Sub-Programa de Sistemas Estadísticos
y de Evaluación

CEPAL-CELADE/R.1

UN METODO GENERAL PARA DESCOMPONER UNA DIFERENCIA ENTRE DOS TASAS EN VARIOS COMPONENTES

Prithwis Das Gupta

(Traducción del artículo publicado en
Demography, 15, N°1, febrero de 1978)

CENTRO LATINOAMERICANO DE DEMOGRAFIA
Santiago, diciembre de 1979

UN METODO GENERAL PARA DESCOMPONER UNA DIFERENCIA ENTRE DOS TASAS EN VARIOS COMPONENTES

Prithwis Das Gupta
Unidad de Demografía, Instituto de Estadísticas de la India
Calcuta 700 035, India

RESUMEN

En 1955 Kitagawa tuvo éxito en el trabajo de investigación sobre los componentes posibles de la diferencia entre dos tasas, al lograr separar esa diferencia en la parte debido al efecto de las propias tasas y el efecto debido al factor considerado en la tabla de clasificación usada. La información usada era una tabla de doble entrada (dos factores), y su análisis implicaba el uso de un factor de interacción de difícil interpretación. En 1973 Retherford y Cho idearon un segundo método que no incluye ningún factor de interacción. Sin embargo este nuevo método tiene algunas limitaciones como el de la dependencia de los resultados del orden en que se consideran los factores analizados. En el presente trabajo se propone un método más general en que se pueda considerar cualquier número de factores, desarrollados según las líneas sugeridas por Kitagawa y Rehterford y Cho, pero sin las limitaciones de esos métodos.

Introducción

Los demografos y otros científicos sociales han sido siempre cautelosos al interpretar la diferencia entre dos tasas brutas, del mismo fenómeno, para dos poblaciones ya que es posible que las diferencias de estructura en las poblaciones puedan explicar total o parcialmente la diferencia. Por ejemplo, una comparación entre dos tasas brutas de natalidad pueden diferir si la comparación se apoya en las correspondientes tasas tipificadas ya sea por edad o por el estado conyugal. Al calcular la diferencia entre dos tasas, considerando varios factores simultáneamente se podría pretender estudiar la contribución de cada uno de esos factores a la posible disparidad observada entre las tasas brutas y la de las tasas tipificadas.

Uno de los primeros trabajos que trata sobre este tipo de problema fué realizado por Goldfield (1948), brevemente descrito y utilizado en el estudio de John Durand de la fuerza laboral de U.S.A., cuyo método de tipificación múltiple asignaba interacciones iguales a todos los factores involucrados en cada interacción. Este tipo de método también fué utilizado por Gibson (1957) en el estudio de los efectos de los cambios en el estado conyugal y fecundidad conyugal sobre el descenso de la fecundidad Americana entre 1961 y 1973. En 1955 Kitagawa fué la primera en presentar una formulación matemática del análisis de componentes que proporciona la relación entre las tasas brutas y tipificadas para dos grupos de poblaciones. Su método fué utilizado por Blake y Das Gupta en 1975 para estudiar los componentes motivacionales y tecnológicos durante el descenso de la fecundidad conyugal americana entre 1960 y 1970. Kitagawa trabajó fundamentalmente con información clasificada de doble entrada (dos factores I y J), sugiriendo para el caso de clasificaciones con más de 2 factores un modo de reducirlos a una tabla de doble entrada y poder utilizar aproximación de dos factores. Que en el caso de dos factores su método implica además de los efectos de cada uno de los factores I y J un efecto de interacción entre ellos que no resultan de interpretación fácil.

En 1973 Retherford y Cho en un trabajo reciente de las tendencias de la fecundidad en los países del Este Asiatico utilizaron una técnica de descomposición de dos factores, I y J, que no involucran términos de interacción. Esto es, sin duda, una ventaja sobre el método de Kitagawa, pero este método posee otras restricciones. Primero, las magnitudes de los efectos de I y de J dependen del orden en que se deducen. Segundo, el método no utiliza información sobre el factor J en la computación del efecto I, si I fuese el factor que se deduce primero. Finalmente, a pesar de que el efecto (residual) de la tasa puede ser interpretado como la diferencia entre dos tasas convencionales tipificadas (como en el caso del efecto de la tasa de Kitagawa), ésta tiene dos expresiones diferentes (como se señala más adelante en la ecuación 9) dependiendo de cuál de los factores, I o J, se considere primero. Por otra parte, el efecto de la tasa de Kitagawa es único, posee una expresión muy simple (señalada más abajo en la ecuación 3, segundo término) y no plantea ningún problema particular para reemplazarla por cualquier otra expresión.

En este trabajo se propone un método general para descomponer una diferencia entre dos tasas siguiendo las líneas sugeridas por Kitagawa y Retherford y Cho, pero eliminando las limitaciones de esos métodos. Más específicamente, el método actual (a) puede aplicarse a la información clasificada por cualquier número de factores, (b) proporciona un efecto en la tasa que es idéntico al que se deduce del método de Kitagawa, (c) no involucra ningún término de interacción, (d) proporciona resultados que son independientes del orden en que consideran los factores, y (e) necesita de toda la información clasificada para el cálculo de todos y cada uno de los factores.

En la última sección, se aplica el método general para estudiar el cambio en la tasa de participación de la fuerza trabajadora de U.S.A. entre 1940 y 1970. Aplicaciones anteriores del método actual pueden también encontrarse en los estudios de Blake y Das Gupta (1976) y Hernández (1976), sobre los componentes del reciente descenso de la fecundidad en varios países.

Información clasificada por una variable I.

Supongamos que existen dos poblaciones llamadas población 1 y población 2. Estas dos poblaciones pueden ser, por ejemplo, la población femenina en edad fértil de los Estados Unidos en 1960 y 1970, o la población de 14 años y más en los estados de California y Alabama (digamos en 1970), o la población total de los Estados Unidos e India (digamos en 1975). Para la población 1 utilizamos las siguientes anotaciones:

N_i = Número de personas en la iesima categoría de I.

E_i = Número de acontecimientos (tales como nacimientos o defunciones) en la iesima categoría de I.

T_i = Tasa correspondiente a personas en la iesima categoría de I $(=E_i/N_i)$.

$N.$ = Número total de personas.

$(= \sum_i N_i)$,

$E.$ = Número total de acontecimientos.

$(= \sum_i E_i)$,

y

$T_.$ = Tasa bruta ($=E_./N_.$)

Para la población 2 utilizamos símbolos análogos a N,E, y T que llamaremos n,e y t respectivamente. Si $N_i = 0$ (tal que $E_i = 0$) podemos establecer que $T_i = t_i$. Igualmente si $n_i = 0$ podemos establecer que $t_i = T_i$.

La diferencia entre las tasas brutas de las poblaciones 1 y 2 pueden expresarse como sigue:

$$t_. - T_. = \sum_i T_i\left(\frac{n_i}{n_.} - \frac{N_i}{N_.}\right) + \sum_i \frac{N_i}{N_.}\left(t_i - T_i\right) + \sum_i \left(t_i - T_i\right)\left(\frac{n_i}{n_.} - \frac{N_i}{N_.}\right) \qquad (1)$$

El primer término del segundo miembro de la ecuación (1) mide el efecto de los cambios en la estructura I, considerando las tasas de la población 1. El segundo término es la diferencia entre las dos tasas tipificadas considerando la población 1 como modelo. Desgraciadamente, estos dos términos no totalizan ($t_. - T_.$), teniéndose un tercer término que considera la interacción entre las tasas y la estructura.

Con el objeto de evitar el término de interacción, Kitagawa sugiere la expresión alternativa

$$t_. - T_. = \sum_i \frac{t_i + T_i}{2}\left(\frac{n_i}{n_.} - \frac{N_i}{N_.}\right) + \sum_i \frac{\frac{n_i}{n_.} + \frac{N_i}{N_.}}{2}\left(t_i - T_i\right), \qquad (2)$$

donde el primer término del segundo miembro de la ecuación (2) mide el efecto de los cambios de estructura de I considerando las tasas promedio de las poblaciones 1 y 2, y el segundo término mide el efecto de los cambios en las tasas, considerando la estructura promedio de las poblaciones 1 y 2.

Por tanto, vemos que en lo que concierne al caso de una variable, Kitagawa encontró una expresión simple, ecuación (2), que descompone la diferencia entre las dos tasas brutas en efecto I y en efecto de tasa. Hacemos notar aquí, que si se distribuye el tercer término del segundo miembro de la ecuación (1) entre el primero y el segundo término se obtienen dos términos en el segundo miembro de la ecuación (2). Por lo tanto, la solución de Kitagawa en el caso de una variable es esencialmente la misma que Goldfield sugirió para la asignación de las interacciones. Kitagawa, sin embargo, plantea el problema de forma diferente. Ella encuentra el efecto del cambio de la estructura I, adoptando las tasas promedios como estándard, así mismo encuentra el efecto del cambio en las tasas adoptando la estructura I como estándard.Y sucede que los dos efectos totalizan la diferenci a entre las dos tasas brutas. Por lo tanto,al método de Kitagawa no se le puede criticar por haber ignorado en la ecuación (1) el término de interacción ni por haber distribuído en partes iguales (y arbitra~~ria~~mente) entre los dos efectos.

<u>Información clasificada por dos variables I y J.</u>

Para la población 1 se usaron las siguientes notaciones:

N_{ij} = Número de personas en la categoría (i, j),

E_{ij} = Número de acontecimientos en la categoría (i, j).

T_{ij} = Tasa correspondiente a personas en la categoría (i,j) $(= E_{ij}/N_{ij}$

$N_{i.}$ = Número de personas en la iesima categoría de I.

$$\left(= \sum_{j} N_{ij}\right),$$

$N_{.j}$ = Número de personas en la jotaesima categoría de J.

$$\left(= \sum_{i} N_{ij}\right),$$

$N_{..}$ = Número total de personas.

$$\left(= \sum_{i} \sum_{j} N_{ij}\right),$$

$E_{i.}$ = Número de acontecimientos en la iesima categoría de I.

$$\left(= \sum_{j} E_{ij}\right),$$

$E_{.j}$ = Número de acontecimientos en la jotaesima categoría de J.

$$\left(= \sum_{i} E_{ij}\right),$$

$E_{..}$ = Número total de acontecimientos.

$$\left(= \sum_{i} \sum_{j} E_{ij}\right),$$

$T_{i.}$ = Tasa correspondiente a personas en la iesima categoría de I ($= E_{i.}/N_{i.}$)

$T_{.j}$ = Tasa correspondiente a personas en la jotaesima categoría de J ($=E_{.j}/N_{.j}$)

y

$T_{..}$ = Tasa bruta ($E_{..}/N_{..}$)

Se utilizarán símbolos análogos para la población 2 sustituyendo a N,E, y T por n,e, y t respectivamente. Igual como en el caso de una variable, si $N_{ij} = 0$ se tiene que $T_{ij} = t_{ij}$; si $n_{ij} = 0$ se tiene que $t_{ij} = T_{ij}$.

El Método de Kitagawa.

Kitagawa encuentra primero el efecto combinado de IJ y el efecto (residual) de la tasa mediante la ecuación (2) considerando las categorías (i,j) como aquellas de una sola variable, de la manera siguiente:

$$t_{.} - T = \sum_i \sum_j \frac{t_{ij} + T_{ij}}{2} \left(\frac{n_{ij}}{n} - \frac{N_{ij}}{N_{.}}\right) + \sum_i \sum_j \frac{\frac{n_{ij}}{n} + \frac{N_{ij}}{N_{..}}}{2} \cdot \left(t_{ij} - T_{ij}\right). \qquad (3)$$

El efecto combinado de IJ, en el primer término del segundo miembro de la ecuación (3), se separa nuevamente en efecto I (independiente de J), efecto J (independiente de I) e interacción (efecto conjunto de I y J):

$$\text{efecto I} = \sum_i \sum_j \frac{t_{ij} + T_{ij}}{2} \; \frac{\frac{n_{.j}}{n} + \frac{N_{.j}}{N_{..}}}{2} \cdot \left(\frac{n_{ij}}{n_{.j}} - \frac{N_{ij}}{N_{.j}}\right), \qquad (4)$$

$$\text{efecto J} = \sum_i \sum_j \frac{t_{ij}+T_{ij}}{2} \; \frac{\frac{x_{i.}}{x} + \frac{N_{i.}}{N}}{2} \cdot \left(\frac{x_{ij}}{x_{i.}} - \frac{N_{ij}}{N_{i.}}\right) \qquad (5)$$

$$\text{efecto de interacción IJ} = \sum_i \sum_j \frac{t_{ij}+T_{ij}}{2} \cdot$$

$$\frac{\frac{N_{ij}}{N_{i.}}\frac{x_{i.}}{x} - \frac{x_{ij}}{x_{i.}}\frac{N_{i.}}{N} + \frac{N_{ij}}{N_{.j}}\frac{x_{.j}}{x_{..}} - \frac{x_{ij}}{x_{.j}}\frac{N_{.j}}{N}}{2} \qquad (6)$$

Se ve que a diferencia del caso de una variable (ecuación (2)), el planteamiento de Kitagawa, en el caso de dos variables, descompone la diferencia entre dos tasas brutas en componentes con términos de interacción.

El método de Retherford y Cho.

Con el objeto de suprimir el término de interacción, Retherford y Cho obtienen primero el efecto I y el efecto de la tasa descartando el factor J, aplicando el planteamiento de Kitagawa para el caso de un factor (ecuación (2)).

$$t_{..} - T_{..} = \sum_i \frac{t_{i.}+T_{i.}}{2}\left(\frac{x_{i.}}{x} - \frac{N_{i.}}{N}\right) + \sum_i \frac{\frac{x_{i.}}{x_{..}} + \frac{N_{i.}}{N_{..}}}{2}\left(t_{i.} - T_{i.}\right) \qquad (7)$$

El efecto de la tasa provisoria , segundo término del segundo miembro de la ecuación (7), nuevamente se separa en efecto J y efecto de la tasa (final) aplicando de nuevo a las diferencias ($t_{i.} - T_{i.}$) el método de Kitagawa para una variable (ecuación (2)).

$$\text{Efecto J} \quad \sum_i \sum_j \frac{t_{ij}+T_{ij}}{2} \cdot \frac{\frac{x_{i.}}{x_{..}} + \frac{N_{i.}}{N_{..}}}{2} \left(\frac{x_{ij}}{x_{i.}} - \frac{N_{ij}}{N_{i.}} \right) \qquad (8)$$

y

$$\text{Efecto de Tasa} = \sum_i \sum_j \frac{\frac{x_{i.}}{x_{..}} + \frac{N_{i.}}{N_{..}}}{2} \cdot \frac{\frac{x_{ij}}{x_{i.}} + \frac{N_{ij}}{N_{i.}}}{2} \left(t_{ij} - T_{ij} \right) \qquad (9)$$

Advertimos que los efectos J en el método Kitagawa y en el método de Retherford y Cho son idénticos al comparar las ecuaciones (5) y (8). A pesar de que las descomposiciones en las ecuaciones (7) y (9) no incluyan ningún término de interacción, es evidente que: (a) el efecto de tasa en la ecuación(9) del método de Retherford y Cho no es idéntico al efecto de tasa en la ecuación (3) obtenido por Kitagawa, y que debe aceptarse el último por razones ya discutidas anteriormente; (b) el efecto I en la ecuación (7) es independiente de la forma en que el factor J se distribuye dentro de las categorías del factor I; y (c) los efectos I y J no serían los mismos en las ecuaciones (7) y (8) si el efecto J se calculara primero.

Una aproximación alternativa.

Nuestro punto de partida será la adecuada descomposición de la diferencia entre dos tasas brutas en un efecto conjunto IJ y un efecto de tasa en la forma dada por Kitagawa en la ecuación (3). Para descomponer posteriormente el efecto conjunto IJ de la ecuación (3) en efecto I y en efecto J, se presentan dos poblaciones, 1' y 2', en las que el tamaño de la población en las categorías (i, j) son los mismos que en las poblaciones 1 y 2, respectivamente, pero en ambas la tasa en la categoría (i, j) es $(t_{ij} + T_{ij})/2$. Denotando con R y r las tasas de las poblaciones 1' y 2' se tendrá:

$$R_{ij} = r_{ij} = \frac{r_{ij} + R_{ij}}{2} = \frac{t_{ij} + T_{ij}}{2}, \quad (10)$$

$$\frac{r_{i.} + R_{i.}}{2} = \frac{1}{2}\left(\sum_j \frac{r_{ij} x_{ij}}{x_{i.}} + \sum_j \frac{R_{ij} N_{ij}}{N_{i.}}\right.$$

$$= \sum_j \frac{t_{ij} + T_{ij}}{2} \; \frac{\frac{x_{ij}}{x_{i.}} + \frac{N_{ij}}{N_{i.}}}{2}, \quad (11)$$

$$\frac{r_{.j} + R_{.j}}{2} = \sum_i \frac{t_{ij} + T_{ij}}{2} \; \frac{\frac{x_{ij}}{x_{.j}} + \frac{N_{ij}}{N_{.j}}}{2}, \quad (12)$$

y

$r_{..} - R_{..}$ = es lo mismo que el efecto conjunto de IJ en la ecuación (3), (13)

Aplicando las ecuaciones (7) y (9), del método de Retherford y Cho, para descomponer $(r_{..} - R_{..})$, para las poblaciones 1' y 2', en efectos I, J y efecto de tasa se ve que el efecto de tasa desaparece al aplicar la ecuación (10). Esta descomposición debido a la ecuación (13) será también una descomposición del efecto conjunto IJ en las poblaciones 1 y 2 en efecto I y en efecto J. Además, ya que las ecuaciones (7) y (9) pueden aplicarse tanto en el orden (I,J) o en el orden (J,I), denotando por I (I,J) y por J(I,J) los efectos I y J cuando se calcula en el orden I, J, y por I(J,I) y J(J,I) cuando el efecto J se obtiene primero, al aplicar las ecuaciones (7) y (11) se obtiene lo siguiente:

$$I(I,J) = \sum_i \sum_j \frac{t_{ij}+T_{ij}}{2} \; \frac{\frac{x_{ij}}{x_{i.}} + \frac{N_{ij}}{N_{i.}}}{2} \cdot \left(\frac{x_{i.}}{x} - \frac{N_{i.}}{N_{..}}\right) \qquad (14)$$

y al aplicar las ecuaciones (8) y (10), se obtiene:

$$J(I,J) = \sum_i \sum_j \frac{t_{ij}+T_{ij}}{2} \cdot \frac{x_{i.}}{x_{..}} + \frac{N_{i.}}{N_{..}} \left(\frac{x_{ij}}{x_{i.}} - \frac{N_{ij}}{N_{i.}}\right) \qquad (15)$$

El efecto J en la ecuación (8) y el efecto J (I,J) en la ecuación (15) son obviamente idénticos. Por simetría J(J,I) e I(J,I) tienen expresiones análogas para I(I,J) y J(I,J), respectivamente. Los únicos cambios necesarios son los reemplazos de $n_{i.}$ y $N_{i.}$ por $n_{.j}$ y $N_{.j}$, respectivamente.

Como es lógico tanto calcular primero el efecto J como el efecto I, se definen finalmente los efectos I y J como la media aritmética de los valores correspondientes obtenidos en los dos posibles órdenes. En otras palabras,

$$\begin{aligned} \text{Efecto I} &= 1/2 \; I(I,J) + I(J,I) \\ \text{Efecto J} &= 1/2 \; J(I,J) + J(J,I) \end{aligned} \qquad (16)$$

Los efectos I y J en la ecuación (16) conjuntamente con el efecto de tasa, segundo término del segundo miembro de la ecuación (3), constituyen el método propuesto de descomposición de la diferencia entre dos tasas brutas para información clasificada por dos factores.

Relación entre el método de Kitagawa y el método de Das Gupta.

Se ve que el efecto J en la ecuación (5) y el efecto J(I,J) en la ecuación (15) son idénticos. Por simetría el efecto I en la ecuación (4) es también idéntico con el efecto I(J,I). Por lo tanto, la descomposición de Kitagawa del efecto conjunto IJ (digamos C) es:

$$C = I(J,I) + J(I,J) + IJ \text{ Interacción} \qquad (17)$$

También de las ecuaciones (14) y (15).

$$C = I(I,J) + J(I,J) = I(J,I) + J(J,I) \qquad (18)$$

Por lo tanto, de las ecuaciones (16) y (18) y por lo tanto de la ecuación (17) obtenemos:

$$\text{Efecto I} = 1/2 \left[\{C - J(I,J)\} + I(J,I)\right] = 1/2\left[\{I(J,I) + IJ \text{ int.}\} + I(J,I)\right]$$
$$= I(J,I) + \frac{IJ \text{ int.}}{2} \qquad (19)$$

Argumentos similares nos conducen a:

$$\text{Efecto J} = J(I,J) + \frac{IJ \text{ int.}}{2} \qquad (20)$$

Al comparar la ecuación (17) con las ecuaciones (19) y (20) es evidente que si el término de interacción IJ de Kitagawa en la ecuación (6) se asigna a los efectos I y J en las ecuaciones (4) y (5) se obtienen los efectos I y J de la ecuación (16) del método de Das Gupta. Esto también es consecuente con la aproximación de Goldfield y con el planteamiento de Kitagawa en la ecuación (2) para información clasificada por un factor. También se puede decir que la idea subyacente en el presente método (lo mismo que en el caso de un factor de Kitagawa) es la de encontrar el efecto del cambio en un factor manteniendo constantes los otros factores a un nivel promedio y que afortunadamente todos estos efectos totalizan la diferencia entre las dos tasas brutas. Por lo tanto, la pregunta de que por qué el término de interacción de Kitagawa en la ecuación (6) se distribuyó equitativamente entre los factores y no en otras proporciones, no parece ser válida.

Información clasificada por tres factores I, J y K.

Por la complejidad de las expresiones, Kitagawa sugiere que, cuando existen tres factores I, J y K, uno de ellos, digamos I, puede considerarse como un factor y la clasificación cruzada de J y K como un segundo factor. Esto nos permitiría utilizar las fórmulas de dos factores de Kitagawa en una situación de tres factores. El método alternativo sugerido en la sección anterior, para dos factores, puede, sin embargo, extenderse directamente a tres factores. Utilizando anotaciones análogas primero, se separa la diferencia entre las dos tasas brutas ($t_{...} - T_{...}$) en los siguientes componentes, como lo realizara Kitagawa:

$$\text{Efecto Combinado IJK} = \sum_i \sum_j \sum_k \frac{t_{ijk} + T_{ijk}}{2} \left(\frac{n_{ijk}}{n_{...}} - \frac{N_{ijk}}{N_{...}} \right) \qquad (21)$$

$$\text{Efecto de Tasa} = \sum_i \sum_j \sum_k \frac{\frac{n_{ijk}}{n_{...}} + \frac{N_{ijk}}{N_{...}}}{2} \left(t_{ijk} - T_{ijk} \right) \qquad (22)$$

De nuevo utilizando notaciones similares para los efectos en un orden particular de I, J y K, se encuentran expresiones análogas a las ecuaciones (14) y(15) como se indica a continuación:

$$I(I,J,K) = \sum_i \sum_j \sum_k \frac{t_{ijk} + T_{ijk}}{2} \cdot \frac{\frac{n_{ijk}}{n_{i..}} + \frac{N_{ijk}}{N_{i..}}}{2} \left(\frac{n_{i..}}{n_{...}} - \frac{N_{i..}}{N_{...}} \right) \qquad (23)$$

$$J(I,J,K) = \sum_i \sum_j \sum_k \frac{t_{ijk} + T_{ijk}}{2} \; \frac{\frac{n_{ijk}}{n_{ij.}} + \frac{N_{ijk}}{N_{ij.}}}{2} \cdot \frac{\frac{n_{i..}}{n_{...}} + \frac{N_{i..}}{N_{...}}}{2} \left(\frac{n_{ij.}}{n_{i..}} - \frac{N_{ij.}}{N_{i..}} \right) \qquad (24)$$

y

$$K(I,J,K) = \sum_i \sum_j \sum_k \frac{t_{ijk} + T_{ijk}}{2} \cdot \frac{\frac{n_{i..}}{n_{...}} + \frac{N_{i..}}{N_{...}}}{2} \; \frac{\frac{n_{ij.}}{n_{i..}} + \frac{N_{ij.}}{N_{i..}}}{2} \left(\frac{n_{ijk}}{n_{ij.}} - \frac{N_{ijk}}{N_{ij.}} \right) \qquad (25)$$

Las expresiones para los efectos I, J y K correspondientes a otros órdenes de cálculo resultan directamente de las ecuaciones (23) y (25) mediante una adecuada subatitución de los subíndices. El efecto K (J,K,I), por ejemplo, tiene la misma expresión que la correspondiente al efecto J(I,J,K) en la ecuación (24) excepto que $n_{ij.}$, $N_{ij.}$, $n_{i..}$ y $N_{i..}$ deben reemplazarse por $n_{.jk}$, $N_{.jk}$, $n_{.j.}$ y $N_{.j.}$ respectivamente. Se advierte que I(I,J,K) = I(I,K,J), J(J,I,K) = = J(J,K,I), y K(K,I,J) = K(K,J,I). Al igual que en la ecuación (16) se obtienen finalmente los efectos I,J y K :

$$\text{Efecto I} = 1/6\left[I(I,J,K) + I(I,K,J) + I(J,I,K) + I(J,K,I) + I(K,I,J) + I(K,J,I)\right] \tag{26}$$

$$\text{Efecto J} = 1/6\left[J(I,J,K) + J(I,K,J) + J(J,I,K) + J(J,K,I) + J(K,I,J) + J(K,J,I)\right] \tag{27}$$

y

$$\text{Efecto K} = 1/6\left[K(I,J,K) + K(I,K,J) + K(J,I,K) + K(J,K,I) + K(K,I,J) + K(K,J,I)\right] \tag{28}$$

<u>Información clasificada por cuatro factores I,J,K y L.</u>

Cuando hay cuatro factores involucrados, la diferencia ($t_{....} - T_{....}$) primero se separa en dos componentes, como en las ecuaciones (21) y (22) como sigue:

$$\text{Efecto combinado IJKL} = \sum_i \sum_j \sum_k \sum_l \frac{t_{ijkl} + T_{ijkl}}{2}\left(\frac{n_{ijkl}}{n_{....}} - \frac{N_{ijkl}}{N_{....}}\right) \tag{29}$$

$$\text{Efecto de Tasa} = \sum_i \sum_j \sum_k \sum_l \frac{\frac{n_{ijkl}}{n_{....}} + \frac{N_{ijkl}}{N_{....}}}{2}\left(t_{ijkl} - T_{ijkl}\right) \tag{30}$$

El efecto conjunto IJKL en la ecuación (29) se divide luego en los efectos de los cuatro factores individuales I, J, K y L:

$$\text{Efecto I} = 1/24\left[I(I,J,K,L) + I(I,J,L,K) + \dots + I(L,K,I,J) + I(L,K,J,I)\right] \tag{31}$$

$$\text{Efecto J} = 1/24\left[J(I,J,K,L) + J(I,J,L,K) + \dots + J(L,K,I,J) + J(L,K,J,I)\right] \tag{32}$$

$$\text{Efecto K} = 1/24\left[K(I,J,K,L) + K(I,J,L,K) + \dots + K(L,K,I,J) + K(L,K,J,I)\right] \tag{33}$$

$$\text{y Efecto L} = 1/24\left[L(I,J,K,L) + L(I,J,L,K) + \dots + L(L,K,I,J) + L(L,K,J,I)\right] \tag{34}$$

Cada uno de los efectos anteriores es el promedio de 24 efectos que corresponden a 24 (=4!) posibles órdenes en los que el efecto de estos cuatro factores pueden considerarse. Los efectos para cualquier orden particular pueden obtenerse de las siguientes expresiones estándard, similares a la ecuación (23) y (25).

$$I(I,J,K,L) = \sum_i \sum_j \sum_k \sum_l \frac{t_{ijkl}+T_{ijkl}}{2} \cdot \frac{\frac{x_{ijkl}}{x_{i...}}+\frac{N_{ijkl}}{N_{i...}}}{2} \left(\frac{x_{i...}}{x_{....}}-\frac{N_{i...}}{N_{....}}\right) \qquad (35)$$

$$J(I,J,K,L) = \sum_i \sum_j \sum_k \sum_l \frac{t_{ijkl}+T_{ijkl}}{2} \frac{\frac{x_{ijkl}}{x_{i...}}+\frac{N_{ijkl}}{N_{ij..}}}{2} \cdot \frac{\frac{x_{i...}}{x_{....}}+\frac{N_{i...}}{N_{....}}}{2} \left(\frac{x_{ij..}}{x_{i...}}-\frac{N_{ij..}}{N_{i...}}\right) \qquad (36)$$

$$K(I,J,K,L) = \sum_i \sum_j \sum_k \sum_l \frac{t_{ijkl}+T_{ijkl}}{2} \frac{\frac{x_{ijkl}}{x_{ijk.}}+\frac{N_{ijkl}}{N_{ijk.}}}{2} \cdot \frac{\frac{x_{i...}}{x_{....}}+\frac{N_{i...}}{N_{....}}}{2} \frac{\frac{x_{ij..}}{x_{i...}}+\frac{N_{ij..}}{N_{i...}}}{2} \left(\frac{x_{ijk.}}{x_{ij..}}-\frac{N_{ijk.}}{N_{ij..}}\right) \qquad (37)$$

$$L(I,J,K,L) = \sum_i \sum_j \sum_k \sum_l \frac{t_{ijkl}+T_{ijkl}}{2} \cdot \frac{\frac{x_{i...}}{x_{....}}+\frac{N_{i...}}{N_{....}}}{2} \frac{\frac{x_{ij..}}{x_{i...}}+\frac{N_{ij..}}{N_{i...}}}{2} \cdot \frac{\frac{x_{ijkl}}{x_{ij..}}+\frac{N_{ijkl}}{N_{ij..}}}{2} \left(\frac{x_{ijkl}}{x_{ijk.}}-\frac{N_{ijkl}}{N_{ijk.}}\right) \qquad (38)$$

Si el número de categorías de los factores I, J, K y L son respectivamente m_i, m_j, m_k y m_l y si cualquiera de estos números, digamos m_l, es igual a uno, entonces la información involucrada en estos cuatro factores I, J, K y L cae en un caso de tres factores I, J y K. En tal situación, los símbolos tales como N_{ijk}, N_{ijkl} o N_{ijk} se refieren al mismo número. Si se tiene en cuenta lo anterior al sustituir $m_l = 1$ en las fórmulas de cuatro factores (29) y (38) obtenemos las fórmulas (21) y (28) que corresponden al caso de tres factores. De la misma forma, en las fórmulas de cuatro factores (29) y (38) si sustituimos $m_k = 1$ y $m_k = m_l = m_j = 1$ resultan las fórmulas de dos factores (14) y (16) respectivamente y la fórmula (2) de un factor de Kitagawa. Esta es otra razón, además de las mencionadas anteriormente (p. 104) de que por qué el presente método es consistente con, y puede considerarse como una generalización del método de Kitagawa para información clasificada por un factor. También el hecho que una simple sustitución de uno o más valores de m por uno, en las fórmulas, lleve a las fórmulas equivalentes a aquellas correspondientes a información clasificada de menor orden - hace posible elaborar un programa general de computación para el análisis de componentes, de digamos, información de cuatro factores y utilizarlo para cualquier grupo de información involucrando uno, dos, tres o cuatro factores (ver la sección siguiente).

En la mayoría de los casos, ya sea por que no se dispone de información clasificada por más de cuatro factores o por el número insuficiente de casos en cada celda, si esta información estuviera disponible, creemos que la consideración de fórmulas apropiadas para cinco o más factores permanecerá siendo una materia de amplio interés academico. Sin embargo, si la situación lo requiere, es fácil, debido a la simetría de la fórmula de cuatro factores (29) y (38), como expresa el efecto de la tasa y el efecto de los factores cuando se trata de cualquier número de factores. El número de cálculos comienza a ser mayor mientras más factores se introducen. Sin lugar a dudas que fue una tarea extraordinaria considerar hasta 3 factores en el artículo de Kitagawa publicado en 1955. Debido al aparecimiento de las facilidades de la computación electrónica, el cálculo para un número razonable de factores dejó de ser un problema y se pudo desarrollar un método general dentro de un programa compacto.

Una ilustración numérica.

Se ilustrará ahora, la aplicación del presente método con información de los Estados Unidos para 1940 y 1970. El porcentaje de población (14 años de edad y más) en la fuerza trabajadora era 52,22 en 1940 (población 1) y 55,49 en 1970 (población 2) de modo que la diferencia entre las dos tasas brutas de participación de la fuerza laboral es de 3.27. Para ambos años se dispone de la información cruzada para la población total de 14 años de edad y más (personas) y para la población en la fuerza de trabajo (acontecimientos), por edad, sexo, estado civil y región (Apéndice, Cuadros 1, 2, 3, y 4). Las categorías consideradas para estos factores son las siguientes:

I (Edad) : 14 a 24, 25 a 34, 35 a 44, y 45 y más

L (Sexo) : 1 = masculino 2 = femenino

K (Estado Civil): 1 = soltero 2 = casado 3 = otros

J (Región) : 1 = urbano 2 = rural -no agricola 3 = rural agrícola

Es posible realizar, con la información anterior disponible,y sus totales marginales, cuatro análisis de componentes de un factor, seis análisis de dos factores, cuatro análisis de tres factores y un análisis de cuatro factores de la diferencia de 3.27 entre las dos tasas brutas. Se ha editado un programa de computación para el análisis de cuatro factores pero si las tarjetas de instrucciones se confecciocan de acuerdo a las indicaciones simples dadas en el programa, este puede también aplicarse adecuadamente a los casos de uno, dos o tres factores (el programa se puede obtener del autor, si se solicita). En realidad el programa se ha utilizado una sola vez con los quince posibles conjuntos de información agrupados, cuyo resumen se presenta en el Cuadro 1. Todos estos resultados se señalan solamente a modo de ilustración. Sin embargo, solo estaremos satisfechos con aquellos resultados que comprendan la utilización máxima de la información (en el caso presente los resultados de cuatro factores del Cuadro 1). En este caso las tasas estandarizadas por edad-sexo-estado conyugal-región son de 51,39 y de 55,81 para 1940 y 1970, respectivamente y su diferencia (4.42) se debe al efecto de la tasa. Vale la pena mencionar que en la computadora 7600 CDC se realizaron en menos de 1.5 segundos todos los cálculos para los quince conjuntos de datos.

Cuadro 1 - Análisis de Componentes de la Diferencia entre Tasas Brutas de Participación en la Fuerza Laboral (por ciento) para Personas de 14 años y más, U.S.A. 1940 y 1970.

Número de Factores	Efectos de los factores: Edad	Sexo	Estado Civil	Región	Efecto de Tasa	Efecto Total
1	-1.28				4.55	3.27
		-0.94			4.21	3.27
			0.06		3.21	3.27
				0.85	2.42	3.27
2	-1.33	-0.98			5.58	3.27
	-1.48		-0.16		4.91	3.27
	-1.34			0.86	3.79	3.27
		-0.61	0.38		3.50	3.27
		-0.78		1.01	3.04	3.27
			0.12	0.94	2.21	3.27
3	-1.51	-0.63	0.19		5.22	3.27
	-1.41	-0.84		0.96	4,56	3.27
	-1.56		-0.11	0.82	4.12	3.27
		-0.54	0.37	1.01	2.43	3.27
4	-1.62	-0.58	0.16	0.89	4.42	3.27

Fuentes: Oficina de Censos de los Estados Unidos, Censos de los Estados Unidos 1940, Población Vol. 3 La Fuerza Laboral, Parte I, Resumen Estados Unidos (Washington D.C., Oficina de Prensa del Gobierno de los Estados Unidos, 1943), Cuadro 6; Oficina de Censos de los Estados Unidos, Censos de los Estados Unidos: 1940, Población Vol. 4, Características por Edad, Parte I, Resumen Estados Unidos (Washington D.C.; Oficina de Prensa del Gobierno de los Estados Unidos, 1943), Cuadros 6,8 y 9; Oficina de Censos de los Estados Unidos, Censos de Población: 1970, Características Detalladas, Informe Final PC (1)-D1, Resumen Estados Unidos (Washington D.C.: Oficina de Prensa del Gobierno de los Estados Unidos, 1973), Cuadros 203, 215 y 216.

En cuanto a los resultados del Cuadro 1, se ve que los efectos de los factores varían dentro de un pequeño margen dependiendo de cuantos y cuales factores se consideren. Por otra parte, el efecto de tasa muestra fluctuaciones considerables, teniendo un peso relativamente alto o bajodependiendo , respectivamen de si los factores con efectos positivos o negativos sean los que predominen.

En la Tabla 1 (donde por tasa se entiende tasa especifica IJKL) se detallan las conclusiones más importantes encontradas en el análisis de cuatro factores presentado en el Cuadro 1.

Tabla 1

Y si las siguientes componentes fueran las mismas (e iguales al promedio de 1940 y 1970) en 1940 y 1970	Y solamente las siguientes componentes hubieran cambiado como lo hicieron desde 1940 a 1970	El cambio porcentual de la población en la fuerza de trabajo desde 1940 a 1970 sería de
Sexo, estado civil, región, tasa	Edad	un descenso de 1.62
Edad, estado civil	Sexo	un descenso de 0.58
Edad, sexo Religión,tasa	Estado civil	un aumento de 0.16
Edad, sexo Estado civil Tasa	Región	un aumento de 0.89
Edad, sexo Estado civil Región	Tasa	un aumento de 4.42
	Edad, sexo, estado civil, región, tasa	un aumento de 3.27

Kitagawa hizo dos observaciones generales sobre el análisis de componentes que vale la pena repetir. Primero, no necesariamente los efectos de los factores implican una relación causal. Simplemente indican la naturaleza de la asociación de los factores con el fenómeno que se está midiendo. Pueden existir algunas fuerzas ocultas detrás de los factores que sean las verdaderas responsables de los valores que les atribuímos a los factores como efectos, pero al identificar estas fuerzas está más allá de las posibilidades del análisis de componentes. Segundo, no necesariamente se puede explicar mejor el efecto total aumentando el número de factores. En otras palabras, no se espera que con un aumento en el número de factores descienda gradualmente el efecto de tasa (ver Cuadro 1). La diferencia entre dos tasas estandarizadas no es necesariamente menor que la diferencia entre sus correspondientes tasas brutas, mientras que la variancia residual en un problema de regresión es necesariamente menor que la variancia total. Por lo tanto, el análisis de componentes no es como el análisis de regresión donde la suma de cada variable independiente a la ecuación explica en forma creciente la variación en la variable dependiente. Otra gran diferencia entre el análisis de componente y el análisis de regresión es que, dada la información, el criterio que separa las dos poblaciones en un análisis de componentes puede considerarse como una variable independiente en un análisis de regresión. Por ejemplo, si se dispone de la información respecto del CI de niños y también de la de su raza (blanca/negra), los ingresos de la familia y la educación de la madre, se puede usar el análisis de componentes para tratar de encontrar en que medida el promedio de 15 puntos de diferencia entre el CI de los niños blancos y negros pueden explicarse en términos de ingreso familiar y educación. Por otra parte, en un análisis de regresión cabe preguntarse cuanto de la variación en el CI de los niños puede explicarse por raza, entrada familiar y educación.

REFERENCIAS

Blake, Judith, and P. Das Gupta, 1975. Reproductive Motivation Versus Contraceptive Technology: Is Recent American Experience an Exception? Population and Development Review 1:229 - 249.

__________, and P. Das Gupta. 1976. Components of the Decline in American Marital Fertility Between 1960 and 1970. Manuscript. Bekerley: University of California.

Gibson, Campbell. 1975. Changes in Marital Status and Marital Fertility and Their Contribution to the Decline in Period Fertility in the United States: 1961 - 1973. Paper presented at the annual meeting of the Population Association of America. April 1975, in Seattle, Washington.

Goldfield, E.D. 1948. Appendix B: Methods of Analyzing Factors of Labor Force Change. Pp. 219 - 236 in John D. Durand, The Labor Force in the United States: 1890 - 1960. New York: Social Science Research Council.

Hernandez, Donald J. 1976. Policy Vs. Other Factors in Fertility Decline After 1950. Unpublished PhD dissertation. Bekerley: University of California.

Kitagawa, E. M. 1955. Components of a Difference Between Two Rates. Journal of the American Statistical Association. 50: 1168 - 1194.

Retherford, R.D. and L.J. Cho. 1973. Comparative Analysis of Recent Fertility Trend in East Asia. Pp. 163 - 181 in International Union for the Scientific Study of Population (ed.) Proceedings of the 17th General Conference of the IUSSP, August 1973. Vol. 2. Liege, Belgium. International Union for the Scientific Study of Population.

Apéndice Cuadro 1 - Población de 14 años de edad y más por Edad (I), Región (J), Estado Conyugal (K) y Sexo (L): Estados Unidos, 1940.

J	K	L	I 14 - 24	I - 25 - 34	I - 35 - 44	I - 45 +
1 [a]	1	1	6,077,395	1,864,473	804,881	1,077,309
2	1	1	2,231,505	524,485	226,033	399,335
3	1	1	3,090,353	625,359	253,080	419,171
1	2	1	767,715	4,048,445	4,328,747	7,373,783
2	2	1	370,162	1,597,307	1,472,711	2,455,969
3	2	1	396,285	1,351,121	1,387,888	3,108,490
1	3	1	85,707	324,938	460,291	1,705,112
2	3	1	37,167	113,235	143,841	645,567
3	3	1	34,372	71,611	87,322	592,583
1	1	2	5,538,511	1,499,021	706,200	1,117,055
2	1	2	1,731,657	297,398	134,495	247,419
3	1	2	2,118,950	253,371	110,181	181,608
1	2	2	1,785,075	4,676,236	4,250,469	5,731,361
2	2	2	860,972	1,753,808	1,367,892	1,895,980
3	2	2	860,715	1,503,101	1,441,182	2,390,921
1	3	2	207,030	599,014	846,383	3,815,524
2	3	2	70,527	146,017	198,655	1,133,611
3	3	2	62,990	90,086	112,969	832,792
Total			26,327,088	21,339,026	18,333,220	35,103,590
Gran Total				101,102,924		

a Para explicación de los números de estas categorías, ver la última sección del texto.

Apéndice Cuadro 2 - Personas de la Fuerza Laboral por Edad (I), Región (J), Estado Conyugal (K) y Sexo (L); Estados Unidos, 1940.

J	K	L	I = 14 - 24	I = 25 - 34	I = 35 - 44	I = 45 +
1	1	1	3,077,342	1,707,329	706,852	718,406
2	1	1	1,088,644	438,263	164,911	204,606
3	1	1	1,778,602	572,784	226,616	312,910
1	2	1	744,377	3,973,494	4,215,065	6,138,377
2	2	1	358,946	1,563,217	1,424,570	1,911,770
3	2	1	385,299	1,328,310	1,358,939	2,786,518
1	3	1	69,357	287,721	403,894	965,244
2	3	1	26,821	82,134	100,745	292,919
3	3	1	28,813	62,079	76,688	361,678
1	1	2	2,400,191	1,293,244	575,797	579,371
2	1	2	512,254	214,531	84,412	86,707
3	1	2	407,490	116,490	39,662	39,325
1	2	2	385,603	1,030,946	770,353	568,925
2	2	2	98,418	258,181	192,616	162,336
3	2	2	57,777	103,939	87,641	90,331
1	3	2	122,277	422,525	559,159	897,820
2	3	2	31,422	81,443	103,098	206,504
3	3	2	21,017	39,941	52,682	150,831
Total			11,594,650	13,576,571	11,143,700	16,474,578
Gran Total				52,789,499		

Apéndice Cuadro 3 - Población de 14 años de edad y más por Edad (I), Región (J), Estado Conyugal (K) y Sexo (L) : Estados Unidos, 1970.

J	K	L	I - 14 - 24	I - 25 - 34	I - 35 - 44	I - 45 +
1	1	1	11,708,759	1,509,681	679,413	1,371,986
2	1	1	3,315,162	321,785	165,719	422,472
3	1	1	712,049	73,149	42,940	127,322
1	2	1	2,479,694	6,846,531	6,860,497	15,988,504
2	2	1	775,569	2,275,328	2,223,362	5,107,775
3	2	1	57,846	242,136	388,784	1,347,368
1	3	1	441,681	702,882	714,339	3,026,023
2	3	1	110,210	178,194	173,014	894,550
3	3	1	11,164	18,117	20,630	147,269
1	1	2	10,218,862	1,052,330	545,479	1,859,618
2	1	2	2,628,069	178,925	103,411	352,084
3	1	2	557,317	30,188	17,916	60,862
1	2	2	4,024,955	7,211,102	6,937,038	13,929,463
2	2	2	1,305,839	2,449,820	2,203,546	4,364,818
3	2	2	103,614	306,312	444,886	1,164,627
1	3	2	817,755	1,201,362	1,328,989	9,269,213
2	3	2	205,512	228,904	260,583	2,220,179
3	3	2	24,773	19,201	22,275	266,458
Total			39,498,830	24,845,947	23,132,821	61,920,591
Gran Total				149,398,189		

Apéndice Cuadro 4 - Personas de la Fuerza Laboral por Edad (I), Región (J), Estado Conyugal (K) y Sexo (L): Estados Unidos, 1970.

J	K	L	I - 14 - 24	I - 25 - 34	I - 35 - 44	I = 45 +
1	1	1	5,371,325	1,260,249	548,076	737,910
2	1	1	1,268,759	233,006	108,207	172,830
3	1	1	287,642	61,972	35,240	88,432
1	2	1	2,285,504	6,653,752	6,705,567	12,183,629
2	2	1	729,432	2,223,153	2,153,680	3,499,502
3	2	1	52,902	235,590	376,438	1,051,983
1	3	1	357,493	606,757	610,218	1,386,585
2	3	1	77,165	132,788	132,049	329,303
3	3	1	8,530	16,396	18,193	76,406
1	1	2	3,785,363	852,652	426,049	946,915
2	1	2	674,622	117,398	60,606	123,865
3	1	2	128,977	19,198	9,863	19,750
1	2	2	1,866,498	2,763,511	3,187,296	5,135,841
2	2	2	510,330	930,666	1,023,095	1,417,442
3	2	2	32,543	92,148	161,369	299,215
1	3	2	454,669	766,538	920,999	3,043,713
2	3	2	100,727	134,203	163,309	559,686
3	3	2	11,694	10,951	12,429	56,640
Total			18,004,175	17,110,928	16,652,683	31,129,647
Gran Total				82,897,433		