

Statistics By Jim

Making statistics intuitive

[Basics](#)[Hypothesis Testing](#)[Regression](#)[ANOVA](#)[Fun](#)[Glossary](#)[Blog](#)[My Store](#)

Why Are P Values Misinterpreted So Frequently?

By [Jim Frost](#) — [2 Comments](#)

P values are commonly misinterpreted. It's a very slippery concept that requires a lot of background knowledge to understand. Not surprisingly, I've received many questions about P values in statistical hypothesis testing over the years. However, one question stands out. *Why* are P value misinterpretations so prevalent? I answer that question in this blog post, and help you avoid making the same mistakes.

The Correct Way to Interpret P Values

First, I need to be sure that we're all on the right page when it comes to interpreting P values. If we're not, the rest of this blog post won't make sense!

P values are the probability of observing a sample statistic that is at least as different from the null hypothesis as your sample statistic when you assume that the null hypothesis is true. That's a pretty convoluted but technically correct definition—and I'll come back to it later on!

In a nutshell, P value calculations assume that the null hypothesis is true and use that assumption to determine the likelihood of obtaining your observed sample data. P values answer the question, "Are your sample data unusual if the null hypothesis is true?"



Here's a quick way to tell if you are misinterpreting P values in hypothesis tests. If you interpret P values as the probability that the null hypothesis is true or the probability that rejecting the null hypothesis is a mistake, those are *incorrect* interpretations. In fact, these are the most common misinterpretations of P values that I'm addressing specifically in this post. Why are they so common?

Historical Events Made P Values Confusing

The problem of misinterpreting P values has existed for nearly a century. The origins go back to two rival camps in the early days of hypothesis testing. On one side, we have Ronald Fisher with his measures of evidence approach (P values). And, on the other side, we have Jerzy Neyman and Egon Pearson with their error rate method (alpha). Fisher believed that you could use sample data to learn about a population. However, Neyman and Pearson thought that you couldn't learn from individual studies but only a long series of hypothesis tests.

Textbook publishers and statistics courses have squished together these two incompatible approaches. Today, the familiar hypothesis testing procedure of comparing P values to alphas seems to fit together perfectly. However, they're based on irreconcilable methods.

Much can be said about this forced merger. For the topic of this blog post, an important outcome is that P values became associated with the Type I error rate, which is incorrect. A P value is NOT an error rate, but alpha IS an error rate. By directly comparing the two values in a hypothesis test, it's easy to think they're both error rates. This misconception leads to the most common misinterpretations of P values.

Fisher spent decades of his life trying to clarify the misunderstanding but to no effect. If you want to read more about the unification of the two schools of thought, I highly recommend this [article](#).

P Values Don't Provide the Answers that We Really Want

Let's be honest. The common misinterpretations are what we really want to learn from hypothesis testing. We'd love to learn the probability that a hypothesis is correct. That *would* be nice. Unfortunately, hypothesis testing doesn't provide that type of information. Instead, we obtain the likelihood of our *observation*. How likely is our sample if the null hypothesis is true? That's just not as helpful.

Think about this logically. Hypothesis tests use data from one sample exclusively. There is no outside reference to anything else in the world. You can't use a single sample to determine whether it represents the population. There's no basis to draw conclusions like that. Consequently, it's not possible to evaluate the probabilities associated with hypotheses. To do that, we'd need a larger perspective than a single sample can provide. As an aside, Bayesian statistics attempt to construct this broader framework of probabilities.

P values can't provide answers to what we really want to know. However, there is a persistent temptation to interpret them in this manner anyway. Always remember, if you start to think of P values as the probability of a hypothesis, you're barking up the wrong tree!

P Values Have a Torturous Definition

As I showed earlier, the correct definition of P values is pretty convoluted. It is the probability of observing the data that you actually did observe in the hypothetical context that the null hypothesis is true. Huh? And, there is weird wording about being at least as extreme as your observation. It's just not intuitive. In fact, it takes a lot of studying to truly understand it all.

Unfortunately, the incorrect interpretations sound so much simpler than the correct interpretation. There is no straightforward *and* accurate definition of P values that can compete against the simpler sounding misinterpretations. In fact, **not even scientists can explain P values!** And, so the errors continue.

Is Misinterpreting P values Really a Problem?

Let's recap. Historical circumstances have linked P values and the Type I error rate incorrectly. We have a natural inclination to want P values to tell us more than they are able. There is no technically correct definition of P value that rolls off the tongue. There's nothing available to thwart the simpler and more tempting misinterpretation. It's no surprise that this problem persists! Even Fisher couldn't fix it!

You might be asking, "Is this really a problem, or is it just semantics?" Make no mistake; the correct and incorrect interpretations are very different. If you believe that a P value of 0.04 indicates that there is only a 4% chance that the null hypothesis is correct, you're in for a big surprise! It's often around 26%!

To better understand the correct interpretation of P values, I've written three blog posts that focus on interpreting and using P values. The first post describes the correct and incorrect ways to interpret P values. It goes into detail about the substantial problems associated with the incorrect interpretations. The second post uses concepts and graphs to explain how P values and significance levels work. I find that these charts are a lot easier to understand than convoluted definitions! Finally, the third post provides tips to avoid being fooled by false positives and other misleading results.

- [How to Correctly Interpret P Values](#)
- [How Hypothesis Tests Work: Significance Levels and P Values](#)
- [Five P value Tips to Avoid Being Fooled](#)

Share this:

Share 35

Share

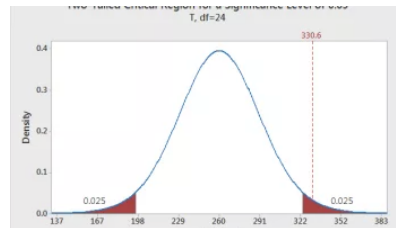
Tweet

Salvar

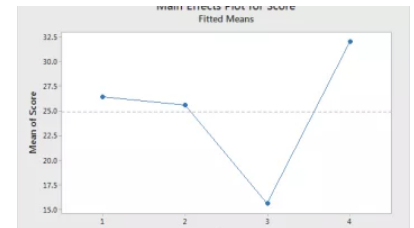
Related



[Interpreting P values](#)
In "Hypothesis Testing"



[How Hypothesis Tests Work: Significance Levels \(Alpha\) and P values](#)
In "Hypothesis Testing"



[Statistical Hypothesis Testing Overview](#)
In "Hypothesis Testing"

Filed Under: [Hypothesis Testing](#)

Tagged With: [conceptual](#)

Comments



James says

April 2, 2019 at 4:52 am

Hi Jim,