



1 de septiembre de 2014 | Vol. 15 | Núm. 9 | ISSN 1607 - 6079

ARTÍCULO

¿QUÉ DIFERENCIA HAY ENTRE LA PROBABILIDAD Y LA ESTADÍSTICA?

José Luis Ángel Rodríguez Silva

¿QUÉ DIFERENCIA HAY ENTRE LA PROBABILIDAD Y LA ESTADÍSTICA?

Resumen

En este artículo se detalla la distinción que existe entre la probabilidad y la estadística, y porqué es importante precisar la distinción entre ellas. Además de un contexto conceptual en el cual se especifican la definición, objeto y método de ambas disciplinas, se ofrece un par de ejemplos que ayudan a clarificar la diferencia que existe entre estas dos áreas. Con lo anterior, se aporta a la comunidad académica interesada en disipar la confusión constante que hay entre la probabilidad y la estadística y así minimizar la posibilidad de cometer errores operativos en la aplicación de ambas ramas de estudio.

“ A la estadística le es indispensable el aparato probabilístico para poder expresar sus ideas y resultados. ”

Palabras clave: matemáticas, estadística, método de estudio, objeto de estudio, probabilidad.

WHAT IS THE DIFFERENCE BETWEEN PROBABILITY AND STATISTICS?

Abstract

In this article it is detailed the distinction between probability and statistics, and it is depicted the importance of such distinction. Besides, in a contextual sense, it is specified the definition, object and method of both disciplines and it is provided some examples to clarify their difference. It is intended to contribute to the interested academic community to dispel this constant confusion between probability and statistics to minimize the possibility to perform operative mistakes in the application of both study branches.

Keywords: *matemáticas, statistics, study method, study object, probability, study purpose.*

¿QUÉ DIFERENCIA HAY ENTRE LA PROBABILIDAD Y LA ESTADÍSTICA?

Introducción

En las aplicaciones científicas y tecnológicas de la actualidad se ha vuelto indispensable la utilización, en alguna medida, de la probabilidad y la estadística, motivo por el cual estas dos ramas de conocimiento ocupan un lugar relevante en la formación académica a nivel superior. Existen numerosos ejemplos que ilustran esta afirmación.

Diariamente, en diversos noticiarios, se da la numeralia más importante con respecto al comportamiento de diversos **indicadores macroeconómicos**, como la tasa de inflación, el desempleo, la cotización peso-dólar, el índice de confianza del consumidor, entre muchos otros. Un conocimiento básico de estadística para conocer e interpretar dichos datos resulta indispensable, por ejemplo, en el desarrollo de alguna gráfica que nos indique el comportamiento de un determinado indicador a lo largo de algún intervalo de tiempo que sea de nuestro interés.

En cuanto a **fenómenos meteorológicos**, para muchos, es de gran interés conocer la probabilidad de lluvia durante un día en específico o en un periodo de una semana. Las implicaciones prácticas son evidentes puesto que así podemos tomar las previsiones que sean necesarias. Para entender tal información, es necesario conocer, aunque sea mínimamente, algunos elementos de probabilidad.

Finalmente, en ciertos trabajos como los que realiza el Instituto Nacional de Estadística y Geografía (INEGI) en México, como los **estudios de población**, se tienen resultados que implican comprobar ciertas conjeturas (hipótesis) o establecer el valor estimado de ciertas características de nuestro interés, lo que puede llevar a desarrollar los así llamados "intervalos de confianza". Todo esto implica un conocimiento mixto tanto de probabilidad y estadística para poder tener una interpretación correcta y explotar mejor tales resultados.

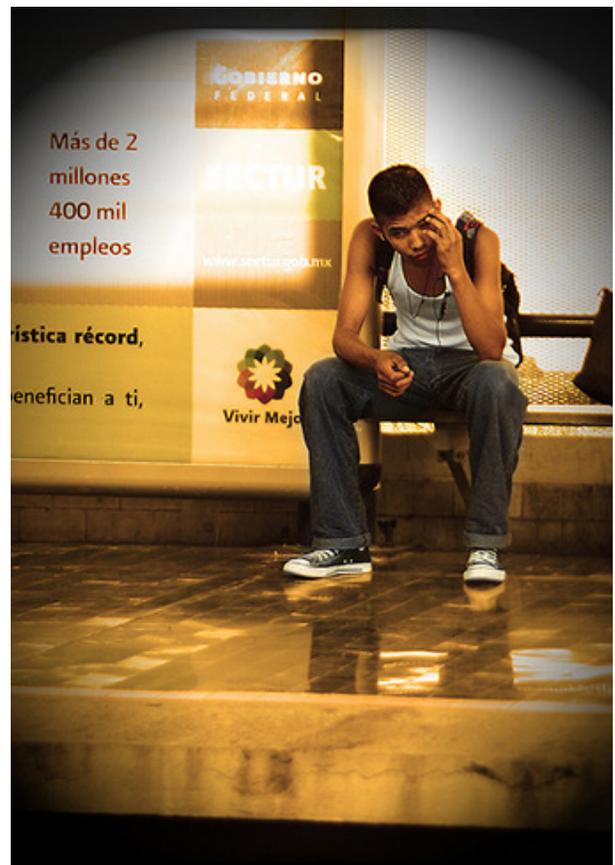


Imagen: Daniel Cristán

Confusiones entre probabilidad y estadística

Existen, sin embargo, confusiones acerca de la diferencia, similitud y nexo entre la probabilidad y la estadística, y con mucha frecuencia se piensan como conceptos sinónimos o equivalentes. Este asunto no es de menor importancia y adquiere una gran relevancia cuando esto se traduce en potenciales errores conceptuales y operativos, ya que si no conocemos cual es el nicho y la gama de aplicación de cada una de estas áreas se pueden cometer importantes errores al respecto.

Un ejemplo es el siguiente: supongamos que se llevará a cabo un ejercicio de acopio de información, mediante un proceso de aplicación de encuestas a una muestra de estudiantes de la UNAM con la finalidad de conocer su punto de vista acerca de la implementación de una nueva política académica, por ejemplo, subir el estándar académico de una nota mínima aprobatoria de 7 a 8. El proceso del diseño de muestreo es una tarea en estricto sentido probabilística, y dependiendo del esquema que se proponga (muestreo aleatorio simple, muestreo estratificado, muestreo por conglomerados, muestreo en varias etapas, etc.), será como se acopie la información, lo cual impacta a la calidad misma de los datos.

Ahora bien, la disposición resumida de los resultados de la encuesta, dados en tablas, gráficas, u otros formatos, es de naturaleza eminentemente estadística. Si el analista, estudiante o el panel de encargados de dicho estudio no tuvieran un marco conceptual claro acerca de estas diferencias, podría por ejemplo, cometerse el error de tratar de aplicar herramientas de naturaleza estadística en el proceso de diseño de muestras (cuando todavía no es el tiempo para ello) o tratar de aplicar resultados de la probabilidad en pasos que ya no la requieran, y que incluso hasta entorpecerían el resultado integral del estudio.

Imagen: Kpta Estrategia Educativa (2014)



Las preguntas clave que contestamos en este estudio son: ¿cuál es la verdadera distinción entre una y otra disciplina, dónde comienzan, en qué punto terminan y cómo se relacionan? El objetivo es contribuir con la comunidad académica a disipar las confusiones frecuentes que hay en la utilización de estos dos términos.

Conceptos clave

La **probabilidad** es una **rama de las matemáticas**, cuyo objeto de estudio son **variables aleatorias** (que son valores que dependen básicamente del azar o de la posibilidad de que puedan o no ocurrir), que **busca establecer las características y propiedades matemáticas** (definiciones, teoremas y consecuencias) de tales variables. Su método es de tipo deductivo, esto es, partiendo de ciertas definiciones y propiedades básicas establecidas de antemano, conocidas como **axiomas**, se van deduciendo las propiedades de los objetos de interés, y cuyos resultados se establecern como teoremas, que son proposiciones ciertas o verdaderas que pueden y deben ser demostradas. En este sentido, la probabilidad es afín a otras ramas de las matemáticas tales como el álgebra, el análisis matemático, la geometría o la topología.

Por su parte la estadística es otra rama de las matemáticas cuyo objeto de estudio son los datos, es decir valores o atributos que tienen los objetos del estudio de nuestro interés, por ejemplo: personas, máquinas o números de una acción financiera, siendo su propósito describir la estructura de los datos, mediante el desarrollo de números resumen, tales como promedios o varianzas, así como mediante el establecimiento de cierto tipo de gráficas, como histogramas o diagramas de pastel, que son materia de estudio de la estadística descriptiva, y realizar inferencias, es decir generalizaciones, sobre alguna variable de interés de una muestra hacia la población analizada. Por ejemplo, si tomamos una muestra de estudiantes de la UNAM para conocer el ingreso promedio de todos los alumnos, en este caso se lleva a cabo un estudio de estadística inferencial.

Imagen: Jorge Franganillo



El método de la estadística se basa en el proceso científico, el cual es básicamente inductivo, que va hasta cierto punto en el orden opuesto al deductivo; este método parte de un proceso de observación, generación de hipótesis, experimentación y potencialmente el pronóstico sobre el comportamiento del fenómeno bajo interés (una finalidad adicional podría ser el establecimiento de leyes o incluso de teorías, de validez general en amplios campos de aplicación).

De este modo, una distinción clave entre la probabilidad y la estadística es que la primera usa el método deductivo, mientras que la segunda es un campo de estudio fáctico y experimental, y se basa en un proceso inductivo, el cual debe de contrastarse en todo caso con la experiencia o la experimentación.

Para discutir con mayor precisión estas ideas, se considerará la definición del matemático ruso Andréi Nikoláyevich Kolmogórov la cual es aceptada actualmente por la mayoría de los expertos en el ramo. Kolmogórov definió un espacio de probabilidad recurriendo a tres conceptos: a) el conjunto de todos los posibles resultados de un experimento o "espacio muestral" (un ejemplo puede ser el conjunto de todos los mexicanos, ya sea que residan en México o en el extranjero); b) el conjunto de todos los posibles combinaciones de objetos tomados del espacio muestral y sobre los cuales tenemos cierto interés (recurriendo al ejemplo de los mexicanos, aquí podríamos estar interesados en saber, por ejemplo, cuantas mujeres mexicanas hay en total); c) finalmente, el establecer que tan probable o factible es que resulte una determinada combinación de elementos en la población, de este modo, según datos del INEGI, al 12 de julio del 2010 había 112'336,538 habitantes, de los cuales 57'481,307 eran mujeres, por lo que la probabilidad de elegir una mujer de entre todos los mexicanos era, en ese momento, de 57'481,307 entre 112'336,538 que es aproximadamente igual a 0.5117, o lo que es lo mismo, había en ese momento una proporción mujer-hombre de casi un 51.17%, (INEGI, 2011).

Ejemplos para clarificar la distinción entre la probabilidad y la estadística

En los estudios de probabilidad se conocen cada uno de los elementos de la definición de Kolmogórov y lo que básicamente se pretende es obtener y probar las propiedades de diversos objetos probabilísticos mediante la particularización de cada uno de los elementos.

Considérese, por ejemplo, una población en la cual cada uno de los elementos tiene asignado un número, el 1 o el 0, que corresponde a un "éxito" o a un "fracaso", respectivamente, detallando aún más esta ejemplificación, supóngase que la población es un millón de tornillos producidos en alguna fábrica, cada uno de los cuales tendría un 1 si pasa las pruebas de calidad –un éxito– y un 0 en caso contrario –un fracaso–. En este caso, las combinaciones posibles que nos podrían interesar sería la colección de todos los posibles subconjuntos de tornillos que podamos formar (cuyo número es igual al cual es gigantesco, y dicho sea de paso, mucho mayor a todos los átomos que hay en el universo, que algunos estiman que es del orden de 10^{80} a 10^{100}). A su vez, la medida de que tan factible puede ser alguna combinación específica que nos interese, por ejemplo,

que existan 100 tornillos en el lote completo, se podrá determinar tal probabilidad. Así, observemos que en un estudio probabilístico todo es conocido: el espacio muestral, las combinaciones de interés que pueden hacerse sobre los elementos de la población, y la medida de probabilidad o factibilidad que asignamos a cada combinación de elementos que sea de nuestro interés.

Por otro lado, en los estudios de estadística, se consideran conocidos o dados sólo los dos primeros elementos, la población y las combinaciones de elementos de dicha población, pero no así la medida de probabilidad o factibilidad de las combinaciones de interés, y tal probabilidad es precisamente la "incógnita" a resolver en los análisis estadísticos. Es decir al no conocer de antemano la forma en que se distribuyen las probabilidades en algún estudio que involucre el azar, el objetivo de la estadística es caracterizar tales propiedades mediante procesos de observación y de ahí derivar conclusiones que sean, en principio, razonables y de utilidad. Por ejemplo, si consideramos la información de la Bolsa Mexicana de Valores (BMV) es posible conocer con toda precisión las empresas que cotizan en dicha institución (osea se conoce la población), y a su vez, se pueden establecer todas las combinaciones o subconjuntos que pudieran ser de nuestro interés, todas las empresas que cotizan en el sector de la construcción, por citar alguno. Sin embargo, de antemano no es posible conocer sin incertidumbre la forma en que pudieran comportarse los datos de una determinada empresa y en todo caso lo podremos saber hasta que se de el cierre de la bolsa al día de hoy.

El problema estadístico de organizar la información e inferir acerca del comportamiento de naturaleza propiamente estadística, mientras que el estimar la posibilidad de que ocurra un alza o una baja es un problema de la teoría de probabilidades.

Es por lo anterior que hemos afirmado que en los estudios de probabilidad se procede de un método deductivo (del "todo" se deducen propiedades de asignaciones específicas), mientras que en la estadística se aplica un procedimiento inductivo (de hechos particulares, se pretenden establecer las propiedades del "todo").

Imagen: German Tenorio



Conviene ofrecer un ejemplo adicional para aclarar más aún esta situación. Considérese que en una población tuvieramos el interés de analizar el sexo de los hijos en un conjunto de mujeres (dicho sexo, para fines de nuestro ejemplo, lo supondremos desconocido). Observamos que en esta ejemplificación se conoce el espacio muestral que consta de todas las combinaciones hombres-mujeres que sumen igual que el conjunto de mujeres (suponiendo, por simplicidad aunque sin pérdida de generalidad, que todos los niños nacen y no hay nacimiento de varios hijos en una sola mujer). A su vez, la colección de todas las combinaciones o sobconjuntos de interés lo podemos conocer (por ejemplo, si hay cinco mujeres, podría ser de nuestro interés la combinación de que todos los nacimientos resulten en mujeres, o bien, que sean 3 mujeres y dos varones).

En este punto se tendría que tomar la siguiente decisión, dependiendo del objetivo del investigador: si se supone que la proporción hombre-mujer es conocida (ya sea por los registros administrativos, por consideraciones de simetría, como una proporción del tipo 50-50, por el valor prevaleciente a nivel nacional, estatal o municipal que podría obtenerse de los estudios censales del INEGI), en este caso las preguntas serían netamente probabilísticas, por ejemplo: ¿cuál es la probabilidad de que todos los nacimientos resulten mujeres?, o ¿cuál es la probabilidad de que el 80% de los niños resulten varones?

Por otro lado, si la finalidad fuera conocer en función de los próximos nacimientos la proporción hombre-mujer en esta población (debido a que no se tenga acceso a los registros municipales, a que se considere que los resultados censales del INEGI aquí no podrían ser aplicables por algún motivo demográfico o bien simplemente se quiere contrastar o verificar alguna cierta conjetura al respecto), entonces las preguntas serían de naturaleza estadística. Por ejemplo, si la proporción hombre-mujer en esta muestra es de 48-52, esta evidencia llevaría a preguntarse si realmente una proporción simétrica de 50-50 es razonable o no en esta localidad.

Conclusiones

Con estos ejemplos lo que busca enfatizarse es: que las **propiedades probabilísticas** resultan estar presentes cuando se supone un conocimiento de la medida de probabilidad, de las combinaciones de elementos que nos interesan tomados de una población, mientras que las **características estadísticas** resultan cuando no puede suponerse como razonable un conocimiento completo de dicha medida de probabilidad. Adicionalmente, según su metodología, la probabilidad y la estadística difieren puesto que la primera deriva sus conclusiones y hallazgos tomando en consideración básicamente un método deductivo y axiomático, mientras que la segunda precisa de un contraste empírico para establecer, finalmente, la veracidad de sus conclusiones.

A la estadística, cuyo objeto de estudio son los datos, le es indispensable el aparato probabilístico para poder expresar sus ideas y resultados, y de manera más precisa, para encontrar cotas a los errores de los experimentos o procesos de observación que resultan en ciencia o casi en cualquier actividad humana. Lo contrario es también verdadero, puesto que los problemas estadísticos impulsan el desarrollo teórico de la probabilidad, creando así un círculo virtuoso, aunque la distinción, como se ha puntualizado en este documento, debe ser clarificada. Un hecho importante es que no podemos ni debemos separar ambas disciplinas, pues una se nutre de la otra, aunque también debe-

mos reconocer las diferencias que hay entre tales disciplinas para evitar confusiones que pudieran llevarnos a un mal entendimiento o errores que deriven en efectos nocivos a nuestros análisis. ❄

Bibliografía

- [1] BEN-ZVI, D. y Garfield, J. (Eds.) (2005). The challenge of Developing Statistical Literacy, Reasoning and Thinking. Estados Unidos de América: Kluwer Academic Publishers.
- [2] BOX, G., Hunter, W. y Hunter, J. (1978). Statistics for Experimenters. An Introduction to Design, Data Analysis, and Model Building. Estados Unidos de América: John Wiley & Sons.
- [3] FRESS.co (2014). Fotografía de tormenta. Recuperado de <http://fress.co/wp-content/uploads/2014/03/tormenta1hx6.jpg>.
- [4] FREUND, J., Miller, I. y Miller, M. (2000). Estadística Matemática con Aplicaciones. Sexta Edición. México, D.F.: Prentice Hall.
- [5] HALMOS, P. (1974). Measure Theory. United States of America: Springer-Verlag.
- [6] INEGI: Instituto Nacional de Estadística y Geografía. (2011). Principales resultados del Censo de Población y Vivienda 2010. Recuperado de http://www.inegi.gob.mx/prod_serv/contenidos/espanol/bvinegi/productos/censos/poblacion/2010/princi_result/cpv2010_principales_resultadosl.pdf.
- [7] INEGI: Instituto Nacional de Estadística y Geografía. (2014). Logotipo del Censo de Población y Vivienda en su edición del año de 2010. Recuperado de <http://www.inegi.org.mx/est/contenidos/proyectos/ccpv/cpv2010/Default.aspx>.
- [8] Kpta Estrategia Educativa (2014). Fotografía de alumnos de la UNAM. Recuperado de <http://kpta.mx/wp-content/uploads/2014/05/alumnos-unam.jpg>.
- [9] KOLMOGOROV, A. y Fomin, S. (1975). Elementos de la Teoría de Funciones y del Análisis Funcional. Unión de Repúblicas Socialistas Soviéticas: Ed. MIR.
- [10] MOORE, D. (1997). New Pedagogy and New Content: The Case of Statistics. *International Statistical Review*, 65(2), pp. 123-165.
- [11] MOORE, D. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association*, 93(444), pp. 1253-1259.