

UN ANÁLISIS DE SENSIBILIDAD DE ENCUESTAS DE OPINIÓN ANTE LA VARIACIÓN DE UN FACTOR

Santos del Cerro, Jesús*

jsantos@jur-to.uclm.es

Facultad de Derecho y Ciencias Sociales

Universidad de Castilla-La Mancha

Palabras clave: encuestas de opinión, coeficientes de sensibilidad, ANOVA.

RESUMEN

Las encuestas de opinión constituyen un instrumento que ha ido adquiriendo un creciente grado de implantación en ámbitos sociales, económicos, académicos, etc. Son frecuentes las encuestas realizadas sobre opiniones políticas acerca de múltiples temas, sobre intenciones de voto, sobre opiniones de expectativas empresariales, sobre evaluación de la calidad docente del profesorado universitario, etc. Los objetivos que se plantean al llevar a cabo un estudio muestral de este tipo se resumen en el de obtener información de determinadas cuestiones cuyo conocimiento es utilizado en la toma de decisiones o actuaciones en entornos específicos. En este trabajo se pretende medir distintos aspectos de la incidencia de un cambio en un determinado factor o circunstancias sobre la opinión mantenida por los encuestados acerca de una realidad concreta. Existen técnicas tradicionales de análisis tales como el Análisis de la Varianza que resultan insuficientes en lo relativo a ciertas cuestiones analíticas de interés. Por este motivo hemos propuesto y discutido unas medidas originales orientadas a medir aspectos tales como la intensidad de la influencia del cambio del factor, la redistribución de los encuestados en la expresión de su opinión ante dicho cambio, etc.

* El autor es profesor del Área de Estadística de la UCLM y además es Asesor Estadístico de la Oficina de Evaluación de la Calidad de la misma universidad.

TEXTO DE LA PONENCIA:

Las encuestas de opinión constituyen un instrumento que ha ido adquiriendo un creciente grado de implantación en ámbitos sociales, económicos, académicos, etc. Son frecuentes las encuestas realizadas sobre opiniones políticas acerca de múltiples temas, sobre intenciones de voto, sobre opiniones de expectativas empresariales, sobre evaluación de la calidad docente del profesorado universitario, etc. Los objetivos que se plantean al llevar a cabo un estudio de este tipo se resumen en el de obtener información de determinadas cuestiones cuyo conocimiento es utilizado en la toma de decisiones o actuaciones en entornos específicos.

El carácter de las respuestas de este tipo de encuestas es, en general, subjetivo, siendo dichas respuestas el resultado de las concepciones y percepciones que de algún aspecto de la realidad posee la persona a la que va dirigido el cuestionario. Muchos de los fenómenos que se plantean en el análisis de encuestas no tienen acomodo bajo una escala cuantitativa. Sin embargo, a pesar de esta circunstancia se suelen establecer escalas numéricas que reflejan cierto carácter ordinal de las respuestas posibles. No cabe duda que en este proceso se pierde una gran cantidad de matices y peculiaridades que pudieran resultar de interés. Por otra parte, la percepción de las cuestiones y de las posibles respuestas no es, en general, coincidente debido a causas tales como las distintas connotaciones semánticas de los términos utilizados, sistemas de valores diferentes, etc. Sin embargo, asumiendo estos aspectos y otros que se pudieran añadir, se establecen en gran cantidad de encuestas de opinión valoraciones que se numeran del 0 al 10, del 1 al 5, etc. Precisamente, en este tipo de encuestas es al que dedicaremos el siguiente análisis.

La información obtenida, después de realizado los procesos de diseño, trabajo de campo y depuración, se presenta según una distribución de frecuencias en la que la variable toma los distintos valores posibles a la cuestión planteada y las frecuencias son el número (o proporción) de encuestados que señala cada uno de dichos valores. Habitualmente este conjunto de información se resume mediante medidas de posición, central o no central, medidas de dispersión, medidas de forma, etc., y además se le puede aplicar distintos análisis estadísticos multivariantes como regresiones, análisis factoriales, etc.

El problema que nos plantearemos en este trabajo es el de analizar la sensibilidad de las opiniones manifestadas ante un cambio ocurrido en un determinado factor, variable o circunstancia, así como proponer algunas medidas de la misma.

Estableceremos como hipótesis de trabajo que la población muestreada es la misma o de análogas características en los distintos escenarios que obedecen a un valor diferente del factor o distintas circunstancias bajo las que se realiza la encuesta. Por ejemplo, en el caso de encuestas sobre opiniones políticas ante una reforma fiscal, consideraremos que, aunque no respondan las mismas personas, es de esperar y así lo supondremos que la población encuestada tiene características similares en una y otra situación. Otro ejemplo, en el caso de las encuestas de evaluación de la calidad docente del profesorado de una universidad si se produce un cambio en un determinado factor, supondremos que los alumnos de un curso que responden no difieren en su modo de opinar respecto de la de los alumnos del siguiente.

En general, el cambio del factor viene expresado a través de una variable dummy en la que sus posibles valores son 0 ó 1, el primer valor representa la situación del factor en la primera fase o etapa y 1 representa la situación desde que se produce la transformación del factor. En estos casos resulta de interés, suponiendo que se verifiquen las hipótesis necesarias, aplicar una análisis ANOVA. Esta técnica, si bien tiene la ventaja de su sencillez y fácil interpretación presenta el inconveniente de no proporcionar una medida precisa de la intensidad con que dicho cambio ha afectado a la opinión sobre la cuestión planteada. Efectivamente, aplicando el contraste de la F podemos averiguar si el factor influye significativamente sobre la variabilidad de las respuestas a la cuestión formulada. Si rechazamos la hipótesis nula, fijado un cierto nivel de significación, de inexistencia de influencia del factor sobre la variable querrá decir que el cambio ocurrido en el factor ejerce un efecto sensible sobre la variable. Se podría, como mucho, calcular el nivel de significación crítico asociado (p-valor) que podremos utilizar como “medida” de la intensidad con la que se rechaza la hipótesis nula. Creemos, sin embargo, que esto resulta insuficiente y por tanto es preciso realizar un análisis adicional que complete al anterior. En este sentido presentaremos, de modo paralelo al desarrollo teórico, una aplicación práctica, cuya evidencia empírica puede servirnos para ilustrar el contenido conceptual de las medidas que aquí se propongan. En este

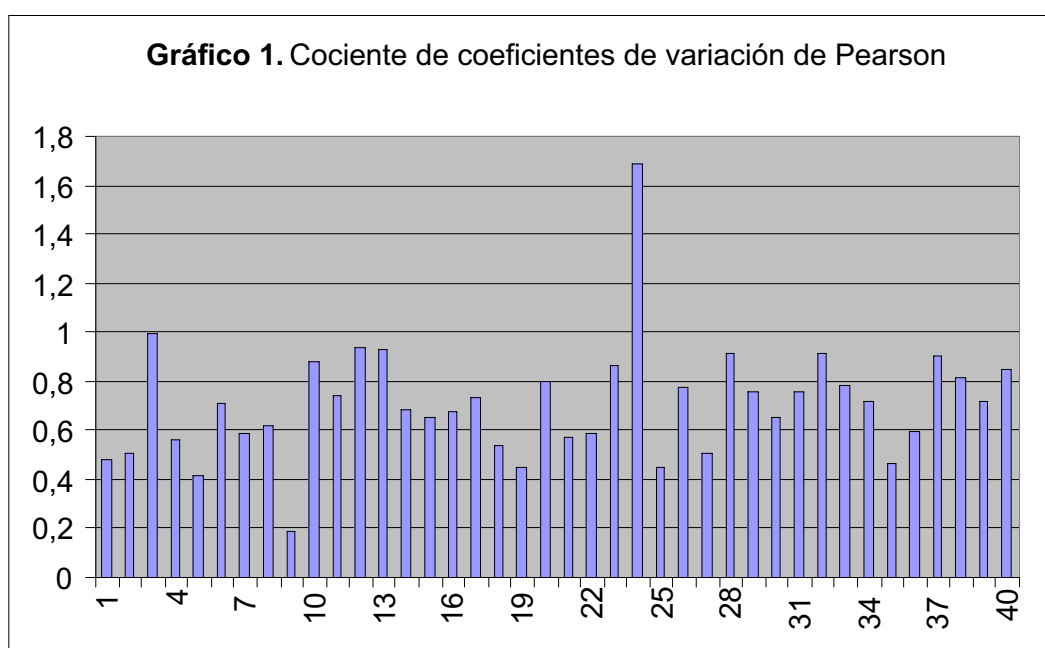
sentido, tomaremos la información relativa a la evaluación de la calidad docente del profesorado de la Universidad de Castilla-La Mancha (UCLM) durante dos años consecutivos, en los que para ciertas asignaturas en uno de esos años los alumnos habían sido examinados y en el siguiente el paso de la encuesta se realizó en el período lectivo antes de comenzar los exámenes. Los alumnos deben responder a una serie de cuestiones valorando cada una de ellas desde el valor 1 (Totalmente en desacuerdo) hasta el valor 5 (Totalmente de acuerdo). Como ya hemos destacado, los alumnos que respondieron a las encuestas de una misma asignatura y profesor difieren, en su mayor parte, de un año respecto de otro, sin embargo establecemos la hipótesis de trabajo, que creemos razonable, consistente en considerar a los alumnos de una asignatura de dos años consecutivos suficientemente homogéneos para realizar el análisis propuesto. Se han tomado las evaluaciones correspondientes a 40 profesores de distintos centros de la UCLM relativas a otras tantas asignaturas durante los cursos 1999/2000 y 2000/2001, realizándose el paso de encuesta en el primer curso después del examen final y antes en el caso del segundo. Todas ellas han sido seleccionadas siempre que el número de encuestas fuese suficientemente alto y, además, en todos ellos se rechaza la hipótesis nula de igualdad de medias en los dos cursos a un nivel de significación del 1%. Las diferencias entre las puntuaciones medias de un curso y otro se recogen en la siguiente tabla.

Tabla 1

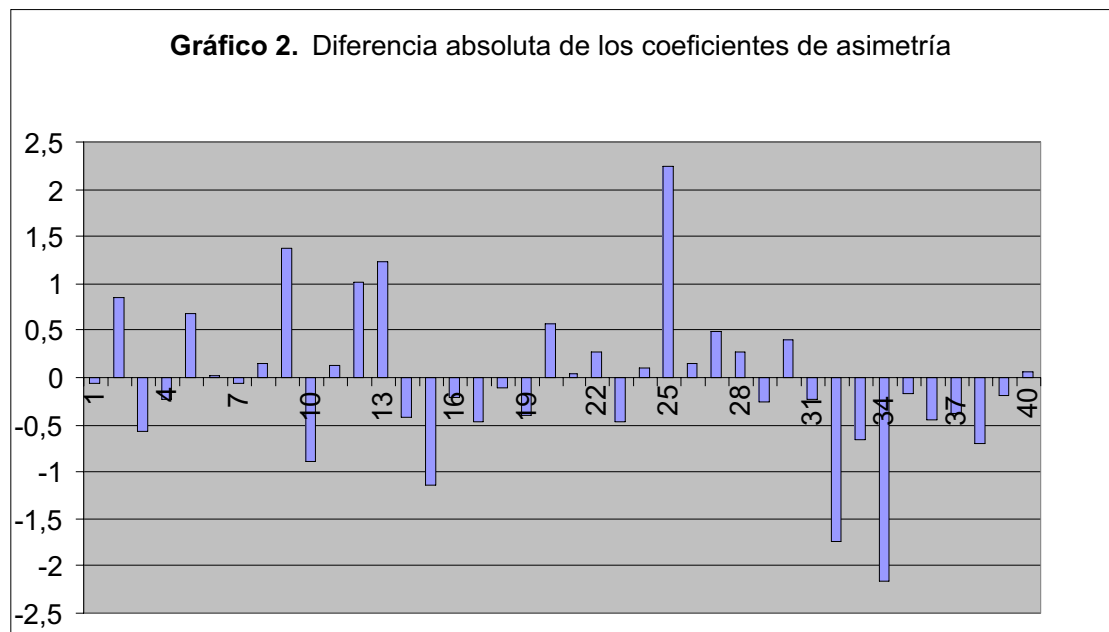
Número de profesor	Diferencia de medias	Número de profesor	Diferencia de medias
1	1.03	21	0.66
2	1.19	22	0.69
3	0.69	23	0.39
4	0.59	24	0.39
5	1.18	25	0.95
6	0.68	26	0.56
7	0.54	27	1.06
8	0.51	28	0.62
9	2.51	29	0.98
10	0.47	30	0.97
11	0.37	31	0.72
12	0.64	32	1.22
13	0.36	33	0.55
14	0.74	34	0.86
15	1.02	35	0.78
16	1	36	0.51
17	0.31	37	0.81
18	0.89	38	0.83
19	1.54	39	1.04
20	1.03	40	0.77

Aparte de lo anterior, desde un punto de vista descriptivo, resulta de interés comparar, entre las distribuciones de frecuencias relativas a ambos cursos, medidas de dispersión, asimetría y forma. Calcularemos respectivamente los coeficientes de variación de Pearson $V = \frac{S}{\bar{x}}$, el coeficiente de asimetría de R.A. Fisher $g_1 = \frac{m_3}{S^3}$ y el coeficiente de apuntamiento $g_2 = \frac{m_4}{S^4} - 3$.

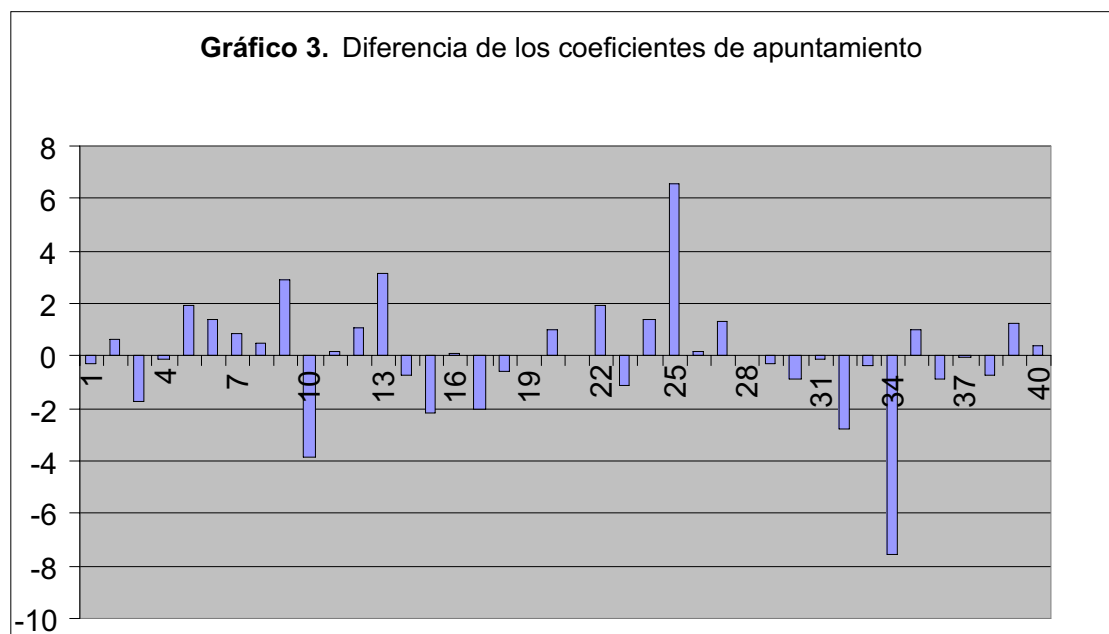
Realizando el análisis para los tres coeficientes anteriores obtenemos las siguientes conclusiones. La dispersión relativa de las distribuciones de frecuencias han descendido, salvo en un solo caso, del curso 99/00 al 00/01, como puede apreciarse en la representación gráfica del cociente $\frac{V_{00/01}}{V_{99/00}}$, que en todos los casos salvo uno es inferior a la unidad (Véase Gráfico 1).



Para comparar la variación del grado de asimetría hemos calculado $|(g_1)_{00/01}| - |(g_1)_{99/00}|$ obteniéndose un comportamiento heterogéneo en cuanto a la evolución sufrida por el carácter de simetría de las distribuciones de frecuencias ante el cambio del factor. El Gráfico 2 recoge dicha diferencia.



Algo parecido ocurre con la evolución sufrida con el coeficiente de apuntamiento. El Gráfico 3 recoge la diferencia $(g_2)_{00/01} - (g_2)_{99/00}$, obteniéndose situaciones dispares según el caso.



Proponemos, además, en este trabajo ciertas medidas originales que tratan de recoger ciertas características de las distribuciones de frecuencias no tenidas en cuenta por las medidas anteriormente vistas. Denotamos por x_i los valores de la variable, que en

general le asignaremos las cifras 1, 2, ..., k, f_i las frecuencias relativas antes del cambio y f_i' las frecuencias relativas después del cambio.

x_i	f_i	f_i'	$f_i' - f_i$
x_1	f_1	f_1'	$a_1 = f_1' - f_1$
x_2	f_2	f_2'	$a_2 = f_2' - f_2$
.	.	.	.
.	.	.	.
x_k	f_k	f_k'	$a_k = f_k' - f_k$

Obsérvese que $0 \leq \sum_{a_i \geq 0} a_i \leq 1$ y $-1 \leq \sum_{a_i \leq 0} a_i \leq 0$ y además $\left| \sum_{a_i \leq 0} a_i \right| = \sum_{a_i \geq 0} a_i$. Si definimos

$\lambda = \left| \sum_{a_i \leq 0} a_i \right| = \sum_{a_i \geq 0} a_i$ es fácil comprobar que oscila entre 0 y 1. Si $\lambda=0$ significa que la

distribución no ha sufrido ninguna transformación y si $\lambda=1$ la distribución de frecuencias queda “totalmente transformada”. Considérese el siguiente ejemplo en el que se recoge esta última situación extrema.

x_i	f_i	f_i'
1	0.1	0
2	0	0.7
3	0.5	0
4	0.4	0
5	0	0.3

Cualquier valor de λ comprendido entre 0 y 1 recogerá transformaciones en la distribución de frecuencias más o menos intensas según que dicho coeficiente se acerque más a 1 ó a 0, respectivamente.

Por otra parte, las ganancias o las pérdidas de frecuencia de los valores de la variable se redistribuyen, en general, de modo no uniforme. Podría, sin embargo, ocurrir una correspondencia biunívoca de idénticas intensidades entre ganancias y pérdidas. Por ejemplo, para $k=5$, sea $a_1=0.1$, $a_2=0$, $a_3=-0.2$, $a_4=-0.1$ y $a_5=0.2$, observamos una

correspondencia biyectiva entre ganancias y pérdidas de frecuencias relativas de la misma intensidad: $a_1 \leftrightarrow a_4$; $a_3 \leftrightarrow a_5$. Otro caso extremo sería aquel en el que las ganancias (o pérdidas) en un conjunto de valores se compensasen por la pérdida (o ganancia) de la frecuencia relativa de un solo valor. Por ejemplo, para $k=5$, sea $a_1=-0.1$, $a_2=-0.05$, $a_3=0$, $a_4=0.35$ y $a_5=-0.2$. Para el primer caso la norma euclídea del vector de ganancias es el mismo que el vector de pérdidas pues $(-0.1)^2 + (-0.2)^2 = 0.1^2 + 0.2^2$, o

lo que es lo mismo $\frac{(-0.1)^2 + (-0.2)^2}{0.1^2 + 0.2^2} = 1$, pero en el segundo caso

$$\frac{(-0.1)^2 + (-0.05)^2 + (-0.2)^2}{0.35^2} = \frac{0.0525}{0.1225} = 0.4285. \text{ A este cociente lo denominaremos } \gamma:$$

$$\gamma = \frac{\sum_{a_i < 0} a_i^2}{\sum_{a_i > 0} a_i^2} \text{ para los casos en que } \sum_{a_i < 0} a_i^2 \leq \sum_{a_i > 0} a_i^2 \text{ y}$$

$$\gamma = \frac{\sum_{a_i > 0} a_i^2}{\sum_{a_i < 0} a_i^2} \text{ para los casos en que } \sum_{a_i > 0} a_i^2 \leq \sum_{a_i < 0} a_i^2$$

Este coeficiente alcanzará el valor 1 cuando la redistribución de las frecuencias en pérdidas y ganancias sean de la misma intensidad, y tomará valores inferiores a 1 en el resto de los casos. Por otra parte, esta doble definición tiene su justificación por su comodidad interpretativa. Por tanto $0 < \gamma \leq 1$ (excluimos el caso trivial de inexistencia de transformación). No obstante el límite inferior de γ dependerá del valor de k . En concreto, la cota inferior de γ (siendo $\lambda > 0$) para $k=2, 3, \dots$ es igual a

$$\frac{(k-1) \left(\frac{\lambda}{k-1} \right)^2}{\lambda^2} = \frac{1}{k-1}$$

Si construimos una tabla para distintos valores de k tendremos las siguientes cotas inferiores de γ .

k	2	3	4	5	6	7	8	9	10
Cota inferior de γ	1	0.5	0.333	0.25	0.166	0.142	0.125	0.111	0.1

Nos encontramos con el inconveniente de tener cotas inferiores diferentes para valores distintos de k . Para tratar de solucionar este problema y proporcionar una medida homogénea desde un punto de vista interpretativo realizaremos la siguiente transformación de γ , de modo que obtengamos un coeficiente γ^* definido por

$$\gamma^* = \frac{k-1}{2-k}(\gamma - 1)$$

de manera que si

$$\gamma = 1 \Rightarrow \gamma^* = 0$$

$$\gamma = \frac{1}{k-1} \Rightarrow \gamma^* = 1$$

Este coeficiente γ^* representa una corrección de γ por el número de respuestas posibles (k), cuya interpretación será:

$\gamma^*=0$ la redistribución de pérdidas y ganancias de frecuencias relativas son de la misma intensidad.

$\gamma^*=1$ la redistribución de las frecuencias relativas es la más descompensada posible.

Cuanto más cerca esté γ^* de 0 tanto menor descompensación existirá en la redistribución y cuanto más cerca de 1 ocurrirá lo contrario.

Hasta este punto hemos definido dos coeficientes, uno de los cuales (λ) mide la intensidad total del cambio sufrido por la distribución de frecuencias relativas de una variable ante una variación de cierto factor, mientras que el segundo (γ^*) cuantifica el efecto redistributivo, en cuanto a las pérdidas y ganancias de frecuencias relativas tras el cambio, de los valores de la variable sometida a estudio.

En las tablas 2 y 3 aparecen calculados para los 40 casos el coeficiente λ y el coeficiente γ^* . La intensidad media ($\bar{\lambda}$) es igual a 0.38375, oscilando este coeficiente desde 0.17 hasta 0.881. Como sabemos su campo de variación oscila entre 0 y 1, observándose que en media el factor examen ha afectado al 38.375% de las frecuencias relativas de la distribución de la evaluación de la calidad docente de los profesores seleccionados.

Tabla 2

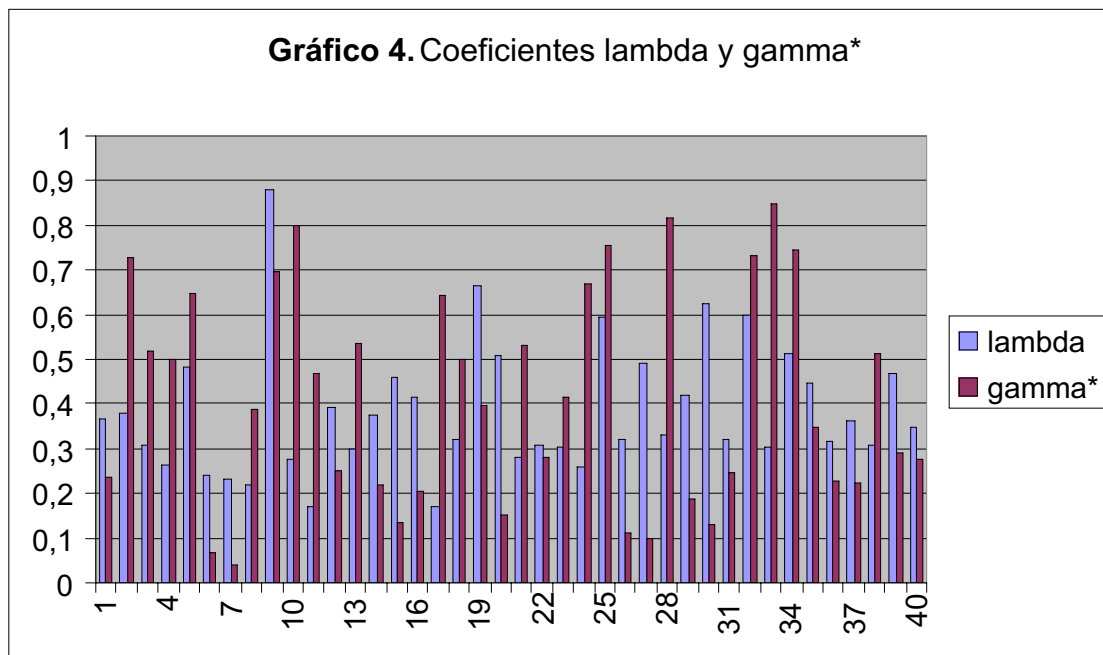
Número de profesor	λ	Número de profesor	λ
1	0.365	21	0.281
2	0.379	22	0.309
3	0.310	23	0.302
4	0.262	24	0.258
5	0.483	25	0.594
6	0.241	26	0.322
7	0.231	27	0.489
8	0.219	28	0.331
9	0.881	29	0.420
10	0.278	30	0.625
11	0.170	31	0.320
12	0.392	32	0.600
13	0.299	33	0.305
14	0.376	34	0.514
15	0.462	35	0.448
16	0.417	36	0.315
17	0.168	37	0.362
18	0.322	38	0.309
19	0.663	39	0.470
20	0.511	40	0.347

Por otra parte, la media de los valores γ^* ha resultado ser igual a 0.41434. Se puede apreciar en la Tabla 3 la existencia de valores cercanos a 1, que significa que ha habido muchos valores (o pocos) que han perdido o ganado mucha frecuencia y pocos valores (o muchos) que la han ganado o perdido, respectivamente.

Tabla 3

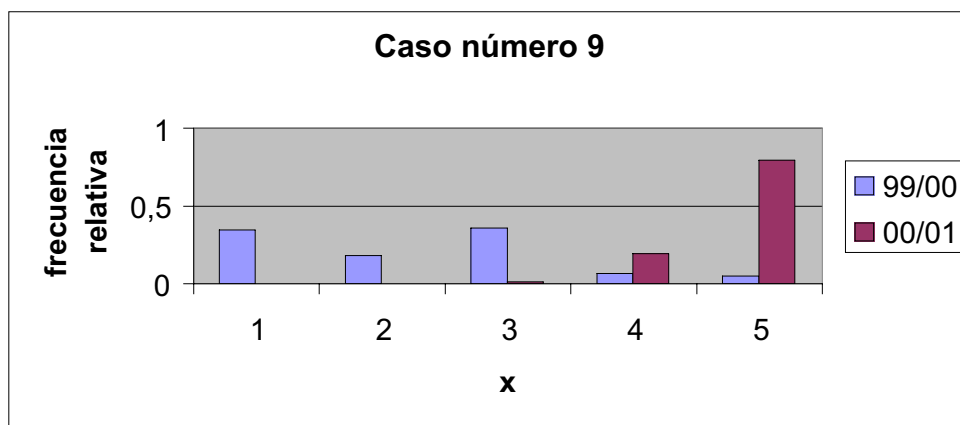
Número de profesor	γ^*	Número de profesor	γ^*
1	0.236	21	0.530
2	0.726	22	0.281
3	0.516	23	0.416
4	0.500	24	0.667
5	0.647	25	0.754
6	0.068	26	0.112
7	0.041	27	0.097
8	0.386	28	0.817
9	0.697	29	0.186
10	0.800	30	0.129
11	0.470	31	0.246
12	0.248	32	0.733
13	0.534	33	0.850
14	0.217	34	0.746
15	0.134	35	0.349
16	0.205	36	0.226
17	0.644	37	0.221
18	0.500	38	0.512
19	0.397	39	0.291
20	0.150	40	0.276

En el Gráfico 4 hemos representado los dos coeficientes, λ y γ^* , donde se observa una gran variedad de situaciones en las que algunos casos λ es mayor que γ^* y otras sucede lo contrario. En un estudio posterior se podrían agrupar los casos por Centros o Estudios, o incluso por tipo de asignatura, y analizar posibles peculiaridades respecto del comportamiento de estos coeficientes.

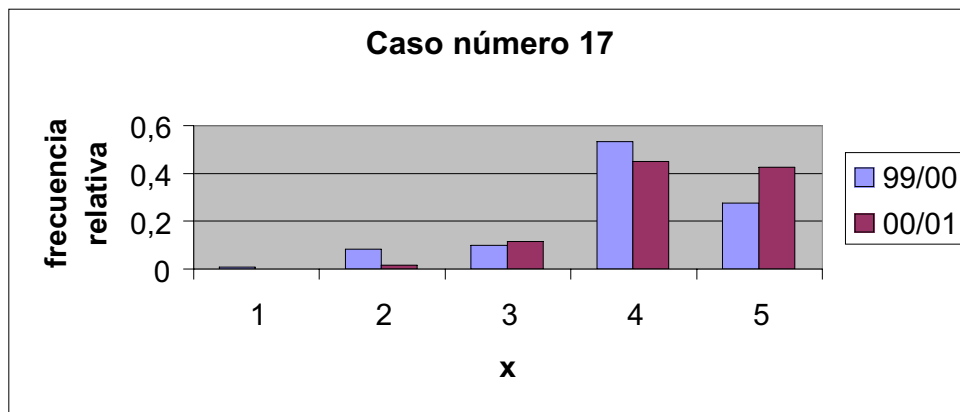


Resulta de interés representar gráficamente la distribución de frecuencias relativas en ambos cursos para los valores extremos de los coeficientes λ y γ^* .

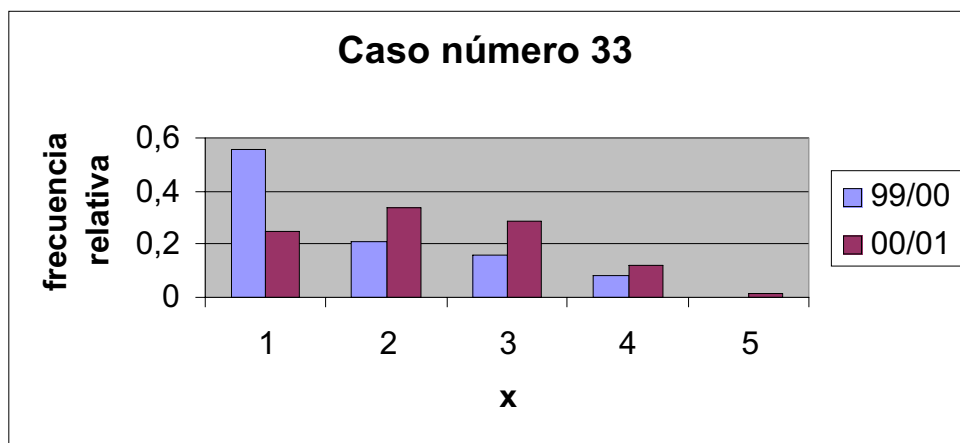
En primer lugar, obsérvese el caso en el que el coeficiente λ es el mayor de todos cuyas distribuciones de frecuencias relativas son las del siguiente gráfico.



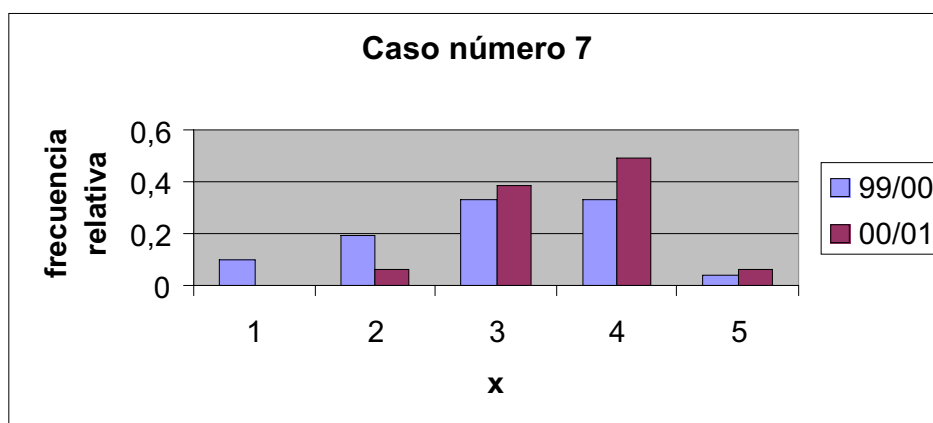
El gráfico siguiente es el correspondiente al caso con el menor valor del coeficiente λ .



El gráfico siguiente es el correspondiente al caso con el mayor valor del coeficiente γ^* .



El gráfico siguiente es el correspondiente al caso con el menor valor del coeficiente γ^* .



CONCLUSIONES

La estadística descriptiva tradicional nos proporciona ciertas herramientas para evaluar la transformación de una distribución de frecuencias ante el cambio de un determinado factor. Resulta de interés los coeficientes λ y γ^* , que miden respectivamente la intensidad del cambio y el efecto redistributivo de la frecuencia de la situación final respecto de la inicial. Como hemos comprobado, no siempre valores del primer coeficiente altos o bajos van acompañados de valores altos o bajos del segundo, como consecuencia de que dichos coeficientes miden cosas distintas y de la propia naturaleza del caso analizado. Por otra parte la acotación en su campo de variación que presentan ambos proporciona una gran comodidad interpretativa.