Introduction to the fourth Regional Comparative and Explanatory Study [Estudio Regional Comparativo y Explicativo] (ERCE 2019) with R.

A large-scale assessment study of Latin-American countries.

Carrasco, D., PhD Centro de Medición MIDE UC, Pontificia Universidad Católica

Natalia López Hornickel, PhD (c) University of Bath, Doctoral Student

Carlos Cayumán Cofré Educational Measurement and Assessment Specialist at OREALC/UNESCO, Santiago

Andrés Sandoval-Hernández University of Bath

Ernestro Treviño Pontificia Universidad Católica

CIES 2023: Workshop
Washington DC, February 18th, 2023

Centro UC
Medición - MIDE

Introduction

Programme

Blocks of this session

February 18th

- First Block (3 hrs)
  - Introduction
  - Data analysis with R
  - Descriptives
- Second Block (2 hrs)
  - Regressions
  - Multilevel Models
- Lunch/coffee break (1 hrs)

Where to download slides and workshop materiales

https://github.com/dacarras/cies_2023_erce_2019

Preparation

Installing libraries

Preparation for the workshop

Installing packages

Where to download slides and workshop materiales

https://github.com/dacarras/cies_2023_erce_2019

```r
#------------------------------------------------
# installing library for the workshop
#------------------------------------------------



# we need this library first
install.packages('devtools’)

# the ERCE library was design for the present workshop
devtools::install_github('dacarras/erce',force = TRUE)


# Note: the second library will give us access to ERCE 2019 data
#       we will install more libraries as we need to.
```

Workshop

Data analysis of ERCE 2019 using R

Introduction to large scale assessment studies

Large scale studies include different examples

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

IEA

OECD

UNESCO

THE WORLD BANK

unicef

Workshop

Large Scale Assessment Studies

How to access to data files and studies documentation

How to access to data files and studies documentation

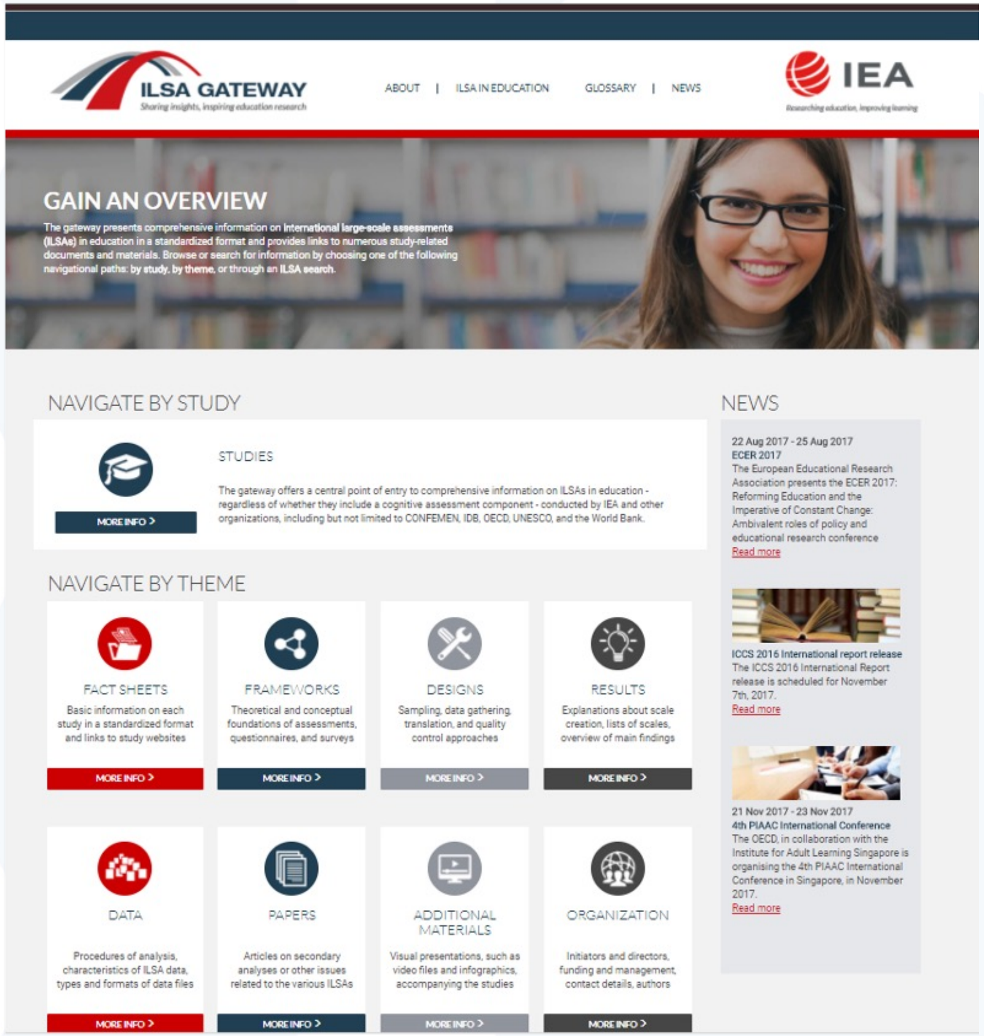
https://ilsa-gateway.org/

The the **ILSA gateway** is web platform where you can find several large-scale assessment studies data files, and study data documentation.

Its an initiative from the **National Center for Education Statistics (NCES)**, developed and maintained by the **International Association for the Evaluation of Educational Achievement (IEA)**.

This is good starting point, to explore different large scale assessment studies.

From this web platform users can get summarized study documentation (e.g., target sample, study design, and data link). However, to have access to the ERCE 2019 study data files, you need to go to the UNESCO website, in spanish.



Workshop

ERCE 2019 data files

How to access ERCE 2019 data files

How to access to data files and studies documentation

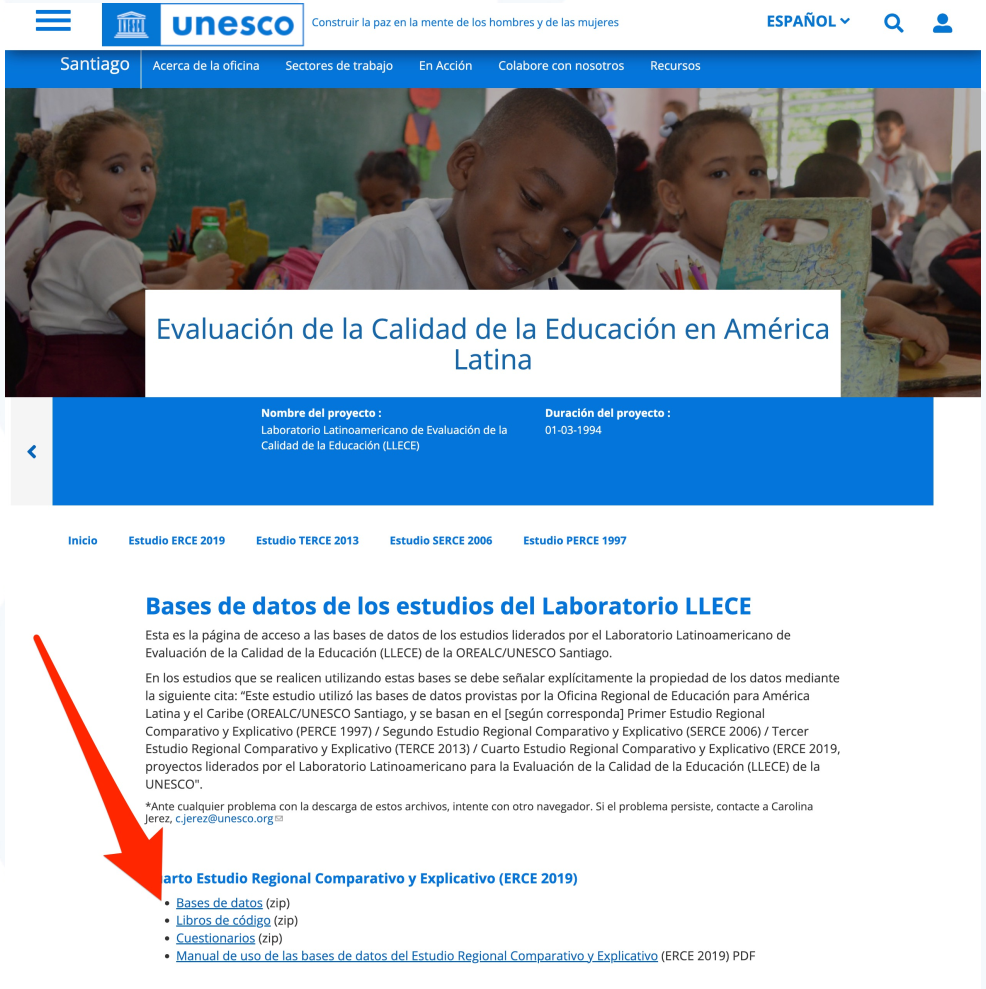
https://es.unesco.org/fieldoffice/santiago/projects/llece/bases

UNESCO website is in spanish.

The link to Access the ERCE 2019 data files is highlighted in the next figure. “Data files” are “Base de datos” en Spanish.

The minimal study documentation for ERCE 2019 data use is in the link “Manual de uso de base de datos ….”.

In the present workshop, we will cover the basic from the ERCE 2019 study.



How to access the ERCE 2019 data files during this workshop

https://github.com/dacarras/erce

The erce library in R, contains a copy of the ERCE 2019 data files.

We develop the present library to conduct workshops and make it easier for users to access the data from the study.

```
#------------------------------------------------
# installing library for the workshop
#------------------------------------------------

devtools::install_github('dacarras/erce')

# Note: this library gives access to a copy of the ERCE 2019
#       data files

#------------------------------------------------
# loading erce data files
#------------------------------------------------

# loading students third grade data
erce_a3 <- erce::erce_2019_qa3

# loading students sixth grade data
erce_a6 <- erce::erce_2019_qa6
```

Using the present code, we can load the student's data on an R session.

In the following slides we will review the main characteristics of the ERCE 2019 study. These are shared feature with other most common large scale assessment studies.

Workshop

Main features of ILSA

Sampling design and plausible values

Main feature of ILSA

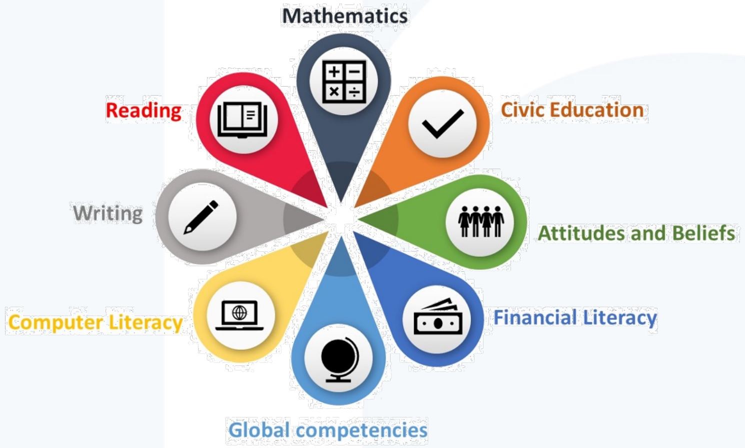
There are two main features of large-scale assessment studies (ILSA), all secondary users need to consider, before generating results.

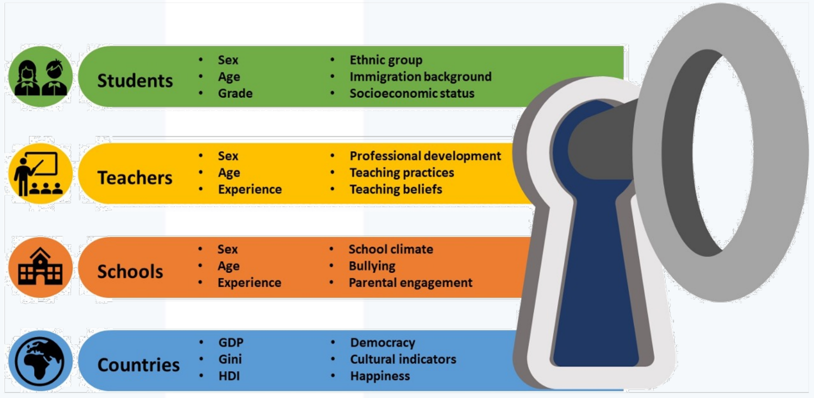
The first feature is the sampling design. Often called "**complex sample design**" in the literature (e.g., Rutkowski et al., 2010). Shared data files contains sampling design information so generated results can be expanded to the sampling frame and make inferences to the population.

In very concrete terms, there is a set of additional variables in the study data files, such as **survey weights, stratification and replicate weights**, that we need to use to generate descriptive, regressions and other results to make inferences to the population of students.

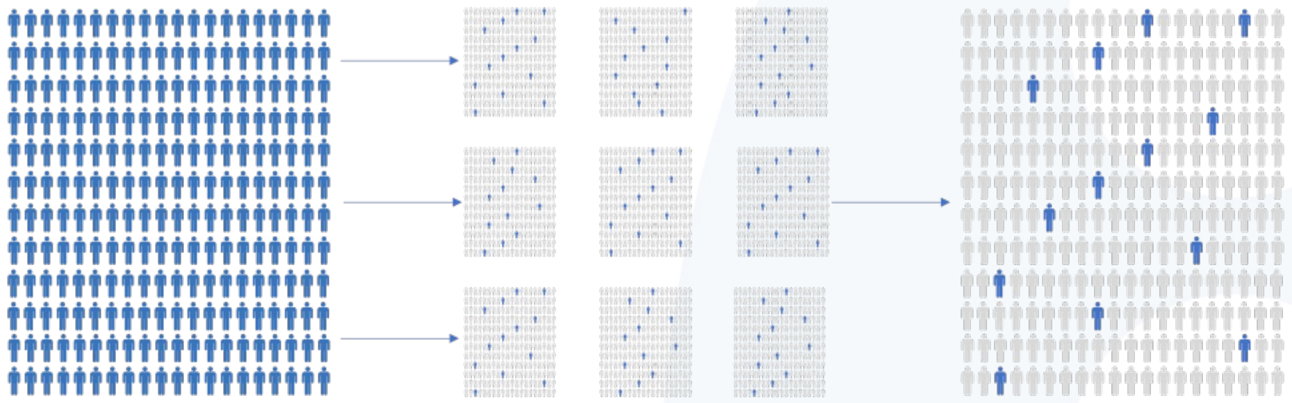
The second feature ILSA studies is the use of plausible values. **Plausible values** are imputed values used to represent the measurement uncertainty of the test. If we use these variable to generate results, we can be sure to make inferences to the population of students, while considering measurement error, and survey error simultaneously.

In essence, most of **ILSA study do not use a single variable to represent test scores, but five different variables or more**. ERCE 2019 includes five different variables for each test, and five values for achievement levels.





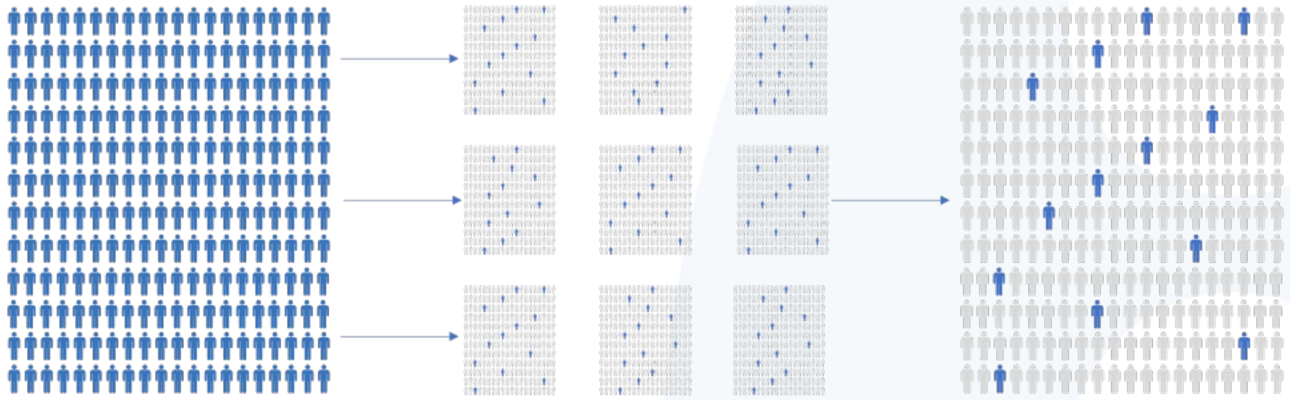
Main feature of ILSA



complex sample

Centro UC
Medición - MIDE

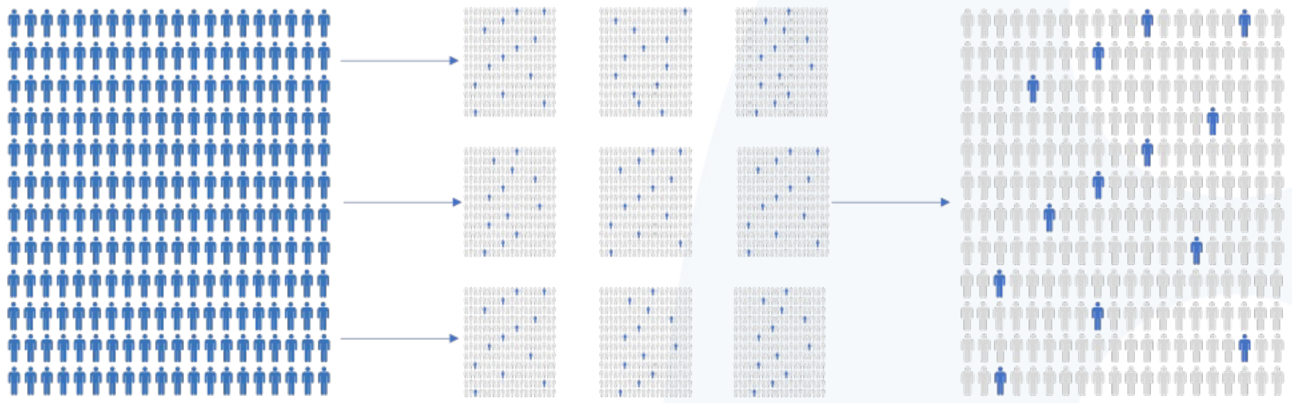
Main feature of ILSA



complex sample

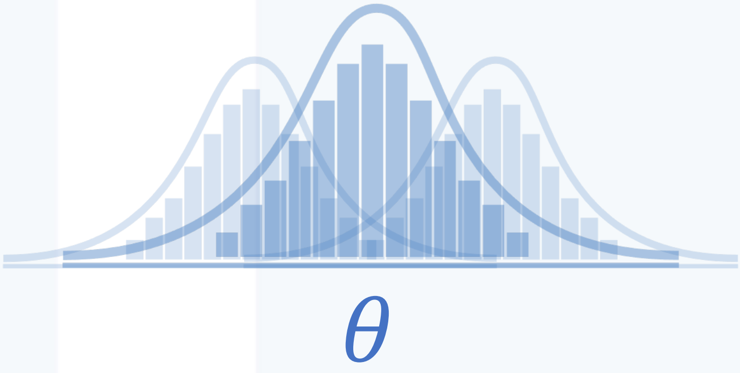
tests

Main feature of ILSA



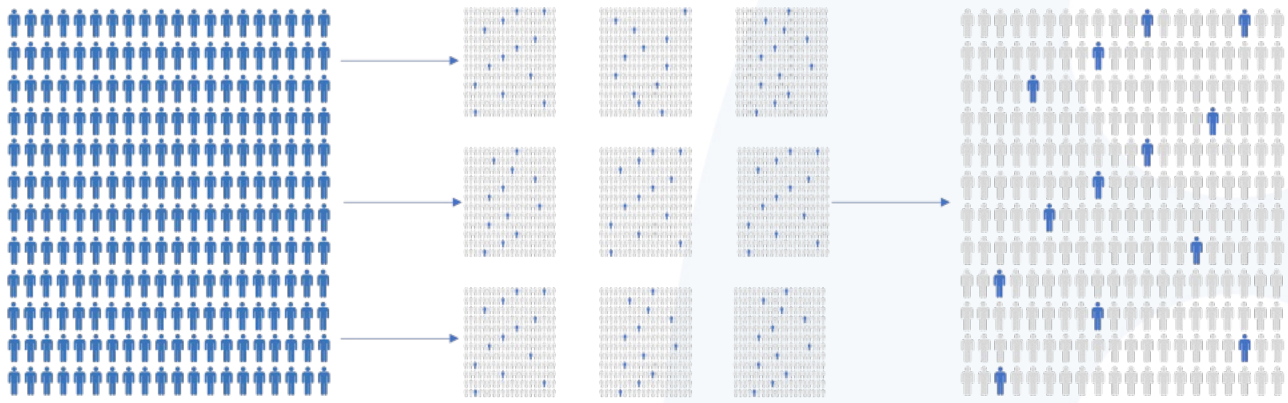
complex sample

tests



plausible values

Main feature of ILSA



complex sample

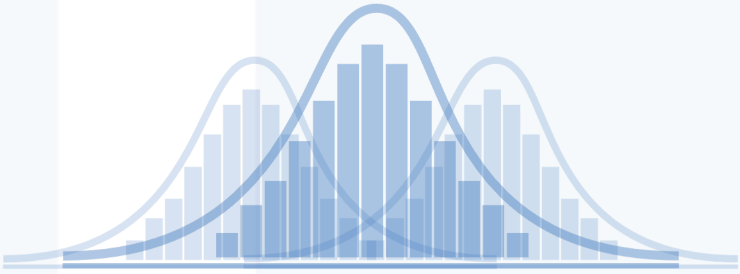
tests

students' questionnaires

families questionnaires

teachers' questionnaires

school principals' questionnaires



θ

plausible values

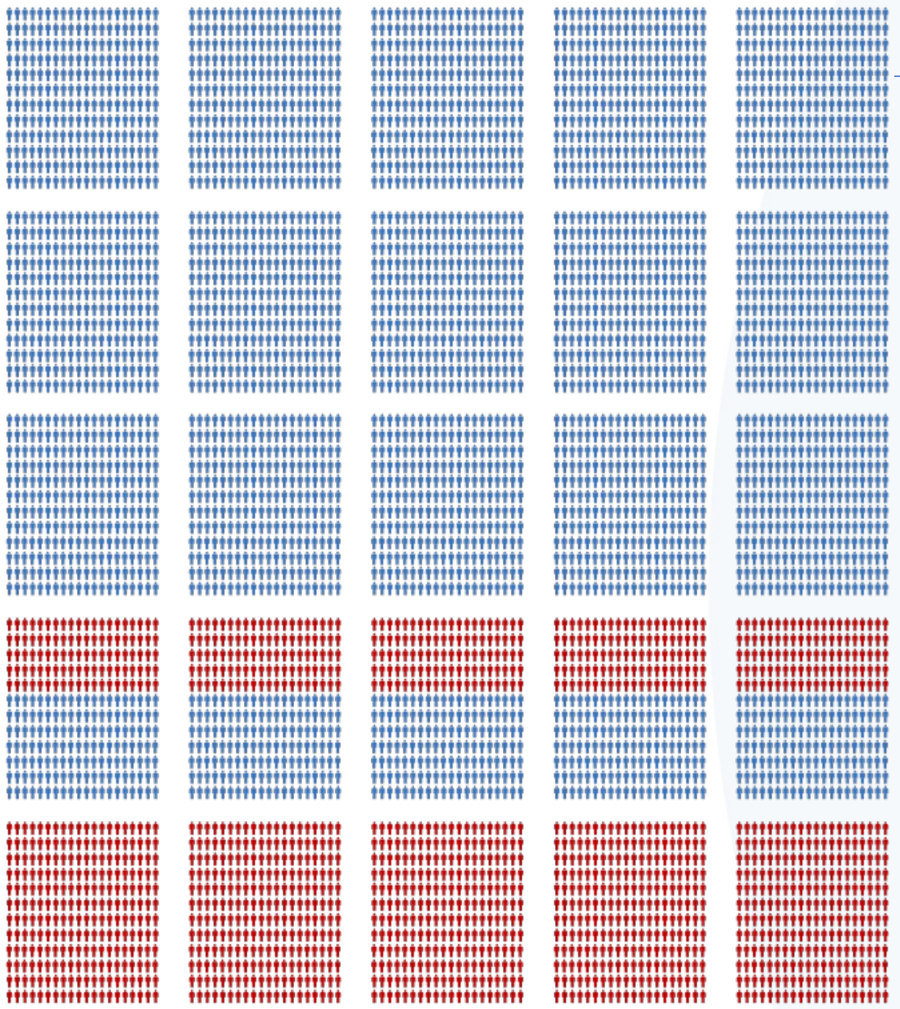
Centro UC
Medición - MIDE

Taller

Muestreo y estimados

Diseño muestral y estimaciones

Sampling from a population



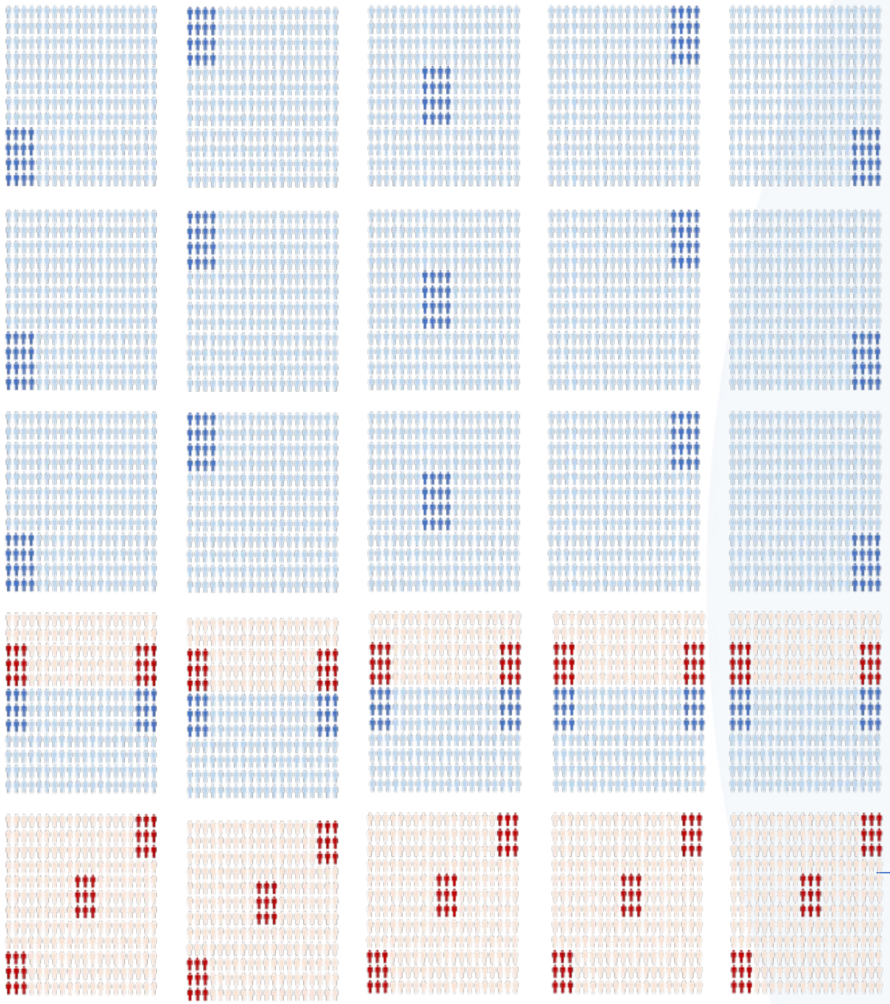
población

μ

When we have a all the observations of a population we can get a population parameter directly, for example the proportion of students with parents with tertiary education)

Centro UC
Medición - MIDE

Sampling from a population



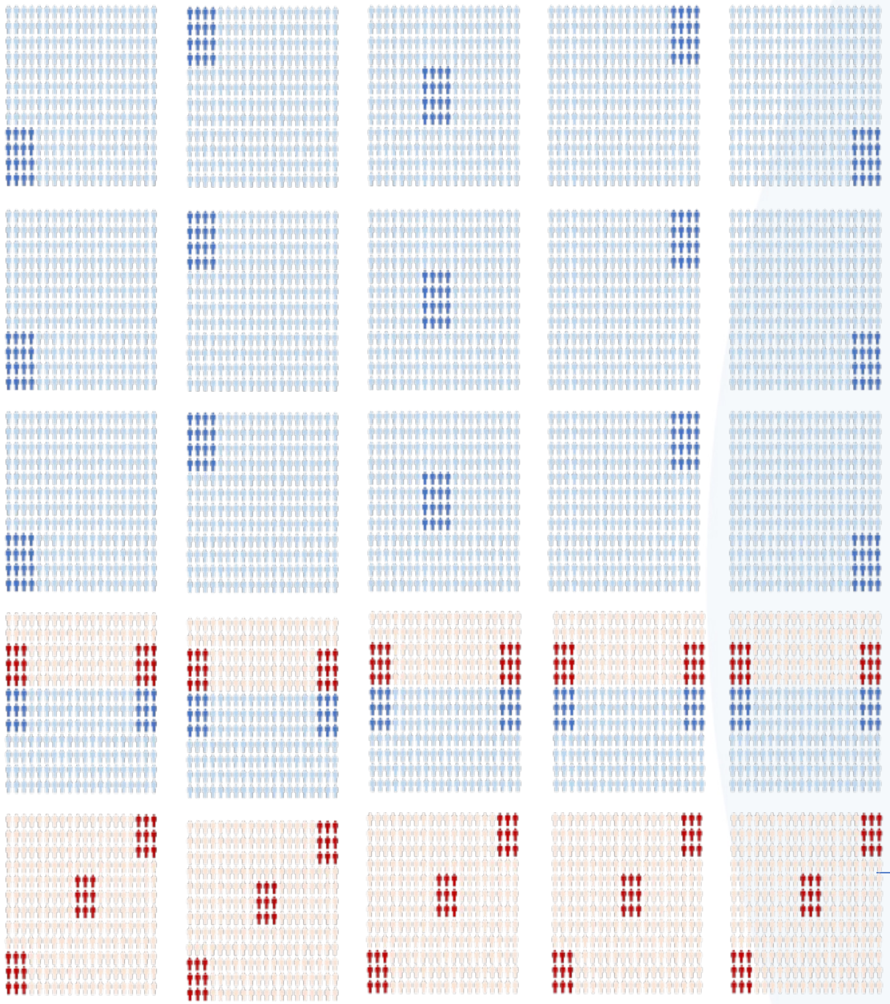
muestra compleja

μ

$\bar{x}$

With a sample of students per schools, we can get an estimate of the proportion of interest. However, is uncertain how close is the estimate to the population parameter.

Sampling from a population



muestra compleja

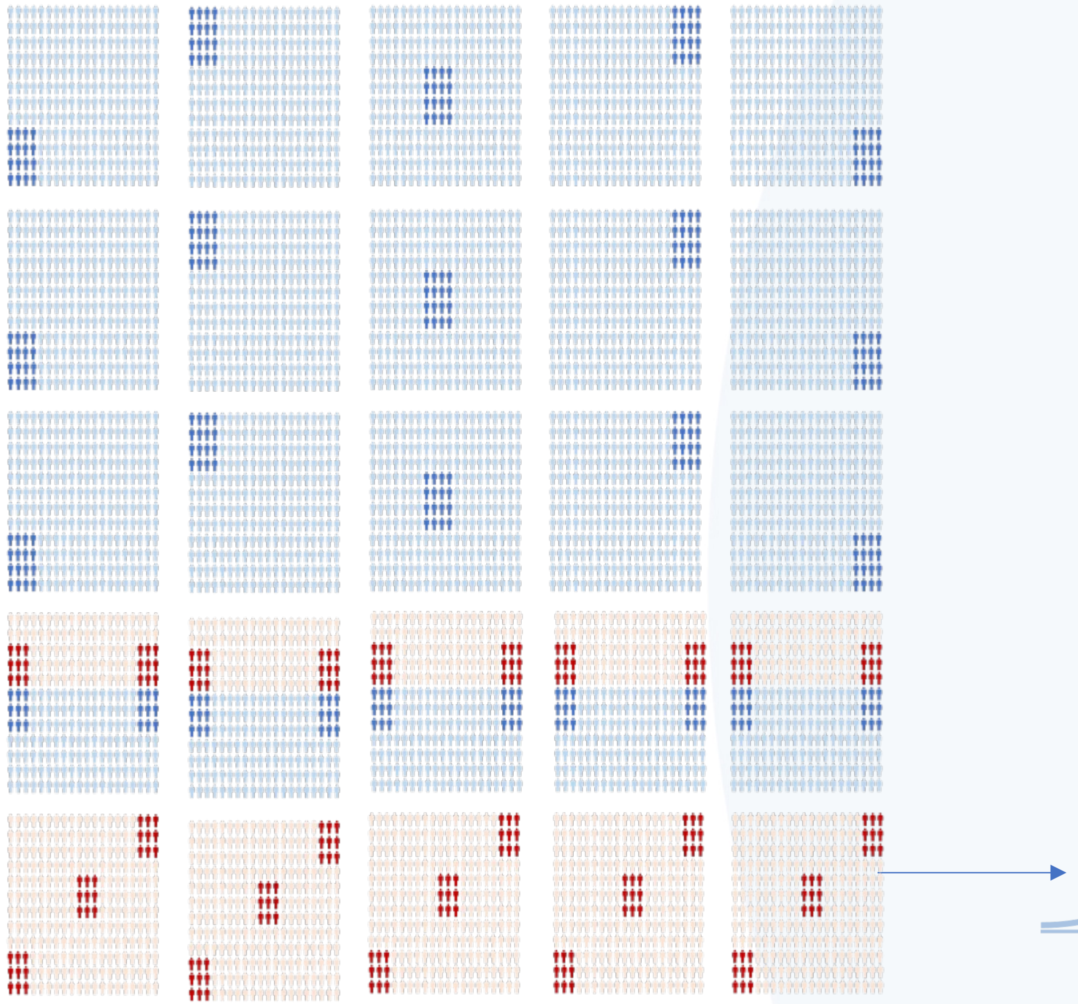
$\bar{x}$ $\hat{x}$

μ

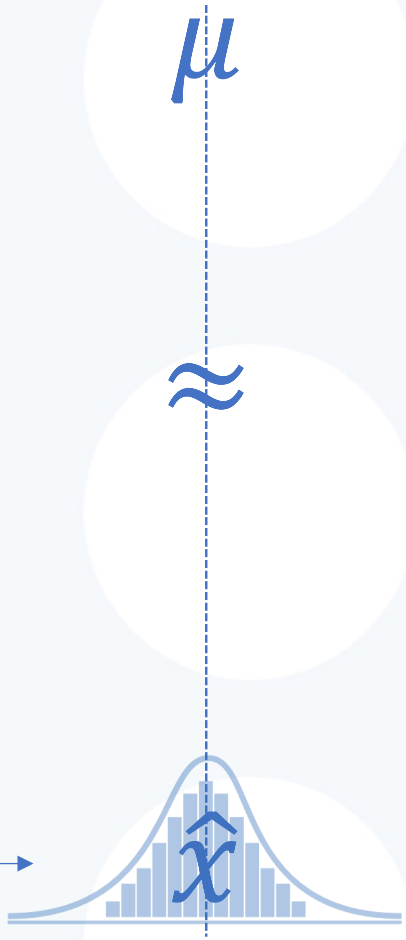
wt
With survey weights, which are the known probability of selection (in its inverse form), we can move the estimate closer to the parameter.

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

Sampling from a population



muestra compleja

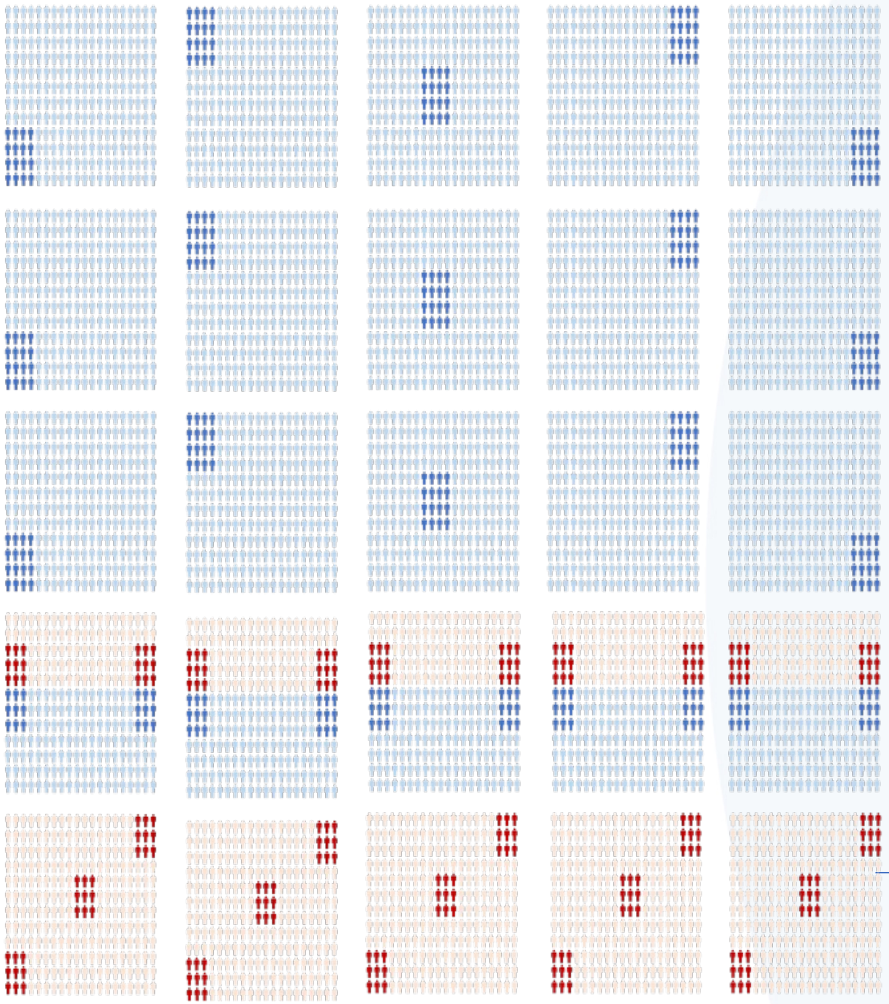


brr, tsl, jkn
Complex sample methods, allows us to to get standard errors and obtain measures of uncertainty.

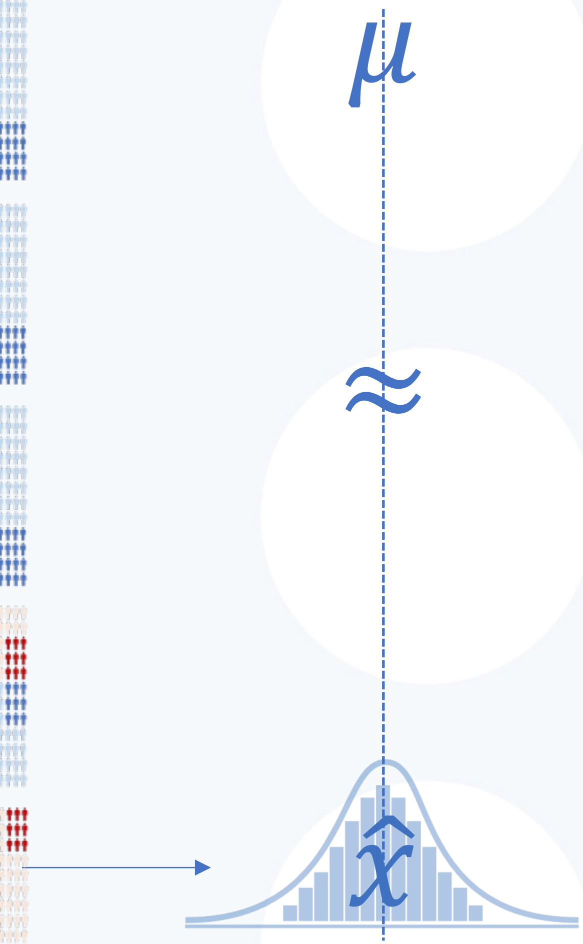
wt
[…] With survey weights we can correct the location of the estimate

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

Sampling from a population



muestra compleja



Using complex sample methods, we can to reasonable estimates of the population parameters.

brr, tsl, jkn
Complex sample methods, allows us to to get standard errors and obtain measures of uncertainty.

wt
Survey weights corrects the location of the estimates

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

Workshop

Estimates with and without design

What happens if we ignore the sample design

Descriptives with and without design

- When estimate results without the sampling design we cannot make inferences to the population. The generated estimates referred to the observed sample, but not to the population of interest.
- **Location estimates such as a means, medians, and proportions can be biased**. Survey weights re-locate the estimates.
- Results without sampling design **underestimate standard errors of our figures**. It will make you believe we have more precision.
- Complex sample methods get us corrected standard errors, and reasonable uncertainty measures of the generated results. **Using the sampling design, we can make inferences to the population of interest.**

Description	Population	Nominal	Delta	Interval Range (with design/without design)
ed. non-tertiary	.69 CI95%[.67, .71]	.74 CI95%[.73, .74]	0.05	4.00
ed. tertiary	.26 CI95%[.24, .28]	.21 CI95%[.21, .21]	0.05	6.67
missing	.05 CI95%[.04, .05]	.05 CI95%[.05, .06]	0.00	1.00

Note: descriptives for students with at least one parent with tertiary education (ERCE 2019, Perú, sixth grade students).

Descriptives with and without design

- When estimate results without the sampling design we cannot make inferences to the population. The generated estimates referred to the observed sample, but not to the population of interest.
- **Location estimates such as a means, medians, and proportions can be biased**. Survey weights re-locate the estimates.
- Results without sampling design **underestimate standard errors of our figures**. It will make you believe we have more precision.
- Complex sample methods get us corrected standard errors, and reasonable uncertainty measures of the generated results. **Using the sampling design, we can make inferences to the population of interest.**

Description	Population	Nominal	Delta	Interval Range (with design/without design)
ed. non-tertiary	.69 CI95%[.67, .71]	.74 CI95%[.73, .74]	0.05	4.00
ed. tertiary	.26 CI95%[.24, .28]	.21 CI95%[.21, .21]	0.05	6.67
missing	.05 CI95%[.04, .05]	.05 CI95%[.05, .06]	0.00	1.00

Note: descriptives for students with at least one parent with tertiary education (ERCE 2019, Perú, sixth grade students).

Biased estimates
In this example, ignoring the sampling design leads us to underestimate and over-estimate the proportion of students with parents with less or more education, respectively.

Descriptives with and without design

- When estimate results without the sampling design we cannot make inferences to the population. The generated estimates referred to the observed sample, but not to the population of interest.
- **Location estimates such as a means, medians, and proportions can be biased**. Survey weights re-locate the estimates.
- Results without sampling design **underestimate standard errors of our figures**. It will make you believe we have more precision.
- Complex sample methods get us corrected standard errors, and reasonable uncertainty measures of the generated results. **Using the sampling design, we can make inferences to the population of interest.**

Description	Population	Nominal	Delta	Interval Range (with design/without design)
ed. non-tertiary	.69 CI95%[.67, .71]	.74 CI95%[.73, .74]	0.05	4.00
ed. tertiary	.26 CI95%[.24, .28]	.21 CI95%[.21, .21]	0.05	6.67
missing	.05 CI95%[.04, .05]	.05 CI95%[.05, .06]	0.00	1.00

Note: descriptives for students with at least one parent with tertiary education (ERCE 2019, Perú, sixth grade students).

Underestimate of errors
Interval range of our estimates are smaller in the nominal estimates (without design). It underestimation is of 4 and 6.67 times the Interval range of the population estimates.

Descriptives with and without design

- When estimate results without the sampling design we cannot make inferences to the population. The generated estimates referred to the observed sample, but not to the population of interest.
- **Location estimates such as a means, medians, and proportions can be biased**. Survey weights re-locate the estimates.
- Results without sampling design **underestimate standard errors of our figures**. It will make you believe we have more precision.
- Complex sample methods get us corrected standard errors, and reasonable uncertainty measures of the generated results. **Using the sampling design, we can make inferences to the population of interest.**

Description	Population	Nominal	Delta	Interval Range (with design/without design)
ed. non-tertiary	.69 CI95%[.67, .71]	.74 CI95%[.73, .74]	0.05	4.00
ed. tertiary	.26 CI95%[.24, .28]	.21 CI95%[.21, .21]	0.05	6.67
missing	.05 CI95%[.04, .05]	.05 CI95%[.05, .06]	0.00	1.00

Note: descriptives for students with at least one parent with tertiary education (ERCE 2019, Perú, sixth grade students).

Estimates using the sampling design
These are estimates expanded to the sampling frame using the sampling design information. It includes uncertainty measures and provides figures allowing inferences to the population.

Workshop

Estimates with plausible values

What happens if we use a single plausible value

Results with plausible values

- When we estimate results with a single plausible value, we are risk providing **bias estimates**, and inconsistent results (results are conditional to the chosen plausible value).
- If we average the plausible values before producing estimates (i.e., mean score), we can retrieve the expected point estimate. However, **we will underestimate errors.**
- If we estimate a relationship with a covariate, we will get **inconsistent** results (the obtained results are conditional to the chosen plausible value, and some can result significant, and some others might not (see Rutkowski et al., 2010).
- In general, plausible values were generated to retrieve population estimates. **We need to estimate results for each plausible value, and then combined these estimations** to get the population estimate.

	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
Mean	759.31	3.64	752.1	766.5	14.4
Combined	759.31	3.88	751.7	766.9	15.2

Note: E = estimate, SE = standard error, LL = lower limit of 95% confidence interval, UL = upper limit of 95% confidence interval, CI range = Interval distance between limits. Population mean estimates with sampling design (ERCE 2019, Perú, sixth grade students).

Results with plausible values

- When we estimate results with a single plausible value, we are risk providing **bias estimates**, and inconsistent results (results are conditional to the chosen plausible value).
- If we average the plausible values before producing estimates (i.e., mean score), we can retrieve the expected point estimate. However, **we will underestimate errors.**
- If we estimate a relationship with a covariate, we will get **inconsistent** results (the obtained results are conditional to the chosen plausible value, and some can result significant, and some others might not (see Rutkowski et al., 2010).
- In general, plausible values were generated to retrieve population estimates. **We need to estimate results for each plausible value, and then combined these estimations** to get the population estimate.

	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
Mean	759.31	3.64	752.1	766.5	14.4
Combined	759.31	3.88	751.7	766.9	15.2

Note: E = estimate, SE = standard error, LL = lower limit of 95% confidence interval, UL = upper limit of 95% confidence interval, CI range = Interval distance between limits. Population mean estimates with sampling design (ERCE 2019, Perú, sixth grade students).

Plausible values
Each of the plausible values are imputations. These values are different between each other, and we need to use their combined information to make population inferences.

Results with plausible values

- When we estimate results with a single plausible value, we are risk providing **bias estimates**, and inconsistent results (results are conditional to the chosen plausible value).
- If we average the plausible values before producing estimates (i.e., mean score), we can retrieve the expected point estimate. However, **we will underestimate errors.**
- If we estimate a relationship with a covariate, we will get **inconsistent** results (the obtained results are conditional to the chosen plausible value, and some can result significant, and some others might not (see Rutkowski et al., 2010).
- In general, plausible values were generated to retrieve population estimates. **We need to estimate results for each plausible value, and then combined these estimations** to get the population estimate.

	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
Mean	759.31	3.64	752.1	766.5	14.4
Combined	759.31	3.88	751.7	766.9	15.2

Note: E = estimate, SE = standard error, LL = lower limit of 95% confidence interval, UL = upper limit of 95% confidence interval, CI range = Interval distance between limits. Population mean estimates with sampling design (ERCE 2019, Perú, sixth grade students).

Underestimation of errors
Using the average of plausible values leads us to underestimation of standard errors.

Results with plausible values

- When we estimate results with a single plausible value, we are risk providing **bias estimates**, and inconsistent results (results are conditional to the chosen plausible value).
- If we average the plausible values before producing estimates (i.e., mean score), we can retrieve the expected point estimate. However, **we will underestimate errors.**
- If we estimate a relationship with a covariate, we will get **inconsistent** results (the obtained results are conditional to the chosen plausible value, and some can result significant, and some others might not (see Rutkowski et al., 2010).
- In general, plausible values were generated to retrieve population estimates. **We need to estimate results for each plausible value, and then combined these estimations** to get the population estimate.

	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
Mean	759.31	3.64	752.1	766.5	14.4
Combined	759.31	3.88	751.7	766.9	15.2

Note: E = estimate, SE = standard error, LL = lower limit of 95% confidence interval, UL = upper limit of 95% confidence interval, CI range = Interval distance between limits. Population mean estimates with sampling design (ERCE 2019, Perú, sixth grade students).

Estimates with plausible values
To get population estimates one must generate estimates for each plausible values and combined the obtained estimates following Rubin-Schafer rules (Schafer, 1997). Generally, this combination is implemented in statistical software design for these purposes.

Muchas gracias!

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

Centro de Medición MIDE UC
Pontificia Universidad Católica de Chile

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.