

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

Isabel Parra Frutos, Lourdes Molera Peris, Fuensanta Arnaldos García,  
MªTeresa Díaz Delfa, Úrsula Faura Martínez

*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, más versátiles que el software puramente estadístico (SPSS, Statgraphics, S-PLUS,...). En particular, centramos nuestra atención en Microsoft® Excel, dada su amplia implantación actual.

La hoja de cálculo permite no sólo realizar cálculos de forma rápida y sencilla, incluso con grandes masas de datos, sino que también posibilita incidir en conceptos, como los intervalos de confianza o los contrastes de hipótesis, que a los alumnos les suelen resultar complicados de entender. Con este objetivo, es posible crear hojas de cálculo en las que los alumnos pueden llevar a cabo cierta “experimentación” para que, a través de la misma, los conceptos les resulten mucho más claros.

**Palabras clave:** inferencia estadística, contrastes de hipótesis, hoja de cálculo.

## **1. INTRODUCCIÓN**

En la actualidad es habitual generar y registrar gran cantidad de datos de todo tipo que son susceptibles de un tratamiento estadístico, revelándose este tipo de análisis fundamental para conocer la realidad en la que se está inmerso y poder así tomar decisiones en cualquier contexto, pero quizás aún más en el ámbito de la Economía y la Administración de Empresas. Sin embargo, el elevado número de datos invalida el uso de la calculadora y precisa de alguna herramienta más potente, capaz de manejar esa gran cantidad de información de forma rápida y sencilla. Las herramientas informáticas constituyen la solución, pero su manejo requiere, lógicamente, un entrenamiento para que el futuro usuario se sienta capaz y confiado en el uso de las mismas.

La descentralización informática, como consecuencia del auge del ordenador personal, hace que los profesionales puedan abordar por su cuenta el tratamiento estadístico de la información, modificando la tradicional dependencia de los departamentos de informática. En este nuevo contexto en el que trabajan los profesionales de hoy en día, y en el que trabajarán en el futuro nuestros alumnos, las hojas de cálculo constituyen una ayuda inestimable tanto a la hora de organizar y hacer más inteligibles grandes masas de datos como de llevar a cabo análisis estadísticos de los hechos económicos y de la información disponible en las empresas.

En la actualidad la herramienta informática más difundida en el mercado es Microsoft® Office, que incluye la hoja de cálculo Excel. De hecho, es frecuente que forme parte del software disponible en empresas e instituciones públicas. Por este motivo ésta es la herramienta principal que utilizamos en nuestras clases, tanto como apoyo para una mejor comprensión de los conceptos teóricos como para agilizar los cálculos involucrados en los ejercicios prácticos.

La alternativa a la hoja de cálculo para enseñar inferencia estadística es utilizar paquetes estadísticos específicos, como SPSS, Minitab, STATGRAPHICS, SAS, S-PLUS, etc., pero estos últimos son excesivamente caros y, por tanto, no suelen ser adquiridos para el uso generalizado del personal de las empresas y de las administraciones públicas. Además, el software puramente estadístico no sirve para hacer más comprensibles los conceptos, sino únicamente para aplicar diversas técnicas inferenciales sin mostrar los cálculos necesarios para conseguir el resultado final. Desde nuestro punto de vista, consideramos que es más adecuado trabajar los conceptos y realizar los ejercicios prácticos con una hoja de cálculo y, en todo caso, dar al final del

curso una visión general de cómo se utiliza un paquete estadístico y qué técnicas inferenciales estudiadas lleva incorporadas, comparando sus salidas con los resultados obtenidos mediante Excel en las diferentes prácticas resueltas.

## **2. LA ENSEÑANZA DE LA INFERENCIA ESTADÍSTICA APOYADA EN UNA HOJA DE CÁLCULO**

En la asignatura dedicada al estudio de la Inferencia Estadística el uso de una hoja de cálculo como Microsoft® Excel nos va a ayudar a conseguir una mejor comprensión de ciertos conceptos estadísticos que, por nuestra experiencia, sabemos que los alumnos tienen dificultades para asimilar, ya que facilitará, en la mayoría de las ocasiones, la presentación de explicaciones intuitivas y/o gráficas de los mismos. Creemos que este enfoque es más didáctico, dado que permite que los alumnos visualicen y experimenten con los conceptos, favoreciendo la comprensión e interiorización de los mismos. Recordemos que lo que uno aprende de forma activa es mucho más difícil de olvidar.

Desde una perspectiva más aplicada, trabajar con Excel agiliza considerablemente los cálculos involucrados en los ejercicios que se proponen y resuelven en las clases prácticas. Esta reducción del tiempo se pone de manifiesto sobre todo en aquellas situaciones en las que se dispone de muchos datos, en cuyo caso el uso de la calculadora se hace muy lento, y/o en la aplicación de algunas técnicas estadísticas que involucran un gran número de operaciones, algunas de las cuales se repiten varias veces. Pueden utilizarse entonces datos reales sin ningún problema, algo que se suele evitar si se trabaja con las herramientas tradicionales (calculadora y pizarra), puesto que suele ser muy engorroso. Se consigue así potenciar la parte práctica de la estadística, que es lo verdaderamente importante para los alumnos de nuestras titulaciones, ya que si estudian esta disciplina es precisamente para poder analizar datos del mundo real. Si, por el contrario, se pierden en disquisiciones teóricas y no saben cuándo y cómo aplicar los diferentes métodos de análisis que la estadística proporciona e interpretar sus resultados, habremos fracasado en nuestro intento de dotarles de herramientas adecuadas para el análisis de datos económicos.

Con todo esto pretendemos que la estadística y, en particular, la inferencia, sea más asequible para nuestros alumnos, menos abstracta, para que realmente puedan entender cuáles son sus objetivos, su potencial y sus limitaciones. Se pretende hacer

más cercana esta materia, en un entorno que será muy habitual en su futuro puesto de trabajo, con lo cual esperamos que quede más clara su utilidad y se consiga que verdaderamente hagan uso de ella en su vida laboral.

No conviene olvidar que la estadística forma parte de las asignaturas instrumentales de titulaciones aplicadas como “Economía” y “Administración y dirección de empresas”, y que, por lo general, los estudiantes no muestran excesivo interés por ella. Una forma de motivarles es hacer que realmente la entiendan y comprueben su utilidad para resolver problemas prácticos reales, y, desde nuestra experiencia, el empleo de una hoja de cálculo influye de forma positiva en la consecución de estos objetivos.

Finalmente, como beneficio añadido, al realizar prácticas de estadística con la hoja de cálculo Excel nuestros alumnos descubren las posibilidades que esta herramienta informática les ofrece en general para su futuro profesional, no sólo desde el punto de vista de la estadística.

## **2.1. Metodología utilizada**

Como ya hemos comentado, el uso de una hoja de cálculo en las asignaturas de inferencia estadística es muy útil, y puede hacerse desde dos perspectivas diferentes: (1) facilitar la resolución de casos prácticos con un número elevado de datos y (2) aportar una visión intuitiva y/o gráfica de conceptos fundamentales.

En cuanto al enfoque recogido en (1), por nuestra experiencia sabemos que, en muchas ocasiones, el alumno se “ahoga” en la maraña de cálculos involucrados en un problema de inferencia estadística, sin saber exactamente adónde quiere llegar y cómo conseguirlo. Por este motivo, una herramienta como Excel puede ayudarles, puesto que los cálculos se reducen al mínimo –usando, si fuera necesario, las funciones estadísticas implementadas y/o las macros de la herramienta *Análisis de datos*, así como la opción de llenado automático–, y el esquema básico de la práctica queda al descubierto.

En este trabajo nos centramos en el segundo enfoque. En este sentido utilizamos dos tipos de ficheros Excel. En primer lugar, como apoyo en las clases de teoría –y gracias a la disponibilidad de ordenador y cañón de video en la mayoría de las aulas de nuestra facultad– usamos ficheros bastante elaborados, que nos permitan visualizar conceptos básicos (nivel de confianza de un intervalo, función de potencia en un contraste de hipótesis,...). Los alumnos tienen acceso a estos ficheros (en la página web

de la asignatura) para que experimenten con ellos si así lo desean, pero no se les exige que sean capaces de construir otros similares. En segundo lugar, en las clases prácticas en el aula de informática se intenta afianzar los conceptos anteriores con ficheros de Excel más sencillos que sí deben construir los alumnos (por ejemplo, obtención de la función de potencia de un contraste concreto, reproducción de las tablas de niveles críticos para algún contraste en que esto sea posible, como en el caso del test no paramétrico de rachas,...). De nuevo, en este caso, el uso de las funciones estadísticas de Excel y el llenado automático facilitarán enormemente la tarea.

Para entender mejor el funcionamiento de este enfoque didáctico vamos a aplicarlo al estudio de un concepto concreto, como es la función de potencia de un contraste paramétrico.

### **3. USO DE EXCEL EN EL ESTUDIO DE LA FUNCIÓN DE POTENCIA DE UN TEST DE HIPÓTESIS PARAMÉTRICO**

En el contexto de los contrastes de hipótesis paramétricos, la función de potencia es una función definida en el conjunto de valores posibles para el parámetro desconocido  $\theta$  (espacio paramétrico) que asigna a cada posible valor de  $\theta$  la probabilidad de rechazar la hipótesis nula. Dada una muestra aleatoria simple extraída de una población cuyo modelo de distribución es conocido pero depende de algún parámetro desconocido, planteadas una hipótesis nula y otra alternativa sobre el valor de dicho parámetro, y establecida una regla de rechazo, el alumno debería ser capaz de calcular la función de potencia asociada a dicho contraste a partir de las explicaciones dadas en clase.

No obstante, el cálculo de cada una de las probabilidades que dicha función involucra suele ser muy tedioso, y en muchas ocasiones no es posible obtenerlas todas debido al infinito número de valores existentes en el espacio paramétrico. Así que, en un aula tradicional con pizarra, lo normal es limitarse al cálculo de algunos de los valores de la función de potencia a partir de los cuales “intuir” su forma. Además, muchas veces el alumno, inmerso en el cálculo manual de estas probabilidades, pierde un poco de vista el significado global de dicha función. En cambio, el uso de una hoja de cálculo como Excel permite la obtención rápida de muchos valores de la función de potencia, siendo también posible obtener una representación gráfica de la misma bastante exacta.

Como ya hemos comentado, estamos convencidos de que una visión intuitiva y gráfica de los conceptos facilita en gran medida la comprensión de los mismos. En el caso particular de la función de potencia que aquí nos ocupa, la hoja de cálculo Excel nos servirá, por ejemplo, para calcular y representar gráficamente la función de potencia asociada a un contraste dado, así como para ver e interpretar la variación de dicha función ante cambios en el tamaño muestral o en otro parámetro cuyo valor es conocido, si éste existe en la distribución poblacional.

En concreto, centrándonos en el caso de una población normal con varianza conocida en el que se desea contrastar el valor de la media, vamos a estudiar:

|   |   |   |
|---|---|---|
| $H_0: \mu \leq \mu_0$<br>$H_1: \mu > \mu_0$ | $H_0: \mu \geq \mu_0$<br>$H_1: \mu < \mu_0$ | $H_0: \mu = \mu_0$<br>$H_1: \mu \neq \mu_0$ |
|---|---|---|

### 3.1. Uso de Excel como apoyo en las clases de teoría mediante ficheros elaborados

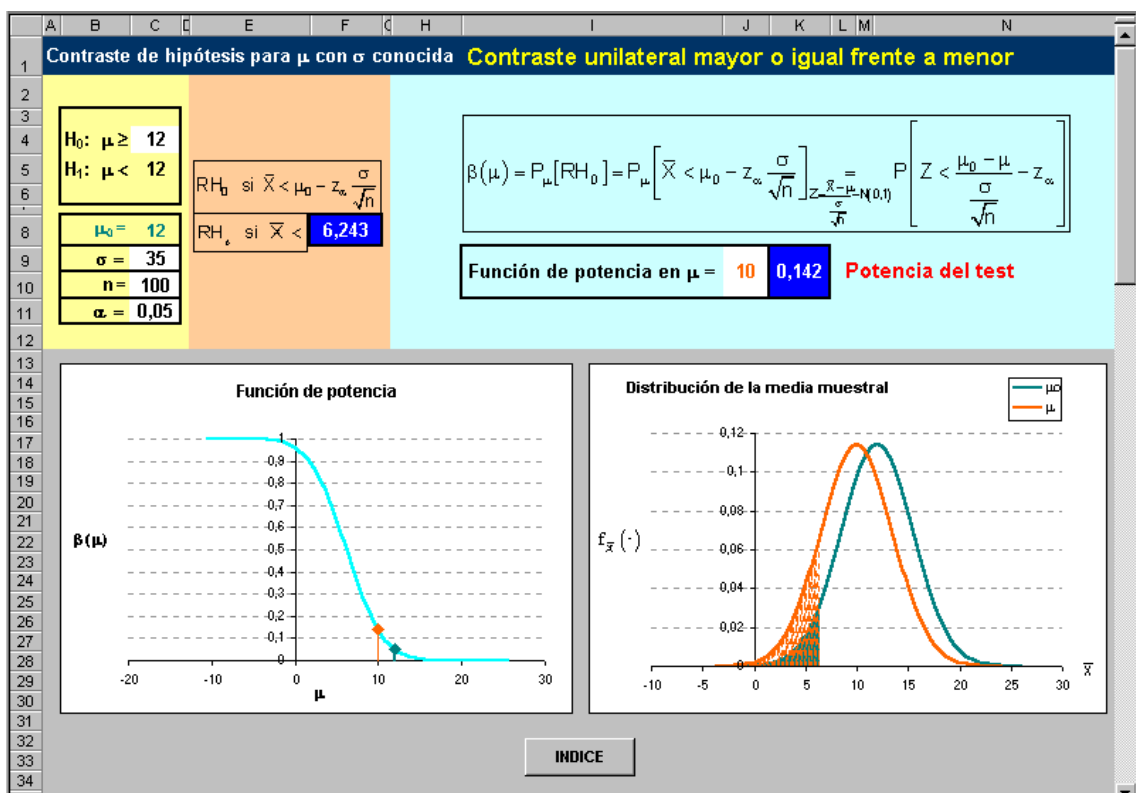
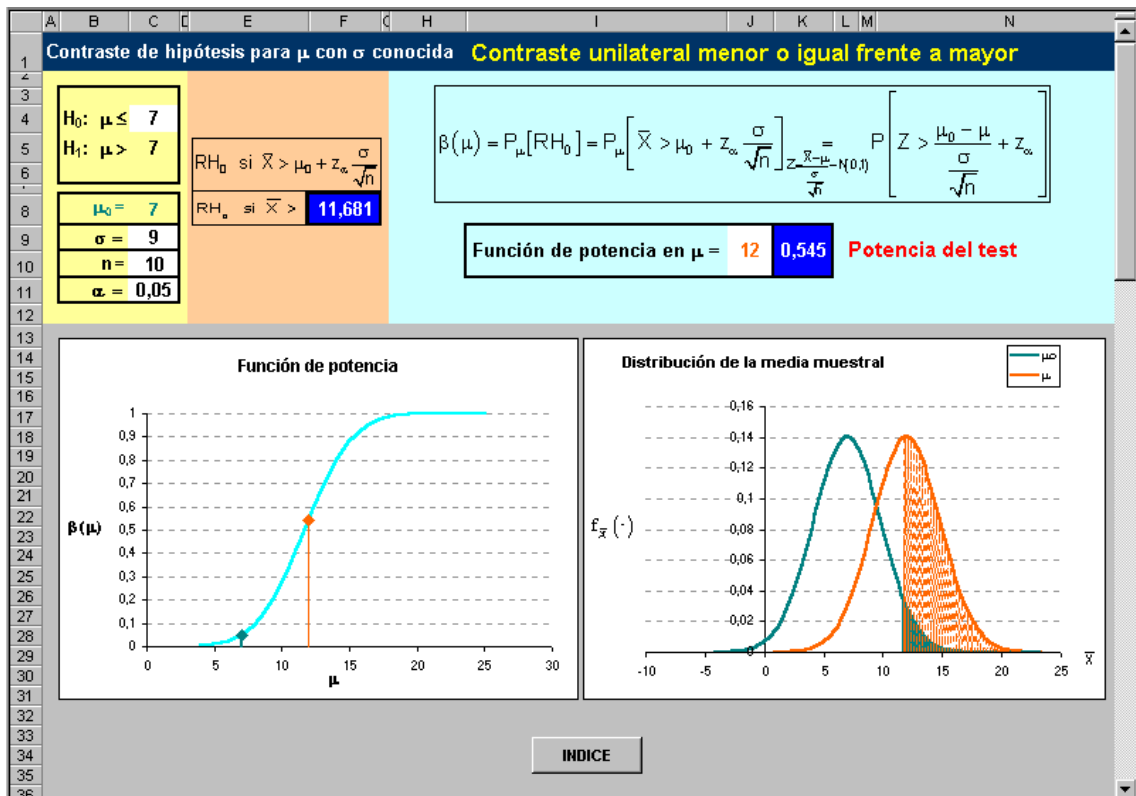
Se ha preparado un fichero de Excel<sup>1</sup> en el que se calcula, para cada uno de los contrastes anteriores y a partir de la información disponible ( $\mu_0$ , tamaño de la muestra  $n$ , varianza de la población  $\sigma^2$  y tamaño del test  $\alpha$ ), la regla de rechazo del correspondiente contraste, así como la función de potencia en un punto concreto del espacio paramétrico. Además, se construyen dos gráficos para cada test que se actualizan al variar alguno de los cuatro elementos anteriores ( $\mu_0$ ,  $n$ ,  $\sigma$ ,  $\alpha$ ) y que representan, respectivamente, la función de potencia y la distribución del estadístico media muestral (en el que se basa el contraste) tanto para  $\mu_0$  como para el valor concreto del parámetro en el que se ha calculado la función de potencia. En el primer gráfico se señala el tamaño del test (valor de la función de potencia en  $\mu_0$ ) y el otro valor de la función de potencia que deseamos destacar. En el segundo gráfico esos valores de la función de potencia se corresponden con las respectivas áreas determinadas por la regla de rechazo para cada una de las distribuciones dibujadas.

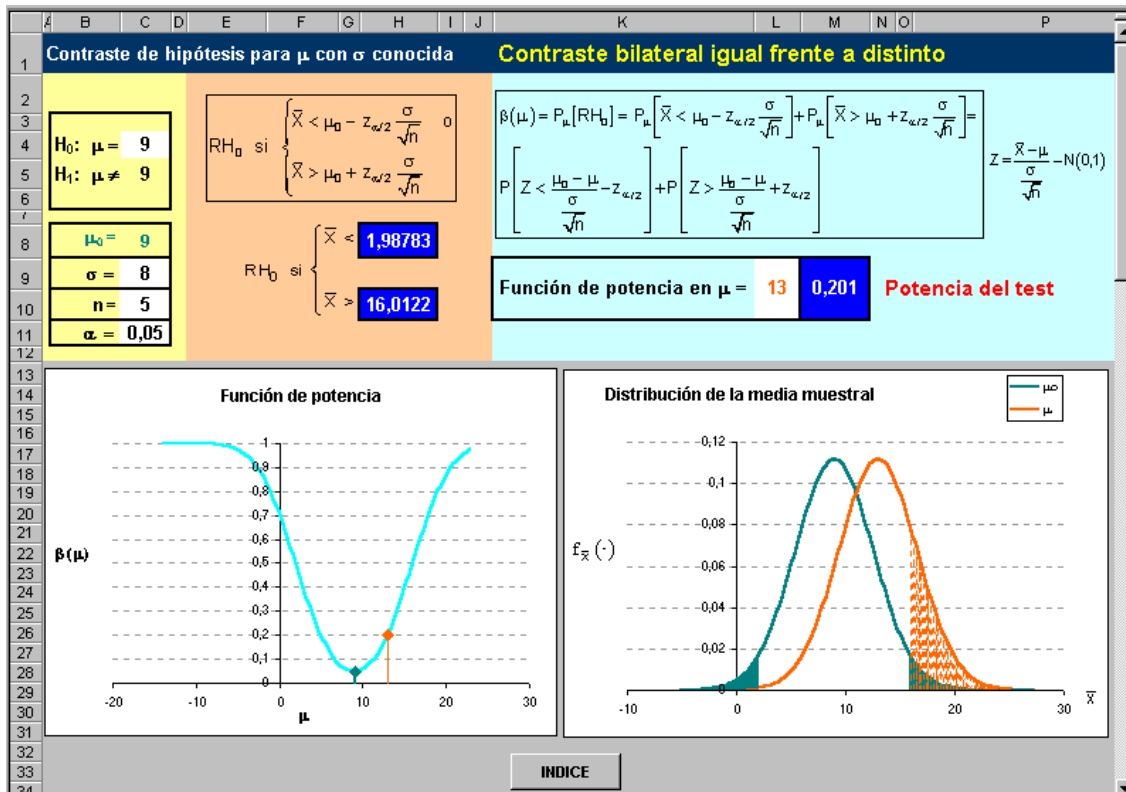
Estos gráficos permiten visualizar la forma de la función de potencia (y cómo dependen tanto la regla de rechazo como la función de potencia de  $\mu_0$ ,  $n$ ,  $\sigma$  y  $\alpha$ ), a la vez que recordar que cada uno de los valores de esta función se corresponde con una probabilidad, representando gráficamente esta última. Por ejemplo, se puede destacar el

---

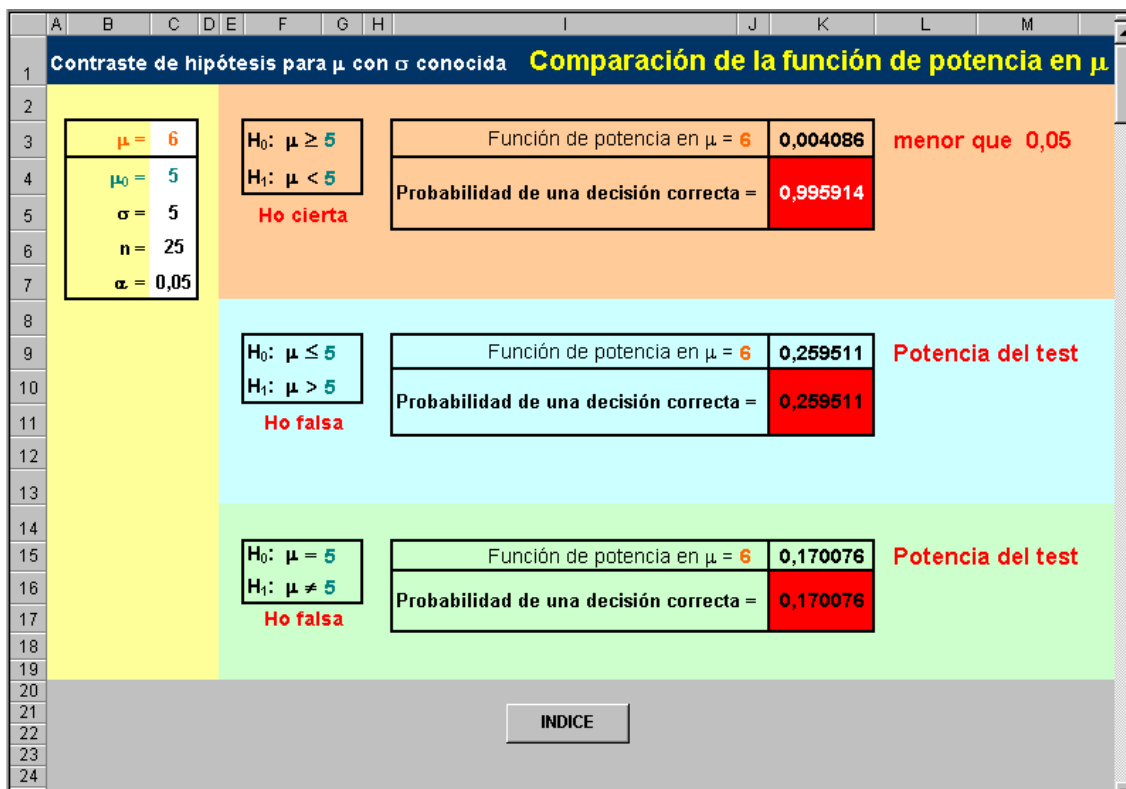
<sup>1</sup> Este fichero incluye una hoja de cálculo con el índice de todo su contenido y con botones de enlace a las distintas partes del mismo.

hecho de que, a mayor separación entre los valores incluidos en la hipótesis nula y el verdadero valor del parámetro, mayor es el valor de la función de potencia, es decir, mayor es la probabilidad de rechazar la hipótesis nula (supuesto que  $\mu_0, n, \sigma$  y  $\alpha$  permanecen fijos).





Otro aspecto interesante consiste en estudiar la probabilidad de tomar la decisión correcta en cada uno de los tests de hipótesis objeto de estudio para un mismo valor de  $\mu_0$ . Los alumnos pueden observar que cuando el verdadero valor del parámetro está en la hipótesis nula la probabilidad de tomar la decisión correcta es mayor.





En la hoja que hemos preparado se resalta con letra blanca la probabilidad más alta y se observa que siempre corresponde al contraste que incluye el verdadero valor del parámetro en su hipótesis nula. Además, el alumno puede experimentar qué ocurre con estas probabilidades para distintas varianzas, tamaños muestrales y tamaños de test. Siempre se obtiene la misma solución cuando  $\alpha < 0,5$ , aunque las probabilidades se incrementan, en general, cuando se aumenta el tamaño muestral y/o la varianza se reduce.

### 3.2. Uso de Excel en las clases prácticas para afianzar los conceptos vistos en clase de teoría

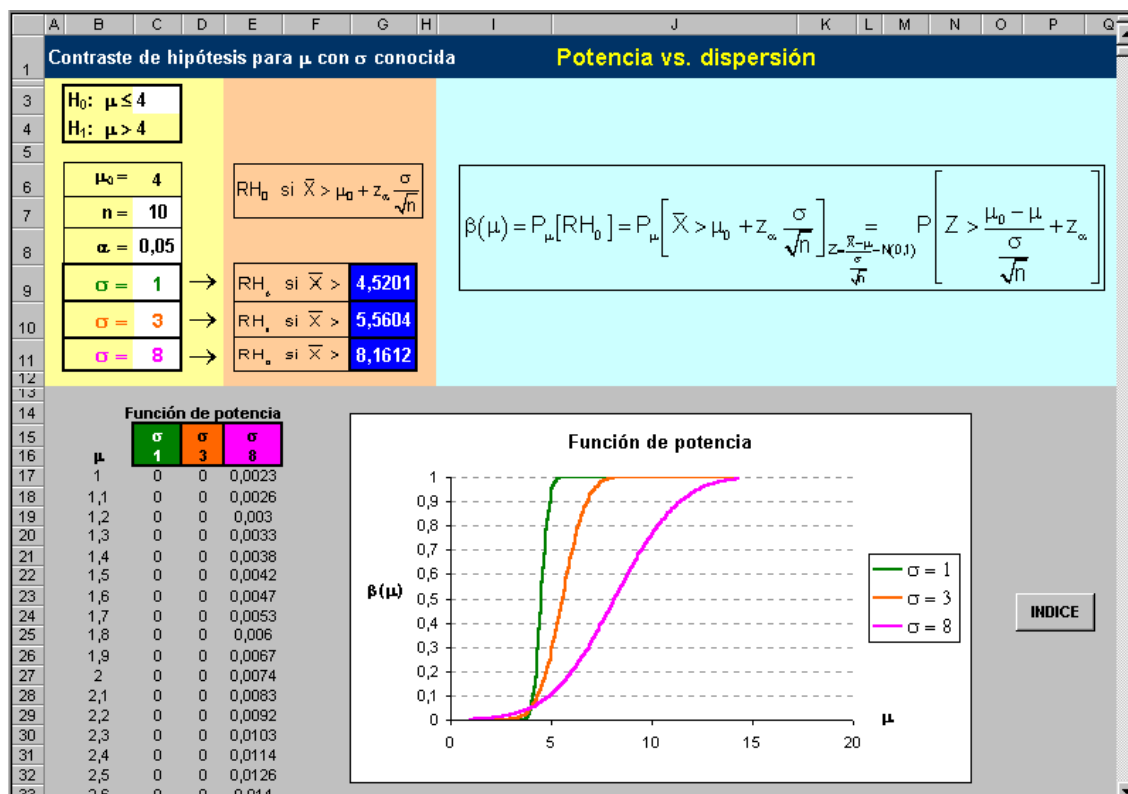
Como ya se ha dicho, para conseguir los gráficos anteriores que cambian con los valores de  $\mu_0$ ,  $n$ ,  $\sigma$  y  $\alpha$  se construyen ficheros que exceden los conocimientos sobre hoja de cálculo que transmitimos a los alumnos en nuestras clases prácticas. En estas últimas creemos que es suficiente con que sepan construir y dibujar la función de potencia asociada a un contraste dado.

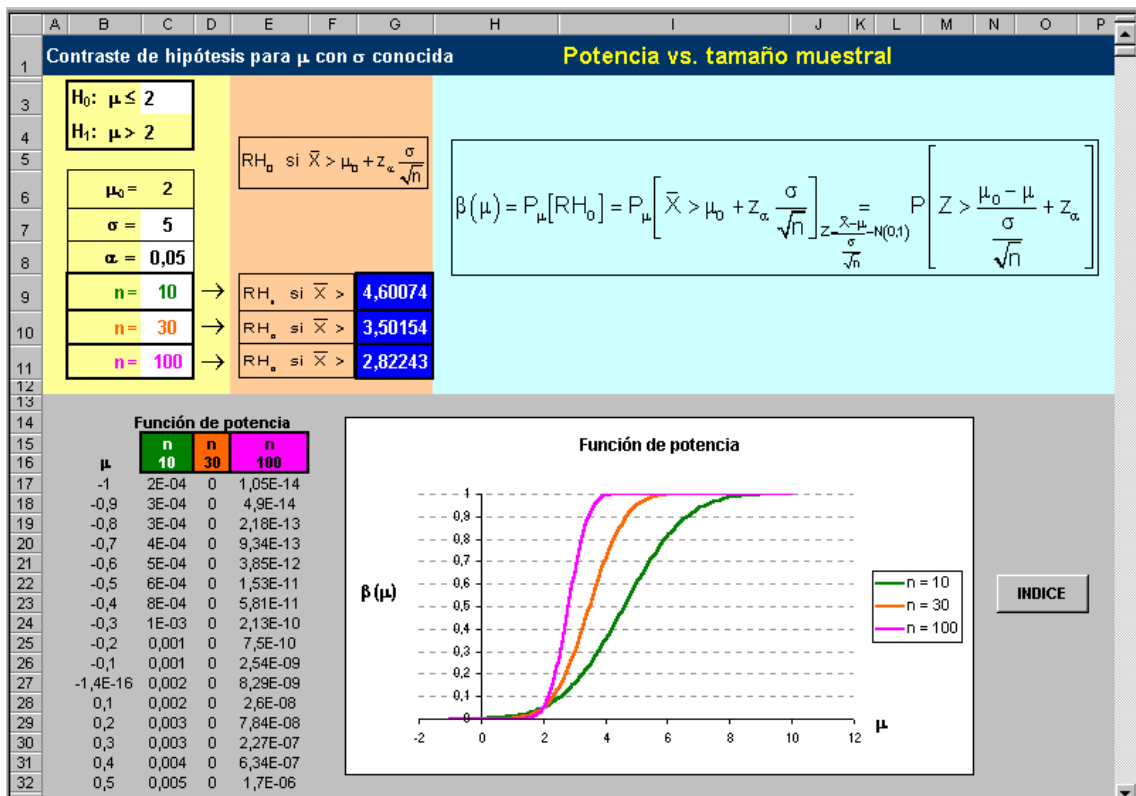
En el caso particular que estamos estudiando de contrastes sobre la media de una población normal con varianza conocida, para obtener la región de rechazo y la función de potencia asociada a cada uno de los contrastes anteriores para valores determinados de  $\mu_0$ ,  $n$ ,  $\sigma$  y  $\alpha$  sólo sería necesario utilizar dos funciones de Excel relacionadas con la distribución normal y la opción de llenado automático que permite repetir tantas veces como queramos el cálculo de la probabilidad de rechazar la hipótesis nula para diferentes valores de la media poblacional. En concreto, se usan dos funciones. Por un lado, **DISTR.NORM(x;media;desv\_estándar;acum)** proporciona el valor de la función de distribución de una normal con media el valor especificado para el argumento **media** y con desviación típica igual al valor en **desv\_estándar** si el argumento **acum** es VERDADERO y la función de densidad de la misma normal si dicho argumento es FALSO. Por otro lado, muy relacionada con la anterior, **DISTR.NORM.INV(probabilidad;media;desv\_estándar)** da como salida aquel valor que, bajo la curva normal especificada, deja a la izquierda la probabilidad detallada en el argumento **probabilidad**. Estas funciones fueron explicadas en las prácticas desarrolladas en la asignatura dedicada a la Teoría de la Probabilidad, por lo que los alumnos conocen perfectamente cómo funcionan.

En cuanto al gráfico de la función de potencia, se construye con una de las versiones del *diagrama de dispersión* que proporciona Excel, y también ha sido ya estudiado por los alumnos, concretamente en la asignatura de Estadística Descriptiva.

Además, es interesante calcular la función de potencia (en cualquiera de los tres contrastes que estamos considerando) para diferentes valores de cada uno de los elementos que determinan la distribución del estadístico media muestral (concretamente  $n$  y  $\sigma$ ), y representar conjuntamente esas funciones en un mismo gráfico, para ver cómo afectan dichos cambios a la función de potencia. Por ejemplo, se observa que la potencia del test aumenta cuando aumenta el tamaño muestral o disminuye la varianza poblacional.

Finalmente, se puede instar a los alumnos para que aborden el estudio de la función de potencia para otros contrastes en poblaciones normales, fomentando así el autoaprendizaje.





#### 4. CONCLUSIONES

Las nuevas tecnologías, y en particular la hoja de cálculo Excel, desempeñan un papel destacado no sólo para los futuros profesionales, sino también para el docente porque permite plantear de una forma nueva la didáctica de la estadística, sobre todo en estudios universitarios aplicados como los nuestros. Con el apoyo de esta herramienta se consiguen distintos objetivos: (1) Mejor comprensión de los conceptos (idea intuitiva, visualización,...), así como manejo más ágil de los mismos; (2) aplicaciones prácticas más realistas (integración de la estadística en la vida cotidiana de un profesional de la economía y la empresa); (3) menor grado de rechazo de nuestras asignaturas (mejor predisposición de los alumnos, más motivación, mayor participación en clase); (4) fomentar el aprendizaje activo y la experimentación.

En lo que se refiere a la función de potencia, el uso de Excel permite que el alumno experimente y trabaje de una forma muy práctica y que asimile la utilidad de la misma.

## **5. 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<sup>®</sup> Excel. Madrid: AC.
- CASAS SÁNCHEZ, J.M. (1996). Inferencia estadística para economía y administración de empresas. Madrid: Centro de estudios Ramón Areces.
- PARRA FRUTOS, I. (2003). Estadística empresarial con Microsoft<sup>®</sup> Excel: Problemas de Inferencia. Segunda edición. Madrid: AC.