სოციალური კვლევის რაოდენობრივი მეთოდები

Quantitative social research methods

რიდერი

მომზადებულია თეონა მატარაძის მიერ

2

Setting Up Sociological Research

Geoff Payne

It is not a novel thing to say that research needs to be a reflective process, but it would be something of an innovation if we were to take the social context of research activity more seriously. Despite, or perhaps indeed because of, our impressive technical developments in recent years, we are too busy doing research to bother over-much about the conditions of production of sociological knowledge. Our collective research quality could benefit from reviewing the process through which people end up as researchers, and how 'junior' researchers become involved in carrying out projects often with little awareness of what really goes on in data capture, what informants are doing when we research them, or how the data captured and analysed with our sophisticated techniques actually relate to theoretical questions. In discussing the way research in general is situated in sociology, rather than a particular technique, this chapter is less a resolution than a prolegomenon to such a review.

Of course, this begs the initial question that sociology, even less than sociological research, is not a unitary discipline. The chapter suggests several ways of conceptualising the discipline that also impinge on how we do research. Starting with illustrations from unpublished works – after all, why should we only look at our best products? – which show authors with considerable abilities who have not grasped the basic relationship between evidence and theory, sections on 'sociology as publications' and 'sociology as taught' demonstrate the ways that research techniques have become isolated from sociological thinking. Taking the example of the interviewee as an active, knowing and disobligingly complex social being, the latter part of the chapter suggests that we over-estimate the quality of the data we capture, and our capacity to make sense of it, instead placing excessive faith in our technical expertise and the shared conventions of our sociological schoolings. In particular, our assumptions about what our informants tell us, and what may be missing from our data, need re-consideration.

The following sections therefore treat research not as a separate act but as located in a longer process of training, professionalisation and continuing practice. This will allow us to question some taken-for-granted assumptions about how we carry out our research. The common theme connecting the discussion is what actually takes place within research projects, rather than what we like to present to the outside world, and to ourselves, as happening. Researchers of a more qualitative, and feminist, disposition would probably object to the implication that we lack reflexivity and self-awareness, and in part they would be justified. However, a life-time of social research has left me sceptical about the capacity of many colleagues (if one may generalise for the moment) to confront their own short comings – and I do not exclude myself from this concern.

FOUR UNFAIR ILLUSTRATIONS

This can be illustrated by examples of unpublished sociological work; those papers submitted to journals but not accepted. There appears to have been no systematic research into the content of empirical research in such papers. Even assuming that a proportion of the submission 'traffic' consists of works re-circulating in search of a home, the rejection rate of the major journals suggests that for every published article there are at least two more which are not part of the visible published record. That is a considerable quantity of sociological activity: the fact that it is unsuccessful is inadequate grounds for ignoring it. Indeed, one might even argue that its greater volume makes it more significant than the minority of work which is published.

Articles rejected for publication, or work-in-progress papers at small, specialist conferences of like-minded colleagues, are admittedly by definition minor, unsuccessful, or 'unacceptable' by the mainstream canons of the discipline. On the other hand, these papers have been produced by sociologists who believe that their work is suitable for publication on the basis of many months of hard labour. The deficiency of the end-product is not so much a problem of individual failings, but a question of what social processes have generated deficient work. What was lacking in the training and management of these (generally less-experienced) researchers? Why were they allowed to engage in poor research and produce deficient work, and not discouraged from attempting to get it published?

One is constrained in discussing this opaque body of sociological production both by the anonymity of authorship rules of submission and the conventions of confidentiality of reviewing and editorial board membership, and yet it is here that we best see the soft underbelly of the profession. Several recent articles (from among about three dozen seen in the last year) are described here as anecdotal illustration (and not as a proper sample). Because they were anonymised and not subsequently accepted for publication, I obviously cannot give bibliographic references for them: if this lack of chapter and verse troubles the reader, I can only suggest that the examples are regarded as fictional inventions, because their function is heuristic rather than strictly evidential. As their purpose is general illustration, rather than any criticism of the individual authors, it is perhaps better that they remain unidentified. The following accounts contain deliberate minor distortions intended to sustain anonymity. The minor breach of the ethics of confidentiality will I hope be compensated by adhering strictly to the principle of anonymity.

Example A is a study of that old chestnut, the relationship between education and occupational outcome. The authors explored the extent to which educational attainment explains occupational outcomes (a classic question in social mobility analysis) drawing on a large national American survey data-set collected for other purposes, and deploying MVA techniques. Despite some neat and indeed quite impressive innovation in the statistical analysis, the results could be anticipated well in advance because the topic has already been so well-researched. Predictably, it was found that, as in all the other cases, education and occupation are associated but not to the extent that common sense might suggest. Nevertheless, the fact that the topic is an old chestnut presumably justifies its selection as a topic, and the authors explained their work clearly and structured their argument in a cogent fashion. Why then was I unhappy with the explicit finding claimed by the research team? Because in fact their research design turned out to not sustain this conclusion.

In the first place, the experience of education was operationalised as ‘number of years of completed full-time education’. There was no discussion of alternative indicators such as part-time schooling, vocational training undertaken as part of employment, differences in school or college environment, or subject specialisation. There was almost no discussion of qualification level achieved, and no comment on the convenient fact that number of years of completed full-time education is a quantitative measurement. The reason for the simplistic definition of education was given in a single sentence: the large and highly-regarded data-set being deployed in the secondary analysis did not contain sufficient information for a more elaborated conception of what ‘education’ means.

But worse was to follow. The researchers – who can by no means be described as lacking ability, given the sophistication of their analytical techniques – explain briefly that because women’s experience of employment differs from that of men, and that it is not easy to assign women to categories of occupational class, no females were included in the analysis. Similarly, as people of colour and immigrants are subject to specific discriminations in the markets for both education and employment, they too needed to be excluded. The claimed finding, that education is moderately associated with occupation, is therefore nothing of the kind. What was actually established was that there is a moderate association between number of years of full-time education and occupational achievement among the American-born white males who comprise at best 40% of the adult population, which is a very different thing from what was explicitly envisaged at the outset. What makes this actual finding sociologically interesting is not that the moderate statistical association which applies to only a sub-set of the population, but rather, first, why it apparently does not apply to females and people of colour, and second, what market ramifications does the situation of these other groups have on the occupational achievement of white males? Even if markets are gendered and racialised, it makes little sense to study one part of the market in isolation from the rest.

This may seem an extreme case but it is not totally atypical. Much of mainstream social mobility research in the last quarter of the twentieth century dealt only with survey data on males. The major European CASMIN project is a classic example. Certainly most of my own early work in this field is open to this stricture. Thus, the illustration is part of a wider failure to recognise the limitations of one's data: even in the case of the specific experiences of distinctive minority groups, these must always be seen in the context of the majority group experience (e.g. Iganski et al., 2001).

A second example is a work-in-progress paper on voluntary association membership. This began with an excellent discussion of the literature, leading to a clever re-casting of the issues in terms of social capital in post-modern society, which was truly impressive. However, the data consisted of the number of voluntary association memberships at time $t = 1$, and the number of voluntary associations joined or left between $t =1$ and $t = 2$. In other words, the design treated all voluntary associations (e.g. trade union, church, residents group, sports club, learned body, gym, parents support group and political party), all levels of membership (passive member; active participant, committee member or officer) and all geographical areas (neighbourhood, region, nation, or global) as identical. As in the first example above, the justification offered for this confounding of activities was that the data-set being used in the secondary analysis did not have any further evidence to offer. In other words, the research question, despite its impressive elaboration, could never be answered with the information available. Why was the research worker being allowed to waste their time (and public funds) on a project which manifestly could never be successful?

Both of these examples come from the quantitative tradition, but equivalent cases can be found in qualitative research. Example C claimed to show that the capitalist class had imposed new conditions of exploitation on labour by demanding that workers maintain high levels of physical fitness and a healthy appearance while at work. The data consisted of a small number of semi-structured interviews with the staff and gym users at two gyms located in a Financial Services district. Informants (mainly gym staff) were quoted as agreeing that more people these days were concerned about how they looked when at work, and that they felt physical fitness was important for career progression. These self-interested replies of the gym staff were taken at face value. No interviews were obtained with any members of the capitalist class or their service class managers. There was no specification of which 'workers' were being subjected to this new form of exploitation, or how many were suffering it, at what stages of their careers, or at what sites. No time frame for the new trend was indicated. No interview quotations referred directly to the core research hypothesis. To adapt a recent British advertising slogan, the research in question did not do 'what it says on the tin'.

In example D, the researcher's explicit aim was to explore how gender determined conditions of employment and academic output in universities. The article consisted of a thoughtful and insightful discussion of the content of extensive interviews completed with six married female academics, although one case was given most attention. The six women, all employed by one high status university, reported how difficult and drawn-out had been the process of obtaining tenured posts. They felt that their careers had been held back by this experience. This was taken to show how gender discrimination works in higher education, an area where universalistic criteria of performance are supposed to apply.

As an explication of some females' experience of academic life, this research may be acceptable. However, as an account of gender it is obviously deficient. The high status university that was the research site is distinctive in having very few tenured posts: first, men as well as women experience difficulty and delays in moving from temporary funded posts onto faculty (even if the degree of difficulty is not the same). Second, as the university is within easy commuting distance of at least half a dozen other high status institutions, let alone an even larger number of less prestigious universities, what was really interesting was the neglected question of why the six women made the choice to restrict their ambitions to the one campus, and therefore whether their 'difficulties' were less imposed upon them than to some extent self-inflicted. Third, the choice of that university as a research site was not discussed (it happened to be the one at which the researcher was a woman on 'soft money'). Its lack of typicality may have shone light on the specific issue of tenured and non-tenured employment, but that was not how the paper was framed. Instead it claimed to be about gender inequalities in higher education, which is not at all the same thing.

We are not talking here about articles submitted to journals that are then rejected for the conventional reasons that they are badly written, about obscure topics, show a lack of knowledge of the field, offer no evidence or contain computation errors, or do not sufficiently advance sociological knowledge. The point about these four examples is that their methods are fundamentally flawed. The data discussed in them do not relate to the research questions at their hearts. Nor have these been the only cases recently encountered, which suggests that this problem extends beyond the illustrations given here.

METHODS AS TAUGHT

All four of these examples show an incapacity to connect sociological questions with appropriate selection of data and awareness of what data really are. This basic methodological problem can be found to varying degrees across a range of sociological research, where issues have not been thought through, or there has been an over-reliance on the principle of *faute de mieux*. This has its roots in the way students come to think about sociology, and particularly in how we see the choice of informants and their answers. It is this set of problems that we need to address.

Such a swinging critique of actual research practice may be accused of over-generalisation. Making statements about what is taking place in sociology is no easy task: exceptions to every attempted rule abound. On what basis can one characterise the research practices of an entire discipline? In earlier discussions of sociological research methods in the UK, my co-researchers and I made several attempts to capture the discipline's contemporary geist, first examining undergraduate programme documentation, arguing that 'we are what we teach'. This approach sees sociology, and the way it researches, as its public face; in that most sociology students will not become professional sociologists but carry away an image of the discipline into their subsequent lives as members of the public. It also treats undergraduate education as the first phase of a professional socialisation process, in which undergraduate experience of research (or its lack) largely determines choices of topic and method in graduate school, while this in turn locates young sociologists on specific career tracks in which their methodological preferences are effectively pre-set for life.

In our 1980s exploration of the teaching of research the choice of data was essentially serendipitous. Our content analysis of undergraduate programme documentation in Britain was part of a wider project (Eggins, 1988) initiated by the Council for National Academic Awards (CNAA). The coverage was governed by the CNAA's remit as custodian of academic standards in the UK's institutions of higher education which subsequently became the country's 'new' or 'post-1992' universities (Payne et al., 1988, 1989). There was nothing special about the year in which the research was carried out, nor which years were included in the study: we simply used the documentation available for current courses. We did not debate questions of sampling, taking all available data for a time period that was externally determined. Our 'sample' was therefore time-specific and institutionally-constrained, rather than a response to any supposed methodological crisis within the discipline or dependent on the logic of some concern purely internal to British sociology.

This phase of research found research methods to be a core element in all institutions, covering both qualitative and quantitative approaches, but with an emphasis on the former. The most common teaching pattern was one (or two) free-standing module(s) in the second year of the degree programme. Qualitative methods were taught by sociologists who made other contributions to the programme, whereas quantitative methods were more often taught by statisticians from outside of the sociology department. Only a minority of programmes involved the students in hands-on researching (final year dissertations were then uncommon), and there was no evidence of research methods being taught as part of other modules. We coined the phrase 'the ghetto-isation of methods' to describe the way that doing research, and how sociological 'knowledge' has been generated, was not integrated into the mainstream programmes.

In a later phase of this work, we asked students and lecturers directly about their perceptions of what was being taught, rather than relying on module documents (Williams et al., 2004, 2008). Although this exposed us to low response rates among undergraduates it gave a more direct and informative access to what happens in class. Among our findings were that qualitative and quantitative methods continue to be covered in most programmes, with teaching now repatriated to sociology departments. Dissertations have become standard for Honours degrees. Research methods otherwise still tend to sit in isolation, and the depth of knowledge of statistical techniques is questionable. Most students perceive sociology as a humanities discipline, and prefer writing essays to analysing sociological data.

Although there are national differences, these patterns are not unique to the UK: the Netherlands is an example of paying greater attention to the development of quantitative skills (Parker, 2011). In the US, the American Mathematics Association has promoted the adoption of minimum quantitative skills for undergraduates, an approach endorsed by the National Science Foundation which funded the ASA-supported Integrating Data Analysis project (Gillman, 2006; Howery and Rodriquez, 2006). The full impact of these initiatives is yet to be seen, with many American students still gaining little experience of doing research. Parker's transnational survey indicates that the key issue is increasingly seen as the extent to which data analysis and hands-on research can be embedded across the curriculum so that undergraduate research practice is systematically encouraged, rather than leaving this to graduate school (Healey, 2008). What matters is not the specific research method used, but rather the practice of researching *per se*.

Two conclusions follow from this work. First, in most countries it is thought proper to introduce undergraduates to research methods, so that research would appear to be part of the sociological canon. However, apart from the dissertation, students are still not called upon to use their methods knowledge, and their learning of methods is separated from the rest of 'sociology'. Right from the start there is a gap between methods as technique and methods as sociological reasoning. Second, the level of skills acquired is low. For example, students probably know what a probability sample is but could not design and set up such a sample: their experience of working with the strict logic of probability is insignificant. Even after graduate school, young sociologists will have practised only a limited range of skills, and acquired only a 'black box' comprehension of others (typically, quantitative techniques that can be learned in a formal way). What they are less likely to have acquired is a rounded understanding of how data collection and data analysis need to be embedded in sociological reasoning.

Our commentaries on the content of undergraduate education, and the methods used in recent research, not least by 'junior researchers', in which we treat sociology 'as what we publish' (Payne et al., 2004) have been taken in some quarters as an implicit antagonism towards qualitative methods (May, 2005). However, our intention has always been to promote the positive value of quantitative research without denigrating other approaches, and to insist on the need for skills in a range of methods – in short, for methodological pluralism. This stance of measured neutrality allows us to criticise shortcomings in both qualitative and quantitative practice. Not least it fosters a degree of circumspection when it comes to the interpretation and presentation of one's research findings.

The way in which students are introduced to a discipline says a lot about how its practitioners perceive it. The isolation of research methods from the remainder of sociological thinking and analysis makes it harder for young sociologists early in their careers fully to comprehend data collection and research design. This is excaberated by teaching organised around the convenient but spurious dichotomy between qualitative and quantitative styles of research. Although both modes of research have deficiencies in how they see sampling and interviewing, they could each learn a great deal from the other.

OBLIGING AND DISOBLIGING INFORMANTS

I recently found myself living in an area selected by the UK's Office of National Statistics for testing new surveys. In rapid succession I was interviewed at length about health provision; employment; and twice on household consumption in terms of both diet and expenditure. By coincidence, I had also signed up as a panel member of an internet survey company so that I could volunteer my attitudinal prejudices, and curiosity stirred, I had begun to reply to every market research questionnaire disguised as prize draws that came my way (as yet without pecuniary success!). I even started to agree to be interviewed on the street by market research interviewers in search of quotas. In this way, after many years of carrying out social research as an investigator, I encountered social research extensively for the first time from the perspective of informant.

Up to this point in my career, I had followed the British Market Research Society Guidelines not to participate in data collection as a respondent. Following the Society's logic, as a 'knowing person' I was not a typical informant – but then on the other hand, the population includes such knowing people, and despite what we may or may not contribute, we are part of the universe from which samples are drawn. The experience of my new participation suggests that the trouble lies less in what we may do to the research findings (where we are likely to be a tiny minority of possible outliers), or as commercial rivals who might gain from seeing potentially advantageous new ways of data collection, but rather in what the experience may do to us as researchers.

Readers may like to ask themselves how often they have been the subject of research in, say, the last three years? If my informal (non-representative sample) soundings of colleagues are anything to go by, your answer is likely to be 'not (or hardly) at all'. None of my acquaintances has been interviewed more than once, and almost all had not experienced being interviewed, within that time-frame. However, to be interviewed once, or from time to time at long intervals, as I guess most sociologists have been, is not enough. Such single events may make one aware of specific technical shortcomings and the substantive nature of the project in question, but they are insufficiently intense to generate a deeper awareness of what 'doing being interviewed' is like, or to provide a platform on which to build a revision of what we mean by 'interviewee'.

As practitioners, our dominant experiences of data collection have been as student, research assistant, investigator, writer, teacher and critic. We therefore conceive of data collection as an active process of doing unto others, not as done to ourselves. Our primary conceptualisation is not as 'respondent', 'subject', 'interviewee', 'informant' or 'participant', but as researcher. This stipulation also extends to retrospective reflexive accounts of how research was actually carried out. In short, sociological research can be defined as an activity that social scientists practice on other people, even in 'collaborative' projects.

This assertion is a prelude neither to a naïve insistence that first-hand experience of 'being on the receiving end' is a pre-requisite for carrying out research, nor a polemic in favour of partnership or co-designed research. My point is that seeing social research from the other side of the questionnaire, laptop, or tape recorder can be a chastening experience: at least it has been for me. The challenge was not so much the discovery that other researchers' techniques, notably in question phrasing, questionnaire design and interviewing skills, leave great room for improvement – that was simply confirmation of something of which I, like many colleagues, have long been aware. Indeed, for academics and especially teachers of research methods, other researchers' inadequacies are our stock-in-trade. Rather, my experience made me realise that the way many professionals conceptualise the interviewee is inadequate. If I could be a thinking social agent during interview, why should I attribute to other informants a lesser state of consciousness? The logical conclusion of this insight is that social researchers should be much more cautious and humble about what they claim to have found in their collection of data.

Of course, there is no single model of 'the interviewee' in sociology. Conceptions range from the highly personalised to the faceless statistical unit, and from the downtrodden co-victim to the obdurate and misguided refuser. Unlike much of research in social psychology (conveniently practiced on a captive audience of psychology undergraduates, particularly in America), few sociologists take the view that subjects are always essentially co-operatively re-active. My reading of standard textbooks, the specialist literature (much of it written as guidance for inexperienced researchers) and more general accounts of findings based on empirical study, suggests that we unconsciously operate with one or more of the following models of the respondent, depending on the conditions of what is being researched:

- The refusenik
- The negative deviant

- The elite member
- The passive
- The simple-minded and potentially confused
- The co-operatively re-active
- The keen to please or excessively deferential
- The potential victim/ethically protectable
- The co-participant (feminist)
- The excessively garrulous.

I have deliberately omitted the special method of reflexive auto-ethnography, where the 'interviewer' and 'interviewee' is the same person, which I accept represents a limiting case. I anticipate that my outline typology is not an exhaustive list, but the identification of additional types does not invalidate its underlying principle, but rather strengthens my point. If informants can play these different roles – and change from one style to another during the course of data collection – this must surely raise doubts about probability sampling which treats respondents as inter-changeable uniform units, and how informants can bias the readings that qualitative researchers take from the interview encounter.

My interest here lies mainly in the first two categories of the typology, because the other categories are better represented in the literature. 'Refuseniks' are those people who decline to be interviewed. There is an extensive literature about ways of reducing the number of refusals in survey research, as well as statistical techniques for adjusting results to compensate for low response rates. Both of these tend to treat refuseniks as if they are either a uniform category, or sit on a monotonic continuum between acceptance and refusal. By definition we have little knowledge of them, although of course if we have some alternative parameters it is possible to identify their socio-economic position, gender, age, etc. through comparisons which by subtraction show their collective characteristics. The development of post coding geo-demographics and similar ecological descriptors offers an even more precise way of typifying those who we cannot successfully interview.

However, the key word here is 'typifying'. The underlying assumption is that refuseniks resemble those who we *do* interview, except for the single quirk of their refusal. If we can know some of their face sheet social features, we assume we can interpolate values for them from the sub-sample which they most resemble.

Alternatively, provided that our achieved sample is not seriously deficient in some social category, we need not worry unduly. But this begs the question of why some people refuse interviews. What if the refuseniks do not differ in most obvious respects (class, gender, age, etc.) but only in some other respects, one of which is their propensity to refuse? The propensity to refuse may be an indicator of a different set of social dispositions. Refusal may simply be a matter of non-co-operation due to inconvenience, an unwillingness to modify the refusenik's plans for spending their next hour. But does this indicate a personality type, or an actor whose embedding in their social routine is distinctive from other people? It is not implausible to suggest that refusal may be associated with one or more of attitude sets reflecting anti-authoritarianism; individualism and conservatism; anti-intellectualism; Asberger's Syndrome; mild paranoia; or even anti-social attitudes or socio-pathological tendencies.

In case this seems excessive, consider the second category of 'negative deviant', a category in which I would place myself – or rather, a category that I find myself in despite having started interviews in a 'co-operatively re-active' frame of mind. The longer the interview, the less co-operative I become. The poorer the question design, the more irritated and obstructive I become. To date, I have never terminated an interview, but my patience is not limitless. For example, where a series of questions ask me to make fine discriminations (or even express degrees of agreement/disagreement with Likert-like statements) or to engage in acts of careful recollection or calculation, I progressively find myself *unable to be bothered to be accurate*. I see no reason why similar reactions should not be experienced by other respondents. If that is so, the data we often collect is likely to be unreliable.

Nor do I find that this shift in response mode occurs only in survey research. Although my experience of being on the receiving end of qualitative research is more limited, I would argue that the same underlying process also operates. One's personal relationship with, say, a participant observer shifts over time: it may grow into stronger rapport and therefore 'openness', or decay into antagonism. When it comes to semi-structured interviews, the scope for destructive replies is even greater than with structured questionnaires.

A particular case in point is what informants 'volunteer' during interview. The conventional assumption seems to be that with good interviewing, informants will talk comprehensively about topics; and in providing answers, will depict their sense of issues. If the informant has any ideas or feelings about an issue, they will become expressed and conversely, the expression of responses means that the researcher can tell what issues matter to the informant. In other words, the *salience* of an issue can be evaluated by the informants replies (see for example Savage et al.'s (2001) explanation of class identity). However, it may be that where there are intense feelings or strong but controversial opinions are held, informants adopt precisely the opposite tack by remaining completely silent. It all depends on what image of the informant we bring to the table. There seems to be no simple way of knowing whether silence may also be an indicator of salience.

The model of the informant that I am sketching is of a knowing being, whose orientation cannot be tied down to some standardised social identity which is accessed via core sociological variables. This informant is knowing in both, the sense of being able to react to the interview, and to change their orientation during interview; a social person who not only reacts to the researcher in a dynamic way during the specific setting of the data capture process, but whose social persona is complex and multi-faceted in a way that cannot easily be reduced to a simple social type. Because people share certain social characteristics, such as class, gender, ethnicity and age does not entitle us to *assume* a consistent similarity of experience, values, attitudes or responses among them. Our task as researchers is to *discover* to what extent such patterns exist, but informants are under no obligation to assist us. Statistical techniques intended to compensate for gaps in our data due to informant non-co-operation, but which actually involve extrapolation or replacement based on the data from other informants which we do have, are both tautological and open to the charge of reliance on an 'over-socialized conception of man'.[1]

OBLIGING AND DISOBLIGING SAMPLES

Until recently, the problem of selecting who we study has not received much attention outside of survey research, or at least so one might believe from reading most published accounts. In the quantitative tradition, the mechanics of sampling are conventionally made a matter of record: in pre-analysis working papers (increasingly available on the internet); in methodological appendices to monographs; and in end-of-project reports to stake-holders. Journal articles may give cursory technical details, but they normally provide links through bibliographic referencing that allow readers to follow up any sampling questions.

Some big surveys have established such a reputation that they can be invoked almost as brand names, thus removing any need for further discussion because the product's qualities (and even any limitations) are universally acknowledged: many sources available from Statistics Canada, the ONS and the ESDS in the UK, and ICPSR in the US provide examples. In the UK, a growing trend towards cost-effective secondary analysis of large data-sets, particularly of omnibus or repeated surveys, is allowing junior researchers to dispense with the impedimenta of primary data collection, although this often still embroils them in statistical issues of weighting or problems of data definition. While there is no shortage of statistical literature on how to compensate for sampling defects, this typically shows little sensitivity to the initial character of the data capture. Because there is an expectation that sampling needs to be explained, researchers are required to give the matter some consideration, albeit only in a highly technical way.

This framework provides considerable comfort for the researcher, but it is not without its own constraints once one moves beyond synchronic analysis. The UK National Child Development Study (NCDS) offers a case in point. Probably the most widely-deployed data-set in educational, health and social inequality trend research, the prominence of the NCDS primarily depends on the absence of alternative, comprehensive and large data-sets for people born in the late 1950s. However, by its seventh re-survey sweep in 2004, attrition and non-response had reduced its original sample of 17,634 to 9,175, i.e. 52% (and if one were interested in comparing data from all seven sweeps, only 33.8% of the original sample can be directly used). While much of this 'missingness', to use a term from Carpenter and Plewis (this volume, Chapter 23) is not a total loss of cases, but rather partial loss of items or single sweeps, it remains reasonable to speculate that the survivors, in their adult status, differ sociologically from those who have disappeared from view, *even if on certain specified variables (with or without compensatory re-weighting) they do not*. How else are we to explain their non-compliance?

In distinguishing between the impact of various causes of missingness in the NCDS (e.g. for the case or item levels) Carpenter and Plewis show how in certain circumstances, such as 'Missing at Random', less biased estimates can be recovered from surviving data. However, their ingenious methods of tackling data which are 'Missing Not at Random' remains based on the assumption that there are no substantial differences between survivors and the missing. If we posit that some individuals have a lower propensity to co-operate with social surveys – especially repeated longitudinal studies – than others (due, say, to their social location, identity, or personality), the question remains whether they differ in the other social characteristics represented by the variables in the survey. If so, the recovery of estimates from the survivors is 'the most difficult situation to adjust for', as Carpenter and Plewis acknowledge in their introduction to this topic.

Attrition is of course a problem for all longitudinal studies, as is the need to rely on original research techniques that become out of date. Those re-using the NCDS have remained silent about the awkward matter that professionally-trained survey interviewers were not used to collect NCDS data until the 1981 sweep. Prior to this, data were collected by and from Health Visitors and school personnel who, with all due respect, can hardly be described as sociologically literate. Almost all secondary analyses of the NCDS have ignored this limitation, presumably on the grounds that there is no alternative data-set. Even if those collecting the data had been better trained in the collection of sociological data, the basic research tools of the 1950s were still primitive. A quarter of a century later, Hope and Goldthorpe (1974) were still attempting to establish the first coherent, explicit and all-inclusive method of scoring social class in the UK. It is little wonder that during the early 1990s, in examining nearly 80 multivariate analysis assignments using unweighted NCDS data (presented as part of the University of Ulster's excellent Social Research Masters degree requirements, and reviewed in my role as External Examiner), I found almost no statistically significant associations between the NCDS variables. The quantitative tradition may be better placed than other methods when it comes to information about sampling, but that does not mean that its record is beyond reproach, especially when it comes to *faute de mieux* justifications.

Nonetheless, at least there is a framework for discussing the sampling. In qualitative research the question has hardly been raised outside of the extended case study method in ethnography, so that the conventional discourse provides less scope for exploring it. It is no defence to argue that qualitative methods do not involve generalisation (Denzin, 1983; Lincoln and Guba, 1985). As an earlier close inspection of a series of journal articles showed (Payne and Williams, 2005), sociologists who base their interpretations on data collected from small numbers of informants (say, less than 50 individuals) do make explicit generalising claims that their findings have society-wide application and implications (Fahrenberg (2003) makes a similar point for interpretative methodologies in psychology and other social sciences).

This presents an almost intractable difficulty. It is no solution to train the users of such methods to eschew generalising statements (although that might not in itself be a bad thing). Such a policy would merely obscure the underlying fact that we need to know the logical status of the small-scale study. If the point of selecting a research site and its human occupants is *not* to generalise in respect to other similar cases, does the choice matter? On the one hand, if each study is self-contained and free-standing, its relationship with previous and subsequent research is problematic. In this view, sociology should be as non-cumulative as the field of literature where one short story is no guide to the plot or characters of another. If, as the constructivists would have us believe, all social phenomena are unique because they are context and time specific, what is the purpose of even exploring similarities of subjective, individual perspectives when our own explorations are themselves unique? But on the other hand, if each study is unique, then as readers we still need to know why the site and its informants were selected because this is an essential part of the unique events being catalogued.

Nor is it easy to connect the small-scale to larger populations by using principles from probability theory. This requires no elaborate statistical reasoning because any attempt falls at the first hurdle: there is usually minimal information about the supposed population, typically manifested in the absence of a sampling frame, a point conceded by Mayring (2007; para. 16) despite his optimism about probabilistic approaches. Indeed, where the research is concerned with small, unusual categories of persons, as in the case of hard-to-access minorities, the actual number of such people is open to dispute. If the topic involves studying what happens within an organisation, we seldom have a complete list of similar organisations, and even here it is necessary to specify the geographical unit – locality, state, region or nation – containing the sites. However, even if this were not so, the populations tend to be small, and so the sample size has to be large relative to the population size.

Thus, if we think of a type of organisation or a hard-to-access category of persons, conventional standards of sampling are going to present a logistical challenge. Sampling conventions (Krejcie and Morgan, 1970) dictate that an adequate sample of a population consisting of 10 or less organisations requires a sample that is a census. If there are 100 organisations or cases, we need to sample 80 of them, or if the hard-to-access set numbers 1000, we need to access well over a quarter of them. There are very few instances of qualitative work that covers either 10 organisations or in excess of 200 in-depth interviews. In cases which do come readily to mind like the life-work of Michael Burawoy (e.g. 1991), or Savage et al.'s (2001) study of class consciousness in Manchester, the relationship between sample and population is not entirely clear. To express it the other way round, if we have the resources for, say, 40 in-depth interviews, the generalising power under probability is to a population no bigger than 50.

If qualitative research cannot draw on probability theory, how should field sites be selected? Under present conventions, there is little rigour in sampling. Where this involves access to an organisation or an established group, the 'selection' is normally done on the basis of a previous contact. Supervisors of postgraduate students act as patrons, activating their social networks to sponsor their protégées to gate-keepers. Researchers' prior work experience or contacts provide another pragmatic route to access, but usually with little consideration of the theoretical implications of researching a specific site. A similar pragmatism applies in contacting hard-to-reach informants, where snow-balling is seldom seriously questioned in terms of how non-typical the resulting sample is likely to be. We are rarely told in detail about the participants in focus groups.

This does not matter if we assume that the sample of persons makes no difference to the information collected, but this is hardly a strong position to adopt. The findings of studies are very commonly used to generalise – or what in qualitative discourse is more often referred to as 'developing theories' – as Payne and Williams (2005) have shown. Who we choose for our sample determines our findings:

> Questions of generalisation are tied to those of *sampling* because the sample is the bearer of those characteristics that it is wished to infer to a wider population. (Williams, 2000: 216, original emphasis)

To give one illustration, if a study site has been chosen 'because of the willingness of local authority access officers to take part in the research' (Edwards and Imrie, 2003: 245), it is not unreasonable to ask whether the same results would have been found had access somehow been possible to sites where access officers were *unwilling* to co-operate? Is willingness to co-operate a social characteristic associated with others like support for evidence-based practice, openness to new ideas, and confidence in the high quality of one's organisation?

Most qualitative research falls into this category described by Mason (1996: 92–93), as being based on small selections of units which are acknowledged to be part of wider universes but not chosen primarily to represent them directly. There are far too few cases where the case selection – especially of the single case study – is explicitly justified by the need for very detailed scrutiny. It would be much better to reflect on generalization *ex ante*, which means to select the single case, or the site, following prior considerations. Looking for a typical case, a representative case, a frequent occurring case or a theoretically interesting case would be a good strategy (Mayring, 2007: 6). That means that it would be good to formulate a case definition as part of the research design rather than burying what turns out to be a 'pragmatic solution' in a brief discussion of methods. A sample need not be typical of the wider population, but its a typicality needs to be confronted, justified in advance, and subsequently integrated into any attempt to 'generate theory'.

SOME CONCLUDING REMARKS

Doing empirical research is central to the being and identity of most sociologists. That does not mean that we are as good at it as we like to think, or that we follow through the implications of our focus on research. I have illustrated several ways in which we can and sometimes do fall short: by using data which are not fit for purpose (particularly with respect to the current fashion for secondary analysis of large data-sets); by not exercising our responsibilities as teachers and supervisors of research; by making excessive claims from limited studies; by over-estimating informants willingness to co-operate and under-estimating the significance of non-response, because we have under-theorised the interview encounter and over-theorised the atypical case. This is a case for methodological innovation less as specific new techniques, and more as basic thinking about interviewing and sampling.

As one example of where this might lead, (I have tried in my exposition to indicate the limits of my own *illustrations* and sampled evidence when I have been making broad generalisations) is something that sociologists should be learning this kind of caution while they are undergraduates. The best long-term investment in producing good research is to reform the undergraduate curriculum, placing research practice not only at its heart but in every corner of its corpus. The issue is not which methods should be taught, but that the research itself should occupy its proper place. Such an innovation would not only improve our technical performance, but also provide a framework in which future generations of sociologists could start reflecting about the research act, in fundamental ways.

It seems to me that there is currently a need for a little more humility among researchers. While we need endlessly beat ourselves up about the limitations of our efforts, much of our writing tends too far in the opposite direction. Faced with perceived threats from rival methodological stances, we not only make excessive claims for our own findings, but become needlessly drawn into denigrating the work of others based in those different methodological camps. Much of this dialogue of the deaf fails to see the simple point that we can describe the social at different *levels of detail.* The broad view of national patterns offers one perspective, gaining society-wide knowledge and generality at the expense of detail. The interpretation of small-scale interactions provides us with accounts of fine social detail and process, but cannot at the same time characterise the wider picture. Temperamentally, we may prefer one level of analysis to the other, or become interested in topics best-suited for this or that method of research, but that should not become routinised in our discipline's culture as exclusionary or hegemonic practice.

Core elements such as sampling and interviewing, which are fundamental to a range of research styles, raise problems throughout sociology that extend beyond the narrow confines of our own preferred approaches. Unless we continue to refine our tools and perspectives by drawing on our *collective* expertises, we stand to lose our claims to provide valuable and uniquely sociological knowledge.

Indeed, when we take seriously the rise of geodemographic systems, transactional databases, and the extension of State-controlled inter-connected personal records (in the name of national security and administrative efficiency), the need to be more aware of what we know and how we know it becomes more acute. How else are we to acquire the sophisticated descriptions required to challenge the results of the 'rules of thumb' (Savage and Burrows, 2007: 887) routinely applied by the managers of commercial databases or the security services? This is a task which can only be achieved by greater co-operation within the discipline. There are many rooms in sociology's mansion, and we can still learn a lot from talking with the guys next-door.

NOTES

1 Readers who share my moderately obsessive disposition might care to check on the frequent mis-referencing of Wrong's original article: I estimate about 80% of Google hits (January 2009) write 'oversocialized' as two words or hyphenated; mis-spell socialized with an s not a z; refer to the concept, not the conception; or even more remarkably attribute the bibliographic source to the AJS instead of the ASR).

REFERENCES

Burawoy, M. (1991) 'The extended case study method'. In M. Burawoy (ed.) *Ethnography Unbound.* Berkley: University of California Press.

Denzin, N. (1983) 'Interpretive interactionism'. In G. Morgan (ed.) *Beyond Method: Strategies for Social Research*. Beverly Hills: Sage.

Edwards, C. and Imrie, R. (2003) 'Disability and bodies as bearers of value'. *Sociology*, 37(2): 239–56.

Eggins, H. (ed.) (1988) *Review of Sociology Courses and Teaching.* London: CNAA.

Fahrenberg, J. (2003) 'Interpretation in psychology and social science – New approach or a neglected tradition?' *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 4(2), Art. 45. http://www.qualitative-research.net/fqs-texte/2-03/2-03fahrenberg-d.htm (Accessed 22.12.2008).

Gillman, R. (2006) *Current Practices in Quantitative Literacy.* Washington, D.C.: Mathematics Association of America.

Healey, M. (2008) *Linking Research and Teaching: A Selected Bibliography.* University of Gloucestershire. http://www.glos.ac.uk/shareddata/dms/5B8F006BBCD42A039BA6832547DBAC47.pdf (accessed 16.7.2008).

Hope, K. and Goldthorpe, J. (1974) *The Social Grading of Occupations.* Oxford: Oxford University Press.

Howery, C. and Rodriguez, H.(2006) 'Integrating Data Analysis (IDA): Working with Sociology Departments to address the quantitative literacy gap'. *Teaching Sociology*, 16: 23–38.

Iganski, P., Payne, G., and Payne, J. (2001) 'Social Inclusion? the changing position of Britain's Minority Ethnic Groups'. *International Journal of Sociology and Social Policy*, 21(4–6): 184–211.

Krejcie, R. and Morgan, D. (1970) 'Determining sample size for research activities'. *Educational and Psychological Measurement*, 30: 607–10.

Lincoln, Y. and Guba, E. (1985) *Naturalistic Inquiry*. Beverly Hills, CA: Sage.

Mason, J. (1996) *Qualitative Researching.* Thousand Oaks, CA: Sage.

May, C. (2005) 'Methodological pluralism, British sociology and the evidenced-based state'. *Sociology*, 39(3): 519–28.

Mayring, P. (2007) 'On generalization in qualitatively oriented research'. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 8(3): Art. 26. http://nbn-resolving.de/urn:nbn:de:0114-fqs0703262. (accessed 22.12.2008).

Parker, J. (2011) 'Best practices in quantitative methods teaching: comparing social science curricula across countries '. In G. Payne and M. Williams (eds) *Teaching Quantitative Methods*. London: Sage (in press).

Payne, G. and Williams, M. (2005) 'Generalisation in quantitative research'. *Sociology*, 39(2): 295–314.

Payne, G., Lyon, S., and Anderson, R. (1988) 'The teaching of social research methods'. In H. Eggins (ed.) *Review of Sociology Courses and Teaching.* London: CNAA.

Payne, G., Lyon, S., and Anderson, R. (1989) 'Research methods in the public sector curriculum'. *Sociology*, 23(2).

Payne, G., Williams, M., and Chamberlain, S. (2004) 'Methodological pluralism in British sociology'. *Sociology*, 38(1): 153–63.

Savage, M., Bagnall, G., and Longhurst, B. (2001) 'Ordinary, ambivalent and defensive: class identities in the north-west of England'. *Sociology,* 35(4): 875–92.

Savage, M. and Burrows, R. (2007) 'The coming crisis of empirical sociolgy'. *Sociology,* 41(5): 885–99.

Williams, M. (2000) 'Interpretivism and generalisation'. *Sociology*, 34(2): 209–24.

Williams, M., Hodgkinson, L., and Payne, G. (2004) 'A crisis of number?' *Radical Statistics*, 85: 40–54.

Williams, M., Payne, G., Hodgkinson, L., and Poade, D. (2008) 'Does British sociology count? Sociology students' attitudes toward quantitative methods'. *Sociology*, 42(5): 1003–21.

3

Feminist Methodology

Gayle Letherby

INTRODUCTION

As many feminist researchers have argued and Sandra Harding (1987: 3) succinctly puts, 'it is not by looking at research methods that one will be able to identify the distinctive features of the best feminist research' (see also, e.g. Roberts, (1981) 1990; Stanley, 1990; Reinharz, 1992; Maynard and Purvis, 1994; Stanley and Wise, 1993; Millen, 1997; Ribbens and Edwards, 1998; Ramazanoglu with Holland, 2002; Letherby, 2002, 2003a; Hesse-Biber, 2007a). So, there is no such thing as a feminist method, rather what is distinctive about feminist research is a sensitivity to the significance of gender within society and a critical approach to the research process. Thus, rather than focusing on the methods (tools for gathering evidence/collecting data), feminists are concerned with the methodological reflection of the researcher(s). Methodology, concerned as it is with the 'getting of knowledge', is key to understanding the relationship between knowledge and power (Stanley and Wise, 2008: 222) and is thus an essential part of the feminist project:

> Within feminism, the term 'feminist methodology' is also used to describe an ideal approach to doing research – one which is respectful of respondents and acknowledges the subjective involvement of the researcher. This leads us to a question which Cook and Fonow (1990: 71) ask: 'is feminist methodology that which feminists *do* or that which we *aim for*'? (Letherby, 2003a: 5)

> Feminist methodology is at the heart of the feminist project of changing the world because it is the focal point for bringing together theory, practical research methods, and the production of new knowledge. (Stanley and Wise, 2008: 221)

so on; (b) informants are given a list of possible answers to a question and asked to choose among them (this is *aided recall*); and (c) informants are interviewed periodically, reminded what they said last time in answer to a question, and asked about their behavior since their last report (this is *bounded recall*).

Having informants consult their records does not always produce the results you might expect. Horn (1960) asked people to report their bank balance. Of those who did not consult their records, 31% reported correctly. Those who consulted their records did better, but not by much. Only 47% reported correctly (reported in Bradburn, 1983:309).

Aided recall appears to increase the number of events recalled, but also appears to increase the telescoping effect (ibid.). Bounded recall corrects for telescoping but does not increase the number of events recalled, and in any event is only useful in studies where the same informants are interviewed again and again. The problem of informant accuracy remains an important issue and a fruitful area for research in social science methodology. For more on this problem, see Neter and Waksberg (1964), Linton (1975), Moss and Goldstein (1979), Bernard et al. (1984), Bradburn et al. (1987), Freeman and Romney (1987), and McNabb (1990).

11

Structured Interviewing

wichtig: viele Methoden müssen
durch nachfolgende
Interviews ergänzt werden

Structured interviewing involves exposing every informant in a sample to the same stimuli. The stimuli may be a set of questions or they may be a list of names, a set of photographs, a table full of artifacts, a garden full of plants, etc. The idea is to control the input that triggers each informant's responses so that the output can be reliably compared. how?

The most common form of structured interviewing is the questionnaire. A questionnaire may be self-administered, or it may be administered over the phone or in person, but in all cases the questions posed to informants are the same. I'll deal with the building and administering of questionnaires in the next chapter. This chapter is an introduction to some exciting new systematic interviewing techniques that are being used in *cultural domain analysis*.

Two things make these methods very productive for anthropological fieldwork. First, they are fun to use and informants find them fun to respond to. Second, the ANTHROPAC computer software (Borgatti, 1992a) has made it much easier than it was just a few years ago to collect and analyze data using these techniques.

Cognitive Anthropology and Cultural Domains

Def.

Cognitive anthropology is the study of how peoples of different cultures acquire information about the world (cultural transmission), how they process that information and reach decisions, and how they act on that information in ways that other members of their culture consider appropriate.

Modern cognitive anthropology traces its roots to 1956 with Ward Goodenough's application of the *emic* and *etic* principle from linguistics to other areas of culture. The emic-etic principle in linguistics was named by the linguist Kenneth Pike (1956, 1967). It illustrates the fact that human beings distinguish phon*emes* (the basic set of underlying constructs that generate the sounds of a language) from their phon*etic* representations (what we actually hear). Many phonetic outcomes might be accepted by native speakers of a language as being representative of a single underlying phoneme.

In English, for example, we have an aspirated *t*, as in "tough," and an unaspirated *t*, as in "spit." (You can distinguish the aspiration by putting your hand up to your mouth and feeling the breath of air that the *t* in "tough" makes as you say it. The *t* in "sit" doesn't do that.) There are no contexts in English in which the acoustical feature of aspiration changes the meaning of a word.

Suppose, though, that in another language the *t* in "tough" and the *t* in "spit" were the *only* difference in the two words "t^{h}ao" and "tao," where the first meant "one million" and the second meant "the axle of an oxcart." (The raised *h* is for the aspiration.) In that case, the *distinctive feature* of aspiration would be meaningful in that particular language.

Goodenough's insight was that this principle could be applied to areas of culture other than phonology. An adequate ethnographic description of the named category "cousin," for example, would consist of stating the emic rules that people use when they decide whether two people are cousins (1956:195).

Etically, you have eight kinds of cousins: The male and female offspring of the male or female siblings of your male or female parents are all your "cousins" in English. In other cultures, the male and female offspring of the male or female siblings of your father are one kind of relative, while the male or female offspring of the male or female siblings of your mother are another kind of relative.

By emic rules, there are many different possible packages of cousins that might be defined. The general research strategy that grew from this insight was dubbed *ethnoscience*—the search for the grammars of behavior in the cultures of the world, and the underlying principles that govern how those grammars differ.

Grammars consist of rules that people carry around in their heads—rules that let them understand brand new sentences they've never heard before and make up new ones that other people understand. This fundamental idea continues to capture the imagination of many ethnographers. The messy, noisy cultural behavior at the observable surface is treated as being driven by a relatively clean set of underlying rules, just as the infinite number of grammatical utterances can be accounted for by a large, but finite set of grammatical rules.

Soon after this principle was articulated, anthropologists began to apply it to what was called "cultural domains"—kinship terms, plants, animals, occupations, and so on—anything that could be listed by informants (Tyler, 1969).

The spectrum of colors, for example, has a single etic reality (you can see the spectrum on a machine) and many emic realities. Several American Indian peoples identify a color we gloss as "grue." The word covers all the colors across the etic spectrum of green and blue.

This does not mean that people who use a word like "grue" fail to see the difference between things that are the color of grass and things that are the color of a clear blue sky. They just label different chunks of the etic spectrum of colors than we do. If this seems exotic to you, get a list of, say, 100 lipstick colors and ask college women and men to describe all those colors. You may find that, on average, women recognize many more colors than do men.

Today, anthropologists are studying many interesting domains—things people do on weekends, ways people believe they can succeed in business, traits that people think of when they think of particular ethnic groups, categories of fast foods, and more. The goal is to understand what people think, how they think it, and how they organize the material. The challenge is to devise methods that get at these things and that produce data that can be checked for their reliability and validity. The most common techniques for gathering data in cognitive anthropology are: free listings, frame elicitations, triad tests, pile sorts, paired comparisons, and rank order tests (Weller & Romney, 1988).

Free Listing

Free listing is a deceptively simple but powerful technique. It is generally used to study a cultural domain. The list of the days in a week is a cultural domain, with no intracultural variation. Everyone knows the same list. "Things to do on vacation" is a much richer domain, with lots of intracultural variation. Common domains studied by anthropologists are things like diseases, plants, occupations, and animals. But you can just as easily study how people classify names of movie stars, brands of computers, types of machines, or titles of anthropology articles.

In free listing, you tell informants: "Please list all the X you know about" or ask them "What kinds of X are there?"

Henley (1969) asked 21 adult Americans (students at Johns Hopkins University) to name as many animals as they could in 10 minutes. You'd be surprised at how much Henley learned from this simple experiment. First of all, there is an enormous variety of expertise in the culture when it comes to naming animals. In just this small group of informants (which didn't even represent the population of Johns Hopkins University, much less that of Baltimore or of the United States), the lists ranged in length from 21 to 110, with a median of 55.

In fact, those 21 people named 423 different animals, and 175 were mentioned just once. The most popular animals for this group of informants were: dog, lion, cat, horse, and tiger, all of which were named by more than 90% of informants. Only 29 animals were listed by more than half the informants, but 90% of those were mammals. By contrast, among the 175 animals named only once, just 27% were mammals.

But there's more. Previous research had shown that the 12 most commonly talked about animals in U.S. speech are: bear, cat, cow, deer, dog, goat, horse, lion, tiger, mouse, pig, and rabbit. There are $N (N - 1)/2$, or 66 possible unique pairs of 12 animals (dog-cat, dog-deer, horse-lion, mouse-pig, etc.). Henley examined each informant's list of animals, and for each of the 66 pairs found the difference in order of listing.

That is, if an informant mentioned goats 12th on her list and bears 32nd, then the distance between goats and bears, for that informant, was $32 - 12 = 20$. This distance was standardized: It was divided by the length of the informant's list and multiplied by 100. Then Henley calculated the mean distance, over all the informants, for each of the 66 pairs of animals.

The lowest mean distance was between sheep and goats (1.8), and the highest was between cats and deer (56.1). Deer are related to all the other animals on the list by at least 40 units of distance, except for rabbits, which are only 20 units away from deer. Cats and dogs are only 2 units apart, while mice and sheep are nearly 52 units from each other. This experiment, too, needs to be replicated in other components of American culture and in other cultures.

Robert Trotter (1981) reports on 378 Mexican-Americans who were asked to name the *remedios caseros* (home remedies) they knew, and what illnesses the remedies were for. Informants listed a total of 510 remedies for treating 198 illnesses. However, the 25 most frequently mentioned remedies (about 5% of the total) constituted about 41% of all the cases, and the 70 most frequently mentioned illnesses (about 36%) constituted 84% of the cases.

Trotter's free-list data reveal a lot about Mexican-American perceptions of illness and home cures. He was able to count which ailments were reported more frequently by men and which by women; which ailments were reported more frequently by older people and by younger people; which by those born in Mexico and those born in the United States; and so on.

Free listing is often a prelude to cluster analysis and multidimensional scaling, which we'll get to in Chapter 20. But you can learn an awful lot from *just* a free list. Gatewood (1983a) asked 40 adult Pennsylvanians to name all the trees they could think of. Then he asked them to check the trees on their list that they thought they could recognize in the wild. Thirty-seven people (out of 40) listed "oak," 34 listed "pine," 33 listed "maple," and 31 listed "birch." I suspect that the list of trees and what people say they could recognize would look rather different in, say Wyoming or Mississippi. We could test that.

Thirty-one of the 34 who listed "pine" said they could recognize a pine. Twenty-seven people listed "orange," but only four people said they could recognize an orange tree (without oranges hanging all over it, of course). On average, the Pennsylvanians in Gatewood's sample said they could recognize about 50% of the trees they listed. Gatewood calls this the *loose talk* phenomenon. He thinks that many Americans can name a lot more things than they can recognize in nature.

Does this loose-talk phenomenon vary by gender? Suppose, Gatewood says, we ask Americans from a variety of subcultures and occupations to list other things besides trees. Would the 50% recognition rate hold?

Gatewood and a group of students at Lehigh University interviewed 54 informants—27 men and 27 women, all university students. The informants free listed all the musical instruments, fabrics, hand tools, and trees they could think of. Then the informants were asked to check off the items in each of their lists that they thought they would recognize in a natural setting.

All of Gatewood's hypotheses were supported: Men and women named about the same number of musical instruments; women named more fabrics; men named more hand tools; and both men and women named more trees than they could identify (Gatewood, 1984).

Romney and D'Andrade asked 105 U.S. high school students to "list all the names for kinds of relatives and family members you can think of in

English" (1964:155). They were able to do a large number of analyses on these data. For example, they studied the order and frequency of recall of certain terms and the productiveness of modifiers, such as "step-," "half-," "-in-law," "grand-," "great," and so on. They assumed that the nearer to the beginning of a list that a kin term occurs, the more salient it is for that particular informant. By taking the average position in all the lists for each kin term, they were able to derive a rank order list of kin terms, according to the variable's saliency.

They also assumed that more salient terms occur more frequently. So, for example, "mother" occurs in 93% of all lists and is the first term mentioned on most lists. At the other end of the spectrum is "grandson," which was only mentioned by 17% of the 105 informants, and was, on average, the 15th, or last term to be listed. They found that the terms "son" and "daughter" occur on only about 30% of the lists. But remember, these informants were all high school students. It would be interesting to repeat Romney and D'Andrade's experiment on many different U.S. populations. We could then test the saliency of English kin terms on the many subpopulations.

Finally, free listing can be used to find out where to concentrate effort in applied research—that is, as part of a rapid assessment approach. In a recent project, a team of anthropologists and sociologists studied how people on the North Carolina coast viewed the possibility of offshore oil drilling. During regular interviews, the field researchers asked people "What are the things that make life good around here?"

The researchers decided to ask this question after preliminary ethnographic research (hanging out, talking informally to a lot of people) in seven small, seaside towns. People kept saying what a "nice little town this is" and "What a shame it would be if things changed around here." Informants had no difficulty with the question, and after just 20 interviews, the researchers had a list of over 50 "things that make life good around here." The researchers chose the 20 items mentioned by at least 12 informants and explored the meaning of those items further (ICMR et al., 1993).

The humble free list has many uses. Use it a lot.

The True-False/Yes-No and Sentence Frame Techniques

Another common technique in cultural domain analysis is called the *sentence frame* or *frame elicitation* method (see also Chapter 16 on qualitative analysis and native taxonomies). The method produces true-false, or yes-no, data.

The frame elicitation method has been used extensively in anthropology [t]o study the distribution of beliefs about the causes of and cures for [i]llnesses (Fabrega, 1970; D'Andrade et al., 1972).

Linda Garro (1986) used the frame elicitation method to compare the [k]nowledge of curers and noncurers in Pichátaro, Mexico. She used a list [o]f 18 illness terms and 22 causes, based on prior research in Pichátaro (Young, 1978). The frames were questions, like "Can ______ come from ______?" Garro substituted names of illnesses in the first blank, and [t]hings like "anger," "cold," "overeating," and so on in the second blank. (ANTHROPAC has a routine for building questionnaires of this type.) This produced an 18 × 22 yes-no matrix for each of the informants. The [m]atrices could then be added together and submitted to analysis by multidimensional scaling (see Chapter 20).

James Boster and Jeffrey Johnson (1989) used the frame substitution method in their study of how recreational fishermen in the United States categorize ocean fish. They asked 120 fishermen to consider 62 belief frames, scan down a list of 43 fish (tarpon, silver perch, Spanish mackerel, etc.), and pick out the fish that fit each frame. Here are a few of the belief frames:

The meat from ________ is oily tasting.
It is hard to clean ________ .
I prefer to catch ________ .

That's 43 × 62 = 2,666 judgments by each of 120 informants, but informants were usually able to do the task in less than half an hour (Jeffrey Johnson, personal communication). The 62 frames, by the way, came straight out of ethnographic interviews where informants were asked to list fish and to talk about the characteristics of those fish.

Gillian Sankoff (1971) studied land tenure and kinship among the Buang, a mountain people of northeastern New Guinea. The most important unit of social organization among the Buang is the *dgwa*, a kind of descent group, like a clan. Sankoff wanted to figure out the very complicated system by which men in the village of Mambump identified with various dgwa and with various named garden plots.

The Buang system was apparently too complex for bureaucrats to fathom, so in order to save administrators a lot of trouble, the men of Mambump had years earlier devised a simplified system that they presented to outsiders. Instead of claiming that they had ties with one or more of five different dgwa, they each decided which of the two largest dgwa they would belong to, and that was that as far as the New Guinea administration knew.

To unravel the complex system of land tenure and descent, Sankoff made a list of all 47 men in the village and all 140 yam plots that they had used over the recent past. Sankoff asked each man to go through the list of men and identify which dgwa each man belonged to. If a man belonged to more than one, then Sankoff got that information, too. She also asked her informants to identify which dgwa each of the 140 garden plots belonged to.

As you might imagine, there was considerable variability in the data. Only a few men were uniformly placed into one of the five dgwa by their peers. But by analyzing the matrices of dgwa membership and land use, Sankoff was able to determine the core members and peripheral members of the various dgwa.

She was also able to ask important questions about intracultural variability. She looked at the variation in cognitive models among the Buang for how land use and membership in descent groups were related. Sankoff's analysis was an important milestone in our understanding of the measurable differences between individual culture versus shared culture. It supported Goodenough's notion (1965) that cognitive models are based on shared assumptions, but that ultimately they are best construed as properties of individuals.

True-false and yes-no tests that generate nominal data are easy to construct (especially with computer programs) and can be administered to a large number of informants. Frame elicitation in general, however, can be quite boring, both to the informant and to the researcher alike. Imagine, for example, a list of 25 animals (mice, dogs, antelopes . . .), and 25 attributes (ferocious, edible, nocturnal . . .).

The structured interview that results from such a test involves a total of 625 (25 × 25) questions to which an informant must respond—questions like "Is an antelope edible?" "Is a dog nocturnal?" "Is a mouse ferocious?" Informants can get pretty exasperated with this kind of silliness. Be careful, therefore, about cultural relevance when doing frame elicitations and true-false tests. It is essential to have a good ethnographic grounding in the local culture in order to select domains, items, and attributes that make sense to people.

Triad Tests

In a *triad test*, you show informants three things and tell them to "choose the one that doesn't fit" or "choose the two that seem to go together best," or "choose the two that are the same." The "things" can be photographs,

actual plants, 3 × 5 cards with names of people on them, or whatever. (Informants often ask "what do you mean by things being 'the same' or 'fitting together'?" Tell them that you are interested in what *they* think that means.) By doing this for all triples from a list of things or concepts, you can explore differences in cognition among individuals, and among cultures and subcultures.

Suppose you ask a group of Americans to "choose the item that is least like the other two" in each of the following triads:

WHALE DOLPHIN MOOSE
SHARK DOLPHIN MOOSE

All three items in the first triad are mammals, but two of them are sea mammals. A few people would choose "dolphin" as the odd item because "whales and moose are both big mammals and the dolphin is smaller." Most people I know, however, would choose "moose" as the most different. In the second triad, many of the same people who chose "moose" in triad 1 will choose "shark" because moose and dolphins are both mammals, and a shark is not.

But some people who chose "moose" in triad 1 will choose "moose" again in triad 2 because sharks and dolphins are sea creatures, while moose are not. Giving informants a judiciously chosen set of triad stimuli can help you understand interindividual similarities and differences in how people think about the items in a cultural domain.

The triads test was developed in psychology (see Kelly, 1955; Torgerson, 1958) and was introduced into anthropology by Romney and D'Andrade (1964). They presented informants with triads of U.S. kinship terms and asked them to choose the term that was most dissimilar in each triad. For example, when they presented informants with the triad "father, son, nephew," 67% selected "nephew" as the most different of the three items. Twenty-two percent chose "father," and only 2% chose "son."

They also interviewed informants and asked them about their reasons for choosing an item on a triad test. For the triad "grandson, brother, father," for example, one informant said that a "grandson is most different because he is moved down further" (Romney & D'Andrade, 1964:161). There's a lot of ethnographic wisdom in that statement.

By studying which pairs of kinship terms their informants chose most often as being similar, Romney and D'Andrade were able to isolate some of the salient components of the American kinship system (components such as male versus female, ascending versus descending generation, etc.).

They were able to do this, at least, for the group of informants they used. Repeating their tests on other populations of Americans, or on the same population over time, would yield interesting comparisons of anthropological significance.

Lieberman and Dressler (1977) used triad tests to examine intracultural variation in ethnomedical beliefs on the Caribbean island of St. Lucia. They wanted to know if cognition of disease terms varied with bilingual proficiency. They used 52 bilingual English-Patois speakers and 10 monolingual Patois speakers. From ethnographic interviewing and cross-checking against various informants, they isolated nine disease terms that were important to St. Lucians.

Here's the formula for finding the number of triads in a list of N items:

$$\text{The number of triads in } N \text{ items} = \frac{N(N-1)(N-2)}{6}$$

In this case, N is nine, so there are 84 possible triads.

Lieberman and Dressler gave each of the 52 bilingual informants two triad tests, a week apart: one in Patois and one in English. (Naturally, they randomized the order of the items within each triad and also randomized the order of presentation of the triads to informants.) They also measured how bilingual their informants were, using a standard test. The 10 monolingual Patois informants were simply given the triad test.

The researchers counted the number of times that each possible pair of terms was chosen as most alike among the 84 triads. (There are $N(N-1)/2$ pairs or $(9 \times 8)/2 = 36$ pairs). They divided the total by seven (the maximum number of times that any pair appears in the 84 triads). This produced a similarity coefficient, varying between 0.0 and 1.0, for each possible pair of disease terms. The larger the coefficient for a pair of terms, the closer in meaning are the two terms. They were then able to analyze these data among English-dominant, Patois-dominant, and monolingual Patois speakers.

It turned out that when Patois-dominant and English-dominant informants took the triad test in English, their cognitive models of similarities among diseases was similar. When Patois-dominant speakers took the Patois-language triad test, however, their cognitive model was similar to that of monolingual Patois informants.

This is a very interesting finding. It means that Patois-dominant bilinguals manage to hold on to two distinct psychological models about diseases, and that they switch back and forth between them, depending on what language they are speaking. By contrast, the English-dominant group

displayed a similar cognitive model of disease terms, irrespective of the language in which they are tested.

The Balanced Incomplete Block Design for Triad Tests

Anthropologists have used the triad test to study occupations (Burton, 1972), personality traits (Kirk & Burton, 1977), and other domains of culture. Typically, the terms that go into a triad test are generated by a free list, and typically the list is much too long for a triad test. There are 84 stimuli in a triad test containing 9 items. But with just 6 more items the number of decisions an informant has to make jumps to 455. At 20 items, it's a mind-numbing 1,140.

Free lists of illnesses, ways to prevent pregnancy, advantages of breast-feeding, places to go on vacation, and so on easily produce 60 items or more. Even a selected, abbreviated list may be 20 items.

This led Burton and Nerlove (1976) to develop the *balanced incomplete block* design, or BIB, for the triad test. BIBs take advantage of the fact that there is a lot of redundancy in a triad test. Suppose you have just four items, 1, 2, 3, 4, and you ask informants to tell you something about *pairs* of these items (e.g., if the items were vegetables, you might ask "Which of these two is less expensive?" or "Which of these two is more nutritious?" or whatever). There are exactly six pairs of four items (1-2, 1-3, 1-4, 2-3, 2-4, 3-4), and the informant sees each pair just once.

But suppose that instead of pairs you show the informant triads and ask which two out of each triple are most similar. There are just four triads in four items (1-2-3, 1-2-4, 2-3-4, 1-3-4), but each item appears $(N-1)(N-2)/2$ times, and each pair appears $N-2$ times. For four items, there are $N(N-1)/2$ = six pairs; each pair appears twice in four triads, and each item on the list appears three times.

It is all this redundancy that reduces the number of triads needed in a triads test. If you want each pair to appear just once (called a *lambda 1* design), instead of seven times in a triads test involving nine items, then, instead of 84 triads, only 12 are needed. If you want each pair to appear just twice (a *lambda 2* design), then 24 triads are needed. For analysis, lambda 2 designs are much better than lambda 1's. Table 11.1 shows the lambda 2 design for 9 and 10 items.

For 10 items, a lambda 2 design requires 30 triads; for 13 items, it requires 52 triads; for 19 items, 114 triads; and for 25 items, 200 triads. In literate societies, most informants can respond to 200 triads in less than 15 minutes.

TABLE 11.1
Balanced Incomplete Block Designs for Triad Tests Involving 9 and 10 Items

For 9 items, 24 triads are needed, as follows:

Items:	1, 5, 9	1, 2, 3
	2, 3, 8	4, 5, 6
	4, 6, 7	7, 8, 9
	2, 6, 9	1, 4, 7
	1, 3, 4	2, 5, 9
	5, 7, 8	3, 6, 8
	3, 7, 9	1, 6, 9
	2, 4, 5	2, 4, 8
	1, 6, 8	3, 5, 7
	4, 8, 9	1, 5, 8
	3, 5, 6	2, 6, 8
	1, 2, 7	3, 4, 9

For 10 items, 30 triads are needed, as follows:

1, 2, 3	9, 3, 10	7, 10, 3	5, 6, 3
2, 5, 8	10, 6, 5	8, 1, 10	6, 1, 8
3, 7, 4	1, 2, 4	9, 5, 2	7, 9, 2
4, 1, 6	2, 3, 6	10, 6, 7	8, 4, 7
5, 8, 7	2, 4, 8	1, 3, 5	9, 10, 1
6, 4, 9	4, 9, 5	2, 7, 6	10, 5, 4
7, 9, 1	5, 7, 1	3, 8, 9	
8, 10, 2	6, 8, 9	4, 2, 10	

SOURCE: Burton and Nerlove (1976).
NOTE: These are lambda 2 designs. See text for explanation.

Unfortunately, there is no easy formula for choosing *which* triads in a large set to select for a BIB. Fortunately, Burton and Nerlove (1976) worked out various lambda BIB designs for up to 21 items, and Stephen Borgatti has incorporated BIB designs into ANTHROPAC (1992a). You simply tell ANTHROPAC the list of items you have, select a design, and tell it the number of informants you want to interview. ANTHROPAC then prints out a randomized triad test, one for each informant. (Randomizing the order in which the triads appear to informants eliminates *order-effects*—possible biases that come from responding to a list of stimuli in a particular order.)

Boster et al. (1987) used a triad test and a pile sort in their study of the social network of an office. There were 16 employees, so there were 16 "items" in the cultural domain ("the list of all the people who work here" is a perfectly good domain). A lambda 2 test with 16 items has 80 distinct triads. Informants were asked to "judge which of three actors was the most different from the other two."

Triad tests are easy to create with ANTHROPAC, easy to administer, and easy to score, but they can only be used when you have relatively few items in a cultural domain. Also, informants sometimes find triad tests to be boring. Use the triad test method when you have just a few items in a domain. Use the pile sort method to look at the cognitive organization of a large cultural domain. I find that informants easily handle lambda 2 triad tests with 9 to 15 items, and pile sorts with 40 to 60 items.

Pile Sorts

Typically, pile sorts are done with cards or slips of paper. Each card has the name of a thing or a concept written on it. Once again, the items are gleaned from a free list that defines a cultural domain. Informants are asked to "sort these cards into piles, putting things that are similar together in a pile."

Two questions that informants often ask are: "What do you mean by 'similar'?" and "Can I put something in more than one pile?" The answer to the first question is "Well, whatever you think is similar. We want to learn what you think about these things. There are no right or wrong answers."

The easy answer to the second question is "no," just because there is one card per item and a card can only be in one pile at a time. This answer cuts off a lot of information, however, because informants can think of items in a cultural domain along several dimensions at once. For example, in a pile sort of consumer electronics, an informant might want to put a VCR in one pile with TVs (for the obvious association) and in another pile with camcorders (for another obvious association) but might not want to put camcorders and TVs in the same pile. You can make up a duplicate card on the spot if you want and, fortunately, the ANTHROPAC software makes easy work of handling this during analysis. An alternative is to ask the informant to do *multiple pile sorts* of the same object.

Free Pile Sorts and the Lumper and Splitter Problem

Most researchers use the *free* or *unconstrained* pile sort method, where informants are told that they can make as many piles as they want, so long as they don't make a separate pile for each item or lump the items into one pile. Like the triad test, the free pile sort presents a common set of stimuli to informants. But here, the informants manage the information and put the items together as they see fit. The result is that some informants will make many piles, others will make few. This is known as the *lumper-splitter problem* (Weller & Romney, 1988:22).

In a pile sort of animals, for example, some informants will put all the following together: giraffe, elephant, rhinoceros, zebra, wildebeest. They'll explain that these are the "African animals." Others will put giraffe, elephant, and rhino in one pile, and the zebra and wildebeest in another, explaining that one is the "large African animal" pile and the other is the "medium-sized African animal" pile.

Some informants will have singleton piles, explaining that each singleton is unique and doesn't go with the others. It's fine to ask informants why they made each pile of items, but wait until they finish the sorting task so you don't interfere with their concentration. And don't hover over informants. Find an excuse to walk away for a couple of minutes after they get the hang of it.

Pile Sorts with Objects

Although pile sorts are typically done with cards or slips of paper, they can also be done with objects. James Boster (1987) studied the structure of the domain of birds among the Aguaruna Jívaro of Peru. He paid people to bring him specimens of birds and he had the birds stuffed. He built a huge table out in the open, laid the birds on the table, and asked the Aguaruna to sort the birds into groups.

Carl Kendall led a team project in El Progreso, Honduras, to study beliefs about dengue fever (Kendall et al., 1990). Part of their study involved a pile sort of the nine most common flying insects in the region. They mounted specimens of the insects in little boxes and asked people to group the insects in terms of "those that are similar." Some anthropologists have used photographs of objects as stimuli for a pile sort.

Borgatti (1992b:6) points out that asking an informant to sort photographs of objects (or actual objects) rather than cards with the names of object can produce different results. Imagine sorting 30 photographs of automobiles—sports cars, pickup trucks, minivans, etc. Seeing the photos, you might classify the vehicles on the basis of physical form or function. If you sorted cards with stimuli like "Alpha Romeo coupe," "Dodge minivan," "Mercedes sedan," and so on, you might do the sort on other criteria, like price, prestige, desirability. "If you are after shared cultural beliefs," says Borgatti, "I recommend keeping the stimulus as abstract as possible" (ibid.).

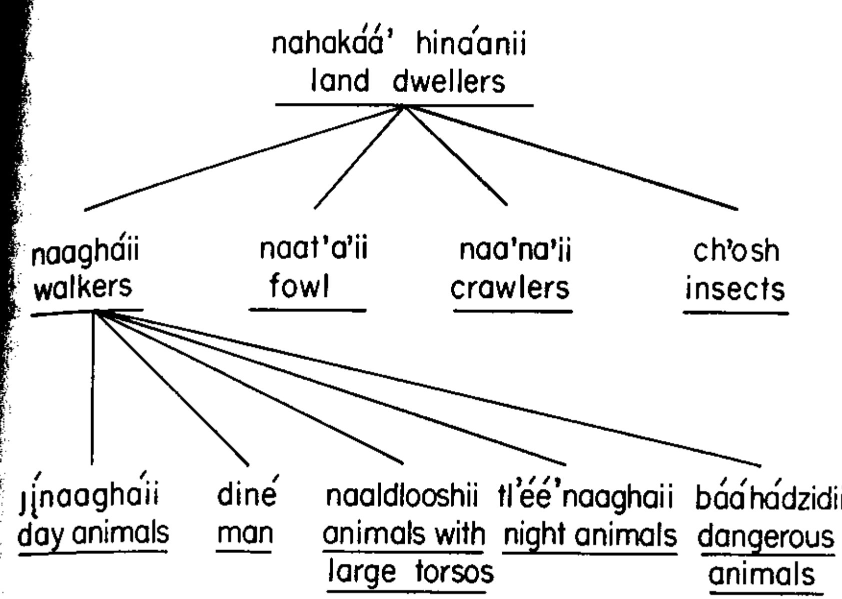

Figure 11.1. Part of the Navaho animal kingdom, derived by Perchonock and Werner (1969) from a pile sort.

SOURCE: "Navaho Systems of Classification" by N. Perchonock and O. Werner, 1969, *Ethnology, 8,* 229-242. Reprinted by permission.

Pile Sorts and Taxonomic Trees

Pile sorting is a an efficient method for generating taxonomic trees (Werner & Fenton, 1973). Simply hand informants the familiar pack of cards, each of which contains some term in a cultural domain. Informants sort the cards into piles, according to whatever criterion makes sense to them. After the first sorting, informants are handed each pile and asked to go through the exercise again. They keep doing this until they say that they can not subdivide piles any further. At each sorting level, informants are asked if there is a word or phrase that describes each pile.

Perchonock and Werner (1969) used this technique in their study of Navaho animal categories. After an informant finished doing a pile sort of animal terms, Perchonock and Werner built a branching tree diagram (such as that shown in Figure 11.1) from the data. They would ask the informant to make up sentences or phrases that expressed some relationship between the nodes. They found that informants intuitively grasped the idea of tree representations for taxonomies.

Pile Sorts and Networks

I've used pile sorts to study the social structure of institutions such as prisons, ships at sea, and bureaucracies, and also to map the cognitively defined social organization of small communities. I simply hand people a deck of cards, each of which contains the name of one of the people in the institution, and ask informants to sort the cards into piles, according to their own criteria. The results tell me how various components of an organization (managers, production workers, advertising people; or guards, counselors, prisoners; or seamen, deck officers, engine room personnel; or men and women in a small Greek village) think about the social structure of the group. Instead of "what goes with what," I learn "who goes with who."

Informants usually find pile sorting fun to do. Asking informants to explain *why* people appear in the same pile produces a wealth of information about the cognitively defined social structure of a group.

Rankings

Rank ordering produces interval-level data, though not all behaviors or concepts are easy to rank. Hammel (1962) asked people in a Peruvian village to rank order the people they knew in terms of prestige. By comparing the lists from different informants, Hammel was able to determine that the men he tested all had a similar view of the social hierarchy. Occupations can easily be rank ordered on the basis of prestige or lucrativeness.

Or even accessibility. The instructions to informants would be "Here is a list of occupations. Please rank them in order, from most likely to least likely that your son will have this occupation." Then ask informants to do the same thing for their daughters. (Be sure to assign informants randomly to doing the task for sons or daughters first.) Then compare the average ranking of accessibility against some independent variables and test for intracultural differences among ethnic groups, genders, age groups, and income groups.

Weller and Dungy (1986) studied breast-feeding among Hispanic and Anglo women in southern California. They asked 55 informants for a free list of positive and negative aspects of breast-feeding and bottle-feeding. Then they selected the 20 most frequently mentioned items in this domain and converted the items to neutral, similarly worded statements. A few examples: "a way that doesn't tie you down, so you are free to do more things"; "a way that your baby feels full and satisfied"; "a way that allows you to feel closer to your baby."

Next, Weller and Dungy asked 195 women to rank the 20 statements. The women were asked which statement was most important to them in selecting a method of feeding their baby; which was the next most important to them; and so on. In the analysis, Weller and Dungy were able to relate the average rank order for Hispanics and for Anglos to independent variables like age and education.

Paired Comparisons

The method of paired comparisons is an alternative way to get rank orderings of a list of items in a domain. If you have a list of 14 items, there are $N(N-1)/2$, or $14(13)/2 = 91$ pairs of items. You write out a list of all the pairs (A and B, A and C . . . B and C . . . F and J . . ., etc.). Show informants each pair and ask them to circle the item that conforms to some criterion. You might say: "Here are two animals. Which one is the more ______?" where the blank is filled in by "vicious," or "wild," or whatever. You could ask informants to choose the "illness in this pair that is more life threatening" or "the food in this pair that is better for you."

In a list of 14 items, each item appears 13 times (A and B, A and C, A and D, etc., down through A and N). To find the rank order of the list for each informant, you simply count up how many times each item "wins"—that is, how many times each item was circled. If cancer is on a list of illnesses and the criterion is "life threatening" you expect to find it circled each time it is paired with another illness—except, perhaps, when it is paired with AIDS. By contrast, the rank ordering of diabetes and high blood pressure on the life-threatening criterion is not so predictable, especially across cultures and ethnic groups.

The paired comparison technique has a lot going for it. Informants make one judgment at a time, so it's much easier on them than asking them to do a rank ordering of a list of items. Also, you can use paired comparisons with nonliterate informants by reading the list of pairs to them, one at a time and recording their answers. Weller and Dungy (1986) did that in their study of breast-feeding.

Like triad tests, paired comparisons can only be used with a relatively limited number of items in a domain. With 20 items, for example, informants have to make 190 judgments. Fortunately, there are BIB designs for paired comparisons, just as there are for triad tests (see above), and ANTHROPAC has a routine for producing individual paired comparison tests.

Ratings

Rating scales produce ordinal data and are easy to administer. Combined with pile sorts and unstructured interviews, ratings are powerful data generators. In a series of papers, John Roberts and his co-workers used pile sorts and rating tasks to study how people perceive various kinds of behaviors in games (see, e.g., Roberts & Chick, 1979; Roberts & Nattrass, 1980).

One "game," studied by Roberts et al. (1981) is pretty serious: searching for foreign submarines in a P3 airplane. The P3 is a four-engine, turboprop, low-wing aircraft that can stay in the air for long periods of time and cover large patches of ocean. It is also used for search-and-rescue missions. Making errors in flying the P3 can result in death or injury at worst, and career damage and embarrassment at least.

Roberts et al. (ibid.) isolated 60 named pilot errors, through extensive unstructured interviews with Navy pilots of the P3. Here are a few of the errors: flying into a known thunderstorm area; taking off with the trim tabs set improperly; allowing the prop wash to cause damage to other aircraft; inducing an autofeather by rapid movement of power level controls. (This is the equivalent of extracting a "free list" from your interviews.) The researchers asked 52 pilots to do an unrestricted pile sort of the 60 errors and to rate each error on a 7-point scale of "seriousness."

They also asked their informants to rank a subset of 13 errors on four criteria that were chosen on the basis of unstructured interviews: (a) how much each error would "rattle" a pilot; (b) how badly each error would damage a pilot's career; (c) how embarrassing each error would be to commit; and (d) how much "fun" it would be to commit each error.

Flying into a thunderstorm on purpose, for example, could be very damaging to a pilot's career, and extremely embarrassing if he had to abort the mission and turn back in the middle. But if the mission turned out to be successful, then taking the risk of committing a very dangerous error would be a lot of fun for pilots who are "high self-testers" (Roberts, personal communication).

Inexperienced pilots rated "inducing an autofeather" as more serious than did highly experienced pilots. Inducing an autofeather is more embarrassing than it is dangerous, and is the sort of error that experienced pilots just don't make. On the other hand, as the number of air hours increased, so did pilots' view of the seriousness of "failure to use all available navigational aids to determine position." Roberts et al. (1981) suggested that inexperienced pilots might not have had enough training to assess the seriousness of this error correctly.

Sydel Silverman (1966) used a pile sort as a rating device to study prestige rankings in an Italian village. During the first few months of her fieldwork, she noticed that people showed deference to one another. She learned that the term *rispetto* (respect) was the important quantity to measure, and that some people had more rispetto than others. The more someone had, the more deference they could expect from others who had less. Everyone in the village was expected to know, more or less, how much rispetto each person had, so that proper interpersonal relations could be maintained.

Ehrerbietung

Occupation didn't predict people's deference to one another, so Silverman worked intensively with three key informants, adult men between 43 and 65 years of age. All three were lifelong residents of the village and had expert knowledge of all families in the village. Silverman gave each key informant a deck of 175 cards containing the names of the families in the village and asked them to sort the cards into piles, according to how much rispetto each family had.

The three informants produced seven, six, and four piles respectively. Silverman asked them to look at a number of paired comparisons between cards in one pile and cards in another. Informants did this exercise until they were satisfied that they had produced a set of internally consistent piles—that each family in a pile belonged in that pile with other families that had the same amount of rispetto.

When the pile-sorting task was done, each informant had created a 3-point, or 6-point, or 7-point ordinal scale, depending on how many piles he wound up creating. Silverman was then able to ask her informants about the sizes of the gaps between the piles. In other words, she tried to understand the intervals between the ordinal ranks. Silverman did not do a statistical analysis of these data. Instead, she used the results of her rating exercise to create a working hypothesis concerning the relative prestige of persons in the village—a model that she could (and did) check against behavioral observations and reports of behavior from informants.

As I said at the beginning of this chapter, I consider the techniques reviewed here to be among the most fun and most productive in the repertoire of anthropological method. They can be used in both applied and basic research; they are attractive to informants; and they produce a wealth of information that can be compared across informants and across cultures. Thirty years in development, the field of cognitive anthropology is becoming increasingly important in anthropology. For further reading, consult Werner and Schoepfle (1987) and Weller and Romney (1988); for handling the actual chores of data collection and analysis, consult the ANTHROPAC manual (Borgatti, 1992a).

19

Bivariate Statistics: Testing Relationships

This chapter is about finding and describing relationships between variables—*covariations*—and testing the significance of those relationships.

We hear the concept of covariation used all the time in ordinary conversation, as when someone asserts that "If kids weren't exposed to so much TV violence, there would be less crime." Ethnographers also use the concept of covariation in statements like: "Most women said they really wanted fewer pregnancies, but claimed that this wasn't possible so long as the men required them to produce at least two fully grown sons to work the land." Here, the number of pregnancies is said to covary with the number of sons husbands say they need for agricultural labor.

The concept of statistical covariation is more precise than that used in ordinary conversation or in ethnographic writing. There are four things we want to know about a statistical relationship between two variables:

1. How big is it? In other words, how much better could we predict the score of a dependent variable in our sample if we knew the score of some independent variable?
2. Is the covariation due to chance, or does it exist in the overall population to which we want to generalize (is it significant)?
3. What is its direction? Is it positive or negative?
4. What is its shape? Is it linear or nonlinear?

Testing for significance is a mechanical affair—you look up in a table whether a statistic showing covariation between two variables is or is not significant. I'll discuss how to do this for several of the commonly used statistics that I introduce below. Interpreting the substantive importance of statistical significance, though, is anything but mechanical. Establishing the theoretical significance of covariations requires thinking, and that's *your* job.

Direction and Shape of Covariations

The concept of *direction* refers to whether a covariation is positive or negative. For example, the amount of cholesterol you have in your blood and the probability that you will die of a heart attack at any given age are positive covariants: the more cholesterol, the higher the probability. By contrast, if you are a native speaker of an Indian language in Mexico, and if you speak Spanish with a strong Indian accent, then the chances are better that you are poor than if you didn't have a strong accent. The higher your score on accent, the lower your wealth.

The various shapes of bivariate relationships are shown in Figure 19.1. Suppose that Figure 19.1(a) were a plot of the number of yams produced by men on a certain Melanesian island, and their height, in centimeters. As you can see, the dots are scattered haphazardly, and there is *no relationship* between the two variables. In Figure 19.1(b), comparing the number of yams produced and the number of wives supported, the relationship is *linear* and positive. The more yams the men produce, the more wives they support.

The third scattergram, Figure 19.1(c), lays out the amount of debt that men have with the amount that others have toward them. The relationship is linear and negative (the more they owe, the less others owe them).

Finally, in Figure 19.1(d), there is clearly a strong relationship (the data are not scattered around randomly), but it is just as clear that the relationship is *nonlinear*—that is, it's not in a single direction. The relationship

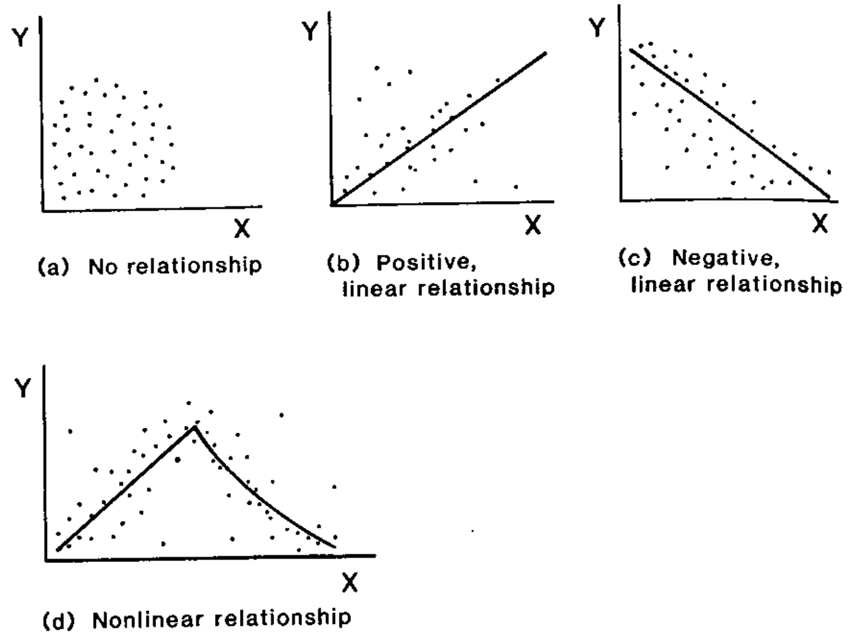

Figure 19.1. Four scattergrams showing the common shapes of bivariate relationships.

between age and the number of people one knows is nonlinear. Early in life the number of friends, kin, and acquaintances is small, but that number grows as you get older. This relationship is linear and positive. The longer you live, the more people you get to know.

Up to a point. If you live long enough, a lot of the people you know start dying, and your network shrinks. There is a strong, negative relationship between age and number of people in your network after age 70. It requires a special kind of statistic, called eta, to test for nonlinear covariation, and I'll discuss it at the end of this chapter when we deal with regression.

Tests for Bivariate Relationships and the PRE Principle

Table 19.1(a) is a hypothetical 2 × 2 (read: two-by-two) table showing the breakdown, by gender, of monolingual Indians and bilingual Indian/Spanish speakers in a Mexican village in 1962. A 2 × 2 table is also called a four-fold table. Any table comparing data on two variables is called a

TABLE 19.1a
Bivariate Table Showing Monolingual and Bilingual Speakers by Gender in a Mexican Village, 1962 (counts)

	Males	Females	
Bilingual	61 (82)	24 (36)	85
Monolingual	13 (18)	42 (64)	55
	74	66	140

old error = 55
new error = 13 + 24 = 37

$$\lambda = \frac{\text{old error} - \text{new error}}{\text{old error}} = .33$$

bivariate table, but not all bivariate tables are 2 × 2, since variables can take more than just two values.

TABLE 19.1b
Bivariate Table Showing Monolingual and Bilingual Speakers by Gender in a Mexican Village, 1962 (percentages)

	Males	Females
Bilingual	82%	36%
Monolingual	18%	64%
	100%	100%
	N = 74	N = 66

λ = .33

NOTE: This table is set up according to the conventions generally followed in reporting data. Dependent variables on the rows, independent down the columns. Column percentages total to 100%, with N given for each column. Only column marginals are given. A statistic describing the table is provided.

In Table 19.1(a), the numbers in parentheses are the column percentages for each cell. Thus, 61 of 74 males, or 82%, are bilingual. Many researchers display only the column percentages in a bivariate table, along with the column totals, or N's, and a summary statistic that describes the table.

This convention is shown in Table 19.1(b). Tables are less cluttered this way, and you get a better understanding of what's going on from percentages than from raw numbers in a bivariate table. As long as the column N's are given, the interested reader can easily reconstruct the N's for each cell. Wherever appropriate, I will provide both the raw cell N's for the tables in this chapter, along with the column percentages for each cell in parentheses, so you can see how tables are constructed.

Numbers along the right side and below a table are called the *marginals*. The marginal in the lower right-hand corner of Table 19.1(a) is the total frequency of elements in the table. The sum of the marginals down the right-hand side and the sum of the marginals across the bottom are identical.

Note that the column percentages sum to 100% and that since we have percentaged the table down the columns it makes no sense to total the percentages in the right margin. In constructing bivariate tables, no matter what size (2×2, as in this case, or larger tables), the common convention is to put the dependent variable in the rows and the independent variable in the columns. Then we percentage down the columns and interpret the table across the rows.

Of course, you can switch the dependent and independent variables around if you like (this is usually done when the independent variable has too many categories to fit conveniently on a narrow page), but be sure to reverse the percentaging also and remember to be consistent. I will follow the convention followed by many scientific journals: *percentage down, read and compare across*.

Reading across Table 19.1(a), we see that 82% of the males were bilingual speakers, compared to 36% of the females. Clearly, gender is related to whether someone is a bilingual Indian/Spanish speaker, or whether they are monolingual in the Indian language only.

Suppose that for the 140 persons in Table 19.1(a) you were asked to guess whether they were bilingual or monolingual, but you didn't know their gender. What would you do? Since the mode for the dependent variable in this table is "bilingual" (85 bilinguals compared to 55 monolinguals), you should guess that everybody is bilingual. If you did that, you'd make 55 mistakes out of the 140 choices, for an error rate of 55/140, or 39%. We'll call this the *old error*.

Since both variables in Table 19.1(a) are nominal, the best measure of central tendency is their modes (recall from Chapter 18 that there is no way to calculate the average sex). Suppose you possessed the data in Table 19.1(a) and knew the mode for each independent variable (each column). The mode for males is bilingual, and the mode for females is monolingual. Knowing this, your best guess would be that every male is bilingual and every female is monolingual. You would still make some mistakes, but fewer than if you just guessed that everyone is bilingual.

Table 19.1(a) shows the aggregate result of this natural experiment on the recruitment of bilingual and monolingual people by gender. It tells you nothing about individual cases, but it contains all the data you need to find out how many fewer mistakes you'd make if you knew the modes of the independent variables.

When you guess that every male is bilingual, you make exactly 13 mistakes, and when you guess that every female is monolingual, you make 24 mistakes, for a total of 37 out of 140 or 37/140 = 26%. This is the *new error*. The difference between the old error (39%) and the new error (26%), divided by the old error is the *proportionate reduction of error*, or PRE. (The PRE principle is well described by Freeman, 1965, and by Mueller et al., 1970.) Thus,

$$\text{PRE} = \frac{55 - 37}{55} = .33 \quad \text{reduction in error, or}$$

$$\text{PRE} = \frac{39\% - 26\%}{39\%} = 33\% \quad \text{reduction in error}$$

This PRE measure of association for nominal variables is called lambda, written either L or λ. Like all PRE measures of association, lambda has the nice quality of being intuitively and directly interpretable. A lambda of .33 means that if you know the scores on an independent variable, you can guess the scores on the dependent variable 33% more of the time than if you didn't know anything about the independent variable.

The PRE principle is very powerful and is the basis for a large group of the most commonly used measures of association. PRE measures are all determined by calculating

$$\text{PRE} = \frac{\text{Old Error} - \text{New Error}}{\text{Old Error}}$$

Lambda can be used for tables larger than 2×2 and for analyzing relationships between nominal and ordinal variables. This is shown in Table 19.2.

TABLE 19.2
Calculating Lambda on a 3 × 3 Table for a Nominal and an Ordinal Variable

	Subsistence Type			
	Hunters	Pastoralists	Agriculturalists	Total
Warfare				
often	(10%) 2	(40%) 8	(55%) 11	21
sometimes	(20%) 4	(45%) 9	(30%) 6	19
never or rarely	(70%) 14	(15%) 3	(15%) 3	20
	N = 20	N = 20	N = 20	

old error = 19 + 20 = 39

new error = (4 + 2) + (8 + 3) + (6 + 3) = 26

$$\text{PRE} = \frac{\text{old error} - \text{new error}}{\text{old error}} = \frac{39 - 26}{39} = .33$$

In this hypothetical example, 60 societies were selected from the Human Relations Area Files—20 hunting and gathering societies, 20 pastoral societies, and 20 irrigation agriculture societies. (I'm treating subsistence technology as a nominal variable here, although from a social evolutionary perspective hunters, pastoralists, and agriculturalists could form an ordinal scale.) Each society was graded on an ordinal scale as to how often it engaged in warfare with its neighbors. "Often" is once a year, or more; "sometimes" is at least once every 7 years, but less than once a year; and "never or rarely" is less often than once every 7 years.

If you didn't know the subsistence technology for each society, your best guess would be that all 60 societies engaged often in warfare—in which case you'd be correct on 21 guesses (35%) and wrong on 39 out of 60 guesses (65%). However, if you *knew* the subsistence technology, your best guess would be that hunters *rarely* engage in warfare (you'd be wrong 4 + 2 = 6 out of 20 times); that all pastoralist societies *sometimes* engage in warfare (you'd be wrong 8 + 3 = 11 out of 20 times); and that agriculturalists are *often* involved in war (you'd be wrong 6 + 3 = 9 out of 20 times).

Guessing this way, you'd make a total of 26 mistakes (a 43.3% error rate) instead of 39 (a 65% error rate). Making 21.7% fewer errors (65% − 43.3% = 21.7%) is a 33% improvement (21.7/65), and this is just what lambda shows in Table 19.2.

TABLE 19.3
Same as Table 19.1, but for 1987

	Males	Females	Total
Bilingual	63 (93)	46 (76)	109
Monolingual	5 (7)	14 (24)	19
	N = 68	N = 60	128

old error = 19

new error = 19

$$\lambda = \frac{19 - 19}{19} = \frac{0}{19} = 0$$

The Problems With Lambda

While lambda demonstrates the intuitively compelling PRE principle, there are three problems with lambda. First of all, there is no way to test whether any value of lambda shows a particularly strong or weak relationship between variables. Second, it is very awkward (even dangerous) to have a statistic that can take different values depending on whether you set up the dependent variable in the rows or the columns.

Third, lambda can be zero (indicating no relationship between the variables), even when there is clear and strong covariation between variables. This is especially likely in certain 2 × 2 tables, where more than 50% of the observations on the independent variable are contained in the cells for the same category of the dependent variable. Look at Table 19.3 to understand this.

These are the hypothetical follow-up data, 30 years later, from the Mexican Indian village study of bilingualism shown in Table 19.1. A new sample of 128 persons has been observed, consisting of 68 males and 60 females. A lot has changed in 30 years. There are hardly any monolingual males left (just 7% in the sample), and there has been a significant reduction in the proportion of monolingual females since a policy of universal, mandatory schooling for both boys and girls was implemented in the mid-1960s. Still, the relationship between gender and bilingualism continues to be obvious: There are many more monolingual females than males.

Despite this clear relationship in the variables, lambda is now zero. The mode for the dependent variable is still "bilingual," so you'd make 19 errors if you guessed that everyone in the sample was bilingual. But the mode for both columns of the independent variable, gender, is on the same row of the dependent variable. More than 50% of both the males and females are bilingual in Table 19.3. Having that table in front of you, then, you'd guess "bilingual" for males, making 5 errors, and "bilingual" for females, making 14 errors, for a total of 19 errors—and lambda would be zero.

Chi-Square

There are two ways to deal with this problem. One way is to play that trick of reversing the dependent and independent variables. The other way is to use a *non-PRE* measure of association for testing covariation between two nominal variables. The most popular of these measures is chi-square, written χ^2. It is very easy to compute, and there are standardized tables for determining whether a particular χ^2 value is significant.

Chi-square will tell you whether or not a relationship exists between or among variables. It will tell you what the probability is that a relationship is the result of chance. But it is *not* a PRE measure and won't tell you the *strength* of association among variables. It is very important to keep this in mind when interpreting this statistic. (You can use lambda as a way to get a better feel for what a particular χ^2 value means, and vice versa.)

The principal use of χ^2 is for testing the *null hypothesis*, i.e., that there is no relationship between two nominal variables. Say you suspect that there is a relationship between two variables in your data—variables like gender and bilingualism in Table 19.1. Using the null-hypothesis strategy, rather than trying to show the relationship, you would try as hard as you can to prove that you are dead wrong—that, in fact, no such relationship exists at all.

If, after a really good faith effort, you *fail to accept* the null hypothesis, you can reject it. Using this approach, you never prove anything; you just fail to disprove it.

Type I and Type II Errors

Chi-square is a particularly good statistic for this conservative, null-hypothesis approach to data analysis. It helps you avoid making either Type I or Type II Errors—that is, either inferring that a relationship exists when it really doesn't, or inferring that a relationship doesn't exist when it really does. Both types of error are serious, but most researchers are more fearful of making a Type I error than a Type II error.

A Type II error is the result of caution and a conservative approach to data analysis—an approach that I fully endorse. When it comes to scientific data analysis, calling someone a "conservative" is to pay them a pretty strong compliment.

Type I errors are the result of what I call "buccaneer data analysis," of being too eager to find relationships, and of engaging in wishful thinking. On the other hand, all of science is based on making mistakes and learning from them. Expect to make a lot of mistakes; try to make more Type II mistakes than Type I, but be ready to engage in a little swashbuckling when you think you're on to something really important.

Calculating χ^2

The formula for χ^2 is

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where O represents the observed number of cases in a particular cell of a bivariate table, and E represents the number of cases you'd expect for that cell *if there were no relationship* between the variables in that cell.

For each cell in a bivariate table, simply subtract the expected frequency from the observed and square the difference. Then divide by the expected frequency and sum the calculations for all the cells. Clearly, if all the observed frequencies equal all the expected frequencies, then chi-square will be zero; that is, there will be no relationship between the variables. While chi-square can be zero, it can never have a negative value. The more the O's differ from the E's (i.e., something nonrandom is going on), the bigger chi-square gets.

Finding the expected frequency for each cell is quite simple. As a first example, let's take a univariate distribution: the amount of land that people own. Suppose there are 14 families in a village and they own a total of 28 hectares of land. If the land were distributed equally among the 14 families, we'd expect each to own 2 hectares. Table 19.4 shows what we would expect, *ceteris paribus* (all other things being equal), compared to what we might find in an actual set of data. The χ^2 value for this distribution is 40.56.

TABLE 19.4
Chi-Square for a Univariate Distribution

Expected Land Holding, in Hectares per Family

Family #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
	2	2	2	2	2	2	2	2	2	2	2	2	2	2	= 28

Observed Land Holding, in Hectares per Family

Family #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
	.2	.4	6.6	1.2	2.1	5.1	.5	.4	.2	.4	.3	3.2	7.1	.3	= 28

(Observed-Expected)²

3.24	2.56	21.16	.64	.01	9.61	2.25	2.56	3.24	2.56	2.89	1.44	26.01	2.89

$\frac{(\text{Observed-Expected})^2}{E}$

1.62	1.28	10.58	.32	.005	4.81	1.13	1.28	1.62	1.28	1.45	.72	13.01	1.45

$$\Sigma \frac{(O-E)^2}{E} = 1.62 + 1.28 + 10.58 + \ldots . \; 1.45 = 40.56$$

Finding the Significance of χ^2

To determine whether this value of χ^2 is significant, first calculate the *degrees of freedom* (abbreviated df) for the problem. For a univariate distribution

$$\text{df} = \text{number of cells} - 1$$

or $14 - 1 = 13$ in this case. For a 2×2 table, there is just one degree of freedom because you know the marginals and once you fill in one of the cells, all the other cell values are determined. The degrees of freedom for any size χ^2 table are calculated by

$$\text{df} = (r - 1)(c - 1)$$

That is, subtract 1 from the number of rows and columns and multiply the two numbers.

Next, go to Appendix E, which is the distribution for χ^2, and read down the left-hand margin to 13 degrees of freedom and across to find the *critical value* of chi-square for any given level of significance. The levels of significance are listed across the top of the table.

By custom (and only by custom) social researchers generally accept as *significant* any relationship that is not likely to occur by chance more than five times in 100 samples. This *p*-value (probability value) is called the .05 level of significance. A *p*-value of .01 level is usually considered *very significant*, and .001 is often labeled *highly significant*.

More and more social researchers are using asterisks instead of *p*-values in their writing. A single asterisk signifies a *p*-value of .05, a double asterisk signifies a value of .01 or less. If you read: "Men were more likely than women** to report dissatisfaction with local schoolteacher training," you'll know that the double asterisk means that the difference between men and women on this variable was significant at the .01 level or better.

The greater the significance of a chi-square value, the less likely it is that you'll make a Type I error. But remember: These customary levels of significance are simply artifacts of our culture. Whether to risk inferring the existence of a relationship that doesn't exist in the population is always a judgment call, for which *you*, not the χ^2 table, take responsibility.

In exploratory field research, you might be satisfied with a .10 level of significance. In evaluating the side effects of a medical treatment you might demand a .001 level. Considering the χ^2 value for the problem in

Table 19.4, I'd say we're on pretty safe ground. A χ^2 value of 34.528 is significant at the .001 level, with 13 degrees of freedom; with a χ^2 of 40.56 we can comfortably assert that inequality of land ownership in the village is significant.

If you are in the field, away from a table of χ^2 values, such as Appendix E, you can estimate the critical value of χ^2 at the 5% level of significance with the formula

$\chi^2 \approx$ (more or less equals)

$1.55(\mathrm{df} + 2)$, for df $\leq$ (equal to or less than) 10

$1.25(\mathrm{df} + 5)$, for $10 < \mathrm{df} \leq 35$

(for df greater than 10, but equal to or less than 35)

For a 3×3 table there are $2 \times 2 = 4$ degrees of freedom. As you can see from Appendix E, the critical value of χ^2 at the 5% level of significance for 4 df is actually 9.488. The rough field formula (Goodman, 1960) produces a critical value of $1.55(6) = 9.30$. It is very unusual to encounter a chi-square problem in anthropology with more than 35 degrees of freedom.

Calculating the Expected Frequencies for χ^2

The test for χ^2 can be applied to any size bivariate table. The expected frequencies are calculated *for each cell* with the formula

$$F_e = \frac{(R_t)(C_t)}{N}$$

where F_e is the expected frequency for a particular cell in a table; (R_t) is the frequency total for the row in which that cell is located; (C_t) is the frequency total for the column in which that cell is located; and N is the total sample size (the lower right-hand marginal). (It is inappropriate to use χ^2 if F_e is less than 5 for any cell.)

Chi-square can be used on bivariate tables comparing observations on nominal variables, or on observations comparing nominal and ordinal

TABLE 19.5

Observed and Expected Frequencies for Chi-Square

Observed Frequencies

Tribe	Religion 1	2	3	Totals
1	150	104	86	340
2	175	268	316	759
3	197	118	206	521
4	68	214	109	391
	590	704	717	2,011

Expected Frequencies

Religion	Tribe 1	2	3
1	99.75	119.03	121.22
2	222.68	265.71	270.61
3	152.85	182.39	185.76
4	114.71	136.88	139.41

$$\chi^2 = \Sigma \frac{(O - E)^2}{E} = \frac{(150 - 99.75)^2}{99.75} + \frac{(104 - 119.03)^2}{119.03} \ldots + \frac{(109 - 139.41)^2}{139.41} = 166.26$$

variables. Table 19.5 shows the observed adherents, in four Native American tribes, of three competing religions. Reading across the top of the table, in tribe #1, there are 150 Mormons, 104 Evangelical Protestants, and 86 members of the Native American Church. The lower half of Table 19.5 shows the expected frequency of each religion in each tribe. Chi-square is computed across the bottom of the table.

Unlike lambda, no matter how you set up a chi-square table (no matter which variable you make the independent one), the value of χ^2 will always be the same. In this case it's a walloping 166.26, with (4 − 1 rows)(3 − 1 columns) = 6 degrees of freedom. In the field, you can trust χ^2 to tell you that something is going on, and you can trust Appendix E (or the rough-and-ready field formula above) to tell you whether a particular distribution of observations is likely to have occurred by chance. Once you have a significant χ^2, a PRE measure like lambda can tell you *how much* the variables are associated.

TABLE 19.6
Dobkin de Rios's (1981) Data on Experience With Witchcraft by Gender

	Personal Experience With Witchcraft, or Close Family With Personal Contact	No Personal Experience With Witchcraft	
Male clients	12	15	27
Female clients	63	5	68
Totals	75	20	95

SOURCE: Reprinted from *Social Science and Medicine, 15B,* M. Dobkin de Rios, "Personal experience of witchcraft and sex of adult interviewed," p. 61, © 1981, with kind permission from Pergamon Press Ltd, Headington Hill Hall, Oxford OX3 OBW, UK.

Cramer's *V*

Cramer's *V* is based on chi-square and is a direct test of the association of two variables. Once you know the χ^2 for a table, you can calculate Cramer's *V* easily. The formula is:

$$V = \sqrt{\frac{\chi^2}{N(k-1)}}$$

where k is the number of rows or columns in your table, whichever is less. If you have three rows and four columns, then $k = 3$ and $k - 1 = 2$. Multiply the number of cases in your sample by $k - 1$. Next, divide χ^2 by that number. Finally, take the square root of *that* number. It's that simple. Cramer's *V* is not a PRE measure, but it will give you an idea of how strong the association is between nominal variables.

Marlene Dobkin de Rios (1981) studied the clientele of a Peruvian folk healer. She suspected that women clients were more likely to have had personal experience with witchcraft (or to have a close family member who has had personal contact with witchcraft) than were men clients. Table 19.6 shows Dobkin de Rios's data.

The χ^2 for this table is 26.89. Consulting Appendix E, we see that, with one degree of freedom, this value of χ^2 is significant at the .001 level. We would not expect this distribution of cases by chance more than once in 1,000 tries.

Cramer's *V* for this χ^2 is the square root of $26.89/(95 \times 1) = 0.28$. Notice that when $k = 2$, the denominator in the formula for Cramer's *V* is simply *N*. Lambda for Dobkin de Rios's table is 0.47. Both Cramer's *V* and lambda show that there is a moderate association in these data between gender and having some personal experience with witchcraft.

a	b	a + b
c	d	c + d
a + c	b + d	TOTAL

Figure 19.2. The cells in a 2 × 2 chi-square table.

The Special Case of the 2 × 2 Table

There is an easy-to-use formula that gives a good approximation of χ^2 for 2 × 2 tables. Since many of the bivariate tables you'll run in the field are of this variety, it pays to get comfortable with this formula:

$$\chi^2 = \frac{N(|ad - bc| - N/2)^2}{(a+b)(c+d)+(a+c)+(b+d)}$$

where *a*, *b*, *c*, and *d* are the individual cells shown in Figure 19.2, and *N* is the total of all the cells (the lower right-hand marginal).

The straight bars inside the parentheses mean that you take the absolute value of the operation $ad - bc$ (that is, you ignore a negative sign, if there is one), and you subtract $N/2$ from *it*. Then you square that number and multiply it by *N* and divide that by the denominator. It takes a little practice to keep track of all the numbers, but this formula is easy to implement in the field with just a simple calculator. Of course, if you have a computer with you in the field, any statistics program will calculate χ^2 for you.

As an example, I've used this formula to compute χ^2 for the data in Tables 19.1 and 19.3. The results are in Table 19.7.

As you'd expect, the χ^2 value for the 1962 data on the relationship between gender and bilingualism is much higher than the 1987 value.

Nevertheless, the 1987 value remains significant. Two-by-two tables have one degree of freedom. Any χ^2 value greater than 3.841 is significant

TABLE 19.7
Raw Frequency Data from Table 19.1 and Table 19.3 Computed for χ^2 Using the Formula for 2 × 2 Tables

	Male	Female	Total
Bilingual	61	24	85
Monolingual	13	42	55
	74	66	140

(a) 1962

	Male	Female	Total
Bilingual	63	46	109
Monolingual	5	14	19
	68	60	128

(b) 1987

$$\chi^2\ (a) = \frac{140\ (|(61)\ (42) - (24)\ (13)| - 70)^2}{(61 + 24)\ (13 + 42)\ (61 + 13)\ (24 + 42)} = 29.14$$

$$\chi^2\ (b) = \frac{128\ (|(63)\ (14) - (46)\ (5)| - 64)^2}{(63 + 46)\ (5 + 14)\ (63 + 5)\ (46 + 14)} = 5.24$$

at the .05 level, and any value over 6.635 is significant at the .01 level. We would expect to get the distribution of data in Table 19.7(b) less than 5 times in 100 tries. Not as good as the χ^2 for Table 19.7(a) (you'd expect *that* distribution less than once in 1,000 tries), but still pretty good.

Fisher's Exact Test

Fisher's exact probability test is used for 2 × 2 tables whenever the *expected* number of frequencies for any cell is less than 5. (With fewer than 5 expected occurrences in a cell, χ^2 values are generally not trustworthy.) Table 19.8 shows data from 20 rural women. After a year of working with these informants, your data show the following: When the 8 women who have equal to or less than 6 years of education need medical help, they rely mostly on traditional healers; when the 12 women who have at least 8 years of education need medical help, they rely on local clinics.

There are thousands of ways to throw 20 cases into four cells, but there are fewer ways to do it if you have to make the right-hand marginals add to 10 each and the bottom marginals add to 8 and 12. Fisher's exact test is based on the fact that, given the marginals for a 2 × 2 table, the number of configurations for achieving that table is fixed. The exact probability

TABLE 19.8
Hypothetical data on labor migration and education for Fisher's Exact Test

	Education		
Medical Help	≤ 6 Years	≥ 8 Years	Total
Healers	7	3	10
Clinics	1	9	10
Total	8	12	20

of getting the *particular* configuration in Table 19.8 is less than 1 in 100 ($p \leq .01$). It is quite safe to say that the women with more education rely on clinics rather than on traditional healers.

Fisher's test does not tell you *why* the distribution turned out the way it did. It could be that education directly influences the choice of medical help; or it could be that education leads to greater wealth and greater wealth leads to reliance on more costly medical services.

Fisher's exact test is handy for small samples (the kind anthropologists often work with), but it is very difficult to calculate by hand. To overcome this problem, Finney (1948) published a long set of tables, listing the exact probability of getting every possible 2 × 2 configuration for samples up to 30 in size. Those tables were extracted and republished by Siegel (1956), who also provided a plain English explanation on how to use them.

The best way to calculate Fisher's test, though, is with a computer program like SPSS, SAS, or SYSTAT. Those top-of-the-line programs automatically calculate Fisher's exact test along with chi-square.

Gamma: The All-Purpose PRE Measure of Association for Ordinal Variables

Once you understand the PRE principle, a lot of things in statistics fall into place. Suppose you test a group of men on how much knowledge they have about the use of wild plants. You rank them as having high, medium, or low knowledge, and you also rank them on a 3-point scale (high, medium, low) for the prestige they hold in their community. We'll assume that you've developed an appropriate index for each of these ordinal variables.

If the two variables are perfectly related, then every man who ranks high on knowledge will also rank high on prestige; every man who ranks low on knowledge will also rank low on prestige; and so on. Of course, things

TABLE 19.9
Plant Knowledge and Prestige Among Male Gardeners in Amazonian Society

Prestige	Plant Knowledge High	Medium	Low	Total
High	18	8	5	31
Medium	9	18	6	33
Low	7	12	12	31
	N = 34	N = 38	N = 23	95

Gamma = .41

never work out so neatly, but if you knew the *proportion of matching pairs* among your informants, you'd have a PRE measure of association for ordinal variables. The measure would tell you how much more correctly you could guess the rank of one ordinal variable for each informant if you knew the score for the other ordinal variable in a bivariate distribution. The raw frequency data for these two variables might look like those in Table 19.9.

What we would like is a PRE measure of association that tells us whether knowing the ranking of pairs of people on one variable increases our ability to predict their ranking on a second variable, and by how much. To do this, we need to understand the ways in which pairs of ranks can be distributed. This will not appear obvious at first, but bear with me.

The number of possible pairs of observations (on any given unit of analysis) is

$$\text{No. of Pairs of Observations} = N(N-1)/2$$

where N is the sample size. There are (95)(94)/2 = 4,465 pairs of observations in Table 19.9.

There are several ways that pairs of observations can be distributed if they are ranked on two ordinal variables.

1. They can be ranked in the same order on *both* variables. We'll call these "same."
2. They can be ranked in the opposite order on both variables. We'll call these "opposite."

3. They can be tied on either the independent or dependent variables, or on both. We'll call these "ties."

In fact, in almost all bivariate tables comparing ordinal variables, there are going to be a lot of pairs with tied values on both variables. Gamma, written G, is a popular measure of association between two ordinal variables because it *ignores* all the tied pairs. The formula for gamma is

$$G = \frac{\text{No. of Same-Ranked Pairs} - \text{No. of Opposite-Ranked Pairs}}{\text{No. of Same-Ranked Pairs} + \text{No. of Opposite-Ranked Pairs}}$$

Gamma is an intuitive statistic; it ranges from −1 (for a perfect negative association) to +1 (for a perfect positive association), through 0 in the middle for complete independence of two variables. At best, the number of opposite-ranked pairs would be 0, in which case gamma would equal 1.

For example, suppose we measured income and education ordinally, such that anyone with less than a high school diploma is counted as having low education, and anyone with at least a high school diploma is counted as having high education. Similarly, anyone with an income of less than $10,000 a year is counted as having low income, while anyone with at least $10,000 a year is counted as having high income.

Now suppose that *no one* with at least a high school diploma earned less than $10,000 dollars a year. There would be no pair of observations, then, in which low income and high education (an opposite pair) co-occurred.

In the worst case for gamma, the number of same-ranked pairs would be 0, in which case gamma would equal −1. For example, suppose that *no one* who had high education also had a high income. This would be a perfect negative association, and gamma would be −1.

The number of same-ranked pairs in a bivariate table is calculated by multiplying each cell by the sum of all cells *below it and to its right.* The number of opposite-ranked pairs is calculated by multiplying each cell by the sum of all cells *below it and to its left.* This is diagrammed in Figure 19.3.

In Table 19.9, the number of same-ranked pairs is

18(18 + 6 + 12 + 12)		= 864
+ 8(6 + 12)		= 144
+ 9(12 + 12)		= 216
+ 18(12)		= 216
	Total	1,440

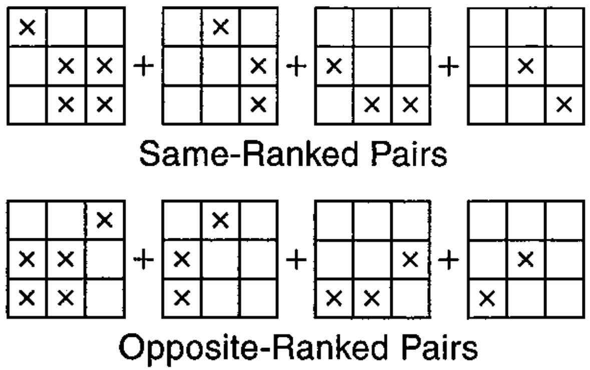

Figure 19.3. Calculating gamma. To calculate the same-ranked pairs in this 3×3 table, multiply each score by the sums of all scores below it and to the right. Then sum the totals. To calculate the opposite-ranked pairs, multiply each score by the sums of the scores below it and to the left. Then sum the totals.

The number of opposite-ranked pairs is

$5(18+9+7+12)$	$= 230$
$+\ 8(9+7)$	$= 128$
$+\ 6(12+7)$	$= 114$
$+\ 18(7)$	$= 126$
Total	598

Gamma for Table 19.9, then, is

$$G = \frac{1{,}440 - 598}{1{,}440 + 598} = \frac{842}{2{,}038} = .41$$

So, plant knowledge and prestige are positively related.

Is Gamma Significant?

If you have more than 30 elements in your sample, you can test for the probability that gamma is due to sampling error using a procedure developed by Goodman and Kruskal (1963). A useful presentation of the procedure is given by Loether and McTavish (1974:552). First, the gamma value must be converted to a z-score, also called a "standard score." (I won't

deal here with how z-scores are derived.) The formula for converting gamma to a z-score is

$$z = (G - \gamma)\sqrt{(N_s + N_o)/N(1 - G^2)}$$

where G is the *sample* gamma, γ is the gamma for the *population*, N is the size of your sample, N_s is the number of same-ranked pairs, and N_o is the number of opposite-ranked pairs.

As usual, we proceed from the null hypothesis, and assume that γ for the entire population is zero—that is, that there really is no association between the variables we are studying. If we can reject that hypothesis, then we can assume that the gamma value for our sample probably approximates the gamma value, γ, for the population. For the gamma value in Table 19.9,

$$z = (.41 - 0)\sqrt{(1{,}440 + 598)/95(1 - .41^2)} = 2.08$$

Appendix F is a z-score table. It lists the proportions of area under a normal curve that are described by various z-score values. To test the significance of gamma, look for the z-score in column 1 of the table. Column 2 shows the area under a normal curve between the mean (assumed to be zero for a normal curve) and the z-score. We're interested in column 3, which shows the area under the curve that is *not* accounted for by the z-score.

A z-score of 2.08 accounts for all but .0188 (1.88%) of the area under a normal curve. To be conservative, we'll round this up to 2%. Now we can reject the null hypothesis at about the 2% level, and we can presume that there is a real relationship between plant knowledge and prestige among the women in the general population from which we took our data. Chi-square for Table 19.8 confirms this finding. The value is $\chi^2 = 13.84$, with 4 degrees of freedom. Appendix E shows that, with 4 degrees of freedom, χ^2 has to exceed 13.277 to be significant at the 1% level.

Kendall's Tau-*b*

Because gamma ignores tied pairs in the data (and there might be a lot of them), some researchers prefer a statistic called Kendall's tau-*b* (written τb) for bivariate tables of ordinal data. The formula for τ_b is

$$\tau_b = \frac{N_s - N_o}{\sqrt{(N_s + N_o + N_{td})(N_s + N_o + N_{ti})}}$$

where N_s is the number of same-ranked pairs, N_o is the number of opposite-ranked pairs, N_{td} is the number of pairs tied on the dependent variable, and N_{ti} is the number of pairs tied on the independent variable. You can calculate the tied pairs as follows: For pairs tied on the dependent variable, multiply the first cell of the *row* of the table by the sum of the cells across the row. In Table 19.9

$$18(8+5)+8(5)$$
$$+\,8(18+12)+18(12)$$
$$+\,5(6+12)+6(12)=892$$

Pairs tied on the independent variable are calculated by multiplying the first cell of the *column* of the table by the sum of the cells down the columns. For Table 19.9

$$18(9+7)+9(7)$$
$$+\,8(18+12)+18(12)$$
$$+\,5(6+12)+6(12)=969$$

For Table 19.9, then,

$$\tau_b=\frac{1{,}440-598}{\sqrt{(1{,}440+598+892)(1{,}440+598+969)}}=.28$$

Kendall's τ_b will nearly always be smaller than gamma, because gamma ignores tied pairs while τ_b uses almost all the data (it ignores the relatively few pairs that are tied on both variables). Gamma is known as an intuitive, friendly statistic, easily interpreted as a PRE measure of association, and easy to evaluate using z-tables. On the other hand, many researchers like τ_b because it is a conservative statistic that doesn't inflate relationships between variables by ignoring data (tied pairs of observations). However, it is very difficult to test the significance of τ_b in the field—the formula is a beast, and there are no convenient tables that you can look up.

Yule's Q

A lot of work done by anthropologists in the field results in 2×2 tables of ordinal variables, like "high" versus "low" prestige, salary, education,

hunting prowess, etc. In these cases, you can use a statistic called Yule's Q. This statistic is a modified, quick form of gamma (but without gamma's precise interpretation), and it can be calculated on frequencies or on percentages. Like the quick formula for χ^2, you can only use Yule's Q on 2×2 tables. The formula for Q is

$$Q=\frac{(ad)-(bc)}{(ad)+(bc)}$$

Yule's Q is a handy, easy-to-use statistic, and that's probably why it's so popular. Unlike a true gamma, however, you can not calculate its significance. A good rule of thumb for interpreting the significance of Q is given by James Davis (1971): When Q is 0, the interpretation is naturally that there is no association between the variables. When Q ranges from 0 to −.29, or from 0 to +.29, you can interpret this as a negligible or small association. Davis interprets a Q value of ±.30 to ±.49 as a "moderate" association; a value of ±.50 to ±.69 as a "substantial" association; and a value of ±.70 or more as a "very strong" association.

What to Use for Ordinal Variables

My advice is this: Since Yule's Q is not easily interpreted, and since the significance of τ_b is very difficult to calculate, you should use these last two statistics only for special purposes. Specifically, Q is a useful statistic for getting a quick feel for the potential relationship of two ordinal variables in a 2×2 table. τ_b is a conservative statistic that lets you check how much stock to place in a marginally significant gamma.

In general, however, I recommend using chi-square and gamma on tables of nominal and ordinal data, respectively. Since 2×2 ordinal tables are usually chock full of tied pairs of ranked observations, try not to make up ordinal variables with only two ranks. This does *not* mean that you should make up artificial ranks just to fill out a variable. On the contrary, it means that you should work as hard as you can to *understand* the ordinal variables you are working with, so that you can make legitimate distinctions among at least three ranks.

For purposes of data analysis, an ordinal variable with seven legitimate ranks can be treated exactly as if it were an interval variable. Many researchers treat ordinals with just five ranks as if they were intervals, because association between interval-level variables can be analyzed by the most powerful statistics—which brings us to correlation and regression.

Correlation: The Powerhouse Statistic for Covariation

Where at least one of the variables in a bivariate relationship is interval or ratio level, we use either *Pearson's product moment correlation*, written simply as *r*, or a statistic called eta, written η, depending on the shape of the relationship. (Go back to the section on "shape of relationship" at the beginning of this chapter if you have any doubts about this concept.)

Pearson's *r*

Pearson's *r* is an intuitive PRE measure of association for linear relationships between lots of different types of variables. It is generally used to test for associations between interval variables, but it can also be used for an interval and an ordinal variable, or even for an interval and a nominal variable. It tells us how much better we could predict the scores of a dependent variable, if we knew the scores of some independent variable.

Consider two interval-level variables, like income (measured in some monetary unit like pesos or drachmas) and education (measured in years). Table 19.10 shows hypothetical data on 10 informants in a small village in Brazil. Now, suppose you had to predict the income level of each person in Table 19.10 *without knowing anything about their education.* Your best guess would be the mean, 45,600 escudos per month. If you have to make a wild guess on the particular scores of any interval-level variable, your prediction error will always be smallest if you pick the mean for each and every informant.

You can see this in Figure 19.4. I have plotted the distribution of income and education for the 10 informants shown in Table 19.10 and have drawn in the line for the mean (the dashed line).

Each dot is physically distant from the mean line by a certain amount. The sum of the squares of these distances to the mean line is the smallest sum possible (that is, the smallest cumulative prediction error you could make), given that you *only* know the mean of the dependent variable. The distances from the dots *above* the line to the mean are positive; the distances from the dots *below* the line to the mean are negative. The sum of the actual distances is zero. Squaring the distances gets rid of the negative numbers.

But suppose you *do* know the data in Table 19.10 regarding the education of your informants. Can you reduce the prediction error in guessing their income? Could you draw another line through Figure 19.4 that "fits" the dots better and reduces the sum of the distances from the dots to the line?

TABLE 19.10
Education and Income for Ten Rural Villagers in Brazil

Person	x Education in Years	y Income in Escudos per Month
1	0	32,000
2	0	42,000
3	3	35,000
4	4	38,000
5	6	43,000
6	6	37,000
7	6	39,000
8	8	54,000
9	12	58,000
10	12	78,000
	$\bar{x} = 5.7$	$\bar{y} = 45{,}600$

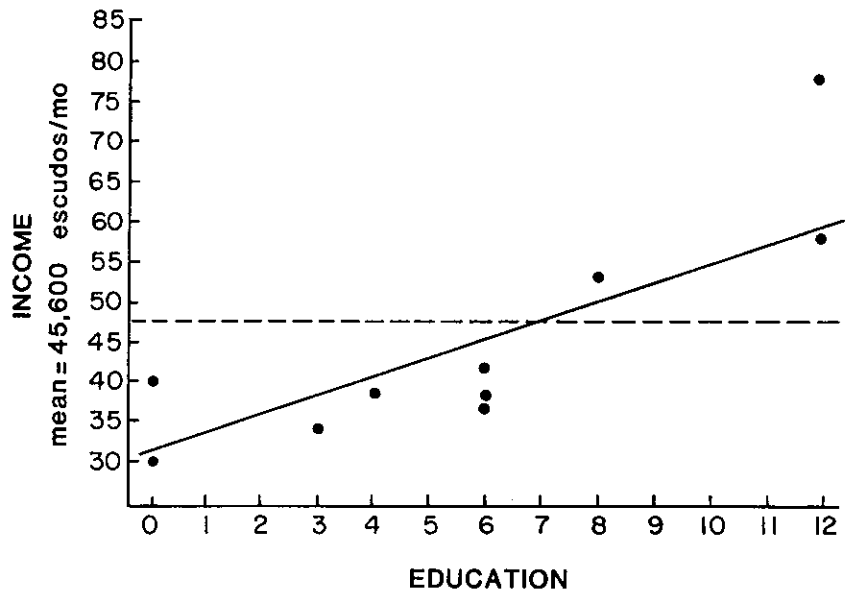

Figure 19.4. A plot of the data in Table 19.10. The dotted line is the mean. The solid line is drawn from the regression equation $Y = 30.10 + 2.72X$.

You bet you can. The solid line that runs diagonally through the graph minimizes the prediction error for these data. This line is called the *best-fitting* line, or the *least squares* line, or the *regression* line. When you understand how this regression line is derived, you'll understand how correlation works.

Regression

The formula for the regression line is

$$y = a + bx$$

where y is the variable value of the dependent variable, a and b are some constants (which you'll learn how to derive in a moment), and x is the variable value of the independent variable. The constant, a, is computed as

$$a = \bar{y} - b\bar{x}$$

and b is computed as

$$b = \frac{N\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{N\left(\sum x^2\right) - \left(\sum x\right)^2}$$

Table 19.11 shows the data needed for finding the regression equation for the raw data in Table 19.10.

To reduce clutter, I have listed income in thousands of escudos/month. The constant b is

$$b = \frac{10(3{,}035) - (57)(456)}{10(485) - (57^2)} = \frac{4{,}358}{1{,}601} = 2.72$$

and the constant a is then

$$a = 45.65 - 2.72(5.7) = 30.10$$

The regression equation for any pair of scores on income (y) and education (x), then, is

TABLE 19.11

Computation of Pearson's r Directly From Data in Table 19.10

Person	x Education	y Income (in thousands)	xy	x^2	y^2
1	0	32	0	0	1024
2	0	42	0	0	1764
3	3	35	105	9	1225
4	4	38	152	16	1444
5	6	43	258	36	1849
6	6	37	222	36	1369
7	6	39	234	36	1521
8	8	54	432	64	2916
9	12	58	696	144	3364
10	12	78	936	144	6084
	$\Sigma x = 57$	$\Sigma y = 456$	$\Sigma xy = 3{,}035$	$\Sigma x^2 = 485$	$\Sigma y^2 = 22{,}560$
	$\bar{x} = 5.7$	$\bar{y} = 45.6$			

$$r = \frac{N\Sigma xy - \Sigma x\,\Sigma y}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2]\,[N\Sigma y^2 - (\Sigma y)^2]}}$$

$$= \frac{10\,(3035) - (57)\,(456)}{\sqrt{[10\,(485) - (57^2)]\,[10\,(22{,}560) - (456)^2]}} = .82$$

$$r^2 = (.82)^2 = .67$$

$$y = a + bx = 30.10 + 2.72x$$

Suppose we want to predict the dependent variable y (income) when the independent variable y (years of education) is 5. In that case,

$$y = 30.10 + 2.72(5) = 37.82$$

or 37,820 escudos a month. As you can see, the regression equation lets us estimate income for education levels that are not even represented in our sample.

The regression equation also lets us draw the solid line through Figure 19.4, such that the squared deviations (the distances from any dot to the line, squared) add up to less than they would for any other line we could draw through that graph. The mean is the least squares point for a *single* variable. The regression line is the least squares line for a plot of *two* variables. That's why the regression line is also called the "best-fitting" line.

Drawing the Regression Line

When you are doing data analysis in the field, I recommend that you actually plot out your data and draw in the regression lines on bivariate plots like Figure 19.4. There is no substitute for the "feel" that you get about covariation from looking at actual plots and regression lines.

To draw these lines, come up the y axis to the point where a (30.10 in Figure 19.4) intercepts it. This is called the "y intercept." Then, for every increment in x, simply apply the formula $y = a + bx$, and connect the dots. Actually, you only need to plot two points for the regression line, connect those points, and extend the line as far as you need to in both directions.

How Regression Works

To give you a clear idea of how the regression formula works, here are all the predictions along the regression line for the data in Table 19.10.

For person	whose education is	predict his or her income is
1	0 years	30.1 + 2.72(0) = 30,100
2	0	30.1 + 2.72(0) = 30,100
3	3	30.1 + 2.72(3) = 38,260
4	4	30.1 + 2.72(4) = 40,980
5	6	30.1 + 2.72(6) = 46,420
6	6	30.1 + 2.72(6) = 46,420
7	6	30.1 + 2.72(6) = 46,420
8	8	30.1 + 2.72(8) = 51,860
9	12	30.1 + 2.72(12) = 62,740
10	12	30.1 + 2.72(12) = 62,740

We now have two predictors of income: (a) the mean income, which is our best guess when we have no data about some independent variable like education; and (b) the values produced by the regression equation when we *do* have information about something like education. Each of these predictors produces a certain amount of error, or *variance*.

You'll recall from Chapter 18 that in the case of the mean, the total variance is the average of the squared deviations of the observations from the mean $(1/N)[\sum(x-\bar{x})^2]$. In the case of the regression line predictors, the variance is the sum of the squared deviations from the regression line. Table 19.12 compares these two sets of errors, or variances, for the data in Table 19.10.

We now have all the information we need for a true PRE measure of association between two interval variables. Recall the formula for a PRE measure: the old error minus the new error, divided by the old error. For our example in Table 19.12:

$$\text{PRE} = \frac{1{,}766.40 - 584.69}{1{,}766.40} = .67$$

In other words: The proportional reduction of error in guessing the income of someone in the sample displayed in Table 19.10, given that you know the distribution of education and can apply a regression equation, compared to just guessing the mean of income, is 67%.

This quantity is usually referred to as "r-squared" (written r^2), or the amount of variance accounted for by the independent variable. The Pearson product moment correlation, written as r, is the square root of this measure, or, in this instance, .82. Most researchers calculate Pearson's r directly from data, using the formula

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{\left[N\sum x^2 - \left(\sum x\right)^2\right]\left[N\sum y^2 - \left(\sum y\right)^2\right]}}$$

Calculating r and r^2

Table 19.11 showed you the calculation of r and r^2 for the data in Table 19.10. As you can see, the procedure is simple and can be handled conveniently in the field without calculating y-intercepts, regression constants, and so on. But I've given you this grand tour of regression and correlation because I want you to see that Pearson's r is not a direct PRE measure of association; its *square* (written r^2) is.

There is a controversy in social statistics over whether Pearson's r or r^2 better describes the relationship between variables. Pearson's r is easy to

TABLE 19.12
Comparison of the Error Produced by Guessing the Mean Income for Each Informant in Table 19.10 and the Error Produced by Applying the Regression Equation for Each Guess

Person	x Education	y Income	Old Error $(y - \bar{y})^2$	Guess Using Regression Equation	New Error $\left[y - \text{Guess Using Regression Equation}\right]^2$
1	0	32	184.96	30.10	3.61
2	0	42	12.96	30.10	141.61
3	3	35	112.36	38.26	10.63
4	4	38	57.36	40.98	8.88
5	6	43	6.76	46.42	11.70
6	6	37	73.96	46.42	88.74
7	6	39	43.56	46.42	59.60
8	8	54	70.56	51.86	4.58
9	12	58	153.76	62.74	22.47
10	12	78	1,049.76	62.74	232.87
		$\bar{y} = 45.6$	$\Sigma(y - \bar{y})^2 =$ 1,766.40		$\Sigma\left[y - \text{guess using regression equation}\right]^2 = 584.69$

compute from raw data and it varies from −1 to +1, so it has direction and an intuitive interpretation of magnitude. It's also almost always bigger than r^2. By contrast, r^2 is a humbling statistic. A correlation of .30 looks impressive until you square it and see that it explains just 9% of the variance in what you're studying.

The good news is that if you double a correlation coefficient, you quadruple the variance accounted for. For example, if you get an *r* of .25, you've accounted for 6.25% of the variance, or error, in predicting the score of a dependent variable from a corresponding score on an independent variable. An *r* of .50 is twice as large as an *r* of .25, but four times as good, because .50*r* means that you've accounted for 25% of the variance.

Testing the Significance of *r*

Just as with gamma, it is possible to test whether or not any value of Pearson's *r* is the result of sampling error, or reflects a real covariation in the larger population. In the case of *r*, the null hypothesis is that, within certain confidence limits, we should predict that the real coefficient of correlation in the population of interest is actually zero. In other words, there is no relation between the two variables.

We must be particularly sensitive in anthropology to the possible lack of significance of sample statistics because we often deal with small samples. The procedure for testing the confidence limits of *r* is a bit complex. To simplify matters, I have constructed Table 19.13, which you can use in the field to get a ball-park reading on the significance of Pearson's *r*. The top half of Table 19.13 shows the 95% confidence limits for representative samples of 30, 50, 100, 400, and 1,000, where the Pearson's *r* values are .1, .2, .3, etc. The bottom half of Table 19.13 shows the 99% confidence limits.

Reading the top half of Table 19.13, we see that at the 95% level, the confidence limits for a correlation of .20 in a sample of 1,000 are .14 and .26. This means that in fewer than 5 tests in 100 would we expect to find the correlation smaller than .14 or larger than .26. In other words, we are 95% confident that the true *r* for the population (ρ, which is the Greek letter rho) is somewhere between .14 and .26.

By contrast, the 95% confidence limits for an *r* of .30 in a representative sample of 30 is not significant at all; the true correlation could be 0, and our sample statistic of .30 could be the result of sampling error.

The 95% confidence limits for an *r* of .40 in a representative sample of 30 is statistically significant. We can be 95% certain that the true correlation

TABLE 19.13

Confidence Limits for Pearson's *r* for Various Sample Sizes

Pearson's r	Sample Size				
	30	50	100	400	1,000
.1	ns	ns	ns	ns	.04–.16
.2	ns	ns	.004–.40	.10–.29	.14–.26
.3	ns	.02–.54	.11–.47	.21–.39	.24–.35
.4	.05–.67	.14–.61	.21–.55	.32–.48	.35–.45
.5	.17–.73	.25–.68	.31–.63	.42–.57	.45–.54
.6	.31–.79	.39–.75	.45–.71	.53–.66	.56–.64
.7	.45–.85	.52–.82	.59–.79	.65–.75	.67–.73
.8	.62–.90	.67–.88	.72–.86	.76–.83	.78–.82
.9	.80–.95	.83–.94	.85–.93	.88–.92	.89–.91
			(95% Confidence Limits)		
.1	ns	ns	ns	ns	.02–.18
.2	ns	ns	ns	.07–.32	.12–.27
.3	ns	ns	.05–.51	.18–.41	.23–.45
.4	ns	.05–.80	.16–.59	.28–.50	.33–.46
.5	.05–.75	.17–.72	.28–.67	.40–.59	.44–.56
.6	.20–.83	.31–.79	.41–.74	.51–.68	.55–.65
.7	.35–.88	.46–.85	.55–.81	.63–.76	.66–.74
.8	.54–.92	.62–.90	.69–.88	.75–.84	.77–.83
.9	.75–.96	.80–.95	.84–.94	.87–.92	.88–.91
			(99% Confidence Limits)		

in the population (ρ) is no less than .05 and no larger than .67. This is a significant finding, but not much to go on insofar as external validity is concerned. You'll notice that with very large samples (like 1,000), even very small correlations are significant at the .01 level. Just because a statistical value is significant doesn't mean that it's important or useful in understanding how the world works.

Looking at the lower half of Table 19.13, we see that even an *r* value of .40 is insignificant when the sample is as small as 30. If you look at the spread in the confidence limits for both halves of Table 19.13, you will notice something very interesting: A sample of 1,000 offers some advantage over a sample of 400 for bivariate tests, but the difference is small and the costs of the larger sample in the field are very high.

Recall from Chapter 4 that in order to halve the confidence interval you have to quadruple the sample size. Where the unit cost of data is high, as in research based on direct observation or personal interviews, the point of diminishing returns on sample size is reached quickly. Where the unit cost of data is low, as in much questionnaire research, a larger sample is worth trying for.

Nonlinear Relationships

All the examples I have used so far have been for linear relationships where the best-fitting "curve" on a bivariate scattergram is a straight line. Whenever long periods of time constitute one of the variables in a pair, however, there is a good chance that the relationship is nonlinear.

Consider political orientation over time. The Abraham Lincoln Brigade was a volunteer, battalion-strength unit of Americans who fought against the rightist forces of Francisco Franco during the Spanish Civil War, 1936-1939. The anti-Franco forces were supported by leftist groups and by the Soviet Union. On the 50th anniversary of the start of the Spanish Civil War, surviving members of the Lincoln Brigade gathered at Lincoln Center in New York City.

Covering the gathering for *The New York Times* (April 7, 1986:B3), R. Shepard noted that "While some veterans might still be inspired by their youthful Marxism, many, if not most, have broken with early orthodoxies" and had become critical of the (then) Soviet Union since their youth. There are many examples of leftist activists in modern society who are born into relatively conservative, middle-class homes, become radicals in their 20s, and become rather conservative after they "settle down" and acquire family and debt obligations. Later in life, when all these obligations are over, they may once again return to left-wing political activity. This back-and-forth swing in political orientation probably looks something like Figure 19.5.

Nonlinear relationships are everywhere, and you need to be on the lookout for them. Munroe et al. (1983) conducted four time allocation studies: two in horticultural peasant communities in Kenya, one in a highland community in Peru, and one on a sample of middle-class urbanites in the United States. They examined the relationship between the amount of time spent in productive labor and the technoeconomic level of the society.

The relationship between these two variables in their data is curvilinear. Labor inputs rise from moderate to very high levels as you go from hunter/gatherers and horticulturists to intensive agriculturalists. But labor inputs fall as you go from agricultural to industrial societies.

As I mentioned in Chapter 3, Lambros Comitas and I studied Greek labor migrants who had returned to Greece after more than 5 years in Germany. We found a nonlinear relationship between socioeconomic class and attitudes

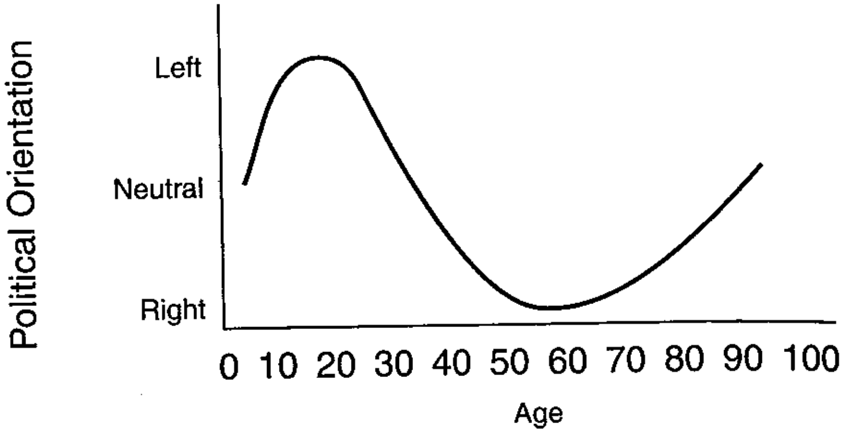

Figure 19.5. A nonlinear relationship. Political orientation through time.

toward dowry. Lower-class and upper-class men and women supported the traditional view that women need dowry in order to get married. Those in the middle class felt that dowry was an outmoded custom, that women should work outside the home, and that the income from that work should substitute for dowry.

If you get a very weak r or r^2 for two variables that you believe, from ethnographic evidence, are strongly related, then draw a scattergram (all statistical analysis packages produce scattergrams) and check out the relation more carefully. Scattergrams are packed with information. For sheer intuitive power, there is nothing like them. If a scattergram looks anything like either of the shapes in Figure 19.1(d) or Figure 19.5, or like any other complex curve, then r is not the right statistic to use because r is based on the concept of *linear* regression. An alternative is eta.

Calculating Eta

Eta, written η, is a very useful statistic. It is a PRE measure that tells you how much better you could do if you predicted the separate means for *chunks* of your data than if you predicted the mean for all your data. Figure 19.6 shows hypothetical data for a sample of 20 informants, ages 12 to 89, on the variable "number of friends and acquaintances." It is based on the data displayed in Table 19.14.

TABLE 19.14
Hypothetical Data on Number of Friends by Age

Person	Age	Number Friends	
1	12	40	
2	18	140	$\bar{y}_1 = 182.50$
3	21	300	
4	26	250	
5	27	560	
6	30	430	
7	36	610	
8	39	410	
9	42	820	$\bar{y}_2 = 570.0$
10	45	550	
11	47	700	
12	49	750	
13	51	410	
14	55	380	
15	61	650	
16	64	520	
17	70	220	
18	76	280	$\bar{y}_3 = 238.0$
19	80	110	
20	89	60	
		$\bar{y} = 409.5$	

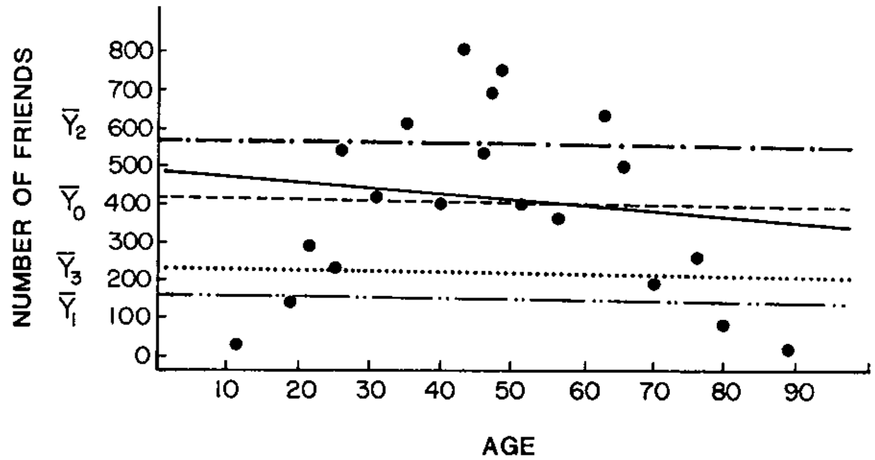

Figure 19.6. Number of friends by age.

In Figure 19.6, the large dots are the data points from Table 19.14. Informant #10, for example, is 45 years of age and was found to have approximately 550 friends and acquaintances. The horizontal dashed line marked $\overline{Y}_0$ is the global average for these data, 409.5. Clearly, (a) the global average is not of much use in predicting the dependent variable; (b) knowing an informant's age *is* helpful in predicting the size of his or her social network; but (c) the linear regression equation, $y = 451.45 - .89x$, is hardly any better than the global mean at reducing error in predicting the dependent variable. You can test this by comparing the mean line and the regression line (the slightly diagonal line running from upper left to lower right in Figure 19.6) and seeing how similar they are.

What that regression line depicts, of course, is the correlation between age and size of network, which is a puny .08. But if we inspect the data visually, we find that there are a couple of natural "breaks." It looks like there's a break in the late 20s, and another somewhere in the 60s. We'll break these data into three age chunks from 12 to 26, 27 to 61, and 64 to 89, take separate means for each chunk, and see what happens. I have marked the three chunks and their separate means on Table 19.14.

Unlike r, which must be squared to find the variance accounted for, eta is a direct measure of this and is calculated from the following formula:

$$\eta = 1 - \frac{\sum (y - \bar{y}_c)^2}{\sum (y - \bar{y})^2}$$

where $\bar{y}_c$ is the average for each chunk and $\bar{y}$ is the overall average for your dependent variable. For Table 19.14, eta is

$$\eta = 1 - \frac{395,355}{1,033,895} = .79$$

which shows a very strong relationship between the two variables, despite the very weak Pearson's r.

Eta varies between zero and one. It is a good statistic to use when you are testing covariation between an interval and a nominal variable—such as age and any yes-no variable like acculturated versus nonacculturated. (According to Freeman, 1965, eta is the *only* statistic to use in that case). It can also be used to compare interval and ordinal variables, and it allows you to test for nonlinear relationships between two interval variables. Eta is an all-around, varsity statistic.

Statistical Significance, the Shotgun Approach, and Other Issues

To finish this chapter, I want to deal with four thorny issues in social science data analysis: (a) measurement and statistical assumptions, (b) eliminating the outliers, (c) significance tests, and (d) the shotgun method of analysis.

Measurement and Statistical Assumptions

By now you are comfortable with the idea of nominal, ordinal, and interval-level measurement. This seminal notion was introduced into social science in a classic article by Stevens in 1946. Stevens said that statistics like t and r, because of certain assumptions that they made, required interval-level data, and this became an almost magical prescription.

Thirty-six years later, Gaito (1980) surveyed the (by then voluminous) mathematical statistics literature and found no support for the idea that measurement properties have anything to do with the selection of statistical procedures. Social scientists, says Gaito, confuse measurement (which focuses on the meaning of numbers) with statistics (which doesn't care about meaning at all) (p. 566). So, treating ordinal variables as if they were interval, for purposes of statistical analysis, is almost always a safe thing to do, especially with five or more ordinal categories (R. P. Boyle, 1970; Labovitz, 1971a).

The important thing is measurement, not statistics. As I pointed out in Chapter 2, many concepts, such as gender, race, and tribe, are much more subtle and complex than we give them credit for being. Instead of measuring them qualitatively (remember that assignment of something to a nominal category is a qualitative act of measurement), we ought to be thinking hard about how to measure them ordinally.

Durkheim was an astute theorist. He noted that division of labor became more complex as the complexity of social organization increased. But he, like other evolutionist theorists of his day, divided the world of social organization into a series of categories (gemeinschaft versus gesellschaft, or mechanical versus organic solidarity, or savagery, barbarism, and civilization).

When anthropologists rejected these simplistic schemes of social evolution, they did not substitute better measurement. Surely, what we really want to know is the *relationship* of the division of labor to social complexity in general. This requires some hard thinking about how to measure these two variables with more subtlety. The meaning of the measurements is crucial.

Eliminating the Outliers

Another controversial practice in data analysis is called "eliminating the outliers," that is, removing extreme values from data analysis. If there are clear indications of measurement error (a person with a score of 600 on a 300-point test turns up in your sample), you can throw out the data that are in error. If you decide to restrict the applicability of your sample, you can get rid of extreme cases—defining your population as "all cities in New York State under 2 million," for instance, eliminates New York City.

The problem is that outliers (so-called freak cases) are sometimes eliminated just to "smooth out" data and achieve better fits of regression lines to data. A single millionaire might be ignored in calculating the average net worth of a group of blue-collar workers on the theory that it's a "freak case." But what if it isn't a freak case? What if it represents a small proportion of cases in the population under study? Eliminating it only prevents the discovery of that fact.

Or suppose you counted the number of separate living quarters among five polygynous households, and found that one man had 11 wives, while the others had 2, 3, 2, and 4 wives, respectively. You might be tempted to eliminate the man with 11 wives from the data, at least for purposes of computing the average number of wives in the sample. But where do you stop? If the data were 2, 3, 4, 2, and 7, would you eliminate the man with 7 wives? On what basis would you make the decision?

Trivially, you can always achieve a perfect regression fit to a set of data if you reduce it to just two points. But is creating a good fit what you're after? Don't you really want to understand what makes the data messy in the first place? In general, you can not achieve understanding by eliminating outliers. Still, as in all aspects of research, be ready to break this rule, too, when you think you'll learn something by doing so.

Tests of Significance

This is one of the hottest topics in quantitative social science. Some researchers argue that statistical tests of significance are virtually useless (Labovitz, 1971b). I wouldn't go that far, but tests of significance aren't magical, either. If you do not have a representative sample, for example, then a test of statistical significance is not much evidence of support for a hypothesis—it doesn't allow you to generalize beyond your particular sample of data. On the other hand, if you get significant results on a nonrandom sample, at least you can rule out the operation of random properties *in your sample* (Blalock, 1979:239-242).

Nor are the .01 and .05 levels of significance sacred, either. These numbers are simply conventions that have developed for convenience over the years. If you want to be especially cautious in reporting correlations, you can apply a severe test known as the *Bonferroni adjustment.* Pick a level of significance for reporting findings in your data—say, .05. If you have 66 variables in your analysis, then there are (66)(65)/2 = 2,145 tests of covariations in your matrix. Simply divide .05 by 2,145 and look for correlations of .00002 in the matrix (these will be reported as .000 on most computer output).

The Bonferonni inequality states that if you report these correlations as significant at the 5% level (the level you chose originally), then your report will be valid (see Koopmans, 1981, and Kirk, 1982). This is a very, very conservative test, but it will prevent you from making those dreaded Type I errors and reporting significant relationships that aren't really there.

On the other hand, this will increase your chance of making Type II errors—rejecting some seemingly insignificant relationships when they really *are* important. You might fail to show, for example, that certain types of exposure are related to contracting a particular disease, and this would have negative public health consequences. There's no free lunch.

Consider the study by Dressler (1980). He examined a sample of 40 informants in St. Lucia, all of whom had high blood pressure, on nine variables having to do with their ethnomedical beliefs and their compliance with a physician-prescribed treatment regimen. He reported the entire matrix of $(9 \times 8)/2 = 36$ correlations, 13 of which were significant at the 5% level or better.

Dressler might have expected just $(36 \times .05) = 1.8$ such correlations by chance. Three of the 13 correlations were significant at the .001 level. According to the Bonferonni inequality, correlations at the 36/.05 = .0014 level would be reportable at the .05 level as valid. Under the circumstances, however (13 significant correlations with only about 2 expected by chance), Dressler was quite justified in reporting all his findings, and not being overly conservative.

I feel that anthropologists who are doing fieldwork, and using small data sets, should be comfortable with tests of significance at the .10 level. On the other hand, you can always find significant covariations in your data if you lower the level of significance that you'll accept, so be careful. Remember, you're using statistics to get hints about things that are going on in your data. I can not repeat often enough the rule that real analysis (building explanations and suggesting plausible mechanisms that make sense out of covariations) is what you do *after* you do statistics.

The Shotgun Approach

A closely related issue concerns "shotgunning." This involves constructing a correlation matrix of all combinations of variables in a study and then relying on tests of significance to reach substantive conclusions. It is quite common for anthropologists to acquire measurements on as many variables as they have informants—and sometimes even *more* variables than informants.

There is nothing wrong with this. After a very short time in the field collecting ethnographic interview data, you will think up lots and lots of variables that appear potentially interesting to you. Include as many of them as you have time to ask on a survey without boring your informants.

The result of effective data collection is a large *matrix* of items-by-variables, like that in Figure 19.7(a). The items are the units of analysis. Most of the time, these units are people, but they could just as easily be cultures or schools. (If this is unfamiliar to you, see the section on units of analysis in Chapter 2.)

The matrix in Figure 19.7(a) is called a *profile matrix*. Each row is a profile of a single unit of analysis. In Figure 19.7(a), informant #1 is profiled by the following facts: She is 27 years old, is female (recorded as a 2 under the variable sex), has completed high school, has 1.5 hectares of land, comes from a household that had nine people in it, and is now part of a household that has six people in it.

For any profile matrix you can compare pairs of rows or pairs of columns to see how similar they are. If you compare rows, you find out how similar the units of analysis are to one another. If you compare columns, you find out how similar the variables are to one another.

Figure 19.7(b) shows a *similarity matrix* of variables. Imagine the list of variable names stretching several feet to the right, off the right-hand margin of the page, and several feet down, off the lower margin. That is what would happen if you had, say 100 variables about each of your informants. For each and every pair of variables in the matrix of data, you could ask: Are these variables related?

Now, if the matrix is symmetrical, then if x and y covary, so do y and x; that gets rid of half the pairs right there. Furthermore, no variable covaries with itself, so the entries in the diagonal have to be discounted. That still leaves $N(N-1)/2$ unique pairs of variables in a symmetric matrix. For 100 variables there are 4,950 pairs to consider.

Even a small matrix of 20 variables contains 190 unique pairs. It would take forever to go through each pair and: (a) decide whether it was worth spending the time to test for covariation in each case; (b) decide on the proper test to run (depending on the level of measurement involved in each case); (c) run the test; and (d) inspect and interpret the results.

Informant	Age	Sex	Education	Land Holding	Natal Household size	Current Household size
1	27	2	3	1.5	9	6
2	31	1	2	1.3	6	7
3	.	.	.	.	.	.
4	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Figure 19.7(a). Profile matrix of persons by variables.

	Age	Sex	Education	Land Holding	Household size	Ethnic Group	...
Age	—						
Sex		—					
Education			—				
Land Holding				—			
Household					—		
Ethnic Group						—	
⋮							

Figure 19.7(b). Similarity matrix of the variables (columns) in Figure 19.7(a).

There are two ways out of this fix. One way is to think hard about data and ask only those questions about covariation that seem plausible on theoretical grounds. It may not be important, for example, to test whether an informant's rank in a sibling set (first child, second child, etc.) covaries with blood pressure. On the other hand, how can we be sure?

The other way out of the fix is the shotgun strategy. You simply use a computer to transform your data matrix into a correlation matrix in which each cell is occupied by a Pearson's r. Then you look through the correlation matrix in search of significant covariations.

Kunitz et al. (1981) studied the determinants of hospital utilization and surgery in 18 communities on the Navaho Indian Reservation during the 1970s. They measured 21 variables in each community, including 17 independent variables (the average education of adults, the percentage of men and women who worked full time, the average age of men and women, the percentage of income from welfare, the percentage of homes that had bathrooms, the percentage of families living in traditional hogans, etc.) and four dependent variables (the rate of hospital use and the rates for the three most common types of surgery). Table 19.15 shows the correlation matrix of all 21 variables in this study

Kunitz et al. (ibid.) point out in the footnote to their matrix that, for $N = 18$, the 0.05 level of probability corresponds to $r = 0.46$ and the 0.01 level corresponds to $r = 0.56$. They could have expected

$$21 \times (20/2) \times .05 = 10.5$$

correlations significant at the 0.05 level and

$$21 \times (20/2) \times .01 = 2.1$$

correlations significant at the 0.01 level by chance. There are 73 correlations significant at the 0.05 level in Table 19.15, and 42 of those correlations are significant at the 0.01 level.

Kunitz et al. (ibid.) examined these correlations and were struck by the strong association of the hysterectomy rate to all the variables that appear to measure acculturation. I'm struck by it, too. This interesting finding was not the result of deduction and testing; it was the result of shotgunning. The finding is not proof of anything, of course, but it sure seems like a strong clue to me about how acculturation affects the kind of medical care that Navaho women receive.

The Problem With the Shotgun Approach

The problem with shotgunning is that you might be fooled into thinking that *statistically* significant correlations are also *substantively* significant.

TABLE 19.15
Correlation Matrix of All 21 Variables in Kunitz et al.'s Study of Hospital Use on the Navaho Reservation

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	Near hospital																				
2	Near surgery	0.67																			
3	Wage work	−0.24	−0.09																		
4	Welfare	0.52	0.46	−0.54																	
5	Education of men	−0.42	−0.43	0.73	−0.49																
6	Education of women	0.01	−0.21	0.67	−0.32	0.81															
7	Hogans	0.07	0.37	−0.26	0.72	−0.28	−0.40														
8	Bathrooms	−0.44	−0.57	0.63	−0.63	0.70	0.64	−0.47													
9	Household size	0.03	−0.16	−0.48	0.24	−0.34	−0.07	0.04	0.12												
10	Working women	−0.35	−0.36	0.68	−0.65	0.65	0.57	−0.48	0.62	−0.22											
11	Working men	−0.24	−0.18	0.73	−0.37	0.73	0.63	−0.16	0.41	−0.45	0.45										
12	Vehicles	0.34	−0.08	0.40	−0.17	0.22	0.52	−0.53	0.29	−0.06	0.31	0.40									
13	Median income	−0.45	−0.51	0.66	−0.60	0.67	0.60	−0.38	0.79	−0.07	0.83	0.48	0.27								
14	*Per capita* income	−0.25	−0.20	0.68	−0.46	0.61	0.40	−0.15	0.29	−0.68	0.48	0.51	0.30	0.47							
15	Age of women	−0.23	−0.03	−0.47	−0.15	−0.36	−0.66	−0.02	−0.46	−0.30	−0.37	−0.32	−0.56	−0.43	−0.04						
16	Age of men	0.28	0.15	−0.77	0.31	−0.60	−0.55	0.01	−0.67	0.06	−0.51	−0.68	−0.35	−0.63	−0.45	0.60					
17	Age of patients	0.13	0.47	0.13	−0.14	−0.21	−0.15	−0.18	−0.22	−0.36	−0.04	−0.26	−0.15	−0.26	−0.002	0.29	0.32				
18	Hysterectomies	−0.41	−0.48	0.62	−0.55	0.57	0.46	−0.23	0.75	−0.09	0.45	0.46	0.39	0.62	0.59	−0.37	−0.62	−0.28			
19	Appendectomies	−0.31	−0.40	0.32	−0.14	0.44	0.33	0.16	0.51	0.13	0.07	0.27	0.01	0.42	0.35	−0.31	−0.49	−0.56	−0.62		
20	Cholecystectomies	−0.23	−0.70	0.15	−0.43	0.35	0.34	−0.46	0.68	0.22	0.19	0.12	0.31	0.40	0.14	−0.17	−0.17	−0.35	0.70	0.49	
21	Hospital rate	−0.49	−0.24	0.02	−0.26	0.02	−0.25	−0.09	0.16	−0.19	−0.04	−0.34	−0.33	0.01	0.22	0.43	0.16	0.45	0.17	0.10	0.18

$N = 18$
0.46, $P = 0.05$.
0.56, $P = 0.01$.

SOURCE: Reprinted from *Social Science and Medicine, 15B*, S. J. Kunitz, D. Temkin-Greener, D. Broudy, and M. Haffner, "Determinants of hospital utilization and surgery on the Navajo Indian Reservation, 1972-1978," p. 74, © 1981, with kind permission from Pergamon Press Ltd, Headington Hill Hall, Oxford OX3 OBW, UK.

This is a real danger, and it should not be minimized (Labovitz, 1972). It results from two problems.

1. First of all, it might not be appropriate to analyze some pairs of variables using Pearson's r. Some pairs of variables are more appropriately analyzed using gamma, or chi-square, or some other statistic. Any particular significant correlation in a matrix may be an artifact of the statistical technique employed and not of any substantive importance. Running a big correlation matrix of all your variables may produce some statistically significant results that would be insignificant if the proper test had been applied.

2. Second, there is a known probability that any correlation in a matrix might be the result of chance. The number of expected significant correlations in a matrix is equal to the level of significance you choose, times the number of variables. If you are looking for covariations that are significant at the 5% level, then you only need 20 tests of covariation to find one such covariation by chance. If you are looking for covariations that are significant at the 1% level, you should expect to find one, by chance, once in every 100 tries. In a matrix of 100 variables with 4,950 correlations, you might find around 50 significant correlations at the 1% level by chance.

This does not mean that 50 correlations at the 1% level in such a matrix *are* the result of chance. They just *might* be. There can easily be 300 or more significant correlations in a matrix of 100 variables. If 50 of them (4,950/100) might be the result of chance, then how can you decide which 50 they are? You can't. You can never know for sure whether any particular correlation is the result of chance. You simply have to be careful in your interpretation of *every* correlation in a matrix.

Use the shotgun. Be as cavalier as you can in looking for statistically significant covariations, but be very conservative in interpreting their substantive (as opposed to their statistical) importance. Correlations are hints to you that something is going on between two variables. Just keep in mind that the leap from correlation to cause is often across a wide chasm.

If you look at Table 19.13 again, you can see just how risky things can be. A correlation of .60 is significant at the 1% level of confidence with a sample as small as 30. Notice, however, that the correlation in the population is 99% certain to fall between .20 and .83, which is a pretty wide spread. You wouldn't want to build too big a theory around a correlation that just might be down around the .20 level, accounting for just 4% of the variance in what you're interested in.

Remember these rules:

1. Not all significant findings at the 5% level of confidence are equally important. A very weak correlation of .10 in a sample of a million persons would be statistically significant, even if it were substantively trivial. By contrast, in small samples, substantively important relations may show up as statistically insignificant.

2. Don't settle for just one correlation that supports a pet theory; insist on several, and be on the lookout for artifactual correlations.

Thirty years ago, when most of the current generation of senior anthropologists were trained, there were no easy-to-use statistical packages. It was a real pain to run statistical tests. It made a lot of sense to think hard about which of the thousands of possible tests one really wanted to run by hand on an adding machine.

Computers have eliminated the drudge work in data analysis, but *they haven't eliminated the need to think critically about your results*. If anything, computers have made it more important than ever to be self-conscious about the interpretation of statistical findings. But if you *are* self-conscious about this issue, and dedicated to thinking critically about your data, then I believe you should take full advantage of the power of the computer to produce a mountain of correlational hints that you can follow up.

Finally, by all means, use your intuition in interpreting correlations; common sense and your personal experience in the field are powerful tools for data analysis. If you find a correlation between the distance from an African farmer's house to credit agencies, and whether the farmer's family brews its own beer in the home, you might suspect that this is just a chance artifact.

On the other hand, maybe it isn't. There is just as much danger in relying slavishly on personal intuition and common sense as there is in placing ultimate faith in computers. What appears silly to you may, in fact, be an important signal in your data. The world is filled with self-evident truths that are not true, and self-evident falsehoods that are not false. The role of science, based on solid technique and the application of intuition, is to sort those things out.

20

Multivariate Analysis

Most of the really interesting dependent variables in the social world—variables such as personality type, amount of risk-taking behavior, level of wealth accumulation, attitudes toward women or men—appear to be caused by a large number of independent variables, some of which are dependent variables themselves. The goal of multivariate analysis is to explain *how* variables are related and to develop a theory of causation that accounts for the fact that variables are related to one another.

There are two strategies for conducting multivariate analysis. One is called the *elaboration method*, developed by Paul Lazarsfeld and others at the Bureau of Applied Social Research at Columbia University (see Hyman, 1955; Rosenberg, 1968; Lazarsfeld et al., 1972; Zeisel, 1985). It involves nothing more than careful construction and inspection of percentage tables and the use of bivariate statistics.

The other kind of multivariate analysis involves an array of advanced statistical procedures. You will run into these procedures again and again as you read journal articles and monographs—things like multiple regression, partial regression, factor analysis, multidimensional scaling, analysis of variance, and so on. I'll discuss the elaboration method at length here and I'll just touch on the conceptual basis of the more complex procedures.

The Elaboration Method: Multivariate Percentage Tables

The elaboration method was popular in sociology up to about the mid-1970s but fell out of favor in some circles when computers became common. I still think that the method is excellent for use in anthropological fieldwork. This is not because elaboration analysis can be done without a computer. It *can* be done without a computer, but you know how I feel about that: Once you understand how any statistical method works, you should *only* do it by computer. The elaboration method is wonderful for field researchers because it keeps you close to your data.

It's going to take you a couple of hours to get through the next half-dozen pages on the elaboration method. The writing is clear and there's no heavy math, so they're not tough going; they're just plain tedious. But bear with me. If I give you 10 five-digit numbers to multiply, you'd probably use a calculator to make short work of the exercise, and quite properly, too. But in the fourth grade, you learned to do the operation by hand, with a pencil and paper, and it was an important learning experience.

The same applies here. Eventually, you'll simply give a computer a list of what you think are possible independent variables, specify a dependent variable, and let the machine do the rest. The next few pages will give you an appreciation of just what a multivariate analysis does. They will also give you the skills you need to conduct a multivariate analysis in the field, while your thoughts are fresh and you still have time to collect any data you find you need. So, be patient, pay close attention to the tables, and stay with it.

Building Tables

Suppose you are working in Peru and you suspect that Indians who move to Lima are no better off than Indians who remain in the villages. The Indians claim that they are seeking better jobs and better opportunities for

TABLE 20.1
Wealth by Residence for a Sample of 500 Peruvian Indians

Wealth	Residence: Rural	Residence: Urban	
Not poor	84 (34)	91 (36)	175
Poor	166 (66)	159 (64)	325
	250	250	500

χ^2 = .56 NS (Not Significant)

their children, but you think that they are not getting what they came to the city to find. You conduct a survey of 250 village residents from a particular region and 250 migrants who have gone to the city from the same region. Table 20.1 shows the relation between residence and accumulated wealth status for your sample.

Chi-square for this table is not significant. Assuming that you have measured wealth status using an appropriate index for both the urban and village environments, residence appears to make no difference in wealth accumulation among these informants.

After 5 years or more in the city, 74% of the sample remained poor. On the other hand, 26% managed to rise out of poverty in that time. Table 20.2 shows that the correlation between time in the city and the chance of remaining poor is .83, but the chance of climbing out of poverty rises with each year spent in the urban environment, and using the regression formula introduced in the last chapter, the projected chance of still being poor after 10 years in the city is .50.

Given that time won't cure poverty at the village level, the Indians' perception that time *might* work in their favor in the cities is substantially correct from these data.

Just as a significant bivariate relation can be rendered spurious by a common third variable, so can an apparently trivial relation become significant when you control for the right intervening variable. From other studies, we know that education is related to both residence and wealth; urban people tend to be both more wealthy than rural people, and more educated.

TABLE 20.2
Wealth Status by Time in the City for 250 Indian Migrants

Wealth Status	Years in City: < 1	≥ 1 < 3	≥ 3 < 5	≥ 5
Not poor	0	2	5	11
Poor	83	68	49	32

Time in City (in years)	Chance of Being Poor
1	83/83 = 1.0
2	68/70 = .97
4	49/54 = .91
5+	32/43 = .74
	r = .83

for 10 years, projected chance of remaining poor = .5

Tables 20.3 and 20.4 show the results of cross-tabulating wealth by education, and education by residence. Forty-four percent of those who completed the eighth grade have a family income above the poverty level, while just 26% of those who did not finish the eighth grade come from families whose income is above the poverty level. Chi-square for this table is highly significant.

TABLE 20.3
Wealth by Education for the Data in Table 20.1

Wealth	Education: Completed Eighth Grade	Education: Did Not Complete	
Not poor	113 (44)	62 (26)	175
Poor	146 (56)	179 (74)	325
	259	241	500

χ^2 = 17.28 Q = .38
$p < .001$

TABLE 20.4
Education by Residence for the Data in Table 20.1

Education	Residence: Rural	Residence: Urban	
Completed 8th grade	100 (40)	159 (64)	259
Did not complete	150 (60)	91 (36)	241
	250	250	500

$\chi^2 = 26.95$ Q = .45
$p < .001$

These tables indicate that urban people receive more education, and that this leads to greater wealth. We test this hypothesis by *elaborating* the relationship (in Table 20.3) of wealth by education *controlling for* residence. This is done in Table 20.5, which really consists of two separate tables, each of which can be analyzed statistically. (Place the control variables above the independent variable when constructing multivariate tables.)

Things are a bit more complex than we imagined at first. Among rural people, those who had completed the eighth grade are more than twice as likely (50% versus 23%) to have risen above poverty as those who had not finished school. Among urban people, by contrast, education doesn't make a significant difference in wealth status of poor migrant families. What's going on here? To find out, we continue to elaborate the analysis, looking at other variables and how they may be magnifying or suppressing relationships.

TABLE 20.5
Wealth by Education, Controlling for Residence

Wealth	Rural: ⩾ 8th-Grade Education	Rural: < 8th-Grade Education		Urban: ⩾ 8th-Grade Education	Urban: < 8th-Grade Education	
Not poor	50 (50)	34 (23)	84	63 (40)	28 (31)	91
Poor	50 (50)	116 (77)	166	96 (60)	63 (69)	159
	100	150	250	159	91	250

Rural: $\chi^2 = 18.89$ $p < .001$, Q = .55 Urban: $\chi^2 = 1.60$ NS

TABLE 20.6
Family Size by Education

Family Size	Education: ⩾ 8th Grade	Education: < 8th Grade	
> 4 children	170 (66)	129 (54)	299
⩽ 4 children	89 (34)	112 (46)	201
	259	241	500

$\chi^2 = 7.12$ $p = < .01$ Q = .25

As you add variables, of course, (as you make the multivariate analysis more elaborate), the number of tables required goes up, *as does the required sample size*. Adding a third variable, residence, to the analysis of wealth by education, requires two additional tables: residence by wealth and residence by education. Adding family size to the model, we need *three* additional tables.

Tables 20.6, 20.7, and 20.8 show the breakdown for family size by education, wealth by family size, and family size by residence.

In Table 20.6 we see that people with more education tend to have smaller families. In Table 20.7 we see that smaller families are 17% more likely to be above the poverty line. And Table 20.8 shows that rural families tend to be larger than urban families. It appears from these tables that economic status is related to family size more strongly than to education or to residence.

To disentangle things we look at the original relationship between wealth and residence, controlling for family size. This is shown in Table 20.9.

Now things are becoming much clearer. When we control for family size, the effect of residence on economic status remains insignificant for rural people, but it makes a big difference for urban residents. We can elaborate further by looking at the relationship between wealth and education, controlling for family size. As Table 20.10 shows, the influence of education on wealth is insignificant for large families, but is highly significant for small families.

TABLE 20.7
Wealth by Family Size

Wealth	Family Size > 4 Children	≤ 4 Children	
Not poor	84 (28)	91 (45)	175
Poor	215 (72)	110 (55)	325
	299	201	500

$x^2 = 16.56$ $p < .001$ Q = .36

To get the full picture, we now produce Table 20.11, which shows the bivariate relationship between wealth status and education, now controlling for *both* family size and residence simultaneously. From Table 20.11, it is obvious why sample size is so crucial. The more cells you have in an elaboration table, the larger the sample you need if you want to ensure that you don't have empty cells.

Sample Size—Again

A good way to plan your sample size requirements is to mock up the analytic tables you intend to produce (without any numbers in them) and see how many control variables you intend to use simultaneously. The total

TABLE 20.8
Family Size by Residence

Family Size	Residence Rural	Urban	
> 4 Children	167 (67)	132 (53)	299
≤ 4 Children	83 (33)	118 (47)	201
	250	250	500

$x^2 = 9.62$ $p < .01$ Q = .29

TABLE 20.9
Wealth by Residence, Controlling for Family Size

Wealth	Family Size: Rural > 4 Children	Rural ≤ 4 Children		Urban > 4 Children	Urban ≤ 4 Children	
Not poor	54 (32)	30 (36)	84	30 (29)	61 (52)	91
Poor	113 (68)	53 (64)	166	102 (71)	57 (48)	159
	167	83	250	132	118	250

$x^2 = .55$ NS $x^2 = 23.85$ $p < .001$ Q = .57

number of cells in a multivariate table depends on the number of control variables and the complexity of the variables. Dichotomous variables such as we're using here (e.g., large family versus small family) create fewer cells than do more complex variables (e.g., large, medium, small families).

Count the number of cells in the largest, most complex table you think you'll create in your analysis and, if you have the resources, make your sample large enough so that there are likely to be at least 20 values in each cell and 100 or more for each major control variable you intend to use. It

TABLE 20.10
Wealth by Education, Controlling for Family Size

Wealth	Family Size: > 4 Children ≥ 8th-Grade Education	> 4 Children < 8th-Grade Education		≤ 4 Children ≥ 8th-Grade Education	≤ 4 Children < 8th-Grade Education	
Not poor	54 (32)	30 (23)	84	59 (66)	32 (29)	91
Poor	116 (68)	99 (77)	215	30 (34)	80 (71)	110
	170	129	299	89	112	201

$x^2 = 2.22$ NS $x^2 = 26.98$ $p < .001$ Q = .66

TABLE 20.11
Wealth by Education, Controlling for Family Size and Residence

Wealth	Rural Residence					Urban Residence				
	Family Size					Family Size				
	> 4 Children		≤ 4 Children			> 4 Children		≤ 4 Children		
	≥ 8th-Grade Education	< 8th-Grade Education	≥ 8th-Grade Education	< 8th-Grade Education		≥ 8th-Grade Education	< 8th-Grade Education	≥ 8th-Grade Education	< 8th-Grade Education	
Not poor	34 (49)	20 (21)	16 (53)	14 (36)	84	20 (20)	10 (31)	43 (73)	18 (31)	91
Poor	36 (51)	77 (79)	14 (47)	39 (64)	166	80 (80)	22 (69)	16 (27)	41 (69)	159
	70	97	30	53	250	100	32	59	59	250
	$\chi^2 = 20.82 \quad p < .01$					$\chi^2 = 46.78 \quad p < .01$				

is often impossible to achieve these numbers in field research. This only means that you'll have to either (a) avoid making your analysis too elaborate or (b) settle for lower significance levels in your test results (.10 instead of .05, or .05 instead of .01, for example).

Reading a Complex Table

Reading across Table 20.11, we see that among urban families with at least an eighth-grade education and four or fewer children, 73% are above the poverty line. Among urban families with at least an eighth-grade education and *more than* four children, only 20% are above the poverty line. Among rural families with at least an eighth-grade education and with four or fewer children, 53% are above the poverty line. Among rural people with at least an eighth-grade education and more than four children, 49% are above the poverty line.

In other words, for rural people, education is the key to rising above poverty. So long as they increase their education, they are about as likely (49% versus 53%) to increase their economic status, whether or not they limit natality. This is not true for urban migrants. Unless they limit their natality *and* increase their education, they are not likely to rise above poverty (20% versus 73%). However, if they *do* limit their family size, *and* increase their education, then urban migrants are 20% more likely (73% versus 53%) than their rural counterparts to rise above poverty.

Taking Stock: What Do We Know So Far?

The lesson from this elaboration is clear. We saw from Table 20.2 that the longer the urban migrants remained in the city, the greater the likelihood that they would rise above poverty. But now we know a lot more. Unless they are prepared to both lower their natality and increase their education, poverty-stricken villagers in our sample are probably better off staying home and not migrating to the city. If they remain in their villages and just increase their education, they stand about a 50-50 chance of rising above poverty. But if they migrate to the city and only increase their education level, then the chances are very great (80%) that they will remain poor.

It is true that people in the urban areas get more education. That much is clear from Table 20.4. But if the urban migrants in our sample (all of whom started out as poor villagers) fail to limit natality, they lose the

TABLE 20.12
Family Size by Education, Controlling for Residence

	Rural Residence			Urban Residence		
Family Size	≥ 8th-Grade Education	< 8th-Grade Education		≥ 8th-Grade Education	< 8th-Grade Education	
> 4 children	70 (70)	97 (65)	167	100 (63)	32 (35)	132
≤ 4 children	30 (30)	53 (35)	83	59 (37)	59 (65)	118
	100	150	250	159	91	250
	x^2 = .55 NS			x^2 = 16.76 $p < .001$ Q = .52		

advantage that education would otherwise bring them. Rural people keep this advantage, irrespective of family size.

Explaining this finding, of course, is up to you. That's what theory is all about. A causal connection between variables requires a *mechanism* that explains how things work (see Chapter 2 if you need to go over the issue of covariation and causality). In this instance, we might conjecture that rural people have lower overall expenses, especially if they own their own land and homes. They usually have extended families that cut down the cost of child care and that provide no-interest loans during emergencies. They grow much of their own food, and having more children may help them farm more land and cut down on expenses.

Urban people get more education, and this gets them better paying jobs. But if they have many mouths to feed, and if they have to pay rent, and if they lack the financial support of kin close by, then these factors may vitiate any advantage their education might otherwise bring.

Taking the Elaboration Another Step

We can look for clues that support or challenge our theory by elaborating the model still further, this time using family size as the dependent variable. Table 20.12 shows the result of cross-tabulating family size by education, controlling for residence.

Chi-square for the left half of this table is insignificant, but for the right half it is highly significant. Rural informants with less than an eighth-grade education are almost twice as likely as urban informants with less than an eighth-grade education to have more than four children (65% versus 35%). Among rural informants, in fact, level of education has little or no effect on family size (70% of those with higher education have large families versus 65% of those with lower education).

Among urban informants, however, the effect of education on family size is dramatic. Highly educated urban informants are much more likely than less educated informants to have *large* families, from these data. This throws new light on the entire subject, and begs to be explained. We know that higher education without small families does not produce an increase in economic status for these poor migrants. We know, too, that most people, whether urban or rural, keep having large families, although large families are less prevalent among urbanites than among rural residents (132 out of 250 versus 167 out of 250).

To understand this case still further, consider Table 20.13, which cross-tabulates family size by wealth, controlling for both education and residence.

This table is illuminating. It shows that neither wealth nor education influences family size among rural informants. For urban residents, however, the story is quite different. As expected, those urban informants who have both increased their education and increased their wealth have small families.

Go through Table 20.13 carefully and make the appropriate comparisons across the rows and between the two halves. Compare also the results of this table with those of Table 20.11, in which wealth status was the dependent variable.

From these tables, we can now hazard a good guess about how these variables interact. A conceptual model of the process we've been looking at is shown in Figure 20.1. Most people in our sample are poor. Sixty-six percent of rural informants (166/250) and 64% of urban informants (159/250) are below the poverty line by our measurements. Among rural informants, education provides an edge in the struggle against poverty, irrespective of family size, but for urban migrants, education only provides an edge in the context of lowered family size.

Among those who remain in the villages, then, education may lead either to accumulation of wealth through better job opportunities, or it may have no effect. The chances are better that it leads to more favorable economic circumstances. Once this occurs, it leads to control of fertility. Among urban informants, education leads either to control of natality or not. If not, then education has practically no effect on the economic status of poor migrants. If it leads to lowered natality, then it may lead, over time, to a favorable change in economic status.

TABLE 20.13
Family Size by Wealth, Controlling for Education and Residence

Family Size	Rural Residence: ≥ 8th-Grade Education Poor	Not Poor	< 8th-Grade Education Poor	Not Poor		Urban Residence: ≥ 8th-Grade Education Poor	Not Poor	< 8th-Grade Education Poor	Not Poor	
> 4 children	34 (68)	36 (72)	20 (59)	77 (66)	167	20 (32)	80 (83)	10 (56)	22 (35)	132
≤ 4 children	16 (32)	14 (28)	14 (41)	39 (34)	83	43 (68)	16 (17)	18 (44)	41 (65)	118
	50	50	34	116	250	63	96	28	63	250
	$\chi^2 = 1.63$ NS					$\chi^2 = 58.46$ $p < .001$				

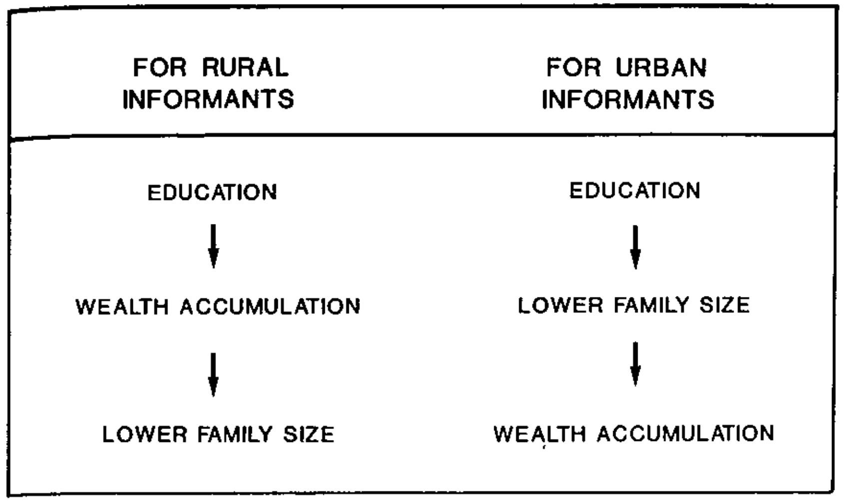

Figure 20.1. Model of how wealth, education, and family size interact in urban and rural environments for informants in Tables 20.11 and 20.13.

We can check this model by going back to our data on wealth status by number of years in the city to see if those migrants who are economically successful over time have both increased their education *and* lowered their natality. Plausible assumptions about time ordering of variables are crucial in building causal models. Knowing, for example, that wealthy villagers never move to the city, rules out some alternative explanations for the data presented here.

You get the picture. The elaboration method can produce subtle results, but it is quite straightforward to use and depends only on your imagination, on simple arithmetic (percentages), and on basic bivariate statistics. Using this method, you can actually get started on data analysis while you're still in the field.

A Recent Real Example

Keith and Wickrama (1990) studied how 136 female household heads in two rural Sri Lankan villages felt about the health care services in their district. Table 20.14 shows the results.

Try going over the data in Table 20.14. See if you can ferret out the key differences in reported use of the various medical facilities by women in

TABLE 20.14
An Elaboration Table Showing the Percentages of Reported Use and Opinion of Health Services Among Rural Sri Lankan Female Household Heads, by Age of Household Head

Use and opinion	Karametiya (N = 67)				Horawinna (N = 69)			
	Under 65 yr. (N = 44)	65 yr. or over (N = 23)	Under 65 yr. (N = 44)	65 yr. or over (N = 23)	Under 65 yr. (N = 50)	65 yr. or over (N = 19)	Under 65 yr. (N = 50)	65 yr. or over (N = 19)
Use of services	Rarely		Frequently or more often		Rarely		Frequently or more often	
Dispensary	38.6	60.9	61.4	39.1	24.0	57.9	76.0	42.1
Hospital	52.3	78.3	47.7	21.7	42.0	52.6	48.0	47.4
Ayuverda	45.5	60.9	54.6	39.1	20.0	10.5	80.0	89.5
Opinion of services	Poor		Satisfactory or good		Poor		Satisfactory or good	
Dispensary	29.5	39.1	71.5	60.9	48.0	47.4	52.0	52.6
Hospital	31.8	43.5	68.2	56.5	78.0	31.6	22.0	68.4
Ayuverda	45.5	56.5	54.6	43.5	6.0	—	94.0	100.0

SOURCE: From P. M. Keith and K. A. S. Wickrama, "Use and evaluation of health services by women in a developing country: Is age important?" *The Gerontologist,* 30, p. 267. Copyright © 1990 by The Gerontological Society of America. Used with permission.

the two villages, Karametiya and Horawinna. Which group uses the hospital more? Which group relies more on the *ayuverda* (traditional healer)?

Some General Advice on Data Analysis

How you actually conduct an elaboration analysis is up to you. There is no formula for deciding which variables to test. My advice is to follow every hunch you get. Other researchers insist that you have a good theoretical reason for including variables in your design and that you have a theory-driven reason to test for relationships among variables. They point out that anyone can make up an explanation for any relationship or lack of relationship after seeing a table of data or a correlation coefficient.

I consider this approach too restrictive, for three reasons.

1. First, I think that data analysis should be lots of fun, and it can't be unless it's based on following hunches. Most relationships are easy to explain, and peculiar relationships beg for theories to explain them. You just have to be very careful not to conjure up support for every significant relation, merely because it happens to turn up. There is a delicate balance between being clever enough to explain an unexpected finding and just plain reaching too far. As usual, there is no substitute for thinking hard about your data.

2. Second, it is really up to you during research design to be as clever as you can in thinking up variables to test. Just because you have no theory is no reason to avoid including variables in your design that you think might come in handy later on. Of course, you can overdo it. There is nothing more tedious than an interview that drones on for hours without any obvious point other than that the researcher is gathering data on as many variables as possible.

3. Third, the source of ideas has no necessary affect on their usefulness. You can get ideas from a prior theory or from browsing through data tables. The important thing is whether you can test your ideas and create plausible explanations for your findings. If others disagree with your explanations, then it is up to them to demonstrate that you are wrong, either by reanalyzing your data or by producing new data. Stumbling onto a significant relationship between some variables does nothing to invalidate the relationship.

So, when you design your research try to think about the kinds of variables that might be useful in testing your hunches. Use the principles in Chapter 5 and consider internal state variables (attitudes, values, beliefs); external state variables (age, height, gender, race, health status, occupation, wealth status, etc.); physical and cultural environmental variables (e.g., rainfall, socioeconomic class of a neighborhood); and time or space variables (Have attitudes changed over time? Do the people in one village behave differently from those in another otherwise similar community?).

In applied research, important variables are the ones that let you "target" a policy—that is, focus intervention efforts on subpopulations of interest (the rural elderly, victims of violent crime, overachieving third graders, etc.)—or that are more amenable to policy manipulation (knowledge is far more manipulable than attitudes, for example). No matter what the purposes of your research, or how you design it, the two principal rules of data analysis are:

1. If you have an idea, test it.
2. You can't test it if you don't have data on it.

Other Techniques for Multivariate Analysis

There are many multivariate techniques for finding subtle and complex relations in data. I will not deal with them at length in this book, but I do want to give you an idea of the range of tools available and enough information so you can read and understand research articles in which these techniques are used. I hope that this will arouse your curiosity enough so that you'll study these methods in more advanced classes.

Partial Correlation

Like the elaboration method, partial correlation allows you to control for the effects of a third (or fourth or fifth . . .) variable on a bivariate relationship. Making up cross-tabulated tables (cross-tabs, for short) is an intuitively satisfying way to go about data analysis, especially for a first cut where you're trying to get a "feel" for what's going on. The advantages of partial correlation are that it is a *direct* way to control for effects, and it can be applied even when your sample is very small. (Cross-tabs require larger samples in order to make sure that all the cells are adequately represented. Running chi-square on a table with empty cells will play havoc with your statistics.)

Suppose you have measured three variables for a sample of informants: Variable 1 is their perceived quality of life (PQOL), variable 2 is their

score on a test of "locus of control," and variable 3 is their income. Locus of control refers to a well-known scale that measures the extent to which people feel they are in control of their own lives. A low score signals that the informant feels that the so-called locus of control for his or her life is "out there" in the hands of others.

The measurements for PQOL and locus of control show a correlation r=.41, which is to say that 17% ($.41^2$) of the dependent variable (the score on the PQOL test) is accounted for by the independent variable (locus of control). What happens to this correlation when you control for income?

Suppose that PQOL and income have a correlation of .68 and that the correlation between locus of control and income is .31. The formula for partial correlation is

$$r_{12 \cdot 3} = \frac{r_{12} - (r_{13})(r_{32})}{\sqrt{(1 - r_{13}^2)(1 - r_{32}^2)}}$$

where $r_{12.3}$ = means "the correlation between variable 1 (PQOL) and variable 2 (locus of control), *controlling* for variable 3 (income) is . . ." (Partial correlation can be done on ordinal variables by substituting a statistic like tau for r in the formula above.) Substituting in the formula, $r_{12.3}$ = .29. Thus, just 8% ($.29^2$) of the variance in the mean PQOL is explained by locus of control, after removing the effect of income.

The test of significance for a partial correlation is based on the scores from the t-test table in Appendix D.

$$t = r_{12 \cdot 3}\sqrt{\frac{N-3}{(1 - r_{12 \cdot 3}^2)}}$$

You can use Appendix D to find the critical value of t with $N-3$ degrees of freedom.

A simple correlation is referred to as a *zero-order* correlation. The formula above is for a *first-order* correlation. The formula for a *second-order correlation* (controlling for two variables at the same time) is

$$r_{12 \cdot 34} = \frac{r_{12 \cdot 3} - (r_{14 \cdot 3})(r_{24 \cdot 3})}{\sqrt{(1 - r_{14 \cdot 3}^2)(1 - r_{24 \cdot 3}^2)}}$$

For a thorough review of partial correlation, see Thorndike (1978) and Blalock (1979).

Multiple Regression

In simple regression, we derive an equation that expresses the relationship between the independent and dependent variable. On the left-hand side of the equation, we have the unknown score for y, the dependent variable. On the right-hand side, you'll remember, we have the y-intercept (the score for y if the dependent variable were zero), and a constant that tells by how much to multiply the score on the independent variable for each unit change in that variable. So, a regression equation like

Starting Annual Income = \$16,000 + \$2000 × Years of College

or

Dep. Var. y = Constant + (Another Constant) (Ind. Var. x)

predicts that, on average, people with a high school education will start out earning \$16,000 a year; people with a year of college will earn \$18,000; and so on. A person with a Ph.D and 9 years of university education would be predicted to start at \$34,000.

In multiple regression, we build more complex equations that tell us how much each of *several* independent variables contributes to predicting the score of a single dependent variable. In simple regression, if height and weight are related variables, we want to know "How accurately can we predict a person's weight if we know their height?" A typical question for a multiple regression analysis might be "How well can we predict a person's weight if we know their height, *and* their gender, *and* their age, *and* their ethnic background, *and* their parents' income?" Each of those independent variables contributes something to predicting a person's weight.

Many computer programs used today produce what is called a *stepwise multiple regression.* You specify a dependent variable and a series of independent variables that you suspect play some part in determining the scores of the dependent variable. The program looks for the independent variable that correlates best with the dependent variable. Then it adds in the variables one at a time, accounting for more and more variance, until all the specified variables are analyzed, or until variables fail to enter because incremental explained variance is lower than a preset value, e.g., 1%.

In stepwise multiple regression, the program prints out the correlation coefficient for each independent variable with the dependent variable and also prints out a multiple correlation coefficient, represented by a capital letter R. The square of *that* statistic, R-squared, is the amount of variance

accounted for in the scores of the dependent variable, taking into account all the independent variables you specified. The programs will also print out the multiple regression equation. (If you are interested in learning how to derive multiple regression equations yourself, consult Blalock, 1979.)

Here are three examples of how multiple regression is actually used. John Poggie (1979) was interested in whether the beliefs of Puerto Rican fishermen about the causes of success in fishing were related to their actual success in fishing. He measured success by asking six key informants to rank 50 fishermen on this variable. Since his research was exploratory, he had a wide range of independent variables, three of which he guessed were related to fishing success: the fishermen's expressed orientation toward delaying gratification (measured with a standard scale), their boat size, and their years of experience at the trade.

The deferred gratification measure accounted for 15% of the variance in the dependent variable; years of experience accounted for another 10%; and boat size accounted for 8%. Together, these variables accounted for 33% of the variance in the success variable. Poggie's guess about which variables to test was pretty good.

Korsching et al. (1980) used a shotgun or shopping technique in their multivariate study of a group of families that were relocated when the land they lived on in Kentucky became part of a reservoir project. Their multiple regression found seven social and economic factors that accounted for at least some of the variance in relative satisfaction with new and old residences among those relocated. Those factors were: change in social activities (accounting for 18%); education (accounting for 4%); total family income before relocation (another 4%); change of financial situation (3%). Three other variables (satisfaction with resettlement payments, tenure status on the land, and length of residence in the old house) each accounted for 1% or less. All together, the seven independent variables accounted for 31% of the variance in satisfaction with the move.

Mwango (1986) studied small farming households in Malawi. He was interested in what made farmers decide to devote part of their land to growing new cash crops (tobacco and hybrid maize) rather than planting only the traditional crop, called "maize of the ancestors." His units of analysis were individual farms; his dependent variable was the ratio of land planted in tobacco and hybrid maize to the total land under plow.

Mwango's independent variables were (a) the total cultivated land area, in hectares; (b) the number of years a farmer was experienced in using fertilizers; (c) whether the farming household usually brewed maize beer for sale; (d) whether farmers owned any cattle at all; (e) whether farmers had had any training in animal husbandry practices from the local extension agents; (f) whether the family had an improved house (this required an index consisting of items such as a tin roof, cement floor, glass windows, and so on); (g) whether the farmer owned a bicycle; and (h) whether the farmer owned a plow and oxcart. All these independent variables together accounted for 48% of the variance in the dependent variable.

In social science research, multiple regression typically accounts for between 30% and 50% of the variance in any dependent variable, using between three and eight independent variables. In a list of six or eight independent variables accounting for, say, 40% of the variance, you will probably find that the first variable accounts for 10% to 20%. After that, the amount of variance in the dependent variable that is accounted for by any independent variable gets smaller and smaller. It is customary not to include independent variables that account for less than 1% of the variance in a multiple regression table (but there is no law against doing so).

If accounting for just 30% or 40% of the variance in what you're interested in seems puny, consider these two facts:

1. In 1983 the average white male had a life expectancy of 71.4 years in this country, or 26,061 days. The life expectancy for the average black male was 66.5 years, or 24,273 days. The *difference* is 1,788 days.

2. There were approximately 2.5 million births in Mexico in 1986 and around 47,500 infant deaths—that is, about 19 infant deaths per 1,000 live births. Compare these figures to the United States, where there were 3.7 million births and approximately 30,000 infant deaths, or about 8 per 1,000 live births. If the infant mortality rate in Mexico were the same as that in the United States, the number of infant deaths would be 20,000 instead of 47,500. The *difference* would be 27,500 infant deaths.

Suppose you could account for 10% of the *difference* in longevity among white and black males in the United States (179 days) or 10% of the *difference* between the United States and Mexico in infant deaths (2,750 children). Would that be worthwhile? How about 1%? To the extent that knowledge about phenomena leads to more effective control over those phenomena, I'd try to account for every percent I could.

Analysis of Variance

Analysis of variance, or ANOVA, is a statistical technique that applies to a set of averages. It is particularly popular in psychology and education

TABLE 20.15
A Typical Experiment in Which ANOVA Is Used in Educational Research

	Average Score on Pretest	Average Score on Posttest
Classes using new program	X_1	X_2
Classes not using new program	X_3	X_4

where groups of people are administered *tests* on which they get some kind of *score*. Each group, then, has an *average score* and these averages can be compared to see if they are significantly different.

For example, suppose educational researchers want to know whether a new method for teaching reading skills to fifth graders really makes a difference. They might divide the fifth-grade classes in a school district into two groups—one group that uses the new program and one group that does not. Both groups would be tested before the program is adopted and after the program is finished. (You'll recognize this method from Chapter 3 on experimental design.) Then the scores would be compared. Table 20.15 is a schematic of the scores that the researchers would be working with.

X_1, X_2, X_3, and X_4 are average scores. The question is: Are all the differences in these scores significant? Put another way (the null hypothesis), despite differences in the scores, are they really from identical populations? Does it make any real difference in their reading skills if fifth graders are exposed to the new program? There are four comparisons to make: between X_1 and X_2; X_3 and X_4; X_1 and X_3; and X_2 and X_4. Each of these comparisons can be done with a *t*-test, which is an analysis of the variance between two means.

But things can be much more complex. Suppose that each of the four cells in Table 20.15 is composed of several separate scores. That is, suppose that five classes are chosen for the new program and five are chosen not to participate, and that each of the groups of five classes is tested before and after the program. An analysis of the variance between more than two means requires the ANOVA technique.

Camilla Harshbarger (1986) investigated the relationship between the productivity of coffee farmers in one region of Costa Rica and their sources of credit. Her raw results are shown in Table 20.16.

Seven (16%) of the 44 farmers she interviewed did not use credit at all, and produced 21 *fanegas* of coffee per hectare (1 fanega = 1.58 bushels). Farmers who depended on commercial bank loans averaged 18.8 fanegas. Farmers who used one of the two cooperatives as credit sources averaged 26.6 and 17.6 fanegas. An analysis of variance showed that there was no

TABLE 20.16
Coffee Production by Credit Source for Four Costa Rican Farmers

	Beneficio	CSV	Bank	None
Number (%) of borrowers	3 (6.8)	12 (27)	22 (50)	7 (16)
Number of fanegas/ha	26.6	17.6	18.8	21

SOURCE: Harshbarger (1986).

significant difference in productivity among those farmers in Harshbarger's sample, no matter where they obtained credit, or even if they did not use credit.

Carole Jenkins (1981) surveyed 750 children in Belize for protein-calorie malnutrition (PCM). Her results are shown in Table 20.17.

An analysis of variance showed that there was a very strong relationship between ethnic group and the likelihood of suffering from childhood PCM.

Sokolovsky et al. (1978) compared the average number of "first-order relations" and the average number of "multiplex relations" among three groups of psychiatric patients who were released to live in a hotel in midtown New York City. (First-order relations are primary relations with others; multiplex relations contain more than one kind of content, such as relations based on visiting *and* borrowing money from, for example.)

One group of patients had a history of schizophrenia with residual symptoms; a second group had a history of schizophrenia without residual symptoms; and the third group had no psychotic history. An analysis of variance showed clearly that the average network size (both first-order and multiplex networks) was different among the three groups. From these data (and from field observation and in-depth interviews) Sokolovsky was able to draw strong conclusions about the ability of members of the three groups to cope with deinstitutionalization.

TABLE 20.17
PCM in Four Ethnic Groups in Belize

	Protein-Calorie Malnutrition		
	Yes	No	Total
Creole	28	170	198
Mestizo	43	184	227
Black Carib	38	144	182
Maya	47	96	143
Total			750

SOURCE: From Patterns of growth and malnutrition among preschoolers in Belize by C. L. Jenkins, *American Journal of Physical Anthropology, 56,* 175. Copyright © 1981 by Wiley-Liss. Reprinted by permission of Wiley-Liss, a division of John Wiley & Sons, Inc.

Whenever you observe three or more groups (age cohorts, members of different cultures or ethnic groups, people from different communities) and *count* anything (e.g., some behavior over a specific period of time, or the number of particular kinds of contacts they make, or the number of kilograms of fish they catch), then ANOVA is the analytic method of choice. If you are interested in the causes of morbidity, for example, you could collect data on the number of sick days among people in various social groups over a given period of time. Other dependent variables in which anthropologists are interested, and which are amenable to ANOVA, are things like blood pressure, number of minutes per day spent in various activities, number of grams of nutrients consumed per day, and scores on tests of knowledge about various cultural domains (plants, animals, diseases), to name just a few.

When there is one dependent variable (such as a test score) and one independent variable (a single intervention like the reading program), then no matter how many groups or tests are involved, a *one-way* analysis of variance is needed. If more than one independent variable is involved (say, several competing new housing programs, and several socioeconomic backgrounds), and there is a single dependent variable (a reading test score), then multiple-way ANOVA, or MANOVA, is called for. When two or more dependent variables are correlated with one another, then *analysis of covariance* (ANCOVA) techniques are used.

Multiple-way ANOVA allows you to determine if there are interaction effects among independent variables. Earlier in this chapter, we saw that independent nominal and ordinal variables (like level of education, family size, and residence) all *individually* affect wealth status, but that those independent variables also interacted with one another. The problem was that we could not tell *how much* they interacted. With interval-level scores on independent variables, we can use ANOVA to actually measure the interaction effects among variables—to determine if a variable has different effects under different conditions.

Like all popular multivariate techniques, ANOVA is available in the packaged computer programs that you are likely to deal with. Many research questions can be addressed using ANOVA in the field, especially these days with notebook computers and inexpensive statistical programs available. Consult Elifson et al. (1990) or Runyon and Haber (1991) for instructions on the fundamentals of analysis of variance.

Factor Analysis

Factor analysis is a technique for information packaging and data reduction. It has been around for about 60 years in the social sciences, although b.c. (before computers) it required truly Herculean efforts to use this technique. Factor analysis is based on complex statistics, but the principle behind the technique is simple and compelling. (For an introduction to factor analysis, see Rummel, 1970.)

In multiple regression, there is one dependent variable and several independent, or predictor variables. In factor analysis, all the variables in a matrix are considered together for their interdependence. The original, observed variables are thought of as reflections of (dependent on, in some way) some underlying dimensions (the so-called factors). The factors are thought to be reflections of the observed variables. The idea is to package and summarize the information contained in many variables (often dozens, or even hundreds) with a few underlying dimensions that covary with clumps of the variables in the original data. This reduces the original long list of variables to a shorter list that is easier to manipulate (e.g., to use in a regression analysis) and to interpret.

Since factors are extracted from a matrix of correlations among the variables in a study, they are really just new variables themselves. Factors consist of several "old" variables—variables in a correlation matrix that are closely related to one another. Some correlation matrices are very dispersed—they have very few significant correlations—while others are very dense. Dispersed matrices tend to have many factors, whereas dense matrices (where many variables are highly correlated with one another) tend to produce only a few factors.

The notion of variance is very important here. Factors account for chunks of variance—the amount of dispersion or correlation in a correlation matrix. Factors are extracted from a correlation matrix in the order of the amount of variance that they explain in the matrix. Some factors explain a lot of variance, while others may be very weak and are discarded by researchers as not being useful. In a dense matrix, then, only a few factors may be needed to account for a lot of variance; in a dispersed matrix, many factors may be needed.

The most common statistical solution for finding the underlying factors in a correlation matrix is called the *orthogonal solution*. In orthogonal factor analyses, factors are found that have as little correlation with each other as possible. Other solutions, which result in intercorrelated factors, are also possible (the various solutions are options that you can select in all the major statistical packages, like SAS and SPSS). Some researchers say that these solutions, although messier than orthogonal solutions, are more like real life.

So-called factor loadings are the correlations between the new factors and the old variables that are replaced by factors. All the old variables *load* on each new factor. The idea is to establish some cutoff (say, a correlation of 0.4) below which you would not feel comfortable accepting that an old variable "loaded onto" a factor. Then you simply go through the list of old variables and pick out those that load sufficiently high on each new factor. Finally, you look at the list of variables that constitute each factor and decide what the factor *means*.

As you saw in Chapter 13, factor analysis is widely used in building reliable, compact scales for measuring variables in the field. Typically, anthropologists find either: (a) that there are no existing, well-tested scales they can use in the field for the things in which they are interested, or (b) that if scales do exist, the instruments are not transportable to another culture.

Suppose, for example, that you are interested in attitudes about gender role changes among women. From ethnographic work you suspect that the underlying forces of role changes have to do with premarital sexuality, working outside the home, and development of an independent social and economic life among women. You make up 50 attitudinal items in the local language and collect data on those items from a sample of informants.

Factor analysis will help you decide whether the 50 items you made up really test for the underlying forces you think are at work. If they do, then you could use a few benchmark items (that load high on the factors), and this would save you (and others) from having to ask all informants about all 50 items you made up. You would still get the information you need—or much of it, anyway. The amount would depend on how much variance in the correlation matrix each of your factors accounted for. An example should make all this a lot clearer.

Marchione (1980) used factor analysis in his study of the nutritional status of 1-year-olds in Jamaica. He measured the height and weight of 132 children and compared these measurements with international standards to determine the nutritional status of the children in his sample. He also collected data on 31 measures relating to households. These included household size, income, and food expenditures; diet variety; presence of mother or father; mother's age; distance to piped water; and so on.

Nineteen of the 31 measures had some statistically significant relation with the height and weight measurements. Marchione reported that: "Although every possible bivariate . . . association was examined, a problem of interpretation remains—*the problem of interrelationships among the household measures themselves*. Examination of the matrix of interrelationships . . . displays a bewildering array of intercorrelations" (ibid.:242, italics added).

Marchione found that some household measures had no direct relationship with either height or weight, but were significantly correlated with other household measures that *were*. For example, employment history was related to both income and food expenditure. The latter two variables are significantly correlated with weight status of 1-year-old children, but the first variable is not.

Marchione wanted a way to use all his data on household measures, without risking throwing away potentially useful information. He factor analyzed the matrix of 132 households and the variables he had studied. Twelve factors emerged, of which 8 seemed to have some intuitive appeal. The first factor was composed of five variables, as follows:

Household Measure	Factor Loading
Father present	.87
Father support	.72
Mother present	.50
Mother's employment	−.30
Mother's age	.25

Marchione's task was to determine what this factor (this package of variables) represented. He decided that the variables in this factor were all related to family stability and integrity, and he labeled the factor "family cohesion." (The negative loading for mother's employment means that when fathers support the family, then mothers are less likely to.)

Marchione labeled the seven other factors he extracted "guardian maturity," "clinic case demand," "household diet," "age transition," "agricultural subsistence," "dependency stress," and "monetary wealth." Then he treated each factor as if it were a new independent variable and he looked at how they correlated with the two dependent variables in his study—weight status and height (or length) status of the 132 1-year-olds. Table 20.18 shows the results.

Overall, the eight factors accounted for about a quarter of the variation in weight or length status among the 1-year-olds studied. (The total variance is found by squaring the separate correlations and adding them together. The multiple correlation, *R*, is the square root of that result).

Factor analysis has become popular in anthropological research because it leaves a lot of room for interpretation by researchers (or informants) of the results. For example, Marchione noticed that there was a negative relationship between nutritional status of 1-year-olds and the degree to which households live off of subsistence agriculture. He explained this by the fact that plots are too small for household size.

Also, in Marchione's data, child growth was retarded as dependency stress increased. Dependency stress was what Marchione labeled a pack-

TABLE 20.18
Correlation Between Factors and Dependent Variables in Marchione's Study of 1-Year-Olds in Jamaica

Factor		Nutritional Status	
		Weight Status N = 132	Length Status N = 114
1		−.25*	−.28*
2		−.22*	−.35*
3		.23*	.15
4		.22*	.11
5		−.08	−.14
6		−.03	−.14
7		.12	.07
8		.09	.03
	Multiple R	.49	.53
	Variance accounted for	24%	28%

SOURCE: Marchione (1980: 153).
*Correlation significant at the .05 level.

age of variables having to do with competition for resources among preschool children, and between them and older children in a household. Marchione also noticed that a child's weight-for-age improved as family cohesion improved. In each case, Marchione was led by the factor analysis to some insights about the phenomenon he was studying.

In some cases, a *lack* of correlation between factors and dependent variables may require an explanation and lead to insights. For example, Marchione found that neither the household diet factor nor the household wealth factor were significantly related to child growth. He interpreted this as a methodological problem. Diet was measured by a single 24-hour recall, which doesn't reflect any diversity and which is also highly unreliable and invalid. Wealth was measured by asking people about their income "last week." These unreliable self-reported data were too crude, according to Marchione, to provide a meaningful correlation with much of anything.

Multidimensional Scaling Analysis (MDS)

MDS is another multivariate data reduction technique. Like factor analysis, it is used to tease out underlying relationships among a set of observations. Also like factor analysis, MDS requires a matrix of measures of associations—

e.g., a correlation matrix based on things like r, tau, gamma, etc. But unlike factor analysis, MDS can handle *metric* and *nonmetric data*.

When you measure something like how strongly people feel about something, the numbers you assign to their feelings don't have the same meaning as, say, numbers that express distances in miles or kilograms of game meat killed per month. The latter numbers are metric because they are grounded in well-understood units of measurement. Most attitude and cognition data are nonmetric. MDS is particularly useful for anthropologists, since a lot of the measurements we make are nonmetric. Also, MDS produces a graphic display of the relationship among a set of items in a cultural domain and this, too, makes it really suited to the kinds of things anthropologists study. (See Romney et al., 1972, for an excellent introduction to the use of MDS in anthropology.)

How MDS Works

Suppose you measure three variables, A, B, and C, using Pearson's r. The association matrix for these three variables is in the *inside box* of Table 20.19.

TABLE 20.19

Matrix of Association among Four Variables

	A	B	C	D
A	x	.50	.80	.30
B		x	.40	.65
C			x	.35
				x

Clearly, variables A and C are more closely related to one another than are A and B, or B and C. You can represent this with a triangle, as in Figure 20.2(a).

In other words, we can place points A, B, and C on a plane in some position relative to each other. The distance between A and B is longer than that between A and C (reflecting the difference between .5 and .8); and the distance between B and C is longer than that between A and C (reflecting the difference between .4 and .8). The numbers in this graph are *similarities*: The lower the correlation, the longer the distance; the higher the correlation, the shorter the distance.

With just three variables, it is easy to plot these distances in proper proportion to one another. For example, the distance between B and C is

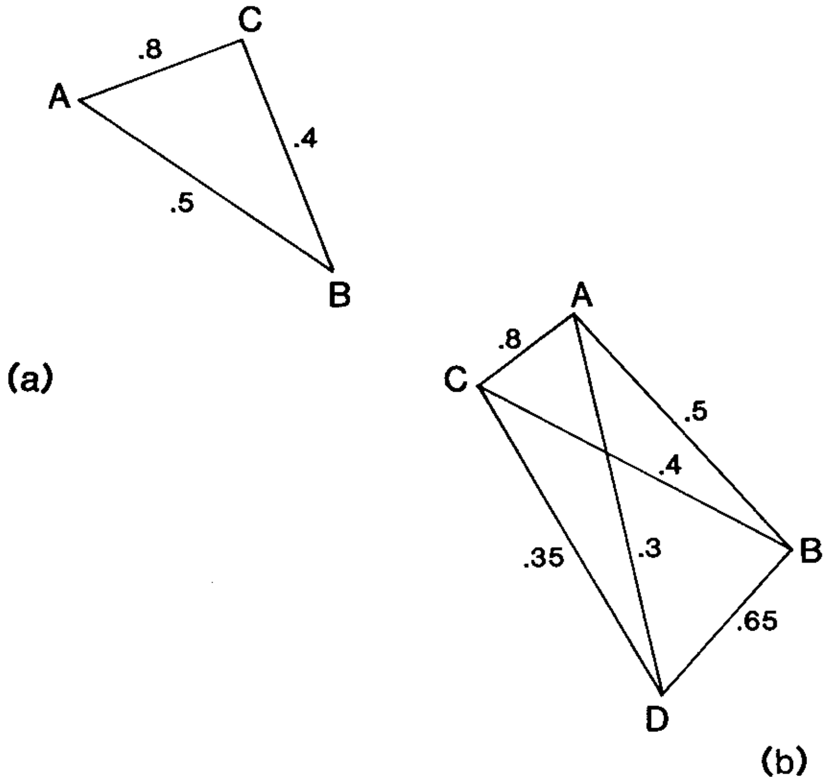

Figure 20.2. Two-dimensional plot of the relationship among three variables (a) and among four variables (b).

twice that of A and C in Figure 20.2(a). Figure 20.2(a) contains *precisely* the same information as the inside box of Table 20.19—but in graphic form.

With four variables, things get considerably more complicated. With four variables there are *six* relations to cope with. These relations are shown in the large box of Table 20.19. Only one two-dimensional graph (apart from rotations and enlargements) can represent the relative distances among the six relations in Table 20.19. The graph is shown in Figure 20.2(b).

Figure 20.2(b) is a two-dimensional graph of six relations in *almost* proper proportions. It is often impossible to achieve perfect proportionality in a graph of six relations if we have only two dimensions to work with. One way out of this is to depict the six relations in Table 20.19 in three dimensions, instead of only two. The extra dimension would give us plenty of room to move around, and we could better adjust the proportionality of the distances between the various pairs of variables.

In principle, you can perfectly represent the relative relations among N variables in $N - 1$ dimensions, so that any graph of six variables can be perfectly represented in five dimensions. But even a three-dimensional graph is sometimes hard to read. What would you do with a five-dimensional graph?

Most researchers specify a two-dimensional solution when they run an MDS computer analysis, and hope for the best. MDS programs produce a statistic that measures the *stress* in the graph produced by the program. This is a measure of how far off the graph is from one that is perfectly proportional. The lower the stress, the better the solution. This means that a cluster of variables in an MDS graph with low stress is likely to reflect some reality about the cognitive world of the people being studied.

A Physical World Example

An example will make this clearer. Table 20.20 shows the actual distance in miles between all pairs of nine cities in the United States (this example is from Borgatti, 1992b).

TABLE 20.20
Distances Between Nine U.S. Cities, in Miles

	Boston	NY	DC	Miami	Chicago	Seattle	SF	LA	Denver
Boston	0								
NY	206	0							
DC	429	233	0						
Miami	1504	1308	1075	0					
Chicago	963	802	671	1329	0				
Seattle	2976	2815	2684	3273	2013	0			
SF	3095	2934	2799	3053	2142	808	0		
LA	2979	2786	2631	2687	2054	1131	379	0	
Denver	1949	1771	1616	2037	996	1037	1235	1059	0

SOURCE: Borgatti (1992a:24).

Note two things about the numbers in this table. First, the numbers are *dissimilarities*. Bigger numbers mean that things are farther apart—less like each other. Smaller numbers mean that things are more similar. Similarity and dissimilarity matrices are known collectively as *proximity* matrices because they tell you how close or far apart things are. Second, the numbers are reasonably accurate measures of a physical reality—distance between points on a map—so they are metric data.

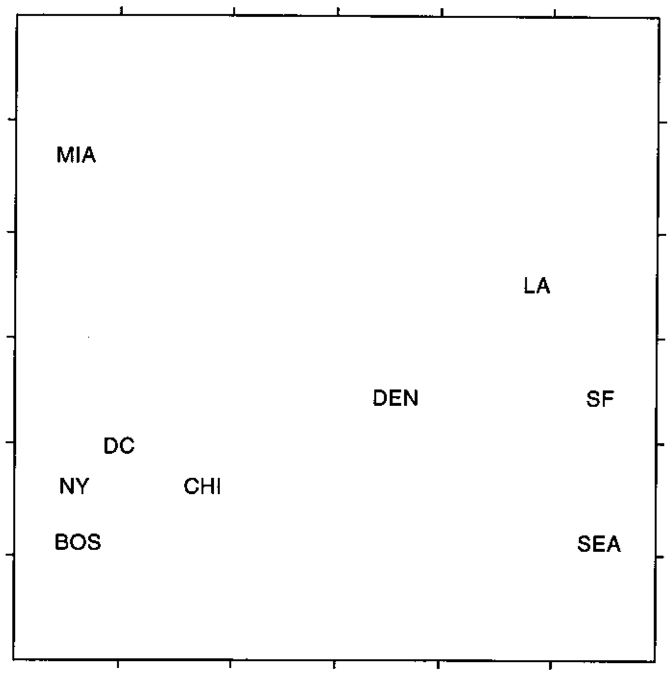

Figure 20.3. Two-dimensional MDS solution for the numbers in Table 20.20.

In principle, there should be a two-dimensional solution with low stress that fits the numbers in this table. I used ANTHROPAC (Borgatti, 1992a) to run MDS on these data, and the solution is shown in Figure 20.3.

Figure 20.3 looks suspiciously like a map of the United States. All nine cities are placed in proper juxtaposition to one another, but the map looks sort of upside-down and backward. If we could only flip the map over from left to right and from bottom to top. . . . Multidimensional scaling programs are notoriously unconcerned with details like this. So long as they get the juxtaposition right, they're finished. Figure 20.3 shows that the program got it right. You can rotate any MDS graph through 360° and it will still be the same graph.

A Cognitive World Example

Here's an example using nonmetric data. Susan Weller (1983) studied perceptions of illness among rural and urban Guatemalan women. She asked 20 women to list as many illnesses as they could think of. Then she took the 27 most frequently named illness, put each named illness on a card, and asked

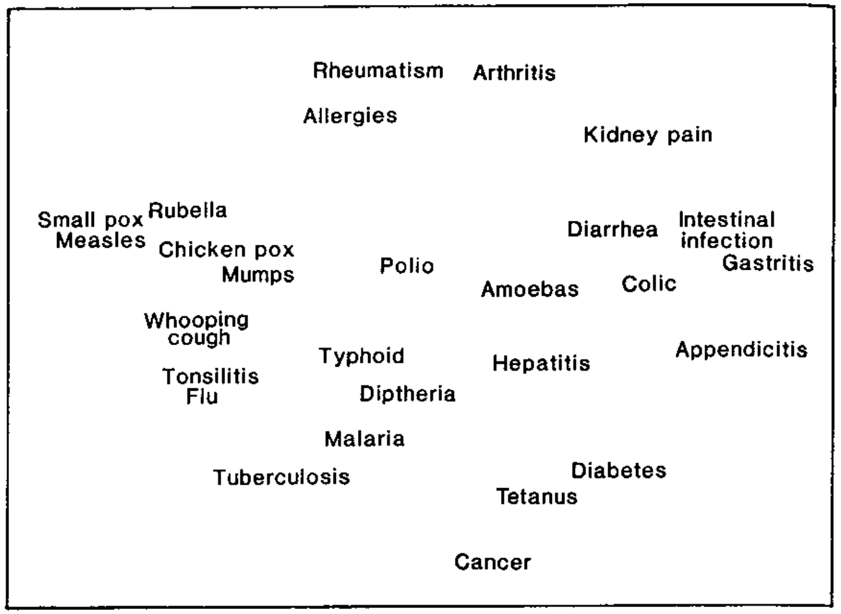

Figure 20.4. MDS representation of 27 illnesses for urban Guatemalan women.
SOURCE: From "New data on intracultural variability" by S. C. Weller, 1983, *Human Organization, 42*, p. 250, Copyright 1983. Reprinted with permission of the Society for Applied Anthropology.

24 other women to sort the cards into piles, according to similarity. The women were allowed to use any criteria they wished for making the piles.

Weller created a correlation matrix from the similarity data. That is, she produced a 27×27 illness-by-illness correlation matrix for the illnesses. The more any pair of illnesses had been placed in the same piles by the informants, the more similar the illnesses.

Then Weller did a multidimensional scaling analysis to represent how her informants collectively perceived the 27 illnesses. Weller said that "the two-dimensional solution was considered adequate because addition of a third dimension only decreased the stress from .142 to .081" (ibid.:249). There are no rules for deciding what "low" or "high" stress is in MDS. As with many things in research, it's a matter of judgment.

Figure 20.4 shows the graph solution that Weller found for her urban sample. As you can see, the MDS program converts similarities (correlations) into graphic distances. The illness terms that were judged to be similar, then, are closer together in Figure 20.4, and the terms judged to be dissimilar by informants are farther apart.

Try naming the clusters in Figure 20.4. There is a clump of illnesses on the right that might be called "gastrointestinal disorders." On the left there is a clump of "childhood disorders." Those, at least, are the "labels" that struck Weller as appropriate. I agree with her intuition about this. What do *you* think?

Remember: All I've done is *label* the group of illnesses on the right in Figure 20.4. The fact that I can come up with a label says absolutely nothing about whether I understand what is going on. It is possible to label anything, once you're confronted with the task. This means that you must be especially careful in the use of factor analysis, MDS, and other dredging techniques that present you with something to explain.

On the other hand, the mere fact that I might make a mistake in my interpretation of the results doesn't stop me from using these techniques. Use every technique you can think of in data analysis, and let your experience guide your interpretation. Interpretation of results is where data analysis in all science ultimately becomes a humanistic activity.

Cluster Analysis

Like factor analysis and MDS, cluster analysis is a descriptive tool for exploring relations among items in a matrix—for finding what goes with what. You start with a similarity matrix, like a matrix of Pearson correlation coefficients. If you factor the matrix, you find underlying variables that may encompass the variables in the original data. If you scale the matrix (MDS), you get a map that shows you graphically the relations among the items. Clustering tells you which items go together and in what order. Thus, in the MDS example from Weller (1983) above, a cluster analysis would let us check the guess about that chunk of illnesses labeled "gastrointestinal disorders."

I'm going to explain cluster analysis in some detail. It's a very important descriptive tool, but as we go through the next few paragraphs, keep in mind two things: First, clustering is just a technique for finding the similarity chunks. It doesn't label those chunks. That part is a Rorschach test, and *you* get to do it—just as you would name the factors in factor analysis or the clumps and dimensions in an MDS graph. Second, as with so many methods, different treatments of your data produce different outcomes. The next few pages will make it very clear just how much *you and only you* are responsible for every choice you make in data analysis.

How Cluster Analysis Works

Consider the following example from De Ghett (1978:121):

1 3 7 9 14 20 21 25

The distance between 1 and 3 is 2. The distance between 21 and 25 is 4. So, in a numerical sense, 1 and 3 are twice as similar to one another as 21 and 25 are to one another. Table 20.21 shows the dissimilarity matrix for these numbers.

TABLE 20.21
Dissimilarity Matrix for Clustering

	1	3	7	9	14	20	21	25
1	0							
3	2	0						
7	6	4	0					
9	8	6	2	0				
14	13	11	7	5	0			
20	19	17	13	11	6	0		
21	20	18	14	12	7	1	0	
25	24	22	18	16	11	5	4	0

SOURCE: From "Hierarchical cluster analysis" by V. J. De Ghett, in *Quantitative Ethology* (P. W. Colgan, Ed.), p. 123. Copyright © 1978. Reprinted by permission of John Wiley & Sons, Inc.

There are several ways to find clusters in this matrix. Two of them are called *single-link* or *closest-neighbor* analysis and *complete-link* or *farthest-neighbor* analysis (there are others, but I won't go into them here). In single-link clustering, we use only the numbers adjacent to the diagonal: 2, 4, 2, 5, 6, 1, 4. The two clustering solutions (again, done with ANTHROPAC) are shown in Figure 20.5.

In the single-link solution, the two closest neighbors are 20 and 21. They are exactly one unit of distance apart, and there is a 1 in the diagonal of the original matrix where 20 and 21 come together. In Figure 20.5(a), 20 and 21 are shown joined at level 1. The numbers 1, 3 and the numbers 7, 9 are the next closest neighbors. They are both two units apart. Figure 20.5(a) shows them joined at level 2.

Once a pair is joined, it is considered a unit. The pairs 1, 3 and 7, 9 are joined together at level 4 because they are four units apart (the nearest neighbor to the pair 1, 3 is 7, which is four units from 3). The pair 21, 25 are also four units apart. However, 20, 21 are already joined, so 25 joins this pair at level 4. The connections are built up to form a tree.

Figure 20.5(b) shows the complete-link (or farthest neighbor) clustering solution for the data in Table 20.21. In complete-link clustering, all the numbers in Table 20.21 are used. Once again, the pair 20, 21 is joined at level 1 because the pair is just one unit apart. The pairs 1, 3 and 7, 9 join at level 2.

(a) Single-Link Clustering

```
                      1 2 2 2
          1 3 7 9 4 0 1 5
Level   -----     - - - - - - - - - -
     1  . . . . . XXX .
     2  XXX XXX . XXX .
     4  XXXXXXX . XXXXX
     5  XXXXXXXXX XXXXX
     6  XXXXXXXXXXXXXXX
```

(b) Complete-Link Clustering

```
                      1 2 2 2
          1 3 7 9 4 0 1 5
Level   -----     - - - - - - - - - -
     1  . . . . . XXX .
     2  XXX XXX . XXX .
     5  XXX XXX . XXXXX
     7  XXX XXXXX XXXXX
    13  XXXXXXXXX XXXXX
    24  XXXXXXXXXXXXXXX
```

Figure 20.5. Cluster analysis of data in Table 20.21.

At this point, the complete-link and single-link solutions are identical. At the next level, though, things change. The neighbors of 20, 21 are 14 and 25. The farthest neighbor from 14 to 20, 21 is 21. The distance is seven units. The farthest neighbor from 25 to 20, 21 is 20. The distance is five units. Since five is less than seven, 25 joins 20, 21 at level 5. But the two pairs 1, 3 and 7, 9 are not joined at this level.

The only number not yet joined to some other number is 14. It is compared to its farthest neighbors in the adjacent clusters: 14 is 11 units away from 25 (which is now part of the 20, 21, 25 cluster) and it is 7 units away from the 7, 9 cluster. So, at level 7, 14 is joined to 7, 9. The same game is played out with all the clusters to form the tree in Figure 20.5(b).

Clusters of Cities

The complete-link method tends to produce more clusters than the single-link method. The method you choose determines the results you get. Figure 20.6 shows what happens when we use the single-link and complete-link clustering methods on the data in Table 20.20.

To me, the complete-link method seems better with these data. Denver "belongs" with San Francisco and Los Angeles more than it belongs with

(a) Complete-Link

```
      M B     C     S D
      I O N D H S L E E
      A S Y C I F A A N
Level 4 1 2 3 5 7 8 6 9
----  - - - - - - - - -
 206  . XXX . . . . . .
 379  . XXX . . . XXX .
 429  . XXXXX . . XXX .
 963  . XXXXXXX . XXX .
1131  . XXXXXXX XXXXX .
1307  . XXXXXXX XXXXXXX
1504  XXXXXXXXX XXXXXXX
3273  XXXXXXXXXXXXXXXXX
```

(b) Single-Link

```
      M S       B     C D
      I E S L O N D H E
      A A F A S Y C I N
Level 4 6 7 8 1 2 3 5 9
----  - - - - - - - - -
 206  . . . . XXX . . .
 233  . . . . XXXXX . .
 379  . . XXX XXXXX . .
 671  . . XXX XXXXXXX .
 808  . XXXXX XXXXXXX .
 996  . XXXXX XXXXXXXXX
1059  . XXXXXXXXXXXXXXX
1075  XXXXXXXXXXXXXXXXX
```

Figure 20.6. Complete-link and single-link cluster solutions for the data in Table 20.20.

Boston and New York. But that may be my own bias. Coming from New York, I think of Denver as a western U.S. city. I've heard people from San Francisco talk about "going back east to Denver for the weekend."

Discriminant Function Analysis (DFA)

Discriminant function analysis (DFA) is used to predict membership in categorical (nominal) variables from ordinal and interval variables. For example, we may want to predict which of two (or more) groups an individual belongs to: male or female; those who have been labor migrants versus those who have not; those who are high, middle, or low income; those in favor of something and those who are not; and so on.

DFA is the technique developed for handling this problem. It has been around for a long time (Fisher, 1936) but, like most multivariate techniques, it is not feasible to do DFA on significant amounts of data without a computer.

DFA can be very useful for research in anthropology. Gans and Wood (1985) used this technique for predicting whether Samoan women informants were "traditional" or "modern" with respect to their ideal family size. (If informants stated that they wanted three or fewer children, then Gans and Wood placed those informants in a category they labeled "modern." Informants who said they wanted four or more children were labeled "traditional.") DFA showed that just six of the many variables that Gans and Wood had collected allowed them to predict correctly which category a woman belonged to in 75% of all cases. The variables were such things as age, owning a car, level of education, etc.

In another case, Lambros Comitas and I surveyed two groups of people in Athens, Greece: those who had returned from having spent at least 5 years in West Germany as labor migrants, and those who had never been out of Greece. We were trying to understand how the experience abroad might have affected the attitudes of Greek men and women about traditional gender roles (Bernard & Comitas, 1978). Our sample consisted of 400 persons: 100 male migrants, 100 female migrants, 100 male nonmigrants, and 100 female nonmigrants. Using DFA, we were able to predict with 70% accuracy whether an informant had been a migrant on the basis of just five variables.

There are some things you need to be careful about in using DFA, however. Notice that our sample in the Athens study consisted half of migrants and half of nonmigrants. That was because we used a disproportionate, stratified sampling design to ensure adequate representation of returned migrants in the study. Given our sample, we could have guessed whether one of our informants was a migrant with 50% accuracy, without any information about the informant at all.

Now, only a very small fraction of the population of Athens consists of former long-term labor migrants to West Germany. The chances of stopping an Athenian on the street and grabbing (at random) one of those returned labor migrants was less than 5% in 1977 when we did the study.

Suppose that, armed with the results of the DFA that Comitas and I did, I asked random Athenians five questions, the answers to which allow me to predict 70% of the time whether any respondent had been a long-term labor migrant to West Germany. No matter what the answers were to those questions, I'd be better off predicting that the random Athenian was *not* a returned migrant. I'd be right more than 95% of the time.

Furthermore, why not just ask the random survey respondent straight out: "Are you a returned long-term labor migrant from West Germany?" With such an innocuous question, presumably I'd have gotten a correct answer at least as often as our 70% prediction based on knowing five pieces of information.

The answer is that DFA can be a powerful descriptive device, even if you don't use it as a prediction technique. Gans and Wood, for example, felt that it was inappropriate to ask Samoan women directly whether they (the informants) were "traditional" or "modern." Combined with ethnography, the DFA gave them a good picture of the variables that go into Samoan women's desired family size.

Similarly, Comitas and I were able to describe the attitudinal components of gender role changes by using DFA. If you are careful about how you interpret a discriminant function analysis, then it can be a really important addition to your statistical tool kit.

Path Analysis

Path analysis is a technique for testing conceptual models of multivariate relationships. It was developed by the geneticist Sewall Wright in 1921, and has been popular in sociology since the 1960s (see Duncan, 1966). Path analysis has been used increasingly in anthropology since 1974, when Hadden and De Walt discussed it in an excellent review article. I expect that discriminant analysis and path analysis will become important multivariate techniques in anthropology.

In multiple regression, we know (a) which independent variables help to predict some dependent variable, and (b) how much variance in the dependent variable is explained by each independent variable. But multiple regression is an inductive technique: It does not tell us how *much* a particular independent variable influences the outcome of a dependent variable. And it doesn't tell us which are the antecedent variables, which are the intervening variables, and so on.

Those are things that researchers have to decide. Path analysis is a technique for deductive analysis. It allows us to test a model of how the independent variables in a multiple regression equation may be influencing each other—and how this ultimately leads to the dependent variable outcome.

In one sense, path analysis really doesn't add anything to a multiple regression analysis. It is simply a measure of the "direct influence along each separate path" in a system of multivariate relations, and a way to find "the degree to which variation of a given effect is determined by each particular cause" (Wright, 1921). Path analysis relies on knowing "the correlation among the variables in a system" and on any knowledge that the researcher happens to have about the causes of those correlations (ibid.).

In other words, path analysis is a statistical technique that depends crucially on the researcher's best guess about how a system of variables really works. It's a nice combination of quantitative and qualitative methods. Here's an example.

Thomas (1981) studied leadership in Niwan Witz, a Tojalabal Mayan village. He was interested in understanding what causes some people to emerge as leaders, while others remain followers. From existing theory, Thomas thought that there should be a relationship among leadership, material wealth, and social resources. He measured these complex variables for all the household heads in Niwan Witz (using well-established methods) and tested his hypothesis using Pearson's *r*. Pearson correlations showed that, indeed, in Niwan Witz, leadership is strongly and positively related to material wealth and control of social resources.

Since the initial hypothesis was supported, Thomas used multiple regression to look at the relationship of leadership to *both* types of resources. He found that 56% of the variance in leadership was explained by just three variables in his survey: wealth (accounting for 46%), family size (accounting for 6%), and number of close friends (accounting for 4%). But, since multiple regression does not "specify the causal structure among the independent variables" (ibid.:132), Thomas turned to path analysis.

From prior literature, Thomas conceptualized the relationship among these three variables as shown in Figure 20.7. He felt that leadership was caused by all three of the independent variables he had tested, that family size influenced both wealth and the size of one's friendship network, and that wealth was a factor in determining the number of one's friends.

I won't discuss here the mechanics of determining the value of the path coefficients. A computer program like SPSS or SAS will take care of that for you. If you are interested in learning more about path analysis, consult Heise (1975). Suffice to say here that the *path coefficients* in Figure 20.7 are standardized values: They show the influence of the independent variables on the dependent variables in terms of standard deviations. The path coefficients in Figure 20.7, then, show that "a one standard deviation increase in wealth produces a .662 standard deviation increase in leadership; a one standard deviation increase in family size results in a .468 standard deviation increase in leadership; and so on" (Thomas, 1981:133).

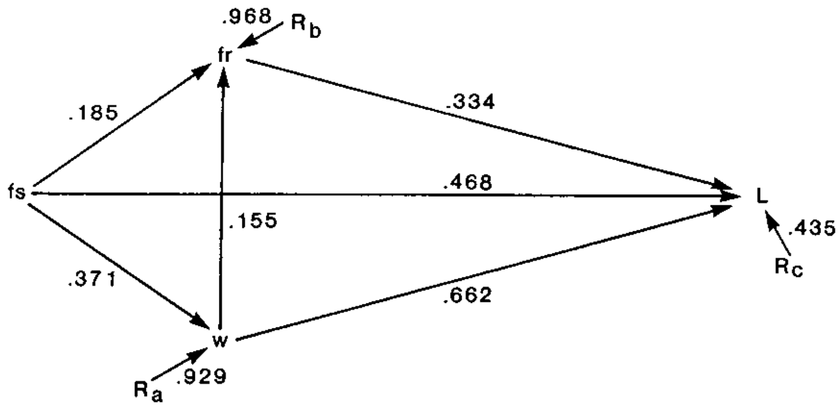

Figure 20.7. Path analysis of effects of wealth, friendship, and family size on leadership in Niwan Witz.

SOURCE: Thomas (1981); reproduced by permission of the American Anthropological Association from *American Ethnologist*, 8:1 (1981). Not for further reproduction.

Four things are clear from Figure 20.7. (a) Among the variables tested, wealth is the most important cause of leadership in individuals. (b) Family size has a moderate causal influence on wealth (making wealth a dependent, as well as an independent variable in this system). (c) The size of a person's friendship network is only weakly related to either family size or wealth. (d) The combined direct and indirect effects of family size, wealth, and friendship network on leadership account for 56% (1.435) of the variance in leadership scores for the household heads of Niwan Witz. Thomas concludes from this descriptive analysis that if one wants to become a leader in the Mayan village of Niwan Witz one needs wealth, and the best way to get that is to start by having a large family.

Path analysis is a tool for testing a particular theory about the relationships among a system of variables. Path analysis does *not* produce the theory; that's *your* job. In the case of Niwan Witz, for example, Thomas specified that he wanted his path analysis to test a particular model in which wealth causes leadership. The results were strong, leading Thomas to reject the null hypothesis that there really is no causal relation between wealth and leadership. Thomas noted, however, that despite the strength of the results, an alternative theory is plausible. It might be that leadership in individuals (wherever they get it from) causes them to get wealthy rather than the other way around. In fact, path analysis is often used to test which of several plausible theories is most powerful.

Conclusion

Once you have mastered the logic of multivariate analysis, you should seek out courses that take you more deeply into the use of these powerful tools. Most departments of sociology, psychology, education, and public health offer courses in multivariate analysis to their students. They often permit students from other disciplines to enroll. The examples used in those courses are usually not from research conducted by anthropologists. By now, however, you must have gathered that that doesn't make much difference.

All multivariate techniques require caution in their use. It is easy to be impressed with the elegance of multivariate analysis and to lose track of the theoretical issues that motivated your study in the first place. On the other hand, multivariate techniques are important aids to research, and I encourage you to experiment and learn to use them. Try out several of these techniques; learn to read the computer output they produce when used on your data.

And don't be afraid to play and have a good time. If you hang around social scientists who use complex statistical tools in their research, you'll hear people talk about "massaging" their data with this or that multivariate technique, of "teasing out signals" from their data, and of "separating the signals from the noise." These are not the sort of phrases used by people who are bored with what they're doing. Enjoy.

APPENDIX A

Table of Random Numbers

10097	32533	76520	13586	34673	54876	80959	09117	39292	74945
37542	04805	64894	74296	24805	24037	20636	10402	00822	91665
08422	68953	19645	09303	23209	02560	15953	34764	35080	33606
99019	02529	09376	70715	28311	31165	88676	74397	04436	27659
12807	99970	80157	36147	64032	36653	98951	16877	12171	76833
66065	74717	34072	76850	36697	36170	65813	39885	11199	29170
31060	10805	45571	82406	35303	42614	86799	07439	23403	09732
85269	77602	02051	65692	68665	74818	73053	85247	18623	88579
63573	32135	05325	47048	90553	57548	28468	28709	83491	25624
73796	45753	03529	64778	35808	34282	60935	20344	35273	88435
98520	17767	14905	68607	22109	40558	60970	93433	50500	73998
11805	05431	39808	27732	50725	68248	29405	24201	52775	67851
83452	99634	06288	98083	13746	70078	18475	40610	68711	77817
88685	40200	86507	58401	36766	67951	90364	76493	29609	11062
99594	67348	87517	64969	91826	08928	93785	61368	23478	34113
65481	17674	17468	50950	58047	76974	73039	57186	40218	16544
80124	35635	17727	08015	45318	22374	21115	78253	14385	53763
74350	99817	77402	77214	43236	00210	45521	64237	96286	02655
69916	26803	66252	29148	36936	87203	76621	13990	94400	56418
09893	20505	14225	68514	46427	56788	96297	78822	54382	14598
91499	14523	68479	27686	46162	83554	94750	89923	37089	20048
80336	94598	26940	36858	70297	34135	53140	33340	42050	82341
44104	81949	85157	47954	32979	26575	57600	40881	22222	06413
12550	73742	11100	02040	12860	74697	96644	89439	28707	25815
63606	49329	16505	34484	40219	52563	43651	77082	07207	31790
61196	90446	26457	47774	51924	33729	65394	59593	42582	60527
15474	45266	95270	79953	59367	83848	82396	10118	33211	59466
94557	28573	67897	54387	54622	44431	91190	42592	92927	45973
42481	16213	97344	08721	16868	48767	03071	12059	25701	46670
23523	78317	73208	89837	68935	91416	26252	29663	05522	82562
04493	52494	75246	33824	45862	51025	61962	79335	65337	12472
00549	97654	64051	88159	96119	63896	54692	82391	23287	29529
35963	15307	26898	09354	33351	35462	77974	50024	90103	39333
59808	08391	45427	26842	83609	49700	13021	24892	78565	20106
46058	85236	01390	92286	77281	44077	93910	83647	70617	42941
32179	00597	87379	25241	05567	07007	86743	17157	85394	11838
69234	61406	20117	45204	15956	60000	18743	92423	97118	96338
19565	41430	01758	75379	40419	21585	66674	36806	84962	85207
45155	14938	19476	07246	43667	94543	59047	90033	20826	69541
94864	31994	36168	10851	34888	81553	01540	35456	05014	51176
98086	24826	45240	28404	44999	08896	39094	73407	35441	31880
33185	16232	41941	50949	89435	48581	88695	41994	37548	73043
80951	00406	96382	70774	20151	23387	25016	25298	94624	61171
79752	49140	71961	28296	69861	02591	74852	20539	00387	59579
18633	32537	98145	06571	31010	24674	05455	61427	77938	91936

(continued)

APPENDIX A, Continued

74029	43902	77557	32270	97790	17119	52527	58021	80814	51748
54178	45611	80993	37143	05335	12969	56127	19255	36040	90324
11664	49883	52079	84827	59381	71539	09973	33440	88461	23356
48324	77928	31249	64710	02295	36870	32307	57546	15020	09994
69074	94138	87637	91976	35584	04401	10518	21615	01848	76938
90089	90249	62196	53754	61007	39513	71877	19088	94091	97084
70413	74646	24580	74929	94902	71143	01816	06557	74936	44506
17022	85475	76454	97145	31850	33650	75223	90607	15520	39823
24906	46977	78868	59973	61110	13047	84302	15982	72731	82300
50222	97585	15161	11327	66712	76500	81055	43716	93343	02797
60291	56491	75093	71017	92139	21562	67305	33066	60719	20033
31485	66220	71939	23182	44059	00289	17996	05268	97659	02611
16551	13457	83006	43096	71235	29381	93168	46668	30723	29437
90831	40282	48952	90899	87567	14411	31483	78232	52117	57484
19195	94881	99625	59598	33330	34405	45601	39005	65170	48419
06056	81764	46911	33370	35719	30207	61967	08086	40073	75215
46044	94342	04346	25157	73062	41921	82742	70481	83376	28856
03690	95581	83895	32069	94196	93097	97900	79905	79610	68639
23532	45828	02575	70187	64732	95799	20005	44543	08965	58907
81365	88745	79117	66599	32463	76925	70223	80849	48500	92536
57660	57584	14276	10166	82132	61861	63597	91025	76338	06878
13619	18065	33262	41774	33145	69671	14920	62061	42352	61546
07155	33924	34103	48785	28604	75023	46564	44875	07478	61678
19705	73768	44407	66609	00883	56229	50882	76601	50403	18003
04233	69951	33035	72878	61494	38754	63112	34005	82115	72073
79786	96081	42535	47848	84053	38522	55756	20382	67816	84693
76421	34950	98800	04822	57743	40616	73751	36521	34591	68549
28120	11330	46035	36097	93141	90483	83329	51529	94974	86242
45012	95348	64843	44570	26086	57925	52060	86496	44979	45833
45251	99242	98656	72488	35515	08968	46711	56846	29418	15329
97318	06337	19410	09936	28536	08458	90982	66566	30286	27797
55895	62683	25132	51771	70516	05063	69361	75727	48522	89141
80181	03112	21819	10421	35725	92004	36822	18679	51605	48064
39423	21649	18389	01344	36548	07702	85187	75037	89625	39524
37040	87608	46311	03712	42044	33852	52206	86204	99714	82241
72664	17872	02627	65809	17307	97355	60006	18166	51375	79461
71584	11935	87348	22204	93483	37555	31381	23640	31469	92988
87697	30854	25509	22665	31581	12507	53679	26381	48023	47916
73663	27869	40208	40672	83210	48573	22406	46286	46987	12017
51544	01914	17431	97024	09620	54225	44529	90758	11151	98314
82670	82296	96903	45286	85145	60329	27682	64892	75961	19800
30051	16942	17241	93593	75336	48698	48564	76832	29214	84972
23338	01489	39942	06609	14070	07351	28226	51996	31244	10725
08739	21034	57145	25526	58145	72334	87799	95132	70300	88277
76383	52236	07587	14161	82994	22829	72713	70265	88650	56335

(continued)

APPENDIX A, Continued

05933	81888	32534	56269	12889	05092	84159	40971	46430	86981
10347	07364	51963	31851	45463	41635	10195	18961	17515	34021
36102	55172	25170	81955	25621	25030	19781	48300	79319	34377
70791	56165	64310	28625	26760	82203	26535	99580	77676	91021
88525	67427	59554	42220	27202	18827	33362	90584	99516	72258
41221	71024	99746	77782	53452	52851	35104	20732	16072	72468
40771	10858	31707	46962	71427	85412	49561	93011	64079	38527
09913	14509	46399	82692	05526	19955	02385	85686	62040	39386
00420	06149	01688	72365	12603	83142	98814	66265	98583	93424
90748	19314	55032	64625	47855	32726	69744	54536	16494	33623

Reprinted from pp. 118-119 of *A Million Random Digits with 100,000 Normal Deviates,* by The Rand Corporation (New York: Free Press, 1955).

APPENDIX B

STATEMENT ON PROFESSIONAL AND ETHICAL RESPONSIBILTIES SOCIETY FOR APPLIED ANTHROPOLOGY

This statement is a guide to professional behavior for the members and fellows of the Society for Applied Anthropology. As members or fellows of the Society we shall act in ways that are consistent with the responsibilities stated below irrespective of the specific circumstances of our employment.

1. To the people we study we owe disclosure of our research goals, methods, and sponsorship. The participation of people in our research activities shall only be on a voluntary and informed basis. We shall provide a means throughout our research activities and in subsequent publications to maintain the confidentiality of those we study. The people we study must be made aware of the likely limits of confidentiality and must not be promised a greater degree of confidentiality than can be realistically expected under current legal circumstances in our respective nations. We shall, within the limits of our knowledge, disclose any significant risk to those we study that may result from our activities.

2. To the communities ultimately affected by our actions we owe respect for their dignity, integrity, and worth. We recognize that human survival is contingent upon the continued existence of a diversity of human communities, and guide our professional activities accordingly. We will avoid taking or recommending action on behalf of a sponsor which is harmful to the interests of a community.

3. To our social science colleagues we have the responsibility to not engage in actions that impede their reasonable professional activities. Among other things this means that, while respecting the needs, responsibilities, and legitimate proprietary interests of our sponsors we should not impede the flow of information about research outcomes and professional practice techniques. We shall accurately report the contributions of colleagues to our work. We shall not condone falsification or distortion by others. We should not prejudice communities or agencies against a colleague for reasons of personal gain.

4. To our students, interns, or trainees we owe nondiscriminatory access to our training services. We shall provide training which is

(continued)

APPENDIX B, Continued

informed, accurate, and relevant to the needs of the larger society. We recognize the need for continuing education so as to maintain our skill and knowledge at a high level. Our training should inform students as to their ethical responsibilities. Student contributions to our professional activities, including both research and publication, should be adequately recognized.

5. To our employers and other sponsors we owe accurate reporting of our qualifications and competent, efficient, and timely performance of the work we undertake for them. We shall establish a clear understanding with each employer or other sponsor as to the nature of our professional responsibilities. We shall report our research and other activities accurately. We have the obligation to attempt to prevent distortion or suppression of research results or policy recommendations by concerned agencies.

6. To society as a whole we owe the benefit of our special knowledge and skills in interpreting sociocultural systems. We should communicate our understanding of human life to the society at large.

Approved by SFAA, March 1983, superceding earlier published statements.

APPENDIX C

Codes From the Outline of Cultural Materials (Murdock, 1971)

000 MATERIAL NOT CATEGORIZED

10 ORIENTATION
- 101 Identification
- 102 Maps
- 103 Place Names
- 104 Glossary
- 105 Cultural Summary

11 BIBLIOGRAPHY
- 111 Sources Processed
- 112 Sources Consulted
- 113 Additional References
- 114 Comments
- 115 Informants
- 116 Texts
- 117 Field Data

12 METHODOLOGY
- 121 Theoretical Orientation
- 122 Practical Preparations
- 123 Observational Role
- 124 Interviewing
- 125 Tests and Schedules
- 126 Recording and Collecting
- 127 Historical Research
- 128 Organization and Analysis

13 GEOGRAPHY
- 131 Location
- 132 Climate
- 133 Topography and Geology
- 134 Soil
- 135 Mineral Resources
- 136 Fauna
- 137 Flora

14 HUMAN BIOLOGY
- 141 Anthropometry
- 142 Descriptive Somatology
- 143 Genetics
- 144 Racial Affinities
- 145 Ontogenetic Data
- 146 Nutrition
- 147 Physiological Data

15 BEHAVIOR PROCESSES AND PERSONALITY
- 151 Sensation and Perception
- 152 Drives and Emotions
- 153 Modification of Behavior
- 154 Adjustment Processes
- 155 Personality Development
- 156 Social Personality
- 157 Personality Traits
- 158 Personality Disorders
- 159 Life History Materials

16 DEMOGRAPHY
- 161 Population
- 162 Composition of Population
- 163 Birth Statistics
- 164 Morbidity
- 165 Mortality
- 166 Internal Migration
- 167 Immigration and Emigration
- 168 Population Policy

17 HISTORY AND CULTURE CHANGE
- 171 Distributional Evidence
- 172 Archeology
- 173 Traditional History
- 174 Historical Reconstruction
- 175 Recorded History
- 176 Innovation
- 177 Acculturation and Culture Contact
- 178 Sociocultural Trends

18 TOTAL CULTURE
- 181 Ethos
- 182 Function

(continued)