8 HYPOTHESIS TESTING

In the last two chapters, we discussed various statistical distributions and basic concepts in inferential statistics including the CLT. In this chapter we seek to build on that knowledge and to apply it in hypothesis testing.

8.1 TERMINOLOGY

Making assumptions is a crucial aspect of decision making. Generally, we make assumptions about specific issues based on our experience, knowledge or information. **Hypothesis testing** is a scientific approach that makes use of statistics to test these assumptions. This is a terminology heavy topic, so let us get to know some common terms in hypothesis testing.

Null and Alternative Hypotheses

A **null hypothesis** is an assertion that we hold as true unless we have sufficient evidence to conclude otherwise. It is often denoted by H_0.

The **alternative hypothesis** is a negation of null hypothesis and is denoted by H_1. This is usually the hypothesis we are trying to prove.

Test Statistic

The evidence to reject or not reject the null hypothesis is based on data from a sample. A **test statistic** is a single standardized metric that captures the information contained in the sample data. From standardized we mean that its probability distribution is known and is easy to utilize.

For example, we know from the CLT that the sample mean $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. We have already learned that any normally distributed variable can be transformed into a standard normal format using the following transformation:

$$Z = \frac{\text{Original variable} - \text{mean}}{\text{standard deviation}} \sim N(\mathbf{0}, \mathbf{1})$$

Applying this to the sample mean $\bar{x}$, we get

$$\text{Test statistic} = \text{Z statistic} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(\mathbf{0}, \mathbf{1}) \text{ i.e. } \textbf{the standard normal distribution}.$$

This test statistic is called the Z statistic and is used often in hypothesis testing.

p-Value

Once you have decided on the null hypothesis and have the test statistic, you need a reference to base your decision on. The strength of evidence in support of a null hypothesis is measured by the p-value.

The **p-value** is the probability of observing a more extreme value than that of the test statistic, provided the null hypothesis is true.

Significance Level (α)

Significance level is used to calculate the critical value against which we compare the p-value in order to accept or reject a null hypothesis. **Significance level**, denoted by "α", is a probability threshold below which the null hypothesis is rejected.

Remember, that unlike p-value, significance level is not a calculated quantity. Instead, it is arrived at by consensus among the researchers. It is set in advance before the hypothesis testing begins. Conventionally, it is taken to be 1%, 5% or 10% i.e. 0.01, 0.05 or 0.10 respectively.

Type I and Type II Errors

Significance level also denotes the probability of making a Type I error. **Type I error** is the false rejection of the null hypothesis. In contrast, **Type II error** is the failure to reject a false null hypothesis.

Reality/ States of the World

Outcome of the test	**Null hypothesis is True**	**Null hypothesis is False**
Reject null hypothesis	**Type I error** (False positive)	Correct outcome (True Positive)
Fail to reject the null hypothesis	Correct outcome (True Negative)	**Type II error** (False negative)

Few things to note about Type I error -

- It is the incorrect rejection of the null hypothesis
- Its maximum probability is set in advance as alpha
- It is not affected by sample size as it is set in advance

Few things to note about Type II error –

- It is the incorrect acceptance of the null hypothesis
- It's probability is referred to as beta
- Beta depends upon sample size and alpha. Beta gets smaller as the sample size gets larger

Of these two errors, the Type I error is considered a more serious error and hence setting the value of significance level as a small fixed number 1%, 5% or 10% helps with this.

Confidence level (1- α)

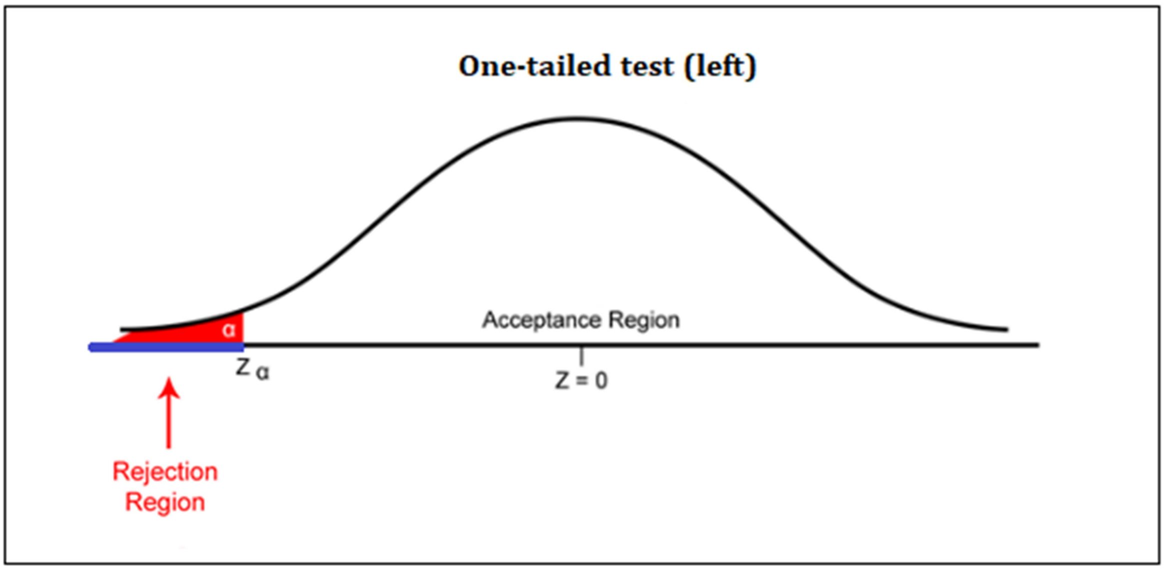
Fixing the probability of type I error at α allows us to have (1- α) percent confidence in the conclusion of our hypothesis test. This value is called the **confidence level**.

One-Tailed & Two-Tailed Tests

The alternative hypothesis determines whether to place the level of significance in one or both tails of the sampling distribution.

In a **one-tailed test**, the region of rejection is only on one side of the test statistic's distribution. For example when the alternative hypothesis is:

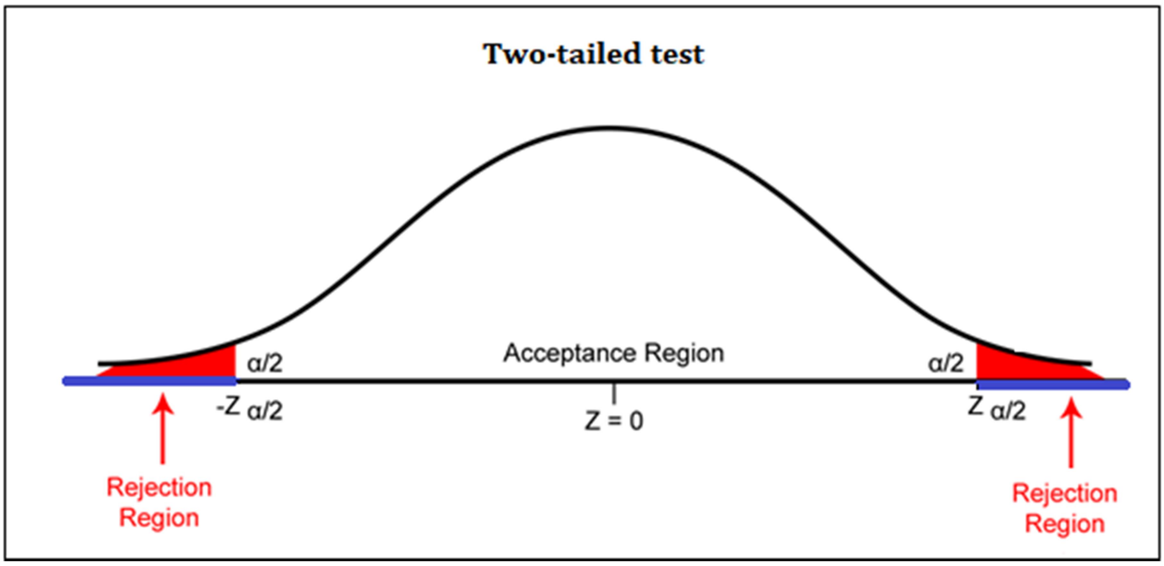
$\mathbf{H_1}: \mu > \mathbf{10}$ (One tailed-right)
or
$\mathbf{H_1}: \mu < \mathbf{10}$ (One tailed-left)



If the test statistic value lies in the region, the p-value is less than significance level, leading to the rejection of null hypothesis. In case it is not, we fail to reject the null hypothesis.

On the other hand, in a **two-tailed test**, the rejection region is on both the sides and the alternative hypothesis has '≠' sign in it. For example when the alternative hypothesis is:

$\mathbf{H_1}: \mu \neq \mathbf{10}$ (Two-tailed test)



8.2 Steps in Hypothesis Testing

While conducting a hypothesis test, we go stepwise. In this section we list the key steps in hypothesis testing:

Step 1: State the null and the alternative hypotheses

Step 2: State the significance level (α)

Step 3: State the decision criteria based on α and the alternate hypothesis (One-tailed/Two-tailed)

Step 4: Determine and calculate the test statistic (Z statistic, t statistic etc.)

Step 5: Calculate the p-value

Step 6: Compare the p-value with α, accordingly reject or not reject the null hypothesis

Step 7: Interpret the result of the test

In the next section, we will try to understand all these steps with the help of an example.

8.3 An Example

Suppose Sophie is looking to invest in a monthly income investment scheme that promises variable monthly returns. However, she will invest in it only if she is assured of an average $180 monthly income. Her financial advisor has informed that the standard deviation of monthly returns from this scheme is $75.

Sophie has a sample of 300 months' returns which has a mean of $190.
Should she invest in this scheme?

Sophie's just begun learning about hypothesis testing and is quite excited to try it on the data she's collected. So she decides to follow the steps mentioned above.

Step 1: State the null and the alternative hypotheses

Sophie will invest in the scheme if she is assured of her desired $180 average return. Thus, she frames the null and alternate hypotheses as follows:

$\mathbf{H_0}$: Null Hypothesis: mean = 180
$\mathbf{H_1}$: Alternative Hypothesis: mean > 180

Step 2: State the significance level (α)

She decides to set the significance level at 5% i.e. 0.05.

Step 3: State the decision criteria based on α and the alternate hypothesis (One-tailed/Two-tailed)

As the alternative hypothesis is in the form of a '>' type inequality, this is a one-sided test (right tailed). The rejection region will be on the right side of the distribution.
The null hypothesis will be rejected if the value of the test statistic exceeds a particular value such that the corresponding p-value becomes smaller than the significance level of 0.05.

Step 4: Determine and calculate the test statistic (Z statistic, t statistic etc.)

Relevant information about the sample is already provided as follows:
Sample size = n = 300
Sample mean = $\bar{x}$ = 190
Population standard deviation = σ = 75
Hypothesized population mean = μ = 180
As the sample size is large (n is greater than 30) and the population standard deviation is known, the appropriate test statistic is the Z statistic.

$$\textbf{Test statistic} = \textbf{Z statistic} = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{190-180}{\frac{75}{\sqrt{300}}} = 2.309$$

Step 5: Calculate the p-value

We know that p-value is defined as:
P(Observing a more extreme value than the test statistic assuming $\mathbf{H_0}$ is true)
= P(Z > 2.309) = 0.01044

Step 6: Compare the p-value with α, accordingly reject or not reject the null hypothesis

As p-value of 0.01044 is less than the significance level of 0.05, we reject the null hypothesis with 95% confidence.

Step 7: Interpret the result of the test

The interpretation of the result of this test is that with 95% confidence, Sophie can infer that the average monthly return for the scheme will be higher than $180, and thus has a very strong case for investing in the scheme.