



¿QUÉ APORTAN LOS ORDENADORES A LA ENSEÑANZA Y APRENDIZAJE DE LA ESTADÍSTICA?

Juan D. Godino

1.-INTRODUCCIÓN

En la actualidad se reconoce la importancia de la enseñanza de la estadística en los diferentes centros educativos a los estudiantes sobre ideas probabilísticas y estadísticas y su mutua interdependencia.

La interpretación, representación y tratamiento de la información, tratamiento del azar en la formación profesional y en las universidades los cursos sobre estadística aplicada se incluyen prácticamente en todas las especialidades.

Este fenómeno no es exclusivo de los países desarrollados. La enseñanza de los contenidos referidos a la estadística y probabilidades se incrementa en los nuevos planes de estudio de diferentes países. Un indicador significativo del interés por la formación estadística a todos los niveles es el hecho de que el último congreso internacional sobre enseñanza de la estadística, hubo un grupo de trabajo especial dedicado a “El currículo estadístico”

Este interés se explica por la importancia que la estadística ha alcanzado en nuestros días, tanto como cultura básica, como en el trabajo profesional y en la investigación, debido a la abundancia de información a la que el ciudadano, el técnico y el científico deben enfrentarse en su trabajo diario. El rápido desarrollo de la estadística y su difusión en los últimos años se ha debido a la influencia de los ordenadores, que también han contribuido a la acelerada cuantificación de nuestra sociedad y al modo en que los datos son recogidos y procesados.

Hasta hace pocos años, el análisis de datos reales estaba reservado a estadísticos profesionales, quienes debían escribir sus propios programas de ordenador para realizar los cálculos. Posteriormente, el uso de los paquetes potentes de análisis de datos requería el conocimiento de los comandos y sintaxis de los mismos. Esta situación, aparentemente, ha sido superada

Por un lado aparecen los entornos operativos “amistosos”, que permiten acceder directamente al manejo de cualquier de los módulos de un paquete estadístico y, con la ayuda del “ratón”, explorar sus posibilidades. Por otro lado, existen programas “de consulta” a los cuales se puede recurrir para obtener un “consejo” sobre el método de análisis que se debe aplicar en función del tipo de datos y nuestra hipótesis sobre los mismos. ¿Quiere esto decir que hemos resuelto definitivamente el problema de la estadística? ¿se debe reducir esta enseñanza a los alumnos el usos de este tipos de programas informáticos? Si no es así, ¿Cómo debemos reconsiderar los contenidos, objetivos y metodología de aprendizaje en función de las nuevas tecnologías?

En este trabajo discutimos estas cuestiones, aportando nuestra respuesta sobre las mismas: por un lado, la capacitación estadística incluye hoy en día el conocimiento del modo de procesar datos mediante un programa estadístico, por los que deberíamos, en la medida de los posible, ofrecer a nuestros alumnos un primer contacto con este tipo de programas.

Por otro, el ordenador no es sólo un recurso de cálculo, sino un potente útil didáctico, que nos permite conseguir una aproximación más exploratoria y significativa en la enseñanza de la estadística

2.- ORDENADORES Y OBJETIVOS EDUCATIVOS

Como hemos razonado, la creciente disponibilidad de programas de ordenador para el análisis de datos nos obliga a una reflexión sobre sus implicaciones en la enseñanza de esta materia.



En primer lugar, el ordenador puede y debe usarse en la enseñanza como instrumento de cálculo y representación gráfica para analizar datos recogidos por el alumno o proporcionados por el profesor. Nos enfrentamos a diario a la necesidad de recoger, organizar e interpretar sistemas complejos de datos y esta necesidad aumentará en el futuro, debido al desarrollo de los sistemas de comunicación y las bases de datos. Uno de los objetivos que debiera incluirse en un curso de estadística es capacitar al alumno para recoger, organizar, depurar, almacenar, representar y analizar sistemas de datos de complejidad accesibles para él. Este objetivo comienza por la comprensión de la idea básica de sistema de datos.

Este término es más adecuado que el conjunto de datos, para describir las estructuras de datos en las aplicaciones reales. Un conjunto no tiene por qué ser ordenado, mientras que un sistema de datos ha de organizarse para poder ser procesado. Organizamos un sistema de datos al identificar en él la unidad de análisis, las variables y las categorías de las mismas. Un conjunto no tiene elementos repetidos, mientras que una de las características de las variables de un sistema de datos es que cada uno de sus valores se dará con una cierta frecuencia. No tendría ningún interés estadístico un sistema de datos en que todos sus elementos fuesen diferentes. Es precisamente las regularidades globales, dentro de la variabilidad individual el objeto de estudio de la estadística. En la mayor parte de los sistemas de datos hay al menos tres componentes: la descripción de las variables, los valores de la variable, que es el cuerpo principal de los datos, y los resúmenes estadísticos de cada variable. Los campos pueden ser de longitud fija o variables, y puede haber campos vacíos. Asimismo, clasificamos las variables según diversas tipologías cualitativas o cuantitativas; discretas, continuas; nominales, datos de intervalo, de razón.

Sobre cada una de estas componentes pueden realizarse operaciones o transformaciones internas (clasificación, remodificación, agrupamiento) y externas (insertar, borrar, seleccionar...). Podemos clasificar variables, clasificar los casos dentro de una variable o clasificar los resúmenes estadísticos, por ejemplo, por su magnitud, podemos seleccionar casos por los valores de una variable, o seleccionar variables porque sus valores coinciden en una serie de casos. También es posible determinar relaciones entre estos componentes, por ejemplo, de dependencia, implicación, similitud (dependencia entre variables; similitud de sujetos; similitud de variables,...). Estos tipos de operaciones deben ser presentados para casos sencillos a los estudiantes, de modo que sean comprendidas. Aunque parezca muy simple, nuestra experiencia personal en el trabajo de análisis de datos no ha mostrado que la principal dificultad de muchos investigadores es precisamente el definir de una forma adecuada sus unidades de análisis y variables.

El punto de comienzo de la estadística debería ser el encuentro de los alumnos con sistemas de datos reales: resultados deportivos de sus equipos favoritos, precios de las golosinas que compran en el recreo, medios de transporte usados para ir a sus estudios, temperaturas máximas y mínimas a lo largo de un mes; color o tipo de vehículo que pasa por delante de la ventana etc.

De este modo podrán ver que construir un sistema de datos propio y analizarlo no es lo mismo que resolver un problema de cálculo rutinario tomado de un libro de texto. Si quieren que el sistema de datos sea real, tendrán que buscar información cuando les falte, comprobar y depurar los errores que cometen al recoger los datos, añadir nueva información a la base de datos cuando se tenga disponible, aprenderán a comprender y apreciar más el trabajo de los que realizan las estadísticas para el gobierno y los medios de comunicación. Si comprenden la importancia de la información fiable, se mostrarán más dispuestos a colaborar cuando se les solicite colaboración en encuestas y censos.



Estos sistemas de datos pueden ser la base de trabajo interdisciplinarios en geografía, ciencias sociales, historia, deportes, etc.

En el caso de que los datos se tomen de los resultados de experimentos aleatorios realizados en la clase, estaremos integrando el estudio de la estadística y probabilidad. Una vez construido un sistema de datos el siguiente paso sería analizar con ayuda del ordenador. El manejo de un paquete es un objetivo importante ya que, en la actualidad, el uso de las técnicas estadísticas está ligado a los ordenadores. Un problema tradicional en la enseñanza de la estadística ha sido la existencia de un desfase entre la comprensión de los conceptos y los medios técnicos de cálculo para poder aplicarlos. La solución de los problemas dependía en gran medida de la habilidad de cálculo de los usuarios, que con frecuencia no tenían una formación específica en matemáticas. Hoy en día la existencia de programas fácilmente manejables permite salvar este desfase. Esta mayor facilidad actual de empleo de procedimientos estadísticos, implica, sin embargo, el peligro de los usos no adecuados de la estadística. En el trabajo de consultoría estadística no es difícil encontrar a investigadores que, habiendo recogido un conjunto de datos sin ningún tipo de consulta con un estadístico profesional en la etapa del diseño de la investigación, piensan que el análisis consiste simplemente en la elección de un programa adecuado que automáticamente dará una interpretación a sus investigaciones.

Acostumbremos, pues, a los alumnos a planificar el análisis que quieren realizar incluso antes de finalizar la construcción de su sistema de datos. Si por ejemplo, quieren hacer un estudio en su instituto para comparar la intención de voto de jóvenes y señoritas en las próximas elecciones al centro de estudiantes, deben recoger una muestra suficientemente representativa de ambos sexos en los diferentes cursos y deben recoger datos sobre las principales variables que influyen en esta intención de voto.

De otro modo, sus conclusiones pudieran estar sesgadas o ser poco explicativas. Debemos también hacer conscientes a los alumnos de que un mismo problema estadístico puede ser resuelto por diferentes procedimientos y las respuestas que se obtienen pueden ser complementarias y a veces poco adecuadas. No todos los procedimientos estadísticos se adaptan bien para todos los problemas. Por ejemplo, la media aritmética no sería un representante adecuado de un conjunto de datos bimodal o con valores atípicos muy acusados.

Finalmente está el problema de la interpretación de los resultados y la generación de hipótesis sobre el problema investigado a partir de los resultados de los análisis. Esta cuestión se incluye dentro de la filosofía del análisis exploratorio de datos.

3.-ANÁLISIS EXPLORATORIO DE DATOS

Los ordenadores actuales, con sus posibilidades interactivas, favorecen la introducción, desde los primeros niveles de enseñanza, de una nueva filosofía en los estudios estadísticos introducida por Tukey (1977): el análisis exploratorio de datos. Esta filosofía no solo se aplica a nivel de estadística elemental. Por el contrario, muchos de los métodos de análisis de datos multivariados son empleados en la actualidad dentro de esta misma filosofía, para analizar fenómenos físicos o sociales complejos.

El análisis de datos tradicional tenía como función principal la confirmación de hipótesis, que se establecían de antemano a la recogida de datos. Estaba basado fundamentalmente en el cálculo de estadísticos en muestras recogidas con el fin de poner las hipótesis a prueba y su comparación con los valores supuestos para los parámetros de la población.



La representación gráfica de los datos no era muy importante, porque la mayor parte de las veces se suponía que los datos provenían de una distribución normal o aproximadamente normal.

En el análisis exploratorio de datos se concede una importancia similar a los dos componentes que componen los datos estadísticos: la “regularidad” o “tendencia” y las “desviaciones” o “variabilidad”. La regularidad es la estructura simplificada del conjunto de observaciones: la media o mediana en su distribución, la línea de regresión en una nube de puntos, Las diferencias de los datos con respecto a esta estructura (diferencia respecto a la media, respecto a la línea de regresión, etc., son las desviaciones o residuos de los datos.

En la inferencia clásica se supone que estas desviaciones siguen un patrón aleatorio. De acuerdo con la teoría de errores, la distribución de estos residuos sería normal con media cero.

El estudio se centraba en buscar un modelo, dentro de una colección dada, para expresar la regularidad de las observaciones. Por ejemplo, en un estudio de regresión lineal se trata de elegir la recta que representa lo mejor posible la nube de puntos. Asimismo, se definen unos ciertos coeficientes cuyo fin es probar la “bondad” de ajuste del modelo mediante un contraste de hipótesis, en este caso, sobre el coeficiente de correlación

Por el contrario, en el análisis exploratorio de datos se concede una importancia similar a los dos componentes de los datos que hemos citado. En lugar de imponer un modelo a las observaciones, se genera dicho modelo desde las mismas.

Por ejemplo, cuando se estudian las relaciones entre dos variables, el investigador no solamente necesita ajustar los puntos a una línea, sino que estudia los estadísticos, compara la línea con los residuos, estudia la significación estadística de la razón de correlación u otros parámetros para descubrir si la relación entre las variables se debe o no al azar. Si se piensa que es posible extraer nueva información de los residuos, se reanalizan éstos, tratando de relacionarlos con otras variables

Un punto importante en el análisis exploratorio de datos es que no se trata de un conjunto de métodos aunque se han creado algunas técnicas, especialmente gráficas, asociadas a ella -sino de una filosofía de aplicación de la estadística. Esto lo podemos ver en el ejemplo anterior en el que una misma técnica la regresión lineal la podemos aplicar con enfoque exploratorio o confirmatorio. Esta filosofía consiste en el estudio de los datos desde todas las perspectivas y con todas las nuevas, en el sentido de conjeturar sobre las observaciones de las que disponemos. Como contrapartida, tales “hipótesis” no quedan contrastadas en el sentido estadístico del término al finalizar el análisis, por lo que sería preciso la toma de nuevos datos (una replicación) sobre fenómenos y efectuar sobre ellos un análisis estadísticos confirmatorio con el fin de contrastarlas.

En la educación secundaria, la enseñanza de la estadística descriptiva debe hacerse desde la perspectiva del análisis exploratorio de datos. A título de ejemplo, describimos, a continuación, una situación problemática que podría proponerse a los alumnos para que tengan la oportunidad de aplicar esta filosofía y aprender de las técnicas gráficas asociadas al análisis exploratorio de datos.

Ejemplo ¿Cuáles son las principales diferencias entre los chicos y chicas en la clase de educación física?

Un proyecto fácil de plantear a los alumnos en clase consiste en proponerles la investigación de la influencia del sexo en la clase de educación física. Una tarea similar puede proponerse eligiendo otro tema, como los rasgos físicos, las preferencias



sobre empleo del tiempo libre, el tipo de música a lectura que prefieren, etc. En el ejemplo podemos recoger datos sobre rendimiento en pruebas de salto de altura y longitud, flexibilidad, número de abdominales por minuto, tiempo en recorrer 100 metros y 500 metros, etc. Estas marcas podrían recogerse a principios y finales de un año, con el objeto de comprobar cuál de los dos sexos han mejorado más sus marcas iniciales. También podríamos recoger datos de algunas variables físicas como el peso y la altura. Si los alumnos tienen acceso a los usos de un paquete de programas para el análisis de datos el profesor puede proponer la grabación de un fichero con estas variables y la realización de los análisis estadísticos necesarios para dar respuesta a las siguientes interrogantes.

- a) ¿Existe diferencias entre hombres y mujeres en las características estadísticas de alguna de estas variables? ¿en cuál es esta diferencia más o menos acusada?
- b) ¿Cuáles de estas características son independientes del sexo y en cuáles existe asociación estadística con dicha variable? ¿Cuál es la intensidad de la asociación?

Una respuesta concluyente a estas preguntas no es posible darla en el marco de una actividad estudiantil en cursos de iniciación a la estadística, ya que precisaría la recogida de datos representativos de la población y la aplicación de procedimientos de contraste de tipo confirmatorio.

Pero sí es posible e instructivo que los alumnos traten de formular hipótesis plausibles apoyadas en los datos empíricos disponibles y utilizando diversas técnicas de análisis exploratorio de datos. A título de ejemplo citamos algunos procedimientos que los estudiantes puedan aplicar usando un paquete de programas y que el profesor puede sugerir.

Un primer nivel de análisis se referirá al estudio de las distribuciones de frecuencias y resúmenes estadísticos de cada una de las variables para todo el conjunto de datos. Esto puede hacerse mediante opciones del paquete que permita calcular los estadísticos elementales (promedios, dispersión y medida de simetría) y las tablas y gráficas de distribución de frecuencias como histogramas y polígonos de frecuencia, gráficos del “tronco” o la “caja”. La representación gráfica de estas distribuciones, con la posible presencia de varias modas y de asimetría mas o menos acusada, nos ayudará también a decidir, para cada variable, cuál es el promedio media, mediana o moda y medida de dispersión más adecuada como base para la comparación.

El disponer de ordenadores hace posible una actitud más crítica y analítica hacia los datos recogidos. Si se detectan algunos valores aislados, que se desvían demasiado de los restantes, podemos optar, tras una indagación de las posibles causas, por suprimirlos y repetir el estudio uní variante. Asimismo, la selección de una amplitud de intervalos de clase en los histogramas, que se adapte bien a los datos, pueden hacerse probando con distintos valores y viendo su efecto en la pantalla.

También se pueden comparar las características de estas distribuciones con otras disponibles, procedentes de tablas de medidas físicas correspondientes a la, población de la cual se ha extraído la muestra, en nuestro caso de una población de jóvenes de edades similares a la de nuestros estudiantes, para juzgar la posible representatividad de la misma

Comparación de subpoblaciones La opción de selección de casos permite hacer los análisis estadísticos para distintas submuestras, en nuestro caso, por ejemplo, para hombres y mujeres. Los que hace posible la comparación de los estadísticos y las distribuciones de frecuencias. Los “gráficos de cajas paralelas” permiten una



comparación simultánea y visual de los valores de las medianas, cuarteles, recorrido intercuartílicos, asimetría y valore atípicos.

En algunos casos, las diferencias entre grupos serán menores para otras variables, llegando incluso el caso en que tengamos duda en la interpretación. Esto nos llevará a dos cuestiones: por un lado, nos interesará dar un coeficiente o medida de la intensidad de la relación entre variables, lo que nos conduce a la idea de asociación. Además, necesitaríamos la herramienta del contraste estadístico de hipótesis, para poder tomar una decisión sobre la existencia de diferencias significativas en las poblaciones. La introducción de estas técnicas para alumnos podría ser motivada con este ejemplo, en las poblaciones. La introducción de estas técnicas para alumnos mayores también podría ser motivada con este ejemplo

Estudio de la asociación estadística: tablas de contingencia, medidas de asociación y regresión el estudio de las tablas de contingencia, de las variables agrupadas en intervalo, frente al sexo, nos permite comparar las distribuciones condicionales y la diferencia entre frecuencias observadas y esperadas y estudiar los conceptos de asociación e independencia entre variables estadísticas.

Además, puede ser necesario el estudio de la nube de puntos de pares de variables tales como tiempo en recorrer 100 y 500 metros, salto de longitud y altura, que nos da una primera visión gráfica de la existencia o no de asociación entre variables numéricas y de la intensidad de la misma. El estudio del coeficiente de correlación junto con el de la recta de regresión puede precisar aún más esta idea de asociación.

Puesto que algunas de las variables que queríamos relacionar con el sexo están relacionadas entre sí, surgirá la idea de hasta qué punto la asociación detectada entre, por ejemplo, el salto de longitud y la altura depende del sexo. Podemos discutir aquí la idea de cuando una variable debe o no jugar el papel de dependiente o independiente, o si sus papeles son intercambiables. También se verá la necesidad de buscar modelos más complejos que incluyan las relaciones entre efectuado, servirá como base intuitiva para la comprensión posterior de estos conceptos

4.- PROGRAMAS UTILIZABLES EN LA ENSEÑANZA

Actualmente existe una gran variedad de programas estadísticos, tanto los de tipo profesional, como los desarrollados especialmente con fines educativos. A continuación describimos brevemente los principales tipos de software utilizable:

Paquetes estadísticos profesionales como B.M.D.P., S.P.S.S. , SYSTAT, SATVIEW, STATGRAPHICS, especialmente las versiones para entorno Windows o Mac Intosh, que no requieren el aprendizaje de los comandos, también los desarrollados especialmente para ser usados en la enseñanza, como MINITAB. La principal finalidad es el Cálculo y representación gráfica. Son también un recurso profesional y permiten el aprendizaje a diversos niveles de complejidad.

Hojas electrónicas, disponible en diferentes paquetes integrados. Aunque más incompletas, permiten comprender los algoritmos de cálculo y pueden servir para otras aplicaciones diferentes de la estadística, concretamente el Microsoft Office Excel, en la tabla de herramientas tiene el comando Análisis de Datos.

Software didáctico para fines especiales, como los siguientes: Statlab (Inferencia), Gasp (Procesos Estocásticos) Tabletop(Exploración de contextos multivariantes para alumnos muy jóvenes), Quercus(Curs o de autoaprendizaje de bioestadística)



5.- Bibliografía

Batanero, C., Estepa, A. y Godino, J. D. (1991a). Análisis exploratorio de datos: Sus posibilidades en la enseñanza secundaria. *Suma*, 9, pp. 25-31.

Batanero, C., Godino, J. D. y Estepa, A. (1991b). Laboratorio de estadística. Uso del paquete de programas PRODEST. Granada: Dpto de Didáctica de la Matemática. Universidad de Granada.

Estepa, A. (1994). Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores. Tesis Doctoral. Universidad de Granada.

Godino, J. D. y Batanero, C. (1993). Ordenadores y enseñanza de la estadística. *Revista de Educación de la Universidad de Granada*, 7, 173-186.

Godino, J. D. y Batanero, C. (1994). Enfoque exploratorio en el análisis multivariante de los datos educativos. *Epsilon*, 10 (2), pp. 11-22.

Batanero (De.): Research on the Role of Technology in Teaching and Learning Statistics, 1996 IASE Round Table Conferenc Papers (pp. 255-264). Universidad de Granada.

