Universidad de Chile
Facultad de Cs. Físicas y Matemáticas
Departamento de Ingeniería Civil Matemática

UNA PEQUEÑA HISTORÍA DE LA ESTADÍSTICA

Nancy Lacourly

Julio 2000

1. Introducción

Muchas de las decisiones que usted, su medico o los ministros de gobierno toman a diario se basan en *estadísticas*.

Usted elige la pasta de dientes *X* para su hijo porque la publicidad le dice que le va a disminuir de 20% las caries. El candidato a la elección presidencial cambia su campaña en vista de los resultados de una encuesta de opinión. El médico le prescribe un tratamiento porque las estadísticas le informan que el tratamiento cura su enfermedad en 70% de los casos. Una muestra de sangre es suficiente para decir miles de cosas sobre la totalidad de la sangre de su cuerpo. Los *People Meters* instalados en algunos televisores permiten a los directores de canales de TV evaluar la popularidad de sus programas.

¿Las estadísticas obtenidas a partir de una gota de sangre, algunos telespectadores o electores permiten tomar siempre decisiones acertadas? La repuesta puede influir decisiones relativas a su vida. No se puede ignorar el uso generalizado de la Estadística en todos las actividades de nuestra sociedad.

Actualmente el gobierno de cada país recolecta sistemáticamente datos relativos a su población, su economía, su recursos naturales y su condición política y social para tomar decisiones. En las actividades industriales o comerciales las estadísticas son parte de la organización así como en los sectores agrícola y forestal, donde se requiere predicciones de la producción. En la investigación científica (medicina, física, biología, ciencias sociales, etc.) el rol de la Estadística es primordial.

> *Hay tres tipos de mentiras: mentiras, condenadas mentiras y la Estadística*
> Atribuido a Mark Twain por el primer ministro
> inglés Benjamin Disraeli (1804-1881)

1 Historia del azar y del desarrollo de la Estadística

El desarrollo de la computación trastornó los progresos de la Estadística y su enseñanza. Vamos a ver aquí como y por quién se desarrollo la Estadística, desde la prehistoria hasta la actualidad. Es difícil separar la evolución de la Estadística sin considerar la de las Probabilidades. El progreso de ambas disciplinas puede verse como la historia de una única ciencia: ***la ciencia del azar***.

2.1 La prehistoria

La Estadística Descriptiva tiene su origen mil o dos miles años antes de Cristo, en Egipto, China y Mesopotamia, donde se hacían censos[1] para la administración de los imperios. Los egipcios tuvieron el barómetro económico más antiguo: un instrumento llamado *"Nilometro"*, que medía el caudal del Nilo y servia a definir un índice de fertilidad, a partir del cual se fijaba el monto de los impuestos. Con la variabilidad del clima ya conocían el concepto de incertidumbre.

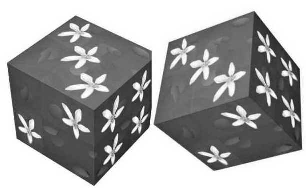
Paralelamente, el concepto de azar es tan antiguo como los juegos (los dados y los juegos con huesos que en Chile llamamos “payayas” son antiquísimos) y motivó desde antaño las reflexiones de los filósofos. En las ideas de Aristóteles (384-322) se encuentran tres tipos de nociones de probabilidad, que definen más bien actitudes frente al azar y la fortuna, que siguen vigentes hasta nuestros días: (1) el azar no existe y refleja nuestra ignorancia; (2) el azar proviene de causas múltiples y (3) el azar es divino y sobrenatural. Sin embargo, pasó mucho tiempo antes de que alguien intentara cuantificar el azar y sus efectos.

---

[1] La palabra *censo* viene de la palabra latina *censere* que significar *fijar impuestos.*

Venus y la suerte
Nada es más impredecible que el lanzamiento de un dado, y cada hombre que juega obtiene de vez en cuando "Venus": puede sacarla dos o tres veces seguidas. Seremos tan estúpido para decir que las cosas ocurren gracias a la intervención de Venus y no por pura suerte.
Cicero (106-43 AC) en De Divinatione

Durante la edad media hubo una gran actividad científica y artística en Oriente y el nombre de *azar* parece haber venido desde Siria a Europa. La flor de a*zahar*, que aparecía en los dados de la época podría ser el origen de la palabra. Las compañías aseguradoras iniciaron investigaciones matemáticas desde tiempos muy antiguos, y en siglo XVII aparecieron los primeros famosos problemas de juegos de azar. En la sociedad francesa, el juego era uno de los entretenimientos más frecuentes. Los juegos cada vez más complicados y las apuestas muy elevadas hicieron sentir la necesidad de calcular las probabilidades de los juegos de manera racional. El caballero de Méré, un jugador apasionado, escribiendo a Blas Pascal (1623-1662) sobre ciertos juegos de azar, dio origen a una correspondencia entre algunos matemáticos de la época. Las preguntas de De Méré permitieron, en particular, iniciar una discusión entre Pascal y Pierre Fermat (1601-1665) y así el desarrollo de la teoría de las Probabilidades. En el siglo anterior, los italianos Tartaglia (1499-1557), Cardano (1501-1576), e incluso el gran Galileo (1564-1642) abordaron algunos problemas numéricos de combinaciones de dados.



En cada juego de azar, dados, cartas o ruleta, por ejemplos, cada una de las jugadas debe dar un resultado tomado de un conjunto finito de posibilidades (números de 1 a 6 para el dado, 52 posibilidades para las cartas o 38 para la ruleta). Si el jugo de azar es "correcto", no se puede predecir de antemano el resultado que se obtendrá en una jugada. Es lo que define el azar del juego. Se observa una cierta simetría en los posibles resultados: son todos igualmente posibles, es decir que el riesgo para un jugador es el mismo cualquier sea lo que juega. De aquí surgió la primera definición de una medida de probabilidad para un determinado suceso:

$$p = \frac{a}{b}$$

donde a es el número de casos *favorables* (el número de casos que producen el suceso) y b el número de casos *posibles*. Por ejemplo, la probabilidad de sacar un *6* en el lanzamiento de un dado es $p = \frac{1}{6}$, de sacar un corazón de un paquete de 52 cartas es $p = \frac{13}{52} = \frac{1}{4}$ o un número par en la ruleta (considerando que *0* y *00* son ni pares y ni impares) es $p = \frac{18}{38}$.

El caballero de Méré, que jugaba con frecuencia, había acumulado muchas observaciones en diversos juegos y constató una cierta regularidad en los resultados. Esta regularidad, a pesar de tener su base de un hecho empírico, permitió relacionar la frecuencia relativa de la ocurrencia de un suceso y su probabilidad. Si f es la frecuencia absoluta de un suceso (el número de veces que ocurrió) en n jugadas, como el número de casos favorables debería ser aproximadamente igual a na, $f \approx \frac{an}{b}$ y entonces la probabilidad de que ocurra el suceso será:

$$p = \frac{a}{b} \approx \frac{f}{n}$$

En un juego, de Méré encontraba una contradicción en su interpretación de la probabilidad a partir de la frecuencia relativa que obtuvo empíricamente. Pascal y Fermat pudieron mostrarle que sus cálculos eran erróneos y que la interpretación propuesta era correcta. De Méré siguió planteando problemas que no pudieron resolver los matemáticos de su época. Sin embargo, Jacques de Bernoulli (1654-1705), el primero de una famosa familia de matemáticos suizos, dio una demostración de la ley de los Grandes Números y Abraham de Moivre enunció el teorema de la regla de multiplicación de la teoría de la probabilidad.

Según Richard Epstein, la ruleta es el juego de casino más antiguo que está todavía en operación. No se sabe a quien atribuirlo: puede ser Pascal, el matemático italiano Don Pasquale u otros. La primera ruleta fue introducida en París en 1765.

Problema de los Puntos
Supongamos que dos jugadores Abel y Bertrán interrumpen un juego secuencial en el cual a Abel le falta X y a Bertrán le falta Y para ganar. ¿Cómo tienen que repartirse las apuestas?
Uno de los famosos problemas propuestos por de Méré y resueltos por Pascal y Fermat

2.3 La demografía

Las reglas de cálculo desarrolladas hasta entonces para los juegos de azar vieron sus aplicaciones en otras disciplinas. Los censos demográficos, que se hacían desde la antigüedad, requieren recolectar muchos datos. La demografía y los seguros de vida aprovecharon del desarrollo de la teoría de las probabilidades. Consideramos, por ejemplo, el sexo de una sucesión de niños recién nacidos. Se puede ver como una repetición de lanzamientos de una moneda, con niño y niña en vez de cara y sello. De la misma manera, podemos considerar un conjunto de hombres mayores de 50 años. Al final del año, una cierta proporción sigue viva. Durante el siglo XVIII con Pierre Simon, Marqués de Laplace (1749-1827), estos problemas fueron reconocidos como similares a los de un juego, y se encontraron las correspondientes frecuencias relativas, lo que permitió determinar la probabilidad que nazca una niña, o que un hombre mayor que 50 años muera en el año.

Si bien la extensión de los juegos de azar a la demografía o a la matemática actuarial fue extremadamente importante, su planteamiento tiene grandes limitaciones debido a que considera todos los resultados posibles simétricos. ¿Qué pasa cuando una situación real puede expresarse como un juego de azar? Por ejemplo, Daniel Bernoulli, careciendo de datos sobre la mortalidad producida por la viruela a distintas edades, supuso que el riesgo de morir de la enfermedad era el mismo a toda edad. Lo que evidentemente es muy discutible.

2.4 La teoría de los errores

Durante los siglos XVIII y XIX la Estadística se expandió sin interrupción mientras la teoría de las probabilidades no mostró progreso. Una de las aplicaciones importante fue desarrollada al mismo tiempo por Gauss (1777-1855), Legendre (1752-1833) y Laplace: el análisis numérico de los errores de mediciones en física y astronomía. ¿Cómo determinar el mejor valor leído por un instrumento que entrega diferentes mediciones del mismo fenómeno? Si tenemos $\{x_1, x_2,...,x_n\}$ n mediciones de un mismo fenómeno, deberíamos tener $x_1= x_2=...=x_n$ si no hubiera errores. En su Anexo sobre el método de los mínimos cuadrados, de "Nuevos métodos para la determinación de las órbitas de los cometas", Legendre propone determinar el valor único z de la medición de manera que una función de los errores $\varepsilon_i = x_i - z$ sea mínima:

$$\underset{z}{Min}\sum_{i=1}^{n}(x_i - z)^2$$

La solución es el promedio de las mediciones.

Esta función cuadrática encuentra su justificación en la distribución normal con Gauss y Laplace, aunque la distribución de los errores fue estudiada mucho antes por Thomas Simpson (1710-1761), que hizo los supuestos que esta distribución tenia que ser simétrica y que la probabilidad de errores pequeños debería ser más grande que la de los errores grandes. En 1840, Sir Francis Galton (1822-1911) partió de una distribución discreta y la fue refinando hasta llegar en 1857 a una distribución continua muy parecida a la distribución normal. Galton inventó incluso una maquina llamada *quincunx*, que permite ilustrar la distribución normal (ver anexo). Galton trabajo en meteorología y en herencia. Era el primo de Charles Darwin.

La distribución normal es la ley en la cual todo el mundo cree
Los experimentadores creen que es un teorema de la Matemática,
y los matemáticos que es un hecho experimental.
El astrónomo Lippman

2.5 Nacimiento de la Estadística Moderna

Es con la introducción de nuevas aplicaciones que la teoría de las probabilidades del siglo XVIII funda la Estadística Matemática. El término de *Estadística* se debe posiblemente a G. Achenwall (1719-1772), profesor de la Universidad de Göttingen, tomando del latín la palabra *status*.

Aparte de la demografía y la matemática actuarial, otras disciplinas introdujeron la teoría de las probabilidades. Fue el inicio de la mecánica estadística, debido a Maxwell (1831-1879) y Bolzmann, quienes dieron también una justificación de la distribución normal en la teoría cinética de los gases.

La Estadística se empezó a usar de una manera u otra en todas las disciplinas, a pesar de un estancamiento de la teoría de las probabilidades. En particular, muchos vieron la dificultad de aplicar el concepto de simetría, o de casos igualmente posibles, en todas las aplicaciones. Hubo que esperar a que Andrey Nickolaevich Kolmogorov (1903-1987) separara la determinación de los valores de las probabilidades de sus reglas de cálculo.

Los primeros resultados importantes de la *Estadística Matemática* se deben al ingles Karl Pearson (1857-1936) y a otros investigadores de la escuela biométrica inglesa.

2.6 La segunda mitad del siglo XX: la revolución computacional

Los científicos, especialmente los ingleses, desarrollaron métodos matemáticos para la Estadística, pero en la práctica manipularon cifras durante medio siglo sin disponer de verdaderas herramientas de calculo. La llegada de los computadores revolucionó el desarrollo de la Estadística. En Francia (Benzécri) y en los Estados unidos (Tuckey) fueron los pioneros en repensar la Estadística en función de los computadores. Mejoraron, adaptaron y crearon nuevos instrumentos para estudiar grandes volúmenes de datos: nuevas técnicas y herramientas gráficas.

El modelo tiene que adaptarse a los datos y no al revés.
Jean-Paul benzécri, 1965

2.7 Calculo de Probabilidades y Estadística

Algunas palabras para concluir. Si bien la historia de la Estadística no se puede separar de la historia de Calculo de las Probabilidades, la Estadística no puede considerarse como una simple aplicación del Calculo de las Probabilidades. Podemos comparar esta situación a la de la Geometría y la Mecánica. La Mecánica usa conceptos de la geometría, y sin embargo es una ciencia a parte.

El Calculo de las Probabilidades es una teoría matemática y la Estadística es una ciencia aplicada donde hay que dar un contenido concreto a la noción de probabilidad. Como ilustración citemos el experimento de Weldon (1894), que lanzó 315.672 veces un dado (bajo la supervisión de un juez) y anotó que 106.602 veces salió un 5 o un 6. La frecuencia teórica debería ser 0.3333... si el dado hubiera sido perfectamente equilibrado. La frecuencia observada aquí fue 0.3377. *¿Deberíamos concluir que el dado estaba cargado?* es una pregunta concreta que es razonable considerar. El Calculo de las Probabilidades no responde a esta pregunta y es la Estadística la que permite hacerlo.

> *El geómetra no se interesa por saber si existen en la práctica objetos que puedan considerarse como líneas rectas. Hay que tener cuidado cuando se razona por analogía con otras ramas de las matemáticas aplicadas, porque a este nivel no nos preocupamos solamente de las relaciones entre calculo y razonamiento. Admitamos el derecho del matemático de desinteresarse al problema, como matemático, pero tenemos que asumir la responsabilidad de resolver la dificultad, como psicólogo, lógico o estadístico, a menos que estemos dispuestos a poner la probabilidad en el campo de la matemática pura y sus aplicaciones en el frontis de nuestras academias.*
>
> Kendall, 1949