

La necesidad de enseñar argumentos persuasivos en cursos de Bioestadística **The Need to Teach Persuasive Arguments in Biostatistics Courses.**

José Renato De Nóbrega

Unidad Docente de Bioestadística y Evolución del Departamento de Ecología, Escuela de Biología, Facultad de Ciencias, Universidad Central de Venezuela.

renato.nobrega@ciens.ucv.ve

Recibido 23/05/2014 – Aceptado 12/12/2015

Resumen

La enseñanza de competencias para elaborar argumentos persuasivos sobre un fenómeno a partir de evidencia cuantitativa, ha sido un aspecto descuidado en los cursos universitarios de Bioestadística. Este descuido, en parte, es resultado de la rígida aplicación de las pruebas de significación estadística. En este artículo se exponen las principales críticas sobre el uso y aplicación de esta técnica estadística. Se desarrolla como propuesta que promueva estas competencias, el exponer las limitaciones e interpretaciones erróneas de dichas pruebas y difundir enfoques de análisis estadísticos alternativos.

Palabras Claves: Bioestadística, Pruebas de Significación, Análisis Exploratorio y Confirmatorio

Abstract

Teaching competences to present persuasive arguments about a phenomenon based on quantitative evidence, have been a neglected aspect in Biostatistics undergraduate courses. This oversight is due in part from the rigid application of statistical significance testing. Major criticism of the use and application of this statistical technique is presented in this article. A proposal to promote these competences is developed: expose the limitations and misinterpretations of such testing and diffuse alternative approaches in statistical analysis.

Keywords: Biostatistics, Significance testing, Exploratory and Confirmatory Analysis

Introducción

Abelson (1998) señala que en las asignaturas y textos básicos de Estadística se descuida un aspecto fundamental: la naturaleza argumentativa de las afirmaciones estadísticas. Se aprenden técnicas detalladas que conducen a declarar conclusiones mecánicamente, en detrimento de una buena retórica o argumentación. Desde su visión, uno de los objetivos de la enseñanza de la Estadística debería ser el desarrollar en el aprendiz la competencia para exponer argumentos persuasivos sobre un fenómeno, a partir

de la evidencia cuantitativa. El autor, plantea cinco criterios para catalogar un argumento como persuasivo, que denomina criterios MAGIC - Magnitud, Articulación, Generalidad, Interés y Credibilidad – definidos así:

- *Magnitud*: un argumento estadístico será más robusto mientras mayor sea la magnitud cuantitativa en la que se apoya la afirmación cualitativa declarada.

- *Articulación*: un argumento estadístico tendrá mayor articulación mientras mayor sea el detalle con que el que expresamos las conclusiones.

- *Generalidad*: el argumento será más general mientras más amplio sea el espectro de casos o situaciones a los cuales se puedan extender las conclusiones.

- *Interés*: el argumento será más interesante en la medida en que genera un cambio de opinión sobre el tema, en términos del número de proposiciones teóricas y aplicadas que deben modificarse ante los resultados.

- *Credibilidad*: un argumento tendrá mayor credibilidad mientras mayor sea su coherencia teórica y la rigurosidad del método empleado.

En este trabajo, se plantea que parte de la explicación de este descuido en la elaboración de conclusiones bien argumentadas, se debe a que la mayoría de los cursos y textos básicos de Estadística dirigen la atención principalmente a la clásica prueba de significación de una hipótesis nula; ésta suele presentarse a manera de receta, en pasos ordenados que conducen a una conclusión simple: rechazar o no la hipótesis, declarando el resultado observado como estadísticamente significativo o no. No se establece – ni en los textos ni en clases – un juicio sobre el resultado que vaya más allá de esta declaración. De allí la pobreza en la argumentación. Se plantean, entonces sugerencias para intentar mejorar esta situación.

Procedimiento

Se procedió, en primer lugar, a una revisión - no exhaustiva - de los principales aspectos considerados en artículos críticos sobre el uso y aplicación de técnicas estadísticas en la investigación. Se revisaron también textos básicos - de uso frecuente en cursos de Bioestadística - con miras a establecer si trataban estos aspectos críticos. A partir de estas revisiones, se propone la inclusión de algunos conceptos y procedimientos en las unidades temáticas a desarrollar en un curso de Estadística dirigido a estudiantes de Biología; en algunos casos nos centramos en la Ecología, nuestra especialidad.

Resultados y discusión

De la revisión de la literatura crítica se seleccionaron tres aspectos que consideramos fundamentales: 1. la necesidad de reconocer el carácter exploratorio y/o confirmatorio de una investigación; 2. las críticas a la prueba de significación estadística; 3. la consideración de enfoques alternativos al representado por estas pruebas.

La finalidad del estudio: ¿exploratorio o confirmatorio?

Tukey (1980) señala la necesidad de reconocer los dos enfoques - exploratorio y confirmatorio - que se presentan en una investigación, resaltando que ninguno de ellos es suficiente por sí solo. Considera que intentar reemplazar uno por otro es incorrecto. Señala que el enfoque exploratorio no es mera estadística descriptiva, sino más bien una actitud flexible en el juicio de los datos y en el uso de sus técnicas, con el fin de detectar posibles patrones que conduzcan a la generación de hipótesis de investigación de interés. Por otra parte, el enfoque confirmatorio apunta principalmente a la puesta a prueba de una clara y bien establecida hipótesis de investigación y demanda un diseño muestral y/o experimental riguroso. En opinión del autor no hay dilema en cuanto a enseñar uno u otro enfoque: se requiere enseñar ambos.

Para Abelson (1998), el estilo del argumento estadístico detrás de estos enfoques presenta matices distintos: el liberal - orientado a la exploración y detección de posibles efectos sistemáticos - y el conservador - manifestado en una actitud confirmatoria hacia los resultados.

La prueba de significación de una hipótesis nula y sus limitaciones.

Existe un abundante literatura crítica sobre las limitaciones inherentes al paradigma clásico de la inferencia estadística - la prueba de significación de una hipótesis nula - su uso indebido y erróneas interpretaciones. Sólo como una muestra, pueden verse algunos de estos planteamientos en Yoccoz (1990), Cox (1982), y Morgan (2003). El análisis de los datos de una investigación no debe reducirse a la búsqueda obsesiva y exclusiva de significación estadística; deben considerarse otros aspectos igualmente importantes como la cualificación de la magnitud de la diferencia observada (tamaño del efecto) y su importancia práctica o científica en el campo de conocimiento particular. Significación estadística no implica necesariamente significación práctica. Sobre medidas de tamaño de efecto estandarizado existen varias revisiones importantes como las realizadas por Cohen (1992), Valera Espín y Sánchez Meca (1997), Frías *et al* (2000), Kotrlik y Williams (2003), Nakagawa y Cuthill (2007) y Ledesma *et al* (2008).

Los enfoques alternativos

En la literatura crítica se recomienda la necesidad de difundir la existencia de otros enfoques alternativos al paradigma clásico de la inferencia estadística: el enfoque bayesiano, el enfoque de la teoría de información, el enfoque de la razón de verosimilitud. Para comprender estos enfoques alternativos, el concepto de *verosimilitud condicional* sobre la data es fundamental. Presentaciones de estos enfoques pueden verse en Cohen (1994) Anderson *et al* (2001), Ellison (2004), y Burnham *et al* (2010).

Recomendaciones

Con base en este diagnóstico, se sugiere incorporar y hacer énfasis en los siguientes conceptos y aspectos en el temario de los cursos:

1. En la Introducción:

Se deben destacar las fases exploratorias y confirmatorias que pueden estar presentes en una investigación y en las que el análisis estadístico puede intervenir; también resaltar la importancia de la primera para generar potenciales hipótesis de investigación, y de la segunda para la confirmación de hipótesis preestablecidas; y por último, enfatizar sobre el carácter flexible que debe tenerse en un estudio exploratorio, y de la importancia de diseños adecuados de recolección de datos en un estudio confirmatorio.

2. En el tema de Estadística Descriptiva:

Aparte de presentar los estadísticos descriptivos y técnicas de representación gráfica básicas, esta unidad representa una buena oportunidad para trabajar la noción de búsqueda exploratoria de patrones, en particular con casos de estudios comparados de dos muestras de observaciones.

Recomendamos un ejercicio en el que solicitamos al estudiante seleccionar aleatoriamente una muestra de 30 observaciones (mediciones) de cada una de dos poblaciones grandes de individuos humanos (500 por población). Las poblaciones son bivariadas: una variable dicótoma (cero-uno) dada por el sexo del individuo y una variable cuantitativa continua, dada por el peso corporal en kilogramos. De esta forma, el estudiante aborda desde el inicio una experiencia aleatoria - generada por sí mismo - y enfrenta una situación de dos tipos de datos distintos que requerirán estadísticos particulares para su descripción y exploración. Algunas de las técnicas de análisis exploratorio de datos - expuestas y popularizadas por Tukey (1977) - caben muy bien aquí. Dos de ellas han sido particularmente importantes en nuestra experiencia: el gráfico en tallos y hojas (stem and leaf) y el gráfico en caja y bigotes (box and whisker plot), con inclusión de vallas internas y externas para detectar valores inusitados o anómalos, moderados a extremos. Entre los textos básicos en ciencias biológicas y médicas revisados, que incluyen estos dispositivos en el capítulo dedicado a la Estadística Descriptiva, tenemos el de Susan Milton (1994) - aunque los señala como opcionales - y Daniel (2005) en su cuarta edición - el cual destaca inclusive que son ejemplos de las denominadas técnicas de análisis exploratorio, las cuales permiten revelar tendencias y relaciones en los datos, aparte de facilitar su descripción y resumen. Nuestra experiencia nos indica que para el estudiante es mucho más sencillo elaborar los intervalos y marcas de las clases de una distribución de frecuencia, a partir de la construcción previa del gráfico en tallos y hojas. Este mismo instrumento les permite también ubicar fácilmente los estadísticos de posición - caso de los cuartiles - que utilizarán para elaborar el gráfico de caja y bigotes.

Culminadas las representaciones, es oportuno ensayar argumentos para calificar las diferencias encontradas. Solemos iniciar solicitando a los estudiantes que califiquen una diferencia de 5 Kg en el peso promedio de las dos muestras, de acuerdo a las categorías

siguientes: grande, moderada o pequeña. Agregamos que la calificación debe ser justificada. Ante esta petición que los obliga a ensayar un argumento que persuada, solicitan información extra acerca de otras variables, generalmente la edad: "no es lo mismo una diferencia de 5 Kg. entre grupos de adultos que entre grupos de niños de la misma edad" como nos señaló un estudiante. Esto es un avance en la búsqueda de articulación y credibilidad en la argumentación. Aprovechamos para indicarles que la calificación de la diferencia dependerá también de la magnitud de la variabilidad del peso dentro de cada muestra de observaciones e introducimos los estadísticos estandarizados de tamaño de la diferencia o efecto observado: un intento de calificación de una diferencia entre promedios de grupos al compararla con las medidas de la dispersión dentro de cada grupo. Estas magnitudes pueden evaluarse siguiendo los criterios establecidos por Cohen (1992). De esta forma, el estudiante se inicia calificando las diferencias según criterios prácticos relativos, antes de sumergirse en la comprensión de la significación estadística de estas mismas diferencias en el tema o unidad de inferencia.

3. En el tema de los fundamentos básicos de la Teoría de Probabilidades.

En este tema, en cursos de pregrado, ya parece necesario incorporar el concepto de *función de verosimilitud*, una vez que el estudiante haya ejercitado y comprendido el concepto de *función de probabilidad*. Se requiere que el docente destaque la sutil diferencia entre ambos. Un ejemplo de aplicación de estas dos funciones en un experimento binomial sencillo (revoleo repetido de una moneda) nos ha servido en ese sentido.

En la presentación y ejercitación del Teorema de Bayes vale la pena destacar al estudiante que la conexión entre probabilidades previas y posteriores se establece a través de una función de verosimilitud. Recomendamos ensayar un ejemplo de aplicación del teorema en el que la partición previa del espacio de posibilidades corresponda a diferentes valores (discretos) de un parámetro (introduciendo así la noción de hipótesis sobre dicho parámetro), y que las probabilidades asignadas fueron establecidas a partir de información previa sobre el fenómeno. Consideramos que de esta forma se siembra la semilla del enfoque bayesiano y se afianza el concepto de verosimilitud. Ninguno de los libros básicos examinados tratan el tema de la verosimilitud; a lo sumo se señala, en el tema de Inferencia, la existencia del método de máxima verosimilitud para obtener estimadores. El Teorema de Bayes se presenta en todos pero de forma escueta, sin destacar su importancia fundamental como centro de otro enfoque posible. Sin embargo, en el texto de Wonnacot y Wonnacot (2004) ya se presentan capítulos completos sobre inferencia bayesiana y verosimilitud. Lamentablemente, en la edición en español, se tradujo el término "likelihood" como "probabilidad" en vez del apropiado "verosimilitud", lo cual no contribuye en nada en la ya comentada diferencia entre los conceptos de función de probabilidad y función de verosimilitud.

4. En el tema de la Inferencia Estadística.

Luego de exponer sus fundamentos, los ejercicios de aplicación de la prueba de significación de una hipótesis nula deben retomar lo aprendido previamente acerca de la importancia o significación práctica de la diferencia detectada. Los textos básicos revisados

no comentan sobre las críticas que se han formulado a esta prueba, ni sobre su uso incorrecto. Es labor entonces del docente el hacerlo. Recomendamos algunos puntos a considerar.

El someter a prueba una hipótesis estadística nula de no diferencia - diferencia exactamente igual a cero - en la respuesta promedio de una variable bajo dos condiciones (experimentales o dadas por el ambiente) es un ejercicio trivial: no es cierta de entrada. Se podría alegar que no es más que un modelo base, de comparación, pero muy simple. El revelar esto al estudiante lo obliga - de nuevo - a la reflexión y a preguntarse por aquellas diferencias de importancia práctica en su particular disciplina. Dicho de otra manera, a pensar en aquellos valores en la respuesta promedio, que si bien son distintos, pueden considerarse equivalentes desde un punto de vista práctico o científico. La exposición aquí de las denominadas pruebas de equivalencia sería interesante, aunque tal vez no haya suficiente tiempo para ejercitarlas en profundidad en un curso de pregrado. Un excelente texto en español, con abundantes ejemplos de aplicación en el campo biológico, es el de Clifford Blair y Taylor (2008). A nivel de postgrado, el manejo de las pruebas de equivalencia debería ser obligatorio. Un buen ejemplo de aplicación en el estudio de tendencias poblacionales, que utilizamos en este nivel, puede verse en Dixon y Pechmann (2005).

El valor de probabilidad del estadístico de prueba observado (valor-p) debe ser utilizado de manera flexible, como una ayuda a la evaluación de la hipótesis estadística: mientras más pequeño, menor será la consistencia o compatibilidad del resultado con lo que plantea la hipótesis nula, pero si es un valor alto, esto no nos debe llevar a concluir sobre la veracidad de dicha hipótesis. La declaración "aceptar la hipótesis nula" debería ser definitivamente erradicada, y en algunos textos ya no se utiliza. Debe señalarse que los valores de significación convencionales - tales como 0,05 y 0,01 - son arbitrarios; en la mayoría de los casos no hay detrás de ellos un procedimiento objetivo que haya determinado que deban ser estos y no otros. Estas convenciones pueden ser razonables, pero no deben entenderse como preceptos o mandamientos. Un punto de reflexión y discusión con los estudiantes cabe aquí: el seguir las convenciones libera de la responsabilidad de un juicio propio.

Debe repetirse varias veces la correcta interpretación del valor-p; ayuda en este sentido una declaración negativa: no es la probabilidad de que la hipótesis sea cierta dada la diferencia observada. Si se quiere una medida tal, o una medida del soporte que ofrece la data a dos o más hipótesis en disputa, otros enfoques serán necesarios. Aquí cabe dar entrada a estas alternativas, como el enfoque bayesiano.

En ejercicios en los que abordamos con los estudiantes la aplicación simultánea de la prueba de significación y los enfoques alternativos como el bayesiano, a un mismo conjunto de datos, solemos preguntar sus opiniones al culminar la experiencia. Algunas de sus declaraciones se exponen a continuación: "descubrí que la hipótesis nula no es el único o principal protagonista"; "el argumento se enriquece cuando podemos asignar diferentes grados de creencia o probabilidades a distintas hipótesis de interés"; "la seguridad que creí tenía cuando aplicaba la prueba de significación era ficticia"; "mientras la prueba de

significación pone todo en dos colores (blanco o negro), los enfoques alternativos generan una gama de tonos grises”.

Reflexiones Finales

La enseñanza de la Estadística no debe descuidar el modo y estilo en que se argumentan las conclusiones de una investigación, a partir de la evidencia cuantitativa examinada. Debe darse cabida al estilo de argumentación flexible en el análisis exploratorio. El juicio sobre un resultado debe considerar no sólo su significación estadística sino también su significación práctica; una adecuada argumentación requiere reconocer las limitaciones de las pruebas de significación y sobre todo evitar las erróneas interpretaciones que de ellas hacemos. Conocer las alternativas a estas pruebas puede ayudar en ese sentido por lo que debe irse hacia una mejora a nivel curricular que dé cabida a estas alternativas en los programas de estadística dirigidos a estudiantes de biología.

Referencias Bibliográficas

- Abelson, R.P. (1998). *La estadística razonada: reglas y principios*. Barcelona: Paidós.
- Anderson, D.R., Link, W.A., Jonhson, D.H. y Burnham, K.P. (2001). Suggestion for presenting the results of data analysis. *The Journal of Wildlife Management*, 65: 373-378.
- Burnham, K.P., Anderson D.R. y Huyvaert K.P. (2010). AIC model selection and multimodel inference in behavioral ecology: some background, observations and comparisons. *Behav Ecol Sociobiol*, 65: 23-35.
- Clifford Blair R. y Taylor R.A. (2008). *Bioestadística*. México: Pearson-Prentice Hall.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112: 155-159.
- Cohen, J. (1994). The earth is round ($p < 0.05$). *American Psychologist*, 49: 997-1003
- Cox, D.R. (1982). Statistical significance tests. *Br.J. clin. Pharmac.* 14: 325-331.
- Daniel, W.W. (2005). *Bioestadística. Base para el Análisis de las Ciencias de la Salud*. Cuarta Edición en Español. México: Limusa Wiley.
- Dixon, P.M. y Pechmann, J. H. K. (2005). A statistical test to show negligible trend. *Ecology* 86: 1751-1756.
- Ellison, A.M. (2004). Bayesian inference in Ecology. *Ecology Letters*, 7: 509-520.
- Frias, N. M. D., Llobel J.P. y Garcia Pérez J.F. (2000). Tamaño del efecto del tratamiento y significación estadística. *Psicothema*, 12: 236-240.
- Kotriik, J.W. y Williams, H.A. (2003). The incorporation of effect size in information technology, learning and performance research. *Information Technology, Learning and Performance Journal*, 21: 1-7.
- Ledesma, R., Macbeth G. y Cortada de Kohan, N. (2008). Tamaño del efecto: revisión teórica y aplicaciones con el sistema estadístico ViSta. *Revista Latinoamericana de Psicología*, 40:425-439.
- Morgan P.L. (2003) Null hypothesis significance testing: philosophical and practical considerations of a statistical controversy. *Exceptionality*, II: 209-221.
- Nakagawa, S. y Cuthill, I.C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, 82: 591-605.

- Susan Milton, J. (1994). *Estadística para Biología y Ciencias de la Salud*. 2ª Edición. México: Interamericana- McGraw Hill.
- Tukey, J.W. (1977). *Exploratory Data Analysis*. London: Addison-Wesley.
- Tukey, J.W. (1980). We need both exploratory and confirmatory. *The American Statistician*, 34: 23- 25.
- Valera Espín, A. y Sánchez Meca, J. (1997). Pruebas de significación y magnitud del efecto: reflexiones y propuestas. *Anales de Psicología*, 13: 85-90.
- Wonnacott, T.H. y Wonnacott, R.J. (2004). *Introducción a la Estadística*. México: Limusa Wiley.
- Yoccoz, N. (1990). Use, overuse and misuse of significance tests in evolutionary biology and ecology. *Bull. Ecol. Soc. Am.*, 72:106-111.