Hypothesis Testing

Jian Zhang

University of Kent

Sampling to test a theory

- In a test of a hypothesis, we test a certain given theory or belief about a population parameter.
- We may want to find out, using some sample information, whether a given claim (or statement) about a population parameter is true.
- If our population has a mean of μ and a variance of σ^2, then we know that the sample mean will be approximately normally distributed with a mean of μ and a variance of ${\sigma^2}/{n}$
- We can therefore evaluate whether the sample is consistent with a claim.

Motivating Example

- As an example, a soft drink company may claim that, on average its cans contain 12 ounces of soda.
- A government agency may want to test whether such cans do contain, on average 12 ounces of soda.

Example continued

- Suppose we take a sample of 100 cans of the soft drink under investigation.
- We then find out that the mean amount of soda in these 100 cans is 11.89 ounces.
- Based on this result, can we state that all such cans contain, on average, less than 12 ounces of soda and that the company is lying to the public?
- Not until we perform a test of hypothesis can we make such an accusation.

Analysing results

- The difference between the 12 ounces and then 11.89 ounces may have occurred only because of the sampling error.
- Another sample of 100 cans may give us a mean of 12.04 ounces.
- Therefore, we perform a test of hypothesis to find out how large the difference between 12 ounces and 11.89 ounces is and whether this difference has occurred because of chance alone.

Two Hypotheses

Consider as a non-statistical example a person who has been indicted for committing a crime is being tried in a **court**. Based on the available evidence, the judge or jury will make one of two possible decisions:

1. The person is not guilty.
2. The person is guilty.

At the outset of the trial, the person is presumed not guilty. The prosecutor's job is to prove that the person has committed the crime and, hence, is guilty.

Statistical Hypotheses

- In statistics, the **person is not guilty** is called the **null hypothesis** and the **person is guilty** is called the **alternative hypothesis**.
- The null hypothesis is denoted by H_0, and the alternative hypothesis is denoted by H_1.
- At the beginning of the trial, it is assumed that the person is not guilty.
- The null hypothesis is usually the hypothesis that is assumed to be true to begin with.

Court Hypotheses

The two hypotheses for the court case are written

Null hypothesis	H_0: The person is not guilty
Alternative hypothesis	H_1: The person is guilty

Returning to the soda example, let μ be the mean amount of soda in all cans. The company's claim will be true if $\mu = 12$, thus the null hypothesis will be written as

H_0: $\mu = 12$ ounces

In this example, the null hypothesis can also be written as $\mu \geq 12$ ounces because the claim of the company will still be true if the cans contain, on average, more than 12 ounces of soda. The company will be accused of cheating the public only if the cans contain, on average, less than 12 ounces of soda. The alternative hypothesis is written as

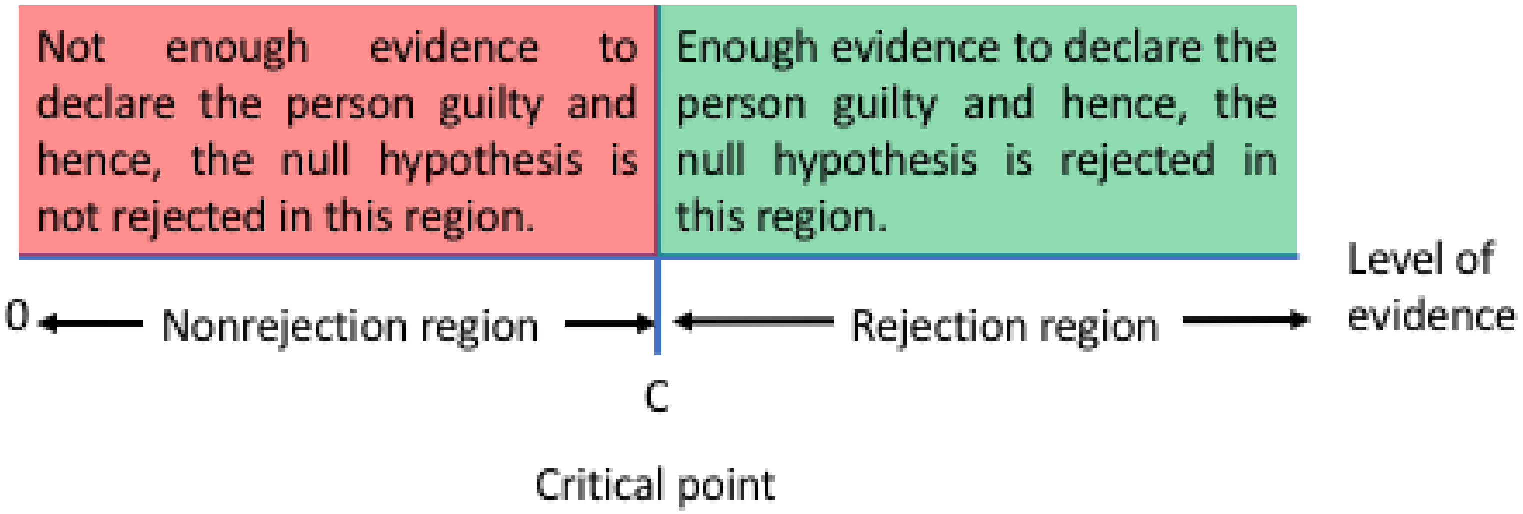
H_1: $\mu < 12$ ounces

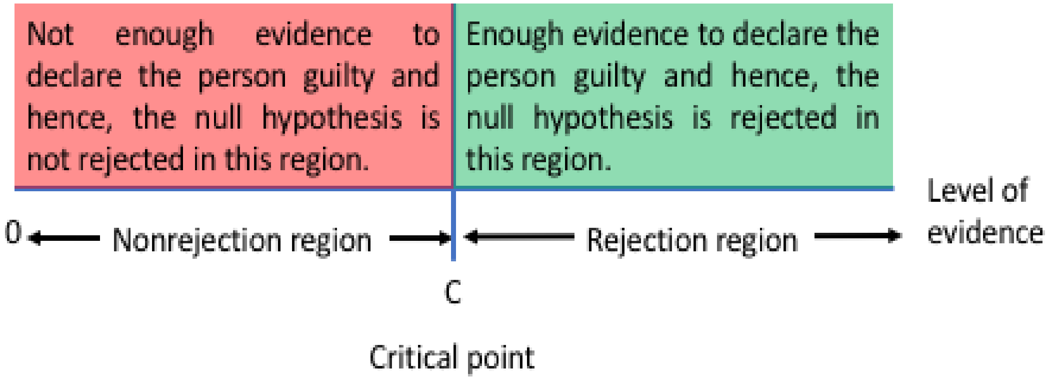
Starting assumption

- Returning to the court case example, the trial begins with the assumption that the null hypothesis is true – that is the person is not guilty.
- The prosecutor assembles all the possible evidence and presents it in court to prove that the null hypothesis is false, and the alternative hypothesis is true.
- In the case of the statistics example, the information obtained from the sample will be used as evidence to decide whether the claim of the company is true.

Rejection and Non-rejection regions

Let us consider the level of evidence against the individual accused of the crime in the court case example.





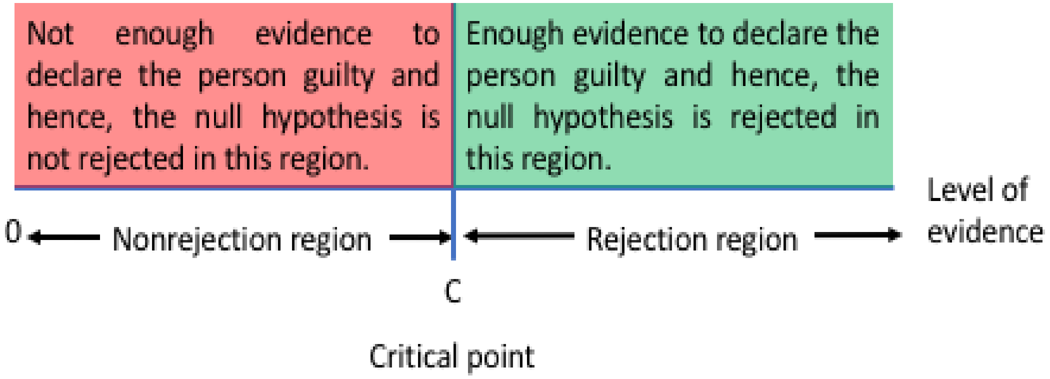
The point marked zero indicates that there is no evidence against the person being tried.

The farther we move toward the right on the horizontal axis the more convincing the evidence is that the person has committed the crime.

We have arbitrarily marked a point C on the horizontal axis.

Let us assume that a judge (or jury) considers any amount of evidence from point C to the right to be sufficient and any amount of evidence to the left of C to be insufficient to declare the person guilty.

Point C is called the critical point (or critical value) in statistics.



If the amount of evidence falls in the area to the left of point C, the verdict will reflect that there is not enough evidence to declare the person guilty. Consequently, the person will be declared not guilty. In statistics, this decision is stated as **do not reject H_0**. It is equivalent to saying that there is not enough evidence to declare the null hypothesis false.

If the amount of evidence falls to the right of C, the verdict will be that there is sufficient evidence to declare the person guilty. In statistics, this decision is stated as **reject H_0.** Rejecting H_0 is equivalent to saying that the alternative hypothesis is true.

Types of Errors

- We know that a court's verdict is not always correct.

		Actual situation	
		Person is not guilty	Person is guilty
Court's decision	Person is not guilty	Correct decision	Type II error, β
	Person is guilty	Type I error, α	Correct decision

- A **Type I error** will occur when we reject H_0, even though H_0 is true. The value of α, called the **significance level** of the test, represents the probability of making a Type I error.
- A **Type II error** will occur when H_0 is not true, but there is insufficient evidence to reject H_0. The Type II error is denoted by β. The value of **$1-\beta$** is called the **power of the test**. It represents the probability H_0 is rejected when it is false.

Tails of a Test

- In the court case example, the rejection region is to the right side of the critical point.
- However, in statistics the rejection region for a hypothesis testing problem can be on both sides, or it can be on just the left side or just the right side, depending on the structure of the hypotheses.
- To determine whether a test is two-tailed or one-tailed we look at the sign in the alternative hypothesis.
- If the alternative hypothesis has a not equal to sign, $\neq$, it is a two-tailed test, whilst if the sign is “>” or “<” it will be a one-tailed test.

Tails for soda example

Return to the earlier soda example. Recall, our hypotheses were

H_0: μ = 12 ounces
H_1: μ < 12 ounces

This will result in a left-tail test – the rejection region is in the left tail of the distribution curve and the area of this rejection region is α.

For a two-tailed test, the rejection region will occur both in the left tail and the right tail and the area of each rejection region will be $\alpha/2$.

P-values

When conducting a hypothesis test we can calculate a p-value, which is defined as:

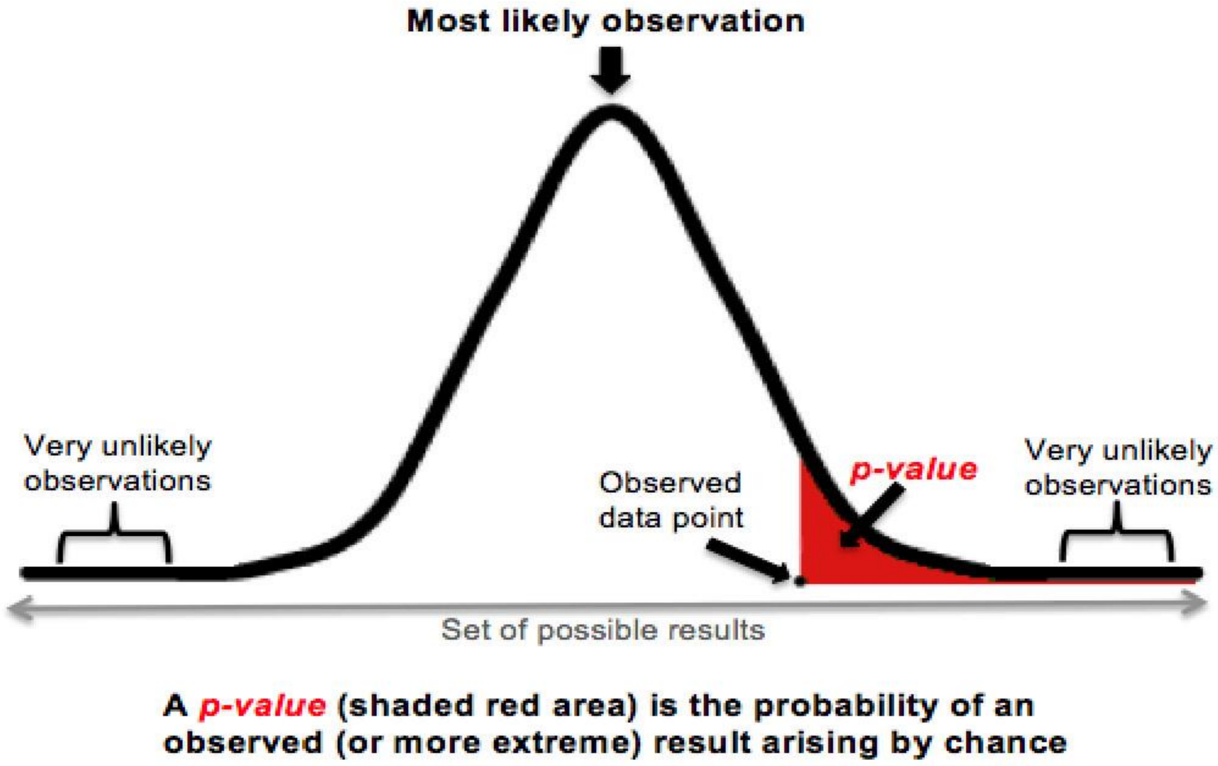
the smallest level of significance at which the given null hypothesis is rejected.

Using this p-value, we state the decision. If we have a predetermined value of α, then we compare the value of p with α and make a decision.

We reject the null hypothesis if the p-value $\leq \alpha$.

We do not reject the null hypothesis if the p-value $> \alpha$.

For a one-tailed test the p-value is given by the shaded red area in the diagram below:



For a two-tailed test, the p-value is twice the area in the tail of the sampling distribution.