

Evaluación estadística en *Medicina (Buenos Aires)*

Hemos realizado una evaluación estadística de los trabajos publicados como artículos originales en los Volúmenes 56 y 57 de *Medicina (Buenos Aires)* correspondientes a los años 1996 y 1997, respectivamente, con el objeto de evaluar su calidad y de proponer algunas ideas para mejorarla (la calidad). Comparamos los resultados obtenidos en esta ocasión, con los de un estudio similar correspondiente al volumen 47¹ de la misma revista.

La participación del estadístico en la investigación interdisciplinaria es diversa. En algunos casos, participa desde el comienzo, en otros, se le consulta a medida que la investigación avanza, y, en la mayoría de los casos, no hay un estadístico involucrado en la investigación. Numerosos errores evidencian la ausencia de una participación activa del estadístico en dichos estudios y el desconocimiento de los investigadores de la metodología adecuada.

Nuestro propósito es analizar el efecto que este problema produce y proponer algunas ideas para mejorarlo, con el mismo espíritu de Amadeo. P. Barousse cuando dice: «el Comité de Redacción cree que aún debe mantener jueces calificados para el análisis de los manuscritos, vigilar el uso de la complicada estadística en creciente y casi abusiva utilización en función de forzar resultados, y procurar que se use correctamente el idioma castellano»².

Analizamos 103 artículos, realizando una intensa lectura crítica con el fin de evaluar la validez del diseño, el método estadístico utilizado y las conclusiones obtenidas. No hacemos referencia a los diversos métodos de diseño de un experimento, tampoco se trata de determinar si el investigador utiliza el mejor diseño posible, el único propósito es observar si las conclusiones obtenidas son válidas en términos del diseño del experimento,

del tipo de análisis empleado y la aplicabilidad de los tests estadísticos utilizados. No analizamos aquellos artículos que aún estaban en etapa de estudio.

La lista de errores se hizo a medida que los mismos aparecían ya que no se los podía prever con anticipación, debido a que algunos correspondían a la obtención de conclusiones erróneas.

Los artículos fueron evaluados como se indica en la Tabla 1.

Se hallaron 6 tipos de errores:

1) *Utilización inadecuada de las técnicas estadísticas:*

- Todo procedimiento estadístico tiene supuestos que deben verificarse para saber si el modelo es o no aplicable.

- Cualquier estudio en el que el investigador aplica tests estadísticos que no están justificados para sus datos, está incluido en esta categoría.

Ejemplo de estas aplicaciones incorrectas son:

i) El uso del test de Chi cuadrado cuando la frecuencia teórica de alguna celda es menor que 5.

ii) El uso del test de Chi cuadrado para datos continuos.

iii) Falta de homogeneidad de varianzas.

iv) Utilización del test de "t" para muestras independientes cuando en realidad se tiene que utilizar un test pareado.

v) Utilización de un test de "t", luego de aplicar un ANOVA, cuando en realidad corresponde utilizar un test de Tuckey.

vi) Aplicación a puntajes pruebas que corresponden a variables continuas.

vii) Utilización del test de Tuckey sin haber realizado un ANOVA, cuando en realidad lo correcto es utilizar un test de Student.

2) *Obtención de conclusiones erróneas como resultado de aplicar un test estadístico inapropiado:*

- Conclusión de diferencias no significativas entre 2 proporciones, siendo este resultado poco confiable debido al pequeño tamaño de la muestra.

- Expresión incorrecta de existencia de diferencias significativas, cuando en realidad, al aplicar la prueba estadística se obtuvo una $p < 0.025$.

3) *El diseño del estudio no es apropiado para resolver el problema planteado:*

- Empleo repetidas veces del test de "t" cuando en realidad lo correcto es utilizar un análisis de la varianza de un factor.

- Confusión entre pruebas para independencia y homogeneidad, en el análisis de tablas de contingencia.

TABLA 1.- *Artículos con tratamiento estadístico*

	Con Responsable Estadístico	Sin Responsable Estadístico
Con Errores	9	37
Sin Errores	6	29
	Sin Tratamiento Estadístico	
Tratamiento Innecesario		18
No Evaluables		1

4) *Conclusión de la existencia de diferencias significativas sin describir el tipo de test aplicado o el nivel de significación utilizado, indicación del valor de p, sin informar la prueba utilizada.*

5) *Otros errores:*

- Abuso del coeficiente de correlación lineal, empleándolo aún cuando la nube de puntos no sugiere tal relación, decidir que hay o no hay correlación entre dos variables sin analizar la significación del mismo. Conclusiones equivocadas confundiendo regresión por correlación.
- Omisión de las tablas a las que se aplica la prueba de X^2 por lo que no se puede evaluar lo adecuado del análisis.
- Falta de consideración de cuestiones de muestreo, principalmente en lo concerniente al tamaño de la muestra. Se obtienen conclusiones de no-detección de diferencias a partir de muestras muy pequeñas o de gran disparidad de tamaño.
- En otro caso, se comparan grupos de distinto sexo para tamaños de muestras considerablemente diferentes.
- Falta de precisión en los métodos utilizados.
- Se informa el error estándar y no se da el tamaño de la muestra.
- Incoherencia entre los resultados y las tablas presentadas.
- Errores de notación, de terminología y de ortografía, por ejemplo, en un trabajo figura $p = 0.025$, y, en otro $p < 0.95$.
- Hay trabajos en los cuales hay una apariencia de exhaustivo análisis estadístico con profuso detalle de exhibición de tablas, software y de test utilizados en el que precisamente se evidencian numerosos errores, desde la redacción "los pacientes fueron randomizados al azar", hasta la aplicación aparente de ANOVA para muestras independientes cuando los datos corresponden a medidas repetidas en el tiempo.
- En cuanto a los gráficos, su función es aclarar los conceptos, hecho que en ocasiones no ocurre ya que los mismos presentan:
 - a) inadecuadas notas al pie.
 - b) confusión entre gráfico de regresión y correlación.
- En la mayoría de los trabajos no se indica la bibliografía ni el software utilizado para realizar el análisis estadístico.

Tal como notáramos en la revisión de 1987¹, los errores más frecuentes conciernen a nociones estadísticas elementales como el test de "t", X^2 , violación de supuestos, tamaño de muestras. Hay contados estudios donde se aplican métodos estadísticos más complejos, y en esos casos no notamos errores, lo que nos hace suponer la intervención de algún especialista.

Dado que la investigación biomédica atañe a la salud de la población, es deseable evitar errores en la metodología que conduzcan a conclusiones falsas.

Para evitar estos errores y elevar la calidad de los trabajos sugerimos, en concordancia con Shuster²:

- 1.- La inclusión de un estadístico en los grupos de investigación, y su incorporación real al trabajo interdisciplinario.
- 2.- La revisión de los análisis estadísticos de las publicaciones científicas, por parte de bioestadísticos responsables.
- 3.- Para cumplimentar 2, las publicaciones deberían incluir un bioestadístico como editor responsable⁴.
- 4.- Dar instrucciones a los revisores sobre los criterios a seguir e incluirlos en las mismas publicaciones.
- 5.- El mejoramiento de la formación estadística de los investigadores mediante cursos.
- 6.- La publicación de trabajos de divulgación estadística en revistas biomédicas.
- 7.- Ateneos de análisis de casos y revisión crítica de análisis estadístico de trabajos publicados.

Es interesante notar que los errores que encontramos en nuestra revisión están entre los 12 errores más frecuentes encontrados por Schor⁵ en 67 publicaciones médicas y publicados en 1966 y si comparamos los resultados de esta evaluación con la que hiciéramos en 1987, notamos que los errores más frecuentes son sobre los mismos ítems, podemos suponer entonces, la no participación del estadístico en los grupos de investigación.

Myriam C. Núñez, Ana S. Haedo.

Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires

1. Sarchi MI, Garrido D, Haedo AS. Estadística en Medicina. Actas del Coloquio de la Sociedad Argentina de Estadística, 1987; 193-9.
2. Barousse AP. Ingelfinger ¿por qué no Lanari? *Medicina (Buenos Aires)* 1996; 56: 533.
3. Shuster JJ, Binion J, et al. Statistical Review Process. *JAMA*, 1976; 235: 534-5.
4. Robinson GK. Instructions to referees. Do we need some? *Austral J Statist* 1987; 29: 220-1.
5. Schor S, Karten MA. Statistical Evaluation of Medical