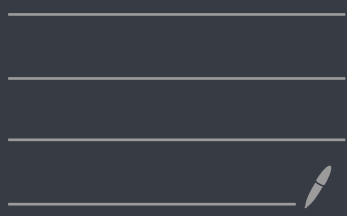
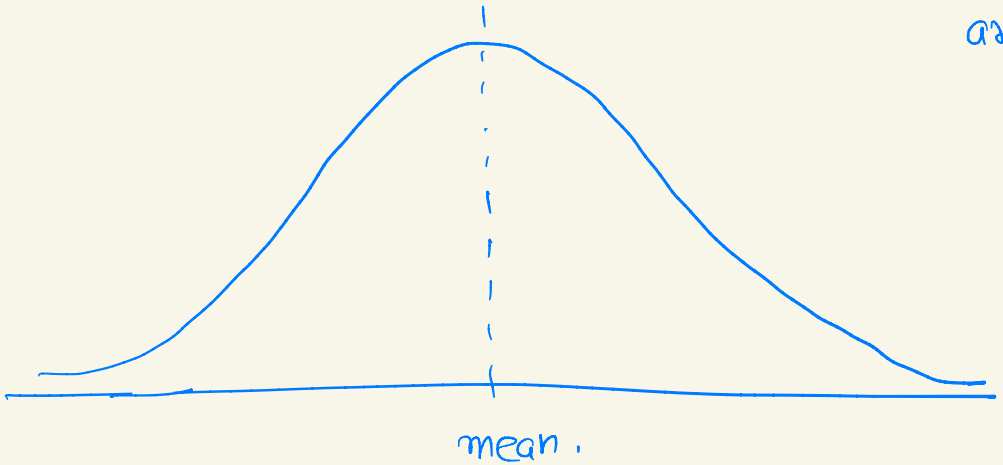




area under the curve
gives me the prob.

Q-18, pg-28

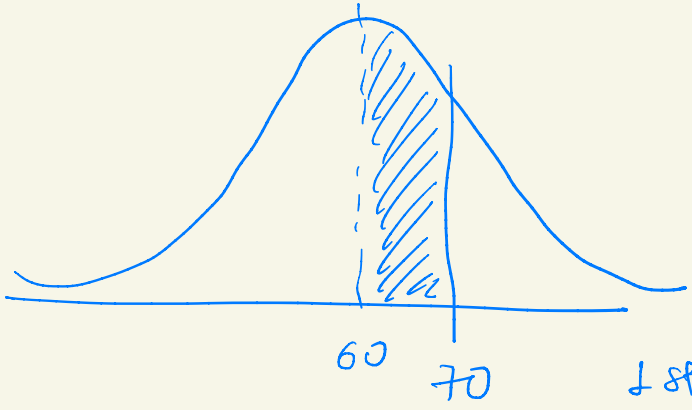
we can cal. z-score

$$z\text{-score} = \frac{\text{value} - \text{mean}}{\text{std}}$$

$$\Rightarrow \frac{70 - 60}{10} = 1$$

value is 1 std. dev. away from mean.

now, I need to find the area, right.
and I have a table which gives me
area b/w mean and z-score 'z'



1 std. away from mean

0.3413

Also, if I know X. I can predict f(x) is it follows Gaussian dist with mean μ and σ

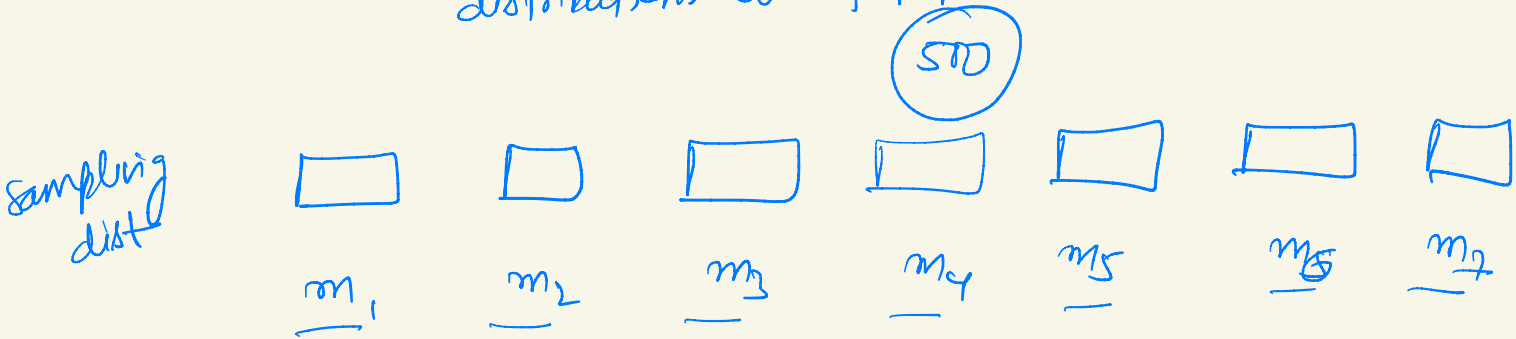
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

suppose I asked heights of 500 people

sample = 500 people

now I'll sample multiple distributions (sampling distributions) of size 100 out of this samples

and I'll take mean of all the sampling distributions. even if population don't



now these means will follow a normal distribution

$N(\mu, \sigma)$ and $\mu = \mu_s$ and $\sigma = \frac{\sigma_s}{\sqrt{n}}$

So, now, I can make a lot of comments and predictions about the height of people in India.

Take ex. of Heights of Students in 2 classes.

I want to know whether there is any diff. in the heights of students of class 1 & 2

I can form a hypo :-

Null :- $(\mu_2 - \mu_1) = 0$

Alt - $\mu_2 - \mu \neq 0$

I want to accept Alt. by Rejecting Null

Now, the point is can I reject null given this data and whether or not we can reject the alternate hypo. is based on p-value

p-value is the prob. of observing test statistic $(\mu_2 - \mu_1)$ if my null hypo. is true.

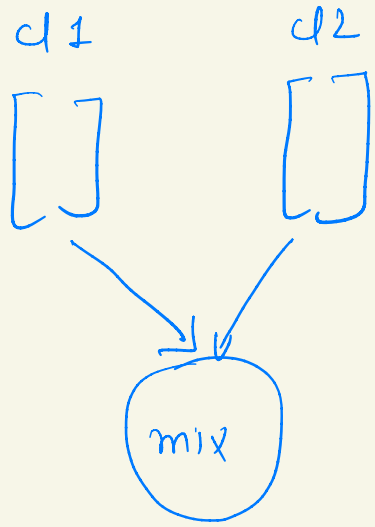
→ So first I assume null hypo is true and cal. $(\mu_2 - \mu_1)$ from the data given

If the prob. of obs. $(\mu_2 - \mu_1)$ is large → I accept null

but if the prob. of observing $(\mu_2 - \mu_1)$ is very small, that means we reject null thus accepting alternate hypo.

now, how to compute p-value :-

Permutation testing :-



and then shuffle. → then I randomly sample 50 points and divide the data again

the cal $\Delta_1 = \mu_1 - \mu_2$

resample and cal Δ_2

Δ_3

⋮

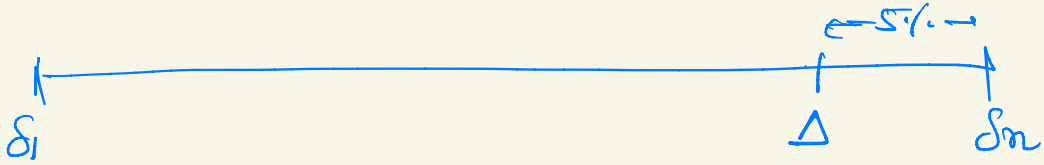
Δ_n

repeat the process till n times.

I'll sort my deltas

$\delta_1, \delta_2 \quad \cdots \quad \delta_n$

and now, my test statistic = $\Delta = \mu_2 - \mu_1$ (from the original dataset)



If only S.I. values are greater than Δ, then my p-value = 0.05

What is the significance of this method

So, combining the points means that we're assuming there is no diff. in heights

If actually there is no difference, then all these δ's should be small.

now, assuming that there is no diff. I got a distribution of deltas.

now, if my actual diff (test statistic) is at the extreme end, I can say that my null hypo is incorrect bcz data can't be wrong. so it means I'll reject null hypo.

and if my p-value is 0.05, then the confidence is or 5%.

$1 - p$

$= 95\%$.

I can say with 95% confidence that the heights of students in both the classes are different.