Why do randomisation tests in the experiments standard?

The biggest ideas in the experiments standard AS 3.11 are about using random allocation of experimental units (e.g. people) to treatment groups as the most reliable method we know of designing experiments that provide fair tests of causal (cause and effect) claims or hypotheses. The resulting designs are standardly called "randomised experiments". Students have also to conduct randomised experiments and then analyse the resulting data. We need useful and accessible forms of analysis to enable students to experience the whole investigation from questions to conclusions.

Why teach statistical tests at all?

While random allocation is the best method we know for getting "fair comparisons", it is not perfect.

If we simply randomly allocate group labels to units to form artificial "groups" we can get quite large apparent differences between the "groups", even though there are no real differences at all.

So whenever we look at data from randomised experiments we need a means of addressing an obvious objection that argues that the differences we are seeing could quite easily have been produced by chance alone (just the luck of who got what group label). It is important to address this to prevent our prematurely jumping to conclusions.

Why make it the randomisation test?

The randomisation test provides the most easily understood testing mechanism (because it involves the smallest number of concepts) and its validity is also easiest to see (because it has the most direct relationship between the problem we are trying to solve and the solution we use[1]). It has other advantages too. No assumptions need to be made about underlying distributions (e.g. Normal). Neither do we need to worry about lack of correctness when we have small group sizes.

Difficulties with more traditional approaches

The traditional hypothesis testing methods many of us met at university come from random sampling theory.

Most experiments, however, use "convenience samples" – employing people or units the experimenters can easily get their hands on. They are not random samples from any meaningful population.

The use of the theoretical methods on convenience samples is justified only to the extent that they approximate the appropriate randomisation test. Their use became widespread because their mathematical approximations gave a work around for lack of computational power. Understanding the relevance of these approaches to a particular experiment, however, involves many concepts and rather convoluted reasoning.

Side benefits

An excellent side effect of the randomisation test in AS3.11 is that it gives a first introduction to the basic ideas underlying significance (hypothesis) tests and *p*-values in their simplest possible form.

Chris Wild (U. Auckland), Jennifer Brown (U. Canterbury), Anna Martin *(Avondale Col.),* *8 Nov.2012*
http://www.stat.auckland.ac.nz/~wild, http://www.math.canterbury.ac.nz/~j.brown

[1] "Can random allocation of group labels acting alone ('chance alone') explain this?"