

Comunicación científica (X). Conocimientos básicos para elaborar un artículo científico (5): Los aspectos estadísticos (más que números)

J. González de Dios¹, M. González-Muñoz², A. Alonso-Arroyo³, R. Aleixandre-Benavent⁴

¹Servicio de Pediatría. Hospital General Universitario de Alicante. Departamento de Pediatría. Universidad «Miguel Hernández». Alicante. ²Facultad de Medicina y Odontología. Universidad de Valencia. ³Departamento de Historia de la Ciencia y Documentación. Facultad de Medicina y Odontología. Universidad de Valencia. ⁴IHMC López Piñero (CSIC-Universidad de Valencia). Unidad de Información e Investigación Social y Sanitaria-UISYS. Valencia

Resumen

La estadística es un apartado esencial, tanto para los autores como para los lectores de literatura científica, que fundamenta el rigor científico del estudio y la validez de las conclusiones. Conocer el tema de un artículo científico ayuda a entender el trabajo, y conocer los principios de estadística ayuda a entender la metodología.

Las principales premisas de la estadística que se deben tener en cuenta en la elaboración de un artículo científico son las siguientes: 1) describir los métodos científicos con suficiente detalle; 2) cuantificar los resultados; 3) conocer el significado del valor de p ; 4) no depender exclusivamente de la p ; 5) analizar los criterios de inclusión y exclusión; 6) proporcionar los detalles del proceso de aleatorización; 7) facilitar los detalles del proceso de enmascaramiento; 8) informar sobre las complicaciones del tratamiento; 9) especificar el número de observaciones e indicar las pérdidas; 10) detallar el programa estadístico y las referencias sobre estadística; 11) definir los términos y evitar el uso no técnico de términos de la estadística, y 12) controlar los errores de la investigación clínica.

La estadística no debe convertirse en una religión que conduce a la salvación a través del ritual de la búsqueda de valores significativos de la p , alcanzando entonces los autores el paraíso prometido de la publicación del trabajo en una revista de alto impacto, sino que es una herramienta útil para el diseño y la realización de estudios médicos, siempre que se utilice adecuadamente.

©2014 Ediciones Mayo, S.A. Todos los derechos reservados.

Palabras clave

Comunicación científica, estadística, formación, información, medicina, pediatría, publicación científica

Introducción

La estadística es la parte más ardua de un artículo científico y generalmente entraña dificultades tanto a los autores como a los lectores. Sin embargo, es un apartado tan esencial que fun-

Abstract

Title: Scientific communication (X). Basic knowledge to develop a scientific paper (5): Statistical aspects (rather than numbers)

Statistics is an essential section for both, authors and readers of scientific literature, that offers the basis for the scientific rigor of the study and the validity of the conclusions. Knowing the subject of a scientific article helps to understand the work and knowing the principles of statistics helps to understand the methodology.

The main assumptions of the statistic that should be taken into account in the preparation of a scientific paper are: 1) describe the scientific methods with enough detail; 2) quantify the results; 3) know the meaning of the p value; 4) does not depend exclusively of p ; 5) analyze the inclusion and exclusion criteria; 6) provide the details of the randomization process; 7) provide the details of the masking process; 8) report on the complications of the treatment; 9) specify the number of observations and indicate losses; 10) specify the statistical software and statistical references; 11) define the terms and avoid using non-technical statistical terms in statistics, and 12) handle errors in clinical research.

Statistics should not become a religion that leads to salvation through the ritual of finding significant values of p , and then the authors reaching the promised land of publication of the paper in a high-impact journal, but properly used is a useful tool for the design and conduct of medical research.

©2014 Ediciones Mayo, S.A. All rights reserved.

Keywords

Scientific communication, statistics, training, information, medicine, pediatrics scientific publication

damenta el rigor científico del estudio y, por ende, la validez de las conclusiones.

El uso de la estadística se ha incrementado y cuidado mucho en las principales revistas científicas en el último medio siglo.

Pero una consecuencia desgraciada de este fenómeno ha sido que se ha puesto demasiado énfasis en examinar las pruebas de hipótesis con un enfoque excesivamente «estadístico» y menos «clínico»: los datos se examinan en relación con una hipótesis nula, práctica que ha llevado a la creencia errónea de que el objetivo de los estudios debe ser obtener una diferencia «estadísticamente significativa», cuando de hecho el objetivo real debe ser determinar la magnitud de algún factor de interés y la diferencia «clínicamente importante».

En realidad, la capacitación formal en bioestadística de los investigadores médicos no siempre ha sido la deseable: cuando estudiamos esta materia al inicio de nuestra formación pregrado en medicina no somos capaces de entender su trascendencia, y es la formación posgrado posterior la que lima algunas de estas debilidades. Por ello, el hecho de publicar un artículo en una revista biomédica, previa revisión por correctores, no implica que éste no presente errores en la metodología estadística. Salvo honrosas excepciones (hay revistas de prestigio que disponen de un consejo de revisores centrados en la metodología y la estadística), es una parte del trabajo que tenemos que realizar (como autores) y entender (como lectores) con sumo rigor y precaución, pues las conclusiones erróneas de algunos estudios clínicos mal realizados pueden exponer a los pacientes al riesgo y los gastos asociados a tratamientos inútiles, así como a una demora innecesaria del tratamiento apropiado.

Conviene tener presente una serie de premisas en la información estadística que deben proporcionar los artículos científicos, de forma que sea comprendida por el lector promedio que conoce la disciplina general, pero que no es especialista en el tema específico de la investigación^{1,2}. Conocer el tema de un artículo científico ayuda a entender el trabajo, y conocer la estadística ayuda a entender la metodología.

Por ello, conviene recordar algunas premisas relacionadas con los aspectos estadísticos que deben tenerse en cuenta en la elaboración de un artículo científico^{3,4} y que exponemos a continuación.

Premisas para sobrevivir a la estadística

1) Describir los métodos científicos con suficiente detalle

Es básico realizar una buena descripción de los métodos científicos utilizados en los estudios para que un lector que tenga acceso a los datos originales pueda verificar los resultados notificados.

Los estudios estadísticos se clasifican en diversas categorías en función de los criterios de Emerson y Colditz (tabla 1), publicados en *The New England Journal of Medicine* hace 30 años⁵ y que siguen siendo válidos, tal como se confirma en varios estudios realizados en España⁶⁻⁸. Estos mismos autores introdujeron un término útil, la «accesibilidad estadística», en-

tendido como un índice de complejidad de las técnicas estadísticas empleadas en los trabajos científicos. Así, se considera que el umbral básico para un lector estándar (el que haya superado un curso básico de bioestadística durante su currículo biomédico) se situaría en la estadística bivariable (hasta la regresión simple).

Si la prueba estadística es de uso común (p. ej., t de Student, χ^2 , Wilcoxon, Mann-Whitney, etc.), sólo cabe citarla, pero las pruebas estadísticas de mayor complejidad (p. ej., análisis de la variancia, estudio multivariante, etc.) sí deben especificarse de forma clara. Conviene reseñar si las observaciones son no pareadas (p. ej., comparación de 2 medias entre 2 grupos de pacientes que reciben diferentes tratamientos) o pareadas (p. ej., comparaciones de 2 medias en el mismo grupo de pacientes en 2 circunstancias diferentes, antes y después del tratamiento), porque variará el tipo de prueba estadística. Así, la prueba de Mann-Whitney utilizada para datos no pareados se sustituye por la prueba de Wilcoxon para datos pareados; la prueba de la χ^2 para datos no pareados se sustituye por la prueba de McNemar para datos pareados.

La descripción general de los métodos estadísticos debe aparecer de forma clara en el apartado de «Material y métodos», habitualmente en la parte final. Y cuando se resuman los datos en la sección de «Resultados», se deben especificar los principales métodos estadísticos que se emplearon para analizarlos.

2) Cuantificar los resultados

Es clave presentar los resultados con los indicadores apropiados y siempre con las oportunas medidas de centralización y de dispersión (porque toda cifra debe ir acompañada siempre de los indicadores adecuados de error o incertidumbre de la medición).

Toda relación de cifras obtenida de un experimento se puede sustituir por tres valores: 1) el número de componentes o tamaño de la muestra; 2) el valor medio como medida de centralización, que informa sobre alrededor de qué valor se agrupan los datos, y 3) la desviación estándar (DE), como medida de dispersión que da idea de si este agrupamiento de datos es muy disperso o muy concentrado.

Porque nunca se debe exponer una media sin la medida de variabilidad o precisión de la media: la DE se usa para conocer la variabilidad entre individuos, y el error estándar de la media (EE) para conocer la variabilidad de la muestra. Así pues, la DE y el EE son dos aspectos distintos que no cuantifican lo mismo⁹. Cuando la variable que se observa se comporta de tal manera que las observaciones se distribuyen igualmente por encima o por debajo de la media (distribuciones simétricas), tiene sentido cuantificar la dispersión de los datos mediante la DE. Para distribuciones asimétricas deberemos pensar en otro tipo de medidas.

El EE de la media no cuantifica la variabilidad de las observaciones, como hace la DE, sino la precisión con la que una

TABLA 1

Categoría de análisis estadístico

Categoría	Descripción
0. Ningún estudio estadístico	
1. Sólo estadística descriptiva	Ej.: porcentajes, medias, desviación estándar, error estándar, etc.
2. Prueba de la t de Student y pruebas z	Para 1 o 2 muestras (datos apareados y/o independientes)
3. Tablas bivariadas	Ej.: prueba de la χ^2 , prueba exacta de Fisher, prueba de McNemar
4. Pruebas no paramétricas	Ej.: prueba de la U de Mann-Whitney, prueba de la T de Wilcoxon
5. Estadísticos demoepidemiológicos	Ej.: riesgo relativo, <i>odds ratio</i> , sensibilidad, especificidad
6. Correlación lineal de Pearson	Correlación clásica producto-momento (r)
7. Regresión simple	Regresión de mínimos cuadrados con una variable predictora y una variable respuesta
8. Análisis de la variancia	Análisis de la variancia y de la covariancia: pruebas F
9. Transformación de variables	Uso de transformaciones (p. ej., logarítmicas)
10. Correlación no paramétrica	Ej.: pruebas rho de Spearman, tau de Kendall, pruebas de tendencia
11. Regresión múltiple	Ej.: regresión polinómica y regresión paso a paso
12. Comparaciones múltiples	Ej.: pruebas de Bonferroni, Scheffé, Dunca, Newman-Keuls
13. Ajustes y estandarización	Estandarización de tasas de incidencia y prevalencia
14. Tablas multivariadas	Ej.: procedimiento de Mantel-Haenszel, modelos log-lineales
15. Potencia y tamaño muestral	Determinación del tamaño muestral en función de una diferencia detectable (o útil)
16. Análisis de la supervivencia	Ej.: tablas de vida (Kaplan-Meier), regresión de supervivencia (logística, de Cox) y otros (Breslow, Kruskal-Wallis, <i>long-rank test</i>)
17. Análisis económicos	Ej.: coste-efectividad, coste-beneficio, coste-utilidad
18. Otros análisis diversos	Ej.: análisis de sensibilidad, análisis discriminante, análisis <i>cluster</i> u otros modelos matemáticos

Modificada de Jiménez⁴.

media muestral estima la verdadera media poblacional. El EE cuantifica la certeza con que puede estimarse la verdadera media poblacional a partir de la muestra. Por tanto, es la DE y no el EE de la media la que debe usarse al resumir los datos. Confundir el EE con la DE es un error (o una trampa), dado que en el EE de la media la variabilidad es siempre menor que en la DE. Si un autor indica el EE y el tamaño de la muestra (n), el lector puede calcular la DE mediante la fórmula: $DE = EE \times \sqrt{n}$.

Debe quedar claro qué medida de variabilidad se presenta, por lo que el signo \pm debe ser sustituido preferentemente por las expresiones DE o EE, o dicho aspecto debe especificarse claramente en el apartado «Material y métodos». Porque la expresión $9,3 \pm 2,5$, por ejemplo, no tiene el mismo valor si se presenta como 9,3 (DE= 2,5) que como 9,3 (EE= 2,5).

Al cuantificar los resultados, siempre conviene previamente tener muy claro lo que no se debe hacer en la presentación de datos numéricos (tabla 2)⁴.

3) Conocer el significado del valor de p

Las pruebas estadísticas no sólo sirven para compendiar los datos, sino también para que los investigadores evalúen si las observaciones son compatibles con sus hipótesis. El resultado de dicha comprobación es el llamado nivel de significación, o

valor p^9 . El principal objetivo de las pruebas estadísticas es responder a la pregunta: ¿cuál es la probabilidad de que la diferencia observada se dé al azar?^{10,11}.

Lo primero que hay que realizar es una prueba de hipótesis: se plantea una hipótesis nula (H_0), es decir, $A = B$, y una hipótesis alternativa (H_1), es decir, $A \neq B$. De acuerdo con el resultado del estudio, aceptaremos o rechazaremos la hipótesis nula.

En segundo lugar, hay que cuantificar la diferencia observada entre las dos muestras (A y B) mediante un número denominado estadístico de la prueba (p. ej., t de Student). Cuanto mayor sea la diferencia entre las muestras, mayor será el valor del estadístico de la prueba.

En tercer lugar, hay que calcular el nivel de significación (valor de p). El valor de p es la probabilidad de equivocarse al afirmar que existe una diferencia verdadera. Un valor de $p < 0,05$ indica que en menos de 5 veces de cada 100 que repitiéramos el mismo estudio, nuestro resultado sería falso o se debería al azar. La significación estadística a partir de $p > 0,05$ es una convención, pero es también de interés la información que se obtiene de los resultados con un valor de p algo mayor de 0,05. Cuanto mayor sea el valor de p, más fuerte será la evidencia a favor de la hipótesis nula.

TABLA 2

Puntos que deben evitarse en la presentación de datos numéricos

1. Presentar demasiados datos: simplificar es la clave
2. Presentar sólo los datos estadísticamente significativos: lo no significativo puede ser clave
3. Confundir al expresar los datos con números o letras: recordar que es preferible expresarlos con números, pero entre el 1 y el 10 se pueden expresar también con todas sus letras
4. Presentar los datos con excesiva precisión: la precisión debe ser adecuada, no excesiva
5. Olvidar el denominador en los datos: principalmente en porcentajes, tasas, cocientes, etc.
6. Usar el signo \pm sin indicar a qué medida precede: desviación estándar, error estándar o intervalo de confianza
7. Confundir con el uso incorrecto de guiones: especial atención a los intervalos, por lo que es preferible a veces indicar -3 a 12% que $-3-12\%$ (y evitar el uso de guion)
8. No indicar todas las unidades de todos los datos
9. No presentar todos los datos de todos los sujetos: es importante el flujograma de seguimiento de los sujetos
10. Uso incorrecto de palabras con significado técnico: en estadística, «al César lo que es del César y a Dios lo que es de Dios»...

Modificada de Jiménez⁴.

Muchos creen que el valor de p indica la probabilidad de cometer un error. Pero, obviamente, hay dos formas de cometer un error a partir de los datos. Puede afirmarse que el tratamiento tiene un efecto cuando en realidad no lo tiene y también negarse que tenga un efecto que en realidad sí tiene. El valor de p sólo cuantifica la probabilidad de cometer el error tipo I, o alfa: afirmar erróneamente que el tratamiento tiene un efecto cuando en realidad no lo tiene (falso positivo). El valor de p no proporciona ninguna información sobre la probabilidad de cometer el error tipo II, o beta: concluir que el tratamiento no tiene efecto cuando en realidad sí lo tiene (falso negativo)⁹. En la tabla 3 se esquematizan estos tipos de error y los riesgos asociados en un contraste de hipótesis.

¿Cuándo elegir una p con una o dos colas? Generalmente, es conveniente elegir una p con dos colas (A puede ser mejor o peor que B). La p con una cola se utilizará cuando consideremos que A nunca podrá ser peor que B. Esto es importante, ya que, por ejemplo, un valor de $p=0,04$ con una cola es estadísticamente significativo, mientras que la equivalencia con dos colas ($p=0,08$) no lo sería.

4) No depender exclusivamente de la p

Contar sólo con el uso de los valores de p como prueba para comprobar la hipótesis estadística no nos transmite toda la información cuantitativa relevante. Los niveles de significación (valores de p) se suelen dar a conocer para demostrar que la diferencia observada en relación con un patrón (la hipótesis nula) tendría muy pocas probabilidades de ocurrir meramente por azar; sin embargo, dar a conocer tan sólo el valor de p de una prueba de significación estadística de las diferencias escatima información.

Las típicas aseveraciones $p > 0,05$ o $p < 0,05$ dan poca información sobre los resultados del estudio y se basan en el consenso arbitrario de utilizar el nivel de significación estadística del 5% para definir dos posibles resultados: no significativo ($p > 0,05$) o significativo ($p < 0,05$). Siempre que sea posible, hay que anotar los valores exactos de p (es más informativo), aunque para simplificar la información se puede expresar la significación estadística en tres intervalos ($p < 0,05$; $p < 0,01$; $p < 0,001$).

En cualquier caso, se debe hacer una clara distinción entre la significación estadística y la significación clínica: un resultado significativo no indica necesariamente un efecto real y, por el contrario, un resultado no significativo no indica ausencia de efecto, sino sólo que con los datos actuales no existe evidencia de efecto. Debe haber cierta flexibilidad en la interpretación de la significación estadística de los test. Usamos generalmente un punto de corte en 0,05, pero unos valores de $p=0,04$ (significativo) y $p=0,06$ (no significativo) no son muy diferentes pese a la interpretación tan distinta. Designar cualquier valor de $p > 0,05$ como «no significativo» no es recomendable, pues puede enmascarar resultados con un sugestivo efecto si se describe el valor exacto de p ¹².

Los intervalos de confianza (IC) constituyen una forma más informativa de presentar las pruebas de significación que un simple valor de p . Es más útil presentar los valores muestrales como estimaciones de los resultados que se obtendrían si se estudiara a toda la población. La falta de precisión de un valor muestral (p. ej., la media), que depende tanto de la variabilidad del factor que se investiga como del tamaño limitado del estudio, puede ilustrarse muy bien mediante un IC. Un IC nos lleva de un solo valor estimado (p. ej., la media muestral, la diferencia entre medias muestrales, etc.) a un recorrido de valores que se consideran plausibles para la población. El tamaño muestral influye en la magnitud del EE y éste, a su vez, influye en la amplitud del IC¹². Frente a la interpretación de la significación estadística que sólo nos indica la probabilidad de que la diferencia encontrada se deba al azar, el IC nos traduce la incertidumbre en valores reales, por lo que facilita entender su significado clínico. Por ello, en la medicina basada en pruebas se utiliza a menudo el aforismo de «muera la p , viva el IC».

Sería preferible utilizar la media con un IC razonable de incertidumbre, habitualmente del 95% (que comprende 2 EE por debajo y por encima de la media); por ejemplo, 8,2 (IC del 95%: 6,3-10,1). Esta forma de presentar los datos es particularmente útil si los resultados del estudio son «negativos», porque, a diferencia del término «no significativo» de los test de significación, el IC refleja el amplio intervalo de efectos que son compatibles con los datos observados.

¿Qué es la precisión estadística? Es la inversa de la variancia (a mayor precisión, menor variancia) que aumenta con el aumento del tamaño de la muestra. Una manera de medir la precisión es mediante el IC: cuanto más amplio es el IC, menor es la precisión, hay menor confianza en que el resultado no se deba al azar y pueda extrapolarse a la población^{11,13}.

TABLA 3**Tipos de errores y riesgos asociados en un contraste de hipótesis**

	Realidad		
	Existe diferencia (Ho falsa)	No existe diferencia (Ho verdadera)	
Resultado de la prueba	Diferencia significativa (rechazo de Ho)	No error	Error tipo I (riesgo α)
	Diferencia no significativa (aceptación de Ho)	Error tipo II (riesgo β)	No error

Algunas revistas recomendaron en su momento el uso del IC para sus publicaciones en lugar de los valores de p, como fue el caso de *Bristish Medical Journal*¹², *The Lancet*¹⁴, *Annals of Internal Medicine*¹⁵ y otras^{16,17}, y esta recomendación se ha convertido en un aspecto generalizado y conveniente.

5) Analizar los criterios de inclusión y exclusión

Estos criterios son aplicables tanto a los sujetos de experimentación como al periodo de estudio. Deben definirse de forma precisa, evitando el sesgo de selección.

6) Proporcionar los detalles del proceso de aleatorización

Los principales tipos y métodos de aleatorización son bien conocidos y están claramente descritos en la literatura¹⁸. La asignación aleatoria de las intervenciones en los ensayos clínicos (EC) pretende conseguir que todos los sujetos participantes tengan la misma probabilidad de recibir cada uno de los tratamientos. Los EC deben emplear métodos explícitos y verificables que garanticen que no existe manipulación externa o sesgos en este proceso.

Existen diferentes métodos de aleatorización; los más usados son los siguientes: 1) asignación aleatoria simple; 2) aleatorización en bloques balanceados; 3) aleatorización estratificada, y 4) aleatorización en conglomerados (grupos, o *clusters*). En general, puede decirse que la aleatorización por bloques es preferible a la simple.

7) Proporcionar los detalles del proceso de enmascaramiento

Si en el estudio se utilizaron métodos de enmascaramiento (o cegamiento), deben describirse, diferenciando el simple ciego (tratamiento desconocido por el paciente), el doble ciego (también desconocido para el médico), el triple ciego (también para el monitor) y el cuádruple ciego (también en el análisis estadístico).

8) Informar sobre las complicaciones del tratamiento

Tales complicaciones deben estudiarse con tanta asiduidad como los efectos beneficiosos, y darlas a conocer objetivamente y en detalle. Si no se identifican efectos adversos, debe indicarse, con una explicación de lo que se hizo para buscarlos.

9) Especificar el número de observaciones e indicar las pérdidas

Se deben especificar las pérdidas de sujetos de observación y estudiarlas a fondo a causa de la probabilidad de que dichos individuos sean atípicos de una manera crucial para el estudio. Es importante considerar las pérdidas prealeatorización (por negativa de los pacientes a entrar en el estudio, o bien por ser considerados no elegibles de acuerdo con los criterios de inclusión/exclusión) y postaleatorización (abandonos y retiradas de pacientes durante el estudio).

Conviene detallar bien el diagrama de flujo del estudio: valorar los pacientes que no realizan el tratamiento inicialmente asignado y los que se pierden durante el seguimiento. Una pérdida en el seguimiento superior al 20% compromete la validez de un estudio.

10) Especificar el programa estadístico y las referencias sobre estadística

Deben concretarse las referencias sobre el programa estadístico informático utilizado (licencia disponible, indicando la versión), así como las principales referencias sobre el diseño del estudio y los métodos estadísticos empleados, especialmente en los estudios estadísticos sofisticados, complejos o de uso poco frecuente. Existen múltiples pruebas estadísticas, por lo que es importante conocer y especificar las que se utilizan (tabla 4).

Se ha observado que este hecho, junto con la colaboración en el artículo de expertos en estadística, se relaciona con una mejora en la metodología de los trabajos^{8,19}. Según el grado de participación de los expertos en estadística en el diseño del trabajo y/o análisis e interpretación de los datos, dichos expertos podrían constar como autores o bien ser reseñados en el apartado de «Agradecimientos»²⁰.

11) Definir los términos y evitar el uso no técnico de términos de la estadística

La ciencia es precisa y la estadística más. Por ello, la notificación estadística es una parte importante de un artículo, por lo que no se deben emplear términos estadísticos con su significado común, como «al azar», «normal», «significativo», «correlación» o «muestra»¹.

«Al azar» implica el empleo de un método de aleatorización. «Normal» es un conjunto de distribución de probabilidad descrita por una fórmula específica, también referido como distribución gaussiana. «Significativo» quiere decir que el resultado

TABLA 4

Pruebas estadísticas según el objetivo y la variable de respuesta

Objetivo		Variable de respuesta			
		Cuantitativa normal	Ordinal o cuantitativa no normal	Binomial	Tiempo de supervivencia
Comparar 2 grupos	Independientes	t de Student	U de Mann-Whitney	χ^2 , prueba de Fisher	Long-rank Mantel-Haenszel
	Apareados	t de Student para muestras pareadas	Prueba de Wilcoxon	Prueba de McNemar	Modelos de fragilidad
Comparar 3 o más grupos	Independientes	Análisis de la variancia (ANOVA)	Prueba de Krukall-Wallis	χ^2	Riesgos proporcionales de Cox
	Apareados	Análisis de la variancia (ANOVA)	Prueba de Friedman	Q de Cochran	Modelos de fragilidad
Evaluar la fuerza de la asociación		Correlación de Pearson	Correlación de Spearman		Riesgos proporcionales de Cox
Predecir el valor de una variable en función de otra		Regresión lineal	Regresión no paramétrica	Regresión logística simple	Riesgos proporcionales de Cox
Describir la relación entre una variable de respuesta y diversas variables predictoras		Regresión lineal múltiple		Regresión logística múltiple	Riesgos proporcionales de Cox

Adaptada de Jiménez⁴.

de la prueba formal de una hipótesis estadística se encuentra fuera de una zona determinada de antemano. «Correlación» es la manera específica de medir la asociación entre dos o más variables. «Muestra» suele referirse a una observación o conjunto de observaciones reunidas de una manera bien definida.

Por otro lado, los símbolos más usuales en la expresión de términos estadísticos están basados en las normas dictadas por la Organización Internacional de Estandarización. Pero como la realización tipográfica no es sencilla en cuanto a los símbolos estadísticos, es permisible la utilización de abreviaturas alternativas; así, la desviación estándar se puede definir como DE (en lugar de s), y el error estándar de la media como EE (en lugar de Sx).

12) Controlar los errores de la investigación clínica

Algunas reglas básicas pueden ayudar al lector a detectar posibles errores, dado que la mayoría de éstos se deben al mal uso de técnicas elementales que los investigadores deberían aprender a utilizar por sí mismos⁹.

Un resultado estadísticamente significativo (la consabida p) no supone ninguna garantía de que sea un resultado válido, que pueda extrapolarse a la población general, si no hemos cuidado el control de los errores estadísticos que se producen en los estudios. Centremos este aspecto en el «patrón de oro» de la investigación: el EC.

La muestra de pacientes incluidos en un EC proviene de una población de muestreo y ésta, a su vez, proviene de la población objetivo. Se pretende que los resultados obtenidos en la muestra de pacientes puedan ser extrapolables (generalizados) a la población objetivo. Por ello, deben controlarse al máximo los errores aleatorios y los errores sistemáticos (o sesgos)¹³, con los siguientes objetivos:

1. La muestra seleccionada debe ser representativa de la población. Para ello, deben controlarse los errores aleatorios, mediante el muestreo al azar y el aumento del tamaño muestral¹⁰.
2. La asignación de los pacientes de la muestra a grupos de tratamiento debe dar lugar a una distribución similar de factores pronóstico. Para ello, deben controlarse los errores sistemáticos (sesgos). El grado de ausencia de sesgos va a determinar la validez del estudio.

Existen dos tipos de sesgos principales:

1. El sesgo de selección de pacientes en el muestreo y en la asignación de los grupos de tratamiento. Para su control se deben definir correctamente los criterios de inclusión/exclusión y, posteriormente, hacer una asignación aleatoria de los pacientes a los grupos de tratamiento (aleatorización), considerando las pérdidas pre/postaleatorización.
2. El sesgo de evaluación de los resultados. Para su control se debe recurrir al enmascaramiento y realizar los estudios con método ciego.

Estadística, editores, autores y lectores

La mayor utilización de técnicas estadísticas cada vez más complejas en los artículos biomédicos plantea una serie de temas que debe considerar cada uno de los protagonistas de la publicación científica: editores, autores y lectores^{3,8,21}.

Editores

A los editores, la inclusión de estas técnicas les dificulta el trabajo de selección editorial. En este sentido, las revistas biomédicas de más prestigio cuentan con un comité de revisores especialistas en metodología y estadística que asesora al comité editorial.

Un área de preocupación para los editores radica en la creciente extensión de contenido eminentemente metodológico dentro de la brevedad deseable de un manuscrito. El dilema consiste en encajar dentro de la brevedad deseable de los manuscritos un espacio suficiente dedicado a la descripción de la metodología. Este equilibrio exige a los editores un gran discernimiento (aunque, hoy en día, la versión electrónica de la revista permite volcar como contenido *on-line* lo que no puede ser volcado totalmente en versión escrita).

Autores

A los autores se les exige un mayor conocimiento de los diseños y las técnicas de análisis. Este hecho (sabida la insuficiencia de contenidos metodológicos en el currículo de medicina)²² impone la necesidad de constituir equipos multidisciplinarios compuestos por clínicos y expertos en metodología como mejor estrategia para obtener el máximo rendimiento de la investigación biomédica²³.

Es indudable que la eclosión de las tecnologías de la información y su interacción con las ciencias de la salud debería verse reflejada en la formación de los profesionales de la medicina^{24,25}. En este sentido, es bienvenida la oferta de diplomaturas en estadística por las universidades españolas^{26,27}, la formación a través de revistas biomédicas²⁸⁻³⁰ o las plataformas de formación virtual de las sociedades científicas, entre las cuales Continuum está empezando a tener una especial relevancia³¹.

Lectores

La mayor utilización de técnicas estadísticas disminuye la accesibilidad de los lectores a un porcentaje creciente de originales. Al lector, que no se le supone un experto en metodología en ciencias de la salud, se le exige un grado creciente de conocimientos que le permitan entender y analizar críticamente la literatura científica que consulta³².

Friedma y Phillips³³ encuentran una deficiente formación en estadística de los residentes de pediatría en hospitales norteamericanos, que no mejora con los años de residencia ni con la práctica médica de los adjuntos, por lo que resulta evidente la necesidad de una mejor formación pre/posgraduada en bioestadística médica^{27,34}. Esta conclusión también estaba presente en un estudio sobre el uso de procedimientos estadísticos en artículos originales publicado en una revista pediátrica española⁹ y en varias encuestas realizadas en nuestro país en otros entornos sanitarios²².

Conclusiones

La estadística, tal como la identifica Salsburg³⁵, no debe convertirse en una religión que conduce a la salvación a través del ritual de la búsqueda de valores significativos de la *p*, alcanzando entonces los autores el paraíso prometido de la publicación del trabajo en una revista de alto impacto, sino que es una herramienta útil para el diseño y la realización de estudios médicos, siempre que se utilice adecuadamente.

TABLA 5

Errores estadísticos frecuentes en los artículos de investigación biomédica

1. Presentar datos con precisión innecesaria
2. No presentar todos los datos y todos los pacientes
3. Dividir datos continuos en categorías ordinales sin explicar el porqué ni el cómo
4. Presentar medias de grupos para datos apareados sin exponer los cambios intraparejas
5. Utilizar incorrectamente los datos descriptivos
6. Utilizar el error estándar de la media como estadístico descriptivo o como medida de precisión de una estimación
7. Presentar sólo los valores de *p* de los resultados
8. No utilizar el intervalo de confianza en cada uno de los estadísticos
9. No confirmar que los datos cumplen las asunciones de las pruebas estadísticas utilizadas para analizarlos
10. Utilizar el análisis de regresión lineal sin establecer si la relación es realmente lineal
11. No informar de si se han realizado ajustes para las comparaciones múltiples o cómo se han hecho
12. Presentar innecesariamente comparaciones estadísticas de la línea basal en los ensayos clínicos
13. No definir «normal» o «anormal» cuando se informa de los resultados de pruebas diagnósticas
14. No explicar cómo se han tratado los valores inciertos (equivocos) de las pruebas diagnósticas al calcular sus características (como sensibilidad y especificidad)
15. Utilizar tablas y figuras sólo para almacenar datos y no para ayudar a los lectores
16. Utilizar tablas y figuras en los que el mensaje visual no concuerde con los datos del texto
17. Interpretar estudios con resultados no significativos y baja potencia estadística como «negativos» cuando, de hecho, son no concluyentes
18. No distinguir entre estudios pragmáticos (efectividad) y explicativos (eficacia) al diseñar e interpretar la investigación
19. No presentar los datos en unidades clínicamente útiles
20. Confundir significación estadística con importancia clínica

Modificada de Jiménez⁴.

Un correcto manejo de la estadística por parte del autor mejorará la calidad científica del artículo y, secundariamente, facilitará que éste sea aceptado por el editor y difundido por los lectores. Los textos sobre normas de publicación científica incluyen algunos capítulos sobre cómo hacer un buen uso y evitar el abuso y mal uso de la estadística⁴, e incluyen útiles listas de errores estadísticos. Porque ya se sabe que, a veces, es más fácil empezar por lo que no debemos hacer (tabla 5).

Bibliografía

1. Ballar JC III, Mosteller F. Guidelines for statistical reporting in articles for medical journals: amplifications and explanations. *Ann Intern Med.* 1988; 108: 266-273.
2. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *BMJ.* 1983; 286: 1.489-1.493.
3. González de Dios J. Conocimientos prácticos para elaborar un artículo científico (IV): manejo de la estadística. *Rev Pediatr Esp.* 1998; 54: 340-345.

4. Jiménez J. Presentación de resultados estadísticos y elaboración de tablas. En: *Publicación científica biomédica: cómo escribir y publicar un artículo de investigación*. Barcelona: Elsevier España, 2010; 57-77.
5. Emerson JD, Colditz GA. Use of statistical analysis in "The New England Journal of Medicine". *N Engl J Med*. 1983; 309: 709-713.
6. Mora Ripoll R, Ascaso Terren C, Sentis Villalta J. Tendencias actuales en la utilización de la estadística en medicina. Estudio de los artículos originales publicados en «Medicina Clínica» (1991-1992). *Med Clin (Barc)*. 1995; 104: 444-447.
7. Mora Ripoll R, Ascaso Terren C, Sentis Villalta J. Uso actual de la estadística en investigación biomédica: una comparación entre revistas de medicina general. *Med Clin (Barc)*. 1996; 106: 451-456.
8. González de Dios J, Moya M. Evaluación del uso de procedimientos estadísticos en los artículos originales de «Anales Españoles de Pediatría»: comparación de dos periodos (1989-90 y 1994-95). *An Esp Pediatr*. 1996; 45: 351-360.
9. Glantz SA. Biostatistics: how to detect, correct and prevent errors in the medical literature. *Circulation*. 1980; 61: 1-7.
10. Plasencia A, Porta Serra M. La calidad de la información clínica (II): significación estadística. *Med Clin (Barc)*. 1988; 90: 122-126.
11. Porta Serra M, Plasencia A, Sanz F. La calidad de la información clínica (y III): ¿estadísticamente significativo o clínicamente importante? *Med Clin (Barc)*. 1988; 90: 436-438.
12. Gardner MJ, Altman DG. Confidence interval rather than p values: estimation rather than hypothesis testing. *BMJ*. 1986; 292: 746-750.
13. Rubio Terres C. Diseño estadístico de ensayos clínicos. *Med Clin (Barc)*. 1996; 107: 303-309.
14. Bulpitt CJ. Confidence interval. *Lancet*. 1987; 1: 494-497.
15. Simon R. Confidence intervals for reporting results of clinical trials. *Ann Intern Med*. 1986; 105: 429-435.
16. Altman DG. Statistics: necessary and important. *Br J Obstet Gynaecol*. 1986; 93: 1-5.
17. Morgan PP. Confidence interval: from statistical significance to clinical significance. *Can Med Assoc J*. 1989; 141: 881-883.
18. Zelen M. The randomization and stratification of patients to clinical trials. *J Chron Dis*. 1974; 27: 365-375.
19. Nolasco A, Gascón E, Mur P, Ferrándiz E, Álvarez Dardet C. Utilización de la estadística en publicaciones médicas: una comparación internacional. *Med Clin (Barc)*. 1986; 86: 841-844.
20. Huth EJ. Guidelines for authorship of medical papers. *Ann Intern Med*. 1986; 104: 269-274.
21. Segú Juan JL, Cobo Valeri E. La estadística en medicina clínica. *Med Clin (Barc)*. 1995; 104: 456-457.
22. González de Dios J, Aleixandre Benavent R. Formación e información en pediatría: aproximación a la necesidad de los profesionales. *Acta Pediatr Esp*. 2010; 68: 235-240.
23. Altman DG, Martin Bland J. Improving doctor's understanding of statistics. *J R Statist Soc*. 1991; 154: 223-267.
24. Sancho JJ, González JC, Patak A, Sanz F, Sitges Serra A. Introducing undergraduates to medical informatics. *Med Educ*. 1993; 24: 479-483.
25. González de Dios J, Polanco Allué I, Díaz Vázquez CA. De las facultades de medicina a la residencia de pediatría, pasando por el examen MIR: ¿algo debe cambiar? Reflexiones de una encuesta de residentes de pediatría de cuarto año. *An Pediatr (Barc)*. 2009; 70: 467-476.
26. González JC, Sancho JJ, Sanz F, Sitges Serra A. Conocimientos informáticos de los estudiantes de medicina. Estudio en la Unidad Docente del Hospital Universitario del Mar (Barcelona). *Med Clin (Barc)*. 1991; 97: 598-599.
27. Martín M, Sanz F, Andreu F. Efecto de la introducción de la bioestadística en el currículum de los estudios de medicina. Análisis de una década en la revista «Medicina Clínica». *Med Clin (Barc)*. 1982; 79: 273-276.
28. Seone T, Martín J, Martín Sánchez E, Lurueña Segovia S, Alonso Moreno F. Curso de introducción a la investigación clínica. Capítulo 5: Selección de la muestra: técnicas de muestreo y tamaño muestral. *Medicina de Familia-Semergen*. 2007; 33: 356-361.
29. Seone T, Martín J, Martín Sánchez E, Lurueña Segovia S, Alonso Moreno F. Curso de introducción a la investigación clínica. Capítulo 7: Estadística descriptiva y estadística inferencial. *Medicina de Familia-Semergen*. 2007; 33: 466-471.
30. Gómez González C, Pérez Castán J. Curso de introducción a la investigación clínica. Capítulo 8: Pruebas diagnósticas. Concordancia. *Medicina de Familia-Semergen*. 2007; 33: 509-519.
31. González de Dios J, Hijano Bandera F, Málaga Guerrero S; coordinadores de Continuum. Presentación de Continuum: el portal de formación continuada de la Asociación Española de Pediatría. *An Pediatr (Barc)*. 2013; 79: 343-345.
32. González de Dios J, Buñuel Álvarez JC, Ochoa Sangrador C. La valoración crítica de documentos científicos y su aplicabilidad a la práctica clínica: aspecto clave en la toma de decisiones basada en las mejores pruebas científicas. *Evid Pediatr*. 2006; 2: 37.
33. Friedman SB, Phillips S. What's the difference? Pediatric resident and their inaccurate concepts regarding statistics. *Pediatrics*. 1981; 68: 644-646.
34. Weiss ST, Samet JM. An assessment of physicians knowledge of epidemiology and biostatistics. *J Med Educ*. 1980; 55: 692-697.
35. Salsburg DS. The religion of statistics as practiced in medical journals. *Am Statistician*. 1985; 39: 220-223.