

Análisis de datos como desarrollo

Modelos Multinivel con diseño complejo

Carrasco, D., PhD & Miranda, D., PhD

Centro de Medición MIDE UC

Castillo, C., Mg.

Estudiante de Doctorado Educación UC

LLECE: Taller de Análisis III

Santiago, Marzo 03 de 2022

Taller

Modelos Multinivel

Ajuste de modelos multinivel sobre datos de gran escala incluyendo al diseño muestral

Modelos Multinivel

- En la siguiente sección de este taller, veremos que alternativas hay para ajustar modelos multinivel, incluyendo al diseño muestral de los datos.
 - En particular, revisaremos como ajustar estos modelos en software especializado como:
 - Mplus, Stata, y en R (con la librería WeMix)
 - Vamos a comparar los resultados obtenidos por estos diferentes programas estadísticos.
 - Finalmente, vamos a indicar los diferentes aspectos a considerar cuando se ajustan modelos de este estipo, tales como:
 - Preparación de los pesos muestrales (i.e., desagregación de pesos)
 - Rol de los pesos muestrales
 - Preparación de las variables de estratificación
 - Rol de las variables de estratificación
 - Informatividad del diseño

Modelos Multinivel

- En la siguiente sección de este taller, veremos que alternativas hay para ajustar modelos multinivel, incluyendo al diseño muestral de los datos.
 - En particular, revisaremos como ajustar estos modelos en software especializado como:
 - Mplus, Stata, y en R (con la librería WeMix)
 - Vamos a comparar los resultados obtenidos por estos diferentes programas estadísticos.
 - Finalmente, vamos a indicar los diferentes aspectos a considerar cuando se ajustan modelos de este tipo, tales como:
 - Preparación de los pesos muestrales (i.e., desagregación de pesos)
 - Rol de los pesos muestrales
 - Preparación de las variables de estratificación
 - Rol de las variables de estratificación
 - Informatividad del diseño

Para el caso de ERCE 2019, todas las estimaciones multinivel fueron calculadas empleando **Mplus** incluyendo a los estratos en la estimación, y a los pesos de forma desagregada, re-escalados a la muestra efectiva (Stapleton, 2013). En Mplus, este método fue implementado con el siguiente código:

```
[...]  
VARIABLE =  
STRATIFICATION = id_s;  
CLUSTER = id_j;  
WEIGHT = wi;  
BWEIGHT = wj;  
WTSCALE = ECLUSTER;  
BWTSCALE = SAMPLE;
```

```
[...]
```

En términos generales, los modelos empleados en las presentaciones, informes nacionales e informes de resultados incluyeron a:

Modelo Nulo (Snijker & Bosker, 2012)

$$y_{ij} = \alpha + u_j + \varepsilon_{ij}$$

Modelo Desagregado (Rights et al, 2019)

$$y_{ij} = \alpha + \beta_w x_w + \beta_b x_b + u_j + \varepsilon_{ij}$$

Modelo de medias como outcome (Geiser, 2012)

$$y_{ij} = \alpha + \delta v_{.j} + u_j + \varepsilon_{ij}$$

El resto de los modelos empleados son versiones un poco más complejas a estas especificaciones, en las cuales se incluye a nivel socioeconómico como covariable, en la forma del modelo desagregado. A continuación revisaremos un ejemplo que integra estos modelos.

Taller

Resultados de Modelos Multinivel

Reproducción de resultados presentados

**Resumen de resultados de las relaciones de factores asociados al logro de las pruebas ERCE 2019
en Paraguay (estimados no estandarizados).**

Factores asociados a los estudiantes y sus familias	Factores asociados	Tercer grado		Sexto grado	
		Lectura	Matemática	Lectura	Matemática
Nivel socioeconómico de la familia	29.49***	11.89*	42.97***	29.88***	32.38***
Asistencia del estudiante a educación preescolar	27.61***	23.81**	40.08***	29.89***	40.33***
Pertenencia del estudiante a etnia o pueblo indígena			-76.64***	-57.04***	-66.47***
Repitencia	-74.29***	-46.67***	-76.24***	-48.71***	-57.63***
Inasistencia a la escuela	-12.52**	-6.84	-4.32	-10.41*	-6.92
Días de estudio a la semana	46.20***	37.79***	24.89***	14.96**	21.16***
Involucramiento parental en el aprendizaje	12.54***	4.54	12.16***	6.12**	9.90***
Expectativas educativas de los padres	57.81***	23.95**	84.30***	56.35***	69.09***
Expectativas educativas de los profesores	19.66	6.84	16.54	15.47	28.81**
Interés de los docentes por el bienestar de los estudiantes	36.88*	55.53***	11.90	16.53	18.65
Apoyo al aprendizaje de los estudiantes por parte del docente	42.35**	60.67***	4.86	22.13	21.14
Organización de la enseñanza por parte del docente	18.68	39.23*	10.11	7.50	11.84
Disrupción en el aula	-23.59	-36.03*	-7.06	-5.96	-2.34
Nivel socioeconómico de la escuela	19.78*	-10.33	39.77***	21.76**	18.27*
Administración escolar privada	47.41	56.27	67.28***	58.54**	67.75**
Escuela en lugar urbano (10 mil o más habitantes)	-11.18	-13.24	25.63*	22.51*	28.55*

Nota: *** p <.001, ** p <.01, * p <.05. Los factores asociados a las estudiantes y sus familias incluyen resultados de inferencia poblacional. La segunda parte de la tabla incluye resultados de inferencia condicional a las escuelas, controlando por nivel socioeconómico de los estudiantes y las escuelas.

Resumen de resultados de las relaciones de factores asociados al logro de las pruebas ERCE 2019 en Paraguay (estimados no estandarizados).

Factores asociados a los estudiantes y sus familias	Tercer grado		Sexto grado		
	Lectura	Matemática	Lectura	Matemática	Ciencia
	29.49***	11.89*	42.97***	29.88***	32.38***
Nivel socioeconómico de la familia					
Asistencia del estudiante a educación preescolar	27.61***	23.81**	40.08***	29.89***	40.33***
Pertenencia del estudiante a etnia o pueblo indígena			-76.64***	-57.04***	-66.47***
Repitencia	-74.29***	-46.67***	-76.24***	-48.71***	-57.63***
Inasistencia a la escuela	-12.52**	-6.84	-4.32	-10.41*	-6.92
Días de estudio a la semana	46.20***	37.79***	24.89***	14.96**	21.16***
Involucramiento parental en el aprendizaje	12.54***	4.54	12.16***	6.12**	9.90***
Expectativas educativas de los padres	57.81***	23.95**	84.30***	56.35***	69.09***
Proceso escolar y prácticas docentes	Expectativas educativas de los profesores	19.66	6.84	16.54	15.47
	Interés de los docentes por el bienestar de los estudiantes	36.88*	55.53***	11.90	16.53
	Apoyo al aprendizaje de los estudiantes por parte del docente	42.35**	60.67***	4.86	22.13
Factores de Escuelas	Organización de la enseñanza por parte del docente	18.68	39.23*	10.11	7.50
	Disrupción en el aula	-23.59	-36.03*	-7.06	-5.96
	Nivel socioeconómico de la escuela	19.78*	-10.33	39.77***	21.76**
	Administración escolar privada	47.41	56.27	67.28***	58.54**
	Escuela en lugar urbano (10 mil o más habitantes)	-11.18	-13.24	25.63*	22.51*
					28.55*

Nota: *** p <.001, ** p <.01, * p <.05. Los factores asociados a las estudiantes y sus familias incluyen resultados de inferencia poblacional. La segunda parte de la tabla incluye resultados de inferencia condicional a las escuelas, controlando por nivel socioeconómico de los estudiantes y las escuelas.

El estimado 22.51* corresponde al estimado de δ del siguiente modelo formal:

$$y_{ij} = \alpha + \delta v_{.j} + \pi_w z_w + \pi_b z_b + u_j + \varepsilon_{ij}$$

Este coeficiente expresa la diferencia promedio entre las escuelas que se encuentran en lugares con mayor densidad poblacional (10 mil o más habitantes), y el resto de las escuelas, para los puntajes de Matemáticas, de los estudiantes de Sexto Grado (ERCE 2019) de Paraguay.

En las siguientes tablas compararemos 5 modelos diferentes, de los cuales solo uno reproduce los resultados incluidos en los informes: el **Modelo 2**.

	Modelo 1	Modelo 2
W Y<-Z_W	12.69 (1.79) ***	12.51 (2.25) ***
B Y<-V_B	11.96 (5.98) *	22.51 (10.39) *
B Y<-Z_B	42.73 (8.74) ***	28.10 (6.91) ***
B Y<-STRAT_2	27.09 (8.12) **	
B Y<-STRAT_3	36.99 (11.94) **	
B Y<-STRAT_4	22.30 (11.54)	
B Y<-STRAT_5	38.33 (8.40) ***	
B Y<-STRAT_6	35.57 (33.14)	
B Y<-STRAT_7	10.38 (36.81)	
B Y<-Intercepts	637.36 (3.36) ***	655.07 (4.60) ***
W Y<->Y	4579.00 (121.29) ***	4490.40 (167.57) ***
B Y<->Y	1534.32 (276.14) ***	2480.01 (471.29) ***

*** p < .001; ** p < .01; * p < .05

	Modelo 3	Modelo 4	Modelo 5
W Y<-Z_W	12.51 (2.26) ***	12.51 (2.26) ***	12.51 (2.26) ***
B Y<-V_B	20.52 (7.90) **	20.52 (7.90) **	20.52 (7.90) **
B Y<-Z_B	25.99 (9.36) **	25.99 (9.36) **	25.99 (9.36) **
B Y<-STRAT_2	30.39 (10.46) **	30.39 (10.46) **	30.39 (10.46) **
B Y<-STRAT_3	53.32 (15.42) **	53.32 (15.42) **	53.32 (15.42) **
B Y<-STRAT_4	39.95 (12.97) **	39.95 (12.97) **	39.95 (12.97) **
B Y<-STRAT_5	48.30 (9.77) ***	48.30 (9.77) ***	48.30 (9.77) ***
B Y<-STRAT_6	33.90 (40.72)	33.90 (40.72)	33.90 (40.72)
B Y<-STRAT_7	-15.98 (17.46)	-15.98 (17.46)	-15.98 (17.46)
B Y<-Intercepts	629.46 (4.21) ***	629.46 (4.21) ***	629.46 (4.21) ***
W Y<->Y	4492.12 (169.39) ***	4492.12 (169.39) ***	4492.12 (169.39) ***
B Y<->Y	2155.03 (452.29) ***	2155.03 (452.29) ***	2155.03 (452.29) ***

*** p < .001; ** p < .01; * p < .05

Nota:
 Modelo 1: sin diseño
 Modelo 2: con diseño (Mplus wb2)
 Modelo 3: estratos como efectos fijos (Mplus wb2)
 Modelo 4: estratos como efectos fijos (wb2)
 Modelo 5: estratos como efectos fijos (wa2)

Analicemos las diferencias principales entre estos modelos.

Taller

Resultados de Modelos Multinivel

Comparación de estimados con y sin diseño

Comparación 1: Resultados de matemáticas sexto grado, condicionado por Nivel Socioeconómico (z_w , z_b), y por Densidad Poblacional (10 mil habitantes >) (v_b) (ERCE 2019, Paraguay)

	Modelo 1	Modelo 2
W Y<-Z_W	12.69 (1.79) ***	12.51 (2.25) ***
B Y<-V_B	11.96 (5.98) *	22.51 (10.39) *
B Y<-Z_B	42.73 (8.74) ***	28.10 (6.91) ***
B Y<-STRAT_2	27.09 (8.12) **	
B Y<-STRAT_3	36.99 (11.94) **	
B Y<-STRAT_4	22.30 (11.54)	
B Y<-STRAT_5	38.33 (8.40) ***	
B Y<-STRAT_6	35.57 (33.14)	
B Y<-STRAT_7	10.38 (36.81)	
B Y<-Intercepts	637.36 (3.36) ***	655.07 (4.60) ***
W Y<->Y	4579.00 (121.29) ***	4490.40 (167.57) ***
B Y<->Y	1534.32 (276.14) ***	2480.01 (471.29) ***

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Modelo 1: sin diseño

Modelo 2: con diseño (Mplus wb2)

Comparación 1: Resultados de matemáticas sexto grado, condicionado por Nivel Socioeconómico (z_w , z_b), y por Densidad Poblacional (10 mil habitantes >) (v_b) (ERCE 2019, Paraguay)

	Modelo 1	Modelo 2
W Y<-Z_W	12.69 (1.79) ***	12.51 (2.25) ***
B Y<-V_B	11.96 (5.98) *	22.51 (10.39) *
B Y<-Z_B	42.73 (8.74) ***	28.10 (6.91) ***
B Y<-STRAT_2	27.09 (8.12) **	
B Y<-STRAT_3	36.99 (11.94) **	
B Y<-STRAT_4	22.30 (11.54)	
B Y<-STRAT_5	38.33 (8.40) ***	
B Y<-STRAT_6	35.57 (33.14)	
B Y<-STRAT_7	10.38 (36.81)	
B Y<-Intercepts	637.36 (3.36) ***	655.07 (4.60) ***
W Y<->Y	4579.00 (121.29) ***	4490.40 (167.57) ***
B Y<->Y	1534.32 (276.14) ***	2480.01 (471.29) ***

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Modelo 1: sin diseño

Modelo 2: con diseño (Mplus wb2)

La diferencia más grande entre los resultados de los cinco modelos comparados, corresponde a la comparación entre el modelo 1, y el modelo 2.

El **modelo 1** corresponde a los resultados de un modelo mixto, o modelo multinivel, en el cual se emplea como covariable desagregada a los puntajes de nivel socioeconómico, y además agregamos como factor de escuelas una variable dummy, que distingue entre las escuelas ubicadas en localidades de 10 mil habitantes o más; y el resto de las escuelas ubicadas en lugares de menor densidad poblacional.

El **modelo 2**, es estructuralmente muy similar al modelo 1, con la salvedad de que incluye a los pesos muestrales al interior de las escuelas, y entre las escuelas. Y además, considera la información de los estratos del diseño, para realizar las estimaciones.

- El aspecto que explica en mayor medida la diferencia entre estos modelos, es la inclusión de los pesos muestrales. En particular, los pesos de las escuelas. Estos *relocalizan* a las observaciones e influyen sobre los puntos estimados. Este tipo de influencia se espera, sea mayor sobre los factores escuelas, debido a que el diseño muestral de ERCE 2019, es particularmente informativo entre las escuelas.
- El segundo aspecto que influye sobre los resultados obtenidos, es la inclusión de los **estratos** en las estimaciones. Este efecto es menos obvio en esta comparación (Modelo 1 vs Modelo 2). Ignorar la estratificación del diseño, tiene consecuencias para los errores estandar obtenidos, pero no para los puntos estimados (Heeringa et al., 2009).

En resumen, ignorar el diseño por completo, tiene consecuencias notorias respecto a los estimados que podemos obtener. Especialmente entre factores de escuelas, donde el diseño es más informativo.

A continuación revisaremos como cambian los estimados del modelo 2 si ignoramos los estratos.

Taller

Resultados de Modelos Multinivel

Comparación de estimados ignorando los estratos

Comparación 2: Resultados de matemáticas sexto grado, condicionado por Nivel Socioeconómico (z_w , z_b), y por Densidad Poblacional (10 mil habitantes >) (v_b) (ERCE 2019, Paraguay)

	Modelo 2	Modelo 2b
W Y<-Z_W	12.51 (2.25) ***	12.51 (2.26) ***
B Y<-V_B	22.51 (10.39) *	22.51 (10.52) *
B Y<-Z_B	28.10 (6.91) ***	28.10 (7.04) ***
B Y<-Intercepts	655.07 (4.60) ***	655.07 (4.78) ***
W Y<->Y	4490.40 (167.57) ***	4490.40 (168.40) ***
B Y<->Y	2480.01 (471.29) ***	2480.01 (471.79) ***

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Modelo 2: con diseño (Mplus wb2)

Modelo 2b: con diseño (Mplus wb2), ignorando estratos

	Wemix	Stata
W Y<-Z_W	12.51 (2.26) ***	12.51 (2.26) ***
B Y<-V_B	22.48 (10.55) *	22.48 (10.55) *
B Y<-Z_B	28.12 (7.00) ***	28.10 (7.00) ***
B Y<-Intercepts	655.07 (4.78) ***	655.07 (4.78) ***

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Wemix: con pesos (wb2), sin estratos

Stata: con pesos (wb2), sin estratos

Comparación 2: Resultados de matemáticas sexto grado, condicionado por Nivel Socioeconómico (z_w , z_b), y por Densidad Poblacional (10 mil habitantes >) (v_b) (ERCE 2019, Paraguay)

	Modelo 2	Modelo 2b
W Y<-Z_W	12.51 (2.25) ***	12.51 (2.26) ***
B Y<-V_B	22.51 (10.39) *	22.51 (10.52) *
B Y<-Z_B	28.10 (6.91) ***	28.10 (7.04) ***
B Y<-Intercepts	655.07 (4.60) ***	655.07 (4.78) ***
W Y<->Y	4490.40 (167.57) ***	4490.40 (168.40) ***
B Y<->Y	2480.01 (471.29) ***	2480.01 (471.79) ***

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Modelo 2: con diseño (Mplus wb2)

Modelo 2b: con diseño (Mplus wb2), ignorando estratos

	Wemix	Stata
W Y<-Z_W	12.51 (2.26) ***	12.51 (2.26) ***
B Y<-V_B	22.48 (10.55) *	22.48 (10.55) *
B Y<-Z_B	28.12 (7.00) ***	28.10 (7.00) ***
B Y<-Intercepts	655.07 (4.78) ***	655.07 (4.78) ***

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Wemix: con pesos (wb2), sin estratos

Stata: con pesos (wb2), sin estratos

Como regla general, ignorar los estratos del muestreo entre escuelas tiene como costo la obtención de estimados con errores de mayor tamaño, que los que obtendríamos si incluyeramos la información de los estratos en las estimaciones (Heeringa et al., 2009; Stapleton 2006, 2008).

Si comparamos los resultados obtenidos por **WeMix** y por **Stata**, estos resultados son equivalentes entre sí, y similares a los obtenidos por Mplus. Se diferencian sólo por .03 en los puntos estimados, del coeficiente de interés ($\delta = 22.51$, SE = 2.25, $p < .05$).

Mplus es de los pocos software que posee opciones específicas para integrar la información de los estratos en sus estimaciones. En cambio, **WeMix** y **Stata** no poseen opciones específica para la inclusión de la información de los estratos.

En estos escenarios, una opción es incluir la información de los estratos como efectos fijos en el modelo (ver Wes et al., 2014, p54). Es decir, incluir a cada uno de los estratos, como variables dummy.

En las siguientes láminas compararemos los resultados obtenidos por Mplus, WeMix y Stata, incluyendo esta última opción.

Taller

Resultados de Modelos Multinivel

Comparación de estimados incluyendo a los estratos como efectos fijos

Comparación 3: Resultados de matemáticas sexto grado, condicionado por Nivel Socioeconómico (z_w , z_b), y por Densidad Poblacional (10 mil habitantes >) (v_b) (ERCE 2019, Paraguay)

	Modelo 2	Modelo 2c
W Y<-Z_W	12.51 (2.25) ***	12.51 (2.26) ***
B Y<-V_B	22.51 (10.39) *	20.52 (7.90) **
B Y<-Z_B	28.10 (6.91) ***	25.99 (9.36) **
B Y<-Intercepts	655.07 (4.60) ***	629.46 (4.21) ***
W Y<->Y	4490.40 (167.57) ***	4492.12 (169.39) ***
B Y<->Y	2480.01 (471.29) ***	2155.03 (452.29) ***
B Y<-STRAT_2		30.39 (10.46) **
B Y<-STRAT_3		53.32 (15.42) **
B Y<-STRAT_4		39.95 (12.97) **
B Y<-STRAT_5		48.30 (9.77) ***
B Y<-STRAT_6		33.90 (40.72)
B Y<-STRAT_7		-15.98 (17.46)

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Modelo 2: con diseño (Mplus wb2)

Modelo 2c: con diseño (Mplus wb2), estratos como efectos fijos

	Wemix	Stata
W Y<-Z_W	12.51 (2.26) ***	12.51 (2.26) ***
B Y<-V_B	20.51 (7.91) *	20.51 (7.91) *
B Y<-Z_B	26.01 (9.33) ***	26.01 (9.33) ***
B Y<-Intercepts	629.47 (4.20) ***	629.47 (4.20) ***

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Wemix: con pesos (wb2), estratos como efectos fijos

Stata: con pesos (wb2), estratos como efectos fijos

Comparación 3: Resultados de matemáticas sexto grado, condicionado por Nivel Socioeconómico (z_w , z_b), y por Densidad Poblacional (10 mil habitantes >) (v_b) (ERCE 2019, Paraguay)

	Modelo 2	Modelo 2c
W Y<-Z_W	12.51 (2.25) ***	12.51 (2.26) ***
B Y<-V_B	22.51 (10.39) *	20.52 (7.90) **
B Y<-Z_B	28.10 (6.91) ***	25.99 (9.36) **
B Y<-Intercepts	655.07 (4.60) ***	629.46 (4.21) ***
W Y<->Y	4490.40 (167.57) ***	4492.12 (169.39) ***
B Y<->Y	2480.01 (471.29) ***	2155.03 (452.29) ***
B Y<-STRAT_2		30.39 (10.46) **
B Y<-STRAT_3		53.32 (15.42) **
B Y<-STRAT_4		39.95 (12.97) **
B Y<-STRAT_5		48.30 (9.77) ***
B Y<-STRAT_6		33.90 (40.72)
B Y<-STRAT_7		-15.98 (17.46)

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Modelo 2: con diseño (Mplus wb2)

Modelo 2c: con diseño (Mplus wb2), estratos como efectos fijos

	Wemix	Stata
W Y<-Z_W	12.51 (2.26) ***	12.51 (2.26) ***
B Y<-V_B	20.51 (7.91) *	20.51 (7.91) *
B Y<-Z_B	26.01 (9.33) ***	26.01 (9.33) ***
B Y<-Intercepts	629.47 (4.20) ***	629.47 (4.20) ***

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Wemix: con pesos (wb2), estratos como efectos fijos

Stata: con pesos (wb2), estratos como efectos fijos

Si incorporamos los estratos como efectos fijos, obtenemos errores estándar más pequeños para el coeficiente de interés:

- (Mplus) sin estratos $V_B = 22.51$, $SE = 10.52$.
- (Mplus) con estratos $V_B = 22.52$, $SE = 7.90$
- (WeMix, Stata) con estratos $V_B = 22.51$, $SE = 7.91$

	Modelo 2	Modelo 2b
W Y<-Z_W	12.51 (2.25) ***	12.51 (2.26) ***
B Y<-V_B	22.51 (10.39) *	22.51 (10.52) *
B Y<-Z_B	28.10 (6.91) ***	28.10 (7.04) ***
B Y<-Intercepts	655.07 (4.60) ***	655.07 (4.78) ***
W Y<->Y	4490.40 (167.57) ***	4490.40 (168.40) ***
B Y<->Y	2480.01 (471.29) ***	2480.01 (471.79) ***

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Modelo 2: con diseño (Mplus wb2)

Modelo 2b: con diseño (Mplus wb2), ignorando estratos

	Wemix	Stata
W Y<-Z_W	12.51 (2.26) ***	12.51 (2.26) ***
B Y<-V_B	22.48 (10.55) *	22.48 (10.55) *
B Y<-Z_B	28.12 (7.00) ***	28.12 (7.00) ***
B Y<-Intercepts	655.07 (4.78) ***	655.07 (4.78) ***

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Wemix: con pesos (wb2), sin estratos

Stata: con pesos (wb2), sin estratos

Taller

Ajustando modelos multinivel

Qué precauciones tener para reproducir los resultados de ERCE 2019

Comparación 3: Resultados de matemáticas sexto grado, condicionado por Nivel Socioeconómico (z_w , z_b), y por Densidad Poblacional (10 mil habitantes >) (v_b) (ERCE 2019, Paraguay)

	Modelo 2	Modelo 2c
W Y<-Z_W	12.51 (2.25) ***	12.51 (2.26) ***
B Y<-V_B	22.51 (10.39) *	20.52 (7.90) **
B Y<-Z_B	28.10 (6.91) ***	25.99 (9.36) **
B Y<-Intercepts	655.07 (4.60) ***	629.46 (4.21) ***
W Y<->Y	4490.40 (167.57) ***	4492.12 (169.39) ***
B Y<->Y	2480.01 (471.29) ***	2155.03 (452.29) ***
B Y<-STRAT_2		30.39 (10.46) **
B Y<-STRAT_3		53.32 (15.42) **
B Y<-STRAT_4		39.95 (12.97) **
B Y<-STRAT_5		48.30 (9.77) ***
B Y<-STRAT_6		33.90 (40.72)
B Y<-STRAT_7		-15.98 (17.46)

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Modelo 2: con diseño (Mplus wb2)

Modelo 2c: con diseño (Mplus wb2), estratos como efectos fijos

	Wemix	Stata
W Y<-Z_W	12.51 (2.26) ***	12.51 (2.26) ***
B Y<-V_B	20.51 (7.91) *	20.51 (7.91) *
B Y<-Z_B	26.01 (9.33) ***	26.01 (9.33) ***
B Y<-Intercepts	629.47 (4.20) ***	629.47 (4.20) ***

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Wemix: con pesos (wb2), estratos como efectos fijos

Stata: con pesos (wb2), estratos como efectos fijos

Comparación 3: Resultados de matemáticas sexto grado, condicionado por Nivel Socioeconómico (z_w , z_b), y por Densidad Poblacional (10 mil habitantes >) (v_b) (ERCE 2019, Paraguay)

	Modelo 2	Modelo 2c
W Y<-Z_W	12.51 (2.25) ***	12.51 (2.26) ***
B Y<-V_B	22.51 (10.39) *	20.52 (7.90) **
B Y<-Z_B	28.10 (6.91) ***	25.99 (9.36) **
B Y<-Intercepts	655.07 (4.60) ***	629.46 (4.21) ***
W Y<-Y	4490.40 (167.57) ***	4492.12 (169.39) ***
B Y<-Y	2480.01 (471.29) ***	2155.03 (452.29) ***
B Y<-STRAT_2		30.39 (10.46) **
B Y<-STRAT_3		53.32 (15.42) **
B Y<-STRAT_4		39.95 (12.97) **
B Y<-STRAT_5		48.30 (9.77) ***
B Y<-STRAT_6		33.90 (40.72)
B Y<-STRAT_7		-15.98 (17.46)

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Modelo 2: con diseño (Mplus wb2)

Modelo 2c: con diseño (Mplus wb2), estratos como efectos fijos

	Wemix	Stata
W Y<-Z_W	12.51 (2.26) ***	12.51 (2.26) ***
B Y<-V_B	20.51 (7.91) *	20.51 (7.91) *
B Y<-Z_B	26.01 (9.33) ***	26.01 (9.33) ***
B Y<-Intercepts	629.47 (4.20) ***	629.47 (4.20) ***

*** p < 0.001; ** p < 0.01; * p < 0.05

Nota:

Wemix: con pesos (wb2), estratos como efectos fijos

Stata: con pesos (wb2), estratos como efectos fijos

Para reproducir los resultados incluidos en los informes y presentaciones de ERCE 2019 tenemos que asegurarnos de los siguientes aspectos:

- ajustar el modelo especificado
 - En este respecto es crucial el centrado de variables de cada modelo.
- incluir la información del diseño de forma adecuada
 - Esto implica incluir los pesos desagregados, escalados a la muestra efectiva
- para obtener los mismos resultados de forma exacta, se requiere emplear un software que aplique el mismo método de cálculo que el software empleado.
 - WeMix y Stata producen los mismos resultados entre sí.
 - Mplus, Wemix, y Stata producen resultados que son similares sustantivamente, pero no exactamente iguales.
 - En caso de no tener acceso a Mplus, Wemix y Stata son buenas opciones.
 - La inclusión de los estratos como efectos fijos es una opción razonable para producir resultados con errores estándar de menor tamaño.
 - Sin embargo, la interpretación del intercepto del modelo va a cambiar producto de como sean ingresados los estratos al modelo.
 - Si se quiere conservar la interpretación del intercepto del modelo, los efectos fijos de los estratos al modelo, pueden ser ingresados como variables centradas. De este modo el intercepto del modelo ajustado, puede seguir siendo considerado la gran media de todas las escuelas.

Taller

Preparación del entorno R

Librerías que emplearemos para ajustar modelos multínivel

Librerías en uso

- A la derecha de la presente lámina incluimos el conjunto de paquetes de R que empleamos para esta sesión.
- En particular, la librería **library(erce)**, la cual nos da acceso a los datos que serán liberados y que nos permiten replicar los resultados generados para los informes.
- Además en este libreríaa se incluyen unas pocas funciones auxiliares para operar con los datos de ERCE
 - **remove_labels()** que la necesitamos para remover toda meta-data de los datos que estamos empleando
 - **combine_reg()** ocuparemos esta función para producir los output de resultados de regresiones lineales
 - **combine_log()** ocuparemos esta función para producir los output de resultados de regresiones logísticas
- Para ajustar modelos multinivel, las funciones que son de especial importancia son:
 - **c_mean()** ocuparemos esta función para generar los centrados de variables
 - **lsa_weights()** ocuparemos esta función para crear los pesos escalados

Librerías en uso

- A la derecha de la presente lámina incluimos el conjunto de paquetes de R que empleamos para esta sesión.
- En particular, la librería **library(erce)**, la cual nos da acceso a los datos que serán liberados y que nos permiten replicar los resultados generados para los informes.
- Además en este libreríaa se incluyen unas pocas funciones auxiliares para operar con los datos de ERCE
 - **remove_labels()** que la necesitamos para remover toda meta-data de los datos que estamos empleando
 - **combine_reg()** ocuparemos esta función para producir los output de resultados de regresiones lineales
 - **combine_log()** ocuparemos esta función para producir los output de resultados de regresiones logísticas
- Para ajustar modelos multinivel, las funciones que son de especial importancia son:
 - **c_mean()** ocuparemos esta función para generar los centrados de variables
 - **lsa_weights()** ocuparemos esta función para crear los pesos escalados

```
# -----
# librerias
# -----
# Nota: instalar librerias manualmente.

#-----
# librerias para instalar librerias remotas
#-----

install.packages('devtools')
install.packages('remotes')
install.packages('credentials')

#-----
# libreria ERCE con datos
#-----

# needed for some computers with git installed
credentials::set_github_pat()

# library is in development and in private mode
devtools::install_github(
  'dacarras/erce',
  auth_token = 'ghp_0qXfVqkIi4AAZeV984H0GieflB45IN33iIEX',
  force = TRUE)

#-----
# librerías en uso
#-----

install.packages('tidyverse')
install.packages('mitools')
install.packages('survey')
install.packages('srvyr')

#-----
# para ajustar modelos
#-----

install.packages('MplusAutomation') # requiere tener Mplus instalado
install.packages('RStata')          # requiere tener Stata instalado
install.packages('WeMix')

#-----
# mostrar estimados
#-----

install.packages('texreg')
install.packages('miceadds')
```

Taller

Mplus

Código empleado para ajustar modelos multinivel con diseño en ERCE 2019

```
DATA:  
FILE = "mlm.dat";  
TYPE = IMPUTATION;  
  
VARIABLE:  
NAMES = id_s id_j wi wj z_w z_b v_b y;  
MISSING=.;  
  
STRATIFICATION = id_s;  
CLUSTER = id_j;  
WEIGHT = wi;  
BWEIGHT = wj;  
WTSCALE = ECLUSTER;  
BWTSCALE = SAMPLE;  
  
USEVARIABLES =  
y  
z_w  
v_b  
z_b  
;  
WITHIN = z_w;  
BETWEEN = v_b z_b;  
  
ANALYSIS:  
TYPE = COMPLEX TWOLEVEL;  
ESTIMATOR = MLR;  
PROCESSORS = 4;  
  
MODEL:  
%WITHIN%  
y on z_w;  
%BETWEEN%  
y on v_b;  
y on z_b;  
  
OUTPUT:  
STANDARDIZED  
CINTERVAL  
RESIDUAL  
;
```

```

DATA:
FILE = "mlm.dat";
TYPE = IMPUTATION;

VARIABLE:
NAMES = id_s id_j wi wj z_w z_b v_b y;
MISSING=.:;

STRATIFICATION = id_s;
CLUSTER = id_j;
WEIGHT = wi;
BWEIGHT = wj;
WTSCALE = ECLUSTER;
BWTSCALE = SAMPLE;

USEVARIABLES =
y
z_w
v_b
z_b
;
WITHIN = z_w;
BETWEEN = v_b z_b;

ANALYSIS:

TYPE = COMPLEX TWOLEVEL;
ESTIMATOR = MLR;
PROCESSORS = 4;

MODEL:

%WITHIN%
y on z_w;
%BETWEEN%
y on v_b;
y on z_b;

OUTPUT:

STANDARDIZED
CINTERVAL
RESIDUAL
;

```

Código Mplus

- A la izquierda de la presente lámina, se encuentra un ejemplo de código de Mplus, para producir los resultados reportados de Paraguay.
- Las líneas resaltadas son las encargas de escalar los pesos condicionales de las observaciones.
 - **wj** son los pesos de las escuelas
 - **wi** son los pesos de los estudiantes al interior de las escuelas
 - De modo tal que el multiplo de ambos pesos, conforma al peso total de las observaciones. Lo cual implica que **wt = wj * wi**.
- Para que Mplus puede generar las estimaciones sobre diferentes valores plausibles, se requiere crear tantas bases de datos como valores plausibles hayan. En el caso de ERCE 2019, se generan 5 bases de datos, y estas son declaradas como una lista en el archivo **mlm.dat**:

mlm_imp_1.dat
 mlm_imp_2.dat
 mlm_imp_3.dat
 mlm_imp_4.dat
 mlm_imp_5.dat

- En cada una de estas bases de datos, se incluye una columna **y**, sobre la cual se alojan los valores plausibles de la variable de interés. En este ejemplo, los puntajes de matemáticas.
- A continuación incluimos un código en R, que ajusta el modelo de interés en Mplus (ver Código 1.4, en [06_multinivel_codigos.rmd](#).)

```

# [...]
# modelo multinivel con mplus
# ----
#-- crear datos para mplus
#--

data_mixed <- data_model %>%
  dplyr::filter(ctry_name == 'Paraguay') %>%
  rename_all(tolower) %>%
  dplyr::select(
    mat_1, mat_2, mat_3, mat_4, mat_5,
    id_k, id_s, id_j, id_i,
    wb1, wb2, wi, wj,
    z_w, z_b, v_b
  )
#-- crear lista de datos imputados
#--

data_1 <- data_mixed %>% mutate(y = mat_1)
data_2 <- data_mixed %>% mutate(y = mat_2)
data_3 <- data_mixed %>% mutate(y = mat_3)
data_4 <- data_mixed %>% mutate(y = mat_4)
data_5 <- data_mixed %>% mutate(y = mat_5)

data_list <- list(
  data_1,
  data_2,
  data_3,
  data_4,
  data_5
)
#-- crear objeto mplus
#--

mplus_model <- MplusAutomation::mplusObject(
  MODEL = '
  %WITHIN%
  y on z_w;
  %BETWEEN%
  y on v_b;
  y on z_b;
  ,
  ANALYSIS = '
  TYPE = COMPLEX TWOLEVEL;
  ESTIMATOR = MLR;
  PROCESSORS = 4;
  ,
  VARIABLE ='
  STRATIFICATION = id_s;
  CLUSTER = id_j;
  WEIGHT = wi;
  BWEIGHT = wi;
  WTSCALE = ECLUSTER;
  BTSCALE = SAMPLE;
  USEVARIABLES =
  y
  z_w
  v_b
  z_b
  ;
  WITHIN = z_w;
  BETWEEN = v_b z_b;
  ,
  OUTPUT ='
  STANDARDIZED
  CINTERVAL
  RESIDUAL
  ;
  ,
  data = data_list,
  imputed = TRUE
)

```

```

# [...]
#----#
# lista de datos
#----#
# mlm_model_data_list <-
# mlm_imp_1.dat
# mlm_imp_2.dat
# mlm_imp_3.dat
# mlm_imp_4.dat
# mlm_imp_5.dat
#
write(mlm_model_data_list, file="mlm.dat", ncolumns=1)
#----#
# fit models
#----#
fit_4 <- MplusAutomation::mplusModeler(mplus_model,
  modelout = 'mlm.lnp',
  dataout = 'mlm.dat',
  run = 1L,
  hasfilename = FALSE,
  writeData = 'always'
)
#----#
# extraer estimados
#----#
ci_00 <- fit_4 %>%
  purrr::pluck('results') %>%
  purrr::pluck('parameters') %>%
  purrr::pluck('ci.unstandardized') %>%
  tibble::tibble() %>%
  mutate(term = param) %>%
  mutate(level = tolower(BetweenWithin)) %>%
  mutate(ll = low2.5) %>%
  mutate(ul = up2.5) %>%
  mutate(index = paste0(paramHeader, "_", term, "_", level)) %>%
  dplyr::select(index, ll, ul)
r2_00 <- fit_4 %>%
  purrr::pluck('results') %>%
  purrr::pluck('parameters') %>%
  purrr::pluck('r2') %>%
  tibble::tibble() %>%
  mutate(level = tolower(BetweenWithin)) %>%
  mutate(term = c('r2_w', 'r2_b')) %>%
  mutate(object = c('r2', 'r2')) %>%
  mutate(e = as.numeric(est)) %>%
  mutate(se = as.numeric(se)) %>%
  mutate(p = as.numeric(pval)) %>%
  mutate(missing = as.numeric(rate_missing)) %>%
  mutate(index = paste0(term, "_", level)) %>%
  mutate(control = 'yes') %>%
  dplyr::select(index, level, object, term, e, se, p, missing, control)
tabla_1_4 <- fit_4 %>%
  purrr::pluck('results') %>%
  purrr::pluck('parameters') %>%
  purrr::pluck('unstandardized') %>%
  tibble::tibble() %>%
  mutate(term = param) %>%
  mutate(level = tolower(BetweenWithin)) %>%
  mutate(e = est) %>%
  mutate(se = se) %>%
  mutate(p = pval) %>%
  mutate(missing = rate_missing) %>%
  mutate(index = paste0(paramHeader, "_", term, "_", level)) %>%
  dplyr::left_join(., ci_00, by = 'index') %>%
  mutate(control = 'yes') %>%
  mutate(object = c('pi_w', 'epsilon',
  'gamma_b', 'pi_b', 'alpha', 'mu')) %>%
  dplyr::bind_rows(., r2_00) %>%
  dplyr::select(level, object, term, e, se, p, missing, ll, ul, control)
#----#
# mostrar tabla
#----#
tabla_1_4 %>%
knitr::kable(., digits = 2)

```

Taller

WeMix

Código R para ajustar modelos multínivel con diseño empleando datos ERCE 2019

```

# -----#
# modelo multinivel con WeMix
# -----#
# configurar consola
#
options(scipen = 999)
options(digits = 4)
#
# survey object
#
data_wemix <- data_model %>%
  dplyr::filter(ctry_name == 'Paraguay')
#
# fit model
#
wemix_pv_results <- mitools::withPV(
  mapping = mat ~ MAT_1 + MAT_2 + MAT_3 + MAT_4 + MAT_5,
  data = data_wemix,
  action = quote(
    WeMix::mix(
      mat ~ 1 + z_w + z_b + v_b + as.factor(id_s) + (1 | id_j),
      data = data_wemix,
      weights = c('wb1','wb2'),
      cWeights = TRUE
    ),
    rewrite=TRUE
  )
)

# tabla
#
tabla_1_6 <- summary(
  miceadds::pool_mi(
    qhat = list(
      wemix_pv_results[[1]]$coef,
      wemix_pv_results[[2]]$coef,
      wemix_pv_results[[3]]$coef,
      wemix_pv_results[[4]]$coef,
      wemix_pv_results[[5]]$coef
    ),
    se   = list(
      wemix_pv_results[[1]]$SE,
      wemix_pv_results[[2]]$SE,
      wemix_pv_results[[3]]$SE,
      wemix_pv_results[[4]]$SE,
      wemix_pv_results[[5]]$SE
    )
  ) %>%
  tibble::rownames_to_column("term") %>%
  tibble::as_tibble()
)
#
# mostrar tabla
#
tabla_1_6 %>%
  knitr::kable(., digits = 2)

```

```

# -----#
# modelo multinivel con WeMix
# -----#
# configurar consola
#
options(scipen = 999)
options(digits = 4)
#-----#
# survey object
#
data_wemix <- data_model %>%
  dplyr::filter(ctry_name == 'Paraguay')
#-----#
# fit model
#
wemix_pv_results <- mitools::withPV(
  mapping = mat ~ MAT_1 + MAT_2 + MAT_3 + MAT_4 + MAT_5,
  data = data_wemix,
  action = quote(
    Wemix::mix(
      mat ~ 1 + z_w + z_b + v_b + as.factor(id_s) + (1 | id_j),
      data = data_wemix,
      weights = c('wb1','wb2'),
      cWeights = TRUE
    ),
    rewrite=TRUE
  )
)
#-----#
# tabla
#-----#
tabla_1_6 <- summary(
  miceadds::pool_mi(
    qhat = list(
      wemix_pv_results[[1]]$coef,
      wemix_pv_results[[2]]$coef,
      wemix_pv_results[[3]]$coef,
      wemix_pv_results[[4]]$coef,
      wemix_pv_results[[5]]$coef
    ),
    se = list(
      wemix_pv_results[[1]]$SE,
      wemix_pv_results[[2]]$SE,
      wemix_pv_results[[3]]$SE,
      wemix_pv_results[[4]]$SE,
      wemix_pv_results[[5]]$SE
    )
  )
) %>%
  tibble::rownames_to_column("term") %>%
  tibble::as_tibble()
#-----#
# mostrar tabla
#-----#
tabla_1_6 %>%
  knitr::kable(., digits = 2)

```

Código WeMix

- A la izquierda se incluye un ejemplo de cómo ajustar un modelo multinivel en Stata, con los paquetes "mixed" y "pv".
- Para ajustar un modelo multinivel, incluyendo valores plausibles, empleamos la función `mitools::withPV()` que ya hemos visto en las sesiones anteriores.
- Un aspecto a considerar, es que los tiempos de estimación de WeMix, tienden a ser mayores, en contraste a Mplus, y Stata. Y por tanto, esto debe ser considerados para casos en que se requiera ajustar muchos modelos.
- Otro aspecto a considerar, es que el producto de las estimaciones luego de pasar por la función `withPV()`, se produce una lista de *outputs*. Estos outputs, deben ser procesados de cierta forma para obtener los estimados combinados. En este ejemplo, estamos resolviendo este problema con la función `miceadds::pool_mi()`
- Para detalles técnicos de esta librería ver Bailey et al. (2019).
- Para ver un ejemplo desarrollado ver Código 1.7 en [06_multinivel_codigos.rmd](#) y [06_multinivel_codigos.html](#).

```
* =====.
* define folder and open data
* =====.

cd "/Users/d/"
use "erce_paraguay_a6.dta"

* =====.
* fit mlm model with @pv
* =====.

pv, pv(MAT_1 MAT_2 MAT_3 MAT_4 MAT_5): ///
mixed @pv
z_w
z_b
v_b
i.id_s
[pw=wb1] || id_j:, ml pweight(wb2)
estimates store m01

* =====.
* variance
* =====.

/*return residual variance */
_diparm lnsig_e, f(exp(@)^2) d(2*exp(@)^2)
/*return between variance */
_diparm lns1_1_1, f(exp(@)^2) d(2*exp(@)^2)

* =====.
* estimates
* =====.

esttab m01, b(%9.2f) se(%9.2f) wide transform(ln*: exp(2*@) exp(2*@))


```

```

* =====.
* define folder and open data
* =====.

cd "/Users/d/"
use "erce_paraguay_a6.dta"

* =====.
* fit mlm model with @pv
* =====.

pv, pv(MAT_1 MAT_2 MAT_3 MAT_4 MAT_5): ///
mixed @pv
z_w
z_b
v_b
i.id_s
[pw=wb1] || id_j:, ml pweight(wb2)
estimates store m01

* =====.
* variance
* =====.

/*return residual variance */
_diparm lnsig_e, f(exp(@)^2) d(2*exp(@)^2)
/*return between variance */
_diparm lns1_1_1, f(exp(@)^2) d(2*exp(@)^2)

* =====.
* estimates
* =====.

esttab m01, b(%9.2f) se(%9.2f) wide transform(ln*: exp(2*@) exp(2*@))


```

Código Stata

- A la izquierda se incluye un ejemplo de cómo ajustar un modelo multinivel con diseño en **Stata**.
- Para ajustar un modelo multinivel, incluyendo valores plausibles, empleamos la función **pv**, **pv()**: del paquete **pv**. Similar a como **withPV()** funciona, basta con incluir las variables que incluyen a los diferentes valores plausibles dentro de la función, y esta se encarga de generar las estimaciones para cada valor plausible.
 - Para instalar el paquete **pv** en Stata, se debe ejecutar el comando **ssc install pv**
- Para ajustar un modelo multinivel dentro de la función **pv**, empleamos la función **mixed**. Dentro de esta función podemos incluir los pesos escalados, en la línea **[pw=wb1] || id_j:, ml pweight(wb2)**, las variables **wb1** y **wb2** son los pesos escalados a la muestra efectiva.
 - Por su parte **id_j** corresponde al vector que aloja a los cluster o identificadores únicos de escuelas
 - **ml** especifica que el estimador empleado es *maximum likelihood*
- Por defecto, **pv** se encarga de producir los estimados para cada valor plausible, y tambien producir las combinaciones de los estimados siguiendo las reglas de Rubin-Schafer
- Cómo vimos en las comparaciones anteriores, los resultados generados por Stata y Wemix, son equivalentes.
- Para ver un ejemplo desarrollado ver Código 1.7 en [06_multinivel_codigos.rmd](#) y [06_multinivel_codigos.html](#).

Taller

Preparación de Pesos muestrales

Cómo escalar pesos muestrales para ajustar modelos multinivel con diseño

Escalar Pesos

- Para crear los pesos muestrales de los estudiantes y las escuelas, se emplea la siguiente equivalencia:

$$w_{ij} = w_j + w_{i|j}$$

- Lo anterior quiere decir, que el peso total de una observación w_{ij} , puede descomponerse entre el peso de la escuela w_j , y el peso del estudiante al interior de la escuela $w_{i|j}$.
- En el caso de ERCE 2019, los vectores que incluyen estos valores son las siguientes variables
 - WT = w_{ij} peso total del estudiante
 - WJ = w_j peso de la escuela
 - WI = $w_{i|j}$ peso del estudiante al interior de la escuela
- Teniendo esta información ya incluida en la base de datos, basta con crear las variables genéricas de diseño (i.e., `id_k`, `id_j`, `id_i`), y empleando la función `erce::lsa_weights()` podemos generar los pesos muestrales escalados.
 - wa1 y wa2 son los pesos normalizados
 - wb1 y wb2 son los pesos escalados a la muestra efectiva

Escalar Pesos

- Para crear los pesos muestrales de los estudiantes y las escuelas, se emplea la siguiente equivalencia:

$$w_{ij} = w_j + w_{i|j}$$

- Lo anterior quiere decir, que el peso total de una observación w_{ij} , puede descomponerse entre el peso de la escuela w_j , y el peso del estudiante al interior de la escuela $w_{i|j}$.
- En el caso de ERCE 2019, los vectores que incluyen estos valores son las siguientes variables
 - $WT = w_{ij}$ peso total del estudiante
 - $WJ = w_j$ peso de la escuela
 - $WI = w_{i|j}$ peso del estudiante al interior de la escuela
- Teniendo esta información ya incluida en la base de datos, basta con crear las variables genéricas de diseño (i.e., `id_k`, `id_j`, `id_i`), y empleando la función `erce:::lsa_weights()` podemos generar los pesos muestrales escalados.
 - $wa1$ y $wa2$ son los pesos normalizados
 - $wb1$ y $wb2$ son los pesos escalados a la muestra efectiva

- Para replicar los resultados de ERCE 2019 incluidos en presentaciones, informes y reportes, los modelos deben incluir los pesos escalados a la muestra efectiva ($wb1$, $wb2$).
- En términos prácticos, ajustando modelos multinivel sobre los datos ERCE 2019, no se producen diferencias sustantivas al emplear los pesos $wa1$ y $wa2$; en contraste a los pesos $wb1$ y $wb2$. Literatura previa indica que los pesos escalados a la muestra efectiva son más convenientes para producir estimaciones de mejor calidad de varianzas y covarianzas (Stapleton, 2008), y estudios de simulación tienden a favorecer a los pesos escalados a la muestra efectiva bajo variadas condiciones (Bertolet, 2008).
- Para una discusión de mayor detalle respecto al escalamiento de pesos ver: Snijder & Bosker (2012, p239) y Heck & Scott (2020, p346).
- Para ver un ejemplo desarrollado ver Código 1.1 en [06_multinivel_codigos.rmd](#) y [06_multinivel_codigos.html](#).

```
# ejemplo de como crear los pesos escalados para producir modelos multinivel
library(dplyr)
data_model <- erce::erce_2019_qq6 %>%
  # remover meta data
  erce::remove_labels() %>%
  # variables genéricas de clustering
  mutate(id_k = as.numeric(as.factor(paste0(IDCNTRY, " ", STRATA)))) %>%
  mutate(id_s = as.numeric(as.factor(paste0(IDCNTRY, " ", STRATA)))) %>%
  mutate(id_j = as.numeric(as.factor(paste0(IDCNTRY, " ", IDSCHOOL)))) %>%
  mutate(id_i = seq(1:nrow(.))) %>%
  # agregar pesos escalados
  erce::lsa_weights(.,
    id_i = 'id_i',
    id_j = 'id_j',
    id_k = 'id_k',
    wt = 'WT',
    wi = 'WI',
    wj = 'WJ')
```

Muchas gracias!

Referencias

- Bailey, P., Kelley, C., & Nguyen, T. (2019). Introduction to Weighted Mixed-Effects Models With WeMix. https://cran.r-project.org/web/packages/WeMix/vignettes/Introduction_to_Mixed_Effects_Models_With_WeMix.pdf
- Bertolet, M. (2008). To weight or not to weight? Incorporating sampling design into model-based analyses. Carnegie Mellon University.
- Heck, R. H., & Thomas, S. L. (2020). An Introduction to Multilevel Modeling Techniques (4th ed.). Routledge.
- Heeringa, S. G., West, B., & Berglund, P. A. (2009). Applied Survey Data Analysis. Boca Raton, London, New York: Taylor & Francis Group.
- Geiser, C. (2012). Data Analysis with Mplus. The Guilford Press.
- Snijders, T. A. B., & Bosker, R. J. (2012). Multilevel Analysis (2nd ed.). SAGE Publications Ltd.
- Stapleton, L. M. (2013). Incorporating Sampling Weights into Single- and Multilevel Analyses. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), Handbook of International Large scale Assessment: background, technical issues, and methods of data analysis (pp. 363–388). Chapman and Hall/CRC.
- Stapleton, L. M. (2006). An Assessment of Practical Solutions for Structural Equation Modeling with Complex Sample Data. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(1), 28–58.
http://doi.org/10.1207/s15328007sem1301_2
- Stapleton, L. M. (2008). Variance Estimation Using Replication Methods in Structural Equation Modeling With Complex Sample Data. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(2), 183–210.
<http://doi.org/10.1080/10705510801922316>
- Rights, J. D., Preacher, K. J., & Cole, D. A. (2019). The danger of conflating level-specific effects of control variables when primary interest lies in level-2 effects. *British Journal of Mathematical and Statistical Psychology*, 4, bmsp.12194.
<https://doi.org/10.1111/bmsp.12194>
- West, B. T., Welch, K. B., & Galecki, A. T. (2014). Linear Mixed Models: A Practical Guide Using Statistical Software (2nd ed.). Chapman and Hall/CRC.

Carrasco, D., PhD

Centro de Medición MIDE UC,
Pontificia Universidad Católica de Chile
<https://dacarras.github.io/>