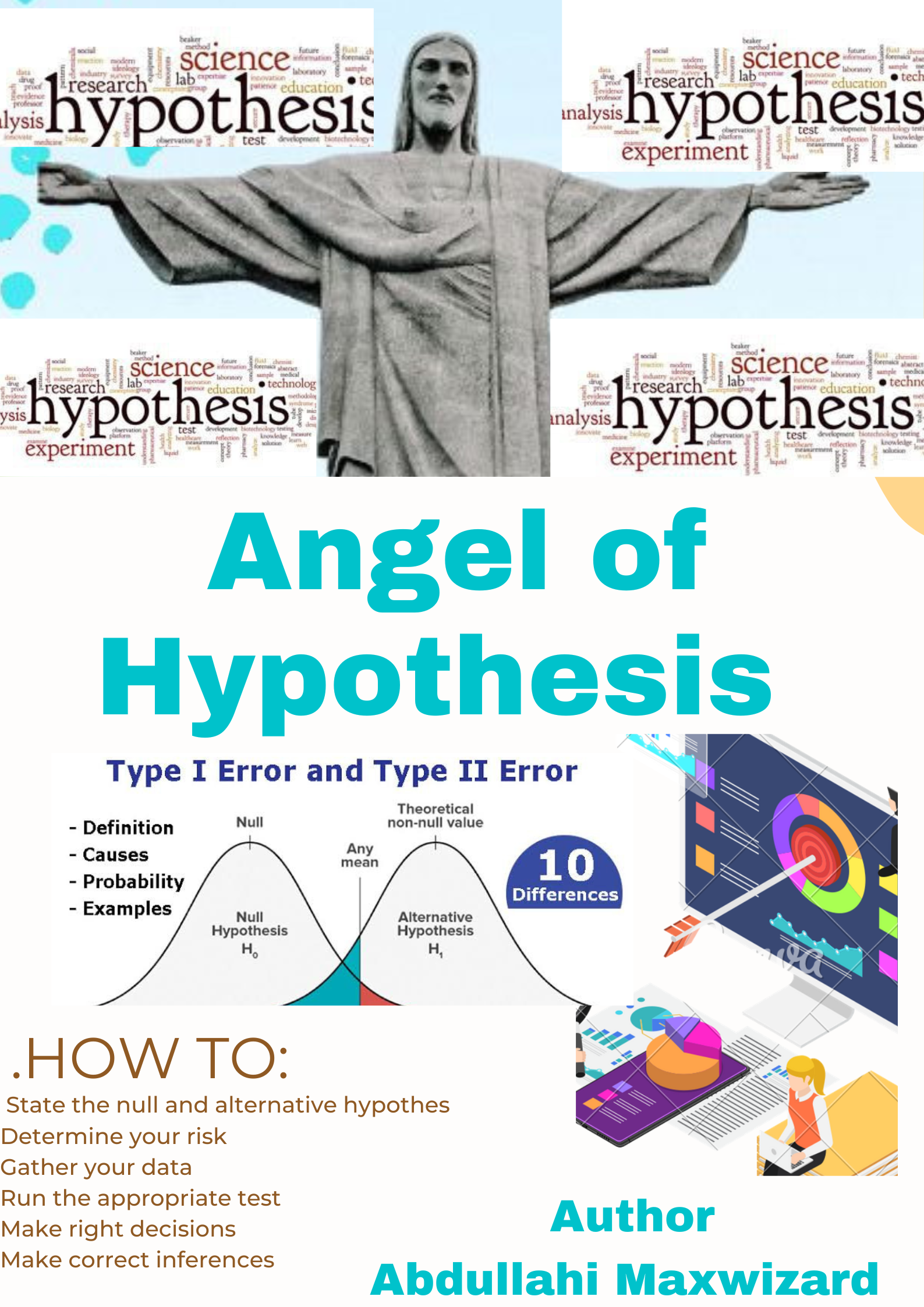
**

Table of Contents

**Introduction to Hypothesis1**

What is hypothesis2

Type chapter title (level 3)3

**Type chapter title (level 1)4**

Type chapter title (level 2)5

Type chapter title (level 3)6

Type chapter title (level 2)2

Type chapter title (level 3)3

**Type chapter title (level 1)4**

Type chapter title (level 2)5

Type chapter title (level 3)

**What is hypothesis???**

Hypothesis can be simply defined as a supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation. In other word a hypothesis is a speculation or theory based on insufficient evidence that lends itself to further testing and experimentation. With further testing, a hypothesis can usually be proven true or false.

Let's look at an example. My friend (James) hypothesizes to me that “the flowers I plants on my land will grow faster than flowers on your land”. She plants on my land and his land and waters each plant daily for a 3month (experiment) and proves her hypothesis true!

Here, the statement made by my friend is called hypothesis.

**What is Hypothesis testing???**  
Hypothesis testing is an act in statistics where by an analyst test an assumption regarding a population parameter. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis.

In the example above, the experiment perform by James is actually what we called hypothesis testing.

Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data. Such data may come from a larger population, or from a data-generating process. The word "population" will be used for both of these cases in the following descriptions.

Hypothesis testing is used to assess the likelihood of a hypothesis by using sample data.

The test provides evidence concerning the likelihood of the hypothesis, given the data.

Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analyzed.

**How Hypothesis Testing Works???**

In hypothesis testing, the statistician tests a statistical sample, with the aims to provide evidence to the likelihood of the null hypothesis. The test is carried out by measuring and examining a random sample of the population being analyzed. All analysts usually used a random population sample to test two different hypotheses: null and the alternative hypothesis.

**1**

**What is null Hypothesis??**

A null hypothesis is a hypothesis that says there is no statistical significance between the two variables in the hypothesis. It is the hypothesis that the researcher is trying to disprove. It is usually denoted by Ho

In the example above, null hypothesis would be something like this: There is no statistically significant relationship between the lands I used to plant the flowers and growth of the flowers.

A researcher is challenged by the null hypothesis and usually wants to disprove it, to demonstrate that there is a statistically-significant relationship between the two variables in the hypothesis.

**What Is an Alternative Hypothesis?**

An alternative hypothesis simply is the inverse, or opposite, of the null hypothesis. So, if we continue with the above example, the alternative hypothesis would be that “there is indeed a statistically-significant relationship between what type of land used to plant the flower and its growth. To make it more easily the following show the two hypothesis. Alternative Hypothesis is usually denoted by Ha

**Null:** a flower plant on my land and the one planted in James’ land will have no difference in growth rate. (Bear it in mind that null always means nothing extraordinary is going on.)

**Alternative**: a flower plant on my land and the one planted in James’ land will be different in tern of growth. i.e. (the flower on James’ land will grow faster than mine).

Mathematically the same Hypothesis can be written as

Ho: Hj = Hm

Ha : Hj > Hm (where Hj and Hm represent the mean height of James’ land plant and my land plant respectively.

**Note:** when no indicator for alternative hypothesis use ‘not equal to ‘

**What are the steps to be taken for Hypothesis Testing?**

All hypotheses can easily be tested using the following four steps:

1. State the two hypotheses i.e. (Determine H0 and Ha. Remember, they are contradictory)

2. Formulate an analysis plan, which only outlines how the data will be evaluated.

3. Calculate the test statistic.

4. Use the right table to calculate the critical value.

5. Compare the calculated test statistic with the Z critical value determined by the level of significance required by the test and make a decision ( reject H0 or fail to reject H0), and write a clear conclusion using English sentences.

**2**

**1.0 How to write the null and alternative hypothesis**

The way your hypothesis will be written depend on what you are trying to test, so we can categorized all different ways you might come across in to 4 Categories

**Category A:** Test for the average of one a population

**Category B:** Test for the average of two populations.

**Category C:** Test for significant difference between means of two population.

**Category D:** Test for significant difference between proportions of two population.

**1.1 Category A (Test for average of a population)**

Here, all we are going to be dealing with will be on just one population and the mean will always be our estimator. In this category your null and alternative will always look like each of the following three patterns.

|  |  |  |
| --- | --- | --- |
| *Right and left(2-tailed)* | *Right(1-tailed)* | *Left(1-tailed)* |
| *H0: μ = mean Value* | *H0: μ ≤ mean Value* | *H0: μ ≥ mean Value* |
| *Ha: μ ≠ mean Value* | *Ha: μ > mean Value* | *Ha: μ < mean Value* |

Note: equal to sign is common to all null hypothesis. Hence the words that will be helping you to know the null statement are: ‘is’, ‘is equal to’, ’not less than’, ’not more than’, ‘not greater than’, ‘at least’, ‘at most’, ‘equal to’, ’more than or equal to’, ‘less than or equal to’, ’not up to’.

Alternative hypothesis indicator are: ‘less than’, ‘greater than’, ‘not equal to’, ‘more than’, ‘is not’, ’higher than’, ‘lower than’ etc.

**3**

**Example1.1.1**

The average life of a car battery of a certain brand is six years. This information is gathered using data obtained from people who have purchased and used this brand of battery over a period of several years. A researcher at the battery company develops a new type of car battery and claims that the average life of this battery is more than six years. To determine whether this claim is true, one would need to do some hypothesis testing.

What would be the null hypothesis and the alternative hypothesis for this hypothesis test?

Assuming the researcher claimed that the average life is not five year. Write out the null and alternative hypothesis.

**Solution**

From the question above you can easily notice the null statement and the alternative. The researcher claim is the alternative as you can see alternative indicator “more than”. While the first statement in the question is the null for the null indicator “is” was used.

Hence:

**Null hypothesis:** ''The average life of the new car battery is six years.''

**Alternative hypothesis:** ''The average life of the new car battery is more than five years.''

**Mathematically**

H0: μ = 6years

Ha: μ > 6years

(Where μ represent the average life of the new car battery)

Null hypothesis: ''The average life of the new car battery is six years.''

Alternative hypothesis: ''The average life of the new car battery is not five years.''

Mathematically

H0: μ = 6years

Ha: μ ≠ 6years

**4**

**Example1.1.2.**

 Past research data from a period of over several years states that the average life expectancy of whales is 85 years. A researcher at a laboratory wishes to test this hypothesis. To that end they procure a sample of life spans of several whales. What is the null hypothesis and the alternative hypothesis that this researcher will establish?

**Solution**

 The null hypothesis for the researcher will state that, ''The average life expectancy of whales is exactly equal to 85years.''

The alternative hypothesis will read: ''The average life expectancy of whales is not equal to 100 years.''

**Mathematically**

H0: μ = 85years

Ha: μ ≠ 85years

**More Explanation:** Here our null indicator is ‘is’ which was use in the first statement. We are not given any indicator for the alternative here. But we know that the new researcher only want to know if the statement propose is likely to be true. Since the previous research state It is 85 then opposite of that will be ‘it is not 85’.

Assuming in the question above that the researcher believes that it should more than 85 then it should be the following:

H0: μ = 85years

Ha: μ > 85years

Note: when no indicator for alternative hypothesis use ‘not equal to ‘

5

5

**Example1.1.3**

A company producing groundnut chip claims that each nylon contains 30 groundnut chips on average but I have a feeling that it is not up to that so what will be my null hypothesis and alternative hypothesis??

**SOLUTION**

 The null hypothesis: ''The average number of groundnut chips in a nylon is 30.''

Alternative hypothesis will read: '' the average number of groundnut chips in a nylon is less than 30.'' (Remember I said not up to)

Mathematically

H0: μ = 30

Ha: μ <30

Null indicator here is ‘contain’ which is similar to ‘is’

Alternative indicator is ‘not up to’ which is same as less than

**Example1.1.4**

A particular brand of tires claims that its deluxe tire averages at least 20,000 miles before it needs to be replaced. From past studies of this tire, the standard deviation is known to be 1,000. A researcher believes this is a lie and conduct survey.

**Solution**

 The null hypothesis: ''The average is at least 20,000.'' i.e more than or equal to 20,000

The alternative hypothesis will read: “it is less than 20,000

Mathematically

H0: μ ≥ 20000

Ha: μ <20000

Null Indicator: at least; Alternative indicator: Not stated

Here: the alternative will be the opposite of the null which is greater than.

Did you understand this?? Yeah, it is as easy as that.

**6**

**Example1.1.5**

Suppose that a recent article stated that the mean time spent in jail by a first–time convicted rapist is 14 years. Your team conducted a research to sees if the mean time has increased last year. What will be the null and alternative hypothesis for this research??

**Solution**

Null: that the mean time spent in jail by a first–time convicted rapist is 14 years.

Alternative: the mean time has increased

When something increased what does it mean?? Yeah it means it is more than before. Hence

Mathematically

H0: μ=14years

Ha: μ >14years

**Summary**

In short your null hypothesis is very easy to write as it will always include equal to or

Specific value. E.g H0: μ = 85years, H0: μ = 35kg. it can be in for any of the three below

Two-tailed test one-tailed test one-tailed test

H0: μ = mean ValueH0: μ ≤ mean ValueH0: μ ≥ mean Value

Ha: μ ≠ mean ValueHa: μ > mean ValueHa: μ < mean Value

The Alternative Hypothesis can be of any of the three below.

1. Ha:  μ ≠ 85years (when we are only testing to say it is not equal i.e 2tails test.)
2. Ha:  μ > 85years (when we are testing if it is greater than i.e. 1tail test.)
3. Ha:  μ > 85years (when we are testing if it is less than i.e 1tail test)

**Note:** if the null is form of first pattern i.e H0: μ = mean Valuethen its alternative can be any of the three pattern but if it is any of the remaining two, then it should go with specific alternative written above.

That is how to write the null and alternative hypothesis for the first Category.

Now try to deal with the exercise.

7

**Exercise1.1**

**Exercise**

1. Imagine you were carrying out research for a fluorescent bulbs company, you are to test the hypothesis that the mean lifetime of the bulbs produced by the company is 506days. What will be your null and alternative hypothesis statement?
2. A school owner want to know if increasing the amount of light during the studying will increase the performance of the participant in the test score so he asked for your help to construct the hypothesis.
3. Your friend posted on his status that “on average Nigerian females lost their virginity at the latest by 15years or less. You think Nigerian females are not that worst hence you try to test the virginity of some females. What will be your hypothesis for this research?
4. A company has stated that their straw machine makes straws that are 4mm diameter. A worker believes the machine no longer make straws of this size. To perform the hypothesis test what will be the null and alternative hypothesis for this worker’s belief
5. A doctor believes that the average teen sleeps on average no longer than 9hours per day. A researcher believes it should be that teens on average sleep longer than that. Write the null and alternative hypothesis for the research.
6. It is believed that a candy machine makes chocolate bars that are on average 5.5g. A worker claims that the machine after maintenance is no longer makes 5.5g bars. Write the null and alternative hypothesis for this claim.
7. You are testing that the mean speed of your cable Internet connection is more than three Megabits per second. State the null and alternative hypotheses..
8. The mean entry level salary of an employee at a company is $58,000. You believe it is higher for IT professionals in the company. State the null and alternative hypotheses.
9. Suppose that a recent article stated that the mean time spent in jail by a first–time convicted burglar is 2.5 years. A study was then done to see if the mean time has increased in the new century. A random sample of 26 first-time convicted burglars in a recent year was picked. The mean length of time in jail from the survey was 3 years with a standard deviation of 1.8 years. Suppose that it is somehow known that the population standard deviation is 1.5. If you were conducting a hypothesis test to determine if the mean length of jail time has increased, what would the null and alternative hypotheses be? The distribution of the population is normal

H0: \_\_\_\_\_\_\_\_ Ha: \_\_\_\_\_\_\_\_

1. A random survey of 75 death row inmates revealed that the mean length of time on death row is 17.4 years with a standard deviation of 6.3 years. If you were conducting a hypothesis test to determine if the population mean time on death row could likely be 15 years, what would the null and alternative hypotheses be?

H0: \_\_\_\_\_\_\_\_ Ha: \_\_\_\_\_\_\_\_

**8**

**Example1.2.4**

**1.2. How to write hypothesis of proportion (Category B)**

Here we are going to be dealing with proportion and percentage our test here will focus on just one population. To write the test will be similar to the previous category, only that we are no more using mean instead we will be using proportion.

**Example1.2.1**

The school board claims that at least 60% of students brings a phone to school. A teacher believes this number is too high and randomly selected 25students. What will be the null and alternative hypothesis??

**Solution**

**Null:** at least 60% of students brings a phone to school.

**Alt:** Not at least 60% of students brings a phone to school.

Here it should be understand that at least in the statement is the null indicator also it can be interpreted as greater than or equal to. The opposite of at least will always be less than.

Hence;

H0: P ≥ 60%

Ha: P<60%

**Example1.2.2**

Your friend wrote the following quote on his whatsapp status.

“Not less than 80% of beautiful ladies in Nigeria usually have low IQ”. You decided to do research if this is true what will be the null and alternative hypothesis.

**Solution**

Null: Not less than 80% of beautiful ladies in Nigeria usually have low IQ

**Alt:** Not less than 80% of beautiful ladies in Nigeria usually have low IQ.

Here it should be understand that **not less than** in the statement is the null indicator also it can be interpreted as greater than or equal to. The opposite of **not less than** is **less than**.

Hence;

H0: P ≥ 80%

Ha: P<80%

**9**

In a population of 1950 air force, it was stated that 500 of them are found to be female last year. You want to test if by chance it is possible that the proportion of females is still the same this

**Example1.2.3**

Your friend wrote the following quote on his whatsApp status.

“Not less than 80% of beautiful ladies in Nigeria usually have low IQ”. You decided to do research if this is true what will p be the null and alternative hypothesis.

**Solution**

Null: Not less than 80% of beautiful ladies in Nigeria have low IQ

**Alt:** Not less than 80% of beautiful ladies in Nigeria usually have low IQ.

Here it should be understand that **not less than** in the statement is the null indicator also it can be interpreted as greater than or equal to. The opposite of **not less than** is **less than**.

Hence;

H0: P ≥ 80%

Ha: P<80%

Where P represent the percentage of beautiful ladies in Nigeria with low IQ

**Example1.2.4**

It was believe that out of 1950 people admitted to air force every year, 500 of them are female. Then a researcher want to test the claim. Establish the hypothesis for this research.

**Solution**

Here also we are dealing with proportion only that the figures are given. Hence we need to calculate the proportion by ourselves.

Total=1950

Female=500

Proportion of female (P)=500/1950=0.256

We are testing whether it is the same or not. This shows that our hypothesis we look like

Null: the proportion of female air force is 0.256

**Alt:** the proportion of female air force is not 0.256

Hence;

H0: P= 0.256

Ha: P≠0.256

**10**

**Example 1.2.5**

The last article stated that the poll taken among 25faculty member at a university shows that 60% of those polled favor a longer break between semester (and a shorter summer vacation) As a researcher you think this is too high and decided to conduct a new poll to justify your belief what will be your null and alternative hypothesis.

**Solution**

This question is straight and direct, we already know what the alternative and the null will look like looking at the word “too high” the word mean we believe it should be less.

**Null**: the percentage of those who favor a longer break between semesters is 60%

**Alt**: the percentage of those who favor a longer break between semesters is less than 60%

H0: P=60%

Ha: P<60%

**Example 1.2.6**

It was stated that ‘not more than 30% of Nigerian Youths have more than #10,000 in their bank account since 2018 till date’. A business man believes Nigeria has not gotten worst to that extent and ask you to carry out research on that. Write out your Hypothesis.

**Solution**

Here the word “not more than” is null indicator here. And the opposite which the business man believed is “more than”

Null: not more than 30% of Nigerian Youths have more than #10,000

Alt: more than 30% of Nigerian Youths have more than #10,000

H0: P≤30%

Ha: P>30%

Note: Not more than can be interpreted as less than or equal to.

Now try the following Exercise

**11**

**Exercise1.2**

1. Marketers believe that 92% of adults in the United States own a cell phone. A cell phone manufacturer believes that number is actually lower. Write out the hypothesis if you were to test for this claim.
2. The National Institute of Mental Health published an article stating that in any one-year period, approximately 9.5 percent of American adults suffer from depression or a depressive illness. Suppose that in a survey of 100 people in a certain town, seven of them suffered from depression or a depressive illness. If you were conducting a hypothesis test to determine if the true proportion of people in that town suffering from depression or a depressive illness is lower than the percent in the general adult American population, what would the null and alternative hypotheses be?

H0: \_\_\_\_\_\_\_\_

Ha: \_\_\_\_\_\_\_\_

1. In a population of fish, approximately 42% are female. A test is conducted to see if, in fact, the proportion is less. State the null and alternative hypotheses
2. Read the following poem very well and answer the question in it.

My dog has so many fleas,

They do not come off with ease.  
As for shampoo, I have tried many types  
Even one called Bubble Hype  
Which only killed 25% of the fleas,  
Unfortunately I was not pleased.

Until I had given up hope  
Until one day I saw  
An ad that put me in awe.

A shampoo used for dogs  
called GOOD ENOUGH to Clean a Hog  
guaranteed to kill more fleas.

With the shampoo I gave Fido a bath  
and started doing the math.  
But I need you to write me the Hypothesis

Before I can do the analysis  
so that I can figure out,

Whether to use the new shampoo or go without?

12

I've used all kinds of soap,  
**1.3 Category C (Dealing with the significance of two population Homogeneous or Heterogeneous)**.

/

1. A research shows that about 35% of students of university of Ibadan pass the UTME on the first try. We want to test if more than 35% pass on the first try. Write out the Null and alternative hypothesis for the test with correct symbols.
2. Maxwizard (An analyst) claims that the proportion of students going to love garden in the University of Ibadan to meet someone He/she loves is 0.8. You want to test to see if the claim is correct. State the null and alternative hypotheses.
3. It was claimed long ago that in a population of fish, approximately 42% are female. A test is conducted to see if, in fact, the proportion is less. State the null and alternative hypotheses.
4. A recent global report on insurance claims payment says that, “a less thansatisfactory claims experience prompts one in five customers to switch insurance providers.”
5. The Pew Research center just reported that Muslims have an average birthrate of 2.9 children per woman. You want to test if the claim is true for your country what will be null and alternative hypothesis.
6. Dr. West Crenshaw published an article in “Your Teen Magazine “that indicated that 47% of high school students in the US have had sex. You want to test if this proportion is the same for high school students in Nigeria what will be the null and Alternative hypothesis
7. From Dr. West Crenshaw’s article it was also stated that 70% teens in the world have had sex or a make out experience at least. You think the world has not bad to that extent what will be the null and alternative hypothesis

**13**

Here we will be dealing with both mean and proportion of two different population or one population of two categories. Our research will focus on comparison between two things. The question of the research will be; is there a significant differences or not? If yes which one is higher or less?? To understand this well let’s take a look at some examples.

**Example 1.3.1**

Suppose that a recent article stated that the mean time spent in jail by a first–time convicted rapist in US is 14 years. You client just has a case of raping, so He would either get justice in Nigeria or US that depends on his choice. Your team conducted a research to sees if the mean time spend in jail by first-time convicted rapist in US is the same as in Nigeria or not.

A) What will be your null and alternative hypothesis?

B) Assuming the research is to know if the mean time spent in US prison is less than Nigeria’s. What will be the null and alternative?

**Solution**

**A) H0:** μu=μ**N=14years**

**Ha:** μu≠μ**N**

**Explanation**

Here we are considering two population, rapist in US and rapist in Nigeria. As you can see they are two different countries. The hypothesis indicator in the question above is ‘is the same as’ in the statement **the mean time spend in jail by first-time convicted rapist in US is the same as in Nigeria or not.** This shows that the alternative will be not equal to. In another word we can write it as the following

**Null:** the mean time spend in jail by first-time convicted rapist in US is the same as in Nigeria

**Alt:** the mean time spend in jail by first-time convicted rapist in US is not the same as in Nigeria

**B) H0:** μu≥μ**N**

**Ha:** μu<μ**N**

**Explanation**

The hypothesis indicator here has been changed to **less than**. So it isvery easy to understand this.

**Null:** the mean time spend in jail by first-time convicted rapist in US is not less than mean time in Nigeria.

**Alt:** the mean time spend in jail by first-time convicted rapist in US is less than mean time in Nigeria.

**Example 1.3.2**

You were given a projects to do research on a project labelled “Do younger U.S. males weigh more on average than older males?” write out your null and alternative hypothesis for this research.

**Solution**

**Hypothesis Indicator: ‘more’** more here is telling us that our alternative will be written with more than while the null will be written with the opposite. Which will be less than or equal to.

**H0:** μy≥μ**A**

**Ha:** μy>μ**A** (where μy, μ**A** represent the weight of younger U.S male and older U.S male on average respectively)

**Example 1.3.3**

There was a belief that the proportion of students happened to be Yahoo boys in University of Ibadan is the same as that of LAUTEC. If you were to test the claim what will be your null and alternative hypothesis??

**Solution**

**H0:** PI=P**L**

**Ha:**  PI=P**L** (where PI and P**L** represent the proportion of yahoo boys in U.I and LAUTEC respectively)

**Example 1.3.4**

A tattoo magazine claimed that the percent of men that have at least one tattoo is greater than the percent of women with at least one tattoo. To test this claim what will be the hypothesis.

**Solution**

**H0:** Pm=P**F**

**Ha:**  Pm>P**F** (where Pm and P**F** represent the percentage of men and women that have at least one tattoo respectively)

**Example 1.3.5**

A health magazine claims that marriage status is one of the most telling factors for a person's happiness. To test the claim that the percent of married people that are unhappy is lower than the percent of single or divorced people that are unhappy what will be your null and alternative.

**Solution**

**H0:** Pm=P**sd**

**Ha:**  Pm<P**sd** (where Pm and P**sd** represent the percentage of unhappy married people and unhappy single or divorced people respectively)

**More exercise on Hypothesis**

**Write the null and alternative hypothesis for the following**

1. The mean number of years Americans work before retiring is 34.
2. At most 60% of Americans vote in presidential elections.
3. The mean starting salary for San Jose State University graduates is at least $100,000 per year.
4. Twenty-nine percent of high school seniors get drunk each month.
5. Fewer than 5% of adults ride the bus to work in Los Angeles.
6. The mean number of cars a person owns in her lifetime is not more than ten.
7. About half of Americans prefer to live away from cities, given the choice.
8. Europeans have a mean paid vacation each year of six weeks.
9. The chance of developing breast cancer is under 11% for women.
10. Private universities' mean tuition cost is more than $20,000 per year.
11. **How to determine the type of test to use.**

There are two important test commonly used although there are still more which we are not interested in for now they are:

1. Z-test or normal distribution test

2. T-test

**Rule:** Use the Z-test if and only if the data is normally distributed or the sample size is more than or equal to thirty i.e. when n >30. Otherwise use T-test.

**Rule2:** use Z-test (normal distribution table) when the population standard deviation is known

Rule3: whenever you are dealing with proportion use Z-test (normal distribution table) if and only if np≥30 and n(1-p) ≥30 otherwise use Binomial distribution

**Example**

**Exercise**

1. Which distribution do you use when you are testing a population mean and the population standard deviation is known? Assume a normal distribution, with n ≥ 30.
2. Which distribution do you use when the standard deviation is not known and you are testing one population mean? Assume sample size is large.
3. A population mean is 13. The sample mean is 12.8, and the sample standard deviation is two. The sample size is 20. What distribution should you use to perform a hypothesis test? Assume the underlying population is normal.
4. A population has a mean of 25 and a standard deviation of five. The sample mean is 24, and the sample size is 108. What distribution should you use to perform a hypothesis test?
5. It is thought that 42% of respondents in a taste test would prefer Brand A. In a particular test of 100 people, 39% preferred Brand A. What distribution should you use to perform a hypothesis test?
6. You are performing a hypothesis test of a single population mean using a Student’s t-distribution. What must you assume about the distribution of the data?
7. You are performing a hypothesis test of a single population mean using a Student’s t-distribution. The data are not from a simple random sample. Can you accurately perform the hypothesis test?
8. You are performing a hypothesis test of a single population proportion. What must be true about the quantities of np and nq?
9. You are performing a hypothesis test of a single population proportion. You find out that np is less than five. What must you do to be able to perform a valid hypothesis test?
10. You are performing a hypothesis test of a single population proportion. The data come from which distribution?

**2.2 How to know which tail test to be used.**

To choose between two tails or one tail test is very easy, all we need to inspect is the hypothesis statement written and follow the following principle.

**Principle**: Use one tail test when your alternative hypothesis has greater than or less than sign else use two tails.

What is the Principle trying to explain? If you have any of the following as your alternative hypothesis you are expected to use one tail

Ha: μ > mean ValueHa: μ < mean Value

Ha: P > proportion ValueHa: μ < proportion Value

Here greater than > denote the right tail test while less than < sign denote left tail test.

**Note:** The only time you should use two tails is when your alternative hypothesis is written with not equal to sign or you were instructed to use it.

**Example 2.2**

**For each of the following question**

a) Write the null and alternative hypothesis.

b) Label whether the null or the alternative is the original claim.

c) Tell whether this is a left tail test, a right tail test, or a two tail test.

1. A company producing groundnut chip claims that each nylon contains 30 groundnut chips on average but you have a feeling that it is not up to that so you want to test the claim.
2. A particular brand of tires claims that its deluxe tire averages at least 20,000 miles before it needs to be replaced. From past studies of this tire, the standard deviation is known to be 1,000. A researcher believes this is a lie and conduct survey.
3. A health magazine claims that marriage status is one of the most telling factors for a person's happiness. Test the claim that the percent of married people that are unhappy is lower than the percent of single or divorced people that are unhappy.
4. In a population of 1950 air force, it was stated that 500 of them are found to be female last year. You want to test if by chance it is possible that the proportion of females is still the same this year. Then establish the hypothesis for this research.

**Solution**

**1.** H0: μ = 30

Ha: μ<30

b) The original claim is null. (As you can see here that what was written for null is the company claim)

c) This is left tail since the alternative hypothesis is less than.

2. H0: μ ≥ 20000

Ha: μ <20000

b) The null is the claim

c) It is left tail

3 **H0:** Pm=P**sd**

**Ha:**  Pm<P**sd** (where Pm and P**sd** represent the percentage of unhappy married people and unhappy single or divorced people respectively)

b) The alternative hypothesis is the claim.

c) It is left tail.

4. Since 500/960 =0.256

H0: P= 0.256

Ha: P≠0.256

b) The null is the claim

c) It is two tail test (it the alternative has not equal to)

**Exercise 2.2**

a) Write the null and alternative hypothesis.

b) Label whether the null or the alternative is the original claim.

c) Tell whether this is a left tail test, a right tail test, or a two tail test.

1. According to a CNN report, besides cell phones, 93% of Americans also own a traditional phone. But has that percentage decreased as more and more Americans opt to only use a cellphone and throw away their traditional phones.

2. More and more Americans are becoming financially sound and opting to not own a credit card. According to an article in USA Today, 74% of Americans still have at least one credit card. But this claim seems a little on the low side. We think that more than 74% of Americans own a credit car

3. According to a recent Newspaper article, people in California spend 1.25 hours a day eating and drinking. Suppose we want to test the claim that the number of hours spent eating and drinking is really 1.25 hours.

4. The standard deviation for the heights of men was thought to be 2.9 inches. New studies disagree with this. Test the claim that the standard deviation for heights of men is not 2.9 inches.

5. It has long been thought that normal body temperature is really 98.6 degrees Fahrenheit. Recent study is now claiming that normal body temperature is really lower than 98.6 degrees.

6. Wikipedia suggests that at least 10% of the world population is left handed. Wikipedia may not be very accurate. Test the claim that at least 10% of the world population is left handed

7. The percent of women that hold CEO level jobs is lower than the percent of men that hold CEO level jobs

**State the type of test to be used for the following (1tail, 2tails, right tail, left or tail )**

8. Assume H0: μ = 9 and Ha: μ < 9. Is this a left-tailed, right-tailed, or two-tailed test?

9. Assume H0: μ ≤ 6 and Ha: μ > 6. Is this a left-tailed, right-tailed, or two-tailed test?

**10.** Assume H0: p = 0.25 and Ha: p ≠ 0.25. Is this a left-tailed, right-tailed, or two-tailed test?

11. A bottle of water is labeled as containing 16 fluid ounces of water. You believe it is less than that. What type of test would you use?

12. Your friend claims that his mean golf score is 63. You want to show that it is higher than that. What type of test would you use?

13. A bathroom scale claims to be able to identify correctly any weight within a pound. You think that it cannot be that accurate. What type of test would you use?

14. You flip a coin and record whether it shows heads or tails. You know the probability of getting heads is 50%, but you think it is less for this particular coin. What type of test would you use? If the alternative hypothesis has a not equals ( ≠ ) symbol, you know to use which type of test?

15. Assume the null hypothesis states that the mean is at least 18. Is this a left-tailed, right-tailed, or two-tailed test?

16. Assume the null hypothesis states that the mean is at most 12. Is this a left-tailed, right-tailed, or two-tailed test?

17. Assume the null hypothesis states that the mean is equal to 88. The alternative hypothesis states that the mean is not equal to 88. Is this a left-tailed, right-tailed, or two-tailed test?

1. **How to calculate the Zstatistics**

I believe you should already know how to calculate the Zstatistics but should you don’t I am going to give you some example on it but it will be inform of the hypothesis questions. But before we go in to that let me give you some important formula you need to know when trying to calculate the Zstatistics and when to use them.

**What are the most important formula used in calculating Zstatistics in hypothesis??**

The most important formula needed for calculating the Zstatistics are of four category.  
1. Using sample mean, sample size and standard Deviation

2. Using the proportion and sample size

3. Using two sample mean, size and standard deviation from two different population

4. Using two proportion and sample size from two different population

**CASE 1**

**Formula for calculating Zstatistics when Sample mean, standard deviation and sample size are given.**

x̄ = Sample mean

μ = population mean

σ = Population standard deviation

n = Sample size

Note: S.E=σ/√n where S.E represent the standard error. Hence we can also rewrite the formula as

**CASE 2.**

Formula for calculating Zstatistics when proportion and sample size is given

Where P0 is the population proportion

P is the proportion from sample

n is the size of the sample

In this case mean and standard deviation can be easily calculated as following;

μx **=** np and

**CASE 3.**

Formula for calculating Zstatistics when Sample mean, size and standard deviation from two different population is given.

Where  is the mean of the first sample,

is the mean of the second sample

is the size of the first sample from the first population­­­

is the size of the second sample from the second population

is the difference between the two proportions.

**CASE 4.**

Formula for calculating Zstatistics when two proportion and sample size from two different population is given*.*

Where P is the average calculated proportion

is the proportion of the first sample

is the proportion of the second sample

is the size of the first sample from the first population

is the size of the second sample from the second population.

is the difference between the two proportions.

All of the formula above needed to be master well before attempting any question we shall work on their example later

1. **How to check the critical value from statistical table**

**4.1 How to make use of statistical table to find the critical value from normal distribution table**

Firstly it is important that you download the table here. Then move on to make use of it.

The Pdf you downloaded has two different normal distribution table in which one are labeled with negative numbers while the other is positive number. Now let make use of the negative one.

**Example 4.1.1**

What is the critical value when α is 5% assuming it is two tailed test.

**Solution**

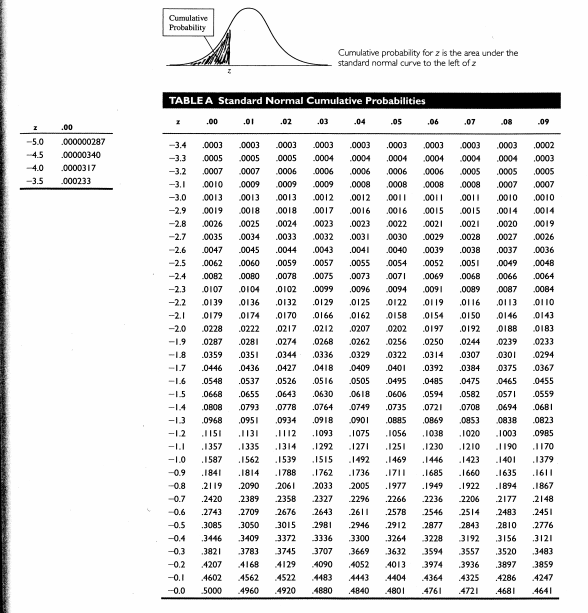
Step1

Here α=5% =0.05

α /2 = 0.05/2 =0.025 (since it is two tailed test)

Step2

Now that we already know the important value go in to the table and look for wherever you have (0.025) or any number very closed to that.



Step3: As you can see above we see exactly the number we are looking for. Now match the row and column of the place where the number is located.

Here if we go to the leftmost we can see (-1.9) and to the topmost we can see (0.06).

Step 4: we merge both numbers together and neglect the negative i.e 1.9 + 0.06 = 1.96

Hence the critical value is ±1.96.

**Example 4.1.2**

What is the critical value of the question given in example1 assuming it is one tail test?

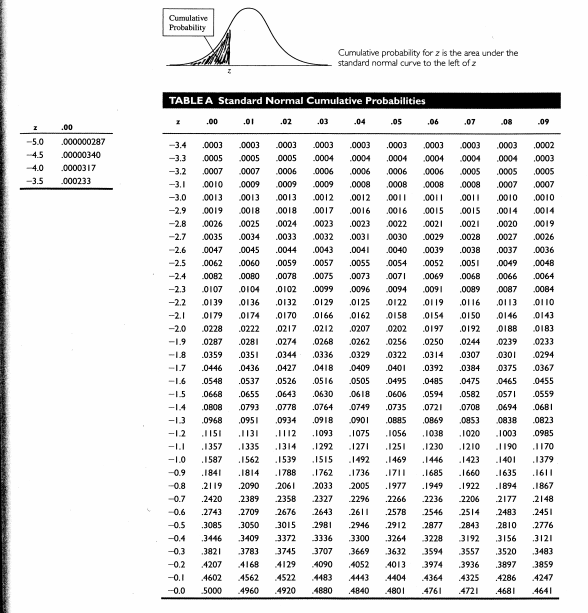
Solution

Step1

Here α=5% =0.05 (no need to divide b 2 since it is one tailed test)

Step2

Now that we already know the important value, go to the table and look for wherever you have (0.05) or any number very closed to that.

s

Step3: here we do not see the exact value (0.05), hence we will pick first number closest to it. Here we can pick either (0.0505) or (0.0495) but I prefer the first one.

Now match the row and column of the place where the number is located. If we go to the leftmost we can see (-1.6) and to the topmost we can see (0.04).

**Step 4:** we merge both numbers together and neglect the negative i.e 1.6+ 0.04 = 1.64

Hence the critical value is 1.64.

**Note:** only one value is gotten for one tailed test while plus or minus is written for two tailed test.

**Example 4.1.3**

What is the critical value of a normal distribution data given 98% as confidence level assuming you are doing left tail test??

**Solution**

Step1

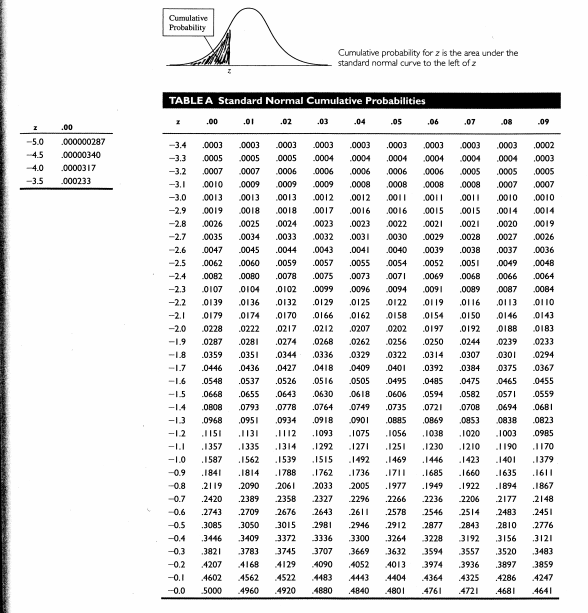
CL=98% =0.98

α =1-CL =1-0.98 =0.02

Here α=0.02 (no need to divide b 2 since it is one tailed test)

Step2

Now that we already know the important value, go to the table and look for wherever you have (0.02) or any number very closed to that.



**Step3:** As you can see above we don’t see exact number we are looking for. So we are choosing the leftmost value between the two closed value of our target i.e. (0.0202 and 0.0197), so we choose 0.0202. Now match the row and column of the place where the number is found.

Here if we go to the leftmost we can see (-2.0) and to the topmost we can see (0.05).

**Step 4:** we merge both numbers together and neglect the negative i.e. 2.0 + 0.05=2.05

Hence the critical value is -2.05. (Whenever you are given left tail our answer will be negative)

Note: we would choose positive answer if the question has not told us that it is left tail.

**Example 4.1.4**

What is the critical value of a normal distribution data given 98% as confidence level??

**Solution**

Step1

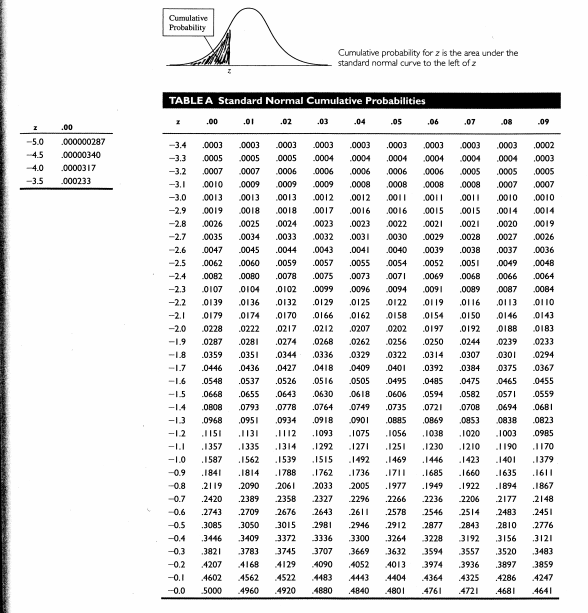
CL=98% =0.98

α =1-CL =1-0.98 =0.02

Here α=0.02 /2 =0.01 (since the type of tail is not stated we will choose two tails)

Step2

Now that we already know the important value, go to the table and look for wherever you have (0.01) or a number very closed to that.



Step3: As you can see above we don’t see exact number we are looking for. So we are choosing the leftmost value between the two closed value of our target i.e. (0.0102 and 0.0099), so we choose 0.0102. Now match the row and column of the place where the number is found.

Here if we go to the leftmost we can see (-2.3) and to the topmost we can see (0.02).

Step 4: we merge both numbers together and neglect the negative i.e 2.3 + 0.02=2.32

Hence the critical value is ±2.05. (Remember two answers are gotten for 2tails)

* 1. **How to check the critical value from t-table**

**Example 4.2.1**

What is the critical value of a data with sample size 9 and level of significant of 1% for 1 tail test.

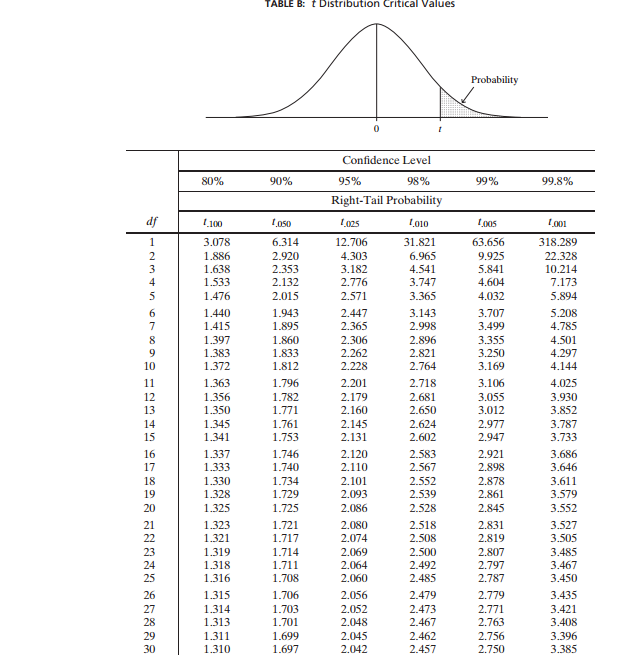
**Solution**

Here α = 1% =0.01

n=9, df=n-1

df=9-1=8

Now look for the 1% or 0.01 on the top of our table and 8 from 'df' column. Match it together as you can see below



As you can see the answer is 2.896, please don’t be deceived b 98% on the top, it is nothing what is written as decimal is what Is important if you have this kind of table with you.

**Example 4.2.2.**

Assuming we decide to do two-tail test for example1 above what will be the critical value??

**Solution**

We would follow the same steps here too.

α = 1%

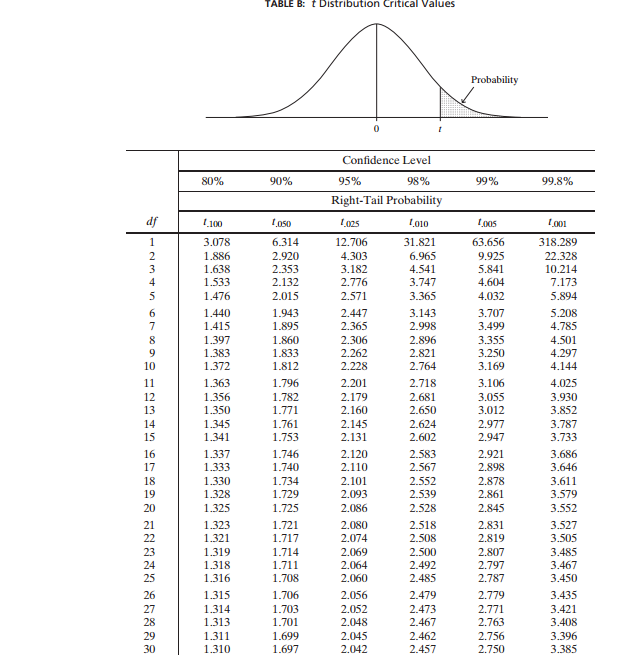
α /2 = ½ %=0.5%

α /2 = 0.5/100=0.005 (to Decimal)

n=9, df=n-1

df=9-1=8

Now look for the 0.5% or 0.005 on the top of our table and 8 from 'df' column. Match it together as you can see below



Now it is very clear that the result from the table is 3.355

Example4.2.3

What will be the critical value of sample of size 18 with confidence level of 95%??

Solution

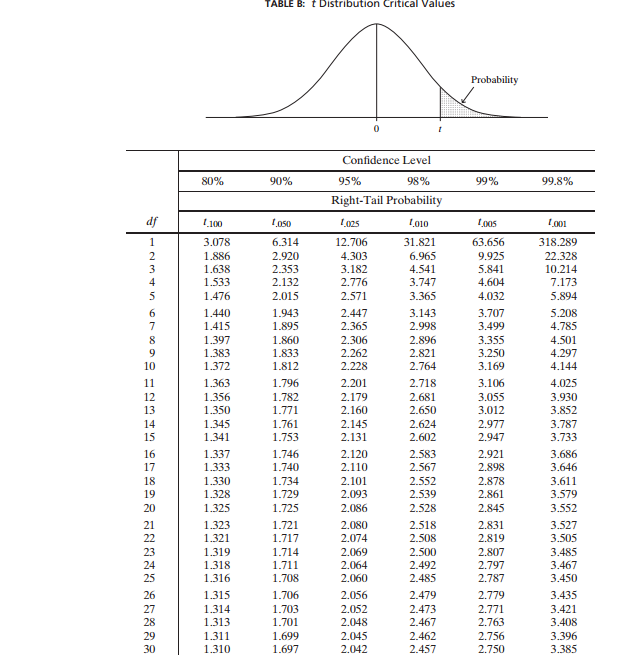
CL=95% =0.95

0.05/2 =0.025 (to Decimal)

n=9, df=n-1

df=18-1=17

Now look for the 2.5% or 0.025 on the top of our table and 17 from 'df' column. Match it together as you can see below



Now it is very clear that the result from the table is 2.110

**Example4.2.4**

What will be the critical value of sample of size 28 with confidence level of 90% for one tail test??

**Solution**

CL=95% =0.90

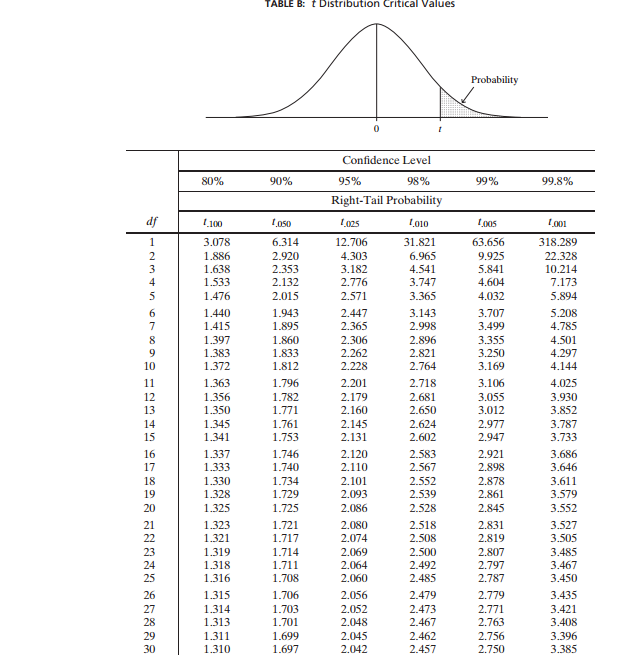
α =1-CL= 1-0.90=0.1

α = 0.1 (to Decimal)

n=28, df =n-1

df=28-1=27

Now look for the 10% or 0.1 on the top of our table and 27 from 'df' column. Match it together as you can see below



Now it is very clear that the result from the table is 1.314

1. **How to make decision Using Rejection Rule.**

To make your decision on the null hypothesis is within the following three rules

**Rule1**: if the alternative hypothesis is written with greater than then reject the null hypothesis if and only if the Zstatistics is greater than the Z-value gotten from the statistical table i.e. the critical value.

**Rule2**: if the alternative hypothesis is written with less than then reject the null hypothesis if and only if the Zstatistics is less than the Z-value gotten from the statistical table i.e. the critical value.

**Rule3**: if the alternative hypothesis is written with equal to then Reject the null hypothesis if and only if the absolute value of the Zstatistics is greater than the Z-value gotten from the statistical table i.e. the critical value.

**Note**: before you do comparison you must let your Z-value from table correspond with the Z-statistic i.e. negate the Z-value whenever your Z-statistics is negative value before you do the comparison.

Now that you have learned all the important procedure let’s take a look at some examples

1. **Testing Hypothesis and making conclusion**

**6.1 Testing for mean of one population**

**Example 6.1.1**

In the population, the average IQ is 100 with a standard deviation of 15. A team of scientists want to test a new medication to see if it has either a positive or negative effect on intelligence, or not effect at all. A sample of 30 participants who have taken the medication has a mean of 150. Did the medication affect intelligence?

**Solution**

H0: (the IQ is not affected by the medication i.e. it is still 100 as it is believed)

Ha:  (the IQ is affected by the medication i.e. it is no more 100 as it is before)

**Mathematically**

H0: μ = 100years (this is usually noted from the population)

Ha: μ ≠ 100years (either there is any effect)

**Step2**

α=0.5 (whenever it is not given we use this

Since the sample size is up to 30 then we will need to use Z-table also it must be two tails test and rule3 should be apply for decision since the alternative test is with “not equal to” sign.

Now we need to proceed to step3

**Step3**

Zcal= 50/2.7386

Zcal=18.2574

**Step4**

α=5% = 0.05 (since the level of significant is not given.)

Zα/2=Z0.05/2 = Z0.025 (since we are dealing with two tails test)

We need to look for the value of 0.025 from the normal distribution table and that give 1.96

**Step5**

Comparing the Zcal and Ztable we see that Zcal >Ztable therefore:

**Decision**: we reject the null hypothesis.

**Conclusion:** The result of the test shows that at 5% level of significant the medication is consider to have effect on the IQ. (Since the null hypothesis is rejected)

**Example 6.1.2**

Past research data from a period of over several years states that the average life expectancy of whales is 85 years. A researcher at a laboratory wishes to test this hypothesis. To that end, they procure a sample of the life spans of 35 whales. If the mean and standard deviation of the 40 whales is 90 and 2.5 respectively is the past research still right at 1% level of significant?

**Solution**

**Step1**: The null hypothesis for the researcher will state that ‘’the average life expectancy of whales is exactly equal to 85 years.’’

Alternative hypothesis will read: ‘’the average life expectancy of whales is not equal to 100 years.’’

**Mathematically**

H0: μ = 85years

Ha: μ ≠ 85years

**Step2:**

α=0.01 (it is not given in the question as 1%)

Since the sample size is more than 30 then we will need to use Z-table also it must be two tails test since the alternative test is with “not equal to sign”.

**Step3**

 μ = 85, σ = 2.5, n= 40, x̄= 90

Zcal= 5/0.3952

Zcal=12.6518

**Step4**

α= 1% = 0.01 (since the level of significant is not given.)

Zα/2=Z0.01/2 = Z0.005 (since we are dealing with two tails test)

We need to look for the value of0.025 from the normal distribution table and that give 2.57

**Step5**

Comparing the Zcal(12.6518) and Z-table (2.57) we see that Zcal >Ztable which follow rule3 of the Rejection rules, therefore :

**Decision**: we reject the null hypothesis.

**Conclusion**

The result above give solid evidence that the average life expectancy of whales is not 85 years at 1% level of significance therefore it can be stated that the past research is no more right

**Example 6.1.3**

A company producing groundnut chips claims that each nylon contains 30 groundnut chips on average but I have a feeling that it is not up to that so I inspected a package containing 60 nylons. If the mean and standard deviation of the package is 33 and 2 respectively should I sue the company at 99% confidence???

**Solution**

**Null hypothesis**: ''The average number of groundnut chips in a nylon is 30.''

**Alternative hypothesis**: '' the average number of groundnut chips in nylon is less than 30.'' (Remember I said not up to)

**Mathematically**

H0: μ = 30

Ha: μ <30

**Step2**

CL=99% =0.99 (this is what is given)

α=1-CL=1-0.99 =0.01

**Step 3**

**Step4**

α= 1% = 0.01

Zα=Z0.01 (since we are dealing with one tails test)

We need to look for the value of 0.01 from the normal distribution table and that gives 2.326

**Step5**

Comparing the Zcal (11.6189) and Z-table(2.326) we see that Zcal >Ztable which follows rule2 of the Rejection rules because we are using less than sign in alternative hypothesis, therefore :

**Decision**: we do not reject the null hypothesis.

**Conclusion**: There is not enough evidence to claim of that the chips on average is not up to 30 . Therefore I would be advised not to sue the company.

**Example 6.1.4**

A researcher examined 200 students on academic performances. He discovered that the average score of the students in major courses is 68 with a variance of 8. It is believed that the average score of these students in major courses is 65 but the researcher thinks that is too high. Test the hypothesis for this claim at 1% level of significant.

**Solution**

The initial belief will be our null hypothesis i.e

**Null:** the average score of students in major courses is 65

**Alternative**: the average score of these students in major courses is less than 65.

Notice since it is not stated what the researcher thinks “that is too high, meaning it should be less than 65.

**Mathematically**

H0: **μ** = 65

Ha: **μ** < 65

**Step2**

α=1% = 0.01

Since the sample size is more than 30 then we will need to use Z-table also it must be one tails test since the alternative test is with less than to sign.

**step3**

μ = 65, σ = 8, n= 200, x̄= 68

Zcal= 3/1.7889

Zcal=5.3033

**Step4**

α= 1% = 0.01 (since the level of significant is not given.)

Zα=Z0.01 (since we are dealing with one tails test)

We need to look for the value of 0.01 from the normal distribution table and that gives 2.326

**Step5**

Comparing the Zcal (5.3033) and Z-table (2.326) we see that Zcal >Ztable which follows rule2 of the Rejection rules because we are using less than sign in alternative hypothesis, therefore:

**Decision:** we do not reject the null hypothesis.

**Conclusion:** we do not have sufficient evidence to state that the average score of the students in major courses is less than 65.

**Note:** we can also Use Pvalue to make decision too like b

Since our Zcal= 5.3033. we check for the Pvalue from the table

P (Z<5.3033) =0.000000287

Now comparing the Pvalue and the level of significant.

(0.000000287) Pvalue < alpha (0.01) Hence we do not reject the null hypothesis.

**Example 6.1.5**

The mean life of a sample of 13 fluorescent bulbs produced by a company is found to be 497days with standard deviation of 47days. Test the hypothesis that the mean lifetime of the bulbs produced by the company is 506days at 1% level of significance.

**Solution**

The hypothesis statement here is clear. The question state we should test that mean lifetime is 506days that means we will have the following as our hypothesis.

**Null:** the mean lifetime of the bulbs produced by the company is 506days

**Alternative:** the mean lifetime of the bulbs produced by the company is not 506days

**Mathematically**

H0: μ = 506

Ha: μ ≠ 506

**Step2**

α=1% = 0.01

Since the sample size (13) is less than 30 then we will need to use T-table also it must be two tails test since the alternative test is with no equal to sign.

**step3**

μ = 506, σ = 47, n= 13, x̄= 497

tcal= (497-506)/(47/√13)

tcal = -9/13.035

tcal = -0.69

**Step4**

α= 1% = 0.01 (since the level of significant is not given.)

Tα/2=T0.01/2=0.005 (since we are dealing with two tails test)

Now To get the value from the table we need the degree of freedom and the level of significance

Recall that df = n-1

df=13-1 =12

We need to look for the value of 12 under 0.005 or 0.5% from T-distribution table and that gives 4.221

**Step5**

Comparing the Zcal (0.69) and Z-table (3.055) we see that Zcal < Ztable applying rule3 of the Rejection rules because we are using not equal sign in alternative hypothesis, therefore:

**Decision:** we do not reject the null hypothesis.

**Conclusion:** we do not have sufficient evidence to state that the mean lifetime of the bulbs produced by the company is not 506days.

Note: we compare with 0.69 and not -0.69 as it is stated in rule3 that we should compare with absolute value

**Example 6.1.6**

The farmer claims that the weight of the yam on his farm on average is not less than 0.5 kg. A buyer took a random sample of 14 yams from the farm and observed. After a few observations, the buyer believes it shouldn’t up to 0.5kg. if the weight of the samples are 0.5kg, 0.4kg, 0.55kg, 0.3kg, 0.45kg, 0.45kg, 0.35kg, 0.6kg, 0.5kg, 0.55kg, 0.4kg, 0.5kg, 0.4kg and 0.4kg. Carry out a hypothesis test on the claim.

**Solution**

Remember the farmer claim not less than 0.5 i.e it is either 0.5 or greater this will be our null hypothesis and the alternative will be that it is less than that.

Null: the weight of the yam on his farm on average is not less than 0.5 kg

Alternative: the weight of the yam on his farm on average is less than 0.5 kg

Mathematically

H0: μ ≥ 0.5kg

Ha: μ <0.5kg

**Step2**

α=5% = 0.05

Since the sample size (14) is less than 30 then we will need to use T-table also it must be one tail test since the alternative test is with less than sign.

**step3**

**H**ere the mean is not given but we are given the data itself hence we need to calculate the mean and standard deviation by ourselves.

For Mean

x̄=

x̄=6.3/14=0.45kg

For standard Deviation

|  |  |  |  |
| --- | --- | --- | --- |
|  | x | x-x | (x-x)^2 |
|  | 0.5 | 0.05 | 0.0025 |
|  | 0.4 | -0.05 | 0.0025 |
|  | 0.55 | 0.1 | 0.01 |
|  | 0.3 | -0.15 | 0.0225 |
|  | 0.45 | 0 | 0 |
|  | 0.45 | 0 | 0 |
|  | 0.35 | -0.1 | 0.01 |
|  | 0.6 | 0.15 | 0.0225 |
|  | 0.5 | 0.05 | 0.0025 |
|  | 0.5 | 0.05 | 0.0025 |
|  | 0.4 | -0.05 | 0.0025 |
|  | 0.5 | 0.05 | 0.0025 |
|  | 0.4 | -0.05 | 0.0025 |
|  | 0.4 | -0.05 | 0.0025 |
| sum | 6.3 | 0E+00 | 0.085 |

sd=0.078

Sd= 0.0781.

Now our parameters has been completed.

μ = 0.5kg σ = 0.0781, n= 14, x̄= 0.45kg

Zcal= (x̄-μ )/(σ/√n)

Zcal= (0.45-0.5/ (0.0781/√14)

Zcal= -0.05/0.0209

Zcal=- -2.3923

**Step4**

α= 5% = 0.05 (since the level of significant is not given.)

T0.05 (since we are dealing with two tails test)

Now To get the value from the table we need the degree of freedom and the level of significance

Recall that df = n-1

df =14-1 =13

We need to look for the value of 13 under 0.05 or 5% from T-distribution table and that gives 4.221

**Step5**

Comparing the Zcal (0.69) and Z-table (1.771) we see that Zcal < Z-table. By applying rule2 of the Rejection rules therefore:

**Decision:** we do not reject the null hypothesis.

**Conclusion:** we do not have sufficient evidence to state that the mean lifetime of the bulbs produced by the company is not 506days.

**Note:** we compare with 0.69 and not -0.69 as it is stated in rule3 that we should compare with absolute value.

**Example6.1.7**

Doctor believes that the average teen sleeps on average no longer than 10hrs per day. A researcher believes that teens on average sleep longer. To prove the claim the researcher sample 40 teens and found out that they sleep for 9hrs on average and their standard deviation is 2. Is the doctor claim right at 1% level of significant??

**Solution**

**Null:** the average teen sleeps on average no longer than 10hrs per day

**Alternative**; the average teen sleeps on average no longer than 10hrs per day

Notice since it is not stated that “no longer than 10”, meaning it should be less than or equal to 10.

**Mathematically**

H0: μ ≤ 10hrs per day

Ha: μ > 10hr per day

**Step2**

α=1% = 0.01

Since the sample size is more than 30 then we will need to use Z-table also it must be one tails test since the alternative test is with greater than to sign.

**step3**

μ = 10, σ = 2, n= 40, x̄= 9

= -3.1626

Zcal=-3.1626

**Step4**

α= 1% = 0.01 (since the level of significant is not given.)

Zα=Z0.01 (since we are dealing with one tails test)

We need to look for the value of 0.01 from the normal distribution table and that gives 2.326

**Step5**

Comparing the Zcal (3.1626) and Z-table (2.326) we see that Zcal >Ztable using Rejection rules

**Decision:** we do not reject the null hypothesis.

**Conclusion:** we do not have sufficient evidence to state that the average score of the students in major courses is less than 65

**6.2 Testing for proportion of a population**

**Example1**

In a simple random sample of 600 students taken from a university, 400 are found to be smokers. In another simple random sample of 900students from another university 450 are smokers Do the data indicate that there is a significant difference in the habit of smoking in the two universities.

**Solution**

**Here** we are dealing with proportion from two different population, It is important to note that whenever two proportion has no significance difference they must be the same i.e 0.5 or 50%.

**Step1**.

Null: there is no significant difference in the habit of smoking in the two universities i.e they are both 0.5

Alternative: there is a significant difference in the habit of smoking in the two universities i.e they are not 0.5

**Mathematically**

H0: P = 0.5

Ha: P≠ 0.5

**Step2**

α=0.5 (whenever it is not given we use this)

Since the sample size is up to 30 then we will need to use Z-table also it must be one tails test since the alternative test is with less than to sign.

Now we need to proceed to step3

**49**

**Step3**

N1=600, smoker1=400, P1=400/600=0.6667 (from the first population)

N2=900, smoker2=450, P2=450/900=0.5 (from the second population)

Since we are dealing with two populations given the proportion we need to use the fourth formula stated above.

Firstly we have to calculate the P

P= (600\*0.6667+0.5\*900)/ (900+600) = 850/1500

P=0.5667

6.36

**51**

**Step4**

α= 5% = 0.05 (since the level of significant is not given.)

(Since we are dealing with two tails test)

We need to look for the value of 0.025 from the normal distribution table and that give 1.96

**Step5**

Comparing the Zcal(0.8165) and Ztable(1.96) we see that Zcal >Ztable which follows rule3 of the Rejection rules therefore :

**Decision:** we reject the null hypothesis.

**Conclusion**: at 5% level of significant we can say that there is enough evidence to conclude that there is a significant difference in the habit of smoking between the two universities.

**Example 6.2.1**

School board claims that at least 60% of student brings a phone to school. A teacher believes this number is too high and randomly selected 25students to test at a level of significance of 0.02 if 15 of them brings phone to school. Does the school claim likely to be truth at 99% confident??

**Solution**

Here the percentage given is our proportions i.e that 60% means 60/100 which is 0.6. The following are our hypothesis

Null : at least 60% of student brings a phone to school i.e starting from 60% upward which means it can be 60% or greater

Alternative: it is less than 60% (not at least 60% of student brings a phone to school)

H0: P ≥ 0.6  
Ha: P < 0.6

**Step2**

α=1-0.99=0.01 (whenever it is not given we use this)

From the question, we have n=25 and α=0.05

**To determine the distribution to be used, we check n×P and n×(1−p).**

For the value of p, we use the claim from the null hypothesis (p=0.6).

n×p=25×0.6=15>5

n×(1−p)=25×(1−0.6)=10>5

Since both n×p≥5 and n(1−p)≥5 we use a normal distribution to calculate the critical value.

also it must be one tails test since the alternative test is with less than to sign.

Now we need to proceed to step3

**Step3**

n=25, P=15/25 =0.6(since 15 brings phone out of 25), α=0.01, Po=0.6

Zcal=0

23

**Step4**

α= 1% = 0.01

We need to look for the value 0.01 from the tabe and that gives 2.325

**Step5**

Comparing the Zcal(0) and Ztable(2.325) we see that Zcal <Ztable By applying rule1 of the Rejection rules

**Decision:** we do not reject the null hypothesis.

**Conclusion**: at 99% confidence level, we don’t have enough evidence to state that the school board claim is unlikely to be truth. In fact seeing the result of the test there is high chances that the school claims is truth.

**Example 6.2.2**

A new flu vaccine claims to prevent a certain type of flu in 75% of people who are vaccinated. If out of 50 people tested 40 of them were prevented is this claim too high at 2% level of significance??

**Solution**

Step1.

**Null:** it prevents 75%

**Alternative:** 75% is too high i.e. it should be less than 75%

**Mathematically**

H0: P = 0.75

Ha: P < 0.75

**Step2**

α=0.5 (whenever it is not given we use this)

**To determine the distribution to be used, we check n×P and n×(1−p).**

For the value of p, we use the claim from the null hypothesis (p=0.75).

n×p=50×0.75=37.5>5

n×(1−p)=50×(1−0.75)=12.5>5

Since both n×p≥5 and n(1−p)≥5 we use a normal distribution to calculate the critical value.

also it must be one tails test since the alternative test is with less than to sign.

Now we need to proceed to step3

**Step3**

Po = 0.75, n = 50, P=40/50=0.8

Zcal= (P- Po)/ √(Po(1-Po)/n)

Zcal= (0.8-0.75)/√ (0.75(1-0.75)/50)

Zcal= (0.05)/√ (0.75\*0.25/50)

Zcal= 0.05/0.0612

Zcal=0.8165

**Step4**

α= 2% = 0.02 (since the level of significant is not given.)

Zα=Z0.02 (since we are dealing with one tails test)

We need to look for the value of 0.02 from the normal distribution table and that give 2.05

**Step5**

Comparing the Zcal (0.8165) and Ztable (2.05) we see that Zcal < Ztable which follows rule2 of the Rejection rules therefore:

**Decision:** we reject the null hypothesis.

**Conclusion**: There is enough evidence that the claim is high.

**Example 6.2.3**

Prof Kuye believes that 40% of first-time brides in the Nigeria are younger than their grooms.  He performs a hypothesis test to determine if the percentage is the same or different from 40%.  He samples 100 first-time brides and 50 reply that they are younger than their grooms.  Use a 5% significance level.

**Solution**

**Hypotheses:**

H0: 40% of first-time brides are younger than the groom

Ha: it is not 40% of first-time brides that are younger than the groom

H0: P=40%

Ha: ≠40%

**To determine the distribution to be used, we check n×P and n×(1−p).**

From the question, we have n=100 and α=0.05

To determine the distribution, we check n×P and n×(1−p).

For the value of p, we use the claim from the null hypothesis (p=0.4).

n×p=100×0.4=40 > 5

n ×(1−p)=100×(1−0.4)=60>5

Since both n×p≥5 and n(1−p)≥5 we use a normal distribution table will be used.

n=25, P=15/25 =0.6 (since 15 brings phone out of 25), α=0.01, Po=0.6

Zcal =2

**Step4**

α= 5% so, α/2= 0.05/2 =0.25

We need to look for the value 0.025 from the table and that gives 1.96

**Step5**

Comparing the Zcal (2) and Z table (1.96) we see that Zcal > Z **table** By applying rejection rule

We need to reject the null

**Decision:** we reject the null hypothesis

**Conclusion:** .  At the 5% significance level there is not enough evidence to suggest that the proportion of first-time brides that are younger than the groom is different from 50%.

**Note:**  It should be noted that we can also make the decision using P-value like below.

We check the P(Z<2) from the table we get 0.0228. Since it is two tail test we will multiply the result by 2. So Pvalue= 0.0228\*2 = 0.0456.

Now comparing the Pvalue and level of significant we see that **Pvalue (0.0456) < α (0.05)** Hence **we reject the Null hypothesis.**

**Example 6.2.4**

A group of researchers claimed that not less than 80% of adults in a community have BMI greater than 28kg/m2. A young Professor believes that this is not true and sampled 50peoples in the community then take their measurement. The following table shows the summary of all the measurements.

|  |  |
| --- | --- |
| BMI(kg/m2) | No of People |
| 0 – 13 | 10 |
| 14 – 27 | 10 |
| 28 – 41 | 30 |

**Solution**

Hypothesis: H0: P ≥ 80%, Ha: P < 80%

We need to get the proportion of adults that have BMI greater than 28kg/m2 from the table.

No of People greater than 28kg/m2= 30

Proportion of People greater than 28kg/m2= 30/50 =0.6

**Step2**: α=0.05

**To determine the distribution to be used, we check n×P and n×(1−p).**

From the question, we have n=100

For the value of p, we use the claim from the null hypothesis (p=8%=0.8).

n×p=50×0.8=40 > 5

n ×(1−p)=50×(1−0.8)=10>5

Since both n×p≥5 and n(1−p)≥5 we use a normal distribution table will be used.

**Step3**

n=50, P=30/50 =0.6 (since 15 brings phone out of 25)

**45**

α= 5% so, α/2= 0.05/2 =0.25

We need to look for the value 0.025 from the table and that gives 1.96

**Step5**

Comparing the Zcal (-5) and Z table () we see that Zcal > Z **table** By applying rejection rule

We need to reject the null

**Decision:** we reject the null hypothesis

**Conclusion:**  At the 1% significance level there is not enough evidence to suggest that the proportion of first-time brides that are younger than the groom is different from 50%.

**Using P-value like below.**

We check the P(Z<-5) from the table we get 0.0228. Since it is two tail test we will multiply the result by 2. So Pvalue= 0.0228\*2 = 0.0456.

Now comparing the Pvalue and level of significant we see that **Pvalue (0.0456) < α (0.05)** Hence **we reject the Null hypothesis.**

**Example 6.2.4**

A health magazine claims that marriage status is one of the most telling factors for a person's happiness. Use the table below to test the claim that the percent of married people that are unhappy is lower than the percent of single or divorced people that are unhappy. The data was collected randomly. (Use a 10% significance level.)

|  |  |  |
| --- | --- | --- |
|  | Married | Single or Divorced |
| Happy | 75 | 90 |
| Unhappy | 125 | 110 |
| Total | 200 | 200 |

.

A tattoo magazine claimed that the percent of men that have at least one tattoo is greater than the percent of women with at least one tattoo. Test this claim with the following sample data. A random sample of 794 women found that 137 of them had at least one tattoo. A random sample of 502 men found that 110 of them had at least one tattoo. (Use a 5% significance level.)

Data from the Center for Disease Control estimates that about 30.4% of American teenagers were overweight in 2008. The definition of overweight is a body mass index (BMI) of over 25. The percentage was very similar for boys and girls.

A professor in public health at a major university wants to determine whether the proportion has changed since 2008. He samples 800 randomly selected incoming freshman at universities around the country. Using the BMI measurements, he finds that 210, or about 26%, of them are overweight.

The professor tests the hypotheses H0: p = 0.304 versus Ha: p ≠ 0.304. The P-value is about 0.011. If the professor uses a significance level of 0.05, what conclusion can he draw?

.

**Example 6.2.4**

A health magazine claims that marriage status is one of the most telling factors for a person's happiness. Use the table below to test the claim that the percent of married people that are unhappy is lower than the percent of single or divorced people that are unhappy. The data was collected randomly. (Use a 10% significance level.)

|  |  |  |
| --- | --- | --- |
|  | Married | Single or Divorced |
| Happy | 75 | 90 |
| Unhappy | 125 | 110 |
| Total | 200 | 200 |

.

A tattoo magazine claimed that the percent of men that have at least one tattoo is greater than the percent of women with at least one tattoo. Test this claim with the following sample data. A random sample of 794 women found that 137 of them had at least one tattoo. A random sample of 502 men found that 110 of them had at least one tattoo. (Use a 5% significance level.)

Data from the Center for Disease Control estimates that about 30.4% of American teenagers were overweight in 2008. The definition of overweight is a body mass index (BMI) of over 25. The percentage was very similar for boys and girls.

A professor in public health at a major university wants to determine whether the proportion has changed since 2008. He samples 800 randomly selected incoming freshman at universities around the country. Using the BMI measurements, he finds that 210, or about 26%, of them are overweight.

The professor tests the hypotheses H0: p = 0.304 versus Ha: p ≠ 0.304. The P-value is about 0.011. If the professor uses a significance level of 0.05, what conclusion can he draw?

.

Newborn babies are more likely to be boys than girls. A random sample found 13,173 boys were born among 25,468 newborn children. The sample proportion of boys was 0.5172. Is this sample evidence that the birth of boys is more common than the birth of girls in the entire population?

Solution

Here, we want to test

The test statistic

We will reject the null hypothesis or equivalently if Z > 1.645

Here's a picture of such a "critical region" (or "rejection region"):

It looks like we should reject the null hypothesis because:

or equivalently since our test statistic Z = 5.49 is greater than 1.645.

**Our Conclusion**: We say there is sufficient evidence to conclude boys are more common than girls in the entire population.

1. (10 points) On July 8th, 2004, the U.S. Census Bureau reported on their "population clock" that the population of the United States was 293,683,456 people. The National Endowment of the Arts conducted a survey — titled "Reading at Risk" — on the reading habits of approximately 17,000 adults. Of those surveyed, only 57% read a book in 2002. a) What is the population under investigation? b) What is the parameter of interest? c) What is the sample? d) What is the sample statistic?

2. (10 points) Most — if not all — statistical software have t-distribution probability calculators that allow you to determine the appropriate t-multiplier when finding a confidence interval for a population mean µ. It entails telling the software the cumulative probability — that is, the probability to the left of the tmultiplier — and the degrees of freedom. Use Minitab to determine the appropriate t-multiplier for calculating a confidence interval for a population mean µ in the following situations. a) For a 95% confidence interval based on a sample size of n = 15. b) For a 95% confidence interval based on a sample size of n = 26. c) For a 91% confidence interval based on a sample size of n = 15.

3. (10 points) In order to monitor the weight changes in a herd of calves, six randomly chosen calves were weighed initially, individually identified with ear tags, and then weighed again one month later. The resulting data were analyzed in Minitab: a) Use the reported sample mean and sample standard deviation, Minitab's t-distribution calculator and the formula for the confidence interval for a population mean µ to verify the confidence interval reported by Minitab. b) Interpret the interval.

4. (10 points) The Center for Disease Control had determined that the mean weight of all eleven-year old Caucasian boys is 88 pounds. A complaint is made that 25 such boys living in a county children's home are underfed. As one piece of evidence, the boys weighed an average of 83 pounds with a standard deviation of 10 pounds. a) State an appropriate null and alternative hypothesis for this situation. b) Choose an appropriate significance level α for this situation. c) Perform the appropriate hypothesis test using the critical value approach. In doing so, use Minitab's t-distribution calculator to find the critical value. d) Perform the appropriate hypothesis test using the P-value approach. In doing so, use Minitab's tdistribution calculator to find the P-value.

5. (10 points) Suppose a random sample of 216 patients that have a skin disease are classified into the four age categories. The frequencies are summarized in the following table: Age category 1 2 3 4 Total Severity Moderate 15 32 18 5 70 mildly severe 8 29 23 18 78 severe 1 20 25 22 68 Total 24 81 66 45 216 Conduct a test to determine if the severity of the disease is independent of the age of the patient.

6. (10 points) suppose the Pennsylvania population is 55% female and 45% male. Then, if a sample of 100 persons yields 53 females and 47 males, can we conclude that the sample is (random and) representative of the population?

7. (10 points) Let X denote the weight loss of women after taking an exercise program. Assume that X is normally distributed with unknown mean μ and standard deviation 4. Weight loss of 14 women are recorded. To see if the women dose lose their weight, we can test the null hypothesis H0: μ = 0 against the alternative hypothesis that HA: μ >0. What is the power of the hypothesis test if the true population mean were μ = 5? (α=0.05)

8. (10 points) In Problem 7, Find the sample size n that is necessary to achieve 0.90 power at the alternative μ = 5.

9. One of the largest problems on college campuses is alcohol abuse by underage students. Universities are acutely aware of the problem of binge drinking, defined as consuming five or more drinks in a row three or more times in a two-week period. An extensive survey of college students reported that 44% of U.S. college students engaged in binge drinking during the two weeks before the survey. The president of a university stated publicly that binge drinking was not a problem on her campus of 25000 undergraduate students. A survey was conducted on 2500 undergraduate students attending the university. The survey showed that 1200 of the 2500 students had engaged in binge drinking. Is there sufficient evidence to indicate that the percentage of students engaging in binge drinking at the university is greater than the percentage found in national survey?( Use Critical Value Approach, α=0.05)

10. (10 points) A sample survey funded by National Science Foundation asked a random sample of American adults about biological evolution. One question asked subjects to answer “True” or “False” to the statement “Human beings, as we know them today, developed from earlier species of animals.” Of the 1484 respondents, 713 said “True”. Does the sample give good evidence to support the claim “Larger than half of American adults think that humans developed from earlier species?

Marketers believe that 92% of adults own a cell phone.  A cell phone manufacturer believes that number is actually lower.  In a sample of 200 adults, 87% own a cell phone.  At the 1% significance level, determine if the proportion of adults that own a cell phone is lower than the marketers’ claim.

***Solution:***

**Hypotheses:**

H0:p=92% of adults own a cell phone

Ha:p<92% of adults own a cell phone

p***-* value:**

From the question, we have n=200�=200, ^p=0.87�^=0.87, and α=0.01�=0.01.

To determine the distribution, we check n×p�×� and n×(1−p)�×(1−�).  For the value of p�, we use the claim from the null hypothesis (p=0.92�=0.92).

n×p=200×0.92=184≥5n×(1−p)=

200×(1−0.92)=16≥5�×�=

200×0.92=18×(1−�)

=200×(1−0.92)=16≥5

Because both n×p≥5�×�≥5 and n×(1−p)≥5�×(1−�)≥5 we use a normal distribution to calculate the p-value.  Because the alternative hypothesis is a <<, the p-value is the area in the left tail of the distribution.

|  |  |  |
| --- | --- | --- |
| **Function** | norm.dist | **Answer** |
| **Field 1** | 0.87 | 0.0046 |
| **Field 2** | 0.92 |  |
| **Field 3** | sqrt(0.92\*(1-0.92)/200) |  |
| **Field 4** | True |  |

So the p-value=0.0046=0.0046.

**Conclusion:**

Because p-value=0.0046<0.01=α=0.0046<0.01=�, we reject the null hypothesis in favour of the alternative hypothesis.  At the 1% significance level there is enough evidence to suggest that the proportion of adults who own a cell phone is lower than 92%.

### NOTES

1. The null hypothesis p=92%�=92% is the claim that 92% of adults own a cell phone.
2. The alternative hypothesis p<92%�<92% is the claim that less than 92% of adults own a cell phone.
3. The p-value is the area in the left tail of the sampling distribution, to the left of ^p=0.87�^=0.87.  In the calculation of the p-value:
   * The function is norm.dist because we are finding the area in the left tail of a normal distribution.
   * Field 1 is the value of ^p�^.
   * Field 2 is the value of p� from the null hypothesis.  Remember, we run the test assuming the null hypothesis is true, so that means we assume p=0.92�=0.92.
   * Field 3 is the standard deviation for the sample proportions √p×(1−p)n�×(1−�)�.
4. The p-value of 0.0046 tells us that under the assumption that 92% of adults own a cell phone (the null hypothesis), there is only a 0.46% chance that the proportion of adults who own a cell phone in a sample of 200 is 87% or less.  This is a small probability, and so is unlikely to happen assuming the null hypothesis is true.  This suggests that the assumption that the null hypothesis is true is most likely incorrect, and so the conclusion of the test is to reject the null hypothesis in favour of the alternative hypothesis.  In other words, the proportion of adults who own a cell phone is most likely less than 92%.

## EXAMPLE

A consumer group claims that the proportion of households that have at least three cell phones is 30%.  A cell phone company has reason to believe that the proportion of households with at least three cell phones is much higher.  Before they start a big advertising campaign based on the proportion of households that have at least three cell phones, they want to test their claim.  Their marketing people survey 150 households with the result that 54 of the households have at least three cell phones.  At the 1% significance level, determine if the proportion of households that have at least three cell phones is less than 30%.

***Solution:***

**Hypotheses:**

H0:p=30% of household have at least 3 cell phonesHa:p>30% of household have at least 3 cell phones�0:�=30% of household have at least 3 cell phones��:�>30% of household have at least 3 cell phones

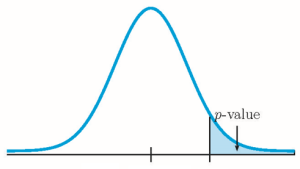
**p-value:**

From the question, we have n=150�=150, ^p=54150=0.36�^=54150=0.36, and α=0.01�=0.01.

To determine the distribution, we check n×p�×� and n×(1−p)�×(1−�).  For the value of p�, we use the claim from the null hypothesis (p=0.3�=0.3).

n×p=150×0.3=45≥5n×(1−p)=150×(1−0.3)=105≥5�×�=150×0.3=45≥5�×(1−�)=150×(1−0.3)=105≥5

Because both n×p≥5�×�≥5 and n×(1−p)≥5�×(1−�)≥5 we use a normal distribution to calculate the p-value.  Because the alternative hypothesis is a >>, the p-value is the area in the right tail of the distribution.

[](https://ecampusontario.pressbooks.pub/app/uploads/sites/2831/2022/08/right-tail-normal-distribution.png)

|  |  |  |
| --- | --- | --- |
| **Function** | 1-norm.dist | **Answer** |
| **Field 1** | 0.36 | 0.0544 |
| **Field 2** | 0.3 |  |
| **Field 3** | sqrt(0.3\*(1-0.3)/150) |  |
| **Field 4** | True |  |

So the p-value=0.0544=0.0544.

**Conclusion:**

Because p-value=0.0544>0.01=α=0.0544>0.01=�, we do not reject the null hypothesis.  At the 1% significance level there is not enough evidence to suggest that the proportion of households with at least three cell phones is more than 30%.

### NOTES

1. The null hypothesis p=30%�=30% is the claim that 30% of households have at least three cell phones.
2. The alternative hypothesis p>30%�>30% is the claim that more than 30% of households have at least three cell phones.
3. The p-value is the area in the right tail of the sampling distribution, to the right of ^p=0.36�^=0.36.  In the calculation of the p-value:
   * The function is 1-norm.dist because we are finding the area in the right tail of a normal distribution.
   * Field 1 is the value of ^p�^.
   * Field 2 is the value of p� from the null hypothesis.  Remember, we run the test assuming the null hypothesis is true, so that means we assume p=0.3�=0.3.
   * Field 3 is the standard deviation for the sample proportions √p×(1−p)n�×(1−�)�.
4. The p-value of 0.0544 tells us that under the assumption that 30% of households have at least three cell phones (the null hypothesis), there is a 5.44% chance that the proportion of households with at least three cell phones in a sample of 150 is 36% or more.  Compared to the 1% significance level, this is a large probability, and so is likely to happen assuming the null hypothesis is true.  This suggests that the assumption that the null hypothesis is true is most likely correct, and so the conclusion of the test is to not reject the null hypothesis.  In other words, the claim that 30% of households have at least three cell phones is most likely correct.

A teacher believes that 70% of students in the class will want to go on a field trip to the local zoo.  The students in the class believe the proportion is much higher and ask the teacher to verify her claim.  The teacher samples 50 students and 39 reply that they would want to go to the zoo.  At the 5% significance level, determine if the proportion of students who want to go on the field trip is higher than 70%.

**Click to see Solution**

**Hypotheses:**

H0:p=70% of students want to go on the field tripHa:p>70% of students want to go on the field trip�0:�=70% of students want to go on the field trip��:�>70% of students want to go on the field trip

**p-value:**

From the question, we have n=50�=50, ^p=3950=0.78�^=3950=0.78, and α=0.05�=0.05.

n×p=50×0.7=35≥5n×(1−p)=50×(1−0.7)=15≥5�×�=50×0.7=35≥5�×(1−�)=50×(1−0.7)=15≥5

Because both n×p≥5�×�≥5 and n×(1−p)≥5�×(1−�)≥5 we use a normal distribution to calculate the p-value.  Because the alternative hypothesis is a >>, the p-value is the area in the right tail of the distribution.

|  |  |  |
| --- | --- | --- |
| **Function** | 1-norm.dist | **Answer** |
| **Field 1** | 0.78 | 0.1085 |
| **Field 2** | 0.7 |  |
| **Field 3** | sqrt(0.7\*(1-0.7)/50) |  |
| **Field 4** | true |  |

So the p-value=0.1085=0.1085.

**Conclusion:**

Because p-value=0.1085>0.05=α=0.1085>0.05=�, we do not reject the null hypothesis.  At the 5% significance level there is not enough evidence to suggest that the proportion of students who want to go on the field trip is higher than 70%.

### NOTES

1. The null hypothesis p=70%�=70% is the claim that 70% of the students want to go on the field trip.
2. The alternative hypothesis p>70%�>70% is the claim that more than 70% of students want to go on the field trip.
3. The p-value of 0.1085 tells us that under the assumption that 70% of students want to go on the field trip (the null hypothesis), there is a 10.85% chance that the proportion of students who want to go on the field trip in a sample of 50 students is 78% or more.  Compared to the 5% significance level, this is a large probability, and so is likely to happen assuming the null hypothesis is true.  This suggests that the assumption that the null hypothesis is true is most likely correct, and so the conclusion of the test is to not reject the null hypothesis.  In other words, the teacher’s claim that 70% of students want to go on the field trip is most likely correct.

## EXAMPLE

Joan believes that 50% of first-time brides in the United States are younger than their grooms.  She performs a hypothesis test to determine if the percentage is the same or different from 50%.  Joan samples 100 first-time brides and 56 reply that they are younger than their grooms.  Use a 5% significance level.

***Solution:***

**Hypotheses:**

H0:p=50% of first-time brides are younger than the groomHa:p≠50% of first-time brides are younger than the groom�0:�=50% of first-time brides are younger than the groom��:�≠50% of first-time brides are younger than the groom

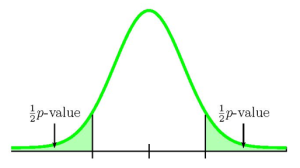
**p-value:**

From the question, we have n=100�=100, ^p=56100=0.56�^=56100=0.56, and α=0.05�=0.05.

To determine the distribution, we check n×p�×� and n×(1−p)�×(1−�).  For the value of p�, we use the claim from the null hypothesis (p=0.5�=0.5).

n×p=100×0.5=50≥5n×(1−p)=100×(1−0.5)=50≥5�×�=100×0.5=50≥5�×(1−�)=100×(1−0.5)=50≥5

Because both n×p≥5�×�≥5 and n×(1−p)≥5�×(1−�)≥5 we use a normal distribution to calculate the p-value.  Because the alternative hypothesis is a ≠≠, the p-value is the sum of area in the tails of the distribution.

[](https://ecampusontario.pressbooks.pub/app/uploads/sites/2831/2022/08/two-tail-normal-distribution.png)

Because there is only one sample, we only have information relating to one of the two tails, either the left or the right.  We need to know if the sample relates to the left or right tail because that will determine how we calculate out the area of that tail using the normal distribution.  In this case, the sample proportion ^p=0.56�^=0.56 is greater than the value of the population proportion in the null hypothesis p=0.5�=0.5 (^p=0.56>0.5=p�^=0.56>0.5=�), so the sample information relates to the right-tail of the normal distribution.  This means that we will calculate out the area in the right tail using **1-norm.dist**.  However, this is a two-tailed test where the p-value is the sum of the area in the two tails and the area in the right-tail is only one half of the p-value.  The area in the left tail equals the area in the right tail and the p-value is the sum of these two areas.

|  |  |  |
| --- | --- | --- |
| **Function** | 1-norm.dist | **Answer** |
| **Field 1** | 0.56 | 0.1151 |
| **Field 2** | 0.5 |  |
| **Field 3** | sqrt(0.5\*(1-0.5)/100) |  |
| **Field 4** | true |  |

So the area in the right tail is 0.1151 and  1212(p-value)=0.1151=0.1151.  This is also the area in the left tail, so

p-value=0.1151+0.1151=0.2302=0.1151+0.1151=0.2302

**Conclusion:**

Because p-value=0.2302>0.05=α=0.2302>0.05=�, we do not reject the null hypothesis.  At the 5% significance level there is not enough evidence to suggest that the proportion of first-time brides that are younger than the groom is different from 50%.

### NOTES

1. The null hypothesis p=50%�=50% is the claim that the proportion of first-time brides that are younger than the groom is 50%.
2. The alternative hypothesis p≠50%�≠50% is the claim that the proportion of first-time brides that are younger than the groom is different from 50%.
3. In a two-tailed hypothesis test that uses the normal distribution, we will only have sample information relating to **one** of the two tails.  We must determine which of the tails the sample information belongs to, and then calculate out the area in that tail.  The area in each tail represents exactly half of the p-value, so the p-value is the sum of the areas in the two tails.
   * If the sample proportion ^p�^ is less than the population proportion p� in the null hypothesis (^p<p�^<�), the sample information belongs to the**left tail**.
     + We use **norm.dist(**^p�^**,**p�**,**sqrt(p∗(1−p)/n)sqrt(�∗(1−�)/�)**,true)** to find the area in the left tail.  The area in the right tail equals the area in the left tail, so we can find the p-value by adding the output from this function to itself.
   * If the sample proportion ^p�^ is greater than the population proportion p� in the null hypothesis (^p>p�^>�), the sample information belongs to the**right tail**.
     + We use **1-norm.dist(**^p�^**,**p�**,**sqrt(p∗(1−p)/n)sqrt(�∗(1−�)/�)**,true)** to find the area in the right tail.  The area in the left tail equals the area in the right tail, so we can find the p-value by adding the output from this function to itself.
4. The p-value of 0.2302  is a large probability compared to the 5% significance level, and so is likely to happen assuming the null hypothesis is true.  This suggests that the assumption that the null hypothesis is true is most likely correct, and so the conclusion of the test is to not reject the null hypothesis.  In other words, the claim that the proportion of first-time brides who are younger than the groom is most likely correct.

Watch this video:[Hypothesis Testing for Proportions: z-test](https://www.youtube.com/watch?v=yWVbkJJPDeA&list=PLrRPvpgDmw0m3oqpp1XcPuaxyM4Bpi0dN&index=53)by ExcelIsFun [7:27]

## EXAMPLE

An online retailer believes that 93% of the visitors to its website will make a purchase.   A researcher in the marketing department thinks the actual percent is lower than claimed.  The researcher examines a sample of 50 visits to the website and finds that 45 of the visits resulted in a purchase.  At the 1% significance level, determine if the proportion of visits to the website that result in a purchase is lower than claimed.

***Solution:***

**Hypotheses:**

H0:p=93% of visitors make a purchaseHa:p<93% of visitors make a purchase�0:�=93% of visitors make a purchase��:�<93% of visitors make a purchase

**p-value:**

From the question, we have n=50�=50, x=45�=45, and α=0.01�=0.01.

To determine the distribution, we check n×p�×� and n×(1−p)�×(1−�).  For the value of p�, we use the claim from the null hypothesis (p=0.93�=0.93).

n×p=50×0.93=46.5≥5n×(1−p)=50×(1−0.93)=3.5<5�×�=50×0.93=46.5≥5�×(1−�)=50×(1−0.93)=3.5<5

Because n×(1−p)<5�×(1−�)<5 we use a binomial distribution to calculate the p-value.  Because the alternative hypothesis is a <<, the p-value is the probability of getting at most 45 successes in 50 trials.

|  |  |  |
| --- | --- | --- |
| **Function** | binom.dist | **Answer** |
| **Field 1** | 45 | 0.2710 |
| **Field 2** | 50 |  |
| **Field 3** | 0.93 |  |
| **Field 4** | true |  |

So the p-value=0.2710=0.2710.

**Conclusion:**

Because p-value=0.2710>0.01=α=0.2710>0.01=�, we do not reject the null hypothesis.  At the 1% significance level there is not enough evidence to suggest that the proportion of visitors who make a purchase is lower than 93%.

### NOTES

1. The null hypothesis p=93%�=93% is the claim that 93% of visitors to the website make a purchase.
2. The alternative hypothesis p<93%�<93% is the claim that less than 93% of visitors to the website make a purchase.
3. The p-value is the binomial probability of getting at most 45 successes (the number in the sample with the characteristic of interest) in 50 trials (the sample size) with a probability of success of 93% (the value of p� in the null hypothesis).  In the calculation of the p-value:
   * The function is binom.dist because we are finding the probability of at most 45 successes.
   * Field 1 is the number of successes x�.
   * Field 2 is the sample size n�.
   * Field 3 is the probability of success p�.  This is the claim about the population proportion made in the null hypothesis, so that means we assume p=0.93�=0.93.
4. The p-value of 0.2710 tells us that under the assumption that 93% of visitors make a purchase (the null hypothesis), there is a 27.10% chance that the number of visitors in a sample of 50 who make a purchase is 45 or less.  This is a large probability compared to the significance level, and so is likely to happen assuming the null hypothesis is true.  This suggests that the assumption that the null hypothesis is true is most likely correct, and so the conclusion of the test is to not reject the null hypothesis.  In other words, the proportion of visitors to the website who make a purchase adults is most likely 93%.

## EXAMPLE

A drug company claims that only 4% of people who take their new drug experience any side effects from the drug.  A researcher believes that the percent is higher than drug company’s claim.  The researcher takes a sample of 80 people who take the drug and finds that 10% of the people in the sample experience side effects from the drug.  At the 5% significance level, determine if the proportion of people who experience side effects from taking the drug is higher than claimed.

***Solution:***

**Hypotheses:**

H0:p=4% of people experience side effectsHa:p>4% of people experience side effects�0:�=4% of people experience side effects��:�>4% of people experience side effects

**p-value:**

From the question, we have n=80�=80, ^p=0.1�^=0.1, and α=0.05�=0.05.

To determine the distribution, we check n×p�×� and n×(1−p)�×(1−�).  For the value of p�, we use the claim from the null hypothesis (p=0.04�=0.04).

n×p=80×0.04=3.2<5�×�=80×0.04=3.2<5

Because n×p<5�×�<5 we use a binomial distribution to calculate the p-value.  Because the alternative hypothesis is a >>, the p-value is the probability of getting at least 8 successes in 80 trials.  (Note:  In the sample of size 80, 10% have the characteristic of interest, so this means that 80×0.1=880×0.1=8 people in the sample have the characteristic of interest.)

|  |  |  |
| --- | --- | --- |
| **Function** | 1-binom.dist | **Answer** |
| **Field 1** | 7 | 0.0147 |
| **Field 2** | 80 |  |
| **Field 3** | 0.04 |  |
| **Field 4** | true |  |

So the p-value=0.0147=0.0147.

**Conclusion:**

Because p-value=0.0147<0.05=α=0.0147<0.05=�, we reject the null hypothesis in favour of the alternative hypothesis.  At the 5% significance level there is enough evidence to suggest that the proportion of people who experience side effects from taking the drug is higher than 4%.

### NOTES

1. The null hypothesis p=4%�=4% is the claim that 4% of the people experience side effects from taking the drug.
2. The alternative hypothesis p>4%�>4% is the claim that more than 4% of the people experience side effects from taking the drug.
3. The p-value is the binomial probability of getting at least 8 successes (the number in the sample with the characteristic of interest) in 80 trials (the sample size) with a probability of success of 4% (the value of p� in the null hypothesis).  In the calculation of the p-value:
   * The function is 1-binom.dist because we are finding the probability of at least 8 successes.
   * Field 1 is x−1�−1 where x� is the number of successes.  In this case, we are using the compliment rule to change the probability of at least 8 successes into 1 minus the probability of at most 7 successes.
   * Field 2 is the sample size n�.
   * Field 3 is the probability of success p�.  This is the claim about the population proportion made in the null hypothesis, so that means we assume p=0.04�=0.04.
4. The p-value of 0.0147 tells us that under the assumption that 4% of people experience side effects (the null hypothesis), there is a 1.47% chance that the number of people in a sample of 80 who experience side effects is 8 or more.  This is a small probability compared to the significance level, and so is unlikely to happen assuming the null hypothesis is true.  This suggests that the assumption that the null hypothesis is true is most likely incorrect, and so the conclusion of the test is to reject the null hypothesis in favour of the alternative hypothesis.  In other words, the proportion of people who experience side effects is most likely greater than 4%.

### NOTES

1. The null hypothesis p=50%�=50% is the claim that the proportion of first-time brides that are younger than the groom is 50%.
2. The alternative hypothesis p≠50%�≠50% is the claim that the proportion of first-time brides that are younger than the groom is different from 50%.
3. In a two-tailed hypothesis test that uses the normal distribution, we will only have sample information relating to **one** of the two tails.  We must determine which of the tails the sample information belongs to, and then calculate out the area in that tail.  The area in each tail represents exactly half of the p-value, so the p-value is the sum of the areas in the two tails.
   * If the sample proportion ^p�^ is less than the population proportion p� in the null hypothesis (^p<p�^<�), the sample information belongs to the**left tail**.
     + We use **norm.dist(**^p�^**,**p�**,**sqrt(p∗(1−p)/n)sqrt(�∗(1−�)/�)**,true)** to find the area in the left tail.  The area in the right tail equals the area in the left tail, so we can find the p-value by adding the output from this function to itself.
   * If the sample proportion ^p�^ is greater than the population proportion p� in the null hypothesis (^p>p�^>�), the sample information belongs to the**right tail**.
     + We use **1-norm.dist(**^p�^**,**p�**,**sqrt(p∗(1−p)/n)sqrt(�∗(1−�)/�)**,true)** to find the area in the right tail.  The area in the left tail equals the area in the right tail, so we can find the p-value by adding the output from this function to itself.
4. The p-value of 0.2302  is a large probability compared to the 5% significance level, and so is likely to happen assuming the null hypothesis is true.  This suggests that the assumption that the null hypothesis is true is most likely correct, and so the conclusion of the test is to not reject the null hypothesis.  In other words, the claim that the proportion of first-time brides who are younger than the groom is most likely correct.

An online retailer believes that 93% of the visitors to its website will make a purchase.   A researcher in the marketing department thinks the actual percent is lower than claimed.  The researcher examines a sample of 50 visits to the website and finds that 45 of the visits resulted in a purchase.  At the 1% significance level, determine if the proportion of visits to the website that result in a purchase is lower than claimed.

***Solution:***

**Hypotheses:**

H0:p=93% of visitors make a purchaseHa:p<93% of visitors make a purchase�0:�=93% of visitors make a purchase��:�<93% of visitors make a purchase

**p-value:**

From the question, we have n=50�=50, x=45�=45, and α=0.01�=0.01.

To determine the distribution, we check n×p�×� and n×(1−p)�×(1−�).  For the value of p�, we use the claim from the null hypothesis (p=0.93�=0.93).

n×p=50×0.93=46.5≥5n×(1−p)=50×(1−0.93)=3.5<5�×�=50×0.93=46.5≥5�×(1−�)=50×(1−0.93)=3.5<5

Because n×(1−p)<5�×(1−�)<5 we use a binomial distribution to calculate the p-value.  Because the alternative hypothesis is a <<, the p-value is the probability of getting at most 45 successes in 50 trials.

|  |  |  |
| --- | --- | --- |
| **Function** | binom.dist | **Answer** |
| **Field 1** | 45 | 0.2710 |
| **Field 2** | 50 |  |
| **Field 3** | 0.93 |  |
| **Field 4** | true |  |

So the p-value=0.2710=0.2710.

**Conclusion:**

Because p-value=0.2710>0.01=α=0.2710>0.01=�, we do not reject the null hypothesis.  At the 1% significance level there is not enough evidence to suggest that the proportion of visitors who make a purchase is lower than 93%.

### NOTES

1. The null hypothesis p=93%�=93% is the claim that 93% of visitors to the website make a purchase.
2. The alternative hypothesis p<93%�<93% is the claim that less than 93% of visitors to the website make a purchase.
3. The p-value is the binomial probability of getting at most 45 successes (the number in the sample with the characteristic of interest) in 50 trials (the sample size) with a probability of success of 93% (the value of p� in the null hypothesis).  In the calculation of the p-value:
   * The function is binom.dist because we are finding the probability of at most 45 successes.
   * Field 1 is the number of successes x�.
   * Field 2 is the sample size n�.
   * Field 3 is the probability of success p�.  This is the claim about the population proportion made in the null hypothesis, so that means we assume p=0.93�=0.93.
4. The p-value of 0.2710 tells us that under the assumption that 93% of visitors make a purchase (the null hypothesis), there is a 27.10% chance that the number of visitors in a sample of 50 who make a purchase is 45 or less.  This is a large probability compared to the significance level, and so is likely to happen assuming the null hypothesis is true.  This suggests that the assumption that the null hypothesis is true is most likely correct, and so the conclusion of the test is to not reject the null hypothesis.  In other words, the proportion of visitors to the website who make a purchase adults is most likely 93%.

**EXAMPLE(change)**

A drug company claims that only 5% of people who take their new drug experience any side effects from the drug.  A researcher believes that the percent is higher than drug company’s claim.  The researcher takes a sample of 100 people who take the drug and finds that 10% of the people in the sample experience side effects from the drug. Test if the proportion of people who experience side effects from taking the drug is higher than claimed.

***Solution:***

**Hypotheses:**

H0:p=5% of people experience side effects

Ha:p>5% of people experience side effects

p**-value:**

From the question, we have n=100=,

 p=10%=0.1, and α=0.05�=0.05.

To determine the distribution, we check n×p�×� and n×(1−p)�×(1−�).  For the value of p�, we use the claim from the null hypothesis (p=0.04�=0.04).

n×p=80×0.04=3.2<5

Because n×p<5we use a binomial distribution to calculate the p-value.  Because the alternative hypothesis is a >>, the p-value is the probability of getting at least 8 successes in 80 trials.  (Note:  In the sample of size 80, 10% have the characteristic of interest, so this means that 80×0.1=880×0.1=8 people in the sample have the characteristic of interest.)

|  |  |  |
| --- | --- | --- |
| **Function** | 1-binom.dist | **Answer** |
| **Field 1** | 7 | 0.0147 |
| **Field 2** | 80 |  |
| **Field 3** | 0.04 |  |
| **Field 4** | True |  |

So the p-value=0.0147=0.0147.

**Conclusion:**

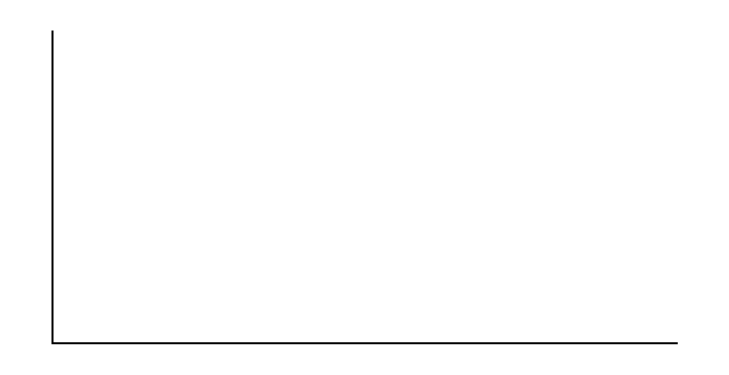
Because p-value=0.0147<0.05=α=0.0147<0.05=�, we reject the null hypothesis in favour of the alternative hypothesis.  At the 5% significance level there is enough evidence to suggest that the proportion of people who experience side effects from taking the drug is higher than 4%.

### NOTES

1. The null hypothesis p=4%�=4% is the claim that 4% of the people experience side effects from taking the drug.
2. The alternative hypothesis p>4%�>4% is the claim that more than 4% of the people experience side effects from taking the drug.
3. The p-value is the binomial probability of getting at least 8 successes (the number in the sample with the characteristic of interest) in 80 trials (the sample size) with a probability of success of 4% (the value of p� in the null hypothesis).  In the calculation of the p-value:
   * The function is 1-binom.dist because we are finding the probability of at least 8 successes.
   * Field 1 is x−1�−1 where x� is the number of successes.  In this case, we are using the compliment rule to change the probability of at least 8 successes into 1 minus the probability of at most 7 successes.
   * Field 2 is the sample size n�.
   * Field 3 is the probability of success p�.  This is the claim about the population proportion made in the null hypothesis, so that means we assume p=0.04�=0.04.
4. The p-value of 0.0147 tells us that under the assumption that 4% of people experience side effects (the null hypothesis), there is a 1.47% chance that the number of people in a sample of 80 who experience side effects is 8 or more.  This is a small probability compared to the significance level, and so is unlikely to happen assuming the null hypothesis is true.  This suggests that the assumption that the null hypothesis is true is most likely incorrect, and so the conclusion of the test is to reject the null hypothesis in favour of the alternative hypothesis.  In other words, the proportion of people who experience side effects is most likely greater than 4%.

**Television Survey**In a recent survey, it was stated that Americans watch television on average four hours per day. Assume that *σ* = 2. Using your class as the sample, conduct a hypothesis test to determine if the average for students at your school is lower.

1. *H0*: \_\_\_\_\_\_\_\_\_\_\_\_\_
2. *Ha*: \_\_\_\_\_\_\_\_\_\_\_\_\_
3. In words, define the random variable. \_\_\_\_\_\_\_\_\_\_ = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. The distribution to use for the test is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.
5. Determine the test statistic using your data.
6. Draw a graph and label it appropriately.Shade the actual level of significance.
   1. Graph:



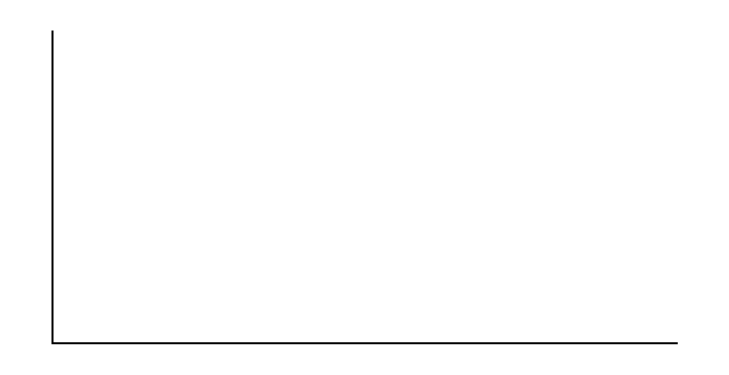
**Figure 9.21**

* 1. Determine the *p*-value.

1. Do you or do you not reject the null hypothesis? Why?
2. Write a clear conclusion using a complete sentence.

**Language Survey** About 42.3% of Californians and 19.6% of all Americans over age five speak a language other than English at home. Using your class as the sample, conduct a hypothesis test to determine if the percent of the students at your school who speak a language other than English at home is different from 42.3%.

1. *H0*: \_\_\_\_\_\_\_\_\_\_\_
2. *Ha*: \_\_\_\_\_\_\_\_\_\_\_
3. In words, define the random variable. \_\_\_\_\_\_\_\_\_\_ = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. The distribution to use for the test is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
5. Determine the test statistic using your data.
6. Draw a graph and label it appropriately. Shade the actual level of significance.
   1. Graph:



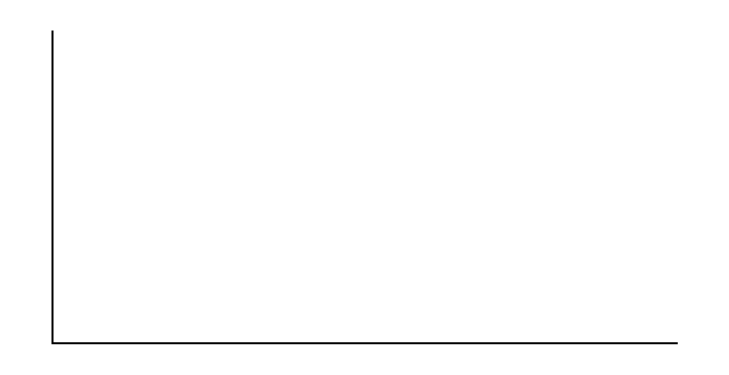
**Figure 9.22**

* 1. Determine the *p*-value.

1. Do you or do you not reject the null hypothesis? Why?
2. Write a clear conclusion using a complete sentence.

**Jeans Survey**Suppose that young adults own an average of three pairs of jeans. Survey eight people from your class to determine if the average is higher than three. Assume the population is normal.

1. *H0*: \_\_\_\_\_\_\_\_\_\_\_\_\_
2. *Ha*: \_\_\_\_\_\_\_\_\_\_\_\_\_
3. In words, define the random variable. \_\_\_\_\_\_\_\_\_\_ = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. The distribution to use for the test is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.
5. Determine the test statistic using your data.
6. Draw a graph and label it appropriately. Shade the actual level of significance.
   1. Graph:



**Figure 9.23**

* 1. Determine the *p*-value.

1. Do you or do you not reject the null hypothesis? Why?
2. Write a clear conclusion using a complete sentence.

**Chi-Square Test of Independence**

Do you remember how to test the independence of two categorical variables? This test is performed by using a Chi-square test of independence.

Recall that we can summarize two categorical variables within a two-way table, also called a *r* × *c* contingency table, where *r* = number of rows, *c* = number of columns. Our question of interest is “Are the two variables independent?” This question is set up using the following hypothesis statements:

**Null Hypothesis**

The two categorical variables are independent

**Alternative Hypothesis**

The two categorical variables are dependent

**Chi-Square Test Statistic**

�2=∑(�−�)2/�

where *O* represents the observed frequency. *E* is the expected frequency under the null hypothesis and computed by:

 �=row total×column totalsample size

We will compare the value of the test statistic to the critical value of ��2 with degree of freedom = (*r* - 1) (*c* - 1), and reject the null hypothesis if �2>��2.

**Example S.4.1**

Is gender independent of education level? A random sample of 395 people were surveyed and each person was asked to report the highest education level they obtained. The data that resulted from the survey is summarized in the following table:

|  | **High School** | **Bachelors** | **Masters** | **Ph.d.** | **Total** |
| --- | --- | --- | --- | --- | --- |
| **Female** | 60 | 54 | 46 | 41 | 201 |
| **Male** | 40 | 44 | 53 | 57 | 194 |
| **Total** | 100 | 98 | 99 | 98 | 395 |

**Question**: Are gender and education level dependent at 5% level of significance? In other words, given the data collected above, is there a relationship between the gender of an individual and the level of education that they have obtained?

Exercise

My dog has so many fleas,

They do not come off with ease.  
As for shampoo, I have tried many types  
Even one called Bubble Hype,  
Which only killed 25% of the fleas,  
Unfortunately I was not pleased.

I've used all kinds of soap,  
Until I had given up hope  
Until one day I saw  
An ad that put me in awe.

A shampoo used for dogs  
called GOOD ENOUGH to Clean a Hog  
guaranteed to kill more fleas.

I gave Fido a bath  
and after doing the math  
His number of fleas  
started dropping by 3's!  
Before his shampoo  
I counted 42.

At the end of his bath,   
I redid the math  
and the new shampoo had killed 17 fleas.  
So now I was pleased.

Now it is time for you to have some fun  
With the level of significance being .01,  
You must help me figure out

Use the new shampoo or go without?

**Error in Hypothesis**

When like old saying “Nobody is above mistake and no one can 100% perfect” Therefore even though we have evidence or reason for any decision taken in statistics yet we are liable to make mistake and wrong prediction what an unpredicted world!

We cannot be always right hence it will be nice if we could know the kind of error we might make while making decision and who doesn’t want to know that??? Well this leads us to type of error!

While you are dealing with hypothesis there are two main type of error you are likely to commit they are:   
a. Type 1 error

b. Type 2 error

**What is Type-I error??**  
A type 1 error is committed when the truth is rejected i.e Here the null hypothesis is rejected when it is in fact true. It is committed when the rejection region is more than what is supposed to be. It is said to occurred when what is supposed to be in acceptance region is being place in rejection region. It is also positive statement fact scenario in real life.

The probability or chances of committing type I error is denoted by alpha (α)

**Note:** rejecting the null hypothesis means taking risk of **type I error**

**Example of Type1 Error in real life**

* Imagine I told you that keep visiting this website and follow me up is the only way you could pass your statistics. You failed to do so and many people do as I say then at end your exam you failed but everyone who did as I said passed the exam. You have committed Type1 because you rejected the truth.
* Sending an innocent defendant to jail is also type1 error because the judge must have rejected the truth.

**What is Type 2 error??**

Type 2 error is an error committed by not rejecting the false. It occurred when the null hypothesis is not rejected when it is actually false. It happens when what ought to have been in rejection region has been place to acceptance region. Hence it exist when acceptance region is larger than what it ought to be. It also negative false statement.

**Note:** failure to reject the null hypothesis means taking risk of type II error

**Example of Type2 Error in real Life**

A friend of mine told me that everyone would like to listen to voice explanation I agree with him and provide voice note Unfortunately no one ever listen to it after I did the review then here. Here I have committed type 2 error because I failed to disagree with my friend opinion which is in fact wrong.

Another example of type two error is freeing a guilty defendant to jail

The following table shows how each of the error is committed

|  |  |  |  |
| --- | --- | --- | --- |
| Decision | Null (H0) | Error | Probability |
| Reject Null | True | Type I | α (alpha) |
| Do not reject Null | False | Type II | β (beta) |
| Reject Null | False | No error | 1- β |
| Do not reject Null | True | No error | 1- α |

α= probability of a Type I error = probability of rejecting the null hypothesis when the null hypothesis is true.  
β = probability of a Type II error = probability of not rejecting the null hypothesis when the null hypothesis is false.

α and β should be as small as possible because they are probabilities of errors. They are rarely zero. The Power of the Test is 1- β. ideally, we want a high power that is as close to one as possible. Increasing the sample size can increase the Power of the Test.

Note: (1-β) is also refers to as Power of test. The kind of error statistician always avoid most is type1 error because it usually comes with a great risk. However, you might need to take risk of type1 just to avoid type2 in some scenario if type2 happened to comes with greater risk.

The following are examples of Type I and Type II errors and the one with greater risk

**Example1**

John was ask to take gun shot in his shirt with his bullet proof! What type of error committed by John if He thinks that

The bullet proof can’t save him from the bullet when it can actually save him?

The bullet proof will save him when in fact it can’t save him?

What type of error comes with greater risk

**Solution**

H0: the bullet can actually save him

Ha: the bullet cannot save him

Type1 error

Type 2 error

Type II error (He will go on and use the bullet proof and take gunshot which might make him loose his life)

**Example2**

You were told that sta114 is very easy to pass what type of error you committed if you believed

1. It is easy to pass when it is actually hard
2. It is hard when it is in fact hard
3. What type of error comes with greater risk

**Solution**

H0: It is easy to pass sta114

Ha: It is not easy to pass sta114

1. Type 2 error
2. No error (You make the right decision)
3. Type 2 ( you believed it is easy Hence you might not take it serious which might leads you to failure)

Example 3

Assuming the null hypothesis, is: The victim of a fire accident is alive when he arrives at the emergency  
room of a hospital. What case can the emergency crew commit type I and type II?? Hence which of the two has the greater consequence???

**Solution**

1. Type I error: If the emergency crew thinks that the victim is dead when, in fact, the victim is alive.
2. Type II error: The emergency crew thinks that the victim is alive when, in fact, the victim is dead.
3. Type I. (Here the emergency will not give any treatment again Hence the victim will later give up to ghost when He probably could have survived)

**Example4**

The mean price of Toyota cars in a Nigeria is 2millions. A test is conducted to see if the claim is true. State the Type I and Type II errors in complete sentences. Hence which one has the greater risk?

**Solution**

H0: The mean price of Toyota cars in a Nigeria is #2millions

Ha: The mean price of Toyota cars in a Nigeria is not 2millions

Type I: The mean price Toyota cars is #2millions, but we conclude that it is not #2millions.

Type II: The mean price of mid-sized cars is not #2millions, but we conclude that it is #2millions.

Type II

**Example5**

In a population of students in a certain school, approximately 64% are female. A test is conducted to see if, in fact, the proportion is less. State the null and alternative hypotheses. If it is concluded that the female proportion is not 64% when it is in fact 64% what type of error committed??

**Solution**

H0: the female proportion is 64%

Ha: the female proportion is less than 64%

Type I error

**Example6**

A student claims that He scored more than 40 in sta114 and complain to his lecturer if in fact what He scored was 40 what type of error has it committed??

**Solution**

Ho: He scored 40

Ha: He scored more than 40

Type I error (He rejected the truth i.e he scored 40)

**Example7**

The power of a test is 0.95. What is the probability of a Type II error?

**Solution**

Power of test = 1-Prob (type II error)

Therefore Prob(type II error) = 1-power of test

Prob (type II error) = 1-0.95

Prob (type II error) = 0.05

**Example8**

The school board claims that at least 60% of student brings a phone to school. A teacher believes this number is too high and randomly selected 25students to test at a level of significance of 0.02 if the out 15 of them brings phone to school. If the teacher remain on his stand at the end of his research what type of error has he committed if in fact the not less than 60% of the student brings phone to school??

**Solution**

Here the percentage given is our proportions i.e that 60% means 60/100 which is 0.6. the following are our hypothesis

Null : at least 60% of student brings a phone to school i.e starting from 60% upward which means it can be 60% or greater

Alternative: it is less than 60% (not at least 60% of student brings a phone to school)

H0: P ≤ 0.6  
Ha: P < 0.6

Type 1 error!

At this Juncture I believed you should have understand Type 1 and Type II error and how to know which one you should avoid in any case. However there are some other Errors which are not generally known They are Type III and Type IV error. I bet it that Your Lecturer might not even tell you this too. We are not interested in this but let us have a little discussion about Type III

Type III error

Type III error is the error committed when you give a correct answer to a wrong question! When you reject the null hypothesis when it is false but your thought or evidence is actually wrong. What a tricky error!

Example

A student score 45 in statistics exam but He argues with the lecturer that He scored more than that when the paper was remark it was discovered that He actually scored 40 what type of error committed by the student??

Answer: Type3 error.

In a population of students in a certain school, approximately 64% are female. A test is conducted to see if, in fact, the proportion is less. State the null and alternative hypotheses.  
Answer  
a. Suppose that a recent article stated that the mean time spent in jail by a first–time convicted rapist is 4 years. A  
study was then done to see if the mean time has increased in the new century. A random sample of 26 first-time  
convicted rapist in a recent year was picked. The standard deviation and the mean length of time in jail from the survey of 30 first-time rapist in this recent year was 3 years and 1.5years respectively. If you were conducting a hypothesis test to determine if the mean length of jail time has increased, what would the null and alternative;

**Exercise**

The mean price of mid-sized cars in a region is $32,000. A test is conducted to see if the claim is true. State the Type I and Type II errors in complete sentences.

A sleeping bag is tested to withstand temperatures of –25 °F. You think the bag cannot stand temperatures that low. State when you could commit Type I and Type II errors in complete sentences.

A group of doctors is deciding whether or not to perform an operation. Suppose the null hypothesis, H0, is: the surgical procedure will go well. State the Type I and Type II errors in complete sentences.

Which is the error with the greater consequence?

The power of a test is 0.981. What is the probability of a Type II error?

A group of divers is exploring an old sunken ship. Suppose the null hypothesis, H0, is: the sunken ship does not contain buried treasure. State the Type I and Type II errors in complete sentences.\

A microbiologist is testing a water sample for E-coli. Suppose the null hypothesis, H0, is: the sample does not contain E-coli. The probability that the sample does not contain E-coli, but the microbiologist thinks it does is 0.012. The probability that the sample does contain E-coli, but the microbiologist thinks it does not is 0.002. What is the power of this test?

**5.0 Rejection Rule Using Pvalue.**

You already learned earlier that you need to calculate for Zstatistics and compare it with the critical value gotten from your statistical table make know whether to reject or not. However there another way to decide. You can also make decision by comparing the Pvalue and the level of significant. You need to get your Pvalue by the use of your Zstatistcs or Binomial distribution.

It all depends on the question.

To make your decision on the null hypothesis is within the following three rules

Rule1: if the alternative hypothesis is written with greater than then reject the null hypothesis if and only if the Pvalue is less than the level of significant.

Rule2: if the alternative hypothesis is written with less than then reject the null hypothesis if and only if Pvalue is greater than the level of significant.

Rule3: if the alternative hypothesis is written with equal to then Reject the null hypothesis if and only if Pvalue is less than the level of significant.

If you observe the rule it is just the inverse of the Previous rule you learned before

ONE WAY ANOVA

In order to do Analysis of variance you have to take the following steps

**Step 1: write the null and alternative hypotheses**

**H**0: µ1 = µ2 = µ3 = µ4 = µ5 …= µk (The mean times served are equal.)

**H**a: Not all of the means are equal.

**Step 2: State the significance Level α = 0.01**

**Step 3: Construct the One-way ANOVA Table and fill the table till you get F-statistics**

**Step 4: Calculate the critical Value Fα (df1,df2)**

**Step 5: Make decision by comparing the Fstats and Critical value**

**Step 6: State the conclusion**

Example1

Is there any significant difference in the administrators’ performance of the academic administrators when they group according to socioeconomic status? Use alpha 0.05.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Socio-economics status** | | |
| **No** | **Low (x1)** | **Average (x2)** | **High (x3)** |  |
| 1 | 4 | 6 | 7 |  |
| 2 | 3 | 2 | 6 |  |
| 3 | 6 | 5 | 4 |  |
| 4 | 3 | 2 | 2 |  |
| 5 | 9 | 5 | 6 |  |
| 6 | 5 | 7 | 5 |  |
| 7 | 8 | 4 | 7 |  |

**Solution**

If you observe the table you will realized that it was categorized base one category (Department) It is only the column that was labeled so, this shows that you will be doing ONE WAY ANOVA

**Step 1: write the null and alternative hypotheses**

**H**0: µlow = µavg = µhigh

**H**a: Not all of the means are equal.

**Step 2:**

Level of significant α = 0.01

**Step 3: Construct the One-way ANOVA Table and fill the table till you get F-statistics**

**Step I:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Low (x1)** | **Average (x2)** | **High (x3)** |  | **Low** | **Average** | **High** |
|  | 4 | 6 | 7 |  | 2.04081633 | 2.469387755 | 2.938776 |
|  | 3 | 2 | 6 |  | 5.89795918 | 5.897959184 | 0.510204 |
|  | 6 | 5 | 4 |  | 0.32653061 | 0.326530612 | 1.653061 |
|  | 3 | 2 | 2 |  | 5.89795918 | 5.897959184 | 10.79592 |
|  | 9 | 5 | 6 |  | 12.755102 | 0.326530612 | 0.510204 |
|  | 5 | 7 | 5 |  | 0.18367347 | 6.612244898 | 0.081633 |
|  | 8 | 4 | 7 |  | 6.6122449 | 0.183673469 | 2.938776 |
| **Total** | **38** | **31** | **37** | SSW | 33.7142857 | 21.71428571 | 19.42857 |

5.4285714

4.4285714

5.2857143

SSW=33.7142857+21.71428571+19.42857

SSW= 74.85714286

**Step II: calculate mean and the sum square between the mean of all the three department.**

GrandMean and SSB the three classes

**SSB=4.0952**

**Step III**: calculate the degree of freedom between and within the group and construct the summary table.

N= 21, c=3

dfw = N-c = 21-3 = 18

dfb =N-1 = 3-1=2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variation | SS | Df | MS=SS/df | F=MSB/MSW | P-value | Fcrit |
| Between Groups | 4.095 | 2 |  |  |  |  |
| Within Groups | 74.8571 | 18 |  |  |  |  |
| Total | 75.4421 | 20 |  |  |  |  |

**Step IV:** use the necessary formula to calculate the remaining parameters and fill the remaining table

For degree of freedom column (DF)

MSB= SSB/df= 4.095/2= 2.0475 (Mean Square Between the group)

MSW=SSW/df= 74.8571/18 =4.1587 (Mean Square Within the group)

MST = 2.0475+4.1587= 6.2062 (Mean Square Total)

F=MSB/MSW=0/7 = 0

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variation | SS | Df | MS=SS/df | F=MSB/MSW | P-value | Fcrit |
| Between Groups | 4.095 | 2 | 2.0475 | 0.4924 | 0.05 | 3.55 |
| Within Groups | 74.8571 | 18 | 4.158730159 |  |  |  |
| Total | 75.4421 | 20 |  |  |  |  |

**Step 4: Calculate the critical Value F**α(df1,df2)

Note: our df1=2 and df2=18, from the F-table our Fcrit=3.55

**Step 5: Make decision by comparing the Fstats and Critical value**

**Decision:** we reject the null hypothesis if the Fstatistics >Fcrit

Hence we do not reject the null hypothesis since Fstats(2.0475) < Fcrit(3.55)

**Step 6: State the conclusion**

There is no enough evidence to say that there is any significant difference in the administrators’ performance of the academic administrators when they group according to socioeconomic status

Example2

The times required by three workers to perform an assembly-line task were recorded on five randomly selected occasions. Here are the times, to the nearest minute.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Hank | 9 | 10 | 6 | 11 | 10 | 9 |
| John | 8 | 9 | 9 | 8 | 10 | 7 |
| Sean | 9 | 9 | 8 | 11 | 9 | 9 |

Compute ONE-WAY ANOVA table for the given data and test wether the time spend by the three workers on average are the same.

**Solution**

**Step 1: write the null and alternative hypotheses**

**H**0: µlow = µavg = µhigh

**H**a: Not all of the means are equal.

**Step 2:**

Level of significant α = 0.01

**Step 3: Construct the One-way ANOVA Table and fill the table till you get F-statistics**

**Step I:** Calculate the mean and sum of square of each categories

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **StepI A** |  |  | **StepI C** |  |  |
|  | **Hank**  **(x1)** | **John**  **(x2)** | **Sean**  **(x3)** |  | **Hank** | **John** | **Sean** |
|  | 9 | 8 | 9 |  | 0.02777778 | 0.25 | 0.027778 |
|  | 10 | 9 | 9 |  | 0.69444444 | 0.25 | 0.027778 |
|  | 6 | 9 | 8 |  | 10.0277778 | 0.25 | 1.361111 |
|  | 11 | 8 | 11 |  | 3.36111111 | 0.25 | 3.361111 |
|  | 10 | 10 | 9 |  | 0.69444444 | 2.25 | 0.027778 |
|  | 9 | 7 | 9 |  | 0.02777778 | 2.25 | 0.027778 |
| **Total** | **55** | **51** | **55** | **SSW** | **14.8333333** | **5.5** | **4.833333** |

**StepI B**

9.1667

8.5

9.1667

**StepI D**

SSW=14.833+5.5+4.8333

SSW=25.1667

**Step II:** calculate mean and the sum square between the mean of all the three department.

Mean and SSB the three classes

n1=6 n2=6 and n3=6

**SSB=1.7778**

**Step III**: calculate the degree of freedom between and within the group

N= 18, c = 3

dfw = N-c = 18-3 = 15

dfb = c-1 = 3-1=2

**Step IV:** calculate the Mean Square Between the group and Mean Square Within the group, then F-statistics, Critical value and construct the ANOVA table.

MSB= SSB/df= 1.778/2= 2.0475 (Mean Square Between the group)

MSW=SSW/df= 74.8571/18 =4.1587 (Mean Square Within the group)

MST = 2.0475+4.1587= 6.2062 (Mean Square Total)

F=MSB/MSW=0/7 = 0

For critical value **F** α(df1,df2)

Note: our df1=2 and df2=18, from the F-table our Fcrit=3.55

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variation | SS | Df | MS=SS/df | F=MSB/MSW | P-value | F crit |
| Between Groups | 1.7778 | 2 | 0.88889 | 0.529801325 | 0.05 |  |
| Within Groups | 25.1667 | 15 | 1.6778 |  |  |  |
| Total | 26.9444 | 17 |  |  |  |  |

**Step 5: Make decision by comparing the Fstats and Critical value**

**Decision:** we reject the null hypothesis if the Fstatistics >Fcrit

Hence we do not reject the null hypothesis since Fstats (0.5298) < Fcrit (6.36)

**Step 6: State the conclusion**

At 1% level of significant here is no enough evidence to say that there is different between the times spent by the three workers on average.

**Example3**

A study of firefighters in a large urban area centered on the physical fitness of the engineers employed by the Fire Department. To measure the fitness, a physical therapist sampled five engineers each with 5, 10, 15, and 20 years’ experience with the department. She then recorded the number of pushups that each person could do in 60 seconds. The results are listed below. Perform an analysis of variance to determine if there are differences in the physical fitness of engineers by time with department group. Use α=0.01.

|  |  |  |  |
| --- | --- | --- | --- |
| **5-years** | **10-years** | **15-years** | **20-years** |
| 56 | 64 | 45 | 42 |
| 55 | 61 | 46 | 39 |
| 62 | 50 | 45 | 45 |
| 59 | 57 | 39 | 43 |
| 60 | 55 | 43 | 41 |

**Solution**

Step1: Writing the Hypothesis

**H**0: µ5yrs = µ10yrs = µ15yrs = µ20yrs

**H**a: Not all of the means are equal.

**Step 2: State the level of significant**

Level of significant α = 0.01

**Step 3: Construct the One-way ANOVA Table and fill the table till you get F-statistics**

**Step I:** Calculate the mean and sum of square of each categories

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **5-yrs**  **X1** | **10-yrs**  **X2** | **15-yrs**  **X3** | **20-yrs**  **X4** |  | **5-yrs** | **10-yrs** | **15-yrs** | **20-yrs** |
| 56 | 64 | 45 | 42 |  | 5.76 | 43.56 | 1.96 | 0 |
| 55 | 61 | 46 | 39 |  | 11.56 | 12.96 | 5.76 | 9 |
| 62 | 50 | 45 | 45 |  | 12.96 | 54.76 | 1.96 | 9 |
| 59 | 57 | 39 | 43 |  | 0.36 | 0.16 | 21.16 | 1 |
| 60 | 55 | 43 | 41 |  | 2.56 | 5.76 | 0.36 | 1 |
| **292** | **287** | **218** | **210** | **SSW** | **33.2** | **117.2** | **31.2** | **20** |

**StepI B**

58.4

57.4

43.6

42

**StepI D**

SSW=33.2+117.2+31.2+20

SSW= 201.6

**Step II:** calculate mean and the sum square between the mean of all the three department.

n1=5, n2=5, n2=5 and n3=5

**SSB=1148.95**

**Step III**: calculate the degree of freedom between and within the group and construct the ANOVA table.

N= 18, c = 4

dfw = N-c = 20-4 = 14

dfb = c-1 =4-1=3

**Step IV:** use the necessary formula to calculate the remaining parameters and fill the remaining table

MSB= SSB/df= 1148.95/3= 382.93 (Mean Square Between the group)

MSW=SSW/df= 201.6/16 =12.6 (Mean Square Within the group)

MST = 382.93+12.6= 6.2062 (Mean Square Total)

F=MSB/MSW=382.93/12.6 = 30.3955

F 0.1(3, 16) =5.29

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Source of variation** | **SS** | **Df** | **MS=SS/df** | **F=MSB/MSW** | **Pvalue** | **Fcrit** |
| **Between Groups** | 1148.95 | 3.00 | 382.93 | 30.39550265 | 0.01 | 5.29 |
| **Within Groups** | 201.6 | 16 | 12.6 |  |  |  |
| **Total** | 1350.55 | 19 | 395.583 |  |  |  |

**Decision:** We reject the null hypothesis since (30.3955)Fsta t>Fcrit(5.29)

**Conclusion:** At 1% level of significant we have enough evidence to conclude that there are differences in the physical fitness of engineers by time with department group.

Example 4

Complete the following one way Anova table by filling the unknown value.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variation | SS | df | MSS=SS/df | F=MSB/MSW | P-value | F crit |
| Between Groups | W?? | 3.00 | 278.2666667 | 9.850147493 | 0.05 |  |
| Within Groups | 452 | X?? | Y?? |  |  |  |
| Total | 1286.8 | Z?? |  |  |  |  |

Solution

We solve for “W” first.

Since we know that MSS=SS/df,

MSS =278.2666667 and df =3.00 then we can get SS

278.2666667 =SS/3.00 (cross multiply)

SS=278.2666667\*3

SS=

Hence W=

We solve for “Y” now.

Since we know that F=MSB/MSW

MSB =278.2666667and F=9.850147493 then we can get SS

9.850147493=278.2666667/MSW (cross multiply)

9.850147493MSW=278.2666667

MSW=278.2666667/9.850147493

MSW=28.25

Hence Y=28.25

We solve for “X”

Since we know that MSS=SS/df,

MSS =28.25 and SS=452

28.25 =452/df (cross multiply)

28.25df=452

Df=452/28.25 = 16

Hence X=16

Z= 16+3 =19.

The table will look like below

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source of variation | SS | Df | MS=SS/df | F=MSB/MSW | F crit |
| Between Groups | 834.8 | 3 | 278.2666667 | 9.850147493 |  |
| Within Groups | 452 | 16 | 28.25 |  |  |
| Total | 1286.8 | 19.00 | 306.5166667 |  |  |

**Example5**

Using the following summary data, perform a one-way analysis of variance using alpha=0.01

|  |  |  |
| --- | --- | --- |
| **N** | **Mean** | **Sd** |
| 21 | 40 | 10 |
| 21 | 45 | 12 |
| 21 | 50 | 11 |

**Solution**

In the question above we are given three means and it shows that it is just 3categories.

Now we need to get the Sum of Square within each categories.

**First category (Say category A)**

SS=2000

**Second category (Say category B)**

SS=2880

**Third category (Say category C)**

SS=2420

**So, Sum of Square within the group**

**SSW= 2000+2880+2420=7300**

**Now let’s calculate GrandMean between the Group and Sum of Square Between the group**

n1=21, n2=21 and n3=21

**SSB=1050**

**Step III**: calculate the degree of freedom between and within the group and construct the ANOVA table.

N= 63, c = 3

dfw = N-c = 63-3 = 60

dfb = C-1 = 3-1=2

**Step IV:** calculate the Mean Square Between the group and Mean Square Within the group, then F-statistics, Critical value and construct the ANOVA table.

MSB= SSB/df= 1050/2= 525 (Mean Square Between the group)

MSW=SSW/df= 7300/60 =121.6667 (Mean Square Within the group)

F=MSB/MSW=525/121.667 = 4.315

For critical value **F** α(df1,df2)

Note: our df1=2 and df2=18, from the F-table our Fcrit=3.55

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source of variation** | **SS** | **Df** | **MS=SS/df** | **F=MSB/MSW** | **Fcrit** |
| **Between Groups** | 1050 | 2 | 525 | 4.315 | 4.98 |
| **Within Groups** | 7300 | 60 | 121.667 |  |  |
| **Total** | 1350.55 | 62 |  |  |  |

**Decision:** We do not reject the null hypothesis since (4.315)Fstat < Fcrit(4.98)

**Conclusion:** At 1% level of significant we don’t have enough evidence to conclude that there are differences in the mean of the three groups.

**TWO Way anova**

Example

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| VARIETIES | CHEMISTRY | | | |
| I | I | III | IV |
| A | 8 | 5 | 5 | 7 |
| B | 7 | 6 | 4 | 4 |
| C | 3 | 6 | 5 | 4 |

SOUTION

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | I | I | III | IV | **Total (Ti)** |
| A | 8 | 5 | 5 | 7 | 25 |
| B | 7 | 6 | 4 | 4 | 21 |
| C | 3 | 6 | 5 | 4 | 18 |
| **Total (Tj)** | 18 | 17 | 14 | 15 | **64** |

**From the tab**l**e above**

**G=64 (the grand tota**l**)**

**N= 12 (number of overa**ll **data given)**

341.3333

**Sum of Square Between Rows**

5

6.1667

**Sum of Square Between Co**l**umn**

Where Tj and nj represent the total and number of data in each column respectively

**Sum of Square Total**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Col | Col | Col | Col | Total |
| Row | 64 | 25 | 25 | 49 | 163.00 |
| Row | 49 | 36 | 16 | 16 | 117.00 |
| Row | 9 | 36 | 25 | 16 | 86.00 |
| Total | 122.00 | 97.00 | 66.00 | 81.00 | **366.00** |

**For Error Within**

Since SStota=SScoumn+SSrow+errorwithin

So, ErrorWithin=SStota-SScoumns-SSrows

Error = 24.67-6.17-3.33=15.17

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sources Of Variation** | **SS** | **df** | **MSS= SS/df** | **F** |
| **Btw Rows** | 6.17 | 2 | 3.08 | 1.22 |
| **Btw Cols** | 3.33 | 3 | 1.11 | 0.44 |
| **Between Groups** | 15.17 | 6 | 2.53 |  |
| **Total** | 24.67 | 11 |  |  |

### Rows Variation

a=0.05  
F0.05(2,6)=5.1433  
**Decision**: We do not reject the null hypothesis since (1.22)�����<�����(5.1433)

**Conclusion:** at 0.05 level of significant The data do not have enough evidence to indicate that the given variable are differ

### Columns Variation

�=0.05  
�0.05(3,6)=4.7571  
**Decision**: We do not reject the null hypothesis since (0.44)�����<�����(4.7571)

**Conclusion:** at 0.05 level of significant the data do not have enough evidence to indicate that the given variable are differ

**Analysis of Variance**

Also referred to as ANOVA, is a method of testing whether or not the means of three or more populations are equal. The method is applicable if:

* All populations of interest are normally distributed.
* The populations have equal standard deviations.
* Samples (not necessarily of the same size) are randomly and independently selected from each population.

The test statistic for analysis of variance is the *F*-ratio.

**One-Way ANOVA**

A method of testing whether or not the means of three or more populations are equal; the method is applicable if:

* All populations of interest are normally distributed.
* The populations have equal standard deviations.
* Samples (not necessarily of the same size) are randomly and independently selected from each population.
* There is one independent variable and one dependent variable.

The test statistic for analysis of variance is the *F*-ratio.

**Variance**

Mean of the squared deviations from the mean; the square of the standard deviation. For a set of data, a deviation can be represented as *x* – x¯�¯ where *x* is a value of the data and x¯�¯ is the sample mean. The sample variance is equal to the sum of the squares of the deviations divided by the difference of the sample size and one.

**CHI SQURE**

**Example**

A man observe 100 people to see who deposits garbage in the can and who litters. You want to see if there is a difference base on gender.

A person can fall in one of four category

* Male, deposits garbage
* Male, litters
* Female deposits garbage
* Female litters.

The following shows the result observe.

|  |  |  |
| --- | --- | --- |
| **Gender** | Deposit | litter |
| Female | 13 | 37 |
| Male | 27 | 23 |

Given the data above, is there a significance in littering behavior between men and women?

**Solution**

|  |  |  |  |
| --- | --- | --- | --- |
| **Gender** | Deposit | Litter | Total |
| Female | 13 | 37 |  |
| Male | 27 | 23 |  |

Example 2

In an experiment you publish fliers in three different colors, and see how many people take them or not, the following shows your observation for 200 people

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Deposit | White | Pink | Green | Blue |
| Take | 13 | 40 | 29 |  |
| Do not take | 27 | 13 | 21 |  |
|  |  |  |  |  |

Is there a significant effect of color?

Example 3

In an Animalia campaign In Nigeria Quinine was administered to 1000 people out of a total population of 4000 the number of fever cases is shown below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Treatment | Fever | No Fever | Total |  |
| Quinine | 200 | 800 | 1000 |  |
| No Quinine | 400 | 200 |  |  |
| Total |  |  | 4000 |  |

Discuss the usefulness of Quinine in checking malaria.

In the 2000 Nigeria census the age of individual in soma town were found to be the fooling

|  |  |  |  |
| --- | --- | --- | --- |
| Ages | Less than 18 | 18-30 years | Above 30 |
| Percentage | 27 | 40 | 33 |

In 2010 ages of n=500 individual were sampled. Below are the results.

|  |  |  |  |
| --- | --- | --- | --- |
| Ages | Less than 18 | 18-30 years | Above 30 |
| No of individual | 27 | 40 | 33 |

2006 Census

|  |  |
| --- | --- |
| Male | Female |
| 7 134V488 | 6 9,08VI, 302 |
|  |  |

### Kebbi

(Number)

|  | 2006 |
| --- | --- |
| Fire-Wood | 412,891 |
| Electricity | 63,810 |
| Kerosene | 51,615 |
| Coal | 22,888 |
| Gas | 6,076 |
| /Animal dung/Sawdust/ Coconut Husk | 3,263 |
| Solar | 1,218 |
| Other | 1,066 |

### Fire-Wood, Cooking Fuel

(Number)

Example

Suppose a random sample of 216 patients that have a skin disease are classified into the four age categories. The frequencies are summarized in the following table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| AGE CATEGORY | | Less than 18 | 18-30 years | 31-V0  years | Above  years |
| SEVERITY | Moderate | 15 | 32 | 18 | 5 |
| mildly severe | 18 | 8 | 29 | 23 |
| severe | 25 | 22 | 1 | 20 |

Conduct a test to determine if the severity of the disease is independent of the age of the patient.

Exampe

A car manufacturer expects the order coors of their mode to be distributed as foows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coor | White | Back | Siver | Other |
| Customer(%) | 18% | 24% | 34% | 24% |

A random sampe of 200 orders reveaed the fowing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coor | White | Back | Siver | Other |
| No of Customers | 40 | 70 | 40 | 50 |

Is there a statistical significant difference between the manufacturers expected proportion and the observed? are the customers’ orders different between the color categories?

The studies shows that the higher number of fight tickets bought by male in comparison to female is 2:1. Out of 200 tickets 140tickets were bought by male find out if the experiment manipulation causes changes the result or we have a chance variation.

|  |  |  |  |
| --- | --- | --- | --- |
| Troublesome | Being Drinking | | |
| Never | Occasionally | Frequently |
| Trouble with Police | 71 | 192 | 398 |
| No trouble with poice | 4929 | 2802 | 2602 |
| Tota | 000 | 3000 | 3000 |

1. A study of 11,000 acoho drinkers on university campuses reveaed;

Ho: binge drinking and troube with poice are independent

Ha: they are dependent.

### Observed values (O)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Never | Occasionally | Frequently | Total |
| Trouble with Police | 71 | 192 | 398 | 661.00 |
| No trouble with poice | 4929 | 2808 | 2602 | 10339.00 |
| Tota | 5000.00 | 3000.00 | 3000.00 | 11000.00 |

### Rows and Columns Total

### R1= 71+192+398 = 661.0000

### R2= 4929+2808+2602= 10339.0000

### C1= 71+4929 = 5000.0000

### C2= 192+2808 = 3000.0000

### C3= 398+2602 = 3000.0000

### GrandTota(G) = 661.0000+10339.0000 =11000.00

### Expected values (E)

### RiCi=Ri\*Ci/G

### R1,C1=66\*1.00500\*0.001100/11, 000 = 30\*0.45

### R1,R2=R1\*C2/G

### R1,2=66\*1.00300\*0.001100/11,000 = 18\*0.27

### R1,C3=R1\*C3/G

### R1,3=66\*1.00300\*0.001100/11,000 = 18\*0.27

### R2,R1=R2RR1R

### R2,C1=1033\*9.00500\*0.00110/11,000 = 469\*9.55

### R2,C2=R2\*C2/G

### R2,C2=1033\*9.00300\*0.00110/11,000 = 281\*9.73

### R2,R3=R2\*C3/G

### R2,3=1033\*9.00300\*0.001100/11,000 = 281\*9.73

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **C 1** | **C 2** | **C 3** | **Total** |
| **R1** | 300.45 | 180.27 | 180.27 | 660.99 |
| **R2** | 4699.55 | 2819.73 | 2819.73 | 10339.01 |
| **Total** | 5000.00 | 3000.00 | 3000.00 | **11000.00** |

### Rows and Columns Total

R 1�����= 300.45+180.27+180.27

R 1�����= 660.9900

R 2�����= 4699.55+2819.73+2819.73

R 2�����= 10339.0100

C 1�����= 300.45+4699.55

C 1�����= 5000.0000

C 2�����= 180.27+2819.73

C 2�����= 3000.0000

C 3�����= 180.27+2819.73

C 3�����= 3000.0000

����������=660.9900+10339.0100

����������=11000.00

### Deviation (O-E)

#### ���=���−���

#### �1,1=�1,1−�1,1

�1,1=71−300.45  
�1,1=-229.45

#### �1,2=�1,2−�1,2

�1,2=192−180.27  
�1,2=11.73

#### �1,3=�1,3−�1,3

�1,3=398−180.27  
�1,3=217.73

#### �2,1=�2,1−�2,1

�2,1=4929−4699.55  
�2,1=229.45

#### �2,2=�2,2−�2,2

�2,2=2808−2819.73  
�2,2=-11.73

#### �2,3=�2,3−�2,3

�2,3=2602−2819.73  
�2,3=-217.73

**H0: Variable1 (column) and variable2 (Rows) are not related in the population;  
  
Ha: Variable1 (column) and variable2 (Rows) are related in the population**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Rows** | **Colums** | **O** | **E** | **D** | **D2** | **D2/E** |
| **Row 1** | **Col 1** | 71 | 300.45 | -229.45 | 52647.30 | 175.23 |
| **Row 1** | **Col 2** | 192 | 180.27 | 11.73 | 137.59 | 0.76 |
| **Row 1** | **Col 3** | 398 | 180.27 | 217.73 | 47406.35 | 262.97 |
| **Row 2** | **Col 1** | 4929 | 4699.55 | 229.45 | 52647.30 | 11.20 |
| **Row 2** | **Col 2** | 2808 | 2819.73 | -11.73 | 137.59 | 0.05 |
| **Row 2** | **Col 3** | 2602 | 2819.73 | -217.73 | 47406.35 | 16.81 |
| **Total** |  | **11000.00** | **11000.00** | **0.00** | **200382.48** | **467.02** |

### �2=∑(�−�)2� �2=467.0200

��=(�−1)(�−1)  
  
�=2,�=3  
  
��=(2−1)(3−1)  
  
��=1∗2=2  
�=0.05

�����=5.99

**Decision:** Reject the null hypothesis since (467.02)�����>�����(5.99)

**Conclusion:**at 0.05 level of significant We have enough evidence to conclude that the variables are related