



ANÁLISIS CUANTITATIVO DE DATOS CUALITATIVOS

Forrest W. Young

Traducción: Elkin Castaño V.

Abstract. This article presents a general vision of an approach to the quantitative analysis of qualitative data with theoretical and methodological explanations of basic principles of the approach, the Alternating Square Minimums and the Optimal Quantification. Using these two principles, a variety of procedures originally proposed for the quantitative data analysis (of interval or reason) to the qualitative data analysis has extended (nominal or ordinal), including the analysis of additive and analysis of variance; multiple and canonical regression, component main; common factor; and analysis of multidimensional scaling.

Resumen. Este artículo presenta una visión general de una aproximación al análisis cuantitativo de datos cualitativos con explicaciones teóricas y metodológicas de los dos principios básicos de la aproximación, los Mínimos Cuadrados Alternantes y la Cuantificación Óptima. Usando estos dos principios, se ha extendido una variedad de procedimientos originalmente propuestos para el análisis de datos cuantitativos (de intervalo o de razón) al análisis de datos cualitativos (nominal u ordinal), incluyendo el análisis de aditividad y análisis de varianza; regresión múltiple y canónica, componentes principales; factor común; y análisis de escalamiento multidimensional.

Presentación.

Este artículo presenta una visión general de una aproximación al análisis cuantitativo de datos cualitativos con explicaciones teóricas y metodológicas de los dos principios básicos de la aproximación, los Mínimos Cuadrados Alternantes y la Cuantificación Óptima. Usando estos dos principios, mis colegas y yo hemos extendido una variedad de procedimientos originalmente propuestos para el análisis de datos cuantitativos (de intervalo o de razón) al análisis de datos cualitativos (nominal u ordinal), incluyendo el análisis de aditividad y análisis de varianza; regresión múltiple y canónica, componentes principales; factor común; y análisis de escalamiento multidimensional. La aproximación tiene dos ventajas: (a) Si se conoce un procedimiento de mínimos cuadrados para analizar datos cuantitativos, este puede ser extendido a datos cualitativos; y (b) el algoritmo resultante será convergente. Se presentan tres ejemplos completamente trabajados del procedimiento de análisis de aditividad y las etapas concernientes a los procedimientos de regresión.

Palabras claves: Análisis exploratorio de datos, análisis descriptivo de datos, análisis multivariado de datos, análisis no métrico de datos mínimos cuadrados alternantes, cuantificación, teoría de datos.

Introducción

Quizás uno de los principales impedimentos del progreso rápido en el desarrollo de las ciencias sociales, de comportamiento y biológicas es la presencia constante de datos cualitativos. Muy frecuentemente es imposible obtener datos numéricos; el investigador tiene que elegir entre usar datos cualitativos o no usar nada. Muchas veces solamente es posible determinar la categoría en la cual el dato cae. Por ejemplo, la socióloga, obtiene información categórica acerca de la afiliación religiosa de sus encuestados; el botánico obtiene información categórica sobre la familia a la que pertenecen sus

plantas; la sicóloga obtiene información categórica sobre el tipo de sicosis de su paciente. Aún en el mejor de los casos, a menudo es imposible obtener más que el orden en el cual caen las categorías. Cuando nuestra socióloga observa la educación de los encuestados en su muestra, sabe que las categorías de observación están ordenadas, pero no es capaz de asignar un valor numérico preciso a las categorías. Cuando la sicóloga obtiene juicios de escala clasificados, los juicios pueden ser razonablemente considerados como ordinales, pero no siempre como numéricos.

Dada la omnipresencia de los datos cualitativos, podemos entender el largo y persistente interés en su cuantificación. Si de algún modo pudiéramos desarrollar un método para asignar “buenos” valores numéricos a las categorías de los datos, entonces los datos serían cuantificados y serían susceptibles de un análisis más significativo. Curiosidad sobre el tema aparece en el trabajo clásico de Yule (1910), y los primeros métodos para cuantificar empiezan a aparecer alrededor de 1940. Probablemente, el primer procedimiento ampliamente diseminado fue el método del “puntaje apropiado” (Fisher, 1938, pp.285-298), el cual fue introducido aproximadamente al mismo tiempo que el método propuesto por Guttman (1941). Varios autores trabajaron en el problema en los 50 (Burt, 1950, 1953; Hayashi, 1950; Guttman, 1953), y su trabajo fue resumido por Torgerson (1958, pp. 338-345). Mucho trabajo se ha hecho recientemente (Benzicri, 1973, 1977; de Leeuw, 1973; Mardia, Kent y Bibby, 1979; Saito, 1973; Saporta, 1975; Tenenhaus, Nota 1 y Nota 2).

En este artículo nos referimos al proceso de cuantificar datos cualitativos como “Cuantificación Óptima”, término que fue introducido primero por Bock (1960). Esta es nuestra definición:

Cuantificación óptima es una técnica de análisis de datos la cual asigna valores numéricos a las categorías de observación en una forma tal que maximicen la relación entre las observaciones y el modelo de análisis de los datos, mientras respeta el carácter de medición de los datos.

Observar que ésta es una definición muy general: No hay una especificación precisa de la naturaleza del modelo, ni una especificación precisa del carácter de medición de los datos. Trabajando con esta definición de cuantificación óptima, hemos desarrollado un grupo de programas para cuantificar datos cualitativos (vea Tabla 1). Los programas permiten que los datos tengan una variedad de características de medición, y permiten el análisis de los datos con una variedad de modelos. A estos programas los llamamos los programas ALSOS puesto que usan la aproximación de los Mínimos Cuadrados Alternantes (**A**lternating **L**east **S**quares) a la Cuantificación Óptima (**O**ptimal **S**caling).

Los programas ALSOS describen los datos cualitativos por medio de modelos cuantitativos dentro de tres clases generales: (a) el Modelo Lineal General; (b) El Modelo de Componentes (Factores); y (c) el Modelo Euclideo General. Como se puede ver en la Tabla 1, los programas GLM están específicamente orientados hacia el análisis de varianza (MANOVALS, ADDALS WADDALS), análisis de regresión (MORALS, CORALS, OVERALS) y análisis de rutas (PATHALS). Los programas de componentes realizan análisis de componentes principales (PRINCIPALS y HOMALS); análisis de componentes (factores) de tres modos (ALSCOMP y TUCKALS); y Análisis de Factores Comunes (FACTALS). El Modelo Euclideo General está conformado por ALSCAL y GEMSCAL.

Para la mayoría de los programas ALSOS los datos pueden estar definidos en niveles de medición binaria, nominal, ordinal o de intervalo (y el nivel de razón con los programas del Modelo Euclideo General), y puede suponerse que han sido generados ya sea por un proceso discreto o uno continuo. Todos estos programas también permiten patrones arbitrarios de datos “missing”. Algunos permiten restricciones de rango o de cotas sobre los valores asignados a las categorías de observación, y algunos permiten el empleo de órdenes parciales con datos ordinales. La información sobre la obtención de estos programas se encuentra en la Tabla 1.

Como mostraremos en este artículo, la aproximación ALSOS a la construcción de los algoritmos tiene una implicación muy importante para el análisis de datos:

Si se conoce que un procedimiento de mínimos cuadrados produce la descripción de datos numéricos (de nivel de medición en intervalo o de razón), entonces un algoritmo ALSOS puede ser construido para obtener una descripción de mínimos cuadrados de datos cualitativos (los cuales tienen una gran variedad de características de medición).

1. Visión general

Cada uno de los programas ALSOS optimiza una función objetivo usando un algoritmo basado en los principios de Mínimos Cuadrados Alternantes (ALS) y Cuantificación Óptima (OS).

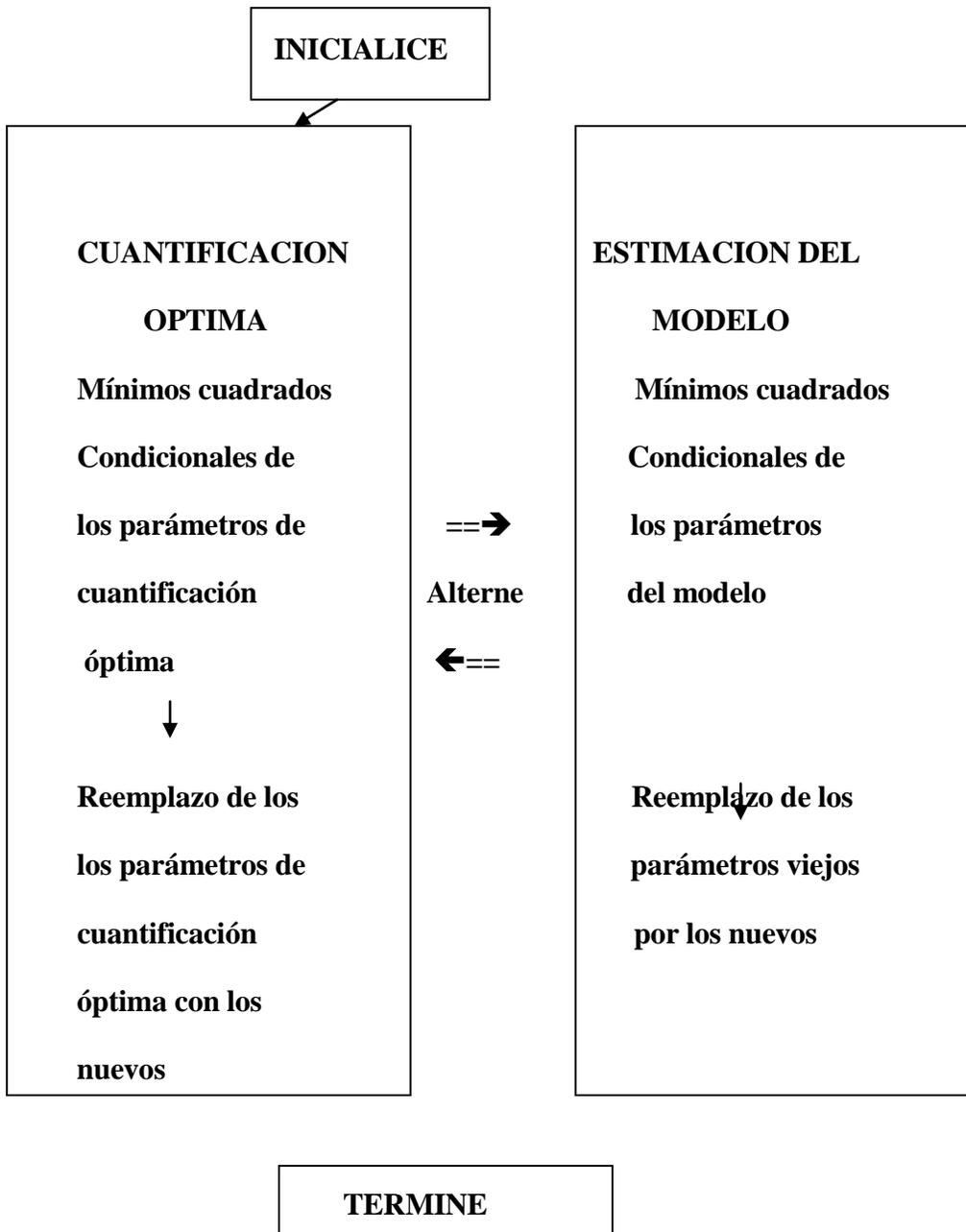
El principio OS implica considerar las observaciones como categóricas, y luego representar la observación por medio de un parámetro. Este parámetro está sujeto a las restricciones impuestas por las características de medición de la variable (por ejemplo, restricciones de orden para las variables ordinales).

El principio ALS implica dividir todos los parámetros en dos subconjuntos mutuamente excluyentes y exhaustivos: (a) los parámetros del modelo; y (b) los parámetros de los datos (llamados parámetros de cuantificación óptima). A continuación se procede a optimizar una función de pérdida optimizando alternativamente con respecto a un subconjunto, y luego al otro (vea la Figura 1). Observe que cada subconjunto mismo puede contener varios subconjuntos los cuales son mutuamente excluyentes y exhaustivos. Por ejemplo, en ALSCAL el modelo tiene varios subconjuntos de parámetros, y en los programas multivariados hay un subconjuntos de parámetros para cada variable.

La optimización prosigue consiguiendo las estimaciones de mínimos cuadrados de los parámetros en un subconjunto mientras asume que los parámetros en todos los demás subconjuntos son constantes. Llamaremos a estas las estimaciones de mínimos cuadrados condicionales, puesto que la naturaleza de los mínimos cuadrados es condicional a los valores de los otros parámetros en los otros subconjuntos. Una vez hemos obtenido las estimaciones de mínimos cuadrados condicionales inmediatamente reemplazamos las estimaciones viejas de estos parámetros por las nuevas. Luego vamos a otro subconjunto de parámetros y obtenemos sus estimaciones de mínimos cuadrados condicionales. Alternativamente obtenemos las estimaciones de mínimos cuadrados condicionales de los parámetros en los subconjuntos del modelo, y luego en los subconjuntos de los datos, hasta que haya convergencia (la cual se asegura bajo ciertas condiciones que serán discutidas posteriormente).

El flujo de un procedimiento ALSOS está diagramado en la Figura 1.

Figura 1. Flujo de los Algoritmos ALSOS



Existen ciertas correspondencias fuertes entre el procedimiento ALSOS y la aproximación NILES para construcción de algoritmos desarrollada por Wold y Lyttkens (1969), y la clase de algoritmos de análisis numérico conocidos como algoritmos de bloques sucesivos (Hageman y Porsching, 1975). La principal diferencia entre estos algoritmos métricos y el algoritmo no métrico ALSOS son las características de cuantificación óptima del algoritmo ALSOS. La característica de cuantificación permite el análisis de datos cualitativos, mientras que los procedimientos mencionados sólo permiten analizar datos cuantitativos.

Hay otras conexiones fuertes entre los algoritmos no métricos desarrollados por Kruskal (1964, 1965), Roskam (1968), Young (1972), y otros. La principal diferencia entre estos procedimientos de gradiente (no ALSOS) y los algoritmos ALSOS es la característica de mínimos cuadrados de la etapa de estimación del modelo.

Una de las principales ventajas de combinar los principios ALS y OS es que la etapa OS de un algoritmo ALSOS no necesita conocer el tipo de modelo involucrado en el análisis. Una ventaja paralela e igualmente importante es la de que la fase de estimación del modelo tampoco necesita saber nada sobre las características de medición de los datos.

El efecto práctico de estos aspectos de los procedimientos ALSOS es enorme: Si existe un procedimiento de mínimos cuadrados para ajustar un modelo particular a datos numéricos (de intervalo o de razón), entonces podemos usar ese procedimiento junto con los procedimientos OS que serán discutidos, para desarrollar un algoritmo ALSOS que permita ajustar el modelo a datos cualitativos. Eso es todo lo que hay que hacer! Si podemos obtener una descripción de mínimos cuadrados de datos numéricos entonces podemos obtener una descripción de mínimos cuadrados de datos cualitativos. Todo lo

que tenemos que hacer es alternar el procedimiento de mínimos cuadrados numérico con el procedimiento OS adecuado a las características de medición de los datos que están siendo analizados.

Hay una desventaja: El procedimiento ALSOS no garantiza convergencia sobre la solución de mínimos cuadrados global, en su lugar garantiza convergencia sobre un tipo particular de solución local de mínimos cuadrados. El óptimo local particular sobre el cual el proceso ALSOS converge está determinado solamente por una cosa, el proceso de inicialización. Es posible que para dos tipos diferentes de inicialización obtengamos que el procedimiento ALSOS nos lleve a dos óptimos locales, quizás con resultados radicalmente diferentes. Por esta razón, y puesto que cada etapa de un procedimiento ALSOS es una solución de mínimos cuadrados condicionales (condicional sobre los valores actuales de los parámetros en el otro subconjunto), nos referimos al punto de convergencia de un procedimiento ALSOS como el “óptimo global condicional”, una manera espectacular de señalar que el punto de convergencia es más que simplemente un óptimo local, pero que puede no ser el óptimo global. (las propiedades de convergencia de un algoritmo ALSOS han sido discutidas por de Leeuw, Young y Takane (1976), y de Leeuw (Nota 3), quienes probaron que tal procedimiento es convergente si: (a) La función que está siendo optimizada es continua; y (b) si cada etapa o subetapa del algoritmo optimiza la función).

Puesto que el procedimiento de inicialización es de tal importancia en el proceso completo, es importante usar la mejor inicialización que se encuentre disponible. En todos los programas ALSOS definimos la mejor inicialización para dar a entender que deberíamos optimizar el ajuste del modelo a los datos originales. De este modo, cada programa ALSOS se inicializa aplicando un procedimiento de mínimos cuadrados ordinarios a los datos originales bajo el supuesto que los datos originales son

cuantitativos, *en la forma que el usuario los ha codificado*. (Observe que una codificación diferente de los datos, mientras que sea consistente con las características de medición de los datos, puede proporcionar un mejor inicio. La inicialización no es “mejor” en este sentido. Hay evidencia que para ALSCAL, este procedimiento reduce la frecuencia de soluciones de mínimos locales (Young y Null, 1978)).

Una vez que el proceso se ha iniciado, el procedimiento para obtener las estimaciones de mínimos cuadrados condicionales de los parámetros del modelo es el procedimiento usado para obtener estimaciones de mínimos cuadrados cuando los datos son numéricos. La *única* diferencia es que procedimiento se aplica a los datos óptimamente cuantificados (que después de todo son numéricos) en vez de aplicarse a los datos originales. Puesto que estamos aplicando el procedimiento de estimación a los datos óptimamente cuantificados, no estamos violando los supuestos de medición de los datos originales, cualesquiera que ellas sean. No estamos aún ni usando los datos originales en la fase de estimación del modelo, así que no necesitamos conocer sus características de medición. Igualmente importante, no tenemos que pensar en una nueva manera de ajustar el modelo a los datos cualitativos, simplemente usamos los procedimientos existentes para ajustarlo a los datos cualitativos.

2. Cuantificación Óptima (OS)

Puesto que las características ALS de nuestro trabajo son por ahora bastante tradicionales (Wold y Lyttkens, 1969), no gastaremos esfuerzos en explicarlas. En su lugar, discutiremos ampliamente los aspectos OS en el resto de este capítulo.

Estos aspectos únicos del OS de nuestro trabajo permiten que nuestros algoritmos ALSOS sean muy flexibles en cuanto a los supuestos que el usuario puede hacer sobre las características de medición de los datos.

2.1 Teoría de medición

Para apreciar completamente las características OS de nuestro trabajo primero discutiremos los fundamentos teóricos del proyecto, nuestro punto de vista de la teoría de medición.

Empezamos enfatizando un concepto el cual es crucial en nuestro trabajo: desde nuestro punto de vista todas las observaciones son categóricas. Es decir, consideramos las observaciones de una variable como conjuntos de observaciones que caen en una variedad de categorías, de forma tal que todas las observaciones en una misma categoría particular son empíricamente equivalentes. Además, tomamos esta definición de “categórica” sin tener en cuenta las características de medición de la variable.

Para decirlo de manera más simple, consideramos que el proceso de observación produce datos que son categóricos debido a la precisión finita del proceso de medición y de observación, si no es por otras razones. Por ejemplo: si estamos midiendo la temperatura en un termómetro ordinario (el cual probablemente generará observaciones de nivel de intervalo que reflejan razonablemente un proceso continuo) las mediciones son dudosas si los grados son reportados con más precisión que grados enteros. De ésta manera la observación es categórica.; hay un número muy grande de temperaturas (realmente infinitas) las cuales podrían ser reportadas como, por ejemplo, 40 grados. Por tanto, decimos que la observación 40 grados es categórica.

En este punto necesitamos definir un vector columna con las n observaciones originales. Denotamos este vector observación como \mathbf{o} con elementos o_i . (Letras minúsculas en negrilla hacen referencia a vectores columna, y letras minúsculas itálicas a escalares.) También definimos un vector \hat{z} de estimaciones del modelo, cuyo elemento general es \hat{z}_i y un vector \mathbf{z}^* con las observaciones óptimamente cuantificadas \mathbf{z}^* cuyo elemento general es z_i^* .

Los elementos de \mathbf{o} están organizados de forma tal que las observaciones que pertenecen a una categoría en particular están contiguas. Los elementos de $\hat{\mathbf{z}}$ y \mathbf{z}^* están organizados de modo que están en correspondencia uno a uno con los elementos de \mathbf{o} . El elemento z^*_i es el parámetro que representa la observación o_i . El vector $\hat{\mathbf{z}}$ es llamado “estimaciones del modelo” debido a que contiene las estimaciones del modelo, en el sentido de mínimos cuadrados, de los datos óptimamente cuantificados \mathbf{z}^* .

Con estas definiciones podemos representar formalmente el problema OS como un problema de transformación, como sigue. Queremos obtener una transformación t de las observaciones originales la cual genera las observaciones óptimamente cuantificadas,

$$t[\mathbf{o}] = [\mathbf{z}^*] \quad (1)$$

donde la definición precisa de t es una función de las características de medición de las observaciones, y es tal que existirá una relación de mínimos cuadrados entre las estimaciones del modelo de los datos cuantificados ($\hat{\mathbf{z}}$) y los datos reales cuantificados (\mathbf{z}^*), dado que las características de medición de \mathbf{o} se mantienen estrictamente. El valor numérico asignado a z^*_i es el valor óptimo del parámetro para la observación o_i .

Varias clases de restricciones son colocadas a las transformaciones t , donde el tipo de restricción depende de las características de medición de los datos. Distinguimos tres tipos restricciones de medición, denominados *nivel de medición*, *proceso de medición* y *condicionalidad de medición*. Como veremos, estos tienen que ver con tres diferentes aspectos de las categorías de observación. El proceso de medición tiene que ver con las relaciones entre todas las observaciones *dentro* de una sola categoría; el nivel de medición: corresponde a las relaciones de todas las observaciones *entre* las diferentes categorías; y condicionalidad de la medición: hace referencia a las relaciones *dentro de*

los conjuntos de categorías. Cada uno de los distintos tipos de procesos, niveles y condicionalidades implica un diferente conjunto de restricciones colocadas sobre la transformación t .

En las Tablas 2, 3 y 4 resumimos los seis tipos de medición resultantes de combinar tres niveles con dos procesos. La Tabla 2 presenta la descripción verbal, las restricciones matemáticas sobre t se encuentran en la Tabla 3, y los métodos de cuantificación óptima están dados en la Tabla 4. La condicionalidad de medición se discutirá al final de ésta sección.

Tabla 2. Características de medición para 6 tipos de mediciones

| Nivel | Proceso | |
|----------|---|---|
| | Discreto | Continuo |
| Nominal | Categorías de observación Representadas por un solo Número real. | Categorías de observación representadas por un Intervalo cerrado de Números reales |
| Ordinal | Categorías de observación Están ordenadas y las Observaciones empatadas Permanecen empatadas | Categorías de observación Están ordenadas pero las Observaciones empatadas llegan a desempatarse |
| Numérico | Categorías de observación Están relacionadas | Categorías de observación Están relacionadas |

| | |
|------------------------|-------------------------|
| Funcionalmente y todas | Funcionalmente pero |
| Las observaciones son | todas las observaciones |
| Precisas | son imprecisas |

Tabla 3. Restricciones matemáticas para los 6 tipos de mediciones

| Nivel | Proceso | |
|----------|--|---|
| | Discreto | Continuo |
| Nominal | $t^d: (o_i \sim o_m) \rightarrow (z^*_i = z^*_m)$ | $t^c: (o_i \sim o_m) \rightarrow (z^-_i = z^-_m) \leq \{z^*_i, z^*_m\} \leq (z^+_i = z^+_m)$ |
| Ordinal | $t^{do}: (o_i \sim o_m) \rightarrow (z^*_i = z^*_m)$ $(o_i < o_m) \rightarrow (z^*_i \leq z^*_m)$ | $t^{co}: (o_i \sim o_m) \rightarrow (z^-_i = z^-_m) \leq \{z^*_i, z^*_m\} \leq (z^+_i = z^+_m)$ $(o_i < o_m) \rightarrow (z^*_i \leq z^*_m)$ |
| Numérico | $t^{dp}: (o_i \sim o_m) \rightarrow (z^*_i = z^*_m)$ $z^*_i = \sum_{q=0}^p \delta_q o_i^q$ | $t^{cp}: (o_i \sim o_m) \rightarrow (z^-_i = z^-_m) \leq \{z^*_i, z^*_m\} \leq (z^+_i = z^+_m)$ $z^*_i = \sum_{q=0}^p \delta_q o_i^q$ |

Tabla 4. Métodos de cuantificación óptima para los 6 tipos de mediciones

| Nivel | Proceso | |
|-------|----------|----------|
| | Discreto | Continuo |
| | | |

| | | |
|----------|--|---|
| Nominal | Medias de elementos del | Medias de elementos del |
| Modelo | Modelo, seguidas por la transformación monótona primaria | |
| Ordinal | Transformación monótona Secundaria de Kruskal | Transformación monótona primaria de Kruskal |
| Numérica | Regresión simple lineal O no lineal | Regresión simple lineal o no lineal seguida por Estimación acotada. |

Proceso de medición.

Hay dos tipos de restricciones en el proceso de medición según sea el proceso generatriz: discreto o continuo. Una u otra hipótesis siempre debe ser hecha. Si creemos que es *discreto* (por ejemplo, el sexo), entonces todas las observaciones en una categoría particular (mujer u hombre) deberían estar representadas por el mismo número real después de la transformación t^d (el super índice indica que es discreta) ha sido hecha. Por otro lado, si adoptamos el supuesto de continuidad (como sería, por ejemplo, para una variable de peso), entonces cada una de las observaciones dentro de una categoría particular (97.2 Kg, por ejemplo), debería estar representada por un número real seleccionado desde un intervalo cerrado de números reales. En el primer caso la naturaleza discreta del proceso se refleja por el hecho de que escogemos un sólo (discreto) número para representar todas las observaciones en esa categoría; mientras que el último caso la continuidad del proceso se refleja por el hecho de que escogemos números reales desde un intervalo (continuo) cerrado de números reales.

Formalmente, definimos las dos restricciones de la siguiente manera: La restricción discreta es

$$t^d: (o_i \sim o_m) \rightarrow (z^*_i = z^*_m) \quad (2)$$

donde \sim indica equivalencia empírica (es decir, están en la misma categoría). La restricción continua es representada como:

$$t^c: (o_i \sim o_m) \rightarrow (z^-_i = z^-_m) \leq \{z^*_i, z^*_m\} \leq (z^+_i = z^+_m) \quad (3)$$

donde z^-_i y z^+_i son los límites inferior y superior del intervalo de números reales. Observe que una de las implicaciones de la equivalencia empírica (categórica) es la de que las cotas superior e inferior de todas las observaciones de una categoría particular son las mismas. De esta manera podemos pensar mejor que las cotas se aplican a las categorías en lugar de las observaciones, pero denotar esto exigiría un sistema notacional más complicado. Observe además que para todas las observaciones que caigan en una categoría particular, la correspondientes observaciones cuantificadas óptimamente caerán en el intervalo pero no necesariamente serán iguales.

Nivel de medición. Volvamos ahora al segundo conjunto de restricciones sobre las distintas transformaciones de medición, las restricciones de nivel. Con estas restricciones determinamos la naturaleza de las transformaciones t permitidas, de manera que correspondan al nivel de medición supuesto de las variables observadas. Por supuesto hay varias restricciones que podrían ser de interés, pero aquí solamente mencionamos tres. Con ellas podemos satisfacer las características de los cuatro

niveles de medición de Stevens, como también las características del nivel de medición de los datos “missing”, y de los datos binarios.

Para el nivel *nominal* de medición, y para los datos “missing” o binarios, no se introducen restricciones de nivel de medición. Las características de estos tres tipos de mediciones son completamente especificadas por las restricciones del proceso de medición. La razón por la cual no necesitamos restricciones adicionales es que para estos niveles solamente conocemos la categoría de la observación. No conocemos nada sobre la relación de las observaciones en las diferentes categorías. De ésta forma, estos niveles están completamente especificados por las restricciones impuestas dentro de las categorías de observación, no habiendo restricciones sobre las relaciones que pueden existir entre observaciones de las diferentes categorías.

La diferencia entre los datos nominales, “missing” y binarios está en el número de categorías de observación. Para datos nominales debe haber al menos tres categorías. Cuando hay solamente dos categorías, los datos son binarios (Los datos binarios son algo anómalos puesto que podemos pensar que están en cualquier nivel de medición. Sin embargo, su descripción más parsimoniosa corresponde al nivel nominal, puesto que todos los niveles más altos de medición implican restricciones entre-categorías adicionales).

Los datos “*missing*”, por otro lado, pueden ser considerados como un tipo particular de datos sobre los cuales solamente sabemos una cosa: que está perdido. De esta forma, podemos considerar los datos “missing” como nominales, pero con una sola categoría de (no)observación, la categoría “missing”. Parecería que podríamos simplemente llamar datos “missing” a una categoría adicional a nuestros datos nominales. Sin embargo, esto no es suficiente cuando existen datos “missing” en un conjunto de observaciones definidas en algún nivel de medición diferente al nominal. Cuando introduzcamos el concepto de condicionalidad al final de ésta sección, seremos capaces de clarificar completamente la forma en la cual consideramos los datos “missing”. Hasta entonces debemos contentarnos simplemente con considerar que los

datos “missing” están definidos en el nivel nominal, y que sólo tienen una categoría de observación.

Debería mencionarse que los datos que sólo contienen una categoría de observación son, lógicamente, equivalentes a los datos “missing” en sus características de nivel de medición. Es decir, no tienen nivel de medición. Esto es cierto, independientemente del supuesto nivel de medición de los datos. De este modo, para definir el nivel de medición de un conjunto de observaciones, el número mínimo de categorías de observación es dos, y debe haber al menos tres para cualquier nivel de medición arriba del nivel nominal.

Para el nivel de medición nominal, y para datos binarios y “missing”, no hay restricciones de nivel: las características de estos datos están completamente especificadas por las restricciones del proceso de medición. Puesto que hay dos tipos de procesos, existirán dos tipos de datos nominales, binarios y “missing”. El nivel discreto nominal es bastante común, siendo, el sexo de una persona una de tales variables. Es claro que es una variable nominal (binaria), y que es razonable suponer que las dos categorías de observación (hombre y mujer) son generadas por un proceso discreto. Un ejemplo de una variable continua nominal es la de los colores en palabras. Las diferentes categorías de observación pueden ser azul, rojo, amarillo, verde, etc., las cuales a pesar de que son nominales, realmente representan un proceso continuo subyacente (la longitud de onda). Incluso los datos “missing” vienen en dos tipos: Los datos “missing” discretos implican que el observador cree que todas las observaciones deberían haber sido idénticas si ellas hubieran sido observadas; mientras datos “missing” continuos, implican que no necesariamente habrían sido idénticas. Al final de la sección hablaremos más sobre esto.

Para el nivel *ordinal* de medición, requerimos que, adicional a las restricciones del proceso, los números reales asignados a las diferentes categorías representen el orden de las observaciones empíricas:

$$f^p: (o_i \prec o_m) \rightarrow (z^*_i \leq z^*_m) \quad (4)$$

donde el super índice en f^p indica restricción de orden y donde \prec indica orden empírico. El problema de qué hacer con los empates ya ha sido manejado por la noción de proceso. Si la variable es discreta ordinal (f^{do}), entonces las observaciones empatadas permanecen empatadas después de la transformación, mientras que para una variable continua ordinal (f^{co}), las observaciones empatadas pueden ser desempatadas después de la transformación. El caso discreto ordinal queda bien ilustrado por los datos obtenidos de personas quienes ordenan $n-1$ términos afines, de acuerdo a su similaridad con el n -ésimo término. Una variable continua ordinal podría ser el nivel de ingreso del padre, como usualmente se obtiene en una encuesta. Las categorías de observación podrían ser “menos de 50000”, “5000-10000”, “10000-20000” y “más de 20000”, y podemos imaginar el proceso continuo por medio del cual dichas categorías son generadas.

Para variables *numéricas* (de intervalo o de razón), necesitamos que los números reales asignados a las observaciones estén funcionalmente relacionados con las observaciones. Por ejemplo (otros ejemplos se pueden construir fácilmente), podríamos requerir que las observaciones óptimamente cuantificadas y las originales estén relacionadas por alguna regla polinomial:

$$f^p: z^*_i = \sum_{q=0}^p \delta_q o_i^q \quad (5)$$

Por ejemplo, si $p=2$, tenemos una relación cuadrática entre las observaciones cuantificadas óptimamente y las originales. Cuando $p=1$ obtenemos la conocida

relación lineal usada con variables de intervalo (y con variables de razón cuando $\delta_0=0$) Es importante observar que con variables numéricas el papel jugado por la distinción discreta-continua es el de precisión de la medición. Si pensamos que las observaciones son perfectamente precisas, entonces queremos que todas las observaciones estuvieran exactamente relacionadas con las observaciones óptimamente cuantificadas por medio de la función especificada por (5). Sin embargo, si creemos que hay alguna falta de precisión en la situación de medición, quisiéramos permitir que las observaciones óptimamente cuantificadas se “tambaleen” un poco alrededor de la función especificada en (5). El primer caso corresponde al caso discreto-intervalo o discreto-razón, en el cual no permitimos variación dentro de la categoría de observación, y el último caso corresponde al caso continuo-intervalo y continuo-razón donde permitimos alguna variación dentro de la categoría. Observe que esta noción es razonable aún cuando haya una sola observación dentro de una categoría de observación particular, como es usualmente el caso.

Condicionalidad de Medición. El último tipo de restricciones colocadas sobre las transformaciones de medición t son llamadas restricciones de condicionalidad. Estas restricciones operan sobre las relaciones que pueden existir entre las observaciones *dentro* de los conjuntos de categorías de observación. Puede ocurrir que, para un conjunto de datos en particular, las características de medición de las observaciones sean condicionales sobre algunos aspectos de la situación empírica (Coombs, 1964). Cuando este es el caso, algunas de las observaciones no pueden compararse significativamente con las otras observaciones. Por tanto, deberíamos subdividir todas las observaciones en grupos tales que aquellas observaciones dentro de un grupo sean aquellas que pueden ser comparadas significativamente con cada una de las otras. Luego debemos restringir el nivel de medición y las transformaciones de los procesos de modo que solamente apliquen a las observaciones dentro de un grupo, no entre grupos. Tales grupos son llamados *particiones*, debido a que particionan los datos en

subconjuntos, y las restricciones dadas por (2) hasta (5) son redefinidas de manera tal que ellas sólo hagan cumplir las relaciones que deseamos que operen entre las observaciones dentro de una partición. No hay restricciones impuestas sobre las relaciones que puedan existir entre observaciones que están en diferentes particiones.

Se pueden distinguir varios tipos de condicionalidad, las cuales son relevantes según sea la clase de análisis de datos que se realice. No vamos a entrar en detalle sobre ellos, sino que mencionaremos sólo dos.

Un tipo, la *matriz condicional*, se encuentra en el siguiente ejemplo. Si hemos preguntado a varios sujetos que juzguen la similaridad de todos los pares de un conjunto de estímulos, generalmente no estamos interesados en comparar las respuestas de un sujeto con las de otro. Es decir, la respuesta 7 de un sujeto (sobre, por ejemplo, una escala de similaridad de 1 a 9) puede no representar más similaridad que la respuesta 6 de otro sujeto. Además, no podemos decir que la respuesta 6 de un sujeto significa lo mismo que la respuesta 6 de otro sujeto. No estamos seguros de que las diferentes personas estén usando la escala de respuestas de la misma manera. En efecto, estamos bastante seguros de que no usan la escala de manera idéntica. De esta manera, decimos que el “significado” de las mediciones son condicionales sobre el sujeto (matriz) que está respondiendo. Llamamos a ésta clase de datos matriz-condicional. Además, para datos matriz-condicional, cada matriz es una partición de los datos.

Otro tipo común de condicionalidad, llamado *columna condicional*, es ejemplificado por los datos multivariados. En datos multivariados cada columna de datos representa una variable medida (tal como el sexo de una persona, el nivel socio-económico, la estatura

y el peso de una persona, su IQ, etc.). La idea importante es que los datos multivariados son (esencialmente siempre) columna condicional.

Este es un claro ejemplo del hecho de que las características de medición de una partición no tienen que corresponder a aquellas de otra partición. Una de las variables es el sexo, que es binaria; otra es el nivel socioeconómico, que es probablemente continua ordinal; las siguientes son la estatura y el peso, las cuales son de razón, y pueden ser consideradas razonablemente como discretas; y finalmente IQ, puede ser una variable ordinal o de intervalo.

Formalmente, establecemos que el dominio de la transformación de medición t es independiente del tipo de condicionalidad. Para datos matriz-condicional el dominio es una sola matriz de datos y la transformación es denotada por t_k para indicar que hay una transformación separada para cada matriz k . Para datos columna-condicional el dominio es una sola columna de una sola matriz, y la transformación es denotada por t_{jk} . La discusión anterior de medición de nivel y proceso estaba implícitamente en términos de datos no condicionales. A pesar de que todas las definiciones de nivel y proceso deben ser modificadas apropiadamente, no explicaremos estas modificaciones en la medida que son tanto extensas como obvias.

Datos "missing". Ahora podemos explicar ampliamente nuestro concepto de datos "missing". Vimos que los datos "missing" pueden ser considerados como si estuvieran definidos en el nivel nominal, y con una sola categoría de observación (la de no observación). Observamos, sin embargo, que necesitamos un concepto adicional, el de condicionalidad, para explicar completamente nuestra idea de datos "missing". De esta forma, ya puede establecerse la idea completa de lo que es un dato "missing". Consideramos los datos "missing" como celdas de observación las cuales forman sus

propias particiones separadas, llamadas las particiones de los datos “missing”. Todas las observaciones “missing” de una partición de datos “missing” caen en una categoría de observación, la categoría “missing”.

Puesto que la partición de datos “missing” tiene una sola categoría de observación, y puesto que la noción de nivel de medición se refiere a las restricciones entre las categorías, ninguna de las restricciones de nivel de medición (4) y (5) se pueden aplicar. De este modo, no hay nivel de medición para los datos “missing”. Sin embargo, puesto que el concepto de proceso de medición hace referencia a las restricciones dentro de las categorías de observación, la noción de proceso de medición sí aplica a los datos “missing”. Aunque esto puede sonar algo extraño, en realidad corresponde al interés frecuente que tiene el investigador cuando se enfrenta con qué hacer cuando tiene varias observaciones “missing”. Él se pregunta qué ha causado las observaciones “missing”. Una de las posibilidades es asumir que las observaciones “missing” han sido causadas por el mismo proceso subyacente, mientras que otra posibilidad es suponer que las observaciones “missing” han sido causadas por varios procesos.

El primer supuesto, el cual dice que un solo hecho causó los datos “missing”, implica que a todas las observaciones missing debería asignárseles un mismo número real como cuantificación óptima. Este supuesto corresponde a lo que llamamos datos “missing” discretos. El último supuesto, que varias causas contribuyeron a las observaciones “missing”, implica que un continuo de números deberían ser asignados. Esto corresponde a lo que consideramos datos “missing” continuos.

Finalmente, cuando decimos que hay varias particiones de datos “missing”, queremos señalar que los datos “missing” pueden ser condicionales. Por ejemplo, si tenemos datos multivariados y hay datos “missing” en dos variables diferentes, probablemente

asignaríamos las observaciones “missing” a dos particiones separadas, una para cada variable.

2.2 Matrices Indicador

En la sección anterior discutimos la teoría de medición desde la perspectiva de restricciones impuestas sobre transformaciones de datos. En ésta sección discutiremos nuestra teoría desde una perspectiva diferente, aquella proporcionada por la idea de que los datos pueden ser representados por parámetros cuyos valores queremos estimar.

Por ahora puede sonar algo inusual discutir sobre parámetros de los datos. Después de todo, siempre asociamos parámetros con modelos. Sin embargo, con datos cualitativos es útil pensar que cada categoría de observación puede ser representada por un parámetro, cuyo valor queremos estimar de alguna manera óptima (Giffi, 1981, explora la posibilidad de usar varios parámetros por categoría). El valor asignado al parámetro de cada categoría de observación es la “cuantificación” de la categoría. Después de determinar los mejores valores de los parámetros hemos “óptimamente cuantificado” los datos.

Redefinamos el objetivo de la cuantificación óptima desde esta nueva perspectiva: Queremos estimar valores para los parámetros de los datos (categorías de observación) de forma tal que se cumplan dos características: Primera, la estimación debe satisfacer perfectamente las restricciones de medición; y segundo, debe proporcionar una relación de mínimos cuadrados para el modelo, dado que las restricciones de medición se cumplen perfectamente, y dadas ciertas consideraciones de normalización.

Para discutir la cuantificación óptima desde el punto de estimación de los parámetros de los datos, se debe introducir la noción de *matriz indicador*. Esta matriz representa

los datos de una partición particular en una forma tal que indica la categoría en la cual la observación cae. Existe una matriz indicador para cada partición.

Por ahora, la matriz indicador U_p la definimos como una matriz binaria de $n \times n_c$, con una fila para cada una de las n observaciones en la partición p , y una columna para cada una de las n_c categorías (más adelante se generalizará esta definición). Los elementos de U_p indican la pertenencia a una categoría:

$$u_{pic} = 1 \text{ si } o_i \text{ pertenece a la categoría } c \\ = 0 \text{ en otro caso.}$$

En el resto de esta sección eliminaremos el subíndice p de U_p , pero se entenderá que estamos discutiendo los datos de una partición específica, y que la discusión es válida para cualquiera y para todas las particiones sin pérdida de generalidad.

Datos Nominales. Con la anterior definición de U podemos ver que la transformación t^{dn} (discreta nominal) es un proceso muy simple de estimación de parámetros. En efecto, corresponde a la técnica del “optimal scoring” de Fisher (1938), la cual consiste en estimar los datos óptimamente cuantificados z_i^* como la media de las estimaciones \hat{z}_j del modelo las cuales corresponden a aquellas observaciones o_j que están en la misma categoría que o_i . Puesto que los z_i^* son las medias de los \hat{z}_j apropiados, su obtención se puede lograr minimizando los residuales $\|\hat{z} - \mathbf{z}^*\|$. Sin embargo, el índice que queremos minimizar es un índice *normalizado* de residuales denominado el Índice de Stress de Kruskal definido por

$$S = \frac{\|\hat{z} - \mathbf{z}^*\|}{\|\mathbf{z}^*\|} \quad (6)$$

Debido a que \mathbf{z}^* aparece tanto en el numerador como en el denominador, S no se minimiza con los valores que minimizan el numerador. Como se discutirá en la sección 2.4, debemos hacer una normalización de los \mathbf{z}^* para minimizar a (6). Para resaltar este aspecto, introducimos los datos cuantificados *no normalizados* \mathbf{z}^u los cuales minimizan los residuales no normalizados $\|\hat{\mathbf{z}} - \mathbf{z}^u\|$, y reservamos \mathbf{z}^* para los datos cuantificados normalizados los cuales minimizan (6). Formalmente, \mathbf{z}^u está definido como

$$\ell^{dn}: \mathbf{z}^u = U(U'U)^{-1}U' \hat{\mathbf{z}} \quad (7)$$

donde U fue definida antes. Observe que $U'U$ es una matriz diagonal de $n_c \times n_c$ con una fila y una columna para cada categoría de observación, y con el número de observaciones en cada categoría sobre la diagonal. Además, $U' \hat{\mathbf{z}}$ es un vector columna de n_c elementos con la suma de los \hat{z}_j como sus componentes. Finalmente, $(U'U)^{-1}U' \hat{\mathbf{z}}$ es un vector columna de n_c elementos con las medias de los \hat{z}_j como sus componentes. *Estas* son las estimaciones de mínimos cuadrados no normalizadas de cada uno de los n_c parámetros de los datos para la partición bajo consideración.

El caso continuo nominal es más complejo que el discreto nominal. Esto se debe a que la continuidad nominal, no involucra restricciones de medición. Para la transformación ℓ^{cn} imponemos las restricciones del proceso continuo (3) de que cada observación óptimamente cuantificada cae en algún intervalo, sin colocar restricciones sobre la formación de ellos. De esta manera, podríamos seleccionar cotas inferior y superior arbitrariamente grandes que permitirían que todas las observaciones óptimamente cuantificadas sean iguales a todas las observaciones originales, minimizando así las diferencias cuadráticas trivial y totalmente.

Naturalmente, este proceso carece de sentido. Por tanto, se propone un proceso alternativo el cual proporciona intervalos disjuntos y contiguos, evitando así las posibilidades triviales permitidas por el proceso anterior.

El proceso de estimación para la transformación continua nominal t^{cn} , posee dos etapas: En la primera etapa tratamos los datos como si fueran discretos nominales y realizamos un análisis ALSOS completo basados en este supuesto. Cuando este proceso ha terminado entramos a la segunda etapa en la cual se tratan los datos como si fueran continuos ordinales (lo veremos más adelante) y se realiza un segundo análisis ALSOS completo. Observe que en ninguna de las etapas asumimos realmente que los datos son continuos nominales. Sin embargo, las hipótesis que son usadas no violan la naturaleza continua nominal de los datos. En la primera etapa usamos información categórica para obtener la cuantificación de mínimos cuadrados de cada categoría. En la segunda, la cuantificación es usada para definir un orden en las categorías de observación, el cual es a su vez empleado para definir las cotas del intervalo.

Hay tres cosas importantes para resaltar en este procedimiento. La primera, proporciona una cuantificación de mínimos cuadrados, la cual es consistente con, pero más estricta que, las restricciones del proceso continuo nominal discutido antes. Específicamente, el proceso proporciona intervalos disjuntos, mientras las restricciones discutidas antes podrían permitir intervalos no disjuntos. Segundo, el procedimiento descrito aquí *no* es el procedimiento pseudo-ordinal discutido por de Leeuw, Young y Takane (1976), sino que es un procedimiento más reciente que evita los problemas de divergencia mencionados en dicho artículo. Tercero, el procedimiento es convergente pero no es estrictamente de mínimos cuadrados, debido a que puede converger a un orden de intervalo no óptimo.

Datos ordinales. Los procedimientos de estimación para las transformaciones ordinales t^{do} y t^{co} , necesitan extender la definición de matriz indicador dada antes. U sigue siendo una matriz binaria, pero en este caso es una matriz de $n \times n_b$, donde n_b es el número *bloques* necesitados para imponer una restricción ordinal. Un elemento de U indica pertenencia a un *bloque* en una forma paralela a la indicación de pertenencia a una categoría para el nivel nominal.

Para el caso discreto ordinal n_b nunca es mayor que n_c (el número de categorías) y U representa una fusión de categorías de observación. Dado el U adecuado, Young (1975^a) mostró que:

$$t^{do}: \mathbf{z}^u = U(U'U)^{-1}U' \hat{\mathbf{z}} \quad (8)$$

Donde U está basa en la transformación *secundaria* monótona de mínimos cuadrados de Kruskal (1964). U señala cuales categorías deberían ser unidas (ablocadas) de forma tal que se satisfagan las restricciones ordinales. Observe que $U'U$ es una matriz diagonal ($n_b \times n_b$) que contiene en su diagonal el número de observaciones en cada bloque. Además, $(U'U)^{-1}U' \hat{\mathbf{z}}$ es un vector columna con las n_b cuantificaciones óptimas no normalizadas, que son estimadores de mínimos cuadrados que preservan las características de medición discreta ordinal de los datos. En la sección 3.1 se da un ejemplo.

Para el caso continuo ordinal, n_b puede ser o no mayor que n_c . U indica cuales *observaciones* (no categorías) deberían ser unidas, de forma que se preserven las restricciones ordinales. Dada la U apropiada, Young (1975^a) mostró que:

$$t^{do}: \mathbf{z}^u = U(U'U)^{-1}U'P \hat{\mathbf{z}} \quad (9)$$

Donde U y P están basadas en la transformación *primaria* monótona de mínimos cuadrados de Kruskal (1964) (Vea de Leeuw, 1975, para una prueba y de Leeuw, 1977^a para una transformación ordinal adicional).

La matriz P es una matriz de permutación (nxn) binaria y diagonal por bloques. Posee n_b bloques, cada uno de los cuales tiene un orden igual al correspondiente elemento de U'U. Cada bloque es una matriz de permutación con un sólo uno en cada fila y columna. P tiene ceros fuera de los bloques. La matriz U'U se interpreta igual que antes (número de observaciones en cada bloque). La matriz $U(U'U)^{-1}U'P \hat{\mathbf{z}}$ contiene los parámetros estimados de mínimos cuadrados para cada categoría de observación.

En la sección 3.1 presentamos ejemplos detallados de U para t^{dn} y para t^{do} , como también de U y P para t^{co} . Estos ejemplos también presentan el proceso por medio del cual se construyen U (y P) para los casos ordinales. Es muy importante observar que solamente para t^{dn} podemos conocer U antes de que el análisis empiece: Ella es simplemente la estructura categórica de los datos. En los casos ordinales se debe determinar a U (y a P) de modo que las propiedades ordinales de los datos se cumplan. En estos casos U (y P) *no* son conocidas antes del análisis, sino que debemos resolverlas! Ellas son *variables* que debemos resolver, mientras que U en el caso discreto nominal es una constante.

Esta diferencia tiene varias implicaciones. Una implicación es que la solución para \mathbf{z}^u es mucho más lenta y más compleja cuando los datos son ordinales. Otra implicación es que la habilidad para determinar los grados de libertad se pierde con datos ordinales.

La última implicación entraña que los procedimientos inferenciales son más difíciles de determinar para ordinales.

Datos “missing”. Cuando tenemos datos “missing”, las celdas de las observaciones vacías son removidas de cualquier partición que estén y son colocadas en una o más particiones llamadas particiones de datos “missing”. Hay una partición de datos “missing” para datos “missing” no condicionales, y más de una para datos “missing” condicionales.

En el caso de datos “missing” discretos, consideramos que todos los datos missing en una partición residen en una sola categoría. Por tanto, U es un vector de n unos, donde n es el número de datos “missing” en la partición. Se aplica la ecuación (7) para calcular \mathbf{z}^u (los datos missing óptimamente cuantificados) el cual, debido a la naturaleza de U , es un vector cuyos elementos son todos iguales a la media de \hat{z} , las estimaciones del modelo de los datos \hat{z} “missing”.

Para el caso de los datos “missing” continuos, los datos “missing” son codificados, en U , como si fueran cada uno una categoría separada. Entonces el número de categorías es igual al número de observaciones “missing”, y U es una matriz identidad de $n \times n$. Por tanto, (7) se simplifica como $\mathbf{z}^u = \hat{z}$, y cada dato “missing” es óptimamente cuantificado como su estimación por el modelo. Observe que este tratamiento de los datos “missing” continuos no concuerda con la discusión al final de la sección anterior. Sin embargo, es matemáticamente equivalente y más simple para usar en la definición de U .

Datos Cuantitativos. A pesar de que el enfoque del artículo es sobre datos cualitativos, es importante gastar un momento en el proceso de estimación de datos cuantitativos. Con la definición apropiada para U la transformación t^d puede ser escrita, en notación matricial, como,

$$t^d: \mathbf{z}^u = U \delta \quad (10)$$

Donde U es una matriz con una fila para cada observación y $p+1$ columnas, y una columna para cada potencia entera del vector de observaciones, \mathbf{o} . La primera columna es la potencia cero (es decir, todos son unos), la segunda columna es la primera potencia (es decir, los mismos \mathbf{o}), la tercera columna es la segunda potencia \mathbf{o}^2 , etc. La estimación de mínimos cuadrados no normalizados de \mathbf{z}^u es:

$$t^{do}: \mathbf{z}^u = U(U'U)^{-1}U' \hat{\mathbf{z}} \quad (11)$$

Observe que en este caso, otra vez, como en el caso discreto nominal, U es conocida antes de que se haga el análisis.

2.3 Proyección Cónica

Es importante observar que para todas las clases de mediciones discutidas aquí, la correspondiente transformación t puede ser considerada, para cada partición, como si regresáramos las estimaciones $\hat{\mathbf{z}}$ del modelo sobre las observaciones originales \mathbf{o} en un sentido de mínimos cuadrados no normalizados y bajo las restricciones de medición apropiadas. En particular, cada t puede ser representada como un operador proyección de la forma:

$$E = U(U'U)^{-1}U \quad (12)$$

Donde la definición particular de U depende de las características de medición. Esto significa que podemos señalar el importante hecho de que

$$\mathbf{z}^u = E(\hat{\mathbf{z}}) \quad (13)$$

Cuando formalmente observamos que la idea de mínimos cuadrados está definida como:

$$\phi^2 = \|\mathbf{z}^u - \hat{\mathbf{z}}\|^2 = (\mathbf{z}^u - \hat{\mathbf{z}})'(\mathbf{z}^u - \hat{\mathbf{z}}) \quad (14)$$

y definimos $F = I - E$, entonces vemos que

$$\phi^2 = \hat{\mathbf{z}}' F \hat{\mathbf{z}} \quad (15)$$

lo cual enfatiza el hecho de que cada transformación optimiza el producto vector de las estimaciones del modelo y alguna combinación lineal de las mismas estimaciones del modelo, donde la combinación lineal está determinada por las restricciones de medición. Este punto ha sido subrayado en una situación más restringida por Young (1975^a), y en el presente contexto primero fue observado por Young, de Leeuw y Takane (1976).

Geoméricamente, el operador proyección E , proyecta las estimaciones \hat{z} del modelo sobre la más próxima superficie de un cono de datos \mathbf{o} . La proyección son los datos óptimamente cuantificados no normalizados \mathbf{z}^u .

Geoméricamente hablando, las estimaciones del modelo, los datos óptimamente cuantificados, y los datos originales, pueden ser considerados como subespacios de un espacio cuya dimensión es muy alta. También podemos dibujar los parámetros del modelo dentro de un espacio paramétrico.

La figura 2 representa las relaciones geométricas entre los subespacios del modelo, los datos y la cuantificación óptima, al igual que la del espacio paramétrico. Observe que los subespacios del modelo, los datos y la cuantificación óptima son subespacios de un solo espacio “*problema*” de dimensión n , donde cada observación está representada por una dimensión del espacio. Este espacio es llamado espacio “*problema*”, puesto que es en este espacio en el que caracterizamos y resolvemos el problema de análisis de datos bajo consideración. Observe que el espacio problema es un espacio de números reales, y que tiene una dimensión para todas las observaciones en todas las particiones incluyendo las particiones de datos “missing” (si ellos existen).

Enfatizamos que el espacio de los parámetros *no* es un subespacio del espacio problema. Dicho espacio es de dimensión p , con una dimensión para cada uno de los p parámetros. Generalmente p es mucho menor que n , reducción de dimensión que representa la parsimonia de la descripción de los datos del modelo. También enfatizamos que los subespacios del modelo, cuantificación y datos tienen menos de n dimensiones, pero son subespacios del espacio problema. Además, cada partición está representada por su propio único subespacio de modelo, cuantificación y datos, de modo que cuando hay m particiones el espacio problema contiene m subespacios de

modelo, m subespacios de cuantificaciones óptimas y m subespacios de datos. En la figura 2 solamente representamos la geometría de una partición, por simplicidad y sin perder generalidad.

En el espacio problema tenemos representados geoméricamente los subespacios de las estimaciones modelo y de la cuantificación óptima como vectores y el subespacio de los datos originales como un cono. Además, los dos vectores y el cono se intersectan en el origen del espacio problema. Hemos escogido el tipo de representaciones para cada uno de los tres subespacios por razones específicas. Representamos el subespacio de cuantificación óptima como un vector geométrico que va desde el origen, para enfatizar el hecho de que los elementos del vector algebraico z^u definen un punto en el espacio problema, y que si formamos el vector geométrico que conecta ese punto con el origen del espacio problema, entonces todos los otros puntos sobre el vector geométrico son equivalentes a z^u en el nivel de razón de medición. En términos de las restricciones discutidas antes, cualquier punto en el subespacio de cuantificación óptima de la figura 2 es equivalente a cualquier otro punto. (Como será discutido en la sección 2.4, las restricciones de normalización seleccionarán un punto específico z^* sobre el vector de cuantificación óptima.) Por el mismo tipo de razones, representamos el subespacio del modelo como un vector geométrico.

Por otro lado, representamos el subespacio de los datos como un cono geométrico, no como un vector geométrico. Aunque la representación es diferente, el razonamiento es el mismo: para el subespacio de los datos un cono representa adecuadamente las características de medición, mientras que para los subespacios del modelo y cuantificación óptima, un vector geométrico es la representación adecuada. Si reflexionamos sobre las restricciones dadas en (2) hasta (5), veremos que todas ellas

pueden ser representadas geoméricamente como conos (algunas restricciones implican ciertos conos degenerados, como por ejemplo, vectores). Este punto ha sido discutido por Young y Takane (1976) y de Leeuw (1975, Nota 3).

Se observará que el vector de cuantificación óptima está representado sobre la superficie del cono. Puesto que los subespacios de cuantificación óptima y datos son completamente equivalentes en términos de las características de medición de los datos, el vector de cuantificación óptima debe estar contenido en el cono de los datos. Puesto que los subespacios del modelo y la cuantificación óptima están tan cercanos como sea posible en el sentido de los mínimos cuadrados, el vector de cuantificación

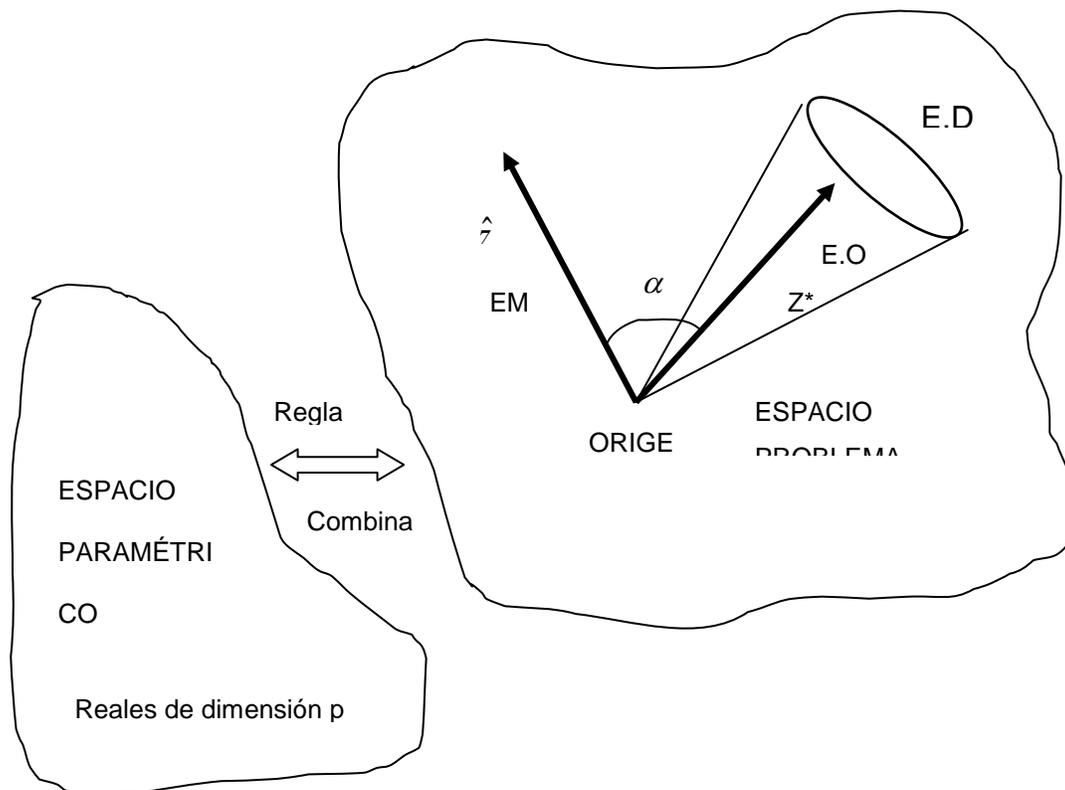


Figura 2

Fundamentos geométricos de los algoritmos ALSOS con énfasis en la proyección cónica óptima debe estar cerca del vector del modelo. Por esto, generalmente se da el caso de que el vector de cuantificación óptima esté sobre la superficie del cono, puesto que la superficie del cono es la parte del cono que está más cerca del subespacio del modelo. (La única vez en la que el vector de cuantificación óptima está dentro del cono es cuando el subespacio del modelo está también dentro del cono, lo cual sucede solamente cuando el modelo se ajusta perfectamente a los datos).

Finalmente, observe el ángulo α entre los vectores del modelo y de la cuantificación óptima. El ángulo α representa el buen ajuste entre los dos vectores, siendo medido el ajuste por (14). Mientras más pequeño sea el ángulo, mejor el ajuste. Cuando el ángulo es cero el ajuste es perfecto (esto usualmente significa que el vector del modelo y el de cuantificación óptima caen dentro del cono, pero también significar que los dos estén sobre la superficie del cono). Observe que hay una dificultad asociada con un subespacio de modelo que solamente contenga ceros. En este caso, el ajuste entre los vectores de modelo y de cuantificación óptima es perfecto y el ángulo α en la figura 2 es cero. Sin embargo, el ajuste es perfecto solamente en un sentido trivial y sin interés. Por tanto, debemos asegurar que cualquiera que sea el procedimiento que adoptemos la solución no esté en el origen del espacio problema. Tales soluciones se evitan normalizando la longitud de los vectores de modelo y cuantificación óptima a alguna longitud no cero arbitraria.

2.4 Normalización

Tal como lo hemos mencionado, una forma trivial y no deseable de minimizar a (14) es hacer que el subespacio del modelo \hat{z} sea igual a cero. Entonces \mathbf{z}^u también es cero para todas las transformaciones, y por tanto ϕ^2 es cero para cada partición. Por esta razón los comentarios con respecto a la normalización fueron hechas precisamente antes de (6). En esta sección discutiremos la normalización.

En los programas ALSOS se usan diferentes normalizaciones. Se introducen las normalizaciones para evitar soluciones representadas por el origen del espacio problema (Vea Figura 2) u otros tipos de soluciones triviales. Las distintas condiciones de normalización han sido discutidas por Kruskal y Carroll (1969), Sands y Young (1980), Young (1972); y de Leeuw (Nota 3). Dos de estas condiciones son equivalentes a definir ya sea

$$\phi^2_a = (\mathbf{z}^*_a - \hat{z})' (\mathbf{z}^*_a - \hat{z}) / (\hat{z}' \hat{z}) \quad (16)$$

o

$$\phi^2_b = (\mathbf{z}^*_b - \hat{z})' (\mathbf{z}^*_b - \hat{z}) / (\mathbf{z}^*_b' \mathbf{z}^*_b) \quad (17)$$

donde \mathbf{z}^*_a y \mathbf{z}^*_b son las versiones normalizadas de \mathbf{z}^u las cuales optimizan a ϕ^2_a y ϕ^2_b respectivamente.

Ahora debería ser claro que \mathbf{z}^u minimiza a (16), puesto que de la sección 2.3 sabemos que ella minimiza el numerador de (16), y puesto que \mathbf{z}^u no se encuentra en el denominador de (16).

De esta manera,

$$\mathbf{z}^*_{a} = \mathbf{z}^u. \quad (18)$$

Además, por las características de medición de \mathbf{z}^u , y como fue dibujado en la Figura 2,

$$\mathbf{z}^*_{b} = b\mathbf{z}^u = \mathbf{z}^* \quad (19)$$

donde b es un “valor de normalización”, el cual debe ser determinado. Observe que estamos usando específicamente \mathbf{z}^* para referirnos a la normalización del \mathbf{z}^u que minimiza a (17).

Observando la Figura 3 podemos entender las relaciones entre ϕ^2 y ϕ^2_b y las relaciones entre \mathbf{z}^u y \mathbf{z}^* . Ésta figura presenta, en más detalle, una parte del espacio problema presentado en la Figura 2. Concretamente, estamos mirando abajo, en una parte de la superficie del cono de datos, donde la superficie está representada por el área de forma irregular. Arriba de la superficie del cono se presenta el vector $\hat{\mathbf{z}}$ del modelo. Observe que el vector sale del origen del espacio problema y del cono de los datos, el origen denotado como 0.0. La proyección ortogonal del vector del modelo sobre la superficie del cono da \mathbf{z}^u , es decir los datos óptimamente cuantificados *no normalizados*. Como vimos en la sección 2.3, esta proyección está representada por el operador E (12) el cual minimiza ϕ^2 (14), el índice de ajuste *no normalizado*.

Geoméricamente, la proyección minimiza el ángulo entre \hat{z} y \mathbf{z}^u , y de ésta manera la longitud del vector de residuales \mathbf{r}^u , y de ésta forma a (14) el cual es simplemente el cuadrado del longitud del vector de residuales.

Sin embargo, \mathbf{z}^u no minimiza ϕ^2_b , aunque si minimiza ϕ^2 , como lo demostraremos a continuación. Recuerde que α , el ángulo entre \hat{z} y \mathbf{z}^u , ha sido minimizado proyectando \hat{z} sobre la superficie del cono. Es fácil ver que la proyección define un triángulo rectángulo tal que

$$\text{sen}^2 \alpha = (\mathbf{r}^u)^2 / \hat{z}^2 = (\mathbf{r}^u \mathbf{r}^u / \hat{z} \hat{z}) = \phi^2_a, \quad (20)$$

Puesto que la proyección ortogonal de \hat{z} sobre la superficie del cono necesita de un ángulo recto en la superficie del cono (indicado por el ángulo de 90° más bajo). Sin embargo, proyectamos ortogonalmente desde \hat{z} (indicado por el ángulo superior de 90°) sobre la superficie del cono en el plano definido por \hat{z} y \mathbf{z}^u , obtenemos una proyección \mathbf{z}^* y un nuevo triángulo rectángulo tal que:

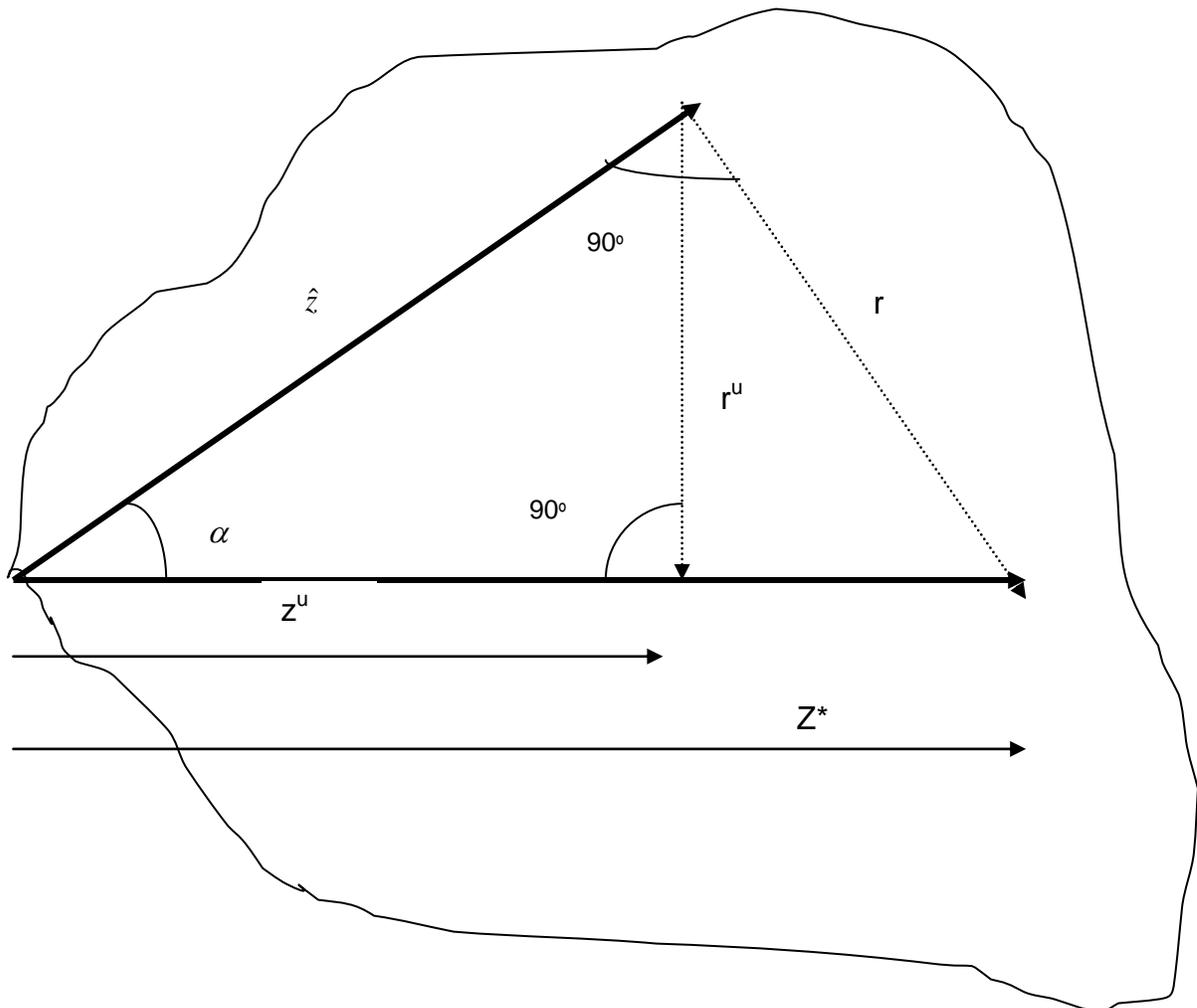
$$\text{sen}^2 \alpha = (\mathbf{r}^2 / \mathbf{z}^{*2}) = (\mathbf{r} \mathbf{r} / \mathbf{z}^* \mathbf{z}^*) = \phi^2_b, \quad (21)$$

y, puesto que,

$$\mathbf{r} = \mathbf{z}^* - \hat{z} \quad (22)$$

vemos que

$$\text{sen}^2 \alpha = (\mathbf{z}^* - \hat{z})'(\mathbf{z}^* - \hat{z}) / \mathbf{z}^* \mathbf{z}^* \quad (23)$$



0.0

Figura 3

Representación geométrica de la normalización en los algoritmos ALSOS

De este modo \mathbf{z}^* minimiza ϕ^2_b (17). Además cuando usamos \mathbf{z}^u en ϕ^2_a y \mathbf{z}^* en ϕ^2_b , ocurre (de (20) y (23)) que:

$$\phi^2_a = \phi^2_b \quad (24)$$

Por tanto, estas dos fórmulas aparentemente distintas son en efecto iguales, y no importa cual normalización se escoja.

Sin embargo, no hemos descubierto como obtener \mathbf{z}^* de \mathbf{z}^u ; es decir, aún necesitamos determinar el valor de b en (19). El valor de b se obtiene observando que

$$\cos^2 \alpha = (\mathbf{z}^u, \mathbf{z}^u) / (\hat{\mathbf{z}}, \hat{\mathbf{z}}) \quad (25)$$

y que

$$\cos^2 \alpha = (\hat{\mathbf{z}}, \hat{\mathbf{z}}) / (\mathbf{z}^*, \mathbf{z}^*) \quad (26)$$

De esta forma

$$(\hat{\mathbf{z}}, \hat{\mathbf{z}}) / (\mathbf{z}^*, \mathbf{z}^*) = (\mathbf{z}^u, \mathbf{z}^u) / (\hat{\mathbf{z}}, \hat{\mathbf{z}}) \quad (27)$$

Y

$$(\mathbf{z}^* \mathbf{z}^*) = (\hat{z}' \hat{z})(\hat{z}' \hat{z}) / (\mathbf{z}^u \mathbf{z}^u) \quad (28)$$

Observando que los valores entre paréntesis son escalares, vemos que

$$\begin{aligned} (\mathbf{z}^* \mathbf{z}^*) &= (\hat{z}' \hat{z})(\mathbf{z}^u \mathbf{z}^u)(\hat{z}' \hat{z}) / (\mathbf{z}^u \mathbf{z}^u) (\mathbf{z}^u \mathbf{z}^u) \\ &= [(\hat{z}' \hat{z})\mathbf{z}^u / (\mathbf{z}^u \mathbf{z}^u)] [\mathbf{z}^u(\hat{z}' \hat{z}) / (\mathbf{z}^u \mathbf{z}^u)] \end{aligned} \quad (29)$$

De este modo, se sigue que

$$\mathbf{z}^* = \mathbf{z}^u(\hat{z}' \hat{z}) / (\mathbf{z}^u \mathbf{z}^u) \quad (30)$$

Por tanto, en (19) vemos que

$$\mathbf{b} = (\hat{z}' \hat{z}) / (\mathbf{z}^u \mathbf{z}^u) \quad (31)$$

Un rápido análisis de la Figura 3.1 revelará que

$$\begin{aligned} \mathbf{b} &= 1 / \cos^2 \alpha \\ &= 1 / (1 - \phi_b^2) \end{aligned} \quad (32)$$

De ésta manera observamos que

$$\phi^2_a = \phi^2_b = 1 - 1/b \quad (33)$$

Finalmente, La ortogonalidad de \mathbf{z}^u y \mathbf{r}^u nos permite mostrar también que

$$b = (\hat{\mathbf{z}}' \hat{\mathbf{z}}) / (\mathbf{z}^u' \mathbf{z}^u) \quad (34)$$

Estas relaciones entre las diferentes expresiones para b fueron observadas primero por Sands y Young (1980). El hecho de que optimizar la función de pérdida no normalizada por medio de un operador proyección está simplemente relacionado con el problema más difícil de optimizar una función de pérdida normalizada fue primero discutido por de Leeuw (1975) y de Leeuw, Young y Takane (1976) y probado por de Leeuw (Nota 3).

Tanto Sands y Young (1980) como de Leeuw (Nota 3) también discuten las relaciones entre las dos versiones de la segunda fórmula de Stress de Kruskal (1965)

$$\phi^2_c = (\mathbf{z}_c^* - \hat{\mathbf{z}})' (\mathbf{z}_c^* - \hat{\mathbf{z}}) / (\hat{\mathbf{z}} - \bar{\bar{\mathbf{z}}})' (\hat{\mathbf{z}} - \bar{\bar{\mathbf{z}}}) \quad (35)$$

$$\phi^2_d = (\mathbf{z}_d^* - \hat{\mathbf{z}})' (\mathbf{z}_d^* - \hat{\mathbf{z}}) / (\mathbf{z}_d^* - \bar{\bar{\mathbf{z}}}_d)' (\mathbf{z}_d^* - \bar{\bar{\mathbf{z}}}_d) \quad (36)$$

donde la barra sobre una letra indica un vector constante de medias del vector indicado. Razonando en forma similar a la anterior concluimos que

$$\mathbf{z}_c^* = \mathbf{z}^* \quad (37)$$

y que
$$\mathbf{z}_d^* = (\mathbf{z}^u - \hat{\mathbf{z}})[(\hat{\mathbf{z}} - \bar{\mathbf{z}})'(\hat{\mathbf{z}} - \bar{\mathbf{z}})/(\mathbf{z}^u - \bar{\mathbf{z}}^u)'(\mathbf{z}^u - \bar{\mathbf{z}}^u)] + \hat{\mathbf{z}} \quad (38)$$

2.5 Particiones

El propósito final de ésta sección tiene que ver con lo que llamamos “particiones de medición”. En algunos conjuntos de datos, se puede pensar que todas las observaciones son generadas por una sola fuente de medición. Además, con algunos de estos conjuntos de datos, la fuente de medición genera datos de una forma tal que podemos asumir razonablemente que todas las observaciones tienen las mismas características de medición. Por ejemplo, cuando un sujeto hace juicios de similitud sobre pares de estímulos, entonces podemos pensar razonablemente que todos los juicios han sido generados por una sola fuente (el sujeto) y que todos tienen las mismas características de medición (ordenamiento discreto de juicios de similitud). Sin embargo, para otros tipos de datos será claro que aparecen desde varias fuentes de medición, o varias escalas.

Por ejemplo, cuando tenemos datos multivariados con variables tales como sexo, edad, color del cabello, ingreso, educación y preferencia política obtenidas de un conjunto de personas, probablemente pensaríamos de cada variable como una única fuente de medición que tiene sus propias características de medición: sexo es binaria; edad de razón; color del cabello es nominal; ingreso puede ser de intervalo; educación puede ser ordinal; y preferencia es ordinal. En este caso quisiéramos particionar el espacio de

datos dentro de un conjunto de subespacios mutuamente excluyentes y exhaustivos (uno para cada variable).

Mientras que la idea de partición se relaciona más claramente con los datos multivariados, dicha noción también es útil para otro tipo de datos. Por ejemplo, el concepto datos de similaridad condicional de Coombs (1964) (para la cual un sujeto ordena la similaridad de $n-1$ “comparación de estímulos” con respecto al n -ésimo estímulo estándar, y luego lo hacen veces, cada vez con un diferente estímulo como estándar) es un caso en el cual una sola fuente de medición (el sujeto) genera n diferentes escalas de medición (los rangos de orden). Para ésta clase de datos las particiones de medición son de gran uso.

Cuando los datos están particionados, la etapa OS de un procedimiento ALSOS es ligeramente más complicada que cuando no lo están. La diferencia está en que debemos realizar el OS y la normalización separadamente para cada partición, una partición a la vez. Puesto que las particiones son mutuamente excluyentes, y debido a que el OS es hecho separadamente para cada partición, las características de medición de una partición no necesitan conllevar una relación especial con aquellas de otra partición. Esto significa, por ejemplo, que muchos de los procedimientos orientados hacia el datos multivariados (vea la Tabla 1) pueden analizar datos con mezclas de características de medición. Estos datos, los cuales los llamaremos datos con nivel mixto de medición, pueden tener un conjunto de características de medición para una variable, y una completamente diferente para otra variable. En la Sección 3.2 se discute un procedimiento multivariado (MORALS).

Observe que para los datos particionados, la función de pérdida total está definida como la raíz cuadrada del cuadrado medio de las funciones de pérdida de cada

partición. De ésta manera, si ϕ^2_i denota la función de pérdida estandarizada para la i -ésima de las p particiones, definimos la función de pérdida total como

$$\phi = (1/p) \sum_{i=1}^p \phi^2_i \quad (39)$$

Aquí hay una consideración importante que no debe ser pasada por alto. Es absolutamente perentorio que exactamente después de realizar la cuantificación óptima para una partición particular, reemplacemos la cuantificación óptima vieja por la nueva. Como se verá claro en la próxima sección de este artículo, el reemplazo inmediato es imperativo cuando las particiones son dependientes. (Las particiones son “dependientes” si los valores de la cuantificación óptima para al menos una partición, suponiendo que las otras están fijas, son una función de los valores de cuantificación óptima de al menos una de las otras particiones.) Puesto que la dependencia es en general una característica de los datos multivariados, los programas para analizarlos implican el inmediato reemplazo. Este punto ha sido resaltado en Young, de Leeuw y Takane (1976).

Si, en efecto, las particiones son dependientes, entonces hay una consideración adicional. Supongamos que, para datos multivariados, hemos completado un ciclo de cuantificación óptima y que hemos hecho el reemplazo para cada variable. Supongamos ahora que repetimos la cuantificación de una de las variables. Si lo hacemos, la segunda cuantificación óptima de la variable no producirá la misma cuantificación que la primera cuantificación óptima. Porqué? Se debe a que las variables son dependientes. La cuantificación obtenida por medio de la cuantificación óptima de una variable depende de la cuantificación de cada una de las demás variables.

A pesar de que todo esto parece complicado, puede ser probado (de Leeuw, Young y Takane, 1976) que si realizamos iteraciones “internas” (“internas” con respecto al esquema de la Figura 1) del ciclo de cuantificación óptima y reemplazo, entonces este proceso convergería a un punto donde las cuantificaciones no deberían cambiar bajo la repetición de la cuantificación óptima. Sin embargo, en nuestro trabajo, no realizamos tales iteraciones internas de la cuantificación óptima, solamente realizamos el proceso una vez para cada una de las variables (o partición) antes de conectar con la etapa de estimación del modelo (vea la Figura 1). Nuestra experiencia ha mostrado que tales iteraciones internas sólo sirven para decrecer la eficiencia global del procedimiento, y de Leeuw (Nota 3) probó que el número de iteraciones internas no tiene efecto sobre el punto eventual de convergencia.