Statistics in ecology is for making a “principled” argument

Song S. Qian

Received: 29 April 2014 / Accepted: 6 May 2014 / Published online: 16 May 2014
© Springer Science+Business Media Dordrecht 2014

In this guest editorial, I discuss the controversy between Fisher and Neyman-Pearson in the context of statistical inference in ecology and environmental sciences. I argue that statistical evidence represented in p -values and AIC is secondary to scientific evidence supporting the underlying statistical model when analyzing observational data. Accordingly, I recommend that ecologists emphasize model justification and model interpretation over the dichotomy of hypothesis testing.

The recent *Ecology* forum on the use of p -values and AIC (Ellison et al. 2014) and the revised significance level of 0.005 by Johnson (2013) renewed the discussion of the role of statistics in ecological research. However, they did not mention the controversy between R.A. Fisher and J. Neyman and E. Pearson over the role of statistics in scientific research, intensified after the publication of Neyman (1935). A revisit of the underlying philosophical issue of the controversy can help us better understand issues related to the use of p -values and AIC in ecological and environmental sciences.

I argue that statistics should be used by ecologists to develop a “principled argument” (Abelson 1995), in which the role of the p -value or other statistical concepts should be determined by the nature of the ecological problems, not the mathematical characteristics of these statistics.

I summarize the *Ecology* forum, the revised significance level, and the Fisher–Neyman–Pearson controversy in an online blog¹. Briefly, Fisher views statistics as a tool for inductive reasoning through hypothetical deduction. Under Fisher’s line of thinking, a model is hypothetical and serves as a mediator between mathematics and the real world. Comparisons of model predictions and observations can expose weaknesses of a model and provide information for model improvement. Neyman and Pearson view the

S. S. Qian (✉)
Department of Environmental Sciences, The University of
Toledo, Toledo, OH 43606, USA
e-mail: song.qian@utoledo.edu

¹ <http://songqiansblog.blogspot.com/2014/04/statistics-is-more-than-p-values-and-aic.html>.

role of statistics as a tool for deduction. In their framework, a model is an integral part of the inferential structure that is not subject to change. New data can lead to different conclusions, but not different models.

Statistical evidence versus scientific evidence

Much of the debate in the *Ecology* forum and in Johnson (2013) is about the use of statistical evidence for making a decision. The statistical evidence is summarized in the p -value and AIC and the decision is often about whether or not some coefficients should be 0, thereby removing the variables associated with these coefficients from further consideration. The *Ecology* forum is focused on model selection using either the p -value or AIC, and Johnson (2013) is focused on an quantitative interpretation of the p -value as evidence in a hypothesis testing situation.

As mentioned above, the statistical model in the Neyman-Pearson framework is fixed and unchanging. Statistical evidence is about model coefficients. In Fisher's framework, we need not only statistical evidence about model coefficients, but also evidence about the validity of the model. Early applications of statistics often focus on the analysis of experimental data. This emphasis can be largely attributed to Fisher's work in developing analysis of variance (ANOVA) models for analyzing data from agricultural experiments. ANOVA models are based on a set of assumptions about the data. By conducting experiments in certain ways, we can ensure that the resulting data meet most assumptions of the model. The main task of data analysis is then to quantify the model's coefficients, which are used for further inferences (e.g., testing whether means from various treatments are different from the control mean).

But focusing only on experimental data is a shortcut of the statistical analysis process defined by Fisher (1922), which divides statistical analyses into three groups of problems—problems of specification, problems of estimation, and problems of distribution.

Problems of specification represent a step of formulating a model parameterized with unknown coefficients. The problems of estimation represent a step of estimating model coefficients from observed data. I interpret the problems of distribution as a step of model evaluation. In Fisher's words, once we

selected a model, "the adequacy of our choice may be tested a posteriori."

In analyzing agricultural experimental data, we can arrange experiments such that the underlying model is known. The focus is then the problem of estimation and the subsequent inference, which is also the focus of Neyman-Pearson. The Neyman-Pearson approach reformulates a statistical inference problem as a confrontation of two hypotheses. Under this framework, the decision (accepting one of the two hypotheses) is a mathematical process of evaluating the relative risks. They introduced the concepts of significance level and power.

Statistics such as p -value and AIC are used under the precondition of a known underlying model. As such, statistical evidence is defined by the likelihood function or likelihood ratio in the case of model selection, whether one uses AIC or p -value is a matter of style.

Questions facing landscape ecologists usually have to be addressed at scales and complexity beyond those of typical controlled experiments. As a result, problems of model specification and evaluation are prominent. Statistical evidence alone is not sufficient. Statistical evidence must be used in conjunction with subjective matter evidence.

Implications to ecologists

When applying statistics to observational data, where we are uncertain about the underlying model, we face conceptual difficulties arising from the need of validating the proposed model, as well as making statistical inference about model coefficients. Much of the conceptual and mathematical difficulties discussed in the *Ecology* forum can be traced to the model validation problem. In other words, we should use p -value or AIC as one piece of information, not the only information in analyzing observational data. Other information should also be sought. Statistical analysis can help us in searching for other relevant evidence, perhaps using the MAGIC criteria of Abelson (1995): a statistical inference should be a principled argument and the strength of the inference should be measured by Magnitude, Articulation, Generality, Interestingness, and Credibility, not just a p -value or AIC or any other single statistic.

When observational data are the primary source of information, statistical evidence measured by p -values or AIC is secondary to evidence supporting the statistical model itself. When applying a statistical model, we need to understand the underlying assumptions and in what ways these assumptions may be violated. When a model is selected and quantified, we should always compare the model predicted to the observed using various model checking methods. Furthermore, we should always be ready to make changes to our models when potential weaknesses are exposed. This process can be facilitated by adding the following steps to a typical statistical analysis process:

- Verbal description of a model—a clear description of the model using non-statistical terms should be a first step. When describing the model in clear scientific terms, we can better judge whether the model is sensible and whether the real world can be reasonably represented by the model. Even for a simple model such as a t -test or ANOVA, a verbal description can be helpful.
- Verifying model assumptions—plots, plots, and more plots (Cleveland 1993).
- Verbal description of estimated model coefficients—before finalizing the model, we should describe the estimated model coefficients in words.

This should be done even in a simple two sample t -test.

With these considerations in mind, we should carefully balance all sources of information when preparing our manuscripts. When reviewing our colleagues work, we should look beyond p -values, AIC, and R^2 . In addition, we should consider making changes in our statistics courses, so that students can learn about the importance of a balanced approach.

Acknowledgments Jianguo Wu encouraged me to pursue this project and provided constructive comments on an earlier draft. I thank Daryl Moorhead, Johan Gottgens, and Craig Stow for their helpful comments and discussions.

References

- Abelson RP (1995) Statistics as principled argument. Psychology Press, New York
- Cleveland WS (1993) Visualizing data. Hobart Press, Summit
- Ellison AM, Gotelli NJ, Inouye BD, Strong DR (2014) P values, hypothesis testing, and model selection: it's déjà vu all over again. *Ecology* 95(3):609–610
- Fisher RA (1922) On the mathematical foundations of theoretical statistics. *Philos Trans R Soc Lond Ser A* 222:309–368
- Johnson VE (2013) Revised standards for statistical evidence. *Proc Natl Acad Sci* 110(48):19313–19317
- Neyman J (1935) Statistical problems in agricultural experimentation. *J R Stat Soc* 2(Supplement):107–180