Hypothesis Testing and the Westminster System

Alan McLean
Monash University, Department of Econometrics and Business Statistics
900 Dandenong Road, Melbourne, Australia
Alan.McLean@buseco.monash.edu.au

1. Introduction

In a paper presented at ICOTS-5 (McLean, 1998) I argued that the underlying theme in statistics is to use sample data to establish a probability model for one or more random variables; this model is used to predict the results of future measurements of the variable(s), and on the basis of these predictions decisions are made. The probability model 'describes the population' if it is sufficiently close to 'reality' that it can be used to provide useful predictions.

As an approach to teaching statistics, this 'predictive voice' emphasises the need for a model to work in practice. It also emphasises that the task of the statistician is to make predictions in situations of uncertainty, using models of the real world.

In a paper under preparation, the role of hypothesis testing under this paradigm is discussed. A test of hypothesis is a choice between two possible models, choosing that which is expected to give the better predictions. It is not an act of deciding which of two possibilities is 'true'.

Hypothesis testing differs from other forms of 'model selection' in several ways. First, only two alternative models are considered, one of which is privileged. This 'null' model is assumed initially to be the better model, and this assumption is to be rejected only if the sample evidence against it is strong. Second, the consequences of each choice, and of making the wrong choice, are unequal. Finally, the null model is fully specified but the alternative typically is not, so that only the null model can be tested on the sample. If it performs sufficiently badly, the alternative is accepted as superior. In a test on a mean, for example, this is done by using the null model to predict the mean of a sample. The criterion used to assess how well the null model works is the conditional probability of getting the sample result, or 'worse'.

Choosing which of the two models is the null is described in a variety of ways, usually saying that if the question is decided in favour of the null, no action will be taken, no change will be made, or the simpler model will be used. These all amount to taking account of the consequences of choosing the 'wrong' model. It is also pointed out that the null provides a fully specified model.

2. Hypothesis testing, court cases, and the scientific method

In teaching, it is common to point out that carrying out a test of hypothesis is functionally equivalent to a criminal trial in a court of law under the Westminster system. The evidence assembled by the prosecution must be sufficiently strong to make it reasonable to reject the truth of the privileged model. In countries following the British tradition, this 'null hypothesis' is that the charged person is innocent. Traditional hypothesis testing is also analogous to the scientific method, in which a new development is expected to prove itself against the current body of theory. If the new theory fails to be accepted, the old theory remains current.

In both these cases one is in some sense seeking objective truth, so the idea of a privileged null hypothesis is natural, and can be seen as morally correct. In a statistical analysis this approach is more arbitrary.

3. Advantages of the model approach

The recognition that in testing hypotheses one is choosing between workable models rather than identifying objective truth has a number of advantages. First, it reduces the psychological pressure imposed by the thought of making a mistake. It also makes absurd the distinction between

the phrases 'accept H_0' and 'not reject H_0': a model is by definition a tentatively accepted picture of reality which provides a current basis for action.

Finally, it answers the inevitable question: if $\mu \neq \mu_0$, what is it equal to? If the null predicts the sample mean to be 750, and the observed value is 748, the natural way to phrase the question is: is the mean approximately 750, as it should be, or approximately 748, as it appears to be?

4. An alternative approach

An alternative process which can be used to decide between the models is the following, illustrated with a test on a mean. Using the sample data to test all possible models (that is, all values of μ, including μ_0) on the criteria of zero mean error and minimal mean squared error, the observed sample mean performs best. Suppose $\mu_0 = 750$, and the sample mean is 748. On the basis of the sample evidence, it is likely that using a model with $\mu = 748$ will work best. The real choice then is between the null model ($\mu = 750$) and this sample based model ($\mu = 748$).

The question now is whether it will matter which of the two models is used. This introduces the concept of 'significant difference', with which many students have difficulties. It can be expressed as follows: two models are significantly different if it is highly probable that one model will work (substantially) better than the other. To measure this probability, calculate, for each of the selected models, the probability of getting the other result, or worse. Note that both probabilities are one sided, and will be equal. If this *p* value is sufficiently small, the sample model is significantly better, so is accepted.

In this balanced approach, the significance level is the probability of making a wrong choice between two fully specified models, not that of wrongly rejecting the null.

It can be argued that the sample based model is automatically favoured by using the sample evidence, and that the choice should be biased in favour of the null model. But this bias need not enter as part of the statistical decision. It only becomes relevant when the sample based model is not significantly better than the null model. Then it does not matter which of the two models is used, so other criteria should be used to decide. In the court case, if the two models cannot be separated, the choice is based on a notion of fairness. In a production line example, it is based on cost and convenience. In regression analysis, it may be based on model simplicity.

Conclusion

Hypothesis testing, as generally taught, is very much influenced by notions of fairness, conservatism, and the cultural concerns of the Westminster system, presumably because the techniques were developed by statisticians who had grown up with these. If hypothesis testing is concerned with identifying objective truth, it is reasonable to assume the truth of the null and insist that the alternative 'prove itself'. Recognising that hypothesis testing is a form of model selection, it becomes clear that this approach is merely optional.

REFERENCES

McLean, A.L. (1998), The forecasting voice: A unified approach to teaching Statistics, In Proceedings of the Fifth International Conference on Teaching of Statistics (eds L. Pereira-Mendoza, et al.), 1193-1199. Nanjing University, Singapore

RESUME

On enseigne souvent le test d'hypothèse comme analogue à une affaire en jugement criminelle. Dans la dernière, objective vérité est recherché; l'hypothèse testant cependant est une technique pour choisir entre deux modèles. Une fois vu dans cette lumière, il apparaît clairement que l'approche traditionnelle, avec le modèle nul favorisé, n'est pas la seule possible. Unweighted la variante sur la méthode habituelle est présenté.