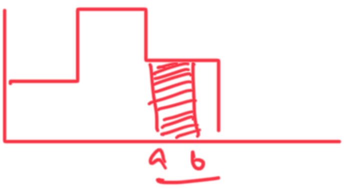
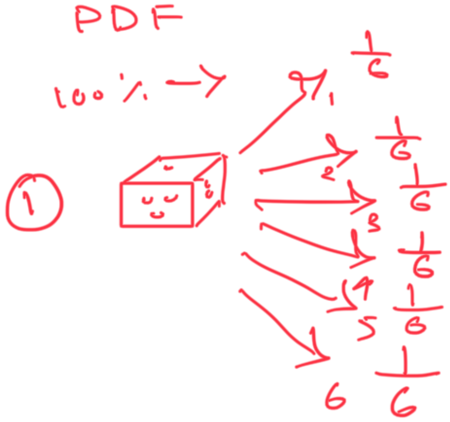
Statistics

data → continuous ⇓ PDF



discrete ⇓ PDF

100% →



(i) can't be negative

(ii) overall area has to be 1

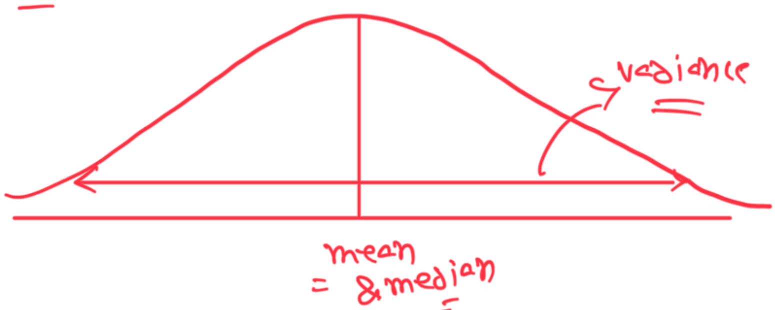
Next??

Normal

→ most ~~any~~ real life data

2 ⇒ mean / median

variance



$X \rightarrow N(\mu, \sigma^2)$

$\mu \Rightarrow$ mean

$\sigma^2 \Rightarrow$ variance

For Normal distribution we only have continuous data

mode isn't applicable

$X + c$ → what changes mean will change

cX → what change

$$\boxed{\begin{aligned} X &\to N(\mu, \sigma^2) \\ X+c &\to N(\mu+c, \sigma^2) \\ cX &\to N(c\mu, c^2\sigma^2) \end{aligned}}$$

cool??

Standard Normal

$N(\mu, \sigma^2)$

μ → anything σ^2 → anything

$[1, 1.2, 1.3, \ ]$ $(1, 1.2, \ldots \ )$

Should we study all these Infinite Normal Distributions??

All possible normal Distribution

$N(\mu, \sigma^2) \longrightarrow$ one specific normal distribution

$Z \to N(0,1)$

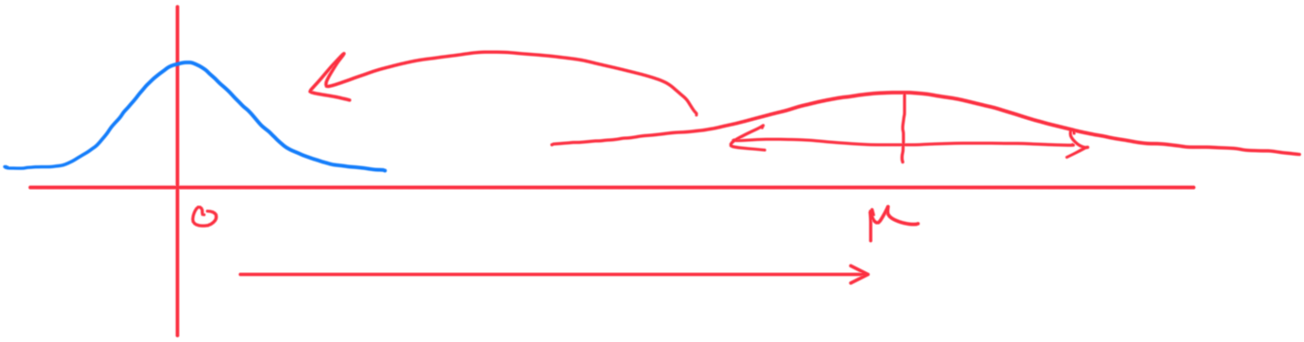
2-3

$X \to N(\mu, \sigma^2)$

$\frac{X-\mu}{\sigma} \to N(0,1)$

Z

$X \to N(\mu, \sigma^2) \to N(0, \sigma^2) \to \boxed{N(0,1)}$



$X \longrightarrow X-\mu \longrightarrow$

$N(0, \sigma^2)$ $\qquad N(0,1)$

$X-\mu \to N(0, \sigma^2) \to N(0, 1)$

$\frac{X-\mu}{\sigma} \Rightarrow ??$

1 2 3

mean → 2

var → 1

$1 \rightarrow \frac{1 - 2}{1} = -1$

$2 \rightarrow \frac{2 - 2}{1} = 0$

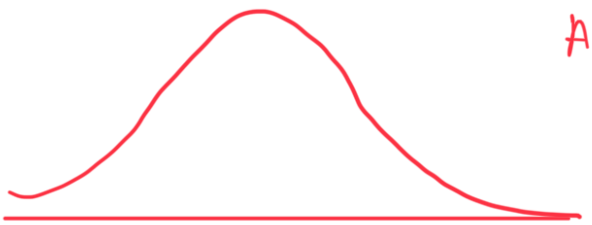
$3 \rightarrow \frac{3 - 2}{1} = 1$

1 2 3 → $\boxed{-1,\ 0,\ 1}$

mean = 0   1

$\boxed{X \rightarrow N(\mu, \sigma^2) \qquad \frac{X - \mu}{\sigma} \left(\frac{\text{data} - \text{mean}}{\text{std}}\right) \rightarrow \text{Std Normal}}$ → Table

Any dist. → Std Nor



IT

Salary of IT → $N(60k, \sigma^2)$

$\sigma = 10k$

Table

$$\boxed{P(x > 90k)} \quad 0$$

$$P(z > 3)$$

$$P\left(\frac{X-\mu}{\sigma} > \frac{90k-60k}{10}\right)$$

$$P(z > 3)$$

$x \rightarrow$ Uber eats $\rightarrow N(30m, \sigma^2)$

$\sigma = 10m$

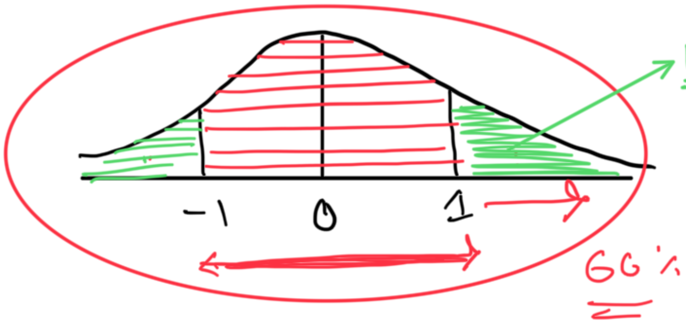
$$\boxed{P(x > 45)} \quad ??$$

$$P\left(\frac{x-\mu}{\sigma} > \frac{45-30}{10}\right)$$

$$P(z > 1.5)$$

Normal distribution

1. continuous distribution
2. μ, σ^2
3. $\frac{x-\mu}{\sigma} \rightarrow z$ (std)
4. some specific properties of std Normal distribution



17%

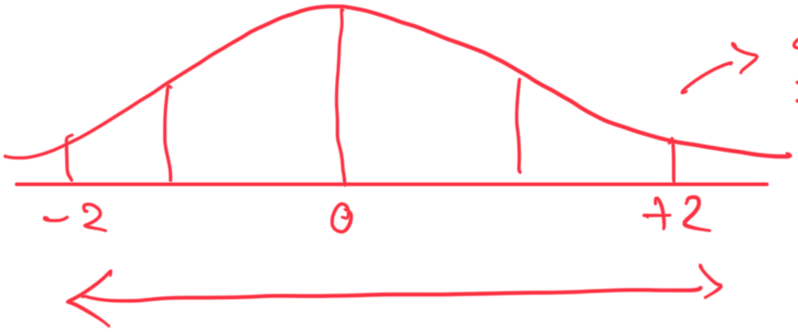
mean = 0

std = 1

34%

Normal Table

Z



→ 95%

$P(Z>2)$

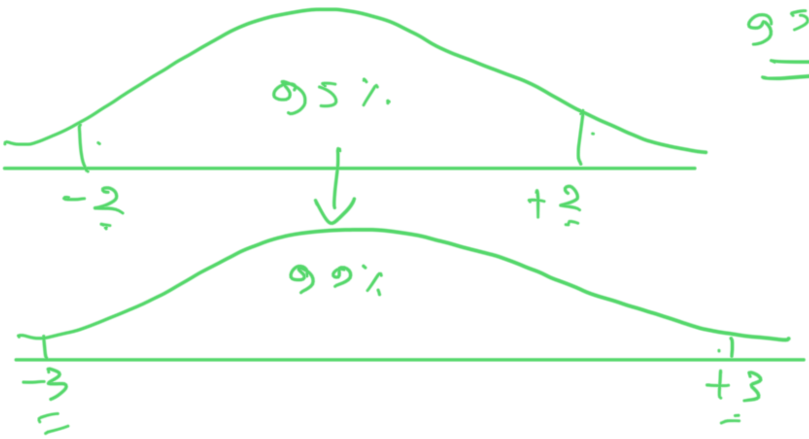
$P(Z>1)$

$P(2<Z<3)$??

$2 \to 3$

95% → 99%

4%



$$\frac{9\%}{2} = 2\%$$

$$\boxed{P(1<Z<3)}$$

$-2 \quad +2$

$-3 \quad +3$

$\boxed{CI}$

Testing of Normal dis

Confidence Interval:

Avg life Iphone

→ 2.71 year

continuous unknown data ⇒ single number

range

Confidence Interval

Confida [ ]

Avg ht Men

[0, 10 ft]

100%

90% [5.4 5.6]

??

(i) [0, 10] 100% → a exam

(ii) [5.4, 5.6) 90%

Sample

data

200

7.8 hr

Population

All students of BLR

$\boxed{427.14 \text{ mm}}$ ×

[400, 500]

How to calculate CI vs probabilities??

User eats → N(30, σ^2) $\sigma = 10$

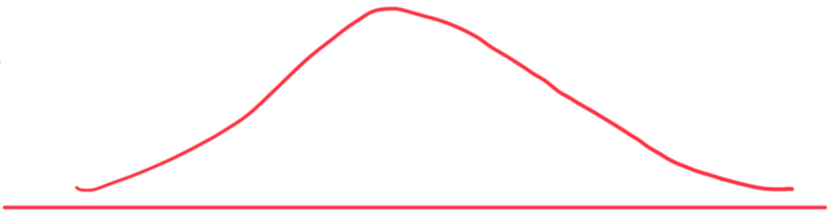
What is the Avg time for the delivery

User ear → (31, 28, 29, 27, 37, 41,)

What is the Avg CI ?? ⇒ P(Avg = 31.5)

Distribution

Data $\longrightarrow$ Distri... $N(\vec{\mu}, \vec{\sigma}^2)$



CI $\rightarrow$ Data $\rightarrow$ Distribution

Point estimate $\Rightarrow$ 4.5

Interval " $\Rightarrow$

0, 1, 2, 3,

[___]

body weight

How to calculate CI ??

95% 90%

$\Rightarrow N(\mu, \sigma^2)$

$x_1, x_2, \ldots x_n$

$\bar{x} \Rightarrow y$

?? $N\left(\mu, \frac{\sigma^2}{n}\right)$

Avg => variance

2   2.3   2.9 → 212.1..

clear??

S_1	1.7	1.4	1.3	1.6	=> 1.5
S_2	1.3	1.1	1.4	1.9	=> 1.425
S_3	1.5	1.2	1.6	1.5	=> 1.45
S_4	1.6	1.4	1.3	1.6	=> 1.475

1.3 → 1.7 =>

1.425 → 1.5

1st → 29 → var..

2nd → 41

3rd => 35

⇓

Max

JAN: 30 | Feb: 31.2 | 30.5

Sample data $\longrightarrow$ population

5.613

90% [5.6 5.8]

How will we calculate the CI??

sample $\longrightarrow$ population

$x_1 \quad \dots \quad x_n \longrightarrow N(\mu, \sigma^2)$

$\boxed{\text{case 1}}$ $\quad \mu$ ✗ $\quad \sigma^2$ ✓

case II $\quad \mu$ ✗ $\quad \sigma^2$ ✗

$\boxed{\text{case 1}}$

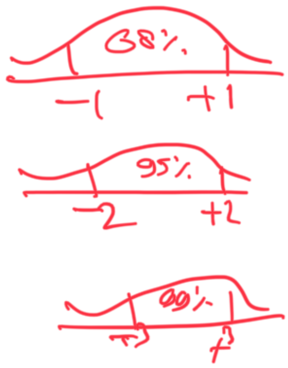
μ ✗ $\quad \sigma^2$ ✓

$\boxed{x_1 \quad \dots \quad x_n} \longrightarrow N(\mu, \sigma^2)$

reduce

$\bar{x} \longrightarrow N\left(\mu, \frac{\sigma^2}{n}\right)$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = Z \Rightarrow N(0, 1)$$



95%

$$-2 \le Z \le +2$$

$$-2 \le Z \le +2$$

$$-2 \le \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le 2$$

$$-\frac{2\sigma}{\sqrt{n}} \le \bar{X} - \mu \le \frac{2\sigma}{\sqrt{n}}$$

$$-\frac{2\sigma}{\sqrt{n}} - \bar{X} \le -\mu \le -\bar{X} + \frac{2\sigma}{\sqrt{n}}$$

$$\bar{X} + \frac{2\sigma}{\sqrt{n}} \ge \mu \ge \bar{X} - \frac{2\sigma}{\sqrt{n}}$$

$$\mu \quad \left[\bar{X} - \frac{2\sigma}{\sqrt{n}} \quad \bar{X} + \frac{2\sigma}{\sqrt{n}}\right] \quad 95\%$$

(i) Uber eats delivery time

$n = 225$ days Avg duration $\rightarrow$ 32

$\sigma = 5$

95% CI of μ

$\bar{x} = 32$

$\mu = \left[\bar{x} - \frac{2\sigma}{\sqrt{n}} \quad \bar{x} + \frac{2\sigma}{\sqrt{n}}\right]$

$\left[32 - \frac{2*5}{15} \quad \bar{x} + \frac{2*5}{\sqrt{15}\,15}\right]$ $\sqrt{225} = 15$

$[31.33, 32.67]$

$[\bar{x} - 0 \quad \bar{x} + 0]$

n

Sample

$\bar{x} \to$ sample mean

$s \to$ sample ~~var~~ SD

$s^2 \to$ " var

Popu...

$\mu \to$ population mean

$\sigma \to$ " SD

$\sigma^2 \to$ " var

metro every day

house $\longrightarrow$ office

$\sigma = 5$ min

95%

length 1 min or less

12 m

$$\left[\bar{x} - \frac{2\sigma}{\sqrt{n}} \quad \bar{x} + \frac{2\sigma}{\sqrt{n}} \right]$$

$$\cancel{\bar{x}} + \frac{2\sigma}{\sqrt{n}} - \left(\cancel{\bar{x}} - \frac{2\sigma}{\sqrt{n}} \right)$$

$$\frac{4\sigma}{\sqrt{n}} < 1 \qquad \frac{4\sigma}{\sqrt{n}} < \frac{1}{2}$$

$$20 < \sqrt{n} \qquad \sqrt{n} > 40$$

$$400 < n \qquad n > 1600$$

case 1

μ unknown

σ known $\longrightarrow$

case II

σ not known

"s"

1 2 3

$$\frac{(1-3)^2 + (2-3)^2 + (3-3)^2}{3-1} = 1$$

$$= \sqrt{1}$$

$= 1$

$\boxed{\text{T distribution}}$