



**DIRECCIÓN DE PROYECTOS
INTERNACIONALES Y ESPECIALES**

**Guía para el análisis de datos del
Segundo Estudio Regional Comparativo y Explicativo
(SERCE)**

18 DE MAYO 2010

ÍNDICE

INTRODUCCIÓN.....	5
MEDIAS DE DESEMPEÑO Y SUS INTERVALOS DE CONFIANZA.....	7
PORCENTAJE DE LOS NIVELES DE DESEMPEÑO Y SUS ERRORES ESTÁNDAR.....	11
DIFERENCIAS DE MEDIAS DE DESEMPEÑO SIGNIFICATIVAS.....	15
DIFERENCIAS DE MEDIAS DE DESEMPEÑO SIGNIFICATIVAS ENTRE PAÍSES	19
CREACIÓN DEL ARCHIVO <i>PLANSERCE.CSPLAN</i> EN SPSS	21
NOTAS SOBRE LOS PESOS MUESTRALES	25

INTRODUCCIÓN

El *Segundo Estudio Regional Comparativo y Explicativo* (SERCE), organizado y coordinado por el *Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación* (LLECE), se enmarca en las acciones de la UNESCO tendientes a asegurar el derecho de todos los estudiantes latinoamericanos y caribeños a recibir una educación de calidad.

El Instituto Nacional para la Evaluación de la Educación (INEE), fue la institución encargada de la coordinación de la aplicación de SERCE en México. En 2008 el INEE dio a conocer los resultados de la evaluación en el reporte *Resultados Nacionales. Segundo Estudio Regional Comparativo y Explicativo 2006*, difundiendo la comparación de los aprendizajes encontrados entre los países participantes, así como los hallazgos encontrados en el estudio.

Continuando la difusión de la información del estudio, el INEE pone a la disposición de los investigadores en educación y del público en general una guía sobre los procedimientos utilizados en la obtención de los resultados nacionales que se presentaron en el reporte antes mencionado. También se proporcionan algunas notas relativas al muestreo estadístico del estudio.

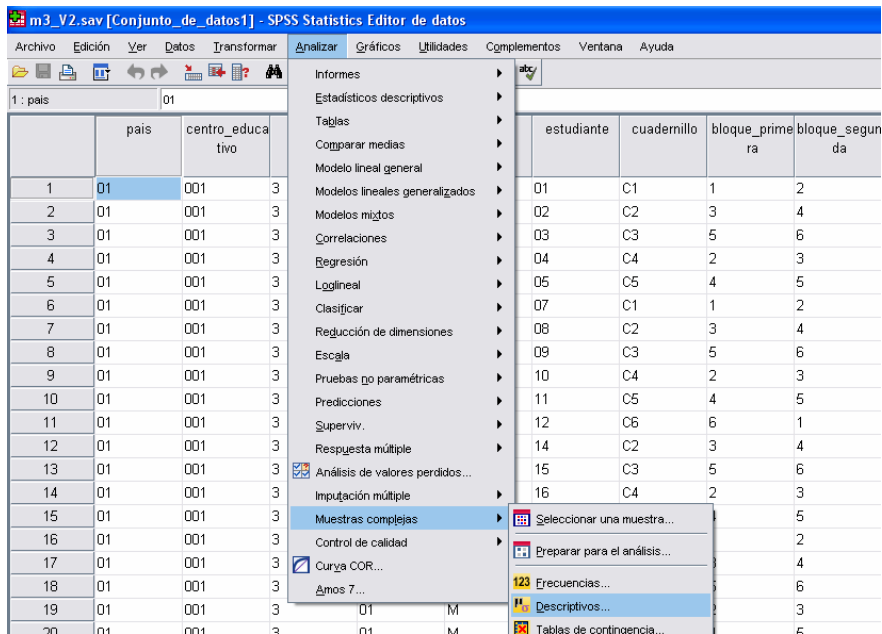
Es importante mencionar que para poder llevar a cabo cada uno de los pasos, es necesario obtener las bases de datos del SERCE así como el archivo *PlanSERCE.csplan* el cual fue preparado por el INEE y que contiene la muestra compleja utilizada en el estudio. Estos archivos están disponibles en la página del Instituto: www.inee.edu.mx Además, es necesario utilizar el software *SPSS* en su versión 15 ó superior, el cual debe contar con el módulo de muestras complejas.

Se espera que al conocer y realizar estos procedimientos, el usuario cuente con las herramientas necesarias que le permitan aprovechar de mejor forma las bases de datos del estudio en el desarrollo de sus propias investigaciones. El objetivo finalmente es incentivar trabajos de investigación que favorezcan la comprensión de los factores que inciden en la calidad de la educación de los estudiantes de Latinoamérica y del Caribe.

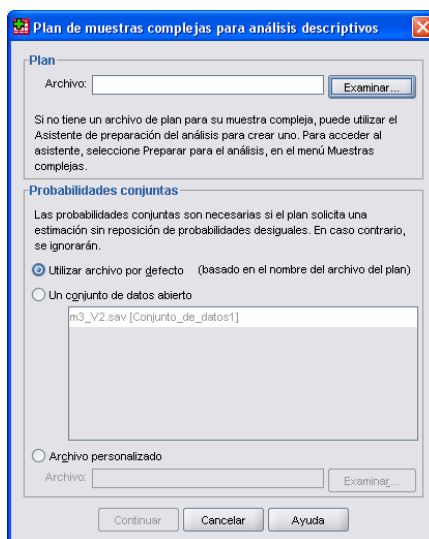
MEDIAS DE DESEMPEÑO Y SUS INTERVALOS DE CONFIANZA

Para la obtención de las medias de desempeño con sus respectivos intervalos de confianza se realiza el siguiente procedimiento:

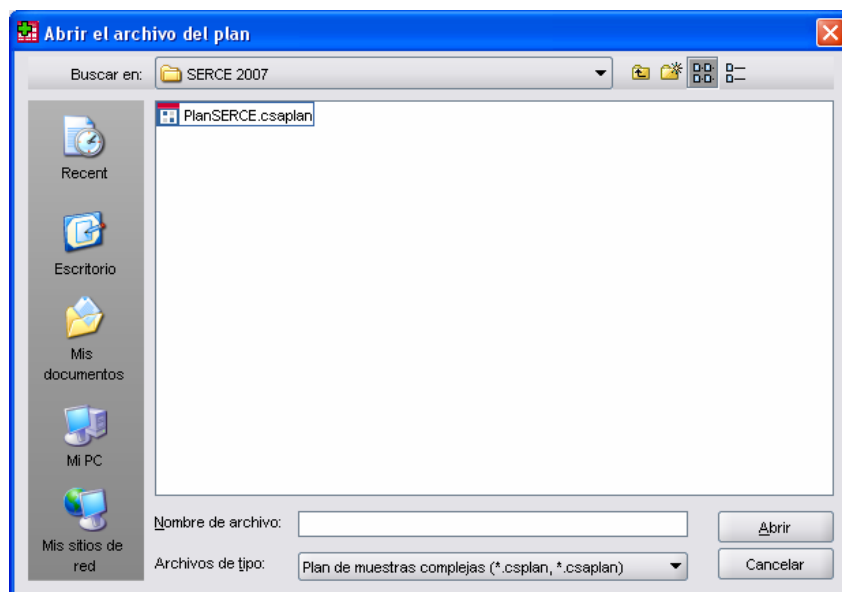
1. Abrir en SPSS la base a partir de la cual se obtendrán los promedios de los puntajes (L3.sav, M3.sav, L6.sav, M6.sav).
2. Seleccionar el menú **Analizar > Muestras Complejas > Descriptivos**.



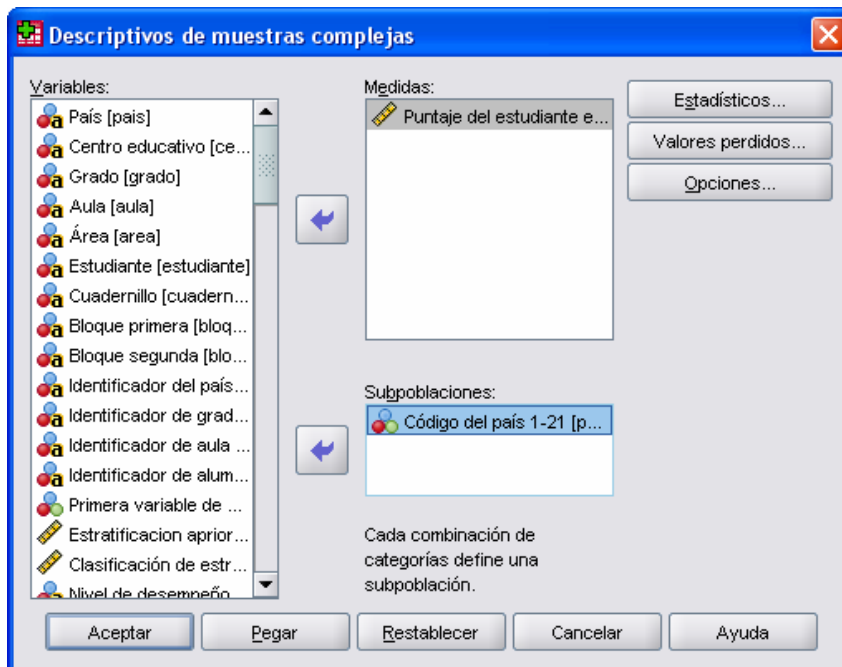
3. En el cuadro de diálogo **Plan de muestras complejas para análisis descriptivos** dar clic en *Examinar*.



4. Localizar el archivo **PlanSERCE.csaplan** (el cual está disponible junto con las bases) seleccionarlo y dar clic en **Abrir**.

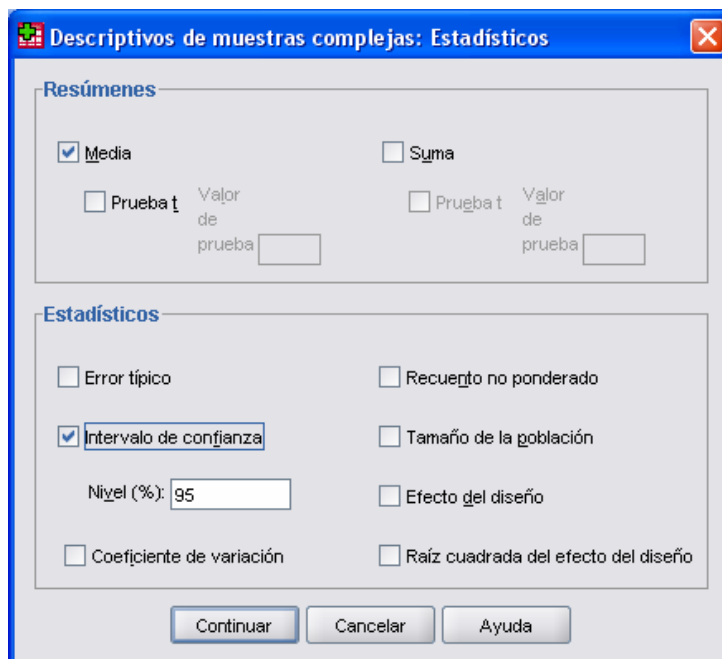


5. Posteriormente dar clic en **Continuar** y en el cuadro de diálogo **Descriptivos de muestras complejas** agregar la variable **Puntaje del estudiante estandarizado** a la lista de **Medidas**. En la lista de **Subpoblaciones** agregar la variable **Código del país**.



6. Enseguida dar clic en **Estadísticos**.

7. En el cuadro de diálogo **Descriptivos de muestras complejas: Estadísticos**, seleccionar la **Media** en el apartado de **Resúmenes** y en el de **Estadísticos** elegir **Intervalo de confianza** e indicar un **Nivel (%)** de **95**.



8. Dar clic en *Continuar* y posteriormente en *Aceptar*.

9. El procedimiento concluye al obtener la tabla de resultados que contiene las medias y sus intervalos de confianza.

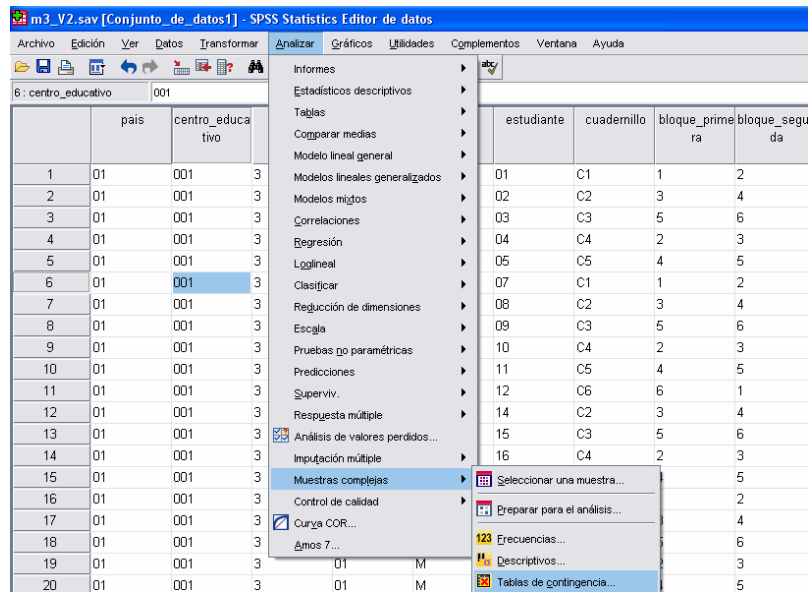
Estadísticos univariantes

Código del país 1-21			Estimación	Intervalo de confianza al 95%	
				Inferior	Superior
Argentina	Media	Puntaje del estudiante estandarizado	505.3636	497.1592	513.5680
Brasil	Media	Puntaje del estudiante estandarizado	505.0331	493.9788	516.0873
Colombia	Media	Puntaje del estudiante estandarizado	499.3539	489.9761	508.7317
CostaRica	Media	Puntaje del estudiante estandarizado	538.3225	532.3411	544.3040
Cuba	Media	Puntaje del estudiante estandarizado	647.9332	629.7540	666.1123

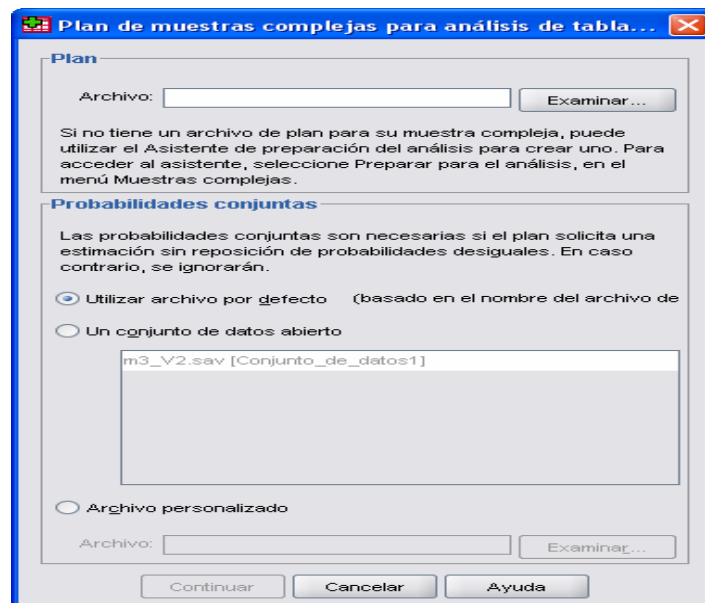
PORCENTAJE DE LOS NIVELES DE DESEMPEÑO Y SUS ERRORES ESTÁNDAR

Para la obtención de los porcentajes de los niveles de desempeño y sus errores estándar, se realiza el siguiente procedimiento:

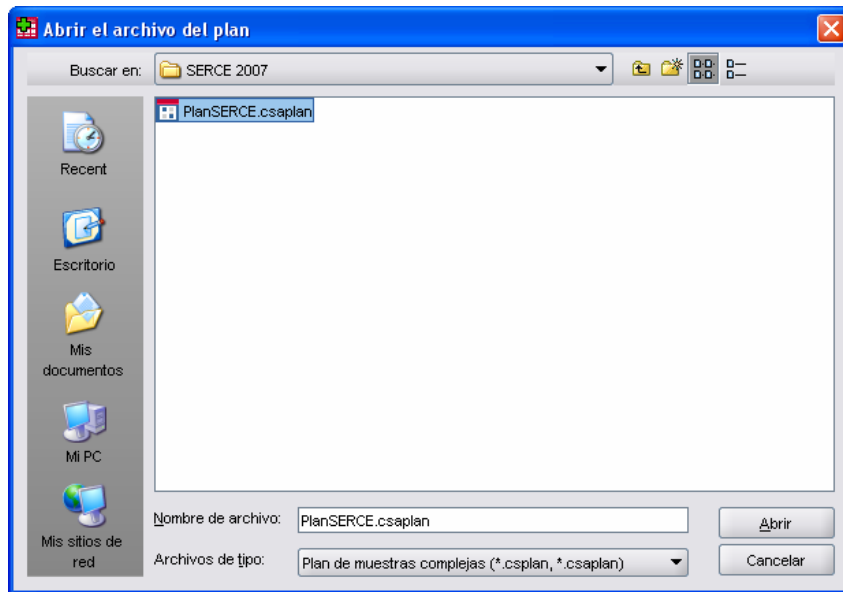
1. Abrir en SPSS la base a partir de la cual se obtendrán los porcentajes de los niveles de desempeño (L3.sav, M3.sav, L6.sav, M6.sav).
2. Seleccionar el menú **Analizar > Muestras Complejas > Tablas de Contingencia**.



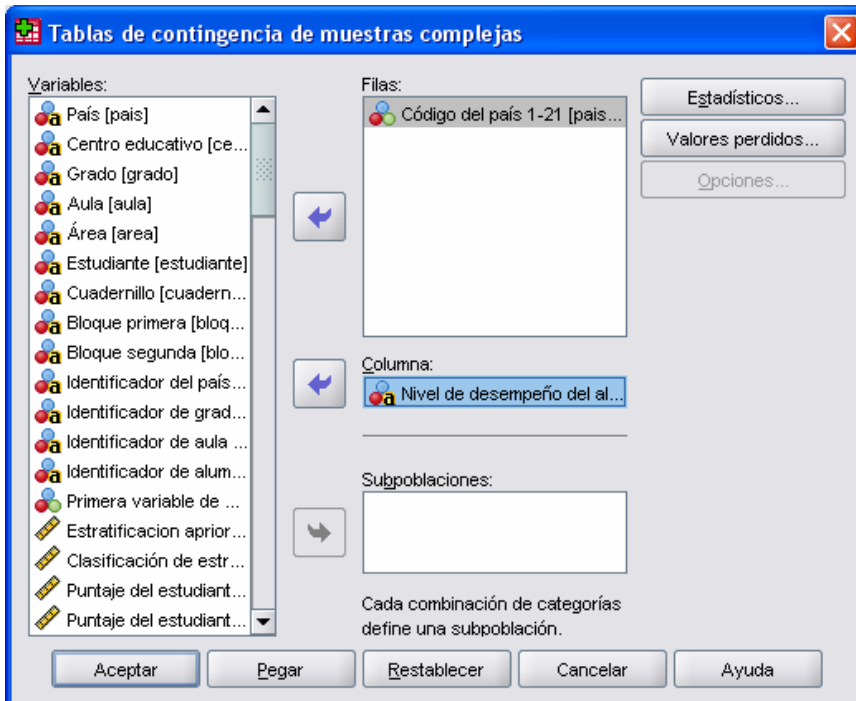
3. En el cuadro de diálogo **Plan de muestras complejas para análisis de tablas de contingencia** dar clic en *Examinar*.



4. localizar el archivo **PlanSERCE.csaplan** (el cual está disponible junto con las bases) seleccionarlo y dar clic en *Abrir*.

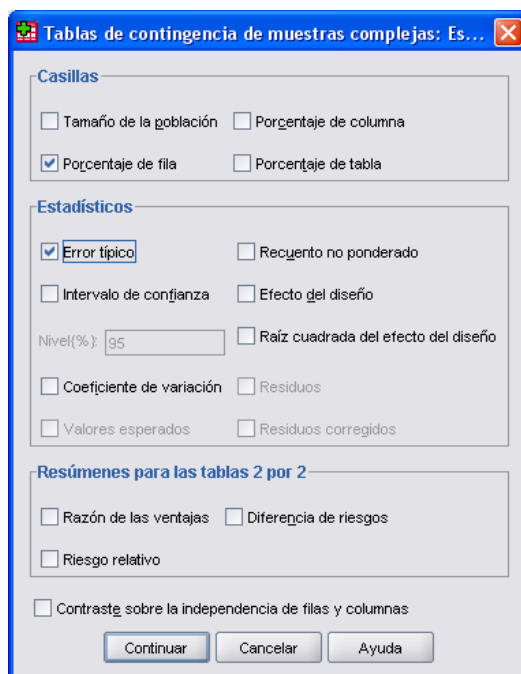


5. Posteriormente dar clic en *Continuar* y en el cuadro de diálogo **Tablas de contingencia de muestras complejas** agregar en la lista **Filas** la variable **Código del país** y en la lista **Columnas** la variable **Nivel de desempeño del alumno**. En este ejemplo no se utilizan variables en la lista **Subpoblaciones**.



6. Enseguida dar clic en *Estadísticos*.

7. En el cuadro de diálogo **Tablas de contingencia de muestras complejas: Estadísticos**, seleccionar **Porcentaje de fila** del apartado **Casillas** y **Error típico** de **Estadísticos**. En este ejemplo no se utiliza las opciones del apartado **Resúmenes para las tablas 2 por 2**.



8. Dar clic en *Continuar* y posteriormente en *Aceptar*.

9. El procedimiento concluye al obtener la tabla de resultados con los porcentajes en los niveles de desempeño y sus errores estándar.

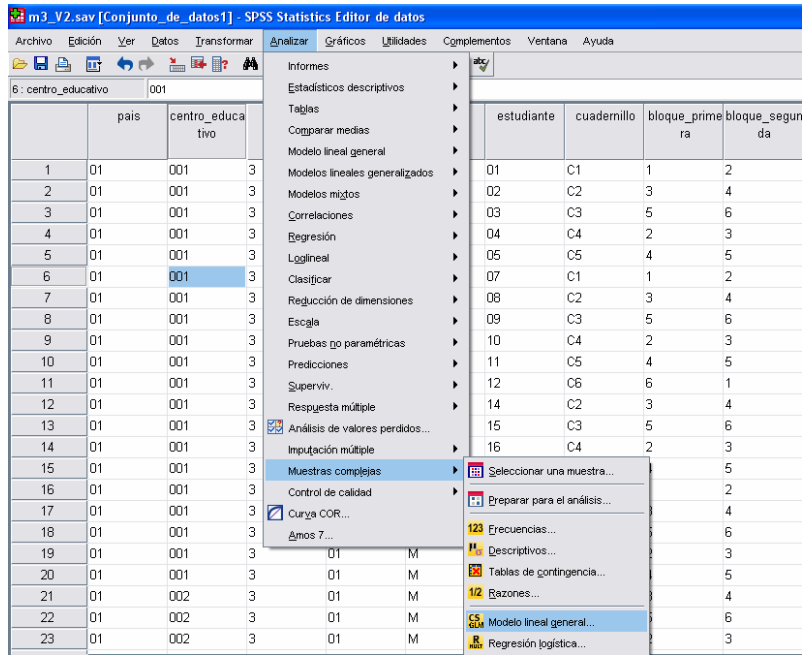
Código del país 1-21 * Nivel de desempeño del alumno

Código del país 1-21			Nivel de desempeño del alumno					Total
			Debajo de I	I	II	III	IV	
Argentina	% de Código del país 1-21	Estimación	10.5%	32.8%	31.1%	15.2%	10.5%	100.0%
		Error típico	.8%	1.2%	.9%	.9%	.9%	.0%
Brasil	% de Código del país 1-21	Estimación	10.3%	36.6%	26.7%	14.3%	12.1%	100.0%
		Error típico	1.0%	1.8%	1.0%	1.0%	1.3%	.0%
Colombia	% de Código del país 1-21	Estimación	8.6%	38.6%	33.2%	13.0%	6.7%	100.0%
		Error típico	.9%	1.7%	1.1%	1.0%	1.0%	.0%
Costa Rica	% de Código del país 1-21	Estimación	2.6%	24.4%	37.0%	22.3%	13.6%	100.0%
		Error típico	.4%	1.3%	.9%	.9%	.9%	.0%
Cuba	% de Código del país 1-21	Estimación	1.1%	10.2%	17.0%	17.4%	54.4%	100.0%
		Error típico	.3%	1.3%	1.5%	1.0%	3.3%	.0%

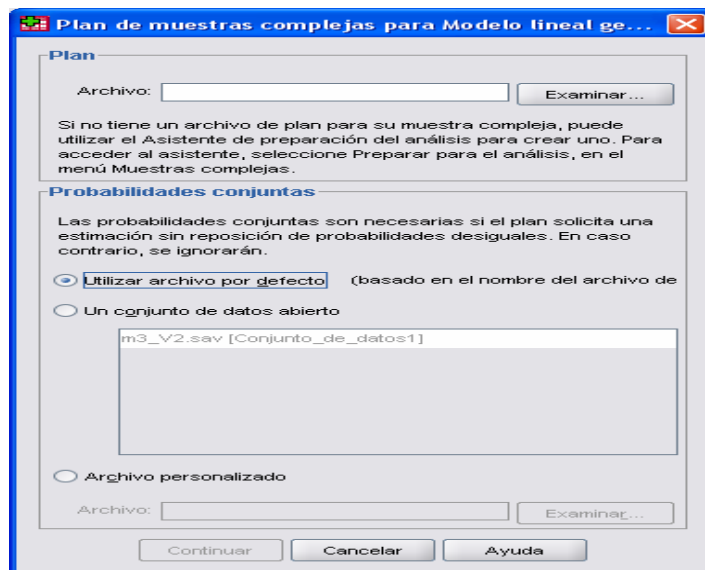
DIFERENCIAS DE MEDIAS DE DESEMPEÑO SIGNIFICATIVAS

Para obtener la diferencia de medias de desempeño y comprobar si es significativa, se realiza el siguiente procedimiento:

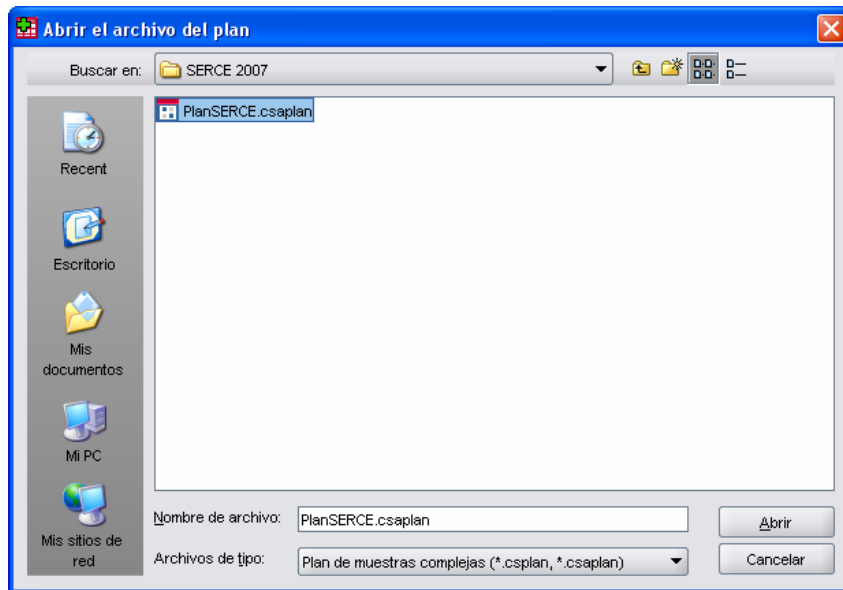
1. Abrir en SPSS la base a partir de la cual se obtendrán las diferencias de las medias de desempeño (L3.sav, M3.sav , L6.sav , M6.sav).
2. Seleccionar el menú **Analizar > Muestras Complejas > Modelo lineal general**.



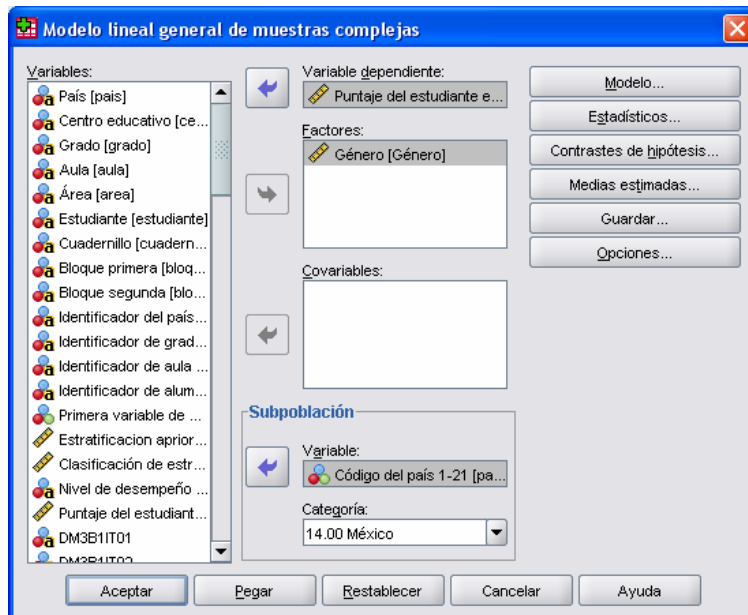
3. En el cuadro de diálogo **Plan de muestras complejas para Modelo lineal general** dar clic en *Examinar*.



4. Localizar el archivo **PlanSERCE.csaplan** (el cual está disponible junto con las bases) seleccionarlo y dar clic en **Abrir**.

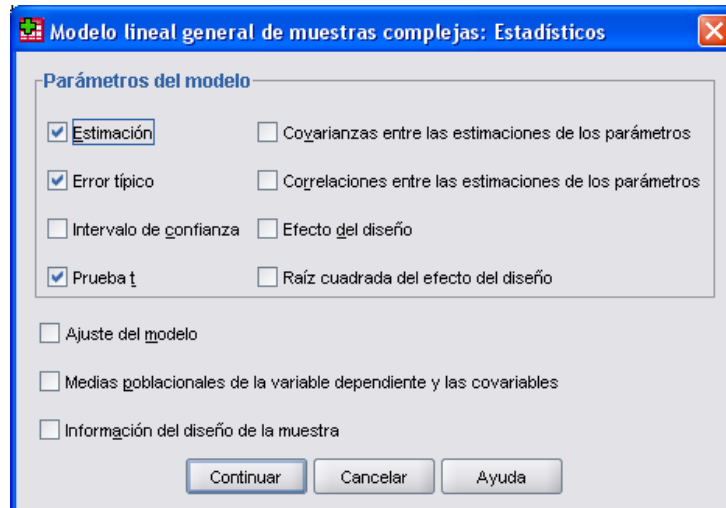


5. Posteriormente dar clic en el botón **Continuar** y en el cuadro de diálogo **Modelo lineal general de muestras complejas**, seleccionar como **Variable dependiente** el **Puntaje del estudiante estandarizado**, en **Factores** seleccionar la variable para la que se requiera obtener las diferencias de medias (por ejemplo **Género**), como **Subpoblación** seleccionar la variable por la que se segmentará la base (**Código del país**), en **Categoría** se definirá cuál es la población para la que se requiere obtener el estadístico (por ejemplo **14.0 México**). En este ejemplo no se agregan variables en la lista **Covariables**.

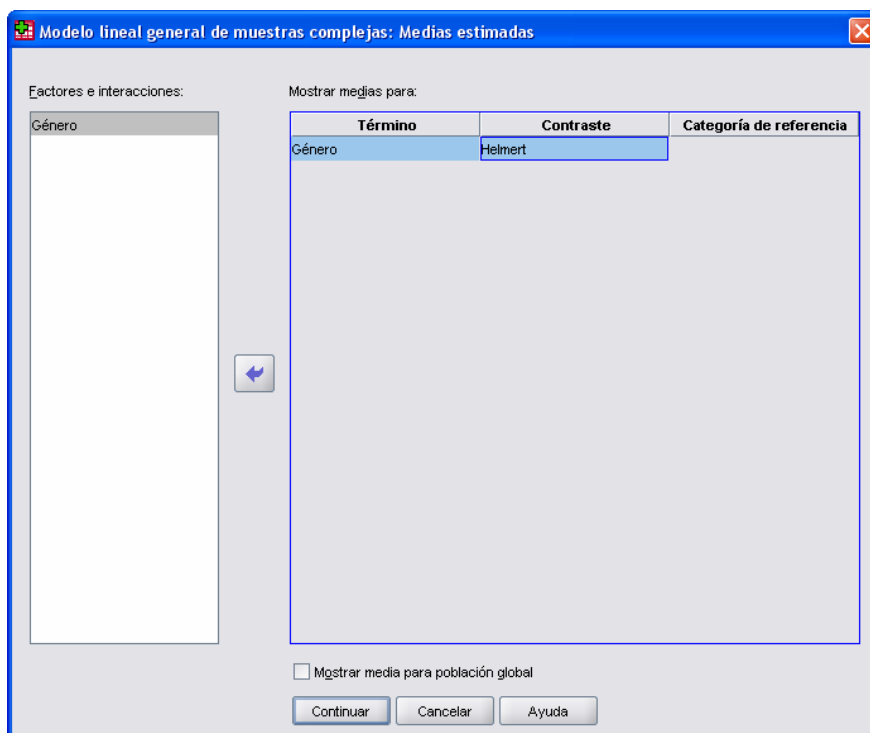


6. Enseguida dar clic en *Estadísticos*.

7. En el cuadro de diálogo **Modelo lineal general de muestras complejas: Estadísticos**, seleccionar **Estimación**, **Error típico** y **Prueba t**. Posteriormente dar clic en *Continuar* y después dar clic en *Medias Estimadas*.



8. En el cuadro de diálogo **Modelo lineal general de muestras complejas: Medias estimadas** seleccionar en la lista de **Factores e interacciones** a la variable **Género** y agregarla a la lista **Mostrar medias para**, en el listado de **Contraste** seleccionar **Helmert**, dar clic en *Continuar* y después en *Aceptar*.



9. Revisar en el cuadro **Resultados de los contrastes individuales**, la columna **Estimación de contraste** que muestra la diferencia entre los dos grupos y la columna **Sig** que presenta el estadístico de significancia.

Resultados de los contrastes individuales

Contraste de Helmert Género	Estimación del contraste	Valor hipotetizado	Diferencia (estimación - hipotetizado)	Error típico	gl1	gl2	F de Wald	Sig.
Nivel 1.00 Femenino frente a último	.087	.000	.087	2.642	1.000	2587.000	.001	.974

10. Al establecer la hipótesis nula como la igualdad entre las medias, si el estadístico en la columna **Sig** es menor que 0.05, como regla se rechaza la hipótesis nula, lo que implica que hay diferencia significativa entre las medias. En caso contrario, no se tienen evidencias para rechazar la hipótesis nula, lo que indica que no existen diferencias significativas entre las medias.

11. El procedimiento concluye con la toma de decisión de rechazar o no la hipótesis nula.

En el ejemplo se muestra **Sig =.974** > 0.05 lo que indica que no hay una diferencia significativa entre las medias de desempeño de hombres y mujeres.

DIFERENCIAS DE MEDIAS DE DESEMPEÑO SIGNIFICATIVAS ENTRE PAÍSES

Para obtener la diferencia de medias de desempeño entre países y comprobar si es significativa, se realiza el siguiente procedimiento:

1. Obtener las medias y los errores estándar, siguiendo los pasos descritos en el procedimiento **Medias de desempeño y sus intervalos de confianza**, con excepción del paso 7, en el que debe seleccionar a la **media** y el **error típico** en lugar del **intervalo de confianza**

2. Calcular el estadístico de prueba T dado por la siguiente fórmula:

$$T = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}}$$

Donde $\hat{\mu}_1$ y $\hat{\mu}_2$ son las medias de desempeño de los países a comparar y $\hat{\sigma}_1^2$ y $\hat{\sigma}_2^2$ son los errores estándar al cuadrado de las medias de desempeño.

3. Establecer la hipótesis nula como la igualdad entre las medias, si el estadístico de prueba T es mayor que 1.96, como regla se rechaza la hipótesis nula, lo que implicaría que hay diferencia significativa entre las medias. En caso contrario, no se tienen evidencias para rechazar la hipótesis nula, lo que indica que no existen diferencias significativas entre las medias.

4. El procedimiento concluye con la toma de decisión de rechazar o no la hipótesis nula.

Por ejemplo, para comparar las medias de desempeño logradas por los estudiantes de tercer año de México y Cuba en Lectura, los estadísticos de ambos países se obtienen como se indica en el paso 1 y se presentan en la siguiente tabla:

Estadísticos univariantes

Código del país 1-21			Estimación	Error típico
Cuba	Media	Puntaje del estudiante estandarizado	626.8889	7.28858
México	Media	Puntaje del estudiante estandarizado	530.4426	4.12067

Con los datos de la tabla el estadístico de prueba T se calcula como:

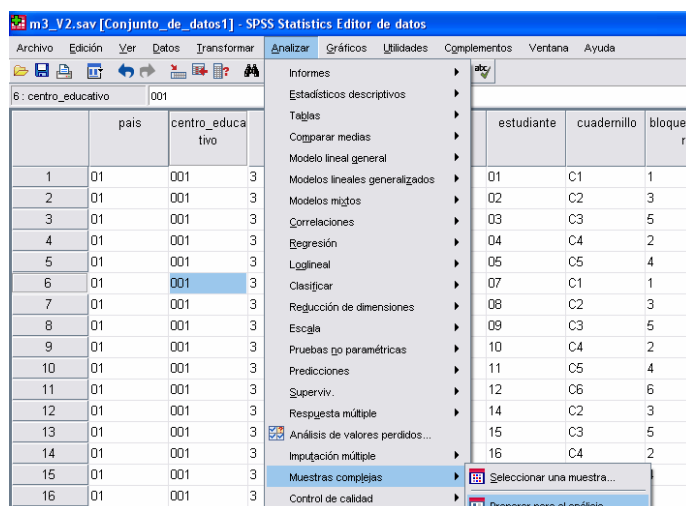
$$T = \frac{626.8889 - 530.4426}{\sqrt{7.28858^2 + 4.12067^2}} = 11.52$$

Como $T > 1.96$ se rechaza la hipótesis nula, es decir, existen diferencias significativas entre las medias de México y Cuba en Lectura de tercer año.

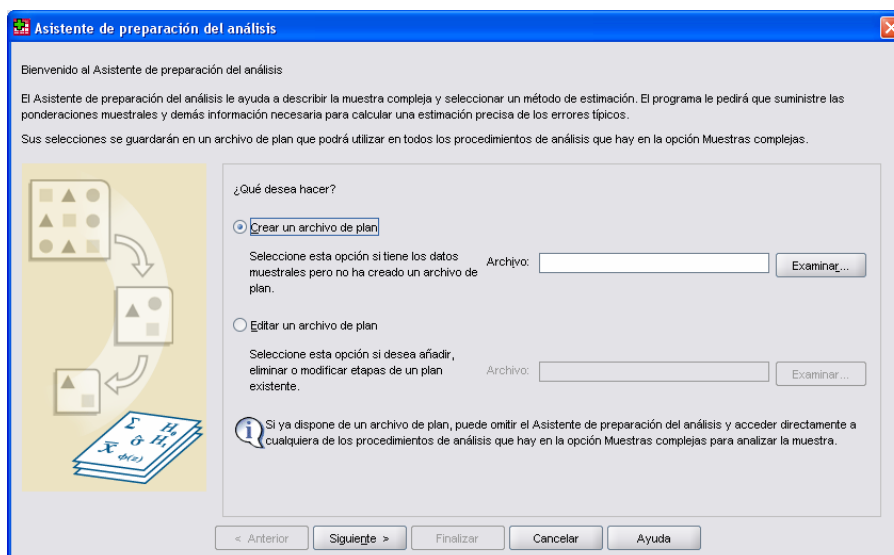
CREACIÓN DEL ARCHIVO *PLAN*SERCE.CSPLAN EN SPSS

A continuación se muestra cómo se generó el plan *PlanSERCE.csplan*:

1. Abrir en SPSS cualquiera de las bases de datos SERCE (L3.sav, M3.sav, L6.sav, M6.sav).
2. Seleccionar el menú **Analizar > Muestras Complejas > Preparar para el análisis**.

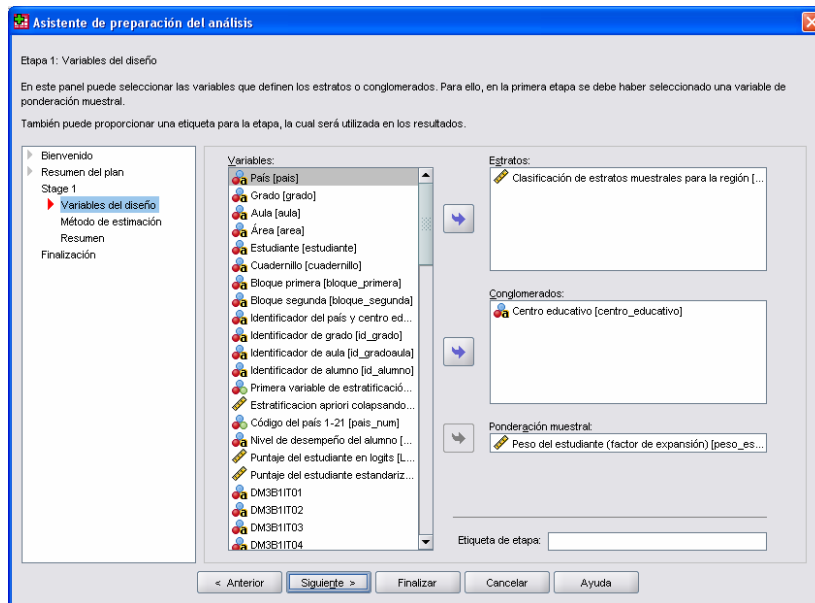


3. En el cuadro de diálogo **Asistente de preparación del análisis** elegir **Crear un archivo de plan**. Posteriormente dar clic en **Examinar**. En el cuadro de diálogo **Guardar datos como** indicar dónde se guardará el plan así como su nombre. En este caso se llama *PlanSERCE.csplan*, dar clic **Guardar** y después en **Siguiente**.

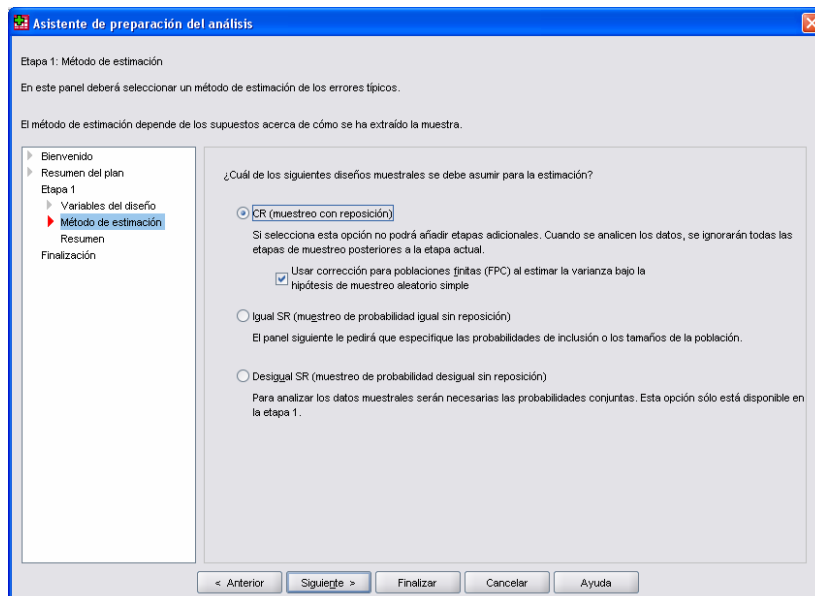


4. En el cuadro de diálogo **Asistente de preparación del análisis (Etapa 1: Variables del diseño)** introducir la información sobre el diseño muestral complejo utilizado en el SERCE, es decir los estratos, los conglomerados y los pesos muestrales. En las bases del SERCE los estratos están disponibles en la variable ***estratoregional***, los conglomerados están definidos por la variable ***Centro_educativo***, los pesos muestrales en la variable ***Peso del estudiante***.

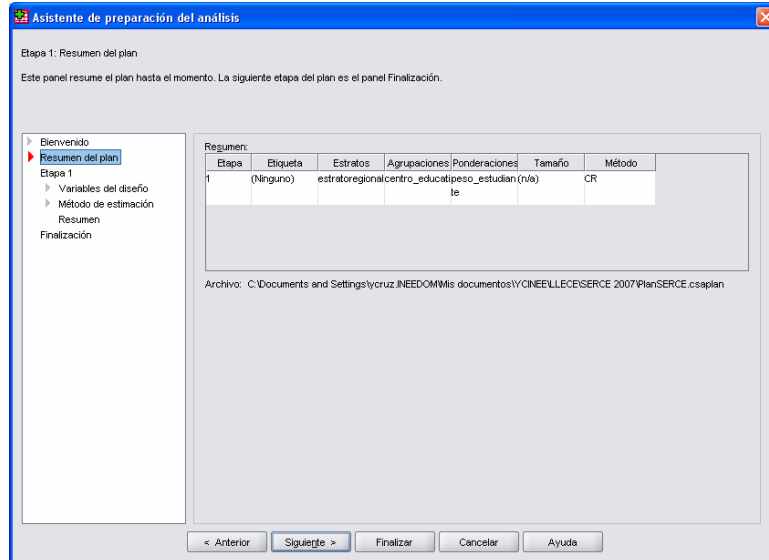
Las variables deben ser colocadas en los listados correspondientes, realizado esto dar clic en *Siguiente*.



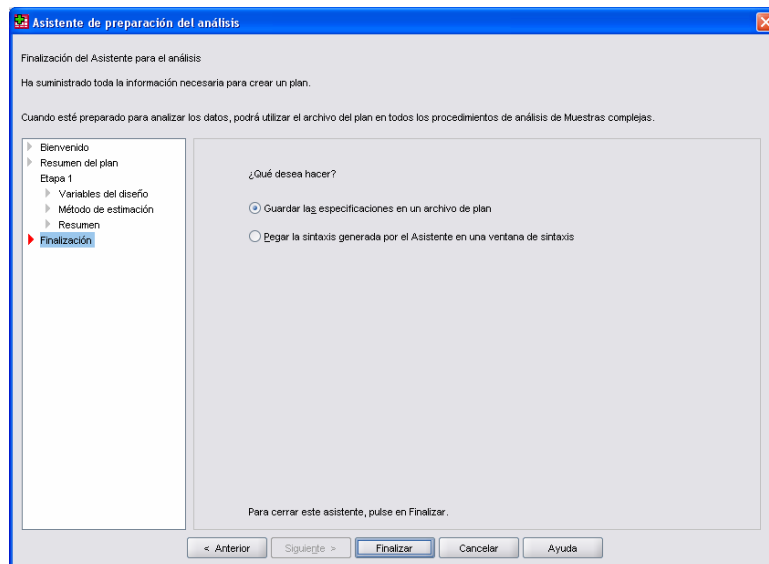
5. En siguiente cuadro de diálogo seleccionar **CR (muestreo con reposición)** y dar clic en *Siguiente*.



6. En el cuadro de diálogo **Asistente de preparación del análisis**, se muestran las variables del diseño y el método de muestreo en el apartado **Etapa 1: Resumen del plan**. Dar clic en *Siguiente*.



7. En siguiente cuadro de diálogo seleccionar **Guardar las especificaciones de un archivo de plan** y posteriormente dar clic en *Finalizar*, con lo cual se concluye el procedimiento.



Es importante mencionar que si se requiere realizar un análisis con alguna de las bases de los cuestionarios de estudiantes, de profesores o de escuelas, es necesario crear un nuevo plan siguiendo el procedimiento descrito, poniendo como estratos la variable **AdmRural**, como conglomerados la variable **centro_educativo** y en la ponderación seleccionando la variable correspondiente (**peso_estudiante**, **peso docente** y **peso_escuela**) de acuerdo con la base del análisis que se desea.

NOTAS SOBRE LOS PESOS MUESTRALES

Para complementar la comprensión de la estructura de las bases de datos del SERCE, a continuación se proporcionan algunas notas referentes a los pesos muestrales, las cuales tienen como objetivo dar un apoyo para poder realizar análisis no descritos en este documento.

Nota 1: Cada una de las bases de datos cuenta con los pesos muestrales para realizar estimaciones poblacionales.

Nota 2: Los pesos están calculados como factores de expansión, es decir, como el inverso de la probabilidad de selección. En el caso de los estudiantes, por ejemplo, el peso o factor de expansión puede interpretarse como el número de estudiantes de su estrato que un alumno representa.

En términos técnicos, se usó la siguiente lógica para calcular los ponderadores a nivel del alumno:

$$w_{hij} = \frac{N_{nl} N_{hi}}{n_{hl} n_{ni}}$$

N_{hl} = Cantidad de escuelas en el estrato h

N_{hi} = Cantidad de alumnos en la escuela i del estrato h

n_{hl} = Cantidad de escuelas en muestra del estrato h

n_{hi} = Cantidad de alumnos en muestra de la escuela i en el estrato h

Para que los estimadores describan de mejor manera las características de las poblaciones analizadas, es necesario hacer un ajuste a los ponderadores que considere la cantidad de población de los estratos en los que se dividió la muestra. A continuación se presentan las fórmulas correspondientes.

$$w^*_{hij} = w_{hij} \frac{N_h}{\hat{N}_h}$$

$$\hat{N}_h = \sum_{S_h} w_{hij}$$

N_h = Cantidad de alumnos en el estrato h

\hat{N}_h = Estimado de la cantidad de alumnos en el estrato h

Nota 3: Un procedimiento similar se utilizó para calcular los factores de expansión a nivel de escuela, realizando un ajuste para asegurarse que los ponderadores representen al total de la población.

Nota 4: Los factores de expansión para las variables de docentes se calcularon considerando el número de aulas, así como el número total de docentes en la escuela, dado que se carecía de información sobre el número total de docentes en cada estrato y dominio. Por lo tanto, este método debe tomarse con las reservas del caso, ya que en contextos donde hay más de un docente por aula pueden subestimarse las características

de los profesores, así como sobrestimarse en el caso de que un docente esté en más de un aula.

Nota 5: Los factores de expansión de las bases de logro son distintos a los factores de las bases de cuestionarios de alumnos. Dado que todos los estudiantes en las bases de resultados de aprendizaje tienen una puntuación estimada, se optó por calcular el factor de expansión en función al número de estudiantes en cada una de las bases, para salvaguardar que se hiciera una estimación adecuada del logro de la población en cada país. Por ello pueden existir algunas mínimas diferencias entre los pesos de las bases de resultados (L3.sav, M3.sav, L6.sav, M6.sav) del mismo grado pero distinta área. En cuanto a las bases de alumnos, los factores de expansión se calcularon independientemente para cada grado. Se tomó como base el número total de estudiantes que tuvo puntajes válidos en al menos una de las pruebas.

Nota 6: Para realizar las estimaciones se debe usar la variable ***estratoregional***, que es la que define los dominios de la muestra y la variable ***centro_educativo*** que es la que define las UPM. Ambas se encuentran en las bases de resultados (L3.sav, M3.sav, L6.sav, M6.sav).

Nota 7: La variable ***ESTRATOA3ORIGINALR_TCCSC*** de las bases de cuestionarios corresponde a la estratificación del Uruguay, donde se consideró una sobremuestra de escuelas de condición social crítica.