

5.3. Accounting for multiple comparisons

Thus far, we have assumed that we are investigating two categories of an explanatory variable or experimental treatment (i.e. comparing a treatment group with a control group). However, the objective may instead be to compare multiple levels of an explanatory variable (e.g. different concentrations of a pesticide) or multiple independent kinds of the same sort of explanatory variable (e.g. competing manufacturers of protein substitutes). In addition, one may be interested in testing multiple explanatory variables at the same time (e.g. effects of three different humidity levels and honey bee age on susceptibility to the tracheal mite *Acarapis woodi*). More complex statistical models warrant increased sample sizes for all treatments. Consider the case where one has one control and one treatment group; there is a single comparison possible. Yet if one has one control and 9 treatments groups, there are $9 + 8 + \dots + 1 = 55$ possible comparisons. If one rigorously follows the cut-off of $P = 0.05$, one could obtain $0.05 * 55 = 2.8$ significant results by chance or in other words the probability of at least one significant by chance alone is $1 - 0.95^{55} = 0.9405$, so one is likely to incorrectly declare significance at least once (in general, 5% of statistical results will have $p \geq 0.05$ if there are no true differences among treatments, this is what setting $\alpha = 0.05$ represents). *Post hoc* tests or *a posteriori* testing, such as Bonferroni corrections, attempt to account for this excessive testing, but in so doing can become very conservative, and potentially significant results may be overlooked (i.e. correctly control for Type I error, but have inflated Type II errors; Rothman, 1990; Nakagawa, 2004). Less conservative corrections, such as the False Discovery Rate, are now typically favoured as they represent a balance between controlling for Type I and Type II errors (Benjamini and Hochberg, 1995). Other ways to avoid or minimise this problem include increasing sample size and simplifying experimental design by reducing the number of treatments and variables.