

Epidat 4.1: una herramienta para la enseñanza de la Estadística

María Isolina Santiago Pérez; Gael Naveira Barbeito; Xurxo Hervada Vidal en nombre del grupo Epidat*

email: soly.santiago.perez@sergas.es; estadistica.sp@sergas.es;
xurxo.hervada.vidal@sergas.es

Dirección Xeral de Innovación e Xestión da Saúde Pública. Xunta de Galicia

*Grupo Epidat: Luis Carlos Silva Ayçaguer; Enrique Vázquez Fernández; Óscar Mujica Lengua; Rubén Manrique Hernández; Jorge Bacallao Gallestey.

RESUMEN

Epidat es un programa de libre distribución desarrollado por la Dirección Xeral de Innovación e Xestión da Saúde Pública (Xunta de Galicia) que incluye, entre otros, varios módulos de estadística: análisis descriptivo, muestreo, inferencia sobre parámetros y distribuciones de probabilidad. Sus características y facilidad de manejo lo convierten en una herramienta muy útil para apoyar la enseñanza de la estadística. En este taller se pretende mostrar el funcionamiento general de Epidat 4.1 y de los módulos enumerados mediante una práctica basada en datos reales recopilados por alumnos de ESO.

Programa informático, Estadística, Análisis descriptivo, Muestreo, Inferencia, Distribuciones de probabilidad

Introducción

Actualmente, la enseñanza del análisis de datos debe complementarse con el uso de algún tipo de software estadístico, pues la posibilidad de trabajar con conjuntos grandes de datos reales, resumir la información de estos datos, o visualizar gráficamente conceptos teóricos favorece el aprendizaje de los alumnos. En este trabajo se describen las potencialidades, para ese objetivo, de un programa llamado Epidat que está desarrollado por la Dirección Xeral de Innovación e Xestión da Saúde Pública de la Xunta de Galicia y que está orientado al análisis estadístico y epidemiológico de datos. También colaboran en su desarrollo la Organización Panamericana de la Salud (OPS-OMS) y la Universidad CES de Medellín, Colombia.

A lo largo de sus cuatro versiones, desarrolladas durante veinte años, Epidat ha ido evolucionando en cuanto al lenguaje de programación, sistema operativo, contenido, prestaciones y enfoque de la ayuda [1,2]. La última versión, 4.1, posee una serie de características que lo convierten en una herramienta muy adecuada para la enseñanza de la estadística en secundaria y para la investigación en educación:

- 1- Es de libre distribución y está disponible en español.
- 2- No necesita instalación y funciona en distintos sistemas operativos.
- 3- Tiene un entorno sencillo e intuitivo, que facilita el manejo del programa.
- 4- Aunque el programa nació con orientación epidemiológica, la variedad de métodos que incluye actualmente permite su uso en distintos contextos.

El programa se puede descargar libremente de la página web de la Dirección Xeral de Innovación e Xestión da Saúde Pública [3]: <http://dxsp.sergas.es> (apartado de Aplicaciones).

En este trabajo se describen las características de Epidat 4.1 y las opciones que se utilizarán en el taller del programa.

Características de Epidat 4.1

La versión 4 de Epidat está desarrollada en Java, lo que permite su funcionamiento en distintos sistemas operativos, tales como Windows, Linux y OS X, entre otros. Además, el programa no necesita instalación, solo es necesario tener instalada en el ordenador una versión actual de Java. En la página web de descarga del programa se incluyen las instrucciones necesarias para ejecutar Epidat correctamente.

En cuanto a sus características, en Epidat 4.1 se pueden destacar tres elementos: el entorno general del programa, el contenido, es decir, los métodos que incluye, y la ayuda.

Entorno general

El entorno de trabajo de Epidat es una ventana, con una barra de menús y otra de herramientas, y que está dividida en tres partes: *Resultados*, donde se muestran todos los resultados obtenidos durante una sesión de trabajo, *Índice de cálculos* y *Módulos en uso*, que muestra y permite acceder a los módulos abiertos (Figura 1).

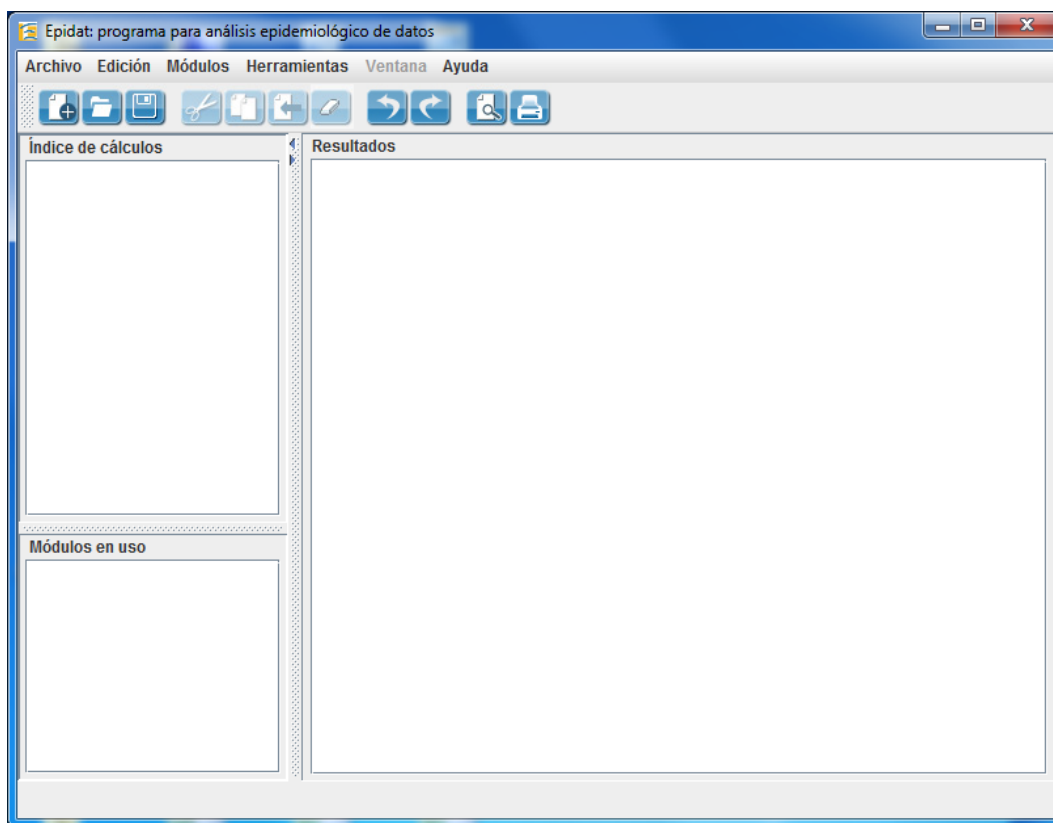


Figura 1.- Ventana principal de Epidat 4.1.

El menú *Módulos* contiene todos los módulos temáticos del programa, que en la versión 4.1 son 11, y que está previsto aumentar hasta 19. Cada uno de los 11 módulos actuales tiene submódulos que en conjunto suponen 105 ventanas diferentes. En el apartado sobre el contenido se describen los módulos que incluyen los métodos estadísticos con los que se va a trabajar en el taller del programa.

Una característica de Epidat que se debe tener en cuenta a la hora de trabajar con él, es que no gestiona bases de datos, lo cual significa que no es posible abrir desde el programa un archivo de datos, visualizarlo y modificar o crear variables. Por tanto, las opciones del menú *Archivo* no son para ese propósito, sino para manejar los resultados, que pueden ser tanto numéricos como gráficos. En consecuencia, el primer paso para empezar a trabajar con Epidat

no es abrir una base de datos, como se hace habitualmente en los paquetes estadísticos, sino que hay que abrir la ventana correspondiente al método que se quiere utilizar e introducir, de forma manual o automática, los datos que dicha ventana solicite. La entrada automática consiste en importar una tabla de datos que previamente debe prepararse en una hoja de cálculo, como Excel o Libreoffice Calc.

En cuanto a los resultados, Epidat 4.1 incluye un editor de gráficos, sencillo e intuitivo, que permite personalizar los gráficos que se generan con el programa. Además, todos los resultados que se obtienen durante una sesión de trabajo se pueden guardar en un archivo con formato propio de Epidat, o como texto (rtf) o pdf.

Ayuda

La ayuda de Epidat 4 es un elemento distintivo del programa. Está disponible un documento de ayuda general y un documento para cada módulo elaborado con un enfoque didáctico y crítico, de forma que no solo describe el manejo del módulo sino que incluye los fundamentos de los métodos, recomendaciones para su aplicación y numerosos ejemplos resueltos y comentados; también incluye un anexo con las fórmulas de cada módulo y la bibliografía utilizada.

La ayuda está integrada en el programa y se accede a ella a través del menú *Ayuda*, pero también está disponible en formato pdf en la web de descarga de Epidat.

Contenido

Como ya se comentó, Epidat 4.1 tiene 11 módulos temáticos que incluyen métodos estadísticos y epidemiológicos. En este apartado se va a describir el contenido de los 4 módulos específicamente estadísticos que pueden ser de utilidad como herramienta en el aula: análisis descriptivo, muestreo, inferencia sobre parámetros y distribuciones de probabilidad.

Módulo de muestreo:

Este módulo tiene un conjunto de opciones que permiten calcular el tamaño de muestra necesario para realizar un estudio por muestreo (Figura 2), opciones para realizar la selección de la muestra y técnicas para analizar los resultados teniendo en cuenta el diseño muestral. El esquema del módulo es el siguiente:

- Cálculo de tamaños de muestra:
 - Intervalos de confianza
 - Contrastes de hipótesis
- Selección de muestras:
 - Muestreo aleatorio simple
 - Muestreo sistemático
 - Muestreo aleatorio estratificado
 - Muestreo por conglomerados monoetápico / bietápico
 - Muestreo por conglomerados monoetápico / bietápico estratificado
- Estimación con muestras complejas

Las ventanas del submódulo de cálculo de tamaños de muestra son muy sencillas y solo requieren unos pocos datos que se introducen manualmente (Figura 2). En selección de muestras, se pueden obtener números aleatorios entre 1 y N, siendo N un tamaño de población, o se puede seleccionar una muestra a partir de un archivo que contiene el listado de los elementos de la población (marco muestral). El resultado será también un archivo con la misma información que el de partida, pero incluyendo solo los registros seleccionados.

Figura 2.- Ventana para calcular el tamaño de muestra para estimar una proporción. Módulo de *Muestreo*.

Módulo de análisis descriptivo:

Este módulo permite hacer el análisis descriptivo de un conjunto de datos mediante tablas, gráficos o medidas características. El esquema del módulo es el siguiente:

- Tablas de frecuencias
- Tablas de contingencia
- Estadísticos descriptivos
- Coeficiente de correlación
- Gráficos:
 - Gráfico de barras
 - Gráfico de sectores
 - Gráfico de líneas
 - Gráfico de dispersión
 - Histograma
 - Diagrama de cajas
 - Intervalos de confianza

Todas las ventanas de este módulo, exceptuando el gráfico de líneas y el de intervalos de confianza, tienen una estructura similar (Figura 3) y comparten las siguientes características:

- La entrada de datos solo se puede realizar de forma automática a partir de un archivo en el que cada fila represente a un individuo o elemento y cada columna a una variable.
- Es posible establecer filtros en los datos, definiendo condiciones lógicas a partir de las variables del archivo.
- Los resultados se pueden *segmentar* en función de las categorías de una variable cualitativa de forma que se obtengan resultados para cada una de las subpoblaciones definidas por dichas categorías.

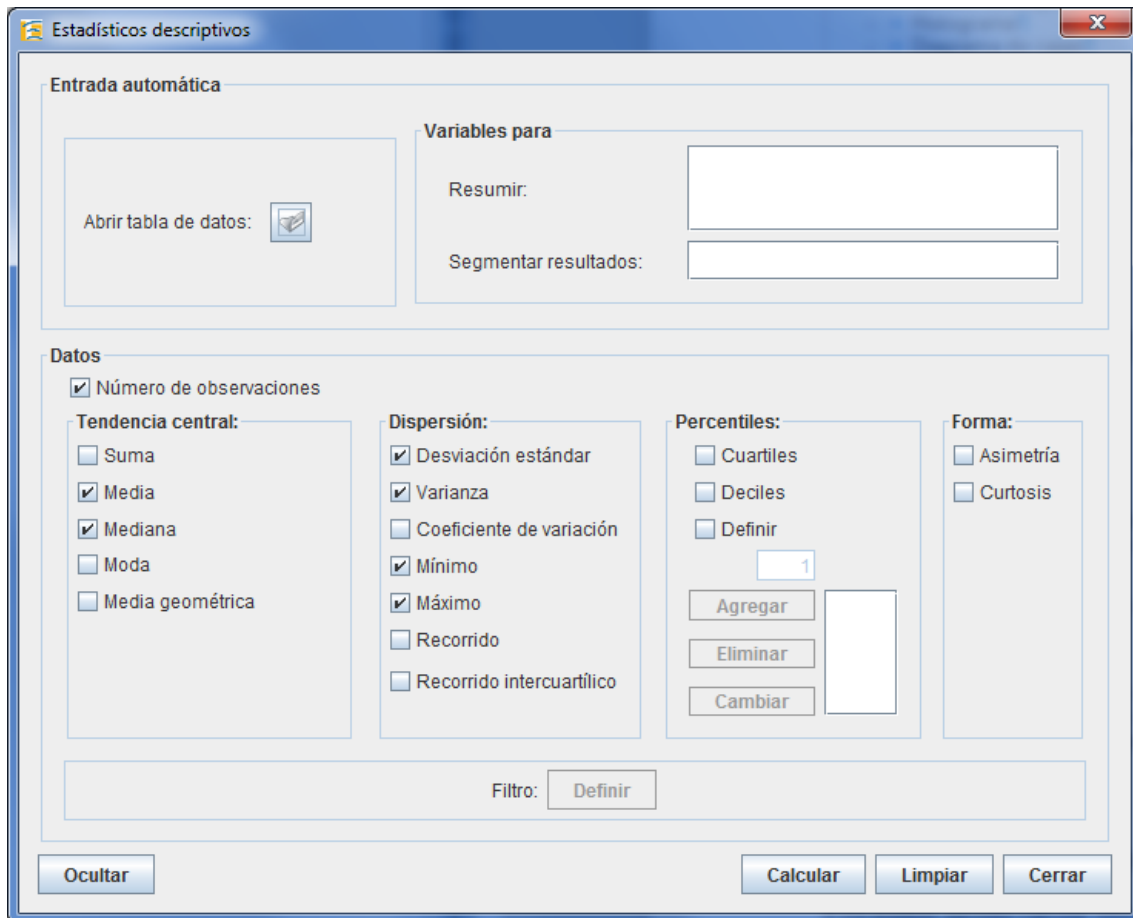


Figura 3.- Ventana para calcular estadísticos descriptivos en el módulo de *Análisis descriptivo*.

Módulo de inferencia sobre parámetros:

Este módulo permite calcular intervalos de confianza y hacer pruebas de hipótesis sobre medias, proporciones y tasas de incidencia, con datos de una o dos muestras. El esquema del módulo es el siguiente:

- Una población:
 - Media
 - Proporción
 - Percentiles
 - Correlación
 - Tasa de incidencia
 - Índice de posición
- Dos poblaciones:
 - Medias independientes
 - Medias emparejadas
 - Proporciones independientes
 - Proporciones emparejadas
 - Tasas de incidencia
- Comparación no paramétrica:

- Medias independientes
- Medias emparejadas
- Contraste de normalidad

Las ventanas de una y dos poblaciones permiten introducir los datos de dos formas: manualmente, ya resumidos, o de forma automática a partir de un archivo con la información individual (Figura 4). En este último caso es posible establecer filtros en los datos, como en el módulo de *Análisis descriptivo*.

Inferencia. Comparación de medias independientes

Datos

Datos resumidos

	Muestra 1	Muestra 2
Media	0,000	0,000
Desviación estándar	0,000	0,000
Tamaño de muestra	0	0

Datos individuales

Abrir datos

Filtro:

Nivel de confianza %

Calcular

Intervalo de confianza

Contraste de hipótesis

Bilateral ($H_0: \mu_1 - \mu_2 = 0$ vs. $H_1: \mu_1 - \mu_2 \neq 0$)

Unilateral izquierdo ($H_0: \mu_1 - \mu_2 = 0$ vs. $H_1: \mu_1 - \mu_2 < 0$)

Unilateral derecho ($H_0: \mu_1 - \mu_2 = 0$ vs. $H_1: \mu_1 - \mu_2 > 0$)

Figura 4.- Ventana para contrastar la hipótesis de igualdad de medias de dos poblaciones independientes en el módulo de *Inferencia sobre parámetros*.

Módulo de distribuciones de probabilidad:

Este módulo permite calcular funciones de distribución y sus inversas para un conjunto de distribuciones discretas y continuas, las representa gráficamente, y también genera muestras simuladas de estas distribuciones. El esquema del módulo es el siguiente:

- Cálculo de distribuciones:
 - Distribuciones discretas
 - Distribuciones continuas
- Generación de distribuciones:

- Distribuciones discretas
- Distribuciones continuas

En las cuatro ventanas de este módulo se presenta una lista de distribuciones, discretas o continuas según el caso, entre las que se debe seleccionar una e introducir sus parámetros (Figura 5).

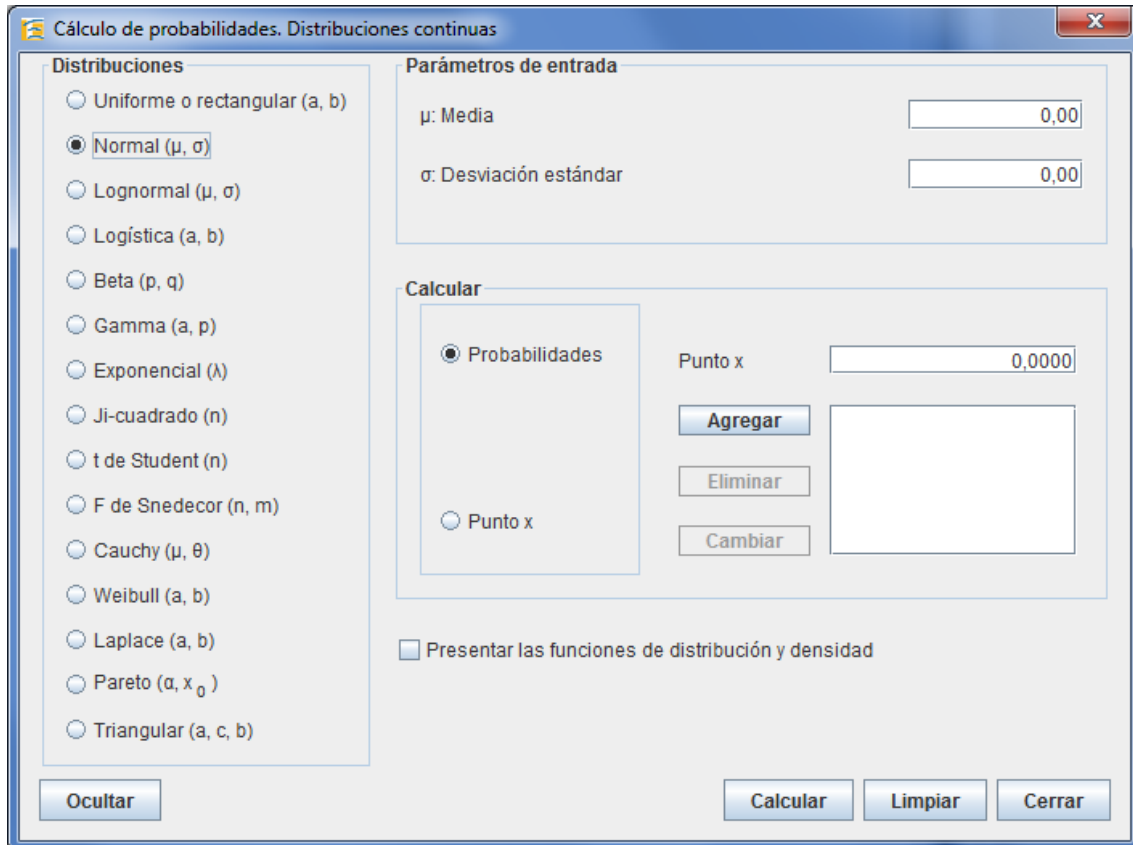


Figura 5.- Ventana de cálculo de probabilidades para distribuciones continuas en el módulo de *Distribuciones de probabilidad*.

Práctica con Epidat 4.1

Para llevar a cabo la práctica de Epidat se van a utilizar los datos de la encuesta realizada por un grupo de alumnos de 1º de ESO del IES Sánchez Cantón de Pontevedra para el proyecto “Y tú ¿qué desayunas?” que presentaron a la III edición del concurso Incubadora de Sondaxes e Experimentos (fase local de Galicia 2013) y con el que obtuvieron el premio de su categoría. Además, representaron a Galicia en la Fase Nacional del concurso, que tuvo lugar en Santiago. El trabajo completo está disponible en la página web de la SGAPEIO (www.sgapeio.es).

Entre los ejercicios del taller se incluirá el cálculo de un tamaño de muestra y una selección de muestra a partir de datos simulados que corresponderían al listado de alumnos del centro por curso y clase. Con los datos de la encuesta, recogidos mediante un cuestionario diseñado por los alumnos y grabados en una tabla de Excel, se realizará un análisis descriptivo y alguna prueba de hipótesis. Por último, se mostrará un ejemplo de cálculo de probabilidades con el módulo de *Distribuciones de probabilidad*. En la Figura 6 se muestran algunos de los resultados que se obtendrán durante el taller.

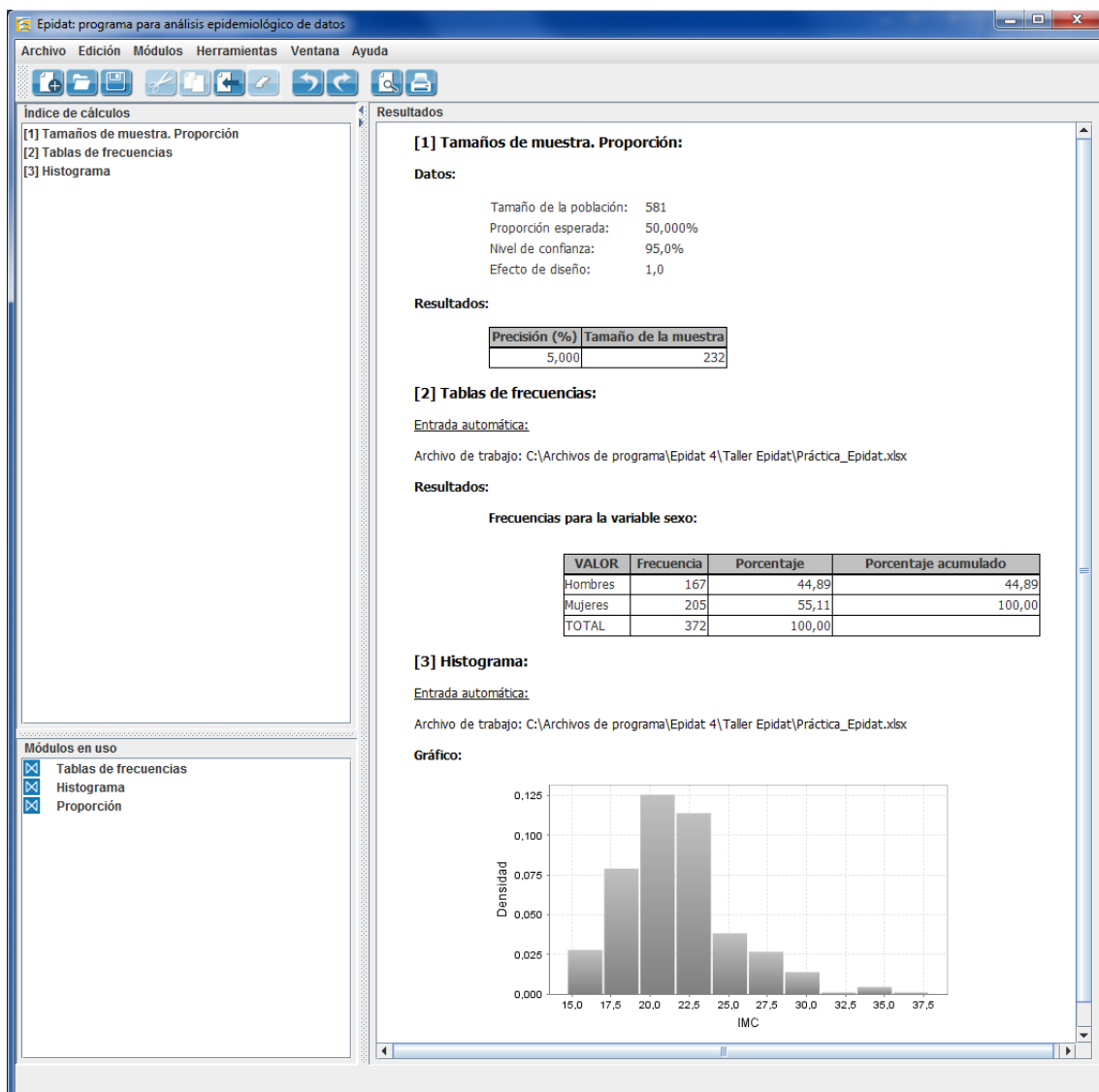


Figura 6.- Resultados obtenidos con Epidat 4.1.

Referencias bibliográficas

[1] Santiago Pérez MI, Hervada Vidal X, Naveira Barbeito G, Silva LC, Fariñas H, Vázquez E, Bacallao J, Mujica OJ. El programa Epidat: usos y perspectivas [Carta al editor]. Rev Panam Salud Publica. 2010;27(1):80-82.

[2] Hervada Vidal X, Santiago Pérez MI, Vázquez Fernández E, Castillo Salgado C, Loyola Elizondo E, Silva Ayçaguer LC. Epidat 3.0: Programa para el análisis epidemiológico de datos tabulados. Versión 3.0. Rev Esp Salud Pública 2004;78(2):277-80.

[3] Epidat: programa para análisis epidemiológico de datos. Versión 4.1, octubre 2014. Consellería de Sanidade, Xunta de Galicia, España; Organización Panamericana de la Salud (OPS-OMS); Universidad CES, Colombia. Disponible en: <http://dxsp.sergas.es>.