Statistical Laboratory

annual report

1944 — 1945

IOWA STATE COLLEGE AMES, IOWA

TABLE OF CONTENTS

CONTENTS

PERSONNEL

The professional staff of the Statistical Laboratory, including resident collaborators, during the past fiscal year included:

G. W. Snedecor, Professor and Head	University of Michigan, Mich.
Bernice Brown, Research Associate	Iowa State College, Iowa
Mary Clem, Technician	Iowa State College, Iowa
W. G. Cochran, Research Professor	Cambridge University, England
S. Lee Crump, Research Associate	Cornell University, New York
Dorothy S. Cooke, Research Associate	University of Chicago, Illinois
W. T. Federer, Resident Collaborator	Kansas State College, Kansas
J. R. Goodman, Resident Collaborator	Friends University, Kansas
Paul G. Homeyer, Res. Asst. Professor	A. & M. College, Texas
R. J. Jessen, Resident Collaborator	University of California, Calif.
A. J. King, Resident Collaborator	University of Wyoming, Wyoming
M. V. Lindquist, Res. Collaborator	Gustavus Adolphus, Minnesota
*A. M. Mood, Res. Assoc. Professor	Princeton University, New Jersey
Abigail Stone, Research Associate	Massachusetts State College, Mass.
H. C. S. Thom, Resident Collaborator	George Washington Univ., Wash., D. C.

* On leave

The following persons were assigned during the year to the Laboratory on a temporary basis for the performance of special duties:

Jarvis Barnes	J. E. Nuquist	C. F. Reid
W. S. Crawford	E. H. Oberholtzer	Joseph Steinberg
W. D. Farwell	Jack Ogus	R. E. Straszheim
L. R. Fletcher	H. M. Peers	Benjamin Tepping
J. R. Grant	Louis Persh	Glen Vergeront
E. M. Miller	W. H. Peterson	R. B. Willey
G. W. Morris	T. J. Reed	

During the past year the work of the professional staff was implemented by a highly trained corps of about twenty clerical workers. In addition, from 150 to 200 clerks and supervisors have been employed throughout the year to prepare the Master Sample of Agriculture and its extension to a Master Sample of Population.

ACTIVITIES OF THE STATISTICAL LABORATORY

The activities of the Statistical Laboratory during the year have been affected by the war. During the early part of the year Mr. Cochran joined a staff of mathematical statisticians at Princeton University engaged with special problems of the Navy. Later he returned to Iowa State College and was assigned to Europe on a Strategic Bombing Survey. The increased need for survey data by the government, which was in part due to the war, led to an expansion of the service work at the Laboratory. The designing and preparation of sample surveys was increased. A Master Sample for the Department of Agriculture and the Bureau of the Census was completed. The agricultural sample is being extended for the Bureau of the Census so that the government will have a sample designed and prepared for a sample census of population by 1946. The materials assembled for designing and preparing the Master Sample has made it possible for the Laboratory conveniently to design and prepare other samples for the federal government.

The demands for the statistical planning of experiments, many of which are directly connected with the war effort, have continued to increase. Demands during the year for this kind of help exceeded the trained manpower available at the Laboratory. Research in the design of experiments, however, has continued about the same as in previous years.

With Mr. Cochran devoting his time to war work, research in mathematical statistics suffered accordingly.

FINANCE

During the last fiscal year the programs of the Statistical Laboratory were financed and supported by at least eight different departments or agencies. The cooperating agencies together with the purpose and amount of contribution to the Laboratory during the past year are shown in the table below.

Source and Amount of Funds
of the Statistical Laboratory
July 1, 1944 - June 30, 1945

Cooperative agency	Purpose	
	Research	Special services
Iowa State College		
Agricultural Experiment Station	$ 7,000	
Department of Mathematics	2,600	
Industrial Science Research Institute	10,000	
Other departments of the College (Computing service)		$ 2,500
United States Department of Agriculture		
Agricultural Research Administration* (Bankhead-Jones Act)	15,000	
Bureau of Agricultural Economics (General fund)	15,000**	30,000
United States Department of Commerce		
Weather Bureau	13,000	
Bureau of the Census	18,000	179,000
TOTAL	$80,000	$211,500

* Through the Bureau of Agricultural Economics

** $9,000 research and $30,000 special services from BAE general fund through the Division of Agricultural Statistics. $6,000 personnel and travel provided by the Division of Program Surveys.

TEACHING AND TRAINING

While the Statistical Laboratory is not a department of instruction, members of the staff devote part of their time to teaching statistics in the Department of Mathematics and the Department of Economics and Sociology. Other members of the College staff teach courses involving specialized application of statistics. One function of the Laboratory is to correlate these activities, to promote interest in statistical scholarship and instruction, to prevent duplication of effort and to assist in organizing such courses in statistics as may be needed in the College. The following courses in statistics have been taught by the indicated instructors in the various departments of the College during the past year:

Analytic Geometry and Statistics. Rectangular co-ordinates, graphs, equations of loci, straight line, conics, elements of statistics applied to problems in forest mensuration. - Bancroft

Statistical Methods. Sampling from biological populations, statistics and experimentation, averages and tests of significance, linear, multiple and curvilinear regression, analysis of variance and covariance, individual comparisons. - Snedecor

Statistics Laboratory. Machine calculation of statistics, use of punched card tabulating equipment. - Brown

Elementary Mathematical Statistics. Summations, moments; normal, binomial, and Poisson distributions; least squares; regression and correlation; analysis of variance. - Bancroft

Mathematical Statistics. Useful functions in statistics including Gamma and Beta functions, moments and semi-invariants, general distribution theory, Gram-Charlier series, joint distributions, correlation surfaces, multiple and partial correlation, fundamentals of sampling theory, students distribution, χ^2 distribution, Fisher's t- and Z- distributions, mathematical theory of tests of significance, Sheppard's corrections. - Bancroft

Design of Experiments. Factorial and quasi-factorial designs, confounding, incomplete blocks, balanced lattice, and lattice squares. - Cochran

Sampling Methods. Design and analysis of sampling investigations; stratification; estimation; fiducial limits. - Snedecor

Advanced Mathematical Statistics. Estimation, distribution theory, least squares. - Cochran

Advanced Mathematical Statistics. General theory of statistical estimation and tests of significance; statistical inference; applications and special problems. - Cochran

Elementary Economic Statistics. Principles and methods of gathering, analyzing, presenting, and interpreting economic data. - Homeyer

Statistical Analysis. Correlation analysis; methods of analysis of prices, production data, and similar series of time variables. - Tintner

Educational Statistics. Statistical concepts and procedures for teachers and rural school administrators. - Wert

In conjunction with the Statistical Laboratory the teaching departments of the College offer training for two groups: first, graduate and undergraduate students specializing in either mathematical statistics or its applications; and second, young statisticians in government departments who are assigned to the Laboratory to work on specialized problems arising in their respective divisions. Students are thus brought into intimate contact with the statistical research of the Laboratory, and are trained not only in the theory but also in its current applications.

CONSULTATION

During the year the staff of the Laboratory has written over one hundred technical discussions of statistical problems answering requests from research workers. The large majority of these were directed to staff members of the national and state agricultural experiment stations. Commercial research organizations came next, with a scattering of letters to other persons in this and other countries.

These discussions sometimes involve extensive computations of data and often call for research on new problems of mathematical statistics. Several weeks may be required for the completion of a single reply, and one letter often leads to another.

Many requests for help involve the salvaging of information from costly experiments that have been damaged through unavoidable circumstances. Others concern the design of experiments that cost thousands of dollars and may require ten or more years to complete.

Some research workers frequently come to the Statistical Laboratory at Ames for conferences of a day or more. This may require much of the staff's time but in the long run it is usually much more efficient than correspondence.

As soon as war-time restrictions are lifted, the extension of both conferences and correspondence will increase the opportunities for the Laboratory to improve the efficiency of this and other nation's researches in agriculture, biology, climatology, social and other sciences.

SPECIAL SERVICES

Design of Experiments

With the growth of knowledge in any subject, the questions whose answers are sought by experimentation are of increasing intricacy and elusiveness. The simpler designs fail to distinguish between sampling variation and the phenomena under investigation. Fortunately, accumulated information leads to the design of more efficient experiments so that the more difficult problems may often be attacked without undue expenditure of time and money.

In other cases, it is necessary to go into the field or laboratory to collect data which will guide the experimenter toward the better designs. This can often be done with little more expense than is put on the routine conduct of the experiment. The analysis of these special data may result in designs that give more information and at the same time cost less than the ones that have been in use.

The mathematical statistician is needed to help design these experiments because there are statistical principles that must be used to guarantee that the results will not be ambiguous. Also, the eventual statistical analysis of the data must be furnished along with the design.

In the course of the experiment accidents may happen so that some of the data may not be obtained or may be faulty. The statistician may be called in again to help salvage the maximum amount of information available from the results.

The cooperation centering in the Statistical Laboratory is designed to make such services available to members of the College staff and to research men in various experimental groups fostered by the government.

Design of Samples

During the war both the federal government and private industry have increased the use of surveys requiring personal visits with the respondent. The cost of obtaining these data has emphasized the importance of developing sampling designs which provide useful results at a minimum cost. A new kind of sampling design has recently been popularized using an area of land as the sampling unit. It has many important advantages for farm and population surveys. The Statistical Laboratory has pioneered in the development and the testing of the designs based upon the area principle and for this and other reasons has been called upon to design many different kinds of surveys for the federal agencies using these principles.

A Master Sample of Agriculture

The increased demand on the part of the federal and state agencies for more accurate and timely statistics on agricultural and population problems increased the need for more precise and integrated methods of obtaining data. The statistician has attempted to meet this increased demand by the adoption of inquiries which require personal interviews with farmers and persons in general on some sort of sample basis. This study was designed to develop a scheme for obtaining a large number of different types of desired information with a single type of sampling scheme so that various government and other agencies may be able to utilize more effectively the information obtained by each, when desired, and to prevent duplication when that is desired. To meet these needs a Master Sample has been developed and its success has been demonstrated by the use to which it has been put since its development.

The sample for the Master Sample consists of a selection of small areas, averaging about 2.5 square miles in size, distributed within every one of the 3,000 counties in the United States. These sample areas or "segments," 67,000 in all, constitute a sample of one-eighteenth of the entire area of continental United States. When appropriately used, they will provide a sample of one-eighteenth of all the farms in the United States. In order to minimize errors on the part of the enumerator in identifying the area selected for the sample, maps and aerial photographs were obtained for indicating as accurately as possible the boundaries of each of these small segments or sample areas. In addition to the drawing of the particular set of 67,000 segments, operations were carried out in such a manner that various kinds of other samples can be conveniently, accurately, and cheaply constructed for many specific inquiries from the same basic materials. In effect, this means that we have information on 1,200,000 small areas. By having statistical information on these small areas available, it is possible to construct a variety of different kinds of samples for a variety of different kinds of inquiries.

The designation of the 67,000 sample segments on maps and aerial photographs was completed November 1, 1944, in time to be used for the 1945 Census of Agriculture. A study is now underway to determine the expected sampling fluctuations of items surveyed on the Master Sample. Preliminary findings indicate that for most of the major items the accuracy of the sample estimate will be suitable for published estimates by states.

The Master Sample has already been used by several state and government agencies. It is an integral part of the 1945 Census of Agriculture. The Bureau of Agricultural Economics has chosen this sample for two of its current national surveys of agriculture. Local surveys in Tennessee, Georgia and Iowa have used Master Sample materials in the construction of their samples. The Master Sample is also the base for a proposed sample census of population to be taken by the Bureau of the Census in 1946.

<u>A Sample Census of Population</u>

Considerable interest has been expressed by both government officials and the public for information of the kind obtained in the Decennial Census of Population. The Bureau of the Census has proposed to obtain this information by means of a sample instead of a complete enumeration. The Statistical Laboratory was assigned the job of preparing such a sample to be ready for field operations January 1, 1946. A sample of this type, if it is to be adequate and yet economical, requires considerable preparatory work based on detailed maps of both the open-country and the densely populated areas. The Master Sample project already had available the required maps and materials for the open-country areas and a complete listing and designation of the unincorporated areas. To provide, for sampling purposes, materials for the densely populated areas, the Bureau of the Census obtained a complete set of maps from the Sanborn Map Company, at a cost of $385,000, and sent them to the Statistical Laboratory for processing. The design of the proposed sample census of population is therefore, in a sense, an extension of that of the Master Sample of Agriculture (which was completed November 1, 1944), for the same general method was followed. In the open-country areas the Master Sample areas were adopted for the population sample (for which purpose they were previously designed). Like the Master Sample of Agriculture this sample used the area method of sampling and sample areas were made as small as was consistent with statistical efficiency and the problems of the field. The sample comprises small areas representing 1/27 of the population; in certain states of relatively small populations a greater rate of sampling is used to obtain suitable accuracy for state estimates. In general, incorporated places over 2500 population will be covered adequately by the detailed Sanborn maps, certain information from which will be recorded on International Business Machine cards and sent to Washington, D. C. for sampling and filing for future sampling. In the unincorporated areas use is being made of Sanborn map coverage where available. Sanborn "spot-in" maps are being used wherever available and the remaining areas are sampled from aerial photographs obtained on large scale aerial photographs.

As a part of the project and also as a part of the program for the 1950 Census of Population, aerial photographs are being obtained for all unincorporated places in the Master Sample listing and all incorporated places of a 1940 population of less than 2,500. Boundaries for these unincorporated places will be delineated on these photos so that they can be identified by the enumerator.

Although this sample was designed for sampling population, plans have been made to use it, with some modification, for proposed special inquiries, such as some of those in the government's Basic Economic Statistics program. These inquiries included cross-sectional samples of the population for obtaining statistics on income, expenditures and savings and more detailed information on the nation's labor force. Because of the flexibility of the sample and because of the kinds of materials that were prepared, it is possible to provide a wide variety of samples for many of the needs that are arising during the close of the war and the period following.

Special Sample Surveys

The Master Sample materials have found considerable use in preparing special sample surveys requested by the several government agencies cooperating with the Statistical Laboratory. These special samples are either parts of the Master Sample or consist of a new selection of areas according to which is more appropriate for the particular inquiry at hand.

One such sample was that for a proposed quarterly survey of the labor force for the Bureau of the Census. This sample, which is in a sense a forerunner of the population sample census, required a cross-sectional sample of the population within approximately 315 selected counties in the United States. Sample areas within the selected counties for the open-country areas and unincorporated places were designated by the Statistical Laboratory. The Bureau of Agricultural Economics has in the field at the present time two sample surveys: a quarterly survey of agriculture and a survey of farm wage rates. The sample for the quarterly survey of agriculture comprises 101 counties selected for general purpose agricultural inquiries within which about 3,500 farms are to be interviewed. The farms designated for this sample are located on the Master Sample areas. The aerial photos purchased for the Master Sample of Agriculture and covering the designated areas for this survey were used by the field enumerators for identifying the selected segments (sample areas). The sample for the farm wage rate survey comprises not only the 101 general purpose counties selected by the quarterly survey of agriculture but in addition 57 other counties. The total of 158 counties appears to be adequate providing wage rate estimates for each of the four major geographic regions in the United States. The areas within these counties on which farms are enumerated were selected so as to properly exclude Master Sample areas and were made larger in order to lessen the need for aerial photographic coverage. Master Sample materials provided a convenient means for selecting the sample segments within the selected counties. In the wage rate survey Master Sample segments were purposely avoided (to minimize interview traffic on Master Sample farms). In addition to these national surveys, research workers at Iowa State College have also made use of the Master Sample materials for some of their sample inquiries.

Computation Services

A very essential part of the organization of the Statistical Laboratory is the computing service. A corps of expert computers is maintained, trained to carry through all the regular calculational procedures required by members of the professional staff. Their services are available to any member of the College staff for computations approved by any professional consultant in the Laboratory. Modern equipment includes an adequate set of electric punched card sorting and tabulating machines.

RESEARCH

Mathematical Statistics

The analysis of components of variance in "all or none" data

In "all or none" data, i.e. samples drawn from a binomial population where each individual can only be classified in one of two classes, the animal breeder often wishes to have a measure of the variation among the probabilities of success of several different groups of animals; e.g. in some work with chickens an estimate of the variation in the proportion of dead chicks between dam families or between sire families was desired. A technique has been worked out following the usual analysis of variance schemes whereby unbiased estimates of this variance may be obtained from linear combinations of the mean squares from the lines of the analysis of variance. The results will be extended to the case of unequal sub-class numbers and to tables of multiple classification.

A test of sampling errors and bias in mailed inquiry samples

The reduction in the sampling error and in the bias of estimates of means made from maldistributed samples, through various weighting schemes, has been studied. The data used for illustrative purposes consist of a sample of 8,193 farms selected from the 1942 South Dakota Assessor's books. This sample is considered as a population from which a subsequent sample of 1,866 farms, indicated by return of mailed questionnaire, was drawn. Estimates of means of several items in the large sample from data on the 1,866 farms of the small sample were made by various weighting schemes. The results showed that in order to reduce the bias component of total error very substantially, weighting must be carried to such small divisions as to be practically impossible. The implication here is that efforts should be directed towards getting well distributed samples rather than towards attempting to eliminate bias in the present samples.

The mathematical justification for partitioning Chi-square from a contingency table into its component parts

More precise methods of analysis of data are very important to research workers in all fields. Such tests enable the researcher to obtain the maximum amount of information per unit of time, expense and space. A considerable amount of data is analyzed by the Chi-square method of analysis. Various groups of data, such as enumeration data and variances, are analyzed by this method. Chi-square has been partitioned into its component parts for particular cases. It was desired to determine the mathematical justification for the specific instances and to extend to the general case the partitioning of Chi-square from a contingency table into its component parts.

This problem has been solved by setting up the hypotheses assumed when Chi-square is partitioned into its component parts. In a 2 x 2 table, $\begin{array}{c|c} a & b \\ \hline c & d \end{array}$, the hypotheses tested are (1) $a + b = c + d$, (2) $a + c = b + d$ and (3) $a + d = b + c$. Consequently the stringent hypothesis that $a = b = c = d$ is tested in three steps while the ordinary method of χ^2 tests the hypothesis that $ad = bc$. Publication of the results is anticipated in the near future and will contain an example with computational directions. For data suitable for Chi-square analysis this method yields more information per unit of effort expended.

A textbook in agricultural statistics

Mr. Snedecor is revising his textbook, "Statistical Methods Applied to Experiments in Agriculture and Biology." This book has found wide use among research workers in the United States Department of Agriculture and its associated state agricultural experiment stations. The revision is greatly amplified to bring it up to date. Many of the findings of this cooperative project are incorporated in order to make them generally available. It is expected that the revised edition will be ready for use during the fall of 1945.

Design of Experiments

Sampling methods for yield determination in experimental plots

Work was continued during the past year to determine the optimum sampling methods for yield determinations in experimental plots of corn, oats, wheat, soybeans, hemp and hay. Since the work of sampling must usually be carried out under pressure of time, it is important to investigate the most efficient methods of sampling these crops.

Uniformity data by individual stalks and ears obtained from 24 plots of drilled corn during 1944 again showed that the commonly used 8-foot segment as the sampling unit is less than half as efficient as individual bearing stalks taken systematically throughout the plots even when adjustments were made for stand or number of ears.

An analysis based on uniformity data by hills from 80 plots of corn, each containing 80 hills, indicates that a 2 by 2 hill sampling unit is about 25 percent more efficient than the commonly used 2 by 5 hill sampling unit and that a sampling unit of every k^{th} hill is about 200 percent more efficient when adjustments were made for number of producing stalks.

One-hundred-four plots of oats were sampled to compare the efficiency of using quadrats 3 feet by 3 feet and 2 feet by 2 feet as sampling units. The statistical efficiency of the 2 by 2 quadrat was 49 percent greater than for the 3 by 3, and is in agreement with the 52 percent obtained in 1943. On the other hand, there are two other factors favoring the 3 by 3 quadrat, more than offsetting the greater statistical efficiency of the 2 by 2. Estimated plot yields based on the 2 by 2 samples were rather consistently higher than those based on the 3 by 3 samples, again indicating a significant bias. Also, time records for the sampling of the 104 plots show that it takes only about 1.4 times as long to take a given number of 3 by 3 samples as it does to take the same number of 2 by 2 samples. Considering both time and statistical efficiency the results favor the use of a 3 by 3 quadrat as the sampling unit. Uniformity data totaling one thousand 3 by 3 samples were obtained in each of two oat fields, one in which the seed were planted by broadcasting and the other by drilling. The analysis of these two thousand samples has been started and when completed will indicate the most efficient method of taking samples using 3 by 3 quadrats. The data for wheat are in close agreement, as would be expected, with those obtained for oats.

Uniformity data using 4-linear-foot segments as the sampling unit were obtained from a field of soybeans, giving a total of 1,000 samples. Data from these samples are being analyzed to determine what multiple of the size sampling unit used gives the highest efficiency and what is the most efficient method for taking the samples.

The data available on alternative methods of sampling hemp and hay are as yet too limited to draw any conclusions. These two crops will be given considerable attention this year.

Recommendations based on this study have been extremely useful to experimenters in Iowa who sample plots for yield determinations. This study will be continued and after the sampling data for this crop year are analyzed, a manuscript for publication will be written.

Machine card methods for analyzing lattice designs

During the past few years there has been a considerable increase in the use of lattice designs to test large numbers of varieties of various crops. These designs are particularly useful to plant breeders in isolating new and improved varieties of hybrid corn, disease resistant oats, etc. The analysis of data from these designs is rather burdensome because the method is more complicated than for simplier designs and because the number of varieties being tested is usually large. Procedures were developed to analyze data from lattice designs with the use of electric-punched card tabulating machines. These procedures materially decrease the time and effort required for the analyses. A manuscript giving these procedures has been written and will be turned in for publication as a bulletin. The bulletin is designed to be helpful also to those who wish to analyze designs without Hollerith machines.

An analysis of the A.O.A.C.*chick method of assay for vitamin D

In view of the extremely variable results obtained from the A.O.A.C. chick method of assay for vitamin D, the Animal Vitamin Research Council requested the Iowa State College Statistical Laboratory to analyze the method to evaluate the relative magnitude of the variability by sources. Data from a total of over 8,000 chicks were supplied by the council from 31 laboratories located throughout the United States. Based on our analysis, we recommended a change in the levels of vitamin D fed in assaying an unknown oil and a change in the method of estimating the potency of an oil, which were included in a report to the council. Also, suggestions were made for redesigning future experiments to make it possible to evaluate additional sources of variability, which was not possible with the present data. The report that was prepared is to be revised and submitted for publication as a bulletin.

* Association of Official Agricultural Chemists

Optimum size and shape of plot for fertilizer experiments

Uniformity data obtained for corn, oats and soybeans are useful to discover the optimum size of experimental plot. Two areas of uniformity data for corn have been divided into plots of various sizes and shapes to find the optimum size of plot for fertilizer experiments. Totals for yield have been computed for these plots and the variances are now being calculated. The same type of analysis is to be made of data for oats and soybeans. Uniformity data previously obtained by the Iowa Agricultural Experiment Station are also to be utilized in this study.

On the efficiency of tillage experiments

Many of the experiments in Iowa on such problems as the response of crops to fertilizers and to different methods of tillage are conducted on farms throughout the state. Based on the results, the experimenters wish to make general recommendations to farmers by soil types. One important question in designing the experiments is: for a given amount of equipment, money and time how many locations should be chosen and how many times should the treatments be replicated on each farm to get the maximum amount of information? An analysis was made to estimate the relative efficiency of using varying numbers of locations on a given soil type and varying numbers of replicates at each location. The results, for example, indicated that more information could be obtained from four locations each with three replicates than from three locations each with 15 replicates. Similar analyses are to be made of other experiments conducted in Iowa. In addition, studies are to be made to estimate the number of years that such experiments should be run. Methods of analyzing the data from group experiments will also be studied.

The efficiency of the soil sampling in tillage experiments

During the past several years experiments have been conducted by the Iowa Agricultural Experiment Station to study the effect of different tillage methods and crop management practices upon soil moisture. Since such a study requires a relatively large number of soil samples taken at various depths at different times of the growing season during a number of years, it is important to use an efficient sampling method. Statistical analyses have been made from time to time to study the efficiency of the sampling methods used. Based on these findings, changes in the sampling method have been made at different times.

All of the soil sampling data obtained during the past four years are to be summarized. Certain additional analyses will be made. The findings are to be published in a journal paper.

Evaluation of efficiency of rotation experiments

In order to discover the best systems of crop rotation for an area, experiments are being conducted to compare several crop rotations which differ in length and percentage soil-depleting crops. Since an experiment of this type must be continued over a period of 10 to 20 years, an inefficient design causes great loss of time and labor. Much information is needed on the principles which should govern the method of layout for rotation experiments and on methods for analyzing the results. The data available so far from rotation experiments conducted by the Iowa Agricultural Experiment Station have been obtained. An analysis of these data has been started and will be continued this coming year. Where possible, data from rotation experiments in other parts of the United States will also be used.

Studies in sample size and number of replications for guayule

The number of samples per plot as related to the number of replicates in a replicate variety trial on guayule was studied by the method of component analysis. It was desired to determine the point at which more samples per plot should be taken in preference to more replicates in order to obtain the highest precision per unit sampled. It was found that little additional information was gained by sampling more than 8 to 12 plants per plot in studying dry weight of shrub, percent of rubber and weight of rubber per plant. This study was initiated on the Guayule Research Project, Salinas, California, and completed at the Statistical Laboratory, Ames, Iowa. A paper on the results has been accepted for publication in the Proceedings of the American Society of Agronomy.

Design of Samples

A sample of counties for national and local agricultural surveys

National and other sample surveys involving large areas are more conveniently and economically made if the activities of the field force can be concentrated within relatively small areas, such as counties, rather than having them scattered thinly over the whole area under inquiry. Sampling for these surveys, therefore, is usually done in two stages--first, a sample of counties is selected, then, within those sample counties a sample of farms is taken. It is important that the manner in which these counties is selected is such that the survey made on them will reflect the characteristics of the universe (that is, all counties for which the inquiry is to apply) as accurately as possible.

For a selected group of agricultural items on which 1935 and 1940 published census data for counties are available, the effectiveness of different methods of stratification, of different probabilities of choosing counties and of different methods of estimation are being studied. Schemes of stratification being tested include one chosen by the Bureau of Agricultural Economics and one developed on this project. The Bureau of the Census is cooperating on this study.

Findings at the present time indicate that: (1) For the nation as a whole, a large number of important agricultural items or characteristics can be adequately estimated from surveys consisting of only 100 of the total 3,000 counties. (2) The accuracy of estimates can be greatly increased by using a method of expansion based on the ratio of the item for the current year to some previous base year rather than the use of the reciprocal of the sampling ratio which is now the general practice. (3) Regression estimates will be no improvement over ratio estimates for most of the agricultural items studied. (4) For different methods of estimation different methods of choosing counties are better, i.e. when expanding by use of reciprocal of the sampling ratio for most of the items studied, it is better to choose the counties with probabilities proportional to a value item, whereas expansion by ratio to the same item to a base year, indicates that counties should be drawn with probabilities proportional to the 1940 number of farms minus croppers. (5) Differences in efficiency of estimates made by using regional ratios and national ratios are not consistent from item to item, i.e. in some cases the use of regional ratios is to be preferred to use of a national ratio, and in others the national ratio is preferable. (6) Methods have been developed to supplement selections of counties for general purpose inquiries in order to obtain more adequate estimates of special items usually largely concentrated in certain localities.

Findings from this study have been helpful in setting up sampling designs for a national 100-county Bureau of Agricultural Economics sample survey of farmers' production intentions in April 1944, a national 38-county sample survey to supplement the Census' special surveys counties for a general farm inquiry taken in April 1944, a national 158-county sample survey of farm wage rates for Bureau of Agricultural Economics and a national 101-county sample survey of agriculture for the Bureau of Agricultural Economics. Findings are immediately considered for application in the design and utilization of sample surveys with which the Statistical Laboratory is concerned. Part of this study has been written in a thesis by D. E. McCarty.

Optimum size and kind of sampling unit for sampling within counties

During the year some additional work was done to determine how much gain in statistical efficiency results from using larger sampling units with sub-sampling within units rather than the relatively small ones now being used in the Master Sample with complete enumeration within units. The answer to this question is expected to vary with the items chosen and the type of variability existing within counties, with the stratification used, with the method of choosing the primary sampling units and the farms within units, with the method of estimation used and with the rate of sampling. In addition to the statistical aspects, cost considerations are very important and will be studied.

Until recently, the data available for studying this problem have been limited. The present analysis is based on a number of agricultural items for practically all farms in Hancock County, Iowa, for 1936. The basic data were collected and tabulated by the Bureau of Agricultural Economics in 1937-38. Also, maps were prepared showing the geographic location of each farm within the county, which are essential for conducting this study.

The county was divided into 32 geographic strata and into 16. Also, the county was divided into primary sampling units of six different sizes, containing an average of 3.5, 7.1, 14.2, 28.4, 56.8 and 113.7 farms, respectively. Two methods of choosing the primary sampling units were studied: with equal probability and with probability proportionate to the number of farms. When the smallest size of primary sampling units was used, all of the farms within were enumerated. When the larger ones were used, a constant number was subsampled. The variances computed for each sampling design were for a sample based on one primary sampling unit chosen from each stratum. This gave two over-all rates of sampling, 1 in 16 or 1 in 32. The "per farm" method of estimation was used.

When the county was divided into 32 strata and a sampling rate of 1 in 16 was used, the gains in sampling efficiency from using the largest size of primary sampling unit and subsampling rather than the smallest size with complete enumeration of farms within were only from 3 to 16 percent. Choosing the primary sampling units with equal probability or with probability proportionate to size had practically no influence on the efficiencies of the sampling designs studied. When the county was divided into 16 strata and the sampling rate used was 1 in 16, the smallest size of primary sampling unit studied contained an average of 7.1 farms instead of 3.6 farms. Under these conditions the relative gains from using the largest size of primary sampling units and subsampling varied from 17 to 39 percent. These greater gains are attributed to the fact that an average of 7.1 farms were taken per stratum with 16 strata instead of 3.5 with 32 strata. They are not attributed to the change in degree of stratification because when an average of 3.5 farms per stratum were taken (reducing the sampling rate to 1 in 32), the gains from increasing the size of the primary sampling units were only from 7 to 14 percent.

Plans have been inaugurated to extend the study to include a sample of the counties being used to estimate the sampling errors for the Master Sample.

Results obtained so far indicate that very little statistical efficiency was lost from using clusters of farms in regions similar to Hancock County in the Master Sample. Also, the results have been useful in designing other samples. When the study is completed, we hope to be able to state rather specifically the conditions under which it pays to use larger sampling units with subsampling within units.

Estimated sampling errors in the Master Sample of Agriculture

The supplementary schedule for the 1940 agricultural census was taken on a sample basis on all of the Master Sample areas. For the estimates from this schedule to be of the most use, it is necessary to know at least approximately the sampling errors of these estimates. A study was started this past year to estimate these sampling errors. A total of 272 census enumeration districts containing 26,475 farms located in 56 counties and 34 states was used. For each enumeration district maps were available showing the location of each of the farms contained therein. These maps and the 1940 agricultural census schedules for the corresponding farms were supplied by the Bureau of the Census.

The Master Sample counting and sampling units were designated on these maps. The counting units not included in the Master Sample were also divided into sampling units following the same rules used in the Master Sample. A total of 79 items from the 1940 agricultural census schedule was listed for each of the 26,475 farms. Each farm was identified by the sampling and counting unit, enumeration district, county, state and region in which it was located. The listing sheets were sent to the Bureau of the Census where nine punch cards were punched for each farm including all of the information listed. Tabulations of these cards by a number of classifications are being made with tabulating machines. From these tabulations, the sampling errors are to be estimated and are expected to be completed within a short time.

The same data used in this study will be extremely useful in studying alternative methods of sampling within counties. In listing and tabulating these data and punching the cards, not only the immediate job of estimating the sampling errors of items estimated from the Master Sample was considered but also a long range research project to study within county sampling.

Relationship between shelling percentage and moisture content of corn

The object was to determine a method by which bushels of shelled corn of a specified moisture content can be estimated from measurements of moisture content and shelling percentage of green corn.

The data were obtained from Corn Storage Investigations, Agricultural Adjustment Administration in cooperation with the Iowa Agricultural Experiment Station in 1940. In the fall of 1940 about three fields in each county were sampled from 3 to 6 times at different dates. Each sample from the field consisted of about 15 ears of corn. Determinations of moisture contents and shelling percentages of green corn, kernel and cob were made on each individual ear, making available data on about 6,000 ears of corn, sampled at different dates from fields in all sections of the state.

It has been the custom to make use of a curve showing the relationship between kernel moisture and the number of pounds of ear corn required to make 56 pounds of shelled corn at 15.5 percent moisture. The curve has been used in corn yield contests, in experimental plots and in determining the number of bushels for which to pay a farmer for his hybrid seed corn.

From the data it was possible to get the actual bushel weight of the sample from each field and a measure of its variability. The fields sampled differed by districts, by varieties and perhaps even by dates of sampling. In some fields, there was as much as 6 or 8 pounds per bushel difference between the actual weight of the shelled corn of the sample and the weight as estimated from a commonly used curve. In terms of money involved, an error of as much as 3 pounds per bushel may mean $120 to a farmer who is being paid $1.00 per bushel for seed corn and has 40 acres which yield 90 bushels per acre. In corn yield contests, an error in the curve may actually affect the placement of the entries. An examination of the 1940 data would indicate that this error might be reduced to about one pound per bushel by the process of actually sampling the grain from each farmer's field and estimating the bushel weight from the sample.

The investigation of the corn storage data will be continued and findings made available to all coordinating agencies through the Iowa Agricultural Experiment Station.

<u>Rural community surveys in Hamilton County</u>

Hamilton County, Iowa, was chosen by the Division of Farm Population and Rural Welfare and by Iowa State College as a key county in this state for various community and social studies. Previous sampling methods used in these studies have often given serious biases with respect to several factors that were checked. With the use of the Master Sample materials two samples of this county were drawn: farm sample survey and Webster City trading area sample survey.

In the farm sample survey twenty of the Master Sample counting units were chosen as the primary sampling units using a stratified random method of selection. A pre-list of all the farms and individuals living on these farms was made with respect to several economic and social factors. A subsample of about eight farms was then chosen from each of the primary sampling units giving a total of 160 farms for the county. On a number of factors checked the sample indicated no bias. These were some of the same factors that showed serious biases when the quota method of sampling was used. The farms were pre-listed at a rate of about 25 per day. The pre-list information will be useful in drawing subsequent samples and makes it possible to stratify by any of the factors pre-listed.

In the Webster City trading area sample survey the Master Sample materials were used to draw a sample of families living within Webster City and its trading area. No obvious biases were detected; however, a detailed check has not been completed. The information obtained from this sample survey has been used as the basis of several studies by the sociologists, both federal and state.

Sample surveys of post war rehabilitation needs of Iowa farmers

The governor of Iowa appointed a "Commission on Post-war Rehabilitation" early in the fall of 1944 to study the Iowa farmers' plans and needs in the immediate post war years. The Iowa Agricultural Experiment Station and the Extension Service were asked to cooperate in collecting information. The sample had to be designed and maps prepared for the enumerators in two days.

It would have been impossible to design an efficient sample in so short a time had materials from the Master Sample not been available. The counting maps from the Master Sample were used and more than 300 sampling units were chosen within the 99 counties of Iowa with the expected number of farms totaling about 1,100.

State estimates were made of the need of Iowa farmers as revealed in the survey. The results of the study were publicized in newspaper and magazine articles. The governor and his advisers have made use of the results in making recommendations to the various agricultural committees of the State legislature.

Sampling by cores for yield of scoured wool

In cooperation with the Wool Division, Livestock and Meats Branch, War Food Administration, extensive analyses were made of sampling for percent yield of scoured wool. Three types of corers were used with from 3 to 10 cores per bag. Each bag of wool was scoured at College Station, Texas, while the samples were scoured individually in Washington, D. C.

While considerable variation was encountered, there was some evidence that the wide corer gave more consistent and more accurate results than the narrow. The source of a small bias in all the estimates was later discovered and rectified.

From the data examined it was plain that estimates of yield in 11-bag clips could be made within 1.5 percent of the correct value by taking 3 cores per bag, and within 2.5 percent with a single core. But in subsampling, where only part of the bags were cored, the estimates were not so accurate. This brings the investigation to the stage where administrative problems must be considered. In field practice, what proportion of bags is it feasible to sample? What are the changes in moisture content during storage and shipping? Data for the solution of such problems were sought in the field during the season of 1945, but are not yet available for analysis.

STATISTICAL EXAMINATION OF CLIMATOLOGICAL DATA

Network Problem of Observation Stations

The theory of sampling for analysis of errors in climatological networks have been partially analyzed. Results seem to indicate that ordinary networks, such as that in existence in Iowa and other states, are dense enough for the measurement of totals for periods of one month or longer. This suggests that the critical lower density may be that for the synoptic network and studies of the sampling problem associated with this network have been initiated.

The monthly rainfall for 5- and 27-station networks in Ohio and 5- and 13-station networks in central Iowa have been analyzed. This analysis has included an examination of the distributions, the autocorrelation in time and theoretical physical factors in random processes associated with the precipitation of moisture.

Drought Problem

A study has been initiated to test the theory that the atmosphere is a storehouse of weather and that once a process, such as drought, is initiated, its duration is related in some way to the initial potentiality for producing the condition. There has been much previous work done on the drought problem in general. One notable example is the work of Blumenstock. Several criticisms can be leveled at this work, the most important of which are: (1) It is too complicated for even the technically trained agriculturist to use satisfactorily. (2) The minimum rainfall amount employed, 0.1 inch, is too small. It is not believed that this amount of rainfall is important agriculturally under any conditions. For example, we know that the normal Iowa rainfall required is rather ideal as measured by the success of agriculture. Hence, if the normal crop season rainfall is divided by the number of days with rain, the mean rainy-day rainfall is found to be 0.4 inch. This is an ideal rainy-day amount and would be too high for breaking drought. Therefore, the Russian value of 0.2 inch was adopted on the basis of which non-disjunct 20-, 30- and 40-year records have been tabulated in frequency tables for Ames and Des Moines. These are now being subjected to statistical treatment.

Rainfall Frequency

Due to the growing importance of the control of soil and water losses on small drainage areas, an increasing need has been felt for adequate short duration rainfall frequencies for Iowa. The tabular data for these are being compiled in the Weather Division of the Iowa State Department of Agriculture, Des Moines, Iowa, while the theoretical work is being done at Iowa State College. The latter consists in extending work previously done by Mr. Thom on the frequency distribution of rainfall amounts at a point. It is hoped that important portions of these studies will be completed for use in connection with imminent agricultural flood control projects.

The Relationships of Sunshine and Cloudiness

There are daily records of estimated cloud amount at some 150 stations over Iowa. It was reasoned that if a satisfactory relation could be found between sunshine percentage and cloud amount, sunshine percentage could be described areally on a weekly basis for relating to the condition of crops. Such a relationship has been found for Des Moines which agrees to a considerable extent with theoretical considerations. This study would have been completed but for the present unavailability of records for other large stations in Iowa where actual sunshine records are kept.

There is an important aspect of this problem which should be investigated as soon as assistance becomes available. It is known that photosynthesis takes place under illumination not necessarily produced by the sun. Further, it is known that illumination can be more intense under certain conditions when the sun is covered by clouds. How reliable then are percent sunshine records in relation to plant growth? Perhaps illumination should be measured instead of sunshine or perhaps the two are related in a usable manner.

Frost Occurrence in Iowa

Very extensive studies of the occurrence of frost in Iowa have been made since the beginning of the year. These were the outgrowth of an attempt to find weather factors which might be harbingers of the termination of frost in spring and the occurrence of first frost in autumn. No results in the form of useful relationships were found although the lack of relationship or the independence of dates of occurrence of spring and autumn frosts are useful. Thus, if the spring frost is late, this apparently gives no indication of whether the autumn frost will be late or early.

As soon as one statistical problem is resolved, extensive frost data will be prepared for publication in the form of risk isolines for agricultural operations. These may be interpreted to indicate the chances of damage from frost in spring or fall when planting is done on a certain date.

Because of the fact that killing frost (not measured by thermometer) is a localized and rather subjective observation, no close relationship was found between the date of killing frost and the date of occurrence of 32°. The frost statistics were, therefore, based on the occurrence of 32°. Here again physical experiments are necessary to learn more about Iowa conditions and should be undertaken as soon as assistance is available.

Climatological Normals

Methods have been devised for obtaining daily normals from monthly normals. These have already been applied with success at the Iowa Weather Bureau Section Center, Des Moines, Iowa. The general problem of normals is being studied, and it is hoped that a theoretical discussion of this problem can be completed in the near future.

The Agronomy Farm at Iowa State College now has some seven years of soil temperatures. These have been summarized and the preparation of normals for the various depths (especially the deeper ones where variation is small) is being studied. A physical analysis of the accompanying heat conduction and radiation problems for the several soil types under various conditions of tillage, moisture, etc., might result in the extrapolation by means of air temperatures of a relatively few and well placed soil temperature measurements.

MISCELLANEOUS ACTIVITIES

Cooperation with the North Carolina Agricultural Experiment Station

During the week of October 16-21, 1944, Mr. Snedecor lectured to a conference of 20 animal nutritionists assembed from the agricultural experiment stations of all the Southeastern states, from the United States Department of Agriculture Office of Experiment Stations and from Cornell University. The conference was devoted to the design and statistical analysis of experiments for research in animal nutrition.

Activities Related to Statistical Organizations

Mr. Cochran is a vice-president of the Institute of Mathematical Statistics and cooperates in the editing of the "Annals of Mathematical Statistics"; he is chairman of the Committee on Membership of the Institute and is chairman of the Committee on Post-war Development. Mr. Cochran is a vice-president of the American Statistical Association. He and Mr. Snedecor are members of the Biometrics Section Committee in the Association, and Mr. Snedecor is a member of the Committee on Sampling and edits a section of the "Biometrics Bulletin" entitled "Queries."

Participation in Public Discussion

These papers were read by staff members before scientific meetings during the past year. "The Master Sample of Agriculture: I. Development and Use" by A. J. King and "II. Design" by R. J. Jessen was read before the Amterican Statistical Association in Washington, D. C. A paper titled "The Statistical Laboratory at Ames" by G. W. Snedecor was read before the American Association for the Advancement of Science meeting in Cleveland, Ohio.

The Master Sample, its design and use was discussed by Jessen and King before staff members of the Bureau of Agricultural Economics, Washington, D. C., Department of Economics and Sociology, Iowa State College and staff members of the Bureau of Agricultural Economics at a regional meeting in Lincoln, Nebraska. King and Jessen led a discussion of modern sampling techniques and their application to agricultural economics research attended by members of the Department of Economics and Sociology, Iowa State College.

PUBLICATIONS

Brown, Bernice. A design for sampling Iowa families. Iowa Academy of Science, Proceedings 51: 476-482. 1944.

Cochran, W. G. "Student's" collected papers. Annals of Math. Stat. 15: 435-438. 1944.

Crump, S. Lee. A test of sampling error and bias in mailed inquiry samples. Bureau of Agri. Econ. C.R.P. #19. 1944.

Federer, Walter T. Studies on sample size and number of replications for guayule investigations. Jour. Amer. Soc. of Agron. Vol. 37. 1945.

Jessen, R. J. and Houseman, E. E. Statistical investigations of farm sample surveys taken in Iowa, Florida and California. Iowa Agri. Exp. Sta. Res. Bul. 329: 263-338. 1944.

Jessen, R. J. The Master Sample of Agriculture: Design. Jour. Amer. Stat. Assn. 40: 46-56. 1945.

King, A. J. The Master Sample of Agriculture: Development and use. Jour. Amer. Stat. Asson. 40: 38-45. 1945.

McCarty, Dale E. Sampling counties for United States estimates of agricultural items. Unpublished thesis. Library, Iowa State College, Ames, Iowa. 1944.

Snedecor, George W. and Breneman, W. R. A factorial experiment to learn the effects of 4 androgens injected into male chicks. Iowa State College Jour. of Sci. Vol. 19. 1945.

Snedecor, George W. and King, A. J. The Statistical Laboratory at Ames. The Agricultural Situation 29: 9-11. 1945.