N. MIHĂIȚĂ
Department of Economic Cybernetics
The Academy of Economic Studies

Sample Processing as a Marketing Experiment

The difference between experiment and observation is given by the degree of the researcher's intervention, as well as by the design, planning and construction of the experimental situation.

The study presents a modality of tackling marketing sampling as a statistical experiment, an argumentation based on the creation and implementation by the author within the Laboratories of the Department of Economic Cybernetics of an automatic processing system of data from samples. The creation of the experimental situations requires the manipulation of conditions in which surveys can be made to measure their influence on the results of the sample.

The analysis by experiments suggested in the study, made on a reduced scale and for a limited length of time has as an objective-function a multiple data processing, the economy of computer time and of effort.

The results listed by the afore-mentioned system are data useful to the segmentation techniques, to the multiple discriminant analysis χ^2 as well as in the application of experimental techniques as latin squares, factorial experiments, simulation, etc.

The obtaining of information from the samples designed and implemented as marketing experiments starts with what is established by the information structures of the questionnaires, it is continued with what is socially established and with what is selected by the experimenter according to the aim, and one's imagination and creation and one returns to the beginning of this cycle. Thus, statistical guarantees are obtained for the decisions on the working out of the marketing strategies.

The market research for ensuring the sales and the entire production requires a wide-scale use of the marketing problems linked with the exami-

nation of the market characteristics and with the processing of the resulting data on the basis of modern computing techniques.

As part of the marketing researches, the fastest and adequate data are provided by the application of the method of survey and samples. By the sampling method detailed and differentiated individual data are obtained and at the same time, the representative value is preserved necessary for the conclusions and significance of global statistics [6].

The differentiations appearing among surveys (the aim, number and nature of characteristics, their form of expression, the types of questions, the amplitude of the variation of numerical variables) require the generalization of certain processing systems. These systems must process any type of questionnaire used in research by sampling, make explicit the contingency tables resulting from the grouping orders, select certain parts from the researched collectivities and ensure accessibility to processing regardless of the degree of training in the field of informatics [3].

The extension of the use of surveys as efficient investigation instruments in various fields simultaneously with the improved means of electronic data processing require the reconsideration of the working methodologies in all phases of design and implementation of such research, especially in those regarding the systematization of the obtained material and the necessity of increasing the degree of processing of the collected data, the use of complex statistical-mathematical methods, etc.

Segmentation, closely linked with the retrieval and analysis of the multiple discriminant, a statistical technique very often used in the processing of data from samples requires the listing of a number of contingency tables containing simple or multiple groups resulting from series of dichotomies. The "operating" processing systems achieve these tables by recording and by successive passages through the information file created by sampling. Therefore, a type of control is necessary which should result in the automatic recording and listing on segmentation characteristics.

Just like the regression, segmentation is seeking a relation of dependence between a group of independent variables and a dependent variable, but unlike the regression it does not seek a formula, a function, but a maximum precision in knowing the value assumed by the dependent variable in the event of the existence of values given to all independent variables or the precision with which one can provide the set of possible values (the elements of the independent vector) for a stable value of the dependent variable.

From the systemic point of view, one can check properties of observation-control, accessibility and sensibility by using multi-dimensional analysis techniques in market sampling. Such values of modern experiment provided by the general systems theory and cybernetics require a change in the processing methods in keeping with the extremely complex requirements of these sciences.

In connection with the design of experiments in the conception of the systems theory, Thomas H. Naylor, a specialist in computer simulation, points to the necessity of synthetizing the studied system in simple relations. At the same time, synthesis means "to identify the major variables affecting the answer", after which the most important variables are associated with the answers, with a view to seeking the possible empiric relations [10]. If

the models are global, more simple applications will be used to identify the fields in which more detailed models are required. "As soon as an understanding is acquired", the author says "the formulation of general laws and theories becomes possible" ... These laws and theories can or cannot be applied to similar systems.

The sample planned as an experiment allows for an integrated processing of information collected from various sources and with different techniques, a fact which ensures their best turning to account in the analysis process. We outline the common elements of the samples for the analysis of the elements contained in various experiments.

The field of manifestation of the economic phenomena is created by various statistical populations of the type of the set of inhabitants of various countries — regions, customers, the set of counties in a region, the set of economic transactions of a group of products, etc.

The set of the examined characteristics defining the state of a system is made up of characteristics specific of dependent variables (endogeneous/explained/criteria) and independent (exogeneous/explicative/predicative). Processing as a system represents a set of operations of a synthesis type on data about the examined statistical population starting from individual information (answers to questions, options, primary calculus), coded, provided by the units involved in research. This processing means the calculus of differentiated frequencies according to membership to the characteristics studied on alternatives. Thus, an important common element of research is noticed, the fact that processings have the same content.

Of interest after the construction of the contingency tables resulting from the processing of the entire information file is the degree of homogeneity of the surveyed population or subpopulation and the significant differentiations appearing between the segments represented by the intersected alternatives of the characteristics.

Thus, one can immediately make an option for the calculus of the values χ^2 on lines and columns of the matrices associated with the tables—for testing the significance of the differentiations or for the calculus of the values of the information energy index — with a view to studying the degree of homogeneity.

The aim of processing is that to "compel" the object of observation (data collection) to reveal certain laws, causal relationships between the examined phenomena. More than the analysis and synthesis of data collected from various primary and secondary sources, the investigator wishes anticipations of phenomena in the event of changes, influences, interventions, and experimenting of anticipations.

The pre-eminently statical-observation-character of a sample is surmounted by treating the entire process of obtaining the information as a planned experiment and the processing of the collected data is directed towards the exploitation of statistical methods of experiment programming.

The advantage of using the experiment now, in the 20th century technical and scientific revolution is revealed by the transformations which have taken place in the very structure and logic of its evolution. The experiment taken over from physics, and chemistry, a forerunner of systemic research "wide spread also today" gradually surveys the results of the variation of

an input variable into the system while the other ones are constant [7]. This way of approach is isomorphous with that of processing data resulting from samples. The information sorted out according to criteria of membership to the essential characteristics comprised in the processing model become partitions of the selected variables. Univocal causal dependences are established between the dependent and independent variables, just like in the case of the classical experiment. For instance, the correlation is sought between the dependent variables which have the character of consequence, such as: attitudes, motivations, tastes, preferences, the independent values and variables which have the character of influence: sex, age, socio-professional category, training, income, geographical localization, stage in the living cycle of the family etc.

Dependence which is checked with classical and rigorously calculated statistical techniques suffers because of the conditions of excessive artificialization in processing even in the conditions of an outsized sample and of the existence of a powerful automatic data processing [8], because priority is given to the operation of separation, disassociation and isolation; in this way, the interdependence, intersection relation is not comprised in model.

The speciality literature provides solutions for a remedy of the absence of statistical conditions of experiment design [4]. Among these solutions, we mention: the replacement of direct manipulation in the experiment by comparing groups subject to contrasting experiments or the establishment of the time order in which the variables are manifested;

— the random assignment of subjects to the "control" and "experimental" groups by carrying out a cross sample of the population and by combining statical with dynamic analysis; in general, the sample automatically requires the observance of this condition;

— the control over the exterior variables, measurements made by repeated surveys for testing the "start" hypotheses, for seeking similar circumstances in various time and space moments or the use for comparisons of analyses of dynamic series from which the same relations in test-situations should result as those noticed in comparative statical analyses (survey of family budgets, panels, samples, etc.).

For example the samples treated as experiments, by using the information basis created by marketing surveys work with the samples created by alternative division at experimental and control groups. Thus:

a) within the study on the population's supply with electronic household devices and the trends of their marketing, the followings were used:

— a way of sampling which fulfils the requirements of a random sampling, thus it can calculate the probabilities according to which the established differences are significant;

— the "experimental" group made up of the electronic device owners who intend to replace them and the control group, the potential buyers who intend to get such a device;

— the independence variable — the supply with devices, the control group being "on the first supply";

b) within the surveys as regards food consumption:
— customers were questioned in the order of coming out from the shop — a selection method — mechanical step;
— the "experimental" group, the buyers of groups of products;
— the control group, the nonbuyers of some groups of products;
c) within the total surveys on the market of certain chemical products:
— the observation unit — the importing country;
— the experimental group — the countries with a low import rate;
— the control group — the countries with a high import rate.

To check the influence of several factors simultaneously implied in the survey, processings are necessary for obtaining groups on parts of the surveyed collectivity or multiple groups both on collectivities and subcollectivities. It is the case of the groups created by the variables taken in the case of experiments of latin squares type and factorial experiments.

The design of experiments known as latin squares is used when it is necessary to examine the variation of the effects created at various levels by certain variables whereas others remain at constant levels. By this experimental design, two or several independent variables can be controlled simultaneously.

If the observation unit is the consummer of a product or of a homogeneous group of products, the significant differences between the options on different age categories, socio-professional categories, groups of income by using the latin squares, can be examined simultaneously. In this case, the lines of the latin square will be filled with the variants of the age characteristic, the columns — with the socio-professional variants and "the treatments" — with the income groups. The single combinations are selected from a mass of information structured and segmented according to the above characteristics and the variation on lines and columns between treatments is established by applying statistical techniques-test F.

In this processing, it is necessary to order the list of the collectivity examined according to three characteristics simultaneously on a number of n_1, n_2, n_3 groups, where n_i is the number of variants of characteristic i. These groups may appear in the form: for each variable of the first characteristic a contingency table with associated matrices of dimensions (n_2, n_3) (figure 1). The matrices are obtained by series of multiple groupings on the whole surveyed collectivity or on certain parts of the collectivity.

Figure 1 presents the way in which a latin square containing 16 elements of 64 possible (single combinations) is filled in, in four stages (dimension 4 for the latin square).

The (frequent) groupings c_{ij} are obtained within contingency tables listed according to the variants of characteristic 1 — c_{1j} (where $j = 1, 2, 3, 4$), the subject including characteristic 2 and the predicate characteristic 3. The 16 combinations are singular (the alternatives of the characteristics will intersect only once within the table).

The conclusions in connection with the significant differences between the options on age, socio-professional and income categories are also valid for the 15 other latin squares possible to construct on the basis of the matrices associated with the contingency tables. If the observation unit is the family

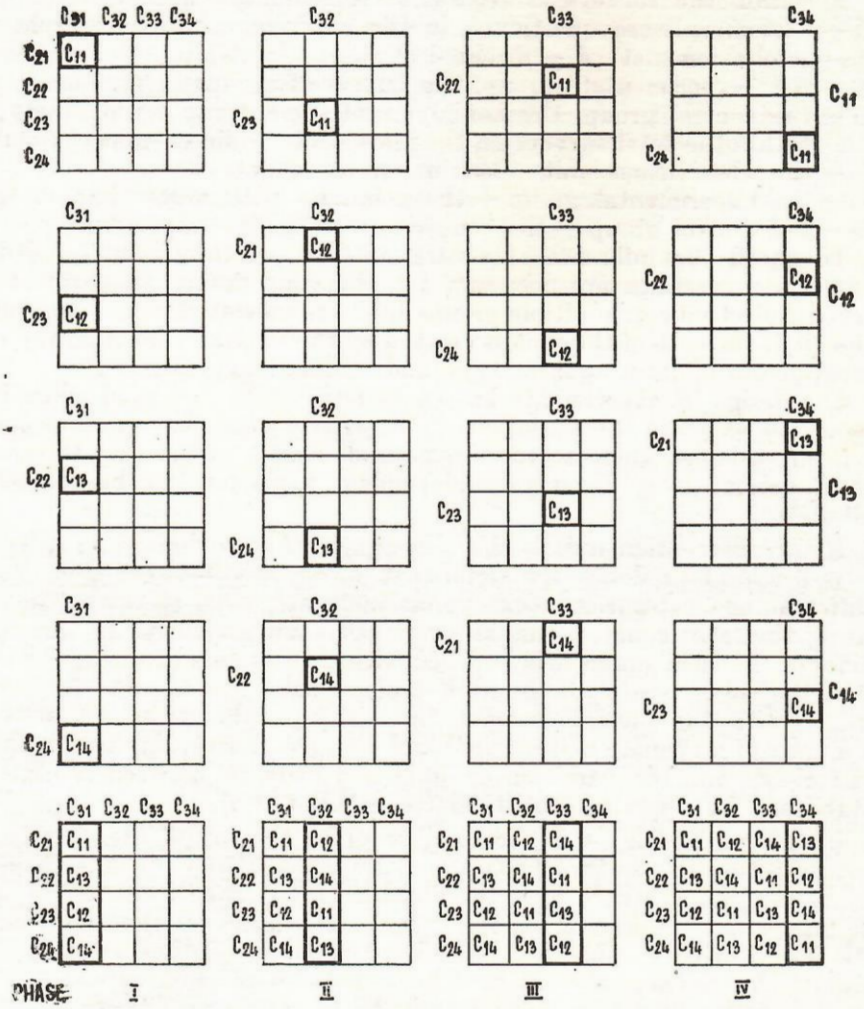


Fig. 1

or house-hold latin square can be constructed which should take into account the place of survey (alternatives), the size of the assigned reference (number of inhabitants, industrial production, other statistical indices) and other independent characteristics which will direct the processings on the created file of data.

Factorial design envisages the carrying out of observations within all possible combinations of experimental factors. The simplest application is given by the construction of a contingency table containing a dichotomized variable in the subject of the table and a dichotomized one in the predicate; this implies the existence of a factorial experiment of type 2^2, thus two variables of two levels each, resulting four combinations of the effects of the

factors. The general form of type m^n signifies the existence of n different factors, each at m levels.

The factorial designs allow the examination of the possible interactions among the variables of interest.

In order to establish the relations presented above, it was necessary to build a system of multiple processing of data from samples, extensible on fields of marketing of total survey type (office). By using orders of multiple groups, the system was used as the **Modules** for the checkings required by the application or verifying of known segmenting techniques, such as: the Belson segmenting method, the AID method (Automatic Interaction Detector-Morgan-Sonquist), the Multiple Classification Analysis, or of the Morgan-Staelin methods (AID-MCA), of those based on the analysis χ^2, multiple discriminant, the canonical analyses or the information theory [4] as well as in the case of the application of experimental techniques, such as: latin squares, factorial experiments.

Being catalogued within the Laboratories of the Economic Cybernetics Department of the Bucharest Academy of Economic Studies, the system of automatic processing of data resulting from samplings was cheked in applications oriented towards the diagnosis and prognosis of markets of industrial goods, chemical and electronic products, of light industry, an important economy of time being obtained in comparison with other systems in the preparation and processing of data, the clear explanation of results and new order facilities of the contingency tables.

The system proposed lists the contingency tables in absolute figures (frequency type groupings) the calculus of the shares of groups expressed in percentage in keeping with the characteristic in the subject of the contingency table or with that in its predicate, the calculus of the shares of groupings in comparison with the total amount in the table, the calculus of the information energy index and the calculus of indes χ^2 of checking the homogeneity of data [9].

The system is used in the simultaneous processing of several samplings on the same magnetic tape (fixed formats in defining the records), of "omnibus" type surveys, by carrying out separate processings on parts of collectivities or on the whole file of data. In keeping with the coded answers and questions, the matrices associated with the tables in percentage may serve in certain cases as a Markovian type matrix, useful to simulations in marketing. The multiple groupings, the segments of collectivities tested as homogeneous may become parameters or a calculus basis for the economic forecasting models.

The system's diagram of exploitation of the file of data specific of the marketing samplings is presented in figure 2. The information deriving either from closed questions (alternatives of answer known previously and precoded) or free are coded in the growing order of natural numbers and together with the number of variants (alternatives of answer) create a distribution of form $\binom{i}{n_i}$ where i is the order number of each characteristic and n_i is the number of variants of each characteristic. If we denote by l_i and L_i the lower and respectively upper limit of the variation interval of interest for esta-

blishing the contingency tables on subcollectivities, we can explain in table 1 the exploiting possibilities of the analysed system.

Mention should be made of the main characteristic of the system, namely the economy of time. The "operating" systems perform, at present, only the order of type *C* unless successive recording and sorting are carried out. Figure 2 underlines some of the possible achievements by using specific segmenting techniques, factorial experiments, latin and Greek-Latin squares, contingency tables comprising multiple groupings, etc.

This arsenal unused so far in the exploitation of data from samplings except for the contingency tables and segmentations deepens the analysis of the relations established by the combinations of characteristics from samplings. The fact that such a method of "promoting" data in a file created on the basis of a sampling or of an office survey was not used so far is mainly

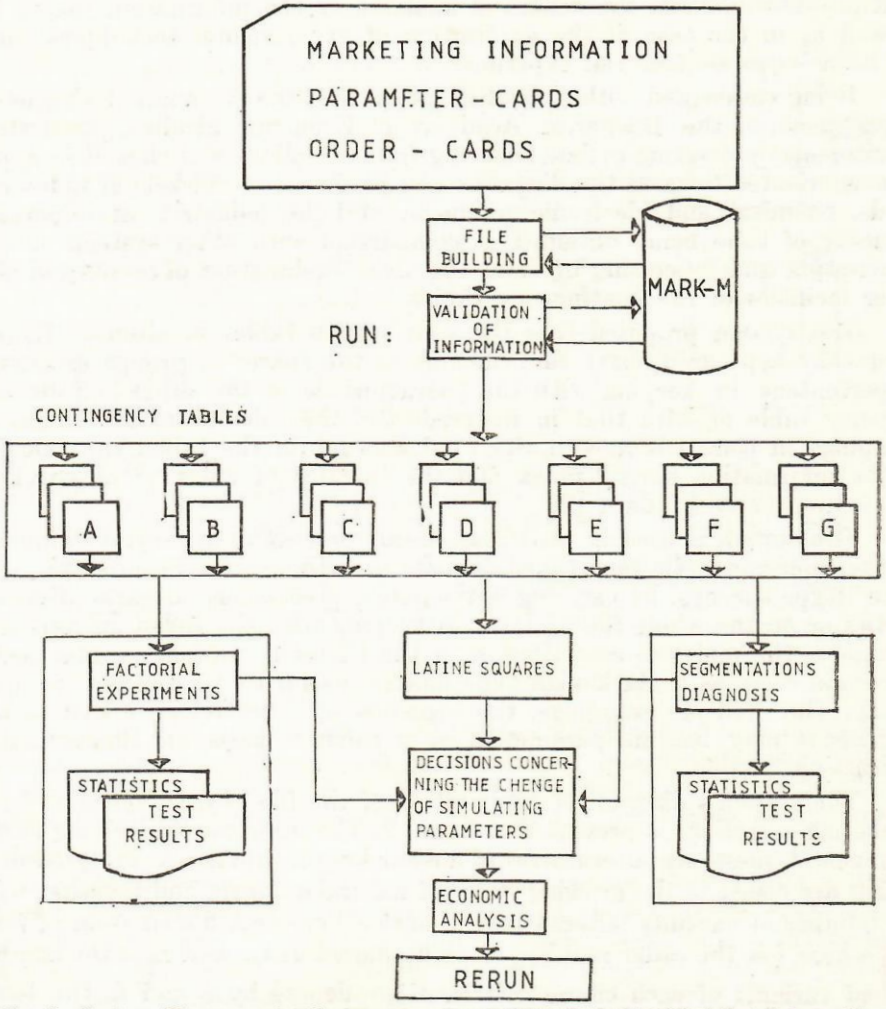


Fig. 2. System Diagram of File Working Possibilities Including Markting Information

due to the absence of a rapid processing system which should lead to reasonable processing costs.

The segmenting orders lead to the listing and construction of reports necessary for the analysis or check ups through the segmenting techniques described above. All types of orders contained in table 1 can be used for segmenting.

Table 1

Significance and action of the orders given to the multiple processing system

No.	Code Type Operation	No. characteristics in the subject of table	The segmenting limits are specified	ACTION
1	A_2	1	yes	Dichotomic segmenting
2	A_n	1	yes	Segments by combining the elementary segments
3	B	b	yes	Segments connected by effects of dichotomic interaction
4	C	c	no	Several contingency tables with the same predicate
5	D	d	no	Simultaneous options for several characteristics with the same predicate
6	E	e	yes	On subcollectivities-several tables with the same predicate
7	F	f	yes	Multiple segmenting with chaining of all alternatives of the characteristics
8	G	g	yes	Segmentations of "F" type with sets of chained alternatives

The "listed" results of the action of the orders given to the system to process the file of coded data are presented in table 2.

Tabel 2

No.	Order Code	RESULTS	Number of tables
1	A_2	Simple groupings	1
2	A_n	Simple regroupings with several alternatives	1
3	B	Multiple groupings — series of dichotomies	2^{b-1}
4	C	Simple groups for the entire collectivity	c
5	D	Simple groups on binary coded alternatives	1
6	E	Simple groupings on parts of collectivity	e
7	F	Series of multiple groupings on the whole collectivity	$\prod_{i=1}^{f-1} n_i$
8	G	Series on multiple groupings on subcollectivities	$\prod_{i=1}^{g-1} (L_i - l_i + 1)$

Both the systematic surveys, the constant research and polls among the users of industrial products imply considerable funds. These funds and expenditures are justified by the obvious increase of factors of socio-economic development, following the adaptation of production, in due time, to the requirements of some "segments" of consumers of industrial goods, in the conditions of the optimal solution of the problem of stocks, the cut in expenses, the waste of energy and in the conditions of the constant rise in the living standards.

The processing of data on the market of indutrial units is necessary both to the planning and control activities and for a precise estimation of the position of a whole branch, both nationally and internationally.

REFERENCES

1. * * * *Directivele Congresului al XII-lea al P.C.R., cu privire la dezvoltarea economico-socială a României în cincinalul 1981—1985 şi orientările de perspectivă pînă în 1990. (The Directives of the Twelfth Congress of the Romanian Communist Party on the Socio-Economic Development of Romania in the 1981—1985 Five-year Plan the Long-Term Guide-Lines Until 1990).* Editura politică, 1979.
2. NICOLAE CEAUŞESCU: *Raport prezentat la cel de al XII-lea Congres al Partidului Comunist Român. (Report Presented at the Twelfth Congress of the Romanian Communist Party).* Editura politică, 1979.
3. BOGDAN V., LUPU D., MIHĂIŢĂ N.: *Tendinţe actuale în prelucrarea informaţiilor din sondajele comerciale. (Topical Trends in Processing Data from Trade Samplings).* Comerţul modern, no. 4, 1977, p. 26—29.
4. BOSS J. F.: *Metode de segmentare în marketing intern şi internaţional. (Segmenting Methods in Home and International Marketing).* In: Marketing intern şi internaţional, Editura politică, Bucureşti, 1976, pag. 239—262.
5. DEMETRESCU M. C.: *Metode cantitative în marketing. (Quantitative Methods in Marketing).* Editura ştiinţifică, Bucharest, 1971, p. 215.
6. DEMETRESCU M. C.: *Marketing.* Editura politică, Bucharest, 1973, p. 21.
7. MIHAIL R.: *Introducere în strategia experimentării cu aplicaţii în tehnologia chimică. (Introduction to the Experimenting Strategy with Applications in Chemical Technology).* Editura ştiinţifică şi enciclopdică, Bucharest, 1976, p. 9.
8. MIHĂIŢĂ N.: *Sistem de prelucrare automată a informaţiilor privind sondajele de piaţă. (System of Automatic Processing of Data on Market Samplings),* OPTIMAL, no. 2, 1979, p. 46.
9. MIHĂIŢĂ N.: *Working out Multi-Processing System for Sample Information Necessary to Simulate Marketing Decisions.* Paper presented at the 6th Conference on probability Theory. Braşov, September, 1979.
10. NAYLOR TH. H.: *Computer Simulation Experiments with Models of Economic System.* John Wiley and Sons, New York, 1971, p. 180.
11. * * * *Marketing.* Editura didactică şi pedagogică. Bucharest, 1975, p. 130.

TRAITEMENT DU SONDAGE COMME EXPERIMENT DE MARKETING

L'auteur présente une modalité d'approche du sondage de marketing comme expériment statistique, son argumentation ayant pour base l'élaboration d'un système de traitement multiple des informations receuillies. Les résultats des traitements selon ce système deviennent utiles aux techniques de segmentation, aux analyses X^2, discriminent multiple, ainsi que pour l'application des techniques expérimentales telles que les carrés latins, expériments factoriels, simulation, etc.

ОБРАБОТКА РЕЗУЛЬТАТОВ ВЫБОРОЧНОГО ОБСЛЕДОВАНИЯ В КАЧЕСТВЕ ЭКСПЕРИМЕНТА В МАРКЕТИНГЕ

Работа содержит определённый способ подхода к маркетинговому выборочному обследованию в качестве статистического эксперимента, на основе разработанной автором системы многократной обработки данных из анкет. Результаты, полученные при использовании данной системы являются полезными для сегментации, для анализа χ^2, многократного дискриминанта, а также для случая применения экспериментальных способов как латинских квадрат, факториальных экспериментов, симуляционных методов и т. д.

25. 05. 1980