

1.01 Tests for two groups: Null hypothesis testing

Inferential statistics or statistical hypothesis testing is all about using a limited sample from a population to make inferences about that entire population. We don't just describe some population parameter - some aspect of the population we're interested in; we use a sample statistic to *decide* something about that population parameter. To do this we specify **statistical hypotheses** about the population, a **null hypothesis** and an **alternative hypothesis**. In this video I'll explain why we need a null hypothesis.

Let's start with an example. I recently started reading about raw meat diets for cats. Suppose I want to test whether a raw meat diet is healthier for cats than regular canned food with cooked meat and grains.

I could ask all my students with cats to participate in a study where half is randomly assigned to feed their cat raw meat, while the other half provides canned food. After two months a veterinarian determines the cats' health by analyzing blood work and stool samples, resulting in a health rating between zero and ten.

If diet is related to health, the two groups will come from distinct populations, one with a higher mean health value than the other. I expect the mean of the raw group minus the mean of the canned group to be larger than zero if raw meat truly results in better health.

Suppose I find a sample difference between group means of 1.12. This difference looks pretty big, so why can't we just infer there's a real effect of diet in the population?

Well it's possible that this is a chance finding. The true difference may be over- or underestimated because the samples aren't perfectly precise representation of the population, simply due to chance.

So we want to take chance - or lack of perfect precision - into account. If our sample statistic is less precise, our decision has a larger probability of being incorrect.

We introduce probability by asking: If we were to sample indefinitely, what would the distribution of our sample statistic look like? This is useful because if we know what the distribution looks like, we can say what the probability of a certain range of sample statistic values is, by looking at the area under the curve. So if we know the distribution we can determine probabilities and make a more informed decision.

The shape of the distribution is determined by the type of statistic we're interested in. Statisticians like Fisher, Pearson, Gosset and



Neyman have shown that proportions, means, differences between two means, differences between more than two means - they're all associated with differently shaped distributions.

The shape is also affected by the sample size and variation in the population, which influence the precision of the sample statistic. Bigger samples provide more precise estimates of the population parameters. Large variation in the population results in less precise estimates, so a wider sampling distribution. By dividing the sample statistic by the standard error, we correct for variation and sample size; the sampling distribution for our particular sample is turned into a standardized probability distribution. Now we know the shape and the scale of the probability distribution.

It would be great if we could now say for example, that based on the sample difference of 1.12, the probability that a true population difference favoring a raw diet exists is 0.32.

Except it doesn't work that way, at least not using frequentist statistics. We simply can't calculate this probability, because we don't know where the probability distribution lies. We know the shape and scale, but not the location, which is determined by the population mean. The only way to pin it down is to make assumptions about its exact location, which means at best we can calculate the probability of finding a sample difference of 1.12, *assuming the true population value is exactly X*. But what is X? The true difference could lie at 10 or 4, or -1, we don't know. As long as we can't pin it down, we can't calculate probabilities.

As an aside - unlike frequentist statistics, Bayesian statistics *do* allow us to calculate the probability of a hypothesis given the data. They require additional assumptions about the hypothesized population value, which some people find problematic. For now we'll stick to classical statistics, which are still most commonly used. Bayesian methods are gaining popularity however, so keep an eye out for courses on this topic!

Ok, back to our problem of the floating probability distribution. How do we pin it down without going Bayesian? We have to assume an exact value for the population parameter. Assuming there even is an effect of diet, we don't know the size of the effect. If we guess and choose the wrong value we over- or underestimate the probability and draw the wrong conclusion.

But we do know what the exact population value is if there is no effect of diet. There's no chance of guessing it wrong; if diet is not effective, the difference in the population will be exactly zero. This is what we call the null hypothesis. It provides an unambiguous, exact value we can use to pin down the distribution. We call this final version of the distribution the test statistic distribution. If the null hypothesis is true, the difference in the



population will be zero, which is the most likely value for our test statistic, so the mean of the test statistic probability distribution lies at zero.

So instead of arbitrarily guessing an exact value for the true difference and trying to show this guess is very likely to be true, we formulate a null hypothesis and try to show that this hypothesis is very *unlikely* to be true. We hope to find a difference in our sample that is so large that the value lies far in the tail, with a very small corresponding probability. This way we can reject the null hypothesis.

To determine the probability we have to specify beforehand what our alternative hypothesis is, whether we expect the population value to be larger or smaller, or just different from the null value, so that we know in which tail to look. In the next video I'll discuss how we find the probability associated with our sample test statistic value and how we decide whether to reject the null hypothesis.