Advance Statistics : 02

Topic :

1. Uniform Distribution
2. Z stats and Z table
3. Central Limit Theorem

Uniform Distribution:

→ ① Continuous Uniform Distribution (PDF)

→ ② Discrete Uniform Distribution (PMF)

① Continuous Uniform Distribution: (Also called rectangular Distribution)

The distribution describes an experiment where there is an arbitary outcome that lies between certain bounds.

The bounds are defined by parameters a and b which are the minimum and maximum value.

Notation: $U(a,b)$

Parameters:

→ $0 < a < b < \infty$

Pdf:

$$\begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

CDF:

$$\begin{cases} 0 & \text{for } x<a \\ \frac{x-a}{b-a} & \text{for } x\in[a,b] \\ 1 & \text{for } x>b \end{cases}$$

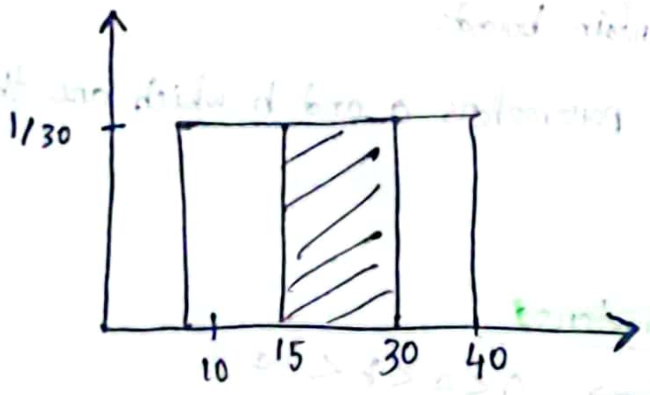
Mean: $\frac{1}{2}(a+b)$ **Median:** $\frac{1}{2}(a+b)$

Variance: $\frac{1}{2}(b-a)^2$ **Std:** $\sqrt{\frac{1}{2}(b-a)^2}$

Example: The number of candy sold daily at a shop is uniformly distributed with a maximum of 40 and a minimum of 10

1) What is the probability of daily sales to fall between 15 and 30?

Ans:



$$Pr(15 \le x \le 30) = (x_2 - x_1) \times \frac{1}{b-a}$$

$$= (30-15)\,\frac{1}{\cancel{30}\,40-10}$$

$$= 0.5$$

② Discrete Uniform Distribution:

It is a symmetric probability distribution where in a finite number of values are equally likely to be observed. Every one of n values has equal probability $1/n$. Here, $n = (b-a+1)$

| $a = min$
| $b = max$

Example: Rolling a dice $\{1, 2, 3, 4, 5, 6\}$

$a = 1$ (min) $\quad Pr(1) = \frac{1}{6} \quad Pr(4) = \frac{1}{6}$

$b = 6$ (max) $\quad Pr(2) = \frac{1}{6} \quad Pr(5) = \frac{1}{6}$

$Pr(3) = \frac{1}{6} \quad Pr(6) = \frac{1}{6}$

Notation: $U(a,b)$

Parameters: a, b with $b \geq a$.

PMF: $1/n$

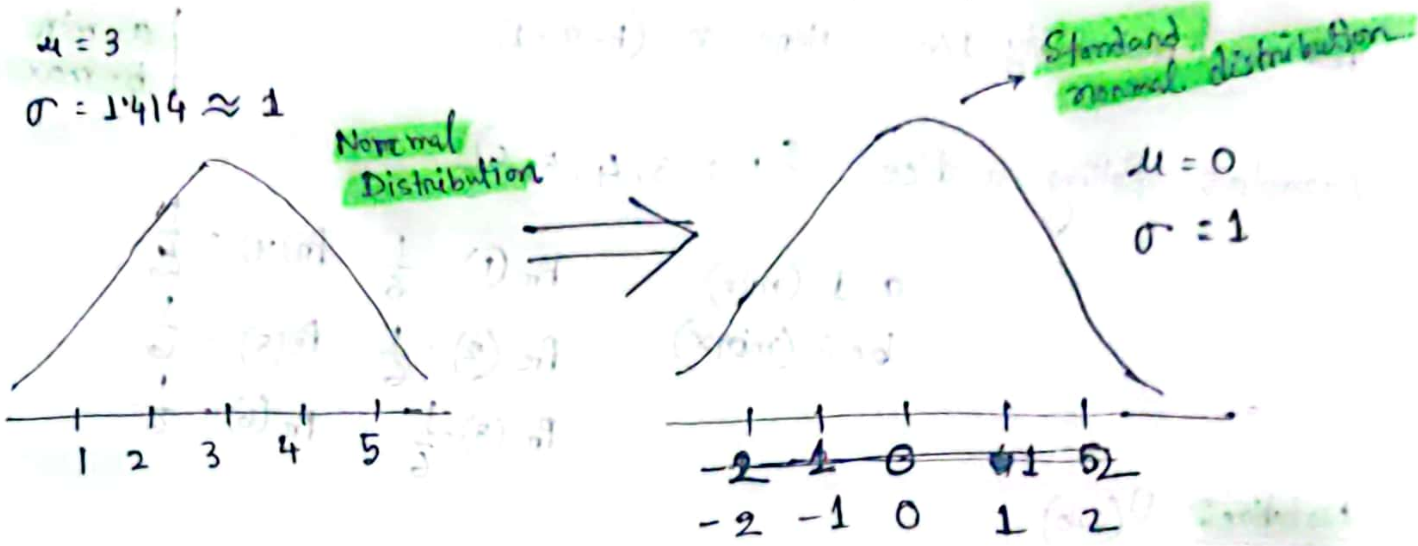
Mean, Median: $\frac{a+b}{2}$ $\quad$ Variance $\frac{1}{2}(b-a)^2$

Standard Normal Distribution and Z-score: (Z-stats)

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

$\mu = 3$

$\sigma = 1.414 \approx 1$



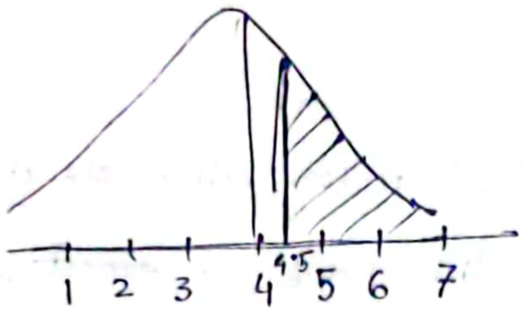
To convert Normal distribution to Standard normal distribution, we need Z-score formula.

$$Z\text{-score} = \frac{x_i - \mu}{\sigma} \qquad [x_i = x = \{1, 2, 3, 4, 5\}]$$

$$x = \{1, 2, 3, 4, 5\} \Rightarrow Z\text{-score} \Rightarrow \{-2, -1, 0, 1, 2\}$$

In short, z-score defines what std ~~difference~~ distance a random variable is from a mean value

Example :



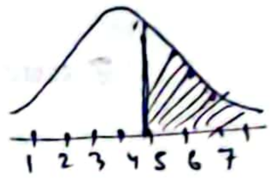
$\mu = 4$
$\sigma = 1$

question: How many std 4·5 is away from mean? $\longrightarrow$ $x_i = 4{\cdot}5$

$$Z\text{-Score} = \frac{x_i - \mu}{\sigma} = \frac{4{\cdot}5 - 4}{1} = 0{\cdot}5$$

Ans: 0·5 std away

question: What percentage of data is falling above 4·5?



Z-score = 0·5 $\longrightarrow$ for 0·5 area from Z-table = 0·69146

Area under the curve ($\geq$ 4·5) = ~~0·69146 [found from Z table sheet]~~

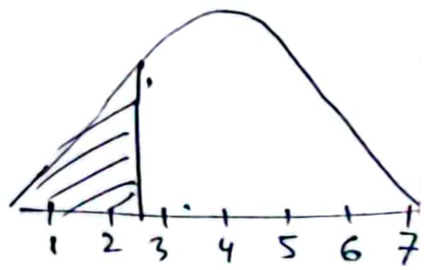
1 − 0·69146

= 0·30854

= 30·85 %

[For positive Z score Area would be 1 − Ztable value]

question: What percentage of data is falling below 2·5



$$Z\text{-score} = \frac{x_i - \mu}{\sigma} = \frac{2{\cdot}5 - 4}{1} = -1{\cdot}5$$

for −1·5, z table value is 0·06681

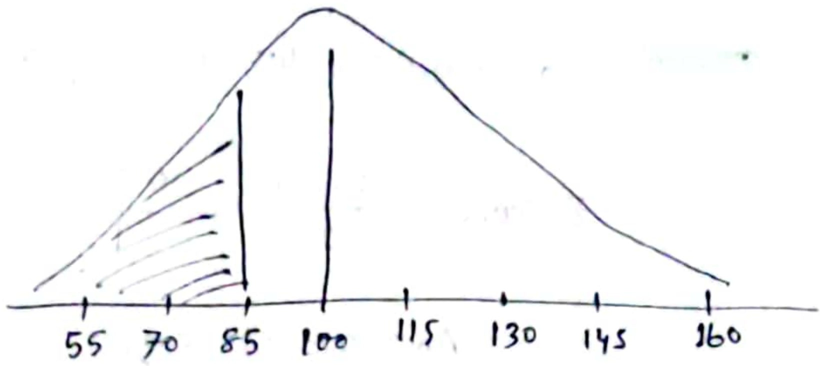
∴ Area under curve ($\leq$ 2·5) = 0·06681

= 6·6%

[For negative Z score Area would be just Ztable value]

Problem:

In India, the average IQ is 100, with a std of 15. What is the pecentage of the population would you expect to have an iq lower than 85?



Here, $\mu = 100$

$\sigma = 15$

$x_i = 85$

$$\text{Z-score} = \frac{\cancel{100}\,85-100}{15}$$

$$= \frac{-15}{15} = -1$$ (random variable is -1 std away from mean)

Area under curve according to z-table where Zscore = -1

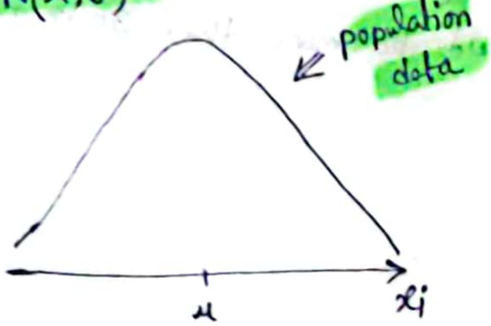
= 0.15866

= 15.866 %

Central Limit Theorem:

The sample distribution of the mean will always be normally distributed as long as the sample size is large enough. Regardless of whether the population has a normal, binomial or any other distribution, the sampling distribution of the mean will be normal.

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



If we take samples from population data.

$S_1 = \{x_1, x_2, x_3, \ldots, x_{20}\} = \bar{x}_1$

$S_2 = \{x_4, x_2, x_3, \ldots, x_{20}\} = \bar{x}_2$

$\vdots$

$S_m = \{\ldots\ldots\} \quad \bar{x}_m$

$\bar{X} = \{\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4 \ldots \bar{x}_m\}$ [$n \geq 30$ for each sample]

If you have a population data which is normally distributed and you take several samples from that and then calculates the sample mean from that, ~~the plot~~ and plot that data would also be normally distributed.

$\bar{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ [n can be any value]

The greate the better

CamScanner

if population

② ~~X~~ not belongs to Normal distribution

means ~~you~~ population data is not normally distributed, in that case if you take several samples from that and then calculates the sample mean from them, that would be normally distributed But you have to make sure while taking each sample, n should be >= 30 [mean each sample data should be more than 30 in number]