Exon thesis on Rudolf Schwarz

I was asked to consider the statistic in the PhD thesis by Charlotte Exon. I have seen pages 126-132 of chapter 3 of the thesis; and Appendix 11.1 'Musical Emigré' study data, 11.1.1, 'Notes on the data' 11.1.5, 'UK Musical Emigré's: Summary of data used in Statistical Study', Appendix 1.2. 'The Statistical philosophy behind the Schwarz study'. Detailed comments follow.

The statement on pg 126 'In truth, Schwarz defied (sic) all emigratory trends' is not supported by evidence. The statement on pg 128 'Moreover, statistical modelling confirms Schwarz's antithesis to the overall trend' is false.

The first error made is to assume that one can understand emigration by ignoring both non-UK immigrants and those who did not emigrate. The 'Notes on the data' state that 240 emigrant musicians had been located. I have not seen an explanation of the choice to use data on only the 110 UK emigrants. There is insufficient information on the bounds of search, and inadequate details. For example, I would expect a summary of which musicians were multiply sourced. The author acknowledges that the study is not exhaustive.

Even using the data chosen by Exon, the statistics do not support these claims about Schwarz. The first error in the statistical modelling is that the prediction of year of emigration is based only on year of birth, status in Germany, confirmed Jewish links and specialty. It is very silly to use this linear regression (which I deduce has been used) to assess whether a departure date of 1945 is different from other departure dates. The fact that war was declared in September 1939 is ignored: at this point Germans in UK were classified as 'enemy aliens'. The statistician Sir Claus Moser was one of many German Jews who were interned in the UK in 1940. Other historical events (e.g. in 1938) are also ignored. This is a failure to allow for systematic differences.

Perhaps Exon used year of birth rather than age at emigration, which is equivalent, because it is embarrassingly obvious that a three year old's 'decision' (and her data includes three year olds as emigrants) to emigrate cannot be modelled as exactly as a fifty year old's decision.

Exon claims the model 'calculated that Schwarz should have left in mid-1936'. However, it is obvious from Figure 3.1 (pg 129) that very few people emigrated in 1936. Exon states that the standardised residual is 2.8: from this one can calculate that the fitted standard error is about 3 years. Therefore, the model which Exon assumes implies that 2/3rds of emigrants left between three years before and after 1936 i.e. between 1933 and 1939. This adds nothing to Figure 3.1 which shows that about 90% of emigrants left between 1933 and 1939. From Figure 3.1, it is also obvious that there was a relatively large emigration in 1933 and 1934 (at least 30% of the study subject) and a further substantial emigration in 1938-39 (24%).

Figures 3.1 and 3.3 clearly demonstrate the falsity of the claim that 'Schwartz represented a unique case'. In both figures, it is obvious that the five people who left in 1943-46 left at a different time from the majority who left before the war. Even if Exon's model were acceptable, there would be no grounds for distinguishing between the four people (points) with residuals in the range 7.7 to 8.8. Figures 3.1 and 3.2 provoked a couple of my colleagues to remark 'is this a witch-hunt?'.

Some other errors deserve comment. From the information given, it is very likely that no account was taken of family members moving together. The point for Alexander Goehr, aged 5 when he left is not connected with that of Walter Goehr, aged 34 when he left. I do not know whether all repeated surnames indicated relatives, but many must do so.

Even when a linear regression is valid, it is essential to distinguish between the population mean value, and the values which an individual might take. It is ridiculous to select nine people who happen be close to the fitted line and then claim, as Exon does on pg 130, that this shows the model has good predictive properties. This is like saying that a student who gets nine out of 110 questions right is a good student.

The 'Statistical Philosophy' includes some reasonable descriptions and illustration, but also serious omissions and errors.

The first sentence has a serious omission: the author does not discuss the essential role of statistics in choosing what data to collect and by what method. This leads to the erroneous claim that statistics do not investigate causality: randomised controlled experiment focus precisely on cause.

The author shows little understanding of descriptive statistics and their role. They can be very effective in drawing conclusions about the data. For example, the range of a continuous variable can prove that it is not what the investigation claims it is. A simple illustration is marks for an assignment. A colleague noticed the marks were wrong because the highest mark was 68%, too low for a cohort of 140 capable students.

The author is completely wrong to claim 'descriptive statistics are essentially useless when trying to analyse data'. On the contrary, exploratory data analysis is essential before using statistical models (and derived tests), to see whether the assumptions behind the statistical modules are sensible. Exon's failure to consider the validity of assuming that decisions of three year and fifty year people are equivalent is a case in point. Inferences should be data driven. Theory should change to fit data, not vice versa.

Inferential statistics do not remove the problem of post hoc explanation unless the data collection is sound, and the statistical models are correct. The author's lack of understanding of the role of random selection and random allocation, and of simple reality, is demonstrated by her example of dividing people by the letter of their last name. Exon thinks it would be absurd to conclude that differences between the first and second halves of the alphabet can be ascribed to their last names. I presume Exon imagines that she has created two random subjects of the data, but she is wrong. If we divided my department's students this way, the ethnic mix of the two halves would be very different, as the majority of our Chinese students have surnames beginning with w, x, y, or z. The higher average test results in foundations of mathematics of the second half of the alphabet can be understood by paying attention to the one difference between the groups which the investigator chose. In fact, Exon changes from last name to given name part way through her example, but given names also contain cultural and ethnic information.

The critical omission from Exon's description of inferential statistics is the assumptions which underpin the inferences. Exon does acknowledge bias, as in the case of the emigrant study, and that this can ruin the analysis. She also acknowledges that 'the failure to allow for systematic differences can lead to misleading and invalid results'. She is wrong to claim that 'statistics, based on the principles of randomness, *automatically*' (sic) allows for such randomness. Randomness takes many forms, and allowing for one form means not allowing for a different form. If Exon's model treated all people as independent individuals, it does not distinguish the difference between randomness among the members of a family from the randomness between unrelated individuals.

The significance level (p-value) does not denote the probability that the results have occurred merely due to chance variation. It gives the probability of an event at least as extreme as the event observed occurring by chance under a hypothesis with a very precisely specified distribution of measurements.

Exon also fails to distinguish between observations and parameters, does not mention the statistical models which are used, and does not specify which tests are used.

Further technical errors can be noticed in the thesis. The model appears to assume that errors follow a Gaussian distribution, which would at best be an approximation as the outcome variable, year of departure, is essentially discrete. The coefficients of the variables, and associated standard errors, should have been given. Checks of the distribution of the residuals should have been made and reported.

The information I have received provides strong evidence that Schwarz has been unreasonably singled out. The statistics presented bring to my mind the joke about a drunk's use of a lamp-post.

Professor J L Hutton
Department of Statistics
The University of Warwick
Coventry
CV4 7AL

6 February 2007