

La declaración de la edad Un análisis comparativo de su calidad en los censos generales de población y vivienda

*Rodrigo Pimienta Lastra**

*Marta Vera Bolaños***

33



1999

La colección *Documentos de Investigación* difunde los avances de trabajo realizados por investigadores de El Colegio Mexiquense, A.C., con la idea de que los autores reciban comentarios antes de la publicación definitiva de sus textos. Se agradecerá que los comentarios se hagan llegar directamente al (los) autor(es). Los editores han mantenido fielmente el texto original del presente documento, por lo que tanto el contenido como el estilo y la redacción son responsabilidad exclusiva del(de los) autor(es). D.R. © El Colegio Mexiquense, A.C., Ex hacienda Santa Cruz de los Patos, Zinacantepec, México. Teléfonos: (72) 18-01-00 y 18-00-56; fax: 18-03-58; E-mail: ui@cmq.colmex.mx Correspondencia: Apartado postal 48-D, Toluca 50120, México.

***Profesor-Investigador del Area de Clases Sociales y Estructura Agraria, UAM-Xochimilco.**

E-mail: plrd6334@cueyatl.uam.mx

**** Investigadora de El Colegio Mexiquense. E-mail: vera@cmq.colmex.mx.**

1. INTRODUCCIÓN

En un país como México donde las estadísticas en muchos casos no son suficientemente confiables, la evaluación de su calidad y su corrección se convierten en una tarea fundamental, previa al análisis de cualquier fenómeno. Esto con el fin de identificar la magnitud, origen y sentido de los errores que contienen, así como las incompatibilidades temporales y espaciales, para hacer las modificaciones y los ajustes necesarios que respondan a los requerimientos analíticos del trabajo que se pretende realizar.

En el caso particular de la variable edad, utilizada en un gran número de investigaciones de tipo aplicado, se ha encontrado que presenta serias deficiencias, debido principalmente a la mala declaración que hacen las personas cuando se les interroga sobre ella.

Con el fin de detectar este tipo de errores se han desarrollado diferentes técnicas, con diversos grados de eficiencia, que permiten evaluar la calidad de esta variable cuando existe duda sobre su confiabilidad. En este trabajo se hace el diagnóstico de la información censal, por edad y sexo de la población, del período de 1930 a 1990, así como del primer recuento censal realizado en el año de 1995.

2. ERRORES Y OMISIONES EN LOS DATOS CENSALES DE POBLACIÓN

Al examinar la calidad de los datos censales de población por edad, pueden encontrarse tres tipos de errores: edad ignorada, omisión diferencial por edad y mala declaración. Los que se presentan con mayor frecuencia son los dos últimos, siendo el más relevante el segundo de éstos, principalmente cuando se trata de ancianos y niños; las edades de los ancianos -hombres y mujeres- con frecuencia se manifiestan más altas que la real y en los niños son mal declaradas o bien no se registran.

La edad ignorada corresponde a aquellas personas para las cuales no se tiene información precisa de sus años, en otros términos, es un problema de *no respuesta* el cual en los censos mexicanos se presenta como una categoría aparte, la de *no especificado*. En este caso el Manual III de Naciones Unidas (1956, p. 16), recomienda que los datos sean prorrateados por sexo en cada una de las edades, aunque en la práctica algunos investigadores suelen eliminar tal cantidad. En ambos casos, en general el análisis no se ve afectado.

La omisión diferencial por edad, es un error de cobertura difícil de detectar. Se presenta cuando no se ha enumerado, por diferentes razones, a una o más personas de las unidades sujetas a investigación, por lo que no se llega a tener datos del total de la población. Un ejemplo de esto es el

grupo de edad 0-4 años, el cual tiene graves problemas de omisión como consecuencia del subregistro.

Si la cobertura mejora de un censo a otro, es posible observar un crecimiento ficticio de la población, debido a que este tipo de error no provoca una distorsión sistemática en la estructura por edad y sexo de la población, sino que afecta su distribución relativa. Si la omisión ocurre en una población que no tiene grandes irregularidades, se puede detectar fácilmente al verse perturbada su estructura por edad, aunque en una población con este tipo de problemas es difícil establecer si se debe a esta causa o a su mala declaración. Una forma adecuada de examinarlo es comparando datos censales con otras fuentes o bien con el *índice de masculinidad* por edad (definido posteriormente en la ecuación 5) el cual establece la relación de los efectivos masculinos de una cohorte con los efectivos femeninos de la misma.

La mala declaración de la edad se refiere a la manifestación de una edad equivocada, voluntaria o involuntariamente, por parte del informante. Es un fenómeno que ocurre en mayor o menor grado en todos los países del mundo, aunque en los denominados en desarrollo tal situación se acentúa haciendo imposible el uso inmediato de la información. Para su estudio se clasifica en tres categorías: preferencia de dígitos, traslado de edades y preferencia o rechazo de una edad específica.

La preferencia de dígitos es un fenómeno ligado a aspectos culturales de cada región o país, se refiere al redondeo sistemático en el dígito final que hacen las personas al declarar su edad. La tendencia general es declarar edades terminadas en cero y cinco, aunque en este último caso es un poco menor, lo que provoca un aumento importante de personas con edades terminadas en las cifras citadas, en perjuicio de las restantes. La atracción de las edades redondeadas con frecuencia puede determinarse trazando la gráfica de las relaciones de masculinidad por edad, sin perder de vista la selectividad de la migración por sexo y edad que podría estar influyendo en esto.

En el traslado de edades ésta se declara sistemáticamente por debajo o por arriba de la ver-

dadera; por ejemplo, las mujeres a partir de cierta edad se declaran más jóvenes, los ancianos y las adolescentes -principalmente cuando estas últimas son casadas y cuentan con varios hijos- tienden a manifestar una edad mayor a la que tienen. Este error además de ser más complicado de detectar y cuantificar es de consecuencias más graves, porque modifica la estructura y el promedio de edad de la población, al provocar un rejuvenecimiento o envejecimiento de algunos de sus grupos de edad. Nótese que la preferencia al dígito también es un traslado pero entre edades próximas, mayores o menores, hacia el dígito de atracción.

La preferencia o rechazo de una edad específica también es un fenómeno de tipo cultural. Por diferentes razones la población tiene cierto rechazo por algunas edades y una atracción especial por otras, en forma independiente de la preferencia de dígitos. Esto se puede observar en la edad de 60 años, la cual es de alta predilección de las personas mayores, o bien, las edades de 13 y 41 años que son de reconocido rechazo.

3. EVALUACIÓN DE LOS DATOS

La evaluación de la declaración de la edad es un proceso de estudio y análisis a través de métodos gráficos y analíticos, para emitir un juicio sobre su calidad. Con ellos se detectan la magnitud, origen y sentido de la omisión, así como las fallas e inconsistencias que presentan los datos, con el fin de establecer los mecanismos adecuados para su corrección.

En cuanto a las técnicas de evaluación de tipo analítico, algunas de las cuales van acompañadas de su respectivo método de corrección, se pueden mencionar los índices de Whipple (Naciones Unidas, 1955, p. 45), Myers (Myers, 1960, pp. 45-46) y Naciones Unidas (Naciones Unidas, *op. cit.* pp. 47-49), que son las más comúnmente usadas y sobre las que se tiene una amplia experiencia en su aplicación e interpretación. Además de éstas existen otros métodos de evaluación y ajuste como el de poblaciones estándar de Coale y Demeney (1983), los indicadores de Michalup (1950), Carrier (1959), Zelnik (1961), Ramachan-

dran (1967), Bachi (1973) y Das Gupta (1975); algunos de ellos menos usuales en su aplicación y más complicados en sus cálculos e interpretación.

En este trabajo para el tratamiento de los datos censales por sexo y edad de la población total del país se utilizaron para evaluar, además de los métodos gráficos, los índices de Whipple, Myers y Naciones Unidas. En el caso del índice de Myers el análisis se hizo con diferentes rangos de edades desplegadas, con el fin de visualizar los cambios del indicador. El de Naciones Unidas se obtuvo tanto con edades desplegadas como con grupos quinquenales, de la misma manera que el anterior, utilizando diferentes rangos de éstas.

3.1 Evaluación Gráfica

El redondeo en la edad se puede detectar con la pirámide de edades desplegadas y a través de algunos indicadores. Aunque la revisión visual de la pirámide mencionada permite detectar en forma clara irregularidades de la estructura por edad de la población y localizar los rangos de las edades afectadas, resulta útil disponer de índices resumidos.

Para el análisis gráfico de la mala declaración de la edad, se han construido las pirámides de edades que presenta la población del país en los últimos cinco de los siete censos considerados y la del conteo de población y vivienda de 1995, con el fin de detectar las preferencias y el rechazo de dígitos. Este tipo de análisis no es posible realizarlo en los censos de 1930 y 1940 por no haberse publicado la información de edades desplegadas.

Como un primer señalamiento sobre los censos considerados, se observa que la atracción y el rechazo se muestran en forma sistemática para las mismas edades durante todo el período, y casi con la misma intensidad en hombres y mujeres, aunque en forma un poco más marcada en éstas últimas, situación que se comprueba más adelante con los indicadores que evalúan la calidad de los datos, Gráficas 1 a 6 del apéndice.

En las gráficas citadas se observa que a partir de los 20 años, en términos generales en los años de 1950 a 1995, la preferencia al dígito cero es la más acentuada tanto en hombres como en mujeres, en detrimento de las edades terminadas principalmente en uno. Le sigue en importancia la atracción al cinco, ocho y dos respectivamente. En las edades de 21, 31, 41 y 51 años se observa un fuerte rechazo, el cual se acentúa en la tercera de las edades referidas. Es importante señalar que en la base de las pirámides descritas, a pesar de notarse ciertas irregularidades de los 2 a los 9 años, no se aprecia una marcada preferencia por algún dígito en particular.

A partir de 1970, aunque hay una mejora sustancial en la declaración de la edad, mantienen su hegemonía como dígitos de atracción el cero y el cinco, conservando sus niveles el ocho y el dos, situación que se repite en 1980. En este último período la base de la pirámide se empieza a hacer más estrecha, como consecuencia de las políticas de población implantadas a principios de los años setentas.

En los datos de 1990 y 1995 se observa un comportamiento muy parecido, las pirámides de edades continúan con la reducción de su base, el dígito de mayor atracción sigue siendo el cero, el cinco pierde su importancia relativa -como dígito de atracción- siendo superado por el ocho y prácticamente igualado por el dos.

Resumiendo, se puede establecer lo siguiente: El cero es un dígito de atracción que se mantiene en todo el período considerado, el cinco lo hace en los cuatro primeros censos y es desplazado por el ocho en los últimos dos. El dos conserva su importancia relativa a lo largo del período de estudio aunque en los dos últimos censos casi iguala al cinco. En todos los censos hay un marcado rechazo por el dígito uno y una amplia desproporción entre las edades de cero y un año, consecuencia del subregistro de nacimientos en la primera edad.

3.2 Índices de Evaluación

Una vez hecho el análisis gráfico de los datos es conveniente ampliarlo y reforzarlo a través de algunas de las técnicas analíticas existentes. En este caso particular se utilizaron los índices de Whipple, Myers y Naciones Unidas. Las dos primeras basadas en la distribución desplegada de las edades y la última en grupos quinquenales de edad.

a) Índice de Whipple

Este indicador, denotado por **IW**, tiene por objeto medir el nivel de atracción que ejercen uno o dos dígitos en las edades. Aunque la propuesta inicial de Whipple fue para medir las preferencias en las edades terminadas en cero y cinco conjuntamente, el índice se ha adaptado para medir éstas tanto en uno como en dos dígitos cualesquiera.

Para el caso de las edades terminadas en cero, el método consiste en sumar los efectivos por año de edad y los de edades terminadas en cero, para comparar esta suma con el quinto de la primera. En estos términos el método queda definido por la ecuación (1).

$$IW = \frac{10 \sum_{x=3}^{x=6} P_{10x}}{\sum_{x=23}^{x=62} P_x} 100 \quad (1)$$

donde P_x es la población de edad x .

En ausencia de concentración el índice vale 100, si es inferior a este valor indicaría que hay repulsión de la edad involucrada y si todos los efectivos están concentrados en la edad cero tomaría el valor 1000.

Partiendo del análisis gráfico y las conclusiones sacadas en el apartado anterior, en este trabajo se utilizará la versión más común, la que mide la atracción de los dígitos cero y cinco. En estos términos la ecuación (1), bajo el supuesto de que la población varía en forma lineal dentro

de los grupos de edades 23-27, 28-32, ..., 58-62, toma la forma siguiente:

$$IW = \frac{5 \sum_{x=5}^{x=12} P_{5x}}{\sum_{x=23}^{x=62} P_x} 100 \quad (2)$$

Los valores extremos entre los que se mueve el índice son: 100 en ausencia de concentración y 500 cuando todos los efectivos se encuentran dentro de las edades cero y cinco. Índices inferiores a 100 indican repulsión en la edades terminadas en dichas cifras.

Para calificar la preferencia de dígitos y la calidad de los datos censales con este indicador, Naciones Unidas propone la escala siguiente: valores del índice entre 100 y 105 señalarán datos muy precisos, de 105 a 110 relativamente precisos, de 110 a 125 aproximados, de 125 a 175 malos y de 175 y más, muy malos.

En este trabajo el cálculo del indicador se hizo para edades desplegadas utilizando dos rangos, de 25 a 60 años y de 25 a 80 en el numerador y de 23 a 62 y de 23 a 82 en el denominador de la ecuación (2) respectivamente. En el cuadro 1 se presentan sólo los primeros resultados, ya que en el segundo caso al aumentar la edad, la declaración de ésta y con ello el índice se distorsiona aún más. También se observa en el cuadro que a través del tiempo se va dando una baja sistemática en el índice, como ya se dijo en el apartado anterior, lo que indica que la declaración de la edad y con esto la información han mejorado. También se alcanza a ver que en todos los censos considerados, la declaración de la edad de la mujeres presenta siempre los mayores índices de atracción a los dígitos cero y cinco.

De acuerdo a los resultados obtenidos con la ecuación (2), la información tanto de hombres como de mujeres, se puede clasificar de la manera siguiente: para el censo de 1950 como de *muy mala calidad*; en 1960 los datos de las mujeres mantienen la clasificación anterior, mientras que los hombres pasan a ser clasificados como *malos* únicamente. Para 1970 y 1980 tanto en

CUADRO 1
ÍNDICES DE EVALUACIÓN DE WHIPPLE A NIVEL NACIONAL POR SEXO DE LOS CENSOS DE 1950 A 1995

	1950	1960	1970	1980	1990	1995
HOMBRES	177.05	166.29	142.88	130.52	123.53	128.5
MUJERES	199.79	181.49	152.90	136.07	126.81	130.6

Fuente: Cálculos propios con base en información censal.

hombres como en mujeres la información cae en el rango de *mala calidad*. En el censo de 1990 los indicadores de los dos sexos muestran una mejoría pasando a ser considerados como *datos aproximados*, pero en el conteo de 1995 hay un retroceso en la calidad de la información captada, los datos de ambos géneros nuevamente son considerados como de *mala calidad* (Cuadro 1).

b) Índice de Myers

El índice de Myers (**IM(j)**) es uno de los más fáciles de evaluar, mide la preferencia o repulsión en forma individual de cada dígito, así como el nivel de atracción general, al sumar los valores absolutos de los índices individuales de los diez dígitos. A este último se le denomina índice resumido o combinado (**IRM**).

En su cálculo se utilizan edades simples, sobre las que se debe suponer un comportamiento lineal en su variación. Para su construcción se suman todos los efectivos de las edades que terminan en los dígitos del 0 al 9, primero de los individuos de 10 y más años de edad y posteriormente de los que tienen 20 y más. Estas dos series se ponderan y se agregan una con la otra para llegar a una población ficticia, cuya distribución debe ser uniforme cuando no existe preferencia por algún dígito, es decir, en una situación de distribución equitativa de las edades se esperaría que en cada dígito la suma mencionada fuera el 10% de la suma total, es decir,

$$IM(j) = \frac{S_{1,j}a_{1,j} + S_{2,j}a_{2,j}}{\sum_{i=0}^9 (S_{1,i}a_{1,i} + S_{2,i}a_{2,i})} 100 = 10 \quad (3)$$

donde

$$j = 0, 1, 2, \dots, 9$$

$$S_{1,j} = \sum_{i=0}^8 P_{(10+j)+10i}$$

$$S_{2,j} = \sum_{i=0}^7 P_{(20+j)+10i}$$

en este caso **S_{1,j}** y **S_{2,j}** son las sumas de las poblaciones de 10 y más años y 20 y más respectivamente, cuyas edades terminan en el dígito **j**; y **P_{(10+j)+10i}** y **P_{(20+j)+10i}** son las poblaciones de 10 y más y 20 y más años. Los valores de los coeficientes **a_{1,j}** y **a_{2,j}** dados por Myers, son los que ponderan las desviaciones de la declaración de cada una de las edades respecto al 10% (ver cuadro 2).

CUADRO 2
COEFICIENTES DEL ÍNDICE DE MYERS

Dígito j	a _{1,j}	a _{2,j}
0	1	9
1	2	8
2	3	7
3	4	6
4	5	5
5	6	4
6	7	3
7	8	2
8	9	1
9	10	0

Fuente: Myers (1960).

Posteriormente, para obtener el índice resumido se suman los valores absolutos de las desviaciones de cada dígito respecto al 10% que debería

haber en cada uno de ellos, bajo el supuesto de uniformidad, esto es,

$$\text{IRM} = \sum_{j=0}^7 |d_j| \quad (4)$$

donde

$$d_j = \text{IM}(j) - 10.0$$

representa las desviaciones de **IM(j)** respecto al 10.0%, en otras palabras, son los índices de atracción cuando d_j es positiva o rechazo es negativa del dígito **j**.

Los valores extremos que puede tomar el indicador **IRM** son cero y 180. El primer caso se tiene cuando no hay preferencia, es decir, si todas las edades fueron declaradas con exactitud las sumas combinadas resultan aproximadamente iguales, por lo que las desviaciones respecto a 100 son insignificantes; aunque se debe tener presente que si no hay preferencia de dígitos, la proporción teórica del 10% no será rigurosamente respetada, pero si hay diferencias importantes en algún dígito en particular estas se reflejarán en el indicador. El segundo se presenta cuando todas las edades declaradas terminan en la misma cifra, lo que indica que el 100% del total de la suma combinada correspondería a un sólo dígito, por lo cual la suma de los valores absolutos de las desviaciones sería igual a 180.

Para interpretar el nivel de atracción Stockwell y Wicks (1974) proponen los rangos siguientes: índices entre 0.0 y 5.0, bajo; entre 5.1 y 15.0, intermedio; entre 15.1 y 30.0, alto; y entre 30.1 y más, muy alto. El método puede adaptarse a otros grupos de edad, por ejemplo: 0 y más con 10 y más años, o bien, 20 y más con 30 y más años.

Finalmente, se puede mencionar como una ventaja de este método la de evitar el sesgo producido por la mortalidad, al considerar que los efectivos de las edades que terminan en cero generalmente son superiores a los de edades posteriores, debido a la forma triangular de la estructura por edad.

Para este trabajo el indicador se calculó relacionando las edades desplegadas en los rangos 10-69 y 20-79 en todos los censos; también se obtuvo para aquellos censos que reportan un ran-

go mayor de edades como son el de 1960, para el cual se tomaron las edades de 10-79 y 20-89 y en los de 1980 y 1990 así como en el conteo de población de 1995 donde fueron considerados rangos de edades de 10-89 y 20-99. En el cuadro 3 se presentan los resultados correspondientes a los primeros cálculos mencionados, los otros no se reportan ya que en todos los casos los índices fueron superiores a los primeros, hecho que prueba una vez más la gran distorsión que se da en la declaración de las edades más avanzadas. Adicionalmente, en el Cuadro 1 del Apéndice se presentan los índices de atracción y rechazo para cada uno de los dígitos de las fuentes de información mencionadas.

En el cuadro 3 puede apreciarse, a través del índice combinado de Myers, que ha habido una mejora sistemática en la declaración de la edad a nivel nacional de los censos considerados, de ser clasificada entre *muy alta* y *alta* atracción, en los dos primeros, pasó a un nivel *intermedio* de atracción en los dos últimos. En el conteo de 1995 ocurrió algo inesperado, el índice subió a 12.9, comparado con el indicador censal de 1990 que era de 9.88.

CUADRO 3
ÍNDICES DE EVALUACIÓN DE MYERS A NIVEL NACIONAL POR SEXO DE LOS CENSOS DE 1950 A 1990

	1950	1960	1970	1980	1990	1995
HOMBRES	27.29	23.20	15.45	12.09	9.88	12.9
MUJERES	35.65	28.76	18.62	13.77	10.84	13.4

Fuente: Cálculos propios con base en información censal.

Analizando la atracción y el rechazo de cada dígito en particular a lo largo del período de estudio tanto hombres como mujeres presentaron situaciones similares, sólo que en estas últimas la mala declaración es más acentuada, Cuadro 1 del apéndice. El dígito de mayor atracción -como se había visto en el análisis gráfico- en ambos sexos es el cero, seguido por el cinco y el ocho, aunque este último en menor grado. Los de mayor rechazo son el uno, el nueve y el siete en el caso

particular de las mujeres y cuando se trata de los hombres el nueve y el cuatro.

Finalmente es importante hacer notar que en los diferentes censos estudiados se va dando una pequeña reducción en los índices de atracción y rechazo, lo cual indica que la declaración de la edad y con ello la información censal había ido mejorando sistemáticamente, hasta antes del recuento censal de 1995 donde se tuvo un pequeño repunte en la mala declaración. En este último la atracción al cero, cinco y ocho se presenta más acentuada así como el rechazo al nueve, siete y cuatro.

c) Índice de Naciones Unidas

Este es un indicador que además de mostrar la mala declaración de la edad y la preferencia de dígitos, presenta la omisión diferencial de individuos que se da en ciertas edades y con ello las irregularidades que presentan los datos por sexo y edad.

El índice puede ser calculado en poblaciones clasificadas por sexo tanto para edades desplegadas (Corona *et al.* 1978; pp. 78-79), como para grupos quinquenales y decenales (Naciones Unidas, 1955; p. 47). Su evaluación tiene como base la regularidad de sexos y los cocientes de edades de cada uno de éstos. Su obtención se divide en tres etapas: regularidad de sexos, cociente de edades e índice resumen, cada una con diferentes posibilidades de análisis.

En este trabajo el índice de Naciones Unidas se obtuvo para edades desplegadas y en grupos quinquenales. En el primer caso, para los rangos de 0-64 y 0-84 años; en el segundo para los grupos de edad de 0-4 a 65-69 y del 0-4 al 84-89. Se reportan los resultados que se obtuvieron para grupos quinquenales en el rango de edades de 0-4 a 65-69, para los censos de 1930 a 1990 (véase Cuadro 4). En forma adicional en el apéndice se presentan los cocientes de edades para hombres y mujeres (véase Cuadro 2) y los índices de masculinidad (véase Cuadro 3).

i) Regularidad de Sexos

La regularidad de sexos se mide a partir de las diferencias sucesivas del índice de masculinidad de la población observada, entre un grupo de edad y el siguiente, ecuaciones (6), (7) y (8). El cálculo del índice de masculinidad es:

$${}_4IM_x = \frac{{}_4P_x^m}{{}_4P_x^f} 100 \quad (5)$$

$$x = 0,5,\dots,75$$

donde ${}_4IM_x$ es el índice de masculinidad del grupo de edad $x, x+4$, ${}_4P_x^m$ y ${}_4P_x^f$ son las poblaciones masculina y femenina en edades cumplidas $x, x+4$ respectivamente.

La medida resumen de este primer indicador está dada por el promedio de la suma de valores absolutos de las primeras diferencias (ver ecuación 6).

$${}_4RS_x = \frac{\sum_{i=1}^{k-2} |{}_4\Delta IM_{i,x}|}{k-2} \quad (6)$$

$$x = 5,9,\dots,75$$

$$i = 1,2,\dots,k-2$$

donde ${}_4RS_x$ es el índice de regularidad de sexos del grupo de edad $x, x+4$, ${}_4\Delta IM_{i,x}$ son los valores absolutos de las diferencias sucesivas del índice de masculinidad y k es el número de intervalos de edad considerados, en este caso 17. Las primeras diferencias de cada grupo de edad están definidas en la ecuación (7).

$${}_4\Delta IM_x = {}_4P_{x+5} - {}_4P_x \quad (7)$$

$$x = 5,10,\dots,75$$

Bajo condiciones regulares del comportamiento demográfico, se espera una evolución gradual de los valores absolutos de estas diferencias.

Valores de RS cercanos a 1.0 indican información de buena calidad mientras que mayores a 15 señalan datos muy deficientes.

ii) Cocientes de Edades

En los cocientes de edades se relacionan los efectivos de un grupo de edad con la media aritmética de los dos grupos adyacentes, esto es, el total de personas de un determinado grupo de edad, entre la media aritmética de los dos grupos contiguos (véase ecuación 8).

$${}_4CE_x = \frac{{}_4P_x}{1/2 ({}_4P_{x-5} + {}_4P_{x+5})} - 100 \quad (8)$$

$$x = 5, 10, \dots, 75$$

donde ${}_4CE_x$ es el cociente de edades cumplidas del grupo $x, x+4$; ${}_4P_x$ es la población en edades cumplidas $x, x+4$ y ${}_4P_{x-5}$ y ${}_4P_{x+5}$ son las poblaciones a edades $x-5, x-5+4$ y $x+5, x+5+4$ respectivamente.

El supuesto del modelo establece que la población real decrece al aumentar la edad, en progresión aritmética de primer grado, esto es:

$$\begin{aligned} {}_4P_{x-5} &= {}_4P_o + (x-5)d \\ {}_4P_{x-5} &= {}_4P_o + xd \end{aligned} \quad (9)$$

$${}_4P_{x+5} = {}_4P_o + (x+5)d$$

$$x = 5, 10, \dots, 75$$

De las ecuaciones (9) y (10) se desprende que:

$$\begin{aligned} {}_4CE_x &= \frac{2 \cdot {}_4P_x}{{}_4P_{x-5} + {}_4P_{x+5}} - 100 \\ &= \frac{2({}_4P_o + xd)}{2({}_4P_o + xd)} - 100 \\ &= 100 \\ x &= 5, 10, \dots, 75 \end{aligned} \quad (10)$$

Calculando las desviaciones del indicador respecto a 100 se mide la calidad de la información (ver ecuación 11). Si la hipótesis se cumple el cociente de edades debe ser 100 y las desviaciones próximas a cero, en caso contrario las perturbaciones pueden haber sido causadas por cambios en la fecundidad o por movimientos migratorios. Desviaciones positivas indican exceso de personas y negativas un déficit.

$${}_4DC_x = {}_4CE_x - 100 \quad (11)$$

$$x = 5, 10, \dots, 75$$

Una medida resumen del indicador es la suma de los valores absolutos de las desviaciones entre el número de desviaciones calculadas (véase ecuación 12).

$$PDC = \frac{\sum_{i=1}^{k-2} |{}_4DC_{i,x}|}{k-2} \quad (12)$$

$$x = 5, 10, \dots, 75$$

$$i = 1, 2, \dots, k-2$$

En información de buena calidad el promedio de las desviaciones calculado es un valor próximo a cero y si es de muy mala calidad a 30.

iii) Indice Resumen

Combinando los dos índices anteriores se obtiene el de Naciones Unidas:

$$INU = 3 \cdot RS + PDC^H + PDC^M \quad (13)$$

Donde **RS** es el índice de regularidad de sexos y **PDC^H** y **PDC^M** representan los promedios de desviaciones respecto a 100, de las poblaciones masculina y femenina respectivamente.

La mayor ponderación de este indicador está en **RS** con un coeficiente de tres, ya que los índices de regularidad en cada edad son más sensibles a perturbaciones en la fecundidad, movimientos migratorios o alguna otra causa.

Debido a que el indicador no tiene límites precisos su interpretación se basa en la experiencia, para lo cual se establece que índices mayores de 40 indican información deficiente, entre 20 y 40 de calidad intermedia y menores a 20 satisfactoria.

Considerando lo anterior, únicamente se analizarán los resultados obtenidos del INU para grupos quinquenales de edad en los rangos de 0-4 a 65-69 años. El valor del índice así como de cada uno de sus tres componentes se presentan en el Cuadro 4 de esta sección.

CUADRO 4

INDICES DE EVALUACION DE NACIONES UNIDAS A NIVEL NACIONAL DE LOS CENSOS DE 1930 A 1990 DE LOS GRUPOS DE EDAD 0-4 A 65-69

	1930	1940	1950	1960	1970	1980	1990	1995
NACIONAL								
INU	56.01	33.76	36.91	26.83	21.74	13.67	12.26	14.47
COC. EDADES								
M.	13.25	11.51	10.68	9.12	6.92	3.58	3.77	4.43
COC. EDADES								
F.	19.14	12.43	12.07	9.59	5.91	3.47	3.63	4.36
PROM. 1a.								
DIF.	7.87	3.28	4.72	2.71	2.97	2.21	1.62	1.89

Fuente: Cálculos propios con base en información censal.

Como se observó en los indicadores calculados anteriormente, en éste ocurre lo mismo, todos disminuyen en forma sistemática durante el período de estudio (véase Cuadro 4). En 1930 se ubica el indicador más alto donde la información se clasifica como **deficiente**, entre 1940 y 1970 se le puede clasificar como de **calidad intermedia** y de 1980 en adelante ya es considerada como **satisfactoria**. En 1995 se incrementa ligeramente el valor de este índice, señalando un pequeño retroceso en la calidad de los datos obtenidos.

Finalmente deben hacerse dos anotaciones: la primera es que al agrupar los datos quinquenalmente se introduce una corrección en la mala declaración de la edad; y segunda, que las perturbaciones en algunos casos pueden ser reales y no causadas por errores de declaración.

4. CONCLUSIONES

Del análisis realizado se desprende que si bien el estudio de los datos a través de técnicas analíticas es importante, una inspección previa a través de gráficas resulta de gran utilidad para detectar un buen número de inconsistencias de manera fácil y rápida.

De los resultados que se obtuvieron vale la pena comentar el hecho de que la mala declaración de la edad aumenta con el envejecimiento de las personas, así como la permanencia en edades avanzadas del cero como principal dígito de atracción.

En el caso de las mujeres la mala declaración de la edad en general es mayor que en los hombres. Una explicación de esto podría ser que el traslado de edades es muy acentuado y va más allá de las edades próximas a la verdadera, situación que se debe tener presente cuando los datos van a ser corregidos a través de algún método.

Finalmente se considera importante resaltar la mejora sistemática que se da en la calidad de la información en el período de 1930 a 1990 y el tropiezo que se tuvo en 1995, lo que invita a establecer alguna explicación sobre las causas de tal evento. Al respecto se puede establecer que esto no es causa de la mala declaración por parte del entrevistado sino al llenado del cuestionario por parte del entrevistador, el cual probablemente contestó la cédula de la entrevista en ausencia del informante. Aunque sobre esto habría que profundizar para que no vaya a repetirse en el censo de población del año 2000.

APÉNDICE

CUADRO 1
ÍNDICES DE ATRACCIÓN Y REPULSIÓN DE DÍ-
GITOS DE MYERS A NIVEL NACIONAL POR
SEXO, 1950-1995

DIGITO	1950	1960	1970	1980	1990	1995
HOMBRES						
0	7.75	6.71	4.10	3.27	2.48	2.88
1	-4.62	-4.05	-3.22	-2.72	-2.71	-2.49
2	0.06	0.13	0.45	0.45	0.46	0.78
3	-1.90	-1.32	-0.86	-0.48	-0.38	-0.28
4	-2.08	-1.56	-0.95	-0.60	-0.32	-0.62
5	4.28	3.34	2.36	1.49	1.23	1.81
6	-1.00	-1.10	-0.84	-0.48	-0.30	-0.80
7	-1.80	-1.52	-0.98	-0.74	-0.47	-0.85
8	1.56	1.41	0.81	0.84	0.77	1.00
9	-2.25	-2.05	-0.87	-1.03	-0.77	-1.43
MUJERES						
0	10.21	8.46	5.33	3.96	2.97	3.27
1	-5.35	-4.60	-3.63	-3.00	-2.84	-2.64
2	-0.52	-0.37	-0.03	0.13	0.26	0.55
3	-2.55	-1.80	-1.13	-0.66	-0.43	-0.34
4	-2.31	-1.73	-1.04	-0.61	-0.28	-0.65
5	5.67	4.37	2.95	1.83	1.41	1.90
6	-1.48	-1.31	-1.09	-0.61	-0.41	-0.75
7	-2.69	-2.17	-1.35	-0.99	-0.71	-0.99
8	1.93	1.56	1.03	0.96	0.77	0.99
9	-2.92	-2.40	-1.04	-1.00	-0.77	-1.34

Fuente: Cálculos propios.

CUADRO 2
COCIENTES DE EDADES A NIVEL NACIONAL DE
HOMBRES Y MUJERES POR GRUPOS DE EDAD
0-4 ... 65-69, DE 1930 A 1995

EDAD	1930	1940	1950	1960	1970	1980	1990	1995
HOMBRES								
0 - 4								
5 - 9	108.7	106.9	103.6	104.7	106.0	111.6	102.8	101.6
10 - 14	90.0	103.4	102.8	100.5	101.8	102.3	103.6	102.6
15 - 19	98.0	97.6	93.7	95.6	95.8	99.8	106.1	101.0
20 - 24	99.3	85.9	95.7	95.7	94.9	97.6	95.7	104.6
25 - 29	107.6	109.5	111.2	99.1	98.0	95.7	96.6	95.0
30 - 34	91.4	89.1	80.8	93.6	91.5	94.5	98.0	97.6
35 - 39	108.5	123.7	116.3	114.0	110.1	102.6	103.2	105.3
40 - 44	93.3	87.1	91.5	85.9	92.9	97.1	93.1	95.2
45 - 49	98.8	98.8	107.8	101.6	107.1	99.9	101.3	98.2
50 - 54	106.9	100.1	101.8	103.8	88.6	97.8	98.0	99.7
55 - 59	74.2	84.0	78.0	90.1	96.4	100.7	95.1	92.3
60 - 64	146.0	129.3	124.5	122.2	106.5	94.3	103.6	105.8
65 - 69	70.7	78.4	87.0	76.3	99.7	94.7	97.5	92.9
70 - 74	110.7	101.8	99.8	109.5	104.1	105.0	93.2	105.1
75 - 79	73.8	81.3	81.2	83.2	74.1	97.0	96.8	88.8
80 - 84								
MUJERES								
0 - 4								
5 - 9	109.9	107.9	104.0	105.2	106.1	111.5	102.5	101.6
10 - 14	79.7	95.8	94.6	96.4	98.4	100.4	101.9	100.6
15 - 19	108.4	104.6	100.9	98.0	98.1	101.0	106.0	101.1
20 - 24	101.1	86.6	101.8	99.3	99.0	99.9	99.1	107.0
25 - 29	108.7	112.5	105.6	101.3	98.7	96.6	97.2	95.8
30 - 34	89.2	88.9	79.8	91.8	88.5	92.5	98.2	97.8
35 - 39	104.9	119.6	117.8	111.2	111.7	104.4	102.9	106.3
40 - 44	100.5	89.0	93.1	86.7	93.5	94.8	92.2	93.0
45 - 49	89.6	98.3	103.1	101.9	102.4	101.1	100.5	98.9
50 - 54	119.8	103.1	105.0	105.3	91.4	99.4	98.8	98.6
55 - 59	66.6	82.6	74.9	86.9	95.5	96.2	94.1	92.3
60 - 64	160.0	128.3	132.5	123.1	107.6	96.2	105.7	106.5
65 - 69	62.8	77.2	81.4	77.4	100.2	97.6	96.7	95.7
70 - 74	125.0	105.0	108.1	112.0	100.5	103.1	93.2	101.6
75 - 79	63.2	76.2	73.8	79.7	76.8	96.3	95.6	87.3
80 - 84								

Fuente: cálculos propios

CUADRO 3
 ÍNDICES DE MASCULINIDAD A NIVEL NACIONAL POR GRUPOS DE EDAD 0-4 ... 65-69, DE 1930 A 1995

EDAD	1930	1940	1950	1960	1970	1980	1990	1995
0 - 4	102.2	102.3	101.5	103.4	103.4	101.1	102.5	103.3
5 - 9	103.9	103.9	103.1	103.6	103.9	101.2	102.2	103.1
10 - 14	109.7	107.9	105.9	105.2	104.7	101.2	101.4	102.6
15 - 19	88.8	94.4	90.2	96.8	97.2	96.8	97.1	98.1
20 - 24	87.2	91.5	86.5	91.1	91.8	93.4	91.4	93.4
25 - 29	88.9	89.6	94.6	91.4	93.5	93.8	91.0	92.2
30 - 34	93.4	92.7	95.4	96.8	98.1	96.6	91.8	92.4
35 - 39	94.7	95.7	93.7	99.8	96.8	95.5	93.3	93.0
40 - 44	88.9	92.3	94.3	98.1	98.5	98.1	95.1	96.1
45 - 49	97.7	91.8	99.2	98.0	102.8	96.1	95.6	95.4
50 - 54	88.0	89.8	95.8	98.4	97.9	96.0	94.3	96.0
55 - 59	100.4	93.6	98.0	102.7	98.3	99.9	94.2	94.3
60 - 64	93.0	95.3	91.8	99.8	96.6	94.5	91.5	91.8
65 - 69	104.9	96.4	97.2	96.6	96.7	91.0	92.1	89.6
70 - 74	93.4	93.9	89.3	93.7	98.3	92.7	91.0	93.3
75 - 79	106.9	98.0	95.1	94.3	89.9	90.9	88.8	91.1
80 - 84								

Fuente: cálculos propios

BIBLIOGRAFÍA

- Bachi, Roberto; 1973. *Medición de la Tendencia a Preferir Ciertos Dígitos en la Declaración de Edades*. CELADE, Serie DS, Núm. 16. San José de Costa Rica. Traducción de *Measurement of the Tendency to Round off Age Returns del Bulletin de l'Institut International de Statistique*; Proceeding of the 28 Session, Rome Sept. 6-12, 1953; 34(3) pp. 129-138.
- Bocaz, Albino; 1974. *El Uso de la Ley de Makeham como Función Demográfica*. En *Notas de Población*, Año II, Vol. 6; CELADE. Santiago de Chile.
- Brass, William, 1973. *Ajuste e Interpretación de Datos Demográficos*; Trad. *Disciplining Demographic Data*. CELADE, Serie DS, Núm. 8. San José de Costa Rica.
- Carrier, Norman H; 1959. *A Note on the Measurement of Digital Preference in Age Recording*. In *Journal of the Institute of Actuaries*, 85. Cambridge, Eng. pp. 71-85.
- Chackiel, Juan y Guillermo Macció, 1978. *Evaluación y Corrección de Datos Demográficos*, Serie B, No. 39. Centro Latinoamericano de Demografía, Santiago de Chile.
- Coale, Ansley J. y P. Demeny, 1983. *Regional Model Life Tables and Stable Populations*. 2nd. Ed. Academic Press, Inc. New York.
- Corona, V. Rodolfo, Alberto Minujin Z. y Gabriel Vera F, 1978. *Manual de Técnicas de Evaluación y Ajuste de Información Estadística*, Serie: Manuales 1. Centro Nacional de Información y Estadísticas del Trabajo, Secretaría del Trabajo y Previsión Social, México.
- Das, Gupta P; 1975. *A General Method of Correlation for Age Misreporting in Census Population*. In *Demography*, 12(2); pp. 303-312. Center for Demography and Ecology, University of Wisconsin. Madison Wisconsin.
- Hobcraft, John, 1977. *Seminario sobre Estimaciones Demográficas*. Del 1o. al 9 de Agosto de 1974, San José de Costa Rica; Trad. de J. L. Somoza; CELADE, Serie D, Núm. 94. Santiago de Chile José de Costa Rica. pp. 99-120.
- INEGI, 1930-1990. *Censos Generales de Población y Vivienda*. Dirección General de Estadística, México.
- Michalup, E; 1950. *Propuesta de un Coeficiente de Exactitud*. En *Estadística*, Vol. VIII, Núm. 26. Washington, D. C. pp. 49-58.
- Myers, Robert J; 1960. *Error and Bias in the Reporting of Ages in Census Data*. En U.S. Bureau of the Census, Washington, D. C. pp. 115-125. Reproducido de *Handbook of Statistical Methods for Demographers, Transactions of the Actuarial Society of America*, 41 Part II (104) pp. 395-415, 1940.
- Naciones Unidas, 1955. *Métodos para Evaluar la Calidad de los Datos Básicos Destinados a los Cálculos de Población*, Manual II, ST/SOA/Serie A, No. 23. Nueva York.
- Naciones Unidas, 1956. *Métodos para Preparar Proyecciones de Población por Sexo y Edad*, Manual III, ST/SOA/Serie A, No. 25. Nueva York.
- Naciones Unidas, 1963. *Modelos de Mortalidad por Sexo y Edad para Países Insuficientemente Desarrollados*. ST/SOA/Serie A, Núm. 22, Nueva York.
- Ramachandran, K. V; 1967. *An Index to Measure Digit Preference Error in Age Data*. In

Summary in the United Nations, World Population Conference, Vol. III. 1965. Belgrade. pp. 202-203.

Spiegelman, Moritmer; 1972. *Introducción a la Demografía*. 1a. ed. en español. Fondo de Cultura Económica. México.

Stockewell, Edward y H. Wicks, 1974. *Age Heaping in Recent National Censuses*. En *Social Biology Review*, Vol. 21, No. 2, 1974.

Wunsch, Guillaume, 1992. *Técnicas para el Análisis de Datos Demográficos Deficientes*. Primera edición en español; El Colegio de México, México.

Zelnik, Melvin; 1961. *Age Heaping in the United States Census: 1880-1950*. En *Milbank Memorial Fund Quarterly*, Vol. 39(3); pp. 540-573. New York.