

Macro versus micro: La población desde el individuo, pasando por los datos.

[Reseña con indicaciones para el profesorado]

En cualquiera de los cursos que conforman las opciones de Bachillerato, el alumnado ya posee las suficientes nociones de Estadística como para abordar esta Unidad Didáctica de ampliación. Partiendo de sus conocimientos sobre estadística descriptiva, se propone un análisis de datos obtenidos de forma real y, posteriormente, se propone el análisis crítico de diferentes situaciones en las que la Estadística juega un papel relevante.

Además de la lectura del documento principal por parte del alumnado, las actividades propuestas para trabajar durante dos sesiones son:

- **Visita alguna web con estudios estadísticos** como por ejemplo la [web del INE](#) (Instituto Nacional de Estadística) o la [web del CIS](#) (Centro de Investigaciones Sociológicas), elige alguno de sus estudios y trata de determinar sus elementos más importantes a partir del **análisis de la ficha técnica**.

Aquí debe ahondarse en la fiabilidad del estudio en función del tamaño muestral así como en los diferentes métodos de selección de una muestra, localizando esta información en la ficha técnica de cada estudio. Además, las herramientas online que facilitan tanto el INE como el CIS permiten realizar la visualización de los datos en diferentes gráficos que pueden analizarse detenidamente poniendo de relevancia los elementos que conocen los alumnos. En caso de conocer los rudimentos de las hojas de cálculo, puede hacerse uso de ellas también para calcular los correspondientes parámetros estadísticos.

- **Busca información sobre la Ley de los Grandes Números**, trata de comprender su enunciado y explícalo con tus propias palabras. Ídem para el **Teorema del Límite Central**.

El enunciado debe traducirse a terminología comprensible por los alumnos, resaltando que es un enunciado de tipo probabilístico. Simplificando, lo que asegura la Ley de los Grandes Números es que cuanto mayor es la muestra, más se parecerán los datos recabados a los reales de la población. En particular, la media muestral representará con precisión razonable a la media poblacional y las frecuencias relativas de la muestra se asemejarán a la distribución de probabilidad de la población. En el caso del Teorema del Límite Central, la repetición de variables con media y desviación conocidos tiene a distribuirse de forma normal. El paso de la distribución Binomial a la distribución Normal puede verse cualitativamente a través del experimento de Galton o si los alumnos conocen el comportamiento de los parámetros estadísticos, hablando también de éstos.

- **Realiza un análisis comparativo** entre los resultados esperados y los que efectivamente se dieron en las elecciones.

Puede utilizarse tanto como una introducción a los intervalos de confianza o como actividad de cierre de ese mismo bloque. La disonancia, se debe al margen de error que tiene todo estudio y que se especifica en la ficha técnica a través del p-valor y el nivel de confianza. Obviamente, también influyen cuestiones de tipo sociológico, como que los encuestados no sean totalmente transparentes en sus respuestas o la componente subjetiva de quien interpreta los datos.

Para la tercera sesión se propone el análisis crítico de gráficos estadísticos:

- **Analiza los dos ejemplos siguientes** y discute, argumentando tu postura, si transmiten una información visual adecuada.

En multitud de ocasiones se transmite información sesgada, ya sea intencionadamente o por error. En los gráficos propuestos no se respeta la proporción y/o no se plantea un eje referencia. En no pocas ocasiones el hecho de romper el eje (y que no comience en 0) acentúa subjetivamente la diferencia entre dos barras que, objetivamente, son similares. Además de analizar los gráficos, se puede pedir que los alumnos hagan una representación correcta de la situación, tanto con los ejemplos sugeridos o cualesquiera otros, como los que aparecen en [esta comunicación](#) en el XVI CEAM, [este post](#) del blog Xataka o en este [artículo periodístico](#) de elDiario.es que a su vez contiene redirecciones a otros enlaces de El País o El Mundo.

- **Observa el siguiente plano** de los alrededores de Broad Street. ¿Qué conclusiones puedes extraer a partir de la representación? ¿Cómo crees que finalmente se resolvió el brote? Tras realizar algunas conjeturas, busca la información del caso real.

Las defunciones se aglutinan principalmente alrededor de la boca de agua de Broad Street. Snow sugirió cerrar el suministro de ésta, y con ello se consiguió controlar el brote.

La cuarta y última sesión se sugiere dedicarla a la toma de conciencia sobre la multitud de información que hay tras nuestras sesiones de navegación y el uso que de ella se hace:

- **¿Podrías buscar algunos ejemplos de uso de Big Data?** ¿Son Machine Learning?

Ejemplos de Big Data pueden ser los análisis de clustering/agrupamientos para realizar campañas de publicidad dirigida a un sector concreto de la población, estudio de grafos y sus (súper) nodos para determinar influencias en redes sociales, el diseño de entrenamientos personalizados en función del rendimiento de un deportista para la prevención de lesiones...

Ejemplos de Machine Learning son los algoritmos que archivan automáticamente correo electrónico en carpetas de spam, la interacción con asistentes virtuales, el reconocimiento facial en fotografías, las sugerencias de contenidos/productos/amistades en plataformas tipo Netflix, Amazon y redes sociales...

Cabe la posibilidad de seleccionar algunos de los ejemplos y estudiar más en profundidad su modelo de funcionamiento.

- **Debate sobre los límites legales y/o éticos** que deberían existir en torno al desarrollo de algoritmos Big Data.

Además de la hiperconexión y la obsesión constante por mantenerse actualizado – tanto en el sentido tecnológico como en el de los contenidos – para evitar quedar desfasados, otra de las ideas centrales del debate debería ser si realmente se mejora la experiencia del usuario o se manipulan sus pulsiones. Aquí puede consultarse el Libro Blanco de la Unión Europea en materia de [Inteligencia Artificial](#) del año 2020, así como una comunicación con respecto al [tratamiento global de datos](#).