de covarianzas $(k \times k)$ de las variables latentes exógenas y Θ_δ es la matriz de covarianzas de los errores de medición o factores específicos.

Este análisis puede ser llevado a diferentes niveles, por ejemplo podemos realizar el AFC a nivel estudiante donde se establezca a priori x factores y a nivel colegio realizar un nuevo AFC estableciendo y factores. De esta manera a distintos niveles se pueden plantear diferentes factores comunes que expliquen las variables observadas en los diferentes niveles.

CAPÍTULO 3

Estimación Bayesiana usando Cadenas de Márkov Monte Carlo

3.1 Introducción

La modelización multinivel y los métodos de Cadenas de Márkov Monte Carlo (CMMC), son dos áreas de la estadística que han venido creciendo recientemente debido a las mejoras en las capacidades computacionales en almacenamiento y en velocidad de operación.

Las principales características que motivan la utilización de CMMC es que obtiene las distribuciones de los parámetros de interés y se puede extender el procedimiento a modelos complejos con datos que no siguen las distribuciones estándar. Los algoritmos de CMMC son un grupo de los métodos Bayesianos [8] que pueden ser usados para ajustar los modelos multinivel. En este capítulo se describirán dos de los algoritmos de CMMC más comunes usados hoy en día que son el muestreador de Gibbs y el algoritmo de Metropolis-Hastings (MH). Estos métodos incorporan suposiciones de distribuciones a priori y, en base a sucesivas distribuciones a posteriori de los parámetros del modelo, se produce una “cadena”, que se puede utilizar para la construcción de estimaciones de intervalos.

Estos algoritmos CMMC son iterativos y en cada iteración están diseñados para producir una muestra de una distribución (multivariante) posterior conjunta de los componentes o parámetros del modelo. Estos parámetros serán coeficientes de regresión, matrices de covarianza, residuos, etc.

Sea el modelo de nivel 2 de componentes de la varianza (modelo con ordenada

aleatoria, con una variable explicativa), propuesto en el capítulo 1.

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{0ij}$$

En la formulación Bayesiana combina la información a priori de los parámetros fijos y aleatorios, con la probabilidad basada en los datos. Estos parámetros son considerados como variables aleatorias descritas por distribuciones de probabilidad, y la información a priori para un parámetro se incorpora en el modelo a través de una distribución a priori.

Después de ajustar el modelo, la distribución producida se conoce como la distribución a posteriori; esto se explica en el teorema de Bayes.

Teorema de Bayes: En la teoría de la probabilidad el teorema de Bayes es un resultado enunciado por Thomas Bayes en 1763, que expresa la probabilidad condicional de un evento aleatorio A dado B en términos de la distribución de probabilidad condicional del evento B dado A y la distribución de probabilidad marginal de sólo A [9].

Sea $Y = (y_1, y_2, \dots, y_n)$ un vector de n respuestas observadas cuya distribución de probabilidad depende de k parámetros en el vector $\theta = (\theta_1, \theta_2, \dots, \theta_k)$. Entonces la distribución conjunta de θ e Y es:

$$Pr(y, \theta) = Pr(y|\theta)Pr(\theta) = Pr(\theta|y)Pr(y)$$

de donde la distribución de probabilidad condicional de θ dado el vector de observaciones Y resulta:

$$Pr(\theta|y) = \frac{Pr(y|\theta)Pr(\theta)}{Pr(y)} \quad (3.1)$$

con la $Pr(y) \neq 0$

A la ecuación (3.1) se le conoce como el teorema de Bayes, donde la $Pr(y)$ es la distribución de probabilidad marginal de Y y puede ser expresada por:

$$Pr(y) = \begin{cases} \int_{-\infty}^{\infty} Pr(y|\theta)Pr(\theta)d\theta & \text{si } \theta \text{ es continuo} \\ \sum_{j=1}^n Pr(y|\theta_j)Pr(\theta_j) & \text{si } \theta_j \text{ es discreto} \end{cases}$$

donde la suma o integral es tomada sobre el espacio paramétrico de θ . De este

modo, el teorema de Bayes [6] puede ser escrito como :

$$Pr(\theta|y) = cPr(y|\theta)Pr(\theta) \approx Pr(y|\theta)Pr(\theta)$$

donde $Pr(\theta)$ es la distribución a priori de θ , la $Pr(\theta|y)$ es la distribución a posteriori de θ dado Y y c es una constante normalizadora necesaria para que $Pr(\theta|y)$ sume o integre según el caso.

Dado que el vector de datos Y es conocido a través de la muestra, $Pr(Y|\theta)$ es una función de θ y no de Y . En este caso la $Pr(Y|\theta)$ se le denomina función de verosimilitud de θ dado Y y se le denota por $l(\theta|y)$. Entonces la formula de Bayes puede ser expresada como:

$$Pr(y|\theta) \approx l(\theta|y)Pr(\theta)$$

Ejemplo: Sea el parámetro θ que a priori tiene una distribución uniforme en el intervalo $[0, 1]$ y la variable aleatoria Y tiene una distribución de probabilidad binomial con parámetros m y θ , con m conocido. Entonces se tienen las siguientes funciones de distribución:

$$Pr(\theta) = 1 \quad 0 \leq \theta \leq 1$$

$$Pr(y|\theta) = \binom{m}{y} \theta^y (1 - \theta)^{m-y} \quad y = 0, 1, \dots, m$$

Para una muestra aleatoria de tamaño n la función de verosimilitud viene dada por:

$$l(\theta|y) = \prod_{i=1}^n \binom{m}{y_i} \theta^{\sum y_i} (1 - \theta)^{nm - \sum y_i} \quad y_i = 0, 1, \dots, m$$

Aplicando el Teorema de Bayes, la distribución a posteriori de θ es:

$$Pr(\theta|y) = c \frac{n(m!)}{\prod_{i=1}^n y_i! \prod_{i=1}^n (m - y_i)!} \theta^{\sum y_i} (1 - \theta)^{mn - \sum y_i}$$

La ecuación anterior puede reescribirse de la siguiente manera,

$$Pr(\theta|y) = c \frac{n(m!)}{\prod_{i=1}^n y_i! \prod_{i=1}^n (m - y_i)!} \theta^{(\sum y_i + 1) - 1} (1 - \theta)^{(mn - \sum y_i + 1) - 1}$$

que tiene la forma de una distribución beta con parámetros $(\sum y_i + 1)$ y

$(nm - \sum y_i + 1)$ y la constante normalizadora c será:

$$c = \frac{\Gamma(nm + 2)}{\Gamma(\sum y_i + 1)\Gamma(nm - \sum y_i + 1)} \frac{\prod_{i=1}^n y_i! \prod_{i=1}^n (m - y_i)!}{n(m)!}$$

La distribución a priori es muy importante en el análisis bayesiano ya que mide el grado de conocimiento inicial de los parámetros en estudio. El uso de una u otra distribución a priori determinará diferencias en la distribución a posteriori. Si se tiene un conocimiento previo sobre los parámetros, este se traducirá en una distribución a priori.

Sin embargo, cuando nada es conocido sobre los parámetros, la selección de una distribución a priori adecuada adquiere una connotación especial pues será necesario elegir una distribución a priori que no influya sobre ninguno de los posibles valores de los parámetros en cuestión. Estas distribuciones a priori reciben el nombre de *difusas o no informativas*.

3.2 Distribuciones a priori por defecto

En estadística Bayesiana, todo parámetro desconocido tiene que tener una distribución a priori. Esta distribución describirá toda la información conocida sobre el parámetro. A menudo poco se conoce sobre los parámetros a priori, y por esta razón las distribuciones a priori por defecto son requeridas para que expresen esta falta de conocimiento. Las distribuciones a priori por defecto aplicadas en el paquete estadístico MLwin [17] con estimaciones de CMMC, son difusas para todos los parámetros. A continuación se describen las difusas a priori que son usadas:

- Para los parámetros fijos $p(\beta) \propto 1$. Esta uniforme a priori es equivalente a una Normal a priori con varianza c^2 , donde c es extremadamente grande con respecto a la escala del parámetro.
- Para las varianzas escalares, $p(\frac{1}{\sigma^2}) \sim \Gamma(\varepsilon, \varepsilon)$, donde ε es muy pequeño.
- Para las matrices de varianza $p(\Omega^{-1}) \sim Wishart_p(p, \hat{\Omega})$ donde p es el número de filas en la matriz de varianza y $\hat{\Omega}$ es una estimación de los verdaderos valores de Ω .

3.3 Muestreador de Gibbs

El primer método de CMMC es el muestreador de Gibbs que trabaja simulando un nuevo valor para cada parámetro (o bloque de parámetros) de turno, de la distribu-

ción condicional se asume que los valores actuales de los parámetros son verdaderos.

Idealmente si se muestrean todos los parámetros juntos en un solo bloque, se va a tener un muestreo independiente. Muestrear los parámetros individualmente induce a la dependencia de las cadenas de los parámetros produciendo autocorrelación entre los parámetros.

Primeramente se necesitan inicializar los valores de cada parámetro; en MLwin estos son tomados de los valores actuales almacenados antes de iniciar la estimación con CMMC-Gibbs. Por esta razón es importante correr la estimación con MCGI antes de correr la estimación con CMMC para dar a este método valores iniciales buenos. El método funciona mediante un muestreo de la siguiente distribución posterior condicional. Por ejemplo:

1. $p(\beta_0|y, \beta_1(0), \sigma_e^2(0))$ para generar $\beta_0(1)$, y entonces de
2. $p(\beta_1|y, \beta_0(1), \sigma_e^2(0))$ para generar $\beta_1(1)$, y entonces de
3. $p(\sigma_e^2|y, \beta_0(1), \beta_1(0))$ para generar $\sigma_e^2(1)$

Habiendo realizado estos pasos se han actualizado todos los valores desconocidos en el modelo. Este procedimiento es repetido muchas veces usando el conjunto de valores de los parámetros generados previamente para generar el siguiente conjunto. La cadena de valores generada por este procedimiento es conocida como *Cadena de Markov*, donde cada nuevo valor generado para un parámetro solo depende de su valor previo.

Se han inicializado las cadenas con valores iniciales particulares, por lo general tomará un tiempo para que las cadenas se estabilicen (converjan). El periodo cuando las cadenas se estabilizan es normalmente llamado *periodo de quemado*. En MLwin por defecto se corre un periodo de quemado de 500 iteraciones.

En MLwin los a priori por defecto son: $p(\beta_0) \propto 1$, $p(\beta_1) \propto 1$, $p(1/\sigma_e^2) \sim \Gamma(\varepsilon, \varepsilon)$, donde $\varepsilon = 10^{-3}$. A continuación el método de CMMC-Gibbs procede a través de una serie de pasos que se detalla a continuación:

Paso 1: Tomar una muestra del conjunto de efectos fijos (β), la distribución posterior condicional viene dada por:

$$p(\beta|y, \sigma_u^2, \sigma_e^2, u) = l(y; \beta, u, \sigma_e^2)p(\beta)$$

un adecuado difuso a priori es la $p(\beta) = 1$, de esta manera se tiene,

$$p(\beta|y, \sigma_u^2, \sigma_e^2, u) = \left(\frac{1}{\sigma_e^2}\right)^{N/2} \prod_{ij} \exp\left[-\frac{1}{2\sigma_e^2}(y_{ij} - u_j - (X\beta)_{ij})^2\right]$$

de manera que la muestra de β sigue una distribución normal multivariante (por sus siglas en inglés MVN), con media $\hat{\beta}$ y varianza \hat{D} ,

$$\hat{\beta} = \left[\sum_{ij} X_{ij}^T X_{ij}\right]^{-1} \left[\sum_{ij} X_{ij}^T (y_{ij} - u_j)\right]$$

$$\hat{D} = \sigma_e^2 \left[\sum_{ij} X_{ij}^T X_{ij}\right]^{-1}$$

Otras suposiciones respecto a la información a priori se pueden considerar, como por ejemplo que $p(\beta) \sim MVN(0, V)$, con lo que tenemos:

$$\beta \sim MVN(\hat{\beta}^*, \hat{D}^*)$$

$$\hat{\beta}^* = \left[\sum_{ij} X_{ij}^T X_{ij} + \sigma_e^2 V^{-1}\right]^{-1} \left[\sum_{ij} X_{ij}^T \tilde{y}_{ij}\right]$$

$$\hat{D}^* = \sigma_e^2 \left[\sum_{ij} X_{ij}^T X_{ij} + \sigma_e^2 V^{-1}\right]^{-1}$$

Para cualquiera de las dos suposiciones respecto a la información a priori, la distribución a posteriori existe.

Paso 2: Tomar una muestra de un nuevo conjunto de residuos.

Cada residuo u_j se asume que tiene una distribución a priori $u_j \sim N(0, \sigma_u^2)$, que conduce a la siguiente distribución a posteriori, donde n_j es el número de unidades de nivel 1 en la j-ésima unidad de nivel 2.

$$p(u_j|y, \sigma_u^2, \sigma_e^2) = \left(\frac{1}{\sigma_e^2}\right)^{n_j/2} \prod_{i=1}^{n_j} \exp\left[-\frac{1}{2\sigma_e^2}(y_{ij} - (X\beta)_{ij} - u_j)^2\right] \left(\frac{1}{\sigma_u^2}\right)^{1/2} \exp\left[-\frac{1}{2\sigma_u^2}u_j^2\right]$$

de manera que la muestra de $u_j \sim N(\hat{u}_j, \hat{D})$,

$$\hat{u}_j = [n_j + \sigma_e^2 \sigma_u^{-2}]^{-1} \left[\sum_{i=1}^{n_j} (y_{ij} - (X\beta)_{ij})\right]$$

$$\hat{D} = \sigma_e^2 [n_j + \sigma_e^2 \sigma_u^{-2}]^{-1}$$

Paso 3: Tomar una muestra de una nueva varianza de nivel 2.

Surge un problema al momento de la elección de una adecuada distribución a priori. A continuación mencionaremos dos posibilidades, una a priori uniforme, en el caso de los parámetros fijos y la otra posibilidad es una gamma inversa a priori. La elección de una a priori uniforme tiende a producir estimaciones sesgadas, debido a que no es una distribución a priori informativa. En lugar de tomar una muestra de la varianza directamente, asumiremos una distribución a priori Gamma, es decir $p(\sigma_u^{-2}) \sim \text{gamma}(\varepsilon, \varepsilon)$. De manera que la muestra de $\sigma_u^{-2} \sim \text{gamma}(a_u, b_u)$ con,

$$a_u = (J + 2\varepsilon)/2 \quad b_u = (\varepsilon + \sum_{j=1}^J u_j^2/2)$$

donde J es el número de unidades de nivel 2. Para una a priori uniforme para σ_u^2 se tiene una distribución Gamma con,

$$a_u = (J - 2)/2 \quad b_u = \sum_{j=1}^J u_j^2/2$$

Cuando los coeficientes aleatorios son mayores que uno, se debe muestrear una nueva matriz de covarianza Ω_u ; para lo que se tiene dos opciones para la estimación de la varianza una a priori uniforme o una Wishart inversa a priori. Para lo cual la muestra de $\Omega_u^{-1} \sim \text{Wishart}(v_u, S_u)$, con

$$v_u = J + v_p \quad S_u = \left(\sum_{j=1}^J u_j^T u_j + S_p \right)^{-1}$$

donde u_j es el vector fila de residuos para la j-ésima unidad de nivel 2 y el a priori $p(\Omega_u^{-1}) \sim \text{Wishart}(v_p, S_p)$.

Se puede utilizar una estimación aproximada de máxima verosimilitud para S_p . Escoger $v_p = -3$, $S_p = 0$ es equivalente a escoger una a priori uniforme para Ω_u .

Paso 4: Tomar una muestra de una nueva varianza de nivel 1.

Es similar al procedimiento realizado para la varianza de nivel 2. La muestra de $\sigma_e^{-2} \sim \text{gamma}(a_e, b_e)$ con

$$a_e = (N + 2\varepsilon)/2 \quad b_e = (\varepsilon + \sum_{ij} e_{ij}^2/2)$$

Para una a priori uniforme se tiene una distribución Gamma con:

$$a_e = (N - 2)/2 \quad b_e = \sum_{ij} e_{ij}^2/2$$

Los primeros valores de la Cadena de Markov Monte Carlo generalmente serán descartados, y la inferencia se basa en el conjunto obtenido después de la cadena que ha convergido a la distribución posterior conjunta. La cadena de parámetros se trata como una muestra aleatoria de la distribución posterior conjunta de los parámetros.

3.4 Muestreador de Metropolis - Hastings (MH)

El muestreador de Gibbs se puede utilizar siempre y cuando se pueden escribir analíticamente las distribuciones condicionales posteriores para cada parámetro o grupo de parámetros a la vez. Cuando esto no se puede hacer, existen enfoques alternativos que pueden ser utilizados.

Un caso importante donde utilizar el muestreador MH es con respuestas discretas, o los modelos multinivel lineales generalizados. Las distribuciones condicionales de los parámetros fijos y los residuales implican funciones exponenciales y estas no son fáciles de muestrear, por lo que otro tipo de muestreo como MH es necesario.

Para uno o más parámetros en una fase determinada del ciclo iterativo, MH procede formando una distribución propuesta de los parámetros y toma una muestra un nuevo conjunto de parámetros de esta distribución. Se aplica una regla para aceptar o rechazar este nuevo conjunto, y si el conjunto propuesto es rechazado se retienen los valores actuales. Esto se repite para cada conjunto de parámetros de turno, debido a que se producen cadenas para los distintos parámetros. Sería ideal tener una distribución propuesta que minimice el número de iteraciones para una precisión dada. Un enfoque sencillo a la creación de una distribución propuesta es una muestra de una distribución normal multivariante. Cuando esta distribución es centrada alrededor de los valores de los parámetros actuales, esto se conoce como *muestreador Metropolis de marcha aleatoria*.

Consideremos el *Paso 1* del muestreador de Gibbs para los coeficientes fijos. Dado los valores actuales de los coeficientes fijos se forma una distribución propuesta y se toma una muestra de un nuevo conjunto de coeficientes a partir de esta distribución. A continuación, se aplica una regla para aceptar este nuevo conjunto

o rechazarlo y permanecer con los valores actuales. Idealmente, se desea tener una distribución propuesta que minimice el número de iteraciones para una precisión dada. Se va a considerar cómo esto puede conseguirse después de describir el principio básico. Un enfoque sencillo es tomar una muestra de una distribución normal multivariante de la forma siguiente:

$$\beta_{(t)} \sim MVN(\beta_{(t-1)}, D)$$

donde D , por ejemplo, es una estimación de la matriz de covarianza de las estimaciones de los coeficientes. Después de haber obtenido un nuevo conjunto β^* , necesitamos una regla de aceptación y esto se basa en la siguiente relación

$$r_t = \frac{p(\beta^*, \sigma_u^2, \sigma_e^2, u|y)}{p(\beta_{(t-1)}, \sigma_u^2, \sigma_e^2, u|y)}$$

donde estas densidades de probabilidad tiene la misma forma que las probabilidades condicionales descritas en el paso 1. Este procedimiento, con dichas condiciones de la distribución propuesta de los valores actuales de los parámetros es conocida como *muestreador Metropolis de marcha aleatoria*.

Además se tiene que la distribución propuesta normal multivariante es simétrica en $\beta_{(t)}, \beta_{(t-1)}$; esto es $p(\beta_{(t)} = a | \beta_{(t-1)} = b) = p(\beta_{(t)} = b | \beta_{(t-1)} = a)$. Donde distribuciones asimétricas se utilizan en la relación anterior y es multiplicada además por la llamada *relación Hastings*.

$$\frac{p(\beta_{(t-1)} | \beta^*)}{p(\beta^* | \beta_{(t-1)})}$$

Se define la probabilidad de aceptación para el nuevo conjunto como

$$\alpha_t = \min(1, r_t)$$

Si es 1, entonces el nuevo conjunto es aceptado; si es menor que 1, entonces se acepta con probabilidad α_t . Por ejemplo, generar un número aleatorio uniforme en $(0, 1)$ y la aceptación de si este es menor que α_t . Si el conjunto nuevo no se acepta entonces el conjunto $(t - 1)$ se utiliza. Así, a diferencia del muestreador de Gibbs, es posible que la cadena permanezca con el mismo conjunto de valores por varias iteraciones.

Un muestreador MH eficiente evita esto en lo posible, a pesar de que el muestreador MH, en principio, producirá estimaciones satisfactorias si corre un tiempo suficiente; la clave es encontrar una distribución propuesta oportuna. Un procedi-

miento algo más satisfactorio [6] para especificar una tasa de aceptación deseada (por ejemplo 50 %) y ajustar la escala de factores de la distribución propuesta para cada uno de los parámetros. Esto se puede hacer durante una etapa preliminar de adaptación de la estimación de la cadena CMMC antes o durante el proceso.

En general, el muestreo de varianzas y covarianzas se puede llevar a cabo utilizando Gibbs y esto da un algoritmo híbrido que es una mezcla de MH y pasos de muestreo de Gibbs.

CAPÍTULO 4

Aplicación de los modelos

4.1 Introducción

Para la realización de la aplicación se procedió a tomar pruebas de Matemática, Lenguaje y Razonamiento a estudiantes de décimo año de Educación Básica y para el tercer año de Bachillerato, adicionalmente de las pruebas mencionadas anteriormente se añadió un complemento de comprensión lectora en la prueba de Lenguaje. Estas pruebas fueron tomadas en los establecimientos del cantón Otavalo de la provincia de Imbabura. El número de establecimientos donde fueron aplicadas las pruebas son 12, de los cuales 5 establecimientos son particulares y 7 son fiscales.

Las pruebas fueron aplicadas a 838 estudiantes del cantón Otavalo, de los cuales 607 son de décimo año de Educación Básica y 231 de tercer año de Bachillerato.

Las pruebas fueron elaboradas en función de los contenidos que dispone el Ministerio de Educación para tercer año de Bachillerato y décimo año de Educación Básica. Estas pruebas pasaron a ser revisadas y aprobadas por la Dirección Provincial de Educación de Imbabura antes de su aplicación (Anexo B).

Cada una de las pruebas estuvo conformada por diez preguntas y para su contestación los estudiantes tuvieron: 60 minutos para la prueba de Matemáticas, 20 minutos para la prueba de Lenguaje-Comprensión lectora y 10 minutos para la prueba de Razonamiento. Estas pruebas (Anexo C) fueron aplicadas por el autor de la presente investigación a los estudiantes de las diferentes instituciones educativas en el cantón Otavalo, y fueron aplicadas durante un mes y medio.

Una vez recolectada toda la información se procedió a construir los modelos jerárquicos de nivel 2 y nivel 3, además de un modelo de ecuaciones estructura-

les multinivel. Esto se realizó con distintas variables explicativas a diferentes niveles hasta llegar a conformar modelos que permitan observar la variabilidad de los resultados de las pruebas de los distintos estudiantes entre los establecimientos educativos y dentro de ellos. Las variables son descritas a continuación:

Variable	Descripción
Colegio	Identificador numérico del colegio. Variable de segundo nivel.
Estudiante	Identificador numérico del estudiante. Variable de primer nivel.
Prueba mat	Calificación de la prueba del estudiante. Variable dependiente.
Prueba raz	Calificación de la prueba del estudiante. Variable explicativa de primer nivel.
Prueba len	Calificación de la prueba del estudiante. Variable explicativa de primer nivel.
Prueba lec	Calificación de la prueba del estudiante. Variable explicativa de primer nivel.
Mujer	1=verdadero, 0=falso.
Tipo colegio	Tipo de colegio (0=estatal, 1=privado). Variable explicativa de segundo nivel.
Constante	Columna de unos.

Tabla 4.1: Descripción de las variables.

Para la formulación de los modelos multinivel se utilizó como variable dependiente la calificación de matemáticas (Prueba mat) y como variable explicada la calificación en la prueba de razonamiento (Prueba raz), debido a la relación lineal que existe entre las variables que se muestra en la figura (4.17). Las variables de las calificaciones de lenguaje y lectura se las utilizó para la aplicación del modelo de ecuaciones estructurales que se presenta más adelante.

El objetivo del análisis es establecer qué colegios han tenido mejor rendimiento en los resultados de las aplicaciones de las pruebas, además de estudiar las distintas variaciones que pudiesen existir en cuanto al rendimiento de los estudiantes en función del género, tipo de colegio, etc.

Para el procesamiento de los datos se utilizó el paquete estadístico MLwin [18] versión 2.26, que es un programa especializado en Modelización Multinivel y para el modelo de ecuaciones estructurales multinivel se utilizó el programa LISREL [7] versión estudiantil 9.1.

4.2 Modelos de nivel 2

A continuación se van a presentar los modelos que fueron expuestos en el primer capítulo del presente Proyecto de Titulación, con el objetivo de encontrar el mejor modelo que se ajuste a la realidad de los datos obtenidos y comparar los resulta-

dos obtenidos con dos tipos de estimaciones: Mínimos Cuadrados Generalizados Iterativos (MCGI) y Métodos de Cadenas de Markov Monte Carlo (CMMC).

Se considera que la población tiene una estructura jerárquica de dos niveles: las unidades de nivel 1 son 607 *estudiantes* de décimo año de Educación Básica y las unidades de nivel 2 son 12 *colegios*.

Estos modelos considerarán como variable dependiente la *Calificación de la prueba de matemática* y se comenzará planteando el modelo vacío y en el transcurso del desarrollo de los distintos modelos se va a ir incorporando cada una de las variables descritas anteriormente hasta llegar a estructurar un modelo que se ajuste a la realidad de los datos.

4.2.1 Modelo con ordenada aleatoria, sin variable explicativa

También conocido como *Modelo vacío* que fue explicado en el capítulo 1, se lo indica a continuación:

$$y_{ij} = \beta_{0j} + e_{0ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

de manera simplificada:

$$y_{ij} = \beta_0 + u_{0j} + e_{0ij} \quad (4.1)$$

Los resultados en MLwin utilizando el Método de los Mínimos Cuadrados Generalizados Iterativo (MCGI) de estimación son:

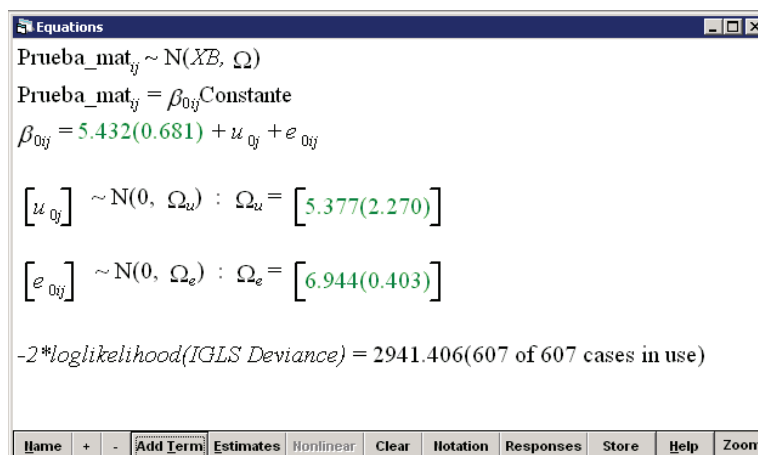


Figura 4.1: Resultados del modelo con ordenada aleatoria, sin variable explicativa con estimación de MCGI.

La primera línea especifica la distribución Normal que sigue la variable dependiente. El vector respuesta tiene una media específica en notación matricial denotado por XB , que representa la parte fija del modelo y la parte aleatoria consiste en un conjunto de variables aleatorias descritas por la matriz de covarianza. Esta matriz de covarianza incorpora matrices de covarianza separadas de los coeficientes aleatorios de cada nivel.

La variable $Prueba_Mat_{ij}$ es el valor de la prueba de matemática para el estudiante i en el colegio j . La media general $\hat{\beta}_0$ es igual a 5.432, que es interpretado como el valor esperado de la prueba de matemática para un estudiante aleatorio de un colegio escogido al azar. Las medias para los diferentes colegios se distribuyen sobre su media general con una varianza de 5.377, es decir, la varianza entre colegios ($\hat{\sigma}_{u0}^2$) es igual a 5.377 y la varianza entre estudiantes ($\hat{\sigma}_{e0}^2$) es 6.944. A continuación se procede a calcular la correlación intraclase del modelo (4.1) que es:

$$\widehat{Corr}(y_{ij}, y_{kj}) = \frac{5.377}{6.944 + 5.377} = 0.4364 \quad (i \neq k)$$

Lo que nos indica que 43.6 % de la varianza del modelo se encuentra explicada por variabilidad de los colegios y el 56.4 % restante, es la variabilidad de los estudiantes. Con esto podemos decir que las agrupaciones por colegios conduce a una similitud importante entre los resultados de las pruebas de matemáticas de los diferentes alumnos del mismo colegio, aunque las diferencias dentro de los colegios es levemente mayor que las diferencias de colegio a colegio.

A continuación se presentan los resultados de las estimaciones de los parámetros del modelo (4.1) utilizando el método de CMMC, con el muestreador de Gibbs.

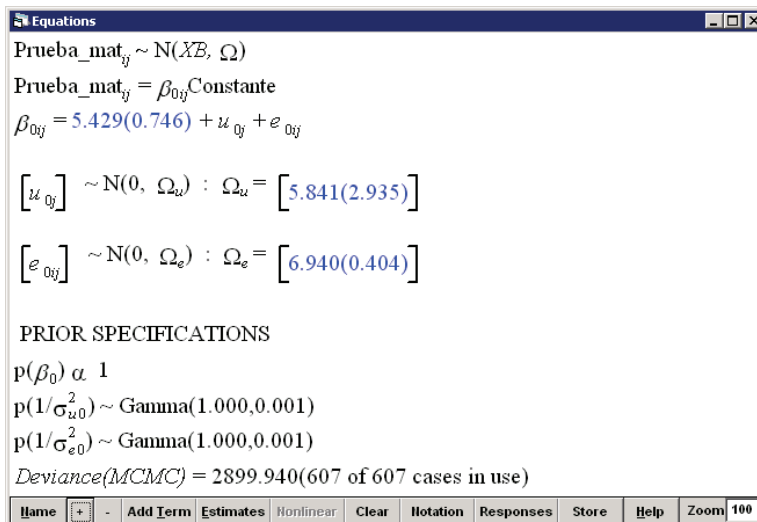


Figura 4.2: Resultados del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.

En este caso la selección a priori para la estimación con CMMC-Gibbs fue $p(\beta_0) \propto 1$, $p(\sigma_{u0}^{-2}) \sim \text{Gamma}(a, b)$ y $p(\sigma_{e0}^{-2}) \sim \text{Gamma}(a, b)$, con $a = 1$ y $b = 0.001$. Se realizaron 5,000 iteraciones en la estimación realizada para cada parámetro. En la siguiente figura se observan las últimas 500 iteraciones.

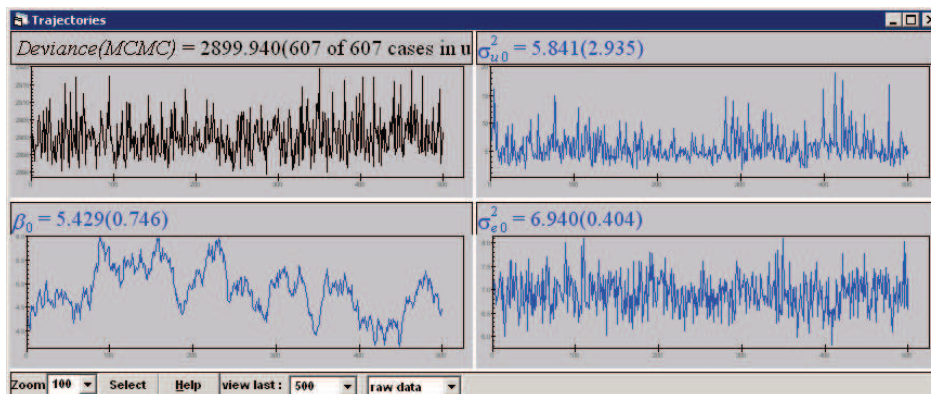


Figura 4.3: Resultados de la primera convergencia del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.

Se puede apreciar que para el parámetro $\hat{\beta}_0$, cada valor en el gráfico está altamente correlacionado con el valor precedido. Un diagnóstico más detallado sobre la información del parámetro se tiene a continuación.

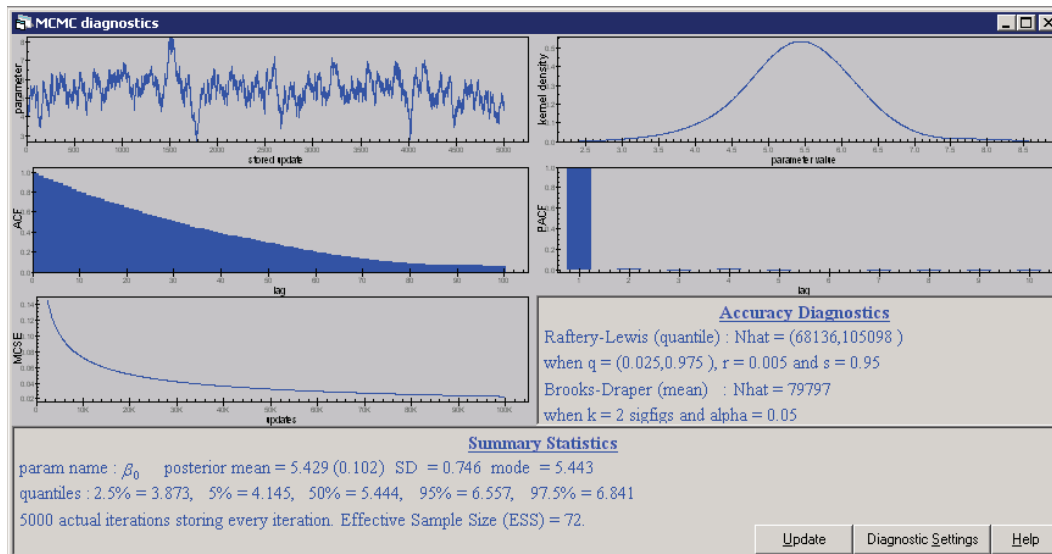


Figura 4.4: Resultados del primer diagnóstico de $\hat{\beta}_0$ del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.

El gráfico superior izquierdo en la figura (4.4), presenta las 5,000 iteraciones en la estimación del parámetro $\hat{\beta}_0$. El gráfico superior derecho es la *densidad del núcleo*¹ estimación de la distribución posterior conjunta; como se puede observar, la densidad se aproxima a una distribución normal.

Los segundos gráficos en la figura (4.4), presentan la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) respectivamente. El gráfico de ACF nos indica que existe una alta autocorrelación en la estimación de $\hat{\beta}_0$.

Por último, el gráfico inferior izquierdo en la figura (4.4), indica el error estándar Monte Carlo estimado (MCSE) de la estimación posterior de la media, contra el número de iteraciones. Como se puede apreciar a medida que el número de iteraciones es más grande, el error estándar es más pequeño.

El cuadro en la figura (4.4), presenta dos contrastes de diagnósticos de precisión, el de Raftery-Lewis que es un diagnóstico del control de la longitud de ejecución basado en el criterio de precisión de la estimación de los cuantiles (2.5% y 97.5%), formando un intervalo central estimado del tamaño de la cadena (Nhat) para $\hat{\beta}_0$ en este caso es de 68,136.11. El diagnóstico de Brooks-Draper es basado en la media de la distribución y es usado para la estimación del tamaño de la cadena

¹Es una manera no paramétrica de estimar la función de densidad de probabilidad de una variable.

de Márkov requerida para producir una estimación media para k ($k = 2$) con una precisión dada. En este caso el tamaño de la cadena (N_{hat}) requerido es de 79,797, que resulta un tamaño mucho más grande que el anterior sugerido.

La interpretación de los números $q = (0.025; 0.975)$, $r = 0.005$ y $s = 0.95$ en el diagnóstico de Raftery-Lewis es que el $100(0.975 - 0.025)\% = 95\%$ del intervalo estimado para el parámetro dado no va a diferir en más del $100(2r)\% = 1\%$.

El último cuadro en la figura (4.4), nos presenta los estadísticos de $\hat{\beta}_0$ y los diferentes intervalos de confianza a nivel α . Por ejemplo, al 95% el intervalo central (intervalo Bayesiano) corre de 3.873 a 6.841.

Dados los resultados obtenidos se procede a realizar una nueva estimación (CMMC-Gibbs), pero con 70,000 iteraciones, de lo que se obtiene:

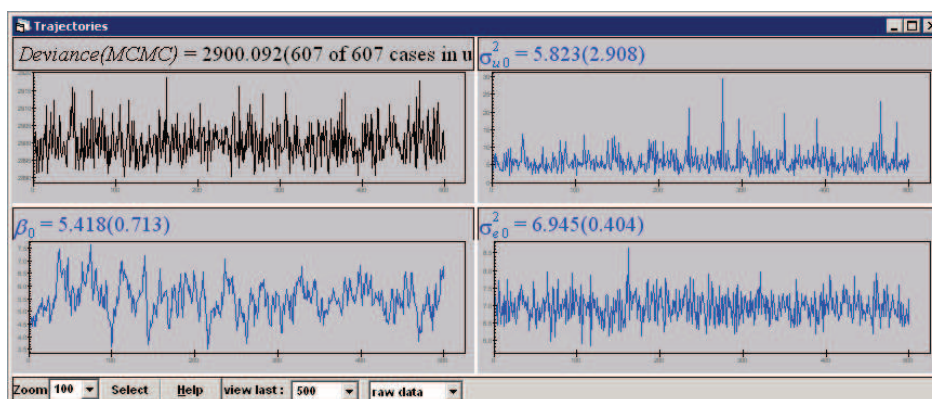


Figura 4.5: Resultados de la segunda convergencia del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.

Se puede apreciar las últimas 500 iteraciones en la estimación de los parámetros. Los valores para la estimación de $\hat{\beta}_0$ se observa que aún existe correlación con el valor anterior en la cadena de estimación del parámetro. Los diagnósticos de $\hat{\beta}_0$ de la estimación con CMMC-Gibbs son los siguientes:

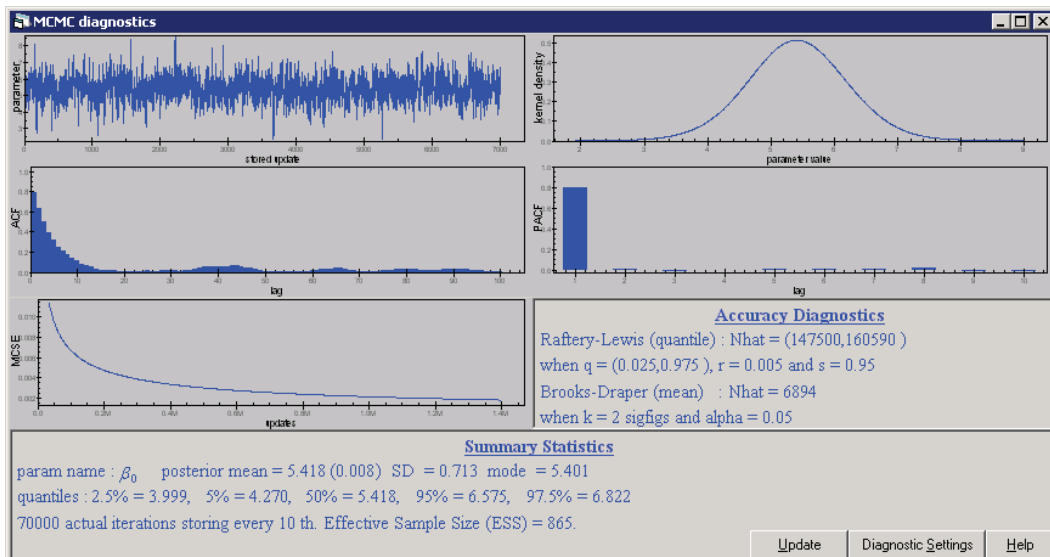


Figura 4.6: Resultados del segundo diagnóstico de $\hat{\beta}_0$ del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.

En la figura (4.6) de la función de autocorrelación (ACF) se observa que existe una disminución en la autocorrelación pero aún sigue siendo importante. El diagnóstico de Raftery-Lewis sugiere un tamaño de cadena de Markov de 147,500.16, por tal razón se realiza una nueva estimación con 150,000 iteraciones.

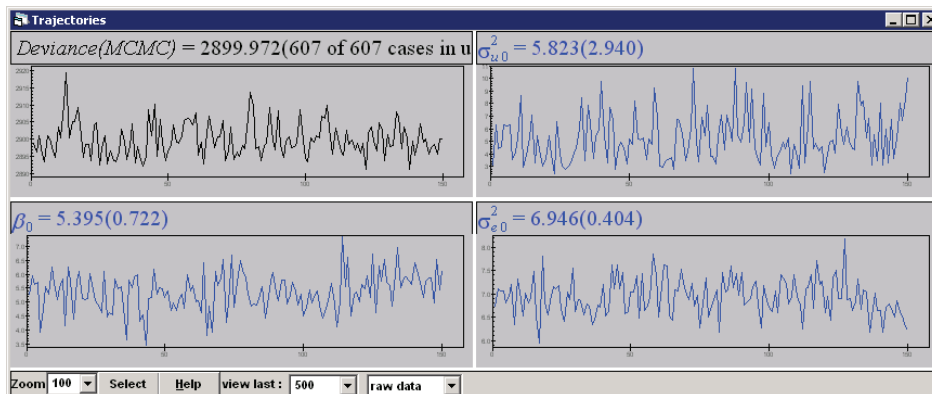


Figura 4.7: Resultados de la tercera convergencia del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.

En los gráficos de trayectorias de las estimaciones de los parámetros ya no tiene una correlación en los valores para la estimación del parámetro $\hat{\beta}_0$ ni de ningún parámetro. Esto se lo constata a continuación con los diagnósticos de la estimación.

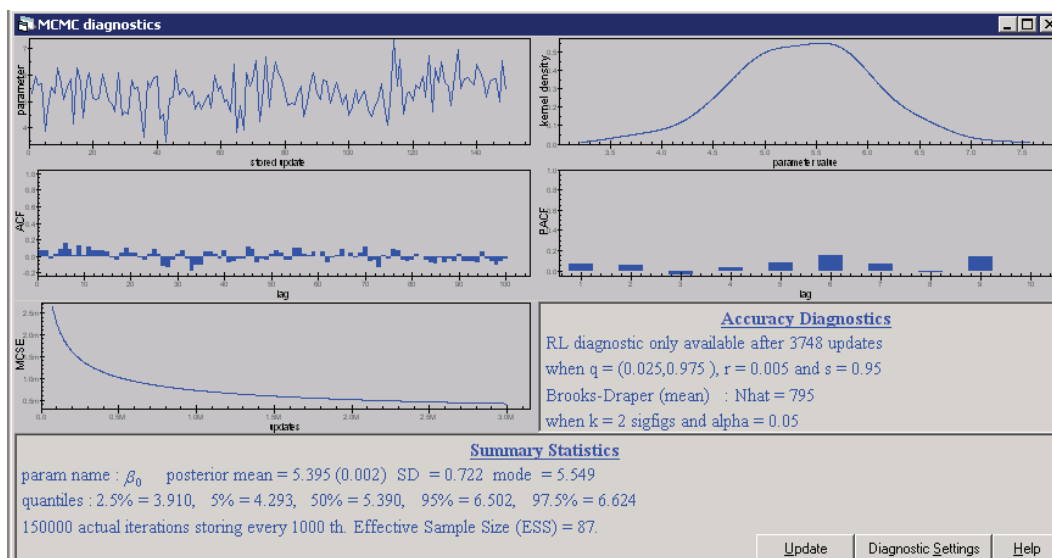


Figura 4.8: Resultados del tercer diagnóstico de $\hat{\beta}_0$ del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.

Con esto podemos decir que la serie de tiempo autoregresiva presenta una pequeña autocorrelación menor a 0.2. Con lo que se concluye que el AFC es consistente; esto sugiere que la cadena es adecuadamente cercana a una distribución independiente e idénticamente distribuida (autocorrelación 0). De la misma manera a continuación se presenta la estimación de la varianza de nivel 2 ($\hat{\sigma}_{u0}^2$).

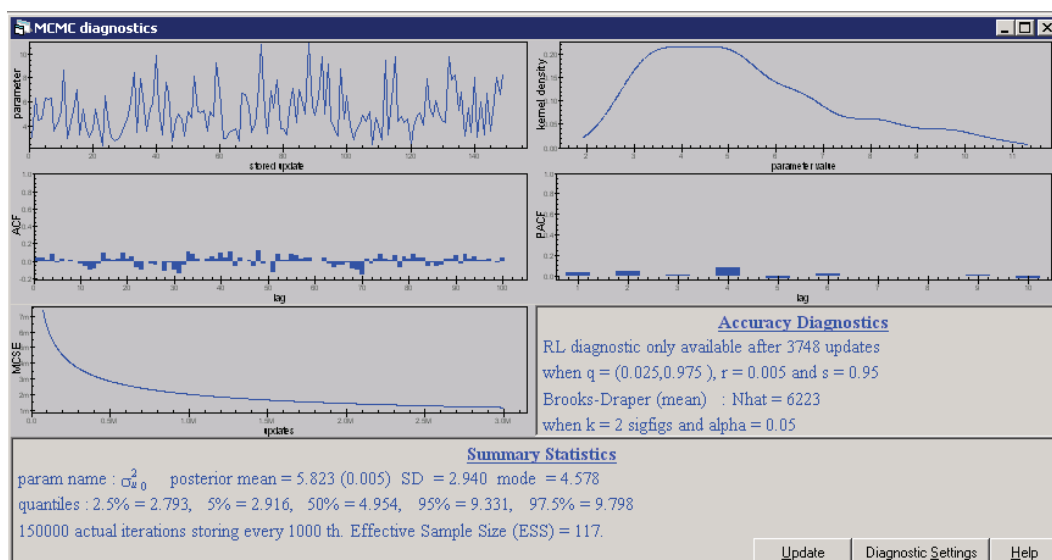


Figura 4.9: Resultados del diagnóstico de $(\hat{\sigma}_{u0}^2)$ del modelo con ordenada aleatoria, sin variable explicativa con estimación de CMMC-Gibbs.

Se puede ver en la figura (4.9) que la densidad del núcleo en la distribución posterior no es simétrica, que es de esperar para el parámetro de la varianza. Los gráficos de ACF y PACF indican que no existe autocorrelación en la estimación del parámetro ($\hat{\sigma}_{u0}^2$).

Ahora se procederá a realizar una nueva estimación con CMMC-Gibbs considerando como distribución a priori de la varianza una distribución Uniforme.

Se corrió la estimación inicial con un tamaño de cadena de 5,000 iteraciones que dio como resultado un alta correlación de los valores para la estimación de $\hat{\beta}_0$. Luego de algunas corridas del modelo con distintos tamaños de cadena, se eliminó la autocorrelación del término $\hat{\beta}_0$ con una cadena de tamaño de 240,000. Así mismo, ningún parámetro presentó autocorrelación en esta estimación final. A continuación se presenta un cuadro comparativo de las estimaciones obtenidas hasta el momento, donde e.e. es el error estándar de cada parámetro.

	CMMC-Gibbs				MCGI	
	$\Gamma^{-1}(1; 0.001)$		Uniforme			
Parámetro	media	e.e.	media	e.e.	media	e.e.
$\hat{\beta}_0$	5.395	0.722	5.419	0.922	5.432	0.681
$\hat{\sigma}_{u0}^2$	5.823	2.94	9.525	6.210	5.377	2.270
$\hat{\sigma}_{e0}^2$	6.946	0.404	6.990	0.408	6.944	0.403
Desviación	2,899.972		2,899.860		2,941.406	
DIC	2,912.54		2,912.61			
Cadena	150,000		240,000			

Tabla 4.2: Comparación de estimaciones CMMC-Gibbs con dos distribuciones a priori y MCGI del modelo con ordenada aleatoria, sin variable explicativa.

Las estimaciones del modelo (4.1) con el método de CMMC-Gibbs presenta mejores resultados, ya que el valor de la desviación se ve disminuida con la estimación con CMMC-Gibbs para con las dos distribuciones a priori. Respecto a la selección a priori de la distribución Uniforme, tiende a obtener un valor más grande de la varianza, casi el doble de su valor; de la misma manera tiene el mismo efecto el error estándar del parámetro $\hat{\sigma}_{u0}^2$, además que el tamaño de la cadena para la estimación de los parámetros fue mayor que con la distribución Gamma inversa a priori.

También se procedió a realizar una estimación con CMMC-MH. Primeramente considerando como distribución a priori: $p(\beta_0) \propto 1$, $p(\sigma_{u0}^{-2}) \sim \text{Gamma}(\varepsilon, \varepsilon)$ y $p(\sigma_{e0}^{-2}) \sim \text{Gamma}(\varepsilon, \varepsilon)$. Se inició la estimación con una cadena de tamaño 5,000, pero la presencia de autocorrelación en la estimación de $\hat{\beta}_0$, sugirió realizar una

nueva estimación con una cadena de tamaño 730,000. Se planteó además para otra estimación del modelo (4.1) como distribución a priori, una Uniforme para todos los parámetros. El tamaño final de la cadena para su estimación fue de 370,000. A continuación el cuadro resumen con todas las estimaciones:

Parámetro	CMMC-Gibbs				CMMC-MH				MCGI	
	$\Gamma^{-1}(1; 0.001)$		Uniforme		$\Gamma^{-1}(1; 0.001)$		Uniforme		media	e.e.
$\hat{\beta}_0$	5.395	0.722	5.419	0.922	5.423	0.709	5.427	0.978	5.432	0.681
$\hat{\sigma}_{u0}^2$	5.823	2.940	9.525	6.210	5.800	2.868	9.663	6.646	5.377	2.270
$\hat{\sigma}_{e0}^2$	6.946	0.404	6.990	0.408	6.947	0.405	6.990	0.407	6.944	0.403
Desviación	2,899.972		2,899.860		2,900.107		2,899.912		2,941.406	
DIC	2,912.54		2,912.61		2,912.82		2,912.72			
Cadena	150,000		240,000		730,000		370,000			

Tabla 4.3: Comparación de estimaciones CMMC-Gibbs con dos distribuciones a priori, CMMC-MH con dos distribuciones a priori y MCGI del modelo con pendiente aleatoria, sin variable explicativa.

Antes de realizar un análisis general de los resultados, se considera necesario una prueba de significación de los parámetros del modelo (4.1) que se indica a continuación.

$$y_{ij} = \beta_0 + u_{0j} + e_{0ij}$$

Prueba de Hipótesis

Para el parámetro $\hat{\beta}_0$ se plantean las siguientes hipótesis:

$$H_0 : \hat{\beta}_0 = 0$$

$$H_1 : \hat{\beta}_0 \neq 0$$

Se rechaza H_0 al nivel 5 %, ya que para todos los métodos de estimación $\frac{\hat{\beta}_0}{ee(\hat{\beta}_0)} \geq 1.96$.

Para la prueba de significación del parámetro $\hat{\sigma}_{u0}^2$, es decir, para probar que la variabilidad entre colegios es significativa, primeramente se realiza una nueva estimación del modelo (4.1) pero sin considerar el término u_{0j}^2 , es decir, sin la variación entre colegios. Con lo que tendríamos el siguiente modelo:

$$y_{ij} = \beta_0 + e_{0ij} \quad (4.2)$$

A continuación se presentan los resultados de la estimación del modelo (4.2)

con MCGI, debido a que el valor de la desviación con la estimación de CMMC para este modelo, no existe variación.

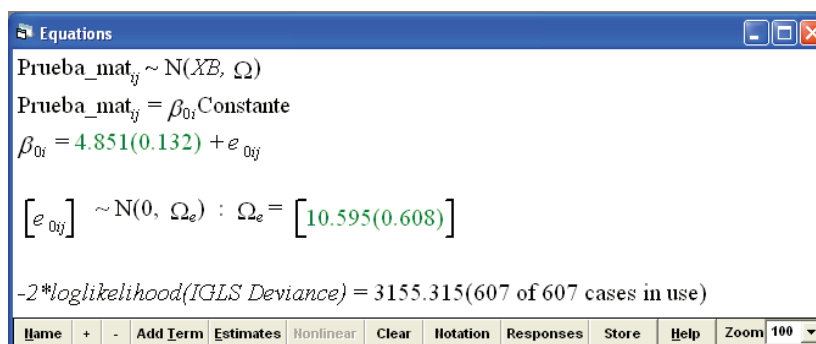


Figura 4.10: Resultados del modelo con ordenada aleatoria, sin variable explicativa y sin término u_{0j} .

En este caso tendríamos los resultados de un modelo de regresión lineal, debido a que no existe un segundo nivel en el modelo, ya que fue removido. Como se puede apreciar los términos $\hat{\beta}_0$ y $\hat{\sigma}_{e_0}^2$ son significativos al nivel 5%. Con esto, se contrastan las siguientes hipótesis:

H_0 : Los datos se ajustan al modelo (4.2)

H_1 : Los datos se ajustan al modelo (4.1)

Modelo (4.1)	$y_{ij} = \beta_0 + u_{0j} + e_{0ij}$
Modelo (4.2)	$y_{ij} = \beta_0 + e_{0ij}$

Se rechaza H_0 al nivel 5%, ya que la diferencia de las desviaciones de los modelos (4.2) y (4.1) es mayor que el cuantil de la distribución χ^2 con un grado de libertad, es decir, $213.91 > 3.84(\chi^2(1))$. Con lo que se puede concluir que las diferencias entre grupos son fuertemente significativas con todos los métodos de estimación.

Una vez conociendo los parámetros que resultaron significativos con los distintos tipos de estimación, se puede dar un mejor criterio de los resultados obtenidos. Es así que los resultados con la estimación de CMMC con una distribución a priori Uniforme presentó los valores de la varianza $\hat{\sigma}_{u_0}^2$ altos, de igual manera que sus errores estándar; sin embargo, sus valores fueron significativos al nivel 5% en todos los casos. Para el resto de parámetros, los resultados estimados fueron significativos y similares con los distintos métodos de estimación. Los valores de la desviación con los métodos de CMMC son menores que el de MCGI, lo que nos indica un mejor ajuste del modelo. Respecto al tamaño de la cadenas para la estimación de los parámetros, el método de CMMC-MH necesitó cadenas más grandes que con

CMMC-Gibbs.

Con esto se puede concluir que los mejores resultados de las estimaciones de los parámetros se obtuvo con el método de MCCM con una Gamma inversa como información a priori, y fue con el método de CMMC-Gibbs que se logró una estimación más rápida con una cadena de tamaño de 150,000.

Residuos

A continuación se procederá a analizar los residuos para detectar casos atípicos que pueden influenciar en el resultado del modelo.

Primeramente se realiza una inspección de los residuos de nivel 1 o nivel estudiante en la figura (4.11) presentada a continuación, donde los residuos se encuentran ordenados de menor a mayor y cada uno con intervalo de confianza del 95%.

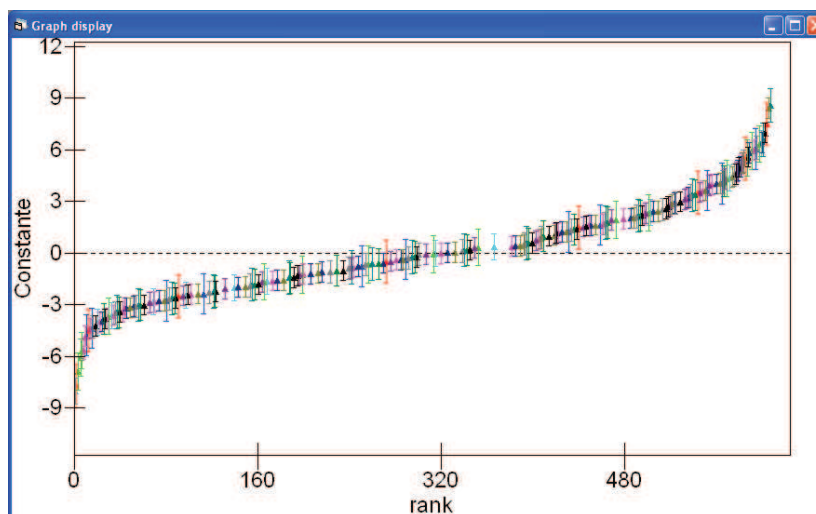


Figura 4.11: Residuos de nivel 1 del modelo con ordenada aleatoria, sin variable explicativa.

Las líneas verticales indican los intervalos de confianza de la media residual de nivel 1. Los residuos toman valores de 9 a -9, pero no se puede apreciar de una forma clara si existen puntos *singulares*² o *atípicos* [4]. Es por esta razón que se procedió a realizar un análisis de residuos por cada colegio, que se presenta a continuación.

²Se lo llama punto singular si se aleja notablemente del resto de puntos.

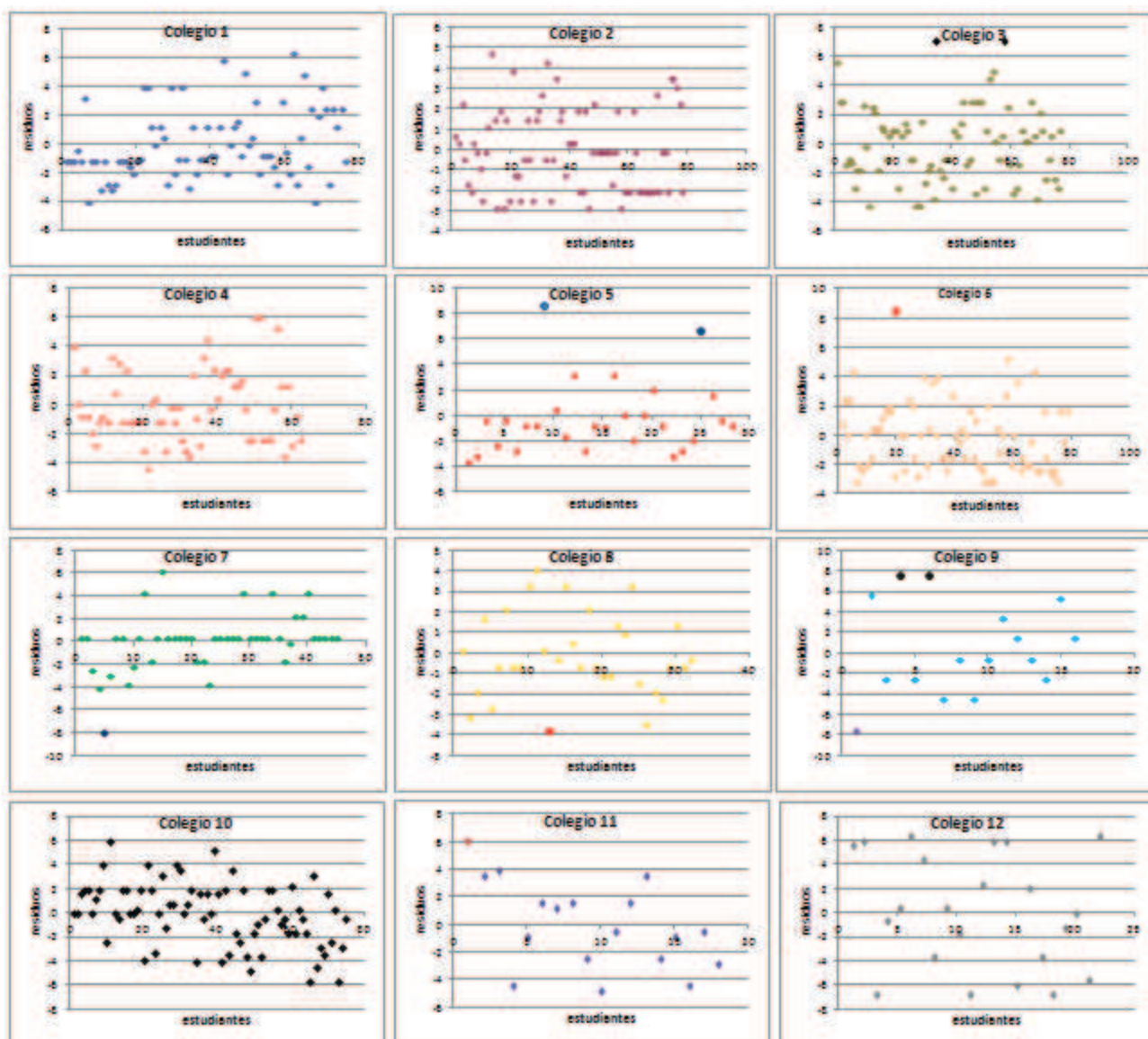


Figura 4.12: Residuos de nivel 1 por colegio del modelo con ordenada aleatoria, sin variable explicativa.

En la figura (4.12) se presentan los residuos por colegio (Anexo D), donde se puede apreciar que los puntos están dentro de una franja horizontal simétrica al eje X, lo que nos indica que la varianza es constante, cumpliendo la hipótesis del modelo. Además, se observa que en los colegios: 3, 5, 6, 9 y 11 existen puntos sobre la nube, que pudieran ser considerados como puntos singulares, pero no es el caso, debido a que estos puntos representan los estudiantes que obtuvieron las notas más altas en la prueba de matemática. Esta es la razón del por qué estos puntos se encuentran alejados del grupo. De la misma manera en los colegios: 7, 8 y 9 se tiene puntos bajo la nube, que representan a los estudiantes con las más bajas notas en la prueba de matemática.

Por las razones expuestas anteriormente dichos puntos no serán descartados, debido a que estos estudiantes representan la realidad que existe en los distintos colegios de estudiantes aplicados y estudiantes con bajo rendimiento.

Comprobar la hipótesis de normalidad de los residuos es sumamente importante; por tal razón, a continuación se presenta el gráfico de los residuos de nivel 1 estandarizados.

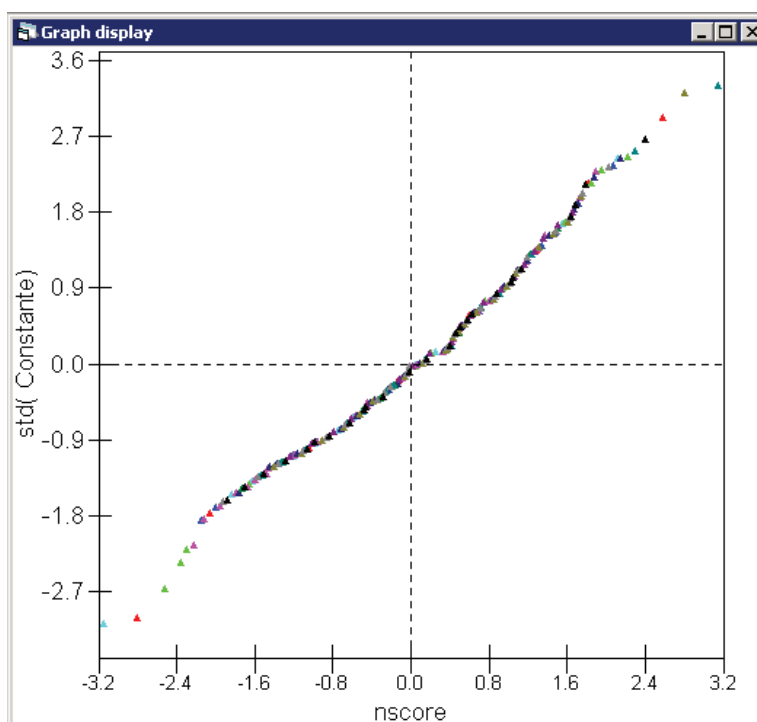


Figura 4.13: Gráfico de probabilidad normal de los residuos estandarizados de nivel 1 del modelo con ordenada aleatoria, sin variable explicativa.

El gráfico de probabilidad normal de los residuos estandarizados de nivel 1 de la figura (4.13), se ajusta de forma lineal, lo que implica que la hipótesis de normalidad se cumple para los residuos de nivel 1 (e_{ij}) del modelo (4.1).

Ahora, se procederá a analizar los residuos de nivel 2, es decir, al nivel colegios que se presenta a continuación.

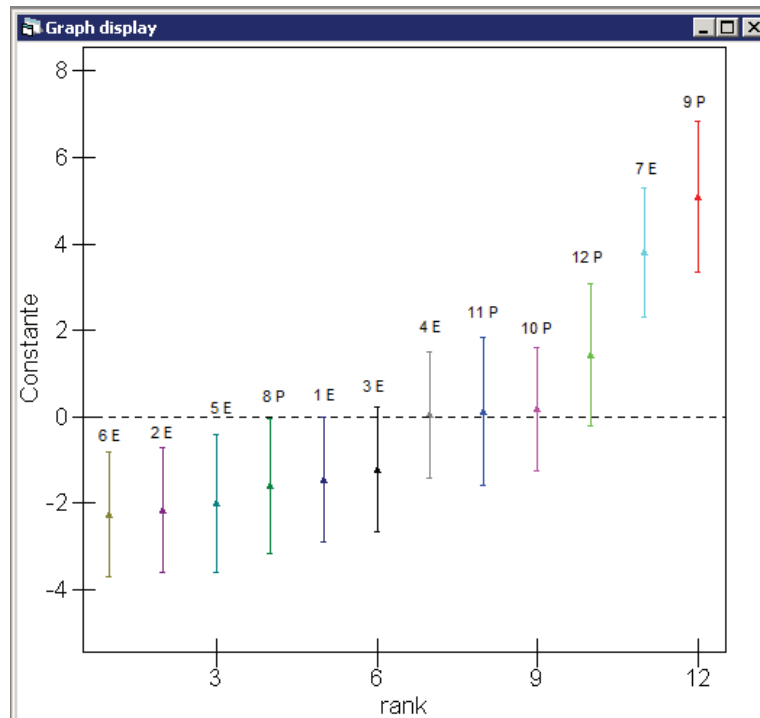


Figura 4.14: Residuos de nivel 2 del modelo con ordenada aleatoria, sin variable explicativa.

La figura (4.14) es un gráfico oruga, donde se tienen los 12 residuos de nivel 2, uno por cada colegio, y cada uno de ellos se encuentran identificados por número (Anexo C), el colegio al que representan y la letra P y E, que significa particular o estatal. Como se puede apreciar los colegios 7 y 9 son los colegios que tienen residuos más grandes y por ende son los colegios que presentan distancias mayores a la regresión común.

A continuación se presenta el gráfico de los residuos.

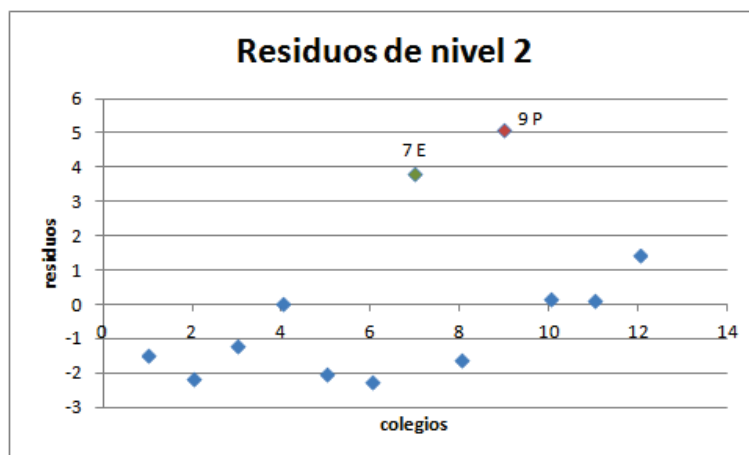


Figura 4.15: Residuos de nivel 2 en función del número de colegios del modelo con ordenada aleatoria, sin variable explicativa.

Como se puede apreciar en la figura (4.15) existe dos puntos alejados del resto, que pudiesen ser considerados como puntos singulares. Pero dichos puntos representan a los colegios 7 y 9 que son los que obtuvieron las mejores calificaciones en las pruebas de matemática; es por esta razón que estos puntos se encuentran alejados del resto. Por tal motivo, estos puntos no van a ser descartados en el modelo.

El gráfico de normalidad de los residuos de nivel 2 se presenta a continuación.

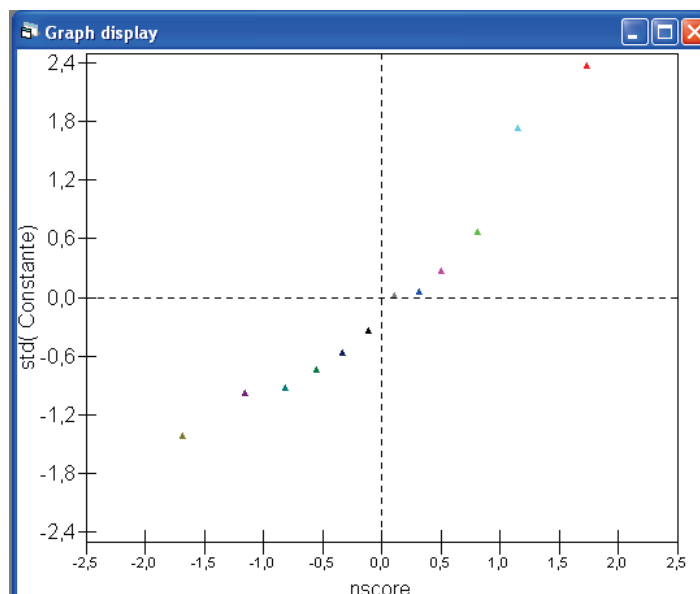


Figura 4.16: Gráfico de probabilidad normal de los residuos estandarizados de nivel 2 del modelo con ordenada aleatoria, sin variable explicativa.

En la figura (4.16) del gráfico de probabilidad normal no se ajusta de una manera clara a una forma lineal, es por esta razón que a continuación se procede a realizar la prueba de Kolmogorov Smirnov (K-S) para verificar si los residuos de nivel 2 siguen una distribución normal. Con lo que se contrastan las siguientes hipótesis:

H_0 : Los residuos de nivel 2 se ajustan a una distribución normal.

H_1 : Los residuos de nivel 2 no se ajustan a una distribución normal.

La prueba se la realizó en el paquete estadístico SPSS 16.0 dando los siguientes resultados:

		Residuos
N		12
Normal Parameters ^a	Mean	.0000
	Std. Deviation	2.38017
Most Extreme Differences	Absolute	.220
	Positive	.220
	Negative	-.171
Kolmogorov-Smirnov Z		.763
Asymp. Sig. (2-tailed)		.605

a. Test distribution is Normal.

Figura 4.17: Prueba de normalidad de K-S de los residuos de nivel 2 del modelo con ordenada aleatoria, sin variable explicativa.

como el nivel de significación es mayor a 0.05, se acepta la hipótesis H_0 , es decir, los residuos de nivel 2 siguen una distribución normal.

4.2.2 Modelo con ordenada aleatoria, con una variable explicativa

El modelo viene formulado de la siguiente manera:

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + e_{0ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

de manera simplificada:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{0ij} \quad (4.3)$$

Para este modelo se ha incorporado como variable explicativa *Prueba raz* para explicar parte de la variabilidad de *Y*. A continuación se muestra en la figura (4.18) la correlación positiva que existe entre las variables calificación de matemática (*Prueba mat*) y calificación de razonamiento (*Prueba raz*).

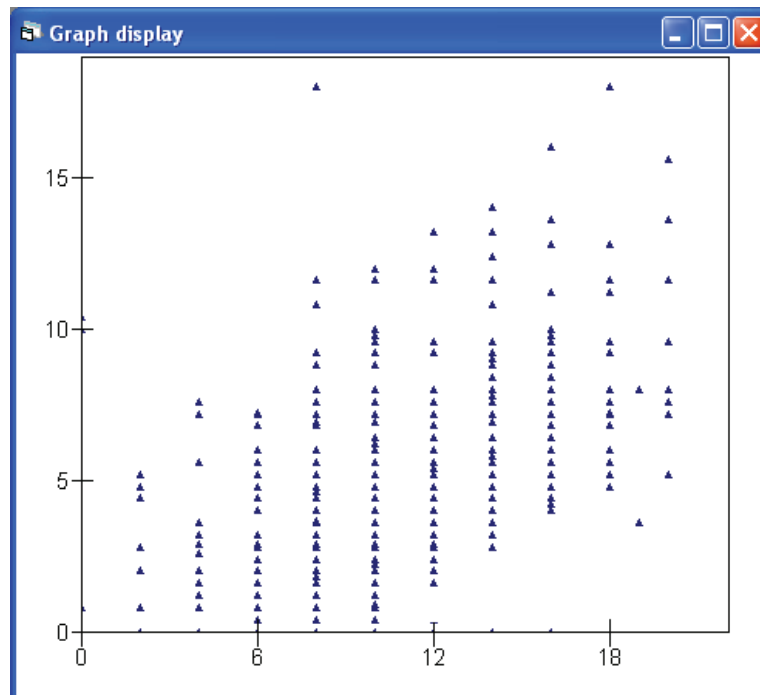


Figura 4.18: Variable Prueba mat vs. Variable Prueba raz.

Los resultados del modelo (4.3) con estimación de MGLI en MLwin son los siguientes:

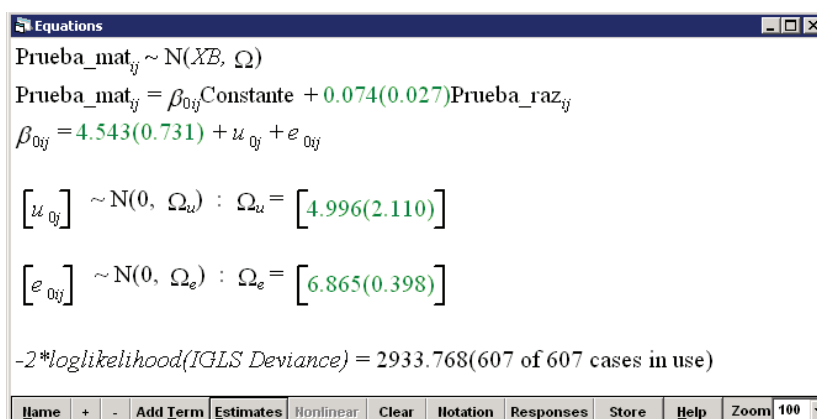


Figura 4.19: Resultados del modelo con ordenada aleatoria, con una variable explicativa con estimación de MGLI.

Podemos dar la siguiente interpretación del modelo: si la *Prueba raz* se incrementa en un punto, la *Prueba mat* se incrementa en promedio 0.074 puntos para un

estudiante aleatorio de un colegio escogido al azar.

Para constatar la significación de la inclusión de la variable *Prueba raz*, se realiza la prueba de significación del parámetro $\hat{\beta}_1$, con lo que se contrastan las siguientes hipótesis:

H_0 : Los datos se ajustan al modelo (4.1)

H_1 : Los datos se ajustan al modelo (4.3)

Modelo (4.1)	$y_{ij} = \beta_0 + u_{0j} + e_{0ij}$
Modelo (4.3)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{0ij}$

Se rechaza H_0 al nivel 5 %, debido a que la diferencia de las desviaciones de los modelos (4.1) y (4.3) es mayor que el cuantil de la distribución χ^2 con un grado de libertad, ya que $7 > 3.84(\chi^2(1))$.

El modelo (4.3) tiene cuatro parámetros: coeficientes de regresión β_0 y β_1 y los componentes de la varianza $\sigma_{u_0}^2$ y $\sigma_{e_0}^2$, donde u_{0j} puede ser considerado como los residuos a nivel de grupo o los efectos de grupo que no fueron explicados por x . La varianza residual $\hat{\sigma}_{e_0}^2$ como la varianza $\hat{\sigma}_{u_0}^2$ son menores en este modelo (4.3) que en el modelo con ordenada aleatoria sin variable explicativa (4.1). La varianza residual es menor debido a que las diferencias entre estudiantes son parcialmente explicadas por la variable x y la varianza $\hat{\sigma}_{u_0}^2$ es menor porque los colegios difieren en promedio en las notas de *Prueba raz*, entonces el nivel estudiante también explica parte de las diferencias entre colegios. La correlación intraclase residual es estimada por:

$$\widehat{Corr}(y_{ij}, y_{kj} \mid x_{ij}, x_{kj}) = \frac{4.996}{6.865 + 4.996} = 0.4212 \quad (i \neq k)$$

Lo que indica que existe una correlación del 42.12 % en las notas de la *Prueba mat* de dos estudiantes escogidos aleatoriamente en un colegio extraído al azar.

Gráficamente cada colegio esta representado por una recta con pendiente β_1 y ordenada en el origen $\beta_0 + u_{0j}$, de manera que constituye un conjunto de rectas paralelas, como se muestra a continuación.

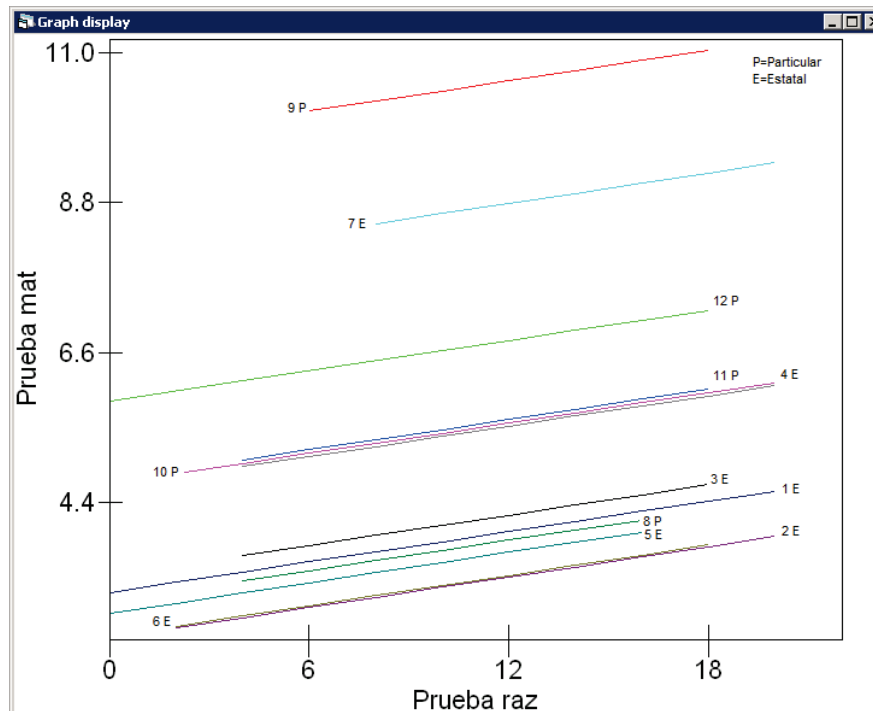


Figura 4.20: Regresiones de 12 colegios según el modelo con ordenada aleatoria, con una variable explicativa.

El parámetro β_{0j} varía aleatoriamente de grupo a grupo. El residuo de segundo nivel u_{0j} modifica el término de la ordenada, pero el coeficiente β_1 es fijo. Entonces las líneas de predicciones para los 12 colegios deben ser paralelas como se muestra en la figura (4.20). Además se encuentran identificadas con un color distinto cada regresión por colegio y con el número correspondiente a cada colegio (Anexo D). Como se puede apreciar el colegio con las más altas puntuaciones es un colegio Privado, seguido de un colegio Estatal y los resultados más bajos en su mayoría es de estudiantes en colegios Estatales.

Ahora se presentan los resultados de las estimaciones de los parámetros del modelo (4.3) utilizando el Método de Cadenas de Markov Monte Carlo (CMMC), con el muestreador de Gibbs y el método de Metropolis-Hastings (MH). Utilizando para cada una de ellas información a priori para la estimación de los parámetros, de una distribución Gamma inversa y una distribución Uniforme. A continuación el cuadro resumen de las estimaciones:

	CMMC-Gibbs				CMMC-MH				MCGI	
	$\Gamma^{-1}(1; 0.001)$		Uniforme		$\Gamma^{-1}(1; 0.001)$		Uniforme			
Parámetro	media	e.e.	media	e.e.	media	e.e.	media	e.e.	media	e.e.
$\hat{\beta}_0$	4.515	0.756	4.543	0.948	4.527	0.730	4.547	0.938	4.543	0.731
$\hat{\beta}_1$	0.075	0.027	0.073	0.027	0.075	0.027	0.073	0.027	0.074	0.027
$\hat{\sigma}_{u0}^2$	5.406	2.626	8.871	6.169	5.331	2.607	8.845	5.753	4.996	2.110
$\hat{\sigma}_{e0}^2$	6.876	0.403	6.922	0.404	6.881	0.400	6.922	0.405	6.865	0.398
Desviación	2,894.100		2,893.891		2,893.991		2,893.910		2,933.768	
DIC	2,907.660		2,907.580		2,907.420		2,907.610			
Cadena	22,000		465,000		200,000		240,000			

Tabla 4.4: Comparación de estimaciones CMMC-Gibbs con dos distribuciones a priori, CMMC-MH con dos distribuciones a priori y MCGI del modelo con ordenada aleatoria, con una variable explicativa.

A continuación se procederá a realizar un test de significación de los parámetros fijos y aleatorios obtenidos con cada tipo de estimación del modelo (4.3) que se presenta a continuación:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{0ij}$$

Prueba de Hipótesis

Para el parámetro $\hat{\beta}_0$ se plantea las siguientes hipótesis:

$$H_0 : \hat{\beta}_0 = 0$$

$$H_1 : \hat{\beta}_0 \neq 0$$

Se rechaza H_0 al nivel 5 %, ya que para todos los métodos de estimación $\frac{\hat{\beta}_0}{ee(\hat{\beta}_0)} \geq 1.96$.

La prueba de significación del parámetro $\hat{\beta}_1$, contrastan las siguientes hipótesis:

$$H_0 : \hat{\beta}_1 = 0$$

$$H_1 : \hat{\beta}_1 \neq 0$$

Se rechaza H_0 al nivel 5 %, ya que $\frac{\hat{\beta}_1}{ee(\hat{\beta}_1)} \geq 1.96$ para todos los métodos de estimación.

Para cerciorar que la variabilidad entre colegios es significativa, se procede a comparar los resultados obtenidos con una nueva estimación mediante el método de MCGI del modelo (4.3) pero sin considerar la variabilidad entre los colegios, es

decir, la varianza de u_{0j} es cero; con lo que se tendría el siguiente modelo:

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij}$$

$$\beta_{0i} = \beta_0 + e_{0ij}$$

de manera simplificada:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + e_{0ij} \quad (4.4)$$

La estimación del modelo tiene como resultados:

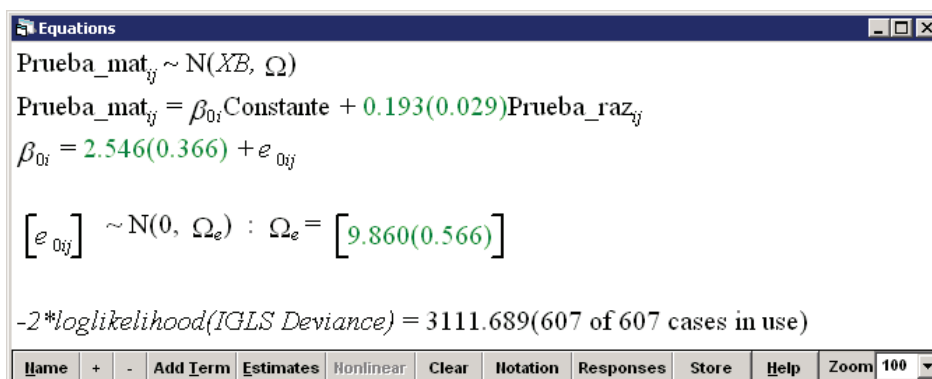


Figura 4.21: Resultados del modelo con ordenada aleatoria, con una variable explicativa sin término u_{0j} con estimación de MCGI.

La exclusión del término u_{0j} del modelo (4.3), tiene como efectos un mayor valor de la varianza $\hat{\sigma}_{e0}^2$ y un valor de la ordena $\hat{\beta}_0$ menor. Estos resultados se presentan por no considerar la variabilidad entre grupos. Además se tiene que los parámetros $\hat{\beta}_0$, $\hat{\beta}_1$ y $\hat{\sigma}_{e0}^2$ son significativos al nivel 5%.

Ahora se procede a realizar la prueba de significación del parámetro $\hat{\sigma}_{u0}^2$ con lo se contrastan las siguientes hipótesis:

H_0 : Los datos se ajustan al modelo (4.4)

H_1 : Los datos se ajustan al modelo (4.3)

Modelo (4.4)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + e_{0ij}$
Modelo (4.3)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{0ij}$

Se rechaza H_0 al nivel 5%, debido a que la diferencia de las desviaciones de los modelos (4.4) y (4.3) es mayor que el cuantil de la distribución χ^2 con un grado de libertad, es decir $177.92 > 3.84(\chi^2(1))$. Con lo que se puede concluir que las diferencias entre grupos son significativas con todos los métodos de estimación.

Conociendo los resultados del test de significación de los parámetros, se procede a realizar un análisis en conjunto de los parámetros. Los resultados con la estimación de CMMC con una distribución a priori Uniforme presentó valores de varianza $\hat{\sigma}_{u0}^2$ altos al igual forma que sus errores estándar; pero las estimaciones fueron significativas al nivel 5 %. Para los demás parámetros los resultados estimados fueron parecidos y significativos con los diferentes métodos de estimación.

Los valores de la desviación con los métodos de CMMC son menores que el de MCGI, lo que nos indica un mejor ajuste del modelo. Respecto al tamaño de las cadenas para la estimación de los parámetros, el método de CMMC-MH necesitó cadenas más grandes que con CMMC-Gibbs, especialmente con una información a priori de una distribución Uniforme.

Con lo anteriormente descrito se puede concluir que la inclusión de la variable *Prueba raz* en el modelo (4.3) fue significativa y presentó mejor ajuste que en el modelo anterior (4.1). Además, la estimación de los parámetros con el método de CMMC tuvo mejores resultados que el método de MCGI. En especial el método de CMMC con una distribución a priori de una Gamma inversa, fue el que presentó los mejores resultados tanto con el muestreador de Gibbs como con el método de MH; pero cabe señalar que el muestreador de Gibbs fue el que resolvió la estimación de los parámetros en un tiempo muy corto, ya que utilizó una cadena de tamaño de 22,000.

Residuos

A continuación se presenta un gráfico oruga de los residuos de nivel 1, que como se puede apreciar no distorciona del gráfico de residuos de nivel 1 del modelo anterior (4.1).

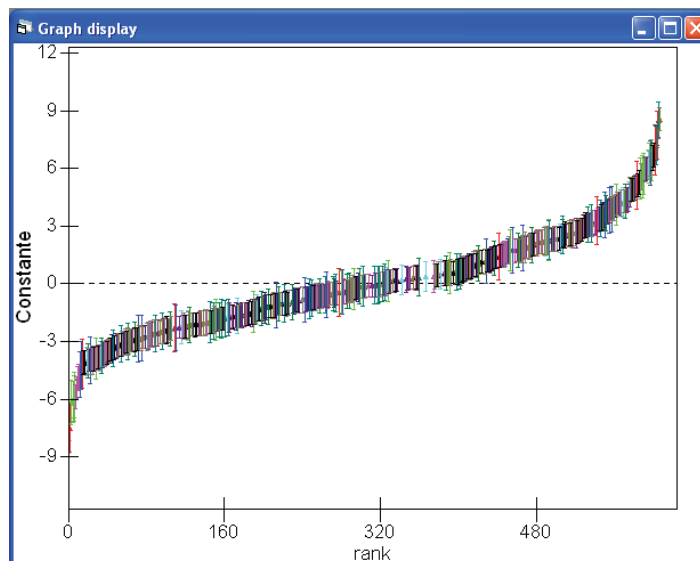


Figura 4.22: Residuos de nivel 1 del modelo con ordenada aleatoria, con una variable explicativa.

Debido a que en el modelo anterior (4.1) se realizó un análisis exhaustivo de los residuos y probables puntos singulares, para este modelo solamente vamos a constatar que se cumplan las hipótesis de los residuos que siguen una distribución normal.

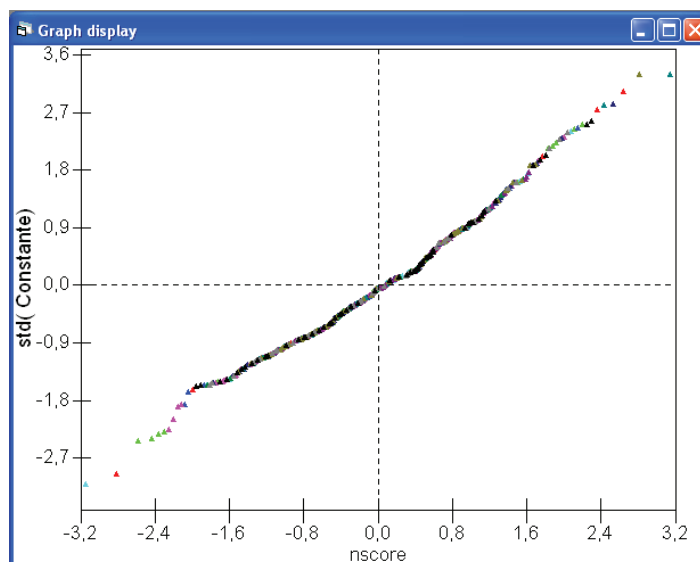


Figura 4.23: Gráfico de probabilidad normal de los residuos estandarizados de nivel 1 del modelo con ordenada aleatoria, con una variable explicativa

En el gráfico de probabilidad normal para los residuos de nivel 1 (figura (4.23)), se puede observar que los puntos se aproximan a una línea recta, lo que valida la

hipótesis de normalidad.

A continuación se presenta el gráfico de residuos de nivel 2.

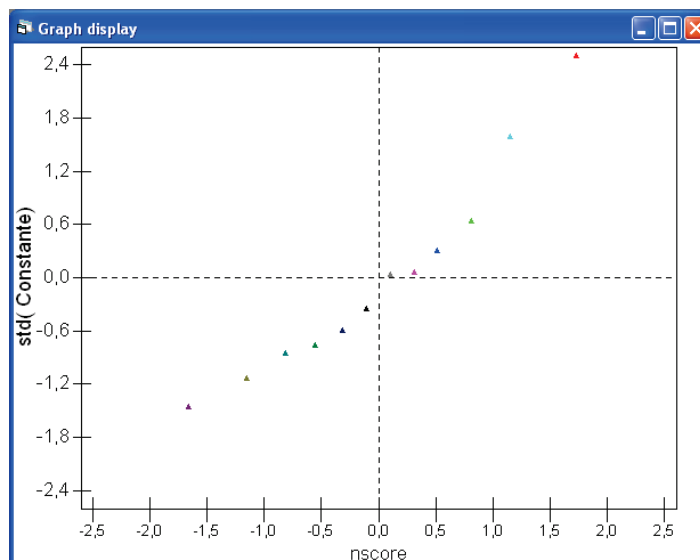


Figura 4.24: Gráfico de probabilidad normal de los residuos estandarizados de nivel 2 del modelo con ordenada aleatoria, con una variable explicativa.

Como se puede apreciar los puntos no se ajustan a una recta, por lo que se procede a realizar una prueba de (K-S) contrastando las siguientes hipótesis:

H_0 : Los residuos de nivel 2 se ajustan a una distribución normal.

H_1 : Los residuos de nivel 2 no se ajustan a una distribución normal.

One-Sample Kolmogorov-Smirnov Test

		Residuos
N		12
Normal Parameters ^a	Mean	.0000
	Std. Deviation	2.29197
Most Extreme Differences	Absolute	.218
	Positive	.218
	Negative	-.177
Kolmogorov-Smirnov Z		.754
Asymp. Sig. (2-tailed)		.621

a. Test distribution is Normal.

Figura 4.25: Prueba de normalidad de (K-S) de los residuos de nivel 2 del modelo con ordenada aleatoria, con una variable explicativa.

como el nivel de significación es mayor a 0.05, se acepta la hipótesis H_0 , es decir, los residuos de nivel 2 se ajustan a una distribución normal.

4.2.3 Modelo con ordenada y pendiente aleatorias

El modelo con ordenada aleatoria se presenta:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{0ij}$$

Si se tiene que β_1 depende del grupo j:

$$\begin{aligned} y_{ij} &= \beta_{0j} + \beta_{1j}x_{ij} + e_{0ij} \\ \beta_{0j} &= \beta_0 + u_{0j} \\ \beta_{1j} &= \beta_1 + u_{1j} \end{aligned}$$

de manera simplificada:

$$y_{ij} = \beta_0 + \beta_1x_{ij} + u_{0j} + u_{1j}x_{ij} + e_{0ij} \quad (4.5)$$

El modelo planteado (4.5) en MLwin es el siguiente:

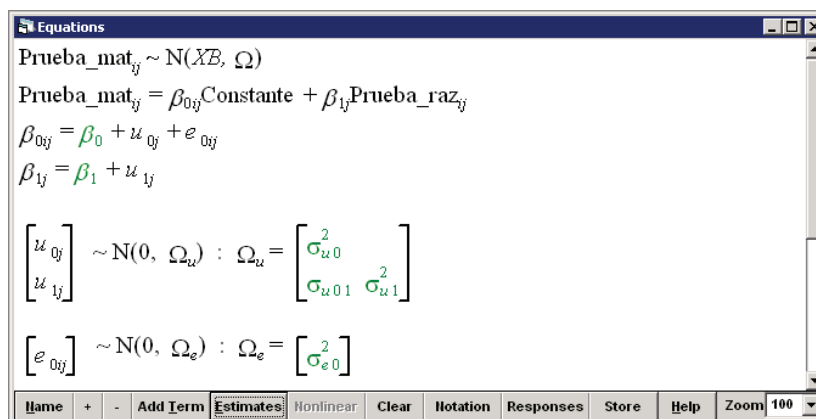


Figura 4.26: Modelo con ordenada y pendiente aleatorias.

Este modelo viene a ser más complejo que los modelos vistos anteriormente, ya que la pendiente β_{1j} varía aleatoriamente entre los colegios. Los parámetros β_0 y β_1 son la parte fija del modelo. Los términos u_{0j} y u_{1j} son las variables residuales al nivel colegio y e_{0ij} la variable residual al nivel estudiante. Los términos u_{0j} y u_{1j} siguen una distribución Normal multivariante (en este caso sería bivalente) con media cero vector y matriz de covarianza Ω_u , de tamaño 2 por 2. Donde σ_{u0}^2 representa

la variación entre las ordenadas de los colegios, $\sigma_{u_1}^2$ representa la variación entre las pendientes de los colegios y σ_{u_01} es la covarianza entre pendientes y ordenadas de los colegios. Los resultados del modelo (4.5) con estimación de MCGI son:

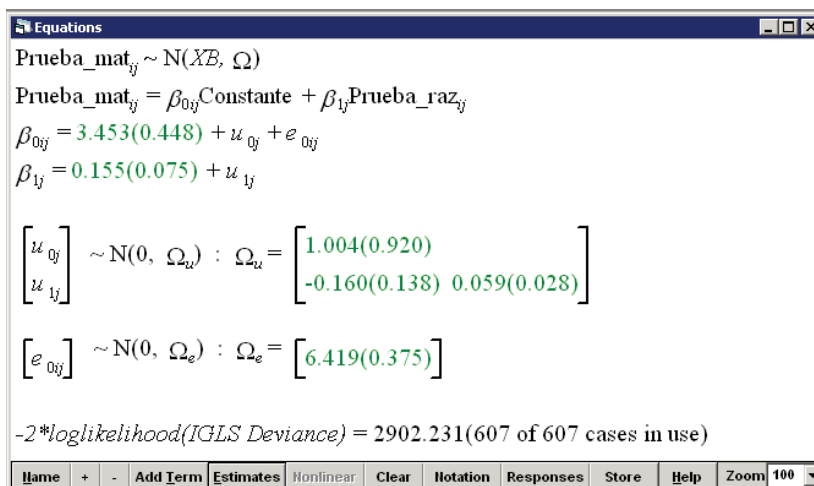


Figura 4.27: Resultados del modelo con ordenada y pendiente aleatorias con estimación de MCGI.

Debido que el modelo (4.5) presenta dos parámetros adicionales al modelo anterior (4.3), la varianza de la pendiente residual u_{1j} y la covarianza entre la ordenada y u_{0j} , se contrastan las siguientes hipótesis para cerciorar su significación.

H_0 : Los datos se ajustan al modelo (4.3)

H_1 : Los datos se ajustan al modelo (4.5)

Modelo (4.3)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{0ij}$
Modelo (4.5)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + e_{0ij}$

Se rechaza H_0 al nivel 5%, debido a que la diferencia de las desviaciones de los modelos (4.3) y (4.5) es mayor que el cuantil de la distribución χ^2 con dos grados de libertad (debido a la inclusión de dos términos en el modelo), ya que $31.54 > 5.99(\chi^2(2))$.

El valor de la pendiente $\hat{\beta}_1$ es 0.155 con error estándar de 0.075. Las pendientes de los colegios varían alrededor de la media de 0.059 ($\hat{\sigma}_{u_1}^2$). Las ordenadas de los colegios también difieren, con media de 3.453 (error estándar de 0.448) y varianza $\hat{\sigma}_{u_0}^2$ de 1.004 (error estándar de 0.920).

Para la interpretación de la varianza de las pendientes aleatorias, $\hat{\sigma}_{u_1}^2$, se debe considerar en primer lugar que el promedio de las pendientes aleatorias es $\hat{\beta}_1$ y que

la pendiente que representa al grupo j es $\beta_1 + u_{1j}$. Debido a que u_{1j} , es una variable aleatoria que sigue una distribución normal con media $\hat{\beta}_1 = 0.155$ y desviación estándar $\sqrt{\hat{\sigma}_{u1}^2} = 0.24$, entonces aproximadamente el 95 % de los grupos tienen pendientes entre $\hat{\beta}_1 - 1.96 \hat{\sigma}_{u1}$ y $\hat{\beta}_1 + 1.96 \hat{\sigma}_{u1}$. Es decir, que las pendientes aleatorias se encuentran entre los valores de -0.3154 y 0.6254 . Esto implica que el efecto de la *Prueba raz* es positivo para algunos colegios y negativo para otros.

La covarianza ($\hat{\sigma}_{u01}$) entre las ordenadas y las pendientes estimadas es de -0.160 (error estándar de 0.138), el signo negativo expresa la tendencia en la relación lineal entre las ordenas y las pendientes, lo que sugiere que los colegios con ordenadas más grandes tienden a tener pendientes menos pronunciadas y esto corresponde la correlación entre la ordena y la pendiente (entre colegios). La correlación entre la pendiente aleatoria y la ordenada aleatoria es:

$$\frac{-0.160}{\sqrt{(1.004)(0.059)}} = -0.66$$

La correlación negativa entre la pendiente y la ordenada significa, que el rendimiento promedio más alto tiende a ser logrado más, por las altas notas en matemáticas de los estudiantes menos racionios que por las más altas notas de los estudiantes más racionios.

Los resultados de los estudiantes varían alrededor de los colegios, residuos de nivel 1 (\hat{e}_{0ij}), con varianza estimada es 6.419 (error estándar de 0.375).

La desviación estándar de la variable *Prueba raz* es de 4.66 , de todas las pruebas tomadas en los diferentes colegios. Debido a que la variable sigue una distribución normal, alrededor del 95 % de las notas se encuentran entre los valores de 4.30 y 22.56 . Para estudiantes con notas en la prueba de razonamiento entre los valores anteriores se tiene:

$$\begin{aligned} \text{Var}(y_{ij} | \text{Prueba raz}_{ij} = 4.30) &= \hat{\sigma}_{u0}^2 + 2x_{ij}\hat{\sigma}_{u01} + x_{ij}^2\hat{\sigma}_{u1}^2 + \hat{\sigma}_{e0}^2 \\ &= 1.004 + 2(-0.160)(4.30) + (0.059)(4.30)^2 + 6.419 \\ &= 7.14 \end{aligned}$$

$$\begin{aligned} \text{Var}(y_{ij} | \text{Prueba raz}_{ij} = 22.56) &= \hat{\sigma}_{u0}^2 + 2x_{ij}\hat{\sigma}_{u01} + x_{ij}^2\hat{\sigma}_{u1}^2 + \hat{\sigma}_{e0}^2 \\ &= 1.004 + 2(-0.160)(22.56) + (0.059)(22.56)^2 + 6.419 \\ &= 30.23 \end{aligned}$$

Como se observa la varianza del modelo se encuentre entre 7.14 y 30.23. A continuación se procede a calcular la correlación que existe entre dos estudiantes del mismo colegio.

$$\begin{aligned} \text{Cov}(y_{ij}, y_{kj} | \text{Prueba raz}_{ij} = 4.30, \text{Prueba raz}_{kj} = 22.56) &= \hat{\sigma}_{u_0}^2 + \hat{\sigma}_{u_01}(x_{ij} + x_{kj}) + \hat{\sigma}_{u_1}^2 x_{ij} x_{kj} \\ &= 1.004 + 26.86(-0.160) + 0.059(97.10) \\ &= 2.44 \end{aligned}$$

Por lo tanto el coeficiente de correlación es:

$$\begin{aligned} \rho(y_{ij}, y_{kj} | \text{Prueba raz}_{ij} = 4.30, \text{Prueba raz}_{kj} = 22.56) &= \frac{2.44}{\sqrt{(7.14)(30.23)}} \\ &= 0.17 \end{aligned}$$

Esto implica que existe una correlación positiva entre las dos variables, aunque esta correlación es débil.

Gráficamente, a cada grupo se le asocia una recta con ordenada en el origen $\beta_0 + u_{0j}$ y pendiente $\beta_1 + u_{1j}$, de manera que conforman un conjunto de rectas que se cruzan como se muestra a continuación.

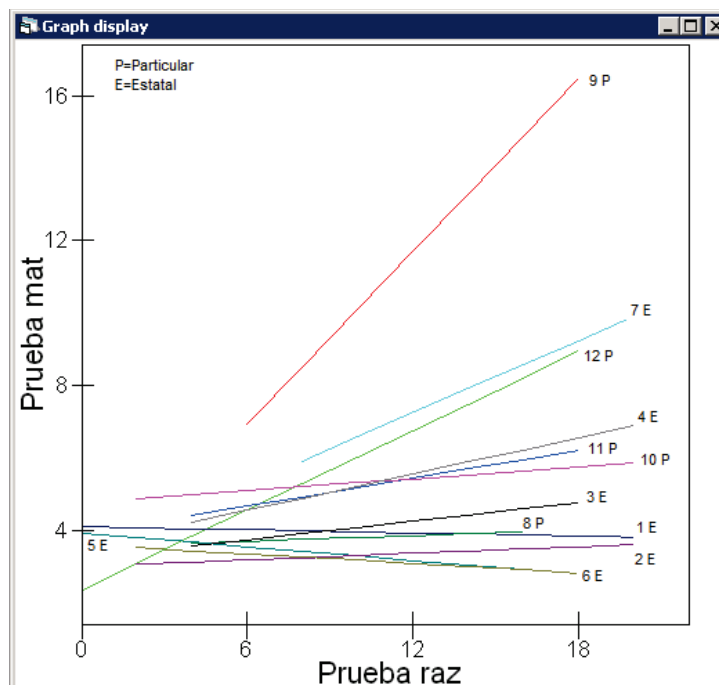


Figura 4.28: Regresiones de 12 colegios según el modelo con ordenada y pendiente aleatorias.

En la figura (4.28), el parámetro β_{0j} y β_{1j} , varían aleatoriamente de colegio a colegio, debido a que los residuos de segundo nivel u_{0j} modifica el término de la ordenada y u_{1j} modifica el término de la pendiente. Es por esta razón que las regresiones se muestran como rectas cruzadas. Además se observa la identificación de cada colegio por un color distinto, con el número correspondiente a cada colegio (Anexo D) y las letras E o P, para indicar si este colegio es Estatal o Privado. Además se puede identificar que para pocos colegios existe una fuerte correlación entre las variables *Prueba mat* y *Prueba raz*, ya que estudiantes que obtuvieron altas calificaciones en matemática también lo obtuvieron en razonamiento, esto se tiene en los colegios: 9P, 7E y 12P. A diferencia del resto de colegios que como se puede observar esta correlación es débil, debido a que se tiene estudiantes que obtuvieron altas calificaciones en razonamiento y bajas calificaciones en matemática.

A continuación se presenta el gráfico comparativo del modelo (4.5) con el modelo anterior (4.3), con el fin de observar la importancia de la variabilidad de la pendiente en los modelos.

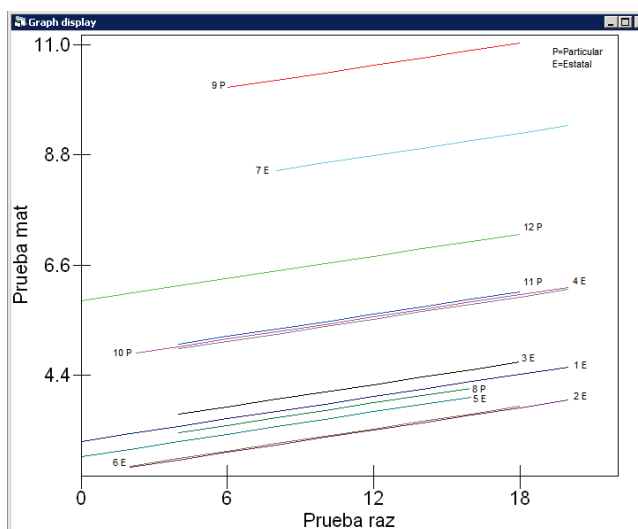


Figura 4.29: Modelo con ordenada aleatoria y pendiente fija.

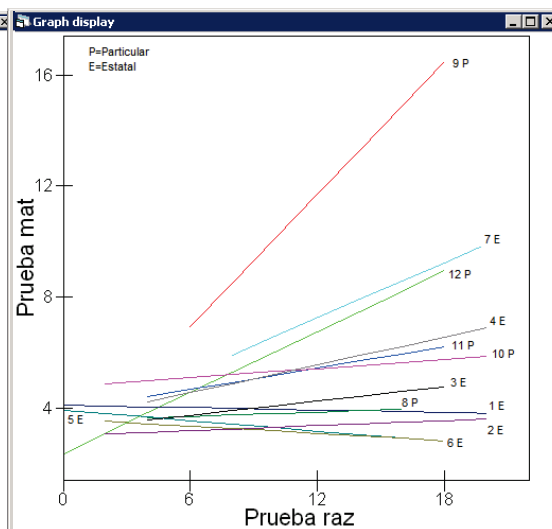


Figura 4.30: Modelo con ordenada y pendiente aleatorias.

Como se puede apreciar el modelo con ordenada y pendiente aleatorias refleja el real desempeño de los estudiantes en las pruebas en los distintos colegios, a diferencia del modelo con pendiente fija en donde no se puede observar la variabilidad que existe entre los colegios.

Heteroscedasticidad

Como se observa no solamente los estudiantes dentro de un mismo colegio tie-

nen valores de la variable respuesta correlacionados, sino que también la varianza del modelo (4.5) depende de los valores de x , es decir, que el modelo presenta heteroscedasticidad:

$$Var(y_{ij}|Prueba\ raz_{ij} = x_{ij}) = 1.004 - 0.32x_{ij} + 0.059x_{ij}^2 + 6.419$$

La expresión de la varianza para el modelo (4.5) es un polinomio cuadrático con respecto a x_{ij} . Derivando esta ecuación con respecto a x e igualándola a cero, se tiene que alcanza su mínimo en:

$$\begin{aligned} x_{ij}^* &= -\frac{\hat{\sigma}_{u01}}{\hat{\sigma}_{u1}^2} \\ &= \frac{0.160}{0.059} \\ &= 2.71 \end{aligned}$$

El gráfico de la varianza del modelo se presenta a continuación:

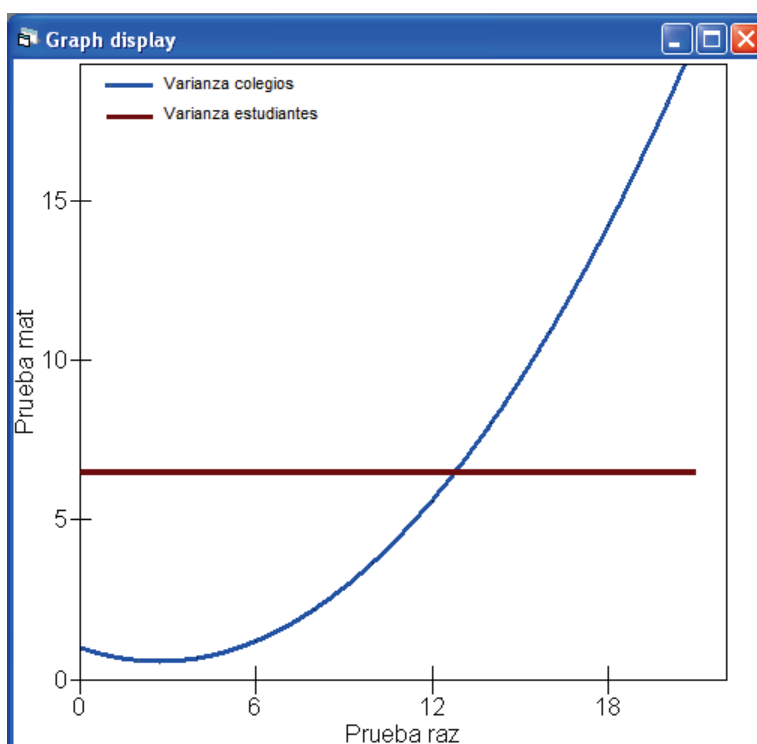


Figura 4.31: Función de la varianza del modelo con ordenada y pendiente aleatorias.

La línea curva de color azul es la variación entre colegios y la línea recta de color rojo es la variación entre los estudiantes, que como se puede observar es constante. Como se determinó anteriormente la varianza es mínima para $x = 2.71$, fuera

del rango de valores para la variable *Prueba raz* que va de: 4.30 a 22.56; es decir, la varianza residual a nivel colegios es una función creciente de x_{ij} . Esto implica que los colegios son más variables para estudiantes con altas notas en razonamiento, mientras que la variación entre estudiantes es constante con la variable *Prueba raz*.

A continuación se presentan los resultados de las estimaciones de los parámetros del modelo (4.5) utilizando el método de CMMC con el muestreador de Gibbs y el método de Metrópolis-Hastings (MH), con información a priori para la estimación de los parámetros, como se detalla en la tabla.

Par.	CMMC-Gibbs				CMMC-MH				MCGI	
	$\Gamma^{-1}(1; 0.001)$		Uniforme		$\Gamma^{-1}(1; 0.001)$		Uniforme		Media	e.e.
β_0	3.445	0.530	3.424	0.762	3.439	0.507	3.455	0.762	3.453	0.448
β_1	0.159	0.081	0.162	0.121	0.160	0.079	0.149	0.128	0.155	0.075
σ_{u0}^2	1.650	1.371	5.168	5.701	1.624	1.324	5.178	5.875	1.004	0.920
σ_{u01}	-0.249	0.231	-0.637	0.817	-0.244	0.222	-0.643	0.832	-0.160	0.138
σ_{u1}^2	0.085	0.049	0.167	0.152	0.084	0.048	0.171	0.155	0.059	0.028
σ_{e0}^2	6.432	0.376	6.456	0.379	6.431	0.377	6.458	0.380	6.419	0.375
Des.	2, 853.020		2, 851.716		2, 853.342		2, 851.730		2, 902.231	
DIC	2, 871.520		2, 872.910		2, 872.080		2, 872.920			
Cad.	150, 000		155, 000		267, 000		350, 000			

Tabla 4.5: Comparación de estimaciones CMMC-Gibbs con dos distribuciones a priori, CMMC-MH con dos distribuciones a priori y MCGI del modelo con ordena y pendiente aleatorias.

Antes de realizar un análisis general de los resultados de las estimaciones se procede a realizar un test de significación de los parámetros fijos y aleatorios del modelo (4.5) que se presenta a continuación:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + e_{0ij}$$

Prueba de Hipótesis

Para el parámetro $\hat{\beta}_0$ se plantean las siguientes hipótesis:

$$H_0 : \hat{\beta}_0 = 0$$

$$H_1 : \hat{\beta}_0 \neq 0$$

Se rechaza H_0 al nivel 5 %, ya que para todos los métodos de estimación $\frac{\hat{\beta}_0}{ee(\hat{\beta}_0)} \geq 1.96$.

La prueba de significación del parámetro $\hat{\beta}_1$, contrastan las siguientes hipótesis:

$$H_0 : \hat{\beta}_1 = 0$$

$$H_1 : \hat{\beta}_1 \neq 0$$

Se rechaza H_0 al nivel 5 %, ya que $\frac{\hat{\beta}_1}{ee(\hat{\beta}_1)} \geq 1.96$ para los métodos con CMMC-MH y CMMC-Gibbs con una distribución a priori de una gamma inversa y para el método de MCGI.

Para la parte aleatoria del modelo (4.5), se lo compara con el modelo anterior (4.3) con sus respectivos tipos de estimación y se contrastan las siguientes hipótesis:

H_0 : Los datos se ajustan al modelo (4.3)

H_1 : Los datos se ajustan al modelo (4.5)

Modelo (4.3)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{0ij}$
Modelo (4.5)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + e_{0ij}$

Se rechaza H_0 al nivel 5 %, debido a que la diferencia de las desviaciones de los modelos (4.3) y (4.5) con sus respectivos tipos de estimación es mayor que el cuantil de la distribución χ^2 con dos grados de libertad. Es decir, la pendiente aleatoria del modelo es significativa con todos los tipos de estimación.

De los resultados obtenidos en los test de significación de la estimación de los parámetros, se tuvo que el parámetro β_1 no es significativo con la estimación mediante CMMC con el método de Gibbs y MH utilizando como distribución a priori una uniforme. Ya que los resultados con esta distribución fueron altos valores de sus errores estándar, lo que nos indica que la variabilidad del parámetro con esta distribución a priori es alta. Este comportamiento de una alta variabilidad en la estimación de los parámetros que componen la matriz $\hat{\Omega}$ también se presentó utilizando como distribución a priori una uniforme utilizando CMMC.

Con esto podemos concluir, que la utilización de la distribución uniforme como información a priori para estimación de los parámetros con los dos tipos de métodos, no obtuvieron buenos resultados. Al contrario, la estimación con una distribución de una gamma inversa como distribución a priori obtuvo los mejores resultados de las estimaciones con los dos tipos de métodos de CMMC.

En general los resultados de las estimaciones con los métodos de CMMC fueron mejores, que con el método de MCGI, debido q que se obtuvo un mejor ajuste del

modelo.

Residuos

Primeramente se presenta un gráfico oruga de los residuos de nivel 1 del modelo (4.5). Como se realizó un análisis detallado de los residuos a nivel estudiantes en el modelo (4.1), donde se concluyó que presuntos puntos atípicos no sería descartados del análisis ya que estos representan el alto y bajo rendimiento de los estudiantes en las diferentes pruebas. Por tal motivo únicamente se constatarán sus hipótesis de normalidad en este modelo.

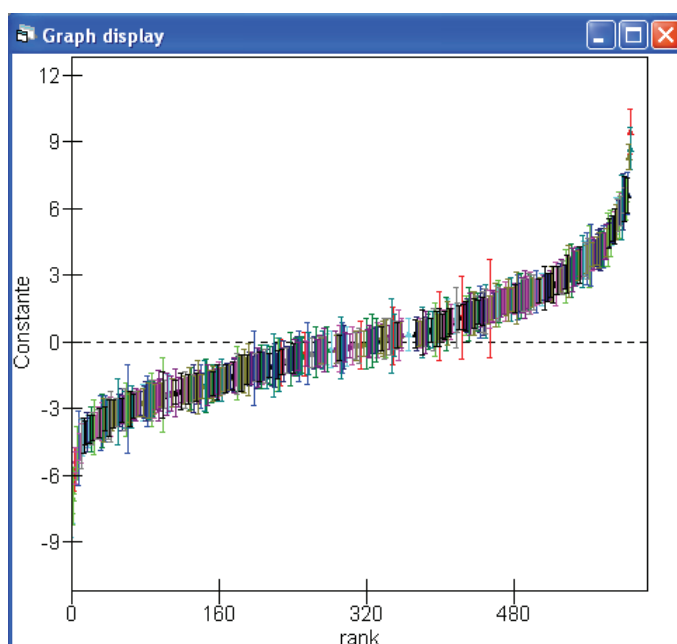


Figura 4.32: Residuos de nivel 1 del modelo con ordenada y pendiente aleatorias

A continuación se presenta el gráfico de normalidad de los residuos de nivel 1, que es lineal, lo que sugiere que la hipótesis de normalidad es aceptable.

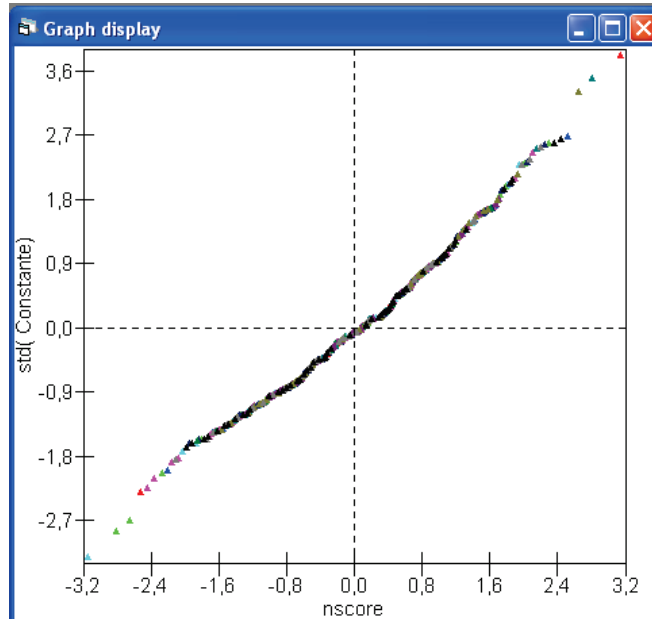


Figura 4.33: Gráfico de normalidad de los residuos de nivel 1 del modelo con ordenada y pendiente aleatorias.

Los residuos de nivel 2 \hat{u}_{0j} y \hat{u}_{1j} se presentan a continuación:

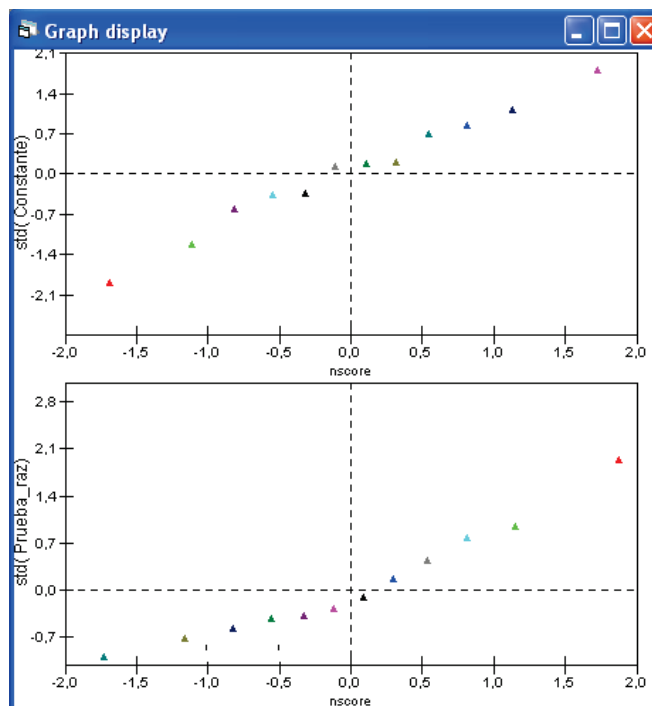


Figura 4.34: Gráfico de normalidad de los residuos de nivel 2 del modelo con ordenada y pendiente aleatorias.

Como se puede apreciar no se ajustan de manera exacta a una recta lineal, ra-

zón por la cual se procede a realizar la prueba de (K-S) contrastando las siguientes hipótesis:

H_0 : Los residuos \hat{u}_{0j} de nivel 2 se ajustan a una distribución normal.

H_1 : Los residuos \hat{u}_{0j} de nivel 2 no se ajustan a una distribución normal.

		Residuos_u0j
N		12
Normal Parameters ^a	Mean	.0000
	Std. Deviation	.72372
Most Extreme Differences	Absolute	.135
	Positive	.100
	Negative	-.135
Kolmogorov-Smirnov Z		.468
Asymp. Sig. (2-tailed)		.981

a. Test distribution is Normal.

Figura 4.35: Prueba de normalidad de (K-S) de los residuos u_{0j} de nivel 2 del modelo con ordenada y pendiente aleatorias.

como el valor de significación es mayor a 0.05, se acepta la hipótesis que los residuos \hat{u}_{0j} siguen una distribución normal. Para los residuos \hat{u}_{1j} se tiene las siguientes hipótesis:

H_0 : Los residuos \hat{u}_{1j} de nivel 2 se ajustan a una distribución normal.

H_1 : Los residuos \hat{u}_{1j} de nivel 2 no se ajustan a una distribución normal.

		Residuos_u1j
N		12
Normal Parameters ^a	Mean	.0000
	Std. Deviation	.24305
Most Extreme Differences	Absolute	.230
	Positive	.230
	Negative	-.184
Kolmogorov-Smirnov Z		.798
Asymp. Sig. (2-tailed)		.548

a. Test distribution is Normal.

Figura 4.36: Prueba de normalidad de (K-S) de los residuos \hat{u}_{1j} de nivel 2 del modelo con ordenada y pendiente aleatorias.

de igual manera no se rechaza la hipótesis H_0 , con lo que se concluye que los residuos \hat{u}_{1j} se ajustan a una distribución normal.

4.2.4 Modelo lineal jerárquico de nivel 2

Se comenzó inicialmente con un modelo estructural básico (modelo vacío) y poco a poco se fueron incorporando variables fijas y aleatorias al modelo inicial, tratando que cada vez este modelo lineal jerárquico se ajuste de mejor manera a los datos hasta llegar a un modelo final que reúna todas las variables medidas y nos permita evaluar el rendimiento de los estudiantes del cantón Otavalo.

Para el modelo lineal de nivel 2 se van a incorporar las variables categóricas *Mujer* y *Privado*, donde es 1 si es verdadero y 0 caso contrario. La variable *Privado* se refiere al tipo de colegio.

Antes de llegar al modelo que se plantea posteriormente, se comprobó que la variable *Mujer* no es significativa al nivel estudiante; además las variables *Mujer* y *Privado* tampoco resultaron significativas al nivel colegio, es decir, su variabilidad entre los colegios no es considerable. En el Anexo E se presentan los modelos que fueron descartados. Luego de los análisis hechos anteriormente se llegó al siguiente modelo planteado.

Figura 4.37: Modelo lineal jerárquico de nivel 2.

El modelo que se presenta es el siguiente:

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + \beta_2 x_{ij} + \beta_3 x_{ij} + e_{0ij}$$

donde β_1 depende del grupo j y β_3 depende del estudiante i .

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

$$\beta_{3i} = \beta_3 + e_{3ij}$$

de manera simplificada se tiene:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + \beta_2 x_{ij} + \beta_3 x_{ij} + e_{0ij} + e_{3ij} x_{ij} \quad (4.6)$$

Este modelo se diferencia del modelo anteriormente visto (4.5), ya que incorpora la pendiente β_2 que es fija y la pendiente β_{3i} que varía aleatoriamente entre los estudiantes. De esta manera los parámetros: β_0 , β_1 , β_2 y β_3 son la parte fija del modelo. Los términos u_{0j} y u_{1j} son las variables residuales a nivel colegio; e_{0ij} y e_{3ij} son las variables residuales al nivel estudiante. Los términos u_{0j} , u_{1j} y e_{0ij} , e_{3ij} siguen una distribución Normal bivalente con media cero vector y matriz de covarianza Ω_u y Ω_e de tamaño dos por dos respectivamente. A continuación se presentan los resultados del modelo (4.6) con estimación de MCGI:

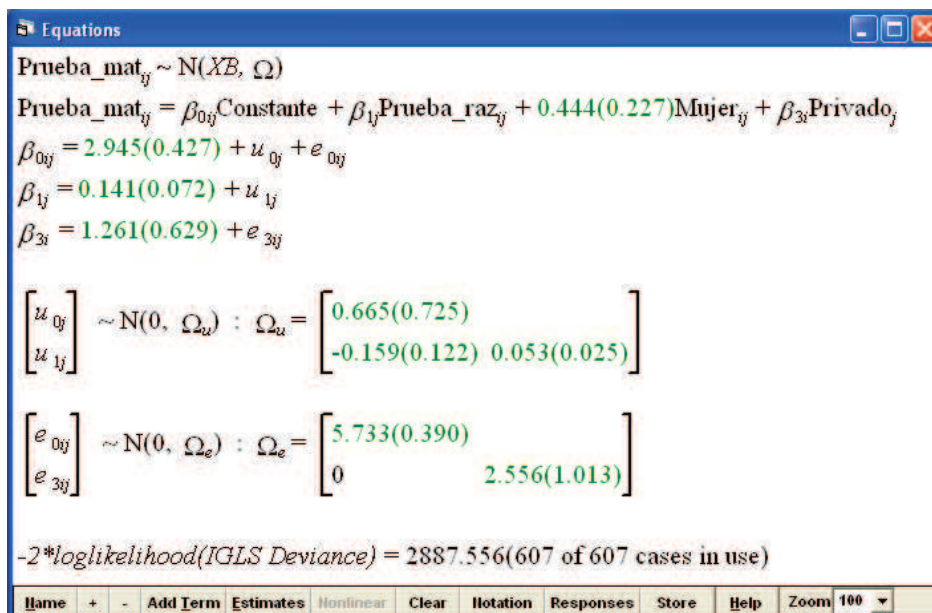


Figura 4.38: Resultados del modelo lineal jerárquico de nivel 2 con estimación de MCGI.

Antes de proceder a un análisis de los valores en la estimación de los parámetros, a continuación se realiza un test de significación de la inclusión de los parámetros añadidos al modelo anterior.

El modelo (4.6) presenta tres parámetros adicionales al modelo anterior (4.5), las pendientes β_2 , β_3 y la varianza residual e_{3ij} . Con esto, se contrastan las siguientes hipótesis para comprobar su significación.

$$H_0 : \text{Los datos se ajustan al modelo (4.5)}$$

H_1 : Los datos se ajustan al modelo (4.6)

Modelo (4.5)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + e_{0ij}$
Modelo (4.6)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + \beta_2 x_{ij} + \beta_3 x_{ij} + e_{0ij} + e_{3ij} x_{ij}$

Se rechaza H_0 al nivel 5 %, debido a que la diferencia de las desviaciones de los modelos (4.5) y (4.6) es mayor que el cuantil de la distribución χ^2 con tres grados de libertad (debido a la inclusión de tres términos en el modelo), ya que $14.68 > 7.81(\chi^2(3))$.

Para este modelo se han incorporado dos variables categóricas. La variable *Mujer* no fue significativa al nivel estudiante y al nivel colegio (Anexo E), con lo que podemos concluir que los resultados de las pruebas tanto en hombres como en mujeres no difieren de manera significativamente entre colegios y los estudiantes. A diferencia la variable *Privado* que resultó significativa su variabilidad a nivel estudiante.

Por dichas razones se llegó al planteamiento del modelo (4.6) y que presenta un mejor ajuste a los datos como se comprobó anteriormente. Debido a que el modelo presenta dos variables categóricas se tendría cuatro posibles subgrupos como se presenta en el cuadro a continuación.

Subgrupos	Valores de variables categóricas		Valor promedio
	mujer	privado	
hombre en estatal	0	0	2.945
mujer en estatal	1	0	$2.945 + 0.444$
hombre en privado	0	1	$2.945 + 1.261$
mujer en privado	1	1	$2.945 + 0.444 + 1.261$

Tabla 4.6: Cuadro de predicciones del modelo lineal jerárquico de nivel 2.

Analizando estos resultados podemos decir que las mujeres en colegio estatal es 0.444 puntos mejor que los hombres en colegio estatal. Las mujeres en colegio privado tienen 1.261 puntos más que las mujeres en colegio estatal y $(0.444 + 1.261)$ puntos mejor que los hombres en colegio estatal. Los hombres en colegio privado son 1.261 puntos mejor que los hombres en colegio estatal. Estos resultados fueron interpretados cuando la variable (*Prueba raz*) es igual a cero.

De manera general podemos concluir que se obtuvieron mejores resultados de las pruebas en colegios privados que en colegios estatales y fueron las mujeres las

que tuvieron mejores notas que los hombres, aunque esta diferencia es muy leve.

La desviación estándar de todas las pruebas de razonamiento tomadas en los diferentes colegios es de 4.66. Debido a que dicha variable sigue una distribución normal, alrededor del 95 % de las notas se encuentran entre los valores de 4.30 y 22.56.

$$\begin{aligned} Var(y_{ij} | Prueba\ raz_{ij} = 4.30) &= \hat{\sigma}_{u0}^2 + 2x_{ij}\hat{\sigma}_{u01} + x_{ij}^2\hat{\sigma}_{u1}^2 + \hat{\sigma}_{e0}^2 + 2x_{ij}\hat{\sigma}_{e03} + x_{ij}^2\hat{\sigma}_{e3}^2 \\ &= 0.665 - 1.367 + 0.980 + 5.733 + 47.260 \\ &= 53.27 \end{aligned}$$

$$\begin{aligned} Var(y_{ij} | Prueba\ raz_{ij} = 22.56) &= \hat{\sigma}_{u0}^2 + 2x_{ij}\hat{\sigma}_{u01} + x_{ij}^2\hat{\sigma}_{u1}^2 + \hat{\sigma}_{e0}^2 + 2x_{ij}\hat{\sigma}_{e03} + x_{ij}^2\hat{\sigma}_{e3}^2 \\ &= 0.665 - 3.587 + 26.975 + 5.733 + 1,300.885 \\ &= 1,330.67 \end{aligned}$$

Es decir, la varianza del modelo se encuentra entre 53.27 y 1330.67.

Heteroscedasticidad

Como se vio anteriormente la varianza del modelo (4.6) depende de los valores x , es decir, que el modelo presenta heteroscedasticidad a nivel estudiante y a nivel colegio. A nivel estudiante se tiene:

$$\begin{aligned} Var(y_{ij} | x_{ij}) &= \hat{\sigma}_{e0}^2 + 2x_{ij}\hat{\sigma}_{e03} + x_{ij}^2\hat{\sigma}_{e3}^2 \\ &= 5.733 + 2.556x_{ij}^2 \end{aligned}$$

La heteroscedasticidad a nivel colegio es:

$$\begin{aligned} Var(y_{ij} | x_{ij}) &= \hat{\sigma}_{u0}^2 + 2x_{ij}\hat{\sigma}_{u01} + x_{ij}^2\hat{\sigma}_{u1}^2 \\ &= 0.665 - 0.318x_{ij} + 0.053x_{ij}^2 \end{aligned}$$

La expresión de la varianza del modelo (4.6) es un polinomio cuadrático con respecto a x_{ij} . Derivando esta ecuación con respecto a x e igualándola a cero, se tiene que alcanza su mínimo en $x_{ij}^* = 5.30$.

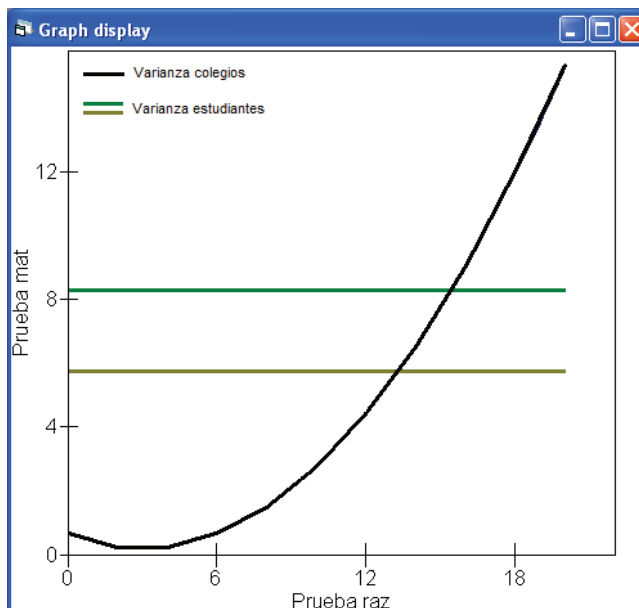


Figura 4.39: Función de la varianza del modelo lineal jerárquico de nivel 2.

El gráfico de la varianza del modelo, presenta la variación entre colegios con la línea curva de color negro y las líneas de color verde es la variación entre los estudiantes. Anteriormente se determinó que la varianza mínima a nivel de los colegios es para $x = 5.30$ fuera del rango de valores para la variable *Prueba raz* que va de 53.27 a 1330.67, es decir, la varianza residual a nivel colegio es una función creciente de x_{ij} . Esto implica que los colegios son más variables para estudiantes con altas notas en razonamiento, mientras que la varianza a nivel estudiante es 5.733 en colegio estatal y 8.289 en colegio privado, lo que nos indica que los estudiantes en colegio privado tienen notas más variables.

A continuación se presenta una tabla resumen de las estimaciones de los parámetros del modelo (4.6) con distintos tipos de estimación.

	CMMC-Gibbs				MCGI	
	$\Gamma^{-1}(1; 0.001)$		Uniforme			
Parámetro	Media	e.e.	Media	e.e.	Media	e.e.
β_0	2.970	0.520	3.003	0.813	2.945	0.427
β_1	0.144	0.073	0.143	0.090	0.141	0.072
β_2	0.446	0.228	0.440	0.234	0.444	0.227
β_3	1.255	0.638	1.243	1.033	1.261	0.629
σ_{u0}^2	1.285	1.116	5.169	6.815	0.665	0.725
σ_{u01}	-0.254	0.203	-0.596	0.843	-0.159	0.122
σ_{u1}^2	0.079	0.044	0.143	0.128	0.053	0.025
σ_{e0}^2	5.786	0.390	5.755	0.393	5.733	0.390
σ_{e3}^2	2.807	1.081	2.717	1.109	2.556	1.013
Desviación	2,844.308		2,847.399		2,887.556	
DIC	2,866.920		2,867.610			
Cadena	5,000		5,000			

Tabla 4.7: Comparación de estimaciones CMMC-Gibbs con dos distribuciones a priori y MCGI del modelo lineal jerárquico de nivel 2.

Antes de realizar un análisis general de los resultados obtenidos procedemos a realizar un test de significación de los parámetros del modelo (4.6) que se detalla a continuación:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + \beta_2 x_{ij} + \beta_3 x_{ij} + e_{0ij} + e_{3ij} x_{ij}$$

Prueba de Hipótesis

Para los parámetros fijos $\hat{\beta}_i$, con $i = 0, \dots, 3$ se plantean las siguientes hipótesis:

$$H_0 : \hat{\beta}_i = 0$$

$$H_1 : \hat{\beta}_i \neq 0$$

Se rechaza H_0 al nivel 5%, para todos los $\hat{\beta}_i$, con $i = 0, \dots, 3$ para las estimaciones con CMMC-Gibbs con una gamma inversa como distribución a priori y MCGI. A diferencia del método de CMMC-Gibbs con una distribución uniforme a priori donde se rechaza H_0 al nivel 5%, únicamente para $\hat{\beta}_0$.

Para la parte aleatoria del modelo, se procede a comparar los resultados obtenidos con una nueva estimación del modelo (4.6) pero sin considerar la variabilidad a nivel estudiante de la variable *Privado*.

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + \beta_2 x_{ij} + \beta_3 x_{ij} + e_{0ij} \quad (4.7)$$

Los resultados son los siguientes:

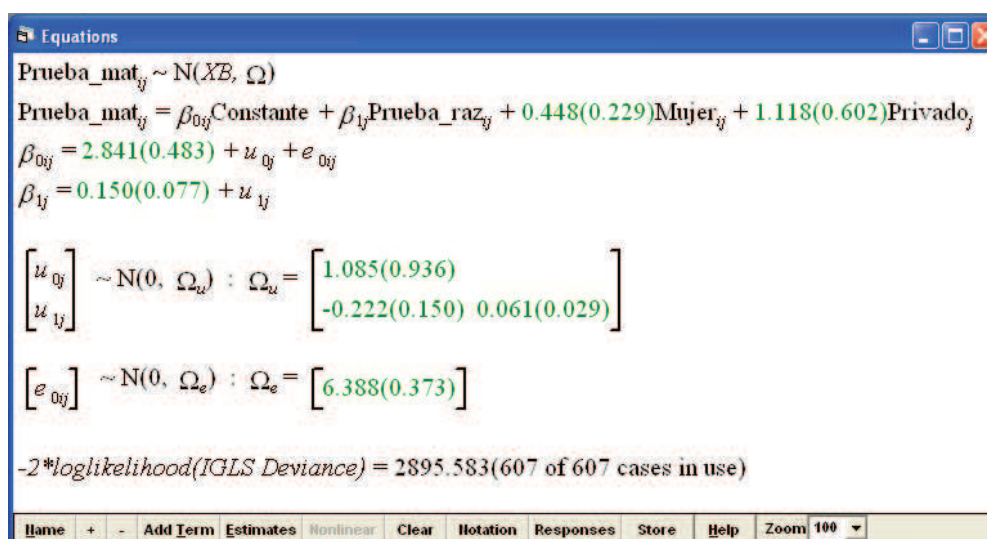


Figura 4.40: Resultados del modelo lineal jerárquico de nivel 2 sin el término e_{3ij} .

La exclusión del término e_{3ij} tiene como efecto mayores valores de la varianza $\hat{\sigma}_{e0}$ y de la matriz Ω_u .

Ahora se procede a realizar la prueba de significación donde se contrastan las siguientes hipótesis:

H_0 : Los datos se ajustan al modelo (4.7)

H_1 : Los datos se ajustan al modelo (4.6)

Modelo (4.7)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + \beta_2 x_{ij} + \beta_3 x_{ij} + e_{0ij}$
Modelo (4.6)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + \beta_2 x_{ij} + \beta_3 x_{ij} + e_{0ij} + e_{3ij} x_{ij}$

Se rechaza H_0 al nivel 5%, debido a que la diferencia de las desviaciones de los modelos (4.7) y (4.6) con sus respectivos tipos de estimación es mayor que el cuantil de la distribución χ^2 con un grado de libertad. Es decir, la variabilidad de la variable *Privado* a nivel estudiante es significativa con todos los tipos de estimación.

Con esto se puede decir, que con los métodos de MCGI y CMMC-Gibbs (con una distribución gamma inversa a priori) todos los parámetros estimados, tanto la parte fija del modelo como la parte aleatoria, son significativos. Además que la estimación con el método de CMMC obtuvo un mejor ajuste que con el método de MCGI.

A diferencia de la estimación con CMMC-Gibbs (con una distribución uniforme a priori) la parte fija del modelo no es significativa excluyendo la ordenada ($\hat{\beta}_0$). Gran parte de estos resultados es debido a que la utilización de la distribución uniforme como información a priori tiende a elevar los valores de los errores estándar para cada parámetro estimado y de manera especial para los elementos de la matriz de varianza estimada $\hat{\Omega}_u$ y $\hat{\Omega}_e$.

Residuos

El gráfico de normalidad de los residuos de nivel 1 presenta una tendencia lineal,

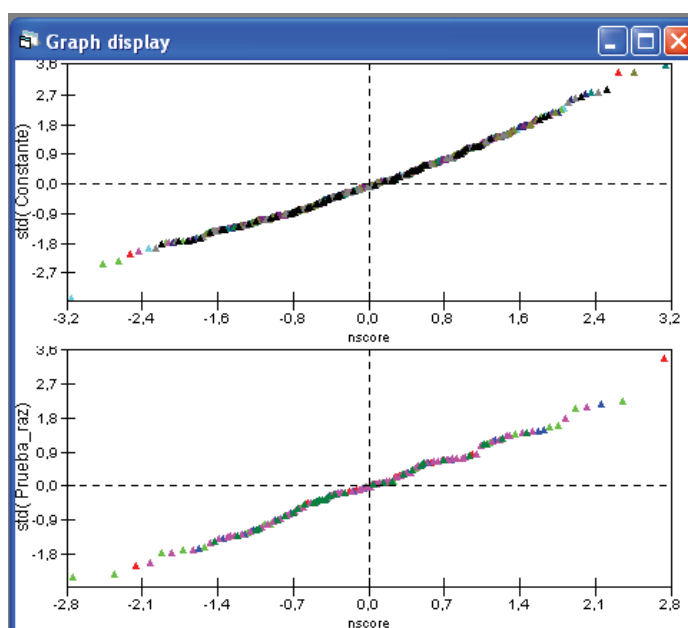


Figura 4.41: Gráfico de normalidad de los residuos de nivel 1 del modelo lineal jerárquico de nivel 2.

verificando la hipótesis de normalidad.

Los residuos de nivel 2 \hat{u}_{0j} y \hat{u}_{1j} se presentan en la figura a continuación:

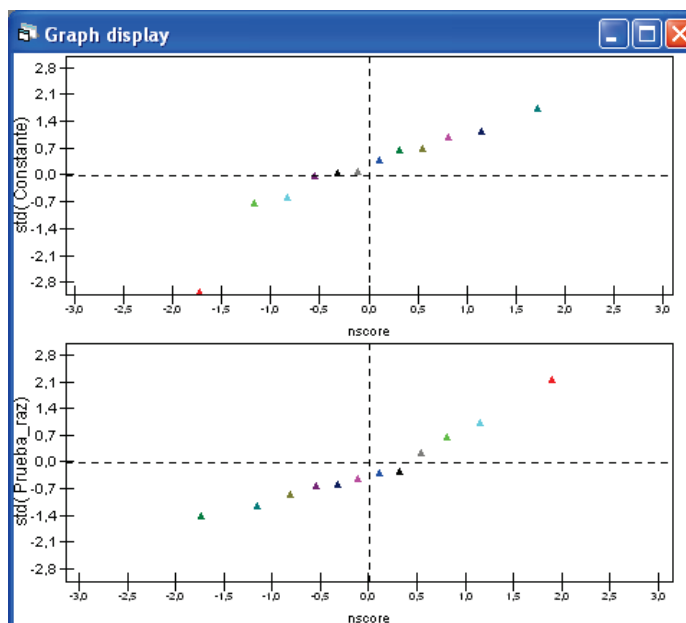


Figura 4.42: Gráfico de normalidad de los residuos de nivel 2 del modelo lineal jerárquico de nivel 2.

Con los gráficos de residuos no se podría afirmar que siguen una distribución normal, por lo que se procede a realizar una prueba de hipótesis:

H_0 : Los residuos \hat{u}_{0j} de nivel 2 se ajustan a una distribución normal.

H_1 : Los residuos \hat{u}_{0j} de nivel 2 no se ajustan a una distribución normal.

One-Sample Kolmogorov-Smirnov Test		Residuos_u0j
N		12
Normal Parameters ^a	Mean	.0000
	Std. Deviation	.66775
Most Extreme Differences	Absolute	.242
	Positive	.171
	Negative	-.242
Kolmogorov-Smirnov Z		.838
Asymp. Sig. (2-tailed)		.484

a. Test distribution is Normal.

Figura 4.43: Prueba de normalidad de (K-S) de los residuos \hat{u}_{0j} de nivel 2 del modelo lineal jerárquico de nivel 2.

Como el valor de significación es mayor que 0.05, se concluye que los residuos \hat{u}_{0j} siguen una distribución normal. De la misma manera se plantea la prueba de

hipótesis para los residuos u_{1j} .

		Residuos u1j
N		12
Normal Parameters ^a	Mean	.0000
	Std. Deviation	.23145
Most Extreme Differences	Absolute	.254
	Positive	.254
	Negative	-.202
Kolmogorov-Smirnov Z		.878
Asymp. Sig. (2-tailed)		.424

a. Test distribution is Normal.

Figura 4.44: Prueba de normalidad de (K-S) de los residuos \hat{u}_{0j} de nivel 2 del modelo lineal jerárquico de nivel 2.

Con esto se confirma que los residuos \hat{u}_{1j} siguen una distribución normal.

4.3 Modelos de tres niveles y estructuras jerárquicas más complejas

4.3.1 Modelo lineal jerárquico de nivel 3

Para el modelo de nivel 3, se considera una estructura jerárquica de 3 niveles: las unidades de nivel 1 son los 607 *estudiantes* de los décimos años de Educación Básica, las unidades de nivel 2 son los *paralelos* y las unidades de nivel 3 son los 12 *colegios*.

Del modelo lineal jerárquico de nivel 2 obtenido anteriormente, se incorporó para este modelo de nivel 3 un nuevo nivel que son los paralelos de los distintos colegios de los décimos años y la variación de la variable *Mujer* a nivel estudiante. La obtención del modelo presentado a continuación es el resultado de una serie de pruebas y validaciones de los parámetros que intervienen en el modelo.

A continuación se presenta el modelo lineal jerárquico de nivel 3 planteado.

Equations

Prueba_mat_{ijk} ~ N(XB, Ω)

Prueba_mat_{ijk} = β_{0ijk}Constante + β_{1k}Prueba_raz_{ijk} + β_{2i}Mujer_{ijk} + β_{3i}Privado_k

β_{0ijk} = β₀ + v_{0k} + u_{0jk} + e_{0ijk}

β_{1k} = β₁ + v_{1k}

β_{2i} = β₂ + e_{2ijk}

β_{3i} = β₃ + e_{3ijk}

$\begin{bmatrix} v_{0k} \\ v_{1k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} \sigma_{v0}^2 & \\ & \sigma_{v1}^2 \end{bmatrix}$

$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 \end{bmatrix}$

$\begin{bmatrix} e_{0ijk} \\ e_{2ijk} \\ e_{3ijk} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma_{e0}^2 & & \\ & \sigma_{e2}^2 & \\ & & \sigma_{e3}^2 \end{bmatrix}$

Name + - Add Term Estimates Nonlinear Clear Notation Responses Store Help Zoom 100%

Figura 4.45: Modelo lineal jerárquico de nivel 3.

El modelo que se plantea es el siguiente:

$$y_{ijk} = \beta_{0jk} + \beta_1 x_{ijk} + \beta_2 x_{ijk} + \beta_3 x_{ijk} + e_{0ijk}$$

donde β_1 depende del grupo k, β_2 y β_3 dependen del estudiante i.

$$\beta_{0jk} = \beta_0 + v_{0k} + u_{0jk}$$

$$\beta_{1k} = \beta_1 + v_{1k}$$

$$\beta_{2i} = \beta_2 + e_{2ijk}$$

$$\beta_{3i} = \beta_3 + e_{3ijk}$$

de manera simplificada se tiene:

$$y_{ijk} = \beta_0 + \beta_1 x_{ijk} + v_{0k} + u_{0jk} + v_{1k} x_{ijk} + \beta_2 x_{ijk} + e_{2ijk} x_{ijk} + \beta_3 x_{ijk} + e_{0ijk} + e_{3ijk} x_{ijk} \quad (4.8)$$

Los parámetros β_0 , β_1 , β_2 y β_3 son la parte fija del modelo. Los términos v_{0k} y v_{1k} son las variables residuales a nivel 3 (colegio); u_{0j} es la variable residual a nivel 2 (paralelo); e_{0ijk} , e_{2ijk} y e_{3ijk} son las variables residuales a nivel 1 (estudiante). Los términos (v_{0k}, v_{1k}) , (u_{0j}) y $(e_{0ijk}, e_{2ijk}, e_{3ijk})$ siguen una distribución Normal con media cero vector y matriz de covarianza Ω_v , Ω_u y Ω_e respectivamente. A continuación se presentan los resultados del modelo (4.8) con estimación de MCGI:

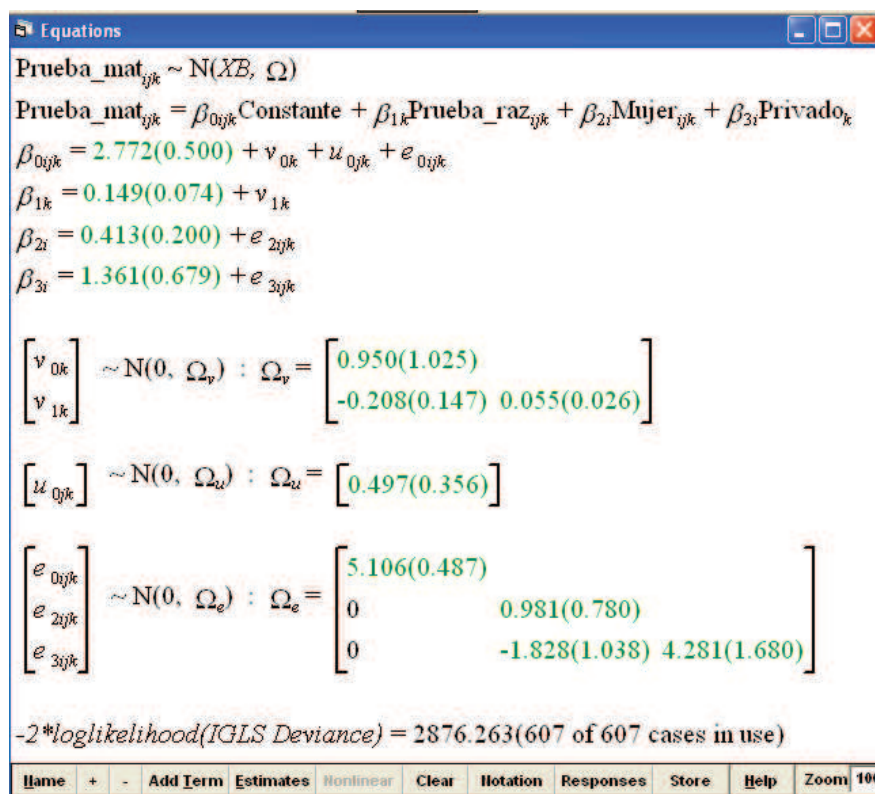


Figura 4.46: Resultados del modelo lineal jerárquico de nivel 3 con estimación de MCGI.

Con los resultados del modelo (4.8), se procede a realizar un test de significación de este modelo con el modelo lineal jerárquico de nivel 2 (modelo (4.6)).

El modelo (4.8) presenta tres parámetros adicionales al modelo anterior (4.6), la varianza residual de nivel 2 (paralelo) u_{0k} , la varianza residual e_{2ijk} de nivel 1 (estudiante) y la covarianza e_{23ijk} . Con esto, se contrastan las siguientes hipótesis para comprobar su significación.

H_0 : Los datos se ajustan al modelo (4.6)

H_1 : Los datos se ajustan al modelo (4.8)

(4.6)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + \beta_2 x_{ij} + \beta_3 x_{ij} + e_{0ij} + e_{3ij} x_{ij}$
(4.8)	$y_{ijk} = \beta_0 + \beta_1 x_{ijk} + v_{0k} + v_{1k} x_{ijk} + u_{0k} + \beta_2 x_{ijk} + e_{2ijk} x_{ijk} + \beta_3 x_{ijk} + e_{0ijk} + e_{3ijk} x_{ijk}$

Se rechaza H_0 al nivel 5 %, debido a que la diferencia de las desviaciones de los modelos (4.6) y (4.8) es mayor que el cuantil de la distribución χ^2 con tres grados de libertad, ya que $11.29 > 7.81(\chi^2(3))$.

Debido a que el modelo presenta dos variables categóricas se tendría cuatro posibles subgrupos como se presenta en el cuadro a continuación.

Subgrupos	Valores de variables categóricas		Valor promedio
	mujer	privado	
hombre en estatal	0	0	2.772
mujer en estatal	1	0	2.772 + 0.413
hombre en privado	0	1	2.772 + 1.361
mujer en privado	1	1	2.772 + 0.413 + 1.361

Tabla 4.8: Cuadro de predicciones del modelo lineal jerárquico de nivel 3.

De igual forma que en el modelo anterior (4.6), los resultados generales obtenidos son los mismos en este modelo de nivel 3. Las mujeres presentaron mejores resultados en las pruebas que los hombres, tanto en colegios privados como en colegios estatales. Los estudiantes de colegios privados obtuvieron mejores calificaciones que los estudiantes de colegios estatales. Estos resultados fueron interpretados cuando la variable *Prueba raz* es igual a cero.

La desviación estándar de todas las pruebas de razonamiento tomadas en los diferentes colegios es de 4.66. Debido a que dicha variable sigue una distribución normal, alrededor del 95 % de las notas se encuentran entre los valores de 4.30 y 22.56.

$$\begin{aligned} Var(y_{ijk}|Prueba\ raz_{ijk} = 4.30) &= \hat{\sigma}_{v0}^2 + 2x_{ijk}\hat{\sigma}_{v01} + x_{ijk}^2\hat{\sigma}_{v1}^2 + \hat{\sigma}_{e0}^2 + 2x_{ijk}\hat{\sigma}_{e02} + x_{ijk}^2\hat{\sigma}_{e2}^2 \\ &\quad + 2x_{ijk}\hat{\sigma}_{e03} + 2x_{ijk}\hat{\sigma}_{e23} + x_{ijk}^2\hat{\sigma}_{e3}^2 + \hat{\sigma}_{u0}^2 \\ &= 87.35 \end{aligned}$$

$$\begin{aligned} Var(y_{ij}|Prueba\ raz_{ij} = 22.56) &= \hat{\sigma}_{v0}^2 + 2x_{ijk}\hat{\sigma}_{v01} + x_{ijk}^2\hat{\sigma}_{v1}^2 + \hat{\sigma}_{e0}^2 + 2x_{ijk}\hat{\sigma}_{e02} + x_{ijk}^2\hat{\sigma}_{e2}^2 \\ &\quad + 2x_{ijk}\hat{\sigma}_{e03} + 2x_{ijk}\hat{\sigma}_{e23} + x_{ijk}^2\hat{\sigma}_{e3}^2 + \hat{\sigma}_{u0}^2 \\ &= 2,620.79 \end{aligned}$$

Es decir, la varianza del modelo se encuentra entre 87.35 y 2,620.79.

Heteroscedasticidad

La heteroscedasticidad a nivel estudiante es:

$$\begin{aligned} Var(y_{ijk}|x_{ijk}) &= \hat{\sigma}_{e0}^2 + 2x_{ijk}\hat{\sigma}_{e02} + x_{ijk}^2\hat{\sigma}_{e2}^2 + 2x_{ijk}\hat{\sigma}_{e03} + 2x_{ijk}\hat{\sigma}_{e23} + x_{ijk}^2\hat{\sigma}_{e3}^2 \\ &= 5.106 + 0.981x_{ijk}^2 - 3.656x_{ijk} + 4.281x_{ijk}^2 \end{aligned}$$

$$= 5.106 - 3.656x_{ijk} + 5.262x_{ijk}^2$$

La heteroscedasticidad a nivel colegio es:

$$\begin{aligned} Var(y_{ijk}|x_{ijk}) &= \hat{\sigma}_{v0}^2 + 2x_{ijk}\hat{\sigma}_{v01} + x_{ijk}^2\hat{\sigma}_{v1}^2 \\ &= 0.950 - 0.416x_{ijk} + 0.055x_{ijk}^2 \end{aligned}$$

A continuación se presenta la siguiente tabla resumen con las estimaciones obtenidas con los distintos métodos.

Par.	CMMC-Gibbs				CMMC-MH				MCGI	
	$\Gamma^{-1}(0.001; 0.001)$		Uniforme		$\Gamma^{-1}(0.001, 0.001)$		Uniforme		Media	e.e.
β_0	2.725	0.667	2.693	1.105	2.594	0.287	2.964	0.318	2.772	0.500
β_1	0.169	0.108	0.154	0.132	0.152	0.021	0.105	0.024	0.149	0.074
β_2	0.421	0.236	0.446	0.236	0.423	0.215	0.415	0.205	0.413	0.200
β_3	1.245	0.782	1.332	1.151	1.602	0.256	1.656	0.245	1.361	0.679
σ_{v0}^2	1.839	1.655	8.048	9.381	1.244	0.590	2.651	1.854	0.950	1.025
σ_{v01}	-0.344	0.274	-0.909	1.083	-0.251	0.125	-0.593	0.425	-0.208	0.147
σ_{v1}^2	0.086	0.054	0.173	0.159	0.063	0.030	0.139	0.101	0.055	0.026
σ_{u0}^2	0.615	0.569	1.056	0.831	0.714	0.260	0.327	0.103	0.497	0.356
σ_{e0}^2	5.289	0.525	5.204	0.512	5.099	0.502	5.245	0.497	5.106	0.487
σ_{e2}^2	0.900	0.806	1.063	0.855	1.257	0.793	1.048	0.782	0.981	0.780
σ_{e23}	-1.806	1.170	-2.218	1.250	-1.802	1.052	-1.909	1.081	-1.828	1.038
σ_{e3}^2	4.557	1.798	4.961	2.020	2.245	1.686	4.336	1.798	4.281	1.6890
Des.	2, 827.257		2, 822.289		2, 817.507		2, 823.003		2, 876.263	
DIC	2, 856.480		2, 854.580		2, 826.350		2, 832.680			
Cad.	5, 000		5, 000		5, 000		5, 000			

Tabla 4.9: Comparación de estimaciones CMMC-Gibbs con dos distribuciones a priori, CMMC-MH con dos distribuciones a priori y MCGI del modelo lineal jerárquico de nivel 3.

Con estos resultados se procede a realizar un test de significación de los parámetros estimados del modelo (4.8), que se presenta a continuación:

$$y_{ijk} = \beta_0 + \beta_1x_{ijk} + v_{0k} + u_{0k} + v_{1k}x_{ijk} + \beta_2x_{ijk} + e_{2ijk}x_{ijk} + \beta_3x_{ijk} + e_{0ijk} + e_{3ijk}x_{ijk}$$

Prueba de Hipótesis

Para los parámetros fijos $\hat{\beta}_i$, con $i = 0, \dots, 3$ se plantean las siguientes hipótesis:

$$\begin{aligned} H_0 &: \hat{\beta}_i = 0 \\ H_1 &: \hat{\beta}_i \neq 0 \end{aligned}$$

Se rechaza H_0 al nivel 5 %, para todos los $\widehat{\beta}_i$, con $i = 0, \dots, 3$ para las estimaciones con CMMC-MH con una gamma inversa, una uniforme como distribuciones a priori y MCGI. A diferencia del método de CMMC-Gibbs donde se rechaza H_0 al nivel 5 %, únicamente para $\widehat{\beta}_0$.

Para la parte aleatoria del modelo (4.8) se lo compara con el modelo lineal jerárquico de nivel 2, con sus respectivas estimaciones y se contrastan las siguientes hipótesis:

H_0 : Los datos se ajustan al modelo (4.6)

H_1 : Los datos se ajustan al modelo (4.8)

(4.6)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + \beta_2 x_{ij} + \beta_3 x_{ij} + e_{0ij} + e_{3ij} x_{ij}$
(4.8)	$y_{ijk} = \beta_0 + \beta_1 x_{ijk} + v_{0k} + u_{0k} + v_{1k} x_{ijk} + \beta_2 x_{ijk} + e_{2ijk} x_{ijk} + \beta_3 x_{ijk} + e_{0ijk} + e_{3ijk} x_{ijk}$

Se rechaza H_0 al nivel 5 %, debido a que la diferencia de las desviaciones de los modelos (4.6) y (4.8) con sus respectivos tipos de estimación es mayor que el cuantil de la distribución χ^2 con tres grados de libertad. Es decir, la inclusión de un nuevo nivel (paralelo) y la variabilidad de la variable *Mujer* a nivel estudiante son significativos con todos los tipos de estimación.

Los resultados obtenidos con la estimación de CMMC-Gibbs de la parte fija del modelo no fue significativa con ninguna de las dos distribuciones a priori. A diferencia de las estimaciones obtenidas con el método de CMMC-MH y MCGI donde si fueron significativos todos los parámetros del modelo (4.8). Las estimaciones con el método de Cadenas de Markov Monte Carlo (CMMC) obtuvieron mejores ajustes que con el método de MCGI, con lo que se puede concluir que la estimación con CMMC-MH con una Gamma inversa presenta el mejor ajuste.

La modelización de los datos en una estructura jerárquica de nivel 3 ha permitido tener mejores resultados en el ajuste de los datos al modelo. Esto se ve reflejado en la disminución del valor de la desviación en los diferentes modelos planteados desde el inicial.

Finalmente se puede decir que los modelos jerárquicos multinivel constituyen uno de los análisis más adecuados para el tratamiento de datos jerarquizados o anidados; convirtiéndose en una estrategia clave para la investigación educativa.

La mejora en la calidad de los resultados de los modelos multinivel a diferencia de la regresión lineal tradicional (a un solo nivel) es clara, debido a que en los mode-

los multinivel se puede estimar la aportación de cada nivel en análisis, considerando que los estudiantes de un mismo grupo comparten una serie de características diferentes a los estudiantes de otros colegios. Esto no se podría analizar en el modelo tradicional debido a que consideraría a los estudiantes como unidades individuales de análisis e independientes ignorando la agrupación de los mismos; y esto se vería reflejado en la subestimación de los parámetros del modelo.

Otro aporte importante de los modelos multinivel es que contienen parámetros fijos y parámetros aleatorios. En los modelos de regresión clásicos los parámetros que se estiman son la ordenada (o punto de corte) y las pendientes, que son parámetros fijos, es decir, comunes a todos los sujetos y son estimados a partir de los datos.

Estas son algunas de las mejoras que se tienen en la utilización de los modelos multinivel.

4.3.2 Modelo de ecuaciones estructurales multinivel

A continuación se presenta la aplicación de los modelos de ecuaciones estructurales multinivel realizando un AFC multinivel. Se consideró los 231 estudiantes de los terceros años de Bachillerato de los 12 colegios que se les aplicó las pruebas de matemática (*MAT*), razonamiento (*RAZ*), lenguaje (*LEN*) y comprensión de lectura (*LEC*).

En la siguiente figura se muestra el diagrama de senderos propuesto para el AFC a nivel estudiante y a nivel colegio.

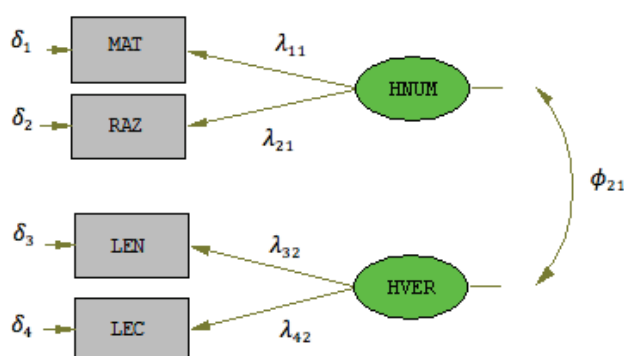


Figura 4.47: Análisis Factorial Confirmatorio a nivel estudiante-colegio.

El modelo planteado presenta cuatro variables observadas exógenas *MAT* y *RAZ* que explican su variabilidad en el factor común no observado (variable latente exógena) llamada *HNUM* que representa la habilidad numérica de los estudiantes. Las otras dos variables observadas exógenas son *LEN* y *LEC* que se encuentran explicados por el factor común (variable latente exógena) *HVER* que es la habilidad verbal. La ecuación del AFC es la siguiente:

$$x = \Lambda_x \xi + \delta$$

donde x son las variables observadas exógenas del modelo, ξ los factores comunes (variables latentes exógenas), Λ_x la matriz de coeficientes de variables exógenas (cargas factoriales) y δ son los errores de medición de las variables exógenas (factores específicos). En forma matricial se tiene:

$$\begin{pmatrix} MAT \\ RAZ \\ LEN \\ LEC \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ 0 & \lambda_{32} \\ 0 & \lambda_{42} \end{pmatrix} \begin{pmatrix} HNUM \\ HVER \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{pmatrix}$$

Ahora se procede a calcular los grados de libertad del modelo planteado:

$$\begin{aligned} g.l. &= \frac{1}{2}[(4 + 0)(4 + 0 + 1)] - 5 \\ &= 5 \end{aligned}$$

Como se explicó en el capítulo 2, los grados de libertad del modelo son mayores a cero, entonces el modelo es sobreidentificado, lo que permite una adecuada estimación de los parámetros.

Los resultados obtenidos mediante la estimación de Mínimos Cuadrados Generalizados se muestran a continuación, el AFC a nivel estudiante es:

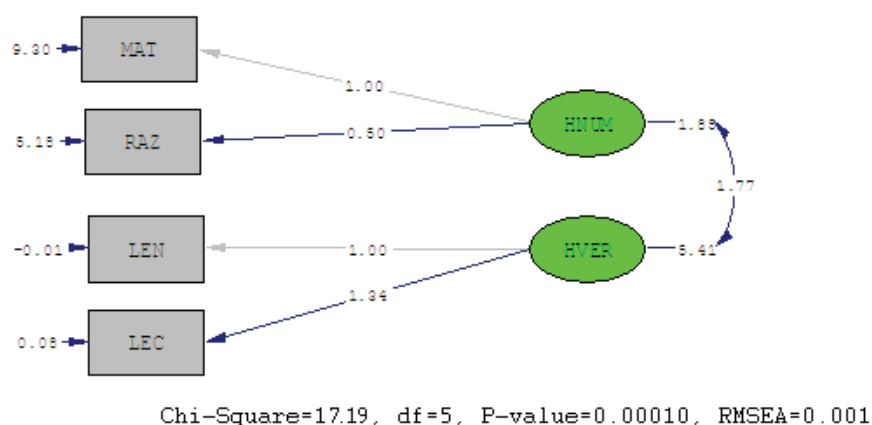


Figura 4.48: Resultados del AFC a nivel estudiante

Para el AFC a nivel colegio se tiene:

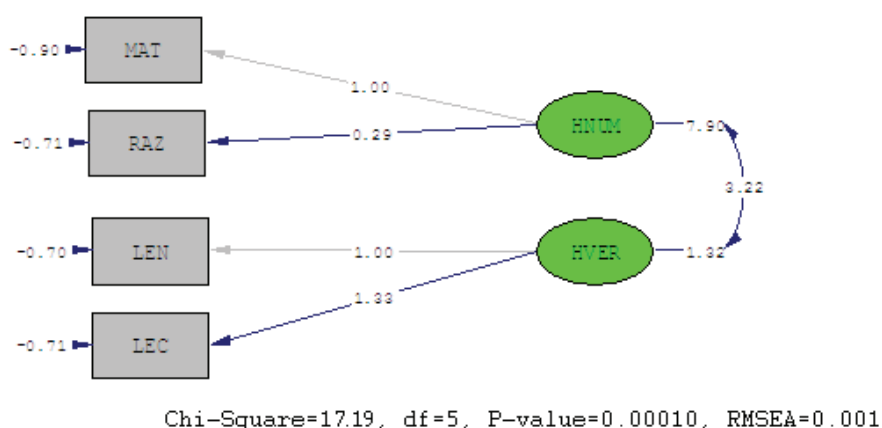


Figura 4.49: Resultados del AFC a nivel colegio

Todos los parámetros estimados resultaron significativos al nivel 5%. De los resultados obtenidos se aprecia que los errores de medición (factores específicos) de las variables observadas exógenas en el modelo a nivel estudiante son mayores que en el modelo a nivel colegio, esto se debe a que los estudiantes presentan mayor variabilidad en sus notas de manera general a diferencia del modelo a nivel colegio donde esta variabilidad se ve disminuida debido a que la medición en este nivel es de grupo y no existen mayores diferencias en las notas con los diferentes colegios de forma general.

Las cargas factoriales de las variables *MAT* y *LEC* respectivamente fueron las de mayor impacto para los factores comunes *HNUM* y *HVER* respectivamente

tanto en el modelo a nivel estudiante como en el modelo a nivel colegio.

Con respecto a los factores comunes podemos decir que la habilidad verbal a nivel estudiante es más fuerte que la habilidad numérica, a diferencia a nivel colegio donde la habilidad numérica es el factor más fuerte, debido a que presentan valores más altos.

El índice RMSEA de calidad de ajuste, el cual un valor cerca a cero indica un mejor ajuste y valores altos indica mal ajuste. Como el valor del $RMSEA \leq 0.05$ indica que el modelo propuesto tiene un buen ajuste de los datos. De igual forma el test estadístico χ^2 es significativo.

CAPÍTULO 5

Conclusiones y Recomendaciones

5.1 Conclusiones

- El análisis multinivel es una herramienta muy importante y poderosa al momento de identificar las relaciones dentro de los grupos y entre los grupos.
- Los modelos jerárquicos permiten obtener estimaciones estadísticamente eficientes de los coeficientes de regresión ya que la información de las agrupaciones permite corregir los errores estándar que son más conservadores que los modelos de regresión tradicionales que ignoran la agrupación.
- Los modelos de ecuaciones estructurales son una potente herramienta cuando las unidades de observación forman una jerarquía de grupos anidados y algunas variables de interés en el estudio deben ser medidas por una o conjunto de variables no observadas.
- La estimación mediante Cadenas de Markov Monte Carlo debido a que incorporan suposiciones de distribuciones sobre los parámetros a estimar, lo vuelve un método muy efectivo sobre otros métodos de estimación al momento de la obtención de resultados.
- La hipótesis planteada en el plan de titulación respecto a la obtención de mejores resultados en la estimación de los parámetros con el método de CMMC fue verdadera.
- En la aplicación de los modelos jerárquicos multinivel en la medición del rendimiento de los colegios en el cantón Otavalo, dio como resultados que los estudiantes de los colegios privados obtuvieron mejores notas que los colegios estatales y en especial el género femenino obtiene notas más altas que el género masculino.

- La estimación con CMMC-Gibbs presentó mejores resultados que con CMMC-MH, debido a que las estimaciones de los parámetros fueron significativas y el ajuste de los modelos en forma general fue mejor.
- El Análisis Factorial Confirmatorio (AFC) multinivel permite conocer los factores comunes que tienen mayor peso en los diferentes niveles.

5.2 Recomendaciones

- Debido al poder del análisis multinivel para identificar la variabilidad de los grupos, dentro y entre ellos, se recomienda el uso de esta herramienta para estudiar problemas sociales y llevar este mismo tipo de análisis a una aplicación a nivel nacional, debido a su eficiente resultado como herramienta de análisis.
- La eficiencia de los modelos de ecuaciones estructurales multinivel debido a que utiliza tres tipos de análisis como son: multinivel, análisis factorial y análisis de senderos, se recomienda su uso con la finalidad de describir las relaciones entre variables en una estructura jerárquica para posibles resoluciones de problemas de tipo social, científico, educacional, etc.
- Cuando se realiza un planteamiento de un problema se puede proponer varios modelos y al momento de realizar la especificación de los mismos y pruebas de hipótesis ver cuál de ellos es el que presenta el mejor ajuste.
- En la elaboración de modelos multinivel enfocados a medir el rendimiento de instituciones educativas, la inclusión de variables como: perfil del educador, infraestructura física, políticas de la institución, padres de familia, entre otras, serían muy adecuadas para mejorar los resultados en estos modelos.
- Con la implementación por parte del Ministerio de Educación de aplicar pruebas homogéneas a nivel nacional en las diferentes instituciones educativas, la elaboración de un modelo multinivel sería la mejor herramienta que se pudiera aplicar para poder medir el rendimiento de los estudiantes en las diferentes regiones del Ecuador tanto en las instituciones públicas como en las instituciones privadas.

Referencias

- [1] J. Aldas-Manzano. Análisis factorial confirmatorio.
- [2] J. Andrade. Modelización multinivel. *XII Encuentro de Matemática y sus aplicaciones*, 2010.
- [3] W. Browne. Mcmc estimation in mlwin. 2009.
- [4] A. Castro. *Regresión Lineal*. Quito, 2007.
- [5] M. Coba. Modelización de ecuaciones estructurales. Master's thesis, Escuela Politécnica Nacional, Quito, 2006.
- [6] D. Draper. Bayesian modeling, inference and prediction. *Department of Applied Mathematics and Statistics University of California*, 2005.
- [7] S. du Toit y otros. Lisrel for windows: Prelis user's guide. *Scientific Software International, Inc.*
- [8] W. Gilks. *Markov Chain Monte Carlo in Practice*. CHAPMAN, New York, U.S.A., 1st edition, 1996.
- [9] H. Goldstein. *Multilevel Statistical Models*. WILEY, United Kingdom, 4th edition, 2011.
- [10] C. Hidalgo. Estimación bayesiana de ecuaciones estructurales. Master's thesis, Escuela Politécnica Nacional, Quito, 2012.
- [11] J. Hox. *Multilevel Analysis: Techniques and Applications*. LAWRENCE ERLBAUM ASSOCIATES, London, England, 2002.
- [12] R. Johnson. *Applied Multivariate Statistical Analysis*. Prentice Hall, New York, 2001.
- [13] D. Kaplan. *Structural Equation Modeling: Foundations and Extensions*. SAGE, California, 2000.

- [14] R. Kline. *Principles and Practice of Structural Equation Modeling*. The Guilford Press, New York, 2005.
- [15] T. Luque. *Técnicas de Análisis de Datos en Investigación de Mercados*. Ediciones Pirámide, Madrid, 2000.
- [16] R. Sierra. *Ciencias Sociales: Análisis Estadístico y Modelos Matemáticos*. Paraninfo S.A., Madrid, 1981.
- [17] J. Rasbash y otros. *A user's guide to mlwin*. 2004.
- [18] J. Rasbash y otros. *Manual supplement mlwin*. 2012.
- [19] S. Searle y otros. *Variance Components*. Jhon Wiley, New Jersey, 2006.
- [20] T. Snijders y R. Bosker. *Multilevel Analysis*. SAGE, London, England, 4th edition, 2003.

Anexos

ANEXO A

Demostración que la varianza observada entre grupos es insesgada

Por demostrar:

$$E(S_{entre}^2) = \sigma_{u0}^2 + \frac{\sigma_{e0}^2}{w} \approx \sigma_{u0}^2, \quad \text{cuando } w \text{ es grande}$$

La varianza observada entre grupos de diferentes tamaños es:

$$\begin{aligned} S_{entre}^2 &= \frac{1}{w(N-1)} \sum_{j=1}^N n_j (\bar{y}_{.j} - \bar{y}_{..})^2 \\ &= \frac{1}{w(N-1)} \sum_{j=1}^N n_j (\bar{y}_{.j}^2 - 2\bar{y}_{.j}\bar{y}_{..} + \bar{y}_{..}^2) \\ &= \frac{1}{w(N-1)} \left(\sum_{j=1}^N n_j \bar{y}_{.j}^2 - 2\bar{y}_{..} \sum_{j=1}^N \bar{y}_{.j} n_j + \bar{y}_{..}^2 M \right) \\ &= \frac{1}{w(N-1)} \left(\sum_{j=1}^N n_j \bar{y}_{.j}^2 - 2\bar{y}_{..} (M\bar{y}_{..}) + \bar{y}_{..}^2 M \right) \\ &= \frac{1}{w(N-1)} \left(\sum_{j=1}^N n_j \bar{y}_{.j}^2 - 2M\bar{y}_{..}^2 + M\bar{y}_{..}^2 \right) \\ &= \frac{1}{w(N-1)} \left(\sum_{j=1}^N n_j \bar{y}_{.j}^2 - M\bar{y}_{..}^2 \right) \end{aligned}$$

Las hipótesis planteadas del modelo multinivel de nivel 2 (capítulo 1) son: $E(u_{0j}) = 0$, $E(u_{0j}^2) = \sigma_{u0}^2$ y $E(u_{0j}, u_{0j'}) = 0$ para $j \neq j'$; y $E(e_{0ij}) = 0$, $E(e_{0ij}^2) = \sigma_{e0}^2$ y

$E(e_{0ij}, e_{0i'j'}) = 0$ para $i \neq i'$ y $j \neq j'$, y $E(u_{0j}, e_{0i'j'}) = 0$. Con lo que tenemos:

$$\begin{aligned}
E(M\bar{y}_{..}^2) &= ME \left(\beta_0 + \frac{\sum_{j=1}^N n_j u_{0j}}{M} + \bar{e}_{..} \right)^2 \\
&= ME \left(\beta_0^2 + \frac{\sum_{j=1}^N n_j^2 u_{0j}^2 + \sum_{j \neq j'}^N \sum_{j' \neq j}^N n_j n_{j'} u_{0j} u_{0j'}}{M^2} \right) \\
&\quad + ME \left(\frac{\sum_{j=1}^N \sum_{i=1}^{n_j} e_{0ij}^2 + \sum_{j \neq j'}^N \sum_{j' \neq j}^N \sum_{i=1}^{n_j} \sum_{i'=1}^{n_{j'}} e_{0ij} e_{0i'j'}}{M^2} \right) \\
&\quad + ME \left(\frac{2\beta_0 \sum_{j=1}^N n_j u_{0j}}{M} \right) + ME(2\beta_0 \bar{e}_{..}) + ME \left(\frac{2 \sum_{j=1}^N n_j u_{0j} \bar{e}_{..}}{M} \right) \\
&= M\beta_0^2 + \frac{M \sum_{j=1}^N n_j^2 \sigma_{u0}^2}{M^2} + 0 + \frac{M \sum_{j=1}^N \sum_{i=1}^{n_j} \sigma_{e0}^2}{M^2} + 0 + 0 + 0 + 0 \\
&= M\beta_0^2 + \frac{\sigma_{u0}^2 \sum_{j=1}^N n_j^2}{M} + \sigma_{e0}^2 \tag{A.1}
\end{aligned}$$

de manera similar se tiene:

$$\begin{aligned}
E \left(\sum_{j=1}^N n_j \bar{y}_{.j}^2 \right) &= \sum_{j=1}^N n_j E(\beta_0 + u_{0j} + \bar{e}_{.j})^2 \\
&= M\beta_0^2 + M\sigma_{u0}^2 + N\sigma_{e0}^2 \tag{A.2}
\end{aligned}$$

restando (A.1) de (A.2) y multiplicando por el término $\frac{1}{w(N-1)}$ se tiene:

$$\begin{aligned}
E(S_{entre}^2) &= \frac{1}{w(N-1)} \left[\sigma_{u0}^2 \left(M - \frac{\sum_{j=1}^N n_j^2}{M} \right) + \sigma_{e0}^2 (N-1) \right] \\
&= \frac{1}{w} \left[\frac{\sigma_{u0}^2}{(N-1)} \left(M - \frac{\sum_{j=1}^N n_j^2}{M} \right) + \sigma_{e0}^2 \right]
\end{aligned}$$

es así que $w = \frac{1}{N-1} \left(M - \frac{\sum_{j=1}^N n_j^2}{M} \right)$, lo que resulta:

$$E(S_{entre}^2) = \sigma_{u0}^2 + \frac{\sigma_{e0}^2}{w} \approx \sigma_{u0}^2, \quad \text{cuando } w \text{ es grande}$$

ANEXO B

Certificación de la Dirección Provincial de Educación de Imbabura



DIRECCION DE EDUCACION HISPANA DE IMBABURA

0001985

Señor
Mario Fernando Coba Cisneros
Ciudad.-

3 MAYO 2012

De mi consideración:

En referencia al oficio sin número del 30-04-2012, comunico a usted que esta dirección autoriza el ingreso a los establecimientos educativos del nivel medio del cantón Otavalo, con la finalidad de que aplique test de Matemática, Lenguaje y Razonamiento a los estudiantes; en virtud de que dichos test se encuentran validados por la Facultad de Ciencias de la Escuela Politécnica Nacional previa el otorgamiento de la Ingeniería Matemática.

Particular que informo.

Atentamente,

Dr. Angel Castillo Rueda
DIRECTOR PROVINCIAL DE EDUCACIÓN DE IMBABURA

Jairo Terán/SE-EJ/SJ
Fernando Guerra Ref. 3774
2012-05-02



ANEXO C

Pruebas aplicadas

ESCUELA POLITÉCNICA NACIONAL FACULTAD DE CIENCIAS

Test de Lenguaje y Comunicación Décimo Año de Educación Básica

Establecimiento: Paralelo:
Género: Edad:

AGUDAS, GRAVES Y ESDRÚJULAS

1. Coloque la tilde en las siguientes palabras que la llevan.

Si me amas de verdad, demuéstramelo, así me dijo Monica cuando apareció ese miércoles por la tarde. Me quedé extático, eso, sin contar que por naturaleza soy tímido. Nada le conteste y enmudecí hasta el próximo sábado.

Vi un condor que volaba placidamente sobre los riscos de la cordillera. Yo también dijo Andrés y se subió sobre un peñasco desde donde divisó la verdura de los árboles dispersos en la llanura tropical.

Unas chicas simpáticas vienen con Verónica, Julian, Micaela y Juan Sebastián; seguramente irán a dar un paseo con Maripaz, que su contemporánea. Cristian, que es padre de las gráciles jóvenes, pidió a Damian que las acompañara.

cancion	Alcala	oceanio	rodapie	Juan
devoto	domingo	oseo	estomago	humildad
devocion	Bolivia	linea	brujula	institutriz
Bolivar	Bogota	lineal	visceversa	jamás
sicologo	biologo	horoscopo	visera	Francisco
sensacional	dificil	frejol	requiem	dubitacion
dificiles	placidas	doctor	arboles	pobreza
Pacifico	oceanico	imagen	agradecelo	olvidalo
tropico	seis	dieciseis	vigesimo	Mediterraneo

2. En cada espacio escriba la letra que falta. Ayúdense para ello de los significados o términos sinónimos de la columna intermedia. Con la palabra obtenida, complete las oraciones de la columna de la derecha.

ába__	Aparato para calcular.	Hice una suma con el _____.
a__a__a	Iglesia, monasterio.	El abad está en la _____.
a__arro__ar	Llenar por completo.	No debemos _____ la bodega.
a__ism__	Profundidad grande.	Cayó en el _____.
a__o__ornar	Avergonzar.	No debes _____ a tus padres.
a__omi__ar	Aborrecer o condenar.	Tienes que _____ los vicios.
a__ori__en	Nativo, indígena.	Descubrimos un _____ en la isla.
a__ov__d__r	Cubrir con una bóveda.	El ingeniero ordenó _____ la cúpula.
a__sc__s__	Acumulación de pus.	Se le ha formado un _____ en el brazo.
a__sci__a	Coordenada horizontal.	La línea roja es la _____ en el gráfico.
au__enti__o	Falta o ausencia.	Debemos evitar el _____ de los obreros.
a__ste__i__	Que no beba alcohol.	Ricardo es un _____, no ingiere licor.
a__alan__a	Alud.	Se produjo una _____ en la montaña.
a__asa__ar	Atropellar, arrollar.	No te dejes _____ por nadie.
a__enta__ar	Sacar ventaja a otros	Este competidor va a _____ a todos.
a__ergon__ar	Producir vergüenza.	No te debes _____ de tus actos.
a__erí__	Daño o rotura.	El coche sufrió una _____.
a__eri__uar	Indagar, preguntar.	Ve a _____ su dirección.
a__er__ión	Antipatía.	Siento _____ por ese individuo.
a__icul__ura	Cría de aves domésticas.	El ingeniero se dedica a la _____.
á__i__o	Ambicioso, hambriento.	Estoy _____ de noticias buenas.
a__i__ar	Poner más vivo, animar.	Hay que _____ el fuego.

ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE CIENCIAS

Test de Lenguaje y Comunicación
Tercer Año de Bachillerato

Establecimiento: Paralelo:
Género: Edad:

SINÓNIMOS

1. **Escoja la palabra que tiene significado semejante de la escrita con mayúscula. Subraye la palabra escogida.**

VERTER	PROPONER	IRRADIAR	PREVISOR
a) llenar	a) producir	a) solear	a) confiado
b) coger	b) tender	b) brillante	b) premeditado
c) observar	c) proclamar	c) energizar	c) asediado
d) vaciar	d) reducir	d) apagar	d) precavido
e) coser	e) plantear	e) centellar	e) astuto

2. **Señale el sinónimo que puede sustituir a la palabra en negrita y subrayada, según el contexto de la oración. Subraye la palabra escogida.**

- Los hombres inteligentes no abandonan sus **ideales**, luchan por ellos.
a) perfecto b) ejemplo c) sueño d) magnífico e) modelo
- Pese al dolor, el herido no pronunció ningún **lamento**.
a) alborozo b) latente c) queja d) gemido e) ayayay
- Al recibir la noticia, tu hermana sintió una gran **aflicción**.
a) tristeza b) arrepentimiento c) alegría d) afluencia e) era
- El cazador **sentía** acercarse el momento.
a) advertía b) percibía c) lamentaba d) presentía e) comprendía

ANTÓNIMOS

3. **Escoja la palabra que es antónimo de escrita con mayúscula. Encierre la palabra escogida.**

DELIBERADO	OBSOLETO	INEFABLE	ESCARNIO
a) volitivo	a) lánguido	a) sustentable	a) ofensa
b) inusual	b) apropiado	b) manifiesto	b) fauna
c) imprevisto	c) sosegado	c) decible	c) alabanza
d) involuntario	d) taimado	d) inestable	d) afrenta
e) ocasional	e) arcaico	e) descifrable	e) catilinaria

4. **Señale el antónimo de las palabras escritas en negrita y subrayada. Encierre la palabra elegida.**

- En los países en **desarrollo** es alto el índice de desnutrición infantil.
a) retroceso b) atraso c) empobrecimiento d) limitación e) subdesarrollo

- La desnutrición infantil origina seres **tarados**, resentidos, inadaptados...
a) normal b) formado c) limpio d) proporcionado e) correcto
- El tercer mundo carece de una **política** concorde al crecimiento poblacional.
a) ineptitud b) incapacidad c) inhabilidad d) torpeza e) incorrección
- La desnutrición infantil **genera** en todo tipo de defectos y malformaciones.
a) regenera b) reforma c) pureza d) virtud e) mejora

TERMINO EXCLUIDO

5. **Identifique la palabra que no comparte el núcleo de significado de la palabra guía. Tache la palabra fuera de contexto.**

DESARRAPADO	OBJETAR	INDAGACIÓN	AÑEJO
a) harapiento	a) impugnar	a) investigación	a) añoso
b) humilde	b) debatir	b) búsqueda	b) viejo
c) andrajoso	c) contradecir	c) cuestionamiento	c) averiado
d) pingajoso	d) refutar	d) averiguación	d) vetusto
e) haraposos	e) rebatir	e) exploración	e) antiguo

FRASES INCOMPLETAS

6. **Elija la palabra o frase que al insertar en la oración complete mejor su significado. En el espacio ponga el literal elegido.**

- Una gran artista no debe ser juzgada _____, sino por su creación total.
a) solo por su música b) solo por un poema c) solo por uno de sus cuadros
d) por una sola de sus obras e) por una única escultura
- Aunque las calificaciones de Jorge eran _____, sus padres estaban _____.
a) excelentes – felices b) malas – furiosos c) deficientes – descontentos
d) aceptables – molestos e) muy buenas – interesados
- Los animales pueden _____ de muchas formas los problemas _____ por los cambios estacionales.
a) soportar – comunitarios b) afrontar – causados c) rechazar – proporcionados
d) esquivar – esperador e) someter – propiciados
- _____ son los que acuden en lo bueno al ser llamados y en las _____ vienen solos.
a) Amigos – adversidades b) Rateros – cantinas c) Humanos – desgracias
d) Niños – Navidades e) Jóvenes – vejez

ANALOGÍAS

7. **Seleccione el par de palabras que mejor expresa la relación manifestada por el par original. Encierre el literal correcto.**

QUEMAR: FUEGO	CUADRADO: CUBO	MADERA: CARPINTERO
A viajar: agua	A círculo: cilindro	A fusil: soldado
B mojar: agua	B círculo: elipse	B hierro: herrero
C esparcir: agua	C círculo: óvalo	C bisturí: cirujano
D mitigar: agua	D círculo: parábola	D andamio: albañil
E reunir: agua	E círculo: circunferencia	E arcilla: alfarero
TESTIGO: VERAZ	AVIÓN: PASAJE	CASA: ESCALERA
A árbitro: neutral	A pantalones: correa	A calle: acera
B condenado: culpable	B rifa: boleto	B virtud: gloria

- | | | |
|--------------------|-------------------|----------------------|
| C hombre: sabio | C baile: pareja | C monte: camino |
| D discusión: lucha | D caminar: calles | D edificio: ascensor |
| E baile: pareja | E hombre: sabio | E punta: eje |

- | | | |
|---------------------|-------------------------|------------|
| BRILLO: BRILLANTE | ALIMENTO: HAMBRE | 369: 123 |
| A color: rojo | A comida: sed | A 386: 122 |
| B amarillo: rojo | B calor: saco | B 896: 232 |
| C ventana: azul | C gusano: tierra | C 927: 373 |
| D luz: fuego | D vitamina: vitaminosis | D 639: 213 |
| E feliz: embelesado | E feliz: embelesado | E 369: 133 |

COMPREENSIÓN LECTORA

8. Lea con atención los siguientes fragmentos de lecturas y escoja la única respuesta correcta de las cinco afirmaciones que siguen a cada pregunta. Tache completamente el literal correspondiente.

LECTURA 1.

Magia adormecedora vierte el río
en la calma monótona del viaje
cuando borra lo lejos del paisaje
la sombra que se extiende en el vacío.
Oculta en sus negruras al bohío
la maraña tupida, y el follaje
semeja los caldos de un encaje,
al caer el crepúsculo sombrío.

1. La idea principal del texto es
 - A. Hay muchas sombras en el camino
 - B. Los bohíos negros están ocultos en el camino.
 - C. Cae el crepúsculo como un encaje.
 - D. La calma hace aburrido el viaje y las sombras no dejan ver el paisaje.
 - E. Cuando anochece hay muchas sombras.
2. La forma en que está presentada esta lectura es:
 - A. En prosa
 - B. En prosa y en verso
 - C. En verso
 - D. En forma dialogada
 - E. Es una síntesis
3. Se afirma que en la primavera:
 - A. El viaje es muy placentero.
 - B. El viaje es borrascoso.
 - C. El viaje es aburrido.
 - D. El viaje es peligroso.
 - E. El viaje es muy instructivo.

LECTURA 2.

Una de las mayores estatuas que se ha construido hasta hoy se realizó en Francia. Después fue embarcada y cruzó el Atlántico. Se la conoce popularmente con el nombre de "Estatua de la Libertad", pero su verdadero nombre es "La libertad iluminando al mundo". Se encuentra en la isla de la Libertad, en la bahía de New York. Es la figura de una mujer vestida con larga túnica, que sostiene, con su brazo derecho levantado, una antorcha. Representa la libertad. El autor fue Frederic Auguste Bartholdi, y para fabricarla se moldearon grandes planchas de cobre que se montaron y soldaron después sobre una armadura de hierro.

En Paris, Francia hay una reproducción de la estatua de la Libertad en un puente sobre el río Sena.

4. El título de la lectura es:

- A. Un paseo por la isla de la Libertad.
- B. La fabricación de planchas de cobre.
- C. Recorrido sobre el río Sena.
- D. La Estatua de la Libertad
- E. Historia de las estatuas de Francia.

5. De acuerdo al contenido del texto, es verdad que:

- A. La isla de la Libertad se encuentra en Francia.
- B. La estatua fue construida en New York.
- C. La estatua de la Libertad está ubicada en una Bahía de New York.
- D. La estatua representa la lealtad.
- E. Sostiene con su brazo derecho una espada.

6. También podemos afirmar que:

- A. La estatua fue un regalo hecho a Paris.
- B. La estatua de la Libertad es una de las más grandes que se ha construido hasta hoy.
- C. La estatua fue construida sobre una base de cobre.
- D. La estatua de la Libertad está ubicada en el río Sena.
- E. Esta estatua fue transportada por aire.

7. Además se puede afirmar que:

- A. Es la estatua más antigua del mundo.
- B. Su autor viajó a América para construir la estatua.
- C. El nombre con el que se la conoce popularmente a la estatua no es el verdadero.
- D. La estatua se puede ver desde cualquier parte de New York.
- E. También se denomina "Paso a la libertad".

8. El texto desarrolla como tema central:

- A. Una alegoría a la libertad
- B. Una historia ficticia
- C. Moraleja a la humanidad
- D. Un informe cultural
- E. Una narración científica

LECTURA 3

La fotografía solo comenzó a existir realmente hace unos 140 años. Dos cosas son necesarias: la producción de una imagen y un método para fijarla, haciéndola permanente. Ya en el siglo IV a.C. se sabía que los rayos de luz del exterior pueden penetrar por un orificio en un cuarto oscuro y proyectar una imagen sobre una de sus paredes ennegrecidas. Este método de producir imágenes se llama cámara oscura. Se utilizó por primera vez en Arabia durante el

siglo XI, para la observación de los eclipses. En 1970 Tomas Wedgewood (inglés), intentó hacer por primera vez una fotografía, con algún éxito.

9. El título de la lectura es:

- A. La cámara oscura
- B. los eclipses
- C. Las imágenes ennegrecidas
- D. La historia de la fotografía
- E. Los rayos de luz

10. De acuerdo con el contenido del texto se afirma que:

- A. Para la producción de una fotografía se necesitan varias imágenes.
- B. El método de producir imágenes se llama cámara oscura.
- C. la observación de los eclipses se hizo por primera vez en Arabia.
- D. El inglés Tomas Wedgewood fracasó en su intento de realizar una fotografía.
- E. La fotografía comenzó a existir hace 40 años.

11. Según el trozo podemos deducir que:

- A. Los rayos de luz al atravesar un orificio proyectan imágenes.
- B. La producción de imágenes se llaman eclipses.
- C. La fotografía comenzó en el siglo XIII.
- D. La cámara oscura no produce imágenes.
- E. Los rayos de luz proyectan imágenes oscuras.

12. El texto presenta como tema central:

- A. La descripción de un paisaje
- B. Un relato social.
- C. Un suceso narrativo
- D. Una composición verbal
- E. Un informe técnico

LECTURA 4

La penicilina es un antibiótico, una sustancia producida por un ser vivo, que puede acabar con bacterias y otros gérmenes. La palabra antibiótico significa "contra la vida". El Dr. Fleming había descubierto la penicilina en un moho, que es una forma primitiva de vida vegetal, al igual que las bacterias. Los científicos empezaron inmediatamente el estudio de los mohos y de las bacterias en busca de otras sustancias matamicrobios. Se encontraron antibióticos nuevos y potentes donde menos se esperaba. La estreptomocina se encontró en organismos llamados actinomocitos. La terramicina se obtuvo de un moho del suelo de Indiana. La cloromicetina, en una muestra de tierra de Caracas, Venezuela. Para encontrar estos antibióticos se

analizaron miles de plantas, y se estudiaron y seleccionaron decenas de miles de muestras de tierra.

13. El título del texto es:

- A. La terramicina
- B. Los antibióticos
- C. La penicilina
- D. Las bacterias
- E. Las plantas

14. De acuerdo al contenido del texto se afirma que:

- A. Las bacterias fueron examinadas para encontrar los antibióticos.
- B. La palabra antibiótico significa contra la vida.
- C. La penicilina no es una sustancia.
- D. No existen los antibióticos.
- E. El moho es un parásito

15. La forma de presentar este texto es:

- A. En versos de arte mayor
- B. En versos exasílabos
- C. En prosa
- D. En forma dialogada
- E. En versos de arte menor

16. Según el texto se afirma que:

- A. La estreptomocina, la terramicina y la cloromicetina son antibióticos.
- B. Los antibióticos son formas de vida primitiva.
- C. Las bacterias son una clase de mohos.
- D. la penicilina es producida por una planta.
- E. La terramicina se obtuvo en Inglaterra.

17. Podemos deducir que los antibióticos:

- A. No son potentes.
- B. Acaban con las bacterias.
- C. Producen enfermedades.
- D. Fueron descubiertos en Alemania.
- E. Son bacilos alargados.

18. El moho es:

- A. Un anticuerpo
- B. Una infección
- C. Una forma de vida vegetal
- D. Una epidemia
- E. Un virus

19. El texto presenta como tema central:

- A. La vida primitiva
- B. Los seres vivos
- C. La existencia de las plantas
- D. Una narración científica
- E. La historia de la tierra

ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE CIENCIAS

Test de Matemática
Décimo Año de Educación Básica

Establecimiento: Paralelo:

Género: Edad:

1. Responder con verdadero (V) o falso (F).

- a) $\sqrt[6]{2^4} = \sqrt[8]{2^2}$
- b) $\sqrt[3]{2^6} = 2\sqrt[4]{2^2}$
- c) $\sqrt{3} \cdot \sqrt[3]{3} = \sqrt[6]{3^3} \cdot \sqrt[6]{9}$
- d) $\frac{1}{2}\sqrt[4]{5} \cdot \frac{1}{3}\sqrt[3]{5} = \frac{1}{2}^{12}\sqrt[12]{125} \cdot \frac{1}{3}^{12}\sqrt[12]{5^4}$
- e) $5\sqrt[3]{250} = 25\sqrt[3]{2}$

2. Resolver la siguiente ecuación y encontrar el valor de x.

$$\frac{x+2}{x^2+2x-3} - \frac{x+1}{x^2-9} = \frac{1}{x^2-4x+3}$$

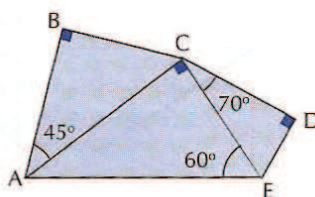
- a) x=4 b) x=-4 c) x=3 d) x=5

3. Encontrar la solución de x, que satisface la siguiente desigualdad.

$$\frac{4x-5}{8} + \frac{1-3x}{2} > \frac{x}{2} - \frac{7}{4}$$

- a) $x > \frac{12}{13}$ b) $x > \frac{3}{8}$ c) $x < \frac{2}{5}$ d) $x < \frac{13}{12}$

4. En la siguiente figura, $\overline{BC} = 10\sqrt{2}$. Calcular \overline{CE} .

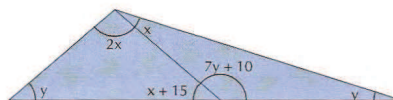


a) $\overline{CE} = 5\sqrt{2}$ b) $\overline{CE} = \sqrt{3}$ c) $\overline{CE} = \frac{20}{\sqrt{3}}$ d) $\overline{CE} = \frac{\sqrt{2}}{2}$

5. Calcular la altura de un cono cuyo radio mide 12 cm. y su volumen es 4500 cm^3 .

a) $h = \frac{375}{4\pi} \text{ cm}$ b) $h = \frac{125}{2\pi} \text{ cm}$ c) $h = \frac{215}{\pi} \text{ cm}$ d) $h = \frac{375}{\pi} \text{ cm}$

6. Calcular el valor de x y y .



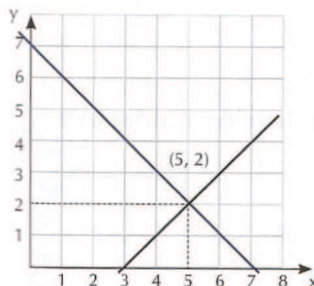
a) $x = 50^\circ$
 $y = 15^\circ$ b) $x = 40^\circ$
 $y = 25^\circ$ c) $x = 65^\circ$
 $y = 10^\circ$ d) $x = 30^\circ$
 $y = 15^\circ$

7. Resolver el siguiente sistema de dos ecuaciones con dos incógnitas.

$$\begin{cases} 2ax - 3by = 2 \\ 5ax + 4by = 5 \end{cases}$$

a) $x = \frac{6}{a}, y = b$ b) $x = \frac{1}{2a}, y = 0$ c) $x = \frac{2}{a}, y = \frac{6}{2b}$ d) $x = \frac{1}{a}, y = 0$

8. ¿De qué sistema es solución el gráfico?



a) $x+2y=1$
 $2x+3y=3$ b) $2x-y=7$
 $x-2y=3$ c) $2x+3y=5$
 $x-y=2$ d) $x+y=7$
 $x-y=3$

9. Determinar que función es exponencial. Cuando $x = \{1, 2, 3, 4, 5\}$

a) $f(x) = \left(\frac{1}{2}\right)^x$ b) $f(x) = (-3)^x$ c) $f(x) = \left(-\frac{1}{3}\right)^x$ d) $f(x) = (-2)^{3x}$

10. Se extrae una carta de una baraja de 52 cartas. Calcular la probabilidad de que sea roja.

a) $\frac{1}{2}$ b) $\frac{1}{4}$ c) 1 d) $\frac{3}{4}$

ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE CIENCIAS

Test de Matemática

Tercer Año de Bachillerato

Establecimiento: Paralelo:

Género: Edad:

1. Calcular el perímetro del triángulo cuyos vértices son: A (0,4), B (-4,1), C (3,-3).

a) $P_{\Delta ABC} = 20,67$ b) $P_{\Delta ABC} = 35,17$ c) $P_{\Delta ABC} = 10,14$ d) ninguna

2. Hallar la ecuación de la circunferencia de centro en el punto (-4,3) y tangente al eje y.

a) $x^2 + y^2 - 6x + 9y - 4 = 0$ b) $x^2 + y^2 + 12x + 5y - 7 = 0$

c) $x^2 + y^2 + 8x - 6y + 9 = 0$ d) ninguna

3. Calcular el dominio de la función $f(x)$.

$$f(x) = \sqrt{4x^2 - 1}$$

a) $Dom(f) =]-\infty, -\frac{1}{2}] \cup [\frac{1}{2}, +\infty[$ b) $Dom(f) = [\frac{1}{2}, +\infty[$

c) $Dom(f) = [\frac{1}{4}, +\infty[$ d) ninguna

4. Calcular el dominio de la función $g(x)$.

$$g(x) = \frac{x+2}{x^2-9}$$

a) $Dom(g) = [3, \infty[$

b) $Dom(g) =]-\infty, -3] \cup [3, \infty[$

c) $Dom(g) = \{\mathbb{R} - \{3, -3\}\}$

d) ninguna

5. Calcular el siguiente límite.

$$\lim_{x \rightarrow \infty} \frac{5 - 2x^2}{3x + 5x^2}$$

a) $-\frac{2}{3}$

b) $-\frac{2}{5}$

c) $\frac{2}{5}$

d) ninguna

6. Calcular el siguiente límite.

$$\lim_{x \rightarrow \frac{\pi}{2}} \frac{\cos x}{1 + 2 \sin x}$$

- a) $\frac{1}{2}$ b) 1 c) 0 d) ninguna

7. Calcular la derivada con respecto a x de la siguiente expresión:

$$y = \frac{2 - x}{1 + 2x^2}$$

- a) $\frac{4x^2 - 9x + 2}{(1 + 2x^2)^2}$ b) $\frac{6x^2 + 12x + 1}{(1 + 2x^2)^2}$
 c) $\frac{2x^2 - 8x - 1}{1 + 4x^2 + 4x^4}$ d) ninguna

8. Calcular la derivada con respecto a x de la siguiente expresión:

$$y = \sin 2x + \ln(x^2 + 1) + e^{2x} + 3^x$$

- a) $2 \sin 2x + \frac{2x}{x^2 - 1} + e^{2x}$ b) $2 \cos 2x + \frac{2x}{x^2 + 1} + 2e^{2x} + 3^x \ln 3$
 c) $2 \cos 2x + \frac{x}{x^2 + 1} + 2e^{2x} + 3^x$ d) ninguna

9. Calcular la siguiente integral.

$$\int_0^1 (x + 2)^3 dx$$

- a) $\frac{65}{4}$ b) $\frac{45}{3}$ c) $\frac{67}{2}$ d) ninguna

10. Calcular la siguiente integral.

$$\int_0^\pi (\sin x + 7 \cos x - 1) dx$$

- a) $2 - \pi$ b) $\pi - 1$ c) $2 - \pi$ d) ninguna

ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE CIENCIAS

Test de Razonamiento
Décimo Año de Educación Básica

Establecimiento: Paralelo:

Género: Edad:

En los siguientes ejercicios escoger el cuadro correcto entre las opciones a, b y c para completar cada una de las series.

1.

1	3	7	13	???
---	---	---	----	-----

 a:

20

 b:

21

 c:

23

2.

99	1	98	2	97	???
----	---	----	---	----	-----

 a:

96

 b:

3

 c:

0

3.

					???
--	--	--	--	--	-----

 a:

--

 b:

--

 c:

--

4.

				???
--	--	--	--	-----

 a:

--

 b:

--

 c:

--

5.

				???
--	--	--	--	-----

 a:

--

 b:

--

 c:

--

6.

	???

 a:

--

 b:

--

 c:

--

7.

A	D	G	J	???
---	---	---	---	-----

 a:

K

 b:

L

 c:

M

8.

A1	A2	B1	B2	???
----	----	----	----	-----

 a:

C1

 b:

C2

 c:

B3

9.

			???
--	--	--	-----

 a:

--

 b:

--

 c:

--

10.

		???
--	--	-----

 a:

--

 b:

--

 c:

--

ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE CIENCIAS

Test de Razonamiento
 Tercer Año de Bachillerato

Establecimiento: Paralelo:
 Género: Edad:

En los siguientes ejercicios escoger el cuadro correcto entre las opciones a, b y c para completar cada una de las series.

1.		a:	b:	c:
2.		a:	b:	c:
3.		a:	b:	c:
4.		a:	b:	c:
5.		a:	b:	c:
6.		a:	b:	c:
7.		a:	b:	c:
8.		a:	b:	c:
9.		a:	b:	c:
10.		a:	b:	c:

ANEXO D

Instituciones Educativas

INSTITUCIONES FISCALES		INSTITUCIONES PARTICULARES	
1	Instituto Tecnológico "Otavalo"	8	Colegio "San Luis"
2	Instituto Tecnológico "República del Ecuador"	9	Colegio "La Inmaculada"
3	Colegio Experimental "Jacinto Collahuazo"	10	U.E. "Santa Juana de Chantal"
4	Colegio Técnico Agropecuario "Carlos Ubidia Albuja"	11	Academia "General Carlos Machado"
5	Colegio Técnico Agropecuario "Fernando Chavez Reyes"	12	Unidad Educativa "School of Tomorrow"
6	Colegio Federico Páez		
7	Unidad del Milenio "Jatun Kuraka"		

ANEXO E

Modelos multinivel descartados

Modelo lineal jerárquico de nivel 2

Para llegar al modelo lineal jerárquico de nivel 2 planteado se comprobó que la variable *Mujer* no es significativa al nivel estudiante. A continuación se presentan los resultados del modelo:

Equations

Prueba_mat_y ~ N(XB, Ω)

Prueba_mat_y = β_{0y}Constante + β_{1y}Prueba_raz_y + β_{2y}Mujer_y + β_{3y}Privado_y

β_{0y} = 2.935(0.440) + u_{0y} + e_{0y}

β_{1y} = 0.139(0.073) + u_{1y}

β_{2y} = 0.414(0.229) + e_{2y}

β_{3y} = 1.230(0.626) + e_{3y}

$$\begin{bmatrix} u_{0y} \\ u_{1y} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.805(0.788) & -0.178(0.130) \\ -0.178(0.130) & 0.054(0.026) \end{bmatrix}$$

$$\begin{bmatrix} e_{0y} \\ e_{2y} \\ e_{3y} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 5.335(0.504) & & \\ 0.416(0.396) & 0.000(0.000) & \\ 0 & -1.593(1.068) & 4.262(1.709) \end{bmatrix}$$

-2*loglikelihood(IGLS Deviance) = 2885.129(607 of 607 cases in use)

Home + - Add Term Estimates Nonlinear Clear Notation Responses Store Help Zoom 100

Realizando la prueba de hipótesis del parámetro fijo $\hat{\beta}_2$ de la variable *Mujer*, considerando la variabilidad a nivel estudiante se tiene:

$$H_0 : \hat{\beta}_2 = 0$$

$$H_1 : \hat{\beta}_2 \neq 0$$

No se rechaza H_0 al nivel 5 %, para $\hat{\beta}_2$. Ahora, se presentan los resultados del modelo considerando la variabilidad a nivel colegio de la variable *Mujer*.

De manera similar a la prueba anterior planteada, se tiene que el parámetro $\hat{\beta}_2$ no es significativo al nivel 5%.

Para la variable *Privado* considerando la variabilidad a nivel colegio el modelo que se tiene es el siguiente:

La prueba de hipótesis del parámetro fijo $\hat{\beta}_3$ de la variable *Privado* es:

$$H_0 : \hat{\beta}_3 = 0$$

$$H_1 : \hat{\beta}_3 \neq 0$$

No se rechaza H_0 al nivel 5%, para $\hat{\beta}_3$.