



UNIVERSIDAD
COMPLUTENSE
MADRID

LECCIÓN INAUGURAL
Curso Académico 2015/2016

Momentos estelares de la Probabilidad y de la Estadística

Conrado M. Manuel García

Catedrático de Escuela Universitaria
de la Facultad de Estudios Estadísticos

Madrid, 2015

Momentos estelares de la Probabilidad y de la Estadística

Conrado M. Manuel García

Catedrático de Escuela Universitaria
de la Facultad de Estudios Estadísticos

Corrección, edición, diseño y maquetación

Departamento de Estudios e Imagen Corporativa. UCM

Impresión

Grafo Industrias Gráficas

Sumario

Momentos estelares de la Probabilidad y de la Estadística.....	7
Los monos mecanógrafos de Borel	8
Galton, el hombre que medía todo	11
La significación estadística de la Guinness	16
El carácter de Fisher	20
Los Big Data.....	24
Bibliografía	29

*“To consult the statistician after an experiment is finished
is often merely to ask him to conduct a post mortem examination.
He can perhaps say what the experiment died of”.*

R. A. Fisher (1938)
Presidential Address to the First Indian Statistical Congress

Magnífico Sr. Rector, Dignísimas autoridades, Señores claustales, queridos compañeros, queridos alumnos y personal de la Universidad Complutense, Señoritas, Señores.

Dice el dicho que es de bien nacidos ser agradecidos. No podría, por tanto, comenzar esta Lección Inaugural sin mostrar mi más profunda gratitud a la persona que me ofreció dictarla, la Ilma. Sra. Decana de mi centro, la Facultad de Estudios Estadísticos, así como a mis compañeros que ratificaron su decisión en la Junta de Facultad. Me llena de orgullo y satisfacción esta distinción por un doble motivo: poder participar en este solemne y singular acto académico que anualmente abre la puerta a todos los demás, así como apreciar el reconocimiento profesional y el afecto personal que me profesan mis compañeros. Muchas gracias, por tanto, de nuevo.

Momentos estelares de la Probabilidad y de la Estadística

Este privilegio, no obstante, tiene aristas cortantes, pues no es fácil atraer el interés de colegas tan doctos y con conocimientos tan dispares como los presentes y, menos aún, si esto se intenta desde la aridez de una disciplina como la Estadística y sus alrededores. He descartado la opción de presentar un tema de investigación con formulaciones teóricas y modelos matemáticos o estadísticos que, aún pudiendo ser más al uso, hubiera sido, sin duda, escasamente agradecido. En su lugar, he preferido sustraer a mi admirado **Stefan Zweig**, grande entre los más grandes escritores del siglo pasado, el título y la estructura de una de sus más celebradas obras, *Momentos estelares de la Humanidad*, para presentar aquí varias miniaturas históricas en las que el protagonismo corresponde a científicos, ideas o situaciones que han supuesto **Momentos estelares de la Probabilidad y de la Estadística**.

En ellas pretendo divulgar, si es que ello es en realidad necesario, o si no tal vez solo recordar, varios episodios que por su singularidad, por la brillantez de sus protagonistas, por las consecuencias que tuvieron o tendrán para el futuro, o por el hecho de que todos somos un poco adictos a los chismes, espero resulten atractivos. Aunque algunos de los protagonistas aparecen aquí por su vinculación con la Estadística, en realidad fueron científicos en el sentido renacentista del término y posiblemente muchos entre los presentes

los asocien con otras disciplinas. Tanto mejor, porque así me resultará más fácil conseguir su complicidad.

Los monos mecanógrafos de Borel

Émile Borel (1871-1956), junto con **René-Louis Baire** y **Henri Lebesgue**, estableció las bases de la Teoría de la Medida. Borel, además, fue pionero en introducir aplicaciones de dicha teoría a la Probabilidad. También publicó investigaciones en Teoría de Juegos y relacionó la Geometría Hiperbólica con la Relatividad Especial. Sin duda fue un hombre válido y polifacético. En 1913 ideó una parábola para ilustrar un acontecimiento extraordinariamente improbable. En su libro *Mecanique Statistique et Irreversibilité (Mecánica Estadística e Irreversibilidad)*, Borel afirmó que aunque un millón de monos mecanografiaran diez horas diarias era extremadamente improbable que llegaran a escribir algo legible. Y utilizó la metáfora para argumentar sobre la validez de las leyes de la Estadística, de las que afirmó que el hecho de que fueran violadas, siquiera someramente, era aún más inverosímil.

Con el paso del tiempo la idea original de Borel se fue refinando y desde los años 70 fue dada la vuelta para afirmar, en el coloquialmente conocido como “*Teorema de los infinitos monos*” o “*Teorema del mono infinito*”, que si un número infinito de monos inmortales aporrean máquinas de escribir (o teclados de ordenador, no nos vamos a poner exigentes) durante toda la eternidad, alguno acabará escribiendo *El Quijote*. Los ingleses lógicamente piensan que escribirá *Hamlet* y los franceses *El Médico a palos* o *Los miserables*. En realidad no harían falta infinitos simios; bastaría uno solo. En la infinitud del aporreo da igual uno que infinitos. Aparentemente nos ahorraríamos muchos plátanos, pero con la teoría de cardinales en la mano, si los monos son numerables, que lo son, a largo plazo no nos ahorraríamos nada. Si ese mono golpea a lo loco durante un tiempo infinito, llegará a escribir cualquier cosa, incluida esta lección inaugural, la partida de nacimiento de cada uno de ustedes o todos los contratos del fichaje de Neymar.

Pensemos en la demostración de este teorema. Escribir *El Quijote* pulsando teclas al azar es extraordinariamente improbable como afirmó Borel, pero no imposible. Supongamos que el número de símbolos diferentes que aparecen en *El Quijote*, incluyendo mayúsculas, números, blancos, etc. es de 50. Supongamos también que hemos entrenado al mono para que teclee verdaderamente al azar. Esto puede no ser sencillo. Los que diseñan algoritmos para obtener números aleatorios lo saben. Pero pongamos que el mono se deja amaestrar para elegir las sucesivas

teclas de manera aleatoria. Entonces la probabilidad de que las dos primeras teclas que golpee sean la E mayúscula y la n minúscula es $(1/50) \times (1/50)$, o sea 2% por 2% o 0,0004. Teclear “En un” con su blanco sería obviamente más difícil y lo conseguiría con probabilidad 0,0000000032. Esta probabilidad todavía es mayor que la de ganar la primitiva con solo una apuesta. Obtener “En un lugar” tiene una probabilidad del orden de 10 elevado a menos 19. Cada vez es más difícil pero como tiene todo el tiempo del mundo...

Obviamente esta cuestión de la inmortalidad no es baladí. Si teclea a ritmo de una pulsación por segundo (imaginemos que comer plátanos no le hace perder tiempo), cada día tecleará $24 \times 60 \times 60 = 86.400$ caracteres, o sea unas 28 páginas en DIN A4, con letra tamaño 11 puntos e interlineado de 1,15 que es un formato bastante estándar. Por ejemplo, es el que se me pidió para escribir esta lección inaugural. Por lo tanto, en los escritos del mono, que serán sin duda muy indigestos, podemos esperar que aparezca “En” después de pulsar 2.500 teclas o sea después de aproximadamente 42 minutos de trabajo. Para encontrar “En un lugar” tendríamos que esperar casi 62.000 millones de años sin que el simio desfallezca.

Augusto Monterroso escribió un cuento, “*El dinosaurio*”, que pasa por ser el más breve jamás escrito. Tiene 50 caracteres y reza:

“Cuando despertó, el dinosaurio todavía estaba allí”.

Aunque un simio hubiera empezado a teclear el día del origen del universo y al ritmo que hemos descrito hasta hoy mismo, aún le faltarían muchos, muchos periodos de tiempo similares para escribir tal cuento. O cualquier otro de 50 caracteres. Obviamente podemos reducir la espera aumentando el número de monos.

Tratando de llevar las cosas a la práctica, en 2003, unos científicos en Paignton Zoo y en la Universidad de Plymouth en Devon, Inglaterra, dejaron un teclado de ordenador a merced de unos monos durante un mes. Solo consiguieron 5 páginas consecutivas de la letra “S” y que luego apedrearan el teclado y orinaran y defecaran reiteradamente sobre él.

En el mismo año otros científicos prefirieron simular informáticamente una gran población de monos tecleando verdaderamente al azar. Crearon el sitio web *The Monkey Shakespeare Simulator*. Año y medio después encontraron 24 letras consecutivas que formaban un pequeño fragmento de *Enrique VI*. Posteriormente también se obtuvieron 30 letras del *Julio César*.

Esta idea de Borel ha hecho correr tantos ríos de tinta que se ha especulado con la posibilidad de que su metáfora estuviera inspirada en un agrio debate que mantuvieron en Oxford **Thomas Henry Huxley** y el obispo anglicano de esa ciudad, **Samuel Wilberforce**, en la reunión del 30 de junio de 1860 de la British Association for the Advancement of Science. Wilberforce era su vicepresidente y no había encajado muy bien los contenidos de *El origen de las especies*, de Charles Darwin, publicada tan solo unos meses antes. Por su parte, Huxley era un biólogo conocido como el bulldog de Darwin por su interés y habilidad en la defensa de la tesis evolutiva y al que se atribuye el invento del término “agnóstico” así como el promover el estudio de la ciencia desde tal punto de vista. No existe prueba alguna de tal influencia en la parábola de Borel pues el debate no se transcribió en su totalidad. Las notas de Huxley permiten saber que hubo referencia a monos y probablemente de ahí nace la especulación. Tratando de ridiculizar la teoría, el obispo le preguntó a Huxley si descendía del mono por parte de padre o de madre, a lo que Huxley contestó que tener una ascendencia simiesca le parecía preferible a descender de alguien tan descortés como el obispo. Parece que el impacto de sus palabras motivó incluso el desmayo de una dama presente. Es, sin embargo, muy poco verosímil que Huxley hiciera ninguna referencia a máquinas de escribir porque a pesar de que se habían patentado en 1714, no empezaron a comercializarse hasta unos años después del referido debate, con lo que eran prácticamente desconocidas.

Sin embargo, **Jonathan Swift** sí anticipa la idea en su obra *Viajes de Gulliver* donde un profesor de la Academia de Lagado desea crear una lista de todo el conocimiento existente obligando a sus estudiantes a crear textos aleatorios de forma permanente a base de girar las manivelas de un extraño mecanismo.

Otros escritores han explotado la idea con éxito. **Borges** situó en los anaqueles de su *Biblioteca de Babel* todos los libros que pudieran escribirse con un número fijo de símbolos. Recuerdo la gran impresión que me causó esta obra la primera vez que la leí. Cito de memoria pero venía a decir más o menos que habría textos con sentido, pero también salas enteras de obras ilegibles. Habría muchos libros que se diferenciarían en apenas un símbolo o en muy pocos. El saber de la humanidad quedaría garantizado incluso ante guerras o invasiones que devastaran la biblioteca pues por muchos libros que desaparecieran siempre quedarían otros prácticamente iguales.

Borges también vuelve a esta idea en su cuento “*El inmortal*”, argumentando que la inmortalidad de Homero le conduce inexorablemente a escribir *La Odisea*:

“Homero compuso La Odisea; postulado un plazo infinito, con infinitas circunstancias y cambios; lo imposible es no componer, siquiera una vez, La Odisea”.

Michael Ende también utiliza la idea en su exitosa obra *La historia interminable*. Describe un país fantástico cuyos habitantes vigilados por un mono juegan el “*juego de la arbitrariedad*”. Este juego consiste en lanzar dados de letras y recopilar los resultados. Periódicamente surgen textos legibles:

“Si se sigue jugando cien años, mil años, cien mil años, con toda probabilidad saldrá una vez, por casualidad, un poema. Y si se juega eternamente, tendrán que surgir todos los poemas, todas las historias posibles, y luego todas las historias de historias, incluida ésta de la que precisamente estamos hablando”.

Hasta en un episodio de los Simpson, **Montgomery Burns** tenía en su villa mil monos encerrados en una habitación aporreando máquinas de escribir. Y se castigaba a uno por escribir mal una letra de la obra de Dickens, *Historia de dos ciudades*.

Una de las veces que hice referencia en clase a esta metáfora, un alumno, biólogo, afirmó que ya hubo un *mono* que acertó a escribir *El Quijote*. Se llamaba Cervantes.

Galton, el hombre que medía todo

Sir Francis Galton (1822-1911) fue un hombre polifacético donde los haya: antropólogo, biólogo, geógrafo, meteorólogo, explorador, inventor, psicólogo, eugenista británico y estadístico. Destacó desde pequeño pues a los dos años y medio aprendió a leer y a los cinco podía leer cualquier libro en inglés con tremenda soltura. En su historial académico de juventud sobresalió especialmente en matemáticas, no brillando excesivamente en las demás áreas. Al enorme valor de sus investigaciones hay que añadir el hecho de que las hizo siempre por su cuenta ya que nunca tuvo un puesto en ninguna universidad. No es menos cierto que procedía de una familia acomodada y nunca pasó por apuros económicos.

Las ideas de Galton fueron tan brillantes e influyentes que abrieron el camino a lo que hoy son varias disciplinas diferentes. Era primo por parte de madre de **Charles Darwin** y los planteamientos sobre la evolución de éste condicionaron en

gran medida los intereses investigadores de Galton, las inteligentes preguntas que se hizo y las magníficas respuestas que se dio. Fundó junto con sus discípulos **Karl Pearson** y **Walter Weldon** la revista *Biometrika* para promover el estudio de la Bioestadística.

Regresión a la media

Galton sentó las bases de la ciencia que hoy conocemos como Estadística. Fue el primero en estudiar la vinculación entre variables introduciendo el uso de la recta de regresión y explicando el fenómeno de la regresión a la media en un artículo de finales del siglo XIX. En este trabajo Galton estableció una relación de tipo lineal entre la altura de los padres y la de los hijos, enfatizando el hecho de que los hijos de padres altos son, en media, de menor estatura que sus padres, y que los hijos de padres bajos son, en promedio, más altos que sus padres. Inicialmente este fenómeno recibió el nombre de regresión (vuelta atrás) a la mediocridad y hoy lo conocemos como regresión a la media.

La regresión a la media es observable con frecuencia en nuestras vidas (González, Internet) y nos referimos a ella en múltiples situaciones quizá sin conocer la terminología científica. Cuántos de los aquí presentes, brillantes académicamente desde su infancia, se han frustrado porque sus hijos no tienen mayor rendimiento académico. Sin embargo, los padres que no acabaron el bachillerato tienen en media hijos que les superan en formación.

Solemos decir con frecuencia en referencia a la continuación de una película que nos deslumbró: *“Segundas partes nunca fueron buenas”*. Esta afirmación se puede explicar en términos de la regresión a la media. Si representamos en una nube de puntos las taquillas de la primera y de la segunda parte (utilizando la recaudación como una medida de calidad, lo cual puede ser desde luego discutible) en general, podremos observar el fenómeno. Entre las primeras partes cuya recaudación fue muy alta, la de las segundas partes, en promedio será menor. Pero también lógicamente hay segundas partes que fueron mejor que la primera y eso muchas veces ocurre cuando la primera fue menos exitosa.

Similarmenete, si nos trasladamos al entorno del deporte, los efectos que Galton detectó en las relaciones entre la altura de padres e hijos se pueden observar en la transmisión de la habilidad. El hijo de Messi o el de Cristiano tiene muy difícil superar el listón que le va a dejar su padre y será, si se dedica a ello, casi con toda seguridad, un futbolista menos dotado.

En nuestro entorno investigador, todos nos hemos sorprendido alguna vez por la que considerábamos escasa calidad de un artículo publicado en una revista de alto impacto. **Rousseuw** (1991) analiza el fenómeno. Si un artículo es revisado por tres *referees* y estos emiten una valoración muy positiva del *paper*, el editor decidirá su publicación. Ahora bien, si ese mismo artículo es revisado a continuación por otros tres evaluadores, en promedio, la valoración será menor.

Y si acudimos a nuestro entorno docente, con frecuencia comentamos con compañeros por los pasillos que *“el grupo de este año es infernal, con lo bueno que fue el del año pasado”*. O también: *“menos mal que los alumnos de este año no son tan malos como los del año pasado”*.

La aproximación normal a la distribución binomial

Galton también inventó la máquina que lleva su nombre y que ayuda a comprender la aproximación normal a la distribución binomial. Tal máquina, por contarlo de forma intuitiva, hace algo similar a lo que ocurre cuando una cosechadora, después de segar, lanza el grano desde la tolva a través de su brazo en una era. Si la cosechadora no se mueve hasta que terminan de caer los granos, éstos formarán el típico montón completamente simétrico que recuerda a la distribución normal bivalente y que adorna los aledaños del caserío de cada pueblo de Castilla durante los meses estivales.

Coefficiente de correlación lineal

Galton, por fin, introdujo el concepto de coeficiente de correlación lineal de **Pearson**. Curioso, ¿verdad? Curioso, digo, que lleve el nombre de Pearson. Como hemos dicho, Pearson fue su discípulo y precisó algunas deficiencias en la definición original de Galton. Con él podemos valorar la intensidad y el sentido de la relación lineal entre parejas de variables. El término correlación forma hoy en día parte de discusiones coloquiales.

Los datos estadísticos, que con frecuencia son tan difíciles de obtener, los conseguía en su laboratorio antropométrico inaugurado en la *International Health Exhibition (Feria Internacional de la Salud)* de 1884 y que mantuvo en funcionamiento durante 6 años en Londres. Este laboratorio le permitió recoger una ingente cantidad de datos y cobrar por los informes que realizaba. Un auténtico *consulting* estadístico que le convirtió en ser el primero capaz de cobrar a los sujetos experimentales.

Su interés por la medición fue quizá la característica más notable de todas sus investigaciones. Lo medía todo de manera obsesiva. Era lo que en lenguaje coloquial llamaríamos hoy un auténtico *friki*. Un excéntrico. Y si no me creen, presten atención a algunas de sus inquietudes intelectuales diferentes de las relatadas hasta ahora. Juzguen ustedes.

Mapa de la belleza de las mujeres

En 1859, que ya no era un niño pues contaba con 37 años, se planteó con su primo Darwin (¡que tenía 50!) obtener un mapa de belleza de las mujeres de las islas británicas. **Hugh Aldersey-Williams** en su libro *Anatomías* (Parra, 2013) detalla cómo hacían para tomar los datos de manera discreta. No olvidemos que estamos en la Inglaterra victoriana.

Para obtener los datos de lo que llamaba su “Mapa de la belleza”, cortaba un trozo de papel en forma de un crucifijo. Utilizando una aguja montada en un dedal, perforaba agujeros en el papel para clasificar las “muchachas junto a las que pasaba en las calles o en cualquier otro lugar, como atractivas, indiferentes o repelentes”. Los orificios para las chicas atractivas los hacía en la parte superior de la cruz, los correspondientes a las mujeres corrientes en el palo horizontal, y los de las feas en la base de la cruz. La ventaja de ello era que podía notar al tacto cada parte de la plantilla de papel en su bolsillo y registrar sus datos sin que las féminas de la ciudad que fuera le vieran ni sospecharan que las estaba evaluando.

Como supongo estarán interesados en saber los resultados de un estudio tan peculiar que hoy cerraría un telediario, hemos de decir que ganó Londres y cerró la clasificación Aberdeen, en el noreste de Escocia. Si sus mujeres no eran bellas para los primos, la ciudad es preciosa.

La eficacia de la oración

Otra de sus investigaciones más curiosas y que no estuvo exenta de polémica fue *Statistical Inquiries into the efficacy of prayer (Investigaciones Estadísticas sobre la eficacia de la oración)* de 1872. Galton trató de correlacionar los efectos de la oración con la duración de la vida. Como los clérigos se pasaban la vida orando y laborando, una mayor esperanza de sus vidas podía ser atribuida a los beneficios del rezo. Hay que recordar que Galton era profundamente religioso. Después de recopilar cientos de datos biográficos

llegó a la conclusión de que las expectativas de vida de médicos y abogados eran mayores que las de los clérigos. Hay que decir en su honor que no manipuló los datos para obtener lo que él deseaba, práctica que no acaba de caer en desuso.

De la longitud de las condenas al agua de cada día y otros asuntos

También analizó diez mil sentencias de jueces de la corte británica para hacer una distribución de frecuencias de la **longitud de las condenas**. Además determinó con exactitud y utilizando complicadas fórmulas matemáticas la **cantidad de agua que debía ingerirse** en los diferentes momentos del día para estar perfectamente hidratado.

Cuando asistía a las carreras de caballos, y lo hacía con frecuencia, obtenía distribuciones de frecuencias del **cambio de color en la cara de los asistentes** cuando los caballos se acercaban a la meta. Además diseñó un sombrero cuya parte superior se alzaba con una pera de goma permitiendo ventilar la cabeza en los días en los que el calor acosaba a los espectadores.

Creó un **índice de aburrimiento en los actos sociales**, tan habituales en la Inglaterra victoriana. Concluyó que los asistentes atentos se sentaban erguidos mientras que los aburridos se movían constantemente de delante a atrás y la frecuencia de estos movimientos podía relacionarse con el aburrimiento.

Para analizar las **relaciones y afinidades entre sus invitados** a casa, ideó un artilugio tipo sensor que, colocado bajo las sillas del salón, permitía cuantificar el número de veces que cada persona orientaba sus movimientos hacia cada uno de los demás. Es evidente que ya pensaba en las redes sociales, tan de moda ahora.

También llevó a cabo estudios sobre el **número de optimistas y pesimistas** con cuestionarios apropiados a los que respondían los *clientes* de su laboratorio.

En 1897 publicó nada más ni nada menos que en *Nature* un artículo estableciendo la **longitud que debía tener la soga del ahorcado** para fracturarle el cuello sin decapitarlo. Suena siniestro pero los intereses científicos en diferentes épocas han sido también muy distintos. A partir de los resultados de Galton y de otros científicos y con datos de ejecuciones fallidas, existe una tabla que relaciona peso del condenado con longitud de la soga. Y en la misma revista publicó un artículo sobre como cortar una tarta redonda de una manera científica.

Su afán por descubrir características individuales de las personas le llevó a demostrar que personas diferentes tienen **huellas dactilares diferentes**. Inmediatamente este proceso de identificación, popular hoy en todo el mundo, fue adoptado por Scotland Yard y permitió resolver un elevado número de delitos pendientes en aquel momento.

Ferviente defensor de las ideas de su primo, argumentó con tenacidad que las diferencias de aptitud en los seres humanos eran debidas a la herencia genética. Propuso un método de intervención social, la eugenesia, según el cual la reproducción debía planificarse para maximizar la inteligencia de los nacidos. Socialmente debía aceptarse que los “eminentes” tuvieran muchos hijos y los demás menos. Figuras como **Graham Bell**, **Bernard Shaw** o **Winston Churchill** apoyaron sus ideas. Hoy en día se asocia el uso de ellas con el nazismo y otros regímenes totalitarios. Desde 1970 ningún país del mundo admite la eugenesia en sus políticas.

La significación estadística de la Guinness

William S. Gosset (1876-1937), más conocido por su sobrenombre de **Student**, abrió nuevos caminos en la Estadística para investigadores empíricos. Dio con preguntas clave a las que acertó a dar respuesta. Algunas de ellas fueron posteriormente reformuladas matemáticamente por **Ronald A. Fisher**.

La distinción entre un parámetro poblacional como la media teórica y sus estimadores y sus estimaciones es algo que incluso hoy en día atormenta a los que se acercan por primera vez a la Inferencia Estadística. A principios del siglo XX los estadísticos no se habían planteado esta diferencia. Trabajaban a nivel teórico con muestras tan grandes que podían permitirse el lujo de identificar parámetro y estimación. Cuando se quisieron hacer trabajos empíricos, saltaron las alarmas.

Gosset era el mayor de los cinco hijos de un coronel de ingenieros reales del Imperio Británico. Siguió los pasos de su padre para convertirse en ingeniero real antes de ser rechazado por su gran miopía. Mucho de lo que sigue nos lo cuenta **Jane Fisher Box**, hija de Fisher, en su artículo de 1987. Respecto a la personalidad de Gosset afirma:

“Muy brillante, con altos ideales y un elevado sentido del humor. Tenía el carácter más atractivo que pueda existir; tranquilo, muy amigo de sus amigos, siempre dispuesto a ayudar, paciente y leal. Todo el mundo le

quería y confiaba en él. En el proceloso mundo de los estadísticos se movía siendo amigo de todos”.

Jamás tuvo un empleo como estadístico. Después de graduarse en Química en 1899 obtuvo un puesto como químico cervecero en la fábrica de Guinness de Dublín. Cuando murió en 1937, dirigía a todos los químicos de Guinness. Desempeñando este trabajo es como se convirtió en realidad en un gran estadístico.

En el último cuarto del siglo XIX, Guinness había pasado de ser una empresa familiar irlandesa a producir más de millón y medio de barriles con exportaciones al mundo entero. Irlanda es un país pequeño y todos los que vivían alrededor de la producción y comercio de cebada, lúpulo y malta, amplias poblaciones en el sur y en el este del país, prosperaban y tenían influencia social. En esta época **Cecil Guinness**, gerente de la firma, decidió acabar con el carácter artesanal transmitido de generación en generación y, con tintes de magia negra, en la fabricación de cerveza para introducir una metodología apoyada en el cientifismo. Y para ello puso mucho dinero sobre la mesa. Comenzó a contratar a los químicos más brillantes egresados de Oxford y Cambridge. Otras firmas empleaban químicos ocasionalmente, pero Guinness les concedió un gran peso específico dentro de la empresa. Y una vez que pasaban dos años auspiciados por químicos más veteranos, se les concedía una sección a su cargo con el correspondiente trabajo de investigación. Así que en 1893 se contrató a **Thomas Case**; en 1895 a **Alan McMullen**; en 1897 a **Arthur Jackson**; en 1899 a **E. G. Peake** y a **Gosset**; en 1900 a **Geoffrey Phillpotts** y así sucesivamente. Mientras permanecían solteros, vivían solos en la casa que a tal efecto tenía Guinness en St. James' Gate al lado de la propia factoría de Dublín. Durante las horas de trabajo incluso comían juntos en el salón comedor. Practicaban deportes en los alrededores y conversaban mucho entre ellos. La vida en el fondo seguía siendo muy similar a la de los *colleges* de los que procedían.

Una de las primeras cuestiones que planteó la empresa era cómo mejorar las materias primas, cebada y lúpulo, de manera que además resultaran más baratas. El típico reto realista empresarial que busca conseguir lo imposible. Además se necesitaba determinar qué variedades, qué tipo de cultivo y manipulación de la cosecha, qué condiciones de secado y almacenado proporcionarían una mayor calidad de la cerveza. Parece que realmente hasta entonces la cosa era pura magia y, en esta cerveza, obviamente negra.

Así que en 1899 Guinness empezó un proyecto de experimentación con selección de cultivos, variedades de cebada y lúpulo, así como de fertilizantes. Se consideró

la variedad de ensayos como el más prometedor de los caminos y se fueron incorporando al estudio más y más agricultores de las zonas productivas. La cuestión se convirtió en interés nacional (para entonces Guinness era una empresa pública aunque siguiera dirigida por la familia fundadora) y el propio Ministro de Agricultura irlandés participaba en la toma de decisiones.

La caducidad de la cerveza era otra cuestión de interés. Guinness era una cerveza natural, sin aditivos ni conservantes y sin pasteurizar, que debía mantenerse imperecedera durante su transporte a largas distancias así como durante su almacenamiento a muy diferentes temperaturas, incluso en países de climas tórridos. Todos estos eran los retos que tenían ante sí nuestros jóvenes talentos.

Imaginemos entonces a media docena de investigadores jóvenes, brillantes y enérgicos con mano ancha para tomar decisiones sobre cuestiones como las referidas pero desde una posición de casi total ignorancia. Eso sí, con campos de cebada y lúpulo, laboratorios y hasta una destilería a su disposición.

Lo primero que hicieron fue ponerse a revisar la literatura existente. Y se enteraron de que el lúpulo almacenado a temperatura ambiente se deterioraba al endurecerse su resina, así como de que el lúpulo americano era más barato y con mejor resina que el europeo. Leyeron estudios en los que se argumentaba que la calidad de la cebada dependía del contenido de nitrógeno.

Con estas mimbres empezaron a acumular datos obtenidos de las parcelas de experimentación y de su producción, pero los resultados que encontraban no eran del todo concluyentes o, lo que es peor, con frecuencia no eran capaces de discernir si las diferencias encontradas eran realmente debidas a los tratamientos o podían considerarse puramente aleatorias. Dos dificultades confluían: alta variabilidad y muestras pequeñas. Los enemigos número uno de los estadísticos.

Pero los jóvenes investigadores trabajaban bien juntos. Algunos se habían convertido en auténticos amigos. Y para todos era natural acudir con los problemas de índole numérica a Gosset. Este había cursado asignaturas de matemáticas en Oxford y los demás le consideraban el único dotado para resolver la parte estadística de sus problemas. Por su parte, la bonhomía de Gosset le hacía escuchar con atención sus inquietudes, ir a la raíz de las cuestiones y desde ella tratar de encontrar una solución. Además era perseverante y obsesivo. No cejaba hasta que la obtenía. Cogió el libro de **Aisy** sobre la teoría de errores y

anotó muchas observaciones y dudas en los márgenes. El libro era suyo. En 1904 escribió un artículo de régimen interno sobre los errores estándar. Pero seguía muy confuso y decidió consultar a **Karl Pearson**, que era la voz autorizada de la época en Biometría.

El escaso tamaño muestral seguía siendo su talón de Aquiles. Los experimentos con la cebada se habían llevado a cabo con cuatro parcelas y sembrando una variedad en cada una. La estimación de la media en tal caso es obviamente inexacta y el error de la desviación estándar no puede ignorarse. En 1905 le planteó sus dos dudas principales a Pearson y éste no pudo resolverlas:

1. *¿Cuál debería ser el valor máximo aceptable del error para poder sustituir los parámetros poblacionales por sus estimaciones muestrales?*
2. *¿Qué nivel de probabilidad debemos considerar significativo?*

Se encontraron en la casa de verano de Pearson en julio de 1905. Gosset había vuelto a pasar unos días de vacaciones a Inglaterra con sus padres y pedaleó los 30 Km. que separaban la casa de éstos de la de Pearson. La conversación, no obstante, fue fructífera para Gosset pues Pearson le aclaró todos los métodos de decisión estadística existentes por entonces. **Joan Fisher** (1987) nos recuerda el comentario que Gosset dedicó a Pearson tras la conversación de ambos:

“He was able in about half an hour to put me in the way of learning the practice of nearly all the methods then in use”.

El verano acabó, Gosset volvió a Irlanda y empezó a poner en práctica esos métodos ya como jefe de la destilería experimental, en la que el análisis de los datos era el objetivo principal. El encuentro entre Pearson y Gosset se valoró tanto incluso dentro de la empresa Guinness que ésta le permitió ir a formarse con Pearson en el departamento de éste durante el curso académico 1906-1907 en la University College de Londres. Allí trabajó en la distribución del estadístico *t* de Student para tamaños muestrales de dos a diez y obtuvo las tablas de probabilidad correspondientes.

Regresó a Guinness justo a tiempo para aportar en el informe sobre siete años de experimentación todo el análisis estadístico. Y Guinness obtuvo su recompensa. Descubrieron que la variedad de cebada Archer era la más apropiada para ser cultivada en Irlanda y decidieron hacerla crecer por toda la isla. Mil barriles de Archer puro danés estaban en venta, los compraron todos y a lo largo de dos cosechas produjeron semillas para toda Irlanda.

En 1908 y utilizando mayoritariamente su tiempo libre, Gosset escribió sus dos famosos artículos sobre el error probable de la media y sobre el coeficiente de correlación (1908a, 1908b). Contra la opinión extendida, según J. Fisher (1987) Guinness sí aceptaba que sus investigadores publicaran artículos científicos. Lo que no se permitía era que aparecieran sus nombres. Las normas de la empresa rezaban:

“It was decided that such publication might be made without the brewer’s names appearing. They would be merely designated Pupil or Student”.

Así que Gosset decidió elegir Student. Posteriormente Fisher se dio cuenta de que faltaba alguna constante en la definición del estadístico t pero esa es ya otra historia.

Lo que hay detrás de una pinta de Guinness...

El carácter de Fisher

Ronald Aylmer Fisher (1890-1962) es considerado el fundador de la Estadística moderna por sus grandes contribuciones en este campo: el análisis de la varianza, la idea de la máxima verosimilitud, la inferencia estadística y la derivación de varias distribuciones asociadas al muestreo. Pero también destacó como biólogo evolutivo (se dice que en su época fue el mejor después de Darwin) y por sus aportaciones a la genética y a la eugenesia.

Tuvo una juventud desgraciada, pues a los 14 años perdió a su madre y unos años después su padre se arruinó, lo que dificultó el pago de los gastos de formación de él y de sus hermanos. Sin embargo, su excelente rendimiento académico en Harrow, donde incluso llegó a ganar un premio en matemáticas, motivó que Cambridge le becara con 80 libras de la época para cursar estudios de matemáticas y astronomía. Era octubre de 1909. Permanece allí hasta su graduación en 1912. Unos meses antes de graduarse conoce a Gosset (Student). Para entonces Fisher ha publicado su primer artículo: *“On an Absolute Criterion for Fitting Frequency Curves”*, en el que introduce la idea de función de verosimilitud y el método de máxima verosimilitud aunque sin citar el término verosimilitud (*likelihood*). Fisher pide a Gosset que lo lea y, a pesar del carácter bondadoso de Gosset, su valoración fue:

“Hasta donde puedo entenderlo es poco práctico y una manera inútil de mirar a las cosas”.

En sus propias palabras, como cita J. Fisher Box (1987):

“As far as I could understand it, quite unpractical and unserviceable way of looking at things”.

Aquí empezaba la incompreensión que siempre sintió que le profesaban sus evaluadores.

Contadas todas estas cosas por la hija de Fisher, uno lo imagina como un estudiante arrogante que al acercarse al profesor lo predispone, de entrada, en contra. No obstante, era perseverante y brillante y unos meses después mantuvo correspondencia frecuente con Gosset hasta que le hizo ver el error que éste había cometido en la formulación inicial del estadístico t . Era un intercambio de ideas complejo porque Fisher utilizaba un prolijo aparato matemático y Gosset no lo entendía del todo. Sin embargo, de esta correspondencia epistolar nació la afición de Fisher por la teoría de los errores (leyó el manual de Airy sobre el tema como lo había hecho Gosset) y prolongó su interés por los temas estadísticos.

Al acabar en Cambridge, Fisher carecía de recursos para afrontar el futuro y decidió viajar a Canadá, donde encontró trabajo en una granja. Cosas parecidas hacen algunos de nuestros jóvenes hoy en día y nos extraña. Regresó a Londres y obtuvo una plaza de estadístico en la *Mercantile and General Investment Company*. Poco después estalló la Primera Guerra Mundial. De todos es conocido el entusiasmo y la fascinación que generó en la juventud inglesa el alistamiento para defender una noble causa que en Navidad quedaría resuelta. Afortunadamente para él y seguramente para la ciencia, Fisher fue rechazado por escasa agudeza visual. Esta falta de aptitud para el ejército y el motivo mismo es un elemento común entre su biografía y la de Student. Seguro que alguna vez lo comentaron entre ellos. Esquivó así trincheras, barro, enfermedades y descargas de ametralladora..., los iconos de aquella conflagración.

Durante la guerra sirvió a su país enseñando Física y Matemáticas en varias escuelas. Pero nunca destacó por sus habilidades docentes y se consideraba llamado a más altas metas. Así que, interesado profundamente en la eugenesia, y con su experiencia canadiense decidió comprar unos terrenos y montar una granja. Por esta época, todavía en plena Primera Guerra Mundial, conoció a través de la esposa de un amigo de juventud a la hermana de ésta, **Ruth Gratton**, que solo contaba 16 años. Se enamoraron como solo a esas edades

sucede y planearon casarse. La madre de Ruth, viuda, puso el grito en el cielo y decidieron casarse en secreto unos días antes de que la novia cumpliera 17 años. El amor duró y fue fructífero, pues tuvieron 9 hijos, uno de los cuales pereció durante la infancia.

En 1917, unos meses después de la boda, empieza una de las disputas más famosas entre científicos. Pearson publicó un artículo argumentando que Fisher no había sido capaz de distinguir entre verosimilitud y “*probabilidad inversa*” en su trabajo de 1915. Fisher solo estaba empezando su carrera pero aceptó de muy mal talante la crítica y el hecho de que Pearson no le hubiera comentado previamente nada. Por otro lado consideró que la crítica no se ajustaba a la realidad. Se quejó siempre, con cierta petulancia, pero con razón a veces, de las injusticias que sufría cuando sus artículos eran rechazados...

“por matemáticos que no saben de biología y biólogos que no entienden matemáticas”.

Dijo esto en 1918 después de someter a la Royal Society su artículo *On the correlation between relatives on the supposition of mendelian inheritance*. Ninguno de los dos revisores (Pearson era uno de ellos) rechazó el artículo pero expresaron sus reservas y reconocieron que no eran competentes para juzgar algunos aspectos. Fisher decidió enviar el artículo a *Transactions of the Royal Society of Edinburgh*, donde fue aceptado. La relación se volvió todavía más amarga cuando Fisher (como editor de *Biometrika*) desacreditó el uso del test *chi-cuadrado* de Pearson. Fue aún más lejos y afirmó que Fisher había hecho un flaco favor a la Estadística diseminando resultados erróneos. Como consecuencia de ello la Royal Statistical Society se negó a publicar artículos de Fisher y, en represalia, Fisher la abandonó. También Fisher aprovechaba cada oportunidad que tenía para atacar a Pearson incluso después de muerto éste en 1936. Hay odios que se llevan a la tumba.

No obstante, en 1919 cuando ya sus relaciones eran turbias, no sé si porque los enemigos se necesitan, Pearson ofreció a Fisher el puesto de jefe estadístico del laboratorio Galton. Fisher lo rechazó y prefirió aceptar otra oferta que tenía sobre la mesa como estadístico en la Estación Experimental Agrícola de Rothamsted, que era el más antiguo centro de investigación agrícola del Reino Unido. Su afición por la agricultura, así como su animadversión con Pearson, estuvieron en el fondo de esta elección. Aquí hizo sus grandes contribuciones al análisis de la varianza y al diseño de experimentos utilizando datos como los clásicos de la flor iris y sus variedades.

Irónicamente, cuando Pearson se retiró en 1933 de su puesto en la University College, Fisher ocupó su cátedra. En realidad fue peor: la cátedra se duplicó, ocupando Fisher una y la otra el hijo de Pearson, **Egon**. A veces pensamos que cosas así solo ocurren por estos pagos. El resentimiento de Fisher con Pearson (padre) afloraba continuamente y, estando presente el hijo por allí, la atmósfera en la University College se hizo bastante irrespirable (O'Connor y Roberston, 2003).

Fisher se retiró de la Universidad en 1957. En 1959 se trasladó a Adelaida (Australia), donde un cáncer de colon terminó con su vida en 1962.

En Bennett (1989) se describe el carácter de Fisher así:

“Era muy capaz y encantador, cálido con sus amistades. Pero así mismo era víctima, como él mismo reconocía, de un temperamento incontrolable y de su devoción por la verdad en la ciencia, que demostró apasionadamente siendo enemigo implacable de los que le acusaban de propagar errores”.

También Bennett (1989) afirma en relación con sus habilidades y deficiencias:

“Su capacidad de penetración era espectacular, pero sus escritos son difíciles para muchos lectores. Realmente muchos de sus resultados han sido mejor divulgados por otros que fueron capaces de modificar su formulación. Como profesor, Fisher también fue difícil para el estudiante medio. Sus clases rápidamente se vaciaban hasta que solo tres o cuatro estudiantes podían seguir, eso sí, fascinados, el paso. Tampoco fue un gestor muy exitoso; quizá era incapaz de aceptar las limitaciones de las personas normales. Pero con su amplio espectro de interés y la capacidad de penetración de su mente era un conversador muy estimulante y simpático”.

Finalmente su afición a fumar en pipa le hizo perder la perspectiva al negar la relación entre cáncer de pulmón y tabaquismo. También pudo estar influido por sus conflictos personales y profesionales. Además trabajaba como consultor para las tabacaleras. En fin, reproduzco el *abstract* del artículo de **Stolley** (1991):

“R. A. Fisher’s work on lung cancer and smoking is critically reviewed. The controversy is placed in the context of his career and personality. Although Fisher made invaluable contributions to the field of statistics, his analysis of the causal association between lung cancer and smoking was flawed

by an unwillingness to examine the entire body of data available and prematurely drawn conclusions. His views may also have been influenced by personal and professional conflicts, by his work as a consultant to the tobacco industry, and by the fact that he was himself a smoker”.

Me quedo con dos frases de Fisher que en algún momento del curso siempre cito a los alumnos. La primera hace referencia a las consecuencias de la falta de control sobre el diseño del experimento y recopilación de datos por parte del estadístico. Y dice:

“Consultar a un estadístico después de que un experimento haya terminado es, a menudo, como pedirle que haga una autopsia. Quizá lo más que pueda decir es de qué murió el experimento”.

La segunda resulta muy clarificadora sobre el verdadero sentido de la hipótesis nula en los contrastes. Fisher afirmaba:

“La hipótesis nula no se considera aprobada pero puede ser rechazada en el curso de la experimentación. La razón de existir de un experimento es conceder a los hechos la oportunidad de rechazar la hipótesis nula” .

Los Big Data

Borel, Galton, Student, Pearson, Fisher y muchos otros injustamente ignorados en esta lección inaugural fueron los artífices de la primera revolución estadística. Algunos como Student y Fisher desarrollaron procedimientos de análisis de datos a partir de las dificultades que planteaba la escasez de información, el bajo tamaño muestral. Exactamente un siglo después tenemos el problema inverso: el análisis de los datos masivos, más habitualmente conocidos por su terminología inglesa, los Big Data.

Hoy en día la sociedad y el mundo empresarial están reclamando la aparición de los científicos que lleven a cabo la segunda revolución estadística, la de los datos masivos. Big Data es un término difuso pero ilustra una nueva era. Los manuales de estadística empezaban introduciendo el concepto de muestra para justificar la reducción del estudio a unos cuantos individuos representativos de la población dado que acceder a toda ella resultaba, en general, impracticable. Pero hoy en día la toma de datos empieza a ser tan accesible y las posibilidades de análisis

tan grandes que ya estamos tratando de estudiar el conjunto masivo de datos para obtener información más precisa, valiosa y sorprendente.

Si una característica define a las personas hoy en día es que somos generadoras de datos. Pasamos el día enviando *whatsApps*, correos electrónicos, haciendo búsquedas en Internet (de las que se registra el tiempo que pasamos en cada página y cuándo las visitamos), *twiteando*, pagando con tarjeta y diciendo nuestro código postal en la caja del comercio donde acabamos de comprar. Si cambiamos de país nos llegan automáticamente mensajes al móvil sobre las tarifas de *roaming* y, al alquilar un coche, estamos informando permanentemente a la compañía sobre dónde estamos (o por lo menos sobre dónde está su coche). La mayor parte de los *parkings* reconoce nuestra matrícula y, cuando inmigramos, en algunos países nos escanean la retina.

Se habla de que Google puede disponer de un volumen de datos que ocupa un millón de discos duros de gran capacidad. Sinceramente no soy capaz de imaginar el volumen de datos almacenados en los servicios de seguridad, defensa e inteligencia de los estados.

La información es el nuevo metal precioso. De hecho, algunas de las técnicas de análisis de Big Data se conocen como Minería de Datos en referencia a la carrera codiciosa con que las empresas buscan hoy metodologías y aplicaciones para obtener las pepitas de la correlación de la misma manera que California se desarrolló por la fiebre del oro. Si Guinness tuvo parte activa en la primera revolución estadística, otras empresas van a entrar en la historia de la segunda.

UPS, la empresa mundial de mensajería y paquetería, ha sido una de las primeras en explotar la información masiva con resultados económicos espectaculares. Desde los 80 maneja datos de millones y millones de clientes y paquetes. Sin embargo, en los últimos años sus ventajas competitivas han procedido de la información proporcionada por los sensores colocados en unos 50.000 vehículos de reparto. Estos sensores recolectan los datos sobre velocidad, dirección y trayectos. El análisis de ellos ha permitido a la empresa un rediseño de las rutas de distribución en aras de una mayor eficiencia. El proyecto, que en su génesis se llamó ORION (On Road-Integrated Optimization and Navigation), está permitiendo incluso modificar las rutas y distribución en tiempo real. En 2011, supuso un ahorro para la empresa de 32 millones de litros de combustible a base de recortar las rutas en unos 135 millones de kilómetros. Encontrar la pepita de oro para ellos fue conseguir que cada repartidor, en promedio, redujera en un par de kilómetros su ruta diaria.

En Wikipedia podemos leer otras aplicaciones del análisis de los Big Data, algunas de las cuales, confieso que me intimidan. Oracle dispone de una herramienta que integra información procedente de Facebook, Twitter y LinkedIn. Los departamentos de recursos humanos la utilizan para obtener perfiles de candidatos basados no solo en sus aptitudes profesionales sino también en las sociales. Podemos llegar a ser transparentes para ellos.

Amazon basa su éxito en ventas cruzadas relacionando los patrones de compra de los usuarios lo que permite crear anuncios personalizados y boletines electrónicos ofertando al instante lo que quiere el usuario. Mejor dejar la tarjeta de crédito en otra habitación y cruzar los dedos para que se nos pase el impulso de compra mientras nos levantamos de la silla y vamos a por ella.

La cantidad de información que generamos los usuarios en la red es utilizada por las empresas llamadas corredores de datos para segmentar los individuos actualmente en unas 70 tipologías en función de sus aficiones, estilo de vida y actividades.

Nuestro móvil se ha convertido en un detective implacable que registra nuestras diferentes ubicaciones, cuándo llegamos a ellas y cuándo nos vamos. Cuidado con lo que usted hace. Orwell anticipó un poco el Gran Hermano pero 30 años después ha llegado. Las ilusiones por prosperar profesional o socialmente o las metas políticas pueden estrellarse brutalmente contra las huellas dejadas en una red social o en el móvil.

Después de la moralina bueno es reconocer otras virtudes de los Big Data. Y una de las principales es que es un excelente proveedor de puestos de trabajo. Los analistas de datos tienen un futuro prometedor en el que su creatividad se va a ver puesta a prueba. La formación en este contexto, también. Tenemos por delante retos como el almacenamiento de los datos masivos de forma más eficiente y prolongada, la mejora de su contextualización, reducción y monitorización así como el desarrollo de herramientas para su visualización. Los profesionales más valorados de un futuro que ya ha llegado son expertos en computación en la nube (*Cloud Computing*), en extracción de información y visualización de los datos y en la ciencia de los datos (*Data Science*). Empiezan a ser conocidos como los científicos de los datos. Deben atesorar conocimientos de programación y estadística, tratamiento y análisis de datos, teoría de grafos y saber presentar los resultados bajo una visualización eficaz. Los hay ya pero debe haber más.

Bien, sólo he pretendido con estas miniaturas históricas hacer más simpáticas y próximas algunas cuestiones relacionadas con la Probabilidad y la Estadística. Citar el Teorema de Borel-Cantelli los hubiera hecho removearse en sus asientos, pero imaginar unos monos escribiendo textos literarios espero les haya arrancado una discreta sonrisa. Así mismo hemos visto que hasta los más estrafalarios experimentos, que le pregunten a Galton, pueden impulsar la ciencia. No hemos inventado las luchas intestinas en los departamentos, ni en las facultades ni en la universidad, pero no estaría de más tratar de evitarlas. En el imaginario colectivo, ni Pearson ni Fisher salen reforzados con sus relaciones tempestuosas. Imaginar a Student con sus gafitas pedaleando 30 Km. para preguntar unas dudas a un profesor, además de ser una imagen muy *british*, nos reconcilia con el valor de nuestra profesión y pensar en el futuro de la Estadística así como el de las demás disciplinas, nos dinamiza e impulsa para afrontar cada curso académico. Bien es cierto que, seguro, ustedes ya sabían todo esto.

Nada más por hoy. Muchas gracias.

Bibliografía

BENNETT, J. H. (ed.) (1989) Statistical inference and analysis: selected correspondence of R. A. Fisher. Oxford.

Big Data. En Wikipedia. Recuperado el 1 de julio de 2015 de https://es.wikipedia.org/wiki/Big_data

El teorema de los infinitos monos (2009). Recuperado el 24 de junio de 2015 de <http://www.neoteo.com/el-teorema-de-los-infinitos-monos>

FISHER BOX, J. (1987) Guinness, Gosset, Fisher, and Small Samples. *Statistical Science*, vol. 2,1, pp. 45-52.

FRANCIS GALTON. En Wikipedia. Recuperado el 10 de julio de 2015 de https://es.wikipedia.org/wiki/Francis_Galton

GALTON, F. (1872) Statistical inquiries into the efficacy of prayer. *Fortnightly Review* vol. 12, pp. 125-35.

GONZÁLEZ, J. J. Regresión a la media: Un fenómeno estadístico con historia y repercusión social. Recuperado el 10 de julio de 2015 de <http://www.ugr.es/~jmcontreras/thales/1/ComunicacionesPDF/RegresionMedia.pdf>

HUXLEY, THOMAS HENRY. En Wikipedia. Recuperado el 26 de junio de 2015 de https://es.wikipedia.org/wiki/Thomas_Henry_Huxley

LÓPEZ NICOLÁS, J. M. (2011) Sir Francis Galton..., el hombre capaz de medirlo todo. Recuperado el 27 de junio de 2015 de <http://scientiablog.com/2011/07/19/sir-francis-galton-el-hombre-capaz-de-medirlo-todo>

O'CONNOR, J. J. y ROBERSTON, E. F. (2003) Sir Ronald Aylmer Fisher. Recuperado el 8 de julio de 2015 de <http://www-history.mcs.st-andrews.ac.uk/Biographies/Fisher.html>

PARRA, S. (2013) Mapa de la belleza: ¿Dónde viven las chicas más guapas... según Darwin? Recuperado el 27 de junio de 2015 de <http://www.diariodelviajero.com/cajon-de-sastre/mapa-de-la-belleza->

ROUSSEUW, P. J. (1991) Why the wrong papers get published. *Chance*, vol. 4, pp. 41-43.

STOLLEY, P. D. (1991) When genius errs: R. A. Fisher and the lung cancer controversy. *American Journal of Epidemiology*, vol. 133, 5, pp. 416-425.

STUDENT (1908a) The probable error of a mean. *Biometrika*, vol. 6,1, pp.1-25.

STUDENT (1908b) Probable error of correlation coefficient. *Biometrika*, vol. 6,2, pp. 302-310.

Teorema del mono infinito. En Wikipedia. Recuperado el 24 de junio de 2015 de https://es.wikipedia.org/wiki/Teorema_del_mono_infinito

