

# Yendo de la experiencia a la ciencia: Uso práctico de las herramientas estadísticas básicas de Excel

Héctor A. Cámara, Ricardo L. Videla

## Resumen

En el presente trabajo, los autores revisan algunos conceptos básicos sobre variables y procesamiento estadístico para obtener el valor "p" como el Test t de Student y el método de Chi-Cuadrado, así como la forma de calcular Intervalos de Confianza con las herramientas estadísticas del programa informático Excel (Microsoft), con ejemplos y un ejercicio para el lector.

**Palabras Clave:** Bioestadística. Valor "p". Test t Student. Chi-cuadrado. Intervalos de confianza. Excel.

## Abstract

*Going from the experience to the science: Practical use of the basic statistical tools from Excel*

*Authors of this paper review some basic concepts about variables and statistical processing to obtain "p" value, such as Student t-Test and Chi-Square test, as well as a way of obtaining Confidence Intervals with statistical tools from Excel software (Microsoft), providing examples and an exercise to the reader.*

**Key words:** Biostatistics. "p" value. Student t-Test. Chi-Square Test. Confidence intervals. Excel software.

## INTRODUCCIÓN

*"Después de todo, las estadísticas más elevadas, sólo son la expresión numérica del sentido común"*

*(Karl Pearson)*

Los médicos y, dentro de ellos, los radiólogos, estamos acostumbrados a leer continuamente artículos científicos. Pero, al preparar los trabajos propios, sobre todo cuando se acercan las fechas límite de las reuniones científicas, estamos siempre muy preocupados por ilustrar con muchas imágenes y la presentación de los resultados suele limitarse a lo anecdótico, con importantes defectos en el análisis estadístico de los datos obtenidos.

Esta carencia en la formación en bioestadística impacta negativamente en nuestras producciones y dificulta la interpretación de los resultados publicados en investigaciones clínicas<sup>(1)</sup>.

No es objetivo de este artículo avanzar demasiado sobre los conceptos teóricos concernientes a los parámetros estadísticos necesarios para efectuar un correcto análisis de la información obtenida, lo que puede encontrarse en numerosas fuentes bibliográficas. Particularmente remitimos al lector a los excelentes artículos del Dr. Rubén Michaux publicados en este medio<sup>(2,3)</sup>.

Muchas veces nos encontramos ante frases como la siguiente: "En los casos de obstrucción arterial, a 20 pacientes le aplicamos el tratamiento A y tuvimos buenos resultados finales, donde la permeabilidad arterial fue de 75%, mientras que en los 14 pacientes sometidos al tratamiento B, la permeabilidad fue de

sólo 57%". Si estos resultados fueran de nuestra propia casuística, uno inmediatamente debería preguntarse: esta diferencia encontrada, ¿se debe realmente a los diferentes tratamientos efectuados?, es decir, ¿es posible en base a esta demostración tomar conductas para seleccionar un tratamiento determinado?

Este ejemplo y otros que se presentarán más adelante servirán de base para que intentemos explicar en la forma más concreta posible el uso de las herramientas estadísticas de que dispone el programa informático Excel (Microsoft®).

## Variables

Lo primero a tener en cuenta es qué tipo de datos estamos estudiando en nuestros pacientes. Las variables representan indicadores medibles de una característica que puede adquirir más de un valor entre una observación y otra. Estos valores pueden ser diferentes en diferentes lugares, en distintas personas o distintos momentos<sup>(4)</sup>. Existen básicamente dos tipos de variables o dos clases diferentes de datos:

- 1: Categóricos o Cualitativos
- 2: Numéricos o Cuantitativos

Los datos categóricos o cualitativos son aquellos que pueden clasificarse pero no medirse. Las categorías deben ser claramente definidas para que sean excluyentes entre sí (ej: sí/no, éxito/fracaso).

Las variables categóricas o cualitativas se subdividen en: nominales y ordinales.

Se llaman nominales a las categorías que no tienen un orden especial. Ejemplo: varón/mujer; TC/RM/ECO/Rayos X

Se llaman ordinales a las categorías que siguen

	A	B
1	Medición de nódulo tiroideo	
2	Paciente	Diámetro
3	1	6.2
4	2	7.5
5	3	5.6
6	4	8.3
7	5	8.5
8	6	6.8
9	7	6.9
10	8	8.4
11	9	5
12	10	6.4
13	11	5.6
14	12	5.5

Fig. 1. Disposición de los datos en formato de tabla.

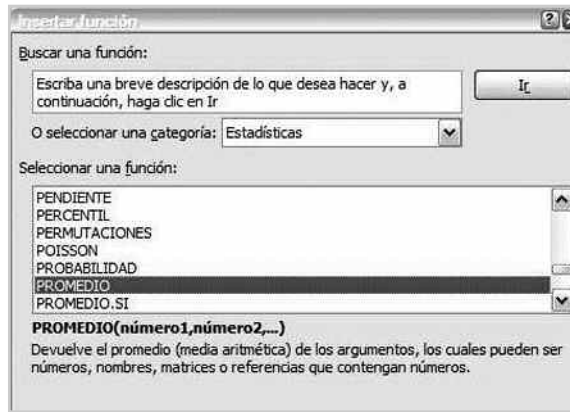


Fig. 2. Cuadro de diálogo emergente para insertar función. De la pestaña Categoría seleccionar las funciones estadísticas y buscar PROMEDIO.

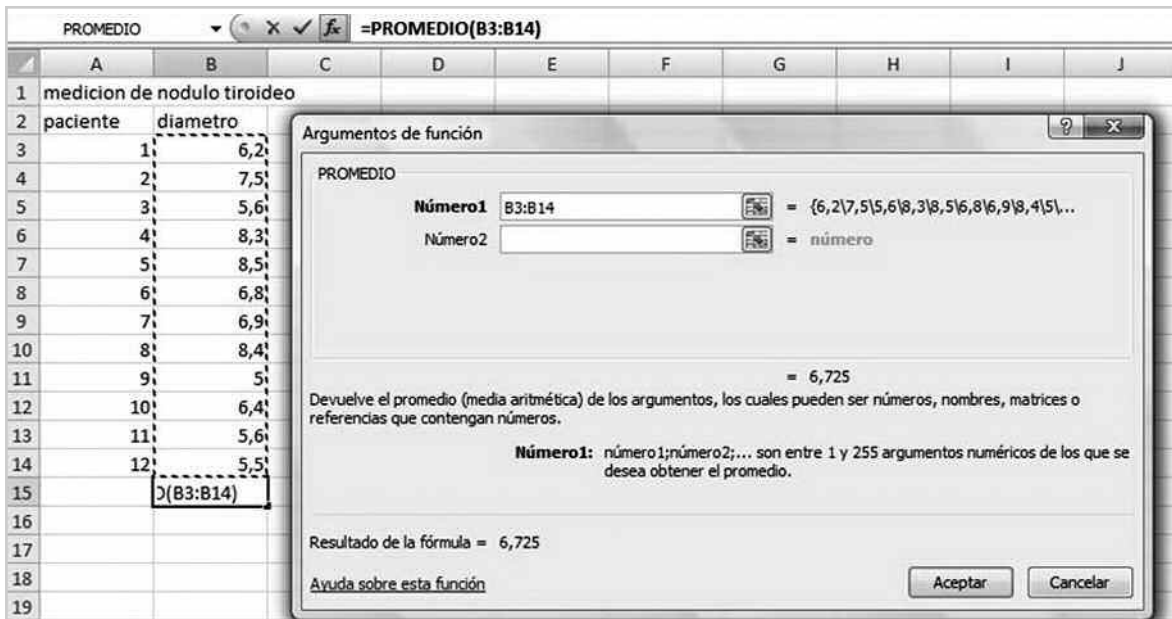


Fig. 3. Cuadro de diálogo de la función PROMEDIO, mostrando la selección de celdas incluidas para promediar.

una escala jerárquica. Ejemplo: muy bueno, bueno, regular, malo; el sistema BiRADS del American College of Radiology para los informes de imágenes mamarias, que establece categorías de malignidad creciente excluyentes entre sí (una lesión no puede ser BIRADS 1, “benigna”, y BIRADS 4, “sospechosa de malignidad” al mismo tiempo).

El otro gran grupo de datos se llama numéricos o cuantitativos, porque pueden tener un valor cualquiera de una escala numérica.

Las variables numéricas o cuantitativas se subdividen en: continuas y discretas.

Variables continuas son aquellas que pueden tomar cualquier valor dentro de una escala. Por ejem-

plo: el peso corporal (58,3 kg, 65,8 kg) o el diámetro de una lesión (2,5 cm, 4 cm, 2,8 cm). Variables discretas son aquellas que no pueden tomar cualquier valor dentro de una escala: uno no puede tener 1,50 hijos <sup>(4)</sup>.

Las estrategias de comparación y las herramientas necesarias son diferentes en cada caso.

Con variables cuantitativas, expondremos los pasos para determinar el promedio y la desviación estándar de una serie de medidas y, posteriormente, demostraremos cómo comparar dos series de datos para obtener el valor “p” y los Intervalos de Confianza. Para las variables cualitativas como en el ejemplo del principio (permeables vs. obstruidos), explicaremos como realizar la prueba de Chi Cuadrado.

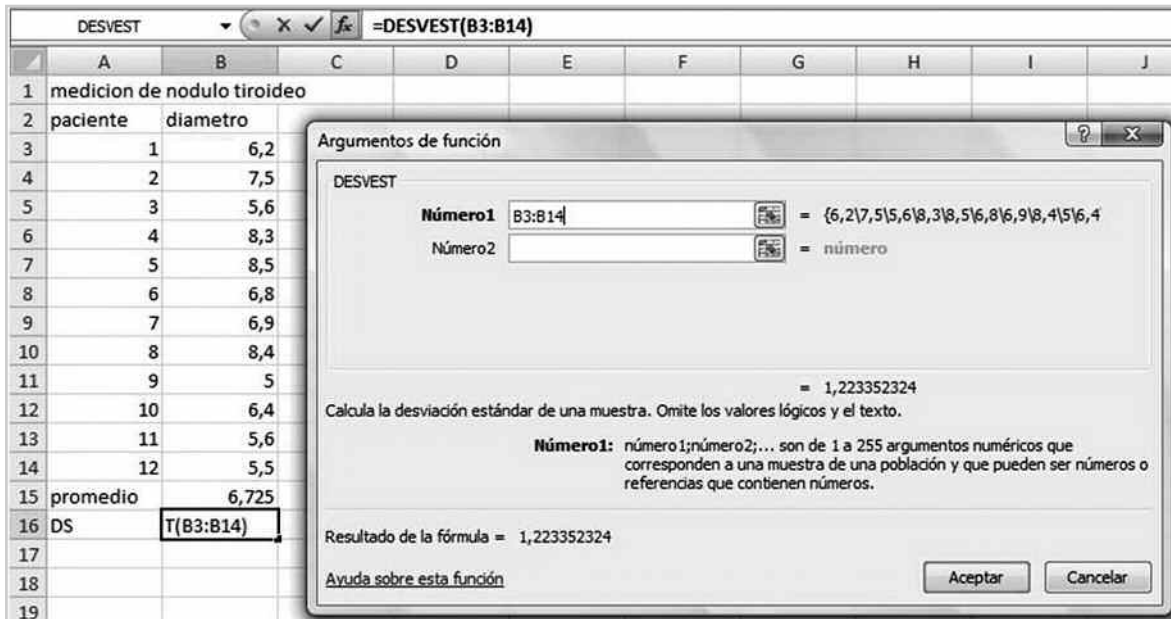


Fig. 4. Función DESVEST, que establece la selección de celdas correspondientes a mediciones ecográficas, para calcular la desviación estándar.

	A	B	C
1	Diámetro quiste menor		
2	TC	ECO	
3	9	9	
4	5	7	
5	5	8	
6	11	11	
7	6	9	
8	8	11	
9	10	10	
10	8	9	
11	7,75	9,25	Promedio
12	2,25198325	1,38873015	DS

Fig. 5. Tabla de datos perteneciente a dos grupos hipotéticos de datos de medición del quiste renal más pequeño visible mediante Ecografía y TAC.

## PROCESAMIENTO DE VARIABLES NUMÉRICAS O CUANTITATIVAS

### Determinación del Promedio y la Desviación Estándar (DS)

Si tenemos una serie de valores medidos, es necesario disponerlos en una tabla, como se muestra en la Figura 1.

A continuación, se debe seleccionar “insertar función” y se abre un cuadro de diálogo como en la Figura 2.

Luego de seleccionada la función PROMEDIO, será necesario determinar cuáles son los números a

promediar, lo que se realiza seleccionando las celdas necesarias (Fig. 3). En esta ventana emergente, una vez determinadas las celdas, Excel ya propone el resultado de la fórmula (6,72, en este caso).

Para determinar la Desviación Estándar (DS) hay que buscar entre las funciones estadísticas la función DESVEST y proceder de similar manera (Fig. 4) (1,22, en este caso).

Con esto hemos alcanzado las medidas de uso más frecuentes para la representación de una variable en la población en estudio: el promedio, que representa una medida de tendencia central, y la DS, que expresa la dispersión de los valores.

En este caso, los nódulos tiroideos miden de promedio  $6,72 \pm 1,22$ .

### Prueba T de significación estadística

Cuando queremos comparar los resultados de dos muestras diferentes, como por ejemplo datos provenientes de ecografías vs. TC o valores obtenidos en diferentes momentos (por ejemplo, control de crecimiento de un nódulo), podemos utilizar el test t de Student, que nos brindará el valor “p”<sup>(5)</sup>.

Vamos a proponer el siguiente ejemplo, con números ficticios a los fines didácticos. En una serie de 8 pacientes con poliquistosis renal, decidimos evaluar el diámetro del quiste menor distinguible con ecografía y tomografía. Nuestra hipótesis es que la TC es capaz de visualizar quistes más pequeños que la ecografía. Procedemos entonces a registrar el diámetro del quiste más pequeño que se observa con cada metodología. Los resultados son los siguientes: En ECO:  $9,25 \pm 1,38$  y en TC:  $7,75 \pm 2,25$  (Fig. 5). Con estos resultados, ¿se podría afirmar que la TC es capaz de visualizar quistes más pequeños que la ecografía? Disponemos,

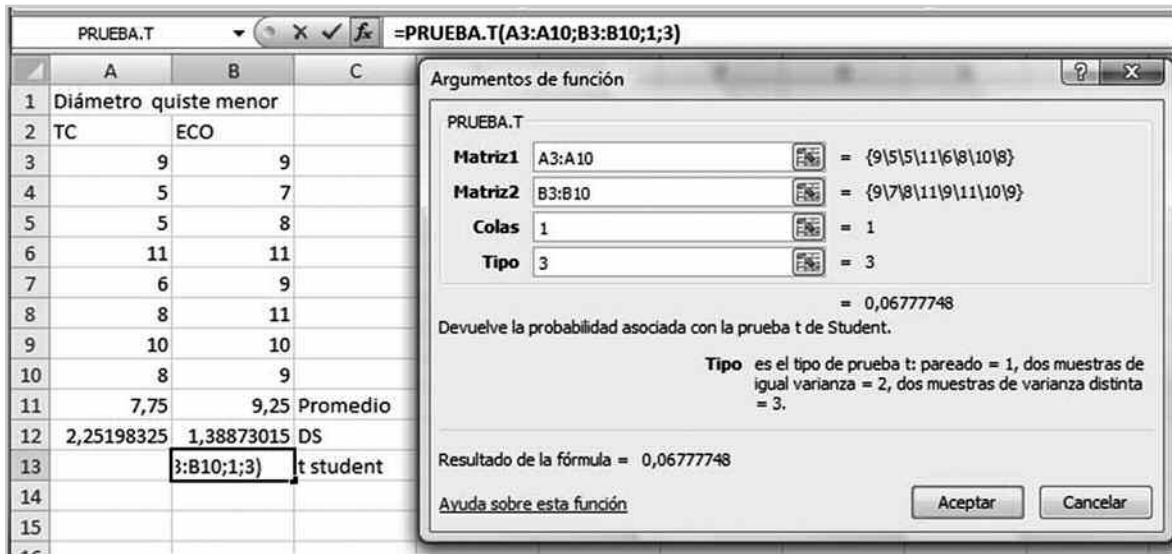


Fig. 6. Cuadro de diálogo de la función PRUEBA T. Ver descripción en el texto.

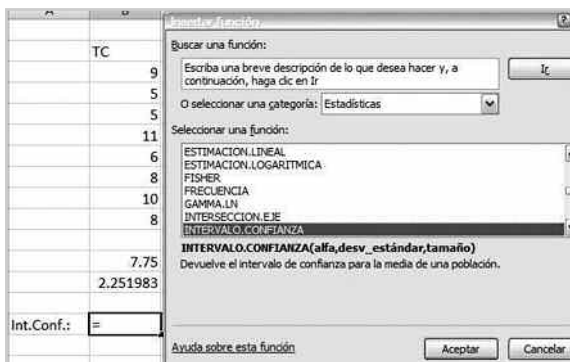


Fig. 7. Cuadro de diálogo de insertar función. En la pestaña Categoría seleccionar funciones estadísticas y a continuación buscar la función INTERVALO.CONFIANZA.

entonces, los datos en forma de tabla, como se ha visto anteriormente.

Para determinar la significación estadística de esta diferencia, utilizaremos la función PRUEBA T. El cuadro de diálogo que se abre (Fig. 6) solicita que ingresemos varios parámetros. En primer lugar, los datos correspondientes a cada muestra. En nuestro ejemplo, matriz 1 corresponde a las medidas obtenidas con TC, y matriz 2, a las medidas obtenidas con ECO. El siguiente parámetro: colas, puede tomar valor uno o dos, de acuerdo con la dirección que se supone que el efecto debe presentar. En nuestro ejemplo, hemos seleccionado una cola, ya que el planteo original supone que la tomografía encuentra quistes más pequeños que la ecografía (es decir, que importa verificar si la hipótesis se cumple sólo en esta dirección).

El último parámetro: tipo de prueba, dependerá

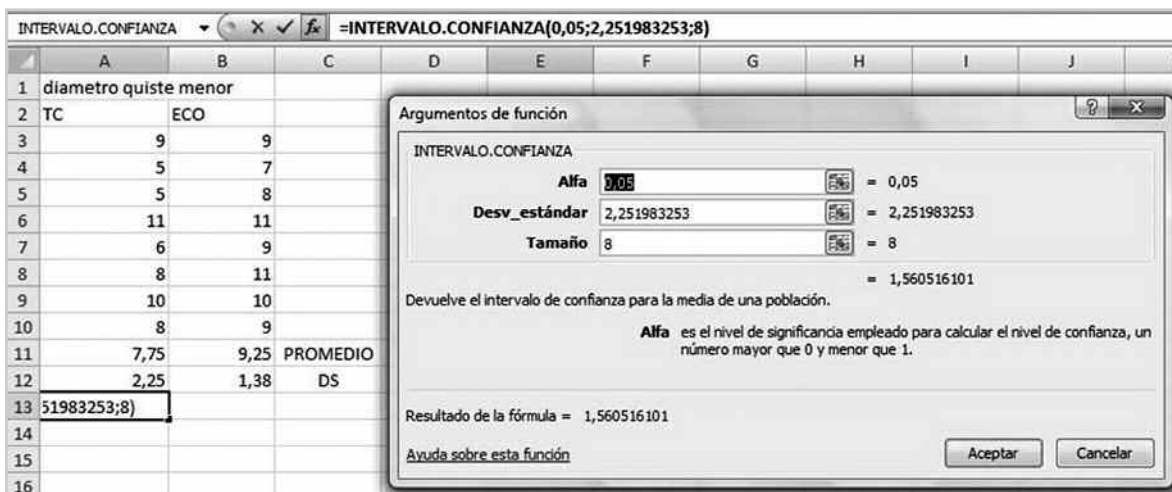


Fig. 8. Cuadro de diálogo de la función INTERVALO.CONFIANZA, ver descripción en el texto.

	A	B	C
1	Diámetro quiste menor		
2	TC	ECO	
3	9	9	
4	5	7	
5	5	8	
6	11	11	
7	6	9	
8	8	11	
9	10	10	
10	8	9	
11	7,75	9,25	PROMEDIO
12	2,251983	1,389	DS
13	1,560516	INTERVALO DE CONFIANZA (IC)	
14			
15	7,75-1,56=6,19	PROMEDIO MENOS IC	
16	7,75+1,56=9,31	PROMEDIO MÁS IC	

Fig. 9. El resultado de la función INTERVALO.CONFIANZA deberá ser sumado y restado al promedio, para obtener los valores extremos del intervalo.

G13				
	A	B	C	D
1	Diámetro quiste menor			
2	TC	ECO		
3	9	9		
4	5	7		
5	5	8		
6	11	11		
7	6	9		
8	8	11		
9	10	10		
10	8	9		
11	7,75	9,25	PROMEDIO	
12	2,251983	1,389	DS	
13	1,560516	0,962	INTERVALO DE CONFIANZA (IC)	
14				
15	6,19	8,29	PROMEDIO MENOS IC	
16	9,31	10,21	PROMEDIO MÁS IC	

Fig. 10. Intervalos de Confianza 95% de los grupos TC y ECO.

Valores Encontrados	Trat. A	Trat. B	Total
Permeables	15	8	23
Obstruidos	5	6	11
Total	20	14	34

	Trat. A	Trat. B
Permeables	75%	57%
Obstruidos	25%	43%
Total Paciente	20	14

Fig. 11. Tabla de sumario de los datos en los grupos en estudio. A la derecha se presenta la expresión en porcentajes y a la izquierda la tabla original de los datos propiamente dichos.

	A	B	C	D
1	Valores encontrados	Trat. A	Trat. B	Total
2	Permeables	15	8	23
3	Obstruidos	5	6	11
4	Total	20	14	34
5				
6	Valores esperados	Trat. A	Trat. B	Total
7	Permeables	(20x23)/34=14	(14x23)/34=9	23
8	Obstruidos	(20x11)/34=6	(14x11)/34=5	11
9	Total	20	14	34
10				
11	Valores esperados	Trat. A	Trat. B	Total
12	Permeables	14	9	23
13	Obstruidos	6	5	11
14	Total	20	14	34
15				

Fig. 12. Tabla de valores esperados. Cada celda interna se calcula mediante la operación: fila x columna / total global. Se presentan los números enteros disminuyendo a cero el número de decimales.

del tipo de datos recolectados. En una prueba de tipo pareada, los datos representan una variable “antes y después”, es decir que, en realidad, contamos con un par de datos por cada paciente (ej: evaluación de crecimiento de un nódulo, etc.). Tipo de prueba: 1.

En caso de que nuestros datos provengan de diferentes pacientes, se deberá decidir el tipo de prueba

en base a la varianza de cada grupo, lo que lograremos mediante la comparación de las respectivas DS (recordemos que  $VARIANZA=DS^2$ ). Tipo de prueba 2 cuando la varianza es la misma.

En caso de dos muestras con diferente varianza, como en el ejemplo presentado, es necesario seleccionar Tipo de prueba 3.

El resultado de la función PRUEBA T es la probabilidad o valor “p”. En nuestro ejemplo, esta probabilidad es > 0,05 lo que permite rechazar la hipótesis planteada, es decir, que la TC no mostró quistes más pequeños que la ECO.

El rechazo de la hipótesis alternativa en este caso probablemente se encuentra relacionado con el escaso número de casos, pero el objeto de esta demostración apunta al uso de la función estadística y no a la interpretación del resultado de la comparación.

### Intervalos de Confianza

El valor “p” sólo informa sobre la significación estadística, es decir, cuál es la probabilidad de que las diferencias encontradas sean debidas al azar. Lamentablemente, la interpretación de los resultados basados en el valor “p” carece por completo del significado clínico de los hallazgos<sup>(6)</sup>. Por ejemplo, un estudio de un dispositivo médico en experimentación animal que informe un tratamiento con efecto estadísticamente significativo ( $p < 0,05$ ) no tendría ningún valor

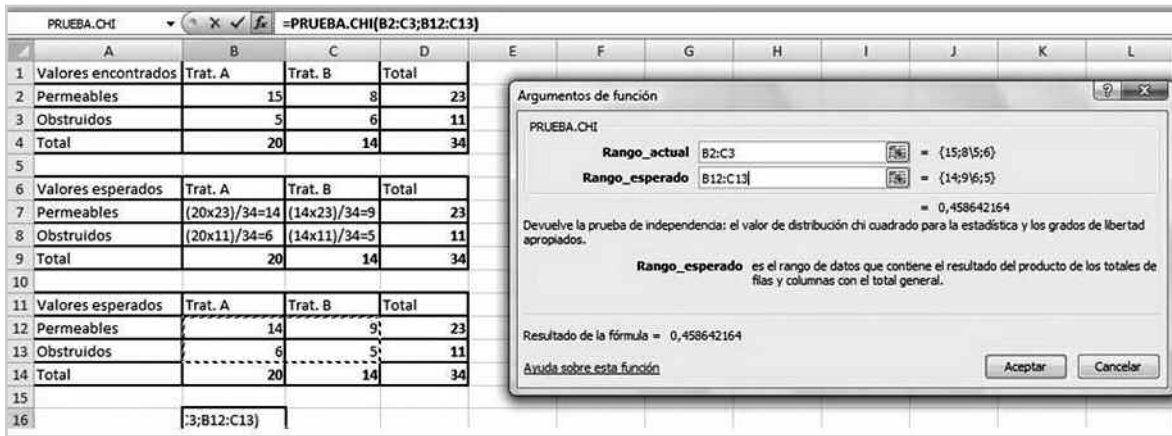


Fig. 13. Función PRUEBA.CHI. Cuadro de diálogo: los rangos actual y esperado corresponden a las matrices de los datos provenientes de las tablas (incluir sólo los datos centrales, sin los totales marginales). El resultado corresponde al valor p.

	A	B	C
1	Estenosis acalasia primaria	Estenosis acalasia maligna	
2		3	2,8
3		2	2,7
4	2,2		4,2
5	2		4
6	3		3,8
7	4		4,8
8	3		3,5
9	2,6		3,6
10	1,9		2,9
11	1,8		3,8
12			
13	2,55	3,61	Promedio
14	0,701189466	0,665749369	DS
15	0,434593747	0,412628152	IC
16		0,002761196	Prueba t
17			
18	La longitud de la estenosis es un parámetro útil. La p es < 0,05		

Fig. 14. Solución a la actividad sugerida.

clínico si el modelo animal no reproduce de manera precisa las condiciones en el ser humano <sup>(7)</sup>.

El uso de los intervalos de confianza (IC) permite una mejor visualización del significado clínico de los resultados y nos da un enfoque alternativo a la tradicional "comprobación de hipótesis" y su consecuente valor "p". Un IC se expresa con dos valores extremos de las unidades medidas y representa un intervalo de forma tal que una serie de valores específicos de la muestra estadística probablemente contenga el valor real de la población con un nivel de confianza específico, habitualmente 95% <sup>(6)</sup>.

En forma similar al valor "p", los IC informan sobre la significación estadística: si los grupos son diferentes, los intervalos no se superponen, es decir, que si estamos 95% seguros de que la diferencia entre dos grupos es real y no debida al azar, entonces podemos decir que la diferencia es "significativa". Pero, además, los IC dan una idea de la precisión (porque se

informa el rango de los intervalos) y principalmente del significado clínico (traducido por la magnitud de los valores) <sup>(8, 9)</sup>.

La determinación de IC comienza con la función INTERVALO CONFIANZA del submenú de funciones estadísticas (Fig. 7). Una vez seleccionada la función, se presenta un cuadro de diálogo que solicita tres parámetros para calcular el intervalo de confianza de la tabla de datos correspondiente. Como muestra la Figura 8, es necesario indicar la celda que contiene la DS, establecer el error Alfa, que es el margen de confianza de nuestro intervalo (habitualmente 5%, por lo que hay que introducir 0,05), y por último, el número de casos o tamaño de la muestra.

El resultado de la función (Fig. 9) corresponde al valor que es necesario adicionar y restar al promedio para obtener los extremos del intervalo. En nuestro ejemplo, el promedio por TC es 7,75. El resultado de la función IC es 1,56. Entonces, a 7,75 le restamos 1,56 y obtenemos el margen inferior del IC, mientras que el margen superior resulta de sumar al promedio 7,75 el mismo valor 1,56. Con esto podemos afirmar que el 95% de los quistes visualizados con TC miden entre: 6,19 y 9,31 cm.

Una vez obtenido el resultado de los márgenes del intervalo para cada grupo (Fig. 10), pueden expresarse los resultados en la forma correspondiente: en este ejemplo, los grupos no mostraron diferencias significativas (TC: 6,16-9,31cm / ECO: 8,21-10,21cm, IC 95%).

## PROCESAMIENTO DE VARIABLES CUALITATIVAS

### Prueba de Chi Cuadrado

Para las variables cualitativas existen varias pruebas estadísticas, pero la más común y tal vez la más sencilla es la comprobación de Chi cuadrado. Esta comprobación evalúa si la distribución encontrada de la variable puede deberse al azar. Retomemos el ejem-

plo del principio. De 20 pacientes con tratamiento A, se encontró permeable al 75%, y en 14 pacientes sometidos al tratamiento B, se encontró permeable al 57%. Podemos construir entonces la tabla de la Figura 11.

Es necesario a continuación calcular la tabla de valores teóricos esperables, o sea, la distribución (permeables/obstruidos) que existiría si la distribución dependiera exclusivamente del azar. Cada uno de los "valores esperables" se obtiene en base a los totales de nuestra observación, multiplicando filas por columnas y dividiendo por el total general. Es decir, para obtener el valor de cada celda se multiplica el total correspondiente a su fila por el total correspondiente a su columna dividiendo el resultado por el total general (Fig.12).

Una vez dispuestas ambas tablas, la de los valores encontrados (nuestros hallazgos) y la de los valores esperados (la tabla calculada), podemos entonces activar la función PRUEBA CHI. El cuadro de diálogo emergente (Fig. 13) muestra dos parámetros básicos a completar. La matriz rango \_ actual corresponde a las celdas de los valores reales encontrados por nuestra investigación <sup>(5, 6, 8)</sup>. La matriz rango\_esperado corresponde a los datos de la tabla calculada <sup>(5, 6, 9)</sup>. En ambos casos, la matriz debe incluir sólo los datos, sin los totales marginales. El resultado de la función PRUEBA CHI, expresa el valor "p" o probabilidad de que los resultados se encuentren debidos al azar. En nuestro ejemplo, el resultado es 0,45, es decir, la  $p > 0,05$  por lo que podemos afirmar que no existen diferencias significativas entre los tratamientos (es decir, que las diferencias en los tratamientos puede ser debida al azar).

#### Actividad sugerida

Contamos con una muestra de 10 pacientes portadores de acalasia primaria y 10 pacientes con pseudoacalasia maligna. Las estenosis en los pacientes con acalasia primaria midieron: 3; 2; 2,2; 2; 3; 4; 3; 2,6; 1,9 y 1,8 cm. En los pacientes con pseudoacalasia, las estenosis midieron: 2,8; 2,7; 4,2; 4,0; 3,8; 4,8; 3,5; 3,6; 2,9 y 3,8cm. Con estos parámetros, ¿Es posible afirmar que la longitud de la estenosis es un indicador útil (aunque no el único) para determinar si la lesión puede ser

benigna o maligna? Habrá que calcular Promedio, DS, IC y valor "p" con Test t de Student (Figura 14).

## CONCLUSIÓN

Hemos presentado en la forma más práctica posible la manera de obtener el promedio y la Desviación Estándar de un grupo de mediciones, expusimos el significado y la forma de obtener Intervalos de Confianza y planteamos dos formas diferentes de obtener el valor "p" de acuerdo con el tipo de variable en estudio; Test t de Student para variantes cuantitativas y la prueba de Chi cuadrado para variantes cualitativas, con las herramientas disponibles en Excel de Microsoft.

Agradecimientos: A Lucía Videla que nos dio una mano para tratar de hacer entendible un tema complejo.

## Bibliografía

1. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA*, 2007;298:1010-1022.
2. Michaux RP. Conceptos estadísticos básicos: Una aproximación teórico-práctica (parte 1). *Rev Argent Radiol* 2004;68:423-427.
3. Michaux RP. Conceptos estadísticos básicos: Una aproximación teórico-práctica (parte II). *Rev Argent Radiol* 2005;69:57-63.
4. Applegate KE, Crewson PE. An introduction to biostatistics. *Radiology* 2002;225:318-322.
5. Castiglia VC, Castiglia NI. El significado de la "p" y el consenso de la comunidad científica. *Síntesis* 2001;12:18-20.
6. Cordova ML. Giving clinicians more to work with: let's incorporate confidence intervals into our data. *J Athl Train* 2007;42:445.
7. Cloft HJ. The value of a "p" value. *Am J Neuroradiol* 2006;27:1389-1390.
8. Kocher MS, Zurakowski D. Clinical epidemiology and biostatistics: a primer for orthopaedic surgeons. *Bone Joint Surg Am* 2004;86:607-620.
9. Godwin M. Hypothesis: the research page. Part 2: Confidence intervals and P values. *Can Fam Physician* 2001;47:1044-1045 .