

RAZONAMIENTO ACERCA DEL SIGNIFICADO DE LOS PARÁMETROS EN LOS MODELOS DE PROBABILIDAD EN ESTUDIANTES UNIVERSITARIOS

Álvaro Cortínez, Norma Alamilla, Armando Albert, José Guadalupe Ríos

Universidad de Tarapacá (Chile)

Universidad Politécnica del Centro. (México)

Tecnológico de Monterrey. (México)

acortinezp@uta.cl, norma.alamilla@gmail.com, albert@itesm.mx, jrios@itesm.mx

Palabras claves: variable aleatoria, parámetro, inferencia bayesiana

Keywords: random variable, parameter, bayesian inference

RESUMEN

En las últimas décadas la estadística bayesiana está siendo usada cada vez más en áreas de la salud, investigación en ingeniería y negocios, entre otros. Esto demanda una necesidad de formación inicial desde por lo menos el nivel de pregrado. Diversas universidades del mundo están ya incluyendo en sus programas de estudio elementos de estadística bayesiana. En nuestro caso, la inclusión de ideas bayesianas en el aula universitaria nos representa un reto didáctico importante por tener sus propias dificultades. Y el parámetro es un concepto básico y fundamental para la construcción de los modelos probabilísticos y sus derivaciones, por ejemplo la inferencia estadística. En esta investigación nos proponemos hacer una exploración sobre concepciones y dificultades de estudiantes universitarios de segundo año de carrera del área de administración de empresas, en la asignatura de Probabilidad y Estadística, alrededor de este concepto clave de la estadística que es parámetro.

ABSTRACT

In last decades the Bayesian statistics is applied in different areas: health, engineering, business, etc. This generates the necessity of training in undergraduate level. Some universities are including in their courses elements of Bayesian statistics. In our case, the inclusion of Bayesian ideas is an important didactic challenge. And the parameter is a basic and fundamental concept for the construction of probability models and its derivations, for example the statistical inference. In this research we propose to make an exploration about conceptions and difficulties of second level students from business studies, in Probability and Statistics subject, about the statistical key concept of parameter.

■ Introducción

Abordar la viabilidad de la estadística bayesiana en el aula, no es algo nuevo. Diversos autores se han dado a la tarea de iniciar su estudio. Tal es el caso de Albert (1995) con su propuesta universitaria de enseñanza de la inferencia acerca de proporciones usando Bayes y Modelos discretos. Iglesias, Leiter, Mendoza, Salinas y Varela (2000) en Chile realizan una Mesa Redonda sobre enseñanza de la estadística bayesiana y hacen una propuesta curricular. Díaz (2007) en diversos trabajos da su respuesta a ¿Cómo puede el método bayesiano contribuir a la investigación en psicología y educación?

Los modelos de probabilidad en los cursos de pregrado suelen enseñarse con la siguiente estructura: nombre del modelo, suceso de interés que describen, identificación de parámetros, función de probabilidad y elementos adicionales (esperanza, varianza, gráfica, etc.). El estudiante consigue, de esta forma, mecanizar la utilización de dichos modelos y basta con que se haga un esquema para sus múltiples ejercicios. Sin embargo, con relación al concepto de parámetro, éste no es presentado formalmente ni abordada didácticamente. Al hablar, por ejemplo, de una probabilidad π de un modelo Binomial, el estudiante busca en el enunciado del problema la información y con ella trabaja utilizando las fórmulas correspondientes. Este procedimiento no da cabida a la posibilidad de que no se tenga información acerca del parámetro. Menos aún existe la posibilidad de incorporar información propia del parámetro. Por ejemplo, si se está utilizando un modelo aplicado para estimar poblaciones de cierta especie de animales marinos, en base a sucesivos muestreos, siempre tendremos el mismo parámetro, cuando en realidad lo razonable es que cada vez que se realicen los muestreos, el valor del parámetro se actualiza. Desde esta perspectiva, el parámetro puede ser tratado como variable aleatoria, actualizando sus probabilidades en función del Teorema de Bayes. Una consecuencia directa de la adecuada formulación de los modelos de probabilidad es la realización de inferencias sobre parámetros y poblaciones. Resulta, por tanto, razonable formularse cuestiones acerca de los parámetros y buscar una respuesta basándose precisamente en los modelos de probabilidad correspondientes. De esta forma, surgen de forma natural las estimaciones puntuales (de hecho, el modelo de probabilidad a posteriori del parámetro es de por sí una estimación del mismo), de intervalo y los contrastes de hipótesis. Esta investigación se centra en un estudio inicial que describa la naturaleza del parámetro desde su complejidad epistémica, cognitiva y didáctica.

■ Acercamiento teórico-metodológico

El acercamiento teórico de esta investigación es desde el enfoque sistémico que considera que el fenómeno didáctico no puede ser visto de manera aislada sino en la interacción de sus dimensiones que le dan explicación (Cantoral, Farfán, Lezama y Sierra, 2006). En este sentido, esta investigación se propone hacer un acercamiento epistemológico, didáctico y cognitivo del parámetro como objeto didáctico con el propósito de tener elementos para comprender la naturaleza del problema didáctico del parámetro para futuras investigaciones tanto de intervención didáctica como del estudio de su emergencia en prácticas de investigación de sociedades profesionales que usan estadística para la generación de conocimiento y toma de decisiones.

■ Aproximación epistemológica

Desde la perspectiva epistemológica, se puede constatar la existencia de dos paradigmas en que se desarrolla la estadística que determinan significados no siempre similares del parámetro. Un primer

paradigma suele llamarse *frecuentista o clásico*. Los procedimientos se desenvuelven en torno a las diferentes muestras aleatorias que se pueden obtener. Las probabilidades, por lo tanto no necesariamente están relacionadas con una muestra aleatoria en particular. El paradigma alternativo es el *bayesiano*. Éste aplica las leyes de probabilidad directamente al problema en cuestión, intentando ser una interpretación directa del fenómeno que se está observando. La perspectiva bayesiana se basa en que como tenemos incertidumbre acerca de los valores de los parámetros, los consideraremos variables aleatorias. Las reglas de probabilidad se utilizan directamente para hacer inferencias sobre los parámetros. Las probabilidades de los parámetros se interpretan como "grado de creencia". La distribución a priori se plantea como subjetiva. Cada persona puede tener su propia creencia *a priori*, lo cual contiene los pesos relativos que esa persona le da a todos los valores posibles parámetros. Así mide cuán plausible considera la persona que el parámetro pueda tomar un determinado valor antes de observar los datos (Bolstad, 2007). De esta forma, se hace una distinción entre las variables aleatorias observables, típicamente "datos", y las variables aleatorias inobservables, típicamente "parámetros" (Bernardo, 2002).

■ Aproximación didáctica

Aunque un aproximación didáctica puede tener muchos elementos importantes tales como el profesor, los materiales para la clase, los programas de estudio, las prácticas educativas, entre otros, sin embargo, nuestro acercamiento en este trabajo será sólo a los libros de texto.

Como punto de partida, se considera necesario desglosar algunas de las definiciones que se pueden encontrar del parámetro en libros de texto.

Degroot y Schervish (2012) define el parámetro de la siguiente forma:

En un problema de inferencia estadística, cualquier característica de la distribución que genera los datos experimentales que tenga un valor desconocido, como la media μ o la varianza σ^2 , se llama parámetro de la distribución. El conjunto de todos los posibles valores del parámetro θ o de un vector de parámetros $(\theta_1, \theta_2, \dots, \theta_k)$ se llama espacio paramétrico. (Degroot y Schervish, 2012, p.378)

Devore (2012) presenta el concepto de parámetro dentro de la definición de distribuciones de probabilidad:

Suponga que $p(x)$ depende de una cantidad a la que se le puede asignar cualquiera de varios valores posibles, donde cada valor diferente determina una distribución de probabilidad distinta. A esta cantidad se le llama parámetro de la distribución. La colección de las distribuciones de probabilidad para diferentes valores del parámetro se llama familia de distribuciones de probabilidad. (Devore 2012, p. 94).

La problemática, como puede observarse en las definiciones anteriores, se presenta desde el comienzo, en su definición. La mayoría de los libros de texto de estadística toman el punto de vista clásico que se basa en que los parámetros, las características numéricas de la población, son constantes fijas pero desconocidas; las probabilidades se interpretan como frecuencias a largo plazo; los procedimientos estadísticos son juzgados en función de cómo se comportan en el largo plazo sobre infinitas hipotéticas

repeticiones del experimento. Las probabilidades sólo tienen sentido para cantidades aleatorias. Los parámetros son fijos, no aleatorios, de modo que no se les puede asignar probabilidades.

Cabe destacar que en los libros de texto y en bibliografía en general, las alusiones al parámetro van dirigidas prácticamente directas a su operacionalidad. Se mencionan en un contexto muy general, como características de una población y se apunta hacia las formas de estimarlo y su utilidad.

Otras fuentes de influencia en los estudiantes son las obtenidas vía Internet. Entre las más populares se entra Wikipedia y Vitutor. En el caso de Wikipedia, ésta afirma que:

En estadística, un parámetro es un número que resume la gran cantidad de datos que pueden derivarse del estudio de una variable estadística. El cálculo de este número está bien definido, usualmente mediante una fórmula aritmética obtenida a partir de datos de la población.

A continuación intenta precisar acerca de su utilidad: "Los parámetros estadísticos son una consecuencia inevitable del propósito esencial de la estadística: crear un modelo de la realidad".

En Vitutor: "Un parámetro estadístico es un número que se obtiene a partir de los datos de una distribución estadística."

Es importante señalar que estas definiciones difuminan las ideas de población y muestra, al considerar que los parámetros se calculan a partir de los datos o de una gran cantidad de datos. De la revisión, no se observa, además, otros puntos de vista distintos para conceptualizar el parámetro.

■ Aproximación cognitiva

Aunque ha habido interés por parte de investigadores en educación estadística por conocer las concepciones, errores y dificultades en la comprensión de conceptos de estadística (Batanero, Vallecillos, Green, y Holmes, 1994), sin embargo, en lo relativo a parámetro en educación estadística no se reportan investigaciones a la fecha. Por esa razón nos hemos dado a la tarea de hacer un primer estudio exploratorio con el fin de determinar que ideas tienen los estudiantes sobre el parámetro tras un primer acercamiento de ellos al concepto. Para ello se trabajó con un grupo de estudiantes de segundo año de Ingeniería Comercial de la Universidad de Tarapacá, en el norte de Chile, que llevan el curso de Probabilidad y Estadística. Se introdujeron los conceptos de probabilidad desde este punto clásico y subjetivista. Se hizo hincapié, en el caso subjetivista, en que las probabilidades se asignan dependiendo de la información que cada individuo tiene acerca del problema que se está tratando, siendo un caso extremo aquél en que no hay información. En una primera fase, se diseñó una situación didáctica, con el fin de recabar la idea de parámetro que los estudiantes luego de las clases y basándose en bibliografía, habían conseguido recordar, en lo posible asociándola con las situaciones de la vida diaria, externa al aula.

■ Antecedentes de los estudiantes

La experiencia se aplicó en un curso de Probabilidades y Estadística (PyE) de la carrera de Ingeniería Comercial (Administración de Empresas), de la Universidad de Tarapacá, en Arica, Chile. Dicho curso se

imparte en segundo año, luego de haber cursado durante el primer año, un curso de Álgebra, además de Cálculo I. En paralelo a nuestra asignatura, los estudiantes cursan Cálculo II (cálculo integral), junto a asignaturas iniciales de su área. Por lo tanto, podemos suponer razonablemente que los estudiantes cuentan con los recursos matemáticos suficientes para enfrentar el curso PyE, así como la actividad que se les presentó en esta investigación.

■ La experiencia

La experiencia se aplicó durante una clase. Anteriormente se les pidió a los estudiantes que formaran equipos de entre 3 y 5 personas. Es así como el curso completo se agrupó en 7 equipos: cinco de 5, uno de 4 y uno de 3 estudiantes. Dos días antes de la experiencia, se les envió a los "jefes de grupo" un conjunto de preguntas en torno al parámetro, para que ellos analizaran de forma grupal:

Actividad en el Aula

ACTIVIDAD EN EL AULA	
Ingeniería Comercial, primer semestre de 2014	
1.-	¿Cuál es la diferencia entre estadística descriptiva y estadística inferencial?
2.-	¿Puede un parámetro ser aleatorio?
3.-	¿Tiene sentido que un parámetro sea fijo?
4.-	¿Para qué sirve el teorema de Bayes?
5.-	¿Qué se entiende por aleatorio?
6.-	¿Qué entiende por variable aleatorio?
7.-	¿Para qué sirve la función de distribución?
8.-	¿Qué es un parámetro?
9.-	¿Qué es una variable aleatorio?
10.-	¿Qué es una población?
11.-	¿Puede definirse un modelo sin parámetro?
12.-	¿Cómo puedo incorporar la información que tenga el investigador acerca del parámetro en el modelo?
13.-	Un parámetro, ¿podría tener distribución propia?
14.-	¿Qué sentido tiene hablar de la probabilidad asociada a un parámetro?

El día de la experiencia, y delante de los estudiantes, con ayuda de una calculadora electrónica, se seleccionó el orden aleatorio de los equipos para que hicieran su presentación, nombrándose los equipos del 1 al 7. A continuación, para cada equipo se escogieron, también con ayuda de una calculadora electrónica, dos de las 15 preguntas para que ellos en unos 3 minutos pusieran en orden sus ideas y plantearan la estrategia de la presentación. Llegado el momento de dicha presentación, frente a una cámara, el profesor les formula las dos preguntas, y los estudiantes comienzan a responderlas, en el formato que ellos decidan. Después, el profesor les plantea algunas preguntas adicionales, relacionadas con lo que el equipo contestó.

Es importante señalar que las preguntas directa o indirectamente, tenían como fin hablar del concepto de "parámetro" ("¿Qué es un parámetro?"), caracterizarlo y, en lo posible, definirlo dentro de un contexto fijo o aleatorio, con sus respectivas consecuencias. En este reporte destacaremos algunas de las

respuestas dadas por los estudiantes, a los cuales identificaremos con el grupo y al orden en que responde dentro del mismo.

El concepto de parámetro genera muchas confusiones entre los estudiantes de los diferentes equipos. En general, se asocia con la estadística descriptiva, y se dan definiciones en torno a los datos mismos. "Un parámetro nosotros lo podemos definir o podemos tener en cuenta como un dato base, un dato característico" (Alumno 3.1); "Un parámetro es un número fijo que es un resultado de un conjunto de datos" (Alumno 6.2). Al consultársele a un equipo respecto a la diferencia entre parámetro y variable, la respuesta es (Alumno 2.1):

"El parámetro es un número resultante de un experimento previo. Y este número caracteriza una característica de la población, o sea interpreta una característica propia de la población, por ejemplo la media o la moda. Y una variable es un dato, no necesariamente una unidad o un número, sino que un dato que también es resultante de un experimento, pero no necesariamente representa a la población".

Al parámetro se le caracteriza como un modelo de la realidad y se entra en contradicciones respecto a su naturaleza de fijo o aleatorio: "Voy a partir por si tenemos un experimento con datos duros, podemos llegar a un parámetro concreto, pero si a ese experimento le ponemos datos variables, el parámetro va a ser variable"(Alumno 5.3). En el Equipo 7, hay una clara discusión, pues mientras el alumno 7.3 afirma que "Puede variar", el alumno 7.2 asegura que "Para mí el parámetro es fijo, no puede ser aleatorio. Para mí". Por su parte, del equipo 3 extraemos la siguiente afirmación, abriéndose a la posibilidad de que el parámetro sea aleatorio: "Dependiendo de la variable, porque si tenemos la variable aleatoria, el parámetro podría ser un parámetro desde una variable aleatoria" (Alumno 3.2).

A pesar de que en general, los equipos afirman que el parámetro es fijo, plantean la posibilidad de que tenga distribución de probabilidades: "De ese modo sí, porque el parámetro va en función de los valores de la muestra. Si los valores de la muestra son aleatorios, entonces el parámetro sería aleatorio solamente. Si es que no es un parámetro de datos duros"(Alumno 1.1) ; " Sí, tiene sentido, porque al llevar ese parámetro, el resultado es determinante para la toma de decisiones del modelo de la realidad que se quiere implementar" (Alumno 2.1). El alumno 3.2 argumenta un poco más:

"A mi parecer no podría tener una distribución propia como tal, porque para un parámetro necesitamos datos anteriores, o sea variables. Y de las variables, ellas pueden ser o tener distribución. Pero el parámetro no puede tener su distribución, debido a que la función de distribución lo que hace es formar una condición respecto a la variable."

Un obstáculo que ha quedado en evidencia durante esta experiencia, es la comprensión de dos grandes ejes de los programas de estadística ¿Cuál es la diferencia entre Estadística Descriptiva y Estadística Inferencial?. Se observa inmediatamente la inclinación a asociar la descripción de datos a la población y la inferencia a las muestras. Los estudiantes conocen los conceptos de analizar, estudiar, describir, estimar, predecir, pero los ubican en los lugares incorrectos. Por ejemplo, el alumno 4.1 afirma:

"La Estadística Descriptiva tiene que ver más que nada con la población y la Estadística Inferencial tiene que ver con la muestra. La Estadística Descriptiva analiza, estudia y describe la totalidad de los individuos de una población. Y la Estadística Inferencial solamente trabaja con la muestra y su conjunto que forman parte del individuo de una población."

"La Estadística Descriptiva se puede decir que su respuesta es más fiable que la Estadística Inferencial, ya que la inferencial sólo trabaja con una muestra en cambio la otra trabaja con la totalidad de la población".

Este problema lleva a confundir el contexto en que analizamos el parámetro, e incluso su naturaleza. Alumno 4.4:

"En la Descriptiva me planteo preguntas sobre el parámetro, porque en la Estadística Descriptiva analizamos y describimos, podemos sacar una totalidad de la población. Y ahí podría ser más un parámetro fijo. Pero si tenemos un parámetro aleatorio, lo podemos sacar de la Estadística Inferencial, ya que la Estadística Inferencial utiliza sólo la muestra de la población para sacar un análisis".

■ Conclusiones

Una importante dificultad que suelen tener los estudiantes es el paso a los primeros elementos de Teoría de Probabilidades. A pesar de ser un concepto que no requiere elementos matemáticos muy complejos, al menos en su introducción, suele presentar serias dificultades en los estudiantes. Por ejemplo, el concepto de aleatorio es asociado al azar o a la suerte. Sin embargo no queda claro que realmente lo entiendan ("por cada suceso que ocurra, las probabilidades son diferentes"). Es asociado a elementos de la vida real que se suelen conocer como aleatorios (como la secuencia aleatoria en un aparato reproductor de música). Dada esta visión de lo "aleatorio", y el concepto de parámetro que manejan, es razonable que no puedan llegar a la idea de parámetro aleatorio. Eso, a pesar de que la duda les provoca disentir con sus propios argumentos. Por ejemplo, "no puede ser aleatorio porque es un número fijo", pero "sí se le puede asignar una distribución de probabilidades". Se utiliza el concepto de "datos duros" y "datos variables", sin entenderlo mayormente y muy posiblemente a partir de expresiones habituales que mencionan los "datos duros" para dar una precisión sobre un argumento o para refutarlo. En ningún caso se manejan en estadística dichas expresiones.

Cabe señalar que para los estudiantes de este nivel, es complicado aún el contexto aleatorio de un parámetro, sobretodo porque es muy predominante en bibliografía el concepto del parámetro fijo. Si bien, en sus argumentaciones, ellos buscan "explicar la realidad", no perciben que quizás modelando el parámetro como una variable aleatoria, se puedan explicar mecanismos de razonamiento que sí se utilizan en la vida diaria. Es posible que la complicación parta incluso debido a la estructura de los programas de asignaturas, en que se diferencian habitualmente tres partes: estadística descriptiva, probabilidad y variable aleatoria, y cada una de esas partes son tratadas prácticamente de forma independiente y en muchas ocasiones como un conjunto de recetas.

Por otra parte, la aproximación epistemológica permitió entender aún más la aproximación didáctica porque dio explicación a lo que sucede en la enseñanza de la estadística en pregrado. Ésta está

básicamente centrada en la estadística clásica, dejándose los elementos subjetivistas casi como una simple curiosidad científica. De hecho, Bernardo (2002) hace un análisis comparado de los programas en los que se basa la enseñanza universitaria de la estadística matemática en los primeros ciclos de la enseñanza universitaria, poniendo de manifiesto un inmenso desfase entre lo que se enseña como 'teoría establecida' y lo que se investiga y (sobre todo) se aplica en el mundo real. Los programas suelen ignorar las limitaciones asociadas al uso exclusivo del paradigma clásico, eludiendo cualquier comentario sobre el paradigma bayesiano. Esto mismo se ve reflejado en los estudiantes cuando tienen dificultades para determinar si se le puede asociar o no al parámetro probabilidades cuando se confrontan con situaciones donde es razonable pensar en la variabilidad del parámetro. El detenerse a analizar los conceptos claves de la estadística, relacionándolos con los contextos externos al aula, permitirá que los estudiantes tengan menos barreras para comprender la estadística en general y su utilidad en su formación. Y eso requiere no obviar los diferentes paradigmas modernos.

■ Referencias bibliográficas

- Albert, J. (1995). Teaching Inference about Proportions Using Bayes and Discrete Models. *Journal of Statistics Education*, 3(3).
- Batanero, C., Godino, J. D., Vallecillos, A., Green, D. R., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematical Education in Science and Technology*, 25(4), 527-547.
- Bernardo, J. M. (2002). Un programa de síntesis para la enseñanza universitaria de la Estadística Matemática contemporánea. *Revista de la Real Academia de Ciencias*, 95(1-2), 81-99.
- Bolstad, W. M. (2007). *Introduction to Bayesian statistics*. Hoboken, New Jersey, USA: John Wiley & Sons.
- Cantoral, R., Farfán, R. M., Lezama, J., y Sierra, G. M. (2006). Socioepistemología y representación: algunos ejemplos. *Revista latinoamericana de investigación en matemática educativa*, 9(1), 83-102.
- DeGroot, M. y Schervish, M. (2012). *Probability and statistics*. Wilmington, Delaware, USA: Addison-Wesley.
- Devore, J. (2012). *Probabilidad y Estadística para Ingeniería y Ciencias*. México: Cengage Learning.
- Díaz, C. (2007). *Viabilidad de la enseñanza de la inferencia bayesiana en el análisis de datos en Psicología*. Tesis de doctorado no publicada. Universidad de Granada, Granada, España.
- Iglesias, P., Leite, J. G., Mendoza, M., Salinas, V., & Varela, H. (1997). Mesa Redonda Sobre la Enseñanza de la Estadística Bayesiana. *Revista Sociedad Chilena de Estadística*, 16-17(1-2), 105-120.
- Plataforma de Teleformación Vitutor. Última consulta: 28 de septiembre de 2014. http://www.vitutor.com/estadistica/descriptiva/a_7.html
- Wikipedia. La Enciclopedia Libre. Última consulta: 28 de septiembre de 2014. http://es.wikipedia.org/wiki/Par%C3%A1metro_estad%C3%ADstico