<https://www.youtube.com/watch?v=KEofcJ1tfkI&list=PLblh5JKOoLUK0FLuzwntyYI10UQFUhsY9&index=28>

Hello and welcome to the very first stat quickie.

Stat quickies are a little short video it's not a full-on stat quest where it's only gonna take a few minutes where we address a question that someone's asked me in the comments or sent me an email or if I just ran into someone in the hallway .

And they said hey I got a question.

So that's what's that quickies are for.

This very first one, we're gonna talk about what a good threshold for significance is ?

People often ask me is point zero five the best threshold for significance ?

Should I use it for everything all the time ?

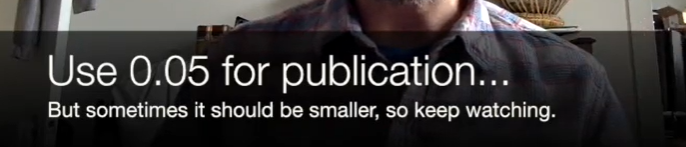
No matter what or are there exceptions, or what would how do I deal with this.

Okay !!!

So what I'm going to talk about is one the original threshold of point O five was randomly selected.

Well I'm not necessarily randomly but it's relatively arbitrary.

There's no biological or natural reason at point zero five is the optimal threshold for significance.



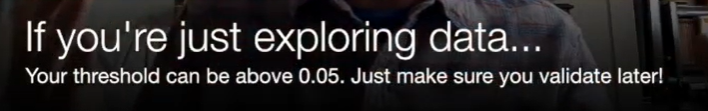
The one that we use mostly in science in publications and whatnot we even used it for a long time.

And for the most part it doesn't okay job.

Lately there have been some high-profile articles about people doing p-hacking and kind of getting away with murder in terms of statistics, but for the most part zero point four oh five has been sort of a good way to hedge the bets five percent of the time, you're wrong.

But 95 percent of the time you're right and that seems to be a cost-effective threshold for most of at least biological science.

Okay !!!



So now let's talk about what good thresholds are other than 0.05.

Okay so if you're got data that you're gonna publish, you've done your analysi,s you're gonna publish it.

What's a good threshold for that well almost always point oh five is a great threshold for that because that's what the editors are expecting.

That's what the reviewers are expecting.

If you can go below that, that's awesome.

Okay !!!

But there are exceptions to that.

Down before you run off let's listen to all those exceptions because sometimes it needs to be smaller and we're going to get to that in just a second.

Okay !!!

So here's another threshold for significant we're going to talk about one is when you're not necessarily gonna publish the data.

You say like you've downloaded a dataset from the internet, you've got a hypothesis that seems reasonable, you've tested that hypothesis, the p-value is not point over five.

Okay, it's a little bit higher.

Okay !!!

Is that a bad thing ?

Well it depends on what you're gonna do.

If you're gonna publish immediately yes you want it to be lower than pointer heart but if instead you're just using this limited data set that you got off the internet for free, and you're just trying to test some basic hypothesis, that you're later going to confirm or validate using alternative methods, well in that case, who cares what the p-value is as long as it's kind of small.

That's good enough, you don't need a massively tiny p-value.

When you're just exploring some data set that you found, you're generating new hypotheses that you're later going to test and validate using all additional data and additional methods.

Ookay !!!

So if that's the case, you know don't don't feel like you have to stick to 0.05.

Ookay !!!

Now here's another thing you need to think about say like you've done your looks like I just got a text message next time I do this I'm gonna turn off I guess alerts or something like that.

Okay !!!

So what's the next thing we need to talk about we need to talk about the effect size.

Okay.

Say like you did an R squared and I've got a whole static West on R squared.

So check this out because I go into this in detail but say like you did an R squared and your R squared value is isn't very good.

Okay, you've got like a you know, you know r-squared of 10 percent.

Right that means that 10 percent of the variation in the data is explained by whatever it is you're interested in that's not very much.

Okay.

So even if you have a tiny tiny tiny tiny p-value.

Okay it doesn't actually matter unless you know in certain circumstances maybe that's good enough but generally speaking you want a good r-squared you want your you want a good correlation you want whatever you're studying to explain the data.

Okay.

And and so if you have a tiny p-value and a tiny effect size and not much explanation of what's go on then who cares how small your p-value is you know I've seen these things in publications, where the p-value is zero point zero zero zero zero zero zero on the r-squared is horrible and I say who cares.

Okay.

So you want to have a decent effect size.

Okay.

So up the small p-value is not enough.

Okay.

One last thing.

Okay.

Extraordinary claims need extraordinary data.

So say like you've done a great experiment proving that there are extraterrestrials flying around New York City in UFOs.

Okay.

That's an extraordinary claim.

Okay you've done your thing and you've got a p-value less than 0.05.

It's point oh four seven.

Okay.

It's less than that threshold of significance is that good enough absolutely not extraordinary claims need extraordinary data.

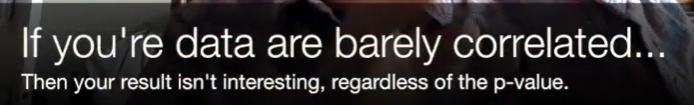
If you're gonna go out and publish a thing sending extraterrestrials are flying around New York City and UFOs your p-value better be so crazy small it's unbelievable and then people say well that p-value is so small I can't even believe it.

Therefore I have to believe that there are extraterrestrials flying around New York City and UFOs.

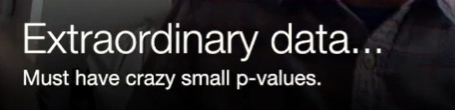
Okay that's it.

So to summarize if you're publishing your p-values you want a small p-value a PETA threshold of 0.05 it's okay.

Unless you're the thing you're studying only explains a little bit and your r-squared is kind of crummy you want a good r-squared okay .



So it's you know point oh five is fine if you're if you've got a good correlation um if you're just using it doing statistical stuff just to explore stuff that you're going to validate later on who really cares what your threshold is you're just trying to go find something of interest don't worry about 0.05.



And the other thing is extraordinary claims require extraordinary data.

If you're gonna say that there are extraterrestrials flying around New York City and UFOs you better have a darn small p-value.