

La (significativa) importancia biológica de la no-significancia estadística

ALEJANDRO G FARJI-BRENER ✉

Laboratorio Ecotono, CRUB, Universidad del Comahue, Bariloche, Argentina

RESUMEN. El rechazo de hipótesis biológicas incorrectas como consecuencia de resultados estadísticamente no-significativos es comúnmente despreciado como un aporte para el crecimiento del conocimiento científico. Los argumentos previos contra esta creencia se han basado principalmente en las consecuencias negativas de la escasa divulgación de resultados estadísticamente no-significativos más que enfocarse sobre el mérito lógico de descartar ideas erróneas. En este ensayo, discutiré cómo los métodos estadísticos clásicos (i.e., estadística frecuentista), y la base epistemológica de la cual derivan, están basados en la eliminación de ideas falsas más que en el descubrimiento de ideas potencialmente verdaderas. El menosprecio de buenas investigaciones y/o su eventual rechazo editorial debido a su no-significancia estadística niega, entonces, la principal ventaja de los métodos estadísticos tradicionales: refutar ideas incorrectas. Rechazar una hipótesis biológica basándose en una investigación de buena calidad y alto poder estadístico es una de las formas más robustas de aprender sobre el funcionamiento de la naturaleza.

[Palabras clave: hipótesis biológicas, hipótesis estadísticas, resultados no-significativos]

ABSTRACT. **The (significant) biological significance of statistical non-significance:** Rejecting incorrect biological hypotheses as a consequence of statistically non-significant results is commonly undervalued as a step in the growth of our knowledge. Previous arguments against this incorrect belief had been largely based on the negative consequences associated with not publishing papers with statistically non-significant results rather than on the intrinsic epistemological merits to discarding erroneous ideas. In this essay, I will discuss how classical statistical methods and the epistemological approach from which these tools are derived are largely based on the elimination of falsehood rather than on the discovery of truth. The rejection of researches with statistically non-significant results denies the main advantage of classical hypothesis-testing methods. In fact, the rejection of an incorrect biological hypothesis based on high-quality research is one of the most powerful ways to understand nature.

[Keywords: biological and statistical hypotheses, hypothesis testing, non-significant results]

Aprender de nuestras ideas incorrectas es un paso básico para conocer con cierta certeza algún aspecto del funcionamiento de los sistemas naturales. Para ello, los científicos ponemos a prueba diferentes hipótesis biológicas que compiten entre ellas para expli-

car un fenómeno. Para determinar cuál de ellas es la más correcta, se proponen diferentes predicciones derivadas de cada hipótesis las cuales son contrastadas con las observaciones. Cuando los resultados obtenidos difieren de los predichos por una hipótesis biológica, ésta

✉ Laboratorio Ecotono, CRUB, Universidad del Comahue, (8400), Bariloche, Argentina.
alefarji@crub.uncoma.edu.ar

Recibido: 31 de marzo de 2005; Fin de arbitraje: 1 de julio de 2005; Revisión recibida: 13 de julio de 2005; Segunda revisión recibida: 28 de noviembre de 2005; Aceptado: 11 de febrero de 2006

es normalmente considerada incorrecta y, en consecuencia, descartada como posible causa del fenómeno que deseamos explicar. En este proceso de aprender sobre nuestras ideas incorrectas necesitamos métodos que nos permitan analizar objetivamente el ajuste entre las predicciones y las observaciones. Estos métodos comprenden el campo de la estadística, la cual es una herramienta que nos permite, indirectamente, descartar hipótesis biológicas incorrectas.

La estadística llamada "frecuentista", pese a ser criticada en los últimos años (Johnson 1999, 2002; Anderson et al. 2000; Eberhart 2003, pero ver una resurrección en Stephens et al. 2005), es el método más utilizado en ecología para evaluar la concordancia entre las predicciones biológicas y las observaciones. Brevemente, este método consiste en construir la denominada hipótesis estadística nula (H_0), la cual es básicamente un resultado esperado en ausencia de la causa invocada en la hipótesis biológica, e intentar rechazarla. Para eso, calculamos la probabilidad de que nuestras observaciones (representadas en una fórmula, o "estadístico") ocurran bajo el supuesto que H_0 sea verdadera (o sea, que no exista efecto de "tratamiento"). Si esta probabilidad es muy pequeña, se rechaza la hipótesis estadística nula (ya que es muy poco probable que nuestras observaciones ocurran si la hipótesis estadística nula es verdadera). En general, se considera como una probabilidad pequeña aquella $< 5\%$ (lo cual es arbitrario pero no el objetivo de este debate, ver Stoehr 1999; Scheiner & Gurevich 2001). En consecuencia, obtener valores de $P < 0.05$ genera alegría en los investigadores, porque al rechazar la hipótesis estadística nula, la hipótesis estadística alternativa (H_a) queda por defecto como la única opción válida *por el momento* (Carver 1978). Dado que H_a es una deducción de nuestra hipótesis biológica (Farji-Brener 2004), mantenerla como la opción válida sugiere que la idea que teníamos sobre el funcionamiento sobre un determinado fenómeno es transitoriamente verdadera. Contrariamente, si aquella probabilidad no es demasiado pequeña (usualmente $P > 0.05$, comúnmente llamados resultados estadísticamente no-significativos), genera tristeza en los investiga-

dores. Al no poder rechazar la veracidad de la hipótesis estadística nula, la hipótesis estadística alternativa es normalmente considerada falsa (que no siempre es acertado, ver supuestos más adelante), lo cual nos lleva a inferir que la hipótesis biológica que la generó también lo es (Fig. 1).

Pese al frecuente uso de este tipo de estadística, descartar hipótesis biológicas incorrectas como consecuencia de resultados estadísticamente no-significativos es paradójicamente subvalorado como aporte al proceso de adquirir conocimiento. En consecuencia, los autores y editores comúnmente rechazan enviar o publicar investigaciones con resultados estadísticamente no-significativos (Rosenthal 1979; Csada et al. 1996). Algunos de los problemas potenciales de este sesgo incluyen el rechazo de artículos, retardo en la tasa de publicación, publicación en revistas de menor circulación y meta-análisis inexactos, lo cual genera una representación sesgada del funcionamiento de los sistemas naturales (Csada et al. 1996; Lortie 1999; Palmer 1999; pero ver Koricheva 2003). Sin embargo, estos argumentos han enfatizado la importancia de la escasa divulgación de este tipo de resultados y no su relevancia lógica para comprender la naturaleza.

Para evaluar una hipótesis biológica, es esencial establecer y poner a prueba predicciones sobre eventos que deberían suceder si dicha hipótesis fuese correcta (Bunge 1997; Farji-Brener 2003). Sin embargo, existen varias limitaciones para probar la veracidad de una idea. Déjenme ilustrar este problema con un ejemplo. Para entender porqué hay más plántulas de un determinado árbol debajo que lejos de arbustos, supongan que luego de exprimirme el cerebro formulo dos hipótesis biológicas, que los arbustos: (1) ofrecen protección contra los herbívoros de las plántulas, y (2) funcionan como "nodrizas", disminuyendo temperaturas del suelo letales para las plántulas. El paso siguiente es elaborar una serie de predicciones para cada hipótesis que me permitan discriminar entre estas dos ideas. Luego planifico un buen diseño experimental (incluyendo alguna que otra manipulación con un buen número de réplicas), y llevo a cabo la investigación. Luego de analizar mis datos

Ejemplo

Proceso lógico

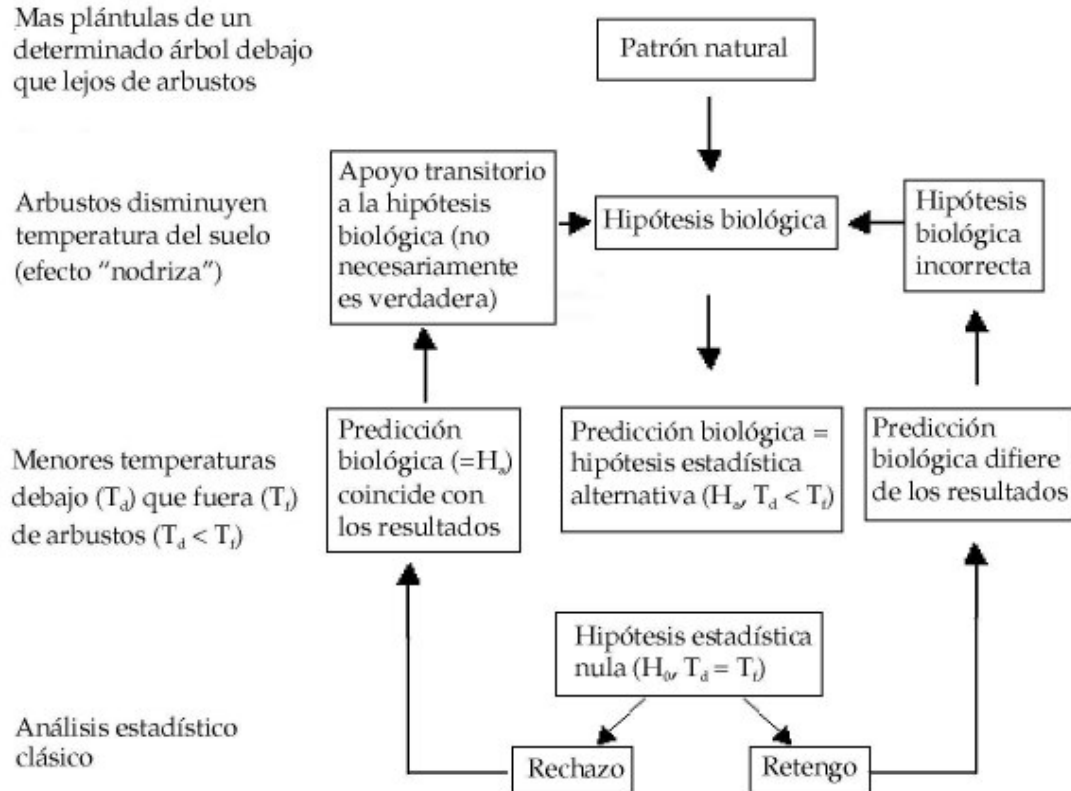


Figura 1. Representación esquemática de los pasos lógicos en el proceso de probar hipótesis y un ejemplo asociado. Solo una hipótesis biológica y una predicción asociada es desarrollada para simplificar el esquema, pero su planteo es similar cuando existen varias predicciones biológicas producto de hipótesis biológicas contrastantes. Nótese que la predicción biológica (i.e., los resultados esperados bajo una hipótesis biológica verdadera) es idéntica a la hipótesis estadística alternativa. Nótese también que bajo este esquema de pensamiento y usando estadística clásica, rechazar una hipótesis biológica incorrecta reteniendo la hipótesis estadística nula (i.e., resultados estadísticamente no-significativos) es el resultado más robusto que puede suceder.

Figure 1. Schematic representation of the logical steps in the process of testing biological hypotheses and an associated example. Only one biological hypothesis and one associated prediction is described to simplify the scheme, but this scheme is similar when different biological predictions are derived from contrasting biological hypotheses. Note that the biological prediction (= expected results if the biological hypothesis is true) is the statistical alternative hypothesis. Note also that under this approach, discarding the biological hypothesis by retaining the statistical null hypothesis (i.e., a statistical non-significant result) is the most powerful result.

usando los métodos estadísticos clásicos de prueba de hipótesis, veo que los resultados apoyan la hipótesis biológica (1) y que están en desacuerdo con las predicciones de la hipótesis biológica (2). En consecuencia, escribo mi manuscrito enfatizando el que considero mi resultado más importante: que la herbivoría

es la presión de selección que mejor explica la alta abundancia de plántulas bajo arbustos. Sin embargo, hay varias razones por las cuales esta afirmación puede no ser del todo verdadera. Primero, dado que soy un ser humano, mi propia imaginación pone límites a las hipótesis biológicas que se me puedan ocurrir para

evaluar las causas de un patrón (Carver 1978). Seguramente hay otros posibles motivos, que no he evaluado, por los cuales las plántulas pueden ser más abundantes debajo que lejos de arbustos (e. g., los arbustos pueden actuar como trampa pasiva de semillas). Segundo, el hecho de probar que una hipótesis biológica es verdadera debería estar basado en todas las observaciones posibles, o asumir que lo que pasa en las circunstancias no observadas puede ser inferido por el muestreo (Underwood 1997). Dadas estas limitaciones, demostrar la veracidad de una hipótesis biológica no parece ser una tarea sencilla, o incluso, posible.

Para resolver este problema, Popper (1959) postuló un enfoque epistemológico en donde no se pretende probar la verdad de una hipótesis, sino su falsedad, lo cual es evidentemente más sencillo. Una vez que una hipótesis biológica es rechazada -si el estudio está bien planificado y correctamente llevado a cabo- no hay necesidad de asumir nada respecto de las observaciones realizadas por el investigador, ni preocuparse por el universo de ideas que no se nos ocurrieron. El "falsificar" hipótesis, o poner a prueba la falsedad en vez de la veracidad de las ideas, es la filosofía sobre la cual se basa la estadística frecuentista. Como expliqué anteriormente, este tipo de estadística presenta dos hipótesis estadísticas contrastantes: H_a , que representa el resultado esperado si nuestra hipótesis biológica es verdadera, y H_0 , que representa un resultado esperado si dicha idea es incorrecta. La estadística clásica intenta rechazar H_0 , no aceptar H_a . El argumento se basa en que rechazar la hipótesis estadística nula -proceso que como discutí recién no presenta los problemas lógicos de aceptar-, deja *transitoriamente* como única opción a H_a , resultado que indirecta y temporalmente apoya nuestra hipótesis biológica (Carver 1978; Farji-Brener 2004, ver Fig. 1). Por el contrario, si H_0 es retenida (i.e., si el P es $> 5\%$, y cuan mayor debe ser es terreno de discusión, no es lo mismo un $P = 0.08$ que un $P = 0.95$), H_a es normalmente considerada falsa, y consecuentemente, la hipótesis biológica de la cual deriva, incorrecta (ver discusiones sobre este tópico en Gotelli & Ellison 2004 y aclaraciones sobre la importancia de una buena potencia estadística más

adelante). Por ejemplo, si no podemos rechazar estadísticamente la H_0 que la temperatura debajo y fuera de los arbustos son iguales, es lógico deducir que la hipótesis biológica de un efecto amortiguador de la temperatura de los arbustos para explicar la mayor abundancia de plántulas debajo de su copa es incorrecta (Fig. 1). En este esquema de pensamiento, nótese que cualquier resultado de las pruebas estadísticas tradicionales implica el rechazo o el no rechazo, pero nunca la aceptación de las hipótesis, sean éstas estadísticas o -indirectamente-, biológicas.

Bajo esta visión de ciencia, el rechazo de ideas incorrectas es uno de los resultados más sólidos que puede generar una buena investigación. En consecuencia, en el ejemplo descrito anteriormente, el mayor mérito de mi trabajo sería el rechazo de la hipótesis de que los arbustos actúan como plantas nodrizas, más que el apoyo a la idea de que actúan como barrera para los herbívoros. Dado las tendencias actuales en la publicación de manuscritos, sospecho que si en vez de discutir largamente la relevancia de la hipótesis (1) destaco la importancia de rechazar mi hipótesis (2), esto enfatizará la probabilidad de que rechacen mi artículo. Resumidamente, pese a que estoy empleando el clásico método estadístico de pruebas de hipótesis, estoy evadiendo su principal fortaleza: su capacidad de rechazar ideas incorrectas. Un caso extremo es cuando se pone a prueba una sola hipótesis biológica y ésta es rechazada. Pese a que los patrones naturales generalmente son producto de varios procesos simultáneos (y por eso es difícil entenderlos poniendo a prueba una sola hipótesis), este tipo de investigación es muy común. Bajo esta circunstancia, el trabajo no apoya ninguna explicación alternativa para el patrón original, y queda la sensación que todo ha sido inútil. Yo disiento. Descartar una idea incorrecta no necesariamente es frustrante si se discute bajo el contexto adecuado. El proceso de descarte simplifica la comprensión de un problema porque nos acerca un poco más a las ideas correctas.

Al desarrollar una investigación, todos -consciente o inconscientemente-, adherimos a una filosofía de cómo funciona la naturaleza.

La estadística también. Si usted utiliza pruebas de *t*, análisis de varianza, correlaciones, regresiones u otras pruebas de la estadística frecuentista, debería saber que las mismas se basan en la filosofía bajo la cual rechazar ideas es uno de los mecanismos más robustos para comprender la naturaleza. En consecuencia, yo le sugiero que no eluda enviar buenos trabajos basados en resultados estadísticamente no-significativos a revistas de alto impacto. Y cuando le toque el papel de revisor, le sugiero que recuerde que rechazar ideas incorrectas –no rechazar artículos por la presencia de resultados no-significativos–, es una de las prácticas más poderosas para aprender de la naturaleza cuando se usa la estadística clásica. Finalmente, mi sugerencia para ambos, autores y revisores, es recordar que la calidad de un manuscrito es completamente independiente de sus resultados estadísticos. La presencia de resultados estadísticamente no-significativos no es sinónimo de una mala investigación, de la misma forma que la abundancia de resultados estadísticamente significativos no indica que el trabajo sea bueno. En consecuencia, bajo ninguna circunstancia la significancia estadística debería ser un criterio para evaluar la calidad de una investigación o de un manuscrito.

Es muy importante enfatizar tres cosas a esta altura del debate. Primero, los resultados estadísticamente no-significativos pueden ser consecuencia de una hipótesis biológica incorrecta, de un pobre diseño experimental, o de la falta de poder estadístico (i.e., la falta de fuerza en detectar un efecto cuando éste realmente existe). Es posible que una hipótesis biológica sea temporalmente correcta pero los resultados sean estadísticamente no-significativos debido a una alta probabilidad de cometer un error de tipo II. Esta falta de poder estadístico puede hacer que retengamos incorrectamente la hipótesis estadística nula, y consecuentemente arrastrarnos a conclusiones biológicas erradas. No es mi intención discutir en este debate la importancia (y el escaso uso) de las evaluaciones del poder estadístico al usar la estadística clásica. Este debate se basa en suponer que los resultados de una investigación provienen de un sólido diseño experimental con un alto poder estadís-

tico, y en consecuencia que el rechazo de la hipótesis biológica solo responde a su falsedad. Segundo, es obvio que el proceso de aprender de la naturaleza se sostiene en el aprendizaje tanto de los errores como de los aciertos. Ambas evidencias, a favor o en contra de una idea, son importantes para la construcción del conocimiento. Sin embargo, este artículo pretende enfatizar que valorar y aprender de los errores (i.e., de los resultados estadísticamente no-significativos que descartan una hipótesis biológica) es una herramienta subvalorada pero fundamental cuando utilizamos la estadística tradicional. Finalmente, el espíritu de este debate no es discutir las ventajas o debilidades de falsificar hipótesis ni del uso de la estadística clásica versus otras aproximaciones filosóficas (e.g., Lakatos 1978; Mayo 1996), u otros métodos estadísticos (e. g., estadística bayesiana, ver Ellison 2004, teoría de la información, ver Burnham & Anderson 2002). Estos tópicos ya han sido muy bien discutidos en otros trabajos (Hilborn & Mangel 1997; Ellison 2004; Stephens et al. 2005). Evidentemente, no hay una única filosofía ni una sola metodología estadística para entender cómo funciona la naturaleza (McIntosh 1987; Mayo 1996; Weber 1999; Brown 2001; Stephens et al. 2005). Más aun, la imaginación y el sentido común pueden ser muchas veces más importantes para el avance del conocimiento que la afiliación a una filosofía o el uso inflexible de una sola metodología (Oksanen 2004; Stephens et al. 2005). Sin embargo, debo enfatizar que es necesario conocer las bases filosóficas de las herramientas estadísticas que solemos emplear, para minimizar sus debilidades y aprovechar sus fortalezas. En consecuencia, si usted está desarrollando o evaluando trabajos en donde se utiliza estadística clásica, recuerde que la fortaleza de esta metodología está basada principalmente en la eliminación de ideas incorrectas más que en la aceptación de ideas potencialmente verdaderas.

El progreso en ecología puede ser considerado como la expansión de nuestra comprensión acerca de cómo funcionan los sistemas naturales, y el rechazo de ideas incorrectas es un factor determinante en este proceso de

aprendizaje (Graham & Dayton 2002). Consecuentemente, uno de los mayores logros de un trabajo científico no es ratificar una idea preconcebida sino contradecirla, ya que encontrar evidencias sólidas en contra de ideas es lo que mantiene la evolución del conocimiento en un estado dinámico (Brown 2001; Graham & Dayton 2002; Hulbert 2004). Bajo esta perspectiva, un resultado estadísticamente no-significativo, producto de un trabajo científico de alta calidad, debería ser considerado como un aporte significativo para comprender mejor cómo funciona la naturaleza.

AGRADECIMIENTOS

A W. Eberhard, B. Wcislo, D. Vázquez y a solo uno de los dos revisores, quienes con sus comentarios mejoraron la calidad de este manuscrito.

BIBLIOGRAFÍA

- ANDERSON, DR; KP BURNHAM & WL THOMPSON. 2000. Null hypothesis testing: problems, prevalence, and an alternative. *J. Wildl. Manag.* **64**:912-923.
- BURNHAM, KP & DR ANDERSON. 2002. *Model Selection and Multimodel inference: A Practical Information-Theoretic Approach*. Springer-Verlag, N. Y.
- BROWN, J. 2001. Ngongas in ecology: on having a worldview. *Oikos* **94**:6-16.
- BUNGE, M. 1997. *La Ciencia, su método y su filosofía*. Segunda edición. Editorial Panamericana. Bogotá, Colombia.
- CARVER, RP. 1978. The case against statistical significance testing. *Harvard Educational Review* **48**:378-399.
- CSADA, R; P JAMES & R ESPIE. 1996. The "file drawer problem" of non-significant results: does it apply to biological research? *Oikos* **76**:591-593.
- EBERHART, LL. 2003. What should we do about hypothesis testing? *J. Wildl. Manag.* **67**:241-247.
- ELLISON, A. 2004. Bayesian inference in ecology. *Ecology Letters* **7**:509-520.
- FARJI-BRENER, AG. 2003. Uso correcto, parcial e incorrecto de los términos "hipótesis" y "predicciones" en ecología. *Ecología Austral* **13**:223-227.
- FARJI-BRENER, AG. 2004. ¿Son hipótesis las hipótesis estadísticas? *Ecología Austral* **14**:201-203.
- GOTELLI, N & A ELLISON. 2004. *A Primer of Ecological Statistics*. Sinauer Assoc, Inc. Mass, USA.
- GRAHAM, M & P DAYTON. 2002. On the evolution of Ecological ideas: paradigms and scientific progress. *Ecology* **83**:1481-1489.
- HILBORN, R & M MANGEL. 1997. *The Ecological Detective: confronting models with data*. Princeton University Press, New Jersey.
- HULBERT, S. 2004. On misinterpretation of pseudoreplication and related matters: a reply to Oksanen. *Oikos* **104**:591-597.
- JOHNSON, D. 1999. The insignificance of statistical significance testing. *J. Wildl. Manag.* **63**:763-772.
- JOHNSON, D. 2002. The role of hypothesis testing in wildlife science. *J. Wildl. Manag.* **66**:272-276.
- KORICHEVA, J. 2003. Non-significant results in ecology: a burden or a blessing in disguise? *Oikos* **102**:397-401.
- LAKATOS, I. 1978. *The Methodology of Scientific Research Programmes*. Cambridge University Press, New York.
- LORTIE, J. 1999. Over-interpretation: avoiding the stigma of non-significant results. *Oikos* **87**:183-184.
- MAYO, D. 1996. *Error and the Growth of Experimental Knowledge*. University of Chicago Press, Chicago.
- MCINTOSH, R. 1987. Pluralism in ecology. *Ann. Rev. Ecol. Syst.* **18**:321-341.
- OKSANEN, L. 2004. The devil lies in details: reply to Stuart Hulbert. *Oikos* **104**:598-605.
- PALMER, AR. 1999. Detecting publication bias in meta-analyses: a case study of fluctuating asymmetry and sexual selection. *Am. Nat.* **154**:220-233.
- POPPER, K. 1959. *The logic of Scientific Discovery*. Hutchinson, London.
- ROSENTHAL, R. 1979. The "file drawer problem" and tolerance for null results. *Psychol. Bull.* **86**:638-641.
- SCHEINER, S & J GUREVITCH. 2001. *Design and Analysis of Ecological Experiments*. Second edition. Oxford University Press, New York.
- STEPHENS, PA; SW BUSKIRK; GD HAYWARD & C MARTÍNEZ DEL RÍO. 2005. Information theory and hypothesis testing: a call for pluralism. *J. Appl. Ecol.* **42**:4-12.
- STOEHR, A. 1999. Are significance thresholds appropriate for the study of animal behavior? *Animal Behavior* **57**:F22-F25.
- UNDERWOOD, A. 1997. *Experiments in Ecology: their logical design and interpretation using analysis of variance*. Cambridge University Press, New York.
- WEBER, T. 1999. A plea for diversity of scientific styles in ecology. *Oikos* **84**:526-529.