can then be formed by a series of table derivation instructions of types (*b*) and (*c*).

It may be asked why computations of this type are not included in standard form as part of the general program. We have indeed incorporated the computation of sampling errors in a program which deals with stratified random samples, but to deal in a general manner with all the complexities that arise in the many, often complicated, types of sampling that are met with in practice would unduly complicate the program, and would make it too unwieldy for a machine the size of the 401. Moreover, we have found that, in practice, evaluation of sampling errors is not very frequently required, except when considering the appropriate size of sample in future surveys, and the efficiency of different sampling methods; in the latter case the sampling errors of different sampling methods have to be evaluated. In the interpretation of the results of a single survey an adequate assessment of the sampling errors can usually be made from internal evidence of the tabular material (its regularity, or lack of regularity, over secondary classifications).

An Autocode Form of the Instructions

Douglas and Mitchell (1960) have recently described a program for the analysis of surveys of the market research type, which was written for Pegasus. In it they have adopted an autocode form of language. A somewhat similar language has been adopted in the scheme outlined in *Sampling Methods for Censuses and Surveys*, and could be used in the 401 program: all that would be required would be a preliminary interpretive routine which would translate the autocode instructions into the coded instructions of the 401 program. Storage space is not a major problem, since the translation has to be performed before the actual analysis is begun. The writing of such interpretive routines is, however, a lengthy and involved task, and it was decided that the more direct instruction code adopted in our program would be adequate for our needs.

References

Douglas, A. S., and Mitchell, A. J. (1960). "AUTOSTAT: a Language for Statistical Data Processing," *The Computer Journal*, Vol. 3, p. 61.

Yates, F. (1960). *Sampling Methods for Censuses and Surveys* (3rd edn.), London: Griffin.

Market Surveys with a Small Computer

By R. L. Cook

This paper, which was presented at the Harrogate Conference of The British Computer Society on 5 July 1960, describes a general-purpose Market Survey program for use on an 803 computer which has been developed from a specification jointly prepared by Social Surveys (Gallup Poll) Limited and Elliott Brothers (London) Limited.

Introduction

The 803 is a medium-speed transistorized computer with a magnetic-core store of 4,096 words of 39 bits. Input is by paper tape or punched card. The Elliott card reader is used, and is arranged to read each column both as a 12-bit number, when reading cards punched in the manner described in this paper, and as a 6-bit character for use with data punched in a conventional punched-card code.

The requirements of the market survey program are as follows:

1. To read data from punched cards without requiring the user to conform to any particular punching convention.
2. To produce a set of double-entry frequency tables from the input data.
3. To produce results in a form directly available for reproduction.
4. To retain simplicity of programming and operation, but also to allow sufficient flexibility to deal with the widest variety of problems possible.
5. To provide a facility for checking the data cards for consistency.

Data Recording

The results of a market survey are assumed to be punched on cards. More than one card may be required to hold the results of one interview, and the word *Block* is used here to mean the one or more cards which contain the answers to one interview.

Questions posed in a survey are of two types: the question may require a single integer answer, or may request the interviewee to indicate which of one or more attributes on a selected list he possesses. An example of the first type is "how old are you," and of the second type "to which of the following age groups do you

belong." The answer to the first question will be punched in a 2-column field, whilst, for the second question, particular digit positions on a card will represent the chosen age groups, and a single hole will be punched in the position corresponding to the answer obtained from the interviewee. In the example given the question was exclusive and can only provide a single answer. More complex questions of the type "which of the following washing powders have you in the house at present?" may produce any number of holes punched in the data card, one hole corresponding to each positive reply.

One card will generally contain many unrelated facts about the interviewee, and the word *Domain* is used to describe the assembly of digit positions which contain a particular piece of information. A domain may vary in size from a single digit position corresponding to one Yes/No fact, to the whole of a card representing 80 × 12 = 960 bits of information. Our investigation showed that three types of domain would cover all requirements. These are:

(*a*) Any set of digit positions belonging to not more than 3 different columns of the same card. For example, positions 1 and 4 of column 6, the whole of column 20, and position *Y* of column 80.

(*b*) The whole of any number of consecutive columns of the same card. For example, columns 8–18 inclusive.

(*c*) Any number of consecutive columns of a card in which is punched a single integer according to the standard punched-card conventions, i.e. the normal punched-card "field."

Domains of types (*a*) and (*b*) may contain any number of punched holes, whilst domain type (*c*) must not have more than one hole punched in each column. A domain need not correspond to a particular question on the interview form, and each digit position of a card may belong to any number of domains.

Program

The basic purpose of the 803 program may now be explained. It is the construction of a set of double-entry frequency tables for each of which the row and column titles correspond to the digit positions of two defining domains.

Three separate programs have been developed and will be briefly described. For each program the requirements of the particular survey analysis are fed into the 803 by means of a few control cards: this information controls the working of the program and is printed out by the computer as a set of near-English statements. This makes a permanent record of the analysis performed, and checks that the control cards have been correctly punched.

Scrutiny

The input-data cards may be checked for consistency before being used to form the frequency tables. It may, for example, be important to check that the rules for completion of the interview forms have been complied with. The interview form may require a certain question to be asked only of particular people, and completion by the wrong class of person may produce erroneous results.

The scrutiny program checks the validity of statements concerning the number of holes in a domain. Three types of statement may be made:

not more than n holes
exactly n holes
at least n holes

where n is an integer.

These statements can be joined into a complex sentence by means of the words "If," "then," "and," "or." A typical sentence may read "*If* there are exactly 2 holes in domain A_1 *and* at least 3 holes in domain A_2 *then* there are at least 2 holes in domain A_3 *or* exactly 0 holes in domain A_3."

To avoid the difficulties of the associative and commutative rules of logic, the words "and," "or," may not both occur in the same half of a scrutiny sentence. Such requirements can always be catered for by a succession of simpler sentences.

The scrutiny program prints out the serial number of the card and of each sentence that is not true.

Accumulation

The accumulation program forms the set of frequency tables. The row and column titles of each table are each defined by a domain. A complete frequency breakdown including row and column totals will be formed, each block of cards contributing either 0 or 1 to every cell position of each table.

A weighted table may also be formed in which the contribution is an amount which is either punched directly on the block of cards or is obtained, *via* a look-up table, from an attribute of each interviewee. For example, if males and females are to be separately weighted, the sex of each person may be punched on the card and the weight derived, or the weight itself may be punched.

Each cell total occupies one word of 803 storage, and 3,000 words are available for storage of tables in any one run through the data cards.

A block of cards with a given attribute may be completely omitted from a table, or from the row and column totals, or from the breakdown cell totals. If, for example, a table referring to men only is required, as well as a table involving both sexes, then the first table may be formed subject to the condition "omit females."

More complex conditions may be imposed on the formation of a table, causing the program to add holes to the card pattern held in the store, and the imaginary columns 81–90 are used for this purpose. If, for example, a breakdown were required showing readers of "*Times* or *Financial Times*" as well as the separate readership, the sentence "If *Times* or *Financial Times* punch Y in column 81" is used, and digit position Y of column 81 is added to the domain specifying the table breakdown.

Output

The frequency tables formed in the 803 can be coded on tape for subsequent output, or directly printed. The output program provides many calculation facilities, including grouping and weighting of rows and columns: it is, therefore, often convenient to have more analysis columns in a table than it is expected will be required, and to group them at the output stage. The program includes facilities for page layout and descriptive row and column titles of tables.

The principal output facilities provided are as follows:

(*a*) Percentage or actual frequencies may be printed.

(*b*) Rows and columns may be printed in any order, grouped, weighted, or omitted altogether.

(*c*) Rows and columns may be interchanged.

Conclusion

The programs described in this paper have been developed during the last six months: some of the complex editing facilities of the accumulation program are not yet complete. The programs have successfully catered for the analysis of several surveys, although in one case a modification was made to the output program to print special results. The programs have been written to provide general-purpose facilities, with the result that they are somewhat slower than tailor-made programs written for a particular survey. Our experience clearly indicates that the cost of the additional machine time is more than offset by the programming effort that would be required to write special-purpose programs, except possibly in the case of very large surveys which are in continuous use.

Market Research Applications on LEO

By J. A. Gosden

This paper was presented at the Harrogate Conference of The British Computer Society on 5 July 1960.

Introduction

The Market Research work carried out by LEO Computers Ltd. has, to a very large extent, been carried out for the Market Research section of their parent company, J. Lyons and Company Ltd., who are engaged in marketing a wide range of consumer products. Three basic applications have been carried out: first an application producing concomitants; second, multiple-regression analysis; and third, an application producing cross-tabulations of answers to questionnaires. The first two are straightforward uses of standard techniques, but the third is a special program designed to meet the needs of the particular types of survey carried out.

Concomitants

The concomitants application takes the answers to questionnaires and seeks out the dominant attitudes of consumers with respect to their eating habits, as affected by such factors as the food itself or the environment in which it is consumed.

The questionnaires may be completed by interviewers in the field, or by observation and interview during experiments in tasting laboratories. The number of questions answered varies between 20 and 45, and the sample sizes vary between 100 and 1,000.

In this application the answer to any question can only be made from a choice of two. For example, sex is either male or female. Other questions are framed so that there are only two possible answers; for example, "Are you over or under 35 years of age?" Wherever possible, questions are framed so that the number of replies of each type is nearly equal.

The procedure to compute the dominant attitudes is a series of steps. In the first step the answers are counted and the correlation between each pair of answers is calculated. In the second step the sets of answers of a small part of the sample are searched for the most frequent patterns of inter-correlated answers. From these totals it is possible to reduce the range of answers which need to be considered in the last step. The last step assembles the correlations between the remaining more highly inter-correlated answers into a matrix, and then extracts the principal latent vectors. These vectors define the dominant attitudes of the sample. An alternative procedure, when there are few questions, is to take the most frequent patterns discovered in step 2 and count their occurrences in the full sample, providing a profile analysis of the sample.

Regression

The second application is the straightforward calculation of multiple regression coefficients. This program is used to monitor the relationships between sales of various commodities and variables which influence sales. It is in two parts: the first calculates the regression constants from actual sales, and the second uses these coefficients to estimate future sales. These assist in the compiling of production plans both for day-to-day and long-term