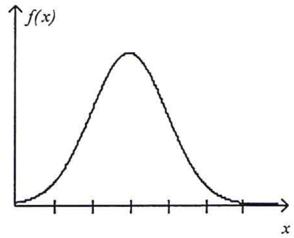
15.3 The normal distribution

Let's look at the investigation on page 538.

Your histogram from the investigation is probably roughly symmetrical and the curve is bell-shaped with the majority of measurement around a central value.



If more measurements were taken, a histogram plotted and the midpoints of the tops of the bars joined with a curve, then it could become more symmetrical and bell-shaped until it would look like the curve shown.

This is the normal distribution.

- Probably the most important distribution is statistics
  - It is a suitable model for many naturally occurring variables.
  - Such as physical attributes of people, animals and plants
  - Can be applied as an approximation for test scores, times to complete a project, IQ scores, etc

- In each case the curve is bell-shaped
- It is symmetrical about the mean μ
- The mean, mode and median are the same.

Characteristics of any normal distribution

There is no single normal curve, but a family of curves, each one defined by its mean, μ, and standard deviation, σ.

> If a random variable, X, has a normal distribution with mean μ and standard deviation σ, this is written $X \sim \mathrm{N}(\mu, \sigma^2)$.

> Note that in the expression $X \sim \mathrm{N}(\mu, \sigma^2)$, σ^2 is the variance.
>
> Remember that the variance is the standard deviation squared.

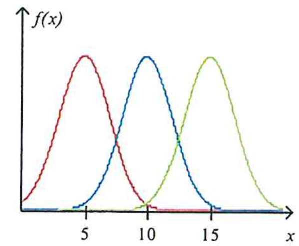
μ and σ are called the **parameters** of the distribution

- The mean is the central point of the distribution.

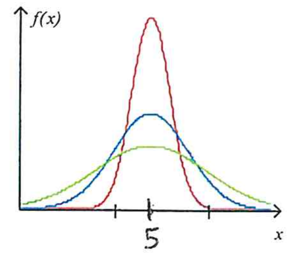
- The standard deviation describes the spread of the distribution.
  - The higher the standard deviation, the wider the normal curve will be.

These three graph show $X_1 \sim N(5, 2^2)$, $X_2 \sim N(10, 2^2)$, and $X_3 \sim N(15, 2^2)$.

The standard deviation are all the same, so the curves are all the same width but $\mu_1 < \mu_2 < \mu_3$



The three graphs show $X_1 \sim N(5, 1^2)$, $X_2 \sim N(5, 2^2)$, and $X_3 \sim N(5, 3^2)$.



Here the means are all the same and all the curves are centered around this but $\sigma_1 < \sigma_2 < \sigma_3$ so curve X_1 is narrower than X_2, and X_2 is narrower than X_3.

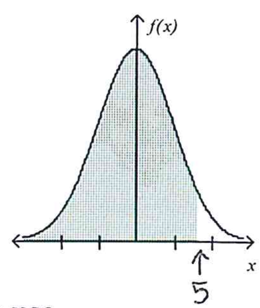
**The curves may have different mean and/or different standard deviations but they all have the same characteristics.

The area beneath the normal distribution curve

NO matter what the value of μ and σ are for a normal probability distribution,

- The total area under the curve is always the same
- **AND equal to 1** (FYI the area under any distribution curve being normal or not is 1)

- We can therefore consider partial areas under the curve as representing probabilities.



So in this normal distribution we could find the probability $P(X < 5)$ by finding the shaded area on the diagram.

Unfortunately (OH shucks!!) the probability function (the equation of the curve) for the normal distribution is very complicated and difficult to use.

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{\left(\frac{-(X-\mu)^2}{2\sigma^2}\right)} \qquad -\infty < X < \infty$$

It would be too hard for us to use integration to find areas under this curve! (especially since most of us haven't had calculus yet) However, there are other methods we can use.

The standard normal distribution {Z = # of standard deviations from mean}

The **standard normal distribution** is the normal distribution where $\mu = 0$ and $\sigma = 1$.

- The random variable is called Z.
- It uses "z-values" to describe the number of standard deviations any value is away from the mean. *

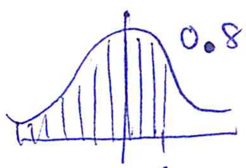
The standard normal distribution is written $Z \sim N(0,1)$

We can use a Z-table or GDC to calculate the areas under the curve of $Z \sim N(0,1)$ for value between a and b and hence $\mathrm{P}(a < Z < b)$.

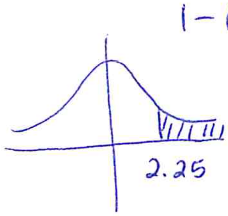
Example: Given that $Z \sim N(0,1)$

Use the z-table to find…

a. $P(Z<1)$	b. $P(Z>2.25)$	c. $P(-.5<Z<2)$



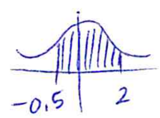
0.8413



$1-P(Z<2.25)$

$1-0.9878$

0.0122

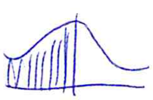


$P(Z<2) - P(Z<-0.5)$

$0.9772 - 0.3085 = 0.6687$

Use the GDC to find normalcdf(min, max, μ, σ)

d. $P(Z<0)$	e. $P(Z>-1.5)$	f. $P(\lvert Z\rvert > 0.2)$



normalcdf$(-10000, 0, 0, 1)$

or normalcdf

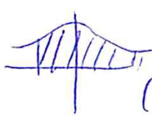
lower: $-10{,}000$

upper: 0

μ: 0

σ: 1

$\boxed{0.500}$



normalcdf$(-1.5, 10000, 0, 1)$

or normalcdf

lower: -1.5

upper: 10000

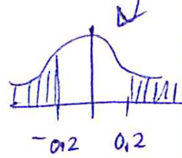
$\mu = 0$

$\sigma = 1$

$\boxed{0.933}$

$1 - P(-0.2 < z < 0.2)$

Note that $P(Z=a)=0$. You can think of this as a line having no width and therefore no area. This means that
$P(a<Z<b)=P(a\le Z\le b)$
$=P(a<Z\le b)=P(a\le Z<b)$



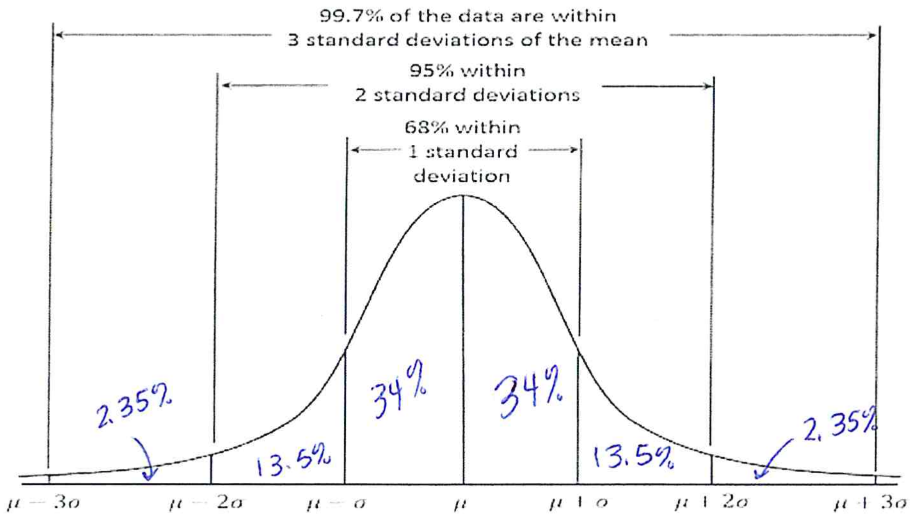
$1-\text{normalcdf}(-0.2, 0.2, 0, 1)$

[illegible]

.8415

Exercise 15H

In question 1 of Exercise 15H you will find the probability that Z lies within one standard deviation of the mean, two standard deviations of the mean and three standard deviations of the mean respectively. You can now see that most of the data for a normal distribution will lie within three standard deviations of the mean. The diagram below represents what is known as the Empirical Rule.



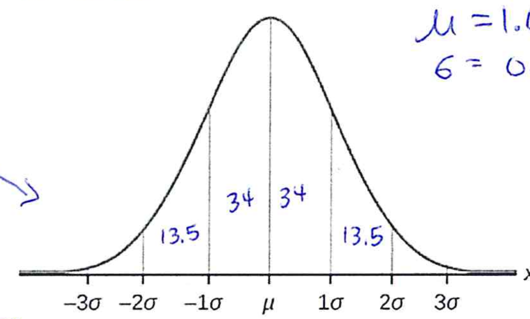
This can be used to compute some probabilities.

Example: The heights of 250 twenty-year-old women are normally distributed with a mean of 1.68m and a standard deviation of 0.06m.

$n = 250$

$\mu = 1.68$
$\sigma = 0.06$

a. Sketch a normal distribution diagram to illustrate this information, clearly indicating the mean, and the heights within 1, 2, and 3 standard deviations of the mean.



1.5 1.56 1.62 1.68 1.74 1.8 1.86

b. Find the probability that a woman has a height between 1.56m & 1.74m.

$13.5 + 34 + 34 = 81.5\%$ or $.815$

c. Find the expected number of women with a height greater than 1.8m.

less than 1.8 = $50 + 34 + 13.5$
$= 97.5$

$250(.975) = 243.75 = 243$ women

Probability for other normal distributions

Very few real-life variables are distributed like the standard normal distribution (with a mean of 0 and a standard deviation of 1.

But you can transform any normal distribution $X \sim \mathrm{N}(\mu, \sigma^2)$ to the standard normal distribution, because all normal distributions have the same basic shape but are merely shifts in location and spread.

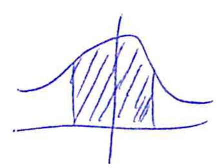
To transform any given value of *x* on $X \sim \mathrm{N}(\mu, \sigma^2)$ to its equivalent *z*-value on $Z \sim N(0,1)$ use the formula $z = \dfrac{x-\mu}{\sigma}$.

We can use the *z*-table and or GDC to find the required probability.

If $X \sim \mathrm{N}(\mu, \sigma^2)$ then the transformed random variable $Z = \dfrac{X-\mu}{\sigma}$ has a standard normal distribution.

Example: The random variable $X \sim \mathrm{N}(10, 2^2)$ $\mu = 10$ $\sigma = 2$

a. Find $P(9 < X < 10.5)$ using the *z*-table.



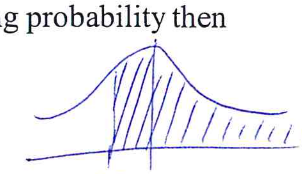
$z = \dfrac{x-\mu}{\sigma}$ $\dfrac{9-10}{2} = -0.5$ $\dfrac{10.5-10}{2} = 0.25$

$P(-0.5 < z < 0.25) = P(z < 0.25) - P(z < -0.5)$
$0.5987 - 0.3085 = 0.2902$

b. Using your GDC use the Z-scores to find the above probability.

$P(-0.5 < z < 0.25)$ normalcdf$(-0.5, 0.25, 0, 1) = 0.2902$
$\uparrow$ min $\uparrow$ max $\uparrow$ μ $\uparrow$ σ

c. You can also use the numbers given. Draw a picture for the following probability then use the *z*-table and your GDC to find $P(X > 8.3)$



8.3

$\dfrac{8.3-10}{2} = -0.85$

$P(z > -0.85)$

$1 - P(z < -0.85)$

$1 - 0.1977 = 0.8023$

normalcdf
$(8.3, 10000, 10, 2)$
$\uparrow$ min $\uparrow$ max $\uparrow$ μ $\uparrow$ σ

0.8023

Exercise 15I

Example:

Eggs laid by a chicken are known to have the mass normally distributed, with mean 55 g and standard deviation 2.5 g. What is the probability that…

a. An egg weighs more than 59 g (Use the table)
b. An egg is smaller than 53 g (use the z-scores and the GDC
c. An egg is between 52 and 54g (use your GDC with these numbers)

Let W = weight of egg

$W \sim N(55, 2.5^2)$

ⓐ $P(W > 59)$ $\quad \dfrac{59-55}{2.5} = 1.6$

$P(z > 1.6) = 1 - P(z < 1.6) = 1 - 0.9452 = 0.0548$

ⓑ $P(W < 53)$ $\quad \dfrac{53-55}{2.5} = -0.8$

$P(z < -0.8) =$ normalcdf $(-10000, -0.8, 0, 1) = 0.2119$

ⓒ $P(52 < W < 54) =$ normalcdf $(52, 54, 55, 2.5)$

0.2295

Exercise 15J

The Inverse normal distribution

Here you want to find the value in the data that has a given cumulative probability.

For example a company fills carton of juice to a nominal value of 150 ml. 5% of cartons are rejected for containing too little juice. The owner of the company may wish to find the cut-off point for the minimum volume of a carton.

To do this we can again use the *z*-table or GDC to find this value.

In these examples we will return to the standard normal distribution $Z \sim N(0,1)$.

Remember probability = area under the curve

Example: Given that $Z \sim N(0,1)$…

a. Use the z-table to find a (use it backwards)

1. $P(Z<a)=0.7$

Look in the probabilities for 0.7

$a = 0.52 \rightarrow 0.6985$ ← closer to 0.7

$a = 0.53 \rightarrow 0.7019$

$\boxed{a = 0.52}$

2. $P(Z>a)=0.1$

$1-0.1=0.9$

$P(z<a)=0.9$

Look for closest to 0.9

$\boxed{a = 1.28}$

To use the table to find the standard deviation that corresponds to the given cumulative probability, find the proportion that is closest to the given proportion and choose the appropriate z-score.

on GDC
invNorm(prob, μ, σ)
or invNorm
area = prob

b. Use the GDC to find

3. $P(Z<a)=0.234$

invNorm(0.234, 0, 1)

$\boxed{a = -0.7257}$

4. $P(Z>a)=0.8$

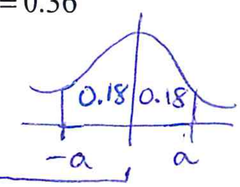
$1-0.8$

$P(z<a)=0.2$

invNorm(0.2, 0, 1)

$\boxed{a = -0.8416}$

5. $P(-a<Z<a)=0.36$



0.5

$P(z<a) = 0.5 + 0.18$

$P(z<a) = 0.68$

$\boxed{0.4677}$

Exercise 15K

Once again, however, we are more likely to be dealing with distributions that are not the standard normal distribution.

Example: Given that $X \sim N(12,3^2)$ determine x where… $\mu = 12$ $\sigma = 3$

(Use the table)

a. $P(X<x)=0.67$

$P(z<z)=0.67$

closest z to 0.67

$z = 0.44$

$$z = \frac{x-12}{3}$$

$$0.44 = \frac{x-12}{3}$$

$$1.32 = x - 12$$

$\boxed{x = 13.32}$

Use the GDC

b. $P(X<x)=0.45$

$x =$ InvNorm(0.45, 12, 3) $= 11.62$

check

$$P(X<11.62) = \frac{11.62-12}{3} = -0.12\overline{6}$$

$P(z<-0.12\overline{6}) \approx 0.45$ ✓

Examples: Cartons of juice are such that their volumes are normally distributed with a mean of 150 ml and a standard deviation of 5 ml. 5% of carts are rejected for containing too little juice. Find the minimum volume, to the nearest ml, that a carton must contain if it is to be accepted. (Find the answer with the tables and with the GDC) Let $J \sim N(150, 5^2)$

$P(J < j) = 0.05$

GDC

invNorm(0.05, 150, 5)

141.776 mL

minimum value

142 mL

table

$z = \frac{J - 150}{5}$

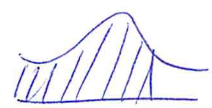
find z with prob 0.05 in table z = -1.645

$-1.645 = \frac{J - 150}{5}$

$J = 141.775$

minimum value: 142 mL

Example: The GPA of a high school (on a non-weighted 4.0 scale) is Normally distributed with a mean of 2.5 and a standard deviation of 0.5. What GPA does Joey have if he is in the 84th percentile of the school? $X \sim N(2.5, 0.5^2)$



84

calc

invNorm (0.84, 2.5, 0.5)

2.997

table

look for prob 0.84 in table... z = 0.99

$.99 = \frac{X - 2.5}{0.5}$

2.995

Exercise 15L

You may also be given cumulative probabilities and asked to find either the mean (if σ is known) or the standard deviation (if μ is known) or both.

Example: Sacks of potatoes with mean weight 5kg are packed by an automatic loader. In a test it was found that 10% of bags were over 5.23 kg. Use this information to find the standard deviation of the process.

$\left(z > \frac{5.23 - 5}{\sigma}\right) = 0.1$

so $\left(z < \frac{5.23 - 5}{\sigma}\right) = 0.9$

use calc invNorm (0.9, 0, 1) = 1.28

or look for 0.9 in table → 1.28

$1.28 = \frac{5.23 - 5}{\sigma}$

$1.28 = \frac{0.23}{\sigma}$

$\sigma = 0.180$

Example: The videos at Hollywood Video are roughly normal in length with a standard deviation of 17.25 minutes. 15% of the movies are shorter than 1 hour and 45 minutes. What is the mean of a video at the store?

$\underbrace{\text{1 hour and 45 minutes}}_{105 \text{ minutes}}$

$$\left(z < \frac{105-\mu}{17.25}\right) = 0.15$$

in calc or table invNorm (0.15, 0, 1) = −1.04

$$-1.04 = \frac{105-\mu}{17.25}$$

$$-17.94 = 105 - \mu$$

$$\mu = 122.94$$

Example: A manufacturer does not know the mean and standard deviation of the diameters of ball bearings she is producing. However a sieving system rejects all ball bearing larger than 2.4 cm and those under 1.8 cm in diameter. It is found that 8% of the ball bearing are rejected as too small and 5.5% as too big. What is the mean and standard deviation of the ball bearing produced?

under 1.8

$$P\left(z < \frac{1.8-\mu}{\sigma}\right) = 0.08$$

invNorm (0.08, 0, 1)

$$-1.41 = \frac{1.8-\mu}{\sigma}$$

$$-1.41\sigma = 1.8 - \mu$$

$$\mu = 1.8 + 1.41\sigma$$

over 2.4

$$P\left(z > \frac{2.4-\mu}{\sigma}\right) = 0.055$$

$$P\left(z < \frac{2.4-\mu}{\sigma}\right) = 0.945$$

invNorm (0.945, 0, 1)

$$1.60 = \frac{2.4-\mu}{\sigma}$$

$$1.60\sigma = 2.4 - \mu$$

$$\mu = 2.4 - 1.6\sigma$$

Set the equations equal

$$1.8 + 1.41\sigma = 2.4 - 1.6\sigma$$

$$3.01\sigma = 0.6$$

$$\boxed{\sigma = 0.199}$$

Plug that in to find μ

$$\boxed{\mu = 2.081}$$

Exercise 15M