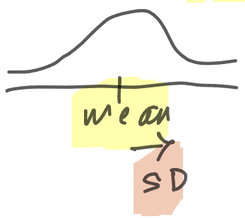
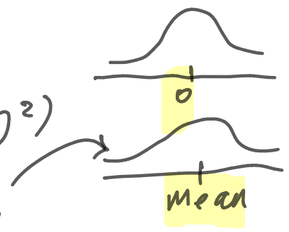
History: Abraham de Moivre (1720)

Why famous curve? — Naturally occuring (eg heights)
— CLT (Module 3)
— Defined by mean & SD $N(\text{mean}, SD)$ ('parameters')



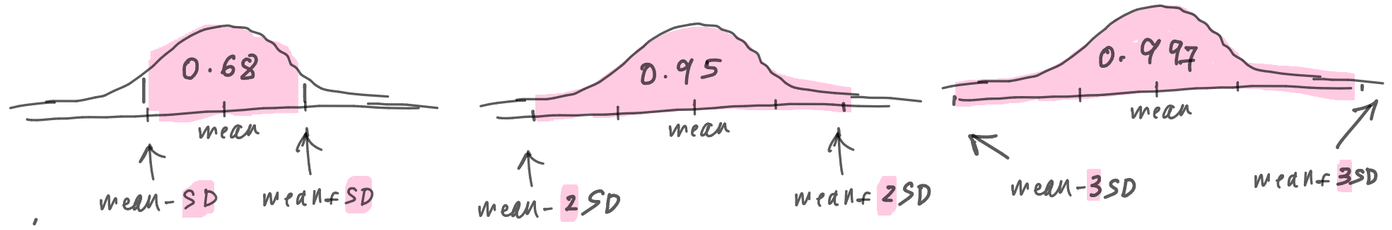
2 types:
Standard $N(0,1)$
General $N(\text{mean}, SD^2)$



$$f(x) = \frac{1}{\sqrt{2\pi}SD} e^{-\frac{(x-\text{mean})^2}{2SD^2}} \quad -\infty \le x \le \infty$$

Special Properties:

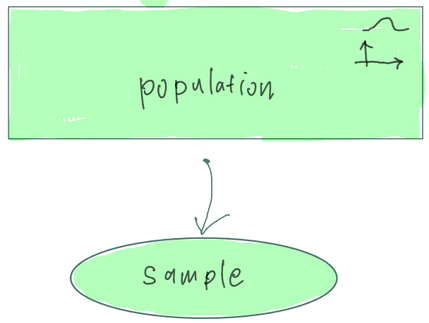
① 68|95|99.7% Rule



② General Normal → re-scaled → Standard Normal

$$Z = \frac{\text{data point} - \text{mean}}{SD}$$

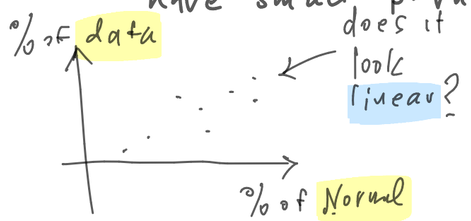
Module 2: Modelling Data



T4: Normal Model
T5: Linear Model

LO4: Apply the Normal approximation to data, with consideration of measurement error.

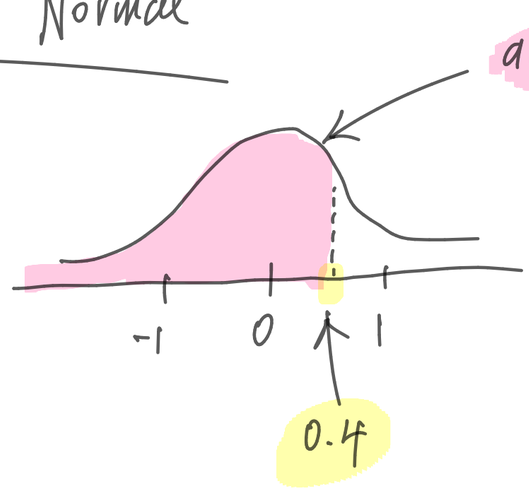
"All models are wrong but some models are useful."
(Box et al, 2009)

Model	Variable	When?	What?	Diagnostics	Predictions
Normal	1 Quant	 looks bell-shaped	 Normal approximation $N(\text{mean}, SD^2)$	① Is the histogram in the proportions 68|95|99.7% (1|2|3 SDs) ② Is the QQ plot linear? ③ Does the Shapiro-Wilk test have small p-value? 	 • integration • tables • pnorm

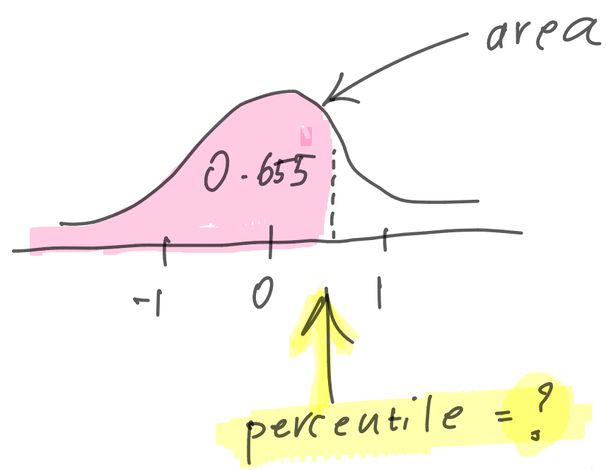
How to use R for the Normal

$X \sim N(0, 1)$

mean = 0 ; SD = 1



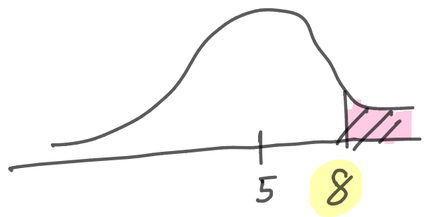
pnorm (0.4) ≒ 0.655



quorm (0.655) ≒ 0.4

$X \sim N(5, 3^2)$

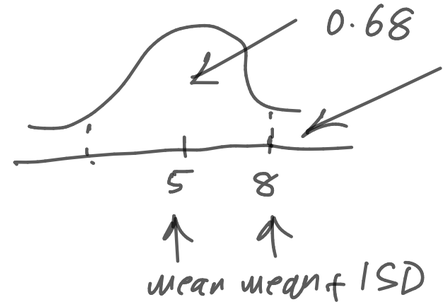
mean = 5 ; SD = 3



pnorm (8, 5, 3, lower.tail = F)

≒ 0.159

check:



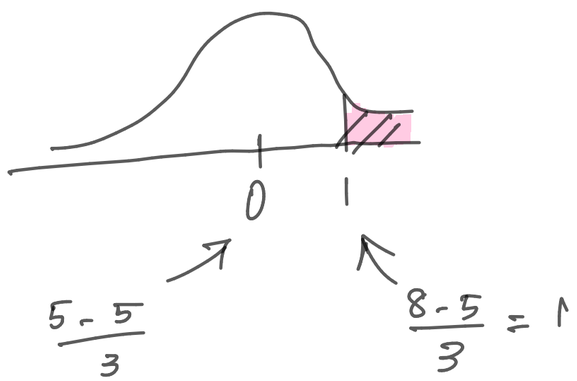
(1 - 0.68)/2 = 0.16

⋮ rescale

⋮

V

$Z \sim N(0, 1^2)$



pnorm (1, lower.tail = F)

≒ 0.159