Standardization

A class of pupils took a test in Mathematics and a test in English. Some of the results (marked out of 100) are shown below.

Mathematics (X) 37 39 58 * * * * .... **80** .... $\bar{x}$ = **70** s = **4**

English (Y) 36 42 43 48 * * * ....**75** .... $\bar{x}$ = **63** s = **6**

One pupil named Lonton, achieved a mark of 80 in Maths and 75 in English and she wished to be able to say, "I did better at ……..".

She could say, "At a first glance I did better at Maths because the mark of **80** is better than the mark of **75**.

If she knew the **means**, she could say "I did better at English because my English mark was further from the mean".

Maths $x - \bar{x} = 10$

Eng $x - \bar{x} = 12$

Now, if she also knew the **standard deviations** of the two sets of data she could divide by the respective standard deviations.

Maths $10 \div 4 = 2.5$

English $12 \div 6 = 2$

In Maths, her mark was 2.5 standard deviations from the mean. In English, her mark was only 2 standard deviations from the mean.

She can say "I did better at Maths". Her mark, in standard deviation units, was better in Maths.

All the pupils' marks may be standardised using **z** = $\dfrac{scores - mean}{s\tan dard.dev} = \dfrac{x-\bar{x}}{s}$

This gives us standardised scores from which we can make comparisons between the tests.

The **mean** of the **standardised scores** will be **0** and the **standard deviation** will be **1.**

A certain pupil got 58 in Maths and 50 in English. 58 standardises to $\dfrac{58-70}{4}$ = **-3** 50 standardises to $\dfrac{50-63}{6}$ = **-2.16**

The pupil actually did **worse** in Maths since his mark was 3 standard deviations less than the mean as opposed to 2.16 standard deviations less than the mean.

<u>Probability distributions</u>

When throwing a die, the outcomes may be modelled by a **discrete rectangular distribution** which looks like this:

OUTCOMES (X)	1	2	3	4	5	6
PROBABILITIES (P)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

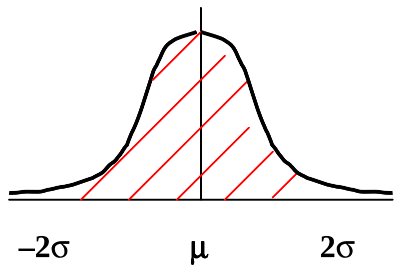
The sum of the probabilities is 1.

The probability of throwing a number greater than six may be calculated by adding Pr(5) + Pr(6) $= \frac{1}{6} + \frac{1}{6} = \frac{2}{3}$

<u>The Normal Distribution</u>

The normal distribution is used as a model for continuous data such as height, weight, time etc. The probability cannot be set out in a table as there are an infinite number of values the variable can take. Instead, it is given by a formula and an associated sketch graph.

f(x) = $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ $-\infty \le x \le \infty$



All normal distributions are symmetric and have bell-shaped probability density curves with a single peak.

The distribution is defined by its mean μ {mew} and its standard deviation σ {sigma}.

Almost all the graph lies within 2 standard deviations of the mean. (About 95%)

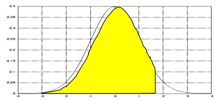
The area under the graph is 1.

So if a particular type of fruit has mean weight 150g and standard deviation 10g. The variance would be $\sigma^2 = 10^2$.

X is the random variable **weight of fruit**. We write: $X \sim N(150, 10^2)$ putting the **mean** and **standard deviation squared** in the bracket.

We now wish to evaluate a probability just like the probability found with the die.

<u>The probability that a fruit picked at random has a weight under 165g.</u>



Pr(x $\leq$ 165): One way to find this probability would be to find the area to the left of 165 by integration.

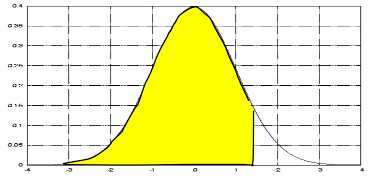
A better way would be to standardize the weight of 165 by subtracting the mean and dividing by the standard deviation, thereby converting to a z – value and looking up this probability in Standardized Normal Tables.

The graph connected to the tables would be the Standardized Normal graph and it would have a mean of zero with a standard deviation of 1.

The value in the left column is the z - value

z	Φ(z)
1.49	0.9319
1.50	0.9332
1.51	0.9342

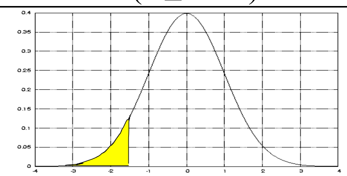
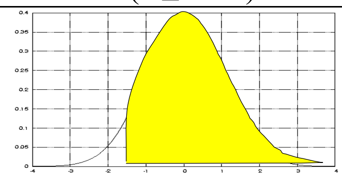
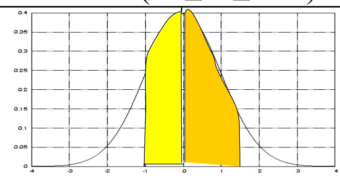
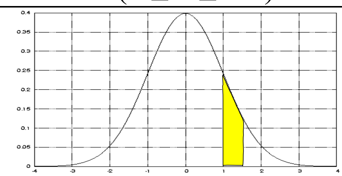
The value in the right column is the probability of a value less than the z - value



$z = \dfrac{165 - 150}{10} = \mathbf{1.5}$ The tables would give us Pr(z $\leq$ 1.5) = **0.9332**

If we want the probability of a fruit weighing over 165g. we look at the other side of the graph. Pr(z $\geq$ 1.5) = 1 - 0.9332 = **0.0668**

<u>Using the Normal tables.</u> All problems may be converted to a probability which may be read off from Normal tables. Typical examples:

	Pr(z $\leq$ - 1.5)	Pr(z $\geq$ - 1.5)	Pr(-1 $\leq$ z $\leq$ 1.5)	Pr(1 $\leq$ z $\leq$ 1.5)
Sketch the graph, bearing in mind the mean is zero.	 . 0	 . 0	 . 0	 . 0
Φ(k) means area to the left of k. Ω(k) is taken as area to the right of k.	Φ(- 1.5) = Ω(1.5) = 1 - Φ(1.5) = 1 – 0.9332 = **0.0668**	Ω(- 1.5) = Φ(1.5) = **0.9332**	We know the area to the left of zero is 0.5. Find the area using two parts Φ(1.5) - 0.5 + Φ(1) - 0.5 = 0.9332 – 0.5 + 0.8413 – 0.5 **= 0.7745**	Φ(1.5) - Φ(1) = 0.9332 - 0.8413 = **0.0919**

Example

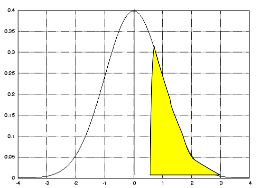
A consultancy undertakes systems development projects. From previous experience it has been ascertained that the mean profit per project is£24,000 with a standard deviation of 8,000. Assuming the Normal distribution can be applied:

(a) What percentage of projects will have a profit in excess of £30,000?

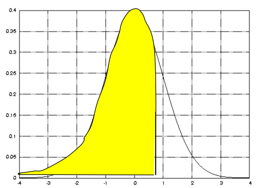
Profit P~ $N(24,000, 8,000^2)$ The question requires $\Pr(p \geq 30,000)$

Convert to a standardized question: $\Pr(z \geq \frac{30,000 - 24,000}{8,000})$ = $\Pr(z \geq 0.75)$

Sketch a graph: N(0, 1) The area is to the **right** so find "1 – the area to the left": 1 - $\Phi(0.75)$ = 1 - 0.7734



1 -



= 0.2266 ie <u>22.66%</u>

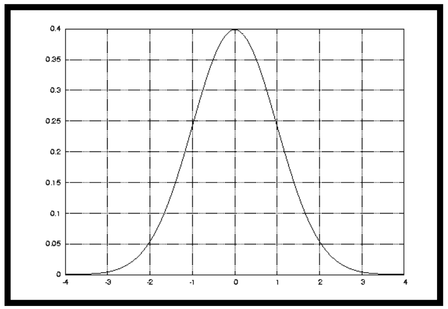
Probabilities for the Normal Distribution

Use this HANDY CALCULATOR to help you calculate normal probabilities.

How-To:

Enter beginning and ending z-scores in the Start and End boxes. The figure is redrawn and the new probability is calculated after hitting the return key while the cursor is in one of the boxes.

For open-ended ranges (i.e., finding the proportion of observations that will lie below a given value or above a given value), use a very large number for the other end (10 is large in this context).



Change the numbers in the Mean and StDev boxes as appropriate and then enter the appropriate Start and End values. Remember to hit the Return key in the last box you enter to start the calculations and redrawing.