

## *Big data*

Vamos a comparar dos situaciones habituales en el contexto de la compra de un producto. Por modernidad, elegimos comprar un móvil de última generación.

Si nos dirigimos a una tienda física tendremos que mirar en los expositores qué móviles tienen a la venta en esa tienda, pudiendo consultar algunas de sus características (las que quepan en un pequeño letrero) y su precio. Después de ojear y tomar una primera decisión procederemos a esperar nuestro turno para que nos atiendan. El consejo de un profesional puede ayudarnos a acabar de decidirnos por ese modelo que no podemos pagar pero que necesitamos o ese otro más modesto, menos llamativo pero con un precio más comedido. Cuando todo está decidido pagamos y nos vamos con nuestro móvil nuevo.

Ahora centremos nuestra mirada en el equivalente de ese proceso de compra en una tienda on-line. Después de elegir la tienda, proceso en el que tenemos una gran variedad, podemos introducir algunas palabras clave en el buscador y la web nos ofrecerá una lista de opciones. Nos fijamos en algunas de las que nos parezcan más interesantes y procedemos a consultar las características que nos resultan más adecuadas. Si dudamos siempre podemos buscar una comparativa en blogs o webs especializadas o algún vídeo en la red. Cuando hemos decidido, pagamos y esperamos a que nos llegue el móvil.

Ahora cambiemos el punto de vista para fijarnos en la perspectiva del vendedor. En la tienda física todo está organizado para ayudarnos en el proceso de compra, pero tienen limitaciones para obtener información sobre nuestros gustos, hábitos, nivel adquisitivo, etc. Y es difícil que el vendedor se pueda fijar en los móviles que hemos estado mirando o si necesitamos algún complemento concreto. Tampoco es sencillo que tenga información sobre nuestras compras anteriores, de móviles o de otros productos.

En cambio, si pensamos en el proceso de compra en una web observamos que la cantidad de información que se genera y puede ser recogida (y posteriormente analizada) es descomunal. Para empezar podrán listar los móviles que hemos ido consultando hasta elegir, la hora a la que lo hemos hecho y el tiempo de consulta con cada uno de ellos, si hemos consultado fuentes externas (para eso sirven las maravillosas cookies que nos regalan las webs) o las compras anteriores. También será fácil que nos ofrezcan productos complementarios. De hecho, lo que ya pasa es que nos ofrecen los productos complementarios que otros usuarios han comprado y nos listan los móviles en función de su relevancia en esos momentos.

De esta forma parece claro que los servicios on-line tienen una ventaja substancial aparte de no necesitar un local situado en cada ciudad o personal que atienda al público: nos pueden monitorizar.

Ventaja o desventaja porque tal vez con este tipo de servicios nos estamos cargando el comercio de proximidad y acabamos montando una cena con unos aguacates que vienen de México, unos tomates de Ecuador, una merluza de Sudafrica y unos kiwis de Nueva Zelanda. En cualquier caso, bueno es saberlo y estar informado para tomar decisiones sobre qué compras y porqué.

También pueden monitorizarse todos los movimientos que hace un conductor en un viaje en coche para determinar el momento en el que el cansancio hace mella en sus capacidades o todas las transacciones económicas que puedan resultar sospechosas de comportar actividad terrorista.

En este contexto, lo que sigue es una lista de palabras no escogidas al azar: Minería de datos, Ciencia de Datos, Datos masivos. O sus equivalentes en inglés que son más conocidos: Data mining, Data Science, Big Data. Parece claro que la Estadística es un campo en auge, por sus implicaciones comerciales pero también por sus usos en investigaciones de todo tipo. También parece que va a tener un mayor peso en nuestras vidas en un futuro no muy lejano aunque no seamos conscientes de ello. Y parece que estas nuevas disciplinas se encuentran en un momento interesante, porque no están bien definidos todos los métodos a utilizar en el futuro. Por su naturaleza matemática y la potencialidad de poder organizar todo tipo de información o datos, la Estadística se perfila como un conocimiento clave para las próximas generaciones.

Estamos super controlados, es como si actuáramos dentro de un programa del *Gran Hermano*. Ya lo cantaba *La Pegatina* hace unos cuantos años.

Sales de tu casa, y te pones a pasear.  
Hoy en día ni en la calle, ni en la calle se puede estar.  
Cámaras por todos lados, no paran de grabar.  
Porque el sistema te quiere, te quiere controlar.  
Gran hermano, gran hermano. Entra en acción.  
Dice que tiene la solución. Aprieta el botón.  
Gran hermano, gran hermano. Tiene el control.  
Hoy, se acabó la libertad,  
Ya nadie tiene, nadie tiene intimidad.  
Si las cámaras me vigilan yo ya sé que debo hacer,  
taparles el objetivo y que no me puedan ver.

Plantéate actuar, tu vida como actor acaba de empezar.  
Dejó de ser lo mismo ir al *mercao*,  
yo me siento mejor si me han *grabao*.  
Se inventan los motivos, no dan información.  
Te controlan por la calle, ¡luces!, ¡cámara!, ¡acción!  
Ponte guapo, que sales en la foto. Ponte guapo, que te miran de reojo  
Sonríe y baila...

En un plan un poco más serio, parece pues que debemos conseguir que nuestros alumnos tengan criterio y no se dejen ahogar o apabullar por el exceso de datos que nos rodea. Bueno será también que sepan manejarse en la red sin dejar más datos e información que la estrictamente imprescindible.

En este sentido, destacamos el punto de vista de Carmen Batanero, que ha presentado diversas de sus investigaciones en *Suma*, y que define y defiende el concepto de *Sentido Estadístico* como la unión de la cultura estadística y el razonamiento estadístico y lo añade a la idea de alfabetización matemática que entendemos que deben alcanzar nuestros alumnos para ser ciudadanos. Y es que no se trata exclusivamente de hacer medias y medianas mientras se leen gráficos de barras, el objetivo es que los alumnos creen una red de conocimientos estadísticos que les permitan entender su entorno a partir de elementos más elaborados como entender una distribución de datos para llegar a conclusiones sobre una población.

Pero para ello es necesario cambiar algunos clichés que tienen (tenemos) algunos profesores de matemáticas. Existe una percepción general de la Estadística como una parte menor de las matemáticas, posiblemente por su orientación práctica y alejada del núcleo de conocimientos matemáticos que siempre colocamos en primer plano (como la aritmética, el álgebra, la geometría...) y porque existen algunos estigmas sociales al respecto. En efecto, muchas veces hemos escuchado que con un estudio estadístico y los datos adecuados se puede defender cualquier afirmación o hecho. La lástima es que nadie responde que ese es un problema relacionado con las intenciones de quien dirige el estudio y que la Estadística, dentro de sus limitaciones, es una herramienta muy potente para poner orden donde solo vemos confusión.

En el entorno escolar tampoco ayuda que los temas de estadística, como decían los compañeros de Ademgi (Girona), en las JAEM de Palma (2013) siempre está en capítulo 13 y raramente se hace, no se sabe si porque está al final del libro o por llevar un número que tradicionalmente lleva mala suerte.

Lo anterior viene a cuento de que en este número de *Suma* encontraréis las conclusiones del Seminario Federal *Estadística y probabilidad en la Educación Matemática*, que se celebró en Castro Urdiales en noviembre de 2014 donde encontraréis una buena lista de ideas para reenfocar las clases de Estadística y Probabilidad y colocarlas en el lugar en el que necesitamos que estén. De lo contrario nos podemos encontrar con que otros sepan interpretar datos mejor que nosotros (como individuos, como sociedad, como país), y eso, en el momento histórico en el que se genera un mayor volumen de datos de la historia, no es la mejor de las perspectivas.