Metodología de Encuestas
Vol 4, Núm 1, 2002, 45-61

ISSN: 1575-7803

VARIACIÓN Y POSICIÓN DE VARIABLES: ESTRATEGIAS PARA SU ESTUDIO EXPLORATORIO .

Vicente Manzano Arrondo

RESUMEN

En la medición de preferencias y en la comparación de ítems entre sí, surge la necesidad de interpretar las variables resultantes con respecto a diversos criterios. Los dos más habituales son una medida de variación o dispersión y una medida de posición o tendencia central. En este trabajo se proponen algunos modelos de representación gráfica, bajo la denomiación común de "diagrama de posiciones y variaciones" donde el conjunto de variables se dispone en un espacio bidimensional. Una de las dimensiones está representada por una medida transformada de tendencia central. La otra dimensión representa una medida transformada de dispersión. Como estrategias de transformación se recurre a la estandarización y a la acotación en el intervalo (0,1). Esta representación gráfica permite comparar las variables en cada dimensión por separado y extraer conclusiones exploratorias en cuanto a la relación entre ambas dimensiones, a la vez que establecer impresiones sobre el conjunto de variables, objetos o ítems.

Palabras clave: *representación gráfica, diagrama de posiciones y variaciones transformadas, acotación de varianzas, análisis exploratorio*.

Análisis previos

Es generalizado el consejo de realizar determinados análisis con los datos recogidos de un estudio antes de proceder a los procesos estadísticos que se derivan de sus objetivos (por ejemplo, Freixa y otros, 1992; Rojas, 2000). Esta etapa previa ha recibido nombres muy diversos, como análisis exploratorio de datos (Tuckey, 1977) o análisis previo de los datos (Chatfield, 1985). Pero las características que la definen son las mismas. Consiste en una fase del proceso de análisis donde se procede a revisar la información, habitualmente con el objetivo de encontrar errores o incoherencias y depurarlas, señalar valores ausentes e imputarlos, identificar casos peculiares o raros y tomar decisiones sobre las condiciones de su inclusión, comprobar supuestos y, tal vez, realizar transformaciones de variables implicadas, etc. Pero también es la situación ideal para comenzar a familiarizarse con los datos, con su comportamiento univariable, con sus posibles relaciones, etc.

La riqueza de esta fase de análisis es impresionante y puede consistir en la aportación más relevante del estudio. Rao (1994) muestra, por ejemplo, situaciones donde estos procesos pueden demostrar la manipulación intencionada de los datos. Y Chatfield (1985), incluso, expone algún ejemplo donde el análisis descriptivo previo a los estudios de inferencia puede desaconsejar éstos.

Lamentablemente, es una costumbre muy extendida realizar los análisis de inferencia con prontitud, sin llevar a cabo las operaciones previas que se acaban de mencionar. Utilizando una analogía fácil, realizar análisis para tomas de decisión sin considerar la fase de exploración previa, es como ir de vacaciones a una isla desierta, acompañado por un grupo de desconocidos. Esta circunstancia tal vez sea debida a una mezcla entre la formación o preparación insuficiente y la facilidad para la realización de los análisis, aún complejos, gracias a las herramientas informáticas (Solanas y otros, 2001).

¿Qué se llevaría usted a un día de campo?

Para plantear el recurso que se presenta en este trabajo, asentado sobre necesidades previas, vamos a utilizar un ejemplo concreto, real, pero arbitrario para el objetivo de mostrar el diagrama de posiciones y variaciones transformadas:

20 profesores universitarios fueron entrevistados. La tarea que debían realizar consistía en la ordenación de un conjunto de 8 tarjetas. Las instrucciones fueron las siguientes:

> "Vas a pasar el día en el campo, en soledad. Alguien te llevará por la mañana y te recogerá por la noche. Te permiten llevar de 1 a 8 cosas, además de tu vestimenta. Ordena los siguientes elementos (del más preferido al menos preferido) según tu interés por llevártelos contigo."

Los elementos fueron:

1. Un botiquín
2. Un reloj.
3. Papel y lápiz.
4. Una hamaca.
5. Un perro.
6. Un libro.
7. Unos prismáticos.
8. Un almuerzo.

Cada elemento ocupaba una tarjeta independiente. El conjunto de tarjetas era barajado de una entrevista a otra.

7	6	5	8	4	3	2	1
3	8	7	6	4	5	2	1
7	8	4	5	2	3	6	1
6	7	3	5	8	2	4	1
6	4	3	7	8	2	5	1
7	8	4	5	6	3	2	1
5	7	2	4	8	1	6	3
5	7	4	2	8	3	6	1
7	3	5	4	8	2	6	1
8	6	5	1	7	2	4	3
6	8	3	7	2	1	4	5
6	4	2	7	8	1	5	3
7	8	1	6	5	3	2	4
6	8	5	7	1	3	4	2
7	4	2	6	5	1	8	3
6	5	7	3	8	2	4	1
4	3	5	6	8	2	7	1
2	3	6	7	8	5	4	1
1	8	5	7	3	4	6	2
2	4	7	8	6	3	5	1

Cuadro 1: *respuestas de preferencia.*

Los resultados numéricos obtenidos, una vez reorganizada la información, se muestran en el cuadro 1. Cada columna representa a cada uno de los ocho objetos, en el orden en que han sido expuestos. Cada fila constituye el vector de respuestas por persona entrevistada. Las cantidades numéricas se refieren al orden de preferencia. Así, por ejemplo, la primera columna, que se refiere al objeto “Un botiquín”, ha sido preferido en séptimo lugar por el profesor 1, en tercer lugar por el profesor 2, etc.

Posiciones y variaciones

No es nuestra intención aquí reproducir un proceso completo de estudio exploratorio. Incluso, algunos de los tópicos carecen de sentido aquí, como la ausencia de respuesta parcial (harto difícil en tareas de preferencia) o problemas diversos de depuración.

En lo que se refiere al tanteo de interpretaciones univariables, un interrogante básico es obtener alguna medida que represente el grado de preferencia de cada objeto. Dado que contamos con las respuestas de varios entrevistados, una estrategia esperable es acudir a una medida promedio, como puede ser, por ejemplo, la media aritmética. En nuestro ejemplo, los valores de la media aritmética para cada objeto se encuentran en el cuadro 2. Se puede observar, al respecto, que el objeto preferido, por término medio, es el 8 (Un almuerzo, μ=1.85), mientras que el objeto menos preferido es el 2 (Un reloj, μ=5,95) seguido, muy de cerca, por el 5 (Un perro, μ=5,85).

El cuadro 2 contiene también otra medida obligada: un índice de variación. Sabemos que el grado en el que un promedio representa al conjunto del que proviene es una propiedad relativa al grado de variación. Es fácil aceptar que conforme es mayor la dispersión de los datos, es menor el poder de representación de cualquier valor único. Así, se puede observar que el objeto 5 (Un perro) muestra una elevada desviación (σ=2,372), al menos en comparación con el conjunto de los valores de variación. Esta circunstancia indica que se trata de un objeto más y menos preferido (según los profesores entrevistados) que el objeto 2 (σ=1,936), a pesar de constar ambos con un promedio similar.

Variable	Media	Desviación Tipo
1	5.400	1.934
2	5.950	1.936
3	4.250	1.728
4	5.550	1.883
5	5.850	2.372
6	2.550	1.161
7	4.600	1.685
8	1.850	1.195

Cuadro 2: *medias y desviaciones tipo para las variables del estudio*.

Medidas estandarizadas

La comparación entre variables es un caso habitual de estudio exploratorio donde las características de un elemento se interpretan en función del grado en que el resto de los elementos participan de la misma característica. En nuestro estudio, la variable “preferencia por llevar un almuerzo” es peculiar, puesto que acapara el extremo inferior de la escala (es el objeto más preferido, por término medio). Pero también resulta peculiar la variable “preferencia por llevar un perro”, debido, esta vez, a su marcada dispersión: señala la existencia de importantes desacuerdos entre entrevistados.

La caracterización de las variables por varias propiedades estadísticas abre el camino para interpretar éstas considerando varias dimensiones. De hecho, algo se ha hecho ya en ese sentido cuando se ha interpretado el valor promedio de la preferencia por “Un perro” en función de la dispersión o grado de desacuerdo entre los entrevistados. No obstante, se puede ir más allá, considerando los valores tipificados para las dimensiones “posición de la distribución” o “tendencia central” (en este caso, media aritmética) y “variación” o “dispersión” (en este caso, desviación tipo).

La estandarización o tipificación permite expresar las mediciones en términos relativos, de tal forma que resulten independientes a las características de variación y tendencia de la escala. Este efecto se consigue gracias a expresar el valor original como número de desviaciones tipo que lo alejan del promedio de su distribución. El cuadro 3 muestra los resultados de la tipificación para el ejemplo.

La observación de estas dos columnas de puntuaciones estandarizadas nos muestra que lo más sobresaliente es el alto grado medio de preferencia por el objeto “Un almuerzo”, seguido por el elevado desacuerdo (relativo, puesto que es un valor *elevado* en comparación con el resto de las dispersiones) con respecto a la preferencia por el objeto “Un perro” y por el elevado acuerdo o coincidencia en la preferencia por el objeto 6 (Un libro).

Variable	Media	Desviación Tipo
1	.622	.524
2	1.003	.529
3	-.173	-.022
4	.726	.390
5	.933	1.689
6	-1.348	-1.531
7	.069	-.137
8	-1.832	-1.441
Mínimo:	-1.832	-1.531
Máximo:	1.003	1.689

Cuadro 3: *medidas estandarizadas para la media y la desviación tipo*.

Un análisis algo más pormenorizado de la tabla nos puede mostrar la circunstancia de que los valores positivos de variación van acompañados de valores también positivos de promedio. Otro tanto puede decirse de los valores negativos. Esta circunstancia sólo posee la excepción de la preferencia por el objeto 7, "Unos prismáticos"; sin embargo, sus valores de tendencia y dispersión se encuentran tan cercanos a cero que el cambio de signo no parece ser relevante. Esta relación es fácilmente identificable si la comparación se realiza entre ambas dimensiones estandarizadas, y permite establecer una regla de relación entre ambas: conforme aumenta la preferencia por un objeto, lo hace también el grado de acuerdo o coincidencia entre entrevistados. Recordemos que nos encontramos en una fase de *tanteo*, de conocimiento previo sobre los datos y sus posibles relaciones. La afirmación de tendencia entre desviaciones tipo y medias aritméticas, se establece únicamente ante el espacio de competencia de la muestra.

Extensión e información gráfica

Las reflexiones realizadas hasta el momento cuentan con la ventaja de realizarse sobre un conjunto simple de datos que comparten la misma escala. Existe una gran diversidad de situaciones donde la comparación entre medidas estandarizadas de promedios y desviaciones es muy fructífera, pero muestra ciertas limitaciones. Algunos de los puntos que justifican esta afirmación son:

- El contexto de respuestas de preferencias sirve de excusa para mostrar beneficios asociados a la comparación de las dimensiones estandarizadas *posición del conjunto de respuestas* y *dispersión del conjunto de respuestas*. El abanico es mucho más amplio. En general, cualquier contexto donde medie un cuestionario, en sentido amplio, es susceptible de ser abordado con estas estrategias exploratorias.
- En situaciones que se alejan de los datos de preferencia, como las respuestas a ítems independientes, las escalas de medida de éstos pueden variar sensiblemente. Así, por ejemplo, es frecuente que un mismo cuestionario,

utilizado en una encuesta, contenga bloques de ítems con formatos diferentes. En tales casos, las comparaciones entre variables pueden generar sesgos en la interpretación. Un ítem con formato Likert de 5 puntos, por ejemplo, mostrará unos valores esperados para el promedio y la dispersión menores que un formato Likert de 7 puntos. Tal circunstancia obliga a equiparar la escala de todos los ítems considerados en este procedimiento descriptivo. En términos generales, una variable $v \in (a,b)$ puede ser recodificada en otra, $v' \in (a',b')$ siguiendo la transformación:

$$v' = (v - a)\frac{b' - a'}{b - a} + a'$$

La media aritmética queda transformada del mismo modo, mientras que la desviación tipo resultante se ve afectada de forma:

$$S_{v'} = S_v \frac{b' - a'}{b - a}$$

Estas operaciones deberían realizarse con antelación a la estandarización de las medidas.

- En la mayoría de las situaciones reales, el número de variables generadas a partir de los ítems de los cuestionarios es muy superior a las 8 que hemos utilizado en el ejemplo. En tal caso, la utilización de tablas numéricas no es recomendable, puesto que la abundancia de información es inversamente proporcional a la capacidad para obtener conclusiones por parte de quien las realiza.

Para establecer conclusiones sobre la posible relación entre las medidas de posición y de variación, así como para obtener una impresión *de conjunto*, es necesario algo más que un número manejable de variables. Así, por ejemplo, el cuadro 3 ha permitido la posibilidad de mostrar una relación más o menos lineal positiva, pero esta conclusión no es inmediata, ni salta a la vista. Resultaría mucho más fructífero ordenar las medidas estandarizadas en función de alguna de ambas dimensiones, puesto que no es posible establecer una ordenación bidimensional al mismo tiempo.

Sin embargo, sí existe un recurso para organizar la información de tal forma que: se permita un criterio de dos ordenaciones simultáneas, sea compatible con un número más extenso de variables y facilite las conclusiones no sólo por cada dimensión por separado, sino también con respecto a la relación entre dimensiones y a la impresión global: la representación gráfica en un espacio, en este caso, bidimensional.

La figura 1 muestra un diagrama de posiciones y variaciones estandarizadas que aún debe ser matizada.

El diagrama muestra con claridad cuatro grupos de variables y una tendencia aproximadamente lineal positiva. El eje de abcisas representa la medida de posición del conjunto de datos en cada variable (en este caso, media aritmética). El eje de ordenadas se refiere a la medida de dispersión de los datos suministrados por cada variable (en este caso, desviación tipo). La tendencia lineal positiva se refiere aquí a la circunstancia de que un aumento en el valor del promedio de la variable va acompañado de un aumento en el valor de la dispersión. En términos más ajustados al ejemplo: conforme existe un objeto más preferido por término medio, también lo es por término general, es decir, existe mayor acuerdo en escogerlo como objeto preferido.

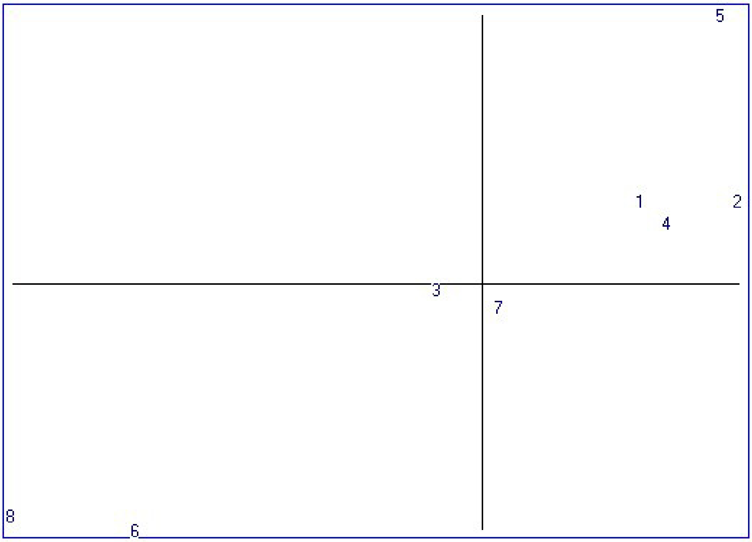

Figura 1: *diagrama de posiciones y variaciones estandarizadas*.

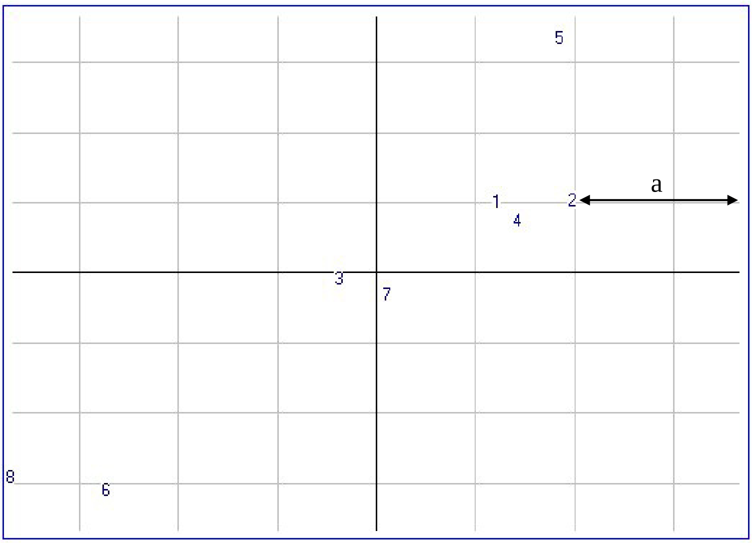

Figura 2: *diagrama cuadriculado de cotas constantes.*

Los cuatro grupos de variables quedan definidos como: (1) alta preferencia y acuerdo para los objetos 6 y 8 (un libro y un almuerzo, respectivamente), (2) niveles medios de preferencia y acuerdo para los objetos 3 y 7 (papel y lápiz, y unos prismáticos, respectivamente), (3) nivel ligeramente bajo de preferencia y de acuerdo para los objetos 1, 2 y

4 (un botiquín, un reloj y una hamaca, respectivamente) y una preferencia ligeramente baja, con un marcado desacuerdo, para el objeto 5 (un perro).

Las conclusiones podrían matizarse o, incluso, extraer aún más información de estas dimensiones, si se operara con una representación gráfica de características algo diferentes. Dos cambios sencillos permitirían estas posibilidades:

- Cuadricular el diagrama siguiendo las unidades estandarizadas como sistema de medida.
- Mantener constantes las cotas de los ejes.

Ambos recursos facilitan la obtención de nuevas conclusiones. Así, por ejemplo, se observa que el objeto más extremo lo representa el 8 en la dimensión “posición”. Lo más sobresaliente es, pues, el alto nivel de preferencia (puntuación más baja) por este objeto. Su puntuación tipificada en la dimensión promedio (-1,832) es la que define la cota en los cuatro extremos. Gracias a esta estrategia, resulta evidente que el grado en que este objeto es preferido es mucho más relevante que el grado en que no lo son los objetos 2 y 5, por ejemplo (obsérvese, al respecto, la distancia *a* que separa el objeto menos preferido, el 2, del extremo derecho de este espacio bidimensional). Por otro lado, el cuadriculado (establecido, en este caso, para ½ desviación tipo) permite observar, entre otros aspectos, que el valor de la distancia *a* se acerca a 1 desviación tipo.

Valores acotados

El recurso de comparar las medidas según sus versiones estandarizadas, tiene la ventaja y el inconveniente de resultar dependiente de las medidas del resto de variables. No obstante, ésta como cualquier otra transformación lineal no modificará las relaciones entre pares comparados, sino más bien las conclusiones generales con respecto a las distancias absolutas. Así, en la figura 2 se observa que el extremo se encuentra *tocado* por el objeto 8 en la dimensión horizontal. Ello no significa que sea imposible un promedio inferior, implica únicamente que no hay valor presente que sea más extremo. Es, pues, una cota empírica, no teórica.

Una estrategia alternativa implicaría re-expresar las medidas utilizadas acotándolas en un intervalo constante, como (0,1), de tal forma que ambos extremos señalen las cotas máximas absolutas o teóricas, en el sentido de que pueden no ser alcanzadas en la práctica. Si la expresión implica una transformación lineal, la configuración de relaciones no sufrirá ningún cambio, pero sí la interpretación con respecto a las cotas. Puede llegar a observarse, por ejemplo, que las distancias entre promedios y variaciones son poco importantes a la luz de lo lejos que se sitúan de los márgenes teóricos.

Una función φ puede ser acotada en (0,1) si se opera con su valor máximo, *max*(φ), y su valor mínimo, *min*(φ), de tal forma que

partiendo de:

$$\min(\varphi) \leq \varphi \leq \max(\varphi)$$

restando el valor mínimo:

$$0 \leq \varphi - \min(\varphi) \leq \max(\varphi) - \min(\varphi)$$

y dividiendo entre la distancia de cotas:

$$0 \leq \frac{\varphi - \min(\varphi)}{\max(\varphi) - \min(\varphi)} \leq 1 \qquad para \;\; \max(\varphi) > \min(\varphi) \tag{1}$$

Si el ítem o la variable permiten respuestas en el intervalo (a,b), cualquier valor promedio, μ, puede ser acotado, siguiendo (1), mediante la estrategia:

$$\mu^{a} = \frac{\mu - a}{b - a} \qquad para \;\; b > a$$

En el caso de la desviación tipo, pueden seguirse dos criterios: la especificación de un máximo absoluto (qué valor máximo es posible obtener para la variación, considerando los extremos a y b) o éste en consonancia con el valor promedio obtenido (qué valor máximo es posible obtener para la variación, considerando los extremos a y b y un promedio de valor M).

Dado que el valor de la varianza es tanto mayor cuanto mayores son también las distancias a la media, su máximo se encontrará cuando todos los datos muestren únicamente como valores a o b. Si se cuenta con n datos para la variable X (de media M y varianza D_1^2), de los que k datos muestran el valor a, entonces:

$$M = \frac{\sum x_i}{n} = \frac{ka + (n-k)b}{n} = k\frac{a-b}{n} + b \tag{2}$$

y, por tanto:

$$D_1^2 = \frac{\sum x_i^2}{n} - M^2 = \frac{ka^2 + (n-k)b^2}{n} - \left[k\frac{a-b}{n} + b\right]^2 =$$

$$= \frac{k}{n}(a-b)^2\left(1 - \frac{k}{n}\right) \tag{3}$$

El valor mínimo para D_1^2 ocurrirá cuando las distancias sean nulas, es decir, cuando no exista variación y, por tanto, $D_1^2 = 0$. El valor máximo tendrá lugar cuando su derivada parcial con respecto a k se anule, es decir:

$$\frac{\partial D_1^2}{\partial k} = \frac{(a-b)^2}{n}\left[1 - \frac{2k}{n}\right] \;\Rightarrow\; \frac{2k}{n} = 1 \;\Rightarrow\; k = \frac{n}{2} \tag{4}$$

Así, sustituyendo (4) en (3):

$$D_1^2 = \frac{(a-b)}{4} \;\Rightarrow\; D_1 = \frac{b-a}{2} \tag{5}$$

No obstante, para n impar, las soluciones más cercanas son

$k = (n+1)/2$ y $k = (n-1)/2$

En ambos casos y partiendo de (3), se obtiene:

$$D_1^2 = \frac{(a-b)\ (n^2-1)}{4n^2} \quad \Rightarrow \quad D_1 = \frac{b-a}{2}\sqrt{1-\frac{1}{n^2}} \tag{6}$$

Las expresiones (5) y (6) pueden unificarse mediante la estrategia:

$$D_1 = \frac{b-a}{2}\delta \qquad \textit{donde}\ \ \delta = \begin{cases} \sqrt{1-\dfrac{1}{n^2}}, & \textit{para n impar} \\ 1, & \textit{para n par} \end{cases}$$

Sin embargo, las expresiones coinciden rápidamente con el aumento del valor para n. Así, para $n = 9$, $\delta = 0{,}994$; por lo que, en la práctica, la distinción entre valores pares o impares de n es intrascendente.

Con ello, siguiendo a (5), el valor máximo que cabe esperar para la desviación tipo es la mitad de la amplitud total. De esta forma, en función de (1), un índice de variación acotado en (0,1) será:

$$\sigma_1^a = \frac{2\sigma}{b-a} \qquad \textit{para}\ \ b > a$$

Si se considera un valor M dado para el promedio, la máxima varianza (D_2^2) tendrá una cota superior en D_1^2, puesto que ésta sólo es posible cuando se cumple la condición (4), es decir, cuando los datos están repartidos mitad y mitad en ambos extremos de la amplitud total. En tal caso, $M = (a+b)/2$. Por el contrario, considerando M como dado, el máximo valor posible para D_2^2 en el recorrido de la variable en (a,b), será tal que mantendrá la cantidad suficiente de datos en los extremos (a,b) como para respetar el valor promedio en M. Así, de (2) se sigue que:

$$M = k\,\frac{a-b}{n} + b \Rightarrow k = n\,\frac{M-b}{a-b} \tag{7}$$

y sustituyendo (7) en (3):

$$D_2^2 = n\frac{M-b}{a-b}\,\frac{a^2-b^2}{n} + b^2 - M^2 = -M^2 + M(a+b) - ab = (b-M)(M-a) \tag{8}$$

Se puede seguir, además, que la expresión $M = (a+b)/2$ aplicada en (8) permite llegar igualmente a (5).

Luego, siguiendo a (1):

$$\sigma_2^a = \frac{\sigma}{\sqrt{(b-M)(M-a)}} \qquad \textit{para}\ \ b > a$$

En definitiva, contamos con tres medidas acotadas en (0,1) que pueden ser utilizadas en la construcción del diagrama de posiciones y variaciones:

$$0 \leq \mu^{a}, \sigma_{1}^{a}, \sigma_{2}^{a} \leq 1$$

La figura 3 muestra el mismo ejemplo utilizado hasta el momento, recurriendo al par μ^{a}, σ_{1}^{a}, mientras que la figura 4 muestra el resultado al acudir al par μ^{a}, σ_{2}^{a}. Por otro lado, el cuadro 4 presenta los resultados numéricos para ambos casos.

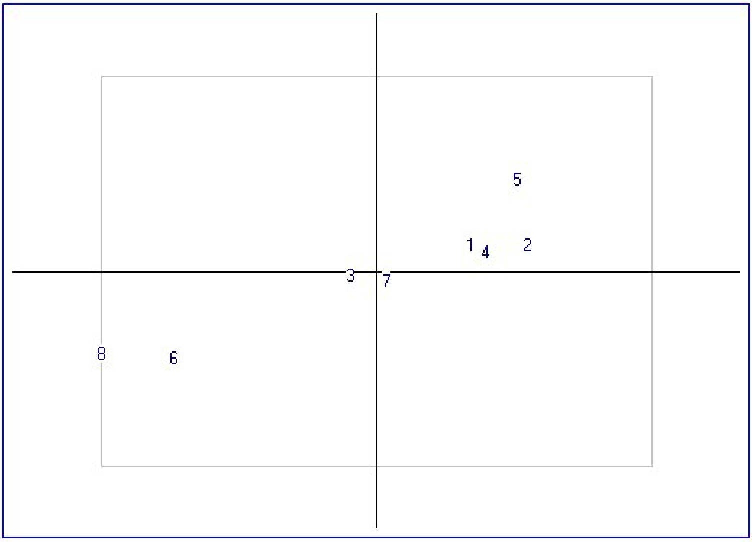

Figura 3: *diagrama para cotas máximas*.

Variable	μ^{a}	σ_{1}^{a}	σ_{2}^{a}
1	,629	,553	,572
2	,707	,553	,608
3	,464	,494	,495
4	,650	,538	,564
5	,693	,678	,735
6	,221	,332	,399
7	,514	,481	,482
8	,121	,341	,523

Cuadro 4: *valores para las medidas acotadas en (0,1).*

En términos generales, las nuevas transformaciones no implican una modificación en cuanto a la relación entre las variables, pero sí en cuanto a la apreciación de la importancia de las diferencias entre variables en posición y variación. Las figuras 3 y 4 consideran *todo* el espacio de posibles valores para los objetos representados. El punto izquierdo inferior señala la coordenada (0,0), mientras que el punto derecho superior se referiere a (1,1). Por otro lado, se ha utilizado la estrategia de remarcar un *espacio simétrico empírico* para mostrar el espacio de máximas dimensiones simétricas que es realmente utilizado. Como se puede observar, es el objeto 8 el que define las dimensiones de este rectángulo, al constituir el valor acotado más extremo (0,121;0,341).

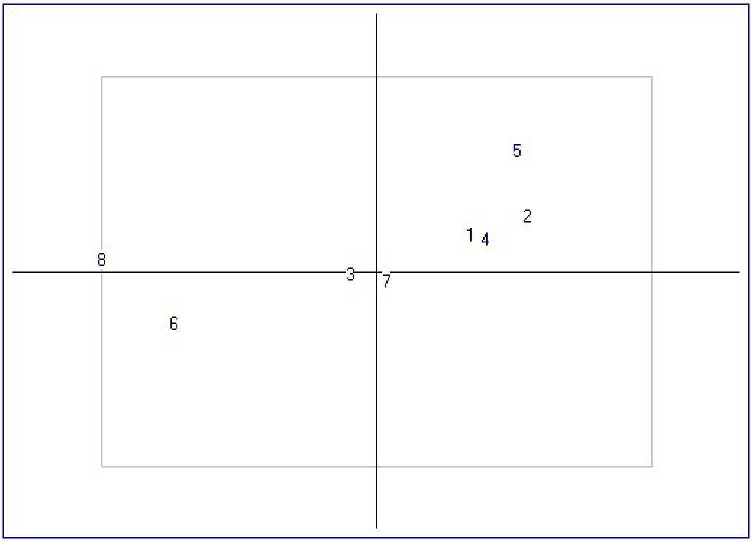

Figura 4: *diagrama para cotas de dispersión relativas a M.*

En términos generales y debido a que $D_2{}^2 \geq D_2{}^2$ entonces $\sigma_1^a \geq \sigma_2^a$. Por ello, el diagrama de cotas de dispersión relativas a la media, mostrará valores superiores para la dimensión de ordenadas y esto será tanto más así cuanto más extremo sea el promedio, puesto que la distancia entre ambos índices de variación acotada aumenta conforme el promedio se aleja del centro de la amplitud. Al respecto, es llamativo el efecto que tiene lugar con el objeto 8, donde la figura 4, que considera σ_2^a muestra una dispersión relativa de mayor cuantía que la figura 3, basada en σ_1^a.

La variación relativa al valor de la media es un recurso de interés al considerar el intervalo real de posibles concreciones para la varianza. Sin embargo, las respuestas no siguen la lógica del promedio. Las personas entrevistadas se disponen libremente a lo largo del continuo de posibles respuestas. Al igual que existe cierta coincidencia alrededor de un valor medio extremo, también podría observarse cierta coincidencia alrededor de un valor medio centrado. Ello añade incertidumbre a la hora de escoger la representación gráfica *más adecuada*. Aún así, se podría aspirar a cierta confluencia entre los recursos de las representaciones que se muestran en las figuras 3 y 4 si se considera la variación acotada sin considerar el valor del promedio, pero la representación gráfica muestra el intervalo de posibles valores para la variación. De esta forma, ambas informaciones pueden tenerse en cuenta para interpretar los resultados. La figura 5 contiene este nuevo recurso.

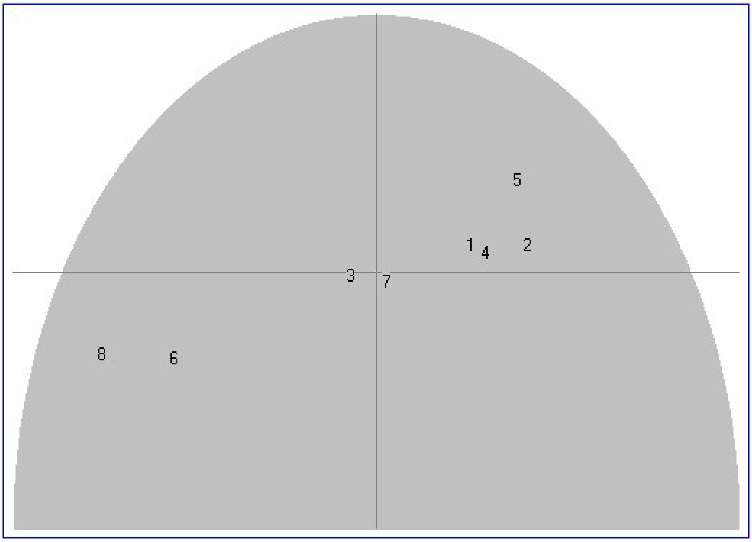

Figura 5: *diagrama para cotas de variación absolutas, sobre el área de posibles valores para la variación.*

¿Relación lineal ascendente?

Uno de los revisores de este trabajo planteaba la posibilidad de que las medidas de posición de la distribución y las de variación mostraran una relación proporcional positiva. En tal circunstancia, las representaciones gráficas deberían mostrar una tendencia lineal ascendente en cualquier caso. Esto no es necesariamente así.

Hemos visto que la desviación tipo está acotada no sólo por el recorrido de la variable sino también por el valor concreto de la media aritmética. Pero esta circunstancia no sugiere una relación lineal ascendente, sino curvilínea, como bien se muestra en la figura 5. El límite superior del área sombreada marca el máximo valor para la desviación tipo que, es muy claro en la figura, sigue una relación curvilínea con máximo en el promedio central.

Sin embargo, pensando en términos psicológicos, no algebraicos, cabe sospechar que sea más fácil encontrar relaciones lineales entre ambos estadísticos.

En un estudio, donde se entrevistó a 150 personas, se les planteaba la tarea de manifestar su predilección por diferentes destinos turísticos, a la vez que debían estimar la predilección del resto de la población. Se recurrió a cuatro destinos: playa, montaña, nieve y ciudad. Para cada uno de estos destinos, el entrevistado disponía de un continuo donde debía situar su posición y la estimación promedio del resto. La figura 6 muestra los resultados, dispuestos numéricamente en el cuadro 5. Se puede observar con facilidad que la relación es, en este caso, lineal descendente: conforme mayor es la predilección media por un destino turístico, menor es la dispersión. Tal circunstancia no difiere de la situación que hemos ejemplificado a lo largo de este trabajo, puesto que en función de cómo se articulen los pares preferencia/posición y dispersión/acuerdo, la representación gráfica puede sufrir cambios artificiales de dirección, como pasar de un sentido ascendente a otro descendente.

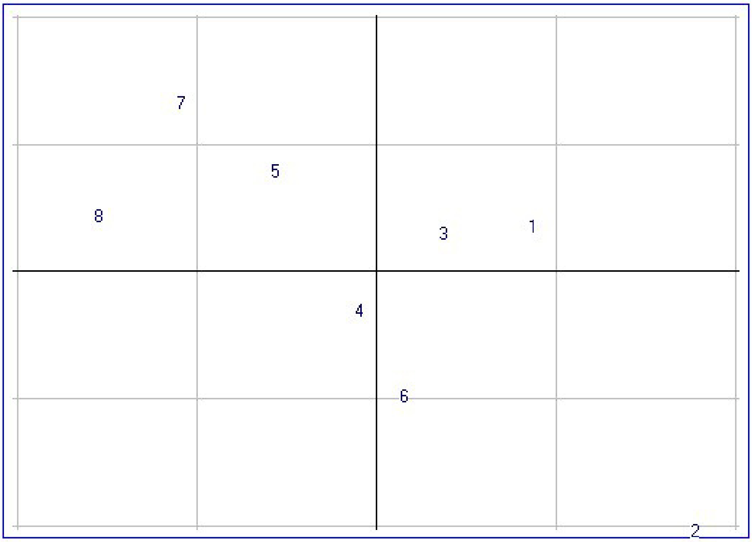

Figura 6: *diagrama estandarizado de cotas constantes (cuadriculado a Z=1)*.

Variable	Media	Desviación Tipo	
1	7.471	2.513	playa propio
2	8.855	1.636	playa demás
3	6.699	2.491	montaña propio
4	5.990	2.272	montaña demás
5	5.261	2.670	nieve propio
6	6.370	2.026	nieve demás
7	4.460	2.867	ciudad propio
8	3.761	2.543	ciudad demás

Cuadro 5: *preferencias por destinos turísticos*.

En otro estudio, donde se depuraba un cuestionario compuesto por 32 ítems, los resultados mostraban, como se observa en las figuras 7 y 8. La figura 7 señala la existencia de una disposición difusa en torno al centro estandarizado (con una muy ligera apariencia lineal ascendente), que permite identificar los ítems especiales en cuanto a su dispersión y acuerdo promedio (ítems tipo Likert de 5 puntos). En ella puede observarse, por ejemplo, la elevada dispersión del ítem 27 (falta de coincidencia en las respuestas de los entrevistados), que podría justificar su promedio central; la alta coincidencia en los ítems 30, 23 y 2, el alto nivel de desacuerdo con el enunciado en los ítems 25 y 30 o el alto nivel de acuerdo promedio con el enunciado de los ítems 29 ó 15. Esta figura muestra los ítems en un espacio donde cada cuadrícula representa una desviación tipo.

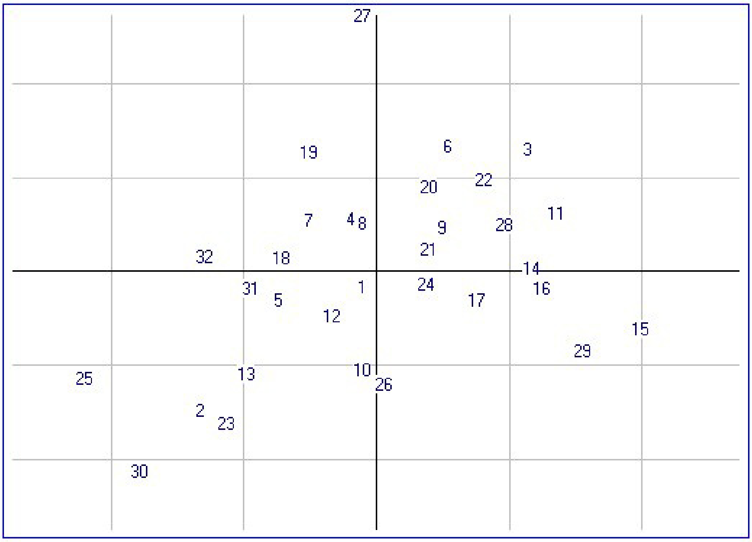

Figura 7: *diagrama estandarizado de cotas constantes (cuadriculado a Z=1).*

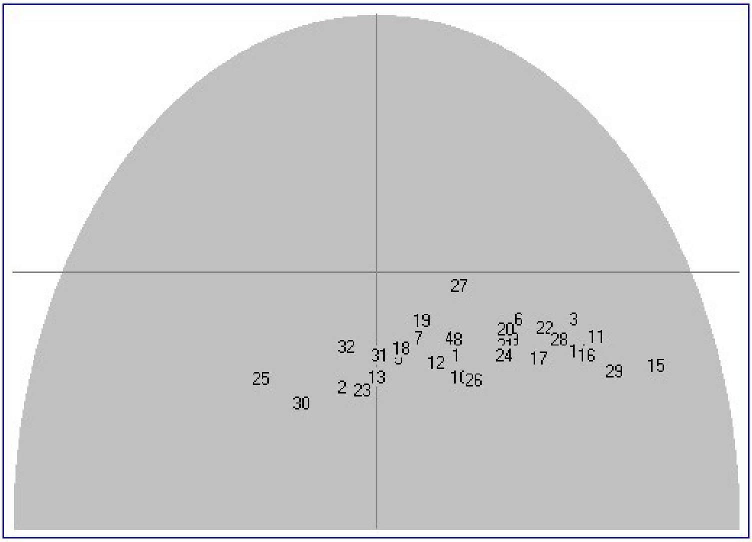

Figura 8: diagrama de medidas acotadas en (0,1).

Pero la figura 8, que recurre a la recodificación en el intervalo (0,1) muestra cómo los ítems se encuentran aglutinados en la zona de baja dispersión (no son elementos muy discriminativos entre sujetos) y con un nivel de acuerdo más bien alto que podría hacer sospechar de un control insuficiente de la aquiescencia.

Conclusiones

Los diagramas de posiciones y variaciones transformadas, especialmente si se mantiene una cota constante para los extremos de ambas dimensiones, permiten obtener *impresiones* sobre las variables del estudio que son más difíciles de concretar mediante otros recursos. Por un lado, se establecen conclusiones que se refieren a cada dimensión por separado, contemplando la interpretación de las variables más extremas. Por otro lado, se posibilita el establecimiento de conclusiones en términos de la relación que puede llegar a ocurrir entre la medida de tendencia o posición de la distribución y la medida de dispersión que se hayan considerado. Y, por último, facilita el establecimiento de impresiones asentadas sobre la configuración del conjunto de las variables, observando, como ha ocurrido en el primer ejemplo expuesto, agrupaciones entre variables similares en sus medidas de variación y posición.

Sabemos que las conclusiones generales sobre el comportamiento promedio de una variable sólo son aceptables si se cuenta con una dispersión moderada o reducida, lo que, en términos gráficos se corresponde con los cuadrantes inferiores. Pero la cuantía de la dispersión es en sí también una característica de gran interés. La existencia de una amplia variación (cuadrantes superiores) muestra la posibilidad de agrupar los casos o individuos, de identificar ítems discriminativos. Así, en los resultados mostrados con motivo de las preferencias por los ocho objetos, el diagrama aconseja realizar un estudio específico para explicar las variaciones observadas con respecto al objeto 5. En estudios de mercado, por ejemplo, tales resultados abren la puerta para el establecimiento de segmentaciones.

En definitiva, pues, los diagramas que se han presentado en este trabajo, a la vez que las medidas acotadas en las que se asientan, constituyen una herramienta muy sencilla y simple con amplias posibilidades para la interpretación de las variables o ítems de un estudio, localizada en esa importante e imprescindible etapa inicial de los análisis donde se pretende conocer los datos, familiarizarse con ellos y comenzar a ensayar conclusiones.

Referencias

Chatfield (1985) The initial examination of data. *Journal of the Royal Statistical Society*, 148, 214-253.

Freixa, M.; Salafranca, L.; Guàrdia, J.; Ferrer, R. y Turbany, J. (1992) *Análisis exploratorio de datos: nuevas técnicas estadísticas*. Barcelona: PPU.

Rao, C.R. (1994) *Estadística y verdad*. Barcelona: PPU.

Rial, A.; Varela, J. y Rojas, A. (2001) *Depuración y análisis de datos con SPSS*. Madrid: Ra-Ma.

Solanas, A.; Batista, J.M.; De Jover, Ll.; López, A.; Núñez, M.I.; Ocaña, J.; Salafranca, Ll.; San Martín, R.; Simó, J.; Valle, B. y Visauta, B. (2001) ¿Existe algo más fácil que un análisis estadístico?. *Metodología de Encuestas*, 3 (2) 227-250.