

Creación de una interfaz de R para la enseñanza de la Estadística

Hernández Martín, Z.¹, San Martín Pérez, M.², Guzmán Delgado, H. U.³

¹zenaida.hernandez@unirioja.es, Universidad de La Rioja

²montse.sanmartin@unirioja.es, Universidad de La Rioja

³unai86@gmail.com, La Rioja

Resumen

Uno de los principales objetivos que nos hemos marcado en la enseñanza de la Estadística, es conseguir trabajar con software libre de manera habitual, pero el manejo de este software suele resultar muy complicado para los estudiantes. Con independencia de otros proyectos, como R UCA, que ya están en marcha, hemos querido desarrollar una interfaz amigable de R para la enseñanza de la Estadística, adaptada a los contenidos que se enseñan en los Grados impartidos en esta Universidad. Otros interfaces como R-Commander, incluyen análisis avanzados en los que los estudiantes se pierden y nuestro objetivo es que los estudiantes (sobre todo los de titulaciones de «letras»), puedan realizar con facilidad análisis estadísticos completos y correctos, manejando un software libre y en un entorno que les resulte «cómodo». Presentamos el desarrollo que ha realizado, como proyecto fin de carrera, un estudiante de la doble titulación Matemáticas e Ingeniería Técnica en Informática de Gestión.

Palabras clave: Interfaz de R

Clasificación AMS: 62

1. Introducción

Cada día vemos cómo las aplicaciones informáticas se van haciendo imprescindibles en casi todos los campos.

La Universidad, como institución de enseñanza, no sólo no es ajena a este hecho sino que contribuye a la difusión y manejo de estas herramientas que generalmente nos facilitan el trabajo y nos permiten una mejor y más rápida comunicación y acceso a la información.

En este sentido los docentes hemos ido introduciendo, en la mayoría de las asignaturas que impartimos, distintos programas informáticos que facilitan las tareas propias de cada materia, desde el manejo de un software de escritura o una hoja de cálculo, a un programa de tratamiento de imágenes, de geografía o de música.

Prácticamente en todas las materias impartidas actualmente en un campus universitario, se utiliza algún software informático.

Esto ha supuesto, en primer lugar la concienciación, aprendizaje y esfuerzo por parte de los docentes y después las modificaciones en los planes docentes para adaptar los contenidos y utilizar de forma eficaz los nuevos recursos.

Por otra parte, también los estudiantes han ido viendo cómo cambiaba la docencia recibida y han ido adaptando su aprendizaje de forma progresiva a las nuevas tecnologías y en particular a la gran cantidad de programas informáticos que deben utilizar a lo largo de sus estudios

En este momento podemos pensar que la primera fase, la de concienciar y «alfabetizar» a nuestros estudiantes en el uso de herramientas informáticas adecuadas a su formación específica, está prácticamente concluida.

Sin embargo, esta apuesta es muy cara para las universidades ya que actualmente se utiliza, para la docencia, mucho software comercial; por eso en las universidades se está empezando a concienciar al profesorado sobre este tema y se ha empezado a promover el uso de software gratuito y/o libre.

¿Por qué no usamos más software libre?

En la actualidad existen versiones libres, o al menos gratuitas, de prácticamente todo el software que utilizamos. Estas versiones tienen, en general, las mismas prestaciones, calidad y fiabilidad que las comerciales, entonces ¿por qué decidimos pagar por las aplicaciones que usamos?

Hay fundamentalmente dos razones por las que la gente prefiere utilizar software comercial: la facilidad de uso (normalmente entornos más amigables) y asistencia técnica.

La asistencia técnica, el poder recurrir a alguien que te solucione los problemas que se te puedan presentar, es algo que se valora mucho aunque, cada vez más, sobre todo los jóvenes son más autosuficientes y son más partidarios de entrar en foros especializados para solucionar sus problemas.

De todas formas, sigue pesando mucho el hecho de que ni todos los problemas se pueden solucionar en los foros ni todo el mundo es capaz de solucionar problemas técnicos solamente con las indicaciones desinteresadas de los colaboradores de los foros. Y además, muchos de estos foros están en otro idioma lo que supone una dificultad añadida.

2. Nuestra apuesta para la enseñanza de la Estadística

Los profesores de Estadística de la Universidad de La Rioja no hemos sido ajenos a toda esta situación y de hecho, aunque llevamos muchos años utilizando software estadístico en nuestra docencia (SYSTAT, SPSS, Statistica) todo este software es comercial y hasta hace 2 años no hemos empezado a utilizar software libre (R) y sólo en algunas asignaturas (en las demás estamos usando Statistica, actualmente).

Al principio la evolución fue impuesta, tanto SYSTAT como SPSS eran los programas estadísticos para los que había licencia de campus. La universidad decidió, supongo que por solicitudes de los investigadores principalmente, la adquisición de este software y como era el disponible y además era muy bueno, pues es el que se utilizaba en la docencia.

Sin embargo, todos los años teníamos problemas con las renovaciones de las licencias (en medio del curso). Por otra parte, los estudiantes sólo podían utilizar este software en

la Universidad y había problemas con la ocupación de las aulas informáticas. Por ello, hace unos años decidimos solicitar la compra de Statistica; este software no es ni mejor ni peor que otros, pero tiene la ventaja, para la docencia, de que tiene una buena versión de estudiantes (gratuita), que les permite estudiar y practicar en sus casas y a su ritmo.

Aún así, seguimos con el problema de las licencias. Y por otra parte, somos conscientes de que este software es caro y que hay software libre y/o gratuito que nos permitiría hacer lo mismo. Por ello decidimos hacer alguna prueba y empezar a usar R.

Hace dos años comenzamos a usar R en una asignatura de Estadística de las titulaciones de Matemáticas e Ingeniería Técnica en Informática de Gestión. Se ha utilizado RCommander para facilitar el manejo y la aceptación de los estudiantes ha sido bastante buena. Sin embargo, teníamos muchas dudas sobre la posibilidad de usar este mismo software en otras titulaciones no tan familiarizadas con la informática o las matemáticas. A pesar de las facilidades proporcionadas por RCommander, teníamos la sensación de que muchos estudiantes «se nos iban a perder».

En este momento, estamos impartiendo asignaturas de Estadística en los Grados de: Matemáticas, Ingeniería Informática, Química, Enología, Ingeniería Agrícola, Administración y Dirección de Empresas, Turismo, Relaciones Laborales y Trabajo Social, por lo que tenemos alumnos de todo tipo y, por lo dicho anteriormente, la apuesta por el uso de R en todas las titulaciones no estaba nada clara.

Sabemos que en la mayoría de las titulaciones, las asignaturas de Estadística son «difíciles» para los estudiantes y, en la medida de nuestras posibilidades, procuramos acercarlos a la Estadística y hacérsela comprensible y amena. Por lo tanto nuestro objetivo era que el software estadístico debería servir para facilitarles el trabajo y no suponerles un problema añadido al que ya tienen con la propia asignatura, que es una de las que mayor esfuerzo les requiere.

Por lo tanto, salvo para los alumnos de Matemáticas e Informática, para poder usar R en nuestras clases debíamos conseguir un entorno «mucho más amigable».

3. La idea del proyecto

Mientras le dábamos vueltas a este tema, un estudiante de la doble titulación Matemáticas e Ingeniería Técnica en Informática de Gestión nos sugirió que quería hacer el proyecto de fin de carrera (para ITIG) en algo relacionado con la Estadística y así fue como decidimos proponer como proyecto de fin de carrera la «Creación de una interfaz amigable para la docencia con R», al estudiante Héctor Unai Guzmán Delgado.

El proyecto consiste en la creación de una interfaz amigable para la docencia con R. Esto es muy importante. Nuestro objetivo principal no es poder utilizar siempre esta interfaz para el uso de R, sino poder usar R en las aulas sin «asustar» a nuestros estudiantes.

Esta interfaz se va a usar sólo para los contenidos estadísticos que manejamos en los Grados, por lo que los menús no irán más allá.

Por último, la interfaz se ha creado para trabajar en Windows, ya que este es el sistema operativo que está instalado en la mayoría de las aulas del campus en la Universidad de

La Rioja.

Evidentemente, si esto funciona, la idea es extenderlo a otros sistemas.

Utilización en los distintos Grados

El uso de la interfaz es muy sencillo, tendremos abierto R y usaremos la interfaz para «escribir» las órdenes, que se envían a R y es en su consola en la que aparecerán los resultados.

Esto nos permitirá un aprovechamiento diferente según las titulaciones:

- En las titulaciones más cercanas a la informática como el Grado en Matemáticas o el Grado en Ingeniería Informática, la interfaz nos servirá para introducir las órdenes, pero además, como tenemos a la vista la consola de R, podemos introducir modificaciones y los estudiantes no sólo la usará para obtener resultados estadísticos sino que aprenderán a usar R, como venían haciendo hasta ahora.
- En el resto de las titulaciones, el principal objetivo no es aprender a usar R sino obtener e interpretar los resultados del análisis estadístico. En este sentido, vamos a usar R para obtener nuestros resultados, pero sin tener que introducir las órdenes sino usando unos menús muy sencillos.

Como los alumnos tienen a la vista la consola de R, pueden aprender a utilizarlo, igual que en el caso anterior, pero este no será nuestro principal objetivo.

4. Resultado del proyecto

El resultado del proyecto ha sido una interfaz llamada **Interactúa R** [2].

A continuación voy a presentar los distintos menús y características de los análisis que se pueden realizar usando esta interfaz.

A la vez iremos viendo cómo debemos trabajar en R.

Abrimos la interfaz y nos aparece la siguiente ventana:



Figura 1: Pantalla inicial

Podemos comprobar la autoría y la licencia de esta aplicación en:

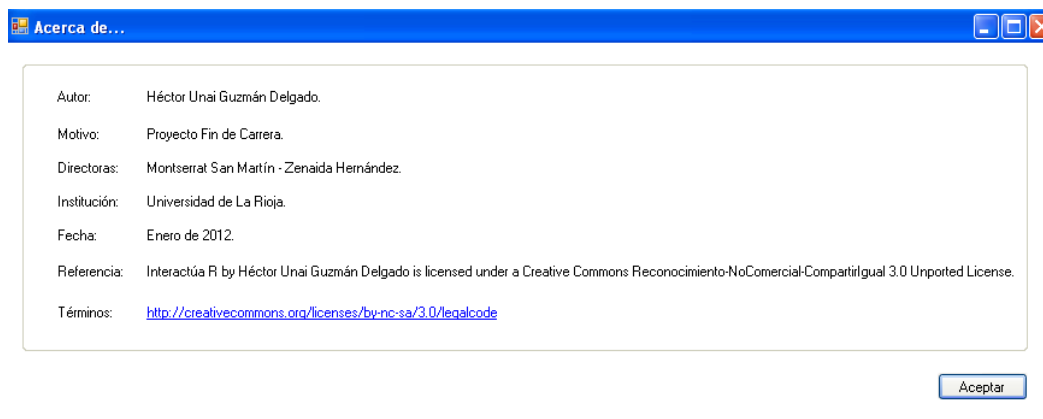


Figura 2: Acerca de

Para poder trabajar, tenemos que abrir, si no lo hemos hecho antes, la consola de R. A continuación se irán viendo y comentando los distintos menús.

4.1. El menú Archivo

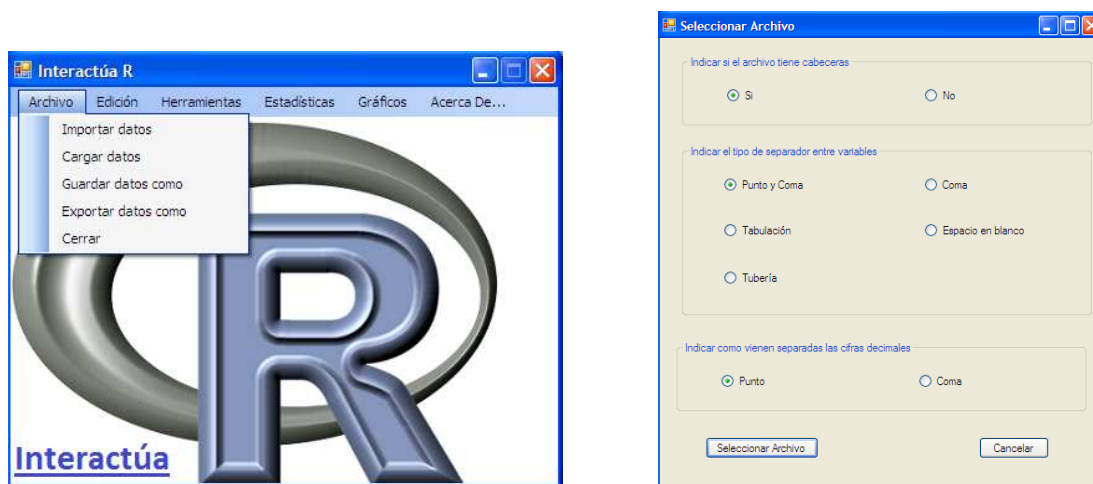


Figura 3: Menú Archivo y ventana Importar datos

En este menú se encuentran las operaciones básicas para importar o cargar datos, guardarlos, exportarlos o cerrar la aplicación.

Nota: cuando se importa un archivo, éste debe ser un archivo de texto (.txt) o de datos separados por comas (.csv). Y del mismo modo, la exportación de datos se hará a uno de estos dos formatos.

Antes de empezar a hacer cálculos estadísticos, se cargarán automáticamente los paquetes necesarios:

```

>
> ##Cargar paquetes necesarios
> library(PASWR)
Loading required package: e1071
Loading required package: class
Loading required package: MASS
Loading required package: lattice
> library(psych)
> library(nortest)
> library(car)
Loading required package: nnet
Loading required package: survival
Loading required package: splines

```

Figura 4: Carga automática de paquetes (vista de la consola de R [1])

4.2. El menú Edición



Figura 5: Menú Edición

Nos permitirá hacer pequeñas modificaciones en el fichero de datos que estemos manejando.

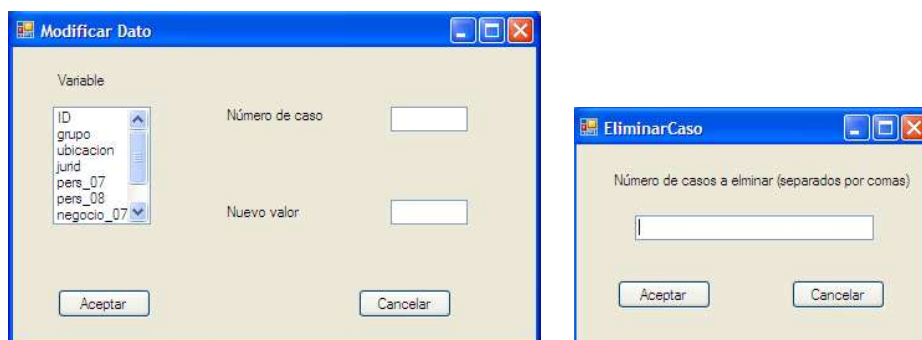


Figura 6: Ventanas del Menú Edición-1

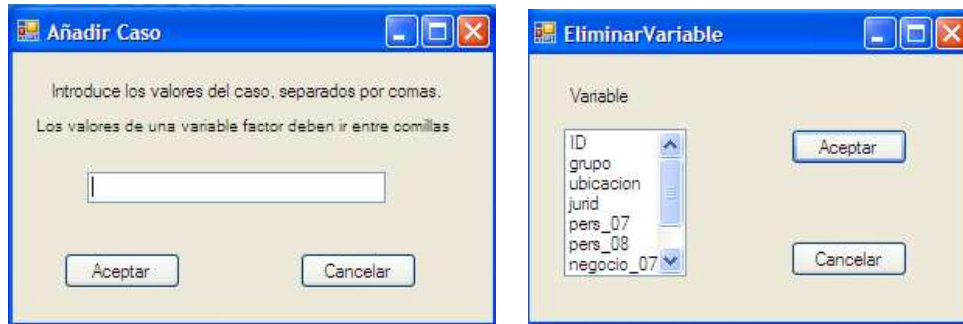


Figura 7: Ventanas del Menú Edición-2



Figura 8: Ventanas del Menú Edición-3

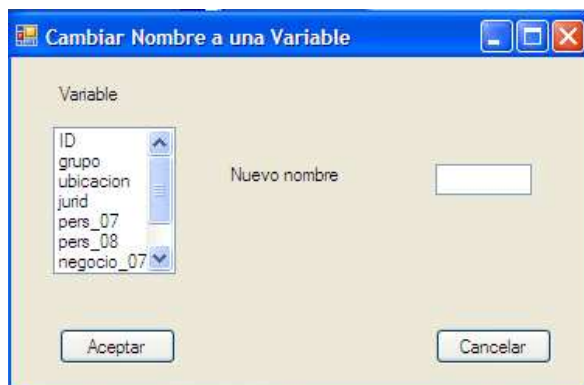


Figura 9: Ventanas del Menú Edición-4

4.3. El menú Herramientas

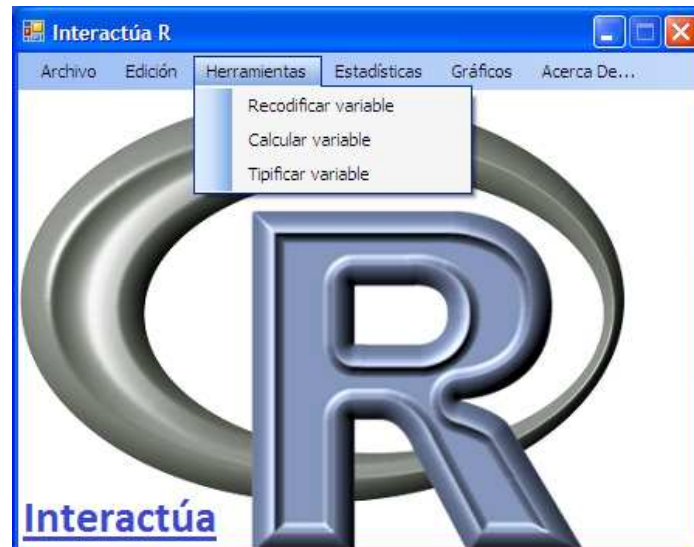


Figura 10: Menú Herramientas

Las únicas herramientas que hemos considerado para esta aplicación son las de Recodificar, Calcular y Tipificar.

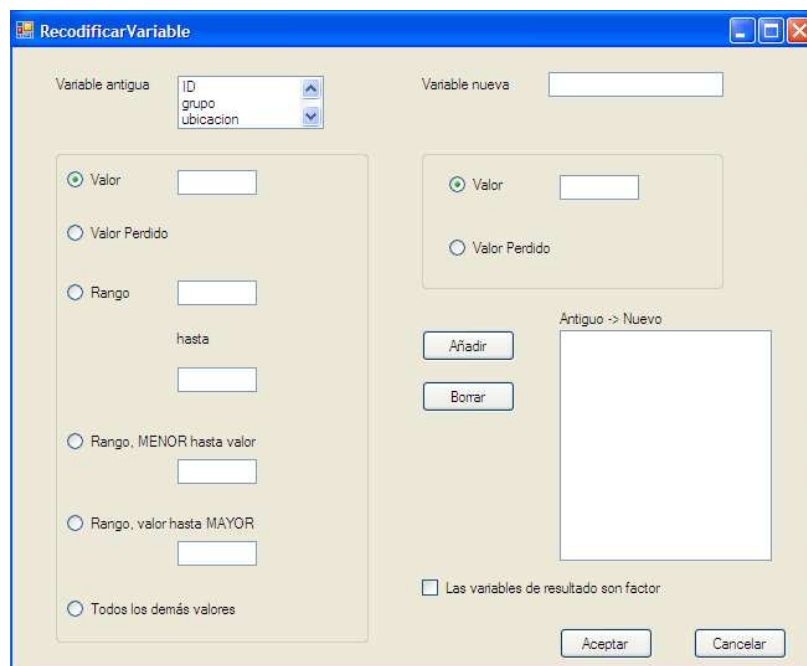


Figura 11: Recodificar

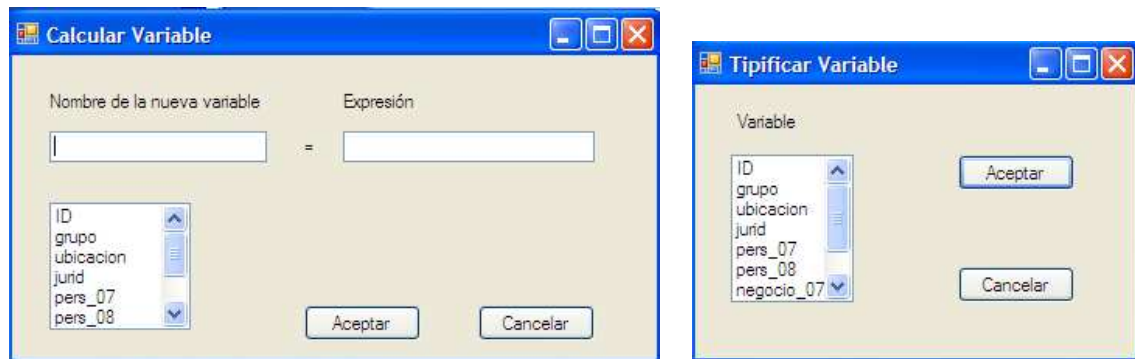


Figura 12: Ventanas de Calcular y Tipificar

A continuación se presentan los menús de cálculo de estadísticas y gráficos, adaptados a la docencia que se imparte en los Grados.

4.4. El menú de Estadísticas



Figura 13: Menú Estadísticas

Dentro de este menú se encuentran seis submenús, cada uno de ellos con sus correspondientes opciones.

Vamos a ir comentándolas una a una:

4.4.1. Unidimensionales



Figura 14: Unidimensionales

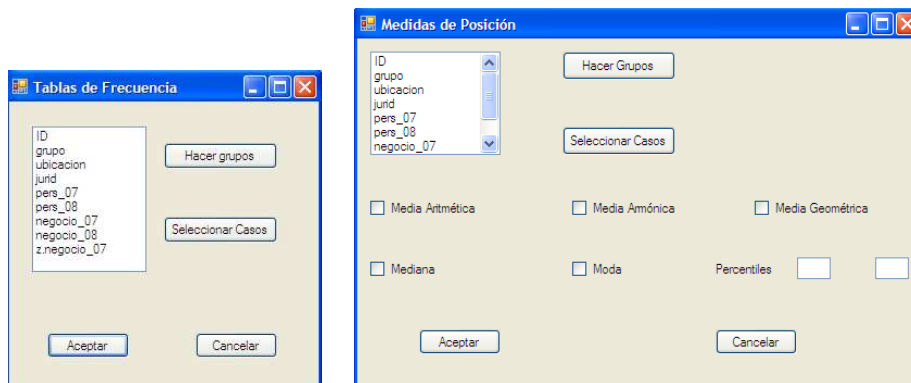


Figura 15: Tablas de frecuencias y medidas de posición

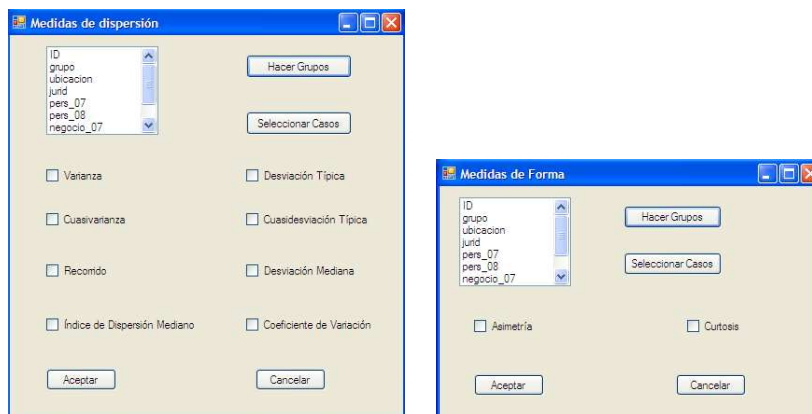


Figura 16: Medidas de dispersión y de forma

Como se puede ver en los menús de estos análisis, se contemplan también las opciones de seleccionar algunos casos para el análisis, así como la de realizar un análisis por grupos (repetiendo el mismo análisis en cada uno de los grupos que determinan los valores de una variable)



Figura 17: Seleccionar casos y Hacer grupos

También se ha incluido el análisis de la concentración en el reparto de los valores de una variable:

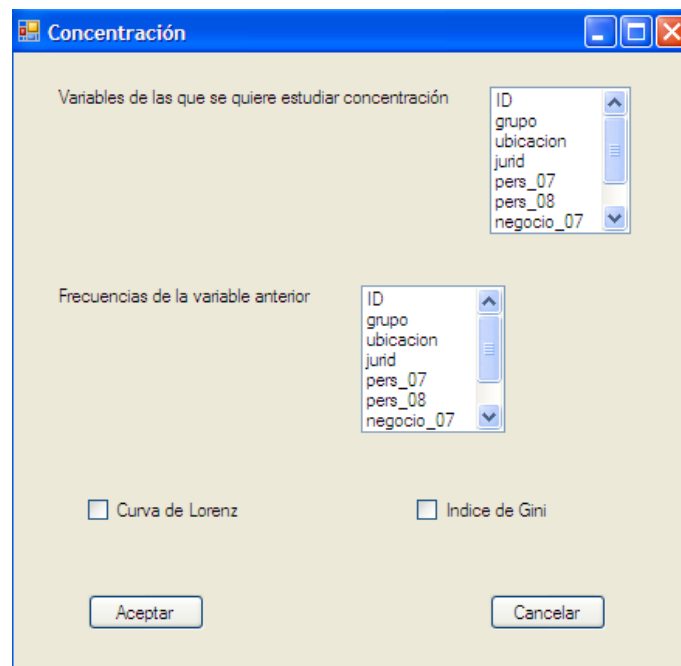


Figura 18: Concentración

4.4.2. Bidimensionales

Para el análisis de las variables bidimensionales sólo se considera el siguiente cuadro que incluye las Medidas bidimensionales más habituales:

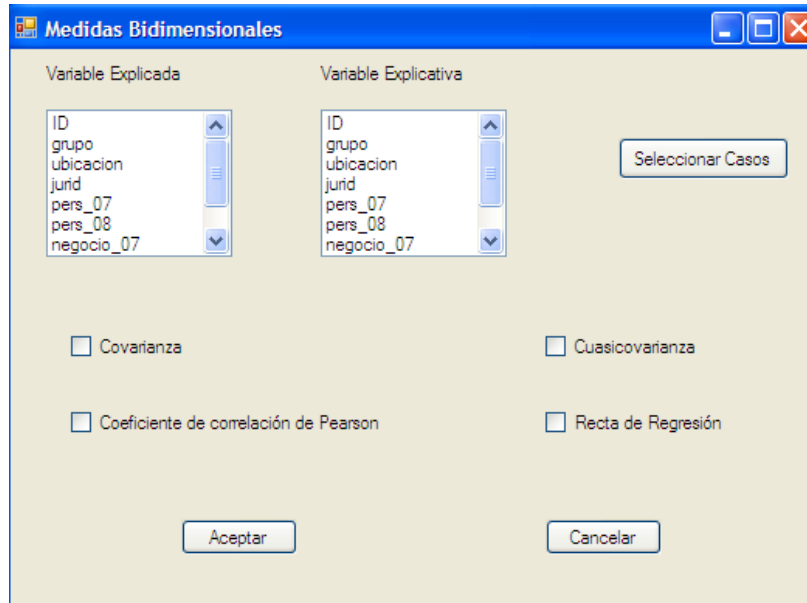


Figura 19: Bidimensionales

4.4.3. Distribuciones de probabilidad

En cuanto a las distribuciones de probabilidad tenemos el siguiente menú:

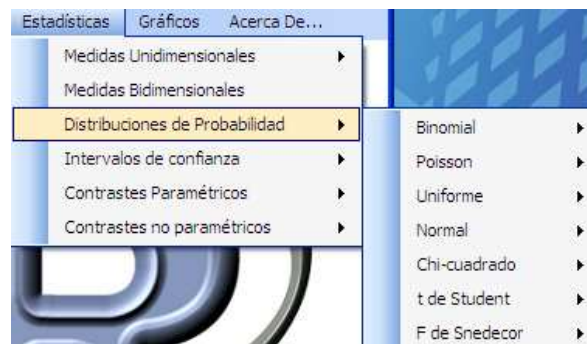


Figura 20: Distribuciones de probabilidad

y para cada una de las distribuciones tenemos distintas opciones, dependiendo además de que estemos trabajando con distribuciones de variables discretas o continuas:

Para las discretas:

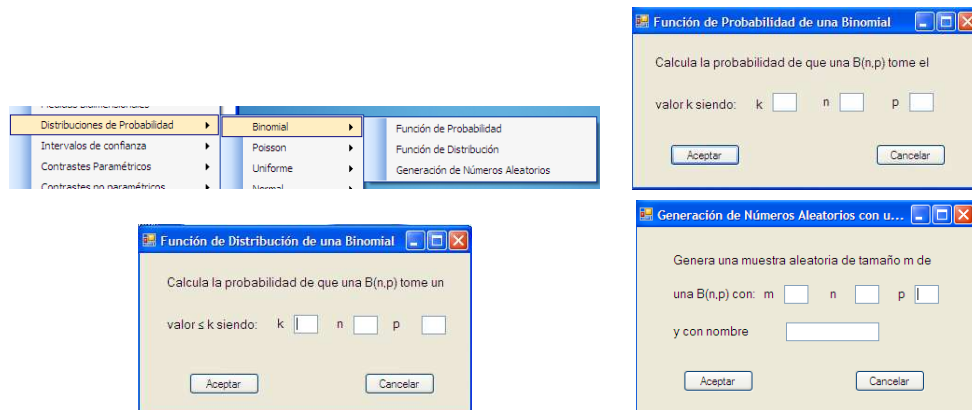


Figura 21: Opciones para variables aleatorias discretas

Para las continuas, además de la Función de distribución y la generación de valores aleatorios, tenemos la función inversa de probabilidad:

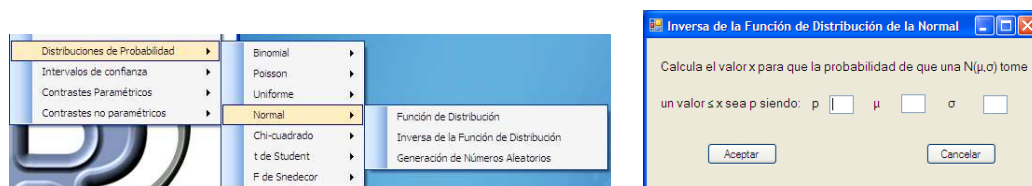


Figura 22: Opciones para variables aleatorias continuas

En el análisis inferencial se incluyen análisis que otros programas no realizan. En este sentido, esta es una adaptación específica para la docencia.

4.4.4. Intervalos de confianza

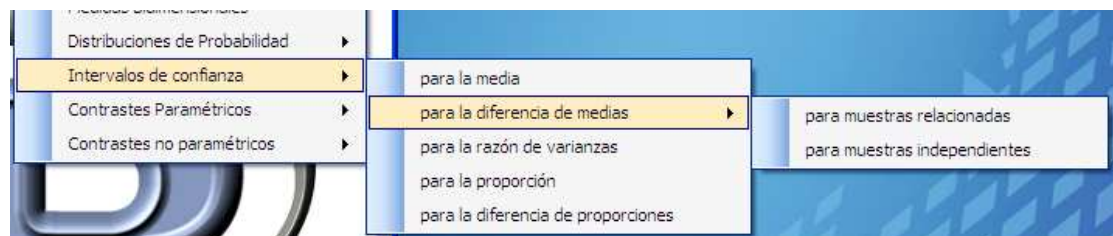


Figura 23: Intervalos de confianza

Los menús correspondientes son:

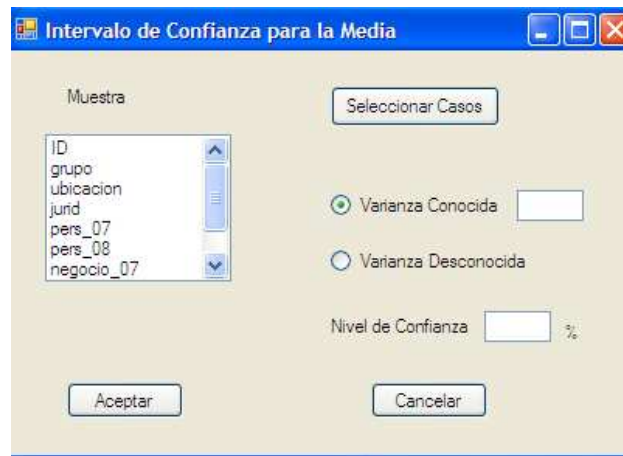


Figura 24: Intervalo de confianza para una media

En este caso se puede observar que se ha incluido también el intervalo para el caso de varianza conocida (que no suelen contemplar los programas estadísticos habituales).

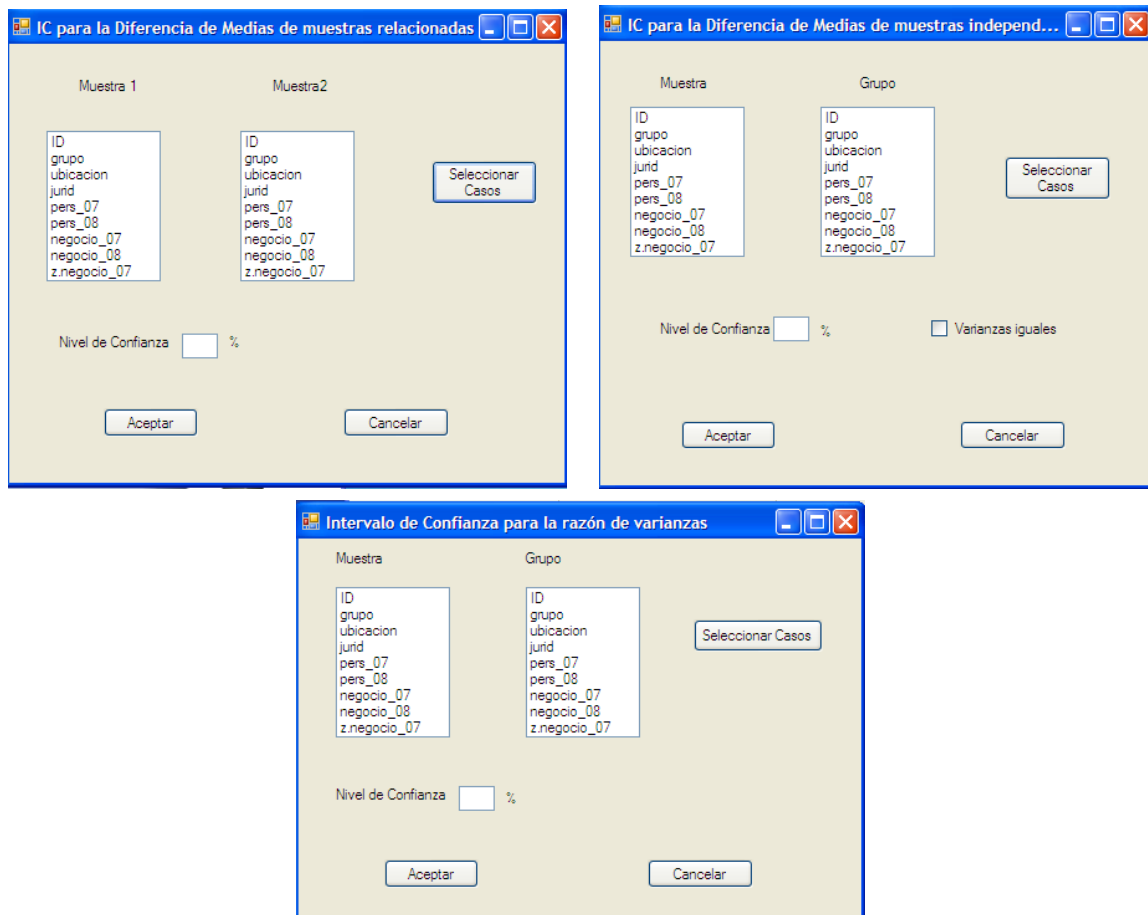


Figura 25: Intervalos de confianza para la igualdad de medias y de varianzas

También se han incluido los intervalos de confianza para una proporción y para la diferencia de proporciones:

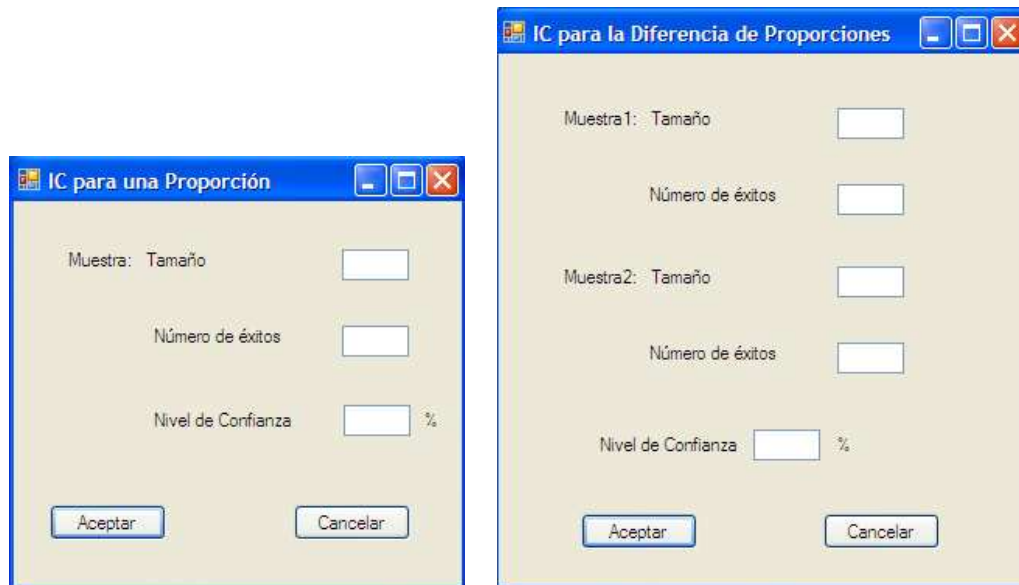


Figura 26: Intervalos de confianza para la proporción y la diferencia de proporciones

En cuanto a los contrastes de hipótesis, se realizan los contrastes paramétricos y no paramétricos más elementales.

4.4.5. Contrastes paramétricos

Igual que para los intervalos de confianza, tenemos:

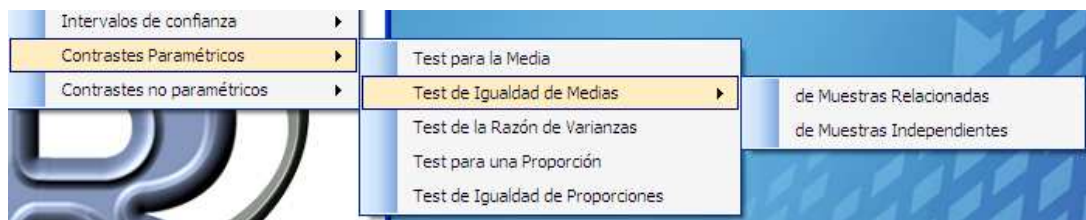


Figura 27: Los distintos contrastes paramétricos

En el caso de los contrastes, además de considerar la posibilidad de varianza poblacional conocida en el caso de la media, también se contempla en todos ellos la opción de los contrastes unilaterales.

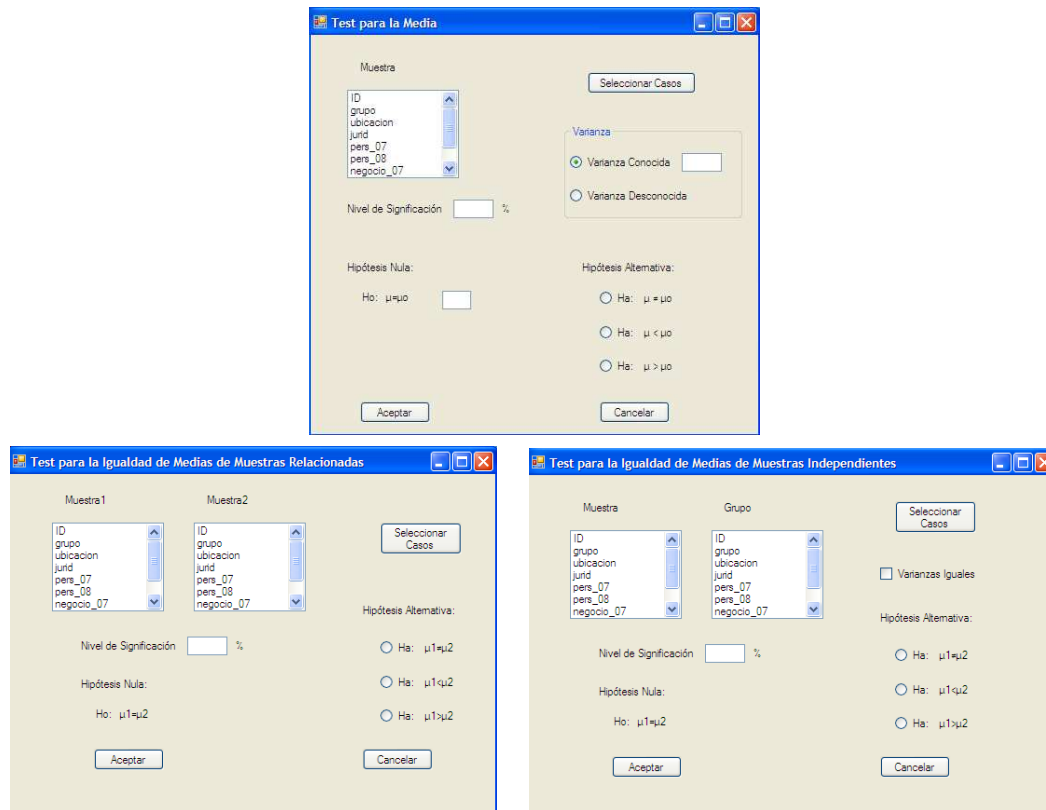


Figura 28: Contrastes de hipótesis para la media y la diferencia de medias

Y los restantes contrastes:

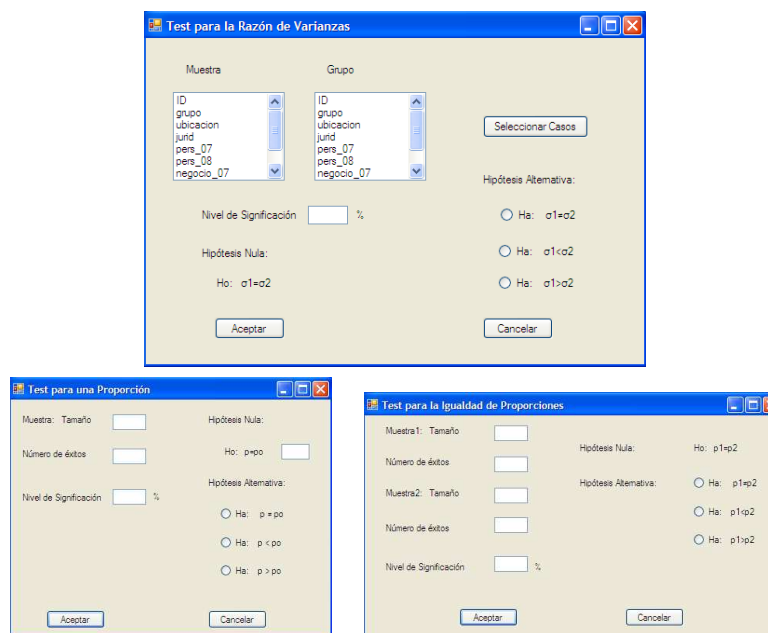


Figura 29: Contrastes de hipótesis para la razón de varianzas, la proporción y la diferencia de proporciones

4.4.6. Contrastes no paramétricos

Para los contrastes no paramétricos, se han considerado las siguientes opciones:

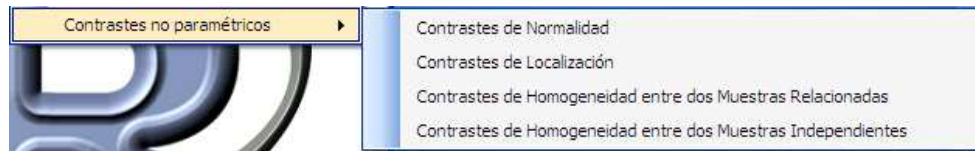


Figura 30: Los distintos contrastes no paramétricos

Cuyos menús son los que se muestran a continuación:

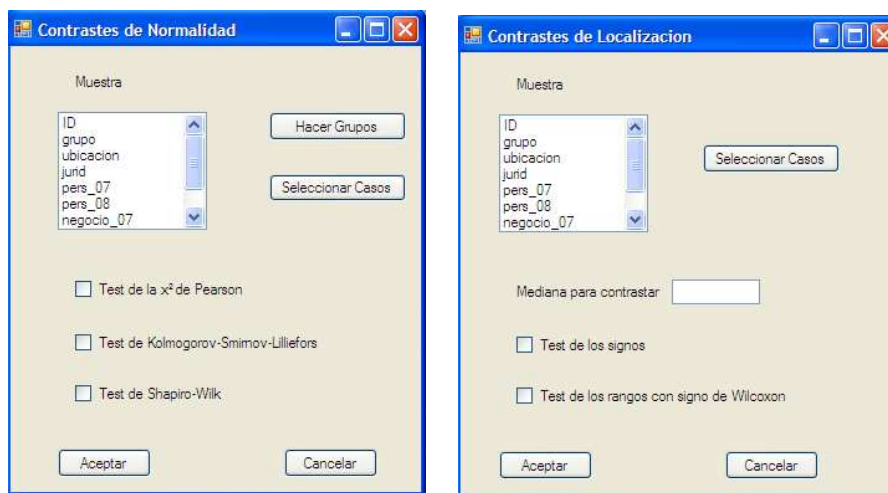


Figura 31: Contrastes de Normalidad y Localización

Y los contrastes de Homogeneidad:

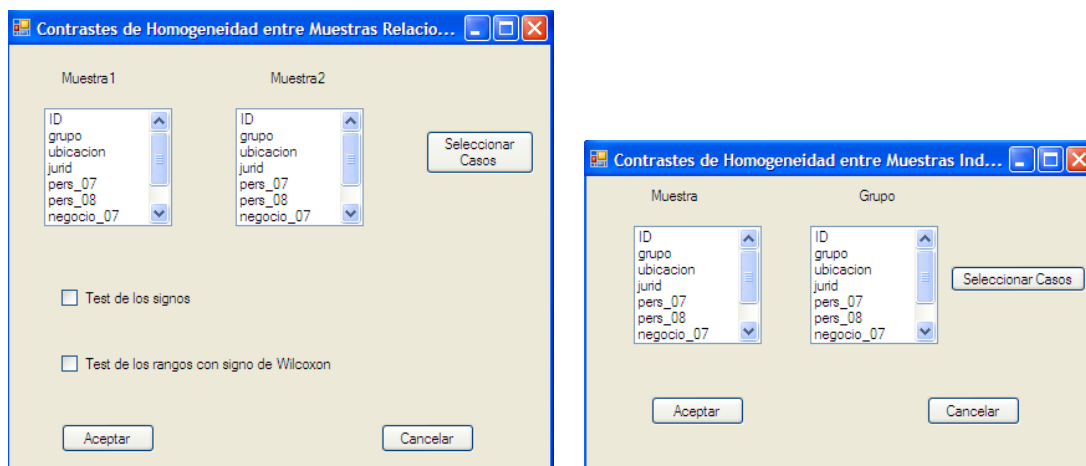


Figura 32: Contrastes de Homogeneidad

4.5. Gráficos

Por último, se consideran los gráficos más elementales así como las funciones de densidad de las distribuciones utilizadas en los distintos Grados.



Figura 33: Menú Gráficos

Los menús correspondientes a gráficos de variables unidimensionales incluyen, al igual que el cálculo de las distintas medidas, las opciones de seleccionar casos y repetir el gráfico en todos los grupos que determinan los valores de una variable.

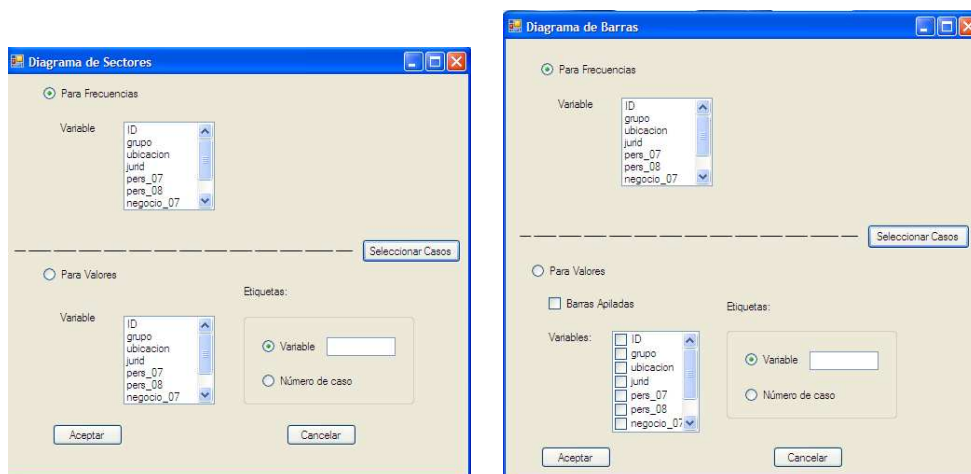


Figura 34: Gráficos de Sectores y Barras

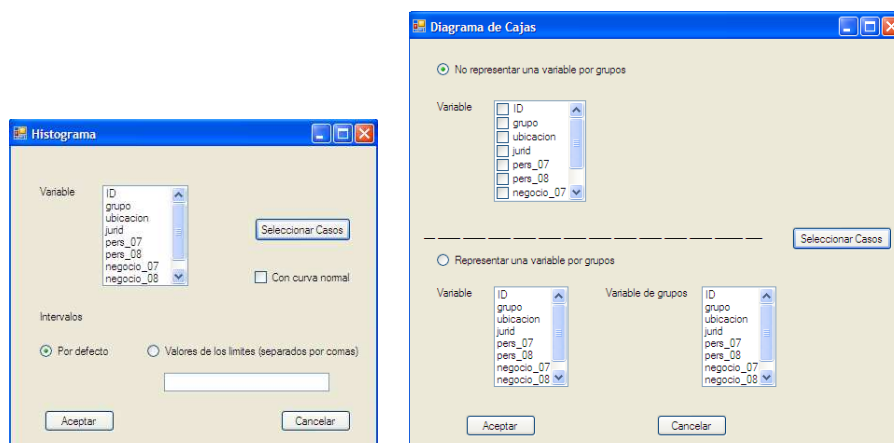


Figura 35: Histograma y gráfico de Cajas

En el caso de los histogramas, estos representan la densidad de frecuencia.

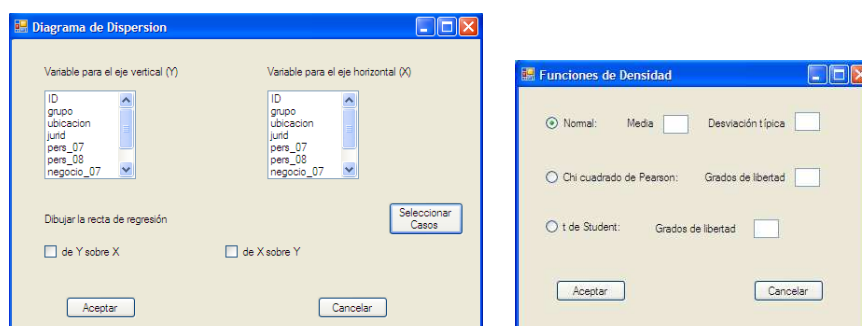


Figura 36: Gráficos de dispersión y Funciones de densidad

5. Conclusiones

Hasta aquí hemos presentado los menús de la nueva interfaz.

Con cada uno de estos menús, enviamos a R las instrucciones adecuadas para el análisis, y obtenemos la respuesta en la consola de R.

Con esto pretendemos alcanzar un doble objetivo, dependiendo de la titulación en la que se utilice:

1. Hacer más sencillo el uso de R, para la obtención de resultados estadísticos en las titulaciones menos familiarizadas con este tipo de software.
2. Permitir el aprendizaje de este lenguaje, a medida que se va viendo y utilizando en el aula, para aquellos estudiantes más familiarizados con este tipo de lenguajes y a los que no les supone un gran inconveniente.

Nuestro propósito, como ya se ha comentado, es empezar a utilizarlo en todos los Grados en los que se imparten asignaturas de Estadística, el próximo curso 2012-2013.

6. Bibliografía

- [1] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- [2] Interactúa R by Héctor Unai Guzmán Delgado is licensed under a Creative Commons Reconocimiento-NoComercial-CompartirIgual 3.0 Unported License. (2012), URL <http://www.unirioja.es/cu/zehernan>