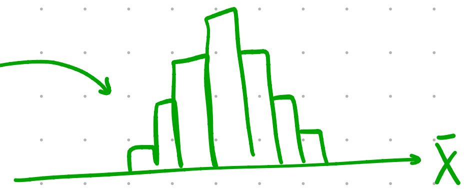
CENTRAL LIMIT THEOREM

If we sample from some population (regardless of its distribution) m times (each of sample size n)

$$
\begin{array}{llll}
X_1^{(1)}, & X_2^{(1)}, & \dots & X_n^{(1)} \longmapsto \bar{X}^{(1)} \\
X_1^{(2)}, & X_2^{(2)}, & \dots & X_n^{(2)} \longmapsto \bar{X}^{(2)} \\
\vdots & \vdots & \ddots & \vdots \qquad\quad \vdots \\
X_1^{(m)}, & X_2^{(m)}, & \dots & X_n^{(m)} \longmapsto \bar{X}^{(m)}
\end{array}
$$

If we do this, we get a LOT of means!

If n is large ($n \to \infty$), $\bar{X}$ is distributed normally. (approximately)



Regardless of initial distribution of population!

This pops up if $m \to \infty$

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

The more non-normal the dist$^{\underline{n}}$ of the population is, the larger n needs to be for the Central Limit Theorem (aka CLTh$^{\underline{m}}$) to be effective.

Ex 5A

③ Length of bolts: $\mu = 3.03\text{cm}$ $\quad \sigma = 0.2\text{cm}$

a) Sample of 100 bolts: $\bar{X} \stackrel{\cdot}{\sim} N\left(3.03, \frac{0.2^2}{100}\right)$

$P(\bar{X} < 3) = 0.0668$

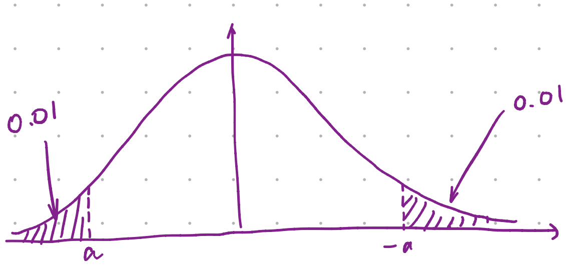
b) $\bar{M}$ is the mean length of bolts $\quad \bar{M} \stackrel{\cdot}{\sim} N\left(3.03, \frac{0.2^2}{n}\right)$

When finding minimum sample sizes, ALWAYS standardise!

Find n s.t. $P(\bar{M} < 3) < 0.01 \qquad Z \sim N(0,1)$

$$P(\bar{M} < 3) = P\left(\frac{\bar{M} - 3.03}{\frac{0.2}{\sqrt{n}}} < \frac{3 - 3.03}{\frac{0.2}{\sqrt{n}}}\right) = P\left(Z < -\frac{0.03\sqrt{n}}{0.2}\right)$$

(convert to standard normal)



if $p = 0.01$, $z = 2.3263 = -a$ (from table)

$$2.3263 = \frac{0.03\sqrt{n}}{0.2}$$

$$\sqrt{n} = 2.3263 \times \frac{0.2}{0.03} = 15.5$$

$$n = 15.5^2 = 240.5\dots$$

for $P(\bar{M} < 3) < 0.01$, $\quad n > 240.5\dots$, $\quad \therefore n = 241$

(5) Fair dice rolled 35 times.

We know the $dist^{\underline{n}}$ of dice.
Use it to find μ and σ^2

a) X = score on the dice

x	1	2	3	4	5	6
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\mu = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6}$$
$$= 3.5$$

$$\sigma^2 = \frac{1+4+9+16+25+36}{6} - 3.5^2$$
$$= \frac{35}{12} = 2.916\ldots \approx 2.917$$

$$\bar{X} \sim N\left(3.5, \frac{35}{12\times 35} = \left(\frac{1}{\sqrt{12}}\right)^2\right) \qquad P(\bar{X} > 4) = 0.0416$$

b) probability that the total of 35 rolls is less than 100.

$$P\Big(\underbrace{\sum_{i=1}^{35} X_i}_{\text{total of 35 rolls}} < 100\Big) = \underbrace{P\Big(\underbrace{\frac{1}{35}\sum_{i=1}^{35} X_i}_{\text{this is the mean}} < \frac{100}{35}\Big)}_{\text{divide both sides by 35}} = P\left(\bar{X} < \frac{100}{35}\right) = 0.0130$$

(8) Fair dice rolled n times.

The mean of the scores should not differ by more than 0.1 (from 3.5) more than 1% of the time.

$$\bar{X} \sim N\left(3.5, \frac{2.917}{n}\right) \qquad P(\bar{X} > 3.6) \le 0.005$$
$$Z \sim N(0,1) \qquad P(\bar{X} < 3.4) \le 0.005$$

convert to standard normal

$$P(\bar{X} > 3.6) = P\left(\frac{\bar{X} - 3.5}{\sqrt{\frac{2.917}{n}}} > \frac{3.6 - 3.5}{\sqrt{\frac{2.917}{n}}}\right) = P\left(Z > \frac{0.1\sqrt{n}}{\sqrt{2.917}}\right) \le 0.005$$
$$P(Z > 2.5758) = 0.005$$

Therefore, $\frac{0.1\sqrt{n}}{\sqrt{2.917}} \ge 2.5758$

$$n \ge \left(\frac{2.5758 \times \sqrt{2.917}}{0.1}\right)^2 = 1935.3553\ldots$$

$\therefore$ minimum sample size is 1936

$X_1, X_2, X_3, \ldots, X_n$ are generated from some distribution with mean μ and variance σ^2 ← Summary of CLThm

If n is large, $\bar{X} \dot\sim N(\mu, \frac{\sigma^2}{n})$

If $X_i \sim N(\mu, \sigma^2)$ where $i = 1, 2, 3, \ldots, n$ (if the original population is normally distributed)

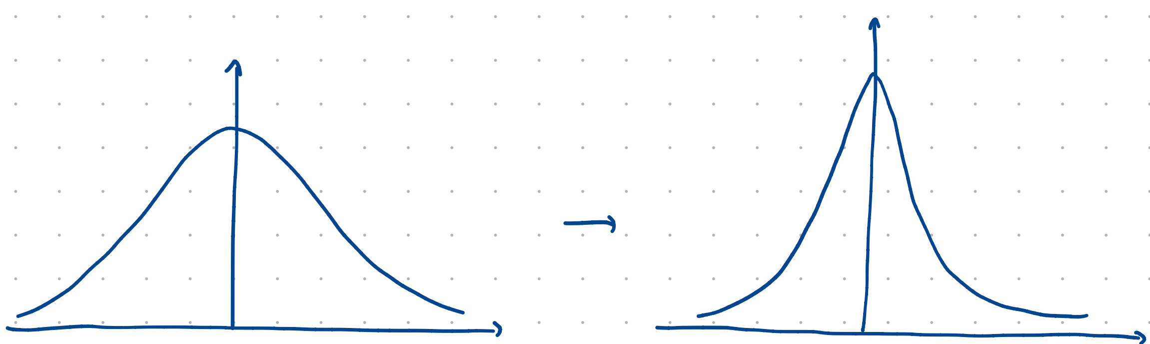
$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = E\left[\frac{1}{n}X_1 + \frac{1}{n}X_2 + \ldots + \frac{1}{n}X_n\right] = \frac{1}{n}E[X_1] + \frac{1}{n}E[X_2] + \ldots + \frac{1}{n}E[X_n]$$

$$= \frac{1}{n}\left(E[X_1] + E[X_2] + \ldots + E[X_n]\right) = \frac{1}{n}(n\mu) = \mu$$

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\left(Var(X_i)\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

Therefore, if the population is normally distributed, the sample mean $\bar{X}$ is distributed perfectly normally (no approximations)

$$\boxed{\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)}$$ ← NOT $\dot\sim$, but actually $\sim$



Data normally distributed → Mean normally distributed