Revista Venezolana de Análisis de Coyuntura, 1997, Vol. III, Nº 1 (ene-jun), pp. 132-145

SOBRE LA COMPARACIÓN ESTADISTICA DE INDICES DE POBREZA*

Alberto Camardiel y Dulce María Mesa
ESCUELA DE ESTADÍSTICA, UCV

RESUMEN

Esta comunicación tiene por objeto llamar la atención sobre la comparación de índices de pobreza estimados con datos provenientes de muestras probabilísticas. ¿Podemos afirmar que ha disminuido la pobreza en un universo determinado, si la estimación de la prevalencia de este fenómeno pasó, por ejemplo, de 35% a 32% en dos años sucesivos? ¿O deberíamos concluir que tal diferencia se debe atribuir al efecto del muestreo? El trabajo que ahora presentamos, aborda el problema de la comparación estadística de índices de pobreza y presenta consideraciones de carácter metodológico que deben tenerse en cuenta cuando se llevan a cabo análisis que contemplan este tipo de comparaciones.

Palabras claves: pobreza, comparaciones estadísticas, intervalos confidenciales.

1. INTRODUCCIÓN

En este artículo se presenta el problema de cómo comparar índices de pobreza entre regiones y en el tiempo, cuando estamos en presencia de variabilidad estadística. El conocimiento del coeficiente de variación de un índice de pobreza, calculado para varios agregados de población, puede ayudar en la comparación de la variabilidad relativa asociada a los distintos índices, pero por sí solo no puede resolver el problema de la comparación estadística. Los intervalos confidenciales que, como sabemos, incorporan la incertidumbre contenida en el proceso estadístico de selección, observación o medición y estimación, proporcionan un camino que se puede emplear para resolver el problema mencionado.

---

* Los autores agradecen los comentarios y las útiles sugerencias formuladas a una primera versión del trabajo por un revisor arbitro anónimo y por el profesor Félix Seijas Zerpa. Igualmente expresan su agradecimiento a Raquel Fernández, por la colaboración prestada en la preparación de las bases de datos necesarias para la realización de los cálculos que se presentan en este trabajo.

En este trabajo presentamos básicamente el problema de la construcción de intervalos confidenciales para un índice o un contraste de índices, señalando aspectos que deben ser tenidos en cuenta cuando se utilizan muestras complejas en la recolección de los datos. La mayor dificultad que se enfrenta en la construcción de estos intervalos es el cálculo del error de muestreo del índice o del contraste. Cuando no se cuenta con los medios que faciliten la solución de este problema (equipo lógico y físico apropiados), mostramos como se puede utilizar un resultado simple para la detección de situaciones en las que no se haya presentado cambios en los niveles de pobreza. Finalmente, para ilustrar algunos de los procedimientos considerados en el artículo, calculamos el coeficiente de variación y un intervalo confidencial del 95% para el índice de prevalencia de pobreza en el Area Metropolitana de Caracas para dos semestres consecutivos, a saber, el segundo de 1995 y el primero de 1996.

2. ALGUNOS INDICES DE POBREZA

Consideremos un conjunto, de n personas u hogares, que denotaremos por A, con ingresos ordenados ascendentemente que representaremos por

$$y_1, y_2, \ldots, y_n,$$

para un momento determinado del tiempo. Consideremos además que z define una línea de pobreza para el mismo conjunto de unidades determinada convencionalmente. Se define como la prevalencia de pobreza en A a todas las unidades de A con ingresos no superiores a z y se calcula a través de la fórmula

$$H = \frac{q}{n},$$

en donde q representa el número de unidades con ingresos menores o iguales a z.

En esta definición no se utiliza el nombre de incidencia de pobreza con el que usualmente se traduce al índice (Head Count Poverty Index), porque seguimos la convención de la epidemiología moderna que distingue entre prevalencia e incidencia (Ver, por ejemplo, Kleinbaum et. al., 1982), y designa con este último término al número de nuevos casos de una condición determinada, en nuestro caso, pobreza, surgidos en un período dado. Para ambos conceptos se pueden concebir diversas definiciones operatorias que proporcionan mediciones de la frecuencia de la pobreza, pero cada uno sirve a fines distintos. El conocimiento de la prevalencia es esencial para la planificación de políticas de alivio de la pobreza, mientras que el concepto de incidencia es útil en la investigación causal del fenómeno. Ambos conceptos, requieren también de datos primarios obtenidos por medio de estudios estadísticos de naturaleza distinta. La prevalencia usualmente se calcula con datos obtenidos mediante estudios transversales y la incidencia, mediante el uso de estudios de cohortes.

El índice de prevalencia se comprende e interpreta fácilmente, pero presenta desventajas cuando se pretende utilizar para evaluar el impacto de políticas sociales de alivio de la pobreza (ibíd.). Un segundo índice que incorpora a la prevalencia una medida de la "profundidad" o "intensidad" de la pobreza (Ravallion, 1994) es el denominado índice de Brecha de Pobreza (Poverty Gap Index), dado por

$$PG = \frac{1}{n}\sum_{i=1}^{q}\left(1 - \frac{y_i}{z}\right).$$

Este índice se puede escribir como

$$PG = HI,$$

en donde I se define por

$$I = 1 - \frac{\mu^z}{z},$$

con μ^z representando el consumo promedio de los pobres.

Se le critica a esta medida que no captura la severidad de la pobreza. Para entender este aspecto consideremos (ibíd.) dos agregados A_1 y A_2, formados por cuatro unidades, personas u hogares, con distribuciones de ingreso dadas por (1,2,3,4) y (2,2,2,4) respectivamente. Se puede verificar que el índice PG toma el mismo valor para ambos agregados, a saber, 0,25, pero la pobreza es más "severa" en A_1 que en A_2, puesto que el más pobre de A_1 recibe la mitad del ingreso que recibe el más pobre de A_2. Sen (1976) propuso un índice que toma en cuenta la "severidad" de la pobreza, incorporando en su construcción la desigualdad de la distribución del ingreso entre los pobres, medida por el coeficiente de desigualdad de Gini. Este índice se define como

$$S = H\,[I - k\,(1 - I)G^p],$$

en donde $k = \dfrac{q}{q+1}$ y G^p es el índice de Gini (Alker, 1965) para los pobres.

Una propiedad de mucho interés, sobre todo para la construcción de perfiles de pobreza es la aditividad (Ravallion, 1994). Los índices H y PG son aditivos, pero no el índice de Sen. Definimos la aditividad de la siguiente forma. Consideremos un agregado A formado por n unidades y consideremos una partición de A dada por $\{A_i\}$ para $i = I,...,m$. Si n_i es el número de unidades en A_i, entonces decimos que un índice de pobreza P es aditivo si y sólo si

$$P = \sum_{i=1}^{m}\frac{n_i}{n}P_i,$$

en donde P_i es el índice de pobreza P aplicado en A_i.

Una medida aditiva de pobreza la proporciona el índice de Foster-Greer-Thorbecke, (1984) definido como

$$P_2 = \frac{1}{n} \sum_{i=1}^{q} \left(1 - \frac{y_i}{z}\right)^2$$

Este índice no tiene una interpretación tan directa como H, pero a su favor vale la afirmación de Ravallion (1994) "...lo que hace útil (a una medida de pobreza) es su habilidad para ordenar distribuciones en una forma mejor que las (medidas) alternativas...".

Resulta ilustrativo comparar el desempeño de las medidas consideradas en la medición y ordenación de la pobreza para varias distribuciones hipotéticas del ingreso para cuatro unidades, que se presentan en el cuadro siguiente

Distribución	$(H)_{z=3}$	$(I)_{z=3}$	$(PG)_{z=3}$	$(G^P)_{z=3}$	$(S)_{z=3}$	$(\mathbf{P_2})_{z=3}$
1,1,1,4	3/4	6/9	2/4	0	2/4	12/36
1,1,3,4	3/4	4/9	(4/3)(1/4)	4/15	(5/3)(1/4)	8/36
1,2,3,4	3/4	3/9	1/4	4/18	(4/3)(1/4)	5/36
2,2,2,4	3/4	3/9	1/4	0	2/4	3/36
3,3,3,4	3/4	0	0	0	2/4	0

Es conveniente advertir que podemos pensar en los índices de pobreza agregada como casos particulares de la forma general (ibíd.)

$$P = \frac{1}{n} \sum_{i=1}^{n} P(z,y_i),$$

en donde $P(z,y_i)$ representa una medida de pobreza para la unidad i-ésima que satisface las siguientes propiedades

1) $P(z,y_i) > 0$ si $y_i \leq z$
2) $P(z,y_i) = 0$ si $y_i > z$
3) $P(z,y_i) \geq P(z',y_i)$ si $z \leq z'$
4) $P(z,y_i) \leq P(z',y_i)$ si $z \geq z'$

En todas las definiciones dadas anteriormente no hemos hecho ningún tipo de consideración sobre la naturaleza de las observaciones empleadas, esto es, si el agregado de unidades considerado constituye una población o una muestra de una población determinada. De ahora en adelante supondremos que los datos básicos se han obtenido mediante el empleo de una muestra probabilística y en consecuencia estaremos considerando por un lado índices de pobreza como parámetros poblacionales y por otro lado, índices de pobreza como estimadores muestrales de esos parámetros. Denotaremos los parámetros mediante letras mayúsculas y sus estimadores mediante letras minúsculas.

Cuando nos planteamos realizar comparaciones estadísticas entre índices de pobreza, podemos representar cada comparación como una combinación lineal de los índices de pobreza definidos para agregados de interés de la forma siguiente

$$\bar{V} = \sum_{i=1}^{m} c_i \bar{V}_i \text{ con } \sum_{i=1}^{m} c_i = 0.$$

Bajo esta fórmula general, se contemplan casos particulares como por ejemplo: a) comparaciones $\bar{V}_i - \bar{V}_j$ para el momento o región i-ésima y el momento o la región j-ésima y b) comparaciones $\bar{V}_i - \frac{1}{m}\left(\sum_{k=1}^{m} \bar{V}_k\right)$ del índice para la región o momento i-ésimo contra el promedio simple global. Hay que destacar además, que la consideración simultánea de varias comparaciones presenta más complejidad que la consideración de una sola comparación. El primer problema habría que plantearlo en el contexto de la inferencia estadística simultánea. En este trabajo nos ocuparemos únicamente del segundo tipo de problema.

3. UN PROBLEMA CLÁSICO DE INFERENCIA ESTADÍSTICA

La realización de una comparación estadística de índices de pobreza se puede abordar como un problema clásico de inferencia estadística cuya solución es bien conocida, aún cuando no está exento por lo menos de dificultades de cálculo, cuando los datos han sido generados por medio de muestras de diseño complejo.

Si denotamos $\bar{v}$ por el estimador del índice de pobreza de una población en un momento determinado o de la comparación de interés entre varios valores de un índice para regiones o momentos distintos, sabemos que para tamaños de muestra grande y formas de $\bar{v}$ tales como medias, proporciones o comparaciones de medias o proporciones, podemos afirmar que el intervalo confidencial cubrirá al valor poblacional que estima $\bar{v}$, con una probabilidad igual a $1-\alpha$, en donde $se(\bar{v})$ es el error estándar calculado de $\bar{v}$ basado en *p* grados de libertad y $t_p^{\alpha/2}$ es el percentil $100(1-\alpha/2)$ de la distribución *t* de Student con *p* grados de libertad, que tomaremos como el percentil correspondiente de la distribución normal si *n* es suficientemente grande.

El fundamento matemático de este resultado se basa en la aproximación a la distribución normal de la distribución en el muestreo del estadístico $\bar{v}$. Este resultado se ha verificado teórica y prácticamente para tamaños de muestra grandes, formas de $\bar{v}$ tales como medias o diferencias de medias y muestras complejas (Kish, 1965). Hay que destacar (ibíd.) que la calidad de la aproximación depende de: 1) los valores de $t_p^{\alpha/2}$ que se utilicen; 2) el tipo de variables para las que se calcula $\bar{v}$ y; 3) la consideración de subclases, si es que estas se utilizan para las estimaciones de $\bar{v}$. Así, para valores de $t_p^{\alpha/2}$ muy grandes, que corresponden a probabilidades $1-\alpha/2$ muy grandes, la aproximación pierde calidad. Esto no es un problema serio para valores "razonables" de $1-\alpha$, tales como 0,95. Tampoco representa, en el caso que nos ocupa, una amenaza seria la advertencia del apartado 2), por cuanto los índices más comunes que consideramos en la cuantificación de la pobreza son proporciones o medias de brechas de ingreso para los pobres, que deberían comportarse con asimetrías positivas mucho menos acentuadas que el ingreso para la población total, una situación que perjudicaría la aproximación. El punto 3) puede llegar a constituir un problema serio para cierto tipo de comparaciones, por cuanto se tiene evidencia de que la consideración de ciertas subclases, por ejemplo, subclases segregadas, tales como regiones, aún para muestras grandes no asegura necesariamente la aproximación normal (ibíd.)

Los términos dominio y subclase se utilizan en la literatura sobre muestreo de poblaciones finitas con el fin de denotar subdivisiones o particiones de la población y las subdivisiones inducidas por los dominios en la muestra, respectivamente, con propósitos de estimación separada (Kish, 1980). Se distingue entre dominios de diseño, dominios mixtos y clases cruzadas. Los primeros se utilizan para planificar la selección de muestras separadas, por ejemplo, los sectores urbano y rural. Los segundos corresponden a situaciones intermedias que tienden a una concentración desigual en unidades de muestreo y estratos, por ejemplo, ocupaciones en la agricultura y en la industria segregadas por fuerzas naturales y sociales. Los últimos son clasificaciones que se presentan en los estratos, en la muestra y aún en las unidades de muestreo, por ejemplo, el sexo. De estos tres tipos de dominios, las clases cruzadas son las que presentan mayores problemas cuando se utilizan muestras complejas.

Las comparaciones de pobreza a excepción de las diferencias de prevalencias entre subclases, corresponden a la denominación de estadísticos complejos (Kish y Frankel, 1974) con errores de muestreo sólo conjeturados para la selección estratificada de elementos y difícil para el muestreo por conglomerados. En este último caso hay que recurrir al cálculo de los errores de muestreo por métodos aproximados.

La forma general considerada permite generar soluciones para casos particulares. Por ejemplo bajo muestreo aleatorio simple (lo cual es poco realista en evaluaciones de pobreza), si intentamos determinar la prevalencia de pobreza en una población y momento determinado con una muestra de tamaño, el intervalo confidencial del 100(1 – α)% asume la forma

$$\left(p - t_p^{\alpha/2}\sqrt{\frac{pq}{n}}; p + t_p^{\alpha/2}\sqrt{\frac{pq}{n}}\right),$$

en donde p es la proporción en la muestra de hogares o individuos con ingreso o consumo no mayor que una línea de pobreza determinada convencionalmente y $q = 1 - p$. Este intervalo nos permite entonces, formular declaraciones en relación a la prevalencia de pobreza P en la población del tipo "P estará comprendida entre $p \pm t_p^{\alpha/2}\sqrt{\frac{pq}{n}}$ con una confianza del 100(1 – α)%." La construcción de un intervalo semejante para la misma población en otro momento nos permitiría verificar estadísticamente la hipótesis de cambio en la prevalencia de pobreza en los dos momentos considerados, puesto que una dócima de la hipótesis nula de igualdad de prevalencias de pobreza versus la hipótesis alternativa de diferencias con un nivel de significación α, es equivalente a la construcción de los intervalos confidenciales del 100(1 – α)% para cada momento y la constatación consecuente de su solapamiento o de su disyunción, es decir del no rechazo de la hipótesis nula o del rechazo de la hipótesis nula respectivamente.

En forma análoga se podría comparar la igualdad de prevalencias entre dos regiones diferentes de un mismo país en un momento dado.

En ambos casos hubiéramos podido abordar el problema como la construcción de un intervalo confidencial del 100(1 – α)% para la diferencia de prevalencias mediante la expresión

$$\left((p_1 - p_2) - t_p^{\alpha/2}\sqrt{\frac{p_1q_2}{n_1} + \frac{p_2q_2}{n_2}}; (p_1 - p_2) + t_p^{\alpha/2}\sqrt{\frac{p_1q_2}{n_1} + \frac{p_2q_2}{n_2}}\right),$$

en donde p_i representa la prevalencia de pobreza en la muestra para la región i-ésima y n_i el número de sujetos en la muestra en la misma región (Ravallion, 1994). La docimasía de la hipótesis nula de igualdad de prevalencias versus la alternativa de no igualdad, se lleva a cabo verificando si el intervalo contiene o no contiene al cero respectivamente.

4. DIFICULTADES PRÁCTICAS

El intervalo confidencial

$$\left(\bar{v} - t_p^{\alpha/2} se(\bar{v}); \bar{v} + t_p^{\alpha/2} se(\bar{v})\right)$$

requiere para su determinación del cálculo del error estándar de $\bar{v}$. La estimación del error estándar de $\bar{v}$ para muestras complejas de acuerdo a expresiones analíticas puede resultar de extrema dificultad, sino imposible. Kish (1965) presenta varios métodos que proporcionan estimaciones aproximadas cuando no se dispone de métodos formales de estimación.

El intervalo confidencial también se puede expresar en términos del coeficiente de variación de $\bar{v}$ como se detalla a continuación

$$\left\{\bar{v}\left[1 - t_p^{\alpha/2} cv(\bar{v})\right]; \bar{v}\left[1 + t_p^{\alpha/2} cv(\bar{v})\right]\right\}$$

o en términos del Efecto de Diseño de $\bar{v}$, que denotaremos por Deff, como

$$\left(\bar{v} - t_p^{\alpha/2} se_{MAS}(\bar{v})\sqrt{Deff}; \bar{v} + t_p^{\alpha/2} se_{MAS}(\bar{v})\sqrt{Deff}\right)$$

Este estadístico se define como

$$Deff = \frac{se(\bar{v})}{se_{MAS}(\bar{v})}.$$

Puesto que el Efecto de Diseño es en general, mayor que uno para muestras complejas, podemos considerar al intervalo

$$\left(\bar{v} - t_p^{\alpha/2} se_{MAS}(\bar{v}); \bar{v} + t_p^{\alpha/2} se_{MAS}(\bar{v})\right)$$

como una cota inferior del intervalo apropiado. En caso de no contar con las facilidades para el cómputo de los errores de muestreo expresados en $se(v)$, podríamos utilizar ventajosamente este resultado para llevar a cabo las comparaciones de interés. En consecuencia, si se acepta la hipótesis nula de igualdad de índices de pobreza en la población mediante el intervalo simple, también se tendría que haber aceptado con el intervalo apropiado. Lo que no podemos saber es cual hubiera sido el resultado según el intervalo apropiado, si en lugar de aceptar la hipótesis de igualdad con el primer intervalo, la hubiéramos tenido que rechazar.

5. UNA VÍA ALTERNATIVA DE SOLUCIÓN

Con el poder de cálculo que ofrecen los ordenadores actuales es tentador pensar en el uso del muestreo autogenerado (Bootstrap) ideado por Efron

(Efron, 1982) como una vía alternativa de construcción de intervalos confidenciales para $\bar{v}$.

El empleo del muestreo autogenerado para la construcción de intervalos de confianza de comparaciones de pobreza, cuando se dispone de una muestra probabilística que produjo datos $y_1, y_2, \ldots, y_n$, puede utilizar la aproximación de Monte Carlo, aplicando los siguientes pasos (Ramírez, 1986)

1) Construir la distribución muestral empírica $\hat{F}$ asignando probabilidad $1/n$ a cada punto $y_1, y_2, \ldots, y_n$.

2) Seleccionar una muestra aleatoria simple con reemplazo de tamaño n de $\hat{F}$ que denotaremos por $y_1^*, y_2^*, \ldots, y_n^*$.

3) Calcular $\hat{v} = \bar{v}(\hat{F})$.

4) Repetir los pasos 2) y 3) independientemente un número grande de veces, digamos B, para obtener replicaciones autogeneradas

$$\hat{v}_1^*, \hat{v}_2^*, \ldots, \hat{v}_B^*.$$

5) Calcular las estimaciones autogeneradas del índice de pobreza y de su error estándar, mediante las fórmulas

$$\hat{v}_.^* = \frac{1}{B}\sum_{i=1}^{B} \hat{v}_i^*$$

y

$$se(\hat{v}) = \sqrt{\frac{\sum_{i=1}^{B}\left(\hat{v}_i^* - \hat{v}_.^*\right)^2}{B-1}}.$$

6) Construir el intervalo confidencial aproximado del 100(1 – α)%, por el método del percentil (Efron, 1982)

$$\left[\hat{v}_{INF}(\frac{\alpha}{2}); \hat{v}_{SUP}(\frac{\alpha}{2})\right],$$

en donde $$\hat{v}_{INF}(\frac{\alpha}{2}) = C\hat{D}F^{-1}(\frac{\alpha}{2}) \text{ y } \hat{v}_{SUP}(\frac{\alpha}{2}) = C\hat{D}F^{-1}(1-\frac{\alpha}{2})$$

y $$C\hat{D}F(t) = \Pr ob_* \{\hat{v}_.^* \leq t\}.$$

En la actualidad estamos planeando un proyecto de investigación que aborda esta línea de trabajo y esperamos contar prontamente con resultados que permitan evaluar este enfoque alternativo para facilitar las comparaciones estadísticas de índices de pobreza de cualquier tipo.

6. ALGUNOS CÁLCULOS PARA PREVALENCIAS DE POBREZA CON DATOS DE LA ENCUESTA DE HOGARES POR MUESTREO DE VENEZUELA

Para este trabajo se llevaron a cabo algunos cálculos con los datos del segundo semestre del año 1995 y del primer semestre del año 1996 de la Encuesta de Hogares por Muestreo (EHM) para el Area Metropolitana de Caracas (AMC), que es uno de los dominios de estudio de la EHM que lleva a cabo regularmente la Oficina Central de Estadística e Informática (OCEI). El número de hogares entrevistados en el AMC para el II-95 fue de 2.257 y de 2.271 para la muestra correspondiente al I-96. Los cálculos de la prevalencia de pobreza se hicieron basados en el costo de las canastas de consumo per-cápita nacionales elaboradas por la OCEI, a saber, Bs. 24.076,94 para el II-95 y Bs.33.633,44 para el I-96 (aún cuando esta cifra todavía no se considera oficial).

La EHM es una investigación cuyo objetivo básico es el de "...proporcionar información oportuna y confiable sobre la estructura y evolución de la Fuerza de Trabajo y en particular sobre el fenómeno del desempleo...(OCEI, 1987), pero también proporciona datos demográficos, sociales y económicos que posibilitan cálculos de pobreza, claro está, con las limitaciones inherentes por no ser éste el objetivo principal de la investigación. La muestra tiene carácter continuo con periodicidad semestral y proporciona información a nivel nacional y a nivel de otros dominios de diseño. El tipo de muestreo utilizado es probabilístico trietápico. En la primera etapa se seleccionan segmentos censales, esto es, unidades compuestas por 200 viviendas en promedio en las ciudades, con probabilidad proporcional al número de viviendas reportadas en el Censo de Población y Vivienda realizado en 1981. En la segunda etapa se seleccionan subdivisiones de los segmentos constituidas por 50 viviendas aproximadamente, denominadas áreas, también con probabilidad proporcional al número de viviendas resultantes del conteo rápido que se lleva a cabo como parte del proceso de actualización del marco de muestreo y, finalmente, en la tercera etapa se eligen viviendas con probabilidad igual dentro de cada área seleccionada y, se entrevista a todos los hogares que habitan las viviendas elegidas.

La fórmula empleada para estimar la prevalencia de pobreza con los datos de la EHM viene dada por:

$$\hat{p} = \frac{\hat{Y}}{\hat{X}}$$

en donde $\hat{Y}$ y $\hat{X}$ representan el número de hogares pobres y el total de hogares en la muestra respectivamente, ponderadas con los inversos de las probabilidades de selección. En la EHM también se aplican ajustes a los datos para compensar la no respuesta y para conformar la distribución estimada de personas por edad y sexo por la vía de ponderaciones adicionales. Puesto que el total de hogares que caen en la muestra es también una variable aleatoria, este estadístico es un estimador de razón.

Los errores de muestreo para estadísticos en la EHM se estiman por el Método de las Diferencias Pareadas Sucesivas, consistente en el cálculo de diferencias sucesivas de estos estadísticos, pero entre segmentos contiguos ordenados según criterios geográficos para la muestra seleccionada (Documento mimeografiado, sin autor ni fecha). Para la prevalencia de pobreza tenemos entonces que

$$\text{Varianza }(\hat{P}) = \frac{1}{\hat{X}_2}\left[\text{Varianza }(\hat{Y}) + \hat{P}^2\,\text{Varianza }(\hat{X}) - 2\hat{P}\,\text{Covarianza }(\hat{X},\hat{Y})\right]$$

y

$$\text{Se }(\hat{P}) = \sqrt{\text{Varianza }(\hat{P})}$$

en donde

$$\text{Varianza }(\hat{X}) = \sum_{h=1}^{L}(x_{ha} - x_{hb})^2,$$

$$\text{Varianza }(\hat{Y}) = \sum_{h=1}^{L}(y_{ha} - y_{hb})^2,$$

$$\text{Covarianza }(\hat{X},\hat{Y}) = \sum_{h=1}^{L}(x_{ha} - x_{hb})(y_{ha} - y_{hb}),$$

L : Número de estratos formados por selecciones pareadas de segmentos,

x_{hi} : Total poblacional estimado de la variable X en el segmento i del estrato h ($i = a,b$),

y_{hi} : Total poblacional estimado de la variable Y en el segmento i del estrato h ($i = a,b$).

La prevalencia de pobreza estimada para el II-95 fue del 40,32% con un coeficiente de variación de 3,14%. El intervalo confidencial del 95% para la prevalencia de pobreza, basado en el error estándar para la muestra compleja, calculado según el método de las diferencias pareadas, resultó igual a

(37,84%;42,80%)

En contraste, el intervalo confidencial del 95% basado en el error estándar de la prevalencia de pobreza para el muestreo aleatorio simple resultó igual a

(38,29%;42,34%)

Se confirma que este último está contenido en el anterior, proporcionando por lo tanto una cota inferior del intervalo apropiado. El efecto de diseño para este semestre fue de 1,23.

Similarmente el estimador de la prevalencia de pobreza para el primer semestre de 1996 fue de 38,90 con un coeficiente de variación de 3,29%. El intervalo confidencial del 95% basado en el error estándar calculado según el método de las diferencias pareadas es:

(36,39%;41,41%)

que contiene también al intervalo

(36,90%;40,91%)

basado en el cálculo del error bajo el supuesto de muestreo aleatorio simple. Para este semestre el efecto de diseño resultó igual a 1,25.

Debemos destacar que todos los cálculos sobre los que se basan los intervalos confidenciales son propios y, en consecuencia, no comprometen a la OCEI.

La estimación puntual de la prevalencia de pobreza para el primer semestre del 1996 fue menor que para el segundo semestre de 1995, pero puesto que los dos intervalos confidenciales basados en $se(\bar{v})$ se solapan, no podemos rechazar la hipótesis nula de igualdad de prevalencias para los semestres considerados al nivel de significación del 95%. Los valores obtenidos son compatibles con la variabilidad en el muestreo y otras características propias del diseño y ejecución de la investigación. Notemos que la misma conclusión pudimos haberla obtenido empleando los intervalos basados en $se_{mas}(\bar{v})$, porque estos también se solapan, pero no hubiéramos podido concluir nada en caso de solapamiento de estos últimos.

CONCLUSIONES

1) Se debe distinguir entre significación estadística y significación sustantiva. Es posible que una diferencia entre índices de una determinada magnitud no resulte estadísticamente significativa, pero sí tenga importancia social.

2) Las comparaciones entre índices de pobreza se pueden llevar a cabo siguiendo procedimientos clásicos de la estadística inferencial.

3) La aproximación normal en la que se fundamenta la facilidad y la conveniencia de los métodos de comparación basados en intervalos confidenciales normales, no está automáticamente asegurada. Hay que prestar atención, sobre todo cuando se quiere comparar índices para subclases que no se contemplaron explícitamente en el diseño de la muestra.

4) El cálculo del coeficiente de variación no facilita necesariamente la comparación estadística de índices de pobreza. La construcción de intervalos confidenciales o el empleo de pruebas de hipótesis depende en una última instancia de la estimación del error de muestreo del índice de pobreza considerado.

5) Los intervalos confidenciales basados en el error estándar correspondiente al muestreo aleatorio simple, aún cuando se utilicen muestras complejas, sirven para detectar la igualdad de índices de pobreza, no así lo contrario.

6) El uso del muestreo autogenerado (Bootstrap) se presenta como una alternativa viable para el análisis estadístico de las comparaciones de pobreza y se debería explorar su uso en este sentido.

BIBLIOGRAFÍA

Alker, Hayward (1965), *El uso de la matemática en el análisis político*, Amorrortu Editores, Buenos Aires.

Efron, Bradley (1982), *The Jacknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia.

Foster, James, J. Greer y E. Thorbecke (1984), "*A Class of Decomposable Poverty Measures*," Econometrica, Vol. 52, 761-765.

Kish, Leslie (1965), *Survey Sampling*, John Wiley & Sons, Nueva York.

Kish, Leslie (1980), "*Design and Estimation for Domains*," The Statistician, Vol. 24, Nº 4, 209-222.

Kish, Leslie y Martin Richard Frankel (1974), "*Inference from Complex Samples*," The Journal of the Royal Statistical Society Series B (Methodological), Vol. 36, Nº 1, 1-37.

Kleinbaum, David. G., L. Kupper y H. Morgenstern (1982), *Epidemiologic Research. Principles and Quantitative Methods*, Van Nostrand Reinhold, Nueva York.

Oficina Central de Estadística e Informática, *20 Años de la Encuesta de Hogares por Muestreo en Venezuela*, Presidencia de la República, Caracas, 1987.

Ramírez, Guillermo (1986), *Tres métodos de estimación del error estadístico: el Jacknife, el Bootstrap y la Validación Cruzada*, mimeografiado, Caracas.

Ravallion, Martin (1994), *Poverty Comparisons*, Hardwood Academic Publishers, Suiza.

Sen, Amartya (1976), "*Poverty: An Ordinal Approach to Measurement*," Econométrica, Vol. 46, 437-446.

Sin Autor. *Cálculo del error de muestreo de la encuesta de hogares*, mimeografiado, sin referencias.