

INTERPRETING FAILURE TO REJECT A NULL HYPOTHESIS

Like many members of the Ecological Society, I review manuscripts for *Ecology* from time to time. As a teacher of biostatistics courses, I pay special attention to applications of statistics as I do so, and because of a personal interest in how scientists *interpret* statistical results, that is what I attend to most. Uncomfortably often, I review papers in which the statistical analyses seem to be misinterpreted. The most recent paper sent to me was one of these, and it has inspired this note. The points I will make seem obvious to me, but there may be ecologists who disagree with them. If so, I hope they will respond in this *Bulletin*, since the issues are important ones.

My argument is with the interpretation, in hypothesis testing, that failure to reject the null hypothesis is proof of the null hypothesis. Failing to reject a null hypothesis is distinctly different from proving a null hypothesis; the difference in these interpretations is not merely a semantic point. Rather, the two interpretations can lead to quite different biological conclusions, as will be seen below.

Consider the following example, which I have made up to illustrate several points. The example is logically similar to the statistical analysis in the manuscript referred to above, but the biological content has been changed to prevent identification of that paper.

Table 1. Masses and areas for seeds of two species of maple. Sample size is 10 for each species.

	Mass (mg)	Area (cm ²)
Species A		
Mean	295	1.8
Variance	4160	4.25
Species B		
Mean	333	2.8
Variance	2080	3.27
Null hypothesis, H_0	$\mu_A = \mu_B$	$\mu_A = \mu_B$
$P(\text{data} H_0)$.04	.12
Statistical significance	*	NS†

* Difference detectable at the $P = .05$ level.

† NS = not significant.

A Synthetic Example

Suppose we are interested in the transport by wind of seeds from two species, A and B, of maple trees. We collect one seed from each of 10 trees of each species, with the trees and seeds being chosen with careful attention to randomization. We measure the mass and projected surface area of each seed, and obtain the results summarized in Table 1. The probability values and statistical significance levels result from t tests applied to each column of data.

Discussion

The manuscript that elicited this note did not present the data nor the details of any statistical tests, and it was based on a different set of biological materials. But it did contain a statement that was logically equivalent to the following interpretation of the tests above (the italics indicate direct quotation; the rest is paraphrased to fit the example):

"Since measurements of seed areas showed no statistical difference between species (data not presented), any difference in the distance their seeds may be transported by the wind must be due entirely to weight effects."

Such a biological interpretation of statistical results like these is obviously unwarranted. Consider the following problems with that interpretation.

First, as is recognized in almost all statistics texts, there is little or no relationship between the biological importance and the statistical significance of a given result. For example, in the constructed maple seed data, the mean masses differed by only 13%, whereas the mean areas differed by 56%. The masses were statistically detectable as different because of the within-species similarity, but the areas were not because of their great variability. Yet the large difference in average areas would probably be of greater ecological importance than the smaller difference in masses.

Table 2. Decision rules for further study of a phenomenon, based on one hypothesis test.

Bio- logi- cal impor- tance	Statistical significance	
	Not significant	Significant
Low	Go on to other questions	Ask, "Does it matter?"
High	Conduct a better experiment	Go on to further studies of this phenomenon

Second, failure to reject a null hypothesis can (and very often does) result as easily from an inadequate experiment as from lack of a large or important effect. Such inadequacy can result from a combination of small sample size, careless control of extraneous factors, and measurement error, as well as from intrinsic variability in the phenomenon under study.

Third, regardless of the results of any hypothesis tests, the sample means for the masses and areas of the two species are the best estimates available from the data for the true population values of those quantities. In fact, parameter estimation with accompanying confidence limits is often a more logical way to interpret data than is hypothesis testing.

Because of their apparent biological importance, the area data in Table 1 suggest the need to obtain a larger sample, if one feels compelled to test hypotheses. Indeed, one might work according to a set of decision rules like those shown in Table 2. Of course, deciding on the degree of biological importance requires subjective scientific judgment, which some workers would rather not face.

The main point is that the above data certainly do not prove the mean seed areas

of populations A and B to be identical, and indeed, no statistical test could possibly prove such an assertion. In other words, *failure to disprove a null hypothesis does not prove that null hypothesis*. Put another way, one should not think of "accepting the null hypothesis," but rather of failure to reject it. The greater accuracy of the latter interpretation far outweighs the fact that it is a more cumbersome phrase.

Unfortunately, the notion of accepting the null hypothesis is suggested by at least one important textbook (Sokal and Rohlf 1981, e.g., pages 172, 190, and 224), and the idea seems to be widely held. (For example, I frequently find graduate students who learned that notion as undergraduates.) In the interest of scientific progress, it is time to reject this interpretation of hypothesis-test results.

Acknowledgments

I thank G. Suter, J. Lauer, H. Shugart, and L. Barnhouse for helpful comments on the manuscript.

David F. Parkhurst
Sabbatical address:
Environmental Sciences Division
Oak Ridge National Laboratory¹
Oak Ridge, TN 37831

Permanent address:
School of Public and Environmental
Affairs, and Biology Department
Indiana University
Bloomington, IN 47405

Literature Cited

Sokal, R. R., and F. J. Rohlf. 1981. *Biometry*. Second edition. W. H. Freeman, San Francisco, California, USA.

¹ Operated by Martin Marietta Energy Systems, Incorporated, under Contract Number DE-AC05-84OR21400 with the United States Department of Energy.