Chapter 9: Hypothesis Testing

Progress Questions

1. When dealing with any statistical association between variables, caution must be exercised before attempting to make statements about causality. Association is not a guarantee of causality. By the same argument, the fact you cannot find a statistical association is itself no guarantee that there is no causal association. The first two chapters of the book set out to establish the problems in this area. As you analyse your data and observe statistical associations (e.g. the mean of one variable increases in proportion to the mean of another) you will want to create hypotheses that attempt to explain the observations. In this chapter, three types of apparent associations were described.

   If the association is ***spurious*** there is no causal relationship. The fact that starlings migrate and oak leaves start turning brown at about the same time does not mean that one causes the other. The more likely explanation is that the two events are coincidental and have the same causal source – i.e. the changing season.

   If the association is the result of an ***intervening variable*** there is an indirect causal relationship. For example, the author undertook a small-scale study several years ago that looked at the employment of older newly qualified teachers (NQT's). There was a perception that older NQT's did have greater difficulty in finding long term employment in teaching. What the data actually showed for one cohort was something a little more complicated than that. Older women NQT's did indeed have a statistically significantly lower employment rate, but older men did not. The key appeared to be whether or not a teacher was geographically mobile and could apply for work across a wide geographical area. So, while there was a link between age and employment, there was an ***intervening variable***, geographical mobility, that had a very important effect. Older women were more likely to be with partners and family who were firmly established and rooted into the local area, whereas older men were more likely to be able to have the choice to move away because they were the main financial provider for the family.

   To make things even more complex, there were other likely factors that were behaving as ***interacting variables***. Clearly, gender had an effect in more than one way. Older women tended to train for the primary age group while older men were often training for the secondary age group, often in shortage subject areas like maths and technology. Also, the geographical location of the institution was in an area where there was relative stability in the teaching labour market with relatively low staff turnover. Therefore, all these factors – and probably others, too - were interacting to produce a very complex set of potential causal relationships.

2. While you may observe statistical differences between two or more variables in your research, probability theory tells us that this could happen within a sample because of sampling error or simply because the act of randomness allows this situation to arise. The question we ask of the data is this: is the observed result so far removed from the result we would expect if there were no relationship (i.e. the *null hypothesis*) between two variables that the probability of getting it is very low. If the probability of gaining this result if there were no relationship between the two variables is very low, we might take the gamble and say that there is a relationship. In other words, we report that there is a ***statistically significant*** set of observations and make certain decisions based on that. By convention, there are three levels of sigificance:

   - $p<0.05$ is the probability of obtaining observations showing a relationship is less than 5% if there actually is no relationship.
   - $p<0.01$ is when this probability falls to below 1%.
   - $p<0.001$ is when the probability falls to below 0.1% (i.e. extremely low).

Ó Ian Hosker & Studymates – 2002

The level of significance selected is dependent on the level of risk the researcher is willing to take of making a mistake. In the chapter two types of error are identified

- *Type I* errors occur if the *null hypothesis* is rejected when it is correct.
- *Type II* errors occur if the *null hypothesis* is accepted when it is actually incorrect.

Making the wrong choice may have little impact – so 0.05 can be selected – or it can have devastating effects by making social policy decisions on the basis of a Type I or Type II error – in which case the lower thresholds would be selected.

3. ***Parametric*** tests assume *normal distribution* of the data used in the test and that it is *interval* in nature. This makes them very strong tests because they are based on common patterns. ***Non-parametric*** tests, on the other hand, do not make these assumptions and so are weaker because there is less control over 'allowable' parameters.

4. For the sake of the answer to this question, assume the level of statistical significance has been set at $p<0.05$. Normally, when you test a research hypothesis, this involves testing a prediction that 'A' is related to 'B' in some definite way. For example, you may predict that the duration of unemployment of a particular group of skilled workers is related to their age, i.e. older workers spend longer on the unemployment register. This research hypothesis has a direction in that you are predicting increasing duration of unemployment. A ***one-tailed test*** of significance is used and if an increase is observed and its probability of occurring if the *null hypothesis* is less than 5%, the *research hypothesis* will be tentatively accepted.

   Of course, it may occur that the research hypothesis states that a relationship occurs, but cannot identify the direction of that prediction. This doesn't happen too often, as it is an indication of a weak hypothesis anyway, but if when it does, a ***two-tailed test*** is applied. Here, you have to share the 5% probability threshold between both possible directions (i.e. a possible predicted increase and a possible predicted decrease). In this case, the *research hypothesis* can only be accepted if its probability of occurring in a *null hypothesis* situation is less the 2.5%.

5. A ***t-test*** is *a* parametric test that compares the *means* of a variable in two samples to see if they are statistically significantly different. For example, you may have a research hypothesis stating that the mean age of male doctors in general practice is significantly lower than that of female doctors. You would use the ***t-test*** to test that hypothesis. There is, however, a slight difficulty with this. The value of the mean will be affected by a number of other factors, including the age distribution and there are two versions of ***t-test***. One version is used where the *variance* of the data distribution is similar in the two samples. The other version assumes the two samples to be of unequal *variance*. We therefore link a ***t-test*** with an ***F-test***. The latter compares the *variance* in the distribution of the variable values in the two samples and determines if they are statistically significant. This will then allow you to apply the correct version of the ***t-test***. Of course, an ***F-test*** is a test of significance in its own right when a research hypothesis is concerned with overall distribution. For example, you may want to test the hypothesis that while mean ages of male and female GPs are not statistically significant, the distribution of ages is greater amongst one gender than the other. Here you would use the ***F-test*** to see if the variance in ages between the two genders is statistically significant. With this test, the mean ages are not relevant, only the spread. By using both tests, though, you are able to say more about this variable in the two groups tested.

6. ***Chi-square*** is a non-parametric test. It is usually applied to frequency tables and cross-tabulations. It works by comparing the observed frequency of distribution amongst the categories with that you would expect to see if the *null-hypothesis* were correct. As identified in the chapter, and in the answer to question 2, probability theory tells us that any combination of frequencies is possible, but is it likely? ***Chi-square*** answers this question by calculating the probability of obtaining the observed frequency when the *null hypothesis* is actually correct and by applying the appropriate level of significance (either as a one- or two-tailed test) you can make a decision about what appears to be occurring within your sample.

Ó Ian Hosker & Studymates – 2002