



ORIGINAL

Utilización de la estadística en cirugía ortopédica y traumatología

Statistic in Orthopaedic Surgery and Traumatology

Fernández-Carreira J M ¹, Hernández-Vaquero D ², Sánchez-Torres M ³

¹ Unidad de Apoyo a La Investigación. Hospital San Agustín, Avilés. ² Departamento de Cirugía. Facultad de Medicina, Oviedo.

³ Licenciada en Psicología.

Resumen

Los estudios de investigación epidemiológica se sustentan en los métodos de la estadística. Para poder alcanzar conocimiento sobre poblaciones frecuentemente se estudian muestras representativas de las mismas. Estos estudios se basan en la inferencia estadística. La inferencia se puede realizar mediante la estimación de parámetros por intervalos de confianza o bien mediante el cálculo de la significación estadística –valores p– mediante contrastes de hipótesis. En un intervalo de confianza se busca una solución que proporcione seguridad y precisión. Con los contrastes de hipótesis se cuantifica la probabilidad de que las observaciones muestrales puedan ser explicadas por el azar del muestreo, midiendo los errores tipo I y tipo II. En cada hipótesis de investigación existe un contraste apropiado en función de los tipos de variables estudiadas, su distribución, el número de sujetos estudiados y el número de grupos que se comparan. La utilización de valores p está sujeta a controversia, ante lo que se ofrecen alternativas como la utilización de métodos bayesianos o el enfoque a intervalos de confianza. Los estudios dirigidos a la medición de eventos a lo largo de periodos de observación se denominan estudios de supervivencia. Existen distintos métodos en función del tipo de periodos estudiados y de las variables a analizar. Los métodos estadísticos multivariantes permiten analizar dos o más variables a la vez. Existen métodos exploratorios que pueden simplificar el número de variables o de sujetos y métodos confirmatorios como la regresión logística que permite analizar el papel de variables confusoras y de interacciones.

Palabras clave:

Cirugía ortopédica, estadística, inferencia, significación estadística.

Abstract

Objetivo: Epidemiologic research is based on statistical methods and usually studies representative samples to attain knowledge about population. These studies are based on statistical inference. There are two ways to make inference. One is the parameter estimation with confidence intervals, another is the hypothesis test. A confidence interval allow us to control the precision (a range of the estimated variable) and the confidence (the probability of the population parameter to be included in that range) of the results. A hypothesis test quantifies the chance that a result could be explained by the random nature of the sample measuring type I and type II errors. Every research question has a matching hypothesis test according to the types of the variables, their distribution, the number of sampled units and the groups to be compared. The use of p-values is subject to controversy. Other methods like Bayesian statistics or confidence intervals offer alternatives. Survival studies measure events along time intervals. There are different methods regarding the type of intervals or the data to be analyzed. Multivariate statistics allow us to analyze two or more variables at a time. The exploratory methods can simplify the data, both the number of subjects and the variables. Confirmatory methods like logistic regression enable the analysis and importance of confounding and interacting variables.

Key words:

Orthopaedic surgery, statistical inference, statistical significance.

Correspondencia

D. Hernández Vaquero
Apartado de correos 341. 33400 Avilés (Asturias)
dhernandezv@meditex.es



I Introducción

La actividad médica en cualquiera de sus aspectos, asistencial o investigador, genera de forma continuada un enorme volumen de datos. De la observación de estos datos depende en gran medida el desarrollo de la práctica médica. Este hecho no es exclusivo de las ciencias de la salud, cualquier faceta del conocimiento científico se fundamenta en el método hipotético deductivo.

El desarrollo de hipótesis de investigación y su resolución depende en gran medida de la aplicación de técnicas de análisis numérico, métodos que nos facilitan la síntesis de información a partir del análisis de datos y finalmente la obtención de conocimiento. El Diccionario de la Real Academia de la Lengua Española define la Estadística como: «1. Estudio de los datos cuantitativos de la población, de los recursos naturales e industriales, del tráfico o de cualquier otra manifestación de las sociedades humanas. 2. Conjunto de estos datos. 3. Rama de la matemática que utiliza grandes conjuntos de datos numéricos para obtener inferencias basadas en el cálculo de probabilidades».

Las dos primeras definiciones corresponden a la estadística descriptiva, el tercero a la estadística inferencial. El esquema descripción-inferencia se adapta perfectamente a la utilización real de los métodos estadísticos en la investigación biomédica siguiendo un esquema lógico o incluso cronológico donde, en primer lugar se estudian de forma descriptiva las variables de una muestra, para en un paso posterior establecer una proyección inductiva o inferencia sobre la población total de donde parte la muestra. En este modelo, las variables estudiadas en la muestra nos proporcionan datos estadísticos descriptivos que, por el proceso inferencial, nos permiten aproximar el valor de los valores correspondientes en la población o parámetros.

I Inferencia estadística

Cuando se pone en marcha un estudio de investigación debemos definir su ámbito en términos de las variables clásicas de la epidemiología: persona, lugar y tiempo, lo que conducirá a los distintos tipos de diseños epidemiológicos observacionales, descriptivos y analíticos o experimentales. En la definición del ámbito del estudio se concretará la población «diana» o de referencia, que será el objetivo de las conclusiones, tanto si se estudia a todos los individuos que componen la población como si se limita a un subconjunto representativo de la misma, o muestra.

La posibilidad de conocer una población estudiando una muestra de la misma posibilita la realización de estudios que no serían factibles a nivel poblacional, aunque los resultados de los mismos irán acompañados siempre de cierto gra-

do de incertidumbre, en función de la variabilidad de los fenómenos estudiados y del tamaño de las muestras seleccionadas. Esta es la razón de la existencia de los métodos de inferencia estadística: inducir o inferir los valores descriptivos de variables de la población o parámetros a partir de la medición de las mismas variables en la muestra (estadísticos).

La validez de la inferencia estará condicionada por la representatividad de la muestra. Los factores determinantes de esta representatividad son el tamaño muestral que ha de ser predefinido para las condiciones fijadas por el investigador minimizando el error aleatorio y por tanto aumentando la precisión y también la utilización de métodos de selección probabilísticos que minimicen el error sistemático (sesgos de selección) aumentando la exactitud.

Precisión y exactitud son los dos aspectos que definen la capacidad de un estudio estadístico para llegar a conclusiones válidas sobre una población «diana» [1-4]. Los estudios de inferencia pueden enfocarse desde dos puntos de vista. En ambos se comparte el hecho de que se produce un razonamiento inductivo; de la parte al todo, de la muestra a la población. Por una parte podemos realizar un estudio de estimación de parámetros donde, a partir de la medición de variables en una muestra, establecemos un intervalo de confianza para el valor del parámetro. Un intervalo de confianza se construye a partir de un rango de la variable y una probabilidad de que el parámetro se encuentre dentro de dicho intervalo. La construcción de intervalos de confianza permite delimitar la incertidumbre inherente a la inferencia y por tanto establecer un marco de referencia para el valor del parámetro estudiado. La geometría asintótica de las distribuciones de probabilidad subyacentes en el proceso de inferencia nos obliga a tomar decisiones sin tener una seguridad absoluta, de ahí que se utilicen rangos de probabilidad del 95% o del 99% y nunca del 100% ya que este último requeriría intervalos infinitos [5].

El otro tipo de inferencia responde a la formulación de hipótesis sobre los datos. Después de diseñar un procedimiento experimental y recoger los datos muestrales, los resultados de dicha muestra se someten a un procedimiento matemático denominado contraste cuyo resultado será un valor de probabilidad o significación estadística, que en último término servirá para tomar una decisión sobre la hipótesis previamente planteada.

I Hipótesis alternativa e hipótesis nula

El planteamiento clásico de la estadística inferencial propone la creación de una hipótesis principal, objeto de investigación que denominaremos hipótesis alternativa o H1 y de otra hipótesis opuesta que denominaremos hipótesis nu-



la o H0. La H0 corresponderá al conocimiento que exista en ese momento y se formulará con la intención de poder rechazarla una vez contrastadas las pruebas obtenidas.

Supongamos que hemos llegado al final de una investigación y hemos encontrado alguna asociación, lo que habitualmente llamaríamos un resultado «positivo». En este caso caben plantearse dos posibilidades: que realmente existe la asociación (habríamos acertado) o que estamos cometiendo un error de tipo I (α). Por esto en los estudios positivos (los que más se publican) siempre debemos fijarnos en el error alfa teórico aceptado *a priori* (habitualmente 5%, y en la p , que es equivalente a alfa pero en los datos estudiados). Al finalizar una investigación también podría ocurrir que no encontremos asociación (estudio negativo), en cuyo caso también caben dos posibilidades: que realmente no exista asociación (acertaríamos) o que estemos cometiendo un error de tipo II o beta (β), o lo que es lo mismo, que el estudio tenga poca potencia. Si el estudio tiene una potencia mayor del 80% cabe pensar que tiene la suficiente potencia para ser tenido en cuenta, aun siendo negativo, por lo que debería pasar a formar parte del conocimiento global y sobre todo, formar parte de la evidencia incluida en los metaanálisis.

Debemos tener en cuenta que cuando rechazamos una hipótesis alternativa lo hacemos siempre por falta de pruebas. No se debe rechazar con el argumento de que la hipótesis alternativa es falsa, puesto que nuestro experimento no se ha diseñado para soportar esa conclusión. No es lo mismo la ausencia de evidencia que la evidencia de ausencia.

El planteamiento clásico de errores de tipo I y tipo II desde el punto de vista del error aleatorio puede extenderse con los errores sistemáticos, que pueden producir el mismo tipo de alteraciones en la toma de decisiones sobre las hipótesis. Según este modelo podríamos hablar de sesgos de tipo I como aquellos que inducen a aceptar hipótesis alternativas cuando deberían ser rechazadas y errores sistemáticos de tipo II cuando un sesgo de selección, clasificación o confusión oculta la existencia de alguna asociación inclinando la decisión a favor de rechazar la hipótesis alternativa (Tabla 1).

Significación estadística

La significación estadística o valor p es la probabilidad de encontrar diferencias iguales o mayores a las que encontramos en nuestra investigación si la hipótesis nula fuese cierta en el ámbito de la población de referencia. Si el valor p resulta suficientemente bajo, este resultado nos hará dudar de la verosimilitud de la hipótesis nula (H0) o lo que es lo mismo, nos inclina a aceptar la hipótesis alternativa [6,7].

Un ejemplo, comparamos la incidencia de complicaciones de dos técnicas quirúrgicas. Se observa un 6% de complicaciones con una técnica y un 4% con el otro, siendo esta diferencia estadísticamente significativa ($p=0,03$). La interpretación correcta de este resultado es que si ambas técnicas tuvieran la misma incidencia de complicaciones, la probabilidad de encontrar una diferencia igual o mayor a la observada es 0,03. En este escenario dudaríamos de que H0 (ambas técnicas son iguales respecto a las complicaciones) sea cierta. Aceptaríamos la H1: ambas técnicas son diferentes.

Frecuentemente se malinterpreta el valor p resultante de un contraste de hipótesis. Un valor muy significativo de p no significa que existan grandes diferencias entre las muestras, sino que las diferencias encontradas en el estudio muestral existen a nivel poblacional con una gran seguridad. Los valores p son muy dependientes del tamaño muestral, de forma que incluso pequeñas diferencias pueden producir valores p muy significativos solo con aumentar el tamaño de la muestra. Este hecho es una de las razones que apoyan las duras críticas a la utilización de valores p como único criterio para interpretar los estudios de investigación biomédicos.

Rothman et al [8], e incluso el grupo de editores del comité de Vancouver, sugieren que los resultados de los estudios epidemiológicos no deberían depender únicamente de los resultados de las pruebas de significación estadística. Estos autores sugieren la utilización en su lugar de la estimación de parámetros por intervalos de confianza. La mayor parte de las hipótesis de investigación formuladas para

Tabla 1. Errores aleatorios y sistemáticos

| | Existe asociación (H1 verdadera) | | No existe asociación (H1 falsa) | |
|---|-------------------------------------|------------------------|---------------------------------------|------------------------|
| Se encuentra asociación (Se acepta H1) | Correcto | | Error | |
| | Aleatorio Potencia (1- β) | Sistemático Validez | Aleatorio Tipo I (α) | Sistemático Sesgo |
| No se encuentra asociación (Se rechaza H1) | Error | | Correcto | |
| | Aleatorio Tipo II (β) | Sistemático Sesgo | Aleatorio Seguridad (1- α) | Sistemático Validez |



Tabla 2. Pruebas de hipótesis en función de los tipos de variables, los grupos de comparación y sus distribuciones

| V. explicadora → | Cualitativa | | | |
|---------------------------------|--------------------------|---------------------------------|-----------------|-------------------------|
| | 2 categorías | | ≥2 categorías | |
| Variable explicada ↓ | Independientes | Pareadas | Independientes | Pareadas |
| Cuantitativa | t Student Anova, GLM* | t Student pareada Anova, GLM | Anova, GLM | Anova medic rep, GLM |
| Ordinal o cuantitativa con n<30 | Mann Whitney | Wilcoxon | Kruskall Wallis | Friedman |
| Cualitativas | Chi cuadrado, Fisher | Mac Nemar | Chi cuadrado | Q Cochran |

* GLM: Modelo general lineal.

ser respondidas con una toma de decisión dicotómica pueden ser planteadas para responder mediante intervalos de confianza de una forma más flexible [9].

■ Métodos bayesianos

Como alternativa a los métodos frecuentistas clásicos se presentan los métodos bayesianos, que frente al planteamiento clásico de H1-H0 con las mismas probabilidades de partida, plantea un razonamiento en el que se parte de distintas probabilidades para ambas hipótesis, y que mediante la investigación iremos aportando más o menos evidencia de credibilidad a una o a otra de las hipótesis. Este tipo de razonamiento es utilizado cuando se interpretan los resultados de una prueba diagnóstica donde el conocimiento previo se cuantifica como la probabilidad pre-prueba o prevalencia de la enfermedad. Los resultados de la prueba diagnóstica con mayor o menor capacidad discriminante medida en términos de sensibilidad y especificidad aportarán mayor o menor grado de verosimilitud a la probabilidad final de que realmente el individuo esté enfermo en caso de resultar positivo para la prueba; esta sería la probabilidad post-prueba o valor predictivo positivo. La interpretación de los resultados de las pruebas diagnósticas como vemos, está siempre modulada por la prevalencia y por supuesto por las características operacionales de la prueba diagnóstica y el modelo probabilístico que rige estas relaciones es precisamente el incluido en el teorema de Bayes. Lo que se propugna desde la escuela Bayesiana es utilizar este esquema de razonamiento en lugar del modelo frecuentista clásico donde no se tiene en cuenta la probabilidad preprueba [10][11].

■ Contraste de hipótesis

Existen múltiples procedimientos matemáticos para realizar contraste de hipótesis, en función de los tipos de variables estudiadas, el número de categorías o grupos que se

comparan y sus relaciones de dependencia y las distribuciones de probabilidad de los parámetros. En general podemos hablar de procedimientos paramétricos y no paramétricos en función del conocimiento de la distribución de las variables en la población de referencia. En este sentido, los métodos paramétricos asumen que las variables poblacionales siguen una distribución normal, mientras que los métodos no paramétricos no hacen ninguna asunción sobre los parámetros, por lo que podemos utilizarlos en situaciones que no reúnen las condiciones de aplicación de las pruebas paramétricas. A priori, podría parecer que nunca podemos conocer si la distribución muestral de medias de nuestras variables es normal; sin embargo sabemos que cuando una muestra tiene el tamaño suficiente ($n > 30$), independientemente del tipo de distribución de la población, la distribución muestral de medias se ajustará a la normal y por tanto serán aplicables los métodos paramétricos. En este sentido, si la muestra es pequeña, aunque las variables a estudio sean de tipo cuantitativo, puesto que no se puede determinar la distribución de los parámetros ni asumir la normalidad de la distribución muestral de medias por el teorema central de límite, debemos recurrir a las correspondientes pruebas no paramétricas, por ejemplo la prueba de Mann Withney en lugar de la prueba t de Student para muestras independientes o la prueba de Kruskal Wallis en lugar de la prueba de análisis de la varianza (ANOVA).

En general podemos enfocar la utilización de contrastes de hipótesis desde una perspectiva bivalente con una variable a estudio o variable explicada y otra variable de agrupamiento o explicativa de tipo cualitativo con dos o más categorías relacionadas o no [12-15]. En la tabla 2 se muestran algunos de los contrastes de hipótesis de uso más frecuente siguiendo el esquema de variable explicada y variable explicadora. En general estas pruebas de hipótesis están disponibles en la mayoría de paquetes estadísticos tanto comerciales como de dominio público.

I Estudios de supervivencia

Entre los métodos de mayor aplicación en los estudios en cirugía ortopédica y traumatológica encontramos los estudios de supervivencia donde se analiza la ocurrencia de determinados eventos a lo largo del tiempo [16-18]. En estos estudios hablamos de «eventos» para los fenómenos a estudiar, por ejemplo la aparición de una complicación, y de casos censurados para los individuos que no han producido eventos, bien porque ha finalizado el estudio sin haber ocurrido o porque esos sujetos abandonaron el seguimiento antes de tiempo; en este segundo caso hay que tener especial cuidado con la presencia de sesgos de selección en el caso de que el abandono ocurra por algo relacionado con el fenómeno estudiado.

En un primer enfoque tendríamos las tablas de vida donde se estudia en periodos fijos de tiempo el número de eventos ocurridos. Los gráficos correspondientes a este tipo de análisis se caracterizan por tener «escalones» de la misma longitud pero de diferente profundidad, ya que en cada periodo puede ocurrir un número distinto de eventos. Otra forma de estudiar la supervivencia es el método de Kaplan Meier donde cada vez que ocurre un evento se tiene en cuenta el tiempo total que el individuo ha aportado al estudio antes de producirse dicho evento. Por esto podemos observar en las gráficas de supervivencia de este tipo escalones de distinta longitud pero siempre de la misma profundidad, a no ser que se solapen varios eventos en el mismo instante. En el método de Kaplan Meier es posible comparar la evolución de diferentes grupos mediante el método de Log Rank (Figura 1). Por último el modelo de regresión de Cox

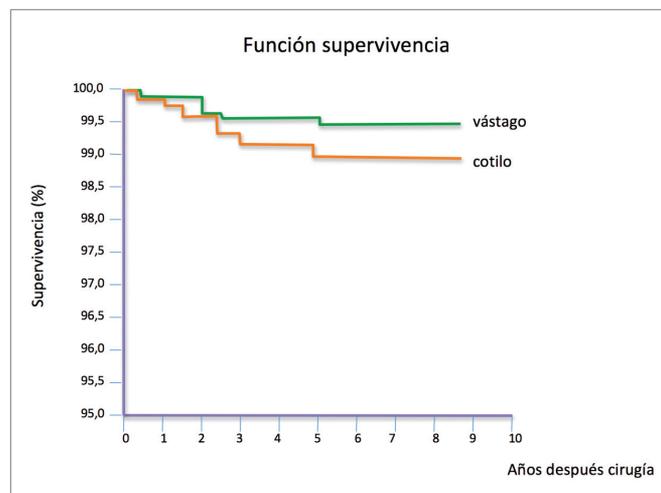


Fig. 1. Gráfico de supervivencia. Método de Kaplan Meier. Se muestra la supervivencia de los componentes en una artroplastia total de cadera con un seguimiento de 9 años.

Tabla 3. Clasificación de los métodos de análisis multivariante

Técnicas descriptivas o exploratorias

- Técnicas representativas
 - Análisis de componentes principales
 - Análisis de correspondencias
 - Análisis de proximidades
- Técnicas clasificatorias
 - Análisis cluster: clasificación de individuos y clasificación de variables

Técnicas confirmatorias

- Análisis discriminante
- Regresión múltiple
- Modelos Log-Linear
- Análisis múltiple de la varianza
- Modelo estructural lineal

se utiliza cuando existen covariables relacionadas con la aparición del evento.

I Métodos de análisis multivariante

Finalmente, debemos considerar la utilidad de los métodos de análisis multivariante [19][20]. Podemos establecer una clasificación funcional (Tabla 3) de estos métodos entre las técnicas descriptivas o exploratorias generalmente enfocadas a simplificar el escenario del experimento, bien reduciendo el número de variables o agrupando a los individuos y las técnicas confirmatorias dirigidas a verificar hipótesis en forma similar a los métodos clásicos bivariantes. Destaca su utilidad en los estudios epidemiológicos en los que la confusión es difícilmente controlable cuando no se pueden realizar procedimientos de aleatorización, métodos como la regresión logística que nos ayudan a clarificar el papel de ciertas variables como interacciones o confusoras.

Evidentemente en estas líneas no podemos resumir una ciencia tan amplia como la estadística y su aplicación a la medicina, específicamente a la cirugía ortopédica. Los investigadores, incluso los clínicos que deseen publicar en revistas de impacto, necesitaran la ayuda de un bioestadístico desde el mismo momento del planteamiento del trabajo. ■

REFERENCIAS BIBLIOGRÁFICAS

1. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing Clinical Research, 3rd Edition. Philadelphia: Lippincott Williams & Wilkins; 2007.



2. Van Belle G, Fisher LD, Heagerty PJ, Lumley T. Biostatistics: a methodology for the health sciences. 2 Edition. Malden: Wiley-Interscience; 2004.
3. Randy M. Pag, Basic Epidemiological Methods and Biostatistics: A Practical Guidebook. Sudbury: Jones & Bartlett Publishers; 1995.
4. Glantz S. Primer of Biostatistics 5th Edition, New York: McGraw-Hill; 2001.
5. Polansky AM. Observed Confidence Levels: Theory and Application. Boca Raton: Chapman & Hall / CRC; 2007.
6. Moyé LA, Statistical Reasoning in Medicine: The Intuitive P-Value Primer. 2nd edition, New York: Springer; 2006.
7. Silva Aycaguer LC. Cultura estadística e investigación científica en el campo de la salud: una mirada crítica. Madrid: Díaz de Santos; 1997.
8. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. Philadelphia: Lippincott Williams & Wilkins; 2008
9. Chernick MC, Friis RH. Introductory Biostatistics for the Health Sciences: Modern Applications Including Bootstrap. Malden: Wiley-Interscience; 2003.
10. Congdon P. Bayesian Statistical Modelling (Wiley Series in Probability and Statistics) Malden : Wiley; 2007.
11. Bolstad WM. Introduction to Bayesian Statistics Malden: Wiley-Interscience; 2007.
12. Aerts M. The Biostatistics Cookbook: The Most User-Friendly Guide for the Bio/Medical Scientist. New York: Springer; 1996.
13. Armitage P, Berry G, Matthews JNS. Statistical Methods in Medical Research. Malden: Wiley-Blackwell; 2001.
14. Bruce N, Pope D.. Quantitative Methods for Health Research: A Practical Interactive Guide to Epidemiology and Statistics. Malden: Wiley-Interscience; 2008.
15. Tutorials in Biostatistics. Volume 1 - Statistical Methods in Clinical Studies» Malden: Wiley; 2004.
16. Lee ET, Wenyu Wang J. Statistical Methods for Survival Data Analysis, 3rd edition. Malden: John Wiley & Sons; 2003.
17. Collet D. Modelling Survival Data in Medical Research, Second Edition. Texts in Statistical Science. Boca Raton: Chapman & Hall/CRC; 2003.
18. Survival Analysis for Epidemiologic and Medical Research. Practical Guides to Biostatistics and Epidemiology. New York: Cambridge, University Press; 2008.
19. Meyers L, Gamst G, Guarino AJ. Applied Multivariate Research. Design and Interpretation. Newbury Park: Sage Publications; 2005.
20. Harris RJ. A Primer of Multivariate Statistics, 3rd Ed. Mahwah: Lawrence Erlbaum Associates; 2001.

Conflicto de intereses

Los autores no hemos recibido ayuda económica alguna para la realización de este trabajo. Tampoco hemos firmado ningún acuerdo por el que vayamos a recibir beneficios u honorarios por parte de alguna entidad comercial. Ninguna entidad comercial ha pagado, ni pagará, a fundaciones, instituciones educativas u otras organizaciones sin ánimo de lucro a las que estamos afiliados.