suma+ 74
Artículos
Noviembre 2013
pp. 5-<#>

Peces, taxis, epidemias... Ideas para crear una actitud positiva hacia la estadística

Pere Grima Cintas

Este artículo presenta algunas ideas que pueden ser útiles a profesores de enseñanza secundaria para discutir casos y plantear actividades que motiven interés por la estadística a la vez que se transmiten nuevos conocimientos. Se hace uso de una hoja de cálculo como instrumento para realizar representaciones gráficas, cálculos y simulaciones.

Palabras clave: Didáctica, estadística, casos prácticos, motivación, secundaria

Fish, taxis, epidemics ... Ideas to create a positive attitude towards statistics

This article presents some ideas that may be useful to high school teachers to discuss cases and propose activities that motivate interest in statistics while transmitting new knowledge. It makes use of a spreadsheet as a tool for graphics, calculations and simulations.

Keywords: Didactics, statistics, practical cases, motivation, secondary

Walter Lewin es un profesor emérito de física del Instituto Tecnológico de Massachusetts (MIT) que se hizo famoso por la originalidad y la pasión con que desarrollaba sus clases, muchas de las cuales han sido vistas por millones de personas en YouTube (en una de las más famosas se cuelga él mismo de un péndulo para mostrar que el periodo no depende de la masa). Ha escrito un libro muy interesante, tanto para los aficionados a la física como a la enseñanza, donde explica sus clases comentando demostraciones prácticas y golpes de efecto. Respecto a la forma de enseñar dice:

> Mi objetivo es conseguir que les guste la física y que miren el mundo de manera diferente, ¡eso es para toda la vida! Amplías sus horizontes, lo que les permite plantearse preguntas que nunca antes se habían hecho. Lo importante es descubrirles el mundo de la física de tal forma que conecte con sus intereses reales. (Lewin, 2013: 293)

Creo que este planteamiento sirve también para la enseñanza de la estadística (y de otras disciplinas, por supuesto), tanto en la enseñanza secundaria como en los cursos generales que se imparten en muchas carreras universitarias[1]. Los objetivos no deben estar solo orientados a los conocimientos sino también a crear actitudes positivas hacia esta materia. Seguramente lo más importante es que cuaje la idea de que la estadística es una herramienta

Artículo recibido en *Suma* en ****** de 2011 y aceptado en **** de 2011

imprescindible para conocer mejor la realidad (natural, física, social,...) y para obtener información que permita tomar mejores decisiones.

Una forma de crear esa actitud es incorporando a las clases casos prácticos que despierten el interés de los alumnos y que pongan de manifiesto como el uso de la estadística (entendida en un sentido muy amplio) permite obtener información de utilidad. Por supuesto, también deben servir para sacar lecciones (conocimientos) respecto a las técnicas usadas y sobre sus posibilidades de aplicación en otras situaciones similares.

A continuación se comentan algunos ejemplos que se pueden adaptar para desarrollar en clase o servir de inspiración para crear otros similares. Están divididos por temas y, como es habitual, empezamos por la estadística descriptiva.

Estadística descriptiva

Una idea importante es que utilizar la estadística no necesariamente implica el uso de técnicas sofisticadas. Un ejemplo paradigmático es el análisis sencillo pero muy eficaz que permitió acabar con las epidemias de cólera en Europa. Por otra parte, en internet se pueden obtener muchos datos que permiten mostrar información interesante aplicando técnicas sencillas.

Cólera en Londres

En 1854 se desató una terrible epidemia de cólera en el centro de Londres que ocasionó más de 500 muertos en apenas 10 días. En aquella época imperaba la teoría de la miasma como vía de transmisión de la enfermedad (se suponía que «el veneno» estaba en los vapores que se desprendían de los cuerpos en descomposición). Pero al Dr. John Snow esta teoría no le cuadraba con sus observaciones y sospechó que el problema podía estar en el agua. Para estudiar la situación marcó sobre un plano de la zona el lugar donde vivían los fallecidos y observó que las marcas se distribuían en torno a una fuente (la de Broad Street). Los datos y la forma de presentarlos era tan clara y convincente que, a pesar de lo estrafalaria que parecía su teoría, las autoridades ordenaron inutilizar la fuente y a partir de aquel momento los contagios cayeron en picado. Aunque no se entendía muy bien cómo el agua podía contagiar la enfermedad, las ciudades de los países desarrollados prestaron mucha atención a la higiene en la distribución de las aguas y el cólera dejó de ser una amenaza en esos países.

¿Qué se aprende?: La importancia de basar las decisiones en base a datos. Un buen diseño de la recogida de datos facilita su análisis posterior.

Otras actividades: Además de buscar información en internet sobre este tema (hay mucha) se pueden discutir las posibilida-

Figura 1. Localización de los muertos por cólera realizada por John Snow sobre un plano de la época (vista parcial)
Fuente: *Wikipedia:* «1854 Broad Street cholera outbreak»

des de uso de este tipo de representación gráfica en otros ámbitos: Identificación de puntos negros en las carreteras o la localización más frecuente de los defectos en un producto.

Más información en: Wikipedia: «John Snow» (más amplio en inglés) y «1854 Broad Street cholera outbreak». Página web dedicada al Dr. Snow elaborada por el Departamento de Epidemiologia de la UCLA: <http://www.ph.ucla.edu/epi/snow.html>, y también en Tufte (1997, 27-37).

Coches caros, coches baratos

Usando datos que se pueden conseguir a través de internet, la estadística descriptiva puede poner de manifiesto aspectos interesantes y no evidentes en áreas que despiertan interés, como características de coches, motos u ordenadores.

Por ejemplo, se pueden obtener datos de coches que cumplan unas determinadas condiciones en la página web

www.autopista.es/coches/nuevos

Buscando coches con combustible «gasóleo» y tipo «berlina» se obtienen datos de 628 coches (29 de agosto de 2013) y se puede analizar la relación entre precio y potencia.

Pasar los datos de la página web a una hoja de cálculo tipo Excel y colocarlos de forma que resulten fáciles de analizar no es inmediato y esta es una de las actividades que a menudo resultan más tediosas y consumidoras de tiempo, pero quizá por eso mismo conviene adquirir experiencia en este terreno. Una forma rápida de hacerlo es copiar los datos de la página web y pegarlos en el bloc de notas para, desde ahí, importarlos con Excel marcando el tabulador como separador de campos.

Analizando los datos (figura 2) se observa, por ejemplo, que la relación precio-potencia

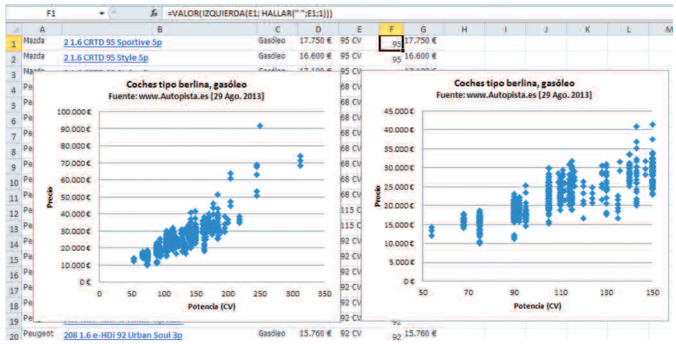


Figura 2. Relación entre precio y potencia de un conjunto de 628 coches de gasóleo tipo berlina.
A la izquierda están representados todos los datos y a la derecha solo los coches con potencias entre 50 y 150 CV

no se ajusta bien mediante una recta (en los coches con potencia muy alta el precio se dispara). Adaptando la escala de los ejes para que solo aparezcan coches con potencias entre 50 y 150 CV (Figura 2, derecha) se descubre que, por ejemplo, para 105 CV existe un coche que vale en torno a 15 000 € (Fiat Bravo 1.6 Multijet 105 Pop 5p) y otro que con la misma potencia vale aproximadamente el doble (Audi A3 Sportback 1.6 TDI 105 Ambition 5p S tronic).

¿Qué se aprende?: Técnicas de análisis gráfico de datos. Uso de la hoja de cálculo para construir gráficos.

Otras actividades: Pueden realizarse investigaciones similares para analizar la relación entre distintas características técnicas de coches, motos, e incluso aviones. También se puede analizar la relación entre el precio y características relevantes de otros productos como ordenadores o cámaras fotográficas.

Estimación de características de la población a partir de una muestra

Esta es una de las actividades típicas con que se relaciona la estadística. A menudo se trata de estimar una proporción en la población (proporción de individuos que consume un determinado producto, que vio un programa de televisión o que piensa votar a un determinado partido), pero en los ejemplos que planteamos a continuación lo que se pretende es estimar el número de elementos que contiene la población.

¿Cuántos peces hay en un lago?

Parece imposible que se pueda responder a esta pregunta, especialmente si el lago es grande y las aguas turbias, pero utilizando métodos razonables (no hace falta vaciar el lago, ni pescar todos los peces, como proponen al principio algunos estudiantes) se puede tener una aproximación suficientemente buena.

Se trata del método de la pesca y repesca: se capturan M peces, se marcan y se devuelven al agua; se deja pasar un tiempo para que peces marcados se dispersen por el lago y se capturan otros C, de los cuales R aparecen marcados. Si N es el número total de peces, es razonable considerar que la proporción M/N (total de peces marcados respecto al total de peces del lago) será parecida a R/C (peces que aparecen marcados en la segunda captura respecto al número de peces capturados) y ya podemos despejar N para tener una primera aproximación del número de peces que hay en el lago:

$$\frac{M}{N} \cong \frac{R}{C}, \quad N \cong \frac{M \cdot C}{R}$$

Con los datos de la figura 3 nuestra estimación para el número total de peces sería

$$N = \frac{M \cdot C}{R} = \frac{15 \cdot 15}{3} = 75$$

Obviamente no se trata de un método exacto, en realidad en el dibujo hay 62 peces y, por tanto, cometeríamos un error del 12 %, pero permite tener un valor útil a efectos prácticos y que se puede obtener con un esfuerzo razonable[2].

La calidad de la estimación obtenida mediante este procedimiento u otros similares es tratada con detalle por la teoría estadística pero también puede indagarse por simulación utilizando *applets* como la incluida en un CD que acompaña al libro «Ecología con números» de J. Piñol y J. Martinez-Vilalta (figura 4) y que también se puede descargar de internet[3]. Otra opción puede ser una hoja de cálculo como se indica en la figura 5.

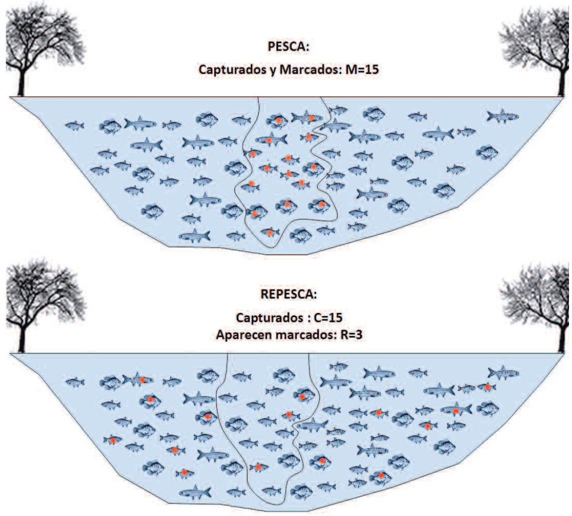


Figura 3. Estimación del número de peces por el método de la pesca y repesca

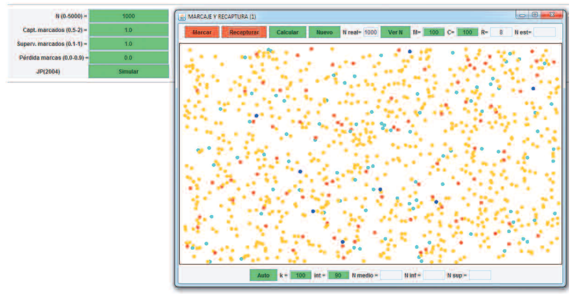
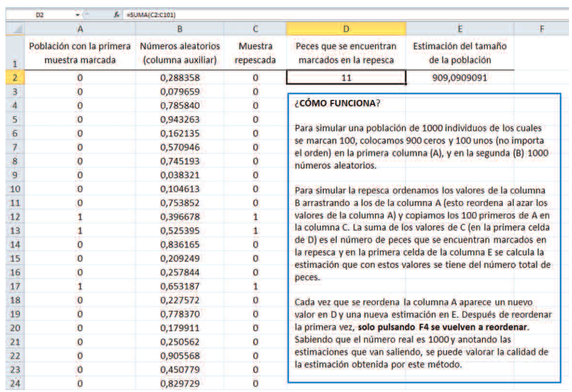


Figura 4. *Applet* para simular el proceso de pesca y repesca (de Piñol y Martinez-Vilalta, 2006)



D2 =SUMA(C2:C101)

	A	B	C	D	E	F
1	Población con la primera muestra marcada	Números aleatorios (columna auxiliar)	Muestra repescada	Peces que se encuentran marcados en la repesca	Estimación del tamaño de la población	
2	0	0,288358	0	11	909,0909091	
3	0	0,079659	0			
4	0	0,785840	0			
5	0	0,943263	0			
6	0	0,162135	0			
7	0	0,570946	0			
8	0	0,745193	0			
9	0	0,038321	0			
10	0	0,104613	0			
11	0	0,753852	0			
12	1	0,396678	1			
13	1	0,525395	1			
14	0	0,836165	0			
15	0	0,209249	0			
16	0	0,257844	0			
17	1	0,653187	1			
18	0	0,227572	0			
19	0	0,778370	0			
20	0	0,179911	0			
21	0	0,250562	0			
22	0	0,905568	0			
23	0	0,450779	0			
24	0	0,829729	0			

¿CÓMO FUNCIONA?

Para simular una población de 1000 individuos de los cuales se marcan 100, colocamos 900 ceros y 100 unos (no importa el orden) en la primera columna (A), y en la segunda (B) 1000 números aleatorios.

Para simular la repesca ordenamos los valores de la columna B arrastrando a los de la columna A (esto reordena al azar los valores de la columna A) y copiamos los 100 primeros de A en la columna C. La suma de los valores de C (en la primera celda de D) es el número de peces que se encuentran marcados en la repesca y en la primera celda de la columna E se calcula la estimación que con estos valores se tiene del número total de peces.

Cada vez que se reordena la columna A aparece un nuevo valor en D y una nueva estimación en E. Después de reordenar la primera vez, **solo pulsando F4 se vuelven a reordenar**. Sabiendo que el número real es 1000 y anotando las estimaciones que van saliendo, se puede valorar la calidad de la estimación obtenida por este método.

Figura 5. Simulación del proceso de pesca-repesca con una hoja de cálculo. Puede bajarse de <http://www-eio.upc.es/~grima/PecesSimulacionExcel.xls>

También puede hacerse una demostración práctica invitando a los estudiantes a estimar el número de bolas que hay en un recipiente. Nuestro método será coger 100 bolas (por ejemplo) marcarlas —aunque en este caso es más cómodo sustituirlas por bolas rojas— mezclar bien, sacar otras 100 y contar cuantas rojas aparecen (figura 6)[4].

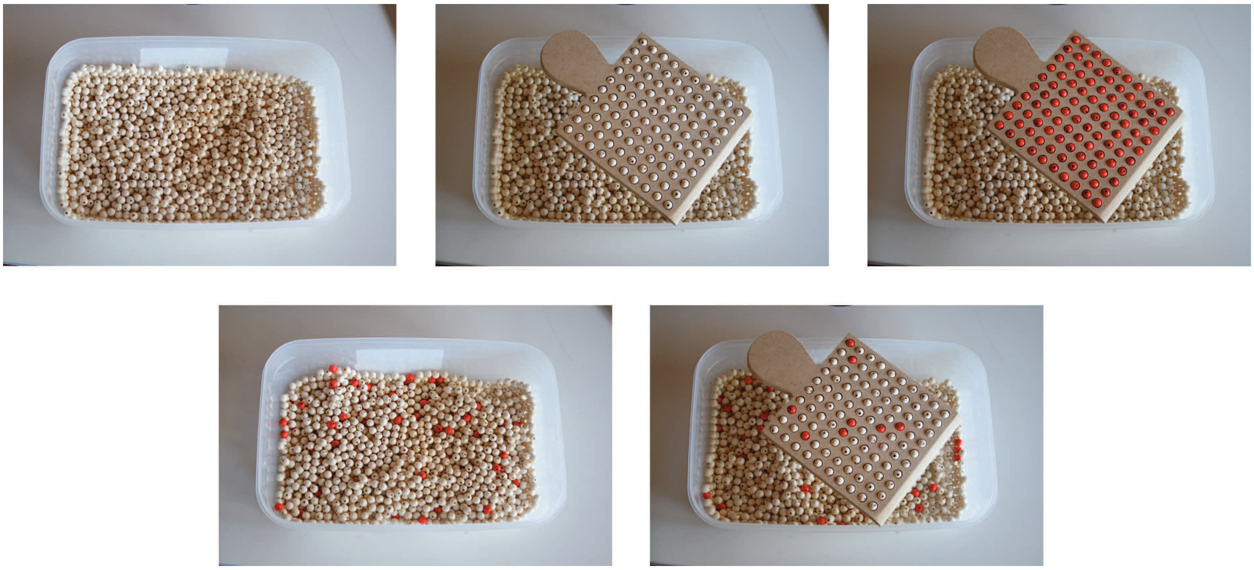


Figura 6. Uso de bolas para ilustrar el método de la pesca y repesca

¿Qué se aprende?: Existen técnicas «ingeniosas» que permiten estimar valores que parecen imposibles de conocer. Es importante conocer la calidad de la estimación realizada y sin grandes conocimientos matemáticos se puede estudiar por simulación.

Otras actividades: ¿Cómo se podría tener una buena estimación de cuantos granos de arroz hay en un kilo si no tuviéramos ningún aparato de medida (no se puede pesar ninguna cantidad)?.

Más información en: Una buena descripción, en tono de divulgación, de esta técnica se encuentra en el artículo: «How Many Fish are in the Pond?» de Roger W. Johnson. En general, todos los artículos de esta página (selección de los mejores artículos de la revista *Teaching Statistics*) contienen buenas ideas para la enseñanza de la estadística a nivel introductorio.

¿Cuántos taxis hay en una ciudad?

Lo habitual es que los taxis lleven a la vista un número de licencia (al menos en las grandes ciudades) y que sea un número correlativo desde el 1 hasta un valor igual al número de licencias que existen. En estos casos en que la población está numerada no es necesario utilizar el método de la pesca y repesca para estimar el número total de elementos.

Por ejemplo, si μ es el valor medio de los elementos numerados siempre ocurre que el número total es igual a $2\mu-1$ y utilizando la media de la muestra $\bar{x}$ como estimador de la media de la población podemos proponer que el estimador sea $2\bar{x}-1$. Si los valores que se han recogido son: 16, 28, 45, 48, 68, 72 y 81, como $\bar{x}=51{,}14$, nuestra estimación podría ser $\bar{x}=2\cdot 51{,}14-1\cong 101$.

Parece un buen método pero tiene un punto débil muy evidente: supongamos que los valores de la muestra son 3, 4, 6 y 15, la media es 7 y por tanto nuestra estimación sería 13, un valor claramente incorrecto (como mínimo hay 15). Una opción que no presenta este problema es considerar que, por simetría, el número de valores que irán a continuación del último será similar al de los que hay antes del primero. Esta estrategia también es razonable y nunca dará resultados evidentemente falsos, pero tiene el inconveniente de que no aprovecha toda la información disponible.

En realidad, la única incógnita es saber cuántos elementos hay a continuación de la última observación. Una buena idea es añadirle el promedio que se calcula con los que hay antes del primero, entre el primero y el segundo, entre el segundo y el tercero, etc. Si n es el número de observaciones y X_i representa el valor de la observación i-ésima, este promedio será:

$$\frac{\left(X_1-1\right)+\left(X_2-X_1-1\right)+\left(X_3-X_2-1\right)+\cdots+}{n}$$

$$\frac{+\cdots+\left(X_n-X_{n-1}-1\right)}{n}=\frac{X_n}{n}-1$$

Y, por tanto, la estimación del número total de elementos queda:

$$\hat{N}=X_n+\frac{X_n}{n}-1$$

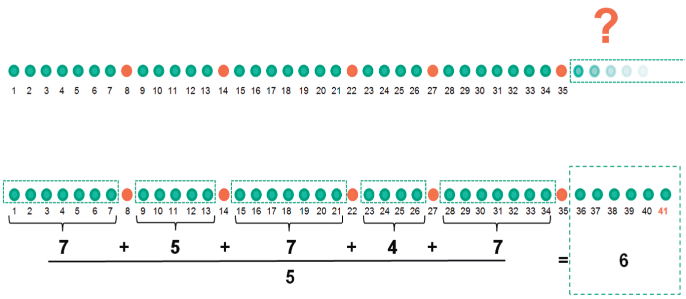


Figura 7. Ejemplo de estimación del número de elementos de una población numerada de la que se tienen 5 observaciones

Este estimador tiene un comportamiento excelente, con todas las propiedades deseables. Y no solo sirve para contar taxis, también puede servir para estimar, por ejemplo, el número de participantes en una carrera popular si los dorsales son números correlativos.

Se puede practicar en clase construyendo pequeñas tarjetas de cartulina numeradas (del orden de 1000 o 1500) introducirlas en un recipiente y sacando unas pocas deducir cuantas hay. Es relativamente fácil imprimir las tarjetas si se utiliza una hoja de Excel y de forma automática se coloca un número en cada celda (hay que cambiar el tamaño, 2×4 cm está bien), a continuación se imprimen y se recortan con una guillotina[5] (figura 8).

¿Qué se aprende?: El método de razonamiento seguido para deducir la expresión del estimador. Que en una situación como ésta es posible realizar una buena estimación con muy poco esfuerzo.

Más información en: La idea está sacada del artículo *Estimating the Size of a Population* de Roger W. Johnson. Puede descargarse de internet, para localizarlo lo más cómodo es poner el título del artículo en Google.

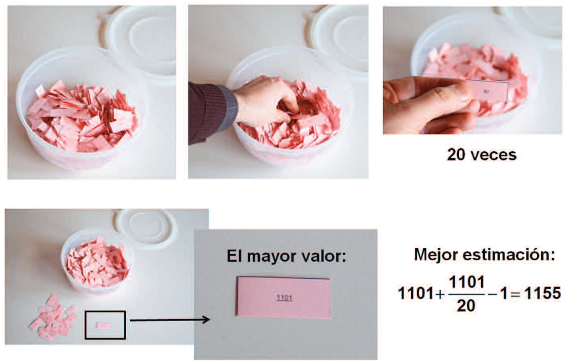


Figura 8. Demostración práctica de cómo estim el número de elementos de una población numerada

Cálculo de probabilidades

El cálculo de probabilidades es un tema que muchas veces se considera difícil y antipático. Los problemas con resultados sorprendentes y el uso de simulaciones puede ayudar a hacerlo interesante.

Coincidencia en las fechas de cumpleaños

En una clase de 30 estudiantes, ¿cuál es la probabilidad de que dos o más celebren su cumpleaños el mismo día? Empezamos calculando la probabilidad de que dos personas no hayan nacido el mismo día. La primera no tiene restricciones, ha podido nacer cualquier día del año (365 casos favorables sobre 365 posibles), pero la segunda ha podido nacer cualquier día menos el que ha nacido el primero (364 días favorables sobre 365 posibles):

$$\frac{365}{365}\cdot\frac{364}{365}=0{,}9973$$

De forma análoga, la probabilidad de que tres hayan nacido en días distintos será:

$$\frac{365}{365}\cdot\frac{364}{365}\cdot\frac{363}{365}=0{,}9918$$

Y la probabilidad de que 30 hayan nacido en días distintos:

$$\frac{365}{365}\cdot\frac{364}{365}\cdot\frac{363}{365}\cdot\ldots\cdot\frac{336}{365}=0{,}2937$$

Solo hay dos posibilidades: todos han nacido en días distintos o al menos dos han nacido el mismo día. Luego la probabilidad de que en un grupo de 30 al menos dos hayan nacido el mismo día será:

$$1-0{,}2937=0{,}7063$$

La probabilidad es mayor de lo que intuitivamente se supone. Con ayuda de una hoja de cálculo se puede construir la curva

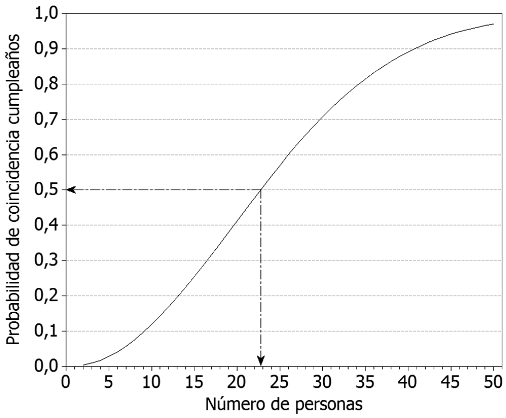


Figura 9. Probabilidad de que dos o más personas de un grupo celebren su cumpleaños el mismo día, en función del número de integrantes del grupo

que relaciona el número de personas y la probabilidad de coincidencia.

¿Qué se aprende?: Las reglas del cálculo de probabilidades permiten deducir probabilidades que evaluadas de forma intuitiva conducen a resultados erróneos. (Moraleja: las reglas para el cálculo de probabilidades son útiles)

Otras actividades: En un grupo de 23 personas la probabilidad de que dos o más celebren su cumpleaños el mismo día ya es ligeramente mayor del 50 % (exactamente del 50,7 %). ¿Dónde hay 23 personas?: en un campo de futbol, 11 + 11 + 1 (el árbitro). En la prensa deportiva es fácil encontrar la alineación de los partidos, y también es fácil encontrar las fechas de nacimiento de los jugadores y del árbitro (en las páginas web de los clubs o incluso en la Wikipedia). Si se elige una jornada de la primera división hay 20 equipos y, por tanto, se juegan 10 partidos casi seguro que en algún partido dos personas de las que estaban en el campo celebraban su cumpleaños el mismo día.

Lotería Primitiva: ¿Cuál es la probabilidad de que una misma combinación ganadora salga dos veces?

Una persona juega toda su vida adulta (pongamos 50 años) a la lotería primitiva, si se realizan dos sorteos por semana, ¿cuál es la probabilidad de que durante ese periodo salga más de una vez la misma combinación ganadora?

En esta lotería el jugador elige 6 números del 1 al 49. Hay 13 983 816 formas de hacerlo[6] de las cuales solo una es la ganadora.

Suponiendo que esta persona juegue 100 veces al año (se realizan dos sorteos semanales) jugará 5 000 veces a lo largo de su vida y resulta que el problema planteado es análogo al del cumpleaños: es como si tuviéramos un año con 13 983 816 días y 5 000 personas cada una de las cuales ha nacido uno de esos días ¿Cuál es la probabilidad de que dos o más hayan nacido el mismo día? Aplicado la fórmula que hemos visto (una hoja de cálculo es imprescindible) resulta que esta probabilidad es del 59 %. Luego no es raro que, si lleva bien las cuentas, descubra que ha tocado dos veces la misma combinación.

¿Qué se aprende?: Como en el caso anterior, que a veces la intuición falla al evaluar probabilidades. También se tiene un ejemplo de problema que, aunque aparentemente es muy distinto al anterior, en esencia es el mismo y se resuelven de la misma forma.

Otras actividades: En España se empezó a realizar este tipo de sorteo en 1985. En

http://www.loteriasyapuestas.es/primitiva/

se pueden consultar todas las combinaciones ganadoras hasta hoy ¿Ha salido alguna vez repetida? Buscarlo es laborioso aunque con una hoja de cálculo se puede hacer de forma eficiente7. La respuesta es sí (figura 10).

Toma de decisiones

La recogida cuidadosa de datos, y su análisis con ayuda del cálculo de probabilidades, permite tomar decisiones más razonables que apelando a intuiciones, sensaciones o corazonadas.

El caso de la catadora de te

Ronald Fisher, uno de los personajes más relevantes de la estadística en el siglo XX, explica en uno de sus libros (Fisher, 1951) que una señora comentó que era capaz de distinguir si en una taza de té con leche se había puesto primero el té o la leche. Se puede plantear una discusión sobre si eso es posible utilizando argumentos del mundo de la física y de la química: la composición del producto resultante, el estado final de las partículas disueltas, el gradiente de temperaturas..., pero Fisher propuso un procedimiento que en aquella época (años 20 del siglo XX) se podía considerar revolucionario: simplemente «hacer la prueba».

De entrada se supone que no sabe distinguir un caso de otro, eso es lo que nos parece normal, y sólo creeremos en su habilidad si los datos recogidos en un experimento bien diseñado y controlado están en contra de esa hipótesis inicial. Estar en contra significa que los resultados sean muy poco probables en el caso de que esa hipótesis sea cierta, y lo que significa poco probable lo decidimos nosotros mismos: que ocurra menos del 5 % de las veces, menos del 1 %, o cualquier otro valor.

Naturalmente no se podía hacer la prueba con solo una taza de cada tipo porque la probabilidad de acertar al azar sería de ½ y si acertaba no se sabría si había sido por casualidad o si realmente sabía distinguir. Si sólo estamos dispuestos a creerla cuando el resultado del experimento pueda ocurrir por azar (por casualidad, sin que sea mérito de la señora) menos del 5 % de las veces, tampoco servirá un experimento dándole a probar dos o tres tazas de cada tipo. Si se le dan tres de cada tipo hay un probabilidad del 5% de que acierte al azar[8]. Si tenemos 4 y 4 de cada tipo el número de maneras de seleccionar 4 es igual a 70 como solo hay un conjunto de cuatro en que se ha puesto primero el té que la leche, la probabilidad de acertar por azar cual es este conjunto

Figura 10. El 22 de agosto de 2002 y el 10 de diciembre de 2009 apareció la misma combinación ganadora en la Lotería Primitiva

es de 1 entre 70, es decir, del 1,4 %. Si de las 4 tazas que elige se equivoca en una, ya no será razonable considerar que sabe distinguir, ya que la probabilidad de que esto ocurra por azar es casi del 23 %.

Pero no se nos tienen que ir todas las energías tras los razonamientos matemáticos. También hay que estar muy atentos a todos los detalles en la realización del experimento, en no dar pistas a la catadora… Fisher lo describe claramente insistiendo en que las tazas se le deben presentar en un orden aleatorio: «Nuestro experimento consistirá en preparar ocho tés con leche, cuatro de un modo y otros cuatro del otro, y llevar las tazas (en un orden escogido al azar) a la catadora para que indique su opinión. Se le habrá explicado antes en qué va a consistir la prueba: se le darán a probar ocho tazas, cuatro de cada tipo, en un orden elegido al azar (por dados, ruleta, cartas etc., o simplemente por números publicados en alguna forma); su tarea consiste en separar las tazas en dos grupos de cuatro clasificándolos, si puede, según se haya puesto primero el té o la leche.»

¿Qué se aprende?: Cómo analizar si somos capaces de distinguir un producto de otro. El tipo de razonamiento que se sigue cuando se realizan tests estadísticos y que se conoce como «contraste de hipótesis».

Otras actividades: En clase se puede organizar una actividad para analizar si se es capaz de distinguir entre dos marcas de refresco.

Más información: En

http://www-eio.upc.es/~grima/CataDeAguas.pdf

puede descargarse una propuesta de actividad para comparación de aguas.

Figura 11: Ejercicio en clase para analizar si se sabe distinguir agua del grifo filtrada de agua embotellada de una marca muy conocida

¿Están los dados bien equilibrados?

Rudolf Wolf (1816-1893) fue un astrónomo suizo que realizó estudios pioneros sobre las manchas solares y que también nos ha dejado los resultados de un curioso experimento que consistió en lanzar 20.000 veces un par de dados (uno blanco y otro rojo), al parecer para estudiar el comportamiento de las leyes estadísticas. Los resultados están en la figura 12.

Como es natural, en ninguno de los dados aparecen los posibles resultados (de 1 a 6) con la misma frecuencia. Es curioso que en ambos dados los resultados 5 y 6 presenten frecuencias por encima del valor teórico (3.333) mientras que 3 y 4 están por debajo. La pregunta que nos planteamos es si esas diferencias permiten afirmar que los dados no estaban equilibrados.

Fijándonos en el dado blanco observamos que la máxima discrepancia entre el valor teórico y el que realmente ha salido es 599 (3932-3333). ¿Es normal esa discrepancia? Las opiniones basadas en impresiones o intuiciones tienen poco valor. Una manera fundamentada de responder a esta pregunta es simulando[9] y para hacerlo basta una hoja de cálculo (Figura 13).

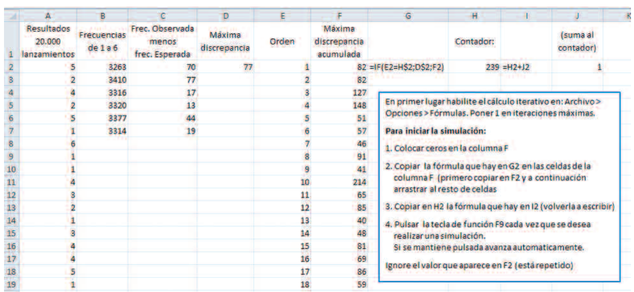
En la columna A se simulan los resultados de los 20.000 lanzamientos [=aleatorio.entre(1,6)], en la B se colocan las frecuencias de aparición de cada resultado (para 1: =contar.si(A$2:A$20001; 1), en C colocamos las diferencias entre la frecuencia observada y la esperada y en D la máxima discrepancia. Basta pulsar la tecla de función F9 para obtener una nueva simulación (nuevos números aleatorios y nuevo valor de la máxima discrepancia). También se puede disponer un sistema que vaya acumulando los valores de la máxima discrepancia, en la hoja de la figura 13 se acumulan en la columna F, y ya solo falta comparar nuestra máxima discrepancia con la distribución que presenta cuando los dados están equilibrados (a esta distribución le llamamos «distribución de referencia»). Claramente los valores que corresponden a los dados de Wolf no pertenecen a esa distribución. Son valores mucho mayores que los obtenidos con los datos equilibrados. Podemos decir con una probabilidad de equivocarnos prácticamente nula que los dados no estaban bien equilibrados.

¿Qué se aprende?: A razonar siguiendo el esquema habitual de los tests estadísticos: se compara el valor

The results of 20,000 throws with two dice (data from Czuber, 1903)

		White die 1	2	3	4	5	6	Total	Proportion
Red die	1	547	587	500	462	621	690	3407	·170
	2	609	655	497	535	651	684	3631	·182
	3	514	540	468	438	587	629	3176	·159
	4	462	507	414	413	509	611	2916	·146
	5	551	562	499	506	658	672	3448	·172
	6	563	598	519	487	609	646	3422	·171
Total		3246	3449	2897	2841	3635	3932	20,000	1·000
Proportion		·162	·172	·145	·142	·182	·197	1·000	

Figura 12. Resultados obtenidos por R. Wolf al lanzar dos dados 20 000 veces (tomado de M. G. Bulner, 1979)



	A	B	C	D	E	F	G	H	I	J
1	Resultados 20.000 lanzamientos	Frecuencias de 1 a 6	Frec. Observada menos frec. Esperada	Máxima discrepancia	Orden	Máxima discrepancia acumulada		Contador:		(suma al contador)
2	5	3263	70	77	1	82	=IF(E2=H$2;D$2;F2)	239	=H2+J2	1
3	2	3410	77		2	82				
4	4	3316	17		3	127				
5	2	3320	13		4	148				
6	5	3377	44		5	51				
7	1	3314	19		6	57				
8	6				7	46				
9	1				8	91				
10	1				9	41				
11	4				10	214				
12	3				11	65				
13	2				12	85				
14	1				13	40				
15	3				14	48				
16	4				15	81				
17	4				16	69				
18	5				17	86				
19	1				18	59				

En primer lugar habilite el cálculo iterativo en: Archivo > Opciones > Fórmulas. Poner 1 en iteraciones máximas.

Para iniciar la simulación:

1. Colocar ceros en la columna F
2. Copiar la fórmula que hay en G2 en las celdas de la columna F (primero copiar en F2 y a continuación arrastrar al resto de celdas
3. Copiar en H2 la fórmula que hay en I2 (volverla a escribir)
4. Pulsar la tecla de función F9 cada vez que se desea realizar una simulación. Si se mantiene pulsada avanza automaticamente.

Ignore el valor que aparece en F2 (está repetido)

Figura 13. Hoja de cálculo para simular 20 000 lanzamientos de un dado y anotar la máxima discrepancia entre las frecuencias observadas y la esperada obtenido con la distribución a que debería pertenecer si la hipótesis realizada (en este caso que los dados están equilibrados) fuera cierta. La conclusión está muy clara para los dos dados aunque no es evidente a primera vista.

Más información: La hoja de cálculo para realizar la simulación puede descargarse de:

http://www-eio.upc.es/~grima/SimulacionDadosWolf.xls

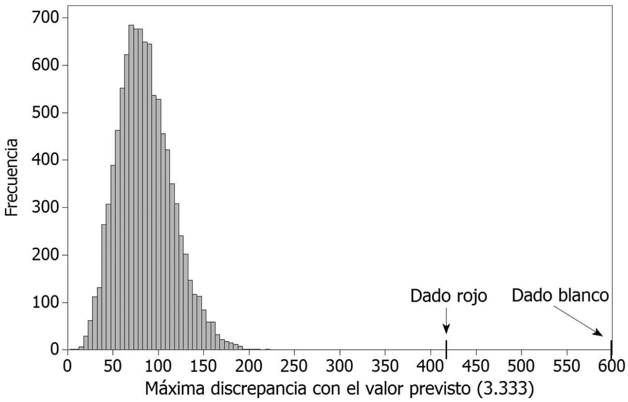


Figura 14: Comparación de la máxima discrepancia obtenida en los dados de Wolf y su distribución de referencia si los dados están equilibrados

Nota final: Concurso de sondeos y experimentos

Existe la posibilidad de presentar los trabajos de estadística realizados por estudiantes de ESO, Bachillerato y ciclos formativos a los certámenes denominados «Incubadoras de sondeos y experimentos» que tras una iniciativa de la Universidad Complutense de Madrid se organizan en diversas comunidades autónomas. A la edición catalana de este año se han presentado 166 trabajos con 598 alumnos participantes, muchos de ellos muy interesantes. También hay una final nacional convocada por la Sociedad de Estadística e Investigación Operativa (SEIO), que este año tuvo lugar en Santiago de Compostela a primeros de julio (más información en la página web de la SEIO).

Referencias bibliográficas

Bulner, M. G. (1979), *Principles of Statistics*, Dover Publications, Nueva York.

Fisher, R. A. (1951), *The Design of Experiments*, Oliver and Boyd, Londres.

Johnson, R. W. (1996), «How Many Fish are in the Pond?», *Teaching Statistics,* vol.18, n.º 1, 2-5.

— (1994), «Estimating the Size of a Population», *Teaching Statistics,* vol. 16, n.º 2, 50-52.

Krebs, Ch. J. (1999), *Ecological Methodology*, Addison-Wesley, USA.

Lewin, W. (2013), *Por amor a la física*, Debolsillo, Barcelona.

Tufte, E.R. (1997), *Visual Explanations*, Editorial Graphics Press, USA.

Piñol, J. y J. Martinez-Vilalta (2006), *Ecología con números*, Lynx, Barcelona.

Pere Grima Cintas
Universitat Politècnica de Catalunya — BarcelonaTech
<pere.grima@upc.edu>

1 La estadística es una de las asignaturas que aparece con más frecuencia en los planes de estudio de titulaciones universitarias. Incluso carreras que no suelen incluir las matemáticas en su currículum, como medicina, sociología, ciencias políticas o psicología, sí incluyen la estadística y tienen en esta materia una de sus herramientas de trabajo más importantes.

2 Existen métodos de estimación más precisos, incluso los hay que tienen en cuenta la posible tasa de mortalidad debida a las marcas o la pérdida de ellas. Ver, por ejemplo, Kreps (1999) o Piñol y Martinez-Vilalta (2006).

3 http://www.ecologiaconnumeros.uab.es/Llibre/AppletsLlibre/3-MuestreoPoblaciones/Applet_3_1/MarcajeRecaptura.html

4 Detalles sobre los materiales en <http://www-eio.upc.es/~grima/CuantosPeces.pdf>.

5 Más información en <http://www-eio.upc.es/~grima/CuantosTaxis.pdf>.

6 Para el primer número existen 49 valores posibles, 48 para el segundo… y para el sexto 44. Por tanto existen $49 \cdot 48 \cdot \ldots \cdot 44$ resultados distintos al extraer los 6 números. Pero entre estos resultados muchos están repetidos porque solo se diferencian en el orden de aparición. El número de ordenaciones posibles para 6 valores es 6! por tanto el número de resultados posibles para la combinación ganadora es: $49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44/6!$

7 Puede descargarse de <http://www-eio.upc.es/~grima/ResultadosPrimitiva.xls>.

8 Hay 20 formas de elegir tres objetos entre seis ($6 \cdot 5 \cdot 4/3!$, razonamiento similar al ejemplo de la lotería primitiva) y solo una es la forma correcta de hacerlo

9 En realidad no sería necesario simular. Existen tests estadísticos para analizar la igualdad de varias proporciones.