the cooperating states. In the southeastern region there are approximately 35 P-R and 30 D-J projects which have statistical sampling and analysis problems. A minimum amount of coordination among states in cooperation with the Institute of Statistics will permit evaluation of techniques on population and harvest appraisal methods of regional significance. For example, several states have farm fish pond sampling projects which are testing methods of appraising these populations. Here minimum coordination and similar statistical design will provide comparable data from which suggestions suitable for the entire region may be made.

The remaining half time will be devoted to basic research into statistical methods needed by all of our states in fish and wildlife research and management. Different types of model populations, and methods to sample them, may be constructed and then with state cooperation, field tested. To us, this is the heartwood of the program, and the area from which, in the long run, we can expect to realize the greatest returns.

The Institute of Statistics will undertake theoretical and methodological investigation of such basic problems as those dealing with response and nonresponse errors in mail surveys, sampling and estimation problems in field checking sportsmen, and population indices and their value in estimating population totals or changes in population levels. The latter includes sampling of fish ponds, reservoirs, and streams by electric seines, nets and chemicals, and sampling of wildlife populations by sight and auditory counts, tracks, pellets and other field signs. These are the indices upon which our recommendations for management are based.

We see a productive future for this program. It will take time to build understanding because, to a certain degree, the biologist must become statistician, and the statistician, biologist. Reasonable continuity has been provided in the Cooperative Statistical Project Design because application of quantitative methods of known precision to fish and wildlife problems will require both new statistics and new biology, neither of which is developed in a minute.

Efficient handling of suitable statistics in research is a mark of maturity in a profession. Perhaps this will become the outstanding contribution of the Cooperative Statistics Project. Thanks.

USE OF MACHINE METHODS IN PROCESSING FISHERY DATA

By VICTOR W. LAMBOU
Louisiana Wild Life and Fisheries Commission
Baton Rouge, Louisiana

INTRODUCTION

Most fishery management agencies are faced with the problem of an ever-increasing workload and a demand for more accurate and reliable information (Leeper, Stern, and Lambou, 1958). Research programs are being expanded and their standards are rapidly rising, necessitating more detail and complex analyses of data. Because of this, there is a need for: (1) machine manipulation of data and (2) improvement in the design and analyses of experiments. These are two distinct problems; however, they are related. This report will be primarily concerned with the need for machine manipulation of data.

ADVANTAGES OF MACHINE METHODS OF HANDLING DATA

Why are machine methods of handling data desirable? As previously stated, research programs are expanding and their standards are rising; while at the same time, there is a shortage of trained fishery personnel. Therefore, it is desirable that the efficiency of the individual fishery worker be increased. This can often be accomplished by using machine methods of handling data. Machines

can do most of the drudgery of routine calculating and computing and can accomplish all of the scientist routine work, allowing him to devote his attention to other important aspects of this job. Also machines, because of their tremendous speed, can solve problems which it would be impossible or impractical to do otherwise. Theoretically, any problem worked out by a machine can be worked out by the scientist manually, *i. e.*, if he is able to take the time, which is very often impossible. Also machines are much more accurate in handling masses of data than are humans.

One of the greatest advantages of mechanical manipulation of data is the relative ease with which exploratory studies can be made. When a tally of ten thousand data cards can be made in less than half an hour, studies of questionable value can be investigated without a feeling of great waste when that particular avenue proves fruitless, and this gives courage to investigate fields of uncertain potential.

SOME MACHINE METHODS

One of the most useful methods to increase the efficiency of processing data is to introduce data into punch cards for machine manipulation. After the data are entered on the cards the data can then be rapidly tabulated or classified in many different ways. The speed at which the data are processed, of course, depends upon the type of tabulating machine being used.

The fastest and probably overall the most efficient machine method of processing data involves the use of automatic electronic digital computing systems, sometimes called electronic brains. The data to be processed by these systems are normally first entered on punch cards, paper tape and/or magnetic tape.

LIMITATIONS OF MACHINE METHODS

It is important to remember that all machines used in tabulating data, even the most complex ones, do not think. They have to be told what to do, step by step. They are only servants which follow our exact directions, even though they are capable of tremendous speed. Moreover, if man, the master, directs these servants to solve a problem in an erroneous manner, they will without question oblige.

Sometimes, it is assumed that if we take data from a poorly designed experiment and/or which was erroneously or sloppily collected and subject this data to some highly refined method of tabulating (such as with punch card methods or with a digital computer) and then let a biostatistician apply some extremely complicated and mystical formula to the data, we will end up with some world shaking results. Of course, this is not true. It can not be over-emphasized that no matter how the data are tabulated (machine, desk, caculator or by hand) the tabulations can be no more accurate than the original observations. Erroneously or sloppily collected data will always be erroneous and inaccurate no matter how refined the methods of tabulating or analyzing the data are. Statistical results can never be better than the data they are based on.

BASIC PRINCIPLES OF USING PUNCH CARDS

The punch card itself is the basic instrument; its dimensions are maintained to very close physical tolerances. Because of these tolerances, careless handling, sorting and subjection to extremes of humidity may greatly hamper smooth machine operation. The Remington-Rand punch cards, which we use, have ninety columns, *i. e.*, positions for ninety digits of information. Each column contains six punching positions and combinations of these six possible positions give all basic numerical and alphabetical values. The punch card methods described here are for the Remington-Rand punch card system; however, these methods are suitable for other systems with minor adaptations.

The type of data collected and the design of the experiment will determine how the data will be programmed for machine tabulation. Under no circumstances should the opposite condition exist.

In order to insure a smooth and efficient operation, the punch card and the field form should be designed together. This is basic, but is very often overlooked. The data should appear in the same sequence on the punch card and the field report.

The most efficient arrangement of the punch cards and the source documents are not always easily determined, but a lot of study devoted to the symmetry and efficiency of field reports and card forms can yield great dividends when several thousand entries are being handled.

The data taken in the field should be arranged, if practical, in a series of blocks, in sequence either vertically or horizontally; and those items not coded in the field (such as body of water, parish, etc.) should be in a space to the left of the block provided for their codification. Whenever possible, codes should be included on the field form and decoded clearly so there will be no confusion.

The columns on the punch cards should be grouped into "Fields" of various sizes from one column up. They must be titled and must always be in the same position on all cards in the group. These fields have two general classifications: designative and accumulative. The designative fields will contain codes indicating descriptive data such as location, sex, etc.; and the accumulative fields will contain quantities whose totals have significance, such as weight, number, etc.

Field reports, or source documents, should be serially numbered and this number entered on the punch card in one of the designative fields. This makes it possible to compare at any time the data on the punch cards with the source document entries. Comparison is often necessary when correcting punching errors or when it is suspected some of the punch cards have been lost.

Any necessary editing, extending, or coding of data on the creel census form should be accomplished in its entirety before the forms are given to the punch operator for punching. The punching operation should be completely automatic, and the operator should not have any personal choice as to what should be punched in the cards. The moment the punch operator begins to think rather than automatically transcribe by touch, it tends to introduce more errors into the work and slows down production.

The punch operator transfers the data contained on the field forms to the punch cards through the use of a machine known as the automatic key punch. The operator by depressing keys on the key punch causes numerical and/or alphabetical values to be punched into the punch cards.

In most research, it is necessary that the punch cards be almost 100 percent accurate. In order to approach this accuracy, it is necessary to verify all data on the cards. There are many different ways of determining the accuracy of entries made on punch cards, most of which depend upon reading back the data either from the cards or printed tabulations. However, the most rapid method and the one most independent of human error is to have the data repunched into the same cards by another operator using the verifying attachment of the punch machine operative. When verifying the machine punches the holes in the card at a slightly lower position. The cards are then run through a verifying machine which checks all punching, and if any card does not have both the original punch and the verify punch hole in any position, the card is flagged as an error card.

When the cards have been punched and verified the resultant documents can be rapidly rearranged and counted. For small installations, most of the tabulations can be handled by a mechanical counting sorter which sorts cards at the rate of 420 cards per minute per column. When large volumes of data are being handled, other machines such as electronic sorters and printing tabulators are more efficient and rapid. With the counting sorter, only one column can be counted at a time; however, the use of a tabulator makes it possible to count and record one or more fields of information simultaneously. Also the tabulator prints the resulting values on a form, whereas with the sorter the values have to be manually copied.

The most helpful of all the auxiliary machines in punch card analyses are the electronic computers such as the Univac 120 Computer. The computer can be programmed to make any mathematical calculation in any sequence and proceed or change its procedure in a predetermined way as a result of any calculation. These computations are carried on at almost unbelievable speeds.

BASIC PUNCH CARD INSTALLATION

In the course of our work with punch cards, we have found that the basic installation needed is a punch machine and a counting sorter. As the volume of work increases and when funds become available, a tabulator, verifer and interpreter should be added to the set-up. Probably very few individual fishery research organizations will be able to justify much more equipment than those already mentioned. The next step would probably be to add a small electronic computer such as the Univac 120 Computer which processes data on punch cards. It is possible that groups of fishery research organizations could combine their resources and justify some type of large automatic electronic digital computing system such as a system built around the Univac Solid State Computer.

OWNERSHIP OF MACHINES

Undoubtedly the best ownership arrangement is for the research section to own the basic machines. In this way, the installation is directly under their control, and close supervision can be maintained over the personnel. Better scheduling of the workload and increased accuracy results from this arrangement.

Some research agencies are dependent upon their accounting sections for the use of machines. If this becomes necessary, it is far better than no machines at all. However, it is necessary to fit the work into their schedules; and therefore, considerable time lapses are involved between punching and the verification and computing of punched data. Also, if the punch operators are not closely supervised, there will be too many errors in the work since many of these operators feel that research work is just an added burden.

BORROWING THE USE OF MACHINES

In our operation, we find it necessary to borrow the use of machines from other State agencies. These machines include the Verifier, the Interpreter, and the Univac 120 Computer. Excellent cooperation has existed between our Commission and the Department of Highways and the Department of Labor in this matter. However, it is necessary to fit our work into their schedules.

USE OF SERVICE BUREAUS IN PROCESSING DATA

Most of the larger companies, such as Remington-Rand Corporation and IBM, have service bureaus which will process data. There are many small independent service bureaus, as well as some universities, which will process research data.

No hard-fast rules can be given as to the desirability of using Service Bureaus in the processing of data. I am of the opinion that it is desirable for the research organization to do as much of their own work as possible, especially the basic operation of punching, verifying and tabulating with the counting sorter and/or printing tabulator. This allows the organization to have more control over the operation and to maintain the desired accuracy in the work. Under some conditions, it might be desirable to farm out the basic operations to a service bureau, *e. g.*, when a backlog of work exists or when it is necessary to process an unusual volume of data. However, if work is to be farmed out, the research organization should be certain the service bureau can process the data in the desired manner and with the necessary accuracy. I know of several unfortunate instances in trying to have a service bureau process data. One of these instances involved punching a large backlog of data. Upon obtaining the punch cards, the research organization found so many errors in the cards that

it took them almost as long to correct the mistakes as it would have taken them to originally punch the cards.

Probably the most desirable use of service bureaus is contracting for the solving of complicated formulas involving large numbers of calculations (*e. g.*, multiple regression formulas involving many observations and variables) on large scale digital computers, such as the Univac Scientific and the IBM 704 Electronic Data Processing Machine.

Grosenbaugh (1958) describes a regression analysis program for the IBM 704 Electronic Data Processing Machine which will automatically handle up to 500 observations of Y along with observations of as many as 9 associated columns of X's. The computer, in about 11 minutes of operation, will furnish information concerning all possible least-square formulas which predict Y, using every possible linear combination of 9 or fewer variables. The program generates just about any information you would want to know about the regressions including which is the best formula for predicting Y and which X's do not appreciably improve the prediction. According to Grosenbaugh (*ibid.*) the cost of contracting for the use of this program with the Service Bureau Corporation, a subsidiary of IBM, lies between $50.00 and $250.00, depending upon the number of observations and variables and whether data sheets or punched cards are submitted. Of course, the above does not include the cost of developing the program. The magnitude of labor involved would prohibit the computation of such a regression analysis with desk calculators or with less efficient tabulating machines or computers.

EXAMPLES OF PUNCH CARD METHODS USED IN LOUISIANA

We have applied punch card methods primarily to the processing of creel census and fish sampling data. Our use of punch cards in processing creel census data are described in detail by Leeper, Stern, and Lambou (*op. cit.*), and Lambou (1959). Examples of the punch cards we have found useful in processing fishery research data are available from the author on request.

ACKNOWLEDGMENTS

In preparing this report, I used freely material contained in an earlier report by Leeper, Stern and Lambou (1958). Acknowledgments are made to the many employees of the Remington-Rand Corporation and Louisiana Wild Life and Fisheries Commission who helped with the development of the punch card procedures used by the Louisiana Wild Life and Fisheries Commission. This report is a contribution of Louisiana Federal Aid in Fish Restoration Project F-1-R.

LITERATURE CITED

Grosenbaugh, L. R. 1958. The elusive formula of best fit: a comprehensive new machine program. Occasional paper 158, Southern Forest Experiment Station, Forest Service, U. S. Department of Agriculture.

Lambou, Victor W. 1959. Simplified creel census program for punch card tabulation. Louisiana Wild Life and Fisheries Commission, Dingell-Johnson Project F-1-R, mimeo.

Leeper, Ben, Herbert Stern, Jr. and Victor W. Lambou. 1958. Some uses of punch card methods in the tabulating and analyses of fishery research data. Proceedings of the Eleventh Annual Conference, S. E. Association of Game and Fish Commissioners (1957).

EDITOR'S NOTE: Copies of two other papers—"Simplified Creel Census Program for Punch Card Tabulation" and "A Contribution Toward a Bibliography on Use of Machine Methods in Processing Fishery Data"—are available from Mr. Lambou on request.