

ACERCA DE LA CALIDAD DE LOS DATOS

En general, cuando se efectúa la descripción y el análisis de un conjunto de datos y de los indicadores que se derivan, es necesario tener en cuenta los posibles errores que afectan a la información básica.

Se presentan dos tipos de errores: de cobertura y de contenido.

Los errores de **Cobertura** están relacionados con la omisión de personas, dado que existen tendencias; por ejemplo, a no declarar un recién nacido o un niño de corta edad, lo mismo sucede con personas de edad avanzada; afectando, en estos casos, ciertas edades de los extremos de la vida (sin distinción de género). Otras veces, no se cuenta algún miembro del hogar que se encuentra ausente por trabajo, en el momento del levantamiento de la información; lo que provoca omisión, particularmente de hombres en edades activas. Cuando se realizan censos de población, una de las tareas importantes es determinar, con la mayor aproximación posible, el grado de omisión de la estadística censal; en el caso del Censo de Población de 1995, se efectuó poco tiempo después, una encuesta de Cobertura y Sesgo, arrojó una omisión de personas cercana al 3.5 por ciento.

En el caso de los errores de **Contenido**, se capta a la persona pero se cambia alguno de sus atributos, siendo el más común *la edad*, dado que existe la tendencia, por parte del declarante o la persona que suministra la información, de redondear la edad, generalmente, a dígito terminado en 0 ó 5, en menor medida en dígitos pares, esto en desmedro de los impares (exceptuado el 5). Otros errores de contenido dependen de los temas investigados; por ejemplo, la persona se dice alfabeta, cuando no lo es, soltera cuando está separada, casada cuando es unida, etc. Lógicamente, estos errores afectan la calidad de los datos.

Es posible, a partir de ciertos índices, evaluar la calidad de la declaración de la edad; y esta evaluación se tiene en cuenta para calificar al conjunto de datos obtenidos en las investigaciones. Los índices más utilizados, recomendados por Naciones Unidas, son el de Myers y el de Whipple¹.

ÍNDICE DE MYERS

Evalúa los dígitos finales en la declaración de la edad de las personas, informando de la atracción o el rechazo de cada uno de ellos, además se obtiene un índice resumen. Si no hay atracción, el valor de este índice resumen es cero, por el contrario, si todas las personas declaran su edad con el mismo dígito final (por ejemplo 0, 10 años, 20, ... 60, etc.), el índice alcanza el valor de 180; por lo tanto, los valores bajos son indicativos de poca atracción de los dígitos, implicando que la

¹ Naciones Unidas, *Manual II - Métodos para evaluar la calidad de los datos básicos destinados a los cálculos de la población*. Nueva York, 1955

declaración de la edad es aproximadamente correcta y, como consecuencia la información en general, se interpreta como de buena calidad. A la inversa, valores elevados del índice, son demostrativos de una declaración deficiente de la edad, lo que afecta la calidad de los datos y hace que se deban tomar con precaución, los resultados que se derivan.

Para calificar el nivel de atracción de los dígitos, a partir del índice de Myers, se utiliza la siguiente escala:

Valor del índice	Nivel de atracción
0.0 a 5.0	Bajo
5.1 a 15.0	Mediano
15.1 a 30.0	Alto
30.1 y más	Muy alto

ÍNDICE DE WHIPPLE

Este índice mide la declaración de la edad, respecto a la preferencia hacia ciertos dígitos (0 y 5); su rango de variación se extiende desde un mínimo de 100 hasta un máximo de 500. El valor inferior es indicativo de que no existe atracción por los dígitos y que se tiene una buena declaración de la edad; por el contrario, el valor máximo se obtiene cuando todas las edades han sido declaradas en dígitos terminados en 0 ó 5.

La escala de valores adoptada para este índice, es la siguiente:

Valor del índice	Calidad de los datos
100 a 105	Muy precisos
105 a 110	Relativamente precisos
110 a 125	Datos aproximados
125 a 175	Datos malos
175 y más	Datos muy malos

Los datos provenientes de distintas fuentes de información demográfica, permiten visualizar el comportamiento de estos índices en el país; los valores obtenidos en las aplicaciones correspondientes son los siguientes:

FUENTE DE DATOS	ÍNDICE DE MYERS	ÍNDICE DE WHIPPLE
CENSO de 1971	27.8	185.2
ESDENIC 1985	15.5	143.5
EMNV 1993	7.9	116.4
CENSO de 1995	7.8	117.9
EMNV 1998	6.0	111.3

Son evidentes los logros alcanzados en el mejoramiento de la declaración de la edad, aspecto que está relacionado íntimamente con los progresos en el campo educativo. De acuerdo con el índice de Myers, se pasó de un nivel de atracción "Alto" (27.8), a uno "Mediano" (6.0), pero cerca del límite superior, para llegar al nivel "Bajo" (5.0); con respecto al índice de Whipple, de unos datos calificados como "Muy malos", a partir del Censo de 1971, se llega a una información calificada de "Aproximada" en la EMNV 1998, (con un valor que se acerca a una calificación de "Aproximadamente precisos").

Los resultados mostrados corresponden a las aplicaciones efectuadas en la población de ambos sexos. Pero, también se investigaron estos índices considerando por separado a hombres y mujeres; es decir, se determinó como es la declaración de la edad, en cada uno de ellos. Estas aplicaciones realizadas dieron índices de Myers y de Whipple que no ofrecen diferencias significativas entre sexos, la calidad de los datos, tiene un comportamiento similar al observado anteriormente (índices a partir de la población de ambos sexos).

Esta presentación de índices que miden la declaración de la edad, permite afirmar que los datos provenientes de fuentes de información más cercanas en el tiempo, son de mejor calidad; dentro de esta calificación se encuentran las Encuestas de Medición de Nivel de Vida de 1993 y 1998. También es claro que aún es posible obtener datos más precisos, en las investigaciones futuras.

Esta nota, relativa a la calidad de los datos, se anexó en el Informe Comparativo de las Encuestas de Medición de Vida de 1993 y 1998²; ahora es posible incorporar los índices correspondientes a la EMNV 2001. A continuación se muestran estos índices, comparándolos con las encuestas similares efectuadas.

FUENTE DE DATOS	ÍNDICE DE MYERS	ÍNDICE DE WHIPPLE
EMNV 1993	7.9	116.4
EMNV 1998	6.0	111.3
EMNV 2001	4.5	103.3

Resulta grato observar los logros alcanzados, a índices que ya eran relativamente bajos, en las encuestas anteriores, se incorporan los de la EMNV- 2001 que se ubican en la parte superior de las escalas propuestas, tanto para el índice de Myres, como para el de Whipple; es decir, un nivel de atracción de los dígitos "bajo", y datos "muy precisos", respectivamente.

Esta mejoría en la declaración de la edad, está vinculada estrechamente a la cobertura, cada vez más elevada, del proceso de cedulaación impulsado por el Consejo Supremo Electoral; que posibilita a los encuestadores, solicitar el documento de identidad a la mayor parte de personas; lo cual permite captar de

² INEC. Informe Comparativo. Encuesta Nacional de Hogares sobre Medición de Nivel de Vida, 1993 y 1998. Septiembre, 2001.

manera más adecuada la edad. Sin embargo, persisten edades que son aproximadas, sobre todo en casos de personas de edad avanzada; también está la circunstancia de los menores de 16 años, que no disponen de cédula de identidad, pero en ellos es más fácil conocer la edad.

Se planteó anteriormente, la posibilidad de contar con datos más precisos, todavía queda espacio para ello; pero se ha llegado a una situación, donde los descensos de los índices van a ser necesariamente menores; más aún, se puede considerar como un logro que en las encuestas futuras, se mantengan los valores obtenidos en la EMNV 2001.