



NACIONES UNIDAS

CEPAL

Limitaciones de las encuestas de hogares en la medición de indicadores sociales

Taller regional sobre desagregación de estadísticas sociales mediante metodologías de estimación en áreas pequeñas

Andrés Gutiérrez

Unidad de Estadísticas Sociales

División de Estadísticas

La necesidad de las desagregaciones en las encuestas sociales



OBJETIVOS DE DESARROLLO SOSTENIBLE

1 FIN DE LA POBREZA

2 HAMBRE CERO

3 SALUD Y BIENESTAR

4 EDUCACIÓN DE CALIDAD

5 IGUALDAD DE GÉNERO

6 AGUA LIMPIA Y SANEAMIENTO

7 ENERGÍA ASEQUIBLE Y NO CONTAMINANTE

8 TRABAJO DECENTE Y CRECIMIENTO ECONÓMICO

9 INDUSTRIA, INNOVACIÓN E INFRAESTRUCTURA

10 REDUCCIÓN DE LAS DESIGUALDADES

11 CIUDADES Y COMUNIDADES SOSTENIBLES

12 PRODUCCIÓN Y CONSUMO RESPONSABLES

13 ACCIÓN POR EL CLIMA

14 VIDA SUBMARINA

15 VIDA DE ECOSISTEMAS TERRESTRES

16 PAZ, JUSTICIA E INSTITUCIONES SÓLIDAS

17 ALIANZAS PARA LOGRAR LOS OBJETIVOS

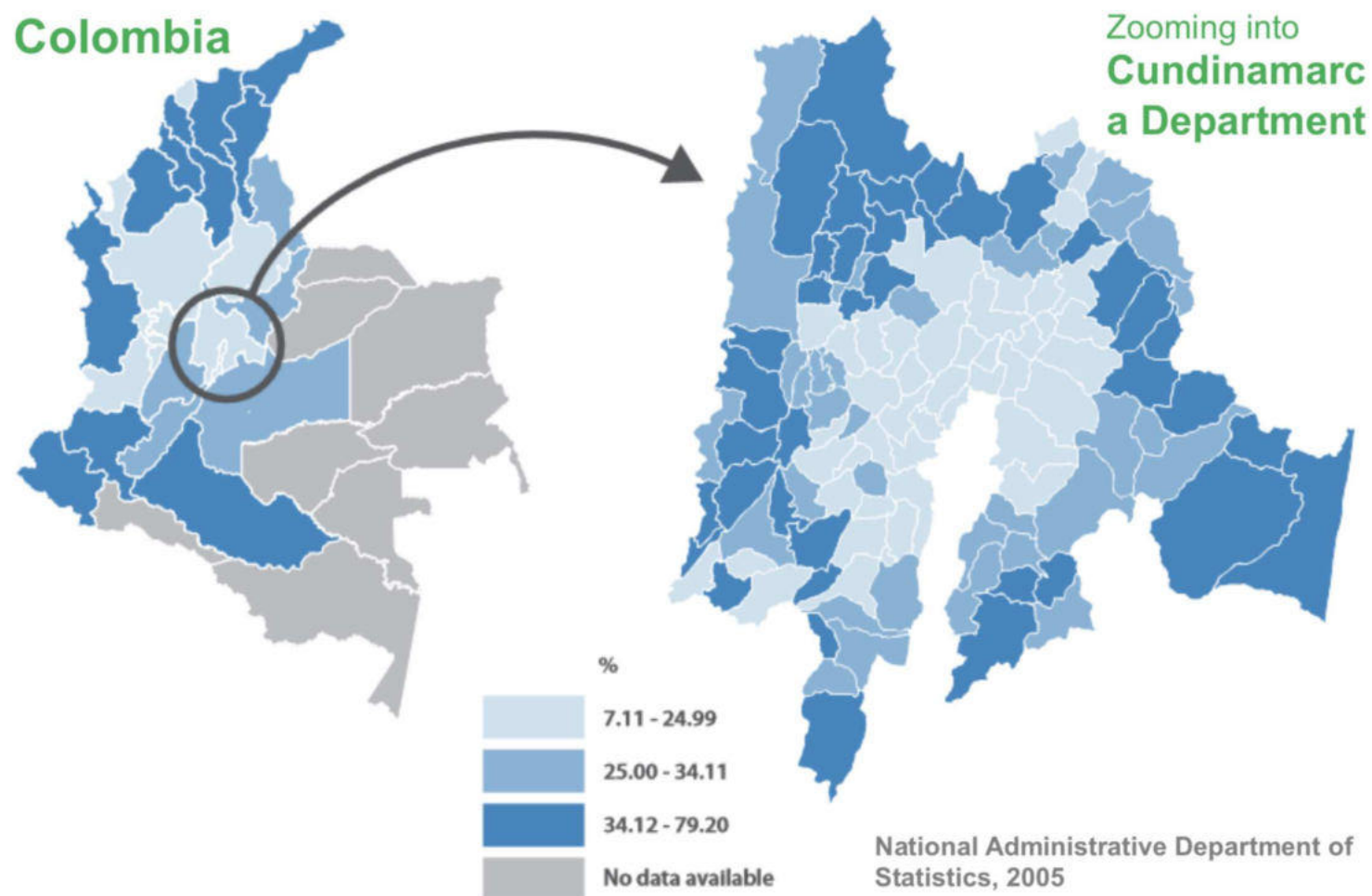
OBJETIVOS DE DESARROLLO SOSTENIBLE

ODS 17 - Meta 18 (*Alianzas para lograr los objetivos*)

*De aquí a 2020, mejorar el **apoyo** a la **creación de capacidad** prestado a los **países en desarrollo**, incluidos los países menos adelantados y los pequeños Estados insulares en desarrollo, para aumentar significativamente la disponibilidad de datos oportunos, fiables y de gran calidad **desglosados** por:*

- ingresos, sexo, edad, raza, origen étnico, estatus migratorio, discapacidad, ubicación geográfica y,*
- otras características pertinentes en los contextos nacionales.*

Share of households per « Basic Unmet Needs » index, Colombia



Desagregación de un indicador en áreas pequeñas en Colombia. Fuente: UNSD

Revisión de los estándares internacionales en desagregación de información

¿Qué es el coeficiente de variación?

El coeficiente de variación es una medida de error relativo a un estimador, se define como:



$$cve(\hat{\theta}) = \frac{se(\hat{\theta})}{\hat{\theta}}$$

Muchas veces se expresa como un porcentaje, aunque no está acotado a la derecha, y por eso es conveniente a la hora de hablar de la precisión de una estadística que viene de una encuesta.

Uso del coeficiente de variación

Sarndal et. al. (2003) expresa que *un estadístico puede expresar su opinión de que un valor del coeficiente de variación del 2% es **bueno**, considerando las restricciones de la encuesta, mientras que un valor del coeficiente de variación de 9% puede ser considerado **inaceptable**.*

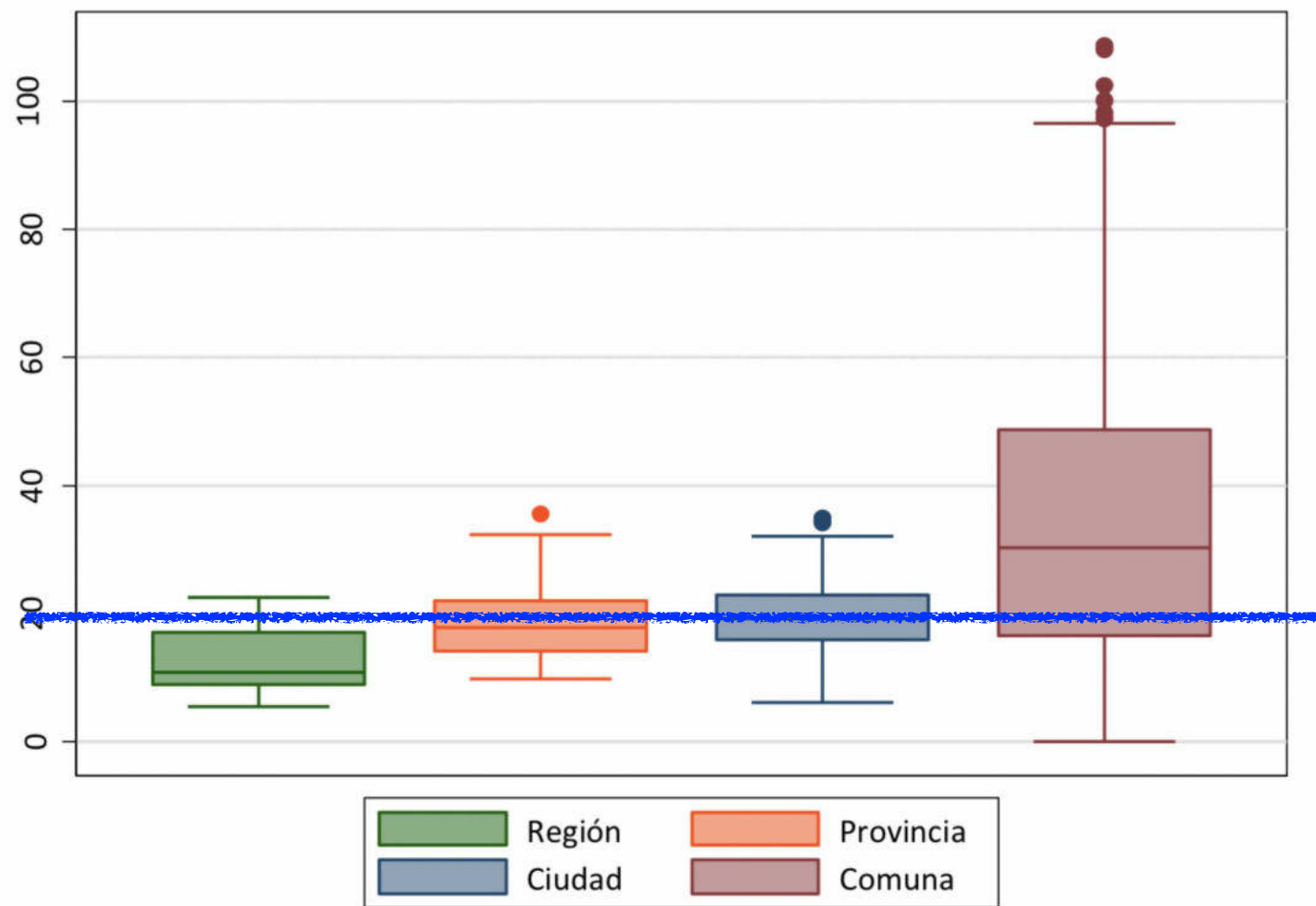
De esta forma, muchos institutos nacionales de estadística alrededor del mundo han considerado que las precisiones de las estadísticas resultantes de una encuesta estén supeditadas al comportamiento de su coeficiente de variación.

Interpretación	Semaforización	Viviendas / Hogares DGES / DGE GSPyJ
Buena		[0%, 15%)
Aceptable		[15%, 25%)
Con reserva		$\geq 25\%$

Alertas debidas al coeficiente de variación. Fuente: INEGI - México

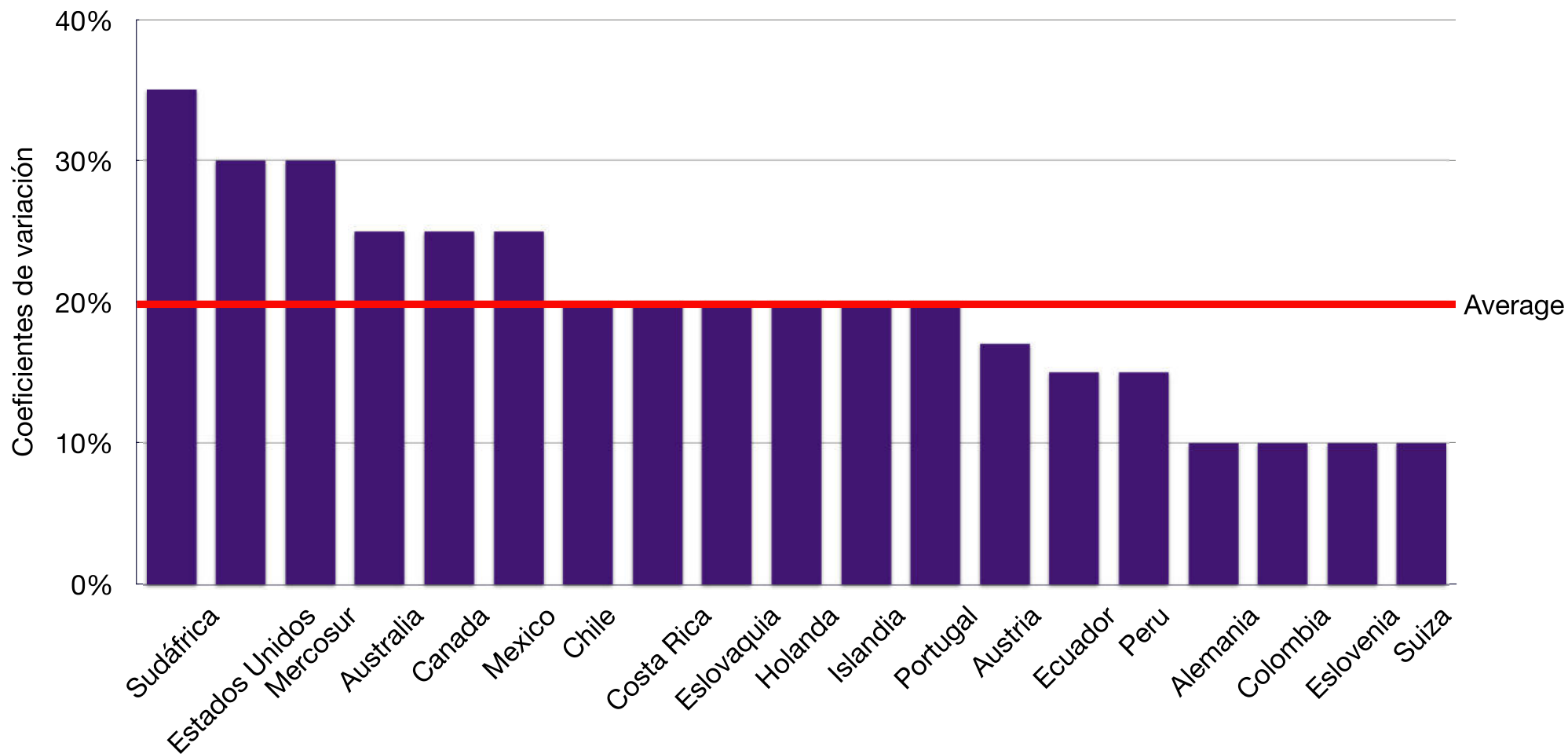
Coeficiente de variación (%)	Número de Observaciones	
	Bajo	Alto
[20 , 100]	Estimador no confiable	Estimador no confiable
[15 , 20)	Estimador no confiable	Descriptivo
[5 , 15)	Descriptivo	Estimador confiable
(0 , 5)	Estimador confiable	Estimador confiable

Alertas debidas al coeficiente de variación. Fuente: INE - Chile

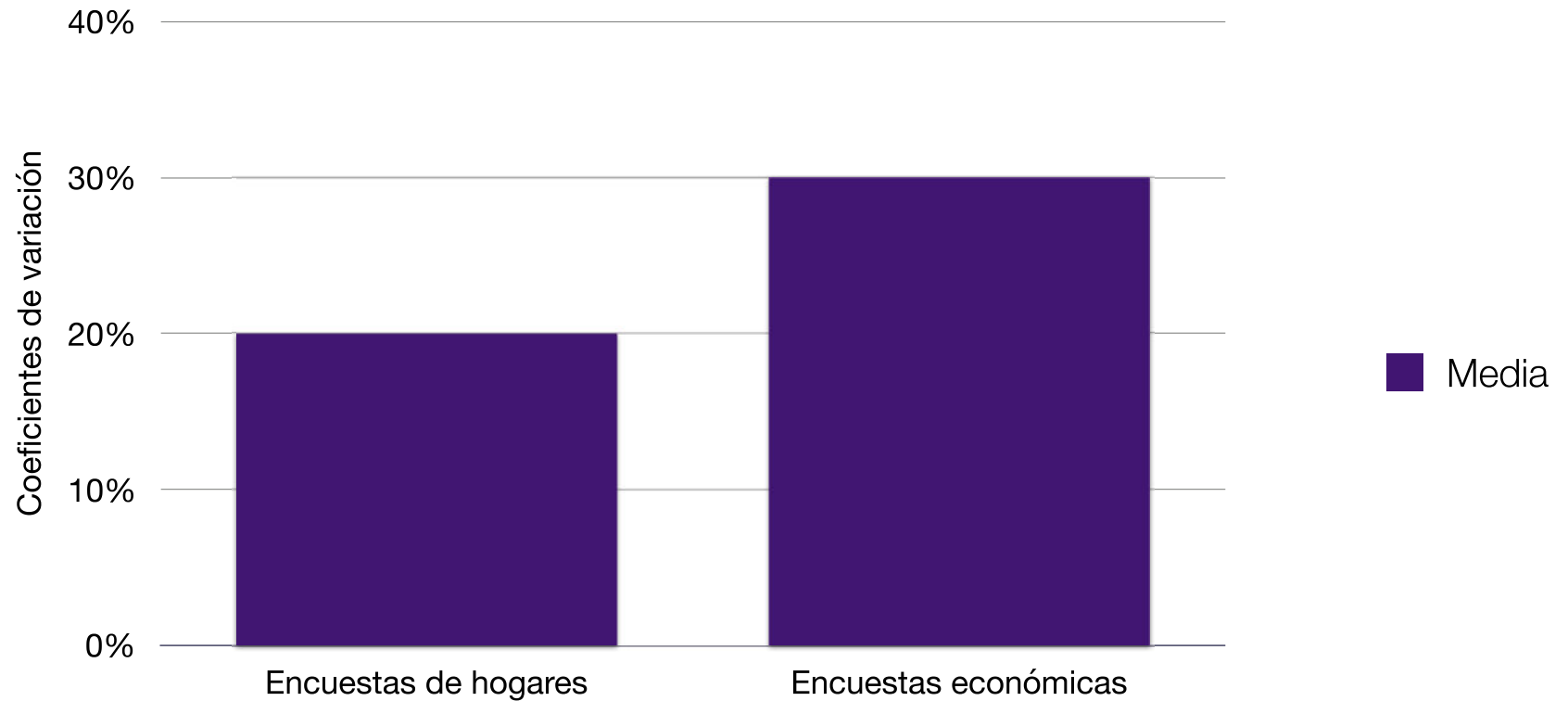


Coeficientes de variación en diferentes desagregaciones. Fuente: INE - Chile

Estándares de alerta en algunos países (encuestas de hogares)



Estándares de alerta en algunos países



**Coeficientes de variación para encuestas de hogares y empresas.
Fuente: INEGI (2018)**

Algunas alertas definidas en la publicación

Cuando se sobrepasa el umbral del coeficiente de variación aparecen algunas de las siguientes alertas:

- **No se publica**
- Usar con precaución.
- Los coeficientes de variación mayores a 20, significan que los ingresos estimados están sujetos a una alta variabilidad, lo que limita su alcance para fines analíticos.
- Las estimaciones requieren revisiones, no son precisas y se deben usar con precaución.
- Poco confiable, menos preciso. Se usa la letra M o () para que se use con cuidado.
- No cumple con los estándares de publicación.
- Con reserva, referencial, cuestionable.
- Valores muy aleatorios, estimación pobre.

**Otros elementos que intervienen en la
confianza de las estadísticas sociales**

Dominios de estudios y subpoblaciones de interés

Una encuesta se **planea** con el fin de generar información precisa y confiable en los dominios de estudio que se han predefinido. Sin embargo, existen subgrupos poblacionales que la encuesta no abordó en su diseño, y sobre los cuales se quisiera una mayor precisión.

- Incidencia de la pobreza desagregado por **departamento** o provincia (tamaño de muestra conocido y planificado).
- Tasa de desocupación desagregada por **sexo** (tamaño de muestra aleatorio, pero planificado).
- Tasa de asistencia neta estudiantil en primaria desagregada por **quintiles de ingreso** (tamaño de muestra aleatorio).

Precisión de los estimadores

Debido a que la encuesta es una investigación parcial sobre una población finita, es necesario saber que:

- A partir de una encuesta, no se **calculan** indicadores, sino que se **estiman** con ayuda de los datos de la encuesta.
- Es necesario calcular el grado de error que se comete al no poder realizar una investigación exhaustiva. Este error es conocido como el **error de muestreo**.
- La precisión de un estimador está supeditada al **intervalo de confianza**. Entre más angosto sea el intervalo, más precisión se genera y por ende se tiene un menor error de muestreo.

El intervalo de confianza en subpoblaciones

Si el parámetro de interés sobre el cual se busca realizar la inferencia es θ_d , y se ha definido una subpoblación de interés U_d , entonces un intervalo del 95% de confianza sobre esa subpoblación está dado por la siguiente expresión:

$$(\hat{\theta}_d - t_{0.975,gl} * se(\hat{\theta}_d), \hat{\theta}_d + t_{0.975,gl} * se(\hat{\theta}_d))$$

El uso del coeficiente de variación como indicador de la confiabilidad de las estadísticas provenientes de encuestas de hogares debería ser complementado con algunas otras medidas que permitan crear reglas de confiabilidad y precisión.

El intervalo de confianza

Nótese que la longitud de los intervalos de confianza induce la seguridad de que un estimador es preciso:

- La incidencia de la pobreza en el departamento del país se estimó en 5.2%, con un intervalo de confianza de (5.15%, 5.25%). ✓
- La tasa de desocupación en el país para los hombres se ubicó en en 7.5%, con un intervalo de confianza de (7.1%, 7.9%); mientras que para las mujeres se ubicó en 9.2%, con intervalo de confianza de (8.8%, 9.6%). ✓
- La tasa de asistencia neta estudiantil en primaria para el último quintil de ingreso se estimó en 85%, con un intervalo de confianza de (48.2%, 100.0%). ✗

A. El tamaño de muestra

- El tamaño de muestra afecta de manera indirecta la amplitud del intervalo de confianza, a través del error estándar que generalmente decrece a medida que el tamaño de muestra se hace más grande.
- Un tamaño de muestra adecuado garantiza la convergencia en distribución de los estimadores a la distribución teórica de donde se calculan los percentiles.
- Por ejemplo, es posible plantear que todas las estimaciones basadas en un tamaño de muestra menor a un umbral predefinido deberían ser suprimidas o marcadas como no confiables.

B. El tamaño de muestra efectivo

- En las encuestas de hogares, con diseños de muestreo complejos, **no existe** una sucesión de variables que sean independientes e idénticamente distribuidas.
- La muestra (y_1, \dots, y_n) no es un vector en el espacio n -dimensional, donde se asume que cada componente del vector puede variar por sí mismo.
- La dimensión final del vector (y_1, \dots, y_n) es mucho menor que n , puesto que existe una forma jerárquica en la selección de los hogares y a la interrelación de la variable de interés con las UPMs

B. El tamaño de muestra efectivo

El tamaño de muestra efectivo se define como sigue:

$$n_{efectivo} = \frac{n}{Deff}$$

En donde *Deff* es el efecto de diseño que depende de:

1. El número de encuestas promedio que se realizaron en cada UPM.
2. La correlación existente entre la variable de interés y las mismas UPMs.

Es posible considerar que si el tamaño de muestra efectivo no es mayor a umbral, entonces la cifra no debería ser considerada para publicación.

C. Grados de libertad

Son una medida de cuántas unidades independientes de información se tienen en la inferencia. Nótese que:

- En el caso extremo de realizar un censo en cada UPMs, sin importar el número de individuos que componen el conglomerado, el número de unidades independientes será únicamente el número de UPMs seleccionadas en la primera etapa de muestreo.
- En las encuestas de hogares, la variabilidad de la estimación es la contribución del conglomerado a la gran media más una contribución (considerada insignificante) de la segunda etapa de muestreo.

C. Grados de libertad

En las subpoblaciones los grados de libertad no se consideran fijos sino variables.

$$gl_{sub} = \sum_{h=1}^H v_h * (n_{Ih} - 1)$$

Note que v_h es una variable indicadora que toma el valor uno si el estrato h contiene uno o mas casos de la subpoblación de interés, n_{Ih} es el número de UPMs en el estrato. En el caso más general, los grados de libertad se reducen a la siguiente expresión:

$$gl = \#UPMs - \#Estratos$$

C. Grados de libertad

Por ejemplo, considere por ejemplo el percentil 0.975 para el cual los valores críticos de la distribución t varían con respecto a sus grados de libertad

- $t\text{-student}_{gl = 1} = 12.7$
- $t\text{-student}_{gl = 2} = 4.30$
- $t\text{-student}_{gl = 2} = 4.30$
- $t\text{-student}_{gl = 5} = 2.57$
- $t\text{-student}_{gl = 40} = 2.02$
- $t\text{-student}_{gl = \infty} = Z = 1.96$

Es posible considerar que si los grados de libertad inducidos por la subpoblación son menores a un umbral predefinido, la cifra debería ser suprimida.

Algunos ejemplos de las limitaciones en las encuestas

País grande: desocupación urbana

Quintiles de ingreso - Sexo

Quintil Urbano	sexo	n	Deff	n.eff	gl	Desocupación%	Li %	Ls %	cv %	Alerta
Quinto	Hombre	29158	1.2	24293	4275	1.4	1.3	1.6	5.7	
Quinto	Mujer	30848	1.3	24639	4215	1.8	1.7	2.0	4.9	
Cuarto	Hombre	29245	1.4	21654	4584	2.3	2.1	2.5	4.8	
Cuarto	Mujer	30094	1.3	22583	4574	2.7	2.4	2.9	4.3	
Tercero	Hombre	29359	1.4	21385	4413	3.2	3.0	3.5	4.1	
Tercero	Mujer	32121	1.5	21757	4542	4.0	3.7	4.3	3.6	
Segundo	Hombre	30419	1.4	21806	4053	6.3	5.9	6.7	3.1	
Segundo	Mujer	33619	1.4	23275	4169	7.6	7.2	8.0	2.6	
Primer	Hombre	30381	1.5	20245	3531	12.7	12.1	13.3	2.4	
Primer	Mujer	35233	1.5	23360	3700	12.5	12.0	13.0	2.1	

País pequeño: desocupación urbana

Quintiles de ingreso - Sexo

Quintil Urbano	sexo	n	Deff	n.eff	gl	Desocupación%	Li %	Ls %	cv %	Alerta
Quinto	Mujer	2055	1.2	1757	309	1.0	0.4	1.6	30.6	*
Quinto	Hombre	1969	1.1	1738	335	1.1	0.5	1.7	26.3	*
Cuarto	Hombre	2245	1.2	1807	347	2.2	1.4	3.0	19.3	
Cuarto	Mujer	2301	1.6	1466	357	4.1	2.7	5.5	17.5	
Tercero	Mujer	2421	1.5	1646	336	6.1	4.3	7.9	15.1	
Segundo	Hombre	2280	1.4	1654	295	5.9	4.3	7.5	13.8	
Tercero	Hombre	2351	1.2	2025	331	4.6	3.4	5.8	13.3	
Segundo	Mujer	2541	1.6	1547	310	10.8	8.0	13.6	13.1	
Primero	Mujer	2862	2.0	1466	266	20.0	15.4	24.6	11.8	
Primero	Hombre	2562	1.6	1610	263	11.9	9.4	14.5	10.9	

País grande: desocupación rural

Quintiles de ingreso - Sexo

Quintil Rural	sexo	n	Deff	n.eff	gl	Desocupación%	Li %	Ls %	cv %	Alerta
Quinto	Hombre	5615	1.7	3279	282	0.7	0.4	1.0	22.1	*
Quinto	Mujer	4637	1.6	2812	268	1.0	0.6	1.3	20.2	*
Cuarto	Hombre	5506	1.3	4197	299	1.1	0.8	1.5	15.5	
Segundo	Hombre	5328	1.3	4040	238	2.1	1.6	2.7	13.3	
Cuarto	Mujer	4892	0.9	5176	291	1.4	1.1	1.8	12.5	
Tercero	Hombre	5436	1.1	4946	284	2.1	1.7	2.6	11.2	
Primer	Mujer	5100	1.3	4039	196	3.3	2.6	4.0	11.2	
Tercero	Mujer	5145	1.3	4051	277	2.9	2.3	3.5	10.6	
Primer	Hombre	5301	1.2	4292	193	4.2	3.4	5.0	9.7	
Segundo	Mujer	5190	1.3	4101	238	3.9	3.2	4.7	9.6	

País pequeño: desocupación rural

Quintiles de ingreso - Sexo

Quintil Rural	sexo	n	Deff	n.eff	gl	Desocupación%	Li %	Ls %	cv %	Alerta
Primer	Mujer	1788	0.6	2754	140	0.8	0.1	1.5	44.5	*
Cuarto	Hombre	2112	1.7	1223	178	1.8	0.8	2.7	26.7	*
Segundo	Hombre	2281	1.8	1236	156	2.6	1.3	3.9	25.7	*
Primer	Hombre	1704	1.3	1324	137	2.8	1.4	4.2	25.5	*
Quinto	Mujer	1780	1.3	1391	166	2.6	1.4	3.9	23.7	*
Segundo	Mujer	2195	1.4	1579	158	5.3	2.9	7.6	22.7	*
Quinto	Hombre	2127	1.4	1553	171	1.8	1.0	2.6	22.1	*
Tercero	Hombre	2180	2.0	1068	169	3.7	2.1	5.2	21.4	*
Tercero	Mujer	2023	1.4	1411	164	6.5	3.8	9.2	20.8	*
Cuarto	Mujer	1942	1.6	1225	174	7.3	4.6	10.0	19.1	

País grande: desocupación

Departamentos (24) - Sexo

Departamento	Sexo	n	Deff	n.eff	gl	Desocupación%	Li %	Ls %	cv %	Alerta
3	Mujer	318	1.7	182	4	3.5	0.5	6.6	43.4	*
6	Hombre	594	1.1	560	5	1.8	0.5	3.0	36.1	*
24	Hombre	1680	0.6	2963	25	0.4	0.1	0.6	33.8	*
3	Hombre	318	1.4	232	4	4.6	1.6	7.5	33.1	*
24	Mujer	1430	1.1	1309	25	1.0	0.4	1.6	31.3	*
6	Mujer	545	1.0	529	5	2.6	1.1	4.2	30.2	*
10	Mujer	2629	1.1	2297	37	1.1	0.6	1.6	21.8	*
18	Hombre	2652	1.0	2607	31	1.1	0.6	1.5	21.2	*
10	Hombre	2853	1.0	2825	37	1.0	0.6	1.4	20.5	*
21	Hombre	6413	2.0	3149	74	1.1	0.6	1.5	20.5	*

País pequeño: desocupación

Departamentos (8) - Sexo

Departamento	Sexo	n	Deff	n.eff	gl	Desocupación%	Li %	Ls %	cv %	Alerta
7	Hombre	201	0.9	212	20	1.8	0.0	4.2	69.3	*
7	Mujer	172	0.9	185	19	5.2	0.0	12.3	68.8	*
3	Mujer	70	0.9	79	8	5.3	0.0	12.4	67.9	*
10	Hombre	2031	0.8	2402	35	0.3	0.0	0.7	53.7	*
6	Mujer	268	0.9	307	21	5.8	0.4	11.2	47.0	*
9	Hombre	148	1.0	151	14	4.7	0.4	9.1	46.7	*
10	Mujer	2031	0.9	2201	35	0.8	0.1	1.5	43.2	*
3	Hombre	69	0.4	166	8	6.6	1.5	11.8	39.7	*
9	Mujer	143	1.0	149	14	18.1	6.7	29.5	32.1	*
6	Hombre	278	1.1	256	21	9.2	4.5	13.8	25.8	*

País grande: pobreza

Departamentos (24) - Sexo

Departamento	Sexo	n	Deff	n.eff	gl	Pobreza %	Li %	Ls %	cv %	Alerta
3	Hombre	318	2.8	115	4	7.9	2.9	12.8	32.0	*
3	Mujer	318	2.4	135	4	11.3	6.0	16.7	24.1	*
6	Hombre	594	3.4	172	5	10.4	5.9	15.0	22.3	*
6	Mujer	545	1.9	287	5	9.5	6.1	12.9	18.2	*
24	Hombre	1680	4.0	425	25	9.0	6.3	11.7	15.4	
24	Mujer	1430	3.6	395	25	10.1	7.1	13.1	15.0	
10	Mujer	2629	4.2	633	37	8.1	6.0	10.2	13.4	
10	Hombre	2853	4.3	665	37	7.8	5.8	9.8	13.3	
8	Hombre	3622	3.1	1181	159	4.9	3.7	6.1	12.8	
8	Mujer	3814	2.9	1336	159	5.2	4.0	6.4	11.7	
22	Hombre	4573	4.2	1097	212	8.6	6.9	10.3	9.8	
22	Mujer	4815	4.9	978	211	9.8	7.9	11.6	9.7	

País pequeño: pobreza

Departamentos (8) - Sexo

Departamento	Sexo	n	Deff	n.eff	gl	Pobreza %	Li %	Ls %	cv %	Alerta
3	Hombre	69	1.4	48	8	2.3	0.0	6.5	94.2	*
3	Mujer	70	0.9	76	8	1.5	0.0	4.3	91.8	*
9	Hombre	148	6.8	22	14	12.0	0.0	25.7	58.3	*
9	Mujer	143	4.5	32	14	11.2	0.2	22.1	49.9	*
6	Hombre	278	6.5	43	21	16.6	5.5	27.7	34.2	*
6	Mujer	268	5.3	50	21	19.1	8.3	29.9	28.9	*
7	Mujer	172	5.7	30	19	35.0	17.9	52.0	24.8	*
7	Hombre	201	7.1	28	20	37.5	19.8	55.3	24.1	*
5	Hombre	1472	5.5	265	151	11.2	7.4	15.0	17.3	
5	Mujer	1511	3.3	452	151	11.4	8.5	14.4	13.1	

Conclusiones

Acerca de las encuestas

- Las encuestas son instrumentos valiosos que permiten obtener información desagregada y precisa por los dominios de diseño.
- A veces es posible obtener información desagregada y precisa en dominios que no fueron considerados en el diseño.
- Muchas veces no es posible obtener información desagregada y precisa en dominios de interés para el seguimiento de los ODS.
- Algunas Oficinas Nacionales de Estadística utilizan el coeficiente de variación como una medida única de información sobre la precisión de la cifra.

Acerca del coeficiente de variación

- Existen otros elementos adicionales al coeficiente de variación que permiten tener una mejor panorámica de la precisión de las cifras.
- Es posible extender el criterio de publicación y/o supresión de cifras a otras medidas que complementan el coeficiente de variación.
- El coeficiente de variación no es simétrico con respecto a las proporciones.
- Cuando las proporciones estimadas son muy pequeñas, es posible que el intervalo de confianza sea angosto y aún así tener coeficientes de variación que se salgan del umbral predefinido para publicación.
- Es posible realizar una transformación sobre el coeficiente de variación que permita obtener mejores resultados en proporciones cercanas a cero.

Acerca de los países

- Para las mismas desagregaciones y los mismos indicadores, las muestras de los países grandes tienen un mayor potencial de precisión en las estimaciones.
- Para aumentar la precisión de los estimadores en las desagregaciones, los países podrían aumentar el tamaño de muestra de la encuesta.
- Para aumentar la precisión de los estimadores en las desagregaciones, los países podrían apoyarse en información auxiliar externa a la encuesta.
- Ante las limitaciones de las encuestas, los métodos de estimación en áreas pequeñas son una excelente opción para hacer frente a este reto.



NACIONES UNIDAS



¡Gracias!

andres.gutierrez@un.org