Desenmañarando la estadística y la probabilidad

Brun Battistini, Dominique

2017-03

Desenmañarando la Estadística y la Probabilidad

D. Brun Battistini (1)

(1) *Universidad Iberoamericana Ciudad de México*
Departamento de Física y Matemáticas
dominique.brun@ibero.mx

***Resumen*—Se muestran ejercicios específicos que han resultado eficaces en hacer que el estudiante universitario logre la comprensión de temas tradicionalmente difíciles para ellos, con la comprensible preocupación de los docentes. En este enfoque, se presentan al estudiante diferentes ejemplos de la vida cotidiana que lo hacen capaz de escribir por sí mismo conclusiones relativas a conjuntos de datos. En particular, se tratan los temas de dispersión, promedio de promedios, muestreo, la relación entre el análisis combinatorio y los cálculos de probabilidad y el teorema de Bayes.**

Palabras clave—conceptualización, aprendizaje activo, experimento educacional.

I. Introducción

El estudio de la Probabilidad representa un cambio de paradigma para el estudiante, suele encontrarse perdido en el contexto: ¿Cómo que si saco una carta roja? ¿De dónde? ¿Por qué es esa la probabilidad? En un principio las conclusiones no parecen lógicas, ni claros los procedimientos.

El docente debe recordar su propia experiencia en un primer curso de este tipo y compartir con sus estudiantes sus inquietudes de entonces; así se crea un vínculo más estrecho con ellos, quienes se sienten libres de preguntar cualquier duda que les surja.

En los planes de estudio vigentes, tanto de ingenierías como de carreras de negocios en la Universidad Iberoamericana Ciudad de México, la primera experiencia de los estudiantes no se trata de dos cursos sino de uno solo que abarca los temas de estadística y probabilidad. El primero también tiene conceptos de difícil apropiación, algunos de los cuales se tratarán más adelante.

La tendencia internacional de los enfoques pedagógicos insiste en que se debe abandonar la figura del docente que lo sabe todo y que abusa del método expositivo. En este artículo se comparten algunos ejercicios y dinámicas de clase que han funcionado para que los estudiantes, en su mayoría de áreas administrativas, logren construir conceptos estadísticos y probabilísticos. La mayoría de las veces esta construcción se ha logrado mediante el método socrático, también conocido como *mayéutica* [1], en conjunto con la pedagogía ignaciana [2]. La intención de este artículo es invitar a los docentes a diseñar y compartir este tipo de dinámicas, para el beneficio de todos los involucrados.

En la sección II se presenta el caso de temas socorridos por la Estadística Descriptiva: dispersión y promedios. En la sección III se abordan los temas de teoría de la probabilidad. En la sección IV se presentan resultados y conclusiones. Las referencias se encuentran en la sección V.

II. Algunos temas de la estadística descriptiva

En esta sección se describe la forma de abordar el tema de dispersión de datos, que usualmente se ve inmediatamente después de las medidas de tendencia central (sub-sección A) y un tema que parece se explica por sí solo, puesto que no aparece en los textos, la media de las medias aritméticas (o promedio de promedios), en la sub-sección B.

A. Dispersión

Tal y como lo remarca la pedagogía ignaciana [2], en el salón de clase resulta indispensable contextualizar los temas en términos de expresiones, palabras y conceptos de la vida cotidiana de los estudiantes. Uno de dichos temas son las calificaciones. En los últimos años se les ha presentado una tabla que representa diferentes grupos de una misma materia, con el mismo número de alumnos pero con distribuciones diferentes de calificaciones (Tabla I). Se les pide calcular la media aritmética, la varianza y la desviación estándar.

TABLA I
Dispersión

	Gpo 1	Gpo 2	Gpo 3	Gpo 4	Gpo 5
	6	10	10	9.9	9.99
	7	10	10	10	10
	8	10	10	10	10
	7	2	10	10	10
	6	2	10	10	10
Media	6.8	6.8	10	9.98	9.998
Varianza	0.70	19.20	0.00	0.00	0.00002
Desv std	0.84	4.38	0.00	0.04	0.00447

Los alumnos suelen identificar rápidamente la columna de cero dispersión, luego se regresan a las dos primeras columnas y ya se les hace muy fácil describir qué sucede en las últimas dos. En este sentido es sumamente importante que los estudiantes escriban la observación y el análisis de cada una de las columnas. Y más importante aún es la exhaustiva retroalimentación del docente.

Por otro lado, se considera que no se hace suficiente hincapié en la importancia de la dispersión en el tema de la representatividad de las medidas de tendencia central. Son particularmente útiles los diagramas que presentan ciertos textos (véase por ejemplo [3] y [4]) donde comparan una curva leptocúrtica y una meso o platicúrtica, que representan, por decir un tema, el tiempo que tarda una compañía de mensajería en entregar un paquete de México a Puebla. Lo ideal es exhortar al estudiante a pensar qué le conviene más a la compañía, y naturalmente abordar el tema de la medida de tendencia central acompañada de su dispersión.

B. Promedio de promedios

Este tema no se toca en los temarios ni en los libros de texto convencionales. Sin embargo, es uno de los errores más frecuentes que se cometen en el ámbito profesional. Nuevamente, sugiero el tema de las calificaciones, esta vez tomando como datos las calificaciones de las diferentes materias del historial académico de un estudiante. Este tipo de cálculos los suelen hacer los estudiantes, muchas veces de manera errónea, al calcular su promedio de calificaciones: calculan el promedio de cada periodo académico y lo dividen entre el número de periodos, en lugar de obtener la media de todas las calificaciones juntas, esto es fundamental para cuidar el promedio de beca, por ejemplo.

Se puede presentar la situación como en la Tabla II, donse se indican las calificaciones de cada materia durante tres periodos académicos.

TABLA II
PROMEDIO DE PROMEDIOS

Periodo 1	Periodo 2	Periodo 3
9	7	9
7	8	6
8	7	10
6	8	8
6	6	
6	9	
COMPARACIÓN DE LOS PROMEDIOS DE LOS TRES GRUPOS		

Se les pide a los estudiantes que calculen la media aritmética de cada grupo (7.0, 7.5 y 8.25) y luego la media aritmética de las tres medias de grupo (7.583). Por otro lado, se les pide calcular la media aritmética de todos los datos juntos (7.5). Se dan cuenta de que no es igual. Luego se agregan dos calificaciones en el tercer periodo (pueden ser dos dieces para un resultado significativamente diferente) y se les pide volver a calcular la media de las tres medias (7.77) y compararla con la media de todos los datos juntos (7.77) ¡Ahora los resultados sí coinciden! Se les pide a los estudiantes que indaguen la razón de esos resultados y que redacten una conclusión al respecto.

III. ALGUNOS TEMAS DE LA TEORÍA DE LA PROBABILIDAD

En muchos de los temarios de las primeras asignaturas de Estadística y Probabilidad, se comienza con Estadística Descriptiva, luego se ven conceptos básicos de Probabilidad y se termina con Estadística Inferencial. El preámbulo de esta última es el muestreo, tema al que curiosamente no suele brindarse la importancia debida y que se estudiará en la sub-sección A. Otro tema curioso es la relación entre el análisis combinatorio y el cálculo de probabilidades, mismo que se estudiará en la sub-sección B. Finalmente, la sub-sección C se dedicará al muchas veces problemático teorema de Bayes.

A. Muestreo

En las referencias [3] y [5] se presentan diferentes ejemplos en los que, con un tamaño de población relativamente pequeño (basta con N=7) y un tamaño de muestra de n=2 se puede construir la distribución muestral de medias. La variable aleatoria puede ser, por ejemplo, número de pizzas entregadas por cada repartidor. Usando el análisis combinatorio se obtiene el número total de muestras de tamaño 2, de una población de 7 elementos ($C_2^7 = 21$). Se calculan las medias de cada una de las 21 muestras y se compara la media de la distribución muestral de medias con la media de la población, concluyendo que son iguales ($\mu_x = \mu$). Es factible que calculen las desviaciones estándar de ambas distribuciones, concluyendo que la dispersión de la distribución muestral de medias es menor que la dispersión de la distribución normal. Una desventaja es que con estos valores pequeños no siempre se cumplirá la relación $\sigma_x = \sigma/\sqrt{n}$. O bien, pueden los estudiantes graficar ambas distribuciones, con la misma escala en el eje x y podrán comparar gráficamente la dispersión entre ambas. Calcular el rango, que es mucho más directo, los lleva a la misma conclusión. En este punto es conveniente preguntarles a los estudiantes ¿Por qué es recomendable una dispersión menor? Y se pueden recordar algunos ejercicios realizados al estudiar el tema de medidas de tendencia central y de dispersión. Estas gráficas permiten, además, comparar la forma de ambas distribuciones; y de nuevo, preguntar a la clase, ¿en qué nos favorece que la distribución sea acampanada?

El paso siguiente consiste en que el estudiante comprenda la diferencia entre las ecuaciones (1) y (2), con las cuales se calcula la variable estandarizada *z*.

$$z = \frac{x - \mu}{\sigma} \quad \text{ec}(1)\,; \qquad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad \text{ec}(2).$$

Finalmente se culmina con un concepto muy práctico e ilustrativo del teorema del límite central, fundamental en la Estadística. Con esta dinámica, haciéndolo por lo menos con dos ejercicios, los estudiantes son capaces de enlistar las tres propiedades fundamentales que conducen al teorema:

- igualdad de las medias entre ambas distribuciones,
- menor dispersión de la distribución muestral de medias y,
- tendencia de forma acampanada de la distribución muestral de medias,

de tal manera que ellos mismos lo postulan.

B. El análisis combinatorio y los cálculos de probabilidad

Es sorprendente que ni en los temarios más comunes ni en los libros de texto más populares el orden de estos temas parece tener lógica: primero se ve todo lo referente a la probabilidad, conceptos, enfoques, axiomas, probabilidades marginales, conjuntas, condicionales, teorema de Bayes. Al final del capítulo o del tema, aparece la sección de análisis combinatorio, también llamada reglas de conteo, que parece no tener ninguna relación con probabilidad.

Muchos profesores de Estadística y Probabilidad han cambiado el orden de estos temas en sus clases. Al analizar los diferentes enfoques del concepto de probabilidad e introducir la probabilidad clásica, se resalta la importancia de contar el número de éxitos y el número total de probabilidades, para lo cual se debe aprender a contar y entonces se introduce el análisis combinatorio.

Un ejemplo típico es el siguiente: Se tienen doce chocolates blancos y quince negros, ¿Cuál es la probabilidad de formar paquetes de cinco chocolates con dos blancos y tres negros?

Partimos de la definición clásica de probabilidad:

$$P = \frac{\#\ de\ éxitos}{\#\ total\ de\ posibilidades}\, ec(3),$$

y se hace uso del análisis combinatorio para calcular numerador y denominador: $P = \frac{C_2^{12}\ C_3^{15}}{C_5^{27}}$,

el resultado es 0.3720, es decir “existe una probabilidad de 37.20 % de formar paquetes de cinco chocolates con dos blancos y tres negros”.

C. El teorema de Bayes

Muchos profesores reportan la dificultad que representa el teorema de Bayes para los estudiantes. Como primera aproximación es recomendable contextualizar el trabajo de Thomas Bayes con respecto de los individuos que padecen una cierta enfermedad (E) y los análisis o pruebas positivas de la misma (A). Previamente se debió haber trabajado el concepto de probabilidad condicional. Diferenciar muy bien entre los condicionales P(A/E) y P(E/A), escribiendo o diciendo en voz alta el significado de cada uno (en este caso es muy importante la verbalización del concepto). Para afianzar la diferencia, es muy importante imaginar el experimento. Esto es, para P(A/E), se elige a todos los enfermos, se les aplica la prueba de análisis, y se escogen a los que dieron positivo. A diferencia de P(E/A), donde se elige a todos los que dieron positivo a la prueba de análisis y se escogen a los que están enfermos.

En los ejercicios de libro de texto que ejemplifican el uso del teorema de Bayes, hay un mundo de posibilidades. El que nos ha parecido más práctico y transparente para los estudiantes es aquel que sigue los siguientes pasos (ver referencias [3] y [5]):

1) Leer y re-leer el problema
2) Responder a la pregunta ¿Qué conozco del problema?
3) Organizar los datos en probabilidades simples y condicionales
4) Tener claro qué preguntan y expresarlo matemáticamente
5) Elaborar un diagrama de árbol
6) Generar la tabla de probabilidad a posteriori o probabilidad inversa
7) Identificar las respuestas solicitadas
8) Expresar las respuestas con palabras.

Tomando como ejemplo el problema de los enfermos [4], podemos seguir los pasos:

¿Qué conozco?

Probabilidades simples

P(E) =0.03 P(E’) =0.97

Donde E’ es el complemento de E.
Probabilidades condicionales

P(A/E) =0.90 P(A/E’) =0.02

Aquí es importante que el estudiante haga la distinción entre ambos condicionales. ¡Que lo verbalicen!, de manera oral o escrita.

¿Qué me preguntan?

P(E/A) =?

El diagrama de árbol se presenta en la Figura 1.

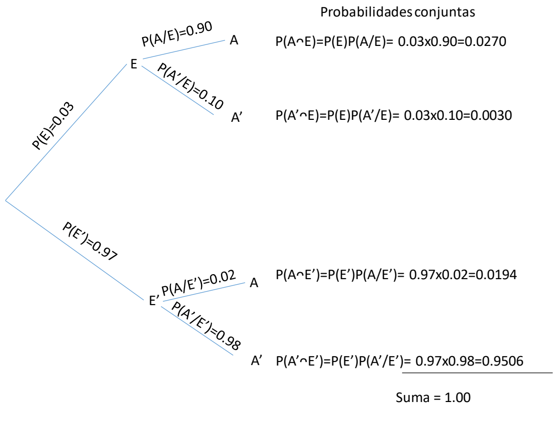

Fig. 1. Diagrama de árbol.

El cálculo de la probabilidad a posteriori se presenta en la Tabla III. Básicamente se ordena la información contenida en el diagrama de árbol, sólo que se sugiere realizarlo para sólo uno de los eventos de la segunda rama: o para análisis positivo o para el negativo; en este caso, como la pregunta tiene que ver con el análisis positivo, se realiza la tabla para el evento "A".

TABLA III
TABLA DE PROBABILIDAD A POSTERIORI PARA "A"

Prob. sim-ples	**Prob. condicio -nales**	**Prob. con-juntas**	**Prob. a Poste-riori**
P(E) =0.03	P(A/E)= 0.90	P(**A**⌒E) =0.0270	P(E/A)= 0.5819
P(E') =0.97	P(A/E')= 0.02	P(**A**⌒E') =0.0194	P(E'/A)= 0.4095
	Suma	**P(A)= 0.0464**	**0.99~ 1**

Es importante resaltar que en la tercera columna de la Tabla III se obtiene un sub-producto y es la probabilidad del evento "A", resultado de la suma de las probabilidades conjuntas de A con E y con E'. Por otro lado, la probabilidad inversa o a posteriori se calcula de manera muy sencilla, dividiendo el elemento de la celda de la izquierda entre la suma de la columna de la izquierda. En este punto cabe expresar en su forma conocida, el teorema de Bayes y mostrar a los estudiantes que la están usando, de una manera lúdica.

$$P(E/A) = \frac{P(A/E)P(E)}{P(A/E)P(E) + P(A/E')P(E')}$$

$$= \frac{(0.9)(0.03)}{(0.9)(0.03) + (0.02)(0.97)} = 0.58 \; ec(4),$$

donde finalmente se retoma la definición operacional de la probabilidad condicional:

$$P\left(\frac{E}{A}\right) = \frac{P(E\frown A)}{P(A)} \; ec(5),$$

El punto importante es que el reverendo Thomas Bayes no conocía P(A), esto es, la probabilidad de que el análisis resultara positivo ni la probabilidad de que un individuo se encuentre enfermo, siendo que la prueba resultó positiva.

IV. RESULTADOS Y CONCLUSIONES

En las secciones anteriores se mostraron una serie de ejercicios para realizar en clase, de una manera interactiva con los estudiantes universitarios de las materias de Estadística y Probabilidad básica a nivel superior. Los resultados han sido efectivos, medidos a partir de los resultados del examen departamental, que es un examen único que se aplica a todos los grupos de una misma materia. La mediana del grupo pasó de 4.5 a 7.0 a lo largo de tres periodos académicos.

Se piensa que el cambio en el desempeño de los estudiantes se basó tanto en el tipo de ejercicios y en la secuencia de los temas por un lado, y por el otro lado en un ambiente de clase y abordaje que conjunta el método socrático y la pedagogía ignaciana, en el sentido de plantear situaciones que los estudiantes deben analizar y de las cuales deben establecer sus propias conclusiones, todo dentro de un contexto que les sea familiar, como son las calificaciones, las encuestas políticas y las encuestas de calidad en el servicio, entre muchas otras. La Estadística tiene la gran ventaja de poder aplicarse casi en cualquier ámbito por lo que no será difícil hacer la inmersión en algún tema que los estudiantes sientan próximo.

Por otro lado, en este artículo se ha insistido en la verbalización de las respuestas. Muchos psicólogos han analizado la relación entre el proceso cognitivo y el habla (véase por ejemplo las referencias [6] y [7]). Es por ello que después de realizar cada ejercicio se le pide al estudiante que exprese oralmente la respuesta, y que la escriba. Se cree que eso garantiza la adquisición del conocimiento.

Un consejo importante es el no incluir como anexo el análisis combinatorio en un curso de probabilidad sino volverlo parte integral del cálculo probabilístico.

V. AGRADECIMIENTO

Agradezco a mis alumnos quienes a lo largo de estos años me han permitido mejorar el abordaje de algunos temas.

VI. REFERENCIAS

[1] Brunschwig, "Socrate (-469-399) et les écoles socratiques", Encyclopædia Universalis [en línea], consultado el 15 marzo 2017. URL : http://www.universalis.fr/encyclopedie/socrate-et-ecoles-socratiques/

[2] Metts, R., *Ignacio lo sabía: la pedagogía jesuita y las corrientes educativas actuales*. México: ITESO, 1999.

[3] Lind, D. y W. Marchal, *Estadística Aplicada a los Negocios y a la Economía*. México: Mc Graw Hill, 2015.

[4] Berenson, M. y Levine, D. *Estadística Básica en Administración: conceptos y aplicaciones*. México: Prentice Hall, 1996, pp. 225-228.

[5] Miller, I. y Freund, J., *Probabilidad y Estadística para Ingenieros*. México: Prentice Hall, 1987.

[6] J. Piaget, *El lenguaje y el pensamiento del niño pequeño*. Barcelona: Ediciones Paidós Ibérica, 1984.

[7] L. Vygotsky, *Pensamiento y lenguaje*, en Cognición y desarrollo humano vol. XX. Barcelona: Ediciones Paidós, 1995.