A demonstration: the Central Limit Theorem

Ian Jacobs: Physics Advisor, KVIS, Rayong, Thailand.

The *Central Limit Theorem* lies at the foundation of statistics and explains why random errors follow a normal (Gaussian) distribution. The point is important because the standard deviation is a well-defined measure of likely error that can be carried through calculations by adding errors and fractional errors in quadrature.

> The theorem states: *when independent random variables are added, their sum tends toward a Gaussian distribution (commonly known as a bell curve or normal distribution) even if the original variables themselves are not normally distributed.* In other words: data that are influenced by many small, unrelated effects (like the height of boys at KVIS or the mass of a can of beans) are approximately normally distributed. See the History link for details …
> https://en.wikipedia.org/wiki/Central_limit_theorem#History

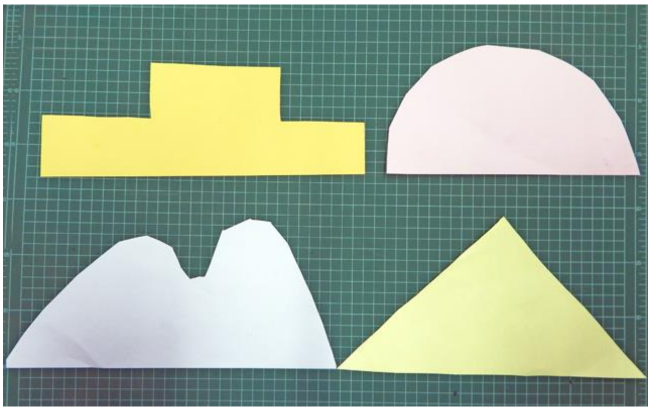

Fig 1 – four non-Gaussian probability distributions cut from sheets of coloured paper.

Twenty sheets of paper, five for each distribution, were cut into strips. The piles of strips were mixed. Random selections of more than one hundred strips of each colour were then arranged on the bench in lines in random order.

Fig 2 – paper strips.

Note: Random events tend to cluster. A truly random selection is difficult to achieve by hand because the natural tendency is to select for a uniform (or non-uniform) distributions. Selections from the piles of strips were made without taking lengths into consideration and the selected strips were then arranged in lines with their lengths obscured so that the selector was not aware of length when placing the strips in line. The lines were then combined into selections of four strips of different colours.

Fig 3 – a section of the lines of four-strip samples.

The lengths of the four-strip samples were measured to the nearest cm and the distribution of sample lengths was plotted in Logger Pro. Figure 3).

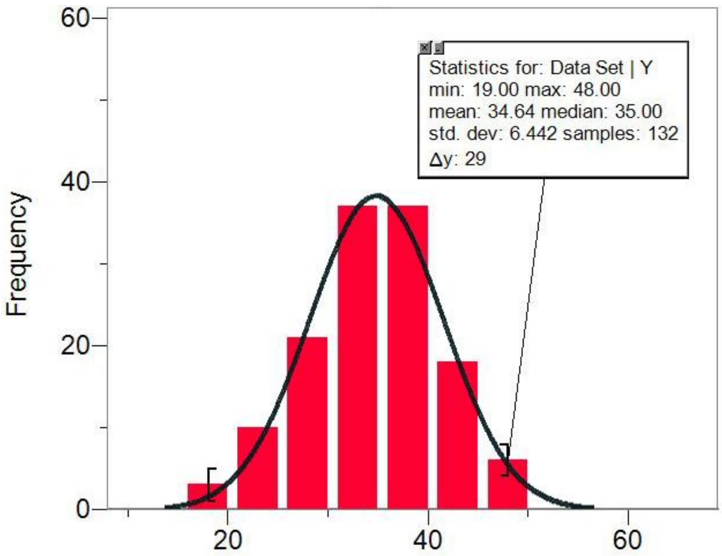

Fig 4 – the histogram shows the distribution of lengths of 132 four-strip samples.

A Gaussian function has been overlaid on the histogram (figure 4) by hand. With a bin size of 5 cm and 132 samples the distribution is close to *normal*, in spite of the four component distributions being very different.

The demonstration illustrates the truth of the central limit theorem in practice and shows why we can be confident that random errors are normally distributed.