Probability

18 May 2024 12:08

What is probability?
Calculating about the event and it percentage of outcomes

Ex: 70% chances are there to win the game

Ex: Throwing a coin

Practice, skills, experience, projects, assignments, dedication, communication, resume -> 95%

X	---> P(X)
Head	1/2
Tail	1/2

→ probability distribution table

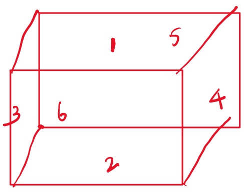
Sum = 1

Why should ?

To take the right decisions?

X ----> random variable

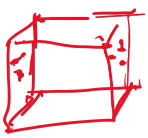
Throwing a unbiased dice



X	P(x)
1 →	1/6
2 →	1/6
3 →	1/6
4 →	1/6
5 →	1/6
6 →	1/6
	1

Throwing a unfair die

X : {1, 3, 4,5, 6}



X	P(X)
1 →	1/6
3 →	2/6
4 →	1/6
5 →	1/6
6 →	1/6

→ $P[x=3] = 2/6$

$6 \rightarrow 1/6$

1

$\rightarrow P[x=3] = \underline{2/6}$

$P[x \neq 3] = \frac{4}{6}$

Throwing a two dice at a time

[1,1] [1,2] [1,3] [1,4] [1,5] [1,6]

[2,1] [2,6]

[3,1] [3,6]

[4,1] [4,6]

[5,1] [5,6]

[6,1] [6,6]

Sum (1,1) = 2

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

X		P(x)
2	—	1/36
3	→	2/36
4	→	3/36
5	→	4/36
6	→	5/36
7	→	6/36 ✓
8	→	5/36 ✓
9	→	4/36
10	→	3/36
11	→	2/36
12	→	1/36

We have a fixed formulae to calculate only probabilities for above problem

$$\boxed{P(x) = \frac{6 - \sqrt{(7-x)^2}}{36}}$$

2 12

$$P(x=8) = \frac{6 - \sqrt{(7-8)^2}}{36}$$

$$= \underline{\underline{5/36}}$$

Random Variable

Continous	Discrete
Normal	Bernouli
Exponential	Binomial
Weibul	Poisson
Gamma	Geometirc
Logistic	Hyper Geometric
Log Normal	Negative Binomial

function

probability density function [pdf]

probability mass function [pmf]

==

Bernoulli Random variable:

Any experiments when we do, if the outcomes exists with only two possible outcomes

Yes / No, True /False, Head / False,

Yes ----> Success (p), No --> Failure (q)

P + q = 100% -----> entire probability will falls under 0 to 1 only.

If we do repeatedly the same experiment for "n" number of times , then those X Random variable is following
Binomial distribution:

pmf for B.D

$$\boxed{P[x=r] = {}^{n}C_{r} \cdot p^{r} \cdot q^{n-r}}$$

$$ {}^{n}C_{r} = \frac{n!}{[n-r]! \cdot r!} $$

$B.D \sim (n, p)$

n = # of times

p = % of success

$q = 1 - p$

$r \Rightarrow$ Any event

==

5 People are applied for home loan to a Bank in a Month. As per previous records only 60% are eligible for issuing home loan because of bank rules.

What is the probability of issuing home loan exactly for 3 people?

X -->	P(X)
0	—
1	—
2	—
3	→
4	—
5	—

1

$p = 60\% \rightarrow 0.6, \quad q = 0.4$

$n = 5$

$$P[x=3] = {}^5C_3 \times [0.6]^3 \times [0.4]^2$$

$$\Rightarrow \frac{5!}{[5-3]! \times 3!} \times (0.6)^3 \times (0.4)^2$$

$$P[x=3] = 34.6\%$$

There is 34.6% of chance that exactly 3 people will get the approved of home loan 5 people are applied with 60% of success in past.

What is the probability of exactly for 2 people?

1) What is the probability of issuing home loan at most for 2 people?

$$P[x \leq 2] = P[x=0] + P[x=1] + P[x=2]$$

cdf ⇒ Cumulative distri. function

2) What is the probability of issuing home loan **atleast** for 3 people?

$$P[x \geq 3] = P[x=3] + P[x=4] + P[x=5]$$

cdf ⇒ 0 → starts

Sum of all probabilities = 1

Let's say that 80% of all business startups in the IT industry report that they generate a profit in their first year. If a sample of 10 new IT business startups is selected, find the probability that exactly seven will generate a profit in their first year.

$n = 10$ $\qquad$ $P[X = 7] = 20.1\%$

$p = 0.8$

$r = 7$

Example 2:

Suppose that 70% of adults who are with corona positive reports giving the feedback that they got a relief with a specific medication.

a. **If the same medication given to another 250 patients, what is the probability the medication will effect at least 160 patients**

$n = 250$ $\qquad$ $P[X \geq 160] = \underline{?}$

$p = 0.7$

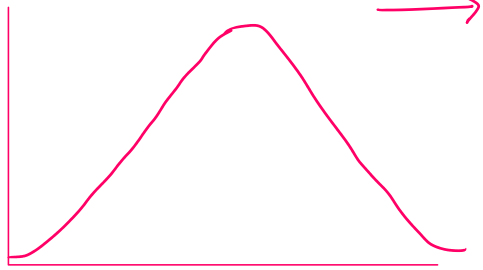
$r = 160$

==

Normal distribution / Gaussian Distribution :

If x is random variable follow continuous data type --> we can assume that is it follows N.D ?

Construct the histogram



→ Normal distribution

Continuous R.V follows -inf to + inf

Collect the height of the people

~~165~~ $\qquad$ **P(x = 170) --> 170.00000009 --> very less --> no usage**
168 $\qquad$ **P(X>= 170)**
170 $\qquad$ **P(X <=170)**
166 $\qquad$ **P(160< x < 174) -->between**

168 P(X>= 170)
170 P(X <=170)
166 P(160< x < 174) -->between
169
174
172 n -
. σ -
.
.
174

$N D \sim [\mu, \sigma]$

Formula

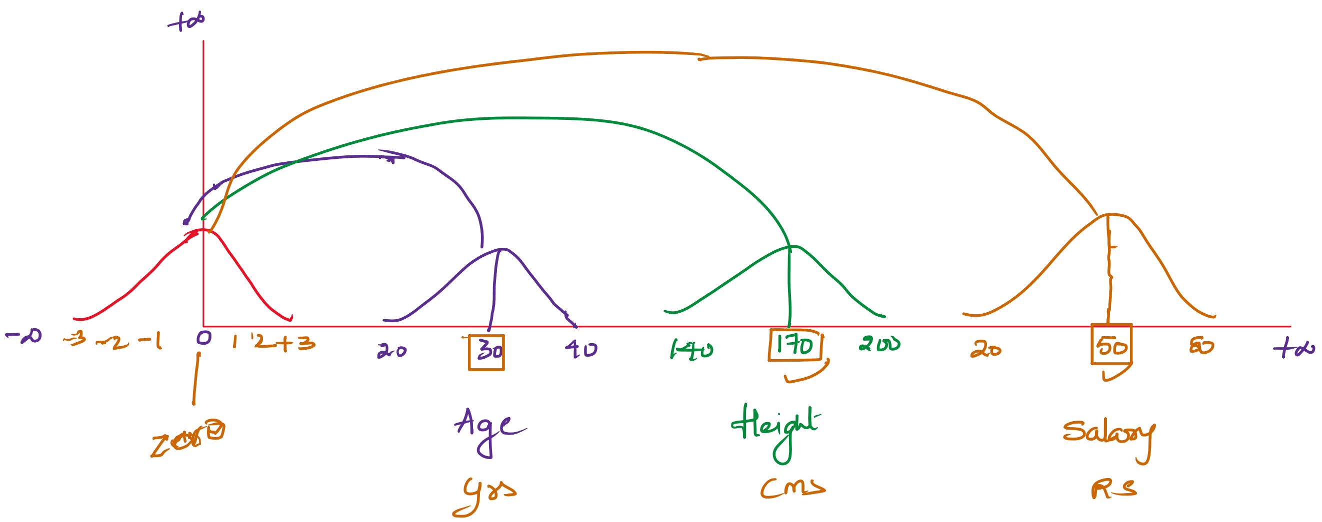
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$f(x)$ = probability density function

σ = standard deviation

μ = mean

Screen clipping taken: 19-05-2024 10:46



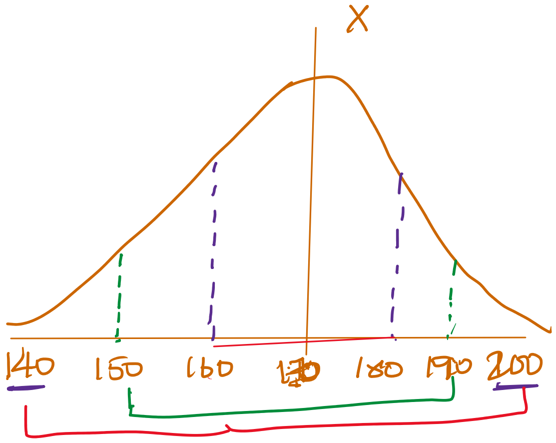
Standardization:

$$Z = \frac{X - \mu}{Sigma}$$

After using the formulae It will transforms its originality to somewhere in between -3 to + 3 sigma

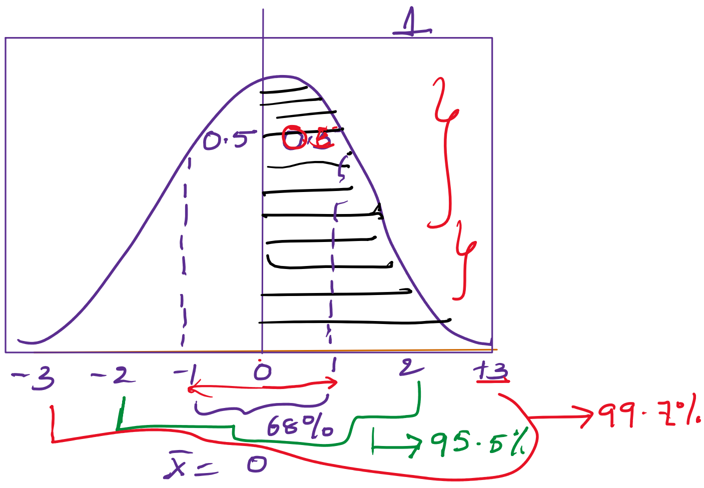
We are removing the units from the data such that it becomes a normal numbers

If we put our data in between -3 to + 3 , we have already calculated the entire probabilities of those range and those values are saved in the form a statistical table book . When we use python, our programming language will auto calculates from the tables book and directly gives the estimates of given atleast, atmost, between cases.



$\bar{x} = 170$

$SD = 10$



$\bar{x} = 0$

$SD = 1$

$\bar{x} = 170, \quad SD = 10$

$\bar{x} = 0 \quad SD = 1$

$P[X > 170] = \, ? \Rightarrow 50\%$

$P[x < 170] = 50\%$

$P[160 < x < 180] = \, ? \Rightarrow 68\%$

$P[150 < x < 190] = \Rightarrow 95.5\%$

$P[140 < x < 200] = \Rightarrow 99.7\%$

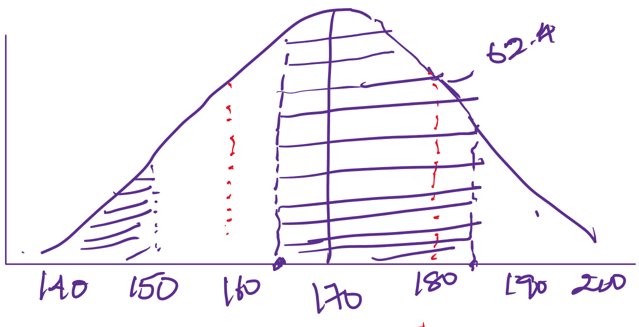
It is approximately estimated that 68% of data lies in between - 1 sigma to + 1 sigma limits under the normal distribution

It is approximately estimated that 95.5% of data lies in between - 2 sigma to + 2 sigma limits under the normal distribution

It is approximately estimated that 99.7% of data lies in between - 3 sigma to + 3 sigma limits under the normal distribution

==

	A	B	C	D	E	F	G	H	I
	id	age	height	salary			Z_age	Z_height	Z_salary
	1	24	170	25000			-0.89903	0.4159	-0.9983
	2	28	160	35000			-0.25686	-0.62385	-0.52292
	3	40	180	40000			1.669619	1.455651	-0.28523
	4	26	165	80000			-0.57795	-0.10398	1.616301
	5	30	155	50000			0.064216	-1.14373	0.190153
	mean	29.6	166	46000			0	0	0
	sd	6.228965	9.617692	21035.68			1	1	1

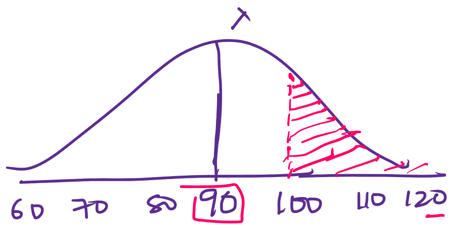


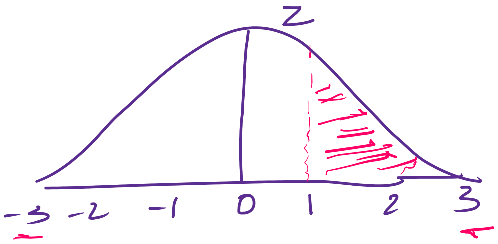
P[x < 150] = ____

P[165 < x < 185] = ____

b. A radar unit is used to measure speeds of cars on a highway. The speeds are normally distributed with a mean of 90 km/hr and a standard deviation of 10 km/hr. What is the probability that a car picked at random is travelling at more than 100 km/hr?

80
100
120
90
65
75
95





P[x ≥ 100] =

σ = 10 → 99.7% → [60 — 120]

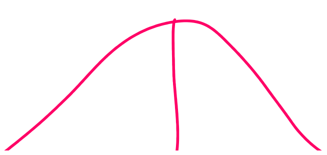
There is 15.9% still chances are there that any vehicles are travelling more than 100

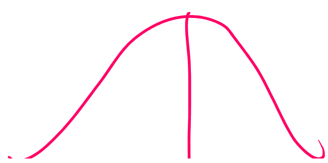
He collected the data, mean = 90, sd = 5

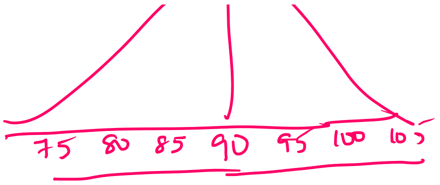
σ = 5

99.7% of the vehicles are travelling in between what range? → [75 — 105] kms

−3σ to +3σ   −3x5 to +3x5 ⟹ −15 to +15 from mean







$P[X > 100] = \quad \mu = 90, SD = 5$

There is 2.27% still chances are there that any vehicles are travelling more than 100

==

c. For a certain type of mobiles, the length of time between charges of the battery is normally distributed with a mean of 10 hours and a standard deviation of 20 minutes. John owns one of these mobiles and wants to know the probability that the length of time will be between 10 and 11 hours.

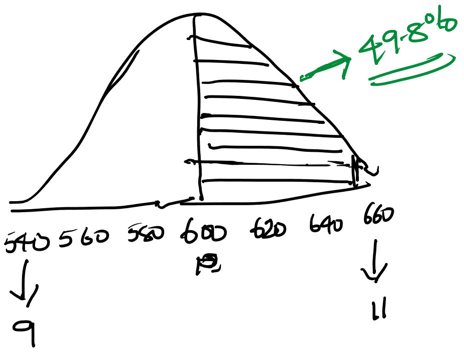
8
9
10
11
-.3
7.5
10 hrs

$SD = 20 min$

$\left.\begin{array}{l}\bar{x} = 10 \text{ hrs} \\ \sigma = 20 \text{ min}\end{array}\right\} \longrightarrow 99.7\% \; [9, 11]$

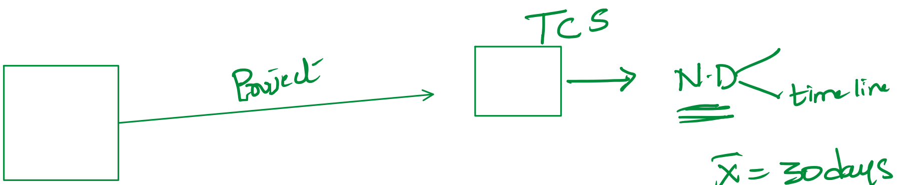
$\bar{x} = 600 \text{ min}$

$\sigma = 20$



==

Normal distribution itself a process that most of the industries will follow in delivering the products the clients whenever they are doing business



$\bar{x} = 30 \text{days}, \; SD = 1.$

$\Rightarrow 99.7\% \rightarrow [27 - 33]$

Any continuous variables , is it zero s.d possible?

Warranty period

Level of significance: alpha =5% --> standard value.

Test of hypothesis:

==