COMPUTER APPLICATIONS IN MARKETING RESEARCH:

A PROPOSAL

212-66

David B. Montgomery

COMPUTER APPLICATIONS IN MARKETING RESEARCH:

A PROPOSAL

David B. Montgomery*

The purpose of this note is to outline the need for a clearing house for computer programs of use in marketing research and to propose a viable mechanism to meet this need.

THE PROBLEM

The 1960's have seen an accelerated interest in empirical research in marketing. One need only peruse the relatively new journals in the field -- the <u>Journal of Marketing Research</u> and the <u>Journal of Advertising Research</u> -- to perceive this trend. The advent of the third generation computer with its greatly enlarged high speed storage as well as time sharing capability will no doubt give further impetus to this trend, as will the hoped for development of a marketing data bank and clearing house.[1]

For purposes of discussion it would seem that computer needs in marketing research may be usefully categorized as follows:

 I. Data Reduction Systems

   A. Cleaning and Editing Systems

   B. Retrieval and Summary Systems

 II. Statistical Systems

   A. Univariate and Multivariate Analysis

   B. Estimation Systems for Models

 III. Special Purpose Languages

These categories will be considered below in terms of the need for general availability of programs, current examples, and future potential.

Data Reduction Systems

The examples cited in this section will tend to be drawn from experience with panel data files. The basic notions and systems needs are, however, applicable to any large, disaggregated data file.

Any researcher who has sought to use a large, disaggregated raw data file is well aware of what a horrendous task it is to prepare these data for further analysis. The development of systems to prepare the data involves considerable manpower and financial costs to say nothing of psychological costs in the frustrating task of debugging a system. No doubt these problems have limited the use made of such data files in the past.

The first step in processing a raw data file is to clean and edit the data. In spite of the careful and rigorous standards of quality control maintained by the MRCA and the Chicago Tribune panels, it is important that the data be cleaned and edited prior to analysis. By cleaning and editing is meant screening of the data to make certain that the data fields contain information in valid codes, gross keypunching errors haven't occurred,[2] and that the data are in appropriate,

packed[3] format for efficient processing later.
Professors Ronald E. Frank and William F. Massy [3]
have developed a cleaning and editing package for
MRCA National Consumer Panel and Chicago Tribune Panel
data. The present version is available in FORTRAN II
with certain FAP subroutines. As this system is trans-
lated into FORTRAN IV and the assembly languages
appropriate to the third generation computers, it is
important that an effective mechanism be developed for
making this new package widely available to the
marketing profession.

Large, disaggregated data files are rarely used
as is. Generally, the analyst must first retrieve and
summarize information from the files prior to carrying
out his intended analysis. For example, if his purpose
is to estimate a stochastic model of brand choice, he
must first retrieve the sequence of brands chosen by
each household. In any case, whether the ultimate
analysis is the estimation of a stochastic model or the
development of a multivariate model, he will need to
develop a programming system which will transform the
cleaned and edited data into a format appropriate for
statistical analysis. If the models and statistical
procedures which are developing in marketing are to
receive wider use within the profession, it would seem
imperative that systems for retrieval and summary of

data be developed and made available. Such systems
would reduce the technological, time, and cost barriers
to progress on this front. As such they would free the
analyst to give greater consideration to the empirical
viability and relevance of the various models and
research techniques.

Once again the Frank-Massy panel data system
provides an example of a system for generating a
flexible set of summary measures. Their Household
Purchasing Characteristics Generation System and Time
Series Generation System provide a wide range of summary
data for analytic purposes. There is a need in the
future for more such systems which generate measures
not currently available in the Frank-Massy system.
In any case, there is a need for a mechanism to call
these developments to the attention of the marketing
profession in an efficient manner. At present, most
of this information is disseminated by word of mouth.

Statistical Routines

There is also a need within the profession for
efficient and wide ranging interchanging of information
regarding available statistical routines. In parti-
cular, one can think of three classes of statistical
applications where this interchange would be most
useful.

Class I. Enhanced Efficiency or Generality of Standard
        Procedures

It would be useful to have a clearing house for
information on improved programs for using standard
statistical procedures. For example, the basic multi-
variate analysis programs by Cooley and Lohnes [2]
have been chained and generalized by Professor Charles
King of Purdue and also by Kenneth Jones of Brandeis
[6]. They have extended the utility and efficiency of
the basic programs by expanding the size of the
problems which the system can handle, providing new
features, and chaining the sub-systems for more
efficient processing. It is of clear interest to the
marketing profession to have available the most powerful
and efficient systems for data analysis.

Class II. New Analytic Procedures

As new methods for analyzing data are developed
and found to be relevant in marketing applications,
there will be a need for exchange of information on
available programs. The utility of this interchange
will be especially clear to anyone who has ever been
involved in writing and debugging a program for multi-
variate analysis.

One can foresee certain of these procedures on the
horizon. For example, Claycamp and Massy [1] have
proposed the use of cluster analysis[4] in market segmenta-
tion. They report that work is underway on an empirical

application which will result in an operational cluster analysis program which undoubtedly would be of interest to other researchers.

Another area where new analytic techniques and programs will be likely to develop is in the area of multi-dimensional scaling. The development of multi-dimensional scaling procedures and operational programs for their implementation may be expected to render psychometric procedures more useful and relevant in marketing research in the future. Once again, there is a clear need for a mechanism to disseminate to the marketing profession information regarding computer systems which can carry out such analyses.

Class III. Estimation Systems for Models

One aspect of the increasing interest in empirical research in marketing has been the development of data based models. A wider use of these modeling techniques in the future would be facilitated by the dissemination of systems for estimating the various models. For example, minimum chi square procedures have been proposed for estimating and testing several recent stochastic models of consumer brand choice. See Morrison [9], Massy [7], and Montgomery [8]. It would be useful to have the estimation programs for these models be made available.

Special Purpose Languages

While such general purpose, problem oriented languages as FORTRAN and ALGOL are readily available to a broad spectrum of market researchers, there are certain other useful special purpose languages which are not, perhaps, so visible. For example, there are IPL-V, the list processing language useful in programming decision flow models, and SIMSCRIPT, a general purpose simulation language.

Once again, the advent of the third generation computer with its capacity for on-line interaction between man and machine promises to witness a veritable explosion of new processing languages to aid the researcher in utilizing the enormous capacity of these new computers. A notable beginning has already begun in this direction with the development of the OPS-3 system for on-line computation and simulation by Greenberger, Jones, Morris, and Ness [4] at M. I. T.

For the marketing profession to continue its accelerating development in the area of empirical research and model building -- including simulation, algorithmic, and stochastic models -- it will be necessary for researchers in the field to take advantage of technological advances in the state of the research art. A positive contribution to the rapid and efficient dissemination of information regarding the relevant

dimensions of these technological advances in the computer area could be made by a clearing house for information on these developments.

Summary

In sum, it has been suggested that a very useful contribution to research in marketing could be made by a clearing house for information on data reduction programs, statistical packages, and special purpose computer languages. The contributions would be greater efficiency in empirical research, a broader awareness within the profession of the tools available, and more rapid diffusion of new techniques and models of use in empirical research.

A PROPOSED SOLUTION

Now that the case for a forum for the exchange of information regarding computer programs for data reduction and statistical analysis and special purpose languages has been developed, what can be done to meet this need? The author would like to propose that the Journal of Marketing Research establish a new section entitled something like "Computer Applications in Marketing Research" in order to provide the profession with a forum for this vital exchange. A successful model for this undertaking is the "Computers in Behavioral Science" section of Behavioral Science. Since 1960, Behavioral

Science has provided behavioral scientists with abstracts of computer programs of relevant to the behavioral sciences as well as short article length discussions of major systems and models.

Why suggest the JMR as the vehicle for meeting this need? First, in its relatively short history, the JMR has become a major mechanism for the exchange of research results and procedures among those interested in advancing the scientific approach to marketing. Hence, it has as its audience those who would benefit most from the proposed section. Second, it has the broadest spectrum of interest in marketing problems among the journals in the field. Finally, it would seem that a journal having the editorial and audience characteristics of the JMR would prove to be the most efficient and effective means for a broader and more rapid exchange of information in this area.

Format

The format which such a section would take might be as follows:

    A. Abstracts of available computer programs

        The dimensions of the abstracts are discussed below.

    B. Short article length papers

        1. Input/Output descriptions for data reduction systems which are too extensive to fit in the Abstract

section.

2. On rare occasions it may be useful to
present a short article description of
less well known statistical procedures
and the program available to perform
the computations. In addition, it may
be helpful to provide notes on problems
and pitfalls in the application of
certain programs.[5]

3. Short discussions of the major innovative
features and useful functions of special
purpose languages. These descriptions
should provide enough information to the
researcher to enable him to determine
whether or not it would be worthwhile
for him to explore the use of the language
further.

Abstracts of Computer Programs

In order for a computer program abstract to be
published in this proposed section of the Journal of
Marketing Research there should first of all be a
reasonable likelihood that there will be general
interest in it among researchers. Beyond this the
individual submitting the abstract should also provide
a concrete example of the documentation which will be
made available. Such documentation should include an

expanded version of the abstract material, references to statistical and mathematical sources, flow charts, control card formats and definitions, and a listing of the source program. The format used in the chapters of Cooley and Lohnes [2] would be a useful model.

Editorial responsibility would principally be to screen submitted programs for general relevance and for evidence of adequate documentation. In no way can or should the JMR accept responsibility for the accuracy of the programs. Its function is solely one of disseminating relevant information in an efficient manner. It would bear no responsibility for supplying the programs to prospective users or for maintaining the systems. The responsibility for supplying documentation and program decks must rest with the researcher submitting the program abstract.

The items which should appear in the abstracts along with some clarifying comments are presented below:

A. Program Title

B. Name of Individual(s) Submitting the Abstract

C. Description

A brief description of the program in terms of inputs, computations, and outputs. It may be necessary to present summary comments on any mathematics here.

D. Capacity and/or Limitations

For example, limits on the number of variables or the sample size.

E. Running Time

May be heuristic or a summary of experience on representative problems.

F. Machine and/or Language Specifications

What language is the system written in? For what machine? What systems idiosyncracies will affect the transfer of the source deck between computation centers?

G. Availability

What is available? Source deck? Object deck? Listing? Where and from whom? Is there any charge? If so, how much? Does the source wish to serve as the focal point for program maintenance?

CONCLUSION

This note has outlined a need for a forum for the exchange of information regarding computer applica-

cations in marketing research and has proposed what
would prove to be an efficient, effective, and viable
method to meet this need. The establishment of such
a vehicle should also serve as a considerable incentive
for researchers in the field to document their systems
well and to make them more generally available to other
interested researchers. It is the hope of the present
author that this proposal may serve as a catalyst
which will soon result in such a forum coming into
being.

In conclusion, perhaps it would be well to note
that it would be possible to broaden the scope of the
proposed section to include abstracts of data which
are available for academic research, giving information
as to the source of such data, cost, restrictions,
format, etc. Such abstracts could serve as valuable
adjuncts to a marketing data bank or perhaps even
serve as a surrogate for a data bank.

FOOTNOTES

*The author is Assistant Professor of Management in the Alfred P. Sloan School of Management, Massachusetts Institute of Technology.

[1] See [5].

[2] For example, the price per unit should be reasonable for the product involved. Thus a price of $5.00 per pound of coffee would clearly be in error, and the screening and editing program would reject such a data record.

[3] By packing is meant the procedure of storing several individual records in one logical block on a magnetic tape. Since large scale computers are generally input/output bound when dealing with large data files, considerable computer time savings may be realized by using packed records for input/output purposes.

[4] The growing interest in this technique is exemplified in the fact that Professors Ronald Frank and Paul Green have also proposed an application of cluster analysis in marketing.

[5] For example, the author and several others have had some difficulty with a widely circulated multiple discriminant program which sometimes yields large negative eigenvalues from a positive semi-definite matrix, a result which is theoretically impossible.

REFERENCES

1. H. J. Claycamp and W. F. Massy, "A New Approach to Market Segmentation", Working Paper 104, Graduate School of Business, Stanford University, 1966.

2. W. W. Cooley and P. R. Lohnes, Multivariate Procedures for the Behavioral Sciences, New York: John Wiley & Sons, 1962.

3. R. E. Frank and W. F. Massy, Computer Programs for the Analysis of Consumer Panel Data, Stanford, California: Graduate School of Business, Stanford University, August 1965.

4. M. Greenberger, M. M. Jones, J. H. Morris, Jr., and D. N. Ness, On-Line Computation and Simulation: The OPS-3 System, Cambridge, Massachusetts: The M. I. T. Press, 1965.

5. C. L. Hinckle and J. S. Wolf, "Academic Research and the Data Drought Dilemma", Journal of Marketing Research, 3 (May 1966), pp. 196-8.

6. K. J. Jones, "The Multivariate Statistical Analyzer", Multilithed Operations Manual Available from the Harvard Coop, 1964.

7. W. F. Massy, "Estimation of Parameters for Linear Learning Models", Working Paper No. 78, Graduate School of Business, Stanford University, October 1965.

8. D. B. Montgomery, "A Probability Diffusion Model of Dynamic Market Behavior", Working Paper 205-66, Alfred P. Sloan School of Management, M. I. T., May 1966.

9. D. G. Morrison, "Stochastic Models for Time Series with Applications in Marketing", Technical Report No. 8, Program in Operations Research, Stanford University, August 1965.