

The Biggest Myth About P-Values

By [Ross Pomeroy](#)

November 06, 2016

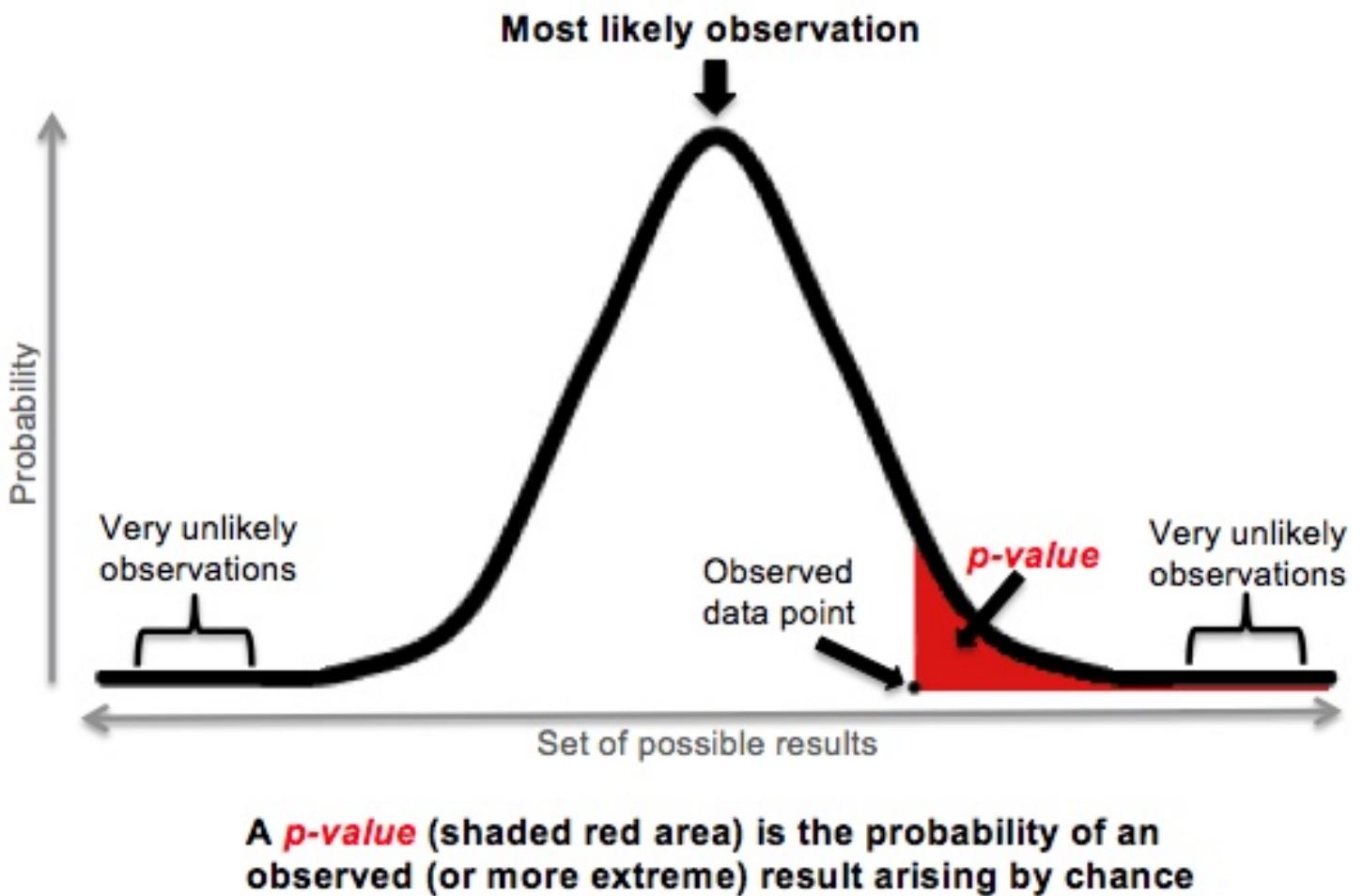
Ninety years ago, Ronald Fisher changed science forever. With his book, [Statistical Methods for Researchers](#), the eminent English statistician popularized P values for measuring statistical significance in a scientific result, noting, almost as an afterthought, "Personally, the writer prefers to set a low standard of significance at the 5 percent point..." **P = 0.05** was born, and, as computer scientist Robert Matthews [critiqued](#) years later, scientists were bestowed with a "mathematical machine for turning baloney into breakthroughs, and flukes into funding."

Researchers in a great many disciplines now operate on Fisher's personal recommendation for significance. If a single finding attains a P value of 0.05 or lower, it's published as a noteworthy discovery and a scientific "truth". But that is not actually what Fisher intended. "A scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give this level of significance," he wrote.

"In other words, the operational meaning of a P value less than .05 was merely that one should repeat the experiment," Johns Hopkins biostatistician Steven Goodman [interpreted](#). "If subsequent studies also yielded significant P values, one could conclude that the observed effects were unlikely to be the result of chance alone. So 'significance' is merely that: worthy of attention in the form of meriting more experimentation, but not proof in itself."

Yet because $P = 0.05$ has become so deeply ingrained and unquestioningly accepted, the scientific literature is now full of flimsy findings, particularly in the fields of [psychology](#), [medicine](#), and [epidemiology](#). Quite simply, a large amount of published research is false.

The largest reason for this sorry state of affairs is a pervasive myth about the P value. Specifically, that it corresponds to the probability that the null hypothesis -- the idea that there is no significant difference between specified populations or study groups -- is true. For example, common thinking suggests that a P value of 0.05 equates to a five percent chance that the null hypothesis is true and thus a 95% chance that the claim is correct. This is false.



"This is, without a doubt, the most pervasive and pernicious of the many misconceptions about the P value," Goodman says. "It perpetuates the false idea that the data alone can tell us how likely we are to be right or wrong in our conclusions." It's like flipping a coin four times, observing four heads (a probability of .125) and concluding that the likelihood of the coin being fair is just 12.5%, he adds. But that's simply not a conclusion that can be drawn from the data.

In actuality, as David Colquhoun **demonstrated** in 2014, the arbitrary level of significance of 0.05 actually produces incorrect results at least 30 percent of the time! He recommends raising the standards for statistical significance. Only results with a P value lower than 0.001 should be labeled discoveries.

Other scientists would prefer to ditch P values altogether. Last year, a psychology journal **banned** their use. The vacancy leaves room for statisticians to come up with new methods for gleaning scientific results. Odds are, they'll come up with something, but getting scientists to adopt new methods will be a much taller task.

(Image: **Cochrane**)