**Hypothesis Testing**

Table of Contents

[Null Hypothesis Significance Testing 4](#_Toc64281698)

[Significance Levels and Power 6](#_Toc64281699)

[Mean of a Normal Population with Known 9](#_Toc64281700)

[Significance Levels 9](#_Toc64281701)

[Defining the Rejection Region 10](#_Toc64281702)

[Making a Decision 11](#_Toc64281703)

[-Value 11](#_Toc64281704)

[Probability of Type II Errors 12](#_Toc64281705)

[Power 12](#_Toc64281706)

[Mean of a Normal Population with Unknown 15](#_Toc64281707)

[Making a Decision 17](#_Toc64281708)

[Equality of Means of Two Normal Populations 18](#_Toc64281709)

[Known Variances 19](#_Toc64281710)

[Unknown Variances 21](#_Toc64281711)

[Paired Test 24](#_Toc64281712)

[Variance of a Normal Population 25](#_Toc64281713)

[Equality of Two Variances 27](#_Toc64281714)

A hypothesis is a statement about a parameter or a set of parameters. The statement may or may not be true, we do not know. Thus, we need to test the hypothesis. For some given data, we check if the data is consistent with the hypothesis or not. If it is consistent, then it is possible that the hypothesis is true, so we keep it. If it is not consistent, then we reject the hypothesis.

We cannot just directly use data we collect with a hypothesis. We have to create a function of the data, i.e. a statistic. We can then check if the probability model for the statistic is consistent with the data.

## Null Hypothesis Significance Testing

We want to know whether a coin is fair or not. Say we toss the coin a hundred times and we get 85 heads. The question is, is the coin unfair? What if we get 65 heads? How about 52? With 85 heads, we can be pretty sure that the coin in unfair. With 65 heads, it is unclear. With 52 heads there is actually no evidence at all.

We cannot just claim this stuff though. We need mathematical proof of it. Afterall, there is always a small possibility that even with a fair coin we really do get 85 heads. This is where Null Hypothesis Significance Testing comes in.

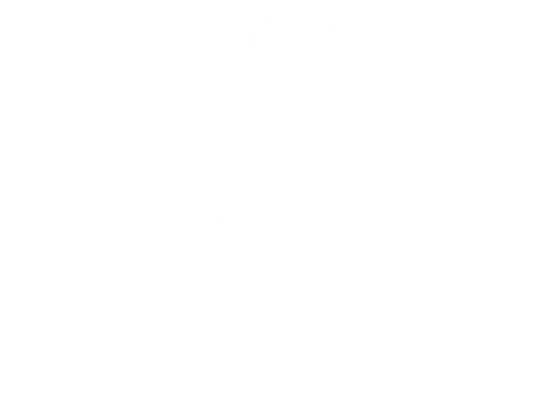
The Null Hypothesis, , is the default assumption about the model, the model generally being the population. Say we are considering the delays of data packets and we assume that maybe the average delay is . This would be . We can reject this hypothesis or we can keep it around (but we still are not accepting it to be 100% true).

If we reject the null hypothesis, it means we have found something that is better, and we are rejecting the null hypothesis in favour of it. This new hypothesis would be the alternative hypothesis, . is the complement to . If we reject , we accept .

Alongside these, we also need a test statistic, . We compute this statistic based on the random sample. Using this, we can define the null distribution, or . Similarly, we can also have an alternative distribution.

Thus, we need to first define a null hypothesis and perhaps an alternative hypothesis. Once we have this, we need to define a test statistic with which we can test the hypothesis. To perform the test, we need the null distribution. Depending on whether the statistic is discrete or continuous, the null distribution may be a PMF or a PDF.

Finally, we need to define a rejection region and a non-rejection region. For example, say we have a normal distribution. We can create a graph for it.



We can define some region on the graph towards the far ends as the rejection region. If the test statistic has a value in that region, we will reject the hypothesis. If is not in the rejection region, it is in the non-rejection region and we do not reject the hypothesis.

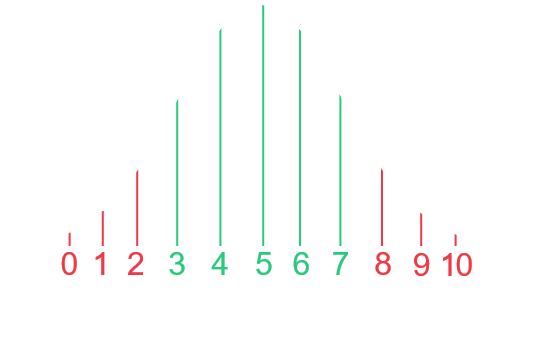
Example

Say we want to test whether or not a coin is fair. Thus, is and is . To test the hypothesis, say we toss the coin times. We can define a random variable which counts the number of heads we get. Thus, the null distribution can be defined as .

For different values of , say we get the following table of probabilities:`

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

We need to define a rejection region and a non-rejection region. We can use a graph for this to help visualize it:



Thus, we decided that the values from to define the non-rejection region. We will get back to how we decided on this region.

### Significance Levels and Power

The question now is, since we will reject the hypothesis if we get values in the rejection regions, meaning we will say that the coin is unfair, what is the probability that we are wrong?

This is just the sum of the probabilities of the values in the rejection regions. We cannot change the values of the probabilities. Those come from the hypothesis itself. All we can do is change the size of the rejection region. The larger the rejection region, the more likely we are to make a mistake. However, the smaller the rejection region, the more likely we are to make the opposite mistake, failing to reject a hypothesis that was not true.

So, how do we decide on the sizes of the regions? The sizes are decided by the person testing. It depends on the sensitivity of the test. If the consequences of rejecting a correct hypothesis are severe, then it is better to allow a few false positives and make the rejection region smaller. The test is said to have a low significance level. On the other hand, if it is more important to know if the coin is fair than to make a correct hypothesis in the first place, we can increase the rejection region. In this case, we will make mistakes, even deliberately. The test is said to have a higher significance level.

Rejecting a correct hypothesis is called a Type I error. Type I errors are false negatives. Failing to reject an incorrect hypothesis is called a Type II error. Type II errors are false positives.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | True State | |
|  |  |  |  |
| Decision | Reject | Type I  Error | Correct  Decision |
| Do not  reject | Correct  Decision | Type II  Error |

The probability of making a Type I error is called the Significance Level () of the test, .

The probability of correctly rejecting when it is not true is called the Power of the test. This is the same as the probability of not making a Type II error, i.e. .

Example

Consider this table showing the probabilities of getting different numbers of heads for different and :

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

In the first case only, is true. We can find , which is the sum of the probabilities of the values of that are in the rejection region.

In the second and third cases, is false. The Power of the test can again be found by summing up the probabilities of the values in the rejection region, since that will tell us the probability of correctly rejecting this value. In the second case, the power will be and in the third case the power will be . The more wrong the hypothesis is, the more likely we are to reject it.

## Mean of a Normal Population with Known

Null Hypothesis, :

Alternative Hypothesis, :

Test Statistic: , where

Null Distribution:

Standardized Null Distribution:

Sample Data:

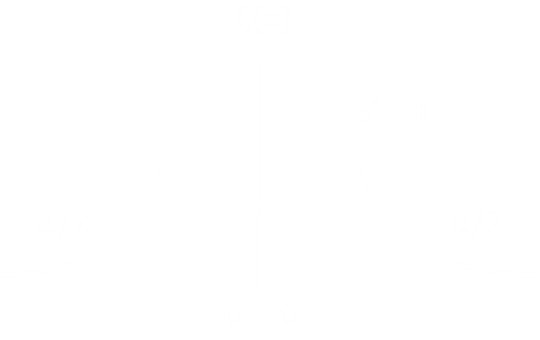
Since this dealt with normal distributions, we were able to standardize the results. Our final decision will be based on the significance level.

### Significance Levels

The significance level could be given as or or , and we could be asked whether or not the hypothesis is acceptable, based on this significance level. Alternatively, the significance level may not be given, in which case we must either assume a value or we could use the p-value, a concept we will be discussing shortly.

### Defining the Rejection Region

Say we are given a significance level of . Now take a look at the graph for the standardized null distribution.



The area on both sides combine to form the total rejection level, so the value of the total area from those two parts should be .

We also need to consider whether the test is two sided or one sided. If the test is two sided, it means that if the hypothesis is not true, the value of could be less than or greater than . In some cases, it may not be possible for to have a negative value, such as if we are dealing with peoples’ weights. In those cases, we have a one-sided test, so the rejection region can only be on one side, so the area on that side (either on the positive side or the negative side) will be double that in the case for two-sided tests. For now, we are assuming this is a two-sided test.

The rejection region also defines , the probability of rejecting a Null Hypothesis that is actually true.

### Making a Decision

Now, we need to know the threshold at which we enter the rejection region, also called the critical values. We need to know the values of for which and . Using the table for standard normal random variables, we will find that for , .

For the given null hypothesis, if we get a value of that is greater than or a value of that is less than , we reject the null hypothesis.

### -Value

The critical values we found depend on the significance level, . Instead of doing this, we could define the test decisions in terms of a value called the -value. The -value defines the maximum possible significance level, above which the null hypothesis will be rejected.

For a two-sided test, -value .

For a one-sided test, -value if the values are positive and -value if the values are negative.

### Probability of Type II Errors

Type II errors are when the alternative hypothesis is true, but we fail to reject the null hypothesis. We will not be getting into the derivation of the equation for the probability of Type II errors.

Here, and .

### Power

The power of the test tells us the probability of the correct decision that the alternative hypothesis is true.

Example

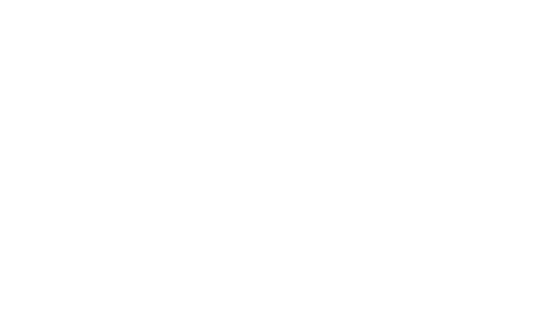
Say data is being transmitted from one station to another, and the mean signal is . The distribution of errors in the signal is , where . The received signal is given by .

Say the null hypothesis is and the alternative hypothesis is . For signal levels, the mean signal level we get is .

The question is, should we reject the hypothesis for a significance level of ? How about for ?

If we standardize the test statistic,

For , from the table for standard normal variables, we will find that occurs at . Thus, we will not reject the null hypothesis.



For , occurs at . Thus, in this case, we will reject the null hypothesis.

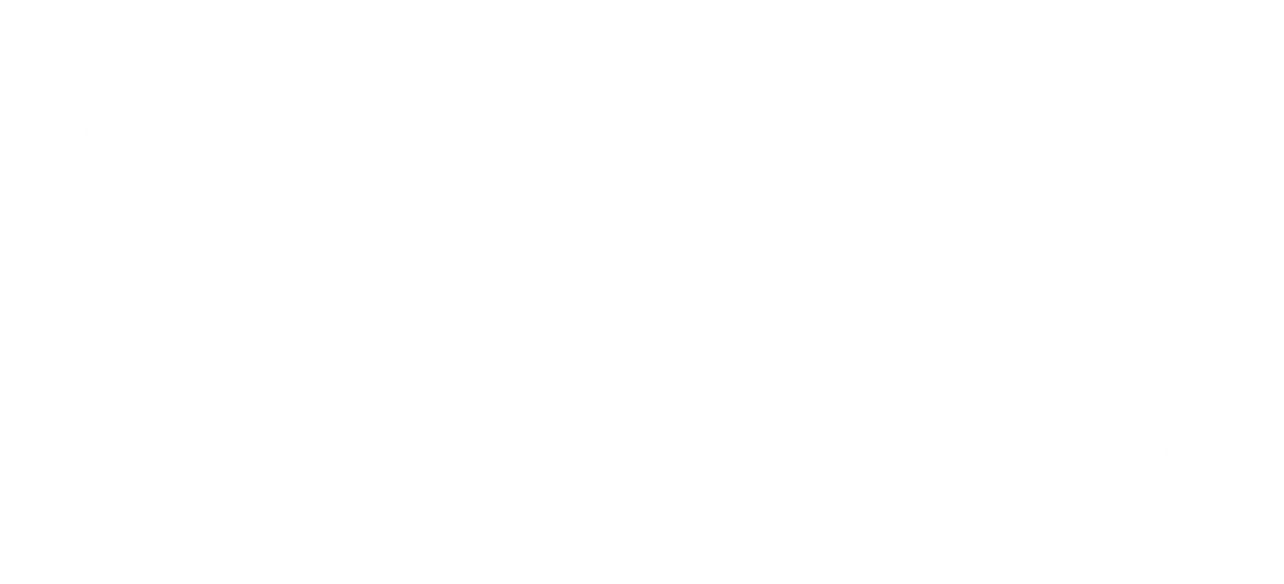
On the other hand, if we use -values,

-value

Thus, for all , we will reject the null hypothesis. For , we do not reject the null hypothesis but for , we will. If we are now given a new value for , we can easily figure out if we should reject the given null hypothesis or not at just a glance.

If we want to calculate the probability of Type II errors, we need to have a specific value for . Knowing that is not enough. Let us assume that .

Using the value of , we can create a curve called the Operating Characteristics (OC) curve, and it defines how clearly we can accept an alternative hypothesis.



The value of depends on the difference between and . If the two hypotheses values are close, the curves of the two hypotheses will overlap. If we reject something as the null hypothesis, we will also end up reject it as the alternative hypothesis. Thus, the value of will be high and the power will be lower.

## Mean of a Normal Population with Unknown

When we know the variance, we are working with normal distribution. Because of this, the hypothesis test is sometimes called a -test. If we do not know the variance, we use the student distribution.

The standardized test statistic is found as

If , we reject the null hypothesis. Otherwise, we do not reject it.

Additionally, the -value is given by

-value

Say we have a random sample where the common distribution is

where is not known.

In this situation, say

The sample mean is given by

Since we do not know the population variance, we must calculate the sample variance from the data as .

The standardized sample mean is thus

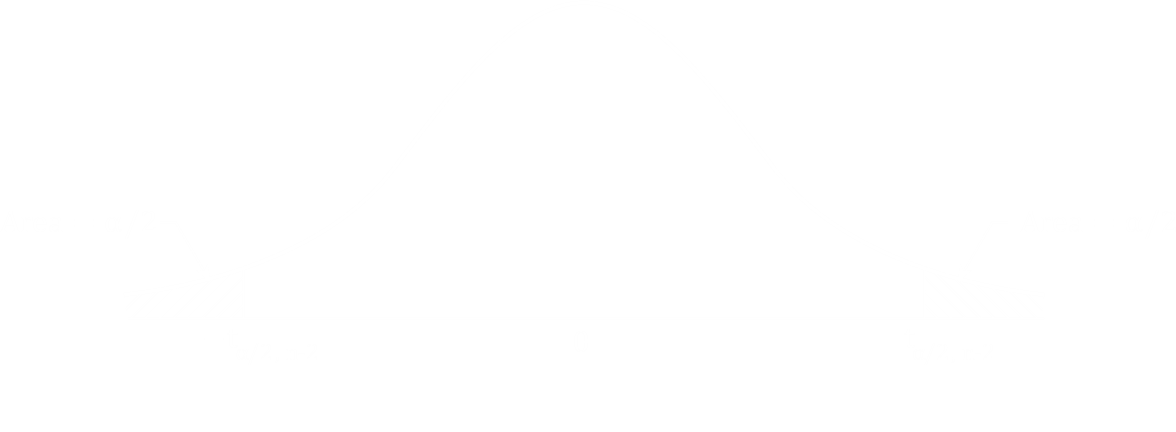
Notice that the standardized sample mean in the case where we had to calculate the variance from the data is a distribution with degrees of freedom.

Finally, the test statistic can be calculated as

where

### Making a Decision

The distribution has a similar curve to the standard normal distribution, except that it has a longer tail. We have to do the same thing, finding the critical points on the curve.



We will reject if . Of course, this is a generalization and we can subdivide this into a few parts if we dig deeper.

If we have a two-sided test, if , then . In this case, for a significance level of , the critical points can be on either side of . That means we must divided into two parts. This is why we find that if , is rejected. In this case, the -value is given by .

If we have a one-sided test where is that is greater than the value suggested by , i.e. , is not divided into two parts, and is reject only if . The -value is given by .

If we a one-sided test where is that is less than the value suggested by , i.e. , is not divided into two parts, and is rejected only if . The -value is given by .

## Equality of Means of Two Normal Populations

Random Samples:

It is assumed that the two random samples are independent.

Common Distributions: .

Sample Means:

We do not know if and will be given. If they are not, we will need to calculate the sample variances.

Sample Variances:

Since we want to test whether the means of the two populations are equal, our hypotheses will be

or

or

The test statistic is given by

Since both the random samples had normal distributions, the test statistic, which is the difference between the sample means, is also normally distributed. The first parameter is the difference between the mean values. The second value is called the pooled variance, which is an assumed value that both the populations had the same variance.

The standardized test statistic is given by

Since our hypothesis is that , we can write the above equation as

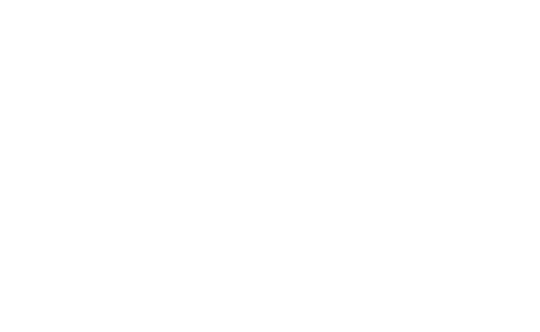
### Known Variances

First, let us consider the case where and are known.

In this case, the pooled variance is given by

And the test statistic is normally distributed.

Accordingly, we can find the critical values from the graph.



We will reject if . Again, this is a generalization under the assumption that the test is two-sided. We can find more subdivisions if we dig deeper.

The -value for this case is given by .

All of this information can be summarized in a table:

|  |  |  |  |
| --- | --- | --- | --- |
| and | Known | | |
| Test Statistics |  | | |
| Null Distribution | where | | |
| Pooled Variance |  | | |
| Alternative Hypothesis |  |  |  |
| Rejection |  |  |  |
| -Value |  |  |  |

### Unknown Variances

When we do not know the values of the variances, we can have one of three cases.

The first case is that the variances are unknown, but are equal, i.e. . In this case, the main points to notice are that the test statistic is a distribution, and the degree of freedom is .

|  |  |  |  |
| --- | --- | --- | --- |
| and | Unknown and Equal | | |
| Test Statistics |  | | |
| Null Distribution | where | | |
| Pooled Variance |  | | |
| Alternative Hypothesis |  |  |  |
| Rejection |  |  |  |
| -Value |  |  |  |

The second case is when the variances are unknown and unequal, but the values of and are large. The fact that they are large means that they are approximately equal to their respective population variances. This is why the test statistic is normally distributed.

|  |  |  |  |
| --- | --- | --- | --- |
| and | Unknown and Unequal with Large and | | |
| Test Statistics |  | | |
| Null Distribution | where | | |
| Pooled Variance |  | | |
| Alternative Hypothesis |  |  |  |
| Rejection |  |  |  |
| -Value |  |  |  |

The last case is when the variances are unknown and unequal, and the values of and are also small. In this case, the test statistic has a distribution. The degree of freedom in this case is

Due to how large this equation is, the degree of freedom is simply denoted as in the table.

|  |  |  |  |
| --- | --- | --- | --- |
| and | Unknown and Unequal with Small and | | |
| Test Statistics |  | | |
| Null Distribution | where | | |
| Pooled Variance |  | | |
| Alternative Hypothesis |  |  |  |
| Rejection |  |  |  |
| -Value |  |  |  |

## Paired Test

In the previous subtopic, we saw hypothesis testing for independent random samples. Now we will see hypothesis testing for dependant random samples.

Say we want to see whether or not the mileage of a set of a type of car decreases after attaching an anti-pollution device. We take random sample before attaching the device and we take a random sample after attaching the devices. In this case, the value of will be directly dependant on the value of .

The first thing we will do is define another set of random variables, , where . For this new set of random variables, our null hypothesis will be

meaning .

The test statistic is given by

where and is the sample variance given by

Lastly, taking a decision is simply checking if . If it is, we reject . Again, this is for the scenario where the test is two-sided and the condition will change accordingly for one-sided tests.

The -value is .

## Variance of a Normal Population

Say we have a random sample .

The null hypothesis is , which means the alternative hypothesis is .

The test statistic is the sample variance.

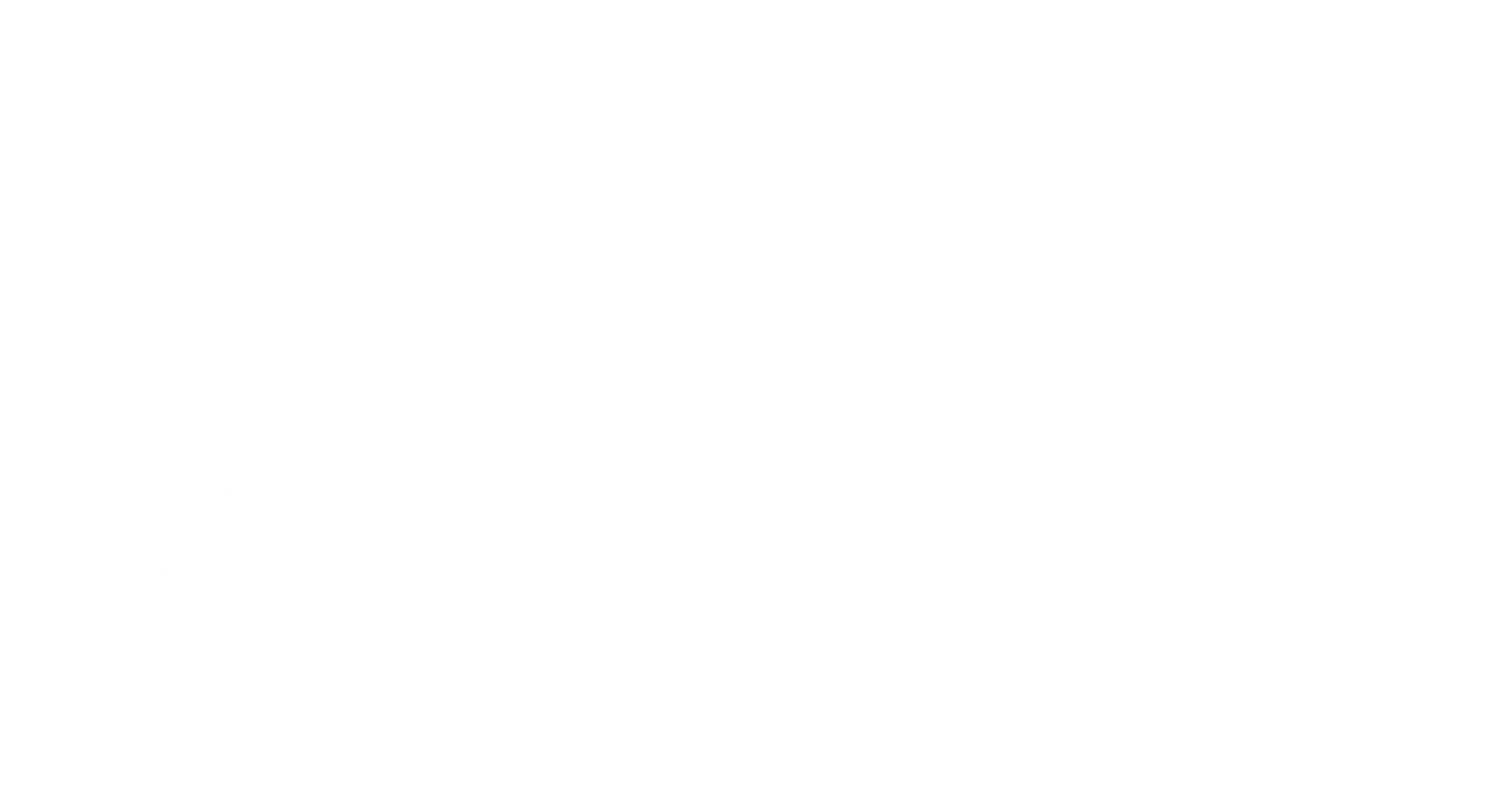
Now, we need to know the distribution of the test statistic. In a previous lecture, we have seen that

Thus, we will convert out test statistic into

Thus, the null distribution is given by

where .

The problem with the chi-square distribution is that it is not symmetric, so we cannot find the critical values as easily. The method basically remains the same though.



From the graph, we find a point such that

Thus, we get a point where the area on the right of that point is .

On the other end, we try to find the point , the point that has an area of on its left.

Finally, we make a decision. is rejected if or .

The -value needs two values in this case, since the graph is not symmetric.

-value

The reason we took the minimum is to be safe, since the -value gives us the extreme value of beyond which will be rejected.

If the test is one-sided, we need to adjust accordingly.

## Equality of Two Variances

Say we have one random sample, with the population mean and population variance and another random sample, with the population mean and the population variance .

Our null hypothesis is .

The sample variances of the two random samples are and respectively.

We know that

In the first lecture of statistics, we saw a rule.

This ratio gives us

According to our hypothesis, . Thus, those two terms get cancelled out. Similarly, and also get cancelled out (I’m assuming they are meant to be equal.) This leaves us with

will be rejected if

The -value is given by

-value

where