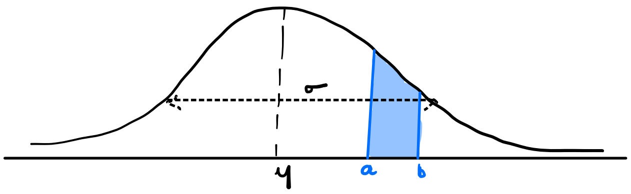
Normal distribution → 1808/9

→ Its a continuous distribution



→ It has 2 parameters i.e mean (μ) and variance (σ)

→ More spread ⇒ more variance and less spread ⇒ less variance

→ total area under This Curve = 1 ∵ probability = 1

→ Area from a to b ⇒ Integration

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

$X + 1 \sim N(\mu+1, \sigma^2)$

i.e

$$\boxed{\begin{aligned} &X + c \sim N(\mu + c, \sigma^2), \quad \underline{\text{given } X \sim N(\mu, \sigma)} \\ &c \cdot X \sim N(\mu \cdot c, \sigma^2 \cdot c^2), \end{aligned}}$$

1 2 3 = 2 $\qquad 1^2 + 0 + 1^2 = \frac{2}{2} = 1$

2 4 6 = $12/3 = 4$ $\qquad 2^2 + 0 + 2^2 = 8/2 = 4 \} \; 2^2$

$2 * 2$

→ If we start with any normal distribution. We can revert to $\boxed{X \sim N(0, 1)}$

⇓

standard distribution

If we study This we can try study everything else.

$$X \sim N(\mu, \sigma^2) \xrightarrow{\text{goal}} X \sim N(0, 1)$$

wkt $X + c \sim N(\mu + c, \sigma^2)$

step 1 $c = -\mu$, $X - \mu \sim N(\mu - \mu, \sigma^2)$

$\Rightarrow X - \mu \sim N(0, \sigma^2)$

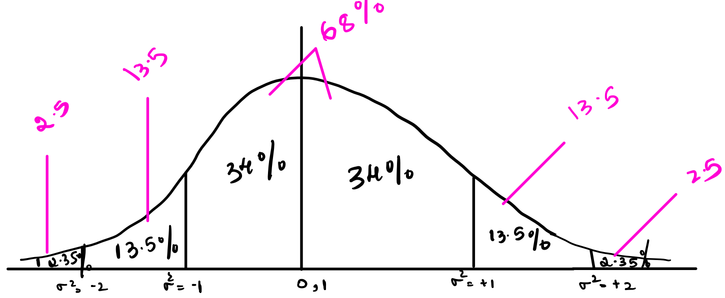
step 2 $c = 1/\sigma$, $c^2 = \frac{1}{\sigma^2}$

$c(x - \mu) \sim N(0 * 1/\sigma, \sigma^2 \times 1/\sigma^2)$

$\boxed{\frac{1}{\sigma}(x - \mu) \sim N(0, 1)}$ given $X \sim N(\mu, \sigma)$ Normal distribution

Normal distribution table

Very imp Bell curve



Problem

① $X \sim N(30, 100)$ $\mu = 30$

what is $P(X < 40)$ $\sigma^2 = 100$, $\sigma = 10$

soln $\frac{x - \mu}{\sigma}$

$= P(x < 40)$

$$= P(x - 30 < 40 - 30)$$

$$= P\left(\frac{x-30}{10} < \frac{40-30}{10}\right)$$

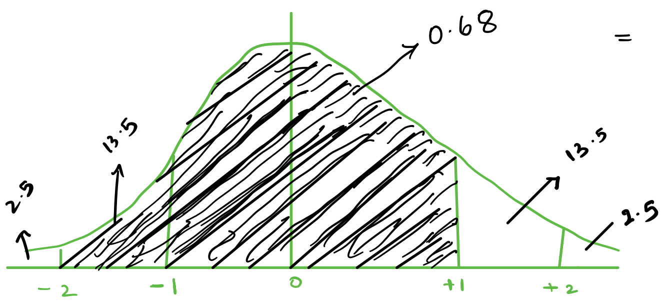
standard normal

z ↓

$$= P\left(\underbrace{\frac{x-30}{10} < 1}\right) \rightarrow P(z<1)$$

$= P(z<1)$ ∴ the area we are looking for is less than +1

i.e



$= \underline{0.68 + 0.16 = 0.84}$

② $P(x<50)$, $x \sim N(30,100)$

$$P\left(\underset{z}{\frac{x-30}{10}} < \frac{50-30}{10}\right)$$

$$P(z<+2)$$

∴ $68 + 13.5 + 13.5 + 2.5$

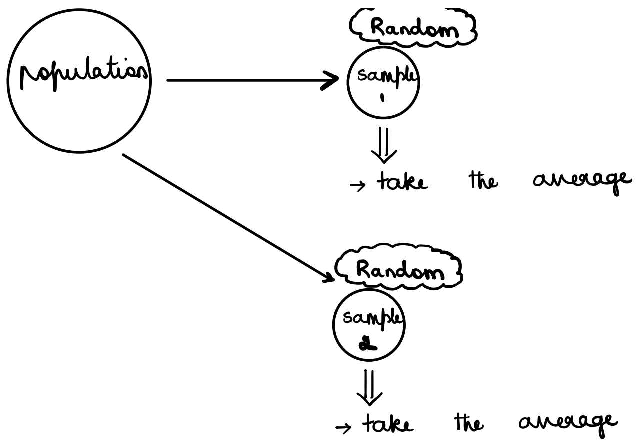
$= 97.5$

<u>Confidence Interval</u> ⇒ Range and confidence percentage

<u>Inferential statistics</u>

population

sample



for continuous variables ⇒ probability is $1/\infty$ ie infinitely many choices

∴ probability is 0

In continuous variable probability of a single number is zero, 0.

But probability of an interval is never zero 0.

↓

This interval is called confidence interval

$$\boxed{\text{Confidence interval} \Rightarrow \text{we describe an interval and rate how confident we are in that interval}}$$

CI

$x_1, x_2, x_3 \ldots, X_n \longrightarrow N(\mu, \sigma^2)$

$\bar{x} = \text{avg} = \frac{x_1 + \ldots + x_n}{n}$

$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$\boxed{CI \Rightarrow \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)}$$

⇓ standardising This

$$\frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$\boxed{\underbrace{\frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma}}_{Z} \sim N(0,1)}$$

$\Rightarrow$ If we say confidence interval is 95% means $-2 \le Z \le +2$

① 95% $\Longrightarrow$ $-2 < \dfrac{\bar{x}_n - \mu}{\sigma/\sqrt{n}} < +2$

$$-2 \ast \overbrace{\frac{\sigma}{\sqrt{n}}}^{a} < \bar{x}_n - \mu < +2 \ast \overbrace{\frac{\sigma}{\sqrt{n}}}^{a}$$

$$\boxed{-2a < \bar{x}_n - \mu < 2a}$$ where $a = \dfrac{\sigma}{\sqrt{n}}$

$$\Downarrow$$

$$\boxed{-2a < \mu - \bar{x}_n < +2a}$$

$$\boxed{x_n - 2a \le \mu \le x_n + 2a} \Rightarrow 95\%$$

$\therefore$ 95% confidence for the value 'mean' μ in the interval $\left(x_n - \dfrac{2\sigma}{\sqrt{n}},\ x_n + \dfrac{2\sigma}{\sqrt{n}}\right)$

② 68% $\Rightarrow$ $-1 \le \dfrac{\bar{x}_n - \mu}{\sigma/\sqrt{n}} \le 1$

$x_n - a \le \mu \le x_n + a$ $\qquad a = \dfrac{\sigma}{\sqrt{n}}$

68% confidence in the interval $\left[x_n - \dfrac{\sigma}{\sqrt{n}},\ x_n + \dfrac{\sigma}{\sqrt{n}}\right]$

$\Big[\quad \sqrt{n} , \qquad \overline{\sqrt{n}} \Big]$

Any parameter about the sample is known & that of population is not known.

$$\frac{\sum (x_i - \bar{x})^2}{N}$$

$\boxed{\dfrac{\bar{x}_n - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)}$ $\Rightarrow$ This cannot be applied to population variance since we do not know σ^2 for sample

$\therefore$ Instead we take 'S' which is the variance of the sample

Sample variance $S^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

Population σ^2 population variance $\longrightarrow$ Sample s^2 known $\quad s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

<u>Student t - distribution</u> → 1875 ~ 1907

→ height of normal distribution is greater than t-distribution.

<u>degree of freedom</u>

↳ as the degree of freedom increases normal distribution becomes equal to t-distribution

σ^2 is known

$x_1, x_2, \ldots, x_n \to N(\mu, \sigma^2)$

$$\frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)$$

σ^2 is unknown

$x_1, x_2 \ldots x_n \longrightarrow N(\mu, \sigma^2)$

$$\frac{\bar{x}_n - \mu}{s/\sqrt{n}} \longrightarrow t_{n-1} \text{ (degrees of freedom)}$$

if $n > 30$, the above distribution is normal

else $n < 30$, the above distribution is 't'.

Hypothesis testing

→ hypothesis is an assumption and testing is basically testing it out.

ex: Employement rote in BLR is 0.9

Is this true ???

this is an hypothesis

Testing this is hypothesis testing

1) Null hypothesis

doesnot agree

- No new information
- go with existing knowledge
- ∴ We negate the the present hypothesis

2) Alternative hypothesis

does agree

- We have new information
- We have new knowledge
- The current huhothesis we are testing

is correct

After creating the hypothesis, we get three hypothesis statements

All these types have different testing procedures

Type I
- $H_0 \rightarrow$ null hypothesis $\rightarrow \mu = \mu_0$
- $H_1 \rightarrow$ alternate hypothesis $\rightarrow \mu \neq \mu_0$

Type II
- $H_0 : \mu \leq \mu_0$
- $H_1 : \mu > \mu_0$

Type III
- $H_0 : \mu \geq \mu_0$
- $H_1 : \mu < \mu_0$