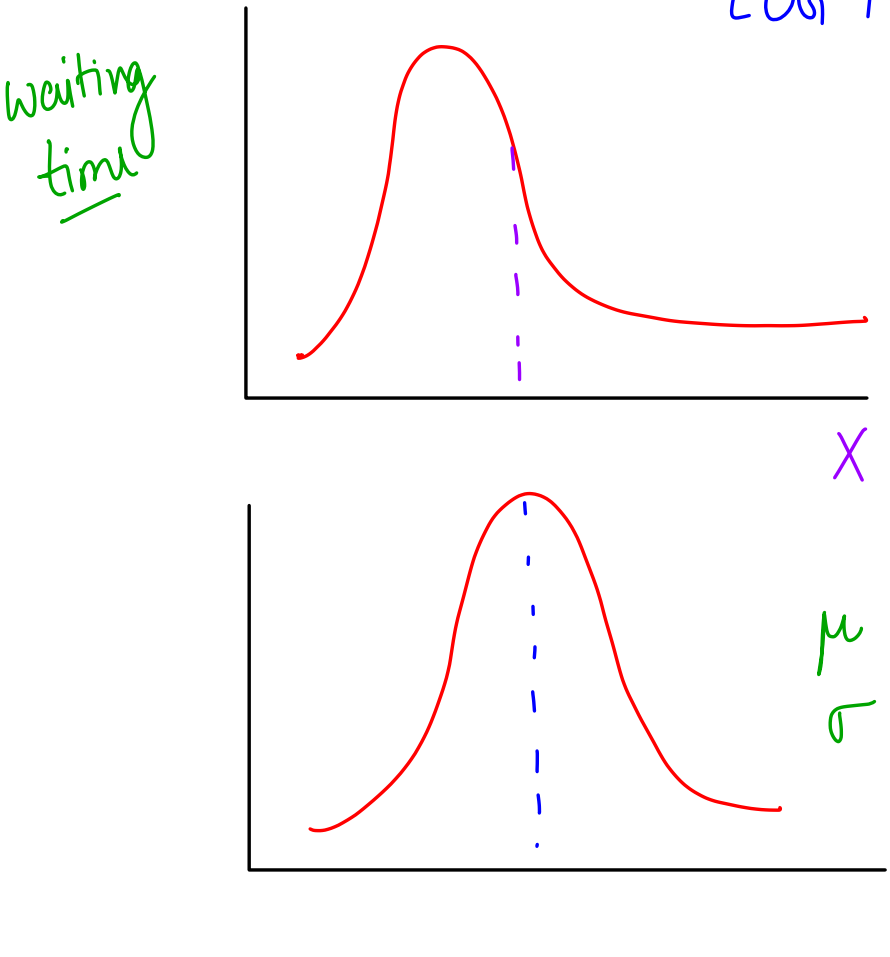
FEATURE ENGINEERING-1

LOG NORMAL DISTRIBUTION



$$F(X) = e^{\left(\mu + \frac{\sigma^2}{2}\right)}$$

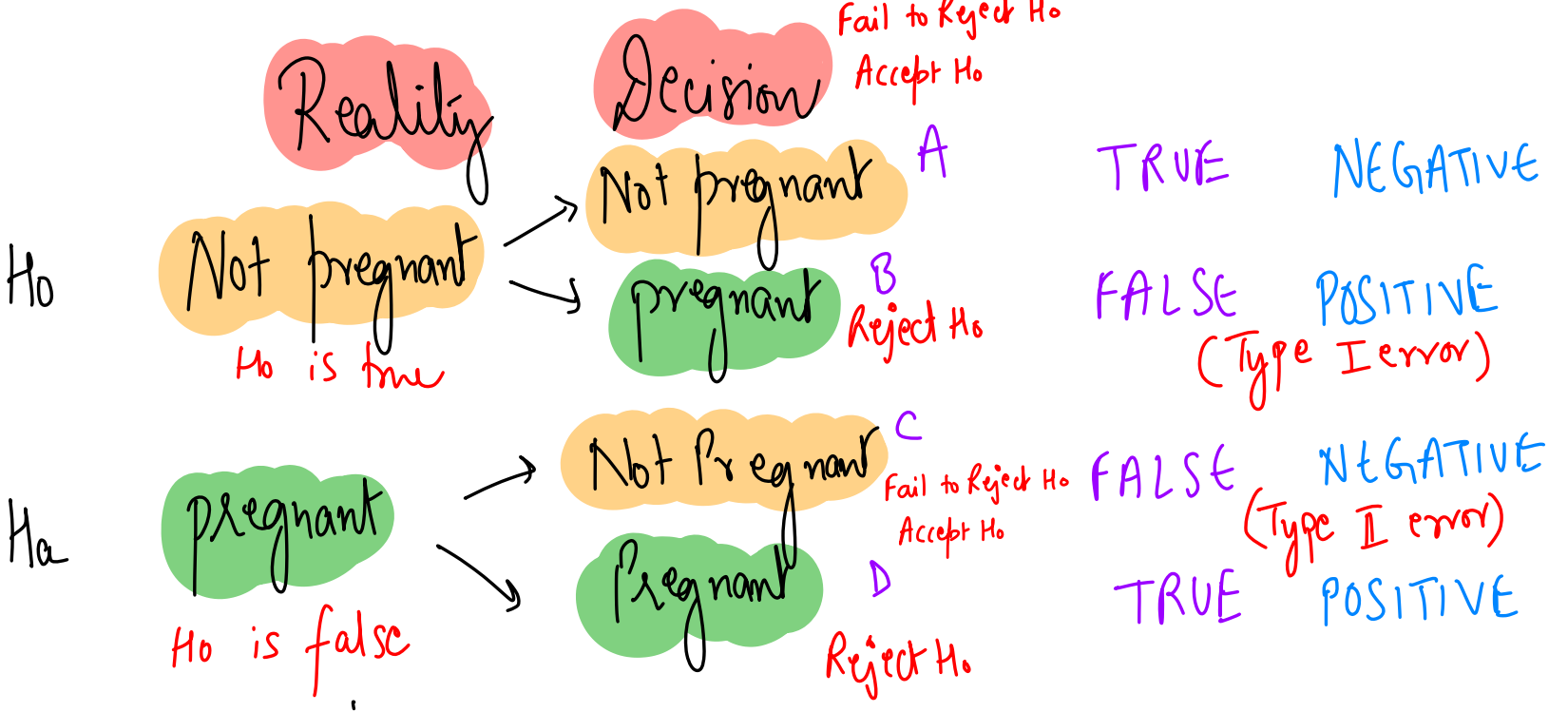
$$Var(X) = \left(e^{\sigma^2} - 1\right)\left(e^{2\mu + \sigma^2}\right)$$

FORTIS

H_0: Not Pregnant

H_a: Pregnant.

* Reject H_0
* Fail to Reject H_0 / Accept H_0



Height of Normal people

$\mu = 65$

$\sigma = 4$

Height of NBA player

$\mu = 72$

$\sigma = 3$

(-ve) H_0 : (Normal) $\mu = 65$

(+ve) H_a : (NBA player) $\mu > 65$

"Right tailed test"

99% Confidence

$\alpha = 0.01$

(1) Sample **2 Individual** $\Rightarrow$ Avg. = 70

pvalue = 0.1638

$1 - t.cdf\left(\frac{70-65}{4/\sqrt{2}}\right), dof = 1$

$\alpha = 0.01$

pvalue $> \alpha$ "Fail to Reject H_0"

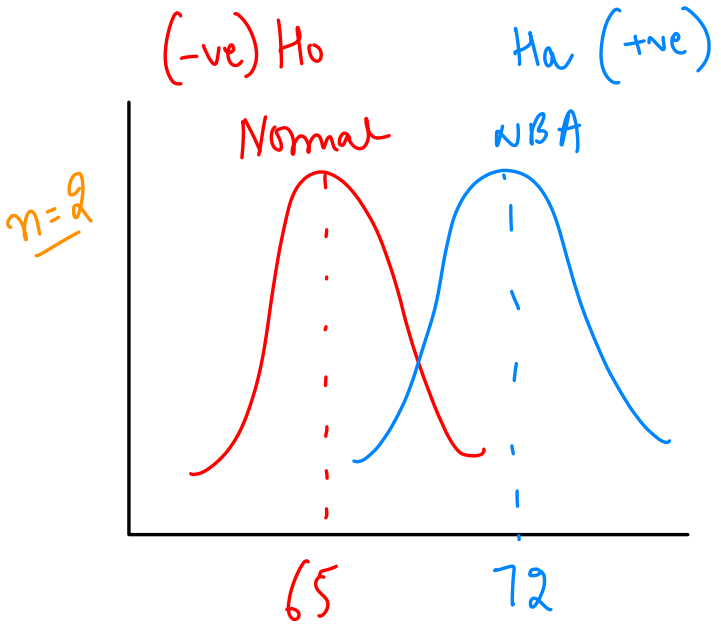
(2) Sample **10 Individuals** $\Rightarrow$ Avg = 70

pvalue = 0.00167

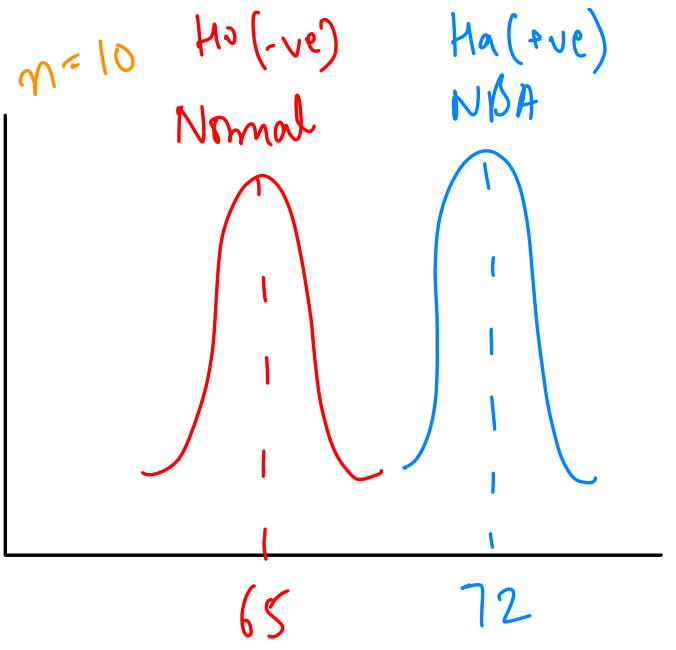
$1 - t.cdf\left(\frac{70-65}{4/\sqrt{10}}\right)$ $dof = 9$

$\alpha = 0.01$

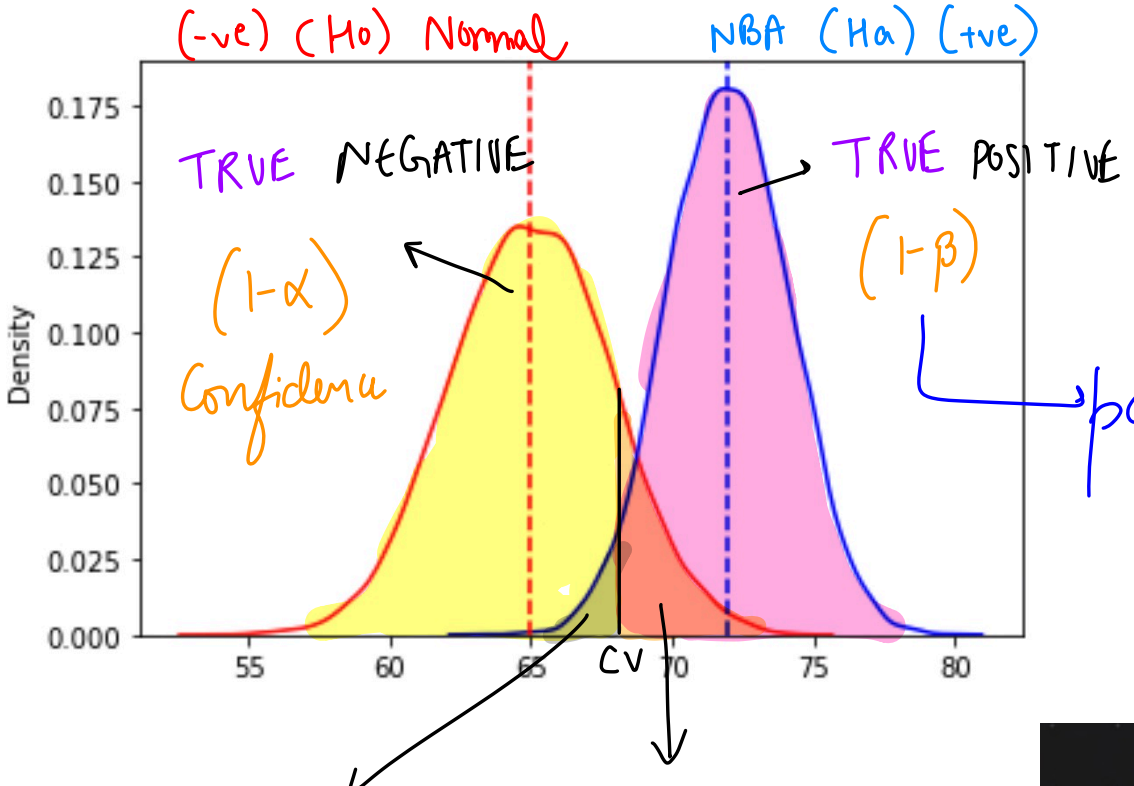
pvalue $< \alpha$ "Reject H_0"



more overlap



less overlap



Confidence = 99%

$\alpha = 0.01$

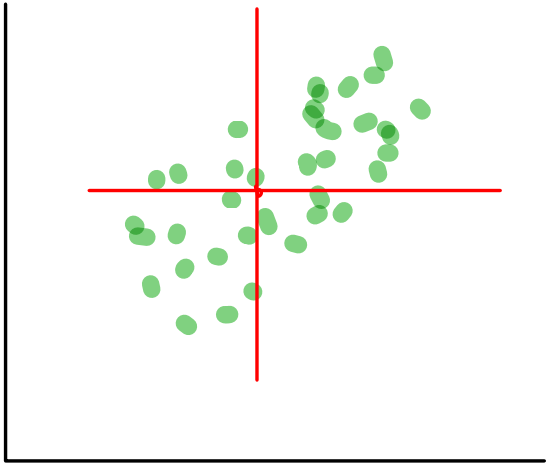
	Decision	
	Accept	Reject
True H_0	True negative $1-\alpha$ Confidence level	False positive α Significance level
False	False negative β	True positive $1-\beta$ Power

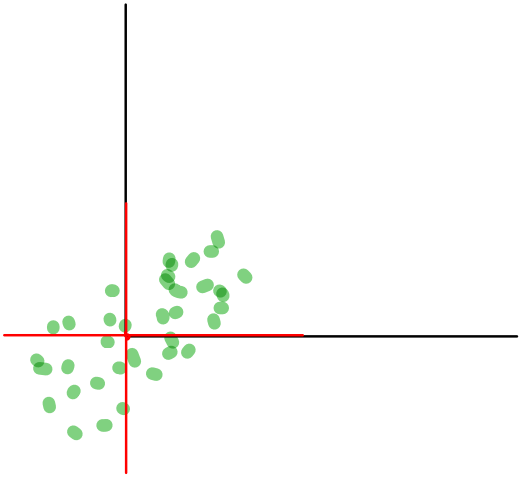
Dataset

Standardisation | Normalisation (Min Max Scaling)

μ - weight
σ - weight

Weights (in g)	Heights (Cms)	Target				
10000	1	CAT	$\frac{x_1 - \mu_w}{\sigma_w}$	-2	$\frac{x_1 - \min(x)}{\max(x) - \min(x)}$	0
20000	2	DOG	$\frac{x_2 - \mu_w}{\sigma_w}$	-3	$\frac{x_2 - \min(x)}{\max(x) - \min(x)}$	0.1
16000	3	DOG	$\frac{x_3 - \mu_w}{\sigma_w}$	-1		0.3
14000	4	CAT		-1.5		0.4
15000	5	CAT		0		0.5
20000	6	CAT		1		0.2
45000	7	DOG		2		0.25
				3		



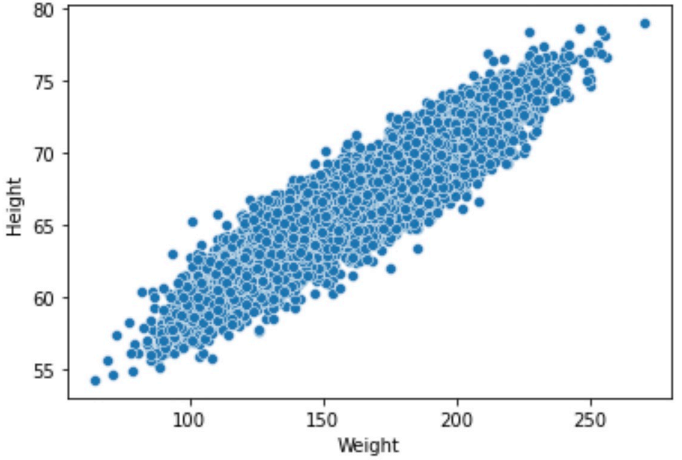


x_1, x_2, x_3, x_4, x_5

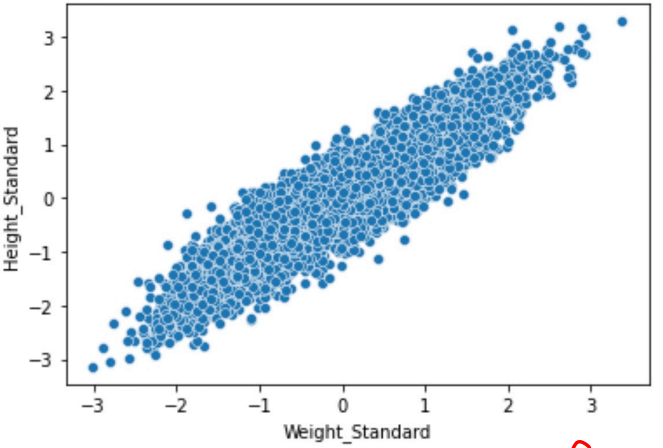
$\bar{x}$

$$\frac{x_1 - \bar{x}}{\sigma_x}, \frac{x_2 - \bar{x}}{\sigma_x}, \frac{x_3 - \bar{x}}{\sigma_x}$$

Original

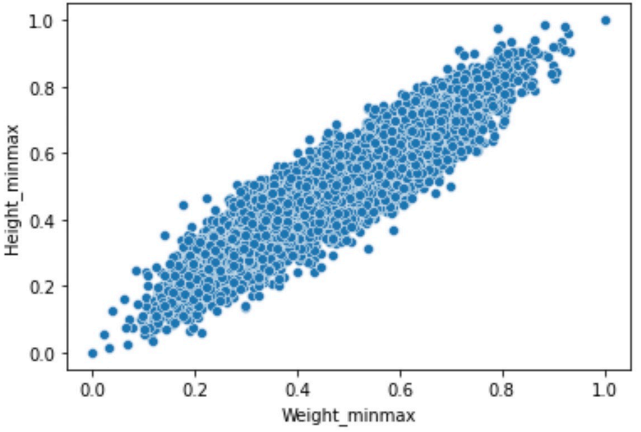


Standardised



Range: Unknown.

MINMAX Scaling



[0,1]

Range. known