



# Serie de Cuadernillos Técnicos

## Generación de bases de datos para reportes por establecimiento



Licenciada Cynthia del Aguila Mendizábal  
**Ministra de Educación**

Licenciada Evelyn Amado de Segura  
**Viceministra Técnica de Educación**

Licenciado Alfredo Gustavo García Archila  
**Viceministro Administrativo de Educación**

Doctor Gutberto Nicolás Leiva Alvarez  
**Viceministro de Educación Bilingüe e Intercultural**

Licenciado Eligio Sic Ixpancoc  
**Viceministro de Diseño y Verificación de la Calidad Educativa**



**Directora** Lcda. Luisa Fernanda Müller

Subdirección de Análisis de Datos

**Autoría**

M.A. José Adolfo Santos Solares  
Lcda. Ana Aidé Cruz Grünebaum  
M. A. Ana Lucía Morales Sierra

**Revisión de texto y diagramación**

Lcda. María Teresa Marroquín Yurrita

**Diseño de portada**

Lic. Roberto Franco Arias

Dirección General de Evaluación e Investigación Educativa

© Dgeduca 2015 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: Santos, J.; Cruz, A. y Morales, A. (2014). *Generación de bases de datos para reportes por establecimiento*. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación.

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

Impreso en Guatemala

[divulgacion\\_dgeduca@mineduc.gob.gt](mailto:divulgacion_dgeduca@mineduc.gob.gt)

Guatemala, 2015

# CONTENIDO

INTRODUCCIÓN .....	5
ELABORACIÓN DE LISTA DE VARIABLES .....	6
PREPARACIÓN DE BASE DE DATOS.....	8
VARIABLES DICOTÓMICAS .....	8
VARIABLES DE RESPUESTA MÚLTIPLE .....	9
VARIABLES CONTINUAS.....	12
VARIABLES DE TEXTO.....	13
LOGRO NACIONAL, DEPARTAMENTAL Y MUNICIPAL.....	15
CREACIÓN DE BASE POR ESTABLECIMIENTO.....	16
MASTERS CON INFORMACIÓN ADICIONAL .....	22
DISPOSICIONES FINALES.....	25
VARIABLES DE APROXIMACIÓN.....	29

## LISTA DE FIGURAS

Figura 1. Sintaxis para desglosar opciones de respuesta en nuevas variables .....	11
Figura 2. Sintaxis para verificar que el desglose de variables esté bien .....	12
Figura 3. Sintaxis para desglosar opciones de respuesta de rama de enseñanza .....	14
Figura 4. Datos de estudiantes en el establecimiento por rama educativa .....	14
Figura 5. Sintaxis para indicar quienes no indicaron forma en la prueba de Matemáticas .....	15
Figura 6. Resultados comparativos en el reporte por establecimiento .....	16
Figura 7. Función agregar datos y elegir el primer dato en SPSS .....	18
Figura 8. Función agregar datos y elegir el promedio en SPSS .....	19
Figura 9. Función agregar datos y elegir el promedio para edad y períodos a la semana en SPSS .....	20
Figura 10. Función agregar datos y elegir suma para variables de conteo en SPSS .....	21
Figura 11. Establecimientos que no se evaluaron en años anteriores .....	24
Figura 12. Sintaxis para obtener resultados de desempeño válido .....	25
Figura 13. Verificación de porcentaje por desempeño válido en SPSS .....	26
Figura 14. Sintaxis para obtener variables de desempeño tomando en cuenta no calificados .....	26
Figura 15. Verificación de porcentaje por desempeño con quienes no se pudo generar calificación en SPSS .....	27
Figura 16. Presentación de porcentajes de desempeño y Logro en el reporte del establecimiento .....	28
Figura 17. Recodificación de porcentaje de Logro, reduciendo la escala a 10 .....	30
Figura 18. Variable nueva reduciendo la escala de porcentaje a 10 categorías en la base de datos .....	30

## LISTA DE TABLAS

Tabla 1. Ejemplo de lista de variables que se incluirán en el informe .....	7
Tabla 2. Tabla de frecuencias de una variable dicotómica .....	9
Tabla 3. Tabla de frecuencias de una variable múltiple .....	10
Tabla 4. Estadísticos de resumen de una variable que se necesita el promedio .....	13

# INTRODUCCIÓN

La elaboración de los reportes individuales por establecimientos de graduandos y tercero básico es un proceso que se realiza conjuntamente con la Unidad de Divulgación de resultados, quienes se encargan de elaborar y diagramar el contenido del reporte. A esta unidad se le proporciona las bases de datos preparadas para la generación de dichos informes.

Estos reportes una vez generados, se entregan a los establecimientos educativos evaluados de manera impresa y son publicados en la página de Internet de la Dgeduca para consulta de la comunidad educativa sobre los resultados obtenidos, que apoyen a la toma de decisiones.

Antes de iniciar el proceso se lleva a cabo una reunión entre las personas que estarán trabajando y quienes serán responsables del proceso, de manera que en la subdirección de Análisis de datos se conozca el contenido que tendrá el informe para graduandos y/o para tercero básico, con el fin de identificar si la información que aparecerá en el informe está disponible en los datos recolectados en la evaluación y en el cuestionario de factores asociados, de manera que todo esté incluido en la base de datos que se generará.

## ELABORACIÓN DE LISTA DE VARIABLES

Al tener definida la información, se realiza una revisión de las variables propuestas para verificar si están disponibles o si necesitan algún procesamiento adicional; por ejemplo, el efecto de las variables en un modelo de factores asociados, si son modificadas con respecto al año anterior o si es información que se puede conseguir por medio de otras fuentes (como *Business Objects*). Las variables se identifican al colocarles el nombre con el que aparecerán en la base de datos que se entregará para mantener un control y orden de las mismas.

La Tabla 1 presenta un ejemplo de la lista de variables que se incluyen en la base de datos, indicando nombre de la variable, su descripción y en observaciones se indica la fuente de donde se adquiere o si tiene consideraciones especiales para ser tomadas en cuenta.

Como se puede observar en la tabla<sup>[1]</sup>, las variables de la 10 a la 16 son generadas con datos del *Business Objects* para cada establecimiento, ya que esa información no se obtiene de las pruebas ni del cuestionario de factores asociados. En el caso que *Business Objects* no cuente aún con la información para el año evaluado, por generar los informes antes que termine el año y porque la Dirección General de Planificación –Diplan– no ha recopilado los registros de datos finales, no se incluyen los indicadores de eficiencia interna en los reportes.

Si las variables solicitadas para generar reportes no se tuvieran o han sufrido modificaciones, se notificará para que en la Unidad de Divulgación se hagan los ajustes al reporte.

---

[1] Todas las tablas y figuras son elaboración propia de la DIGEDUCA, por lo que no se indicará fuente.

**Tabla 1. Ejemplo de lista de variables que se incluirán en el informe**

N.º	Descripción de la variable	Nombre de la variable	Observaciones
1	Código del establecimiento	cod_esc	
2	Nombre del establecimiento	Nombre_del_establecimiento	
3	Dirección del establecimiento	Dirección_del_establecimiento	
4	Departamento	cod_depa	
5	Municipio	cod_muni	
6	Sector	SECTOR	
7	Área	AREA	
8	Plan	PLAN	
9	Jornada	JORNADA	
10	Estudiantes inscritos en el establecimiento en el año 2004	inscritos_2004	Variable generada en <i>Business Objects</i>
11	Estudiantes inscritos en el establecimiento en el año 2005	inscritos_2005	Variable generada en <i>Business Objects</i>
12	Estudiantes inscritos en el establecimiento en el año 2006	inscritos_2006	Variable generada en <i>Business Objects</i>
13	Estudiantes inscritos en el establecimiento en el año 2007	inscritos_2007	Variable generada en <i>Business Objects</i>
14	Estudiantes inscritos en el establecimiento en el año 2008	Inscritos_2008	Variable generada en <i>Business Objects</i>
15	Estudiantes inscritos en el establecimiento en el año 2009	Inscritos_2009	Variable generada en <i>Business Objects</i>
16	Estudiantes inscritos en el establecimiento en el año 2010	Inscritos_2010	Variable generada en <i>Business Objects</i>

Una vez que se tenga completa la lista de variables con las modificaciones propuestas, es necesaria una nueva reunión con los responsables de la base de datos y los reportes, para que se tengan establecidos los cambios necesarios en el formato que se utilizará. Es hasta entonces cuando se procede a preparar la base de datos que contienen

como casos a los establecimientos educativos con la información especificada como variables.

## PREPARACIÓN DE BASE DE DATOS

Los procedimientos para obtener la base de datos de establecimientos son utilizados para generar reportes de estudiantes graduandos como de tercero básico. Se debe recordar que son dos procesos diferentes por lo que al final se deben obtener una base a la vez, aunque en este documento se tomen ejemplos de ambas evaluaciones por ser similares y tener el mismo fin.

Como primer paso, se debe tomar la base de graduandos o la de tercero básico completa y obtener las tablas de frecuencias de las variables que serán utilizadas en los reportes. Esto servirá para corroborar qué variables son dicotómicas, cuáles son de respuesta múltiple, continuas o de texto, además de verificar los rangos de valores que deben contener. Luego se deben agregar los datos adicionales a la base de datos agregada.

### VARIABLES DICOTÓMICAS

Son variables que contienen solamente dos valores. Se usa el valor 1 para indicar que tiene la característica a la que se refiere y 0 para indicar que no la tiene. Al obtener las tablas de frecuencias de las variables dicotómicas, se debe verificar que solo tengan las categorías 1 y 0.

En el reporte, estas variables están identificadas como el porcentaje o proporción de alumnos del establecimiento que poseen esta característica. Si las variables no están codificadas de esta forma, se deben corregir. Esto es por la forma en que se calculan las proporciones, ya que es más práctico verlo como promedio entre cero y uno para obtener la proporción al guardar los nuevos valores en variables agregadas. Si se tienen variables en las cuales hay valores omitidos, se debe verificar que sean utilizados los porcentajes válidos.

**Tabla 2. Tabla de frecuencias de una variable dicotómica**

		<b>¿Asistió a la escuela preprimaria?</b>			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	NO (0)	24679	21.0	21.5	21.5
	SÍ (1)	89942	76.7	78.5	100.0
	Total	114621	97.8	100.0	
Perdidos	Sistema	2637	2.2		
Total		117258	100.0		

## VARIABLES DE RESPUESTA MÚLTIPLE

En estos casos, se deben crear nuevas variables para cada una de las categorías de respuesta. Pero antes, hay que obtener las tablas de frecuencia de dichas variables, para verificar cuáles y cuántas categorías de respuesta hay.

**Tabla 3. Tabla de frecuencias de una variable múltiple**

		Identificación étnica			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Maya	28730	24.5	25.1	25.1
	Ladino	84032	71.7	73.4	98.5
	Garífuna	619	.5	.5	99.0
	Xinka	402	.3	.4	99.4
	Extranjero	727	.6	.6	100.0
	Total	114510	97.7	100.0	
Perdidos	Sistema	2748	2.3		
Total		117258	100.0		

En el ejemplo de la variable “etnia” de la Tabla 3, se cuenta con cinco categorías de respuesta; sin embargo, en el reporte se agregó una opción de **“No reportó etnia”**. Para este caso, se deben crear las siguientes variables:

- Etnia\_maya
- Etnia\_ladino
- Etnia\_garífuna
- Etnia\_xinka
- Etnia\_extranjero
- Etnia\_no\_reportó

Lo más adecuado es crear una sintaxis en SPSS con las instrucciones para generar las variables dicotómicas, de manera que se pueda revisar el procedimiento y determinar que se hará lo que se necesita. Además, el uso de sintaxis permite borrar lo que se ha cambiado si hay errores, corregirla y volver a correrla para generar las variables cuantas veces sea necesario. Un ejemplo de una sintaxis para crear las variables de las opciones de respuesta en la variable “etnia”, se muestra en la Figura 1.

Figura 1. Sintaxis para desglosar opciones de respuesta en nuevas variables

```
RECODE IE_IDENTIFICACION_ETNICA (1=1) (ELSE=0) INTO ETNIA_MAYA.  
EXECUTE.  
RECODE IE_IDENTIFICACION_ETNICA (2=1) (ELSE=0) INTO ETNIA_LADINO.  
EXECUTE.  
RECODE IE_IDENTIFICACION_ETNICA (3=1) (ELSE=0) INTO ETNIA_GARIFUNA.  
EXECUTE.  
RECODE IE_IDENTIFICACION_ETNICA (4=1) (ELSE=0) INTO ETNIA_XINCA.  
EXECUTE.  
RECODE IE_IDENTIFICACION_ETNICA (5=1) (ELSE=0) INTO ETNIA_EXTRANJERO.  
EXECUTE.  
RECODE IE_IDENTIFICACION_ETNICA (SYSMIS=1) (ELSE=0) INTO ETNIA_NO_REPORTO.  
EXECUTE.
```

Teniendo las opciones de respuesta desagregadas en variables dicotómicas, se pueden agregar datos y generar la base por establecimiento. Para estas variables nuevas se obtiene el promedio, que al tener valores 0 y 1, representa la proporción de estudiantes en el establecimiento que eligieron cada opción de respuesta. De allí, para verificar que el proceso de generación de nuevas variables esté bien, se multiplican las proporciones por 100 para obtener los porcentajes, las cuales al sumarlas, deben dar como resultado **100**, ya que son mutuamente excluyentes y colectivamente exhaustivas. La sintaxis en SPSS para realizar este proceso con las variables agregadas se muestra en la Figura 2.

Se debe repetir el proceso para cada una de las variables que tengan respuesta múltiple y que sean incluidas en el reporte final. Esto se aplica también en las variables de desempeño en Matemáticas y Lectura para cada uno de los años que sean incluidos en el reporte.

Figura 2. Sintaxis para verificar que el desglose de variables esté bien

```
COMPUTE ETNIA_MAYA = ETNIA_MAYA_PROMEDIO*100.  
EXECUTE.  
COMPUTE ETNIA_LADINO = ETNIA_LADINO_PROMEDIO *100.  
EXECUTE.  
COMPUTE ETNIA_GARIFUNA = ETNIA_GARIFUNA_PROMEDIO *100.  
EXECUTE.  
COMPUTE ETNIA_XINCA = ETNIA_XINCA_PROMEDIO *100.  
EXECUTE.  
COMPUTE ETNIA_EXTRANJERO = ETNIA_EXTRANJERO_PROMEDIO *100.  
EXECUTE.  
COMPUTE ETNIA_NO_REPORTO = ETNIA_NO_REPORTO_PROMEDIO *100.  
EXECUTE.
```

## VARIABLES CONTINUAS

Las variables continuas contienen valores que están en el rango de los números reales y pueden tener decimales como la variable de habilidad de los estudiantes u otros indicadores calculados a partir de grupos de variables como el índice socioeconómico. Además, se tienen variables como la edad de los estudiantes que contiene valores enteros, pero para agregar los datos en cada escuela se usa el mismo procedimiento que una variable continua y se obtiene el promedio de ellas. En la Tabla 4 se muestran estadísticos de resumen de estas variables para verificar su contenido y el valor que se obtendrá en promedio al crear la base de datos por establecimiento.

**Tabla 4. Estadísticos de resumen de una variable que se necesita el promedio**

<b>Estadísticos</b>		
Edad del estudiante		
N	Válidos	141846
	Perdidos	84
Media		19.51
Mediana		18.00
Moda		18
Desv. típ.		4.126
Mínimo		14
Máximo		69
Suma		2766915

## VARIABLES DE TEXTO

Estas variables contienen texto de origen, evitando los pasos de recodificación y así disminuir el riesgo de errores en manipulación. Se utilizan como están pero se usan códigos para desglosar las opciones de respuesta en nuevas variables. Para calcular la cantidad de evaluados por establecimiento, por ejemplo en **rama de enseñanza**, (en el caso de Graduandos) se deben crear nuevas variables para cada rama reportada. Una sintaxis para obtener las variables nuevas de rama de enseñanza se presenta en la Figura 3.

La Figura 4 muestra cómo se verá el dato de ramas de enseñanza en el reporte de establecimiento, el cual es obtenido sumando los casos que pertenecen a cada rama en cada uno de estos.

Figura 3. Sintaxis para desglosar opciones de respuesta de rama de enseñanza

```

RECODE NOM_RAMA_ENSEÑANZA ('BACHILLERATO' =1) (ELSE= 0) INTO
EVALUADOS_BACHILLERATO.
EXECUTE.

RECODE NOM_RAMA_ENSEÑANZA ('MAGISTERIO' =1) (ELSE= 0) INTO
EVALUADOS_MAGISTERIO.
EXECUTE.

RECODE NOM_RAMA_ENSEÑANZA ('PERITO' =1) (ELSE= 0) INTO EVALUADOS_PERITO.
EXECUTE.

RECODE NOM_RAMA_ENSEÑANZA ('SECRETARIADO' =1) (ELSE= 0) INTO
EVALUADOS_SECRETARIADO.
EXECUTE.

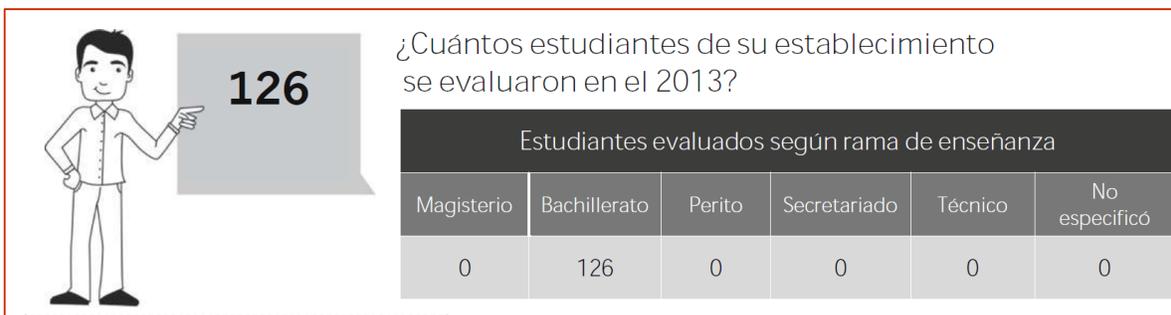
RECODE NOM_RAMA_ENSEÑANZA ('TECNICO' =1) (ELSE= 0) INTO EVALUADOS_TECNICO.
EXECUTE.

El total de evaluados por rama es la suma de estas cinco variables.

COMPUTE TOTAL_EVALUADOS =EVALUADOS_BACHILLERATO + EVALUADOS_MAGISTERIO +
EVALUADOS_PERITO + EVALUADOS_SECRETARIADO + EVALUADOS_TECNICO.
EXECUTE.

```

Figura 4. Datos de estudiantes en el establecimiento por rama educativa



En el caso del cálculo de evaluados por materia, se deben recodificar a los estudiantes que no respondieron alguno de los temarios de evaluación como cero en una nueva variable, para cuadrar la cantidad de evaluados en el desglose de temarios recibidos en cada área curricular.

Figura 5. Sintaxis para indicar quienes no indicaron forma en la prueba de Matemáticas

```
RECODE CODIGO_FORMA_MAT CODIGO_FORMA_LEC (SYSMIS=0) (ELSE=1) INTO  
evaluados_mateevaluados_lect.  
EXECUTE.
```

## LOGRO NACIONAL, DEPARTAMENTAL Y MUNICIPAL

A la base de datos general se agregan las variables que contengan el Logro nacional en Matemáticas y Lectura, así como el Logro por departamento y municipio de cada establecimiento. Para estas dos últimas variables, se puede crear un máster (base de datos que se fundirá para agregar información a otra) que contenga esta información y fundir la información por departamento o municipio, según sea el caso. Se debe verificar que la información de Logro nacional, departamental y municipal coincida con los del Anuario de resultados de evaluación de la Digeduca, que es en donde se tiene la información oficial publicada. En la Figura 6 se muestra la parte del reporte de resultados por establecimiento, donde se presenta el Logro obtenido por este, el municipal, departamental y nacional en cada materia, de manera que se tenga información para comparar. El Logro nacional se agrega como dato único en todos los casos para una variable adicional.

Figura 6. Resultados comparativos en el reporte por establecimiento



## CREACIÓN DE BASE POR ESTABLECIMIENTO

Una vez que se han generado todas las variables necesarias para el reporte, se crea una base de datos agregada en donde cada caso será un establecimiento. Esto se realiza utilizando la función de **Agregar datos en SPSS**, se coloca el código del establecimiento como la variable de segmentación y las variables de resumen deben ser todas las que serán incluidas en el reporte. El procedimiento se especifica en cada uno de los incisos a continuación.

- a. Las variables que se refieren a la información del establecimiento como: nombre del establecimiento, departamento, municipio, área, jornada, plan y dirección, utilizan la función de agregación **PRIMERO**,

lo cual significa que tomará el primer dato del establecimiento que encuentre y eso está bien al tener todos los casos iguales en ese establecimiento. El procedimiento para utilizar el menú de SPSS se muestra en la Figura 7.

- b. Las variables dicotómicas como el porcentaje de estudiantes que asistieron a preprimaria o repitieron algún grado, utilizan la función de agregación de **MEDIA**, las cuales al estar codificadas con 0 y 1, representarán la proporción para luego multiplicarlas por cien y obtener el porcentaje, como se observa en la Figura 8.
- c. El reporte por establecimiento también presenta información de algunas variables continuas como edad<sup>1</sup> y horas que trabajan. Estas también se calculan mediante la función de agregación de **MEDIA** y se usa el valor resultante sin transformarlo (ver Figura 9).

---

<sup>1</sup>Antes de agregar los datos, hay que verificar que se use la variable que ya fue limpiada y que no tiene valores ilógicos o fuera de rango.

Figura 7. Función agregar datos y elegir el primer dato en SPSS

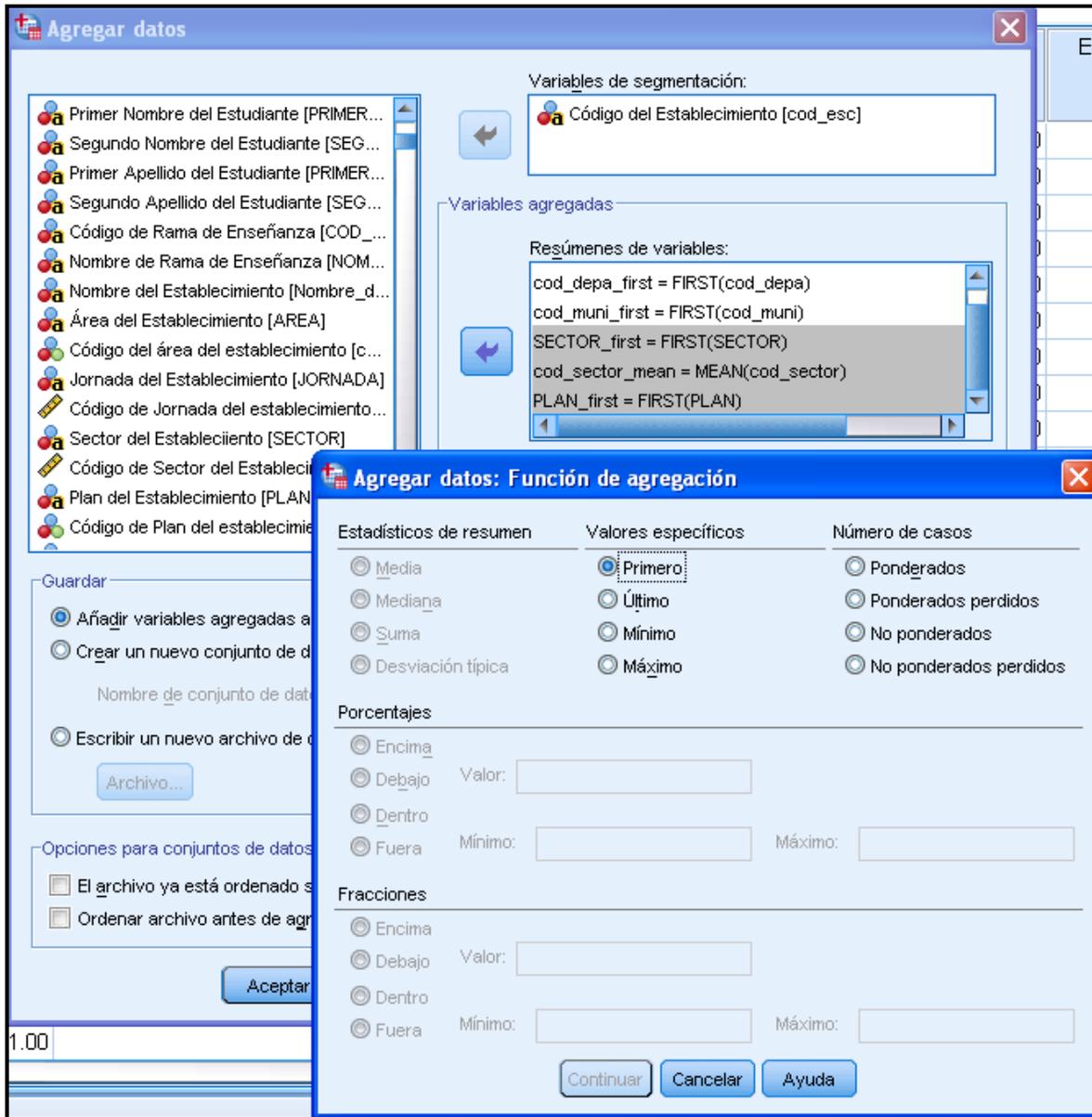
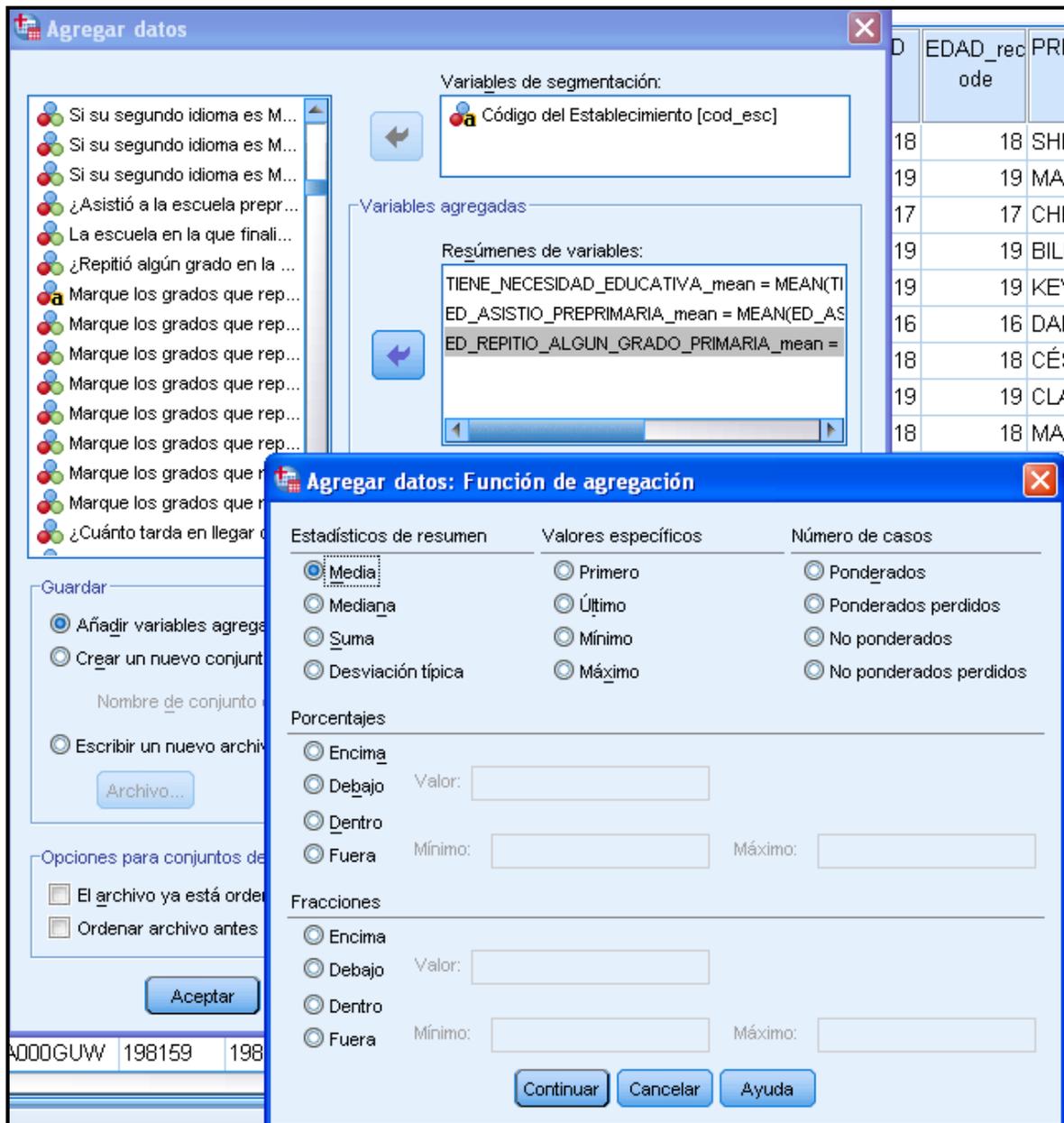
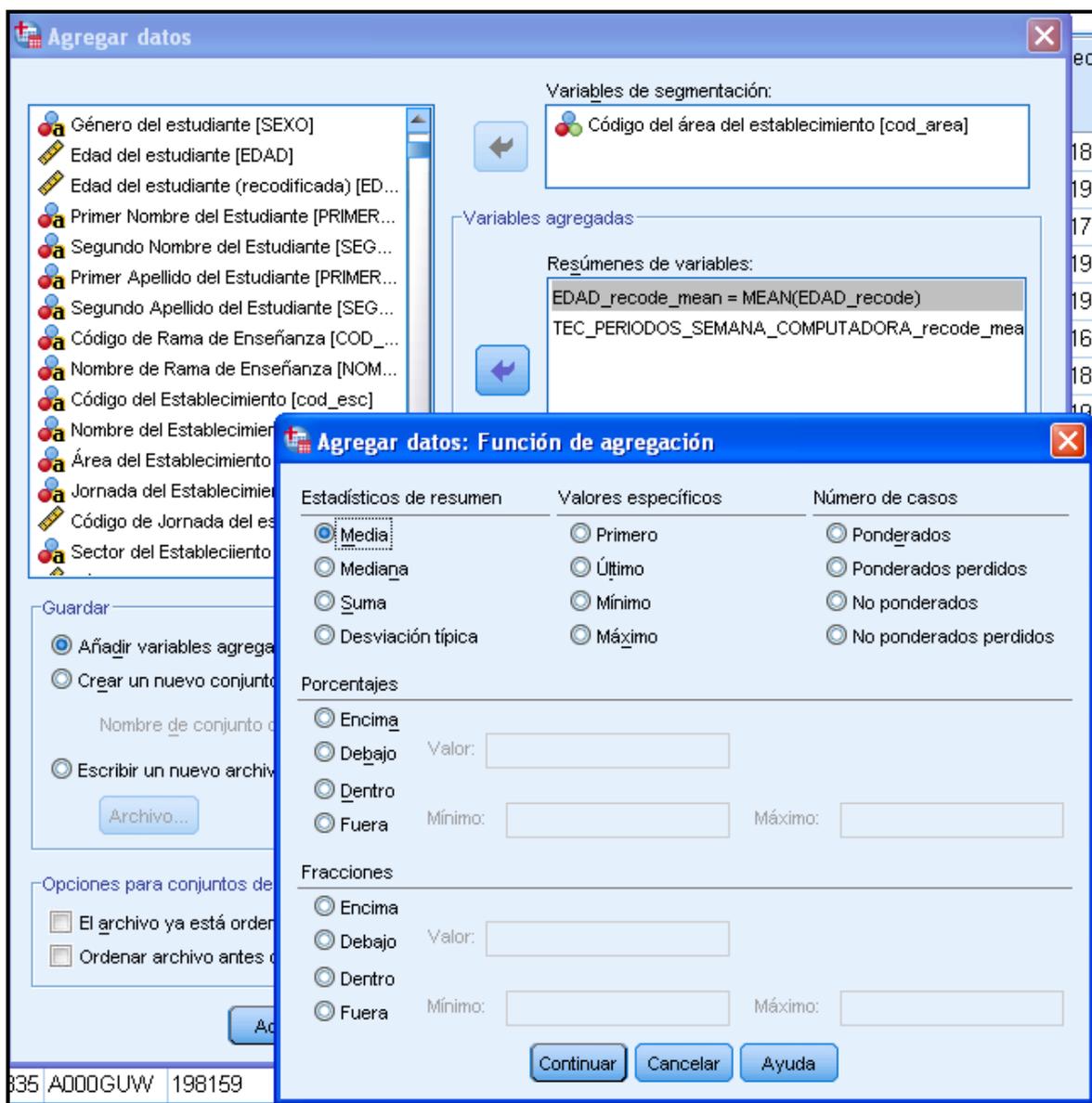


Figura 8. Función agregar datos y elegir el promedio en SPSS



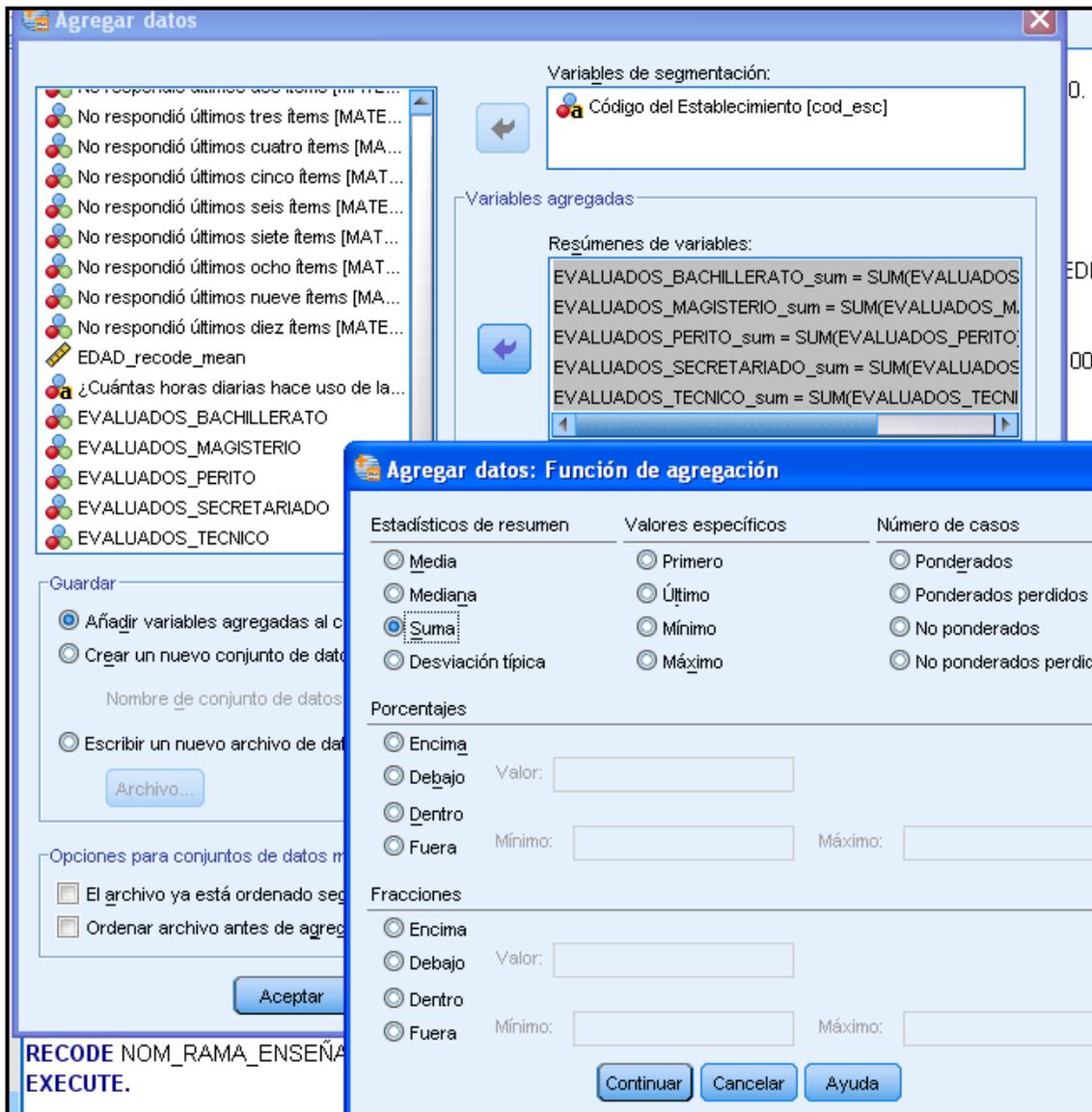
- d. Las variables de respuesta múltiple también deben ser agregadas a la nueva base de datos por establecimiento utilizando la función **MEDIA**, para cada una de las variables nuevas creadas (ver Figura 9).

Figura 9. Función agregar datos y elegir el promedio para edad y períodos a la semana en SPSS



- e. Las variables que indican la cantidad de evaluados por materia y rama de enseñanza, son incluidas mediante la función de agregación **SUMA**, sin otro procedimiento adicional, por tener valores enteros o mostrar solo el conteo de casos, como se observa en la Figura 10.

Figura 10. Función agregar datos y elegir suma para variables de conteo en SPSS



## MASTERS CON INFORMACIÓN ADICIONAL

Se debe crear una base de datos por establecimiento que contenga información adicional a la que aparece en el cuestionario de factores asociados y los resultados de las pruebas. Esta información ha incluido en versiones anteriores:

- Histórico de inscritos
- Histórico de tasas de aprobación
- Histórico de niveles de desempeño en Matemáticas
- Histórico de niveles de desempeño en Lectura
- Histórico de Logro en Matemáticas
- Histórico de Logro en Lectura

Este máster se fusiona con la base final por establecimiento para complementar la información y enriquecer el reporte para que los establecimientos tengan más elementos puestos en un mismo lugar, de manera que les pueda servir en la toma de decisiones o el mejoramiento de la calidad educativa que prestan.

Al tener la base de datos por establecimiento con la información completa, se guarda una copia en formato Excel, ya que este es un formato amigable para visualizar la información sin necesidad de tener el programa SPSS. Este documento en Excel debe ser editado para que los datos aparezcan de forma lógica y que solamente puedan ser direccionados desde un documento en Word que será el reporte utilizado como plantilla; la primera fila será los nombres que identifican la información por columna y el resto de filas contendrá los casos. Se deben verificar todas las variables con sus valores y que no contengan valores Nulos, o si es así, indicar que no aplica valores en esos casos. Especialmente se deben verificar las variables

desagregadas con una muestra de establecimientos para observar si la información es correcta, además de hacer coincidir los valores uno a uno con reglas de verificación fundiendo la base de datos en SPSS con la de Excel.

En el caso de Graduandos, este proceso se debe repetir con la base de datos para establecimientos del ciclo internacional, la cual se entrega debidamente identificada y separada de la base del ciclo regular. Si se solicita, se integrará la base de datos del ciclo internacional con los establecimientos del ciclo regular.

Asimismo, se deberá repetir el procedimiento con las bases de datos de la evaluación extemporánea y otras bases de estudiantes graduandos o tercero básico que se evaluaron después del tiempo de la evaluación regular.

Al haber completado la verificación de valores y formato en la base de datos final en Excel, esta será entregada a la Unidad de Divulgación para que allí se generen los reportes de resultados por establecimiento.

Es preferible que los datos de Logro o información de años anteriores, se generen de las bases de datos oficiales corroborando y verificando que coincidan con las bases de datos de establecimientos del año anterior o de años anteriores.

Se debe tener cuidado de ubicar a los establecimientos nuevos de cada año ya que no tendrán historial de resultados, por lo que en estos casos no es información faltante, sino que se no aplica para el historial (ver Figura 11).

Figura 11. Establecimientos que no se evaluaron en años anteriores

	cod_estab	LOGRO_Lectura_2010	LOGRO_Lectura_2011	LOGRO_Lectura_establecimiento_2012
1	00-01-0158-46	55.56	27.78	40.00
2	00-01-0173-46	55.13	47.52	55.07
3	00-01-0174-46	42.11	48.82	50.00
4	00-01-0175-46	66.67	50.00	54.17
5	00-01-0176-46	61.59	69.44	58.62
6	00-01-0178-46	.	.	34.00
7	00-01-0179-46	21.43	58.33	25.00
8	00-01-0180-46	21.74	6.67	30.00
9	00-01-0182-46	19.35	30.30	18.97
10	00-01-0183-46	59.55	56.36	64.77

Igualmente, hay establecimientos que se someten a la evaluación en algunos años y en otros no (esto es porque algunos años tienen estudiantes graduandos o de tercero básico y en otros años no). En los años no evaluados se coloca N/A que significa que no aplica para esta información. Por esta razón se debe verificar o generar el porcentaje de Logro o los porcentajes de los niveles de desempeño de las bases oficiales para identificar si no aplica para determinar los valores en esas variables.

## DISPOSICIONES FINALES

Cuando se definan las variables, se debe decidir si se requiere reportar el porcentaje válido de Logro y desempeño. El porcentaje válido representa el resultado de los estudiantes que pertenecen a la base de datos y tienen calificación.

Si solo se requiere reportar el porcentaje válido (sin incluir a quienes no se les pudo generar calificación por no haber reportado la forma de la prueba), entonces se utilizará la sintaxis que se presenta en la Figura 12.

**Figura 12. Sintaxis para obtener resultados de desempeño válido**

```
RECODE DESEMPEÑO_MATE (1=1) (SYSMIS=SYSMIS) (ELSE=0) INTO MATE_DESEMPEÑO_INSATISFACTORIO.  
EXECUTE.  
  
RECODE DESEMPEÑO_MATE (2=1) (SYSMIS=SYSMIS) (ELSE=0) INTO MATE_DESEMPEÑO_DEBE_MEJORAR.  
EXECUTE.  
  
RECODE DESEMPEÑO_MATE (3=1) (SYSMIS=SYSMIS) (ELSE=0) INTO MATE_DESEMPEÑO_SATISFACTORIO.  
EXECUTE.  
  
RECODE DESEMPEÑO_MATE (4=1) (SYSMIS=SYSMIS) (ELSE=0) INTO MATE_DESEMPEÑO_EXCELENTE.  
EXECUTE.
```

Es necesario que los datos perdidos (SYSMIS=SYSMIS) corresponden a estudiantes que no colocaron forma de la prueba o estudiantes que por alguna otra razón dejaron respuestas en blanco.

Al obtener resultados válidos de desempeño, se continúa con la verificación en donde se sumará el resultado de las cuatro columnas, el cual deberá quedar el cien por ciento, como se puede observar en la siguiente figura.

Figura 13. Verificación de porcentaje por desempeño válido en SPSS

	DESEMPEÑO_LEC_EXCELENTE_mean_n_2013	DESEMPEÑO_LEC_SATISFAC_mean_2013	DESEMPEÑO_LEC_D_MEJORAR_me_an_2013	DESEMPEÑO_LEC_INSATIS_mean_2013	suma
1	13.64	27.27	22.73	36.36	100.00
2	35.43	25.20	33.07	6.30	100.00
3	31.87	25.27	21.43	21.43	100.00
4	60.00	15.00	17.50	7.50	100.00
5	39.62	30.19	24.53	5.66	100.00
6	21.21	18.18	33.33	27.27	100.00
7	7.69	38.46	7.69	46.15	100.00
8	7.14	14.29	35.71	42.86	100.00
9	12.50	15.63	42.19	29.69	100.00
10	36.36	30.30	21.21	12.12	100.00
11	30.00	20.00	45.00	5.00	100.00
12	37.21	11.63	32.56	18.60	100.00
13	8.77	10.53	50.88	29.82	100.00
14	20.00	30.00	5.00	45.00	100.00
15	22.81	22.81	36.84	17.54	100.00
16	13.97	16.18	33.82	36.03	100.00
17	19.05	28.57	30.95	21.43	100.00
18	8.16	20.41	20.41	51.02	100.00

El porcentaje válido de Logro o desempeño deja afuera los valores perdidos que en este caso serían los estudiantes a los que no se les generó resultado.

Figura 14. Sintaxis para obtener variables de desempeño tomando en cuenta no calificados

```

RECODE DESEMPEÑO_MATE (1=1) (ELSE=0) INTO MATE_DESEMPEÑO_INSATISFACTORIO.
EXECUTE.

RECODE DESEMPEÑO_MATE (2=1) (ELSE=0) INTO MATE_DESEMPEÑO_DEBE_MEJORAR.
EXECUTE.

RECODE DESEMPEÑO_MATE (3=1) (ELSE=0) INTO MATE_DESEMPEÑO_SATISFACTORIO.
EXECUTE.

RECODE DESEMPEÑO_MATE (4=1) (ELSE=0) INTO MATE_DESEMPEÑO_EXCELENTE.
EXECUTE.

RECODE DESEMPEÑO_MATE (SYSMIS=1) (ELSE=0) INTO MATE_DESEMPEÑO_NO_INDICA_O_FORMA_NEE.
EXECUTE.
    
```

En algunos años se ha reportado el porcentaje incluyendo valores perdidos de desempeño. Esto quiere decir que dentro de los resultados se generan cinco columnas de resultados: desempeño excelente, satisfactorio, debe mejorar, insatisfactorio y no indicó forma de la prueba. Para este caso la sintaxis a ser usada será como la mostrada en la Figura 14.

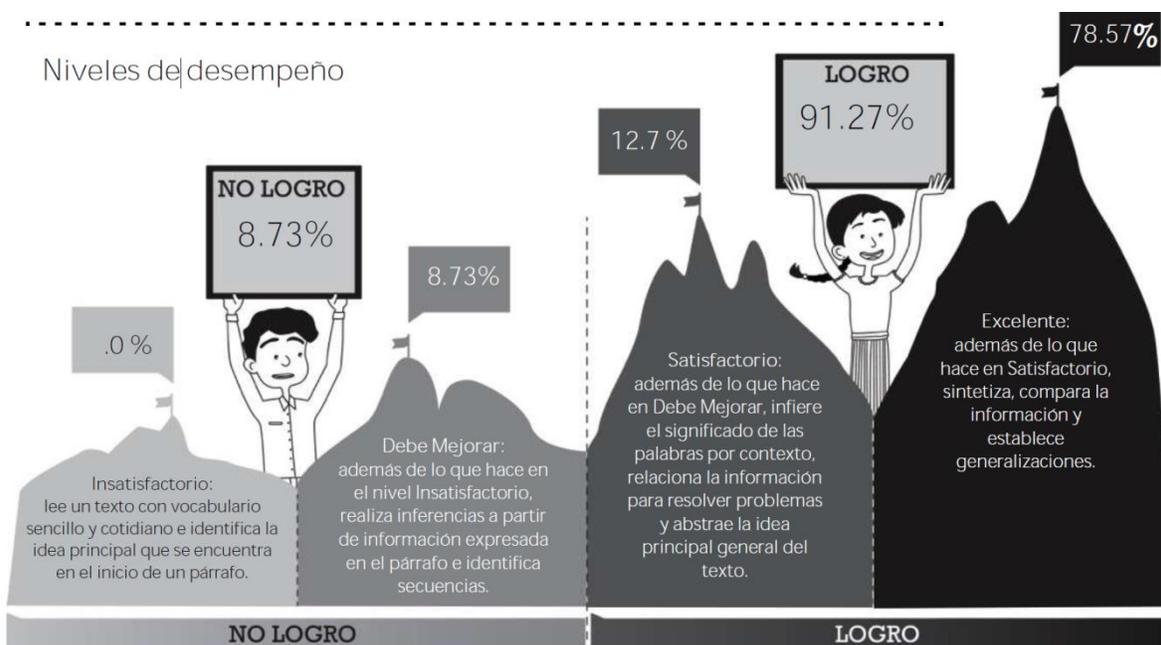
En el caso de haber utilizado la sintaxis como la de la figura anterior, en donde se toman en cuenta a los estudiantes que no colocaron forma de la prueba, la verificación del cien por ciento se hará con las cinco variables generadas en vez de solo los cuatro niveles de desempeño, como se observa en la Figura 15.

**Figura 15. Verificación de porcentaje por desempeño con quienes no se pudo generar calificación en SPSS**

	DESEMPEÑO O_LEC_EXC ELENTE...	DESEMPEÑO O_LEC_SATI SFAC_m...	DESEMPEÑO O_LEC_D_M EJORAR...	DESEMPEÑO O_LEC_INSA TIS_mean...	DESEMPEÑO O_LEC_NO_I NDICA_m...	suma
1	13.64	27.27	22.73	36.36	.00	100.00
2	35.16	25.00	32.81	6.25	.78	100.00
3	31.87	25.27	21.43	21.43	.00	100.00
4	60.00	15.00	17.50	7.50	.00	100.00
5	39.62	30.19	24.53	5.66	.00	100.00
6	21.21	18.18	33.33	27.27	.00	100.00
7	7.69	38.46	7.69	46.15	.00	100.00
8	7.14	14.29	35.71	42.86	.00	100.00
9	12.50	15.63	42.19	29.69	.00	100.00
10	36.36	30.30	21.21	12.12	.00	100.00
11	30.00	20.00	45.00	5.00	.00	100.00
12	37.21	11.63	32.56	18.60	.00	100.00
13	8.77	10.53	50.88	29.82	.00	100.00
14	20.00	30.00	5.00	45.00	.00	100.00
15	22.81	22.81	36.84	17.54	.00	100.00
16	13.97	16.18	33.82	36.03	.00	100.00
17	19.05	28.57	30.95	21.43	.00	100.00
18	8.16	20.41	20.41	51.02	.00	100.00
19	4.35	26.09	21.74	43.48	4.35	100.00
20	52.27	31.82	13.64	2.27	.00	100.00
21	55.00	35.00	10.00	.00	.00	100.00
22	78.57	12.70	8.73	.00	.00	100.00
23	16.67	.00	16.67	66.67	.00	100.00

Es importante considerar esto para generar los resultados usando el mismo criterio para Logro y No logro y para los niveles de desempeño, es decir, que en ambos se incluyan a los que no indicaron forma o en ambos los excluyan. Dependerá si en el informe se crea la casilla correspondiente a estudiantes que no reportaron forma o una nota al pie del cuadro. En la Figura 16 se puede observar la forma como se presentan los resultados de Logro y desempeño en el reporte.

**Figura 16. Presentación de porcentajes de desempeño y Logro en el reporte del establecimiento**



Otra revisión importante en la base de datos a entregar es verificar que todos los establecimientos tengan resultados de Logro. Si hay datos perdidos en esta variable se deberá verificar por qué se da esta situación.

Una razón por la que un establecimiento no tenga resultados de Logro es porque puede haber establecimientos en donde todos sus estudiantes fueron evaluados con alguna forma de necesidades educativas especiales.

Luego de haber corroborado que es así, se debe identificar el establecimiento en la base preparada, de manera que se verifique el llenado de estos resultados ya que estos tomaron pruebas especiales.

## VARIABLES DE APROXIMACIÓN

En el caso que se soliciten valores recodificados o reducción de escala, se deberá tomar en cuenta la consideración de que los datos perderán exactitud al categorizarlos. Una de las variables nuevas que se ha solicitado en ocasiones es el número de estudiantes por cada 10 evaluados que sí alcanzaron el Logro, lo cual ya no es porcentaje y por ende, se debe categorizar la escala.

Por ejemplo, si tiene 50 % de Logro entonces la variable de Logro será recodificada de forma que la nueva variable tendrá el número 5. Porque de cada 10 evaluados 5 alcanzaron el Logro. En la Figura 17 se muestra la sintaxis en SPSS que reduce la escala de porcentaje a una escala de 10 valores.

En la Figura 18 se muestra la nueva variable con escala reducida de cantidad que alcanzaron el Logro por cada 10. Esto presenta pérdida de exactitud por aproximación, por lo que al sumar la cantidad obtenida de los cuatro niveles de desempeño puede ser que el resultado no llegue a 10. Esto debido a que las fracciones de cada desempeño sean iguales y todos aproximen a la decena anterior, por lo que la suma de las aproximaciones sea menor al valor total. Esta observación se debe hacer para que se tome en cuenta.

Figura 17. Recodificación de porcentaje de Logro, reduciendo la escala a 10

```
RECODE LOGRO_LEC_2013_PORCENTAJE (0 thru 4.9999999999999999=0) (5 thru 14.999999999999999=1)
(15 thru 24.999999999999999=2) (25 thru 34.999999999999999=3) (35 thru 44.999999999999999=4) (45 thru 54.999999999999999=5)
(55 thru 64.999999999999999=6) (65 thru 74.999999999999999=7) (75 thru 84.999999999999999=8) (85 thru 94.999999999999999=9)
(95 thru 100=10) INTO NUM_ALUM_10_EVAL_LEC.
```

EXECUTE.

```
RECODE LOGRO_MATE_2013_PORCENTAJE (0 thru 4.9999999999999999=0) (5 thru 14.999999999999999=1)
(15 thru 24.999999999999999=2) (25 thru 34.999999999999999=3) (35 thru 44.999999999999999=4) (45 thru 54.999999999999999=5)
(55 thru 64.999999999999999=6) (65 thru 74.999999999999999=7) (75 thru 84.999999999999999=8) (85 thru 94.999999999999999=9)
(95 thru 100=10) INTO NUM_ALUM_10_EVAL_MATE.
```

EXECUTE.

Figura 18. Variable nueva reduciendo la escala de porcentaje a 10 categorías en la base de datos

	LOGRO_LECTURA_ESTABLECIMIENTO_mean_2013	NUM_ALUM_10_EVAL_LEC
1	41	4
2	61	6
3	57	6
4	75	8
5	70	7
6	39	4
7	46	5
8	21	2
9	28	3
10	67	7
11	50	5
12	49	5
13	19	2
14	50	5

# Serie de Cuadernillos Técnicos



## Generación de bases de datos para reportes por establecimiento