AN INQUIRY INTO THE VALIDITY OF MATHEMATICAL METHODS USED IN EVALUATING THEORIES OF OCCUPATIONAL CHOICE

BY

G J BORIS ALLAN

MAIN THESIS

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE UNIVERSITY OF KEELE - SEPTEMBER 1975

ABSTRACT

An examination of the ontological basis of sociology reveals that sociology is inherently statistical. Various critiques of the use of mathematics and statistics are shown to be based on incorrect views of science and a confusion of different concepts of various ontological bases. As an introduction to some of the later discussion, a case study in inference is made.

The nature of the sampling distribution of a statistic is made clear in detail, arguments against the traditional (frequentist) approach to statistical inference are shown to be compelling. We start therefore from the idea of belief in the value of a parameter being the most important thing, with the cumulation of findings being essential. This is discussed from the point of view of Bayesian, Maximum Likelihood/Support, and Fiducial inference. A way in which frequentist studies may be cumulated is given.

Finally the importance of the ontological status of variables in an equation is shown with respect to two particular types of mathematical manipulation of variates. A analogue is drawn with the importance of dimensions in scientific formulae and it is shown that certain eqüations are not even possible, never mind correct.

The Appendices contain material (some previously published) which amplify the approach contained in the main thesis, and are of varying statusses.

CONTENTS

PREFACE

'... You walk into the room with a pencil in your hand and you see somebody naked and you ask "Who is that man ?"... You don't know what you will say when you get home ... Something is happening here, and you don't know what it is. Do you, Mr Jones ?...'
(Robert Zimmerman. Ballad of a Thin Man.)

'... [The scientist] appears as a realist insofar as he seeks to describe a world independent of the acts of perception; as idealist insofar as he looks upon the concepts and theories as free inventions of the human spirit...; as positivist insofar as he considers his concepts and theories justified only to the extent to which they furnish a logical representation of relations among sensory experiences. He may even appear a Platonist or Pythagorean insofar as he considers the viewpoint of logical simplicity as an indispensable and effective tool of his research...'
(Albert Einstein. Reply to Criticism.)

The following is the result of much interplay between theory and research, (chronicled in Appendices C,E and F) but is mainly concerned with the philosophical aspects of statistics and mathematics and how they impinge on the way we treat and conceptualize our 'data'. I have not delved into the complicated field of causal modelling or into factor analysis partly because I have been on record on the subjects previously, and I did not want to overlap, but mainly because these are the epiphenomena and I search after the core. This explains why after an introduction to my all pervasive philosophy, I concentrate so much on statistical inference - we may not use it properly or frequently , so why ? - and then I take an example of how the seductiveness of a formula can lead us to forget what are the real attributes of that with which we are dealing.

The Appendices are not essential to the understanding of the argument in this thesis, but help to amplify various aspects and fill-out its content: as there is more to life and science than the logic or form of an argument and we need the content to keep a grip on reality - so this thesis is lessened without the broadening content of the Appendices, which show in action the ideas which have been purveyed in the thesis. My approach is aptly described by the opening line to 1.3 '... Not agreeing with Hernandez-Cela that we cannot learn anything about the general through the examination of the particular I shall proceed to show how it may be possible to make some general points about theory-testing from an examination of a particular study ...'

My thanks are due to Roy Mapes, a friend rather than 'supervisor', and also Bill Bytheway , a 'supervisor' but more a superb breed of Devil's Advocate. I have benefitted from discussions with many people at Keele Manchester and Nottingham. In particular I have come of age as a sociologist mainly, I feel, through the agency of Jim, Joel and John of the School of Sociology, Manchester Polytechnic.

I dedicate this thesis to my wife (who has borne the brunt of the noise of my insistent typing) and my children who, despite my faults, love me (or so they say).

CHAPTER 1

THEORIES AND TESTING

Arctic elephants are the same as African ones only they're colder. Feel one. (Spike Milligan. _Arctic Elephant_)

1.1 Sociology as Science

Sociology is a science or, if it is not a science, it should be. Sociologists should not, however, aspire to be as natural scientists because the nature of the reality studied in natural science differs from that reality studied in social science : if the natures of the realities differ why should the methods of the sciences be the same? (Popper disagrees, see Appendix A). The reason sociologists can be scientists is that they use theories, some theories being more powerful than others - the difference between theology and science lies in testing of the implications of theories (I will pursue this below). Various commentators see sociology as a (potentially) scientific concern where the key element is the use of 'concepts' rather than 'observables' in explanation, with concommitant use of theories (eg Allan,1974; Hindess,1973;Willer and Willer,1973). Unfortunately there is a tendency to over-emphasize the difference -

Hindess (p51) says that theoretical entities are not '... objects of immediate human experience and do not appear to be reducible to such objects ...'; and the Willers (p34) suggest that one way of distinguishing between a 'cow' and a 'force' is '... force is not an observable ...' (and where does that place a blind man ?). Grover Maxwell (1962) and Sellars (1963) both make a similar point with which I agree : If a conceptual entity is named in a theory then, if the theory is 'good' theory, it is reasonable to believe in the reality of the entity named. In the Willers' example of a cow and a force there is a further complication : a cow is entity, and force is a characteristic usually associated with some entity, so that they are not comparable (the Aristot an distinction between substance and 'accident' of substance). Below I will give an example of an 'abstract' observable.

Theories have to be tested before we can be sure about them - not an obvious point - Willer-and-Willer discuss testing of theories and experimental design, and Hindess has an Appendix which discusses an experiment in the measurement of time (Koyré's original example). Bachelard (eg 1951,1968) has a phrase 'revolution of concepts' - taken from Nietzsche - and suggests there are crucial experiments and discoveries which totally change the course of people's (ie scientists') thinking - an insubstantial notion which has been surprizingly influential (eg Althusser,Hindess). It is insubstantial because it starts from false premisses, and the creation of a false edifice is shown most clearly in one of his earliest examples - the impact of the Michelson-Morley experiment on the development by Einstein of his Special Theory of Relativity. Bachelard (1951) in Einstein's Festschrift (Schilpp, 1951) dwells at length upon the upheaval (ie revolution) of concepts, and uses the incongruous, unexpected result of the Michelson-Morley experiment on aether drift to account for the development of Einstein's theory. (On p566 we read '... As we know, as has been repeated a thousand times, relativity was born of an epistemological shock; it was born of the "failure" of the Michelson experiment ...' .) Einstein himself in his 'Autobiographical Introduction' to the the collection nowhere mentions the Michelson-Morley experiment in his description of the development of Special Relativity - he does however say (1951a:53) '... By and by <u>I despaired of the possibility of discovering the true laws</u> [relating matter, space and energy] <u>by means of constructive efforts based on known facts</u>. The longer and the more despairingly I tried, the more I came to the conviction that only the discovery of a universal formal principle could lead us to assured results....After ten years of reflection <u>such a principle resulted from a paradox upon which I had already hit at the age of sixteen</u> : [what happens to somebody moving with the speed of light, and what does that person see ?]...in this paradox the germ of the special relativity theory is already contained ...' (my emphasis). (See also M Polanyi's(1962) discussion of the same point .) As with the Copernican

'Revolution' pre-existing ideas are justified by discoveries, and discoveries do not create ideas. Aristarchus (3rd century BC) had proposed that the planets follow circles around the sun, but on the basis of observation it was shown by Hipparchus (2nd century BC) not to be feasible. Instead of improving upon Aristarchus' idea - ellipses instead of circles - the geo-centric scheme replaced it (largely because it was less impious).

In a similar manner the Willers (1973:137) make the the contrast between Science and Empiricism partly to lie in that '...Science is inherently imaginative and radical ...'. In criticizing the practices of empiricists they should be aware of the practices of scientists - unless they have some rarified picture of what Science is like. The Willers are strong proponents of the view I oppose at the start of this section; mine, restated, is that what is good and true for one science is not necessarily true for another science. Two examples will suffice :

they say (pp 21-22) '... Scientific knowledge [they always use examples from the physical sciences] is more precise than empiricist knowledge [they use examples from social science] but not because the means for measurement in science are more precise, <u>for any measurement used in science could also be used in empiricism</u> ...'(My emphasis), it is not too clear what they mean; but, in their section on measurement and scaling, we read (pp113-114) '... One does not measure length by asking a sample of people to rank basketball players, Supreme Court justices [etc]... according to the [people's in the sample] beliefs about which has the greatest length and then summing the result to get a "scale" of the way the average person in the sample thought they should be ranked. No amount of refining could make this procedure even fractionally useful as a yardstick ...Applying this method to determine a "scale" to "measure status" is equally preposterous ...'. I am not too certain what the Willers mean by 'yardstick', but : if they mean a stick a yard long, it is not only preposterous, it is impossible, to use a yardstick to measure the distance to the moon; if they mean by yardstick an absolute definition of units and properties at the macro-level, they will become unhinged at the micro-level.

Bachelard, the Willers and Hindess all seem to be Rationalists in a Platonic Realist sense :eg there are 'true' concepts of 'space-time',or 'length', or 'time' ('universal' Forms of Plato?), these exist only in their mathematical definition; by relating these definitions to various indicators we can perform experiments (or tests of theories) and can obtain perfectly clear-cut results. Bachelard (1951:577) says '... Let us firmly underscore that efficacious thought proceeeds in the direction of rationalism → realism ...', the Willers imply that lemgth only exists in as much as it is a mathematical element in an equation, and Hindess(1973:60) says '... I have insisted that the basis of knowledge is not to be found in human 'experience' but in concepts and rationalist forms of proof and demonstration. In particular, that knowledge is never simply given but is always the product of a determinate practice ... [see Bachelard] for a systematic refutation of the claims of realism in the natural sciences...'. There are many forms of Realism and the arguments of particularly Bachelard and (therefore) Hindess are those of Platonic Realists, or as they are frequently termed Platonic <u>Idealists</u> : Einstein was a self-avowed realist, and he wrote (1951b: 678) approvingly '... there are concepts...which play a dominating role in our thinking, and which, nevertheless, can not be deduced by means of a logical process from the empirically given (a fact which several empiricists recognize, it is true, but seem always again to forget)...'. I would contrast to the above Idealistic Rationalism, a Realism in the Aristotelian sense (see also Smart(1963) on 'Scientific Realism') '... Such a realism may be summed up in two dominant considerations : (1) the consideration that the things of the world simply are what they are in themselves and independently of our attitudes toward them or our opinions about them [though this does not discount the self-fulfilling prophecy]; and (2) the consideration that human beings are capable - subject, of course to all sorts of errors and mistakes that they may commit in the process - of coming to know such things of the world more or less adequately ...'(Veatch,1974:75 - my emphasis). There is another aspect to Aristotle's philosophy that can clarify some confusions in the plethora of 'scientific' critiques of sociology; the distinction made above between 'substance' (eg an electron) and 'accident' of substance (eg the length of an electron) and further the importance of finding why such an accident pertains to such a substance. (Does it make sense to talk of the length of an electron ?). For example, Hindess(1973:51) starts his appendix on observational categories by noting on the first page '... Scientific discourse refers to objects(electrons,electromagnetic fields,imperialism, the capitalist mode of production) that are not objects of immediate human experience and do not appear to be reducible to such objects ...' (so the capitalist mode of production does not really exist?) : and on the final page (p58) he concludes

'... In these examples the measurement of time is a theoretical operation ...'. This is clearly a move from substance (eg, that electron) to an other item 'time' which is not an accident of a substance (what is the time of an electron ?) but something central to all life and science - Aristotle himself (we can agree or disagree) considered that time is inseparable from change; and in the four-dimensional space-time continuum we are only aware of the continuum when motion occurs. The point is ontological - that is, what are the natures of the items under discussion, what is their essential status ? - rather than epistemological - what is the nature of our theory of knowledge ? (See also the discussion in Appendix A).

This again brings us back full-circle to the beginning of this section - scientific methods in Sociology must differ from scientific methods in the natural sciences, because the nature of the realities studied differ (an ontological argument). The Willers and Hindess are arguing epistemologically; we must not only have a theory of knowledge but also a theory of what it is we are attempting to study.

Finally, on the subject of 'crucial' experiments and 'crucial' results, Zetterberg is probably much closer to the truth for sociology when he comments (1966:161-162) upon how <u>hard</u> it is to confirm (or contradict) a single proposition in sociology (Marxian theory is a prime example - see Sorokin, 1927:514-546). He suggests that often we can merely provide some form of 'odds' - the chances of it being true to it not being true. The reasons, which are not really connected to poor measurement, are examined in the next section.

1.2 A Sociological Ontology

In the preceding section I introduced the idea that: what scientists from different sciences did (the tools they used, and the things in which they were interested; that is, their methodologies) would differ. I also showed how the Willers and Hindess had not grasped this important ontological point, because they were principally epistemologists. In this section I will first sketch-out the form of a viable sociological ontology, and then consider how it stands up to Chapter 6 of Willer-and-Willer (1973). (I approach this problem from what might be called 'Scientific Realism' and the philosophical questions involved in this stance are discussed in Appendix B).

1.2.1 Statistics and Parameters

Sociologists are - almost by definition - concerned with groups of individuals, why and how the groups differ. Sociologists continually use <u>Statistics</u> - a statistic (or set of statistics) summarizes a whole series of different values of a group of individuals in one value (or set of values). When an ethnomethodologist (E) says that quantitatively-inclined sociologists are 'Positivists', he or she is providing a statistic. Individuals within the group might <u>never</u> have exactly the value of the statistic, and some statistics (such as those which measure degrees of variation within a group) have no counterpart at the level of individuals. Statistics are totally abstract, and are not observables as such; but, being formed from observables, I suppose we can call them 'abstract' observables. To show, therefore, that any individual does not have the same value as the statistic is not damning, because the statistic pertains to the group. For example, to point to an individual quantitative sociologist who is a Scientific Realist and not a Positivist does not negate E's point; but to show that <u>most</u> quantitative sociologists are Scientific Realists casts doubt on the statistic - which in this case is some form of average. (In fact, when youthful lay sociologists chant "All coppers are narners", they are expressing two statistics, which they feel are sufficient to describe the distribution of intelligence of police officers."The average intelligence of police officers is low, and there is no variation around this average".)

It is possible to query E's statistic in at least two ways. The first query concerns the way in which the value of the statistic was found (ie , given E's data, would we have arrived at the average, 'Positivist' - or, is the 'average' income given by the median or the mean ?); and the second query is concerned with the representativeness of the group examined by E (ie, is it typical of quantitative sociologista in general ?). Suppose that E cannot be faulted on either count, but that, however, another sociologist M (ie, me) finds that the quantitative sociologists he studied were closer to Scientific Realists - M's study also cannot be faulted. Which of the two statistics is correct ? And - this is essential - what is our criterion of 'correctness' ? For something to be incorrect, or to talk of degrees of 'correctness' we must have a notion of what it is to be perfectly correct, (Bachelard: If we have a Philosophy of Yes we must have a Philosophy of No). If all quantitative sociologists were studied we might find that the statistic had the value 'Pragmatist' - a mythical creature akin to the Griffon and half-way between Logical Positivist and Neo-Aristotelian Realist. The name given to a statistic calculated from all possible individuals (ie the 'population') is a Parameter. If the parameter is Pragmatist, and E and M loooked at similar numbers, then probably the statistics Scientific Realist and Positivist are each as likely as the other to be found in subgroups (ie 'samples'). Given the notion that statistics from samples need not agree with the corresponding parameters of populations, we can then employ the abstract, (conceptual ?) notion of a sampling distribution of a statistic. In taking many different samples from a population we will find many different values of statistics, some statistics occuring more frequently than others. The graph of the (theoretical) frequency of occurence of any **value** for a **statistic** given a set value of the parameter is called the sampling distribution of a statistic, (this is usually a purely mathematical enterprize).

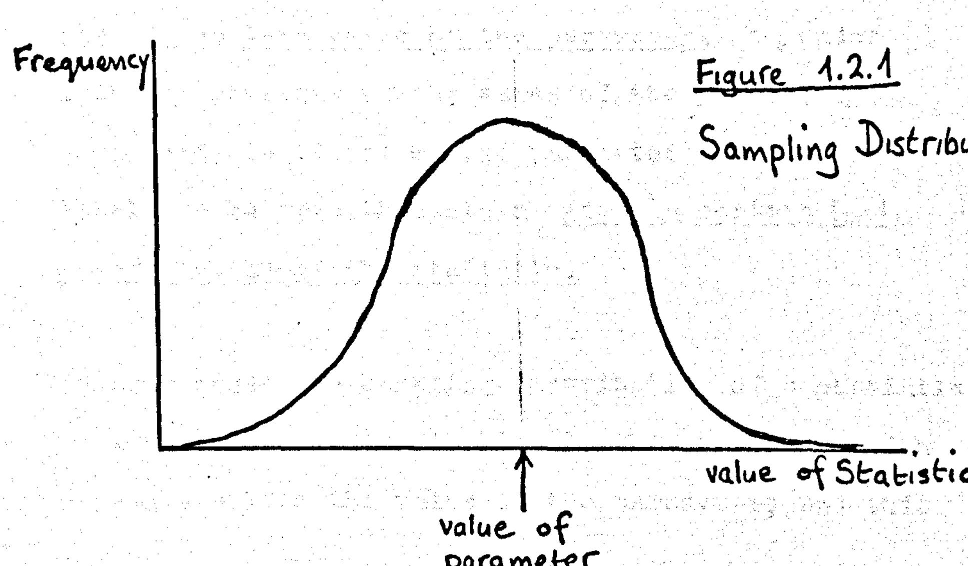


Figure 1.2.1

Sampling Distribution

This theoretical notion has been attacked quite unreasonably, because we can never conceive of a very long run of different samples to have any relevance to the individual sample. As it so frequently is, a valuable notion is lost because it can be used in more than one way - and the way it is commonly used is highly dubious (to wit: 'The Significance Test Controversy'). Some important implications of the sampling distribution of a statistic are discussed below in connection with Bayesian and Fiducial Inference; but for the present suffice it to say that, as some values of statistics are more likely to occur than others given a certain value of the parameter, so, under lack of knowledge of the value of the parameter, some possible values of the parameter are more likely to be true than others given a certain fixed, observed value of the statistic.

In many cases the sampling distribution of a statistic follows a normal distribution (eg Figure 1.2.1) with a mean equal to the value of the parameter; and this distribution, whose mathematical formulation arose from the 'theory of (measurement) errors', has led many to confuse sampling error and measurement error. The nature of a statement to the effect that we are 99.9% certain that the value of parameter is within certain limits, is entirely different from the nature of a statement that the speed of light lies within certain limits. In the first case our limits are set by the size of sample and the recognition that the values of <u>statistics</u> will differ from the value of the parameter; in the second case our limits are set by our technology, eg our wavemeter will only measure to certain level of accuracy, and the existence of a 'population' of possible readings is a non-starter. In a similar sort of way the juxta-positioning in argument of Statistical Mechanics (say) and statistical methods in sociology is a non-starter (though see Nagel(1961:290-293,503-510) for an example of this). In statistical mechanics we observe what is analogous to a population of (say) gas molecules, each individual molecule has its own energy and we wish to predict the value of a quasiparamter - the average energy of the gas per unit surface. The value of the quasi-parameter is constantly fluctuating (by minute amounts) and it is possible to graph variations in value of quasi-parameter against frequency of occurence to obtain a graph somewhat similar to Figure 1.2.1. The similarity is nice, but it is misleading because in Figure 1.2.1 the variation was in value of the statistic, whereas in this case the variation is in the value of the parameter (which is why I termed it a <u>quasi</u>-parameter - it is a parameter but only for a short instant). It is always possible, I suppose, to regard the quasi-parameters as <u>really</u> statistics because the actual statistical <u>methods</u> are the same, but in sampling we are in a different ball-game (and what is the population in the case of quasi-parameters ?).

The question is, again, ontological : A 'true' parameter of a distribution (though it may possibly vary in value over time) is <u>essentially</u> different from the speed of light or the ascension of Venus.

The first refers to a characteristic of a population and the distribution of some varying quantity across this population, and (assuming perfect measurement) any possible variability in estimating the value of this parameter is due to the fact that we cannot completely enumerate this population - a complete enumeration would lead to the same kind of variability in measurement as in the case of the speed of light.

This small point leads us elliptically to my main point - natural science is by nature different from social science - and also stresses the importance of distributions, and therefore statistics and probability theory, in sociology. In fact sociology <u>is</u> statistical (in the widest sense of the term statistical).

1.2.2 The Ontological Status of Probability

In Willer-and-Willer (1973) there is a chapter on "Probability in empiricism and science" written by Hernadez-Cela (henceforth H-C). This is a poor chapter in a trite book (ie H-C,1973); and we have in our possesion a sufficiently capacious armoury of concepts to examine it in a comprehehsive manner. Because 'Probability' is the key element, as the section-head says : What is probability ?

One, probability can be the parameter of a distribution, ie the <u>population</u> equivalent of the statistic we call a proportion. Whittle (1970:28) says '...The <u>probability of A</u> ...is to be regarded as the expected proportion of experiments [ie cases] in which A actually occurs. The motivation for the definition comes from the finite [NB!] population census...'. What Whittle means by 'expected' proportion is the mean of the sampling distribution of the proportion. Now, corresponding to a sample proportion, we have a population probability, where (as I note above) the population is possibly finite : This is in complete contradiction to H-C (1973:97) when he writes '...a relative frequency [ie a proportion] is a probability only if the number of events taken into account is infinite...'(my emphasis - H-C is proposing an alternative definition, as if it were the only definition).

Two, probability can be a probability 'density' : that is, for any distribution we can find the proportion of cases between values X and X + ΔX (say) divided by the difference between the two values, ie ΔX. If $Pr(x : X \leq x \leq X + \Delta X)$ represents the proportion of cases of value x, where x lies between X and X + ΔX and we write

$$g(x) = \frac{Pr(x : X \leq x \leq X + \Delta X)}{\Delta X}$$

letting ΔX tend to zero gives a limiting value of this function $g(x)$. If $F(x)$ is the proportion of cases in which the value of the variable is less than X, the limiting value of $g(x)$ is given by $f(X)$, where

$$f(X) = \frac{dF(X)}{dX} \qquad \text{........(1.2.1)}$$

This limiting value ($f(X)$) is the probability density of X - mathematically it is given by the slope of the graph of the cumulative proportion (ie the ogive, $F(X)$) against X, at any particular point. $f(X)$ is not the probability (or proportion!) of X - assume for the moment that X is continuous-valued - of X occuring, for the probability of observing any particular value of X beforehand is near-zero, though after it has been observed the probability is unity. The probability desity is what is shown by the height of the curve in Figure 1.2.1 - that is the axis labelled 'Frequency'.

Three, probability can be a 'likelihood'. Let us be interested in the sampling distribution of the statistic S, given a fixed value of the parameter, ie P. The probability density of a certain value of S, given this fixed value of P will be written f(S:P). Suppose, however, that, instead of being interested in how likely a value of S is to occur, we use the information about the value of S we have calculated to say something about possible values of the parameter (given the value of the statistic and information about the sampling distribution). Obviously some values of the parameter are more likely to be correct than others, and if the probability of S given a certain value of P is f(X:P), we can turn this around to say that the 'Likelihood' of the value of P given a calculated value of S (ie L(P:S)) is related to the probability density by :

$$L(P:S) = f(S:P) \qquad \text{........(1.2.2)}$$

'Likelihood' (as is 'probability density') is a mathematical abstraction, and is on a different level to probability defined as the value of a parameter. (Maximum Likelihood methods of choosing a parameter involve choosing the value of P for which the likelihood L(P:S) is greatest. See De Finnetti (1972:73-74) for specific comments on the types of probability from a Bayesian viewpoint).

Obviously 'probability', 'probability density' and 'likelihood' are ontologically distinct concepts; and the notion of probabilities before and after the event being distinct is, we will see, of crucial importance. The H-C critique confuses all these and in my discussion I will leave aside such outlandish scepticism as (H-C,1973:97) '...The proportion, however, is not <u>relevant</u> to [those]... not in the sample...'(my emphasis). Surely it must be relevant in some way to others, otherwise we have to contemplate chaos ? Wittgenstein's attitude towards eternal sceptics was : Go ahead and <u>practice</u> your scepticism. In this case : Why worry, what you say about some quantitative sociolgists is not relevant to any other quantitative sociologists ?

The critique commences with a justifiable condemnation of the notion (H-C,1973:96)'...that scientific knowledge is gained by inference from the direct observation of individual facts...' but immediately he falls into a well-populated trap : '...the most that can be said about the number of heads that will turn up when tossing a coin twenty times is that there will be a particular frequency which is unknown until we toss the coin. In other words, the assignment of a value ½ simply because the coin has two sides is an error because we do not know that each side will be equally represented in any empirical case. <u>Equal representation in probability is a mathematical assumption which is violated in finite empiridal</u> cases...'(H-C,1973:98 - my emphasis).

Two mistakes are made : firstly, we cannot know how many heads will arise in twenty tosses <u>but</u> we can say which is the most probable number of heads that will appear - there is <u>no</u> mathematical assumption which is violated in empirical cases, (unless we confuse statistic with parameter); the second mistake is further exemplified in '...We are told that the probability of rain tomorrow is 60% when, in fact, it will either rain or it will not. Such statements are [as usual ?] unjustified, wrong and misleading...' (H-C,1973:98). I too am not all that confident about weather forecasts, but H-C has confused two ontologically distinct items - the former is a measure of belief in the occurrence of a future event, and the latter is an observation made of the future event after the event has happened. H-C would ask us to condemn statements such as : 'The weather tomorrow is likely to be very warm' as '...unjustified,wrong and misleading...'. (Here is a consequence of the Platonic Idealist epistemology - an Aristotelian would never have contemplated this flying in the face of common-sense reason). There is also the point that what we should do or believe before the event is different from what we should do or believe after the event. The point that is reiterated in the critique (with unbelievable deduced consequences), the point is that an observed proportion is not a probability, (ie a statistic is not a parameter) and an observation is not a probability - true but trite.

I will finish these notes on the critique by examining (H-C,1973:101) '...Fisher in his "fiducial" argument claimed that probability can be constructed out of available data. Thus Fisher rejected the notion that there is any difference between a proportion and a probability...'. Clearly H-C has misunderstood Fisher's argument (he is not alone in this !), for in essence - given a value of the statistic S - Fisher asks us to examine the variation in value of the likelihood L(P:S) with the possible values of the parameter P. Fisher then asks us to treat values of L(P:S) as if they were probability densities - this, I suppose, is what H-C means by '...probablities [L(P:S)] can be constructed out of available data [S]...'. Note that the fiducial argument requires not only the value of the statistic but also the form of the sampling distribution of the statistic (to give f(S:P) upon which we base L(P:S)).

Below (in Chapter 2) I elucidate the implications of the distinction between the sampling distribution of a statistic and the fiducial distribution of possible values of a parameter, after consideration of a case study in inference in the next section.

1.3 A Case-Study in Inference

Not agreeing with Hernandez-Cela that we cannot learn anything about the general through the examination of the particular, I shall proceed to show how it may be possible to make some general points about theory-testing from an examination of a particular study. The study is that of Ford and Box(1967), and, leaving for later delectation their theoretical development, I will consider their analysis of results in support of their theory. Actually all that I will examine is contained in part of their Table 1, and (Ford and Box,1967:295) :
'...When all three variables [ie Type of Scientist, Type of Favourable Emplyment Conditions, and Expected Degree Result] are considered together, however, our ability to predict occupational choice [ie Future Employment Preference (University/Industry)] is much improved. Thus four out of five public scientists perceiving university as providing better

professional freedom and expecting good degrees [a 'Type A' scientist] chose academic employment, while less than one in ten instrumental scientists perceiving industry as providing higher salaries than universities and expecting degrees of lower second class standard or less ['Type B' scientists] chose such employment...'

Table 1.3.1

Occupational choice of two types of university science students.

	UNIVERSITY	INDUSTRY
TYPE A	15	4
TYPE B	1	12

This is Ford-and-Box's interpretation of the results shown in Table 1.3.1, and is the culmination of a long discussion of the separate effects of the three variables on occupational choice. In essence, they show that when university science students are split up into various groups (samples, thought to be representative in some way of a theoretical population where the population is <u>finite</u> and ideal-typical), then in these groups the proportions choosing university differ. To establish that the proportions differ they apply chi-squared test of significance - they do not do this for the difference between the two groups which most clearly tests their theory. I will not use a conventional test of difference between proportions; the reason for this is partly explained in these terms - am I interested in whether the two samples come from the same population ? (To quote from a well-known book - written (so we are told) by '...a Fellow of the Royal Statistical Society...' - (Reichmah,1970:327) '...It is sometimes desired to test whether two samples are in fact drawn from the same population or whether there is a significant difference between the samples and therefore also between the populations...For this purpose we may calculate the <u>standard error of the difference</u>...') Venn pointed out that any person, indeed any sample, is part of a vast number of populations; in the case of the Type A and Type B student scientists, both samples are from the population of student scientists! If we twist the test into a test whether the two samples come from different populations but with the same mean, as is sometimes done, we have lost the sampling theory legitimation. This will be explored in more detail later.

Consider this then : of all the students interviewed .3 chose university as their future employment preference; suppose, therefore, that the population probability for Type A scientists was also .3 - given a parameter of this value what is the probability density of obtaining 15 choices of university out a total of 19 ? Refering back to Figure 1.2.1, alternatively, what is the Frequency of a statistic of value .7895 (=15/19) if the value of the parameter is .3 ? Actually Figure 1.2.1 is not too accurate a representation in this case because the sampling distribution is not symmetrical, neither is it continuous (it is discrete). The probability density in this case is given from the theory of binomial sampling, that is, the probability density is (writing factorial n as n!) :

$$\frac{19!}{15!\ 4!}(.3)^{15}(1 - .3)^{4}$$

(this is considered in greater mathematical detail in Appendix D), and has the value 1.34×10^{-5} - a very low figure. We have to consider whether this very low value has any meaning : for example, what is the probability density of this statistic corresponding to a parameter of value .79 ? The probability density in this case is :

$$\frac{19!}{15!\ 4!}(.79)^{15}(1 - .79)^{4}$$

and has the value 2.20×10^{-1} - still a low value, but much greater than the previous density.

What has been done ? If, as in the previous section, we 'invert' the probability density of a statistic to arrive at the likelihood of a parameter (p 1.2.13), then the likelihood of a parameter being .79 is 2.20×10^{-1} and the likelihood of the parameter being .3 is 1.34×10^{-5}. The relative likelihood of the parameter being .79 to it being .3 is thus 1.64×10^{4} to 1 ; alternatively, the likelihood that the sample is drawn from a population with a high (in this case an estimate is .79) probability of choosing university is vastly greater than the likelihood of the sample being drawn from a population with a low (.3) probability of choosing university. Expressing this in percentage terms, we could say we are 99.998% confident: that '... there are about [99998] chances out of [100000] that they have hit upon something really true...'(Zetterberg,1966:161). We may have hit upon the outlandish case but our non-statistical knowledge suggests that this is not the case : probably this is clearly a sample from a high probability population, where this population is more of a theoretical construction than anything else - a theoretical construction, that is, which attempts to portray a real thing, and, indeed, may be real. I suggest that this analysis forms a clear confirmation of Ford-and-Box's prediction, (whether the prediction really follows from the theory is a different - epistemological - question).

What is also shown in Table 1.3.1, is the converse of Type A, the Type B scientist. The observed sample proportion choosing university is is .0769 (= 1/13) : the probability density corresponding to a probability of .3 is :

$$\frac{13!}{1!\ 12!}(.3)^1(1 - .3)^{12}$$

and, likewise, for a population probability of .08 :

$$\frac{13!}{1!\ 12!}(.08)^1(1 - .08)^{12}$$

The relative likelihood of the parameter being .08 to it being .3 is thus :

$$\left(\frac{.08}{.3}\right)^1\left(\frac{1 - .08}{1 - .3}\right)^{12}$$

which is equal to 4.96. This relative likelihood is not very high, which is probably due to two things : the small number of Type B science students interviewed, so that we cannot be very sure of the stability of the statistic as a measure of the parameter; and the small difference between the 'average' probability of .3 and the 'Type B' probability (as hypothesized) of .08. (Below it will be shown how this relative likelihood of 4.96 corresponds to a (one-tailed) p-value of .04, standardized to a normal distribution. The difference would be established rather more 'firmly' in the conventional tests).

What we have are really two separate tests of the theory; as both tests support the predictions - to differing extents - does this mean that we have a greater amount of confidence in their theory than if we had just one test? In any sane world this would be so, but rarely is this done in sociology. Usually tests are made against a null hypothesis, and if p is less than .1 the null hypothesis is accepted, but the results of this test are not used to inform later tests of the same null hypothesis. If we have only small samples it is very difficult not to accept the null hypothesis in most cases, because generally: the smaller the sample the greater the variation in feasible values of the appropriate parameter. However, if in all the different samples the statistic was positive, (and the null hypothesis is that the parameter is less than or equal to zero) surely the null hypothesis is incorrect ? In the conventional tests '...after a time the null hypothesis joins that corpus of hypotheses referred to as 'knowledge', on no positive grounds whatever...'(Edwards,1972:179). Using this RL (Relative Likelihood) method, we simply combine the 'odds' : in this case the odds that the predictions taken together are true, rather than that the results are merely sampling fluctuations from a common average value, these odds are 8.16×10^4 to 1 ($8.16 \times 10^4 = 1.64 \times 10^4 \times 4.96$). The combination of odds derived from conventional tests is incorrect, as will be explained in the next chapter. The specification of .3 as the 'average' probability and choice of .79 and .08 are, of course, open to discussion and perhaps this might lead to further clarity in the predictions.

It might be objected that a test is a test is a test, in the sense that a sample is a sample is a sample. In this example, if Type A were more likely (note how common usage is so free with these technical terms - or is it vice versa), to repeat, if Type A were more likely than Type B to choose university employment in this study, that is that - to Eurydice with the statistical tests? One obvious objection to this is to show that I could form a sub-sample of 4 Type A students who chose industry, and a sub-sample of 1 Type B student who chose university - perhaps Hernandez-Cela's arguments about samples can be seen in this light. If a prediction were couched in terms such as '<u>all</u> X are Z', then a test is a test.

CHAPTER 2

STATISTICAL INFERENCE

'... when we are powerless to discern the truest opinions, we must follow the most probable, and although we see no more probability in some than in others, we must nevertheless settle on some ...'
(René Descartes. Discourse on Method (3)).

2.1 Groups, Populations and Samples

There is a group of individuals - a <u>set</u> called G. The individuals in G are also members of a wider group - a set called U. Techically we say the group G is included in U, in set notation

$$G \subset U \qquad \text{........(2.1.1)}$$

The set G is composed of many elements, where each element is an e_G and we show this by

$$G \equiv \{ e_G \} \qquad \text{........(2.2.2)}$$

How do we know whether a certain individual is in G or not in G (ie $\bar{G}$)? If we have no information at all about an individual, other than that the individual <u>is</u> an individual, we are not able to say whether the individual is in G or $\bar{G}$, (though by definition the individual will always be in U). For any individual e^* (say) we can always write

$$e^* \in U \qquad (2.1.3)$$

(e^* is a member of U), but the question is whether

$e^* \in G$ or $e^* \in \bar{G}$?

In less abstract terms, we know that Bloggs (e^*) was a member of the human race (U), but we do not know whether he was one of the group (G) of people we call psychopaths, or the much larger group ($\bar{G}$) of people who are not psychopaths. Given the existing (nil) information we cannot say; but Bloggs is more likely to be a member of $\bar{G}$ (non-psychopaths) than G (psychopaths), because there are more people in $\bar{G}$ than there are in G. Suppose that Bloggs was a pseudonym for A Hitler : does this change our assessment ? It does, because we are not in possession of <u>no</u> knowledge and most people will think that, given his reported behaviour, he is more likely to be a member of G. (We could be wrong, even so).

This is the same old problem, touched on in Chapter 1 :

if it has rained it is immaterial to the rain that day that somebody said it was to be sunny all day; if we have a psychopath it is immaterial to this observation that there are more non-psychopaths than there are psychopaths (though we might ask why there are only a few psychopaths).

Let, then, each individual be distinguished by a set of characteristics $c_1, c_2, \ldots c_m$, shown by $e(c)$ (ie the individual e has the set of characteristics c). Further let there be two sets of characteristics for every individual in G, ie c_U and c_G ; and two sets for an individual in $\bar{G}$, ie c_U and $c_{\bar{G}}$. More concretely, c_U is the set of characteristics that every individual has indiscriminately, the elements of which are not related to a person being a psychopath or not, (eg hair colour). c_G is the set of characteristics, the elements of which are associated with a person being a psychopath, and $c_{\bar{G}}$ is the complement of these.

To show this symbolically, G is those individuals who have a set of characteristics c_U and c_G

$$G \equiv \left\{ e(c) : c = c_U , c_G \right\} \quad \text{........(2.1.4)}$$

That is, G is the set of individuals whose characteristics are c_U and c_G (and not c_U and $c_{\bar{G}}$). This leads to the question of the ambiguity of the elements of c_G - to what extent does this set clearly define membership of G ? This question re-appears in many guises.

2.1.1 Predicting Group Membership

One of the ways in which the term 'probability' is understood may be called the 'propensity' interpretation - the 'degree of rational belief' (Keynes). That is (given a long diatribe at a conference from a person who accuses the giver of a paper of not being aware of her own auspices) : "He's probably an ethno ...". Given a composite of information about a thing, (where the thing can be a person, group, tec.) and a range of exclusive groups of which the thing could be a member, then : of which of these groups is the thing most likely to be a member ? Of course if we know of which group the the thing <u>is</u> a member then we need not guess of which group the thing is a member. (The deliverer of the diatribe might not be an 'ethno', he might be a well-known comedian). This 'composite of information' is the set of characteristics (c) which we split into : those characteristics which convey no information about the relative propensities of the thing's membership of the various groups (ie c_U); and those characteristics which provide information on the relative propensities (eg c_G and $c_{\bar{G}}$). As I noted above (and this is reflected in the propensity interpretation of probability) to what extent does a set of characteristics clearly define group membership, is a moot point. If a set of characteristics do not distinguish clearly between membership of the groups, we shall begin to believe that these characteristics are not really relevant to whether a thing is in this group or that.

If we decide that the 'thing' in which we are interested is a social individual (ie a person), apart from the inherent cussedness of people, there is the seeming randomness which arises because most characteristics pertaining to social individuals are <u>variates</u>. A variate has differing values and some values are more likely to occur than others (frequently they are termed variables - but variate emphasizes the distributional aspect). Intelligence is an example of a variate, pertaining to people - it is a variate because it can take different values (some people are more clever than others), and the frequencies with which it takes these values follow a normal distribution over the population as a whole. Let G be the sub-group of people we term lecturers : it is well-known (?) that the distribution of intelligence for lecturers is such that there are are relatively more people of high intelligence, than there are in the rest of the population (ie sub-group $\bar{G}$). Thus, a person of high intelligence is more likely to be a lecturer - or is he/she ? If you think of the proportion of lecturers in the population, it is very small : there are sufficient highly intelligent non-lecturers for there to be more of them than highly intelligent lectureres. Prior knowledge of the odds 'number of lecturers' divided by 'number of non-lecturers' informs our guess about whether a person is likely to be a lecturer ; in conjunction with knowledge of the relative likelihood of a person of that intelligence being in the group of lecturers or in the group of non-lecturers. What has just been discussed forms the subject of Bayes' Theorem.

Let us suppose that somebody has found the intelligence of all the lecturers, and the distribution of

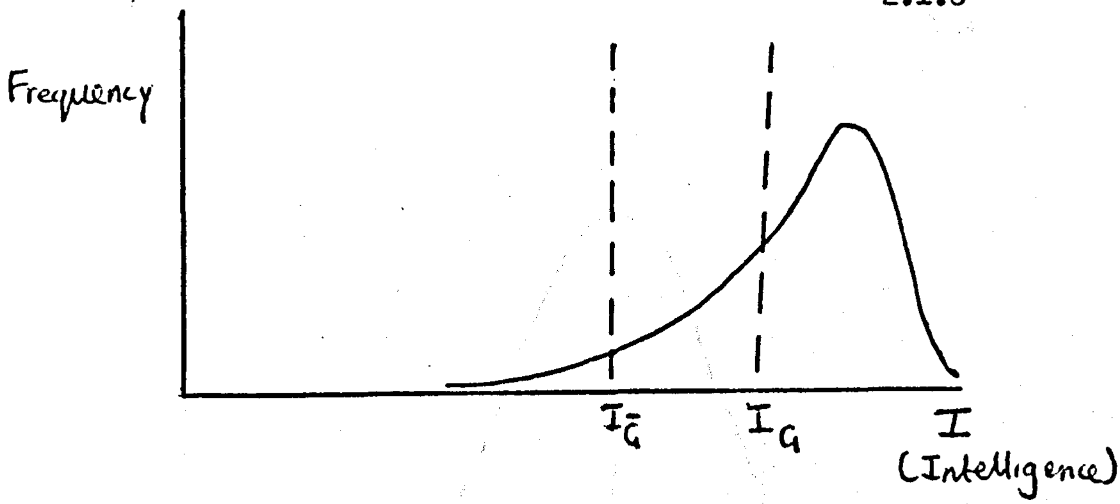


Figure 2.1.1 The distribution of intelligence of Lecturers

---

intelligence is as shown in Figure 2.1.1. On Figure 2.1.1 , $I_{\bar{G}}$ shows the average intelligence of the population-at-large and I_G shows the average intelligence of lecturers. Two comments : (1) The average intelligence of the group of non-lecturers will be almost identical to the average intelligence of the population-at-large; (2) I have <u>not</u> defined the nature of the 'average', it could (eg) be a median or a mean. The shape of the distribution of intelligence for the population-at-large is 'normal', ie it follows the same shape as the sampling distribution of the statistic as shown in Figure 1.2.1, and is shown in

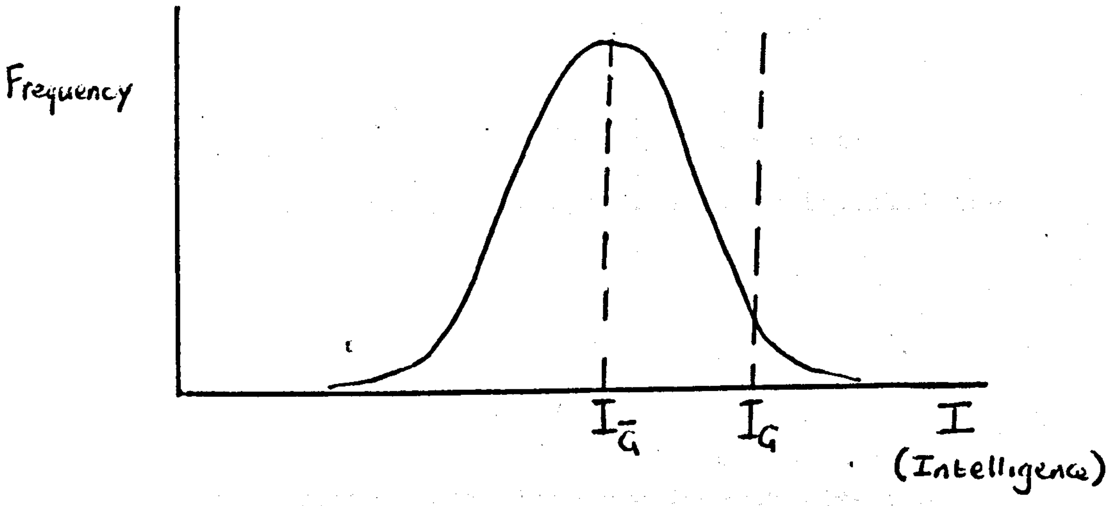


Figure 2.1.2 The distribution of intelligence of the General Population

---

Figure 2.1.2 . It is clear that, as portrayed, the two distributions of intelligence are of different forms.

'Smoothing'-out the curves of Figures 2.1.1 and 2.1.2, allows us to substitute the 'Frequency' of the vertical axis with the 'Probability Density'; the probability density of an individual coming from the group of lecturers with an intelligence of I will be shown by

$$f_G(I:I_G) \qquad \ldots\ldots(2.1.5a)$$

Similarly, for the group of non-lecturers the probability density corresponding to an intelligence I is

$$f_{\bar{G}}(I:I_{\bar{G}}) \qquad \ldots\ldots(2.1.5b)$$

(The subscripts to the 'f's are to emphasize that the shapes of the distribution of intelligence differ in the two groups). The <u>likelihoods</u> of a person coming from these two groups , when his intelligence is I, have the same <u>value</u> as the appropriate densities (without sharing the same ontological status), that is :

$$L(I_G:I) = f_G(I:I_G) \qquad \ldots\ldots\ldots(2.1.6a)$$

$$L(I_{\bar{G}}:I) = f_{\bar{G}}(I:I_{\bar{G}}) \qquad \ldots\ldots\ldots(2.1.6b)$$

The <u>relative likelihood</u> - R - of a person coming from these two groups, given that his intelligence is I, is (in the absence of any other information) :

$$R = \frac{L(I_G:I)}{L(I_{\bar{G}}:I)} \qquad \text{........(2.1.7)}$$

(This relative likelihood is often called the likelihood ratio - sometimes, less exactly, the odds - but I will continue to refer to the relative likelihood.)

The relative likelihood has been interpreted as the odds that either of two exhaustive propostions are true (the two propositions are : 'the person is a lecturer' and 'the person is not a lecturer'), but in calculating R in (2.1.7) we are only explicitly concerned with the relative likelihood of two parameters of two different distributions - given an observed value I.(The value I can be regarded as a statistic calculated by examining one person only). This should be remembered for later; because, though we might seem to be calculating the relative likelihood of propositions, we are gauging the the relative likelihood of parameters (in this case I_G and $I_{\bar{G}}$) given the observed value of a statistic.

Suppose that we did not know a person's intelligence, and also we did not know of which group he was a member. Intuitively, I am sure, most people would agree that, if you choose a person at random, the person chosen would be, by-and-large, a non-lecturer. How can we turn this 'common-sense' into something we can use ? (Forgetting quickly the sceptics who - though they solve them continually in their every-day existence -would not admit such problems ; because, they argue, it rains or it doesn't rain and he/she is a lecturer or he/she is not a lecturer). The approach I set out below is directed towards common-sense and its success or not depends upon whether it seems <u>reasonable</u> to the reader - a mathematical proof which does not make sense to the resource we call common-sense, even though it is an exact and correct proof, is automatically suspect. The common-sense evaluation should be applied to the <u>reasoning</u>, because obviously common-sense is not always ready to evaluate the result without the reasons for the result. Non-Euclidean Geometries give results with do not seem to match common-sense, but the reasoning behind them is eminently in line with common-sense. Some of Bertrand Russell's pyrotechnics concerning the non-existence of Napoleon can be best seen in this light - the reasoning does not even make sense, (see Veatch, 1974:14-20).

If a person has an intelligence I the probability densities are as (2.1.5), but, as there are more non-lecturers than lecturers, we should perhaps be interested in <u>numerical densities</u>. Numerical Densities (my term) are still mathematical abstractions but they are scaled to be in units of people, that is the numerical density is the probability density multiplied by the number of individuals in the groups (ie N_G and $N_{\bar{G}}$). If we did not know the value of I, we would have to consider <u>all</u> possible values of I : Summing all the individual numerical densities (usually by integration) gives an idea of how likely a person of <u>any</u> intelligence is to be in that group. If we show the <u>summed</u> numerical density for group G as $S(*:I_G)$, and for group $\bar{G}$ as $S(*:I_{\bar{G}})$ - where the * shows that <u>all</u> values have been taken into account - then it can be proven mathematically and it seems reasonable that :

$S(*:I_G) = N_G$ and $S(*:I_{\bar{G}}) = N_{\bar{G}}$

Let the total number $N_U = N_G + N_{\bar{G}}$ and define

$$s(*:I_G) = S(*:I_G) / N_U$$

$$s(*:I_{\bar{G}}) = S(*:I_{\bar{G}}) / N_U$$

These standardized summed numerical densities can be themselves inverted to provide thevalues of likelihoods of being a member of the two groups (in the absence of any information). These likelihoods are numerically equal to the proportions (pr(G) and pr($\bar{G}$)) of persons in the two groups, and are shown as $P(I_G:*)$ and $P(I_{\bar{G}}:*)$; that is,

$$P(I_G:*) = s(*:I_G) = pr(G) \qquad ......(2.1.8a)$$

$$P(I_{\bar{G}}:*) = s(*:I_{\bar{G}}) = pr(\bar{G}) \qquad ......(2.1.8b)$$

That there is no equivalent to the :* for the ordinary proportion and/use the group name (eg pr(G)) loses the emphasis, that is in the likelihood $P(I_G:*)$, on the parameter I_G in the absence of any information on the intelligence of individuals. The relative likelihood of a person being a lecturer to the person being a non-lecturer (ie R) is numerically equal to all of the following -

$$R = \frac{P(I_G:*)}{P(I_{\bar{G}}:*)} = \frac{pr(G)}{pr(\bar{G})} = \frac{N_G}{N_{\bar{G}}} \quad \ldots\ldots\ldots(2.1.9)$$

Combining the information in (2.1.7) with the prior odds (prior = before we know any thing else, such as a person's intelligence), this gives the relative likelihood of a person of intelligence I being a lecturer to being a non-lecturer (given the different likelihoods of people in general being lecturers) :

$$R = \frac{L(I_G:I)}{L(I_{\bar{G}}:I)} \times \frac{P(I_G:*)}{P(I_{\bar{G}}:*)} \quad \ldots\ldots\ldots(2.1.10)$$

Now we will suppose that we cannot find the values I_G and $I_{\bar{G}}$, but that we are able to find what <u>proportion</u>

of lecturers and non-lecturers fall into the intelligence <u>category</u> represented by I (I is no longer a single value). This can be shown as pr(I:G) and pr(I:$\bar{G}$), and the relative likelihood of a person in intelligence category.I being in group G of group $\bar{G}$ is shown as

$$R = \frac{pr(G:I)}{pr(\bar{G}:I)} = \frac{pr(I:G)}{pr(I:\bar{G})} \times \frac{pr(G)}{pr(\bar{G})} \qquad \ldots(2.1.11)$$

The term pr(G:I) is called the likelihood of the proposition that the person is a member of group G. Reading from left to right in (2.1.11), the relative likelihood is equal to the <u>posterior</u> <u>odds</u> (of the person being a member of groups G or $\bar{G}$), which is equal to the <u>likelihood</u> <u>ratio</u> multiplied by the <u>prior</u> <u>odds</u> (of the person being a member of groups G or $\bar{G}$). This is one formulation of Bayes' Theorem. (If extra, independent, information is obtained it is possible to use the posterior odds from one calculation as prior odds in another calculation).

As a footnote to this sub-section, suppose that the difference between I_G and $I_{\bar{G}}$ were small. The conclusion would not be that lecturers and non-lecturers came from the same population - they must in any case, because they are both sub-groups of U. Rather the conclusion would be that the two groups differed but slightly in average intelligence. A further, pragmatic, conclusion would be that intelligence (or lack of it) had little to do with a person being a lecturer.

2.1.2 Random Sampling and Experimental Design

Ian Hacking(1965:118) points out that '... Many persons take inference fromsample to population as the very type of all reasoning in statistics ...' and this sub-section takes many of its critical ideas from his chapter on 'Random Sampling' (Hacking,1965 :Ch VIII). I will first consider the relationship between population and sample in terms of my preceding notation. The group U I will now class as the population; that is, it is the totality of individuals about whom I wish to make statements and check predictions - the parameter is calculated on the population of individuals. The group G I will class as the sample; that is, it is the totality of individuals who have been studied - a statistic is calculated on the sample of individuals. The group $\bar{G}$ are those in the population who are not sampled - G is a perfectly representative sample of U if there are no characteristics c_G and $c_{\bar{G}}$ which differentiate G from $\bar{G}$. From a slightly different angle, G is a representative sample of the population U if we are unable to decide upon a Person's membership of G or $\bar{G}$ (in contrast to the way in which we were able to decide if somebody were a lecturer or a non-lecturer). In random sampling it is often implied that a random sample is representative - a random sample is not necessarily representative.

A (nearly) classical definition is '... By a random sample, we mean a sample which has been selected in such a manner that every possible sample has a calculable chance of selection ...' (Kendall and Stuart,1969: 206), and Kendall and Stuart then make two comments upon their definition. One, 'calculable' means 'able to be calculated in principle'; and, two, not every sample need have an <u>equal</u> chance of selection. I would make a further comment : not every randomly-chosen will be representative, and surely we need samples selected in such a manner that every possible representative sample has a calculable chance of selection ? The difficulty with this idea of representativeness lies in the simplest (and most frequently discussed) case of sampling, (and it lies in the most complex). The simplest case to which I refered is that of the bag of balls, some red some black, where the contents of the bag are the population. (I sometimes feel that it would be more in keeping with contemporary ethos for the bag to be a commune and the red and black balls, males and females - perhaps this would have a greater intuitive appeal ?) A certain number of balls are extracted from the bag - this is the ideal-typical random sample, is it also a representative sample ? Though the sample is made up of raund balls (just like the population), the only time it is truly representative is when the proportion of red balls in the sample is the same as the proportion of red balls in the population. Here, then, is the problem : we can only have a representative sample if the statistic has the same value as the parameter, but the reason we calculate the statistics' value is that we do not know the value of the parameter. For the sample of balls we have a characteristic (being a ball) about which we are sure of the sample's representativeness : we also have a characteristic (colour) about which we have no information concerning representativeness. The sampling distribution of a statistic can be seen as an exemplification of the inherent degrees of representativeness of one characteristic over a set of samples representative in <u>all</u> other respects. Always, therefore, the usual way of calculating the variation in statistics (from the sampling distribution) is <u>always on the conservative side</u>, because representativeness is assumed in all other respects.

As an example of this latter point, consider the standard error of the mean. In theory, the variance of the sampling distribution of the mean (a statistic), is equal to the true variance of the variate in the population divided by the size of the samples taken. This is an exact mathematical statistical proof, but what happens in empirical cases is that we do not know the true value of the population variance - it is a parameter which we have to estimate by the sample's variance. What then happens is that, instead of the true value of the population variance, we take the sample variance and divide <u>that</u> by the sample size. Any deductions we make, on the basis of this sampling variance, are about a population with a variance equal to that of the sample. The sample variance is itself a statistic, where the appropriate parameter is the population variance, and so is itself subject to sampling error. Obviously, then, our estimate of the variance of the sampling distribution of the mean will be a 'middle' figure and it quite possibly underestimates the true variance of the sampling distribution of the mean.

Representativeness, like most statistical concepts, is relative and relational - it should be represenitaive-in-aspects-relevant-to-the-estimation-of-the-statistics-under-consideration. Suppose, for example, that we have a clearly defined population 'All the fitters in Shop B', and our interest is in average income from outside sources. We take a sample of fitters and calculate their average income (which is the statistic); we then notice that all of our sample have the fore-name 'George', and nobody else in the shop is called 'George'. Does this mean that we have a sample which is unrepresentative in any aspect <u>relevant to a person's income</u> ? (If all their surnames were Cohen, it might). Probably it is not unrepresentative in this sense, and the question might be answered empirically (using information from other studies) and/or theoretically ('George' is not associated with any type or class in society, whereas 'Cohen' is). Sometimes there is no such thing as an unrepresentative sample; there is an example in Allan (1974) where I point out that, if class and intelligence have distinct and independent effects in the population, the effects should be distinct whatever sample is used.

An associated problem is that of experimental design. Edwards (1972:203-206) discusses two types of advantage supposedly resulting from randomization: the first is that advantage which results from a design which most probably exhibits very little regularity with respect to both forseen and unforseen factors; the second advantage resides <u>in</u> the actual fact of randomization, ie it has the appearance of 'objectivity'. The first advantage is possibly nullified if we have chosen a design at random, and, as frequently happens, it turns out to be regular in some undesirable way - for really we should throw it out, though some would say that it should be kept warts and all. By trying again, we have lost the claim to objectivity; as Edwards (1972:204) comments, '... if we were very percipient we would be able to choose a suitable design without [randomization] ...', and Hacking (1966:Ch VIII) shows that a truly random sequence (and therefore design) is very difficult to construct. Another facet of experimental design is that in an ordinary sample we do not do anything to the sample, apart from select and observe; in an experiment we do something to one or more of the individual samples - eg spray with fertilizer, or increase the temperature in a work-place. This distinction is reflected in the terminology of the analysis of variance: in an experimental set-up, we have ANOVA of the first kind; and in an ordinary sample we have ANOVA of the second kind. In an experiment the independent variables are <u>variables</u>, for we choose the values they take as part of the experiment; in an ordinary sample the independent variables are <u>variates</u> for we cannot choose the values and the values follow a distribution. (Fairly frequently we meet a combination of both types of independent variables). Estimates of the variances of sampling distributions of statistics are more accurate in the case of truly random experiments.

2.2 Frequentist Approaches to Inference

I use the term 'Frequentist' to describe approaches to statistical inference which only consider the sampling distribution of a statistic, and 'areas' under the curve. That is, instead of looking at the probability density for a certain value of the statistic we look at the probability integral corresponding to that certain value - the probability integral gives the (abstract) proportion of cases whose value is less or equal to this certain value of the statistic. The probability integral is an abstract mathematical analogue of the cumulative proportion, and the integral distribution is the analogue of the cumulative distribution (or ogive). My attitude towards this analogy is that the cumulative proportion in an 'abstract' observable (a combination of observable quantities to form something that is not self-evidently there - questions of 'observability' raise their head at this stage), this abstract observable is an attempt to get near - measure ? - the 'true' probability integral. The whole shape (or mathematical form) of the integral distribution can be regarded as a parameter (perhaps many-dimensional) and so the calculated cumulative distribution is, in itself, a statistic of the population distribution.

The frequentist approaches are many but can be reduced to two main classes - hypothesis testing and confidence intervals - but both classes are intimately related, and show their inherent relatedness most clearly in the case of the normal distribution of a statistic.

2.2.1 The Normal Sampling Distribution of a Statistic

The normal distribution of a statistic, s, given a parameter of value p and a sampling variance of d^2, is described by the probability density

$$f(s:p) = (2d\pi)^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}\left(\frac{s-p}{d}\right)^2\right] \qquad (2.2.1)$$

The size of the standard deviation of the sampling distibution of a statistiv (ie d) is the <u>standard error</u> (of estimate), and is usually unknown - it has to be estimated from the sample which has provided the estimate of the original statistic (ie s).

The likelihoods of there being a parameter of value p_1 and p_2, given a statistic of value s, are (where k is a constant) :

$$L(p_1:s) = k \exp\left[-\tfrac{1}{2}\left(\frac{s-p_1}{d}\right)^2\right] \qquad ....(2.2.2a)$$

$$L(p_2:s) = k \exp\left[-\tfrac{1}{2}\left(\frac{s-p_2}{d}\right)^2\right] \qquad ....(2.2.2b)$$

The likelihood ratio (relative likelihood) of the values p_1 and p_2, given the statistic of value s is (from equations (2.2.2) and remembering the form of division of powers) :

$$\frac{L(p_1:s)}{L(p_2:s)} = \exp\left[-\frac{1}{2d^2}\left((s-p_1)^2-(s-p_2)^2\right)\right] \qquad \text{......(2.2.3)}$$

The log-likelihood-ratio is defined as the natural logarithm of (2.2.3) and I will show it by G, so that

$$G = -\frac{1}{2d^2}\left((s-p_1)^2-(s-p_2)^2\right) \qquad \text{......(2.2.4)}$$

Note that if one of the two values of the parameter (as postulated) is equal to s (eg $s = p_1$), then

$$G = \frac{1}{2d^2}(s-p_2)^2 \qquad \text{.......(2.2.5)}$$

$$\frac{L(p_1:s)}{L(p_2:s)} = \exp\left[\frac{1}{2d^2}(s-p_2)^2\right] \qquad \text{.......(2.2.6)}$$

As d is the standard deviation of a normal distribution, and p_1 and p_2 are postulated means of this distribution, the differences $\frac{s - p_1}{d}$ and $\frac{s - p_2}{d}$ are standard normal deviates. (Values for the probabilty density and the probabilty integral - often called the distribution function - can be consulted from tables of the standard normal deviates.) If we write $z_1 = \frac{s - p_1}{d}$ and $z_2 = \frac{s - p_2}{d}$, equations (2.2.3) through (2.2.6) become

$$\frac{L(p_1:s)}{L(p_2:s)} = \exp\left[-\tfrac{1}{2}(z_1^2 - z_2^2)\right] \quad \text{........(2.2.7)}$$

$$G = -\tfrac{1}{2}(z_1^2 - z_2^2) \quad \text{........(2.2.8)}$$

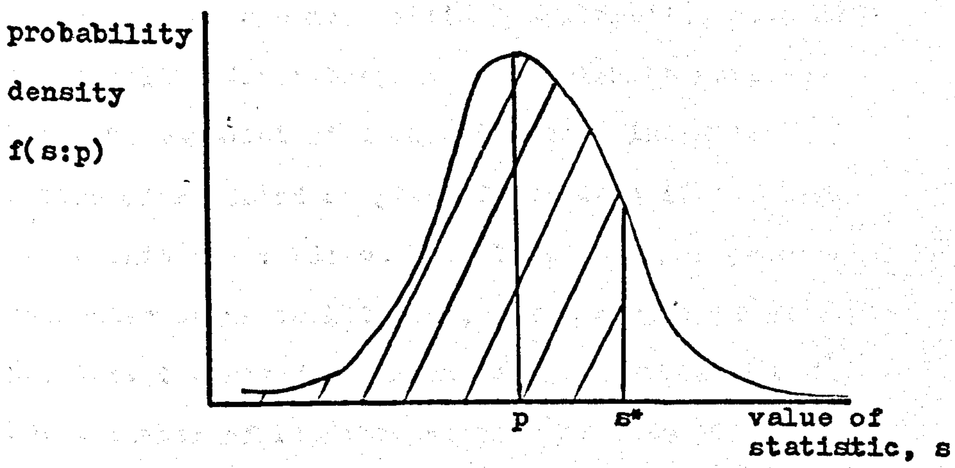
$$G = \tfrac{1}{2}z_2^2 \quad \text{........(2.2.9)}$$

$$\frac{L(p_1 = s : s)}{L(p_2:s)} = \exp\left[\tfrac{1}{2}z_2^2\right] \quad \text{........(2.2.10)}$$

The probability integral for a value s^*, given a parameter of value p, etc, is shown by $F(s^*:p)$ and

Figure 2.2.1

The Normal Sampling Distribution of a Statistic



The shaded area to the left of s* is represented by F(s*:p)

is equal to the (abstract) proportion of cases for which values of the statistic would be less than or equal to s* - this is illustrated on Figure 2.2.1. These results will be utilized in what follows.

2.2.2 Confidence Intervals

The idea behind confidence intervals is natural enough; we have a mathematical abstraction such as the sampling distribution, and it can be seen as being the endpoint of a long (perhaps infinite) series of repeated samples of the same size. Look at it this way : if we know that the true population parameter value is 100; and, for a sample of size 50, that the standard error is 10; then, for a large number of repeated samples of size 50, in .975 of the the repeated samples the value of the statistic would be less than 119.6. But - students often ask when presented with this information - what has this to do with the single (probably unrepeatable) sample ? Students are frequently perspicacious for this is the very point made by many committed statisticians; these statisticians see that the notion of repeated sampling may have a relevance in acceptance procedures, but they would argue that a different approach is needed for the unique sample. Fisher (1955:69-70) writes :

'... I am casting no contempt on acceptance procedures, and I am thankful, whenever I travel by air, that the high level of precision and reliability required can really be achieved by such means ... (p71) where acceptance procedures are appropriate, the source of supply has an objective reality, and the population of lots, or one or more, which could be successively chosen for examination is uniquely defined; whereas if we possess a unique sample ... on which significance tests are to be performed, there is always ... a multiplicity of populations to each of which we can legitimately regard our sample as belonging; so that the phrase "repeated sampling from the same population" does not enable us to determine which population is to be used to define the probability level, for no one of them has objective reality, all being products of the statisticians' imagination ...'.
It is of note that the popular introduction of Moroney(1951) in discussing confidence intervals and hypothesis testing takes examples exclusively from what might be termed acceptance procedures.

In a way there is a logical inversion of the sampling distribution when 'confidence intervals' are set up. We move from a situation where we know p and d and wish to find the value of s which satisfies $F(s:p) = a$ (the proportion of cases, a, for which the value of the statistic is less than s, given a parameter p); we move to the situation where we know s and d and wish to find the value of p, such that $F(s:p) = a$. Likewise we try to find p, so that $F(s:p) = b$. Letting p_a be the value of p in the first case and p_b be the value of p in the second case; we say we are $(a - b)$ <u>confident</u> that the true value of p lies in the interval $p_a \leq p \leq p_b$. This is legitimated by : knowledge of d; and the notion of a long-run of repeated samples. The first is suspect (as I mentioned in section 2.1) on the graunds that we rarely know d and have to estimate it by $\hat{d}$, say. We thus restrict our interest to a <u>special</u> population for which $d = \hat{d}$; and (though we might have an original population with, as Fisher says, 'objective reality') we do not display our confidence in the original population but rather in a population which is a 'product of the statistician's imagination', (ie for which $d = \hat{d}$). The legitimation by repeated samples obviously should have no viability if we cannot conceive of there ever being any repeated samples. It might be argued that the repeated samples are conceptual constructs, and thus have relevance : I would maintain they do have relevance but not in the way they have been used in establishing confidence intervals - the reason lies in the difference between likelihood and probability density.

I think a very clear objection lies in the point I made above, that '... There is a logical inversion of the sampling distribution when "confidence intervals" are set up ...'; for, whereas the likelihood is the inversion of the probability density, we do not calculate the likelihood integral as the inversion of the probability integral in confidence intervals - we use the probability integral as it is. The distribution, for which we use likelihood instead of probability density, was called the <u>fiducial distribution</u> by Fisher (fiducial = shows confidence or trust). The fiducial distribution is a 'distribution' of possible values of a <u>parameter</u>, given an observed value of a statistic. In simple cases (such as the normal distribution) the <u>fiducial</u> <u>limits</u> are the same as confidence intervals, at the same level of confidence (I might be tempted to argue that the reason that confidence intervals have been fairly 'successful' - if that be the word - is because of this equality with fiducial limits). I will discuss the fiducial approach in greater detail below, and I must disagree with Moroney's(1951:240-241) statement that '... Such discussions [of the bases of the fiducial and confidence approaches] are mainly matters for the professional statistician and would be out of place in the present introductory sketch of the subject ...' : the differences between the two approaches are basic philosophical and sociological problems. To understand the reasoning behind the bases might stop claims of the form '... We may therefore say that the population mean [p] lies in the range ... 46.65 [p_a] to 47.35 [p_b] , and express our confidence in this claim by saying that in a long series of estimates of this type, if we were to use exactly the same kind of argument, we would expect to be right 68% [a - b = .68] of the time ...' (Moroney,1951:241).

2.2.3 Hypothesis Testing

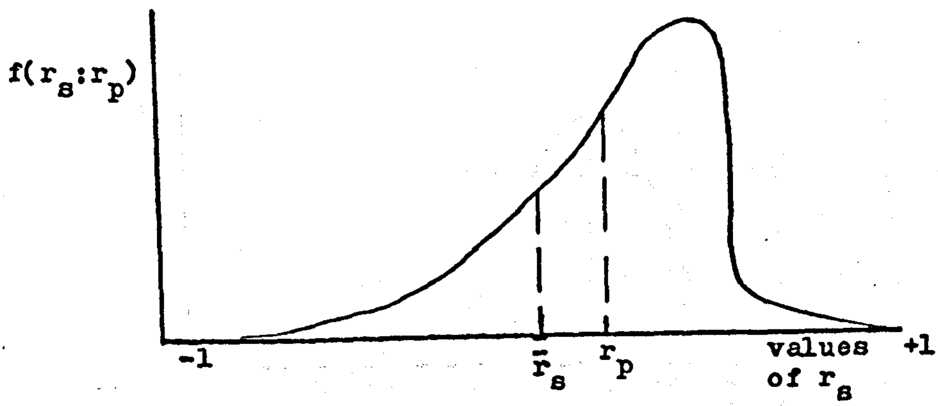
We have observed a statistic of value s, eg the correlation between educational attainment and subsequent occupational attainment, and the value of s is greater than zero, ie the correlation is positive. Perhaps as a result of the inherent conservatism of a certain type of statistician (and a certain type of quantitative sociologist), a test is made to see if this correlation could have occurred 'by chance' - that is, the population correlation parameter is zero. The statistic could not have occured by chance for at least two reasons : The probability of a statistic of any value occurring is zero (though the probability density may not be zero); and, Nothing ever happens by chance. It is also worth commenting that no sociologist worth his salt would ever hypothesize that there was no association between educational attainment and occupational attainment. (For five samples from different types of area, Sewell et al (1970:1018) find correlations of .634, .630, .568, .581 and .618;

the correlation calculated on the five samples taken together is .618).

The sampling distribution of the correlation statistic is distinctly non-normal, particularly for values of the correlation parameter nearer to unity. The distribution is skewed with a long negative tail for positive values of the correlation parameter (and vice versa). This is shown in Figure 2.2.2; note that the mean of the sampling distribution, $\bar{r}_s$, is less than the value of the parameter r_p. This is why the correlation statistic r_s is supposed to be a biased estimator of r_p - in a large series of trials the mean of the statistics, $\bar{r}_s$, is not equal to r_p. Whether this is important is another problem, for why should we worry about the <u>mean</u> of the sampling distribution? Fisher discovered a way of transforming the values of the correlation statistic, so that the transformed values of the statistic were normally distributed around the value of the transformed correlation parameter. The transformation is called Fisher's z-transform, or his arctanh transform of the correlation coefficient, and is equal to t where

Figure 2.2.2

The Sampling Distribution of the Correlation Statistic



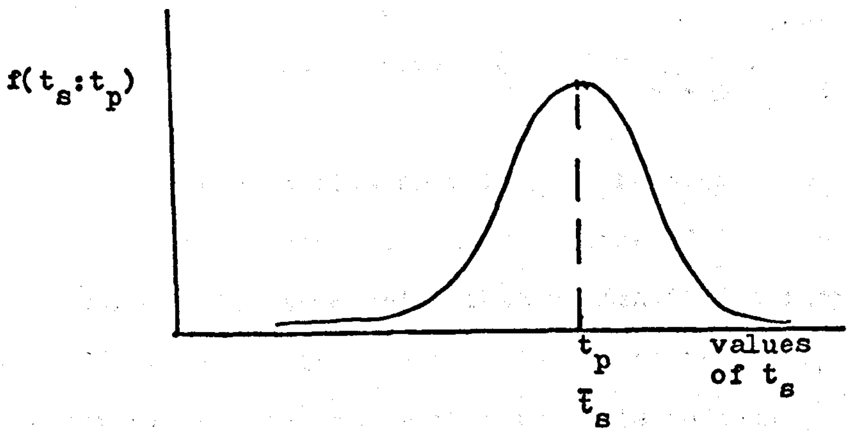
$$t = \text{arc tanh}\,(r)$$

(conversion tables are readily available). The distribution of t_s (ie transformed r_s) is shown in Figure 2.2.3.

It is evident from Figure 2.2.3 that t_s is unbiassed because the mean of the statistics, $\bar{t}_s$, is equal to the value of the parameter t_p. This is a clear indication of the relativism that exists in the

Figure 2.2.3

The Sampling Distribution of the Transformed Correlation Statistic



defintion of bias, for it would seem that in theory we could always find some transform so that the transformed statistic is unbiassed.(Bias is different from consistency which asks that, as the sample size increases, the value of the statistic approaches that of the parameter.) I will remain unworried about accusations of bias, just as many statisticians are unworried about inconsistencies in bias.

A rather more important point arising from the use of the arc tanh transformation is that we are now in the position of knowing the variance of the sampling distribution of t_s exactly, no matter what is the value of t_p. The variance is $\frac{1}{n-3}$ where n is the sample size, (ie $d^2 = \frac{1}{n-3}$).

Previously I gave five correlations between Occupational Attainment and Educational Attainment and came to the conclusion that no sensible person would expect the true correlation to be zero. A reasonable hypothesis is that there is no real difference between the areas, and that any differences are due to sampling fluctuations. The correlation statistic calculated for all the samples together is .618 and this value (being calculated on a total sample of 4388) is probably close to the true common value of the different correlation parameters for the five areas - if such a common value exists. We set up the value .618 as the value in the 'null' hypothesis, (Kendall and Stuart (1973:171) call it the 'hypothesis under test' and refer to the level of significance as the 'size of the test' -

in this case I will use the usual terms, the size and 'power' of tests I will consider anon). The sample of 857 people from a 'Farm' area have a correlation of .634, and we wish to find out whether we can accept the null hypothesis that the population correlation (for the Farm area) could be .618. (Data from Sewall (1970:1018)).

First a sampling distribution is set up, with t_p = arc tanh .618 = .722 and $d^2 = \frac{1}{854} = .001171$. We then find that t_s = arc tanh .634 = .748 , so that the difference $t_s - t_p = .026$; as d = .03422 we can form a standard normal deviate $z = \frac{t_s - t_p}{d}$ = .769 . From tables of the normal distribution function, the probability of getting a value of z less than or equal to .769, is .779. The significance level of this difference is 1 - .779 = .221, and so the null hypothesis that the correlation parameter is .618 is 'accepted'. (Kendall and Stuart (1973: 171) suggest that 'decide that the observations are favourable to' may be used instead of accept, '... If the reader cannot over-come his philosophical dislike of [this] admittedly inapposite expression ...' .)

It might be tempting to say that the relative likelihood of the correlation parameter being .634 to the likelihood of the correlation parameter being .618 is equal $\frac{.779}{.221}$ or about 3.5 . This is not really true, for if we introduce z (= .769) into equation (2.2.10) the likelihood ratio is exp $\frac{.591}{2}$ = 1.34 . We have, then, two estimates of the relative likelihood of two parameter values : the value 1.34 I believe to be easily interpretable, and the status of the value of 3.5 less readily interpretable. (However, the repeated sampling ideal is so heavily entrenched that in the likelihood ratio tests of Neyman and E S Pearson, the likelihood ratio is turned into a statistic itself. E S Pearson (1966) commenting on Barnard's suggestion that the likelihood ratio is enough in itself, said '... We certainly considered [this suggestion] but soon realized that the value of [the likelihood ratio] itself ... could provide no clear guide on which to base conclusions ...' .)

In deciding upon the significance of a result, we are forced, in the traditional tests, to reject a hypothesis because (as Jeffreys (1961:385) remarks)

'... because it has not predicted observable results that have not occurred ...'. Jeffreys is saying that from the sampling distribution (ie the (potentially) observable results) the area in the tail beyond the observed value (ie some of the (potentially) observable results which have not occurred) is used to decide whether the hypothesis is acceptable. As Edwards (1972:177) comments : Why choose the <u>tail</u> area ? (any area of similar size would be as powerful a way of rejecting hypotheses - Fisher says that in rejection an exceptionally rare chance has occured but cf my previous comments). Jeffreys , in the same work, comments that a null hypothesis is merely something set up like a coconut to stand until it is hit : and unfortunately this is only too true, because in cases where we have 'replications' (but also small sample sizes) we might never achieve 'significance', and after a time the null hypothesis will become accepted as 'fact' - for no good reason at all. (An excellent example of this is given by Bill Bytheway (1975), and I would like to acknowledge my debt to him for many valued discussions.) This brings us back to the relative likelihoods and will be extended in sub-section 2.6

Up until now, in this sub-section, though I have treated of the null hypothesis as if it had a complement, so that the relative likelihood (likelihood ratio) was between the complement to null hypothesis, in the Karl Pearson-'Student'-Fisher null hypothesis there is no complement. Fisher fully denied Neyman's principle that there could be no statistical test of a hypothesis without reference to rival hypotheses, (but see Hacking's (1966:81-83) dismissal of Fisher's denial). This has led to one fairly trivial, though over-rated, exegesis of the contemplation of rival hypotheses, that of 'Types of Error' (to call them 'errors' is very revealing). I refer to Type 1 and Type 2 errors. A Type 1 error - an error of the first kind- occurs when we reject the null hypothesis when it is true (the probability that we set, that this will occur, is called the 'size' of the test, ie the confidence level). A Type 2 error - an error of the second kind - occurs when we accept the null hypothesis when it is false (unity minus this probability is called the 'power' of the test). Much is made of this difference in intròductions to the topic of inference; but, again, it only really makes sense in the context of acceptance procedures. Fisher's (1955:73-74) argument is very telling here : firstly, he says that in most milieu '... [the errors] are essentially of one kind only and of equal theoretical importance ...'; he continues by pointing out that in scientific investigations we can from the sampling distribution find the extent of the error of the first kind - here I would disagree, see above - but, we cannot find the extent of the error of the second kind merely from consideration of the 'null' hypothesis because we do not know the distribution of rival hypotheses. Fisher does not actually say distribution, but '... not only on the frequency with which rival hypotheses are in fact true, but also how closely they resemble the null hypothesis ...'. He then makes the point, again, that '... In an acceptance procedure ... acceptance is irreversible, whether the evidence for it was strong or weak ...'(my emphasis): later he makes the idealized point - which is often forgotten in practice, perhaps due to the influence of acceptance procedures - '... the conclusions drawn by a scientific worker from a test of significance are _provisional_, and involve an intelligent attempt to _understand_ the experimental situation ...' (It might also be noted that, because there are often many tests of the same size, the maximization of the power of a test becomes a way of deciding between competing, and often conflicting, tests.)

2.3 Bayesian Methods

In sub-section 2.1 I briefly mentioned Bayes' Theorem, and all the systems of inference called Bayesian start from the idea of a <u>prior</u> distribution of feasible values of a parameter. That is, so we are told, we rarely start out with <u>no</u> idea of the possible values of a parameter : eg a correlation is almost certainly positive ; these two regression coefficients have to be positive and sum to unity; etc. This is turned into a <u>prior</u> distribution of values of the parameter, where the probability densities show how likely we think are these values <u>before we perform our study</u>. In previous terminology, we start out with knowledge $f_1(p:*)$ - which I will shorten to $f_1(p)$. As well as this information we have the result of a study which provides a set of likelihoods of the parameters, given a sample statistic, ie $L(p:s)$. The result of multiplying the prior density of a possible value p by its likelihood is proportional to the <u>posterior</u> density of the value p - that is, what we now think about these possible values of p after we have performed our study. The posterior density $f_2(p)$ is given by

$$f_2(p) \quad = \quad k\, L(p{:}s)\ f_1(p) \qquad .1.......(2.3.1)$$

where k is constant.

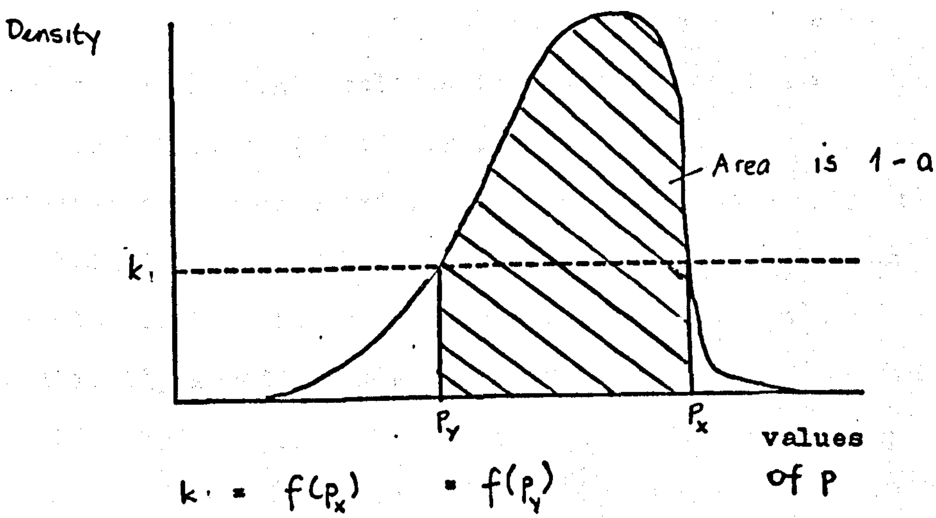
The Bayesian estimate of the parameter is given by the <u>mean</u> of the posterior distribution - although frequently, because of mathematical difficulties, the <u>mode</u> is found instead: the Bayesian confidence intervals are chosen so as to find an interval, (p_x to p_y , say) such that $F_2(p_x{:}s) - F_2(p_y{:}s)$ $= 1 - a$ (where a is the confidence level). This is an impossible task (and this is true for frequentist confidence intervals as well), unless we place an additional restriction on values of p_x and p_y - the restriction is that $f_2(p_x) \; = \; f_2(p_y)$. (See Figure 2.3.1). This interest in the mean on one hand and probability densities on the other is inconsistent. It is perfectly feasible with very skewed distributions to find the best estimate to be the mean, but for the mean to be outside the Bayesian confidence limits ! Ontologically, the mean and the Bayesian limits are as alike as chalk and cheese. Edwards(1972:67) makes the point '... Bayesian solutions are constrained to provide a single [estimate of a parameter] as an answer, we may view the argument over the appropriate prior as equivalent to the interminable arguments over point estimation ...'; and earlier (p 59) he comments that the Bayesian formulation is concerned with the probability of a parameter lying in an <u>interval</u>, but when they concern themselves with probability <u>densities</u> this shows that they have become interested in <u>point</u> comparisons, '... they should come clean and adopt a point formulation ...'.

Perhaps the most common criticism of the idea of a prior <u>distribution</u> is that which comments that the prior distribution of p, is different from the prior distribution of p^2, and this leads to different estimates for p. If, for example, we estimate $q = 1/p$ the Bayesian estimate of q does not equal the reciprocal of the Bayesian estimate of p : this is only the same as saying the harmonic mean is different from the arithmetic mean. Edwards (1972:20) writes (over-strongly I think) '... [this problem with transformation of the parameter] turns out to be one of the most telling points against the use of probability as a measure of belief in hypotheses ...' (my emphasis), obviously this would be without point if we estimated the value of the parameter by the median of the distribution.

Suppose that we have undertaken a study and have found the posterior distribution $f_2(p)$, it is always possible to use this as a prior distribution for use in another study, and this can be repeated over any number of studies. If we assert : at one point in time we were completely unknowledgeable about the likelihoods of the possible values of the parameter; and our prior densities we possess now are the result of previous studies; then we must have a prior distribution which follows the the form of the likelihoods L(p:s). This assertion runs counter to various trends in Bayesian thought, but is the approach taken by the Empirical Bayesians - who, it might be said, are not accepted by Bayesians in general. To use Veatch's(1974:195) graphic prase about Popper, in forming a prior distribution of a parameter, Bayesians '... so far from basing their hypotheses on the data, simply make them up out of whole cloth ! ...'. Jeffreys (1961:37) suggests that disputes about the form of the prior densities be referred to an 'International Research Council' : this needs a touch of Realism and less Idealism, (in fact Jeffreys' Idealism is shown in his preoccupation with, and attempts to measure precisely, 'Simplicity' (1961:47)).

Figure 2.3.1

Bayesian confidence interval p_y to p_x for an unknown parameter p



2.3.1 Bruno de Finnetti and 'Probability does not Exist'

De Finnetti has long been a leading exponent of the subjectivist school of statistical inference - as he says in the Preface to his 'Theory of Probability' (de Finnetti,1975): '... My thesis, paradoxically, and a little provocatively, but nontheless genuinely, is simply this : PROBABILITY DOES NOT EXIST ...'. De Finnetti sees all probability statements in terms of bets - despite being a 'subjectivist' he is almost a 'positivist' in that he clings to this operational definition - if we say that the probability of heads is ½, what we really mean, he says, is that we would be as willing to bet on heads as we were to bet on tails. This relates to the rain problem : the probability of heads is ½, but we only ever observe a head or a tail. He says '... that an objective probability [eg of heads] is always unknown, although hypothetical estimates of its value are made in a not really specifiable sense ...' (quoted in Hacking(1966:213)). De Finnetti would have us betting on nothings; if probability does not exist, why should we ever bother to bet ? De Finnetti cannot get round this basic problem : If there is no such thing as an objective probability of heads being ½, we would be inclined to think it was some other number near to a ½, but we always bet as if it were exactly a half - I hold that the reason for this can be reduced to the physics of a circular disk, and that the probability of a ½ in a certain set of circumstances is a property of the disk (as is its wind-resistance). It is an accident of the coin, that it is metallic : it is part of the essence of a coin (revealed in a conceptual change) that both sides are as likely to appear uppermost. If a coin after a fair number of tosses had a large surfeit of heads, I would still take a lot of convincing that this was not a rare occurence from a fair coin before I questionned its fairness.

At the same time, because I believe that 'probability' can be a characteristic of reality, I am very suspicious of using 'probability' as a device to show belief. There are (as I noted in the first chapter) many different probabilities, and de Finnetti would give ontological priority to the probability as an inclination to bet : I see the various probabilities in an Aristòtlean pros hen equivocal sense (see Appendix A) - all the different probabilities are connected in the end to concept of relative numbers in discrete categories. Bayes' original idea was expressed in proportions and not probability densities!

Note, however, that in Bayesian methods what we think is modified by what happened before the present study, whereas in the frequentist school there is no way of taking such things into account, (but see sub-section 2.6).

2.4 The Method of Support

Maximum likelihood chooses that value of the parameter which maximizes the value of L(p:s). Often maximum likelihood (ML) and frequentist estimates of the same parameter do not agree, though as the sample size increases they tend to the same value in many cases. (For example, the ML estimate of the population variance is the sample variance (the statistic), in the case of the usual unbiassed estimate the population variance is estimated by the sample variance multiplied by $n/(n - 1)$.) The maximization of L(p:s) can be seen by analogy to be as choosing the mode of a distribution : but why choose the mode ? There have been many criticisms made of ML methods : and one is the possible existence of multiple maxima (think of a hilly range); others are due to silliness of ML estimates in some circumstances - or the existence of no possible ML estimate ! It is also worth considering whether the most likely value of the parameter is the best value to choose or whether one should not really be interested in providing an iterval. Analogizing to the distribution of income, say, it is like saying that the best estimate of the average wage is the most frequent wage, is the interquartile range preferable? Maximum likelihood is concerned with point estimation and not with interval estimation (and as such is not an inferential method).

The Method of Support, excellently explained by Edwards (1972) building upon parts of Hacking (1966), tries to take L(p:s) as the foundation of inference. Let L(p*:s) be the maximum likelihood, corresponding to the ML estimate of the parameter, p*; Edwards tells us to take the two(?) values of p which satisfy

$$m = \log\left[\frac{L(p^*:s)}{L(p\ :s)}\right] \qquad \text{.........(2.4.1)}$$

This Edwards (1972:180-197) calls the m-unit support limits, and a m-unit support is equivalent to a relative likelihood (R) of exp(m), (m=2, R=7.39; m=3, R=20.09). For a normal sampling distribution of s, 2-unit support gives the same limits as two-tailed confidence test of size .0455; for 3-unit support, the corresponding size is .0143 (se Table 2.4.1 for some relevant tabulations. It is interesting that the value of R is always much less than the ratio of the included area to the area outside the conventional limits). I do not find the use of support limits defensible, for it is possible with multiple maxima (hills, some bigger than others) to have many values of p which satisfy (2.4.1), partly depending on the value of m. (This is very similar to the Bayesian intervals - see Figure 2.3.1 - when the restriction is not only on included area, but on relative likelihoods.) It is interesting to read '... the object of this book is to describe a method which will enable us to "assess the relative merits of rival hypotheses" ...'(Edwards,1972:54), and this may be why the m-unit limits do not seem to convince - to use Edwards' own point concerning the Bayesian confusion of point and interval estimation, to adopt an interval formulation shows that a point formulation is not sufficient (even though we arrive at the interval by a point formulation).

Edwards innovates to introduce the idea of a <u>prior</u> likelihood, and in essence and practice is almost

identical to the Bayesian formulation of equation (2.3.1), but the difference is supposed to lie in the nature of the prior information. The equivalent to equation (2.3.1) is :

$$L_3(p:s,*) \;=\; L_2(p:s) \times L_1(p:s) \qquad \ldots\ldots\ldots(2.4.2)$$

where this is to read as : the posterior likelihood is equal to the product of the likelihood from the sample and the prior likelihood. (And the posterior likelihood can be used as a prior likelihood in another study). In practice Edwards works with the log-likelihood , which he (after Hacking) calls 'Support' so that

$$S_3(p:s,*) \;=\; S_2(p:s) \;+\; S_1(p:*) \qquad \ldots\ldots\ldots(2.4.3)$$

Again, the ability to incorporate prior information is a valuable facet not available with frequentist approaches.

TABLE 2.4.1

Values for the standardized normal distribution.

Deviate	Size of 2-tail test	Relative likelihood	Logarithm of Relative Likelihood
0.0	1.000	1.000	0.000
0.1	0.920	1.005	0.005
0.2	0.841	1.020	0.020
0.3	0.764	1.046	0.045
0.4	0.689	1.083	0.080
0.5	0.617	1.133	0.125
0.6	0.549	1.197	0.180
0.7	0.484	1.278	0.245
0.8	0.424	1.377	0.320
0.9	0.368	1.499	0.405
1.0	0.317	1.649	0.500
1.1	0.271	1.832	0.605
1.2	0.230	2.054	0.720
1.3	0.194	2.328	0.845
1.4	0.162	2.665	0.980
1.5	0.134	3.079	1.125
1.6	0.110	3.597	1.280
1.7	0.089	4.242	1.445
1.8	0.072	5.053	1.620
1.9	0.057	6.079	1.805
2.0	0.0455	7.389	2.000
2.25	0.0245	12.57	2.53125
2.5	0.0124	22.76	3.125
2.75	0.0060	43.87	3.78125
3.0	0.0027	90.02	4.500
3.25	0.0012	196.6	5.28125
3.5	0.0005	505.3	6.125
∞	0.0	∞	∞

2.5 Fiducial Inference

A common characteristic of the Bayesian and Support methods is the combination of information, the Fiducial approach as originally envisaged by Fisher did not have such a facility. In essence, as I define it, Fiducial inference consists of regarding the likelihood af a parameter as a probability density. Hacking (1966:133) notes '... Apparently the fiducial probability of an hypothesis [eg the possible value of a parameter], given some data, is the degree of trust you can place in an hypothesis if you possess only the given data ... Fisher gave ho general instruction for computing his fiducial probabilities ...'; Prof George A Barnard in privately circulated draft 'Theory of Estimation' takes the view that it is permissible '... <u>in the absence of any other knowledge about</u> [p], to regard [p] as itself a random variable having a distribution [of a specified form] ...'. I agree with Barnard that '... arguments which appear to depend on the fiducial distribution ... in fact do not require its acceptance; and many practical applications ... will be the same, irrespective of whether the argument is accepted or not ...'; and if a distinction is made between the prior distribution of p, the sample (fiducial) distribution of p, and the posterior distribution of p - suggestively we might write

$$f(p:s,*) = f(p:s) \times f(p:*) \qquad \text{.........(2.5.1)}$$

This is not satisfactory in itself, for it implies that p = p+p (equation (2.5.1) is of the form of the sum of varaites, (eg Kendall and Stuart,1969:263-264)). What is needed then is a constant term k(-), which adjusts the area under the curve to unity. We write

$$f(p:s,*) = k(-) \times f(p:s) \times f(p:*) \qquad \text{....(2.5.2)}$$

(To reiterate an earlier point, in the end this is numerically eqivalent to some applications of Bayesian and Support methods.)

The reason for the emphasis on the distribution of the parameter lies in the earlier problem of inconsistencies in point and interval estimation (when both appear in the same analysis). What we are trying to do in point estimation is to locate the middle of the parameter distribution - though in Support the distribution does not exist. We are trying to estimate a <u>super</u>-parameter of the distribution of possible parameter values. I personally believe that the best estimate of the value of the super-parameter is the median, with o' .er quantiles providing the various intervals. (Note that this, in fact, does not commit us to use of a distribution ; areas under curves existed before probability densities.)

Further, I doubt the utility of estimating a probabilty density or likelihood corresponding to a single value (or the contemplation of ratios of such). Expressing interest in a single value (eg, earlier, a single value of a proportion or a single value of a correlation) does not seem of great importance. At what point do we consider a correlation to show the existence of 'skill' in prediction of Football League Tables ? (A point missed by Hill(1974) who assumes any value greater than zero will do - he also seems to assume (by scoring them as one) all values greater than zero, are as equally likely to be really representative of a positive correlation).

A fiducial confidence interval <u>is</u> a confidence interval because it represents confidence in a <u>parameter</u> (not in a sampling distribution of statistic values). Of course the arrangement of the limits can be subject to all the arrangemental criticisms of traditional frequentist intervals. (The estimation of the location of the super-parameter for the probability parameter is examined in Appendix D).

2.6 Combining information in frequentist studies

Up to now we have seen how built-in to the methods of Support , Bayes and fiducial inference there is a facility for combining the results of earlier studies. Fisher (1958:99-101) strangely enough provides such a measure for frequentist methods, and in essence it is very simple.

Fisher suggests that if we have a test of size P (ie a confidence level of P) we convert this to a value of chi-squared equal to a test of this size. It so happens (Fisher,1958) '... in the case of [2 degrees of freedom] , the natural logarithm of the probability [ie the size] is equal to to $-\frac{1}{2}X^2$. If therefore we take the natural logarithm of a probability, change its sign and double it, we have the equivalent value of X^2 for 2 degrees of freedom. Any number of such values may be added together, to give a composite test ...'. He then provides an example of three studies with sizes

(P) equal to .145, .263, and .087 which give values of -log(P) equal to 1.9310, 1.3356, and 2.4419. these latter three values are added together and the sum is 5.7085; this value is then doubled to 11.4170 and as each constiuent chi-squared has two degrees of freedom, this value corresponds to a chi-squared with six degree of freedom. The size of the test is between .05 and not far from .075.

Barnard has commented (personal communication) that there is no reason to use -log(P) other than convenience; he suggests that if we have reason to think that one or more samples are particularly 'good', eg they seem on <u>theoretical</u> grounds to be capable of providing a more accurate answer (which would not influence the calculation of things such as standard errors), then we could find the value corresponding to a certain probability value from the tables of chi-squared with four degrees of freedom, say. The obvious draw-back to this is that it is rather more tedious, but does allow weighting by non-statistical criteria.

I hope I have now shown that when Zetterberg (1966) suggests that we combine results from different studies by (p 162) an unspecified '... well-known law of probability calculus ...', he is describing an unknown animal. Actually, his form of 'confidence appraisal' for the flow-charted example (p 109) is the ubiquitous chi-squared (with (p 105) no clear statement of his statistical null hypothesis), so he could possibly mean Fisher's method - I have a strong suspicion that all he means is the multiplication of likelihood ratios !

CHAPTER 3

MATHEMATICAL MANIPULATION

'... Quantities which can be specified by giving just one number (positive, negative or zero) are called <u>scalars</u>. For example, temperature, density, mass and work are all scalars. Scalars can only compared if they have the same physical dimensions. Two sudh scalars measured in the same system of units are said to be equal if they have the same magnitude (absolute value) and sign ...' (A I Borisenko and I E Tarapov. <u>Vector and Tensor Analysis with Applications</u>).

Earlier I commented that the equalities pf equations (2.1.6a,b) did not imply that a likelihood had the same ontological status as a probability density (we equated values). This can perhaps be made clearer by thinking of the assignment

L := F ; ....(3.1)

in Algol 68. Algol 68 in its 'Informal Introduction' (Lindsey and van der Meulen, 1971) is revealed to be a brilliant jewel of logic - in an area where logic is needed, but to judge by FORTRAN sadly lacking. I will not go into this in great detail but two types of concept are important : there is the way in which we look at the numbers we store; and there is the idea of a 'clause' always delivering a 'value'. We need the ideas of a location with a certain name possessing a variable of a certain value : that is, there is a place in core (the location), and we refer to the location by a destination (the name), within this location there is stored some information (a variable of a certain value is possessed). (3.1) then becomes : 'the value of the variable possessed by location with name F (or simply 'F') is made the value of the variable possessed by L'. L and F thus have the same <u>value</u>, but they are different things. If the types of variable in locations L and F differ (L can only possess an integer variable and F can only possess a string variable) then the value of the variable possessed by F cannot be copied over to be possessed by L.

A further extension of this is

Z := X*Y ....(3.2)

'The value possessed by X is multiplied by the value possessed by Y and the value delivered by the clause X*Y is then possessed by Z'. In the 'Informal Introduction' (p 251) there are only nine possible combinations of variable for which the clause X*Y delivers a result. That is, X can possess a variable of mode integer, real or complex, and Y can also locate a variable of any of these three modes. Although <u>in mathematical notation</u> X and Y could name say, matrices , in Algol 68 the compiler confronted with (5.2) when X and Y were matrices (of mode, 'row-of-row') the compiler would produce an error flag - I sometimes wish we were as choosy as compilers ! All is not lost, for in Algol 68 we can give an <u>extra</u> meaning to a pre-existing operator : in this case we could show how we would wish '*' to be interpreted . if the compiler discovered that X and Y were row-of-rows. (See Lindsey and van der Meulen (1971:319-323) for an eloboration of this.)

My point which I may be making rather circuitously, is that we can write many equations which seem mathematically correct, but are in reality undefined on the variables upon which we are operating. (Synge (1970;Ch 3) makes this point very clearly : mathematical operations only do what we design them to do, and sometimes we have to extend the definitions to cover new realms. In the case af multiplication, nobody is now likely to challenge that usually X*Y $\neq$ Y*X in the case of matrices.)

3.1 Partial Differentiation and Interaction

Macdonald (1972) wrote a paper in which he took the 'interaction' function

$$X_3 = X_1 * X_2 \qquad \text{.........(3.1.1)}$$

and partially differentiated the function by X_1 to find

$$\frac{d\, X_3}{d\, X_1} = X_2 \qquad \text{........(3.1.2)}$$

(I have removed various superfluities from the exposition and haven written d instead of the partial differential duff - ∂). Commenting on this (Allan,1972) I differentiated (3.1.2) as product of functions to obtain

$$\frac{d\, X_3}{dX_1} = X_2 + \frac{d\, X_2}{dX_1} \qquad \text{...........(3.1.3)}$$

and noted that it was difficult to interpret what the partial differentials meant in practival terms. I think Macdonald's inspiration for (3.1.2) comes from Boudon's (1969:221-227) paper in which he makes the statement '... Let us examine a typical linear equation such as

$$x_3 = a_{13}x_1 + a_{23}x_2 + e_3$$

The effect of, say, x_1 on x_3 is given by taking the partial derivative of x_3 with respect to x_1. This derivative may be interpreted as the rate of variation in x_3 when x_1 varies, the third variable x_2 having a fixed value. In the case of [the above linear equation], this derivative is

$$\frac{d\,x_3}{dx_1} = a_{13} \qquad \text{...'}$$

As I noted, Macdonald's ideas, repeated is his (1973:75), are either taken from Boudon or are very similar, so I will only consider the consequences of Boudon's position. Firstly, his derivative is wrong because x_2 is a function of x_1, as much as is x_3, for the values taken by x_1 and x_2 jointly ~~are~~ related by the joint probability density function $f(x_1,x_2)$. Secondly, and this follows from the first, if the value of x_2 is fixed so are the frequencies of the possible values of x_1 and thus the value of the derivative will not be constant, across the board. Thirdly, the derivative - but it is not a derivative - is the derivative of the expected value of y with respect to x_1. (This leads to the mathematical horror $\frac{dx_1}{dy} \neq \left(\frac{dy}{dx_1}\right)^{-1}$; which, despite of all possible reasons to the contrary, is a clear proof that the Boudon partial derivative is not that of the rest of mathematics.) This series of misunderstandings are a direct result of applying mathematics to variables the mode of which is such that the operation of partial differentiation has yet to be defined. Unfortunately, at least one reasonable stricture upon the extension of the meaning of an operation is that in use the new use of the operation on new material looks as similar as possible to pre-existing uses of the operation - otherwise invent a new symbol and concept. Boudon's use falls down here. I do not have to go further to show that the partial differentiation of Boudon and Macdonald does not follow conventional mathematical usage. (Macdonald (1973:75) seems singularly unimpressed by the lack of convention in his use of the operation.)

To go further we have to consider how two variates (variates, note, not mathematical variables) can be functionally related. They are related via the density function f(x,y); with two variates we do not have only the two-dimensional space spanned by the variates we have three dimensions where the third dimension is the probability density - this is called the bivariate surface. If the two variates are jointly bivariate normal, and we 'look' down onto the plane spanned by the two variates, we 'see' a 'hill'. Drawing contours to 'map' the hill we find a series of concentric ellipses - shown in Figure 3.1.1.

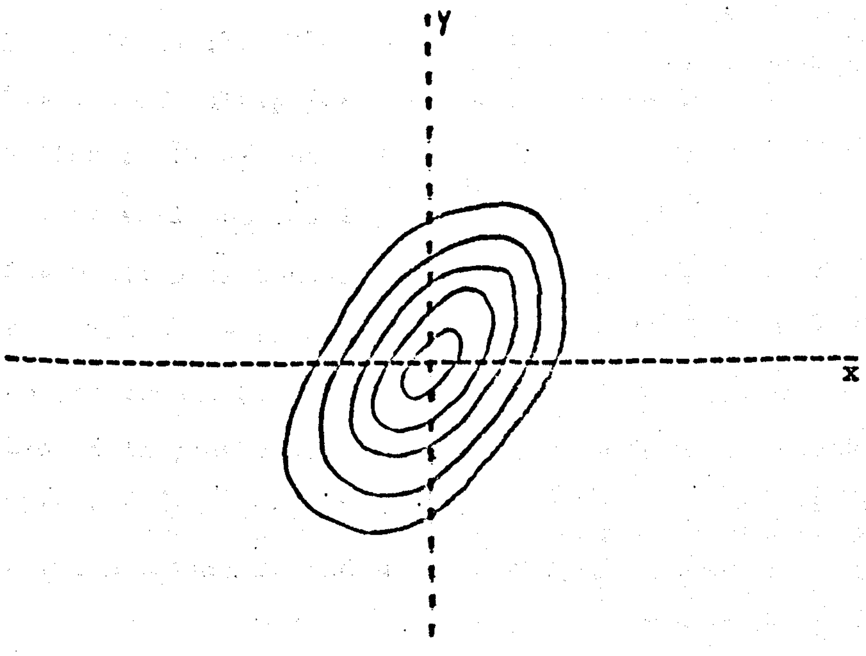
If we examine an ellipse it shows at any point the partial derivative as a tangent to the ellipse. The slope sensed with y as 'up' for any pair of values is

$$\frac{dy}{dx} = \frac{x - ry}{rx - y} \qquad \text{.........(3.1.4)}$$

where r is the correlation coefficient (the proof is in Appendix G). From (3.1.4) we can work out that at the mean of x (which is x=0), the derivative is always ry/y = r for all values of y. This is the derivative which agrees with that of Boudon <u>if</u> he dealing with standardized vardiables taken at their mean value. If $x = 1$, the derivative is $(1 - ry)/(r - y)$ which varies according to the value of y chosen - this can be visually inspected from Figure 3.1.1 . As for what the coefficient a_{123} of the 'interaction' term $a_{123}x_2x_3$ means as a derivative - this is lost in complexities .

Figure 3.1.1

Density contours upon a bivariate normal distribution



3.2 Multiplying Values of Variates

Ford and Box (1967:289) make a bold statement '... In choosing between alternative actions, a person will choose the one for which, as perceived by him, the (mathematical) value of p x v is the greater, where p is the probability of the action being successful in getting a given reward, and v is the value of the reward ...'. It is this powerful mathematical statement that they 'test' by use of mainly 2x2 tables and chi-squared. (In fact, they test for an effect of probability and liking, but p x v is lost). Blalock (1969:155-165) also mentions the use of multiplicative relations in the modelling of interaction; and in discussions of ordinally-scaled variates the impossibility of testing for interaction is bemoaned, and any value is regarded as a label for an ordinal category as long as order is preserved. (See my unpublished paper in Appendix E on this problem). What seems to have happened is that authors do not know, or have ignored, the difference between a variate and a mathematical variable.

Consider this mathematical expression :

$Y = X*X$ ....(3.2.1)

This is a perfectly good expression - Y has a value which is equal to the value of X-squared. Suppose, though, that X is a normally distributed variate : to be true for variates Y must be a chi-squared variate. In practical terms, this means that we cannot entertain the truth of (3.1.5) if X is normally distributed and Y is not distributed as chi-squared.

In essence this means that we cannot, for variates, choose an formula and estimate it, unless the distributional information fits. This is almost the same as the concern about dimension in physical science. Despite the occurence of ideas to the contrary by some sociologists concerned with 'mathematical experimentation' (eg R L Hamblin) ; we would never countenance the expression $E = mc^3$ and think it might be a possible viable equation in the way we think that $E = mc^2$ could be viable. E(energy) has the dimensions L^2MT^{-2} (L = length, M = mass, and T = time); m(mass) has dimension M, and c(speed of light) has dimensions LT^{-1}. If $E = mc^3$ was to be at all feasible

L^2MT^{-2} would have to equal ML^3T^{-3}

which is perfectly unfeasible. (Checking dimensions shows that $E = mc^2$ is at least feasible.)

I have just chosen in (3.2.1) an extremely simple example and, if it is so complex for this simple example, we have to be very careful in more complex formulations to be writing sense. I do not feel any more need be said.

In this thesis I have attempted to produce a rational approach to the use of mathematics and quantitative methods in sociology and , although occupational choice and social mobility have only appeared by the scruff of their respective necks, my attitude has been that I spelt out in detail in the first chapter - one can learn a lot from a few examples, providing one has a cogent philosophical position. In fact, my interest in the mathematics of social theories was originated by the ire raised by my reading of the Ford and Box article. (I could have written a thesis on the various mathematical shortcomings of their paper alone, with lovely words like 'level of measurement' and 'Abelian Groups' freely spattered around. But there is more to the use of mathematics in social theories than the mere structure of the theories - there is the whole panoply of problems of inference and <u>testing</u> these theories. And, before you test, what do you think is the relationship between your concepts and reality ? And what is the nature of reality ?)

I have ended with this chapter with what I consider to be <u>some</u> of the worst examples of uninformed mathematical licence in the literature. They are not the worst; yet at the same time they seem to be damnable because they look mathematically sophisticated to the ingenuous. (I hope I have shown they are themselves distributionally ingenuous). One only has to read some of the examples of variance-splitting which abound in the literature - some into the milliards of order of interaction (eg Hope,1971) - to feel that if only the distributions were studied, many worthless adventures would never leave base.

Something I started with, and something with which I will end. Sociology tries to find out whether "All coppers are narners".

REFERENCES

Allan G J B (1972) ~~A~~ comment on 'Mobility and work satisfaction : a comment'. BSA Mathematics and Computing Group Newsletter.

Allan G J B (1974) Simplicity in Path Analysis. Sociology.

Bachelard G (1951) Philosophical Dialectic of Realativity theory. In Schlipp (1951)

Bachelard G (1968) The Philosophy of No.

Blalock H M (1969) Theory Construction. Prentice-Hall.

Boudon R (1968) A New Look at Correlation Analysis. In Methodology and Social Researdh (ed H M Blaock and A B Blalock).

Byetheway B (1975) Social Class and Self-Reported Delinquency. International Journal of Criminology and Penology.

Edwards A W F (1972) Likelihood. Cambridge U P

Einstein A (1951a) Autobiographical Notes. In Schlipp(1951).

Einstein A (1951b) Remarks to the Essays appearing in this Volume. In Schlipp (1951).

Finnetti B de (1972) Probability, Induction and Statistics. Wiley.

Finnetti B de (1975) Theory of Probability (in 2 vols). Wiley.

Fisher R A (1955) Statistical Methods and Scientific Inference. Journal of the Royal Statistical Society, (Series B).

Fisher R A (1958) Statistical Methods for Research Workers (13th ed). Oliver and Boyd.

Ford J and S Box (1967) Sociological Theory and Occupational Choice. Sociological Review.

Hacking I (1965) Logic of Statistical Inference. Cambridge U P.

Hernandez-Cela C (1973) Probability and Empiricism in Science. (Ch 6 of Willer and Willer,1973).

Hill I D (1974) Statistical Inference and Football Scores. Applied Statistics.

Hindess B (1973) Use of Official Statistics in Sociology. Papermac.

Hope K (1971) Social Mobility and Fertility. American Sociological Review.

Jeffreys H (1961) Theory of Probability. Oxford U P.

Kendall M G and A Stuart (1969) Distribution Theory. Griffin.

Kendall M G and A Stuart (1973) Inference and Relationship. Griffin.

Lindsey C H and S G van der Meulen (1971). An Informal Introduction to Algol 68. North-Holland.

Macdonald K I (1972) Mobility and Work Satisfaction: A comment. BSA Mathematics and Computing Applications Group, Newsletter.

Macdonald K I (1973) Downwardly mobile mothers; and other interaction effects. In Mobility in Britain Reconsidered (ed J M Ridge), Oxford U P.

Maxwell G (1962) Ontological Status of Theoretical Entities. In Minnesotta Studies in the Philosophy of Science, Vol 3.

Moroney M J (1951) Facts from Figures. Pelican Books.

Nagel E (1961) Structure of Science. Routledge and Kegan Paul.

Pearson E S (1966) The Neyman-Pearson Story: 1926-34. In Festscrift for Jerzy Neyman.

Polanyi M (1962) Personal Knowledge. Routledge and Kegan Paul.

Reichman W J (1970) Use and Abuse of Statistics. Pelican Books.

Sclipp P A (1951) Albert Einstein : Philosopher -Scientist. Tudor Bookks.

Sellars W (1963) Science, Perception and Reality. Routledge and Kegan Paul.

Sewell W H , A O Haller and G W Ohlendorf (1970). The educational and early occupational status attainment process : Replication and revision. American Sociological Review.

Smart J J C (1963) Philosophy and Scientific Realism. Routledge and Kegan Paul.

Sorokin R A (1927) Contemporary Sociological Theories. Harper Torchbooks.

Synge J L (1970) Talking about Relativity. North-Holland.

Veatch H B (1974) Aristotle : A Contemporary Appreciation. Indiana U P.

Whittle P (1970) Probability. Pelican Books.

Willer D and J Willer (1973) Systematic Empiricism. Prentice-Hall.

Zetterberg H L (1966) On Theory and Verification in Sociology, (3rd ed) Bedminster Press.

AN INQUIRY INTO THE VALIDITY OF MATHEMATICAL METHODS USED IN EVALUATING THEORIES OF OCCUPATIONAL CHOICE

BY

G J BORIS ALLAN

APPENDICES

APPENDIX A

ON THE DIVERSITY OF METHOD

- ESSENTIALISM <u>SUPRA</u> NOMINALISM

ON THE DIVERSITY OF METHOD - Essentialism _supra_
Nominalism

In his critiques of scientific methods Popper, starting from a confusion of Platonic Idealism and Aristotqlean Realism (eg Popper,1965:31-34) comes to the conclusion that '... all theoretical or generalizing sciences make use of the same method, whether they are natural or social sciences...' (Popper,1961:130-131). I wish to show: one, Popper wrongly describes Aristotlean Essentialism, and we can learn from what is wrong in Popper's description; and,two, Popper's Unity of Method is non-existent and that Methodological Nominalism (if it existed) would lead to no _theoretical_ advance.

1. Aristotlean Realism

Popper defines Methodological Essentialism as '... The school of thinkers ... founded by Aristotle who taught that scientific research must penetrate to the essence of things in order to explain them ...' (Popper,1961:28 - my emphasis, note 'explain' and not 'predict'), Popper's comments on this school are scattered throughout his many works. I will examine part of his chapter 'The Aristotlean Roots of Hegelianism' from Book II of 'The Open Society and its Enemies', (Popper,1966) to illustrate the extent to which he is incorrect. If it were only that Popper were incorrect in his interpretation of Aristotle, that would be relatively unimportant, but the very fact of the mis-interpretation is crucial to his notion of the Unity of Method.

Popper(1966:6) argues '... Aristotle's version of Plato's essentialism shows only unimportant differences ...' , when really the two are (though essentialists) at opposing ends of a continuum. Plato's essentialism consisted of trying to find general 'Forms' within ordinary things, Idealizations which were somehow more Real than the ordinary things. Plato's Forms - such as 'Justice', 'Circularity' - are what philosophers often term Universals, in that many individual things (Particulars) can be described by them. A particular only exists in as much as it partakes of one or more of these Forms. Aristotle was concerned with 'substances' (which are particulars) and 'accidents' of substances, where priority was given to the substances. In the Platonic scheme a particular only existed in as much as it represented one or more Forms: in the Aristotlean scheme a particular existed, and the accidents were derivative of the substance , only existing as accidents OF a substance. For Aristotle there were two questions of Being, of Essence: What is the being of things that exist in themselves (substances)?; and, What is the being of things that have to exist in another (accidents)? (A'dog' is a dog, but'length' can only be applied to a substance such as a dog.) In Aristotle's Logic a substance is that which is always subject and never predicate; for example, 'Karl Popper' cannot be used to describe anything apart from Karl Popper, but there are many things that can be used to describe Karl Popper - 'accidents' such as he is a philosopher. A modern extension of this, which I feel would not be unacceptable to Aristotle, is to say these substances, these entities are real but there is a hierarchy of levels of belief such that it is easier to believe in the reality of Karl Popper than it is to believe in atoms being real - 'Karl Popper' is more credible than a 'muon'.

Closely allied to Aristotle's ideas concerning essence is the analysis of change, and it is the Aristotlean view of change which Popper next considers - I wish, however, to amplify the difference between What-questions and Why-questions. A What-question is (to use an example from Popper(1961:29)) of the form 'What is matter?' and, Aristotle holds, unanswerable by the deductive method - as was noted above a substance cannot be the predicate of a proposition, and to attempt to use other substances to predicate 'matter' will lead to an infinite regress. Popper considers the what-question to be metaphysical, and says matter is what he says it is; Aristotle says the answer comes from induction, knowledge of essences is not innate (otherwise the world would be transparent) but acquired by contemplation on the information available to us. Using the idea of a hierarchy of levels, we can see that the essence of matter contains the information that is composed of atoms, but matter (eg a table) is more to us than a collection of atoms - though Sellars(1963) holds that in accepting the reality of atoms we deny the reality of a table!

A why-question is concerned not with the essence of a a substance, but with an explanation of what things are <u>accidentally</u>, as against what things are essentially - why should a certain accident pertain to certain substance, what are the causes of such a thing happening ? It was for why-questions such as 'Why do men philosophize ?' and 'Why do black-holes absorb all elecro-magnetic radiation ?' that Aristotle developed the other branch of his logic; that is, deductive logic exemplified by the use of the Syllogism.

(Note that we will never really be able to explain why black-holes absorb all radiation untill we know what _is_ a black-hole : part of the essential nature of a black-hole is reflected in the accident 'absorbs all electro-magnetic radiation').

Popper (1961:29) gives a what-question 'What is Justice ?', and Aristotle points out that the being or essence implied in this sort of question (ie, the essence of an accident) is _pros hen_ equivocal. (Literally, it means '_with respect to one_, of many meanings'.) 'Being' is _pros hen_ equivocal because accidents can be said to have being, but solely in that the being of an accident can be understood with reference to the 'proper' being of a substance. Justice does not exist, have its being, on its own, for justice can only be seen in relationship with substances - people : Plato's Form (of Justice) is more real than the people. (Popper (1966:291) points to the many different meanings of 'puppy' - but we can see that the essences of these different substances called 'puppy' are _pros hen_, where the one to whidh they are referred is a young dog.)

2.Aristotle's Study of Change

Aristotle had a doctrine of four causes of change (where 'cause' is not necessarily an accurate rendering of the nature of the four), and he had three principles of change (ie requisite pre-conditions that were necessary if there was to be change of anything). The three principles were : privation, the initial state; the material principle, that which is changed; and the formal principle, the state towards which the change occurs. For example consider the education of a child : the change is a change of the child (material principle) from being in the state of being without knowledge deemed socially necessary (privation) to being with knowledge deemed socially necessary (formal principle). My typifications of the principles for this example are obviously ripe for argument, and this is the value of the Aristotlean approach - we immediately begin to think about what is happening with the change, rather than merely describing educational attainment. Two things to note: the formal principle need not be immutable; and the 'form' in the formal principle is not a Platonic Form, as Popper tends to suggest.

From these three principles Aristotle isolates two causes, the material cause and the formal cause - I shall refer to them as the material factor and the formal state. Aristotle saw these as necessary but not sufficient conditions for change, for you need something towards which to change and something which is changed but you need also an agent of change. (Privation is not really on the same footing as the other principles, indeed often one only discovers privation after change has occured. Miss Eliza Doolittle might not have realized that she was not an English-speaker.) The cause of change was called by Aristotle the efficient cause of change, or the agent of the change, and this was also a necessary condition. All three causes are severally necessary conditions, and only jointly are they sufficient for any change. This is an important point : in the context of historicism it denies the superordinacy of the formal state or, as Popper would have it, Hegel's destiny; and in the study of causality generally, it leads to the stricture that cause always accompanies effect - if the efficient cause is removed there is no change - this notion has been re-asserted as the need for the spatial and temporal contingency of cause and effect.

Compare this to Popper's(1968:59-60) causal analysis. '... To give a causal explanation of an event means to deduce a statement which describes it, using as premisses of the deduction one or more universal laws, together with certain singular statements, the initial conditions ... The initial conditions describe what is usually called the "cause" of the event in question ... And the prediction describes what is usually called the "effect" ...' (Popper gives an example of how a thread breaks when a 2 lb weight is hung from it). Given the scathing criticism he makes of Aristotle's causal analysis, Popper's own causal analysis is remarkably similar. Take his causal explanation of the breaking of a thread : Element (1) is a thread of tensile strength 1 lb; Element (2) is a singular statement 'The weight put on this thread is 2 lbs'; and Element (3) is a universal statement (scarcely a law) 'Whenever a thread is loaded with a weight exceeding that whidh is the tensile strength of the thread, then it will break'. Reworking this within Aristotle's scheme : The material factor is a thread of tensile strenghh 1 lb; privation is the state of being unstretched; the efficient cause is a weight os 2 lb; the formal state is the breaking of the thread; and the final cause is a law of elasticity. The final cause, just noted, has been a favoured aunt-sally because it smacks of teleology, whereas the finality comes from the final cause being a law (of nature, of society). In terms more akin to those of Aristotle they are the regular and characteristic consequences or results that accompany the various efficient causes, (in the context of the other causes and principles). I think that Aristotle's scheme is equally as good as Popper's, if not better, for it also directs us to look for the essence of the thread, 'what is the thread ?' It is through inductive, free-thinking exercises such as this that theories of matter were developed in the pre-crystallographic days - why did things stretch so far, but no further? We should remember that the thread stretches before it breaks, and that, if the weight is removed, the thread need not break. Popper's analysis seems to suggest instantinaity or an immutable progress, whereas Aristotle's analysis emphasizes that the change from privation to the formal state is dependent not only on the efficient cause, but also on the material factor (eg we don't burn the thread).

Within Marxian exegesis there can be discerned two strands which are relevant here. There are those traditionalists who believe, in exactly the way Popper is concerned to destroy, in Destiny - The revolution will come, the capitalist collapse is at hand. There are those who follow what is more akin to a systems/cybernetic model, '... [There] is the fatuous notion of the ideologists that because we [Engels and Marx] deny an independent historical development to the various ideological spheres which play a part in history we also deny them any effect on history. The basis of this is the common undialectical conception of cause and effect as rigidly opposite poles, the total disregarding of interaction. The gentlemen often almost deliberately forget that once an historic element has been brought into the world by other, ultimately economic causes, it reacts, can react on its environment and even on the causes that have given rise to it ...'(Engels,1893).

3.Popper's Unity of Method

This section will,given the previous groundwork, argue against Popper's unity of method, and I will use his discussion in 'Poverty of Historicism' as my starting-point. But first I will examine the Essentialism vs Nominalism debate as conducted by Popper(1961:26-34), where after a similar series of confusions about 'essence' and 'Form' he ostensively defines : '... Methodological essentialists are inclined to formulate scientific questions in such terms as "what is matter ?" ... and they believe that a penetrating answer to such questions, revealing the real or essential meaning of these terms ... is at least a necessary pre-requisite of scientific research ... Methodological nominalists, ... , would put their problems in such terms as "how does this piece of matter behave ?" ... For methodological nominalists hold that the task of science is only to describe how things behave, and suggest that this is to be done by freely introducing new terms wherever necessary ... '(Popper,1961:29). Physics does not inquire, he says, into the essence of atoms : he does not also say that the notion of atom originated through philosophers contemplating upon the essence of matter : The whole panoply of Bohr's ill-fated Atomic Theory was a direct result of attempts to answer 'what is an atom ?', given certain aspects of the behaviour of matter under various forms of excitation.

If methodological nominalism was indeed followed by the successful sciences (I do not believe it has), great awareness would have to be shown, because theoretical aridity could easily arise. Apart from my personal conviction that this would be so, I find support in Gödel's Proof (Gödel,1962) and Craig's Theorem (Craig,1953). Gödel showed that in any closed system the consistency or otherwise of the system could not be proved <u>within</u> the system, and his comments were directed in particular against the highly definitional (nominalist) <u>Principia Mathematica</u> of Russell and Whitehead. I interpret this to mean in this context that it is impossible within nominalism to find whether sets of descriptions are <u>consistent</u> (it is always possible to define an atom in one context as different from an atom in a different context); corroboration is not a check on consistency. Craig's point is different, but at the same time highly associated. Craig showed that any prediction made by use of a theory, could be made just as accurately using solely observable quantities. For example; we do not need to know about the molecular theory of gases to use the Universal Gas Law; and we do not need any theory to perform a Path Analysis. (However; we need the molecular theory of gases to derive Van der Waals equation; and interpreting the path analysis we might be tempted to say that a person's number of siblings 'causes' occupational prestige). Craig's result shows one of the pitfalls of nominalism, the possible neglect of theory (though, as I said, I do not believe science is nominalist).

The classical Aristotlean distinction between induction and deduction is relevant here. Popper says that nominalism is concerned only with deduction (and test of the deduction) and that essentialism is concerned only with induction - Aristotle says that science (like life) is both inductive and deductive, and one informs the other. This brings us to the Unity of

Method, which is '... the methods in the two fields [of natural and social science] are fundamentally the same ... The methods always consist in offering deductive causal explanations, and in testing them ... the method of testing hypotheses is always the same ...'(Popper,1961:131-132). Consider the implications : the 'method' is concerned only with the aspect I have called 'deduction', (Popper conspicuously disavows any interest in how a theory is developed); and if the only criterion is testing, what of a theory that is untestable now, but may be in the future ? (I refer to Popper's 'Refutation of Historicism' - see the Addendum). I can agree with Popper that the testing of theories, if it at all possible, is highly desirable; I cannot agree that the testing of a predicted value of the frequency of the next peak of high intensity radiation (using an oscilloscope), is of the same type as the testing of a predicted difference in the average levels of alienation in two groups (by taking samples from within these groups). Measurement error is <u>not</u> the same as sampling error.

4. The Essential Popper

I find it difficult to take Popper seriously, but one has to, for he has been so influential. Popper dismisses essentialism and offers nominalism, but, because he cannot do otherwise, he makes many essentialist points. The question is the essentialist one 'What is an army ?', and Popper does not explicitly ask it, but : '... Most of the objects of social science, if not all of them, are abstract objects; they are theoretical constructions. (Even 'the war' or 'the army' are abstract concepts, strange as it may sound to some. What is concrete is the many who are killed; or the men and women in uniform, etc) ...' (Popper, 1961:135). Concerning this very point see the discussion above about a hierarchy of levels, and Sellars denying the existence of a table.

The advocacy of methodological individualism is itself an answer to the question 'What is the nature of society? '.

ADDENDUM : The Refutation of Historicism.

Popper presents a proof that we cannot predict the future course of human history, which also contains an assumption that human knowledge is synomynous with scientific knowledge. (See the Preface to 'The Poverty of Historicism').

This proof is suspect on one point; that is the identification of human knowledge and scientific knowledge - for obvious reasons.

The proof is invalidated on another point; if we cannot predict the course of the growth of scientific knowledge then we do not know that, because of some new advance, we cannot at some time in the future predict the pattern of growth. (This proposition is time-bound).

The refutation is itself a historicist theory, it cannot be tested and is damned by itself; was it Epimenides who stated that all Cretans are liars?

REFERENCES

Craig W

1953 On axiomatization within a System. _Journal of Symbolic Logic_, 18:30-32.

Engels F

1893 Letter to E Mehring in Berlin, (London, 14 July). Reprinted in _Marx and Engels - Selected Works_ (in one volume), London : Lawrence and Wishart.

Gödel K

1962 _On formally undecidable propositions of 'Principia Mathematica' and related systems._ Edinburgh : Oliver and Boyd.

Popper K R

1961 _The Poverty of Historicism_. London: Routledge and Kegan Paul.

1965 _The Open Society and its Enemies._(Vol I) London : Routledge and Kegan Paul.

1966 *The Open Society and its Enemies.* (Vol II)
London : Routledge and Kegan Paul.

1968 *The Logic of Scientific Discovery.*
London : Hutchinson.

Sellars W

1963 *Science, Perception and Reality.*
London : Routledge and Kegan Paul.

APPENDIX B

SCIENTIFIC REALISM IN SOCIOLOGY

SCIENTIFIC REALISM IN SOCIOLOGY

Scientific Realism is easily confused at a superficial level with (logical) positivism (the philosophical position that only takes into account positive facts and observable phenomena) so I will discuss the nature of scientific realism and some of the ways in which it differs from positivist approaches. In the course of this I will isolate the *empiricist* error and the *logicist* error in scientific thought, and show how they are present in sociology. My thesis will be partly that; many of the mistakes in quantitative analysis in sociology arise through a mistaken view of science.

1. Scientific Realism and the Empiricist Error

My discussion of scientific realism (below), has been strongly influenced by Smart and Graves, and I wish to make clear my intellectual debt to these two. Scientific realism can be expressed in terms of five main theses (Graves,1971:7) which are :

(1) There is an external world independent of anyone's sensory perceptions of it.

(2) This independent external world may contain entities and processes which differ radically from what might seem to be obvious from sense perception.

(3) We can attain a degree of knowledge of this world though never perfect Knowledge. Things in this world are not by their intrinsic nature unknowable.

(4) Science is an attempt by human beings to understand the structure of this external world. It seeks to go above immediate experiences, whilst at the same time explaining them.

(5) Scientists achieve (4) by developing theories which postulate theoretical entities named by theoretical terms - hoping in doingh this to obtain as close a correspondence as possible between their entities and the actual elements of the real world.

To accept (1) is necessary unless one makes reality dependent upon some perceiving subject, leading to a form of idealism in science - eg the solipistic early phenomenology of Blum and McHugh (see their contributions to Douglas,1970). In sociology the main proponents of (1) have been interested in social structure, functional analysis and other aspects of macrosociology ; an interest in social action can be compatible with thesis (1) but has often led to subjectivism and a reification of what the individual 'sees' to what is'there' - nominalism rather than essentialism-plus. Thesis (1) is held by all realisms, where the Marxian-Engelsian materialist position is a clear example - social man is a product of his social and economic environment, but man's knowledge is usually of the epiphenomena and not of the true basis.

The acceptance of (1) but the denial of (2) is probably the usual attitude of the man in the street. Philosophically the rejection of (2) is trivial, as it has little explanatory power; it requires that the perceptions of any two people must be almost the same, for these perceptions must be images of the perfectly obvious true reality. Thesis (2) encourages dialectic (in the sense of argument). Really phenomenologists assert a modification of (2) : for 'external' substitute 'individual's view of the'; for 'sense' substitute 'the sociologist's' : 'The individual's view of the world may contain entities and processes radically different from those which seem to be directly disclosed in the sociologist's perception'. In essence, then, one version of the phenomenologist's main thses is subsumed in thesis (2).

To deny (3) is typical of Kantianism and and versions of logical positivism. Kant argued that there is no way of showing that anything exists except the contingent causal phenomena of 'experience'. However, he says, we seem compelled to think of ourselves as free agents having a 'real' self lying outside the causal contingent scheme - this led to Kant's plea for a 'higher' morality.

Apart from its moral overtones, Kantianism has been very influential in the natural sciences. As an example, there is the very Kantian notion of the principle of verifiabilty (not to be confused, so Popper says, with refutability); verifiability takes the idea that there are <u>nouema</u> (ie, 'things-in-themselves') whose nature is basically unknowable, and this leads to the idea that the meaning of any proposition is to be shown in its method of verification. This was Kant's way of eliminating as meaningless all references to things not directly accessible to observation; metaphysics was therefore, because unprovable, nonsense.

Logical positivism grew out of a desire by physical scientists of the nineteenth century to do away with metaphysics and anything that smacked of the abstract. For example Kirchoff denied that science explains why things happen : he said (Kirchoff,1874) a scientist sees in every 'why' a 'how', and the scientist dicovers new connections between phenomena but does not refer to underlying reasons. This position is close to the anti-positivist position of Popper, for Popper with his doctrine of methodological nominalism holds that science has advanced through asking questions such as '...how does this piece of matter behave ? ...' (Popper,1961:29) - probably he has confused the practices of scientists with the nature of Science, (and not taken a very represetative conceptual sample). It is also interesting to compare this to Wittgenstein's later position in the 'Philosophical Investigations', in which he classifies philosophy as descriptive rather than analyical. He believed that there were no philosophical problems as such, or - if there were any problems - they were merely problems of linguistic usage. (I think this is interesting because Wittgenstein's earlier work - as exemplified in the 'Tractatùs Logico-Philosophicus' - is usually taken as positivist. Also see later concerning Logical Atomism.)

'The Science of Mechanics' (Mach,1893) tried to apply Kirchoff's principles : as an illustration, Mach said atomic theory can be useful to scientists when considered as a mathematical model, but we must not suppose that atoms have a reality of their own, ('... in Nature there is neither cause nor effect; Nature merely progresses ...'). Following in Mach's footsteps Karl Pearson wrote (in 1892) the very influential 'The Grammar of Science', and in the Everyman edition (Pearson,1937) we read :

'... We know ourselves, and we know around us an impenetrable wall of sense impressions. There is no necessity, nay, there is want of logic, in the statement that behind sense-impressions there are "things-in-themselves" producing sense-impressions. About this super-sensuous sphere we may philosophize and dogmatize unprofitably, but we can never know usefully ...'. It is as well to remember Pearson's influence on the development of statistics, and this may be one reason why statistics in practice has often been anti-theoretical or non-theoretical.

Many logical positivists have toned-down their emphasis on a complete observational/conceptual dichotomy, but the neo-positivist doctrine of operationalization (eg Bridgman,1927,1936) has fooled many social scientists into ignoring the conceptual. Whereas originally psychologists tried to measure a real intelligence factor using tests, a common contemporary position is 'IQ is what IQ tests measure' (and end of thought). Craig(1956) showed that we need not use theoretical terms to predict new events, for we can accomplish the same with merely observational data. This is effectively what is suggested in Allan (1974) : by being able to predict (poorly) observable events from observational data, path analysts think they have achieved 'explanation', <u>which involves the use of concepts</u>. Maxwell (1962) puts the case for retaining theoretical terms partly on the impossibility of <u>direct</u> observation (and many commentators would suggest that in addition to physical prostheses such as microscopes, there are conceptual prostheses derived from the scientists' theories). Maxwell (1962:7) writes : '... the point I am making is that there is, in principle, a continuous series beginning with looking through a vacuum and [going down to] ... a high-power microscope,etc., ... The important consequence is that, so far, we are left without any criteria which would enable us to draw a non-arbitrary line between "observation" and "theory" ...'.

Maxwell discusse a semi-hypothetical example about a scientist called Jones. Jones noticed that a certain disease was transmitted by touch and postulated a mechanism based on 'crobes' - minute little animals carrying the disease. Using the notion of crobes Jones built up a theory that gave testable and verifiable consequences. The positivists dismissed the crobes as being part of the metaphysical sphere : then however the microscope was discovered and crobes were observed. At this some philosophers became realists (in our sense) and others natural phenomenologists. These natural phenomenonologists thought that as far as our <u>senses</u> were really concerned a theoretical entity and an observable physical object have the same status, or nearly the same status. A far more radical contention was that we had not observed crobes, all that we had actually observed were shadows or images - so the crobes did not exist as far we are concerned.

However, to return to the theses. If thesis (4) is denied, those who deny it might be metaphysicians who feel that they have their own key to reality, or perhaps phenomenologists such as Husserl. Husserl felt that it is not possible from the standpoint or methods of science to arrive at a <u>pure</u> theory, a theory independent of contingent empirical facts - they usually identify 'science' with 'natural science', a common stance. In sociology phenomenologists do not tend to essentialists (such as Husserl) but rather they tend to be existenstial phenomenologists (similar to Sartre) and this can lead to certain confused thinking. Note the confusion of the Husserlian emphasis on transituational understanding with the Sartrian emphasis on existence, in Douglas'(1970:x) : '... We must always begin by studying ... meaningful social phenomena on their own grounds, but, true to our goal of creating a science of man's existence we must then seek an ever more general, transituational (objective) understanding of everyday life. This is the program of all phenomenological and existential sociologies... '

Thesis (5) is in part a complement of thesis (4) : it emphasizes that the correspondence between science and reality is one of attention to detail in all respects. This includes the <u>formal</u> aspects of the relationships between the theoretical entities, where these formal relationships must try to match relationships in the real world. And this leads on to the next section.

2. Analysis for a Purpose

The empiricist error noted above is distinct from, yet confounded with, the 'logicist error', (see Graves,1971:Chs 2,3). The logicist error is a very pervasive one in quantitative sociology, (and interpretations of science): it is contained in the notion that a theory must have the same structure as a highly developed formal system of uninterpreted calculus. Generally, proponents of this approach in sociology seem to have an imperfect, simplistic idea of what constitutes a theory in the natural sciences, for example, there is Blalock's(1969:2) : '... Ideally, one might hope to achieve a completely closed deductive theoretical system in which there would be a minimal set of propositions taken as axioms, from which all other propositions could be deduced by purely mathematical or logical reasoning ...' (My emphasis).

Why should some sociologists hold to these beliefs ? Perhaps one reason is the influence of the positivists on the philosophy of science in earlier decades, and another, more easily isolated, is the 'success' of mechanics, in particular, in the physical sciences. Coleman (1964:95) justifies the use of mathematics in sociology partly on the basis that : '... the rise of mechanics depended very greatly on the use of ... mathematical formulae to replace verbal statements like those of [Galileo] ... A random page of any modern text in the mechanics of rigid bodies would indicate just how hopeless it would be to express mechanics in less formal or rigorous language ...'. In direct contast to this statement of Coleman, there is Graves' (1971:35) assertion that: '... Classical mechanics is the first full-scale physical theory with which a student comes into contact, and he <u>ought</u> to be impressed with the simplicity, elegance, and power of its structure. It is naturally tempting to seize on it as an ideal to which all scientific theories should try to conform, and I think that this is precisely what many formalists have done. But to claim as a matter of fact all theories do contain such an isolable calculus would certainly be an unwarranted induction if based on classical mechanics alone, and a careful study of other theories would show it to be false ...'.

Perhaps equally relevant to this logicist emphasis is the status of econometrics, and experimental and behavioural pschology. These subjects and their apparent success have made a large impact on Homans (particularly 1967); Coleman(1964) follows his section on 'Mechanics' by 'Mathematics as a Language for Economics' and Blalock uses the econometricians' recursive equation approach in most of his publications. The strain towards axiomatization (cf Blalock above) can be counterproductive as Smart(1963:30) points out : '... very few physical (or other) theories have been at all rigorously formalized ... The rigorous axiomatization of such a theory as quantum mechanics is an almost infinitely distant goal. Furthermore, in a rapidly developing subject, such as modern physics, any axiomatization would become out of date almost overnight, and in any case if it were actually used and were not a museum piece, it would have a fossilizing effect on physical theory...'.

Some quantitative sociologists give the impression of being modern sociological 'logical atomists'. Logical atomism was a philosophy strongly influenced by Russell and the younger Wittgenstein, which began to decline after

1925, (see Urmston (1956) for historical details). Russell had himself been strongly influenced by the philosophy of Leibnitz with its great emphasis on the basic elements of knowledge, and initial development of symbolic logic, (see Midonick,1965: Ch42). When Whitehead and Russell (1910) had demonstrated that the whole of mathematics could be subsumed under an axiomatic system, Russell came to think that, as this axiomatic system was so perfect, the world would have the structure of this logic, '... the structure of the world would ... resemble the structure of "Principia Mathematica". That is the simple argument of the plot ...'(Urmston,1956:7). Wittgenstein's (1922) 'Tractatus Logico-Philosophicus' was an extension of this idea, in some respects, and was highly esteemed by the logical positivists. The positivists took their notion of verifiability from the 'Tractatus ...' ; but, though Wittgenstein did write '... to understand a proposition means to know what is the case, if it is true ...', he held that he had been grossly mis-understood. In fairness to the positivists, if Wittgenstein was understood, it was his own fault. In the 'Tractatus ...' he starts with '... The world is everything that is the case. The world is the totality of facts, not of things ...'; and later '.. Even if ... every fact consists of an infinite number of atomic facts and every atomic fact is composed of an infinite number of objects, even then there must be objects and atomic facts ...'.

From this very Leibnitzian attitude, Wittgenstein changed to a different posture - philosophy was no longer analytical, it explained nothing for it merely described. This, the 'later' Wittgenstein, had an impact on the phenomenological schools in sociology (particularly his emphasis on language), and in conjunction with the influence of J L Austin there has arisen the spectre of the 'lay sociologist'. In the present ideological climate this chimera has an egalitarian appeal - the lay sociologist is the ordinary person who, because he has the undisputed right to have ideas about the social, is on a par with the professional sociologist (or so some professional sociologists would have us believe). Following on from this reification of the lay sociologist there appears a niche for the professional sociologist -

the professional sociologist is now a lexicographer. Austin saw that the ordinary person could be wrong (of course the eternal relativist - Winch for example - can always maintain that we are all correct) and he saw that the philosopher (ie professional sociologist) would always have to go <u>beyond</u> mere individual usages, (eg Austin,1957).

Wittgenstein's 'descriptive' philosophy is closely akin to the 'positive' philosophy of Kirchoff, Madh and Karl Pearson mention ed above : in fact, much phenomenology and linguistic analysis has a great deal in common with positivism in this respect. In a sense this is hardly surprizing, for both are not realist and also Wittgenstein has been a leading light in both schools. Much has been made of the difference between the younger and older Wittgensteins , as if there were an epistemological break or a 'revolution of concepts'. Many do not agree that there was a clean break, and it is relevant to the realist view-point to examine Gellner's(1959:143) : '... [One mistake common to both the "Tractatus ..." and the "Philosophical Investigations" is] the supposition that there is such a thing as "seeing the world right", an absolute insight without intermediary, so to speak; ...',(compare this to the first three theses of section 1).

Truly atomistical views are rare in sociology and might be of the form : 'If something cannot be expressed in a simple mathematical manner, if it cannot be measured, then it is of no value'. It is possible to read, however, sentiments such as : '... The kinds of verbal theories that now predominate in the social sciences *seem much too complex* to allow for mathematical formulation ...' (Blalock,1969:27,my emphasis); and indeed '... In sociology ... the kinds of verbal theories and research results which have been set forth are so vaguely stated or so weak that it is difficult to translate them into mathematical language and once translated *they often fail (sic) to show an isomorphism with powerful parts of mathematics* ...'(Coleman,1964:3, my emphasis). Rather than impugne mathematics , as a tool for its limitations at present, both the above atomistically use this to question the validity of social theories for their lack of correspondence with 'powerful' (?) parts of mathematics. It is interesting to consider that many parts of mathematics were developed because there were no conventional methods to deal with the problems posed by certain scientific theories - a trivial example is the development of Bose-Einstein and Fermi-Dirac statistics for energy levels of atoms, and occupancy by electrons.

Scientific realism implies intentionality in analysis, (intentionality = for a purpose); and this means that, though scientific realism encompasses <u>all</u> sciences, the results of, and methods resulting from, that application of realism will differ according to the nature of the reality studied by that science.

REFERENCES

Allan G J B

1974 Simplicity in Path Analysis. <u>Sociology</u>, 8 : 197-212.

Austin J L

1957 A Plea for Excuses. <u>Proceedings of the Aristotlean Society</u>,

Blalock H M

1969 <u>Theory Construction</u>. Prentice-Hall, Englewood Cliffs.

Bridgman P W

1927 <u>The Logic of Modern Physics</u>. Macmillan, New York.

1936 <u>The Nature of Physical Theory</u>. Dover Books, New York.

Coleman J S

1964 <u>Introduction to Mathematical Sociology</u>. Free Press, Glencoe.

Craig W

1956 Replacement of Auxiliary Expressions. <u>Philosophical Review</u>, 65 : 38-55.

Douglas J D

1970 *Understanding Everyday Life.* Routledge and Kegan Paul, London.

Gellner E

1959 *Words and Things.* Pelican Books, Harmondsworth.

Graves J C

1971 *The Conceptual Foundations of Modern Relativity Theory.* M I T Press, London.

Homans G C

1967 *The Nature of Social Science.*

Kirchoff G R

1874 *Principles of Mechanics.*

Mach E

1893 *The Science of Mechanics.*

Maxwell G

1962 The Ontological Status of Theoretical Entities. (*Minnesota Studies in the Philosophy of Science*, Vol 3 : 3-27)

Midonick H

1965 *The Treasury of Mathematics : 2*. Pelican Books, Harmondsworth.

Pearson K

1937 *The Grammar of Science*. Everyman, London.

Popper K R

1961 *The Poverty of Historicism*. Routledge and Kegan Paul, London.

Urmston J O

1956 *Philosophical Analysis : Its development between the wars*. Oxford University Press.

Whitehead A N and B Russell

1910 *Principia Mathematica.* Cambridge University Press. (Actually, 1st ed 1910-1913; 2nd ed 1925-1927)

Wittgenstein L

1922 *Tractatus Logico-Philosophicus.* Routledge and Kegan Paul, London. (3rd imp 1947).

B.22

1953 <u>Philosophical Investigations</u>.

Blackwell, Oxford.

APPENDIX C

PUBLISHED WORKS - G J Boris Allan

The following list is composed of publications whose origin was the ideas contained in this thesis. They are rather more empirically-based than the content of this thesis, and are best seen in tandem.

1972

'A comment on "'Mobility and work satisfaction' : A comment" '. BSA Maths and Computing Applications Group Newsletter.

"Laws and Causal Modelling - Some implications of a scientific approach". Essays in Statistical Sociology,1972.

1973

"Some implications of the use of an American Life Satisfaction Rating Scale on British Subjects". Proceedings of the 11th International Congress of Gerontology, Kiev, USSR (with Dr A Bigot).

"Linear models of processes: An attempt at synthesis". Sociological Review Monograph, 19.

"The effects of differential fertility on sampling in studies of intergenerational social mobility". Sociology, (with Bill Bytheway).

"Path Analysis - A cautionary note". _Sociological Review_, (with Roy Mapes).

1974

"Simplicity in Path Analysis". _Sociology_.

1975

"Up the path analysis". _Sociology_.

1976 (forthcoming)

"Ordinal-scaled variables and multivariate analysis". _American Journal of Sociology_.

All joint works are truly joint.

APPENDIX D

ESTIMATING A PROPORTION

ESTIMATING A PROPORTION

1. Notation

A population can split into exclusive sets G and $\bar{G}$ with the population being the set U. The number of individuals in a set is shown by n(U), n(G) and n($\bar{G}$). We characterize the population by a parameter $p = \frac{n(G)}{n(U)}$ and a further parameter $q = \frac{n(\bar{G})}{n(U)} = 1 - p$. A sub-set of U is observed (the sample S), $S \subset U$. The number of individuals in the intersection of S and G is $R = n(S \wedge G)$, (ie the number of individuals in the sample who are members of G). $N = n(S)$ is the number of individuals in the sample S.

2. Problem

Is there a 'best' estimate of p and q on the basis of the information in the sample.

3. The sampling distribution of R

From the binomial theorem, given a population probability/proportion p and a sample size N, the proportion of cases (in the long run) for which we will observe R individuals from set G is pr(R:N,p) where

$$pr(R:N,p) = \frac{N!}{R!\,(N-R)!}\,\underline{p^R(1-p)}^{N-R} \qquad (3.1)$$

(N! is factorial N). This is interpreted to mean that the probability of observing R members of G, given a sample of size N and a population parameter p, is given by the RHS of (3.1).

4. Unbiassed estimates of p

The unbiassed estimate of p is p_u

$$p_u = R/N \qquad (4.1)$$

but in cases of small N it is always possible for R to be equal to zero, when from other considerations the parameter is not zero, (eg equation (3.1) and theoretical knowledge). This is shown when an attempt is made to estimate a function of p, in particular the estimation of the sampling variance (p(1-p) with p_u substituted is zero when $p_u = 0$) and the logit transform θ, where

$$\theta = \log\left(\frac{p}{q}\right)$$

If $p_u = 0$ or $p_u = 1$, the resulting logit transform is plus or minus infinity; for this reason, and considerations of unbiassedness, we estimate the 'empirical' logit θ_u by

$$\theta_u = \log\left(\frac{R+\frac{1}{2}}{N-R+\frac{1}{2}}\right)$$

It requires little arithmetic to show that the estimate of p, derived from the unbiassed logit is p_{1u}

$$p_{1u} = (R+\tfrac{1}{2})/(N+1) \qquad (4.2)$$

This also the estimate which can be found from considerations of 'continuity' corrections.

5. Maximum likelihood estimation of p

The maximum likelihood converts (3.1) into a likelihood

$$L(p^*:N,R) = K\, p^{*R}(1 - p^*)^{N-R} \qquad (5.1)$$

where p^* is now a variable and R fixed. Maximum likelihood methods find that value of p^* which maximizes (5.1), and this value is p_m

$$p_m = R/N \qquad (5.2)$$

p_m is suspect in the same way as p_u.

6. Bayesian estimation of p

The Bayesian methods convert the likelihood of (5.1) into a probability density of p^*,

$$f(p^*) = K()\, p^{*R}(1 - p^*)^{N-R} \qquad (6.1)$$

(I am using a uniform prior distribution, the estimate

p_{lu} can result if a different prior is used - this was shown by Fisher). The constant K() is chosen to make the total density unity. (6.1) is the equation of a beta distribution and the mean of this distribution of p* is

$$p_b = (R+1)/(N+2) \qquad (6.2)$$

(The mode of this distribution gives p_m). p_b has the advantage of not leading to estimates of p_b = 0 or 1, but has the disadvantage that $(p_b)^2$ is not the same as $(p^2)_b$ (the Bayesian estimate of p^2).

7. Middle likelihood or Median Distributional estimates of p

The median of a distribution under a functional one-to-one transformation is exactly transformed. So, therefore, I would suggest that the median of the distribution of (6.1) - or the value of p* in (5.1) which splits the area under the likelihood curve in half - the median is a good measure to use. The median can be found analytically but for most purposes I have found that it can be closely approximated by

$$p_Q = (R+.7)/(N+1.4) \qquad (7.1)$$

This gives estimates which are very close to those of p_{1u}, but is closer to the analytical median.

8. Example

Take the extreme example $R=0$ and $N=5$.

$p_u = 0$
$p_{1u} = .08\dot{3}$
$p_m = 0$
$p_b = .143$
$p_Q = .109$

The median of the beta distribution is .109 (by analytical methods).

9. Final comment

If we have two samples of size N, the only estimates, on the two samples combined which equal the average of the two separate sample estimates, are p_u and p_m. For the other three estimates we find the effects of the coarseness of our observed data, which I think is an admirable virtue.

APPENDIX E

VARIOUS UNPUBLISHED PAPERS - G J Boris Allan

The following papers are concerned with various aspects, one extends my papers on social mobility and path analysis, and two are concerned with what enters the analysis - they play great emphasis, as I have throughout this thesis, on the importance of specifying underlying distributions. All are heavily influenced by my scientific realist philosophy.

They are concerned with practical matters, where this thesis has attempted to argue a cogent theme at a more abstract level - J L Synge in "Talking about Relativity" p 13
'... As a matter of fact, there is always some fuzziness about a concept, and one of the main differences between ordinary life and science is that scientific concepts are less fuzzy. But they are fuzzy nevertheless. It is only in mathematics that we find clear-cut concepts, and it is probably this inhuman characteristic that makes the subject repellent to many people...'

THE BIVARIATE NORMAL CORRELATION OF A CONTINGENCY TABLE

G J Boris Allan

School of Sociology

Manchester Polytechnic

Aytoun Street

Manchester M1 3EH

SUMMARY

For a 2 x 2 table, when the values of ϕ and Yule's $\underline{Y}$ are averaged (to give an estimate of the 'true' value Of Kendall's Tau) and this mean value is transformed by a sine-transformation, the transformed value is a very good estimate of the bivariate normal correlation of the table. For an $\underline{R}$ x $\underline{C}$ contingency table, when the (R - 1)(C - 1) different estimates are averaged the accuracy of the estimate resulting is ~~[illegible]~~. high.

KEYWORDS

BIVARIATE NORMAL CORRELATION

CONTINGENCY TABLES

KENDALL'S TAU

YULE'S Y

PHI

SOMER'S D

ACKNOWLEDGEMENTS

I would like to thank Ian McDowell (Medical School, Nottingham) and William B Tyler (University of Kent at Canterbury) for their encouragement to solve what they both thought was an important practical issue. Part of the work reported herein was supported by the Health Services and Mental Health Administration of the US Govt, under contract HSM 110-70-393 at the Department of Community Health, Medical School, Nottingham).

THE BIVARIATE NORMAL CORRELATION OF A CONTINGENCY TABLE

INTRODUCTION

Often when we examine a two-by-two (2x2) table we can conceive of a bivariate normal distribution upon which we have placed a double dichotomy. That is, we suppose that we have an underlying bivariate normal distribution (eg of X and Z) and that we are only able to distinguish between values of X as greater than or less than a certain (usually unknown) value - and similarly for Z. For specific values of the bivariate normal correlation Karl Pearson (Tables for Statisticians and Biometricians , Vol II - long out of print) provided tables which show, for specific marginal proportions, the proportions in the cells of the table, (actual use involves ~~using~~ interpolation for values between the specified values). This is Pearson's method of tetrachoric correlation, where the proportions in the cells are obtained by expanding the tetrachoric series, (Kendall and Stuart, 1969:160-161; 1973:316-319). However, for two medial splits the tetrachoric series expansion is identical to a trigonometrical expansion; that is, the psychometrician's 'cosine-pi' formula, which estimates the bivariate normal correlation by :

= square root
= circumflex
= Greek pie

$$\hat{r} = \cos\left(\pi \frac{\sqrt{ad}}{\sqrt{ad} + \sqrt{bc}}\right)$$

where a b c and d have their conventional meaning for a 2x2 table (there is a very clear discussion in Guilford and Fruchter(1974:300-306); the psychometric issue can be followed by examining their references).

It is a simple matter to show that this is equivalent to :

$$\hat{r} = \sin\left(\frac{\pi}{2} \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}\right)$$

so that it is immediately obvious that :

$$\hat{r}_Y = \sin\left(\frac{\pi}{2} Y\right)$$

where Y is Yule's coefficient of colligation.

This has a clear family resemblance to the estimate of the bivariate normal correlation based on Kendall's tau, $\underline{T}$:

$$\hat{r}_T = \sin\left(\frac{\pi}{2}T\right)$$

(Kendall,1970:Chs 9,10). In the case of a 2x2 table tau(b) is equal to ϕ (phi-the product-moment correlation for binary data , Kendall(1970:43-45)). It takes very little arithmetic to show, for two medial splits, that $\underline{Y} = \phi$: this reinforces the family resemblance.

THE ESTIMATES FOR 2x2 TABLES

The estimation using $\underline{Y}$ (for non-medial splits) over-estimates ; the 'true' correlation is a lower bound. This is because $\underline{Y}$ is purely a function of the cross-product ratio - and is unaffected by arbitrary multiplications of the numbers of row or column elements, (Yule,1912;Edwards,1963). This, in a sense, leads to the the maximization of the estimate of the association being achieved no matter what the marginal proportions. The estimation using ϕ (for non-medial splits) underestimates ; the true correlation is an upper bound. This is probably because ϕ is a correlation derived from a linear regression with heteroscedastic residuals, so that the estimate of the correlation is not as high as it should be (eg see Goldberger,1972:249-250; Johnston,1972:176-186, 214-221). However, if the estimate of the bivariate normal correlation via the sine-transformation, using $\underline{\emptyset}$ for $\underline{T}$, tends to underestimate, the underestimation is much less than using $\underline{\emptyset}$ as such,(it is as well to remember that $\underline{\emptyset}$ is not only the product-moment correlation of a 2x2 table, but it is also the canonoical correlation for that table). If we have available published values of $\underline{\emptyset}$ (eg Blalock,1964:72-77) we can reasonably transform these values, if the bivariate normal hypothesis fits. (Transformation tables are available on request).

For any 2x2 table we can calculate two correlations, $\hat{r}_Y$ and $\hat{r}_\emptyset$, which are upper and lower bounds to the value of the bivariate normal correlation. Frequently it is not sufficient to provide upper and lower bounds for a correlation, for we need a point estimate. The problem is whether (and how) to average $\underline{Y}$ and $\underline{\emptyset}$ into one value - which we might call an estimate of the 'true' value of tau for the table - or whether to transform $\underline{Y}$ and $\underline{\emptyset}$ and then average this pair of values.

In most cases the approach by transforming the true value of tau agrees closely with the value obtained by use of the second approach. As the first approach gives the true value of tau as a direct by-product, I recommend that approach.

EXTENSION TO AN RxC TABLE

It goes without saying that if we have an RxC table we lose an amount of information by collapsing it into a 2x2 table. No matter how accurate our estimate of the bivariate normal correlation of a 2x2 table, we know very well that differing ways of collapsing tables will give rise to different estimates, perhaps not differing greatly from each other but still differing (eg Lancaster and Hamdan,1964:337). To collapse in one way only is frivolous ~~of~~ with information, really we need to find some way of averaging all these differing values (for an RxC table there are (R-1)(C-1) different values to average).

Some values will be intrinsically of greater accuracy than others : at one end of the continuum is the case of two medial splits and at the other (useless) end there is the insoluble case where one cell is empty, (I say insoluble because it could be empty through not having a large enough sample, or through a form af exact relationship). As the standard errors are unknown for this measure, I posit that, if each estimate of true tau is weighted by the size of the smallest cell number, this is reasonable, for the total sample will remain identical for each, and 'true' taus estimated in the case of zero cells will not contribute to the final average true tau (which is then transformed).

EXAMPLES OF ESTIMATES FOR 2x2 TABLES

Karl Pearson (Tables for Statisticians and Biometricians, Vol II) gives some examples for the calculation of the tetrachoric correlation, and it is these examples I will use in this section.(Because of the difficulty in obtaining these tables, I will supply those interested with copies).The results of applying my true tau method, are compared with the values given by Pearson's tetrachoric method. As can be seen the agreement between the true tau (transformed) estimate and the tetrachoric estimate is very high, even in the case noted where the marginal proportions were very 'strange' - but the robustness of the method is shown more clearly in the next section, (see Table 1).

EXAMPLE OF AN ESTIMATE FOR AN RxC TABLE

The RxC table considered here is a 7x7 table which displays a bivariate normal surface with r = .5. Numbers in the cells are rounded to the nearest whole number so, though almost a bivariate normal surface, it is not quite bivariate normal. This table originally appeared in Pearson and Heron(1913:200), and was one of the examples used by Richie-Scott(1918:125). This table is reproduced here as Table 2, and I would ask the reader to note that six of the cells are empty. In Table 3 I show the estimates of true tau for each 2x2 (collapsed) table, and if it is remembered the value of tau corresponding to r= .5 is .333, then examination of this table in conjunction with Table 2 will reveal a remarkable stability of estimate.

***********TABLES 1,2,3 ABOUT HERE!!!!!

The range in value of estimate of true tau is .311 to .344; or, in terms of correlations, .4[illegible] 71 to .514 -

this, I submit, is as high an accuracy as could be reasonably expected. (Richie-Scott has tetrachoric estimates which range from .498 to .510, but he does not consider all possible 2x2 tables - in fact none of his marginal distributions are really 'strange'. In the cases of the two extreme estimates in Table 3: the marginal distributions; and the numbers in the cells are 'strange'. Richie-Scott (1919;125-129) suggests a method called enneachoric r which varies, in estimate of correlation for the entire table, from .467 to .517) My estimate of the correlation of this table - weighted by the size of the number in the smallest cell - is .497. (Well within the rounding error of the cell-numbers).

THE CASE OF ONE FIXED MARGIN

In some cases, such as the relationship between vaccination and recovery from smallpox, the distribution of the values of one of the variables is not random, ie it is not fixed. For example, the proportion vaccinated depends upon vagaries of chance or choice : however, the conditional recovery dependent upon vaccination is a part of the process, for the number who recover is dependent upon the number who are vaccinated.(This is the crux of the argument between Yule and Pearsón.) In such cases I would suggest that the $\underline{\phi}$ - based <u>conditional</u> measure is used in place of ϕ, ie Somer's (asymmetrical) d-coefficient.

TABLE 1 :

ESTIMATES OF THE BIVARIATE NORMAL CORRELATION

FROM Ø	FROM Y	FROM 'TRUE' TAU	TETRACHORIC
.195	.210	.203	.204
.386	.427	.407	.410
.951	.959	.955	.956
.035	.040	.038	.038
.052	.056	.054	.054
.215	.246	.231	.232
.311	.411	.361	.365*
.430	.472	.451	.452
.457	.465	.461	.461
.524	.548	.536	.537
.789	.810	.800	.800
.929	.957	.943	.946

*Marginal splits for this example are .7/.3 and .15/.85.

TABLE 2 :

BIVARIATE NORMAL SURFACE WITH r = .5

	1	2	3	4	5	6	7	TOTAL
1	7	20	5	2	0	0	0	35
2	21	145	79	36	10	9	1	301
3	6	94	85	54	19	22	4	284
4	2	32	39	31	12	17	4	137
5	0	18	28	25	11	18	5	105
6	0	11	22	24	12	22	7	98
7	0	2	6	8	5	13	7	41
TOT	36	322	264	180	69	101	28	1000

TABLE 3 : [illegible] OF

ESTIMATES OF TRUE TAU WITH POINTS OF DICHOTOMY AT H AND K

K \ H	1,2	2,3	3,4	4,5	5,6	6,7
1,2	.329	.317	.323	*	*	*
2,3	.323	.331	.329	.329	.332	.344
3,4	.333	.330	.332	.330	.330	.314
4,5	*	.333	.332	.334	.336	.312
5,6	*	.329	.326	.330	.331	.331
6,7	*	.339	.321	.334	.342	.341

REFERENCES

BLALOCK HM

1964 *Causal Inferences in Non-experimental Research*

U of N Carolina Press : Chapel Hill.

EDWARDS AWF

1963 The measure of association in a 2 x 2 table.

J R Statist. Soc., A, 126 :109-

GOLDBERGER AS

1964 *Econometric Theory.* Wiley : New York.

GUILFORD JP and B FRUCHTER

1974 *Fundamental Statistics in Psychology and Education.* McGraw-Hill : New York. (5th ed)

JOHNSTON J

1972 *Econometric Methods.* McGraw-Hill : New York (2nd ed).

KENDALL MG

1970 *Rank Correlation Methods.* Griffin : London (4th ed).

KENDALL MG and A STUART

1969 *Advanced Theory of Statistics. Vol I : Distribution Theory.* Griffin : London (3rd ed).

1973 *Advanced Theory of Statistics. Vol II : Inference and Relationship.* Griffin : London (3rd ed).

+ Insert here
* Insert here

YULE GU

1912 On the methods of measuring association between two attributes. *J. R. Statist. Soc.*, 75 : 579-

* PEARSON K and D HERON

1913 On theories of association. *Biometrika*, 9 : 159-315.

* RICHIE-SCOTT A

1918 The correlation coefficient of a polychoric table. *Biometrika*, 12 : 93-133.

+ LANCASTER HO and MA HAMDAN

1964 Estimation of the correlation coefficient in contingency tables with possibly nonmetrical characters. *Psychometrika*, 29 : 383-391.

PATHS OR DIVERSIONS ?

Up the path analysis (1)

G J Boris Allan

School of Sociology
Manchester Polytechnic

This is a revidion of a paper presented to the BSA Maths Computing and Statistics Group at the University of Surrey (April 1974). The paper was entitled simply : "Up the path analysis".

PATHS OR DIVERSIONS ? - UP THE PATH ANALYSIS (1)

In "Simplicity in Path Analysis" (SPA - Allan,1974a) I present a method for the analysis of sets of observable variables, where these observable variables co-vary and are supposed to represent in some way the concepts of an underlying causal process. In one sense the method is akin to path analysis for it uses as its basis a correlation matrix and is a method which assists in the testing of causal inferences (that often path analysis is used to produce causal inferences is a salient point). In most other senses the methods differ crucially, some of which will be examined in more detail in this papaer. One, there is the attitude towards "error" in prediction; two, there are the means for establishing general statements; and, three, the treatment of concepts by their causal orderings. The mode of analysis presented in SPA, (and I am not concerned with the factor analysis based method), did not arise in a vacuum but resulted from my own efforts in analysis to reconcile a conviction that a quantitative approach in sociology is necessary to its advance, with a belief in the power of a theoretical approach. The need for such a reconciliation seemed strange to one who was once a physicist, but it was obvious that the quantitative procedures available were largely anti-theoretical, (see some of the examples in SPA).

I believe that the main example used in SPA was not a particularly good one, and it was only used because it had been used previously by Hope (1970), and part of SPA consisted of a critique of Hope. I will now examine two other examples : these examples are best considered in conjunction with the example in SPA, and are selected from two American Sociological Journals [1].

1. Two American Social Mobility Examples

I decided that additional examples were needed to illustrate the workings of my method. These had to be from published material so that the various types of analysis could be compared, so I decided on a year (1972) and two journals (the American Journal of Sociology and the American Sociological Review). I first 'randomly' selected an article in ASR which used path analysis [2], and came up with : "Achievement Orientations and Socioeconomic Career Attainments" by David L Featherman,(Featherman,1972). This happened to be a very good example, for there was a clear causal ordering in most of the variables,(as I noted in SPA, I do not think that the prstige accorded a job is nessarily a 'cause' of the remuneration of the job-holder, but in this case I will accept the conventional ordering). I decided to abstract a subset from twelve variables used by Featherman (1972,Figure 1) in such a manner that no variables outside the subset were supposed to be "caused" of the variables within the subset. (See my Figure 1).

Figure 1 about here

For the second paper I took AJS Volume 77 - and, as if to prove the randomness of my method of sampling, I selected another Featherman paper(1971). I decided not to reject this second paper because I felt that it was reasonable that I find two papers in a specialist area by the the same author. The title of the second paper was : "A research note : A social structural model for the socioeconomic career", and I chose to examine the correlation matrix of Featherman's Table 1, and the path analysis of Table 2. The path analysis of Featherman's Table 2 is diagrammatized, (accurately I hope) as my Figure 2.

Figure 2 about here

1.1 The ASR paper

The causal (and temporal) ordering of the concepts implied in Featherman's first model (my Figure 1) is, I suggest, this: the type of home social background of an individual (indicated by the NORC score for the father's job) is prior to the individual's type of educational experience (indicated by his number of years of schooling in 1957), which is itself prior to how "good" a job the individual first enters (indicated by the NORC score for his job in 1957), these all being prior to how good a job he has at two later periods (indicated by NORC scores for his jobs in 1960, and some time between 1963-1967).

In path analysis, exactly as used by Featherman, this causal ordering only affects which variables are used as predictor variables in the path regressions. The usual result is evident in Featherman's analysis, trivialization - the only path that is ommitted from Figure 1 is that between observables 6 and 1, which implies that an individual's type of home background does not affect how good a job he had in 1960, but it does affect how good a job he has at other times. In my method this causal ordering is very important, for although there are no strict rules (social science data is far too intricate for the application of rote), in the analysis of the determinants of a certain observable I introduce the observable immediately prior into a regression first. I then introduce the next most immediately prior and see what improvement in prediction has occured (and examine the collinearity effect). In the case of the data portrayed in figure 1, the resulting analysis is very simple and is shown in Figure 3.

Figure 3 about here

This figure shows a simple causal chain ~~chain~~ model. That is, I felt (without any prosthetic statistical inference) that I could account for variations in 8 as easily with 6 alone as with 6,4,3 and 1. For observable 8 the residual path (from path analysis) is .50, whereas using 6 alone it is .51 - I am willing to increase my "error" at the expense of what I see as probable trivialization.

1.2 The AJS paper

The process which is being copied in this paper (Featherman,1971) is obviously similar to that of the previous paper (Featherman,1972), but now the "goodness" of a job is indicated by two observables - its NORC score and the income from the job. I have noted that I am not in agreement with the causal ordering of prestige (status?) and income (worth?) - probably they are highly related dimensions of the goodness of a job ~~gxh~~ according to people's interpretation of the social ethos : however, for this example I will accept Featherman's (conventional) causal ordering.

Reference to Featherman's results (my figure 2) illustrates yet again the fear of 'losing' variance explained : the inclusion of observables merely to boost variance explained by a trifle. Out of a possible (10x9)/2 = 45 paths, only 19 paths are eliminated - on the basis of a test of statistical significance (Featherman,1971:Table 2, 301). In fact Featherman seems lothe to eliminate any path, for his Figure 2 (1971:298) includes all paths even though some are as low as .091 or .099.

Figure 4 about here

Iwould now ask you to examine my analysis, (shown in my Figure 4) where it is especially clear that - if it copies reality accurately - Kelley's (1973a) two causal chain hypothesis is supported. Without wishing to enter into the debate between Featherman(1973) and Kelley (1973b), (as this is not my reason for this paper), I think that Kelley is right in postulating two causal chains, but he wrong to use path analysis to establish his claims (and perhaps the two causal chains are but realizations of one real chain). However Kelley (1973b:791-792) makes the following comment :

> But when corrected for attenuation, the Six-Cities data offer absolutely no evidence for an histórical effect [of occupation at time 1 on the occupation at time 3] . The path is a miniscule -.005 and the increment in variance explained, .000, can be ignored with some safety. The more parsimonious causal chain model fits the data astonishingly well.

In the present context this is endowed with an added significance.

I have not shown how my figure was arrived at, and I will now do so. Firstly, I tokk the observables in Featherman's causal ordering, apart from the assumption that income and occupation at a certain time were of equal causal priority in terms of their <u>effects</u>, though jointly of different causal priorities ih that prestige was a 'cause' of income for a certain time. Secondly, if there was only one observable immediately prior, it was automatically introduced and further observables were only introduced if they added to the variance explained <u>in the order of their causal priorities</u>: if there was more than one observable immediately prior, this set was analyzed according to the incremental method of SPA - further observables were only introduced if they added to the variance explained, <u>in the order of their causal priorities</u>. (If this method seems rather <u>ad hoc</u>, I do not apologize, for as I noted above application of rote can be very misleading, if not wrong. In fact it is not as <u>ad hoc</u> as it seems, for the analysis is motivated a firm philosophical position based on a scientific realist view of sociology - Allan, 1974b).

Two instances may make this clearer. Take the final observable in the sequence, ie income at time 3 (I_3), when the residual path in the analysis of Featherman is .77 (see Figure 2) using four observables. Now the immediately prior observable is the NORC score at time 3 (Y_3), and the introduction of income at time 2 (I_2) into ~~a~~ regression with Y_3 adds .12 to the variance explained, giving a total variance explained of .38 and a collinearity effect of .15 - the residual path is .79. The variance explained in Featherman's path analysis (<u>using 8 observables</u>) .41, only .03 more than my .38, and I will suggest that this means that there is no <u>direct</u> (historical) effect of the concepts indicated by observables X through Y_2 on the concept indicated by I_3.

In this instance only one observable was immediately prior, and, taking the other observables in the order of their causal priority, when one of the next observables had been introduced no more <u>real</u> improvement could be made.

In the next instance, the dependent variable is observable Y_3, and immediately prior are <u>two</u> observables, Y_2 and I_2. If the incremental method is used, the variance explained in Y_3 by Y_2 alone is .54, whereas the variance explained in Y_3 by observables X through Y_2 is .60 - only .06 more. Again I would suggest that only the concept indicated by Y_2 has a direct effect on the concept indicated by Y_3. Note that the results in Figure 3 are contained in the results of Figure 4.

2. Inference : Theoretical or statistical

Suppose I have measured the values of two variates for a <u>population</u>, and the correlation between these two variates is .01. As the population has been studied the correlation of .01 is the true correlation - but should we ignore it, cell it zero? So arises many of the problems and confusions in quantitative analysis - statistical tests are often used merely to eliminate small associations or differences under the rubric of the null (usually zero) hypothesis. This is often displayed in its most extreme (because vocal) forms by those "learning the trade", and from my own experience, particularly postgraduate psychologists (who have never heard of Kuhnian paradigms). These trainee psychologists are, eg, studying a group of children and are interested in the effects of some treatment on the children ; they admit (when pressed) that really the group is their statistical population, for otherwise the tha population to which they would generalize would be so specific - eg remedial readers aged 8 years in school X - that it would be of no real value. When it is pointed out tà them that a statistical test of difference is wrong - because the differences are there or they are not, and that really they should be interested in how large/important are the differences - they say :(1) "How do we know the difference is significant?". (They mean "important"?);and (2) "My supervisor expects it". Authors frequently do not know whether a path coefficient represents a viable causal connection, so they eliminate on the basis of a statistical test - I have much more sympathy with those who choose to ignore path coefficients of, eg, less than .15.

This is only part of the problem : I am interested in how, given a set of singular, time-bound results from different studies, we can advance our knowledge. This advance in knowledge (I echo Zetterberg and many others) can only come through the use of a theoretical perspective, and one which requires looking at other studies in a ~~newx~~ light - as I have said, (SPA:212):

> It is as well to remember what we are doing when we examine a path analysis : we are trying to find 'reverse -operationalizations', ie what have we really measured ?

where the 'we' is understood to be the sociologist-in-general.[3] "Let this be applied then", is the obvious comment, so first examine Figure 5 (the result of analyses in SPA), and compare this to Figure 6 (the analysis of Figure 4, using ~~the~~ terms/concepts compatible to those of Figure 5). When presented in this manner the "overlap" between the two singular studies can be seen, and if a little (sociological?) imagination is used we can combine the two, to provide a new result, (shown in Figure 7). From two studies using a theoretical infrence we have improved our knowledge - which can be compared with knowledge from other studies or act as the basis of further studies. The general, theoretical approach (which is not solely concerned with causal analysis) can enable us to increase knowledge in general by the examination of the particular.[4]

Figure 5 about here

Figure 6 about here

Figure 7 about here

FOOTNOTES

1. The philosophy underlying SPA and the present paper may be described as "scientific realism". I will not expand upon the philosophical aspects in this paper, for this has been done in Allan (1974b). Other relevant references are Smart(1963) and Graves (1971:Chs 2,3).

2. I write "... randomly..." because a truly random selection implies that a book ends at the beginning.

3. Surprizingly, the phenomenologists McHugh,Raffel, Foss and Blum (~~Blum~~ McHugh *et al*,1974:16) state very clearly :

> The idea of theorizing makes necessary a distinction between the concrete and the analytic. In so far and whenever a theorist fails to formulate a distinction between the concrete and the analytic - between concrete and analytic speech - he loses his ability to account for his own activity : for theorizing. Without a distinction between the concrete and analytic we necessarily formulate theorizing as <u>a reproduction or reporting of what appears</u> (My emphasis).

This is far more scientific than the actions of a path analyst who thinks "numbers of siblings" causes "years of education" - it might be a shorthand, but it might not.

4. In Allan (1974c) I examine the nature of enquiries into (particularly) social mobility and shows that the theoretical stance

in such studies has been notably lacking. The crime is not merely & the confounding of observable and conceptual, but an unawareness of what is being measured.

Allan G J B

1974a Simplicity in path analysis, Sociology, 8(2), May :197-212

1974b Sociology and scientific realism. (Mimeo)

1974c The meta-analysis of social mobility. (Mimeo)

Featherman D L

1971 A research note : A social structural model for the socioeconomic career. American J of Sociology, 77(2):293-304

1972 Achievement orientations and socioeconomic career attainments, American Sociological Review, 37, April:131-143

1973 Comments on models for the socioeconomic career, American Sociological Review, 38; 785-791.

Hope K

1970 Path analysis : Supplementary procedures, Sociology, 5(2), May:225-241.

Kelley J

1973a Causal chain models for the socioeconomic career, American Sociological Review, 38:481-493.

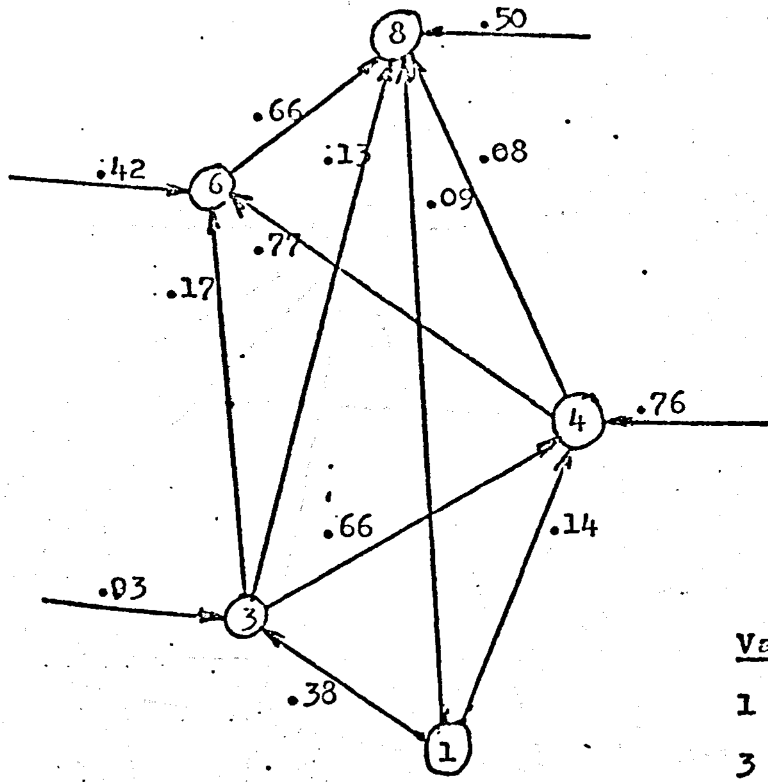
1973b History, causal chains and careers : A reply, American Sociological Review, 38 :791-796.

Graves J C

1971 The conceptual foundations of modern relativity theory. London : MIT Press.

McHugh P;Raffel S;Foss D C ; and Blum A F
1974 *On the beginning of social inquiry*. London : Routledge.

Smart J J C
1963 *Philosophy and scientific realism*. LOndon : Routledge.



Variables

1 - Father's Occ-NORC

3 - Education

4 - Occ-NORC I

6 - Occ-NORC II

8 - Occ-NORC III

FIGURE 1. A Model of the Status Attainment Process

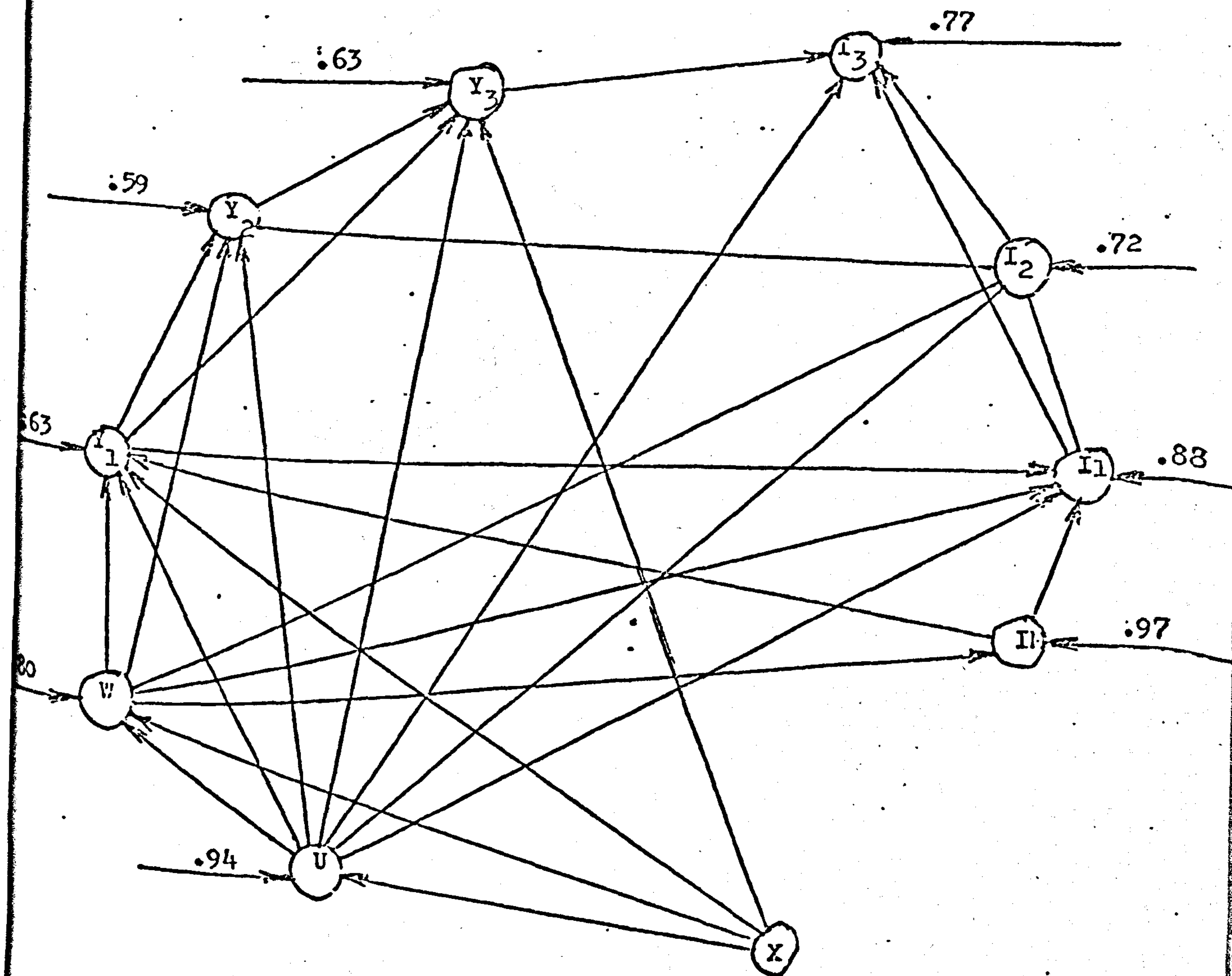


FIGURE 2 : A Social Structural Model for the Socioeconomic Career

<u>Variables</u>

X - Father's Occ-NORC

U - Education

W - Occ-NORC at marriage

IM - Income at marriage

Y_1 Y_2 Y_3 - Occ-NORC at times I II III

I_1 I_2 I_3 - Incomes at times I II III

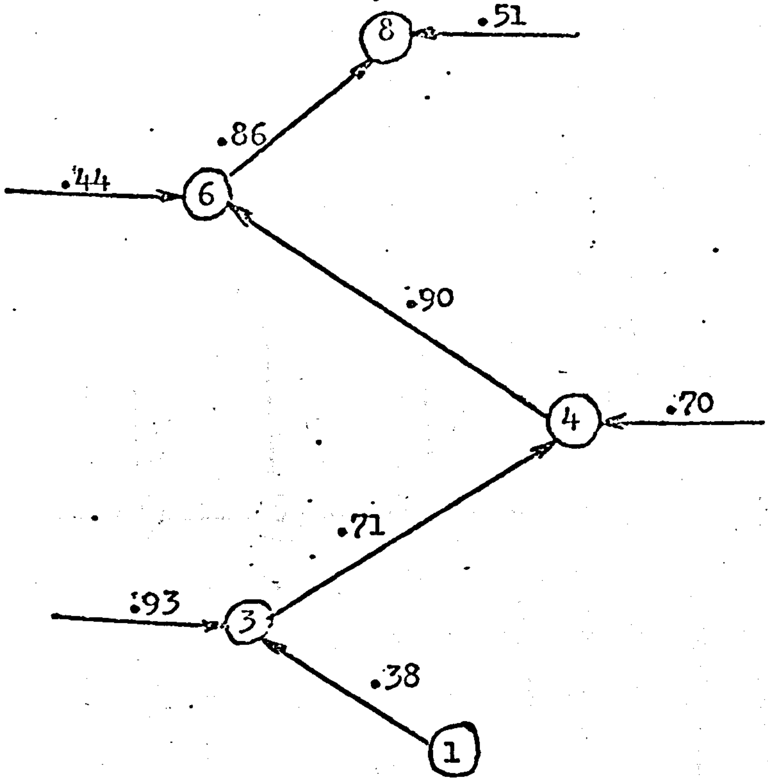


FIGURE 3 : A Model of the Status Attainment Process - Revised

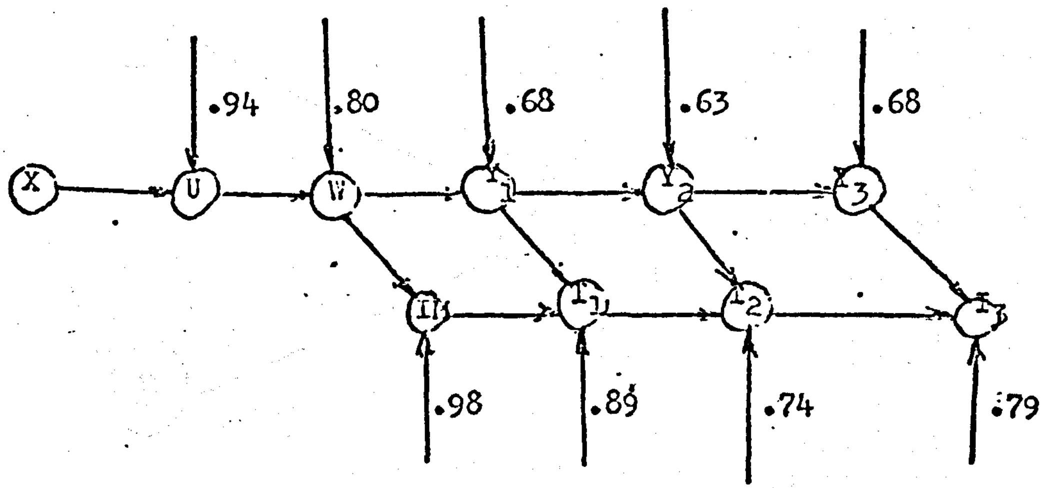
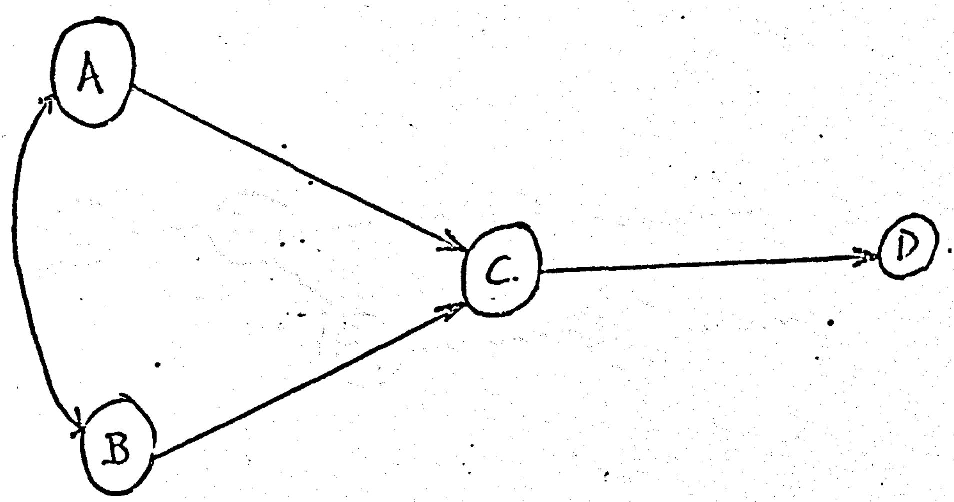


FIGURE 4 : A Social Structural Model - Revised



Concepts

A - Advantageousness of home

B - Mental Ability

C - Level of educational training

D - Level of status of job

FIGURE 5 : The results of SPA

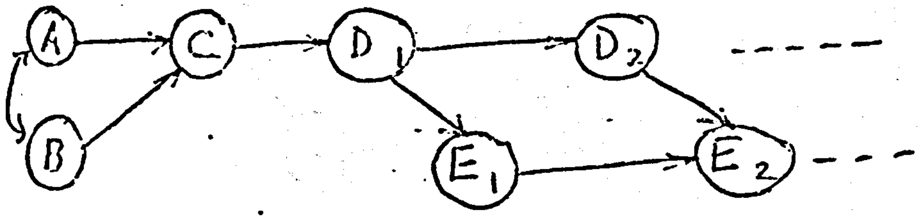
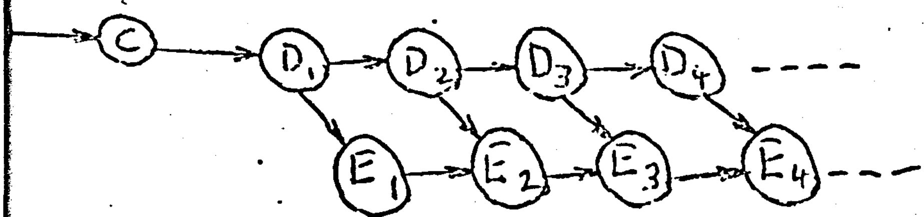


FIGURE 7 : The composite model



Additional concept

E - Monetary value of job

FIGURE 6 : The revised social structural model in new form.

CORRELATING ORDINALLY-SCALED VARIATES

- A note on the numbers game.

G J Boris Allan

School of Sociology

Manchester Polytechnic

Aytoun Street

Manchester

M 1 3 GH

ACKNOWLEDGEMENTS

Part of work reported herein was supported by the Health Services and Mental Health Administration of the US Govt, under contract HSM 110-70-393 at the Department of Community Health, Medical School, Nottingham.

CORRELATING ORDINALLY-SCALED VARIATES - A note on the numbers game.

At one time we assigned numbers 1 through 7 to the categories of the Hall-Jones scale, and found an ordinary Pearson product-moment correlation between this and educational attainment (categories scored 1 through 4 ?). Later Labovitz(1970) and others showed empirically that any order-preserving assignment of numbers to categories was almost the same as any other - in terms of the size of the resulting correlation. Kendall and Stuart (1973:586-588) extend the discussion of Williams(1952) and suggest that rank correlations will not differ by much from the Pearson formed by any order-preserving assignment of numbers. (The maximum Pearson correlation for a contingency table given all possible assignments of numbers is the 'canonical' correlation of the table. Kendall and Stuart give an example for which the canonical correlation is .697 and tau_b - a rank correlation - is .643). The implications are that for continuous-valued data we can use Spearman's Rho (which is no more than a Pearson correlation on ranks), and that for contingency tables it doesn't really matter whether we calculate tau_bs or Pearson correlations. (Hawkes(1971) suggests the use of tau_bs and this we can see has added attractions.)

I do not believe in that which has been proposed.

Firstly : for uhgrouped data by all means calculate rho(R), but then - if an assumption of an underlying bivariate normality is applicable - estimate the Pearson correlation (r) by

o Greek ('pie')

$$r = 2\sin\left[\frac{\pi R}{6}\right]$$

(Kendall,1973:Ch9).

Secondly : for grouped data no such easy transformation exists, and any correlation made by assigning numbers to ranks will underestimate because of grouping-effects. Take the simple case of the 2x2 table in Table 1. This is a double dichotomy (at the medial values) : upon a bivariate normal surface with a correlation of .75 - from tetrachoric tables. The Pearson correlation for this table is .5, as is tau_b ; perfect consistency in estimates of correlation - but all are consistently biased from the true value of .75.

For non-normal surfaces the true correlation will not be .75, but it is highly unlikely that any useful surface will have a correlation of .5, given these marginal proportions.

Table 1 about here

This may seem an extreme example, but many variables are no more than dichotomies : as a further illustration consider the data portrayed in Table 2. Table 2 is a 5x5 contingency table showing the relationship between father's height and son's height - a sample of 1000 observations on two extremely well-behaved variates. The correlation between the raw, un-grouped heights is .5189 (Karl Pearson's data). The Pearson correlation for this table (categories scored 1 through 5) is .466, and tau_b is .399, which is not a great difference in value. (Actually Macdonald (1973:107-108,113-115) shows that correlations can be quite different - changes of around 20% - but the influence of Labovitz is so strong that he thinks he has illustrated that '... providing that two comparable prestige rankings are available, one should not worry about the more difficult task of ensuring comparability of numeric assignments to these categories...'}.

Table 2 about here

We need more ingenuity applied to the problem of estimating grouping effects, for there can be implications for the examination of partial relations. For example, consider the (degenerate) trivariate normal surface with bivariate normal correlations .7071, .7071 and .50 - the surface is degenerate because a partial correlation is zero (.7071x.7071 = .50). The correlations estimated from collapsed tables such as Table 1 will be .50, .50 and .333. In this case .50x.50 $\neq$ .333 and so the partial correlation will not be zero. Checking to see if a partial correlation (or path coefficient) is zero is the cornerstone of causal modelling - yet notice the impact of grouping effects. If all three real correlations are .7071, all three partial correlations will be .4142. The corresponding observed (estimated) correlations of .50 produce partial correlations of .333. Again there is this radical difference in interpretation, a partial of 59% of the correlation as against a partial correlation of 67% of the correlation. (Decisions have been made on smaller

'is not equal to'

differences.)

In a pragmatic vein, it frequently commentd that the measure you use to find whether two things are associated is not too important; in a sense I can agree with this, but difficulties arise when you compare associations. I can think of two main reasons for comparison :(1) You wish to find if social class and educational attainment are more highly related in Britain or the US - this can be extended to many relationships; and (2) You may wish to find whether social class or mental ability is more highly related to educational attainment - which can also be extended. The first is comparative, the second is the nitty-gritty of modelling, and we can do both at once. The message of this note - the intellectual massage -is : we need to concentrate more on the impact of of grouping-effects of correlations and forget the non-problem of the assignment of numbers to rank-order categories; in so doing we will be making fewer mistakes in our multivariate analyses. The impact of grouping effects is probably why we find so many spurious, tiny paths (or beta coefficients) in reported analyses. After, most data can only be collected or used in a categorized form, and it is only rarely that we have a truly continuous-valued variate,(who mentioned Income ?).

TABLE 1

ABIVARIATE NORMAL SUFACE WITH CORRELATION

PARAMETER OF .7071

.375	.125	.500
.125	.375	.500
.500	.500	1.000

TABLE 2

CONTINGENCY OF STATURE OF FATHER AND SON*

	Father's Height					
Son's Height	203	91	26	9	6	335
	95	75	66	22	26	284
	30	36	37	14	20	137
	18	27	26	11	23	105
	12	35	25	13	54	139
	358	264	180	69	129	1000

* Taken from Tables for Biometricians and Statisticians Vol II,(lxxviii) by Karl Pearson

REFERENCES

HAWKES R K

1971 Multivariate analysis of ordinal measures. American Journal of Sociology 76:908-26

KENDALL M G

1970 Rank correlation methods (3rd ed) London: Griffin.

KENDALL M G and A STUART

1973 Advanced theoty of statistics. Vol II (3rd ed) London:Griffin.

LABOVITZ S A

1970 The assignment of numbers to rank-order categories. American Journal of Sociology 35:515-24.

MACDONALD K I

1973 The Hall-Jones scale.:97-115, in J M Ridge (ed) Mobility in Britain reconsidered. Oxford : University Press.

WILLIAMS E J

1952 Use of scores for the analysis of association in contingency tables. Biometrika 39:274-

APPENDIX F

TWO RESEARCH REPORTS - Department of Community Health
Medical School
Nottingham

For part of the period of the registration time for this thesis I have been a consultant in statistics and research methods to the Dept of Community Health. During this period I was most closely involved with planning the analysis of which there are two reports herein. I was totally responsible for the origination of, and actual analysis of, the mode of analysis. I played a large part in the many discussions on the philosophical and practical implications of this type of research, and my scientific realist philosophy will be apparent throughout.

This also enables me to thank Carlos, Jan, Mary and Maurice (in alphabetical order only) for the impact of their various forms of wisdom. If I wasn't a realist when I started this type of area of research turns you into one. After all, an infant death rate is children dying.

HEALTH INDICES SENSITIVE TO MEDICAL CARE VARIATION.

DEPARTMENT OF COMMUNITY HEALTH, MEDICAL SCHOOL, NOTTINGHAM UNIVERSITY.

C.J.M. MARTINI, M.D., M.P.H., M.Sc., M.F.C.M.
G.J.B. ALLAN, B.A.
J. DAVISON, B.Sc., Ph.D.
E.M. BACKETT, B.Sc., M.B.B.S., F.R.C.P., D.P.H.

This work was supported by a grant from the National Center for Health Statistics, HEW, to the Department of Community Health, Nottingham University.

1. The Problem.

The advance of medical science and rising standards of living in the twentieth century have brought about major improvements in health and life expectancy. Some diseases, like poliomyelitis and diphtheria, have almost disappeared in some societies. Others, like tuberculosis and measles, are far less common. In Britain, only twenty-six years ago, (1949), the infant death rate was 32.4 and the stillbirth rate 24 per thousand. At the end of the sixties the infant death rate was already less than 20 per thousand, and the stillbirth rate at 14 per thousand was down by nearly half. These, with crude death rate, birth rate, and maternal death rate, have been the most commonly used indices of community health for many years.

Since the mid-fifties, there have been some gains in health, some losses, and some areas where we are holding our own, or where progress has been uncertain.

An impression of what are the determinants of this mixed picture of gains and losses can be obtained with the aid of statistics routinely collected, or from the many special studies that have accumulated detailed information on specific diseases and conditions. Standards have been established for everything from body weight and height to blood cholesterol concentrations and the level of immunological resistance to infections. The prevalence of many chronic diseases has been described. Unfortunately, many such studies are based on groups selected to conform to special criteria and therefore not representative of the population as a whole. Only a small fraction of the existing statistics give an undisturbed picture of health and disease, and in consequence we are often faced with conflicting evidence, both of what the situation is and of what are its determinants. This is an undesirable situation as statistics are the foundation for programmes of action in this, as in most other areas of human activity.

A feature of much planning of medical services is our ignorance about the extent to which many of these indices which are assumed to reflect the effects of medical care actually do so rather than reflecting other circumstances affecting the lives of the populations concerned.

Although some research has already been done on ways of measuring whether medical care does what it sets out to do (and in the most effective way, giving due regard to considerations of economy and the best use of scarce resources) there is still considerable doubt as to the specificity of some of the measurements employed.

There are a great many "end results" of the interaction of illness and medical care and some (like the examples given above) can be rendered into simple rates or indices of "healthiness" or "unhealthiness" (mortality, morbidity, residual handicap, etc.). There are also, of course, numerous other aspects of this interaction where the end result aimed at is not necessarily health in individual terms but the achievement of the "best" medical care that can be given under the circumstances. Examples would be a shorter rather than a longer time between diagnosis and operation, and the lack of pain during terminal care. Such measures do not reflect health but they are assumed (often without sufficient reason) to be related to the most satisfactory outcome of medical intervention.

What are the advantages and disadvantages of end results (outcomes) as indicators of "good" medical care? The most important advantage is their validity. The values that govern what is a good or bad outcome are generally accepted and end results reflect the contribution of care rendered not only by physicians but by other health professionals as well. Furthermore, the 'results' tend to be concrete and therefore capable of measurement although there are difficulties when a long time-lag is involved before they can be assessed. It is also obvious that medical care which successfully postpones death may result in higher morbidity rates in the survivors; this was demonstrated by Sanders (1).

The main disadvantage of the use of outcomes arises from the fact that they are influenced not only by medical care but also by extraneous circumstances. Housing, occupation, education, air pollution, and the constituents of drinking water are some of the variables whose effects can combine with those of medical care. Certain outcomes which were formerly believed to be related mainly to medical care have now been recognised as predominantly influenced by socio-economic variables. For example, infant mortality contrasts with perinatal mortality in this respect, for whereas the latter is sensitive to medical care, infant mortality is much more influenced by home surroundings (2,3). This finding has led some investigators to propose the use of, for example, birth weight statistics, instead of infant mortality, for many of the purposes to which these indices are applied (4). In addition, the limitations of many individual outcome indices, such as mortality rates, are widely recognized (5) and the difficulties of measuring morbidity and disability well known. Understanding the last two measures is additionally complicated because they have different components such as severity, duration and social impact, that have not been thoroughly explored.

This paper presents the results of a study which attempted to extend the limited area of current knowledge concerning the relative importance of medical care in the determination of outcome indices. Only by identifying many more outcomes which are specifically sensitive to one or other aspect of medical care, and eliminating others, which respond most readily to changes in extraneous variables, can more effective decisions be made in Health Services Planning. This is not to deny that indices which are more sensitive to variation in socio-economic and environmental conditions are not valid measures of health status, but that for the evaluation of medical care, which is the main concern of the authors, interest must be focussed on the former group of outcome measurements.

The general hypothesis underlying this project was that some indices of health are more sensitive than others to variations in the pattern of provision and the resources invested. The essential points are:

1. Health can be measured in terms of outcomes which for most planning purposes are still expressed as mortality,

morbidity and disability indices.

2. Health outcomes are affected by three main types of circumstances: Health services, environmental conditions, and socio-economic patterns.

3. The contributions of each type of circumstances can be distinguished and hence a weighting obtained.

4. Not all measurable health outcome indices are equally influenced by all three kinds of external circumstances.

5. It is therefore possible to detect some which are particularly (though not necessarily exclusively) sensitive to different levels of provision of health services.

6. These indices can also be combined into more comprehensive measurements, which would be less subject to random fluctuations and more able to evaluate the effectiveness of the care provided.

2. <u>The Rationale.</u>

The degree to which health outcome indices are influenced by health services, environmental conditions and socio-economic patterns will be reflected in varying degrees of correlation between indices and the three circumstances.

A necessary step in the development and/or evaluation of health indices is the construction of models in which the relationships between health (as a theoretical concept) and the socio-economic and medical environment of the individual are expressed. Such efforts have, in the past, usually been unsuccessful due to the lack of appropriate methods. Modelling may offer a unique means to the understanding of the forces underlying the interaction of health and ecology, but their relationship is so intricate and complex that, at present, no attempt has been made to develop a really comprehensive model of health, in this investigation.

However, all analysis must be based on a theoretical framework, which may or may not be made explicit. This framework, in our rationale, assumed that there is a meaningful, real entity, the "Health of a Community", which can be seen in several constituents, (a constituent of Health will be called an "H"). For instance, an important constituent of the "Health of a Community" is considered to be the "Health of children under one year of age". The real health of the children is represented by H_{ch}.

This constituent, "Health of children under one year" is not measured perfectly by any of the usual indicators (like perinatal or neonatal mortality) so, by using many differing indicators of this concept, it is hoped to arrive at a closer approximation to the real dimensions. A set of indicators of H_{ch} will be termed I_{ch}'s.

The measure "H_{ch}" is the parameter of a distribution i.e. it is the average health of a set of individuals. It is possible to conceive of the parameters "H_{ch}" themselves following a distribution, and it is postulated on the basis of theoretical and practical considerations that the distribution of the values of "H_{ch}" will be normal. The theoretical reasons include the statistical notion of the Central Limit Theorem and the practical reasons those of maximising correlation and equalising ranges of values.

If values of H_{ch} are normally distributed what is the expected distribution of I_{ch}? Under perfect operationalisation the values of I_{ch} should be normally distributed, but the distribution of the observed values may be non-normal. If the observed distribution is non-normal, the reason for the discrepancy may be two fold. Firstly, our operationalisation may be incorrect either in scaling or conceptualisation, and secondly our assumptions about the distribution of H_{ch} may be invalid. In the absence of any firm theoretical or substantive considerations, a derived scale may be used if the derived scale values for any two variables are in the same order as the observed values. For the reasons given above, it was decided to use a scale based on the normal distribution. This transformation from observed to normal scores is a hypothesis for which it was felt there is a high degree of a priori legitimation and has been used by other workers (6).

If the I_{ch}'s all measure, perhaps imperfectly, H_{ch}, then all the I_{ch}'s, should be highly intercorrelated. This is a test of our I_{ch}'s consistency, i.e. are they measuring nearly the same thing? This type of scheme fits closely the mathematical model underlying factor analysis (7,8) and the relative variance of the first principal factor is a measure of consistency. If the I_{ch}'s are consistent, what they most consistently measure is the first principal factor which is an approximation to H_{ch} (it could be denoted $\hat{H}_{ch}$) and can be used as a combined index*.

In this way it is possible to arrive at a set of approximations to the constituents of the "Health of a Community". These constituents can be seen as the results of (1) random circumstances (2) the differential provision of medical care and (3) social, economic and environmental factors. Average levels of the various constituents of health are chosen in order to minimise the effects of random fluctuations. The problem then arises of measuring "differential provision of medical care" and which are the most relevent "social, economic and environmental factors". In a similar manner to that used to approximate "health", approximations to the medical, social, economic and environmental constituents of the factors affecting health can be derived by the use of selected indices. These approximations may then be analyzed to determine how sensitive are outcome indices to the provision of medical care and the extent to which there is incomplete separation of the main factors (i.e. the 'overlap' or the collinearity effect).

3. Results.

A large number of variables measuring various aspects of outcome (dependent), facets of the medical care process and socio-demographic characteristics (independent) have been extracted from the routinely

---

* Combined indices are much less susceptible to random or sporadic fluctuations than single indicators, and are often much more valid for comparative purposes (9).

published statistical sources for the years 1970 to 1972, in the United Kingdom. Many of these independent variables and all the dependent variables, relating to outcome of care, were standardized by age; sex and diagnosis.

Since it was clearly not feasible to include indicators for all possible diagnoses and operations in the study, a number of criteria were applied to their initial selection and in the selection of age and sex groups. These were that the diagnoses should be firm, well established and simple, the number of cases should be sufficient to enable rates to be calculated (this was the basis for the selection of age and sex groups), several types of health professionals should usually be involved in the management of the patient, the diseases should be important in economic and human terms and, preferably, costly in terms of hospital time and skills and information should be available for the diagnostic groupings from the major sources of data. Seventeen diagnoses for both immediate admissions and all admissions* and four operations were selected. These are listed in Table 1. Only two of these diagnoses were used in the last stages of the analysis, Cerebrovascular disease and Pneumonia, to represent the whole group.

On examination of the available data, it was concluded that the fifteen Regional Hospital Boards of England and Wales would be the most suitable units of analysis. The selection of these units of

---

* In the United Kingdom, a high proportion of admissions are arranged through waiting lists.

analysis necessitated aggregation or splitting of data, in some instances, as comparable areas of the country are not always used by the major government agencies collecting health and health related information. The actual numerical values obtained by such manipulation of the data can only be approximations to the exact values and, for this reason, it was decided to convert the calculated values to ranks and use these ranks in subsequent analyses. Ranking was also used because in many cases, it is uncertain which measure of central tendency should be used, and the values of the mode, median or mean appeared to maintain the same rank order throughout the 15 Hospital Regions.

The variables were grouped on the basis of *a priori* theoretical and substantive considerations. Among the 321 independent variables, one group represented Socio-demographic characteristics and five represented various aspects of the medical care system, namely Community expenditure, Traditional general practice, Hospital resources and performance, Met demand* and Efficiency of care**.

The 409 variables measuring final outcome were represented by five groups: Deaths under one year, Total mortality, Deaths in hospital, Multiple diagnoses (measures of case complication) and Certified incapacity (the latter comprised the only variables available as measures of morbidity in the total population).

---

* This group consists of hospital utilization variables which are population based such as outpatient attendance and discharge rates.

** This group comprises those variables relating to hospital activity such as time on waiting list and duration of stay.

The initial stage of the analysis was concerned with determining the most 'relevant' variables from among the large number selected in the first instance. In a preliminary analysis the outcome (dependent) variables were cross-correlated (using Spearman's rho) with each of the socio-demographic/medical care variables (independent). Following the correlation analysis, it was possible to rank the variables in terms of their number of 'large' correlations (R 0.7). This enabled the most consistently relevant variables to be identified. However, it was recognised that the importance of a particular variable may be masked in a strict count of large correlations. Therefore further criteria were also used to select variables based on their possible relevance in epidemiological and medical care terms.

The rank values for each of the 15 Hospital Regions of 66 independent variables and 38 final outcomes so selected from the larger number introduced into the preliminary cross-correlation analysis were converted to standard scores and the correlation coefficients recalculated. On factor analysing the two groups of variables separately, certain variables were found to be of low communality and were dropped from subsequent runs.

After this further selection process, 44 independent variables and 32 final outcomes were retained. From factor analysis it was possible to identify clusters of variables within the three groups.

The distinctive clusters conformed closely to the <u>a priori</u> groupings that had been used in the first stage of the analysis. Within the socio-demographic set two sub-sets could be distinguished, "status" and "urbanization". The clusters and their constituent variables are shown in Table 2.

A principal factor analysis was then performed for each cluster, each variable weighted by its factor score* and the weighted variables aggregated in an additive manner to form combined indices representing as far as possible the theoretical dimension associated with each cluster of variables.

The combined indices were factor analysed, and three factors extracted using the principal factor solution. (The total variance extracted by the three factors was 87.6%, being 39.1%, 35.4% and 13.1% for Factors (1), (2) and (3) respectively). An oblique transformation of these three factors still resulted in the correlations between them remaining almost zero so that they may be considered as being effectively orthogonal. This is important as no assumption was made of non-relatedness amongst the theoretical dimensions.

Examination of the factor pattern matrix (Table 3) shows that Factor (1) appears to represent an urbanization/medical care dimension, Factor (2) a socio-economic status/community mortality and morbidity dimension and Factor (3) an urbanization/hospital mortality dimension. A number

---

* The factor score for each variable is also shown in Table 2. The number of large correlations associated with each variable is also shown in this table.

of the combined indices are loaded on more than one factor. Thus urbanization appears to contribute equally to Factor (1) and Factor (3). *Although socio-economic* status appears to be mainly loaded on Factor (2), it does also make some contribution to Factor (1). Of the outcome indices, Deaths in hospital appears to be measuring a different dimension of outcome because, although loaded on Factor (2) which is the main factor representing outcome, its largest contribution is to Factor (3).

A multi-stage regression technique developed by one of the authors (10) from a method of Mood (11) was used to partition the variance in each combined outcome index between the socio-demographic and medical care components* shown in Table 3.

The bar diagrams in Figure 1 show the proportion of variance in each of the five combined outcome indices explained by the medical care and socio-demographic components respectively, and the joint contribution (collinearity) of these two components. The collinearity is a measure of the degree of overlap or lack of independence of the two components.

Each combined outcome index consists of a weighted combination of a number of individual indicators in which the weightings are the appropriate factor scores. If, within each combination of variables, there are considerable differences among the individual indicators

---

* The term 'component' is defined as a group of indices and is not to be confused with the terminology used in principal component analysis.

in sensitivity to variation in medical care, which are not reflected in the weightings, the effect of those indicators more sensitive to medical care variation may be masked by those in which the relative contribution from the socio-demographic component is large. A number of individual outcome indicators were therefore selected for a similar analysis from within each of the five groups of variables which constitute the combined outcome indices. The bar diagrams showing the partitioning of variance in each individual indicator between the medical care and socio-demographic components can also be seen grouped with tne representation for the combined index to which they contribute.

Deaths Under One Year.

The proportion of variance explained in this combined index is 0.70. The contribution to the variance explained from the socio-demographic component is 0.36 compared to 0.25 from that measuring medical care. Thus environmental influences appear to make a slightly greater contribution to this index of Deaths under one Year.

The five individual indicators contributing to this index have equal weightings (see Table 2). Two indicators, Infant Mortality per 1000 live births and Perinatal Mortality per 1000 live births, were selected for further analysis. The contribution of the socio-demographic component to the variance explained in Infant Mortality is more than twice that from the medical care component, while, for Perinatal Mortality, the contribution from the two components is almost equal. As Perinatal Mortality comprises stillbirths and deaths during the first week of life, this finding confirms what is currently known about the effects of medical care on infant mortality shortly before, during and shortly after childbirth when environmental factors appear to be of lesser importance.

Total Mortality.

The proportion of variance in this index of mortality explained by the medical and socio-demographic components is 0.75, which is slightly higher than that for Deaths under one Year. Further examination of the partitioning of the variance between these two components shows that the socio-demographic component explains four times as much of the variance in this outcome index as the medical care component, indicating that the socio-demographic variables account for most of the variation in this index.

Three individual indicators were selected from within this group for further analysis, Death Rate (adjusted) per 1000 home population, Standardized Mortality Rate for Cerebrovascular disease in Males and Age-Sex Specific Death Rate for Pneumonia in Males aged 65-74. The proportion of variance in these individual indicators explained by the medical care and socio-demographic components varies from 0.54 for Death Rate (adjusted) to 0.82 for Age-Sex Specific Death Rate. The partitioning of variance between the two components also differs markedly for these three indicators. The indicator, Death Rates (adjusted), appears to be equally sensitive to medical care and environmental influences. The other two indicators in this group do not show a similar sensitivity to medical care. In Age-Sex Specific Death Rate - Pneumonia in Males 65-74, the contribution from the socio-demographic component is very large (0.78), which is probably reflecting deficiencies in the environment of the patients. The joint contribution of the socio-demographic and medical care components to the variance explained in the indicator, Standardized Mortality Ratio for Cerebrovascular disease, is of the same order of magnitude as the individual contribution from the socio-demographic component, implying an incomplete separation of medical care and environmental influences for this indicator.

Deaths in Hospital.

The proportion of variance explained in this combined index is high (0.88), but the contribution of the medical care component is just over half that of the socio-demographic. The collinearity is negative, indicating the existence of suppressor effects.

Examination of the partitioning of variance in the seven individual indicators selected from this group, shown in Figure 1, reveals first the existence of two distinct groups, namely deaths within 48 hours and all deaths (the latter group is divided into immediate admissions* and all admissions). The former group, which describes the mortality experience of patients admitted as emergencies and dying in the first two days following admission, is characterised by a

* In the United Kingdom a large proportion of patients referred to hospitals with less severe diagnoses are not admitted immediately, but through a waiting list, which, in some cases is of many weeks duration.

relatively low proportion of variance explained, a small contribution from the medical care component and either a small positive collinearity or no joint contribution from the socio-demographic and medical care components. The relatively low proportion of variance explained, particularly in the case of Cerebrovascular disease and All diagnoses, implies that certain variables of significance in emergency admissions and very severe cases have not been included. These missing variables are probably related to the seriousness of tha patient's condition, and not to the medical care system, as the distinct contribution of the medical care component to the variance explained is similar for the two individual diagnoses and for all diagnoses. The proportion of variance explained for these three individual indicators parallels the factor score weightings, shown in Table 2, which were used in the construction of the combined index for Deaths in Hospital.

The remaining four variables selected from those contributing to this combined index are measures of case fatality rates for all deaths (both before and after 48 hours of admission). The proportion of variance explained in these individual indicators is relatively high, ranging from 0.73 for Immediate Admissions for pneumonia to 0.93 for Immediate Admissions for all diagnoses. For the two specific diagnoses selected, Cerebrovascular disease and Pneumonia, the distinct contribution from the medical care component is less than half that from the socio-demographic component. The proportion of the variance in case fatality rate for all diagnoses explained by the medical care component alone is the same for all admissions and immediate admissions. However, the contribution from the socio-demographic component is higher for immediate admissions and this also accounts for the higher proportion of total variance explained in this indicator.

Multiple Diagnoses.

The combined index, Multiple Diagnoses, is currently used as a proxy measure of case complication rate, albeit a crude one, as the additional diagnoses may be concurrent conditions and not necessarily complications affecting the principal diagnosis. This index appears to be most sensitive of the five outcome indices to variation in medical care. The contribution from the medical care component is almost four times as large as the contribution from the socio-demographic component. However, it is perhaps questionable whether this indeed can be considered as a true measure of outcome. It appears to be more strictly a measure of process - a function of the available technology, the sophistication of the system of medical care and the attitudes of health professionals. This is borne out by the pattern of the factors extracted on factor analysing the correlations between the combined indices, shown in Table 3. It can be seen that, of the five oombined indices, Multiple Diagnoses loads most highly on Factor 1 which is the factor representing characteristics of the medical care system.

Four individual indicators were selected for further analysis from the group of variables comprising this index. With the exception of Multiple Diagnoses rate (immediate admissions): Cerebrovascular disease, the proportion of variance explained by the socio-demographic and medical components is high, varying from 0.73 to 0.84. For the two individual diagnoses, Cerebrovascular disease and Pneumonia, there is essentially no contribution from the socio-demographic component to the variance explained. The contribution from the medical component is high for Pneumonia, being almost twice that for the other single diagnosis, Cerebrovascular disease. The partitioning of variance between the medical care and socio-demographic components shows a different pattern for immediate admissions and all admissions when calculated over all diagnoses. In the former case the contribution from the medical care component is almost twice that from the socio-demographic component, but, for all admissions, the reverse situation applies (probably, in the case of immediate admissions, for more severe cases, more intensive and dedicated care is common, which may mean that additional problems are more likely to be diagnosed). The collinearity is relatively high for Multiple Diagnoses rate for all diagnoses for both immediate and all admissions, implying an incomplete separation into medical care and socio-demographic components.

Certified Incapacity.

The only measures of morbidity that could be obtained from the routinely published statistics related only to the working population and their days and spells of incapacity for work as certified by general practitioners. For the combined index, Certified incapacity, the contribution of the medical care component is less than half that of the socio-demographic component.

This index comprised two individual indicators. One of these, Days Certified Incapacity per 1000 males, was selected for further analysis. This indicator showed the same degree of sensitivity to medical care as the combined index, but the contribution from the socio-demographic component was almost halved. This reduction in sensitivity to socio-demographic influences accounted for the decrease in the proportion of variance explained.

4. Discussion.

It would be outside the reasonable length of this paper to discuss extensively all the possible implications of each individual finding. However, the overall results of this study seem to indicate that indices constructed from the traditional outcome measures are more sensitive to variations in the socio-economic and environmental circumstances of the population than to the amount and type of medical care provided and/or available. This seems to be especially true for those indices, such as infant mortality or certified incapacity, which are community based. By contrast, those indices, such as case fatality rates (especially case fatality rates: immediate admissions for all diagnoses), which apply to care provided in hospitals, appear to be relatively more sensitive to medical care. Possible implications could be: first, that these indices are not really measuring health outcomes; and, second, that health outcomes are even less affected by medical care than is currently assumed; or a combination of both. However, before statements as condemnatory as these can be justified, it is necessary to consider the problems of validity and reliability of the indices. This is a difficult area due to lack of criteria, external to the indicators themselves. The indices used in our study are a function of the quality of the data input. The quality of the data used, although the 'best' available, was less than perfect; if, for no other reason, than aggregation and splitting of the data was necessary in those cases in which comparable areas were not used by the agencies collecting the information. Deductions made from analyses of existing data, however complete, cannot, in a case like this, be a satisfactory substitute for those based on experimental methods, though they can form the basis of hypotheses. Experimental evidence is required both to sustain the hypothesis and to establish the magnitude of the cause-and-effect relationship between given factors. Causal relationships cannot be proved from the results of such a study; they can only be inferred.

The regression method is sensitive to the strengths or weaknesses of the material used. The more precise and detailed the observations, and the greater the understanding of the structure of the medical care system, the greater the confidence that can be placed in the conclusions. To obtain the best picture, particularly when regression analysis is employed, the numbers of units of analysis and of variables are of crucial importance. In our case, there was a need to obtain enough information (usually produced by different agencies) in as many units of analysis as possible. In our study only 15 units were used, and possibly these are too few for an ideal use of multivariate analysis techniques. However, as incremental contributions to the variance were estimated rather than regression coefficients, a small number of units of analysis is less critical.

There were a number of theoretically important variables in both the independent and dependent groups which could not be incorporated because no information was available. These included survival rates for certain diseases, patient satisfaction, restoration of physical and social function and also the amount of residual morbidity outside the hospital (certified incapacity is not a good indicator of the amount of illness in the community). Indices of health measuring disability presented a particularly intractable problem. There was very little information available and the local registers were incomplete and thus inadequate for our purposes. Among these missing variables were some that may have provided the most satisfactory measurements of outcome of care, but a population survey would have been necessary to provide this information. If special studies of this type were able to show that these missing indicators were strongly associated with the indicators measuring aspects of medical care, then a good case could be made for keeping augmented records routinely.

A possible fallacy which could be of importance and which has caused difficulty in previous social research is that the patients who provided the information for the outcome indices may not have the socio-economic characteristics of the resident population from which some of the independent variables were derived. In general it is assumed that utilization of health services does in fact vary according to social class, (even within a health services system in which there is open access to medical care) so probably the patients are not a representative cross-section of the population. Conclusions are, however, drawn about areas and not individuals and it was assumed, in this study, that the extent of the discrepancy does not differ between the units of analysis.

To eliminate this difficulty would require either specific studies on the patients to determine their socio-economic characteristics, or that such information should be collected routinely about each patient, which is not at present done in the Study Areas.

Most of the other pitfalls of the methods in this study have been mentioned as the relevant methods were explained. However there is one further point that needs discussion: the combination of indicators which may have elements in common.

In a weighted combination of two indicators, which have a common element, the size and sign of the weightings must be taken into account in the interpretation of the combination. If both weightings are of the same sign then the importance of the common element is increased: when the weightings are of opposite sign the importance of the common element is reduced. For example, in the combination of peri-natal mortality (stillbirths and deaths within one week of birth) with infant mortality (deaths within a year of birth), the importance of 'deaths within one week of birth' is accentuated.

A different problem arises when studying individual indicators, one of which is included within the other: for example, the partitioning of variance in the case of hospital deaths within 48 hours due to cerebrovascular disease, and total hospital deaths (within and outwith 48 hours) for the same diagnosis. The variance explained for deaths within 48 hours is low (0.23) with the largest contribution from the medical care component (0.11), whilst the variance explained for total deaths is high (0.83) with a high socio-demographic component (0.63). Two separate issues are apparent in explaining the difference in the partitioning of the variance in these two indicators. Firstly, the low proportion of variance explained for deaths within 48 hours suggest that the variables included in this study were not appropriate to the consideration of more urgent cases (for instance, no data on availability of intensive care units was included). Secondly, the contribution from the socio-demographic component is high for total deaths in comparison to that for deaths within 48 hours indicating that, for those cases in which death does occur within 48 hours, recovery is more dependent on socio-demographic characteristics (probably due to the different case-mix in these patients).

From these findings it would appear that the weakest group of independent variables may be those measuring medical care, while the socio-economic/urban characteristics of the population are better described by the variables included in the socio-demographic group (this last type of information was mainly census based). This is borne out by the partitioning of variance in some of the other indicators of outcome.

At this point it must also be emphasized that, although, in most cases, the 'combined indices' explained a larger proportion of variance, they were not more sensitive to medical care than all their constituent indicators. These combined indices, however, have other advantages, as outlined previously.

After this necessarily brief review of some of the possible methodological shortcomings in this study, let us assume that these results may be expressing, albeit crudely, a real phenomenon. In other words, the impact of medical care on the indices measuring outcome is only secondary to the effect of the socio-economic and environmental circumstances of the population. This does not necessarily mean that medical care is not affecting health but that the traditional measures (with the possible exception of case fatality rates) may be inappropriate for use in at least part of the planning of health services.

Two very important stages in this planning process are: initially, the detailed description of the actual situation in terms of health status and later, after implementation of policy decisions, a periodic focused evaluation of the achievement of objectives.

The first analysis is based mainly on information of the frequency and distribution of health problems, and it is for this purpose that most of the outcome indices analysed in this study could be used. These indices still remain the most readily available proxy measurement of health status of a Community and it must always be borne in mind that the events which comprise rates have an intrinsic value in themselves (for example, the death of a child) which goes beyond the statistical meaning of the rates. One must consider individuals (even single cases) when formulating social policy.

However, it is in the later stage of evaluation, mentioned previously, that a very important problem arises. To be able to evaluate effectiveness and quality of our programmes, it is necessary to focus on those aspects of the health-sickness process that theoretically can be affected by our efforts. This is where the value of many of the indices used in our study is very limited. Most of the important advances of medical care in the last 20 or 30 years are related to the quality of life before death for which we do not have, as yet, any precise measurements. An example of this is osteo-arthritis of the hip which seriously restricts mobility and produces sufficient pain to very severely disable the patient. In many cases regular analgesics are needed even when the patient is resting. However, since the early 1960's, an entirely successful operation can be performed, the arthroplasty of the hip, which changes those patients affected by this gross disablement into practically normal individuals. It is difficult to imagine a more dramatic improvement in the quality of life due to medical care. Measurement of improvements such as these are not to be found amongst the traditional indicators of outcome (case fatality rate of arthroplasty of the hip operations is only 1.2%).

This lack of appropriate indicators is even more pronounced in the area of primary medical care, only a small part of which is concerned with mortality, and where evaluation based on measures of morbidity does not give credit to the work of those health teams, who are concerned also with the patient's social, emotional and psychological well-being.

To conclude, it is important to emphasize that health is not fully describable in terms only of mortality, morbidity and disability, and it is probably in the search for new indicators of quality of life that the way forward lies.

BIBLIOGRAPHY

1. Sanders, B.S. : (1964) Measuring Community Health Levels. Amer J Pub Health, 54, 7, 1063-1070.

2. Shah, F.K., Abbey, H : (1971) Effects of Some Factors on Neonatal and Postnatal Mortality : Analysis by a Binary Variable Multiple Regression Method. Milbank Memorial Fund Q, 49, 1, 33-57.

3. Buckatzsch, E.J : (1947-8) The Influence of Social Conditions on Mortality Rates. Population Studies, 1, 3, 227-248.

4. Lewis, R., Charles, M. and Patwary, K.M. : (1973) Relationships Between Birth Weight and Selected Social, Environmental and Medical Care Factors. Amer J Pub Health, 63, 11, 973-981.

5. Donabedian, A. : (1968) Promoting Quality through Evaluating the Process of Patient Care. Med Care, 6, 3, May 1968.

6. Ashford, J.R., Read, K.L.Q., and Riley, V.A. : (1973) An Analysis of Variations in Perinatal Mortality Amongst Local Authorities in England and Wales. Int J Epid, 2, 1, 31-46.

7. Harman, H.H. : (1967) Modern Factor Analysis. Chicago : University Press.

8. Van de Geer, J.P. : (1971) Introduction to Multivariate Analysis for the Social Sciences. San Francisco : W.H. Freeman & Co.

9. Hertzberg, P. : (1969) Parameters of Cross-Validation. Psychometrika Monograph : Supplement 16, June.

10. Allan, G.J.B. : (1974) Simplicity in Path Analysis. Sociology, 8, 2, 197-212.

11. Mood, A.M., : (1971) Partitioning Variance in Multiple Regression Analyses as a Tool for Developing Learning Models. Am. Educ. Res. J. 8, 2, 191-201.

TABLE 1.

SELECTED DIAGNOSES AND OPERATIONS

ICD CODE *	DIAGNOSIS
140-99	All malignant neoplasms
151	Malignant neoplasm of stomach
162	Malignant neoplasm of trachea, bronchus and lung
174	Malignant neoplasm of breast
185	Malignant neoplasm of prostate
204-7	Leukemia
250	Diabetes mellitus
410	Acute myocardial infarction
411-4	Other ischaemic heart disease
430-8	Cerebrovascular disease
454	Varicose veins of lower extremities
480-6	Pneumonia
531-3	Peptic ulcer (excluding gastro-jejunal ulcer)
540-3	Appendicitis
550-3	Hernia (with or without mention of obstruction)
574-5	Cholelithiasis and cholecystitis
N820	Fracture of neck of femur

GRO CODE **	OPERATION
410-1	Inguinal hernia repair
520-9	Gall bladder operations
441-4	Appendicectomies
893-4	Varicose veins operations

TABLE 2

CLUSTERS AND THEIR CONSTITUENT VARIABLES

		Factor Score	No. of large Correlations ($R \geq 0.7$)
SOCIO-DEMOGRAPHIC CHARACTERISTICS			
1. SOCIO-ECONOMIC STATUS			
Demographic:	Lower quartile age of the total population	0.8	53
	Upper quartile age of the male population	0.6	34
Socio-economic:	Households per car	-0.5	101
	Socio-economic group: Foreman & Supervisors, skilled manual workers, own account workers, (8,9,12,14)	-0.9	91
	Socio-economic group: Employers & Managers (1,2,13)	0.9	89
	% of households having rateable values £100 or less	-0.9	85
	% of students remaining in school after statutory school-leaving age	0.7	44
2. URBANIZATION			
	Population per hectare	1.0	50
	% of females economically active	1.0	50
CHARACTERISTICS OF THE MEDICAL CARE SYSTEM			
1 COMMUNITY EXPENDITURE			
	Executive Council expenditures on General Medical Services	0.7	73
	Local Authority expenditures on Social Services	0.7	55
	Total expenditures on Health and Social Services	0.9	44
2. TRADITIONAL GENERAL PRACTICE			
	% of General Practitioners in solo practice	1.0	84
	Median age of General Practitioners	1.0	48
3 HOSPITAL RESOURCES AND PERFORMANCE			
Hospital manpower:	Nurses & Midwives per 10,000 population	0.9	52
	Total hospital manpower per 100,000 population	0.9	49
	General Medical SHMO and Consultants per 100,000 population	0.8	54
Hospital in-patients:	Average daily occupied beds per 1000 population	0.9	92
	Obstetrics: ante- and post-natal beds per 1000 females, aged 14-44	0.7	68
Teaching hospitals:	Deaths and discharges as a % of all deaths and discharges	0.3	57
Beds per million for:	Malignant neoplasm	0.9	76
	Hernia	0.8	74
	Acute myocardial infarction	0.9	73
	Peptic ulcer	0.7	72
	Cholelithiasis	0.5	60
	Malignant neoplasm of trachea, bronchus and lung	0.9	53

TABLE 2 continued

	Factor Score	No of large Correlations ($R \geq 0.7$)
4 MET DEMAND		
Outpatient attendances per 1000 population	0.8	50
Discharge rate from hospital per 10,000 population for:		
Peptic ulcer	0.7	19
Hernia	0.5	16
All diagnoses	0.9	13
5 EFFICIENCY OF CARE		
Time on waiting list before admission to hospital for:		
Peptic ulcer	-0.6	10
Hernia	-0.8	31
All diagnoses	-0 7	21
Median duration of stay in hospital for:		
Peptic ulcer	0.8	42
Hernia	0 8	30
All diagnoses	0.8	34
Median duration of stay (immediate admissions) for:		
Acute myocardial infarction	0.7	17
Peptic ulcer	0.6	29
Cholelithiasis	0.6	11
Median time on waiting list (Operations) for:		
Gall bladder	-0.8	33
Varicose veins	-0 8	28
Waiting time in hospital before operation for:		
Gall bladder	0 5	12
Duration of stay in hospital (Operations) for:		
Varicose veins in males	0.8	17
Gall bladder in females	0.6	17

TABLE 2 continued

MEASURES OF FINAL OUTCOME	Factor Score	No. of large Correlations ($R \geq 0.7$)
1 DEATHS UNDER ONE YEAR		
Infant mortality per 1000 live births	1.0	8
Perinatal mortality per 1000 live births	1.0	8
Neonatal mortality per 1000 live births	1.0	7
Environmental deaths per 1000 total births	1.0	6
Infant death rate for pneumonia	1.0	10
2. MORTALITY (Total)		
Death rates (adjusted) per 1000 home population	0.9	8
Standardised mortality ratios for:		
Malignant neoplasm of trachea, bronchus and lung in males	0.5	13
Cerebrovascular disease in males	0.9	19
Age-sex specific death rates:		
Malignant neoplasm of trachea, bronchus and lung in males 65-74	0.3	20
Cerebrovascular disease in males 65-74	0.9	17
Pneumonia in males 65-74	0.8	9
3. DEATHS IN HOSPITAL		
Case fatality rate per 100 discharges within 48 hours of admission (all admissions):		
Cerebrovascular disease	0.3	1
Pneumonia	0.8	3
All diagnoses	0.4	3
Case fatality rate net per 100 discharges (all admissions):		
Cerebrovascular disease	0.8	3
Pneumonia	0.8	3
All diagnoses	0.8	2
Case fatality rate per 100 discharges within 48 hours of admission (immediate admissions):		
Cerebrovascular disease	0.3	1
Pneumonia	0.7	8
All diagnoses	0.4	3
Case fatality rate net per 100 discharges (immediate admissions):		
Cerebrovascular disease	0.9	3
Pneumonia	0.9	3
All diagnoses	0.8	3

TABLE 2 continued

	Factor Score	No. of large Correlations ($R \gtrsim 0.7$)
4. MULTIPLE DIAGNOSES (Complicated cases)		
Maternity complications:		
Abortions. therapeutic and other	0.3	6
Multiple diagnoses rate per 100 discharges for:		
Cerebrovascular disease	0 6	0
Pneumonia	0.8	9
All diagnoses	0 7	10
Multip'e diagnoses rate per 100 discharges (immediate admissions) for:		
Cerebrovascular disease	0.7	3
Pneumonia	0.9	5
All diagnoses	0.9	4
5 CERTIFIED CAPACITY		
Inception rate per 1000 males	1.0	12
Days certified incapacity per 1000 males	1.0	13

TABLE 3.

OBLIQUE FACTOR PATTERN MATRIX OF THE COMBINED INDICES AFTER ROTATION

COMBINED INDEX	FACTOR (1)	FACTOR (2)	FACTOR (3)	COMMUNALITY
I. SOCIO-DEMOGRAPHIC INDICES				
Socio-Economic Status	0.42	0.82	-0.13	0.86
Urbanization	0.61	0.06	0.70	0.92
II. MEDICAL CARE INDICES				
Community Expenditure	0.67	0.40	0.34	0.78
Traditional General Practice	0.90	-0.02	0.20	0.87
Hospital Resources	0.97	-0.11	-0.08	0.95
Met Demand	0.91	-0.20	-0.26	0.91
Efficiency of Care	0.91	-0.12	-0.06	0.83
III. FINAL OUTCOME INDICES				
Deaths Under One Year	0.20	-0.90	0.17	0.84
Mortality	0.07	-0.94	0.25	0.92
Deaths in Hospital	-0.19	-0.41	0.90	0.93
Certified Incapacity	0.03	-0.93	0.03	0.84
Multiple Diagnoses	-0.32	0.84	0.19	0.85

FIGURE 1.

THE SENSITIVITY OF COMBINED OUTCOME INDICES AND INDIVIDUAL HEALTH INDICATORS TO MEDICAL CARE

(PROPORTION OF VARIATION EXPLAINED BY MEDICAL CARE AND SOCIO-DEMOGRAPHIC COMPONENTS).

DEPENDENT VARIABLE	MEDICAL CARE	SOCIO-DEMOGRAPHIC	COLLINEARITY	PROPORTION OF VARIANCE EXPLAINED
DEATHS UNDER ONE YEAR	.25	.36	.09	.70
Infant Mortality per 1000 Live Births	.18	.39	.06	.63
Perinatal Mortality per 1000 Live Births	.26	.29	.06	.61
TOTAL MORTALITY	.14	.56	.05	.75
Death Rates (Adjusted) per 1000 Home Population	.29	.24	.01	.54
Standardised Mortality Ratio: Cerebrovascular Disease in Males	.08	.29	.25	.62
Age-Sex Specific Death Rate: Pneumonia in Males 65-74	.07	.78	-.03	.82

FIGURE 1 (continued).

DEPENDENT VARIABLE	MEDICAL CARE
MULTIPLE DIAGNOSES	.50
Multiple Diagnoses Rate: All Diagnoses	.21
Multiple Diagnoses Rate: Immediate Admissions	
a. Cerebrovascular Disease	.37
b, Pneumonia	.63
c. All Diagnoses	.41
CERTIFIED INCAPACITY	.22
Days Certified Incapacity per 1000 Males	.29

SOCIO-DEMOGRAPHIC	COLLINEARITY	PROPORTION OF VARIANCE EXPLAINED
.13	.16	.79
.39	.22	.82
.01	-.01	.37
.03	.06	.73
.21	.22	.84
.50	.05	.77
.26	-.01	.54

COMBINED HEALTH OUTCOME INDICES AND THEIR USE IN RESOURCE ALLOCATION.

C.J.M. Martini, G.J.B. Allan and M.N. Garroway.

Supported under a grant from the National Center for Health Statistics, HEW, to the Department of Community Health, Nottingham University, to evaluate the sensitivity of health indices to variation in medical care.
Address communications to Dr. Carlos Martini, M.Ph., M.Sc., M.F.C.M., Health Services Research Group, Department of Community Health, Nottingham University, England.

SUMMARY

A proposal is being made for the use of combined health outcomes in formulae used for resource allocation between areas or hospitals. A possible methodology is described for the combination of mortality, morbidity and disability indices, and an example provided with information about 15 hospital regions in England and Wales.

Preliminary results support the suggestion that outcomes in some hospital regions differ from those expected on the basis of their medical facilities and social status. The limitations of traditional health indices and the data available are also discussed and future steps proposed.

INTRODUCTION

Major improvements in health and in life expectancy are often attributed to advances in medical science, but changes in the environment and improved standards of living have also played an important part.

Because of this, evaluation of the importance of medical advances is made difficult, since these extraneous factors are largely outside the influence of those responsible for medical care.

Infant mortality rates provide a typical example of confusion in this field, for they have been shown to depend more on such extraneous factors than on the quality of medical care. Traditional health indices are often poor indicators of the true picture of health and disease, probably because they were initiated for purposes of routine administration, and are necessarily somewhat inefficient when used in another context - that of evaluating the health service in terms of its impact on health and disease in the community. Nevertheless, they are in many cases the only measures available. While we should not underestimate their usefulness, we must recognise that they are not sufficiently sensitive to monitor changes in the provision of medical care.

The purpose of this paper is to present some preliminary results in the search for new and better health indices and particularly for those sensitive to changes in the provision of medical care. Such measures, should they be found, could be of considerable value for the planning of health services and the more efficient use of techniques, such as cost-benefit analysis, PPBS, and systems analysis.[1,2,3,4,5,6,7,8,9]

Outcomes and medical care.

Three features of the interrelationship between illness and services have been identified: structure, process and outcome[10]. All are interrelated, and in fact mutually reinforcing.

This paper reports work in progress in the development of sets of combined 'traditional' health outcome indices. By outcome or end results we mean the final health status of individuals after discharge or exit from a medical care system. This operational definition involves the idea of a change occurring between entrance into and exit from the system and the potential importance of individual experiences with health services.

Traditionally these outcomes are measured by morbidity, mortality and disability indices, although awareness of their limitations and the need for new measurements is a constant subject of most literature in health indices. However, since they are still used extensively all over the world, our first efforts were concentrated on them.

Our research objectives were (a) to extend the limited area of current knowledge concerning outcomes of medical care and the effect of extraneous variables in the determination of these outcomes; and (b) to develop a more comprehensive set of simple measurements that could be used as a first "detector" of abnormal situations and be expanded into a continuous monitoring system with automatic sensors for detection of particular areas in which existing knowledge is deficient.

To demonstrate the potential utility of combined outcome indices, consider the example of the current method of allocating funds to hospitals in the U.K. At the present time, revenue is allocated to each of the 15 hospital regions in England and Wales on the basis of its population, age and sex structure, average daily total of occupied beds by specialty, and case-flow. This formula is being progressively introduced. No data about needs, nor about outcomes of medical care for the region's population are included[11] although it might possibly be argued that using adjusted population figures could provide a proxy indicator of medical needs. We propose that the incorporation of a set of outcome indices according to this formula could perhaps ameliorate some of the deficiencies, and supply some of the missing information about the efficiency and effectiveness of care. However, other criteria must also be used to determine the allocation of resources between regions or to decide on additional financing needed for a particular region, and of course, the use of outcome indicators should not exclude using indicators of structure or process of delivery of care (e.g. beds available or discharge rates).

What are the advantages and disadvantages of outcomes as indicators of medical care?

- The most important advantage would be their acceptability. The absence of disease, or reduction of disability are culturally defined, and are assumed to reflect the contribution of care rendered by all health professionals. Furthermore, the results are capable of measurement, though there are difficulties when a long time-lag is involved before results can be assessed.
- One of the main disadvantages of the use of outcomes to show the effects of medical care arises from the fact that they too are influenced by circumstances external to medical care. Housing, occupation, education, air and water pollution are some of the variables whose effects can combine with those of medical intervention to confuse causal understanding. Certain outcomes which were formerly believed to be related to medical care have now been recognised as predominantly influenced by socio-economic variables. For example, we have already mentioned infant mortality as being influenced much more by socio-economic conditions than by medical care.[12,13.] In addition the limitations of many individual outcome indices, such as mortality rates, are widely recognized[14] and the difficulties of measuring morbidity and disability are well known. Understanding the last two factors is made even more complicated because they have different components, such as severity, duration, social impact, that have not been thoroughly explored.

We are aware that combined measures of mortality, morbidity and disability will also be incomplete in two senses: first, the combined indices do not include the positive aspects of health; and second, the development of these measures will require several progressively "finer" definitions from the initial crude measures.

METHODOLOGY.

A necessary step in the development and/or evaluation of health indices is the construction of models in which the relationships between health (as a theoretical concept) and the socio-economic and medical environment of the individual are expressed. Such efforts have, in the past, usually been unsuccessful due to the lack of appropriate methods. Probably modelling offers a unique way of understanding the forces underlying the interaction of health and ecology, but their relationship is so intricate and complex that such an effort would have been outside the resources available for our study. Therefore, no attempt has been made in this study to develop a comprehensive model of health.

However, all analysis must be done with a certain model in mind, which may or may not be made explicit. Typically, the thinking is in terms of regarding a certain social condition, for example health, as depicted by a set of background variables (within and/or without the system of health care) which in turn produce certain behavioural elements, health outcomes in our case. The design would then study how outcome elements vary with changes in the background variables. Here the level of analysis and the value of the results will be a function of how many variables the analyst is able to handle simultaneously and how well it is possible to explain "how the processes run" (or how the variables are related).

The theoretical model on which our study was based is explained below.

In a particular area, a set of characteristics of the individual, the environment and the medical care system (depicted by S, where each element "s" represents one aspect of the social-demographic or medical care characteristics of an individual) results in a set of health outcomes (depicted by H, where each element "h" represents one aspect of the health status of an individual). H and S are theoretical constructs in that they may not be observed directly, but are believed to have a real existence. We have chosen to represent the observed measures of H and S by a set of outcome indicators (HI) and a set of socio-economic, environmental and medical care variables (SI) respectively. The relationships described are shown diagramatically in Figure 1.

(Figure 1)

The elements of HI and SI are not direct observations of the theoretical constructs, so we must recognise that we could be failing to describe the "true" or "real" characteristics of the elements of H and S. If the assumptions are made that the different elements of H and S would be revealed in a factor analysis and that the variables included in the study actually represent outcome and social factors, we can find the extent to which the observed elements of HI and SI fit the model. We attempted to determine the interrelationships between the elements of H and HI by factor analysing the correlations between all elements in the HI set.

Values of the health outcomes in the set H were estimated by the use of weighted combinations of the outcome indicators. A similar approach was used to determine the interrelationship between the sets S and SI. Since the final selection of variables from the HI and SI sets has been based on the interrelations of these two sets, the factors in the H and S sets should be related.

The use of composite outcome indices has been advocated in the past with the argument that, by summarizing closely related measures, the composite indices will greatly facilitate analysis of community health status [15]. However, there is no agreement in the literature about the degree to which aggregation of social indicators is acceptable and the artificial grouping together of quite unrelated or different variables has been deplored[16]. Frequently a variable may measure two different aspects of the same dimension or one variable may be a component of another. An example of the latter would be "Hours spent watching TV" as part of "Total Leisure Hours". Consider what has, in effect, been measured with these two variables.

Let $\underline{T}$ represent the total leisure time, $\underline{W}$ represent time watching TV, and $\underline{U}$ represent leisure time without TV. We can obviously write

$$T = W + U \qquad (1)$$

Suppose that hours spent on leisure activities measures in some way subjective "Leisure", as does hours spent watching TV. Now further suppose that the relative 'leisure-worth' of TV watching is $\underline{w}$, and non-TV is $\underline{u}$ (where w and u are, at present, unknown) and if L is the "value of leisure"

$$L = wW + uU$$

so that, as $\underline{U} = \underline{T}-\underline{W}$ (from 1):

$$\begin{aligned} L &= wW + u\ (T - W) \\ &= wW + uT - uW \qquad (2) \\ &= (w - u)\ W + uT \end{aligned}$$

The weighted combination of $\underline{W}$ and $\underline{T}$ thus is effectively the same as the weighted combination of $\underline{W}$ and $\underline{U}$ (only the weightings differ), and if, in other cases, it is difficult to calculate the equivalent of $\underline{U}$, or many different Us that make up the total value, then it is sensible to let each variable 'find' its own level, through a weighting procedure.

To take another instance, if one attempts to measure an index of 'healthiness' or 'affluence' of an area using various variables whose values may measure the same concept, the aim is to combine such values in a way most likely to be correct. The method used is effectively a weighted least-squares.

Combined health indices are also less susceptible to random or sporadic fluctuations than single indicators and would be more valid for use in geographical areas other than those included in this study.

These weighted combinations of variables are produced by the aggregation of socio-economic and medical care indices, on the one hand, and the outcomes indices on the other. Several combinations need to be prepared to test their differing sensitivities. By the use of regression on the factor scores for health outcomes by the factor scores for characteristics of the individual, the environment and the medical care system, expected values of the areas (H) given the values of the particular variables in the S set, can be predicted.

RESULTS OF PILOT STUDY.

As a test of this approach, a pilot study with very limited objectives was undertaken. The results are preliminary and are presented only to demonstrate the method.

Two groups of 25 outcome indicators and 25 socio-economic and medical care variables were selected according to their relevance in previous work and their theoretical importance. These variables were assumed to represent some dimensions of medical care outcomes and socio-economic characteristics of the patients. Environmental data were not available for inclusion in the preliminary study. Value judgments, based as far as possible on previous research findings, have also been included in the first selection of the variables, and in the weighting (e.g. the relative weighting of immediate admissions or case fatality rate for the first 48 hours for particular diagnoses). (The list of these variables is presented in Appendix A). For these variables, the 15 regions of England and Wales were chosen as the units of analysis since they were the general groupings used for presenting published hospital data; further, the 15 regions provided a sufficient number of units of analysis and events; however, it was realised that other sources of data (especially socio-economic information) would have to be manipulated (aggregated, divided, etc.) in order to use the desired regions. Once the data about each of the 15 regions were obtained, the regions were ranked from highest to lowest and these rankings converted to z scores [17,18] and these z scores were used in the analysis; this technique was preferred because it did not require an analysis of the variance within a region and also because we recognized that the rank ordering was more important for our preliminary work than the actual numerical values.

Using the factor loadings from the factor analysis, the scores for the hospital regions can be seen in Table 1. The factor analysis of the intercorrelations between the elements of HI revealed that we were only able to isolate one element (h) from the set of H. We believe that this is due to the small number of variables and their heterogeneity;

h is probably a global outcome indicator. At the same time, and probably for similar reasons, only one element (s) was isolated from the set S.

(Table 1)

A high score on Outcome (H) means that this area scores better on medical care outcomes and a high score on characteristics of the population and medical care system (S) means that this area scores better on social status and medical care facilities. The scores, converted to ranks, where highest H and S rank 1, is shown in Table 2. (It should be noted that the numbers presented in Table 1 have no meaning in themselves, rather, it is the difference between areas that is important.)

The Pearson correlation between Outcome (H) and Characteristics of the population and the medical care system (S) was 0.69; the Spearman rank correlation between H and S was 0.72, suggesting that the values were almost normally distributed.[19]

(Table 2)

Two groups can be discerned. This division of England into two different areas, North and South, has been a frequent finding in socio-economic studies and reinforces the validity of the analysis developed.

Given areas with equal rankings on social status and medical care facilities, one would expect (all things being equal) equal rankings on outcome. Figure 2 shows the difference in rankings; a positive difference means that, in our study, an area has a better medical care outcome than one would expect, or a worse outcome if the difference is negative.

(Figure 2)

These suggest that areas such as Leeds, Oxford, and East Anglia are doing far better than one would expect on the basis of their combined social status and medical care facilities. Areas like Liverpool, South East, and South West Metropolitan, on the contrary, are much worse than one would expect. No attempt has yet been made to explain these differences.

DISCUSSION

As noted before, this paper presents only some preliminary results of the study of health indicators; further and more difficult steps yet to be taken are:

- The construction of other combined health indices; it is obvious that the same indices of health are not applicable to different groups of the population at different times and places;
- the determination of the indicators most sensitive to medical care variation and those more sensitive to socio-economic and environmental factors;
- the verification that our observed indicators (HI and SI) adequately represent the "true" or underlying factors (H and S). This verification may prove to be especially difficult, and it is not yet clear whether it will be possible to develop a set of relationships in health care similar to the work in input-output analysis economics.

It should also be stated that these results remain preliminary for several reasons:

- Environmental variables were unfortunately not available for inclusion in this first analysis, although we had hypothesized their influence on health outcomes [14]. In fact, we have only been able to obtain three measures of environmental conditions for further analysis: smoke concentration, sulphur dioxide concentration, and solid fuel consumption per household. This important type of information is woefully lacking by comparable geographic or administrative areas. Information on climate, water supply, sewage treatment, air and noise pollution are needed before a satisfactory analysis of their impact on health outcomes can be made.[20]

- There are several severe limitations on the data that we used in the analysis: regularly published data related to health and health status were not generally available in the units of analysis that we desired; two examples are especially relevant:

(1) Data for the London area were only available in aggregated form despite the fact that this area falls into four hospital regions; lacking a better means of approximation, a quarter of the aggregated values was attributed to the respective hospital regions.

(2) Information about teaching hospitals was not as accurate as one would desire. It is presented in many different forms: for instance, covering London and Provincial Teaching Hospitals; or Regional Hospital Boards and their associated Teaching Hospitals; or sometimes merely "Teaching Hospitals". When a method of apportioning values to respective hospital regions was considered justifiable, this was done; however, in the case of the London Teaching Hospitals this approach is not entirely satisfactory although it is probably sufficiently accurate for the "ranking" of the hospital regions.

- In addition, the data used to reflect hospital care and utilization were episode, visit or "body counts" (as in the Hospital Activity Analysis and Hospital In-Patient Enquiry in use in the U.K.) rather than being linked to actual patients. The health status of an individual is a dynamic phenomenon which varies on a continuum and cannot easily be evaluated by cross-sectional data. Further, most of the data presented are comparisons between population groups rather than the characteristics of particular individuals. Of course, the research would have been much more sound if we could have obtained this patient-linked data. However, we were limited by the data available and, unfortunately, these data were more representative of the purposes of hospital management committees than of health care evaluation. There are also limitations due to the type of analysis used, which has not yet been fully explored in community health research. Multivariate techniques, although very appropriate when dealing with multifactorial phenomena, are often feared and with justification, because of the complexity of the procedures involved.
- No attempt was made to explain differences between the hospital regions. Whether they are due to the existence of implicit social criteria that produced a differential resource allocation long ago, and are still maintained in the present circumstances, remains to be seen. More and better data, especially about not only the quantity but also the quality of resources, must be analyzed before an interpretation can be provided.

Finally, it is most important to realise that health is not fully describable in terms of morbidity, mortality, and disability; other factors - though not easily quantifiable - such as quality of life, need to be recognized as components of health, and it is probably in these new dimensions that the way forward lies. To be useful, our indicators must be simple, preferably based on the statistics which are routinely published, and easily calculated by its users.

However, if it does no more than identify a number of outcomes which are specifically sensitive to medical care and eliminate those more sensitive to socio-economic and environmental influences, this type of information can assist and perhaps improve the existing methods of allocating medical and allied resources, and aid in future planning.

We are extremely grateful for the constructive and constant criticisms made by Professor E. Maurice Backett, whose thinking originated this work.

REFERENCES.

1. Bush, J.W., M.M. Chen and D. Patrick. Health Index Project (La Jolla, California : Muir College). Personal communication, undated.

2. Bush, J.W., M.M. Chen and J. Zaremba. Estimating health program outcomes using a Markov equilibrium analysis of disease development. Am. J. Pub. Health 61 : 2362 December 1971.

3. Chiang, C.L. An Index of Health, Mathematical Models. Vital and Health Statistics, PHS No. 1000, Series 2, No. 5. National Centre for Health Statistics, Washington, 1965.

4. Culyer, A.J. Indicators of Health : An Economist's Viewpoint. Institute of Social and Economic Research, University of York, undated.

5. Grogono, A.W. and D.J. Woodgate. Index for measuring health. Lancet 2 : 1024 1971.

6. Sanazaro, P.J. and J.W. Williamson. End results of patient care : a provisional classification based on reports by internists. Med. Care 6 : 123 February 1968.

7. Sanders, B.S. Measuring community health levels. Am. J. Pub. Health 54 : 1063 July 1964.

8. Sullivan, D.F. Disability Components for an Index of Health. Vital and Health Statistics, PHS No. 1000, Series 2, No. 42 National Centre for Health Statistics, Washington, 1971.

9. Swaroof, S. and K. Uemura. Proportional mortality of 50 years and above : a suggested indicator of the component "Health, including demographic conditions" in the measurement of levels of living. Bull. World Health Organisation 17 : 439 1957.

10. Donabedian, A. Promoting quality through evaluating the process of patient care. Med. Care 6 : 3 May 1968.

11. West, P.A. Allocation and equity in the Public Sector : the Hospital Revenue Allocation Formula. Applied Economics 5 : 3 1973.

12. Buckatzsch, E.J. The influence of social conditions on mortality rates. Population Studies 1 : 227 1948.

13. Shah, F.K. and H. Abbey. Effects of some factors on neonatal and postneonatal mortality : analysis by a binary variable multiple regression method. Milbank Mem. Fund Quart. 49 : 33 January 1971.

14. Martini, C.J.M. Health Indices Sensitive to Medical Care Variation, Vol. 1, p. 4. Department of Community Health, Nottingham University, 1972.

15. Levine, D. and D. Yett. A method of constructing proxi measures of health status. In R. L. Berg (ed.), Health Status Indexes p. 12.

16. Allen, D. Social indicators. Social Sci. Res. Newsl. p. 13 November 1968.

17. Fisher, R.A. and F. Yates. Statistical Tables. Edinburgh : Oliver and Boyd, 1970.

18. Phillips, L.D. Bayesian Statistics for Social Scientists. London : Nelson, 1973.

19. Diem K. and C. Lentner (eds.). Geigy Scientific Tables, p. 180, Basle : J.R. Geigy S.A., 1970.

20. Schwing, R. and L. McDonald. Measures of association of some air pollutants, natural ionizing radiation and cigarette smoking with mortality rates. Research Publication GMR - 1573. Michigan : General Motors Corporation, 1974.

Figure 1:

The observed Relationship of Theoretical and Observable Characteristics of the Individual, the Environment, and the Medical Care Process.

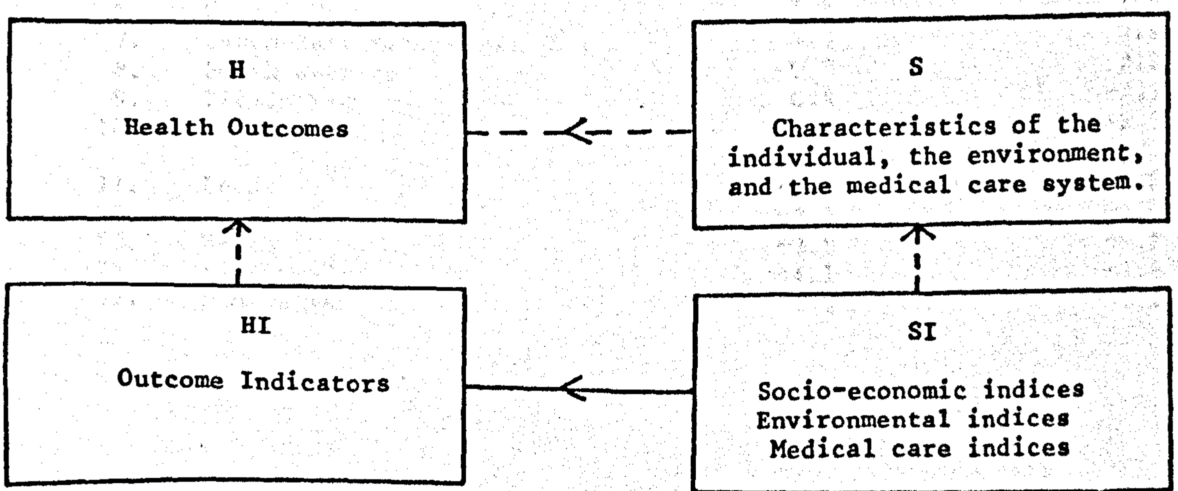


Table 1: Comparison Between Region's Scores

	Area	Raw score of Health Outcomes (H)	Raw score of Characteristics of the Population and Medical Care System (S)
1.	Oxford	5.3	1.8
2.	East Anglian	4.0	-0.9
3.	North West Met.	3.2	5.4
4.	North East Met.	2.5	2.6
5.	South West Met.	2.2	6.2
6.	Wessex	1.6	1.5
7.	South East Met.	1.5	2.9
8.	South Western	0.5	2.2
9.	Birmingham	0.4	-2.3
10.	Sheffield	-1.2	-2.8
11.	Leeds	-3.3	-4.8
12.	Newcastle	-3.4	-3.0
13.	Welsh	-3.5	-4.1
14.	Liverpool	-4.1	-1.8
15.	Manchester	-5.8	-3.1

Table 2: Comparison Between Region's Ranks

Area		Rank of Health Outcome (H)	Rank of Characteristics of the population and Medical Care System (S).
1.	Oxford	1	6
2.	East Anglian	2	8
3.	North West Met.	3	2
4.	North East Met.	4	4
5.	South West Met.	5	1
6.	Wessex	6	7
7.	South East Met.	7	3
8.	South Western	8	5
9.	Birmingham	9	10
10.	Sheffield	10	11
11.	Leeds	11	15
12.	Newcastle	12	12
13.	Welsh	13	14
14.	Liverpool	14	9
15.	Manchester	15	13

Figure 2:

<u>Ranking of 15 Regions in terms of Observed and Expected Health Outcomes.</u>

	7	
	6	East Anglian
	5	Oxford
Health Outcomes	4	Leeds
Better than	3	
Expected	2	
	1	Sheffield, Wessex, Welsh, Birmingham.
	0	Newcastle, North East Met.
	-1	North West Met.
	-2	Manchester
Health Outcomes	-3	South Western
Worse than	-4	South East Met. South West Met.
Expected	-5	Liverpool
	-6	
	-7	

APPENDIX A.

1. The following were included as Socio-economic and Medical Care Variables (SI).

(At this stage there was no attempt to classify them according to these headings):

- Population, total number
- Lower quartile age, total population
- Median age, total population
- Upper quartile age, total population
- % of students remaining in school after school leaving age
- Households per car
- Average rateable value of households in pounds
- Average daily number of available beds per 1000 population
- Discharges and deaths per 1000 population
- Medical - allocated beds per 1000 population
- Surgical - allocated beds per 1000 population
- Obstetrical and General Practice Maternity - allocated beds per 1000 female population 15-44
- Obstetrical and General Practice Maternity - live and stillbirths in hospital as % of total births
- Total outpatient attendances per 1000 population
- Teaching hospital discharges and deaths as a % of total discharges and deaths
- Hospital manpower whole time equivalents per 100,000 population
- Health and Social Services expenditures per capita
- Current revenue expenditures of hospital regions per capita
- Local authority expenditures on health per capita
- Executive Councils - expenditures in General Practice
- General Practitioners median age
- General Practitioners average list size
- % of general practitioners in group practices
- Mean post-natal stay after delivery
- Median length of stay in hospital for all diagnoses
- Median waiting time before hospital admission for all diagnoses

2. <u>The following variables were included as Outcome Indicators (HI)</u>

- Infant mortality rate per 1000 live births
- Neonatal mortality rate per 1000 live births
- Stillbirth rate per 1000 total births
- Perinatal mortality rate per 1000 total births
- Infant environmental death rate per 1000 total births*
- Infant death rate per 1000 live births for enteritis, diarrhoea
- Infant death rate per 1000 live births for pneumonia
- Infant death rate per 1000 live births for congenital anomalies
- Infant death rate per 1000 live births for accidental mechanical suff
- Inception rate (certified incapacity for work) per 1000 males at risk
- Days certified incapacity per 1000 males at risk
- Crude death rate per 1000 home population
- Death rate per 1000 males - all ages
- Death rate per 1000 males 65-74
- Death rate per 1000 females - all ages
- Death rate per 1000 females 65-74
- Death rate for acute myocardial infarction, males 55-64
- Death rate for cerebrovascular disease, males 65-74
- Death rate for pneumonia, females 75 and over
- Discharge rate per 10,000 pop. for all diagnoses
- Case fatality rate per 100 discharges for acute myocardial infarctio
- Case fatality rate per 100 discharges for cerebrovascular disease
- Case fatality rate per 100 discharges for pneumonia
- Immediate admission rate per 10,000 population for acute myocardial infarction

---

* "Infant environmental death rate per 1000 total births" was defined infant deaths due to unexplained prematurity, malformations, antepa haemorrhage and miscellaneous causes for 1000 total births.

APPENDIX G

THE PARTIAL DIFFERENTIAL OF TWO VARIATES

1. Notation

There are two variates x and y. x and y jointly follow a bivariate normal distribution with correlation parameter r. x and y are standardized deviates, so therefore any two values of x and y have a joint probability density f(x,y)

$$f(x,y) = K \exp\left[-\tfrac{1}{2}\left(x^2 - 2rxy + y^2\right)\right] \qquad (1)$$

where K is a constant. Instead of working with f(x,y) I will work with $g(x,y) = \log(f(x,y))$

$$g(x,y) = k - \tfrac{1}{2}(x^2 - 2rxy + y^2) \qquad (2)$$

2. Problem

If g(x,y) is a constant, what is value of the partial derivative $\frac{dy}{dx}$?

3. Solution

From (2), we can write, <u>if g(x,y) is constant</u> ,

$$x^2 - 2rxy + y^2 = C \tag{3}$$

where C is a new constant. Finding the derivative of (3) wrt x leads to

$$2x - 2ry - 2rx \frac{dy}{dx} + 2y \frac{dy}{dx} = 0$$

which simplifies to

$$\frac{dy}{dx} = (x-ry)/(rx-y) \tag{4}$$

4. Discussion

Equation (4) can be interpreted to read that the partial derivative of y by x, given the joint probability density of these values, is given by the RHS of (4). For cases other than the bivariate normal distribution, the generalization of the solution is simple. Simple, that is, if we are in the possesion of a <u>regular</u> joint distribution.