Technical brief: Decision-making with statistics

Many of us loathed statistics at school – but understanding statistics can help marketers make better decisions. In this occasional series of Technical briefs, we explain some essential concepts and how they are applied.

In a previous briefing note, we described how the concept of statistical significance indicates our level of certainty that a result is "real". We say that a result – for example a difference in response rates to a test mailing vs control - is "statistically significant at the 5% level" if there is less than a 5% probability that it arose just by chance. However, we are dealing with probabilities not certainties, and there is still that small possibility that it is just sampling error. Conversely, it is possible to conclude that a result is not significant, when in fact there is a genuine effect. This paper explores those risks and the factors that affect them.

Table 1 illustrates four possible outcomes, depending on whether a test is significant, and whether there is a true effect. Two of these are straightforward - Yes/Yes and No/No – where the decision made on the basis of the significance test plays out as expected in a full scale implementation (although you never verify that with a No/No.....).

The false positive is where the trial says the result is significant – less than 5% probability of it happening by chance – but in fact there is no real effect. The low probability event has happened, and misled us into believing the effect is real. This is known as a Type I error, or α error. This might lead, for example, to the decision to scale up a test mailing, only to find at full scale that the results are no better than the control.

The false negative is the opposite type of error, known as a Type II or β error. That is the one that got away: statistical analysis said we can't rule out that the result was just chance – not significant – but in fact there was a real effect. These are the missed opportunities, such as the campaign that was never run because it didn't look any better than what we already had.

Minimising the risk

Clearly, we want to minimise the chance of making either of these errors. Three concepts are helpful in understanding the issues, but there is no black-and-white answer, and the best approach will vary with the type of decision being made.

- **Sample size**: both types of error are reduced by increasing the number of people in the trial - but practicality intervenes; the need for certainty has to be balanced with the cost and timing.
- **Significance level**: the probability of a false positive is the significance level, which is within your control – if you set (for example) the conventional 5%, you are taking a 1 in 20 chance that you will declare an effect when there isn't one. You could choose a more stringent 1%, or a more relaxed 10%, depending on the type of decision. For example, a marketing initiative that is particularly high cost, or risky in some way, might require a higher level of certainty than a minor cost-saving.

  Importantly, though, the two error types are closely linked. Tightening the significance level to minimise false positives, increases the chance of missing a good one. This might not be problematic in a test of a campaign design – particularly if your creative team is prolific! – but unacceptable in an industry where new ideas are hard-won, for example, or in medical diagnostics, where a missed case has life-changing consequences for the patient.
- **Using confidence intervals**: statistical significance is often used as a cut-off – either a result is or it isn't. However, in terms of the underlying probabilities – the probability of observing that test result if in fact there's no difference from control - it clearly doesn't make sense for probabilities of, say, 4.5% and 5.5%, to drive opposite conclusions. For this reason, it is helpful to consider the actual probability (referred to as the p-value) rather than a simple greater than/less than test, and to calculate a confidence interval – the range of values within which there is (e.g.) a 95% probability that the true value lies.

Table 1: possible outcomes from significance testing

		Actual effect? Yes	Actual effect? No
Trial is significant	Yes	√	False positive
	No	False negative	√