



Research Methods & Technical Writing

Lesson 8 - Week 8

Testing of Hypothesis

Lecturer: Dr. Msagha J Mbogholi, PhD

Flashback from Lesson 7 (1 of 2)

- Parameter implies a summary description of the characteristics of the target population. On the other extreme, the statistic is a summary value of a small group of population i.e. sample.
- Precision is the range within which the population average (or other parameter) will lie in accordance with the reliability specified in the confidence level as a percentage of the estimate \pm or as a numerical quantity.
- The confidence level or reliability is the expected percentage of times that the actual value will fall within the stated precision limits.
- Sampling distribution is a statistic that determines the probability of an event based on data from a small group within a large population: there is sampling distribution of the mean, sampling distribution of proportion, student's t-distribution, F distribution, and chi-square distribution. There is also Sandler's A-test which borrows from t-distribution.

Flashback from Lesson 7 (2 of 2)

- The central limit theorem states that irrespective of a random variable's distribution if large enough samples are drawn from the population then the sampling distribution of the mean for that random variable will approximate a normal distribution. This fact holds true for samples that are greater than or equal to 30.
- The standard deviation of sampling distribution of a statistic is known as its standard error (S.E).
- The sample mean \bar{x} is the best estimator of the population mean, μ , and its sampling distribution, so long as the sample is sufficiently large, approximates the normal distribution.
- There are generally two ways by which we can determine the size of the sample: by using standard deviation together with confidence level, and based on Bayesian statistics

Content

- Introduction
- Basic concepts
- Tests of hypotheses



Part 1

Introduction

Introduction

- In our daily lives we are always testing and comparing, consciously as well as sub-consciously.
- These techniques enable us to make choices; which airline to use to fly from point A to point B, which bus, which restaurant, which clothes, and so on.
- Our choices are guided by parameters that are recognized and accepted by our brains as being suitable for the task at hand, so to speak.
- For example marketing by certain companies play a big role in the choices we make; a billboard showing a MacDonald's burger makes you feel hungry and guess where you'll go eat that burger?
- Research is also about finding out....finding out whether choice A is better than choice B, whether choice A is better than any other choice, and so on. This is what this lesson is all about; how researchers can ask questions and use scientifically accepted methods to find the answers.

Introduction (cont'd)

- In research you are trying to find out something; whether it is by exploration or investigation. You set out to find answers using a research methodology; once you have collected the data then you need to find out what it means. This is where the term inferential statistics come in; you make inferences based on the analysis of your data using known tests.
- In this lesson we give the name to the “idea” or theory that you wish to investigate, and share the tests that will prove/disprove your idea/theory.
- Hypothesis and testing thereof plays an anchor role in the field of research, and it is important to know which test to apply to which situation based on the available information (data).
- In this lesson we shall examine the tests and give simple examples where applicable to help the learner understand the theory, and also to see practical implementation of the tests.



Part 2

Basic concepts

2.1 Definitions

- Hypothesis: A hypothesis is a potential explanation for something that happens or that you observe and think to be true. (Herrity, 2023). Other literature define it similarly; for example,
- “A hypothesis (plural hypotheses) is a precise, testable statement of what the researcher(s) predict will be the outcome of the study. It is stated at the start of the study.” (Mcleod, 2018)
- “A hypothesis is a tentative statement about the relationship between two or more variables. It is a specific, testable prediction about what you expect to happen in a study.” (Cherry, 2019)
- Based on the above definitions we pick out 3 words that stand out: potential explanation, testable statement, and tentative statement. This tells us that a hypothesis is an attempt to explain some phenomena using a testable statement.
- Kothari (2004) states that “a hypothesis may be defined as a proposition or a set of proposition set forth as an explanation for the occurrence of some specified group of phenomena either asserted merely as a provisional conjecture to guide some investigation or accepted as highly probable in the light of established facts.”

2.2 Definitions (cont'd)

- Every research hypothesis must be testable using some accepted scientific method; therefore, the wording of the hypothesis is equally important. It includes a dependent and independent variable and a proposal of the relationship between them. Examples of hypotheses:
- Students who eat breakfast will perform better on a math exam than students who do not eat breakfast. (Cherry, 2019)
- Students who receive counselling will show a greater increase in creativity than students not receiving counselling. (Kothari, 2004)
- If children eat a lot of sugar, then they will develop cavities in their teeth.
- Based on the above and as Kothari (2004) rightly describes, “we may conclude that a hypothesis states what we are looking for and it is a proposition which can be put to a test to determine its validity.”

2.3 Characteristics

- According to Shamil (2023) the characteristics of an effective research hypothesis are as follows:”
 - It must be logical.
 - It must be simple and clear.
 - It needs to be precise.
 - It must identify the research objectives.
 - It must be empirically testable with experimentation and research.
 - It must be manageable.
 - It must be relevant and specific to the theme of the research.
 - It must be predictable.
 - It must be falsifiable.
 - It must be neither specific nor general.
 - It must be considered valuable even if it proves false.”

2.4 Concepts

- Types of hypotheses: Between (Cherry, 2019) and (Herrity, 2023) we identify the following types of hypothesis:
- **Simple hypothesis:** This type of hypothesis suggests that there is a relationship between one independent variable and one dependent variable.
- **Complex hypothesis:** A complex hypothesis looks at the relationship between two or more independent variables and two or more dependent variables.
- **Empirical hypothesis:** An empirical hypothesis, also called a working hypothesis, is one that professionals accept as a basis for future research in order to formulate a theory for testing.
- **Statistical hypothesis:** This hypothesis uses statistical analysis to evaluate a representative sample of the population and then generalizes the findings to the larger group.

2.4 Concepts (cont'd)

- **Logical hypothesis:** A logical hypothesis offers an explanation without extensive evidence.
- **Null hypothesis:** A null hypothesis is the default position that assumes variables have no relation to each other.
- **Alternative hypothesis:** An alternative hypothesis is created to disprove a null hypothesis and adapts its method and prediction according to its results.
- In research we mostly use the null and alternative approach, denoted H_0 and H_a (or H_1) in most literature. Kothari (2004) demonstrates this using a simple example:
- If we are to compare method A with method B about its superiority and if we proceed on the assumption that both methods are equally good, then this assumption is termed as the null hypothesis. As against this, we may think that the method A is superior or the method B is inferior, we are then stating what is termed as alternative hypothesis. The null hypothesis is generally symbolized as H_0 and the alternative hypothesis as H_a . Suppose we want to test the hypothesis that the population mean μ is equal to the hypothesized mean (μ_{H_0}) = 100. Then we would say that the null hypothesis is that the population mean is equal to the hypothesized mean 100 and symbolically we can express as: $H_0 : \mu = \mu_{H_0} = 100$

2.4 Concepts (cont'd)

- If our sample results do not support this null hypothesis, we should conclude that something else is true. What we conclude rejecting the null hypothesis is known as alternative hypothesis. In other words, the set of alternatives to the null hypothesis is referred to as the alternative hypothesis. (Kothari, 2004). In our example the alternatives are that population mean might not be equal to, might be greater than, or less than, the hypothesized mean; these are represented using the appropriate notation (replace the equal sign with the $<$, $>$ or $<>$ as the case may be).
- Other important aspects of the null hypothesis worth noting are (Kothari, 2004):
- Alternative hypothesis is usually the one which one wishes to prove and the null hypothesis is the one which one wishes to disprove. Thus, a null hypothesis represents the hypothesis we are trying to reject, and alternative hypothesis represents all other possibilities.
- If the rejection of a certain hypothesis when it is actually true involves great risk, it is taken as null hypothesis because then the probability of rejecting it when it is true is α (the level of significance) which is chosen very small.
- Null hypothesis should always be specific hypothesis i.e., it should not state about or approximately a certain value.

2.4 Concepts (cont'd)

- Significance level: The significance level, also known as alpha or α , is a 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. The researcher determines the significance level before conducting the experiment. The significance level is the probability of rejecting the null hypothesis when it is true. For example, a significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference. Lower significance levels indicate that you require stronger evidence before you will reject the null hypothesis. (Frost, 2019).
- Decision rule or test of hypothesis: “The decision rule is a statement that tells under what circumstances to reject the null hypothesis. The decision rule is based on specific values of the test statistic (e.g., reject H_0 if $Z \geq 1.645$). The decision rule for a specific test depends on 3 factors: the research or alternative hypothesis, the test statistic and the level of significance. (LaMorte, 2017).

2.4 Concepts (cont'd)

- Type I and Type II errors: Type I error refers to the situation where we incorrectly reject H_0 when in fact it is true. This is also called a false positive result (as we incorrectly conclude that the research hypothesis is true when in fact it is not). When we run a test of hypothesis and decide to reject H_0 (e.g., because the test statistic exceeds the critical value in an upper tailed test) then either we make a correct decision because the research hypothesis is true or we commit a Type I error. When we run a test of hypothesis and decide not to reject H_0 (e.g., because the test statistic is below the critical value in an upper tailed test) then either we make a correct decision because the null hypothesis is true or we commit a Type II error. Table 1 shows these errors in relation to the null hypothesis. Note that we will never know whether the null hypothesis is really true or false (i.e., we will never know which row of the following table reflects reality).

2.4 Concepts (cont'd)

- In the first step of the hypothesis test, we select a level of significance, α , and $\alpha = P(\text{Type I error})$. Because we purposely select a small value for α , we control the probability of committing a Type I error. For example, if we select $\alpha = 0.05$, and our test tells us to reject H_0 , then there is a 5% probability that we commit a Type I error. Most investigators are very comfortable with this and are confident when rejecting H_0 that the research hypothesis is true (as it is the more likely scenario when we reject H_0).
- Beta (β) represents the probability of a Type II error and is defined as follows: $\beta = P(\text{Type II error}) = P(\text{Do not Reject } H_0 \mid H_0 \text{ is false})$. Unfortunately, we cannot choose β to be small (e.g., 0.05) to control the probability of committing a Type II error because β depends on several factors including the sample size, α , and the research hypothesis. When we do not reject H_0 , it may be very likely that we are committing a Type II error (i.e., failing to reject H_0 when in fact it is false). Therefore, when tests are run and the null hypothesis is not rejected we often make a weak concluding statement allowing for the possibility that we might be committing a Type II error. If we do not reject H_0 , we conclude that we do not have significant evidence to show that H_1 is true. We do not conclude that H_0 is true". (LaMorte, 2017).

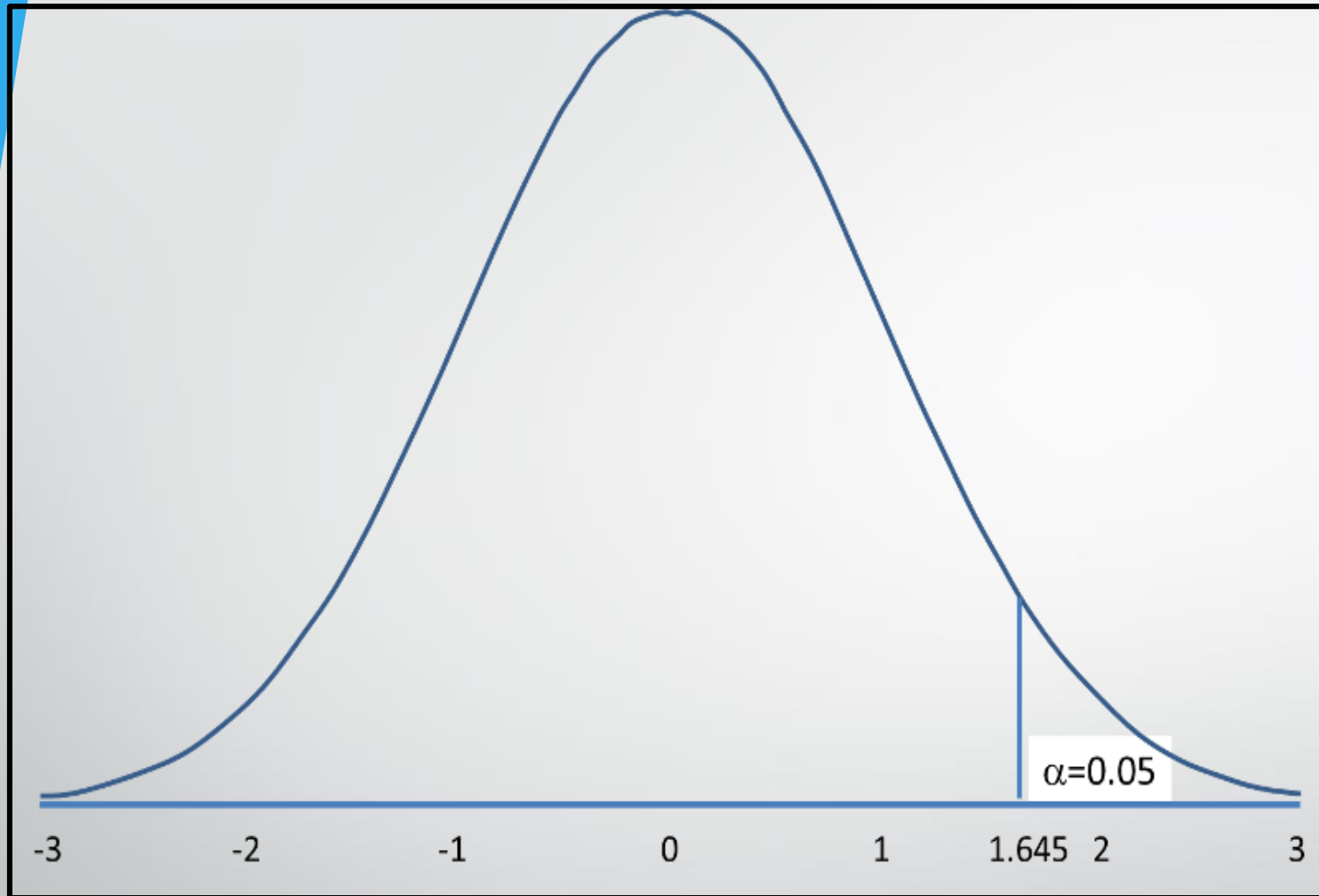
2.4 Concepts (cont'd)

Table 1. Conclusions in tests of hypothesis (LaMorte, 2017)

	Do Not Reject H_0	Reject H_0
H_0 is True	Correct Decision	Type I Error α error
H_0 is False	Type II Error β error	Correct Decision

2.4 Concepts (cont'd)

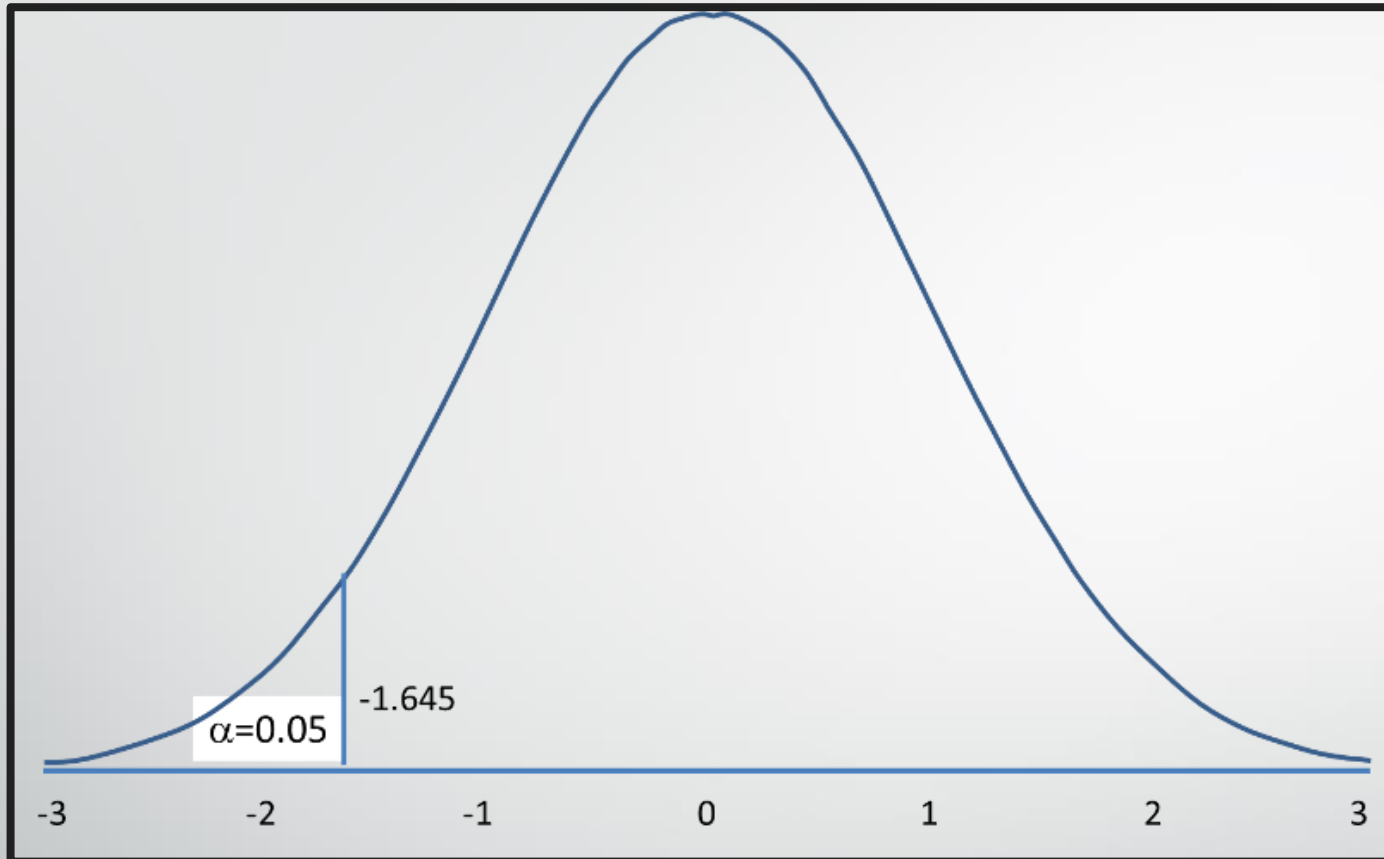
- Upper-tailed, Lower-tailed, Two-tailed Tests: LaMorte (2017) describes these three tests as follows: “The research or alternative hypothesis can take one of three forms. An investigator might believe that the parameter has increased, decreased or changed. For example, an investigator might hypothesize:
- $H_1: \mu > \mu_0$, where μ_0 is the comparator or null value (e.g., $\mu_0 = 191$...) and an increase is hypothesized - this type of test is called an **upper-tailed test**;
- $H_1: \mu < \mu_0$, where a decrease is hypothesized and this is called a **lower-tailed test**;
or
- $H_1: \mu \neq \mu_0$, where a difference is hypothesized and this is called a **two-tailed test**. ”
- Fig 1, 2 and 3 show rejection regions defined by the decision rule for upper-, lower- and two-tailed Z tests with $\alpha=0.05$. Notice that the rejection regions are in the upper, lower and both tails of the curves, respectively. The decision rules are written below each figure.



Upper-Tailed Test	
α	Z
0.10	1.282
0.05	1.645
0.025	1.960
0.010	2.326
0.005	2.576
0.001	3.090
0.0001	3.719

Rejection Region for Upper-Tailed Z Test ($H_1: \mu > \mu_0$) with $\alpha=0.05$
 The decision rule is: Reject H_0 if $Z \geq 1.645$.

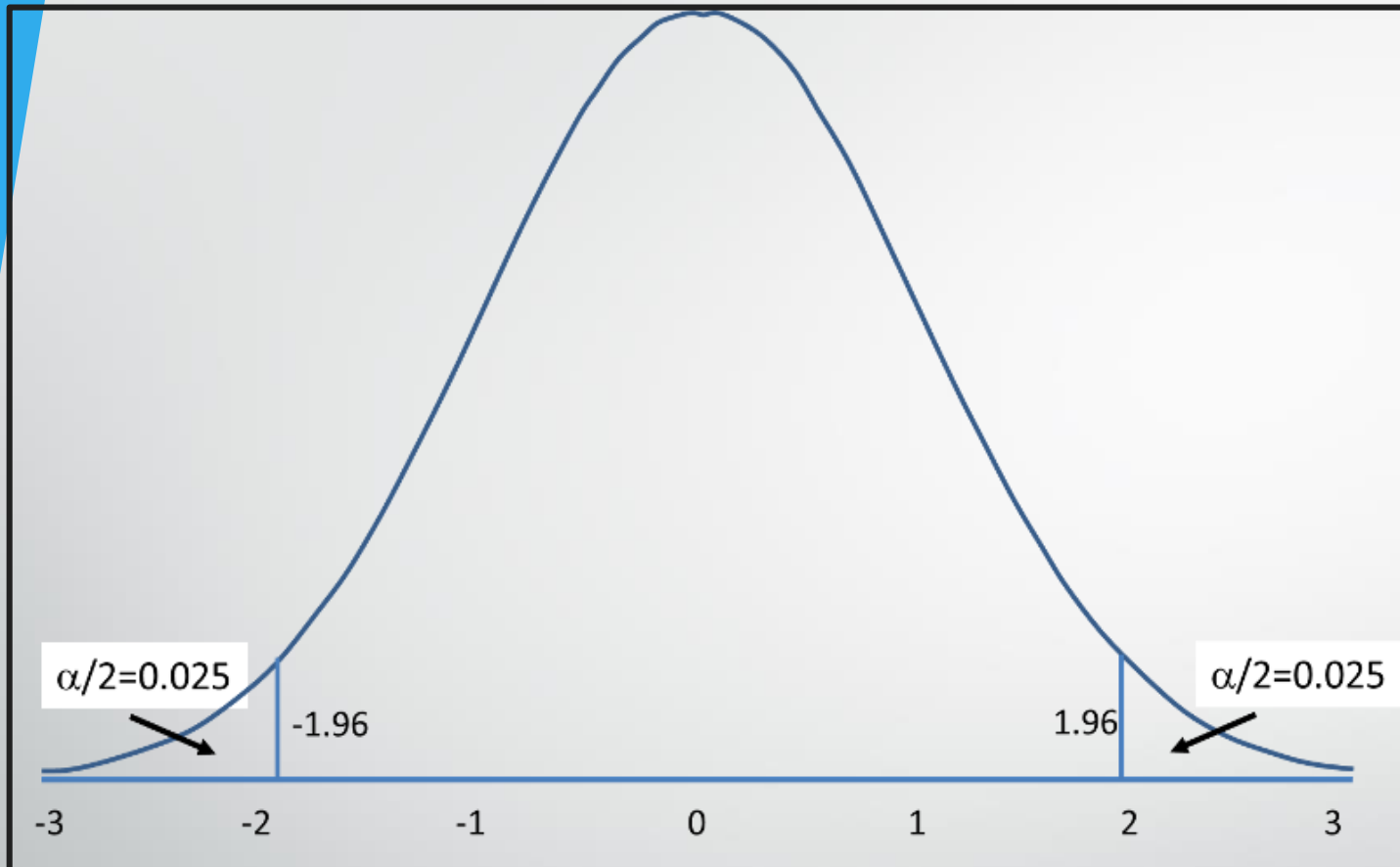
Fig 1. Upper tailed test (LaMorte, 2017)



Lower-Tailed Test	
α	Z
0.10	-1.282
0.05	-1.645
0.025	-1.960
0.010	-2.326
0.005	-2.576
0.001	-3.090
0.0001	-3.719

Rejection Region for Lower-Tailed Z Test ($H_1: \mu < \mu_0$) with $\alpha = 0.05$
 The decision rule is: Reject H_0 if $Z \leq -1.645$.

Fig 2. Lower tailed test (LaMorte, 2017)



Two -Tailed Test	
α	Z
0.10	1.282
0.05	1.645
0.025	1.960
0.010	2.326
0.005	2.576
0.001	3.090
0.0001	3.719

Rejection Region for Two-Tailed Z Test ($H_1: \mu \neq \mu_0$) with $\alpha = 0.05$
 The decision rule is: Reject H_0 if $Z \leq -1.960$ or if $Z \geq 1.960$.

Fig 3. Two-tailed test (LaMorte, 2017)

2.5 Approach to hypothesis testing

- Kothari (2004) suggests a six step approach to hypothesis testing as follows:
- Make a formal statement – declare the null hypothesis and its alternative(s).
- Select significance level – most researchers use 5% , but there are a variety of factors that affect this choice.
- Decide distribution to use – choose the appropriate sampling distribution, for example, normal or t- distribution.
- Select a random sample and compute an appropriate value – concerning the test statistic.
- Calculate the probability - that the sample result would diverge as widely as it has from expectations, if the null hypothesis were in fact true. Compare the probability - with the specified value for α , the significance level. If the calculated probability is equal to or smaller than the α value in case of one-tailed test (and $\alpha / 2$ in case of two-tailed test), then reject the null hypothesis (i.e., accept the alternative hypothesis), but if the calculated probability is greater, then accept the null hypothesis. In case we reject H_0 , we run a risk of (at most the level of significance) committing an error of Type I, but if we accept H_0 , then we run some risk (the size₂₃ of which cannot be specified as long as the H_0 happens to be vague rather than specific) of committing an error of Type II.

2.5 Approach to hypothesis testing (cont'd)

- LaMorte (2017) suggests a 5 step approach to hypothesis testing as follows:
- **Step 1.** Set up hypotheses and determine level of significance
- **Step 2.** Select the appropriate test statistic.
- **Step 3.** Set up decision rule.
- **Step 4.** Compute the test statistic.
- **Step 5.** Conclusion. (draw an inference -If the null hypothesis is rejected, then an exact significance level is computed to describe the likelihood of observing the sample data assuming that the null hypothesis is true. The exact level of significance is called the p-value and it will be less than the chosen level of significance if we reject H_0 .)
- We shall illustrate this using an example in the next part of this lesson.

2.6 Power of hypothesis

- Recall that a Type I Error is rejecting the null hypothesis in favor of a false alternative hypothesis, and a Type II Error is failing to reject a false null hypothesis in favor of a true alternative hypothesis. The probability of a Type I error is typically known as Alpha, while the probability of a Type II error is typically known as Beta.
- (Walmsley & Brown, 2017) describe multiple ways of understanding power:
- Power is the probability of rejecting the null hypothesis when, in fact, it is false.
- Power is the probability of making a correct decision (to reject the null hypothesis) when the null hypothesis is false.
- Power is the probability that a test of significance will pick up on an effect that is present.
- Power is the probability that a test of significance will detect a deviation from the null hypothesis, should such a deviation exist.
- Power is the probability of avoiding a Type II error.
- They further add that mathematically, power is $1 - \beta$. The power of a hypothesis test is between 0 and 1; if the power is close to 1, the hypothesis test is very good at detecting a false null hypothesis. Beta is commonly set at 0.2, but may be set by the researchers to be smaller.

2.6 Power of hypothesis

- Examples of calculations on power can be found at:
- <https://online.stat.psu.edu/stat415/lesson/25/25.2>
- <https://bookdown.org/jkang37/stat205b-notes/lecture12.html>
- Due to time and space considerations the learner is encouraged to go through these links at the end of the lesson.



Part 3

Tests of hypothesis

3.1 The 5-step approach

- In the previous part of this section we outlined the 5 step approach to hypothesis testing outlined by LaMorte (2017). Let us demonstrate this using an example in that literature (LaMorte, 2017):
- Problem: Use the five-step procedure to test the research hypothesis that the mean weight in men in 2006 is more than 191 pounds. We will assume the sample data are as follows: $n = 100$, $\bar{x} = 197.1$, $s = 25.6$
- Solution:
- Step 1: Set up hypotheses and determine level of significance
- $H_0: \mu = 191$ $H_1: \mu > 191$ $\alpha = 0.05$ (5% significance)
- The research hypothesis is that weights have increased, and therefore an upper tailed test is used.

3.1 The 5-step approach

- **Step 2.** Select the appropriate test statistic.
- Because the sample size is large ($n \geq 30$) the appropriate test statistic is
- $Z = (\bar{x} - \mu_0) / (s/\sqrt{n})$
- **Step 3.** Set up decision rule.
- In this example, we are performing an upper tailed test ($H_1: \mu > 191$), with a Z test statistic and selected $\alpha = 0.05$ Reject H_0 if $Z \geq 1.645$.

3.1 The 5-step approach

- **Step 4.** Compute the test statistic.
- We now substitute the sample data into the formula for the test statistic identified in Step 2 (the mean is 197.1 not 195.3; substitute and we still get the same answer.
- $Z = (\bar{x} - \mu_0) / (s/\sqrt{n}) = (195.3 - 191) / (25.6 / \sqrt{100}) = 2.38$

3.1 The 5-step approach

- **Step 5.** Conclusion.
- We reject H_0 because $2.38 \geq 1.645$. We have statistically significant evidence at $\alpha = 0.05$, to show that the mean weight in men in 2006 is more than 191 pounds. Because we rejected the null hypothesis, we now approximate the p-value which is the likelihood of observing the sample data if the null hypothesis is true. An alternative definition of the p-value is the smallest level of significance where we can still reject H_0 . In this example, we observed $Z = 2.38$ and for $\alpha = 0.05$, the critical value was 1.645. Because 2.38 exceeded 1.645 we rejected H_0 . In our conclusion we reported a statistically significant increase in mean weight at a 5% level of significance. Using the table of critical values for upper tailed tests, we can approximate the p-value. If we select $\alpha = 0.025$, the critical value is 1.96, and we still reject H_0 because $2.38 \geq 1.960$. If we select $\alpha = 0.010$ the critical value is 2.326, and we still reject H_0 because $2.38 \geq 2.326$. However, if we select $\alpha = 0.005$, the critical value is 2.576, and we cannot reject H_0 because $2.38 < 2.576$. Therefore, the smallest α where we still reject H_0 is 0.010. This is the p-value. A statistical computing package would produce a more precise p-value which would be in between 0.005 and 0.010. Here we are approximating the p-value and would report $p < 0.010$.

3.2 Example 2: Hypothesis testing of means

- This example is taken from Kothari (2004):
- Suppose we are interested in a population of 20 industrial units of the same size, all of which are experiencing excessive labor turnover problems. The past records show that the mean of the distribution of annual turnover is 320 employees, with a standard deviation of 75 employees. A sample of 5 of these industrial units is taken at random which gives a mean of annual turnover as 300 employees. Is the sample mean consistent with the population mean? Test at 5% level.

3.2 Example

Solution: Taking the null hypothesis that the population mean is 320 employees, we can write:

$$H_0: \mu_{H_0} = 320 \text{ employees}$$

$$H_a: \mu_{H_0} \neq 320 \text{ employees}$$

and the given information as under:

$$\bar{X} = 300 \text{ employees, } \sigma_p = 75 \text{ employees}$$

$$n = 5; N = 20$$

Assuming the population to be normal, we can work out the test statistic z as under:

$$\begin{aligned} z^* &= \frac{\bar{X} - \mu_{H_0}}{\sigma_p / \sqrt{n} \times \sqrt{(N - n) / (N - 1)}} \\ &= \frac{300 - 320}{75 / \sqrt{5} \times \sqrt{(20 - 5) / (20 - 1)}} = -\frac{20}{(33.54) (.888)} \\ &= -0.67 \end{aligned}$$

As H_a is two-sided in the given question, we shall apply a two-tailed test for determining the rejection regions at 5% level of significance which comes to as under, using normal curve area table:

$$R: |z| > 1.96$$

The observed value of z is -0.67 which is in the acceptance region since $R: |z| > 1.96$ and thus, H_0 is accepted and we may conclude that the sample mean is consistent with population mean i.e., the population mean 320 is supported by sample results.

3.3 Example 3: Hypothesis testing of difference of means

- We demonstrate this using an example drawn from Kothari (2004):
- The mean produce of wheat of a sample of 100 fields is 200 lbs. per acre with a standard deviation of 10 lbs. Another sample of 150 fields gives the mean of 220 lbs. with a standard deviation of 12 lbs. Can the two samples be considered to have been taken from the same population whose standard deviation is 11 lbs? Use 5 per cent level of significance.

3.3 Example 3: Hypothesis testing of difference of means

Solution: Taking the null hypothesis that the means of two populations do not differ, we can write

$$H_0 : \mu = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

and the given information as $n_1 = 100$; $n_2 = 150$;

$$\bar{X}_1 = 200 \text{ lbs.}; \quad \bar{X}_2 = 220 \text{ lbs.};$$

$$\sigma_{x_1} = 10 \text{ lbs.}; \quad \sigma_{x_2} = 12 \text{ lbs.};$$

and

$$\sigma_p = 11 \text{ lbs.}$$

Assuming the population to be normal, we can work out the test statistic z as under:

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{200 - 220}{\sqrt{(11)^2 \left(\frac{1}{100} + \frac{1}{150} \right)}}$$

3.3 Example 3: Hypothesis testing of difference of means

- Solution (cont'd):

$$= -\frac{20}{1.42} = -14.08$$

As H_0 is two-sided, we shall apply a two-tailed test for determining the rejection regions at 5 per cent level of significance which come to as under, using normal curve area table:

$$R : |z| > 1.96$$

The observed value of z is -14.08 which falls in the rejection region and thus we reject H_0 and conclude that the two samples cannot be considered to have been taken at 5 per cent level of significance from the same population whose standard deviation is 11 lbs. This means that the difference between means of two samples is statistically significant and not due to sampling fluctuations.

3.4 Example: Hypothesis testing for comparing two related samples

- The paired t-test, (and A-test), is used for this purpose: we demonstrate this using an example from Kothari (2004):

The sales data of an item in six shops before and after a special promotional campaign are:

<i>Shops</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>
<i>Before the promotional campaign</i>	53	28	31	48	50	42
<i>After the campaign</i>	58	29	30	55	56	45

Can the campaign be judged to be a success? Test at 5 per cent level of significance. Use paired *t*-test as well as *A*-test.

3.4 Example: Hypothesis testing for comparing two related samples

Solution: Let the sales before campaign be represented as X and the sales after campaign as Y and then taking the null hypothesis that campaign does not bring any improvement in sales, we can write:

$$H_0 : \mu_1 = \mu_2 \text{ which is equivalent to test } H_0 : \bar{D} = 0$$

$$H_a : \mu_1 < \mu_2 \text{ (as we want to conclude that campaign has been a success).}$$

Because of the matched pairs we use paired t -test and work out the test statistic ' t ' as under:

$$t = \frac{\bar{D} - 0}{\sigma_{diff.}/\sqrt{n}}$$

To find the value of t , we first work out the mean and standard deviation of differences as under:

Table 9.9

<i>Shops</i>	<i>Sales before campaign</i> X_i	<i>Sales after campaign</i> Y_i	<i>Difference</i> $(D_i = X_i - Y_i)$	<i>Difference squared</i> D_i^2
A	53	58	-5	25
B	28	29	-1	1
C	31	30	1	1
D	48	55	-7	49
E	50	56	-6	36
F	42	45	-3	9
$n = 6$			$\Sigma D_i = -21$	$\Sigma D_i^2 = 121$

3.4 Example: Hypothesis testing for comparing two related samples

- Solution (cont'd):

$$\therefore \bar{D} = \frac{\sum D_i}{n} = -\frac{21}{6} = -3.5$$

$$\sigma_{diff.} = \sqrt{\frac{\sum D_i^2 - (\bar{D})^2 \cdot n}{n-1}} = \sqrt{\frac{121 - (-3.5)^2 \times 6}{6-1}} = 3.08$$

Hence,

$$t = \frac{-3.5 - 0}{3.08/\sqrt{6}} = \frac{-3.5}{1.257} = -2.784$$

Degrees of freedom = $(n - 1) = 6 - 1 = 5$

As H_a is one-sided, we shall apply a one-tailed test (in the left tail because H_a is of less than type) for determining the rejection region at 5 per cent level of significance which come to as under, using table of t -distribution for 5 degrees of freedom:

$$R : t < -2.015$$

The observed value of t is -2.784 which falls in the rejection region and thus, we reject H_0 at 5 per cent level and conclude that sales promotional campaign has been a success.

3.5 Hypothesis testing for difference between proportions

- We illustrate how to test for difference between proportions using an example from Kothari (2004):

A drug research experimental unit is testing two drugs newly developed to reduce blood pressure levels. The drugs are administered to two different sets of animals. In group one, 350 of 600 animals tested respond to drug one and in group two, 260 of 500 animals tested respond to drug two. The research unit wants to test whether there is a difference between the efficacy of the said two drugs at 5 per cent level of significance. How will you deal with this problem?

3.5 Hypothesis testing for difference between proportions

Solution: We take the null hypothesis that there is no difference between the two drugs i.e.,

$$H_0: \hat{p}_1 = \hat{p}_2$$

The alternative hypothesis can be taken as that there is a difference between the drugs i.e.,

$H_a: \hat{p}_1 \neq \hat{p}_2$ and the given information can be stated as:

$$\hat{p}_1 = 350/600 = 0.583$$

$$\hat{q}_1 = 1 - \hat{p}_1 = 0.417$$

$$n_1 = 600$$

$$\hat{p}_2 = 260/500 = 0.520$$

$$\hat{q}_2 = 1 - \hat{p}_2 = 0.480$$

$$n_2 = 500$$

We can work out the test statistic z thus:

$$\begin{aligned} z &= \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} = \frac{0.583 - 0.520}{\sqrt{\frac{(.583)(.417)}{600} + \frac{(.520)(.480)}{500}}} \\ &= 2.093 \end{aligned}$$

As H_a is two-sided, we shall determine the rejection regions applying two-tailed test at 5% level which comes as under using normal curve area table:

$$R: |z| > 1.96$$

The observed value of z is 2.093 which is in the rejection region and thus, H_0 is rejected in favour of H_a and as such we conclude that the difference between the efficacy of the two drugs is significant.

3.6 Hypothesis testing for the equality of variances of two populations

- In this instance the F-test is used based on F distribution. Let us illustrate this using an example from Kothari (2004):

Two random samples drawn from two normal populations are:

Sample 1 20 16 26 27 23 22 18 24 25 19

Sample 2 27 33 42 35 32 34 38 28 41 43 30 37

Test using variance ratio at 5 per cent and 1 per cent level of significance whether the two populations have the same variances.

3.6 Hypothesis testing for the equality of variances of two populations

Solution: We take the null hypothesis that the two populations from where the samples have been drawn have the same variances i.e., $H_0: \sigma_{p_1}^2 = \sigma_{p_2}^2$. From the sample data we work out $\sigma_{s_1}^2$ and $\sigma_{s_2}^2$ as under:

Table 9.10

Sample 1			Sample 2		
X_{1i}	$(X_{1i} - \bar{X}_1)$	$(X_{1i} - \bar{X}_1)^2$	X_{2i}	$(X_{2i} - \bar{X}_2)$	$(X_{2i} - \bar{X}_2)^2$
20	-2	4	27	-8	64
16	-6	36	33	-2	4
26	4	16	42	7	49
27	5	25	35	0	0
23	1	1	32	-3	9
22	0	0	34	-1	1
18	-4	16	38	3	9
24	2	4	28	-7	49
25	3	9	41	6	36
19	-3	9	43	8	64
			30	-5	25
			37	2	4
$\Sigma X_{1i} = 220$	$\Sigma (X_{1i} - \bar{X}_1)^2 = 120$		$\Sigma X_{2i} = 420$	$\Sigma (X_{2i} - \bar{X}_2)^2 = 314$	
$n_1 = 10$			$n_2 = 12$		

$$\bar{X}_1 = \frac{\Sigma X_{1i}}{n_1} = \frac{220}{10} = 22; \quad \bar{X}_2 = \frac{\Sigma X_{2i}}{n_2} = \frac{420}{12} = 35$$

$$\therefore \sigma_{s_1}^2 = \frac{\Sigma (X_{1i} - \bar{X}_1)^2}{n_1 - 1} = \frac{120}{10 - 1} = 13.33$$

3.6 Hypothesis testing for the equality of variances of two populations

- Solution: (cont'd)

and

$$\sigma_{s_2}^2 = \frac{\sum (X_{2i} - \bar{X}_2)^2}{n_2 - 1} = \frac{314}{12 - 1} = 28.55$$

Hence,

$$F = \frac{\sigma_{s_2}^2}{\sigma_{s_1}^2} \quad (\because \sigma_{s_2}^2 > \sigma_{s_1}^2)$$
$$= \frac{28.55}{13.33} = 2.14$$

Degrees of freedom in sample 1 = $(n_1 - 1) = 10 - 1 = 9$

Degrees of freedom in sample 2 = $(n_2 - 1) = 12 - 1 = 11$

As the variance of sample 2 is greater variance, hence

$$v_1 = 11; v_2 = 9$$

The table value of F at 5 per cent level of significance for $v_1 = 11$ and $v_2 = 9$ is 3.11 and the table value of F at 1 per cent level of significance for $v_1 = 11$ and $v_2 = 9$ is 5.20.

Since the calculated value of $F = 2.14$ which is less than 3.11 and also less than 5.20, the F ratio is insignificant at 5 per cent as well as at 1 per cent level of significance and as such we accept the null hypothesis and conclude that samples have been drawn from two populations having the same variances.

3.7 Limitations of tests of hypothesis

- Kothari (2004) identifies the following as the limitations associated with hypothesis testing:
“
- The tests should not be used in a mechanical fashion. It should be kept in view that testing is not decision-making itself; the tests are only useful aids for decision-making. Hence “proper interpretation of statistical evidence is important to intelligent decisions.”⁶
- Test do not explain the reasons as to why does the difference exist, say between the means of the two samples. They simply indicate whether the difference is due to fluctuations of sampling or because of other reasons but the tests do not tell us as to which is/are the other reason(s) causing the difference.
- Results of significance tests are based on probabilities and as such cannot be expressed with full certainty. When a test shows that a difference is statistically significant, then it simply suggests that the difference is probably not due to chance.
- Statistical inferences based on the significance tests cannot be said to be entirely correct evidences concerning the truth of the hypotheses. This is specially so in case of small samples where the probability of drawing erring inferences happens to be generally higher. For greater reliability, the size of samples be sufficiently enlarged.
- Most literature are in general agreement with his observations: see <https://nap.nationalacademies.org/read/9686/chapter/7>, <https://www.formpl.us/blog/hypothesis-testing> and https://www.stats.ox.ac.uk/~snijders/Encycl_isb203057.pdf , for example.

3.8 Non-parametric testing

- The tests described so far are parametric, that is, they are based on some kind of distribution. There are also distribution-free tests in research, and those which don't require any sample population for assumptions. These are not often used, and as such we shall mention them here, without going into details (like the parametric ones covered in this lesson). The learner is encouraged to do some further reading on these tests. The common tests as described by (Soni, 2020) are:
- **Wilcoxon Rank Sum Test:** The Wilcoxon test also known as rank sum test or signed rank test. It is a type of non-parametric test that works on two paired groups. The main focus of this test is comparison between two paired groups
- **Mann-Whitney U Test:** The Mann-Whitney U test also known as the Mann-Whitney-Wilcoxon test, Wilcoxon rank sum test and Wilcoxon-Mann-Whitney test. It is a non-parametric test based on null hypothesis. It is equally likely that a randomly selected sample from one sample may have higher value than the other selected sample or maybe less. Mann-Whitney test is usually used to compare the characteristics between two independent groups when the dependent variable is either ordinal or continuous. But these variables shouldn't be normally distributed

3.8 Non-parametric testing (cont'd)

- **Kruskal Wallis Test:** Sometimes referred to as a one way [ANOVA](#) on ranks, Kruskal Wallis H test is a nonparametric test that is used to determine the statistical differences between the two or more groups of an independent variable. The word ANOVA is expanded as Analysis of variance. The test is named after the scientists who discovered it, William Kruskal and W. Allen Wallis. The major purpose of the test is to check if the sample is tested if the sample is taken from the same population or not.
- **Friedman Test:** The Friedman test is similar to the Kruskal Wallis test. It is an alternative to the ANOVA test. The only difference between Friedman test and ANOVA test is that Friedman test works on repeated measures basis. Friedman test is used for creating differences between two groups when the dependent variable is measured in the ordinal.

3.8.1 Distribution free tests

- Distribution free tests are defined as the mathematical procedures. These tests are widely used for testing statistical hypotheses. It makes no assumption about the probability distribution of the variables. An important list of distribution free tests is as follows (Soni, 2020) :
- **Anderson-Darling test:** It is done to check if the sample is drawn from a given distribution or not.
- **Statistical bootstrap methods:** It is a basic non-statistical test used to estimate the accuracy and sampling distribution of a statistic.
- **Cochran's Q:** Cochran's Q is used to check constant treatments in block designs with 0/1 outcomes.
- **Cohen's kappa:** Cohen kappa is used to measure the inter-rater agreement for categorical items.

3.8.1 Distribution free tests

- **Kaplan-Meier test:** Kaplan Meier test helps in estimating the survival function from lifetime data, modeling, and censoring.
- **Two-way analysis Friedman test:** Also known as ranking test, it is used to randomize different block designs.
- **Kendall's tau:** The test helps in defining the statistical dependency between two different variables.
- **Kolmogorov-Smirnov test:** The test draws the inference if a sample is taken from the same distribution or if two or more samples are taken from the same sample.
- **Kendall's W:** The test is used to measure the inference of an inter-rater agreements.
- **Kuiper's test:** The test is done to determine if the sample drawn from a given distribution is sensitive to cyclic variations or not.

3.8.1 Distribution free tests

- **Log Rank test:** This test compares the survival distribution of two right-skewed and censored samples.
- **McNemar's test:** It tests the contingency in the sample and revert when the row and column marginal frequencies are equal to or not.
- **Median tests:** As the name suggests, median tests check if the two samples drawn from the similar population have similar median values or not.
- **Pitman's permutation test:** It is a statistical test that yields the value of p variables. This is done by examining all possible rearrangements of labels.
- **Rank products:** Rank products are used to detect expressed genes in replicated microarray experiments.
- **Siegel Tukey tests:** This test is used for differences in scale between two groups.

3.8.1 Distribution free tests

- **Sign test:** Sign test is used to test whether matched pair samples are drawn from distributions from equal medians.
- **Spearman's rank:** It is used to measure the statistical dependence between two variables using a monotonic function.
- **Squared ranks test:** Squared rank test helps in testing the equality of variances between two or more variables.
- **Wald-Wolfowitz runs a test:** This test is done to check if the elements of the sequence are mutually independent or random.

Summary (1 of 2)

- A hypothesis (plural hypotheses) is a precise, testable statement of what the researcher(s) predict will be the outcome of the study. It is stated at the start of the study.
- Types of hypotheses include simple hypothesis, complex hypothesis, empirical hypothesis, statistical hypothesis, logical hypothesis, null hypothesis, and alternative hypothesis.
- Significance level: The significance level, also known as alpha or α , is a 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.
- Decision rule or test of hypothesis: The decision rule is a statement that tells under what circumstances to reject the null hypothesis.

Summary (2 of 2)

- Type I error refers to the situation where we incorrectly reject H_0 when in fact it is true. This is also called a false positive result (as we incorrectly conclude that the research hypothesis is true when in fact it is not). When we run a test of hypothesis and decide not to reject H_0 (e.g., because the test statistic is below the critical value in an upper tailed test) then either we make a correct decision because the null hypothesis is true or we commit a Type II error.
- The 5 step approach to hypothesis testing as follows: **Step 1.** Set up hypotheses and determine level of significance, **Step 2.** Select the appropriate test statistic, **Step 3.** Set up decision rule, **Step 4.** Compute the test statistic, **Step 5.** Conclusion. (draw an inference -If the null hypothesis is rejected, then an exact significance level is computed to describe the likelihood of observing the sample data assuming that the null hypothesis is true. The exact level of significance is called the p-value and it will be less than the chosen level of significance if we reject H_0 .)

References

- Cherry, K. (2019). *How to Form a Hypothesis Statement for Psychology Research*. Verywell Mind. <https://www.verywellmind.com/what-is-a-hypothesis-2795239>
- Frost, J. (2019). *Significance level - Statistics By Jim*. Statistics by Jim. <https://statisticsbyjim.com/glossary/significance-level/>
- Hayes, A. (2022, October 24). *What Is T-Distribution in Probability? How Do You Use It?* Investopedia. <https://www.investopedia.com/terms/t/tdistribution.asp#:~:text=The%20t%2Ddistribution%2C%20also%20known>
- Herrity, J. (2023, February 4). *What Is a Hypothesis? (With Types, Examples and FAQs)*. Indeed.com; Indeed. <https://www.indeed.com/career-advice/career-development/what-is-a-hypothesis#:~:text=Professionals%20typically%20write%20hypotheses%20as,the%20result%20is%20developing%20cavities.>
- Kothari, C. R. (2004). *Research methodology : methods & techniques* (2nd ed.). New Age International (P) Ltd., Publishers, Cop.

References

- LaMorte, W. (2017). *Hypothesis Testing: Upper-, Lower, and Two Tailed Tests*. Sphweb.bumc.bu.edu; Boston University School of Public Health. https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_hypothesistest-means-proportions/bs704_hypothesistest-means-proportions3.html#:~:text=The%20decision%20rule%20is%20a
- Mcleod, S. (2018, August 10). *What is a Hypothesis? | Simply Psychology*. Simplypsychology.org. <https://www.simplypsychology.org/what-is-a-hypotheses.html>
- Shamil, F. (2023). *Research Hypothesis - Types, Examples Characteristics, and Sources*. T4Tutorials.com; T4Tutorials. <https://t4tutorials.com/research-hypothesis/>
- Soni, P. (2020). *Non-Parametric Statistics: Types, Tests, and Examples |*. Wwww.analyticssteps.com. <https://www.analyticssteps.com/blogs/non-parametric-statistics-types-tests-and-examples>
- Walmsley, A., & Brown, M. (2017). *What Is Power?* Statistics Teacher; American Statistical Association. <https://www.statisticsteacher.org/2017/09/15/what-is-power/#:~:text=Power%20is%20the%20probability%20of>