LA LEY DE LOS GRANDES NÚMEROS COMO PUENTE DIDÁCTICO ENTRE UNA VARIABLE ESTADÍSTICA Y SU VARIABLE ALEATORIA ASOCIADA

Stella Maris Figueroa; María Andrea Aznar; Gloria Prieto
Universidad Nacional de Mar del Plata. Facultad de Ingeniería.
stellafigueroa@gmail.com

Resumen

Una de las dificultades más importantes en la enseñanza tradicional de la estadística surge de no considerar el análisis de datos como parte necesaria del proceso de aprendizaje. En la búsqueda de la contextualización de los datos, esta propuesta utiliza el concepto de aleatoriedad desde el punto de vista formal a través de la simulación de fenómenos aleatorios con el programa GeoGebra. Se pretende favorecer la construcción de los distintos elementos de significado que intervienen en la vinculación entre una variable estadística y su variable aleatoria asociada. En este escenario, la variable aleatoria surge naturalmente como una "ampliación" de su variable estadística asociada. Mientras que la variable aleatoria toma todos los resultados posibles del experimento aleatorio, su variable estadística toma sólo los resultados de una muestra, y, cuanto más se amplíe el tamaño muestral, la variable estadística tomará más valores cuyas frecuencias relativas tendrán mejor aproximación a la probabilidad respectiva.

Palabras clave: Ley de los grandes números, Variable estadística, Variable aleatoria, Simulación, GeoGebra

Abstract

One of the most important difficulties in the traditional teaching arises not consider data analysis as a necessary part of the learning process. In search of contextualization of data, this proposal uses the concept of randomness from the formal point of view through the simulation of random phenomena with the GeoGebra program. The intention is to favors the construction of the different elements of meaning involved in link between a statistical variable and its associated random variable, by means of the law of large numbers. In this scene, the random variable arises naturally as an "extension" of its associated statistical variable. While the random variable takes all possible outcomes of the random experiment, its statistical variable takes only the results of a sample, and when the sample size expands, the statistical variable take more values, whose relative frequencies have better approach to respective probability.

Keywords: law of large numbers - statistical variable - random variable - simulation - GeoGebra

1. Introducción

Una de las dificultades más importantes en la enseñanza tradicional de la estadística surge de la falta de interpretación de los datos en un contexto. Para tratar estas dificultades, expertos en didáctica de la estadística proponen una enseñanza contextualizada de esta ciencia, y una de sus recomendaciones es la presentación de la estadística como complementaria de la probabilidad. Esta complementariedad se puede mostrar a los estudiantes con la visualización de las regularidades en el comportamiento de los fenómenos aleatorios, con la comprobación de la convergencia de las frecuencias relativas a la probabilidad y con la inferencia informal, considerada como la simulación de fenómenos aleatorios a través de recursos informáticos.

El concepto de aleatoriedad se puede interpretar desde un punto de vista formal o informal. Batanero (2001) señala que, desde el punto de vista formal, la aleatoriedad es la sucesión de resultados de un mismo experimento realizado de manera independiente y en forma repetida. Desde el punto de vista informal, la aleatoriedad tiene como idea central el azar, es el patrón que explica efectos para los que no conocemos la causa o que no son predecibles.

Para Rodríguez (2012), el razonamiento inferencial informal, es un proceso que incluye razonar sobre las posibles características de una población basadas en una muestra, sobre las posibles diferencias entre dos poblaciones, observadas entre sus respectivas muestras de datos, para responder si dichas diferencias se deben al azar o si existe alguna diferencia significativa entre ambas, y razonar si es probable que se presente una muestra con determinados datos y con sus estadísticos descriptivos.

El presente trabajo considera estas definiciones y las recomendaciones de expertos en didáctica de la estadística, ya que tiene en cuenta el concepto de aleatoriedad y aplica un acercamiento empírico, la inferencia informal, con la simulación de fenómenos aleatorios. Con estos elementos y, en la búsqueda de contextualizar la enseñanza de la estadística, se presenta una variable estadística con su variable aleatoria asociada.

El marco teórico de la propuesta está sustentado en el Enfoque Ontosemiótico de la Instrucción y la Cognición Matemática (EOS), que define una práctica matemática como toda actuación o manifestación realizada para resolver problemas matemáticos, comunicar a otros la solución, validar la solución y generalizarla a otros contextos. El significado de un objeto matemático está conformado por el sistema de prácticas operativas y discursivas para resolver cierto tipo de problemas. Si estas prácticas corresponden a un determinado individuo se habla de significado personal, y si son compartidas en el interior de una institución se habla de significado institucional. Así, el significado personal es el construido por cada estudiante o aprendiz, y el significado institucional es el dado por los textos matemáticos, las orientaciones curriculares, lo expresado por los expertos. El aprendizaje supone la apropiación de los significados institucionales por parte del estudiante, mediante su participación en las comunidades de prácticas. La comprensión del significado de un objeto matemático por un sujeto está dada cuando es capaz de reconocer sus propiedades y representaciones, relacionarlo con otros objetos y aplicarlo a una diversidad de situaciones problemáticas (Godino, 2003)

En este sentido, la propuesta se focalizó en favorecer la construcción de los distintos elementos de significado que intervienen en la vinculación entre una variable estadística y su variable aleatoria asociada, mediante la ley de los grandes números.

Para ello, se plantean dos situaciones problemáticas del campo de la ingeniería. La primera situación está referida a la cantidad de señales que se pueden transmitir sobre un canal de comunicación digital. La transmisión de dichas señales puede ser correcta o incorrecta, dando lugar a pensar a cada señal transmitida como los resultados de un experimento aleatorio de Bernoulli, con una probabilidad de éxito p. La cantidad de señales erróneas que se producen, en un cierto número n de transmisiones efectuadas, son los resultados simulados de una muestra extraída de una población dada con distribución binomial. La simulación puede realizarse con el programa GeoGebra que a su vez, provee herramientas para el análisis de los datos.

La segunda situación problemática utiliza datos reales de una variable estadística con su distribución de frecuencias. Se refiere a una empresa de la construcción que reporta los accidentes laborales ocurridos a ingenieros jóvenes durante su primer año de trabajo.

La idea que se propone en este caso, es encontrar un modelo de probabilidad que se ajuste a la variable estadística y que permita resolver situaciones reales que puedan predecirse. Para el ajuste, se comparan los valores observados con los valores teóricos esperados. Dichos valores teóricos están determinados por un modelo de probabilidad asociado al experimento aleatorio del problema propuesto, la distribución de Poisson. Se compara luego esta distribución aproximada con la distribución binomial y se resuelve el problema planteado.
Los elementos de significado que intervienen para vincular una variable estadística y su variable aleatoria asociada, están dados por los objetos matemáticos primarios que define el EOS: el *campo de problemas* del que emerge la variable estadística y su variable aleatoria asociada, el *lenguaje* utilizado, referido a la representación de la variable estadística con su distribución de frecuencias relativas mediante tablas y gráficos, los *procedimientos* empleados, como la simulación de muestras aleatorias, el cálculo de estadísticos descriptivos para la interpretación y análisis de los datos, el cálculo de probabilidades, la comparación con las frecuencias relativas a medida que aumenta el tamaño de la muestra, la identificación de las variables binomial y de Poisson, las *definiciones y propiedades* aplicadas en la variable aleatoria asociada con su distribución de probabilidades, los *argumentos* que justifican las relaciones entre frecuencia relativa y probabilidad, como la ley de los grandes números, la verificación de que la distribución de Poisson es una aproximación de la distribución binomial. Sobre todo, se destacaron los argumentos que distinguen una variable estadística de una variable aleatoria, al trabajar con los resultados obtenidos en una sola muestra y al compararlos con todos los resultados posibles de un experimento aleatorio.

2. Descripción de la propuesta didáctica

En principio, se propone analizar el experimento aleatorio de efectuar cinco transmisiones de bits y detectar qué resultados posibles tendría ese experimento. Luego se define la variable aleatoria X: "número de bits transmitidos erróneamente en cinco transmisiones", para que los estudiantes descubran la correspondencia entre el conjunto de resultados posibles del experimento y el conjunto de los números reales. Posteriormente, se busca que los estudiantes calculen la probabilidad correspondiente a cada valor que toma la variable y se registran los resultados en el pizarrón.
Se pondrá énfasis en la aparición de *todos* los valores de la variable por la misma definición de variable aleatoria, que considera *todos* los resultados posibles del experimento aleatorio. Se destacará este hecho como el equivalente a considerar *todas las muestras posibles con reposición* obtenidas de la población dada. Luego se calcularán las probabilidades respectivas para cada valor de la variable.
Se comprobarán estos resultados a través del GeoGebra 5.0, de dos maneras muy diferentes.
La primera comprobación, de naturaleza teórica, eligiendo la opción binomial para n=5 y p=0.4 y verificando la distribución de probabilidades en la gráfica y en la tabla correspondiente. Esta actividad recapitularía lo realizado con los estudiantes en el pizarrón.
Para la otra comprobación, de tipo empírico, se seleccionarán, sin reposición, muestras aleatorias de distintos tamaños con el comando AleatorioBinomial[n,p], siendo n el número de pruebas repetidas independientes que plantea el problema, y p la probabilidad de éxito en cada prueba. Se propondrá el análisis de la variable estadística "número de transmisiones erróneas" y se compararán la distribución de las frecuencias relativas obtenidas en cada muestra, con la distribución de probabilidades de su variable

aleatoria asociada. A continuación se muestran los dos problemas que se propondrán a los estudiantes para trabajar en grupos.
I) Problema simulado: Se transmiten 5 bits sobre un canal de comunicación digital y se define la variable aleatoria X como el "número de bits transmitidos erróneamente en cinco transmisiones".
¿Cuál es el experimento aleatorio? ¿Qué resultado se obtiene en cada transmisión? ¿Las transmisiones son independientes? Dé ejemplos de resultados en las cinco transmisiones. ¿Cuál es el espacio muestral? ¿Cuántos elementos tiene? ¿Son mutuamente excluyentes los resultados (c, c, c, e, e) y (c, e, e, c, c)? Sea p = 0,4 la probabilidad de que una transmisión sea errónea. ¿Cuál es la probabilidad de que una transmisión sea correcta? Defina variable de Bernoulli y variable Binomial. Use el comando BinomialAleatorio de GeoGebra para simular dos muestras, de tamaño 20 y 50 respectivamente de las 5 transmisiones asociadas a la variable estadística con p=0.4. Construya su distribución de frecuencias relativas en cada caso. Grafique.
Compare la distribución de frecuencias relativas de la variable estadística con la distribución de probabilidades de su variable aleatoria asociada X. ¿Qué conclusión extrae? Para la comparación, se construye con los estudiantes un gráfico con la distribución de probabilidades de la variable aleatoria X: "número de bits transmitidos erróneamente en cinco transmisiones" y otro gráfico donde aparecen las actividades desarrolladas con los alumnos en la simulación, como se muestra en el Gráfico 1.

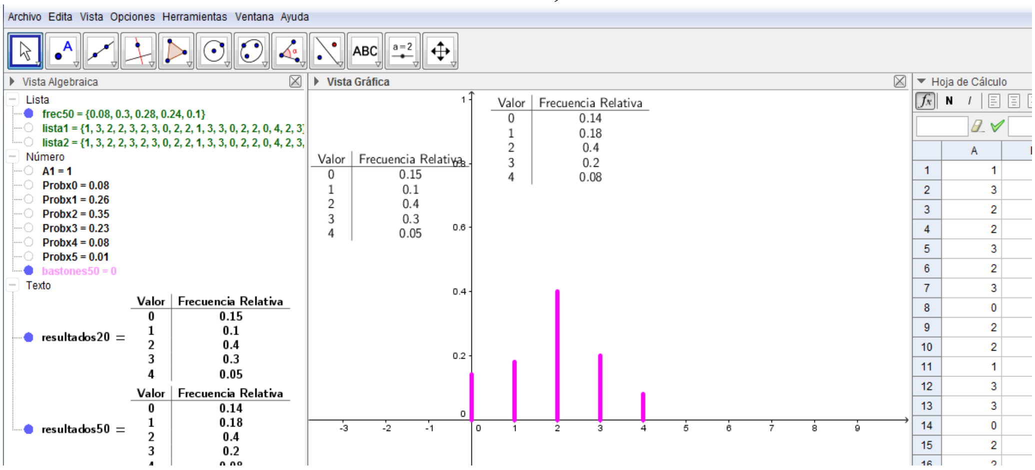

Gráfico 1: Vista algebraica, vista gráfica y parte de la planilla de cálculo de la distribución de la variable estadística del problema

La resolución del segundo problema, cuyo enunciado se muestra a continuación, requiere de los estudiantes ciertos conocimientos previos tales como la identificación de las variables binomial y de Poisson y la relación entre ambas distribuciones de probabilidad
II) Una empresa de la construcción reporta los siguientes accidentes laborales de 155 ingenieros entre 30 y 35 años durante su primer año de trabajo. Los resultados se muestran en la Tabla1:

Accidentes de trabajo	0	1	2	3	4 ó más
Nº de ingenieros	80	61	13	1	0

Tabla1: Número de accidentes laborales de ingenieros en su 1er año de trabajo.

La dirección de la empresa proyecta pagar un plan de seguros médicos a los ingenieros jóvenes, pues su trabajo los expone más a los accidentes. La aseguradora solicitó la información mencionada para entregar un plan de seguros apropiado y desea saber si la frecuencia de accidentes está dentro de límites aceptables.

Para resolver este problema, se requiere relacionar la variable estadística "número de accidentes laborales en su 1er año" con su variable aleatoria asociada, comparando su distribución de probabilidades con la distribución empírica de la variable estadística y sus frecuencias relativas.
Por consiguiente, a partir de la variable que cuenta el número de accidentes laborales en un año, se plantean preguntas como estas: ¿En qué distribución de probabilidades podríamos pensar de acuerdo al fenómeno aleatorio que se está analizando? ¿Cómo calcular su distribución de probabilidades para verificar esta suposición?
Se pretende que los estudiantes identifiquen la variable aleatoria de Poisson al detectar el experimento aleatorio, en este caso es trabajar en la empresa de construcción, y su variable estadística, el "número de accidentes laborales producidos en un año", ya que es una variable discreta que cuenta el "número de casos raros" en un intervalo, en este caso, de tiempo.
Una vez identificada la variable aleatoria asociada a esta variable estadística, los estudiantes se encuentran con el problema de vincular esta distribución de frecuencias con la distribución de probabilidades de Poisson de parámetro λ, donde λ es el promedio de accidentes por año. ¿Cómo encontrar este valor? Los estudiantes efectuarán el cálculo correspondiente al promedio en esta serie de frecuencias. Así obtendrán una estimación de λ= 0,5806 en un año. Conocido λ, ya pueden calcular la distribución de probabilidades del número de accidentes en un año y efectuar comparaciones con las respectivas frecuencias relativas, verificando la ley de los grandes números y responder luego a la problemática planteada.
Pero esta propuesta intenta además verificar que la distribución de Poisson es una aproximación de la distribución binomial. Para ello, apelando a los conocimientos previos de los estudiantes, se les pregunta ¿por qué a la distribución de Poisson se la conoce como la ley de los sucesos raros? Cuando en una distribución binomial, el número de pruebas repetidas e independientes es muy grande y la probabilidad p de que ocurra un determinado suceso es muy pequeña, se considera un "suceso raro" y en ese caso, $\lambda = n.p$, es un valor finito. Para el problema, $0,5806 = n.p$. Esta igualdad permite vincular la distribución de Poisson con la distribución binomial y también obtener la probabilidad de tener un accidente por día, considerando el año en 365 días, por lo que $p = 0,5806/365 = 0,0016$.
Una vez identificados *n* y *p*, se puede encontrar la distribución de probabilidades de esta variable binomial: "número de accidentes laborales que ocurren en 365 días"
En la Tabla2 se pueden ver las relaciones entre los datos reales con sus frecuencias relativas, los ajustes a la probabilidad aproximada aplicando Poisson y la probabilidad exacta con la distribución binomial.

X: nº de accidentes en 1 año	Frecuencias relativas	Prob.Aprox. (Poisson) λ= 0,5806	Probabilidad exacta (binomial) n=365 y p= 0,0016
0	80/155=0,5161	0,5595	$(1-0,0016)^{365} = 0,55930$
1	61/155= 0,393548	0,3249	$C_{365,1}.0,0016^{1}(1-0,0016)^{364.} = 0,326044$
2	13/155 =0,08387	0,0943	$C_{365,2}.0,0016^{2}(1-0,0016)^{363.} = 0,0950963$
3	1/155= 0,00645	0,0183	$C_{365,3}.0,0016^{3}(1-0,0016)^{362.} = 0,018413$
4	0/155=0	0,0027	$C_{365,4}.0,0016^{4}(1-0,0016)^{361.} = 0,00267$
5	0	0,0003	$C_{365,5}.0,0016^{5}(1-0,0016)^{360} = 0,0003094$
6	0	0	$C_{365,6}.0,0016^{6}(1-0,0016)^{359} = 0,000029$

Tabla2: Relaciones entre los datos reales y los ajustes a las distribuciones de Poisson y binomial.

Al haber encontrado la distribución de Poisson que se ajusta a los datos reales, se puede responder a la aseguradora que la frecuencia de accidentes está dentro de los límites aceptables, ya que se puede calcular la probabilidad de su ocurrencia y verificar su aproximación.

3. Reflexiones finales

Con la premisa de proponer una enseñanza contextualizada de la estadística, expertos en didáctica de la estadística sugieren la presentación de la estadística como complementaria de la probabilidad. En ese sentido, esta propuesta didáctica plantea un acercamiento entre el concepto de variable aleatoria y su variable estadística asociada a través de la ley de los grandes números.
Buscando promover en los estudiantes la vinculación entre estas variables, esta propuesta los hace partícipes de la realización de variadas e integradas prácticas con los objetos matemáticos involucrados. Entre ellas se pueden mencionar la selección de muestras aleatorias con la simulación que proporciona GeoGebra, la visualización de los resultados de estas muestras a través de tablas de frecuencias y de gráficos, la comparación de la distribución de frecuencias relativas de la variable estadística con la distribución de probabilidades de su variable aleatoria asociada y la verificación de la ley de los grandes números. También se plantea la verificación de resultados teóricos al comprobar que la distribución de Poisson es una aproximación de la distribución binomial, utilizando la relación entre una variable estadística y su variable aleatoria asociada.
Así mismo, se destacaron los argumentos que distinguen una variable estadística de una variable aleatoria, cuando se consideraron por un lado, los resultados obtenidos en una sola muestra y por otro lado, todos los resultados posibles de un experimento aleatorio. Se enfatizó que, en una variable estadística, se obtiene la distribución de frecuencias relativas de los resultados de una sola muestra y no necesariamente toma todos los valores posibles como lo hace su variable aleatoria asociada.
De esta forma, la variable aleatoria surge naturalmente como una "ampliación" de la variable estadística asociada. Mientras que la variable aleatoria toma todos los resultados posibles del experimento aleatorio, obtenidos de todas las muestras posibles de la población, su variable estadística toma sólo los resultados de una muestra. Además, se pudo comprobar que cuanto más se amplíe el tamaño de la muestra, la variable estadística tomará más valores, cuyas frecuencias relativas tendrán una mejor aproximación a la probabilidad respectiva.
Pensamos que el manejo de conceptos fundamentales como los de población y muestra, variable estadística y variable aleatoria en un mismo contexto, contribuyen a su comprensión. Esta conjetura puede ser la hipótesis de un próximo trabajo de investigación, ya que estos conceptos no son definiciones aisladas, sino que se relacionan con un sentido para la resolución de una situación problemática donde es necesario el aporte de la estadística y de la probabilidad como actividades complementarias.

4. Referencias

Batanero, C. (2001). "Didáctica de la estadística". Granada: Grupo de Investigación en Educación Estadística.

Batanero, C. (2003). "La simulación como instrumento de modelización en probabilidad". *Educación y Pedagogía*, 35: 37-64.

Rodríguez, M., (2012) "Inferencia Informal: Del Análisis de los Datos a la Inferencia Estadística". *FAMAF*. Universidad Nacional de Río Cuarto Córdoba. Versión obtenida el 3/junio/2013. http://www2.famaf.unc.edu.ar/rev_edu/documents/vol_28/28-1_Rodriguez-InferenciaInformal.pdf

Ruiz, B. Batanero, C. y Arteaga, P. (2011). "Vinculación de la Variable Aleatoria y Estadística en la Realización de Inferencias Informales por parte de Futuros Profesores". *Bolema,* Rio Claro (SP), 24(39): 431-449.

Godino, J.D (2003). "Teoría de las funciones semióticas. Un enfoque Ontológico-Semiótico de la Cognición e Instrucción Matemática". Universidad de Granada. Documento publicado en Internet: [http://www.ugr.es/~jgodino/funciones-semioticas/monografiatfs.pdf].

Godino, J. D., Batanero, C. y Font, V. (2009). "Un enfoque ontosemiótico del conocimiento y la instrucción matemática". [Versión ampliada del artículo: Godino, J., Batanero, C. y Font, V. (2007). The ontosemiotic approach to research in mathematics education. ZDM. The Internacional Journal on Mathematics Education, 39 (1-2): 127-135]. Versión obtenida el 10/junio/2010 Disponible en: http://www.ugr.es/~jgodino/funciones-semioticas/sintesis_eos_10marzo08.pdf