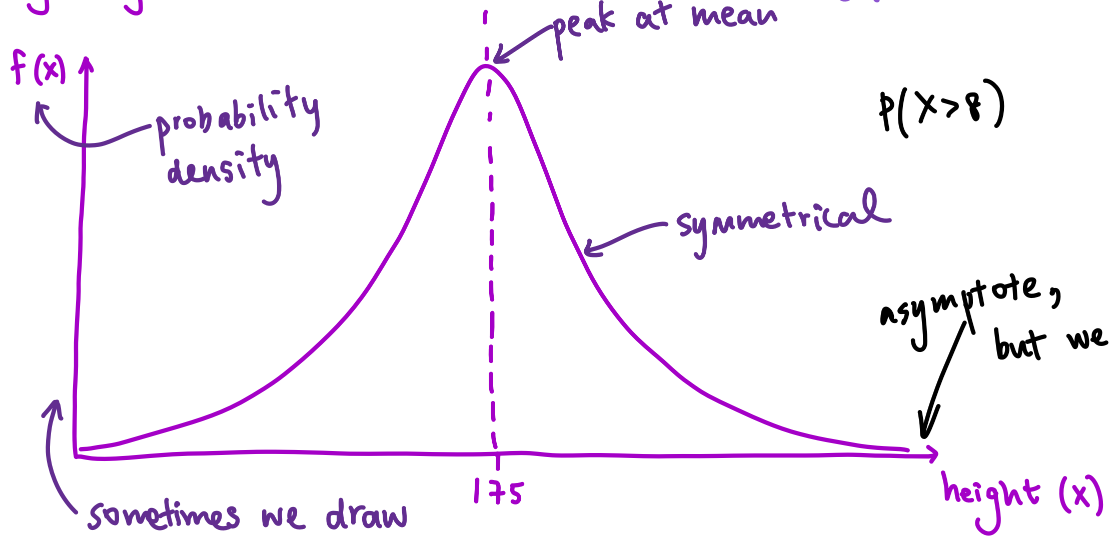
Normal Distribution: Continuous Data

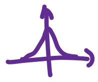
it HAS to be continuous

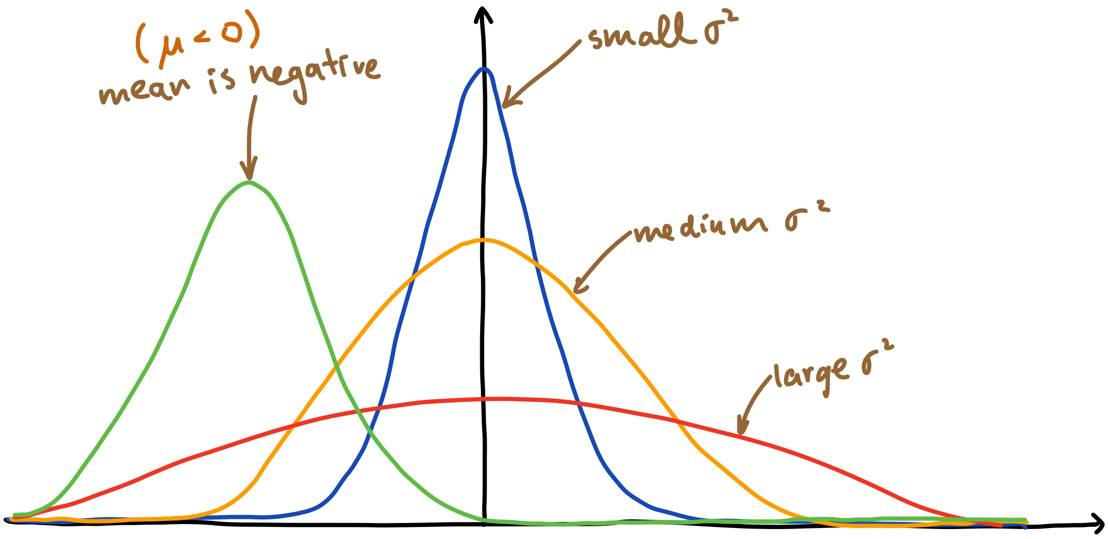
e.g height



$P(X>8)$

sometimes we draw y-axis at the mean





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

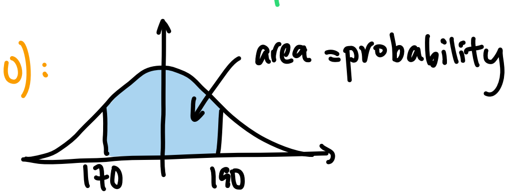
Normal Distribution

mean

variance / standard deviation squared

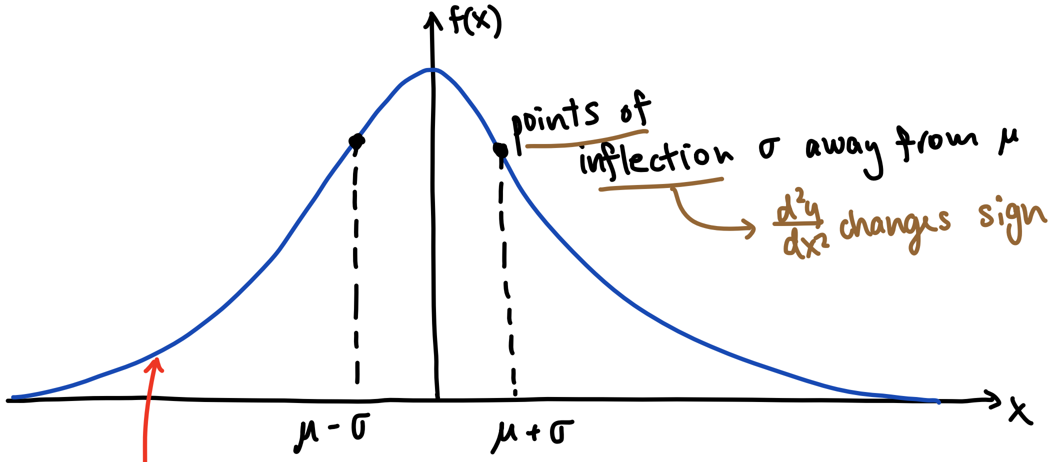
In a normal distribution graph, the area under the graph = 1
just like how sum of discrete data probabilities = 1

Find $P(170 < X < 190)$:



We never try to find $\underbrace{P(X=180)}_{\text{infinitesimally small (continuous data)}}$!

Normal Distribution facts!

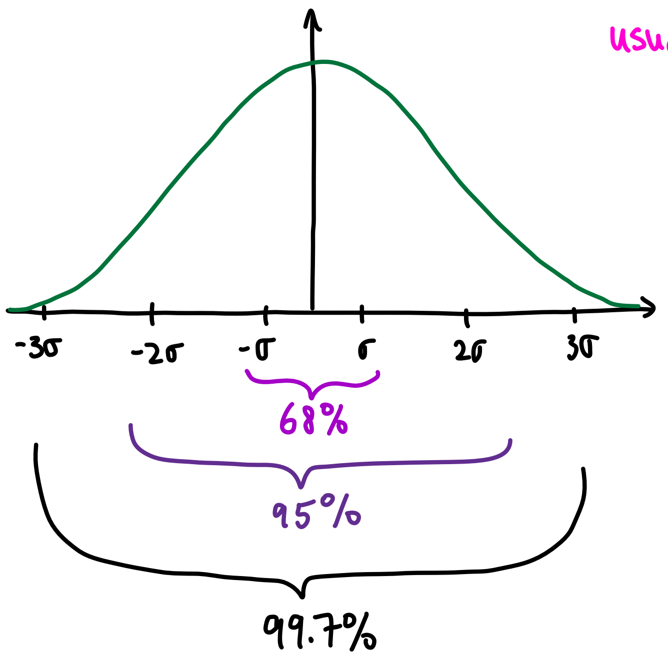


points of inflection σ away from μ

→ $\frac{d^2y}{dx^2}$ changes sign

If symmetrical, mean = median = mode!

The 68 - 95 - 99.7 rule



usually all data lies within 5σ of μ

Calculator

Distribution → Normal $\overbrace{\text{CD}}^{\text{NOT PD}}$

$</>$ and $\le/\ge$ are the same

parameters

- Lower
- Upper
- σ
- μ

Lower, Upper: e.g. $6 < X \le 7$: lower = 6, upper = 7

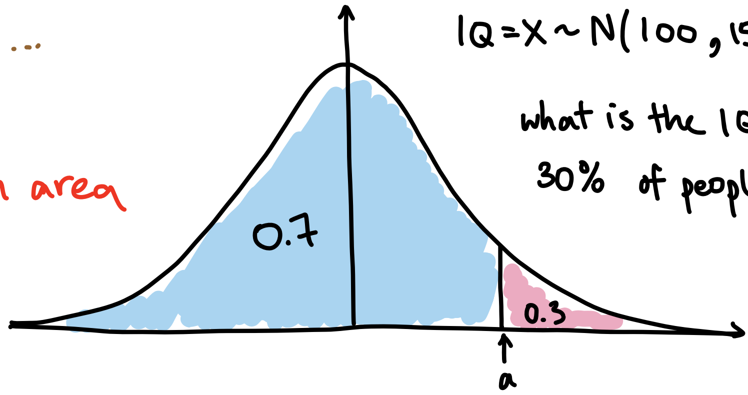
$X > 7$: lower = 7, upper = 100000 (some large number)

σ, NOT σ^2

finding area from range

What if ...

Finding range from area



$IQ = X \sim N(100, 15^2)$

what is the IQ of the top 30% of people?

Remember!

$P(a < X < b)$
$= P(X < b) - P(X < a)$

Calculator

Distribution ⇒ Inverse Normal

$P(X > a) = 0.3$

$P(X < a) = 0.7$

$a = 107.86$

Area : 0.7

μ : 100

σ : 15

Standard Normal Distribution

Z = # of standard deviations from the mean

e.g $IQ = X \sim N(100, 15^2)$

IQ	Z
100	0
130	2
85	-1
165	4.333
62.5	-2.5

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

AND

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

phi

$P(Z < a) = \Phi(a)$

A z-table

p	z
0.5	0
0.4	0.2533
0.3	0.5244
⋮	⋮

"Top p%"

where $P(Z > z) = 1 - \Phi(z) = p$

e.g. top 30% IQ:

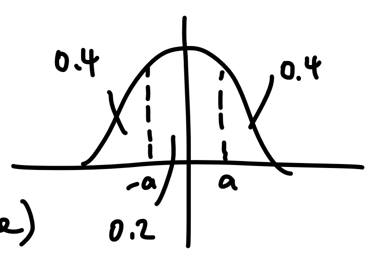
$$0.5244 = \frac{X - 100}{15}$$

$X = 107.866$

Solve for a: $P(-a < Z < a) = 0.2$

$P(Z > a) = 0.4$

$a = 0.2533$ (from table)



Ex 3C

1. $X \sim N(30, 5^2)$

a) $P(X < a) = 0.3$

$a = 27.38$

b) $P(X < a) = 0.75$

$a = 33.37$

c) $P(X > a) = 0.4$

$P(X < a) = 0.6$

$a = 31.27$

d) $P(32 < X < a) = 0.3$

$P(X < a) - P(X < 32) = 0.3$

$P(X < a) = 0.3 + 0.655$

$a = 38.50$

2. $X \sim N(12, 3^2)$

a) $P(X < a) = 0.1$

$a = 8.16$

b) $P(X > a) = 0.65$

$P(X < a) = 0.35$

$a = 10.84$

c) $P(10 \leq X \leq a) = 0.25$

$P(X \leq a) - P(X \leq 10) = 0.25$

$P(X \leq a) = 0.25 + 0.25$

$a = 12.02$

d) $P(a < X < 14) = 0.32$

$P(X < 14) - P(X < a) = 0.32$

$P(X < a) = 0.75 - 0.32$

$a = 11.45$

3. $X \sim N(20, 12)$ or $N(20, \sqrt{12}^2)$

a) i) $P(X < a) = 0.4$

$a = 19.12$

ii) $P(X > b) = 0.6915$

$P(X < b) = 0.3085$

$b = 18.27$

b) $P(b < X < a) = P(X < a) - P(X < b) = 0.4 - 0.3085 = 0.0915$

Finding μ or σ using z

e.g. $X \sim N(\mu, 3^2)$. If $P(X>20)=0.2$

$P(X>20)=0.2$

$0.8416 = \frac{20-\mu}{3}$ } find in table & substitute

$\mu = 17.5$ } solve for μ

} Same method if σ is missing

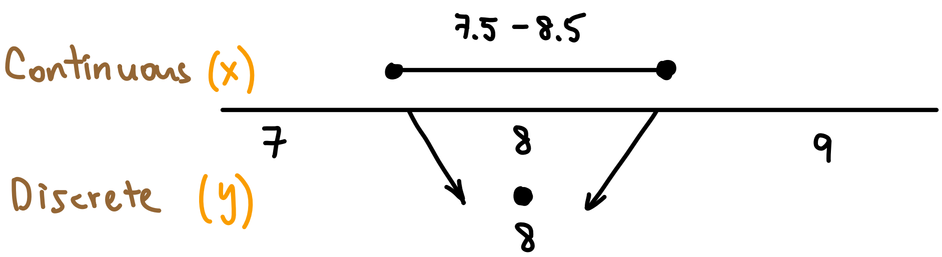
If both μ & σ are missing, simultaneous eqn

Approximating Binomial Distribution with Normal Distribution.

$\mu = np$

$\sigma = \sqrt{np(1-p)}$ where $B(n,p) \Rightarrow n$ is large & $\underbrace{p \approx 0.5}_{\text{because distribution is symmetrical}}$

Continuity Correction



$P(7.5 \le x < 8.5) \approx P(y=8)$

Another example: $P(y<9) = P(y \le 8) \approx P(x<8.5)$

Hypothesis Testing

Same as binomial

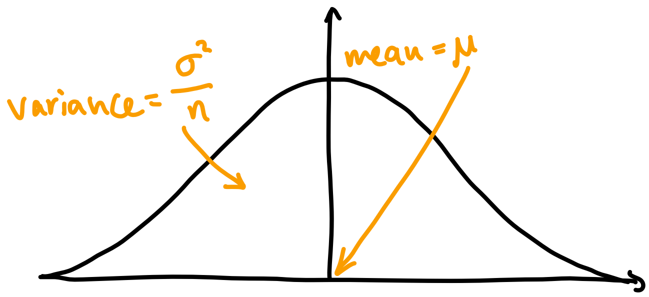
e.g. Children age

Population: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $\mu = 4.5$

Sample 1: 1, 3, 7, 8 $\bar{x}_1 = 4.75$

Sample 2: 6, 2, 0, 9 $\bar{x}_2 = 4.25$

When we have a lot of samples, the sample means ($\bar{x}$) will follow a normal distribution.



When sample size ↑ & # of samples ↑, σ of the normal distribution will ↓ and the mean of the sample means get closer to μ (the population mean)

e.g. juice in cartons ($\sigma = 3$, $\mu = 60$)

$H_0: \mu = 60$ ml X = amount of juice per carton

$H_1: \mu < 60$ ml

Assuming H_0 is true, $X \sim N(60, 3^2)$

Inspection: mean of 16 cartons = 59.1 ml

$$\bar{X} \sim N\left(60, \frac{3^2}{16}\right) \rightarrow \bar{X} \sim N(60, 0.75^2)$$

mean variance = (original variance)/(# of samples)

$P(\bar{x} < 59.1) = 0.1151$

$0.1151 > 0.05$ (significance level)

$\therefore$ Not enough evidence to reject H_0

Machine producing bolts: Diameter D mean = 0.580 cm deviation = 0.015 cm

Claim: mean has changed after machine service

Find critical region Use 1% sig. level

$H_0: \mu = 0.580$ $H_1: \mu \neq 0.580$ ← 2-tailed test, split sig level

$D \sim N(0.580, 0.015^2)$

$\bar{D} \sim N\left(0.580, \frac{0.015^2}{50}\right) \longrightarrow \bar{D} \sim N\left(0.580, \left(\frac{0.015}{\sqrt{50}}\right)^2\right)$

$z = \pm 2.5758$ (from z table: $\underbrace{0.0050}_{\text{significance level}} \rightarrow 2.5758$)

$-2.5758 = \dfrac{\bar{D} - 0.580}{\frac{0.015^2}{\sqrt{50}}} \rightarrow \bar{D} = 0.5745$

$2.5758 = \dfrac{\bar{D} - 0.580}{\frac{0.015^2}{\sqrt{50}}} \rightarrow \bar{D} = 0.5854$

Critical region: $\bar{D} \leq 0.5745$ or $\bar{D} \geq 0.5854$