CONSIDERACIONES A LAS APLICACIONES DE ALGUNOS METODOS ESTADISTICOS

José Francisco Cumsille G. (*)

Cuad. Méd. - Soc.; XXVIII, 4, 1987. 161 - 166

ABSTRACT: *Using simulated situations, common abuses of statistical methods are shown to lead to false conclusions.*

Key words: STATISTICAL METHODS

INTRODUCCION.

Los lectores habituales de publicaciones científicas en el área de la salud observan que una alta proporción de los trabajos presentados incluyen la utilización de diferentes métodos estadísticos que tienen por finalidad describir fenómenos o analizar información de acuerdo a ciertos objetivos e hipótesis planteadas en cada problema particular.

Sin embargo, se puede comprobar con cierta preocupación, que el empleo de dichos métodos no siempre es el más adecuado para la información disponible. Este hecho no sería tan grave a no ser porque su utilización sirve para derivar conclusiones respecto a los objetivos que dan origen a dichas investigaciones. Es decir, es probable que la inapropiada elección de un procedimiento estadístico lleve a inferencias no acordes con la realidad de la información obtenida.

El objetivo de esta publicación es hacer un llamado de atención en relación a abusos que se suele hacer de algunos métodos estadísticos y cómo esta circunstancia puede conducir a la obtención de conclusiones falsas.

Para tal propósito se presentarán algunas situaciones simuladas, las que analizadas de acuerdo a los criterios observados en las publicaciones, conducirán a deducciones erradas. Se ha preferido la vía de la simulación de problemas en lugar de mostrar resultados de publicaciones ya concretadas para evitar identificarlas innecesariamente.

METODO.

Para evidenciar los problemas que puede originar la mala utilización de los métodos estadísticos, se generan tres ejemplos que involucran dos de las metodologías más recurridas en las publicaciones: la prueba de chi-cuadrado (X^2) y la prueba t de Student para muestras independientes.

Situación 1.

Suponga que se desea comparar el comportamiento de cinco grupos independientes respecto de una variable cualitativa que tiene tres niveles posibles de respuesta; este comportamiento se entenderá expresado en la distribución de frecuencias con que se observan los tres niveles de respuesta en cada grupo. La información disponible arrojó los siguientes resultados:

CUADRO (página siguiente).

(*) División de Bioestadística y Biomatemática. Escuela de Salud Pública, Facultad de Medicina, Universidad de Chile.

RESPUESTAS DE LA VARIABLE

GRUPO	1	2	3	TOTAL
1	34	39	27	100
2	37	39	24	100
3	34	39	27	100
4	33	37	30	100
5	51	50	99	200
TOTAL	189	204	207	600

Si no se analiza con detención la información disponible, se podría llegar a una fría conclusión mediante el empleo de la prueba de X^2: "se observaron diferencias significativas de la distribución de los cinco grupo $X^2 = 31{,}025$, $p < 0{,}0005$".

Si bien esta afirmación no es completamente errónea (dado que basta un grupo con comportamiento diferente para que los cinco difieran) podría inducir una conclusión incompleta. En efecto, los grupos a comparar son diferentes porque uno difiere del resto o los cinco entre sí difieren. Mediante la afirmación anterior, este punto no queda para nada claro no especificado.

El valor de X^2 encontrado de 31,025 es el resultado de sumar las dicrepancias entre las frecuencias observadas en la tabla y las frecuencias esperadas producto del planteamiento de una hipótesis.

La tabla siguiente muestra, para cada celda los aportes parciales al X^2 obtenido:

RESPUESTA

GRUPO	1	2	3	TOTAL
1	0.198	0.735	1.630	2.563
2	0.960	0.735	3.190	4.885
3	0.198	0.735	1.630	2.563
4	0.071	0.265	0.587	0.923
5	2.286	4.765	13.040	20.091
TOTAL	–	–	–	31.025

La suma de los valores parciales de X^2 para el grupo 5 es igual a 20,091 lo cual equivale a cerca de un 54 o/o del total.

Si sólo se considera para el análisis a los cuatro primeros grupos y se recalcula el valor de X^2 éste será igual a 1 ($p = 0.98$); esto es, no hay diferencias significativas entre esos (cuatro) grupos.

Al formar un solo grupo con estos cuatro y contrastarlo con el quinto, el valor de X^2 es aproximadamente 30 con $p < 0.0001$.

Por lo tanto, una conclusión algo más completa que la primitiva será que no hay diferencias significativas entre los cuatro primeros grupos y sí hay entre éstos y el quinto; lo cual equivale a afirmar que el comportamiento de éste último grupo difiere significativamente de los anteriores y entre éstos no se observan diferencias significativas en la distribución de la variable. Para mayor información sobre este tipo de análisis puede consultarse los textos de Fleiss (1) y Agresti (2).

Situación 2.

Otro equívoco observado con bastante frecuencia se refiere al estudio de la asociación entre dos variables, una cualitativa y otra continua en la cual se forman intervalos o clases. Análogamente, se podrían comparar dos o más grupos respecto de una variable continua sobre la cual se forman intervalos.

Cualquiera de estos dos casos genera una tabla de contingencia y a partir de allí parece tentador recurrir a la prueba de X^2.

A modo de ejemplo, considérese la comparación de los valores de presión arterial en dos grupos de personas: un conjunto de pacientes sanos y otro con algún tipo de patología (enfermos). Para la presión arterial se forman tres intervalos obteniéndose los siguientes resultados:

PRESION	SANOS	ENFERMOS	TOTAL
100,1 – 120	40	60	100
120,1 – 140	40	60	100
140,1 – 160	40	60	100
TOTAL	120	180	300

Es decir, la distribución de las 120 personas sanas es uniforme en los tres intervalos y lo mismo ocurre con los 180 enfermos.

Al aplicar la prueba de X^2 ésta arrojará un valor de 0 con $p = 1$; lo que estaría indicando que la distribución de la presión arterial es idéntica en ambos grupos.

¿Qué hubiese ocurrido si en lugar de formar 3 intervalos se consideran 6, cada uno con amplitud igual a la mitad de los anteriores?

Del primer intervalo resultaron los subintervalos 100,1 - 110 y 110,1 - 120 con 10 sanos vs. 30 enfermos en el primero y 30 sanos vs. 30 enfermos en el segundo. Aplicando el mismo procedimiento en los otros intervalos se generará la siguiente tabla:

PRESION	SANOS	ENFERMOS	TOTAL
100,1 – 110	10	30	40
110,1 – 120	30	30	60
120,1 – 130	15	30	45
130,1 – 140	25	30	55
140,1 – 150	5	30	35
150,1 – 160	35	30	65
TOTAL	120	180	300

Nótese que esta tabla es esencialmente igual a la primera considerando seis en lugar de tres intervalos; y a través de ambas se está tratando de contrastar la misma hipótesis: diferencia de valores de presión arterial entre sanos y enfermos.

Sin embargo, el valor X^2 para esta nueva tabla es 22,6 con $p < 0.0005$; es decir, la distribución (o valores) de presión arterial no es la misma para ambos grupos.

¡Esta conclusión es diametralmente opuesta a la generada con la primera tabla!

Frente a esta evidente contradicción, es lícito preguntarse las razones que la producen, ya que obviamente es una situación anómala.

Ante este tipo de resultados parecen cobrar fuerza afirmaciones como las siguientes: "la estadística es el arte de mentir científicamente" o "con la estadística se puede probar cualquier cosa". Lógicamente, desde nuestro particular punto de vista, estas u otras afirmaciones similares no tienen ningún asidero y sólo confirman que los usuarios de los métodos estadísticos no prestan debida atención a los conceptos teóricos que sustentan una u otra prueba estadística.

Lo que ha ocurrido en este caso es relativamente simple: la prueba de X^2, basada en una tabla de contingencia, se utiliza para estudiar la relación o asociación entre dos variables ***cualitativas*** (conocida como prueba de asociación) o para estudiar la distribución de dos o más grupos (muestras) frente a una variable ***cualitativa*** (conocida como prueba de homogeneidad). El problema que se está resolviendo ahora, no corresponde a ninguna de las dos aplicaciones descritas anteriormente, debido a que la variable estudiada, presión arterial, es una variable cuantitativa. Por lo tanto no cabe la aplicación de la prueba de X^2 en este caso y se deberá recurrir a otra prueba (por ejemplo, la prueba t de Student para muestras independientes, siempre que los supuestos que sustentan dicha prueba se cumplan).

Por lo tanto, quienes procedan a resolver un problema como el descrito mediante la prueba de X^2, lo que están haciendo es "forzar" la información para que dicha prueba "se pueda aplicar". Pero obviamente, como queda demostrado en el ejemplo, los resultados obtenidos por dicha vía no serían nada confiables.

Situación 3.

Se acaba de mencionar que una de las posibles soluciones al problema planteado en el acápite anterior podría ser la aplicación de la prueba t de Student para muestras independientes.

Esta prueba permite contrastar dos promedios provenientes de sendas muestras independientes.

La correcta aplicación de esta prueba está condicionada por el cumplimiento de dos supuestos básicos: que la variable estudiada tenga distribución normal y que las varianza poblaciones sean iguales (supuesto de homoscedasticidad). Si estos supuestos se cumplen, entonces el valor calculado para la estadística t está dado por:

ECUACION (I)

(página siguiente)

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (\mathrm{I})$$

donde n_1, $\bar{X}_1$ y S_1^2 son el tamaño de muestra, el promedio y varianza para la primera muestra y n_2, $\bar{X}_2$ y S_2^2 los valores equivalentes para la segunda muestra.

Este valor se compara con un valor de la tabla t con $n_1 + n_2 - 2$ grados de libertad para algún nivel de significación prefijado. A partir de esta comparación se decidirá si hay o no diferencias significativas entre los promedios de las poblaciones de donde provienen las muestras.

Sin embargo, si para un problema particular no se satisfacen uno o ambos supuestos, es posible llegar a conclusiones falsas.

Para ilustrar este caso se han generado, computacionalmente, dos muestras independientes con distribución normal de 10 y 50 observaciones cada una, es decir controlando el primer supuesto enunciado. En cuanto a las varianza, la generación de los valores de las muestras se hizo de tal forma que asegurara que el supuesto de igualdad de varianzas no se cumple. Este procedimiento se repitió en dos oportunidades haciendo algunas variaciones en cuanto a los valores de las varianzas muestrales:

a) la información para el primer caso fue:

CUADRO. (abajo)

Al reemplazar estos valores en la expresión (I) se obtiene un valor de t = 2,49 que con 58 grados de libertad entrega un $p < 0.05$, con lo cual se concluye que hay diferencias significativas entre los promedios de las poblaciones.

Sin embargo, si se observan las varianzas obtenidas en las muestras ($S_1^2 = 186{,}7$ y $S_2^2 = 18{,}33$) y se realiza una prueba de igualdad de varianzas se concluye que hay diferencias significativas entre las varianzas poblacionales ($p < 0.001$). Es decir, no se satisface el segundo supuesto que sustenta a dicha prueba y por ello no es aplicable, en este caso, la expresión (I) para comparar los promedios poblacionales.

En esta situación, podría utilizarse una prueba t de Student considerando el no cumplimiento de la homoscedasticidad.

La estadística t en este caso adquiere la forma:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

con grados de libertad (g.1.) dados por:

ECUACION. (ver página siguiente).

Si para los valores dados del problema se aplica la expresión II el valor de t será 1,32 con 9 grados de libertad aproximadamente, con lo cual se tiene $p > 0.10$. Esto implicaría que no hay diferencias

GRUPO	TAMAÑO MUESTRA n	MEDIA $\bar{X}$	VARIANZA S^2	DESVIACION STANDARD S
1	$n_1 = 10$	$\bar{X}_1 = 60{,}763$	$S_1^2 = 186{,}7$	$S_1 = 13{,}664$
2	$n_2 = 50$	$\bar{X}_2 = 55{,}003$	$S_2^2 = 18{,}33$	$S_2 = 4{,}281$

$$\text{g.l.} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

significativas entre los promedios muestrales, conclusión contraria a la obtenida por medio de la utilización de la prueba t bajo el supuesto de igualdad de varianzas.

b) En el caso anterior, la mayor varianza correspondía a la muestra más pequeña. Al simular una situación inversa, se obtuvieron los siguientes resultados:

GRUPO	n	$\bar{X}$	S^2	S
1	10	59,7	12,89	3,59
2	50	55,7	110,67	10,52

Al igual que en el caso anterior, las varianzas son significativamente diferentes ($p < 0.005$) con lo cual no debería recurrirse a la expresión (I) para docimar la igualdad de los promedios. Si se utiliza dicha expresión, el valor t sería 1,182 con $p > 0.10$, con lo cual se concluiría que no hay diferencias significativas entre los promedios de las poblaciones.

En cambio, dado que las varianzas son significativamente diferentes, se debería utilizar la expresión II; así el valor de t con 43 grados de libertad es 2,1375 con $p < 0.02$; es decir hay diferencias significativas entre los promedios poblaciones a dicho nivel.

CONCLUSIONES.

Como fue mencionado en la introducción, el objetivo de este trabajo es ser una alerta respecto de las consecuencias de una mala elección de una determinada metodología estadística frente a un problema dado.

Aquí se han mencionado tres casos que evidencian que la "automedicación" estadística puede conducir a "reacciones adversas o efectos colaterales" impredecibles y peligrosos, como son las conclusiones falsas. Esto ha quedado claramente demostrado con las situaciones descritas.

Sin embargo, hay otras aplicaciones estadísticas no consideradas aquí y que también se observan con alguna frecuencia en la literatura con escasa credibilidad en sus conclusiones: otras aplicaciones de X^2, coeficiente de correlación, repetidas pruebas t para analizar 3 o más grupos, etc.

Por otra parte, hay numerosos ejemplos en los cuales la información disponible podría claramente dar para un buen análisis descriptivo; sin embargo, se insiste en utilizar metodologías inferenciales (particularmente recurrir a un valor p) en la creencia que si no es así el trabajo tiene poco valor. Esto está muy lejos de la realidad: es preferible un buen análisis descriptivo para una cierta información disponible, que un mal análisis basado en pruebas de significación que sólo pueden conducir a conclusiones erradas cuando no se satisfacen los supuestos y requerimientos que las sustentan.

Los comentarios entregados en este trabajo nos debieran hacer reflexionar respecto de la aplicación de los métodos estadísticos en la investigación en el área de la salud, de modo de revisar críticamente lo que hemos ya realizado y planificar cuidadosamente el análisis estadístico de problemas futuros. Soluciones puede haber muchas y muy variadas, siendo la más importante la toma de conciencia de los usuarios de la estadística respecto de la problemática descrita, con el objeto de poder producir un avance científico multidisciplinario de las ciencias de la salud.

RESUMEN

El objetivo de esta publicación es hacer un llamado de atención en relación con abusos en que se suele incurrir de algunos métodos estadísticos y cómo, esta circunstancia, puede conducir a la obtención de conclusiones falsas.

Se presentan algunas situaciones simuladas, para no caer en personalizaciones, las que analizadas de acuerdo con los criterios observados en las publicaciones, conducen a deducciones erradas.

SUMMARY

The aim of this publication is to draw attention to common abuses of some statistical method and how these can lead to false conclusions.

To avoid individual references, simulated situations are presented and these are analysed according to current publication criteria, showing that they lead to wrong conclusions.

RESUME

Le propos de cette publication est celui d'attirer l'attention sur les abus commis avec l'utilisation de certaines méthodes statistiques, ce qui peut mener à tirer des fausses conclusions.

On présente quelques situations simulées, pour ne pas tomber dans des personnalisations, lesquelles, analysées d'après les critères observés dans les publicactions, conduisent à des déductions erronées.

BIBLIOGRAFIA.

1) Fleiss, J. L. Statistical Methods for rates and proportions. 2nd. Ed., John Wiley and Sons Inc, 1981.

2) Agresti, A. Analysis of Ordinal Categorical Data. John Wiley and Sons, 1984.

3) Snedecor, G,W. and Cochran, W.G. Statistical Methods. 6th. Ed. Iowa State University Press, 1967.

4) Cochran, W.G. and Cox, G. Experimental Designs. John Wiley and Sons Inc, 1957.

5) Dunn, O.L. Basic Statistics: A primer of the Biomedical Sciences. John Wiley and Sons Inc. 1978.

6) Siegel, S. Nonparametric Statistics for the Behavioral Sciences, Mc Graw-Hill, 1956.

7) Steel, R. y Torrie, J. Bioestadística: principios y procedimientos. 2a. Ed. Mc Graw-Hill, 1985.

8) Cumsille, F.; Muñoz, S. y Rodríguez. J. Inferencia Estadística. Apunte Mimeo Nº 1-4922, 2a. Ed. Escuela de Salud Pública, Universidad de Chile, 1987.

AGRADECIMIENTOS: A los Profesores Sres. Luis Marchant y Claudio Silva por sus valiosas críticas y sugerencias a este trabajo.