

THE STATISTICAL DESIGN AND ANALYSIS OF AN EXPERIMENT TO MEASURE THE
EFFECTIVENESS AND COSTS OF A HEALTH AND WELFARE PROGRAM

Julius A. Jahn, University of Pennsylvania

As a part of a research and demonstration project concerned with ~~Special Ser-~~ vices for the Aging, in the Community Service Society of New York, an experiment was planned to answer questions concerning the effectiveness and costs of some alternative programs. These questions were concerned with the relative effectiveness of several alternative service programs that could be experimentally tested, and also, with ways in which the effectiveness of these programs could be increased.¹

The major steps or phases involved in this experiment were as follows:

- I. Definition of the Problems for Experimentation.
- II. Formulation of Hypotheses to be tested by Experimentation.
- III. The Statistical, Experimental Tests of the Hypothesis.
- IV. Derivation of Hypotheses from the Experiments.
- V. Application of Hypotheses that have been Experimentally Tested or Derived.

I. DEFINITION OF THE PROBLEMS
FOR EXPERIMENTATION

A. THE GIVEN CONDITIONS

Definition of the Population. The effectiveness and costs of programs of "Services for the Aging" were measured for a population consisting of persons 60 years of age or over in two major sub-populations or strata. The first, referred to as the "Applicant Stratum," included persons in the population for whom applications were made for service to either the Special Services for Aging Office or in the Central Office of the Community Service Society. The other stratum, referred to as the "Non-Applicant Stratum," includes all other persons in the population for whom such applications were not received. The term "Participant" was used to refer to any member of this population. In the "Applicant Stratum" there were four criteria

¹ Footnotes are appended to end of this paper.

ia defined for identifying "Participants:"

- (1) Persons who are 60 years of age or over at the time of application.
- (2) Persons who are making an inquiry or request (or for whom an inquiry or request is made) for service to the Special Services for Aging Office, either for themselves or for another person, and whether in-person, by telephone or by letter.
- (3) Persons whose permanent place of residence at the time of application is not outside Manhattan, Bronx or Queens Borough of New York City, or is not in a "congregate" residence for aged persons; and, who do have a residence within these three boroughs (except for applicants for "Ward Manor").
- (4) Persons who are not included in cases currently opened for service in a Community Service Society district office or cases which have been closed within the current month in the Special Services for Aging office.

In the "Non-Applicant Stratum" only two of these criteria were applied: (1) The ages of persons as of the date they were first listed for sampling during any given month of the period for study, and (2) place of residence.

Definition of Alternative Programs. In this paper only certain distinctive conditions involved in the alternative programs will be indicated, as more complete definitions are provided elsewhere.²

Program 0, the "Basic" program, included those health and welfare services that would be available to persons in the population, excluding those in the two alternative programs.

Program 1, "Short Term Service" included services by either a caseworker or a public health nurse within the Special Services for the Aging unit of the Community Service Society. The service was to be completed within two months from the date of case opening and was not to require more than four in-person interviews

Program 2, "Collaborative Service," included the possibility of services by

either or both caseworker and public health nurse within the Special Services for the Aging unit. No restriction on the time or number of in-person interviews was given.

Measures of Effectiveness of Programs.

Of the many possible criterion variables that might have been selected, the final choices involved the use of both personal judgments and statistical methods of analysis. Two major steps were involved in making the selection. The first stage involved selecting questions to be included in the research schedule. For purposes of both interviewing and analysis, the various questions were classified into seven "content areas," identified as: (1) Housing, (2) Occupation, (3) Financial, (4) Physical and Mental Health, (5) Personal Adjustment, (6) Interpersonal Relations, and (7) Recreational or Social Activities and Interests.

The selection of questions for inclusion in the research schedule was primarily based on the judgments of the research staff. These judgments were made with reference to three sources: First, the particular goals and policies of the Community Service Society with respect to providing services for aging persons. Second, the kinds of problems indicated by the requests for service by persons applying to the Community Service Society. And, third, research and professional literature on problems of aging and programs for aging, and reports of public meetings of professional and lay persons especially concerned with problems of aging persons. The specific questions were selected to provide objective definitions of problems of aging persons which seemed to be explicitly referred to in the various statement found in these sources.

The second stage was carried out in order to reduce the number of criterion variables from the approximately 600 included in the ~~final~~ research schedule to a relatively small number to use for measuring effectiveness of alternative service programs. The second stage was carried out using procedures to avoid the obvious error of selecting those criterion variables for which the differences were in the "right direction."

In this study it has not been possible to arrive at the selection of one particular criterion variable or group of cri-

terion variables which were acceptable to all persons involved. Instead, four different groups of criterion variables were selected by somewhat different conditions and methods:³

C.V. Group 1. Adaptive and Adjustive Status. This is a modification of scales developed and used in previous studies by the Institute of Welfare Research of the Community Service Society.⁴ In the present study, the ratings were based on the direct observation, interviews and judgment of research interviewers, and not on previously recorded service records. The ratings were made and recorded at the time of an initial research interview and again at the time of the follow-up interview six months later.⁵

In order to simplify analysis, "The Overall Level of Adjustment and Adaptation," was selected as a criterion variable in this group. The ratings were recorded using number 1 through 11 to indicate ratings from "Very Poor" to "Excellent." For uniformity in analysis the order of the numerical ratings has been reversed so that 11 represents "Very Poor," and 1 represents "Excellent." This rating was selected as a criterion variable in this experiment because the "Movement Scale," from which it was derived, has been previously used in research studies to measure effectiveness of casework services in the Community Service Society and other social agencies.

C.V. Group 2. The Basic Problems of Aging. The criterion variables in this group were selected by personal judgments of the research staff that these variables would be accepted as problems by most of the persons expected to support programs dealing with "problems of aging." The variables are coded so that "1" represents the occurrence of a "problem" for the aged person, and "0" absence of a problem.⁶

C.V. Group 3. Problems of Aging Applicants for Services. The third group of criterion variables was selected statistically, by comparing Applicants assigned to Program 1 or 2, with Non-Applicants at the time of the initial research interview. The problems which lead persons to apply were expected to occur more frequently among the Applicants than among the Non-Applicants. The variables were

selected from all those which had been coded for IBM tabulation from the research interviews or from combinations of such coded variables. Variables were considered as "potential criterion variables" if the following considerations were satisfied: First, the variable provided a basis for ranking persons from more or less "favorable" conditions; second, the variable was included in both the initial and the follow-up research schedules; and third, the variable was so defined that changes could occur in the rank order of persons between the initial and the follow-up research interview.⁷

Minimum Statistical Conditions for the Selection of Criterion Variables. In order for any of the potential criterion variables to be selected for the specified group, five statistical conditions were defined that were to be satisfied according to the results of the statistical analysis. (1) The percentages of persons within some of the sub-categories of the potential variable must be higher for the applicants than for the non-applicants. Those sub-categories for which this condition is satisfied will be referred to as "problems" of the aging applicants. The presence of such problems was coded by the number 1 and their absence was coded by the number 0. (2) The frequency of persons in the "problem" sub-categories among the non-applicant sample must be lower than 50%. (3) The frequency of the "problem" sub-categories must be higher for persons 65 years of age or over than for persons 60 through 64 years of age in the population sampled. (4) The frequency of the "problem" sub-categories must be higher for the applicants than for the non-applicants within either one of the two groups, those 60 through 64 or those 65 or over. (5) The frequencies of the "problem" sub-categories based on the initial interview must be equal to or less than the corresponding frequencies of these problem sub-categories based on the follow-up interview for the non-applicant sample.

These five conditions were applied to select one criterion variable in each of the seven content areas. If only one of the potential criterion variables in an area satisfied all the conditions, it was selected as the criterion variable

for measuring effectiveness in that area.

The Optimal Statistical Condition for Selecting Criterion Variables. If more than one of the potential criterion variables in an area satisfied these five conditions, one was chosen for which the percentage of applicants included in the "problem" sub-category was highest. In case of ties using this condition, the one chosen was that for which the percentage of non-applicants who were in the "problem" sub-category in the initial interview but were not in this sub-category in the follow-up interview, was the lowest. In order to allow for errors due to sampling, differences of five percent or less were considered to be ties.

C.V. Group 4. Problems of Aging for Non-Residential Services. The fourth group of criterion variables was statistically selected using the same methods as for the third group just described. However, the methods were applied to information obtained for applicants selected for Program 0, who were referred to the district offices of the Community Service Society, and Non Applicants. The persons referred to the district office were selected by random sampling from all applicants after excluding certain of the applicants from the referral because of agency policies. Any applicant for whom the request for service was concerned with the residences for aging operated by the Society or the summer camping program for aging persons, was not referred to the district office. Since persons selected for Program 0 were interviewed only at the time for follow-up, it was necessary to use data based on the follow-up rather than on the initial interview.⁸

With respect to each of these criterion variables, persons were classified into two sub-categories depending on the presence or absence of "problems." The symbol P_{jst} will be used to represent the proportion of persons in the population who would be included in the "problem sub-category" with respect to a specified criterion variable "j" assuming that all of the persons had been selected for a particular program "s" during a particular period for experimentation "t;" each of the measures of effective-

ness can be defined in terms of the following statistical index:

$$E_{jst} = \frac{P_{jot} - P_{jst}}{P_{jot}}$$

Measures of Costs of Programs. In this experiment two measures of cost have been used: (1) The mean length of time between the beginning and termination of services. (2) The mean number of in-person interviews by a member of the professional staff. The symbol C_{kst} will be used to represent any defined measure of costs of programs. The value of this measure will depend on the way in which cost is defined, represented by the subscript "k", as well as upon the particular program represented by the subscript "s", and upon the period of time represented by the subscript "t."

Uses of Measures of Result of Experimentation. These measures of effectiveness and costs had three major purposes: To select one of these alternatives as the "best" for most persons in the population until further experimentation may change this conclusion; to select any one or more of the programs for repeated experimentation; to suggest improved programs which may be expected to satisfy certain minimum levels of effectiveness.

In this experiment the minimum level of effectiveness was defined to be equal to the value computed for a specified criterion variable, assuming the frequency of problems among Non-Applicants 60 and over is reduced to the level occurring for those 60 to 65, and that the frequency of problems among those in the Applicants' sub-population is reduced to the level occurring in the Non-Applicant sub-population.

B. THE QUESTIONS TO BE ANSWERED BY EXPERIMENTATION

In order that the defined measures of effectiveness and costs could be used for the purposes of selecting and improving programs of services, it was necessary for the values of these measures to be determined by hypotheses which have been or can be experimentally tested. The unknown values of the measures of cost and effectiveness that are required can be expressed in the form of questions:

First, with respect to the purpose of selecting alternative possible programs for maximum utilization, two questions had to be answered. (a) Which of given possible alternative programs is most effective with respect to all selected measures of effectiveness (E_{jst})? (b) If more than one of the given programs is most effective, which one is least costly with respect to the measures of cost (C_{kst})?

Second, with respect to selection of programs for further experimentation, which of the experimental programs is more effective than the "Basic" program with respect to reducing the mean number of "problems" for a specified group of criterion variables?

Third, with respect to the development of improved programs the following questions had to be answered. (a) Do any of the existing programs satisfy minimum conditions specified with reference to each of the measures of effectiveness? (b) If not, how can these programs be modified or developed to increase the degree of effectiveness with respect to defined measures?

II. HYPOTHESES TO BE EXPERIMENTALLY TESTED

The hypotheses that were formulated prior to experimentation were so formulated that acceptance of the stated hypothesis would lead to the selection of the "Basic" program pending further experimentation.

1. Hypothesis for the Selection of the "Best" Program

Hypothesis 1(a) - The effectiveness of Programs 1 or 2 will not be greater than that for Program 0 with respect to all of the defined criterion variables.

2. Hypothesis for Selection of Programs for Further Experimentation.

Hypothesis 2(a) - In reducing the mean number of "problems" in criterion variable Group 1, Group 2 and Group 4, the effectiveness of Program 1 will be equal to or greater than Program 0.

Hypothesis 2(b) - The effectiveness of Program 2 will be equal to or greater than Program 0.

3. Hypotheses Concerning the Development of Programs

Before some experimentation was carried out, it was not possible to formulate hypotheses with reference to specific ways in which programs can be improved in effectiveness and costs.

III. STATISTICAL-EXPERIMENTAL TESTS OF HYPOTHESES

Major Stratification: For the purposes of sampling, the defined population was sub-divided into major strata; (1) The "Applicant" stratum consisted of those persons in the defined population for whom an application was made for services in the Community Service Society. This strata also included any other person in the population living in the same household with such "Applicants." (2) "Non-Applicants" consisted of those persons in the population for whom no such application was made at the time selected.

The Master Samples. Within each of these strata, "master samples" were selected by random-probability sampling methods.

The first step in selecting an "Applicant" Sample was listing all persons for whom an application is received in the Special Services for the Aging office during a given month. All persons were selected for whom a case was opened for service between the date listed and the final date for the follow-up interview about six months later. A sample for all other Applicants was selected by simple random sampling within each ten consecutive listed names.

Three Non-Applicant samples were selected using lists of blocks, block segments, or dwelling units available from the United States Bureau of Census reports and maps provided by the New York City Housing Authority.

Multiple Samples for Months of Study. The specified sampling methods were applied to select a series of samples for each of the eight months during which the study was carried out, from April 1 to December 1, 1957. The number of persons sampled in any given month depended on the number of persons who could be served by the available staff. For the Applicants, samples were selected each month from all those for whom applications were received during the month. For the Non-

Applicants, some samples of blocks, block segments, or dwelling units were selected each month and the "Participants" residing in the sampled area were listed and interviewed by the research staff.

Random Selection of Persons for Alternative Service Programs. Within each of the "master samples" of Applicants and Non-Applicants, three sub-samples were selected by random-probability sampling. Each of three alternative service programs were preassigned to one of these three sub-samples. Consequently, each of the programs were made available to persons in one of the "sub-samples" selected from the "master samples" previously selected from the defined population.

Random Selection for Interviewing. Not only were "Participants" sampled for alternative programs, but in addition they were randomly sampled for assignment to various persons in the interviewing staff. Each month a number of persons were selected and preassigned to each one of the interviewers. Some changes in assignment were made after the random assignment under special conditions. No substitutions were made.

Size of Samples and Probabilities of Selection. The size of samples and probabilities of selection were predetermined for each of the monthly periods of interviewing on the basis of two general conditions: (1) The total number of persons sampled during the twelve month period was to be sufficiently large to apply statistical tests that assume large samples: (2) The maximum number to be sampled and interviewed in any given month was not to exceed a number which could be interviewed and served by the research and service staff in conformity to the standards and procedures defined for the experiment.

The number of persons sampled for the initial interviews in the Applicant Sample was 119 and in the Non-Applicant Sample was 133. The number sampled for the follow-up interview in the Applicant Sample was 217, and in the Non-Applicant Sample was 84.

The information used to classify persons according to the criterion variable for measuring effectiveness was obtained by the use of standardized research schedules. The two basic schedules were the Program Evaluation Schedule, used for interviewing the Participant to obtain self-

ratings and objective information concerning the Participant and others in the household, and the Research Interviewer Schedule used by the interviewers to record their observations and judgments.

Definitions and instructions for the administration of the research schedules were used in the training and supervising the interviewing process. In general, they required that the interviewing be attempted and completed within the first month after selection and again within six months after selection. For Applicants sampled for Program 0 (District Office Service) no interviewing was attempted until the time for the six month follow-up. The place for interviewing was to be within the home of the Participant or in the office of the Community Service Society.

An unlimited number of attempts could be made within a specified time period for interviewing. If the interview could not be completed within one month, two weeks additional time was allowed in order to complete the contact if possible. It was not expected that all persons be interviewed. However, it was expected that at least three attempts be made to interview. If the interviewer was unable to complete the interview within the specified time period, the reasons for noncompletion were specified.

The research interviewers were selected from persons with previous social work training and experience, but in this study they were selected, trained and used only for interviewing and not for any service functions.

The general policies for the administration of the service program were those already in practice within the Community Service Society, but some modification of these policies was required by the experimental design. Written instructions were developed for these policies. The service workers for Programs 1 and 2 were selected, trained and supervised in carrying out the services according to these policies for approximately one year before the experimental study was begun. The same persons were involved in carrying out both Programs 1 and 2 in the Special Services for the Aging unit. A different service staff carried out Program 0 in the district offices.

In order for any person to actually receive any service, it was first neces-

sary for an application to be made and for the case to be opened for service according to the usual policies and practices of the Community Service Society. Persons initially interviewed by research staff who were sampled for Program 1 or 2 from the "Non-Applicant" stratum of the population, were "informed" about the availability of these services at the Community Service Society by the research interviewer. However, the research interviewer was not to enter into a service relationship with the persons interviewed nor to encourage them to apply, except in case of special emergencies.

The hypotheses about the effectiveness of alternative programs in this experiment were applied in making two kinds of decisions. (a) The first decision was to select one of the alternative programs or the "best" for maximum utilization by the agency in providing services for persons in the defined population. (b) The second decision is to select one or more of the alternative programs for further experimental testing. Accordingly, two somewhat different rules for testing the hypotheses will be defined and applied.

Statistical Tests of Hypotheses in the Selection of Programs for Maximum Utilization. As the basis for testing this hypothesis, results of research interviews carried out in the six-month follow-up period for Participants sampled within the "Applicant" stratum during April 1 to December 1, 1957 were used. Those Participants identified as "exceptional" cases because they could not be referred to the Program 0 - District Office Service Program were excluded from comparison of Program 1 and Program 2 with Program 0. Results of interviews for the "Non-Applicant" samples were not included because only one sampled Non-Applicant requested service within the six-month period for the follow-up research interview. The numbers of sampled Participants included were: For Program 0, $n_0 = 68$; Program 1, $n_1 = 44$; and Program 2, $n_2 = 22$.

In order to classify persons on the basis of information recorded in the research schedule, it was necessary (1) for the persons to be located, and alive and (2) for a research interview to be at least partially completed. It was expected that the frequencies of persons

for whom information was not obtained for these reasons would be sufficiently large to make possible separate analysis, and that, these frequencies could be effected by the service program as well as by the procedures used in the follow-up interviews. In order to simplify the analysis, information about these reasons were coded in terms of the following: (1) The Participant was located and alive up to a date three months from the date sampled. (2) Research data was obtained that was sufficient to permit classifications of the sampled Participant on some of the possible criterion variables.

Program 0 is to be recommended for maximum utilization unless a Null-hypothesis 1 (a) 1; that the effectiveness of Program 0 is equal to or greater than that of Program 1, and of Program 2 is rejected. This hypothesis was to be rejected if the outcome of the experiment is as follows:

(1) For comparisons between Program 1 and Program 0: (a) The estimated percentage of persons located and alive three months after being sampled for Program 1 is greater than the corresponding percentage of Program 0. That is, $P'(v)_1 > P'(v)_0$; and (b) Among those located and alive, the percentage for whom research interview data was obtained for Program 1 is greater than the corresponding percentage for Program 0. That is, $P'(u/v)_1 > P'(u/v)_0$; and (c) Among those persons for whom research data was obtained, the estimated percentage of persons with a specified problem for Program 1 is less than the corresponding percentage for Program 0, for each of the selected criterion variables in Groups 1, 2 and 4. That is, $P'(j/u)_1 < P'(j/u)_0$; and (d) The computed value for the "Normalized t-test" for all criterion variables in a selected group is equal to or greater than a specified value " C " = 1.28^{10} ; or

(2) For the corresponding comparisons between Program 2 and Program 0:

- (a) $P'(v)_2 > P'(v)_0$,
- (b) $P'(u/v)_2 > P'(u/v)_0$, and
- (c) $P'(j/u)_2 < P'(j/u)_0$

for all criterion variables in a selected group; and

(d) The computed value of the "Normalized t-test" is equal to or greater than specified value " C " = 1.28^{10}

If this hypothesis was rejected, Program 1 was to be recommended for maximum utilization if a Null-hypothesis 1(a) 2, (that is, if the effectiveness of Program 1 is equal to or greater than that of Program 2), was rejected.

Program 2 was to be recommended for maximum utilization if both Hypotheses 1(a) 1 and 1(a) 2 were rejected by the results of the experiment.

Comparisons between Program 1 and Program 2 were to be made with reference to criterion variables in Groups 1, 2 and 3; whereas, comparisons between Program 1 or Program 2 with Program 0 were to be made with reference to criterion variables in Groups 1, 2 and 4.

Statistical Tests of Hypotheses in the Selection of Programs for Further Experimentation. The choice of programs for further experimentation was not restricted to the choice of only one of the alternatives. Any one or more of the alternative programs could be selected for further testing by the following rule:

(1) Whichever program is selected as a result of the application of the rule defined in the preceding part will be initially selected for further experimentation.

(2) Each of the remaining programs will be excluded from further experimental tests if the Null-hypothesis 1(a) 3, (that is, if the effectiveness of this program is equal to or greater than that of the initially selected program), is rejected for each program.

This Null-hypothesis will be rejected if the results of the experiment are as follows:

(a) The estimated mean number of problems among persons sampled for the alternative program is greater than corresponding estimated mean for persons sampled for the initially selected program: That is, $M'(x)_s > M'(x)_{s*}$; and (b) The estimated percentage of persons for whom research data is not obtained for the alternative program is greater than the corresponding estimated percentage for the initially selected program: That is, $P'(u)_s > P'(u)_{s*}$; and (c) The computed value for the Normalized t-test applied to this difference in percentages in condition (b) is greater than " C " = 1.28^{10}

(3) Each of the remaining programs will be included for further experimental testing, if the Null-hypothesis 1(a) 3 is

rejected for that program, according to results of the experiment that have been carried out using the rule for rejection defined in the preceding section.

Results of Statistical Tests of Hypotheses for Selection of the "Best" Program for Maximum Utilization. The results of the application of the defined rules for testing hypotheses to the data obtained in this experiment were as follows:

(1) When effectiveness was measured by C.V. Group 1, (Overall Level of Adjustment and Adaptation), the Null-hypothesis that the effectiveness of Program 0 is equal to or greater than that for both Program 1 and Program 2 was not rejected.

(2) When effectiveness is measured by the criterion variables in Group 2, (Basic Problems of Aging), this Null-hypothesis was also not rejected.

(3) When effectiveness is measured by the criterion variables in Group 4, ("Problems of Applicants for Non-residential Services") this Null-hypothesis was also not rejected.

Accordingly, Program 0 would be selected as the "Best" Program for maximum utilization.

Tests of Hypotheses for Selection of Programs for Further Experimentation. The results of the application of the defined rules for testing these hypotheses are shown in Tables B1 and B2.

(a) The Null-hypothesis, (that the effectiveness of Program 1 is equal to or greater than that of Program 0), was not rejected when the criterion variables in Group 1, Group 2 or Group 4 were selected to measure effectiveness. Accordingly, Program 1 would be selected for further experimentation, based on the criterion variables in Groups 1, 2 or 4.

(b) The Null-hypothesis, (that the effectiveness of Program 2 is equal to or greater than that of Program 0) was rejected when criterion variables in Groups 1, 2 or 4 were selected to measure effectiveness. Accordingly, Program 2 would not be selected for further experimentation based on the criterion variables in these Groups.

FOOTNOTE REFERENCES

1. The work in preparing this report was supported by funds from U.S. Public Health Service Grant 2M-6358. For more complete reports of the experiment see: Blenkner, Jahn and Wasser (1), Blenkner and Sibulkin (2) and Jahn and Blenkner (4). The research was supported by a grant from the Rockefeller Brothers Fund and by funds from the Community Service Society of New York.
2. See Blenkner, Jah and Wasser (1), and Blenkner and Sibulkin (2).
3. See Jahn (4) pp. 37-39 for a discussion of sources, procedures and results of the statistical selection of criterion variables. The sources from which items were selected or derived as potential criterion variables are given in Appendix A-5. For a more generalized discussion of the problems, principles and methods involved see Jahn (5).
4. See Kogan (6) for a discussion of the development of the "Adjustive Status Scales."
5. See Jahn and Blenkner (4) for the form used in making these ratings.
6. See Jahn and Blenkner (4) for definitions of the criterion variables included in Group 2.
7. See Jahn and Blenkner (4) for definitions of the criterion variables included in Group 3.
8. See Jahn and Blenkner (4) for definition of the criterion variables in Group 4.
9. For more complete presentation of basic principles and methods for sampling, estimation and testing hypothesis see Hansen (3) Neyman (7) and Jahn and Blankner (4)
10. See Jahn and Blenkner (4) for a definition of the "Normalized t Test" applied in this experiment. The value .05 has been selected as the probability for rejecting any "Null-hypothesis" assuming it is true, based on the specified research design and procedures.

BIBLIOGRAPHICAL REFERENCES

1. Blenkner, Margaret, Jahn, Julius and Wasser, Edna. Serving the Aging: An Experiment in Social Work and Public Health Nursing. New York: Community Service Society of New York, March 1964.
2. Blenkner, Margaret and Sibulkin, Lil-lain. Experimental Research in Social Work: A Report of a Study in Services to the Aging. Community Service Society of New York, June 1959.
3. Hansen, M.H., et al. Sample Survey Methods: Vol. I. New York, John Wiley & Sons, Inc., 1953.
4. Jahn, Julius and Blankner, Margaret. Serving the Aging: Methodological Supplement - Part I. Experimental Design and Statistical Tests of Hypotheses. New York: Community Service Society of New York, March 1964.
5. Jahn, J.A. "The Statistical Definition and Selection of Criteria for Evaluating Effectiveness in Social Work." University of Pennsylvania School of Social Work, Research Center, 1961.
6. Kogan, Leonard and Shyne, Ann. A Study of Components of Movement. New York: The Community Service Society, 1957.
7. Neyman, Jerzy. First Course in Probability and Statistics. New York: Holt, Rinehart & Winston, Inc., 1950.