

To Print: Click your browser's PRINT button.

NOTE: To view the article with Web enhancements, go to:

<http://www.medscape.com/viewarticle/524206>

Stats for the Health Professional

To P or Not to P: Why Use a P Value, Anyway?

Andrew J. Vickers, PhD

Medscape Business of Medicine. 2006;7(1) ©2006 Medscape

Posted 03/02/2006

My nonstatistical colleagues seem to imagine that I spend my off-hours calculating the mean time it takes to grill a piece of fish, or an exact binomial confidence interval for the proportion of Saturdays that I get to lie in. In truth, of course, I spend my off-hours trying not to think of statistics at all. But let's indulge the fantasy of the 24/7 statistician.

Going home each night, I have a choice between cycling down a busy road or winding through the beautiful backstreets of Brooklyn. Being statistically obsessed, I have recorded how long each route takes me on a number of occasions and have calculated means and standard deviations. Imagine that, one day, I had to get home as soon as possible for an appointment. To choose a route, I conduct a statistical analysis of my travel time data: it turns out that the travel time for the busy road is shorter, but the difference between routes is not statistically significant ($P = .4$). Nonetheless, it would still seem sensible to take what is likely to be the quicker route home, even though I haven't proved that it will get me there fastest.

Now let's imagine that this incident got me fired up, and I spend 2 years randomly selecting a route home and recording times. When I finally analyze the data, I find strong evidence that going home via the busy road is faster ($P = .0001$), but not by much (it saves me 57.3 seconds on average). So I decide that, unless I am in a real rush, I'll wind along the backstreets, simply because it is a more pleasant journey and well worth the extra minute.

We tend to think that P values should determine our actions; in the case of a drug clinical trial, for example, we say: " $P < .05$: use the drug; $P \geq .05$: don't use the drug." Yet, the bicycle example shows the opposite: I chose the busy road when P was .4 but not when P was .0001. This suggests we need to think a little harder about what P values are and how we should use them.

The most important thing to remember about P values is that they are used to test hypotheses. This sounds obvious, but it is all too easily forgotten. A good example is the widespread practice of citing P values for baseline differences between groups in a randomized trial. The hypothesis being tested here is whether there are real differences between groups. Yet we know that groups were randomly selected, so any differences in characteristics such as age or sex must be due to chance alone.

Science is often said to be about testing hypotheses, but in many cases this is not what we want to do at all. When I had to get home in a rush, I wasn't interested in proving which was the quickest way home, I just needed to work out what route was likely to get me to my appointment on time. Moreover, even when we do want to test hypotheses, our conclusion is a necessary but not sufficient guide to action. I eventually proved that using the busy road was quickest but decided to choose a different route on the basis of considerations -- pleasure and quality of life -- that formed no part of the hypothesis test.

An even more difficult problem is when our P value is $> .05$, that is, when we have failed to prove our hypothesis. This is often interpreted as proof that our hypothesis is false. Such an interpretation is not only incorrect, but it can also be dangerous; I'll discuss this in a future column.

Andrew J. Vickers, PhD, Assistant Attending Research Methodologist, Memorial Sloan-Kettering Cancer Center, New York, NY

Disclosure: Andrew J. Vickers, PhD, has disclosed no relevant financial relationships.
