# Hypothesis Testing

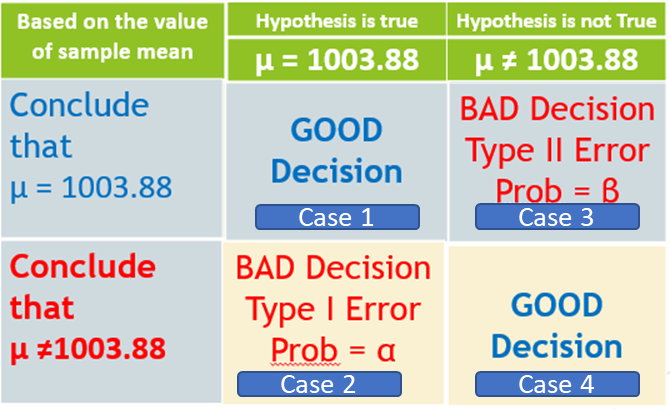
## A revisit to How Much to Pack?

We concluded the following after the case study:

* Akash set the filling machine to fill at an average of 1003.88 gms
* Does it mean that μ = 1003.88?
* We know σ = 5/3 gms
* Take a sample of 100 packets, each of 1000 gms.
* Calculate
* Can we conclude that μ = 1003.88gm and the machine is based on the value of ?

The last question is very important. Suppose Akash set the machine at 1003.88gms. Now also there are below 4 possibilities that can happen due to machine’s internal issues:

1. Akash took out a sample of 100 packets and got a mean weight of 1003.88gm and machine is also producing packets at 1003.88gm mean.
2. Akash took out a sample of 100 packets but didn’t get a mean weight of 1003.88gm while machine is producing packets at 1003.88gm mean. He got a bad sample only.
3. Akash took out a sample of 100 packets and got a mean weight of 1003.88gm but machine is not producing packets at 1003.88gm mean.
4. Akash took out a sample of 100 packets but didn’t get a mean weight of 1003.88gm and machine is also not producing packets at 1003.88gm mean.



## What can be the implications of Akash’s decisions

In each case Akash has choices either to stop the production line all together or continue with the production. Keep in mind:

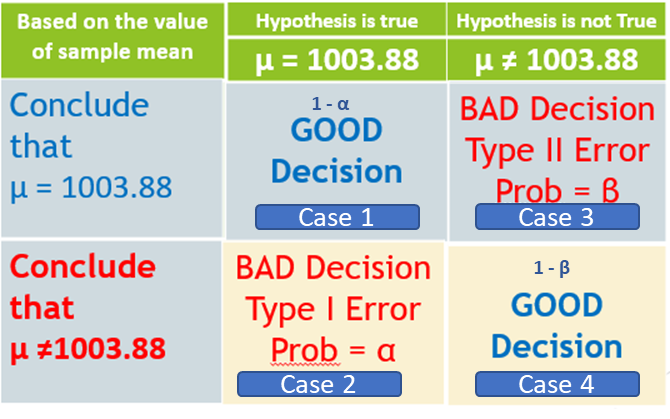
1. Stopping production of a single machine will create a bottleneck of the whole production line due to false panic.
2. Letting production continue casually might risk being caught be FDA as machine might be producing packets below the guidelines.

Following can be the decisions as per the case numbers:

1. Akash let the machine continue.
2. Akash got a panic and stops the machine. But machine was actually ok. So it became a false panic and Akash committed a **type I error.** i.e. machine was ok while a false panic was set in. This will be a bad decision.
3. Akash continue to run the machine but it was not producing output at mean = 1003.88. So Akash committed a **type II error.** i.e. machine was not ok but Akash casually ignored. This might result in FDA guideline violation. This will again be a bad decision.
4. Akash stops the machine. Since he got a sample with a mean not equal to the set value and in actual the machine was producing packets with a different mean. This will be a good decision.

## Nomenclature

1. Type 1 error: Machine was ok decision was taken predicting it to be not ok. The probability of this error is denoted by **α**. So α is the probability that we let the machine stop given that machine was doing all right. **It is the error of rejecting the null hypothesis.**
2. Type 2 error: Machine was not ok decision was taken predicting it to be ok. The probability of this error is denoted by **β**. So **β** is the probability that you let the machine continue given that the machine is not doing all right. **It is the error of not rejecting the null hypothesis**
3. So **α** and **β** are **conditional probabilities** where  are the conditions. So if we run by columns the conditional probabilities of taking the good decisions are as follows:



1. **There is no reason in the world for α and β should sum up to 1.**

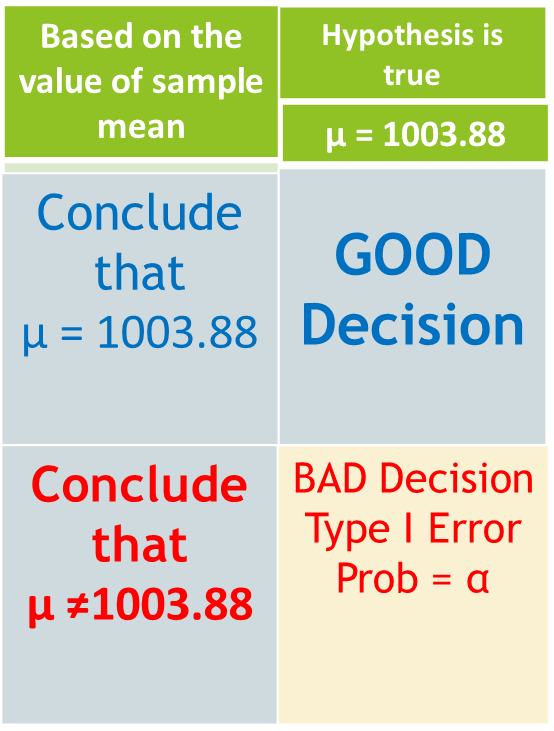
# Steps of Doing Hypothesis testing

## Implication of α

### What is the idea?

Before diving into the steps we need to understand what variable are we going to estimate and why.

1. We start with the first column i.e. Hypothesis is true.

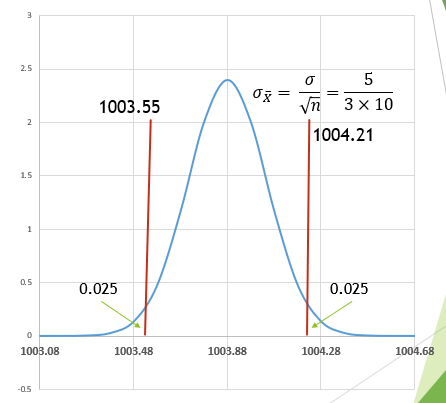


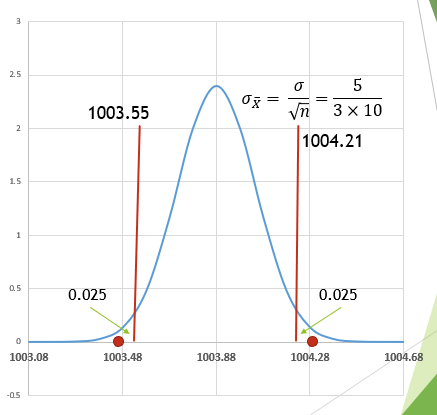
1. When we are checking the first column, there is no possibility of committing type II error.
2. So a sample came out, we calculate the sample mean , and we conclude that the machine is doing all right and let it continue which is a good decision.
3. A sample came out, we calculate the sample mean , and we conclude that the machine is not doing all right and stop the machine while in reality machine is actually doing all right. This will be a bad decision with a probability α.
4. So in other words we need to focus on the range of values of after which we will say to stop the machine. This will be the estimation in this case.

### Steps

Below are the steps:

1. Assume the machine is doing all right (Since we are testing in first column this should be our assumption) and producing packets with the set mean μ = 1003.88 gm and σ = 5/3 gm.
2. Define α as the probability of our assumption from the taken-out sample is wrong to a set value say 0.05. In other words, I am giving a tolerance limit of our estimation going wrong and we wrongly stop the machine. How the α will be set? Depends on use case as below:
   1. Suppose we form a hypothesis that a person is innocent until proven guilty beyond reasonable doubt. The law of the land.
   2. Suppose a biker got in a quarrel with the traffic guard over an unjustified challan issue and the traffic guard seized his bike keys and took him to his boss.
      1. The boss’ work is only hearing the case and say guilty and not guilty and decide the fine.
      2. The boss hears the case from the traffic guard in 1-2 minutes and without hearing the bikers argument he pass a judgement that biker was guilty and fined him Rs. 500
      3. In this case he had a high value of α i.e. the error that he tolerates in making his assumption.
   3. Suppose the next day same traffic guard went to a commotion nearby and sees a man dead in a pool of blood with multiple stab wounds and another man is holding a knife. He took the man holding the knife to the same boss of him. Do you think the boss’ behavior will be the same as in b? So here he will have to make sure he will not establish an innocent man guilty without sufficient proof. In other words “we will tolerate even 100 murderers in the street rather than an innocent being punished”. In another word the traffic police will try to reduce α in his assumption.
3. We create a normal distribution of mean = μ and standard deviation = create the confidence interval of 95% as below:



1. Thus there is a 95% probability that a sample we pickup will have its sample mean within the above red lines.
2. Suppose we get a sample with mean as in the below places: 

Then statistically my decision will be to unnecessarily stop the machine which was doing all right. Thus the probability of this situation arising is 5%. On calculating the above intervals it will come as:

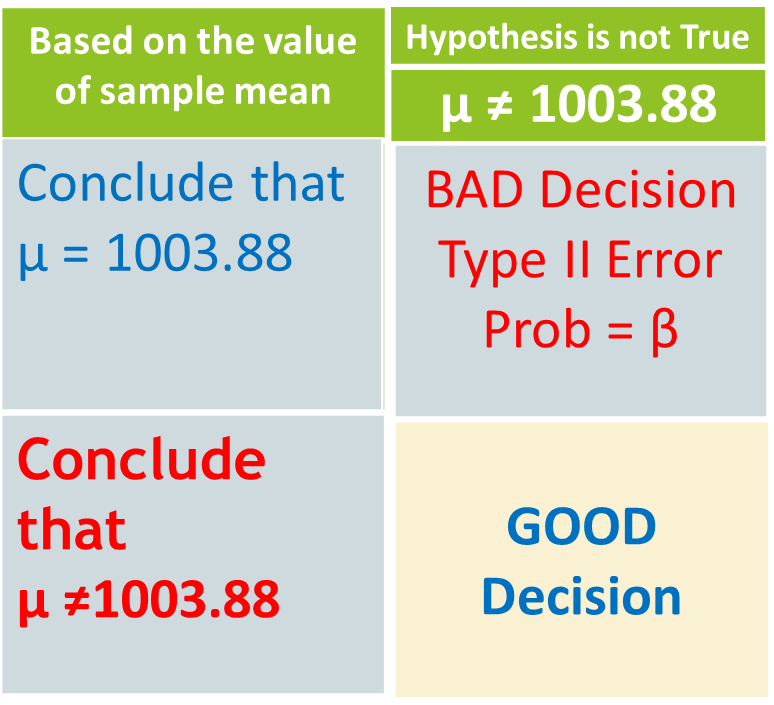
* 1. LL = 1003.88 – 1.96\*5/30= 1003.55
  2. UL = 1003.88 + 1.96\*5/30=1004.21

1. So we form a decision rule that we will let the machine continue if our sample means are between 1003.55 mg and 1004.21 mg.

## Implication of β

### What is the idea?

1. For β understand that we are in the 2nd column as below:

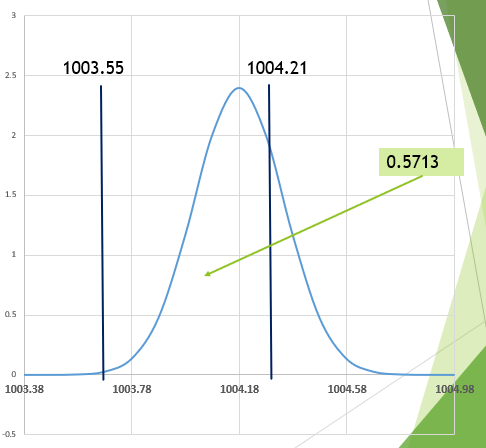


i.e. μ ≠ 1003.88. So how many possible values where μ ≠ 1003.88? Its infinite. This is the first problem of β, there is no unique value we can put as μ.

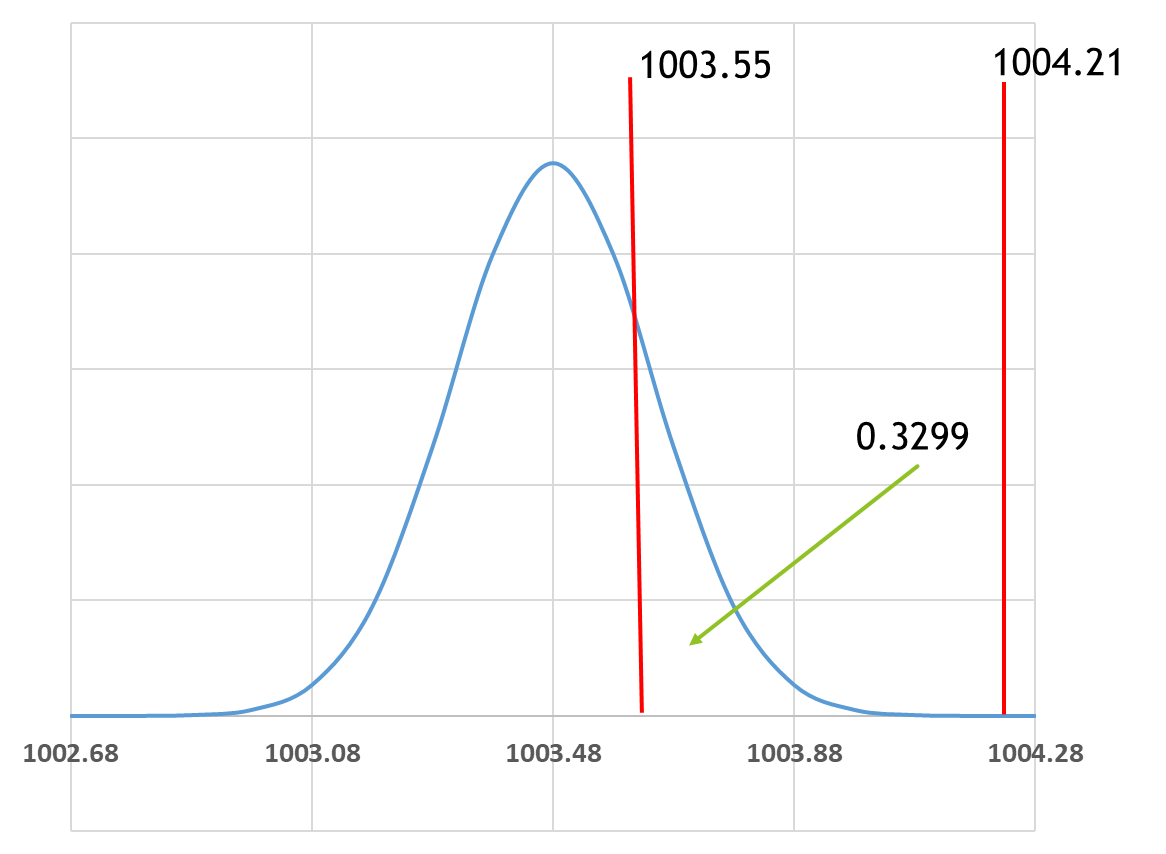
1. Let’s look at when do I commit a Type II error: When I let the machine continue given that machine is not doing all right.
2. Now during Hypothesis is true we had built our decision rule as below:

we will let the machine continue if our sample means are between 1003.55 mg and 1004.21 mg

1. Now I pick a μ = 1004.18 (arbitrary) such that it is not equal to 1003.88mg.The below curve is established:

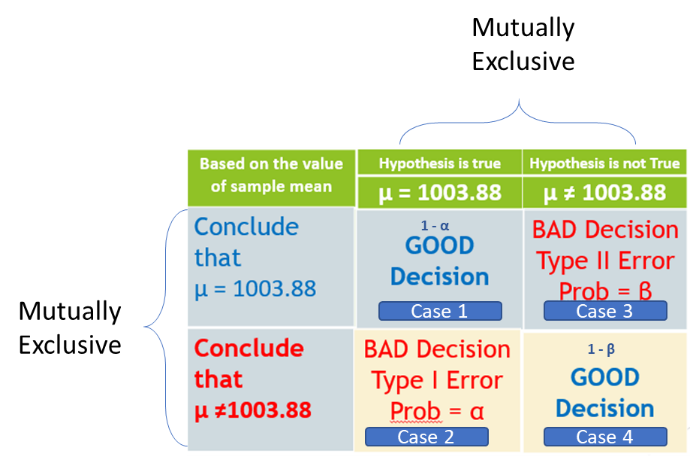


1. The bars are showing the decision rule limits. Now as per our decision rule that we will let the machine run if our sample mean falls within the decision boundaries. In the above case what will be the probability that the sample mean falls within the given decision boundary? It’s 0.5713 as marked in the diagram. So I let the machine run even if it is packing more. This is the idea behind the type II error.
2. Now μ can take another value = 1003.48 and form see the probability of a sample having sample mean within the given range:



We see the probability is 0.3299

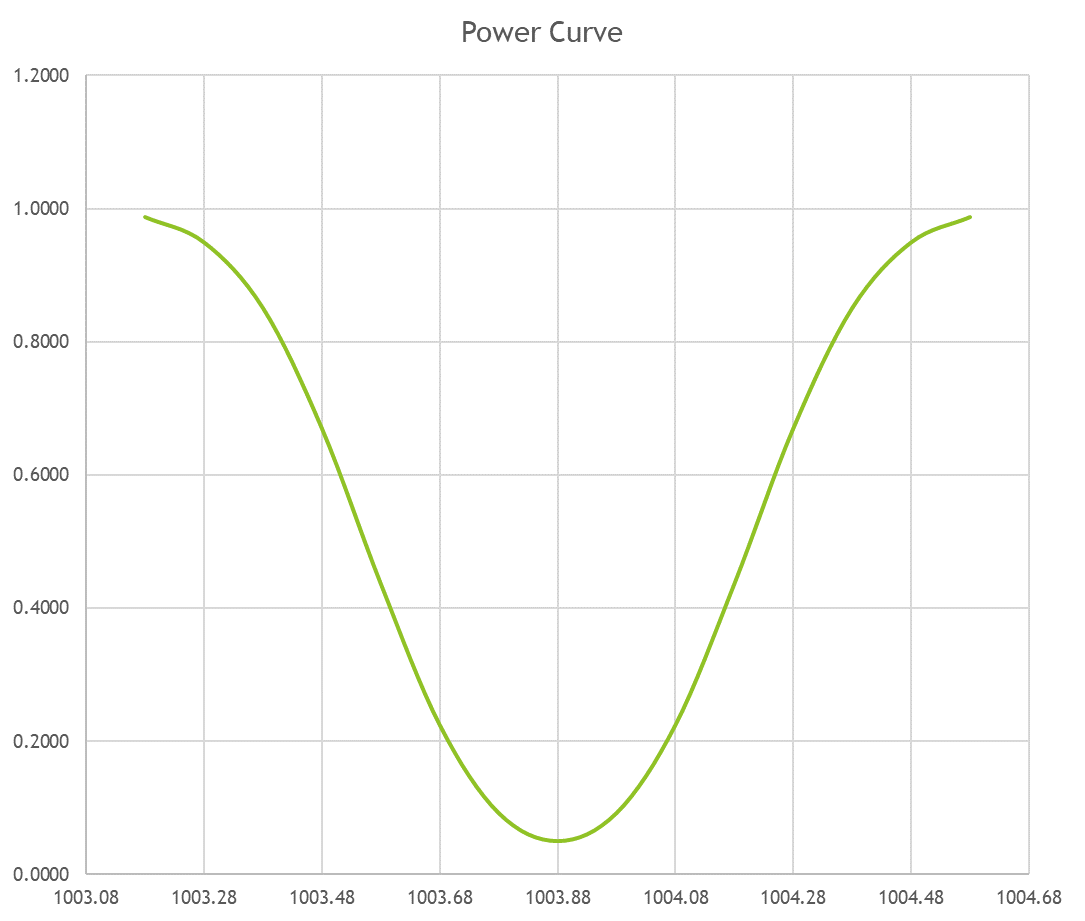
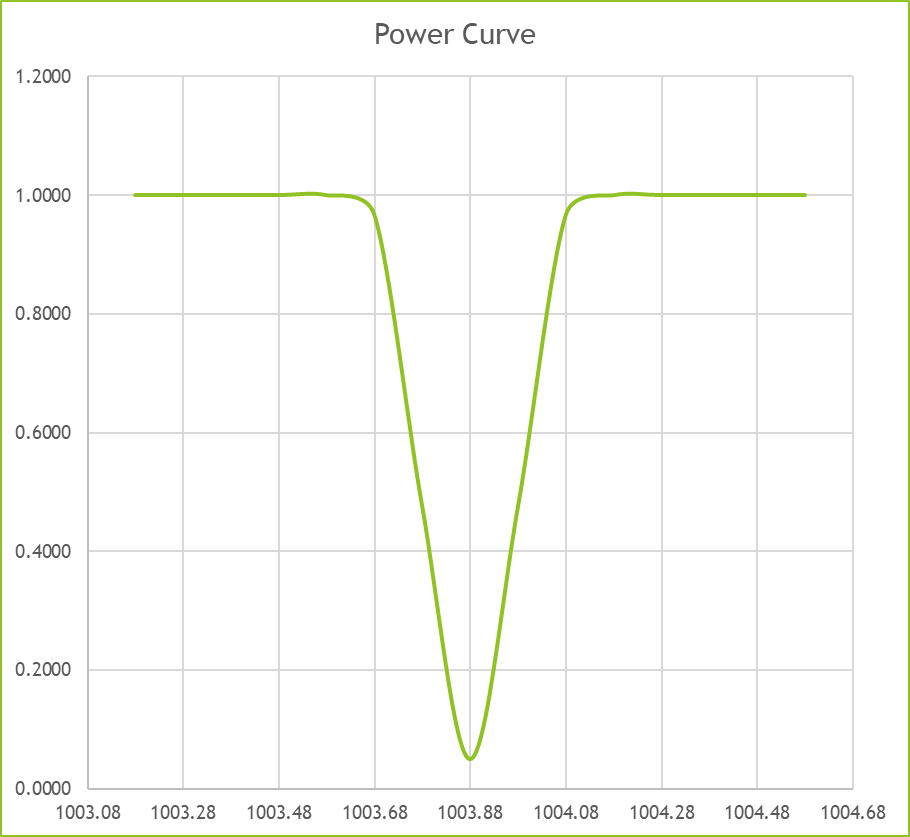
1. Notice that the mutual exclusivity of columns and rows which tells that:
   1. Either I can be on the first column or the 2nd column but not both.
   2. Either I can be on the 1st row or the 2nd row but not both.



1. So the bottom line is:
   1. we don’t know μ. All we know is μ can take infinite values.
   2. We have our decision rule fixed.
   3. Thus there is a corresponding value of β corresponding to each value of μ≠1003.88.
   4. The moment μ = 1003.88mg we are no longer in the 2nd column. I am in first column.
   5. The mutual exclusivity of the rows and columns gives the following information:
      1. If I am on the 1st row I can have the possibility of committing only type II error but not type I.
      2. If I am on the 2nd row I can commit type I error but not type II.
      3. Similarly if I am on the 1st column I can commit type I not type II.
      4. Similarly if I am on the 2nd column I can commit type II not type I.
      5. Thus I can commit either type I error or type II error but not both the errors.
   6. Who decides the value of α? We and in fact we can drive it down to as low as possible.
   7. Who decides β? I can’t set a fixed value of β. All I can say for this μ this is the corresponding β. So I can give you a list of possible values of β. So if I have a choice between these two I would rather go with type I error rather than type II error.
2. **Conclusion: Thus as seen in** [**Implication of α**](#_Implication_of_α) **that setting a high value of α might result in sending an innocent person go to jail. While setting high β will result in letting a guilty person walk free.** So as a judge we can set a value of α but never a value of **β.** Remember these two lines:
   1. “[Let a hundred guilty be acquitted, but one innocent should not be convicted](https://www.quora.com/What-change-does-Let-a-hundred-guilty-be-acquitted-but-one-innocent-should-not-be-convicted-bring)” is there in judicial system.
   2. Also in case of judicial system a judge will never say “I find this person innocent”. He will say “based on the evidences the prosecution failed to prove this person guilty so I am letting him go”. Whether the person is actually guilty or not the judge will never know.

### OC Curve and Power Curve

So we have seen for a particular value of μ there is a corresponding β. So if we plot β vs. μ it gives **Operating Characteristics** curve and if we plot (1- β) vs. μ it gives the **Power Curve** as shown below:

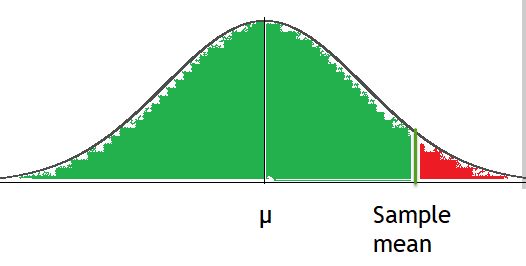
# One sided test with rejection region in right tail

## Quadrants of operation

|  |  |  |
| --- | --- | --- |
|  | **H0** | **H1** |
|  | **μ <= rating** | **μ > rating** |
| **Conclude μ <= rating** | Good decision | β (Type 2 Error) |
| **Conclude μ > rating** | α (Type 1 Error) | Good decision |

## Rejection Region

We will reject null hypothesis that **μ <= rating** if we find a sample whose mean is much bigger than given rating. The how much bigger value is determined by how much α we are willing to tolerate.



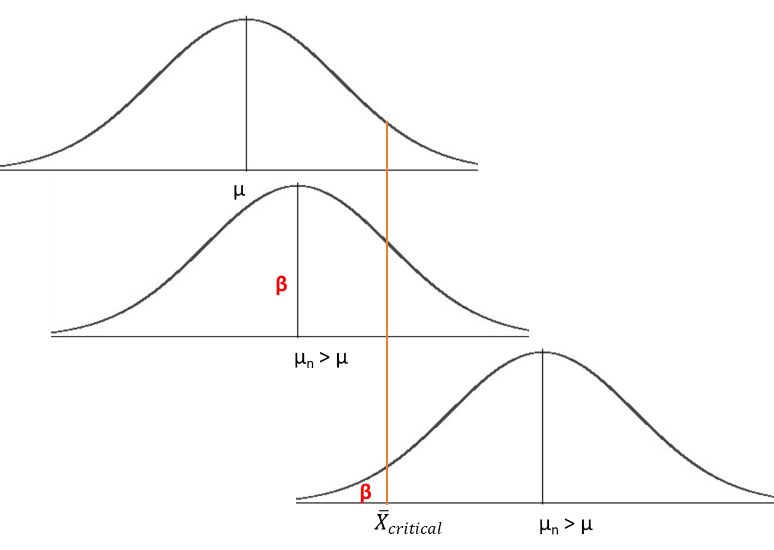
Thus the area of red region is α

## Calculation of critical value of sample mean

To reject null hypothesis sample mean should be > …. (i)

## Calculation of β

To calculate β we have to consider the error that we would be committing if we **Conclude μ <= rating** while **μ > rating**. This conclusion will be done if we get a sample mean << given µn. Thus our operating region will be as follows:



So we fix as calculated from and go on shifting the curve for different **μ > rating.**

|  |  |  |
| --- | --- | --- |
| **μ** | **Z value** | **β** |
| μ1 |  | NORM.DIST(Z1,0,1,cumulative) |
| μ2 |  | NORM.DIST(Z2,0,1,cumulative) |
| **…** |  |  |
| μn |  | NORM.DIST(Zn,0,1,cumulative) |

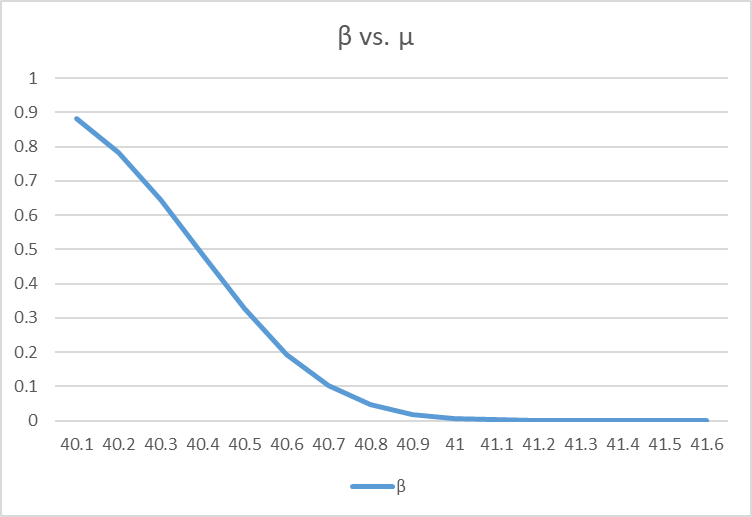
Note: μi+1 > μi

## Sample Example

|  |  |
| --- | --- |
| **Rated μ** | 40 |
| **Rated σ** | 1.5 |
| **Sample mean** | 38 |
| **n** | 40 |
| **α** | 0.05 |
| **Z\_α** | 1.644853627 |
| **Critical sample mean for right tail rejection** | 40.39011129 |

|  |  |  |  |
| --- | --- | --- | --- |
| **sl.no.** | **μ** | **Z value** | **β** |
| 1 | 40.1 | 1.192244 | 0.883417 |
| 2 | 40.2 | 0.781283 | 0.782682 |
| 3 | 40.3 | 0.370322 | 0.644429 |
| 4 | 40.4 | -0.04064 | 0.483792 |
| 5 | 40.5 | -0.4516 | 0.325779 |
| 6 | 40.6 | -0.86256 | 0.19419 |
| 7 | 40.7 | -1.27352 | 0.101417 |
| 8 | 40.8 | -1.68448 | 0.046044 |
| 9 | 40.9 | -2.09544 | 0.018066 |
| 10 | 41 | -2.5064 | 0.006098 |
| 11 | 41.1 | -2.91737 | 0.001765 |
| 12 | 41.2 | -3.32833 | 0.000437 |
| 13 | 41.3 | -3.73929 | 9.23E-05 |
| 14 | 41.4 | -4.15025 | 1.66E-05 |
| 15 | 41.5 | -4.56121 | 2.54E-06 |
| 16 | 41.6 | -4.97217 | 3.31E-07 |

### OC Curve



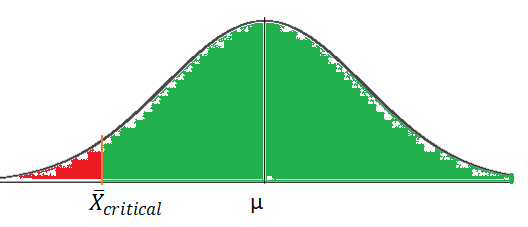
# One sided test with rejection region in left tail

## Quadrants of operation

|  |  |  |
| --- | --- | --- |
|  | **H0** | **H1** |
|  | **μ >= rating** | **μ < rating** |
| **Conclude μ >= rating** | Good decision | β (Type 2 Error) |
| **Conclude μ < rating** | α (Type 1 Error) | Good decision |

## Rejection Region

We will reject null hypothesis that **μ >= rating** if we find a sample whose mean is much smaller than given rating. The how much bigger value is determined by how much α we are willing to tolerate.



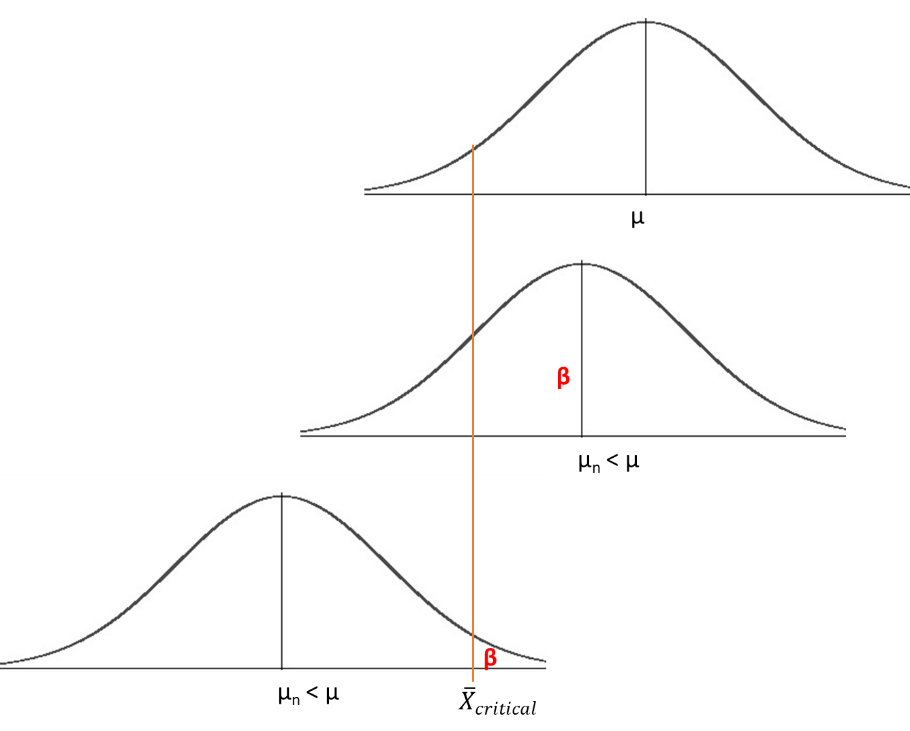
Thus the area of red region is α

## Calculation of critical value of sample mean

To reject null hypothesis sample mean should be < …. (i)

## Calculation of β

To calculate β we have to consider the error that we would be committing if we **Conclude μ >= rating** while **μ < rating**. This conclusion will be done if we get a sample mean >> rated µ. Thus our operating region will be as follows:



So we fix as calculated from and go on shifting the curve for different **μ < rating.**

|  |  |  |
| --- | --- | --- |
| **μ** | **Z value** | **β** |
| μ1 |  | NORM.DIST(Z1,0,1,cumulative) |
| μ2 |  | NORM.DIST(Z2,0,1,cumulative) |
| **…** |  |  |
| μn |  | NORM.DIST(Zn,0,1,cumulative) |

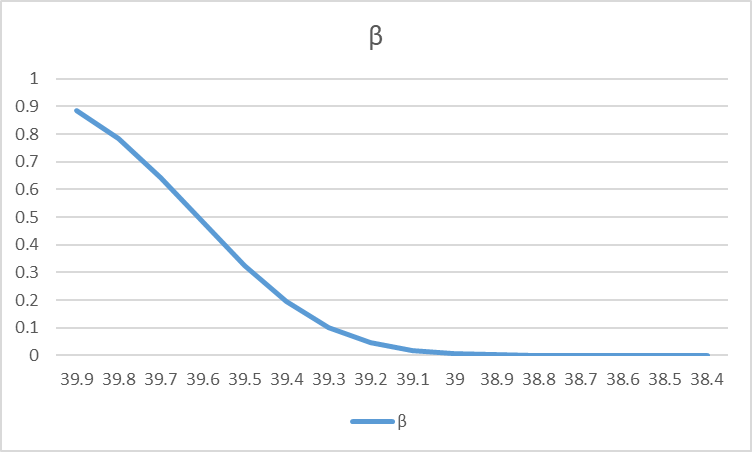
Note: μi+1 < μi

## Sample Example

|  |  |
| --- | --- |
| **Rated μ** | 40 |
| **Rated σ** | 1.5 |
| **Sample mean** | 38 |
| **n** | 40 |
| **α** | 0.05 |
| **Z\_α** | 1.644853627 |
| **Critical sample mean for right tail rejection** | 40.39011129 |
| **Critical sample mean for left tail rejection** | 39.60988871 |

|  |  |  |  |
| --- | --- | --- | --- |
| **μ** | **Z value** | **CI** | **β** |
| 39.9 | -1.19224 | 0.116583 | 0.883417 |
| 39.8 | -0.78128 | 0.217318 | 0.782682 |
| 39.7 | -0.37032 | 0.355571 | 0.644429 |
| 39.6 | 0.040639 | 0.516208 | 0.483792 |
| 39.5 | 0.4516 | 0.674221 | 0.325779 |
| 39.4 | 0.862561 | 0.80581 | 0.19419 |
| 39.3 | 1.273522 | 0.898583 | 0.101417 |
| 39.2 | 1.684482 | 0.953956 | 0.046044 |
| 39.1 | 2.095443 | 0.981934 | 0.018066 |
| 39 | 2.506404 | 0.993902 | 0.006098 |
| 38.9 | 2.917365 | 0.998235 | 0.001765 |
| 38.8 | 3.328326 | 0.999563 | 0.000437 |
| 38.7 | 3.739287 | 0.999908 | 9.23E-05 |
| 38.6 | 4.150248 | 0.999983 | 1.66E-05 |
| 38.5 | 4.561209 | 0.999997 | 2.54E-06 |
| 38.4 | 4.97217 | 1 | 3.31E-07 |

### OC Curve



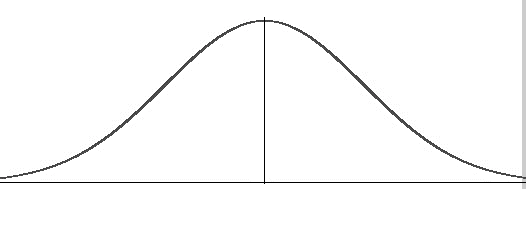
# Two sided test

## Quadrants of operation

|  |  |  |
| --- | --- | --- |
|  | **H0** | **H1** |
|  | **μ = rating** | **μ ≠ rating** |
| **Conclude μ = rating** | Good decision | β (Type 2 Error) |
| **Conclude μ ≠ rating** | α (Type 1 Error) | Good decision |

## Rejection Region

We will reject null hypothesis that **μ ≠ rating** if we find a sample whose mean is outside the threshold of the given rating. The threshold value is determined by how much α we are willing to tolerate.



μ

**α/2**

**α/2**

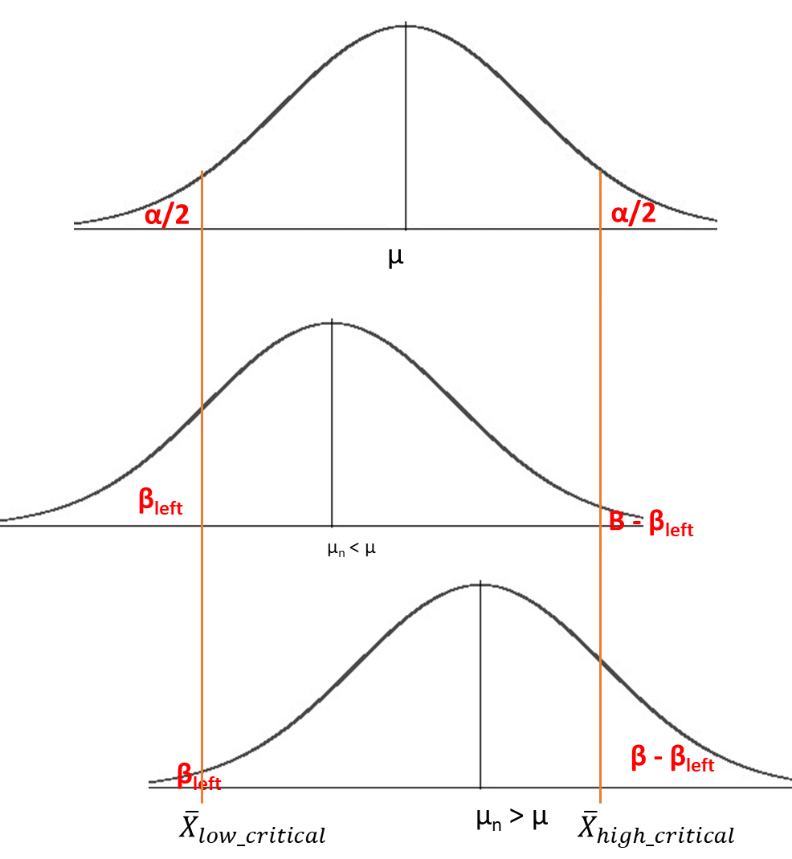
Thus the area of red region is α

## Calculation of critical value of sample mean

To reject null hypothesis sample mean should be within and

## Calculation of β

To calculate β we have to consider the error that we would be committing if we **Conclude μ = rating** while **μ <> rating**. This conclusion will be done if we get a sample mean within threshold of rated µ. Thus our operating region will be as follows:



So we fix and as calculated from and go on shifting the curve for different **μ <> rating.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **μ** | **Z values** | **CI** | | **β** |
| μ1 |  | CI1 = NORM.DIST(Z1\_high,0,1,cumulative) - NORM.DIST(Z1\_low,0,1,cumulative) | | 1 - CI1 |
| μ2 |  | CI2 = NORM.DIST(Z2\_high,0,1,cumulative) - NORM.DIST(Z2\_low,0,1,cumulative) | | 1 - CI2 |
| **…** |  |  | |  |
| μn |  | CIn = NORM.DIST(Zn\_high,0,1,cumulative) - NORM.DIST(Zn\_low,0,1,cumulative) | 1 - CIn | |

Note: μi+1 <> μi

## Sample Example

|  |  |
| --- | --- |
| **Rated μ** | 40 |
| **Rated σ** | 1.5 |
| **Sample mean** | 38 |
| **n** | 40 |
| **α** | 0.05 |
| **Z\_α** | 1.644853627 |
| **Critical sample mean for right tail rejection** | 40.39011129 |
| **Critical sample mean for left tail rejection** | 39.60988871 |
| **Z\_α/2 for two tailed** | 1.959963985 |
| **Lower limit of sample mean for two tailed** | 39.53515373 |
| **Upper limit of sample mean for two tailed** | 40.46484627 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **μ** | **Z\_low** | **Z\_high** | **CI** | **β** |
| 37 | 10.68915 | 14.60907 | 0 | 1 |
| 37.5 | 8.580962 | 12.50089 | 0 | 1 |
| 38 | 6.472776 | 10.3927 | 4.81E-11 | 1 |
| 38.5 | 4.364591 | 8.284519 | 6.37E-06 | 0.999994 |
| 39 | 2.256406 | 6.176334 | 0.012023 | 0.987977 |
| 39.5 | 0.148221 | 4.068149 | 0.44106 | 0.55894 |
| 40 | -1.95996 | 1.959964 | 0.95 | 0.05 |
| 40.5 | -4.06815 | -0.14822 | 0.44106 | 0.55894 |
| 41 | -6.17633 | -2.25641 | 0.012023 | 0.987977 |
| 41.5 | -8.28452 | -4.36459 | 6.37E-06 | 0.999994 |
| 42 | -10.3927 | -6.47278 | 4.81E-11 | 1 |
| 42.5 | -12.5009 | -8.58096 | 4.7E-18 | 1 |
| 43 | -14.6091 | -10.6891 | 5.72E-27 | 1 |

### OC Curve

