VOL 31, NÚMERO 1, AÑO 2018

MEJORA DE UNA PROPUESTA DIDÁCTICA SOBRE LA VINCULACIÓN ENTRE UNA VARIABLE ESTADÍSTICA Y SU VARIABLE ALEATORIA ASOCIADA

Stella Maris Figueroa, Sandra Baccelli
Universidad Nacional de Mar del Plata. Facultad de Ingeniería. (Argentina)
stellafigueroa@gmail.com, sbaccelli@gmail.com

Resumen

Este trabajo muestra la propuesta de mejora de un proceso de instrucción con un problema cuya modelización corresponde a una distribución binomial. Se utilizan herramientas teóricas del Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos. Los datos corresponden a estudiantes de estadística de la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata. El análisis de la idoneidad didáctica provee información y propone un diseño instruccional que considera los conflictos cognitivos encontrados y favorece la vinculación entre una variable estadística con su variable aleatoria asociada para propiciar una enseñanza de la estadística con proyectos de la ingeniería.

Palabras clave: variable estadística, variable aleatoria, simulación con GeoGebra

Abstract

This work shows a proposal to improve a teaching process with a problem whose modelling corresponds to a binomial distribution. Theoretical tools of the Onto-semiotic approach to mathematical knowledge and instruction are used. The data correspond to students of statistics of Engineering Faculty at the National Universidad in Mar del Plata. The analysis of didactic suitability provides information and proposes an instructional design that considers the cognitive conflicts found and improves the link between a statistical variable and its associate random variable to encourage statistics teaching through engineering projects.

Key words: statistical variable, random variable simulation with GeoGebra

Introducción

Los textos de estadística para carreras de ingeniería, en general, se inician con estadística descriptiva, y desarrollan el concepto de variable estadística como aquella *característica medible* que se estudia a partir de los resultados obtenidos de una muestra. Recién con la teoría de probabilidades, se presenta la unidad relativa a variables aleatorias, pero en general, en la mayoría de los textos, no existe una vinculación explícita entre ambas variables. Luego del cálculo de probabilidades, los estudiantes "ya olvidaron" la variable estadística por lo que no la relacionan con la variable aleatoria. La falta de conexión entre estos objetos matemáticos provoca que no se explote la relación entre conceptos claves como el de muestra y población, estimación (como el valor numérico de un estadístico) y parámetro. Ante esta dificultad, los estudiantes no advierten la forma natural de vincular la teoría de probabilidades con la estadística, esto introduce mayores dificultades cuando se tratan los contenidos de inferencia. Acerca de esta problemática, se presentó un trabajo, disponible en las actas del 2do Congreso Internacional Virtual del Enfoque Ontológico Semiótico del Conocimiento y la Instrucción Matemáticos, (2 CIVEOS) que propone un proceso instruccional para favorecer la vinculación entre una variable estadística y su variable aleatoria asociada. En esa propuesta se parte de un problema, en este caso, modelizado por una variable aleatoria binomial, para la comparación posterior de su distribución de probabilidades con la distribución de frecuencias relativas de su variable estadística asociada. Se aplica la ley de los grandes números como una manera de argumentar los resultados de los datos obtenidos al relacionar la estadística y la probabilidad. El acercamiento empírico se produce por la simulación de los datos con el software GeoGebra (https://www.geogebra.org/).

El objetivo de este trabajo es presentar una propuesta de mejora del proceso de instrucción mencionado, para estudiantes que cursan Estadística en la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata, Argentina. Se aplica el marco teórico y metodológico del Enfoque Ontológico Semiótico del Conocimiento y la Instrucción Matemáticos, (Godino, Contreras y Font, 2006) y se realiza la propuesta de mejora a partir del análisis de la idoneidad didáctica de este proceso instruccional en la dimensión epistémica (Godino, 2011). La evaluación de los descriptores junto a la información obtenida de los significados personales de los estudiantes, proporcionan elementos necesarios para la superación de los conflictos semióticos detectados y para el favorecimiento de las relaciones entre los conceptos involucrados, considerados en la nueva propuesta

Descriptores de la idoneidad epistémica en términos de la nueva propuesta

Los descriptores de la idoneidad epistémica están agrupados según los objetos matemáticos primarios del Enfoque Ontológico Semiótico (EOS) (las situaciones problemas, el lenguaje, las definiciones, proposiciones los procedimientos y los argumentos) y están adaptados a la nueva propuesta:

Las situaciones problemas: 1) La muestra representativa y articulada de situaciones de contextualización, ejercitación y aplicación debe incluir mayor cantidad de experimentos aleatorios con sus espacios muestrales correspondientes. 2) La variable aleatoria debe surgir naturalmente de la correspondencia entre los resultados del experimento y el número que se le asigna a cada resultado. Se pretende que el estudiante relacione los resultados del experimento aleatorio con los valores posibles que toma la variable aleatoria al considerar toda la población, y los valores que toma su variable estadística asociada al considerar los resultados de una sola muestra. *El lenguaje*: 3) El uso de diferentes modos de expresión matemática, sea verbal, gráfica, simbólica, y las traducciones y conversiones entre los mismos, debe incluir las distribuciones de frecuencias presentadas para distintos tamaños de muestras. Es conveniente que, para evitar confusiones en la construcción de estas distribuciones, las bases de datos utilizadas sean generadas por los mismos estudiantes, y si las proporciona el profesor, que sean los estudiantes quienes encuentren resultados y resuman la información. 4) Se pretende lograr la comprensión del significado de los elementos que componen estas distribuciones de frecuencias con los tamaños de muestras respectivos, para su comparación posterior con la distribución de probabilidades correspondientes. *Las reglas (Definiciones, proposiciones, procedimientos):* En el proceso de instrucción implementado, no han sido especificadas completamente las reglas, por lo que se propone agregar situaciones donde los estudiantes generen o negocien definiciones, proposiciones o procedimientos. Entre ellos: 5) el manejo del GeoGebra para la simulación de experimentos aleatorios y la recopilación de información. 6) los conceptos de frecuencia relativa y probabilidad, 7) la aplicación de la ley de los grandes números, 8) el reconocimiento de los sucesos mutuamente excluyentes, de los sucesos independientes y el cálculo de las probabilidades respectivas en cada caso, 9) la construcción de las tablas de frecuencias y de frecuencias relativas en forma manual y con un software como el GeoGebra, por ejemplo. 10) Diferenciar el número de veces que se produce el experimento aleatorio de los tamaños de muestras elegidos y de los valores que toma la variable. Para los otros objetos matemáticos, referidos a *los argumentos* y a *las relaciones* entre problemas, definiciones, y proposiciones, se propone relacionarlos y conectarlos al vincular la teoría de probabilidades con el análisis de datos.

En consecuencia, para lograr una mejora en *la idoneidad epistémica* del proceso de instrucción implementado, el objetivo es ampliar el número de situaciones problema de la ingeniería donde los estudiantes de estas carreras tengan más situaciones de argumentación, de comprobación y de justificación, en una muestra representativa y articulada de situaciones de contextualización, ejercitación y aplicación. (Por razones de espacio, se plantean 2 situaciones). Se incluyen experimentos aleatorios con sus espacios muestrales correspondientes, donde deben identificar las variables de Bernoulli y binomial en cada caso. Para la comprobación de resultados, el acercamiento empírico se produce por la simulación de los datos con GeoGebra al comparar las distribuciones de frecuencias relativas y de probabilidad y los estadísticos con sus parámetros respectivos. La justificación teórica es proporcionada por la demostración de la ley de los grandes números, a partir de la desigualdad de Tchebycheff.

Esquema de la propuesta didáctica

Luego de aplicar los descriptores de la idoneidad epistémica al proceso de instrucción implementado, se presenta la nueva estructura que esta propuesta didáctica plantea. Parte de un problema de la ingeniería que se resuelve de dos maneras: una, aplicando el análisis de datos, a través de una enseñanza de la estadística con proyectos, y la otra, aplicando la teoría de probabilidades. Se establece un paralelismo entre estos dos contextos con actividades que las relaciona, como se muestra en el Gráfico 1 y como se ejemplifica en las dos situaciones problemas que se describen a continuación.

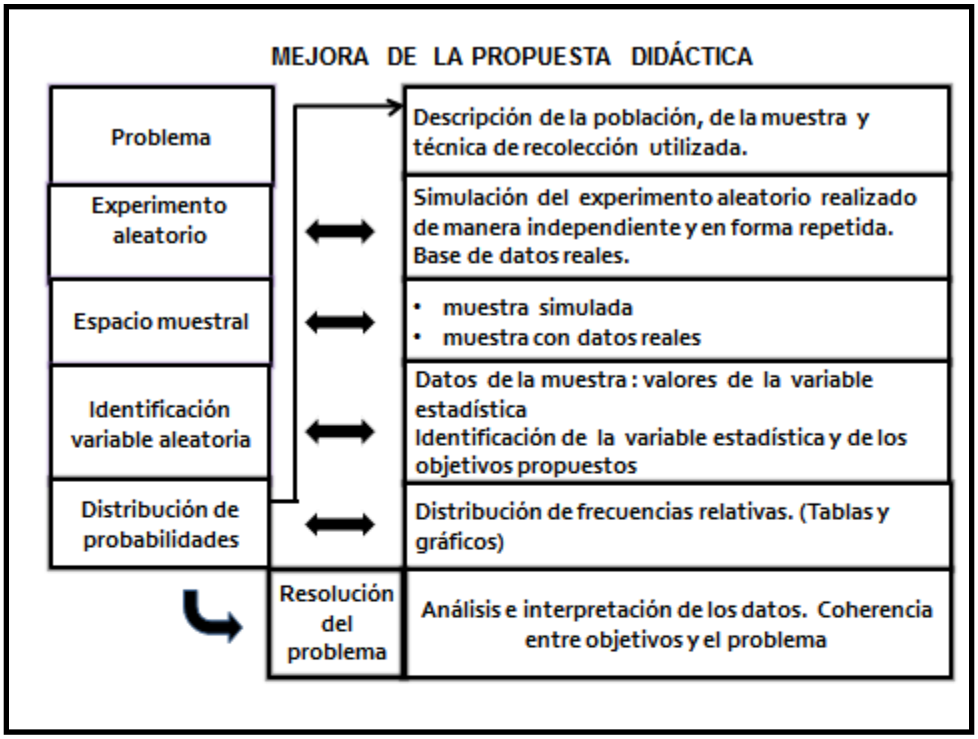

Gráfico 1. Estructura de la propuesta didáctica mejorada.

Situación problema1

a. Observa la tabla siguiente que representa la hoja de registro de artículos defectuosos para una embotelladora X. Define y clasifica las variables intervinientes.

Tabla 1. *Hoja de registro de artículos defectuosos para una embotelladora*

Fecha de inspección: Fecha de fabricación: Nombre del inspector:	Número de lote:30 Número de artículos inspeccionados:1000 Observaciones: muchos envases sucios	
Tipo de defecto	Frecuencia	Subtotal
Nivel fuera de especificaciones. Envases sucios Botellas sin tapas. Envases vacíos. Envases rotos Acabado defectuoso.	IIIII IIII IIIII I IIIII III IIIII II IIIII IIIII IIIII III IIIII IIIII	9 6 8 7 18 10
	Total	58
Total de rechazado		43

a) ¿Cuál es el experimento aleatorio del que surgen estas variables?
b) El resumen de la hoja de registro indica que se inspeccionaron 1000 envases y hubo 43 rechazados. Considera esta información como la de una población. Simular con GeoGebra esta población real dada por un lote de N = 1000 envases con un 4,3% de defectuosos. Para ello:

c) Utilizar el comando =BinomialAleatorio[n,p] de GeoGebra considerando los parámetros n = 1 y p = 0,043. Esta elección simula efectuar *una* inspección (n = 1) que tiene dos resultados posibles. Estos dos resultados dan lugar a una variable, llamada variable de Bernoulli. Si el envase es defectuoso, la variable toma el valor 1, con probabilidad de defectuoso p = 0,043 y si el envase no tiene defectos, la variable toma el valor 0, con probabilidad 1- p = 0,057.

d) Para aumentar el número de inspecciones a 1000, arrastrar el mouse en la planilla de cálculo, desde la casilla B1 a B1000. Se genera una población de unos y ceros de tamaño 1000. Luego listar la población de A1 a A1000. Cada inspección o prueba puede pensarse como un experimento aleatorio repetido e independiente, que dio como resultado un uno o un cero. La suma de variables de Bernoulli define otra variable: "la variable binomial" de parámetros n = 1000 y p = 0,043. (De ahí el nombre del comando de GeoGebra).
e) Seleccionar la población y crear una lista. Nombrarla "Población"

f) Construir la tabla de frecuencias relativas seleccionando el comando TablaFrecuencias[<Lista de datos brutos>, <Factor de escala (opcional)>]. En "lista de datos brutos", se escribe el nombre de la lista creada: Población y en el factor de escala opcional, se escribe 1/1000 porque 1000 es el tamaño de la población, y el factor de escala indica que la tabla es de frecuencias relativas. Si se considera como factor el número 1, la tabla que devuelve el programa es la de frecuencias absolutas.

g) Construir la gráfica de la distribución de frecuencias seleccionando el comando Barras[<Lista de datos brutos>, <Ancho de barras>, <Factor de escala vertical (opcional)>] .Como es una variable discreta, el gráfico correspondiente es el de bastones, por lo que el ancho de barras debe ser cero.

h) Calcular los parámetros poblacionales seleccionando los comandos: Media[<Lista de datos brutos>] ; Mediana[<Lista de datos brutos>]; Moda[<Lista de números>] ; Varianza[<Lista de datos brutos>] y analizar simetría y variabilidad de la población.

i) Seleccionar muestras aleatorias con reposición, de tamaño n = 10, 20, 30, 50 y 80 de la población binomial. Para ello utilizar el comando =Elemento[Población,AleatorioEntre[1,1000]]

j) Tabular y representar gráficamente para cada tamaño de muestra, la variable estadística número de envases defectuosos. Trabajar con frecuencias absolutas y relativas.

k) Calcular las estimaciones de los parámetros de las muestras anteriores y analizar simetría y variabilidad en cada muestra. ¿Encuentras buenas aproximaciones de cada estimación con su parámetro poblacional respectivo? De ser así. ¿En qué casos?

l) Obtener la distribución de probabilidades de la variable aleatoria binomial "número de envases rechazados" de esta población de tamaño N =1000 y proporción de envases rechazados P= 0,043.

Para ello, selecciona la planilla de cálculo de GeoGebra, abre la ventana de cálculos de probabilidad y elige la distribución binomial con esos parámetros.

m) Comparar los valores que toma la distribución de probabilidades de esta variable aleatoria con la distribución de frecuencias relativas de su variable estadística asociada, para los distintos tamaños de muestras obtenidas en h). ¿En qué casos hay mejores aproximaciones?

n) Verificar la desigualdad de Tchevycheff con las variables "número de envases rechazados" de esta población de tamaño N = 1000 y "proporción de envases rechazados" para p = 0,043.

o) Demostrar el teorema de Bernoulli, la ley de los grandes números, para un suceso A. ¿A partir de qué valor de n la frecuencia relativa de un suceso converge probabilísticamente a la probabilidad de dicho suceso?

p) ¿Cuál debiera ser el número de experimentos que debemos realizar para que la frecuencia relativa de un suceso sea una buena estimación de la probabilidad de ese suceso?

q) ¿Cuál debiera ser el número de experimentos que debemos realizar para que la frecuencia relativa de un suceso sea una buena estimación de la probabilidad de ese suceso?

Situación problema 2

En cualquier sistema de comunicaciones, la señal que se recibe difiere de la señal transmitida debido a dificultades sufridas en la transmisión. En las señales analógicas, estas dificultades pueden degradar la calidad de la señal. En las señales digitales, se generarán bits erróneos: un 1 binario se transformará en un 0 y viceversa.

Sabiendo que la probabilidad de generar un bit erróneo es 0,4 y que sobre un canal telefónico se transmiten datos binarios.

a) ¿Cuál es el experimento aleatorio?
b) ¿Cuáles son los valores que toma la variable asociada al experimento aleatorio si se efectúa: 1) una sola transmisión? 2) cinco transmisiones?
c) Escribe el espacio muestral asociado al experimento aleatorio de transmitir cinco señales y observar el número de señales erróneas que se pueden registrar.
d) Utilizar el comando =BinomialAleatorio[n,p] de GeoGebra para simular una transmisión (n=1) con probabilidad p= 0,4 de generar un bit erróneo. Arrastrar el mouse en la planilla de cálculo, desde la casilla B1 hasta B100. De esta manera se genera una población binomial (de unos y ceros) de 100 transmisiones repetidas e independientes. Seleccionar la población y crear una lista. Nombrarla "Población". Para listar la población de A1 a A100, escribir el número de orden.
e) Construir las tablas de frecuencias absolutas y relativas con el comando TablaFrecuencias[<Lista de datos brutos>, <Factor de escala (opcional)>]. En "lista de datos brutos", se escribe el nombre de la lista creada: "Población" y en el factor de escala opcional, se escribe 1/100 si 100 es el tamaño de la población. En este la tabla es de frecuencias relativas. Si se considera 1 como factor de escala, la tabla que devuelve el programa es la de frecuencias absolutas.

f) Construir la gráfica de la distribución de frecuencias seleccionando el comando Barras[<Lista de datos brutos>, <Ancho de barras>, <Factor de escala vertical (opcional)>] Como es una variable discreta, el gráfico correspondiente es el de bastones, por lo que el ancho de barras debe ser cero.
g) Calcular los parámetros poblacionales seleccionando los comandos: Media[<Lista de datos brutos>] ; Mediana[<Lista de datos brutos>]; Moda[<Lista de números>];Varianza[<Lista de datos brutos>] y analizar simetría y variabilidad de la población.
h) Seleccionar muestras aleatorias con reposición, de tamaño n = 10, 20, 30, 50 y 80 de la población binomial. Para ello utilizar el comando =Elemento[Población,AleatorioEntre[1,100]]
i) Tabular y representar gráficamente para cada tamaño de muestra, las dos variables estadísticas: el número y la proporción de señales erróneas para estas muestras. Trabajar con frecuencias absolutas y relativas.
j) Calcular las estimaciones de los parámetros de las muestras anteriores y analizar simetría y variabilidad en cada muestra. ¿Encuentras buenas aproximaciones de cada estimación con su parámetro poblacional respectivo? De ser así. ¿En qué casos?
k) Construye la distribución de probabilidades de las variables aleatorias asociadas a las variables estadísticas respectivas para los distintos tamaños de muestras. Verifica los resultados con GeoGebra.
l) ¿Cuál es el número de señales erróneas más probable de transmitirse?
m) Compara las distribuciones probabilísticas con las distribuciones de frecuencias relativas. ¿En qué casos hay mejores aproximaciones? Extrae conclusiones relativas a:
n) Los valores que puede tomar una variable aleatoria y su variable estadística asociada.
o) Las frecuencias relativas de la variable estadística con las probabilidades respectivas de su variable aleatoria asociada. ¿Qué papel juega el tamaño de la muestra?
p) La relación entre una variable estadística y su variable aleatoria asociada.

Reflexiones Finales

Entre las ventajas más importantes que surgen de esta propuesta didáctica sobre la vinculación de una variable estadística con su variable aleatoria asociada, se encuentra la contextualización de los datos y la aplicación de contenidos difíciles de abordar, en cuanto a su complejidad, como son los de Inferencia Estadística. Entre esos contenidos se encuentran los de población y muestra, estimador y estimación; estimación y parámetro. En esta vinculación, se trabaja en forma paralela con una población y con las muestras aleatorias extraídas de ella, creando el espacio adecuado para integrar estos conceptos. Se logra una ampliación de los significados involucrados al incorporar prácticas matemáticas relativas al análisis de datos y a la teoría de probabilidades en un ambiente computacional. Este abordaje de contenidos conduce a las distribuciones muestrales – tema central dentro de la inferencia estadística- de una manera natural, adaptando los procesos de enseñanza y aprendizaje comenzando por las prácticas matemáticas utilizadas y descriptas en este trabajo.

Referencias bibliográficas

Figueroa, S. y Aznar, M. (2017). *Significados personales sobre la vinculación entre una variable estadística y su variable binomial asociada en el contexto de un problema*. En J. M. Contreras, P. Arteaga, G. R. Cañadas, M.M. Gea, B. Giacomone y M. M. López-Martín (Eds.), Actas del 2do Congreso International Virtual sobre el

Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos. Disponible en: enfoqueontosemiotico.ugr.es/civeos.html

Godino, J. D., Contreras, A. y Font, V. (2006). Análisis de procesos de instrucción basado en el enfoque ontológico-semiótico de la cognición matemática. *Recherches en Didactiques des Mathematiques*, 26 (1), 39-88.

Godino, J. D. (2011). *Indicadores de la idoneidad didáctica de procesos de enseñanza y aprendizaje de las matemáticas*. Actas del XIII CIAEM-IACME, Recife, Brasil. Disponible en http://www.cimm.ucr.ac.cr/ocs/index.php/xiii_ciaem/xiii_ciaem/paper/viewFile/2679/1140

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* Boletim de Educação Matemática. 24 (39), 431-449.