

# Estadística en cirugía, cómo entender y aplicar conceptos básicos

## *Statistics in surgery, how to understand and apply basic concepts*

Eric M. Saucedo-Moreno\* y José Fenig-Rodríguez

Servicio de Cirugía General, Hospital Ángeles, MOCEL, Ciudad de México, México

### Resumen

*Se estima que un lector habitual de publicaciones biomédicas tendrá acceso a un 70% de revistas con procedimientos bivariados con estadística básica, es decir, cualquier profesional de la salud con conocimientos básicos es capaz de entender la gran mayoría de los estudios publicados. Por eso surge la necesidad de escribir este texto que pretende dar una visión general de la estadística aplicada a la cirugía, haciendo referencia a cómo y cuándo utilizar cada procedimiento estadístico y, por supuesto, cómo interpretar los resultados. Manejaremos definiciones de conceptos como variables, hipótesis, intervalos de confianza, análisis multivariados, contraste de hipótesis, pruebas paramétricas y no paramétricas, y regresión logística. La importancia de esta guía es que todos los profesionales de la salud puedan interpretar de manera sencilla y concreta los diferentes estudios estadísticos en su desarrollo profesional.*

**PALABRAS CLAVE:** Estadística. Métodos no paramétricos. Métodos paramétricos.

### Abstract

*It is estimated that a regular reader of biomedical publications will have access to 70% of journals with bivariate procedures with basic statistics, that is, any health professional with basic knowledge is capable of understanding the vast majority of published studies. That is why the need arises to write this text that aims to give an overview of the statistics applied to surgery, making reference to how and when to use each statistical procedure and of course how to interpret the results. We will handle definitions of concepts such as variables, hypotheses, confidence intervals, multivariate analysis, contrast of hypotheses, parametric and nonparametric methods and logistic regression. The importance of this guide is that all health professionals can interpret in a simple and concrete way the different statistical studies in their professional development.*

**KEY WORDS:** Statistics. Nonparametric methods. Parametric methods.

### Introducción

Se estima que un lector habitual de publicaciones biomédicas tendrá acceso a un 70% de revistas con procedimientos bivariados solo con estadística básica,

es decir, cualquier profesional de la salud con conocimientos básicos es capaz de entender la gran mayoría de los estudios publicados<sup>1</sup>. Por eso decidimos escribir esta revisión, que pretende dar una visión básica de la estadística aplicada, tratando de hacer referencia a cómo, cuándo y qué significa utilizar cada

#### Correspondencia:

\*Eric M. Saucedo-Moreno.

Gobernador Gregorio Villa Gelati, 29

San Miguel Chapultepec, 1 Sección, Del. Miguel Hidalgo

C.P. 11850, Ciudad de México, México

E-mail: eric.saucedo.m@gmail.com

Fecha de recepción: 26-05-2018

Fecha de aceptación: 30-01-2019

DOI: 10.24875/CIRU.19000487

Cir Cir. 2019;87:692-697

www.cirugiaycirujanos.com

proceso estadístico en un estudio biomédico. Todos los datos presentados a continuación no son reales y solo se han utilizado para ejemplificar los conceptos.

## **Variables**

De la misma manera que una nefrona es una unidad funcional en el riñón, en estadística la unidad funcional son las variables. Son cualidades o cantidades de cada individuo, y se llaman «variables» porque varían de un sujeto a otro<sup>1-5</sup>.

### ***Variables cualitativas o categóricas nominales***

Hacen referencia a las características que posee, o no, un individuo en un estudio. Pueden dividirse en dicotómicas o policotómicas, y permiten clasificar de forma excluyente a los individuos, pero no es posible ordenar ni jerarquizar. Ejemplo: el sexo, masculino o femenino; si un sujeto es masculino no puede ser femenino, sin tener mayor o menor valor. Las variables cualitativas ordinales son aquellas características que pueden jerarquizarse y ordenarse utilizando escalas ordinales. Ejemplo: queremos conocer si los hombres hacen más ejercicio que las mujeres. La intensidad de la actividad se categoriza en baja, moderada y alta. Los sujetos clasificados en alta presentan mayor actividad que los clasificados en baja, pero esto no es el doble, sino que solo es mayor<sup>1,4,5</sup>.

### ***Variables cuantitativas***

Se llaman variables cuantitativas cuando los números que se utilizan para codificarlas equivalen con exactitud a los verdaderos datos<sup>1,4,5</sup>. Aquí se encuentran las variables discretas, que utilizan números enteros y finitos. Ejemplo: número de días que realizan actividad física a la semana: 1 día, 2 días, 4 días, etc., pero no 1.5 días a la semana, y solo hay 7 días a la semana<sup>1,4,5</sup>. Las variables continuas son aquellas que utilizan números decimales, infinitos y que pueden valer cualquier cantidad entre dos posibles valores. Ejemplo: el tiempo que se realiza actividad física: 60 minutos, 60.30 minutos, 60.40 minutos, etc.<sup>1,4,5</sup>.

## **Intervalos de confianza**

Una de los objetivos al escribir un estudio de investigación es que tanto la información recolectada como la estimada sea lo más exacta posible y con el menor

error posible. Uno de los errores que pueden cometerse al medir es el error accidental o aleatorio, que se debe a causas imposibles de controlar. Este error se estima calculando los intervalos de confianza (IC) y aplicando pruebas de contraste de hipótesis.

Un intervalo de confianza es un rango de valores en el que confiamos que se contenga el parámetro de la población donde se obtuvo la muestra. Supongamos que se realiza un estudio en el que se pretende determinar cuál es el medicamento que presenta mayores índices de mejoría en las náuseas posquirúrgicas. Se encuesta a 1000 pacientes que utilizaron metoclopramida, ondansetrón o palonosetrón. De ellos, 60 pacientes mencionan que el mejor resultado lo obtuvieron con ondansetrón. Interpretamos que el 60% de los pacientes mejoran con ondansetrón. Sin embargo, no es totalmente confiable pensar que el 60% de toda la población mejora con ondansetrón. Por ello, estimamos un intervalo en el cual confiamos que podremos encontrar el valor real<sup>1,2,4</sup>. En el mismo estudio se reporta un intervalo de confianza del 95% de 58-62%; interpretaremos que entre el 58 y el 62% de la población se encuentra el valor real del porcentaje de pacientes que mejoran con el uso de ondansetrón. Esto quiere decir que si repetimos el mismo estudio 100 veces, con 100 pacientes y con el mismo IC, tendríamos 95 estudios que contendrían el mismo resultado y 5 que no<sup>1,2,4</sup>.

## **Contraste de hipótesis**

Seguramente en algún momento hemos leído o escuchado los conceptos de hipótesis nula e hipótesis alterna. La hipótesis nula ( $H_0$ ) hace referencia a que el efecto de interés no existe en la población de donde procede la muestra. Ejemplo: no existe diferencia en la mejoría de las náuseas posquirúrgicas entre los pacientes que utilizan metoclopramida, ondansetrón o palonosetrón. La hipótesis alterna ( $H_1$ ) hace referencia a que el efecto existe en la población de donde procede la muestra. Ejemplo: existe diferencia en la mejoría de las náuseas posquirúrgicas entre los pacientes que utilizan metoclopramida, ondansetrón o palonosetrón<sup>1,4</sup>. Cuando se realiza un estudio de investigación se establecen estas dos hipótesis de manera arbitraria y lo que se busca es compararlas entre sí, con la finalidad de demostrar cuál es cierta.

El siguiente paso es calcular la «p». Interpretar de forma incorrecta el significado de «p» es un error muy común. La definición estricta de «p» es la probabilidad de rechazar la hipótesis nula cuando esta es cierta. En

otras palabras, la probabilidad de observar diferencias entre los datos de la muestra cuando en realidad no los hay. Si el valor de «p» es muy alto quiere decir que no hay datos confiables para rechazar la hipótesis nula, y entonces muy probablemente no haya diferencias. Es importante dejar en claro que la «p» es una forma de comunicar los resultados de un estudio, pero no define por sí misma la importancia práctica de sus resultados<sup>1,2,4</sup>.

Al comparar las hipótesis pueden cometerse dos tipos de errores:

- Tipo 1: decir que existen diferencias cuando en realidad no las hay. La probabilidad de cometer este error se conoce como riesgo alfa<sup>1,2,4</sup>.
- Tipo 2: decir que no existen diferencias cuando realmente las hay (aceptar la  $H_0$  cuando debió rechazarse). Una de las causas para cometer este error es que el número de sujetos estudiados sea escaso para demostrar el efecto, o que el efecto sea pequeño. El riesgo de cometer un error de tipo 2 se conocido como riesgo beta<sup>1,2,4</sup>.

Al hablar de estos errores y riesgos también hay que hablar de la potencia estadística, que es la capacidad de una prueba para detectar una diferencia cuando existe. Una de las causas para que un estudio presente baja potencia es un tamaño de muestra insuficiente; en otras palabras, a mayor tamaño muestral, menor riesgo alfa y beta, mayor potencia estadística e IC mucho más estrechos<sup>1,2,4</sup>.

### Pruebas paramétricas y no paramétricas

Antes de comentar las pruebas paramétricas hay que entender el concepto de «distribución». La distribución que tiene una variable significa cuán lejos se encuentra el valor de la media. Cuanto mayor sea el número, mayor será la variabilidad; cuanto menor sea, más homogéneo. Cuando un estudio tiene una normalidad, igualdad u homogeneidad de varianza entre los grupos se utilizan pruebas paramétricas. En caso de que los resultados no tengan estas características, se utilizan pruebas no paramétricas, en las cuales se transforman los datos originales en rangos<sup>4</sup>. Se entenderá mejor el concepto de varianza y rangos cuando hablemos de medias y medianas.

### Datos categóricos y porcentajes

Las variables categóricas se presentan con proporciones o percentiles<sup>4</sup>. Ejemplo: si leemos que el 60% de los pacientes son hipertensos y el 40% no son

**Tabla 1. Test de ji al cuadrado en el estudio de pacientes con cáncer de piel**

Pacientes	Cáncer de piel	No cáncer de piel	Total
Sin exposición al sol	342	1175	2117
Con exposición al sol	84	238	320
Total	424	2013	2437

**Tabla 2. Test exacto de Fisher en el estudio de los pacientes con cáncer de piel**

Pacientes	Cáncer de piel	No cáncer de piel	Total
Sin exposición al sol	5	6	11
Con exposición al sol	14	5	19
Total	19	11	30

hipertensos, deberíamos interpretar que la proporción es de 0.60 y 0.40. Si lo que queremos es comparar entre dos muestras la proporción de sucesos, las pruebas más comunes para realizar estas comparaciones son el test de ji al cuadrado ( $\chi^2$ ) de Pearson y el test exacto de Fisher<sup>4</sup>. Ejemplo: supongamos que en un estudio se investiga cuál es la relación entre la exposición al sol y el riesgo de desarrollar cáncer de piel (Tabla 1). De 2437 pacientes estudiados, solo 342 (16%) no tuvieron exposición al sol y desarrollaron cáncer. De manera contraria, 84 pacientes (26%) que se encontraron expuestos al sol desarrollaron cáncer de piel. Formularíamos la siguiente  $H_0$ : no existe diferencia en la incidencia de cáncer entre los pacientes expuestos y lo no expuestos al sol. El resultado que se presenta es el siguiente:  $\chi^2 = 17.3$  con una  $p < 0.001$ . Como el resultado es significativo ( $p < 0.05$ ), entonces interpretamos que la  $H_0$  es rechazada y que sí hay diferencia entre los grupos, y afirmamos que existe evidencia de un mayor riesgo de presentar cáncer de piel en los pacientes que se exponen al sol y que, además, esta diferencia encontrada no se debe al azar.

El test exacto de Fisher tiene la misma utilidad que el de  $\chi^2$ , pero se aplica en muestras pequeñas, en las que los valores esperados son  $\leq 5$ <sup>1,4</sup> (Tabla 2).

### Test t de Student

Una de las formas de presentar las variables cuantitativas es con la medias o promedio, que es la suma de todos los valores divididos por el número de observaciones, mientras que la mediana es el valor que se encuentra en la mitad cuando todas las observaciones

se clasifican de menor a mayor (o viceversa)<sup>2,4</sup>. Ejemplo: en un estudio quiere conocerse la diferencia que existe entre la disminución de la presión arterial con un solo medicamento (Tabla 3). Si obtenemos la media de la disminución de la presión arterial con un solo medicamento sería de 15.2 mmHg. En este ejemplo ya se encuentran acomodados de mayor a menor y podemos encontrar que la mediana es exactamente el mismo valor de 15 mmHg. Pero supongamos que agregamos dos pacientes más con valores atípicos, es decir, que no se comportan de la misma forma que los demás (varianza heterogénea), lo que puede alterar el resultado (Tabla 4). Entonces, al calcular la media es de 24 mmHg, y decimos que la media es sensible a los datos atípicos porque aumenta su valor considerablemente en comparación con el anterior. En cambio, si utilizamos la mediana continúa siendo de 15 mmHg, lo que es más adecuado porque la gran mayoría de los resultados se encuentran en este mismo rango.

El test t de Student es el procedimiento estadístico más habitual para comprobar si hay diferencia entre las medias de una variable cuantitativa en dos muestras independientes<sup>1,4</sup>. Cuando se aplica este test, damos por hecho que se cumplieron las siguientes reglas: 1) normalidad o una muestra mayor de 30 sujetos en cada grupo, y 2) homogeneidad de la varianza<sup>1,2</sup>. Ejemplo: queremos comparar dos técnicas quirúrgicas en una muestra de 600 pacientes. En el estudio hay dos grupos: pacientes tratados con ileostomía en asa y pacientes tratados con ileostomía más cierre de muñón. Se compara el tiempo quirúrgico y se encuentra que en el grupo 1 hay 14 sujetos con una media de 48.58 minutos y en el grupo 2 hay 11 pacientes con una media de 47.55 minutos. La idea es comprobar si la diferencia entre las medias presentadas se debe realmente a que es más eficiente una técnica que la otra o si se debe al azar. Una vez calculado, el resultado es una t de Student de 0.30 con una  $p = 0.76$ . Cuanto mayor es la «p», menos argumentos para rechazar la  $H_0$  y menor es el valor de la t de Student. Es decir, el efecto observado (la diferencia entre las dos técnicas) tiene una probabilidad de encontrarse del 76% cuando realmente no existe diferencia entre los grupos, y es una diferencia al azar.

En su contraparte está el test U de Mann-Whitney, para comparar dos grupos con datos no paramétricos. En este test los valores deben ordenarse, por lo que no se comparan realmente medias, sino las dos medianas (rangos)<sup>1,4</sup>. De igual forma, cuando se utiliza

**Tabla 3. Estudio de la disminución de la presión arterial. Datos ordenados de mayor a menor. Mediana = 15 mmHg. Media = 15.2 mmHg**

Pacientes	Disminución de la presión arterial
1	20 mmHg
2	16 mmHg
3	15 mmHg
4	13 mmHg
5	12 mmHg

**Tabla 4. Representación de outliers. Estudio de la disminución de la presión arterial. Datos ordenados de mayor a menor. Mediana = 15 mmHg. Media = 15.2 mmHg**

Pacientes	Disminución de la presión arterial
1	80 mmHg
2	20 mmHg
3	16 mmHg
4	15 mmHg
5	13 mmHg
6	12 mmHg
7	12 mmHg

**Tabla 5. Media de la frecuencia cardiaca, test de ANOVA**

Pacientes	Media	
No hacen ejercicio	0.89	$p = 0.01$
Deportistas	0.67	
Exatletas	0.70	

este test hay que dar por hecho que la muestra tiene menos de 30 sujetos, o que es mucho más pequeña ( $< 10$ )<sup>4</sup>.

Ahora bien, ¿qué sucede cuando lo que queremos comparar son más de dos grupos? Para este tipo de problemas existen otros procedimientos: el test de ANOVA, que es paramétrico, o en su defecto el test de Kruskal-Wallis, que es paramétrico<sup>1,4</sup>. Ejemplo: se estudia la frecuencia cardiaca de pacientes que no hacen ejercicio, de pacientes que son deportistas y de pacientes exatletas. Se obtienen las medias de los grupos (Tabla 5). Al calcular el test de ANOVA se obtiene una  $p = 0.01$  y se interpreta que las medias de los tres grupos son diferentes, por lo que se rechaza la  $H_0$ . Solo sabemos que los grupos presentan diferencias, pero es necesario saber cuáles son estas diferencias comparando los grupos de dos en dos

(parejas). En este ejemplo es muy evidente que hay una diferencia entre los grupos 2 y 3 en comparación con el grupo 1, pero no sabemos si hay diferencias entre los grupos 2 y 3. Posterior a realizar el estudio es necesario saber entre qué grupos hay diferencias, con un procedimiento de contraste<sup>1,4</sup>.

### Significancia clínica vs. significancia estadística

Todos en algún momento hemos caído en el error de realizar un estudio o leer un ensayo cuyo resultado es estadísticamente no significativo y concluimos que el estudio no sirve o tiene poca relevancia, y esto no necesariamente es cierto<sup>4</sup>.

Con frecuencia un valor de «p» se encuentra muy vinculado al tamaño de la muestra de un estudio. Ejemplo: en un estudio de mortalidad, utilizando como predictor la albúmina, se encuentran los datos mostrados en la tabla 6. Si somos realmente estrictos, la diferencia es de 0.1 mg entre los grupos. Al realizar la comparación resulta una  $p = 1.24$ , por lo que se considera estadísticamente no significativo. Sin embargo, si el mismo estudio se realiza en una muestra de 70,000 pacientes, el resultado da una  $p = 0.0003$ . Por lo tanto, aunque estadísticamente el resultado puede no ser relevante, clínicamente sí puede tener utilidad.

### Correlación y regresión

La finalidad de la correlación es explorar la dirección y la magnitud de la asociación entre dos variables cuantitativas. Ejemplo: para buscar la relación que existe entre la elevación de la amilasa y la elevación de la lipasa en un paciente con pancreatitis se calcula el coeficiente de correlación; en este caso, las variables no son dependientes de otra<sup>1,2</sup>.

### Análisis multivariado

Cuando analizamos un estudio en el que se pretende conocer, por ejemplo, la mortalidad de utilizar quimioterapia o quimiorradioterapia, existen algunas otras variables, como la raza, la edad, la comorbilidad, la presencia de metástasis, etc. Todas ellas pueden afectar el resultado. El investigador trata de demostrar la relación existente entre el tratamiento y todas estas variables agregadas (covariables) simultáneamente, y para ello se vale de tres análisis multivariados (el uso de uno u otro dependerá del tipo de variable)<sup>1,2</sup>:

**Tabla 6. Medias de los valores de la albúmina en el estudio de mortalidad**

Pacientes	Media de albúmina
Murieron	3.1 mg
No murieron	3.2 mg

- Variables continuas: regresión lineal múltiple.
- Variables categóricas o dicotómicas: regresión logística múltiple.
- Tiempo de eventos: regresión proporcional *Hazard*.

### Regresión lineal múltiple

En este procedimiento estadístico solo hay una variable cuantitativa, pero más de una variable dependiente<sup>1,2</sup>. Ejemplo: en un estudio se pretende valorar los cambios en la presión sistólica con un medicamento, que es una variable dependiente y cuantitativa, pero existen más variables independientes que afectan la disminución de la presión arterial, como la edad, el sexo, el índice de masa corporal y la comorbilidad. Al aplicar la regresión lineal múltiple podremos saber si el sexo afecta a la disminución de la presión arterial, así como las otras covariables, y la cantidad en que la afectan.

### Regresión logística múltiple

Este procedimiento se utiliza cuando hay una variable dependiente dicotómica y se desea identificar los predictores de la ocurrencia de un determinado fenómeno (que ocurra o que no ocurra)<sup>1,2,4</sup>. Ejemplo: los pacientes con diverticulitis se tratan con cirugía y buscamos conocer los factores asociados con la eficacia del tratamiento. Algunas de las variables serían el sexo, la edad y el tiempo desde el inicio del padecimiento. Se utiliza una regresión logística múltiple porque la variable dependiente (eficiencia) es categórica y dicotómica, y queremos saber si es eficiente o no, y si las covariables afectan la eficacia. Si midiéramos la eficacia por el tiempo de estancia hospitalaria, entonces tendríamos que utilizar una regresión lineal múltiple.

### Porciones de modelo Hazard

Este procedimiento se utiliza cuando se pretende conocer la incidencia que tiene cierto evento en un determinado tiempo. Ejemplo: queremos conocer la

incidencia de recidiva de la diverticulitis en los pacientes tratados con cirugía a 5 años. A diferencia de la regresión logística, en la cual también la variable dependiente es dicotómica (ocurrió o no), queremos saber el tiempo que tarda en producirse<sup>1,2,4</sup>.

### **Conceptos básicos sobre odds, odds ratio y riesgo relativo**

Cualquier persona que utilice procedimientos estadísticos estará de acuerdo en que es mucho más sencillo calcularlo que describirlo o explicarlo. Ejemplo: en 1000 pacientes con diverticulitis complicada se utilizó tratamiento conservador con dos antibióticos y se alcanzó el éxito en 800 pacientes. Si se divide este número entre el de los pacientes en que el tratamiento no fue satisfactorio, equivaldría a 800/200, y entonces obtendríamos una *odds* de 4, lo que significa que por cada paciente en que no se logró el éxito con antibióticos hay cuatro en que sí se consiguió.

#### **Odds ratio**

Este concepto se refiere solo a la división entre una *odds* y otra *odds*. En algunas traducciones puede encontrarse como razón de momios o razón de posibilidades, entre otras<sup>1,2,4</sup>. Ejemplo: 80 pacientes con diverticulitis complicada son tratados por laparoscopia y se alcanza el éxito en 72, con lo que la *odds* sería de 9. Al dividirla entre la *odds* de nuestro estudio previo (9/4) tendríamos una *odds ratio* (OR) de 2. Interpretamos que el éxito terapéutico del tratamiento quirúrgico es mayor que el del tratamiento conservador. Si obtuviéramos una OR = 1, los dos tratamientos serían iguales. Si el resultado fuera < 1, se interpretaría que el desenlace es menos frecuente en la

categoría o grupo de interés. Ejemplo: si queremos comparar el éxito terapéutico del tratamiento conservador sobre el quirúrgico, entonces la división será  $4/9 = 0.44$  y se interpretará que el tratamiento médico es menos efectivo que el tratamiento quirúrgico.

#### **Riesgo relativo**

El riesgo relativo es muy similar a la OR, pero se utiliza para casos raros que ocurren en menos del 10% de los sujetos<sup>1,2,4</sup>.

#### **Conflicto de intereses**

Los autores declaran que no existen conflictos de intereses.

#### **Responsabilidades éticas**

**Protección de personas y animales.** Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

**Confidencialidad de los datos.** Los autores declaran que en este artículo no aparecen datos de pacientes.

**Derecho a la privacidad y consentimiento informado.** Los autores declaran que en este artículo no aparecen datos de pacientes.

#### **Bibliografía**

1. Martínez-González MA. Bioestadística amigable. 2.ª ed. Madrid: Díaz de Santos; 2006. p. 919.
2. Guller U, DeLong ER. Interpreting statistics in medical literature: a vademecum for surgeons. J Am Coll Surg. 2004;198:441-58.
3. Altman DG, Godman. The transfer of technology from statistical journals to the biomedical literature. JAM. 1994;272:19-32.
4. Clifford Blair R, Taylor Richard A. Bioestadística. México: Pearson Educación; 2008. p. 552
5. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Epidemiología clínica. Ciencia básica para la medicina clínica. 2.ªed. Buenos Aires: Panamericana; 1994.