Atoll Population Structure: Some Methods of Analysis

Michael J. Levin

This chapter will develop methods of compiling and analyzing data collected by anthropological inquiry in small atoll populations. The techniques developed here can be used in populations where person specific information has been collected, but advantages and disadvantages of these techniques must be weighed in individual cases before the large amount of time invested in the hand coding is undertaken.

Since time in the field is becoming increasingly more expensive and consequently shorter and fewer trips can be taken, it is necessary to maximize time actually spent in the field for effective use of the data. Since demographic and genealogical data are often useful, if not essential, in anthropological analysis of much socio-cultural, economic and ecological phenomena, it is wise to collect quantitative information about each individual in the population. These data should include sex, parents, spouses and children, birthdate and place, marriage information, and information about death (see Carroll, Chapter ____). With this minimal, basic demographic information population parameters can be established (Feeney, Chapter ____), and the biological relationships among various members of the population can be determined (if similar

information for ancestors is also collected). Further, by collecting other kinds of data useful to the ethnographer - data collected on economic organization (e.g., distribution of land parcels, yields for different years, relationship of land owner to tenant) or political organization (e.g., relationships of succeeding chiefs, relationships of chiefs to other members of the population), correlations with the demographic information, or with other aspects of behavior in the population can be determined.

COLLECTION OF DATA IN THE FIELD

It is probably unwise to take preprinted census forms into the field unless it is impossible to have them sent in later, or unless a fairly good approximation of the kinds of data to be collected can be easily determined in the pre-field situation. However, if preprinted forms are taken with plenty of space on the forms for writing in of additional variables to be studied, then a preprinted form can be created. The form should be large enough to allow for collecting and coding of all important information, but not so large that it would be cumbersome to use in the field situation. Forms printed on 4 x 6 or 5 x 8 index cards would be best since these are thicker than paper, easy to handle and it is possible to buy filing boxes of appropriate size to hold them. Cards with perforated edges also are useful since punching holes in the cards makes possible some preliminary analysis in the field which might aid in encouraging the filling in of blanks in material to substatiate a

preliminary hypothesis. A disadvantage of these cards might be the expense of preprinting variable titles for information to be collected.

Most professional keypunchers can punch from any legible form, but they tend to demand that there be some systematization in the way the information is displayed. This is of use to the ethnographer as well since in doing analysis in the field deficiencies in the data can be spotted quickly, and this can only be done if particular pieces of information are always put in the same places. It is necessary to use a form that is easy to read,but still provides all the useful information necessary to collect the data. For the demographic information the suggested form in figure 1 is complete.

/ Insert Figure 1 here /

A few other points about forms can be made. If the anthropologist were in a field situation which allows for easy and reasonably rapid mail service with a methodologist at a university, it would be worthwhile to try to compile data from the field. It would be necessary to carry carbons, either as part of each card (this could be done if the preprinted forms were made professionally), or as pieces of carbon paper inserted between the cards which would permit sending one copy back to the methodologist for coding, compiling, and printing. The data could be printed in ways which would allow for more efficient use of field time in checking errors or omissions, and could allow for preliminary testing of hypotheses. Testing hypotheses in the field could also easily be done with the perforated cards, particularly if the punching were done with binary notation. In either case systematic

coding is essential for easy analysis.

PREPARING DATA FOR MACHINE ANALYSIS

The advantage of using the coding sheets or forms as I have described them is that the anthropologist can use the forms in the field, and can prepare the data for machine manipulation directly on the forms, thus eliminating a step of transferring all information to computer coding sheets.

To prepare the genealogical and demographic material for card punching, it is advantageous to put all forms in alphabetical order, and then to number all people sequentially in the deck. Of course, with later fieldwork, or perhaps with further research using fieldnotes more people will be found than are initially accounted for in the population originally tabulated, and these people will be added at the end of the data set and assigned new numbers. When the data set is considered complete (when no further fieldwork is contemplated), the anthropologist can have the computer repunch the deck with the numerical order the same as the alphabetical order. The advantage of this is that the numerical order and the alphabetical order will be the same; the disadvantage is that if the anthropologist had used , and become used to, a particular number standing for an individual, confusion results when 'new' and 'old' numbers must be considered.

When all individuals have been numbered, the anthropologist should then go through and number all fathers, mothers, spouses and adopters on the cards (and any other person specific information, e.g., tenants on land, who land was inherited from). This is done by referring to the indentifiaation number (ID) of the individual considered. For the first person's father's ID, it is necessary to find the father in the alphabetical listing and then record his number near his name on the first person's ID listing. Then the mother must be recorded in the same way, as must each spouse and the adopters. If, when recording the father and mother, the father's ID is noted next to his name as spouse to the mother, and if when the mother is coded, her number is noted next to her name as the father's spouse, time is saved later on when it will not be necessary to look up the mother as spouse to the father and the father as spouse to the mother. In fact, for siblings, it is only necessary to find the father's and mother's numbers once; after that it is only necessary to find the number of one or the other.

The numbering process is also useful for discovering liaisons or marriage which were missed either because the informants forgot them, or because the ethnographer unintentionally omitted them. These people inevitably turn up in this way, and can then be added to the spouse of the individuals involved.

It could be argued that this kind of tedious numbering of individuals could be done better by the the computer, but, unfortunately, this is not the case since the amount of similar, but not identical, spellings of names would cause many programming problems

problems. But, what is worse is the number of cases of several people with the same name, or alternate names for the same people, or occasionally a young person who is listed under a parent's name (i.e., Mary's son). Hence, by the time a program is written to take all exceptions into account (assuming such a program could be written), it is easier to code this information by hand. Also, by coding by hand, a greater facility in working with the data should develop, and a better knowledge of the relationship of the data to the computer; one of the most difficult problems in working with a computer programmer who does not anthropology is the necessity of explaining what is wanted. By having a thorough knowledge of the relationship of the data to the computer, many of the problems which would develop might be alleviated.

When all people have been numbered, codes should be set up for other kinds of information to be added. A code should be set up for all place names, and the same number should be assigned for a place whether it is a birthplace, place of death, or marriage. All of the current places should first be put on small pieces of paper and then put in alphabetical order; this close approximation of the alphabetical and numerical orders helps in numbering the place names on the coding forms. Similarly clan names, occupations, special statuses, land holdings, etc., can be numbered. These numbers must be put next to the proper attributes on the coding forms so that punching and compiling can be done. When all forms have been coded, the anthropologist can either hire a keypunch operator, or can keypunch the information to prepare it for the computer. The advantage of hiring a keypunch

operator is greater initial accuracy and faster preparation; if the anthropologist does the keypunching, remaining closer to the data for a longer time, a better understanding of the relationship of the data to the computer continues. Another anxiety many people have about the computer is the fear of the computer's doing violence to the culture by making all people numbers to be manipulated at will, and consequently, removing individuals from the cultural setting. Only by seeing how the data is prepared for computer analysis can these fears be partly put aside.

Whether the anthropologist or a keypunch operator punches the information, a simple program can decode the cards so that the data can be checked against the original coding forms. Although assistants can be hired to do this very tedious work (in fact, this is the most difficult step in the whole process, and is what probably causes many people to give up the idea of using the computer), the anthropologist should not have someone else do the checking. Only he really understands what the information should look like. This gives one last chance for making critical decisions about categories of information, or ascertaining for a final time that the right attribute was chosen, and of determining that keypunching errors have not been made. This checking should be done slowly, since it is very tiring and the efficiency of the person checking quickly deteriorates without frequent rests from the boredom of the computer print and the monotony of the displayed information. It is absolutely necessary that this checking be done since any incorrect information will remain incorrect for all compilations and displays. When the checking has been done, and cards have been corrected, displays can be made.

USES OF THE COMPUTER

There are four general areas of anthropological research that benefit from the use of the computer: lists, displays, statistical tests, and genealogical analyses.

Lists. When information has been punched on cards, a card sorter can be used to divide the punched deck. By sorting in appropriate columns, and then having a lister or a computer print the information from each card, the information can be listed in various forms. For example, if there are photographs of all residents in the population, and a printout of the information is needed for each photograph, after sorting in the columns of the computer cards which have been reserved for photograph number, a printout of the information can be obtained. Similarly any other information available by simple sorting can be made available. Most card sorters also give column totals; that is, as the cards are read through the card sorter, counters note which numbers are recorded in the column, and how many occurrences of each number there are. So, by sorting on the column for sex, the number of males and the number of females can be determined. All the males and all the females can then be listed for later reference.

Displays. In displaying information, we want to be able to look at compiled information for analysis and for testing hypotheses. Essentially, this is little different from sorting cards into piles and then counting the numbers of cards in each pile. The advantage of using the computer comes in its efficiency and its tirelessness. If a single variable, such as sex, is to be sorted, then the card sorter will be adequate since it is free and only one sort is

needed. If the computer is used to total single columns, the procedure is the same as for a card sorter or for the anthropologist sorting by hand. The computer sets aside one location for each attribute of a variable (e.g., for sex, males, females and unknowns), and then reads each card or card image to determine the recorded attribute of each individual in the population. Initially each assigned location is set to zero, and at the end of the sort, the location will have a value of anything from zero to a number of people in the population. The total number of people in the population can be determined by summing the values of the locations. If locations called MALE, FEMALE and UNKNOWN were used, the sum of the three values could be put in location TOTAL (and the value should be the total number of individuals), and printed:

THERE ARE 30 MALES, 30 FEMALES, 0 UNKNOWNS AND 60 TOTAL PEOPLE .

The importance of the computer in sorts, however, comes more in its ability to sort several variables at once. For example, in comparing sex with age, locations are assigned for males aged 0 to 9, females aged 0 to 9, males age 10 to 19, etc. This can be done more efficiently with a two-dimensional matrix (having a horizontal axis for age cohorts and a vertical axis for sexes), but the method of assigning locations and of incrementing values is the same. Matrices of three or more dimensions can be created to examine information for as many variables as may be useful for the hypotheses of the analyst and the number of people in the population. The advantage of the computer over the card sorter is its ability to print column and row headings and totals, as well as display the matrix in a number of different ways for easier analysis.

In displaying information the computer does not take over analysis from the anthropologist. It does not generate theory, nor does it construct tables in any ways other than those explicitly defined by the programmer. The machine cannot think; it cannot correct misconceptions which are programmed. This is another important reason for the anthropologist to do the programming, or to work closely with an anthropological methodologist who has proper training and interest to sometimes second guess the anthropologist.

Statistical tests. Most computing centers have standard programs (frequently called "canned" programs) which have been developed by professional programmers to run statistical tests on generalized data sets. Most demographic data can use these programs; most genealogical data has little use of them. To run the tests it is only necessary to contact consultants at the computer centers. It is wise to have a knowledge of elementary statistics before attempting to use these programs. Frequently, even complete anthropological data of the type described here does not really lend themselves to this analysis, and, if they do, often the results are not as useful as displays. Traditionally anthropologists have been interested in explaining why variation exists from a general pattern, not simply how much variation there is (and what the level of significance is). Although these measures are important, they tell us little about the culture, when taken without anthropological embellishment. One of the anxieties about some of the newer anthropological interests in using such measures as calories is the fear that statistics devoid of the people who make them up often do not provide a very full view

of a society. It is for these reasons that displays are usually more useful for anthropological analysis than statistical testing.

Genealogical analysis. Most of the hypotheses anthropologists can test with the computer are best done by using forms of displays. All of the demographic information can be handled in this way. However, in dealing with genealogical analysis, even in connection with other variables, a slightly different approach is needed. By carefully developing the methods used for this approach, the relationship between the data and the computer can be understood.

If a summary of marriage patterns for a population were wanted, traditionally many days would have had to be taken constructing the ancestors of each spouse, and, if the population were very inbred at all, a number of common ancestors would probably be omitted. For example, figure 2 demonstrates a spouse pair taken from an artificial population used to demonstrate computer techniques in anthropology (Levin,[illegible]).

/ Figure 2 about here /

The figure demonstrates how person number 27 is related to person number 26 through three generations of ancestors. Both individuals have numbers 7 and 8 as ancestors in the second ascending generation. Although person 7 is person 26's mother's father (MF), he is 27's father's father (FF). Therefore person 27 is related to person 26 as mother's father's son's daughter (MFSD). The two people are also related as MMSD,FFFDDD,FMFSDD,FFMDDD, and FFFSDD. All other

relationships are obtained through these; that is, the added relationships of number 7's parents on both sides are not unique. So, in computing inbreeding, or even in determining marriage patterns, these are usually ignored.

In the three generations considered here, there are fourteen possible relationships between as individual and the common ancestor: F,M,FF,FM,MF,MM,FFF,FFM,FMF,FMM,MFF,MFM,MMF, and MMM. These are all ascending terms, so if two people are to be related through a common ancestor, a descending list is also needed: C,C, SC,SC,DC,DC,SSC,SSC,SDC,SDC,DSC,DSC,DDC,and DDC. The C stands for child, since these people will be female (D) for male individuals and male (S) for female egos. Each term of the second list is the reverse (the descending term) of the element in the first list (the ascending list).

This fact makes it possible to relate individuals very rapidly by referring to the position in the array, rather than to the term. Two arrays, one for the individual and one for the spouse, are set up. The father's identification number is put in the first position of the individual's array and the mother's identification number is put in the second position. The father's father's and father's mother's identification numbers are put in positions four and five, and so on. Similarly, the array of the spouse's ancestors is also created. The two arrays for individuals 26 and 27 in the example would be:

Individual 26: 14 14 5 10 7 8 1 2 3 4 1 2 3 4

Individual 27: 15 18 7 8 9 6 1 2 3 4 3 4 1 2 .

Now each member of the spouse array is checked against each member of the ego array. The information stored in location 3 of the spouse array is the same as the information stored in position 5 of the ego array. The two arrays of ancestor terms indicate that position 5 in the array of the individual's ancestors is MF and position 3 in the spouse's ancestor array is SC. Since the individual is male, the relationship is MFSD. By continuing in this way, all biological relationships of the two spouses are found. There is one problem here. When the MFSD relationship is determined, the parents of the common ancestor (person 7) repeat the relationship (as MFFSSD and MFMSSD); these two relationships are not unique. In determining unique biological relationships, these should be omitted. This is done by checking to see if the child of each person who is found as a common ancestor is the same for both arrays (since 7 is the child in both arrays, these relationships would not be counted). Although this would be very tedious and time consuming to do by hand, simple computer programs have been developed to check for relationships among spouses as well as among any members of a population. Summary tables can also be displayed (cf Levin,ms.).

In order to do these tests, it is necessary pnly to have coded the father and mother of each individual (when these people are known), since successive ascending generations can be quickly assembled. The same computer program can also be used to determine the relationship between individuals and their adopters through both the father and mother, and any other coded personal data.

Also, in combination with demographic and other socio-cultural variables, relationships between specific genealogical ties and other types of data can be determined by coding for this information horizontally in a two-dimensional matrix; each individual will have one row.

Since each person usually has only one recorded mother (and rarely more than one father), these data can be coded in his row, However, it would be an enormous waste of space to try to record all of the spouses and appropriate children of a person within this row since some people will have no spouses or children, and others will have several spouses and many children. This problem is solved by creating another matrix for wives and children. In this matrix there is one row for each spouse pair. In each row are the identification numbers of the individual and his or her spouse, and each child with the oldest child first, followed by the next oldest, etc. Also, the coded information should contain the year of marriage, place of marriage, whether the marriage was sanctioned religiously or legally, if the marriage ended, cause (death or divorce) and year of termination of marriage, and perhaps the number in the sequence of marriages. Although there will be empty space in the matrix since it is easiest to construct the matrix row with enough spaces for all the children of the largest sibset, this form greatly simplifies programming. It is feasible to use a linear array to store all this information, but this is more cumbersome, and if new children are born, or when new marriages occur, the array must be redone. Finally, it is necessary to note where in this second matrix the spouses are located for an individual and how many spouses there are. This will take two locations in the first matrix for each individual. In this way it is very simple

to jump back and forth between the two matrices to do various tests.

ATOLL POPULATION: SPECIAL CONSIDERATIONS

In ethnographic fieldwork demographic and genealogical data collection remains the foundation for understanding other aspects of behavior at the person specific level. In using this method of data collection and manipulation, we can find the exact percentage of patrilocal residence or inter-matriclan marriages by coding the information (or by using information which can be constructed for the coded data) for each person. Cross-cultural comparisons can then be made at a level which was not previously possible when the subjective feelings of ethnographers were the measures available for this work. Although cross-cultural research remains very difficult due to differences in application of various anthropological theories and techniques, we can better control some of the variables which have caused problems for earlier anthropologists. Fortunately, atoll populations because of their relatively small size and relatively discrete area, are the prototype of the kind of population that can use the methods efficiently and reliably. Even more fortunate has been the great effort made by Carroll (see Chapter __ and Appendix 1) to bring anthropologists together for the Atoll Populations Conference, but even more, to define populations and field techniques necessary to elicit appropriate responses for this analysis.

The parameters for various populations as Carroll has defined them must be handled with great care to assure that the population that is used can be constructed with the appropriate methodology.

The total population is easily defined, methodologically, as all the listed individuals, and when summaries of the total population are desired, each row in the matrix will be checked for the appropriate information. To construct each of the other populations requires checking each row to determine whether the individual should be included in the test. For example, if analysis of the total ethnic population is to be done, it is necessary to check for each individual the column where ethnicity is recorded, to make certain that the individual should be included in the population. Similarly, if a de facto population for a census were to be used for analysis, each individual in the total population would be checked to see if he or she should be included in this smaller population. Also, it would be simple to construct a de facto ethnic population for a census by ascertaining the ethnicity and the residence at the time of census.

The advantage of using the computer comes not in doing these tests which might be done equally rapidly by hand (if they are only to be done once), but in constructing populations at different time periods, making diachronic analysis possible. Assuming no migration (or, if migration histories are recorded, with migration), it is possible to construct age and sex distributions for any defined population for any year. It is then possible to see yearly fluctuations as well as population trends. Cohorts can be followed through their reproductive years, and variations and trends in fertility, mortality, marriage patterns, migration, and adoption can be followed over time.

Household compositions and locations (plotted either relatively, i.e., household 1, household 2, etc., or spatially using coordinates) are extremely important in understanding the social, political and

economic organization of atolls. By recording the household residence for each census, and by recording changes as they occur, the importance of the demographic and genealogical information becomes emphasized in the total population structure.

Other geographic data can be incorporated into an individual's array to aid in determining interrelationships among people. On atolls where some people are tenants to others, the relationship between holder-tenant pairs may be useful. Also, distance from the household to the land holdings, when considered with the inheritance patterns permits analysis of variation from the "rule" and may demonstrate flexibility within a system. Relationships of size of matri-clan to land holdings can be analyzed with time depth, and, with the aid of a computer plotter, can be visualized also.