

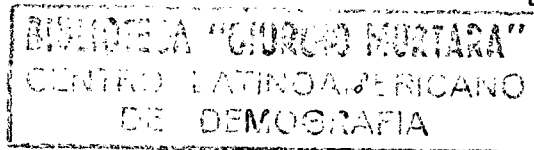
CELADE

SUBSEDE

William Brass

AJUSTE E INTERPRETACION DE

DATOS DEMOGRAFICOS



Traducción del Documento: "Disciplining Demographic Data",
presentado a la Conferencia de Londres, organizada por la
UIECP. Vol. 1. Setiembre, 1969.

CENTRO LATINOAMERICANO DE DEMOGRAFIA

SERIE DS No. 8
Edición provisional

San José, Costa Rica

1971 5606 ✓

AJUSTE E INTERPRETACION DE DATOS

DEMOGRAFICOS 1/

William Brass

INTRODUCCION

Durante los últimos diez a quince años ha habido grandes adelantos en las técnicas para extraer mediciones significativas de datos de población limitados y defectuosos. La mayor parte de estos adelantos se han relacionado con la construcción de modelos (o distribuciones de referencia) de procesos demográficos, pero también han cobrado importancia otros conceptos como comprobaciones internas y robustez de las relaciones. Junto con estos adelantos en la metodología se ha logrado un mayor conocimiento y una comprensión más profunda de la "forma" de las mediciones de población, por ejemplo de las distribuciones por edad, patrones de mortalidad, y cómo éstos cambian con el tiempo. A medida que aumente la exactitud de los datos estadísticos, puede que las técnicas se vayan haciendo anticuadas, pero la experiencia puede conducir a nuevos adelantos.

En vista de estas observaciones, parece paradójico que los resultados de la aplicación de métodos más recientes de análisis a los materiales de una población determinada sean con frecuencia, desalentadores. En primer lugar, los tipos de datos recogidos (o con frecuencia todo lo tabulado) pueden no ser totalmente apropiados para la aplicación de los métodos. A medida que se extiende la comprensión de la potencia de los procedimientos, este obstáculo será menos común. En segundo lugar, el asunto está sufriendo por su propio éxito. Ahora que se ha demostrado cuánto puede hacerse con material que no promete, se tienen mayores esperanzas. A sistemas primitivos de registro se

1/ Disciplining Demographic Data, Documento presentado a la Conferencia General de la Unión Internacional para el Estudio Científico de la Población, Londres, setiembre de 1969.

N de T: Se ha traducido disciplining por ajuste e interpretación.

le exigen mediciones que pueden usarse para una planificación sofisticada. Pero los datos van mejorando tanto en extensión como en exactitud, lo que sorprendentemente conduce a la tercera razón de descontento. Las técnicas necesariamente rígidas para imponer el orden en información muy restringida y dudosa resultan demasiado inflexibles cuando los materiales son mejores. Solamente podrá lograrse más progreso hacia la medición precisa por medio de análisis que tomen en cuenta, en mayor grado, las peculiaridades de la población determinada en vez de moldearlas a los "promedios" de los modelos.

Se han publicado varios compendios tanto de los principios como de la aplicación de nuevos métodos para derivar medidas de datos poco satisfactorios. Si bien ninguno de ellos es exhaustivo, y algunos de los nuevos adelantos todavía tienen que describirse completamente; si se escribiera otra obra, sería necesario repetir mucho de lo que ya se ha dicho antes. Están comenzando a aparecer estudios en los que se han hecho análisis más flexibles, pero solidamente basados, de registros de mejor calidad. Sin embargo, todavía es escasa la experiencia sobre los problemas y la eficacia de las soluciones intentadas, y sería prematuro pretender una evaluación. En lugar de eso, he escrito un documento que explora algunos de los enfoques que parecen ser productivos en el manejo de datos mejorados pero todavía inseguros. Las conclusiones sobre su valor general deben ser provisionales y, a decir verdad, la naturaleza de las metas es tal que siempre tiene que haber ciertas reservas al formular técnicas. Las idiosincrasias de las observaciones sobre una población determinada deben servir de guía al adoptar los métodos.

Este documento tiene tres secciones. En la primera se explican e ilustran algunos de los procedimientos generales que han demostrado ser útiles en la práctica. Las otras dos secciones tratan, cada una, de una zona particular de aplicación en la cual persiste la necesidad de introducir mejores técnicas.

II. METODOS GENERALES

Cuando los datos son realmente malos, hay que atenerse mucho a modelos simples. Lo más que puede esperarse es conseguir determinar unas pocas cifras basadas en las observaciones, con suficiente aproximación como para establecer una probable gama de posibilidades. Lo correcto sería decir que comenzamos con los modelos y adaptamos la información a los patrones ya establecidos. Trabajando en dirección contraria, esto es, comenzando por las características de las observaciones se logra estructurar datos más satisfactorios. Así, la detección y el ajuste de errores procede de los supuestos más débiles sobre la relación fundamental los que generalmente dependen de ideas de uniformidad y orden antes de que ecuaciones funcionales rígidas. Las desviaciones se rechazan por espurias porque ocurren fuera del patrón general o tendencia y no porque no se ajusten a diseños predeterminados. En gran medida, por lo tanto, el "disciplinar" las observaciones depende de la elaboración de procedimientos apropiados para interpretarlas.

Los problemas de corrección de datos son comunes a muchos campos de estudio. En años recientes ha habido un resurgimiento del interés en el tópico, estimulado por las grandes cantidades de cifras desembuchadas por las computadoras electrónicas. El ordenamiento y arreglo de aquellas y el uso de los grandes recursos de la máquina para ayudar en la operación constituyen tareas sustanciales. Sea cual sea el origen de los datos, los principios de 'disciplinar' son muy parecidos y he tomado la palabra misma de un trabajo sobre datos médicos preparado por Healy (1968). No obstante, la demografía difiere de la mayoría de otros estudios en el enorme volumen, detalle y exactitud de sus materiales cuantitativos. Como consecuencia de ello puede hacer un menor uso de estructuras matemáticas simples (distribuciones normales, etc.) en las que las desviaciones se toman como elementos aleatorios. Los modelos (por ejemplo, distribuciones de referencia) se han construido con frecuencia en forma empírica.

Para disciplinar datos demográficos he encontrado cuatro recursos de especial valor, los que se analizan más adelante con ilustraciones tomadas de la corrección de distribuciones por edades, uno de los problemas centrales en estudios de población. Los recursos son:

- (a) Acumulación
- (b) Uso de estándares de referencia
- (c) Transformaciones lineales
- (d) Transformaciones de escala

Por lo general es mejor examinar los resultados de estos procedimientos en representaciones gráficas ya que las características de las medidas se ven más fácilmente. La decisión sobre la siguiente etapa de ajustes depende de la inspección aún cuando se introduzcan técnicas objetivas de suavizamiento o de ajuste.

La acumulación de cantidades se usa extensamente por razones que fácilmente pueden apreciarse. Si los términos en una serie de medidas están todos afectados por la misma clase de sesgos sistemáticos, las sumas acumuladas retendrán esos sesgos, pero la influencia de fluctuaciones erráticas, fortuitas y otras se reducirá por la cancelación de movimientos contrarios. La existencia y la naturaleza de sesgos se hará más clara. Existen series demográficas apropiadas en función del tiempo (tasas de fecundidad y mortalidad, etc.) y la edad (volumen de poblaciones, probabilidades de muerte, etc.) El Gráfico 1 muestra en porcentajes la distribución de varones, por grupos de edades, del Censo de Yugoslavia de 1960, comparada con las proporciones acumuladas a partir de la edad final. Resulta evidente la mayor suavidad de esta última disposición y la consiguiente mejora en la probabilidad de descubrir errores sistemáticos.

La decisión en cuanto a lo que constituye una divergencia inaceptable del patrón o tendencia será más fácil si se puede comparar con una distribu -

ción de referencia sino, tomando en cuenta las características de su forma, resaltar más claramente las anomalías. Con frecuencia conviene trabajar con las diferencias entre las medidas correspondientes a las observaciones y la distribución estándar o de referencia. Basándose en supuestos razonables, estas diferencias mostrarán una tendencia pareja, aparte de los errores fortuitos y sistemáticos. Las distribuciones estándares pueden construirse en diversas formas. En un extremo, puede tratarse de una simple modificación con el fin de hacer que las medidas se presenten en una forma que haga más fácil su inspección. Un ejemplo sería la ojiva transformada de Carrier-Farrag (1959) para examinar distribuciones relativas acumuladas por edad. En efecto, la distribución estándar es una función que cambia linealmente desde cero, al nacimiento, y concuerda con las observaciones a la edad de 60 años. Al restarla de los valores observados da un conjunto de diferencias cuyas variaciones se aprecian mejor. En el otro extremo, la distribución estándar puede provenir de una construcción más elaborada de modelos, por ejemplo, una población estable basada en una tabla modelo de mortalidad; podrá ser matemática o empírica, representar una norma general o referirse específicamente a una situación determinada. En esta última categoría está el caso relevante de distribuciones de referencia derivadas de otros datos correspondientes a la población que se estudia, como el uso de tasas de fecundidad obtenidas a partir de la información retrospectiva de un censo y usada como norma para corregir los registros de inscripción de nacimientos.

Estos métodos se ilustran con una aplicación a la distribución por edades de mujeres en las Islas Gilbert y Ellice, empadronadas en el censo de 1963. Se tomó como referencia una distribución por edades estable basada en una tabla modelo de mortalidad (para los fines actuales, no tienen importancia los detalles de su construcción). Las diferencias entre las proporciones observadas y acumuladas hasta cierta edad y las de la norma se han trazado en el Gráfico 2 A confrontándolas con los valores de referencia. Los puntos en el gráfico siguen un patrón relativamente simple que hace posible distinguir las divergencias sin dificultad. Trazando una curva suave a través de los puntos se podrían obtener resultados ajustados. Por supuesto, es posible que algunas de las desviaciones de esta curva sean reales más que deberse a un error, pero los efectos imputables a error sólo pueden reducirse mediante alguna distorsión de las curvas reales. Es importante hacer notar que aunque la norma de referencia es la de una población estable, la distribución por edades corregida por este medio está, en nuestro ejemplo, lejos de ser estable. Los requisitos de suavidad en las diferencias son mucho menos rígidos que el ceñirse a la forma del modelo estable.

Si bien es posible trazar una línea a través de los puntos con un grado regular de objetividad, el hecho de que las diferencias deban ser pequeñas en las edades bajas y altas hace probable que ocurra un grado considerable de curvatura. El ojo humano aprecia mejor una tendencia más suave, en cuyo caso resultan más eficaces métodos simples y objetivos para adecuar los puntos. Por consiguiente, se pueden lograr otras mejoras, si se consigue encontrar una transformación lineal, es decir, una función de las medidas que le dé a la relación que se traza una forma más cercana a una recta. Dicho recurso es especialmente útil cuando se trata de proporciones, dada la frecuencia con que se presenta, y por la necesaria reducción a cero en los valores extremos de las diferencias entre el valor observado y el valor de referencia.

Hay varias transformaciones que "amplían" la extensión de una proporción. He encontrado que el logito, definido como la mitad de $\ln(p/(1-p))$ donde p es la proporción, es lo más sencillo para trabajar. A medida que p pasa de cero a uno, el logito se traslada de menos infinito a más infinito. El logito ha dado muy buenos resultados en la transformación a formas lineales de las tasas de sobrevivencia, lx de las tablas de mortalidad, y sobre esta base se ha construido un modelo (Brass, 1967). En otras aplicaciones, no necesariamente endereza la curva con la misma intensidad, pero a menudo es útil para la inspección de los datos.

Las distribuciones por edad acumuladas, observadas y estándares que aparecen en el Gráfico 2 A se transformaron ambas con la función logito. Las diferencias entre los dos juegos de logitos para las respectivas edades se han trazado en el Gráfico 2 B, contrastándolos con la norma de referencia transformada. Los puntos siguen una curva suave en las edades de la niñez y se encuentran cerca de la linealidad en los años de madurez. Las desviaciones erráticas de la tendencia general son obvias y es posible efectuar la graduación con confianza, ya sea por una línea hecha a mano alzada o por algún método más objetivo, por ejemplo, adaptando la curva por secciones, o un promedio móvil ponderado. En este caso también, la distribución por edades corregida tendría una forma considerablemente diferente de la de los modelos estables en el sistema que se hubiera usado como referencia.

La transformación lineal cambia el tamaño de las desviaciones causadas por errores, en diversos grados, a lo largo de la escala. Como consecuencia, en algunos casos suele darse una impresión desproporcionada del peso relativo de los errores. Por ejemplo, la transformación logito de las proporciones por debajo de una edad determinada o de la relación de sobrevivencia aumenta el efecto de los errores en edades más avanzadas. Si éstos son grandes, puede que dominen el gráfico (al hacer la inspección o adaptación visual) hasta un grado inconveniente. Un recurso que a veces es útil es lo que yo he llamado transformación de escala, es decir, una disposición de las medidas de modo que las variaciones se expresen en una escala fácilmente comprensible. Una forma de hacer esto es invertir la presentación de la relación entre las observaciones y la distribución de referencia para mostrar las desviaciones en función de la edad en años (o tiempo) en que ambas concuerdan. Las desviaciones sobre el gráfico pueden entonces interpretarse en una forma obvia como errores promedio en datos de edad. Los resultados del procedimiento se ilustran en el Gráfico 2 C correspondiente a los datos por edad de las islas Gilbert y Ellice, que ya se usaron. Si, como en el presente ejemplo, no es razonable atribuir solamente a error las desviaciones de la distribución de referencia, puede adaptarse una línea a los puntos y determinar las correspondientes cantidades ajustadas. Sin embargo, en general se pierden las ventajas de adoptar una transformación lineal.

III. SERIES DE CENSOS

Gran parte de la aplicación del análisis demográfico en países subdesarrollados, estadísticamente hablando, ha estado referido a datos de un censo o estudio más o menos satisfactorio. En un grado creciente se disponen ya de materiales de dos o más censos, y esto significa que hay oportunidades para emplear técnicas más flexibles y completas que actualmente no están sufi-

cientemente aprovechados. Sugeriré algunos enfoques posibles y los ilustraré con base en la serie de censos realizados a intervalos de cinco años en Turquía. Demeny y Shorter (1968) han hecho un estudio excelente de la demografía básica de Turquía, que incluye distribuciones de la población por grupos quinquenales de edad a intervalos de cinco años, los que se han adecuado para tomar en cuenta la migración y los cambios territoriales. Los métodos de análisis que se describen resultan más simples con datos presentados en dicha forma. Es posible aplicarlos en otras circunstancias, por ejemplo con diferentes intervalos censales, pero la adaptación puede ser complicada aunque los principios sean los mismos. En esta breve exposición y para mayor claridad, solamente se discutirá el caso más simple.

En muchos países, los censos periódicos son el mejor medio para calcular la mortalidad. Partiendo de dos series sucesivas de distribuciones por edad (normalmente para cada sexo por separado), puede calcularse la proporción de personas en un grupo quinquenal de edad que sobrevive cinco años más tarde. Esto se hace dividiendo simplemente la población en un grupo quinquenal de edades para la primera fecha, entre los que aparecen en el grupo siguiente, en la fecha posterior. Puede escribirse B_{i+1}/A_i , donde A y B denotan las distribuciones por edad anterior y posterior, y la i el grupo de edad (desde 1, para 0-4 años hacia adelante). En la práctica, a causa de errores en la declaración de la edad, omisiones, etc., las B_{i+1}/A_i fluctúan considerablemente. En las edades más jóvenes, por lo menos, las fluctuaciones tienden a sobrepasar las proporciones de los que fallecen porque éstas son relativamente pequeñas. Las estimaciones se consiguen después de un riguroso suavizamiento. Por lo general no es suficiente o satisfactorio hacer correcciones previas en cada distribución por edades por separado, porque los supuestos que se hagan pueden tener un efecto considerable sobre las proporciones de sobrevivientes, lo cual depende de pequeñas diferencias entre cantidades grandes.

Demeny y Shorter, siguiendo lo propuesto por Coale y Demeny (1967) utilizan tablas modelo de mortalidad para realizar la graduación. Por medio de ensayos se escoge el sistema de tablas de mortalidad que mejor describa el número de personas que sobreviven desde la primera hasta la segunda fecha. En este caso, lo "mejor" se decide por la concordancia de los números observados con los números esperados mayores de 5, 10, ... 45 años.

Dentro del contexto del presente trabajo, el método tiene la desventaja de que se apoya en un sistema de modelo rígido. En un caso particular, es probable que se necesite dicha rigidez dada la magnitud de los errores, pero esto debería desprenderse de las observaciones en vez de predeterminarse. Si se usa un sistema flexible de tablas modelo de mortalidad con más de un grado de libertad --por ejemplo, el que se basa en las relaciones funcionales logito-- el procedimiento de hacer ensayos es sumamente laborioso. Un método simple que podría servir como alternativa sería construir, a partir de los cocientes B_{i+1}/A_i , la distribución por edades de una población estacionaria no ajustada. Comenzando con un número arbitrario S_1 correspondiente al grupo de edades de 0-4 años, al multiplicarlos por B_2/A_1 se obtendrá la estimación de S_2 , la población con edades de 5-9 años; S_2 se multiplica por B_3/A_2 para dar S_3 , y así sucesivamente. El resultado es una población estacionaria.

ria que correspondé al esquema estipulado de la mortalidad por edad, pero aún lleva incorporada la estructura de error de B_{i+1}/A_i . Sí podría ahora disciplinarse por medio de las técnicas indicadas en la Sección II, empleando una distribución por edad estacionaria basada en una tabla modelo de mortalidad como norma de referencia. Una vez hecha la graduación por estos medios, se pueden obtener los índices de mortalidad correspondientes a las mediciones ajustadas, por diversos procedimientos de reversión.

Las distribuciones por edad, por sí solas, pueden dar muy pocas evidencias de la mortalidad de la primera infancia ya que buena parte de esto ocurre en un período demasiado corto después del nacimiento como para que sus efectos sobre los cómputos se separen de los del nivel de fecundidad. No es confiable la estimación de la mortalidad temprana partiendo de tasas de mortalidad a edades mayores, que es, en efecto, lo que implica la aplicación de tablas modelo de mortalidad a este problema, porque la relación no es lo suficientemente fuerte.

Para contar con buenos resultados es necesario obtener mediciones de la mortalidad infantil de otras fuentes tales como informes retrospectivos de muertes de niños en una encuesta o censo. Si se dispone de información de esta clase, S_i , el número de personas entre 0-4 años en la población estacionaria, en vez de ser elegida arbitrariamente puede calcularse a partir de una raíz fija de nacimientos. La población esperada en cada grupo de edades es el producto de esta raíz por la integral de l_x (la probabilidad de supervivencia en la tabla de mortalidad) en el intervalo. Dividiendo entre el producto de la raíz y la amplitud del intervalo (en este caso cinco años) se obtiene la medida que, para efectos prácticos se puede tomar como el valor de l_x en el punto medio del grupo, (excepto para 0-4 años, donde la mortalidad cambia rápidamente con la edad), aunque por supuesto, está expuesto a error. Los logitos de estos valores l_x pueden entonces compararse con los correspondientes valores para una distribución de referencia. La selección obvia es una tabla estándar de mortalidad del sistema logito, ya que hay pruebas de peso de que la relación debería entonces estar cerca de la lineal. La evaluación y la compensación del error puede hacerse entonces trazando las diferencias y contrastándolas con la norma de referencia. El procedimiento descrito se ha aplicado a los datos de hombres de Turquía correspondientes al período 1955-60 con los resultados que se indican en el Gráfico 3A (la medida de referencia a los 0-4 años es el logito de los promedios de l_x por encima del intervalo). La información sobre la cual debe tomarse la decisión acerca de la graduación se indica aquí claramente, así se escoja o no un modelo del sistema logito, representado en el gráfico por una línea recta a través del punto para 0-4 años.

Es interesante especular sobre cómo se afectaría el método mediante un cambio en la completitud de la cobertura entre los dos censos ya que dichos efectos, que no son poco comunes, resultan especialmente difíciles de evaluar. Si la diferencia en cobertura fuera claramente independiente de la edad, los cocientes B_{i+1}/A_i estarían todos modificados por el mismo factor constante. La distorsión de las proporciones de los que fallecen sería relativamente mayor en las edades más jóvenes que en las edades más viejas. Sería de esperar, por lo tanto, que la consecuencia en una representación como

el Gráfico 3 A fuera un aumento de la curvatura en la relación. Que eso es así, puede verse en el Gráfico 3 B, en el cual se usaron los mismos datos para Turquía pero con la imposición de una omisión relativa del 3 por ciento en la distribución por edades de 1955, comparada con la de 1960. En esa forma se alteró la tasa de crecimiento de la población del 2.7 por ciento, hasta una aparente de 3.4 por ciento anual. Teóricamente al menos, al examinar la supervivencia puede dar alguna luz sobre la posibilidad de cambios en la cobertura, pero cabe la duda en cuanto a lo sensible que podría ser el método; fluctuaciones en los errores, omisiones diferenciales por edad y desviaciones del comportamiento lineal en la verdadera relación de los logito podrían complicar la evaluación. Se necesita un estudio empírico adicional.

Cuando los únicos datos sobre la distribución por edades provienen de un solo censo, no hay forma de decidir hasta qué punto una divergencia de un patrón nítido en determinados grupos, se debe a error o a verdaderas fluctuaciones. Con una serie de censos, el alcance para examinar esta cuestión es mucho mayor. Es obvio que si los cómputos en intervalos quinquenales de edad que se refieren a una determinada cohorte de nacimientos parecen ser demasiado grandes en varios censos sucesivos, hay una buena base para considerar que el exceso es verdadero; si se obtiene un resultado similar correspondiente a un intervalo fijo de edad (y por consiguiente, diferentes cohortes) ello sugeriría errores en las edades.

Con frecuencia, esta idea general se ha usado para la adaptación de distribuciones por edades, pero generalmente en una forma arbitraria e intuitiva sin una clara especificación de los supuestos. Demeny y Shorter aplican en las distribuciones por edades de Turquía una técnica ingeniosa y objetiva para separar las fluctuaciones reales de las espurias. El principal supuesto es que, en lo que se refiere a un grupo de edades determinado, el efecto proporcional del error es igual en dos censos sucesivos. La población en el grupo i se puede escribir entonces como $A_i(1 + E_i)$ y $B_i(1 + E_i)$ en los dos censos, donde A_i y B_i son los valores reales y E_i es el error proporcional. Las relaciones de sobrevivencia para intervalos de cinco años se consideran que han sido determinadas con exactitud por un procedimiento de estimación del tipo que ya se ha descrito. El aplicar éstas al primer censo dará el juego de valores $B_{i+1}(1 + E_i)$ que son los verdaderos guarismos para el siguiente grupo de edades, pero con el error del grupo inicial. Existen entonces dos valores desconocidos para cada i (B_i y E_i) pero también dos ecuaciones para cada uno con excepción de B_1 .

La introducción de la condición adicional de que la suma de B_i y $B_i(1 + E_i)$ son iguales (es decir, que sólo se toman en consideración errores relativos) permite encontrar soluciones únicas para E_i y B_i (y por consiguiente, también A_i). Demeny y Shorter dan un método simple para llevar a cabo los cálculos.

A través de las series de censos de Turquía, la técnica da resultados uniformes y convincentes, pero existen dudas sobre si puede aplicarse a otras poblaciones, especialmente donde los intervalos de censos son mayores y los niveles educativos están subiendo rápidamente. ¿Será razonable la hipótesis de que los errores proporcionales siguen siendo iguales? ¿Podrán usarse y justificarse otros supuestos? Los errores se podrían tomar como E_i en el

primer censo y $k_j E_j$ en el segundo, donde k_j es cualquier conjunto de constantes especificadas y, en principio, se podría obtener una solución. En realidad, los cálculos serían muy laboriosos. Una idea más práctica sería hacer que todas las k_j fueran iguales a k , es decir, suponer que cambiaran todos los errores, (normalmente reducidos) en la misma proporción. No parece haber un método relativamente simple para resolver las ecuaciones resultantes cuando k no es igual a uno pero se puede hacer por procedimientos iterativos. Cuando consideramos que k es igual a un medio (por ejemplo) en la aplicación de dos distribuciones por edad sucesivas en los datos turcos, las estimaciones resultantes de A_j , B_j y E_j no muestran características que no sean razonables. Aparentemente, no hay ninguna forma de escoger entre los valores obtenidos sobre los muy diferentes supuestos de que los errores permanecen iguales o se reducen a la mitad. No he podido encontrar ningún elemento de juicio basado en la suavidad o en la minimización de las desviaciones, que lo haga exitosamente.

La pregunta obvia es qué sucede si suponemos mal. Se pueden encontrar expresiones algebraicas para las distorsiones resultantes, pero son complicadas series ponderadas de E_j y difíciles de interpretar. Un enfoque más directo es hacer cálculos representativos. Esto se ha hecho tomando A_j , B_j y E_j calculados por Demeny y Shorter con relación a las distribuciones por edades de 1955 y 1960 de Turquía como base. Partiendo de éstas se construyeron conjuntos artificiales de datos imponiendo cambios en el error entre censos y también proyectando hacia atrás para darse una distribución hipotética para 1950. Luego se calcularon los valores de E_j en el supuesto erróneo de que los errores, en efecto, no habían variado entre los censos. El Gráfico 4 muestra, con relación a dos ejemplos, comparaciones entre los errores estimados y los que se habían incorporado en los datos. En el conjunto, los resultados fueron reconfortantes. En promedio, y aunque hay raras excepciones, las estimaciones tienden a caer entre los errores de ambos censos. Parece justo aceptar que la técnica de ajuste de Demeny-Shorter mejorará, por lo general, la estimación de la distribución por edades aunque haya habido cambios apreciables en la magnitud del error entre dos censos, siempre que el patrón siga siendo parecido.

La dificultad en el caso de dos censos es que hay apenas suficientes observaciones para hacer posible la estimación de valores "verdaderos" y errores si se conocen las relaciones de estos componentes entre las distribuciones por edades. Si hay un tercer censo, eso hace posible comprobar los supuestos. Los recuentos de grupos de edades en los tres censos se podrían escribir $A_j(1 + E_j)$, $B_j(1 + k_j E_j)$, $C_j(1 + h_j E_j)$ donde k_j y h_j son constantes que describen los cambios en los errores. Mediante la aplicación de las probabilidades de sobrevivencia conocidas, se obtienen ecuaciones múltiples que pueden resolverse en teoría poniendo ciertas restricciones a h_j y k_j ; por ejemplo, que $1 - k_j = k_j - h_j$ esto es, que el cambio en el error es igual en los dos intervalos quinquenales. Los cálculos resultarían muy engorrosos y ciertamente no deberían intentarse sin tratar de alcanzar resultados preliminares, cuando menos mediante un enfoque más simple. El procedimiento sugerido es, tratar primero los censos en dos pares sucesivos con la técnica de Demeny-Shorter. Se obtendrán dos conjuntos de estimaciones de error, digamos E_j^1 y

E_i^{II} para la distribución media por edad que aparece en ambos pares. El error se toma entonces como la media aritmética de los dos. El ajuste por medio de éstas da la "verdadera" distribución por edades. Al proyectarla hacia adelante y hacia atrás aplicando las probabilidades de sobrevivencia para intervalos de cinco años proporcionará las distribuciones por edades ya corregidas correspondientes a las fechas anteriores y posteriores, y por consiguiente, los errores a través de la comparación con los valores observados. El método se basa en la creencia de que E_i^I y E_i^{II} , al menos en promedio, estarán más o menos en el medio de los errores correspondientes a los pares respectivos de los censos, y que su media aritmética será aproximadamente la mitad del error correspondiente a la distribución central por edades más una cuarta parte de los errores en cada uno de los dos. Con una tendencia uniforme, esto no debe estar lejos del valor buscado.

El Gráfico 5 compara los errores estimados con los reales correspondientes a una serie de censos preparados artificialmente con los datos de Turquía de 1955 y 1960. No obstante algunas aberraciones, existe un promedio aceptable de concordancia entre los valores derivados y los construidos. En vista de las posibles discrepancias entre las verdaderas probabilidades de sobrevivencia y las estimadas, y la probable inconsistencia en el patrón de cambio de los errores (incluyendo, como en realidad ocurre, los efectos de omisiones, ajustes erróneos para la migración, etc., así como mala declaración de la edad), no parece estar justificada una técnica más refinada y laboriosa para corregir las distribuciones por edades. Con las observaciones actuales, donde no es posible la comparación entre los errores calculados y los errores verdaderos, la representación de las estimaciones del tipo del Gráfico 5 dará indicios en cuanto a la compatibilidad y racionalidad de los resultados en relación con los supuestos.

IV. HISTORIAS REPRODUCTIVAS DE MUJERES, AL MOMENTO DEL REGISTRO DE UN NACIMIENTO

En los lugares donde la inscripción o notificación de nacimientos es incompleta, es difícil sacar medidas útiles de la información numérica, sean cuales fueren las categorías de las sub-divisiones. Los supuestos de que el grado de sub-registro es igual en diferentes segmentos de la población, (por área, ocupación a lo largo del tiempo, etc.) son siempre especulativos y con frecuencia no justificados. Hasta los índices relativos o proporcionales dan lugar a dudas. No obstante los esfuerzos por extender y mejorar los registros, es probable que en muchos lugares éstos serán defectuosos por muchos años. El intento de obtener el mayor provecho posible de los registros existentes, por lo tanto, está justificado.

La forma más prometedora de hacer esto es recopilando más información sobre la historia reproductiva de la madre al momento de la inscripción o notificación de un nacimiento, en especial sobre el número de sus otros hijos, nacidos y sobrevivientes. Aunque se han hecho sugerencias con respecto a ese tipo de enfoque, no parece haberse elaborado ningún programa importante. Un

proyecto experimental de notificación de nacimientos, incluyendo el registro de historiales de reproducción, ha sido iniciado por el Dr. J.D. McGregor, Director de Servicios Médicos, en las Islas Salomón (Británicas). Examinaré algunos de los problemas de análisis con referencia a los registros del proyecto correspondientes a 1967, que abarcaban alrededor de 1700 nacimientos (del 25 al 30 por ciento del total) e informes sobre otros 4 600 niños nacidos vivos previamente.

Intuitivamente, parece claro que el orden del nacimiento debe estar relacionado con la fecundidad; el número medio de hijos nacidos de madres que dan a luz en un período determinado parece un posible índice de fecundidad - que no depende directamente de la integridad del registro. En efecto, índices de este tipo han sido usados por El-Badry (1967) con relación a estudios de diferenciales en la ciudad de Bombay. Sin embargo, las relaciones de tales medidas con las corrientes distan mucho de ser sencillas. Las madres de una edad determinada que tienen un hijo en un año no son, en general, una muestra aleatoria de las madres de esa edad en la población. Existen sesgos obvios, tales como la omisión de mujeres que se han convertido en estériles a una edad más joven, y la sobrerrepresentación de las que tienen una fecundidad sobre el promedio, así como otros más sutiles.

Entre éstos se encuentran variaciones en la fecundidad específica a una edad determinada con la duración del matrimonio y la inclusión de mujeres en el momento en que ocurre el nacimiento en lugar de ubicarlo al azar con relación al mismo. La característica sorprendente es que, al menos en poblaciones de alta fecundidad, estos sesgos parecen tener relativamente poco efecto. El número medio de niños nacidos vivos previamente (con exclusión del nacimiento actual) parece ser similar a los correspondientes valores de madres en la población. Esto se ilustra en el cuadro 1 que muestra, para las Islas Salomón (Británicas), los nacimientos de niños por madre al momento de la notificación y también los promedios 5, 10 y 15 años antes en comparación con las medidas de la población del censo de 1959.

Al examinar los modelos de fecundidad, matemáticamente y por simulación de computadoras, pareciera que la razón de la aceptable concordancia es el predominio de los factores fortuitos en las probabilidades de nacimientos sobre las variaciones entre mujeres. En las edades más avanzadas de maternidad, el sesgo ocasionado por la exclusión de mujeres estériles indudablemente se torna considerable pero la proporción de nacimientos actuales en el grupo es pequeña. La limitada influencia del sesgo lleva a la conclusión de que el número medio de hijos anteriores puede usarse satisfactoriamente como índice comparativo de fecundidad para subgrupos de población, donde los factores que ejercen su influencia pueden aceptarse plausiblemente como similares. Los estudios realizados por El-Badry encuentran estas condiciones.

La conversión de dichos índices en medidas convencionales de fecundidad es, sin embargo, algo mucho más dudoso; cuando las observaciones se tratan como representativas de toda la población de mujeres, el error es demasiado incierto y difícil de evaluar para poder sacar mediciones absolutas con confianza. Hay una relación más sencilla que puede usarse para estimar la fecundidad con base en datos de orden de nacimientos (Brass, 1954). Cuando

los datos se estandarizan a una población con igual número de mujeres en cada grupo quinquenal de edad del período reproductivo, la relación del total de nacimientos con los de primer orden, en el año, es igual a la relación total de fecundidad dividida por la proporción de mujeres que se convierten en madres a las tasas de natalidad actuales. Esto es, en sí, un significativo índice de fecundidad correspondiente, si las tasas se mantienen constantes al tamaño medio de familia completa de madres.

La proporción de mujeres que tienen por lo menos un hijo en el período reproductivo generalmente se puede estimar con relativa exactitud, y su uso como multiplicador dará la relación total de fecundidad. La desventaja del método es que depende en alto grado del registro de los primeros nacimientos, los cuales son especialmente vulnerables a fluctuaciones de efectos periódicos y fortuitos y también a declaración diferencial. La información sobre la distribución del orden de los nacimientos se puede utilizar para vencer estos obstáculos. La idea es ajustar una curva apropiada a las cifras según orden, de lo cual es posible estimar los valores de primeros nacimientos. Se han propuesto funciones teóricas para tal curva, pero las técnicas de ajuste son complicadas y su eficacia para el fin que perseguimos es dudosa. El enfoque que se describe en la Sección II conduce a un procedimiento más simple y más informativo, el cual se ilustra aplicándolo a los registros de las Islas Salomón.

Tomando como base el Censo de las Islas de 1959, se estimó aproximadamente la distribución por edades de las mujeres en el período de procreación a mediados de 1967. Se calcularon factores para elevar los valores en cada grupo quinquenal de edad al mismo nivel que el de los años 15-19. La aplicación de éstos al número de nacimientos por orden para grupos convenientes de edad de la madre, y sumando a través del lapso de procreación dio la distribución estandarizada cuando los resultados se expresaban como proporciones del total de nacimientos ajustados. Se construyó una distribución de referencia con base en los informes de las mujeres de 40-49 años de edad en el censo, del número total de niños nacidos vivos de dichas mujeres. La acumulación dio el número de mujeres con uno o más, dos o más, tres o más hijos, y así sucesivamente, hasta el tamaño máximo de familia. Estos valores, calculados como una proporción de suma, fueron los términos de la escala empírica de referencia.

En el Gráfico 6 se traza la distribución estandarizada por orden de nacimiento contra los valores de referencia. Se ve inmediatamente que el punto de los primeros nacimientos muestra una gran discrepancia con respecto a la tendencia uniforme seguida por los otros. Semejante distorsión podría surgir de varias causas; la más plausible por gran margen es el hecho de que los informes son más completos con relación al primer nacimiento que con respecto a los subsiguientes. No hay indicación de que las diferencias en las inscripciones se extiendan a otros órdenes de nacimiento y, hasta donde llega el grado de exactitud de las observaciones, los puntos representativos no se apartan visiblemente de una línea recta. Entonces el porcentaje de primeros nacimientos se puede ajustar para corresponder a los otros valores adaptando numéricamente una línea recta a dichos valores, o al ojo. La medición resultante de primeros nacimiento leída de la línea, da estimaciones del tamaño

promedio de la familia completa de madres y tasas totales de fecundidad, que serán mucho más lógicas que los niveles sin ajustar.

La evaluación de cuán razonables son los índices derivados como estimaciones correspondientes a la población de las Islas Salomón está fuera de los alcances del presente ensayo. La opinión sobre los resultados obtenidos en estudios particulares dependerá de cómo se recopilaron los datos, su integridad y las características de la población. Aunque en la ilustración el sesgo diferencial sobresaliente está en la inscripción de los primeros nacimientos, no se desprende de ello que hayan otros errores o que siempre ocurrirá la misma configuración. Los nacimientos notificados son un subgrupo seleccionado y el supuesto de que son representativos en cualquier característica determinada no se puede hacer sin justificación. Aún donde exista algún sesgo residual, se podría aplicar la técnica provechosamente para hacer comparaciones entre subgrupos y a través del tiempo.

Aunque en el ejemplo podría fácilmente efectuarse la estandarización por edades y la construcción de una distribución de referencia usando los datos del censo de 1959 de la población particular, ésta no es una parte esencial de la técnica. Por lo general será posible estimar factores de estandarización de edades lo suficientemente exactos con base en un modelo aproximado, y derivar la escala de referencia de cualquier distribución similar de mujeres de fecundidad completa por número de niños nacidos vivos. El objeto es obtener una relación gráfica que tienda a una recta con el fin de poder ver claramente y ajustar las anomalías por medios sencillos. En algunas circunstancias la transformación (por ejemplo, mediante el logito) puede ser útil para examinar órdenes de nacimientos más altos donde las proporciones son pequeñas.

Los informes retrospectivos de hijos nacidos y muertos, por madres, con preguntas de tipo censal son una fuente de información valiosa sobre la mortalidad. Se han ideado y usado extensamente métodos para convertir las observaciones en mediciones convencionales (Brass et al. 1968). La aplicación de las mismas técnicas a informes de muertes de niños en el pasado, obtenidos de madres al momento del nuevo nacimiento, se justifica si se admite la no aleatoriedad de la fecha de inscripción con respecto al espaciamento de los nacimientos. Aparte de esto, el supuesto es que para cada grupo de edad de estas madres, la distribución de intervalos de nacimientos anteriores de niños vivos es efectivamente igual al del grupo correspondiente en la población. La naturaleza de los sesgos en el número de nacimientos ya se comentó anteriormente. Con respecto a los intervalos se pueden hacer observaciones similares, pero como nos preocupa el número relativo de nacimientos por edad más bien que niveles absolutos, es probable que sean menores los efectos del sesgo. En efecto, al hacer el cálculo directo de los datos de las Islas Salomón (Británicas), usando la distribución de intervalos tomados de informes de nacimientos por períodos, se ve que el error introducido por el supuesto es pequeño. Sin embargo, en cualquier punto fijo en el tiempo, las madres de la población se distribuirán a lo largo de un intervalo entre nacimientos y no se concentrarán en un punto extremo. Los datos de madres de nacimientos actuales se pueden hacer aproximadamente equivalentes introduciendo un desplazamiento de medio intervalo.

En el análisis de datos de encuestas se usan factores multiplicadores para derivar probabilidades de morir hasta la edad de uno, dos, tres, cinco, diez años, etc., partiendo de proporciones de hijos muertos a madres en los grupos de 15-19, 20-24 años, etc. respectivamente. Para el caso de madres de nacimientos inscritos se aplica el mismo procedimiento, pero las probabilidades de morir deben referirse a edades, medio intervalo de nacimiento más alto. Puesto que las edades ahora no serán, en general, números redondos, se destruye la nitidez original del método pero se pueden obtener valores interpolados suavizados por referencia a un estándar en la escala logito. En el Gráfico 7 se comparan las probabilidades de morir calculadas con base en los datos de las Islas Salomón por el procedimiento descrito, con valores correspondientes de una estándar después de la transformación logito. El intervalo de nacimiento, estimado sobre la base de la tasa de cambio en el número medio de niños según edad de la madre, durante la parte principal del período reproductivo, fue de 3.33 años.

Hay otro posible enfoque de considerable interés que también se aplicó a estos registros. Las tabulaciones que se necesitan son de nacimientos y muertes de niños, ordenados por el número de intervalos que preceden el intervalo en curso, es decir, del nacimiento anterior, uno antes del anterior, y así sucesivamente. El tiempo medio desde cada una de estas categorías de nacimientos se conoce aproximadamente por la extensión del intervalo. Para los tiempos medios más largos, las proporciones muertas corresponden estrechamente con las probabilidades de morir en los intervalos apropiados desde el nacimiento. Para períodos más cortos hay que tomar en cuenta el cambio rápido en las tasas de mortalidad con la edad en los primeros años de vida, conjuntamente con la asimetría de la distribución de intervalos de nacimientos. El efecto es transformar la proporción de fallecidos en una categoría dada, en la probabilidad de morir hasta una edad algo más temprana que la correspondiente a la mitad del período de los nacimientos. Partiendo de distribuciones de intervalos de nacimientos que representan las características de las Islas Salomón, se obtuvieron edades correspondientes de 2.47 y 5.80 años para el nacimiento precedente y el anterior a ese, comparados con los tiempos medios de 3.33 y 6.66. Para tres intervalos y más, el ajuste tuvo un efecto tan insignificante que no se tomó en cuenta. Las probabilidades de morir estimadas en esta forma también se compararon en el gráfico 7 con el estándar de referencia en la escala logito.

Cada uno de los dos conjuntos de puntos en el gráfico es razonablemente consistente, tomando en consideración el número un tanto pequeño de muertes de las cuales se derivan las medidas. Las probabilidades de morir que se encontraron disponiendo los nacimientos de acuerdo con intervalos anteriores - tienden, sin embargo, a ser más bajas que el conjunto alternativo. La interpretación tiene que ser provisional en vista de lo limitado de la experiencia en aplicar el procedimiento y también la posibilidad de efectos fortuitos. Sin embargo, hay una explicación plausible de algún interés. En el método de intervalos, especialmente de períodos más cortos, los tiempos hasta el nacimiento están más estrechamente dispersos alrededor de la exposición media al riesgo que en el otro enfoque. Se puede demostrar que, como resultado, si las tasas de mortalidad infantil hubieran estado bajando en los años anteriores a la inscripción, el método de intervalos daría estimaciones más

bajas de mortalidad. El hecho de que tales descensos en las tasas de mortalidad estén ocurriendo en las Islas Salomón lo sugiere también el rápido aumento en las probabilidades de morir, por edades, según lo muestran las tendencias de los puntos en el gráfico. Es posible, por lo tanto, que al hacer comparaciones de las mediciones obtenidas por los dos métodos nos den pruebas útiles adicionales acerca de las cambiantes tasas de mortalidad. Aunque los procedimientos descritos son factibles, no resultan convenientes a causa de que los períodos de exposición al riesgo son difíciles de manejar y requieren interpolación en la tabla de mortalidad de referencia y del laborioso cálculo ad-hoc de las correspondientes edades para estimar las probabilidades de morir en la técnica de intervalos. Se necesita mayor refinamiento y simplificación.

Cuadro 1.

COMPARACION DEL PROMEDIO DE NIÑOS NACIDOS POR MUJER, CONFORME A LO INFORMADO EN EL CENSO Y EN EL REGISTRO DE NACIMIENTOS

Medición cuando la madre tenía la edad	Censo 1959	Inscripción de nacimientos de 1967 conforme se informó hasta años antes			
		0	5	10	15
15 - 19	0.16	0.20	0.13	0.15	0.22
20 - 24	1.29	0.98	1.05	1.14	0.78
25 - 29	2.84	2.50	2.82	2.17	2.26
30 - 34	4.67	4.28	3.85	3.90	
35 - 39	5.45	5.30	5.68		
40 - 44	6.40	6.86			

Gráfico 1.

PUERTO RICO 1960: MUJERES

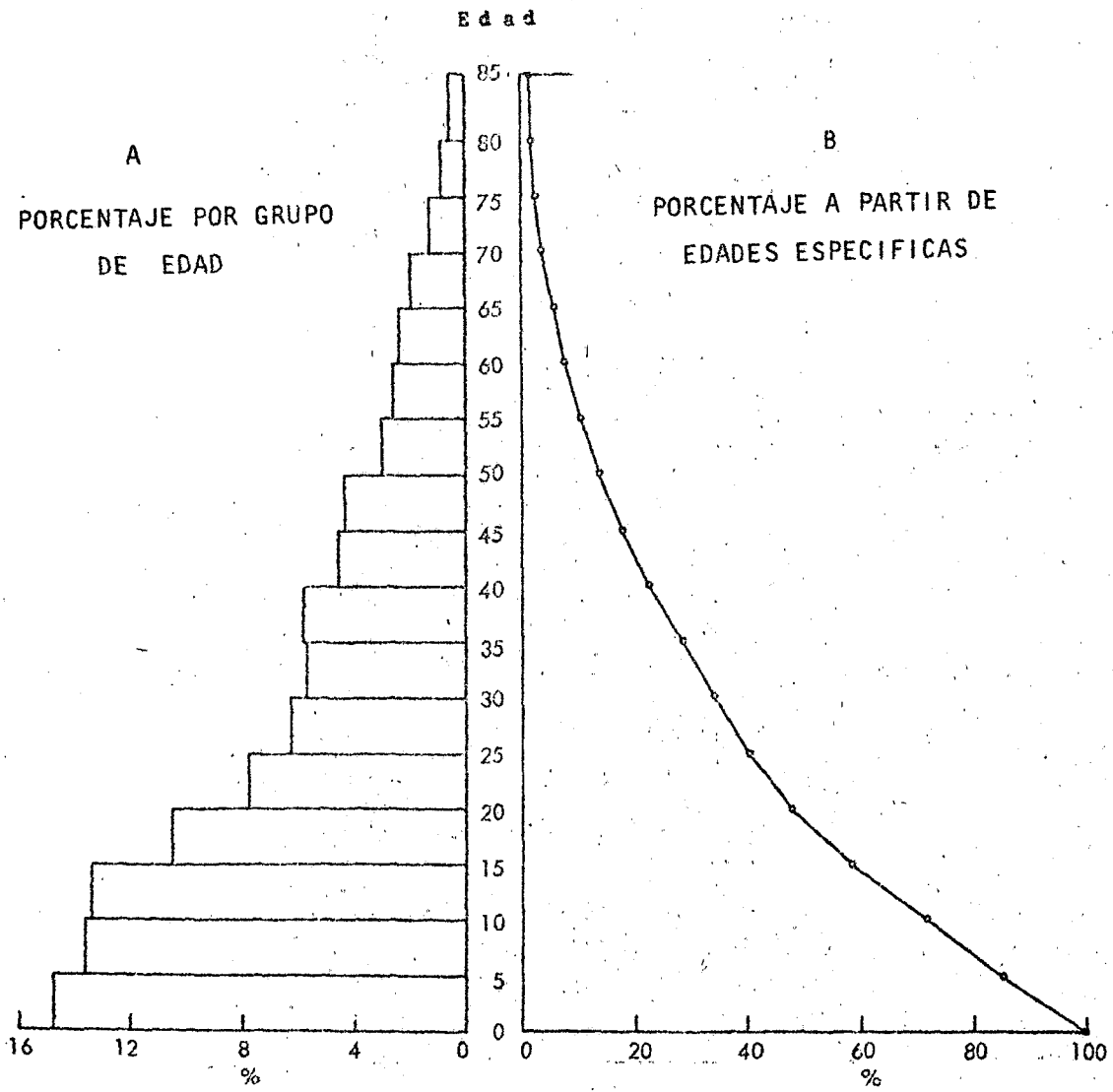


Gráfico 2 A.

DIFERENCIAS ENTRE PORCENTAJES ESTANDARES Y OBSERVADOS
HASTA CIERTA EDAD

Islas Gilbert y Ellice 1963: Mujeres

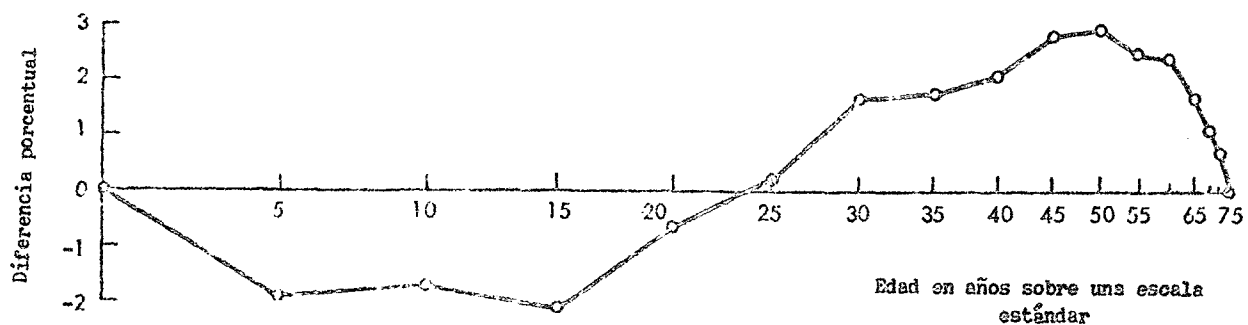


Gráfico 2 B.

DIFERENCIAS ENTRE LOGITOS DE PORCENTAJES ESTANDARES Y OBSERVADOS
HASTA CIERTA EDAD

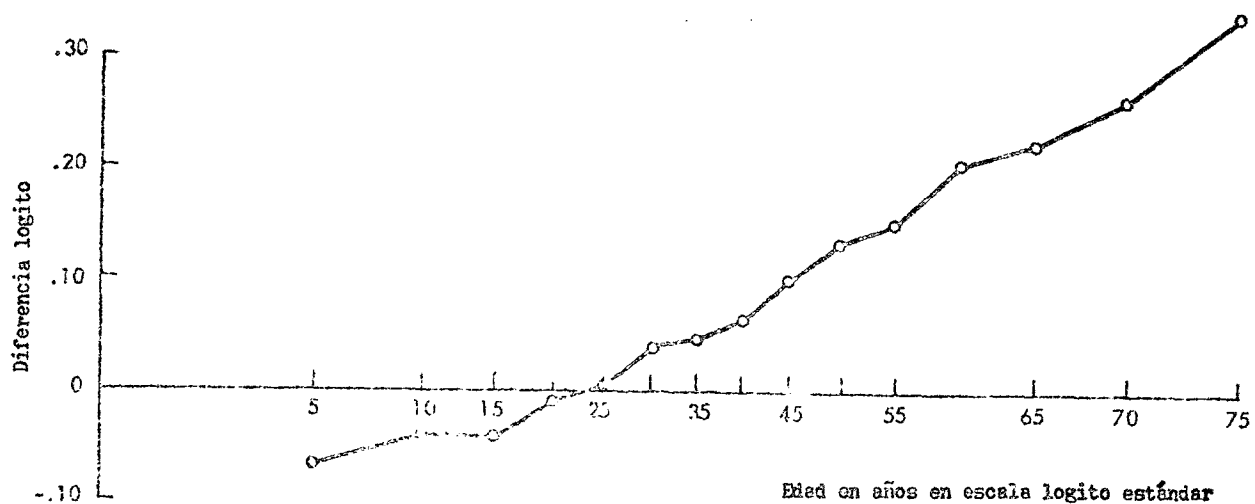


Gráfico 2 C.

ERRORES PROMEDIO EN EDADES INFORMADAS PARA DISTRIBUCION
OBSERVADA IGUAL A LA ESTANDAR

Islas Gilbert y Ellice 1963: Mujeres

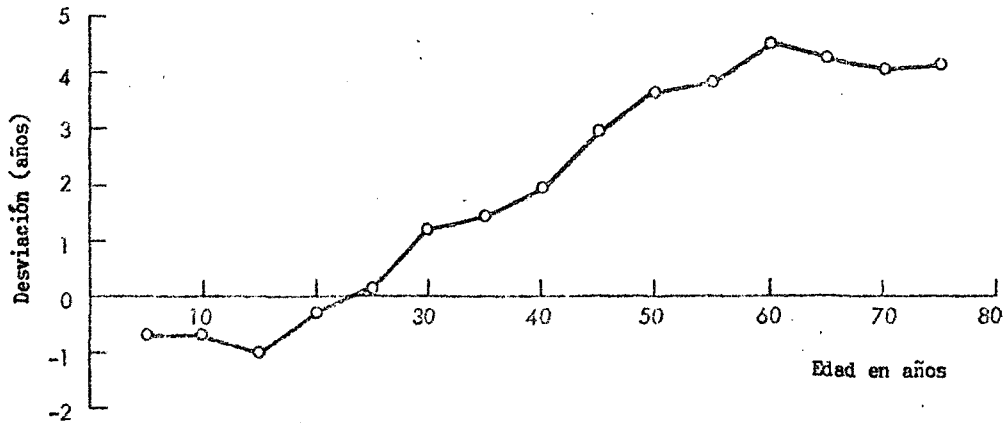


Gráfico 3 A.

COMPARACION DE TASAS DE SOBREVIVENCIA, SEGUN LA TABLA DE MORTALIDAD
ESTIMADA CON LA ESTANDAR

Turquía 1955-60: Varones

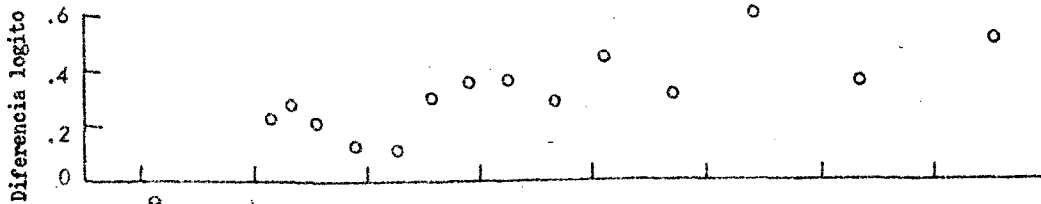


Gráfico 3 B.

COMPARACION DE TASAS DE SOBREVIVENCIA, SEGUN LA TABLA DE MORTALIDAD ESTIMADA,
PARA OMISIONES DEL 3% EN CENSOS ANTERIORES, CON LA ESTANDAR

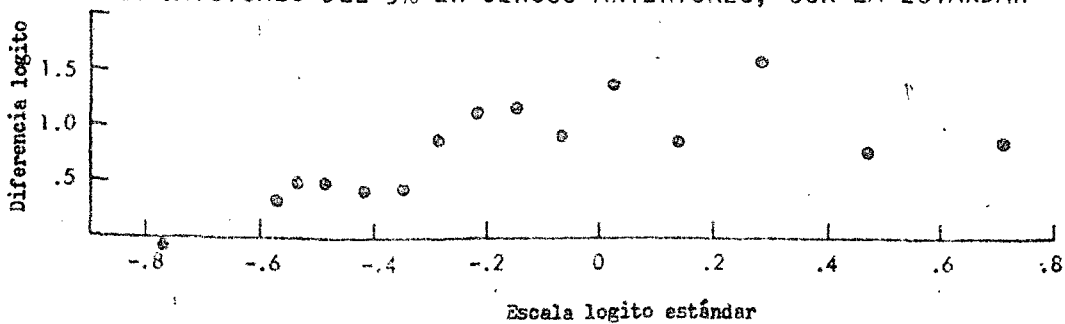


Gráfico 4 A.
 ERROR REDUCIDO A DOS TERCIOS EN EL SEGUNDO CENSO

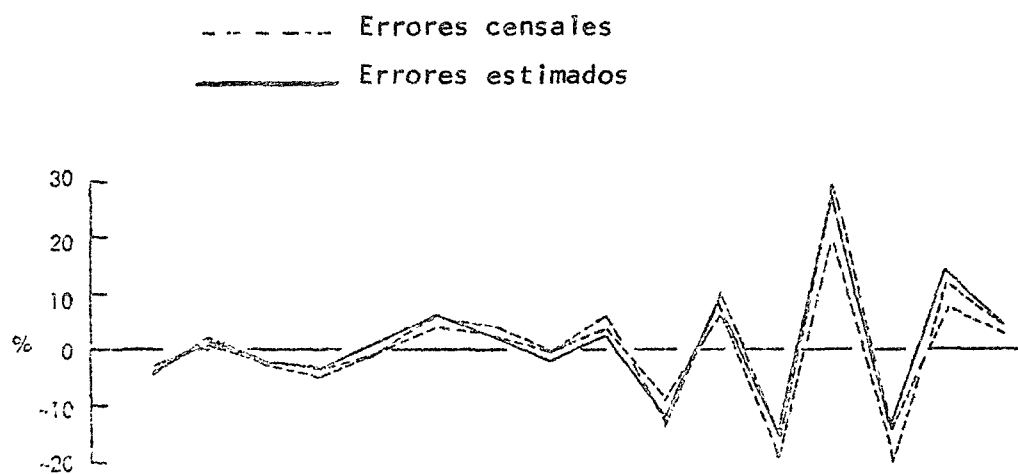
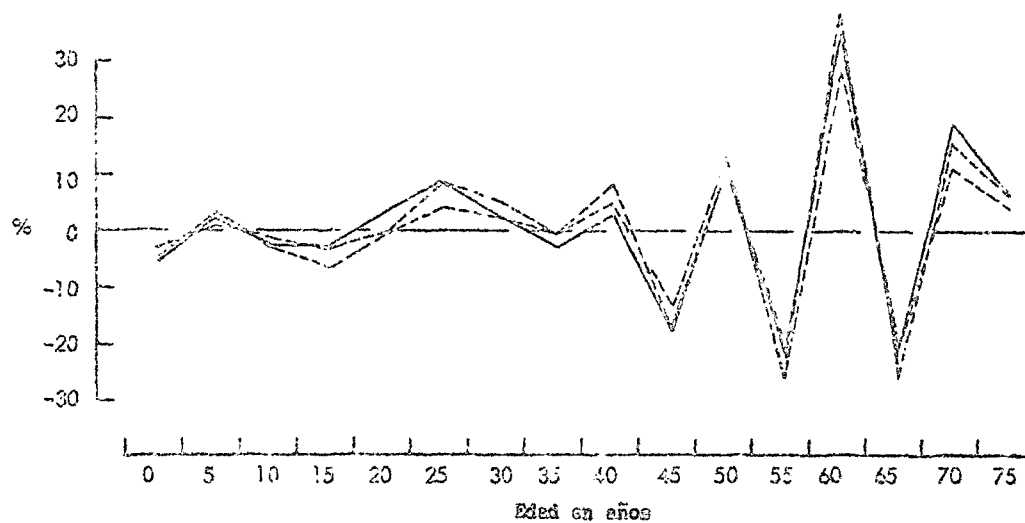


Gráfico 4 B.
 MENOS DE 40 AÑOS: ERROR REDUCIDO A LA MITAD EN EL SEGUNDO CENSO
 MAS DE 40 AÑOS: ERROR REDUCIDO A TRES CUARTOS EN EL SEGUNDO CENSO

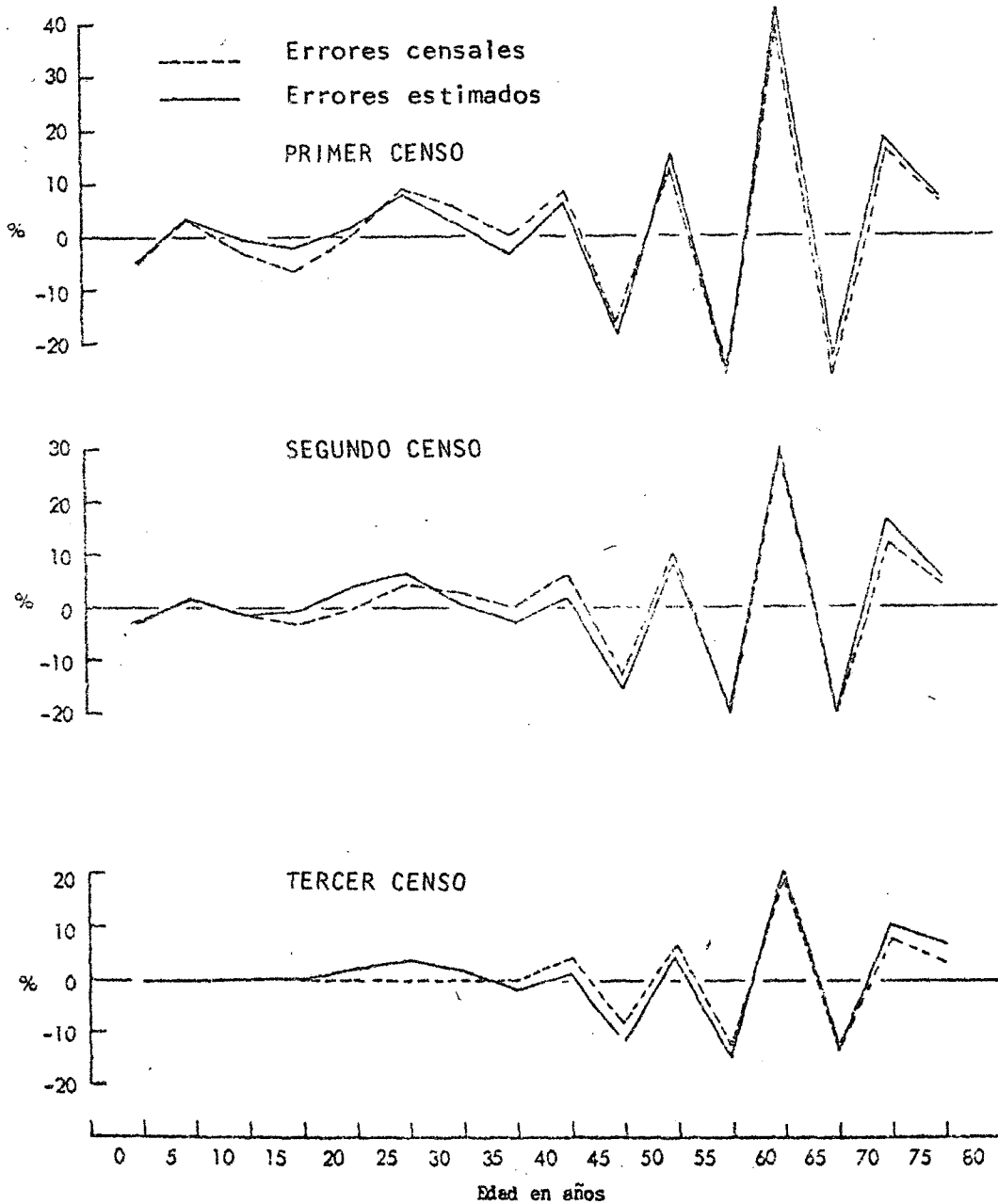


Corrección de la distribución por edades, de dos censos sucesivos.

Gráfico 5.

MENOS DE 40 AÑOS: ERROR REDUCIDO A LA MITAD EN EL SEGUNDO CENSO Y A CERO EN EL TERCERO

MAS DE 40 AÑOS: ERROR REDUCIDO A TRES CUARTAS PARTES EN EL SEGUNDO CENSO Y A LA MITAD EN EL TERCERO



Corrección de la distribución por edades, de tres censos sucesivos.

Gráfico 6.

COMPARACION DE LA DISTRIBUCION DE ORDEN DE NACIMIENTO
CON LA ESTANDAR

Islas Salomón Británicas: 1967

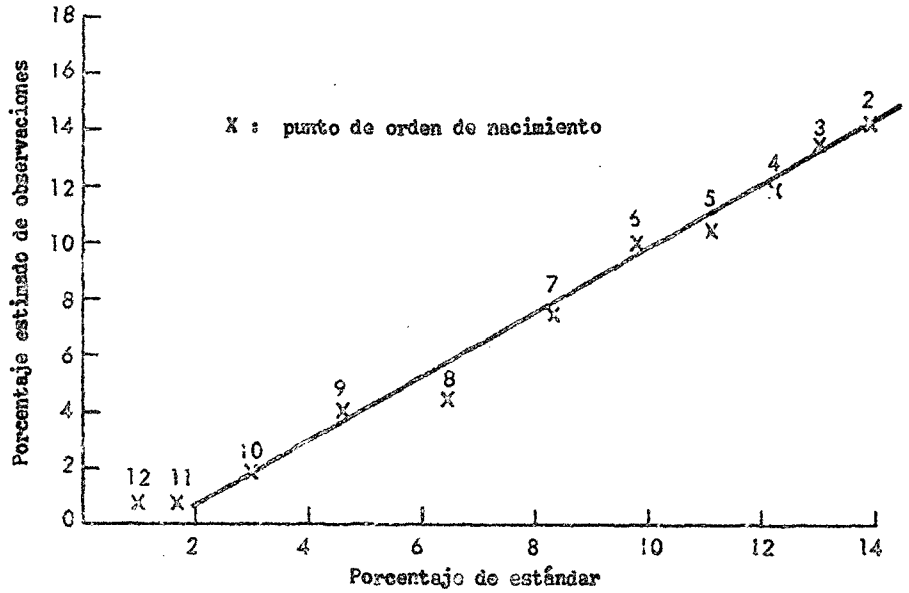


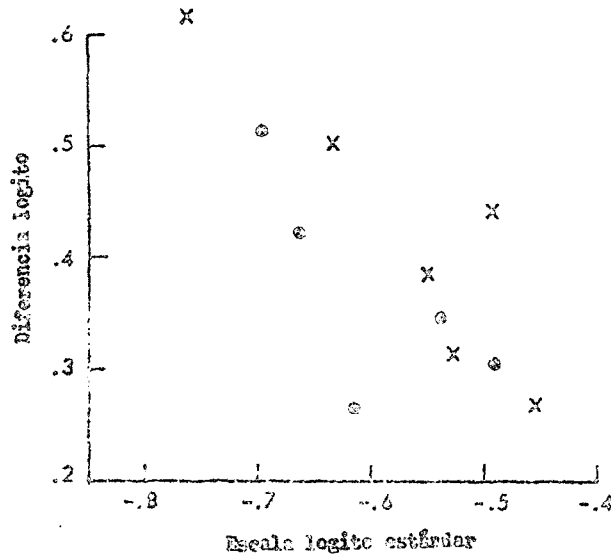
Gráfico 7.

DIFERENCIAS LOGITO DE SOBREVIVENCIA INFANTIL

Islas Salomón Británicas: 1967

X : estimado de intervalos a nacimientos anteriores

o : estimado de proporciones de niños muertos por edad de la madre



1
2
3
4
5

6
7
8

REFERENCIAS

- Brass, W., "The estimation of fertility rates from ratios of total to first births". *Population Studies*, 8(1), 74-87 (1954).
- Brass, W., "On the scale of mortality, in Symposium on biological aspects of demography", *Society for the Study of Human Biology*, (1967).
- Brass, W. et al "The demography of tropical Africa". Princeton University Press, Princeton (1968).
- Carrier, N.H. and Farrag, A.M., "The reduction of errors in census population for statistically under-developed countries". *Population Studies*, 12(3), 240-285 (1959).
- Coale, A.J. and Demeny, P., "Methods of estimating basic demographic measures from incomplete data". Publication ST/SOA/SER.A/42. United Nations, New York (1967).
- Demeny, P. and Shorter, F.C., "Estimating Turkish mortality, fertility and age structure". Istanbul University, Istanbul (1968).
- El-Badry, M.A., "Some aspects of differential fertility in Bombay as assessed from registration data". Contributed papers to the SYDNEY Conference of the International Union for the Scientific Study of Population, Session V(a), 309-318 (1967).
- Healy, M.J.R., "The disciplining of medical data". *British Medical Bulletin*, 24(3), 210-214 (1968).

1
2
3

4
5
6

7
8
9

10
11
12

