Student Learning Centre

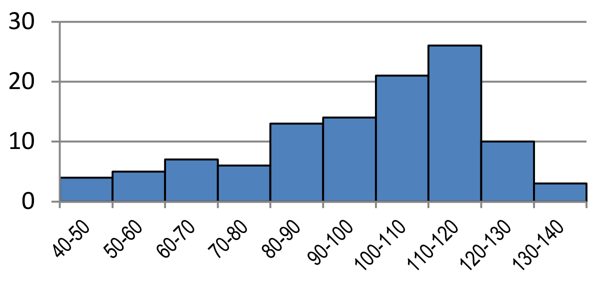
# The Normal Distribution

Flinders University
inspiring achievement

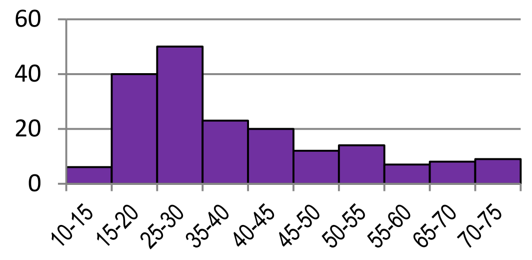
## What is the normal distribution?

The normal distribution refers to a particular way in which data is spread out or distributed. Data can be spread out in many ways. For example:

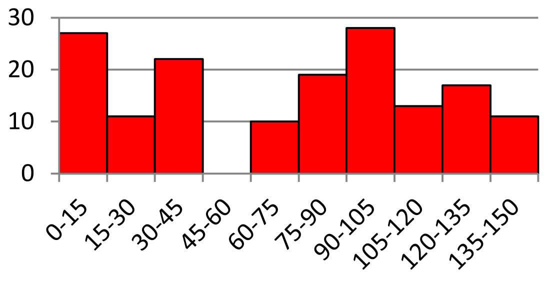
**Data can be spread out to the right**



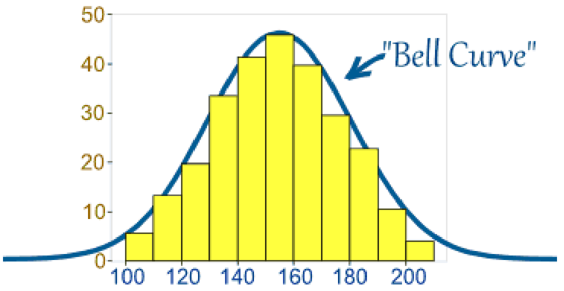

**Data can be spread out to the left**




**Data can be all jumbled up**




But in many situations, data tends to be spread around a central value with no bias left or right, and it gets close to a "Normal Distribution" like this:




So, a normal distribution is a distribution or spread of data where the data is evenly spread in a bell-shaped curve around the mean. In a normal distribution:

- Fifty percent of data is above the mean, and fifty percent below it.
- The distribution of data is symmetric about the mean.
- The mean, median and mode are all equal.

Many distributions are not exactly normal, but approximate a normal distribution, as shown in the graph above, where the histogram is not exactly symmetrical and bell-shaped, but is roughly symmetrical and bell-shaped. Examples of things that approximate a normal distribution are:

- Heights of people
- Size of things produced by machines
- Blood pressure
- Marks on a test
- Errors in measurement

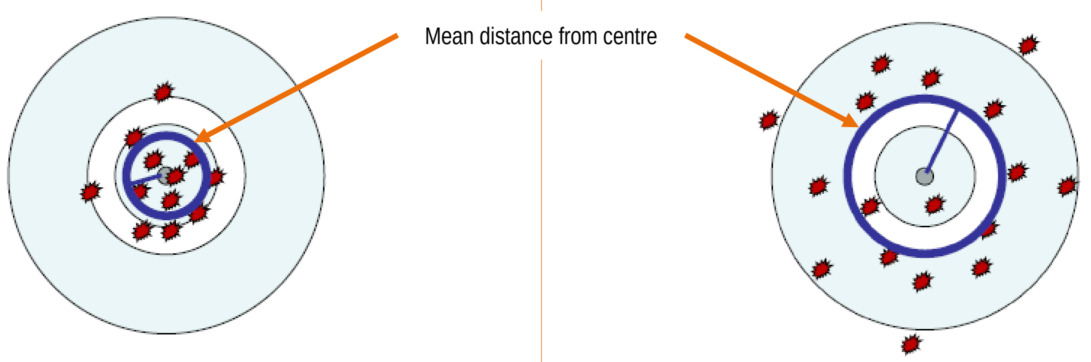
The normal distribution is a continuous distribution, meaning that it describes variables that are continuous. A continuous variable is a variable that can take on any value between two specified values. For example, the measurement of a group of people's heights is continuous because it can be any part of a whole unit: 165.97cm, for example. On the other hand, counting the number of heads/tails in a collection of coin tosses is not continuous (it is discrete) because the result can only be an integer number; it is not possible to have 3.5 heads.

The normal distribution also has other interesting properties, which are discussed in more detail later in this worksheet.

# The Normal Distribution and the Standard Deviation

The standard deviation is a measure of how far individual values are from the mean value. The symbol for the standard deviation is σ (the Greek letter sigma). In both the pictures below the blue ring highlights the mean value. Note that the average distance of arrow shots (shown in red) from the target is greater in Picture B than in Picture A. So, the standard deviation is greater in Picture B than in Picture A:

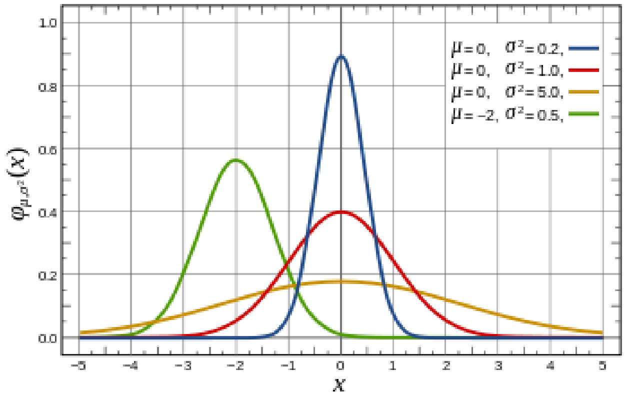
| **Picture A** (smaller standard deviation) | **Picture B** (greater standard deviation) |
| --- | --- |




| | |
| --- | --- |
| Note the hits (red splotches) are mostly quite close to the mean. | Note the hits (red splotches) are mostly further from the mean than hits in Picture B. |

In a normal distribution, the greater the standard deviation, the "flatter" the distribution, as shown in this picture.

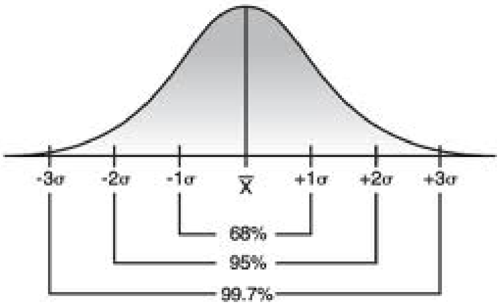
Note that the mean is denoted as $\mu$ and the standard deviation is $\sigma$ (the square root of $\sigma^2$). As $\sigma$ increases, values in the distribution are clustered more closely around the mean, so the distribution appears "taller".




A key property of the normal distribution is that a certain percentage of data values are within 1, 2 and 3 standard deviations of the mean:

- 68% of data is within one standard deviation of the mean
- 95% of data is within two standard deviations of the mean
- 99.7% of data is within three standard deviation of the mean

*This is known as the 68-95-99.7 rule.*




This picture shows a normal distribution and the percentages of data that fall within one, two and three standard deviations ($\pm\sigma$ , $\pm 2\sigma$ , and $\pm\, 3\sigma$ ) from the mean ($\bar{X}$)

# Standardising the Normal Distribution: the Standard Normal Distribution

The number of standard deviations that a value is located away from the mean is also called the "Standard Score", "sigma" or "z-score".

### Formula for a Z-Score

To convert a value to a Standard Score ("z-score"):

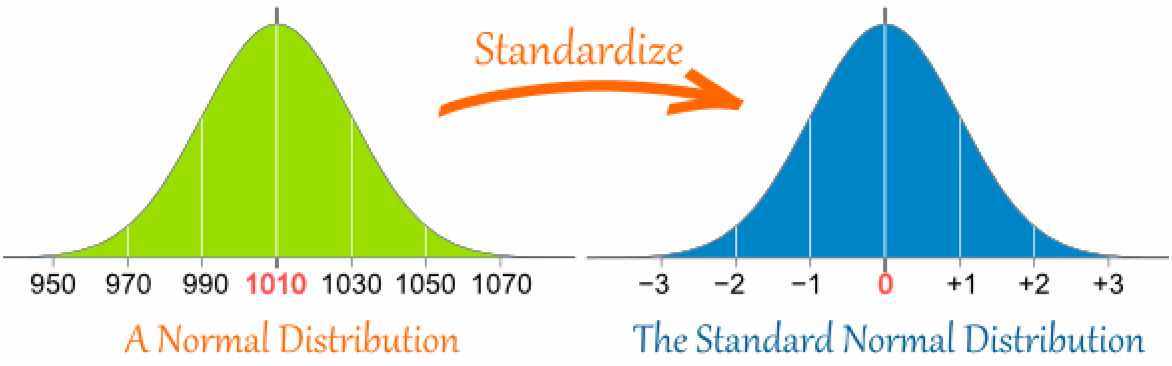
- first subtract the mean
- then divide by the standard deviation

$$z = \frac{x - \mu}{\sigma}$$

where:
x = a particular data value
$\mu$ = mean
$\sigma$ = standard deviation

For example: if the mean test score in a class is 62% with a standard distribution of 8, and Jen has 79%, then her z-score is (79-62)/8 = 2.125. This means Jen's score is roughly two standard deviations above the mean.

Obtaining the standard normal distribution for any distribution means shifting that distribution so that its mean is zero, as shown in the picture below:




Standardizing a normal distribution is useful because is enables values from different normal distributions to be easily compared.

## Practice Questions

Six hundred mathematics students sat for an exam. Their marks were normally distributed with a mean of 63 and standard deviation of 8.

A. If a passing result is 51% or higher and a credit is between 65% and 74% (inclusive), will the proportion of students who fail be more or less than those who get a credit? *(hint: roughly draw the normal curve labelling each standard deviation with the appropriate percentage)*

B. How many students would you expect to have scored:
   i. between 55% and 77%
   ii. more than 69%

C. What would be the z-score of a student who got 78%?

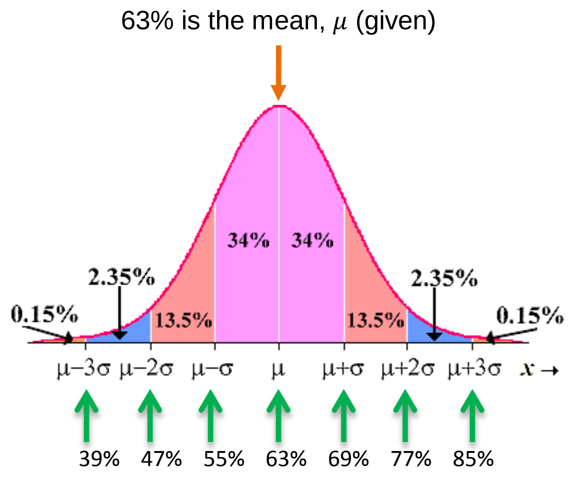
D. In another mathematics exam, the mean was 67 and standard deviation 4. Would a student who got 75% in this exam have done relatively better or worse than the student in part (c)?

Check your answers on the back.

## Solutions to Practice Questions

A. Remember that the 68-95-99.7 rule tells us the percentages of data for a normal distribution within each standard deviation from the mean, and given $\mu$ = 63% and $\sigma$ = 8, you can work out the information you need as follows:

| Point on Normal Distribution | Grade (%) |
|---|---|
| $-3\sigma + \mu$ | 39 |
| $-2\sigma + \mu$ | 47 |
| $-\sigma + \mu$ | 55 |
| $\mu$ | 63 |
| $\mu + \sigma$ | 69 |
| $\mu + 2\sigma$ | 77 |
| $\mu + 3\sigma$ | 85 |




Looking at the graph and estimating, there is clearly a much greater percentage of values from 65% to 74% than there is below 51%.

B.

i. Between 55 and 77% lies 34 + 34 + 13.5 = 81.5% of results. There are 600 students, and 81.5% of 600 is (81.5 x 600) / 100 = 489 students.

ii. Above 69% lies 13.5 + 2.35 + 0.15 = 16% of results. There are 600 students, and 16% of 600 is (16 x 600) / 100 = 96 students.

C. Using the formula for z-score with $\mu$ = 63% and $\sigma$ = 8 gives z = (78 – 63)/8 = 1.875. So the student receiving 78% has scored 1.875 standard deviations above the mean score.

D. Using the formula for z-score with $\mu$ = 68% and $\sigma$ = 5 gives z = (75 – 67)/4 = 2. So the student receiving 75% has scored 2 standard deviations above the mean score. Relatively speaking, this is a slightly better score than the student in part C because it is further above the mean.

**Student Learning Centre**
Registry Building Annexe

Tel: 61-8-8201 2518
E-mail: slc@flinders.edu.au

Internet: http://www.flinders.edu.au/SLC
Postal: PO Box 2100, Adelaide, SA 5001