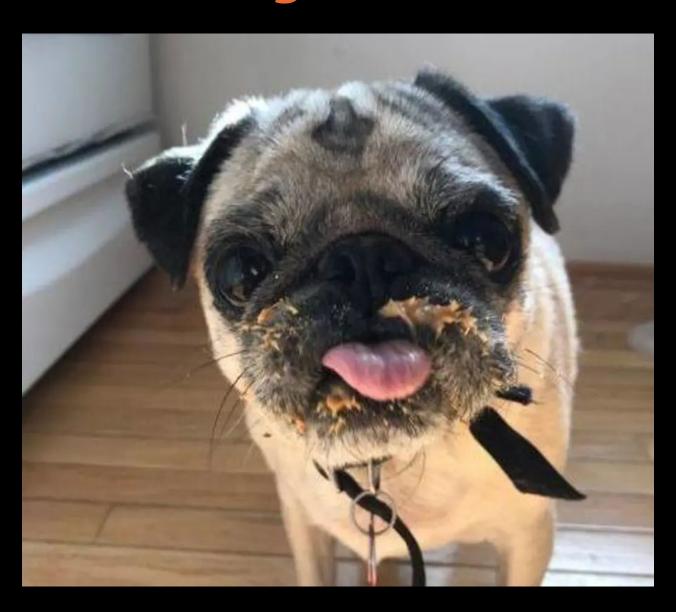
What p-values really mean

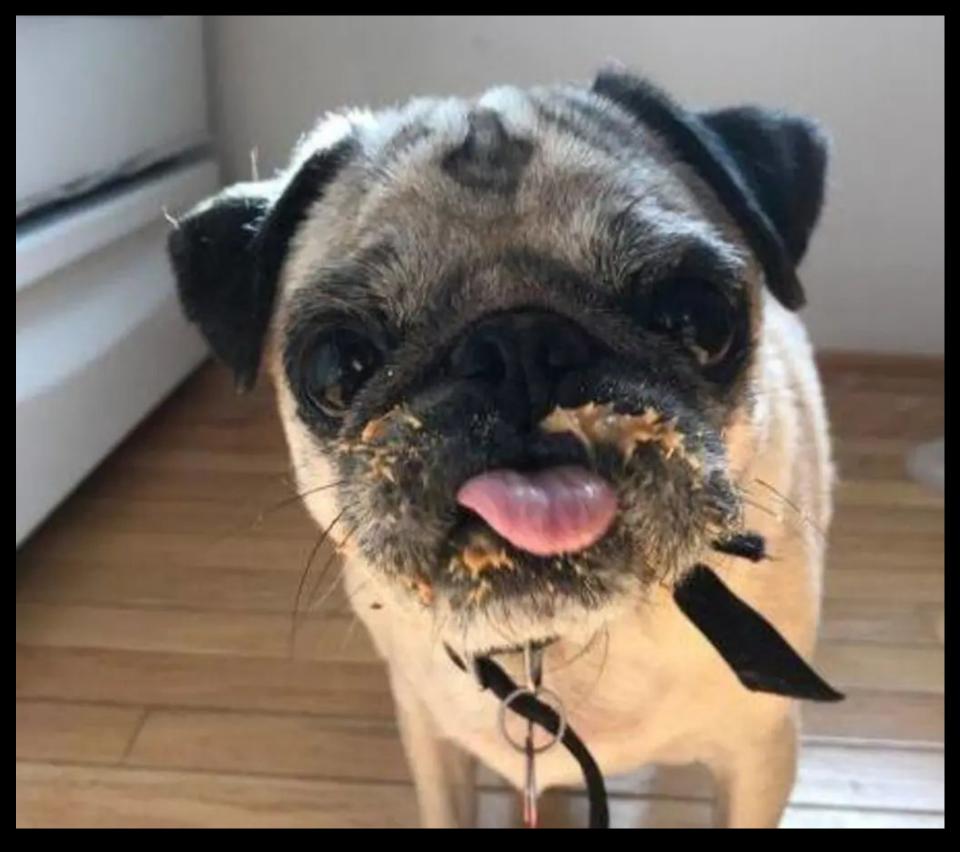


P-values are all over the data science (and the rest of science, if we're at it). But even though p-values are used so extensively, a lot of people struggle to understand what they really mean.

When looking up p-values on Wikipedia we find: A p-value is the probability of obtaining test results at least as extreme as the results actually observed, under the assumption that the null hypothesis is correct.

No wonder that this can cause confusion. Let's have a look at what it actually means.

Imagine one day you come home and see that your jar peanut butter is empty. You look around and see this:



For the sake of argument let's assume that this is your dog Taco in your house. Being a reasonable owner you don't want to scream at Taco right away, but you will put him on trial.

You start with the assumption that Taco is innocent. This is your null-hypothesis.

Now imagine a world where Taco is innocent. This is actually the hard part, describing the world of an innocent Taco.

Now you ask yourself what is the probability of this world creating an event as damning to Tacos innocence as the one observed.

That number is the p-value.

High p-value means there is a good explanation for the observation.

Maybe you have a kid that wreaked havoc, and Taco was an innocent bystander.



Or your partner secretly fed Taco.



Whatever the reason, a high p-value indicates that there are reasonable explanations for the observed situation, assuming Tacos innocence. But even with a low p-value, there might be some explanations.

Maybe a crazy YouTuber put peanut butter on his face and snuggled with Taco.



This scenario is not impossible but quite unlikely. And when you hear it you might find it ridiculous.

That is the whole point of a low p-value, it makes the explanation look ridiculous.

And what do we do if every explanation is sounding ridiculous? We reject the null hypothesis in favour of the better sounding alternative, a guilty Taco.

Bad Taco!

Keep in mind that we have not proved that Taco is guilty, we only gathered data that makes this assumption ridiculous.

This is what we do with hypothesis testing. We ask: does the collected data make our hypothesis look ridiculous, yes or no.

The p-value tells you the probability of that.

Now at which point do we decide that the null-hypothesis is to ridiculous to accept?

Commonly a value of 0.05 is considered to be the threshold of ridicule.

Is this a good value? That is a question for another day.

But keep in mind that with small data sizes your p-value tends to jump around, so don't scream at Taco if you have not collected enough data.

Remember

- 1. The p-value is the probability of an observation (peanut butter on Tacos face) under the assumption of a null hypothesis (Innocent Taco).
- 2. Smaller p-values make the assumption of an innocent Taco ridiculous.
- 3. P-values tend to jump around with small data sets.

Feel free to reach out or to connect with me for more weekly slideshows on visualization, data science and machine learning.