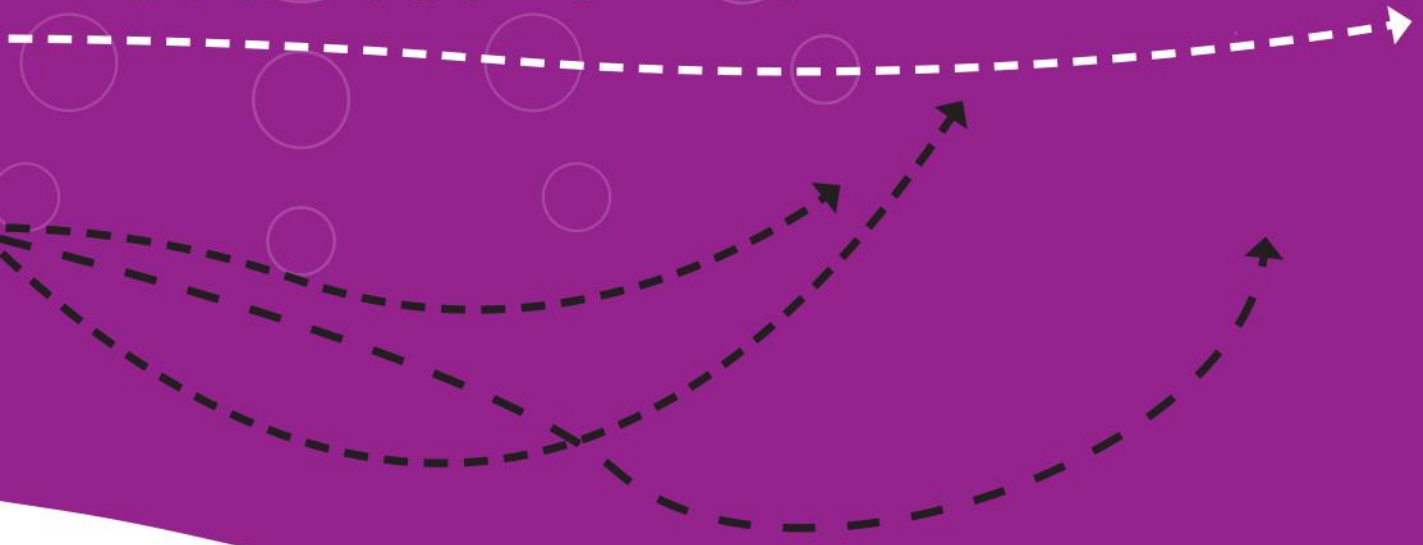
Serie
de Cuadernillos Técnicos

Elaboración del Nivel socioeconómico



GOBIERNO DE LA REPÚBLICA DE GUATEMALA
MINISTERIO DE EDUCACIÓN

Dirección General de Evaluación
e Investigación Educativa
-Digeduca-

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
Ana Aidé Cruz Grünebaum

Revisión de texto y diagramación
María Teresa Marroquín Yurrita
Diseño de portada
Eduardo Avila

Dirección General de Evaluación e Investigación Educativa
© Digeduca 2016 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: Cruz, A. (2016). *Elaboración del Nivel socioeconómico.* 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, 2016

Contenido

Nivel socioeconómico 6
Indicador 7
Índice 8
Escala en investigación educativa 8
Variables que se han utilizado para conformar el Nivel socioeconómico 9
Metodologías utilizadas para la construcción del Nivel socioeconómico 14
Diversos programas para realizar un Análisis factorial 15
Procedimiento de análisis de las variables 15
Bases a utilizar 16
Exploración inicial de la base a utilizar 16
Identificación de variables para el Nivel socioeconómico 20
Recodificación de las variables tipo Cadena 22
Análisis para recodificación de variables 27
Recodificación de los valores de las variables 30
Recodificaciones generales 30
Recodificaciones de variables que conformarán el Nivel socioeconómico 34
Elaboración de bitácora de trabajo 39
Procedimiento de Análisis factorial exploratorio 41
Pasos para realizar un Análisis factorial exploratorio con el *software* SPSS 42
Resumen del procedimiento de Análisis factorial exploratorio 51
Interpretación de las tablas y gráficas resultantes del procedimiento de análisis factorial exploratorio para la conformación del Nivel socioeconómico 52
Matriz de correlaciones 52
KMO y la prueba de Esfericidad de Barlett 53
Comunalidades 54
Gráfico de sedimentación 56
Matriz factorial 57
Matriz de factores rotados 58
Referencias 62
Anexo I. Sintaxis para generar Frecuencias y Correlaciones 64
Anexo II. Sintaxis para recodificar valores de las variables 65
Anexo III. Sintaxis para generar tablas con información de las variables e información de la habilidad que obtuvieron los estudiantes 66

Lista de figuras

Figura 1. Ubicación de carpetas con bases finales 16
Figura 2. Generación de frecuencias 17
Figura 3. Resultados del análisis de frecuencias 17
Figura 4. Histograma de los datos 19
Figura 5. Diagrama de cajas 20
Figura 6. Ejemplo de identificación de variables a utilizar 21
Figura 7. Tipo de variables en la base de datos 21
Figura 8. Tipo de variables según el resultado de frecuencias 22
Figura 9. Observar datos en «Vista de datos» 23
Figura 10. Cambio de variable cadena a numérica parte 1 24
Figura 11. Cambio de variable cadena a numérica parte 2 24
Figura 12. Cambio de variable cadena a numérica parte 3 25
Figura 13. Cambio de variable cadena a numérica parte 4 25
Figura 14. Cambio de variable cadena a numérica parte 5 26
Figura 15. Verificación de frecuencias de variable cadena y variable numérica 26
Figura 16. Base de datos con variables numéricas 27
Figura 17. Valores de las opciones y signo direccional del coeficiente de correlación 31
Figura 18. Formato de preguntas relacionadas con el Nivel socioeconómico parte 1 34
Figura 19. Formato de preguntas relacionadas con el Nivel socioeconómico parte 2 35
Figura 20. Formato de preguntas relacionadas con el Nivel socioeconómico parte 3 35
Figura 21. Formato de preguntas relacionadas con el Nivel socioeconómico parte 4 36
Figura 22. Frecuencias de variables 36
Figura 23. Cruce de variables 39
Figura 24. Selección de Análisis factorial 42
Figura 25. Análisis factorial, Variables 42
Figura 26. Análisis factorial, Descriptivos 43
Figura 27. Análisis factorial, Extracción 45
Figura 28. Análisis factorial, Extracción parte 2 46
Figura 29. Análisis factorial (continuación 1) 47
Figura 30. Análisis factorial, Rotación 48
Figura 31. Análisis factorial (continuación 2) 48
Figura 32. Análisis factorial, Puntuaciones 49
Figura 33. Análisis factorial (continuación 3) 50
Figura 34. Gráfico de sedimentación 57
Figura 35. Factores generados en la base de datos parte 1 60
Figura 36. Factores generados en la base de datos parte 2 61

Lista de tablas

Tabla 1. Variables utilizadas para el Nivel socioeconómico en Graduandos 2011 10
Tabla 2. Variables utilizadas para el Nivel socioeconómico en Graduandos 2011 11
Tabla 3. Variables utilizadas para el Nivel socioeconómico en tercero básico 12
Tabla 4. Variables utilizadas para el Capital cultural de tercero básico 2013 12
Tabla 5. Variables utilizadas para para el Nivel socioeconómico en primaria 2010 13
Tabla 6. *Software* utilizado para realizar análisis factorial 15
Tabla 7. Porcentaje de respuestas y omisiones 18
Tabla 8. Estadísticas descriptivas 19
Tabla 9. Frecuencias de la variable original 23
Tabla 10. Correlaciones de las variables parte 1 28
Tabla 11. Correlaciones de las variables parte 2 28
Tabla 12. Correlaciones de las variables parte 3 29
Tabla 13. Correlaciones de las variables parte 4 30
Tabla 14. Valores para recodificar variables parte 1 32
Tabla 15. Valores para recodificar variables parte 2 33
Tabla 16. Tabla de contingencia con el rendimiento de los estudiantes 38
Tabla 17. Valores de codificación 38
Tabla 18. Observaciones de los cambios en las variables 40
Tabla 19. Opciones de descriptivos del Análisis factorial 44
Tabla 20. Resumen de pasos para el Análisis factorial exploratorio 51
Tabla 21. Matriz de correlaciones 52
Tabla 22. Matriz de significancia de las correlaciones 53
Tabla 23. KMO y prueba de Bartlett 54
Tabla 24. Comunalidades 55
Tabla 25. Varianza total explicada 56
Tabla 26. Matriz de factores 58
Tabla 27. Matriz de factores rotados parte 1 59
Tabla 28. Matriz de factores rotados parte 2 60

Nivel socioeconómico

Los aspectos socioeconómicos se refieren a una combinación de ciertas características del contexto del estudiante que describe su condición social, económica y cultural (OCDE, 2010). Según Willms (2003), el nivel socioeconómico tiene un efecto en los resultados del rendimiento de los estudiantes. El mismo autor afirma que los niños que tienen alto nivel socioeconómico tienen más probabilidades de éxito que aquellos con bajo nivel socioeconómico y que incluso estos son más propensos a abandonar la escuela.

Lamont y Lareau (1988) definen el capital cultural como aquellos recursos que envuelven un conjunto de actitudes, preferencias, conocimientos formales, comportamientos, bienes, entre otros, que son usados para la diferenciación y distinción social. Para la OCDE (2010) el nivel sociocultural está compuesto por actividades culturales y Bordieu (2011) explica este término como «aquella herramienta que al ser apropiada por los individuos les permite la producción y la reproducción de la cultura, estos bienes pueden ser representados por los recursos educativos con que disponen los individuos».

El nivel socioeconómico o sociocultural es una agrupación de variables que conforman una aproximación a las diferencias del nivel socioeconómico y cultural de las familias. Se puede utilizar información referida específicamente a lo económico o una combinación con información de lo cultural.

La Dirección General de Evaluación e Investigación Educativa (Digeduca), en el análisis de factores asociados de las evaluaciones nacionales, ha utilizado la variable nivel socioeconómico conformada por la información sobre características del hogar (piso, paredes, techo), servicios del hogar (agua, electricidad, teléfono, internet, etc.), bienes que poseen en casa, y educación de los padres, este último la literatura lo menciona como característica del capital cultural; sin embargo, para los usos del nivel socioeconómico, en los análisis de factores asociados de Digeduca del año 2006 al 2010, no se separa la información de la educación de los padres ni la información económica.

En el Segundo Estudio Regional Comparativo y Explicativo (SERCE), también se utiliza la variable del nivel sociocultural, este índice cuenta con información directa de los padres del estudiante, de

la lengua de origen, de la infraestructura y servicios que existen en su hogar, así como la disponibilidad de recursos educativos en casa.

Antes de entrar al procedimiento, habrá que establecer la diferencia entre un indicador e índice, debido a que al resultado final de la construcción del nivel socioeconómico algunos lo llaman indicador socioeconómico, índice socioeconómico o escala socioeconómica.

Indicador

Según Cecchini (2005) en general el indicador es un instrumento que se construye con la información, es decir, valores numéricos sintetizando aspectos importantes de algún fenómeno con propósitos de análisis.

Un indicador es una «medida usada para demostrar el cambio que resulta de una actividad, proyecto o programa»; «medidas que ayudan a cuantificar o describir el logro de resultados y monitorear el progreso alcanzado» (Celade, 2002, citado en Cecchini, 2005).

«Un indicador se puede entender como una expresión que sintetiza información cuantitativa y/o cualitativa sobre algún fenómeno relevante» (Mokate, 2003, citado en Cecchini 2005).

Los indicadores pueden ser medidas que se obtienen de la población total, es decir, valores absolutos, pero también pueden construirse a través de un proceso de cálculo que relacione la medida con otras magnitudes (por ejemplo, tasa de aprobación) (Cecchini, 2005).

Asimismo, los indicadores pueden ser simples o compuestos; el indicador simple es una colección de datos básicos por ejemplo, porcentaje de estudiantes en primaria; porcentaje de estudiantes hombres o mujeres (Cecchini, 2005).

Los indicadores están siempre referidos a datos observables, por lo que habrán de reunir dos condiciones esenciales:

a. Estar relacionados con una variable o dimensión concreta.
b. Ser susceptibles de medida, bien de forma descriptiva o analítica.

(García, González y Ballesteros, 2001)

Índice

Los **indicadores compuestos** suelen llamarse índices porque para construirlos hay que utilizar una combinación ponderada de indicadores de los factores subyacentes, y el dato final obtenido es la combinación de más de una variable. Por ejemplo, el Índice de Desarrollo Humano (IDH) del Programa de las Naciones Unidas para el Desarrollo (PNUD), incluye la interacción de variables demográficas, de salud, educación e ingreso y se mide ponderando la esperanza de vida al nacer, la tasa de alfabetismo, la tasa de matriculación primaria, secundaria y terciaria y el ingreso real per cápita (Cecchini, 2005).

Otro ejemplo es el Índice de Vulnerabilidad Social que está siendo desarrollado por la Sede Subregional para el Caribe de la Comisión Económica para América Latina y el Caribe (Cepal) y que incluye a 10 indicadores (Cecchini, 2005).

Los autores García, González y Ballesteros (2001) mencionan que los índices son resultado de la combinación de indicadores. Los mismos autores citan a Gonzalo Blasco (1994) quien señala las siguientes características básicas que deben reunir los índices:

a. Validez o exactitud: grado en que refleja lo que dice medir.
b. Potencia: capacidad para aportar información sobre el concepto que mide y su grado de normalización que permite comparaciones.
c. Reproductibilidad: cualidad de representar el concepto medido.

Escala en investigación educativa

De acuerdo con Stevens (1951) medir es asignar números a los objetos según ciertas reglas. Por tanto, una escala es un conjunto de números que se refiere a un conjunto de objetos, cualidades o propiedades, que mantienen determinadas relaciones entre sí, equivalentes a las relaciones que mantienen los objetos, cualidades o propiedades entre sí y que tienen poder representativo completo por sí solos.

La Subdirección de Análisis de Datos de Digeduca, ha optado por utilizar el término **Nivel socioeconómico**, debido a que el análisis elaborado permite obtener números que ordenados forman parte de una escala.

Variables que se han utilizado para conformar el Nivel socioeconómico

Cada instrumento de evaluación de los aprendizajes que elabora Digeduca, va acompañado con preguntas que recopilan información sobre condiciones socioeconómicas del estudiante. Cada año las preguntas son revisadas, se mejoran, se cambian o se decide no incluirlas en el cuestionario, proceso del que está a cargo el departamento de Desarrollo de Instrumentos de la dirección.

Por esta razón es que cada año la Subdirección de Análisis de Datos, analiza las preguntas del cuestionario de factores asociados de cada grado, y en el estudio generado para conformar el nivel socioeconómico se utilizan variables similares o iguales, dependiendo si las preguntas se mantienen en los cuestionarios o se han modificado.

A continuación se presenta un resumen de las variables empleadas para la creación del **Nivel socioeconómico** en graduandos, tercero básico y primaria.

La **Tabla 1** muestra la información utilizada para generar el nivel socioeconómico de la evaluación de Graduandos 2011. Las variables empleadas en ese año, son similares a las usadas en años anteriores; una de las diferencias es que en el año 2011 no se utilizó la educación de los padres como parte del nivel socioeconómico, y se generó un valor de hacinamiento con las variables: número de personas que viven en la casa y, número de habitaciones en la casa.

Tabla 1. Variables utilizadas para el Nivel socioeconómico en Graduandos 2011

Graduandos 2011
Material del piso de la casa (torta de cemento, piso de granito, piso cerámico, madera, piso de tierra, tabla)
Material de paredes de la casa (block, ladrillo, lámina, madera rústica, madera fina, adobe)
Material del techo de la casa (terraza fundida, duralita, lámina, teja, material perecedero o frágil)
Cómo obtiene el agua que usa para lavar o echar en el baño (tubería, chorro público, fuente natural –río o lago-, pozo)
Agua que utiliza para beber (del chorro, filtrada, pozo, comprada –marca comercial-, fuente natural –río o lago-, cisterna, clorada)
Cuenta con electricidad en la casa
Hay un ambiente separado para la cocina
Tipo de combustible que usa para cocinar: Leña, gas, electricidad
Cuenta con línea telefónica fija
Familia cuenta con uno o más celulares
Cuenta con servicio de televisión por cable en su casa
Cuenta con servicio de internet en su casa
Electrodomésticos en su casa: Televisor, lavadora de ropa, consola de videojuegos, refrigerador, secadora de ropa, equipo de sonido, horno microondas, VHS/DVD, computadora, otros electrodomésticos
Tipo de sanitarios en su casa: No hay, letrinas, inodoro cerámico
Familia cuenta con vehículo propio
Cómo se moviliza para ir al establecimiento educativo: A pie o bicicleta, en transporte público, en motocicleta, en automóvil o bus escolar
Valor de Hacinamiento

Fuente: Informe de Factores Asociados de Graduandos 2011, Digeduca.

En el año 2011, se generó un valor para la variable **Capital cultural**, este valor fue obtenido a través de un análisis factorial, mismo procedimiento que se utiliza para generar el nivel socioeconómico. Los análisis mostraron que utilizar el indicador de capital cultural, aportaba más para la explicación de la varianza, que si se utilizaban las variables separadas relacionadas con este constructo. La **Tabla 2** contiene las variables utilizadas en el Capital cultural de Graduandos 2011.

Tabla 2. Variables utilizadas para el Nivel socioeconómico en Graduandos 2011

Graduandos 2011 Variables del Capital Cultural
Frecuencia de uso de idioma materno: De nunca a mucho
Destreza de uso de idioma materno (habla, entiende, lee, escribe)
Madre asistió a la escuela
Padre asistió a la escuela
Nivel de escolaridad de la madre: Primaria, básicos, diversificado, universidad, posgrado
Nivel de escolaridad del padre: Primaria, básicos, diversificado, universidad, posgrado
Libros completos leídos en el último año: De ninguna a más de 6 libros
Lectura de periódicos
Horas diarias de uso de la computadora en casa: Desde 1 a más de 4 horas diarias
Utiliza internet para hacer investigaciones

Fuente: Informe de Factores Asociados de Graduandos 2011, Digeduca.

En la **Tabla 3** se puede observar que en los tres años en que se generó el nivel socioeconómico de las evaluaciones de tercero básico, hay variables en común; por ejemplo, los electrodomésticos, los servicios en la casa y el material de la infraestructura de la casa.

En el año 2013 se realizó un análisis más detallado y solamente se utilizaron las variables, que según los análisis, mostraban una diferencia clara entre los estudiantes con menor y un mayor nivel socioeconómico.

Asimismo, la información sobre electrodomésticos y bienes que posee la familia se utilizaron en una variable que incluía la suma de todos los bienes, a diferencia de los años 2006 y 2009 en que los electrodomésticos se incluyeron como variables separadas en el análisis factorial.

Tabla 3. Variables utilizadas para el Nivel socioeconómico en tercero básico

Tercero Básico 2006	Tercero Básico 2009	Tercero Básico 2013	
¿Cuenta con electricidad?	¿Cuenta con electricidad?	¿Qué material predomina en el piso de su casa?	
¿Cuenta con línea telefónica fija (cableado)?	¿Cuenta con línea telefónica fija (cableado)?	¿Qué material predomina en las paredes de su casa?	
¿Qué combustible utiliza para cocinar?	¿Hay un ambiente separado par la cocina?	¿Cómo obtiene el agua que usa para lavar o echar en el baño?	
¿Qué material predomina en el piso de su casa?	¿Qué clase de agua utiliza para beber o cocinar?	¿Qué agua utiliza para beber?	
¿Qué material predomina en el techo de su casa?	¿Qué combustible utiliza para cocinar?	¿Cuenta con electricidad en su casa?	
¿Qué material predomina en las paredes de su casa?	¿Qué material predomina en el piso de su casa?	¿Qué combustible utiliza para cocinar?	
¿Su familia tiene vehículo propio?	¿Qué material predomina en el techo de su casa?	¿Tipo de sanitarios de su casa?	
¿Tiene su familia uno o más teléfonos celulares?	¿Qué material predomina en las paredes de su casa?	Marque los electrodomésticos que tiene en su casa (la suma de todos bienes)	
¿Tipo de sanitario de su casa?	¿Su familia tiene vehículo propio?	¿Cuenta con línea telefónica fija (cableado) en su casa?	
Grado más alto aprobado por su mamá	¿Tiene su familia uno o más teléfonos celulares?	¿Tiene su familia uno o más teléfonos celulares?	
Grado más alto aprobado por su papá	¿Tipo de sanitario de su casa?	Suma de los bienes	Marque los electrodomésticos que tiene en su casa: (puede marcar varios)
Tiene electrodoméstico: DVD	Grado más alto aprobado por su mamá		Marque los electrodomésticos que tiene en su casa: Televisor
Tiene electrodoméstico: equipo de sonido	Grado más alto aprobado por su papá		Marque los electrodomésticos que tiene en su casa: Refrigerador
Tiene electrodoméstico: horno microondas	Tiene electrodoméstico: computadora		Marque los electrodomésticos que tiene en su casa: Equipo de sonido
Tiene electrodoméstico: lavadora de ropa	Tiene electrodoméstico: consolas de videojuegos		Marque los electrodomésticos que tiene en su casa: VHS/DVD
Tiene electrodoméstico: otros	Tiene electrodoméstico: equipo de sonido		Marque los electrodomésticos que tiene en su casa: Lavadora de ropa
Tiene electrodoméstico: refrigerador	Tiene electrodoméstico: horno microondas		Marque los electrodomésticos que tiene en su casa: Secadora de ropa
Tiene electrodoméstico: secadora de ropa	Tiene electrodoméstico: lavadora de ropa		Marque los electrodomésticos que tiene en su casa: Horno microondas
Tiene electrodoméstico: televisor	Tiene electrodoméstico: otros		Marque los electrodomésticos que tiene en su casa: Computadora
Tiene electrodoméstico: VHS	Tiene electrodoméstico: refrigerador		Marque los electrodomésticos que tiene en su casa: Consola de videojuegos
	Tiene electrodoméstico: secadora de ropa		Marque los electrodomésticos que tiene en su casa: Otros electrodomésticos
	Tiene electrodoméstico: televisor		¿Su familia tiene vehículo propio?
	Tiene electrodoméstico: VHS/DVD		

Fuente: Bases de datos de la evaluación de tercero básico de los años 2006, 2009 y 2013, Digeduca.

La **Tabla 4** muestra las variables utilizadas en el Capital cultural de tercero básico 2013; en el mismo se utilizaron variables como la educación de los padres, que en los años 2006 y 2009 se habían incluido en el nivel socioeconómico.

Tabla 4. Variables utilizadas para el Capital cultural de tercero básico 2013

Tercero Básico 2013 Variables del Capital Cultural
¿Poseen computadoras en su establecimiento para uso de los estudiantes?
¿Utiliza la computadora de su establecimiento para realizar tareas?
¿Cuántos minutos dura un periodo de Matemáticas en su escuela?
¿Cuántos períodos de Matemáticas recibe a la semana?
¿Cuántos minutos dura un periodo de Idioma/Lenguaje/Lectura en su escuela?
¿Cuántos períodos de Idioma/Lenguaje/Lectura recibe a la semana?
¿Cuál es su identificación étnica?
¿Cuál es su idioma materno?:
Idioma materno de su mamá:
Idioma materno de su papá:
Marque el grado más alto aprobado por su papá.
Marque el grado más alto aprobado por su mamá.

Fuente: Bases de datos de la evaluación de tercero básico 2013, Digeduca.

La **Tabla 5** presenta la información que conforma el nivel socioeconómico de primero, tercero y sexto primaria en la evaluación del año 2010. En este año, para primer grado, se generó al igual que en los años 2008 y 2009, un nivel sociocultural, ya que el cuadernillo de factores asociados no tiene las preguntas suficientes para generar un nivel socioeconómico. Esto es debido a que hay preguntas que es poco probable que los niños de primero primaria contesten, por ser niños pequeños y por esa razón, no se incluyen en los cuestionarios de factores asociados de los años 2006 al 2013.

En tercero y sexto primaria, se generó un nivel socioeconómico y se incluyó la información del nivel educativo de los padres. En ambos grados se utilizaron las mismas variables, ya que el cuadernillo de factores asociados de los años 2006 al 2013, contiene las mismas preguntas para ambos grados.

Tabla 5. Variables utilizadas para el Nivel socioeconómico en primaria 2010

1ro Primaria 2010	3ro y 6to Primaria 2010
¿Tu mamá sabe leer?	¿Hay en tu casa Computadora?
¿Tu mamá fue a la escuela?	¿Hay en tu casa Lavadora de ropa?
¿Cuál fue el último grado de estudios de tu mamá?	¿Hay en tu casa Teléfono de línea fija?
¿Tu papá sabe leer?	¿Hay en tu casa Videograbadora?
¿Tu papá fue a la escuela?	¿Hay en tu casa Auto?
¿Cuál fue el último grado de estudios de tu papá?	¿Hay en tu casa Camión, tractor?
¿Alguien en casa te ayuda a hacer las tareas o a estudiar?	¿Hay en tu casa Televisión?
¿Hay libros en tu casa para leer?	¿Hay en tu casa Plancha eléctrica
¿Alguien de tu familia te lee cuentos o historias?	¿Hay en tu casa Teléfono celular?
¿Cuánto tiempo dedicas a leer en tu casa?	¿Hay en tu casa Equipo de sonido?
Materiales de Lectura que hay en casa	¿Hay en tu casa Reproductor de DVD?
	¿Hay en tu casa Refrigeradora?
	¿Qué material predomina en las paredes de tu casa?
	¿Qué tipo de gas utilizan en la Cocina de tu casa?
	¿Qué material predomina en el piso de tu casa?
	¿Qué material predomina en el techo de tu casa?
	¿Cómo obtienen el agua en tu casa?
	¿Cómo se ilumina tu casa?
	¿Cuál fue el último grado de estudios de tu papá?
	¿Cuál fue el último grado de estudios de tu mamá?

Fuente: Informe de Factores Asociados de primaria 2010, Digeduca.

Metodologías utilizadas para la construcción del Nivel socioeconómico

Existen diferentes procedimientos para la construcción del nivel socioeconómico; el Sistema de Medición de la Calidad de la Educación de Chile (Simce) utiliza la técnica estadística de conglomerados que consiste en dividir un conjunto de observaciones en grupos, de manera que las características de interés (características que corresponden a variables socioeconómicas) sean lo más similares entre sí, y que a la vez entre grupos diferentes sean lo más distintas posibles (Simce, 2011).

Otros autores como Caro (2009) con datos de evaluaciones de Perú y Gil (2011) en su estudio *Medición del nivel socioeconómico familiar en el alumnado de Educación Primaria en Sevilla*; el autor Saz Choxín en su tesis *Análisis de los Factores Asociados con el rendimiento de los estudiantes de tercero básico en Matemática y Lectura, del año 2006 en Guatemala,* así como también en los análisis de la *Infraestructura escolar en las primarias y secundarias de México* por el Instituto Nacional para la Evaluación de la Educación (INEE), utilizan la metodología de reducción de variables, a través de un análisis factorial.

Según el procedimiento de análisis factorial, la construcción del nivel o índice socioeconómico se basa en la integración sucesiva de variables en factores. Técnica de análisis de componentes principales que permite sintetizar la información obtenida de muchas variables en unos cuantos indicadores que representan a constructos empíricos y cuya correlaciones cercana a cero (INEE, 2007).

La Subdirección de Análisis de Datos de la Digeduca, ha utilizado la metodología de **Análisis factorial** para obtener el Nivel socioeconómico, y utilizarlo en sus análisis de Factores Asociados al rendimiento de los estudiantes.

Diversos programas para realizar un Análisis factorial

Existen diferentes tipos de programas de cómputo para realizar el Análisis factorial, algunos de ellos se mencionan en la **Tabla 6**. En Digeduca el *software* con el que se generan análisis factoriales es SPSS.

Tabla 6. *Software* utilizado para realizar análisis factorial

Programas	Observaciones
R	Se necesita elaborar sintaxis.
STATA	Tiene una interfaz poco amigable, pero se puede utilizar sintaxis o realizar el análisis a través de los comandos visuales.
SPSS	Tiene una interfaz amigable, por lo que se puede elegir para realizar el análisis con sintaxis o a través de los comandos visuales.
MPLUS	Se necesita elaborar sintaxis.
SAS	Se necesita elaborar sintaxis.

Fuente: Digeduca, 2014.

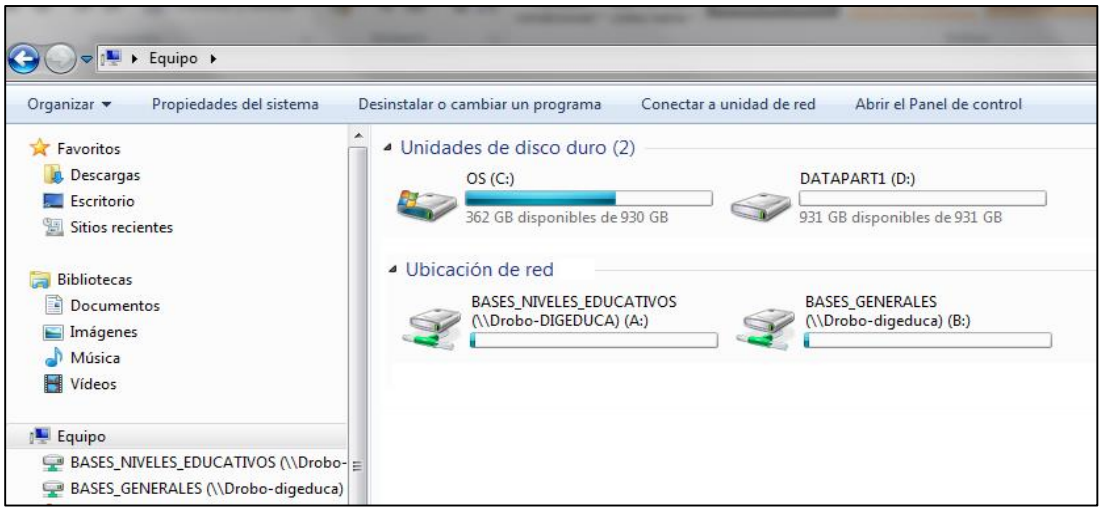
Procedimiento de análisis de las variables

Los procedimientos que se abordarán en el siguiente apartado, no solamente pueden utilizarse para la generación del Nivel socioeconómico así como también para el análisis de variables que puedan conformar otros indicadores como Capital cultural, indicadores que incluyan información sobre la clase de Matemática o información relacionada con la clase de Comunicación y Lenguaje, entre otros indicadores.

Bases a utilizar

Como punto de partida, se debe tener la base final del grado para el cual se generará el Nivel socioeconómico. Las bases finales se encuentran en la carpeta compartida con nombre «BASES_NIVELES_EDUCATIVOS»; si es el caso de que se utilicen bases del cuestionario de directores, docentes de grado o de alguna otra evaluación, el nombre de la carpeta es «BASES_GENERALES» (ver **Figura 1**).

Figura 1. Ubicación de carpetas con bases finales



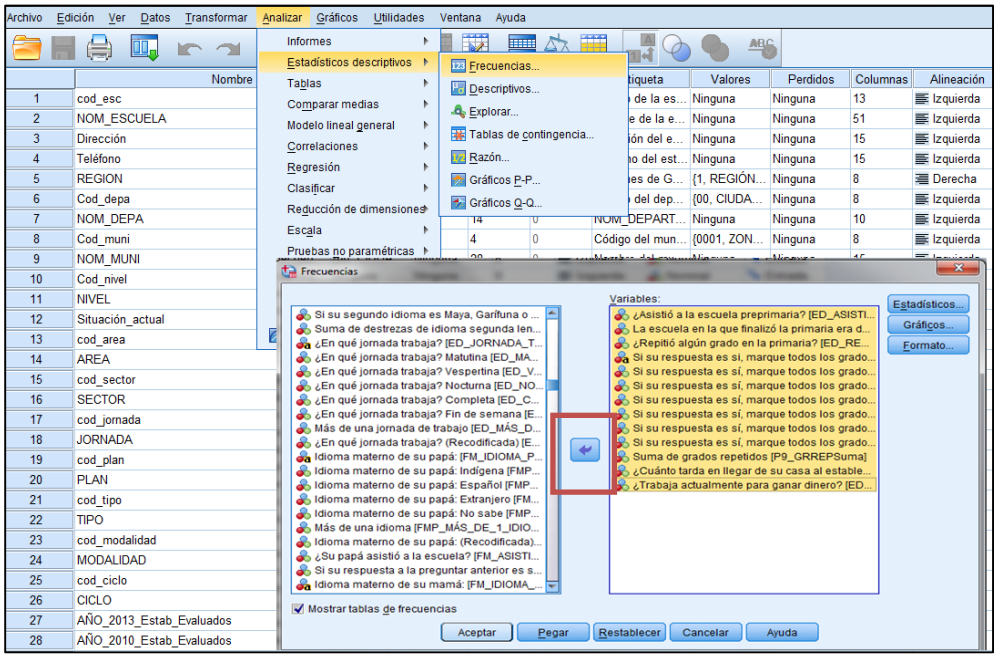
Fuente: Digeduca, 2014.

Exploración inicial de la base a utilizar

Se procede a abrir la base que se estará utilizando y se realiza una exploración inicial, la cual incluirá un **análisis de frecuencias** de cada una de las variables que conforman la base de datos para conocer qué características e información tienen las preguntas o variables que conforman la base de datos que se está usando.

Para generar el análisis de frecuencias, se utiliza la opción de: *Analizar, Estadísticos Descriptivos, Frecuencias*. Al ver el recuadro presentado en la **Figura 2**, se selecciona del listado de variables la información que se quiere explorar y utilizando la flecha, se colocan del lado derecho, en el espacio en blanco con título «Variables».

Figura 2. Generación de frecuencias



Fuente: Digeduca, 2014.

El resultado del análisis de frecuencias son tablas que presentan información sobre la cantidad de respuestas dadas a la pregunta, porcentaje de datos válidos y perdidos de cada variable (**Figura 3**).

Figura 3. Resultados del análisis de frecuencias

Frecuencias

[Conjunto_de_datos1] C:\Users\aacruz\Desktop\3RO BÁSICO 2013 VERSIÓN FINAL.sav

Estadísticos

		TELEVISOR Marque los electrodomésticos que tiene en su casa: Televisor	REFRIGERADOR Marque los electrodomésticos que tiene en su casa: Refrigerador	EQUIPO_SONIDO Marque los electrodomésticos que tiene en su casa: Equipo de sonido	VHS_DVD Marque los electrodomésticos que tiene en su casa: VHS/DVD	LAVADORA_ROPA Marque los electrodomésticos que tiene en su casa: Lavadora de ropa	SECADORA_ROPA Marque los electrodomésticos que tiene en su casa: Secadora de ropa	HORNO_MICROONDAS Marque los electrodomésticos que tiene en su casa: Horno microondas	COMPUTADORAS Marque los electrodomésticos que tiene en su casa: Computadora	CONSOLA_VIDEOJUEGOS Marque los electrodomésticos que tiene en su casa: Consola de videojuegos
N	Válidos	191412	191412	191412	191412	191412	191412	191412	191412	191412
	Perdidos	0	0	0	0	0	0	0	0	0

Tabla de frecuencia

TELEVISOR Marque los electrodomésticos que tiene en su casa: Televisor

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0 NO	18854	9.8	9.8	9.8
	1 SI	172558	90.2	90.2	100.0
	Total	191412	100.0	100.0	

REFRIGERADOR Marque los electrodomésticos que tiene en su casa: Refrigerador

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0 NO	71801	37.5	37.5	37.5
	1 SI	119611	62.5	62.5	100.0
	Total	191412	100.0	100.0	

Fuente: Digeduca, 2014.

Con la información del análisis de frecuencias se identifica el porcentaje de respuestas que tuvo la pregunta y el porcentaje de omisiones, resultados que pueden ser colocados de forma ordenada en una tabla en Excel, que servirá como anexo, para futuros informes de Factores Asociados (ver **Tabla 7**).

Tabla 7. Porcentaje de respuestas y omisiones

Variables	Categorías	3ro P	
		% de Respuestas	% de Omisiones
¿Hay en tu casa Computadora?	1 = Sí 0 = No	99.3	0.7
¿Hay en tu casa Lavadora de ropa?	1 = Sí 0 = No	99.3	0.7
¿Hay en tu casa Teléfono de línea fija?	1 = Sí 0 = No	99.3	0.7
¿Hay en tu casa Videograbadora?	1 = Sí 0 = No	99.3	0.7
¿Hay en tu casa Auto?	1 = Sí 0 = No	99.3	0.7
¿Hay en tu casa Camión, tractor?	1 = Sí 0 = No	99.3	0.7
¿Hay en tu casa Televisión?	1 = Sí 0 = No	99.3	0.7
¿Hay en tu casa Plancha eléctrica	1 = Sí 0 = No	99.3	0.7
¿Hay en tu casa Teléfono celular?	1 = Sí 0 = No	99.3	0.7
¿Hay en tu casa Equipo de sonido?	1 = Sí 0 = No	99.3	0.7
¿Hay en tu casa Reproductor de DVD?	1 = Sí 0 = No	99.3	0.7
¿Hay en tu casa Refrigeradora?	1 = Sí 0 = No	99.3	0.7
¿Qué material predomina en las paredes de tu casa?	1 = Madera Rústica 2 = Lámina 3 = Adobe 4 = Madera Fina 5 = Block 6 = Ladrillo	82.9	17.1
¿Qué tipo de gas utilizan en la Cocina de tu casa?	1 = Cocina de Leña 2 = Cocina queroseno 3 = Cocina gas eléctrica	79.8	20.2

Fuente: Digeduca, 2014.

Es importante generar los estadísticos más relevantes para conocer la población que conforma la base de datos, dentro de ellos están los presentados en la **Tabla 8**, que son el rendimiento de los estudiantes con sus estadísticos descriptivos, media, mediana, moda, curtosis, asimetría, entre otros.

Tabla 8. Estadísticas descriptivas

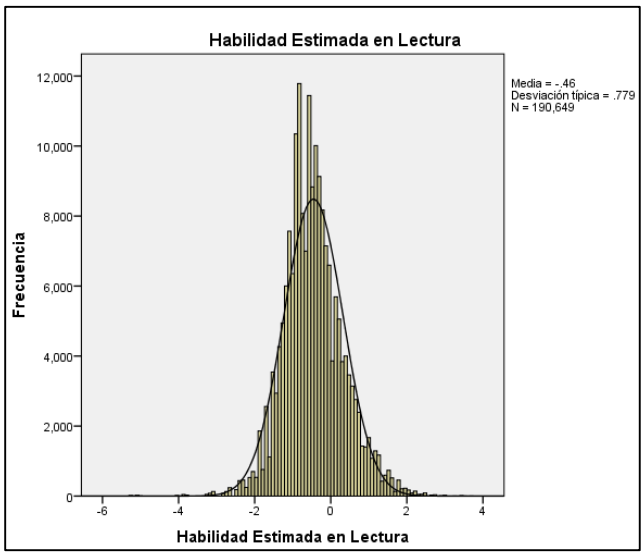
Estadísticos

		Habilidad Estimada en Matemáticas	Habilidad Estimada en Lectura
N	Válidos	190928	190649
	Perdidos	484	763
Media		.00	-.46
Mediana		.04	-.51
Moda		0	-1
Desv. típ.		.662	.779
Asimetría		-.374	.287
Error típ. de asimetría		.006	.006
Curtosis		4.104	1.171
Error típ. de curtosis		.011	.011

Fuente: Digeduca, 2014.

Cuando se realiza la exploración de la base, también es pertinente generar gráficos, los cuales son fuente de información sobre la distribución de los datos (**Figura 4**).

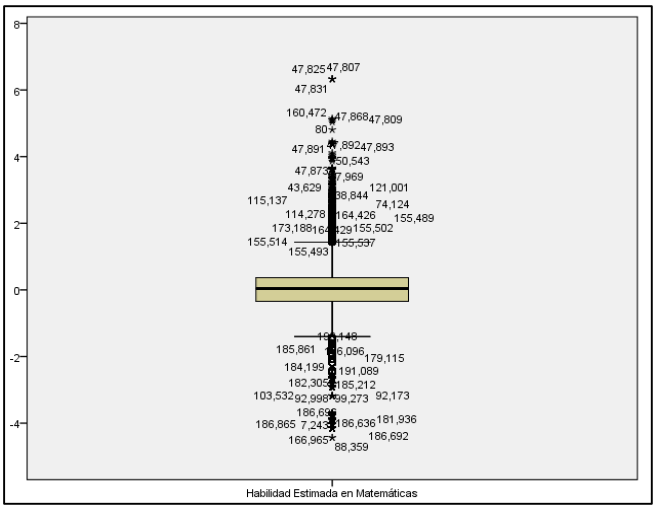
Figura 4. Histograma de los datos



Fuente: Digeduca, 2014.

Otro gráfico que ofrece información importante es el diagrama de cajas (**Figura 5**); en él se pueden observar los datos extremos, lo cual es útil para considerar el tomar decisiones sobre ellos, como por ejemplo, permitir que sean parte de la base de datos para el análisis o decidir eliminarlos con teoría fundamentada en procesos estadísticos que permitan este tipo de acciones.

Figura 5. Diagrama de cajas



Fuente: Digeduca, 2014.

Identificación de variables para el Nivel socioeconómico

Una vez que se haya tenido el primer acercamiento y exploración de los datos, lo siguiente será generar frecuencias para identificar qué variables conformarán el nivel socioeconómico u otro indicador que se desee crear.

Se observará cada una de las variables, la forma de codificación en que fue registrada la información y si son variables numéricas (digitadas en formato de número) o en cadena (digitadas en formato de letras o números). La observación de los datos es importante porque ayudará a determinar qué variables se eliminarán, cuáles se usarán y cuáles necesitan ser recodificadas o transformadas según sea necesario para el análisis factorial, puesto que este solo permite variables numéricas. Para realizar este cambio se realiza lo siguiente:

Al tener generados los cuadros de frecuencias, se pueden ir colocando en un Excel para identificar qué recodificaciones necesita cada una de las variables (**Figura 6**).

Figura 6. Ejemplo de identificación de variables a utilizar

P8 ¿Tu mama fue a la escuela?

		Frecuencia	Porcentaje	válido	acumulado
Válidos		3528	16.8	16.9	16.9
	0 No	4530	21.6	21.8	38.7
	1 Si	12759	60.9	61.3	100.0
	Total	20817	99.4	100.0	
Perdidos	Y Doble respuesta	134	.6		
Total		20951	100.0		

NUMÉRICA

P9 ¿Cuál fue el último grado de estudios de tu mamá?

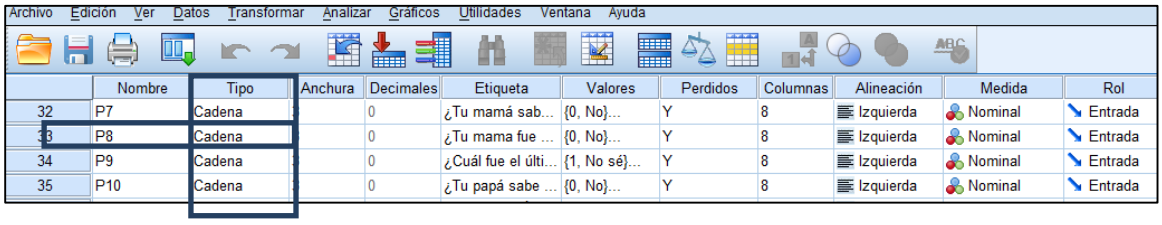
		Frecuencia	Porcentaje	válido	acumulado
Válidos		3216	15.4	16.2	16.2
	1 No sé	5679	27.1	28.6	44.8
	2 No estudió	2825	13.5	14.2	59.0
	3 Primaria	4916	23.5	24.8	83.8
	4 Básicos	1781	8.5	9.0	92.8
	5 Carrera a nivel medio	555	2.6	2.8	95.6
	6 Universitario incompleto	347	1.7	1.7	97.3
	7 Universitario Completo	531	2.5	2.7	100.0
	Total	19850	94.7	100.0	
Perdidos	Y Doble respuesta	1101	5.3		
Total		20951	100.0		

0 No estudió
1Primaria
2Básicos
3Carrera a Nivel Medio
4Universitario Incompleto
5Universitario Completo

Fuente: Digeduca, 2014.

Por ejemplo, la **Figura 7** muestra en la base de datos, en vista de variables, la variable P8, que es una variable de cadena. Esto se puede identificar en la columna de Tipo, que tiene la palabra **Cadena**.

Figura 7. Tipo de variables en la base de datos

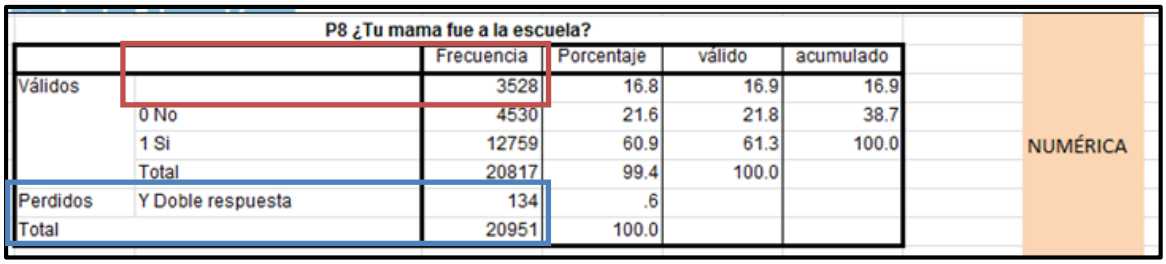


	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
32	P7	Cadena		0	¿Tu mamá sab...	{0, No}...	Y	8	Izquierda	Nominal	Entrada
33	P8	Cadena		0	¿Tu mama fue ...	{0, No}...	Y	8	Izquierda	Nominal	Entrada
34	P9	Cadena		0	¿Cuál fue el últi...	{1, No sé}...	Y	8	Izquierda	Nominal	Entrada
35	P10	Cadena		0	¿Tu papá sabe ...	{0, No}...	Y	8	Izquierda	Nominal	Entrada

Fuente: Digeduca, 2014.

Otra forma de identificar que es una variable en cadena, es observar la tabla de las frecuencias, porque en los datos hay espacios vacíos que no están considerados en valores perdidos como lo presenta la **Figura 8**.

Figura 8. Tipo de variables según el resultado de frecuencias



P8 ¿Tu mama fue a la escuela?

		Frecuencia	Porcentaje	válido	acumulado
Válidos		3528	16.8	16.9	16.9
	0 No	4530	21.6	21.8	38.7
	1 Si	12759	60.9	61.3	100.0
	Total	20817	99.4	100.0	
Perdidos	Y Doble respuesta	134	.6		
Total		20951	100.0		

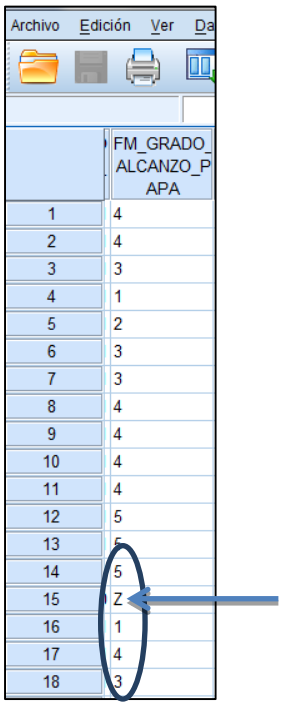
NUMÉRICA

Fuente: Digeduca, 2014.

Recodificación de las variables tipo Cadena

Cuando se hayan identificado las variables tipo Cadena, se procede a realizar los cambios que se requieran. En la base de datos en la pestaña «Vista de datos», se verifica visualmente la variable para observar cómo fueron digitados los números que la componen. Si los números digitados están hacia la izquierda, la variable está codificada como cadena, y puede tener algunos signos o letras que no sean números. Esto es importante porque si tiene letras u otros signos, cuando se recodifique a una variable numérica, se perderán esos datos; automáticamente se borran de la variable (**Figura 9**).

Figura 9. Observar datos en «Vista de datos»



Fuente: Digeduca, 2014.

Luego de observar los datos, también se genera una frecuencia de la variable original, antes de modificarla, para que sirva de comparación con las frecuencias de la variable cuando ya esté modificada en numérica (**Tabla 9**).

Tabla 9. Frecuencias de la variable original

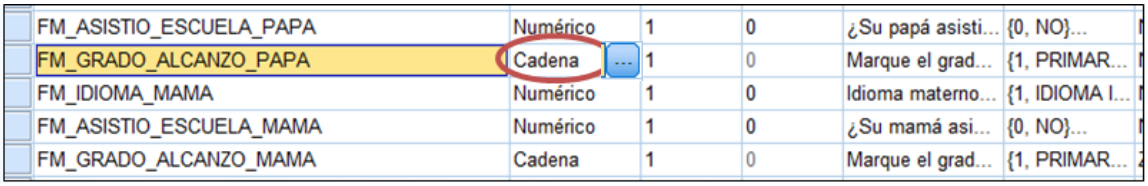
Marque el grado más alto aprobado por su papá

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
1 PRIMARIA	55173	44.7	44.7	44.7
2 BÁSICOS	14675	11.9	11.9	56.6
3 DIVERSIFICADO	19034	15.4	15.4	72
4 UNIVERSIDAD	10869	8.8	8.8	80.8
5 POSGRADO	2282	1.8	1.8	82.6
Z SIN RESPUESTA	21479	17.4	17.4	100
Total	123512	100	100	

Fuente: Digeduca, 2014.

Para cambiar el tipo de variable de cadena a numérica, se presiona con el cursor en la columna con la palabra «cadena», de la variable que se desea cambiar, y cuando aparece un recuadro celeste, también deberá presionarse (**Figura 10**).

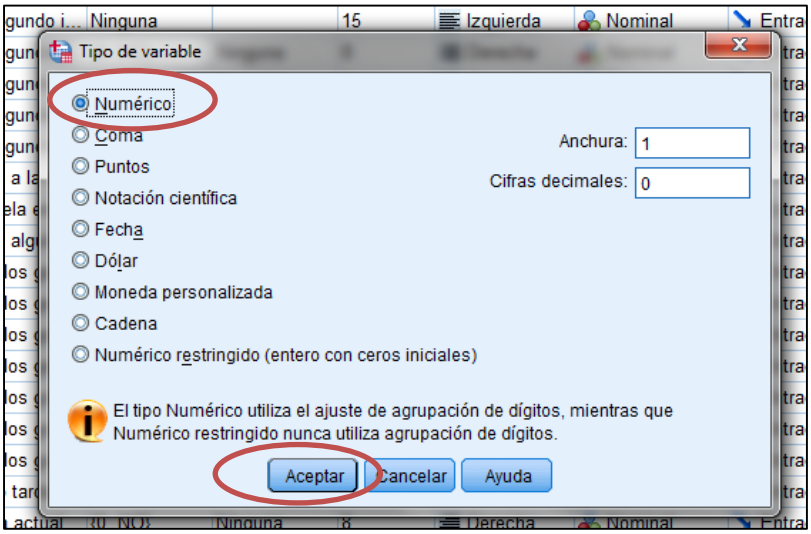
Figura 10. Cambio de variable cadena a numérica parte 1



Fuente: Digeduca, 2014.

En seguida aparecerá la ventana en donde se selecciona la opción «Numérico» y se presiona el botón «Aceptar» (**Figura 11**).

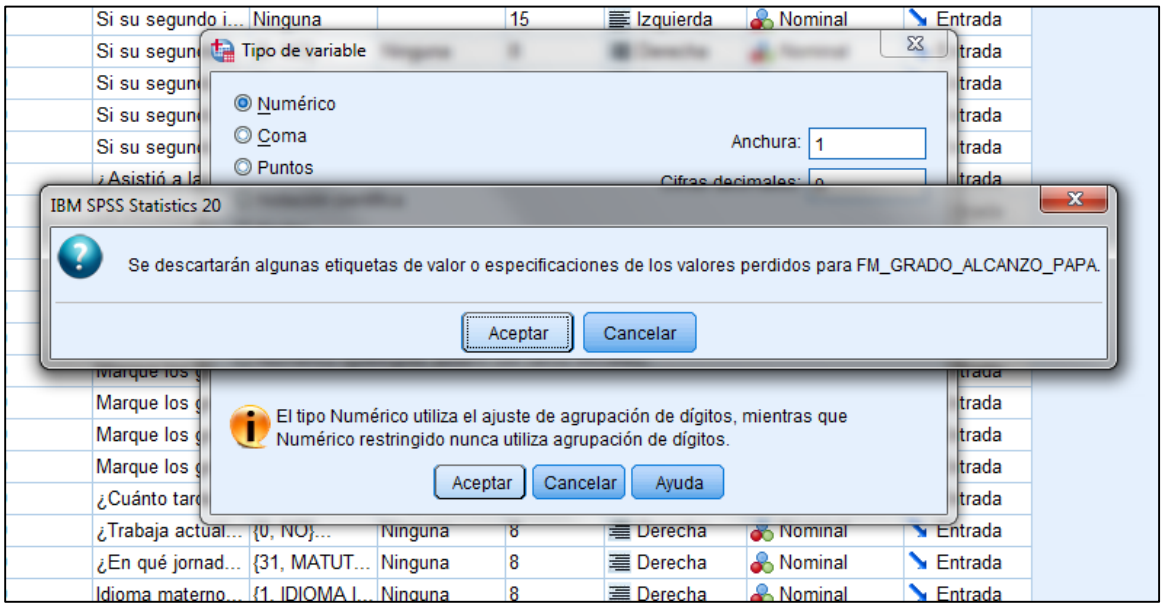
Figura 11. Cambio de variable cadena a numérica parte 2



Fuente: Digeduca, 2014.

Aparecerá una notificación que advierte que los valores de las etiquetas se descartarán. Luego de leer esta notificación hay que presionar el botón «Aceptar». Las nuevas versiones de SPSS informan sobre esta situación, pero las etiquetas de valores o especificaciones de valores perdidos se conservan. En antiguas versiones de SPSS sí se eliminan los valores en la etiqueta (**Figura 12**).

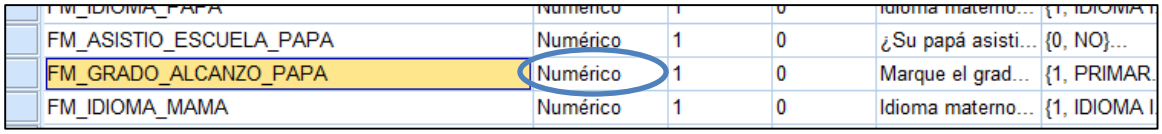
Figura 12. Cambio de variable cadena a numérica parte 3



Fuente: Digeduca, 2014.

Al presionar «Aceptar», la variable que se está cambiando, ahora aparece con la palabra «Numérico» (**Figura 13**).

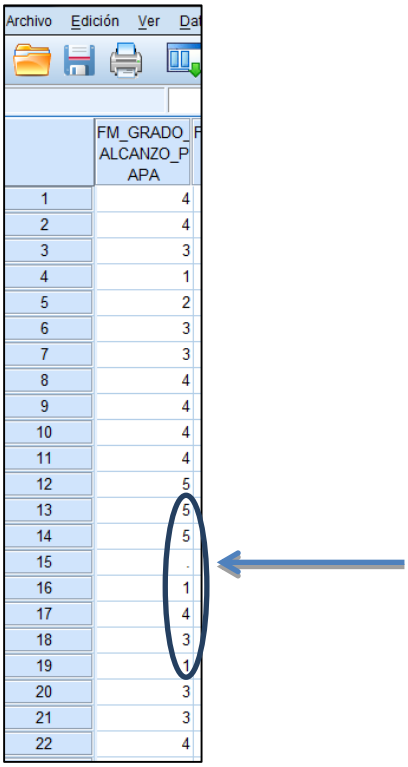
Figura 13. Cambio de variable cadena a numérica parte 4



Fuente: Digeduca, 2014.

Siempre hay que revisar visualmente, en la pestaña de «Vista de datos», verificando que los números siguen allí. Pues hay veces que se borran todos los datos de la variable, entonces se deberá repetir el procedimiento o utilizar otro. Asimismo, si la variable tenía letras u otros signos, el procedimiento utilizado los elimina automáticamente y les coloca un punto como lo muestra la **Figura 14**.

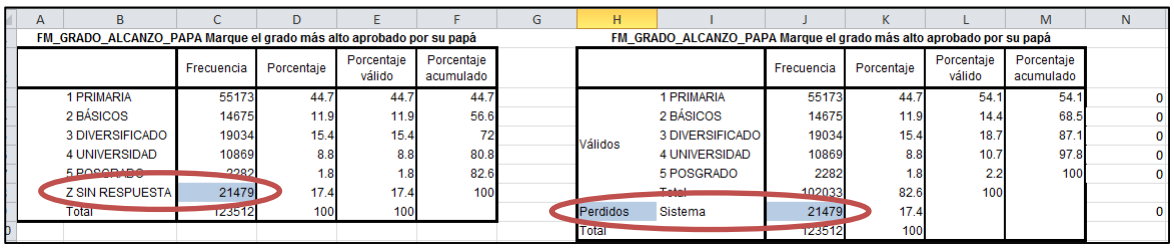
Figura 14. Cambio de variable cadena a numérica parte 5



Fuente: Digeduca, 2014.

Cuando se haya convertido la variable tipo cadena en numérica, se genera nuevamente la frecuencia de la variable numérica para comparar los datos con la frecuencia de la variable original. En la **Figura 15**, se puede observar que los datos como «Z» ahora son datos perdidos y todas las frecuencias de la variable original corresponden en la misma cantidad a las frecuencias de la variable modificada como numérica.

Figura 15. Verificación de frecuencias de variable cadena y variable numérica



FM_GRADO_ALCANZO_PAPA Marque el grado más alto aprobado por su papá

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	1 PRIMARIA	55173	44.7	44.7	44.7
	2 BÁSICOS	14675	11.9	11.9	56.6
	3 DIVERSIFICADO	19034	15.4	15.4	72
	4 UNIVERSIDAD	10869	8.8	8.8	80.8
	5 POSGRADO	2282	1.8	1.8	82.6
	Z SIN RESPUESTA	21479	17.4	17.4	100
	Total	123512	100	100	

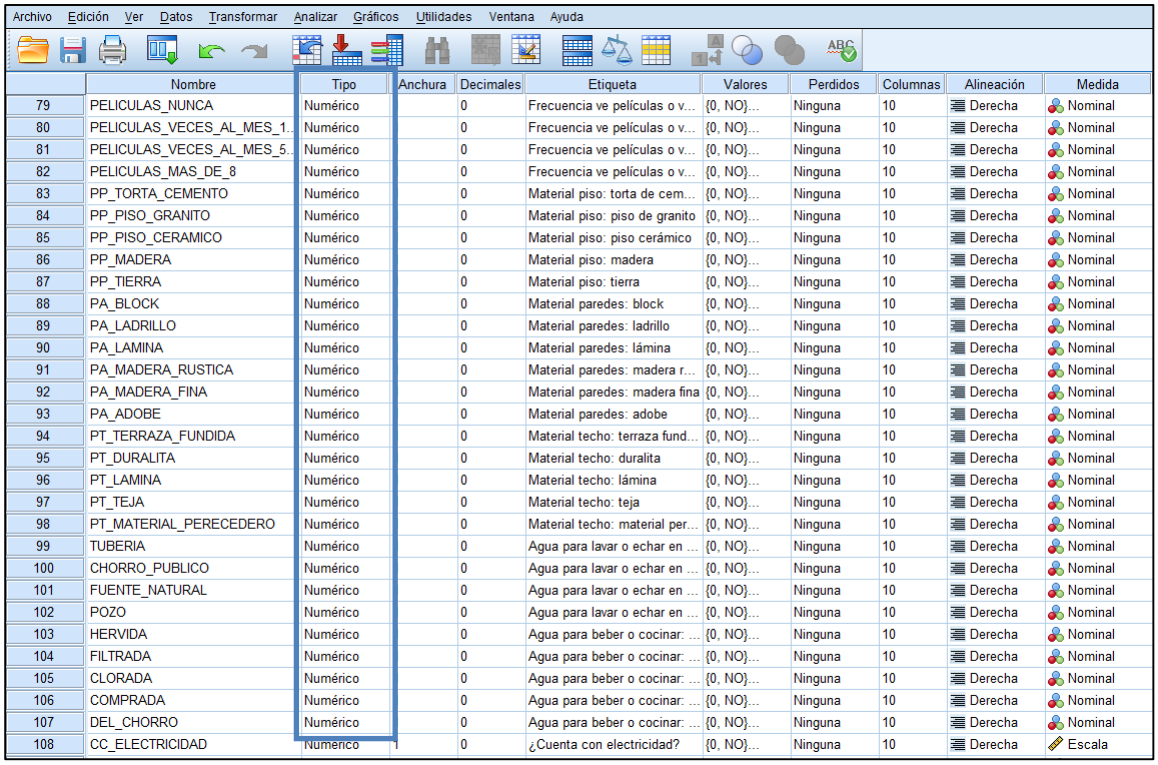
FM_GRADO_ALCANZO_PAPA Marque el grado más alto aprobado por su papá

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado	
Válidos	1 PRIMARIA	55173	44.7	54.1	54.1	0
	2 BÁSICOS	14675	11.9	14.4	68.5	0
	3 DIVERSIFICADO	19034	15.4	18.7	87.1	0
	4 UNIVERSIDAD	10869	8.8	10.7	97.8	0
	5 POSGRADO	2282	1.8	2.2	100	0
	Total	102033	82.6	100		
Perdidos	Sistema	21479	17.4			0
Total		123512	100			

Fuente: Digeduca, 2014.

Todas las variables que se decidan analizar o utilizar para conformar un indicador o nivel socioeconómico quedarán como variables numéricas, como lo muestra la **Figura 16**.

Figura 16. Base de datos con variables numéricas



Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana Ayuda

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
79	PELICULAS_NUNCA	Numérico		0	Frecuencia ve películas o v...	{0, NO}...	Ninguna	10	Derecha	Nominal
80	PELICULAS_VECES_AL_MES_1..	Numérico		0	Frecuencia ve películas o v...	{0, NO}...	Ninguna	10	Derecha	Nominal
81	PELICULAS_VECES_AL_MES_5..	Numérico		0	Frecuencia ve películas o v...	{0, NO}...	Ninguna	10	Derecha	Nominal
82	PELICULAS_MAS_DE_8	Numérico		0	Frecuencia ve películas o v...	{0, NO}...	Ninguna	10	Derecha	Nominal
83	PP_TORTA_CEMENTO	Numérico		0	Material piso: torta de cem...	{0, NO}...	Ninguna	10	Derecha	Nominal
84	PP_PISO_GRANITO	Numérico		0	Material piso: piso de granito	{0, NO}...	Ninguna	10	Derecha	Nominal
85	PP_PISO_CERAMICO	Numérico		0	Material piso: piso cerámico	{0, NO}...	Ninguna	10	Derecha	Nominal
86	PP_MADERA	Numérico		0	Material piso: madera	{0, NO}...	Ninguna	10	Derecha	Nominal
87	PP_TIERRA	Numérico		0	Material piso: tierra	{0, NO}...	Ninguna	10	Derecha	Nominal
88	PA_BLOCK	Numérico		0	Material paredes: block	{0, NO}...	Ninguna	10	Derecha	Nominal
89	PA_LADRILLO	Numérico		0	Material paredes: ladrillo	{0, NO}...	Ninguna	10	Derecha	Nominal
90	PA_LAMINA	Numérico		0	Material paredes: lámina	{0, NO}...	Ninguna	10	Derecha	Nominal
91	PA_MADERA_RUSTICA	Numérico		0	Material paredes: madera r...	{0, NO}...	Ninguna	10	Derecha	Nominal
92	PA_MADERA_FINA	Numérico		0	Material paredes: madera fina	{0, NO}...	Ninguna	10	Derecha	Nominal
93	PA_ADOBE	Numérico		0	Material paredes: adobe	{0, NO}...	Ninguna	10	Derecha	Nominal
94	PT_TERRAZA_FUNDIDA	Numérico		0	Material techo: terraza fund...	{0, NO}...	Ninguna	10	Derecha	Nominal
95	PT_DURALITA	Numérico		0	Material techo: duralita	{0, NO}...	Ninguna	10	Derecha	Nominal
96	PT_LAMINA	Numérico		0	Material techo: lámina	{0, NO}...	Ninguna	10	Derecha	Nominal
97	PT_TEJA	Numérico		0	Material techo: teja	{0, NO}...	Ninguna	10	Derecha	Nominal
98	PT_MATERIAL_PERECEDERO	Numérico		0	Material techo: material per...	{0, NO}...	Ninguna	10	Derecha	Nominal
99	TUBERIA	Numérico		0	Agua para lavar o echar en ...	{0, NO}...	Ninguna	10	Derecha	Nominal
100	CHORRO_PUBLICO	Numérico		0	Agua para lavar o echar en ...	{0, NO}...	Ninguna	10	Derecha	Nominal
101	FUENTE_NATURAL	Numérico		0	Agua para lavar o echar en ...	{0, NO}...	Ninguna	10	Derecha	Nominal
102	POZO	Numérico		0	Agua para lavar o echar en ...	{0, NO}...	Ninguna	10	Derecha	Nominal
103	HERVIDA	Numérico		0	Agua para beber o cocinar: ...	{0, NO}...	Ninguna	10	Derecha	Nominal
104	FILTRADA	Numérico		0	Agua para beber o cocinar: ...	{0, NO}...	Ninguna	10	Derecha	Nominal
105	CLORADA	Numérico		0	Agua para beber o cocinar: ...	{0, NO}...	Ninguna	10	Derecha	Nominal
106	COMPRADA	Numérico		0	Agua para beber o cocinar: ...	{0, NO}...	Ninguna	10	Derecha	Nominal
107	DEL_CHORRO	Numérico		0	Agua para beber o cocinar: ...	{0, NO}...	Ninguna	10	Derecha	Nominal
108	CC_ELECTRICIDAD	Numérico	1	0	¿Cuenta con electricidad?	{0, NO}...	Ninguna	10	Derecha	Escala

Fuente: Digeduca, 2014.

Análisis para recodificación de variables

Antes de asignar nuevos valores a las variables que se utilizarán, se debe realizar un análisis detallado de las mismas; el primer análisis consiste en generar correlaciones entre cada variable que conforma la base de datos (**Tabla 10**).

Tabla 10. Correlaciones de las variables parte 1

Correlaciones

		FM_GRADO_ALCANZO_PAPA Grado más alto aprobado por su papá	FM_GRADO_ALCANZO_MAMA Grado más alto aprobado por su mamá	Piso_recode ¿Qué material predomina en el piso de su casa?	Paredes_recode ¿Qué material predomina en las paredes de su casa?	Techo_recode ¿Qué material predomina en el techo de su casa?	CC_ELECTRICIDAD ¿Cuenta con electricidad?	CC_AMBIENTE_COCINA_SEPARADA ¿Hay un ambiente separado par la cocina?	Combustible_Cocina_recode ¿Qué combustible utiliza para cocinar?	CC_CUENTA_LINEA_TELEFONICA ¿Cuenta con línea telefónica fija (cableado)?	CC_TIENE_FAMILIA_CELULAR ¿Tiene su familia uno o más teléfonos celulares?	TELEVISOR Tiene electrodoméstico: televisor	REFRIGERADOR Tiene electrodoméstico: refrigerador
FM_GRADO_ALCANZO_PAPA Grado más alto aprobado por su papá	Correlación de Pearson	1	.690**	.300**	.205**	.308**	.142**	.065**	.419**	.323**	.177**	.177**	.344**
	Sig. (bilateral)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	111726	107905	106614	98097	98784	108576	102599	109922	105015	107552	111726	111726
FM_GRADO_ALCANZO_MAMA Grado más alto aprobado por su mamá	Correlación de Pearson	.690**	1	.313**	.219**	.311**	.150**	.061**	.452**	.334**	.187**	.190**	.375**
	Sig. (bilateral)	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	107905	110155	105158	96648	97461	107124	101278	108447	103630	106139	110155	110155
Piso_recode ¿Qué material predomina en el piso de su casa?	Correlación de Pearson	.300**	.313**	1	.458**	.216**	.271**	.082**	.401**	.283**	.253**	.307**	.396**
	Sig. (bilateral)	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	106614	105158	111484	98288	99855	108305	102033	109989	104573	107307	111484	111484
Paredes_recode ¿Qué material predomina en las paredes de su casa?	Correlación de Pearson	.205**	.219**	.458**	1	.184**	.258**	.080**	.238**	.202**	.179**	.245**	.259**
	Sig. (bilateral)	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000
	N	98097	96648	98288	102526	90646	99696	94099	101216	96423	98832	102526	102526
Techo_recode ¿Qué material predomina en el techo de su casa?	Correlación de Pearson	.308**	.311**	.216**	.184**	1	.082**	.054**	.269**	.259**	.107**	.097**	.220**
	Sig. (bilateral)	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000
	N	98784	97461	99855	90646	103310	100500	94560	102031	96915	99510	103310	103310
CC_ELECTRICIDAD ¿Cuenta con electricidad?	Correlación de Pearson	.142**	.150**	.271**	.258**	.082**	1	.050**	.223**	.158**	.201**	.422**	.243**
	Sig. (bilateral)	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000
	N	108576	107124	108305	99696	100500	113186	104730	111959	107680	110047	113186	113186
CC_AMBIENTE_COCINA_SEPARADA ¿Hay un ambiente separado par la cocina?	Correlación de Pearson	.065**	.061**	.082**	.080**	.054**	.050**	1	.036**	.075**	.078**	.048**	.078**
	Sig. (bilateral)	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000
	N	102599	101278	102033	94099	94560	104730	106632	105529	101474	103627	106632	106632
Combustible_Cocina_recode ¿Qué combustible utiliza para cocinar?	Correlación de Pearson	.419**	.452**	.401**	.238**	.269**	.223**	.036**	1	.309**	.259**	.258**	.465**
	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000
	N	109922	108447	109989	101216	102031	111959	105529	114896	108144	110921	114896	114896

Fuente: Digeduca, 2014.

La tabla de resultados de SPSS se puede pegar en Excel con el objetivo de seleccionar con filtros solo las significancias, para ver qué correlación resulta significativa. En la **Tabla 11** se están analizando las correlaciones de las variables que contienen información del estudiante con el rendimiento en Lectura y Matemática. La sintaxis para generar las correlaciones se encuentran en el Anexo *Sintaxis de correlaciones*.

Tabla 11. Correlaciones de las variables parte 2

Correlaciones

		HABILIADAD_MATEMATICAS Habilidad en Matemáticas	HABILIDAD_LECTURA Habilidad en Lectura
FM_GRADO_ALCANZO_PAPA Grado	Sig. (bilateral)	0.0	0.0
FM_GRADO_ALCANZO_MAMA Grado	Sig. (bilateral)	0.000	0.000
Piso_recode ¿Qué material predomina	Sig. (bilateral)	.000	0.000
Paredes_recode ¿Qué material	Sig. (bilateral)	.000	0.000
Techo_recode ¿Qué material	Sig. (bilateral)	.000	0.000
CC_ELECTRICIDAD ¿Cuenta con	Sig. (bilateral)	.000	0.000
CC_AMBIENTE_COCINA_SEPARADA	Sig. (bilateral)	.007	.000
Combustible_Cocina_recode ¿Qué	Sig. (bilateral)	0.000	0.000
CC_CUENTA_LINEA_TELEFONICA	Sig. (bilateral)	.000	0.000
CC_TIENE_FAMILIA_CELULAR ¿Tiene	Sig. (bilateral)	.000	0.000
TELEVISOR Tiene electrodoméstico:	Sig. (bilateral)	.000	.000
REFRIGERADOR Tiene	Sig. (bilateral)	0.000	0.000
HABILIADAD_MATEMATICAS Habilidad	Sig. (bilateral)		0.000
HABILIDAD_LECTURA Habilidad en	Sig. (bilateral)	0.000	

Fuente: Digeduca, 2014.

Asimismo, se pueden ver las correlaciones de una variable contra todas las demás. En la **Tabla 12** aparece que utilizando un filtro se fue observando la significancia de cada correlación. Esta estrategia es útil para marcar con color aquellos datos que no fueron significativos en las correlaciones. Por ejemplo, en la información de « ¿Cuántos libros completos ha leído? », el dato de significancia indica que el coeficiente de correlación de Pearson no fue estadísticamente significativo con un valor de p de 0.05, en la correlación con habilidad en Matemática y la variable «Su familia tiene vehículo propio».

Tabla 12. Correlaciones de las variables parte 3

Correlaciones

		FM_GRADO_ALCANZO_PAPA Grado más alto aprobado por su papá	FM_GRADO_ALCANZO_MAMA Grado más alto aprobado por su mamá	DVD Tiene electrodomestico: DVD	LAVADORA_ROPA Tiene electrodomestico: lavadora de ropa	SECADORA_ROPA Tiene electrodomestico: secadora de ropa	HORNO_MICROONDAS Tiene electrodomestico: horno microondas	OTROS_ELECTRODOMESTICOS Tiene electrodomestico: otros	Sanitarios_recode ¿Tipo de sanitario de su casa?	CC_FAMILIA_AUTOMOVIL ¿Su familia tiene vehículo propio?	HABILIADAD_MATEMATICAS Habilidad en Matemáticas	HABILIDAD_LECTURA Habilidad en lectura	LC_CUANTOS_LIBROS_COMPLETOS_HA_LEIDO ¿Cuantos libros leyo el ultimo año?
FM_GRADO_ALCANZO_PAPA Grado	Sig. (bilateral)		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.000
FM_GRADO_ALCANZO_MAMA Grado	Sig. (bilateral)	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.000
Piso_recode ¿Qué material	Sig. (bilateral)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.000	0.000	.000
Paredes_recode ¿Qué material	Sig. (bilateral)	0.000	0.000	0.000	0.000	.000	0.000	.000	0.000	0.000	.000	0.000	.000
Techo_recode ¿Qué material	Sig. (bilateral)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.000	0.000	.000
CC_ELECTRICIDAD ¿Cuenta con	Sig. (bilateral)	0.000	0.000	0.000	.000	.000	0.000	.000	0.000	.000	.000	0.000	.000
CC_AMBIENTE_COCINA_SEPARADA	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.007	.000	.065
Combustible_Cocina_recode ¿Qué	Sig. (bilateral)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.000
CC_CUENTA_LINEA_TELEFONICA	Sig. (bilateral)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.000	0.000	.003
CC_TIENE_FAMILIA_CELULAR	Sig. (bilateral)	0.000	0.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	0.000	.000
TELEVISOR Tiene electrodoméstico:	Sig. (bilateral)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.000	.000	.000	.000
REFRIGERADOR Tiene	Sig. (bilateral)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.000
EQUIPO_SONIDO Tiene	Sig. (bilateral)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.000
VHS Tiene electrodoméstico: VHS	Sig. (bilateral)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.000
DVD Tiene electrodoméstico: DVD	Sig. (bilateral)	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.000
LAVADORA_ROPA Tiene	Sig. (bilateral)	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	.000
SECADORA_ROPA Tiene	Sig. (bilateral)	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	.000	0.000	.000
HORNO_MICROONDAS Tiene	Sig. (bilateral)	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	.000
OTROS_ELECTRODOMESTICOS	Sig. (bilateral)	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	.000
Sanitarios_recode ¿Tipo de sanitario	Sig. (bilateral)	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	.000
CC_FAMILIA_AUTOMOVIL ¿Su familia	Sig. (bilateral)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		.000	0.000	.017
HABILIADAD_MATEMATICAS	Sig. (bilateral)	0.000	0.000	0.000	0.000	.000	0.000	0.000	0.000	.000		0.000	.034
HABILIDAD_LECTURA Habilidad en	Sig. (bilateral)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		.000
LC_CUANTOS_LIBROS_COMPLETO	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000	.000	.017	.034	.000	

Fuente: Digeduca, 2014.

Este análisis también ayuda a realizar correlaciones para identificar variables que como investigadores se cree que son importantes, pero la significancia de la correlación de Pearson indica que el coeficiente no es significativo, por lo que se puede revisar la información de la variable y ver si es necesario recodificarla. Esta puede ser una de las razones del por qué no está correlacionando; por ejemplo en la **Tabla 13** se puede observar que no hay correlación entre la variable «agua» con la información de «paredes» y «techo de la casa».

Tabla 13. Correlaciones de las variables parte 4

Correlaciones

		P9_Grado_Mamá ¿Cuál fue el último grado de estudios de tu mamá?	P12_Grado_papá ¿Cuál fue el último grado de estudios de tu papá?	P36 ¿Hay libros en tu casa para leer?	P14_Piso_recode ¿Qué material predomina en el piso de tu casa?	P15_Paredes_recode ¿Qué material predomina en las paredes de tu casa?	P16_Techo_recode ¿Qué material predomina en el techo de tu casa?	P17_Agua_recode_2 ¿Cómo obtienen el agua en tu casa?	P18_iluminación_recode ¿Cómo se ilumina tu casa?	P19_1 Marca con una x los objetos que hay en tu casa: Auto	measure_M Estimated PERSON Measure: UIMEAN=.00 USCALE=1.00	measure_L Estimated PERSON Measure: UIMEAN=.00 USCALE=1.00
P9_Grado_Mamá ¿Cuál fue el último grado de estudios de tu mamá?	Correlación de Pears		.56	.08	.15	.18	.13	.06	.13	.22	.142	.199
	Sig. (bilateral)		0.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
P12_Grado_papá ¿Cuál fue el último grado de estudios de tu papá?	Sig. (bilateral)	0.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
P36 ¿Hay libros en tu casa para leer?	Sig. (bilateral)	.000	.000		.000	.000	.000	.000	.000	.000	.007	.000
P14_Piso_recode ¿Qué material predomina en el piso de tu casa?	Sig. (bilateral)	.000	.000	.000		.000	.000	.009	.000	.000	.000	.000
P15_Paredes_recode ¿Qué material predomina en las paredes de tu casa?	Sig. (bilateral)	.000	.000	.000	.000		.000	.189	.000	.000	.000	.000
P16_Techo_recode ¿Qué material predomina en el techo de tu casa?	Sig. (bilateral)	.000	.000	.000	.000	.000		.250	.000	.000	.000	.000
P17_Agua_recode_2 ¿Cómo obtienen el agua en tu casa?	Sig. (bilateral)	.000	.000	.000	.009	.189	.250		.000	.000	.000	.000
[illegible]	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000
P19_1 Marca con una x los objetos que hay en tu casa: Auto	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000
P19_Cocina_sumada	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
P19_5 Marca con una x los objetos que hay en tu casa: Camión, tractor	Sig. (bilateral)	.000	.000	.049	.000	.000	.000	.025	.089	.000	.561	.239
P19_6 Marca con una x los objetos que hay en tu casa: Refrigeradora	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000	.000	0.000	.000	.000
P19_7 Marca con una x los objetos que hay en tu casa: Televisión	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000	0.000	.000	.000	.000
P19_8 Marca con una x los objetos que hay en tu casa: Computadora	Sig. (bilateral)	.000	0.000	.000	.000	.000	.000	.000	.000	0.000	.000	.000
P19_9 Marca con una x los objetos que hay en tu casa: Equipo de sonido	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
P19_10 Marca con una x los objetos que hay en tu casa: Lavadora de ropa	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
P19_11 Marca con una x los objetos que hay en tu casa: Plancha eléctrica	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.028	0.000	.000	.000	0.000
P19_12 Marca con una x los objetos que hay en tu casa: Teléfono de línea fija	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
P19_13 Marca con una x los objetos que hay en tu casa: Teléfono celular	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.084	.000	.000	.000	.000
P19_14 Marca con una x los objetos que hay en tu casa: Videograbadora	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
P19_15 Marca con una x los objetos que hay en tu casa: Reproductor de DVD	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
measure_M Estimated PERSON Measure: UIMEAN=.00 USCALE=1.00	Sig. (bilateral)	.000	.000	.007	.000	.000	.000	.000	.000	.000		0.000
measure_L Estimated PERSON Measure: UIMEAN=.00 USCALE=1.00	Sig. (bilateral)	.000	.000	.000	.000	.000	.000	.000	.000	.000	0.000	

Fuente: Digeduca, 2014.

Recodificación de los valores de las variables

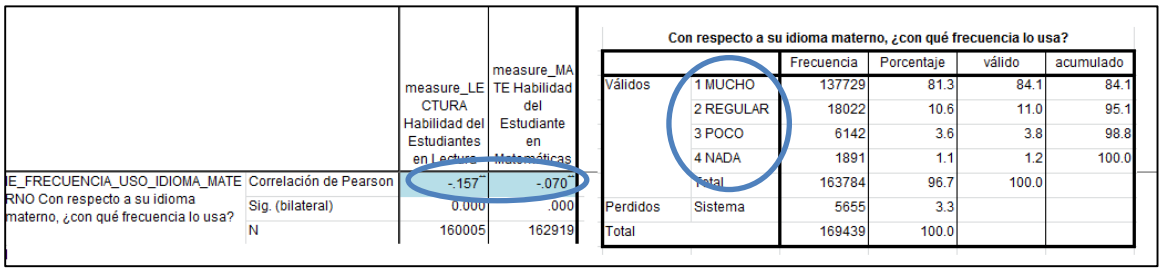
Recodificaciones generales

La dirección del coeficiente de correlación depende de los valores que componen la pregunta; por ejemplo en la **Figura 17**, en la variable de la frecuencia del uso del idioma materno, los coeficientes de correlación son negativos, esto se debe a que los valores que conforman la variable van de forma ascendente de «1 a 4», en donde uno es la característica «MUCHO» y cuatro es «NADA».

Con este ejemplo el coeficiente se interpreta de la siguiente forma: hay una influencia negativa en el rendimiento cuando los estudiantes dicen que usan mucho su idioma materno. Esto sucede porque el número menor pertenece a una opción que corresponde a la condición más alta de todas las otras opciones. Por lo que este análisis también ayuda a determinar que dicha variable deberá ser recodificada para que los valores asignados contribuyan a una fácil interpretación.

De manera que al tener los valores correctos el coeficiente de correlación saldría positivo para que pueda ser interpretado de la siguiente forma: existe una influencia positiva en el rendimiento en Lectura de los estudiantes que dicen que usan «MUCHO» su idioma materno.

Figura 17. Valores de las opciones y signo direccional del coeficiente de correlación



		measure_LECTURA Habilidad del Estudiantes en Lectura	measure_MATE Habilidad del Estudiante en Matemáticas
E_FRECUENCIA_USO_IDIOMA_MATERNO Con respecto a su idioma materno, ¿con qué frecuencia lo usa?	Correlación de Pearson	-.157**	-.070**
	Sig. (bilateral)	0.000	.000
	N	160005	162919

Con respecto a su idioma materno, ¿con qué frecuencia lo usa?

		Frecuencia	Porcentaje	válido	acumulado
Válidos	1 MUCHO	137729	81.3	84.1	84.1
	2 REGULAR	18022	10.6	11.0	95.1
	3 POCO	6142	3.6	3.8	98.8
	4 NADA	1891	1.1	1.2	100.0
	Total	163784	96.7	100.0	
Perdidos	Sistema	5655	3.3		
Total		169439	100.0		

Fuente: Digeduca, 2014.

Antes de recodificar los valores, se realiza un análisis cuidadoso de la variable. Hay variables cualitativas que no son continuas sino variables de clasificación o de atributos, es decir, que no tienen ninguna relación numérica pura. Estas variables pueden ser recodificadas con valores de «0» y «1» y se vuelven dicotómicas (uno o cero) de acuerdo a la aparición –o no– de la característica del atributo.

Hay variables que son fáciles de asignarles este tipo de valores porque la característica de desventaja tendrá valor «0» o porque se desean crear variables separadas para poder compararlas y no necesitan de un análisis a profundidad, y cuando se recodifican como dicotómicas ayudará a la interpretación del coeficiente en los modelos de regresión.

Es importante tomar en cuenta que al recodificar una variable a dicotómica, el valor «1» será asignado por el investigador, según el interés que tenga en el estudio de terminada opción que conforma la pregunta. La **Tabla 14** presenta algunas variables que se han utilizado en los análisis de Digeduca y que se han recodificado como dicotómicas.

Tabla 14. Valores para recodificar variables parte 1

Nombre de la variable	Valor original	Nombre de la variable nueva	Recodificación	Observación
Sector	21 Oficial 22 Privado 23 Municipal 24 Cooperativa	Sector Oficial Sector Privado Sector Municipal Sector Cooperativa	1=Oficial/0=Otro 1=Privado/0=Otro 1=Municipal/0=Otro 1=Cooperativa/0=Otro	Se separa cada variable
Jornada	31 Matutina 32 Vespertina 33 Doble 34 Nocturna 35 Intermedia	Jornada Matutina Jornada Vespertina Jornada Doble Jornada Nocturna Jornada Intermedia	1=Matutina/0=Otra 1=Vespertina/0=Otra 1=Doble/0=Otra 1=Nocturna/0=Otra 1=Intermedia/0=Otra	
Área	11 Urbana 12 Rural	Área	1=Urbana 0=Rural	
Identificación étnica	1 Maya 2 Ladino 3 Garífuna 4 Xinka 5 Otro Perdidos por el Sistema	Ladino_No_Ladino	1=Ladino 0 =Otro (Todas las demás etnias y Otro)	Lo datos perdidos se dejan como datos perdidos no como «0».
Idioma materno	1 Español 2 Maya 3 Extranjero 4 Garífuna 5 Xinka 6 Otro Perdidos por el Sistema	Idioma_Español_y_Otro	1=Español 0=Otro (Todas las demás etnias, otro y extranjero)	Lo datos perdidos se dejan como datos perdidos no como «0».

Fuente: Digeduca, 2014.

Otras variables que son fáciles de recodificar dado la característica y el valor por el que tienen que ser representadas, se pueden observar en la **Tabla 15**.

Tabla 15. Valores para recodificar variables parte 2

Nombre de la variable	Valor original	Nombre de la variable nueva	Recodificación	Observación
Grado que Repitió: **GR_PRIMERO** **GR_SEGUNDO** **GR_TERCERO** **GR_CUARTO** **GR_QUINTO** **GR_SEXTO**	Estas variables están separadas. 1 Sí 0 No	Suma_Grados_Repitió	00=Ninguno 01=Un grado repetido 02=Dos grados repetidos 03=Tres grados repetidos 04=Cuatro grados repetidos 05=Cinco grados repetidos 06=Seis grados repetidos	Se unifica la variable sumando todas.
Grado más alto aprobado por su papá/mamá	1 No sé 2 No estudió 3 Primaria 4 Básicos 5 Diversificado 6 Universidad Incompleta 7 Universidad Completa 1 No sé 2 No estudió 3 Primaria 4 Básicos 5 Diversificado 6 Universidad 7 Posgrado	Grado más alto aprobado por su papá/mamá	0=No estudió 1=Primaria 2=Básicos 3=Diversificado 4=Universidad Incompleta 5=Universidad Completa Perdidos=No sé y casos vacíos 0=No estudió 1=Primaria 2=Básicos 3=Diversificado 4=Universidad 5=Posgrado Perdidos=No sé y casos vacíos	El número 0 en este caso es una condición de ausencia.
Variables con Opciones en forma ascendente -ejemplo 1-	1 Mucho 2 Regular 3 Poco 4 Nada		3=Mucho 2=Regular 1=Poco 0=Nada	El número 0 en este caso es una condición de ausencia.
Variables con Opciones en forma ascendente -ejemplo 2-	1=3-4 Días a la semana 2=1-2 Días a la semana 3=A veces 4=Nunca		3=3-4 Días a la semana 2=1-2 Días a la semana 1=A veces 0=Nunca	El número 0 en este caso es una condición de ausencia.

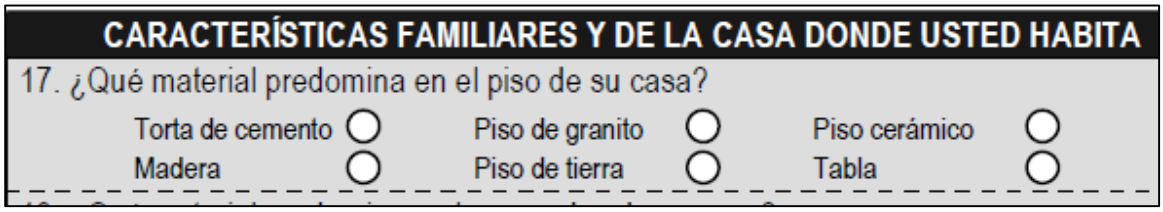
Fuente: Digeduca, 2014.

Recodificaciones de variables que conformarán el Nivel socioeconómico

Existen otras variables que necesitan un análisis un poco más profundo; tal es el caso de las que conformarán el nivel socioeconómico. Por ejemplo, ¿qué material predomina en el piso de su casa?, ¿qué material predomina en las paredes de su casa?, ¿qué material predomina en el techo de su casa?, ¿qué combustible utiliza para cocinar?, ¿qué agua utiliza para beber?, ¿cómo obtiene el agua que usa para lavar o echar en el baño?, entre otras.

Hay variables que se utilizarán en el nivel socioeconómico y que la pregunta original en la hoja para respuestas es presentada tal y como se muestra en la **Figura 18**.

Figura 18. Formato de preguntas relacionadas con el Nivel socioeconómico parte 1

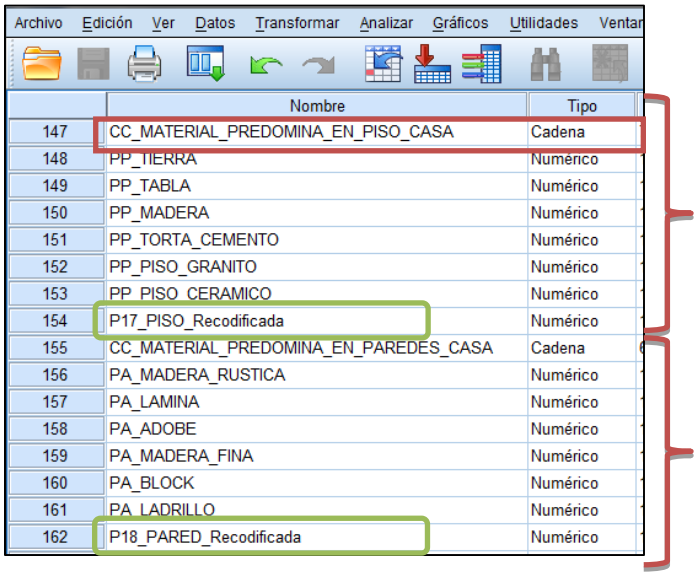


Fuente: Digeduca, 2014.

Cuando se digitalizan las respuestas, la base de datos puede tener varios formatos de la pregunta digitada; por ejemplo, el primer formato consiste en que la pregunta original sea de tipo cadena, como lo presenta la **Figura 19**. Esto quiere decir que si un estudiante marcó varias repuestas, esa variable está conformada por todas las respuestas que dio un solo alumno y están digitalizadas en letras, por lo que sería muy complicado utilizar esa variable en ese formato.

Sin embargo, en la digitalización, esas opciones de la pregunta original están separadas y son variables numéricas (ver la misma figura). Esto es importante tomarlo en cuenta ya que cada una de las opciones separadas, de la variable original, se tienen que unificar y colocarles los valores correspondientes, para utilizar la variable unificada.

Figura 19. Formato de preguntas relacionadas con el Nivel socioeconómico parte 2



Fuente: Digeduca, 2014.

En este aspecto, hay que considerar que algunas bases que se trabajan ya incluyen dentro de la base limpia, la variable recodificada. Únicamente se tendrían que verificar que los valores asignados sean los adecuados (**Figura 20**).

Figura 20. Formato de preguntas relacionadas con el Nivel socioeconómico parte 3

¿Qué material predomina en el piso de su casa? (Recodificada según valor máximo)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0 SIN RESPUESTA	524	.3	.3	.3
	1 PISO DE TIERRA	28910	15.1	15.1	15.4
	2 TABLA/MADERA	3604	1.9	1.9	17.3
	3 TORTA DE CEMENTO	72377	37.8	37.8	55.1
	4 PISO DE GRANITO	19520	10.2	10.2	65.3
	5 PISO CERÁMICO	66477	34.7	34.7	100.0
	Total	191412	100.0	100.0	

Fuente: Digeduca, 2014.

El segundo formato en que las preguntas pueden estar digitalizadas en la base de datos, se puede observar en la **Figura 21**. En este todas las opciones fueron digitadas en una misma variable. Estas variables definitivamente necesitan ser recodificadas, porque aunque estén todas las opciones bien definidas en una variable, los valores no están relacionados con la condición socioeconómica.

Figura 21. Formato de preguntas relacionadas con el Nivel socioeconómico parte 4

¿Qué material predomina en el piso de su casa?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 Cemento	33319	27.4	29.9	27.4
	2 Piso Cerámico	60809	50.0	54.5	77.4
	3 Piso Granito	1700	1.4	1.5	78.8
	4 Madera	2490	2.0	2.2	80.9
	5 Tierra	13166	10.8	11.8	91.7
	Total	111484	91.7	100.0	
Perdidos	Sistema	10110	8.3		
Total		121594	100.0		

Fuente: Digeduca, 2014.

En la **Figura 22** se pueden observar los valores asignados a cada opción. En el caso de la variable «tipo de sanitario en su casa», es necesario realizar una recodificación puesto que los valores asignados no representan el valor de la condición; por ejemplo, el valor 3 fue asignado a la opción «NO HAY», pero esta opción significa ausencia de lo que se está preguntando, por lo que el número que le corresponde sería 0 debido a que tiene menos valor que aquellos que respondieron que tenían «LETRINAS» o «INODORO CERÁMICO».

Figura 22. Frecuencias de variables

¿Tipo de sanitarios de su casa?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 INODORO CERÁMICO	94873	76.8	77.2	77.2
	2 LETRINAS	26803	21.7	21.8	99.0
	3 NO HAY	1243	1.0	1.0	100.0
	Total	122919	99.5	100.0	
Perdidos	Sistema	593	.5		
Total		123512	100.0		

Fuente: Digeduca, 2014.

Pero para determinar qué valor se le asignará a cada categoría, se deberá realizar un análisis previo de las variables. La asignación de un valor o símbolo debe ser realizada de acuerdo a reglas específicas, ya que de lo contrario, las medidas resultantes no podrán ser comparables porque cada una fue tomada desde un criterio específico (García, González & Ballesteros, 2001).

Estas reglas estarán sujetas a la decisión de la Subdirección de Análisis de Datos sobre cómo realizar el procedimiento.

Una regla es que los valores asignados a las categorías pueden estar apegados a la condición socioeconómica o precio en el mercado. Por ejemplo, a la variable de la **Figura 22** «tipo de sanitario en su casa», se le puede asignar los valores realizando el siguiente análisis.

–Razones por las cuáles el inodoro cerámico debe tener un valor mayor que la letrina–

–El inodoro cerámico es más costoso que la letrina. Cuando alguien tiene inodoro cerámico, tendrá también drenaje o tubería de agua, en la mayoría de casos. Por lo que la familia de ese estudiante invierte en la compra de un sanitario de cerámica, pero también podrían estar pagando un servicio de agua–.

En cualquier tipo de análisis hay que considerar las excepciones, por lo que el narrativo, lógico o empírico se deberá complementar con información cuantitativa y no solamente asignar un valor a conveniencia.

Por eso el análisis de los investigadores es importante, ya que puede generar una discusión del valor que se le asignará a cada característica, teniendo como fundamento que al realizar un nivel socioeconómico, se desea determinar qué condiciones van relacionadas con nivel socioeconómico bajo o alto.

Una técnica para el análisis de los valores de las opciones de que conforman la variable, además de generar correlaciones con el rendimiento, es generar una tabla de contingencia (la sintaxis se presenta en los anexos) de manera que se puedan visualizar las opciones de la variable y el promedio del rendimiento obtenido de los estudiantes que marcaron una opción específica.

En la **Tabla 16** se puede observar que los estudiantes que mencionaron tener piso de tierra y de madera, tienen los más bajos resultados en habilidad y los estudiantes que marcaron que tienen piso cerámico tienen resultados más altos en la habilidad. De esta manera se identifica que de alguna forma concuerda en que el niño que menciona tener un piso de mayor precio, puede tener mayores recursos, por lo que tiene mejores resultados.

Tabla 16. Tabla de contingencia con el rendimiento de los estudiantes

		Habilidad en Lectura	Habilidad en Matemáticas
		Media	Media
¿Qué material predomina en el piso de su casa?	1 Cemento	-.10	-.05
	2 Piso Cerámico	.28	.21
	3 Piso Granito	.13	.12
	4 Madera	-29	-.15
	5 Tierra	-.34	-.20

Fuente: Digeduca, 2014.

La información analizada en la tabla de contingencia ayudará a confirmar el valor que se le asignará a las opciones recodificadas (**Tabla 17**).

Tabla 17. Valores de codificación

Codificación original	Nueva codificación
1 Cemento	3
2 Piso cerámico	5
3 Piso granito	4
4 Madera	2
5 Tierra	1

Fuente: Digeduca, 2014.

Esta revisión de recodificación se va realizando con cada variable, asimismo se pueden hacer cruces de variables para observar el comportamiento de los datos, dadas ciertas combinaciones, como por ejemplo la **Figura 23** que muestra una tabla con el resultado en Lectura y Matemática, clasificado según los diferentes materiales que el estudiante reporta que componen la casa donde vive y clasificada según si el estudiante menciona que tiene lavadora o tiene secadora. Esto con el objetivo de analizar los datos y verificar la conformación de la base de

datos, de manera que tenga sentido la información que se observa y que proporcione más para la asignación de valores.

Figura 23. Cruce de variables

				LECTURA	_MATEMATIC
				Media	Media
LAVADORA_ROPA Tiene electrodoméstico: lavadora de ropa	0 NO	CC_MATERIAL_PREDOMINA_EN_PISO_CASA ¿Qué material predomina en el piso de su casa?	1 Cemento	-.12	-.06
			2 Piso	-.04	-.02
			3 Baldosas	.01	.04
			4 Madera	-.48	-.24
			5 Tierra	-.29	-.15
			6 Otros materiales	-.01	.01
	1 SI	CC_MATERIAL_PREDOMINA_EN_PISO_CASA ¿Qué material predomina en el piso de su casa?	1 Cemento	.10	.06
			2 Piso	.33	.18
			3 Baldosas	.50	.35
			4 Madera	.29	.28
			5 Tierra	-.38	-.10
			6 Otros materiales	.35	.19
SECADORA_ROPA Tiene electrodoméstico: secadora de ropa	0 NO	CC_MATERIAL_PREDOMINA_EN_PISO_CASA ¿Qué material predomina en el piso de su casa?	1 Cemento	-.10	-.05
			2 Piso	.04	.02
			3 Baldosas	.16	.13
			4 Madera	-.45	-.22
			5 Tierra	-.29	-.15
			6 Otros materiales	.11	.06
	1 SI	CC_MATERIAL_PREDOMINA_EN_PISO_CASA ¿Qué material predomina en el piso de su casa?	1 Cemento	.02	.08
			2 Piso	.37	.23
			3 Baldosas	.64	.46
			4 Madera	.49	.40
			5 Tierra	-.49	-.28
			6 Otros materiales	.37	.23

Fuente: Digeduca, 2014.

Elaboración de bitácora de trabajo

Simultáneamente al proceso de recodificación y análisis de las variables se deberá llevar una bitácora por escrito de los cambios que se sugieren a cada variable, y de las razones por las que se decidió no incluirla en el nivel socioeconómico. La **Tabla 18** muestra un ejemplo de observaciones escritas de algunas de las variables.

Se debe considerar que para cada base no serán las mismas decisiones de exclusión o conservación de la información, debido a que las preguntas del cuestionario de factores asociados de los estudiantes se van mejorando año con año y pueden ser diferentes en la

formulación y en las opciones presentadas en el cuadernillo para que sean respondidas por los estudiantes.

Tabla 18. Observaciones de los cambios en las variables

Variable	Fundamento de la decisión
Agua para lavar o echar en el baño	Se consideró redundante ya que se tiene una variable que indica si tiene servicio sanitario.
Cómo se moviliza usted para ir a su establecimiento	Se consideró no representativa ya que algunas personas podrían decidir caminar o llegar a su establecimiento en vehículo por preferencia y no por necesidad.
La casa donde vive su familia es...	Se consideró que la condición de propiedad, alquiler o préstamo no es equiparable a capacidad económica.
¿Cuántos niveles de construcción tiene su casa?	Una casa podía ser pequeña aunque tenga varios niveles.
¿Cuántas personas viven en su casa RECODE?	Se decidió utilizar variable de cadena para hacer un índice de hacinamiento cuando se compara con la variable «número de habitaciones de la casa».
Su familia recibe remesas	Recibir remesas no es exclusivo de personas en condiciones de pobreza.
Las variables cuántas personas viven en su casa (cadena) y variable de cuántas habitaciones tiene su casa que se utilizaría para hacer un índice de hacinamiento	No fue posible hacer el índice debido que la variable √habitaciones de la casa» es una variable de intervalo. Se eliminaron ambas variables.

Fuente: Digeduca, 2014.

Procedimiento de Análisis factorial exploratorio

Cuando se tienen las variables seleccionadas y recodificadas, se procede a realizar el procedimiento de reducción de dimensiones a través de un análisis factorial.

El análisis factorial es una técnica que simplifica la información de una matriz de correlaciones (Morales, 2007). El análisis factorial trata de establecer una estructura subyacente entre las variables del análisis, a partir de estructuras de correlación entre ellas; es decir, trata de encontrar correlaciones entre grupos de variables, las cuales agrupa en factores. Este procedimiento estadístico se utiliza para reducir la complejidad de un gran número de variables en un número más reducido (Méndez & Roldón, 2012).

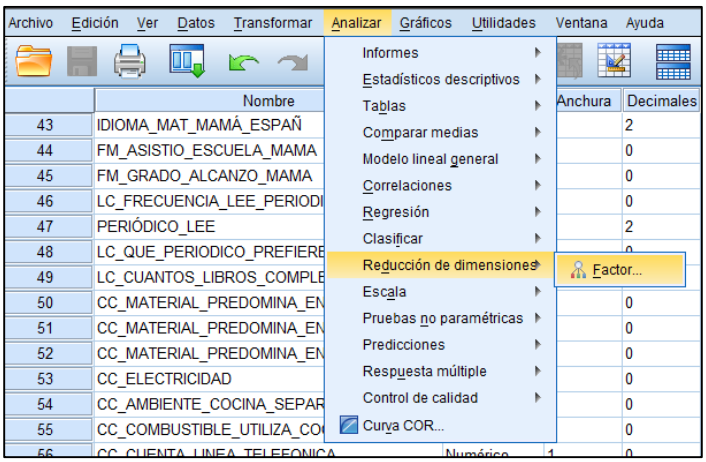
> El análisis factorial consta de cuatro fases características: el cálculo de una matriz capaz de expresar la variabilidad conjunta de todas las variables, la extracción del número óptimo de factores, la rotación de la solución para facilitar su interpretación y la estimación de las puntuaciones de los sujetos en las nuevas dimensiones (Manual SPSS, cap. 20).

A través de los cuestionarios de Factores Asociados, se recopila información acerca de las condiciones socioeconómicas del estudiante. Con esta información es posible utilizar el procedimiento de reducción de dimensiones porque el objetivo es emplear las variables que están relacionadas y que sustentan un constructo como el nivel socioeconómico.

Pasos para realizar un Análisis factorial exploratorio con el *software* SPSS

Con el programa SPSS, en la barra superior se selecciona la etiqueta de «Analizar», (**Figura 24**); cuando despliegan todas las opciones, se elige la opción de »Reducción de dimensiones» y finalmente la opción «Factor».

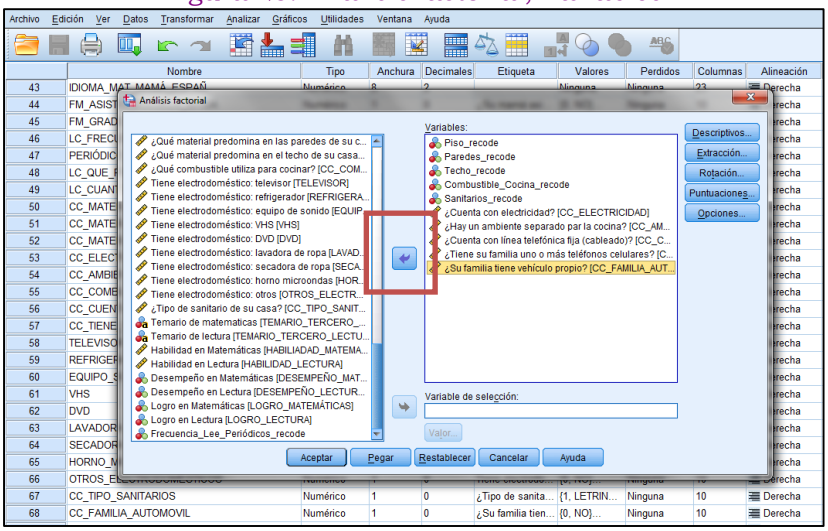
Figura 24. Selección de Análisis factorial



Fuente: Digeduca, 2014.

Surgirá un cuadro de diálogo, en donde se deben seleccionar las variables que conformarán el nivel socioeconómico. Al seleccionarlas, se trasladan, utilizando la flecha indicada en la **Figura 25**, hacia el espacio en blanco de la derecha con título *Variables*.

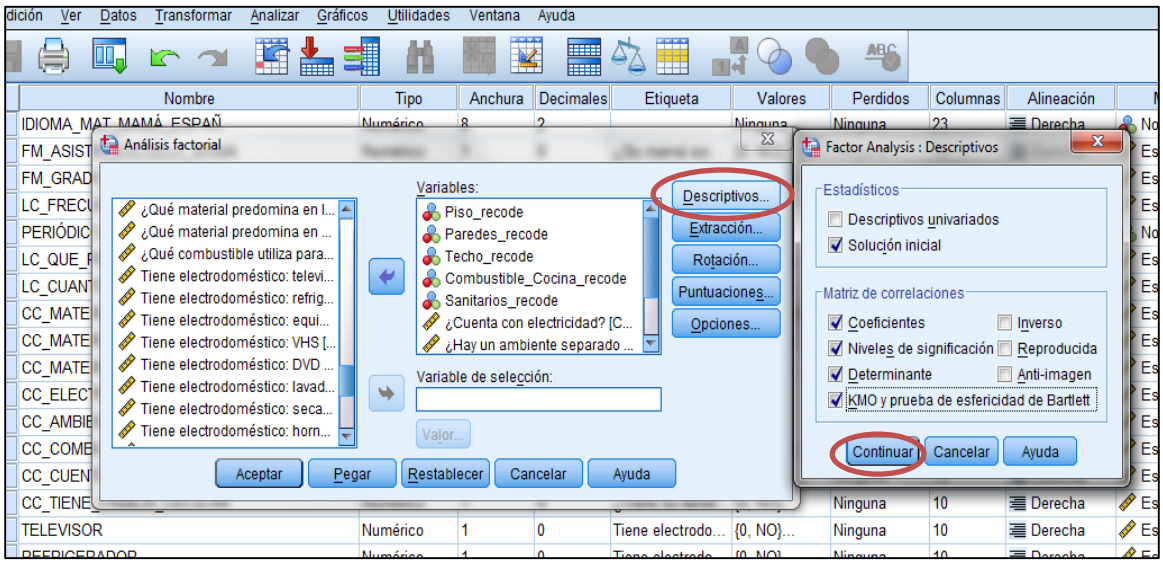
Figura 25. Análisis factorial, Variables



Fuente: Digeduca, 2014.

Se deberán seleccionar algunas opciones para que el análisis factorial proporcione la información que ayudará a realizar el análisis de los resultados. Para ello se inicia presionando el botón «Descriptivos» en donde se eligen las opciones «solución inicial, Coeficiente, Nivel de significancia, Determinante y KMO» y se debe presionar continuar para que las opciones elegidas queden guardadas (**Figura 26**).

Figura 26. Análisis factorial, Descriptivos



Fuente: Digeduca, 2014.

El capítulo 20 del *Manual de SPSS*, describe cada una de las opciones que se seleccionaron. Dicha descripción se presenta en la **Tabla 19**.

Tabla 19. Opciones de descriptivos del Análisis factorial

OPCIONES	DESCRIPCIÓN
La solución inicial	Permite obtener las comunalidades (proporción de la varianza que puede ser explicada por el modelo factorial generado) iniciales, los autovalores de la matriz analizada y los porcentajes de varianza a cada autovalor.
Coeficientes	Muestra la matriz con los coeficientes de correlación de las variables usadas.
Niveles de significación	Incluye en la matriz de correlaciones los niveles críticos unilaterales asociados a cada coeficiente.
Determinante	Muestra un valor que, si es próximo a cero, está indicando que las variables utilizadas están linealmente relacionadas, lo que significa que el análisis factorial es una técnica pertinente para analizar esas variables.
KMO	El KMO es una medida de adecuación muestral de (Kaiser-Meyer-Olkin) contrasta las correlaciones parciales entre las variables. Permite comparar la magnitud de los coeficientes de correlación parcial. El estadístico varía entre 0 y 1, los valores pequeños indican que el análisis factorial no es adecuado porque las correlaciones entre las variables no pueden ser explicadas por otras variables. Los valores menores de 0.5 indican que no se debe usar el análisis factorial con los datos que se analizan.
Prueba de esfericidad de Bartlett	Contrasta la hipótesis nula respecto a que la matriz de correlaciones es una matriz identidad, en cuyo caso no existirían correlaciones significativas entre las variables y el modelo factorial no sería pertinente.
	Fuente de información: Manual SPSS, cap. 20)

Fuente: Digeduca, 2014.

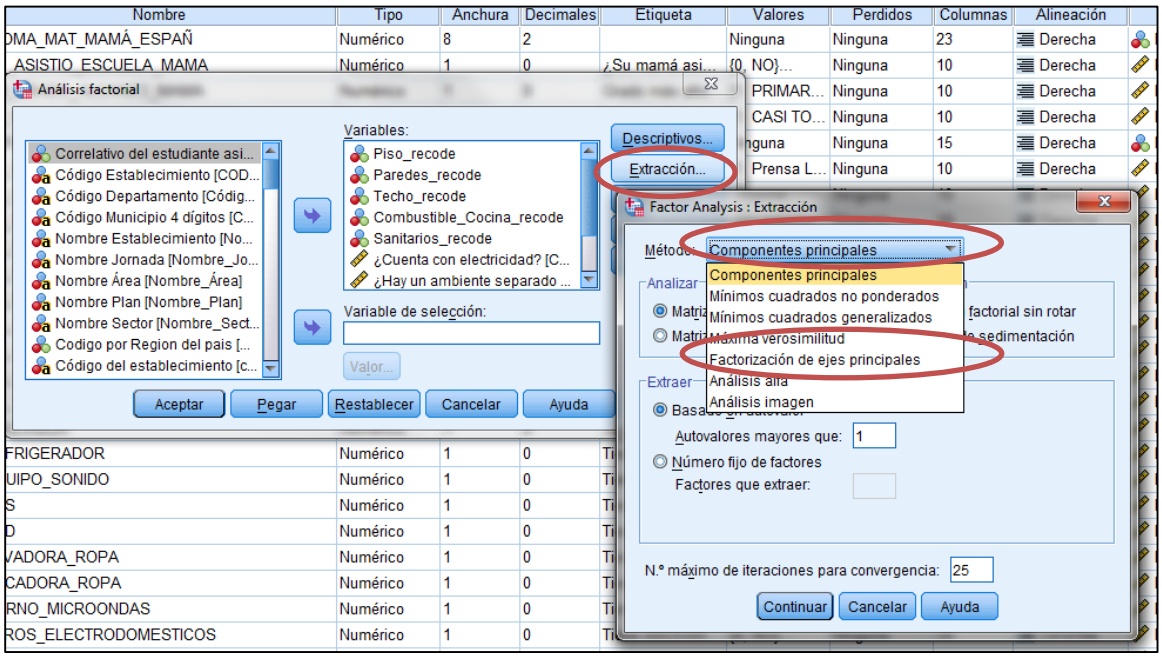
El siguiente paso lo muestra la **Figura 27**. Presionando la opción «Extracción», se elegirá el método por el cual se obtendrá la reducción de dimensiones de las variables. Se pueden utilizar varios métodos para el análisis factorial del nivel socioeconómico, los más usados son los métodos de componentes principales o factorización de ejes principales (Khan, 2006; citado en Pérez & Medrano, 2010).

Una de las ventajas del método de componentes principales es que siempre proporciona una solución, pero la desventaja es que al no estar basado en el modelo de análisis factorial, puede llevar a estimadores muy sesgados de la matriz de cargas factoriales, si existen variables con comunalidades bajas (Fernández, 2011).

La ventaja del método de ejes principales es que tiene una identidad fundamental de análisis factorial, por lo que puede proporcionar mejores estimaciones y la desventaja es que no se garantiza su convergencia (Fernández, 2011).

La mayoría de los investigadores utiliza el método de componentes principales porque siempre proporciona una solución.

Figura 27. Análisis factorial, Extracción



Fuente: Digeduca, 2014.

En el apartado «Analizar» (**Figura 28**), se deben dejar las opciones que ya vienen seleccionadas que son:

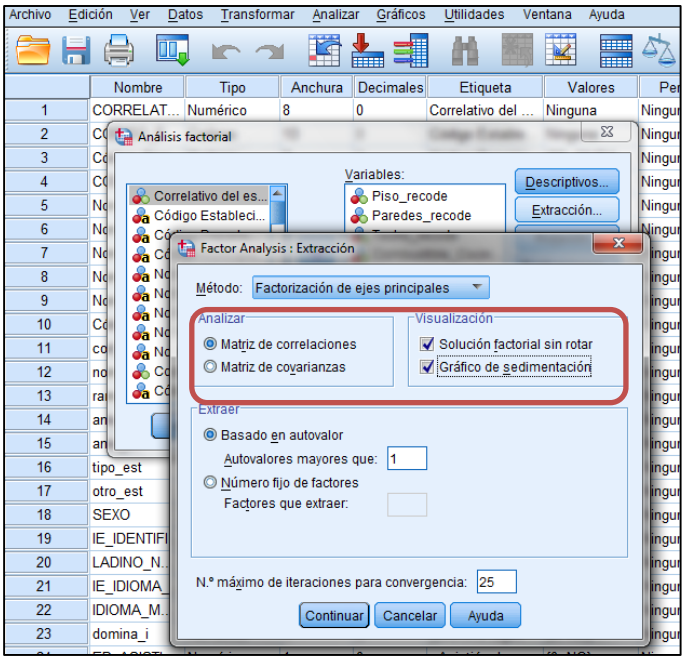
- **Matriz de correlaciones**: El análisis se basa en la matriz de correlaciones, en la matriz de correlaciones reducida, o en la matriz de correlaciones antimagen, según el método seleccionado.

- Y en el apartado «Visualización», la opción «**Solución factorial sin rotar**», la cual muestra las saturaciones factoriales sin rotar (la matriz de componentes o factorial), las comunalidades y los autovalores de la solución factorial (Manual SPSS, cap. 20).

Asimismo, se debe seleccionar la opción «Gráfico de sedimentación» (también llamado prueba de sedimentación de Cattell), el cual muestra una representación gráfica de la magnitud de los autovalores. El corte en la tendencia descendente sirve de regla para la determinación del número óptimo de factores que deben estar presentes en la solución. Siempre se muestra la representación de los autovalores de la matriz de correlaciones (o de covarianzas) originales,

independientemente del método de extracción seleccionado (Manual SPSS, cap. 20).

Figura 28. Análisis factorial, Extracción parte 2



Fuente: Digeduca, 2014.

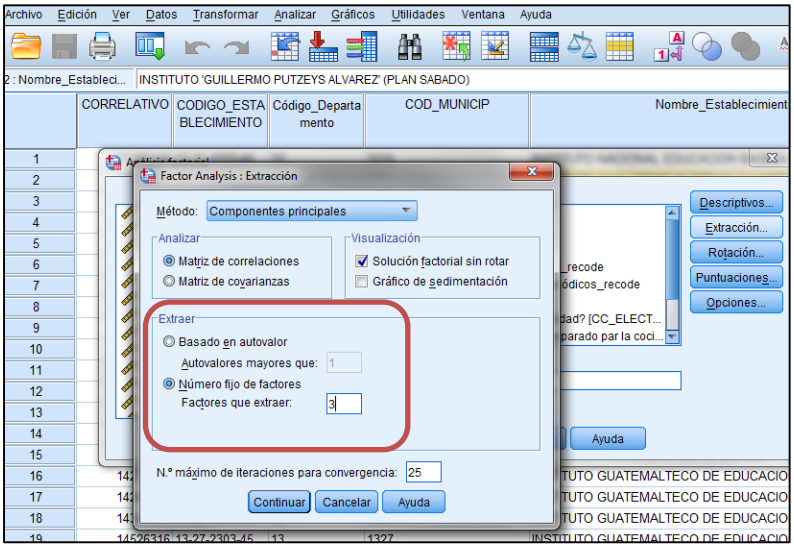
En análisis factorial, el factor constituye una agrupación de variables, y en el apartado «Extraer», las opciones permiten indicarle al programa la cantidad de factores en que tienen que agruparse las variables.

La opción «Basado en autovalor» es la que da libertad al programa de generar la cantidad de factores necesarios y la agrupación de las variables en cada uno de ellos. Sin embargo, hay que tener en cuenta que la razón del análisis factorial es reducir dimensiones por lo que se deben tener solo los factores necesarios que permitan explicar el constructo.

Existen casos por ejemplo en donde la agrupación de variables genera cuatro factores, pero el último factor solo contiene una variable, y dado que ya se decidió que en el análisis que se está realizando es indispensable la inclusión de esta variable, se podría utilizar la opción «Número fijo de factores». La **Figura 29** muestra que una vez seleccionada esa opción, se permite escribir la cantidad de factores que se desea obtener, y de esta manera no dejará aislada una variable en un solo factor, sino que la incluirá en los otros factores. Con esto se

está forzando la reducción de las variables y lo más recomendado sería excluir esa variable del análisis factorial.

Figura 29. Análisis factorial (continuación 1)



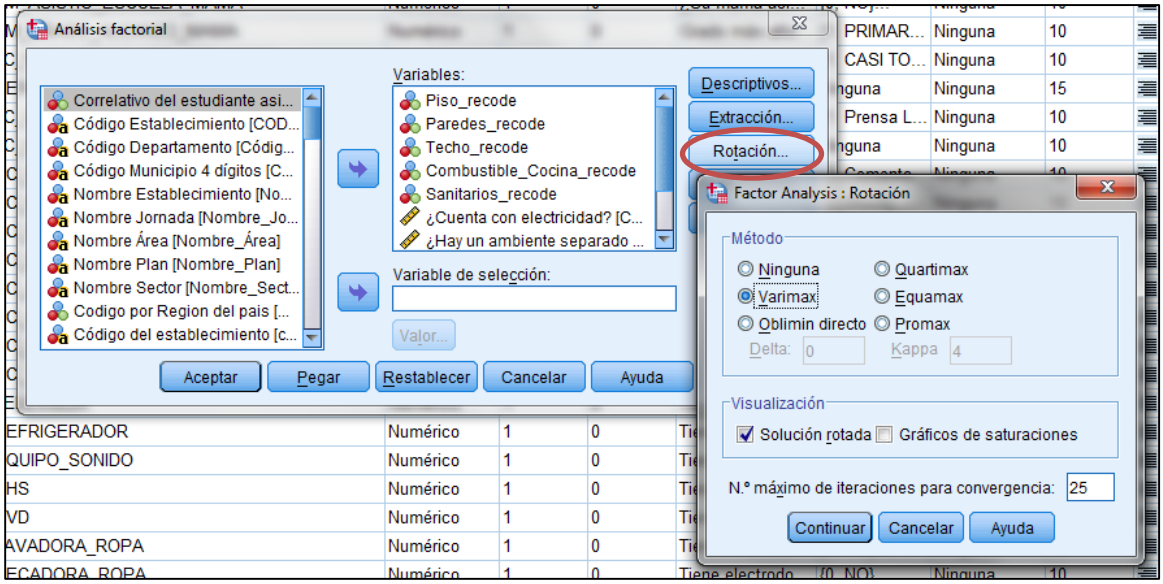
Fuente: Digeduca, 2014.

Rara vez los métodos de extracción de factores proporcionan matrices de cargas factoriales adecuadas para la interpretación. Por eso se utilizan métodos de rotación para hacer más fácil la interpretación.

Con la rotación «Varimax» se intenta simplificar las columnas (los factores; se maximiza la varianza de las columnas). En la construcción de escalas factoriales (con subconstructos bien diferenciados y replicables con más probabilidad en otras muestras) suele recomendarse la rotación Varimax (Nunnally, 1978: pp. 384-385; Kline, 1986: p. 191; Kline, 1994: pp. 68, 71 y 76; Rennie, 1997, citados en Morales, 2007).

Para seleccionar la opción «Varimax», se presiona el botón «Rotación» y en el apartado «Método» se selecciona la opción «Solución rotada». Está opción permite obtener una o más tablas con los resultados del proceso de rotación. Además, en la tabla de porcentajes de varianza explicada aparecen columnas adicionales que contienen la varianza total explicada por los factores rotados (**Figura 30**).

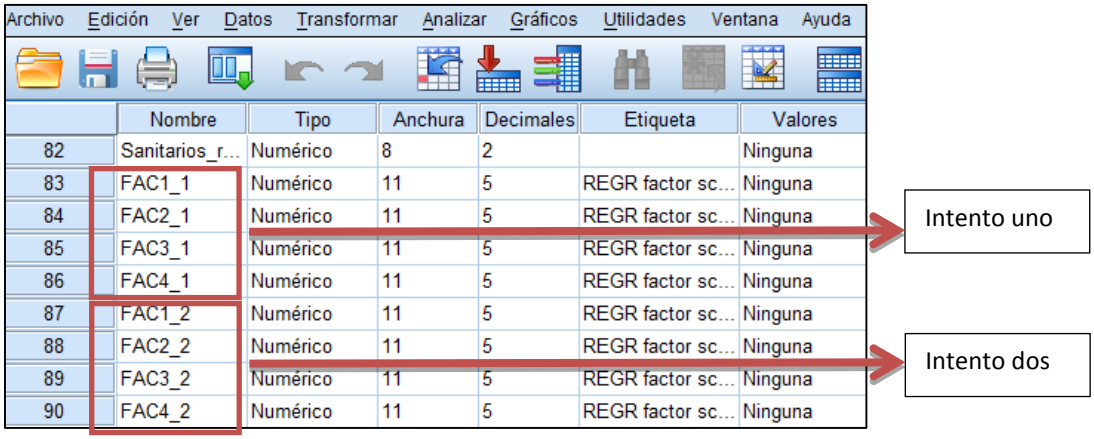
Figura 30. Análisis factorial, Rotación



Fuente: Digeduca, 2014.

Cuando ya se haya alcanzado la solución final, es decir ya se realizó el análisis de las variables que se deben incluir y la cantidad de factores finales que se van a utilizar, entonces se pueden guardar las puntuaciones resultantes del análisis factorial final. De lo contrario, cada vez que se genere el análisis, se guardarán tantos factores como intentos de análisis factorial se realicen y la base de datos tendrá una gran cantidad de nuevas variables como lo muestra la **Figura 31**.

Figura 31. Análisis factorial (continuación 2)



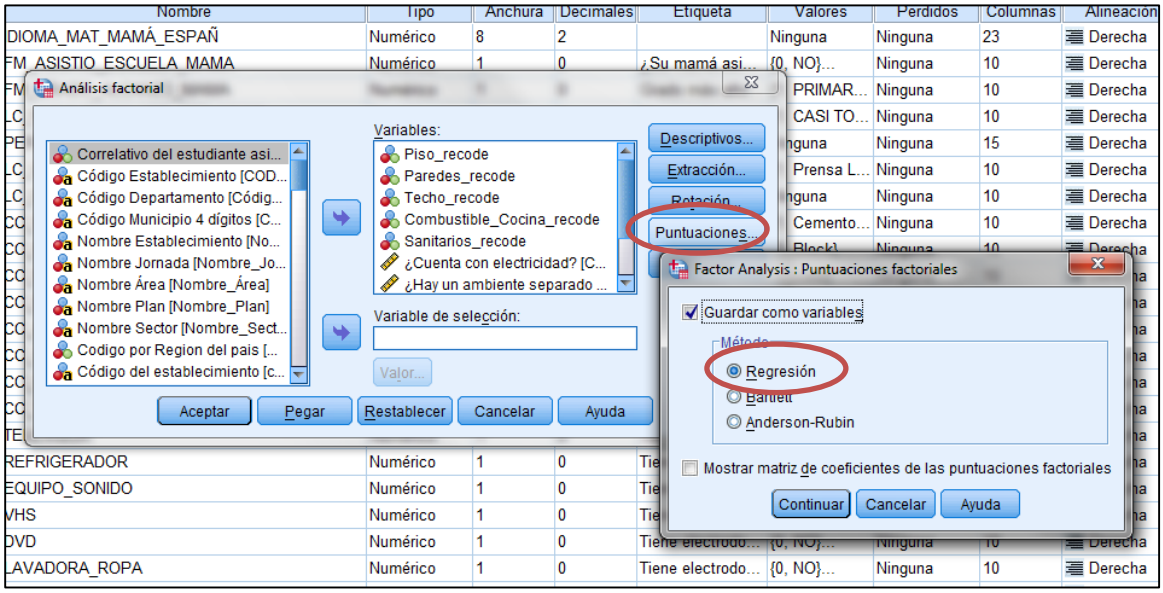
Fuente: Digeduca, 2014.

Para evitar lo anterior, es mejor tener terminado el análisis de las variables y finalmente se procede a guardar las puntuaciones de la siguiente manera (**Figura 32**).

En la opción «Puntuaciones», se selecciona «Guardar como variables» y en el apartado «Método» se elige la opción «Regresión». Esto generará puntuaciones estimadas para cada factor a cada caso de la base.

El método de regresión es un método que estima las puntuaciones factoriales en el que las estimaciones resultantes tienen una media de cero y una varianza igual al cuadrado de la correlación múltiple entre las puntuaciones factoriales estimadas y los valores factoriales verdaderos. Las puntuaciones factoriales estimadas con este método pueden estar correlacionadas incluso cuando los factores son ortogonales (Manual SPSS, cap. 20).

Figura 32. Análisis factorial, Puntuaciones



Fuente: Digeduca, 2014.

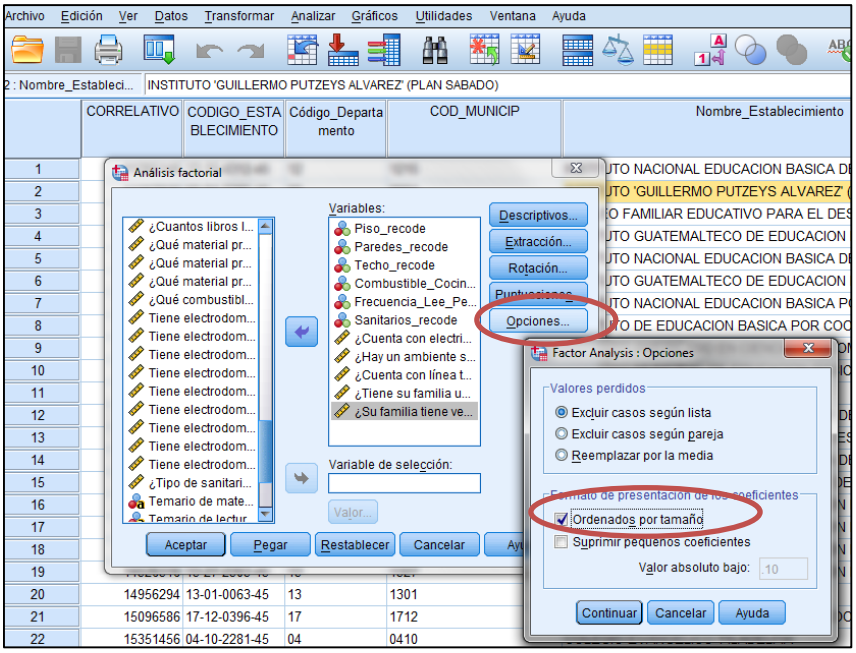
Para finalizar la elección de las opciones, hay que presionar el botón «Opciones» y a continuación «Excluir casos según lista». Esta es la opción que el programa trae por defecto y con ella se excluyen del análisis los sujetos que tengan valores perdidos en cualquiera de las variables trasladadas a la lista «Variables del cuadro de diálogo análisis

factorial», solo se incluyen en el análisis los casos completos (es decir, los casos con puntuación válida en todas las variables seleccionadas). Sin embargo, conviene tener en cuenta que esta forma de tratar los valores perdidos puede suponer la pérdida de un gran número de casos y la consiguiente reducción del tamaño efectivo de la muestra (Manual SPSS, cap. 20).

Asimismo, en el cuadro de diálogo que aparece cuando se presiona el botón «Opciones» en el apartado «Formato de visualización de los coeficientes», se selecciona la opción «Ordenados por tamaño» como se muestra en la **Figura 33**.

Esta opción sirve para ordenar las variables de las tablas de resultados en función de la magnitud (en valor absoluto) de los coeficientes de esas tablas (saturaciones, correlaciones, etc.). La ordenación se realiza de forma ascendente: primero las variables con coeficientes más altos (Manual SPSS, cap. 20).

Figura 33. Análisis factorial (continuación 3)



Fuente: Digeduca, 2014.

Resumen del procedimiento de Análisis factorial exploratorio

En la **Tabla 20** se presenta un resumen de las opciones a seleccionar en un análisis factorial exploratorio para la generación del Nivel socioeconómico.

Tabla 20. Resumen de pasos para el Análisis factorial exploratorio

En programa...	**Observación:**
Analizar	
Reducción de dimensiones	
Factor	Introducir las variables para Nivel socioeconómico
Submenú descriptivos	Seleccionar: • Solución inicial • Coeficiente • Nivel de significancia • Determinante • KMO
Submenú extracción	Seleccionar: • Factorización de ejes principales. (si converge, si no es así, utilizar el de Componentes Principales) • Matriz de correlaciones • Solución factorial sin rotar • Gráfico de Sedimentación
Submenú rotación	• Varimax • Solución rotada
Puntuaciones	• Guardar como variables • Regresión
Opciones	• Reemplazar por la media (si se desea que todos los casos tenga una puntuación para el nivel socioeconómico

Ejemplo de Sintaxis:

```
FACTOR
  /VARIABLES Piso_recodeParedes_recodeTecho_recodeCombustible_Cocina_recode
Frecuencia_Lee_Periódicos_recodeSanitarios_recode CC_ELECTRICIDAD
CC_AMBIENTE_COCINA_SEPARADA
    CC_CUENTA_LINEA_TELEFONICA CC_TIENE_FAMILIA_CELULAR CC_FAMILIA_AUTOMOVIL
  /MISSING LISTWISE
  /ANALYSIS Piso_recodeParedes_recodeTecho_recodeCombustible_Cocina_recode
Frecuencia_Lee_Periódicos_recodeSanitarios_recode CC_ELECTRICIDAD
CC_AMBIENTE_COCINA_SEPARADA
    CC_CUENTA_LINEA_TELEFONICA CC_TIENE_FAMILIA_CELULAR CC_FAMILIA_AUTOMOVIL
/PRINT INITIAL CORRELATION SIG DET KMO EXTRACTION ROTATION
  /PLOT EIGEN ROTATION
  /FORMAT SORT
  /CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PC
  /CRITERIA ITERATE(25)
/ROTATION VARIMAX
/SAVE REG(ALL)
/METHOD=CORRELATION.
```

Fuente: Digeduca, 2014.

Interpretación de las tablas y gráficas resultantes del procedimiento de análisis factorial exploratorio para la conformación del Nivel socioeconómico

Matriz de correlaciones

La primera tabla que se generará en análisis, es la Matriz de correlaciones. La **Tabla 21** muestra un ejemplo. En esta matriz lo que interesa es que las correlaciones entre variables sean lo más altas posibles.

Tabla 21. Matriz de correlaciones

Matriz de correlaciones[a]

		Piso_recode	Paredes_recode	Techo_recode	Combustible_Cocina_recode	Frecuencia_Lee_Periódicos_recode	Sanitarios_recode	CC_ELECTRICIDAD ¿Cuenta con electricidad?	CC_AMBIENTE_COCINA_SEPARADA ¿Hay un ambiente separado par la cocina?	CC_CUENTA_LINEA_TELEFONICA ¿Cuenta con línea telefónica fija (cableado)?	CC_TIENE_FAMILIA_CELULAR ¿Tiene su familia uno o más teléfonos celulares?	CC_FAMILIA_AUTOMOVIL ¿Su familia tiene vehículo propio?
Correlación	Piso_recode	1.000	.430	.229	.344	.037	.290	.252	.074	.255	.222	.263
	Paredes_recode	.430	1.000	.191	.223	.019	.199	.230	.069	.186	.152	.194
	Techo_recode	.229	.191	1.000	.290	.017	.215	.080	.066	.275	.115	.302
	Combustible_Cocina_recode	.344	.223	.290	1.000	.064	.425	.215	.021	.298	.260	.289
	Frecuencia_Lee_Periódicos_recode	.037	.019	.017	.064	1.000	.035	.012	.032	.064	.049	.055
	Sanitarios_recode	.290	.199	.215	.425	.035	1.000	.217	.021	.214	.212	.212
	CC_ELECTRICIDAD ¿Cuenta con electricidad?	.252	.230	.080	.215	.012	.217	1.000	.036	.138	.189	.097
	CC_AMBIENTE_COCINA_SEPARADA ¿Hay un ambiente separado par la cocina?	.074	.069	.066	.021	.032	.021	.036	1.000	.058	.060	.080
	CC_CUENTA_LINEA_TELEFONICA ¿Cuenta con línea telefónica fija (cableado)?	.255	.186	.275	.298	.064	.214	.138	.058	1.000	.136	.350
	CC_TIENE_FAMILIA_CELULAR ¿Tiene su familia uno o más teléfonos celulares?	.222	.152	.115	.260	.049	.212	.189	.060	.136	1.000	.198
	CC_FAMILIA_AUTOMOVIL ¿Su familia tiene vehículo propio?	.263	.194	.302	.289	.055	.212	.097	.080	.350	.198	1.000

Fuente: Digeduca, 2014.

Seguido de la Matriz de correlaciones está la información correspondiente a la significancia de las correlaciones. Lo importante en esta tabla es observar qué correlaciones son significativas, por lo tanto se debe elegir si se utilizará un valor de significancia menor de 0.05 o menor de 0.01. La **Tabla 22** muestra que todas las correlaciones de las variables son significativas al nivel de significancia de 0.01.

Otro dato importante es el que presenta la tabla de significancia en la parte inferior; este dato es llamado «Determinante» y dado que el número es un valor próximo a cero, está indicando que las variables utilizadas están linealmente relacionadas, lo que significa que el análisis factorial es una técnica pertinente para analizar esas variables.

Tabla 22. Matriz de significancia de las correlaciones

Sig. (Unilateral)	Piso_recode		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Paredes_recode	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
	Techo_recode	.000	.000		.000	.001	.000	.000	.000	.000	.000	.000
	Combustible_Cocina_recode	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000
	Frecuencia_Lee_Periódicos_recode	.000	.000	.001	.000		.000	.012	.000	.000	.000	.000
	Sanitarios_recode	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000
	CC_ELECTRICIDAD ¿Cuenta con electricidad?	.000	.000	.000	.000	.012	.000		.000	.000	.000	.000
	CC_AMBIENTE_COCINA_SEPARADA ¿Hay un ambiente separado par la cocina?	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000
	CC_CUENTA_LINEA_TELEFONICA ¿Cuenta con línea telefónica fija (cableado)?	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000
	CC_TIENE_FAMILIA_CELULAR ¿Tiene su familia uno o más teléfonos celulares?	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000
	CC_FAMILIA_AUTOMOVIL ¿Su familia tiene vehículo propio?	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	

a. Determinante = ,059

Fuente: Digeduca, 2014.

KMO y la prueba de Esfericidad de Barlett

El estadístico KMO varía entre 0 y 1. Los valores pequeños indican que el análisis factorial puede no ser una buena idea, dado que las correlaciones entre los pares de variables no pueden ser explicadas por otras variables. Los menores que 0.5 indican que no debe utilizarse el análisis factorial con los datos que se están analizando. Si el nivel crítico (Sig.) es mayor que 0.05, no se puede rechazar la hipótesis nula de esfericidad y, consecuentemente, no se pude asegurar que el modelo factorial sea adecuado para explicar los datos (Kerlinger, 2006). Ambas pruebas, la KMO y la de Esfericidad de Barlett, se utilizan para confirmar si el análisis factorial resulta adecuado para este estudio.

En la **Tabla 23** el valor de la adecuación de la muestra de KMO (Kaiser-Meyer-Olkin), es de 0.94, lo que implica que los datos introducidos se adecúan excelentemente para un análisis factorial. Además, se observa que el nivel crítico es Sig. = 0.000, por lo que el modelo factorial resulta apropiado para explicar los datos. Ambas pruebas, la KMO y la de Esfericidad de Barlett, confirman que el análisis factorial resulta adecuado para derivar información útil.

Tabla 23. KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin		**.940**
Prueba de Esfericidad de Bartlett	Chi-cuadrado aproximado	**721824.026**
	gl	**253**
	Sig.	**.000**

Fuente: Digeduca, 2014.

Comunalidades

La **Tabla 24** contiene las comunalidades asignadas inicialmente a las variables (inicial) y las comunalidades reproducidas por la solución factorial (extracción). Analizando las comunalidades de la extracción, se puede valorar cuáles de las variables son peor explicadas por el modelo.

En este análisis, la variable «Hay un ambiente separado para la cocina» es la peor explicada: el modelo solo es capaz de reproducir el 2.6 % de su variabilidad original, por lo que en función únicamente de esta tabla obtenida, se podría indicar que la variable con menor explicación de la varianza «Hay un ambiente separado para la cocina», podría quedar fuera del análisis; sin embargo, para ello es necesario considerar otros aspectos fundamentales del análisis factorial.

Por último, el porcentaje de la varianza que explica cada factor se aprecia en la **Tabla 25**. La misma indica que el factor uno explica el 11.95 % de la varianza total de todas las variables (la varianza total de la matriz de correlaciones), el segundo factor el 9.716 %, el tercer factor el 8.775 % y el cuarto factor 5.756 %. Se emplea esta columna debido a que se está trabajando con la matriz de componentes rotados.

Tabla 24. Comunalidades

Variables seleccionadas	Inicial	Extracción
GRAD_MAM (Grado más alto aprobado por la madre)	.304	.467
GRAD_PAP (Grado más alto aprobado por el padre)	.316	.467
pared (Material predominante en la pared de la vivienda)	.199	.251
piso (Material predominante en el piso de la vivienda)	.438	.544
techo (Material predominante en el techo de la vivienda)	.304	.354
agua_beb (Fuente de abastecimiento de agua para beber)	.314	.332
ELEC (Cuenta con electricidad)	.252	.357
CO_SEP (Hay un ambiente separado para la cocina)	.030	.026
COMBUST (Combustible que utiliza para cocinar)	.449	.472
LI_TEL (El hogar cuenta con línea telefónica fija)	.229	.249
FAM_CEL (La familia cuenta con uno o más teléfonos celulares)	.104	.119
T_SANIT (Tipo de sanitario de la casa)	.312	.346
FAM_veh (en su hogar se cuenta con automóvil)	.281	.306
Tv (En su hogar cuenta con televisor)	.332	.541
Refri (En su hogar cuenta con refrigeradora)	.411	.454
EQUI_SON (En su hogar cuenta con equipo de sonido)	.231	.279
VHS_DVD (En su hogar cuenta con reproductor VHS o DVD)	.314	.369
LAVAD (En su hogar cuenta con lavadora)	.441	.517
SEC_ROPA (En su hogar cuenta con secadora de ropa)	.209	.247
MICROO (En su hogar cuenta con microondas)	.410	.471
COMPU (En su hogar cuenta con computadora)	.390	.431
VIDEO_JU (En su hogar cuenta con consola de video-juego)	.271	.321
OTROS_EL (En su hogar cuenta con otros electrodomésticos)	.098	.101

Método de extracción: Factorización de Ejes principales.

Fuente: Digeduca, 2014.

La suma de las saturaciones al cuadrado de la rotación es mayor a uno (1), lo que es un indicador desde el punto de vista del número idóneo de factores. Finalmente los cuatro factores extraídos están explicando un 36.20 % la varianza, lo cual se considera bien, tomando en cuenta que Henson y Roberts en el 2006, en una revisión de 60 análisis factoriales, encontraron que la proporción media de varianza explicada por los factores es del 52.03 %.

Tabla 25. Varianza total explicada

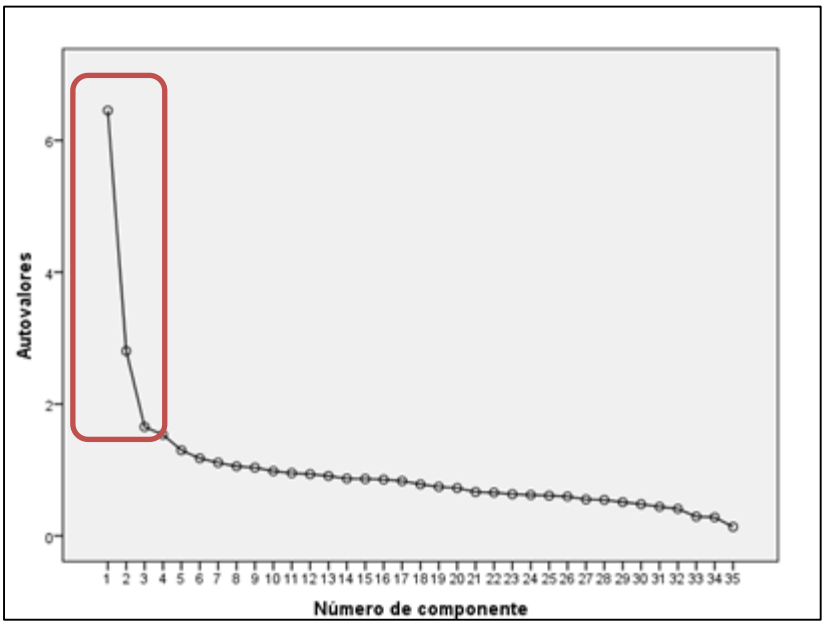
Factor	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6.772	29.445	29.445	6.185	26.891	26.891	2.750	11.958	11.958
2	1.813	7.881	37.325	1.261	5.482	32.374	2.235	9.716	21.674
3	1.088	4.732	42.058	.472	2.051	34.425	2.018	8.775	30.449
4	1.003	4.363	46.420	.409	1.779	36.204	1.324	5.756	36.204
5	.943	4.100	50.521						
6	.934	4.061	54.582						
7	.858	3.731	58.313						
8	.840	3.653	61.966						
9	.783	3.403	65.369						
10	.750	3.262	68.631						
11	.692	3.009	71.639						
12	.682	2.965	74.604						
13	.662	2.880	77.484						
14	.634	2.757	80.241						
15	.624	2.712	82.953						
16	.589	2.563	85.516						
17	.551	2.397	87.913						
18	.516	2.243	90.156						
19	.493	2.145	92.301						
20	.479	2.082	94.383						
21	.464	2.017	96.400						
22	.450	1.955	98.356						
23	.378	1.644	100.000						

Fuente: Digeduca, 2014.

Gráfico de sedimentación

En la **Figura 34** se presenta el gráfico de sedimentación. Cada círculo representa un factor y lo que se debe observar en este gráfico es la pendiente y en dónde se da el punto de corte de la misma. Del lado izquierdo del gráfico sobresalen dos puntos y luego la pendiente cambia drásticamente, por lo que se puede decir que el primer factor explica una cantidad mayor de varianza; los que le siguen son el segundo y tercer factor. Pero a partir del cuarto factor no hay cambios drásticos en la pendiente. Por lo tanto, con este criterio se debería incluir principalmente solo el primer factor (Zamora, Monroy & Chávez, 2009).

Figura 34. Gráfico de sedimentación



Fuente: Digeduca, 2014.

Matriz factorial

La **Tabla 26**, es la Matriz factorial, pero cuando el análisis extrae los factores, estos no son fácilmente interpretables y requiere habilidad del analista para visualizar un patrón o tendencia alguna. Por lo que se emplea frecuentemente una «rotación» de estos factores para superar esta dificultad.

Tabla 26. Matriz de factores

Variables	Factor			
	1	2	3	4
COMBUST	.681	.044	-.080	-.092
PISO	.662	.170	-.196	-.203
MICROO	.657	-.031	-.061	.185
LAVAD	.653	-.258	-.027	.151
COMPU	.647	-.093	-.009	.063
REFRI	.640	.200	-.047	.035
AGUA_BEB	.567	-.008	-.097	-.051
GRAD_PAP	.551	-.418	.250	-.215
T_SANIT	.546	.173	-.073	-.137
TECHO	.545	-.048	-.195	-.138
GRAD_MAM	.544	-.471	.275	-.184
VHS_DVD	.542	.182	.083	.183
FAM_VEH	.542	-.055	-.088	.091
VIDEO_JU	.518	-.104	.028	.201
LI_TEL	.479	-.150	-.035	.054
EQUI_SON	.438	.254	.051	.132
PARED	.412	.091	-.204	-.180
SEC_ROPA	.389	-.250	.061	.170
OTROS_EL	.308	-.042	.026	.054
FAM_CEL	.278	.202	.029	.028
CO_SEP	.140	.076	.007	.032
TV	.426	.508	.312	-.045
ELEC	.328	.418	.257	-.080

Fuente: Digeduca, 2014.

Matriz de factores rotados

Una matriz de factores principales y sus cargas explican la varianza del factor común de las puntuaciones de la prueba; pero en general no proporcionan estructuras con un significado científico.

Son las configuraciones de las pruebas o variables en el espacio factorial las que tienen una importancia fundamental. El proceso de rotación busca lo que Thurstone (1947, citado en la *Guía para el análisis de datos con SPSS*) denominó una estructura simple: variables que saturen, a ser posible, en un único factor, y factores que contengan un número reducido de variables que saturen inequívoca y

exclusivamente en ellos. Al realizar esta rotación en SPSS, se obtiene la siguiente matriz.

La **Tabla 27**, muestra que la matriz de factores rotados brinda una interpretación más simple de los factores. Es posible por tanto, identificar las ocho variables que correlacionan considerablemente con el factor 1, las seis que lo hacen con el factor 2, las siete que correlacionan fuertemente con el 3 y las dos que correlacionan fuertemente con el factor 4, las cuales han sido resaltadas intencionalmente para facilitar su comprensión.

Tabla 27. Matriz de factores rotados parte 1

	Factor			
	1	2	3	4
LAVAD	.604	.274	.084	.265
MICROO	.553	.308	.242	.107
VIDEO_JU	.497	.156	.169	.144
COMPU	.480	.323	.210	.229
SEC_ROPA	.443	.061	.017	.214
FAM_VEH	.427	.306	.156	.113
LI_TEL	.391	.244	.082	.193
OTROS_EL	.242	.122	.118	.114
PISO	.225	.618	.316	.119
TECHO	.274	.494	.103	.165
COMBUST	.350	.483	.289	.204
PARED	.114	.464	.143	.061
T_SANIT	.189	.441	.329	.116
AGUA_BEB	.327	.403	.193	.163
TV	.034	.118	.720	.073
ELEC	-.017	.106	.581	.074
VHS_DVD	.395	.177	.422	.039
EQUI_SON	.272	.174	.414	-.027
REFRI	.356	.393	.410	.061
FAM_CEL	.120	.145	.289	-.013
CO_SEP	.083	.066	.123	-.006
GRAD_MAM	.355	.161	.008	.689
GRAD_PAP	.321	.203	.037	.664

Fuente: Digeduca, 2014.

En la **Tabla 28** de factores rotados, también se puede identificar a qué constructo pertenece cada uno de los factores y las variables agrupadas en ellos. Por ejemplo la tabla siguiente, es una matriz de factores rotados que se pegó en Excel y se remarcaron las cargas factoriales intencionalmente. En ella se puede observar que el primer factor contiene variables pertenecientes a servicios (luz, teléfono, agua) en el segundo factor se agruparon variables de los objetos que tiene la familia (posesiones materiales) y en el tercer factor se encuentran las variables relacionadas con infraestructura de la casa (techo, paredes, piso).

Tabla 28. Matriz de factores rotados parte 2

	Factor		
	1	2	4
P18_iluminación_recode ¿Cómo se ilumina tu casa?	.544	.040	.140
P19_12 Marca con una x los objetos que hay en tu casa: Teléfono de línea fija	.528	.188	.124
P17_Agua_recode_2 ¿Cómo obtienen el agua en tu casa?	.478	.190	.146
P19_6 Marca con una x los objetos que hay en tu casa: Refrigeradora	.159	.520	.082
P19_13 Marca con una x los objetos que hay en tu casa: Teléfono celular	.082	.368	.129
P19_14 Marca con una x los objetos que hay en tu casa: Videograbadora	.044	.334	.001
P19_1 Marca con una x los objetos que hay en tu casa: Auto	.207	.233	.153
P19_8 Marca con una x los objetos que hay en tu casa: Computadora	.153	.220	.199
P19_10 Marca con una x los objetos que hay en tu casa: Lavadora de ropa	-.004	.137	.132
P19_5 Marca con una x los objetos que hay en tu casa: Camión, tractor	.001	.117	.028
P15_Paredes_recode ¿Qué material predomina en las paredes de tu casa?	.241	.042	.572
P16_Techo_recode ¿Qué material predomina en el techo de tu casa?	.050	.159	.365
P14_Piso_recode ¿Qué material predomina en el piso de tu casa?	.187	.050	.252

Servicios

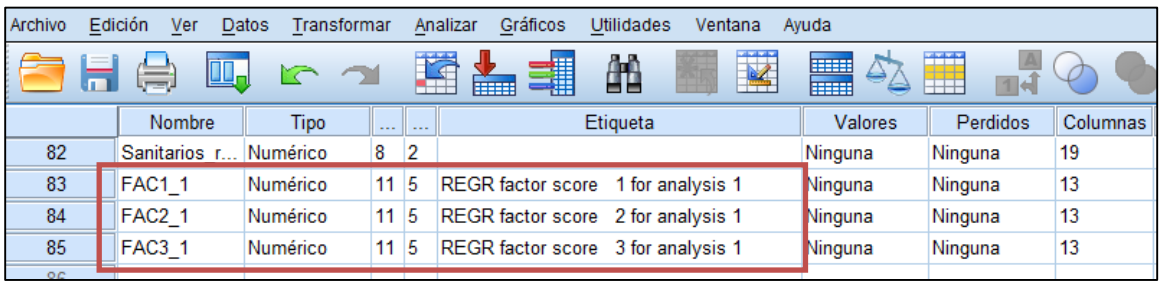
Posesiones Materiales

Infraestructura

Fuente: Digeduca, 2014.

En la base de datos, los factores se visualizarán como se presentan en la **Figura 35**. En la misma se observa que se generaron tres factores los cuales coinciden con los factores de la **Tabla 28**, que fue generada en el archivo de resultados y aparece en parte superior de este párrafo.

Figura 35. Factores generados en la base de datos parte 1

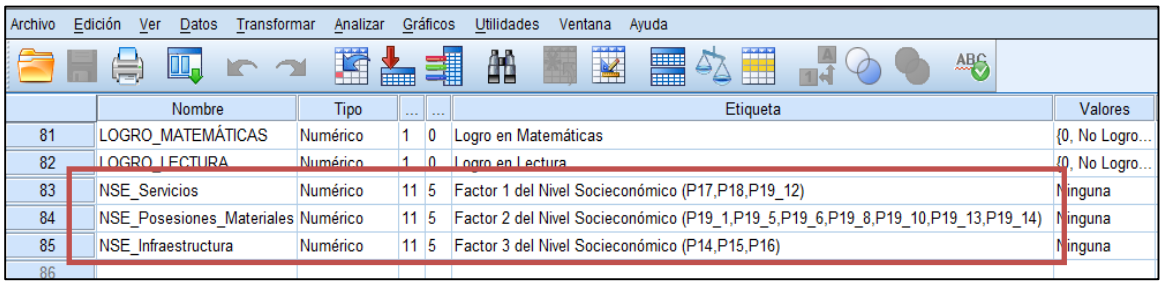


	Nombre	Tipo	...	...	Etiqueta	Valores	Perdidos	Columnas
82	Sanitarios_r...	Numérico	8	2		Ninguna	Ninguna	19
83	FAC1_1	Numérico	11	5	REGR factor score 1 for analysis 1	Ninguna	Ninguna	13
84	FAC2_1	Numérico	11	5	REGR factor score 2 for analysis 1	Ninguna	Ninguna	13
85	FAC3_1	Numérico	11	5	REGR factor score 3 for analysis 1	Ninguna	Ninguna	13

Fuente: Digeduca, 2014.

Para no confundir los factores, es adecuado que se le coloque nombre a cada factor generado con el objetivo de identificar cada uno de ellos. También se recomienda colocarles en la columna de etiquetas, el nombre de las preguntas que conformaron cada factor como se observa en la **Figura 36**.

Figura 36. Factores generados en la base de datos parte 2



	Nombre	Tipo	...	...	Etiqueta	Valores
81	LOGRO_MATEMÁTICAS	Numérico	1	0	Logro en Matemáticas	{0, No Logro...
82	LOGRO_LECTURA	Numérico	1	0	Logro en Lectura	{0, No Logro...
83	NSE_Servicios	Numérico	11	5	Factor 1 del Nivel Socieconómico (P17,P18,P19_12)	Ninguna
84	NSE_Posesiones_Materiales	Numérico	11	5	Factor 2 del Nivel Socieconómico (P19_1,P19_5,P19_6,P19_8,P19_10,P19_13,P19_14)	Ninguna
85	NSE_Infraestructura	Numérico	11	5	Factor 3 del Nivel Socieconómico (P14,P15,P16)	Ninguna
86						

Fuente: Digeduca, 2014.

Con los factores obtenidos, se genera una correlación entre cada factor con la habilidad de Lectura y Matemática, para verificar la relación que existe entre ellos, y finalmente se obtiene solo un factor que representará el Nivel socioeconómico.

Quedará en decisión del investigador o la Subdirección de Análisis de Datos, la forma de obtener un solo factor, ya sea sumando todos los factores generados, eligiendo el primer factor generado, seleccionando el factor que tenga mayor porcentaje de varianza explicada o el que correlacione de mejor forma con el resultado de Matemática y Lectura. Se deberá recordar que toda decisión deberá ser fundamentada con bases teóricas.

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.

Bourdieu, P. (2011). *Capital cultural, escuela y espacio soc*ial. (I. Jiménez, Trad.) México: Siglo veintiuno Editores.

Caro, D. (2009). Socio-economic Status and Academic Achievement Trajectories from Childhood to Adolescence. *Canadian Journal of Education 32, 3* (2009): 558-590. University of New Brunswick.

Cecchini, S. (2005). *Indicadores sociales en América Latina y el Caribe. Serie Estudios estadísticos y prospectivos*. Naciones Unidas. Santiago de Chile.

Cruz, A. & Santos, J. (2014). *Informe de resultados de Factores Asociados al rendimiento de los estudiantes de la Evaluación Nacional de Primaria 2010*. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación.

García, A.; Benítez, T.; Huerta, E.; Medina, I. & Ruiz, G. (2007). *Infraestructura escolar en las primarias y secundarias de México*. Instituto Nacional para la Evaluación de la Educación, INEE. Recuperado de: http://www.odisea.org.mx/Biblioteca/Educacion/Infra_escolar_prims_secs_INEE.pdf

García, J., González, M. & Ballesteros, B. (2001). *Introducción a la investigación en educación*. Madrid: Universidad Nacional de Educación a Distancia. Editorial IMPRESA.

Gil, J. (2011). *Medición del nivel socioeconómico familiar en el alumnado de educación primaria*. DOI: 10-4438/1988-592X-RE-2011-362-162.

Guía para el Análisis de datos con SPSS. (2002). Capítulo 20. Madrid: McGraw Hill.

Henson & Roberts (2006). Use of Exploratory Factor Analysis in Published Research. *Educational and Psychological Measurement. Vol 66 N.° 3,* pp. 393-416.

Hernández, S. (2011). *Análisis Factor*ial. Universidad Autónoma de Madrid.

Kerlinger, F. & Howard, L. (2006). Investigación del comportamiento. (Cuarta Edición). México: Editorial McGraw-Hill.

Lamont, M. & Lareau, A. (1988). Cultural Capital: Allusions, Gaps and Glissandos in Recent Theoretical. *Sociological Theory, 6(2)*, pp. 153-168.

Méndez, C. & Roldón, M. (2012). Introducción al análisis factorial exploratorio. *Red de Revistas Científicas de América Latina, el Caribe, España y Portugal, 41(1),* pp. 197-207.

Morales, P. (2007). *El Análisis Factorial en la construcción e interpretación de test, escalas y cuestionarios*. Madrid: Universidad Pontificia Comillas.

OCDE. (2010). *PISA 2009 Results: Overcoming Social Background – Equity in Learning Opportunities and Outcomes*. Obtenido de: http://dx.doi.org/10.1787/9789264091504-en

Pérez, E. & Medrano, L. (2010) Análisis Factorial Exploratorio: Bases Conceptuales y Metodológicas. *Revista Argentina de Ciencias del Comportamiento, 2010, Vol. 2, N°.1*, pp. 58-66.

Quim, M. & Santos, J. (2013). *Informe de Factores Asociados, ciclo básico, nivel medio 2009*. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación.

Saz, M. (2009) *Análisis de los Factores Asociados con el rendimiento de los estudiantes de tercero básico en Matemática y Lectura, del año 2006*. Tesis para obtener el grado de Master en la Universidad del Valle de Guatemala.

SIMCE. (2010). *Metodología de la construcción de grupos socioeconómicos. Unidad de Currículum y Evaluación*, Ministerio de Educación de Chile.

Stevens, S. (1951). *Handbook of Experimental Psychology*. Nueva York, NY: Wiley.

Willms, D. (2003). *Ten Hypotheses about Socioeconomic Gradients and Community Differences in Children's Developmental Outcomes*. Canadá: Human Resources Development Canada Publication Center.

Zamora, S.; Monroy, L. & Chávez, C. (2009). *Análisis factorial: una técnica para evaluar la dimensionalidad de las pruebas*. Cuaderno técnico N.° 6. México: Centro Nacional de Evaluación para la Educación Superior, A.C. (Ceneval).

Anexo I. Sintaxis para generar Frecuencias y Correlaciones

```
**GENERACIÓN DE FRECUENCIAS DE LAS VARIABLES**

FREQUENCIESVARIABLES=SEXO IE_IDENTIFICACION_ETNICA IE_IDIOMA_MATERNO domina_i
  ED_ASISTIO_PREPRIMARIA ED_AREA_DE_ESCUELA_INICIO_PRIMARIA ED_AREA_DE_ESCUELA_FINALIZO_PRIMARIA
  ED_REPITIO_ALGUN_GRADO_PRIMARIA ED_QUE_GRADOS_REPITIO GR_PRIMERO GR_SEGUNDO GR_TERCERO GR_CUARTO
  GR_QUINTO GR_SEXTO FM_IDIOMA_MATERNO_PAPA FM_ASISTIO_ESCUELA_PAPA FM_GRADO_ALCANZO_PAPA
  FM_IDIOMA_MATERNO_MAMA FM_ASISTIO_ESCUELA_MAMA FM_GRADO_ALCANZO_MAMA LC_FRECUENCIA_LEE_PERIODICO
  LC_QUE_PERIODICO_PREFIERE_LEER LC_CUANTOS_LIBROS_COMPLETOS_HA_LEIDO
  CC_MATERIAL_PREDOMINA_EN_PISO_CASA CC_MATERIAL_PREDOMINA_EN_PAREDES_CASA
  CC_MATERIAL_PREDOMINA_EN_TECHO_CASA CC_ELECTRICIDAD CC_AMBIENTE_COCINA_SEPARADA
  CC_COMBUSTIBLE_UTILIZA_COCINA CC_CUENTA_LINEA_TELEFONICA CC_TIENE_FAMILIA_CELULAR TELEVISOR
  REFRIGERADOR EQUIPO_SONIDO VHS DVD LAVADORA_ROPA SECADORA_ROPA HORNO_MICROONDAS
  OTROS_ELECTRODOMESTICOS CC_TIPO_SANITARIOS CC_FAMILIA_AUTOMOVIL
 /ORDER=ANALYSIS.

**GENERACIÓN DE CORRELACIONES**

CORRELATIONS
/VARIABLES=SEXO EDAD edad_recode TIENE_NECESIDAD_EDUCATIVA DISCAPACIDAD IE_IDENTIFICACION_ETNICA
  IE_IDIOMA_MATERNO IDIOM_MATE_HAB IDIOM_MATE_ENT IDIOM_MATE_LEE IDIOM_MATE_ESC
  IE_FRECUENCIA_USO_IDIOMA_MATERNO IDIOMA_SEG_LENGUA_ESPAÑOL IDIOMA_SEG_LENGUA_IDIOMA_MAYA
  IDIOMA_SEG_LENGUA_IDIOMA_EXTRANJERO IDIOMA_SEG_LENGUA_IDIOMA_GARIFUNA
  IDIOMA_SEG_LENGUA_IDIOMA_XINCA DOMINIO_SEG_LENGUA_ENTIENDO_OIRLO DOMINIO_SEG_LENGUA_ENTIENDO_LEERLO
  DOMINIO_SEG_LENGUA_PUEDO_ESCRIBIR DOMINIO_SEG_LENGUA_LO_HABLO ED_ASISTIO_PREPRIMARIA
  ED_AREA_DE_ESCUELA_FINALIZO_PRIMARIA ED_REPITIO_ALGUN_GRADO_PRIMARIA REPITIÓ_GR_PRIMERO
  REPITIÓ_GR_SEGUNDO REPITIÓ_GR_TERCERO REPITIÓ_GR_CUARTO REPITIÓ_GR_QUINTO REPITIÓ_GR_SEXTO
  ED_TIEMPO_CASA_AL_ESTABLECIMIENTO ED_TRABAJA_ACTUALMENTE ED_HORAS_TRABAJA_SEMANA
  ED_CONFORME_CON_SU_EDUCACION ED_DURACION_PERIODO_MATE_ESCUELA MINUTOS_PERIODO_MATE_RECODE
  ED_PERIODOS_MATE_SEMANA PERIODOS_SEMANA_MATE_RECODE ED_DURACION_PERIODO_LENGUAJE
  MINUTOS_PERIODOS_LENGUAJE_RECODE ED_PERIODOS_LENGUAJE_SEMANA PERIODOS_LENGUAJE_SEMANA_RECODE
  TEC_POSEEN_COMPUTADORAS_ESCUELA TEC_UTILIZA_COMPUTADORA_REALIZAR_TAREAS_ESCUELA
  TEC_HORAS_DIARIAS_USA_COMPUTADORA_EN_CASA TEC_HORAS_DIARIAS_USA_COMPUTADORA_EN_ESCUELA
  TEC_UTILIZA_INTERNET_REALIZAR_TAREAS_INVESTIGACION TEC_HAN_RECIBIDO_CAPACITACION_EN_COMPUTACION
  FM_IDIOMA_MATERNO_PAPA FM_ASISTIO_ESCUELA_PAPA FM_GRADO_ALCANZO_PAPA FM_IDIOMA_MATERNO_MAMA
  FM_ASISTIO_ESCUELA_MAMA FM_GRADO_ALCANZO_MAMA LC_FRECUENCIA_LEE_PERIODICO
  PREFIERE_LEER_PRENSA_LIBRE PREFIERE_LEER_SIGLO_XXI PREFIERE_LEER_EL_PERIODICO
  PREFIERE_LEER_DIARIO_CENTROAMERICA PREFIERE_LEER_NUESTRO_DIARIO PREFIERE_LEER_AL_DIA
  PREFIERE_LEER_LA_HORA PREFIERE_LEER_OTRO LIBROS_NINGUNO LIBROS_1a4 LIBROS_5a10 MAS_DE_10
  PELICULAS_NUNCA PELICULAS_VECES_AL_MES_1_4 PELICULAS_VECES_AL_MES_5_8 PELICULAS_MAS_DE_8
  PP_TORTA_CEMENTO PP_PISO_GRANITO PP_PISO_CERAMICO PP_MADERA PP_TIERRA PA_BLOCK PA_LADRILLO
  PA_LAMINA PA_MADERA_RUSTICA PA_MADERA_FINA PA_ADOBE PT_TERRAZA_FUNDIDA PT_DURALITA PT_LAMINA
  PT_TEJA PT_MATERIAL_PERECEDERO TUBERIA CHORRO_PUBLICO FUENTE_NATURAL POZO HERVIDA FILTRADA CLORADA
  COMPRADA DEL_CHORRO CC_ELECTRICIDAD CC_AMBIENTE_COCINA_SEPARADA CC_COMBUSTIBLE_UTILIZA_COCINA
  CC_CUENTA_LINEA_TELEFONICA CC_TIENE_FAMILIA_CELULAR CC_SERVICIO_TV_CABLE CC_SERVICIO_INTERNET
  TELEVISOR REFRIGERADOR EQUIPO_SONIDO VHS_DVD LAVADORA_ROPA SECADORA_ROPA HORNO_MICROONDAS
  COMPUTADORAS CONSOLA_VIDEOJUEGOS OTROS_ELECTRODOMESTICOS CC_TIPO_SANITARIOS CC_FAMILIA_AUTOMOVIL
  CC_COMO_SE_MOVILIZA_USUALMENTE CC_CASA_DONDE_VIVE_FAMILIA_ES CC_CASA_NIVELES_CONTRUCCION
  CC_CASA_CANTIDAD_DORMITORIOS CC_CUANTAS_PERSONAS_VIVEN_EN_CASA
  CASA_NIVELES_CONTRUCCION_RECODIFICADA CASA_CANTIDAD_DORMITORIOS_RECODIFICADA
  CUANTAS_PERSONAS_VIVEN_EN_CASA_RECODIFICADA CC_RECIBE_REMESAS MINUTOS_MATEMATICA_MODIFICADA
  MINUTOS_LECTURA_MODIFICADA PERIODOS_MATEMATICA_MODIFICADA PERIODOS_LECTURA_MODIFICADA
  TEMARIO_TERCERO_MATEMATICA TEMARIO_TERCERO_LECTURA TRC_LECTURA TRC_MATE PRC_MATE PRC_LECTURA
measure_LECTURAmeasure_MATE
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.
```

Anexo II. Sintaxis para recodificar valores de las variables

```
*RECODIFICACIÓN DE VALORES**

RECODE IE_IDENTIFICACION_ETNICA (2=1) (SYSMIS=SYSMIS) (ELSE=0) INTO LADINO_NO_LADINO.
EXECUTE.

RECODE IE_IDIOMA_MATERNO (1=1) (SYSMIS=SYSMIS) (ELSE=0) INTO IDIOMA_MAT_ESPAÑOL_OTRO.
EXECUTE.

RECODE ED_AREA_DE_ESCUELA_INICIO_PRIMARIA (1=1) (2=0) (SYSMIS=SYSMIS) INTO INICIÓ_PRIMARIA.
EXECUTE.

RECODE ED_AREA_DE_ESCUELA_FINALIZO_PRIMARIA (1=1) (2=0) (SYSMIS=SYSMIS) INTO FINALIZÓ_PRIMARIA.
EXECUTE.

RECODE FM_IDIOMA_MATERNO_PAPA (2=1) (SYSMIS=SYSMIS) (3=SYSMIS) (1=0) INTO IDIOMA_MAT_PAPÁ_ESPAÑ.
EXECUTE.

RECODE FM_IDIOMA_MATERNO_MAMA (2=1) (SYSMIS=SYSMIS) (3=SYSMIS) (1=0) INTO IDIOMA_MAT_MAMÁ_ESPAÑ.
EXECUTE.

*GRADOS QUE REPITIO: SUMA DE GRADOS REPETIDOS*

COMPUTE P9_GRREPSuma=SUM(GR_PRIMERO,GR_SEGUNDO,GR_TERCERO,GR_CUARTO,GR_QUINTO,GR_SEXTO).
EXECUTE.

VALUE LABELS P9_GRREPSuma
00 'NINGUNO'
01 'UN GRADO REPETIDO'
02 'DOS GRADOS REPETIDOS'
03 'TRES GRADOS REPETIDOS'
04 'CUATRO GRADOS REPETIDOS'
05 'CINCO GRADOS REPETIDOS'
06 'SEIS GRADOS REPETIDOS'.

VARIABLE LABELS
P9_GRREPSuma 'Suma de grados repetidos'.

**RECOFICACIÓN DE VARIABLES QUE CONFORMARÁN EL NIVEL SOCIOECONÓMICO**


RECODE CC_MATERIAL_PREDOMINA_EN_PISO_CASA (SYSMIS=SYSMIS) (1=3) (2=4) (3=5) (4=2) (5=1) (6=SYSMIS) INTO Piso_recode.
EXECUTE.

RECODE CC_MATERIAL_PREDOMINA_EN_PAREDES_CASA (SYSMIS=SYSMIS) (1=3) (2=4) (3=1) (4=2) (5=SYSMIS) INTO
Paredes_recode.
EXECUTE.

RECODE CC_MATERIAL_PREDOMINA_EN_TECHO_CASA (SYSMIS=SYSMIS) (1=4) (2=3) (3=1) (4=2) (5=SYSMIS) INTO Techo_recode.
EXECUTE.

RECODE CC_COMBUSTIBLE_UTILIZA_COCINA (SYSMIS=SYSMIS) (1=1) (2=2) (3=3) (4=SYSMIS) INTO Combustible_Cocina_recode.
EXECUTE.

RECODE CC_TIPO_SANITARIOS (SYSMIS=SYSMIS) (1=1) (2=2) (3=SYSMIS) INTO Sanitarios_recode.
EXECUTE.

RECODE LC_FRECUENCIA_LEE_PERIODICO (SYSMIS=SYSMIS) (1=4) (2=3) (3=2) (4=1) (5=0) INTO Frecuencia_Lee_Periódicos_recode.
EXECUTE.
```

Anexo III. Sintaxis para generar tablas con información de las variables e información de la habilidad que obtuvieron los estudiantes

* Tablas personalizadas*

```
CTABLES
  /VLABELS VARIABLES=SEXO LADINO_NO_LADINO IDIOMA_MAT_ESPAÑOL_OTRO domina_i ED_ASISTIO_PREPRIMARIA
    INICIÓ_PRIMARIA FINALIZÓ_PRIMARIA ED_REPITIO_ALGUN_GRADO_PRIMARIA ED_QUE_GRADOS_REPITIO
    P9_GRREPSuma IDIOMA_MAT_PAPÁ_ESPAÑ FM_ASISTIO_ESCUELA_PAPA FM_GRADO_ALCANZO_PAPA
    IDIOMA_MAT_MAMÁ_ESPAÑ FM_ASISTIO_ESCUELA_MAMA FM_GRADO_ALCANZO_MAMA LC_FRECUENCIA_LEE_PERIODICO
    PERIÓDICO_LEE LC_QUE_PERIODICO_PREFIERE_LEER LC_CUANTOS_LIBROS_COMPLETOS_HA_LEIDO
    CC_MATERIAL_PREDOMINA_EN_PISO_CASA CC_MATERIAL_PREDOMINA_EN_PAREDES_CASA
    CC_MATERIAL_PREDOMINA_EN_TECHO_CASA CC_ELECTRICIDAD CC_AMBIENTE_COCINA_SEPARADA
    CC_COMBUSTIBLE_UTILIZA_COCINA CC_CUENTA_LINEA_TELEFONICA CC_TIENE_FAMILIA_CELULAR TELEVISOR
    REFRIGERADOR EQUIPO_SONIDO VHS DVD LAVADORA_ROPA SECADORA_ROPA HORNO_MICROONDAS
    OTROS_ELECTRODOMESTICOS CC_TIPO_SANITARIOS CC_FAMILIA_AUTOMOVIL HABILIDAD_LECTURA
    HABILIADAD_MATEMATICAS
    DISPLAY=BOTH
  /TABLE SEXO + LADINO_NO_LADINO + IDIOMA_MAT_ESPAÑOL_OTRO + domina_i [C] + ED_ASISTIO_PREPRIMARIA
    [C] + INICIÓ_PRIMARIA + FINALIZÓ_PRIMARIA + ED_REPITIO_ALGUN_GRADO_PRIMARIA [C] +
    ED_QUE_GRADOS_REPITIO [C] + P9_GRREPSuma + IDIOMA_MAT_PAPÁ_ESPAÑ + FM_ASISTIO_ESCUELA_PAPA [C] +
    FM_GRADO_ALCANZO_PAPA [C] + IDIOMA_MAT_MAMÁ_ESPAÑ + FM_ASISTIO_ESCUELA_MAMA [C] +
    FM_GRADO_ALCANZO_MAMA [C] + LC_FRECUENCIA_LEE_PERIODICO [C] + PERIÓDICO_LEE +
    LC_QUE_PERIODICO_PREFIERE_LEER [C] + LC_CUANTOS_LIBROS_COMPLETOS_HA_LEIDO [C] +
    CC_MATERIAL_PREDOMINA_EN_PISO_CASA [C] + CC_MATERIAL_PREDOMINA_EN_PAREDES_CASA [C] +
    CC_MATERIAL_PREDOMINA_EN_TECHO_CASA [C] + CC_ELECTRICIDAD [C] + CC_AMBIENTE_COCINA_SEPARADA [C] +
    CC_COMBUSTIBLE_UTILIZA_COCINA [C] + CC_CUENTA_LINEA_TELEFONICA [C] + CC_TIENE_FAMILIA_CELULAR [C] +
    TELEVISOR [C] + REFRIGERADOR [C] + EQUIPO_SONIDO [C] + VHS [C] + DVD [C] + LAVADORA_ROPA [C] +
    SECADORA_ROPA [C] + HORNO_MICROONDAS [C] + OTROS_ELECTRODOMESTICOS [C] + CC_TIPO_SANITARIOS [C] +
    CC_FAMILIA_AUTOMOVIL [C] BY HABILIDAD_LECTURA [MEAN] + HABILIADAD_MATEMATICAS [MEAN]
  /CATEGORIES VARIABLES=SEXO domina_i ED_ASISTIO_PREPRIMARIA ED_REPITIO_ALGUN_GRADO_PRIMARIA
    ED_QUE_GRADOS_REPITIO P9_GRREPSuma FM_ASISTIO_ESCUELA_PAPA FM_GRADO_ALCANZO_PAPA
    FM_ASISTIO_ESCUELA_MAMA FM_GRADO_ALCANZO_MAMA LC_FRECUENCIA_LEE_PERIODICO
    LC_QUE_PERIODICO_PREFIERE_LEER CC_MATERIAL_PREDOMINA_EN_PISO_CASA
    CC_MATERIAL_PREDOMINA_EN_PAREDES_CASA CC_MATERIAL_PREDOMINA_EN_TECHO_CASA CC_ELECTRICIDAD
    CC_AMBIENTE_COCINA_SEPARADA CC_COMBUSTIBLE_UTILIZA_COCINA CC_CUENTA_LINEA_TELEFONICA
    CC_TIENE_FAMILIA_CELULAR TELEVISOR REFRIGERADOR EQUIPO_SONIDO VHS DVD LAVADORA_ROPA SECADORA_ROPA
    HORNO_MICROONDAS OTROS_ELECTRODOMESTICOS CC_TIPO_SANITARIOS CC_FAMILIA_AUTOMOVIL ORDER=A KEY=VALUE
    EMPTY=INCLUDE
  /CATEGORIES VARIABLES=LADINO_NO_LADINO IDIOMA_MAT_ESPAÑOL_OTRO INICIÓ_PRIMARIA FINALIZÓ_PRIMARIA
    IDIOMA_MAT_PAPÁ_ESPAÑ IDIOMA_MAT_MAMÁ_ESPAÑ PERIÓDICO_LEE LC_CUANTOS_LIBROS_COMPLETOS_HA_LEIDO
ORDER=A KEY=VALUE EMPTY=EXCLUDE.
```

Serie
de Cuadernillos Técnicos

Elaboración del Nivel socioeconómico

GOBIERNO DE LA REPÚBLICA DE GUATEMALA
MINISTERIO DE EDUCACIÓN

Dirección General de Evaluación
e Investigación Educativa
-Digeduca-