

STATISTICAL MODELLING – SOME NOTES AND REFLECTIONS (Most of which will be ludicrously familiar)

* Remember the words of Brian Clough!

The Paper Trail

Ensure that all serious work can be reproduced i.e. have a clear ‘paper trail’ in place.

The platinum standard is that if a research assistant/fellow was killed in a freak accident the professor could complete the project.

The gold standard is that all files and notes are correctly and clearly set out so that they can be passed on to someone without much explanation. This will mean that you and the other members of the research team can follow the paper trail and therefore subsequently reproduce and augment material if required. This is particularly important as referees can often ask for minor, and in the case of some of my work major, amendments to statistical analysis.

Working with syntax will tend to help you in these aims.

Making A Start

IT IS ESSENTIAL TO KNOW YOUR DATA.

This includes understanding how concepts have been operationalised (e.g. via the survey instrument). It is worth thinking about how the survey instrument has been applied. Think about all the tiny nuts and bolts, for example the rubric of questions and how the routing has been worked out. These minor issues may have a major impact on your data.

Understanding how variables have been measured and coded is **OBVIOUSLY** essential. It is also worth getting to know the distribution of variables and some simple measures of central tendency (e.g. means and modes).

Make sure that you are working with the best data available. In the case of the BHPS this will be the most recent release of the data.

ALWAYS MAKE BACK-UP FILES. Work with as clean a set of data as possible.

Always start with exploratory analysis.

EVERY recode, compute, re-labelling task should be documented and be traceable in the paper trail.

DON'T START MODELLING TOO SOON!

Statistical Modelling

Always proceed from a position informed by substantive theory. The economists are particularly good at this (although occasionally a little rigid). The modelling building process should (ideally) always be guided at all stages by your substantive theory(s).

REMEMBER – REAL DATA IS MUCH MORE MESSY, BADLY BEHAVED, HARD TO INTERPRET ETC. THAN THE DATA USED IN BOOKS AND AT WORKSHOPS.

In terms of your model consider its formulation deeply. Think about how you can operationalise a suitable empirical inquiry using the variables that you have available. It is essential to think about the form of your outcome variable(s) and the appropriate type of model. It is equally important to think about your explanatory variables.

Always start with simple analysis. I prefer to develop models having first undertaken a great deal of univariate and bivariate analysis.

USE A DEFENSIBLE MODEL FITTING PROCEDURE.

Never use STEPWISE regression because it is a “fool’s errand”. The “cooking pot” approach is not too far behind. In non-linear models the model fitting procedure can substantially affect the results of the model. Think clearly about what you are trying to model. Most people are aware that correlated explanatory variables can cause problems but fewer analysts are aware that non-correlated variables can also be unstable in non-linear models.

In the case of longitudinal analysis spend as much time as possible getting the underlying social process clear before you fit a model. The best way to do this is to build upon well thought out cross-sectional analysis.

Always “guesstimate” the output before you formally estimate (i.e. fit) your model. This will help trap errors or indicate when your data is “behaving badly”.

Always have a notebook handy (or use notepad or your word processor) to help with the paper trail.

Keep a calculator handy.

If a job is incomplete keep a record. For example I frequently e-mail myself at the end of the day so that I am reminded the next time I log on.

Show others your results and try to explain results to others especially when analyses are complicated (here ethnomethodologists or small children can be useful). Use internal forums and select relevant conferences in order to get rapid critical feedback.

Get help from people with expertise. Newsgroups and e-mail lists can be very helpful.

Collaborate with people with relevant expertise.

Relax, statistical modelling can be fun.