

When we need to decide if the evidence is strong enough to reject H_0 , we have seen that the key is whether p-value falls below a prespecified significance level α .

We reject H_0 if p-value $\leq \alpha$
We do not reject H_0 if p-value $> \alpha$

The smaller α is, the stronger the evidence must be to reject H_0 .

Because of sampling variability, decisions in hypothesis testing always have some uncertainty. A decision can be in error.

Hypothesis testing has two types of potential errors

- ① Type I error
- ② Type II error

Type I error:— Type I error occurs when H_0 is true, but it is rejected.

One can think Type-I error as a false positive, because a positive decision is reached by rejecting H_0 , yet the decision is false.

Type II error:— Type II error occurs when H_0 is false, but it is not rejected.

One can think Type-II error as false negative, because a negative decision is reached by not rejecting H_0 , yet the decision is false.

Example: Consider a decision in a legal trial.

H_0 : defendant is innocent

H_1 : defendant is guilty

A Type-I error, rejecting null hypothesis, occurs in convicting a defendant who is actually innocent.

A Type-II error, not rejecting H_0 even though it is false, occurs in acquitting a defendant who is actually guilty.

A potential consequence of a Type-I error is sending an innocent person to jail, whereas a potential consequence of a Type-II error is setting free a guilty person.

The significance level is the probability of Type-I error.

Let us look at a two-sided test about a proportion to see how we get this result.

Recall for a two-sided test, the two-tail probability that forms p-value ≤ 0.05 corresponds to test statistics z satisfies $|z| > 1.96$, which forms rejection region.

Standard normal distribution of z test statistic, if H_0 true

The probability of rejecting H_0 is the probability of observing the z -statistics in the rejection region.

This prob would be

$$P(z < -1.96) + P(z > 1.96) = 0.05$$

we see that the summation of probability is equal to the significance level.

$$\text{Prob}(\text{Type I error}) = \text{significance level}(\alpha)$$

The good news is that we can control the prob of Type-I error by our choice of the significance level.

The more serious the consequences of a Type-I error, the smaller α should be.

Trade-off between Type-I and Type-II errors:

Why don't we make extremely small prob of Type-I error by setting α to be small such as $\alpha = 0.000001$. For example, why don't we make it almost impossible to convict someone who is really innocent.

When we make α smaller, we need a smaller p-value to reject H_0 . It then becomes harder to reject H_0 . Thus, it will also be harder to reject H_0 even if H_0 is false.

In other words, the smaller we make the prob of Type-I error, the larger the prob of Type-II error becomes, that is failing to reject H_0 even though it is false.

As $P(\text{Type I error})$ goes down, $P(\text{Type II error})$ goes up

calculation of $P(\text{Type II error})$ can be complex. In practice, to make decisions, we need to set only $P(\text{Type I error})$, which is the significance level.