Statistical Significance

by Sophia

WHAT'S COVERED

This tutorial will cover statistical significance, which is an important concept in hypothesis testing. Our discussion breaks down as follows:

1. Statistical Significance

When you run a significance test, you need to determine what level of departure is considered a significant departure from what you would have expected to have happened.

IN CONTEXT

Suppose you work in research at Liter O'Cola company. They've developed a new diet cola that they believe is indistinguishable from the classic cola. To test this claim, you obtain 120 individuals to do a taste test. If the claim is true, what percent of people should select the correct cola just by random chance, by guessing?

Well, if Liter O'Cola's claim is correct, about 50% of people would just guess correctly and 50% would guess incorrectly if presented with the two options. So now the question is, at what point are we going to stop believing Liter O'Cola's claim?

Suppose 61 people were able to pick the diet cola. Is this evidence against the claim? Well, 61 is not that different from 60, so you're going to say no. This is not significantly different from what you would expect.

Conversely, suppose 102 people were able to pick the diet cola correctly. Would that be evidence against the company's claim?

In this case, you would probably say yes--102 is significantly more than 60, and 60 is what you would expect had they been randomly guessing. It's fairly unusual that you would see 102 people get it right by randomly guessing out of 120. Therefore, this is evidence that some people can taste the difference.

This is the whole idea of **statistical significance**. The result of 61 out of 120 is not a statistically significant result, meaning that it is not evidenced against the claim or null hypothesis. Conversely, the 102 would be evidence against the null hypothesis, because it's so much higher than what we would have expected. Statistical significance means that you doubt that the results that you obtain are due to chance.

Instead, you believe that it's part of some larger trend. For instance, in the cola example, you don't believe the null hypothesis that people can't distinguish. You believe that the trend is that people, in fact, *can* distinguish.

So, if 61 people correctly identify it, you're not convinced that over half can identify the diet cola. The difference might be only due to chance. In fact, it probably is. On the other hand, the difference of 42 from what you expect is probably not due to chance. That would be called statistically significant.

BIG IDEA

In many cases, it is possible to evaluate a claim without performing a rigorous statistical test. However, in future lessons you will learn how to evaluate claims where the correct conclusion is unclear.

To get a sense of how to detect statistical significance, we'll focus on differences between a population mean and the mean of a sample that was collected from this population.

Recall the following from the Central Limit Theorem: Given a population mean μ and standard deviation σ, a sample with mean $\bar{x}$ based on n values, has mean μ and standard error = $\frac{\sigma}{\sqrt{n}}$.

One main consideration is sample size. If a sample is small, it is hard to say if a result is significant since there is more chance for variation in a small sample. Therefore, for now, if n<30, there is no reliable way to tell if a result is statistically significant. When n≥30, you can use this rule of thumb to evaluate a claim about a population mean.

STEP BY STEP

1. Calculate $\frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$, then take its absolute value (if necessary).

   This is the number of standard errors that the sample mean is away from the population mean.
2. The larger the result, the more likely it is that result is statistically significant.

Large, >4	Result is likely to be statistically significant.
Small, between 0 and 1.5	Result is unlikely to be statistically significant, and therefore the difference between $\bar{x}$ and μ is likely due to chance.

Moderate, between 1.5 and 4	A more formal approach is needed to determine statistical significance. We'll explore this soon!

EXAMPLE A local researcher claims that the monthly rent for a one-bedroom apartment is $1600 with a standard deviation of $200. To test this claim, a sample of 100 one-bedroom apartments with a sample mean of a sample mean of $1720 was collected.

Is this result likely to be statistically significant or due to chance?

Since the sample size is greater than 30, we can use the rule of thumb.

$\bar{x}$=1720, μ=1600, σ=200, n=100	Identify all the given quantities
$\frac{1720-1600}{\left(\frac{200}{\sqrt{100}}\right)}=6$	Compute $\frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$ The result is positive, so no need to take absolute value.

The result is more than 4, so we conclude that this result is likely statistically significant.

Keep in mind that this rule of thumb is used to get a sense of statistical significance vs. chance. There are many more things to consider when evaluating the statistical significance of a claim, which we'll discover later.

TERM TO KNOW

Statistical Significance

The statistic obtained is so different from the hypothesized value that we are unable to attribute the difference to chance variation.

2. Practical Significance

Practical significance refers to whether or not something is meaningful in the real world. With practical significance, we can ask ourselves, in practice, "Does this affect our lives?"

It's important to make the distinction between practical significance and statistical significance. They're not necessarily the same thing.

Suppose you had a large enough sample of coin flips. You would expect a fair coin to yield 50% heads and 50% tail. If the sample size was large enough, even a sample with 50.1% heads could yield a statistically significant result, even though 50.1% is very close to 50%.

The statistical significance argument is based largely on sample size and how far off your sample is from the expected value. If the sample size is very large, you don't need to be very far off. If the sample size is small, you need to be further away to claim statistical significance.

Apply this to our coin flips, where if the sample size is very large, you might get something like 50.1% heads, which could be statistically significant, but is not likely practically significant. In other words, the coin is not fair, but it's so close it doesn't matter for any practical purpose.

IN CONTEXT

A state survey of all high school students finds that 15% of 10th graders drink regularly. A town randomly selects 100 students and finds that 18% of their 10th graders drink regularly.

By doing some statistical test and setting a significance level and if this passes that test, then we can say whether this is statistically significant or not.

Now whether this is practically significant, we need to consider if this affects our lives in the real world. For this town, even if it came back that there was no statistical significance and the 18% result was random, you may still want to do something about this report because it may still hold meaning for you in the real world because this is about something serious.

So without doing a test, we cannot say that this is statistically significant, but it may be practically significant nonetheless.

EXAMPLE Consider the example we worked through earlier, in which a local researcher claims that the monthly rent for a one-bedroom apartment is $1600 and a sample of 100 one-bedroom apartments was collected with a sample mean of $1720.

We deduced that this result is likely statistically significant, but this result is also practically significant since apartment rents do affect our lives. Having a sample mean that is $120 more than the claimed average might make someone hesitant to rent there since there might be concerns about affordability.

TERM TO KNOW

Practical Significance
An arbitrary assessment of whether observations reflect a practical real-world use.

SUMMARY

You learned about statistical significance and how to measure it versus practical significance. You also learned how those two are not necessarily the same. Statistical significance is the extent to which a sample measurement is evidence of a trend, like being able to taste the difference between regular cola and diet cola, and whether the difference can be attributed to chance. Sometimes very small differences can be statistically significant, though not have a lot of real-life meaning, which is practical significance.

Good luck!

Source: THIS TUTORIAL WAS AUTHORED BY JONATHAN OSTERS FOR SOPHIA LEARNING. PLEASE SEE OUR **TERMS OF USE**.

TERMS TO KNOW

Practical Significance

An arbitrary assessment of whether observations reflect a practical real-world use.

Statistical Significance

The statistic obtained is so different from the hypothesized value that we are unable to attribute the difference to chance variation.