Hypothesis Testing

Ellie Trace

Summary

A hypothesis test is a statistical process used to determine whether there is enough evidence in a sample to support or reject a claim about a population. It involves comparing a null hypothesis, which represents the status quo, against an alternative hypothesis using sample data, a chosen significance level, and either critical values or p-values. This guide will focus on the general structure of a hypothesis test, critical values, how to choose which type of test to use, as well as when to reject, or not reject a hypothesis.

Before reading this guide, it is recommended that you read [Guide: Introduction to probability]

What is a hypothesis test?

Hypothesis testing is one of the most important and most useful tools in statistics. They help you to use data from a sample to test whether or not it is reasonable to believe a certain statistical characteristic is true for the whole population.

For example, Cantor's confectionery shop have 50 customers on average one week. Can they say that their daily average customers has now probably increased compared to their previous average of 45 per day?

MOTIVATION

Setting up a hypothesis test

Hypothesis, hypothesis test

A **statistical hypothesis** (shortened to **hypothesis** in this guide) is a statement showing an idea, or something that could possibly be true, about the whole of a particular data set.

A **hypothesis test** allows you to say, with a defined level of certainty, whether or not a hypothesis can be rejected. This determines whether there is enough statistical evidence to show that the original hypothesis is unlikely to be true.

Generally speaking, when you define hypotheses in statistics, they follow a very specific format, and refer to the *entire population* being studied.

- **Null hypothesis (H_0):** This hypothesis represents the 'status quo' or no effect. It is always a statement of equality.
- **Alternative hypothesis (H_1):** This is the hypothesis that you are trying to test. It is always a statement of **inequality**.

Depending on the question, the alternative hypothesis H_1 can either be **one-tailed** or **two-tailed**:

```
- One-tailed means you are you are testing whether the characteristic has increased o
- Two-tailed means you are testing whether the tested characteristic is equal to the
```

Here's some examples of hypotheses that require testing. In these examples, the Greek letter μ (mu) is used for the **population mean**; see [Guide: Means, variance, and standard deviation] for more.

Example 1

Cantor's Confectionery wants to determine whether their waiting times in their corporate headquarters are longer than their target of 15 minutes per person.
You can set up a hypothesis test to evaluate this claim.
Null hypothesis (H_0): The average waiting time is equal to the target waiting time of 15 minutes. (Remember that the null hypothesis is always statement of equality!)
Alternative hypothesis (H_1): The average waiting time is longer than the target waiting time of 15 minutes.
So the null hypothesis is $H_0 : \mu = 15$ and the alternative hypothesis is $H_1 : \mu > 15$.
This is a one-tailed alternative hypothesis.

Example 2

Context: Cantor's Confectionery wants to determine if a new sweet production method produces different average approval scores compared to their traditional method.

Null hypothesis (H_0): The mean approval score for the new method (n) is equal to the mean approval score for the traditional method (t) $(\mu_n = \mu_t)$

Alternative hypothesis (H_1): The mean approval score for the new method is not equal to the mean approval score for the traditional method, indicating a difference in performance. $(\mu_n \neq \mu_t)$

Example 3

Context: The confectionery factory wants to test if the proportion P of defective products produced is lower than last year's average of 2%.

Null hypothesis (H_0): $P = 0.02$

Alternative hypothesis (H_1): $P < 0.02$

Test selection and statistic calculation

Choosing the appropriate test is really important because they all have different inputs depending on the data you have available to you. You can use this interactive flow chart of which test statistic you would want to use in whichever situation! Once you have found the test you want to use, you can refer to their relevant guide to make sure you are doing the correct calculations.

Critical values and p-values

You can use two different methods to decide whether or not to reject your hypothesis.

Critical value

A **critical value** is the boundary that defines the rejection region (or critical region) based on **the significance level** α.

You would then use statistical tables (like Z or t tables) to find the critical value for the chosen test and then compare the test statistic to the critical value. It can be helpful to sketch or visualize the graph whilst doing this to make sure you are not missing any negatives and you are using the right critical region depending on if you are doing a one- or two-tailed test.

p-**value**

The p-**value** is the probability of obtaining a test statistic as extreme as the observed and is equal to the area under the distribution curve that corresponds to the test statistic.

One-tailed test: If you have an upper-tailed test:

$$H_1 : \mu > \mu_0$$

(Where μ_0 is the value you are testing your sample data against), the P-value is the area to the right of the test statistic under the probability distribution curve.

If you have a lower-tailed test:

$$H_1 : \mu < \mu_0$$

the P-value is the area to the left of the test statistic under the probability distribution curve.

Two-tailed test: In the case of a two-tailed test you need to find the area in both tails. Essentially, you half the P-value.

You would then compare the P-value to α: If the P-value is less than α, **you reject** H_0. If the P-value is greater than or equal to α, **you fail to reject** H_0

Significance levels

A **significance level** (often called alpha (α)) is the level of certainty you want to test your hypothesis with, or the line at which you would reject H_0. Most commonly, α is set to 0.05 (5%), 0.01 (1%) or 0.10 (10%). The most common choice is $\alpha = 0.05$ (so a significance level of 5%). **This means you are willing to accept a 5% risk of rejecting a true null hypothesis.**

> Tip
>
> Rejecting a true null hypothesis is known as a **Type I error**, or a false positive! For more information on these [See guide: Errors in hypothesis testing]

This has a different impact based on whether you are performing a one-tail or a two-tail test. As shown in the graph below, for a one-tailed test, that whole 5% rejection region will be at one end of the results. This is because you are only testing whether the characteristic is greater than or equal to our comparative characteristic. In the case of a two-tailed test, that 5% will be split across either end of the distribution to reject the 2.5% at both far left and far right of the curve. This is because you are testing whether the test characteristic is both greater than or less than the comparative characteristic. If your test statistic falls in the critical region at the far end(s) of the probability distribution *See guide: Introduction to probability distributions*, that is when **you reject the null hypothesis** H_0.

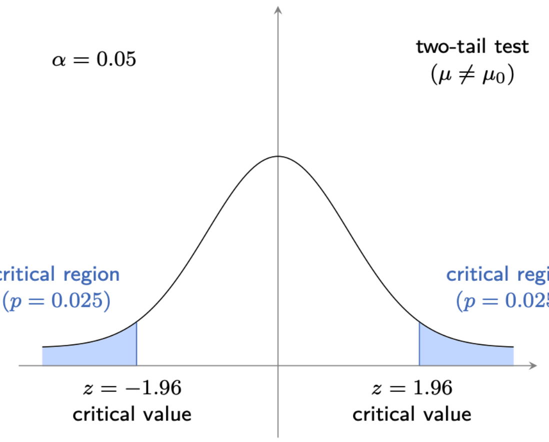

Figure 1: A **two-tailed probability distribution** where $\alpha = 5\%$, the critical value is 1.96 and there is a highlighted area on both the left and right tails of the curve representing a probability area of $p = 0.025$ each. These highlighted areas represent the critical regions or the areas where you would reject the null hypothesis H_0

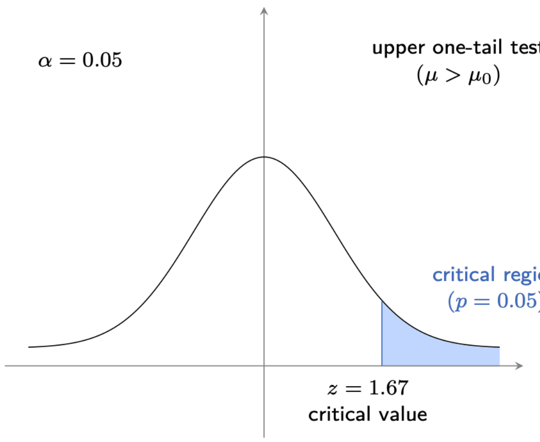

Figure 2: An **upper one-tail probability distribution** where $\alpha = 5\%$, the critical value is 1.67 and there is a highlighted area on the right tail of the curve representing a probability area of $p = 0.05$. This highlighted area represents the critical region or the area where you would reject the null hypothesis H_0

Forming a conclusion

To form a conclusion you then use the test to decide whether or not to reject your null hypothesis H_0.

If the test statistic falls in the critical region or if the P-value is less than α then you would **reject** H_0. This suggests there is enough evidence to support the alternative hypothesis.

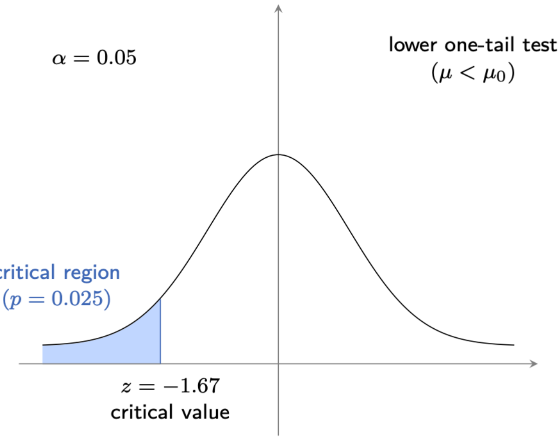

Figure 3: A **lower one-tailed probability distribution** where $\alpha = 5\%$, the critical value is -1.67 and there is a highlighted area on the left tail of the curve representing a probability area of $p = 0.05$. This highlighted area represents the critical region or area where you would reject the null hypothesis H_0

If the test statistic does not fall in the critical region or if the P-value is greater than or equal to α. This means there is not enough evidence to support the alternative hypothesis and you **fail to reject** H_0.

To complete the problem, you then must formally state your conclusion!

> Tip
>
> "I reject H_0 as there is sufficient evidence to conclude that the mean is greater than 50."

> Tip
>
> "I do not reject H_0 as there is not sufficient evidence to conclude that the mean is greater than 50."

> Warning
>
> Hypothesis tests are based on sample data, not on the entire population so you can never accept either hypothesis; it is always a rejection or a failure to reject. This is because you cannot definitively "prove" the null hypothesis to be true.
> If you do not reject H_0, this does not mean that the null hypothesis is necessarily correct; it means that the sample data didn't provide strong enough evidence against it.
> Similarly, if you reject H_0, this does not mean that the null hypothesis is necessarily incorrect; it means that the sample data provided strong enough evidence for the alternative hypothesis.
> For more information about this see [Guide: Errors in hypothesis testing]*

Quick check problems

1. What would your hypotheses be when testing whether the average battery life of a laptop is less than 10 hours?

a)$H_0 : \mu > 10\ H_1 : \mu < 10$

b)$H_0 : \mu = 10\ H_1 : \mu \neq 10$

c)$H_0 : \mu = 10\ H_1 : \mu < 10$

2. What α corresponds to a 10% level of significance?

3. What test would you use to compare means when you have one sample and you know the population standard deviation?

4. You are performing a two tail test with $\alpha = 0.05$ and critical values of 1.96 and -1.96. Your test statistic is -2.34. What conclusion do you draw?

Further reading

For more questions on the subject, please go to Questions: Hypothesis tesing

[For more information on how to perform a Z-test please see guide: Z-testing] [For more information on type I and type II errors please see guide: Errors in hypothesis testing]

Version history

v1.0: initial version created 12/24 by ect6 (as part of a University of St Andrews VIP project)