El Índice de Correlación Ordinal (de Spearman) y su Aplicación al Estudio Estadístico de la Soledad

ÓSCAR URIBE VILLEGAS

0. *Introducción*

El propósito de estas páginas es: (1) obtener, a partir de una fórmula genérica de correlación otra específica, de inmediata utilización en la estadística ordinal; (2) usar la fórmula específica con datos concretos de la realidad social de México; (3) hacer una interpretación tentativa de resultados y (4) mostrar cómo gracias a dicha interpretación se abre una brecha para iniciar el estudio ceñido de un problema social.

En concreto: (1) derivaremos de una fórmula general de correlación, el índice coordinativo de Spearman; (2) aplicaremos ese índice a algunos datos del Anuario Estadístico de México; (3) reuniremos los resultados obtenidos de coordenar los datos que dicho *Anuario* proporciona sobre número de personas solas y los que se refieren a otros aspectos de la vida social de México y (4) haremos un intento de interpretación de ese problema (que se ha tratado más desde el ángulo filosófico que desde el sociológico aunque es de enorme importancia social).

1. DERIVACIÓN DE LA FÓRMULA DE SPEARMAN

Tomaremos como punto de partida lo que, como dato, nos proporciona la estadística matemática: una fórmula genérica de correlación. Ésta, incluso en su aspecto formal, tiene la apariencia de la muy utilizada de Pearson-Bravais.

1.1 *Fórmula genérica de correlación*

Podemos definir en forma genérica el índice de correlación en la forma siguiente:

El índice de correlación es un cociente de: (1) una suma de productos, entre (2) una raíz cuadrada.

Los productos que se suman 1º son de: (1.1) la diferencia de la calificación de cada miembro del conjunto y una primera característica y (1.2) una diferencia análoga de cada miembro y una segunda característica.

La raíz cuadrada lo es de un producto (2.1) de la suma de los cuadrados de las diferencias citadas en (1.1) por (2.2) la suma de los cuadrados de las otras diferencias (1.2).

Si designamos por:

d— las diferencias de calificación, según la primera característica o criterio.

D.— las diferencias, según la segunda, los subíndices *i* y *j*, la especificación de los miembros del conjunto.

d_{ij} representará la diferencia de calificación entre *i* y *j* de acuerdo con el primer criterio.

D_{ij} la diferencia de calificación entre *i* y *j* según el segundo criterio.

De acuerdo con lo anterior, el numerador del índice de correlación se obtendrá de los pares de diferencias multiplicados entre sí, o sea las d_{ij} y D_{ij} y de la suma de la serie de productos resultante.

$$\sum d_{ij}\ D_{ij}$$

El denominador estará dado por la raíz cuadrada del producto de las sumas de los cuadrados de dichas diferencias, o sea, por

$$(\sum d_{ij}^2\ D_{ij}^2)^{1/2}$$

Según esto, la fórmula del índice de correlación, que designaremos por *I* resultará ser:

$$I = \frac{\sum d_{ij}\ D_{ij}}{(\sum d_{ij}^2\ D_{ij}^2)^{1/2}}$$

1.2 *Fórmula específica de correlación ordinal*

Esta fórmula genérica puede especificarse. Una especie de clasificación consiste en dar un número de orden (creciente o decreciente) a cada miembro del conjunto, al ordenar a todos ellos de acuerdo con el valor que les tienen: primero, de acuerdo con una característica y, después, según una segunda característica. Esta especificación es particularmente útil en el campo de las ciencias sociales, pues, conforme se ha reconocido muy generalizadamente, si no siempre es posible cuantificar precisamente los fenómenos sociales, por falta de unidades adecuadas, es mucho más fácil el darles algún orden creciente o decreciente; se usan así evaluaciones que, en otras condiciones, se considerarían burdas, pero que, en nuestro actual nivel de desenvolvimiento académico son las únicas asequibles para la iniciación de un trabajo científico.

Si el ordinal que corresponde a un miembro, de acuerdo con un primer criterio de ordenación (o de acuerdo con el valor de una primera característica) lo designamos por p, y el ordinal que le corresponde de acuerdo con un segundo criterio (o de acuerdo con el valor de una segunda característica) por q, la diferencia de orden entre los miembros i y j del conjunto quedará representada por $d_{ij} = p_i - p_j$ la diferencia de orden entre los miembros i y j del conjunto, de acuerdo con el segundo criterio, quedará representada por $D_{ij} = q_i - q_j$

De acuerdo con lo anterior, la fórmula genérica se convertirá, en cuanto la calificación se haga por ordenación de los miembros del conjunto, en:

$$\frac{\Sigma(p_i - p_j)(q_i - q_j)}{[\Sigma(p_i - p_j)^2 \, \Sigma(q_i - q_j)^2]^{1/2}} = C$$

Es fácil mostrar (en el denominador) que la suma de los cuadrados de las diferencias entre los ordinales correspondientes a la primera ordenación (u ordenación de acuerdo con un primer criterio) es igual a la suma de los cuadrados de las diferencias entre los ordinales correspondientes a la segunda ordenación:

$$\Sigma(p_i - p_j)^2 = \Sigma(q_i - q_j)^2$$

Por lo mismo, puede escribirse:

$$\frac{\Sigma(p_i - p_j)(q_i - q_j)}{\Sigma(p_i - p_j)^2}$$

ya que, en el denominador, aparecería la raíz cuadrada del cuadrado de la suma anotada, que se reduce a dicha suma de cuadrados de las diferencias de orden, de acuerdo con uno de los criterios de ordenación (el primero o el segundo).

Como en el numerador aparece una suma de productos, se pueden desarrollar los productos y aplicar, en seguida, los operadores sigmáticos. El desarrollo del producto de factores binomiales produce:

$$(p_i - p_j)(q_i - q_j) = p_i q_i - p_i q_j - p_j q_i + p_j q_j$$

A los términos resultantes del desarrollo —en los que los campos de variabilidad están dados por los subíndices i y j— habrá que aplicarles un doble operador sigmático que extienda los límites de las sumatorias en dos direcciones (en la dirección $i = 1 \ldots n$ y en la dirección $j = 1 \ldots n$).

El resultado del desarrollo y de la aplicación del operador es:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} p_i q_i - \sum_{i=1}^{n} \sum_{j=1}^{n} p_i q_j -$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} p_j q_i + \sum_{i=1}^{n} \sum_{j=1}^{n} p_j q_j$$

que es el equivalente del numerador.

En este equivalente del numerador, el primer término del desarrollo $p_i q_i$ tiene una variabilidad que se extiende sólo en la dirección i; o sea, que dicho término es constante para el operador sigma mayúscula que se extiende en la dirección j. Como ocurre siempre que se tiene una constante ($p_i q_i$ es constante para el sumador en j) afectada por un sumador, el resultado es la constante ($p_i q_i$) multiplicada por el número de casos (n); o sea que:

$$\sum_{j=1}^{n} p_i q_i = n p_i q_i$$

Consecuentemente, todo el primer término del numerador tiene los siguientes equivalentes:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} p_i q_i = \sum_{i=1}^{n} n\, p_i q_i = n \sum_{i=1}^{n} p_i q_i$$

La última equivalente se justifica en cuanto un factor constante (n) puede entrar o salir del sumador sin alterar la sumatoria.

En forma parecida, y en relación con el último término del desarrollo del numerador, puede demostrarse que:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} p_j q_j = n \sum_{j=1}^{n} p_j q_j$$

Por otra parte, es fácil mostrar que:

$$\sum_{i=1}^{n} p_i q_i = \sum_{j=1}^{n} p_j q_j$$

Según esto, la suma de los dos miembros extremos (primero y último) del desarrollo del numerador se reduce a:

$$2n \sum_{i=1}^{n} p_i q_i$$

En cuanto a la suma de los dos miembros centrales (segundo y tercero) del desarrollo:

$$-\sum_{i=1}^{n} \sum_{j=1}^{n} p_i q_j - \sum_{j=1}^{n} \sum_{j=1}^{n} p_j q_i$$

Cada uno de ellos es igual a:

$$p_1 q_1 + p_1 q_2 + \ldots\ldots\ldots\ldots p_1 q_n = p_1 (q_1 + q_2 + \ldots\ldots q_n)$$

$$p_2 q_1 + p_2 q_2 + \ldots\ldots\ldots\ldots p_2 q_n = p_2 (q_1 + q_2 + \ldots\ldots q_n)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$p_n q_1 + p_n q_2 + \ldots\ldots\ldots\ldots p_n q_n = p_n (q_1 + q_2 + q_n) \ldots\ldots$$

O sea, que cada uno de ellos es igual a:

$$p_1 \sum_{i=1}^{n} q_1 +$$

$$+ p_2 \sum_{i=1}^{n} q_i +$$

$$+ \ldots\ldots\ldots\ldots +$$

$$+ p_n \sum_{i=1}^{n} q_i$$

Y, finalmente, que también es igual a:

$$+ \sum_{j=1}^{n} p_j \sum_{i=1}^{n} q_i$$

Es decir, que la suma de los dos miembros centrales (negativos) del numerador se reduce a dos veces ese producto de sumatorias, afectado el resultado por el signo menos. Es decir, que el índice de correlación fundado en la coordenación tiene como numerador:

$$2n \sum_{i=1}^{n} p_i q_i - 2 \sum_{j=1}^{n} p_j \sum_{i=1}^{n} q_i$$

De estos dos términos, consideraremos —en particular— el segundo. Cada sumatoria del producto que en él aparece es la suma de los ordinales de los miembros, o sea, que cada sumatoria es igual a la suma de los n primeros números naturales. Como es sabido, la suma de los primeros números naturales (como puede demostrarlo la aplicación de la fórmula de la suma de los términos de una progresión aritmética) es igual a la mitad del producto del producto del último número de la serie por el que le seguiría inmediatamente, o sea: n por $n + 1$ sobre 2. En cuanto en el segundo término del numerador de nuestro índice de coordenación existen dos factores iguales a n por $n + 1$ sobre 2, el resultado será n^2 por $(n + 1)^2$ sobre 2^2. Pero, como este resultado aparece multiplicado por 2, el denominador 2^2 se reduce a dos. Según esto, todo el segundo término del numerador es igual a:

$$\frac{-n^2 (n + 1)^2}{2}$$

Para obtener el equivalente del primer término del numerador, comenzaremos por hacer igual a d (sin sufijos) la diferencia entre los ordinales que corresponden a un mismo miembro del conjunto considerando en los dos ordenamientos distintos:

$$d = p_i - q_i$$

La suma de los cuadrados de estas diferencias será:

$$\Sigma d^2 = \sum_{i=1}^{n} (p_i - q_i)^2 = \Sigma p_i^2 - 2\Sigma p_i q_i + \Sigma q_i^2$$

Tanto el primero como el último término son la suma de los cuadrados de los ordinales o sea, que ambos son iguales a la suma de cuadrados de los n primeros números naturales. Como se sabe, ésta es igual a la suma de los números, no elevados al cuadrado (o sea, n por $n + 1$ sobre 2), por $2n + 1$ sobre 3. Como son dos los sumandos iguales, la suma de los dos será igual a $n\ (n + 1)$ por $2n + 1$ sobre 3. Toda la expresión anterior se convertirá en:

$$d^2 = n\ (n + 1\ \frac{2n + 1}{3} - 2 \sum_{j} p_i q_i$$

Si despejamos el último término, tendremos:

$$2\ \Sigma p_i q_i = n\ (n + 1)\ \frac{2n + 1}{3} - d^2$$

Si substituimos en el numerador del índice de coordenación este equivalente del primer término y el que encontramos antes como equivalente del segundo, tendremos:

$$\text{NUMERADOR} = 2n \sum p_i q_i - 2 \sum p_j q_i + \sum q_i =$$

$$= \left[n^2 (n+1) \frac{2n+1}{3} - n \sum d^2 - \right] n^2 (n+1) \frac{n+1}{2}$$

En esa expresión, destacamos entre corchetes el equivalente del primer término, y el equivalente del segundo lo descompusimos en factores, a fin de hacer notar la existencia de un factor común con el primer término de la expresión entre corchetes. Ese factor complejo, común al primero y al tercer término del desarrollo, puede colocarse como factor común, fuera del paréntesis.

$$\text{NUMERADOR} = n^2 (n+1) \left[\frac{2n+1}{3} - \frac{n+1}{2} \right] - n \sum d^2$$

La resta de quebrados incluida en los corchetes, se reduce fácilmente a $n - 1$, sobre 6; multiplicado dicho resultado por el factor complejo común, se obtiene $n^2 (n+1)(n-1) = n^2 (n^2 - 1)$; o sea que, finalmente, el equivalente del numerador es:

$$\frac{n^2 (n^2 - 1)}{6} - n \sum d^2$$

El denominador de nuestro índice de coordenación está constituido por:

$$(p_j - p_i) = \sum_i \sum_j p_i^2 - 2 \sum_i \sum_j p_i p_j + \sum_i \sum_j p_j^2$$

En los términos extremos —como en caso anterior— el sumando es constante para una de las sumatorias y variable para la otra, por lo que la doble sumatoria se convierte en el producto de n por una de las sumatorias. Los dos sumandos extremos son, en última instancia iguales, por lo que su suma es $2\,n \qquad p^2$. El término medio del desarrollo, por su parte, como doble sumatoria de un producto de factores cada uno de los cuales varía en el ámbito determinado por uno de los sumadores, se reduce fácilmente a $-2\sum_i p_i \sum_j p_j$ y, en última instancia a $-2\,(p_i)^2$ con lo cual, el denominador resulta ser:

$$\text{DENOMINADOR:} = 2n \qquad p^2 - (\qquad p)^2$$

El primer término es la suma de los cuadrados de los primeros números y el segundo es el cuadrado de la suma de los n primeros números; o sea, que el denominador resulta ser:

$$\text{DENOMINADOR} = 2\ n^2\,(n+1)\ \left[\frac{2n+1}{6} - \frac{n+1}{4}\right] = \frac{n^2\,(n^2-1)}{6}$$

De acuerdo con lo anterior, el índice de coordenación será:

$$C = \frac{\dfrac{n^2\,(n^2-1)}{6} - n\sum d^2}{\dfrac{n^2\,(n^2-1)}{6}} = 1 - \frac{n\,6\sum d^2}{n^4 - n^2} =$$

$$= 1 - \frac{6\sum d^2}{n^3 - n} =.$$

El índice resultante se designa convencionalmente con la letra griega ro minúscula y se conoce como índice de Spearman de acuerdo con el nombre de su inventor.

De acuerdo con esta fórmula, para calcular el índice de coordenación de Spearman:

1º Se ordenan los miembros del conjunto, universo o población en forma creciente (o decreciente) de acuerdo con los valores que les correspondan en razón de la primera de las características elegidas para la coordenación (o en razón de algún criterio apreciativo y no rigurosamente cuantitativo que se adopte para su ordenación).

2º Se ordenan los miembros del conjunto en la misma forma (creciente si la primera fue creciente, decreciente si así lo fue la primera) de acuerdo con el segundo criterio adoptado para la coordenación.

3º Anotados en sendas columnas, frente a cada uno de los individuos del conjunto, los ordinales que les hayan correspondido en la primera y en la segunda ordenación, se procede a encontrar la diferencia entre cada par de ordinales, a fin de obtener las *des*, que se anotan en una tercera columna.
4º Se elevan al cuadrado las diferencias encontradas, en otra columna.
5º Se suman dichos cuadrados al pie de la columna.
6º Se multiplica la suma por seis.
7º Se divide ese producto entre la diferencia que se obtiene de elevar al cubo el número de miembros de la población coordenada, y de restarle a dicho cubo el número de miembros, y
8º Se reste dicho cociente de la unidad, obteniéndose así el índice de coordenación de Spearman.

Es importante que el principiante note que cuando se manejan datos cuantitativos, y a base de ellos se hacen las ordenaciones, las diferencias a que se refiere el procedimiento deben ser, precisamente, diferencia *entre los ordinales* asignados en cada ordenación a los diversos miembros de la población, y *nunca* diferencias entre los valores o datos iniciales que pudieron servir para hacer la ordenación.

Por otra parte, el resultado final (como en casi todos los casos en que se calculan índices de correlación) puede variar entre + 1 y — 1 pasando por 0. Cuando el índice de coordenación es igual a 1, se dice que la correlación es perfecta, y que no existe correlación cuando el índice vale 0. Por otra parte, la correlación es directa cuando el índice de coordenación produce un valor positivo, e inversa si el índice produce un valor negativo.

2. APLICACIÓN DEL ÍNDICE COORDINATIVO DE SPEARMAN

Para ejemplificar, en concreto, el uso que puede hacerse del índice coordenativo de Spearman en el caso de los fenómenos sociales, tomamos algunos de los datos proporcionados por el *Anuario Estadístico de los Estados Unidos Mexicanos (1951-1952)*. Entre los datos proporcionados por dicho Anuario se encuentran los que se refieren al número de personas solas dentro de la población, que se proporcionan en el Cuadro Número

22. La distribución se hace, en dicho cuadro, por entidades federativas. Así, puede verse que hay 4 050 personas solas en Aguascalientes, para el año de 1950 al que se refieren los datos; que en Baja California Norte eran 16 071 las personas solas en ese año, y así sucesivamente. Los datos contenidos en ese cuadro 22 del Anuario los hemos transcrito en la primera columna de nuestro cuadro 1.

Con base en los datos correspondientes al número de personas solas es posible hacer una ordenación de las entidades federativas. Si comenzamos por buscar la entidad que tiene menor número de solos (1 505 en el caso de Quintana Roo) y le asignamos el número 1 (correspondiente al ordinal "primero"), determinamos cuál es la entidad que le sigue en cuanto a número de personas solas (2 050 en el caso de Baja California Sur) y le asignamos a ésta el número 2, continuando hasta encontrar la entidad que tiene el máximo número de solos (Distrito Federal con 122 039) y asignarle el número 32, último de nuestra ordenación, tendremos los resultados anotados en la segunda columna del cuadro 1.

Una ordenación simple como la que hemos hecho, nos puede enseñar varias cosas, pues mediante la misma, podremos señalar cuáles son las entidades que tienen un número normal de personas solas, en el conjunto (entidades marcadas con los números 9 al 24), cuáles las que tienen un número deficiente de personas solas y cuáles las que tienen un número excedente de solos (las marcadas con los números del 1 al 8 en el primer caso, y del 25 al 32 en el segundo).

Sin embargo, si esta ordenación permite describir, en alguna forma, la situación existente en México en 1950, por lo que se refiere a la distribución de las personas solas en las diferentes entidades federativas, la misma tiene varias limitaciones, y la principal de ellas es la de que en cuanto no relaciona estos datos con otros de la misma realidad social, impide el que se pueda llegar a postular alguna conexión causal entre el fenómeno de la soledad y otros fenómenos sociales.

Una primera relación —la más simple de todas— puede establecerse por medios matemáticos relacionando los valores absolutos registrados, con otros valores semejantes que sirvan de término de comparación. Más concretamente, una primera forma de relación puede establecerse dividiendo el total de personas solas en cada entidad entre la población total de dicha entidad. De este modo, pueden obtenerse una serie de porcientos (que hemos anotado en la columna 3 de nuestro cuadro 2) que dan debida consideración al hecho de que, independientemente de otros factores, es probable que el número de personas en la población de una entidad aumente al aumentar el efectivo demográfico de dicha entidad, y disminuya, al disminuir dicho efectivo. En caso de que no hubiese otros factores el resultado que debería esperarse de la división de la cifra de solos en la entidad entre el total de la población de la misma sería que el porciento correspondiente fuese constante para todas las entidades. El hecho de que los porcientos obtenidos no son constantes, indica, ya de por sí, que el volumen de personas solas no depende exclusivamente del volumen de población de la entidad de que se trate.

Si ordenamos los porcientos correspondientes al número de solos frente al número de habitantes, obtendremos los resultados anotados en la columna cuarta del cuadro 2, que podrá darnos una información menos burda de la situación del país por lo que se refiere al fenómeno de la soledad, pero que aún sigue siendo insuficiente en relación con las posibilidades de hacer imputaciones causales.

Pero, a partir del momento en que de la simple ordenación —u ordenación de acuerdo con un solo criterio— se pasa a la coordenación —u ordenación conjunta de acuerdo con dos criterios— comienzan a vislumbrarse las posibilidades de imputación causal o, por lo menos, de correlación estructural-funcional de los fenómenos sociales.

Así, por ejemplo, si en una columna de nuestro cuadro 3 anotamos los ordinales correspondientes a las entidades federativas de la República Mexicana registrados en el cuadro 2 (última columna) y, por otra parte —tomando en cuenta los datos de superficie proporcionados por el anuario— determinamos el número de orden que corresponda a esas mismas entidades de acuerdo con su extensión territorial, y anotamos dichos ordinales en una segunda columna de nuestro cuadro 3, estaremos en vías de coordenar dichas entidades, de acuerdo con el criterio conjunto de su tasa de soledad y de su superficie.

Determinados los ordinales de las entidades de acuerdo con cada uno de los dos criterios elegidos, bastará —de acuerdo con la fórmula de Spearman— con encontrar las diferencias entre los ordinales correspondientes a cada entidad (9 — 5 = 4 para Aguascalientes; 32 — 21 = 11 para Baja California Norte, etc.) y anotar dichas diferencias en una tercera columna de nuestro cuadro, elevándolas en seguida al cuadro y anotando los cuadrados en una cuarta columna que sumaremos al pie, para obtener como suma el valor 5 444 que substituiremos en vez de d^2 en la fórmula de Spearman, con lo que obtendremos como valor de ro minúscula o sea, del índice de coordenación de Spearman:

$$p = 1 - \frac{6 \times 5\,444}{32^3 - 32}$$

El resultado nos produce un índice de Spearman igual a 0.315. Este índice, positivo nos indica que hay una correlación directa entre la superficie de las entidades federativas y la tasa de soledad de la población, o sea, que en cuanto sea mayor la superficie de la entidad, mayor será el porciento de las personas solas dentro del total de la población, y que, conforme sea menor esa superficie, menor será el número de personas solas en el total de la población. Sin embargo, en cuanto el valor del índice es pequeño, el mismo indicará que si bien esta correlación directa se da entre los dos fenómenos, la misma es débil y, por lo mismo, serán endebles las imputaciones causales que se hagan con base en la coordenación obtenida.

Los cuadros subsiguientes muestran algunas otras coordenaciones. En ellos no se ha hecho otra cosa que repetir el procedimiento anterior para un número reducido de fenómenos que, en seguida se han coordenado con la tasa de soledad. Los resultados se han concentrado en un cuadro que nos da los valores de los diferentes índices de coordenación obtenidos, y en un cuadro más se han ordenado los índices de coordenación en orden creciente (comenzando por el índice negativo de máximo valor absoluto, y terminando con el índice positivo de máximo valor absoluto).

3. INTENTOS DE INTERPRETACIÓN

En el cuadro 22 del *Anuario* —como hemos dicho— bajo el encabezado "personas solas que no forman familia", y frente al rubro "Estados Unidos Mexicanos" se consignan 663 452 individuos. Si se considera que en el momento del recuento censal había, según el Propio Anuario, para el total del país 25 791 017, puede verse que en 1950, las personas solas que no formaban familia en México representaban el 2.57% del total de la población.

Si se ordenan las entidades federativas del país de acuerdo con el volumen total de personas solas existentes en cada una de ellas, se encuentra que Quintana Roo es de todas las entidades la que, en términos absolutos, tiene menor número de personas solas. En el otro extremo, en términos absolutos, el Distrito Federal es, de las entidades federativas mexicanas, la que tiene mayor número de personas solas. En un caso son poco más de mil quinientas, en el otro más de ciento veintidós mil. Ocupan posición intermedia, en este aspecto, Baja California Norte y Nuevo León.

Sin embargo, los valores absolutos y el orden resultante de dichos valores absolutos —como ya indicamos— resultan menos significativos, que ciertos valores relativos y el orden que de ellos resulte. En el caso, conviene calcular el porciento que representan las personas solas de una entidad en relación con la población total de dicha entidad para, en seguida, ordenar las entidades de acuerdo con el valor menos o más grande de dichos porcientos. Este cálculo y esta ordenación nos muestran que, ya relativizados los datos, es Sinaloa la entidad que tiene menor número de personas solas en relación con el total de su población, y que Baja California Norte tiene el máximo de personas solas en relación con el total de su población. Este segundo hecho es particularmente notable, pues bien podría suponerse que al Distrito Federal correspondiera un máximo de soledad relativa, cosa que no ocurre. En efecto, al Distrito Federal le corresponde una alta, pero no máxima, tasa de soledad estadístico-social, pues ocupa el lugar número 30, siendo superado por Baja California Norte, a la que corresponde el máximo, y Quintana Roo, que le sigue. La posición intermedia la ocupan Puebla e Hidalgo. La posición extrema de Baja California Norte debe subrayarse fuertemente pues en dicha entidad, el porciento de solos en la población es de 7.08, y esta cifra destaca incluso en contraste con la ya alta del Distrito Federal (4.00%) o con la de Quintana Roo (5.58%). De paso, conviene señalar el hecho de que en este caso, como en algún otro que hemos estudiado en alguna otra ocasión —porciento de personas originarias de la entidad frente a las procedentes de fuera de ella—, las entidades colocadas en la periferia del país presentan caracteres semejantes a la entidad central de la federación mexicana.

La cuarta parte del número de entidades a las que corresponden los porcientos más bajos de solos en la población comprende:

1. Sinaloa
2. Zacatecas
3. Michoacán
4. Guanajuato
5. México
6. Querétaro
7. Campeche
8. San Luis Potosí

La cuarta parte de entidades a las que corresponden los porcientos más altos de solos en la población está constituida por:

32. Baja California Norte
31. Quintana Roo
30. Distrito Federal
29. Sonora
28. Colima
27. Tamaulipas
26. Baja California Sur
25. Nayarit
24. Morelos.

En este momento, resultaría imprudente el que tratásemos de interpretar estos hechos, pues el nombre de cada una de estas entidades —como de todas las restantes—, ampara realidades sociales complejas que no hay que manejar con ligereza. Es por esto por lo que preferimos empezar explorando algunas posibilidades de vinculación entre el fenómeno de la soledad y otros aspectos, aislados, de la realidad de dichas entidades federativas.

Aun cuando haya que atender el llamado de Durkheim y deba de tratarse siempre de explicar lo social por lo social, se puede empezar preguntando si la soledad puede ponerse en función de la superficie en que se da dicho aislamiento. En efecto, no es imposible o absurdo pensar que a igualdad de número de pobladores, podrá haber mayores posibilidades de que se produzca un aislamiento obligado o buscado (soledad o apartamiento) si la superficie poblada es mayor en un caso que en otro.

3.1 *Soledad y extensión territorial*

Si establecemos una coordinación estadística entre el número de personas solas y la superficie de las entidades, obtendremos el cuadro 3. El cálculo del coeficiente de coordenación (ro minúscula) produce 0.315. O sea, que existe apenas una ligera correlación directa (ligera en cuanto más próxima a 0 que a 1, y directa en cuanto positiva) entre la superficie y el número de personas solas. Es decir que hay una ligera tendencia a que aumente el número de personas solas al aumentar la superficie de la entidad.

Sin embargo, esta conclusión nos enseña muy poco porque la conexión entre la superficie y el número de solos está mediatizada por la población total. En efecto, al aumentar la superficie *puede haber* —no que la haya necesariamente, pues esto depende de muchas otras condiciones como la topografía o la calidad de suelos de territorios que pueden tener igual extensión— una población mayor, y, asimismo, es posible suponer que, conforme aumente la población aumentará el número de aislados.

Si se obtiene la coordenación de la tasa de aislados respecto a la población total y la superficie de las entidades, puede verse que el índice correspondiente disminuye hasta convertirse en 0.003, o sea, que prácticamente la superficie no influye en nada sobre la magnitud de la tasa de aislados de la población. La conexión que se notaba antes entre la superficie y el número absoluto de aislados, aun siendo como era pequeña, no resultaba ser sino puramente aparencial, en cuanto había una mediatización por parte del total de habitantes. Algo parecido ocurriría con otros índices de coordenación que calculásemos utilizando valores absolutos; índices muy altos, que quedan ejemplificados por los dos o tres que figuran en la columna correspondiente del cuadro 2.

3.2 *Soledad y densidad de población*

Pero, si bien la superficie bruta de una entidad no puede explicar los cambios de magnitud de la tasa de soledad, parece que ocurre algo distinto si recurrimos a la densidad de población.

No obstante el enorme grado de simplicidad de este cociente obtenido de dividir sencillamente el total de los habitantes de una entidad entre el número de kilómetros cuadrados de la misma, dicha relación matemática tiene ya —al menos en principio— componentes sociales. En efecto, la población se distribuye en determinada forma y más o menos densamente en un territorio, por causas de orden físico (como las ya señaladas, de una topografía y una edafología diferenciales), pero lo hace así —en proporción no menor— por razones de carácter social y cultural, de mentalidad, de disponibilidades técnicas, etc.

Fuera de cualquier otra consideración, podría esperarse que a mayor densidad de población sería menor la tasa de aislados, pues podría suponerse —sin llegar al absurdo— que conforme la población fuese más densa, mayores serían las oportunidades que sus componentes tendrían de interrelacionarse, y menores las oportunidades que cada uno de ellos tendría de verse obligado al aislamiento o de poder darse el lujo de aislarse. El cálculo de la coordenación respectiva, nos produce como valor de ro minúscula, 0.265 con signo negativo. Esto significa que existe una ligera tendencia a que varíen inversamente la densidad de población de una entidad y la tasa de solos dentro de la población de la misma. El resultado se presta a reflexiones, a lucubraciones de todo tipo —que apuntarían quizás en el mismo sentido del "estar solo en la multitud"—, pero, sobre todo, parece que debe incitar a la realización de una investigación al respecto, sin la cual sería aventurada toda afirmación que se hiciese.

3.3 *Soledad y minoridad*

La coordenación entre la tasa de solos en la población y la proporción de los menores dentro de la misma también puede resultar de interés. Sin forzar mucho las cosas, puede llegar a postularse la hipótesis de que conforme aumente la proporción de menores de edad dentro de la población, disminuirá la proporción de personas solas. Esto parece casi obvio, en cuanto que los menores son generalmente dependientes en lo económico, así como en otros muchos aspectos y es, por lo mismo, altamente difícil que vivan solos (especialmente "solos" en los términos de la definición censal). El cálculo del índice de coordenación entre la proporción de solos y la proporción de menores, produjo un índice ro minúscula de menos 0.497, o sea, de casi 50%. El índice es alto, sobre todo en comparación con los que hemos obtenido previamente, y muestra que se trata de una coordenación inversa. O sea que, con base en los datos de que disponemos, podemos considerar que como válida la hipótesis de que el número relativo de personas solas dentro de la población aumenta si disminuye la proporción de menores en el total de la población, y disminuye cuando esta última proporción aumenta.

3.4 *Soledad y estado civil*

Sin que pueda considerarse en ningún momento que las dos nociones son sinónimas, es posible suponer que existe una vinculación bastante estrecha entre la soledad y la soltería, y que aumentarán o disminuirán una y otra en el mismo sentido. En efecto, si calculamos la coordenación entre el porciento de solteros dentro del total de la población de cada entidad y el porciento de personas solas en la misma, se obtiene un índice de coordenación de 0.474 que también podemos considerar como bastante elevado en relación con los obtenidos anteriormente. En cuanto el índice es positivo, puede afirmarse que hay una relación directa entre la tasa de soltería y la tasa de soledad de las entidades de la República Mexicana. Sin embargo, también es fácil apreciar que no es en ninguna forma la soltería la que da cuenta en forma total y plenaria de la soledad (como ocurriría en caso de que el índice correspondiente hubiese sido de 1). Debe considerarse, en efecto, que no todos los solteros viven solos, pues algunos incluso superada su minoridad, continúan viviendo en su familia de origen, mientras otros viven en comunidades religiosas de célibes y otros más viven con amigos también solteros o casados con otras personas de las que estos últimos viven separados o con las que continúan viviendo. Por otra parte, no todas las personas solas son solteras, pues a más de que en una exploración menos superficial y apresurada que ésta habría que considerar las proporciones de viudos y divorciados de la población para coordenar las entidades de acuerdo con estos criterios y con la tasa de soledad, sería necesario tomar en consideración que muchas personas, legalmente casadas, o casadas mediante cualquier rito religioso, viven separadas de sus cónyuges. Una clasificación cruzada de "solos y convivientes" y de "no casados y casados" sería un mínimo que resultaría deseable obtener de las publicaciones estadísticas mexicanas para la prosecución de estudios de este tipo. Un tercer criterio clasificatorio que tratase de definir la voluntariedad o falta de voluntariedad del aislamiento nos colocaría, probablemente, en el umbral mismo de uno de los estudios sociológicos de máximo interés.

3.5 *Soledad y autoctonía*

El examen de la coordenación de las entidades de acuerdo con la proporción de solos y la proporción de habitantes que son nativos de la entidad parece confirmar una expectativa fácil de formular, de acuerdo con la cual, al aumentar la proporción de habitantes de una entidad que sean oriundos de la misma disminuirá la tasa de solos dentro de la población, mientras que al disminuir la proporción de autóctonos aumentará la tasa de solos. El valor del índice es 0.475 con signo negativo. No hay que decir que una mayor finura en las tabulaciones ofrecidas por nuestras publicaciones estadísticas podría contribuir a precisar estas relaciones. La afinación en las tabulaciones podría consistir, en el caso, en mostrar el número de años que los no autóctonos tienen de establecidos en la entidad, puesto que, para el problema que se estudiará con base en tales datos —y al que damos breve introducción con nuestras líneas—, una permanencia muy prolongada de un no autóctono en la entidad puede llegar a ser equivalente de la misma autoctonía, en cuanto dicha permanencia podría haberle permitido crear vínculos parecidos a los de los autóctonos y distintos de quienes no han tenido oportunidad de permanecer en la entidad sino por breve tiempo.

Los tres índices últimos a que nos referimos, presentan —según creemos— máximo interés, en función de las consideraciones que pueden hacerse al tratar de interpretarlos más conjunta que aisladamente. Nos referimos a los índices de coordenación entre las proporciones de solteros y las de urbanícolas, por una parte; la proporción de solteros y la de analfabetos, por otra, y la de solteros y económicamente activos, finalmente.

3.6 *Soledad y urbanización*

Es relativamente fácil postular la hipótesis de que la tasa de solos en la población aumentará al aumentar el urbanismo de la misma. La hipótesis se ve confirmada por el índice ro de coordenación entre las entidades. Sin embargo, contra lo que hubiéramos podido esperar, la correlación entre soledad y urbanismo no es muy alta, pues el índice apenas si alcanza la magnitud de 0.280. Quizás si la coordenación se establece tomando en cuenta límites distintos (superiores) para la distinción entre la población urbana y no urbana de una entidad, que los que utiliza el censo, el índice de coordenación correspondiente aumente. Una exploración del uso de diferentes niveles delimitativos entre lo urbano y lo no urbano y de su repercusión sobre la magnitud del índice de coordenación entre urbanización y soledad podría descubrir cuál es el punto crítico en el que las poblaciones mexicanas del medio siglo han tendido a favorecer al máximo el aislamiento (ya positivo o ya negativo) de sus individuos. De todos modos, queda el hecho de que el urbanismo sí parece favorecer el aislamiento.

3.7 *Soledad y actividad económica*

Podría pensarse que con el incremento de la actividad económica de la población disminuiría la tasa de solos dentro de la misma; sin embargo, esto no parecen confirmarlo los cálculos puesto que el índice de coordenación es positivo y *superior al 50%*. Según esto, puede afirmarse, a título tentativo —como todo lo anterior, pero con menos inseguridad que en casos previos— que, conforme aumenta la población económicamente activa dentro de la población total, la tasa de soledad aumenta asimismo. ¿Querrá decir esto que la actividad económica, en vez de contribuir a intensificar las relaciones sociales, dentro de las actuales condiciones de la sociedad mexicana, tiende a debilitarlas? De momento no puede responderse, pero este solo hecho abre una interrogante a la que convendría buscar respuesta. Frente a las posibles respuestas no convendría ser ni apresurado ni pesimista. La investigación estadística no alcanza ni de lejos el terreno de las significaciones y, en el caso, no habría justificación para concluir que si el ingreso de un número creciente de pobladores al ejército laboral se conecta con un incremento en la proporción de personas solas, ha de reputarse como socialmente negativo dicho ingreso. En efecto, la soledad tiene un aspecto negativo (aislamiento obligado), pero tiene también un aspecto positivo (aislamiento voluntariamente buscado). En este sentido, en el mismo grado en que el ingreso de mayor número de individuos al ejército laboral favorezca el aislamiento obligado, dicho ingreso será negativo (y habrá que pensar en modificar los sistemas laborales correspondientes a fin de evitar que los mismos obliguen al individuo a aislarse en vez de permitirle intervenir en relaciones humanas creadoras); pero, en el mismo grado en que dicho ingreso de un número creciente de individuos a las filas laborales favorezca el aislamiento voluntariamente buscado, habrá que considerar a dicho ingreso como socialmente positivo, en cuanto, al aumentar la proporción de individuos integrantes de la población económicamente activa, aumenta el número de quienes están en aptitud de liberarse de una serie de dependencias o cadenas.

3.8 *Soledad y alfabetismo*

Si examinamos la coordenación entre soledad y analfabetismo, en la que podríamos esperar que al aumentar el analfabetismo aumentaría la soledad, nos encontramos con el hecho sorprendente de que el índice de coordenación es de 0.425 con signo negativo, o sea, que hay una coordenación relativamente considerable e inversa entre ambos fenómenos, o bien que, conforme aumenta el analfabetismo disminuye la proporción de solos, o, a la inversa, que conforme disminuye el analfabetismo, aumenta la proporción de personas solas en la población. El hecho queda también, por supuesto, abierto a la investigación.

Creemos que incluso una exploración tan superficial como ésta puede mostrar el enorme interés que tendría tanto para la ciencia sociológica como para la problemática y para la política social, la investigación del problema de la soledad, iniciada sobre bases estadístico-sociales.