Un ejercicio sencillo de estadística descriptiva abordado desde un enfoque probabilístico

Félix Núñez Vanegas
Instituto Tecnológico de Costa Rica – Universidad de Costa Rica, Costa Rica.
fnunez@itcr.ac.cr

Giovanni Sanabria Brenes
Instituto Tecnológico de Costa Rica – Universidad de Costa Rica, Costa Rica.
gsanabria@itcr.ac.cr

Resumen: Con el fin de establecer un vínculo entre la estadística descriptiva y el concepto de simulación, se toma un ejemplo con datos agrupados y se calcula el promedio de dos maneras: Usando los datos y la simulación de Montecarlo.

Palabras clave: estadística descriptiva, probabilidad, didáctica, simulación.

1. Introducción

Por lo general, en los cursos de estadística descriptiva, por su naturaleza, los conceptos involucrados en ella se ven en un contexto ajeno al de la probabilidad, dado que la misma está desprovista de situaciones aleatorias. En estadística, se tiene siempre la necesidad de analizar los datos de que se disponen, los cuales constituyen la materia prima de esta disciplina. De acuerdo con los objetivos que se persigan, esta disciplina Estadística, la podemos ver dividida en dos grandes ramas: La estadística descriptiva y la estadística inferencial. De acuerdo con Trejos (2000), en la primera, "se trata de hacer descripciones de los datos, mediante números que resuman la información, cuadros que la presenten adecuadamente y gráficos que sean fáciles de interpretar." Mientras que la estadística inferencial, Trejos (2000) indica que "consiste en inferir o generalizar las propiedades de un todo (llamado población) partiendo de lo observado en una parte de esa población, llamada muestra." Por otro lado, es sabido que en la estadística inferencial, en el proceso de generalización, hay asociada una cierta incertidumbre, y es por ello que sus métodos están basados en la teoría de probabilidades.

No obstante, en una visión integral de educación estadística, la probabilidad debería ser transversal y no verse como algo ajeno a la misma, por lo que hacer un esfuerzo adicional para establecer un vínculo entre ambas ramas de la estocástica, es deseable. Esta visión caracteriza a los Programas de Matemática del Ministerio de Educación Pública de Costa Rica, y en ellos se insiste en el siguiente punto: "…en la acción de aula, se realicen procesos matemáticos, es decir actividades transversales que se asocian a capacidades presentes en cada área para comprender y usar conocimientos, apoyando el desarrollo de la competencia matemática." MEP (2012).

En esa línea, en dichos programas se propone desarrollar cinco formas de acciones cognitivas que pueden generar capacidades: Razonar y argumentar, plantear y resolver problemas, conectar, comunicar y representar, estas acciones cognitivas, corresponden a actividades transversales que deben estimularse cotidianamente.

2. Referentes teóricos

2.1. La teoría de Cuadros de Regine Douady

Douady (1984) indica que los conceptos matemáticos tienen carácter de instrumento y carácter de objeto. El ver un concepto como instrumento para resolver un problema es lo que le da sentido al concepto. En esa línea, Vérgnaud (1990), afirma que, "un conocimiento, si se precia de ser racional, debe ser operatorio."

En dicha resolución, el concepto puede intervenir en uno o varios marcos: geométrico, numérico y algebraico, entre otros. En cada marco el concepto se visualiza en términos de objetos y relaciones, formando significados del concepto en el marco.

El juego de marcos, consiste en establecer correspondencias entre los significados que un mismo concepto adquiere en diferentes marcos. Este juego contribuye a construir la diversidad semántica del concepto, poniendo en evidencia el carácter heterogéneo del conocimiento que varía según el estudiante.

Por lo tanto, Douady (1984) recomienda que para lograr un buen funcionamiento de los conocimientos en los alumnos, el docente debe elegir problemas donde estos intervienen en dos cuadros como mínimo.

Por otro lado, este juego de cuadros da a los docentes nuevas alternativas de abordar un determinado concepto, en el sentido de que, un estudiante que no comprenda la explicación dada a través de un determinado cuadro, pueda pasarse a otro en el que el estudiante se siente identificado. A su vez, dada una situación problema propuesta a un estudiante, este juego de cuadros le da la opción de moverse a través de ellos, dándole una robustez al conocimiento adquirido.

2.2. Simulación de Monte Carlo

De acuerdo con Azofeifa (2004), "La simulación Montecarlo es básicamente un muestreo experimental cuyo propósito es estimar las distribuciones de las variables de salida que depende de variables probabilísticas de entrada." Concretamente, dada una variable aleatoria X, el Método de Monte Carlo consiste en asignar valores a X, de acuerdo con su distribución acumulada y al valor obtenido en un número aleatorio entre 0 y 1.

2.3. Otros referentes

- Estadística descriptiva.
- Probabilidad teórica.

3. Los cuadros conceptuales en un problema de estadística descriptiva

3.1. Cuadro algebraico

La estadística descriptiva desarrolla una serie de fórmulas algebraicas que permiten describir los conceptos involucrados, y de esta forma, los operacionaliza.

3.2. Cuadro geométrico

Además de las fórmulas, la estadística descriptiva brinda diversas representaciones gráficas de los conceptos.

Por ejemplo: leer la media en una gráfica.

3.3. Cuadro probabilístico

Por lo general, un requisito para utilizar la probabilidad como una herramienta para resolver un problema es, desde luego, que el problema posea una situación o experiencia azarosa.

Sin embargo, se pueden presentar problemas desprovistos del azar, donde si aquel que lo resuelve quiere utilizar la probabilidad, debe recrear dentro del problema una situación azarosa, modelando las relaciones de la situación problema con los elementos teóricos de la probabilidad.

Este tipo de situaciones que utilizan la probabilidad como modelo, requiere de un dominio más completo y complejo de la probabilidad, pues ello implica aplicar la probabilidad donde no se ve, construyendo una situación azarosa. En estas situaciones, aplicar el concepto adquirido de probabilidad, implica crear las condiciones necesarias para aplicarlo.

¿Cuál situación problema se debe plantear a los estudiantes? Sobresalen los problemas de dados y monedas, pero ¿será correcto introducir el concepto de probabilidad planteando situaciones al estudiante como que determine la probabilidad de que salga un seis en un dado?

Al respecto, mencionamos que en una ocasión, nos comentaba un ex-estudiante de un curso de probabilidades, que él estaba sorprendido, porque en ese curso lo habían enseñado a calcular probabilidades de ganar en juegos de azar como dados, ruletas, cartas, cálculos que a él no le interesaban, le daban la sensación de estar preparándose para ir a un casino, y su formación cristiana chocaba con eso.

Si bien se puede aprovechar el concepto intuitivo de probabilidad que posee el estudiante, el fin de un ambiente a-didáctico es que el conocimiento a enseñar surja del tratamiento con la situación y no que la misma situación lo mencione.

Entonces, ¿cuáles situaciones problema pueden ser útiles en la enseñanza de la probabilidad? Con base en lo anterior, se proponen tres tipos de situaciones: las situaciones centradas en el cálculo de probabilidades (ampliamente tratadas en los libros de texto), las situaciones sobre toma de decisiones (que serán abordadas en un próximo trabajo) y las situaciones ajenas al azar que utilizan la probabilidad como modelo (a exponer en el presente trabajo).

Así, la modelización probabilística de una situación problema de estadística descriptiva implica la creación de una situación azarosa ficticia.

3.4. Cuadro numérico (y probabilístico)

Este corresponde al cuadro anterior utilizando el método de Montecarlo y la ley de los Grandes Números para estimar probabilidades, con ayuda de la simulación computacional.

4. Un ejercicio sencillo de estadística descriptiva

Esta visión holística de la educación estocástica es muy importante, y el propiciar el desarrollo de estas acciones cognitivas, favorecerá, por ejemplo, ligar el complicado proceso de la modelización con la resolución de problemas. Empero, esta labor no es fácil. En el caso particular de vincular la estadística descriptiva con la probabilidad, en este trabajo, daremos un ejemplo concreto de cómo se puede lograr. Este ejercicio de poner en relieve a la probabilidad, aun cuando se trate de un problema de estadística descriptiva, requiere de un gran esfuerzo, pero dará otra perspectiva de análisis y nuevas revelaciones o conjeturas que enriquecerá algún estudio. Esta acción debe vertebrar todo el proceso de formación estadística y probabilística.

En esa dirección, deberían desarrollarse situaciones de enseñanza que permitan asociar la estadística descriptiva y la probabilidad, e ir generando con ello, experiencias en esa línea.

En ese sentido, con el fin de asociar estas dos ramas de la estocástica, tomamos un problema típico de la estadística descriptiva y hemos formulado las soluciones desde los diferentes cuadros mencionados anteriormente. Para abordarlo desde la probabilidad, dotamos al problema de una situación aleatoria artificial para luego resolverlo desde esa perspectiva. Como ya el problema lo habíamos dotado de una situación aleatoria ficticia, usamos Monte Carlo para simular el comportamiento de la variable, obteniendo más valores.

4.1. El problema

Cien estudiantes de una determinada escuela realizaron la tarea de vender, entre sus amigos y familiares, cinco cartones de un bingo con el fin de recoger fondos para organizar una fiesta a fin de año. La tabla siguiente muestra el número de cartones que vendió cada niño.

Cartones	Niños
0	2
1	6
2	12
3	34
4	32
5	14

En promedio, ¿cuántos cartones vendió cada niño?

4.2. Solución algebraica (estadística descriptiva)

El significado que la media adquiere en este cuadro es la fórmula de media para datos están agrupados.

Notamos que los datos están agrupados, y de la información de la tabla se ve por ejemplo que, 2 niños no vendieron un solo cartón, mientras que 34 niños lograron vender 3. De tal manera que, para obtener el promedio de cartones vendidos por niño, debemos sumar el total de cartones vendidos por los niños y dividir entre el total de niños:

$$\bar{X} = \frac{0 \times 2 + 1 \times 6 + 2 \times 12 + 3 \times 34 + 4 \times 32 + 5 \times 14}{100} = \frac{330}{100} = 3.3$$

De esta manera, el promedio de cartones vendidos por niño es 3.3.

4.3. Solución Geométrica

En esta parte, al hacer el histograma de la variable y realizar en el eje x una misma escala, el estudiante puede aproximar el promedio a través de él.

4.4. Solución probabilística

Crearemos una situación aleatoria artificial.

Situación aleatoria: Elegir al azar un niño de los cien. Se define la variable aleatoria a utilizar: X el número de cartones vendidos por el niño elegido. Note que X es variable aleatoria discreta pues depende de la situación aleatoria. Observar que, el problema planteado, lo hemos transformado en determinar la esperanza de X, es decir, el número esperado de cartones vendidos por niño, o lo que es lo mismo, la esperanza del número de cartones vendido por niño. Así, el significado que la media que adquiere en este cuadro es el de esperanza.

Ahora nuestra variable aleatoria tiene una distribución de probabilidad y podemos ver además que el rango de X es $\{0, 1, 2, 3, 4, 5\}$.

Solución Teórica

Note que la función de distribución de X está dada por

$$f_x(m)\begin{cases}\frac{2}{100} & si\ m=0\\ \frac{6}{100} & si\ m=1\\ \frac{12}{100} & si\ m=2\\ \frac{34}{100} & si\ m=3\\ \frac{32}{100} & si\ m=4\\ \frac{14}{100} & si\ m=5\end{cases}$$

De acuerdo con la distribución de probabilidad de la variable aleatoria X, observamos por ejemplo que,

$$P(X=2)=\frac{12}{100}$$

La esperanza de la variable aleatoria X está dada por

$$E(X)=\frac{2}{100}\times 0+\frac{6}{100}\times 1+\frac{12}{100}\times 2+\frac{34}{100}\times 3+\frac{32}{100}\times 3+\frac{14}{100}\times 5=3.3$$

4.5. Solución numérica probabilística

Vamos a simular el comportamiento de la variable aleatoria X: El número de cartones vendidos por el niño elegido. Para ello, en Excel colocamos la distribución de frecuencias del número de cartones vendidos por los niños. En la siguiente columna colocamos la frecuencia relativa y en la otra columna, la frecuencia acumulada. Generamos ahora intervalos de clases, tal como se muestra en la siguiente tabla:

Cartones	Niños	Frec. Relativ	FrecAcumu.	LímiteInf.	LímSup
0	2	0,02	0,02	0	0,02
1	6	0,06	0,08	0,02	0,08
2	12	0,12	0,2	0,08	0,2
3	34	0,34	0,54	0,2	0,54
4	32	0,32	0,86	0,54	0,86
5	14	0,14	1	0,86	1
Total	100	1			

Ahora agregamos la columna de cartones vendidos por los niños:

Cartones	Niños	Frecuencia	FrecAcumula	Infe	Super	Cartones
0	2	0,02	0,02	0	0,02	0
1	6	0,06	0,08	0,02	0,08	1
2	12	0,12	0,2	0,08	0,2	2
3	34	0,34	0,54	0,2	0,54	3
4	32	0,32	0,86	0,54	0,86	4
5	14	0,14	1	0,86	1	5

En la siguiente columna generamos un numero aleatorio entre 0 y 1, con la función ALEATORIO() de Excel, con el fin de que se busque el valor m cuya probabilidad sea igual a dicho número generado. Así por ejemplo, si el número generado es 0.4, Excel, a través de la función BUSCARV, lo ubicará entre 0.2 y 0.54, y por tanto dirá que el valor m es 3. Lo interpretamos como $\mathrm{p}(\mathrm{X}=3)=0.4$.

Repetimos el experimento unas 1000 veces, y obtendremos muchos valores simulados del número de cartones que vendió cada niño. Luego la esperanza de X será el promedio de esos valores. En nuestro caso obtuvimos 3.306. Y si nos ubicamos en una celda vacía, y tecleamos suprimir, obtenemos nuevos valores y un nuevo promedio, que es muy cercano a 3.3.

5. Conclusiones

El trabajo plantea la solución de problemas de estadística descriptiva por medio de diferentes cuadros conceptuales, según la teoría de Douady, permitiendo la transversalidad de conocimientos.

Así, dado un problema en estadística, puede tener una solución algebraica, geométrica, numérica -probabilística.

La modelación probabilística de un problema en estocástica, no es una cuestión fácil e implica la recreación del problema por medio de una situación azarosa.

El juego de cuadros que propone Douady, permite a los estudiantes tener varias dimensiones de un concepto, y al docente, le brinda la posibilidad de pasarse de cuadro, cuando el estudiante entra en situación de bloqueo, dando así nuevas opciones de establecerlo. De esta manera, ver el problema de estadística descriptiva planteado, desde un enfoque numérico-probabilístico, abre un cuadro nuevo de la dimensión de media, a la vez que evidencia la utilidad de la simulación de Montecarlo.

El abordaje de este tipo de problemas mediante la modelización probabilística, brinda una nueva dimensión semántica del concepto de media.

6. Referencias bibliográficas

Azofeifa, C. E. (2004). Aplicación de la Simulación Monte Carlo en el cálculo del riesgo usando Excel. Tecnología en Marcha, 17(1), 97-109.

Brousseau, G. (1986). Fundamentos y Métodos de la Didáctica de las Matemáticas. Traducción al castellano del artículo "Fondements et méthodes de la didactiques des mathématiques" publicado en la revista Recherches en Didactique des Mathématiques, 7(2):33-115, y realizada por Julia Centeno, Begoña Melendo y Jesús MurilloCarvajal.

Douady, R. (1984). Relación enseñanza –aprendizaje, Dialéctica instrumento – objeto, Juego de marcos. Cuadernos de Didáctica de las Matemáticas Nº 3, IREM de Paris

Espeleta, A. (2002). La enseñanza de la Estadística: una propuesta metodológica. Sistema de Estudios de Posgrado. Maestría Profesional en Planificación Curricular. Universidad de Costa Rica, San José, Costa Rica.

M; Masís, K; Méndez, T. (1998). La estadística y probabilidad en primer y segundo ciclo: tratamiento y propuesta. Memoria del seminario de Graduación presentado para optar por el grado de Licenciatura en Educación Primaria con énfasis en primer y segundo ciclos. Universidad de Costa Rica, San José, Costa Rica.

Ministerio de Educación Pública. (2005). Programas de estudio matemáticas. San José, Costa Rica.

Ministerio de Educación Pública. (2012). Programas de estudio matemáticas. San José, Costa Rica.

Núñez, F. (2008). La Enseñanza y el aprendizaje de la estadística en secundaria: situación actual, aproximación metodológica. Sistema de Estudios de Posgrado. Maestría académica en matemática. Universidad de Costa Rica, San José, Costa Rica.

Núñez, F. (2008). Consideraciones sobre la didáctica de la probabilidad y de la estadística. Memorias del VII CIBEM, Uruguay.

Sanabria, G; Núñez, F. (2010). Una propuesta para introducir el estudio de las probabilidades: Probabilidad Frecuencial. Memorias del III Encuentro Enseñanza de la Matemática, UNED, Costa Rica.

Sanabria, G. (2012). Comprendiendo las probabilidades. Primera. Cartago: Editorial Tecnológica de Costa Rica.

Trejos, J.; Moya, E. (2000). Introducción a la estadística descriptiva. Segunda edición, Sello Latino, San José, Costa Rica.

Vergnaud, G. 1990. "La théorie des champs conceptuels", Recherches en Didactique des Mathémathiques Vol. 10 (23): 133-170.