Errors in Hypothesis Testing and How to Calculate Them

CSE 407 - Class 3

LEC RAIYAN RAHMAN
Dept of CSE, MIST
raiyan@cse.mist.ac.bd

Remember that we have a limited sample?

So, the decisions we give from our test may not always be correct.

“Statistics is useful but not always Perfect!”

But to compensate, in AP stat, we’ll also calculate the chance of our hypothesis test result being wrong and mention that with the test result!

Errors in hypothesis Testing

H_0 (Actual Case) / Test Outcome	"Accepted" / Fail to Reject	Rejected
H_0 True	We're correct!	Type I Error
H_0 False	Type II Error	We're correct!

Type I Error

- It's the probability that we reject the null hypothesis H_0 when it is, in fact, true. Also called **false positive** error.
- It's denoted by **α**. The probability of making type I error is also **α**.
- As you may already guess, it's the more serious error of the two cases.
- The responsibility of choosing **α** depends solely on the research/data analyst.
- So, depending on your test data, you better choose **α** carefully.
- The less confident you are in your test data, the lower the value of **α** should be
- The reason is hopefully apparent from the student t table!. See how critical value is larger for lower **α** with the same sample size.

Type II Error

- It's the probability that we accept (fail to reject) the null hypothesis H_0 when it is, in fact, false. Also called **false negative** error.
- It's denoted by $\boldsymbol{\beta}$.
- It's the less serious error of the two cases.
- The value of $\boldsymbol{\beta}$ depends on the sample data (n and $\boldsymbol{\sigma}$ or s) and also μ and indirectly $\boldsymbol{\alpha}$ as well.
- We can calculate the exact probability of making a type II error ($\boldsymbol{\beta}$).

Power of the Test

- It's the probability that we reject the null hypothesis H_0 when it is, in fact, false.
- or simply, It's the probability that H_0 is false and our test also rejects it.
- So it's the opposite of Type II error.
- And hence, it's calculated as 1-β.

Examples!

Exercise

1. The guard stole the mobile phone (not innocent) and you also thought him to be the thief.
2. The guard was innocent but you assumed he was the thief. Type I error
3. The guard was not innocent yet your probe found him to be a very loyal person! Type II error
4. The guard did steal the phone and you and your friends detected him to be the thief.

Calculating Type II error (Ex 6)

Data: You're working as maintenance engineer at Samsung. The new SSDs that they are making have an advertised read speed of 550mb/s. But after purchasing some users are reporting much lower read speeds with different values.

To verify their claims, you decided to gather 100 SSDs ready to be shipped to stores from your production line. After careful testing, the read speed of these test SSDs were found to be 548.9 mb/s. You know from previous tests that the population std deviation of your SSDs is 6.5 mb/s.

Question: Test the claim by the customers with 95% confidence and establish if the advertised read speed is **less than** 550 mb/s. Also, find the power of the test if, from historical data of similar model, $\mu = 547.5$ mb/s.

Calculating Type II error - Steps

1. Find out the critical value. (for what value of Z we'll reject H_o)
2. Find out for which value of $\bar{X}$ we'll reject H_0.
3. Now, for given μ , find the probability of type II error (β). *
4. Calculate Power of test by 1- β (or directly in step 3)

*In this step remember to choose the overlapping graph area correctly and choose the z table accordingly.

Step 1

1. Here C=0.95. So $\boldsymbol{\alpha}$ = 0.05. Since it's a one tailed test (μ <550), $\boldsymbol{\alpha}$ = 0.05.
2. So from negative z table, the corresponding closest critical value for 0.05 is -1.65.

Step 2

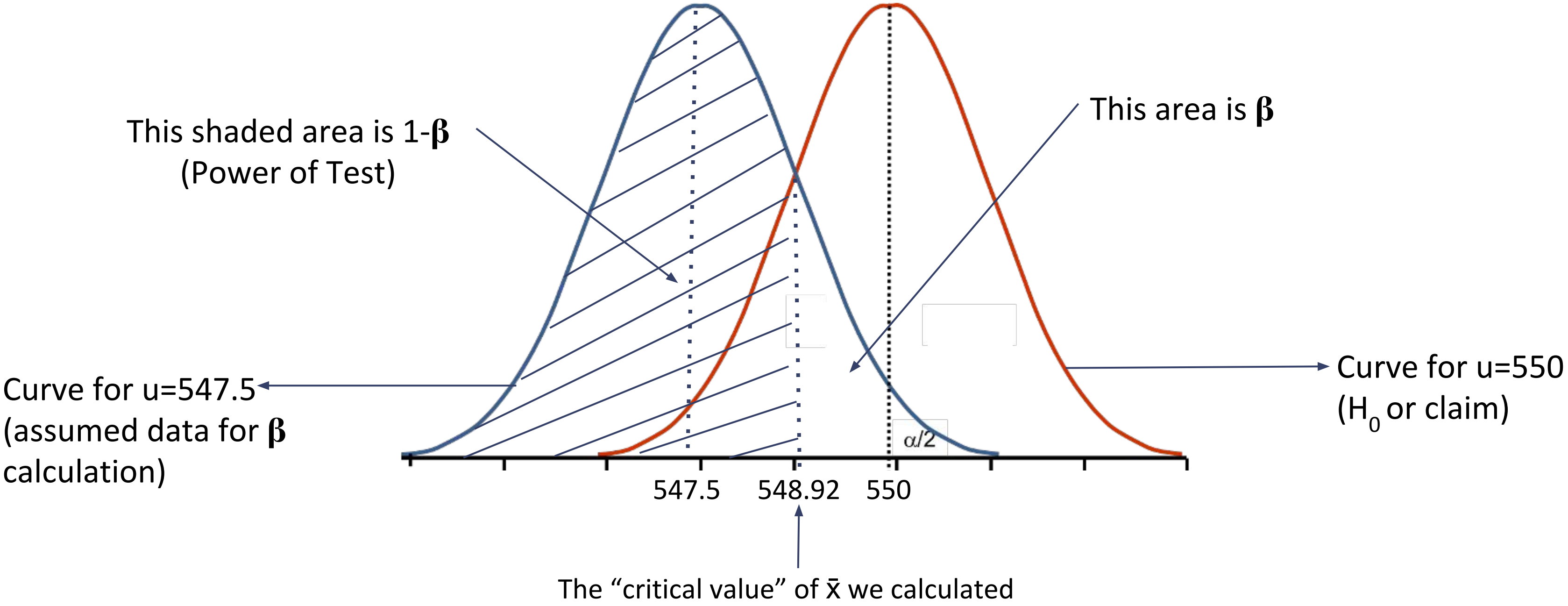
1. Now, let's inversely find for which value of $\bar{x}$ we'll reject H_0 .
2. This can be found by simply rearranging the z equation in previous problem's step 5.

$$\bar{x} = \mu_0 + \frac{\boldsymbol{\sigma}}{\sqrt{n}} z$$

$= 550 + (6.5/\sqrt{100}) * (-1.65)$

$= 548.92$

Step 3 - Graph for β calculation



Step 3

1. Now, let's see the probability of us actually accepting H_o even though it's false. (check out the graph in next slide for better understanding)

P ($\bar{x} > 548.92 \mid \mu = 547.5$) or

P $\left(z > \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right)$ = 548.92-547.5 / (6.5 / √100)

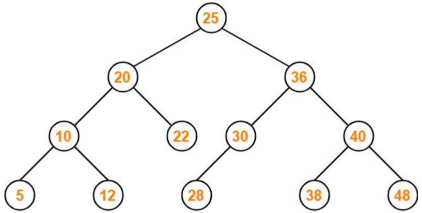
= P (z> 2.19) = 0.0143

Step 4

From negative z table the probability of the critical value being greater than 2.19 is .0143 ($\boldsymbol{\beta}$)

Accordingly, the power of the test is: $1 - \boldsymbol{\beta} = 0.9857$

Exercise 7



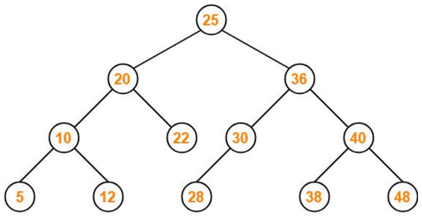
Binary Search Tree

Suppose that, in your undergraduate thesis, you designed an algorithm that searches a BST with a depth of at most 20 **within** an avg of 2.25 ms (max time). However, in your pre-defence the defence team ran your algorithm and found the search time to be 2.51 ms for a particular BST.

You, very confident in your algorithm, claimed that the problem might be with the particular machine's specification. To prove that your algorithm will search a node within 2.25ms, irrespective of machine specifications, you decided to test the code again using the same tree and same search node in 5 different machines of varied specs.

The search time for each of the 5 machines are 2.23ms, 2.21ms, 2.36ms, 2.44ms and 2.55ms. Now, based on this test data, with 90% level of confidence (or $\alpha = 0.1$), will you be able to defend your claimed 2.25ms search time or would you need to revise it in your final defence? **Also, what would you say before the defence board your power of test was when the avg time of such algorithm is usually 2.38 ms?**

Exercise 7 - Soln



Binary Search Tree

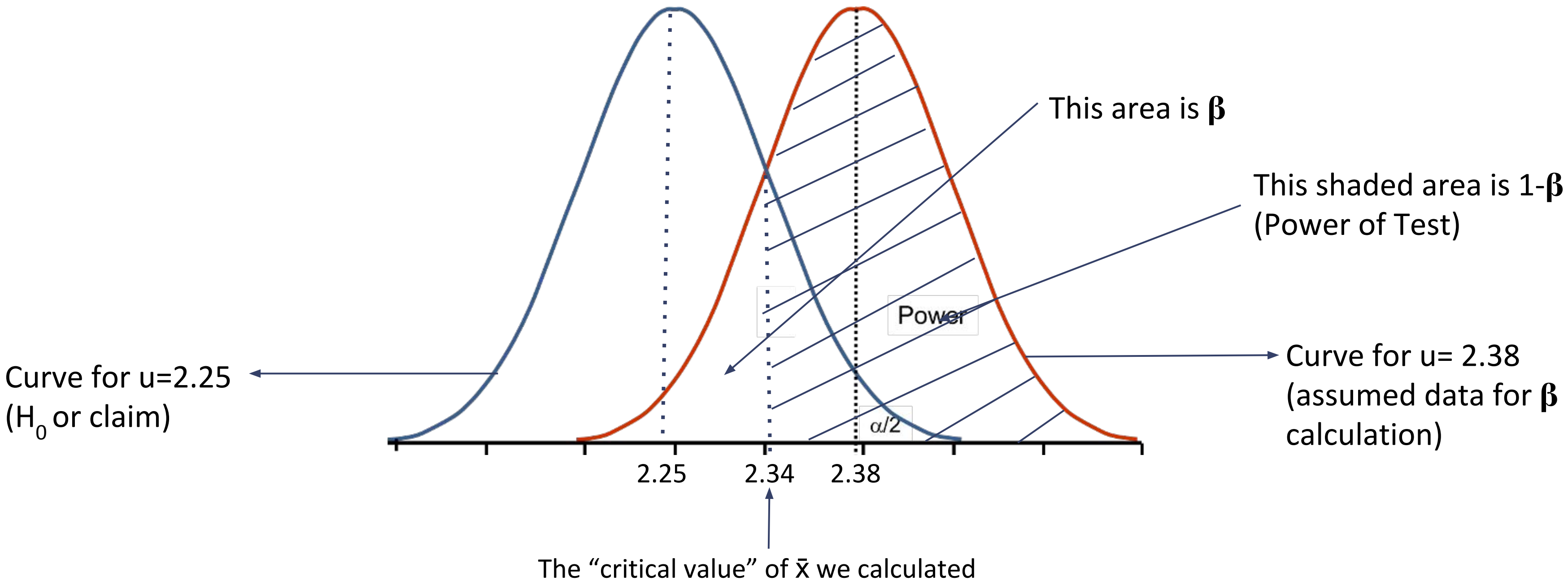
1. Firstly, let's find the critical value. It's a one-tailed test (right tailed) and α = 1 - 0.90 = 0.01. From t table, critical value is 1.533 (for df =4 and $t_{0.90}$). We chose t table since n = 5 and population std deviation is also unknown.
2. Let's now find for which value of $\bar{x}$ we'll accept H_0 (similar to how we accepted or rejected based on z value). From the corresponding equation, it comes to be 2.34.

3. Now, the probability that we'll make a type two error, in this case, is, when we'll accept the null hypothesis (if $\bar{x}$ <2.34) although the average value of similar algorithms is 2.38 (given in question).

   So it's P ($\bar{x}$ <2.34) | μ = 2.38). See the graph in next slide.
   By calculating this in the similar way like the last problem we have P (z < -0.635) which, from the negative z table is 0.263.

4. So, β = 0.263 and Power of test = 1 - β = 0.737.

Graph for β calculation for Ex 7



*from the graph we can see why we should use negative cumulative z table to find the β for both cases by simply looking at the shape of the area for β.

Calculating Type II error for two tailed tests (Ex 8)

Data: You're working as maintenance engineer at Samsung. The new SSDs that they are making have an **advertised read speed of 550mb/s.** But after purchasing some users are reporting much lower read speeds with different values.

To verify their claims, you decided to gather 100 SSDs ready to be shipped to stores from your production line. After careful testing, the read speed of these test SSDs were found to be 548.9 mb/s. You know from previous tests that the population std deviation of your SSDs is 6.5 mb/s.

Question: Test the claim by the customers with 95% confidence and establish if the advertised read speed is **NOT** 550 mb/s. Also, find the power of the test if, from historical data of similar model, **μ = 550.5 mb/s.**

Calculating Type II error for two tailed tests - Steps

1. Find out the critical values. (for what value of Z we'll reject H_o)
2. Find out for which values of $\bar{X}$ we'll reject H_0.
3. Now, for given μ , find the probability of type II error (**β**). This calculation will be different. *
4. Calculate Power of test by 1- **β** (or directly in step 3)

*Here, as we'll soon see from the graph, we'll have to use both positive and negative z table to calculate **β**.

Calculating Type II error for two tailed tests (Full solution)

Click here to check out the full written solution to exercise 8

Some General Notes

1. **α** is set by the researcher based on the sample authenticity and their confidence on the test data.

2. **β** depends on **α**, μ, n and $\boldsymbol{\sigma}$/s. We can calculate the exact value of **β**.

3. Value of **β** is to be calculated slightly differently based on the type of the test (two vs one tailed). Similarly, the range/graph and the tables we'll use to find **β** will be different.

4. Depending on the area of **β**, we may need to use either positive or negative (or a combination of both) z tables to find **β**.