Chapter 4 (Statistics) Normal Distribution (Teacher's copy)

Objectives

At the end of the chapter, students should be able to:

(a) Understand the nature of the normal distribution which include
- a normal distribution is used to model a continuous random variable
- use normal curves to illustrate normal distributions with different μ and σ
- the shape and location of a normal curve are determined by the values of μ and σ
- the area under a normal curve between $x = a$ and $x = b$ is the probability $\mathrm{P}(a \le x \le b)$
- the total area under a normal curve is 1

(b) Understand X and Y have independent normal distributions, then $aX + bY$ has a normal distribution with mean $a\mu_x + b\mu_y$ and variance $a^2\sigma_x^2 + b^2\sigma_y^2$

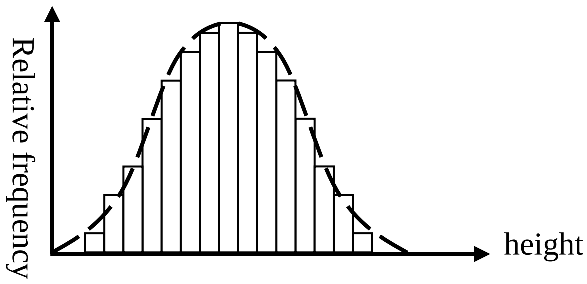
Contents

4.1 Introduction to Normal Distribution

If you sketch a histogram corresponding to the heights of 17 years old male students in a SAJC, what would you observe?

You will realized that majority of the students have about the same height with a few students who are taller than the average height and another few who are shorter than the rest. This is logical since there are fewer very short or very tall boys, and even fewer very very short or very very tall boys.

The outline of the histogram generally shapes like a **bell-shaped curve**.



Sample size = 400

Height	Frequency	Relative Frequency
⋮	⋮	⋮
1.650 – 1.675	112	112/400 = 0.28
1.675 – 1.700	128	128/400 = 0.32
1.700 – 1.725	116	116/400 = 0.29
⋮	⋮	⋮

In fact, such a ‘pattern’ of distribution can be observed for many other physical phenomena as well, such as weight of watermelons in a particular market, I.Q. scores of a population, time taken by students to run 100m, examination results of students in SAJC, household income of Singaporeans etc. Since it is so common, we shall name the distribution a **normal distribution (**and the bell-shape curve **a normal curve**) and study it with some depth.

The normal variable is an example of a continuous random variable.

Note:

$$\text{Relative Frequency} = \frac{\text{Frequency}}{\text{Total number in the sample}}$$

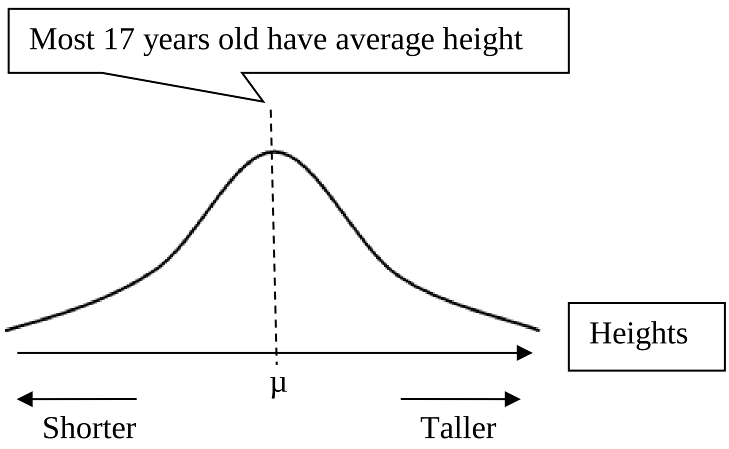
4.2 Continuous Random Variable

Recall from Chapter 3, a continuous random variable (CRV) is a random variable that takes values in an interval. This differentiates from discrete random variable which can only takes a finite or countable infinite set of exact values. In theory, CRVs can be measured to any desired degree of accuracy.

Examples of CRVs are:

- The *height* of a randomly chosen 18-year-old student.
- The actual *mass* of a packet of 1 kg rice.
- The *lifetime* of a new battery.
- The *temperature* of a student in SAJC on a particular day
- The waiting *time* of bus number "142" at the bus-stop outside SAJC on a Monday morning.

4.3 Normal Distribution Curve

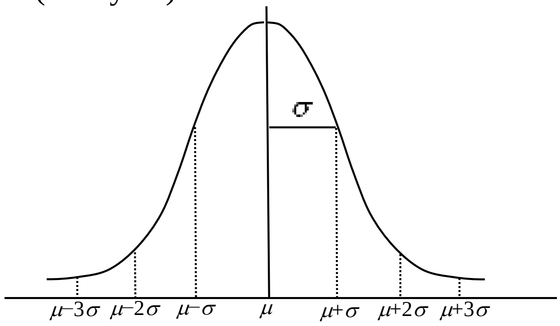


Normal distribution curve of 17 years old students in SAJC

4.3.1 Properties of the Normal Distribution Curve

1. It is bell-shaped and symmetrical about the mean μ.
2. It extends from $-\infty$ to $+\infty$.
3. The total area under the curve is 1.
4. Mean = Median = Mode = μ.
5. The shape of any Normal curve is determined completely by its mean μ and standard deviation σ. The mean gives the position of the axis of symmetry and standard deviation gives the spread. The smaller the standard deviation, the greater the clustering of values near the mean. Hence the curve will have a higher peak at the mean.
6. A continuous random variable X which follows a normal distribution with mean μ and variance σ^2 can be denoted as $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$

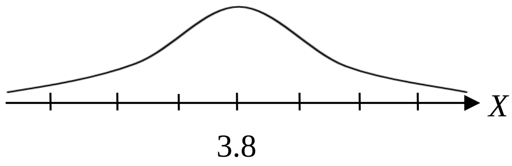
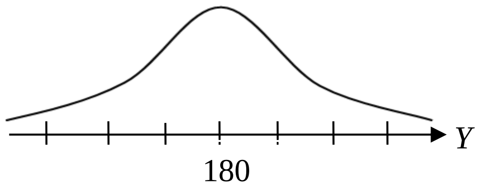
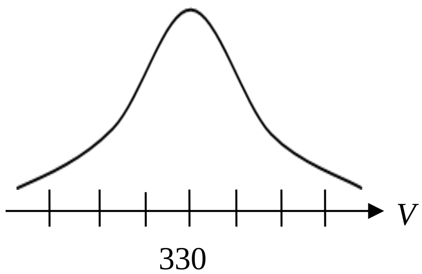
7. For all normal curves
   a. Approx. 68% of the distribution lies within 1 standard deviations of the mean.
   b. Approx. 95% of the distribution lies within 2 standard deviations of the mean.
   c. Approx. 99.7% (nearly all) of the distribution lies within 3 standard deviations of the mean.



Note: The above properties help us to determine whether the distribution of a random variable follows the normal distribution. Another method of checking whether a distribution follows a normal distribution is to use a graphical test, a common way to do so would be to plot the frequency distribution of the random variable. A normal curve would then be overlaid on the frequency distribution, so that we can see whether the normal curve is a good fit.

<u>Example 1</u>
View the video https://www.youtube.com/watch?v=Wqw9cLRMPL0 (up to 07:50).

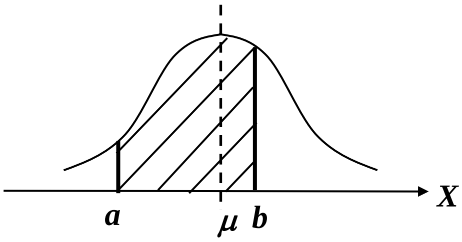
Fill in the blanks. On the Normal curves, put in the markings on the horizontal axis and shade a region corresponding to (a) 68% of the distribution X, (b) 99.7% of the distribution Y, (c) 95% of the distribution V.

	Random Variable	Normal Distribution Curve
(a)	Let X be the r.v. time taken in hours to complete a marathon by a group of males. $\mu = 3.8, \quad \sigma^2 = 0.5^2$ $X \sim \mathrm{N}\left(3.8, 0.5^2\right)$	
(b)	Let X be the r.v. height of a class of 18 year old students, measured in cm. $\mu = 180, \quad \sigma^2 = 10^2$ $Y \sim \mathrm{N}\left(\ 180\ ,\ \ 10^2\ \right)$	
(c)	Let V be the r.v. volume in ml of a particular can of drink. $\mu = 330, \quad \sigma^2 = 4$ $V \sim \mathrm{N}\left(\ 330\ ,\ 2^2\ \right)$	

4.4 Finding Probabilities

Given that $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$, the probability that X lies between a and b is $\mathrm{P}\left(a < X < b\right)$.

$\mathrm{P}\left(a < X < b\right)$ = Area under the normal curve between $x = a$ and $x = b$



Note:

1. Probability that X takes a specific value is zero, i.e. $\mathrm{P}\left(X = c\right) = 0$

   As a result, $\mathrm{P}\left(X \le c\right) = \mathrm{P}\left(X < c\right)$, $\mathrm{P}\left(X \ge c\right) = \mathrm{P}\left(X > c\right)$.

2. $\mathrm{P}\left(a \le X \le b\right) = \mathrm{P}\left(a \le X < b\right) = \mathrm{P}\left(a < X \le b\right) = \mathrm{P}\left(a < X < b\right)$

3. $\mathrm{P}\left(X \le \mu\right) = \mathrm{P}\left(X \ge \mu\right) = 0.5$

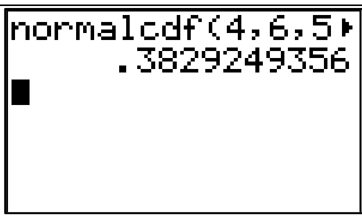
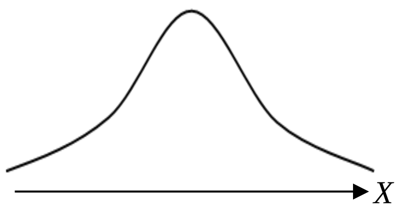
Example 2 (GC: Normalcdf)

Given X follows a normal distribution with mean 5 and standard deviation 2, i.e. $X \sim \mathrm{N}\left(5, 2^2\right)$, calculate

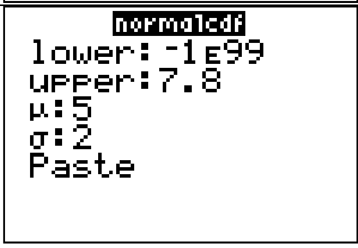
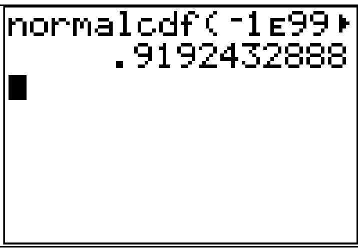
(i) $\mathrm{P}(4 < X \le 6)$; (ii) $\mathrm{P}(X < 7.8)$; (iii) $\mathrm{P}(X > 3)$.

Solution:

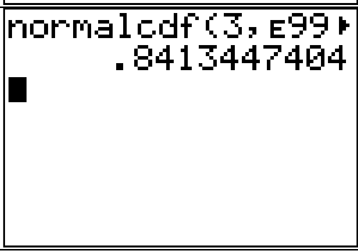
(i) $\mathrm{P}\left(4 < X \le 6\right) = 3.83$

Steps	Screenshot	Remarks
Press [2nd][VARS][2] You should see this screen.		Note the heading is 'normalcdf' If the lower limit, μ and σ are not specified, the default values used by the GC are -1E99, 0 and 1 respectively.
Enter the values of : Lower bound (lower), Upper bound (upper), Mean (μ) and Standard deviation (σ) Press ENTER, ENTER.		
		

(ii) $\mathrm{P}(X < 7.8) = 0.919$

Steps	Screenshot	Remarks
Press $\boxed{\text{2nd}}\boxed{\text{VARS}}\boxed{2}$ You should see this screen.		
Enter the values of : Upper bound (upper), Mean (μ) and Standard deviation (σ) Press ENTER, ENTER.		To find $\mathrm{P}(X < a) = \mathrm{P}(-\infty < X < a)$, use $-$E99 to represent $-\infty$, by pressing $\boxed{(-)}\boxed{\text{2nd}}\boxed{,}\boxed{9}\boxed{9}$ We do not need to change the default lower limit in the GC for part (ii)
Read off the answer		

(iii) $\mathrm{P}(X > 3) = 0.841$

Steps	Screenshot	Remarks
Press $\boxed{\text{2nd}}\boxed{\text{VARS}}\boxed{2}$ You should see this screen.		
Enter the values of : Lower bound (lower), Upper bound (upper), Mean (μ) and Standard deviation (σ) Press ENTER, ENTER.		To find $\mathrm{P}(X > a) = \mathrm{P}(a < X < \infty)$, use E99 to represent $+\infty$.
		

<u>Example 3</u>

The weight of a randomly chosen male student from SAJC is normally distributed with mean 68 kg and standard deviation 3 kg. A male student is randomly chosen. Find the probability that his weight

(i) exceeds 72 kg,
(ii) is at most 70 kg,
(iii) is within one standard deviation from the mean weight.

<u>Solution</u>:

Let X be the random variable that denotes the weight of a randomly chosen male student in SAJC in kg.

$$X \sim \mathrm{N}\left(68, 3^2\right)$$

(i) $\mathrm{P}(X > 72) = 0.0912$

(ii) $\mathrm{P}(X \le 70) = 0.748$

(iii) $\mathrm{P}(|X - 68| < 3) = \mathrm{P}(68 - 3 < X < 68 + 3) = 0.683$

Note:
$\|x\| > b \Leftrightarrow x < -b$ or $x > b$
$\|x\| < b \Leftrightarrow -b < x < b$

<u>Exercise 1</u>

1. If $G \sim \mathrm{N}(100, 80)$, find: (a) $\mathrm{P}(85 < G < 112)$ (b) $\mathrm{P}\left(-\sqrt{80} < G - 100 < \sqrt{80}\right)$

(a)	$\mathrm{P}(85 < G < 112) = 0.863$
(b)	$\mathrm{P}\left(-\sqrt{80} < G - 100 < \sqrt{80}\right) = \mathrm{P}\left(100 - \sqrt{80} < G < 100 + \sqrt{80}\right) = 0.683$

2. The researcher goes out onto the streets and records down the heights of men he met. He found out that the heights of the men are normally distributed with mean 185cm and standard deviation 2 cm. What is the probability that the first man she meets is

(i) between 1.80m to 1.86m tall
(ii) not more than 1.88m tall
(iii) within one standard deviation of the mean height?

If three men are picked at random, find the probability that only 2 of them have heights at most 1.88m.

Let H be the random variable that denotes the height of a man in cm." Then $H \sim \mathrm{N}(185,\ 2^2)$.

(i) $\mathrm{P}(180 < H < 186) = 0.685$

(ii) $\mathrm{P}(H \le 188) = 0.933$

(iii) $\mathrm{P}(183 < H < 187) = 0.683$

Let Y be the random variable that denotes "the number of men, out of 3, who are at most 1.88 m".
Then $Y \sim \mathrm{B}(3,\ 0.93319)$.
$\mathrm{P}(Y = 2) = 0.175$

3. [H1 Maths/N2008/7part]
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.

A normal distribution should accommodate values up to ± 3 standard deviations from the mean. However, the maximum possible mark, 100, is only $\frac{100-72.1}{15.2} \approx 1.84$ s.d. above the mean. $\therefore$ A normal distribution is not a good fit in this case.

Alternative solution:

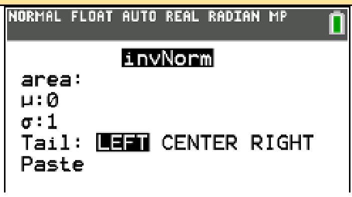
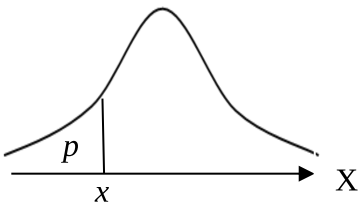
$$\begin{aligned} \mathrm{P}(X > 100) &= 0.0332 \\ &= 3.3\% \end{aligned}$$

This implies that 3.3% of the students scored more than 100 marks which is not impossible.

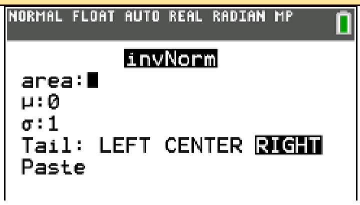
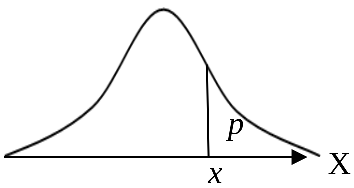
4.5 Use of Inverse Normal

Given $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$, the function **invNorm** is used:

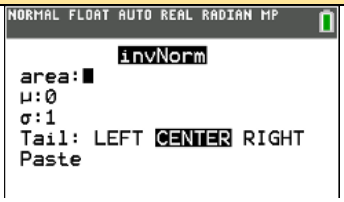
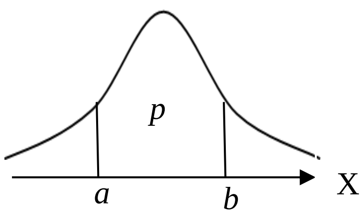
(i) to find x such that $\mathrm{P}(X < x) = p$ if p is given using LEFT tail.

Steps	Screenshot	Remarks
Press $\boxed{\text{2nd}}\boxed{\text{VARS}}\boxed{3}$ You should see this screen.		Note the heading is ‘invNorm’ If μ and σ are not specified, the default values used by the GC are 0 and 1 respectively.
Enter the values of : p (area), Mean (μ) and Standard deviation (σ) Select Tail “LEFT” Press ENTER, ENTER.		

(ii) to find x such that $\mathrm{P}(X > x) = p$ if p is given using RIGHT tail.

Steps	Screenshot	Remarks
Press $\boxed{\text{2nd}}\boxed{\text{VARS}}\boxed{3}$ You should see this screen.		Note the heading is ‘invNorm’ If μ and σ are not specified, the default values used by the GC are 0 and 1 respectively.
Enter the values of : p (area), Mean (μ) and Standard deviation (σ) Select Tail “RIGHT” Press ENTER, ENTER.		

(ii) to find a and b such that $\mathrm{P}(a < X < b) = p$ if p is given using CENTER tail.

Steps	Screenshot	Remarks
Press $\boxed{\text{2nd}}\boxed{\text{VARS}}\boxed{3}$ You should see this screen.		Note the heading is ‘invNorm’ If μ and σ are not specified, the default values used by the GC are 0 and 1 respectively.
Enter the values of : p (area), Mean (μ) and Standard deviation (σ) Select Tail “CENTER” Press ENTER, ENTER.		Note: a and b must be symmetrical about the mean.

<u>Example 4</u> (GC: InvNorm)

Given that $X \sim \mathrm{N}\left(123, 4^2\right)$.

(a) Find x such that

(i) $\mathrm{P}(X \le x) = 0.56$

(ii) $\mathrm{P}(X \ge x) = 0.78$

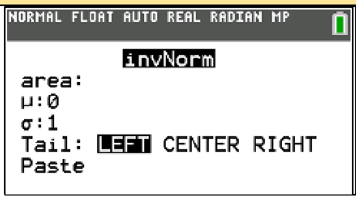
(b) Find a pair of possible values of a and b such that $\mathrm{P}(a < X < b) = 0.8$.

Question:
Are a and b unique?

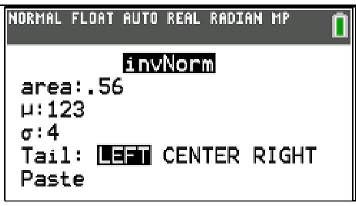
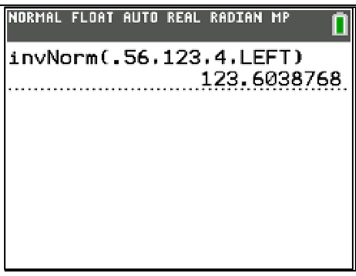
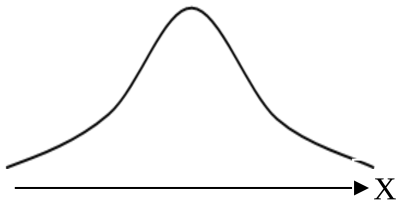
<u>Solution</u>:

(a)(i) $\mathrm{P}(X \le x) = 0.56$

$x = 124$

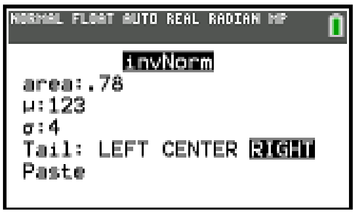
<table>
<tr><th>Steps</th><th>Screenshot</th><th>Remarks</th></tr>
<tr><td>Press 2nd VARS 3
You should see this screen.</td><td>
</td><td>Note the heading is ‘invNorm’

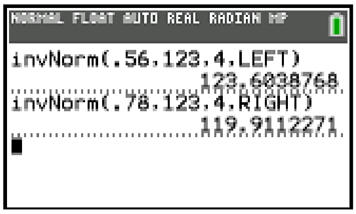
If μ and σ are not specified, the default values used by the GC are 0 and 1 respectively.</td></tr>
<tr><td>Enter the values of :
p (area),
Mean (μ) and
Standard deviation (σ)
Select LEFT

Press ENTER, ENTER.</td><td>
</td><td>If μ and σ are not specified, the default values used by the GC are 0 and 1 respectively.</td></tr>
<tr><td></td><td>
</td><td>
</td></tr>
</table>

(a)(ii) $\mathrm{P}(X \ge x) = 0.78$

$x = 120$

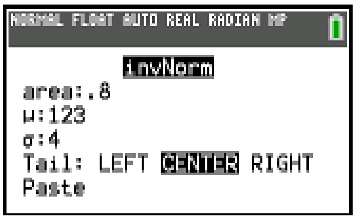


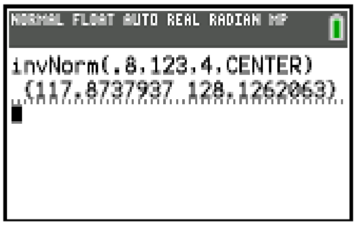


(b) $\mathrm{P}(a < X < b) = 0.8$

By choosing a and b to be symmetrical about μ , using GC,

$a = 118,\ \ b = 128$





Example 5

The marks scored by students in a test are found to be approximately normally distributed with mean 65 and standard deviation 20.

(i) If 15% of the students got an A grade for the test, calculate the least mark scored by students with an A grade,

(ii) What is the passing mark if Mr Lim hopes that 90 % of his students pass the test?

Solution:

Let X be the random variable that denotes the marks scored by a student in a test.

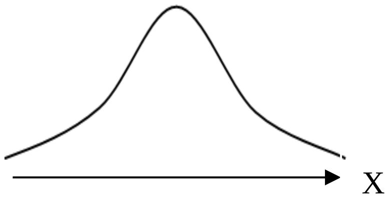
$$X \sim \mathrm{N}\left(65, 20^2\right)$$

(i) Let the marks to obtain an A grade be a.

$P\left(X \geq a\right) = 0.15$

$a = 85.7$ (3 s.f.)

A student needs to score at least 85.7 marks to obtain an A.

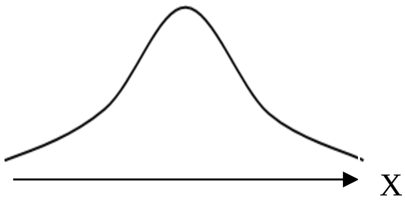


(ii) Let the passing mark be p.

$\mathrm{P}\left(X \geq p\right) = 0.9$

$p = 39.4$ (3 s.f.)

The passing mark for 90% of the students to pass is 39.4.



Exercise 2

1. Given $H \sim \mathrm{N}(185, 4)$, solve for the unknown in the following equations. Give your answer to the nearest 2 decimal places.

(i)	$\mathrm{P}\left(H < h\right) = 0.37$	$h = 184.34$
(ii)	$\mathrm{P}\left(H \geq a\right) = 0.8$	$a = 183.32$
(iii)	$\mathrm{P}\left(182 < H \leq b\right) = 0.71$	Ok… let's break this up: $\mathrm{P}\left(H \leq b\right) - \mathrm{P}\left(H \leq 182\right) = 0.71$ Now $\mathrm{P}\left(H \leq 182\right) = 0.066807$ So, $\mathrm{P}\left(H \leq b\right) = 0.71 + 0.066807$ $b = 186.52$
(iv)	$\mathrm{P}(-c < H - 185 < c) = 0.9$	$\mathrm{P}(-c < H - 185 < c) = 0.9$ $\mathrm{P}(-c + 185 < H < c + 185) = 0.9$ $c + 185 = 188.29$ $c = 3.29$
(v)	$\mathrm{P}\left(H \leq x\right) > 0.05$	Consider $\mathrm{P}\left(H \leq x\right) = 0.05$ first $x = 181.71$ This implies $\mathrm{P}\left(H \leq 181.71\right) = 0.05$ Therefore for $\mathrm{P}\left(H \leq x\right) > 0.05$, $x > 181.71$
(vi)	$\mathrm{P}\left(H \geq y\right) > 0.95$	Consider $\mathrm{P}\left(H \geq y\right) = 0.95$ $y = 181.71$ For $\mathrm{P}\left(H \geq y\right) \geq 0.95$, $y \leq 181.71$

If H is the normal distribution of the heights of men's population in a particular country, find the height that corresponds to the 90th percentile of the population.

Let x cm be the height corresponding to the 90th percentile of the population. $\mathrm{P}(H \le x) = 0.90$ $x = 187.56$

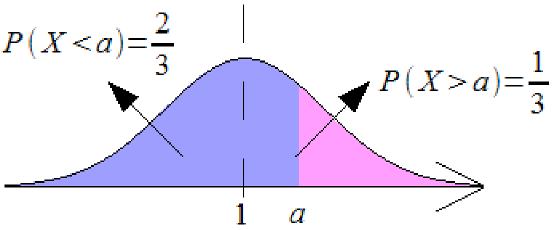
2. [N2003/II/27]
The random variable X has the distribution $\mathrm{N}(1,20)$. Given that $\mathrm{P}(X < a) = 2\mathrm{P}(X > a)$, find a.

$$\begin{aligned}\mathrm{P}(X < a) &= 2\mathrm{P}(X > a)\\ &= 2\left(1-\mathrm{P}(X \le a)\right)\\ &= 2-2\mathrm{P}(X \le a)\\ &= 2-2\mathrm{P}(X < a)\end{aligned}$$

$$3\mathrm{P}(X < a) = 2$$

$$\mathrm{P}(X < a) = \frac{2}{3}$$

$$a = 2.93$$



3. The marks obtained for Mathematics in an examination may be assumed to follow a normal distribution with a mean of 53 and a standard deviation of 7.9.
(i) Estimate, to the nearest integer, the pass mark if 15% of the students failed the examination,
(ii) All students with marks from 60 to 69 received a B grade. Given that 9 students received a B grade, how many students took the examination?

(i) Let X be the random variable that denotes the marks obtained for Mathematics.

$$X \sim \mathrm{N}(53, 7.9^2)$$

Let m be the pass mark.

$$\mathrm{P}(X < m) = 0.15$$

$$m = 44.8$$

$$m = 45$$

(ii) Let n be the number of students who took the test.

$$\mathrm{P}(60 \le X \le 69) \times n = 9$$

$$n = 9 \div \mathrm{P}(60 \le X \le 69)$$

$$n = 54.096$$

$$n = 54$$

4.6 Standard Normal Distribution

A standard normal distribution is a normal distribution with **mean 0** and **standard deviation 1**.

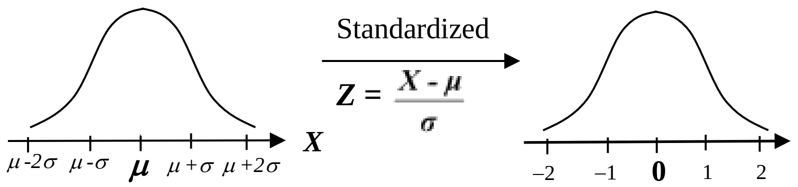
The random variable Z follows the standard normal distribution. We write

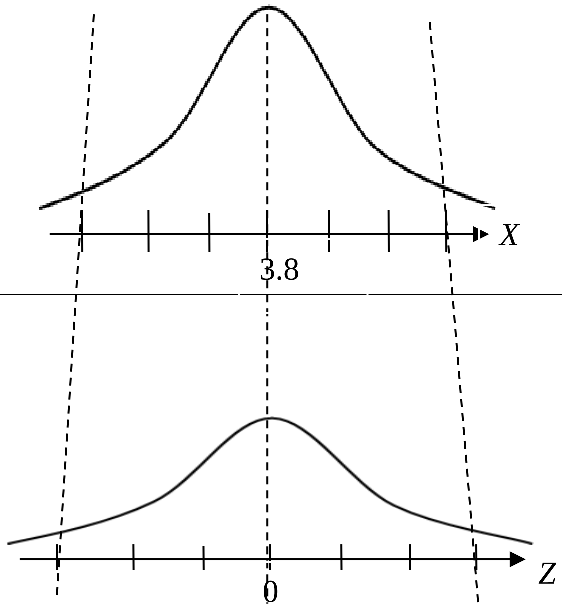
$$Z \sim \mathrm{N}(0,1).$$

Any normal random variable with mean μ and standard deviation σ, ie $X \sim \mathrm{N}(\mu, \sigma^2)$, can be transformed to the standard normal variable Z with mean 0 and standard deviation 1,

i.e. $Z \sim \mathrm{N}(0, 1)$, using $Z = \dfrac{X-\mu}{\sigma}$.

For example,



Let X be the time taken in hours to complete a marathon by a group of males. $X \sim \mathrm{N}(3.8, 0.5^2)$ Find the probability $\mathrm{P}(2.3 \le X \le 5.3)$? **Solution**: 0.997	
Recall that $\mathrm{P}(2.3 \le X \le 5.3) = 0.997$ and $\mathrm{P}(-3 \le Z \le 3) = 0.997$. Is $\mathrm{P}(2.3 \le X \le 5.3) = \mathrm{P}(-3 \le Z \le 3)$? **Solution**: $\mathrm{P}(2.3 \le X \le 5.3)$ $= \mathrm{P}\left(\frac{2.3-3.8}{0.5} \le \frac{X-3.8}{0.5} \le \frac{5.3-3.8}{0.5}\right)$ $= \mathrm{P}(-3 \le Z \le 3)$	

Therefore $\qquad \mathrm{P}(x_1 < X < x_2) = \mathrm{P}\left(\dfrac{x_1-\mu}{\sigma} < Z < \dfrac{x_2-\mu}{\sigma}\right).$

Note:
The process of transforming X to Z is called standardisation.

When do we need to convert a given normal distribution to the standard normal distribution?

When we are given a normal distribution with either an **unknown mean μ and/or unknown standard deviation σ**, it is useful for us to convert the given normal distribution to the standard normal distribution, so as to allow us to calculate the mean and/or variance.
(Refer to the following example.)

<u>Example 6</u> (Problem that involves finding the value of σ)
Given that a continuous random variable $X \sim \mathrm{N}\left(185, \sigma^2\right)$ and $\mathrm{P}(X > 190) = 0.2$, find the value of σ.
<u>Solution</u>:

$$\mathrm{P}(X > 190) = 0.2$$

$$\mathrm{P}\left(\frac{X-185}{\sigma} > \frac{190-185}{\sigma}\right) = 0.2$$

$$\mathrm{P}\left(Z > \frac{5}{\sigma}\right) = 0.2$$

$$\frac{5}{\sigma} = 0.84162 \Rightarrow \sigma = 5.94 \text{ (3 s.f.)}$$

<u>Example 7</u> (Problem that involves finding the values of μ and/or σ)

An examination was taken by 1000 candidates. 67 gained more than 78 % and 23 gained less than 29 %. Assuming the distribution to be normal, estimate the mean mark and the standard deviation.

<u>Solution</u>:
Let X be the random variable that denotes the marks obtained by a candidate.

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

Given that $\mathrm{P}(X > 78) = \dfrac{67}{1000}$ and $\mathrm{P}(X < 29) = \dfrac{23}{1000}$,

$$P(X > 78) = \frac{67}{1000}$$

$$P\left(\frac{X-\mu}{\sigma} > \frac{78-\mu}{\sigma}\right) = \frac{67}{1000}$$

$$P\left(Z > \frac{78-\mu}{\sigma}\right) = \frac{67}{1000}$$

$$\frac{78-\mu}{\sigma} = 1.49851 \ldots \quad (1)$$

$$\mathrm{P}(X < 29) = \frac{23}{1000}$$

$$\mathrm{P}\left(\frac{X-\mu}{\sigma} < \frac{29-\mu}{\sigma}\right) = \frac{23}{1000}$$

$$\mathrm{P}\left(Z < \frac{29-\mu}{\sigma}\right) = \frac{23}{1000}$$

$$\frac{29-\mu}{\sigma} = -1.99539 \qquad (2)$$

Solving equations (1) and (2), $\mu = 57.0$ and $\sigma = 14.0$ (use GC)

Exercise 3

1. (a) Find σ, given that $X \sim \mathrm{N}\left(21, \sigma^2\right)$ and $\mathrm{P}(X < 27) = 0.9332$.

   (b) Find μ, given that $X \sim \mathrm{N}(\mu, 12)$ and $\mathrm{P}(X > 32) = 0.8438$.

(a)

$$\mathrm{P}(X < 27) = 0.9332$$

$$\mathrm{P}\left(Z < \frac{27-21}{\sigma}\right) = 0.9332$$

$$\frac{6}{\sigma} = 1.500$$

$$\sigma = 4.00$$

(b)

$$\mathrm{P}(X > 32) = 0.8438$$

$$\mathrm{P}\left(Z > \frac{32-\mu}{\sqrt{12}}\right) = 0.8438$$

$$\frac{32-\mu}{\sqrt{12}} = -1.0102$$

$$\mu = 35.5$$

2. X is a normal variable with mean μ and standard deviation σ. It is given that $\mathrm{P}(X > 128) = 0.15$ and that $\mathrm{P}(X > 97) = 0.875$. Calculate μ and σ.

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

Since $\mathrm{P}(X > 128) = 0.15$

$$\Rightarrow \mathrm{P}\left(\frac{X-\mu}{\sigma} > \frac{128-\mu}{\sigma}\right) = 0.15$$

$$\Rightarrow \frac{128-\mu}{\sigma} = 1.0364$$

$$\mu + 1.0364\sigma = 128 \ldots\ldots\ldots(1)$$

Since $\mathrm{P}(X > 97) = 0.875$

$$\Rightarrow \mathrm{P}\left(\frac{X-\mu}{\sigma} > \frac{97-\mu}{\sigma}\right) = 0.875$$

$$\Rightarrow \frac{97-\mu}{\sigma} = -1.1503$$

$$\mu - 1.1503\sigma = 97 \ldots\ldots\ldots\ldots(2)$$

Solving (1) and (2) using GC,

$\sigma = 14.2$

$\mu = 113$

3. [MI/2009/BT/P2/Q9a (modified)]

   The random variable X has the distribution $\mathrm{N}(\mu, \sigma^2)$. Find the values of μ and σ given that $\mathrm{P}(X < 85) = \mathrm{P}(X > 101) = 0.0548$.

$$\mu = \frac{85+101}{2} = 93$$

$$\mathrm{P}(X < 85) = 0.0548$$

$$\mathrm{P}\left(Z < \frac{85-93}{\sigma}\right) = 0.0548$$

$$\frac{-8}{\sigma} = -1.599994$$

$$\sigma = 5.00$$

4.7 Linear Combinations of Independent Normal Distributions

4.7.1 Properties of Expectation & Variance

For any **independent** random variables X and Y where a and b are constants,

1.	$\mathrm{E}(a)=a$	$\mathrm{Var}(a)=0$ (Constants do not vary!)
2.	$\mathrm{E}(aX)=a\,\mathrm{E}(X)$	$\mathrm{Var}(aX)=a^2\,\mathrm{Var}(X)$
3.	$\mathrm{E}(aX\pm b)=a\,\mathrm{E}(X)\pm b$	$\mathrm{Var}(aX\pm b)=a^2\,\mathrm{Var}(X)$
4.	$\mathrm{E}(X\pm Y)=\mathrm{E}(X)\pm\mathrm{E}(Y)$	$\mathrm{Var}(X\pm Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$
5.	$\mathrm{E}(aX\pm bY)=a\,\mathrm{E}(X)\pm b\,\mathrm{E}(Y)$	$\mathrm{Var}(aX\pm bY)=a^2\,\mathrm{Var}(X)+b^2\,\mathrm{Var}(Y)$

Example 8

X and Y are two independent continuous random variables with mean 3 and 5 respectively, and variance 12 and 10 respectively. Find the mean and variance of

(i) $4X$
(ii) $5Y-3$
(iii) $9X+2Y$
(iv) $9X-2Y$

Solution:

(i) $\mathrm{E}(4X)=4\mathrm{E}(X)=4\times3=12$

$\mathrm{Var}(4X)=4^2\,\mathrm{Var}(X)=4^2\times12=192$

(ii) $\mathrm{E}(5Y-3)=5\mathrm{E}(Y)-3=5\times5-3=22$

$\mathrm{Var}(5Y-3)=5^2\,\mathrm{Var}(Y)=5^2\times10=250$

(iii) $\mathrm{E}(9X+2Y)=9\mathrm{E}(X)+2\mathrm{E}(Y)=9\times3+2\times5=37$

$\mathrm{Var}(9X+2Y)=9^2\,\mathrm{Var}(X)+2^2\,\mathrm{Var}(Y)=9^2\times12+2^2\times10=1012$

(iv) $\mathrm{E}(9X-2Y)=9\mathrm{E}(X)-2\mathrm{E}(Y)=9\times3-2\times5=17$

$\mathrm{Var}(9X-2Y)=9^2\,\mathrm{Var}(X)+2^2\,\mathrm{Var}(Y)=9^2\times12+2^2\times10=1012$

4.7.2 Sum and Difference of Independent Normal Distributions

Given two **independent** random variables X and Y, if $X \sim \mathrm{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathrm{N}(\mu_y, \sigma_y^2)$, then

$$aX + bY \sim \mathrm{N}\left(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2\right) \text{ and}$$

$$aX - bY \sim \mathrm{N}\left(a\mu_x - b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2\right).$$

Note:

1. $\mathrm{E}(aX \pm bY) = \mathrm{E}(aX) \pm \mathrm{E}(bY) = a\mu_x \pm b\mu_y$

   $\mathrm{Var}(aX \pm bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) = a^2\sigma_x^2 + b^2\sigma_y^2$

2. $\mathrm{E}(aX \pm bY) = \mathrm{E}(aX) \pm \mathrm{E}(bY)$ is true even if X and Y are not independent but the results for variance holds only if X and Y are independent.

Example 9

$X \sim \mathrm{N}(100, 25)$ and $Y \sim \mathrm{N}(120, 20)$ are two independent normal distributions. Find the following distribution

(i) $X + Y$

(ii) $X - Y$

(iii) $5X - 4Y + 14$

Solution:

(i) $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y) = 100 + 120 = 220$

$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 25 + 20 = 45$

$X + Y \sim \mathrm{N}(220, 45)$

(ii) $\mathrm{E}(X-Y) = \mathrm{E}(X) - \mathrm{E}(Y) = 100 - 120 = -20$

$\mathrm{Var}(X-Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 25 + 20 = 45$

$X - Y \sim \mathrm{N}(-20, 45)$

(iii) $\mathrm{E}(5X - 4Y + 14) = \mathrm{E}(5X) - \mathrm{E}(4Y) + \mathrm{E}(14)$

$= 5\,\mathrm{E}(X) - 4\,\mathrm{E}(Y) + 14 = 5(100) - 4(120) + 14 = 34$

$\mathrm{Var}(5X - 4Y + 14) = 5^2\,\mathrm{Var}(X) + 4^2\,\mathrm{Var}(Y) + 0 = 5^2(25) + 4^2(20) = 945$

$5X - 4Y + 14 \sim \mathrm{N}(34, 945)$

Exercise 4

1. [N87/2/9]

X and Y are continuous random variables having independent normal distributions. The means of X and Y are 10 and 12 respectively, and the standard deviations are 2 and 3 respectively.

Find

(i) $\mathrm{P}(Y < 10)$,

(ii) $\mathrm{P}(Y < X)$,

(iii) $\mathrm{P}(4X + 5Y > 90)$,

(iv) the value of x such that $\mathrm{P}(Y + X > x) = \frac{1}{4}$.

(i) $Y \sim \mathrm{N}(12, 3^2)$

$\mathrm{P}(Y < 10) = 0.252$

(ii) $X \sim \mathrm{N}(10, 2^2)$

$\mathrm{P}(Y < X) = \mathrm{P}(Y - X < 0)$

$\mathrm{E}(Y - X) = \mathrm{E}(Y) - \mathrm{E}(X) = 12 - 10 = 2$

$\mathrm{Var}(Y - X) = \mathrm{Var}(Y) + \mathrm{Var}(X) = 3^2 + 2^2 = 13$

$Y - X \sim \mathrm{N}(2, 13)$

$\mathrm{P}(Y - X < 0) = 0.290$

(iii) $\mathrm{E}(4X + 5Y) = 4\mathrm{E}(X) + 5\mathrm{E}(Y) = 4 \times 10 + 5 \times 12 = 100$

$\mathrm{Var}(4X + 5Y) = 4^2\mathrm{Var}(X) + 5^2\mathrm{Var}(Y) = 4^2 \times 2^2 + 5^2 \times 3^2 = 289$

$4X + 5Y \sim \mathrm{N}(100, 289)$

$\mathrm{P}(4X + 5Y > 90) = 0.722$

(iv) $\mathrm{E}(Y + X) = \mathrm{E}(Y) + \mathrm{E}(X) = 12 + 10 = 22$

$\mathrm{Var}(Y + X) = \mathrm{Var}(Y) + \mathrm{Var}(X) = 3^2 + 2^2 = 13$

$Y + X \sim \mathrm{N}(22, 13)$

$\mathrm{P}(Y + X > x) = \frac{1}{4}$

$x = 24.4$

4.7.3 Sum of *n* independent observations from the same Normal Distribution

If $X_1, X_2, \dots, X_n$ are n independent observations of the random variable X where

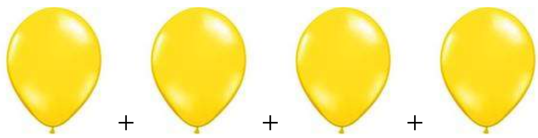
$$X \sim \mathrm{N}\left(\mu, \sigma^2\right),$$

then $X_1 + X_2 + \dots + X_n$ is also a Normal variable such that

$$X_1 + X_2 + \dots + X_n \sim \mathrm{N}\left(n\mu, n\sigma^2\right).$$

Question:
If X is a normal distribution, what's the difference between 4X and $\left(X_1 + X_2 + X_3 + X_4\right)$ *where* X_1, X_2, X_3, X_4 *are independent observations from X?!*

Answer:
Allow us to use the following illustration…
Suppose X represents a balloon with mean mass of 50grams and has a variance of 4grams.

$4X$ means multiplying the mass *4 TIMES*!!:  $\mathrm{E}(4X) = 4 \times \mathrm{E}(X) = 4 \times 50 = 200$g $\mathrm{Var}(4X) = 4^2 \times \mathrm{Var}(X) = 4^2 \times 4 = 64$g	$\left(X_1 + X_2 + X_3 + X_4\right)$ means adding up 4 independent distributions of X.  $\mathrm{E}\left(X_1 + X_2 + X_3 + X_4\right) = 4 \times \mathrm{E}(X) = 4 \times 50 = 200$g $\mathrm{Var}\left(X_1 + X_2 + X_3 + X_4\right) = 4 \times \mathrm{Var}(X) = 4 \times 4 = 16$g

Important Note:
It is important to distinguish between **<u>multiples</u>** of normal variable and **<u>sums</u>** of normal variables. To give another example, suppose X is the random variable "weight of a handphone". Then $X_1 + X_2$ refers to the sum of the weights of two different handphones, while $2X$ refers to the weight of a handphone which is twice as heavy as that of X.

Example 10

The length, in cm, of a rectangular tile is a normal variable with mean 19.8 and standard deviation 0.1. The breadth, in cm, is an independent normal variable with mean 9.8 and standard deviation 0.1.

(i) Find the probability that the sum of the lengths of five randomly chosen tiles exceeds 99.4 cm,

(ii) Find the probability that five times the length of a randomly chosen tile exceeds 99.4 cm,

(iii) Find the probability that the breadth of a randomly chosen tile is less than one half [1]of the length.

Solution:

Let X be the random variable that denotes the length of a tile in cm. $X \sim \mathrm{N}(19.8, 0.1^2)$

Let Y be the random variable that denotes the breadth of a tile in cm. $Y \sim \mathrm{N}(9.8, 0.1^2)$

(i)
$$\mathrm{E}(X_1 + X_2 + X_3 + X_4 + X_5) = 5\mathrm{E}(X) = 19.8 \times 5 = 99$$
$$\mathrm{Var}(X_1 + X_2 + X_3 + X_4 + X_5) = 5\mathrm{Var}(X) = 0.1^2 \times 5 = 0.05$$
$$X_1 + X_2 + X_3 + X_4 + X_5 \sim \mathrm{N}(99, 0.05)$$
$$\mathrm{P}(X_1 + X_2 + X_3 + X_4 + X_5 > 99.4) = 0.0368$$

(ii)
$$\mathrm{E}(5X) = 5\,\mathrm{E}(X) = 5(19.8) = 99$$
$$\mathrm{Var}(5X) = 5^2\,\mathrm{Var}(X) = 25(0.1^2) = 0.25$$
$$5X \sim \mathrm{N}(99, 0.25)$$
$$\mathrm{P}(5X > 99.4) = 0.212$$

(iii) For this part we are finding $\mathrm{P}(Y < 0.5X)$
$$\mathrm{E}(Y - 0.5X) = \mathrm{E}(Y) - 0.5\mathrm{E}(X) = 9.8 - 0.5 \times 19.8 = -0.1$$
$$\mathrm{Var}(Y - 0.5X) = \mathrm{Var}(Y) + 0.5^2\,\mathrm{Var}(X) = 0.1^2 + 0.5^2 \times 0.1^2 = 0.0125$$
$$Y - 0.5X \sim \mathrm{N}(-0.1, 0.0125)$$
$$\begin{aligned}\mathrm{P}(Y < 0.5X) &= \mathrm{P}(Y - 0.5X < 0) \\ &= 0.814\end{aligned}$$

[1] 'One half' refers to $\frac{1}{2}$.

Example 11

Melons are sold by weight at a price of \$1.50 per kilogram. The masses of melons are normally distributed with a mean of 0·8 kg and a standard deviation of 0.1 kg. Pumpkins are sold by weight at a price of \$0.50 per kilogram. The masses of pumpkins are normally distributed with mean of 1.2 kg and a standard deviation of 0.2 kg. Find the probability that the total price of 5 randomly chosen melons and 3 randomly chosen pumpkins exceeds \$8.

Solution:

Let X be the random variable that denotes the mass of a randomly chosen melon in kg
Let Y be the random variable that denotes the mass of a randomly chosen pumpkin in kg

$X \sim \mathrm{N}(0.8,\ 0.1^2)$, $Y \sim \mathrm{N}(1.2,\ 0.2^2)$

Let M be the random variable that denotes the cost of a randomly chosen melon
$M = 1.5X$
$\mathrm{E}(M) = \mathrm{E}(1.5X) = 1.5\,\mathrm{E}(X) = 1.5(0.8) = 1.2$
$\mathrm{Var}(M) = \mathrm{Var}(1.5X) = 1.5^2\,\mathrm{Var}(X) = 1.5^2(0.1^2) = 0.15^2 = 0.0225$
$M \sim \mathrm{N}(1.2,\ 0.15^2)$

Let P be the random variable that denotes the cost of a randomly chosen pumpkin
$P = 0.5Y$
$\mathrm{E}(P) = \mathrm{E}(0.5Y) = 0.5\,\mathrm{E}(Y) = 0.5(1.2) = 0.6$
$\mathrm{Var}(P) = \mathrm{Var}(0.5Y) = 0.5^2\,\mathrm{Var}(Y) = 0.5^2(0.2^2) = 0.1^2 = 0.01$
$P \sim \mathrm{N}(0.6,\ 0.1^2)$

Let T be the random variable that denotes the total cost of 5 randomly chosen melons and 3 randomly chosen pumpkins
$T = M_1 + M_2 + M_3 + M_4 + M_5 + P_1 + P_2 + P_3$

$$\mathrm{E}(T) = \mathrm{E}(M_1+M_2+M_3+M_4+M_5+P_1+P_2+P_3) = 5\,\mathrm{E}(M) + 3\mathrm{E}(P) = 5(1.2) + 3(0.6) = 7.8$$

$$\mathrm{Var}(T) = \mathrm{Var}(M_1+M_2+M_3+M_4+M_5+P_1+P_2+P_3) = 5\mathrm{Var}(M) + 3\mathrm{Var}(P) = 5(0.15^2) + 3(0.1^2)$$
$$= 0.1425$$

$T \sim \mathrm{N}(7.8,\ \sqrt{0.1425}^2)$
$\mathrm{P}(T > 8) = 0.298$

Example 12

Bottles of mineral water are delivered to shops in crates containing 12 bottles each. The weights of the bottles are normally distributed with mean 2 kg and standard deviation 0.05 kg. The weights of the crates are normally distributed with mean 2.5 kg and standard deviation 0.3 kg.

(i) Assuming that all random variables are independent, find the probability that a full crate will weigh between 26 kg and 27 kg,

(ii) Find the probability that the difference in mass between two randomly chosen full crates is more than 1 kg.

(iii) Three bottles are selected at random from a crate. Find the probability two of them weigh more than 2.1 kg and one weighs less than 2.1 kg,

(iv) A crate that weigh between 26 kg and 27 kg is classified as SAFE. Find the probability of finding at least 45 SAFE crates in a container of 60 crates.

Solution:

Let X be the random variable that denotes the weight of a bottle measured in kg.

$$X \sim \mathrm{N}\left(2, 0.05^2\right)$$

Let Y be the random variable that denotes the weight of a crate measured in kg.

$$Y \sim \mathrm{N}\left(2.5, 0.3^2\right)$$

(i) Let W be the random variable that denotes the weight of a full crate measured in kg.

$$W = X_1 + X_2 + X_3 + \cdots + X_{12} + Y$$
$$\mathrm{E}(W) = \mathrm{E}(X_1 + X_2 + X_3 + \cdots + X_{12} + Y) = 12 \times 2 + 2.5 = 26.5$$
$$\mathrm{Var}(W) = \mathrm{Var}(X_1 + X_2 + X_3 + \cdots + X_{12} + Y) = 12 \times 0.05^2 + 0.3^2 = 0.12$$
$$W \sim \mathrm{N}(26.5, 0.12)$$
$$\mathrm{P}(26 < W < 27) = 0.851$$

(ii)
$$\mathrm{E}(W_1 - W_2) = 0$$
$$\mathrm{Var}(W_1 - W_2) = 2(0.12)$$
$$W_1 - W_2 \sim \mathrm{N}(0, 0.24)$$
$$\mathrm{P}(|W_1 - W_2| > 1) = 1 - \mathrm{P}(-1 < W_1 - W_2 < 1)$$
$$= 0.0412$$

(iii) Let A be the random variable that denotes the number of bottles out of 3 that exceed 2.1 kg

$$A \sim \mathrm{B}(\mathbf{3}, \mathrm{P}(X > \mathbf{2.1}))$$
$$\mathrm{P}(A = 2) = 0.00152$$

(iv) Let T be the random variable that denotes the number of SAFE crates out of 60 crates

$$T \sim \mathrm{B}(60, 0.85109)$$
$$\mathrm{P}(T \geq 45) = 1 - \mathrm{P}(T < 45)$$
$$= 1 - \mathrm{P}(T \leq 44)$$
$$= 0.987$$

<u>Example 13 (Self Reading)</u>
The time of arrival of a bus at a bus stop varies in a normal distribution with a mean of 0900h and a standard deviation of 2 min. Independently, a second bus departs from this stop at a time which varies in a normal distribution with a mean of 0901h and a standard deviation of 1 min. Find the probability that

(i) the first bus arrives before the second bus leaves,
(ii) this happens on 5 given consecutive days.

<u>Solution</u>:

(i) Let X be the time of arrival in mins of the first bus after 9 am. $X \sim \mathrm{N}\left(0, 2^2\right)$

Let Y be the time of departure in mins of the second bus after 9 am. $Y \sim \mathrm{N}\left(1, 1^2\right)$

$\mathrm{E}(X - Y) = 0 - 1 = -1$
$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 2^2 + 1^2 = 5$
$\therefore X - Y \sim \mathrm{N}(-1, 5)$
$\mathrm{P}(X < Y) = \mathrm{P}(X - Y < 0) = 0.67264 \approx 0.673$

(ii) $\mathrm{P}\left(\text{happens on 5 consecutive days}\right) = (0.67264)^5 = 0.138$

<u>Exercise 5</u>

1. B and C are the random variables "mass of a banana in kg" and "mass of a carrot in kg" respectively. Assume that the masses of bananas and carrots are all independent of one another.
Express, in terms of B and C, the probabilities that

(i) 4 randomly chosen carrots weigh more than 700g,
(ii) thrice the mass of one randomly chosen carrot is less than the total mass of three randomly chosen carrots,
(iii) 2 randomly chosen bananas are less than 50g apart in mass,
(iv) the difference in mass between 2 randomly chosen bananas is more than 50g,
(v) the mass of a banana exceeds the mass of a carrot by not more than 35g,
(vi) between 2 randomly chosen carrots, one weighs more than 200g and the other weighs less than 150g.
(vii) between three random chosen carrots, one weighs more than 200g and the other two weighs less than 150g.

<u>Solution</u>:

(i) $\mathrm{P}(C_1 + C_2 + C_3 + C_4 > 0.7)$
Qns Prompt: How would the question be phrased if $\mathrm{P}(4C > 0.7)$ is required?
Answer: Find the probability that 4 times the mass of one randomly chosen carrot is more than 700g

(ii) $\mathrm{P}(3C < C_1 + C_2 + C_3)$

(iii) $\mathrm{P}(-0.05 < B_1 - B_2 < 0.05)$
Qns Prompt: Why is the answer not $\mathrm{P}(B_1 - B_2 < 0.05)$?
Answer: The question did not specific which banana is the heavier banana out of the two bananas.

(iv) P($B_1 - B_2 > 0.05$) + P($B_2 - B_1 > 0.05$) i.e. 2P($B_1 - B_2 > 0.05$)
Qns Prompt: What is significance of the word "difference"?
Answer: You have consider two cases when "difference" appears.

(v) P($0 < B - C \le 0.035$)
Qns Prompt: Why is it necessary to ensure $B - C > 0$?
Answer: The mass of a banana exceeds the mass of a carrot
Qns Prompt: What is another way of phrasing the question to get the same mathematical statement?
Answer: Find the probability that the mass of a banana is more than the mass of a carrot by not more than 35g

(vi) P($C_1 > 0.2$)×P($C_2 < 0.15$) + P($C_2 > 0.2$)×P($C_1 < 0.15$)
i.e. 2P($C > 0.2$)×P($C < 0.15$)
Qns Prompt: Why is the answer not P($C > 0.2$)×P($C < 0.15$)?
Answer: You have consider two cases as the question did not specific which carrot is heavier.

(vii) $\mathrm{P}(C_1 > 0.2)\times\mathrm{P}(C_2 < 0.15)\times\mathrm{P}(C_3 < 0.15)+$
$\mathrm{P}(C_2 > 0.2)\times\mathrm{P}(C_1 < 0.15)\times\mathrm{P}(C_3 < 0.15)+$
$\mathrm{P}(C_3 > 0.2)\times\mathrm{P}(C_1 < 0.15)\times\mathrm{P}(C_2 < 0.15)$
i.e. $\binom{3}{1}$P($C > 0.2$)×P($C < 0.15$)×P($C < 0.15$)

2. The length of time which an ordinary light-bulb will last may be taken to have a normal distribution with mean 600 hours and standard deviation 100 hours. The length of time for which a new 'long-life' bulb will last may be taken to have a normal distribution with mean 2000 hours and standard deviation 200 hours.

(a) One ordinary bulb is chosen at random. Find the probability that it will last for more than 450 hours.

(b) Two ordinary bulbs are chosen at random. Find the probability that the sum of the times for which they last will be less than 1100 hours.

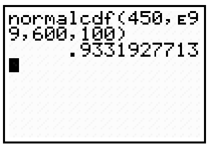
(c) One ordinary bulb and one long-life bulb are chosen at random. Find the probability that the long-life bulb lasts for more than three times as long as the ordinary bulb.

Solution:

Let X and Y be the lengths of time, in hours, an ordinary and 'long-life' light-bulb will last respectively.

Then $X \sim \mathrm{N}(600,\ 100^2)$, $Y \sim \mathrm{N}(2000,\ 200^2)$

a) P(X > 450) = 0.933

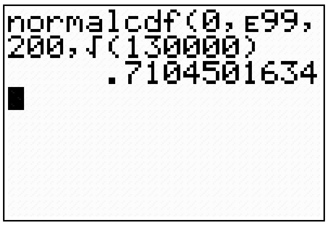


b) $X_1 + X_2 \sim \mathrm{N}(600\times 2, 100^2\times 2)$

$\Rightarrow X_1 + X_2 \sim \mathrm{N}(1200, 20000)$

P($X_1 + X_2 < 1100$) = 0.240

c)
$Y-3X \sim \mathrm{N}(2000-(600\times 3), 200^2+100^2\times 3^2)$
$\Rightarrow Y-3X \sim \mathrm{N}(200, 130000)$
P(Y > 3X) = P(Y – 3X >0) = 0.710



3. **[GCE A Level/N2009/P2/part Q9]** The thickness in cm of a mechanics textbook is a random variable with the distribution $\mathrm{N}(2.5,\ 0.1^2)$. The thickness in cm of a statistics textbook is a random variable with the distribution $\mathrm{N}(2.0,\ 0.08^2)$.
   (i) Calculate the probability that 21 mechanics textbooks and 24 statistics textbooks will fit onto a bookshelf of length 1 m. State clearly the mean and variance of any normal distribution you use in your calculation.
   (ii) Calculate the probability that the total thickness of 4 statistics textbooks is less than three times the thickness of 1 mechanics textbook. State clearly the mean and variance of any normal distribution you use in your calculation.
   (iii) State an assumption needed for your calculations in parts (i) and (ii).

(i)	Let the thickness of a mechanics textbook be M. $M \sim \mathrm{N}(2.5,\ 0.1^2)$ Let the thickness of a statistics textbook be S. $S \sim \mathrm{N}(2.0,\ 0.08^2)$ Let $X = M_1 + ... + M_{21} + S_1 + ... + S_{24}$. $\mathrm{E}(X) = 21\times 2.5 + 24\times 2.0 = 100.5$; $\mathrm{Var}(X) = 21\times 0.1^2 + 24\times 0.08^2 = 0.3636$ $X \sim \mathrm{N}(100.5, 0.3636)$ $\mathrm{P}\left(X \le 100\right) = 0.203$
(ii)	Let $Y = 3M - \left(S_1 + S_2 + S_3 + S_4\right)$ $\mathrm{E}(Y) = 3\times 2.5 - 4\times 2.0 = -0.5$; $\mathrm{Var}(Y) = 9\times 0.1^2 + 4\times 0.08^2 = 0.1156$ $Y \sim \mathrm{N}(-0.5, 0.1156)$ $\mathrm{P}\left(S_1+S_2+S_3+S_4 < 3M\right) = \mathrm{P}\left(3M - \left(S_1+S_2+S_3+S_4\right) > 0\right)$ $= \mathrm{P}\left(Y > 0\right)$ $= 0.0707$
(iii)	Assume that the thickness of each statistics and mechanics textbook is independent of one another.

4. [AJC/2007/Prelim/P2/Q11]
   Large beer cans contain a volume of beer which is normally distributed with mean 500 ml and standard deviation 3.3 ml, while small beer cans contain a volume of beer which is also normally distributed with mean 340 ml and standard deviation 2.4 ml. The volume of beer in any can is independent of the volume of beer in any other can.
   (a) If 5% of the small cans produced contain more than k ml of beer each, find the value of k.
   (b) Find the probability that the volume of beer in two large cans differ by not more than 10 ml.
   (c) Beer is also sold in bottles, each of which contains four times the volume of a large can. Find the probability that a crate of six bottles contain in total more than 12040 ml of beer.

Let the volume of beer in a large and small beer can be L and S respectively. $L \sim \mathrm{N}(500, 3.3^2)$, $S \sim \mathrm{N}(340, 2.4^2)$	
(a) $\mathrm{P}(S > k) = 0.05$ $k = 344$	(c) Let the volume of beer in a bottle be X. $X = 4L \sim \mathrm{N}(2000, 174.24)$ $X_1 + ... + X_6 \sim \mathrm{N}(12000, 1045.44)$ $\mathrm{P}(X_1 + ... + X_6 > 12040) = 0.108$
(b) $L_1 - L_2 \sim \mathrm{N}(0, 21.78)$ $\mathrm{P}(\lvert L_1 - L_2 \rvert \le 10) = \mathrm{P}(-10 \le L_1 - L_2 \le 10)$ $= 0.968$	

5. **[GCE A Level/N2015/P2/Q12]** In this question you should state clearly the values of the parameters of any normal distribution you use.

The masses in grams of apples have the distribution N(300, 20^2) and the masses in grams of pears have the distribution N(200, 15^2). A certain recipe requires 5 apples and 8 pears.

(i) Find the probability that the total mass of 5 randomly chosen apples is more than 1600 grams.

(ii) Find the probability that the total mass of 5 randomly chosen apples is more than the total mass of 8 randomly chosen pears.

The recipe requires the apples and pears to be prepared by peeling them and removing the cores. This process reduces the mass of each apple by 15% and the mass of each pear by 10%.

(iii) Find the probability that the total mass, after preparation, of 5 randomly chosen apples and 8 randomly chosen pears is less than 2750 grams.

(i)	Let X and Y be the mass, in grams, of an apple and a pear respectively. $X \sim \mathrm{N}(300, 20^2)$, $Y \sim \mathrm{N}(200, 15^2)$ $X_1 + ... + X_5 \sim N(5\times 300, 5\times 20^2) = N(1500, 2000)$ $\mathrm{P}(X_1 + ... + X_5 > 1600) = 0.012674 \approx 0.0127$
(ii)	$(X_1 + ... + X_5) - (Y_1 + ... + Y_8) \sim N(5\times 300 - 8\times 200, 5\times 20^2 + 8\times 15^2) = \mathrm{N}(-100, 3800)$ $\mathrm{P}\big((X_1 + ... + X_5) > (Y_1 + ... + Y_8)\big)$ $= \mathrm{P}\big((X_1 + ... + X_5) - (Y_1 + ... + Y_8) > 0\big) = 0.052379 \approx 0.0524$
(iii)	The total mass of 5 apples and 8 pears after preparation is the random variable, $T = 0.85(X_1 + ... + X_5) + 0.9(Y_1 + ... + Y_8)$ $T \sim \mathrm{N}(0.85\times 5\times 300 + 0.9\times 8\times 200, 0.85^2\times 5\times 20^2 + 0.9^2\times 8\times 15^2)$ $T \sim \mathrm{N}(2715, 2903)$ P(T < 2750) = 0.74202 ≈ 0.742

6. The school bus leaves the bus stop near Ben's house at X minutes past 8.00 a.m., where X follows a normal distribution with mean 20 minutes and standard deviation 3 minutes. Ben reaches the bus stop at Y minutes after 8.00 a.m., where Y follows a normal distribution of 15 minutes and standard deviation 2 minutes. Assuming that X and Y are independent, find the probability that Ben misses the bus.

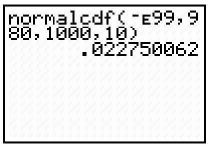
The ride from this bus stop to the school lasts Z minutes, where Z follows a normal distribution of mean 30 minutes and standard deviation $\sqrt{7}$ minutes. W is the number of minutes before 9.00 a.m. at which the bus arrives at the school. Calculate the mean and variance of W and find, to three decimal places, the probability that the school bus arrives at the school after 9.00 a.m.

$X \sim \mathrm{N}(20,3^2)$, $Y \sim \mathrm{N}(15,2^2)$, $Z \sim \mathrm{N}(30,\sqrt{7})$ $X-Y \sim \mathrm{N}\left(20-5,3^3+2^2\right)$ $X-Y \sim \mathrm{N}\left(5,13\right)$ $\mathrm{P}\left(X<Y\right)=\mathrm{P}\left(X-Y<0\right)=0.0828$	$X+Z+W=60 \Rightarrow W=60-X-Z$ $\mathrm{E}(W)=60-20-30=10$ $\mathrm{Var}(W)=9+7=16$ $W \sim \mathrm{N}(10,16)$ Probability that the school bus arrives at the school after 9.00 a.m. $=\mathrm{P}(W<0)=0.006$

Practice Questions

1. A machine produces packets of a manufactured powder. Each packet has a nominal mass of 1 kg, but the packets produced have masses which are normally distributed with a standard deviation of 10 g. Government regulations forbid the sale of packets having a mass less than 980 g. If the mean mass of packets can be adjusted, find
   (a) the percentage of output rejected if the mean mass is set at 1000 g,
   (b) to the nearest gram, the minimum to which the mean may be set if not more than 5% of the output is to be rejected.

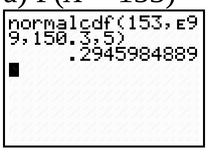
Let X be the mass, in kg, of a packet of manufactured powder.
Then $X \sim \mathrm{N}(1000,\ 10^2)$.

(a) $\mathrm{P}(X<980)=0.0228$  The percentage of output rejected if the mean mass is set at 1000g is 2.28%	(b) Let Y be the mass of a packet of manufactured powder after adjustment. Then $Y \sim \mathrm{N}(\mu,\ 10^2)$ $\mathrm{P}(Y<980)\le 0.05$ $\mathrm{P}(Z<\frac{980-\mu}{10})\le 0.05$ $\frac{980-\mu}{10}\le -1.6449$ $\mu \ge 980+(1.6449\times 10)$ $\mu \ge 996.45$ Least mean is 997g.

2. The height of boys at a particular age follows a normal distribution with mean 150.3 cm and standard deviation 5 cm. Find the probability that a boy picked at random from the group has height
   (a) more than 153 cm,
   (b) between 148 cm and 155 cm.
   If three boys are picked at random, find the probability that only 2 of them have heights less than 153.

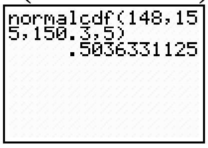
Let X be the random variable denoting "the height of boy at a particular age."
Then $X \sim \mathrm{N}(150.3,\ 5^2)$

a) $\mathrm{P}(X > 153) = 0.295$



b)

P(148 <X< 155) = 0.504



P(X < 153) = 1 – 0.2946 = 0.7054

Let Y be the random variable denoting “the number of boys have heights less than 153”

Then $Y \sim \text{B}(3,\ 0.7054)$

$$\text{P}(Y=2)=\binom{3}{2}(0.7054)^2(0.294)=0.440$$

3. [2016 Promo YJC/Q8]

The radius of a rubber ball manufactured by a factory follows a normal distribution with mean μ cm and standard deviation σ cm. It is known that 95% of them have a radius of at most 10 cm and 70% of them have a radius of at least 7 cm. Show that μ = 7.73 and σ = 1.38, correct to 3 significant figures.

The factory also produces metal pipes, each with inner diameter d cm, which will allow the manufactured balls to pass through. Find the range of values of d which will allow at least 99% of the balls to pass through.

Let X be the radius (in cm) of a rubber ball. $X \sim \text{N}(\mu,\sigma^2)$	
P(X $\le$ 10) = 0.95 $\text{P}(Z \le \frac{10-\mu}{\sigma}) = 0.95$ From GC, $\frac{10-\mu}{\sigma} = 1.64485$ $\mu + 1.64485\sigma = 10$ --- (1)	P(X $\ge$ 7) = 0.7 $\text{P}(Z \le \frac{7-\mu}{\sigma}) = 0.3$ From GC, $\frac{7-\mu}{\sigma} = -0.52440$ $\mu - 0.52440\sigma = 7$ --- (2)

Solving (1) and (2),

$\mu = 7.725 \approx 7.73,\ \ \sigma = 1.383 \approx 1.38$ (shown)

$X \sim \text{N}(7.73, 1.38^2)$

For the balls to pass through the metal pipe, $\text{P}\left(X < \frac{d}{2}\right) \ge 0.99$

Using GC: $\frac{d}{2} \ge 10.940$

$d \ge 21.9$

4. [2016 Promo PJC/Q2]

A random variable Y has the distribution $\text{N}(\mu,\ \sigma^2)$.

Given that $\text{P}(Y<85)=\text{P}(Y>155)=0.05$, find μ and σ^2.

The random variable X has the distribution N(40, 3^2) and is related to Y by the formula $Y = aX + b$. Given that a is a positive real number and b is a negative real number, determine the values of a and b.

Given E(Y) = μ , Var(Y) = σ^2 Consider $\text{P}(Y<85)=\text{P}(Y>155)$ $\Rightarrow\ \text{E}(Y)=\mu=\frac{155+85}{2}=120$ Consider $\text{P}(Y<85)$ = 0.05	Given $Y = aX + b$ and X ~ N(40, 3^2) Consider E(Y) = a E(X) + b $\Rightarrow 120 = 40a + b$ ----------(1) Consider $\text{Var}(Y) = a^2\text{Var}(X)$ $\Rightarrow 452.75 \approx 9a^2$

$\Rightarrow \ P(Z < \frac{85-120}{\sigma}) = 0.05$ $\Rightarrow \frac{85-120}{\sigma} \approx -1.6449$ Solving, $\sigma = 21.278$. Thus Var $(Y) = \sigma^2 = 21.278^2 \approx 453$	$\Rightarrow a = 7.0926$ or -7.0926 (NA since a is a positive real number) Substitute $a = 7.0926$ into (1), $b = -163.704$ Thus $a = 7.09$, $b = -164$

5. **[GCE A Level/N2010/P2/Q9]** In this question you should state clearly the values of the parameters of any normal distribution you use.

Over a three-month period Ken makes X minutes of peak-rate telephone calls and Y minutes of cheap-rate calls. X and Y are independent random variables with the distributions $\mathrm{N}(180, 30^2)$ and $\mathrm{N}(400, 60^2)$ respectively.

(i) Find the probability that, over a three-month period, the number of minutes of cheap-rate calls made by Ken is more than twice the number of minutes of peak-rate calls.

Peak-rate calls cost \$0.12 per minute and cheap-rate calls cost \$0.05 per minute.

(ii) Find the probability that, over a three-month period, the total cost of Ken's calls is greater than \$45.

(iii) Find the probability that the total cost of Ken's peak-rate calls over two independent three-month periods is greater than \$45.

(i) $X \sim \mathrm{N}\left(180, 30^2\right)$, $Y \sim \mathrm{N}\left(400, 60^2\right)$ $Y - 2X \sim \mathrm{N}\left(40, 7200\right)$ $\mathrm{P}\left(Y > 2X\right) = \mathrm{P}\left(Y - 2X > 0\right)$ $= 0.681$	(ii) Let T be the total cost of Ken's calls over a three-month period. $T = 0.12X + 0.05Y \sim \mathrm{N}(41.6, 21.96)$ $\mathrm{P}(T > 45) = 0.234$ (iii) $0.12\left(X_1 + X_2\right) \sim \mathrm{N}\left(43.2, 25.92\right)$ $\mathrm{P}\left(0.12\left(X_1 + X_2\right) > 45\right) = 0.362$

6. [2016 Promo SAJC/ Q4]

A coffee machine dispenses coffee into mugs of two sizes, namely a large mug or a small mug. The volume of coffee, in ml, dispensed on each occasion have independent normal distributions with means and standard deviations as shown in the following table.

Size of mug	Mean	Standard deviation
Large	355	13.6
Small	235	6.8

(i) Two large mugs are randomly chosen. Find the probability that the volume of coffee in one large mug is less than 330 ml and the volume of coffee in the other large mug is more than 330 ml.

(ii) Find the probability that the volume of coffee in three randomly chosen small mugs exceeds twice the volume of coffee in a randomly chosen large mug. State clearly the mean and variance of the distribution used.

(iii) It is now given that a large mug of coffee costs \$0.03 per ml and a small mug of coffee costs \$0.02 per ml. Find the probability that the total cost of three randomly chosen large mugs of coffee and two randomly chosen small mugs of coffee is more than \$40.

(i)	Let L be the amount of coffee in a large mug in ml. $L \sim \mathrm{N}\left(355, 13.6^2\right)$ Let S be the amount of coffee in a small mug in ml. $S \sim \mathrm{N}\left(235, 6.8^2\right)$ Required probability $= 2\mathrm{P}(L_1 < 330)\mathrm{P}(L_2 > 330)$ $= 2(0.0330)(1-0.0330)$ $= 0.0638$ (3 s.f.)
(ii)	$\mathrm{E}(S_1+S_2+S_3-2L) = 3\times 235 - 2\times 355 = -5$ $\mathrm{Var}(S_1+S_2+S_3-2L) = 3\times 6.8^2 + 2^2\times 13.6^2 = 878.56$ $\mathrm{P}(S_1+S_2+S_3 > 2L) = \mathrm{P}(S_1+S_2+S_3-2L > 0)$ $= 0.43302$ $= 0.433$ (3 s.f.)
(iii)	Let C be the cost of three randomly chosen Large mugs and two random chosen Small mugs. Then, $C = 0.03\left(L_1+L_2+L_3\right) + 0.02\left(S_1+S_2\right)$ $C \sim \mathrm{N}(0.03\times 3\times 355 + 0.02\times 2\times 235, 0.03^2\times 3\times 13.6^2 + 0.02^2\times 2\times 6.8^2)$ $C \sim \mathrm{N}\left(41.35, 0.5364\right)$ $\mathrm{P}\left(C > 40\right) = 0.96736$ $\approx 0.967\ \left(3\text{ s.f.}\right)$

7. Beez Bee Farm produces and sells honey in jars of two sizes, small and large. The amount of honey in the jars are normally distributed with mean μ ml, standard deviation σ ml and the cost per jar is as shown in the table below.

Size of jar	μ	σ	Cost per jar
Small	703	2.3	\$12
Large	1407	7.8	\$24

(i) Show that the probability of two large jars of honey containing at least 2.8 litres of honey is 0.898, correct to 3 significant figures. [2]

(ii) Find the probability that the total amount of honey in one large jar and two small jars is at least 2.8 litres. [3]

Scones And Juice Confectionary requires 2.8 litres of honey daily for their cakes and desserts. They have budgeted to spend \$50 on honey every day. In order to ensure freshness of their products, any honey in excess of 2.8 litres will be donated to charity.

(iii) Suggest another combination of small and/or large jars of honey (besides those considered in parts **(i)** and **(ii)**) that can meet the daily requirement of 2.8 litres of honey and is within the budget of \$50. Explain clearly which combination the confectionary should choose.

(iv) In a particular month of 30 days, what is the probability that the daily requirement is met on more than 28 days of the 30 days if the combination in **(iii)** is chosen?

(i) Let X and Y be the amount of honey, in ml, in a small and large jar of honey respectively. $X \sim \mathrm{N}\left(703, 2.3^2\right)$ $Y \sim \mathrm{N}\left(1407, 7.8^2\right)$ $Y_1 + Y_2 \sim \mathrm{N}\left(1407+1407, 7.8^2+7.8^2\right)$ $\therefore Y_1 + Y_2 \sim \mathrm{N}\left(2814, 121.68\right)$ $\mathrm{P}\left(Y_1+Y_2 \geq 2800\right) = 0.89781 = 0.898$ (3 s.f.)	(iii) Suggestion: 4 small bottles $X_1+X_2+X_3+X_4 \sim \mathrm{N}\left(703\times 4, 2.3^2\times 4\right)$ $\therefore X_1+X_2+X_3+X_4 \sim \mathrm{N}\left(2812, 21.16\right)$ Four small bottles cost \$48 which is within the budget of \$50 and is expected to meet the daily requirement. $\mathrm{P}\left(X_1+X_2+X_3+X_4 \geq 2800\right) = 0.99546$ $= 0.995$ (3 s.f.) It's best to order 4 small jars since it has the highest probability of meeting the requirement.
(ii) $X_1 + X_2 + Y$ $\sim \mathrm{N}(703+703+1407, 2.3^2+2.3^2+7.8^2)$ $\therefore X_1+X_2+Y \sim \mathrm{N}(2813, 71.42)$ $\mathrm{P}(X_1+X_2+Y \geq 2800) = 0.93801$ $= 0.938$ (3 s.f.)	(iv) Let W be the number of days in a month, out of 30 days, that the requirement is met. $W \sim \mathrm{B}\left(30, 0.99546\right)$ $\mathrm{P}\left(W > 28\right) = 1 - \mathrm{P}\left(W \leq 28\right)$ $= 0.99176 = 0.992$ (3 s.f.) *If students are unable to answer (iii) and if chosen combination is 1 large and 2 small jars of honey:* $W \sim \mathrm{B}\left(30, 0.93801\right)$ $\mathrm{P}\left(W > 28\right) = 1 - \mathrm{P}\left(W \leq 28\right)$ $= 0.43734 = 0.437$ (3 s.f.)

4.8 Summary and Checklist

1 Properties of expectation and variance

Properties of Expectation	**Properties of Variance**
Given any constants a and b and random variables X and Y	Given any constants a and b
1) $\mathrm{E}(a)=a$	1) $\mathrm{Var}(a)=0$
2) $\mathrm{E}(aX)=a\,\mathrm{E}(X)$	2) $\mathrm{Var}(aX)=a^2\,\mathrm{Var}(X)$
3) $\mathrm{E}(aX\pm b)=a\,\mathrm{E}(X)\pm b$	3) $\mathrm{Var}(aX+b)=a^2\,\mathrm{Var}(X)$
	For independent random variables X and Y
4) $\mathrm{E}(X\pm Y)=\mathrm{E}(X)\pm\mathrm{E}(Y)$	4) $\mathrm{Var}(X\pm Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$
	5) $\mathrm{Var}(aX\pm bY)=a^2\,\mathrm{Var}(X)+b^2\,\mathrm{Var}(Y)$

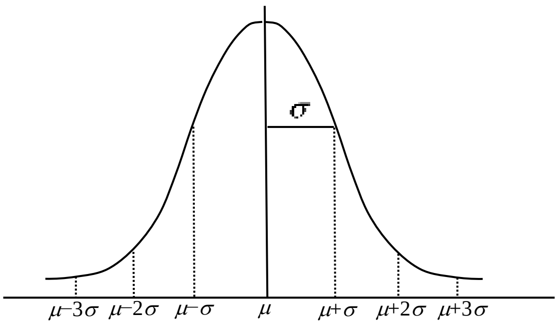
2 Properties of a Normal Distribution

a) A random variable X which is normally distributed with mean μ and variance σ^2, is denoted by $X\sim \mathrm{N}(\mu,\sigma^2)$.

b) The distribution is bell shaped and symmetrical about the vertical line $x=\mu$.

c) Mean = Median = Mode $=\mathrm{E}(X)=\mu$

d) On any Normal curve, the mean determines the position of the axis of symmetry and the standard deviation gives the spread.



3 Standard Normal Distribution

a) A Standard Normal distribution is a Normal distribution with mean 0 and variance 1, and is denoted by $Z\sim \mathrm{N}(0,1)$.

b) Given any Normal distribution X with mean μ and variance σ^2, i.e. $X\sim \mathrm{N}(\mu,\sigma^2)$, X can be standardised to the normal distribution Z with mean 0 and standard deviation 1.

i.e. $Z\sim \mathrm{N}(0, 1)$, using $Z=\dfrac{X-\mu}{\sigma}$.

Then, $\mathrm{P}(x_1<X<x_2)=\mathrm{P}\left(\dfrac{x_1-\mu}{\sigma}<Z<\dfrac{x_2-\mu}{\sigma}\right)$.

4 Inverse Normal Distribution

For any Normal distribution $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$, the Inverse Normal function is used to find *a* where $\mathrm{P}\left(X < a\right) = p$

5 Linear combinations of Normal Distributions

a) If $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$, then $aX + b \sim N\left(a\mu + b, a^2\sigma^2\right)$.

b) If $X \sim \mathrm{N}\left(\mu_x,\ {\sigma_x}^2\right)$ and $Y \sim \mathrm{N}\left(\mu_y,\ {\sigma_y}^2\right)$ are two independent random variables, then $X + Y \sim \mathrm{N}\left(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2\right)$ and $X - Y \sim$ $\mathrm{N}\left(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2\right)$.

c) If $X_1,\ X_2,\ ...,\ X_n$ are *n* independent observations of $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$, then $X_1 + X_2 + ... + X_n \sim \mathrm{N}\left(n\mu, n\sigma^2\right)$.

<u>Checklist</u>

I am able to

☐ understand that a normal distribution is used to model a continuous random variable

☐ know the properties of a normal curve and hence able to determine whether a given random variable follows the normal distribution

☐ understand that the total area under a normal curve is 1

☐ understand that the area under a normal curve between $x = a$ and $x = b$ is the probability $\mathrm{P}\left(a \le x \le b\right)$

☐ use a GC to find probabilities associated with normal distributions

☐ use the invNorm function in GC to find the value of *x* given the probability

☐ know when to standardise a normal random variable so that the mean is 0 and standard deviation is 1

☐ solve problems where the mean and/or the standard deviation of a normal distribution is unknown

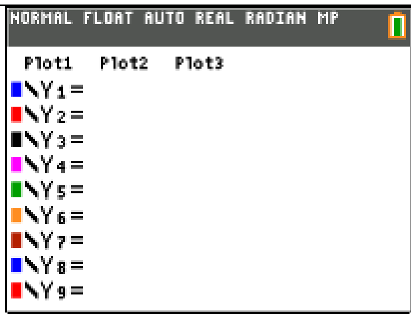
☐ know the properties of expectation and variance for any continuous random variable

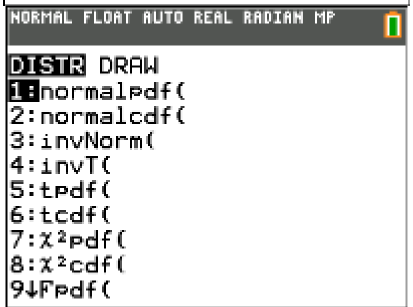
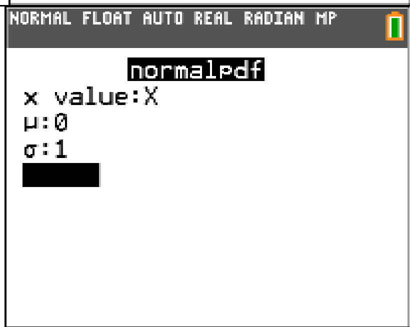
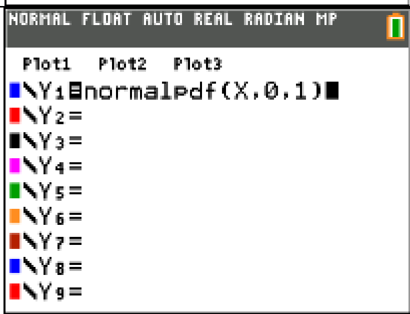
☐ know the distribution of
multiple of a normal random variable;
sum and difference of independent random variables and
sum of *n* independent observations of a normal random variable

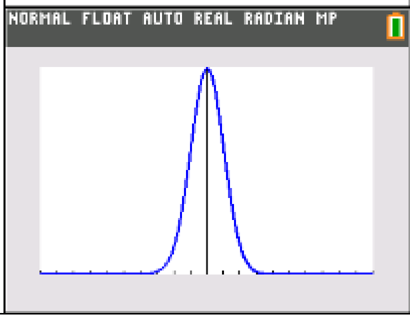
<u>4.9 Learning Experience</u>

(A) Use of GC to plot Normal Curves

We can use GC to plot the graph of $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$. For example, we can use GC to plot the standard normal curve, ie. $Z \sim \mathrm{N}(0,1)$ by follow the steps outline below.

<table>
<tr><th>Steps</th><th>Screenshot</th></tr>
<tr><td>Press [Y=], [2nd][vars], [1]</td><td>


</td></tr>
<tr><td>Key in the relevant values</td><td>
</td></tr>
<tr><td>Press [enter], [graph], [zoom] and [ZoomFit]</td><td>


</td></tr>
</table>

(B) Properties of Normal Distribution

1. Sketch the following normal curves with the given mean and standard deviation on the same diagram using your GC.
   (You may need to adjust the WINDOW for proper viewing of the three curves)
   (a) $\mu = 1,\ \sigma = 1$
   (b) $\mu = 2,\ \sigma = 1$
   (c) $\mu = 3,\ \sigma = 1$

2. By observing the three graphs and the given values of mean and standard deviation, what conclusion can you draw with regard to the shape and the location of the graphs?

3. Sketch the following normal curves with the given mean and standard deviation on the same diagram using your GC.
   (You may need to adjust the WINDOW for proper viewing of the three curves)
   (a) $\mu = 1,\ \sigma = 1$
   (b) $\mu = 1,\ \sigma = 1.5$
   (c) $\mu = 1,\ \sigma = 2$

4. What do you observe of the Normal graphs when the standard deviations are smaller?

5. By observing the three graphs and the given values of mean and standard deviation, what conclusion can you draw with regard to the shape and the location of the graphs?

6. What are the factors that determine the shape of the Normal Curve?