

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

Juan D. Godino

Versión revisada del artículo publicado en UNO, 5, 45-56, 1995

RESUMEN

En este trabajo analizamos la influencia de los ordenadores en el desarrollo y difusión de la estadística en los últimos años. Como consecuencia, razonamos la necesidad de propiciar una renovación de los contenidos y metodología en la enseñanza de esta materia, haciéndola más significativa. Finalmente presentamos ejemplos de las diversas modalidades de uso de los ordenadores como ayuda para el aprendizaje de conceptos y métodos estadísticos.

INTRODUCCIÓN

En la actualidad nos encontramos en un proceso de reforma curricular en los diversos niveles de enseñanza, en el que se reconoce una gran importancia a la formación de los estudiantes en estocástica, esto es, sobre ideas probabilísticas y estadísticas y su mutua interdependencia. Un ejemplo significativo lo constituye el currículo de la Enseñanza Secundaria Obligatoria. De los cinco bloques de contenidos que propone el MEC (1992) para este nivel de enseñanza dos están relacionados con la Estadística: el 4, "Interpretación, representación y tratamiento de la información" y el 5 "Tratamiento del azar". Igualmente, en la formación profesional y en la universidad los cursos sobre estadística aplicada se incluyen prácticamente en todas las especialidades.

Este fenómeno no es exclusivo de España: la enseñanza de los contenidos referidos a la estadística y probabilidad 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 en el último Congreso Internacional sobre la enseñanza de la Estadística (ICOTS IV, celebrado en Marruecos en 1994) hubo un grupo de trabajo especial dedicado a "El curriculum estadístico del año 2000".

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 (Batanero, Godino y Vallecillos, 1992). Esta situación, aparentemente, ha sido superada: Por un lado aparecen los entornos operativos "amistosos", que permiten acceder directamente al manejo de cualquiera de los módulos de un paquete estadístico y, con la ayuda del "ratón", explorar sus posibilidades. Por otro lado, existen programas "de consulta" (Brent y Mirelli, 1991) 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 nuestras hipótesis sobre los mismos, ¿Quiere ésto decir que hemos resuelto definitivamente el problema de la estadística? ¿Se debe reducir esta enseñanza a enseñar a los alumnos el uso de este tipo 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 día el conocimiento del modo de procesar datos mediante un programa estadístico, por lo que deberíamos, en la medida de lo 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.

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 accesible para él. Este objetivo comienza por la comprensión de la idea básica de *sistema de datos*.

Este término es más adecuado que el de conjunto de datos, según Vasco (1994), 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 el mismo las unidades 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 cada uno de sus valores se darán 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 las variable (campos), que es el cuerpo principal de los datos, y los resúmenes estadísticos de cada variable. Los campos pueden ser de longitud fija o variable, y puede haber campos vacíos. Asimismo, clasificamos las variables según diversas tipologías: cualitativas o cuantitativas; discretas, continuas; nominales, ordinales, datos de intervalo, de razón.

Sobre cada una de estas componentes pueden realizarse operaciones o transformaciones internas (clasificación, recodificación, 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, similaridad (dependencia entre variables; similaridad de sujetos; similaridad de variables,... Estos tipos de operaciones deben ser presentadas 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 nos 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 la escuela, temperatura máxima y mínima 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 trabajos 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 analizarlo 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. (Biehler, 1991). 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 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 del uso no adecuado 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 de 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 (Vallecillos y Batanero, en prensa). 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 escuela para comparar la intención de voto de chicos y chicas en las próximas elecciones al consejo escolar, deben recoger una muestra lo suficientemente representativa de chicos y chicas en los diferentes cursos escolares y deben recoger datos sobre las principales variables que influyan 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, que merece una nueva sección.

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 multivariantes son empleados en la actualidad dentro de esta misma filosofía, para analizar fenómenos físicos o sociales complejos (Godino y Batanero, 1994).

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.

Como exponemos en Batanero y cols. (1991a) 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 una 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 desviación 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 represente 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 re-analizan 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 herramientas posibles, incluso las ya existentes. El propósito es extraer cuanta información sea posible, generar "hipótesis" 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 el fenómeno y efectuar sobre ellos un análisis estadístico 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 algunas de las técnicas gráficas asociadas al análisis exploratorio de datos.

EJEMPLO 1: ¿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 el empleo del tiempo libre, el tipo de

música o 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 y 500 metros, tiempo que es capaz de mantenerse haciendo el pino, etc. Estas marcas podrían recogerse a principios y final de un trimestre, con objeto de comprobar cuál de los dos sexos ha mejorado más sus marcas iniciales. También podríamos recoger datos de algunas variables físicas como el peso y la altura.

Si los alumnos tienen acceso al uso 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 cuestiones:

- (1) ¿Existen diferencias entre varones y hembras en las características estadísticas de alguna de estas variables? ¿En cuál es esta diferencia más o menos acusada?
- (2) ¿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 escolar 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 los datos. A título de ejemplo citamos algunos procedimientos que los alumnos pueden aplicar usando un paquete de programas y que el profesor puede sugerir.

a) *Estudio univariante.* 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 permitan calcular los estadísticos elementales (promedios, dispersión y medidas de simetría) y las tablas y gráficas de distribuciones de frecuencias como histogramas y polígonos de frecuencias, gráficos del "tronco" o la "caja". La representación gráfica de estas distribuciones, con la posible presencia de varias modas y de asimetría más 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 univariante. Asimismo, la selección de una amplitud de los intervalos de clase en los histogramas, que se adapte bien a los datos, puede 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 correspondiente 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 alumnos, para juzgar la posible representatividad de la misma.

b) *Comparación de subpoblaciones.* La opción de selección de casos permite hacer los análisis estadísticos para distintas submuestras, en nuestro caso, por ejemplo, para varones y hembras, lo que hace posible la comparación de los estadísticos y las distribuciones de frecuencias. Los "gráficos de cajas paralelos" permiten una comparación simultánea y visual de los valores de las medianas, cuartiles, recorridos intercuartílicos, asimetría y valores atípicos.

En algunos casos, las diferencias entre grupos serán muy claras ya que no se solaparán los intervalos que cubren el 50% de los valores centrales. Estas diferencias 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 mayores también podría ser motivada con este ejemplo.

c) *Estudio de la asociación estadística: tablas de contingencia, medidas de asociación y regresión.* El estudio de las tablas de contingencia, de las variables agrupada en intervalos, 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 altura depende del sexo. Podremos 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 tres o más variables. Estos modelos serán estudiados en cursos más avanzados, pero el estudio exploratorio que hemos efectuado, servirá como base intuitiva para la comprensión posterior de estos conceptos.

SIMULACIÓN DE SITUACIONES ESTOCÁSTICAS

La exploración que permiten los ordenadores en el terreno estocástico no queda restringida a la resolución de problemas, sino que puede ser extendida al de los mismos conceptos abstractos, mediante la creación de micromundos estocásticos con los que se puede interactuar. La facilidad de simulación de experimentos aleatorios hace posible la experimentación, observación y exploración de procesos estocásticos. También existe la posibilidad de variar los parámetros de los que dependen el comportamiento estocástico de tales procesos, de una forma interactiva, y de observar el efecto producido, mediante las representaciones gráficas. De este modo los nuevos "objetos" pierden su carácter abstracto, proporcionando una experiencia estocástica que no es fácil alcanzar en la vida real.

El fundamento de la simulación de fenómenos aleatorios se basa en dos principios (Heitele, 1975). El primero es la concepción frecuencial de la probabilidad, que considera ésta como el número a que tendería la frecuencia relativa de un suceso, al repetir indefinidamente un experimento aleatorio, y viene avalada teóricamente por el teorema de Bernoulli. Aunque es importante estudiar, de un modo teórico, los teoremas del cálculo de probabilidades y a partir de ellos, deducir las técnicas necesarias para el cálculo de la probabilidad asociada a determinados sucesos complejos, estamos a veces interesados en determinarlas experimentalmente. El valor que obtenemos es aproximado, no en el sentido ordinario del análisis numérico, sino en el sentido estocástico, de convergencia de las frecuencias relativas a las correspondientes probabilidades teóricas. El segundo principio es la posibilidad de estudiar las propiedades de un fenómeno aleatorio, sustituyéndolo por otro "isomorfo", es decir, en el que el espacio muestral, y el álgebra de sucesos considerada sean isomorfas y los sucesos elementales, que se corresponden en los dos experimentos, tengan las mismas probabilidades.

El interés de la simulación en la enseñanza de la estadística y de la probabilidad es destacada entre otros en los trabajos de Biehler (1991) y Estepa (1994). Como describe este último autor hay fenómenos que, por sus características espacio - temporales o de otra índole son difíciles de

observar. Por medio de la simulación se construye, con el ordenador, un modelo simplificado de dicho fenómeno, eliminando las variables irrelevantes a su estructura, condensado en el tiempo y manipulable por el alumno. El uso de programas de simulación permite poner en manos del alumno un nuevo instrumento que hace posible la exploración y el descubrimiento de conceptos y principios que de otro modo serían mucho más abstractos, contribuyendo a paliar el problema de la falta de experiencia estocástica y a la mejora de la intuición probabilística que, en general, no se desarrollan espontáneamente (Fischbein, 1975). A continuación mostramos, con un ejemplo, estas posibilidades.

EJEMPLO 2: Distribuciones en el muestreo

El empleo de la simulación es especialmente útil en la enseñanza de conceptos de inferencia estadística. Por ejemplo, haciendo que el ordenador realice muestreos repetidos y nos proporcione un resumen de los resultados, representando gráficamente las distribuciones en el muestreo obtenidas de forma empírica. Los puntos esenciales en estas distribuciones es reconocer que la estimación de un parámetro en la población variará de una muestra a otra (el estimador es una variable aleatoria) y que esta variación se ajustará a un patrón predecible (podemos conocer el tipo de distribución que sigue el parámetro). Si elegimos un estimador insesgado, el valor esperado del estadístico se acercará al del parámetro en la población. Aunque el estadístico varía de una muestra a otra, los valores más probables son próximos al del parámetro.

La idea de distribución en el muestreo es muy compleja y su comprensión no puede disociarse de la comprensión del proceso de muestreo a partir de la cual se forma (Lipson, 1994). El proceso de muestreo es dinámico y depende de la población, el tipo de muestras extraídas, el estadístico que se calcule a partir de estas muestras y su distribución. En una enseñanza tradicional, este proceso se explica verbalmente, y los estudiantes deben imaginarlo, ya que sólo se les muestra el punto final, que es la distribución del estadístico en el muestreo, la cual es deducida algebraicamente. En el mejor de los casos, se les presenta un ejemplo artificial de población (formada por unos pocos elementos) y se generan todas las posibles muestras de dos o tres elementos para mostrar con un ejemplo particular la distribución del estadístico.

Por el contrario, en un entorno informático el alumno tiene la oportunidad de elegir una población teórica. Por ejemplo, si se trata de estimar la altura de los estudiantes de una clase, podríamos elegir una población normal de media y varianza convenientes. Si nos interesamos por la esperanza de vida, la distribución normal no sería adecuada y podríamos sustituirla por una exponencial. En lugar de escoger una distribución teórica, podríamos introducir en el ordenador los datos sobre las alturas de los alumnos de la clase, cuya media conocemos y explorar el efecto del tamaño de la muestra sobre la estimación de la altura media de los alumnos a partir de un muestreo.

Los alumnos podrán observar las diferentes muestras según se van seleccionando y el valor obtenido para el estadístico en cada caso (Shaugnessy y Batanero, 1995). En el caso de paquetes estadísticos como Statlab (Stirling, 1987), podemos tener en pantalla simultáneamente dos ventanas, En una de ellas observamos una representación gráfica de la población y en la otra la de los valores obtenidos del estadístico, esto es, la distribución en el muestreo. En cada una de estas ventanas podemos cambiar la representación gráfica: diagramas de puntos, gráfico de la caja o histogramas, en los que podemos variar la anchura de los intervalos. El ordenador da un papel activo al alumno en la generación de las distribuciones en el muestreo. Variando el tamaño de la muestra podremos observar el efecto sobre la variabilidad del estadístico. Es fácil que el alumno observe la forma aproximadamente normal de la distribución en el muestreo para muestras de suficiente tamaño. De aquí podrá deducir la importancia de elegir un tamaño adecuado de muestra.

En lugar de representar gráficamente la distribución del estadístico, podríamos representar los intervalos de confianza que se van generando, a partir de los valores obtenidos del estadístico en

cada una de las muestras, para un coeficiente de confianza dado. Interviene aquí un nuevo concepto a explorar: el de coeficiente de confianza. Experimentando con ellos, el alumno podrá comprobar que la amplitud del intervalo de confianza depende de tres factores: el coeficiente de confianza, la dispersión en la población y el tamaño de la muestra. Algunos de los intervalos de confianza obtenidos cubren el valor del parámetro y otros no lo cubren. Si hemos elegido un coeficiente de confianza del 95%, veremos que, aproximadamente 5 de cada 100 intervalos no cubren el valor del parámetro. Los alumnos tendrán así una oportunidad de comprender el significado del coeficiente de confianza y no caerán en el error común de creer que, para una muestra particular, la probabilidad de que el parámetro esté comprendido en el intervalo de confianza es el 95%.

El ejemplo que hemos propuesto es sólo uno de los posibles. Cualquier otro experimento estocástico puede ser simulado en el ordenador con alguno de los muchos programas disponibles. Por ejemplo, el paquete GASP (Graphical Aids for Stochastic Processes) de Fisch y Griffeath (1988) permite la simulación de diferentes tipos de procesos estocásticos, desde la caída de bolas en un aparato de Galton hasta los procesos de difusión, incluyendo los procesos de poisson, de nacimiento y muerte, de ramificación y la teoría de colas. En nuestro caso, hemos incorporado al paquete PRODEST (Batanero y cols., 1991b) la posibilidad de simulación de las sucesiones de Bernoulli y los modelos probabilísticos asociados y de la extracción de muestras de diversos tipos de poblaciones.

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.
- Software didáctico para fines especiales, como los siguientes: Statlab (Inferencia), Gasp (Procesos estocásticos), Tabletop (Exploración de contextos multivariantes para alumnos muy jóvenes), Quercus (Curso de autoaprendizaje de bioestadística)

LAS CALCULADORAS GRÁFICAS

Considerados por algunos autores como la tecnología del futuro, debido a que su coste cada vez menor hace creíble que, en el futuro, cada estudiante pueda disponer de su propia calculadora. Entre las nuevas posibilidades que ofrecen a la enseñanza de la estadística, citamos:

- Transmisión de datos (entre calculadoras o calculadora y ordenador). Es posible, por ejemplo, tomar datos de internet, sobre un tema de interés y transmitirlo a la calculadora, sin necesidad de tener que grabarlos a mano.
- Opciones de manejo de listas
- Posibilidad de transformación de los datos
- Cálculos estadísticos básicos para una y varias variables
- Gráficos estadísticos usuales
- Posibilidad de ser programadas

- Generador de números aleatorios y tablas estadísticas básicas.

RECURSOS EN LA INTERNET

Una nueva dimensión en la enseñanza y la práctica estadística está siendo marcada por Internet. En esta sección realizamos un resumen del trabajo de Snell (1996), que presenta algunos de los recursos disponibles en la red.

Cursos y materiales didácticos

El prototipo de los cambios previsibles con las nuevas tecnologías es el curso Chance, desarrollado en cooperación por varias universidades americanas y situado en la dirección (<http://www.geom.umn.edu/locate/chance>). Este curso presenta el uso de los conceptos básicos de estadística en la prensa. Un boletín electrónico proporciona trimestralmente resúmenes de artículos de prensa que usan conceptos de estadística. Adicionalmente una base de datos en WWW contiene planificación de cursos que han utilizado este material y una guía para el profesor.

Las clases de un curso de este tipo se organizan del modo siguiente: se elige un artículo reciente y se preparan algunas preguntas relacionadas. Los estudiantes, en grupos, leen el artículo e intentan contestar las preguntas formuladas u otras relacionadas que surjan durante la discusión. Todo ello se utiliza como base para introducir un tema de estadística relacionado con el contenido del artículo.

The Journal of Statistics Education

Es una revista con sistemas de referee, publicada desde 1993, electrónicamente, cuyo tema es la enseñanza de la estadística a nivel universitario. Se recibe libre de costo por internet y también puede conseguirse los números atrasados o un artículo suelto directamente a partir de la dirección donde están archivados. La universidad de North Carolina mantiene una base de datos relacionada con esta revista (<http://>) donde se contiene otra serie de recursos para la enseñanza de la estadística. Una diferencia con una revista convencional es que es posible a los lectores mandar comentarios a un artículo o hacer búsquedas automatizadas de artículos sobre un cierto tema. Muchos de estos comentarios serán seleccionados para pasar a ser parte del archivo y, por tanto, del propio artículo.

Un recurso interesante es el llamado “*data sets and stories*”, donde se acumulan conjuntos de datos, junto con su descripción y algunas indicaciones de sus posibles usos en la enseñanza. Los datos se pueden recuperar en formato útil para la mayor parte de paquetes estadísticos, hojas de cálculo y calculadoras gráficas. Otro el denominado “teaching bits” que proporciona resúmenes de artículos de interés para los profesores de estadística.

Grupos de discusión o trabajo

Se puede ser miembro activo de un grupo de discusión sobre la enseñanza de la estadística por correo electrónico. Los temas pueden ir desde pedir la solución a un problema o sugerencias sobre como resolver la dificultad que tiene un estudiante, anuncios de nueva bibliografía, intercambio de material didáctico, etc. Estas discusiones también se archivan para posibles búsquedas sobre un tema. Dos grupos relevantes son el grupo de discusión *Edstat-L* mantenido en el servicio JSE y el *International Study group for Research on Learning Probability and Statistics* (Stated-List) mantenido actualmente en la universidad de Granada. Este último publica también un boletín distribuido por correo electrónico con resúmenes de trabajos de investigación de sus miembros y noticias sobre reuniones, proyectos y recursos de interés para la investigación en el área. La dirección desde la que puede recuperarse el boletín es (gopher://jse.stat.ncsu.edu).

Otros recursos:

- (1) The Statistical Instruction Internet Palette (<http://seamonkey.ed.asu.edu/beherns/siip/>).

- (2) Está dividido en cinco secciones: Data gallery, graphing studio, computing studio, equation gallery, classroom gallery y Wide World of Web Data.
- (3) Electronic Textbook: <http://WWW.stat.ucla.edu/>
- (4) Statlib: <http://lib.stat.cmu.edu> (Department of Statistics, Carnegie Mellon). Principal punto de referencia sobre software y datos disponibles en la red.

CONCLUSIONES

En este trabajo hemos analizado algunos de los nuevos objetivos educativos que la disponibilidad de ordenadores plantea sobre la enseñanza de la estadística. También hemos mostrado, mediante algunos ejemplos, las posibilidades didácticas de los ordenadores tanto en el análisis exploratorio de datos como en la exploración de las ideas estocásticas. Los programas a que hemos hecho referencia son sólo una muestra de la gran variedad disponible. Desde paquetes gráficos de utilidad general, hojas electrónicas, paquetes estadísticos profesionales, hasta el "software" didáctico específico para la enseñanza de la estadística, como los ejemplos mostrados.

Estos programas por sí mismos no resuelven el problema de la enseñanza. Será necesario una gran labor de reflexión e investigación para elaborar las correspondientes "guía didáctica" adecuadas a la enseñanza de conceptos específicos. Con ayuda de ellas podríamos proporcionar al alumno un micromundo estocástico, esto es, un entorno operativo que le permite generar, observar, reflexionar e interactuar sobre fenómenos que difícilmente se pueden explorar sin la herramienta informática. En estos micromundos se usa un núcleo de conceptos y procedimientos matemáticos interrelacionados, en situaciones problemáticas particulares, mediante sistemas de representación múltiples, dinámicos y simultáneos. Por lo tanto, es de esperar un énfasis en componentes nuevos del significado de los objetos matemáticos implícitos.

Las situaciones didácticas deben configurarse de tal modo que el alumno tenga que resolver problemas cuya apertura sea graduable y en las cuales se destaquen las fases de acción, formulación de conjeturas y validación (Brousseau, 1986), que el profesor deberá completar con la fase de institucionalización de los conocimientos puestos en juego.

REFERENCIAS

- 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.
- Batanero, C., Godino, J. D. y Vallecillos, A. (1992). El análisis de datos como útil y como objeto en la didáctica de la Matemática.
- Burrill, G. (1996). Technology and the teaching and learning of statistics. En C. Batanero (Ed.), *Research on the Role of Technology in Teaching and Learning Statistics, 1996 IASE Round Table Conferenc Papers* (pp. 25-40). Universidad de Granada.
- Biehler, R. (1991). Computers in probability education. En R. Kapadia (Ed.), *Chance encounters: probability in education. A review of research and pedagogical perspectives*. (pp. 109-211). Amsterdam: Reidel. Amsterdam.
- Brent, E. y Mirielli, E. (1991). *Statistical navigator professional*. Columbia, MO: The Idea Works, Inc.
- Brousseau, G. (1986). Fondements et méthodes de la didactique des mathématiques. *Recherches en Didactique des Mathématiques*, 7(2), 33-115.

- 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.
- Fischbein, E. (1975). *The intuitive sources of probability thinking in children*. Dordrecht: D. Reidel.
- Fisch, B. y Griffeath, D. (1988). *Graphical aids for stochastic processes*. Belmont, CA: Wadsworth & Brooks/Cole.
- 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.
- Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics*, 6, 187-205.
- Lipson, K. (1994). Understanding the role of computer based technology in developing fundamental concepts of statistical inference. En *Proceedings of the IV International Conference on Teaching Statistics* (V.1, pp. 65-72). Marrakech: The National Institute of Statistics and Applied Economics.
- Mc Closey, M. (1996). Quercus and Steps: the experience of two cal projects from Scottish universities. En C. Batanero (Ed), *Research on the Role of Technology in Teaching and Learning Statistics, 1996 IASE Round Table Conferenc Papers* (pp. 101-118). Universidad de Granada.
- M.E.C. (1992). *Matemáticas Secundaria Obligatoria*. Madrid: Ministerio de Educación y Ciencia.
- Shaughnessy, J. M. y Batanero, C. (1995). Un enfoque visual para enseñar la probabilidades binomiales. *UNO*, 5, 103-112.
- Snell, L. (1996). The internet: A new dimension in teaching statistics. En C. 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.
- Stirling, W.D. (1987). *STATLAB. Microcomputer based practical classes and demonstrations for teaching statistical concepts*. Wellington, NZ: The New Zealand Statistical Association.
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison Wesley.
- Vasco, C. (1994). Handling data systems in the curriculum for general basic education. En: *Proceedings of the IV International Conference on Teaching Statistics*, Marrakech: The National Institute of Statistics and Applied Economics, V.1, pp. 8-14
- Vallecillos, A. y Batanero, C. (1995). Algunas dificultades comunes en el empleo de la Estadística en la investigación experimental. *Revista de Educación de la Universidad de Granada*, 8, 5-16.