EVALUACION Y CORRECCION DE DATOS DEMOGRAFICOS

VII. Técnicas de corrección y ajuste de la mala declaración de la edad

Juan Chackiel y Guillermo Macció

Serie B, Nº 39
Agosto, 1979
Santiago de Chile

04912.00=No pedido DOCPAL(NACCESO) 1979=Fecha publ.
CHACKIEL, Juan; MACCIO, Guillermo A (Au)
Evaluacion y correccion de datos demograficos. VII. Tecnicas de correccion y ajuste de la mala declaracion de la edad.
Agosto 1979; Pags:26
Editorial: CELADE. Santiago CL
Serie B 39d
Idioma:Es Distr:General Impresion:Mimeo

Pais/region principal:ZZ Paises tratados:ZZ
Descriptores:<ERROR*> <SUAVIZAR UNA DISTRIBUCION*> <DECLARACION DE LA EDAD*> <AJUSTE DE DATOS*>
Fuente datos: <CENSO DE POBLACION>
Proyecto: <CELADE>
Categ. Revista:<POBL:RECOL>
Fechas datos demogr:1950-1973 No. de Ref= 4

Para corregir los errores que afectan a la informacion de la poblacion por edades, se pueden utilizar los siguientes procedimientos: a).procedimientos demograficos, entre los que se encuentran el agrupamiento quinquenal convencional, agrupamiento quinquenal no convencional centrado en 0 y 5, y agrupamiento no convencional propuesto por Hobcraft; estos 2 ultimos, presentan la informacion en grupos de edades poco usuales, pero existen tecnicas para transformarlos en los grupos que normalmente se emplean. Mediante la aplicacion del indice de Myers, y sus supuestos, se puede determinar cual es el mejor de los agrupamientos; b).ajustes graficos, entre los que se cuentan el ajuste de ojivas naturales que consiste en transferir poblacion de un grupo a otro, corrigiendo errores debidos a traslacion de edades o preferencia por algunos de ellos; y el ajuste de ojivas de ejes oblicuos cuya caracteristica es que supera el problema de escala que tienen las ojivas naturales, logrando hacer mas notorias las irregularidades presentes en ciertos grupos de la poblacion bajo estudio; c).ajustes estadistico-matematicos, donde se incluyen aquellas tecnicas basadas en formulas matematicas cuya aplicacion lleva implicita la aceptacion de ciertos supuestos sobre los errores de la informacion, tales como los promedios moviles, promedios moviles con ponderacion y ajuste de la razon. Para cada uno de los procedimientos se discuten sus ventajas y limitaciones, y se presentan ilustraciones de su aplicacion a traves de ejemplos
Notas:Otros capitulos del documento "Evaluacion y correccion de datos demograficos" publicados individualmente, entrados en DOCPAL: Cap.I y II con NACCESO 2945.00, ISIS 7221; Cap.VI con NACCESO 3611.00, ISIS 8732; Cap.VIII con NACCESO 3336.00, ISIS 8082
(Inf. interna para DOCPAL: ISIS=11799 NRES=41-101 LS -m Cfd)

I N D I C E

Indice de cuadros y gráficos

VII. TECNICAS DE CORRECCION Y AJUSTE DE LA MALA DECLARACION DE LA EDAD

En el capítulo anterior se analizaron los errores que afectan a la información de la población por edades (preferencia de dígitos, traslación de la edad y omisión selectiva según la edad). Ahora corresponde encarar la forma de corregir estos datos, para lo que se cuenta con varios tipos de procedimientos que usualmente se agrupan bajo tres categorías principales: ajustes demográficos, ajustes gráficos, ajustes analíticos o estadístico-matemáticos. Esta clasificación obedece principalmente a fines prácticos, más que a un propósito científicamente riguroso. Los distintos procedimientos no son totalmente excluyentes y tampoco puede decirse que sean mejores unos que otros; todo depende del caso concreto que se tiene entre manos.

1. Procedimientos demográficos

El principio fundamental que toma en consideración este grupo de técnicas es el examen cuidadoso de la información sujeta a corrección. A través del comportamiento observado de las variables de trata de detectar el error y hacer la corrección; por lo tanto está muy ligado a los estudios mencionados en la sección VI.4 en que se hizo referencia al examen minucioso de la información con el fin de determinar los patrones de errores que la afectan. Se trata de tomar conciencia de las irregularidades existentes y corregirlas sin forzar demasiado la información original, respetando el principio de "rehabilitación" enunciado en el capítulo I, evitando la aplicación "ciega" de técnicas de ajuste.

En lo que sigue se describen algunos de los procedimientos que responden a las características mencionadas:

1.1 Agrupamiento quinquenal convencional

Consiste en agrupar la población por edades individuales en grupos quinquenales 0-4, 5-9, 10-14, 15-19, etc. No es estrictamente una corrección,

sino más bien una forma de presentación de los datos que a base de compensaciones ayuda a ocultar ciertos defectos. Obsérvese en el cuadro 1 y en el gráfico 1 del capítulo VII, que la población censal de Guatemala de 1973, que en grupos individuales de edades presentaba con mucha claridad los errores de mala declaración, al agruparla en grupos quinquenales se lograba un comportamiento más suave, detectándose errores solamente en algunos grupos muy particulares.

El supuesto básico de este agrupamiento es que la atracción por los dígitos 0 y 5, o mejor dicho que la preferencia de los dígitos de 0 a 4, es la misma que de 5 a 9 en conjunto. En la mayoría de los casos este agrupamiento se usa como paso previo a la aplicación de ajustes mucho más rígidos de tipo estadístico-matêmaticos o gráficos.

Pueden mencionarse como limitaciones más importantes las siguientes: i) probablemente la mayor atracción sea por el dígito 0 y que por lo tanto los grupos de edades que lo contienen (10-14, 20-24, 30-34, etc.) están sobreestimados frente a los grupos restantes (15-19, 25-29, etc.); ii) si existen errores de "traslado de edades" más allá de los grupos, se desfiguraría la distribución de la población y frente a ello nada se lograría con realizar un agrupamiento como el indicado.

1.2 Agrupamiento quinquenal no convencional centrado en 0 y 5

Se forman los grupos 8-12, 13-17, 18-22, 23-27, etc., que suponen que la alta preferencia por las edades terminadas en 0 y 5, no importa cual sea mayor, proviene de los dos dígitos anteriores y de los dos posteriores. En este caso las edades que finalizan en 0 y 5 estarían exageradas pero habría compensación dentro de cada grupo con el rechazo de los dígitos adyacentes.

El presente agrupamiento tropieza con una dificultad, cual es presentar la información en grupos de edades muy poco usuales. Este inconveniente es salvable mediante la transformación de estos grupos no convencionales en los usualmente utilizados (0-5, 10-14, 15-19, 20-24, etc.). Hobcraft[1/] ha propuesto dos formas al respecto: i) a través del uso del método gráfico de "ojiva de ejes oblicuos" que se describe más adelante en este mismo capítulo;

1/ Hobcraft, J., Seminario sobre Estimaciones Demográficas. Sesión II, CELADE, Serie D, Nº 34.

ii) usando un método de interpolación que permita subdividir grupos de edades. En particular propone usar un polinomio de segundo grado para lo cual ha generado una tabla de multiplicadores.[2/] Por ejemplo, si se dispone de los grupos 43-47 y 48-52, el primero se subdivide en 43-44 y 45-47; el segundo en 48-49 y 50-52. Luego combinando 45-47 y 48-49 se compone el grupo convencional 45-49 y así para los demás.

En el cuadro 1 se presenta una aplicación para el caso particular de la población de ambos sexos del censo de Guatemala de 1973 en el tramo de edades 45-74 años. Para obtener los grupos convencionales de edades se usó el procedimiento de interpolación cuadrática.

Cuadro 1

CENSO DE GUATEMALA DE 1973: AJUSTE DE LA POBLACION POR EDADES USANDO EL AGRUPAMIENTO NO CONVENCIONAL CENTRADO EN 0 Y 5

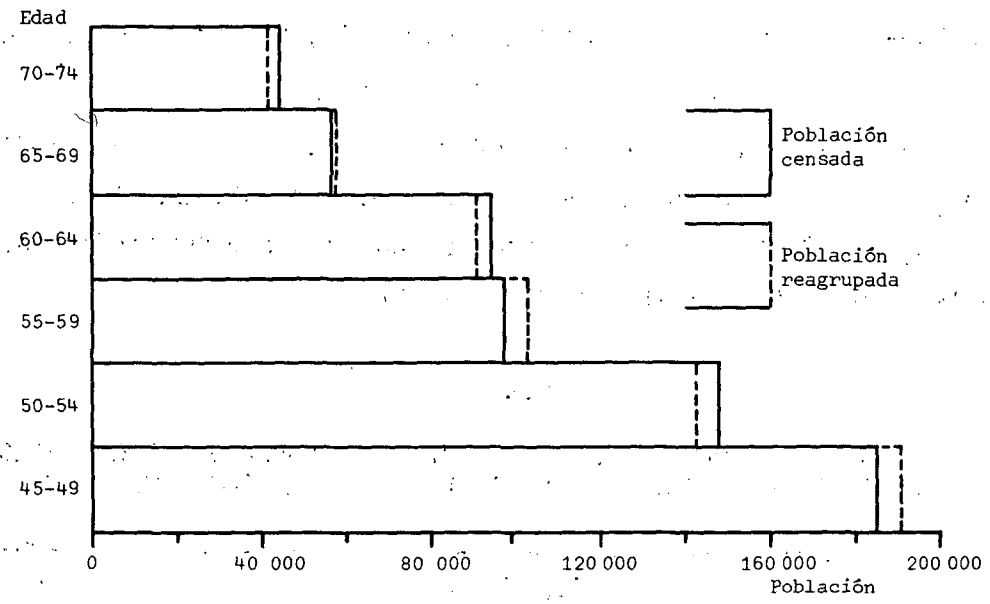
Grupos de edades no convencional	Población	Grupos de edades convencional	Población	
			Reagrupada	Agrupada convencionalmente
43-47	188 247	45-49	191 771	184 502
48-52	167 936	50-54	143 531	147 004
53-57	108 334	55-59	104 874	99 665
58-62	103 712	60-64	89 159	95 026
63-67	65 043	65-69	57 353	56 239
68-72	48 953	70-74	40 939	44 214
73-77	29 445			
78-82	17 623			

Debido a la atracción especial por la edad 60, el grupo 60-64 aparece en los datos con agrupamiento convencional muy próximo al 55-59. El nuevo agrupamiento parece corregir en parte esos defectos, siendo la evolución

2/ Carrier, N. y Hobcraft, J., Estimaciones Demográficas para Sociedades en Desarrollo, CELADE, Serie D, Nº 1026, San José, Costa Rica, 1975, pág. 77.

Gráfico 1

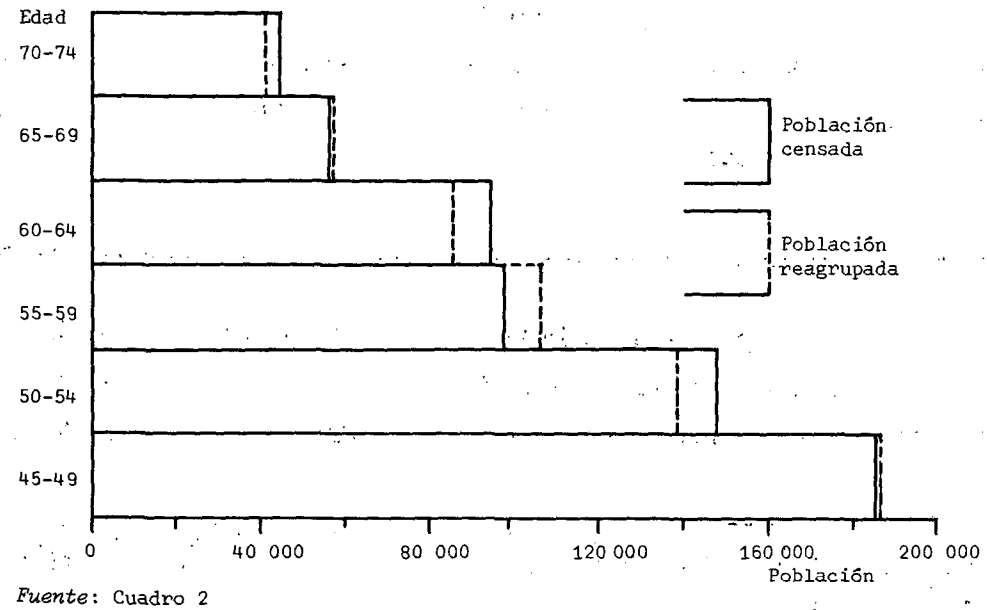
POBLACION CENSADA Y POBLACION REAGRUPADA USANDO EL AGRUPAMIENTO NO CONVENCIONAL CENTRADO EN 0 Y 5.
CENSO DE GUATEMALA DE 1973



Fuente: Cuadro 1.

Gráfico 2

POBLACION CENSADA Y POBLACION REAGRUPADA USANDO EL AGRUPAMIENTO NO CONVENCIONAL PROPUESTO POR HOBCRAFT.
CENSO DE GUATEMALA DE 1973



Fuente: Cuadro 2

de la población ajustada mucho más acorde con el comportamiento esperado. Esto se observa con mayor claridad en el gráfico 1.

Aún subsiste la limitación debida a los traslados de edades más allá de los grupos formados. Además el método no sería aplicable para aquellos países donde existan dígitos con gran atracción, diferentes del 0 y el 5.

1.3 Agrupamiento no convencional propuesto por Hobcraft[3/]

Este agrupamiento se basa en la hipótesis de que los redondeos de edades en 0 y 5 son los más importantes y se produce tanto por traslados de edades cercanas menores como mayores. Habría que agregar además que esos traslados son de la misma magnitud en las dos direcciones. En una edad determinada de preferencia de dígitos, por ejemplo en la edad 60 años, estarían incluidos: a) los que declaran 60 años y realmente los tienen, b) los que siendo menores de 60 años declaran 60 y c) los que siendo mayores de 60 años declaran tener 60.

Si a la población que aparece con 60 años de edad se la divide por la mitad, se tendrían dos sub-grupos: 60-60,5 y 60,5-61. De acuerdo con el supuesto inicial, el primer grupo incluye correctamente a menores de 60,5 años exactos y el segundo a mayores de esa misma edad. Si además se acepta que los traslados, por preferencia de dígitos, provienen de edades cercanas a la preferida, podría plantearse el siguiente tipo de agrupamiento:

55,5 - 60,5
60,5 - 65,5
65,5 - 70,5
etc.

Al igual que en el caso anterior es posible, por los mismos artificios anotados antes, volver a los grupos convencionales. En el cuadro 2 y el gráfico 2 se presenta el ejemplo para el mismo tramo de edades de la población censada de Guatemala en 1973.

3/ Hobcraft, J., Seminario sobre Estimaciones ... op.cit. Sesión I.
Carrier, N. y Hobcraft, J., Estimaciones Demográficas ... op.cit.,Cap.I.

Cuadro 2

CENSO DE GUATEMALA DE 1973: AJUSTE DE LA POBLACION POR EDADES USANDO EL AGRUPAMIENTO NO CONVENCIONAL PROPUESTO POR HOBCRAFT

Grupos de edades no convencional	Población	Grupos de edades convencional	Población	
			Reagrupada	Agrupada convencionalmente
40,5-45,5	216 954	45-49	185 281	184 502
45,5-50,5	181 197	50-54	138 525	147 004
50,5-55,5	134 150	55-59	107 588	99 665
55,5-60,5	105 200	60-64	85 819	95 026
60,5-65,5	83 574	65-69	57 406	56 239
65,5-70,5	54 810	70-74	41 345	44 214
70,5-75,5	40 022			
75,5-80,5	22 621			

Con este procedimiento se ha logrado, en forma más pronunciada que con el anterior, eliminar el efecto de la fuerte preferencia de dígitos, especialmente para la edad 60. Además de observarse un comportamiento lógico de descenso de la población con el aumento de la edad, se ve que sistemáticamente el ajuste disminuyó la cifra de los grupos de edades que contienen el 0 como dígito terminal aumentando los grupos que no lo contienen.

Parecería que este tipo de tratamiento de la información, muy poco usado hasta ahora, es prometedor, aunque se puede mencionar algunas limitaciones a los supuestos básicos:

a) Existe una mayor probabilidad de que las personas que redondean la edad provengan de las edades menores que de las mayores. Esto se debe a que en la pirámide de población el número de personas desciende con la edad y se espera, por ejemplo que el grupo 60-60,5 sea mayor al 60,5-61. Estaría implícito en el supuesto aceptado que la tendencia a rejuvenecerse es más intensa que a envejecerse. Al formar grupos quinquenales habría cierta compensación, pues si por el extremo superior se estuviera subestimando la población en el extremo inferior el efecto sería contrario.

b) Sigue existiendo la posibilidad de que la mala declaración produzca traslados de la edad más allá de los intervalos quinquenales. Estos errores pueden cometerse por los dos extremos y en los dos sentidos, es decir gente que se incluye en un grupo y no debería estar y gente que está en otros grupos y debería estar en el considerado.

c) Otra dificultad surge cuando se trata de información donde la preferencia por los dígitos 0 y 5 no es muy clara. Hobcraft[4/] pone como ejemplo el Censo de Argelia de 1948 en que la preferencia es por el 0,8 y 5 de acuerdo a los siguientes resultados del índice de Myers:

Dígito	Indice (%)	Dígito	Indice (%)
0	18,72	5	13,33
1	6,83	6	9,16
2	8,59	7	6,74
3	8,34	8	15,33
4	7,07	9	5,89

También en América Latina se encuentran algunos casos elocuentes, como Paraguay (1962):

Dígito	Indice (%)	Dígito	Indice (%)
0	12,3	5	10,5
1	8,0	6	9,5
2	12,4	7	9,7
3	9,4	8	10,3
4	9,5	9	8,4

Para superar la última limitación mencionada, podría usarse como criterio cortar por la mitad las edades de mayor rechazo, tratando de compensar las fuerzas que hay entre dos dígitos de atracción. Por ejemplo, al partir

4/ Hobcraft, J., Seminario sobre Estimaciones ... op.cit. Sesión I.

la edad 9 se compensaría la fuerza de los dígitos 0 y 8. En el caso particular de Argelia el agrupamiento propuesto es 9,5-3; 4-6,5; 6,5-9,5. Lo mejor sería formar grupos quinquenales, pero no es necesario que sea así, pues de todas maneras este agrupamiento es un paso intermedio para volver luego a los grupos convencionales. Tampoco es necesario realizar un mismo tipo de agrupamiento para todas las edades; la observación demográfica del comportamiento de las poblaciones por edades simples puede sugerir un tratamiento diferencial en distintos tramos de edades.

1.4 Determinación del mejor agrupamiento

En el cálculo del índice de Myers, para medir preferencias de dígitos, se supone que si la información es correcta debe esperarse que la suma combinada de cada dígito represente un 10 por ciento del total. Si es así, para un agrupamiento quinquenal correcto debe lograrse teóricamente que los índices de Myers, de los dígitos que se agrupen, sumen 50 por ciento.

A continuación se usa este criterio para analizar cuál de los agrupamientos sugeridos daría mejor resultado en el caso del censo de Guatemala de 1973, cuyos índices de Myers se calcularon en el cuadro 3, del capítulo VI:

Dígito	Indice (%)	Dígito	Indice (%)
0	15,63	5	13,65
1	6,08	6	8,53
2	10,90	7	8,64
3	9,51	8	10,89
4	8,66	9	7,51

a) Agrupamiento quinquenal convencional:

Agrupamiento	Porcentaje en suma combinada	
	Real	Esperado
0-4	50,8	50
5-9	49,2	50

b) Agrupamiento quinquenal centrado en 0 y 5:

Agrupamiento	Porcentaje en suma combinada	
	Real	Esperado
3-7	49,0	50
8-2	51,0	50

c) Agrupamiento quinquenal propuesto por Hobcraft:

Agrupamiento	Porcentaje en suma combinada	
	Real	Esperado
0,5-5,5	49,8	50
5,5-0,5	50,2	50

Como se ve, el último es el que da mejor resultado, pues el desvío entre los valores empíricos deducidos del censo y los esperados es el menor. En el primer caso planteado el grupo con dígitos de 0 a 4 resulta de mayor atracción por contener el 0. El defecto se agudiza en el segundo caso debido a que el grupo que contiene los dígitos de 8 a 2 cuenta, además del 0, con el 8, también de atracción.

Este indicador de mejor agrupamiento se refiere a casi la totalidad de la población por edades (Myers se calculó para el intervalo de 10 a 79 años), pero es importante recalcar una vez más que no necesariamente debe darse el mismo tratamiento a toda la población: un agrupamiento que funciona bien en un tramo de edades determinado, puede no ser satisfactorio para otro. El uso de estas herramientas depende en gran medida del buen juicio del demógrafo.

2. Ajustes gráficos

El objetivo de estos métodos es imponer un comportamiento regular a la información mediante algún artificio gráfico, bajo dos supuestos: a) que la

población en ausencia de fenómenos extraordinarios, tiene un comportamiento regular con la variación de la edad y b) que ese comportamiento presenta una forma esperada y más o menos universal.

Son útiles cuando se trata de eliminar grandes irregularidades que pueden ser observadas a simple vista al graficar la información. El éxito de estos procedimientos depende en gran parte de la posibilidad de establecer cuándo una irregularidad se debe a hechos reales o es producto de errores de los datos que se procederá a eliminar. Por ello los ajustes gráficos están cargados con un alto grado de subjetividad. Por otra parte, resultan sumamente flexibles, pudiendo respetarse en mayor o menor medida los datos observados.

2.1 Ajuste de ojivas naturales

Se utiliza el término "ojiva" como sinónimo de "curva de distribución" referida a un gráfico donde la ordenada representa las frecuencias de la población acumuladas y la abscisa, la edad. En el cuadro 3 y el gráfico 3 se presenta la ojiva correspondiente al censo de Haití de 1950 y la forma operativa de realizar el ajuste correspondiente. Luego de acumulada la población se grafican los puntos y a través de un ajuste manual se corrigen las irregularidades que muestra la ojiva. A través de una lectura gráfica se obtienen los valores corregidos de la población acumulada. Luego, desacumulando, se llega a la población por grupos de edades corregida. En el caso expuesto se trabajó con valores absolutos, pero nada impide hacerlo con la distribución relativa de la población y, luego de ajustada, aplicarla al total.

Una característica fundamental de este método es que no se afecta el total de la población; simplemente se transfiere población de un grupo a otro. Con esto se corrigen principalmente los errores debidos a traslación de edades o preferencia por algunas de ellas que producen excesos en determinados grupos en desmedro de otros. Por ejemplo, véase que los datos observados presentaban un grupo 35-39 superior a 30-34 y 60-64 mayor a 55-59; en cambio, al efectuar el ajuste se logra, en esos casos, un comportamiento más regular de acuerdo a lo esperado a medida que avanza la edad.

El problema de escala del gráfico es la principal dificultad con que se tropieza al aplicar esta técnica. En primer lugar, las irregularidades no se perciben con mucha nitidez, a no ser que sean realmente muy importantes

o se utilicen escalas de una dimensión poco frecuente. Por otro lado, la lectura gráfica es muy imprecisa y una pequeña modificación en un punto representa la transferencia de miles de personas.

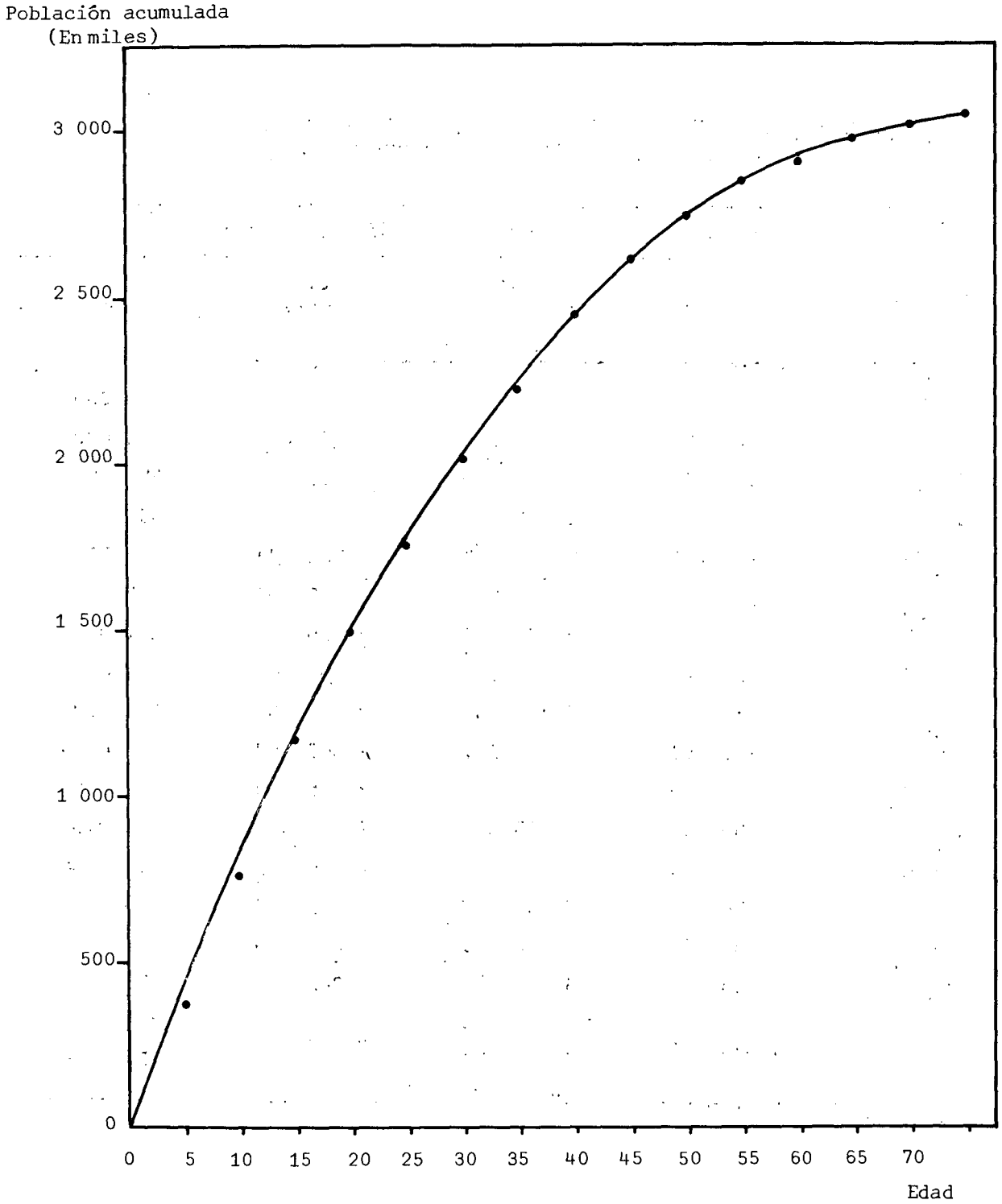
Cuadro 3

CENSO DE HAITI DE 1950: AJUSTE DE LA POBLACION POR GRUPOS DE EDADES MEDIANTE EL METODO DE LA OJIVA NATURAL

Grupos de edades	Población	Ojiva	Ojiva ajustada	Población ajustada
	${}_5N_x$	$P(x)=\sum_{0}^{x} {}_5N_x$	$\hat{P}(x)$	${}_5\hat{N}_x=\hat{P}(x+5)-\hat{P}(x)$
0-4	374 872	0	0	450 000
5-9	400 518	374 872	450 000	375 000
10-14	397 708	775 390	825 000	348 098
15-19	308 026	1 173 098	1 173 098	308 026
20-24	267 401	1 481 124	1 481 124	293 876
25-29	277 177	1 748 525	1 775 000	245 000
30-34	189 144	2 025 702	2 020 000	230 000
35-39	229 644	2 214 846	2 250 000	194 490
40-44	157 697	2 444 490	2 444 490	157 697
45-49	133 451	2 602 187	2 602 187	133 451
50-54	99 389	2 735 638	2 735 638	99 389
55-59	56 828	2 835 027	2 835 027	74 973
60-64	70 954	2 891 855	2 910 000	52 809
65-69	40 390	2 962 809	2 962 809	40 390
70-74	39 566	3 003 199	3 003 199	39 566
75 y más	43 270	3 042 765	3 042 765	43 270
Total	*3 086 035*	*3 086 035*	*3 086 035*	*3 086 035*

Gráfico 3

AJUSTE GRAFICO DE LA OJIVA NATURAL DE
LA POBLACION POR GRUPOS DE EDADES. CENSO DE HAITI DE 1950



Fuente: Cuadro 3

2.2 Ajuste de ojivas de ejes oblicuos

Con el fin de superar el problema de escala que tienen las ojivas naturales, Carrier y Parrag[5/] han propuesto una modificación mediante un sistema de rotación del eje de las abscisas. Para ello se propone restar a la función acumulada original, una recta bisectriz, adoptando esas diferencias como nuevos puntos de ajuste: $Q(x) = P(x) - Ax$ (véase el gráfico 4). A la función $Q(x)$ se la denomina "ojiva de ejes oblicuos" y permite cuadruplicar, en este caso, la escala correspondiente a la ordenada y, por lo tanto, las irregularidades, que apenas eran perceptibles en la ojiva natural, ahora son muchísimo más notorias, como puede apreciarse en el gráfico 4 donde aparecen ambas funciones acumuladas.

Para obtener la ojiva de ejes oblicuos es necesario en primer lugar darse la recta. A continuación se detalla la forma de obtenerla.

El primer paso es acumular la población, es decir, obtener la ojiva natural. El ajuste puede comenzarse en cualquier edad; debido a los errores de omisión que comúnmente afectan al grupo 0-4, es conveniente en muchos casos comenzar a acumular desde la edad 5. Tampoco es necesario acumular hasta el final; puede terminarse en 65 ó 70 años si se piensa que las edades más avanzadas merecen un tratamiento especial o que por ser poblaciones muy pequeñas aporten muy poco a la ojiva y por lo tanto no haríamos otra cosa que reproducir los puntos.

Se define entonces:

l = límite inferior del intervalo a ajustar. En este caso: $l = 0$; $P(l) = 0$.

u = Límite superior del intervalo a ajustar. En este caso se adoptará $u=65$.

La recta $y = a + bx$ se calcula a partir de las ecuaciones:

$a + bl = 0$

$a + bu = P(u)$

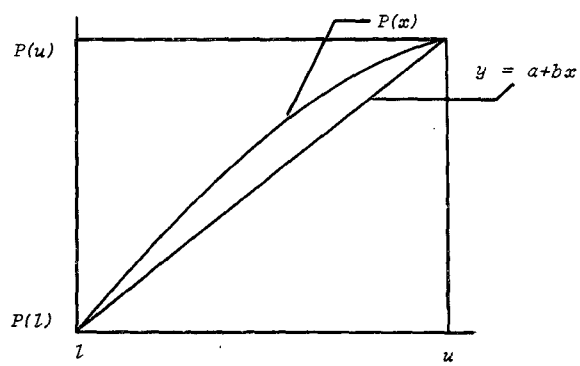
Haciendo la diferencia:

$b(u-l) = P(u)$

de donde:

$$b = \frac{P(u)}{u-l}$$

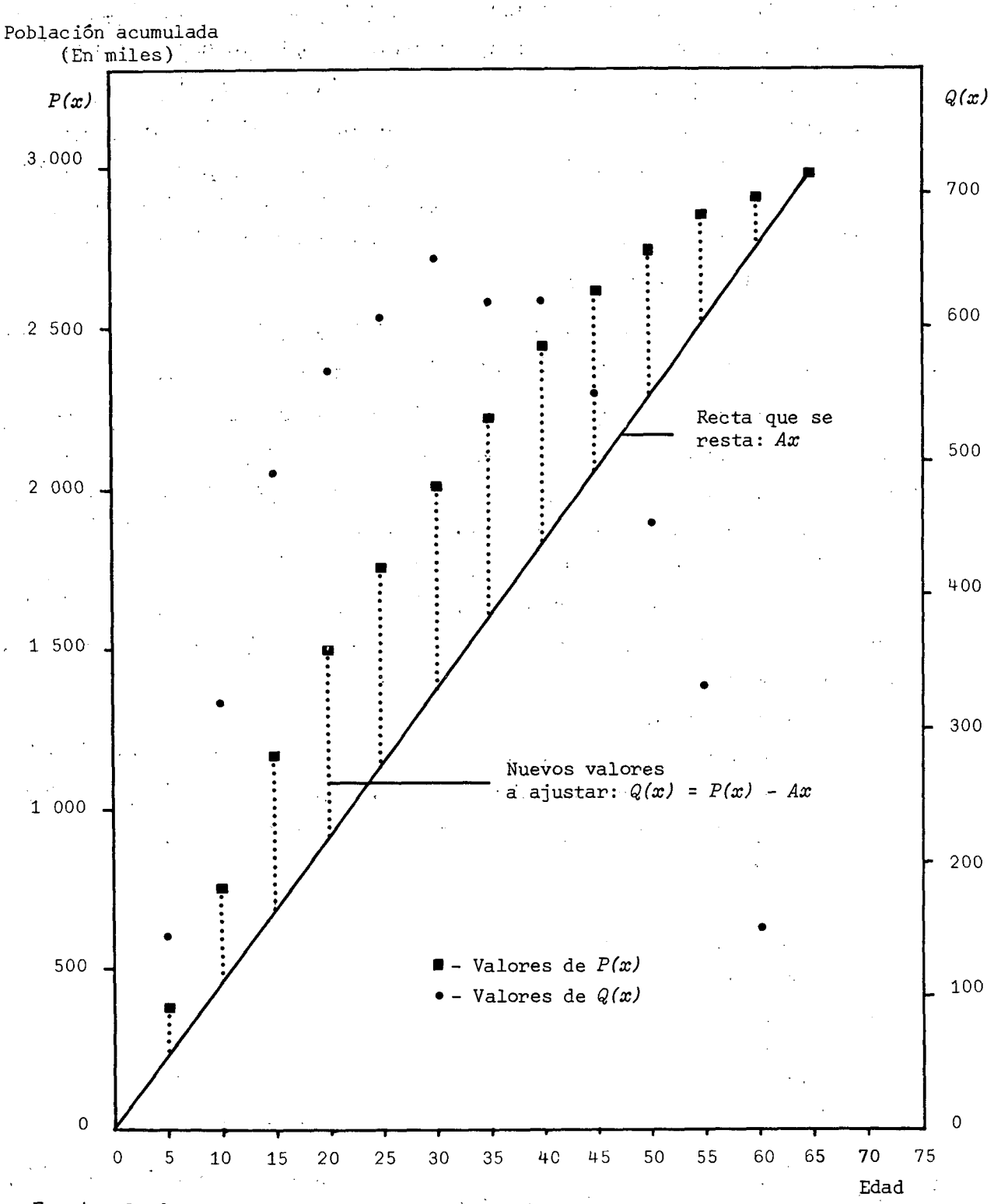
$$a = \frac{-P(u)}{u-l}\, l$$



5/ Carrier, N.H. y Farrag, A.M., "La Reducción de Errores en los Censos de Población para Países Estadísticamente Subdesarrollados", en Revista Estadística. Journal of the Inter-American Statistical Institute, N°71, junio 1961.

Gráfico 4

OJIVA NATURAL Y OJIVA DE EJES OBLICUOS.
CENSO DE HAITI DE 1950



Fuente: Cuadro 4

y por lo tanto la recta es:

$$\boxed{y = \frac{P(u)}{u-l}\,(x-l)}$$

La ojiva de ejes oblicuos será entonces:

$$Q(x) = P(x) - \frac{P(u)}{u-l}\,(x-l)$$

En el cuadro 4 se realiza el cálculo de $Q(x)$ para la población censada de Haití en 1950:

$$Q(x) = P(x) - \frac{2\ 962\ 809}{65}\,x$$

Cuadro 4

CENSO DE HAITI DE 1950: AJUSTE DE LA POBLACION POR GRUPOS DE EDADES MEDIANTE EL METODO DE LA OJIVA DE EJES OBLICUOS

Grupos de edades	Población ${}_5N_x$	Población acumulada $P(x)$	Recta $A(x-l)$	Ojiva de ejes oblicuos $Q(x)$	$\hat{Q}(x)$	$\hat{P}(x)$	${}_5\hat{N}_x$
0-4	374 872	0	0	0	0	0	427 908
5-9	400 518	374 872	227 908	146 964	200 000	427 908	387 909
10-14	397 708	775 390	455 817	319 573	360 000	815 817	357 281
15-19	308 026	1 173 098	683 725	489 373	489 373	1 173 098	308 026
20-24	267 401	1 481 124	911 634	569 490	569 490	1 481 124	278 418
25-29	277 177	1 748 525	1 139 542	608 983	620 000	1 759 542	247 908
30-34	189 144	2 025 702	1 367 450	658 252	540 000	2 007 450	217 909
35-39	229 644	2 214 846	1 595 359	619 487	630 000	2 225 359	200 408
40-44	157 967	2 444 490	1 823 267	621 223	602 500	2 425 767	176 420
45-49	133 451	2 602 187	2 051 175	551 012	551 012	2 602 187	133 451
50-54	99 389	2 735 638	2 279 084	456 554	456 554	2 735 638	99 389
55-59	56 828	2 835 027	2 506 992	328 035	328 035	2 835 027	77 874
60-64	70 954	2 891 855	2 734 901	156 954	178 000	2 912 901	49 908
65-		2 962 809	2 962 809	0	0	2 962 809	

$l = 0 \quad u = 65 \quad P(u) = 2\ 962\ 809 \quad A = \frac{2\ 962\ 809}{65}$

Posteriormente se grafican los puntos y se ajustan manualmente (véase el gráfico 5). En el gráfico se leen los valores ajustados y luego se realiza todo el proceso inverso, es decir se suma la recta para llegar a la ojiva natural ajustada $[\hat{P}(x)]$ y finalmente se desacumula obteniéndose la población ajustada por grupos de edades.

Este tipo de ajuste puede ser utilizado para pasar de grupos de edades no convencionales a grupos convencionales, tal como se mencionó en la sección 1.2 de este mismo capítulo. Esto es posible porque se puede acumular la población con cualquier tipo de agrupamiento. Por ejemplo, si se trata de los grupos 8-12, 13-17, 18-22, 23-27, etc. se parte de $l = 8$ y se obtienen $Q(13)$, $Q(18)$, $Q(23)$, etc. que se grafican y ajustan. La lectura gráfica puede realizarse en las edades que convergen: $\hat{Q}(10)$, $\hat{Q}(15)$, etc. y a partir de estos valores corregidos se obtiene la población para los grupos 10-14, 15-19, 20-24, etc.

2.3 Ventajas y limitaciones del método de la ojiva de ejes oblicuos

Se pueden mencionar como ventajas de este tipo de ajuste, los siguientes aspectos:

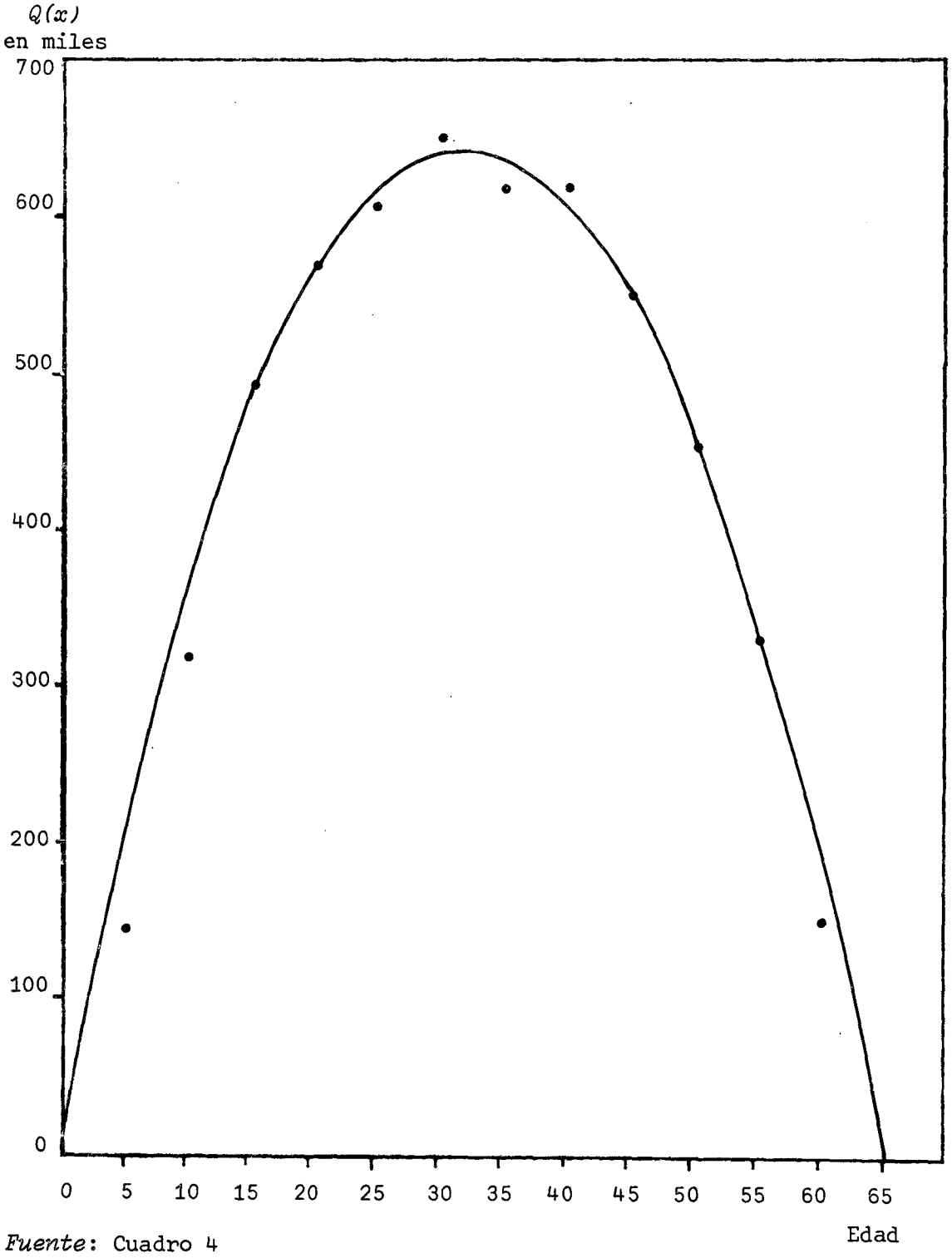
a) La simplicidad de las operaciones que deben realizarse. Es una virtud general de los ajustes de tipo gráfico el hecho de no estar basados en técnicas matemáticas muy sofisticadas, lo que permite a cualquier persona, aun sin conocimientos estadístico-matemáticos muy profundos ponerlos en práctica.

b) Puede usarse cualquier agrupamiento de edades y, es más, no necesariamente se debe trabajar con grupos iguales. Esto permite agrupar la población en forma diferente de acuerdo con el tipo de errores que contengan los datos. Luego se ajusta por medio de la ojiva de ejes oblicuos, formando los grupos de edades convencionales.

c) Al modificar un punto del gráfico, en esa misma medida se modifican los grupos de edades involucrados. Si al corregir a $Q(x)$ le aumentamos 10 000 personas, eso significa agregar esa misma cantidad a la población entre x-5 y x años de edad (${}_5N_{x-5}$) y deducirla del grupo siguiente (${}_5N_x$).

Gráfico 5

AJUSTE GRAFICO DE LA OJIVA DE EJES OBLICUOS DE LA POBLACION POR GRUPOS DE EDADES. CENSO DE HAITI DE 1950



Fuente: Cuadro 4

d) Los ajustes mediante ojivas reproducen el total de la población, es decir, que la corrección consiste en trasladar personas de un grupo de edades a otro, tratando de eliminar los errores en la declaración de la edad.

Como limitaciones puede señalarse:

a) Se trata de una técnica con alto grado de subjetividad y arbitrariedad. Dos demógrafos enfrentados a los mismos puntos, harían un ajuste distinto; es más, un mismo demógrafo enfrentado en diferentes momentos a los mismos puntos no repetiría exactamente el primer ajuste. El resultado final depende más que nada del buen juicio del demógrafo y de su conocimiento del porqué de las irregularidades observadas.

b) Si bien la ojiva de ejes oblicuos mejora la precisión por permitir una mayor escala que la ojiva natural, aun depende de la lectura gráfica que dista mucho de ser exacta.

c) Si bien este procedimiento no recurre a técnicas matemáticas complicadas, la experiencia indica que el uso de gráficos, que deben tener cierta precisión, no resulta tan sencillo. Es necesario poseer alguna destreza para dibujarlos y leerlos, de lo contrario el trabajo se hace lento e impreciso.

d) Las ojivas, en general, solamente son útiles en países con datos muy malos. De no ser así, es difícil percibir las irregularidades. Por lo mismo, en América Latina solamente tienen aplicaciones en los censos de menor calidad. En la mayoría de los casos lo más probable es que no se manifiesten claramente los errores que se desea corregir.

3. Ajustes estadístico-matemáticos

Cuando se habla de métodos analíticos o estadístico-matemáticos para realizar ajustes de la población por grupos de edades, se hace referencia a aquellas técnicas basadas en fórmulas matemáticas cuya aplicación lleva implícita la aceptación de ciertos supuestos sobre los errores de la información. Además, el uso de estos procedimientos, a diferencia de los gráficos, conduce a un único resultado.

Si se hiciera el supuesto de que los errores que contienen los datos son de tipo aleatorio, bastaría realizar un ajuste por mínimos cuadrados mediante una función adecuada a efectos de eliminar las irregularidades debidas al azar. Pero en demografía, y en particular en los países de América Latina, las cosas no son así; por el contrario, las deficiencias de la información se deben en su gran mayoría a errores sistemáticos como, por ejemplo, la preferencia de dígitos, el traslado de edades y la omisión diferencial por edades. De ahí que las fórmulas planteadas procuran generalmente eliminar las irregularidades originadas por ciertos patrones de error que se aceptan como los más importantes o posibles de detectar.

Estos procedimientos son muy simples y fáciles de aplicar pues, luego de elegida la fórmula, su tratamiento es mecánico, no interviniendo el juicio personal del demógrafo.

La limitación fundamental está dada por la rigidez que implica el uso de una expresión matemática igual para todas las edades, lo que conduce a darles el mismo tratamiento, independientemente del tipo de error que las afecte. Es necesario señalar, además, que la mayoría de estas técnicas no reproducen el total de la población que se tiene antes del ajuste.

A continuación se describen algunos de estos procedimientos, haciendo mención de sus supuestos básicos y ejemplificando con el censo de Guatemala de 1973 (véase el cuadro 5).

3.1 Promedios móviles

Se basa en los supuestos de eliminación parcial de los errores aleatorios que pueden ser por exceso o por defecto y en un comportamiento lineal de los grupos considerados en el ajuste. La corrección de errores se hace mediante un promedio de excesos y defectos. Si además de los errores aleatorios existen irregularidades reales o desvíos sistemáticos, la técnica les da el mismo tratamiento. En ese sentido existe el riesgo de sobrecorregir la información violando el principio de "rehabilitación",[6/] lo que constituye un peligro latente en los procedimientos que aquí se consideran.

6/ Véase el Cap. I, pág. 10.

La fórmula puede aplicarse con tres puntos de apoyo, tanto por grupos simples de edades como por grupos quinquenales:

$$\hat{N}_x = \frac{N_{x-1} + N_x + N_{x+1}}{3}$$

donde: $\hat{N}_x$ - población ajustada de edad x

N_x - población censada de edad x

N_{x-1} - población censada del grupo de edades anterior al que se ajusta

N_{x+1} - población censada del grupo de edades posterior al que se ajusta.

Una desventaja de esta fórmula, al trabajar por edades individuales, es que si se ajusta una edad en que no hay atracción por las tres edades involucradas, es posible que el valor estimado sea muy bajo (por ejemplo, N_{22} ó N_{23}). Este problema puede ser resuelto trabajando con cinco valores:

$$\hat{N}_x = \frac{N_{x-2} + N_{x-1} + N_x + N_{x+1} + N_{x+2}}{5}$$

En este caso, cualquier edad que se ajuste contendrá el 0 ó el 5, que son las edades de mayor atracción, pero se sigue considerando como supuesto el que los errores son de tipo aleatorio y en ese sentido el promedio le da el mismo peso a todas las edades. Además, en esta modalidad, el supuesto de linealidad es más exigente. Por otra parte, las edades terminadas en 0 mantendrían un exceso debido a que en el ajuste intervendrían los dígitos 0,8 y 2 que generalmente son 3 de los 4 preferidos.

Es probable que luego de usado este procedimiento, aun no se alcance a suavizar la información, en cuyo caso podría reiterarse su aplicación las veces que se considere necesario (en el ejemplo del cuadro 5 se aplicó una sola vez).

3.2 Promedios móviles con ponderación (16 avos)

Esta fórmula de ajuste ha sido recomendada por las Naciones Unidas[7/] y presentada con más detalle por Carrier y Farrag.[8/] Ha sido diseñada para ajustar la población por grupos quinquenales de edades tratando de eliminar, fundamentalmente, los errores debidos a preferencia de dígitos.

Cuando se presenta la población en grupos quinquenales convencionales, aún persiste la influencia del redondeo de edades. Carrier y Farrag suponen que luego de agrupar la población permanecen presentes las siguientes preferencias:

Dígitos finales de los grupos de edades	Ganancias	Pérdidas
0 a 4	Del 5 al 9 anteriores y del 5 siguiente hacia el 0	Del 3 y 4 hacia el 5 siguiente
5 a 9	Del 3 y 4 anteriores hacia el 5	Del 5 hacia el 0 anterior o hacia el siguiente; del 6 al 9 hacia el 0 siguiente

En resumen, el balance sería favorable a los grupos con dígitos de 0 a 4 que tendrían excesos similares a los defectos del grupo anterior con dígitos de 5 a 9. Aceptando este tipo de patrón de errores, la fórmula del 16 avos se basa en los siguientes supuestos:

a) Las ganancias y pérdidas netas de grupos de edades alternados, que se mencionó anteriormente, se mantienen constantes con la edad.

b) La diferencia cuarta (sucesivas) de los grupos de edades es igual a cero. Este supuesto se realiza como una exigencia de suavidad en la información e implica el uso de 5 grupos de edades.

7/ Naciones Unidas, Manual III. Métodos para Preparar Proyecciones de Población por Sexo y Edad, ST/SOA/Serie A, Nº 25, Nueva York, 1956.
8/ Carrier y Farrag, "La Reducción de Errores ...", op.cit. pág. 313.

El primer supuesto puede expresarse matemáticamente:

$$ {}_5\hat{N}_x = {}_5N_x + (-1)^x e $$

e: son las ganancias y pérdidas netas, que cuando x es par se agregan y cuando es impar se restan.

El segundo supuesto, a su vez, exige que:

$$ \Delta^4 \; {}_5\hat{N}_x = 0 $$

lo que implica:

$$ {}_5\hat{N}_{x-10} - 4 \; {}_5\hat{N}_{x-5} + 6 \; {}_5\hat{N}_x - 4 \; {}_5\hat{N}_{x+5} + {}_5\hat{N}_{x+10} = 0 $$

o sea:

$$ ({}_5N_{x-10} + e) - 4({}_5N_{x-5} - e) + 6({}_5N_x + e) - 4({}_5N_{x+5} - e) + ({}_5N_{x+10} + e) = 0 $$

y despejando:

$$ e = \frac{-{}_5N_{x-10} + 4 \; {}_5N_{x-5} - 6 \; {}_5N_x + 4 \; {}_5N_{x+5} - {}_5N_{x+10}}{16} $$

y como:

$$ {}_5\hat{N}_x = {}_5N_x + e $$

se tiene:

$$ \boxed{{}_5\hat{N}_x = \frac{-{}_5N_{x-10} + 4 \; {}_5N_{x-5} + 10 \; {}_5N_x + 4 \; {}_5N_{x+5} - {}_5N_{x+10}}{16}} $$

que es la fórmula propuesta. Obsérvese que es un promedio móvil, pero con determinadas ponderaciones.

En el cuadro 5 puede verse que, salvo los grupos 30-34 y 40-44, la tendencia es a disminuir la población con dígitos entre 0 y 4 en desmedro de los otros. El único grupo con dígitos 5 a 9 que disminuye es el 35-39 (esto

produciría el aumento de 30-34 y 40-44), lo que podría estar relacionado con una preferencia especial por la edad 35, que parece existir de acuerdo a los análisis efectuados en el capítulo VI. Si bien el ajuste aparece como satisfactorio por la alternación de signos en las diferencias con los valores observados, no se cumpliría el primer supuesto mencionado, ya que estas diferencias son muy variables con la edad.

3.3 Ajuste de la razón K[9/]

Este procedimiento supone que la población agrupada decenalmente en la forma 15-24, 25-34, etc., estaría exenta de los errores de preferencia de dígitos debido a compensaciones internas y por lo tanto se acepta como correcta. El agrupamiento propuesto resulta poco práctico para fines de análisis, por lo tanto se hace necesario subdividir en grupos quinquenales. Para ello se introduce el supuesto de que la relación entre un grupo quinquenal de edades y el siguiente es constante:

$$\frac{{}_5N_x}{{}_5N_{x+5}} = K$$

Bajo tales condiciones, el procedimiento de cálculo para subdividir el grupo decenal ${}_{10}N_x = {}_5N_x + {}_5N_{x+5}$ sería el siguiente:

a) $$K^4 = \frac{{}_{10}N_{x-10}}{{}_{10}N_{x+10}} \quad \therefore \quad K = \sqrt[4]{\frac{{}_{10}N_{x-10}}{{}_{10}N_{x+10}}}$$

b) El grupo quinquenal más viejo:

$${}_5\hat{N}_{x+5} = \frac{{}_{10}N_x}{1+K}$$

c) El grupo quinquenal más joven:

$${}_5\hat{N}_x = {}_{10}N_x - {}_5\hat{N}_{x+5}$$

Por construcción éste es un ajuste que reproduce el total de la población ya que simplemente reasigna personas dentro de cada grupo decenal. Los cálculos y resultados aparecen en el cuadro 5.

9/ Carrier y Farrag, La Reducción de Errores ... op. cit. pág. 317.

A continuación se presenta el desarrollo teórico que conduce a las fórmulas anteriormente planteadas:

Se sabe que:

$$ {}_{5}N_{x} = K \; {}_{5}N_{x+5} \tag{1} $$

y,

$$ {}_{10}N_{x-10} = {}_{5}N_{x-10} + {}_{5}N_{x-5} \tag{2} $$

Considerando que:

$$ {}_{5}N_{x-10} = K \; {}_{5}N_{x-5} \tag{3} $$

sustituyendo en (2) se tiene:

$$ {}_{10}N_{x-10} = (1+K) \; {}_{5}N_{x-5} \tag{4} $$

entonces,

$$ {}_{10}N_{x} = (1+K) \; {}_{5}N_{x+5} \tag{5} $$

y

$$ {}_{10}N_{x+10} = (1+K) \; {}_{5}N_{x+5} \tag{6} $$

Ahora:

$$ {}_{5}N_{x-5} = K \; {}_{5}N_{x} \tag{7} $$

a su vez:

$$ {}_{5}N_{x} = K \; {}_{5}N_{x+5} \tag{8} $$

sustituyendo en (7):

$$ {}_{5}N_{x-5} = K^{2} \; {}_{5}N_{x+5} \tag{9} $$

a su vez:

$$ {}_{5}N_{x+5} = K \; {}_{5}N_{x+10} \tag{10} $$

sustituyendo en (9):

$$ {}_{5}N_{x-5} = K^{3}\ {}_{5}N_{x+10} \tag{11} $$

y como:

$$ {}_{5}N_{x+10} = K\ {}_{5}N_{x+15} \tag{12} $$

se obtiene:

$$ {}_{5}N_{x-5} = K^{4}\ {}_{5}N_{x+15} \tag{13} $$

Combinando las expresiones (4) y (13):

$$ {}_{10}N_{x-10} = K^{4}(1+K)\ {}_{5}N_{x+15} \tag{14} $$

y a partir de (6) y (14):

$$ \boxed{\frac{{}_{10}N_{x-10}}{{}_{10}N_{x+10}} = K^{4}} \tag{15} $$

Hallado K, y con base en (5):

$$ \boxed{{}_{5}N_{x+5} = \frac{{}_{10}N_{x}}{1+K}} \tag{16} $$

y al otro grupo por diferencia:

$$ \boxed{{}_{5}N_{x} = {}_{10}N_{x} - {}_{5}N_{x+5}} \tag{17} $$

Cuadro 5

CENSO DE GUATEMALA DE 1973: AJUSTES DE LA POBLACION POR GRUPOS DE EDADES MEDIANTE FORMULAS ESTADISTICO-MATEMATICAS

Edad	Población censada	Población ajustada		Razón *K*		
		Promedios móviles	16 avos	Población censada	*K*	Población ajustada
0-4	870 377					
5-9	777 735			1 455 895		
10-14	678 160	671 365	674 563			
15-19	560 279	566 741	566 970	1 030 551	1,2331	569 062
20-24	470 272	467 691	460 765			461 489
25-29	347 260	385 721	353 439	629 742	1,1993	343 405
30-34	282 482	319 621	286 691			286 337
35-39	268 313	262 467	262 525	498 091	1,1740	268 978
40-44	229 778	222 416	229 972			229 113
45-49	184 502	185 852	186 511	331 506	1,2647	185 126
50-54	147 004	151 195	142 619			146 380
55-59	99 665	116 487	107 752	194 691	1,3478	101 766
60-64	95 026	88 430	86 416			82 925
65-69	56 239	63 780	62 246	100 453	1,5032	60 323
70-74	44 214	46 723	40 795			40 130
75-79	23 756			38 136		
80-84	14 380					

Este documento forma parte de un texto más amplio, constituido por 10 capítulos, cuyo título es Evaluación y Corrección de Datos Demográficos.

La publicación independiente de este capítulo y de otros que aparecerán posteriormente, se hace con la finalidad de utilizarlos como material docente.

Los autores agradecen, desde ya, los comentarios y sugerencias que les puedan hacer llegar, pues serán de gran utilidad para la versión definitiva.