



ARTÍCULOS

Elipsoides Confidenciales para Comparar dos Estructuras de Proporciones

Fernando Ferrero

Revista de Economía y Estadística, Tercera Época, Vol. 20, No. 1-2-3 (1976): 1º, 2º, 3º y 4º Trimestre, pp. 31-47.

<http://revistas.unc.edu.ar/index.php/REyE/article/view/3718>



La Revista de Economía y Estadística, se edita desde el año 1939. Es una publicación semestral del Instituto de Economía y Finanzas (IEF), Facultad de Ciencias Económicas, Universidad Nacional de Córdoba, Av. Valparaíso s/n, Ciudad Universitaria. X5000HRV, Córdoba, Argentina.

Teléfono: 00 - 54 - 351 - 4437300 interno 253.

Contacto: rev_eco_estad@eco.unc.edu.ar

Dirección web <http://revistas.unc.edu.ar/index.php/REyE/index>

Cómo citar este documento:

Ferrero, F. (1976). Elipsoides Confidenciales para Comparar dos Estructuras de Proporciones. *Revista de Economía y Estadística*, Tercera Época, Vol. 20, No. 1-2-3 : 1º, 2º, 3º y 4º Trimestre, pp. 31-47.

Disponible en: <http://revistas.unc.edu.ar/index.php/REyE/article/view/3718>

El Portal de Revistas de la Universidad Nacional de Córdoba es un espacio destinado a la difusión de las investigaciones realizadas por los miembros de la Universidad y a los contenidos académicos y culturales desarrollados en las revistas electrónicas de la Universidad Nacional de Córdoba. Considerando que la Ciencia es un recurso público, es que la Universidad ofrece a toda la comunidad, el acceso libre de su producción científica, académica y cultural.

<http://revistas.unc.edu.ar/index.php/index>

ELIPSOIDES CONFIDENCIALES PARA COMPARAR DOS ESTRUCTURAS DE PROPORCIONES

FERNANDO FERRERO

En la investigación cuantitativa se presenta a menudo la necesidad de comparar dos estructuras de proporciones a objeto de determinar la significatividad de las diferencias observadas, referidas generalmente a momentos diferentes. La población económicamente activa desocupada, clasificada por sector de actividad, puede presentar cambios entre dos períodos de referencia y es importante, por tanto, disponer de una herramienta estadística que permita discernir acerca de la significatividad de tales cambios. En igual sentido, la clasificación de un segmento de la población por estrato socio-económico puede variar entre dos períodos considerados, por lo que resultará de natural interés indagar en qué medida tales modificaciones obedecen a fluctuaciones puramente muestrales o reflejan —en cambio— la presencia de factores que están produciendo diferencias efectivas. Otro tanto ocurre en ciertos problemas demográficos, donde a veces resulta de interés comparar dos estructuras etarias, separadas en el tiempo o en el espacio.

Para el tratamiento de estos problemas el análisis estadístico provee algunos instrumentos que permiten docimar hipótesis relativas a las magnitudes de las proporciones (pruebas de la bondad de ajustamiento), al igual que otras dócimas no paramétricas, las que a menudo resultan de fácil aplicación e incluso de mejor interpretación.

Sin embargo, dentro del conjunto de pruebas citadas, la mayoría de ellas confluye en el uso de la estadística de Chi-Cuadrado y resulta curioso observar que en la práctica ésta se expresa como una función directamente proporcional al tamaño de la muestra (número de observaciones), con el sorpresivo resultado de que a medida que aumenta el tamaño de la muestra, mayor es la probabilidad de rechazar la hipótesis nula, cualquiera sea el verdadero estado de la naturaleza, existiendo incluso la certeza de que para un cierto número de observaciones la hipótesis nula será rechazada independientemente de su veracidad. Tampoco existe unicidad en lo que respecta a la dócima uniformemente más potente, ya que diferentes métodos de agrupamientos de las categorías conducirán a diferen-

tes funciones de potencia para la décima. Además, y como corolario de la dificultad señalada en primer término, un resultado no significativo en la prueba puede obedecer exclusivamente al tamaño reducido de la muestra, cosa que si bien ocurre con otras pruebas, en la de Chi-Cuadrado se da con mayor dependencia.

En este trabajo se desarrolla un método paramétrico, basado en la teoría normal y fundado también en el uso de estadísticas conjuntamente suficientes. El método propuesto es una extensión natural a k -dimensiones de lo que comúnmente se conoce en Estadística como intervalo de confianza para la diferencia de proporciones. Como lógica consecuencia, la t de Student es reemplazada por la T^2 de Hottelling, al tiempo que los intervalos (espacios unidimensionales) se transforman en elipsoides k -dimensionales. Con todo, la idea directriz de este trabajo consiste en pasar de los vectores estimados a vectores diferencia, $d = P_t - P_{t+1}$, cuyas componentes son funciones de estimadores máximo verosímiles y están también basadas en estadísticas conjuntamente suficientes; razón por la cual dicho vector gozará de las mismas propiedades en cuanto a eficiencia, máxima potencia o, lo que es equivalente, mayor aptitud discriminatoria para detectar cambios significativos entre los períodos de comparación cuando aquéllos efectivamente han sucedido.

El método aquí desarrollado permite no sólo establecer la presencia de cambios significativos en el conjunto, sino que también permite identificar las componentes responsables de tales cambios, sin alterar el nivel de significación de la décima.

La ventaja más notoria de introducir el vector diferencia estriba en la posibilidad de derivar un método de cálculo tal que, cualquiera sea la dimensión del espacio original o el número de categorías que se consideren, la decisión final reposa en la solución de ecuaciones de segundo grado y en las particularidades algebraicas de sus raíces. La mera comparación de las mismas provee así un camino rápido y sencillo para establecer la significatividad de los cambios en las estructuras de las proporciones. Obviamente, se llegaría al mismo resultado planteando el problema como una décima de hipótesis y comparando el valor observado del estadístico con el que surge de una transformación de la T^2 de Hottelling. No obstante, la construcción de una región confidencial posee una ventaja importante al permitir la identificación simultánea de las categorías (componentes del vector diferencia) que son responsables de los cambios significativos operados entre los períodos. En otros términos, la región confidencial no sólo provee un método para detectar la presencia de modificaciones significativas en el conjunto sino que permite además individualizar cuáles son las categorías que de modo efectivo contribuyen a la variación total.

1. Distribución asintótica del vector \hat{P}_t .

Una muestra de tamaño n , extraída de una población constituida por $k+1$ categorías se representa a través de la función de cuantía de tipo multinomial, es decir,

$$g(n_1, n_2, \dots, n_k; p_1, p_2, \dots, p_k) = \frac{n!}{\pi n_1!} \pi p_1^{k+1} \dots \pi p_k^{n_i} \quad (1)$$

donde n_i representa el total de casos que en la muestra pertenecen a la i -ésima categoría. A objeto de estimar (p_1, p_2, \dots, p_k) , las componentes (n_1, n_2, \dots, n_k) , son conjuntamente suficientes. El vector de estimadores máximo verosímiles $(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k)$, donde $\hat{p}_i = n_i/n$, ($i = 1, 2, \dots, k$) está pues basado en estadísticas suficientes.

En muestras grandes se demostrará que dicho vector de estimadores máximo verosímiles se distribuye con función de densidad correspondiente a una normal multivariada, con media (p_1, p_2, \dots, p_k) y matriz de Varianza-Covarianza igual a $n^{-1} \cdot V$, siendo:

$$V = \begin{bmatrix} p_1(1-p_1) & -p_1 p_2 & -p_1 p_3 & \dots & -p_1 p_k \\ -p_2 p_1 & p_2(1-p_2) & & & -p_2 p_k \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ -p_k p_1 & -p_k p_2 & & & -p_k(1-p_k) \end{bmatrix}$$

La demostración no es otra cosa que una generalización a k dimensiones del Teorema de De Moivre-Laplace, por lo que sus detalles se dan en el apéndice que acompaña este trabajo.

Cando $n \rightarrow \infty$, la función de cuantía expresada en (1) se aproxima a una normal multivariada del tipo:

$$g(n_1, n_2, \dots, n_k; p_1, p_2, \dots, p_k) \sim [n/(2\pi^k \pi n p_i)^{k+1}]^{1/2}$$

$$\exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} + \sum_{i=1}^k \frac{(n_i - np_i)}{np_{k+1}} \right]^2 \right\} \quad (3)$$

$$\sim \frac{1}{(2\pi)^{k/2}} \frac{n^{-k/2}}{(\pi n p_i)^{1/2}} \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} + \sum_{i=1}^k \frac{(n_i - np_i)}{np_{k+1}} \right]^2 \right\} \quad (4)$$

efectuando la transformación:

$$n_i = \hat{np}_i$$

el Jacobiano resulta igual a $|J| = n^k$, de donde la anterior expresión (3) se transforma en:

$$f(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k; p_1, p_2, \dots, p_k) = \frac{1}{(2\pi)^{k/2}} \cdot \frac{n^{k/2}}{(\pi p_i)^{k/2}}$$

$$\exp \left\{ -\frac{1}{2} n \sum_{i=1}^k \frac{(\hat{p}_i - p_i)^2}{p_i} + \left[n \sum_{i=1}^k \frac{(\hat{p}_i - p_i)^2}{p_{k+1}} \right] \right\} \quad (5)$$

A su vez en el exponente de (5),

$$\sum_{i=1}^k \frac{(\hat{p}_i - p_i)^2}{p_i} + \frac{[\sum_{i=1}^k (\hat{p}_i - p_i)]^2}{p_{k+1}} = \frac{\sum_{i=1}^k (\hat{p}_i - p_i)^2}{p_i} + \frac{\sum_{i=1}^k (\hat{p}_i - p_i)^2}{p_{k+1}} + \frac{\sum_{i \neq j} (\hat{p}_i - p_i) (\hat{p}_j - p_j)}{p_{k+1}}$$

$$= \sum_{i \neq j} \frac{1}{p_i} + \frac{1}{p_{k+1}} (\hat{p}_i - p_i)^2 + \sum_{i \neq j} \frac{1}{p_{k+1}} (\hat{p}_i - p_i) (\hat{p}_j - p_j) = (\hat{P} - P)' R (\hat{P} - P)$$

Siendo:

$$\hat{P} - P = \begin{bmatrix} \hat{p}_1 - p_1 \\ \hat{p}_2 - p_2 \\ \vdots \\ \hat{p}_k - p_k \end{bmatrix} \quad \text{y} \quad R = \begin{bmatrix} p_1^{-1} + p_{k+1}^{-1} & p_{k+1}^{-1} & \dots & p_{k+1}^{-1} \\ p_{k+1}^{-1} & p_2^{-1} + p_{k+1}^{-1} & \dots & p_{k+1}^{-1} \\ \vdots & \vdots & \dots & \vdots \\ p_{k+1}^{-1} & p_{k+1}^{-1} & \dots & p_k^{-1} + p_{k+1}^{-1} \end{bmatrix}$$

y puesto que $|n R|^{1/2} = n^{k/2} (1/\pi p_i)^{k/2}$:

$$f(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k; p_1, p_2, \dots, p_k) = \frac{1}{(2\pi)^{k/2}}$$

$$\cdot / nR^{1/2} \cdot e^{-(n/2) (\hat{P} - P)' R (\hat{P} - P)} \quad (6)$$

También se puede comprobar que:

$$(nR)^{-1} = n^{-1} \cdot R^{-1} = n^{-1} \cdot V = n^{-1} \cdot \begin{pmatrix} p_1(1-p_1); & -p_1 p_2; & \dots & -p_1 p_k \\ -p_2 p_1; & p_2(1-p_2); & \dots & -p_2 p_k \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ -p_k p_1; & -p_k p_2; & \dots & p_k(1-p_k) \end{pmatrix}$$

Con lo cual queda establecida la distribución asintótica del vector

$$(\hat{p}_1; \hat{p}_2, \dots, \hat{p}_k), \text{ es decir } \hat{P} \sim NM (P; n^{-1} \cdot V)$$

Sobre la base de la distribución asintótica del vector P , Pearson propuso la estadística:

$$v = \sum (n_i - np_{0i})^2 / np_{0i} \quad (7)$$

la cual para muestras grandes se distribuye aproximadamente como una Chi-Cuadrado con k grados de libertad (si es que en la población hay $k+1$ categorías). Ello deriva de la propia distribución del exponente de (6), ya que si $\hat{P} \sim NM (P; n^{-1} \cdot V)$, la forma cuadrática $(\hat{P} - P)' (n^{-1} \cdot V) (\hat{P} - P)$ se distribuye Chi-Cuadrado con un número de grados de libertad igual al rango de la matriz V .

Nótese que (7) se puede expresar también como

$$v = \sum n(p_i - p_{0i})^2 / p_{0i}$$

pudiéndose entonces apreciar que el valor de la estadística propuesta es directamente proporcional al número de observaciones, n .

Igualmente la dócima de la razón de verosimilitud, L , y más precisamente su transformación $-2 \ln L$, posee una distribución asintótica del tipo Chi-Cuadrado con k grados de libertad, dado que el exponente de la normal multivariada (6) se puede expresar del siguiente modo:

$$\sum_{i=1}^k \frac{n(\hat{p}_i - p_i)^2}{p_i} + n \frac{[\sum (\hat{p}_i - p_i)]^2}{p_{k+1}} = \sum_{i=1}^k \frac{n(\hat{p}_i - p_i)^2}{p_i} + n \frac{[1 - \hat{p}_{k+1} - (1 - p_{k+1})]^2}{p_{k+1}}$$

$$\sum_{i=1}^k \frac{n(\hat{p}_i - p_i)^2}{p_i} + n \frac{(\hat{p}_{k+1} - p_{k+1})^2}{p_{k+1}} = \sum_{i=1}^{k+1} \frac{n(\hat{p}_i - p_i)^2}{p_i} \quad (8)$$

Tanto en uno como en otro caso cabe advertir que el test o estadístico depende del tamaño de la muestra, valor que será tanto mayor cuanto más grande sea el número de observaciones. De ahí, pues, las advertencias formuladas en la introducción de este trabajo.

2. Adaptación de la estadística T^2 .

Es sabido que si un vector $\bar{X} \sim NM(\mu; V/n)$, la forma cuadrática:

$$(\bar{X} - \mu)' n S^{-1} (\bar{X} - \mu) \sim T^2$$

con $(n-1)$ y $(n-k)$ grados de libertad, siendo S^{-1} tal que $E(S) = V$.

A su vez $T^2 \sim \frac{(n-1)k}{n-k} F(k; (n-K))$, donde k es el rango de V .

Considérese el vector \hat{P}_t con distribución $NM(P_t; V_1 \cdot n_1^{-1})$ y el vector \hat{P}_{t+1} con distribución $NM(P_{t+1}; V_2 \cdot n_2^{-1})$, (estocásticamente independiente del anterior). Defínase el vector diferencia $\hat{d} = \hat{P}_t - \hat{P}_{t+1}$, cuya distribución será:

$$NM(P_t - P_{t+1}; n_1^{-1} \cdot V_1 + n_2^{-1} \cdot V_2)$$

razón por la cual:

$$[\hat{P}_t - \hat{P}_{t+1} - (P_t - P_{t+1})]' \cdot S^{-1} \cdot [(\hat{P}_t - \hat{P}_{t+1}) - (P_t - P_{t+1})] \quad (9)$$

se distribuirá como una T^2 de $n_1 + n_2 - 2$ y $n-k$ grados de libertad y donde

$$S = (n_1 - 1)^{-1} \cdot \hat{V}_1 + (n_2 - 1)^{-1} \cdot \hat{V}_2$$

$$E(S) = (n_1 - 1)^{-1} \cdot E(\hat{V}_1) + (n_2 - 1)^{-1} \cdot E(\hat{V}_2) = (n_1 - 1)^{-1} \cdot (n_1 - 1)V_1/n_1 + (n_2 - 1)^{-1} \cdot (n_2 - 1) \cdot V_2/n_2 = n_1^{-1} \cdot V_1 + n_2^{-1} \cdot V_2$$

siendo:

$$\hat{V}_1 = \begin{bmatrix} \hat{p}_{1t}(1 - \hat{p}_{1t}) & -\hat{p}_{1t} \hat{p}_{2t} & \dots & \hat{p}_{1t} \hat{p}_{kt} \\ -\hat{p}_{2t} \hat{p}_{1t} & \hat{p}_{2t}(1 - \hat{p}_{2t}) & \dots & -\hat{p}_{2t} \hat{p}_{kt} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \hat{p}_{kt}(1 - \hat{p}_{kt}) \end{bmatrix} \quad (10)$$

$$\hat{V}_2 = \begin{pmatrix} \hat{p}_{1, t+1} (1 - \hat{p}_{1, t+1}) & ; & -\hat{p}_{1, t+1} \cdot \hat{p}_{2, t+1} & \dots & \dots & -\hat{p}_{1, t+1} \cdot \hat{p}_{k, t+1} \\ -\hat{p}_{2, t+1} \cdot \hat{p}_{1, t+1} & ; & \hat{p}_{2, t+1} (1 - \hat{p}_{2, t+1}) & \dots & \dots & -\hat{p}_{2, t+1} \cdot \hat{p}_{k, t+1} \\ & & \cdot & & & \cdot \\ \hat{p}_{k, t+1} \cdot \hat{p}_{1, t+1} & ; & \cdot & & \dots & \hat{p}_{k, t+1} (1 - \hat{p}_{k, t+1}) \end{pmatrix} \quad (11)$$

$(n_1 - 1) \qquad \qquad \qquad (n_2 - 1)$

tal que: $E(\hat{V}_1) = \frac{(n_1 - 1)}{n_1} V_1$; $E(\hat{V}_2) = \frac{(n_2 - 1)}{n_2} V_2$

3. Construcción del elipsoide confidencial

La construcción del elipsoide confidencial está basada en dos muestras independientes obtenidas de dos poblaciones con distribución normal multivariada de parámetros $(P_t; n_1^{-1} \cdot V_1)$ y $(P_{t+1}; n_2^{-1} \cdot V_2)$, respectivamente.

En la muestra de tamaño n_1 , se calcula el vector \hat{P}_t , tal que $\hat{P}'_t = (\hat{p}_{1t}, \hat{p}_{2t}, \dots, \hat{p}_{kt})$, y en la muestra de tamaño n_2 se calcula \hat{P}_{t+1} tal que $\hat{P}'_{t+1} = (\hat{p}_{1, t+1}; \hat{p}_{2, t+1}; \dots, \hat{p}_{k, t+1})$.

Para determinar la región se define el vector diferencia, esto es, $\hat{d} = \hat{P}_t - \hat{P}_{t+1}$, y $d = P_t - P_{t+1}$, con lo que (9) se convierte:

$$[(\hat{P}_t - \hat{P}_{t+1}) - (P_t - P_{t+1})]' \cdot S^{-1} \cdot [(\hat{P}_t - \hat{P}_{t+1}) - (P_t - P_{t+1})] =$$

$$= (\hat{d} - d)' \cdot S^{-1} \cdot (\hat{d} - d) \quad (12)$$

puesto que $(\hat{d} - d)' \cdot S^{-1} \cdot (\hat{d} - d) \sim T^2$ y esto a su vez se distribuye $(n-1)k / (n-k) \cdot F_{\alpha}(k; n_1 + n_2 - k)$ es decir:

$$\frac{(n_1 + n_2 - 2) k}{(n_1 + n_2 - k)} F(k; (n_1 + n_2 - k)),$$

una región confidencial de nivel $1 - \alpha$, puede hallarse determinando:

$$P [(d - \hat{d})' S^{-1} (d - \hat{d}) \leq T^2_{\alpha}] = \int_0^{T^2_{\alpha}} g(T^2) d T^2 = 1 - \alpha \quad (13)$$

y realizando la transformación

$$T^2 = \frac{(n_1 + n_2 - 2)k}{(n_1 + n_2 - k)} \cdot F(k, n_1 + n_2 - 2)$$

$$1 - \alpha = \int_0^{T^2} g(T^2) dT^2 = \frac{(n_1 + n_2 - 2)k}{(n_1 + n_2 - k)} \int_0^{F_\alpha} F(k, n_1 + n_2 - 2) dF(k, n_1 + n_2 - 2) \quad (14)$$

por lo que el elipsoide confidencial quedará determinado por la desigualdad:

$$(\hat{d} - d_0)' S^{-1} (\hat{d} - d_0) < \frac{(n_1 + n_2 - 2)}{(n_1 + n_2 - k)} k F_\alpha(k, n_1 + n_2 - k) \quad (15)$$

la que a su vez genera un elipsoide en el espacio k-dimensional en el cual el vector nulo, $d_0 = 0$, $[0 = (0, 0, \dots, 0)]$, puede o no pertenecer al sólido. En caso afirmativo, si d_0 , representado en el espacio por el punto donde se cortan ortogonalmente los ejes del sistema, pertenece al sólido generado por la desigualdad (15), habrá que concluir que al nivel confidencial especificado no hay diferencias significativas entre ambas estructuras de porcentajes. Es claro que se trata de una condición suficiente pero no necesaria dado que bien puede ocurrir que el vector nulo no pertenezca al sólido sin que ello resulte en la presencia de cambios significativos en las estructuras de porcentajes.

La Fig. 1 representa el caso en el que el elipsoide contiene el punto d_0 , es decir, cuando no hay cambios significativos; la Fig. 2, por el contrario, exhibe una situación diferente en la cual d_0 no pertenece al sólido y donde efectivamente hay cambios significativos en las estructuras de proporciones (semi-espacio positivo de d_2 y negativo de d_1). De todas maneras, el d_0 no pertenece al sólido, la aplicación del método de cálculo desarrollado en el apartado 5) permitirá determinar en definitiva si existen o no cambios significativos en las estructuras de proporciones.

4. Método de cálculo

Para determinar si el punto d_0 está contenido en el elipsoide bastará con asignar valores nulos a todas las componentes del vector excepto una (d_1 , por ejemplo), en cuyo caso y reemplazando en (15) por los valores conocidos del vector $\hat{d}' = (\hat{d}_1, \hat{d}_2, \dots, \hat{d}_k)$ e introduciendo $((n_1 + n_2 - 2) / (n_1 + n_2 - k)) k F(k, n_1 + n_2 - 2)$ en el primer miembro, la desigualdad (15) se puede transformar en una ecuación de segundo grado en la variable d_1 . En esta ecuación de segundo grado las raíces pueden ser:

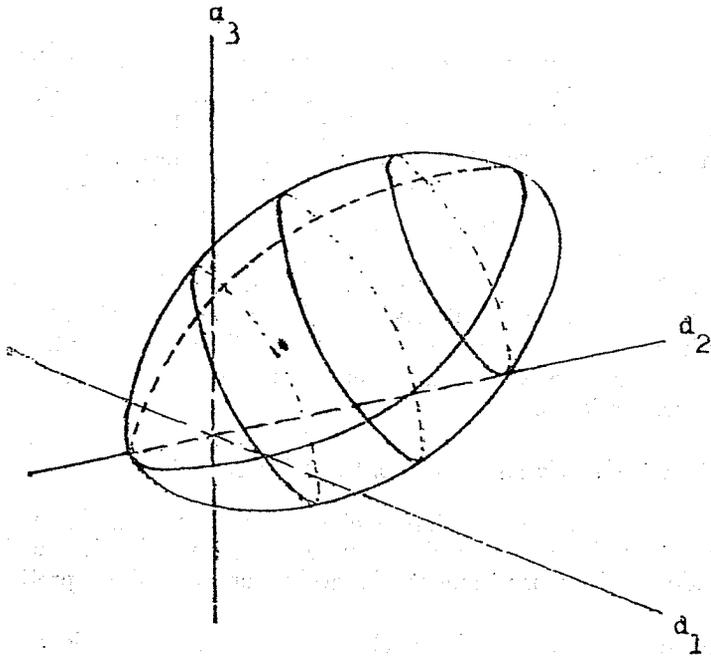


FIG. 1

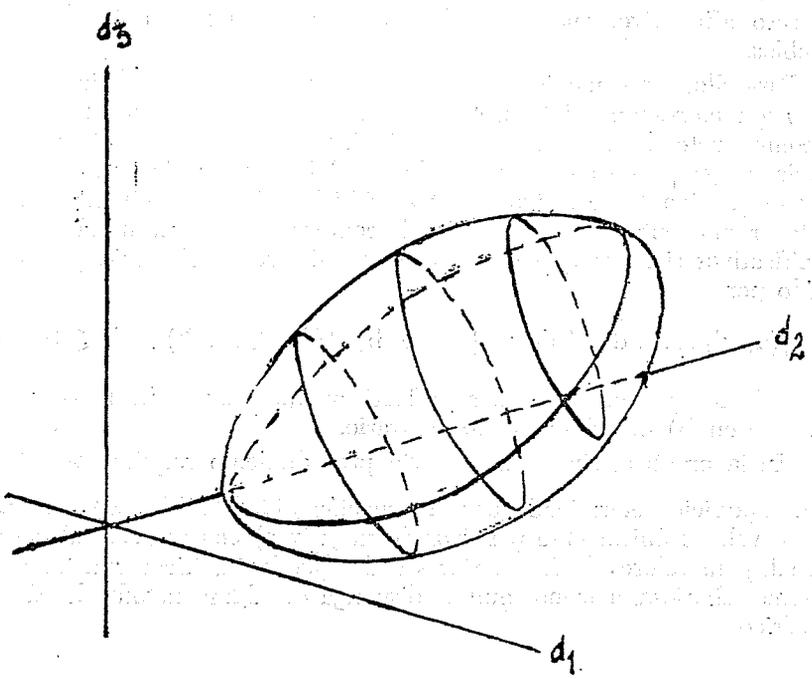


FIG. 2

a) reales y de signos opuestos, b) imaginarias, c) reales y del mismo signo, d) reales e iguales. En a) se da el caso de que d_0 pertenece obviamente al sólido y en d) el punto d_0 pertenece a la superficie del elipsoide en cuyo caso, y puesto que en (15) se establece una desigualdad en sentido estricto, los puntos ubicados en la superficie no pertenecen a la región confidencial, por lo que (d) lleva a las mismas conclusiones que (b) y (c). En estas últimas el punto d_0 no está contenido en el elipsoide. Dado que (a) es sólo una condición suficiente, si se dieran cualquiera de los casos apuntados en (b), (c) o (d) será todavía necesario aplicar las transformaciones que se describen en el punto siguiente para determinar si existen o no diferencias significativas en las estructuras de proporciones.

5. Significatividad de las componentes del vector d

Desde el punto de vista inferencial la determinación de una región confidencial conduce a intervalos simultáneos para todos los parámetros involucrados. Tal es precisamente la razón por la cual se prefirió este método.

Si conforme a lo señalado en el punto anterior se dan algunas de las situaciones descritas en (b), (c), o (d), habrá entonces que establecer si la estructura de proporciones presenta o no cambios significativos y, en caso afirmativo, cuáles son las componentes responsables de tales cambios.

Para ello, y por medio de una simple generalización a k dimensiones de lo que corresponde interpretar en el caso de un intervalo confidencial, la componente d_i exhibirá cambios significativos si es que el elipsoide queda íntegramente comprendido dentro del semi-espacio relativo al intervalo positivo (o negativo) relativo a dicho eje. En otros términos, y sin tener en cuenta los restantes ejes, la componente d_i presentará cambios significativos siempre que el elipsoide se ubique en el semi-espacio definido por

$$a) \{(d_1, d_2, \dots, d_k) / d_i \geq 0\} \quad \text{o} \quad b) \{(d_1, d_2, \dots, d_k) / d_i \leq 0\}$$

En a) la componente P_{ti} será significativamente mayor que $P_{t+1,i}$, mientras que en b) ocurrirá todo lo contrario.

En la práctica, sin embargo, este procedimiento requiere establecer si la superficie determinada por la ecuación $(\hat{d} - d)' S^{-1} (\hat{d} - d) - T_{\alpha}^2 = 0$ asume valores únicamente positivos (o negativos) con respecto al eje d_i , lo cual, para valores de k superiores a dos puede significar una tarea en extremo laboriosa, a menos que se disponga de algún método de cálculo numérico.

Tal dificultad puede ser obviada mediante el procedimiento de trasladar el elipsoide al punto $(\hat{d}_1, \hat{d}_2, \dots, \hat{d}_k)$ y efectuar allí una rotación de los ejes de forma tal que el sólido queda ubicado en su posición canónica. En ésta, la superficie corta al eje i -ésimo en los puntos representados por los vectores

$$\begin{aligned} z'_i &= (0, 0, \dots, a_i, 0, \dots, 0) & y \\ -z'_i &= (0, 0, \dots, -a_i, 0, \dots, 0) \end{aligned}$$

donde $a_i = (T_\alpha^2 / \lambda_i)^{1/2}$, siendo λ_i la i -ésima raíz característica de S^{-1} .

Puesto que los vectores así obtenidos son producto de una transformación ortogonal aplicada a los originales, la utilización de la matriz inversa de dicha transformación permitirá pasar de los ejes transformados a los originales, estableciéndose así una correspondencia biúnívoca entre ambos conjuntos. No obstante, la determinación de la posición del elipsoide en los ejes transformados se puede efectuar con toda facilidad una vez calculados los valores propios de la matriz S^{-1} .

Algebraicamente, de la desigualdad expresada en (15)

$$(\hat{d} - d)' S^{-1} (\hat{d} - d) \leq T_\alpha^2 \quad \text{o} \quad (\hat{d} - d)' n S^{-1} (\hat{d} - d) \leq T_\alpha^2$$

según que S haya sido calculada previamente dividiendo o no por n , se efectúan las siguientes transformaciones

$$h = (n)^{1/2} \cdot (\hat{d} - d) \quad y \quad (16)$$

$P' h = Z$, donde $P' S^{-1} P = D$ (D es una matriz diagonal) (17) con lo cual la forma cuadrática de (15) se transforma en $Z' D Z \leq T_\alpha^2$, quedando definida una superficie cuya ecuación es $Z' D Z - T_\alpha^2 = 0$ (18) y que corta a los ejes transformados en valores que sólo dependen de T_α^2 y de los valores propios de S^{-1} . De estos últimos se vuelve a los ejes originales revirtiendo las etapas, esto es, para determinar la posición del elipsoide con respecto al sistema de ejes representado por \hat{d} , bastan las siguientes operaciones:

a) Para la raíz característica λ_i se calculan dos vectores $h = P Z_i$ para $Z_i = (0, 0, \dots, a_i, 0, \dots, 0)$ y $-Z_i = (0, 0, \dots, -a_i, 0, \dots, 0)$

b) Se determinan dos vectores $\underline{d} = \hat{d} - h (n^{-1})^{1/2}$ (19)

c) Las operaciones anteriores se repiten para las restantes raíces características de S^{-1} , formándose una matriz A que consta de k filas y $2k$ columnas.

Las componentes de cada uno de los vectores columnas de A son los cosenos directores de los vectores que representan (en el sistema original) los puntos en que la superficie corta a los ejes transformados (cada eje corta dos veces a la superficie). En la matriz A , por tanto, una fila donde todas las componentes son no negativas implicará lógicamente que el elipsoide se ubica en el semi-espacio no negativo de dicho eje y viceversa. Si, por el contrario, las componentes de una fila alternan en signos, ello importará que el elipsoide no queda íntegramente comprendido dentro de un semi-espacio. Tanto en el primero como en el segundo caso habrá que concluir que la respectiva componente presenta cambios significativos (positivos en el primer caso, negativos en el segundo).

6. Ejemplos

a) Se considerará en primer lugar un caso en el cual ninguna componente exhibe cambios significativos. Sea, por ejemplo, una población que consta de tres categorías y de la que se obtienen dos muestras independientes de tamaño 100 con los siguientes resultados

\hat{P}_t	\hat{P}_{t+1}	\hat{d}
0.40	0.36	0.04
0.30	0.34	-0.04
0.30	0.30	0.00

A partir de estos resultados se obtienen

$$\hat{V}_1 = \begin{bmatrix} 0.24 & -0.12 \\ -0.12 & 0.21 \end{bmatrix} \quad y \quad \hat{V}_{t+1} = \begin{bmatrix} 0.2304 & -0.1224 \\ -0.1224 & 0.2244 \end{bmatrix}$$

$$S = (1/99) (V_t + V_{t+1}) = \begin{bmatrix} 0.00475 & -0.00245 \\ -0.00245 & 0.00439 \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} 295,6226 & 164,9830 \\ 164,9830 & 319,8650 \end{bmatrix} \quad y \quad S_{t+1}^{-1} = \begin{bmatrix} 2,95623 & 1,64983 \\ 1,64983 & 3,19865 \end{bmatrix}$$

Al nivel del 95 %

$$T_{0.95}^2 = ((n_t + n_{t+1})^{-2} / 2 / (n_t + n_{t+1})^{-2}) F_{0.95} (2, 198) = 6.06$$

De forma tal que la región confidencial (elipse) para el nivel escogido quedará determinada por la desigualdad

$$((0.04 - d_1), (-0.04 - d_2)) \begin{bmatrix} 295,6226 & 164,9830 \\ 164,9830 & 319,8650 \end{bmatrix} \begin{pmatrix} (0.04 - d_1) \\ (-0.04 - d_2) \end{pmatrix} \angle 6.06$$

en tanto que la superficie se representará por medio de la ecuación

$$(0.04 - d_1)^2 295,6226 + 319,8650 (-0.04 - d_2)^2 - 329,966 (0.04 - d_1)(-0.04 - d_2) - 6.06 = 0$$

Las raíces características de S_1^{-1} son 4,7317 y 1,4232, la matriz modal

$$P = \begin{bmatrix} 0.68071 & -0.73255 \\ 0.73255 & 0.68071 \end{bmatrix}$$

la semilongitud del eje mayor es 2.05 y la del menor 1,13. Premultiplicando los vectores

$$\begin{bmatrix} 0,00 \\ 2,05 \end{bmatrix}, \begin{bmatrix} 0,00 \\ -2,05 \end{bmatrix}, \begin{bmatrix} 1,13 \\ 0,00 \end{bmatrix} \text{ y } \begin{bmatrix} -1,13 \\ 0,00 \end{bmatrix}$$

por la matriz P se obtienen los valores de la matriz A,

$$A = \begin{bmatrix} 0.19 & -0.11 & -0.037 & 0.1173 \\ -0.18 & 0.10 & -0.1232 & -0.0432 \end{bmatrix}$$

Tal como era de esperar tanto las filas relativas a d_1 al igual que las correspondientes a d_2 presentan componentes que alternan en signos, no hallándose por consiguiente ningún cambio significativo en el conjunto representado por ambos ejes.

Nótese de paso que, conforme a lo señalado en el punto 4), el vector nulo pertenece en efecto al sólido, ya que al anular d_2 las raíces de la ecuación resultante son, respectivamente 0.156 y -0.121 , es decir, reales y de signos opuestos.

b) Es lógico pensar que, aun cuando una diferencia de 0.04 en las proporciones no sea significativa en una muestra de tamaño 100, puede sin embargo serlo en muestras de tamaños mayores. A tal efecto, con los mismos datos anteriores se considerarán los resultados bajo el supuesto de que los mismos han sido obtenidos a partir de dos muestras independientes de tamaño 1500. Los valores de S_1^{-1} , P y las raíces características son los mismos que en el ejemplo anterior, pero cambian los resultados de la transformación $h = (n)^{1/2} (\hat{d} - d)$, con lo que la nueva matriz A queda conformada del siguiente modo:

$$A = \begin{bmatrix} 0.0203 & 0.05967 & 0.0011 & 0.0784 \\ -0.0186 & -0.06140 & -0.0757 & -0.0757 \end{bmatrix}$$

O sea que en este caso todos los valores correspondientes a la primera fila de A son positivos, indicativo ello de que han existido cambios significativos (positivos) entre ambos periodos en la primera componente; en tanto que la segunda fila de A arroja valores negativos en todas las columnas, por lo que los cambios resultantes son también significativos (negativos) para la segunda componente.

c) Se considerará otro caso en el cual la magnitud de las diferencias en una componente es lo suficientemente pronunciada como para registrar la presencia de cambios significativos aún en muestra de tamaño 100, mientras que la segunda componente no presentará cambio alguno. Supóngase los siguientes resultados

\hat{P}_t	\hat{P}_{t+1}	\hat{d}
0.40	0.20	0.20
0.30	0.30	0.00
0.30	0.50	0.20

Con tales valores, y sobre la base de dos muestras independientes de tamaño 100, se obtiene la matriz

$$S_1^{-1} = \begin{bmatrix} 3,42779 & 2,05894 \\ 2,05894 & 4,57039 \end{bmatrix}$$

$$\lambda_1 = 6.1357 \quad \text{y} \quad \lambda_2 = 1,8625$$

$$P = \begin{bmatrix} 0.60734 & -0.79607 \\ 0.79607 & 0.60734 \end{bmatrix}$$

siendo los correspondientes valores de la matriz A

$$A = \begin{bmatrix} 0.056 & 0.2144 & 0.1396 & 0.2604 \\ -0.1100 & 0.1100 & 0.0792 & -0.0792 \end{bmatrix}$$

de donde se aprecia que sólo la primera componente registra cambios significativos (positivos); en tanto que la segunda —como era de esperar— presenta coeficientes con signos opuestos.

d) Por último, conviene tener en cuenta un caso en el cual hay cambios significativos de distintos signos en dos componentes y ninguno en la tercera. Sean v.gr.: dos muestras independientes de tamaño 101, con los siguientes resultados:

ELIPSOIDES CONFIDENCIALES PARA COMPARAR DOS ESTRUCTURAS

P_t	P_{t+1}	d
0.40	0.20	0.20
0.20	0.40	-0.20
0.15	0.10	0.05
0.25	0.30	-0.05

$$\hat{V}_t = \begin{bmatrix} 0.24 & -0.08 & -0.06 \\ -0.08 & 0.16 & -0.03 \\ -0.06 & -0.03 & 0.1275 \end{bmatrix} \quad \hat{V}_{t+1} = \begin{bmatrix} 0.16 & -0.08 & -0.02 \\ -0.08 & 0.24 & -0.04 \\ -0.02 & -0.04 & 0.09 \end{bmatrix}$$

$$S_t = \begin{bmatrix} 0.40 & -0.16 & -0.08 \\ -0.16 & 0.40 & -0.07 \\ -0.08 & -0.07 & 0.2175 \end{bmatrix}$$

$$S_t^{-1} = \begin{bmatrix} 3,58201 & 1,76264 & 1,88477 \\ 1,76264 & 3,51657 & 1,78006 \\ 1,88477 & 1,78006 & 5,86373 \end{bmatrix} \quad \begin{array}{l} \lambda_1 = 8.19482 \\ \lambda_2 = 2.98228 \\ \lambda_3 = 1,78519 \end{array}$$

$$P = \begin{bmatrix} 0.48122 & 0.51097 & -0.74338 \\ 0.46424 & 0.54072 & 0.66256 \\ 0.74358 & -0.66286 & 0.09167 \end{bmatrix}$$

$$T^2_{.95} = ((n_t + n_{t+1} - 2) 3 / (n_t + n_{t+1} - 3)) F_{.95}(3, 197) \\ = 7.84$$

Semilongitud de los ejes principales,

$$Z_1 ; (7.84/8.19482)^{1/2} = 0.98$$

$$Z_2 ; (7.84/2.98223)^{1/2} = 1.62$$

$$Z_3 ; (7.84/1.78519)^{1/2} = 2.096$$

Premultiplicando los vectores

$$\begin{bmatrix} 0.978 \\ 0.000 \\ 0.000 \end{bmatrix} \begin{bmatrix} -0.978 \\ 0.000 \\ 0.000 \end{bmatrix} \begin{bmatrix} 0.000 \\ 1.620 \\ 0.000 \end{bmatrix} \begin{bmatrix} 0.000 \\ -1.620 \\ 0.000 \end{bmatrix} \begin{bmatrix} 0.000 \\ 0.000 \\ 2.096 \end{bmatrix} \begin{bmatrix} 0.000 \\ 0.000 \\ -2.096 \end{bmatrix}$$

por la matriz P y realizando las transformaciones,

$$h = Pz$$

$$d = h (n^{-1})^{1/2} - \hat{d}$$

se obtienen los coeficientes de la matriz A,

$$A = \begin{bmatrix} 0.1529 & 0.2471 & 0.1170 & 0.2830 & 0.3560 & 0.0440 \\ -0.2454 & -0.1546 & -0.2870 & -0.1124 & -0.3380 & -0.0661 \\ -0.0227 & 0.1227 & 0.1580 & -0.0580 & 0.0310 & 0.0690 \end{bmatrix}$$

Nuevamente se puede advertir que las componentes relativas a d_1 son todas positivas; las referidas a d_2 son todas negativas; mientras que las componentes que corresponden a d_3 (que no registra ningún cambio significativo) son de signos opuestos.

APENDICE

La función de cuantía correspondiente a un vector con distribución multinomial de $k+1$ parámetros ($k+1$ categorías) se expresa:

$$g(n_1, n_2, \dots, n_k; p_1, p_2, \dots, p_k) = (n! / \pi n_i!) \pi p_i^{n_i} \quad (A.1)$$

Aproximando el primer factor de (A.1) por medio de la fórmula de Stirling,

$$n! / \pi n_i! = (n^n e^{-n} n^{1/2} 2\pi^{1/2}) / \pi n_i^{n_i} e^{-n_i} (n_i 2\pi)^{1/2}$$

$$= (1/2\pi)^{-k/2} n^n (n/\pi n_i)^{1/2} = (1/2\pi)^{-k/2} \pi n_i^{n_i} (n/\pi n_i)^{1/2} (1/\pi n_i n_i)$$

de modo que (A.1) se puede aproximar por medio de

$$g(n_1, n_2, \dots, n_k; p_1, p_2, \dots, p_k) = (1/2\pi)^{-k/2} (n/\pi n_i)^{1/2} \pi ((np_i)/n_i)^{n_i} \quad (A.2)$$

Si se considera el último factor de (A.2), con la sustitución

$$u_i = n_i - np_i \quad (i = 1, 2, \dots, k+1), \text{ es decir, } n_i = u_i + np_i, \quad (A.3)$$

$$\pi ((np_i)/(u_i + np_i))^{n_i} = \pi (1/(1 + u_i/np_i))^{n_i}, \text{ pasando a ln}$$

$$\ln \pi (1/(1 + u_i/np_i))^{n_i} = \ln \pi [(1/(1 + u_i/np_i))^{(u_i + np_i)}]$$

$$= -\sum^{k+1} (u_i + np_i) \ln (1 + u_i/np_i) \quad (A.4)$$

$$= -\sum^{k+1} (u_i + np_i) [(u_i/np_i) - (u_i/np_i)^2/2 + (u_i/np_i)^3/3 \dots]$$

resultado que se obtiene al desarrollar por Taylor $\ln(1+u_i/np_i)$

Efectuando el producto indicado y teniendo en cuenta que $u_i/n \rightarrow 0$ y $u_i^3/n^2 \rightarrow 0$ cuando $n \rightarrow \infty$, o sea reteniendo únicamente los términos dominantes, lo anterior es aproximadamente igual a

$$\sum^{k+1} [(u_i^2/np_i) + u_i + (-u_i^2/2np_i) \dots] = \sum^{k+1} ((u_i^2/2np_i) + u_i) \quad (A.5)$$

Sin embargo, de acuerdo a la sustitución realizada

$$\sum^{k+1} u_i = \sum^k u_i + u_{k+1} = \sum^{k+1} n_i - n \sum^{k+1} p_i = n - n = 0, \text{ es decir, } u_{k+1} = \\ = \sum^k u_i, \text{ de modo}$$

$$\text{que } \sum^{k+1} ((u_i^2/2np_i) + u_i) = \sum^{k+1} (u_i^2/2np_i) + \sum^{k+1} u_i = \sum^{k+1} (u_i^2/2np_i) \\ = \sum^k (u_i^2/2np_i) + u_{k+1}^2/2np_{k+1} \\ = \sum^k (u_i^2/2np_i) + (\sum^k u_i)^2/2np_{k+1} \quad (A.6)$$

en definitiva,

$$\ln \pi (1/1 + u_i/np_i)^{n_i} = -(\sum^k (u_i^2/2np_i) + (\sum^k u_i)^2/2np_{k+1})$$

de donde

$$\pi (1/1 + u_i/np_i)^{n_i} = \exp [-(1/2) (\sum^k (u_i^2/2np_i) + (\sum^k u_i)^2/2np_{k+1})] \quad (A.7)$$

Por otra parte, el segundo factor de (A.2) tiende a

$$(n/\pi np_i)^{1/2} \text{ cuando } n \rightarrow \infty, \text{ ya que } n_i \sim np_i \quad (A.8)$$

Con estos resultados (A.2) se convierte en

$$(1/2\pi)^{-k/2} (n/\pi np_i)^{1/2} \exp \{ -(1/2) [\sum^k (n_i - np_i)^2/np_i + \\ + (\sum^k n_i - np_i)^2/np_{k+1}] \} = \quad (A.9)$$

$$(1/2\pi)^{-k/2} (1/n^k \pi p_i)^{1/2} \exp [-(1/2) (U' (n^{-1}R) U)]$$

pero como $n^{-1}R/ = n^{-k} \pi (1/p_i)$, (A.9) puede ser expresada del

siguiente modo

$(1/2\pi)^{-k/2} /n^{-k} \pi^{1/2} \exp (-(1/2) (U' (n^{-1}R) U)$ que es obviamente la función de densidad de una normal multivariada. Nótese que de u_i se puede pasar a $n_i - np_i$ sin necesidad de ningún factor adicional ya que el Jacobiano de la transformación es unitario.