Abstract. In this paper the history of correlation and regression analyses, both in the discipline of statistics generally and in human geography particularly, is examined. It is argued that correlation and regression analysis emerged from a particular social and cultural context, and that this context entered into the very nature of those techniques. The paper is divided into three sections. First, to counter the idea that mathematics and statistics are somehow outside the social, the arguments put forward by David Bloor and Bruno Latour suggesting that mathematical propositions are socially constructed are briefly reviewed. Second, using the ideas of both Bloor and Latour I turn to the development of statistics as an intellectual discipline during the 19th century, and specifically to the invention of correlation and regression at the end of that period. It is argued that the development of statistics as a discipline and its associated techniques are both stamped by, but also leave their stamp on, the wider society in which they are set. Last, the importation of correlation and regression analyses into human geography which occurred in the 1950s is examined. Following my general social constructionist argument, it is suggested that because of the difference in context the correlation and regression analyses devised in the late 19th century were often inappropriate for mid-20th century spatial science.

A history of regression: actors, networks, machines, and numbers

"...it is a mistake to see statistical theory as a field of knowledge developing simply by its own internal logic and giving rise to necessarily value-free techniques. Rather, statistical theory... is a social, historical and ideological product and not merely a collection of neutral techniques."

Donald A MacKenzie (1979, page 48)

"Statistics are 90% science and half luck."

Yogi Berra

Along with death and furniture, numbers and statistics are often used as bottom-line arguments against social constructionism (Edwards et al, 1995). Dead birds, the thumping of sturdy oak tables, and $2 + 2 = 4$ all represent seemingly incontrovertible claims that reality is real and truth is not relative.

But are they so incontrovertible? Seemingly deceased parrots, such as the Norwegian Blue, may only be "resting" after "a good night's squawk". And smacking the furniture around is persuasive only because it is a powerful piece of rhetoric resting on: metonymy—the one part of the table that is cuffed stands for it all; metaphor—the thump of a hand is like a convincing philosophical argument; and irony—producing a nonrepresented, unconstructed bit of external world is undertaken through a dramatic constructed representation. But that still leaves numbers and statistics. Surely they are the real thing? As nature's own language, they are socially untainted.

The purpose of this paper is to argue that they are not. Numbers and statistics are as much social and cultural products as Palladian landscapes, the sayings of the American...
baseball coach, Yogi Berra, references to a classic Monty Python sketch about dead parrots, and death and furniture. In trying to justify this claim I intend to examine the history of correlation and regression analyses both in the discipline of statistics generally and in human geography particularly. I will argue that correlation and regression analyses emerged from a particular social and cultural context, and that this context entered into the very lineaments of those techniques. As a consequence, statistics can be viewed as a social and cultural artifact. Clearly this is not the only way to view statistics, but as I will argue it is a fruitful one in approaching the theme issue of geography's use of quantitative methods from a social and cultural perspective.

The paper is divided into three sections. First, to counter the idea that mathematics and statistics are somehow outside the social I will briefly review arguments put forward by some sociologists of scientific knowledge. In particular, I draw upon first, David Bloor's work to argue that rationalist justifications of mathematics and statistics do not hold up because both are forms of social practice, and second, Bruno Latour's writings to define the nature of those practices. Second, using the ideas of both Bloor and Latour I turn to the development of statistics as an intellectual discipline during the 19th century, and specifically to the invention of correlation and regression at the end of that period. I argue that the development of statistics as a discipline and its associated techniques are both stamped by, and also leave their stamp on, the wider society in which they are set. Last, I examine the importation of correlation and regression analyses into human geography, which occurred in the 1950s. Following my general social constructionist argument, I suggest that because of the difference in context the correlation and regression analyses devised in the late 19th century were often inappropriate for a mid-20th century spatial science.

Towards a social constructionist view of mathematics and statistics

Social constructionism

I begin with the nature of a social constructionist approach to mathematical and statistical knowledge. Notice that I say the sociology of knowledge rather than a sociology of mathematics or statistics. Though there are many very good historical sociological studies of mathematics and statistics, and of individuals working within each discipline, to remain at only the level of describing the broader social and cultural context, including institutional arrangements, career moves, and even the class location and sexual orientation of individuals, is to prosecute a Mertonian sociology of science (Merton, 1973); one that, while recognising the significance of external social relations for individuals, never connects those relations to the production or uses of mathematical and statistical knowledge itself. As a result, the truth of mathematical and statistical knowledge remains unsullied by the often fractured and fractious contexts that form its origin and media of mobilization. More generally, in a Mertonian approach to the sociology of science, a sharp distinction is drawn between the external context of discovery and the internal context of justification. So, although tales of Adolphe Quetelet's travels to Britain in the first part of the 19th century (Hilts, 1978), Francis Galton's private wealth (Pearson, 1914–30), or the enmity between Karl Pearson and Ronald Fisher (Fisher Box, 1978) are interesting stories, they remain just that, stories that have nothing to do with the nature of the mathematical and statistical inventions associated with them.

In contrast, the strong social constructionist approach to mathematics and statistics that is pursued in this paper argues that the context of discovery is inseparable from the context of justification. In laying out this form of social constructionism I draw together two different but related literatures in the sociology of scientific knowledge: the first is associated with Bloor and the Edinburgh School, and the second with
Latour's work on actor-network theory. There are certainly major differences between these two authors, but they are also a useful combination. For on the one hand, Bloor, drawing primarily upon Ludwig Wittgenstein, clears the ground by demonstrating that there are no rationalist foundations for mathematics and statistics; both are just different forms of contingent social practice. Whereas, on the other hand, Latour's concern is with unravelling the nature of those contingent practices. For him, mathematics and statistics are deeply enmeshed in wider webs of social power, control, and authority.

David Bloor and the Edinburgh School

The nub of Bloor's general argument is that knowledge reflects the various social interests of those who propose it: if 19th-century professors at Edinburgh University believed in phrenology it was because of their class location (Shapin, 1975); if 20th-century medical researchers believed that sperm were intrepid, forceful, and resolute it was because of the power of masculinist thought (Harding, 1986); and if Walter Isard believed in the practical efficacy of regional science then it was because he shared a peculiarly 1950s white, masculinist 'can-do' American mindset (Barnes, 1996, pages 130-136). In each case, the world is constructed not as it is, but according to the dictates of the social interests of those putting forward such representations.

For Bloor the corollary of a social constructionist position is a deep scepticism towards any form of rationalism, defined as the belief in a set of universal, asocial logical rules that if followed lead to the truth. For if social interests go all the way down, as Bloor suggests, then there can never be any method that lies outside of them. To show this Bloor takes on what he considers to be the hardest case, mathematics, still conceived as the purest, most distilled form of a universal, asocial logic. If Bloor can show that mathematical knowledge is soaked with social and cultural assumptions, biases, and imperatives, then the softer sciences, and a fortiori the social sciences, must be at least as sodden.

For the sake of brevity it is not possible to provide a comprehensive review of Bloor's meticulous arguments (see Barnes, 1994). Instead let me provide three brief illustrations to give the gist.

First illustration: Bloor is unremitting in his attack on foundationalist justifications of mathematical knowledge (see also Kitcher, 1990). This is necessary because if Bloor is to maintain his position that doing mathematics is a contingent social practice like any other then there must always be something wrong with any argument that justifies mathematics as more than such a practice. A useful example here is Bloor's dismantling of Platonism which posits that the truths of mathematics are guaranteed by some shadow world whose faint traces we copy down when undertaking arithmetic. With reference to Wittgenstein, Bloor shows that such a justification is circular: to claim, for example, that $2 + 2 = 4$ because it preexists in some shadow world is to assume already that that shadow world is right. We can only believe it is right, though, provided we already know the correct answer, but it was the correct answer that the shadow world was supposed to reveal. This is only one instance, but it illustrates Bloor's general strategy of turning foundationalist claims in mathematics against themselves. They always fail because mathematical truths rest not on epistemological but sociological reasons. "The compelling force of mathematics", writes Bloor, "derives from being accepted and used by a group of people. The procedures are not accepted because they are correct or correspond to an ideal; they are deemed correct because they are accepted" (1983, page 92).

Second illustration: Bloor provides a sociological interpretation of mathematical proofs. Typically proofs are conceived as a mechanism by which mathematical truths are justified. Defined as "a sequence of statements such that every member of the
sequence is either a basic a priori statement or a statement which follows from previous members of the sequence in accordance with some a priori preserving rule of inference.” Proofs are a technology for generating truths and rejecting errors (Kitcher, 1983, pages 37 – 38). In this standard view, the validity of a proof is ensured by its asocial cast-iron logic of inference which is there from the moment of a mathematical theorem’s conception. In contrast, Bloor along with a number of others, argues that the irresistible force of a proof is a result of it being a social product (Goodman, 1979; MacKenzie, 1993, pages 55 – 56; Restivo, 1992). Sal Restivo (1992, page 145), for example, writes: “The testing and refinement, the acceptance and rejection of proofs brings other mathematicians into the picture to be a ‘collective’ witness, and to make the question of the proof a collective performance. In the end, if the proof is accepted, it is a result and validation of knowledge under the dictum: ‘It is not I who says this, but all of us’.”

Proofs are not without point, though. But their usefulness should be seen in sociological terms. Proofs should be viewed as a type of theory which reflects the context of those who put them forward, rather than the singular unfolding of some immanent nub of logic. For example, in a brilliant historical account, Imre Lakatos (1976) shows how the proof of Euler’s theorem—that the number of vertices of a solid object minus the number of edges plus the number of faces always equals two—was dickered over for more than a century (see also Bloor, 1990, chapter 7). There was no single seed of inexorable logic within Euler’s initial theorem that proved its rightness; rather the proof was subject to conflict, disputes, and redefinition according to the changing social practices and customs of the time. It was those changing practices and customs that defined what counted as adequate proof, not an invincible logic.

Last illustration: Bloor asks can there be an alternative mathematics? For example, 2 + 2 = 4 seems so self-evident that it is almost impossible to imagine what it would mean to doubt it. But it has been doubted. Winston Smith in his interview with O’Brien is told that in the 1984 world of Big Brother 2 + 2 = 5. What Orwell suggests here are worlds so different that the self-evident truths of one cannot be carried over to another. Bloor argues the same point, albeit in less dramatic terms. For him, “certain conditions have to obtain before a computation has any meaning. These conditions are social in the sense that they reside in the collectively held system of classification and meanings of a culture. Consequently they will vary and as they vary so will the meaning of pieces of mathematics” (1990, page 124). Bloor (1990, chapter 6) provides a number of examples, such as the varying meanings associated with root two, infinitesimals, and the number ‘1’. Likewise, Helen Watson (1990) in her study of the Yoruba people of West Africa shows that, because of the very nature of their linguistic system which does not make a distinction between separate objects of the same kind, natural numbers mean something quite different to them than to us in the West, although both sums add up to the same result. In each of these examples, the alternative meanings are not wrong, but predicated on radically different assumptions that either are no longer held or are not present in Western culture. They represent an alternative mathematics, worlds where 2 + 2 might equal 5.

The usefulness of Bloor’s work is primarily a critical one. He shows over and over again that rationalist justifications for mathematical truths do not hold up. Ideas thought to be sealed from the social and the cultural turn out at critical points to depend upon them. There is no going around about; the great Boyg of culture and society blocks our way at every point and at every number.

Bruno Latour and actor—network theory

Though Bloor is persuasive in the abstract about both the problems of rationalism and the general influence of social interests on knowledge, he is less helpful in marking the
A history of regression

207

often contorted and convoluted processes, material and nonmaterial, involved in both fixing that knowledge and generating those initial social interests. Here Latour's work is germane. Whereas Bloor is concerned with providing a complete analysis of events once all the dust is settled, cleaning and tidying them up with forensic precision, care, and patience, Latour is mainly interested in events while the dust is still flying. For it is then that the various networks of resources, allies and actors that are responsible for all the commotion, and which for Latour are the stuff of both science and society, are best glimpsed.

The problem of science for Latour is one of creating order. This is done through marshalling an array of resources: financial, technical, institutional, conceptual, intellectual, and so on. Those resource flows are held temporarily in place—that is, made to work together in particular kinds of ways—in the form of a network. For example, the British localities project in human geography was a classical network in Latourian terms, absorbing large sums of money from the ESRC, relying on all kinds of technologies in the assembly, communication, and dissemination of results, involving the cooperation of various institutions and groups, drawing upon a range of existing concepts such as Doreen Massey's spatial division of labour thesis, and resting upon countless hours of labour by researchers and staff at various levels of skill, experience, and rank. For Latour if the network can be extended, and made stable, then eventually a black box is formed; that is, the various processes, and the interconnections among the parts of the network, drop out of view and what is left is only the final product: the double helix, a computer, the Pearson product moment correlation coefficient, or an edited book called *Localities* (Cooke, 1989). Once they start circulating, however, those black boxes quickly become components of new networks that then redefine existing social relationships and new kinds of knowledge.

It is this processal view of Latour's that provides the basis of the second gloss on social constructionism: namely, that the processes involved in acquiring knowledge within a network help construct the very nature of society itself. John Law (1994, chapter 1) says that this view is one where there are only verbs and no nouns; processes are both medium and outcome. This does not seem quite right, though, because certain things, however temporarily, remain fixed, for example, Latour's black boxes, or Bloor's social interests. But they are not fixed for ever. The emergence of new networks continually undermines prevailing black boxes and social interests, producing new kinds of both. The social constructionism attempted in this paper will try to include both nouns and verbs. I return to this issue below.

Latour's primary interest has been the networks that branch out from science, although more recently he has become interested in the nonscientific kind as well (Latour, 1993; see also Law 1994). Critical to any scientific network, Latour suggests, is the role of inscription. At the heart of science is writing because without inscription the networks on which science constitutively depends could not be organized, maintained, or extended. One particular form of inscription is mathematics. As a special kind of writing it organizes, maintains, and extends networks in particular ways, and therefore, following Latour's thesis, the nature of society itself.

Inscription is critical to the organization of networks because it is the first act in what Latour calls translation; that is, converting something unmanageable in its original form into something that can be controlled and shaped within a network. For, when inscribed on a page, phenomena, as Latour (1990, page 45) writes, can be "... dominated with the eyes and held by the hands, no matter when and where they come from or what their original size". Furthermore, once dominated the phenomena are then easily manipulated, shuffled, and recombined to create new patterns and structures. "Levi-Strauss' theories of savages", says Latour, for example, "... are an artifact of card
indexing at the College de France” (1990, page 46). The broader point is that scientists only “…start seeing something when they stop looking at nature and look exclusively and obsessively at prints and flat inscriptions” (page 39).

Given this context, mathematics is an especially useful form of inscription. As a language it is especially suited to domination, control, manipulation, and transformation. It dominates by imposing particular characteristics onto phenomena. In order to be counted, for example, phenomena need to be standardized according to some attribute and also to be distinct and separate. Moreover, such standardization is not innocent; if successful it redefines the very way we look at the world. Mary Poovey (1993, page 275) writes because “…of the categories by which statistical representation organizes materials, …meanings are…constructed before the statistics are compiled; they then radiate from the starkest tables”.

Internal control of its practitioners is achieved by the shared rigid and highly formalized nature of mathematical discourse. As Ted Porter argues, the authority of quantitative methods derives from a lack of alternatives: “power minus discretion” as he puts it (1992, page 642). We seemingly have no choice but to apply mathematical and statistical standards to us and to others, not because of any direct dictatorial power (except, perhaps, in the case of an overbearing mathematics teacher), but because that is the way it is done. We maintain autonomy only in so far as we maintain good numbers (Porter, 1992, page 643).

Manipulation and transformation are central attributes of all branches of mathematical inquiry, but are best crystallized perhaps by the equation. Equations shuffle diverse phenomena into new relationships, define new equivalences, summarize the consequence of new combinations, and express and measure the resistance of new associations to disruption. In short, the power of an equation is in drawing different things together. In part, we accept, say, the multiple regression equation because it unifies on one line many different things and relationships.

The maintenance of networks is achieved through a process of gaining new allies, defined as “…convincing someone else to take up a statement, to pass it along, to make it more of a fact, and to recognise the first author’s ownership and originality” (Latour, 1990, page 24). Convincing new allies, however, often rests on rhetorical gestures that are in turn bound up with inscription; “inscription allows conscription” (Latour, 1990, page 50). To any doubter one simply says “I’ll show you’. And, without moving more than a few inches, [one] unfold[s] in front of [their] eyes figures, diagrams, [and] texts” (page 36). As one accumulates more and more allies, so the rhetoric of representation becomes increasingly more brazen: qualifiers, scare quotes, and citations rapidly disappear leaving only an objective fact or idea. No longer, for example, “Most of the work…[applying to geography] methods and techniques of modern science…has been frankly exploratory and the findings somewhat tentative” (McCarty et al, 1956, page iv). But rather, “…any branch of geography that has a need for theory has a need for quantitative techniques” (Burton, 1963, page 159). Through garnering more fruitful allies objectivity is slowly erected. Furthermore, once erected it is extremely difficult to dismantle. Not only does one require massive amounts of resources to mount any kind of challenge, but one has to undo the framework of existing allies that helps support it. As Latour says (1990), opposition is squashed and objectivity made stronger through a process of making dissenters feel lonely.

The role of mathematics in currying favour with allies is particularly important. It is maybe the most successful of all rhetorical tropes, lending instant credibility to those who use it. To any doubter one simply performs the proof, takes out the textbook, or talks about a 10 000-year history of achievement. In this way, mathematics is very difficult to contest not because
it is inherently correct but because of the immense amount of resources that have gone into it, and the way in which we have been socialized into its mode way of thinking. As Latour (1990, page 52) says, "it is in... the training that the anthropology of geometry and mathematics [is] decisive".

Last, extending networks is for Latour an inherently spatial problem. Typically it is not an issue usually discussed because of the belief that once a universal truth is found it will be universally taken up. The history of science, and geography as well, though, is one in which research takes place in very specific local sites: Latour (1987) calls them "centres of calculation"—places such as the university, research laboratories, R&D (research and development) facilities, museums, and archives. With the large resources found at those sites, centres of calculation are places where domination, control, manipulation, and transformation occur, and hence are also the beginning points for network organization and maintenance. How, though, are specific local findings given a wider power and effect? First, it requires what Latour calls "immutable mobiles"—forms of communication that can travel across space, and which do not corrupt the message: maps drawn on the sand are neither immutable nor mobile, but maps found in the Rand and McNally Atlas are both. It is not, however, only the physicality of immutable mobiles that are important, but also their inscriptive forms. Here again mathematics has advantages. Precisely because it is a language of set rules and open to scrutiny, it is, as Porter (1992, page 644) says, "a superb technology of distance", one that "...spans continents, leaps oceans, and links laboratories, factories and governments".

The universality we think we see, then, is only the rhetorical success of particular centres of calculation diffusing their results along extended networks by immutable mobiles. Along with this success, of course, goes power: the power to act and control at a distance. Latour (1987, page 245), in a long passage worth quoting in full, says:

"When people wonder how 'abstract' geometry or mathematics may have some bearing on 'reality', they are really admiring the strategic position taken by those who work inside the centres [of calculation] on forms of form. They should be the weakest since they are the most remote... from any application. On the contrary, they may become the strongest by the same token as the centres end up controlling space and time: they design networks that are tied together in a few obligatory passage points. Once every trace has been not only written on paper, but rewritten in geometrical form, and rewritten in equation form, there is no wonder that those who control geometry and mathematics will be able to intervene almost everywhere."

In sum, whereas Bloor argues that rationalist justifications of mathematics are inadequate, suggesting instead that it must be conceived as a social product, Latour provides a theoretically informed account of the kind of social product it is. As a special form of inscription, mathematics participates in the organization, maintenance, and extension of scientific networks. Such networks for Latour are not simply the precipitates of more fundamental social relations, but they enter into their very definition. Mathematics shapes society as much as it is shaped by it. More generally, the strong form of social constructionism proposed here attempts to be similarly recursive. It involves both reading back from nouns such as social interests and black boxes to the network processes in which they gain meaning, while at the same time also reading forward to the kind of network processes those interests and black boxes help create.

A social history of statistics

Let me now try to apply these general arguments to the discipline of statistics, and in particular to correlation and regression analyses. Specifically, by tracing out the early history of both I will argue after Bloor that they represent the social interests of those who propose them, and after Latour that they are critical forms of inscription
facilitating the organization, maintenance, and extension of various kinds of networks that then participate in the constitution of society.

**Early history**
The origins of the term statistics are with the late 18th-century German descriptive science of nation-states, *Statistik* or *Staatenkunde* (Porter, 1985, page 54). States, of course, have been in the business of gathering numbers since the Book of Numbers, and likely before that. The interest in numerical information, though, snowballed in Continental Europe from the early 19th century onwards. Facilitated by an increasing use of immutable mobiles, the result was, as Hacking (1990a) puts it, an “avalanche of printed numbers” and corresponding “fetishistic tabulations”: Those numbers and tabulations showed an astounding regularity across a range of social statistics: suicides, murders, hangings, and even the number of dead letters in Paris's central post office remained constant even though everything else around them was in flux.

The tabulating fetish also crossed the English Channel. The Statistical Society of London, later the Royal Statistical Society, was inaugurated in 1834, albeit not without some resistance from the British Association for the Advancement of Science (Hilts, 1978; Poovey, 1993). In this formative period only numbers mattered for Society members. In fact, the seal of the Statistical Society of London was that of a wheatsheaf with the inscription “aliis exterendum”: to be threshed by others. The task was only to collect, as one critic at the time put it, the “... dull, dry parade of... figures” (Taylor, quoted in Poovey, 1993, page 260), not to analyze or draw conclusions from them.

But of course it was impossible not to do so. With reference to Latour, the very inscription of numbers and later their joining to probability calculations within a burgeoning set of both commercial and statist networks produced worlds that were to be organized, controlled, and manipulated. There are several points to be made here.

The first is that there was something about the emergence of industrial capitalism and the nation-state that made statistics an appropriate style of thought for the times. For Raymond Williams the critical change was capitalism's growing complexity, size, and unmanageability. He says, “It is very striking that the classic technique devised in response to the impossibility of understanding contemporary society from experience, the statistical model of analysis, had its precise origins within the [19th century]. For without the combination of statistical theory... and arrangements for collection of statistical data a society that was emerging out of the industrial revolution was literally unknowable” (1982, page 170). There is thus something about the very nature of capitalist industrialization and its relationship to the state that drove the formation of networks, and through statistical inscription they were able to maintain and extend themselves. Karl Marx said that doublesided bookkeeping was one of the greatest inventions of capitalism, and likewise the census may be one of the greatest inventions of the state. Keeping track of the numbers helped workshops become the workshops of the world, and the British Foreign Office in Whitehall to rule half the world.

This is not a functionalist argument, however, because, and this is the second point, the very act of mobilizing numbers creates a new world that was not there before. As Talal Asad (1994, page 70) puts it, statistics is “... not merely a mode of representing ... social life but also a mode of constructing it”. This is true in a number of different senses. There is the formation and maintenance of an institutional apparatus charged with collecting and analyzing the numbers. There is the invention of the associated technology of calculation, both physical such as early calculating machines and punch cards, and conceptual such as correlation and regression analyses (Hacking, 1990b, pages 697–698). There are the new categories to organize the numbers collected, which then became the basis for understanding and intervening in the world.
Ian Hacking (1990a, page 8), mischievously asks at one point, for example, whether Marx would have made class the basis of his theoretical system had he not first pored over the tables of English factory inspectors. Last, there is the new reality of the statistical measures themselves. Terms such as the mean, or standard deviation, and later the correlation coefficient become objects in and of themselves, with the same kind of ontological resonance as dead birds and oak tables. In short, the avalanche of number created a new world in the sense that things could be said, sentences written down, that not only would not have been understood earlier, but could not have even been conceived (Hacking, 1992, pages 141–142).

Third, statistical talk led to people having a different sense of themselves in relationship to others and society more generally. Once society is conceived as a statistical distribution, individuals are assessed and assess themselves in relation to the norms of that aggregate. As Hacking (1990a, page 2) writes: “People are normal if they conform to the central tendency of [socially statistical aggregates], while those at the extremes are pathological. Few of us fancy being pathological, so ‘most of us’ try to make ourselves normal, which in turn affects what is normal”. Decisive here was the recognition of the kind of statistical regularities I alluded to earlier, and especially the application to them of a probability distribution, ‘the error curve’ or as it is now known the normal distribution. The error curve had been invented by Abraham De Moivre, applied by Pierre-Simon Laplace, but was best known from Carl Gauss’s work as the law governing the distribution of measurement errors around the true value of astronomical and geodetic observations. The Belgian astronomer cum social tabulator Adolphe Quetelet was the first to apply the idea to humans—in his case, to the chest measurements of more than 5000 Scottish Highland soldiers—and arrived at the notion of l’homme moyen. This ‘average man’ was not simply the arithmetic mean, though, but took on a moral connotation because Quetelet stayed with the language of the error curve and its basis in astronomical measurement. Anyone who did not conform to the average was in some sense in error.

Fourth, and following, once normality is defined as the norm within a wider distribution, there is a clear charter for bureaucratic monitoring and intervention to ensure that it is met. In this sense, “the priority of numbers was a surface effect. Behind it lay new technologies for classifying and enumerating, and new bureaucracies with the authority and continuity to deploy the technology” (Hacking, 1990a, pages 2–3). This resulting intervention could cut two ways, though. At ‘the bad end’ of the distribution there could be an attempt to control the pathological, either by hiding the extremes away, for example, in mad houses or expulsion to Botany Bay, or making them conform closer to the norm through moral reform and education. Conversely, there could be an attempt to move the norm to ‘the good end’ of the distribution, the most extreme example being the use of eugenics.

Fifth, and this brings us back to one of Latour’s central concerns, statistics was a powerful piece of rhetoric in drawing together allies. Clearly numbers connected the different parts of government, branches of the state, various institutions, and quasi-professional associations such as the London Statistical Society. Also interesting here is the way statistics crosses the traditional divide between the sciences and social sciences, cleaving adherents on both sides. Philip Mirowski (1994, page 15) speaks of this “...as a kind of spiral, gyrating back and forth between historically contingent locations of the Natural and the Social, wobbling when the poles themselves shift”. For example, the probability distribution taken up by Gauss to evaluate the accuracy of astronomical observations is projected by Quetelet back onto society as the normal distribution, which is then later picked up by James Clark Maxwell and used to derive his gas laws (Porter, 1981), which yet in turn is utilized by Ysidro Edgeworth to describe the
stability of the market under conditions of uncertainty, and on it goes (Mirowski, 1994, page 15; see also Tribe, 1991, page 417). In this reading the normal distribution is not just a figure of statistics but a figure of speech (Poovey, 1993). Specifically, it is a very pliable metaphor: the distribution of social statistics is like the distribution of astronomical observations which in turn is like the distribution of the velocities of gas molecules. In each case something unfamiliar is turned into something familiar, thereby drawing together adherents and allies.

From this account, it is possible to begin to see the relevance of some of the points made by Bloor and Latour. From Bloor’s perspective, statistics is not the unfolding of an immanent mathematical logic. Rather, statistics emerged as an eminently practical pursuit, reflecting the interests of those who pursued it. But it is not just a matter of statistics reflecting society, but following Latour society being fashioned in the likeness of statistics. To make these points sharper let me examine the history of one particular set of statistical methods, correlation and regression analyses. Both techniques are pivotal historically because: they were the first to combine successfully both probability distributions and numbers; as an advertisement of the power of statistical analysis they drew together allies in such diverse fields as biology, psychology, sociology, economics, and later even human geography; and perhaps most importantly they were the basis for a new kind of explanation that shunned determinism—the statistical law (Hacking, 1987, page 53). And with it, as Hacking (1990a) puts it, chance was tamed.

On correlation and regression

In 1889 the English eugenicist, Victorian polymath, and cousin of Charles Darwin, Francis Galton was taking shelter from a rainstorm during a country walk at Naworth Castle near Carlisle in Northern England, when, as Galton (1908, page 300) tells it, “there the idea flashed across me, and I forgot everything for a moment in great delight”. In one interpretation, at least, what delighted Galton so much on that rainy day ramble was an idea for correlation analysis (this is Pearson’s 1914–30, page 393, interpretation of that event although it is contested by Hilts, 1973, page 233, and Stigler, 1986, page 298). For Galton (1888, pages 135–136) correlation, or ‘co-relation’ as he called it, occurred:

> “when the variation of one [variable organ] is accompanied on the average by more or less variation of the other... It is easy to see that co-relation must be a consequence of the variations of the two organs being partly due to common causes. If they were wholly due to common causes, the co-relation would be perfect... If they were in no respect due to common causes, the co-relation would be nil. Between these two extremes are an endless number of intermediate cases, and it will be shown how the closeness of co-relation in any particular case admits of being expressed by a simple number.”

There had been other attempts to devise measures of ‘co-relation’ before Galton, but none were very successful, and certainly not widely taken up (Stigler, 1986, pages 297–299). What made the difference were the background assumptions in which Galton framed his work, and found in eugenics. This is best elaborated by a historiographical excursus (and one closely informed by Bloor’s 1983, pages 103–109, interpretation).

Karl Pearson, the first Professor of Statistics in Britain, wrote two quite different histories of correlation and regression analyses separated by some twenty-five years. The first in 1895 attributed correlation to the French mathematician Auguste Bravais, a mid-19th-century error theorist, interested in assessing the accuracy of astronomical measurements. For any given point in space, a separate set of measurements could be made for both the $x$ and $y$ coordinates. The overall pattern of error is then given by multiplying together the two independent laws of error associated with each coordinate.
Bravais, however, was interested in the more difficult case where the same set of measurements were used to calculate both $x$ and $y$ together. To do that required him to calculate a joint law of error, the equation of which is remarkably similar to Galton's later equation for correlation.

Specifically, Bravais's formula is:

$$
\frac{k}{\pi} = \exp \left[ -(ax^2 + 2exy + by^2) \right],
$$

where $k$, $a$, $b$, and $e$ are constants to be evaluated. Whereas Galton's equation for the correlation surface of the bivariate normal distribution is:

$$
h(x,y) = \frac{1}{2\pi\sigma_1\sigma_2(1-r^2)^{1/2}} \exp \left\{ -\frac{1}{2(1-r^2)} \left[ \frac{x^2}{\sigma_1^2} - \frac{2rxy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2} \right] \right\}.
$$

Comparing equations (1) and (2) we can see their similarities. They can be translated into one another by writing:

$$
k = \frac{1}{2\sigma_1\sigma_2(1-r^2)^{1/2}},
$$

and finding similar expressions for $a$, $b$, and $e$.

By 1920, however, Pearson (1970) had changed his mind, arguing that it was Galton who was the true originator of correlation because the problem to which Bravais applied his work was completely different from the one Galton was trying to solve. Bravais devised his equation (1) in order to get rid of the amount of statistical variation of error around the true values of the variables, whereas for Galton it was precisely the statistical variation—the error—that needed to be kept. Once explained, the variation could be made the source of intellectual progress: specifically for Galton's purposes geniuses could be bred.

Bloor uses Wittgenstein's work to explain what was happening here. Presented on their own, Bravais's and Galton's correlation equations appear similar, but only because, and this is Wittgenstein's phrase, there is no corresponding "...organism of applications which envelops and connects... them". In this sense, the two equations are like "two bones separated from the surrounding manifold context of the organism in which we are accustomed to think of them" (Wittgenstein in Remarks on the Foundations of Mathematics quoted in Bloor, 1983, pages 103–104). Once reembedded in the prevailing intellectual and cultural context, which is what Pearson does in his later account, it becomes clear that the two equations mean quite different things (incidently also illustrating the point made earlier that proofs are not inexorable but mean different things in different contexts).

So what was the context of Galton's work that made such a difference? In many ways it turns on Galton's interpretation of normality. Whereas Quetelet interpreted the average found at the centre of the normal distribution as a golden mean—an original of which actual individuals are mere tarnished copies—Galton saw the mean as equivalent to the mediocre. Galton's interest was not in celebrating l'homme moyen, but moving beyond her or him; "to get beyond mean values, not to exalt them" (Gigerenzer et al, 1989, page 55). It was this task that eugenics could perform. Galton had already made a theoretical study of Hereditary Genius (1869), and concluded that book by saying that "I do not see why any serious difficulty should stand in the way of mathematicians, in framing a compact formula... to give us... the means of foretelling the average distribution of characteristics among a large multitude of offspring whose parentage was known" (pages 371–373). Without going into the details (for that see
Porter, 1986, pages 280–286, and Stigler, 1986, pages 265–299), the mathematics most suited to the theory of hereditary Galton proposed was one that required variables with well-defined means, with individuals possessing irregular deviations from it but which in aggregate take on a normal distribution. In this way, as Porter (1986, page 286) writes, "...Galton's [theoretical] model [of hereditary] furnished terms of statistical analysis ideally suited to what would become the formulation of regression".

Working out the hereditary relationship first on peas and the size of their progeny, and then later on humans, Galton discovered what he called the law of reversion or regression: the mean for any batch of progeny differed from the general mean of the population in an amount proportional to the displacement of their parents. However, the displacement from the population mean of the offspring was always less than that of their parents, and as a consequence was always reverting or regressing back to the population mean. Such a relationship could be expressed mathematically as a linear equation:

\[ y = rx, \]

(4)

where \( y \) is the average divergence of offspring from the mean of the population; \( x \) the divergence of the parents; and \( r \) is a constant with a value between zero and one. Though this became the basis for the statistical technique of regression, in some sense for Galton it is simply a 'law of hereditary'.

At this point we can begin to see what lay behind Pearson's revisionist historical account. Galton is interested primarily in the deviations around the mean, and not the mean itself. As Galton himself says, "The primary objects of the Gaussian Law of Error were exactly opposed, in one sense, to those which I applied them. They were to get rid of, or to provide a just allowance for errors. But those errors or deviations were the very things I wanted to preserve and know about" (1908, page 305, note 2). Such a statement becomes even more relevant when we turn to Galton's contribution to correlation analysis. For what came to him when he was out in that rain shower in Northern England was the relationship between correlation and regression. The problem precipitating his realization was only indirectly related to his studies on heredity, although as soon as he made the discovery he could see its relevance to that field. Of Galton's many sidelines, one was anthropometricism, the measuring of body parts and assessing their relationship one to another. The French criminologist Alphonse Bertillon had argued earlier in the century that anthropometricism provided one means for classifying criminals: they could and should be identified according to the four measurements of their height, and length of finger, arm, and foot. Though Galton thought that there were better ways both to classify and to control the criminal classes—he invented both fingerprinting and the silent whistle for guard dogs—what he found intriguing in Bertillon's scheme was the possibility of 'entanglements' or correlation between the measurements, for example, tall people likely having longer arms than short people. With the use of anthropometric data collected from his own research, he plotted body height against the length of left cubit, scaling both axes in units of their respective standard deviations. As "the number of marks on the paper grew in number" it suddenly struck him that their form was similar to the one he had found for inheritance (Galton, 1890, page 421, quoted in Porter, 1986, page 292); that is, the problem of correlation was mathematically identical to that of regression.

What was so novel about Galton's work here is not the use of linear approximation but his realization that variation or 'error' could be divided into two parts: a part that could be explained by another variable, and a part that could not (Porter, 1986, pages 293–296). When he carried out his work on inheritance with the use of peas, and later heights of parents and progeny, the partitioning of different types of error had a clear biological interpretation: one part was explained by a 'reversion' back to the population mean,
the other by the inherited characteristics of the parents. The genius of Galton was to recognize that this same partitioning could be generalized and applied to any case where two variables were co-related, and not just to intergenerational transmission of attributes. It is at this point that we first hear the lid of the black box of regression being closed.

By way of a summary let me try to relate Galton's work to the broader conceptual framework set out in the first part of the paper. First, Galton's belief in the possibility of eugenics—his social interest in Bloor's terms—fundamentally affected the statistics that he chose to develop. Correlation and regression thus emerged out of his interest in the variation around the mean, and not like Quetelet in the mean itself. In effect, this was Pearson's realization, which indirectly also gives support to another of Bloor's claims, the importance of 'nonmathematical' background conditions in making sense of a given computation.

Second, Latour's arguments about the importance of inscriptions are also well illustrated. Galton was literally able to see his new truths emerge in front of his eyes: his various measurements of pea sizes, human heights, and body parts were meticulously recorded and then transformed and reshuffled on paper as he applied his statistical techniques. In doing so he created a new tier of reality: the statistical one. The correlation coefficient and the regression line became new bits of being.

Third, all of this was made possible because of the wider network in which Galton was placed involving: money, mainly his own; his heavy involvement with a number of scientists and scientific societies including the Kew Observatory, the Royal Geographical Society, and the British Academy; his use of machines such as the quincunx that showed a normal mixture of different normal distributions was itself normal (Stigler, 1986, pages 275–281); and his reliance on a body of contemporary ideas including those of his cousin, Charles Darwin, Quetelet's error curve which he learnt of from the geographer William Spottiswoode, and John Venn's work on probability (Porter, 1986, page 271).

Fourth, this network was extended and came to affect the very nature of society as the black box of correlation and regression diffused from the centre of calculation. This occurred in various ways: in the institutionalization of statistics, which Galton in effect initiated when he funded Pearson's Chair at University College, London (Gigerenzer et al., 1989, pages 109–122); in the spread of immutable mobiles in the form of scholarly periodicals (Pearson launched both *Biometrika* and the *Annals of Eugenics*); and in the treatment both by governments and by academic disciplines of their respective subjects.

**Human geography and regression**

Geography had the odd early brush with correlation and regression analyses. There was William Spottiswoode’s advice to Galton, and Galton (1872, page 199) himself in a Presidential Address to the Geography Section of the British Academy of Advanced Sciences had said: “The configuration of every land, its soil, its vegetable covering, its rivers, its climate, its animal and inhuman inhabitants act and react upon one another. It is the highest problem of geography to analyse their correlations and to sift the casual from the essential”. Another eighty years or so elapsed, however, before human geography scaled those heights.

**Early history**

When correlation and regression finally came to human geography in the mid-1950s, it was the result of the building of allies, complicated networks, and the linking of disparate centres of calculation. Those first centres were not necessarily the world's metropoles, though.
Iowa City was the backdrop for Harold McCarty’s (1979, page 124) pioneering work. McCarty’s paper (1954a), “McCarty on McCarthy”, which provided a geographical analysis of Senator Joseph McCarthy’s Wisconsin vote, was one of the first applications of correlation and regression analyses by a human geographer [both techniques, however, had been applied earlier by climatologists in geography; see Hooker (1907) and Rose (1936)]. McCarty (1956) followed this up with an abstract in the Annals of the Association of American Geographers in 1956, “Use of certain statistical procedures in geographic analysis”, and a collaboratively written research report on The Measurement of Association in Industrial Geography (McCarty et al, 1956). One of the motivations for McCarty in pursuing this work was to provide a laboratory for economic theory; the associations he found through his correlation and regression analyses would be the empirical grist for the theoretical mill of economics. Specifically, in an early manifesto published in 1954 McCarty (1954b) argued that because of the complexity of geographical phenomena one could not identify causes simply by inspecting the data. The best one could hope for were geographically limited associations that could then be compared with existing theories. The resulting vision of the discipline was later well illustrated in his joint textbook A Preface to Economic Geography (McCarty and Landberg, 1966). There economic geography is metaphorically laid out as if it were stages in a regression equation: classify the data, choose the economic hypothesis to test, select the dependent and independent variables, carry out the correlation and regression analyses, and assess the goodness of fit. It was as simple as that.

Of course, in retrospect it was too simple, and furthermore, it did not even satisfy McCarty’s own quest of making economic geographers the assemblers and testers of economic theory. Put bluntly economists did not need his help. They were quite capable of doing regression analysis on their own, and in many ways have made the technique their’s (see McCloskey, 1985, page 163; and also Mirowski, 1989). The interesting point, though, is that McCarty felt the need to offer up his work to another discipline; that is, he felt the need to look for support and allies outside of his own field to legitimate his methods. In Latour’s term, of course, this is just a variant of the age-old strategy of winning friends and influencing people.

That strategy was plied in a different way in the American Pacific Northwest, and was to influence the next generation perhaps more significantly than McCarty’s work at Iowa (see Berry’s, 1993, account). Under Donald Hudson’s auspices, William Garrison pioneered during the mid-1950s “…the first… advanced courses in statistical methodology…” in the country at the University of Washington at Seattle (Hudson, 1955, page 29). In a 1955 advertisement for the programme Hudson boasts of sending one of the graduate students, Duane Marble, to Stanford’s prestigious Science Institute in Mathematics for Social Sciences, the departmental use of an IBM 604 digital computer, and the achievements of Garrison himself, who had recently returned from special training at the University of Pennsylvania with Walter Isard and William Warntz (Hudson, 1955, pages 28–29). Garrison was also involved in probably the first debate in human geography over the introduction of statistics. Following an anodyne review of the potential of statistics in geography by R B Reynolds (1956), which included a few mild cautionary remarks, Garrison erupted, denying all criticisms and vigorously asserting the power of mathematics. “Statistical methods”, Garrison wrote, “are written in the universal language of mathematics…” (1956, page 428). The intensity of the reply showed that mathematically inclined geographers, if not mathematics itself, were a force to be reckoned with.

In Britain the network of quantification was put together in yet another different way. As an example let me focus on Peter Haggett’s work carried out first at Cambridge and later at Bristol. Haggett’s first published paper that made use of quantitative
A history of regression

techniques appeared in 1961, and drew upon field work carried out in the late 1950s near São Paulo. Despite early resistance to his methods—Sir Dudley Stamp told him “This kind of thing has got to stop” following an RGS (Royal Geographical Society) presentation in which Haggett showed a multiple regression equation—Haggett’s influence in spreading the word about numbers was possibly greater than anyone’s in the discipline (Thrift, 1995, pages 381–382). Leaving aside Haggett’s own manifold abilities, such success was also bound up with his adroitness to extend the network of quantitative geography. Partly this was a result of him being, as Nigel Thrift (1995, page 380) puts it, already a “member of a multiple geographical elite [within British geography].... He was able to mount an assault on the then geographic establishment in part because he was so firmly located within it”. Partly it is a result of his prodigious travelling and the number of contacts he made—he gate-crashed his first regional science meeting at Berkeley in 1962, and later made contact with Hägerstrand in Sweden and geographers in Australasia (Haggett, 1965, preface). And partly because of his influence on the ‘anthropology’ of mathematical training (Latour, 1990, page 52)—with Chorley he was responsible for the Maddingly Hall lectures for secondary schoolteachers out of which both Frontiers in Geographical Teaching (Chorley and Haggett, 1970) and Models in Geography (Chorley and Haggett, 1967) originated, and whose message seeped deeply into the institutional structures of geographical education in England and Wales.

By the mid-1960s, then, the sides of the black box that was to be quantitative geography were beginning to be nailed shut. The network was becoming stable. Peter Taylor’s (1977, page 15) flight plan for the fictional Quantgeog Airlines, as he called it, and which connected human geography’s hubs of calculation, was a perfect visual representation of that network (figure 1). Though it was literally a transportation network—Garrison (1979, page 120) says that “improvements in air transportation made a high level of external contacts thinkable and workable”—it was also an intellectual one too. Actors in that network included the individuals and institutions already mentioned. There are nonhuman inputs such as computers and desk calculators—McCarty (1979, page 123) talks about “…the hum of calculating machines...[in] the department’s...laboratory rooms”—as well as printouts and an increasing number of mimeographed papers and reprints. There are ideas taken from other fields such as inferential statistics, geometry, physics, and economics. And there are the large sums of money from such institutions as the office of Naval Research and the National Science Foundation to pay for the computing costs, the travel to gate-crash conferences, and for sending staff “...to different summer camps for insurgency training in computer methods in geographic research, spatial statistics and regional science” (King, 1979, page 127). In short, an array of heterogeneous elements was organized to form a network that could be then maintained and extended. It was this network that formed the basis of

Figure 1. Quantgeog Airlines flight plan (source: redrawn from Taylor, 1977, page 15).
what Ian Burton announced as the quantitative revolution, and not, as he originally suggested, some inherent rigorosity of the methods themselves or their innate ability to ward off self-deception.

Although the network certainly consisted of these various nouns, it was also a verb, a process, something in flux. This of course is easiest to see in the earlier stages, before the sides of the box go up. At that point it is not clear that techniques such as correlation and regression analyses will work out. It is then that there is the greatest bravado by advocates (witness Garrison’s over-the-top reaction to Reynolds); the greatest skepticism (“the sharp inhalation of breath” by some RGS members at Haggett’s lecture; Thrift, 1995, page 381); the greatest search for allies (in McCarty’s case, economists; in Garrison’s case, regional scientists which through Warnitz and Stewart led to physicists; and in Haggett’s case through Chorley to geomorphologists); and the greatest need for training and education in the new methods (new courses at the University of Washington, and the Maddingly Hall lectures for British schoolteachers).

As the network becomes more entrenched, so the claims become more confident and less shrill. Textbooks such as Stanley Gregory’s (1963) *Statistical Methods and the Geographer* are written codifying the first hesitant forays, specialized journals appear such as *Geographical Analysis* (see Golledge, 1979), and foundationalist epistemological claims are made turning on the essential correctness of mathematics as a representational language (for examples, see Barnes, 1994). More generally, geographical practice was being transformed. What geographers did was beginning to be defined by numbers, equations, punch cards, and FORTRAN manuals; that is, it was becoming a desk-bound form of enquiry rather than field-based. Of course, there had always been armchair geographers, but quantitative methods presaged a discipline of only armchair geographers. No wonder Stamp demanded that this kind of thing had to stop.

But, of course, it did not stop. Stamp had failed to recognize something even more important, perhaps: the power and control bound up with numbers and equations on a page. A good example was the way in which the leading theory of the day, the gravity model, was joined with the leading statistical techniques of the day, correlation and regression analyses.

The standard formula for the gravity model is:

\[
I_{ij} = f(P_i, P_j, D_{ij}) ,
\]

where \(I_{ij}\) is the interaction between places \(i\) and \(j\); \(P_i\) and \(P_j\) are the masses (populations) of, respectively, places \(i\) and \(j\); and \(D_{ij}\) is the distance separating places \(i\) and \(j\).

Equation (5) can be operationalized as:

\[
\ln \frac{I_{ij}}{P_i P_j} = k - b \ln D_{ij} ,
\]

which is the form of a simple regression model with the logarithm of the observed interaction intensity as the dependent variable, and the logarithm of distance as the independent or explanatory variable (Olsson, 1970, pages 226–228).

The very high correlation coefficients associated with the testing of the gravity model led some to speculate that, like Galton’s law of heredity, the gravity model when expressed in the form of the regression equation represented a fundamental law of space. In Waldo Tobler’s words, “everything is related to everything else, but near things are more related than distant things.... [This is] the first law of geography” (Tobler quoted by Olsson, 1970, page 228). Specifically, that first law was manifest on the page as a high \(r^2\) and a negative \(b\) (‘friction of distance’) coefficient.

What was happening here was that diverse events in far off places—shopping trips by Nebraskan consumers, migration choices by rural Swedes, and the travelling patterns
of lorry drivers in Chicago were being translated and enrolled into the network of quantitative geography. Through numbers and mathematical equations those events ended up on 8 ½ - 11 inch paper, and bound between the covers of the Geographical Review, or Economic Geography, or The Canadian Geographer. One reason the network garnered increasing allies was because those numbers, graphs, and equations seemed so effective in communicating over long distances. Events and things were ordered and controlled, shuffled and manipulated, at long range. Moreover, that shuffling and manipulation were not simply the innocent pastimes of those in ivory towers, but as Gunnar Olsson (1980, chapter 14) showed for the gravity model and the associated techniques of correlation and regression, they had an effectivity in the world. By using $r^2$ and $b$ coefficients as a basis to plan the provision of services, the state could make the world conform to the gravity model.

An interpretation
The argument so far has been that correlation and regression analyses in human geography was successful because it was part of a wider network that managed to align a heterogeneous set of elements into a stable order. There were the numbers themselves which were being accumulated at an increasing rate (Haggett, 1965, page 277, says that they were accumulating almost too fast). There were the machines that offered both discipline and prestige (Ambrose, 1969, page 16; Gould, 1984, page 18). There were the contacts with other disciplines. Peter Gould (1984, page 19) writes, “We also found many colleagues in adjacent fields of the social sciences who were employing statistical methodology and we began to share ideas and exchange techniques...There is no question about it: this was an important facet of the development of statistical methodology in the field”. And there were the statistical techniques themselves including correlation and regression analyses. These techniques were not just techniques, though, but powerful pieces of rhetoric that could make dissenters feel lonely and convince potential allies to get on side (Taylor, 1976). And those allies were crucial. Without them the facts of quantitative human geography would not be facts. As Steve Woolgar (1996, page 829) puts it: “The ‘hardness’ of a fact is a reflection of subsequent usage rather than of its correspondence to a pre-existing nature; ... a claim becomes a fact in virtue of the conviction of sufficient numbers of allies”.

As is now known, though, many of those allies from the mid-to-late 1970s increasingly deserted the network. There were many reasons for this (see Cloke et al, 1991; Johnston, 1991), but one of them is the clash between the original biometric context in which inferential statistics such as correlation and regression were developed and the context of use in which they were employed by human geographers. The broader argument is about the socially constructed nature of statistics. If statistical techniques were really the embodiment of “the universal language of mathematics” (Garrison, 1956, page 428) then where they were applied would not matter. In fact, this was a point of contention between Reynolds and Garrison. Whereas Reynolds (1956, page 129) had argued that “…it takes time to translate the language of biology or psychology in which fields most of the [statistical] methods have been developed into that of regional... geography”, Garrison (1956, page 428) countered that because of its universality “…mathematical language may easily be adopted by a non-mathematical field”. Even proponents of quantitative methods would probably now say that Garrison was wrong and Reynolds was right.

The kind of context in which Galton, and later Pearson, and even later still Fisher, Jerzy Neyman, and Pearson’s son, Egon, developed their inferential statistical techniques was one that involved statistical independence, large samples, and the experimental method (Gigerenzer et al, 1989, chapters 2 – 3). For the most part none of that context was
replicated when early human geography quantifiers came to apply those techniques. Rather, the way they defined what they were doing was in direct opposition to all of the pivotal background conditions that formed the constitutive origins of inferential statistics. This is perhaps best illustrated with respect to the independence assumption. Tobler’s ‘first law of geography’ was not only the leitmotif of the gravity model, but it was central to the more fundamental project of spatial science in which quantitative methods were an integral part. For at the heart of spatial science is the idea that there is a geographical order to explain and that it is explained by space itself. In other words, as an enterprise spatial science rests on the assumption of “spatial interdependence” (Hepple, 1974, page 96). But if this is so, then it contradicts one of the classical assumptions made by Galton et al, namely, that of statistical independence. As Les Hepple (1974, page 94) writes, “A glance at any set of maps of geographical phenomena will give a strong disposition against a classical urn-model generation of the pattern [and on which classical inferential statistics is predicated]. As F F Stephen remarked: ‘Data of geographic units are tied together like bunches of grapes, not separate, like balls in an urn’.”

The implication is that the context that shaped classical inferential statistical techniques was quite different from the one in which those techniques were applied in human geography. It is true that a number of people later recognized this incongruity (Bennett and Wrigley, 1981; Gould, 1970; 1984; Poole and O’Farrell, 1971; Silk, 1981), and have sought to devise statistical techniques that avoid the problems (for example, Anselin, 1988; Griffith, 1985). For the purposes of this paper, though, the interesting point is not whether the problem of spatial interdependence can be solved but how it became a problem in the first place. It became a problem because of the context-dependent nature of the statistics themselves; the language of mathematics turned out not to be universal. At that juncture allies began to be lost, and David Harvey (1973, page 128) was taken seriously when he began saying things like, “The quantitative revolution has run its course, and diminishing marginal returns are setting in ...”.

Conclusion

By making use of a set of fragmented historical stories I have argued that numbers and statistics are not the precipitates of some superordinate rationality but are bound up with shifting networks of power, institutions, people, and things. The key word is historical. It is easy to think of statistics as having no history; that in the beginning were numbers.

The argument, though, is that numbers are not eternal perfect forms but have very particular social and cultural biographies. This claim was illustrated by presenting, albeit in a very crude way, the birth of statistics in the mid-19th century as the offspring of both an earlier concern with probability, and a very much earlier concern with state enumeration. Correlation and regression were the first of the inferential statistics, but they too were designed to advance a particular social and ideological agenda, eugenics.\(^1\)

\(^1\)I should make it clear that I am not claiming that regression analysis is an inherently racist technique or anything close to that. Regression techniques can be used for all manner of purposes both good and bad: for developing armaments of mass destruction through to developing medications that prolong the life expectancy of massive numbers of people. My point is that regression analysis originated within a particular intellectual context, in its case eugenics, and carries with it concomitant background assumptions. Those assumptions are not of the order ‘the West is best’. Rather, assumptions such as the need for independent sampling, large numbers of samples, and so on, facilitate control and power over certain bits of the world. Certainly, that control and power might be motivated and used for racist purposes, but it is not inherent in the technique as such. Both theories and statistical techniques ‘travel’. In making this point I am not claiming that regression analysis is neutral either. It is just that the assumptions underlying it cannot be read as equivalent to any simple racism. It is more complicated. I would like to thank Eric Sheppard for making me see the importance of this point.
As a result they were not a set of value-free, neutral techniques, but carried with them the assumptions, conditions, and wider problematics of the intellectual movement in which they arose. As correlation and regression analyses became black boxes, they moved out of biometrics and into other fields, including eventually human geography in the mid-1950s. The problem with this transferral, as I suggested, was a clash of underlying assumptions, background conditions, and disciplinary problematics (see Porter, 1994, who discusses a similar clash in economics). Because statistics is a social and cultural product, its formulae cannot necessarily be transported to a different context without inconsistencies arising.

In making this argument, I am not claiming that all correlation and regression analyses are a mistake, or that the very venture of statistical use here in geography or elsewhere is in error. In any case, to speak in this context of ‘mistake’ or ‘error’ is inappropriate. Statistics is one of the bases by which we now understand the world; it is our reality. If so, then it becomes necessary to illustrate the processes by which the construction of that reality occurs. It is here that a social and cultural perspective on quantitative methods is useful. It can be the crowbar for prising off the lid of the black box of statistics and taking a peek inside.

Acknowledgements. I would like to thank the three anonymous reviewers who provided useful comments on the paper, as well as to Eric Sheppard for his (always) astute suggestions.

References
Ambrose PJ (Ed.), 1969 Concepts in Geography 2: Analytical Geography (Longman, Harlow, Essex)
Anselin L, 1988 Spatial Econometrics: Methods and Models (Kluwer, Dordrecht)
Asad T, 1994, “Ethnographic representation, statistics and modern power” Social Research 61 55–88
Barnes T J, 1994, “Probable writing: Derrida, deconstruction and the quantitative revolution in human geography” Environment and Planning A 26 1021–1040
Barnes T J, 1996 Logics of Dislocation: Metaphors, Models and Meanings of Economic Space (Guilford Press, New York)
Bennett R J, Wrigley N, 1981, “Introduction” in Quantitative Geography: A British View Eds R J Bennett, N Wrigley (Routledge and Kegan Paul, London) pp 3–11
Berry B J L, 1993, “Geography’s quantitative revolution: initial conditions. A personal memoir” Urban Geography 14 434–441
Bloor D, 1983 Wittgenstein: A Social Theory of Knowledge (Macmillan, London)
Bloor D, 1990 Knowledge and Social Imagery second edition (University of Chicago Press, Chicago, IL)
Burton I, 1963, “The quantitative revolution and theoretical geography” The Canadian Geographer 7 151–162
Chorley R J, Haggett P (Eds) 1967 Models in Geography (Methuen, London)
Chorley R J, Haggett P (Eds) 1970 Frontiers in Geographical Teaching (Methuen, London)
Cloke P, Philo C, Sadler D, 1991 Approaching Human Geography: An Introduction to Contemporary Theoretical Debates (Paul Chapman, London)
Cooke P, 1989 Localities (Unwin Hyman, London)
Edwards D, Ashmore M, Potter J, 1995, “Death and furniture: the rhetoric, politics and theology of bottom line arguments against relativism” History of the Human Sciences 8 25–49
Fisher Box J, 1978 R A Fisher: The Life of a Scientist (John Wiley, New York)
Galton F, 1869 Hereditary Genius (Macmillan, London)
Galton F, 1872, “Address as President of the Geography Section” British Academy of Advanced Sciences 42 198–203
Galton F, 1888, “Co-relations and their measurements, chiefly from anthropometric data” Proceedings of the Royal Society of London 45 135–145
Galton F, 1890, “Kinship and correlation” North American Review 150 419–431
Galton F, 1908 Memories of My Life (Methuen, London)
Garrison W L, 1956, “Mathematical geography” Geographical Review 46 427–429
Garrison W L, 1979, “Playing with ideas” Annals, Association of American Geographers 69 118–120
Gigerenzer G, Swijtink Z, Porter T, Daston L, Beatty J, Kruger L, 1989 The Empire of Chance: How Probability Changed Science and Everyday Life (Cambridge University Press, Cambridge)
Golledge R G, 1979, “The development of geographical analysis” Annals, Association of American Geographers 69 151 – 154
Goodman N D, 1979, “Mathematics as an objective science” American Mathematical Monthly 86 (7) 540 – 547
Gould P R, 1970, “Is Statistix Inferens the geographical name for a wild goose chase?” Economic Geography 46 439 – 448
Gould P R, 1984, “Statistics and human geography: historical, geographical and algebraical reflections”, in Spatial Statistics and Models Eds G L Gaile, C J Willmott (D Reidel, Dordrecht) pp 17 – 32
Gregory S, 1963 Statistical Methods and the Geographer (Longman, Harlow, Essex)
Griffith D A, 1985 Spatial Autocorrelation: A Primer Association of American Geographers, 1710 Sixteenth Street NW, Washington DC 20009
Hacking I, 1987, “Was there a probabilistic revolution 1800 – 1930?”, in The Probabilistic Revolution volume 1, Eds L Kruger, L J Daston, M Heidelberger (MIT Press, Cambridge, MA) pp 45 – 58
Hacking I, 1990a The Taming of Chance (Cambridge University Press, Cambridge)
Hacking I, 1990b, “Probability and determinism, 1650 – 1900”, in Companion to the History of Modern Science Eds R C Olby, C J Willmott, R J R Christie, M J S Hodge (Cambridge University Press, Cambridge) pp 590 – 637
Haggett P, 1961, “Landuse and sediment yield in an old plantation tract of the Serra do Mar, Brazil” Geographical Journal 127 60 – 62
Haggett P, 1965 Locational Analysis in Human Geography (Edward Arnold, London)
Harding S, 1986 The Science Question in Feminism (Cornell University Press, Ithaca, NY)
Harvey D, 1973 Social Justice and the City (Edward Arnold, London)
Hepple L W, 1974, “The impact of stochastic process theory upon spatial analysis in human geography”, in Progress in Human Geography volume 6, Eds C Board, R J Chorley, P Haggett, D R Stoddart (Edward Arnold, London) pp 91 – 142
Hilts V L, 1973, “Statistics and social science”, in Foundations of the Scientific Method: The Nineteenth Century Eds R N Giere, R S Westfall (University of Indiana Press, Bloomington, IN) pp 206 – 233
Hilts V L, 1978, “Aliis exterendum, or, the origins of the Statistical Society of London” Isis 69 21 – 43
Hooker R H, 1907, “Correlation of the weather and crops” Journal of the Royal Statistical Society 70 1 – 51
Hudson D, 1955, “University of Washington” The Professional Geographer 7 (4) 28 – 29
Johnston R J, 1991 Geography and Geographers: Anglo-American Geography Since 1945 fourth edition (Edward Arnold, London)
King L J, 1979, “Areal associations and regressions” Annals of the Association of American Geographers 69 124 – 128
Kitcher P, 1983 The Nature of Mathematical Knowledge (Oxford University Press, New York)
Kitcher P, 1990, “The foundations of mathematics”, in Companion to the History of Modern Sciences Eds R C Olby, G N Canto, J R Christie, M J S Hodge (Routledge, London) pp 677 – 689
Lakatos I, 1976 Proofs and Refutations (Cambridge University Press, Cambridge)
Latour B, 1987 Science in Action: How to Follow Scientists and Engineers Through Society (Harvard University Press, Cambridge, MA)
Latour B, 1990, “Drawing things together”, in Representation in Scientific Practice Eds M Lynch, S Woolgar (Cambridge University Press, Cambridge) pp 19 – 68
Latour B, 1993 We Have Never Been Modern (Harvard University Press, Cambridge, MA)
Law J, 1994 Organizing Modernity (Blackwell, Oxford)
McCarty H H, 1954a, “McCarty on McCarthy. The spatial distribution of the McCarthy vote, 1952”, unpublished paper, Department of Geography, State University of Iowa, Iowa City
McCarty H H, 1954b, “An approach to a theory of economic geography” Economic Geography 30 95 – 101
McCarty H H, 1956, “Use of certain statistical procedures in geographical analysis” Annals, Association of American Geographers 46 263
McCarty H H, 1979, “Geography at Iowa” Annals of the Association of American Geographers 69 121 – 124
McCarty H H, Landberg J B, 1966 A Preface to Economic Geography (Prentice-Hall, Englewood Cliffs, NJ)
McCarty H H, Hook J C, Knos D S, Davies G R, 1956 *The Measurement of Association in Industrial Geography* Department of Geography, State University of Iowa, Iowa City

McCloskey D N, 1985 *The Rhetoric of Economics* (University of Wisconsin Press, Madison, WI)

MacKenzie D A, 1979, “Eugenes and the rise of mathematical statistics in Britain”, in *Demystifying Social Statistics* Eds J Irvine, I Mills, J Evans ( Pluto Press, London) pp 39–50

MacKenzie D A, 1993, “Negotiating arithmetic, constructing proof: the sociology of mathematics and information technology” *Social Studies of Science* 23 37–65

Merton R K, 1973 *The Sociology of Science: Theoretical and Empirical Investigations* (University of Chicago Press, Chicago, IL)

Mirowski P, 1989, “The probabilistic counter-revolution, or how stochastic concepts came to neoclassical economic theory” *Oxford Economic Papers* 41 217–235

Mirowski P, 1994, “Doing what comes naturally: four metanarratives on what metaphors are for”, in *Natural Images in Economic Thought: Markets Read in Tooth and Claw* Ed. P Mirowski (Cambridge University Press, Cambridge) pp 3–19

Olsson G, 1970, “Explanation, prediction, and meaning variance: an assessment of distance interaction models” *Economic Geography* 46 223–233

Olsson G, 1980 *Birds in Egg/Eggs in Bird* (Pion, London)

Pearson K, 1914 30 *The Life, Labours and Letters of Francis Galton* 3 volumes in 4 parts (Cambridge University Press, Cambridge)

Pearson K, 1970, “Notes on the history of correlation”, in *Studies in the History of Statistics and Probability* Eds E Pearson, M Kendall (Griffin, London) pp 185–205; first published in *Biometrika* in 1920

Poole M A, O'Farrell P N, 1971, “The assumptions of the linear regression model” *Transactions, Institute of British Geographers* 52 145–158

Poovey M, 1993, “Figures of arithmetic, figures of speech: the discourse of statistics in the 1830s” *Cultural Inquiry* 19 256–276

Porter T M, 1981, “A statistical survey of gases: Maxwell’s social physics” *Historical Studies in the Physical Sciences* 12 77–116

Porter T M, 1985, “The mathematics of society: variation and error in Quetelet’s statistics” *British Journal of the History of Science* 18 51–69

Porter T M, 1986 *The Rise of Statistical Thinking* (Princeton University Press, Princeton, NJ)

Porter T M, 1992, “Quantification and accounting ideal in science” *Social Studies of Science* 22 633–652

Porter T M, 1994, “Rigor and practicality: rival ideals of quantification in nineteenth-century economics”, in *Natural Images in Economic Thought: Markets Read in Tooth and Claw* Ed. P Mirowski (Cambridge University Press, Cambridge) pp 128–170

Restivo S, 1992 *Mathematics in Society and History: Sociological Inquiries* (Kluwer, Dordrecht)

Reynolds R B, 1956, “Mathematical geography” *Geographical Review* 46 129–132

Rose J K, 1936, “Corn yield and climate in the corn belt” *The Geographical Review* 26 88–102

Shapin S, 1975, “Phenomenal knowledge and the social structure of early nineteenth-century Edinburgh” *Annals of Science* 32 219–243

Silk J, 1981, “The general linear model”, in *Quantitative Geography: A British View* Eds R J Bennett, N Wrigley ( Routledge and Kegan Paul, London) pp 75–85

Stigler S M, 1986 *The History of Statistics: The Measurement of Uncertainty Before 1900* (Harvard University Press, Cambridge, MA)

Taylor P J, 1976, “An interpretation of the quantification debate in British geography” *Transactions, Institute of British Geographers* 1 129–142

Taylor P J, 1977 *Quantitative Methods in Geography: An Introduction to Spatial Analysis* (Houghton Mifflin, Boston, MA)

Thrift N J, 1995, “Peter Haggett’s life in geography”, in *Diffusing Geography: Essays for Peter Haggett* Eds A D Cliff, P R Gould, A G Hoare, N J Thrift (Blackwell, Oxford) pp 375–395

Tribe K, 1991, “The economic metric” *Economy and Society* 20 411–422

Watson H, 1990, “Investigating the social foundations of mathematics: natural numbers in culturally diverse forms of life” *Social Studies of Science* 20 283–312

Williams R, 1982 *Politics and Letters* (New Left Review Books, London)

Woolgar S, 1996, “Sociology of science”, in *The Social Science Encyclopedia* second edition, Eds A Kuper, J Kuper (Routledge, London) pp 828–830