Tests of Statistical Significance – their (ab)use in the social sciences – a comment

Many final year students use a questionnaire based survey to collect primary data for their dissertations.   It should be highlighted that the sampling strategy adopted for these surveys is likely to be convenience sampling. These samples commonly comprise of students located on the Collegiate Crescent site on a particular date that were willing to participate by completing a self-completion questionnaire. Whilst this could be viewed as haphazard it cannot make any claims to randomness.   Some students may use quotas to try and boost the representative nature of the sample (in terms of age and gender for example). This is an improvement on accepting what / who happens to be there it still cannot claim any random status.

Because the design of these surveys do not contain any random component (primarily due to practicality as this would require a sampling frame that listed all potential respondents), the analyses conducted will only be reflective of the specific sample under study. Generalisations (for example through tests of statistical significance) are not valid when using such a (non-random) design (De Vaus 1985 p189-197).

A test of statistical significance identifies associations in the sample that are unlikely to be created just through (random) chance. Commonly, if the chance (p-value) of an association in a sample being due to random sampling variation is measured as 5% or below ($p <= 0.05$) the association is identified as statistically significant. It is important to distinguish the clear difference in meaning between statistical significance (a finding is unlikely to be due to chance) and the lay definition of the word significance (a finding is important). A statistically significant finding does *not* necessarily mean it is an important finding.

By drawing upon the theories of statistical inference, random sampling variation is accounted (or controlled) for. This allows systematic pattern (above and beyond what is expected through chance alone) to be identified. If systematic pattern is identified above and beyond what would be expected from random sampling variation it is concluded that the sample finding is likely to be also found in the parent population.   This process of generalisation from random sampled surveys was the conceptual heart of a groundbreaking paper given in 1934 to the Royal Statistical Society (Neyman 1934 discussed in Dale 2006).

Ignorance of this fundamental assumption in statistical inference leads to tests being run on non-random samples. This results in non-existent randomness being accounted for in the analysis and (alarmingly) population generalisations being made from non-random (biased) samples. This will inevitably lead to fallacious research and construction of false 'knowledge'. This concerning trend seems to be on the rise and reflects a low level of statistical literacy in the social science discipline.   Two quotes below emphasise this point:

> *"If the researcher does not use a random sample then traditional (significance testing) statistics are of no use since the probabilities become meaningless. … Researchers using significance tests with convenience, quota or snowball samples, for example, are making a key category mistake"*
> Gorard 2006.

Significance testing is not appropriate for enumerations or non-random samples because it only deals with the chance of Type I error based on a random sample. …. We would like to make inferences for non-random samples, but that is impossible
Garson 2006

Having written this, I must acknowledge that I am perhaps overstating the reliability of random samples. Significance tests also assume 100% response and perfect measurement so that any variations observed are attributed solely to (random) sampling variation. This perhaps reflects the substantive discipline from which significance testing emerged (agricultural experiments) but is clearly a flawed assumption in the social sciences. More about this can be read in the Gorard article referenced below (Gorard 2006).

To conclude, students who undertake primary research using a non-random sample are (**strongly**) advised to adopt a descriptive approach and NOT use ANY tests of statistical significance. They should accept and be honest that their analysis will ONLY be related to patterns within their sample - any attempt to generalise using significance testing will (falsely) overplay the representative nature their data, reveal a fundamental misunderstanding of the survey method and at least to some extent be ethically questionable.

Author: Sean Demack (s.demack@shu.ac.uk)

References

Dale, A (2006) *Quality issues with survey research.* International Journal of Social Research Methodology 9(2) pp143-158

Garson, G.D. *Significance.* Course notes from a module entitled 'Quantitative Research in Public Administration' available online at
http://www2.chass.ncsu.edu/garson/pa765/signif.htm (Last accessed 3/4/07)

Gorard, S. (2006) *Towards a judgement-based statistical analysis.* British Journal of Sociology of Education 27(1) pp67-80

Neyman, (1934) *On the two different aspects of the representative method (with discussion)* Journal of the Royal Statistical Society 97 pp558-625

de Vaus, D. A. (1985) *Surveys in Social Research.* London: Allen & Unwin