



MEMORIA DE LAS ACCIONES DESARROLLADAS
PROYECTOS DE MEJORA DE LA CALIDAD DOCENTE
VICERRECTORADO DE PLANIFICACIÓN Y CALIDAD
XI CONVOCATORIA (2009-2010)



❖ **DATOS IDENTIFICATIVOS:**

Título del Proyecto

Paquete Estadístico: Probabilidad, Econometría y Análisis de Datos. (VIII)

Resumen del desarrollo del Proyecto

En la actualidad existen en el mercado numerosos paquetes informáticos de tratamiento de datos estadísticos. La gran mayoría de estos programas requieren bastantes conocimientos, a priori, de estadística para comenzar a utilizarlos. A todo esto se une la poca disponibilidad (alto coste de licencias de uso) de estos programas restringiéndose su uso a algunas "horas" de prácticas en centros de cálculo.

Las nuevas tendencias en los planes educativos universitarios de convergencia europea en los que se apuesta por el esfuerzo y refuerzo del trabajo individual del alumno, hace necesario el ofrecer a los estudiantes todo el material y herramientas necesarias para completar su proceso de autoaprendizaje, y en esta línea se pretende desarrollar el proyecto.

Por tanto, se hace patente la necesidad de un software de análisis de datos en español, de fácil utilización y amplio sistema de ayudas para el desarrollo de la materia de Estadística en esta Universidad y que además pueda estar en todo momento a disposición del alumnado para completar, de esta forma, su formación en esta disciplina.

En el presente proyecto se pretende seguir con los desarrollos iniciados en convocatorias anteriores proponiendo nuevos módulos de desarrollo que amplían los resultados obtenidos anteriormente, dotando al programa PEAD de mayor potencia y ámbito de aplicación al implementar nuevas técnicas estadísticas y su consecuente aplicación.

Coordinador/a:

Nombre y apellidos	Código del Grupo Docente	Departamento
Roberto Espejo Mohedano	069	Estadística, Econometría e I.O

Otros participantes:

Nombre y apellidos	Código del Grupo Docente	Departamento
José Diz Pérez	017	Estadística, Econometría e I.O
Manuel Jurado Bello	017	Estadística, Econometría e I.O

Asignaturas afectadas

Métodos estadísticos en la Ingeniería	Estadística, Econometría e I.O.	Ingeniero Técnico Industrial en: Electricidad, Electrónica y Mecánica
Estadística	Estadística, Econometría e I.O.	Ingeniero Técnico Informático Especialidad: Gestión y Sistemas
Bioestadística	Estadística, Econometría e I.O.	Ciencias Biológicas
Estadística	Estadística, Econometría e I.O.	Ciencias Ambientales
Técnicas Estadísticas de Investigación Social	Estadística, Econometría e I.O.	Relaciones Laborales
Análisis de datos en Ciencias Sociales	Estadística, Econometría e I.O.	Relaciones Laborales

MEMORIA DE LA ACCIÓN

1. Introducción (justificación del trabajo, contexto, experiencias previas etc.)

Como ya se ha comentado en convocatorias anteriores, la necesidad de disponer de un software específico para completar la docencia en algunas disciplinas es indispensable, pero desgraciadamente, algunas veces ya sea debido al idioma, la dificultad de utilización, la falta de manuales adecuados, la complejidad del propio software o simplemente el elevado coste en las licencias de uso de algunos programas, hacen que el uso de este software específico se limite en cualquier caso a unas "pocas horas" de prácticas en laboratorio con la consecuente falta de disponibilidad del mismo para el alumnado, lo que conlleva una inadecuada preparación de las prácticas, y por tanto de la disciplina genérica.

Se pretende con este proyecto el desarrollo de un software estadístico para ser utilizado en la docencia de las asignaturas de Estadística impartidas en la Universidad de Córdoba, implementando una serie de módulos estadísticos clásicos bajo entorno Windows 9x/NT/XP, aprovechando las ventajas del entorno gráfico que proporciona Visual C++.NET. Estos módulos se documentarán de la forma mas precisa posible de manera que cualquier intento de modificación o ampliación posterior pueda producirse de forma sencilla, como ha estado ocurriendo hasta la actualidad con las versiones anteriores.

Además, los módulos creados contendrán un tratamiento de errores lo mas amplio posible y que requiera una intervención mínima por parte del usuario, junto con el correspondiente sistema de ayuda en línea. Los módulos se desarrollarán de manera que posean un alto nivel de portabilidad y de facilidad de uso, de forma que, aunque el usuario no esté relacionado con el campo de la informática, pueda comprender y sacar el máximo rendimiento a la misma. También se dotará a la aplicación de sistemas de control inteligente de errores tanto en la manipulación de datos como en la técnica estadística a utilizar.

En este punto se puede describir de manera sucinta los progresos que se han conseguido desde la primera versión del programa PEAD a desarrollar:

- En la primera versión se desarrollaron las bases de la aplicación implementando una versión simple con algunas utilidades estadísticas básicas como podrían ser el análisis descriptivo, la regresión y correlación y algunos procedimientos para comparación de muestras.
- En la segunda versión se implemento un nuevo entorno de trabajo permitiendo el tratamiento simultáneo de varios ficheros de datos. Además se desarrollo un nuevo sistema de introducción y manipulación de datos con todas las posibilidades asociadas: introducir y modificar variables y datos, transformación de las mismas, impresión de datos y resultados, etc. Además se puso un interés especial en el tratamiento de errores y el sistema de ayuda.
- La tercera versión se implementaron varios módulos de cálculos estadístico y se mejoraron sensiblemente los ya existentes (esta es una pauta que hemos

seguido a través de las distintas versiones). Los nuevos procedimientos que se pusieron operativos eran los correspondientes a contrastes de hipótesis tanto para una población con dos poblaciones (independientes y relacionadas), estimación de parámetros tanto por punto como por intervalo en una o dos poblaciones, estimación de tamaños muestrales y se comenzó con el tratamiento del análisis de la varianza.

- En la cuarta se terminó el módulo de análisis de la varianza tanto con un factor como con dos o incluso tres. También se introdujo un módulo de consulta interactiva de tablas estadísticas para las principales distribuciones tanto discretas como continuas así como el cálculo de probabilidades asociadas. Además se amplió el módulo de regresión añadiendo nuevos test sobre la validez del modelo y coeficiente de correlación así como mejora sustancial en la interfaz de usuario y presentación de resultados, y finalmente se implementó un módulo dedicado a la realización de test no paramétricos.
- En la quinta versión se implementó un nuevo módulo de gráficos estadísticos interactivos usuales, se ha mejorado del sistema de ayudas y se comenzó la preparación de un sistema de asesoramiento en las técnicas estadísticas para usuarios de análisis de datos no experimentados en las mismas.
- Con la versión VI del paquete Estadístico se ha conseguido crear un Asesor Estadístico que ofrece al usuario la posibilidad de encontrar el procedimiento adecuado para analizar sus datos. Para ello se pedirá información a través de cuestiones sencillas, que apoyadas en ejemplos visuales del propio programa, permitan guiar al usuario hasta una posible solución a su problema. También se ha unificado el sistema de ayuda y se han corregido algunas anomalías detectadas en versiones anteriores.
- Finalmente, con la versión VII se ha conseguido poder analizar una variable (numérica o categórica) para las distintas categorías de otra cualitativa. Implicó revisar y actualizar muchos de los procedimientos ya implementados para adaptarlos a la nueva posibilidad (procedimiento “descriptivos”, estimación de parámetros, contrastes, pruebas no paramétricas y test de ajuste). El segundo objetivo alcanzado con esta versión fue la implementación de un procedimiento de “tablas de contingencia” dentro del módulo cálculos descriptivos. Esta tarea implicó, entre otras cosas, la creación, manipulación y edición de tablas de frecuencias de clasificación doble. A partir de este punto se “aprovechó” para incorporar un procedimiento similar de tablas de frecuencias simples a añadir en cálculos descriptivos univariantes.

2. Objetivos

Desarrollo de un software estadístico para ser utilizado en la docencia de las asignaturas de Estadística impartidas en la Universidad de Córdoba, implementando una serie de módulos estadísticos clásicos bajo entorno Windows 9x/NT/XP/Vista, aprovechando las ventajas del entorno gráfico que proporciona Visual C++.NET. Estos módulos se documentarán de la

forma mas precisa posible de manera que cualquier intento de modificación o ampliación posterior pueda producirse de forma sencilla.

Además, los módulos creados contendrán un tratamiento de errores lo mas amplio posible y que requiera una intervención mínima por parte del usuario, junto con el correspondiente sistema de ayuda en línea. Los módulos se desarrollarán de manera que posean un alto nivel de portabilidad y de facilidad de uso, de forma que, aunque el usuario no esté relacionado con el campo de la informática, pueda comprender y sacar el máximo rendimiento a la misma.

La principal actividad a desarrollar en el presente proyecto es dotar a la aplicación PEAD (Programa Estadístico para análisis de datos), iniciada en convocatorias anteriores, de nuevas funcionalidades, al desarrollar módulos complementarios de aplicación estadística. En particular se trata de añadir un nuevo módulo al programa que permita trabajar con Series Temporales de datos, lo que puede conllevar la adaptación de la estructura interna y externa del mismo para poder satisfacer las necesidades del problema.

Específicamente, en este proyecto se trata de :

- Añadir un nuevo módulo al programa PEAD que permita trabajar con Series Temporales de datos, lo que puede conllevar la adaptación de la estructura interna y externa del mismo para poder satisfacer las necesidades del problema.
- El módulo debe permitir que el usuario elija el intervalo de tiempo en el que desea que se introduzcan los datos (años, meses, semanas, días y horas) y detectar mediante el cálculo algorítmico casos excepcionales como años bisiestos, etc.
- Se han de desarrollar procedimientos gráficos que permitan la representación de series temporales. Estos han de ser versátiles y permitir representar tanto series completas como periodos específicos atendiendo al carácter de la serie particular y escalar las graficas atendiendo a estas peculiaridades.
- Adaptar el módulo de importación de datos a la nueva estructura de datos a implementar para permitir la entrada de datos de aplicaciones de bases de datos de tiempo estándar como por ejemplo:
 - texto (txt): son un tipo de documento sencillo para representar datos en forma de tabla, en las que las columnas se separarán por espacios y las filas por saltos de línea.
 - texto (csv): son un tipo de documento sencillo para representar datos en forma de tabla, en las que las columnas se separarán por comas y las filas por saltos de línea.
 - texto (xer), series de base de datos del Ministerio de Economía y Hacienda: son un tipo de documento donde toda la información contenida en una serie temporal se ofrece en un fichero de texto, en un formato específico, compuesto por una serie de líneas cada una de las cuales indica un campo (Código, Título, Unidades, Fuente, etc.). Los datos pertenecientes a la serie en cuestión se muestran a continuación del último campo y van separados por separadores.

- Excel (xls): son un tipo de documento desarrollado por Microsoft para trabajar con hojas de cálculo.
- Proporcionar una ayuda dinámica y personalizada sobre cómo trabajar con este módulo a los usuarios no familiarizados con el programa.

Una última tarea a realizar y que se incluye como objetivo es revisar algunos módulos ya implementados. En particular, y debido al uso continuado desde su conclusión, se han detectado algunos problemas de falta de operabilidad en ciertos procesos y cálculos numéricos.

3. Descripción de la experiencia

Como ya se comentó en versiones anteriores de este proyecto, la complementación a través de las prácticas en ordenador (software estadístico) de las clases teóricas es un hecho indudable que favorece la comprensión y aplicación de las técnicas desarrolladas en clase. Es por tanto una realidad que las clases prácticas redundan en un beneficio hacia el alumno mejorando claramente su calidad de aprendizaje y mejora la calidad docente desde el punto de vista del profesorado.

Las nuevas tendencias en los planes educativos universitarios de convergencia europea en los que se apuesta por el esfuerzo y refuerzo del trabajo individual del alumno, hace necesario el ofrecer a los estudiantes todo el material y herramientas necesarias, por tanto, y en esta línea se ha pretendido desarrollar el proyecto.

Tradicionalmente las clases prácticas en las asignaturas de Estadística (y en general aquellas relacionadas con métodos numéricos) se han desarrollado en la pizarra a través de la realización de ejercicios y casos prácticos de forma manual, lo que conllevaba una lentitud en el desarrollo de la mismas y la acotación de estos casos prácticos a “pocos” datos numéricos y por tanto su distanciamiento a ciertas aplicaciones reales. En los últimos años estas situaciones pueden corregirse mediante la utilización del ordenador y su potencia de cálculo, y sobre todo dada su amplia difusión no sólo en centros de cálculo científico sino a nivel personal (la mayoría de los estudiantes universitarios posee un ordenador en casa). Los paquetes de programas estadísticos son la herramienta que permite el análisis de gran cantidad de datos de una manera rápida y cómoda, pero tienen algunos inconvenientes para su utilización por parte del alumno que hay que reseñar: en primer lugar el idioma ya que suelen ser en inglés lo que dificulta su utilización y el seguimiento de los sistemas de ayuda, aunque en los últimos tiempos esta tendencia está cambiando. En segundo lugar el lenguaje de control en algunos paquetes que requiere, por tanto, del aprendizaje de un lenguaje de programación de alto nivel y ciertos conocimientos a nivel informático por parte del alumno. La amplia gama de procedimientos estadísticos genéricos de los que suelen constar estos paquetes hace que el alumno, en ocasiones, se “pierda” entre ellos, no sabiendo cuál es el que debe utilizar; Y por último, y más grave, la no disponibilidad de estos paquetes para uso personal dado el alto precio de los mismos, hace que la utilización de estos paquetes por parte del alumno se restrinja a “algunas” horas en los centros de cálculo en el mejor de los casos o bien hay que decantarse por el uso de versiones antiguas ya descatalogadas, con los inconvenientes que este hecho trae consigo.

Son estas razones las que ponen de manifiesto la necesidad de la preparación de un software estadístico a los que puedan acceder libremente y en todo momento los alumnos de esta disciplina. El paquete de programas estadísticos desarrollado PEAD (paquete estadístico de análisis de datos) en sus distintas versiones, se adapta perfectamente a los programas de la asignatura y dispone de un amplio sistema de ayudas en castellano, lo que facilita la comprensión de los conceptos teóricos vistos en clase y facilita la resolución de casos prácticos reales propuestos en las clases de problemas. Además, dada la portabilidad del paquete, el alumno puede disponer del mismo no sólo en los centros de cálculo de esta Universidad sino también en casa para su uso personal, lo que redundará en un beneficio evidente para el alumno, más aun en estos tiempos donde se está implantando una nueva visión de los estudios universitarios a través del nuevo crédito europeo y el apoyo al esfuerzo individual al estudio.

Con la nueva versión del paquete estadístico, además de las virtudes de las anteriores y un nuevo módulo operativo, se dispone de un entorno de trabajo adaptado a las nuevas plataformas operativas, se corrigen algunas deficiencias de las primeras versiones, se amplían los sistemas de corrección de errores tanto en los datos como en la utilización de la técnica estadística a usar, además de dotarlo de un entorno de trabajo más cómodo y estable para el usuario final.

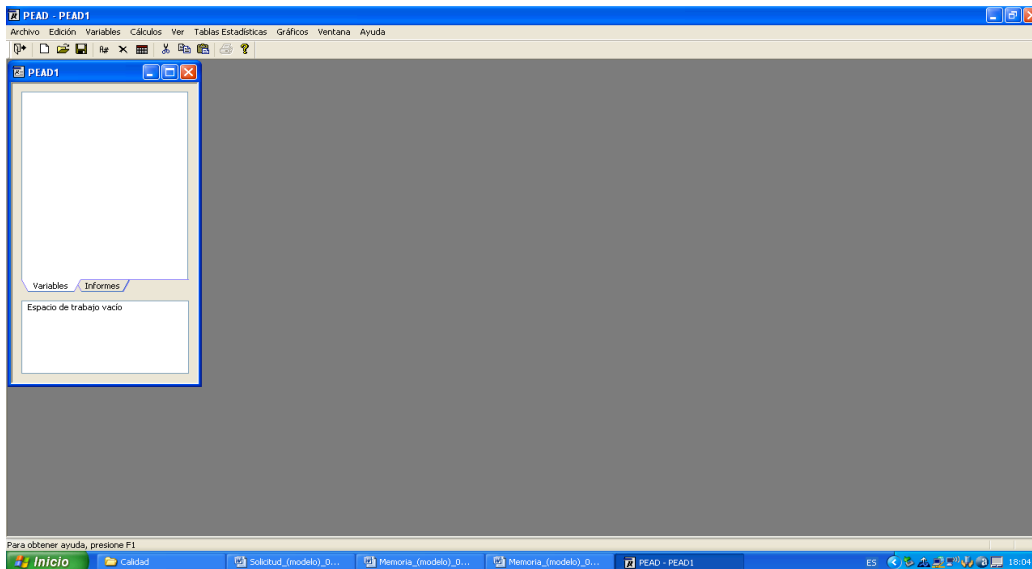
4. Materiales y métodos

La metodología seguida en el desarrollo de esta aplicación se enmarca dentro de la genérica en la producción de software, en donde en primer lugar se atiende el problema a desarrollar (potenciar la herramienta al posibilitar el uso de variables de clasificación), los requerimientos teóricos necesarios para poder llevar a cabo los algoritmos a implementar, el desarrollo de los esquemas de relación pertinentes para su codificación e integración dentro de la aplicación existente, su puesta en funcionamiento y finalmente los procesos de depuración de errores. En este último caso se trata de un procedimiento de retroalimentación que se seguirá hasta su completa validación.

En cuanto al material utilizado en cada fase puede consultarse la bibliografía descrita para el primer paso (componente teórico) y para la algoritmia y codificación. En cuanto a la implementación e integración, igual que en versiones anteriores se usó la aplicación Visual C++ .Net y los documentos y fuentes de las distintas versiones PEAD.

5. Resultados obtenidos y disponibilidad de uso

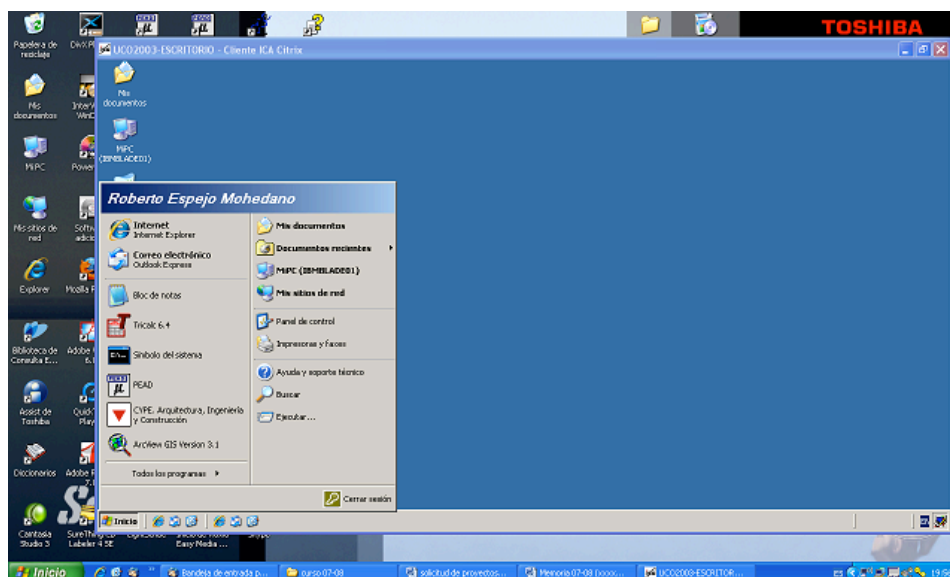
Como se ha comentado, el material elaborado es el programa estadístico PEAD (paquete estadístico de análisis de datos) puesto a disposición tanto de alumnos como profesores de esta Universidad.



Aspecto general de PEAD Cuadro de mensaje

La operatividad de software desarrollado consta de los módulos descritos en proyectos precedentes, y la posibilidad del uso de series temporales. Además, se ha actualizado el sistema de ayudas del programa.

La aplicación esta ya disponible para toda la comunidad universitaria a través de los servidores de la Uco o bien en casa por medio de las conexiones Ica.



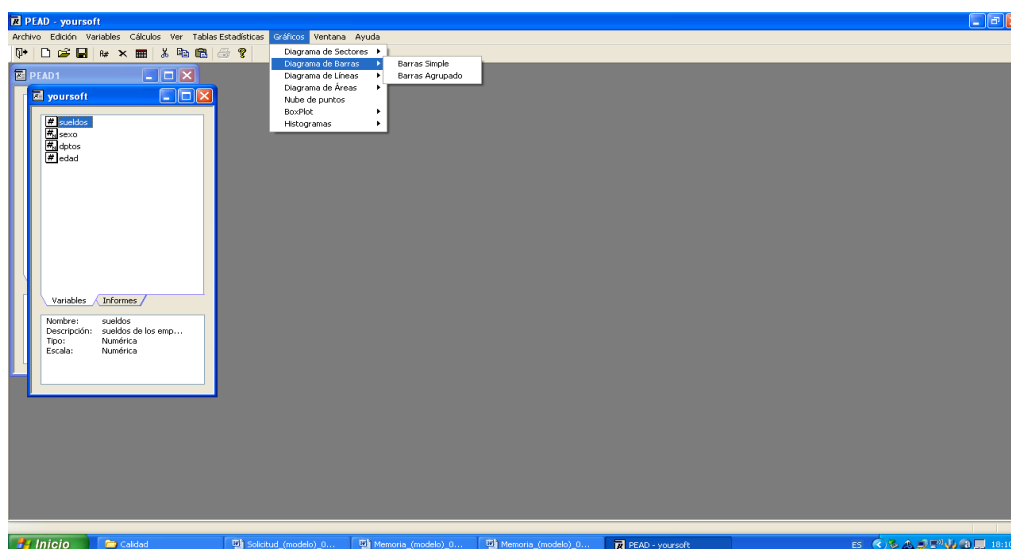
Disponibilidad de PEAD Acceso cliente Ica

Las aplicaciones de PEAD van desde la introducción asistida de datos (sensiblemente mejorada al incluir módulos de control sobre la introducción y datos, no permitiendo la misma si no se corresponden con las características declaradas sobre la naturaleza de los mismos, también se incluyen herramientas y utilidades para transformaciones de variables y manipulación de las mismas) hasta el análisis completo de los mismos a través del uso de los distintos módulos asociados tanto a los espacios de trabajo activos como la tipología de variables en uso.

Una vez introducidos los datos se está en disposición de poder aplicar las técnicas estadísticas correspondientes a las variables implicadas en el estudio. Así, puede realizarse en primer lugar un análisis estadístico descriptivo donde están disponibles las técnicas usuales de descripción uni y bivalente, unidos a los correspondientes gráficos y análisis exploratorio. Además de presentar los resultados de las medidas usuales, el programa

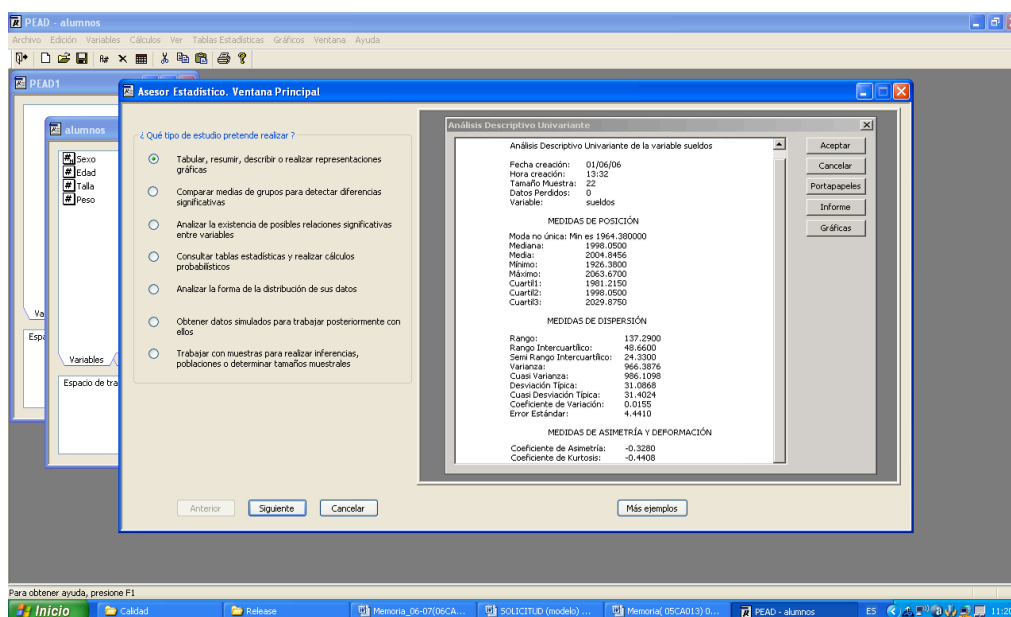
permite la elaboración de informes y el traspaso de información entre aplicaciones win32 usuales (Word, Excel, etc) mediante el portapapeles Windows, así como gráficas adecuadas a los datos.

A partir de este momento se puede proceder a un examen mas exhaustivo de los datos a través de los distintos módulos disponibles.



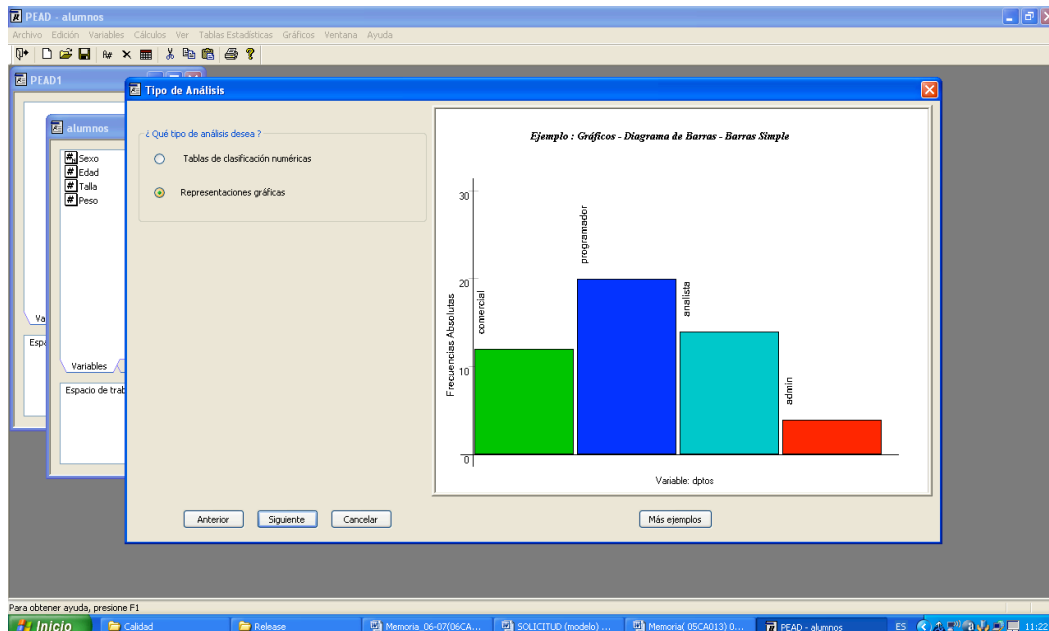
Aspecto general de PEAD Cuadro de mensaje

Como ya se describió en la última versión, se incorporó el módulo de asesoramiento estadístico y mejora del sistema de ayuda. En este sentido, presenta la misma operatividad en cuanto a la salida de resultados, portabilidad de los mismos, etc.



Aspecto general de PEAD Cuadro de resultados

El Asesor Estadístico ofrece al usuario la posibilidad de encontrar el procedimiento adecuado para analizar sus datos. Para ello se pedirá información a través de cuestiones sencillas, que apoyadas en ejemplos visuales del propio programa, permitan guiar al usuario hasta una posible solución a su problema.



Aspecto general de PEAD 7.2 Asesor estadístico

Una vez que el sistema dispone de toda la información necesaria para seleccionar la técnica estadística adecuada para los propósitos requeridos, se muestra (proporcionando una breve descripción de la misma) y se pone de manifiesto la forma de llevarla a cabo mediante la aplicación PEAD.

Aspecto general de PEAD Cuadro de recomendaciones

La nuevas funcionalidades de PEAD se corresponden con las marcadas en los objetivos inicialmente marcadas. La posibilidad del uso y análisis series de tiempo.

Introducir Datos de Serie Temporal

Intervalo

Desde las 0 del 17/06/2009

Hasta las 0 del 17/06/2009

De 1 en 1 Unidades

Temporización

Horas

Días

Meses

Años

Cancelar Aceptar

Aspecto general de PEAD Cuadro de diálogo

Datos de la Serie

Intervalo a Mostrar

Desde

Hora Día Mes Año

0 17 6 2009

Hasta

Hora Día Mes Año

5 17 6 2009

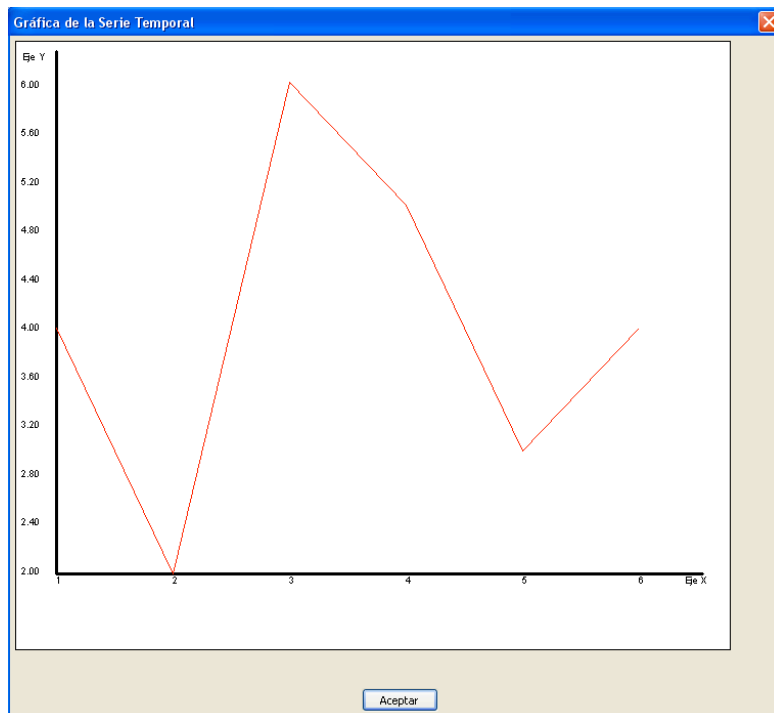
Mostrar Gráfica

Guardar

Cancelar Aceptar

Indice	Hora	Día	Mes	Año	Valor
1	0	17	6	2009	
2	1	17	6	2009	
3	2	17	6	2009	
4	3	17	6	2009	
5	4	17	6	2009	
6	5	17	6	2009	

Aspecto general de PEAD Cuadro de datos



Aspecto general de PEAD Cuadro de resultados

En definitiva el paquete de programas presentado cumple los propósitos inicialmente planteados y permite, de una forma rápida y sencilla, la aplicación de los métodos estadísticos vistos en clase en la mayoría de los programas de Estadística de esta Universidad, y sobre el que los alumnos podrán tener acceso no restringido para su uso.

6. Utilidad

El proyecto desarrollado no cabe duda que ha supuesto un esfuerzo considerable, pero a su vez ha servido de gran utilidad en cuanto al conocimiento y aplicación de los conceptos estadísticos como computacionales por parte de los componentes que forman el equipo, además de una estrecha colaboración entre los mismos que nos mueve a poder abordar nuevos proyectos conjuntos.

7. Observaciones y comentarios

También queremos destacar, a modo de satisfacción personal, los comentarios por parte de los alumnos sobre los que se ha testeado el programa, en cuanto a las ventajas de disponer de un software para realización de prácticas de libre disposición y que se adapta al programa de la asignatura. Como ya hemos comentado desde el principio, éste era el fin último de la experiencia, siempre mejorando y ampliando las versiones anteriores.

8. Autoevaluación de la experiencia

Los resultados obtenidos son satisfactorios en el sentido expuesto en los apartados anteriores. No obstante, debemos insistir en el esfuerzo llevado a cabo (desde el punto de vista de programación informática y tiempo dedicado) que ha ocasionado el desarrollo del presente trabajo.

Hay que comentar que, aún cuando el contenido operativo del software desarrollado se corresponde con la mayoría de los programas generales de Estadística de esta Universidad, y que versión tras versión se han introducido mejoras notables y nuevos módulos de tratamiento estadístico de la información, todavía existen algunos campos o módulos que pueden y deben ser incorporadas a PEAD para cubrir algunos aspectos concretos de la Estadística en algunas especialidades y asignaturas optativas del departamento.

9. Bibliografía

- **[BEC99]** Zaratian, B.; **Microsoft Visual C++ 6.0 Manual del Programador**, "Microsoft Visual C++ 6.0 Programmer's Guide", Traductores: M^a Pilar Quijada Arteaga, José Ignacio Sánchez García. Revisión técnica: Antonio Vaquero Sánchez. Editorial McGraw-Hill, 1999.
- **[BOO00]** Booch, G.; Rumbaugh, J.; Jacobson, I.; **El proceso unificado de desarrollo de software**. Pearson Educación. Madrid, 2000. 464 p. ISBN: 84-7829-036-2.
- **[ESP03]** Espejo Mohedano, A. R; Gallego Segador, A; **Fundamentos de Estadística: Teoría**. Universidad de Córdoba, 2006. 362 p. ISBN 84-689-1627-7
- **[SHE03]** Shepherd, G.; Kruglinski, D.; **Programación avanzada con Microsoft Visual C++ .Net**;, Editorial Mc-Graw Hill, 2003, ISBN: 84-481-3724-8.

Lugar y fecha de la redacción de esta memoria

Córdoba, 30 de Septiembre de 2010.