Facts from Figures

M.J.Moroney

May 3-4, 2019

An old Pelican Book I find in my library. I do not know how it came there. Could my father have bought it, or did I pick it up from some sale in a used bookstore (but there is no hand marked price, but a printed of 5/- i.e. five shillings). The author Moroney is invisible on the internet, except through this book (an excerpt of which was included in the anthology 'The World of Mathematics'). The book is billed a a layman's introduction to statistics, and if so the author takes his task very seriously indeed. No prior mathematical knowledge, except the most elementary is assumed, calculations, graphs and tables are explained, which makes the reading somewhat tedious for the mathematician. Yet, in this extravaganza of number crunching, complicated examples are pursued at length, sometimes for a dozen pages. No one innocent of mathematics would have the stamina to engage with numbers to such an extent as the author takes for granted. And after all, to the general public, mathematics is all about number-crunching.

This is statistics written from the perspective of a true statistician. Mathematics is but a tool, although interesting mathematics is bound to be glimpsed, but dealing with numbers and tabulating them is a delight. The book was written in 1950 so the calculations were done by hand, although there are occasional references to machines, but the machines in mind were certainly not computers but mechanical ones still in use until the early 70's. There is also a reliance on statistical tables and graphs, which by now are long since obsolete, so much of this now easily being part of software. More pointedly, statistics, if anything is geared towards applications, and now we are not primarily speaking of those to natural science, but to commercial manufacture. How to sample products to keep an eye of their quality through taking of samples. Although those examples are just manifestations of some general principle, I find them boring, however, clever they may be. Testing charts and control levels with due regard, they do not engage me, and I realize, as if I have not done it before, that being a industrial statistician would not have been my metier. On the other hand you can get used to many things, especially if there are no alternative options. Anyway, the lack of a mathematical perspective turns the book into a cook-book. Worked-out examples may be excellent to get the point of a general theory, but when the book consists of almost nothing else, it becomes a bit tedious and you long for some mathematical overviews, after all the theory is not so complicated after all and involves few new concepts to master.

So what did I pick up from those four hundred odd pages that I did not know before? That even the general binomial distribution, not only the symmetric case of $p = q$ converges to the Gaussian bell distribution, but the skewer the greater is the need for the size of the population. Thus in fact the Poisson distribution is an intermediate one, but exactly how it fits in, is of course not explained. The Poisson distribution I recall from almost fifty years ago, but I have in the interim forgotten. That the variances vary inversely to the sample size (thus standard deviations inversely to square roots) I probably have known,

but now I really understand the context and how to use it. Now the whole rigmarole with variances and co-variances could have been succinctly explained by scalar-products but this may be just for the mathematician, and his explanation of least squares is messy and confusing to layman and expert alike when we are not told that we are minimizing distances. There is a few interesting combinatorial problems having to do with rankings and comparing different rankings. The most interesting is connected with ranking n objects by giving preferences for each pair, and those may not be consistent. Thus you get directed edges on a complete graph and you are asked how many triangular cycles will there be (corresponding to inconsistent preferences of pairs[1]). How many cyclic triangles in the graph? This can apparently be worked out from the incidence matrix of say ones and zeroes defined by the ordered pairs.

When the author is not busy working out examples he gives you his opinions, which can be quite amusing. He has a thing against indexes, such as the cost-of-living index. Who needs it? And when it comes to time series and predictions of the future, he has but scorn, and considers it the modern variant of astrology, with the difference that astrology was at least excusable at the time. A statistician cannot get something out of nothing, and by putting figures on something does not mean much in general. A complicated mathematical computation which results in a high accuracy, could easily be worthless, in spite of the awe-inspiring accuracy. Statistics is a sharp tool, though, and if you do not know how to handle it, you will cut yourself. Not everyone can be a statistician, but people should at least learn so much as to be somewhat knowledgeable and leave the details to the professionals. If you are young, lose no time in studying some, if you are old, make sure that your younger upcoming colleagues learn. When all is said and done, what comes out of number-crunching has to be interpreted by informed common sense. He regrets that industrial statistics is relatively neglected in Britain as opposed to in the States, and that is detrimental to Britain's industrial growth, which clearly is of concern to him.

May 4, 2019 **Ulf Persson:** *Prof.em, Chalmers U.of Tech., Göteborg Sweden* ulfp@chalmers.se

[1] If you prefer A to B and B to C will you necessarily prefer A to C ? Not necessarily, and the author explains it by judging objects in many different ways, and only letting a few enter into a comparison. A is young, stupid and beautiful, B is young, stupid and rich, while C is old, smart and beautiful. You prefer beauty over wealth, everything else being equal, thus A is better than B, and stupidity over smartness more than beauty over wealth and age over youth combined, thus B over C, and finally age over youth more than stupidity overs smartness, hence C above A