Hypothesis testing

Example:

Jeffrey, as an eight-year-old, established a mean time of 16.43 seconds for swimming the 25-yard freestyle, with a standard deviation of 0.8 seconds. His dad, Frank, thought that Jeffrey could swim the 25-yard freestyle faster using goggles. Frank bought Jeffrey a new pair of expensive goggles and timed Jeffrey for 15, 25-yard freestyle swims. For the 15 swims, Jeffrey's mean time was 16 seconds. **Frank thought that the goggles helped Jeffrey to swim faster than the 16.43 seconds**.

Hypothesis testing is a statistical procedure or form of statistical inference that uses data from a sample to draw conclusions about a population parameter. **OR**

Hypothesis testing is a systematic way to test claims or ideas about a group or population.

Steps of Hypothesis Testing

1. Setting null and alternative hypothesis
2. Setting the level of significance
3. Calculating the Test Statistics
4. Calculating probability value (*p*-value), or finding the rejection region
5. Making a decision about the null hypothesis
6. Stating an overall conclusion

Step 1 Set the null and alternative hypothesis:

Each hypothesis test includes two hypotheses about the population. One is the null hypothesis and is denoted by $\boldsymbol{H_0}$, which is a statement of a particular parameter value. This hypothesis is assumed to be true until there is evidence to suggest otherwise.

The second hypothesis is called the alternative hypothesis. An **alternative hypothesis** (denoted $\boldsymbol{H_1}$**)**, is the opposite of what is stated in the null hypothesis. The alternative hypothesis is a statement of a range of alternative values in which the parameter may fall.

Example:

We want to test whether the mean GPA of students in UIU is different from 2.9 (out of 4.0). The null and alternative hypotheses are:

$H_0 : \mu = 2.9$

$H_1 : \mu \neq 2.9$ $H_1 : \mu < 2.9$ $H_1 : \mu > 2.9$ based on wording the question

In our Case (Starting Example)

$H_0 : \mu = 16.43$ $H_1 : \mu < 16.43$

Step 2 Setting the level of significance

The level of significance is defined as the fixed probability of wrong elimination of null hypothesis when in fact, it is true. The level of significance is preset by the researcher with the outcomes of error. The significance level is denoted by the Greek letter **alpha – α**. This value is used as a probability cutoff for making decisions about the null hypothesis. This alpha value represents the probability we are willing to place on our test for making an incorrect decision regarding rejecting the null hypothesis.

Level of significance measure of the strength of the evidence that must be present in your sample before you will reject the null hypothesis and conclude that the effect is statistically significant.

A 5% level of significance means the chance that you will accept your alternative hypothesis when your null hypothesis is true. The smaller the significance level, the greater the burden of proof needed to reject the null hypothesis, or in other words, to support the alternative hypothesis.

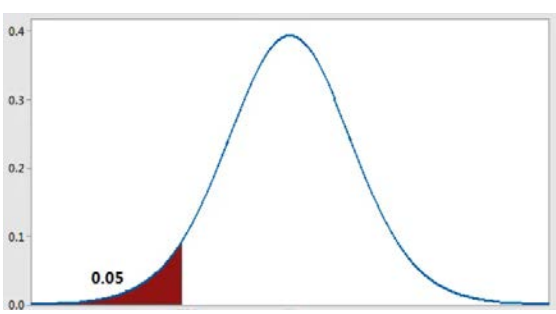
Rejection region: The rejection region is the values of test statistic for which the null hypothesis is rejected.

Acceptance region: The set of all possible values for which the null hypothesis is not rejected is called the rejection region.

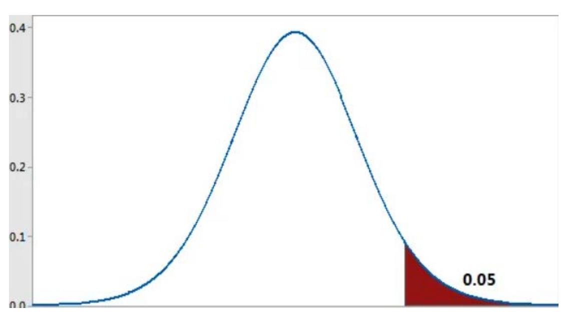
One-tailed hypothesis tests

One-tailed hypothesis tests are also known as directional and one-sided tests because you can test for effects in only one direction

The rejection region for one-tailed test is shown below:



$H_0 : \mu = 16.43$
$H_1 : \mu < 16.43$

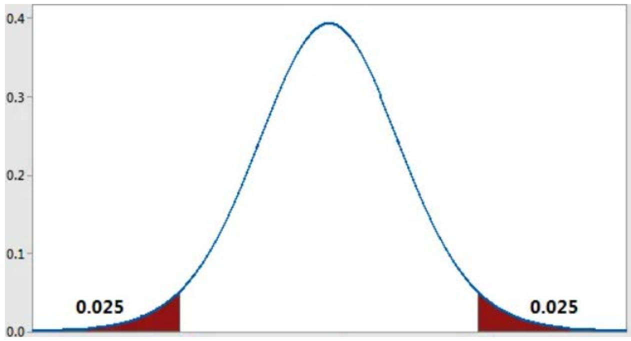


$H_0 : \mu = 16.43$
$H_1 : \mu > 16.43$

Two-tailed hypothesis tests

Two-tailed hypothesis tests are also known as nondirectional and two-sided tests because you can test for effects in both directions.

The rejection region for **two-tailed** test is shown below:



$$H_0 : \mu = 16.43$$
$$H_1 : \mu \neq 16.43$$

Step 3 **Calculating the Test-statistic**

$$p(x < 16) = p\left(z < \frac{16 - 16.43}{0.8}\right) = p(z < -0.5375)$$

$$= 1 - \varphi(0.5375)$$

$$= 1 - 0.7054 = 0.2946$$

Step 4 **Drawing Conclusion**

There exits insufficient evidence to reject the null hypothesis. *i.e.*

Use of goggles has no effect on **Jeffrey** to swim faster.

Type I error and type II error

In hypothesis testing, Type I and Type II errors refer to the two types of mistakes that can occur when making decisions based on sample data.

1. Type I Error (False Positive) : A Type I error occurs when the null hypothesis (H_0) is rejected even though it is actually true. In other words, it is the error of concluding that there is an effect or difference when, in reality, none exists.

Example: In a clinical trial testing the effectiveness of a new drug, a Type I error would occur if the researchers conclude that the drug works reject (H_0) when, in fact, it does not (the null hypothesis that the drug has no effect is true).

Significance Level (α): The probability of making a Type I error is denoted by α, which is the threshold set by the researcher (commonly 0.05 or 5%). This means that there's a 5% risk of rejecting the null hypothesis when it is true. Lowering α reduces the likelihood of making a Type I error, but it increases the risk of a Type II error.

2. Type II Error (False Negative)

A Type II error occurs when the null hypothesis (H_0) is not rejected, even though it is false. In other words, it is the error of failing to detect an effect or difference when one actually exists.

Example: In the same clinical trial, a Type II error would occur if the researchers conclude that the drug does not work (fail to reject (H_0) when, in fact, it does work (the alternative hypothesis H_1 is true).

Probability of Type II Error (β): The probability of making a Type II error is denoted by β. This error is related to the power of the test, where power is defined as 1−β. A higher power indicates a lower chance of making a Type II error. To reduce β, researchers often increase the sample size or adjust the study design.

	Null Hypothesis is TRUE	Null Hypothesis is FALSE
Reject null hypothesis	Type I Error (False positive)	Correct Outcome! (True positive)
Fail to reject null hypothesis	Correct Outcome! (True negative)	Type II Error (False negative)

Q1

The masses of cereal boxes filled by a certain machine have mean 510 grams. An adjustment is made to the machine and an inspector wishes to test whether the mean mass of cereal boxes filled by the machine has decreased.

After the adjustment is made, he chooses a random sample of 120 cereal boxes. The mean mass of these boxes is found to be 508 grams.

Assume that the standard deviation of the masses is 10 grams.

(a) Test at the 2.5% significance level whether the mean mass of cereal boxes filled by the machine has decreased. [5]

Q2

At an election in Menham last year, 24% of voters supported the Today Party. A student wishes to test whether support for the Today Party has decreased since last year. He chooses a random sample of 25 voters in Menham and finds that exactly 2 of them say that they support the Today Party.

Test at the 5% significance level whether support for the Today Party has decreased. [5]

Q3

Every July, as part of a research project, Rita collects data about sightings of a particular kind of bird. Each day in July she notes whether she sees this kind of bird or not, and she records the number X of days on which she sees it. She models the distribution of X by $B(31, p)$, where p is the probability of seeing this kind of bird on a randomly chosen day in July.

Data from previous years suggests that $p = 0.3$, but in 2022 Rita suspected that the value of p had been reduced. She decided to carry out a hypothesis test.

In July 2022, she saw this kind of bird on 4 days.

(a) Use the binomial distribution to test at the 5% significance level whether Rita's suspicion is justified. [5]

Q4

In the past the number of enquiries per minute at a customer service desk has been modelled by a random variable with distribution Po(0.31). Following a change in the position of the desk, it is expected that the mean number of enquiries per minute will increase. In order to test whether this is the case, the total number of enquiries during a randomly chosen 5-minute period is noted. You should assume that a Poisson model is still appropriate.

Given that the total number of enquiries is 5, carry out the test at the 2.5% significance level. [5]