

The significance of the p-value in 2017

Nick Waters

Hot Topic: September 26, 2017

National University of Ireland, Galway
James Hutton Institute, Dundee

The problem with statistics in research

- Lack of reproducibility across fields
- Irresponsible handling of results
 - p-hacking
 - ignoring multiple comparisons
- Publication bias
- Lack of power



July22	•	Redefining Statistical Significance [Benjamin et al].
July24	•	Eliminate The P-Value (and Bayes Factor) Altogether [Briggs].
Sept21	•	Abandon Statistical significance [McShane, et al]

A refresher



NHST: Null hypothesis significance testing A p-value is the odds of seeing your result by accident if there is not effect.

- $p < 0.05$: the null hypothesis might be able to be rejected
- $p > 0.05$: the null hypothesis doesn't deserve rejection

A p-value is not:

- The likelihood of seeing your results again
- The confidence you should have in your results
- Whether your hypothesis is actually true



Bayes factor is the ratio of likelihoods of two models

- Generally more transparent
- Subjective
- Can be difficult to implement

The Minor Problems with P values



p-value - Wikipedia

<https://en.wikipedia.org/wiki/P-value>

In statistical hypothesis testing, the **p-value** or probability value is the probability for a given statistical model that, when the null hypothesis is true, the statistical ...

Null hypothesis · Misunderstandings of p-values · Statistical significance

P Values (Calculated Probability) and Hypothesis Testing - StatsDirect

www.statsdirect.com/help/basics/p_values.htm

P Values. The **P value**, or calculated probability, is the probability of finding the observed, or more extreme, results when the null hypothesis (H_0) of a study question is true – the definition of 'extreme' depends on how the hypothesis is being tested.

What a p-Value Tells You about Statistical Data - dummies

www.dummies.com/education/math/.../what-a-p-value-tells-you-about-statistical-data/

A small **p-value** (typically ≤ 0.05) indicates strong evidence against the null hypothesis, so you reject the null hypothesis. A large **p-value** (> 0.05) indicates weak evidence against the null hypothesis, so you fail to reject the null hypothesis.

3.2 - Hypothesis Testing (P-value approach) | Statistics

<https://onlinecourses.science.psu.edu/statprogram/node/138>

The **P-value** approach involves determining "likely" or "unlikely" by determining the probability — assuming the null hypothesis were true — of observing a more ...

How to Correctly Interpret P Values - Minitab Blog

blog.minitab.com/blog/adventures-in-statistics-2/how-to-correctly-interpret-p-values

Apr 17, 2014 - The **P value** is used all over statistics, from t-tests to regression analysis. Everyone knows that you use **P values** to determine statistical ...

Understanding Hypothesis Tests: Significance Levels (Alpha) and P ...

blog.minitab.com/.../understanding-hypothesis-tests%3A-significance-levels-alpha-an-

Mar 19, 2015 - What do significance levels and **P values** mean in hypothesis tests? What is statistical significance anyway? In this post, I'll continue to focus on ...

P-Value - Investopedia

www.investopedia.com/terms/p/p-value.asp

The level of marginal significance within a statistical hypothesis test, representing the probability of the occurrence of a given event. The **p-value** is used as an ...

The Major Problems with P values

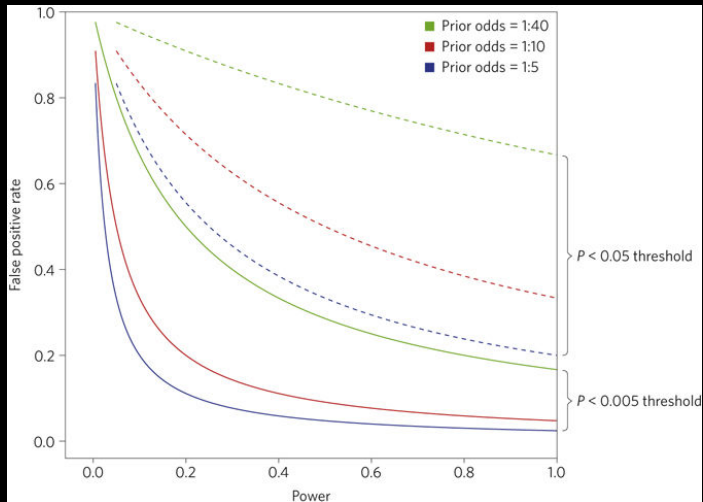


A wee p-value means only one thing: the probability of seeing an ad hoc statistic larger than the one you did see is small given a model you do not believe. This number is as near to useless as any number ever invented, for it tells you nothing about the model you don't believe, nor does it even whisper anything about the model you do believe.

(July 22) Solution 1: $p < 0.005$



- Correlates to a Bayes factor of between 14 and 26 (strong evidence)
- Allows researchers to be more confident with less power
- reduces the false positive rate from 30% to 5%
- Easily adoptable



(July 24) Solution 2: Quantify model likelihood



Anybody can check [the model likelihood]'s predictions, even if they do not know [the data] or [the model]'s details. Given [the model and the data], authors might claim there is a 55% chance Y is true under the new protocol. Any reader can verify whether this prediction is useful for him or not, whether the predictions are calibrated, etc.

Solution 3: Abandon Statistical Significance



- Any thresholds will be circumvented by bad practice
- Higher thresholds don't mean more true
- Higher thresholds breed overconfidence
- NHST promote pairwise comparisons instead of total data
- Statistical tests do not remove personal bias



Proposal:

- Limit p-values to screening
- Include more evidence than just p-values
- Judge the research based on all results