Hypothesis Testing
Guide to the Perplexed

Probabilities Are Not Real

Probabilities are not facts about the world. They describe how our human mind relates to the unknown future. This is the most fundamental concept to understand. Nothing changes in the world today if the probability of it raining tomorrow is 20% or 50%. Before you roll two dice there's a 1/18 chance of getting a 3. You roll and they show 5. Nothing happened to that 1/18 probability – it wasn't created or destroyed. It just exists in your head.

The human factor explains the role of the alpha value* and also why it is so helpful to think of these hypothesis tests as a back-and-forth human conversation between you (the experimenter) and **Ho** (the null hypothesis).

Who's Taller

We want to argue that men are on average taller than women. That is our assertion **Ha**, what we're trying to prove.

Ha: Men's Height > Women's Height

Our opponent is **Ho**. He is the perpetual *doubter*. Anything we try to prove, he takes the opposite side. And he always has the = with him.

Ho: Men's Height <= Women's Height

Ha: We assert than men are taller than women.

Ho: Wrong! Either they have the *same* height, or *women* are taller.

We go collect evidence. We randomly sample 50 men and 50 women – add up their heights and divide. Now we have the men's sample mean and the women's sample mean.

Men's mean height – Women's mean height = 4 inches.

Ha: In a sample size of 50 I found a 4" height difference between men and women.

Ho: Those are outliers. You randomly picked just the right group to make your point.

p-value

The p-value puts a probability on Ho's claim. p-value tells us:

If Ho is true, how likely is this result we just got?

> Ho: Your p-value is .03

p-value = .03
if men and women are the same height (Ho) – your sample results would happen 3% of the time.

p-value = .01
if Ho was true – your sample results would happen 1% of the time

p-value = .005
If Ho were true – your results would happen 0.5% of the time.

Significance (α) *

The p-value alone is not enough to make your case. Afterall – the null hypothesis admits that low-probability events *do* occasionally happen. That's his whole point.

> Ho: I'm right. You just got lucky and picked 50 people who matched your numbers.
>
> Ha: What's the probability of *that* being true? (p-value)
>
> Ho: 3%

This is where the **α alpha value** (significance) comes into play. The alpha is the subjective *human factor*.

Would you take a 3% bet?

Alpha is the minimum threshold that ***WE* the experimenter** will accept to still believe Ho. The sample results are our facts. Should we believe Ho is true if he produces the sample results less than 5% of the time?

$\alpha = .05$ and $p < \alpha$?

Conclusion	p-value	probability of sample (future) results agreeing
Ho is true	.03	3%
Ho is false (Ha)	1 – (p-value)	97%

Which conclusion would you <u>bet on</u> for the *next* sample?

Take the Inverse

Instead of p = 0.03, imagine the opposite scenario.

> Ho: The p-value is .97. If I'm right you'll get those same results 97% of the time!
>
> Ha: *Ugh*….I was ready to believe you at 5%.

$$\alpha = .05,\ p = .97 \rightarrow \alpha < p$$

What can you say? If the p-value is 97% then the sample data practically **confirms** Ho. It doesn't prove it, but it says that 97% of the time the evidence is going to be on Ho's side.

What We Expect to See

We are trying to predict the unknown. All we have is a random sample of it. Hypothesis testing provides a mathematical framework for asking – if Ho's assumption about the world is true, how often should we expect to see data consistent with it? And if that number (p) is below our threshold (α), is it reasonable to believe the assumption at all? Hypothesis testing and probabilities are mental models that help us decide how to place our bets.

Scott Gross
1/31/2025