Metodologías de estimación de varianza con pesos replicados: una mirada desde el usuario externo

División de Estadísticas
CEPAL

9 de septiembre de 2021

RStudio Source Editor

data2

Filter

	id_hogar	id_pers	parentco	persindo	pers	miembro	_feh	_fep	_upm	_estrato	areageo	areageo2	metrop	uf	edad	sexo	edadj	sexoj	ncony
1	1	1	1	1	1	Miembro del hogar	270	270	1	110001	20	1	NA	11	23	Hombre	23	Jefe hombre	0
2	2	1	1	1	1	Miembro del hogar	270	270	1	110001	20	1	NA	11	23	Mujer	23	Jefa mujer	0
3	3	1	1	6	6	Miembro del hogar	270	270	1	110001	20	1	NA	11	35	Mujer	35	Jefa mujer	1
4	3	2	2	6	6	Miembro del hogar	270	270	1	110001	20	1	NA	11	34	Hombre	35	Jefa mujer	1
5	3	3	3	6	6	Miembro del hogar	270	270	1	110001	20	1	NA	11	11	Mujer	35	Jefa mujer	1
6	3	4	3	6	6	Miembro del hogar	270	270	1	110001	20	1	NA	11	7	Mujer	35	Jefa mujer	1
7	3	5	3	6	6	Miembro del hogar	270	271	1	110001	20	1	NA	11	4	Mujer	35	Jefa mujer	1
8	3	6	5	6	6	Miembro del hogar	270	270	1	110001	20	1	NA	11	18	Mujer	35	Jefa mujer	1
9	4	1	1	2	2	Miembro del hogar	271	271	1	110001	20	1	NA	11	46	Hombre	46	Jefe hombre	0
10	4	2	4	2	2	Miembro del hogar	271	270	1	110001	20	1	NA	11	81	Mujer	46	Jefe hombre	0
11	5	1	1	1	1	Miembro del hogar	270	270	1	110001	20	1	NA	11	71	Mujer	71	Jefa mujer	0
12	6	1	1	2	2	Miembro del hogar	270	270	1	110001	20	1	NA	11	47	Mujer	47	Jefa mujer	0
13	6	2	3	2	2	Miembro del hogar	270	271	1	110001	20	1	NA	11	24	Hombre	47	Jefa mujer	0
14	7	1	1	3	3	Miembro del hogar	270	270	1	110001	20	1	NA	11	28	Mujer	28	Jefa mujer	1
15	7	2	2	3	3	Miembro del hogar	270	270	1	110001	20	1	NA	11	50	Hombre	28	Jefa mujer	1
16	7	3	3	3	3	Miembro del hogar	270	270	1	110001	20	1	NA	11	1	Hombre	28	Jefa mujer	1
17	8	1	1	5	5	Miembro del hogar	271	271	1	110001	20	1	NA	11	34	Mujer	34	Jefa mujer	1
18	8	2	2	5	5	Miembro del hogar	271	270	1	110001	20	1	NA	11	35	Hombre	34	Jefa mujer	1
19	8	3	3	5	5	Miembro del hogar	271	270	1	110001	20	1	NA	11	16	Hombre	34	Jefa mujer	1
20	8	4	3	5	5	Miembro del hogar	271	270	1	110001	20	1	NA	11	11	Mujer	34	Jefa mujer	1
21	8	5	3	5	5	Miembro del hogar	271	270	1	110001	20	1	NA	11	3	Mujer	34	Jefa mujer	1
22	9	1	1	3	3	Miembro del hogar	270	270	1	110001	20	1	NA	11	57	Hombre	57	Jefe hombre	1
23	9	2	2	3	3	Miembro del hogar	270	270	1	110001	20	1	NA	11	51	Mujer	57	Jefe hombre	1
24	9	3	3	3	3	Miembro del hogar	270	270	1	110001	20	1	NA	11	16	Hombre	57	Jefe hombre	1
25	10	1	1	1	1	Miembro del hogar	270	270	1	110001	20	1	NA	11	60	Mujer	60	Jefa mujer	0
26	11	1	1	2	2	Miembro del hogar	270	270	1	110001	20	1	NA	11	50	Mujer	50	Jefa mujer	0
27	11	2	3	2	2	Miembro del hogar	270	271	1	110001	20	1	NA	11	30	Mujer	50	Jefa mujer	0
28	12	1	1	3	3	Miembro del hogar	270	270	1	110001	20	1	NA	11	26	Hombre	26	Jefe hombre	1
29	12	2	2	3	3	Miembro del hogar	270	270	1	110001	20	1	NA	11	20	Mujer	26	Jefe hombre	1

Estrategia de muestreo

UNITED NATIONS

ECLAC

Conceptos básicos

Después de que la muestra fue seleccionada y luego de realizar el proceso de medición, es necesario realizar la estimación de los parámetros junto con la estimación de sus errores estándar, definido como la raíz cuadrada de la varianza.

UNITED NATIONS

ECLAC

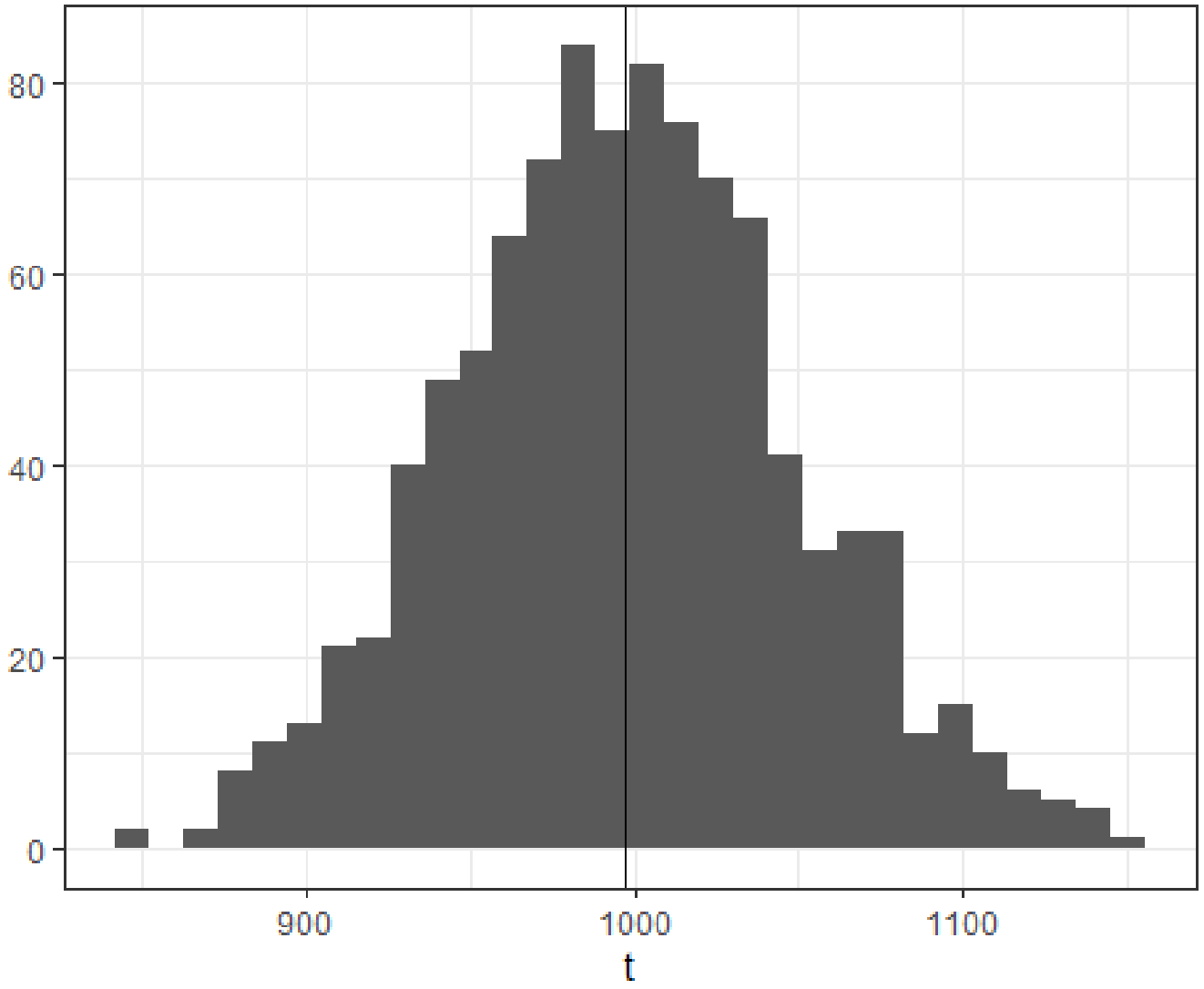
Estimación del error estándar

- La forma de estimar el error estándar depende de:
  - La complejidad del diseño de muestreo: estratificación, selección proporcional al tamaño, múltiples etapas.
  - La complejidad del estimador: ajuste de pesos por ausencia de respuesta, calibración, razón de totales, medias, percentiles, coeficientes de regresión.

UNITED NATIONS

ECLAC

Distribución del estimador



Estimación del error estándar

- Existen tres alternativas para calcular el error estándar de las estimaciones en una encuesta:
  - Fórmulas exactas
  - Linealización
  - Réplicas

UNITED NATIONS

ECLAC

Fórmulas exactas

UNITED NATIONS

ECLAC

Fórmulas

Siempre se pueden encontrar las fórmulas exactas que aplican a cada diseño de muestreo y a cada estimador usado.

- Para diseños simples es posible implementarlas para calcular la estimación de los errores.
- En diseños multietápicos y para estimadores simples se pueden tornar extremadamente complicadas.
- En diseños multietápicos y para estimadores complejos simplemente no son una opción plausible.

UNITED NATIONS

ECLAC

En general

Para cualquier diseño de muestreo sin reemplazo, la fórmula exacta para calcular una varianza de un total está dada por:

$$Var(\hat{t}_y) = \sum_U \sum_U \Delta_{kl} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l}$$

Evidentemente, por razones computacionales es imposible hacer esta estimación para todos los totales de las variables de interés en las encuestas.

UNITED NATIONS

ECLAC

La técnica del último conglomerado

Si el diseño de la encuesta es complejo (estratificado por regiones h, con tres etapas de selección dentro de cada estrato), entonces al utilizar la técnica del último conglomerado, el estimador de la varianza estaría dado por

$$\hat{V}(\hat{t}_y) = \sum_h (1 - f_h) \frac{n_h}{n_h - 1} \sum_{i \in s_h} \left(\hat{t}_{y_i} - \bar{\hat{t}}_{y_h}\right)^2$$

UNITED NATIONS

ECLAC

La técnica del último conglomerado

- Esta última expresión requiere menos conocimiento acerca del diseño de muestreo y de las fracciones de muestreo de las UPM.
- Usualmente las bases de datos públicas solo contienen información de la identificación de la UPM, el estrato y el peso final de muestreo.
- Con estos insumos, los investigadores tienden a considerar la estimación de la varianza suponiendo que el muestreo de las UPM se hizo con reemplazo.

UNITED NATIONS

ECLAC

Anonimización de la información

UNITED NATIONS

ECLAC

Secreto estadístico

La correcta estimación del error de muestreo y la determinación de la precisión de los estimadores requiere de variables sobre el diseño muestral muchas veces no disponible para el público usuario de las encuestas:

- Por tratarse de información que puede afectar el secreto estadístico.
- UPM y estratos pueden revelar información confidencial.
- Los ID de las UPM se construyen muchas veces con base en códigos nacionales geográficos

UNITED NATIONS

ECLAC

Secreto estadístico y estimación de la varianza

Situación difícil para las ONE:

- Proveer al usuario de herramientas para realizar una estimación correcta.
- Al mismo tiempo mantener la información anonimizada.

Pero además:

- El nivel de especialización para estimar errores estándar correctamente es muy alto y específico del personal de las ONE.

UNITED NATIONS

ECLAC

La posición del usuario

En general, el usuario regular de las bases de datos de las encuestas de hogares:

- ***No sabe que no sabe:*** el nivel de especialización para el cómputo de los errores estándar es muy alto: calibración, razones, medianas, ajustes, no respuesta, etc.
- ***Y aunque supiera:*** sin UPM ni estratos en las bases no es posible generar ninguna aproximación al error estándar.
- ***Por lo tanto:*** el procesamiento de datos va continuar sin una buena estimación del error estándar y las conclusiones tendrán una alta probabilidad de ser erróneas.

UNITED NATIONS

ECLAC

Réplicas

UNITED NATIONS

ECLAC

Conceptos básicos

El enfoque de esta técnica computacional se basa en:

- Dividir la toda la muestra en pequeños subconjuntos (réplicas).
- Repetir los mismos procesos de ajuste de ponderadores en cada réplica.
- Hacer la estimación en cada subgrupo.
- La varianza del estimador se calcula de manera simple como la desviación de todas las estimaciones en cada réplica.

UNITED NATIONS

ECLAC

Ventajas

- Al proveer los pesos de las réplicas, los investigadores pueden estimar el error de muestreo de forma automatizada y sin necesidad de intrincadas fórmulas matemáticas.
- Estos métodos ha demostrado ser eficientes y precisos para la mayoría de parámetros de interés.
- No se requiere conocer el diseño de muestreo utilizado en la encuesta.
- No se requiere del conocimiento de los estratos ni de las UPM y esto protege la anonimidad de los respondientes.

UNITED NATIONS

ECLAC

Ejemplos de encuestas y estudios por muestreo

- American Community Survey
- American Housing Survey
- Current Population Survey
- PISA, TIMMS, PIRLS, ICCS

UNITED NATIONS

ECLAC

Jackknife

Conceptos básicos

- Este método provee estimaciones eficientes para estimadores lineales y no lineales (a excepción de los percentiles).
- En su forma más básica, los pesos replicados se crean al retirar una UPM del análisis.
- Se tendrán tantas réplicas como UPM existan en la muestra.
- Todas las unidades dentro de las UPM se retiran del análisis.

UNITED NATIONS

ECLAC

Pesos en Jackkinife

Los pesos de la unidad k que pertenece a la UPM U_i, en el estrato U_h están dados por la siguiente expresión:

$$d_{hk}^{i} = \begin{cases} 0, \text{ si } U_i \in U_h \text{ y } k \in U_i \\ d_k, \text{ si } k \notin U_h \\ \dfrac{m_h}{m_h - 1} d_k, \text{ si } U_i \in U_h \text{ y } k \notin U_i \end{cases}$$

En donde m_h es el número de UPM en el estrato U_h.

Esquema de las réplicas Jackknife

R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
0.00	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11
1.11	0.00	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11
1.11	1.11	0.00	1.11	1.11	1.11	1.11	1.11	1.11	1.11
1.11	1.11	1.11	0.00	1.11	1.11	1.11	1.11	1.11	1.11
1.11	1.11	1.11	1.11	0.00	1.11	1.11	1.11	1.11	1.11
1.11	1.11	1.11	1.11	1.11	0.00	1.11	1.11	1.11	1.11
1.11	1.11	1.11	1.11	1.11	1.11	0.00	1.11	1.11	1.11
1.11	1.11	1.11	1.11	1.11	1.11	1.11	0.00	1.11	1.11
1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	0.00	1.11
1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	0.00

Fuente: PISA Data Analysis Manual

Consideraciones

- Una vez se tengan las réplicas de los pesos de muestreo se tendrán $m + 1$ columnas con pesos.
- En una encuesta con mil UPM, habrían mil réplicas de los pesos de muestreo
- Es posible agrupar las UPM para reducir el número de réplicas y hacer el proceso de cálculo de varianza más eficientemente.

UNITED NATIONS

ECLAC

Esquema del Jackknife agrupado

R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
2	1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	1
1	0	1	1	1	1	1	1	1	1
1	2	1	1	1	1	1	1	1	1
1	1	2	1	1	1	1	1	1	1
1	1	0	1	1	1	1	1	1	1
1	1	1	0	1	1	1	1	1	1
1	1	1	2	1	1	1	1	1	1
1	1	1	1	2	1	1	1	1	1
1	1	1	1	0	1	1	1	1	1
1	1	1	1	1	2	1	1	1	1
1	1	1	1	1	0	1	1	1	1
1	1	1	1	1	1	0	1	1	1
1	1	1	1	1	1	2	1	1	1
1	1	1	1	1	1	1	0	1	1
1	1	1	1	1	1	1	2	1	1
1	1	1	1	1	1	1	1	0	1
1	1	1	1	1	1	1	1	2	1
1	1	1	1	1	1	1	1	1	2
1	1	1	1	1	1	1	1	1	0

Fuente: PISA Data Analysis Manual

Réplicas repetidas balanceadas (BRR)

Conceptos básicos

- Esta técnica se desarrolló para diseños en donde dos UPM son seleccionadas por estrato.
- Funciona consistentemente para la estimación de parámetros lineales y no lineales (incluidos los percentiles).
- Este diseño asegura máxima dispersión de las UPM a través de las regiones geográficas (estratos).

UNITED NATIONS

ECLAC

Conceptos básicos

- Si $m_h = 2$, es posible hacer 2^H posibles réplicas al seleccionar aleatoriamente una UPM en cada estrato.
- Es posible lograr la misma eficiencia reduciendo el número de réplicas utilizando un enfoque ortogonal con matrices de Hadamard.
- El número de réplicas ortogonales será igual al menor múltiplo de 4 mayor o igual al número de estratos.

UNITED NATIONS

ECLAC

Matrices de Hadamard

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$

Matriz de Hadamard con 8 réplicas

UNITED NATIONS

ECLAC

Matrices de Hadamard

- +1 implica que la primera UPM se mantiene como parte de la réplica y la segunda UPM es retirada de la réplica.
- -1 implica que la segunda UPM se mantiene como parte de la réplica y la primera UPM es retirada de la réplica.
- En cada réplica se retira una UPM por estrato.
- Las UPM que se mantienen en cada réplica se llaman half-samples

UNITED NATIONS

ECLAC

Pesos en BRR

El peso de los individuos en la UPM que se mantiene se multiplica por un factor de 2. Por tanto, se tiene que

$$d_k = \begin{cases} 0, & \text{si } k \text{ pertence a la UPM que fue retirada} \\ 2d_k, & \text{en otro caso.} \end{cases}$$

UNITED NATIONS

ECLAC

Esquema del BRR

R1	R2	R3	R4	R5	R6	R7	R8	R9	R 10	R 11	R 12
2	0	0	2	0	0	0	2	2	2	0	2
0	2	2	0	2	2	2	0	0	0	2	0
2	2	0	0	2	0	0	0	2	2	2	0
0	0	2	2	0	2	2	2	0	0	0	2
2	0	2	0	0	2	0	0	0	2	2	2
0	2	0	2	2	0	2	2	2	0	0	0
2	2	0	2	0	0	2	0	0	0	2	2
0	0	2	0	2	2	0	2	2	2	0	0
2	2	2	0	2	0	0	2	0	0	0	2
0	0	0	2	0	2	2	0	2	2	2	0
2	2	2	2	0	2	0	0	2	0	0	0
0	0	0	0	2	0	2	2	0	2	2	2
2	0	2	2	2	0	2	0	0	2	0	0
0	2	0	0	0	2	0	2	2	0	2	2
2	0	0	2	2	2	0	2	0	0	2	0
0	2	2	0	0	0	2	0	2	2	0	2
2	0	0	0	2	2	2	0	2	0	0	2
0	2	2	2	0	0	0	2	0	2	2	0
2	2	0	0	0	2	2	2	0	2	0	0
0	0	2	2	2	0	0	0	2	0	2	2

BRR: 10 estratos y 12 réplicas

Fuente: PISA Data Analysis Manual

Comentarios

- Una desventaja del método BRR es que las unidades en dominios con muestra pequeña pueden estar ausentes en algunas combinaciones de réplicas por el diseño ortogonal.
- Lo anterior conlleva una pérdida de precisión en el cálculo del error estándar.
- Una solución a este problema es modificar los pesos en las réplicas.

UNITED NATIONS

ECLAC

Modificación de Fay

En esta propuesta se siguen los lineamientos basados en la matriz de Hadamard, aunque las UPM no son retiradas completamente, sino que su peso se modifica de la siguiente manera:

$$d_k^a = \begin{cases} \rho * d_k \\ \text{si } k \text{ no pertenece a la } *\text{half–sample}* \\ (2-\rho)d_k \text{en otro caso} \end{cases}$$

En donde $0 < \rho < 1$. Algunos estudios por simulación han mostrado una buena eficiencia para valores de ρ iguales a 0.3, 0.5 o 0.7.

Varianza BRR

Bajo la metodología BRR con el ajuste de Fay, el estimador de la varianza toma la siguiente forma:

$$Var(\hat{\theta}) = \frac{1}{A(1-\rho)^2} \sum_{a=1}^{A} (\hat{\theta}_a - \hat{\theta})^2$$

En donde $\hat{\theta}_a$ es el estimador del parámetro de interés en la réplica a.

Bootstrap

Conceptos básicos

- El Bootstrap es el método basado en réplicas más versátil para el cálculo de errores estándar.
- Es muy eficiente en la estimación de parámetros lineales y no lineales, a diferencia del Jackknife que no es eficiente en la estimación de percentiles.
- Funciona también para tamaños de muestra pequeños (mayores a dos), a diferencia del método BRR que requiere una muestra de mínimo dos UPM por estrato.

UNITED NATIONS

ECLAC

Conceptos básicos

- En cada estrato se selecciona una submuestra mediante un diseño con reemplazo.
- A diferencia de la técnica del último conglomerado en donde el reemplazamiento se asume aunque no sea realmente implementado.
- Es posible que una UPM sea seleccionada más de una vez en las submuestras Bootstrap.
- Este método requiere una gran cantidad de réplicas, usualmente mayor a 200.

UNITED NATIONS

ECLAC

Pesos en Bootstrap

Siendo s_{BS} la submuestra Bootstrap con reemplazo, el peso del individuo k perteneciente a la UPM i del estrato h sigue la siguiente expresión:

$$d_k = \begin{cases} 0, \text{ si la UPM } i \text{ no pertence a } s_{BS} \\ d_k \left[1 - \sqrt{\frac{m_h^*}{m_h - 1}} + \sqrt{\frac{m_h^*}{m_h - 1}} \frac{m_h}{m_h^*} m_{hi}^* \right], \text{ en otro caso} \end{cases}$$

En donde m_h es el número de UPM en la muestra original del estrato h, m_h^* es el número de UPM en la muestra Bootstrap y m_{hi}^* es el número de veces que la UPM i fue seleccionada en la muestra Bootstrap.

UNITED NATIONS

ECLAC

Pesos en Bootstrap clásico

En este caso se selecciona una muestra Bootstrap con $m_h^* = m_h - 1$, y los pesos toman la siguiente forma:

$$d_k = \begin{cases} 0, \text{ si la UPM } i \text{ no pertence a } s_{BS} \\ d_k \left[\dfrac{m_h^*}{m_h - 1} m_{hi}^* \right], \text{ en otro caso} \end{cases}$$

UNITED NATIONS

ECLAC

Varianza Bootstrap

Bajo la metodología Bootstrap, el estimador de la varianza toma la siguiente forma:

$$Var(\hat{\theta}) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_b - \hat{\theta})^2$$

En donde $\hat{\theta}_b$ es el estimador del parámetro de interés en la réplica b inducida por la muestra Bootstrap.

UNITED NATIONS

ECLAC

Es el turno de las ONE

UNITED NATIONS

ECLAC

Weight	Mean estimate	Weight	Mean estimate
Final weight	**49.33**		
Replicate 1	49.44	Replicate 41	49.17
Replicate 2	49.18	Replicate 42	49.66
Replicate 3	49.12	Replicate 43	49.18
Replicate 4	49.46	Replicate 44	49.04
Replicate 5	49.24	Replicate 45	49.42
Replicate 6	49.34	Replicate 46	49.72
Replicate 7	49.13	Replicate 47	49.48
Replicate 8	49.08	Replicate 48	49.14
Replicate 9	49.54	Replicate 49	49.57
Replicate 10	49.20	Replicate 50	49.36
Replicate 11	49.22	Replicate 51	48.78
Replicate 12	49.12	Replicate 52	49.53
Replicate 13	49.33	Replicate 53	49.27
Replicate 14	49.47	Replicate 54	49.23
Replicate 15	49.40	Replicate 55	49.62
Replicate 16	49.30	Replicate 56	48.96
Replicate 17	49.24	Replicate 57	49.54
Replicate 18	48.85	Replicate 58	49.14
Replicate 19	49.41	Replicate 59	49.27
Replicate 20	48.82	Replicate 60	49.42
Replicate 21	49.46	Replicate 61	49.56
Replicate 22	49.37	Replicate 62	49.75
Replicate 23	49.39	Replicate 63	48.98
Replicate 24	49.23	Replicate 64	49.00
Replicate 25	49.47	Replicate 65	49.35
Replicate 26	49.51	Replicate 66	49.27
Replicate 27	49.35	Replicate 67	49.44
Replicate 28	48.89	Replicate 68	49.08
Replicate 29	49.44	Replicate 69	49.09
Replicate 30	49.34	Replicate 70	49.15
Replicate 31	49.41	Replicate 71	49.29
Replicate 32	49.18	Replicate 72	49.29
Replicate 33	49.50	Replicate 73	49.08
Replicate 34	49.12	Replicate 74	49.25
Replicate 35	49.05	Replicate 75	48.93
Replicate 36	49.40	Replicate 76	49.45
Replicate 37	49.20	Replicate 77	49.13
Replicate 38	49.54	Replicate 78	49.45
Replicate 39	49.32	Replicate 79	49.14
Replicate 40	49.35	Replicate 80	49.27

Fuente: PISA Data Analysis Manual

UNITED NATIONS

ECLAC

Weight	Squared difference	Weight	Squared difference
Replicate 1	0.0140	Replicate 41	0.0239
Replicate 2	0.0228	Replicate 42	0.1090
Replicate 3	0.0421	Replicate 43	0.0203
Replicate 4	0.0189	Replicate 44	0.0818
Replicate 5	0.0075	Replicate 45	0.0082
Replicate 6	0.0002	Replicate 46	0.1514
Replicate 7	0.0387	Replicate 47	0.0231
Replicate 8	0.0583	Replicate 48	0.0349
Replicate 9	0.0472	Replicate 49	0.0590
Replicate 10	0.0167	Replicate 50	0.0014
Replicate 11	0.0124	Replicate 51	0.3003
Replicate 12	0.0441	Replicate 52	0.0431
Replicate 13	0.0000	Replicate 53	0.0032
Replicate 14	0.0205	Replicate 54	0.0086
Replicate 15	0.0048	Replicate 55	0.0868
Replicate 16	0.0009	Replicate 56	0.1317
Replicate 17	0.0074	Replicate 57	0.0438
Replicate 18	0.2264	Replicate 58	0.0354
Replicate 19	0.0077	Replicate 59	0.0034
Replicate 20	0.2604	Replicate 60	0.0081
Replicate 21	0.0182	Replicate 61	0.0563
Replicate 22	0.0016	Replicate 62	0.1761
Replicate 23	0.0041	Replicate 63	0.1173
Replicate 24	0.0093	Replicate 64	0.1035
Replicate 25	0.0199	Replicate 65	0.0008
Replicate 26	0.0344	Replicate 66	0.0030
Replicate 27	0.0007	Replicate 67	0.0139
Replicate 28	0.1919	Replicate 68	0.0618
Replicate 29	0.0139	Replicate 69	0.0557
Replicate 30	0.0001	Replicate 70	0.0324
Replicate 31	0.0071	Replicate 71	0.0016
Replicate 32	0.0215	Replicate 72	0.0011
Replicate 33	0.0302	Replicate 73	0.0603
Replicate 34	0.0411	Replicate 74	0.0052
Replicate 35	0.0778	Replicate 75	0.1575
Replicate 36	0.0052	Replicate 76	0.0157
Replicate 37	0.0150	Replicate 77	0.0378
Replicate 38	0.0445	Replicate 78	0.0155
Replicate 39	0.0000	Replicate 79	0.0354
Replicate 40	0.0004	Replicate 80	0.0031
		Sum of squared differences	**3.5195**

Fuente: PISA Data Analysis Manual

UNITED NATIONS

ECLAC

Algunas consideraciones

- En algunas ocasiones las UPM o los estratos se colapsan para formar grupos que permitan el uso de las técnicas (Jackknife, BRR o Fay).
- Lo anterior permite manejar un conjunto de réplicas de un tamaño adecuado, además de permitir la creación de pseudo estratos cuando solo se tiene una UPM en algunos estratos.
- Si se implementa adecuadamente, el agrupamiento producirá estimaciones precisas e insesgadas del error estándar.

UNITED NATIONS

ECLAC

Algunas consideraciones

- Es necesario reflejar el ajuste de los pesos en las réplicas, por esto es necesario trabajar con la muestra originalmente seleccionada.
- La muestra original contendrá unidades no elegibles y unidades que no respondieron.

UNITED NATIONS

ECLAC

Algunas consideraciones

- Los mismos ajustes que se hicieron a la muestra original se deben realizar en cada réplica.
- Si hubo calibración de pesos también debe ser incluida como un proceso en cada réplica.
- Hacer esto asegura que el error estándar inducido por estos métodos incluirá el incremento (o decremento) de la varianza inducida por estos ajustes a los pesos.

UNITED NATIONS

ECLAC

¡Gracias!

Por favor, no dude en ponerse en contacto con la División de Estadísticas de ECLAC si necesita ayuda en el diseño y análisis de encuestas de muestras, detección de sesgos y calibración.

Rolando Ocampo - rolando.ocampo@un.org

Xavier Mancero - xavier.mancero@un.org

Andrés Gutiérrez - andres.gutierrez@un.org

UNITED NATIONS

ECLAC