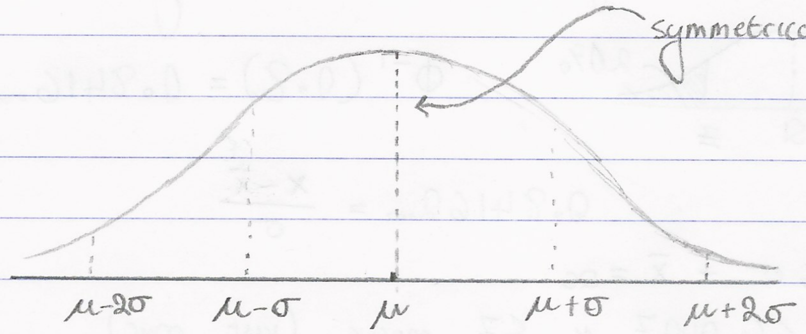
Bell Curve & Normal Distribution & Confidence Intervals

Bell Curve

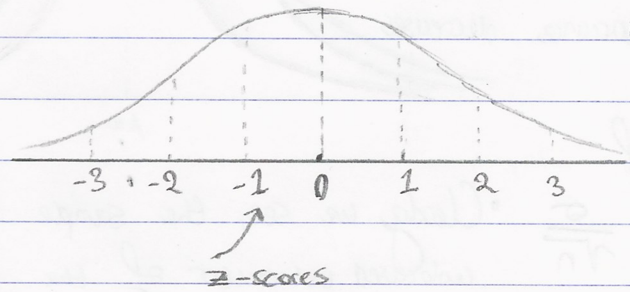
- For normal distribution.



- The majority of data is captured within 3σ's of the mean.

Standardized Normal Distribution

- For normal distribution, $X \sim N(\mu, \sigma^2)$; read as "X is normally distributed with mean μ & standard deviation σ^2."
- The standard normal curve is $Z \sim N(0, 1)$



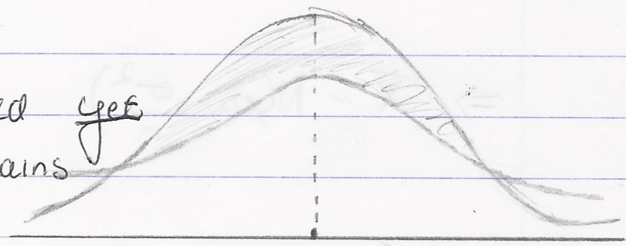
r.v. normal

$\phi(z) = P(Z=z)$

area

$\Rightarrow \phi^{-1}(P(Z=z)) = z$

Doubling σ means greater spread yet reduced height ∵ area remains at 1.



Examples

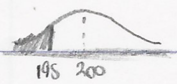
① A machine cuts planks of wood with $\mu = 200$cm & $\sigma = 4$cm. What proportion of planks are under 195cm?

You need to use the standard normal curve to find areas.

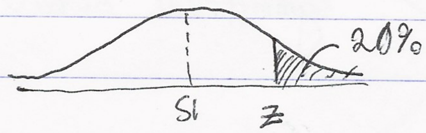
$$z = \frac{X - \bar{X}}{\sigma} = \frac{195-200}{4} = -1.25$$

$$P(X < 195) = P(Z < -1.25)$$

$$\approx 10.6\%$$



② $X \sim N(51, 6^2)$, for which x-value will give the top 20%.



$\Phi^{-1}(0.8) = 0.8416...$

$$0.84162... = \frac{x - \bar{x}}{\sigma}$$

$\Rightarrow 0.8416\sigma + \bar{x} = x$

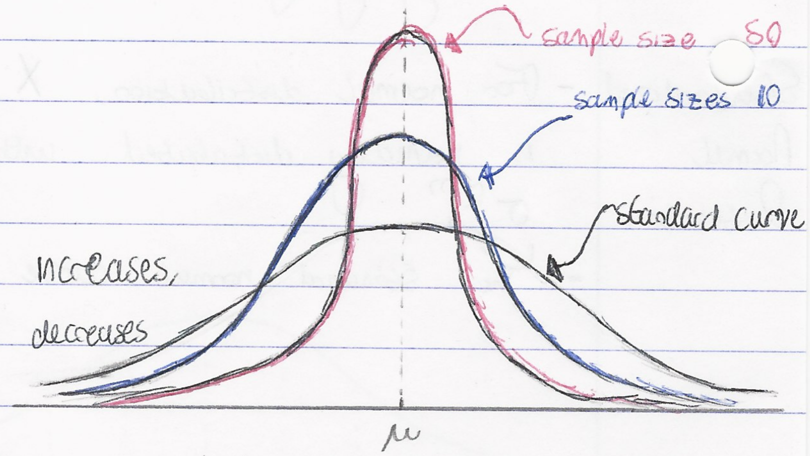
$\Rightarrow x = 56.0497 \approx 57$ marks (just over).

Estimation & Standard Error

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

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

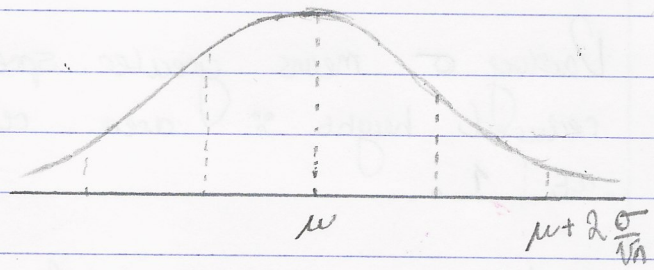
→ as n increases, variance decreases



• The standard deviations of means = standard error = $\frac{\sigma}{\sqrt{n}}$

• Clearly, we see, the sample mean is an unbiased estimator of the population mean.

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



(2 lots of standard error

• The unbiased estimator of the variance is:

Using for finding σ^2 if unknown

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$ unlike $$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$ overestimate

E.g., 40, 35, 37, 42, 48, 39, 38, 39, 31, 42

$s^2 x = 20.544...$ ∴ it is an unbiased estimator of a larger population variance.

Central Limit Theorem

- The larger a sample, the closer it is to a normal distribution (even when sampling from a non-normal distribution):

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \xrightarrow{\delta} N(0, 1) \quad \text{as} \quad n \longrightarrow \infty$$

tends in distribution — as the sample size increases

- The very rough guide for the required sample size is $30 \Rightarrow n \geq 30$

Note

$\bar{x} = \dfrac{\sum x_i}{n}$, this can be easily forgotten.

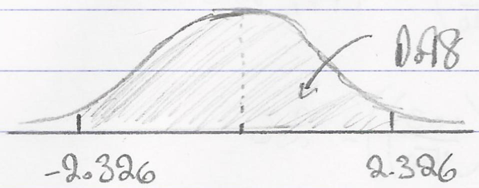
Confidence Intervals

- Based on sample data giving a range of plausible values for a parameter. E.g. 99% confident that x lies in $(2.5, 13.4)$.
- Thus, there is a trade-off between the confidence and width of interval.

Examples

① $X \sim N(\mu, 3.5^2)$ — a sample has values: 116, 152, 128, ...
$\Rightarrow$ unbiased estimator of mean $= 131.1$

Construct a 98% confidence interval.



$\Rightarrow$ By $\bar{x} \pm z\left(\dfrac{\sigma}{\sqrt{n}}\right)$ — no. of standard deviations from mean

$131.1 \pm 2.326 \times \dfrac{3.5}{\sqrt{10}}$

$= [128.53, 133.67]$

② The length of a particular species of snake is normally distributed with μ 28cm & σ 0.6cm. What is the probability that the mean length of a random sample of 8 snakes is > 28.3cm?

$X \sim N(28, 0.6^2)$ & $\bar{X} \sim N\left(28, \frac{0.6^2}{8}\right)$

$$Z\text{-score} = z = \frac{X-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{28.3-28}{\frac{0.6}{\sqrt{8}}} = 1.414$$

Standard error

$\Rightarrow P(Z > 1.414) = 1 - P(Z < 1.414) = 92.1\%$

③ While examining a water company's finances, an auditor selected a random sample of 90 customers, who owed the company money, in order to scrutinise their accounts.

The amounts owed by these 90 customers had a mean of £197 and a standard deviation of £103.

For customers who owed the company money:

(a) Calculate a 95% confidence interval for the mean amount owed.

(b) State the width of the confidence interval that you have calculated in part (a)(i).

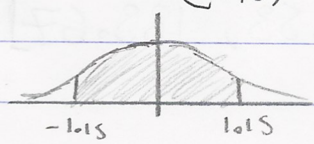
(c) A p% confidence interval based on this sample has a width £25.
Find the value of p.

ⓐ $Z \sim N(0, 1)$. $\phi^{-1}(0.025) = \pm 1.64483$ deviations.

interval $= \bar{x} \pm 1.644853\left(\frac{103}{\sqrt{90}}\right) = 179, 215$

ⓒ $\left(\bar{x} + z\left(\frac{103}{\sqrt{90}}\right)\right) - \left(\bar{x} - z\left(\frac{103}{\sqrt{90}}\right)\right) = 25$

$\Rightarrow 2z\left(\frac{103}{\sqrt{90}}\right) = 25 \Rightarrow z = -1.15,\ 1.15$ deviations



area $= 74.98\% \approx 75\%$