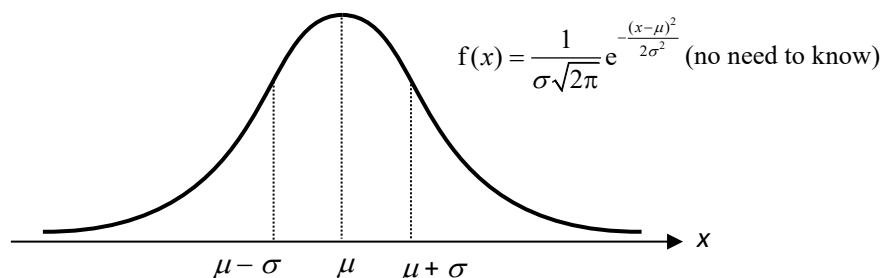


SUMMARY: Normal Distribution

A normal distribution is a good model for continuous random variables such as weight, height, cost, duration, temperature, etc.

If X has a normal distribution with mean μ and variance σ^2 , we write it as $X \sim N(\mu, \sigma^2)$.

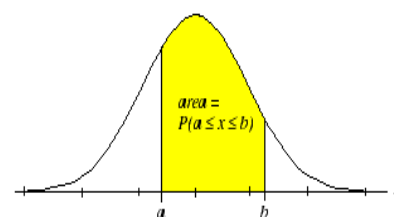
1 Properties of the Probability Density Function



- (i) It is symmetrical about its mean μ . (*This property is used quite often.*)
- (ii) The mean, mode and median coincide at $x = \mu$.
- (iii) The curve lies completely above the x -axis (since $f(x) > 0$ for all real values x) and tapers off rapidly at each “tail” i.e. $f(x) \rightarrow 0^+$ as $x \rightarrow \pm\infty$.
- (iv) Total area under the curve $= \int_{-\infty}^{\infty} f(x) dx = 1$.
- (v) The following probabilities are useful for determining if a random variable can be modelled by a normal distribution.
 - Approximately 68.3% of the data falls within one standard deviation about the mean.
 - Approximately 95.4% of the data falls within two standard deviations about the mean.
 - Approximately 99.7% of the data falls within three standard deviations about the mean.

2 Calculating Probabilities

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$$
$$= \int_a^b f(x) dx = \text{area under the pdf curve between } x = a \text{ and } x = b$$



Probabilities are calculated using GC command **normalcdf** by specifying the lower and upper bounds, (i.e. a, b) and the mean and standard deviation.

3 Finding Unknown x Values Given Probability

When the probability is given, we use the calculator command **invNorm** to find the unknown x value.

Examples

Given that $X \sim N(10, 4)$, find the value of the unknowns given that

- (i) $P(X < a) = 0.2$
- (ii) $P(X > b) = 0.34$
- (iii) $P(|X - 10| < c) = 0.7$

4 Standardisation

Given that $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$. We use standardisation only if either μ and/or σ^2 are unknown or contains unknowns.

Examples

Given that $X \sim N(\mu, \sigma^2)$, $P(X > 5.4) = 0.3$ and $P(X < 2.1) = 0.2$, find μ and σ^2 .

5 Linear Combination of Independent Normal Random Variables

Linear combinations of independent normal variables yield variables that are also normally distributed.

Specifically, if X and Y are any two **independent** variables, and a and b are constants, then $aX \pm bY$ is also normally distributed with the means and variances calculated as follows:

$$\begin{aligned}E(aX \pm bY) &= aE(X) \pm bE(Y) \\ \text{Var}(aX \pm bY) &= a^2 \text{Var}(X) + b^2 \text{Var}(Y)\end{aligned}$$

If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$, and a and b are constants, then

- $aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$
- $aX - bY \sim N(a\mu_1 - b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$

This result can be easily extended to linear combination of any number of independent normal variables.

If $X_1, X_2, \dots, X_n$ are **independent observations** of X where $X \sim N(\mu, \sigma^2)$, then $X_1, X_2, \dots, X_n$ have the same distribution as X . Therefore, the sum of n independent observations of X

$$X_1 + X_2 + \dots + X_n \sim N(n\mu, n\sigma^2).$$

Read and interpret the context carefully to see whether you are taking repeated observations of a variable, or a scalar multiple of a variable.

Examples

Given that the height of boys in cm, $X \sim N(168, 5^2)$ and height of girls in cm, $Y \sim N(162, 3.4^2)$, find the probabilities such that

- (i) the total height of two randomly selected boys exceeds 340 cm,
- (ii) the height of a randomly selected boy and girl differs by more than 5 cm,
- (iii) three times the height of a girl exceeds the total height of three randomly selected boys,
- (iv) the average height of 2 boys and 1 girl exceeds 167 cm.

Solution (focus on formulating and calculating the combined normal variables)

(i) $P(X_1 + X_2 > 340)$ $X_1 + X_2 \sim N(168 + 168, 5^2 + 5^2) = N(336, 50)$

(ii) $P(|X - Y| > 5)$ $X - Y \sim N(168 - 162, 5^2 + 3.4^2) = N(6, 36.56)$

(iii) $P(3Y > X_1 + X_2 + X_3)$

$$3Y - (X_1 + X_2 + X_3) \sim N(3(162) - 3(168), 9(3.4^2) + 3(5^2)) = N(-18, 179.04)$$

(iv) $P\left(\frac{X_1 + X_2 + Y}{3} > 167\right)$ or $P(X_1 + X_2 + Y > 501)$

$$\frac{X_1 + X_2 + Y}{3} \sim N\left(\frac{168 + 168 + 162}{3}, \frac{5^2 + 5^2 + 3.4^2}{3^2}\right) = N(166, 6.84) \quad \text{or}$$

$$X_1 + X_2 + Y \sim N(168 + 168 + 162, 5^2 + 5^2 + 3.4^2) = N(498, 61.56)$$