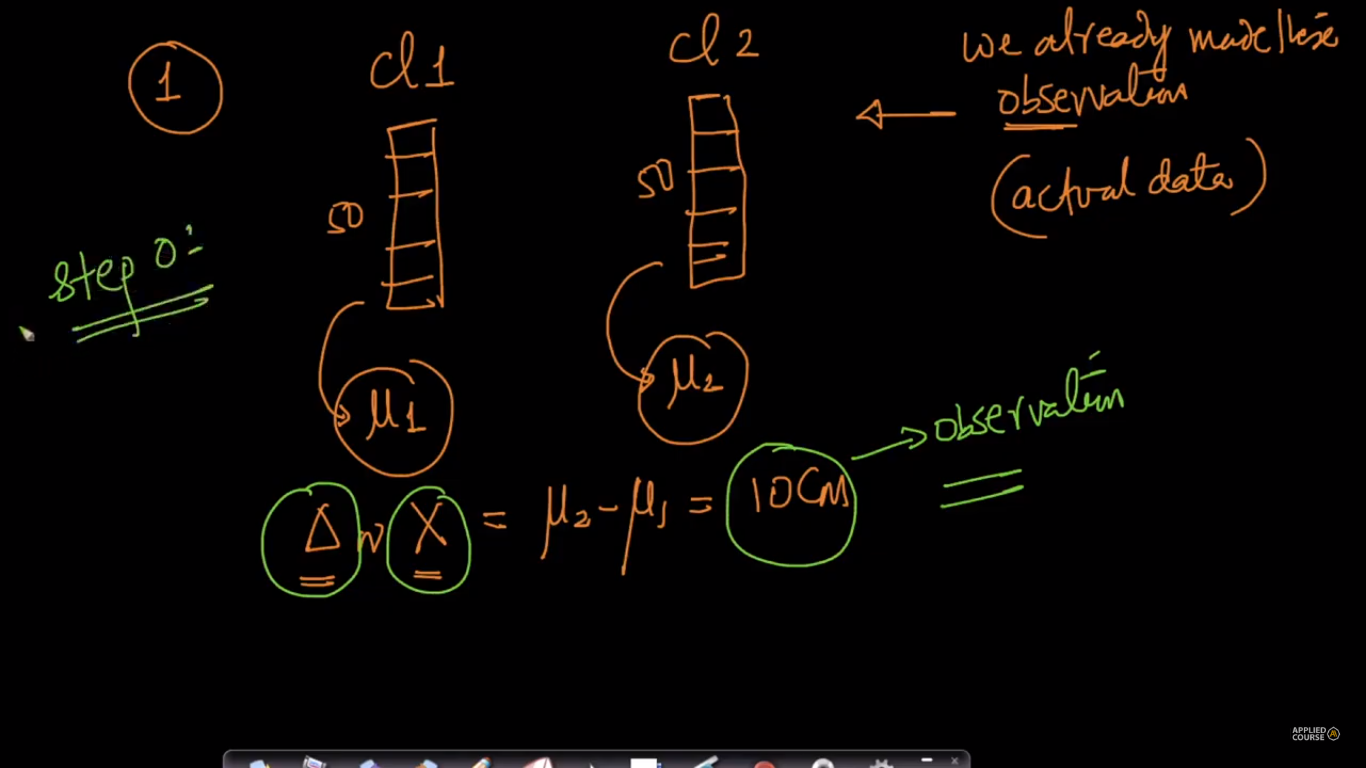
Let’s say we are given heights of two classes and we’ve to determine that there is difference between heights of class’s students.

Given heights of 50 students in each class.

1. First we make observation from given data, to check whether there is difference between heights of two class, we subtract the mean of both class.

Let’s say: mean2 – mean1 = 10cm

Here 10 cm is our observation, which is **truth** for given data, that means in reality this is the value, now we’ve to test for this truth and for the heights to be different the mean should be equal to or more than 12, because as mean becomes closer to 0 there will no difference between heights.



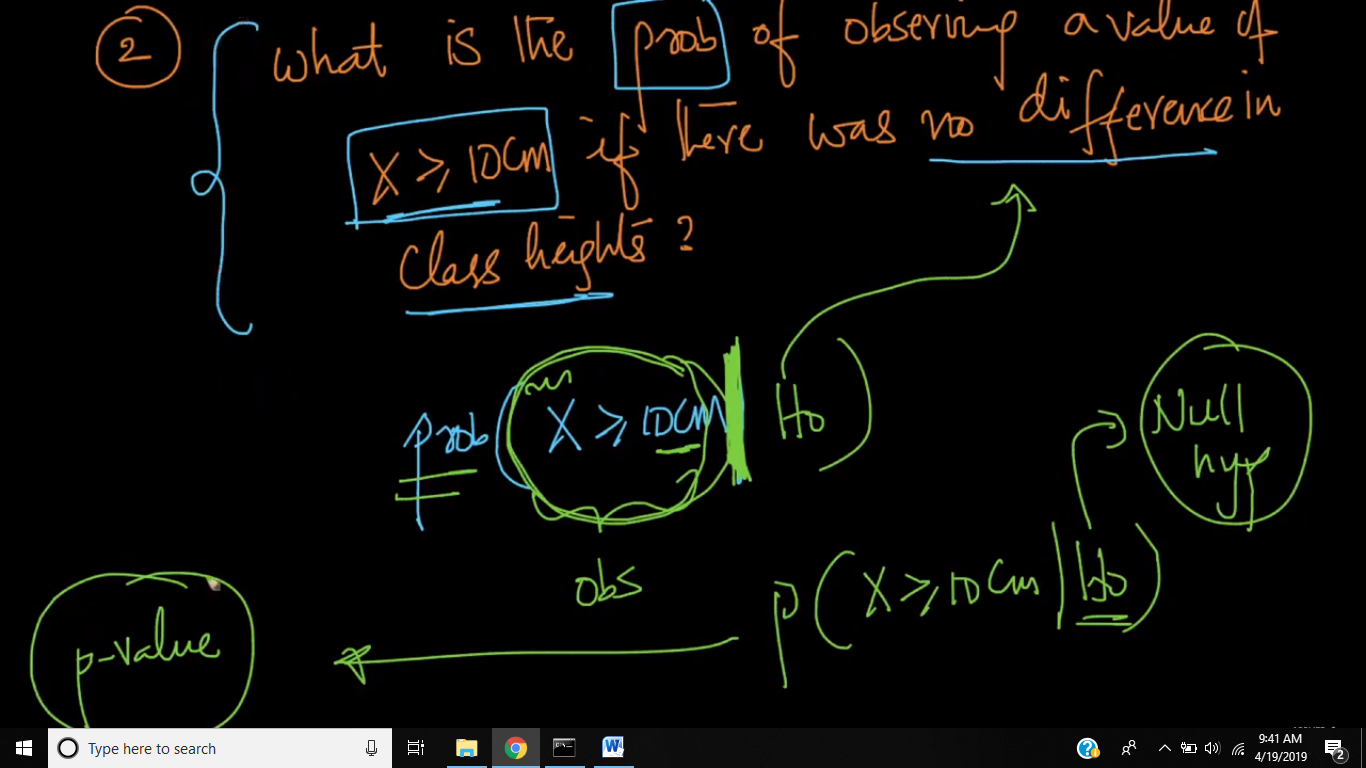
1. Now we choose the null hypothesis (basically it’s a contradiction of statement to determine).

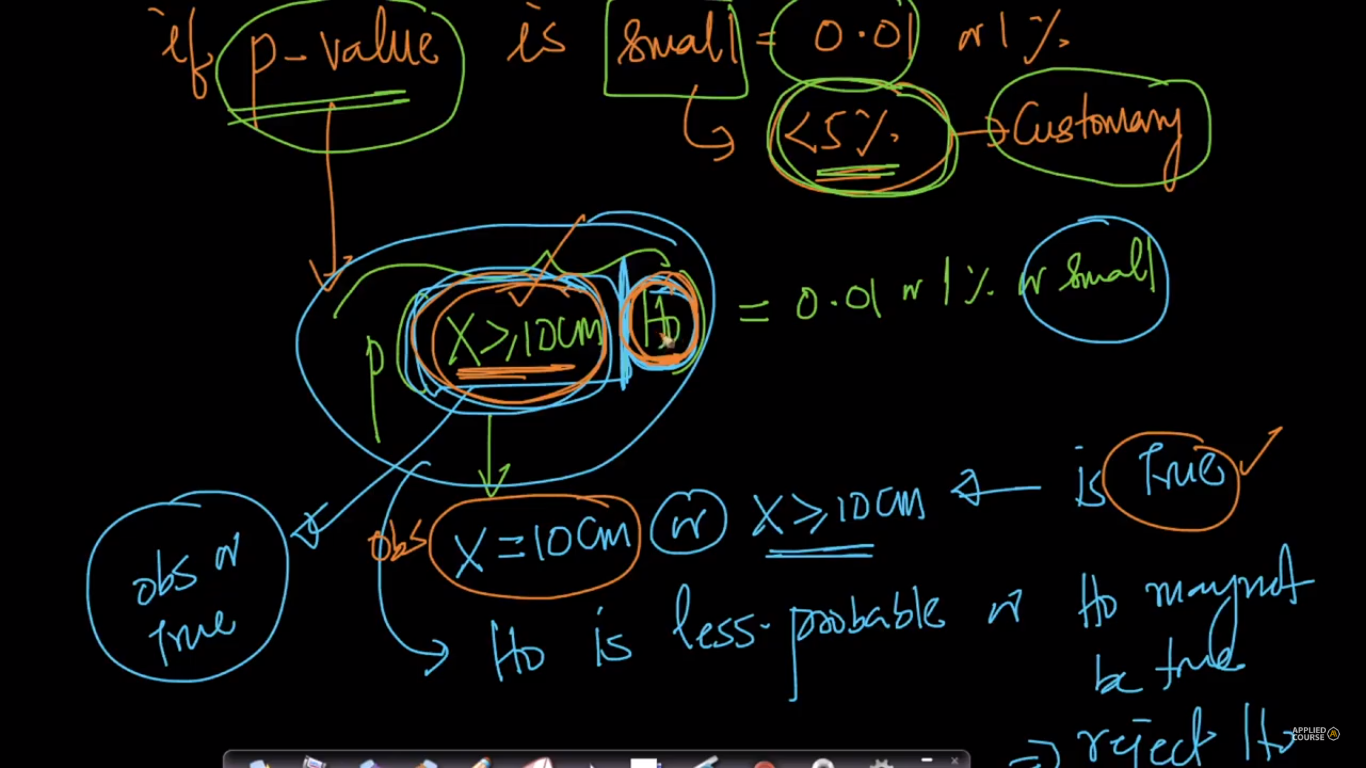
So for current example null hypothesis will be “ there is no difference in class’s heights”

Now we will find the probability of observing a value of x >= 10cm if there was no difference in class heights.

Note here x >= 10 cm is truth, so we are testing “there was no difference in class heights”.

If we get a probability < 0.05 that means null hypothesis is less probable or null hypothesis may not be true, because only null hypothesis can be wrong now as x >= 10 is something we know the truth.

01



1. Now we will do resampling, for bringing the situation according to null hypothesis, or by doing sampling, that means merging data from both class we are stimulating null hypothesis.

We will merge data from both class into a single sample(S) randomly.

And from that sample(S), we will randomly select data into two sample

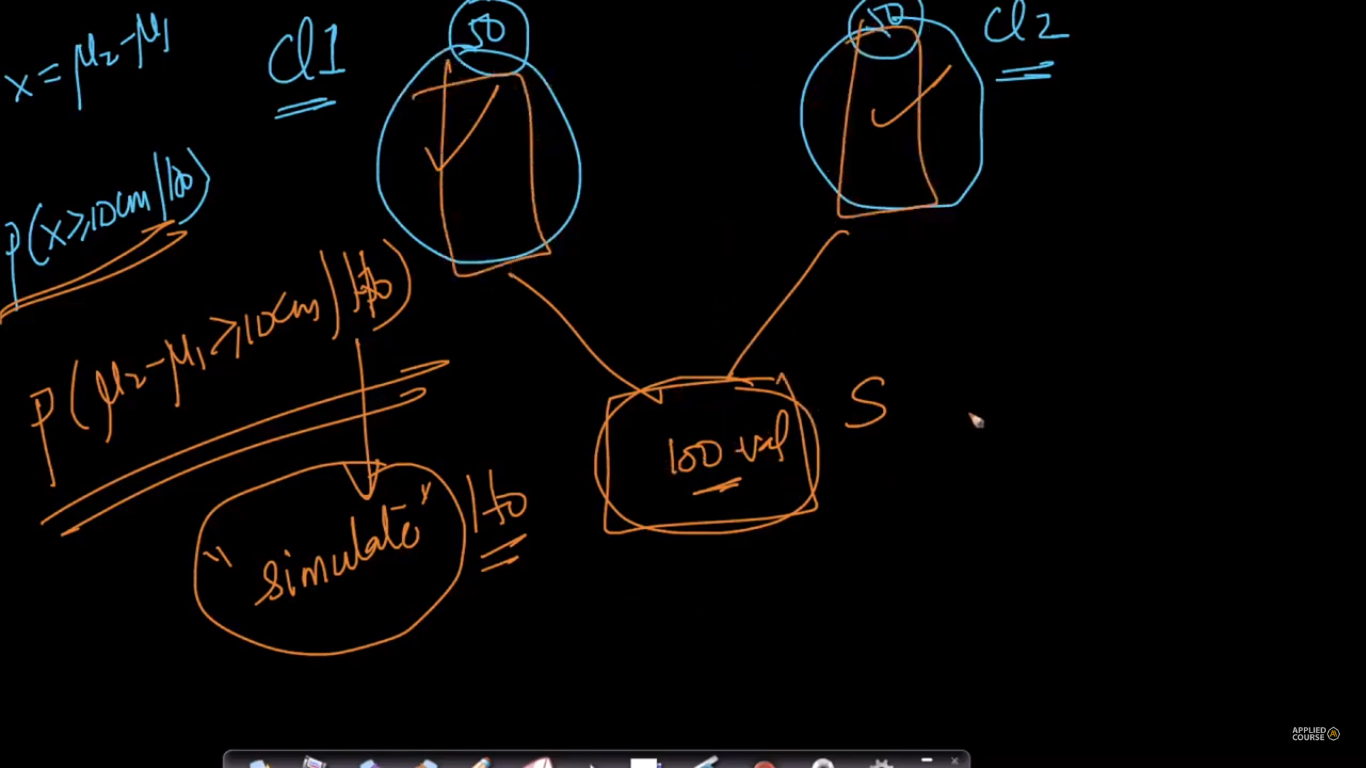
For current example we have 50-50 student’s height in each class, so we merge them randomly in single sample (S), so S have 100 data now.

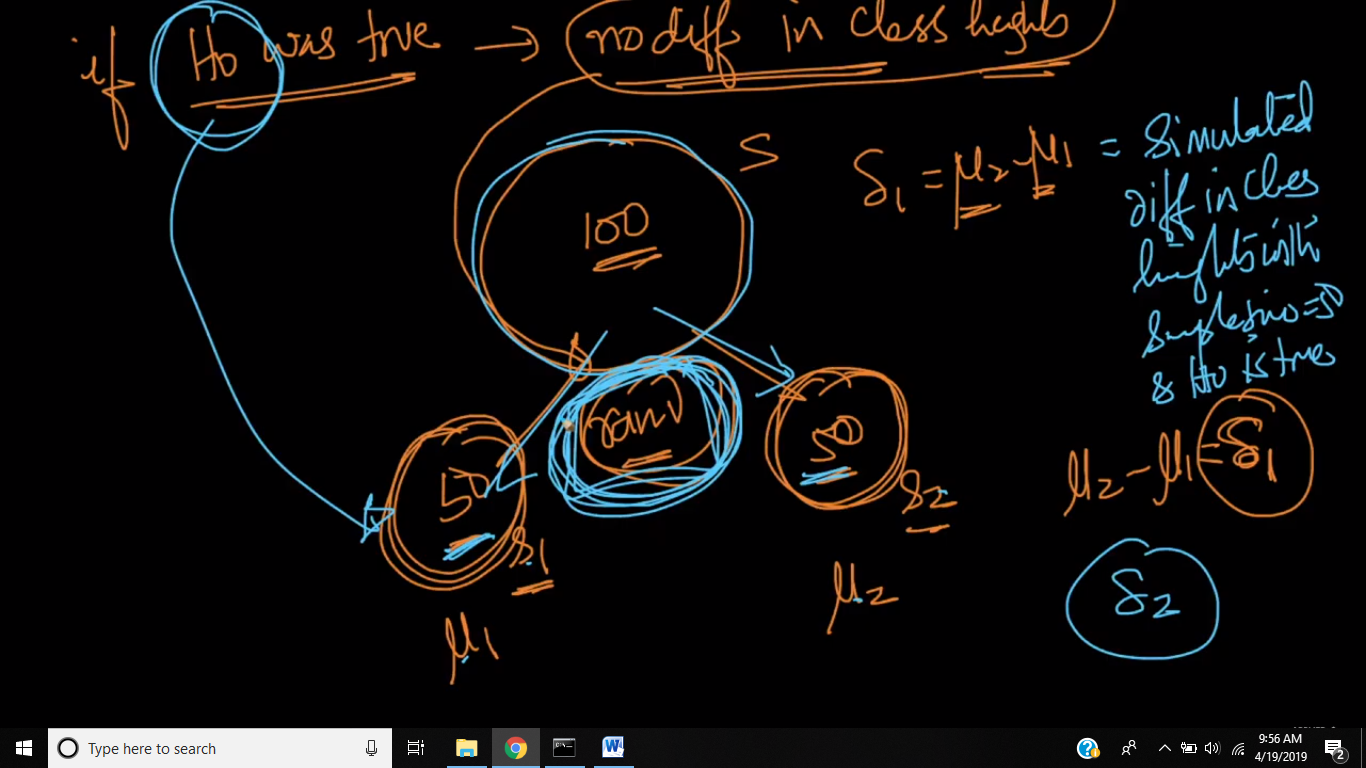
Now we will split them randomly into two samples of each with 50 data, let’s say them s1, s2.

Now we will find the difference in mean of both samples (s2 & s1).

**Repeat this step 3, ‘n’ no. of times let’s take n = 10000 here.**

**Why we are doing resampling:** since in resampling we are taking height of students from both class and merging them and again creating two samples, so this mean that after merging randomly there is no difference in generated two samples, or in other words we are creating situation for our null hypothesis or simulating null hypothesis





1. Now after repeating step 3 we have 10000 mean difference (which is also called stimulated mean),

Now we will sort them in ascending order.

Now we will place the original mean (that is mean of height of class which is 10cm) in this sorted means

And we will find how much percentage of samples mean have value greater than 10cm, which is what we were finding, probability that x>=10cm given that there is no difference in heights of two class. And this percentage is our p-value.

Suppose there are 2k people after 10 cm, therefore

P(x >= 10cm | H0) = 2k / 10k = 0.2 that is 20%.

Since p-value >= 5%, therefore we accept null hypothesis that there is no difference in height of two class, and we reject alternative hypothesis that there is difference In height of two class.

