



Serie de Cuadernillos Técnicos

Manual de procedimiento para el análisis de factores asociados al aprendizaje



Oscar Hugo López Rivas
Ministro de Educación

Héctor Canto Mejía
Viceministro Técnico de Educación

María Eugenia Barrios Robles de Mejía
Viceministra Administrativa de Educación

Daniel Domingo López
Viceministro de Educación Bilingüe e Intercultural

José Inocente Moreno Cámbara
Viceministro de Diseño y Verificación de la Calidad Educativa



Directora

Luisa Fernanda Müller Durán

Subdirección de Análisis de Datos

Autoría

Vivian Yvette Bolaños Gramajo
Ana Aidé Cruz Grünebaum
José Adolfo Santos Solares

Revisión de texto y diagramación

María Teresa Marroquín Yurrita

Diseño de portada

Eduardo Avila

Dirección General de Evaluación e Investigación Educativa

© Digeduca 2017 todos los derechos reservados.

Se permite la reproducción de este documento total o parcial, siempre que no se alteren los contenidos ni los créditos de autoría y edición.

Para efectos de auditoría, este material está sujeto a caducidad.

Para citarlo: Bolaños, V.; Cruz, A. & Santos, J. (2017). *Manual de procedimiento para el análisis de Factores Asociados al aprendizaje*. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación.

Disponible en red: <http://www.mineduc.gob.gt/digeduca>

Impreso en Guatemala

divulgacion_digeduca@mineduc.gob.gt

Guatemala, 2017

Contenido

Introducción	6
Propósito	6
1. Área de aplicación y alcance del procedimiento de análisis de Factores Asociados.....	7
2. Responsables del proceso desde las bases de datos hasta el informe de análisis de Factores Asociados al aprendizaje	7
3. Políticas procedimentales	7
4. Conceptos.....	9
4.1. Factores Asociados al aprendizaje	9
4.2. Habilidad latente.....	9
4.3. Modelo Jerárquico Lineal Multinivel	9
4.4. Teoría de Respuesta al Ítem (TRI) o Teoría de Rasgo Latente (TRL)	22
5. Procedimiento.....	23
5.1. Diseño del marco metodológico del análisis	23
5.2. Revisión del marco teórico.....	24
5.3. Procedimiento técnico previo	26
5.3.1. Procedimiento de preparación de bases de Nivel 1 y 2	27
5.4. Análisis y resultados del modelo	45
5.5. Elaboración del informe	55
6. Flujogramas	56
Glosario.....	63

Lista de tablas

Tabla 1. Marco teórico internacional básico de Factores Asociados	25
Tabla 2 Marco Teórico Nacional básico de Factores Asociados	26
Tabla 3. Sintaxis de recodificación para el análisis de variables de sector, jornada, área y sexo	30
Tabla 4. Variables del modelo Multinivel para el análisis de Factores Asociados de Graduandos 2011	32

Lista de figuras

Figura 1. Procedimiento para prueba de hipótesis	14
Figura 2. Interpretación de la prueba de hipótesis	15
Figura 3. Elegir efectos aleatorios en el modelo (aplica únicamente para investigaciones específicas)	17
Figura 4. Prueba de DEVIANZA que puede utilizarse para elegir efectos fijos o aleatorios de variables en el modelo.....	18
Figura 5. Resultados de la prueba de DEVIANZA	19
Figura 6. Verificación de supuestos	20
Figura 7. Ubicación de carpeta compartida	29
Figura 8. Ubicación de herramienta en SPSS para estandarizar variables de habilidad latente en Matemática y Lectura.....	33
Figura 9. Ubicación de herramienta en SPSS para transformar las variables de habilidad latente en Matemática y Lectura con media 500 y desviación estándar 100	34
Figura 10. Procedimiento opcional para reemplazar el signo “-” en el código de establecimiento.....	35
Figura 11. Ejemplo de análisis de patrones de datos perdidos	36
Figura 12. Ejemplo de análisis de patrones de datos perdidos en tablas de contingencia.....	37
Figura 13. Ejemplo de imputación de datos perdidos	39
Figura 14. Procedimiento en SPSS para obtener la proporción y los promedios de las variables observadas a nivel del establecimiento.....	42

Figura 15. Ejemplo de cuadro resumen del modelo de Factores Asociados al aprendizaje de Matemática de primaria 2010.....	53
Figura 16. Ejemplo de cuadro resumen del modelo de Factores Asociados al aprendizaje de Matemática de Graduandos 2011	54
Figura 17. Flujograma de equipo responsable	57
Figura 18. Flujograma de políticas procedimentales	58
Figura 19. Flujograma de procedimiento de preparación técnica para la base del Nivel 1	59
Figura 20. Flujograma de procedimiento de preparación técnica para la base del Nivel 2	60
Figura 21. Flujograma de procedimiento de análisis y resultados del modelo	61
Figura 22. Flujograma de procedimiento de elaboración del informe.....	62

Introducción



Este documento forma parte de los cuadernillos técnico-administrativos que el personal de la Subdirección de Análisis de Datos de Evaluación e Investigación de la Dirección General de Evaluación e Investigación Educativa (Digeduca), debe conocer y que son referente básico para ejecutar sus responsabilidades. Se complementa con los manuales del digitador, de limpieza de bases de datos, de calificación y de elaboración de Nivel Socioeconómico y otros indicadores, ya que como paso previo del proceso de análisis de Factores Asociados, las bases de datos con las que se trabajan deben cumplir con los lineamientos establecidos en estos manuales.

Se presenta como un manual en el que se explica su propósito, el área de aplicación y alcance de los procedimientos; los responsables del proceso de elaboración del informe de Factores Asociados al aprendizaje, las políticas procedimentales, conceptos básicos así como el procedimiento a detalle: los pasos específicos, responsables y productos según fase procedimental. Consigna los flujogramas del proceso, un glosario y referencias bibliográficas.

Propósito



Este documento de procedimientos es un instrumento diseñado para coadyuvar a la labor del analista de datos de las unidades de Análisis Cuantitativo y Cualitativo con relación a los procedimientos básicos que se realizan en el análisis de Factores Asociados al rendimiento educativo, que son aplicados en la elaboración del informe de Factores Asociados al aprendizaje de las cohortes de estudiantes evaluadas de primero, tercero y sexto grados del Nivel de Educación Primaria y del último año del Ciclo de Educación Diversificada (Graduandos). No se incluyó tercero básico debido a que en el momento de la elaboración del cuadernillo, no se contaba con un análisis de Factores Asociados de este grado que replicara el método y el procedimiento que se presenta en este material. En virtud que en cada nivel educativo se observan factores específicos asociados al aprendizaje de los estudiantes según el grado y nivel que cursan, este documento se concentra en los pasos básicos que el analista debe tomar en cuenta en el proceso de diseño y aplicación del modelo matemático para el análisis de Factores Asociados al aprendizaje del nivel que elabora.

1. Área de aplicación y alcance del procedimiento de análisis de Factores Asociados

Área temática: análisis estadístico e interpretación de resultados.

Fuente de datos:

- Bases de datos de Factores Asociados obtenidos en el cuestionario adjunto de la evaluación de estudiantes (bases de evaluación muestral de primero, tercero y sexto grados de Nivel de Educación Primaria, y las de las evaluaciones censales de tercero básico y graduandos).
- Bases de datos del cuestionario del director de establecimientos evaluados en los grados y niveles correspondientes.

Manuales de consulta:

- Manual del digitador
- Manual de limpieza de bases
- Manual de calificación
- Manual de elaboración del Nivel Socioeconómico y otros indicadores

2. Responsables del proceso desde las bases de datos hasta el informe de análisis de Factores Asociados al aprendizaje

- **Preparación previa:** Coordinación de digitación.
- **Elaboración del análisis:** analistas de la Subdirección de Análisis de Datos de Evaluación e Investigación.
- **Coordinación, verificación, seguimiento y control:** analistas de la Subdirección de Análisis de Datos de Evaluación e Investigación.
- **Supervisión y control:** Coordinación de Análisis Cuantitativo / Coordinación de Análisis Cualitativo.
- **Visto Bueno:** Subdirección de Análisis de Datos de Evaluación e Investigación.

3. Políticas procedimentales

- Se utilizarán como fuente de información las bases de datos de evaluaciones de estudiantes, que incluyen las respuestas de los estudiantes al cuestionario de Factores Asociados y de directores de establecimientos que participaron en la evaluación.
- Las bases a utilizar deben llenar los requisitos previos de calidad, integridad y correspondencia que se garantiza en la fase de digitación y verificación. Para mayor información sobre este proceso consulte el *Manual del digitador*.

- Las bases a utilizar también deben ser resultado de un proceso de verificación y limpieza de datos posterior a la fase de digitación. Para mayor información sobre este proceso consulte el *Manual de limpieza de bases de datos*.
- En virtud de la prioridad de requerimiento de información para la toma de decisiones del Ministerio de Educación y como paso básico, el análisis de Factores Asociados al aprendizaje tendrá lugar una vez que se cumpla con la generación de resultados de la evaluación con Teoría de Respuesta al Ítem (TRI) y de forma inmediata, posterior al procedimiento de calificación (consulte el *Manual de calificación de bases de datos*).
- El enfoque de análisis, variables de interés y formato de informe de resultados de cada grado y nivel / ciclo evaluado, debe ser discutido previamente con la Subdirección de Análisis de Datos.
- Con el aval de la Subdirección de Análisis de Datos, la Coordinación de área coordina la elaboración y tiempos de entrega de los resultados e informes con los analistas.
- La Coordinación dará seguimiento al proceso de análisis y resultados obtenidos, realimentará y reencauzará lo que se estime pertinente; asimismo consultará con la Subdirección de Análisis lo que se considere oportuno.
- Una vez que el analista finalice el análisis, entregará los resultados cuantitativos a la Coordinación que es la responsable de revisar y entregar a la Subdirección para su validación.
- El informe final será entregado por el analista a la Coordinación de área con copia a la Subdirección de Análisis de Datos, quienes revisarán y procederán con los siguientes pasos para finalizar el informe.

4. Conceptos

4.1. Factores Asociados al aprendizaje

Según la Real Academia Española, la palabra «factor» se deriva del latín *factor* que significa «elemento o causa» o bien, se entiende como «cada una de las cantidades o expresiones que se multiplican para formar un producto»; y la palabra «asociado» que significa «que acompaña a otra en alguna comisión o encargo» (RAE, 2001). Se definen como los factores que tienen un efecto o impacto en el aprendizaje de los estudiantes, de tal manera que la existencia o ausencia de este es capaz de predecir el Logro o No Logro de los estudiantes.

En Guatemala el estudio de Factores Asociados al aprendizaje de estudiantes de último año de Ciclo de Educación Diversificada seleccionó variables que se consideró aportan información valiosa para la toma de decisiones de nivel institucional. Por ello es importante que el uso de los Factores Asociados y las oportunidades de aprendizaje, esté enfocado a aquellos en los cuales el Ministerio de Educación tiene incidencia, es decir, en los escolares, pues son áreas en las que se pueden promover cambios (Flores, 2010, citado en *Informe de Factores Asociados de Graduandos 2011*).

4.2. Habilidad latente

Los resultados para cada estudiante son generados por un proceso de calificación utilizando la Teoría de Respuesta al Ítem (TRI) o Teoría de Rasgo Latente (TRL). En las evaluaciones nacionales de graduandos, el rasgo latente que se mide es la habilidad estimada del estudiante en Matemática y Lectura. Al hablar de estimación de la habilidad como rasgo latente se entiende como la medida obtenida por el estudiante que se encuentra expresada en una escala con unidades *Logit* representados por la letra griega *Theta* (Θ) (ver cuadernillo técnico de Análisis Rasch). Esto ha sido previamente estimado, validado y organizado en la prueba según el grado de dificultad de los ítems que el estudiante responde correctamente (Informe de Factores Asociados de Graduandos 2011).

4.3. Modelo Jerárquico Lineal Multinivel

Es un modelo que se elabora con una regresión lineal multinivel. Es de carácter múltiple que utiliza más de dos variables para comparación o control en diferentes niveles. Según Judd y Kenny (1981), la regresión múltiple es una técnica general para estimar los efectos de un modelo lineal general donde intervienen más de dos variables. Las relaciones pueden ser más complejas ya que pueden identificarse variables de los estudiantes que pueden variar según el grupo al que pertenezcan.

Por ejemplo, cuando se estudia un grado académico específico, como primero primaria en un establecimiento cualquiera, los estudiantes pueden ser diferentes (en cuanto a la edad, nivel socioeconómico, si son repitentes, por su idioma materno). Estudiantes del mismo grado de ese establecimiento podrían ser homogéneos o similares debido a que las características de todos tienden a determinado nivel o característica: casi todos puede ser que se encuentren en una edad promedio de siete años, de alto nivel socioeconómico, en general no repitentes y con idioma materno ladino.

Al comparar esta realidad con otro establecimiento y, estos estudiantes son en mayor grado heterogéneos o diferentes, debido a las características entre estudiantes: puede ser que casi todos tengan edad promedio de ocho años, de un nivel menor de nivel socioeconómico, una proporción de cuatro por cada 10 son repitentes, y con idioma materno ladino. Mientras en algún establecimiento los estudiantes podrían promediar un alto nivel socioeconómico y cultural, el clima escolar también podría ser mejor que otro, la proporción del sexo en el establecimiento también puede ser diferente (en algunos la proporción de sexo femenino puede ser mayor o menor), etcétera.

Las similitudes y diferencias también pueden deberse a la variación de características del establecimiento: los estudiantes del primer establecimiento estudian en el área urbana y los del segundo establecimiento en el área rural, estudian en diferentes jornadas, también varían en función del tiempo de instrucción de Matemática que se imparte por los docentes y la metodología que utilizan para ello, si los estudiantes tienen acceso a laboratorio de computación y disponibilidad de tiempo para su uso, si poseen diferentes directores que varían también en sexo, nivel educativo y experiencia que posean, entre otras variables.

Estas relaciones son complejas porque se entiende que los estudiantes pueden ser objeto de estudio no solo en función del grado y establecimiento al que pertenecen sino según su realidad en diferentes niveles. En Guatemala como país, es posible analizar el rendimiento de los establecimientos con las diferencias geográficas como el departamento donde se ubican, entre estos, según los establecimientos que lo componen y dentro de estos últimos, según grado o sección en la que se encuentra el estudiante. Pueden definirse tantos niveles jerárquicos como sean necesarios dependiendo del interés del investigador o propósito del estudio. En el caso de investigaciones exploratorias pueden definirse niveles que corresponden a grupos y subgrupos cada vez más amplios o de mayor especificidad o en función del constructo observado. En el caso de informes de estado de situación del sistema educativo, es posible analizar a nivel internacional, nacional, departamental, local y de niveles jerárquicamente menores.

Murillo (2008) se refiere a que los modelos multinivel son la metodología más adecuada para tratar datos «jerarquizados» o «anidados» (estudiantes en aulas, aulas en escuelas). En Digeduca, el análisis jerárquico multinivel de Factores Asociados al aprendizaje de un estudiante se desarrolla diseñando un modelo para cada nivel o grado. El analista podrá

consultar los informes de Factores Asociados del nivel primario que incluyen los grados de primero, tercero y sexto primaria, también los Factores Asociados de tercero básico y de graduandos. Todos estos informes utilizan la misma metodología (modelo jerárquico multinivel) pero con diferencias de diseño en el modelo. También podrá observar las diferencias de enfoque utilizando este mismo modelo multinivel en investigaciones relacionadas a Oportunidades de Aprendizaje (Pellecer y Santos, 2013), Factores Asociados al rendimiento de Lectura en primaria, entre otros elaborados en Digeduca (Del Valle, 2013).

Con relación al modelo, el informe de Factores Asociados de Graduandos 2011 indica que un análisis de regresión lineal multinivel permite profundizar en el análisis de resultados de Logro, según las características particulares de los establecimientos así como de los estudiantes. De esta cuenta se tiene un análisis simultáneo que involucra información del establecimiento educativo así como del estudiante. Permite analizar el vínculo anidado entre establecimientos y estudiantes de una forma más flexible y llevar a cabo un análisis estadístico en niveles jerárquicos en donde el estudiante representa el Nivel 1 y el establecimiento el Nivel 2. El análisis de Factores Asociados al rendimiento académico del estudiante involucra diferentes procesos que van desde la identificación de las variables que se desean conocer, así como el diseño del instrumento a aplicar, el proceso de recopilación de información y los procesos de análisis de varianza y multinivel, para lo cual es necesaria una adecuada selección de las variables observadas.

Por lo antes expuesto cabe preguntarse ¿por qué el modelo jerárquico multinivel es la herramienta más flexible para el análisis de estas realidades?

Se debe a que se estima la relación entre variables controlando las variaciones en cada nivel observado. Considera la independencia de cada variable, tomando en cuenta el sesgo de cada nivel, lo que deriva en mayor precisión considerando la dependencia de Nivel 2 y la verosimilitud de heterogeneidad de la relación de Nivel 1 que varía a través de las unidades de Nivel 2 (Moreno, Gálvez, Morales, Saz, Arriola, Johnson y Santos, 2009).

Para ilustrar esta aseveración, se considera que con regresión lineal es posible estimar la relación de estudiantes de primer grado en un establecimiento específico y de una sección, con algún nivel de error. Como indica el *Informe de Factores Asociados de Graduandos 2011*:

(...) un error también puede manifestarse debido a la presencia de variables extrañas que tienen un impacto en la investigación pero que no es posible identificarlas o cuantificarlas. O bien como resultado de la diferencia entre el valor esperado y el obtenido como resultado de la medición. Por lo anterior, el error que se considera en estadística no debe entenderse como el resultado de errores de medición sino como resultado de la delimitación de la realidad y la presencia de variables extrañas que no

pueden observarse pero que tienen un impacto en los resultados o como resultado de la presencia de otras variables que no han sido consideradas en el estudio y que podrían afectar o tener un impacto en los resultados obtenidos (Bolaños & Santos, 2013).

Cuando se desea hacer la relación de varios establecimientos educativos que involucran diferentes realidades complejas entre variables, el error de medición es mayor si no se procura el control de variables que llevaron a los estudiantes a cada uno de estos establecimientos. Como hace referencia Judd y Kenny (1981), cuando la asignación no es aleatoria, se requiere el control de la variable de asignación para hacer válidas las comparaciones entre los grupos. Si no se controlan estas variables, el modelo explicará menor porcentaje de la varianza y perderá precisión.

Una manera de obtener mayor precisión es identificar estas variables así como considerar grupos anidados en diferentes niveles, cada grupo y nivel con su propia varianza. Así, con el modelo de regresión lineal multinivel se considera las variaciones del primer nivel en función de la dependencia del segundo (ya que las estimaciones son combinaciones ponderadas de una gran media, más la estimación del parámetro obtenido) y sucesivamente este en función del tercer nivel. De esta manera, el modelo permite observar el efecto del Nivel 2 en las relaciones del Nivel 1 y la variación de todos los parámetros observados del Nivel 1 en el Nivel 2, sucesivamente.

En términos prácticos cuando se estudia la varianza en los grupos anidados a través de los niveles, se determina si hay diferencias significativas. Como ejemplo se propone lo siguiente: si se diseña un modelo que se propone conocer cuál es el efecto de las diferencias de sexo del estudiante, del efecto del capital socioeconómico y cultural, del nivel de educación del director, de la jornada de estudio o de las Oportunidades de Aprendizaje (ODA) en el rendimiento educativo, es posible obtener lo siguiente:

- a) Llegar a la conclusión que permite afirmar que existe un impacto de la variable «sexo masculino», posición en el indicador de capital socioeconómico y cultural, licenciatura como nivel mínimo del director o de la jornada vespertina o de las ODA.
- b) Caracterizar el impacto que se observa develando que el sexo masculino aporta al rendimiento (+), a mayor nivel socioeconómico y cultural se obtiene mayor rendimiento (>), la licenciatura como nivel educativo mínimo del director también tiene un efecto positivo (+), estudiar en jornada vespertina tiene un efecto negativo (-), a mayor ODA, mayor rendimiento (+).
- c) Determinar la significancia estadística de estas diferencias. En otras palabras, confirmar que la diferencia observada puede generalizarse a toda la población observada y debe tomarse como un factor que tiene el poder de caracterizar el rendimiento. En este caso se observa por ejemplo que aunque la jornada vespertina

parece disminuir el rendimiento, no es significativa a nivel de $p = 0.000 < 0.005$ (que se define como el nivel crítico mínimo para asegurarlo).

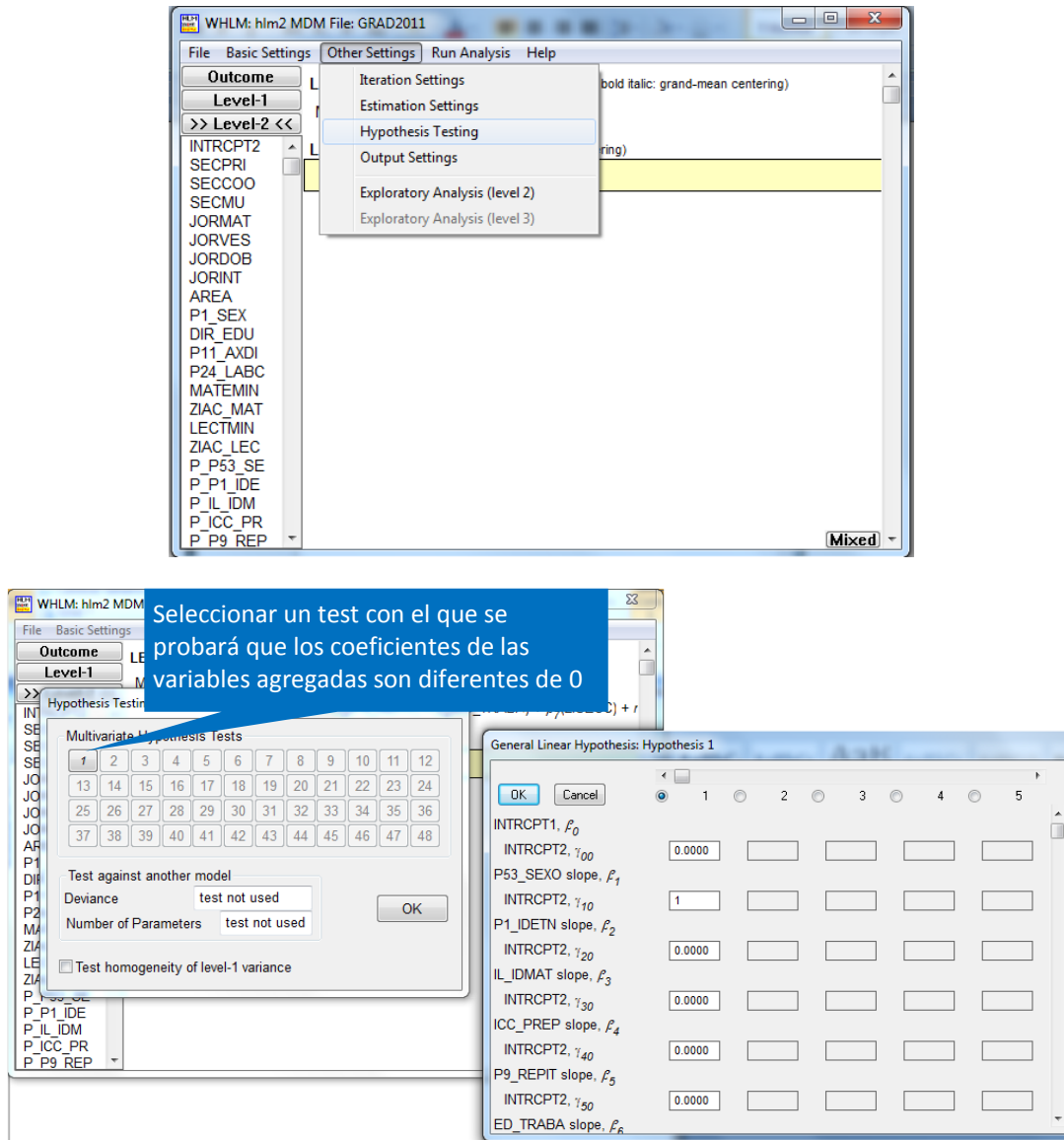
- d) Determinar la magnitud del impacto indicando que el sexo masculino posee mayor rendimiento que el femenino, estimar cuánto impacta la posición socioeconómica y cultural en unidades de rendimiento, cuánto aporta que el director posea nivel de licenciatura o cuánto impacta las ODA en el rendimiento de Matemática o Lectura.

En otras palabras, afirmar el impacto, caracterizar el impacto, determinar la significancia y la magnitud del efecto de la variable en el constructo que se mide. En general permite conocer cuánto varían los estudiantes entre ellos y entre las escuelas así como investigar el efecto de las variables observadas.

Debido a que el Modelo Jerárquico Lineal Multinivel se diseña con el espíritu científico de conocer el impacto de determinadas variables en el constructo observado (en este caso la habilidad latente de Matemática y Lectura), la elección de las variables a observar y controlar, deriva de una revisión responsable del marco teórico y de la observación de la realidad que motiva el interés científico para confirmar, analizar o probar nuevas relaciones entre variables o constructos. Se plantea una pregunta de investigación que busca someterse a prueba mediante una hipótesis y método idóneo. De tal manera que como efecto del diseño, se obtiene un modelo que se espera confirmar y estudiar. Para determinar qué tanto el modelo representa la realidad en función del diseño establecido, es necesario observar los indicadores del modelo.

El *software* comúnmente utilizado en la Digeduca para realizar el análisis multinivel es el (Hierarquical Linear Model) HLM en su versión 7. Es posible realizar una prueba de Hipótesis con el *software* utilizado. Esto permitirá confirmar que las variables del diseño son adecuadas para el modelo propuesto, como se observa en la Figura 1.

Figura 1. Procedimiento para prueba de hipótesis



En la figura se observa la elección del menú para elaborar la prueba de hipótesis. Es necesario seleccionar un test con el que se probará que los coeficientes de las variables son diferentes de 0 e indicarlo en la ventana siguiente. Esto se hace para cada variable que se desea someter a prueba.

En la Figura 2 se muestran los resultados de la prueba de hipótesis que se obtiene al correr el modelo. En esta se observa que existe evidencia estadísticamente significativa para rechazar la hipótesis nula ya que los valores estimados son diferentes de 0, por esta razón las variables deben permanecer en el modelo.

Figura 2. Interpretación de la prueba de hipótesis

Results of General Linear Hypothesis Testing - Test 2

	Coefficients	Contrast					
For INTRCPT1, β_0							
INTRCPT2, γ_{00}	459.126610	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P1_SEX, γ_{01}	-15.229797	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
For P53_SEXO slope, β_1							
INTRCPT2, γ_{10}	34.455922	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
For P1_IDETN slope, β_2							
INTRCPT2, γ_{20}	17.341869	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
For IL_IDMAT slope, β_3							
INTRCPT2, γ_{30}	5.362912	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
For ICC_PREP slope, β_4							
INTRCPT2, γ_{40}	-0.408053	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
Estimate		-15.2298	34.4559	17.3419	5.3629	-0.4081	
Standard error of estimate		2.1307	0.7233	0.7685	1.1868	0.6771	

χ^2 statistic = 2897.312580
 Degrees of freedom = 5
 p-value = <0.001

Se rechaza H0 ya que los valores estimados son diferentes de 0. Estas variables deben permanecer en el modelo.

La prueba de hipótesis permitirá decidir qué variables deben permanecer en el modelo. Esta prueba puede hacerse tanto para variables de Nivel 1 como de Nivel 2. Sin embargo, es importante señalar que el analista que realiza un informe de Factores Asociados y que se basa en un modelo anterior (años anteriores), puede realizar esta prueba de hipótesis para observar el comportamiento de las variables pero no como fundamento para tomar una decisión de eliminación de la variable del modelo. Esto no será así en el caso que el analista se encuentre diseñando un nuevo modelo o realizando una investigación específica y desee incluir únicamente las variables que son significativas. Estas pruebas se validarán con la Coordinación del área de Análisis Cuantitativo y la Subdirección cuando se obtengan los resultados finales.

En la estimación del modelo, el analista podrá observar que existe la posibilidad de centrar las variables de Nivel 1 y Nivel 2 a la gran media o a la media del grupo.

- Centrado en la gran media: se centran las variables en el centro de la distribución, por lo cual la interpretación del intercepto es más precisa. Representa el valor de la media del grupo para una persona.
- Centrado en la media del grupo: se centran las variables en la media del grupo cuando se desea conocer el efecto del grupo al que pertenecen los sujetos. Compara a los sujetos con la media de su propio grupo.

Para la elaboración del informe de Factores Asociados al aprendizaje de las cohortes evaluadas anteriores, se centra en la gran media todas las variables observadas en el Nivel 2 (establecimiento), debido a que produce estimaciones del intercepto que reflejan el promedio

de todos los evaluados. Esta decisión puede variar cuando se trata de análisis específicos o investigaciones que buscan observar el efecto de una sola variable centrada en la gran media, por ejemplo cuando se desea observar el efecto de la proporción de estudiantes de sexo masculino en la población (variable dummy).

Las variables de Nivel 1 no se centran en los informes de Factores Asociados al aprendizaje anteriores, porque lo que se espera es estimar la varianza de Nivel 1 en función de la varianza de Nivel 2. En las variables dummy (dicotómicas), no se centran las variables para hacer interpretaciones directas y conclusiones sobre la población entera.

Es posible estimar efectos composicionales en el modelo. Estos efectos se aplican cuando se conoce que las diferencias del constructo que se mide (habilidad en Matemática o Lectura) pueden ser atribuidas a diferencias en una o más variables observadas (por ejemplo, que el nivel socioeconómico tiene un efecto diferencial en el rendimiento de los estudiantes). Se estudia a manera de cascada, fijando el efecto composicional centrando las variables de Nivel 1 en la media del grupo, posteriormente centrando las variables de Nivel 2 en la gran media.

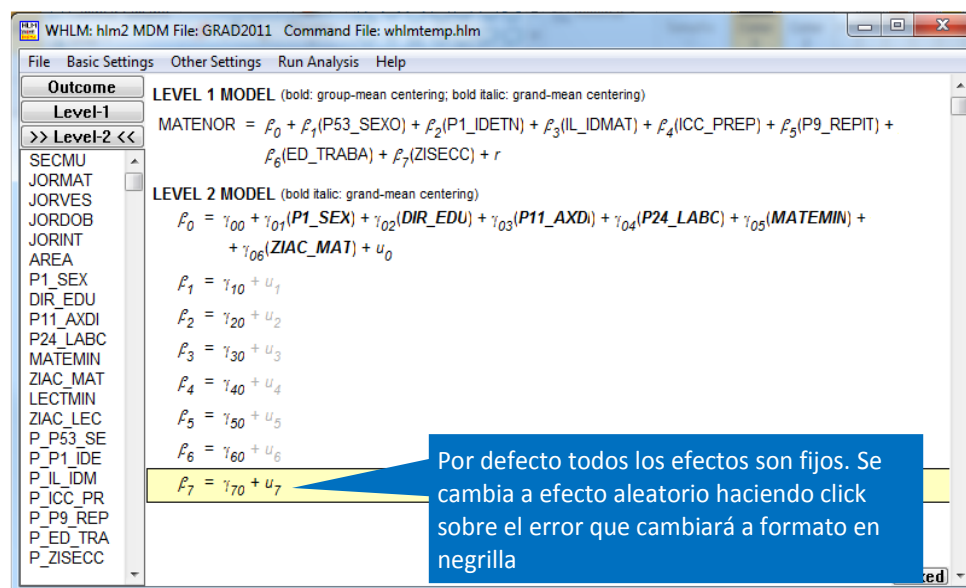
Lo anterior permitirá estimar el efecto promedio de la variable observada en la escuela y su efecto en el rendimiento después de controlar la variable por estudiante. Al respecto, cabe recordar que en los informes de Factores Asociados de estudiantes evaluados según grado, no se fijan efectos en el Nivel 1 y en el Nivel 2 se centran las variables en la gran media. Esto podría no ser así en otros estudios elaborados con otros objetivos.

También es posible observar los efectos de una o más variables seleccionando una de dos opciones, efectos fijos o efectos aleatorios.

- Con efecto fijo: es posible definir un efecto fijo en alguna variable debido a que se desea restringir el mismo valor para todos los grupos (el mismo promedio, la misma proporción en todos los grupos). Es posible decidir por un efecto fijo cuando: a) no existen suficientes sujetos en los grupos como para estimar efectos aleatorios en el Nivel 1 (lo que podría provocar que el modelo no converja), b) cuando se sabe que el efecto es fijo, como resultado de una prueba estadística de comparación de devianzas que indica que un efecto fijo es significativo, c) cuando existe la probabilidad de colinealidad de efectos aleatorios por lo que se decide fijar para evitar colinealidad y, d) cuando el marco teórico indica que el efecto de la variable que se observa es fijo.
- En el Nivel 1, el único efecto fijo es el intercepto debido a que representa el promedio de todos los estudiantes de todas las escuelas.
- En el Nivel 2 todas las variables observadas de Nivel 1 toman un efecto fijo por defecto. Pero puede indicarse un efecto aleatorio activando la variable de error.

En la Figura 3 se muestra el procedimiento para elegir la opción de efectos aleatorios en el modelo.

Figura 3. Elegir efectos aleatorios en el modelo (aplica únicamente para investigaciones específicas)



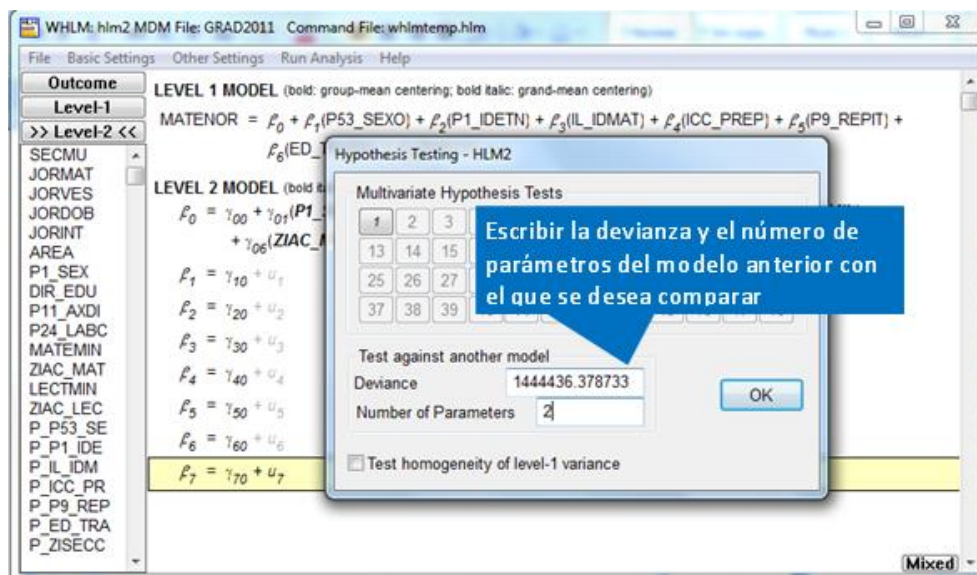
- Con efecto aleatorio: el coeficiente resultante de cada variable tiene pendientes diferentes.

Es importante recalcar que en los informes de Factores Asociados de estudiantes evaluados según grado, se utilizan efectos fijos en el Nivel 2, es decir los coeficientes obtenidos dependen de las pendientes observadas en el Nivel 1. En el Nivel 1, se utiliza el único efecto fijo por defecto (el intercepto), debido a que se estima el rendimiento de todos los estudiantes en todos los establecimientos cuando se controla por las variables observadas en el modelo propuesto.

En el caso de investigaciones específicas podría definirse un efecto aleatorio en una variable del Nivel 2 cuando se tomó la decisión de fijarlo conociendo que el efecto de la variable es producto de la estimación con todas las pendientes de las variables observadas (como pudiera suceder en el caso de clima escolar, ODA). Puede utilizarse una prueba de DEVIANZA para confirmar que la variable de interés posee un efecto aleatorio.

Como ejemplo, la Figura 4 muestra la prueba de DEVIANZA utilizando el Indicador Socioeconómico y de Capital Cultural (ZISECC). Como procedimiento en el menú de prueba de hipótesis se indica la devianza y el número de parámetros del modelo con el que se desea comparar (en el ejemplo se compara con la DEVIANZA del Modelo Nulo).

Figura 4. Prueba de DEVIANZA que puede utilizarse para elegir efectos fijos o aleatorios de variables en el modelo



En los resultados del modelo se encontrará la estimación resultante de la prueba de DEVIANZA. Como se observa en la Figura 5, la diferencia es estadísticamente significativa por lo que podría considerarse que el ISECC es un efecto aleatorio.

Figura 5. Resultados de la prueba de DEVIANZA

MODELO NULO

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, u_0	53.66820	2880.27601	2937	58053.18716	<0.001
level-1, r	83.26316	6932.75375			

Statistics for current covariance components model

Deviance = 1444436.378733

Number of estimated parameters = 2

Devianza del modelo con efecto aleatorio

EFECTOS FIJOS EN EL INDICADOR SOCIOECONÓMICO Y DE CAPITAL CULTURAL

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, u_0	31.11974	968.43795	2888	12618.06128	<0.001
ZISECC slope, u_1	8.67393	75.23705	2894	3256.12654	<0.001
level-1, r	79.28691	6286.41478			

Note: The chi-square statistics reported above are based on only 2895 of 2938 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

Statistics for current covariance components model

Deviance = 1422614.444366

Number of estimated parameters = 4

Devianza del modelo con efecto fijo

Variance-Covariance components test

χ^2 statistic = 21821.93437

Degrees of freedom = 2

p-value = <0.001

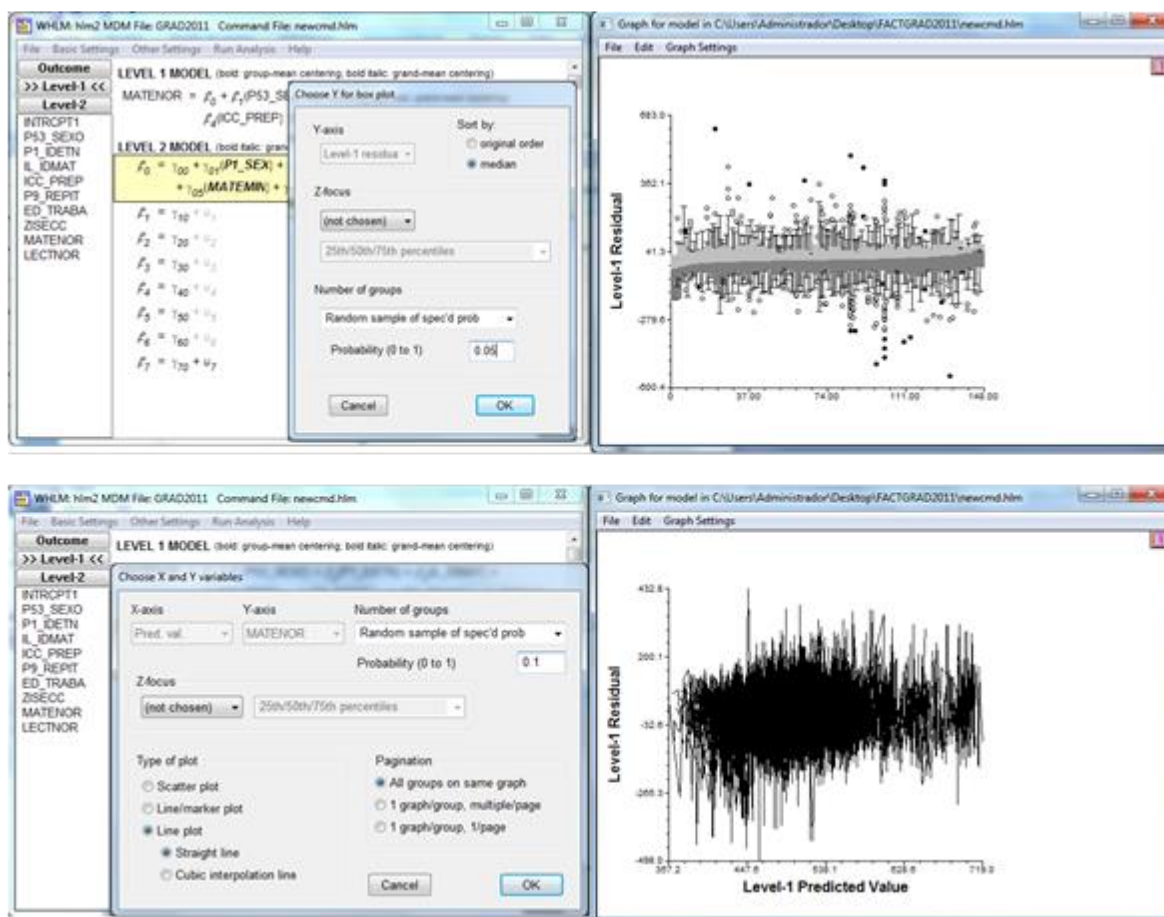
La diferencia es estadísticamente significativa entre los modelos. Se podría considerar que el ZISECC es un efecto aleatorio

También es posible realizar otras pruebas con el *software* HLM como la verificación de supuestos.

- En el Nivel 1: test de homoscedasticidad, verificación de la normalidad de los errores con media cero, verificación de la independencia entre predictores y errores.
- En el Nivel 2: errores normalmente distribuidos con media cero y verificación de independencia de predictores de Nivel 2 del error de ese nivel.

En la Figura 6 se muestra el procedimiento para verificar supuestos utilizando los recursos gráficos del software. Es necesario indicar el orden y el valor de probabilidad.

Figura 6. Verificación de supuestos



Los resultados del modelo proporcionan una serie de indicadores que el analista debe considerar en todo momento, los cuales se detallan en los siguientes incisos.

- La confiabilidad del modelo. En la estimación se obtiene un indicador de confiabilidad que el analista debe verificar y observar.
- El indicador CINTRA permite conocer el efecto de la escuela en el rendimiento educativo, es decir, la variabilidad entre establecimientos que puede interpretarse como un indicador de desigualdades de aprendizaje entre las escuelas y al interior de las escuelas. Se calcula sumando la varianza de los niveles 1 y 2. Se obtiene el porcentaje de la proporción observada de la varianza de Nivel 2 entre la Varianza Total (para mayor información ver Glosario).
- El indicador DEVIANZA permite estimar la bondad de ajuste del modelo. En el modelo Nulo, en donde no se observa el efecto de alguna variable, se obtiene un

coeficiente que puede considerarse el punto de partida del modelo propuesto. La introducción de variables al modelo debiera disminuir este indicador, lo que significa que el modelo se ajusta cada vez mejor. El analista debe observar este indicador ya que en la medida que introduce las variables de estudio, la interacción entre las variables en los diferentes niveles producen un efecto en este indicador que debiera disminuir, develando un mejor ajuste. También debe reflexionar acerca de la magnitud de las diferencias observadas cuando ejecuta el modelo. Este indicador lo calcula el *software* HLM.

- d) El intercepto (γ_{00}). Es la recta de promedio generada en la habilidad latente (rendimiento en Matemática y Lectura), cuando se introducen las variables de interés en el modelo (deriva de la presencia de todas las rectas de variables observadas). Es el único efecto fijo de Nivel 1 por defecto del modelo, que puede interpretarse como el promedio de Logro en Lectura de todos los estudiantes de todas las escuelas. Un modelo aceptable debiera mantener el intercepto inicial, observado en el modelo Nulo o bien, incrementar su valor. El intercepto lo estima el *software* HLM.
- e) La varianza residual del Nivel 1 representa las diferencias observadas entre las realidades que se definan para este nivel. En informes de DigeDuca de Factores Asociados al aprendizaje (según grado), esta varianza representa las diferencias entre estudiantes. De igual manera, la varianza residual de Nivel 2 representa las diferencias observadas entre escuelas. La varianza de Nivel 2 se calcula de igual manera que CINTRA y la varianza de Nivel 1 estimando el porcentaje de la proporción observada de la varianza de Nivel 1 entre la Varianza Total.
- f) En el modelo es posible observar el poder predictivo del modelo (aproximación a R^2). Esto permitirá conocer qué porcentaje de la varianza predice el modelo. Este es otro indicador de bondad de ajuste del modelo que da una estimación de la capacidad predictiva del mismo en cada nivel (uno para el Nivel 1 y otro indicador para el Nivel 2). Para su cálculo se resta la varianza observada del nivel que se estima (por ejemplo, Nivel 1) que es resultante de la introducción de variables, de la varianza obtenida del mismo nivel en el Modelo Nulo (Nivel 1). El resultado se divide dentro de la varianza del mismo nivel (Nivel 1) del modelo Nulo y posteriormente se multiplica por 100. El mismo procedimiento se realiza para los dos niveles y como resultado se obtiene el porcentaje que el modelo predice de la varianza observada.

La ecuación del Modelo Jerárquico Lineal se representa de la siguiente manera:

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

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

4.4. Teoría de Respuesta al Ítem (TRI) o Teoría de Rasgo Latente (TRL)

La Teoría de Respuesta al Ítem (TRI), inicialmente conocida como Teoría o Modelo de Rasgo Latente, enfoca su atención en los ítems más que en el instrumento de evaluación. El análisis que se desarrolla con este modelo puede abarcar desde los resultados obtenidos en los ítems hasta las propiedades psicométricas de estos ítems y de la prueba que los integra. La TRI, como modelo, asume como supuestos lo siguiente:

- Curva Característica de los Ítems (CCI): existe una relación funcional entre los valores de la variable que miden los ítems y la probabilidad de responder correctamente a estos que se denomina CCI (el valor máximo de probabilidad es 1, es decir, 1 al acertar y 0 al fallar). Acertar a un ítem es posible únicamente por la habilidad que el sujeto posee en la variable medida. La habilidad se representa con el símbolo θ y sus puntuaciones comprenden $-\infty$ a $+\infty$. Cada ítem tiene un CCI que utiliza la función logística definida por tres parámetros (Índice de discriminación o “a”, Índice de dificultad o “b” e Índice de probabilidad de acertar al azar o “c”).
- Unidimensionalidad: todos los ítems de la prueba miden una sola variable o rasgo latente (que puede ser un constructo, aptitud o conocimiento). Es decir, la posibilidad de responder de una persona a un ítem o a la prueba, depende de su nivel de habilidad en una sola variable. Por esta razón, para llevar a cabo el análisis en TRI, los ítems deben ser unidimensionales y dicotómicos.
- Independencia local: la probabilidad de responder un ítem es independiente a la probabilidad de responder cualquier otro (no existe correlación de ningún ítem con otro). De forma análoga, el rendimiento de un sujeto es independiente del rendimiento de otro.

5. Procedimiento

5.1. Diseño del marco metodológico del análisis

Dirección técnica, supervisión, control: Subdirección, Coordinación de área

Responsable de elaboración y realimentación: analista

Productos:

- Marco metodológico
- Identificación del modelo(s) de análisis
- Propuesta de estructura del informe

Como paso fundamental de elaboración del análisis e informe de Factores Asociados al aprendizaje, la Subdirección y Coordinación de área son responsables de definir y establecer el marco de interés institucional, considerando las tendencias contemporáneas de análisis de Factores Asociados, los métodos y medios tecnológicos de punta para el análisis.

El analista recibirá la dirección técnica del estudio a través de la Coordinación de área o cuando se estime pertinente, en coordinación con la Subdirección. En este contexto, es necesario que el analista aclare dudas y brinde realimentación. Como resultado del análisis y discusión (comunicación vertical), se fijarán los lineamientos sobre los cuales se diseñará el marco metodológico.

El marco metodológico del estudio será plasmado por el analista en un documento borrador que incluya el propósito del estudio, el (los) modelo (s) de análisis, la estructura del informe y una propuesta de referencias. El documento será presentado a la Coordinación de área para revisión y aprobación.

El modelo se diseñará tomando en cuenta las variables de interés en función del propósito de la investigación: réplica del modelo o investigación específica. Es necesario recordar que en virtud del propósito de la investigación y en el proceso de análisis de los resultados del Modelo Jerárquico Multinivel, algunas variables podrían obtener coeficientes no significativos y por ende ser objeto de eliminación del modelo final. En el proceso de análisis y toma de decisión, la Coordinación de área aportará a la discusión de resultados y aprobará las decisiones que se tomen en cuanto a los cambios que tengan lugar en el modelo propuesto inicialmente.

La Coordinación de área recibirá el documento, solicitará las correcciones necesarias y trasladará a la Subdirección. Ambos niveles definirán el procedimiento más adecuado para la revisión y aprobación del documento marco. Una vez aprobado, se notificará al analista quien continuará con las fases posteriores.

5.2. Revisión del marco teórico

Responsables: Coordinación de área, analista

Producto:

- Borrador de marco teórico (que será enriquecido durante el proceso de elaboración del informe)

El analista debe conocer y explorar el referente teórico sobre Factores Asociados al aprendizaje. La Coordinación de área puede orientar o delimitar esta consulta en función de los objetivos trazados para el análisis. En esta fase, se deben alcanzar los siguientes objetivos:

- a) Revisar y conocer el marco teórico internacional de Factores Asociados.
- b) Revisar y conocer los informes elaborados por Digeduca sobre Factores Asociados al aprendizaje.
- c) Actualizar su conocimiento de investigaciones nacionales e internacionales sobre Factores Asociados al aprendizaje.
- d) Elaborar el borrador del marco teórico.

Es deseable que tanto el Coordinador de área como el analista conozcan los referentes teóricos existentes aun cuando no se refieran con exclusividad a las áreas evaluadas por Digeduca (Matemática y Lectura) y las investigaciones que se realicen en estudiantes de diferentes grados y niveles educativos.

Considérese como referente internacional básico las siguientes fuentes que se indican en la Tabla 1 en la cual se identifican las evaluaciones y los años en los que se llevaron a cabo o se tiene programado evaluar.

Tabla 1. Marco teórico internacional básico de Factores Asociados

	EVALUACIÓN	AÑOS EVALUADOS O POR EVALUAR
Trends in International Mathematics and Science Study (TIMSS) y Progress in International Reading Literacy Study (PISA)	TIMSS	1995, 1999, 2003, 2007, 2011
	TIMSS AVANZADO	1995, 2008
	PIRLS	2001, 2006, 2011
	PIRLS AVANZADO	2011
Program for International Student Assessment (PISA)	MATEMÁTICAS	2003, 2012
	LECTURA	2000, 2009
	CIENCIAS	2006, 2015
Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura	Eficacia escolar y factores asociados en América Latina y el Caribe	2008
	Factores Asociados al Logro Cognitivo de los Estudiantes de América Latina y el Caribe –SERCE-	2010

Fuente: Digeduca.

En el marco teórico nacional e internacional se han elaborado otros referentes teóricos e investigación de Factores Asociados al aprendizaje que se pueden consultar con el propósito de enriquecer el análisis y la discusión, conforme a los objetivos definidos para el informe a elaborar.

El analista también debe conocer los referentes básicos de Digeduca de Factores Asociados. En la Tabla 2 se lista los informes desarrollados por la Dirección, en el marco del análisis de Factores Asociados al aprendizaje.

Tabla 2 Marco Teórico Nacional básico de Factores Asociados

2006	Informe de resultados de Graduandos 2006
2007	Informe de resultados de Graduandos 2007
2008	Informe técnico de Factores Asociados al rendimiento escolar de Graduandos 2008
2009	Informe de Factores Asociados al aprendizaje de tercero básico
2010	Informe de Factores Asociados al aprendizaje de primero primaria
2011	Informe de Factores Asociados al aprendizaje Graduandos 2011
2012	Análisis de las Oportunidades de Aprendizaje (ODA) en estudiantes del ciclo básico del Nivel de Educación Media
2012	La desigualdad social y el aprendizaje de Lectura y Matemática en estudiantes de sexto grado de primaria
2013	Informe de Factores Asociados al aprendizaje de tercero básico
2013	Informe de Factores Asociados al aprendizaje de primero primaria
2013	Informe de Factores Asociados al aprendizaje de directores de tercero básico

Fuente: Digeduca.

Al final de esta fase el analista tendrá listo un borrador del marco teórico, previo al análisis estadístico que enriquecerá durante en el proceso de elaboración del informe.

5.3. Procedimiento técnico previo

Responsable de elaboración: analista

Responsable de supervisión, verificación y control: Coordinación de área

Software: SPSS, EXCEL

Productos:

- **Garantía de calidad de bases de datos a utilizar**
- **Indicador (es) Socioeconómico / Cultural elaborado (s) y estandarizado (s)**
- **Variables de habilidad latente transformadas a media 500 y desviación estándar 100**
- **Variables observadas recodificadas según el (los) modelo (s) de análisis**
- **Bases de datos imputadas (de los nivel 1 y 2)**

En esta fase se trabajará con la base de datos de la evaluación de estudiantes con su respectiva calificación (de aquí en adelante, base del Nivel 1) y de directores de los establecimientos evaluados en ese grado y nivel (de aquí en adelante, base del Nivel 2). Los pasos que se describen a continuación deben ser garantizados antes de correr el análisis estadístico.

Se trabajará con los programas SPSS y EXCEL. Aunque el analista puede elegir trabajar con ambas bases al mismo tiempo (Nivel 1 y Nivel 2), se recomienda que trabaje primero con una y después con otra.

5.3.1. Procedimiento de preparación de bases de Nivel 1 y 2

Responsable de elaboración: analista

Responsable de supervisión, verificación y control: Coordinación de área

Productos finales:

- Bases resultantes de la preparación técnica (de Nivel 1 y Nivel 2) se verificaron y superaron los criterios de calidad, responden al modelo propuesto e integran las variables que por su idoneidad o interés científico serán analizadas con el análisis jerárquico multinivel.
- Archivos de respaldo de las bases resultantes de los niveles 1 y 2.
- Garantía de calidad de bases de datos de los niveles 1 y 2.

Productos intermedios:

- Variables del modelo de los niveles 1 y 2 fueron recodificadas, transformadas y son de tipo numérico.
- Indicador/es de los niveles 1 y 2 elaborado/s y estandarizado/s.
- Variables imputadas en los niveles 1 y 2.
- Variables de habilidad latente estandarizadas y transformadas a una escala de media 500 y desviación estándar 100.
- Variable de código de establecimiento con iguales características en los niveles 1 y 2, numéricas y ordenas de forma ascendente.
- Resultados generados y análisis de todas las variables de los niveles 1 y 2 en cuanto a la relación con habilidad latente, poder de determinación e interacción.

El analista alcanzará los siguientes objetivos:

- a) Verificar que la base de datos a utilizar ha completado el proceso de limpieza y de calificación en TRI para el grado y nivel que se analiza (para mayor información sobre este proceso consulte los Manuales de Limpieza y Calificación).
- b) Conocer y revisar la base, garantizando la calidad de los datos y que cuenta con la calificación final en TRI (Ver Manual de Limpieza).
- c) Garantizar que la variable resultante de calificación TRI (en escala *Logit*) denominada por Dgeduca MEASURE de Matemática y MEASURE de Lectura, se estandarizó y se transformó a una escala con media 500 y desviación estándar 100.
- d) Garantizar que las bases de los niveles 1 y 2 que se utilizarán en el análisis de Factores Asociados, cuentan con las variables que se utilizarán en el (los) modelo (s) de análisis que se utilizará para este grado y nivel.

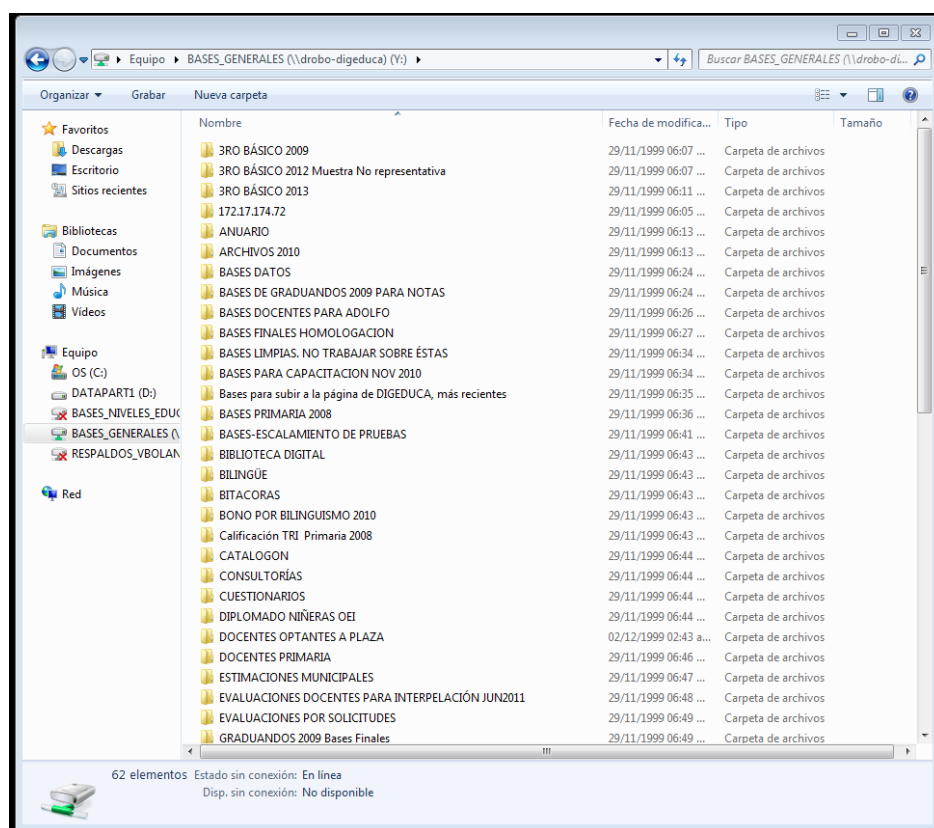
- e) Garantizar que las variables de los niveles 1 y 2 que se utilizarán son numéricas (con excepción del código de estudiante que permanece en la base del Nivel 1 como referencia) y están recodificadas según lo diseñado en el modelo de Análisis.
- f) Garantizar que la variable/s nivel socioeconómico/cultural fueron generadas tomando en consideración los lineamientos en el *Manual de Elaboración de Nivel Socioeconómico y otros indicadores* y se encuentra/n estandarizadas.
- g) Imputar variables de los niveles 1 y 2 como procedimiento posterior a la exploración y selección del método adecuado.
- h) Garantizar que el código de establecimiento posee las mismas características en las bases finales de los nivel 1 y 2, que son numéricas y se encuentran ordenadas de forma ascendente.
- i) Analizar la relación, el poder de determinación de todas las variables de interés de los niveles 1 y 2 con la habilidad latente y su interacción.
- j) Garantizar que las bases resultantes de la preparación técnica del Nivel 1 y Nivel 2 se verificaron y superaron los criterios de calidad, y que responden al modelo propuesto e integran las variables que por su idoneidad o interés científico serán analizadas con el análisis jerárquico multinivel.
- k) Generar un archivo de respaldo de las bases resultantes de los niveles 1 y 2.

Paso 1. Base del Nivel 1. Ubicar la base verificada final. Explorar la base de datos que se utilizará en el análisis y garantizar que posee la calificación TRI final. Para garantizar este paso debe responderse afirmativamente las siguientes preguntas:

- ¿La base de datos es la última versión generada con la calificación TRI de Matemática y Lectura (MEASURE_MATE, MEASURE_LECTURA)?
- ¿La base de datos fue verificada y confirmada como la final por la Subdirección?
- ¿La base de datos finalizó el proceso de limpieza?

El analista puede garantizar que utiliza la última versión, toda vez que copie y pegue la base final que encontrará en la carpeta compartida (en el servidor Y:). En la Figura 7 se muestra la ubicación de la carpeta compartida.

Figura 7. Ubicación de carpeta compartida



Paso 2. Base del Nivel 1. Garantizar la calidad de la base a utilizar. Para garantizar la calidad de la base, el analista identificará las variables que se utilizarán en el Modelo Jerárquico Multinivel (Tabla 4), los códigos del estudiante y del establecimiento así como las

variables que se utilizarán para elaborar el/los indicador/es Socioeconómico y Cultural. Estas variables deben ser exploradas y garantizar:

- Que no existen códigos del estudiante duplicados.
- Que todos los estudiantes poseen código de establecimiento.
- Que las variables a utilizar se encuentran limpias (observar el tipo de variable y valores de las variables).

Paso 3. Base del Nivel 1. Recodificar. Recodificar las variables según el modelo que se ha definido para el análisis multinivel. En la Tabla 3 se observa que el sector, jornada, área y sexo se analizarán como variables dummy. En las bases de calificación estas variables son categóricas, de tal manera que es necesario recodificar las variables.

Tabla 3. Sintaxis de recodificación para el análisis de variables de sector, jornada, área y sexo

RECODIFICAR SECTORES <pre> RECODE cod_sector ('21'=1) ('22'=0) ('23'=0) ('24'=0) INTO OFICIAL. EXECUTE. RECODE cod_sector ('21'=0) ('22'=1) ('23'=0) ('24'=0) INTO PRIVAD. EXECUTE. RECODE cod_sector ('21'=0) ('22'=0) ('23'=1) ('24'=0) INTO MUNI. EXECUTE. RECODE cod_sector ('21'=0) ('22'=0) ('23'=0) ('24'=1) INTO COOP. EXECUTE. </pre> *RECODIFICAR JORNADA* <pre> RECODE cod_jornada ('31'=1) ('32'=0) ('33'=0) ('34'=0) ('35'=0) INTO MATUT. EXECUTE. RECODE cod_jornada ('31'=0) ('32'=1) ('33'=0) ('34'=0) ('35'=0) INTO VES. EXECUTE. RECODE cod_jornada ('31'=0) ('32'=0) ('33'=1) ('34'=0) ('35'=0) INTO DOB. EXECUTE. RECODE cod_jornada ('31'=0) ('32'=0) ('33'=0) ('34'=1) ('35'=0) INTO NOCT. EXECUTE. </pre>	<pre> RECODE cod_jornada ('31'=0) ('32'=0) ('33'=0) ('34'=0) ('35'=1) INTO INTER. EXECUTE. </pre> *RECODIFICAR ÁREA* <pre> RECODE cod_area ('11'=1) ('12'=0) INTO URBANA. EXECUTE. </pre> *RECODIFICAR SEXO* <pre> RECODE SEXO (1=0) (0=1) INTO MASCULINO. EXECUTE. </pre> *RECODIFICAR ETNIA LADINO* <pre> RECODE IE_IDENTIFICACION_ETNICA (2=1) (SYSMIS=SYSMIS) (ELSE=0) INTO E_LADINO. EXECUTE. </pre> *RECODIFICAR IDIOMA MATERNO ESPAÑOL* <pre> RECODE IE_IDIOMA_MATERNO (1=1) (SYSMIS=SYSMIS) (ELSE=0) INTO I_ESPANOL. EXECUTE. </pre>
--	--

Fuente: Digeduca.

La sintaxis de recodificación en SPSS puede correrse utilizando las variables de código de sector, código de jornada tal y como se muestra en la Tabla 4. En el proceso de recodificación no debe olvidarse que como política institucional:

- Toda ausencia de la característica de que indica la variable debe recodificarse como 0. Como ejemplo, en las variables estudió en preprimaria, repitió y el estudiante trabaja la recodificación es 0=NO, 1=SÍ.
- Todo grupo en desventaja se recodifica como 0. Como ejemplo, el grupo indígena, mujer.
- En el caso de variables de las que se desee conocer el impacto del nivel de la característica. La transformación debe considerar un ordenamiento de menor a mayor indicando la ausencia de la característica como 0. El uso de estas variables se conversará con la Coordinación de área en la fase del diseño del modelo y en caso de reconsiderar, incluir una variable de este tipo en esta fase, debe ser analizado y aprobado por la Coordinación. Este extremo se debe a que la inclusión de estas variables demandará mayor habilidad de interpretación por parte del analista una vez que se obtengan los resultados del modelo.

Tabla 4. Variables del modelo Multinivel para el análisis de Factores Asociados de Graduandos 2011

VARIABLES DEPENDIENTES				MODELO NULO O VACÍO: Habilidad de Matemáticas, Habilidad de Lectura			
VARIABLES INDEPENDIENTES				VARIABLES DE NIVEL 2			VARIABLES DE NIVEL 1
				MODELO SECTOR	MODELO ESTRUCTURAL	MODELO COMPOSICIONAL	MODELO DEL ESTUDIANTE
Establecimiento del sector Privado	1	Otro	0	X	X	X	X
Establecimiento del sector por Cooperativa	1	Otro	0	X	X	X	X
Establecimiento del sector Municipal	1	Otro	0	X	X	X	X
Jornada Matutina	1	Otra	0		X	X	X
Jornada Vespertina	1	Otra	0		X	X	X
Jornada Doble	1	Otra	0		X	X	X
Jornada Intermedia	1	Otra	0		X	X	X
Área: Urbana	1	Rural	0		X	X	X
Género del Director: Masculino	1	Femenino	0		X	X	X
Nivel de Educación del Director: Licenciatura mínimo	1	Otro	0		X	X	X
Experiencia del Director	A > años de experiencia				X	X	X
Laboratorio de Computación	1	No existe	0		X	X	X
Tiempo de enseñanza de Matemáticas y Lectura al mes	Cantidad de minutos y períodos al mes				X	X	X
Indicador de Actividades de Aprendizaje					X	X	X
Proporción de estudiantes de género masculino						X	X
Proporción de estudiantes ladinos						X	X
Proporción de estudiantes con Idioma materno español						X	X
Proporción de estudiantes que asistieron a preprimaria						X	X
Proporción de repitentes						X	X
Proporción de estudiantes que trabajan						X	X
Promedio de Indicador Socioeconómico y de Capital Cultural (ISECC)						X	X
Género del Estudiante: Masculino	1	Femenino	0				X
Identificación étnica del estudiante: Ladino	1	Otra	0				X
Idioma materno del estudiante: Español	1	Otro	0				X
Estudiante asistió a preescolar	1	No	0				X
Estudiante repitió en primaria	1	No	0				X
El estudiante trabaja	1	No	0				X
Indicador Socioeconómico y de Capital Cultural (ISECC)							X

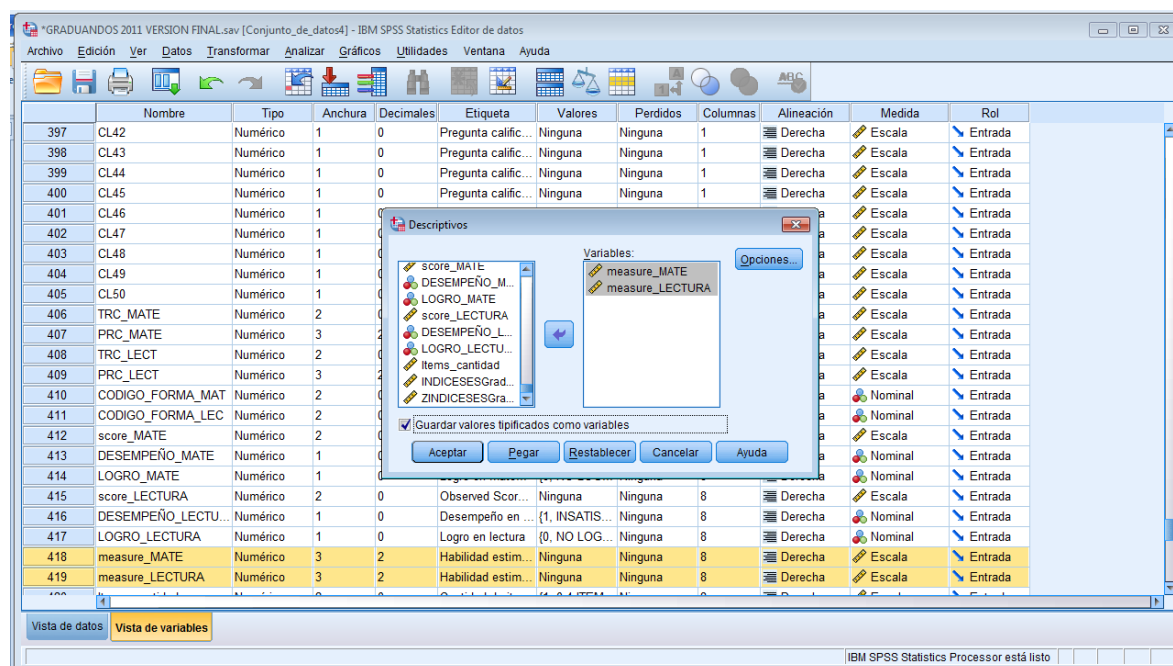
Fuente: Informe de Factores Asociados de Graduandos 2011.

Paso 4. Base del Nivel 1. Garantizar variables numéricas. Garantizar que las variables recodificadas son numéricas.

Paso 5. Base del Nivel 1. Elaborar indicadores. Elaborar el Indicador/es Socioeconómico/Cultural tomando en consideración el paso previo de exploración de variables de interés, recodificación, correlación con la habilidad estimada en Matemática y Lectura así como con el Logro alcanzado en estas áreas; el procedimiento de reducción factorial y estandarización del indicador (Para mayor información consulte el *Manual de elaboración del Nivel Socioeconómico y otros indicadores*).

Paso 6. Base del Nivel 1. Estandarizar habilidad latente. Estandarizar la habilidad en Matemática y Lectura (MEASURE). Para el efecto, el analista puede utilizar la herramienta de SPSS en el menú Analizar/Estadísticos Descriptivos/Descriptivos en donde se desplegará la imagen que se observa en la figura. El analista introduce las variables a estandarizar (MEASURE_MATE, MEASURE_Lectura) y elige la opción «Guardar valores tipificados como variables». Como resultado se guardarán las variables estandarizadas como puntuaciones Z que en la preparación de la base del Nivel 1 de Graduandos 2011, se denominaron: «Zmeasure_MATE» y «Zmeasure_Lectura». El procedimiento para estandarizar las variables de habilidad en Matemática y Lectura se ilustra en la Figura 8.

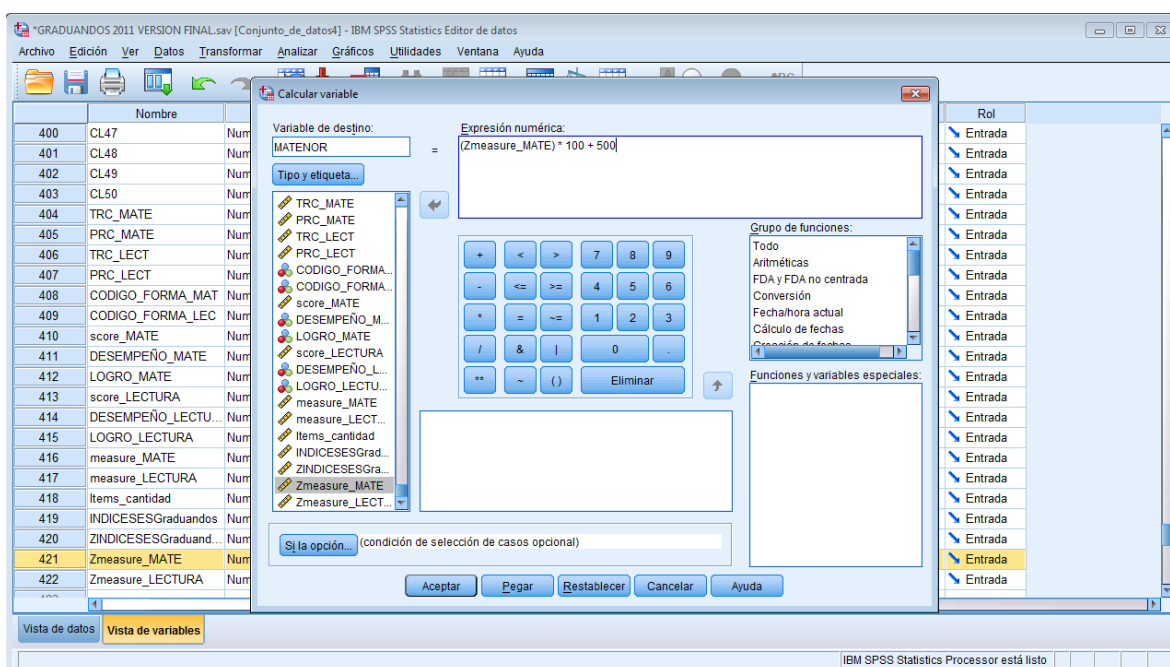
Figura 8. Ubicación de herramienta en SPSS para estandarizar variables de habilidad latente en Matemática y Lectura



Paso 7. Base de Nivel 1. Transformar habilidad latente. El analista transforma la habilidad en Matemática y Lectura (MEASURE) con una media de 500 y desviación estándar de 100. La transformación puede realizarse utilizando la herramienta de SPSS en el menú Transformar/Calcular variable.

En esta opción se indica el nombre de la variable a utilizar en el análisis multinivel y la fórmula de transformación. Como ejemplo, en la Figura 9 se observa el procedimiento utilizado para la transformación de la variable «Zmeasure_MATE».

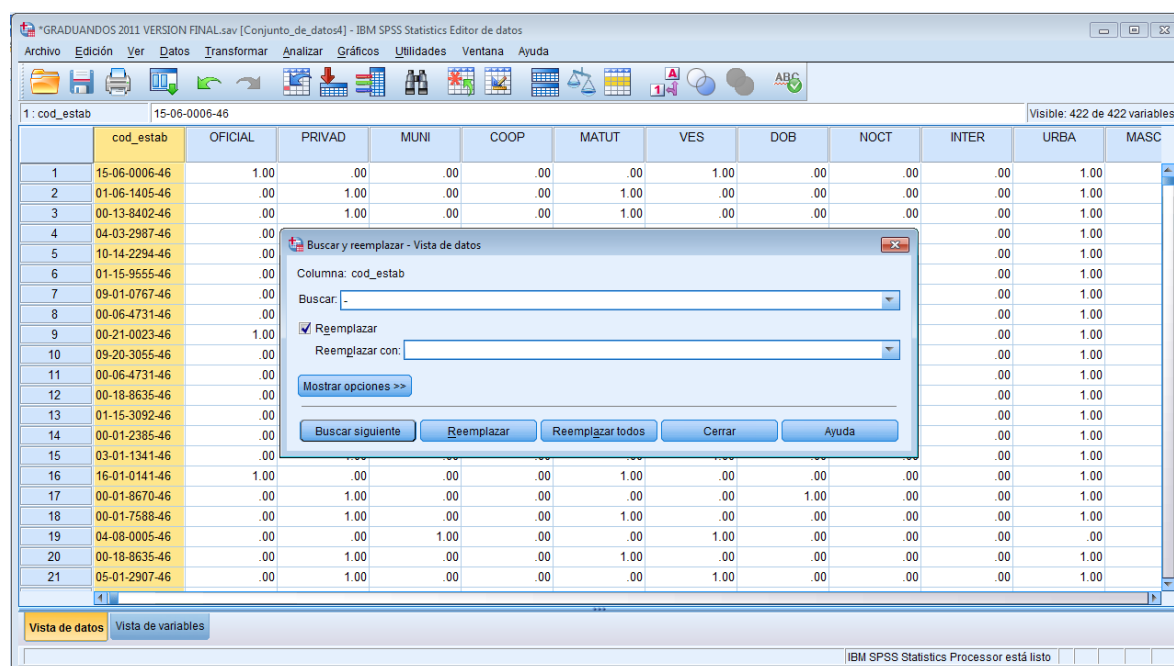
Figura 9. Ubicación de herramienta en SPSS para transformar las variables de habilidad latente en Matemática y Lectura con media 500 y desviación estándar 100



Paso 8. Base del Nivel 1. Transformar el código de establecimiento. Transformar el código de establecimiento a variable numérica. Para ello considere:

- a) Eliminar los guiones en la variable de establecimiento de la base original (de 15-06-0006-46 a 1506000646) y posteriormente transformar la variable de cadena a numérica. En versiones más actuales de SPSS la eliminación del guion se puede realizar en la ventana de vista de datos reemplazando el “-” por “ ”, como se observa en la Figura 10.

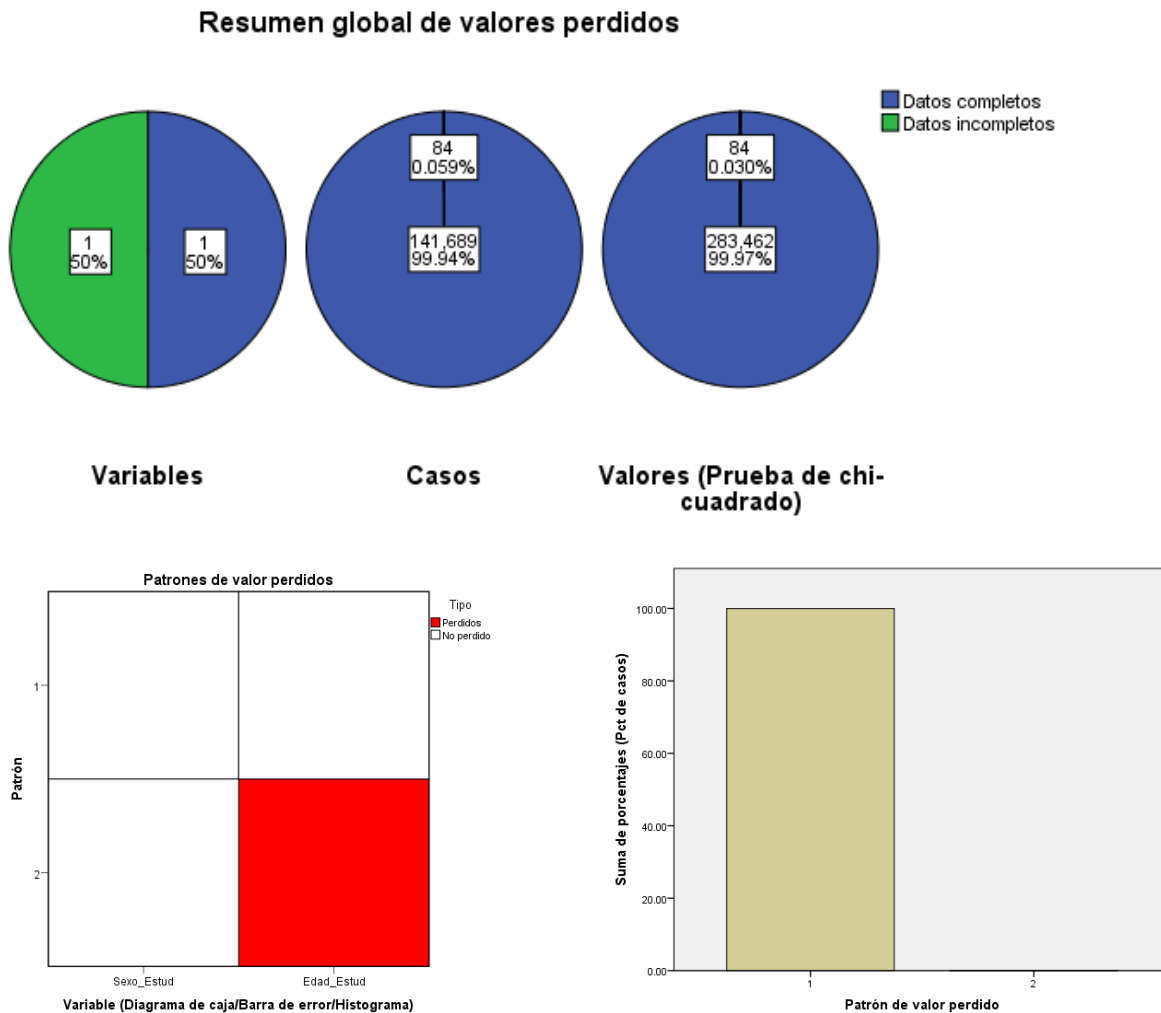
Figura 10. Procedimiento opcional para reemplazar el signo “-” en el código de establecimiento



Paso 9. Base del Nivel 1. Analizar la omisión de respuesta e imputación de variables. Las variables de interés deben pasar por un proceso de análisis de omisión de respuesta. Previo a la imputación debe explorarse las variables con análisis de frecuencias con el propósito de conocer los estadísticos de tendencia central y dispersión. También pueden realizarse otras pruebas para conocer cómo se comportan las variables de interés tales como comparación de medias, correlación y regresión. Como resultado se obtendrá una visión más amplia del comportamiento de las variables que interesan analizar.

Utilizando SPSS pueden explorarse los patrones de datos perdidos de las variables identificando la distribución de valores perdidos según variable. Se sugiere realizar este ejercicio de dos en dos variables. En la Figura 11 se muestra como ejemplo la sintaxis para explorar los patrones de datos perdidos de las variables sexo y edad de la base de Graduandos 2013. Como resultado del análisis se obtendrá una apreciación de patrones según variables, según casos y según valores (prueba de chi-cuadrado) y gráficos relacionados.

Figura 11. Ejemplo de análisis de patrones de datos perdidos



Sintaxis:

```
MULTIPLE IMPUTATION Edad_Estud Sexo_Estud
/IMPUTE METHOD=NONE
/MISSINGSUMMARIES OVERALL VARIABLES (MAXVARS=25 MINPCTMISSING=10) PATTERNS.
```

En el ejemplo (Figura 11) una de las dos variables posee datos perdidos. En la variable edad, 84 casos poseen datos perdidos y que en estos casos se registró un total de 84 datos perdidos. En la gráfica de patrones se observa un patrón donde los sujetos respondieron ambas variables y otro donde respondieron la variable sexo pero no la variable edad. En la gráfica de barras se observa la distribución del porcentaje de datos perdidos según el patrón de valores perdidos.

Seguidamente se podrá generar una tabla de contingencia para obtener mayores elementos de análisis. En la Figura 12 se observan los resultados de una tabla de contingencia entre las variables de edad y sexo. El paquete no generó una tabla de contingencia debido a que las variables no superan el 5 % de datos perdidos. Se confirma que la variable edad posee 84 datos perdidos y se observa la distribución de frecuencias en cada patrón.

Figura 12. Ejemplo de análisis de patrones de datos perdidos en tablas de contingencia

Advertencia

1

No hay ninguna variable con un 5% o más de sus valores perdidos. No se generarán las tablas CROSSTAB.

Estadísticos univariados

	N	Perdidos	
		Recuento	Porcentaje
Edad_Estud	141689	84	.1
Sexo_Estud	141773	0	.0

Patrones tabulados

Número de casos	Patrones perdidos ^a		Completo si... ^b	Edad_Estud ^c						Sexo_Estud ^c	
	Sexo_Estud	Edad_Estud		16 AÑOS O MENOS	17 AÑOS	18 AÑOS	19 AÑOS	20 AÑOS	MAYOR DE 20 AÑOS	FEMENINO	MASCULINO
141689			141689	7424	29453	39885	23211	13661	28055	70494	71195

Los patrones con menos del 1% de los casos (1418 o menos) no se muestran.

a. Las variables se ordenan según los patrones perdidos.

b. Número de casos completos si las variables perdidas en ese patrón (marcado con X) no se utilizan.

c. Distribución de frecuencias en cada patrón único

Sintaxis:

```
MVA VARIABLES= Edad_Estud Sexo_Estud
/MAXCAT=25
/CATEGORICAL=Edad_Estud Sexo_Estud
/CROSSTAB PERCENT=5
/TPATTERN PERCENT=1 DESCRIBE=Edad_Estud Sexo_Estud.
```

Es relevante determinar el porcentaje de datos perdidos y observar la distribución de datos en las variables de interés ya que las variables que exceden el 5 % de datos perdidos y que en sus distribuciones se observa uniformidad, podrían indicar que los datos perdidos no son aleatorios y su ausencia se debe a otra razón. En este contexto es necesario regresar a las variables con las cuales se podrían relacionar para determinar si los datos perdidos se deben a preguntas condicionadas en las que no se tomó en cuenta la respuesta de una variable anterior. Por ejemplo, se descubre que una variable excedió el 5 % de datos perdidos (jornada en la que trabaja) ya que no se tomó en cuenta que en una pregunta anterior el sujeto respondió negando la condición o característica (en la variable trabaja se respondió que no trabaja). En este caso, es necesario recodificar nuevamente la variable jornada en la que trabaja agregando una categoría que indique que no se responde la pregunta porque los sujetos indicaron que no trabajan. De esta cuenta, el porcentaje de datos perdidos disminuirá.

Para la imputación de datos, se solicita trabajar como se indicó anteriormente, con dos o un máximo de tres variables que tienen datos perdidos. Se aplicará una imputación múltiple y posteriormente analizará si los datos imputados transformaron las variables originales. Interesa que la distribución de casos sea lo más parecida a la distribución original. En la Figura 13 se muestra la sintaxis de imputación múltiple de sexo y edad de la base de Graduandos 2013. Se aplicó una regresión lineal utilizando como predictor ambas variables. Como resultado fue imputado por un método de imputación monotonal la variable edad ya que en la variable sexo no se encontró datos perdidos. Se imputaron un total de 84 datos. En los estadísticos descriptivos se observó que el porcentaje de las variables imputadas no cambiaron la distribución original.

Este procedimiento se reproduce en todas las variables que se utilizarán para el análisis, garantizando que las imputaciones no transforman la distribución original. Pueden utilizarse las variables imputadas como predictores de variables a imputar.

Figura 13. Ejemplo de imputación de datos perdidos

Especificaciones de imputación	
Método de imputación	Automático
Número de imputaciones	1
Modelo para variables de escala	Regresión lineal
Interacciones incluidas en modelos	(ninguna)
Porcentaje máximo de valores perdidos	100.0%
Número máximo de parámetros del modelo de imputación.	100

Restricciones de imputación		
	Papel en imputación	
	Dependiente (Regresión logística)	Predictor
Edad del estudiante (Recodificada en intervalos)	Sí	Sí
Género del estudiante	No	Sí

VALORES IMPUTADOS

Resultados de imputación		
Método de imputación		Monotonal
Iteraciones de método de especificación totalmente		n/a
Variables dependientes	Imputado	Edad_Estud
	No imputado (demasiados valores perdidos)	
	No imputado (sin valores perdidos)	Sexo_Estud
Secuencia de imputación		Sexo_Estud, Edad_Estud

Modelos de imputación				
	Modelo (Fiabilidad)		Valores perdidos	Valores imputados
	Tipo	Efectos		
Edad del estudiante (Recodificada en intervalos)	Regresión logística	Sexo_Estud	84	84

Estadísticos descriptivos

Edad_Estud				
Datos	Imputación	Categoría	N	Porcentaje
Datos originales		1	7424	5.2
		2	29453	20.8
		3	39885	28.1
		4	23211	16.4
		5	13661	9.6
		6	28055	19.8
Valores imputados	1	1	4	4.8
		2	20	23.8
		3	18	21.4
		4	10	11.9
		5	6	7.1
		6	26	31.0
Datos completos después de la imputación	1	1	7428	5.2
		2	29473	20.8
		3	39903	28.1
		4	23221	16.4
		5	13667	9.6
		6	28081	19.8

Sintaxis

```

DATASET ACTIVATE Conjunto_de_datos1.
DATASET CLOSE Base_Datos_1.
DATASET DECLARE Base_Datos_1.
MULTIPLE IMPUTATION Edad_Estud Sexo_Estud
/IMPUTE METHOD=AUTO NIMPUTATIONS=1 MAXPCTMISSING=NONE
/CONSTRAINTS Sexo_Estud( ROLE=IND)
/MISSINGSUMMARIES NONE
/IMPUTATIONSUMMARIES MODELS DESCRIPTIVES
/OUTFILE IMPUTATIONS=Base_Datos_1 .

```

El analista deberá garantizar que las variables de interés poseen en la base de datos la medida adecuada, indicación del tipo, anchura y decimales según el tipo de variable, que se encuentran plenamente identificadas y que han sido exploradas.

Entre otros criterios, debe recordar: las variables de habilidad latente no se imputarán, que es necesario determinar con la Coordinación de área lo que se hará con los estudiantes que no indicaron forma en la prueba, que incluirá en el modelo las variables requeridas según los lineamientos institucionales y otras que se consideren relevantes como resultado del análisis teórico y comportamiento en el análisis descriptivo.

Debido a que el proceso de imputación genera bases de datos nuevas al imputar las variables, deberá guardar de forma ordenada y sistemática el registro y archivos generados por el *software*.

Paso 10. Base del Nivel 1. Ordenar según clave. La base final del Nivel 1 deberá ser ordenada de forma ascendente según la clave que el programa utilizará para unir ambos niveles en el análisis multinivel: el código de establecimiento.

Paso 11. Base del Nivel 1. Identificar variables de interés y guardar base. La base se guardará con las variables de interés en este nivel. Se pide al analista guardar una copia de este archivo que se utilice únicamente como referencia y utilizar un duplicado en el análisis multinivel. Como ejemplo, la base del Nivel 1 de Graduandos 2011 se integró por los códigos de identificación de establecimiento (numérica), código de estudiante (cadena), las variables de interés imputadas, el/los indicador/es socioeconómico y de capital cultural así como las variables de habilidad latente de Matemática y Lectura (finales).

Paso 12. Base del Nivel 1. Explorar la relación y predicción de las variables de interés en la habilidad latente en Matemática y Lectura. Como un elemento que aporta al análisis continuo de las variables, el analista explorará el comportamiento de las variables de interés en cuanto a su relación e impacto en las variables de habilidad latente del estudiante. Este ejercicio será de utilidad en todo el proceso de análisis por lo que se sugiere al analista crear y guardar un archivo con los coeficientes de correlación y determinación obtenidos aplicando un método de correlación de Pearson (bilateral) y regresión lineal. Con estos resultados, el analista obtiene una apreciación primaria que le dará elementos de análisis sobre el comportamiento e interacción entre estas variables.

Paso 13. Base del Nivel 1. Analizar y tomar decisiones. Si al llegar al punto de análisis previo, alguna variable obtiene coeficientes no significativos o se comporta de forma que lleve a la duda (tomando como referente el marco teórico existente), el analista, en consulta con la Coordinación de área, puede decidir mantener esta variable para probarla en el modelo multinivel o eliminarla del análisis.

Paso 13.1. Base del Nivel 1. En caso de modificación, guardar nuevamente la base. La decisión de modificación derivará en la revisión y actualización de la base del Nivel 1. En caso de modificar la base, se pide guardar una copia de la base final y trabajar con un duplicado en el análisis multinivel.

Paso 14. Base del Nivel 2. Generar proporción, promedios e identificación de variables a nivel del establecimiento. Este paso tiene lugar en el caso de un modelo que busque analizar las variables en un nivel composicional.

- Como política utilizada en los análisis de Factores Asociados en los diferentes niveles, las variables «características del sector educativo» como área, sector, jornada y plan, se generan desde la base del estudiante –Nivel 1– (una decisión diferente debe consultarse con la Coordinación de área).
- Promedio del establecimiento de la habilidad latente de los estudiantes en Matemática y Lectura: En este paso el analista debe considerar incluir las variables de habilidad latente de los estudiantes en ambas áreas evaluadas, debido a que aunque incluyen en el modelo multinivel (desde la base del Nivel 2), se utilizarán para explorar las variables observadas (coeficiente de correlación, determinación e interacción entre variables).

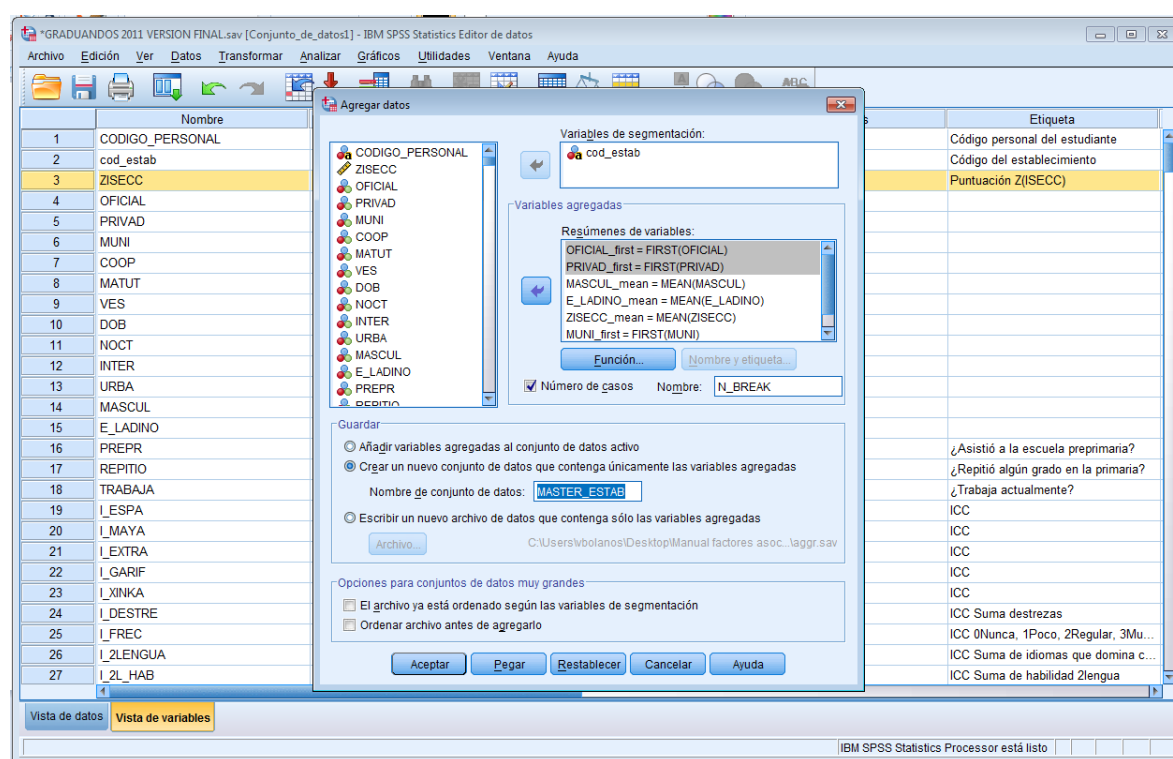
Para llevar a cabo este procedimiento, el analista utiliza la base del Nivel 1 para generar un máster de establecimientos con la proporción y promedios de las variables que son objeto de análisis en el modelo y que caracterizan la varianza de estudiantes dentro del establecimiento. Para ello se utiliza la herramienta de SPSS en el menú de Datos/Agregar datos (Figura 14). Se introduce el código de establecimiento como variable de segmentación y las variables del modelo indicando el estadístico de resumen: Media para variables de proporciones como sexo, autoidentificación étnica ladina, idioma español y para las variables

continuas, como el indicador/es socioeconómico/cultural. Se indica al *software* que obtenga el primer valor para las variables nominales como sector, jornada, plan, etc. El *software* generará una base de establecimientos (en el ejemplo, denominada MASTER_ESTAB).

Estas variables se fundirán posteriormente con la base del Nivel 2 donde se encontrarán las variables observadas en la base del cuestionario del director. No obstante, es necesario garantizar:

- Las variables que caracterizan al establecimiento como sector, jornada y área, son variables dummy: ejemplo, la variable del sector privado codificada 1 = privado, 0 = otro sector; la variable de jornada matutina codificada como 1 = matutina, 0 = otra jornada, etc.
- Las variables que resumen la composición del establecimiento (de proporción y continuas) como el sexo, identificación étnica ladino, idioma materno español, estudió en preprimaria, repitió, estudiante trabaja, indicador/es socioeconómico/cultural, pueden conservar la escala obtenida del proceso.

Figura 14. Procedimiento en SPSS para obtener la proporción y los promedios de las variables observadas a nivel del establecimiento



Paso 15. Base del Nivel 2. Paso «ubicar la base del director y las variables a utilizar». La Coordinación de área entregará al analista la base del director. El analista identificará las variables que se observarán según el modelo a replicar o bien, identificará todas las variables

de interés para el análisis. Es importante recordar que el analista identificó estas variables en el paso de diseño del modelo que fue de conocimiento de la Coordinación de área y Subdirección.

Paso 16. Base del Nivel 2. Garantizar la calidad de variables a utilizar. El analista debe explorar la base de datos para garantizar:

- Existe un código único para cada establecimiento.
- Las variables a utilizar se encuentran limpias y codificadas según los criterios de análisis que requiere el modelo. En caso contrario, recodificará y transformará las variables.
- Las variables de interés para indicadores son resultado de la aplicación de criterios de elaboración de indicadores (ver el *Manual de elaboración del Nivel Socioeconómico y otros indicadores*).
- Como ejemplo, en el análisis de Factores Asociados de Graduandos 2011, se generó un indicador de actividades de Matemática y un indicador de actividades de Lectura.
- Todas las variables a utilizar en el modelo son numéricas.

Para mayor información sobre los procesos asociados, consulte los del pasos 3 al 5 de esta sección.

Paso 17. Base del Nivel 2. Transformar el código de establecimiento. El código de establecimiento será la clave con la que el programa establecerá la relación entre los dos niveles de análisis. De igual manera que en el Nivel 1, es necesario transformar el código de establecimiento a variable numérica y garantizar que posee las mismas características que la variable de establecimiento del Nivel 1. (Para mayor información sobre este proceso, consulte el paso 8 de esta sección).

Paso 18. Base del Nivel 2. Analizar la omisión de respuesta e imputación de variables. De igual manera que en la base de datos del Nivel 1, las variables de interés de este nivel deben pasar por un proceso de análisis de omisión de respuesta. Para imputar las variables deben tomarse en cuenta los siguientes criterios:

- a) Realizar una exploración de todas las variables (observar estadísticos de tendencia central y dispersión).
- b) Las variables a imputar no debieran exceder el 40 % de los datos perdidos. Las variables que excedan este porcentaje deben considerar eliminarse del modelo (no obstante, este extremo debe consultarse con la Coordinación de área).

- c) Puede utilizarse el método de imputación de regresión con uno o múltiples predictores. La decisión debe hacerse de conocimiento a la Coordinación de área.

Para mayor información sobre los procesos asociados, consulte el paso 9 de esta sección.

El analista debe observar el comportamiento final de estas variables y verificar sus resultados.

Paso 19. Base del Nivel 2. Ordenar según clave. De igual manera que en el Nivel 1, la base del Nivel 2 deberá ser ordenada de forma ascendente según el código de establecimiento que será la clave que el programa utiliza para unir ambos niveles en el análisis multinivel.

Paso 20. Base del Nivel 2. Fundir base de proporción, promedios e identificación de variables a nivel del establecimiento con la base del Nivel 2. El analista fundirá las variables de la base del Nivel 2 a la base generada en el paso 14 de esta sección (en el ejemplo, denominada MASTER_ESTAB). Es necesario que para fundir esta base se garantice que:

- La base de proporción, promedios e identificación de variables del nivel de establecimiento (en el ejemplo, denominada MASTER_ESTAB), se encuentra abierta y es la base en donde se agregarán las variables.
- Se elija la opción «El que no es un conjunto de datos activo es una tabla de claves» en el menú de añadir variables cuando se emparejen los casos en las variables clave para los archivos ordenados.
- Se verifique el procedimiento.

Lo anterior debido a que en la base del cuestionario del director podrían participar directores de establecimientos que no existen en la base del Nivel 1 debido a que en ese establecimiento no se evaluaron estudiantes de último grado de ciclo diversificado (Graduandos).

Paso 21. Base del Nivel 2. Identificar variables de interés y guardar base. El analista garantizará que en la base del Nivel 2 se encuentran las siguientes variables: código de establecimiento, las variables que caracterizan el establecimiento (área, sector, jornada), variables de interés del Nivel 2 (del cuestionario del director), variables de proporción y promedio de los estudiantes en el establecimiento educativo e indicadores. Estas variables cumplen con los criterios de calidad para el análisis (verificación de códigos únicos, codificación según el modelo, variables limpias, recodificadas, transformadas e imputadas). El analista podrá guardar dos versiones de esta base: una con las variables antes y después de imputar y otra base (que será la base final) únicamente con las variables imputadas. La base final se guardará y un duplicado se utilizará en el modelo multinivel.

Paso 22. Base del Nivel 2. Explorar la relación y predicción de las variables de interés en la habilidad latente en Matemática y Lectura. De igual manera que en el Nivel 1, el analista realizará la exploración de las variables en cuanto a su relación e impacto con la habilidad latente del estudiante (promedios de habilidad de los estudiantes en Matemática y Lectura por establecimiento, obtenido en el paso 14 de esta sección).

En virtud de su importancia para el análisis de todo el proceso, se solicita al analista crear y guardar un archivo con los coeficientes de correlación y determinación obtenidos aplicando un método de correlación de Pearson (bilateral) y regresión lineal.

Paso 23. Base del Nivel 2. Analizar y tomar decisiones. De igual manera que en el Nivel 1, si al llegar al punto de análisis previo, alguna variable obtiene coeficientes no significativos o se comporta de forma que lleve a la duda (tomando como referente el marco teórico existente), el analista, en consulta con la Coordinación de área, puede decidir mantener esta variable para probarla en el modelo multinivel o eliminarla del análisis.

Paso 23.1. Base del Nivel 2. En caso de modificación, guardar nuevamente la base. De igual manera que en el Nivel 1, la decisión de modificación derivará en la revisión y actualización de la base del Nivel 2. En caso de modificar la base, se sugiere guardar una copia de la base final y trabajar con un duplicado en el análisis multinivel.

5.4. Análisis y resultados del modelo

Responsable de elaboración: analista

Responsable de supervisión, verificación y control: Coordinación de área

Software: HLM, SPSS, EXCEL

Productos:

- Output/s de resultados del modelo de Análisis
- Cuadro resumen de la interacción final de variables en el modelo

Productos intermedios:

- Variables analizadas en función de los coeficientes e indicadores estadísticos
- Registros informáticos del análisis de variables en el modelo

En esta fase se trabajará con las bases de datos finales de los niveles 1 y 2 que fueron resultado de la fase anterior. El analista alcanzará los siguientes objetivos:

- a) Creará un espacio en el ordenador donde ubicará las bases finales de los niveles 1 y 2, el análisis y los resultados.

- b) Cargará las bases de los niveles 1 y 2 en el *software* (HLM) e introducirá las variables para el análisis según del modelo establecido.
- c) Observará los resultados e interacción de las variables introducidas de forma sistemática.
- d) Analizará los resultados en función de los indicadores del modelo, coeficientes e interacción de las variables.
- e) Experimentará de forma análoga buscando el mejor ajuste del modelo.
- f) Analizará sus resultados con la Coordinación del área.
- g) Con el aval de la Coordinación de área, tomará decisiones sobre estos resultados.
- h) Ejecutará en el *software* el modelo final y archivará los resultados.
- i) Desarrollará un cuadro resumen de los resultados finales.
- j) Presentará a la Coordinación del área los resultados finales.

Paso 1. Crear un espacio en el ordenador. Como paso fundamental para el análisis en HLM, el analista deberá crear un espacio (carpeta) en el ordenador donde ubicará las bases finales de los niveles 1 y 2, los resultados del *software* y el análisis. Para realizar esta tarea deberá considerar:

- Garantizar que las bases de los niveles 1 y 2 son los archivos finales.
- Que las bases de los niveles 1 y 2 cumplen con el criterio que la clave a utilizar (código del establecimiento) posee las mismas características en ambas bases de datos, es numérica y se encuentra ordenada de forma ascendente.
- Que todas las variables que se analizarán son numéricas. La única variable de cadena es el código personal en la base del Nivel 1 que permanece como referencia pero que no incluye en el análisis ni se carga en el *software*.

Paso 2. Cargar las bases de los niveles 1 y 2 en HLM e introducir variables del modelo establecido. El analista cargará las bases de los niveles 1 y 2 en el *software* en HLM. Para realizar esta tarea, deberá considerar:

- Que la carga de variables en el *software* se ejecuta de forma sistemática.
- Garantiza la selección de todas las variables de interés en su totalidad.

Paso 3. Procedimiento científico: Experimentación. El analista ejecuta un proceso de análisis científico de las variables que introduce de forma sistemática. Observa los resultados e interacción y en función del análisis prueba las variables de interés del modelo. El analista puede validar con la Coordinación del área los resultados del *software* para enriquecer el análisis y proceso de experimentación. Como resultado de este paso, posee una apreciación holística del comportamiento de las variables en el modelo establecido.

Paso 4. Analizar los resultados en función de los indicadores del modelo, coeficientes e interacción de variables. Los resultados obtenidos durante el proceso de experimentación del paso 3, se complementan con el análisis simultáneo de los indicadores del modelo (DEVIANZA, CINTRA, coeficiente de determinación, varianza y desviación estándar de los niveles 1 y 2, indicador de confiabilidad del modelo), el intercepto del modelo, los coeficientes obtenidos (signo, desviación estándar y p-value). Y el análisis de modelo establecido con relación al modelo Nulo. De igual manera que el paso anterior, el analista puede validar con la Coordinación del área los resultados del *software* para enriquecer el análisis y proceso de experimentación. Como resultado de los pasos 3 y 4, el analista obtendrá los resultados del modelo que se diseñó para el análisis de Factores Asociados y una apreciación holística del modelo en su conjunto.

Paso 5. Procedimiento científico: experimentación en la búsqueda de mejor ajuste del modelo. Consecutivamente, durante el proceso de toma de decisiones, el analista experimentará de forma análoga buscando el mejor ajuste del modelo. Como corolario del análisis cuantitativo y referente teórico existente, tomará decisión sobre el modelo y las variables que lo componen, buscando el mejor ajuste del modelo.

Tomará en cuenta la teoría para el análisis jerárquico multinivel. Como ejemplo de interpretación, se cita textualmente lo que indica el informe de análisis de Factores Asociados de Graduandos 2011 acerca de la interpretación de coeficientes:

- **Correlación intraclase (CINTRA)** o efecto de la escuela en la variable dependiente observada (habilidad estimada en Matemática / Lectura). Este coeficiente se interpreta como un indicador de las desigualdades de aprendizaje entre escuelas y al interior de las escuelas. En el modelo, un coeficiente alto indica que el porcentaje de desigualdad de aprendizajes se debe a la variabilidad o desigualdad de los establecimientos educativos. Un coeficiente bajo sugiere que los resultados de Matemática o de Lectura no se deben tanto a las diferencias entre establecimientos sino a las diferencias entre estudiantes. En otras palabras, los estudiantes acceden a establecimientos de similar calidad educativa, por lo que sus resultados no dependen del tipo de educación que brinda el establecimiento sino de la variación que se produce como resultado del esfuerzo que cada estudiante invierte en su proceso de aprendizaje.
- Como resultado también es posible observar el coeficiente de **DEVIANZA** que puede entenderse como un indicador de bondad de ajuste. En el modelo Nulo, este coeficiente toma un valor que mientras más grande existe se da un peor ajuste del

modelo. En los siguientes modelos, este indicador idealmente debiera disminuirse por la presencia de variables que permiten un mejor ajuste del modelo.

- **Intercepto:** como ya se indicó, este indicador está representado por y_{00} . Una forma de interpretarlo es la recta de promedios generada por la presencia de todas las rectas de las variables observadas en los modelos diseñados.
- **Varianza residual del estudiante (Nivel 1):** significa la varianza que existe dentro de la escuela o en otras palabras, entre los estudiantes. Las diferencias que se observan se deben a las diferencias entre las realidades de los estudiantes.
- **Varianza residual de la escuela (Nivel 2):** significa la varianza que existe entre las escuelas.

Paso 6. Analizar los resultados con la Coordinación del área. El analista presentará a la Coordinación del área los resultados cuantitativos, apreciaciones y análisis sobre el modelo. Presenta el modelo Final y en caso de variantes del modelo establecido, expondrá el fundamento científico de la eliminación o inclusión de otras variables de interés. Los fundamentos pueden enmarcarse en el marco teórico existente, análisis cuantitativo y resultados obtenidos del modelo.

Paso 7. Con el aval de la Coordinación del área, tomar decisión sobre los resultados. Producto del análisis colegiado, con el aval de la Coordinación, el analista tomará decisiones sobre el modelo.

Paso 7.1 Ejecutar en el *software* el modelo final y archivar los resultados. En caso de cambios, ejecutará el *software* el modelo final. En caso de existir cambios o no, archivará los resultados del modelo.

Como ejemplo, se transcribe el modelo utilizado en el *Informe de Factores Asociados al aprendizaje de Graduandos 2008* (Dirección General de Evaluación e Investigación Educativa, Digeduca, 2008):

Modelo Nulo o Vacío. Este modelo se estableció con el objetivo expreso de responder a la pregunta uno, a través de estimar el porcentaje de la varianza total, del rendimiento escolar, asociada a las diferencias individuales de los estudiantes y a las diferencias entre escuelas. Adicionalmente permitió, determinar si el intercepto es una variable aleatoria o fija y sirvió de base para determinar el porcentaje de la varianza explicada debido a las variables que integran los modelos posteriores. Este modelo es el más sencillo de los cinco modelos desarrollados; carece de variables explicativas, asumiendo que la varianza de los puntajes del rendimiento escolar es provocada por una parte de la varianza total atribuida a las diferencias del estudiante y otra atribuida a los establecimientos escolares. Dicho modelo se representa de la siguiente forma:

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

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

Donde el sub índice i se refiere al resultado de los estudiantes dentro de las escuelas, mientras que j representa a las escuelas; y_{ij} es el resultado del rendimiento escolar para el estudiante i dentro de la escuela j; β_{0j} es la media del rendimiento en el escuela j; e_{ij} es el error aleatorio asociado a β_{0j} el cual es una variable aleatoria de distribución normal, con media igual a cero y varianza igual a σ^2 , homogénea para todos los estudiantes. γ_{00} es la gran media y u_{ij} es el error asociado al segundo nivel cuando β_{0j} es aleatorio.

Modelo del sector escolar. Este modelo tuvo como objetivo fundamental determinar el efecto que tiene el sector al que pertenece el establecimiento escolar. El sector escolar estuvo compuesto por establecimientos públicos, establecimientos privados, establecimientos municipales y establecimientos por cooperativa. Con esto se contemplaron tres variables indicadoras (dummy), siendo el sector público el valor de referencia en el modelo. En este modelo el intercepto es considerado como un parámetro aleatorio que depende de un conjunto de variables del segundo nivel. Este modelo se representa así:

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

$$\beta_{0j} = \gamma_{00} + u_{01} Z_{Privado,j} + u_{02} Z_{Municipal,j} + u_{03} Z_{Cooperativa,j} + u_{0j}$$

Donde β_{0j} es la media del rendimiento en el escuela j, ajustada al sector escolar; $Z_{Privado,j}$ es una variable indicadora que toma el valor de 1 si el establecimiento j es privado y 0 si no; $Z_{Municipal,j}$ es una variable indicadora que toma el valor de 1 si el establecimiento j es municipal y 0 si no; $Z_{Cooperativa,j}$ es una variable indicadora que toma el valor de 1 si el establecimiento j es por cooperativa y 0 si no. Los establecimientos públicos fueron tomados como referencia. Representa el intercepto de la media global β_{0j} y representa la influencia de los establecimientos públicos sobre el rendimiento escolar. γ_{01} , γ_{02} y γ_{03} , son los coeficientes de regresión de β_{0j} asociados a las variables del sector escolar y representan el cambio promedio entre los establecimientos públicos, que son el valor de referencias y cada uno de los establecimientos que integran el sector escolar; u_{0j} es el error asociado al segundo nivel del modelo, el cual tienen una distribución normal con media igual a cero y varianza σ^2 y es independiente a través de las escuelas y los alumnos.

Modelo Estructural. Este modelo adiciona, al modelo anterior, las variables propias de los establecimiento escolares que no dependen de los estudiantes. Dichas variables se

suman solamente al segundo nivel del modelo el cual se representa de la siguiente manera:

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

$$\begin{aligned} \beta_{0j} = & \gamma_{00} + u_{01} Z_{Privado,j} + u_{02} Z_{Municipal,j} + u_{03} Z_{Cooperativa,j} + u_{04} Z_{Cooperativa,j} + u_{05} \\ & W_{Matutina,j} + u_{06} W_{Vespertina,j} + u_{07} W_{Intermedia,j} + u_{08} W_{Urbana,j} + u_{09} W_{Lab\ Computaci3n,j} + u_{010} \\ & W_{Educaci3n\ Director,j} + u_{011} W_{Experiencia\ Director,j} + u_{012} W_{G3nero\ Director,j} + u_{013} W_{IAA,j} + u_{014} W_{Minutos\ de} \\ & \text{clases al mes},j + u_{0j} \end{aligned}$$

Donde $W_{Matutina}$ es una variable indicadora que toma el valor de 1 si el establecimiento j funciona en jornada matutina y 0 si no; $W_{Vespertina}$ es una variable indicadora que toma el valor de 1 si el establecimiento j funciona en jornada vespertina y 0 si no; W_{Doble} es una variable indicadora que toma el valor de 1 si el establecimiento j funciona en jornada doble y 0 si no; $W_{Intermedia}$ es una variable indicador que toma el valor de 1 si el establecimiento j funciona en jornada doble y 0 si no; los establecimientos de jornada nocturna fueron tomados como valor de referencia. W_{Urbana} es una variable indicadora que toma el valor de 1 si el establecimiento se encuentra en el 3rea urbana y 0 si no; los establecimientos del 3rea rural tomaron el valor de referencia. $W_{Lab\ Computaci3n}$ es una variable indicara que toma el valor de 1 si en el establecimiento se cuenta con laboratorio de computaci3n y 0 si no; los establecimientos que no tienen laboratorio de computaci3n tomaron el valor de referencia. $W_{Educaci3n\ Director}$ es una variable moderadora que toma el valor de 1 si el director del establecimiento tiene al menos licenciatura y 0 si no.

Los establecimientos que tienen un director con un nivel educativo m3s bajo que licenciatura tomaron el valor de referencia. $W_{Experiencia\ Director}$ es una variable que toma los a3os de experiencia que tienen el director, la experiencia puede ser como director o como docente. $W_{G3nero\ Director}$ (tomando como g3nero el sexo masculino o femenino) es una variable moderadora que toma el valor de 1 si el director del establecimiento j es hombre y 0 si es mujer; los establecimientos con directores de sexo femenino tomaron el valor de referencia. W_{IAA} es una variable que representa el 3ndice de actividades de aprendizaje de la asignatura, Matem3tica o Lectura seg3n sea el caso, a nivel del establecimiento. $W_{Minutos\ de\ clase\ al\ mes}$ es una variable que representa la cantidad de minutos que los alumnos reciben clases de la asignatura, Matem3tica o Lectura, durante un mes. El resto de variables se defini3 en el modelo anterior.

Los γ_{04} , γ_{05} y $\gamma_{06}, \dots, \gamma_{012}$, γ_{013} y γ_{014} , los coeficientes de regresi3n de β_{0j} .

Dichos coeficientes representan la influencia en las medias de aprendizaje de las variables estructurales del segundo nivel. El resto de coeficientes ya fueron definidos en el modelo anterior.

Modelo Composicional. La construcción de este modelo ayudó a determinar la influencia de las características composicionales de los establecimientos sobre el rendimiento escolar, así como también su contribución en la varianza explicada. Las variables composicionales incluyeron aquellas variables agregadas de estudiante, tales como promedios y proporciones a nivel de escuela, provenientes de un atributo del alumno. Este modelo agrega, al sector escolar y las variables estructurales de los establecimientos escolares, las variables composicionales al segundo nivel de modelo, ya que también son factores condicionales a nivel de los establecimientos educativos. El modelo quedó representado así:

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

$$\beta_{0j} = \gamma_{00} + u_{01} Z_{Privado,j} + u_{02} Z_{Municipal,j} + u_{03} Z_{Cooperativa,j} + u_{04} Z_{Cooperativa,j} + u_{05} W_{Matutina,j} + u_{06} W_{Vespertina,j} + u_{07} W_{Intermedia,j} + u_{08} W_{Urbana,j} + u_{09} W_{Lab Computación,j} + u_{010} W_{Educación Director,j} + u_{011} W_{Experiencia Director,j} + u_{012} W_{Género Director,j} + u_{013} W_{IAA,j} + u_{014} W_{Minutos de clases al mes,j} + u_{015} V_{MasculinoE,j} + u_{016} V_{LadinoE,j} + u_{017} V_{EspañolE,j} + u_{018} V_{PreescolarE,j} + u_{019} V_{RepitióE,j} + u_{020} V_{TrabajaE,j} + u_{021} V_{ICC,j} + u_{022} V_{IPSE,j} + u_{0j}$$

Donde $V_{masculinoE}$, $V_{LadinoE}$, $V_{EspañolE}$, $V_{PreescolarE}$, $V_{RepitióE}$, $V_{TrabajaE}$, V_{ICC} y V_{IPSE} , representan las variable composicionales de los establecimientos que se definen de la siguiente forma:

Proporción de estudiantes hombre, proporción de estudiantes autoidentificados como ladinos, proporción de estudiantes cuyo idioma materno es el español, proporción de estudiantes que asistió al preescolar, proporción de estudiantes que repitió al menos un grado en primaria, proporción de estudiantes que trabaja, promedio de índice de capital cultura y promedio de índice de posición socioeconómico.

Los γ_{015} , γ_{016} ,....., γ_{021} , γ_{022} , son los coeficientes de regresión de β_{0j} , asociadas a las variables agregadas del alumno por establecimiento educativo. El resto de variables y coeficientes ya fueron definidos.

Modelo de las características del estudiante. Este modelo incorpora ocho variables individuales del estudiante al Nivel 1 del modelo y queda definido de la siguiente forma:

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{\text{Género } ij} + \beta_{2j}X_{\text{Ladino } ij} + \beta_{3j}X_{\text{Español } ij} + \beta_{4j}X_{\text{Preescolar } ij} + \beta_{5j}X_{\text{Repitió } ij} + \beta_{6j}X_{\text{Trabaja } ij} + \beta_{7j}X_{\text{ICC } ij} + \beta_{8j}X_{\text{IPSE } ij} + e_{ij}$$

$$\begin{aligned} \beta_{0j} = & \gamma_{00} + u_{01}Z_{\text{Privado},j} + u_{02}Z_{\text{Municipal},j} + u_{03}Z_{\text{Cooperativa},j} + u_{04}Z_{\text{Cooperativa},j} + u_{05} \\ & W_{\text{Matutina},j} + u_{06}W_{\text{Vespertina},j} + u_{07}W_{\text{Intermedia},j} + u_{08}W_{\text{Urbana},j} + u_{09}W_{\text{Lab Computación},j} + u_{010} \\ & W_{\text{Educación Director},j} + u_{011}W_{\text{Experiencia Director},j} + u_{012}W_{\text{Género Director},j} + u_{013}W_{\text{IAA},j} + u_{014}W_{\text{Minutos de}} \\ & \text{clases al mes},j + u_{015}V_{\text{MasculinoE},j} + u_{016}V_{\text{LadinoE},j} + u_{017}V_{\text{EspañolE},j} + u_{018}V_{\text{PreescolarE},j} + u_{019} \\ & V_{\text{RepitióE},j} + u_{020}V_{\text{TrabajaE},j} + u_{021}V_{\text{ICC},j} + u_{022}V_{\text{IPSE},j} + u_{0j} \end{aligned}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

$$\beta_{4j} = \gamma_{40}$$

$$\beta_{5j} = \gamma_{50}$$

$$\beta_{6j} = \gamma_{60}$$

$$\beta_{7j} = \gamma_{70}$$

$$\beta_{8j} = \gamma_{80}$$

Donde $X_{\text{Género}}$ representa una variable moderadora que toma el valor 1 si el estudiante es de sexo masculino y 0 si no, el sexo femenino tiene el valor de referencia. X_{Ladino} representa una variable moderadora que toma el valor de 1 si el estudiante se autoidentifica como ladino y 0 sino, el valor de referencia lo tiene el grupo que se autoidentifica como no ladino, en cual se incluyen mayas, garífunas, xinkas y otros. $X_{\text{Español}}$ es una variable moderadora que toma el valor 1 si el idioma materno del estudiante es español y 0 si no, el grupo que no tiene español como idioma materno tiene el valor de referencia. $X_{\text{Preescolar}}$ es una variable moderadora que toma el valor 1 si el estudiante asistió a preescolar y 0 si no, el grupo que no asistió a preescolar tiene el valor de referencia. X_{Repitio} es una variable moderadora que toma el valor 1 si el estudiante repitió algún grado de primaria y 0 si no, el grupo que no repitió tiene el valor de referencia. X_{Trabaja} es una variable moderadora que toma el valor 1 si el estudiante trabaja y 0 si no, el grupo que no trabaja tiene el valor de referencia. X_{ICC} es el índice de capital cultural de los estudiantes y es una variable continua. X_{IPSE} es el índice socioeconómico de los estudiantes y es una variable continua. En este modelo el coeficiente β_{0j} representa la media del rendimiento en la escuela j , a partir de las variables explicativas del alumno y de la escuela incluyendo las variables agregadas del alumno. Los coeficientes β_{1j} , β_{2j} , β_{3j} , β_{4j} , β_{5j} , β_{6j} , β_{7j} y β_{8j} son los coeficientes de regresión

del Nivel 1 del modelo y están asociados a las variables descritas arriba. El resto de variables ya fueron definidas en los modelos anteriores; es importante mencionar que estos coeficientes en el modelo se establecieron como efectos fijos.

Paso 8. Diseñar un cuadro resumen de los resultados finales. El analista diseñará un cuadro resumen en el que registrará los resultados finales del modelo establecido. Tomar en cuenta lo que se indicó en el diseño de modelo de regresión múltiple multinivel, que los modelos pueden ser diferentes en función del grado para el que se elabore el informe, si se trata de una réplica o en virtud del objetivo que se fijó para la elaboración del informe o hipótesis y pregunta de investigación. En la Figura 15 se encontrará un ejemplo del modelo de Factores Asociados al aprendizaje aplicado a primaria 2010 y en la Figura 16 un ejemplo del modelo aplicado en Graduandos 2011.

Figura 15. Ejemplo de cuadro resumen del modelo de Factores Asociados al aprendizaje de Matemática de primaria 2010

	VARIABLE	MODELO NULO		MODELO ESTRUCTURAL		MODELO COMPOSICIONAL		MODELO FINAL	
		Coeficiente	SE	Coeficiente	SE	Coeficiente	SE	Coeficiente	SE
	Intercepto (Measure Mate)	503.92**	2.93	503.89**	2.83	503.91**	2.71	398.55**	14.52
NIVEL ESTABLECIMIENTO	Área			12.84*	5.96	3.70	6.39	1.87	6.75
	Género del Director			-11.36*	5.80	-5.19	5.86	-2.80	5.97
	Nivel Educativo Director			3.85	6.10	4.90	6.07	5.27	6.14
	Capacitaciones del Director			1.89	1.33	1.85	1.23	2.20	1.27
	Períodos de Matemática			4.82	4.50	4.83	4.52	5.82	4.62
	Género del Docente			-20.67**	6.22	-15.20**	6.11	-18.25**	6.09
	Nivel Educativo Docente			43.73**	10.81	38.29**	10.21	36.17**	10.76
	Capacitaciones del Docente			3.10	6.12	-3.15	5.71	-4.99	5.71
	Cantidad de Estudiantes en 1ro Primaria					0.02	0.05	0.02	0.06
	Proporción de Hombres en Primero Primaria					-8.45	17.92	-18.81	17.85
	Proporción de Ladinos en Primero Primaria					39.65**	9.62	28.24**	10.54
	Proporción de los que Repiten en Primero Primaria					-23.43*	7.99	-21.80**	8.52
	¿Estudiaste pre-primaria antes de entrar a primer grado?					-8.41	11.94	-16.99	12.46
	Proporción de estudiantes hablan Español					-16.39	11.50	-33.23**	12.06
	Proporción de estudiantes dicen maestro deja tareas					122.09*	56.78	112.95*	56.84
	Proporción de estudiantes dicen maestro devuelve tareas revisadas					35.01	29.72	20.36	30.09
	Promedio Escuela Nivel Sociocultural					14.06**	5.69	8.24	5.60
NIVEL ESTUDIANTE	Género del estudiante							3.76*	1.84
	Etnia Ladino							12.66**	4.62
	Edad del estudiante							5.50**	1.23
	Número de veces que has repetido este grado							0.84	1.42
	Idioma Español							8.87*	4.26
	Nivel Sociocultural							5.92**	0.93
	¿Tu maestro te deja tareas para hacer en tu casa?							-5.20	7.77
	¿Tu maestro te devuelve tus trabajos ya revisados?							2.93	5.84
	¿Estudiaste pre-primaria antes de entrar a primer grado?							11.02**	2.44
	Recursos Matemática							2.17**	0.49
	¿Te gusta leer?							31.97**	4.62
Componentes de la Varianza		Varianza	DE	Varianza	DE	Varianza	DE	Varianza	DE
Escuela		5329.83	73.01	5016.65	70.83	4637.30	68.10	4437.52	66.61
Estudiante		4719.25	68.70	4719.55	68.70	4719.50	68.70	4083.92	63.91

Figura 16. Ejemplo de cuadro resumen del modelo de Factores Asociados al aprendizaje de Matemática de Graduandos 2011

MATEMÁTICAS		MODELO NULO		MODELO SECTOR		MODELO ESTRUCTURAL		MODELO COMPOSICIONAL		MODELO DEL ESTUDIANTE	
		Coficiente	SE	Coficiente	SE	Coficiente	SE	Coficiente	SE	Coficiente	SE
Intercepto		494.09**	1.06	494.21**	1.06	493.91**	0.81	494.19**	0.62	492.04**	1.37
Establecimiento del sector Privado	1 = Privado, 0 = Otro			7.96**	2.24	4.38	2.35	-8.71**	1.81	-8.72**	1.81
Establecimiento del sector por Cooperativa	1 = Cooperativa, 0 = Otro			-19.21**	3.89	-11.20*	4.26	-5.77	3.27	-5.77	3.28
Establecimiento del sector Municipal	1 = Municipal, 0 = Otro			-15.11*	5.82	-5.44	6.91	8.01	4.67	8.01	4.68
Jornada Matutina	1 = Matutina, 0 = Otra					25.07**	6.38	0.70	5.17	0.92	5.19
Jornada Vespertina	1 = Vespertina, 0 = Otra					5.41	6.23	1.52	5.10	1.65	5.13
Jornada Doble	1 = Doble, 0 = Otra					5.62	6.26	3.64	5.04	3.72	5.07
Jornada Intermedia	1 = Intermedia, 0 = Otra					10.55	10.38	12.44	8.53	12.67	8.54
Área: Urbana	1 = Urbano, 0 = Rural					9.64**	2.70	-0.91	2.18	-0.88	2.17
Género del director: Masculino	1 = Masculino, 0 = Femenino					-7.19**	1.71	-0.29	1.33	-0.35	1.33
Nivel educativo del Director: Licenciatura	1 = Licenciatura, 0 = Menor					5.40**	1.63	0.88	1.24	0.91	1.24
Experiencia del Director	A > años de experiencia					0.10	0.10	-0.10	0.08	-0.10	0.08
Laboratorio de computación	1 = SI, 0 = NO					17.73**	2.34	7.02**	1.81	7.03**	1.81
Tiempo de enseñanza al mes	A > cantidad de periodos y minutos al mes					4.04**	0.19	1.95**	0.16	1.94**	0.16
Índice Actividades de Matemáticas y Lectura						-1.12	0.83	0.39	0.64	0.39	0.64
Proporción de Hombres								37.18**	3.33	1.32	3.40
Proporción de Ladinos								-27.85**	3.38	-35.02**	3.47
Proporción de Idioma Materno: Español								-11.35*	4.98	-6.10	5.12
Proporción asistencia preescolar								-10.07*	4.45	-2.83	4.51
Proporción de Repitentes								-73.03**	5.04	-45.03**	5.05
Proporción de trabajan								-9.47*	3.62	-3.29	3.66
Promedio de Indicador Socioeconómico y de Capital Cultural (ISECC)								47.10**	1.86	29.35**	1.91
Género del estudiante: Masculino	1 = Masculino, 0 = Femenino									35.84**	0.71
Identidad étnica del estudiante: Ladino	1 = Ladino, 0 = Otro									7.47**	0.79
Idioma materno del estudiante: Español	1 = SI, 0 = Otro									-5.23**	1.15
¿Asistió a escuela preprimaria?	1 = SI, 0 = NO									-7.04**	0.66
¿Repitió algún grado en primaria?	1 = SI, 0 = NO									-27.42**	0.62
¿Trabaja actualmente?	1 = SI, 0 = NO									-6.24**	0.59
Indicador Socioeconómico y de Capital Cultural (ISECC)										17.58**	0.45
		Varianza	DE	Varianza	DE	Varianza	DE	Varianza	DE	Varianza	DE
(N2) Escuela		2880.28	53.67	2842.76	53.32	1569.42	39.62	803.82	28.35	827.31	28.76
(N1) Estudiante		6932.75	83.26	6932.90	83.26	6933.76	83.27	6934.78	83.28	6319.63	79.50
CINTRA		29.35		29.08		18.46		10.39		11.58	
DEVIANZA		1444436.38		1444385.39		1442917.01		1441440.91		1430252.72	
R2 NIVEL ESCUELA				1.30		44.79		48.78		-2.92	
R2 NIVEL ESTUDIANTE				0.00		-0.01		-0.01		8.87	

*P value < 0.05

**P value < 0.001

Paso 9 Presentar a la Coordinación de área los resultados finales. Los modelos finales de Matemática y Lectura (sea el caso), se presentarán a la Coordinación de área. Estos cuadros serán incluidos en el informe de Factores Asociados al aprendizaje.

5.5. Elaboración del informe

Responsable de elaboración: analista

Responsable de supervisión, verificación, control y revisión del documento: Coordinación de área, Subdirección

Productos:

- Redacción del informe final
- Revisión del informe y solicitud de correcciones (Coordinación de área y Subdirección)
- Corrección del informe
- Revisión final y entrega a la Coordinación de Divulgación

En esta fase el analista elaborará el informe de Factores Asociados con el formato discutido y aprobado previamente por la Coordinación de área y Subdirección. Alcanzará los siguientes objetivos:

Objetivo general

Redacción del informe final

Objetivos intermedios

Finalizar la elaboración del marco teórico

Redactar la sección de procedimiento y las consideraciones técnicas

Redactar la sección del modelo utilizado

Redactar la sección de resultados

Redactar la sección de discusión

Redactar la sección de conclusiones / implicaciones en política educativa

Agregar al informe resumen, índice de tablas, índice de figuras, referencias y otros que se consideren pertinentes

Las secciones de marco teórico, procedimiento y consideraciones técnicas y del modelo utilizado que han sido objeto de elaboración durante las fases iniciales, son concluidas en un momento previo a la elaboración de los demás resultados. Otros criterios específicos de elaboración del informe, estarán dados por las coordinaciones de área y Subdirección.

Revisión del informe y solicitud de correcciones (Coordinación de área y Subdirección). El analista concluirá el informe y lo entregará a la Coordinación de área y Subdirección quienes revisarán y solicitarán cambios pertinentes.

Corrección del informe. El analista tendrá la oportunidad de validar con la Coordinación de área y Subdirección, puntos que se consideren necesarios validar previo a la modificación. Posteriormente realizará los cambios solicitados y devolverá el informe a la Coordinación de área y Subdirección.

Revisión final y entrega a la Coordinación de Divulgación. La Coordinación de área y Subdirección de Análisis revisarán el informe y cuando se considere oportuno, entregarán a la Coordinación de Divulgación.

6. Flujogramas

A continuación se presentan los siguientes flujogramas:

- Equipo responsable
- Políticas procedimentales
- Procedimiento de preparación técnica para base del Nivel 1
- Procedimiento de preparación técnica para base del Nivel 2
- Procedimiento de análisis y resultados del modelo
- Elaboración del informe

Figura 17. Flujograma de equipo responsable

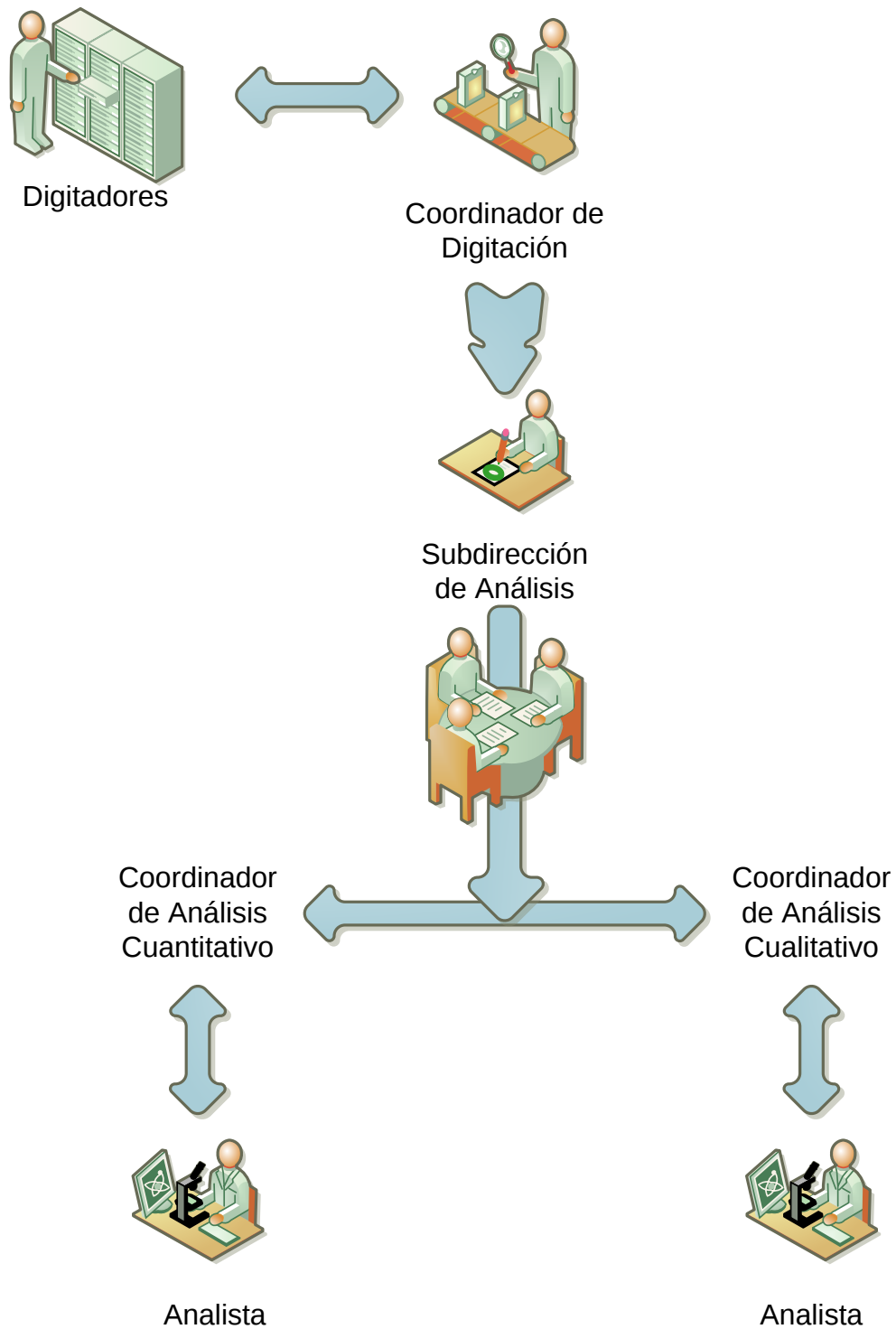


Figura 18. Flujograma de políticas procedimentales

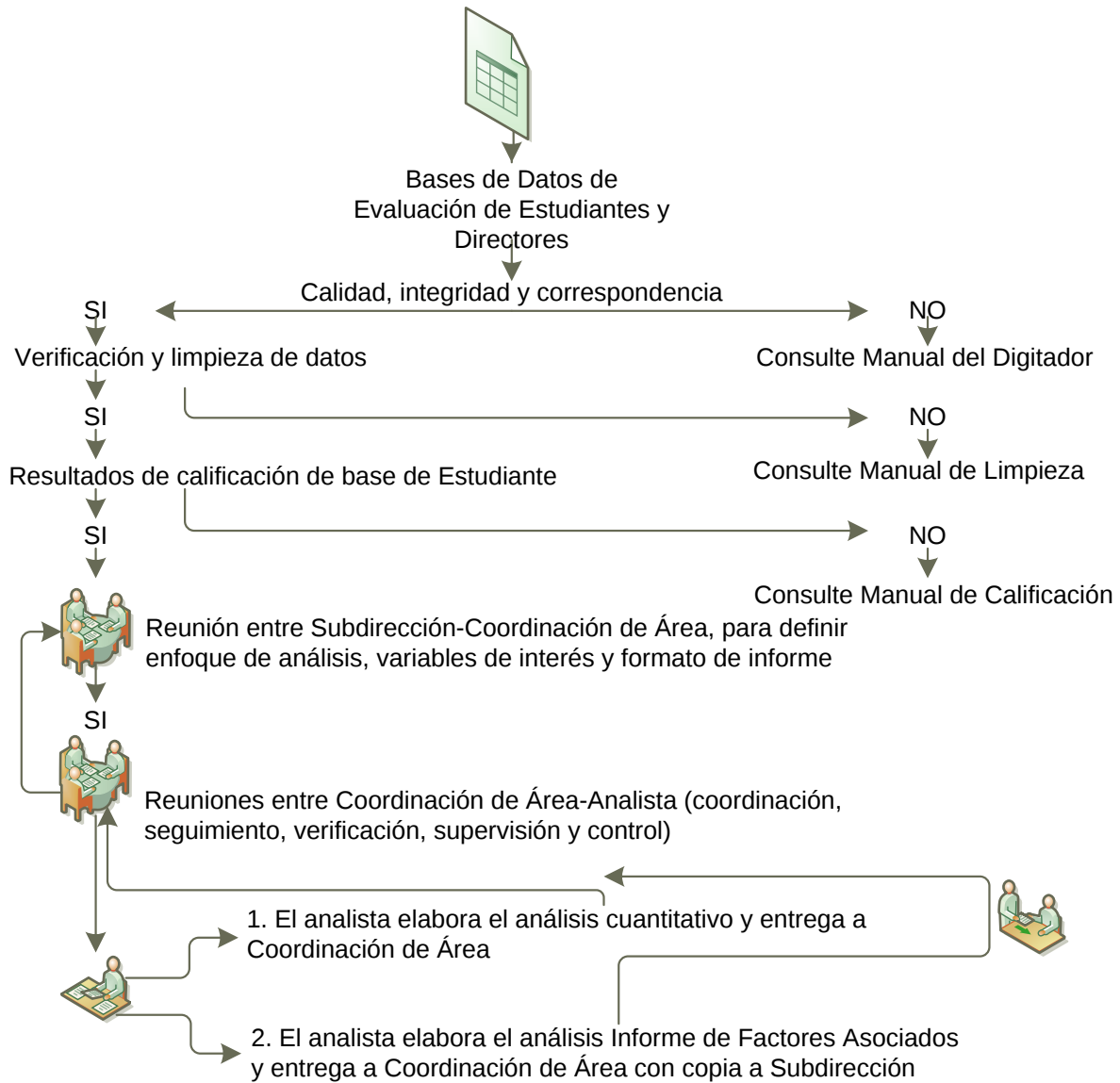


Figura 19. Flujograma de procedimiento de preparación técnica para la base del Nivel 1

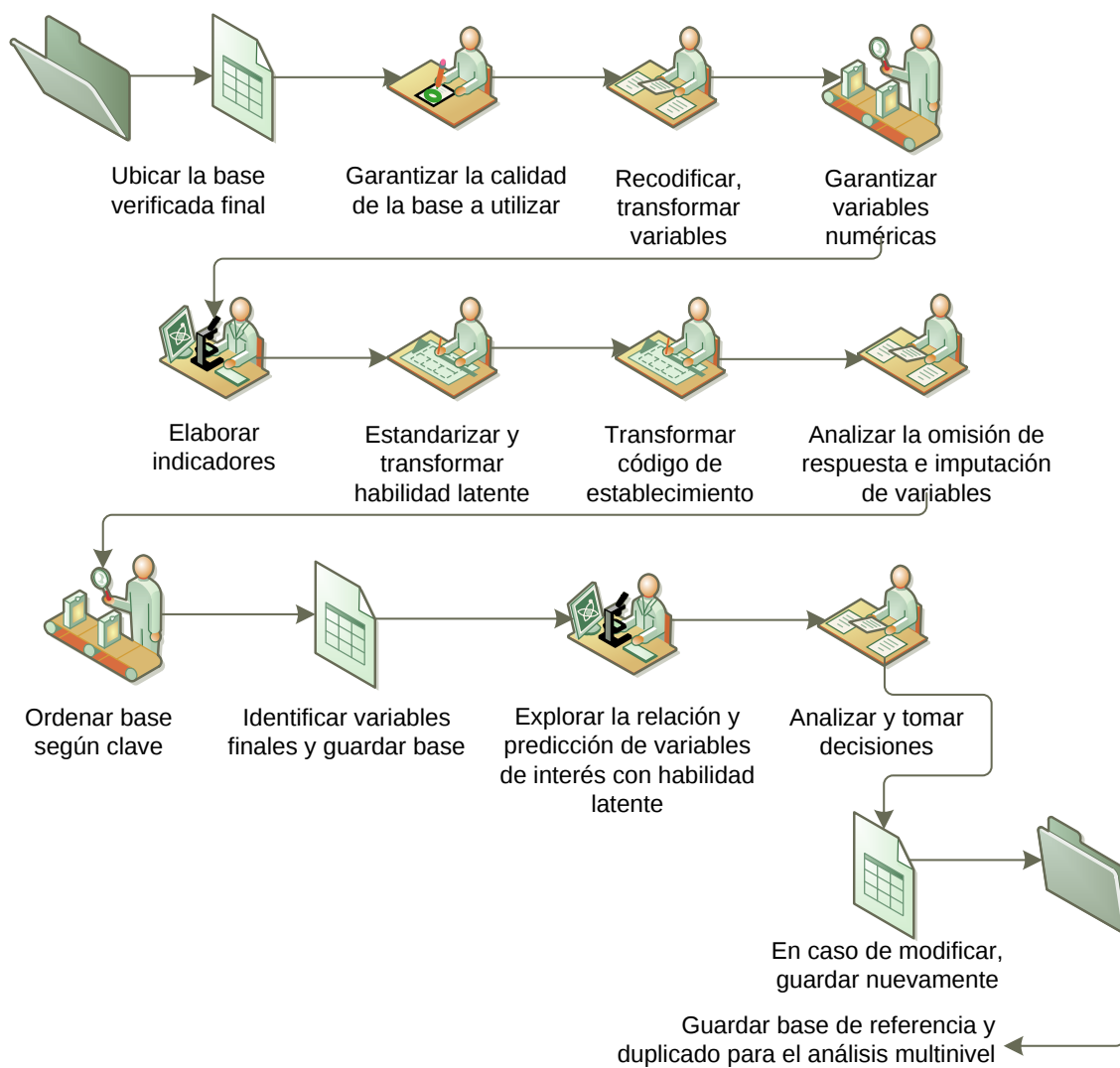


Figura 20. Flujograma de procedimiento de preparación técnica para la base del Nivel 2

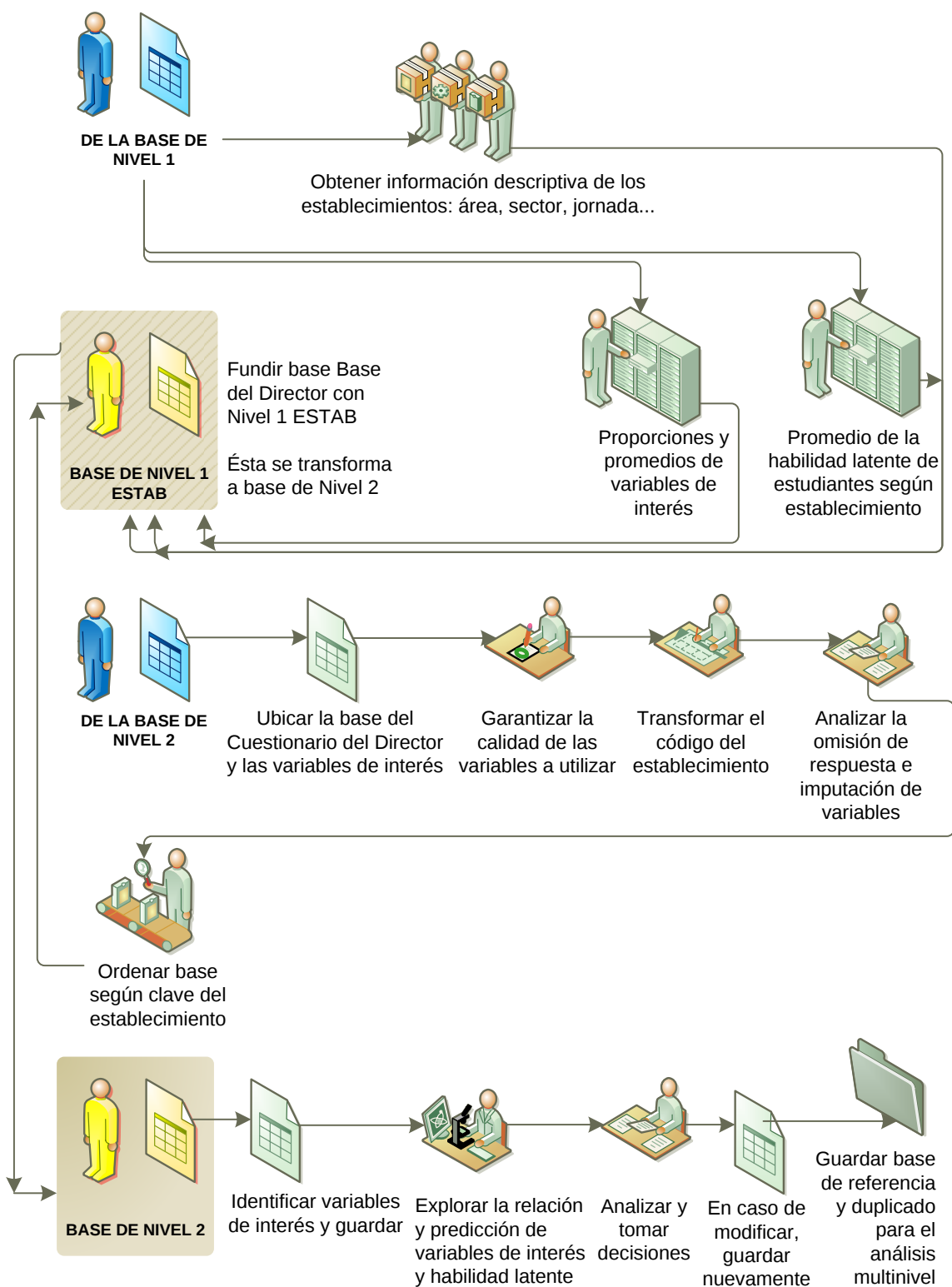


Figura 21. Flujograma de procedimiento de análisis y resultados del modelo

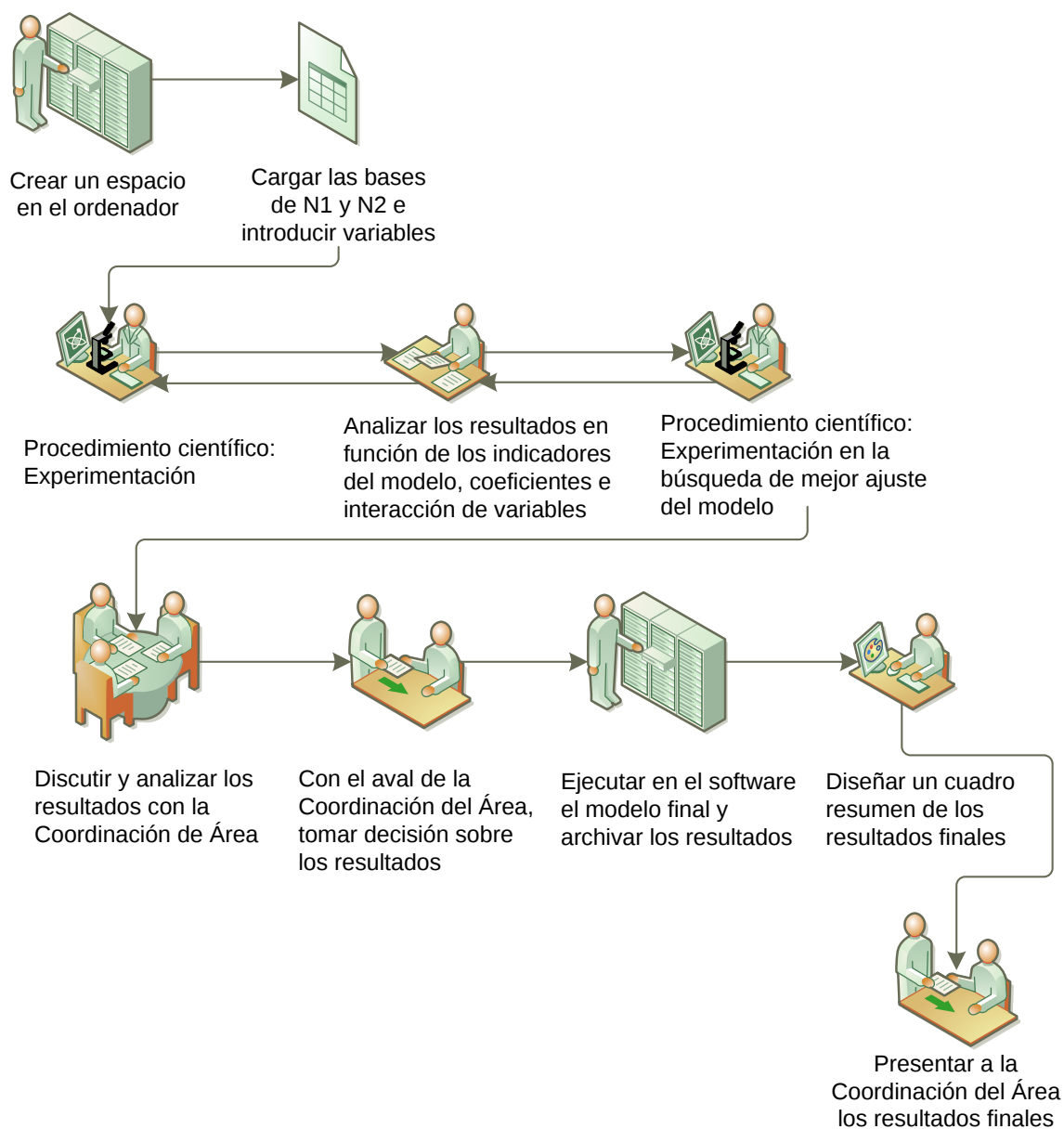
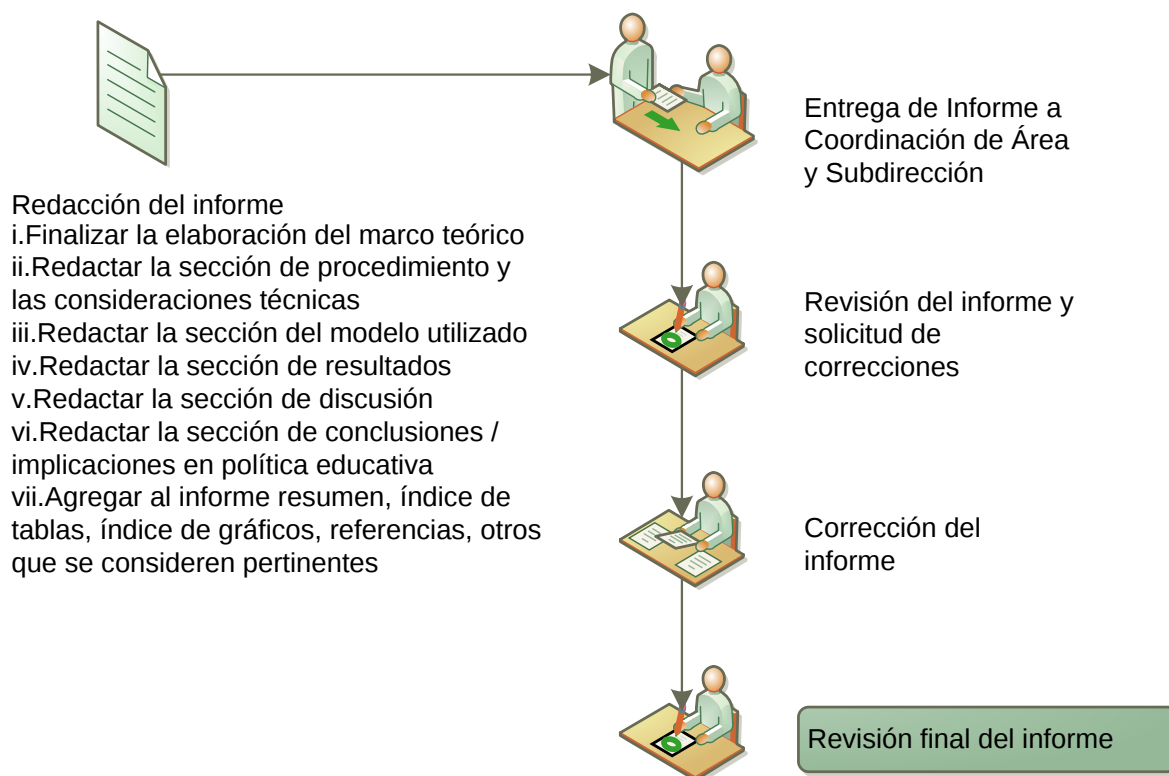


Figura 22. Flujograma de procedimiento de elaboración del informe



Glosario



Bondad de ajuste: es un indicador que permite discernir acerca de qué tan buena es la ecuación obtenida. Para determinar la bondad de un ajuste se utilizan diferentes criterios en la regresión lineal. Unos se refieren a los residuales como son el valor de la sumatoria de residuales al cuadrado, la varianza, la desviación estándar del ajuste y el coeficiente de correlación al cuadrado. Otro indicador de la bondad de ajuste es el realizado mediante el test de bondad de ajuste utilizando la prueba Ji-Cuadrada (X^2), Kolgomorov -Smirnov (K-S) entre otras (Instituto Nacional de Estadística e Informática, INEI, 2006).

Capital cultural: quisiera hoy evocar los mecanismos extremadamente complejos a través de los cuales la institución escolar contribuye (insisto en esta palabra) a reproducir la distribución del capital cultural, y con ello, a la reproducción de la estructura del espacio social. A las dos dimensiones fundamentales de este espacio, del que hablaba ayer, corresponden dos conjuntos de mecanismos de reproducción diferentes. La combinación de ambos mecanismos define el modo de reproducción y hace que el capital vaya al capital y que la estructura social tienda a perpetuarse (no sin sufrir deformaciones más o menos importantes). La reproducción de la estructura de la distribución del capital cultural se opera en la relación entre las estrategias de las familias y la lógica específica de la institución escolar. Esta tiende a proporcionar el capital escolar, que otorga bajo la forma de títulos (credenciales), al capital cultural detentado por la familia y transmitido por una educación difusa o explícita en el curso de la primera educación. El sistema escolar actúa a la manera del demon de Maxwell: al precio del gasto de energía que es necesario para realizar la operación de selección mantiene el orden preexistente, es decir, la separación entre los alumnos dotados de cantidades desiguales —o de tipos diferentes— de capital cultural. Más precisamente, mediante toda una serie de operaciones de selección, separa a los detentores de capital cultural heredado de los que están desprovistos de él. Como las diferencias de aptitud son inseparables de las diferencias sociales según el capital heredado, el sistema escolar tiende a mantener las diferencias sociales preexistentes (Bourdieu, 1998).

Codificación y etiquetación de las bases de datos: es necesario volver a codificar las variables de acuerdo al tipo de análisis que se realizará con cada una de ellas. Se hace nuevamente el análisis de frecuencia y datos descriptivos para asegurar que se realizará la transformación adecuada a los códigos que se necesitan. Se deben etiquetar todas y cada una de las variables para que las bases de datos contengan la información del significado de su contenido e interpretación en la codificación. Las etiquetas son de dos tipos: etiqueta de variable y etiqueta de valor. La etiqueta de variable se refiere a la descripción o información para reconocer de qué pregunta o ítem son los datos. La etiqueta de valor se refiere a la información del dato contenido como código en la variable. Método de verificación: todas las variables deben tener etiquetas que identifican de qué pregunta o ítem fueron obtenidas (ver

Figura 3). Las etiquetas de valor deben ser las que se presentan en las opciones de respuesta de las preguntas o la codificación necesaria para el análisis específico que se realizará con las bases de datos (ver Figura 4) (Digeduca, 2014).

Coeficientes de regresión: son los valores constantes de una ecuación de regresión lineal. En el modelo de regresión lineal siguiente los coeficientes son a y b (Instituto Nacional de Estadística e Informática, INEI, 2006).

Correlación: cuando dos fenómenos sociales, físicos o biológicos crecen o decrecen de forma simultánea y proporcional debido a factores externos, se dice que los fenómenos están positivamente correlacionados. Si uno crece en la misma proporción que el otro decrece, los dos fenómenos están negativamente correlacionados. El grado de correlación se calcula aplicando un coeficiente de correlación a los datos de ambos fenómenos (Rodríguez, 2002).

Correlación intraclass (CINTRA) o efecto de la escuela en la variable dependiente observada (habilidad estimada en Matemática / Lectura). Este coeficiente se interpreta como un indicador de las desigualdades de aprendizaje entre escuelas y al interior de las escuelas. En el modelo, un coeficiente alto indica que el porcentaje de desigualdad de aprendizajes se debe a la variabilidad o desigualdad de los establecimientos educativos. Un coeficiente bajo sugiere que los resultados de Matemática o de Lectura no se deben tanto a las diferencias entre establecimientos sino a las diferencias entre estudiantes. En otras palabras, los estudiantes acceden a establecimientos de similar calidad educativa, por lo que sus resultados no dependen del tipo de educación que brinda el establecimiento sino de la variación que se produce como resultado del esfuerzo que cada estudiante invierte en su proceso de aprendizaje (Digeduca, 2014).

DEVIANZA: puede entenderse como un Indicador de bondad de ajuste. En el modelo Nulo este coeficiente toma un valor que mientras más grande, existe un peor ajuste del modelo. En los siguientes modelos, este indicador idealmente debiera disminuirse por la presencia de variables que permiten un mejor ajuste del modelo (Digeduca, 2014).

Distribución normal o curva normal: llamada también como distribución de Gauss, es la distribución de probabilidad más utilizada en estadística y teoría de probabilidad. Esto se debe a dos razones: su función de densidad es simétrica y con forma de campana lo que favorece su aplicación como modelo a gran número de variables. Es además, límite de otras distribuciones y aparece relacionada con resultados ligados a la teoría de las probabilidades gracias a sus propiedades matemáticas (Instituto Nacional de Estadística e Informática, INEI, 2006).

Factores Asociados al aprendizaje: por ello es importante que el uso de los Factores Asociados y las Oportunidades de Aprendizaje, esté limitado a aquellos en los cuales el Ministerio de Educación tiene incidencia, es decir, en los escolares, pues son áreas en las que se pueden promover cambios (Flores, 2010).

Imputación por método de regresión: este método calcula las estimaciones de regresión lineal múltiple y ofrece opciones que permiten incrementar las estimaciones con componentes aleatorios. Para cada valor pronosticado, el procedimiento puede añadir un residuo de un caso completo seleccionado de manera aleatoria, una desviación normal aleatoria o una desviación aleatoria (escalada por la raíz cuadrada del residuo cuadrático promedio) de la distribución t. El método de regresión estima los valores perdidos utilizando la regresión lineal múltiple. Se muestran las medias, la matriz de covarianza y la matriz de correlaciones de las variables pronosticadas. Corrección de la estimación. El método de regresión puede añadir un componente aleatorio a las estimaciones de regresión. Puede seleccionar residuos, variantes normales, variantes t de *student* O sin corrección: a) Residuo. Los términos de error se eligen al azar de entre los residuos observados en los casos, completos, para añadirlos a las estimaciones de regresión; b) Variantes normales. Los términos de error se escogen al azar de una distribución con valor esperado 0 y desviación típica igual a la raíz cuadrada del termino error cuadrático medio de la regresión; c) Variantes de *student*. Los términos de error se escogen al azar de una distribución t con los grados de libertad especificados y se escalan según la raíz del error cuadrático medio (RMSE); d) Número máximo de predictores. Establece un límite máximo para el número de variables predictoras (independientes) utilizadas en el proceso de estimación; e) Guardar datos completados. Escribe un conjunto de datos en la sesión actual o en un archivo de datos externo con formato IBM® SPSS® Statistics, reemplazando los valores perdidos por los valores estimados mediante el método de regresión. Para especificar las opciones de regresión E en el cuadro de diálogo principal Análisis de valores perdidos, seleccione las variables cuyos valores perdidos desea estimar utilizando el método de regresión. E Seleccione Regresión en el grupo Estimación. E Para especificar las variables predictoras y pronosticadas, pulse en Variables. Si desea obtener más información, consulte el tema Variables pronosticadas y predictoras, p. 11. E Pulse en Regresión. E Seleccione las opciones de regresión deseadas. Por defecto, se utilizan todas las variables cuantitativas para la estimación de regresión y EM. Si es necesario, puede especificar que determinadas variables se utilicen como variables pronosticadas o variables predictoras en las estimaciones. Una determinada variable puede aparecer en ambas listas, pero hay situaciones en las que quizá quiera restringir el uso de una variable. Por ejemplo, a algunos analistas no les resulta cómodo estimar los valores de las variables de resultados. También es posible que quiera utilizar variables diferentes en estimaciones distintas y ejecutar el procedimiento varias veces. Por ejemplo, si tiene un conjunto de elementos que son valoraciones de enfermeras y otro conjunto que son valoraciones de médicos, tal vez quiera ejecutar el procedimiento una vez utilizando el elemento de las enfermeras para estimar los elementos de las enfermeras y otra vez para estimar los elementos de los médicos. También hay que hacer otra consideración al utilizar el

método de regresión. En la regresión múltiple, el uso de un subconjunto grande de variables independientes puede generar valores pronosticados de peor calidad que los que generaría un subconjunto más pequeño. Por tanto, para que se utilice una variable, debe alcanzar un límite de F para entrar de 4,0. Este límite se puede cambiar utilizando la sintaxis. Para especificar las variables pronosticadas y predictoras E. En el cuadro de diálogo principal Análisis de valores perdidos, seleccione las variables cuyos valores perdidos desea estimar utilizando el método de regresión. E Seleccione EM o Regresión en el grupo Estimación E Pulse en Variables. E Si desea utilizar determinadas variables, en vez de todas, como variables pronosticadas y variables predictoras, elija Seleccionar variables y mueva las variables a las listas adecuadas (IBM, 2011).

Indicador: un indicador es una expresión cualitativa o cuantitativa observable, que permite describir características, comportamientos o fenómenos de la realidad a través de la evolución de una variable o el establecimiento de una relación entre variables, la que comparada con períodos anteriores, productos similares o una meta o compromiso, permite evaluar el desempeño y su evolución en el tiempo. Por lo general, son fáciles de recopilar, altamente relacionados con otros datos y de los cuales se pueden sacar rápidamente conclusiones útiles y fidedignas. Un indicador debe cumplir con tres características básicas: 1. Simplificación: la realidad en la que se actúa es multidimensional, un indicador puede considerar alguna de tales dimensiones (económica, social, cultural, política, etc.), pero no puede abarcarlas todas. 2. Medición: permite comparar la situación actual de una dimensión de estudio en el tiempo o respecto a patrones establecidos. 3. Comunicación: todo indicador debe transmitir información acerca de un tema en particular para la toma de decisiones (Departamento Administrativo Nacional de Estadística, DANE, S.f.).

Índice: es la relación expresada en porcentaje entre el precio, cantidad o valor de un bien y servicio o conjunto de bienes y servicios, en un período (Instituto Nacional de Estadística e Informática, INEI, 2006).

Intercepto: este indicador está representado por y_0 . Una forma de interpretarlo es la recta de promedios generada por la presencia de todas las rectas de las variables observadas en los modelos diseñados (Digeduca, 2014).

Limpieza de las bases de datos: al encontrarse datos extraños, fuera de rango o duplicados se confirma con la Dirección de Informática (Dinfo), se revisa el material físico, se verifica la razón por la cual aparece y se corrige. Una vez revisado el contenido de las bases de datos, se ajustan las especificaciones de las variables colocando la extensión adecuada, el tipo de variable, la alineación, ancho de visualización, medida y si habrá datos que se serán excluidos en los análisis. También se determina que no se tengan variables duplicadas o vacías. Método de verificación: las variables deben tener la extensión acorde a los caracteres del dato más largo en su contenido (ver Figura 1). Además, se debe dejar un ancho de columna que sea adecuado para el reconocimiento, ya que se realiza verificación visual en la

vista de datos (ver Figura 2). Por ejemplo, variables como los ítems de la prueba no requieren mucha amplitud y al ser todos con similares características de un carácter, basta una amplitud de un espacio para visualizar su contenido (Digeduca, 2014).

Logit: escala *Logit* representados por Theta (θ) o probabilidad (de 0.5), que el estudiante posee para responder correctamente los ítems de la evaluación (Digeduca, 2014).

Modelo Jerárquico Lineal: un análisis de regresión lineal multinivel permite profundizar en el análisis de resultados de Logro, según las características particulares de los establecimientos así como de los estudiantes. De esta cuenta se tiene un análisis simultáneo que involucra información del establecimiento educativo así como del estudiante. Permite analizar el vínculo anidado entre establecimientos y estudiantes de una forma más flexible y llevar a cabo un análisis estadístico en niveles jerárquicos en donde el estudiante representa el Nivel 1 y el establecimiento el Nivel 2 (Digeduca, 2014). En suma, el Modelo Jerárquico Lineal complementa el conocimiento generado por los análisis estadísticos previos, ya que no se considera la independencia de una variable sino que toma en cuenta el sesgo generado por cada nivel, posee fortaleza en la precisión, considera la dependencia de Nivel 2 y la verosimilitud de heterogeneidad de la relación de Nivel 1 que varía a través de las unidades de Nivel 2 (Moreno, Gálvez, Morales, Saz, Arriola, Johnson & Santos, 2009).

Modelo de Análisis: modelo que se diseña para el análisis multinivel. En los informes de Graduandos 2008 y 2011, el modelo incluye además del modelo Nulo, el modelo Sector, modelo Estructural, modelo Composicional y modelo del Estudiante (Rodríguez, 2002).

Nivel Socioeconómico: el informe PISA España 2006, indica algunos hallazgos que merecen consideración en este espacio, siendo cuatro los que proceden para este estudio. Situación laboral y educativa de los padres: a mayor estatus ocupacional y escolaridad de los padres, los niños obtienen mejores resultados. Recursos económicos: a mejores condiciones del hogar y menor el hacinamiento, mejores resultados. Situación laboral de los estudiantes: los alumnos que trabajan tienen peores resultados y entre mayor es su jornada su rendimiento es más bajo. Costos de la educación: las familias que menos gastan, comúnmente están en dura situación de pobreza, lo que influye en el menor rendimiento académico de los niños (Saz, s.f.).

Normalidad: se puede hablar de normalidad en una distribución de datos, cuando las medidas de tendencia central (media, mediana y moda) coinciden en su valor. Lo cual origina gráficamente una curva simétrica, donde su eje de simetría es el punto donde coinciden las tres medidas de tendencia central (Rodríguez, 2002).

Probabilidad: es un número que se le asigna a un suceso como una medida de su incertidumbre. Este número puede tomar valores entre cero y uno inclusive. Cuando los sucesos son equiprobables, es decir todos tienen la misma probabilidad para calcularla, se utiliza la Regla de Laplace. Sea un suceso A, entonces: $P(A) = \text{Casos favorables} / \text{casos posibles}$. El experimento aleatorio debe cumplir dos requisitos: a) El número de resultados posibles (sucesos) debe ser finito. b) Todos los sucesos deben tener la misma probabilidad. A la regla de Laplace también se le denomina «probabilidad a priori», ya que para aplicarla hay que conocer antes de realizar el experimento cuáles son los posibles resultados y saber que todos tienen las mismas probabilidades (Instituto Nacional de Estadística e Informática, INEI, 2006).

Promedio: es cualquier medida de posición de tendencia central. Cuando se obtiene sumando los datos y dividiendo entre el número de ellos, se conoce como promedio simple (Instituto Nacional de Estadística e Informática, INEI, 2006).

Regresión: es una técnica de análisis para poner de manifiesto la estructura de dependencia que mejor explique el comportamiento de la variable dependiente o explicada (y) a través de un conjunto de variables independientes o explicativas ($x_1, x_2, \dots, x_p$), con las que se supone está relacionada. El método más utilizado es el de los mínimos cuadrados. La ecuación a ajustar puede ser lineal o no lineal. En ambos casos el objetivo es el mismo: encontrar las mejores estimaciones de los parámetros y cuantificarla precisión de los mismos (Instituto Nacional de Estadística e Informática, INEI, 2006).

Variable dicotómica (dummy): son aquellas que, por su propia naturaleza solo pueden manifestarse según dos modalidades (Rodríguez, 2002).

Varianza residual de la escuela (Nivel 2): significa la varianza que existe entre las escuelas (Informe Factores Asociados de Graduandos de 2011).

Varianza residual del estudiante (Nivel 1): significa la varianza que existe dentro de la escuela o en otras palabras, entre los estudiantes. Las diferencias que se observan se deben a las diferencias entre las realidades de los estudiantes (Informe Factores Asociados de Graduandos de 2011).

Verificación de las variables: cada variable es analizada y revisada para determinar que su contenido es el especificado en el libro de códigos y que no contenga valores extraños o fuera de los rangos. Se utiliza la función de frecuencias para obtener una tabla con los valores de resumen, ya sea frecuencia o estadísticos descriptivos (según el tipo de variable) de todas las variables en la base de datos, esto con el fin de comprobar su contenido. Las variables con datos únicos como nombres o identificadores, son verificadas determinando que no existan casos duplicados. Método de verificación: se realiza al tener la cantidad de casos completos de la base de datos distribuidos en las opciones de respuesta de cada pregunta específica, lo

cual se puede observar en la Tabla 1. En la Tabla 2 se observa que los datos estén entre el máximo y mínimo, la media y desviación estándar de una variable continua o de escala. Se revisa material físico de forma aleatoria y se comprueba que la información en la base de datos es la misma que el estudiante proporcionó en el instrumento de evaluación (Digeduca, 2014).

Referencias



- Bolaños, V. & Santos, J. (2013). *Reporte de los resultados de la evaluación de graduandos 2011*. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación.
- Bordieu. (1998). *Capital Cultural. Escuela y espacio social*. Madrid: España: Siglo Veintiuno de España Editores, S.A.
- Departamento Administrativo Nacional de Estadística (DANE). (S.f.). *Guía para diseño, construcción e interpretación de indicadores*. S.d.: Departamento Administrativo Nacional de Estadística (DANE).
- Digeduca. (2014). *Calificación 3ro. Básico 2013*. Guatemala: Guatemala: Digeduca, MINEDUC.
- Digeduca. (2014). *Informe de Factores Asociados al aprendizaje de Graduandos 2011*. Guatemala: Guatemala: Digeduca: MINEDUC.
- Dirección General de Evaluación e Investigación Educativa, Digeduca. (2006). *Informe graduandos 2006*. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación.
- Dirección General de Evaluación e Investigación Educativa, Digeduca. (2007). *Informe final de la evaluación nacional censal de Graduandos 2007. Logros en matemática y lectura*. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación.
- Dirección General de Evaluación e Investigación Educativa, Digeduca. (2008). *Evaluación de Graduandos 2008. Informe técnico de resultados*. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación.
- Figuerroa de Teos, D. (2010). *Calidad educativa y ampliación de la educación secundaria. Proyecto BIRF 7430-GU*. Guatemala: Dirección General de Evaluación e Investigación Educativa. Ministerio de Educación.

- Flores, M. (2010). *Variables utilizadas para el análisis de Factores Asociados al rendimiento de los estudiantes*. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación.
- IBM. (2011). *IBM SPSS Missing Values 20*. Estados Unidos: Copyright IBM Corporation 1989, 2011.
- Instituto Nacional de Estadística e Informática (INEI). (2006). *Glosario básico de términos estadísticos*. Lima: Perú: Instituto Nacional de Estadística e Informática (INEI).
- Moreno, M; Gálvez, A.; Morales, A; Saz, M.; Arriola, P.; Johnson, J. & Santos, A. (2009). *Informe técnico de Factores Asociados al rendimiento escolar de Graduandos, de acuerdo a la evaluación nacional de Lectura y Matemática 2008*. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación.
- OCDE, O. p. (s.f.). *El programa PISA de la OCDE. Qué es y para qué sirve*. Recuperado el 15 de Julio de 2013, de OECD:
<http://www.oecd.org/pisa/publicacionesdepisaenespaol.htm>
- Organisation for Economic Co-operation and Development. (2001). *Conocimientos y aptitudes para la vida. Primeros resultados del Programa Internacional de Evaluación de Estudiantes (PISA) 2000 de la OCDE*. México: Editorial Santillana, S.A. de C.V., para edición en español.
- Organización para la Cooperación y el Desarrollo Económico, OCDE. (2004). *Informe PISA 2003. Aprender para el mundo del mañana*. España: Santillana Educación, S.L.
- PIRLS & TIMSS. (2013). *TIMSS, PIRLS*. Recuperado el 2013, de <http://timssandpirls.bc.edu/>
- RAE, 2. (s.f.). *Real Academia Española*. Recuperado el Julio de 24 de 2013, de <http://www.rae.es/rae.html>
- Rodríguez, T. (2002). *Breve diccionario de estadística*. Colegio Marista Cristo Rey.
- Saz, M. (S.f.). *Influencia del nivel socioeconómico y cultural en el rendimiento de los estudiantes de tercero básico y Graduandos del año 2006*. Digeduca: Digeduca, MINEDUC.
- Subdirección de Análisis de Datos de Evaluación e Investigación Educativa. (2011). *Informe de evaluación de graduandos 2009*. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación.
- Subdirección de Análisis de Datos e Investigación Educativa, Digeduca. (2011). *Informe final de la evaluación a Graduandos del año 2005*. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación.

TIMSS. (2001). *Effective Schools in Science and Mathematics*. Trends in International Mathematics and Science Study (TIMSS).

TIMSS. (2011). *International Results in Mathematics*.TIMSS.

Serie de Cuadernillos Técnicos



Manual de procedimiento para el análisis de factores asociados al aprendizaje

