sas

Print

Statistical Hypothesis Test

In inferential statistics, you use statistics to infer information about population parameters. But inferences aren't exact. As you've seen, there's variability in parameter estimates. In a hypothesis test, you phrase questions as tests of hypotheses about population parameters. Your null hypothesis is usually one of equality, whereas the alternative hypothesis is one of inequality. The alternative hypothesis is typically what you suspect, or what you're attempting to demonstrate.

The conclusions reached from your hypothesis test are usually phrased in reference to the p-value, or the probability of obtaining a test statistic as extreme or more extreme than the one observed in your data given that the null hypothesis is true. When the p-value is low, it provides doubt about the truth of the null hypothesis. But how low does the p-value need to be before you reject the null hypothesis completely? That depends on you. A common significance level is 0.05 (1 chance in 20). If you require a stricter cutoff, you might consider lowering your significance level when planning your analysis. Let's look at an example.

To begin each game of a worldwide soccer tournament, one of the teams chooses heads or tails, and the match referee tosses a coin. The team that wins the toss decides which goal it will attack, and the team that loses the toss takes the kick-off to start the match. But how do you know the coin is fair? You might suspect that the coin is not fair, but you begin by assuming that it is fair.

Next, you select a significance level: if you observe five heads in a row or five tails in a row, you conclude that the coin is not fair. Otherwise, you decide that there is not enough evidence to show that the coin is biased. With a fair coin, a true null hypothesis, the probability of observing 5 heads or 5 tails in a row in five trials is 1 out of 16.

Why 1 out of 16? There are 5 tosses, and each has a 50% probability of being heads. Tosses are independent, and therefore, the probability of 5 heads is $(1/2)^5$ or 1 out of 32. The probability of 5 tails is also 1 out of 32. These probabilities can be added together to give the probability of 5 heads or 5 tails as 2 out of 32, or 1 out of 16. So the significance level is 1/16, or 0.0625.

To collect evidence, you flip the coin five times and count the number of heads and tails. Finally, you decide either that there is enough evidence to reject the assumption that the coin is fair (either all trials are heads or all trials are tails), or that there is not enough evidence to reject the assumption that the coin is fair (meaning not all trials are either heads or tails).

So, you performed a hypothesis test and used a decision rule to decide whether the coin was fair or not. But was your decision correct? You began by assuming that the null hypothesis is true: that the coin is fair. But what if you're wrong?

If you reject the null hypothesis when it's actually true, you've made a Type I error. The probability of committing a Type I error is α. α is the significance level of the hypothesis test. In the coin example, it's the probability that you conclude that the coin is not fair when it is fair.

A Type II error, often referred to as β, is when you fail to reject the null hypothesis and it's actually false. In the coin example, it's the probability that you fail to find that the coin is not fair when it is in fact biased. Type I and Type II errors are inversely related. As one type increases, the other decreases.

Power is the probability that you correctly reject the null hypothesis. The power of a statistical test is equal to $(1 - \beta)$, where again, β is the Type II error rate.

Actual/ Decision	**H_0 is true**	**H_0 is false**
Fail to reject H_0	Correct	Type II error $p(\text{Type II} \mid H_a)=\beta$
Reject H_0	Type I error	Correct

	p(Type I | H_0)=α	(1 - β)=Power

Close