**<http://www.une.edu.au/WebStat/unit_materials/c7_anova/oneway_general_comments.html>**

**Introduction to Analysis of Variance**

Analysis of variance (ANOVA) is a popular tool that has an applicability and power that we can only start to appreciate in this course. The idea of analysis of variance is to investigate how variation in structured data can be split into pieces associated with components of that structure.

We look only at one-way and two-way classifications, providing tests and confidence intervals that are widely used in practice.

**One-way analysis of variance**

One-way analysis of variance looks to see how much of the variation in grouped data comes from differences between the groups, and how much is just random observational error. There can be any number of groups, that may be of different sizes (each group with at least two observations).

A typical application of one-way analysis of variance would be to investigate whether three different types of growing conditions make any difference to the yield of an agricultural crop, and if so, how great those differences are.

The observations would be the yields of many different experimental plots, grouped according to the growing condition that applied to them.

To describe analysis of variance accurately one needs a lot of notation; it is annoying, but worth spending time on. Suppose that we have n random observations classified into k different groups, so that there are ni observations in group *i* for *i = 1, 2, 3. . . , k*.

We shall assume that all the observations are independent of each other, and that the distribution from which those in group *i* are taken is N(μi 1*,* , σ2).

Notice that the population mean μi may be different for each group, but that the variance is the **same** for all observations in all groups.

Since we assume that the observations are from normal distributions, they are not counts - so application of analysis of variance to tables of counted data (contingency tables) will not usually be possible. For counted data the variance is often proportional to the mean, and so not the same for all the counts.

Outliers or wild observations also invalidate the assumption of normal distributions, so one needs to check there are none of those (for instance a wrong recording of 8 instead of 80 ).

***Example:*** We will assume that the observations are independent random observations from normal populations with means μ1, μ2, μ3, μ4, and with the same variance σ2. There are no obvious outliers or wild observations that make this assumption look wrong.