



Cromosomas humanos. NHGRI



**CURIOSIDADES ESTADÍSTICAS**

Por la Sociedad de Estadística e Investigación Operativa ([www.seio.es](http://www.seio.es))

**Lotería y probabilidades**

La estadística utiliza la teoría matemática de las probabilidades para incorporar en sus modelos la incertidumbre asociada al azar. El primero en hacerlo fue Cardano, médico, filósofo y matemático italiano del siglo XVI, jugador compulsivo, quien escribió el tratado *Libro sobre los juegos de azar*, en el que calculaba la probabilidad de ganar en distintos juegos de azar y ¡daba incluso técnicas para hacer trampas! La Lotería Nacional en España existe desde hace 200 años, y es fácil calcular la probabilidad de acertar el Gordo: una entre 100.000. Este sábado el boleto hará un guiño a la estadística en su año internacional. Fuente: Mathieu Kessler, catedrático de Estadística, Universidad Politécnica de Cartagena.

**Estadística y Bush versus Gore**

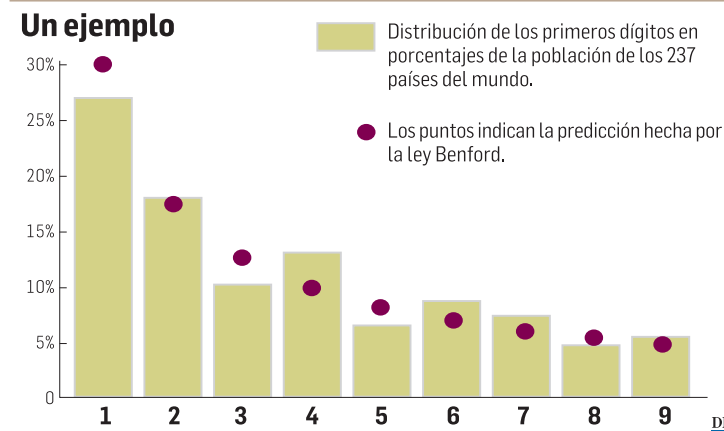
Bush ganó las elecciones presidenciales USA del año 2000 por Bush por un margen muy estrecho de votos. Al Gore hubiera vencido si habría obtenido 537 votos más en Palm Beach (Florida). Técnicas estadísticas elementales permiten comprobar que el uso de un tipo de papeleta de votación (las "papeletas mariposa") confundió a unos 2.500 votantes de Florida que creían estar votando a Gore, cuando en realidad otorgaron su voto a un tercer candidato (Pat Buchanan). Fuente: Greg Adams, *Voting Irregularities in Palm Beach, Florida*, Chance 14, No. 1 (2001), págs. 22-24.

**LOS NÚMEROS DE LA VIDA DIARIA**

**La ley de Benford en los datos estadísticos**

**L**a estadística necesita de datos para obtener la información necesaria para construir hipótesis e inferir resultados para los problemas que pretenda analizar. Los datos son un elemento esencial y es necesario estructurarlos. En ese ámbito, la ley de Benford describe cómo se distribuyen los dígitos 1, 2, 3, ..., 9 dentro de las cifras significativas (las cifras de más importancia dentro de un número excluyendo los ceros) de una colección de datos numéricos. Esta ley fue inicialmente observada por el astrónomo y matemático canadiense Simon Newcomb en 1881, quien manejando manuales que contenían tablas de logaritmos observó que las páginas que contenían los números con primeras cifras 1 y 2 estaban mucho más manoseadas y más oscuras por el uso, que las páginas con las primeras cifras 8 y 9. Esto indicaba que los manuales de tablas de logaritmos eran más empleados para números con cifras significativas pequeñas (1, 2 y 3) que para cifras significativas grandes (8 y 9). En particular, notó que en el 30% de las ocasiones el primer dígito era 1; el 9 sólo lo era un 4,6% de las veces. Esta ley fue formalmente establecida en 1938 por el físico Frank Benford, a quien debe su nombre. Durante varios semestres estuvo recogiendo datos y ese año publicó un artículo en *Proceedings of the American Philosophical Society*, en el que basándose en 20.229 observaciones de origen diverso (cuencas de ríos, estadísticas de la liga americana de béisbol, pesos atómicos de elementos,...) mostró que esta ley se cumplía en una gran cantidad de conjuntos de datos.

No sé conoce aún la razón precisa de por qué grandes cantidades de conjuntos de datos cumplen esta ley mientras otros conjuntos de datos no la cumplen. No



obstante, se han observado algunas características que permiten delimitar la naturaleza de los conjuntos de datos que siguen la ley de Benford:

- 1) Datos provenientes de procesos de crecimiento exponencial. Estos datos suelen tener órdenes de magnitud muy diferentes (es decir, hay datos muy pequeños junto con otros muy grandes: 0,005 y 10.000.000, por ejemplo), lo que suele ser un rasgo muy común de los conjuntos que siguen esta ley. Habitualmente, los conjuntos de datos que presentan valores acotados dentro de un rango (por ejemplo, poblaciones entre 1.000 y 5.000 habitantes) no suelen seguir la distribución de Benford.
  - 2) Datos que presentan invariancia respecto a la escala. Esta situación se presenta cuando los conjuntos de datos incluyen tanto a los que se presentan en una unidad de medida como en otra. Es decir, da lo mismo expresar los datos en centímetros o en metros, por ejemplo, porque en ambas situaciones se cumple la ley de Benford.
- De esta forma, una amplia variedad de conjuntos de datos siguen la ley de Benford: facturas de electricidad, direcciones de calles, precios de acciones en la bolsa, tamaños de poblaciones, ra-

tios de defunciones, longitudes de ríos, constantes físicas y matemáticas, por nombrar algunos.

La ley de Benford puede aplicarse en diferentes casos prácticos: a) Detección de fraude contable. Dado que los datos contables siguen esta ley, cuando éstos han sido manipulados, habitualmente se observa una mayor presencia de primeros dígitos con valores 8 y 9 que la proporción indicada por la distribución de Benford.

b) Análisis de resultados electorales. La ley de Benford sirvió como evidencia de fraude en las elecciones de Irán en 2009.

c) Control de datos macroeconómicos. Los datos macroeconómicos proporcionados por el Gobierno griego antes de ingresar en la Unión Europea en 1980 fueron considerados fraudulentos a causa de esta ley.

Finalmente, es importante señalar que la evidencia de que un conjunto de datos no siga la ley de Benford, cuando se esperaría que sí la siguiera, debe hacerse siempre con la conveniente prudencia, fruto de la aplicación cuidadosa de la inferencia estadística, ya que los datos siempre muestran una realidad parcial de un fenómeno determinado. **Javier Faulín** es catedrático de Estadística e Investigación Operativa de la UPNA.

dos, víctimas, militares... "Cada caso requiere unas medidas de protección diferente, según el uso que se le vaya a dar", explica Katsanis. Según los expertos, todos los programas de almacenamiento de ADN deben ser transparentes y flexibles, para garantizar el máximo beneficio posible y a la vez la protección de los derechos de las víctimas. "No hay evidencia de que alguno de estos programas esté utilizando la información de manera poco ética", opina Katsanis. "Sin embargo los participantes pue-

den pensar lo contrario, en especial en países con gobiernos corruptos". A medida que las bases de ADN se vuelvan más numerosas e internacionales, será cada vez más importante tomar las precauciones necesarias para garantizar la protección de las víctimas, por ejemplo, asegurando que solo autoridades imparciales tengan acceso a los datos. Las posibilidades que abre el estudio de nuestro genoma permitirán avanzar cada vez más en campos que van desde la medicina hasta la criminología. No obstante, la tecnología no puede olvidar la ética y los derechos humanos cuando se habla del ADN. No hay nada más personal que nuestros genes.