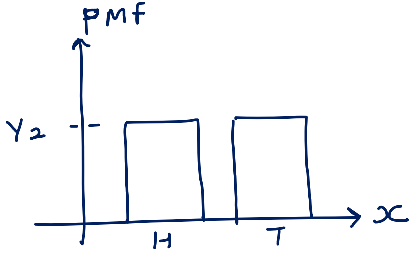
Statistics

1. PMF → Discrete Random variable
2. PDF → Continuous Random variable
3. CDF → cumulative D.F
4. UDF

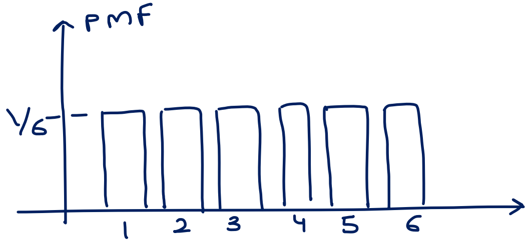
for ex: toss of a coin

$x = \{H, T\}$

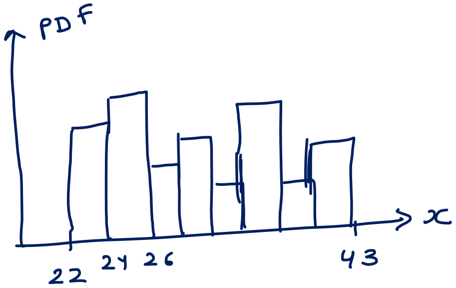


for ex: Roll of a dice

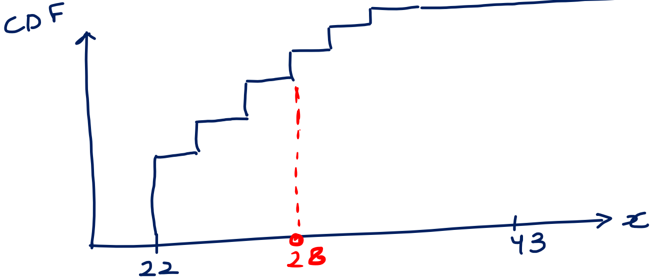
$x = \{1, 2, 3, 4, 5, 6\}$



Ex: Age of a random student in a class



=



$cdf = P(x \le x_i)$

what is the prob that a randomly selected student will have age at most 28 years $= P(x \le 28)$

Do we have any statistical test to confirm if a given random var. follows Normal Dist or not? ⟶ Yes, (next module)

Normal Distribution

normal dist = continuous Rand. var.

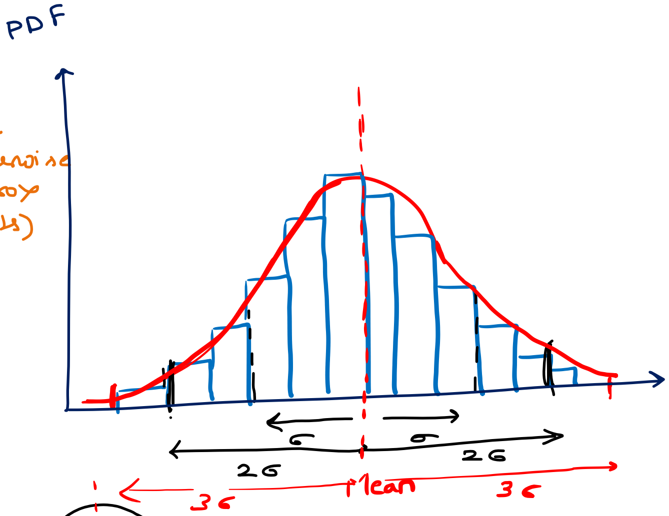
mathematical rep. for normal Dist

$\mathcal{N}$

$$\boxed{\mathcal{N}(\mu, (\sigma)^2)}$$

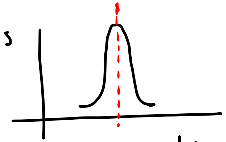
If a PDF represents a bell shape curve when plotted and is approx. symmetrical around mean, then that PDF is also known as the normal dist. or Gaussian Dist

The normal dist. comes with various proven properties that can be applied to any given random var. which appears to be normal and this will help us to better understand the nature of the random var.



properties (these will hold exactly true if PDF is 100% normal otherwise we should get approx results)

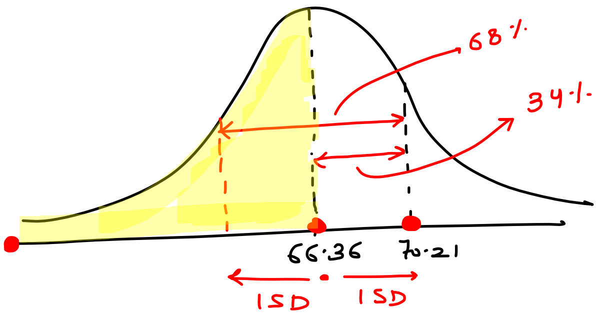
1. 68% of the observations b/w $\mu \pm 1SD$
2. 95% of the observations b/w $\mu \pm 2SD$
3. 99.7% of the obs b/w $\mu \pm 3SD$

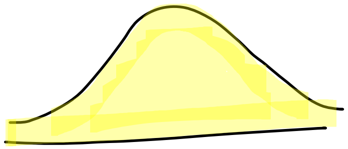


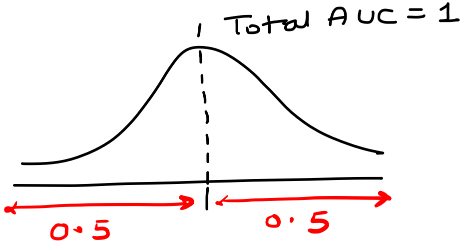
Low value of σ

High value of σ

$$\mathcal{N}(66.36, (3.84)^2)$$







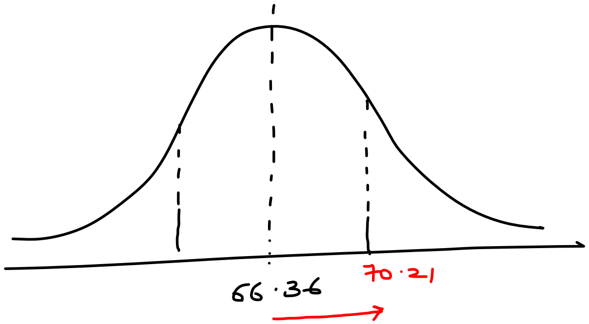
$$P(x \leq 66.36) = 0.5$$

$$x = \mu + 1\cdot\sigma$$
$$= 66.36 + 1 \times 3.84$$
$$= 70.21$$

$$P(x \leq 70.21) =$$
$$P(x \leq 66.36) + P(66.36 - 70.21)$$
$$= 0.5 + 0.34$$
$$= 0.84$$

$P(x \leq 70.21)$ = cdf of the given normal dist

$N(66.36, (3.84)^2)$



$$\{\, 70.21 = 66.36 + 1 \times (3.84) \,\}$$

$$x = \mu + 1(SD)$$

$$x = \mu + z(\sigma)$$

z = no. of std. dev. above or below mean

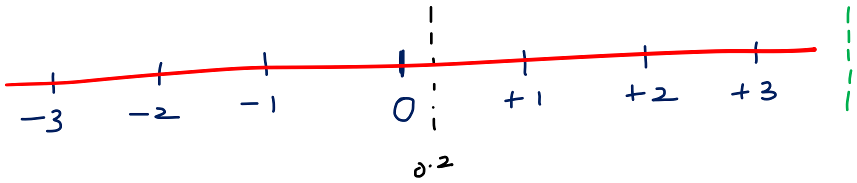
$$x - \mu = z(\sigma)$$

$$\boxed{z = \frac{x-\mu}{\sigma}}$$

Help us to find out how many std. deviation a value is above or below mean.

Range	z	P(x)
$\mu - z(\sigma),\ \mu + z(\sigma)$	1	0.68
$\mu - z(\sigma),\ \mu + z(\sigma)$	2	0.95
$\mu - z(\sigma),\ \mu + z(\sigma)$	3	0.997

Z-Scale



age is normally dist

$$z = \frac{(x - \mu)}{\sigma}$$

⇒ Random person

$\boxed{z = +3.5}$

① $x = \mu$ = $z = 0$

$z = 0.2$

Suppose The height of people is Gaussian with mean 65 inches and standard deviation 2.5 inches.

If I want to know 69.1 inches is how many standard deviation far away from mean?

$\mathcal{N}(65, (2.5)^2)$

$$z = \frac{(x-\mu)}{\sigma}$$

$$= \frac{69.1 - 65}{2.5}$$

$$\boxed{z = 1.64}$$

from scipy.stats import norm

norm.cdf(x = 69.1, loc = 65

scale = 2.5)

norm.cdf(1.64)

z

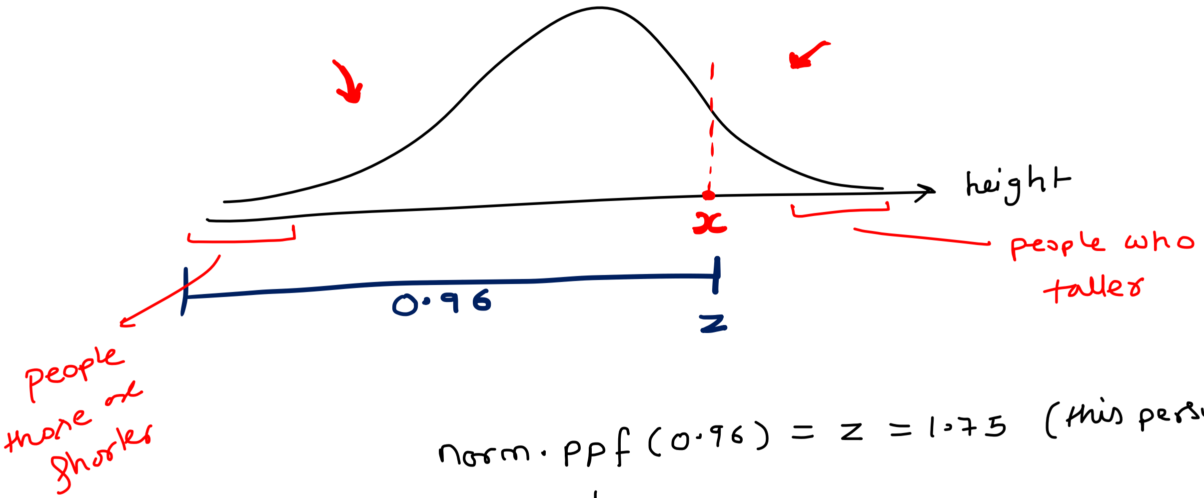
$P(x \leq 69.1) = 0.94$

what's the prob. that the height of a randomly selected person is at most 69.1

The height of people is Gaussian with mean 65 inches and standard deviation 2.5 inches.

One person says:
96% people are shorter than me. What is my height?

$$\mathcal{N}(65, (2.5)^2)$$



$\text{norm.ppf}(0.96) = z = 1.75$ (this person is $1.75\,\sigma$ above mean)

↓

percent point function

$$= \mu + 1.7(\sigma)$$
$$= 65 + 1.75(2.5)$$
$$= 69.37$$

if we want to find AUC given mean, and std. dev of the normal dist

1. norm.cdf(x, loc, scale)

2. $z = (x - \mu)/\sigma$

   norm.cdf(z)

if we want to find z given AUC

1. norm.ppf(AUC)

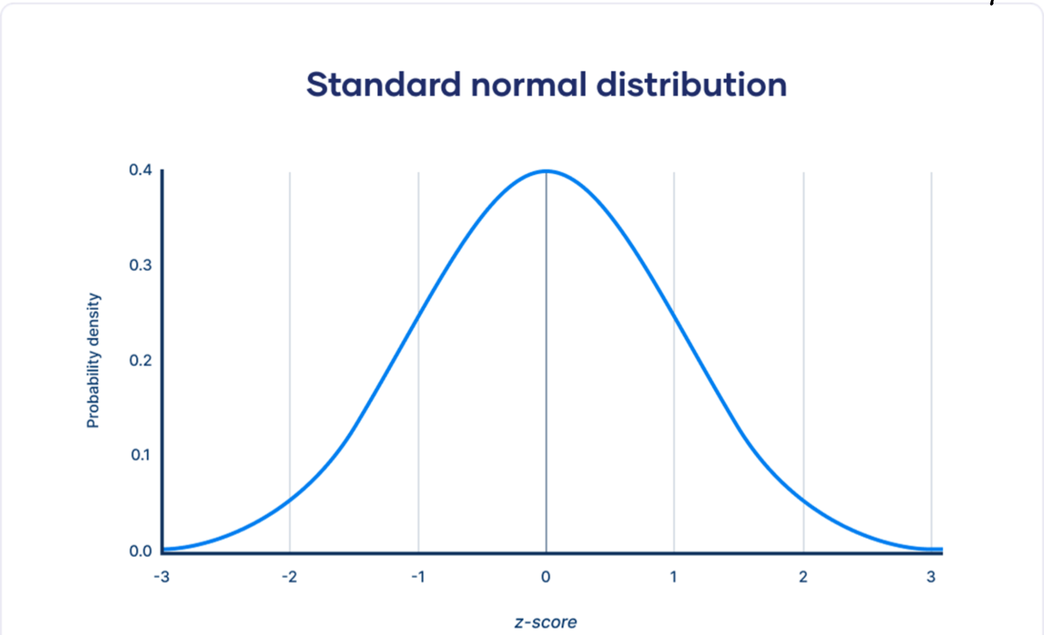
   $\downarrow$

   z-value

Standard Normal Distribution

$N(0,(1)^2)$

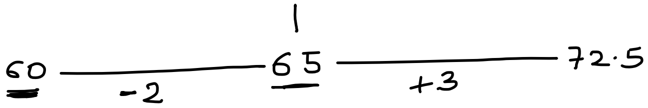
mean = 0
std.dev = 1

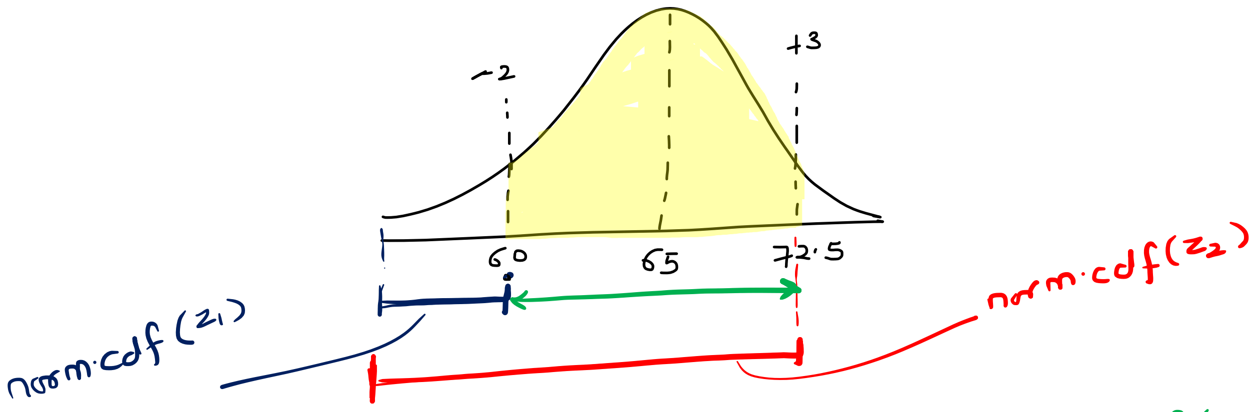


Quiz-1: The height of people is Gaussian with mean 65 inches and standard deviation 2.5 inches.
What is the fraction of people whose height is between 60 and 72.5?

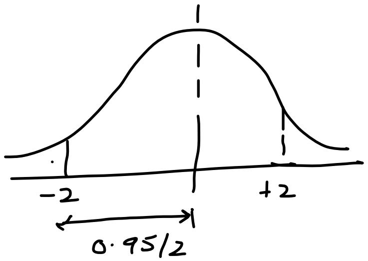
$65 - 2.5 = 62.5$

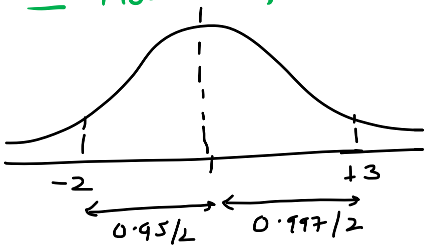
$62.5 - 2.5 = 60$





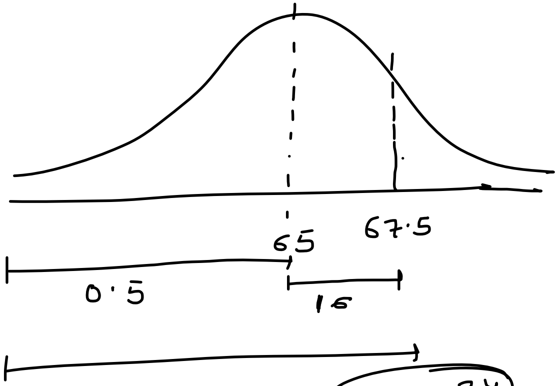
$= \text{norm.cdf}(z_1) - \text{norm.cdf}(z_2)$





$\frac{0.95}{2} + \frac{0.997}{2} =$

Quiz-2: The height of people is Gaussian with mean 65 inches and standard deviation 2.5 inches. What fraction of people are shorter than 67.5?



$0.5 + 0.34 = 0.84$

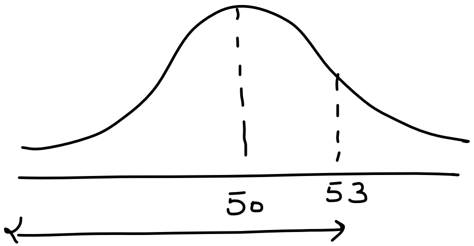
$\mu \pm 1SD = 68\%$

0.34

Quiz-3: Balls produced by manufacturer have mean diameter of 50 mm and std dev 2 mm. What fraction of balls have diameter smaller than 53 mm?

AUC

$\mathcal{N}(50, (2)^2)$

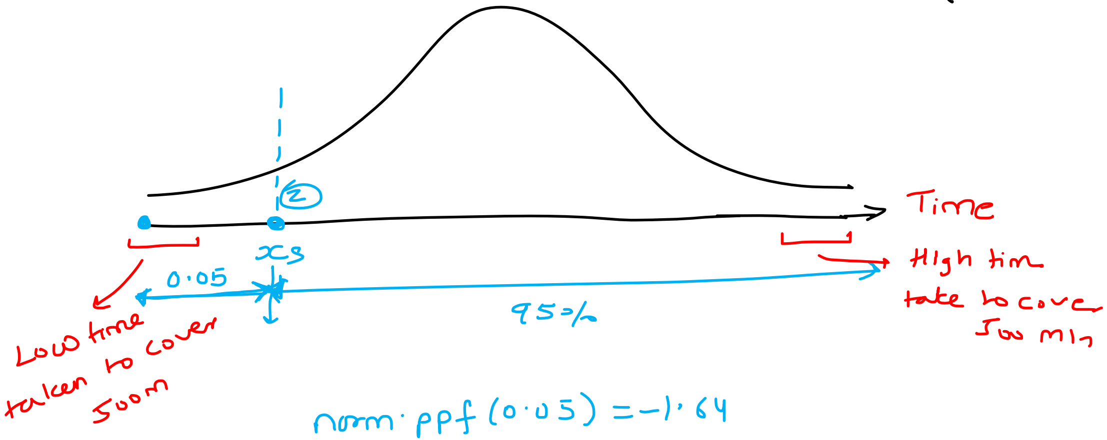


norm.cdf(53, loc = 50, scale = 2)

$\text{norm cdf}\left(\frac{53-50}{2}\right)$

Quiz-4: Skaters take a mean of 7.42 seconds and std dev of 0.34 seconds for 500 meters. What should his speed be such that he is faster than 95% of his competitors? [Microsoft]

$N(7.42, (0.34)^2)$



norm.ppf(0.05) = −1.64

$$\text{time} = \mu - 1.64(\sigma)$$

$$= 7.42 - 1.64(0.34)$$

$$= 6.86 \text{ sec}$$

$$\text{Speed} = \frac{D}{T}$$

$$= \frac{500}{6.86}$$

$$\boxed{= 72.8 \text{ m/s}}$$

Quiz-5: A retail outlet sells around 1000 toothpastes a week, with std dev = 200.

if we have 1300 stock units as our inventory then what fraction of weeks will we go out of stock?