

CELADE

CENTRO LATINOAMERICANO DE DEMOGRAFIA

Distribución interna

Roberto Bachi

5056

BIBLIOTECA "GIORGIO NORTARA"
CENTRO LATINOAMERICANO
DE DEMOGRAFIA

Serie D, n° 61.
Agosto, 1970.
300.

ANALISIS ESTADISTICO DE
SERIES GEOGRAFICAS
(Traducción del artículo aparecido
en Bulletin of the International
Statistical Institute, Proceedings of
the 30th. session, Estocolmo, 1957,
Tomo 36, 2ª edición, 1958.)

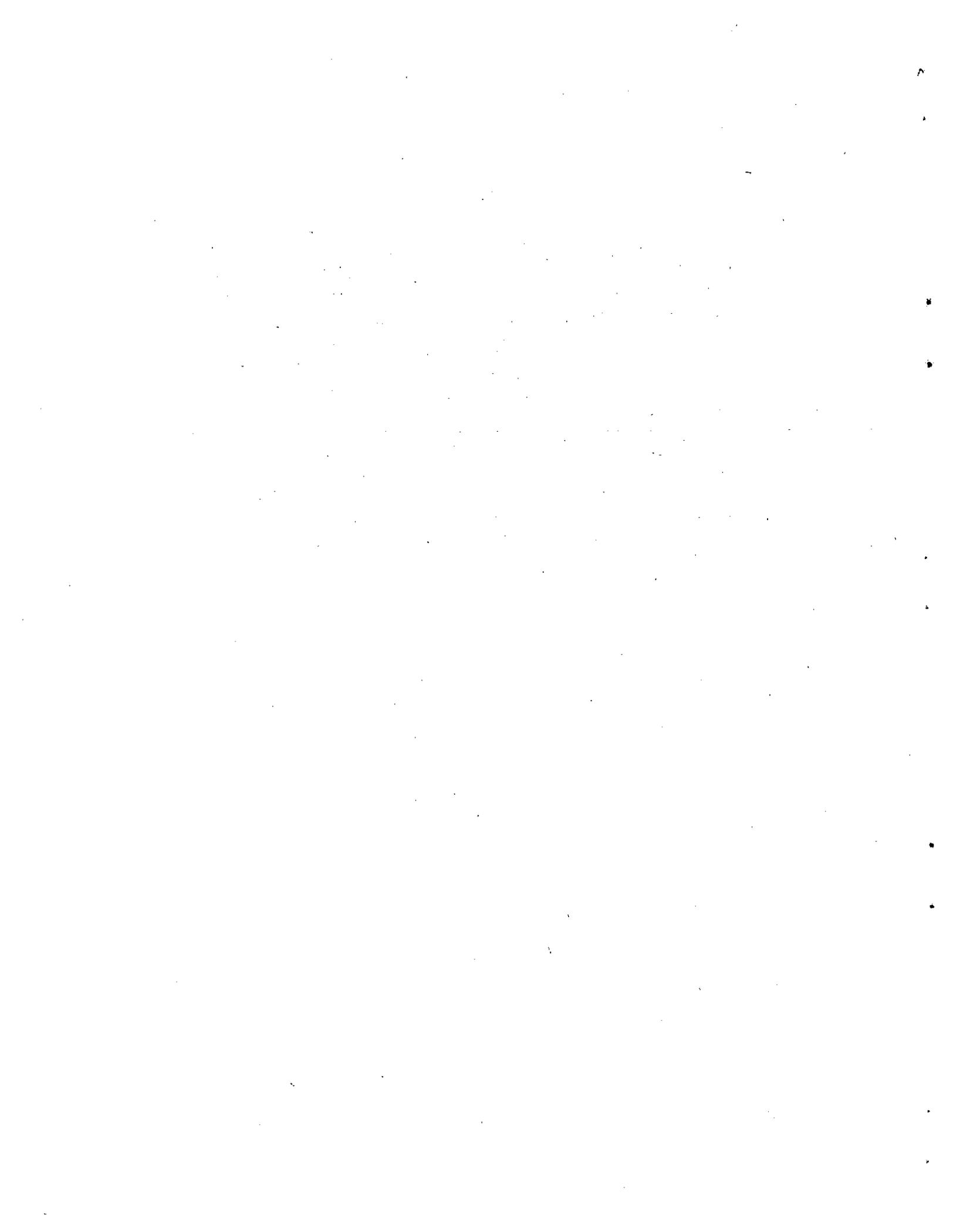
Las opiniones y datos que figuran en este trabajo son responsabilidad del autor, sin que el Centro Latinoamericano de Demografía (CELADE) sea necesariamente partícipe de ellos.

I N D I C E

	<u>Página</u>
1. Series geográficas	1
2. El tratamiento de las series que muestran la distribución de los casos de un fenómeno en el territorio	1
2.1 La distribución de las distancias	1
2.2 La distancia normal	2
2.3 Ejemplos de aplicaciones de la distancia normal	3
2.4 Aplicación al estudio de poblaciones heterogéneas y a la subdivisión de una población por zonas geográficas	4
3. Tratamiento de las series geográficas que muestra las distribuciones de frecuencias o promedios	5
4. El tratamiento de las series que muestran movimientos clasificados por lugares de salida y de destino	6
4.1 Índices de preferencia o de aversión para cada dirección de los movimientos ..	6
4.2 Distancias recorridas por los migrantes	7
4.3 Mediciones sintéticas de los efectos de las migraciones	9

Cuadros

1 Distancias por carretera (en km.) entre los habitantes de Roma y Zurich	2
2 Cuadrados de las distancias normales (standard distances) de dos grupos de población en Israel	4
3 Distancias medias reales y teóricas recorridas en la migración interna (metros) ...	9



1. Serios geográficas

Las series geográficas constituyen una parte considerable de las estadísticas oficiales, y se utilizan ampliamente en la investigación científica. No obstante, los métodos para tratar estas series están menos desarrollados que los de otros tipos de series estadísticas y aún no han sido incluidos en los textos de estadística corrientes. En este trabajo ^{1/} se ilustran algunos métodos elementales para el tratamiento de las series geográficas, aunque en absoluto se pretende que sean totalmente nuevos o que su lista sea completa.

Consideraremos en forma separada los métodos para el tratamiento de las siguientes series: a) series que muestran la distribución de los casos de un fenómeno en el territorio (véase sección 2); b) series geográficas de distribuciones de frecuencia de atributos o variables, y c) series que muestran la intensidad media de un fenómeno en cada sección del territorio (véase sección 3); d) series que muestran los movimientos en el territorio (véase sección 4).

2. El tratamiento de las series que muestran la distribución de los casos de un fenómeno en el territorio

2.1. La distribución de las distancias

Consideremos i) las series que muestran la ubicación exacta de cada caso individual de un fenómeno (v.g. la dirección de cada caso de sarampión que ocurre en el curso de un período dado en una ciudad); o ii) las series que muestran el número de casos de un fenómeno que se presenta en cada lugar o distrito (como por ejemplo la población de cada uno de los lugares habitados de un país).

Entre los métodos de tratamiento empleados para sintetizar estas series geográficas, se utiliza con frecuencia la media aritmética o la mediana de estas series ("centro de gravedad" o simplemente "centro" del fenómeno; "centro mediano"). Estos valores deben, no obstante, ser complementados por valores que miden la diseminación o dispersión geográfica del fenómeno. Veamos cómo se puede calcular y utilizar estos valores.

Que $1, \dots, i, \dots, l, \dots, n$ sean los casos del fenómeno estudiado, $x_1, \dots, x_i, \dots, x_l, \dots, x_n$ las longitudes de los lugares o puntos en que se presenta, $y_1, \dots, y_i, \dots, y_l, \dots, y_n$ sus latitudes, y que d_{il} sea la distancia entre i y l (a ser calculada en forma general por $\sqrt{(x_i - x_l)^2 + (y_i - y_l)^2}$ y a veces como "distancia por carretera", "distancia por ferrocarril", etc.).

Para estudiar la diseminación de los casos en el territorio, podemos, aunque el número de casos sea grande, calcular la distribución de todas las distancias $\frac{1}{2} n(n-1)$ entre ellos. El cuadro 1 da algunos ejemplos de estos cálculos que muestran la diseminación de algunas poblaciones urbanas, por medio de distribuciones de "distancias por carretera" entre los habitantes. Distribuciones de distancias como las que se dan en el cuadro 1 se pueden utilizar no sólo para comparar la diseminación de la población en varias ciudades (o países)

^{1/} Este trabajo es un resumen de un artículo más extenso actualmente en preparación, en el cual se dan detalles de los cálculos y demostraciones de las diversas fórmulas.

Cuadro 1

DISTANCIA POR CARRETERA (EN Km) ENTRE LOS HABITANTES DE ROMA Y ZURICH							
	0-2	2-4	4-6	6-8	8-11	Total	Promedio (metros)
(1) Distancias reales entre habitantes							
Roma 1925	29 567	43 192	22 921	3668	652	100 000	3123
Zurich 1915	33 575	48 270	13 155	-	-	100 000	2688
1925	32 560	48 563	18 876	-	-	100 000	2717
(2) Distancias recorridas por los migrantes							
Roma 1924-26	48 468	32 663	15 051	3169	649	100 000	2470
3) Razón $\frac{(2)}{(1)} \cdot 100$							
Roma 1924-26	164	76	66	86	100	100	79

sino que también para estudiar los cambios en el tiempo de la diseminación de las poblaciones o de otros fenómenos, para comparar la diseminación de diversas clases de la población, y para evaluar la extensión de las distancias recorridas por los migrantes en el territorio (véase sección 4.2).

2.2. La distancia normal (Standard Distance)

La distribución de las distancias puede resumirse tomando su mediana, su media aritmética (véase Cuadro 1), la raíz cuadrada de la media aritmética de los cuadrados de las distancias (o "distancia cuadrática media", etc.) Este último valor puede calcularse directamente por las longitudes y latitudes de los puntos, sin medir de hecho cada distancia.

La distancia cuadrática media con repetición 2/

$$2D_R = \sqrt{\frac{\sum_{i=1}^s \sum_{j=1}^s f_i f_j d_{ij}^2}{n^2}} = \sqrt{2 \left[\frac{\sum_{i=1}^s f_i x_i^2}{n} - \left(\frac{\sum_{i=1}^s f_i x_i}{n} \right)^2 + \frac{\sum_{i=1}^s f_i y_i^2}{n} - \left(\frac{\sum_{i=1}^s f_i y_i}{n} \right)^2 \right]} \quad (2.1)$$

Se ve según (2.1) que la distancia cuadrática media con repetición es igual a $\sqrt{2}$ veces al promedio cuadrático de las distancias de cada uno de los casos desde su centro. Tomando las coordenadas del centro como

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}, \quad \bar{y} = \frac{\sum f_i y_i}{\sum f_i} \quad (2.2)$$

se encuentra que el promedio cuadrático de las distancias, que puede llamarse "distancia normal" (standard distance) entre cada uno de los casos y su centro, es

2/ La fórmula se da aquí para las distancias cuadráticas medias "con repetición" (a saber, incluyendo las distancias entre cada caso y él mismo) y para los datos agrupados en s clases (secciones del territorio, distritos, lugares, etc.)

$$\text{Distancia normal} = d = \sqrt{\frac{\sum_{i=1}^s f_i x_i^2}{n} - \left(\frac{\sum_{i=1}^s f_i x_i}{n}\right)^2 + \frac{\sum_{i=1}^s f_i y_i^2}{n} - \left(\frac{\sum_{i=1}^s f_i y_i}{n}\right)^2}$$

$$= \sqrt{\frac{\sum_{i=1}^s f_i (x_i - \bar{x})^2}{n} + \frac{\sum_{i=1}^s f_i (y_i - \bar{y})^2}{n}}$$

(2.3)

La distancia normal (standard distance) puede calcularse con mucha rapidez tomando la raíz cuadrada de la suma de las variaciones de longitud y latitud de los casos estudiados. Si se dan frecuencias para un gran número de lugares habitados o secciones del territorio, pueden re-ordenarse convenientemente, una vez de acuerdo a las clases de longitud, y una vez de acuerdo a las clases de latitud. Se obtienen así dos distribuciones de frecuencia, sobre las que se ejecutan los cálculos acostumbrados para el segundo momento con respecto a la media.

La distancia normal (standard distance) parece ser la medición más sencilla de la dispersión geográfica a ser asociada con el centro (considerado como un índice de ubicación del fenómeno estudiado). Permite, entre otras cosas, describir sintéticamente la dispersión real de un fenómeno, compararlo con la dispersión que cabe esperar según ciertas hipótesis, evaluar su cambio en el tiempo, y compararlo con la dispersión de otros fenómenos.

2.3. Ejemplos de aplicaciones de la distancia normal (standard distance)

Se ha aplicado la "distancia normal" (standard distance) para evaluar la diseminación de las poblaciones humanas. Por ejemplo, al estudiar la distribución geográfica de la población de Israel en 1953, encontramos que la distancia normal (standard distance) real era de 48,5 kms. en comparación con un valor teórico de 105,5 km, que se hubiera encontrado según la hipótesis de la distribución uniforme de la población en el territorio. Se encontró que esta discrepancia se debía a los siguientes factores: (1) que los lugares habitados (que tienen una distancia normal (standard distance) de 64,1 km) de hecho están esparcidas sólo en parte del territorio; 2) que la población se halla distribuida en forma dispereja entre estos lugares (según la hipótesis de la asignación aleatoria de poblaciones reales a los lugares habitados, la distancia normal (standard distance) hubiera sido de 61,9 km); 3) que las ciudades más grandes están menos diseminadas que los poblados más pequeños.

Otras aplicaciones de la distancia normal (standard distance) parecen sugerir que el método puede ser útil en muchos campos tales como 1) el análisis comparativo de las distribuciones de diferentes grupos étnicos, sociales o políticos, o de oleadas de inmigrantes por el territorio de un país o de una ciudad, y de sus cambios en el tiempo; 2) estudios epidemiológicos; 3) el análisis de estadísticas agrícolas e industriales

(distribuciones de las culturas, de diversas ramas de la producción, etc.); 4) análisis de la distribución geográfica de los servicios de salud, educacionales o sociales en comparación con la distribución geográfica de la población cubierta, etc.

2.4. Aplicación al estudio de poblaciones heterogéneas y a la subdivisión de una población por zonas geográficas

Puede mencionarse otra aplicación del método, que resulta útil para el análisis de la diseminación geográfica de una población heterogénea, es decir, de una población compuesta de grupos étnicos, sociales o económicos distintos. Tomemos otra vez como ejemplo la distribución geográfica de la población de Israel. Esta población puede descomponerse en dos grupos principales: a) los judíos y b) los musulmanes, cristianos y drusos. El cuadrado de la distancia normal (standard distance) de la población con respecto al centro puede subdividirse en dos partes como se muestra en el Cuadro 2: una que se "explica" por la tendencia de cada grupo a concentrarse alrededor de su propio centro, y la otra que permanece "sin explicación". La razón entre la parte "explicada" y el total es en el ejemplo considerado $e_1 = 11,76$ por ciento.

El mismo método también puede aplicarse a las series geográficas que pueden dividirse en sub-series, dividiendo el territorio en zonas de acuerdo con criterios históricos, económicos, sociológicos, geográficos, etc. Para cada subserie (incluida en una zona) se calcula el centro y la distancia normal (standard distance) y se describe toda la serie por medio de estos valores.

Cuadro 2

CUADROS DE LAS DISTANCIAS NORMALES (STANDARD DISTANCES) DE DOS GRUPOS DE POBLACION EN ISRAEL

	Población judía	Poblac. musulmana, cristiana y drusa	Población total
1. Nº habitantes incl. en cál. ^{a/}	1 445 659	148 098	1 593 757
2. Promedio del cuadrado de las distancias desde el centro de la población total (km.)	22 123	37 599	
3. Promedio de los cuadrados de las distancias desde el centro del propio grupo (km)	21 837	10 578	
4. Cuadrado de la distancia entre el centro del grupo y el centro de la población total (km)	286	27 021	
5. Razón del cuadrado de la distancia "explicada", por ciento	1,29 ^{b/}	71,87 ^{b/}	11,76 ^{c/}

^{a/} Datos para fines de 1953. Los centros de recepción de nuevos inmigrantes, la población beduina nómada y ciertos otros grupos fueron excluidos del cálculo.

^{b/} Razón (4)/(2).

^{c/} La razón $\Sigma(4).(1)/\Sigma(2).(1)$ se obtuvo sumando los datos para los dos grupos de población.

A veces, cuando el territorio es subdividido en un cierto número (s) de zonas, surge la pregunta de si el marco particular de división adoptado es mejor que otros marcos de alternativa de división en s zonas.

Al permanecer los demás factores iguales, se debe preferir aquel marco que tiene el valor más alto de e_1 , es decir, el marco que hace mínima la suma del cuadrado de la distancia "dentro de la zona" y que hace máxima la suma del cuadrado ponderado de la distancia entre los centros de las zonas y el centro general.

3. Tratamiento de las series geográficas que muestra las distribuciones de frecuencia o promedios

Se puede considerar que el método que se ilustra en el Cuadro 2 es apropiado al estudio de cualquier serie geográfica de distribuciones de frecuencia de atributos y variables (v.g. la distribución de la población de cada distrito por ocupación, por grado de instrucción o por ingresos, etc.) Para cada atributo (v.g. para cada ocupación) o para cada clase de la variable (v.g. para cada grupo de ingresos) se calculan los centros y las distancias normales (standard distances), y estos valores se comparan con los de toda la población.

En el caso de las variables, pueden introducirse a veces algunos métodos adicionales de tratamiento. Por ejemplo, sobre la base de datos sobre la distribución de los reclutas del Ejército italiano por región geográfica y estatura se ha encontrado que los "centros" de los grupos de reclutas cuya estatura es de "menos de 150 cm.", "151-155 cm.", ..., "181 cm. y más", respectivamente, siguen una línea bastante regular, en que la estatura aumenta de sur a norte y de este a oeste. La distancia normal (standard distance) de todos los reclutas dentro de cada grupo es por término medio algo menor que la distancia normal (standard distance) de todos los reclutas, y se encuentra que e_1 es de 5,76 por ciento. Por interpolación, podemos obtener ecuaciones de cálculo

$$x = 10,47 - 0,064 h; \quad y = 27,51 + 0,092 h$$

de las longitudes (x) y de las latitudes (y) de los centros de los reclutas de estatura h.

El cuadrado de la distancia normal (standard distance) entre los lugares de residencia de los reclutas y su centro general puede dividirse en dos partes: 1) el promedio de los cuadrados de la distancia de cada caso respecto al centro estimado de su clase (se obtiene el centro mediante las ecuaciones de estimación dadas anteriormente); 2) el promedio ponderado de los cuadrados de las distancias de los centros estimados de cada clase con respecto al centro general. Calculando $100 \frac{(2)}{(1) + (2)}$, se obtiene una razón $e_2 = 5,56$ por ciento la que resulta ser algo menor que la razón e_1 .

Mediante el empleo de un método semejante a veces se pueden analizar series que muestran la intensidad promedio de un fenómeno en cada sección del territorio. Por ejemplo, las series que muestran el porcentaje de analfabetismo en cada una de las 92 provincias de Italia (1931) fueron analizadas por medio del cálculo

de los centros y de las distancias normales de la población de grupos de provincias clasificadas de acuerdo al analfabetismo. Uniendo los centros de los diversos grupos de provincias se obtiene una línea que parece proporcionar una indicación útil de la dirección geográfica del analfabetismo que crece de norte a sur y de oeste a este. Las distancias normales muestran que la diseminación es mucho menor dentro de cada grupo de provincias que en el país en conjunto ($e_1 = 78,2$ por ciento). Según las ecuaciones de cálculo de las longitudes (x) y de las latitudes (y) de los centros de provincias con un promedio de analfabetismo f ($x = 0,151 f + 3,568$, $y = -0,164 f + 46,510$) se encuentra un valor de $e_2 = 73,7$ por ciento.

4. El tratamiento de las series que muestran movimientos clasificados por lugares de salida y de destino

Al estudiar estas series, incluso el estadístico más experto puede verse confundido por lo intrincado de los datos básicos. En un país que está subdividido en 200 distritos por ejemplo, las estadísticas sobre los movimientos internos de las personas (artículos de primera necesidad, camiones, etc.) darán 40 000 cifras, que deben tratarse y sintetizarse de alguna manera para que se pueda llegar a alguna conclusión. El tratamiento de los datos puede ayudar tanto i) a describir el fenómeno como ii) a probar algunas hipótesis, que podrían "explicar" ciertos aspectos del fenómeno.

Se discuten a continuación tres grupos de métodos para el tratamiento de las series de movimientos.

Como ejemplo de las series de movimientos consideremos los datos sobre migraciones internas entre las diversas secciones de la ciudad de Roma (1925).

4.1. Índices de preferencia o de aversión para cada dirección de los movimientos

Para estudiar las direcciones de la migración interna, puede resultar útil comparar los datos reales con datos teóricos que se obtendrían de acuerdo a ciertas hipótesis. Las siguientes son dos hipótesis muy sencillas:

i) Supongamos que las migraciones internas no son influidas por los factores geográficos u otros factores que diferencian a las distintas secciones. Entonces a) el número de personas que salen de cada sección y el de personas que llegan a cada sección debiera ser proporcional al número de habitantes de las secciones y b) los migrantes desde y hacia cada sección, como se calculó anteriormente, debieran cambiar sus respectivas residencias al azar. Indiquemos la población de la sección i por P_i ; la de la sección l por P_l ; la población de todo el territorio por P ; el número de migrantes de la sección i a la sección l por i^m_l ; el número de migrantes que provienen de la sección i por i^m ; el número de migrantes que llegan a la sección l por m_l ; y el número total de migrantes por m .

El número esperado de migrantes desde i a l en la hipótesis anterior sería entonces de $m P_i P_l / P^2$. Al calcular las razones

$$100 \frac{i^m_l}{m P_i P_l / P^2} \quad (4.1)$$

entre los números reales y los teóricos de los movimientos desde la sección i a la sección l para la migración interna entre 10 secciones de la ciudad de Roma (1925), encontramos que 1) los movimientos se producen en todas las direcciones; 2) existe una tendencia marcada entre los migrantes a permanecer en la sección en que residieron con anterioridad; 3) los migrantes que cambian de sección muestran preferencia por las secciones que se hallan cercanas a su residencia anterior; 4) las secciones centrales son las menos atractivas, y generalmente pierden frente a las otras secciones; las secciones intermedias ganan parcialmente de las centrales, y pierden frente a las periféricas, y éstas últimas por lo general ganan. El carácter centrífugo de la migración interna es por lo tanto claro.

(i) Una segunda hipótesis de alternativa que puede sugerirse para comparar los datos reales con los datos teóricos es la usada en general en el análisis de las tablas de contingencia. Supongamos que ya hemos estudiado la influencia de las tendencias generales, como las que se mencionaron anteriormente en el número 4), a migrar desde el centro a la periferia, etc. Podemos tomar ahora los totales marginales dados por las estadísticas (es decir, el número de personas que entran y que salen de cada sección) como números fijos. Si el lugar anterior de residencia no influyó en la elección de la nueva residencia, podemos suponer que los migrantes habrían cambiado sus respectivas residencias al azar. Según esta hipótesis el número esperado de personas que migran desde i a l deber ser $i^m \cdot m_l / m$. Si i^m_l sobrepasa este número esperado, se puede calcular un índice de atracción en la dirección $i-l$ tomando:

$$100 \frac{i^m_l - i^m \cdot m_l / m}{i^m \cdot m_l / m} \quad (4.2)$$

Este índice es de +100 por ciento cuando todas las personas que antes residieron en la sección i , tienen su nueva residencia en la sección l . Si i^m_l es menor que el número esperado, puede calcularse un índice de aversión en la dirección $i-l$. Los índices calculados en esta forma para la migración interna en Roma muestran una atracción hacia la sección anterior de residencia y secciones vecinas, y una aversión hacia otras secciones (que crece al aumentar la distancia respecto a las secciones de residencia anterior).

4.2 Distancias recorridas por los migrantes

Las distribuciones reales y teóricas de las migraciones pueden convertirse fácilmente en distribuciones según la distancia recorrida por los migrantes. En el Cuadro 1 se da un ejemplo de estas distribuciones para la migración interna en Roma (1925). Este cuadro muestra que las razones entre los números reales y los actuales van disminuyendo desde aquéllas para las distancias cortas hacia aquéllas de hasta 6 kms., a causa de la fuerte tendencia que muestran los migrantes a permanecer cerca de sus anteriores lugares de residencia. Las razones crecen, por otra parte, después del sexto kilómetro y llegan a 100 al final de la

$$z_D^2 = [(\bar{x}_A - \bar{x}_D)^2 + (\bar{y}_A - \bar{y}_D)^2] + d_D^2 + d_A^2 - 2(r_{D,A,x} \sigma_{D,x} \sigma_{A,x} + r_{D,A,y} \sigma_{D,y} \sigma_{A,y}) \quad (4.10)$$

Esta fórmula nos permite encontrar i) un límite inferior al valor mínimo del cuadrado del promedio cuadrático de las distancias ($z_{D,\min}^2$) suponiendo que tanto $r_{D,A,x}$ como $r_{D,A,y}$ son iguales a 1; ii) un límite superior al valor máximo ($z_{D,\max}^2$), suponiendo que $r_{D,A,x} = r_{D,A,y} = -1$; iii) el valor en el caso de la asociación aleatoria de los puntos de salida y de destino, suponiendo que:

$$r_{D,A,x} = r_{D,A,y} = 0$$

Los valores anteriores nos permiten calcular un índice que mide la no aleatoriedad en la asociación de los puntos de salida y los puntos de destino de los migrantes. El índice se obtiene calculando la razón entre: a) el excedente ^{6/} del cuadrado del promedio cuadrático de las distancias en el caso de la asociación aleatoria por encima del cuadrado del promedio cuadrático de las distancias reales y b) el máximo valor posible de este excedente. Este índice varía entre 0 y +1.

$$\frac{z_D^2 - z_{D,\min}^2}{z_{D,\max}^2 - z_{D,\min}^2} = \frac{r_{D,A,x} \sigma_{D,x} \sigma_{A,x} + r_{D,A,y} \sigma_{D,y} \sigma_{A,y}}{\sigma_{D,x} \sigma_{A,x} + \sigma_{D,y} \sigma_{A,y}} = r_{D,A} \quad (4.11)$$

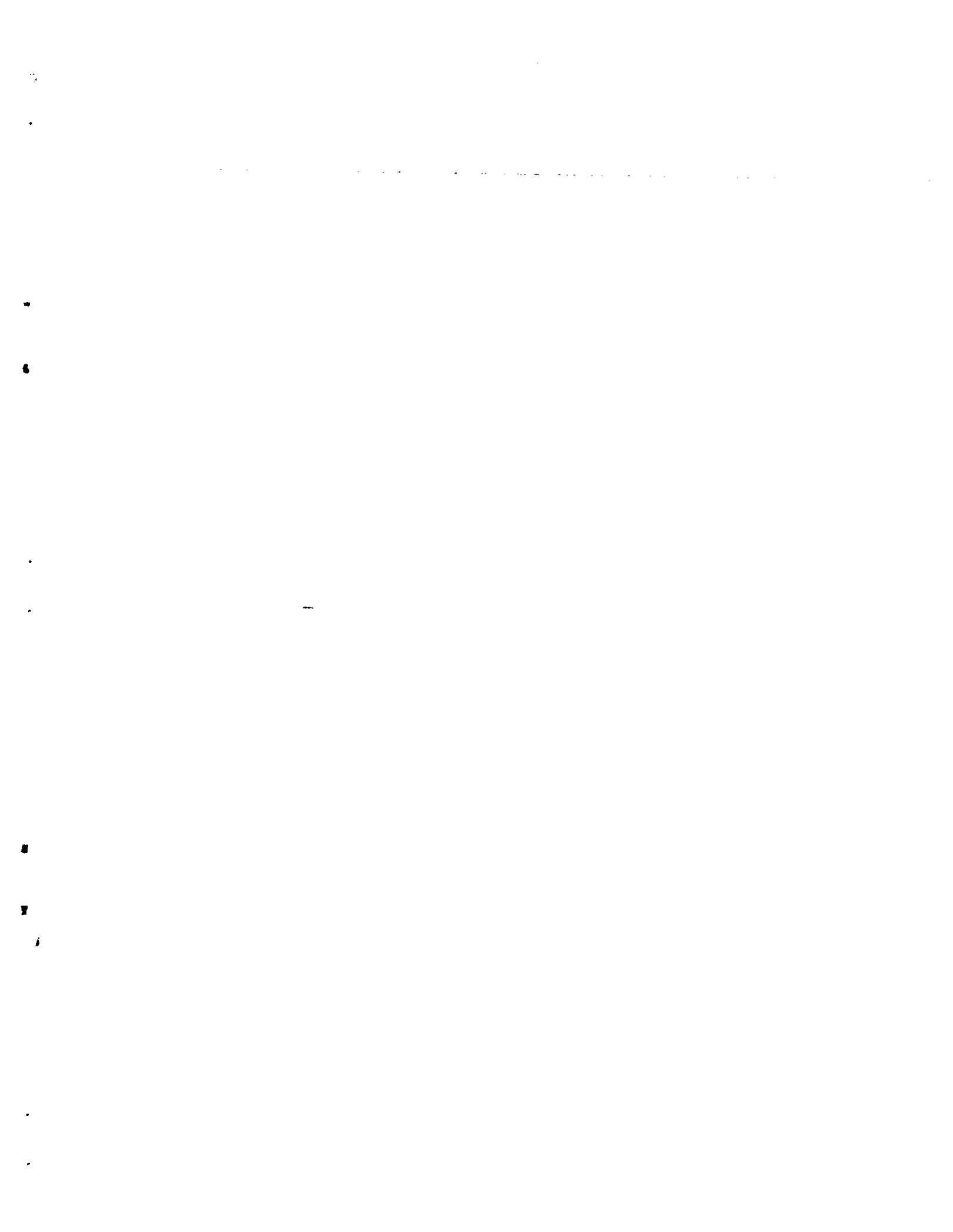
En el caso de las migraciones internas en Roma este índice es de 0,4042.

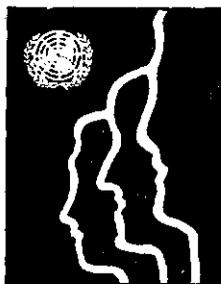
A veces se puede introducir un valor aun más sintético que z_D^2 , a saber z_{D,p_0,p_t}^2 , que mide el cuadrado de la distancia cuadrática media entre el lugar de residencia en el tiempo 0 y el lugar de residencia en el tiempo t para todos los habitantes, tanto migrantes como no migrantes. ^{7/} Puede demostrarse que este valor puede ser dividido en sus componentes con respecto a parámetros de población en el tiempo 0 y en el tiempo t, o en un número mayor de componentes, respecto a los no migrantes, a los migrantes que salen y a los migrantes que llegan, por separado.

^{6/} Suponemos que $r_{D,A}^2 > z_D^2$. Puede interpretarse que la "no aleatoriedad" se debe a la tendencia a elegir nuevos lugares de residencia que se hallen más cercanos a los antiguos, que bajo la hipótesis aleatoria.

En el caso sumamente improbable de que $r_{D,A}^2 < z_D^2$ podría calcularse una razón $(z_D^2 - r_{D,A}^2) / (z_{D,\max}^2 - r_{D,A}^2)$

^{7/} El conocimiento de estas distancias se obtiene a veces de los censos o de las muestras en que se les solicita a los habitantes que consignen, además de su actual lugar de residencia, su lugar de residencia en una cierta fecha en el pasado. Puede aplicarse una fórmula análoga a la distancia entre el "lugar de residencia" y el "lugar de trabajo".





**CENTRO LATINOAMERICANO DE DEMOGRAFIA
CELADE**

**Sede: J.M. Infante 9. Casilla 91. Teléfono 257806
Santiago (Chile)**

**Subsede: Ciudad Universitaria Rodrigo Facio
Apartado Postal 5249
San José (Costa Rica)**