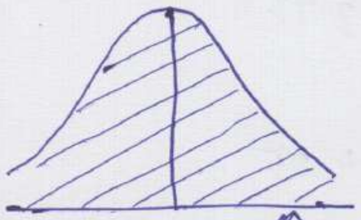
Day 3 - Statistics.

Topic Covers

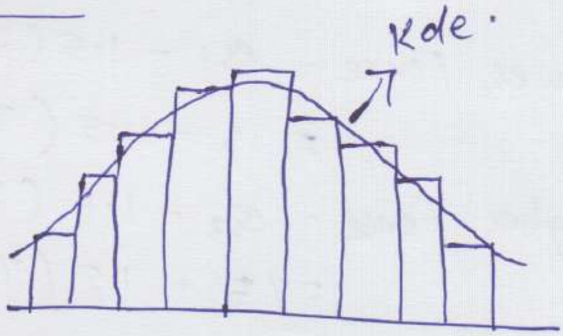
1. Normal Distributions
2. Standard Normal Distribution.
3. Z - score

★ Gaussian / Normal Distribution.



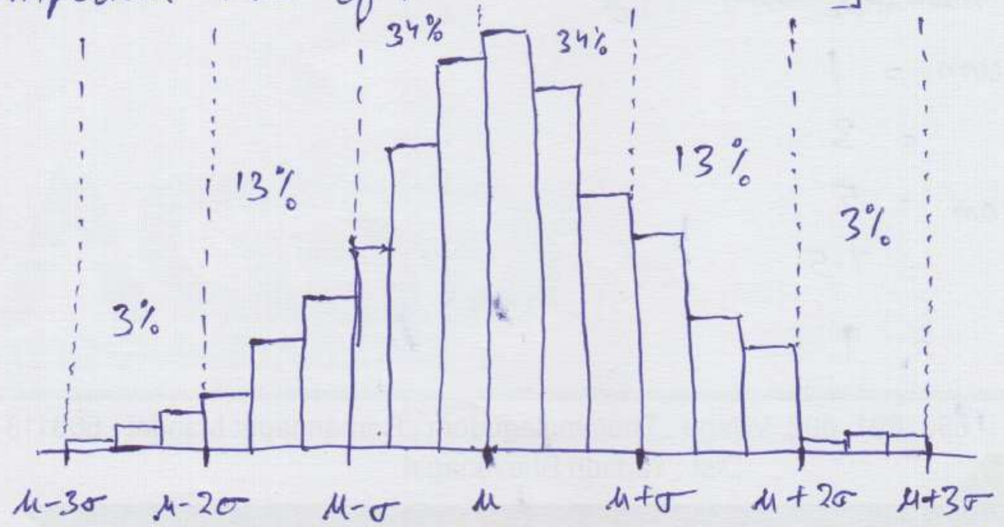
Distribution

Age, weight, height



Kde :- Kernal density estimator.

★ [Emperical Rule of Normal Distribution]



* Emperical Rule of Normal Distribution.

If the distribution is normal / Gaussian the

- 68% of data fall betwen $\mu - \sigma$ & $\mu + \sigma$
- 95% of data fall betn $\mu - 2\sigma$ & $\mu + 2\sigma$
- 99.7% of entire data is present betn $\mu - 3\sigma$ & $\mu + 3\sigma$

This rule is called $\boxed{68 - 95 - 99.7\%}$ => Emperical formula.

$\boxed{\text{Q-Q Plot}}$ → To Know wheather the distribution is Gaussian or Not ?

Standard Normal Distribution.

$$X \approx \text{Gaussian Distribution } (\mu, \sigma)$$
$$\Downarrow$$
$$y \approx \text{SND } (\mu = 0, \sigma = 1)$$

$$\text{Z-score} = \frac{x_i - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$\frac{\sigma}{\sqrt{n}}$ => Standard Error.

If n = 1

$$\boxed{\text{Z-score} = \frac{x_i - \mu}{\sigma}}$$

Example:-

$x = \{1, 2, 3, 4, 5\}$ Here $\mu = 3$ $\sigma = 1.414$.

$$y = \left\{ \frac{1-3}{1.414}, \frac{2-3}{1.414}, \frac{3-3}{1.414}, \frac{4-3}{1.414}, \frac{5-3}{1.414} \right\}$$

$y = \{-1.414, -0.707, 0, 0.707, 1.414\} \Rightarrow$ Here $\mu = 0$ $\sigma = 1$ means SND.

Q. Why we have to convert Gaussian distribution to standard normal distribution.

Age (years)	Weight (kg)	Height (cm)
24	72	150
26	78	160
32	84	165
33	92	170
34	87	150
28	83	180
29	80	175

Normalization :- In Normalization we try to normalize the value betn lower scale to higher scale.

(1) Min max Scaler [0-1]

$$\boxed{x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}}$$

x	y
1	0
2	0.25
3	0.5
4	0.75
5	1

$\frac{2-1}{5-1} = \frac{1}{4} = 0.25$ | $\frac{4-1}{5-1} = \frac{3}{4}$

$\frac{3-1}{5-1} = \frac{2}{4} = 0.5$ | $\frac{5-1}{5-1} = \frac{4}{4} = 1$

① **Standardization**:- $\quad$ $Z\text{-score} = \dfrac{x_i - \mu}{\sigma}$

$x \rightarrow$ Normal Distribution (μ, σ)

$\Downarrow$ Z-score

$y \rightarrow$ SND $(\mu = 0, \sigma = 1)$

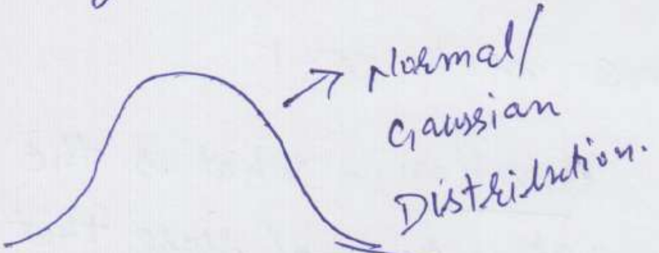
Why do we do Standardization → To Bring the feature in the same scale.

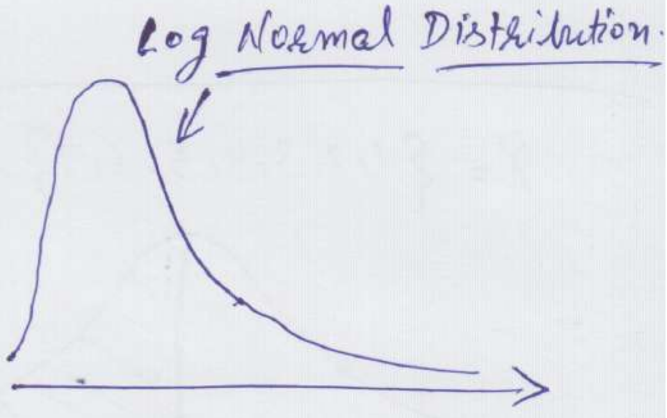
② Normalization → We can normalize the data in range say [0-1]

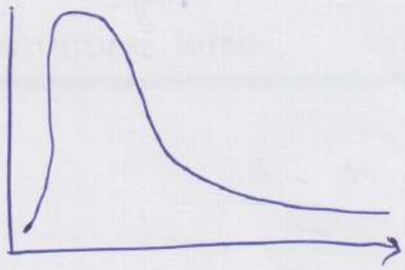
① Min Max Scaler →

---

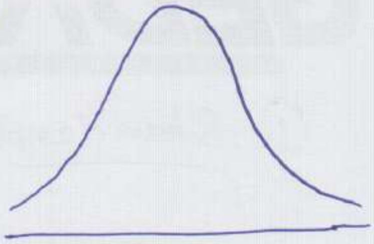
* **Log Normal Distribution.**







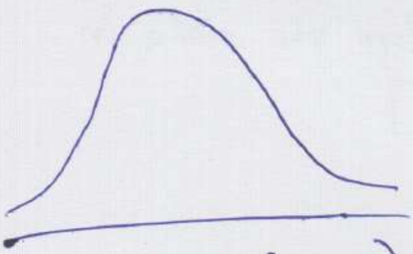
$y = \ln(x)$



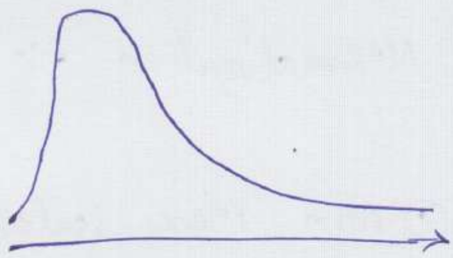
$x \approx$ Log Normal Distribution

Log normal Distribution can be converted to normal distribution by applying $\log_{10}(x)$.

Similarly normal distribution can be converted to log normal distribution by applying $\exp(x)$.

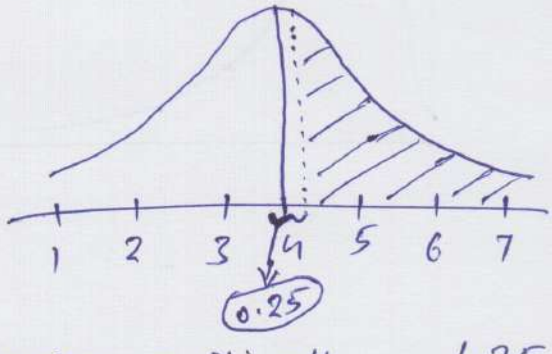


$\Rightarrow \exp(x)$
$\Downarrow$
y



$x \approx ND(\mu, \sigma)$

$X = \{1, 2, 3, 4, 5, 6, 7\}$ Assume $\mu = 4$ $\sigma = 1$



Question: - What is the percentage of score that falls above 4.25?

① Z-score $= \dfrac{x_i - \mu}{\sigma} = \dfrac{4.25 - 4}{1} = 0.25$

① Z-table (Area under the curve) :- This table is used to find the area which is under the curve.

Z - table :- Search on google For Z-table and go to first link. Z table. net.

Z-score = 0.25

Now According to Positive Z score table.

$+0.25 = 0.59871 \approx 0.59 = 50\%$

Total Area of normal distribution is 1

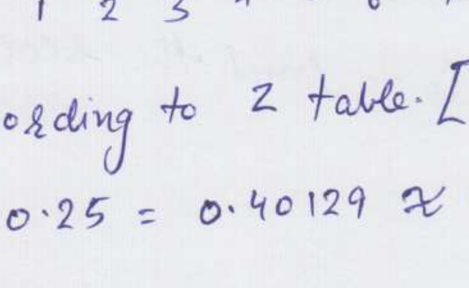
So the area of pattern is ~~1-0.59 = 41~~ $1 - 0.59 = 0.41 = 41\%$

Similar Question :- What is the percentage of score that falls below 3.75?

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

Assume $\mu = 4 \quad \sigma = 1$

we have to find this area.



$$Z\ score = \frac{x_i - \mu}{\sigma} = \frac{3.75 - 4}{1} = -0.25$$

According to Z table. [Negative Z table]

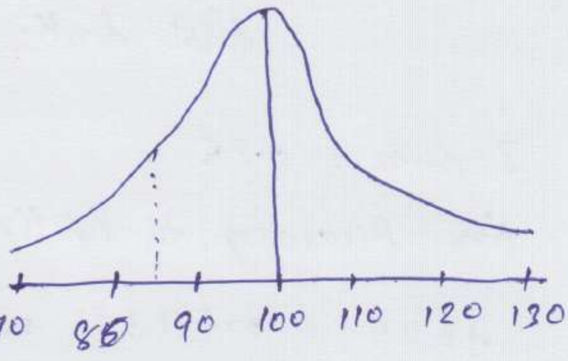
$-0.25 = 0.40129 \approx 0.40 = 40\%$

Question:- In India the average IQ is 100 with a standard Deviation of 15, what is the percentage of population would you expect to have an IQ ~~lower then~~

① Lower then 85

② Higher then 85

③ Between 85 and 100



---

① Lower then 85 $\therefore$ Zscore $= \frac{x_i - \mu}{\sigma} = \frac{85-100}{15} = \frac{-15}{15} = -1$

Here $\mu = 100$

$\sigma = 15$

According to Negative Zscore table

$-1 = 0.15866 = 15\%$

② Higher then 85 $\therefore$ Zscore $= \frac{85-100}{15} = -1$

Acc. to Negative Zscore table $-1 = 0.15866$

Total Area of distribution is 1, So to find the score higher then 85 we have to.

$1 - 0.15866 = 0.84134$

③ Between 85 and 100.

First Find Zscore for area lower then 100

Zscore $= \frac{100-100}{15} = \frac{0}{15} = 0$

According to positive Zscore table $+0 = 0.5$

Between 85 and 100 $= 0.5 - 0.15866 = 0.34134$.