

# **DOCENCIA EN ESTADÍSTICA CON MICROSOFT® EXCEL: PROBABILIDAD**

M<sup>a</sup> Teresa Díaz Delfa, Fuensanta Arnaldos García, Ursula Faura Martínez,  
Lourdes Molera Peris, Isabel Parra Frutos  
*Universidad de Murcia*

## **RESUMEN**

Con la finalidad de que los alumnos comprendan mejor los conceptos teóricos desarrollados en cualquier curso de estadística y, además, dotar a los futuros profesionales de recursos que les ayuden a manejar y tratar datos del mundo real, es conveniente estudiar la estadística desde una perspectiva más práctica apoyándonos en herramientas informáticas como las hojas de cálculo. Estas presentan la ventaja de que son mucho más versátiles a nivel básico que el software puramente estadístico (SPSS, Statgraphics, S-PLUS,...), puesto que, en general, permiten obtener los resultados deseados no sólo utilizando funciones y herramientas específicas, sino también a través de un enfoque más intuitivo y similar al cálculo manual que se realizaría sin ordenador. En particular, centramos nuestra atención en Microsoft® Excel, dada su amplia implantación actual.

Así, se pueden estudiar los modelos de variables aleatorias, tanto discretas como continuas, mediante el uso de un paquete informático, pudiendo calcular desde probabilidades hasta la forma gráfica de las distribuciones, lo que permitiría al alumno comprender mejor las características de los modelos estudiados.

**Palabras clave:** cálculo de probabilidades, modelos de distribuciones, hoja de cálculo.

## **1. INTRODUCCIÓN**

En una materia como la Estadística en las titulaciones de Ciencias Sociales es indiscutible la necesidad de trabajar con programas informáticos. Los cálculos que pretenden sintetizar las masas de datos ya no se realizan con lápiz y papel debido, principalmente, a la gran cantidad de datos, a la rapidez de las herramientas informáticas y a la accesibilidad a los medios informáticos de los estudiantes y profesionales. Sin embargo, es necesario usar estos programas con precaución, ya que nuestro objetivo no se debe limitar a la simple interpretación de las salidas proporcionadas por el software que se decida utilizar, sino que es importante que los alumnos conozcan los fundamentos teóricos y los procedimientos prácticos que conducen a esos resultados, para así lograr un mejor entendimiento de éstos.

La elección de la hoja de cálculo como herramienta informática frente a programas estadísticos específicos estriba sobre todo en su disponibilidad, ya que ésta se puede encontrar en casi todos los ordenadores a los que tiene acceso el alumno mientras cursa sus estudios, así como en la mayoría de puestos de trabajo. Además, presenta la ventaja de que es mucho más versátil que el software puramente estadístico. En general, permite obtener los resultados deseados no sólo utilizando funciones y herramientas específicas, sino también a través de un enfoque más intuitivo y similar al cálculo manual que se realizaría sin ordenador. Se puede comenzar utilizando la hoja de cálculo básicamente como una calculadora potente y versátil, en la que es necesario controlar todo el proceso de obtención de un resultado, tal y como haría el alumno para resolver el problema manualmente. Esta manera de actuar puede complementarse, de una forma más directa, utilizando las funciones y macros disponibles que permiten obtener los resultados sin necesidad de programar todo el proceso.

Las prácticas con ordenador nos ayudan a abordar dos aspectos que consideramos muy importantes:

- Potenciar el papel de la visualización.
- Evitar la limitación que, a la hora de realizar cálculos, supone emplear una gran cantidad de datos. Con el ordenador podemos, por tanto, proponer prácticas a nuestros alumnos mucho más interesantes y realistas.

El uso de este tipo de sesiones prácticas depende de la materia que se imparte. En el caso de los contenidos de Estadística Descriptiva e Inferencia Estadística, la hoja de cálculo facilita enormemente los cálculos a realizar dadas sus prestaciones, entre las que se incluyen disponibilidades gráficas suficientes.

Por otra parte, aunque los contenidos correspondientes a la Teoría de la Probabilidad son posiblemente los que resultan más complicados de enfocar desde un punto de vista práctico, se puede utilizar Microsoft® Excel para estudiar los principales modelos de distribuciones de probabilidad. Esta hoja de cálculo permite obtener los valores de las funciones de probabilidad (en modelos de variables discretas) y de las funciones de densidad (en modelos de variables continuas), así como los de las funciones de distribución en ambos casos.

Por ejemplo, se pueden representar gráficamente dichas funciones para observar y experimentar el cambio de forma de las mismas ante modificaciones en los valores de los parámetros de la distribución correspondiente. Así, se puede incidir en la relación que existe entre los valores de los parámetros y la forma que adquiere la representación gráfica de la distribución. Otras utilidades son visualizar la probabilidad de un intervalo de valores como área encerrada por la función de densidad en el intervalo correspondiente en un modelo de variable continua, y desarrollar distintas ideas intuitivas relacionadas con los teoremas límite.

El objetivo de esta comunicación es, por tanto, ilustrar cómo el uso de esta hoja de cálculo puede ayudar a clarificar conceptos tratados en la docencia en Teoría de la Probabilidad, centrándonos para ello en dos distribuciones muy usuales, la distribución Binomial y la distribución Normal.

## **2. ESTUDIO DE LAS DISTRIBUCIONES BINOMIAL Y NORMAL UTILIZANDO MICROSOFT® EXCEL**

Microsoft® Excel dispone de funciones muy útiles para trabajar con los modelos de variables aleatorias más comunes, tanto discretas como continuas. A modo de ejemplo, introducimos en este caso práctico las funciones relativas a las distribuciones Binomial y Normal.

- **DISTR.BINOM(núm\_éxito;ensayos;prob\_éxito;acumulado).** Proporciona, para una variable aleatoria Binomial de parámetros  $n$  (**ensayos**) y  $p$

(**prob\_éxito**), denotada por  $B(n,p)$ , el valor de la función de probabilidad en el punto  $x$  (**núm\_éxito**), en el caso de **acumulado** igual a FALSO, y el de la probabilidad acumulada en el mismo punto (función de distribución), en el caso de **acumulado** igual a VERDADERO.

- **DISTR.NORM(x;media;desv\_estándar;acum)**. Calcula, para una variable aleatoria Normal de parámetros  $\mu$  (**media**) y  $\sigma$  (**desv\_estándar**), denotada por  $N(\mu,\sigma)$ , el valor de la función de densidad en el punto  $x$ , en el caso de **acum** igual a FALSO, y el de la probabilidad acumulada en el mismo punto (función de distribución), en el caso de **acum** igual a VERDADERO.

Utilizando ambas funciones se puede explicar a los alumnos cómo obtener a través de la hoja de cálculo cualquier probabilidad que nos interese para ambos tipos de distribuciones, mostrando este cálculo como una alternativa al uso de las tradicionales tablas estadísticas. También podemos representar gráficamente las probabilidades que se han calculado y dejar los cálculos referenciados a celdas en las que aparezcan los valores de los parámetros, observando cómo cambian las probabilidades y sus representaciones gráficas cuando se alteran los valores de dichos parámetros.

Esta comunicación está orientada a clarificar conceptos que suelen ser complicados de entender utilizando, básicamente, las funciones de Excel anteriormente citadas y las representaciones gráficas disponibles en la hoja de cálculo. En concreto, la aplicación que presentamos está basada en un fichero de Microsoft<sup>®</sup> Excel que se ha estructurado en cinco hojas. La primera de ellas muestra los objetivos que se persiguen con el mismo, así como el contenido de cada una de las cuatro hojas restantes (véase figura 1). Pulsando en el botón correspondiente al contenido que se desee trabajar se activa la hoja correspondiente. En las restantes hojas aparece el botón **INDICE** que permite, por su parte, volver a esta primera hoja de introducción.

Para mostrar esta aplicación a nuestros alumnos recurrimos al uso combinado del ordenador y el cañón de vídeo, disponible para uso del profesor en las aulas de los centros en los que impartimos docencia. Adicionalmente, el fichero se puede facilitar a los alumnos en la página web de la asignatura para que puedan experimentar por su cuenta.

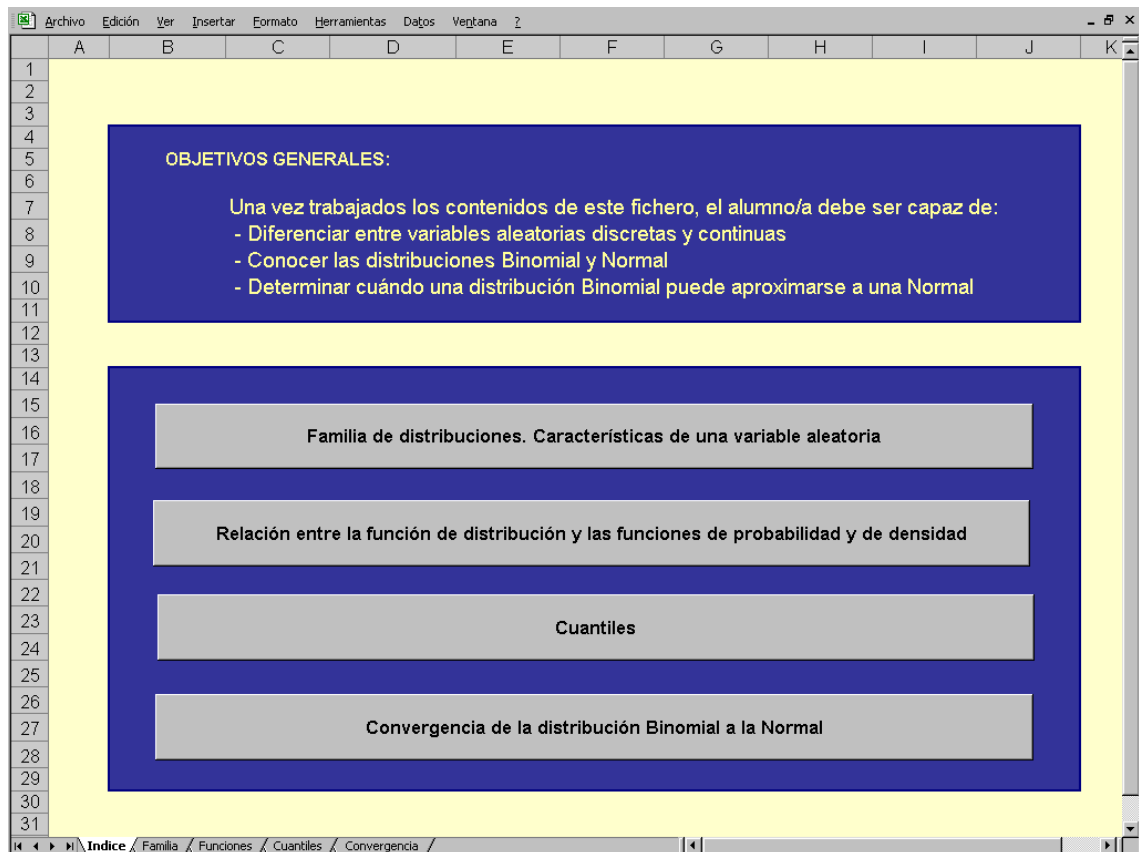


Figura 1: Hoja de cálculo “Índice”

## 2.1. Familia de distribuciones. Características de una variable aleatoria

El objetivo que se persigue con esta hoja de cálculo es que los alumnos comprendan el concepto de familia de distribuciones. Bajo esta denominación se agrupan variables aleatorias que tienen la misma distribución de probabilidad aunque con diferentes valores en uno o más parámetros presentes en la misma.

Cada una de estas familias de distribuciones sigue un patrón de comportamiento probabilístico que se ajusta a un experimento aleatorio genérico y bajo el que tienen cabida numerosos fenómenos aleatorios. Todos los miembros de una misma familia están caracterizados por la misma función de probabilidad (caso discreto) o de densidad (caso continuo) diferenciándose únicamente en una serie de valores denominados parámetros. Según el valor o valores que tomen estos parámetros, la representación gráfica de las distribuciones para una misma familia variará, pudiendo alterarse características como la asimetría.

En esta hoja de cálculo (véase figura 2) buscamos mostrar visualmente cómo influyen los valores de los parámetros en la forma de la gráfica de la función de

probabilidad de la Binomial y de la función de densidad de la Normal, así como en los valores de distintas características, como son la media, la varianza, el coeficiente de asimetría y el de curtosis. En el caso de la Binomial, si el parámetro  $p$  está cercano a cero, la probabilidad de que el suceso de interés ocurra es baja, por lo que los valores pequeños de la variable aleatoria serán más probables que los valores grandes y la distribución será asimétrica a la derecha. Por el contrario, si  $p$  está próximo a uno, los valores de la variable aleatoria grandes serán más probables que los pequeños, y el gráfico corresponderá a una distribución asimétrica a la izquierda. En el caso de la Normal, su representación gráfica corresponde siempre a una distribución simétrica respecto al parámetro  $\mu$  y con forma de campana, y cuya altura en torno a  $\mu$  depende del valor que toma el parámetro  $\sigma$ . Por ejemplo, si éste es pequeño, esto es, existe poca dispersión, los valores próximos a  $\mu$  serán más probables y la representación gráfica será más apuntada.

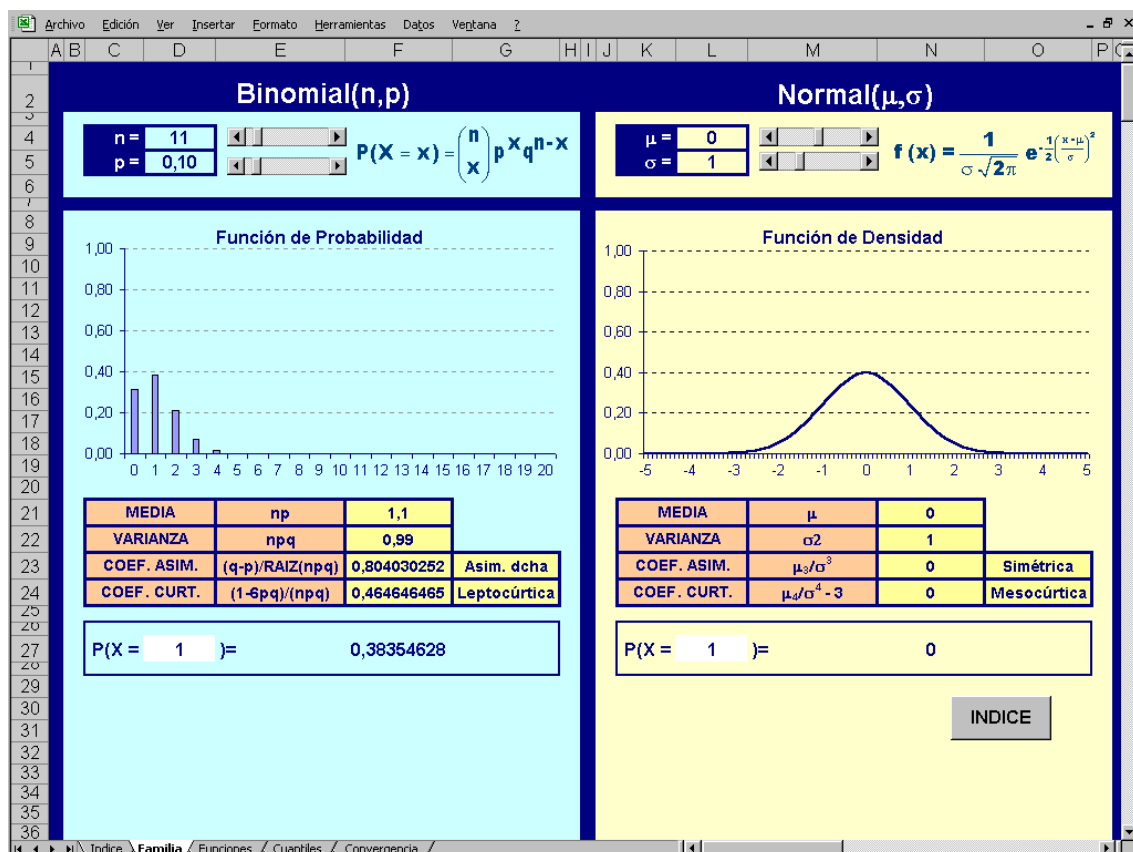


Figura 2: Hoja de cálculo "Familia"

Finalizamos con la obtención de la probabilidad con que una variable aleatoria toma un valor concreto, que en el caso continuo es siempre cero, mientras que en el caso discreto puede no serlo.

## 2.2. Relación entre la función de distribución y las funciones de probabilidad y de densidad

El objeto de esta hoja (véase figura 3) es insistir en la relación que existe entre la función de distribución y las funciones de probabilidad o de densidad, según la variable aleatoria sea discreta o continua. En el primer caso, la función de distribución en un punto es la suma de las probabilidades puntuales de la variable aleatoria hasta ese valor inclusive, mientras que en el segundo caso es el área limitada por la función de densidad y a la izquierda de la recta perpendicular al eje de abscisas en dicho punto.

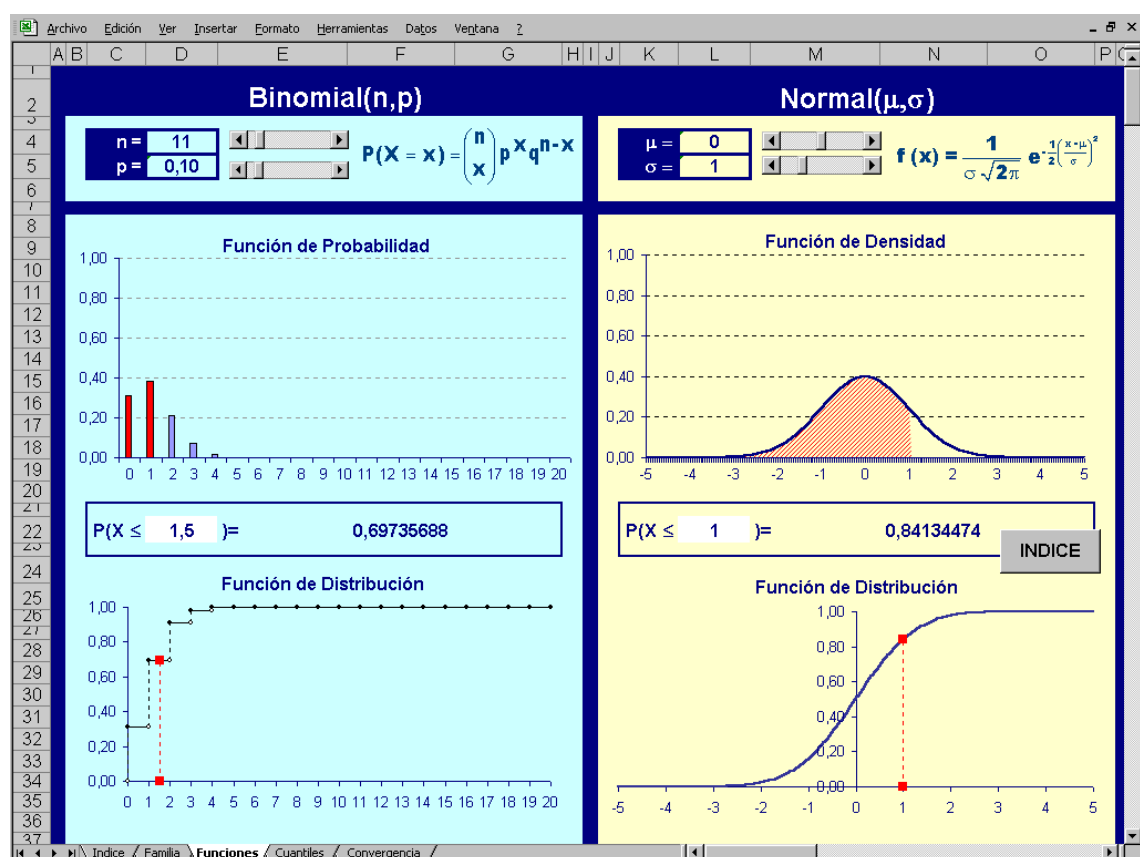


Figura 3: Hoja de cálculo “Funciones”

Además, se puede calcular el valor de la función de distribución en un punto (a modificar en la celda en blanco), tanto para una variable aleatoria  $B(n,p)$  como para una  $N(\mu,\sigma)$ . En el caso de la distribución  $B(n,p)$  el valor de la función de distribución en dicho punto está representado utilizando un color diferente en la función de probabilidad (sería la suma de las columnas de color rojo), así como en la propia función de distribución. Por su parte, en la  $N(\mu,\sigma)$  se puede observar el área rayada bajo la función de densidad y el valor de la función de distribución.

Todos los cálculos de la hoja están referenciados a las celdas que contienen los valores de los parámetros, pudiendo alterar éstos utilizando las barras de desplazamiento disponibles al efecto.

### 2.3. Cuantiles

En esta hoja (véase figura 4) únicamente se pretende mostrar gráficamente cómo en el caso de una variable aleatoria continua siempre es posible encontrar un valor de la variable (cuantil) para el cual la probabilidad acumulada a la izquierda sea igual a un valor fijado de antemano, pero que en el caso de una variable discreta sólo es factible en algunos casos. Por ello, en el caso de variables aleatorias discretas debemos recurrir a localizar el primer valor de la variable para el que la probabilidad acumulada a la izquierda es mayor o igual a la especificada.

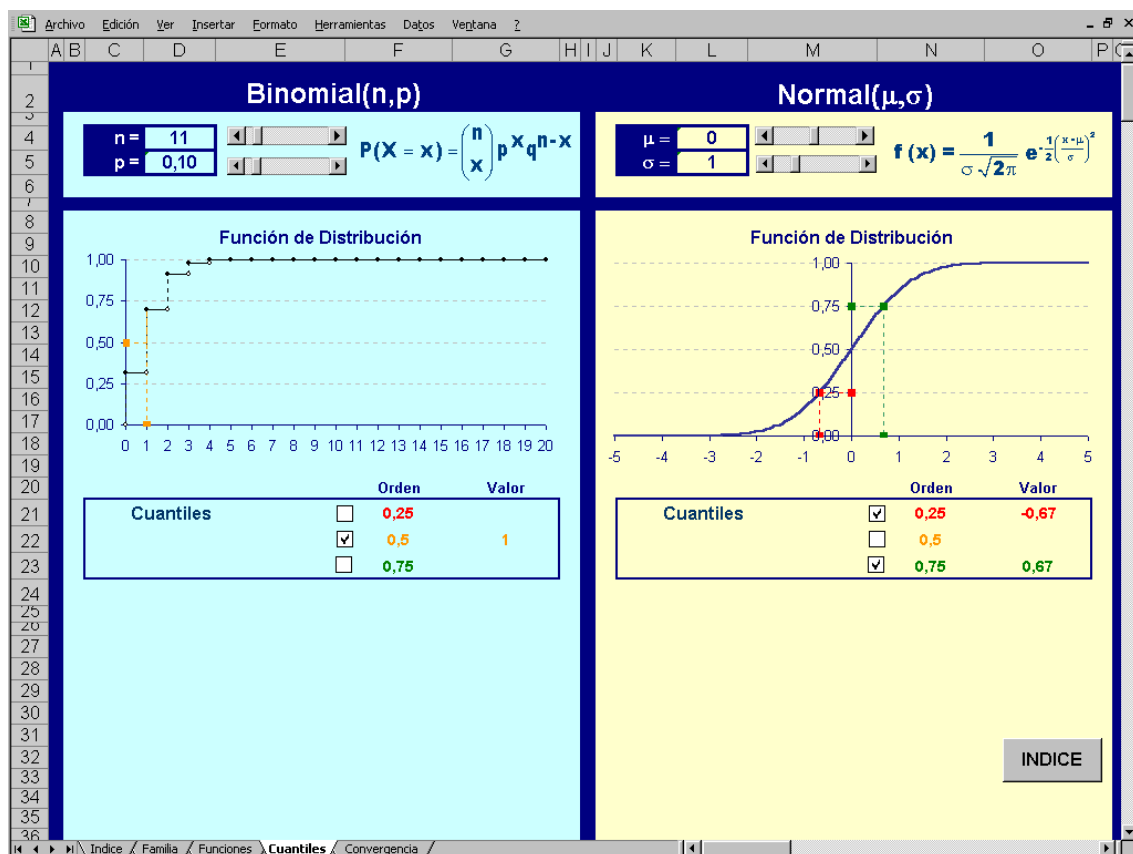


Figura 4: Hoja de cálculo “Cuantiles”



La hoja está diseñada para obtener y representar simultáneamente hasta 3 cuantiles<sup>1</sup> simplemente pinchando en la casilla de verificación correspondiente e introduciendo en la celda el orden de los mismos. También es posible observar los cambios que se producen en los cuantiles para diferentes valores de los parámetros de las distribuciones Binomial y Normal (utilizando las barras de desplazamiento).

#### **2.4. Convergencia de la distribución Binomial a la Normal**

Por último, como una aplicación del teorema central del límite estudiamos la convergencia de una distribución Binomial a una distribución Normal. En estos teoremas se estudia cómo la suma de variables aleatorias con media y varianza finitas se aproxima “bajo ciertas condiciones” a una distribución conocida cuando el número de variables aleatorias es suficientemente grande.

La distribución Binomial de parámetros  $n$  y  $p$  puede entenderse como la suma de  $n$  variables aleatorias independientes, todas ellas Bernouilli de igual parámetro  $p$ . A medida que aumenta el parámetro  $n$ , la distribución Binomial se aproxima a una Normal de media y varianza iguales a las de la Binomial.

En la hoja “Convergencia” se puede comparar gráficamente (véase figura 5) la función de probabilidad de la distribución  $B(n,p)$  con la función de densidad de la distribución  $N(\mu = np, \sigma = \sqrt{npq})$ , observando la mayor o menor semejanza de las mismas para distintos valores de los parámetros de la distribución Binomial (seleccionados utilizando las barras de desplazamiento).

Adicionalmente, se puede activar o desactivar el cálculo y representación gráfica de la probabilidad de un intervalo de valores. El valor de dicha probabilidad se ofrece tanto para la distribución  $B(n,p)$ , como para su aproximación bajo la distribución  $N(\mu = np, \sigma = \sqrt{npq})$ . En este último caso se emplea la corrección por continuidad, pudiendo observarse de forma gráfica esta probabilidad.

---

<sup>1</sup> En la figura 4 se ha elegido como ilustración los cuantiles (tres valores de la variable aleatoria para los que se verifica que la probabilidad acumulada es al menos 0,25; 0,5 y 0,75, respectivamente).

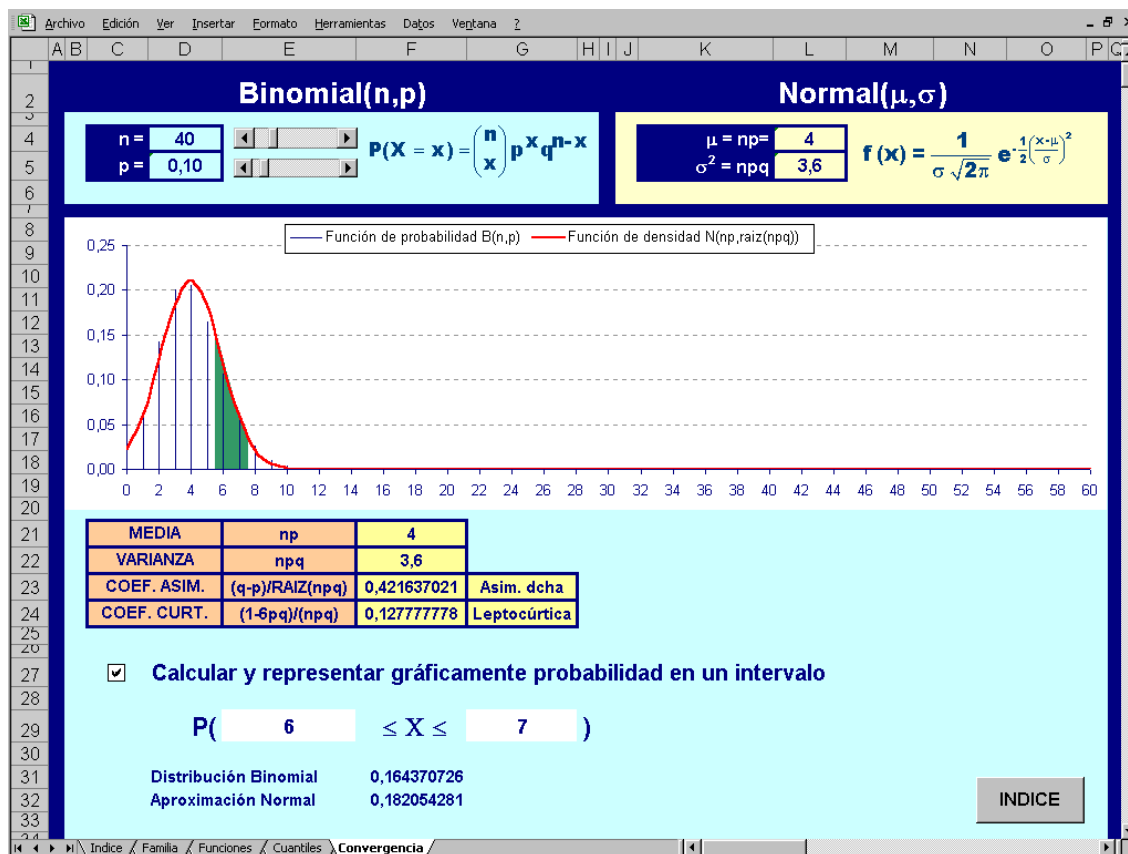


Figura 5: Hoja de cálculo “Convergencia”

### 3. CONCLUSIONES

Actualmente existe una gran variedad de programas informáticos que pueden ser de gran utilidad en la docencia de asignaturas de estadística. Entre ellos nos hemos decantado por la utilización de la hoja cálculo Microsoft® Excel frente a otro software específico de estadística debido, fundamentalmente, a su versatilidad y accesibilidad.

En el caso concreto de la Teoría de la Probabilidad su uso puede sustituir a las tradicionales tablas estadísticas y dar un soporte visual a conceptos teóricos que puedan resultar complicados de entender para los alumnos. Así, y entre otras cosas, permite incidir en: (1) el cambio de forma de las funciones de probabilidad, densidad y distribución ante modificaciones en los valores de los parámetros de las distribuciones; (2) la visualización de la probabilidad de un intervalo de valores como área encerrada por la función de densidad en el intervalo correspondiente en un modelo de variable continua ; (3) el desarrollo de ideas intuitivas relacionadas con los teoremas límite; etc.

El uso de la hoja de cálculo para propiciar la visualización y experimentación nos ayuda a captar la atención de nuestros alumnos, a desarrollar clases más dinámicas, a aumentar su motivación al hacer la estadística más cercana y accesible, y a fomentar un aprendizaje activo y cooperativo.

#### **4. REFERENCIAS BIBLIOGRÁFICAS**

- ARNALDOS, F.; DÍAZ, M.T.; FAURA, U.; MOLERA, L. y PARRA, I. (2003). Estadística descriptiva para economía y administración de empresas: cuestiones tipo test y ejercicios con Microsoft® Excel. Madrid: AC.
- CASAS SÁNCHEZ, J.M. y SANTOS PEÑAS, J. (1996). Introducción a la estadística para administración y dirección de empresas. Madrid: Editorial Centro de estudios Ramón Areces.
- FERNÁNDEZ-ABASCAL, H.; GUIJARRO, M. M.; ROJO, J. L. y SANZ, J. A. (1994). Cálculo de probabilidades y estadística. Barcelona: Ariel.
- PARRA FRUTOS, I. (2003). Estadística empresarial con Microsoft® Excel: Problemas de inferencia. Segunda edición. Madrid: AC.
- PIÑOLE, R.; MORENO, A. y CABALLERO, A. (2002). Análisis de datos y probabilidad. Excel como instrumento de cálculo. Madrid: Civitas.