Chapter **S5**

Normal Distribution

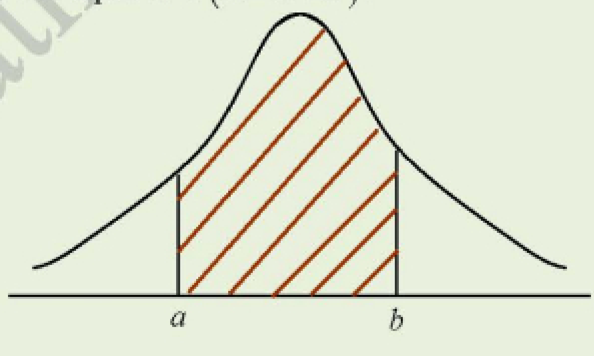
In this chapter, you will learn to

- understand the concept of a normal distribution as an example of a continuous probability model and its mean and variance;
- standardise a normal distribution, and use it to calculate probabilities;
- make use of the symmetry of the normal distribution curve to solve problems;
- solve problems involving the use of E(aX + b) and Var(aX + b);
- solve problems involving two independent normal distributions.

5.1 Continuous Random Variables

A **continuous random variable** is a **random variable** where the data can take infinitely many values. For example, a random variable measuring the time taken for something to be done is continuous since there are infinite number of possible times that can be taken. The random variables may take any value in an interval of the real number line.

For continuous probability distributions, the area under the curve of the **probability density function (p.d.f)** $\mathrm{f}(x)$ between points a and b is equal to $\mathrm{P}(a \le X \le b)$.



Thus $\mathrm{P}(a \le X \le b) = \int_a^b \mathrm{f}(x)\,\mathrm{d}x$ and $\mathrm{P}(X = a) = \int_a^a \mathrm{f}(x)\,\mathrm{d}x = 0$.

Hence $\mathrm{P}(X \le a) = \mathrm{P}(X < a)$

and $\mathrm{P}(a \le X < b) = \mathrm{P}(a < X \le b) = \mathrm{P}(a \le X \le b) = \mathrm{P}(a < X < b)$

For any continuous random variable with probability density function f(x), we have $\int_{\text{all } x} \mathrm{f}(x)\,\mathrm{d}x = 1$.

The **cumulative distribution function** of X is defined as $\mathrm{P}(X \le x)$.

5.1.1 Properties of Expectation and Variance

Recall the following results for mean and variance for discrete random variables with random variable X and Y. They also hold for continuous random variables.

Mean	**Variance**
If a is a constant,	If a is a constant,
(a) $\mathrm{E}(a)=a$	**(a)** $\mathrm{Var}(a)=0$
(b) $\mathrm{E}(aX)=a\,\mathrm{E}(X)$	**(b)** $\mathrm{Var}(aX)=a^2\,\mathrm{Var}(X)$
(c) $\mathrm{E}(X+a)=\mathrm{E}(X)+a$	**(c)** $\mathrm{Var}(X+a)=\mathrm{Var}(X)$
(d) $\mathrm{E}(X-a)=\mathrm{E}(X)-a$	**(d)** $\mathrm{Var}(X-a)=\mathrm{Var}(X)$
	If X and Y are independent, then
(e) $\mathrm{E}(X+Y)=\mathrm{E}(X)+\mathrm{E}(Y)$	**(e)** $\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$
(f) $\mathrm{E}(X-Y)=\mathrm{E}(X)-\mathrm{E}(Y)$	**(f)** $\mathrm{Var}(X-Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$

5.2 Introduction to Normal Distribution

The Normal Distribution is an example of a *continuous* probability model and is one of the most important distributions in statistics. Many statistical problems in real-world situations (e.g. in social sciences) are modelled using the normal distribution. Examples include marks, heights, masses, time and length.

Definition: A continuous random variable X having a probability density function (p.d.f.)

$$\mathrm{f}(x)=\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{for } x\in(-\infty,\infty)\ ,\ \sigma>0$$

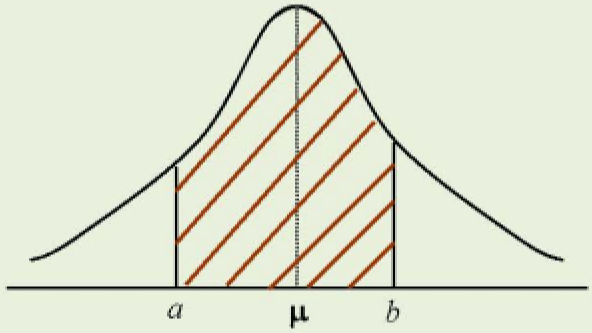
is said to follow a **normal distribution with mean μ and variance σ^2** .

The mean μ and variance σ^2 are the parameters of the distribution, where $\mathrm{E}(X)=\mu$ and $\mathrm{Var}(X)=\sigma^2$.

> If X follows a **normal distribution with mean μ and variance σ^2**, we write
> $X\sim \mathrm{N}(\mu,\sigma^2)$

For $X \sim \mathrm{N}(\mu, \sigma^2)$,

$$\mathrm{P}(a \le X \le b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x$$, which is the shaded area under the p.d.f.



The **normal distribution curve** has the following features:

1. The total area under the curve and above the horizontal axis is equal to 1.
2. The distribution is **symmetrical about** the mean μ.

   Implications: The mode, median and mean are all equal, due to the symmetry of the distribution.

   $$P(X \le \mu) = P(X \ge \mu) = \frac{1}{2}$$

3. The curve is bell-shaped but not every bell-shaped curve is a normal curve. For a normal curve, 68.3%, 95.5% and 99.7% of the values of x are expected to lie within ±1, ±2 and ±3 standard deviations of the mean of X respectively (Figure 1). Thus the normal distribution, while theoretically taking all real values, can be used to model real life random variables such as height which will only take finite range of values. (See Section 5.7)

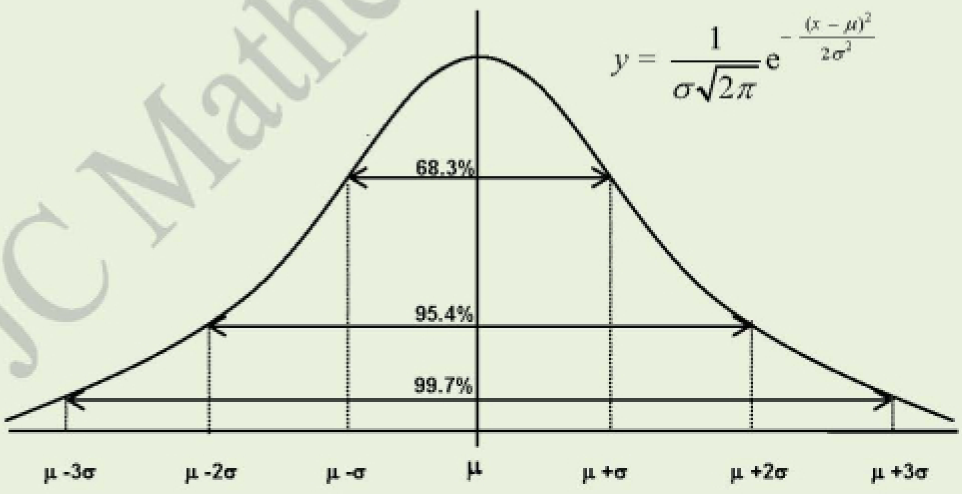
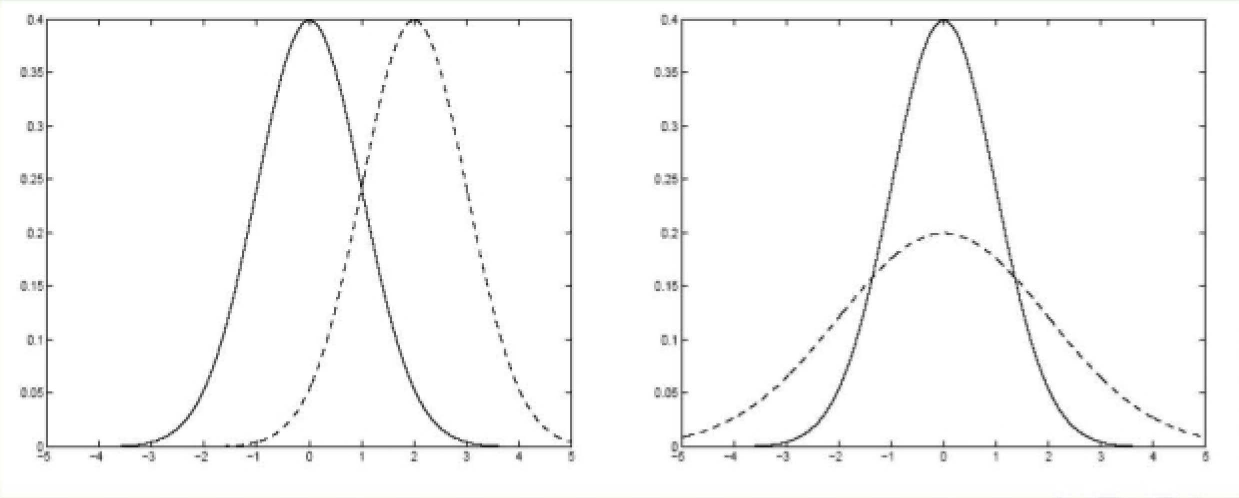


Figure 1: Normal Distribution Curve

4. The location and shape depends on the values of μ and σ respectively.



Same standard deviation (σ =1), different mean (μ =0 for solid, μ =2 for dashed)

Same mean (μ =0) , different standard deviation (σ =1 for solid, σ =2 for dashed)

Figure 2: How the values of μ and σ affect the location and shape of the normal curve

5.3 Finding Probabilities of a Normal Distribution

It is impractical to find the area of the region under graph of $y = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ using integration. As such, we will use the graphing calculator's command "normalcdf(" to provide the probabilities of specified ranges.

When doing so, we may approximate $-\infty$ with -10^{99}, and ∞ with 10^{99}.

Example 1:

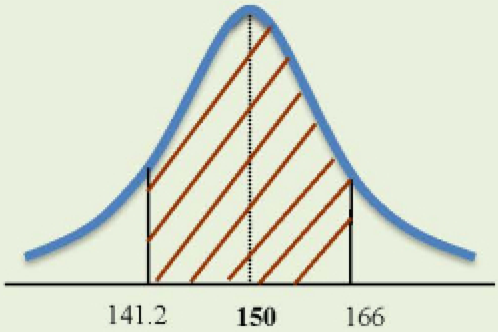
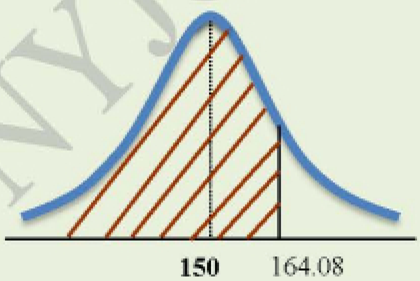
The heights of female NYJCians follow a **normal distribution** with mean 150 cm and standard deviation 22 cm. Find the probability that the height of a randomly chosen female NYJCian is

(i) between 141.2 cm and 166 cm.

(ii) less than 164.08 cm.

(iii) at least 175.5 cm.

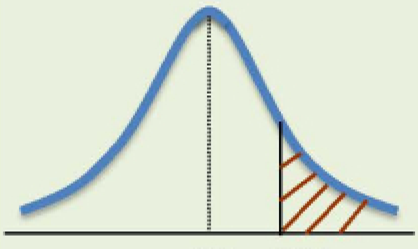
(iv) differs from the mean by more than 10 cm.

Solution:	
(i) Let H be the random variable denoting the height of a randomly chosen female NYJCian. i.e. $H \sim \text{N}(150, 22^2)$ $\text{P}(141.2 < H < 166) \approx 0.42189$ $= 0.422$ (3 s.f.)	Define the random variable and state its distribution with parameters μ and σ^2. **Step 1:** Go to the distribution menu by pressing  **[2ND] [VARS]** (DISTR) and select **2:normalcdf(**
Graphical representation: 	**Step 2:** Key in the values for $\text{P}(141.2 < H < 166)$, i.e. **lower bound: 141.2** **upper bound: 166** **μ: 150** **σ: 22** Select **Paste** and press **ENTER** to evaluate.  For σ, input the value of the **standard deviation**, not the variance.
(ii) $\text{P}(H < 164.08) \approx 0.73891$ $= 0.739$ (3 s.f.) Graphical representation: 	Repeat **Step 1**. **Step 2:** **lower bound: − E99** **upper bound: 164.08** **μ: 150** **σ: 22**  **Note:** Press [2nd] [,] for **E** Select **Paste** and press **ENTER** to evaluate.

(iii) $P(H \geq 175.5) \approx 0.12321$

$= 0.123 \quad (3 \text{ s.f.})$

Graphical representation:



Repeat **Step 1**.

Step 2:

Key in the values for $P(H > 175.5)$, i.e.

lower bound: 175.5

upper bound: E99

μ: 150

σ: 22

Select **Paste** and press **ENTER** to evaluate.

(iv)

$$\begin{aligned} P(|H-\mu|>10) &= P(|H-150|>10) \\ &= P(H-150<-10 \quad \text{or} \quad H-150>10) \\ &= P(H<140 \quad \text{or} \quad H>160) \\ &= P(H<140) + P(H>160) \\ &= 0.649 \end{aligned}$$

Alternative:

$$\begin{aligned} P(|H-\mu|>10) &= P(|H-150|>10) \\ &= 1-P(|H-150|<10) \\ &= 1-P(-10<H-150<10) \\ &= 1-P(140<H<160) \\ &= 0.649 \end{aligned}$$

Example 2

The random variable Y is normally distributed with standard deviation 4, but its mean is unknown. Find the greatest possible value of $P(-3.74 < Y < 5.82)$.

Solution:

Given: $Y \sim N(\mu, 4^2)$

For $P(-3.74 < Y < 5.82)$ to be greatest, the area representing the probability must be symmetrical about the mean μ.

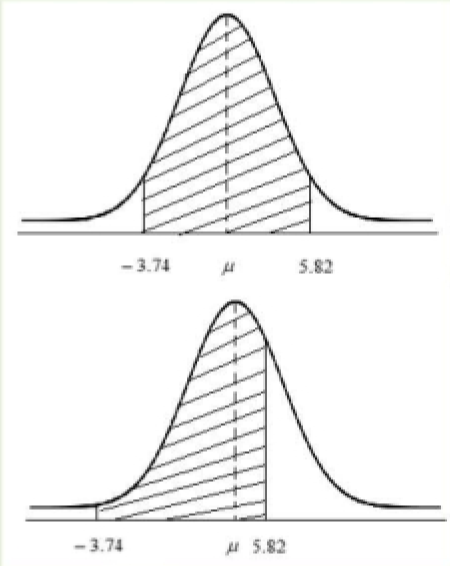
Hence $\mu = \dfrac{-3.74 + 5.82}{2} = 1.04$

$\therefore \; Y \sim N(1.04, 4^2)$

$P(-3.74 < Y < 5.82) = 0.76791$

$= 0.768 \quad (3 \text{ s.f.})$

Which of the following case should it be?



Self-review 1:

Given that $X \sim N(12, 3^2)$, find the probabilities:

(a) $P(X < 12.5)$ **(b)** $P(X > 18.5)$ **(c)** $P(10 < X < 14)$

(d) $P(|X - 4| < 15)$ **(e)** $P(|X - 8| > 3)$

[(a) 0.566 (b) 0.0151 (c) 0.495 (d) 0.990 (e) 0.640]

5.4 Solving Inverse Normal Problems

In certain cases, the distribution $X \sim \mathrm{N}(\mu, \sigma^2)$ is known and you are to find the value of x such that

(a) $\mathrm{P}(X < x) =$ known probability (i.e. area);

(b) $\mathrm{P}(X > x) =$ known probability; and

(c) $\mathrm{P}(-x < X - \mu < x) =$ known probability.

To solve such questions, we will use the Graphing Calculator's function "invNorm".

Example 3:

The heights of male NYJCians are **normally distributed** with mean 169 cm and standard deviation 9cm.

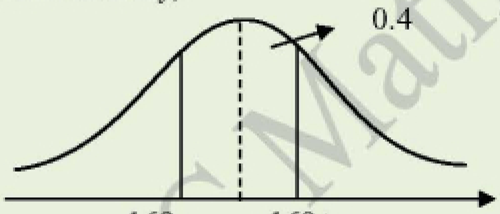
(i) Given that 80% of these male NYJCians have a height less than a cm, find the value of a.

(ii) Given that 60% of these male NYJCians have a height more than b cm, find the value of b.

(iii) Given that 40% of these male NYJCians have a height of within c cm of the mean, find the value of c.

(iv) Given that 20% of these male NYJCians have a height between 160cm and d cm, where $d > 160$, find the value of d.

Solution:	**ThinkZone**
Let X be the random variable denoting the height of a randomly chosen male NYJCian. i.e. $X \sim \mathrm{N}(169, 9^2)$ **(i)** $\mathrm{P}(X < a) = 0.80$ Using GC, $a \approx 176.57 = 177$ (3 s.f.)	**Step 1:** Go to the distribution menu by pressing **[2ND] [VARS]** (DISTR) and select 3:**invNorm(**  **Step 2:** Key in the values for $\mathrm{P}(X < a) = 0.80$, i.e. **area: 0.8** **μ: 169** **σ: 9** **Tail: LEFT** **Paste** and then **ENTER** to evaluate. 

(ii) $P(X > b) = 0.60$ Using GC, $b \approx 166.72 = 167$ (3 s.f.) **Alternatively,** $P(X > b) = 0.60$ $1 - P(X \le b) = 0.60$ $P(X \le b) = 1 - 0.60 = 0.40$ Using GC, $b \approx 166.72 = 167$ (3 s.f.)	Repeat **Step 1** as above. **Step 2:** Key in the values for $P(X > b) = 0.60$, i.e.  **area: 0.6** **μ: 169** **σ: 9** **Tail: RIGHT** **Paste** and then **ENTER** to evaluate. To find x such that $P(X > x) = p$, we can also rewrite the probability as $P(X \le x)$: $P(X > x) = p \Rightarrow 1 - P(X \le x) = p$ $P(X \le x) = 1 - p$
(iii) $P(\|X - 169\| < c) = 0.40$ $P(169 - c < X < 169 + c) = 0.40$ Using GC, $P(164.28 < X < 173.72) = 0.40$ Thus $169 + c = 173.72$. $c = 4.72$ Alternatively,  $P(X < 169 + c) = 0.4 + \frac{1 - 0.4}{2}$ $= 0.7$ Using GC, $169 + c = 173.72$ (or) $P(X < 169 - c) = \frac{1 - 0.4}{2}$ $= 0.3$ Using GC, $169 - c = 164.28$ $\therefore c = 4.72$	Repeat **Step 1** as above. **Step 2:** Key in the values for $P(169 - c < X < 169 + c) = 0.40$ i.e.  **area: 0.4** **μ: 169** **σ: 9** **Tail: CENTER** **Paste** and then **ENTER** to evaluate. 

(iv)	
$P(160 < X < d) = 0.20$ $P(X < d) - P(X < 160) = 0.20$ $P(X < d) = 0.35866$ Using GC, $d \approx 165.74 = 166$ (3 s.f.)	Note: In this case, we cannot use invNorm + center, as we are uncertain if 160 and d are symmetrical about the mean

Note: Given probability p, known parameters μ, and σ^2,

(I) invNorm(p, μ, σ, LEFT) returns the value x such that $P(X < x) = p$, where $X \sim N(\mu, \sigma^2)$.

(II) invNorm(p, μ, σ, RIGHT) returns the value x such that $P(X > x) = p$, where $X \sim N(\mu, \sigma^2)$.

(III) invNorm(p, μ, σ, CENTER) returns the value x such that $P(-x < X - \mu < x) = p$, where $X \sim N(\mu, \sigma^2)$.

To use invNorm + Center, both the lower and upper bound must be symmetrical about the mean.

Self-review 2:

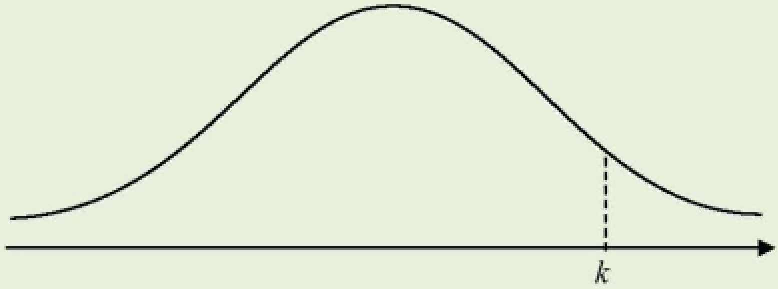
1. Given that $X \sim N(12, 9)$, find the value(s) of k such that

(a) $P(X < k) = 0.55$ **(b)** $P(X > k) = 0.777$ **(c)** $P(6 < X < k) = 0.444$

[(a) 12.4 (b) 9.71 (c) 11.7 5]

5.4.1 Solving For Range of Possible Values

For questions that require a range of x such that $\mathrm{P}(X<x)$ (or $\mathrm{P}(X>x)$) satisfy an inequality, we first solve for the equality case then evaluate the possible values that satisfy the inequality.



$\mathrm{P}(X<k)=p$	$\mathrm{P}(X<x)>p$ $\Rightarrow x>k$	$\mathrm{P}(X<x)<p$ $\Rightarrow x<k$

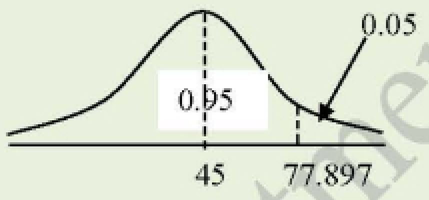
Example 4:
Given that $X \sim \mathrm{N}(12,9)$, find the value(s) of k such that

(a) $\mathrm{P}(X \le k) > 0.222$ **(b)** $\mathrm{P}(X<k) \le 0.6$ **(c)** $\mathrm{P}(X>k)>0.2$ **(d)** $\mathrm{P}(X>k)<0.3$

Consider $\mathrm{P}(X \le k)=0.222$ Using GC, $k=9.7036$ $\therefore k>9.7036$ $k \ge 9.71$	
Consider $\mathrm{P}(X<k)=0.6$ Using GC, $k=12.760$ $\therefore k \le 12.760$ $k \le 12.7$	
Consider $\mathrm{P}(X>k)=0.2$ Using GC, $k=14.525$ $\therefore k<14.525$ $k \le 14.5$	
Consider $\mathrm{P}(X>k)=0.3$ Using GC, $k=13.573$ $\therefore k>13.573$ $k \ge 13.6$	

Example 5:

The marks of 500 candidates in an examination are normally distributed with mean 45 and standard deviation 20. A candidate obtains a distinction by scoring x marks or more. If less than 5 % of the candidates obtain a distinction, find the range of values of x, correct to 1 decimal place.

Solution:	**ThinkZone**
Let M be the random variable denoting the examination marks obtained by a randomly chosen candidate, i.e. $M \sim \mathrm{N}(45, 20^2)$ $\mathrm{P}(M \geq x) < 0.05$ $1 - \mathrm{P}(M \leq x) < 0.05$ $\mathrm{P}(M \leq x) > 0.95$ Let $\mathrm{P}(M \leq k) > 0.95$. Using GC, $k > 77.897$ $\Rightarrow x \geq 77.9$ (to 1 d.p.) Therefore, a candidate will obtain a distinction by scoring 77.9 marks or more.	

Self-review 3:

The mass of a bar of candy is normally distributed with mean 90g and standard deviation 10g.

(i) Find probability that a randomly chosen bar of candy has a mass more than 90.7g.

(ii) The probability that a randomly chosen bar of candy has a mass of less than w g is 0.2. Find the value of w. [0.472, 81.6]

5.5 Standard Normal Distribution

The **standard normal distribution** is a special normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. A random variable following such a distribution is called the **standard normal random variable**. The letter Z is *reserved* to denote the standard normal random variable: **$Z \sim$ N(0, 1)**. The standard normal random variable is also called a **standard score** or a **z-score**.

The p.d.f. of Z is $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$, $z \in (-\infty, \infty)$. The **cumulative distribution function of Z** is defined as $P(Z \le z)$ and is denoted by $\Phi(z)$.

- We can transform a non-standard normal distribution to a standard normal distribution by using the substitution $z = \frac{x-\mu}{\sigma}$, i.e. we can standardise the distribution of X where $X \sim N(\mu, \sigma^2)$ by letting $\boldsymbol{Z = \frac{X-\mu}{\sigma}}$ where $Z \sim N(0,1)$.

For example, consider $X \sim N(100, 10^2)$.

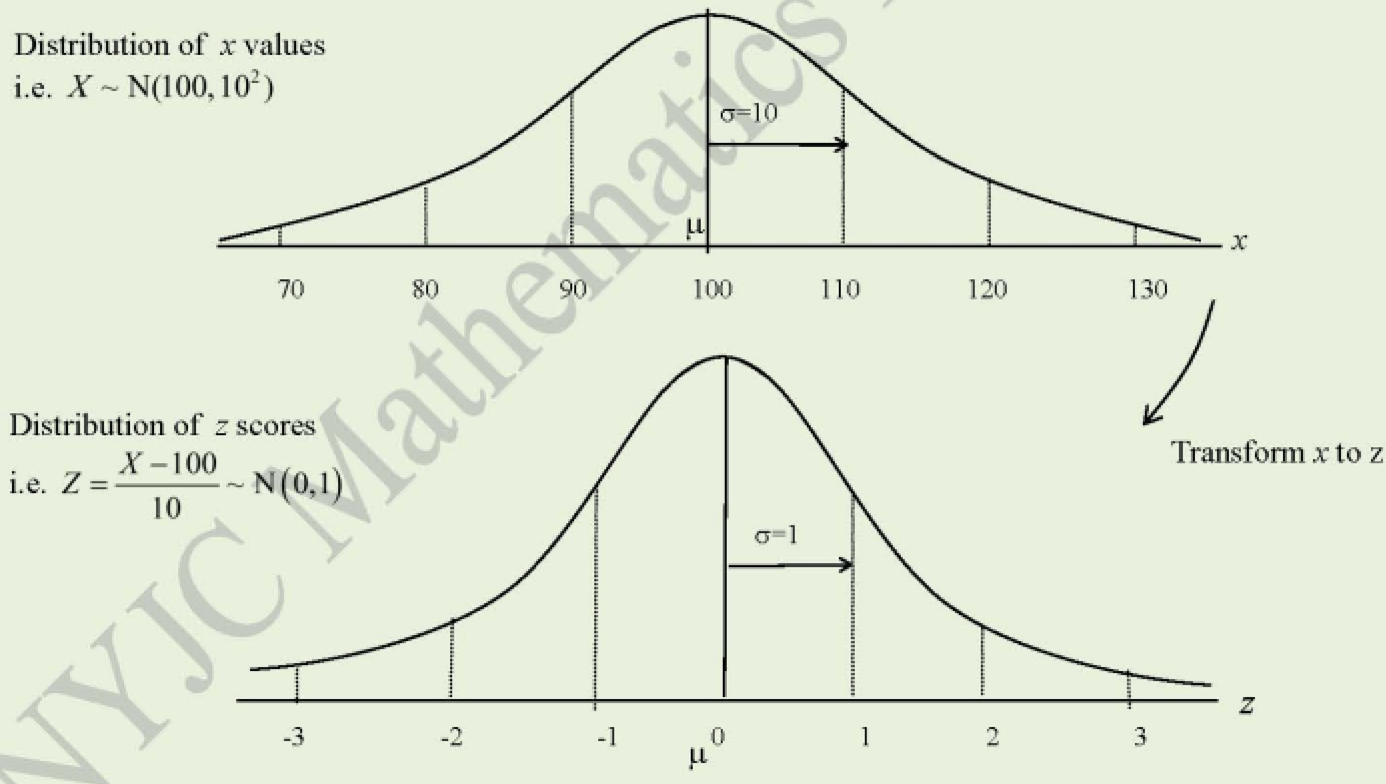


Figure 3: Standardising a Normal Distribution with mean 100 and standard deviation 10

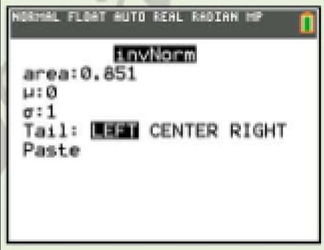
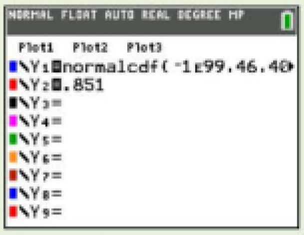
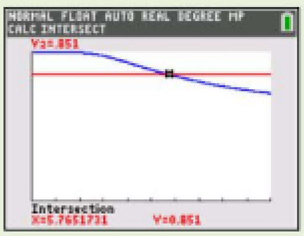
- The area in any normal distribution bounded by some x is the same as the area bounded by the equivalent z-score in the standard normal distribution, i.e. $P(X < x) = P\left(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right) = P\left(Z < \frac{x-\mu}{\sigma}\right)$.

5.5.1 Standardising Normal Distributions for unknown μ or σ

The invNorm(function cannot be used when **μ or σ is unknown**. We need to transform a non-standard normal distribution $X \sim \mathrm{N}(\mu, \sigma^2)$ to a standard normal distribution $Z \sim \mathrm{N}(0, 1)$ through the operation $Z = \dfrac{X - \mu}{\sigma}$ first before using the function.

Example 6:

The length of metal rods produced by a machine has lengths that are normally distributed with mean 40 cm and standard deviation σ cm. A quality control engineer recorded that 85.1 % of the rods have length less than 46cm. Find the value of σ.

Solution:	**ThinkZone**
Let L be the random variable denoting the length of a randomly chosen metal rod i.e. $L \sim \mathrm{N}(40, \sigma^2)$ $\mathrm{P}(L < 46) = 0.851$ $\mathrm{P}\left(\dfrac{L-40}{\sigma} < \dfrac{46-40}{\sigma}\right) = 0.851 \quad \left(\text{standardising } Z = \dfrac{X-\mu}{\sigma}\right)$ $\mathrm{P}\left(Z < \dfrac{6}{\sigma}\right) = 0.851$ Using GC, $\dfrac{6}{\sigma} \approx 1.0407$ (5 s.f.) $\Rightarrow \sigma \approx 5.7654 = 5.77$ (3 s.f.)	 
Alternative Method: We can use a graphical method to solve the equation $\mathrm{P}(L < 46) = 0.851$. We can let σ be the variable and graph the LHS of the equation in GC as show on the right. From the graph, $\sigma = 5.77$ (3 s.f.)	  

Example 7:

The masses of articles produced in a workshop are normally distributed with mean μ and standard deviation σ. 5% of the articles have mass greater than 85g and 10 % have mass less than 25g. Find the value of μ and σ.

Solution:	**ThinkZone**
Let M be the random variable denoting the mass of a randomly chosen article, i.e. $M \sim N(\mu, \sigma^2)$. Given: $P(M > 85) = 0.05$ $\Rightarrow P\left(\frac{M-\mu}{\sigma} > \frac{85-\mu}{\sigma}\right) = 0.05$ $\Rightarrow P\left(Z > \frac{85-\mu}{\sigma}\right) = 0.05$ $\therefore \frac{85-\mu}{\sigma} \approx 1.6449$ $85 - \mu \approx 1.6449\sigma$ .... (1) Given: $P(M < 25) = 0.10$ $\Rightarrow P\left(\frac{M-\mu}{\sigma} < \frac{25-\mu}{\sigma}\right) = 0.10$ $\Rightarrow P\left(Z < \frac{25-\mu}{\sigma}\right) = 0.10$ $\therefore$ $\frac{25-\mu}{\sigma} \approx -1.2816$ $25 - \mu \approx -1.2816\sigma$ .... (2) Solving (1) and (2): mean, $\mu = 51.3$ g (3 s.f.) and standard deviation, $\sigma = 20.5$ g (3 s.f.)	

Self-review 4:

The masses of boxes of oranges are normally distributed with mean μ kg and variance σ^2 kg^2, such that 30 % of them have masses greater than 4 kg and 20% have masses greater than 4.53 kg. Find the mean and standard deviation for the masses of boxes of oranges. [$\mu = 3.13$, $\sigma = 1.67$]

Solution

Let M be the random variable denoting the mass of a randomly chosen box of oranges i.e. $M \sim \text{N}(\mu, \sigma^2)$.

Given: $\text{P}(M > 4) = 0.30$

$$\Rightarrow \text{P}\left(\frac{M-\mu}{\sigma} > \frac{4-\mu}{\sigma}\right) = 0.30$$

$$\Rightarrow \text{P}\left(Z > \frac{4-\mu}{\sigma}\right) = 0.30$$

$$\therefore \frac{4-\mu}{\sigma} \approx 0.52440$$

$$4 - \mu \approx 0.52440\,\sigma \quad \ldots\ldots\ldots\ldots (1)$$

Given: $\text{P}(M > 4.53) = 0.20$

$$\Rightarrow \text{P}\left(\frac{M-\mu}{\sigma} > \frac{4.53-\mu}{\sigma}\right) = 0.20$$

$$\Rightarrow \text{P}\left(Z > \frac{4.53-\mu}{\sigma}\right) = 0.20$$

$$\therefore \frac{4.53-\mu}{\sigma} \approx 0.84162$$

$$4.53 - \mu \approx 0.84162\,\sigma \quad \ldots\ldots\ldots\ldots (2)$$

Solving (1) and (2):

mean, $\mu = 3.13$ kg (3 s.f.) and

standard deviation, $\sigma = 1.67$ kg (3 s.f.)

5.6 Solving Problem Sums

5.6.1 Linear Combinations of normal variables

- If X and Y are two independent normally distributed random variables, i.e. $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$, then
  - $X + Y \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
  - $X - Y \sim \mathrm{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

- For n independent normal variables of X, where $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, 3, \ldots, n$, it is true that the sum of these variables is also normally distributed, i.e.
  $X_1 + X_2 + \ldots + X_n \sim \mathrm{N}(\mu_1 + \mu_2 + \ldots + \mu_n, \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2)$

- For the normal variable such that $X \sim \mathrm{N}(\mu, \sigma^2)$ and for any constant a,
  $aX \sim \mathrm{N}(a\mu, a^2\sigma^2)$

- For two independent normal variables such that $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$ and for any constants a and b,
  - $aX + bY \sim \mathrm{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$
  - $aX - bY \sim \mathrm{N}(a\mu_1 - b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$

Example 8:

The mass of a brand 'NY potato' is normally distributed with mean 50g and standard deviation 6g.

(i) If one potato is chosen at random, find the probability that its mass exceeds 44g.

(ii) If 3 potatoes are chosen at random, find the probability that their total mass will exceed 135g.

(iii) If one potato is chosen at random, find the probability that three times the mass of the potato will exceed 135g.

(iv) Four potatoes are chosen at random. Find the probability that their mean mass is less than 55g.

Solution:

Let M be the random variable denoting the mass of a randomly chosen 'NY potato' i.e. $M \sim \mathrm{N}(50, 6^2)$

(i) $\mathrm{P}(M > 44) = 0.841$ (3 s.f.)

(ii) The required probability is $\mathrm{P}(M_1 + M_2 + M_3 > 135)$

$\mathrm{E}(M_1 + M_2 + M_3) = \mathrm{E}(M_1) + \mathrm{E}(M_2) + \mathrm{E}(M_3) = 3\mathrm{E}(M) = 3(50) = 150$

$\mathrm{Var}(M_1 + M_2 + M_3) = \mathrm{Var}(M_1) + \mathrm{Var}(M_2) + \mathrm{Var}(M_3) = 3\mathrm{Var}(M) = 3(6^2) = 108$

i.e. $M_1 + M_2 + M_3 \sim \mathrm{N}\left(150, \left(\sqrt{108}\right)^2\right)$

$\therefore$ $P(M_1 + M_2 + M_3 > 135) \approx 0.92554 = 0.926$

(iii) The required probability is $P(3M > 135)$

$E(3M) = 3E(M) = 3(50) = 150$

$Var(3M) = 3^2\ Var(M) = 9(6^2) = 324$

i.e. $3M \sim N\left(150, \left(\sqrt{324}\right)^2\right)$

$\therefore$ $P(3M > 135) \approx 0.79767 = 0.798$ (3 s.f.)

Note: Observe that $P(M_1 + M_2 + M_3 > 135)$ is different from $P(3M > 135)$. Why?

(i) $M_1 + M_2 + M_3$ gives the total mass of 3 independent potatoes whereas $3M$ gives 3 times the mass of one potato.

(ii) $Var(M_1 + M_2 + M_3)$ is different from $Var(3M)$.

(iv) The required probability is $P(\bar{M} < 65)$ where $\bar{M} = \frac{M_1 + M_2 + M_3 + M_4}{4}$

$$E(\bar{M}) = E\left(\frac{M_1 + M_2 + M_3 + M_4}{4}\right) = \frac{1}{4}\left[4E(M)\right] = E(M) = 50$$

$$Var(\bar{M}) = Var\left(\frac{M_1 + M_2 + M_3 + M_4}{4}\right) = \left(\frac{1}{4}\right)^2\left[4Var(M)\right] = \frac{1}{4}Var(M) = \frac{1}{4}(6^2) = \frac{36}{4}$$

i.e. $\bar{M} \sim N\left(50, \left(\frac{6}{2}\right)^2\right)$

$\therefore$ $P(\bar{M} < 55) =$ (3 s.f.)

Example 9:

The masses of Nanyang JC boys are normally distributed with mean 60kg and standard deviation 3kg. The masses of Nanyang JC girls are also normally distributed, with mean 50kg and standard deviation 4kg.

(i) Find the probability that the mass of a boy is between 56kg to 62 kg.

(ii) Three boys are chosen. Find the probability that one of them is between 56 kg to 62kg while the other two boys is less than 56kg.

(iii) Find the probability that the total mass of 2 randomly chosen boys exceeds twice the mass of a girl.

(iv) Find the probability that the difference in mass between a randomly chosen Nanyang JC boy and a randomly chosen girl in this age group is more than 3kg.

Solution:

Let B and G be the random variable denoting the mass of a randomly chosen Nanyang JC boy and girl respectively i.e. $B \sim N(60, 3^2)$

$G \sim N(50, 4^2)$

(i) required probability is $P(56 < B < 62) = 0.656$

(ii) The required probability is $= P(56 < B < 62)\left[P(B < 56)\right]^2 \times \frac{3!}{2!}$

(iii) The required probability is $= P(B_1 + B_2 > 2G) = P(B_1 + B_2 - 2G > 0)$

$E(B_1 + B_2 - 2G) = E(B) + E(B) - 2E(G) = 60 + 60 - 2(50) = 20$

$Var(B_1 + B_2 - 2G) = Var(B) + Var(B) + 4Var(G) = 3^2 + 3^2 + 4(4^2) = 82$

$B_1 + B_2 - 2G \sim N(20, 82)$

$\therefore$ The required probability is $= P(B_1 + B_2 > 2G) = P(B_1 + B_2 - 2G > 0) = 0.98640$

$= 0.986$ (3 s.f.)

(iv) The required probability is $P(|B - G| > 3)$

$E(B - G) = E(B) - E(G) = 60 - 50 = 10$

$Var(B - G) = Var(B) + Var(G) = 3^2 + 4^2 = 25$

i.e. $B - G \sim N(10, 5^2)$

$\therefore P(|B - G| > 3) = 1 - P(-3 < B - G < 3)$

$= 0.92390 = 0.924$ (3 s.f.)

The probability that the *difference* in mass between a Nanyang JC boy and a girl in this age group is more than 3kg is 0.924.

Example 10

Watermelons are sold by weight at a price of \$2.00 per kilogram. The masses of watermelons are normally distributed with mean 1.2 kg and standard deviation 0.3 kg. Papaya are sold by weight at a price of \$1.70 per kilogram. The masses of papaya are normally distributed with mean 1.0 kg and standard deviation 0.2 kg.

(i) Find the probability that the total price of 5 randomly chosen watermelons and 3 randomly chosen papayas exceeds \$18.

(ii) In another purchase, 3 randomly chosen papayas are packed in a plastic bag. Each plastic bag costs \$1. Find the probability that the total price of this purchase is less than \$6.

Solution:

Let W be the random variable denoting the mass of a randomly chosen watermelon i.e. $W \sim N(1.2, 0.3^2)$

Let P be the random variable denoting the mass of a randomly chosen papaya i.e. $P \sim N(1.0, 0.2^2)$

(i) Let T be the random variable denoting the total price of 5 watermelons and 3 papayas.

i.e. $T = 2.0(W_1 + W_2 + W_3 + W_4 + W_5) + 1.7(P_1 + P_2 + P_3)$

The required probability is $P(T > 18)$

$E(T) = 2.0E(W_1 + W_2 + W_3 + W_4 + W_5) + 1.7E(P_1 + P_2 + P_3)$

$= 2.0[5E(W)] + 1.7[3E(P)] = 2.0(5)(1.2) + 1.7(3)(1.0) = 17.1$

$\text{Var}(T) = 2.0^2\ \text{Var}(W_1 + W_2 + W_3 + W_4 + W_5) + 1.7^2\ \text{Var}(P_1 + P_2 + P_3)$

$= 2.0^2\ [5\text{Var}(W)] + 1.7^2\ [3\text{Var}(P)] = 2.0^2(5)(0.3^2) + 1.7^2(3)(0.2^2) = 2.1468$

i.e. $T \sim N\left(17.1, \left(\sqrt{2.1468}\right)^2\right)$

$\therefore\ P(T > 18) \approx 0.26952 = 0.270$ (3 s.f.)

The probability that the total price of 5 randomly chosen watermelons and 3 randomly chosen papayas exceeds $18 is 0.270.

(ii) Let X be the random variable denoting the total price 3 papayas and a plastic bag.

i.e. $X = 1.7(P_1 + P_2 + P_3) + 1$

The required probability is $P(X < 6)$

$E(X) = 1.7E(P_1 + P_2 + P_3) + 1$

$= 1.7[3E(P)] + 1$

$= 1.7(3)(1.0) + 1$

$= 6.1$

$\text{Var}(X) = 1.7^2\ \text{Var}(P_1 + P_2 + P_3)$

$= 1.7^2\ [3\text{Var}(P)]$

$= 1.7^2(3)(0.2^2)$

$= 0.3468$

i.e. $X \sim N\left(6.1, \left(\sqrt{0.3468}\right)^2\right)$

$\therefore\ P(X < 6) \approx 0.43258 = 0.433$ (3 s.f.)

The probability that the total price of the purchase is less than $6 is 0.433.

Example 11

The times taken by two runners, Jack and John, to run 400 meters races are independent and normally distributed with means 45.0 and 46.2 seconds and standard deviations 0.5 and 2.0 seconds respectively. The two runners are to compete in a 400 meter race for which there is a track record of 44.0 seconds.

(i) Calculate the probability of Jack breaking the track record. Hence, show that the probability of John breaking the record is greater than that of Jack.

(ii) Find the probability of Jack beating John.

(iii) Find the probability that the total sum of four randomly chosen times of Jack is more than four times a randomly chosen time of John.

Solution:

Let X be the random variable denoting the time taken by Jack to run a 400 meters race i.e.
$X \sim \mathrm{N}(45.0,\ 0.5^2)$

Let Y be the random variable denoting the time taken by John to run a 400 meters race i.e.
$Y \sim \mathrm{N}(46.2,\ 2.0^2)$

(i) P(Jack breaking the record) = $\mathrm{P}(X < 44.0) = 0.022750 = 0.0228$ (3 s.f.)

P(John breaking the record) = $\mathrm{P}(Y < 44.0) = 0.13567 = 0.136$ (3 s.f.)

$\therefore$ The probability that John breaking the record is higher than Jack.

(ii) The required probability is $\mathrm{P}(X < Y) = \mathrm{P}(X - Y < 0)$

> What is the condition on the random variables for Jack to beat John?

$\mathrm{E}(X - Y) = \mathrm{E}(X) - \mathrm{E}(Y) = 45.0 - 46.2 = -1.2$

$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 0.5^2 + 2.0^2 = 4.25$

i.e. $X - Y \sim \mathrm{N}\left(-1.2,\ \left(\sqrt{4.25}\right)^2\right)$

$\therefore\ \mathrm{P}(X < Y) = \mathrm{P}(X - Y < 0) \approx 0.71975 = 0.720$ (3 s.f.)

(iii) The required probability is $\mathrm{P}(X_1 + X_2 + X_3 + X_4 > 4Y) = \mathrm{P}(X_1 + X_2 + X_3 + X_4 - 4Y > 0)$
$= \mathrm{P}(T > 0)$

where $T = X_1 + X_2 + X_3 + X_4 - 4Y$

$\mathrm{E}(T) = \mathrm{E}(X_1 + X_2 + X_3 + X_4 - 4Y) = 4\mathrm{E}(X) - 4\mathrm{E}(Y)$
$= 4(45) - 4(46.2) = -4.8$

$\mathrm{Var}(T) = \mathrm{Var}(X_1 + X_2 + X_3 + X_4 - 4Y) = 4\mathrm{Var}(X) + 4^2\,\mathrm{Var}(Y)$
$= 4(0.5^2) + 16(2.0^2) = 65$

i.e. $T \sim \mathrm{N}\left(-4.8,\ \left(\sqrt{65}\right)^2\right)$

$\therefore\ \mathrm{P}(T > 0) \approx 0.27580 = 0.276$ (3 s.f.)

Self-review 5:

1. If $X \sim \text{N}(15, 4)$ and $Y \sim \text{N}(10, 2^2)$, find

(a) $\text{P}(X < 12)$ **(b)** $\text{P}(Y < X)$ **(c)** $\text{P}(4X + 5Y > 90)$

[(a) 0.0668 (b) 0.961 (c) 0.941]

2. In a cafeteria, baked beans are served either in ordinary portions or in children's portions. The quantity given for an ordinary portion is a normal variable with mean 90g and standard deviation 3g and the quantity given for a child's portion is a normal variable with mean 43g and standard deviation 2g. What is the probability that Paul, who has two children's portion, is given more than his father, who has an ordinary portion? [0.166]

3. Two firms, A and B manufacture similar components with a mean breaking strength of 6 kN and 5.5 kN and standard deviations of 0.4 kN and 0.2 kN respectively. If both distributions are normal and random samples of 100 components from manufacturer A and of 50 from B are tested, find

(i) the probability that the mean breaking strength of the components from manufacturer A will be between 0.45 kN and 0.55 kN more than the mean of those from manufacturer B.

(ii) the probability that the difference in breaking strength between a randomly chosen component from manufacturer A and a component from manufacturer B is more than 0.05kN

[0.693, 0.952]

5.7 Appropriateness of Normal Model

Example 12 [UCLES Specimen Paper/II/Qn5]

A school has a large number of girls. For each year-group it may be assumed that the heights of girls are normally distributed, with average heights and standard deviations as given in the following table.

Year-group	Average height	Standard deviation
11-year old girls	1.1 m	0.2m
16-year old girls	1.7 m	0.3 m

(i) For each of the following cases, state, with a reason, whether or not a normal model is likely to be appropriate.

(a) The number of days since her last birthday of a randomly chosen 11-year old girl from the school. **[1]**

(b) The height of a pupil chosen at random from the combined group of 11-year and 16-year old girls in the school. **[2]**

(c) The average height of a randomly chosen sample of eight 11-year old girls from the school. **[2]**

(ii) One 11-year old girl and one 16-year old girl are chosen at random from the school. Find the probability that the 16-year old girl is at least 0.5m taller than the 11-year old girl. **[3]**

Solution:

(i)(a) Number of days is discrete and not continuous, therefore the model is not normal.

(b) As there are two distributions involved here, the combined distribution is bimodal. Hence the model is not normal.

(c) Since the random sample is from a normal distribution, the average height is a normal random variable. A normal model is appropriate.

(ii) Let X be the random variable denoting the height of a randomly chosen 11-year old girl i.e.

$X \sim \mathrm{N}\left(1.1,\ 0.2^2\right)$

Let Y be the random variable denoting the height of a randomly chosen 16-year old girl i.e.

$Y \sim \mathrm{N}\left(1.7,\ 0.3^2\right)$

The required probability is $\mathrm{P}(Y - X \geq 0.5)$

$\mathrm{E}\left(Y - X\right) = \mathrm{E}\left(Y\right) - \mathrm{E}\left(X\right) = 1.7 - 1.1 = 0.6$

$\mathrm{Var}\left(Y - X\right) = \mathrm{Var}\left(Y\right) + \mathrm{Var}\left(X\right) = 0.3^2 + 0.2^2 = 0.13$

i.e. $Y - X \sim \mathrm{N}\left(0.6, \left(\sqrt{0.13}\right)^2\right)$

$\therefore\ \mathrm{P}(Y - X \geq 0.5) \approx 0.60924 = 0.609$ (3 s.f.)

Example 13 [H1/N2008/1/7]

An examination is marked out of 100. It is taken by a large number of candidates. The mean mark, for all candidates, is 72.1, and the standard deviation is 15.2. Give a reason why a normal distribution with this mean and standard deviation, would not give a good approximation to the distribution of marks.

Solution:

Let X be the random variable denoting the mark obtained by a randomly chosen candidate.

If $X \sim \mathrm{N}(72.1, 15.2^2)$, $\mathrm{P}(X > 100) = 0.0332$.

There is approximately 3% of the population whose mark is greater than 100, which is not possible. Therefore a normal distribution will not provide a good approximation to the distribution of the mark.

Alternatively,

If $X \sim \mathrm{N}(72.1, 15.2^2)$, $\mathrm{P}(\mu - 3\sigma < X < \mu + 3\sigma) = 0.997$

But $\mu + 3\sigma = 72.1 + 3(15.2) = 117.7 > 100$, which is not possible.

Therefore a normal distribution will not provide a good approximation to the distribution of the mark.

5.8 Miscellaneous Examples

Example 14

A fruit grower grows both red and green apples which have masses that are normally distributed. The mass of a randomly chosen red apple has mean 75g and standard deviation 12.5g. The mass of a randomly chosen green apple has mean of 55g and standard deviation 10.5g.

(i) Find the probability that the total mass of 3 randomly chosen green apples exceeds twice the mean mass of 3 randomly chosen red apples.

(ii) A red apple is considered "underweight" if it weighs less than 70g. Red apples are packed into bags of 10 for transportation to a supermarket. A bag is considered to have passed the quality test if it contains less than 2 "underweight" apples. Calculate the probability that a randomly chosen bag of red apples fail the quality test.

Solution:

Let R be the random variable denoting the mass of a randomly chosen red apple i.e. $R \sim \mathrm{N}(75, 12.5^2)$

Let G be the random variable denoting the mass of a randomly chosen green apple i.e. $G \sim \mathrm{N}(55, 10.5^2)$

(i) Let $\overline{R}$ denotes the mean mass of 3 randomly chosen red apples i.e. $\overline{R} = \frac{R_1 + R_2 + R_3}{3}$

The required probability is $\text{P}\left(G_1 + G_2 + G_3 > 2\overline{R}\right) = \text{P}\left(G_1 + G_2 + G_3 - 2\overline{R} > 0\right)$

$$\text{E}(\overline{R}) = \text{E}\left[\frac{1}{3}(R_1 + R_2 + R_3)\right] = \frac{1}{3}\left[3\text{E}(R)\right] = \text{E}(R) = 75$$

$$\text{Var}(\overline{R}) = \text{Var}\left[\frac{1}{3}(R_1 + R_2 + R_3)\right] = \left(\frac{1}{3}\right)^2\left[3\text{Var}(R)\right] = \frac{1}{3}\left(12.5^2\right) = \frac{625}{12}$$

i.e. $\overline{R} \sim \text{N}\left(75,\ \left(\sqrt{\frac{625}{12}}\right)^2\right)$

$$\text{E}\left(G_1 + G_2 + G_3 - 2\overline{R}\right) = 3E(G) - 2E\left(\overline{R}\right) = 3(55) - 2(75) = 15$$

$$\text{Var}\left(G_1 + G_2 + G_3 - 2\overline{R}\right) = 3\text{Var}(G) + 2^2\,\text{Var}\left(\overline{R}\right) = 3\left(10.5^2\right) + 2^2\left(\frac{625}{12}\right) = \frac{6469}{12}$$

i.e. $G_1 + G_2 + G_3 - 2\overline{R} \sim \text{N}\left(15,\ \left(\sqrt{\frac{6469}{12}}\right)^2\right)$

$\therefore\ \text{P}\left(G_1 + G_2 + G_3 - 2\overline{R} > 0\right) \approx 0.74088 = 0.741$ (3 s.f.)

(ii) P(a red apple is considered "underweight") = $\text{P}(R < 70) = 0.34458$ (5 s.f.)

Let X denote the number of red apples out of 10 which are underweight.

i.e. $X \sim \text{B}(10,\ 0.34458)$

Given: P(a bag of red apples passed the quality test) = $\text{P}(X < 2)$

$\therefore$ P(a bag of red apples fails the quality test) = $\text{P}\left(X \geq 2\right)$

$= 1 - \text{P}(X \leq 1) \approx 0.90846 = 0.908$ (3 s.f.)