SUMA 28
junio 1998, pp. 103-107

Simulación de la paradoja de Stein con la hoja de cálculo

Rafael Cortés Mora

IDEAS Y RECURSOS

La hoja de cálculo constituye un potente entorno para la experimentación en clase de estadística, comparable al laboratorio en la de ciencias experimentales. Entre sus múltiples aplicaciones se encuentra la de proporcionar un medio para la comprobación experimental de resultados teóricos. Para ilustrarlo, proponemos un modelo para verificar el teorema de Stein relativo a la estimación óptima de un conjunto de k > 2 medias. El carácter paradójico de este resultado lo convierte en un ejemplo ideal para este tipo de simulaciones.

LA HOJA de cálculo constituye una poderosa herramienta didáctica para la enseñanza de la estadística. Proporciona un entorno ideal para la simulación de experiencias aleatorias, para la comprobación experimental de resultados teóricos y para la investigación.

La posibilidad de comprobar experimentalmente resultados teóricos es especialmente importante cuando no se dispone del aparato matemático requerido para su demostración rigurosa, como ocurre frecuentemente en la enseñanza secundaria. El alumno no tiene más remedio que «creer» los enunciados del profesor y hacer uso de ellos en las aplicaciones y en los ejercicios que se le proponen. En ocasiones, estos resultados chocan con sus «creencias» previas y surge una «necesidad» de comprobación. El objetivo de este trabajo es, precisamente, ejemplificar la situación que acabamos de describir. Se ha elegido un resultado poco conocido, cuya naturaleza paradójica provocará la mencionada «necesidad» de comprobación. Se trata del teorema de Stein, cuyo campo de aplicación es la construcción de estimadores óptimos para un conjunto de $k > 2$ medias.

Conviene recalcar que nuestro objetivo no es presentar una actividad para ser trabajada con los alumnos (de hecho, los contenidos desarrollados exceden el nivel de la enseñanza secundaria) sino mostrar, mediante un ejemplo, la necesidad de tratar la estadística de forma experimental y las posibilidades que ofrece para ello la hoja de cálculo.

Estimación de parámetros

El objetivo esencial de la estadística inferencial o inductiva es obtener conocimiento de grandes conjuntos de datos o poblaciones a partir de subconjuntos del mismo llamados muestras. Uno de sus problemas fundamentales lo constituye la estimación de parámetros que surge cuando la distribución de la característica estudiada depende de un parámetro ω que debe estimarse a partir de los datos obtenidos en una muestra. Para ello, se utilizan los denominados estimadores o funciones de la muestra $T(x_1,\ldots,x_n)$ cuyos valores en el muestreo deben ser próximos a ω.

Una de las propiedades exigidas a estos estimadores es que su esperanza matemática coincida con el valor del parámetro a estimar, es decir, que $E(T) = \omega$. Con ello queda garantizada la convergencia del promedio de las estimaciones hacia el verdadero valor del parámetro. Un estimador que verifica esta propiedad se dice que es *insesgado* o centrado. En caso contrario, es decir, cuando $E(T) = \omega + e(\omega)$, se dice que el estimador es *sesgado*, y a la expresión $e(\omega)$ se la denomina sesgo o *error sistemático* del estimador. Por ejemplo, la media muestral

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

es un estimador centrado de la media poblacional ($E(\overline{x}) = \mu$) mientras que la varianza muestral

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

es un estimador sesgado de la varianza poblacional σ^2 ya que

$$E(s^2) = \frac{n-1}{n}\sigma^2$$

Para corregir este sesgo, suele utilizarse el estimador

$$\hat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

el cuál verifica que.

$$E(\hat{s}^2) = \frac{n}{n-1}E(s^2) = \sigma^2$$

Para valorar la eficiencia de un estimador, suele utilizarse el *error cuadrático esperado* $E(T - \omega)^2$. Cuanto menor sea esta cantidad, más eficiente será el estimador, ya que, en promedio, proporcionará estimaciones más cercanas a ω. Cuando los estimadores son centrados el error cuadrático esperado coincide con la varianza:

$$E(T - \omega)^2 = E(T - E(T))^2 = \mathrm{var}(T)$$

Al comparar pues estimadores insesgados, se preferirá siempre al que tenga menor varianza. Si para una distribución dada existe un estimador insesgado, cuya varianza es inferior a la de cualquier otro, se dice que el estimador es el *más eficiente*. Si la distribución es normal,

El objetivo esencial de la estadística inferencial o inductiva es obtener conocimiento de grandes conjuntos de datos o poblaciones a partir de subconjuntos del mismo llamados muestras.

puede demostrarse que la media muestral $\overline{x}$ es el estimador más eficiente para la media poblacional, es decir, ninguna otra función insesgada de los datos puede estimar la media más exactamente que $\overline{x}$.

Cuando se elimina la hipótesis de estimación insesgada, los estimadores deben compararse mediante el error cuadrático esperado. Supongamos, por ejemplo, los siguientes cuatro estimadores para la media de una población normal $N(\mu, \sigma)$:

$T_1(x_1,\ldots x_n) = \overline{x}$

$T_1(x_1,\ldots x_n) = Md$

$T_1(x_1,\ldots x_n) = \overline{x}/2$

$T_1(x_1,\ldots x_n) = 0$

que son, respectivamente: la media muestral o promedio, la mediana de los datos muestrales, la mitad de la media muestral y la función constante igual a 0.

Al calcular el error cuadrático esperado para cada uno de ellos se obtiene:

$$E(T_1 - \mu)^2 = \frac{\sigma^2}{n}$$

$$E(T_2 - \mu)^2 = \frac{\pi}{2}\frac{\sigma^2}{n} \cong 1{,}57\frac{\sigma^2}{n}$$

$$E(T_3 - \mu)^2 = \frac{1}{4}\left(\frac{\sigma^2}{n} + \mu^2\right)$$

$$E(T_4 - \mu)^2 = \mu^2$$

En el gráfico 1 se representan los errores cuadráticos de los cuatro estimadores en función del parámetro μ. Puede observarse como los errores cuadráticos esperados de la media y de la mediana (estimadores 1 y 2, respectivamente) no dependen del verdadero valor de μ. Entre estos dos estimadores, la media es uniformemente mejor que la mediana ya que, para cualquier valor de μ, su error cuadrático esperado es inferior. En la terminología de la teoría de la estimación, se dice que la mediana es un estimador *inadmisible* de μ pues existe otro uniformemente mejor. Hay que recalcar que se están comparando estimadores para la media de una distribución normal; si la distribución base no es normal, pueden darse otros resultados.

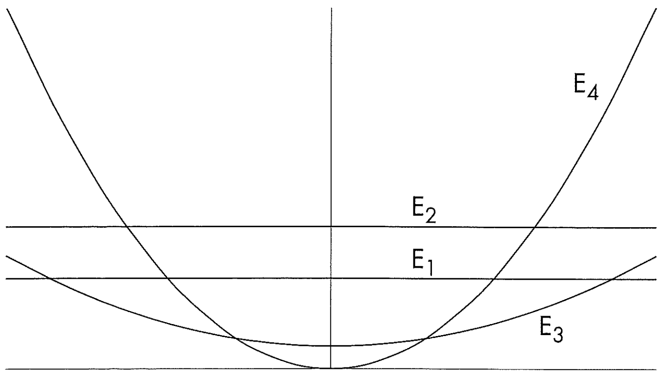

Gráfico 1

Para los estimadores 3 y 4, ambos sesgados hacia el valor $\mu = 0$, las funciones de error no son constantes. Las estimaciones que proporcionan serán muy precisas cuando el verdadero valor de μ se encuentre próximo a 0. En cambio, cuando μ se aleja de 0, crece rápidamente el error cuadrático esperado aunque este crecimiento es más rápido para T_4 que para T_3.

El teorema de Stein

Puede demostrarse que el promedio $\bar{x}$ es un estimador admisible para la media de una población normal; es decir, no existe otro estimador uniformemente mejor. Este resultado sigue siendo válido cuando son dos las medias a estimar. En cambio, Stein (1961), en colaboración con James, demostró que no es cierto en el caso de $k > 2$ medias; en otras palabras, cuando hay que estimar un conjunto de más de 2 medias, hacerlo a partir de sus promedios individuales es un procedimiento inadmisible. La demostración del teorema es constructiva en el sentido que proporciona un ejemplo de estimador para *k* medias cuyo error cuadrático esperado es uniformemente inferior al obtenido con los promedios individuales.

Con la ayuda de un ejemplo, se vislumbra la naturaleza paradójica de este resultado. Supongamos que se quiere estimar la eficacia de varios jugadores de baloncesto en el lanzamiento de tiros libres. De alguna manera, lo que se quiere obtener es una estimación del porcentaje de acierto en dichos lanzamientos que alcanzarán al final de la temporada. Para ello, se observan los 30 primeros lanzamientos de cada uno de los jugadores. Si hay que dar una estimación de la eficacia de los jugadores al final de la temporada con los datos de la muestra, la primera respuesta que a uno se le ocurre es dar como mejor estimación los porcentajes de acierto individuales obtenidos en la muestra. El teorema de Stein afirma que existe otra respuesta mejor y, además, proporciona un método para construirla: los denominados *estimadores de James-Stein.*

La esencia del método de Stein consiste en contraer los promedios individuales hacia el gran promedio, según un factor de contracción que depende del grado de dispersión de los promedios observados.

La esencia del método de Stein consiste en contraer los promedios individuales hacia el gran promedio (es decir hacia el promedio de los promedios), según un factor de contracción que depende del grado de dispersión de los promedios observados. Siguiendo con el ejemplo del baloncesto, el estimador de Stein de un jugador determinado se obtiene contrayendo su porcentaje de aciertos en los 30 primeros lanzamientos hacia el porcentaje medio observado entre todos los jugadores. Si el porcentaje de un jugador es superior al gran promedio, deberá disminuirse; en caso contrario, aumentarse. En cierta forma, el método de Stein postula que, en general, un promedio inicial alto se transformará en uno menos alto y, viceversa, uno inicialmente bajo no será, al final, tan bajo.

Matemáticamente, la expresión de los estimadores de James-Stein es:

$$z_i = \bar{y} + c(y_i - \bar{y})$$

El factor de contracción *c* depende del número de medias a estimar y de su dispersión. Su expresión es:

$$c = 1 - \frac{(k-3)\sigma^2}{\sum (y_i - \bar{y})^2}$$

donde:

$\bar{y}$: gran promedio

y_i: promedios individuales

σ: desviación típica

k: número de medias a estimar

La interdependencia entre las estimaciones resalta aún más el aspecto paradójico del resultado. ¿Por qué el grado de acierto inicial de un jugador ha de influir en la estimación de la eficacia de otro jugador? La respuesta a ésta y a otras preguntas debe hacerse a partir de un análisis exhaustivo del comportamiento de estos estimadores y del error cuadrático esperado para diferentes valores de los parámetros a estimar (las medias verdaderas). Su realización excede los límites de este trabajo y, para ello remitimos al lector al artículo de Bradley Efron (1977). Nuestra propuesta es menos ambiciosa: ya que no podemos demostrar el teorema, ¡intentemos comprobarlo!

Simulación con la hoja de cálculo

El modelo ha sido implementado sobre la hoja de cálculo Microsoft Excel 4.0, aunque es fácilmente trasportable a cualquier otra con semejante nivel de prestaciones. Consiste básicamente en simular 15 experiencias con la distribución binomial de parámetros n = 30 y p variable. Con los resultados obtenidos, se calculan las estimaciones de Stein para las proporciones teóricas *p* conocidas, lo que permite compararlas con los promedios observados y observar cuál de los dos métodos proporciona resultados más cercanos a los valores verdaderos.

La simulación completa puede verse en la tabla 1. En la parte superior de la hoja (filas 1 a 5) figuran los parámetros básicos del modelo: el tamaño de las muestras *n* y el número de medias a estimar k. También, en esa zona, se recogen los valores globales calculados a partir de los datos observados: el gran promedio *my*, la varianza *mvar* (calculada promediando las varianzas de las 15 distribuciones simuladas), el factor de contracción *c* y los errores cuadráticos totales de los dos conjuntos de estimadores (*SC* y para los promedios y *SCz* para los estimadores de Stein).

	A	B	C	D	E	F	G
1	n=	30					
2	k=	15		SCy=	0,1058		
3	my=	0,6422					
4	mvar=	0,0076		SCz=	0,0569		
5	c=	0,52844719					
6							
7	p	y	(y-my)^2	var	z	(p-y)^2	(p-z)^2
8	0,6	0,5333	0,0119	0,0080	0,5847	0,0044	0,0002
9	0,65	0,6667	0,0006	0,0076	0,6551	0,0003	0,0000
10	0,58	0,5000	0,0202	0,0081	0,5671	0,0064	0,0002
11	0,45	0,5000	0,0202	0,0083	0,5671	0,0025	0,0137
12	0,7	0,8000	0,0249	0,0070	0,7256	0,0100	0,0007
13	0,72	0,8333	0,0365	0,0067	0,7432	0,0128	0,0005
14	0,68	0,6333	0,0001	0,0073	0,6375	0,0022	0,0018
15	0,66	0,7333	0,0083	0,0075	0,6904	0,0054	0,0009
16	0,7	0,6333	0,0001	0,0070	0,6375	0,0044	0,0039
17	0,64	0,6333	0,0001	0,0077	0,6375	0,0000	0,0000
18	0,48	0,6667	0,0006	0,0083	0,6551	0,0348	0,0307
19	0,54	0,5000	0,0202	0,0083	0,5671	0,0016	0,0007
20	0,66	0,6333	0,0001	0,0075	0,6375	0,0007	0,0005
21	0,64	0,5333	0,0119	0,0077	0,5847	0,0114	0,0031
22	0,74	0,8333	0,0365	0,0064	0,7432	0,0087	0,0000

Tabla 1

El área A8:G22 contiene los resultados de las 15 simulaciones. El contenido de las columnas es el siguiente:

Columna A: contiene los valores verdaderos p de las proporciones a estimar.

Columna B: incluye los promedios obtenidos en la simulaciones. Por ejemplo, la casilla B8 contiene la fórmula =Binomial(B1;A8)/B1 que devuelve la proporción de éxitos al simular una distribución binomial de parámetros B1=30 y A8=0,6. La función Binomial ha sido implementada utilizando el lenguaje de macros que proporciona Excel. Para ello se ha definido previamente la función Bernoulli(p) que simula una distribución de Bernoulli de parámetro *p* (devuelve 1 con probabilidad *p* y 0 con probabilidad 1–p). La tabla 2 de la página siguiente muestra las implementaciones completas de ambas funciones.

Columna C: contiene los cuadrados de las diferencias entre cada promedio observado y el gran promedio.

Columna D: contiene las varianzas de las distribuciones simuladas. [La varianza de una distribución Binomial (n; p) es p(1–p)/n].

Columna E: contiene las estimaciones de Stein.

Columnas F y G: contienen los cuadrados de las diferencias entre los valores verdaderos y sus correspondientes estimados. La columna F se refiere a los promedios observados y la G a las estimaciones de Stein.

Los gráficos 2 y 3 presentan los resultados de la experimentación. Puede

Bernoulli	Binomial
=RESULTADO(7)	=RESULTADO(7)
=ARGUMENTO("p";15)	=ARGUMENTO("n";15)
=SI(ALEATORIO()<p)	=ARGUMENTO("p";15)
=VOLVER(1)	=ESTABLECER.NOMBRE("suma";0)
=SI.NO()	=PARA("i";1;n)
=VOLVER(0)	=ESTABLECER.NOMBRE("suma";suma+Bernoulli(p))
=FIN.SI()	=SALIR.BUCLE()
	=VOLVER(suma)

Tabla 2

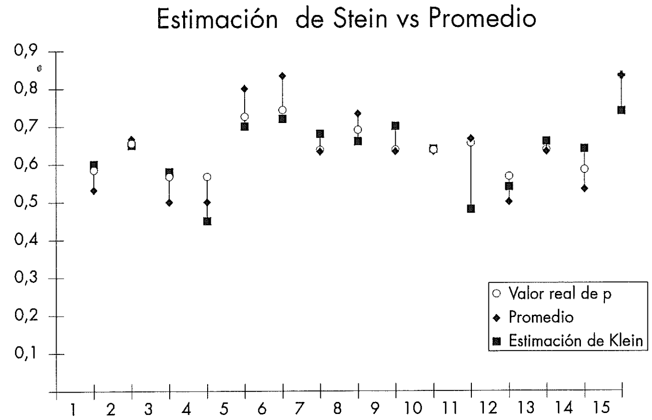

Gráfico 2

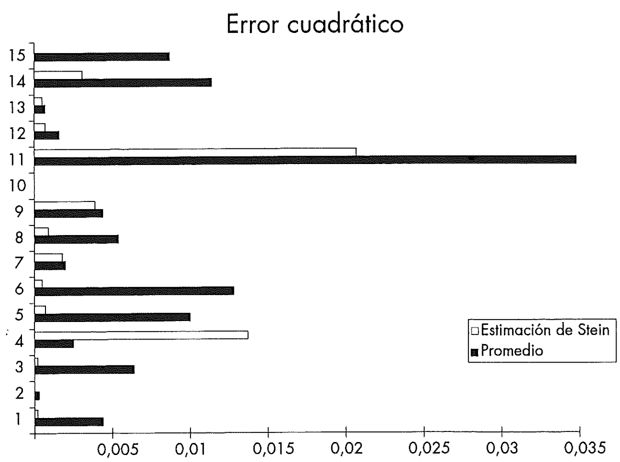

Gráfico 3

observarse como en 12 de las 15 simulaciones las estimaciones de Stein obtienen mejores resultados. Solamente en un caso fue superior el promedio observado. El error cuadrático total es de 0,1058 para los promedios individuales, casi el doble del obtenido con los estimadores de Stein: 0,0569.

Rafael Cortés
Centro de Profesores
de Palma de Mallorca

Bibliografía

EFRON, B. y C. MORRIS (1977): «La paradoja de Stein en estadística», *Investigación y Ciencia*, n.° 10, 94-102.

RÍOS, S. (1977): *Métodos estadísticos* (segunda edición), Ediciones del Castillo, Madrid.