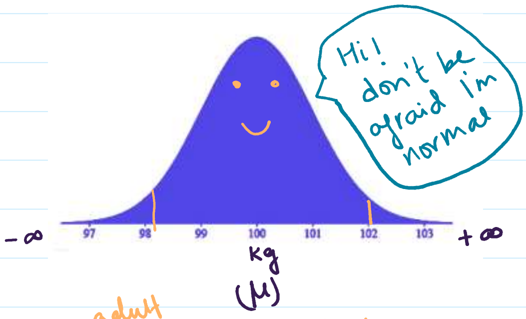
Normal Distribution 16oct22

14 July 2022 01:20 PM

intro → continuous r.v
→ probability distribution

Definition: It is a continuous distribution whose density function is of a bell curve and has 2 parameters namely mean (μ) and variance (σ^2)

Some important features:
1. Used in natural situations ✓
2. notation ✓
3. Symmetric distribution ✓
4. Shape fixed by parameters ✓
5. unimodal ✓
6. Pdf form ✓



When most of our data points are concentrated near the mean value and prob. of getting points farther away from the mean keeps reducing.

most adult giraffes should be in weight category 98 to 102, when mean is 100 kg & giraffes below 98 & above 102 would be very less in no.

[Pop^n: adult giraffes], weight.

pdf: $\boxed{\dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}}$

pop^n parameters: μ, σ^2

notation: data 1 $\sim N(\mu, \sigma^2)$ eg. $\boxed{X \sim N(100, 25)}$ weight of giraffe. 5^2

Range: range of normal distribution is always $(-\infty, \infty)$

If I integrate pdf I get prob. value.

$$\Rightarrow \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \underline{1}$$

$\because$ total prob. is added.

eg. to get P[getting a giraffe whose weight is in b/w 100 to 102] =

$$\int_{100}^{102} \frac{1}{5\sqrt{2\pi}}\, e^{-\frac{(x-100)^2}{50}}\, dx = \underline{0.1554}$$

\# since calculating this integral is a tedious task we will use a shortcut. → std. normal dist^n & it's table.

What is symmetricity? → m=m=m, mean = median = mode.

What is Skewness?

Skewness is a deviation from the symmetrical bell curve or normal distribution in a data set. The curve is said to be skewed if it is shifted to the left or right. Skewness can be measured as an indicator of how much a distribution deviates from the normal distribution.

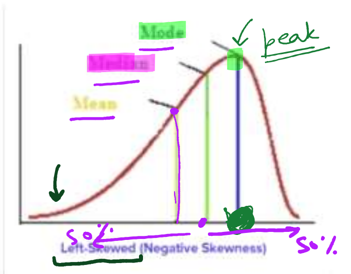
Non-symmetric-ness
Always measure from tail

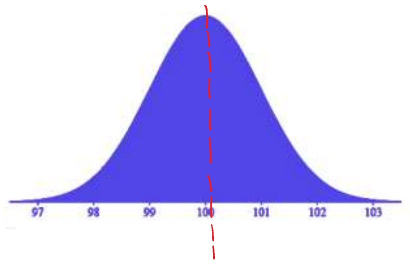
\# Normal distribution is ALWAYS symmetric.
to see skewness always look at tails

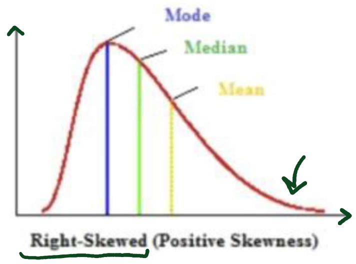
Normal distribution is ALWAYS symmetric.

to see skewness always look at tails

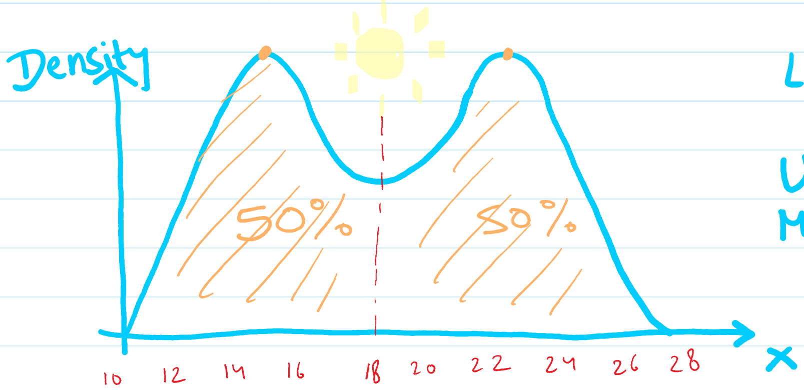
↓ here both tails equal







?



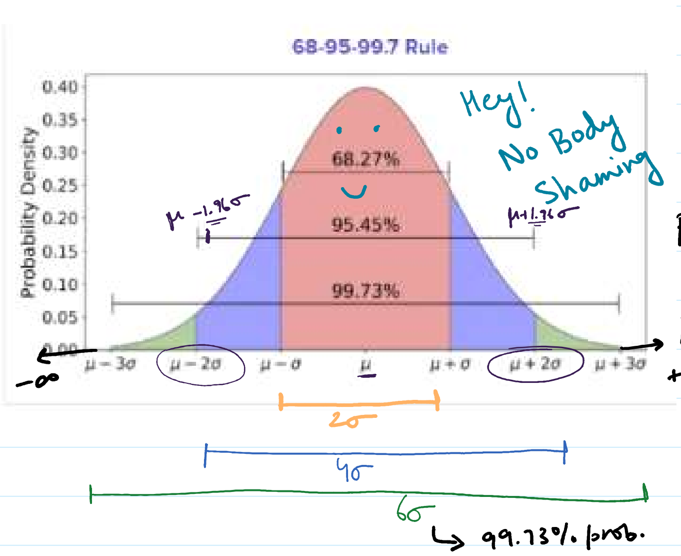
Left	Right	None
Uni	Bi	None

Mark → mean med mode

symmetric about 18

only exception in symmetric curve where mode ≠ (mean = median) where 2 modes exist.

68-95-99.7 Rule



$X \sim N(\mu, \sigma)$

$P[\mu-\sigma < X < \mu+\sigma] = 68.27\%$

$P[\mu-2\sigma < X < \mu+2\sigma] = 0.9545$

$P[\mu-3\sigma < X < \mu+3\sigma] = 0.9973$

Practice Problem

For a normal distribution, 95.5% of students at a school got marks in a test between 32 and 98. Calculate the mean and standard deviation. If the middle 95.5% of the population is considered

If 95.5% (middle) are covered left point is $\mu - 2\sigma$
right point is $\mu + 2\sigma$

ATQ $\quad \mu - 2\sigma = 32 \qquad \mu + 2\sigma = 98$

$4\sigma = 66 \Rightarrow \sigma = \frac{66}{4} = \frac{33}{2} = \boxed{16.5}$

$\mu = 32 + 2(16.5) = 33 + 32 = \boxed{65}$

Also μ is in the centre so centre of 32 & 98 also = 65 verified.

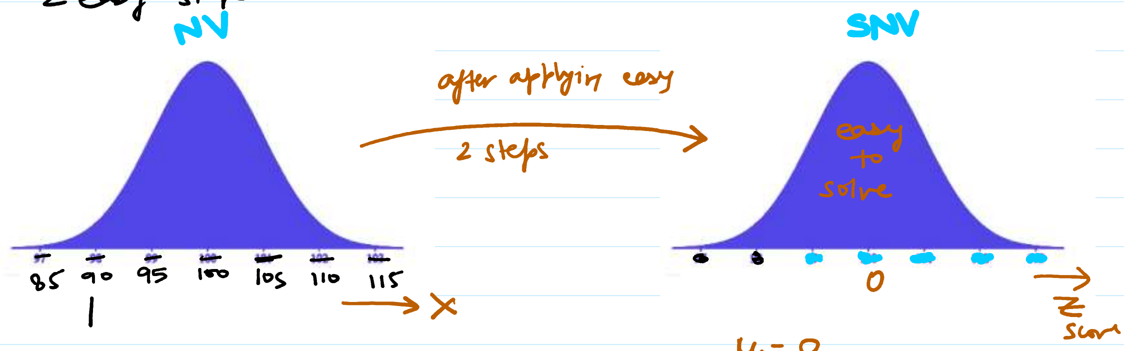
SHORTCUT

Standard Normal Distribution

The **Standard Normal Distribution**, also called the **z-distribution**, is a special normal distribution where the mean(μ) is 0 and the standard deviation(σ) is 1 and is denoted by **Z(0,1)**. or $N(0,1)$

A normal distribution where mean = 0 & std. dev = 1 is called standard normal distribution.

→ Every specific normal can be converted into std. normal in 2 easy steps.



$\mu = 100$

$\sigma = 5$

$\mu = 0$

$\sigma = 1$

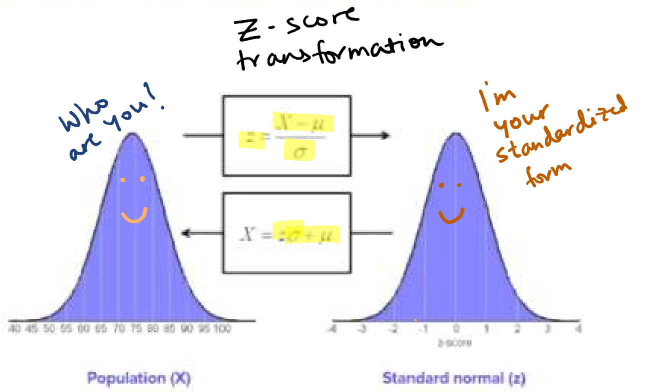
Step 1: subtract μ from x so centre is 0

step 1: $X - \mu$

Step 2: divide by originial std. dev. so that new std. dev = 1

step 2: $\boxed{\frac{X-\mu}{\sigma}}$ = z score

We can take any Normal Distribution and convert it to The Standard Normal Distribution.



Why to add this unnecessary complexity?

This complexity is necessary to avoid integrating PDF for each & every question ∵ we have constructed a z-score probability table to find answer in 10s.

Reading a Z-Score Table

https://www.statology.org/z-table/

If $y \sim N(60, 36)$ find $P[40 < Y < 50]$

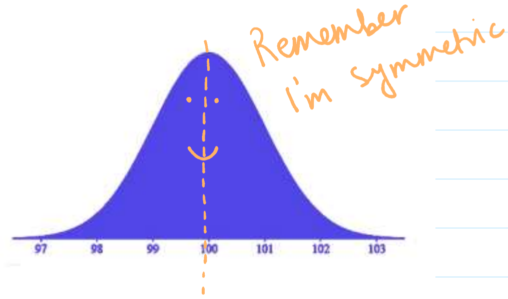
HARD WAY:

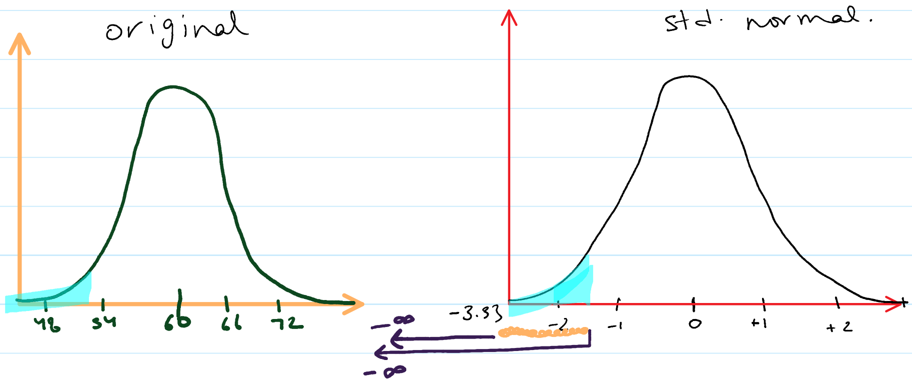
$$\int_{40}^{50} \frac{1}{6\sqrt{2\pi}} e^{-\frac{1}{72}(x-60)^2} dx$$

EASY WAY:

\# Shape remains intact in conversion to std. normal

$$P[40 < Y < 50]$$
$$= P[40-60 < Y-60 < 50-60]$$
$$= P\left[\frac{40-60}{6} < \frac{Y-60}{6} < \frac{50-60}{6}\right]$$
$$= P\left[\frac{-20}{6} < Z < -\frac{10}{6}\right]$$
$$= P[-3.33 < Z < -1.665]$$





$$P[-3.33 < Z < -1.67] = P[Z < -1.67] - P[Z < -3.33]$$

$$0.0475 - 0.0012$$
$$= 0.0463 = 4.63\%$$

Using Standard normal distribution to find the probability

Since the Standard Normal Distribution is a probability distribution, the probability that a variable will take on a range of values is indicated by the area under the curve between two points. The entire curve area under it is 1, or 100%.

Every z-score has a corresponding value that indicates the likelihood that all values below or above that z-score will occur. This is the region under the curve to the left or right of that z-score.

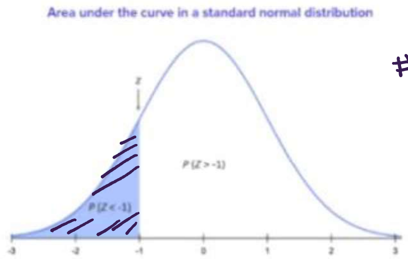


table gives value towards left only

Find area to the left right and between

Q1. If $X \sim N(\underset{\mu}{3}, \underset{\sigma^2}{16})$ find: (a) $P[X<1]$ (b) $P[X \geq 3.2]$
(c) $P[-2 < X < 2]$ (d) $P[-1 < X < 1]$

(a) $P[X<1]$

$$\Rightarrow P[X<1] = P\left[\frac{X-3}{4} < \frac{1-3}{4}\right] = P[Z < -0.5]$$

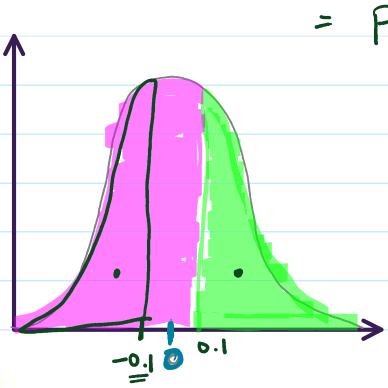
$$= 0.3085$$

(b) $P[X \geq 3.2] = P\left[\frac{X-3}{4} \geq \frac{3.2-3}{4}\right]$

$$= P[Z \geq 0.1]$$

$$= P[Z < -0.1]$$

$$= 0.4602$$



Problem - 1

The test results of students in a class have a mean μ of 70 and a standard deviation σ of 12, normally distributed. What percentage of students scored above 85?

Problem - 3

A particular species of dolphin's weight has a mean μ = 400 pounds and a standard deviation σ = 25 pounds, which is normally distributed. What proportion of dolphins weighs between 410 and 425 pounds on average?

RECAP…