

LLECE: Taller de Análisis II

Introducción

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

Centro de Medición MIDE UC

Castillo, C., Mg.

Estudiante Doctorado Educación UC

LLECE: Taller de Análisis III
Santiago, Marzo 02 de 2022

Introducción

Programa

Bloques de esta sesión

Programa Día 1 (Martes 2 de Marzo de 2022)

• Bloque 1 (120 min)

- Presentador: Diego Carrasco
- Introducción a estudios de gran escala
- Desarrollo

• Bloque 2 (120 min)

- Presentador: Diego Carrasco
- Descriptivos
- Regresiones

Bajar las presentaciones y material del material taller:

https://github.com/dacarras/erce_2022_Isa

Programa Día 2 (Martes 3 de Marzo de 2022)

• Bloque 1 (120 min)

- Presentador: Diego Carrasco
- Resultados
- Ejercicios, preguntas y respuestas

• Bloque 2 (120 min)

- Presentador: Diego Carrasco
- Modelos multinivel
- Ejercicios, preguntas y respuestas

Bajar las presentaciones y material del material taller:

https://github.com/dacarras/erce_2022_Isa

Preparación

Instalación

Preparación para el taller

Instalación

Bajar las presentaciones y material del taller:

https://github.com/dacarras/erce_2022_Isa

```
#-----  
# instalar librería para el taller.  
#-----  
  
# algunos computadores requieren esta opción, cuando tienen instalado git  
credentials::set_github_pat()  
  
# la presente librería se encuentra en desarrollo  
devtools::install_github(  
  'dacarras/erce',  
  auth_token = 'ghp_0qXfVqkIi4AAZeV984H0GieflB45IN33iIEX',  
  force = TRUE)  
  
# Nota: nos permitirá a todos acceder a los datos de ERCE 2019
```

Taller

Análisis de datos de ERCE 2019 empleando R

Introducción a estudios de gran escala

Los Estudios de gran escala incluyen diferentes ejemplos:

TIMSS: Trends in International Mathematics and Science Study
PIRLS: Progress in International Reading Literacy Study
ICCS: International Civic and Citizenship Education Study

PISA: Program for International Student Assessment
PIAAC: Programme for the International Assessment of Adult Competencies
TALIS: Teaching and Learning International Survey

ERCE: Estudio Regional Comparativo y Explicativo (ERCE 2019)

SAQMEC: The Southern and Eastern Africa Consortium for Monitoring Educational Quality
PASEC: The Analysis Programme of the CONFEMEN Education Systems
PRIDI: The Regional Project on Child Development Indicators

SEA-PLM: The Southeast Asia Primary Learning Metrics
PILNA: The Pacific Islands Literacy and Numeracy Assessment





Taller

Estudios de gran escala

Cómo acceder a los archivos de datos y documentación

Cómo acceder a datos y documentación de estudios de gran escala

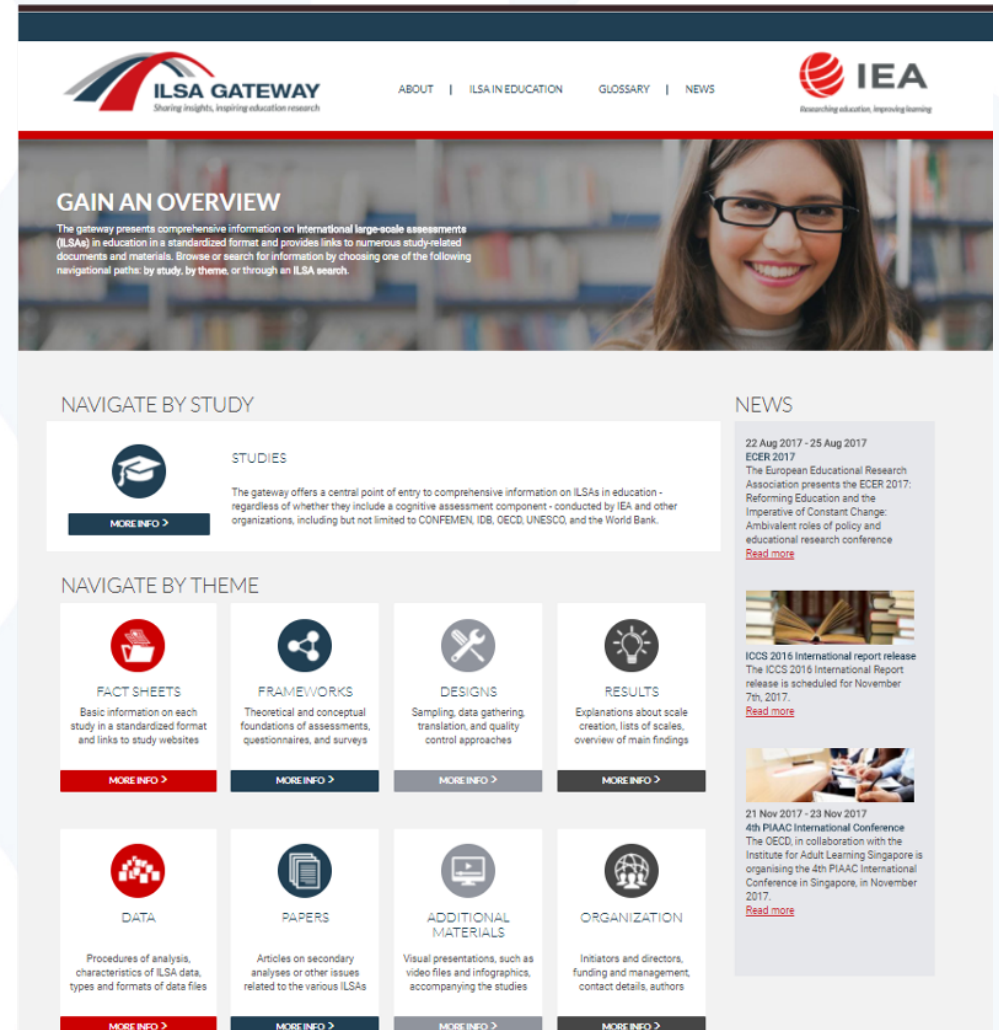
<https://ilsa-gateway.org/>

La página web **ILSA gateway** es un portal web que brinda acceso a diferentes estudios de gran escala disponibles.

Es una iniciativa del **National Center for Education Statistics (NCES)**, el cual fue desarrollado y mantenido por la **International Association for the Evaluation of Educational Achievement (IEA)**.

Es buen punto de partida para explorar los diferentes estudios disponibles.

De esta página los usuarios secundarios pueden acceder a la documentación resumida de los estudios, así como a los links con los cuales bajar los datos de cada estudio.



Taller

Datos de ERCE 2019

Cómo acceder a los datos de ERCE 2019

Cómo acceder a datos de ERCE durante este taller

<https://github.com/dacarras/erce>

Durante este taller emplearemos la librería "erce". Podemos acceder a esta librería, visitando el link anterior.

De todas formas, si han seguido las instrucciones hasta ahora, ya deben tener instalada esta librería.

Esta librería nos permite cargar los datos públicos de ERCE 2019.

Por ejemplo, empleando la librería anterior podremos cargar los datos de tercer y sexto grado, empleando el siguiente código en R:

```
# cargar datos de estudiantes de tercer grado  
erce_a3 <- erce::erce_2019_qa3
```

```
# cargar datos de estudiantes de sexto grado  
erce_a6 <- erce::erce_2019_qa6
```

```
#-----  
# instalar librería para el taller.  
#-----  
  
devtools::install_github('dacarras/erce')  
  
# Nota: nos permitirá a todos acceder a los datos de ERCE 2019.  
  
#-----  
# cargar los datos empleando la librería del taller  
#-----  
  
# cargar datos de estudiantes de tercer grado  
erce_a3 <- erce::erce_2019_qa3  
  
# cargar datos de estudiantes de sexto grado  
erce_a6 <- erce::erce_2019_qa6
```

A continuación, revisaremos las características principales del estudio **ERCE 2019**, que hacen que este estudio califique como un estudio de gran escala (i.e., *international large scale assessment*).

Taller

Diseño muestral y valores plausibles

Características principales de los estudios de gran escala

Características de los estudios de gran escala (i.e., large scale assessment)

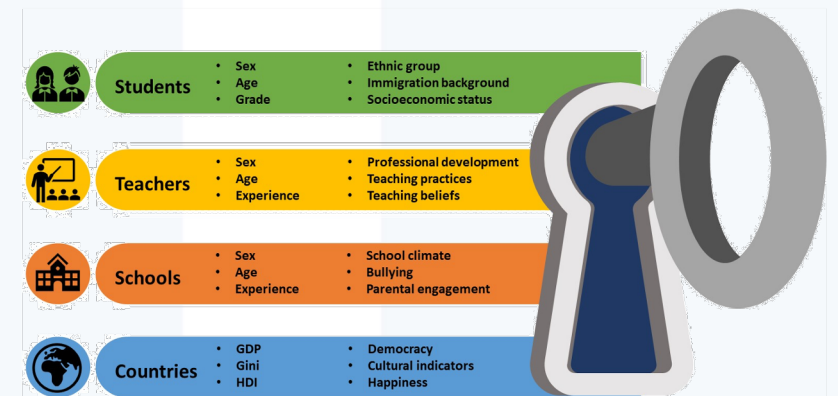
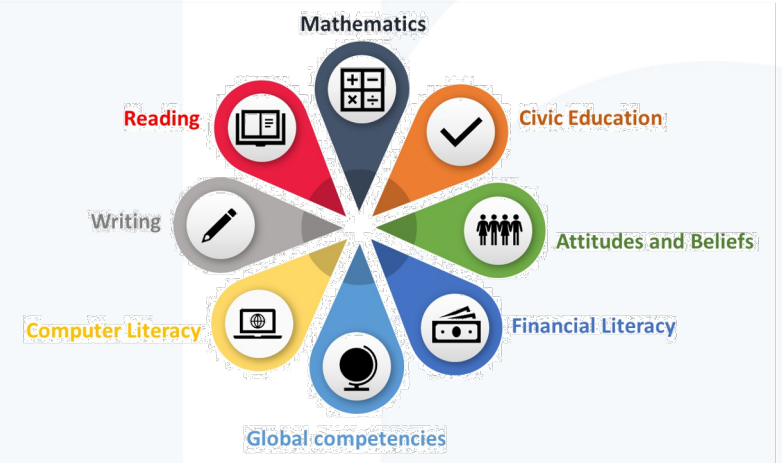
Los estudios de gran de escala se caracterizan por emplear diseños de **muestras complejas**. Esta es la característica principal que nos permite realizar inferencias a la población.

- ERCE 2019 permite realizar inferencias sobre la población de estudiantes de tercer grado, y los estudiantes de sexto grado de todos los países participantes del estudio.
- En la práctica, los usuarios de los datos de ERCE 2019 requieren incluir información del diseño, para producir resultados que son generalizables a la población objetivo del estudio. Lo anterior, incluye a las variables de estratos, escuelas, y pesos muestrales.

Otra característica distintiva de estos estudios, es que poseen algún *outcome* o variable dependiente de interés, la cual se obtiene de forma indirecta. Esta variable es generada mediante un modelo respuesta de variables latentes, y se encuentra representada en las bases de datos mediante diferentes **valores plausibles**.

- ERCE 2019 incluye diferentes valores plausibles (5 valores) que representan la habilidad de los estudiantes en lectura, matemáticas y ciencias respectivamente. En conjunto, estos valores permiten representar la habilidad de interés, y realizar inferencias de resultados incluyendo al error de medición de las pruebas.

Finalmente, este tipo de estudios incluye un conjunto de **cuestionarios de contexto**, los cuales recogen información acerca de diferentes factores asociados a los outcome de interés.



Características de los estudios de gran escala (i.e., large scale assessment)



muestra compleja

Características de los estudios de gran escala (i.e., large scale assessment)



muestra compleja



pruebas

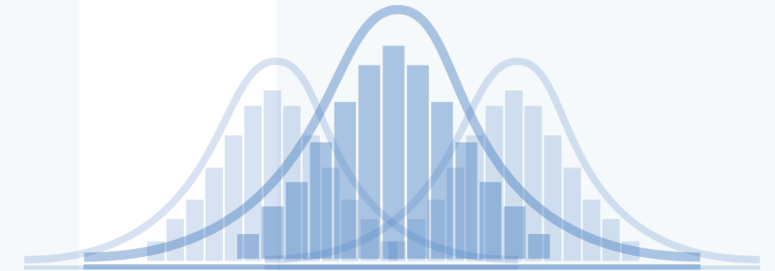
Características de los estudios de gran escala (i.e., large scale assessment)



muestra compleja



pruebas



θ

valores plausibles

Características de los estudios de gran escala (i.e., large scale assessment)



muestra compleja



pruebas



cuestionario de
estudiantes



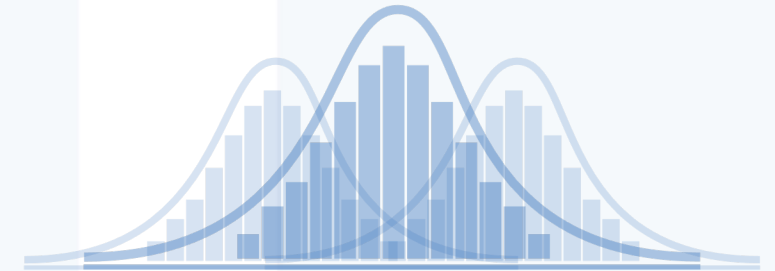
cuestionario de
familias



cuestionario de
profesores



cuestionario de
directores



θ
valores plausibles

Taller

Muestreo y estimados

Diseño muestral y estimaciones

Características de los estudios de gran escala (i.e., large scale assessment)



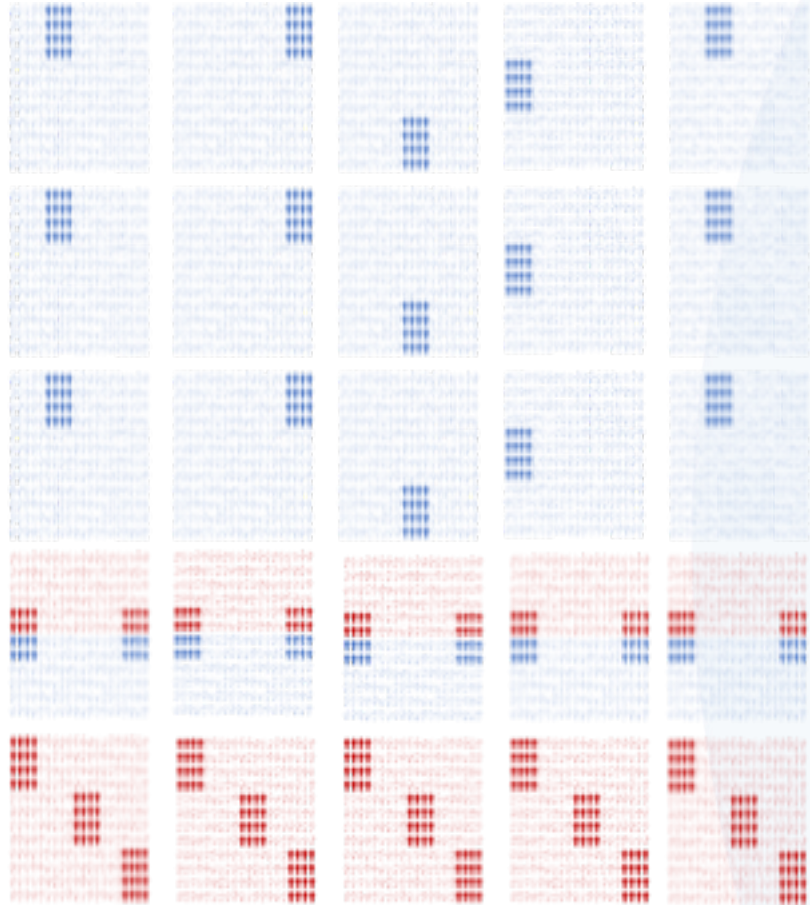
población



μ

Cuando tenemos todos los datos de la población, podemos obtener el parámetro poblacional de forma directa.

Características de los estudios de gran escala (i.e., large scale assessment)



muestra compleja

μ

\bar{x}

Con una muestra, podemos obtener un estadígrafo; pero es incierto que si esta cifra se acerca al parámetro de la población.

Características de los estudios de gran escala (i.e., large scale assessment)



muestra compleja

μ

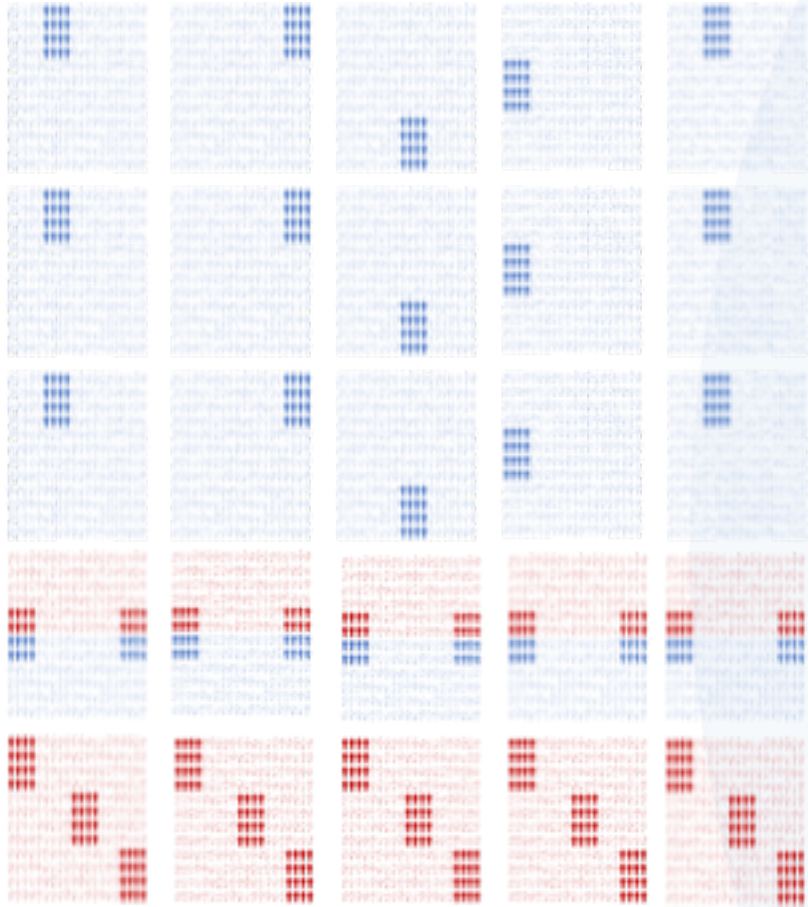
wt

Pesos muestrales corrigen la locación de los estimados.

\bar{x} \hat{x}

Con una muestra, podemos obtener un estadígrafo; pero es incierto que si esta cifra se acerca al parámetro de la población.

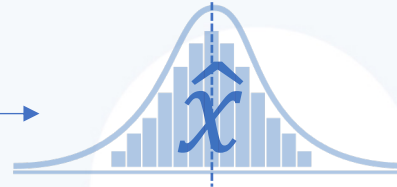
Características de los estudios de gran escala (i.e., large scale assessment)



muestra compleja

μ

\approx



brr, tsl, jkn

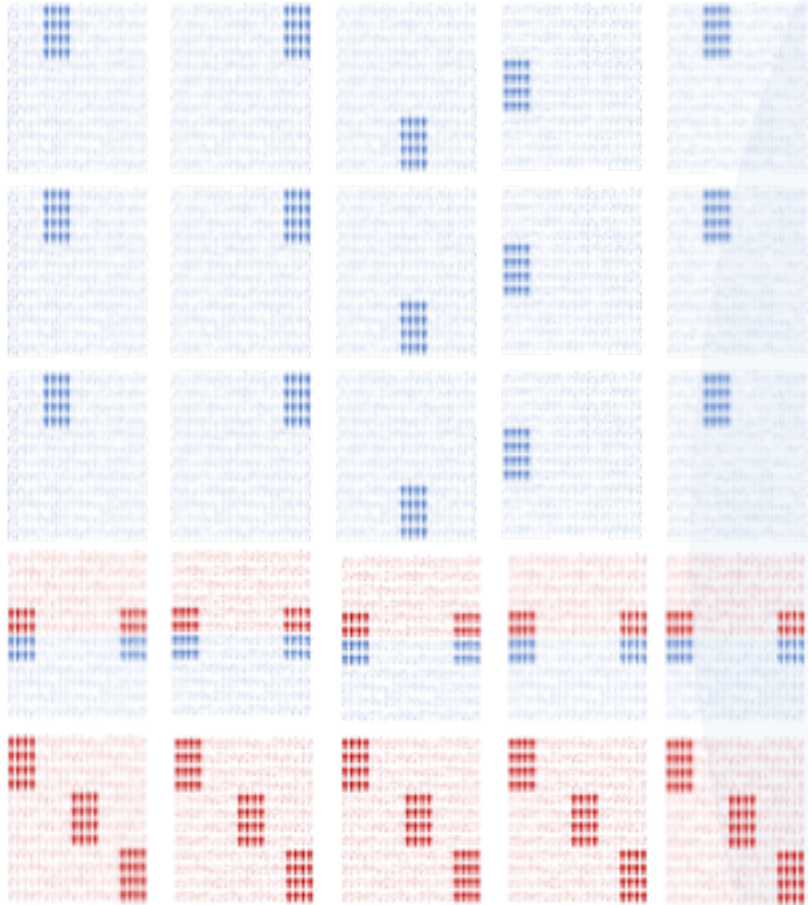
Métodos de muestras complejas, nos permiten calcular errores estándar, y obtener medidas de incertidumbre.

wt

Pesos muestrales corrigen la locación de los estimados.

Con una muestra, podemos obtener un estadígrafo; pero es incierto que si esta cifra se acerca al parámetro de la población.

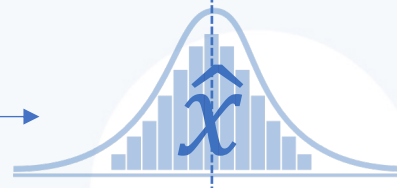
Características de los estudios de gran escala (i.e., large scale assessment)



muestra compleja

μ

\approx



Empleando muestras complejas podemos llegar a estimados razonables acerca de la población.

brr, tsl, jkn

Métodos de muestras complejas, nos permiten calcular errores estándar, y obtener medidas de incertidumbre.

wt

Pesos muestrales corrigen la locación de los estimados.

Con una muestra, podemos obtener un estadígrafo; pero es incierto que si esta cifra se acerca al parámetro de la población.

Taller

Estimados con y sin Diseño

Qué pasa si ignoramos el diseño muestral

Resultados descriptivos con y sin diseño

- Cuando estimamos resultados ignorando el diseño muestral, perdemos la oportunidad de realizar inferencias a la población de interés.
- Nuestros resultados van referir a la muestra de datos observada, y no a la población de interés.
- Los estadígrafos de **locación** pueden estar **sesgados**, porque los pesos se encargan de re-localizar a las cifras como promedios, medianas, cuartiles, y porcentajes.
- Nuestros resultados además van a **subestimar los errores estándar** de nuestras cifras. Nos hacen creer que tenemos mayor precisión de la que, por diseño, efectivamente tenemos.
- Los métodos de corrección de cálculo de errores se encargan de entregarnos medidas de incertidumbre razonables respecto a los resultados calculados

Descripción	Poblacional	Nominal	Delta	Ratio de Rango Intervalar (Con Diseño/Sin Diseño)
ed. no terciaria	.69 CI95% [.67, .71]	.74 CI95% [.73, .74]	0.05	4.00
ed terciaria	.26 CI95% [.24, .28]	.21 CI95% [.21, .21]	0.05	6.67
perdido	.05 CI95% [.04, .05]	.05 CI95% [.05, .06]	0.00	1.00

Nota: descriptivo de estudiantes con al menos un padre con educación terciaria (ERCE 2019, Perú, Estudiantes de Sexto Grado).

Resultados descriptivos con y sin diseño

- Cuando estimamos resultados ignorando el diseño muestral, perdemos la oportunidad de realizar inferencias a la población de interés.
- Nuestros resultados van referir a la muestra de datos observada, y no a la población de interés.
- Los estadígrafos de **locación** pueden estar **sesgados**, porque los pesos se encargan de re-localizar a las cifras como promedios, medianas, cuartiles, y porcentajes.
- Nuestros resultados además van a **subestimar los errores estándar** de nuestras cifras. Nos hacen creer que tenemos mayor precisión de la que, por diseño, efectivamente tenemos.
- Los métodos de corrección de cálculo de errores se encargan de entregarnos medidas de incertidumbre razonables respecto a los resultados calculados

Descripción	Poblacional	Nominal	Delta	Ratio de Rango Intervalar (Con Diseño/Sin Diseño)
ed. no terciaria	.69 CI95% [.67, .71]	.74 CI95% [.73, .74]	0.05	4.00
ed terciaria	.26 CI95% [.24, .28]	.21 CI95% [.21, .21]	0.05	6.67
perdido	.05 CI95% [.04, .05]	.05 CI95% [.05, .06]	0.00	1.00

Nota: descriptivo de estudiantes con al menos un padre con educación terciaria (ERCE 2019, Perú, Estudiantes de Sexto Grado).

Estimaciones sesgadas
En este caso, ignorar el diseño nos puede llevar a sobreestimar o subestimar la cantidad de estudiantes con al menos un padre o madre con educación terciaria.

Resultados descriptivos con y sin diseño

- Cuando estimamos resultados ignorando el diseño muestral, perdemos la oportunidad de realizar inferencias a la población de interés.
- Nuestros resultados van referir a la muestra de datos observada, y no a la población de interés.
- Los estadígrafos de **locación** pueden estar **sesgados**, porque los pesos se encargan de re-localizar a las cifras como promedios, medianas, cuartiles, y porcentajes.
- Nuestros resultados además van a **subestimar los errores estándar** de nuestras cifras. Nos hacen creer que tenemos mayor precisión de la que, por diseño, efectivamente tenemos.
- Los métodos de corrección de cálculo de errores se encargan de entregarnos medidas de incertidumbre razonables respecto a los resultados calculados

Descripción	Poblacional	Nominal	Delta	Ratio de Rango Intervalar (Con Diseño/Sin Diseño)
ed. no terciaria	.69 CI95% [.67, .71]	.74 CI95% [.73, .74]	0.05	4.00
ed terciaria	.26 CI95% [.24, .28]	.21 CI95% [.21, .21]	0.05	6.67
perdido	.05 CI95% [.04, .05]	.05 CI95% [.05, .06]	0.00	1.00

Nota: descriptivo de estudiantes con al menos un padre con educación terciaria (ERCE 2019, Perú, Estudiantes de Sexto Grado).

Errores subestimados
Los rangos intervalares de las estimaciones sin diseño son menores que las estimaciones con diseño. Subestiman entre 4 a 6.67 la incertidumbre alrededor de los estimados.

Resultados descriptivos con y sin diseño

- Cuando estimamos resultados ignorando el diseño muestral, perdemos la oportunidad de realizar inferencias a la población de interés.
 - Nuestros resultados van referir a la muestra de datos observada, y no a la población de interés.
 - Los estadígrafos de **locación** pueden estar **sesgados**, porque los pesos se encargan de re-localizar a las cifras como promedios, medianas, cuartiles, y porcentajes.
 - Nuestros resultados además van a **subestimar los errores estándar** de nuestras cifras. Nos hacen creer que tenemos mayor precisión de la que, por diseño, efectivamente tenemos.
- Los métodos de corrección de cálculo de errores se encargan de entregarnos medidas de incertidumbre razonables respecto a los resultados calculados

Descripción	Poblacional	Nominal	Delta	Ratio de Rango Intervalar (Con Diseño/Sin Diseño)
ed. no terciaria	.69 CI95% [.67, .71]	.74 CI95% [.73, .74]	0.05	4.00
ed terciaria	.26 CI95% [.24, .28]	.21 CI95% [.21, .21]	0.05	6.67
perdido	.05 CI95% [.04, .05]	.05 CI95% [.05, .06]	0.00	1.00

Nota: descriptivo de estudiantes con al menos un padre con educación terciaria (ERCE 2019, Perú, Estudiantes de Sexto Grado).

Estimaciones empleando diseño

Son estadígrafos expandidos a la población empleando la información del diseño, con medidas de incertidumbre adecuadas, que nos brindan información generalizable a la población de interés.

Taller

Estimaciones con valores plausibles

Qué pasaría si empleáramos un solo valor plausible

Resultados con valores plausibles

- Cuando estimamos resultados empleando un solo valor plausible corremos el riesgo de obtener **resultados sesgados** e inconsistentes (los resultados son condicionales al valor plausible elegido).
- Si promediamos los valores plausibles entre los estudiantes (i.e., puntaje promedio), y luego calculáramos resultados podríamos recuperar el punto estimado. Sin embargo **vamos a subestimar los errores**.
- Si estimáramos una relación entre un factor asociado u otra covariable, y alguno de los valores plausibles, los resultados pueden ser **inconsistentes**. Es decir, que es posible que los resultados encontrados sean diferentes condicional al valor plausible empleado (algunos significativos, otros no) (ver Rutkowski et al., 2010)
- En general, los valores plausibles **están diseñados para realizar estimaciones poblacionales** (i.e., realizar inferencias respecto a la población de interés). Y como tal, no es un método recomendable para realizar ranking de estudiantes o de escuelas.

	E	SE	LL	UL	CI range
PV1	758.07	3.73	750.7	765.5	14.8
PV2	760.12	3.61	753.0	767.3	14.3
PV3	758.89	3.72	751.5	766.3	14.8
PV4	760.75	3.64	753.5	768.0	14.5
PV5	758.70	3.74	751.3	766.1	14.8
Promedio	759.31	3.64	752.1	766.5	14.4
Combinados	759.31	3.88	751.7	766.9	15.2

Nota: E = estimado, SE = error estándar, LL = límite inferior de un intervalo de confianza de 95%, UL = intervalo superior del intervalo de confianza de 95%; CI range = distancia del rango intervalar. Medias poblacionales, estimadas con diseño, de matemáticas de estudiantes de sexto grado (ERCE 2019, Perú, Estudiantes de Sexto Grado).

Resultados con valores plausibles

- Cuando estimamos resultados empleando un solo valor plausible corremos el riesgo de obtener **resultados sesgados** e inconsistentes (los resultados son condicionales al valor plausible elegido).
- Si promediamos los valores plausibles entre los estudiantes (i.e., puntaje promedio), y luego calculáramos resultados podríamos recuperar el punto estimado. Sin embargo **vamos a subestimar los errores**.
- Si estimáramos una relación entre un factor asociado u otra covariable, y alguno de los valores plausibles, los resultados pueden ser **inconsistentes**. Es decir, que es posible que los resultados encontrados sean diferentes condicional al valor plausible empleado (algunos significativos, otros no) (ver Rutkowski et al., 2010)
- En general, los valores plausibles **están diseñados para realizar estimaciones poblacionales** (i.e., realizar inferencias respecto a la población de interés). Y como tal, no es un método recomendable para realizar ranking de estudiantes o de escuelas.

	E	SE	LL	UL	CI range
PV1	758.07	3.73	750.7	765.5	14.8
PV2	760.12	3.61	753.0	767.3	14.3
PV3	758.89	3.72	751.5	766.3	14.8
PV4	760.75	3.64	753.5	768.0	14.5
PV5	758.70	3.74	751.3	766.1	14.8
Promedio	759.31	3.64	752.1	766.5	14.4
Combinados	759.31	3.88	751.7	766.9	15.2

Nota: E = estimado, SE = error estándar, LL = límite inferior de un intervalo de confianza de 95%, UL = intervalo superior del intervalo de confianza de 95%; CI range = distancia del rango intervalar. Medias poblacionales, estimadas con diseño, de matemáticas de estudiantes de sexto grado (ERCE 2019, Perú, Estudiantes de Sexto Grado).

Valores Plausibles

Cada uno de los valores plausibles son diferentes "imputaciones" que provienen de un modelo; y como tal, se requiere que sean diferentes.

Resultados con valores plausibles

- Cuando estimamos resultados empleando un solo valor plausible corremos el riesgo de obtener **resultados sesgados** e inconsistentes (los resultados son condicionales al valor plausible elegido).
- Si promediamos los valores plausibles entre los estudiantes (i.e., puntaje promedio), y luego calculáramos resultados podríamos recuperar el punto estimado. Sin embargo **vamos a subestimar los errores**.
- Si estimáramos una relación entre un factor asociado u otra covariable, y alguno de los valores plausibles, los resultados pueden ser **inconsistentes**. Es decir, que es posible que los resultados encontrados sean diferentes condicional al valor plausible empleado (algunos significativos, otros no) (ver Rutkowski et al., 2010)
- En general, los valores plausibles **están diseñados para realizar estimaciones poblacionales** (i.e., realizar inferencias respecto a la población de interés). Y como tal, no es un método recomendable para realizar ranking de estudiantes o de escuelas.

	E	SE	LL	UL	CI range
PV1	758.07	3.73	750.7	765.5	14.8
PV2	760.12	3.61	753.0	767.3	14.3
PV3	758.89	3.72	751.5	766.3	14.8
PV4	760.75	3.64	753.5	768.0	14.5
PV5	758.70	3.74	751.3	766.1	14.8
Promedio	759.31	3.64	752.1	766.5	14.4
Combinados	759.31	3.88	751.7	766.9	15.2

Nota: E = estimado, SE = error estándar, LL = límite inferior de un intervalo de confianza de 95%, UL = intervalo superior del intervalo de confianza de 95%; CI range = distancia del rango intervalar. Medias poblacionales, estimadas con diseño, de matemáticas de estudiantes de sexto grado (ERCE 2019, Perú, Estudiantes de Sexto Grado).

Subestimación de errores

Emplear el promedio de las “imputaciones” nos lleva a subestimar los errores alrededor de los estimados.

Resultados con valores plausibles

- Cuando estimamos resultados empleando un solo valor plausible corremos el riesgo de obtener **resultados sesgados** e inconsistentes (los resultados son condicionales al valor plausible elegido).
- Si promediamos los valores plausibles entre los estudiantes (i.e., puntaje promedio), y luego calculáramos resultados podríamos recuperar el punto estimado. Sin embargo **vamos a subestimar los errores**.
- Si estimáramos una relación entre un factor asociado u otra covariable, y alguno de los valores plausibles, los resultados pueden ser **inconsistentes**. Es decir, que es posible que los resultados encontrados sean diferentes condicional al valor plausible empleado (algunos significativos, otros no) (ver Rutkowski et al., 2010)
- En general, los valores plausibles **están diseñados para realizar estimaciones poblacionales** (i.e., realizar inferencias respecto a la población de interés). Y como tal, no es un método recomendable para realizar ranking de estudiantes o de escuelas.

	E	SE	LL	UL	CI range
PV1	758.07	3.73	750.7	765.5	14.8
PV2	760.12	3.61	753.0	767.3	14.3
PV3	758.89	3.72	751.5	766.3	14.8
PV4	760.75	3.64	753.5	768.0	14.5
PV5	758.70	3.74	751.3	766.1	14.8
Promedio	759.31	3.64	752.1	766.5	14.4
Combinados	759.31	3.88	751.7	766.9	15.2

Nota: E = estimado, SE = error estándar, LL = límite inferior de un intervalo de confianza de 95%, UL = intervalo superior del intervalo de confianza de 95%; CI range = distancia del rango intervalar. Medias poblacionales, estimadas con diseño, de matemáticas de estudiantes de sexto grado (ERCE 2019, Perú, Estudiantes de Sexto Grado).

Estimaciones con valores plausibles

Para realizar estimaciones poblacionales con este tipo de variables, se requiere producir las estimaciones para cada valor, y luego realizar las combinaciones, empleando las reglas de Rubin-Schafer (Schafer, 1997). En general, se emplea software o código que realiza este proceso.

Muchas gracias!

Referencias

- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting. *Educational Researcher*, 39(2), 142-151. <https://doi.org/10.3102/0013189X10363170>
- Schafer, J. L. (1997). Analysis of Incomplete Multivariate Data. In We (Vol. 141, Issue 7). Chapman & Hall/CRC.

*Diego Carrasco, PhD.
Daniel Miranda, PhD.
Carolina Castillo, Mg.*

*Centro de Medición MIDE UC
Pontificia Universidad Católica de Chile*