Statistical measures

6.13 There are many statistics that can be calculated based upon a distribution; however, most are esoteric and are unlikely to contribute much to business people's understanding of risk. The table below lists the most common statistics, and explains when they might be useful:

Statistic	Definition	Use	Dangers
Mean (expected value)	The average of all the generated outputs	Very useful, this is one of the two most common statistics reported. For example the average NPV of a transaction. It also has the useful property that if two (or more) variables are independent, then: mean(a+b)=mean(a)+mean(b), & mean(a*b)=mean(a)*mean(b).	Confusing the mean with the most probable (mode)
Standard deviation (σ)	The square root of the variance (see later under variance)	Another very useful statistic, it gives a measure to the dispersion around the mean of a distribution. It is frequently used in conjunction with normal distributions to give the level of certainty that a value lies within a certain distance from the mean: +/- σ of the mean = 68% +/- 2σ of the mean = 95% +/- 3σ of the mean = 99.7% So, for example, a normally distributed variable with a mean of 1.0 and a σ = 0.05 can be said to have a 95% certainty of lying between 1.1 and 0.9.	Assuming that the standard deviation of the sum of independent components is the sum of the separate standard deviations! In fact it is the square root of the sum of the squares: $\sigma^2_{TOT}=\sigma^2_1 + \sigma^2_2$ The relationship given above is only valid if the distribution is symmetrical. It becomes more of an approximation the more skewed the distributions are.
Variance (V)	The variance is calculated by determining the mean of a set of values, and then summing the square of the difference between the value and the mean for each value: $V = \frac{\sum_{i=1}^{n} (x_i\text{-mean}(x))^2}{(n-1)}$	This is also a measure of the dispersion around the mean, however it is in the units of a quantity squared. Thus the variance of a distribution in NPV (in £s) will be given in £^{2}. The reason it is used is that it is useful for estimating the widths of a sum or multiple of several independent variables: V(a + b) = V(a) + V(b), & V(a * b) = V(a) * V(b). The risks in a model have been successfully disaggregated if the variances of the different significant risks are similar.	As with the standard deviation, the relationships shown to the left are only valid if the distribution is symmetrical. It should be noted that the variance (and thus the standard deviation) is much more sensitive to the values at the tails of the distribution than those close to the mean.
Median	The median is the value at which there is an equal percentage chance of a being above it as below it. In other words, it is the 50th percentile.	Rarely used as it gives no indication as to the range of the values above it or below it. If the mean is not equal to the medium, then the distribution is skewed.	Confusing the median with the mean or mode.
Percentiles	The nth percentile of a variable is that value for which there is a n% chance of the variable lying at or below that value.	A useful concept, used in measuring the range of a variable. For example the range of a distribution might be defined as the difference between the 5th and 95th percentile; this means the width of a distribution if the top 5% and bottom 5% of all values are ignored. It can also be used to answer questions like "What are the chances that the IRR is below 9%?". The answer would be the percentile for which the value was 9%.	Can be confusing as to exactly what is meant, so it is usually a good idea to explain the concept of percentiles in layman's terms.

Statistic	Definition	Use	Dangers
Mode	The most likely value. For a discrete distribution this is the value with the greatest observed frequency, and for a continuous distribution the point of maximum probability.	Sometimes used to describe a Poisson like distribution: the mode is the most likely number of event to occur in the given time period (and is approximately given by the reciprocal of the rate). Also used in describing triangular distributions (the minimum, the mode and the maximum). In general it has little value in uncertainty and risk analysis.	It is difficult to determine precisely, particularly is a distribution is unusually shaped.
Skewness (S)	$S = \frac{\sum_{i=1}^{n} (x_i\text{-mean}(x))^3}{\sigma^3}$	This is a measure of the 'lopsidedness' of a distribution. It is positive if a distribution has a longer right tail (and negative if a more prominent left tail). A zero skewness means the distribution is symmetric. Apart from a general measure it is used to determine how 'normal' a distribution, the closer a distribution is to having a skewness of zero, the more normal it is. Examples of skewness: the skewness of normal distribution is 0, triangular distributions vary between 0 and 0.56, and an exponential distribution has a skewness of 2.	The skewness is even more sensitive to the points in the tail of the distribution than the variance. It therefore requires many iterations to be run before it reaches a stable value.
Kurtosis (K)	$K = \frac{\sum_{i=1}^{n} (x_i\text{-mean}(x))^4}{\sigma^4}$	The kurtosis is a measure of the 'peakedness' of a distribution. Examples of kurtosis: uniform distribution has a kurtosis of 1.8, a triangular distribution 2.4, a normal 3, and an exponential has a kurtosis of 9. If a distribution is approximately bell shaped, and has a skewness of around 0 together with a kurtosis of close to 3, then it can be considered normal.	Stable values of the kurtosis often require even more iterations to be run then skewness. For example a randomly sampled normal distribution required approximately 1500 iterations to be within 2% of 3.
Coefficient of variability (normalised standard deviation) (σ_n)	This is defined as the standard deviation divided by the mean: $\sigma_n = \sigma$ / mean	This is a dimensionless quantity that allows you to compare, for example, the large standard deviation of a large variable with the small standard deviation of a small variable. An example would be comparing the level of fluctuation with time between different currencies.	This is not a meaningful statistic to compare if the mean and standard deviation are unlikely to bear any relation with each other. An example would be the NPV of a project. Here the spread need not be related to the mean value, which could be close to zero. An extreme would be the coefficient of variability of a normal distribution that is centred on zero.
Mean standard error.	This is included purely as it is one of the statistics provided by Crystal Ball.	Crystal Ball calculates this as a measure of the accuracy of the simulation, and whether enough iterations have been run. Specifically it tells you the likely difference between the estimated mean and the actual mean, to a certainty level of 68%.	This is precise **only** for the accuracy of the mean. The accuracy's of the other statistics such as the standard deviation or any percentile value are likely to be considerably less then this figure implies. It should be used only as indicative in rough and ready simulations. For more detail on this subject, see the sub-section on 'how many iterations are necessary' under risk analysis.

Figure15: Table of common statistics