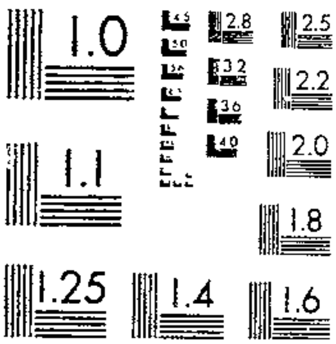
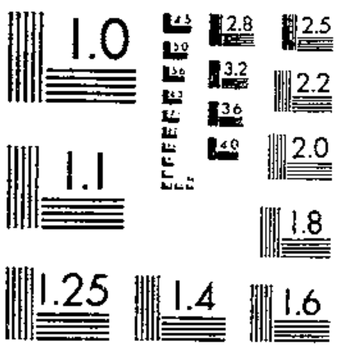
TB 814 (1942) USDA TECHNICAL BULLETINS UPDATA
AN OBJECTIVE METHOD OF SAMPLING WHEAT FIELDS TO ESTIMATE PRODUCTION AND
KING, A.J., MCCARTY, D.E., MCPEEK, M. 1 OF 1

START



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNITED STATES DEPARTMENT OF AGRICULTURE

NOTES TO ACCOMPANY TECHNICAL BULLETIN NO. 814, "An Objective Method of Sampling Wheat Fields to Estimate Production and Quality of Wheat"

Substitute "North Dakota Agricultural College" wherever the term "North Dakota State College" appears.

Page 23, par. 2: The formula should read:

$$\frac{1}{n}\left[\frac{V_b - V_w}{2} + \frac{V_w}{k}\right]$$

Where V_b = mean square among fields in the district

V_w = mean square within fields both mean squares being on a sample basis

Page 25, table 9, third column: Delete "Number."

Page 25, footnote 2: Change (i.e. 1/2 bu., 1/8 percent, or 1/8 lb.) to (i.e. 1 bu., 1/4 percent, 1/4 lb.) Make same change page 26, line 4.

Page 28, line 16: After the statement, "Where V(P) = variance of the estimate of production" insert "P = estimated production;"

Page 28, line 31: Substitute the following line in place of the line as given. "where r = the sampling rate (road miles between fields sampled)"

Page 29, last paragraph, first sentence: Delete the "r and" and change the word <u>are</u> to <u>is</u>.

(over)

Page 29, table 10: Change principal column headings to read:
"Wheat sampling rate [1]", "Total road mileage (M)", and "Fields sampled ($N = \frac{M}{r}$)." Also change footnote to read "Miles of wheat between samples."

The average number of miles of wheat instead of the average number of road miles between fields sampled (r) is presented in table 10. To obtain a solution for district 3, for example, by using the constants on page 30,

$$r = \sqrt{\frac{ad}{Kb}} = 68.81$$

Then, r multiplied by the ratio of miles of wheat metered in the district and total road miles equals the "wheat sampling rate."

$$r\left(\frac{\text{wheat miles}}{\text{road miles}}\right) = 68.81\left(\frac{1.224 \times 104}{895}\right) = 68.81 \times .1421 = 9.772$$

The calculated number of road miles to be driven is:

$$M = \frac{C}{\sqrt{\frac{aKb}{d}} + b} = \frac{C}{\frac{a}{r} + b} = \frac{278}{\frac{1.53}{68.81} + .0919} = 2437$$

from which the calculated number of fields is obtained:

$$\frac{M}{r} = \frac{2437}{68.81} = 35$$

The cost, C = 278, used in solving the equation to obtain M, contained some fixed costs that should have been excluded.

An appropriate value of C to use in district 3 is:

$$C = aN + bM = (1.53)\ (104) + (.0919)\ (895) = 241$$

The calculated road mileages and numbers of fields to be sampled, that appear in table 10, are somewhat large since some fixed costs were included in C when solving the equation to obtain M.

Sufficient data are not given in the bulletin for tracing the solution on the basis of one sample per field. The method, however, is the same, the only change being the use of different values for a and K.

Technical Bulletin No. 814 • February 1942

UNITED STATES DEPARTMENT OF AGRICULTURE
WASHINGTON, D. C.

An Objective Method of Sampling Wheat Fields to Estimate Production and Quality of Wheat

By Arnold J. King, *agricultural statistician*, Dale E. McCarty, *agent*, and Miles McPeek, *agricultural statistician, Agricultural Marketing Service* [1]

United States Department of Agriculture, Agricultural Marketing Service, in cooperation with the Kansas, North Dakota, and Iowa Agricultural Experiment Stations

CONTENTS

INTRODUCTION

For many years, the United States Department of Agriculture has prepared and published information relating to crop production in the United States with special emphasis upon the so-called speculative crops—wheat, cotton, oats, and corn. The Government crop reports, prepared by an independent agency interested in presenting the most reliable information possible, have been made available to both buyers and sellers of agricultural products. These reports, which include forecasts of the probable size of a crop prior to harvest, have met an important need for information as to the volume of crop production.

The importance of quality in the marketing of wheat has created an additional demand for information on the quality of the crop by small geographic areas such as counties. At present little information on quality of the crop is available until wheat reaches the terminal markets. As much of the wheat crop leaves the farmers' hands and moves to market soon after harvest, quality and production

[1] This study was made by the Agricultural Statistics Division when it was a part of the Agricultural Marketing Service. The Division was transferred to the Bureau of Agricultural Economics, Feb. 23, 1942.

327602° 42 —1

information alike should be available at or prior to harvest to be of maximum value to producers and millers. With such information, producers of high quality wheat would be better able to bargain for higher prices that such grain is entitled to, and millers would be better informed regarding the location of the various quality types. By quality is meant chiefly protein content and test weight (pounds of wheat per measured bushel) which are considered to be the best indicators of baking quality and milling quality, respectively.[2] (*6*) As a result of the importance of these quality indicators, changes have taken place in trade practices of evaluating wheat (*1*).

In recognition of the growing need for quality information and with the hope of discovering methods of improving the existing estimates of yield and acreage, the Agricultural Marketing Service with the cooperation of several agencies[3] conducted preharvest wheat surveys during 1938, 1939, and 1940. It is the purpose of this bulletin to report the results obtained from the 1939 and 1940 surveys. These surveys were operated as projects in objective sampling which also provided for currently reporting timely information relating to the wheat crop. The results of the analysis of the 1938 survey have been reported by King and Jebe (*5*).

The first section of this bulletin describes the technique employed in the field, laboratory operation, and the method of publication of the information. The second section presents an appraisal of the survey from the standpoint of its efficiency in estimating quality and production. The third section shows the results of a study designed to evaluate several attributes of yield and the possibilities of successfully forecasting yield therefrom.

DESCRIPTION OF THE 1939 AND 1940 PREHARVEST WHEAT SURVEYS

OBJECTIVES

The objectives of the surveys were:

1. To provide at harvest and prior to marketing accurate, timely information on the quality of wheat by small areas such as counties or groups of counties.
2. To obtain by objective sampling methods an indication of the acreage of wheat for harvest and the yield per acre.
3. To furnish current information on the distribution and acreage of varieties of wheat.
4. To determine the utility of certain attributes of yield per acre for forecasting purposes.

AREAS INCLUDED IN THE SURVEYS

The 1939 survey covered Kansas, North Dakota, and parts of Oklahoma, Nebraska, South Dakota, and Minnesota. The variation in the time of harvest in the Wheat Belt made it possible to begin sampling in southwestern Oklahoma about June 1 and to proceed northward at the rate of about 100 miles a week.

The 1940 survey covered Kansas and the western half of Oklahoma only. Field work in each of the 2 years was done by three crews

[2] Numbers in parentheses refer to Literature Cited, p. 76.
[3] See the preface for a list of cooperating agencies.

of two men each. (Four crews were used during the last 10 days of June 1940.) The crews traveled in automobiles equipped with crop meters [4] for recording the frontage of wheat on each side of the road and the total mileage. Figures 1 to 8, inclusive, show the sampling routes followed in 1939 and 1940.

The sampling, except in Kansas and North Dakota, covered only the more important wheat-producing areas of the States involved. The extent of coverage was affected by the time available for sampling. In 1939, wheat ripened in Nebraska and the southern part of South Dakota so soon after wheat ripened in Kansas that there was not enough time available for sampling all the important wheat counties in Nebraska.

Sampling Procedure

Sampling of wheat in the field to obtain essential data involved certain difficulties. It was necessary to have a practical plan for obtaining samples of the wheat at maturity over extensive areas. Uneven ripening of wheat owing to differences in varietal characteristics, cultural practices, and vagaries of the weather, therefore, presented certain difficulties in field operation, particularly in 1940. Since it was necessary to take the samples before harvest, there was only a relatively short period of about a week in which sampling could be done in an area. Because uneven ripening was prevalent to an unusual degree in Oklahoma and Kansas in 1940, problems in practical operation of the sampling plan were rather perplexing. In a few instances expediency necessitated the taking of somewhat immature grain in some of the fields sampled in order that the crew could finish sampling in a particular county and proceed with the work in other counties.

Another problem involved the manner of selecting the fields to be sampled together with the location of the points within a field from which samples were to be taken. From the standpoint of operation, random selection of fields was not practical. Apparently, some method of route sampling was necessary. The 1938 preharvest wheat survey demonstrated that route sampling of wheat to estimate and forecast yield per acre was a practical and an efficient method (5). Analysis of route sampling by Hendricks [5] indicated that the ratio of crop frontage to total frontage as shown by the crop meter could be successfully expanded to give a reliable estimate of total acreage of the crop. Thus, route sampling with the aid of a crop meter permitted the gathering of samples in a relatively short time and also provided a simple method of estimating acreage for harvest.

Sampling procedure was essentially the same in 1939 and 1940, except for slight improvements in operation in 1940 which were made as the result of experience in 1939. Depending upon the density of wheat acreage in the several counties, samples were taken at

[4] A crop meter is an instrument for recording frontage of crops along highways. It has a number of dials each of which may be engaged and disengaged independently of the others by means of separate control buttons. The instrument is connected to the drive shaft of an automobile in a manner similar to that used for connecting a speedometer. Developed in 1928 by D. A. McCandliss now of the Agricultural Marketing Service, the crop meter has been used since that time to obtain indications of year to year changes in crop acreages. Its use in regular crop reporting work involves the recording separately of the mileages of various crops that front on highways.

[5] Hendricks, Walter A. (A study as yet unpublished.)

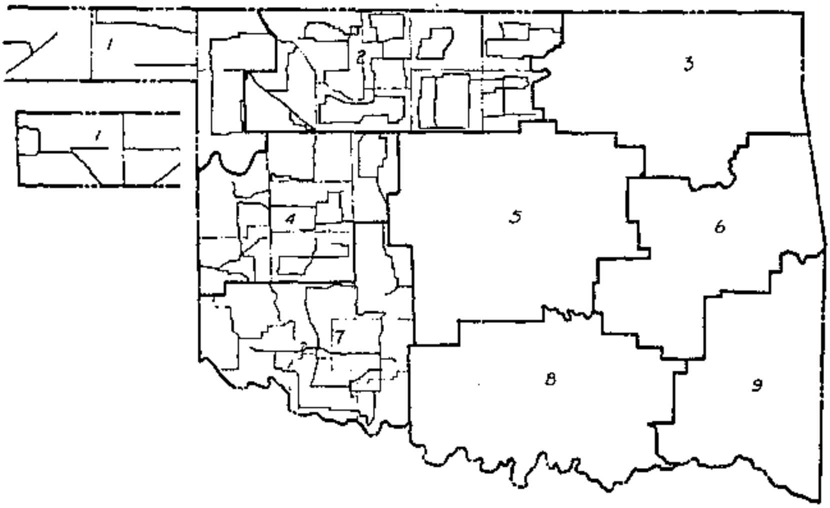


FIGURE 1.—Oklahoma. Preharvest wheat survey sampling routes, 1939.

intervals of 25 to 200 crop meter units of wheat on the right-hand side of the road. (One crop meter unit is equal to $\frac{1}{50}$ mile.) This interval by counties was determined in advance on the basis of the ratio of the preliminary official estimates[6] of wheat acreage to total land area in the county. Wheat acreage on both sides of the road was metered to obtain an indication of the acreage for harvest.

FIELDS SAMPLED

a. Upon entering a county, the first field to be sampled was the field on the right-hand side of the road indicated by the crop meter. The sampling interval used for selecting this first field was that in

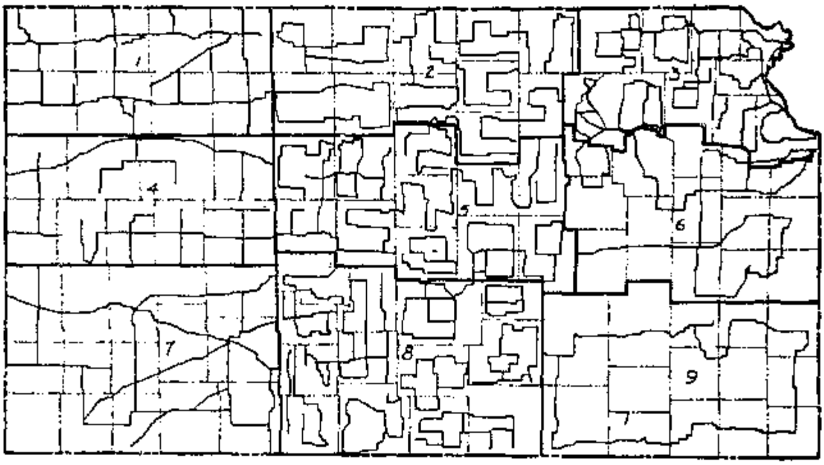


FIGURE 2. Kansas. Preharvest wheat survey sampling routes, 1939.

[6] Estimates prepared by the Agricultural Statistics Division.

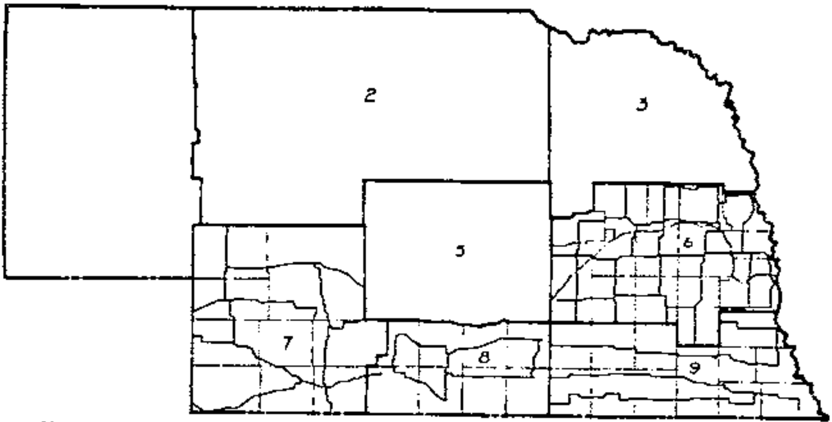


FIGURE 3.- Nebraska. Preharvest wheat survey sampling routes, 1939.

use in the preceding county. The second and following fields sampled in the county were similarly selected by using the sampling interval previously determined for that county.

b. A field was not sampled unless the grain had reached the semihard dough stage. If wheat in a field selected for sampling was not that far developed, the field (if there was one) on the opposite side of the road was sampled. If there was no wheat on the opposite side, the next field encountered on either side was sampled.

POINTS OF ENTRY TO FIELD, LOCATION OF SAMPLING UNITS

Two samples were taken from each field selected. The points at which these samples were taken were located objectively as follows: The car was stopped when the crop meter indicated the mileage at

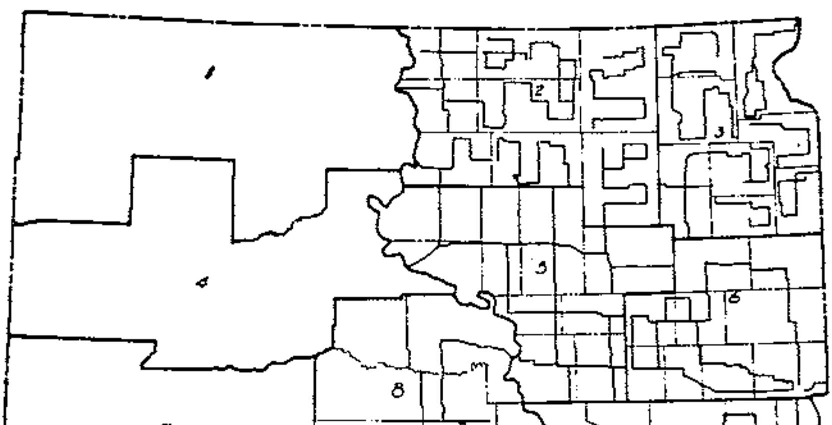


FIGURE 4.—South Dakota. Preharvest wheat survey sampling routes, 1939.

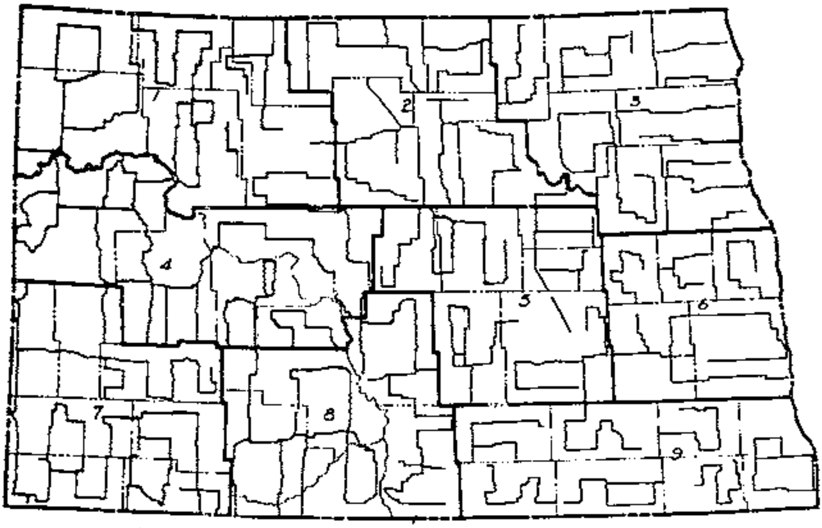


FIGURE 5.—North Dakota. Preharvest wheat survey sampling routes, 1939.

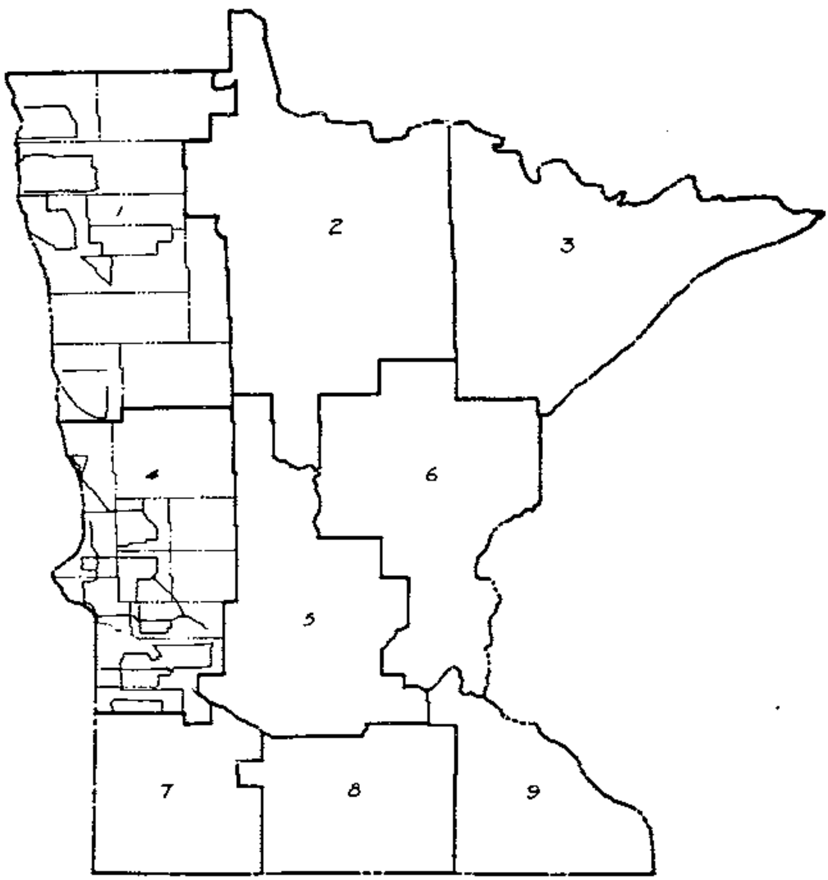


FIGURE 6.—Minnesota. Preharvest wheat survey sampling routes, 1939.

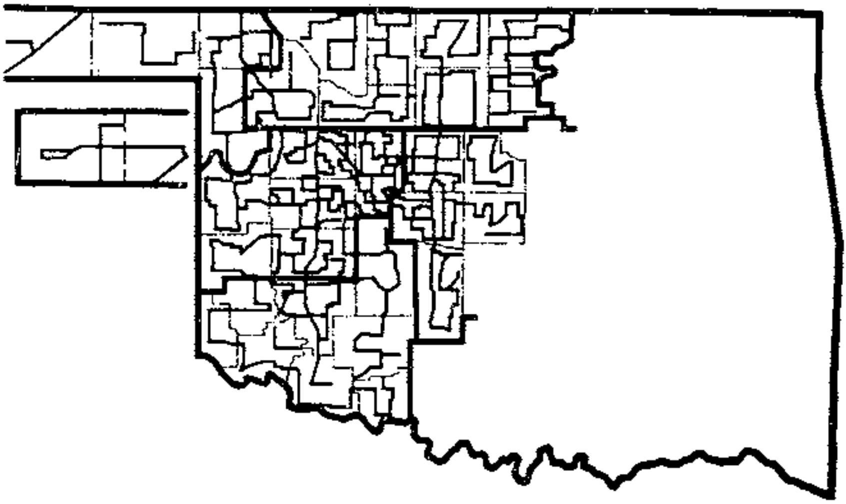


FIGURE 7.—Oklahoma. Preharvest wheat survey sampling routes, 1940.

which a sample should be taken. The two samplers walked from 1 to 100 steps in opposite directions from the car to points of entry to the field. They then walked from 1 to 100 steps into the wheat field to the points where the samples were taken. The two numbers between 1 and 100 used to determine the distances walked from the car and the two numbers used to determine the distances walked into the field had been previously selected from a table of random numbers and entered on the envelopes used for holding the wheat heads that made up the sample. In cases where the car was stopped near one end of a field, it sometimes was necessary to divide the random number involved by (say) 2, 3, 4, or 5. Likewise, the random numbers that determined the distance the samplers were to go into a field were reduced (by division) if the field was very narrow.

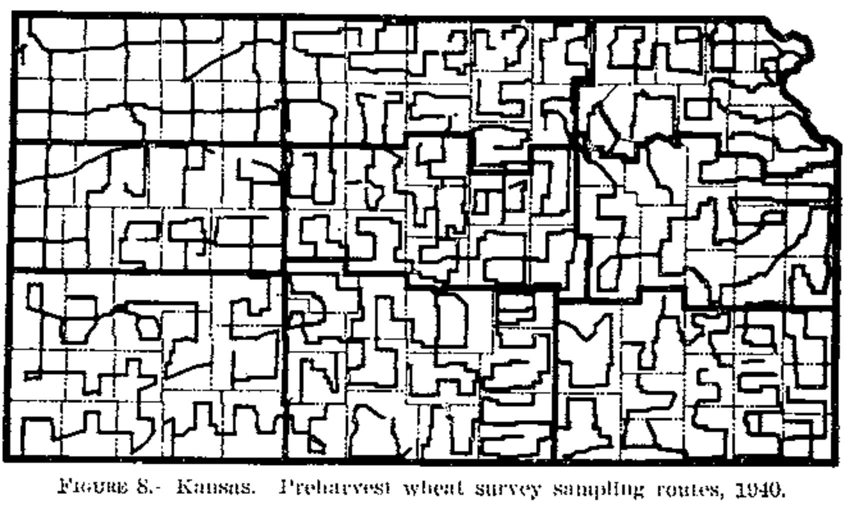


FIGURE 8.—Kansas. Preharvest wheat survey sampling routes, 1940.

SAMPLING UNIT

Findings of Yates and Zacopanay (7) made it appear desirable to use a rectangular shaped area that included sections of a number of drill rows as a sampling unit. Accordingly, this general idea was followed in adopting the sampling unit used in the preharvest wheat surveys. Due to varied widths of drill rows found in commercial wheat fields, however, the sampling unit used in the surveys was not a fixed rectangular area with inflexible dimensions.[7] Instead, the sampling unit consisted of a certain number of contiguous rows, 26.136 inches in length, the number of rows being governed by the distance between drill rows. The number of rows included in a sample together with the area of the sample for given drill-row widths is shown in table 1.

TABLE 1.—*Area of sample taken at different widths of row*

Width of drill row	Rows included in sample	Area of sample
Inches	*Number*	*Acres*
6	8	1/5000
7	8	1/4286
8	6	1/5000
10	6	1/4000
12	4	1/5000
14	4	1/4286

Preparatory to taking a sample a rigid, rectangular-shaped steel frame open on one side was placed on the ground immediately beyond the sampler's foot at the completion of the last step taken to locate the sampling point. This sampling frame was placed in position by slipping the open side through the grain perpendicular to the drill rows. The frame, about 24 by 26 inches, was used to control the length of the rows of wheat included in the sample. (Fig. 9.) After placing the sampling frame in position the sampler measured the distance between drill rows to determine the number of rows of wheat to be included in the sample. The sample heads were clipped and put into a 9½ by 12 inch envelope for transmittal to the laboratory. Heads from the usual size sample of approximately 1/5,000 acre would more than fill the envelope if the field being sampled had unusually heavy, high yielding wheat. Consequently, in such cases the number of drill rows included in the sample was reduced by one-half. The decision as to whether the number of drill rows was to be reduced by one-half was made from general observation of the field while walking to the point where a sample was to be taken rather than being based on observation of the wheat in the sampling area. This was done to avoid the effects of a possible bias in the sample. The number of drill rows included in a sample was recorded on the envelope used as a container for the sample heads.

[7] In the few cases where no drill rows could be distinguished, a fixed area outlined by the sampling frame was used as the sampling unit.

Yields for each sample were calculated separately by applying the conversion factor appropriate to the size of the sampling unit used.

DETERMINATIONS MADE AT THE TIME THE SAMPLE WAS TAKEN

a. *Drill row width*.—Several measurements were often necessary to determine width of drill row. By using hand signals the two samplers agreed on the drill row width before they started to clip off any wheat heads.

b. *Average height* was obtained by measuring a handful of culms selected objectively. Height was the distance from the ground to the tip of the heads, excluding the beards.

c. *Grasshopper damage* was determined by counting the heads that appeared to have been eaten off by grasshoppers within the area sampled.

d. *Rust infestation.—Stem rust* infestation was estimated as a percentage of the total culm surface covered by pustules. This,

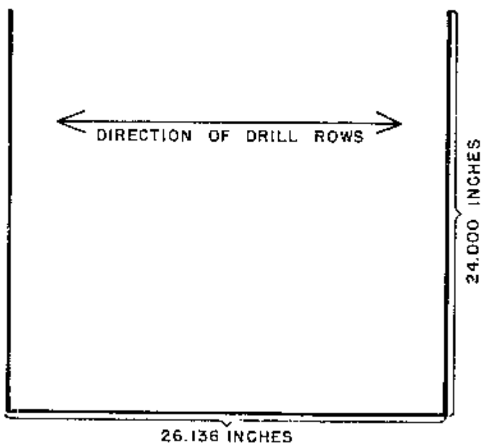


FIGURE 9.—Illustration of sampling frame used to control length of rows of wheat included in sample.

probably, gave only a very rough indication of the extent of damage because other factors, such as stage of development of wheat when rust infestation sets in, kind of weather, amount of available soil moisture, etc., are very important in governing the extent of injury from stem rust.

Leaf rust was estimated as "slightly infected" or "badly infected."

e. *Lodging* was estimated as a percentage of total wheat in the sample.

RECORDING OF DATA

A form for recording data was stamped on the back of each envelope used for holding the sample heads. The laboratory address also was stamped on each envelope, before the sampling was started. Each night the name of the State and county, a field number, date, and random numbers used to determine the points at which samples were to be taken, were entered on a set of envelopes for each county to be sampled the following day. When the car was stopped to take a sample, three crop meter readings (left side wheat dial, right side wheat dial, total route mileage dial) and the time of day were recorded on the envelopes. The samplers were then ready to enter the field to make various measurements and determinations to be recorded on the envelopes. The time of day when the car was started after taking the sample was also recorded. In addition to the detailed information recorded on each envelope, the crop meter readings at the beginning and end of sampling in each county were entered on a separate sheet. The sampling interval, number of samples taken, and any pertinent comments relating to the sampling or to the wheat crop in the county were also shown. The routes covered in each county were shown on a road map. This map and the record sheet were sent to the laboratory after sampling had been completed in a county. The field crew retained copies of the route map and a county summary sheet.

MAILING OF SAMPLES TO LABORATORY

Samples were mailed to the laboratory usually twice each day. Where feasible the larger towns on main railroad lines were selected as points from which samples were mailed.

SUPPLEMENTAL SAMPLES

In addition to the regular measured samples, an equal number of small supplemental samples from other fields were taken in 1940 to give a larger number of samples to be used to estimate the distribution of wheat varieties. Supplemental samples of this kind were not taken in 1939. A supplemental sample included a handful of heads clipped from two or more plants near the edge of a field. In obtaining these samples it was necessary to clip heads from at least two plants because a sample from only one plant would give misleading results in case two or more varieties (mixed) were growing in the field. If near the field selected for regular sampling, there was another wheat field, it was sampled so that both the regular and supplemental samples could be obtained at one stop. Otherwise the supplemental sample was taken at a point between two regular sampling points. In view of the widespread practice among wheat farmers (at least in Oklahoma and Kansas) of growing more than one variety of wheat on the same farm, it was believed that the procedure, whereby regular and supplemental samples were oftentimes obtained during a single stop, would inject but little bias into the survey even if the two samples should be taken from different fields on the same farm. Moreover, in many cases the two samples, being obtained on opposite sides of the road, came from different farms. The person who returned to the car first after taking a regular sample, took the supplemental sample. Thus, the picking up of an extra sample for determining variety actually took but little additional time. The supplemental samples were put in small envelopes which carried the name of the State and county and the same serial number as the regular sample taken nearby.

WHEAT ACREAGE

The wheat acreage that was metered and sampled included only that which the crew believed would be harvested. In 1940, however, the decision as to whether a field in western Kansas would be harvested was in some cases difficult to make. In some instances the samplers would start metering a field and then stop before coming to the end of it because the "second guess" was that it was too poor to harvest. Likewise, it was sometimes decided to meter the latter part of a field after rejecting part of it. Some of this more or less haphazard measuring was due to the crop's being better in some parts of the field than in other parts; although more of it was due to the fact that some of the wheat appeared to be on the border line between that which would be harvested and that which would not. Where it appeared that a harvesting machine would be pulled over only the good spots in a field, the extent of the metering included only those spots.

FRONTAGE

"Frontage" of wheat was recorded on the crop meter when wheat for harvest appeared in a field opposite the car and at a point about 70 yards beyond the shoulder of the road. In other words, it was "read" on a line parallel to the road and about 70 yards in. Wheat frontage on both sides of the road was recorded for use in preparing acreage estimates whereas the frontage on the right side only was used to determine the fields to be sampled. It was felt that a frontage line 70 yards in from the road would be far enough in to reach fields beyond lanes, garden patches, etc., and yet would be close enough so that crops could be easily identified.

In case a railroad right-of-way was near, and parallel to, the sampling route being traveled, both wheat frontage dials were engaged when there was wheat frontage on the side of the highway opposite that on which the railroad was located. This seemed to be a better procedure than any of the following alternatives, since wheat frontage on each side of most roads tends to be about equal:

a. To record frontage of wheat beyond a right-of-way would be likely to inject errors as a result of obstructed view and increased distance to point where wheat would be "read."

b. To check out the strip as "no route," that is disengage the route dial and stop recording wheat frontage would be rather wasteful in areas where considerable highway and railroad mileages are located side by side.

c. To ignore the acreage beyond the railroad tracks in such instances would result in the indication of wheat acreage being biased downward.

If, while recording the frontage of a wheat field, the "frontage line" cut across a farmstead, grove, ravine, patch of waste land, or anything other than wheat, the wheat dial was disengaged until the "frontage line" again reached wheat. Total length of sampling routes in each county was obtained from the readings on the "free" dial which could be engaged and disengaged independently of the other dials on the crop meter. It was kept engaged while traveling over a sampling route and then disengaged when "back tracking" or doing any other driving while not sampling.

ROUTES

Selection of sampling routes in each county was made by the crew. The routes covered a county in a grid-like pattern insofar as permitted by available highways, time, and location of the county to be sampled next. This grid-like coverage was designed to give proper representation to the various soil and topographical sections within the county. Such representation seemed essential in sampling wheat yield and quality as well as wheat frontage for use in estimating acreage. The total frontage in a county was proportional to total area to the extent that practical operating considerations permitted.

WHEAT FIELD

Any continuous area of one variety of wheat having similar soil treatment was called a wheat field. In case of strip cropping, the entire tract was considered as a field, provided the strips of wheat were all of the same variety.

Laboratory Procedure

In order to conduct detailed analyses of the head samples, it was necessary to establish laboratories at centralized points to which the field men could send the samples. For the 1939 survey, two central laboratories were established. One was set up at Manhattan, Kans., in cooperation with the Department of Milling and the Department of Agronomy of Kansas State College. This laboratory served the three winter wheat States included in the survey—Oklahoma, Kansas, and Nebraska. After sampling was completed in this area, a portion of the equipment and personnel was moved to Fargo, N. Dak., under the direction of J. E. Pallesen, who was in charge of the laboratory work. At Fargo, a second laboratory was set up in cooperation with the Departments of Milling and Agronomy of the North Dakota Agricultural College. This laboratory made the analysis of the spring wheat samples from North Dakota, South Dakota, and western Minnesota. As the 1940 survey was confined to the two winter wheat States—Kansas and Oklahoma—the one laboratory at Manhattan was adequate to handle all samples.

The head samples were obtained twice daily from the post office by one of the laboratory personnel. It was quite important that the samples be removed from the bags as soon as possible, since some were taken under damp field conditions and the tight mail bags afforded the samples little opportunity for drying out in transit.

Upon arrival at the laboratory, the envelopes containing the samples were opened and arranged by counties. Thus the wheat samples could dry readily and were handy for analysis when the laboratory crew was ready for them. The next operation consisted of the counting and measuring of heads. In the 1939 survey all samples were counted and measured for length of head. Because a larger volume of samples was handled in Oklahoma and Kansas in 1940, it was necessary to limit the counting and measurement of heads to the even numbered fields in most counties. Otherwise, it would not have been possible to maintain timeliness of the releases. One laboratory assistant did most of the counting and measuring. The number and length of heads were recorded on the envelope which had contained the samples.

After they were counted and measured, the head samples were identified as to variety by two graduate agronomists. If time permitted, each agronomist identified the samples independently of the other and then compared results. If disagreement occurred, the questionable samples were taken to members of the agronomy department and checked for variety. The disagreements were relatively infrequent, however, and occurred mostly with samples of mixed wheat which were grouped together in the reports.

Next, the head samples were weighed on a large pan grain balance, the weight being recorded on the sample envelope. After weighing, the samples were stacked to await their turn to be threshed. The grain was threshed in a small head thresher built by the Agricultural Engineering Department of Kansas State College. In addition to running the heads through the thresher, it was necessary to winnow the chaff from the grain by "panning" the mixed chaff and grain before an electric fan. The cleaned grain was then weighed and placed in small paper bags and labeled to show State, county, field number, sample number, and weight of grain.

Weight per kernel was obtained by taking two 500-kernel counts on each grain sample. The two groups of kernels were each weighed on the torsion balance used for obtaining the test weight of samples. The kernel counter used was designed by Dr. H. H. Laude, Department of Agronomy, Kansas State College. The principal feature of the counter is a brass plate with 500 kernel-size holes. The brass plate is supported on a glass plate under which an electric light is located. A quantity of grain was placed on the brass plate and as the surplus grain was brushed off, a kernel was left in each hole. Then the glass plate was pulled away from beneath the grain, and the 500 kernels were permitted to flow into a "nozzle pan" at the rear of the counter.

The next operation was to measure the test weight of the sample. The small size of the samples made it necessary to use a micro method of determining test weight. A 16 cc. sample of wheat was weighed, the weight in grams being converted to pounds per bushel through the use of a constant conversion factor. The reliability of the micro method will be discussed later. The agronomist who analyzed the sample for test weight also estimated the character of the sample, indicating the proportion of dark, hard, vitreous kernels. This completed the operations performed on the sample before it was taken to the milling laboratory for further analysis. A separate mimeographed form sheet was prepared on which the laboratory and field data for each field were recorded.

In the milling laboratory each sample was analyzed for moisture content, by the use of a Tag-Heppenstall moisture meter according to instructions issued by the Department of Agriculture (*3*). Readings from the meter were converted to moisture content in percent by the use of the moisture tables in the same handbook.

Next the samples were ground in a Hobart coffee mill equipped with special wheat burrs. In the grinding procedure, part of each sample was run through the mill first to clean the grinder. This portion was then discarded and the remainder of the sample was saved for protein analysis. Crude protein was determined by a modified Kjeldahl-Gunning-Arnold method.

After the protein content, variety, etc., had been determined and recorded for each field, the data were ready to be summarized and prepared as a press release.

Publication of Information

The 1939 and 1940 surveys were conducted with the dual objectives of supplying farmers and the grain trade with wheat quality information as well as continuing the research phase of the project. Therefore, it was necessary to adopt means whereby the survey information on quality could be currently released to the public. In 1939, only the "flash" type of report was issued. The 12 flash reports that were released between June 17 and August 17 of that year reported quality by groups of 10 to 20 contiguous counties each. As soon as the laboratory analysis was completed for such a group of counties, the means and frequency distributions of the various quality items were computed. Then a summary of these data was prepared in the form of a flash release which also included interpretative comments and a report of additional observations made by the field crews.

After the release was written, it was rushed to a field office of the Agricultural Marketing Service for mimeographing and mailing. The releases for the winter wheat States in 1939 were handled through the Kansas City, Mo., Market News Office under the direction of Don E. Rogers, Local Representative. The spring wheat reports for 1939 were released from the Minneapolis office, by W. R. Kuehn, Local Representative. These two offices distributed the reports to their own mailing lists which consist chiefly of mills and elevators, farmers, newspapers, banks, and agricultural equipment dealers. An example of a 1939 release is shown in the appendix (exhibit 1).

In 1940, one objective was to demonstrate further the type and amount of information that could be obtained from the preharvest wheat survey and released currently. Therefore, several changes were made in the content, frequency, and distribution of the flash releases. In regard to content, the number of samples taken per county in 1940 was considered adequate to justify the release of information on a county basis. It was not considered necessary to wait until the samples from a group of 10 or 15 counties had been mailed in and analyzed before a report could be prepared and released. Instead, it was thought that data by counties could be released more frequently with fewer counties reported in each release than was possible under the procedure followed in 1939.

Fourteen flash reports were released in 1940 between June 13 and July 15 for the two States, Kansas and Oklahoma. The average amount of time between the date the sample was taken in the field and the date the information was released in mimeographed form was 6.4 days in Kansas in 1940. It would have been possible to reduce this period if less time in the laboratory had been spent in obtaining data for research. At an early stage of the survey, however, it is quite important to obtain necessary data to study the sampling errors of the means released to the public, and to investigate allocation problems. An example of a 1940 flash report is shown in the appendix (exhibit 2).

The flash reports were circulated more extensively in 1940 than in 1939. In addition to the circulation given them by the Market News

Service, the 1940 reports were also released by the Agricultural Marketing Service through the offices of the Agricultural Statisticians for Oklahoma and Kansas. The reports relating to the Oklahoma wheat crop were furnished to a list composed of about 3,400 names whereas those for Kansas were sent to a list of about 8,500 names. Included in the lists were substantial numbers of newspapers, radio broadcasting stations, trade journals, extension workers, and county agents. Thus, the timely information contained in the reports was quickly disseminated to the general public by press and radio facilities. Also included in the lists of names to which the reports were sent direct were farmers, operators of mills and elevators, bankers, and other interested persons who requested them.

In addition to the flash reports, State summaries were issued at the completion of sampling in Oklahoma and Kansas in 1940. Through the cooperation of the Agricultural Statistician for Kansas, State summary reports for the two States were compiled and released to present a combined account of the survey results in each State. Tables were constructed to show means by county and district and graphs and maps were used to delineate further the quality of the season's wheat crop as disclosed by the survey. The Kansas summary is reproduced in the appendix (exhibit 3). Attention is called to the maps that show the district and State frequency distributions of protein and test weight. These frequencies have been designed to indicate the quantity of wheat that exhibits a specific quality characteristic in each district. In other words, the total area of all histograms in the State represents total production, and the total area of a district histogram represents the relative production in that district as a proportion of the State production. Thus, the total production of wheat for the State is catalogued by quality within each district.

EFFICIENCY OF THE SAMPLE

In evaluating the success with which a sampling investigation has been conducted, two important questions need to be answered. First, how could the sampling have been conducted in order to permit an estimate of greater precision? Here consideration rests entirely with the statistical efficiency of the sample. But a second and more useful evaluation needs to be made in terms of cost. This second approach requires an estimate of the manner in which the total funds could have been allocated in order to have obtained a maximum amount of information for each dollar spent. Detailed statistical analysis of the preharvest wheat surveys has provided information upon which the sampling efficiency could be appraised from both the statistical and cost viewpoints.

Statistical Efficiency

STRATIFICATION

In sampling a heterogeneous population, the pattern of variability is sometimes such that the population can be subdivided into more homogeneous groups. This subdivision of a population for sampling purposes is commonly called stratification. In obtaining a random sample from a stratified population, a separate random sample is taken in each stratum, rather than a single random sample over the entire population as is done in unstratified sampling. The estimate of the mean for the entire population is made by weighting the mean per stratum by the total number of units in each stratum. The gain in precision that can be obtained through stratification depends chiefly upon the differences among strata, the precision of the estimates of the weights per stratum (when the weights are not definitely known), and the distribution of the sample among strata.

Previous observations indicated that the acreage, yield, and protein content of wheat differ among crop reporting districts.[8] Therefore, the sample was stratified by districts in order to take advantage of information at hand. From the results of a stratified sample, the analysis of variance technique permits us to estimate the sampling variance of the mean of an unstratified sample distributed at random over the whole State (*2*). By comparing this variance with the corresponding variance for the mean of the stratified sample which was actually taken, the increase in precision due to stratification may be estimated.

After obtaining mean squares for both the stratified and random sample, the percentage of gain in information due to stratification of the sample is calculated as follows:

$$\text{Percent gain} = 100(V_r - V_s)/V_s$$

where V_r = estimated variance of means for a nonstratified sample;
V_s = variance of the weighted mean of a stratified sample, assuming weights are accurate.

At this stage of the analysis it is assumed that the weights are estimated without error. Table 2 shows the gains due to stratification by crop reporting districts in 1939 and 1940.

TABLE 2.—*Percentage of gain in information due to stratification by crop reporting districts, assuming no error in weights*[1]

State	Yield		Protein		Test weight	
	1939	1940	1939	1940	1939	1940
	Percent	*Percent*	*Percent*	*Percent*	*Percent*	*Percent*
Oklahoma	75.0	1.3	33.5	11.1	10.7	22.8
Kansas	19.9	27.5	−9.7	6.8	−9.8	−4.0
Nebraska	20.5	------	−.2	------	27.7	------
South Dakota	28.4	------	−6.6	------	−11.2	------
North Dakota	−.2	------	−6.1	------	−.6	------

[1] Percent gain $= \frac{100(V_r - V_s)}{V_s}$, where V_r = estimated variance of means for a nonstratified sample; V_s = variance of weighted mean.

[8] For the purpose of stratifying data furnished by crop reporters, the Agricultural Marketing Service has divided each State into groups of counties which are known as crop-reporting districts. Figures 1 to 8, inclusive, show the outlines of crop reporting districts in the States included in the preharvest wheat surveys.

It can be seen from table 2, assuming that there are no errors in weights used to combine district means into a State estimate, that for yield a gain would likely result from district stratification in all States sampled in 1939 except North Dakota. The table, however, indicates that in both 1939 and 1940 the actual distribution of the sample was such that in some cases losses rather than gains occurred from stratification. This loss in efficiency due to stratification will be discussed more fully later.

Actually, the weights employed in the computing of a State estimate were subject to error. Therefore, the actual gains due to stratification were even less than the previous paragraph indicates. It is necessary now to compute another variance of weighted mean yield allowing for errors in weights. The approximate increase in variance of a weighted mean due to errors in weights has been calculated as follows (2):

$$\text{Increase in variance, } I_v^2 = \frac{\Sigma(\bar{x}_i - \bar{x}_w)^2 V_{wi}}{(\Sigma_{wi})^2}$$

where $\bar{x}_i$ is the mean of the ith stratum.

x_w is the weighted mean;

w_i is the weight (acreage in this case);

V_{wi} is the variance of w_i.

In applying this formula, it was necessary to calculate the sampling error of the estimate of the acreage of wheat in each district. The wheat acreage used for weighting purposes was estimated from crop meter records by using the following formula:

$$\text{Total land in district (acres)} \times \frac{\text{(Miles of wheat metered)}}{\text{(Total miles driven)}}$$

The sampling error of the ratio of wheat mileage to route mileage was estimated from the crop meter records. There are some objections to this procedure since the route followed was not strictly a "random" one, but it appeared to be the most satisfactory method available. The percent gains due to stratification, assuming no error in weights and adjusted for errors in weights due to sampling variation, are shown in table 3.

Table 3 indicates that the errors in weights resulted in a marked reduction in the percent of gain due to district stratification. The table does not present a strong argument in favor of district stratification, in view of the negative gains and the increase in variability of the weighted means due to errors in weights. After admitting that the distribution of the sample did not permit full advantage of stratification, consideration should be given to the sample distribution which is necessary in order to take advantage of possible gains through stratification.

---

[2] In Cochran's formula "Ni" appears in place of "wi" used here. The formula applies even though the weights used may not be estimates of the total number of sampling units in each stratum, if they are estimated independently.

TABLE 3.—*Percentage of gain in information due to stratification by crop reporting districts, (1) assuming no error in weights and (2) adjusted for error in weights due to sampling variation*

State	No error in[1] weights		With error in[2] weights	
	1939	1940	1939	1940
YIELD				
	Percent	*Percent*	*Percent*	*Percent*
Oklahoma	75. 0	−1. 3	25. 2	−27. 1
Kansas	19. 9	27. 5	1. 3	−17. 8
Nebraska	20. 5		15. 0	
South Dakota	28. 4		9. 4	
North Dakota	−. 2		−. 9	
PROTEIN				
Oklahoma	33. 5	11. 1	1. 4	−29. 0
Kansas	−9. 7	6. 8	29. 0	−34. 1
Nebraska	−. 2		−3. 8	
South Dakota	−6. 6		−10. 4	
North Dakota	−6. 1		−10. 7	
TEST WEIGHT				
Oklahoma	10. 7	22. 8	6. 0	−4. 0
Kansas	−9. 8	−4. 0	−18. 1	−30. 0
Nebraska	27. 7		−17. 9	
South Dakota	−11. 2		−25. 7	
North Dakota	−. 6		−6. 8	

[1] Percent gain $=100\ (V_r-V_s)/V_s$, and
[2] Percent gain $=100\ (V_r-V_s-I)/(V_s+I)$
where V_r = estimated variance of means for a nonstratified sample;
V_s = variance of the weighted mean assuming weights accurate;
I = increase in variance of weighted mean due to errors in weights.

EFFICIENCY OF DISTRIBUTION

In a stratified sample the most efficient distribution (from the statistical standpoint) is that in which the number of samples taken per stratum is proportional to the product of weight for the stratum and the standard deviation. As the preharvest sample was not distributed according to the optimum criteria, it is desirable now to investigate the actual distribution of the sample, as it approximates the optimum. If optimum distribution of the sample is assumed, the variance of mean yield can be calculated as follows:

$$V_o=\frac{1}{N}\frac{(\Sigma w_i\sigma_i)^2}{(\Sigma w_i)^2}$$

where V_o = variance of weighted mean with optimum distribution of sample;
N = the number of samples taken in the State;
w_i = the weight per stratum;
σ_i = the stanard deviation per stratum.

In applying this formula, the values of w_i and σ_i found in the sample were assumed to be the true population values. The optimum mean squares were calculated and compared with the mean square of a random sample to evaluate gains possible through district stratification, provided the sample were distributed in the optimum manner given above. Table 4 shows these gains, which provide an upper limit to the gains attainable by district stratification. It will be noted from the table that all the gains from stratification are positive when the sample is distributed among the strata in optimum proportions.

TABLE 4.—*Percentage of gain in information due to stratification by crop reporting districts, assuming the sample is distributed among strata in optimum proportions*[1]

State	Yield		Protein		Test weight	
	1939	1940	1939	1940	1939	1940
	Percent	*Percent*	*Percent*	*Percent*	*Percent*	*Percent*
Oklahoma	82.3	16.6	44.0	30.1	16.3	33.8
Kansas	40.7	70.7	29.1	75.3	19.4	45.5
Nebraska	22.2		5.6		36.3	
South Dakota	49.8		14.6		10.3	
North Dakota	6.4		4.3		13.5	

[1] Percent gain $=\frac{100(V_r - V_o)}{V_o}$, where V_r is the variance of a random sample and V_o is the variance within district of a sample distributed in optimum proportions.

The actual distribution of the survey samples differed from the optimum in that an attempt was made to distribute the sample according to acreage. This was not successful, however, because of irregular ripening of the grain. It was necessary to hurry through some areas in order to keep ahead of harvesting combines, while in areas where the wheat was maturing less rapidly time was available to take a greater number of samples. It is important to consider whether much accuracy was lost through the failure to distribute strictly according to acreage, because steps could be taken in future years to distribute the sample more strictly according to this criterion if such distribution materially increased the accuracy. The sampling variance of the estimated means, if the distribution of the sample is strictly according to acreage, can be calculated from the data. Table 5 shows the efficiency of distribution of the survey sample and a sample distributed according to acreage, compared with the efficiency of a sample distributed in the optimum manner.

In table 5 it is shown that the efficiency of distribution of the actual sample in 1939 was on the average only 2 or 3 percent below the efficiency of a sample distributed strictly according to acreage. The lack of information concerning the standard deviation did not seriously impair efficiency. In other words, the actual distribution of the 1939 sample did not depart seriously from the acreage criteria that closely approximated the optimum. Additional calculations have indicated that if it had been possible to distribute the protein and test weight samples according to production, the 1939 efficiency of distribution for these factors would have averaged just as high as that of the yield sample distributed according to acreage, that is, about 95 percent efficient. This comparison is academic, however, in that the production of each stratum is not known at harvest time, while an indication of acreage can be obtained by the crop meter.

TABLE 5.—*Efficiency of the survey sample and a sample distributed according to acreage, compared with the efficiency of a sample distributed in the optimum manner*

YIELD

State	Distribution: Acreage [1] 1939	Distribution: Acreage [1] 1940	Distribution: Actual [2] 1939	Distribution: Actual [2] 1940
	Percent	*Percent*	*Percent*	*Percent*
Oklahoma	94	96	96	85
Kansas	95	96	85	75
Nebraska	93		99	
South Dakota	96		86	
North Dakota	98		94	
Average	95	96	92	80

PROTEIN

State	Acreage 1939	Acreage 1940	Actual 1939	Actual 1940
Oklahoma	81	96	93	85
Kansas	79	82	70	61
Nebraska	98		95	
South Dakota	88		82	
North Dakota	98		90	
Average	89	89	86	73

TEST WEIGHT

State	Acreage 1939	Acreage 1940	Actual 1939	Actual 1940
Oklahoma	85	90	95	92
Kansas	84	90	76	66
Nebraska	90		94	
South Dakota	89		80	
North Dakota	97		88	
Average	89	90	87	79

[1] Percent efficiency $= \frac{V_o}{V_a}$, where V_o = variance of weighted mean of sample distributed in optimum proportions;
V_a = variance of the weighted mean of sample distributed strictly according to acreage.

[2] Percent efficiency $= \frac{V_o}{V_s}$, where V_o = variance of weighted mean of sample distributed in optimum proportions;
V_s = variance of weighted mean of survey sample.

The actual distribution of the 1940 sample did not turn out so well. Because of irregularities in sampling conditions mentioned previously, the 1940 sample departed seriously from the acreage criteria. An average of 14 percent efficiency was lost in 1940 by failure to distribute according to acreage, while the 1939 sample showed only an approximate loss in efficiency of 2 or 3 percent from this source. In each survey, we cannot attribute all the inefficiency to the maldistribution of total number of samples among districts. It has been indicated previously that inaccuracies in the estimates of weights per stratum also were responsible for failure to gain fully from stratification. Therefore, more attention will have to be directed toward obtaining a more accurate estimate of the weights, if gains from stratification are to be more fully achieved. The estimation of weights will be reconsidered later in the cost analysis.

DOUBLE STRATIFICATION

In addition to district stratification, another possibility is that varietal differences might be of sufficient magnitude to warrant double stratification by both district and variety. Such procedure would require extra effort to obtain accurate estimates of variety proportions within a district. It is believed that the varietal weights could be estimated with adequate precision if a handful of heads for varietal identification were obtained by samplers stopping at extra fields. This extra varietal information could be obtained at low cost in both field and laboratory, so the extra expense involved in obtaining accurate varietal weights would not be a serious limiting factor if varietal stratification appeared useful.

TABLE 6.—*Mean squares on sample basis: Fields in same district vs. fields in variety in district*

Source of variability	Mean square		
	Yield per acre	Protein content	Test weight per bushel
Oklahoma 1939:	*Bushels*	*Percent*	*Pounds*
Fields in same district	127.39	9.018	9.732
Fields in variety in district	124.81	9.297	10.356
Oklahoma 1940:			
Fields in same district	125.66	7.957	11.324
Fields in variety in district	124.99	8.049	10.931
Kansas 1939:			
Fields in same district	165.06	7.728	9.312
Fields in variety in district	150.65	7.358	9.042
Kansas 1940:			
Fields in same district	171.02	7.953	25.333
Fields in variety in district	164.14	7.715	24.783
Nebraska 1939:			
Fields in same district	149.37	8.750	8.795
Fields in variety in district	130.97	8.862	8.946
South Dakota 1939:			
Fields in same district	58.77	5.636	15.626
Fields in variety in district	57.79	5.174	13.694
North Dakota 1939:			
Fields in same district	73.51	9.034	23.469
Fields in variety in district	73.42	8.250	19.069

In order to obtain an indication of the gain that would likely result from stratification by variety within a district, table 6 has been computed. This table compares the mean squares between fields in the same district and between fields within varieties in a district.

Since the mean square among fields within varieties in a district is not much smaller than the mean square among fields in the same district, the additional refinement of varietal stratification would not have been of sufficient value to warrant the extra effort required to estimate the weights accurately. Table 6 shows that such subdivision of the district would not have substantially reduced the sampling error.

NUMBER OF SAMPLES PER FIELD

When the preharvest wheat surveys were taken there was some doubt about the number of samples that should be taken per field. It was believed that the variation among fields was much greater than the variation within fields. Therefore, it appeared advisable to take only a few samples per field, and place more emphasis upon sampling a greater number of fields. As two men were required to operate the car and crop meter, it was decided to take two samples per field. These samples have given a basis for evaluating the variability within fields. Table 7 shows the relative magnitudes of the mean squares within field and among fields within districts in specified States.

TABLE 7.—*Analysis of variance on sample basis*

Source of variability	Yield		Protein		Test weight	
	Degrees freedom	Mean square	Degrees freedom	Mean square	Degrees freedom	Mean square
Oklahoma:						
1939—Fields within districts	231	127.39	228	9.018	231	9.732
Samples within fields	235	28.18	225	1.933	223	2.170
1940—Fields within districts	619	125.66	619	7.957	619	11.324
Samples within fields	624	28.74	624	2.250	624	3.064
Kansas:						
1939—Fields within districts	651	165.06	651	7.728	651	9.312
Samples within fields	660	33.77	609	1.435	625	2.171
1940—Fields within districts	1,091	171.02	1,091	7.953	1,091	25.333
Samples within fields	1,100	41.72	1,100	1.737	1,100	8.909
Nebraska:						
1939—Fields within districts	133	149.37	133	8.750	133	8.795
Samples within fields	137	28.15	132	2.266	135	1.630
South Dakota:						
1939—Fields within districts	170	58.77	170	5.636	170	15.626
Samples within fields	176	25.60	168	1.833	171	3.837
North Dakota:						
1938—Fields within districts	217	90.28				
Samples within fields	222	19.01				
1939—Fields within districts	535	73.51	535	9.034	535	23.469
Samples within fields	544	27.44	519	2.144	482	4.324

The mean squares among fields within districts are much larger than the mean squares between samples within fields. Apparently it was well to take two samples per field. It will be useful, however, to look into the possibilities of varying the number of samples per field. It will be necessary again to apply some statistical theory to

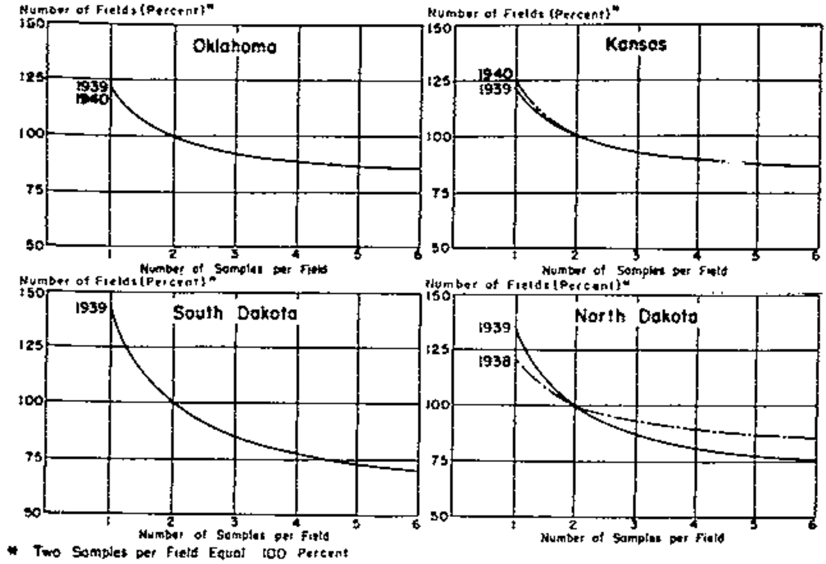


FIGURE 10.—Yield of wheat from preharvest wheat surveys, 1938–40. Relationship between number of samples per field and number of fields (expressed as a percentage) required to be sampled in order to obtain the same degree of accuracy.

estimate the behavior of the sampling error as the number of samples per field varies.

If k samples per field are taken from each of n fields, the estimated variance of the sample mean for a district is

$$\frac{1}{n}\left[\frac{(Vb-Vw)}{2}+Vw/k\right]$$

SEE CORRECTION

where Vb = mean square among fields in the district
Vw = mean square within fields
both mean squares being on a sample basis.

The percent change in the number of fields required to maintain the accuracy of two samples per field will vary according to the percent change in sampling error. The graphs shown in figure 10 illustrate by States the number of fields required to be sampled in order to maintain the same degree of accuracy when the number of samples per field is changed.

To illustrate the situation in Kansas in both 1939 and 1940, a reduction of the number of samples per field from 2 to 1 would require the sampling of 21 percent more fields in order to maintain the same degree of accuracy as was obtained with 2 samples per field. But in this case, the total number of samples required would be reduced 40 percent. If the number of samples per field were doubled from 2 to 4 in Kansas, the number of fields required to be sampled would be reduced 10 percent, but the total number of samples would be increased 79 percent. This would greatly increase costs.

The close agreement between the Kansas and Oklahoma curves in 1939 and 1940 suggests that the relationship between the variability among fields and within fields might be rather constant from year to year. Such information would be useful in planning future surveys.

In table 7, the ratio of the variance within field to that among fields is shown to be highest in South Dakota. It follows, therefore, that additional samples per field would be most useful in that State. The graph for South Dakota also shows a more sloping curve, which means that an increase in the number of samples per field would have been more effective in decreasing the number of fields sampled than in the case of the other States.

How is it possible to estimate yield from samples of 1/5000 acre? Some might think that it is necessary to use a whole field as a sampling unit, but it can be seen from the graphs that increasing the number of these small samples per field beyond 5 or 6 would not have appreciably increased the information. The 2,200 samples of 1/5000 acre each taken in Kansas in 1940 constitute a total area of only 0.44 acre. Yet from this small quantity of grain, the mean yield of the State was estimated with a standard error of 0.396 bushel.

If the fields sampled had been harvested completely, the estimated gains in information would be as shown in table 8.

TABLE 8.—*Percentage of gain in information if fields had been harvested completely*

State and year	Yield	Protein	Test weight
Kansas:	*Percent*	*Percent*	*Percent*
1939	26	23	30
1940	32	28	54
Oklahoma:			
1939	28	27	29
1940	30	39	37
Nebraska, 1939	23	35	23
South Dakota, 1939	76	48	32
North Dakota:			
1938	27	----------	----------
1939	57	31	12

It is, of course, impractical to harvest each field completely. The calculations in table 8 indicate that the loss in information as a result of sampling instead of completely harvesting a field is negligible in relation to the time and money required for complete enumeration of the sampling unit. The discussion of number of samples per field will not be complete until costs are considered. (See p. 27.)

NUMBER OF FIELDS

It is important to consider the total number of samples taken and the relationship of number of samples to precision. The number of samples needed (at two samples per field) to estimate the State yield, protein content, and test weight within 1 bushel, ¼ percent, and ¼ pound, respectively, at fiducial probability of 95 percent, and the precision of the actual sample have been computed (table 9).

TABLE 9.—*Precision of the sample in specified States, 1939 and 1940*

State and year	Samples needed for estimate of given precision 95 percent probability [1]	Samples actually taken	Precision calculated from samples actually taken 95 percent probability [2]
YIELD			
	Number	*Number*	*Number*
Kansas:			
1939	857	1, 320	0. 806
1940	1, 384	2, 200	. 793
Oklahoma:			
1939	650	470	1. 175
1940	796	1, 248	. 799
Nebraska, 1939	533	274	1. 396
South Dakota, 1939	323	332	. 986
North Dakota, 1939	302	1, 088	. 527
PROTEIN			
Kansas:			
1939	866	1, 269	0. 206
1940	1, 625	2, 200	. 215
Oklahoma:			
1939	734	457	. 316
1940	936	1, 248	. 217
Nebraska, 1939	583	269	. 369
South Dakota, 1939	445	324	. 293
North Dakota, 1939	675	1, 063	. 199
TEST WEIGHT			
Kansas:			
1939	825	1, 285	0. 200
1940	3, 334	2, 200	. 308
Oklahoma:			
1939	597	458	. 286
1940	839	1, 248	. 205
Nebraska, 1939	789	272	. 425
South Dakota, 1939	1, 384	327	. 515
North Dakota, 1939	1, 710	1, 026	. 322

[1] The number of samples needed to estimate the State yield, protein content, and test weight within ±1 bushel, ±¼ percent, and ±¼ pound.

[2] Number of samples needed $= n' = n4V/(x-m)^2$

where $(x-m)$ = one-half the range of accuracy (i. e. ½ bu., ⅛ percent, or ⅛ lb.); V = estimated variance of the mean; and n = number of samples in the sample actually taken.

SEE CORRECTION

827602°—42——4

The number of samples (n') needed to obtain given limits of accuracy is

$$n' = n \cdot \frac{4V}{(x-m)^2}$$

where $(x-m)$ = one-half the range of accuracy (i. e. ½ bu., ⅛ percent, or ⅛ lb.)
V = estimated variance of the mean
n = the number of samples actually taken

the limits of accuracy for the sample actually taken are $\pm 2\sqrt{V}$. Table 9 shows that the number of samples needed to estimate yield in Kansas in 1939 with the precision of ±1 bushel at 95 percent prob-

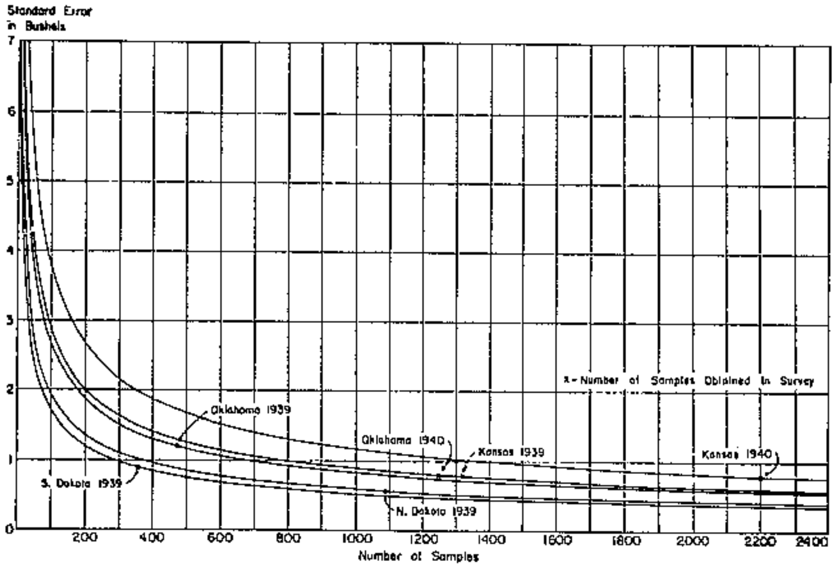


FIGURE 11.—Yield of wheat from preharvest surveys of 1939 and 1940. Ninety-five percent fiducial limits are shown for different size samples. Based upon variance among fields within crop reporting districts.

ability, differed from the number required in 1940 to give a similarly precise estimate. In other words, before the 1940 survey was taken, the best information available—the 1939 data—indicated that a sample of 857 individual samples distributed like the 1939 sample would be the appropriate number to take. Sampling variances, however, change from year to year. Hence, data for additional years will be needed in order to appraise previous data as indications of the number of samples required to obtain an estimate of given precision in a given year.

The relationship between size of sample and precision is shown in figure 11.

The data used for computing the curves were obtained by calculating the number of samples needed to obtain an estimate of given precision, by means of the formula outlined. The purpose of this chart is to indicate graphically the relative effectiveness of taking larger samples than were taken in the preharvest wheat surveys. It is noted that in the 1939 survey in North Dakota and Kansas, and the 1940 survey in Oklahoma and Kansas, the number of samples actually taken was sufficiently large to place the precision of yield estimates in these States at the lower end of their respective curves, where the slope is approaching the horizontal. It is thus indicated that the collection of additional samples would be relatively ineffective in reducing the sampling error of yield estimates in the States mentioned. If more precise estimates are desired, it may be more profitable to direct additional effort toward estimating the weights more accurately, and distributing the sample in the proportions that will permit greater gains from stratification. The efficiency in distribution will be reconsidered in the discussion of costs.

SUMMARY OF STATISTICAL EFFICIENCY

1. It would have been possible to obtain more precise information by district stratification in all States and for all variables.
2. The actual sample did not, however, fully achieve these gains.
3. Lack of gain from stratification can be attributed chiefly to errors in weights and improper distribution of the sample.
4. A distribution according to the product of the weight and standard deviation per stratum would have been optimum from the statistical viewpoint.
5. Approximately a 95 percent optimum distribution could have been obtained by distributing the entire sample according to acreage alone.
6. The efficiency of distribution of the 1939 samples was only 2 or 3 percent below the efficiency of distribution strictly according to acreage.
7. Practical difficulties in field sampling in 1940 were responsible for less efficient distribution than in 1939, and the actual efficiency of the 1940 sample distribution was approximately 15 percent below the acreage criterion.
8. Stratification by varieties would not have been worth the extra effort required to estimate varietal proportions accurately.
9. Substantial reduction of sampling error could not have been practically obtained through increasing either the number of samples per field or the number of fields sampled.

EFFICIENCY OF SAMPLE, CONSIDERING COSTS

The foregoing analysis has appraised the efficiency of distribution of the surveys from the statistical viewpoint only. The distribution of the sample needs also to be studied in connection with relative costs. If a certain amount of money were appropriated to sample the Kansas wheat crop, the question would naturally arise as to what is the most efficient sampling rate, mileage, and total amount of money to apportion to each district in order to obtain a maximum amount of information for each dollar spent. An attempt is made to answer this question for the 1940 Kansas survey. Adequate cost records are not available to permit a cost analysis of the other States sampled. One State, however, furnishes a basis to approach the problem and indicate roughly the changes that are required in sampling method to promote efficiency from the cost viewpoint.

It has been shown previously that yield had a much greater sampling error than either protein or test weight in the States sampled in 1939 and 1940. Accordingly, in the cost study an attempt is made to maximize the efficiency of distribution of the yield sample. It has also been found that it is necessary to estimate the weight per stratum as well as the mean with relatively high precision, if the full advantages of stratification are to be obtained. Therefore, in order to account for both the estimate of the weight (acreage) and the mean (yield), the procedure has been to minimize the variance of production (yield × acreage) for a given cost.

The variance of the estimated production is minimized for a given cost by using the following formula:

$$\frac{V(P)}{P^2}=\frac{1}{N}\cdot\frac{V(y)}{2\bar{y}^2}+\frac{V(A)}{A^2}$$

where $V(P)$ = variance of the estimate of production;
N = number of fields sampled;
$V(y)$ = yield mean square among fields in district, as given by the analysis of variance on a sample basis;
$\bar{y}$ = the district mean yield;
$V(A)$ = variance of the estimate of acreage;
A = estimated acreage.

In the following discussion it will be assumed that the variance of the estimated acreage is inversely proportional to the number of miles driven; then the equation may be written:

$$\frac{V(A)}{A^2}=\frac{d}{M}$$

where d is a constant and
M = total miles driven.

It may also be written:

$$K=\frac{V(y)}{2\bar{y}^2}\text{ and }M=rN$$

where r = the sampling rate (miles between samples)
N = the number of fields sampled.

Then

$$\frac{V(P)}{P^2}=\frac{1}{M}(rK+d)$$

The production variance is to be minimized for a given cost; therefore, it is necessary to estimate how the total cost depends on the number of miles driven and the rate of sampling. After investigating the costs involved in the survey, the following linear function was selected to fit the variable costs:

$$C=aN+bM=M\left(\frac{a}{r}+b\right)$$

where C = total cost;

a = cost per field not including driving. It is composed of the salaries of the samplers for the time actually spent sampling in the field plus the laboratory costs of analyzing each sample.

b = the travel cost per mile, 2.72 cents per mile for the automobile plus the salary cost per mile for the samplers.

The a and b constants were calculated empirically from the cost records. To obtain M and r, the production function was minimized, the total cost remaining fixed. This solution enabled the sampling rate (r) and the miles driven (M) to be evaluated as follows:

$$r=\sqrt{\frac{ad}{Kb}}$$

$$M=\frac{C}{\sqrt{\frac{aKb}{d}}+b}$$

This approach assumes that the density of wheat is rather constant within districts. This is not strictly the case. It is thought, however, that the above approach will be useful in roughly pointing the direction toward improvements in the cost efficiency of the sample distribution.

The r and M calculated to minimize the variance of production in each district are listed in table 10. In calculating these coefficients for each district, the C used in the formula is an estimate of the actual amount of funds spent for sampling each district in 1940. Later, the allocation of funds among districts will be investigated.

TABLE 10.—*Kansas, 1940: Sampling rates and total mileage per district, actual and calculated to minimize production variance according to the cost function*

District	Sampling rate [1]			Total mileage			Fields sampled		
	Calculated		Actual sample	Calculated		Actual sample	Calculated		Actual sample
	2 samples per field	1 sample per field		2 samples per field	1 sample per field		2 samples per field	1 sample per field	
	Miles	*Miles*	*Miles*	*Miles*	*Miles*	*Miles*	*Number*	*Number*	*Number*
1	1.662	1.168	0.936	869	977	544	72	115	80
2	2.864	2.006	2.038	1,340	1,506	969	115	185	117
3	9.772	6.429	1.224	2,437	2,502	895	35	55	104
4	1.686	1.211	1.022	1,039	1,176	680	79	124	85
5	5.717	3.973	2.952	1,438	1,573	1,034	79	124	110
6	4.891	3.383	.770	2,325	2,470	815	69	107	155
7	3.416	2.172	.844	2,341	2,446	906	95	157	149
8	6.104	3.922	2.442	2,305	2,432	1,252	115	189	156
9	3.420	2.232	.722	2,940	2,067	1,074	83	133	144

[1] Miles between samples.

The first general observation to be made from table 10 is that the sampling interval and the total mileage driven should have been increased in each of the districts in order to have utilized the funds most efficiently in estimating production. The reason for this is quite obvious when one looks at the costs of collecting and analyzing each sample, the error of the acreage estimate, and the cost of obtaining sufficient mileage upon which to base an acreage estimate. The sample showed a relatively high error in the acreage estimate and a low cost per route mile as compared with a high cost per sample for laboratory analysis and a relatively low coefficient of variability in yield. For example, the coefficients in the cost function in district 3 were:

a = \$1.53, field and laboratory costs for each field sampled;
b = \$0.0919, cost per route mile, salaries plus automobile costs;
d = 36.9772;
K = 0.1300.

It is indicated that sampling fewer fields as a result of an increase in the sampling interval and driving additional miles to obtain a more accurate estimate of acreage would have resulted in a greater amount of information per dollar spent.

In the preceding discussion an estimate of the actual amount of money spent per district in 1940 in Kansas (C in the formula) was used to estimate the rate of sampling and the mileage which would promote maximum accuracy of the district production estimates. Next the allocation of total funds among districts which will minimize the variance of the State estimate for a given State cost will be considered. As the cost per district is proportional to the miles driven, the approach has been to distribute the total mileage among districts to promote a minimum production variance. The optimum rate of sampling per district calculated from the actual allocation of funds is held constant in the calculation of the optimum allocation.

For a given number of miles driven in the State, the variance of the State production estimate will be minimum, when:

$$M_i = \frac{D_i}{\Sigma D_i} M_s$$

where M_i = the number of miles to be driven in the ith district
M_s = the total number of miles to be driven in the State

$$D_i = P_i \sqrt{r_i K_i} + d_i$$

r_i being the optimum rate of sampling computed above, and P, K, and d as calculated and defined previously.

The only remaining unknown quantity in the expression for M_i is M_s, which is computed as follows:

$$M = \frac{C}{\frac{a}{r'} + b}$$

Where C is the total amount of money spent in the State, a and b are constants calculated empirically as shown previously, and

$$r' = \frac{\Sigma(D_i)}{\Sigma\left(\frac{D_i}{r_i}\right)}$$

After thus computing the total number of miles M_i to drive per district, the variance of production for a district when the sample is allocated according to the cost and variance functions can be computed by

$$\frac{VP}{(P)^2} = \frac{1}{M_i}(r_i K_i + d_i)$$

Table 11 shows the number of miles that would have been driven and the number of fields that would have been sampled at two samples per field under the new allocation.

The high laboratory cost per sample has an important bearing upon the number of samples to be taken in each field. The analysis of variance has indicated that the variability within a field is very small in relation to the variability among fields. To reconsider the problem from the cost standpoint, the analysis of only one sample per field was made in lieu of the analysis of two samples per field. It was necessary to use different (a) and (K) coefficients in the function. The new (a) includes the salary cost of the samplers while they stopped to take the sample, plus the laboratory costs of analyzing one sample rather than two. The K_1 used is calculated from the analyses of variance, on the assumption that only one sample per field was taken. The new K_1 is calculated as follows:

$$K_1 = \frac{V(y) + V(w)}{2\bar{y}^2}$$

where $V(y)$ = yield mean square among fields in district, from analysis of variance;

$V(w)$ = mean square within fields, from analysis of variance.

Table 11 shows the number of miles and number of fields that would have been most efficient with two samples and one sample per field, compared with the actual numbers taken in the 1940 Kansas survey. The reduction to one sample per field would have increased the number of fields needed to maintain precision because of the higher variability among fields. The total number of samples, however, would have been decreased considerably. This decrease in total number of samples would have resulted in lower costs at the laboratory, thus releasing funds for covering additional mileage and thereby making possible a more accurate estimate of acreage. The total mileage driven in each district is increased and the total number of samples taken is decreased, while the number of fields sampled is greater as the number of samples per field is reduced from two to one.

TABLE 11.—*Kansas, 1940: Number of miles and number of fields per district when allocation of funds among districts is optimum*

District	Miles per district			Fields per district		
	Calculated		Actual	Calculated		Actual
	2 samples per field	1 sample per field		2 samples per field	1 sample per field	
	Number	*Number*	*Number*	*Number*	*Number*	*Number*
1	688	722	544	57	85	80
2	1, 436	1, 507	969	123	185	117
3	2, 766	3, 036	895	40	67	104
4	539	565	680	41	60	85
5	3, 571	3, 798	1, 034	196	300	110
6	2, 260	2, 440	815	67	105	155
7	1, 166	1, 269	906	47	81	149
8	3, 585	3, 885	1, 252	179	301	156
9	1, 417	1, 544	1, 074	40	67	144
State	17, 428	18, 766	8, 169	790	1, 251	1, 100

To summarize the discussion on cost efficiency, table 12 shows the increase in efficiency that could have been obtained in Kansas in 1940 if allocation of resources had been on the cost basis. With a better allocation of the funds spent in Kansas, an increase of 94 percent in information could have been obtained for the State production estimate, on the basis of two samples per field. If the number of samples taken per field had been reduced to one, the gain in information due to allocation on the basis of cost would have been 123 percent. In several districts, however, the cost allocation would have reduced the accuracy of the production estimate.

TABLE 12.—*Kansas: Gain in information in 1940 due to distribution according to cost function*

District	Production variance actual	Production variance calculated 2 samples per field	Percent gain	Production variance calculated 1 sample per field	Percent gain
1	498, 656	528, 301	-5. 6	447, 622	11. 4
2	1, 352, 036	1, 100, 766	22. 8	930, 939	45. 2
3	5, 421, 780	2, 116, 698	156. 1	1, 876, 089	189. 0
4	255, 196	411, 875	-38. 0	346, 786	-26. 4
5	7, 257, 300	2, 733, 487	165. 5	2, 346, 631	209. 3
6	3, 451, 691	1, 732, 429	99. 2	1, 507, 781	128. 9
7	797, 935	894, 149	-10. 8	784, 342	1. 7
8	5, 712, 062	2, 737, 720	108. 6	2, 397, 935	138. 2
9	1, 068, 548	1, 085, 826	-1. 6	954, 203	12. 0
State	25, 815, 204	13, 341, 251	93. 5	11, 592, 328	122. 7

SUMMARY OF COST ANALYSIS

a. The sampling interval should have been increased and the total mileage driven in each district should have been increased in order to have utilized the funds most efficiently.

b. Fewer fields thus would have been sampled to estimate yield and quality which have a relatively low coefficient of variability and involve comparatively high laboratory costs of analysis.

c. In sampling fewer fields, additional funds would have been made available for covering more mileage in order to estimate acreage which showed a high coefficient of variability and involved a relatively low cost per mile to measure.

d. Because of high laboratory costs per sample and low coefficient of variability within fields, the efficiency could have been improved through reducing the number of samples per field from two to one.

e. An allocation of the sample on the basis of the cost function would have permitted an estimated gain in information of 94 percent with two samples per field and of 123 percent if the number of samples per field had been reduced to one.

BIAS IN THE ESTIMATES OF ACREAGE, YIELD, PROTEIN, AND TEST WEIGHT

In the discussion of the statistical efficiency of the sample several methods for reducing sampling error were pointed out. In addition to keeping the sampling error small, however, it is also necessary either to eliminate virtually all bias from the sample data or be able to make adjustments in the estimates to counteract the effect of bias. The term "bias" is used here to refer to any systematic error that occurs in the survey data.

BIAS

SOURCES OF BIAS

ACREAGE

The wheat acreage indicated by the preharvest surveys was obtained by multiplying the total land area by the ratio of wheat mileage along the sampling routes to total route mileage. The acreage thus obtained might be subject to several kinds of bias.

In order to measure the frontage of wheat along selected routes in conducting a preharvest wheat survey, it is of course necessary that the personnel doing the field work be able quickly to identify wheat in a field from an automobile moving rather swiftly along a highway. Most of the time this is easy for persons with an agricultural background. When strong winds are blowing, however, or when the samplers must "look into" the sun early in the morning or late in the evening, or when it is difficult to obtain a good view of a field because of trees, banks, hedges, weeds, etc., it is sometimes necessary to stop the automobile and walk to the edge of a field to determine whether it is a wheat field. It is believed that the identification of wheat along the routes was accurate and that no bias in the indicated wheat acreage occurred from this source.

327802°—42——5

A second possible source of bias in the acreage arises from the fact that members of the crews who made the surveys had to decide whether a field of very poor wheat would be harvested. If a field of wheat as a whole is very poor but has some sizeable patches or spots of fair wheat, these patches may be harvested. A field of wheat may appear good enough to warrant harvesting but, after a few circuits with a combine, the farmer may decide that it is inadvisable to finish harvesting the field. These examples of conditions that sometimes prevail with respect to wheat farming in the drier portions of the Great Plains area are cited to show that under some circumstances, the members of field crews working on the preharvest wheat surveys could not know definitely whether certain fields or parts of fields should be regarded as wheat for harvest or wheat that would not be harvested. Instances of this kind were not widespread but, when they occurred, the crews had to make decisions that sometimes were little more than guesses. Obviously, the preharvest wheat surveys lose some of their objectivity in cases of this kind; but, fortunately, the number of such cases is relatively small.

A third source of bias in the wheat acreage estimates from the surveys might result from the failure to record enough wheat frontage, because of the fact that farmsteads are usually located near roads. The frontage line, which is about 70 yards in from the edge of the right-of-way, would likely tend to be weighted too heavily to farmsteads and, therefore, a wheat acreage obtained from this indication would be expected to be too low unless other biases should counteract this tendency. The practice of metering on a line farther in (say, about 150 yards) could have been followed. But such an increase in this distance would make metering more difficult and the results less accurate, particularly where trees, the topography of the land, etc., might impair visibility. Bias due to the "frontage" line, however, should not vary from one year to another.

Failure of the survey routes in a county to pass through areas that were representative of the wheat and nonwheat land in the county would be a fourth source of bias that could affect the estimate of wheat acreage. Comparison of the wheat acreage indicated by the surveys with that shown by the official estimates by counties reveals differences so great that one would be led to believe that in some counties the survey routes traversed a disproportionate share of the areas with concentrated wheat acreage. In some of these counties the survey indicated too much wheat, whereas in others it indicated a wheat acreage that was too small. The routes actually selected were influenced by the time available for sampling in a county, available roads, location of the county to be sampled next, and the desire to cover the county as uniformly as possible in a pattern somewhat resembling a grid. Therefore, the bias resulting from nonrepresentative routes in a county should be compensating when the acreages for individual counties are combined into a total for a crop reporting district or a State.

YIELD

The yield per acre of wheat indicated by the preharvest field sampling was obtained by threshing and weighing the wheat obtained from a small measured area in a commercial wheatfield and then multiplying the weight by the appropriate factor to convert to bushels per acre. Such a yield might be subject to several kinds of bias.

The decision made regarding whether or not a very poor field would be harvested would affect the average yields found in a county or other geographic area. This sort of bias would be inversely related to the bias in acreage from the same source and, therefore, the two should partly compensate each other when acreage and yield are converted into a production figure.

A bias also would result if the sampling routes were selective and yield levels were different over the area being sampled.

Biased yields would be obtained if wheat in the 100-yard strip of the field where samples were obtained differed from the wheat in the remainder of the field. Yield bias might also occur, because it sometimes is necessary to pass up a field selected for sampling if wheat in it is not mature.

Yield from the survey probably would be biased upward by the amount of harvesting losses encountered in commercial wheat production.

Bias in the yield estimates might arise if immature wheat were included in the sample. Precautions were taken to avoid this, however, and it is believed that this source of bias did not appreciably affect the yield estimates.

There may be other sources of bias in addition to those mentioned.

PROTEIN CONTENT

In the absence of adequate check data, it has not been possible to directly check the survey protein estimates for bias. It is known, however, that the protein content of the wheat kernel changes during maturity. Since the preharvest wheat survey samples were taken about 5 days or less before harvest, it is possible that some samples showed a different protein content than would have been found in the grain at harvest time. Actually, the samplers often found combines harvesting in the fields sampled, so the difference between sampling and harvesting dates was usually quite small. In order to obtain an indication of the change in protein content of wheat during a short period immediately preceding harvest, some data taken at Fargo, N. Dak., in 1939 have been examined.[10] The protein content of 4 varieties was found to increase in approximately the same manner throughout the 8-day period studied. The corrections that would have made the protein estimates of preharvest samplings comparable to harvest indications are listed in table 13.

If the preharvest sample estimates were to be corrected for immaturity bias, it would be necessary for the field men to estimate the number of days until harvest for each field. Many of the fields were actually sampled at harvest. Therefore, it would not be necessary to correct all samples. At this time it is not recommended that the preharvest wheat survey estimates be so corrected. It is not felt that this limited investigation necessarily furnishes conclu-

[10] KIRKBRIDE, J. WALLACE. Unpublished Report, "THE EFFECT OF HARVEST DATE ON TEST WEIGHT, YIELD, AND PROTEIN CONTENT OF SPRING WHEAT." The data were obtained under the direction of E. A. Helgeson, North Dakota State College, Fargo, N. Dak., in cooperation with the Agricultural Marketing Service weather-yield project.

sive evidence regarding the change in protein content of wheat during maturity. The data have been used only as an indication of what one might have expected at Fargo, N. Dak., in 1939. To establish more definitely a basis for correction, it will be necessary to conduct more extensive investigations for other States and for additional years and varieties.

TABLE 13.—*Adjustment of protein content of spring wheat when sampled prior to harvest*

Days prior to harvest	Amount to be added
Number	*Percent*
2	0.10
4	.20
5	.26
6	.32
7	.41
8	.50

TEST WEIGHT

Comparable data are not available to furnish a check for bias in the survey test weight estimates; however, at least two sources of bias are believed to exist. One is due to measuring test eight on samples taken prior to harvest. The data collected at Fargo, N. Dak., have provided a basis for investigating the change in test weight accompanying changes in maturity throughout 8 days before harvest. All four varieties showed similar increases in test weight during the period, the relative amount of change being approximately the same as was experienced in protein content. The analysis of variance indicated that the changes in test weight were significant. The corrections which would have been applicable to preharvest samples are shown in table 14.

TABLE 14.—*Adjustment of test weight of spring wheat when sampled prior to harvest*

Days prior to harvest	Amount to be added
Number	*Pounds*
2	0.10
4	.50
5	.80
6	1.20
7	1.70
8	2.30

Although the limited investigation cited above has indicated change in test weight of wheat prior to maturity at Fargo, further study will be required to establish more definitely a basis for correcting test weight in preharvest samples based on similar studies in other States and for additional years and varieties.

The test weight estimates were subject to a second bias which was inherent in the laboratory technique used. As the amount of grain obtained from each head sample was insufficient for 1-quart standard test weight analysis, it was necessary to use a micro method. This method consisted of measuring 16 cc. quantities of grain, and converting the weight in grams to pounds per bushel through the use of a computed conversion factor. As systematic error might arise with such a method, a study was conducted to investigate the agreement between estimates made from the micro method and the standard method.

Ten batches of wheat of various test weights were obtained for this study from the Grain Supervisor of the Agricultural Marketing Service, at Kansas City, Mo. After test weights were determined by the regular method in the Supervisor's office, the 10 batches of wheat were shipped to the Statistical Laboratory, Iowa State College, in air-tight containers to prevent moisture change while in transit. For the experiment, each of the 10 batches of grain was placed in a separate pan and numbered. Micro samplings were taken at random from the batches of grain, weighed, and replaced. One person measured the grain, and another weighed it. Tippett's numbers were used to designate the order in which samplings were taken from each batch. The weigher did not know from which batch a sampling was drawn; thus memory bias in scale reading was reduced. The individuals took turns at measuring and weighing, so that the data could furnish information regarding differences between observers. Two sizes of brass cylinders were used, one 16.08 cc. and one 32.5 cc. Thirty samples were taken at random from each batch of grain by using the smaller cylinder, and 15 were similarly selected from each batch by using the larger cylinder.

Analysis of the data permits an estimate of precision as well as of bias. The precision or sampling error of the micro method will be discussed first. The analysis of variance of the small cylinder data is shown in table 15.

TABLE 15.—*Analysis of variance of test weight obtained through the micro method by using a 16.08 cc. cylinder*

Source of variation	Degrees of freedom	Sum of squares	Mean square [1]	s
Batches of grain	9	94.3307	10.4812	
Samplings within batches	290	6.0526	.0208	0.144
Individual samplers	1	.1262	.1262	
Total [2]	299	100.3833		

[1] Analysis of variance was done on weighings in grams. To convert mean squares to pounds per bushel, multiply by $(4.831)^2$.
[2] Excludes figures for individual samplers.

The standard error of the mean of 30 samplings taken with the 16.08 cc. cylinder from each batch is $(0.144)\ (4.831)\div\sqrt{30}$, or 0.1270 pounds per bushel. Table 15 shows that the micro method is capable of detecting differences in test weight between batches of wheat, with a proportionately insignificant sampling error in the measurement of each batch. The mean square of 0.1262 for individual samplers is significant.

The analysis of variance of the measurements obtained by using the large cylinder is shown in table 16. The sampling error of an estimate from 15 samplings by using the large cylinder is $(0.182)(2.3904) \div \sqrt{15} = 0.1123$. Hence, approximately the same or more information can be obtained from 15 of the 32.5 cc. samples as from 30 of the 16.08 cc. samples. In future surveys, it is suggested that the 32.5 cc. cylinder be adopted for measuring test weight of individual samples.

After the experiment had been completed, the standard test weight measurements on each batch of wheat were obtained from the Grain Supervisor at Kansas City and were compared with the mean estimate indicated by the micro method. Table 17 shows these comparisons.

TABLE 16.—*Analysis of variance of test weight obtained through the micro method by using a 32.5 cc. cylinder*

Source of variation	Degrees of freedom	Sum of squares	Mean square [1]	s
Batches of grain	9	165. 6253	18. 4028	
Samplings within batches	140	4. 6516	. 0332	0. 182
Total	149	170. 2769		

[1] Analysis of variance was done on weighings in grams. To convert mean squares to pounds per bushel, multiply by $(2.3904)^2$.

TABLE 17.—*Estimates of test weight obtained by the use of the standard method and the micro method*

Batch number	Standard method	Micro method 16.08 cc. cylinder	Micro method 32.5 cc. cylinder
	Pounds	*Pounds*	*Pounds*
1	60. 4	59. 7	59. 5
2	58. 6	57. 9	58. 0
3	62. 6	61. 6	61. 8
4	54. 1	53. 7	54. 1
5	58. 0	57. 3	57. 6
6	60. 2	58. 9	58. 8
7	57. 4	56. 5	56. 7
8	61. 4	59. 7	60. 2
9	53. 7	53. 7	53. 7
10	55. 8	55. 1	55. 2
Average	58. 2	57. 4	57. 6

Table 17 shows that estimates obtained by the micro method are lower and apparently biased when compared with estimates by the standard method. The bias in the test weights obtained by the use of the large cylinder is slightly less than that obtained by the use of the small cylinder. In the standard method, a quart of grain is measured, weighed, and the weight is converted into pounds per bushel. It is thought that the weight of this larger quantity of grain tends to compress the mass and thereby creates a larger amount of grain per unit volume than is measured by the micro method where the com-

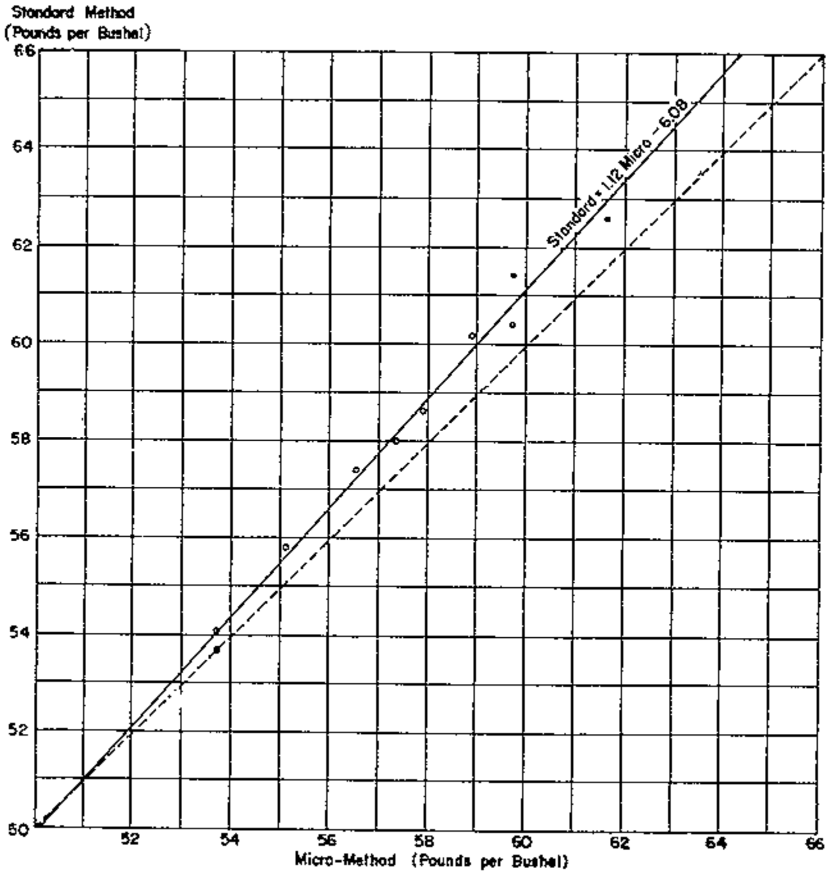


FIGURE 12.—Comparison of micro method and standard method of obtaining test weight of wheat. 16.08 cc. cylinder used in the micro method.

pression effect is decreased, as a result of the reduced volume measured. Figures 12 and 13 show graphically the bias in the micro method of estimating test weight.

The graphs show that the bias of the cylinders is quite similar and is a relative rather than a constant amount. It is interesting to note that the bias approximately disappears at test weight of 50 pounds per bushel. In future surveys the charts shown as figures 12 and 13 could be used in correcting the estimates from the micro method. As the analysis of variance indicated a significant difference between observers, it may be advisable to use a separate correction system for each individual taking the micro measurements. So long as the research phase of the survey is carried on, it will be necessary to collect quality information by individual samples. Therefore, it will be necessary to maintain the micro method. If the research work is dispensed with, however, the samples taken in a county can be pooled and standard methods used to measure test weight.

BIAS FOUND IN THE ACREAGE AND YIELD DATA

In order to obtain statistical evidence regarding the actual occurrence of bias in the acreage and yield data from the survey, the differ-

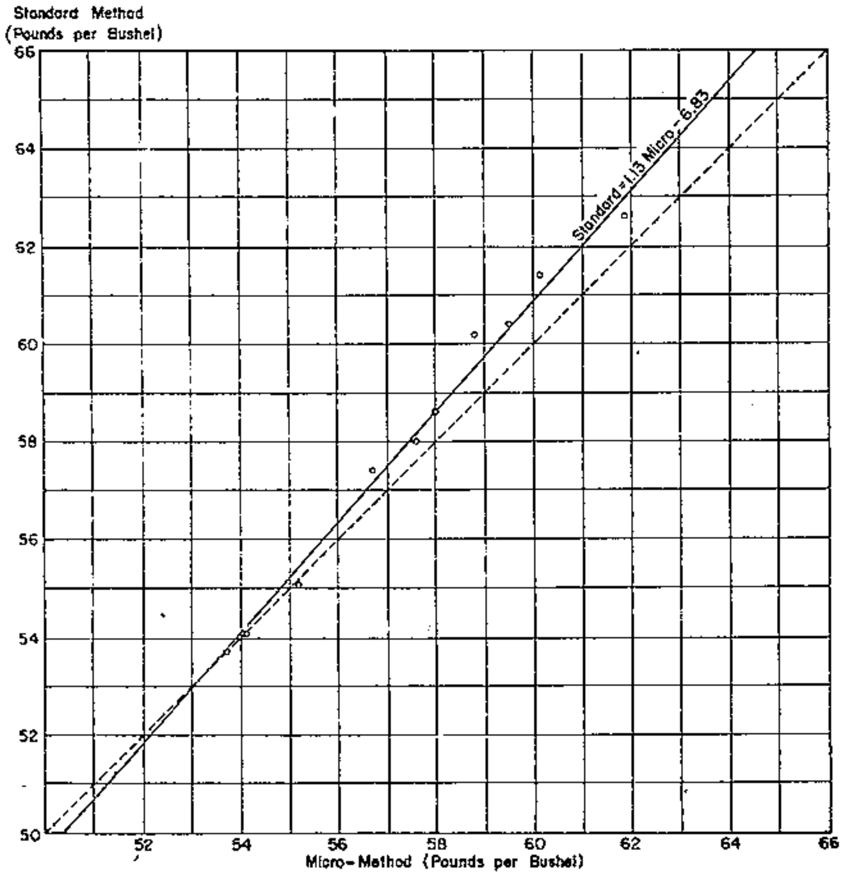


FIGURE 13.—Comparison of micro method and standard method of obtaining test weight of wheat. 32.5 cc. cylinder used in the micro method.

ences between the survey estimates of yield and acreage and the official estimates by crop reporting districts have been computed and compared with twice the sampling error of the survey estimates. If this difference between the estimates is greater than twice the sampling error of the survey estimate, it indicates that the survey estimate is, at the 5 percent probability level, significantly different from the official estimate. As the amounts by which the official estimates may differ from the true acreages and yields are not known, it has been necessary to use only the sampling error of the survey estimates in the test for bias mentioned above. Tables 18 and 19 show the differences between the official and survey estimates by districts in 1939 and 1940.

TABLE 18.—*Comparison of 1939 survey estimates and official district estimates of yield and acreage to detect bias*

State and district	Yield				Acreage			
	Preharvest wheat survey	Official	Difference preharvest minus official	Sampling error of survey estimate	Preharvest wheat survey	Official	Difference preharvest minus official	Sampling error of survey estimate
	Bushels	*Bushels*	*Bushels*	*Bushels*	*1,000 acres*	*1,000 acres*	*1,000 acres*	*1,000 acres*
Kansas:								
1	7.5	6.4	1.1	0.778	701	667	34	72
2	12.3	10.4	[1] 1.9	.656	1,350	1,172	[1] 178	80
3	21.3	16.9	[1] 4.4	1.550	637	583	54	41
4	5.6	4.4	1.2	.631	688	554	134	72
5	14.1	10.2	[1] 3.9	.704	1,754	1,592	162	84
6	20.3	17.3	[1] 3.0	1.334	690	584	106	59
7	6.8	5.0	[1] 1.8	.667	1,406	1,306	100	132
8	19.5	15.3	[1] 4.2	.850	2,698	2,612	86	101
9	23.0	15.5	[1] 7.5	3.645	901	643	[1] 258	121
Av. or total	14.9	11.5	[1] 3.4	.356	10,825	9,713	1,112	------
Oklahoma:								
1	8.2	5.8	[1] 2.4	1.006	1,039	784	[1] 255	105
2	22.1	19.7	[1] 2.4	.679	2,027	1,705	[1] 322	92
4	17.7	14.1	[1] 3.6	.919	820	567	[1] 253	90
7	7.8	8.6	−.8	.914	696	476	[1] 220	87
Nebraska:								
6	13.0	10.0	3.0	1.553	625	715	−90	55
7	16.2	12.6	3.6	4.605	527	416	111	100
8	10.2	8.8	1.4	1.333	515	404	111	97
9	12.7	13.6	−.9	1.156	756	867	−111	47
South Dakota:								
2	8.4	7.2	[1] 1.2	.418	856	733	[1] 123	46
3	14.5	13.2	[1] 1.3	.502	444	442	2	31
5	6.3	7.1	−.8	.824	335	254	[1] 81	31
6	14.4	8.6	5.8	3.228	140	130	10	38
9	18.7	8.5	10.2	5.998	221	156	65	43
North Dakota:								
1	8.8	8.6	.2	.506	1,021	1,078	−57	80
2	10.9	11.8	−.9	.584	988	1,017	−29	85
3	12.5	11.1	[1] 1.4	.495	1,476	1,362	[1] 114	48
4	10.5	10.3	.2	.354	788	697	91	53
5	11.2	9.9	1.3	.776	811	819	−8	61
6	11.9	11.4	.5	1.068	697	730	−33	48
7	12.0	10.6	[1] 1.4	.666	795	688	[1] 107	49
8	9.1	8.4	.7	.489	633	559	74	47
9	12.1	10.0	[1] 2.1	.586	680	703	−23	47
Av. or total	11.1	10.3	[1] .8	.262	7,889	7,653	236	------

[1] Difference more than twice as large as sampling error.

TABLE 19.—*Comparison of 1940 survey estimates and official district estimates of yield and acreage to detect bias*

State and district	Yield				Acreage			
	Preharvest wheat survey	Official	Difference preharvest minus official	Sampling error of survey estimate	Preharvest wheat survey	Official	Difference preharvest minus official	Sampling error of survey estimate
	Bushels	*Bushels*	*Bushels*	*Bushels*	*1,000 acres*	*1,000 acres*	*1,000 acres*	*1,000 acres*
Kansas:								
1	11.0	8.1	[1] 2.9	1.052	692	663	29	62
2	13.7	13.3	.4	.681	1,354	1,259	95	78
3	30.4	21.9	[1] 8.5	1.192	556	572	−16	89
4	7.2	6.4	.8	.754	617	544	73	71
5	19.0	16.0	[1] 3.0	1.129	1,791	1,580	211	146
6	27.2	19.8	[1] 7.4	.798	779	582	[1] 197	78
7	9.2	8.2	[1] 1.0	.217	1,021	1,113	−92	117
8	16.3	15.5	.8	.458	2,131	1,988	143	157
9	20.3	16.3	[1] 4.0	.525	632	556	76	60
Average	16.7	14.0	[1] 2.7	.280				
Oklahoma:								
1	12.3	12.5	−.2	.626	581	583	−2	69
2	19.3	14.9	[1] 4.4	.421	1,741	1,545	[1] 196	97
4	18.2	16.1	[1] 2.1	.551	715	543	[1] 172	76
5 [2]	20.0	15.9	4.1	.714	553	459	94	72
7	13.1	12.9	.2	.193	800	548	[1] 252	116

[1] Difference more than twice as large as sampling error.
[2] Includes only 5 of the 13 counties in the district.

Table 18 shows that of 31 crop reporting districts sampled in 1939, the survey estimates of yield were larger than the official estimates in 27 districts with a difference of twice the sampling error of the survey sample in 15 districts. In 1940, similar bias was experienced with yield estimates from 14 districts in which the survey estimate exceeded the official in 13 districts and 9 of the differences were greater than twice the sampling error of the survey estimates.

As for the acreage estimates, the survey estimates in 24 of the 31 districts sampled in 1939 were above the official estimates with 10 of the differences being more than double the sampling error of the survey estimates. In 1940, the survey acreage indications in 11 of the 14 districts sampled were above the official estimates and 4 of the differences were statistically significant according to the criteria being employed to discern significance of differences.

Because the survey estimates are consistently higher than the official estimates and because half the survey yield estimates and a third of its acreage estimates differ by more than twice the sampling error of the survey indications, evidence points strongly toward bias in the survey indications of yield and acreage for both years. Not all the differences were significant according to the criteria used; however, if the survey estimates are to be of much value in supplementing the early official estimates it will be necessary to reduce these differences or to be able to anticipate them. If they should prove to be relatively consistent for several years, that would indicate that it might be possible to anticipate them. The preceding section regarding statistical efficiency has suggested means whereby the sampling error of the survey estimates might be reduced by improved sampling technique. The systematic error or bias found in the estimates will now be investigated to try to determine the extent to which this component of error can be reduced.

As it has not yet been possible to allocate the total bias to its various assumed sources, our chief concern has been to investigate the possibility of applying a constant adjustment for bias in samples taken each year. If sufficient consistency is found to warrant a constant adjustment for bias in succeeding years, a great portion of the systematic error could thus be eliminated from the estimates. The consistency of bias between States in a given year has also been examined to discover whether different adjustments would be needed for each State. In making the comparisons of bias between States and between years, regression technique has been employed to indicate differences in relationship between survey and official estimates. The relationship between these estimates for the several States sampled in 1939 and 1940 is shown in figures 14 to 17, inclusive.

Figures 14, 15, 16, and 17 indicate that in each State a different regression was found to exist in both the 1939 and 1940 surveys. In figure 14, the slopes of the regression lines of official yield on survey yield are shown to be quite similar for four States, but their levels are different. The South Dakota regression line in figure 14 departs considerably from the relationship shown in the other four States. The relationship between the acreage estimate shown in figure 15 likewise is found to differ among States with the Nebraska line departing from the more consistent relationship found in the other four States. The two States sampled in 1940 show different slopes in the regression between yield estimates in figure 16, whereas in figure 17 the slopes of the regression lines between acreage estimates are found to be quite similar but the elevations of the lines differ slightly. The differences shown above in the relationship between the survey and official estimates of yield and acreage in the different States indicate that in adjusting the survey estimates for bias it may be necessary to use a different adjustment for each State.

After finding that each State appeared to show a different bias, the next step was to investigate the extent to which a bias encountered 1 year in the district estimates within a State would be repeated in the following year. As data for 2 years were available only for Oklahoma and Kansas, it was necessary to limit this phase of the study to these two States. In figure 18 it is noted that the relationship between the official yield and survey yield is quite different in 1940 from that shown in 1939 in Oklahoma. Figures 19, 20, and 21, show, however, that the relationship between survey and official acreage estimates in Oklahoma and of both yield and acreage estimates for Kansas, shows little change from 1939 to 1940.

Table 20 shows the 1940 district estimates of yield and acreage for Oklahoma and Kansas adjusted for bias as indicated in the 1939 estimates. After examining several methods of adjustment for bias, it appeared that probably the most appropriate adjustment could be made by reading the 1940 survey estimates from the 1939 relationship charts such as figures 14 and 15. In making such an adjustment,

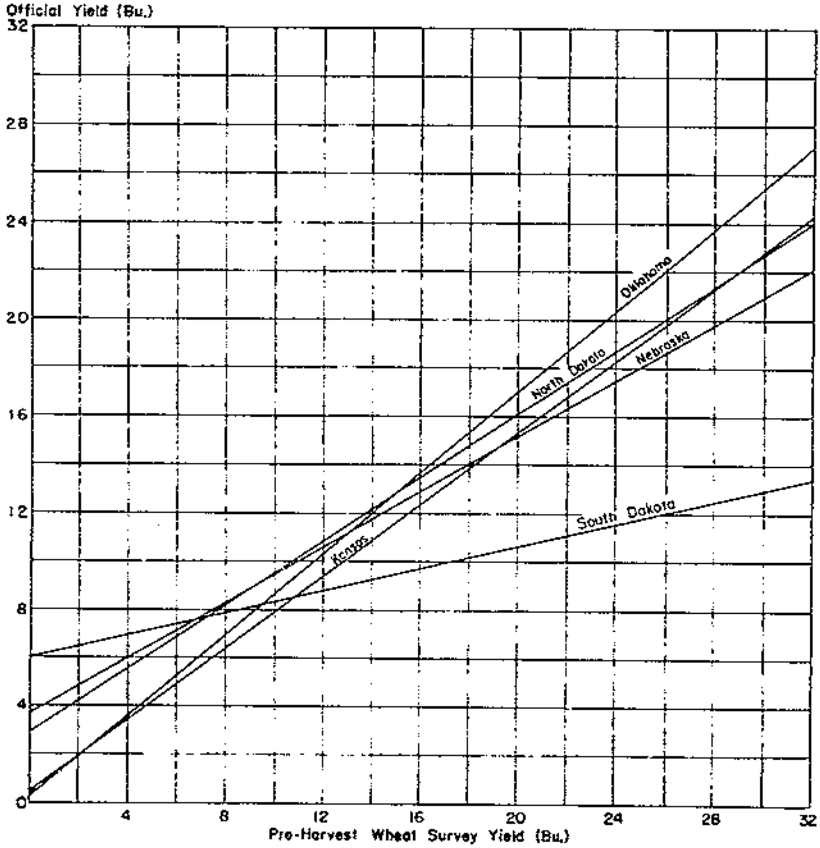


FIGURE 14.—Yield of wheat per harvested acre in five States, 1939. Regression of official yield on yield indicated by preharvest survey, by crop reporting districts.

the procedure followed was to enter the 1939 chart on the x-axis at the estimate of yield or acreage indicated by the survey in 1940, then by a vertical projection of this entry point the regression line was intersected at a point which was extended horizontally to cut the y-axis at the point which was taken as the adjusted figure.

Table 20 shows that in adjusting the 1940 district estimates for bias that was shown in the 1939 estimates, an "overadjustment" occurred. After adjustment, 8 out of 14 survey yield estimates were lower than the official estimates while before adjustment 13 of the 14 survey figures were higher than the official estimates. Eight of the 14 adjusted survey estimates differed from the official estimates by more than twice the sampling errors of the survey estimates. Six of these eight differences were negative and the other two were positive. As for the acreage estimates after adjustment, in 11 of the 14 districts the survey indications were below the official estimates as compared with 11 out of 14 above the official estimates before adjustment.

From the foregoing investigation it has been found that the total amount of bias found in 1940 appeared to differ from that of 1939. If this is true, it might indicate that the bias arises from several sources, some of which are more important in certain years than others. Therefore, to ascertain more accurately the amount of bias likely to occur in a given year, it will be necessary to determine the relative contribution of each of the various sources of bias. Then with each year's sampling it may be possible to estimate the relative possibility that each bias might occur and adjust for total bias accordingly.

TABLE 20.—*Comparison of 1940 survey district estimates of yield and acreage adjusted for bias shown in 1939 with the official estimates*

State and district	Yield				Acreage			
	Preharvest wheat survey	Official	Difference preharvest minus official	Sampling error of survey estimate	Preharvest wheat survey[1]	Official	Difference preharvest minus official	Sampling error of survey estimate
Kansas:	*Bushels*	*Bushels*	*Bushels*	*Bushels*	*1,000 acres*	*1,000 acres*	*1,000 acres*	*1,000 acres*
1	8.6	8.1	0.5	1.052	585	663	−78	62
2	10.5	13.3	[2] −2.8	.681	1,228	1,259	−31	78
3	25.2	21.9	[2] 3.3	1.192	447	572	−125	89
4	5.4	6.4	−1.0	.754	520	544	−24	71
5	15.2	16.0	−.8	1.129	1,649	1,580	69	146
6	22.4	19.8	[2] 2.6	.798	672	582	90	78
7	7.0	8.2	[2] −1.2	.217	910	1,113	−203	117
8	13.2	15.5	[2] −2.3	.458	1,967	1,988	−21	157
9	16.5	16.3	.2	.525	535	556	−21	60
Average	13.4	14.0	−.6	.280				
Oklahoma:								
1	9.7	12.5	[2] −2.8	.626	380	583	[2] −203	69
2	15.6	14.9	.7	.421	1,460	1,545	−85	97
4	14.7	16.1	[2] −1.4	.551	500	543	−43	76
5 [3]	16.3	15.9	.4	.714	355	459	−104	72
7	10.3	12.9	[2] −2.6	.195	580	548	32	116

[1] Adjusted for bias shown in 1939.
[2] Difference more than twice as large as sampling error.
[3] Only 5 of the 13 counties in the district were sampled.

Figures 22 to 38, inclusive, show certain comparisons of wheat acreage, yield, and production as indicated by the preharvest field samples with the official estimates on a crop reporting district basis. These figures show the amount of scatter in the district figures. Where the survey indications have not been adjusted, the extent and direction of their bias and sampling error may be observed. Where the survey indications have been adjusted and the plotted points depart from a one-to-one line of relationship, it indicates that the

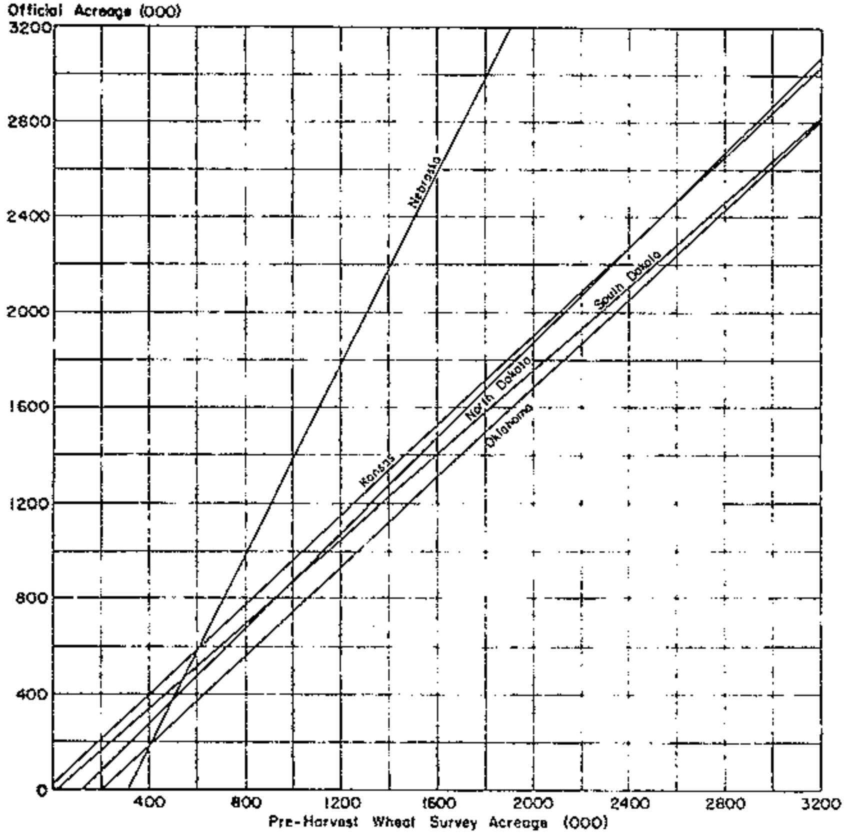


FIGURE 15. Acreage of wheat harvested in five States, 1939. Regression of official acreage on acreage indicated by preharvest survey, by crop reporting districts.

bias in the survey data was not the same in the 2 years or that any error that might be present in the official estimates differed 1 year with the other.

COMPARISON OF PRODUCTION ESTIMATES FROM THE PREHARVEST SURVEY WITH OFFICIAL PRODUCTION ESTIMATES

A comparison of the official estimates of wheat production in Kansas with estimates indicated by the preharvest survey will illustrate the possibility of obtaining greater accuracy in the estimated size of the wheat crop at harvest time. Kansas is used as an example because it is the only State in which the preharvest survey covered all crop reporting districts in both 1939 and 1940. This permits a ready comparison of estimates from the preharvest survey with the preliminary official estimates which in nearly all States are on a State basis only.

It is obvious from tables 18 and 19, as well as from the charts of both acreage and yield, that the unadjusted indications from the surveys are high. When the acreage computed from the wheat

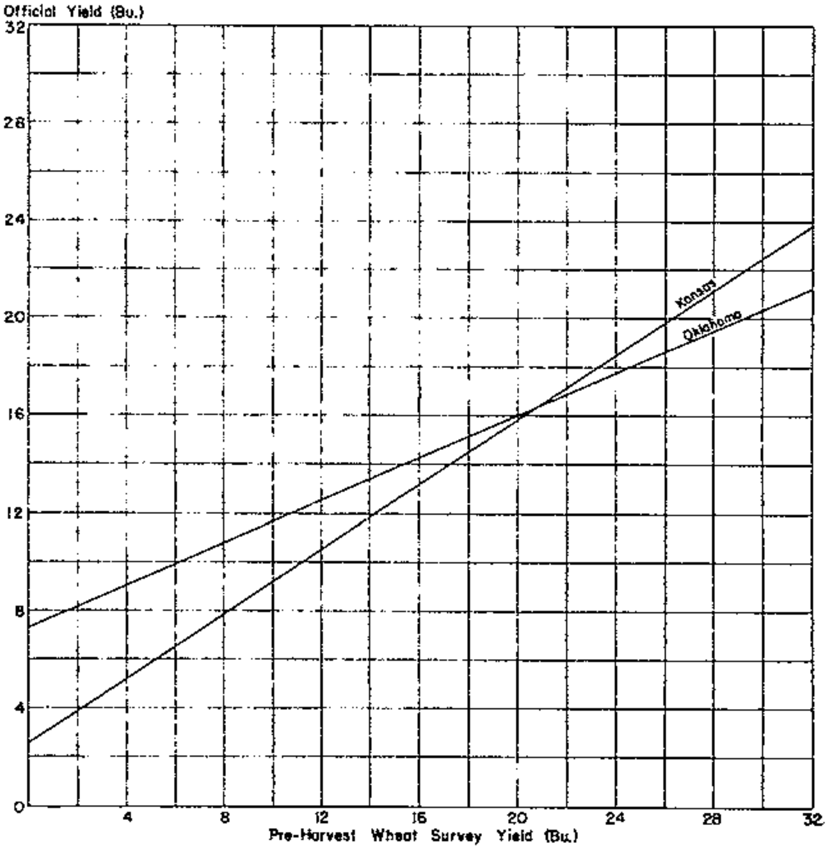


FIGURE 16.—Yield of wheat per harvested acre in Kansas and Oklahoma, 1940. Regression of official yield on yield indicated by preharvest survey, by crop reporting districts.

frontage obtained through use of the crop meter is multiplied by the yield computed from the wheat grain samples, the indicated production exceeds the official estimates as finally determined. Bias of this kind is present in most samples of data relating to crop production, particularly of acreage. Since 1939 was the first year the survey was conducted in Kansas, there was no way of knowing currently whether the survey data would be biased, or the direction and extent of any bias that might be present. The 1940 survey data relating to acreage and yield were adjusted for the bias shown by comparing the 1939 survey indications with the official estimates for 1939. Then.

after the 1940 survey had been conducted, the 1939 survey data were adjusted for the bias shown by the 1940 survey to see what the 1939 survey would have indicated currently if some method of adjustment had been available.

The unadjusted and adjusted survey figures for both years are compared with the official estimates prepared currently as of July 1, August 1, and December 1 in table 21. The wheat production indicated by the preharvest survey is the most nearly comparable

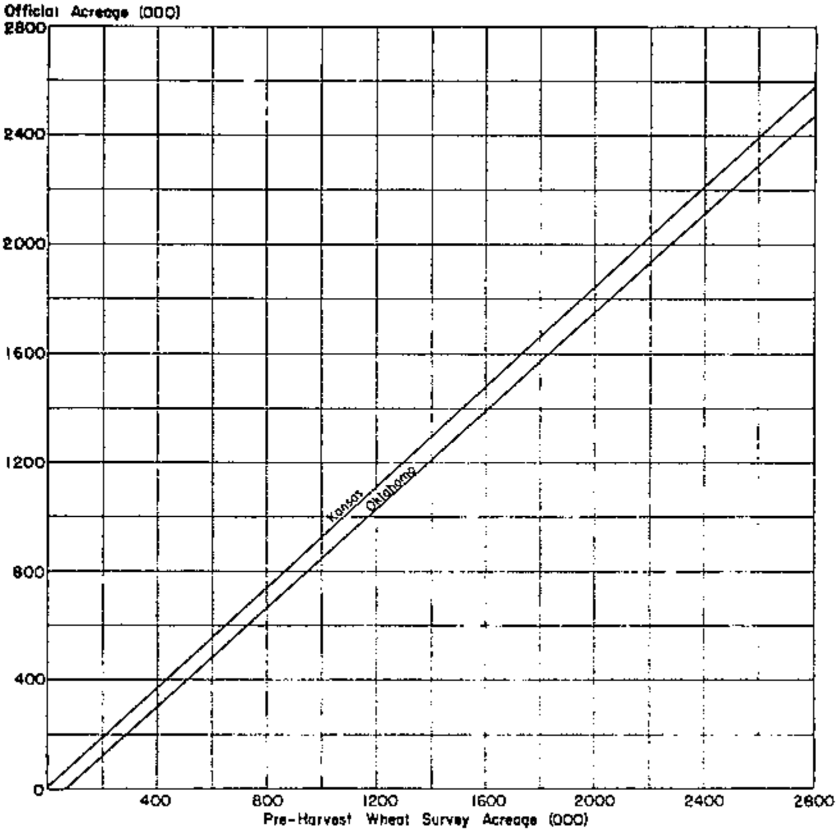


FIGURE 17.—Acreage of wheat harvested in Kansas and Oklahoma, 1940. Regression of official acreage on acreage indicated by preharvest survey, by crop reporting districts.

as to time of preparation and availability with the official estimate for July 1.

The table indicates that in 1939 the preharvest survey (if it had been possible currently to make the adjustment for bias) would have indicated a crop considerably too large and that the official estimate at harvest was practically the same as in December. A preliminary examination of the Federal census enumeration of wheat acreage and production for Kansas in 1939 suggests that the December official estimate may be a little low but still the official estimate as of July 1 seems to be better than the preharvest survey indication adjusted in the manner previously described.

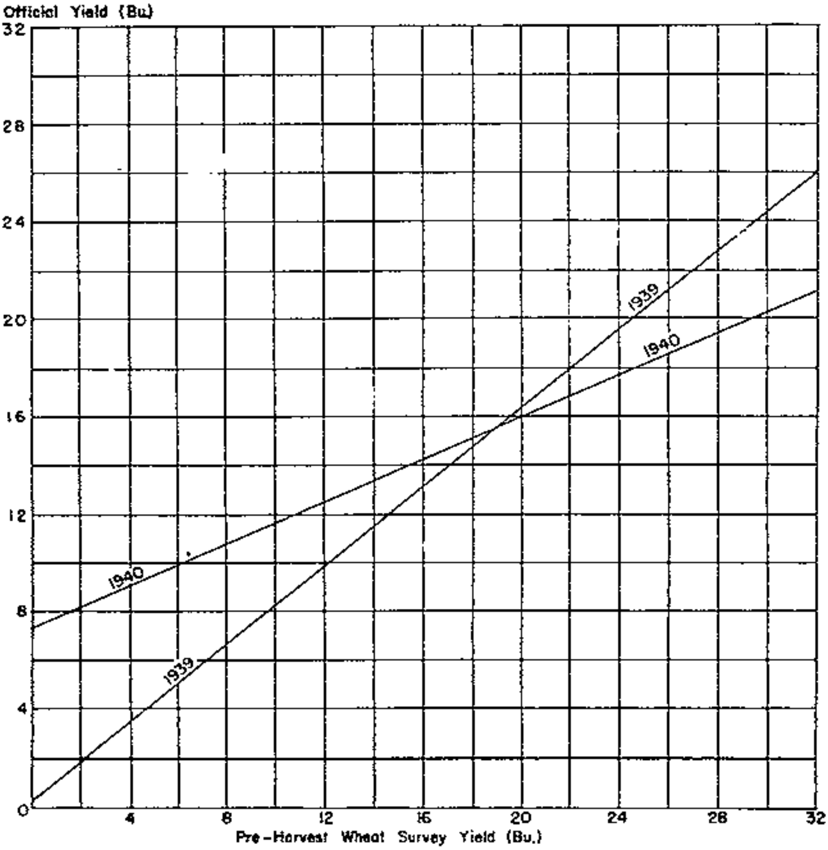


FIGURE 18. Yield of wheat per harvested acre in Oklahoma, 1939 and 1940. Regression of official yield on yield indicated by preharvest survey, by crop reporting districts.

TABLE 21.—*Kansas wheat production, official estimates and indications from survey, 1939 and 1940*

Official estimates and indications from survey	1939	1940
	Million bushels	*Million bushels*
Official estimates:		
July crop report	111	89
August crop report	116	101
December crop report	112	124
Preharvest wheat survey:		
Primary indication	[1] 161	[1] 160
Adjusted indication	[2] 125	[3] 113

[1] Acreage, indicated by crop meter ratios, multiplied by yield per acre indicated by wheat grain samples.
[2] 1939 survey data adjusted by bias shown by 1940 survey.
[3] 1940 survey data adjusted by bias shown by 1939 survey.

In 1940, however, the preharvest survey (adjusted) more nearly indicated the true size of the Kansas wheat crop at harvest than the official estimate on July 1. The survey related to a later date, since the field work on the survey in 1940 was not completed until July 10—the day the official July 1 report was released by the Crop Reporting Board at Washington. Since a few days must necessarily elapse between the time field work is completed and the time when samples can be threshed, weighed, computations made, etc., at the

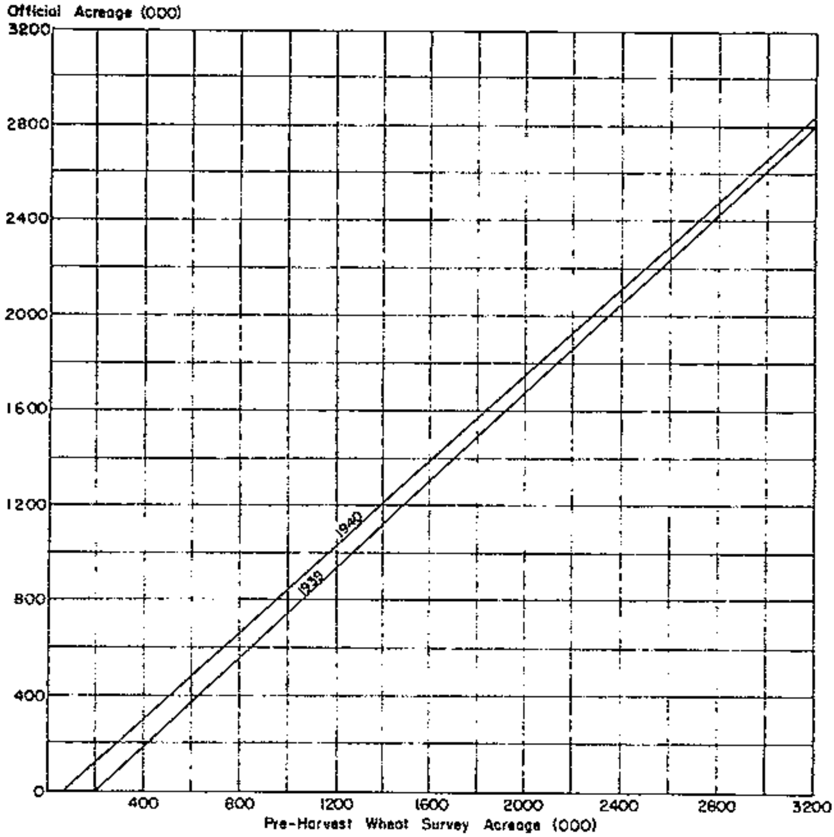


FIGURE 19.—Acreage of wheat harvested in Oklahoma in 1939 and 1940. Regression of official acreage on acreage indicated by preharvest survey, by crop reporting districts.

laboratory, the complete results of the 1940 survey in Kansas were not available in time to have been used as an aid in preparing the official crop report as of July 1. The survey information for part of Kansas could be made available even in years of late harvest as in 1940, however, in time to be utilized in preparing the July crop report.

Since the survey method is still in the experimental stage, it remains to be seen just how effective it is as an indication of acreage and yield of wheat. This will be learned only after observing the results of the survey over a period of years. Based on experience to date, however, it appears to have merit and may prove to be a real contribution to methods of estimating the wheat crop. It now seems that with respect to estimating the size of the wheat crop, the survey may have its greatest utility in years when crop prospects shift radically during the several weeks immediately preceding harvest. The most recent example of such a year in Kansas is 1940 when wheat was seeded under extremely unfavorable conditions and presented a dismal prospect until copious, timely precipitation during the spring permitted

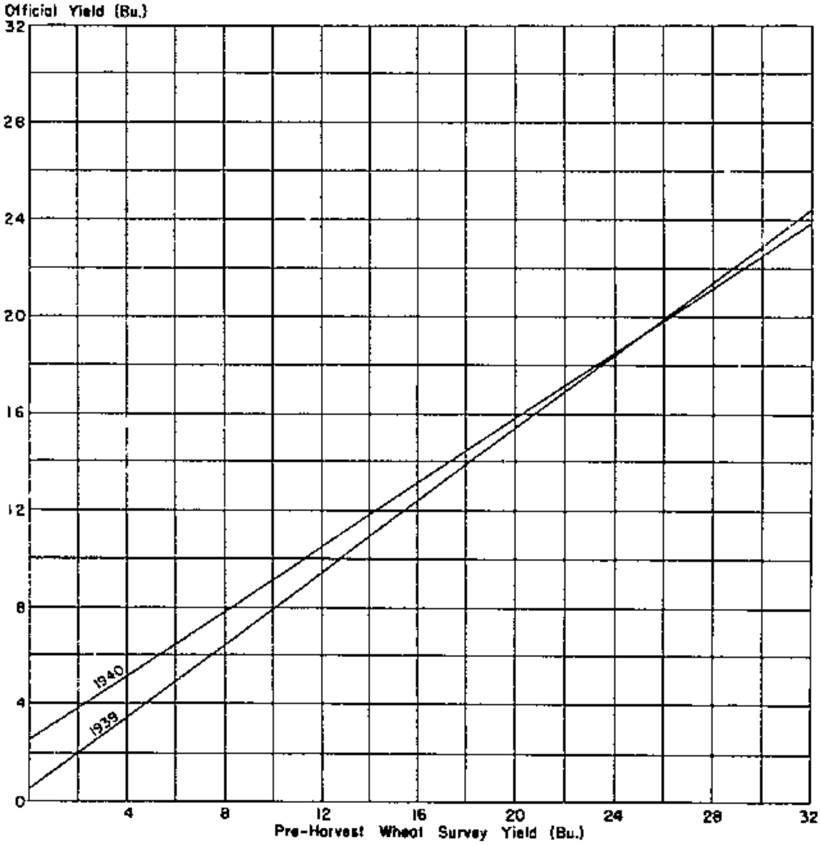


FIGURE 20. Yield of wheat per harvested acre in Kansas, 1939 and 1940. Regression of official yield on yield indicated by preharvest survey, by crop reporting districts.

a remarkable development. As a result, rapid and continuous improvement of the crop persisted until harvest. Since the regular acreage schedules used by the Crop Reporting Board in preparing its July 1 crop report were filled out and sent in by growers in early June, they apparently failed to reflect the full recovery of wheat that earlier appeared to be lost. Nor did other regular sources of information fully indicate the extent of the area of wheat to be harvested. Yield information for the July report was sent in about July 1 but growers even then were not fully aware of the yields that would finally be secured. Under these circumstances, the July 1 official estimate of Kansas wheat production was too low.

The preharvest survey taken at or just before harvest, therefore, should be of most help in years when prospects change rapidly as harvest approaches. The limited amount of evidence now available for purposes of comparison suggests that present methods employed by the Crop Reporting Board for obtaining wheat acreage and yield data by mail inquiry probably will give more accurate estimates of the wheat crop at harvest than the preharvest survey (as conducted so far) in years when crop prospects are relatively stable as harvest

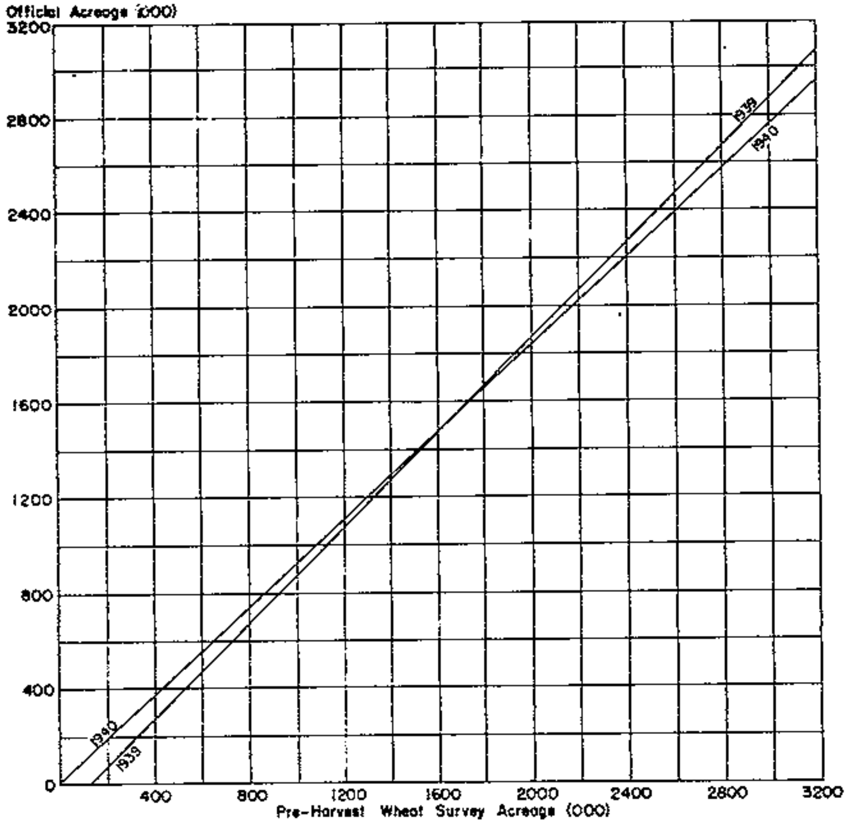


FIGURE 21.—Acreage of wheat harvested in Kansas in 1939 and 1940. Regression of official acreage on acreage indicated by preharvest survey, by crop reporting districts.

approaches. During the period 1931–40, however, the official estimate of Kansas wheat production as of July 1 varied from the final estimate by 15 percent or more in 3 of the 10 years. It seems that in such years the results of the preharvest survey might prove to be valuable as supplemental information to be used in addition to that obtained from other sources in preparing the official wheat estimates.

FORECASTING

Sufficient information relating to plant characteristics was collected to form the basis for a study of the correlation of yield per acre with plant counts and measurements. The plant observations were confined to (a) number of heads, (b) height of plant, and (c) average length of heads within the area of the sampling unit. These observations can be readily obtained prior to wheat harvest. A correlation between yield and plant characteristics would indicate the possibility of developing an objective method of forecasting yield from plant counts and measurements.

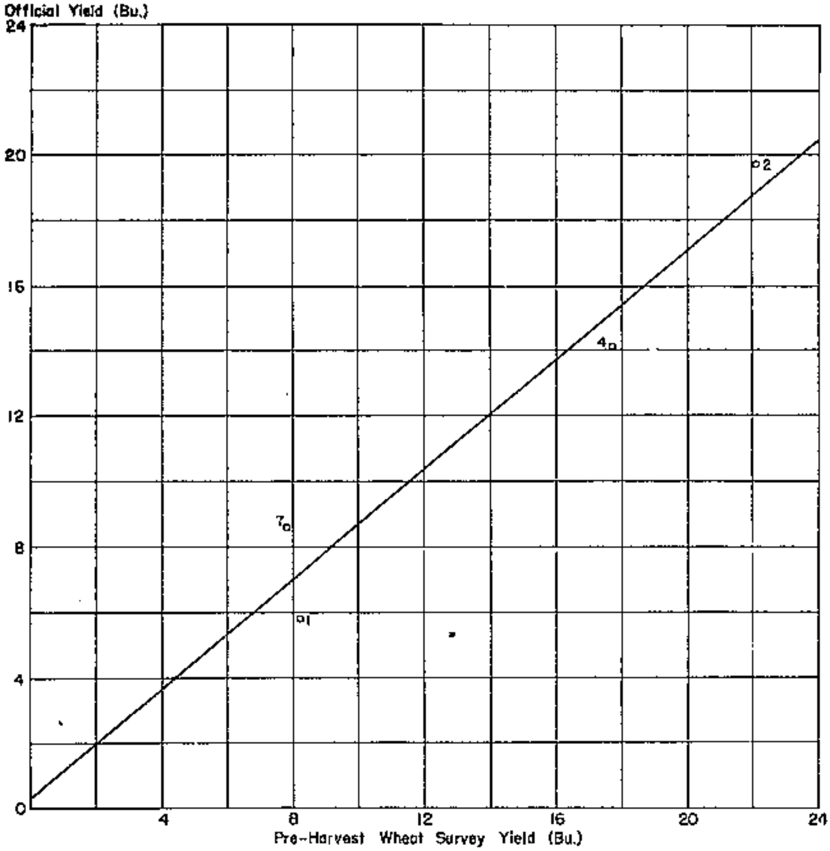


FIGURE 22.—Yield of wheat per harvested acre in Oklahoma, 1930. Regression of official yield on yield indicated by preharvest survey, by crop reporting districts.

A more objective method of forecasting wheat yields than now used might permit the Crop Reporting Board to attain greater accuracy in its forecasts. If it should develop that yields can be forecast from plant observations with greater accuracy than by present methods which utilize reports from growers giving condition and probable yield, the accuracy of the production forecasts would be improved. The production forecasts still would be affected by any error contained in the forecast of the acreage to be harvested. Nevertheless, any method which would permit greater accuracy in the forecasts of either of the components of production (acreage and yield) would prove valuable.

If it should be found that wheat production can be forecast with more precision by preharvest field sampling than by present methods, it might be advisable to issue forecasts for areas smaller than a State because of variation in date of maturity. In this way a wheat forecast for southeastern Kansas based on plant observations could be issued soon after the field sampling in that area was completed, possibly about May 20. On the other hand the forecast for the northwestern part of the State, where wheat matures somewhat later, would

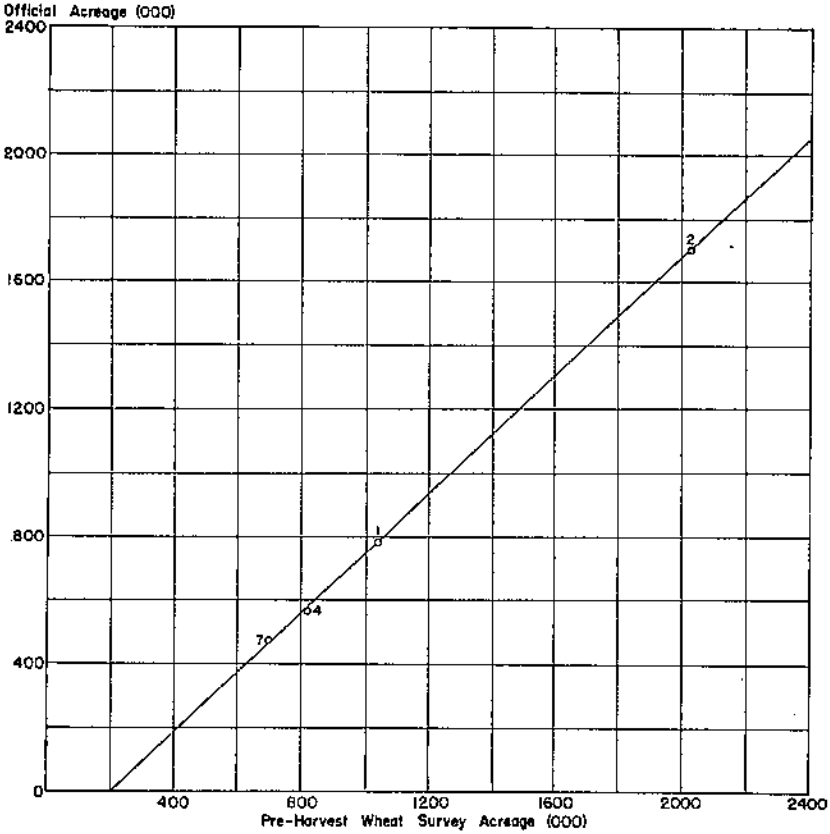


FIGURE 23.—Acreage of wheat harvested in Oklahoma, 1939. Regression of official acreage on acreage indicated by preharvest survey, by crop reporting districts.

not be available until about June 18. Forecasts for other parts of the State could be obtained between those two dates. Similar series of forecasts of the wheat crop for other States could be issued about 3 weeks in advance of harvest in each area, the forecasts to be based on the results of field sampling started about a month prior to harvest.

Such a series of forecasts would have the advantage of timeliness. That is, the forecast for an area smaller than a State could be published as soon as it was possible to make the necessary observations. However, since the effect of weather, insects, diseases, etc., on the wheat crop during the month preceding harvest cannot accurately be predicted, it would seem that the combining of a series of separate forecasts each applying to a different date to secure an over-all total might give misleading results. Since one of the most important functions of the Crop Reporting Board is to furnish forecasts of the total production of a crop for the whole country, it would be

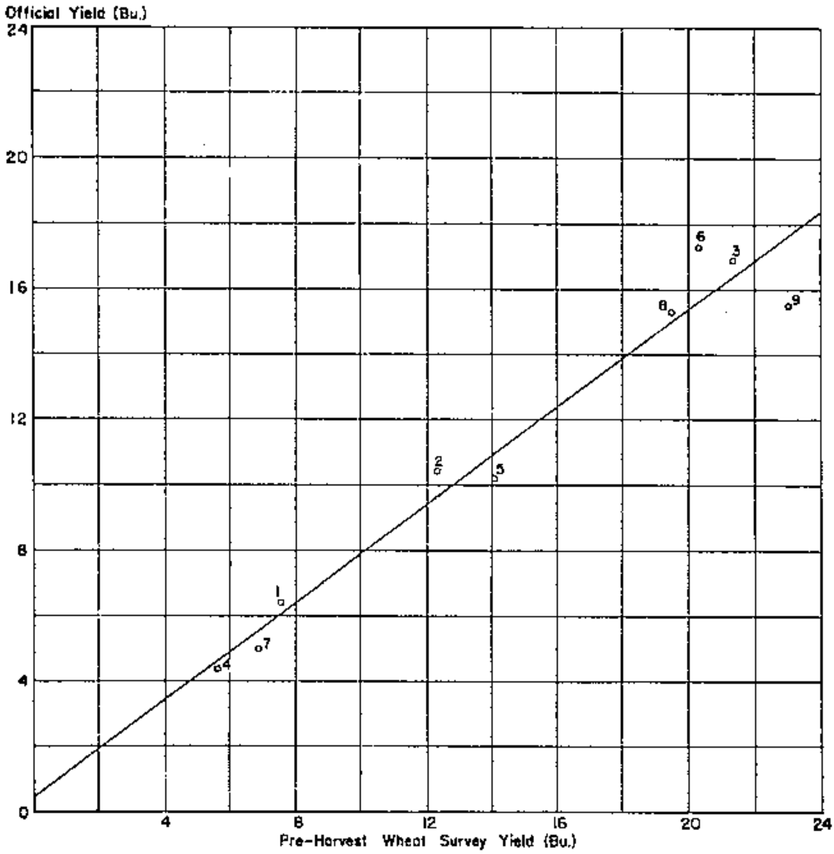


FIGURE 24. Yield of wheat per harvested acre in Kansas, 1939. Regression of official yield on yield indicated by preharvest survey, by crop reporting districts.

necessary to utilize information from sources other than the preharvest survey in forecasting total production as of a given date.

An attempt has been made to point out some of the goals to strive for and difficulties that may be encountered in forecasting wheat acreage and yield from preharvest field sampling and then utilizing such forecasts in a program of crop reports. The possibility of making serviceable forecasts from preharvest wheat survey data will be examined next.

It was found in the 1938 survey (*5*) that out of nine plant observations taken, the number of heads, height of plant, and length of head gave most promise of providing a basis for forecasting the final yield. Consequently, in 1939 and 1940, the plant observations were confined to these three. The observations were made on each of the $\frac{1}{5000}$ acre sampling units that were taken from each field. Since yield also was obtained on each of these units, it was possible to make several correlations. Yield was first correlated with the one variable, number of heads, then with the two variables, number of heads and height, and finally with three variables, number of heads, height of plant, and length of head. The results of the three analyses are shown in table 22.

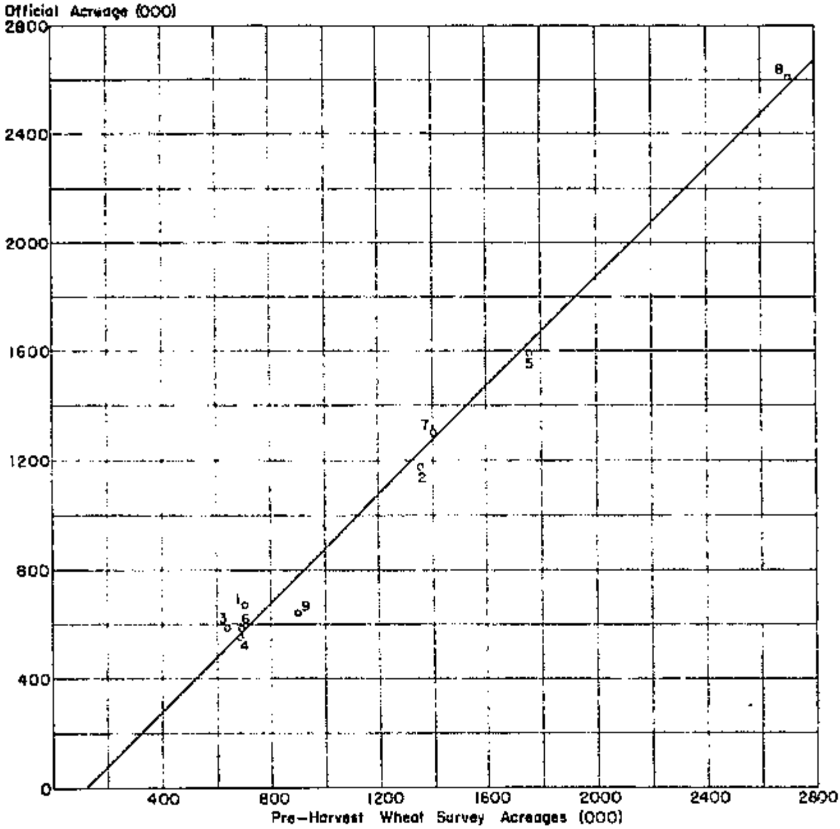


FIGURE 25.—Acreage of wheat harvested in Kansas, 1939. Regression of official acreage on acreage indicated by preharvest survey, by crop reporting districts.

In both years the number of heads accounts for a highly significant amount of the variation in yield and might be considered an important factor in determining the final yield. The addition of another variable, height, however, increases the correlation coefficient (R_2) to such a degree that the error of estimate is again reduced by a highly significant amount. The introduction of a third variable, length of head, also reduces the unexplained variation, but by only a small amount. Therefore, in future surveys, it might not pay to continue to measure length of heads. This can be seen by comparing r^2_1 with R^2_2 with R^2_3 in table 22.

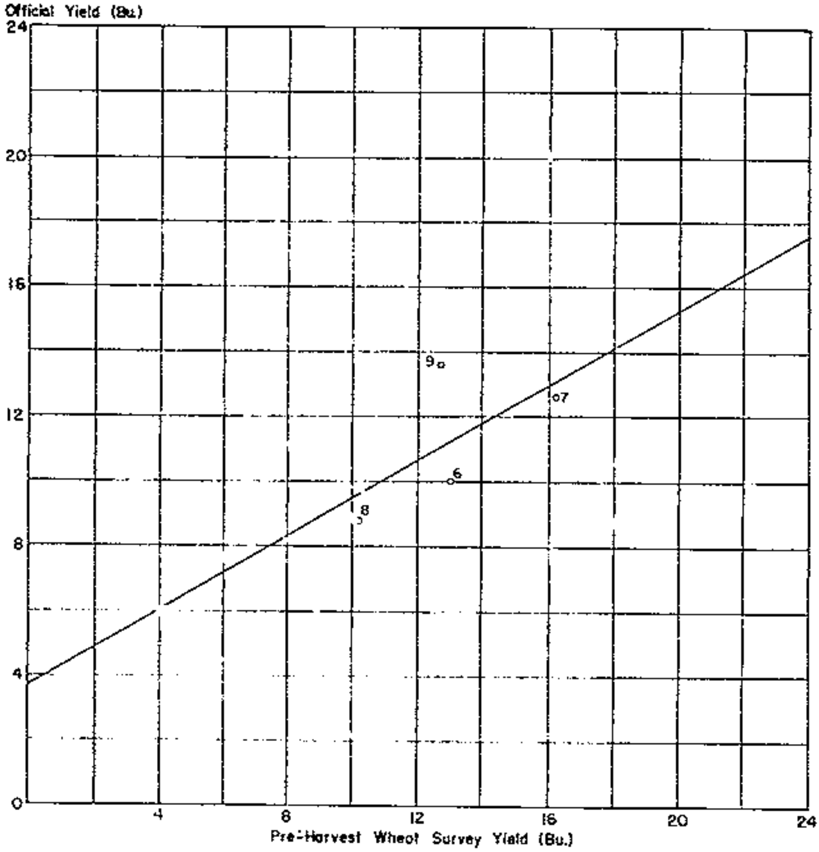


FIGURE 26. Yield of wheat per harvested acre in Nebraska, 1939. Regression of official yield on yield indicated by preharvest survey, by crop reporting districts.

TABLE 22. *Multiple regression analysis, with variation among States removed, based upon individual samples, 1939 and 1940*

Item	1939[1]	1940[2]
Yield *vs.* number of heads:		
Correlation coefficient (r_1)	0.760	0.507
Square of correlation coefficient (r_1^2)	.580	.257
Regression coefficient (b)	.077	.094
Yield *vs.* number of heads and height of plant:		
Coefficient of multiple correlation (R_2)	.859	.730
Square of coefficient of multiple correlation (R_2^2)	.738	.533
Standard partial regression coefficient ($\beta_{YN \cdot H}$)	.568	.365
Standard partial regression coefficient ($\beta_{YH \cdot N}$)	.443	.544
Yield *vs.* number of heads, height of plant, and length of head:		
Coefficient of multiple correlation (R_3)	.861	.721
Square of coefficient of multiple correlation (R_3^2)	.741	.535
Standard partial regression coefficient ($\beta_{YN \cdot HL}$)	.572	.361
Standard partial regression coefficient ($\beta_{YH \cdot NL}$)	.419	550
Standard partial regression coefficient ($\beta_{YL \cdot NH}$)	.060	−.043

[1] Data represent 5 States: Oklahoma, Kansas, Nebraska, South Dakota, and North Dakota.
[2] Data for Kansas and Oklahoma only.

To forecast yield of wheat based on a single variable, such as number of heads, it would be necessary to use a regression equation such as $y=\bar{y}+b(x-\bar{x})=a+bx$, the regression equation to be based on data secured from previous investigations. In this equation y is the estimated yield, x is the number of heads, b is the regression coefficient of y on x from the previous investigation, and a is a constant denoting the level of the regression and is obtained by subtracting the weighted mean number of heads times the regression coefficient from the weighted mean yield, all obtained in the previous investigation. The weighted mean number of heads obtained in the year for which the forecast is to be made would be substituted in this equation for x and the forecasted yield would result. The weighted mean number of heads is obtained by weighting the mean number of heads in each district by the acreage in the district as shown by the sample after it has been corrected for bias. If, however, the relationship between yield and number of heads is found to be curvilinear the equation would need to be modified accordingly. The best results would be expected from a forecasting equation based upon

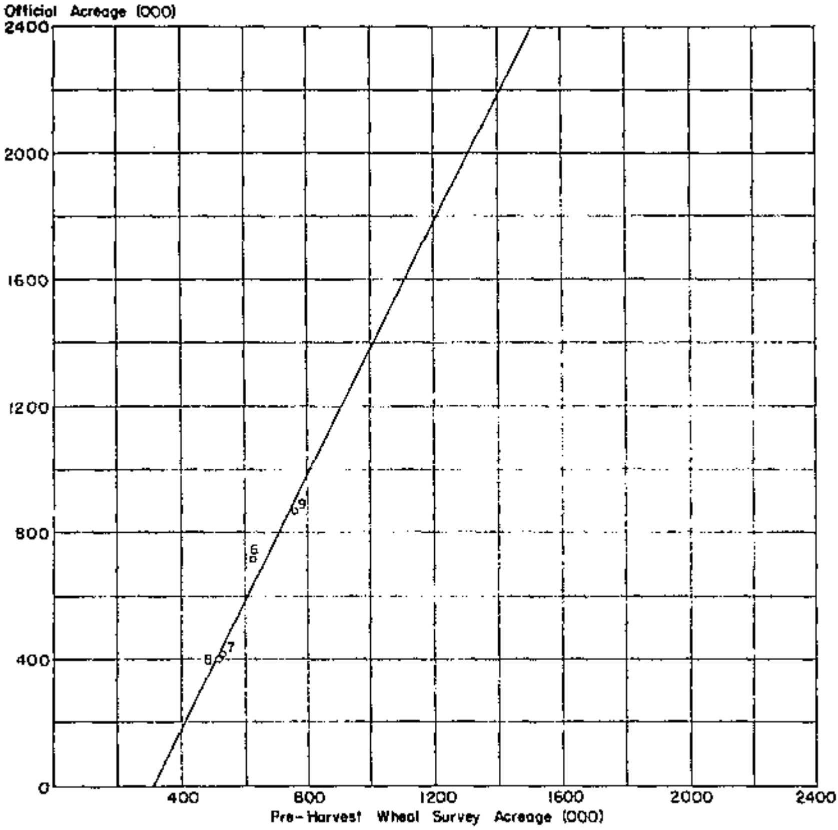


FIGURE 27. Acreage of wheat harvested in Nebraska, 1939. Regression of official acreage on acreage indicated by preharvest survey, by crop reporting districts.

data secured in a year when conditions of crop growth were similar to the one to which the forecast applies. Since any variation in the *a* constants and *b* coefficients from year to year would result in an error in the forecasts, it would be of interest to examine the differences between corresponding constants and coefficients in the 2 years for which we have data.

There have been three kinds of regressions studied. In one case the regressions were obtained by pooling all the samples in each State.

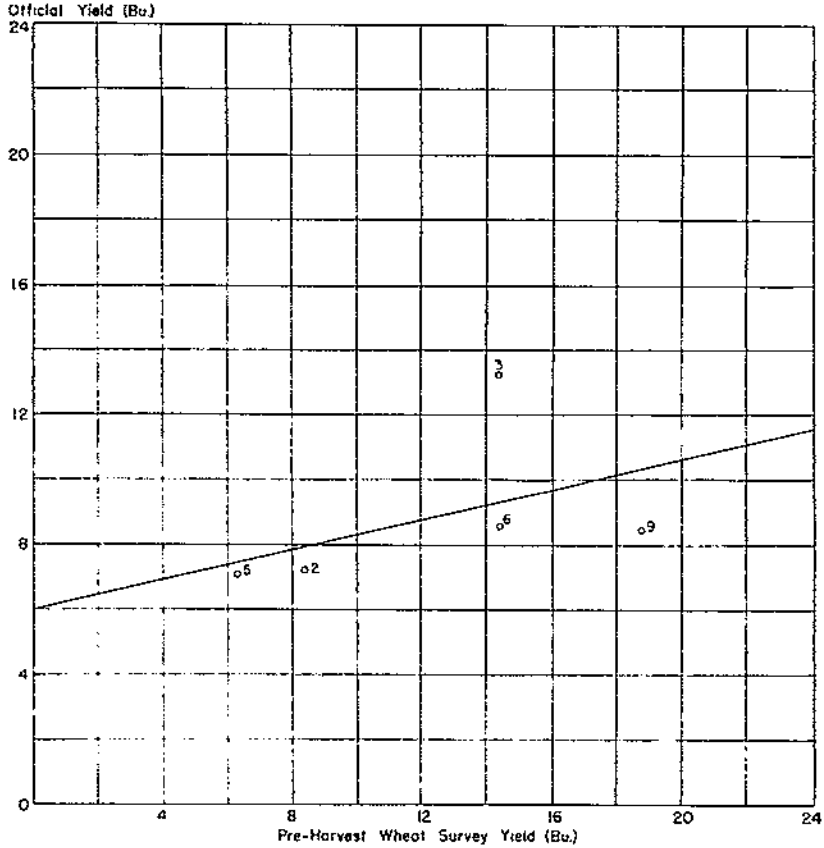


FIGURE 28. Yield of wheat per harvested acre in South Dakota, 1939. Regression of official yield on yield indicated by preharvest survey, by crop reporting districts.

In a second case the effect of district variation was removed from the regression equation. In the third case the samples were segregated into the four most prevalent varieties with the remainder thrown into a miscellaneous variety group. Then for each variety classification the effect of the county variation was removed from the regression equation.

All three kinds of regressions have been calculated and the resulting constants and coefficients are given in table 23.

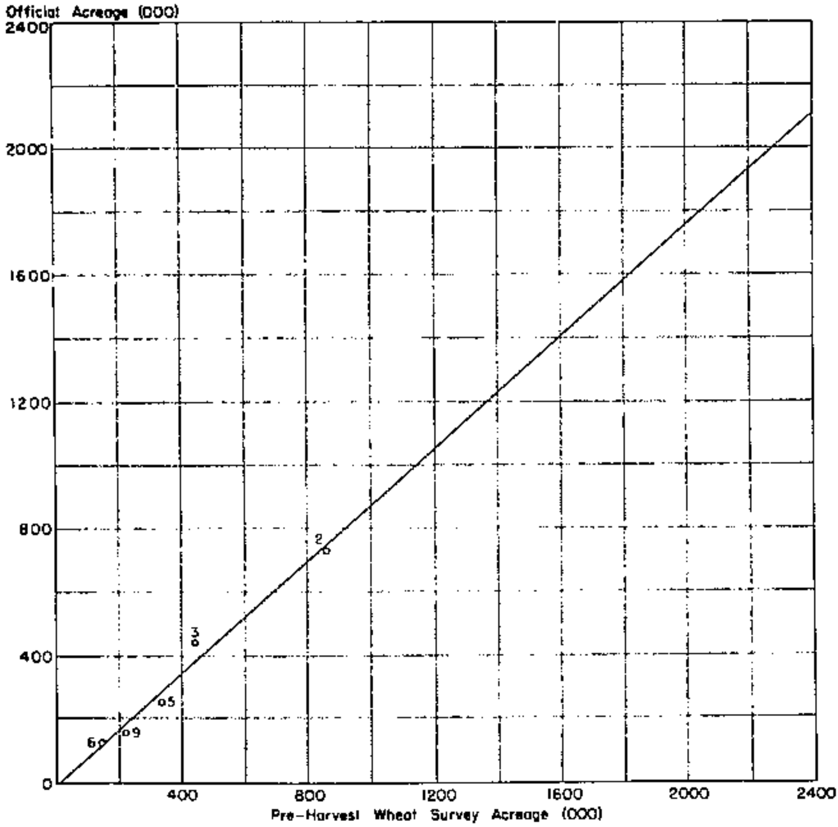


FIGURE 29.—Acreage of wheat harvested in South Dakota, 1939. Regression of official acreage on acreage indicated by preharvest survey, by crop reporting districts.

TABLE 23.—*Regression and correlation of yield on number of heads*

State and year	No variation removed				District variation removed				County variety variation removed			
	b [1]	a [1]	r	r^2	b [1]	a [1]	r	r^2	b [1]	a [1]	r	r^2
Oklahoma:												
1939	0.094	−2.02	.794	.631	0.080	−0.06	.768	.590	0.078	0.37	.761	.570
1940	.065	6.02	.450	.202	.008	5.40	.478	.229	.040	10.17	.372	.138
Kansas:												
1939	.088	−1.40	.811	.658	.070	−.08	.807	.651	.081	−.40	.859	.738
1940	.127	−2.49	.589	.346	.108	.29	.599	.359	.093	2.32	.588	.345
Nebraska, 1939	.074	−1.64	.812	.659	.074	−1.64	.815	.660	.052	2.59	.732	.535
South Dakota, 1939	.047	1.60	.628	.395	.030	3.20	.548	.300	.073	.47	.609	.371
North Dakota, 1939	.055	1.50	.693	.480	.058	.93	.686	.471	.060	−.35	.746	.556

[1] a and b are the constants in the regression equation $y = a + bx$.

From table 23 it can be seen that the *a* constants and *b* coefficients differed both between years for a given State and between States in the same year. This indicates that in the 2 years both the slope and level of the regression have changed. Weather factors affecting the growth of wheat in Kansas and Oklahoma in the 2 years studied were dissimilar; and the differences in the constants and coefficients of the equations for the 2 years are probably greater than usually would be found.

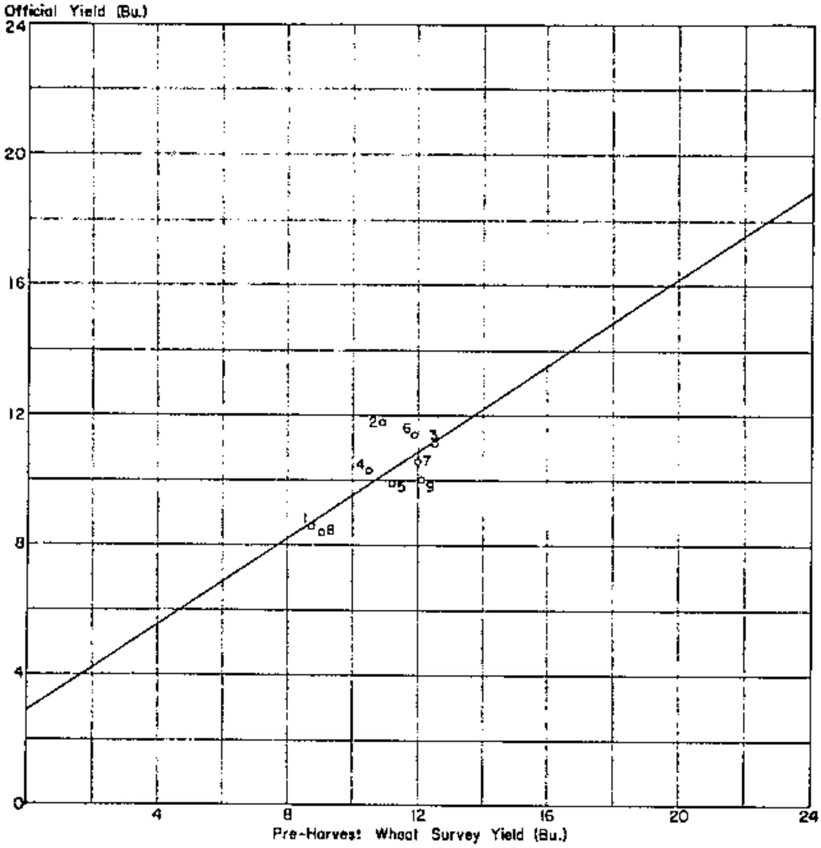


FIGURE 30. Yield of wheat per harvested acre in North Dakota, 1939. Regression of official yield on yield indicated by preharvest survey, by crop reporting districts.

Although wheat seeded in these two States for the 1939 crop had a poorer than usual prospect as it went into the winter, it was relatively in much better condition at this stage of development than the 1940 crop. The 1939 crop, however was forced to early maturity by abnormally high temperatures, at and after filling time. The 1940 crop on the other hand was favored by timely rains and almost ideal temperatures at filling time which permitted a remarkable development (begun in the spring) to continue right up to harvest. Wheat harvest in 1940 was from 7 to 10 days later than in 1939. The difference between growing conditions immediately preceding harvest in the 2 years probably contributed much toward the large differences in the constants and coefficients of the regression equations.

After data for several years have been secured it would seem likely that more accurate forecasts would be made if the *a* constant and *b* coefficient used in forecasting the yield were obtained from an average of several years' regressions. Another possibility for increasing the accuracy of the forecasts would be to include a second variable in the forecasting equation, such as depth of soil moisture.

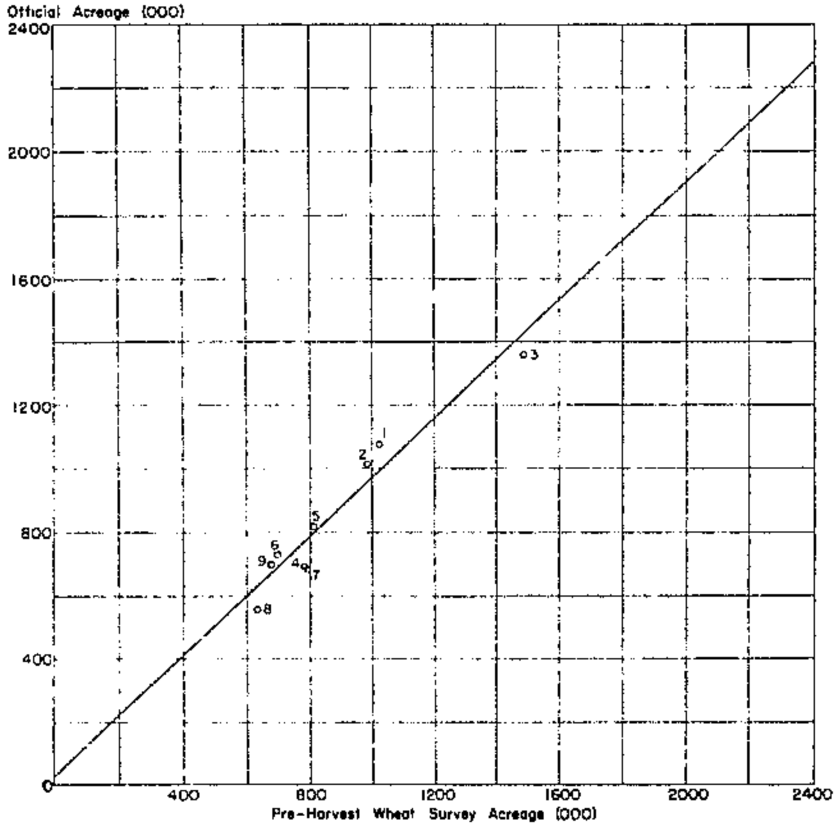


FIGURE 31.—Acreage of wheat harvested in North Dakota, 1939. Regression of official acreage on acreage indicated by preharvest survey, by crop reporting districts.

It is evident from the large differences in the coefficients by States that the data for all the States should not be pooled and a forecast made from a common regression. A more accurate forecast would likely be obtained if based upon a separate regression for each State.

From table 23 it can be seen that the differences between the *b* coefficients obtained from the 1939 and 1940 regressions for both Kansas and Oklahoma are smaller with district variation removed than for the *b* coefficients in the regressions obtained by pooling all sampling units in either State. For Kansas, the difference between the *b*'s in the regressions based upon the data with county variety variation removed is smaller than with only district variation removed while the reverse is true for Oklahoma. In testing the significance of differences between the b's in 1939 and 1940, it was found that those based on the within district regression for Oklahoma and those based on the within county variety group regression for Kansas showed significance at the 5 percent point while the others were highly significant at the 1 percent point. It would seem from this that the regressions in which the differences in the b's were the least

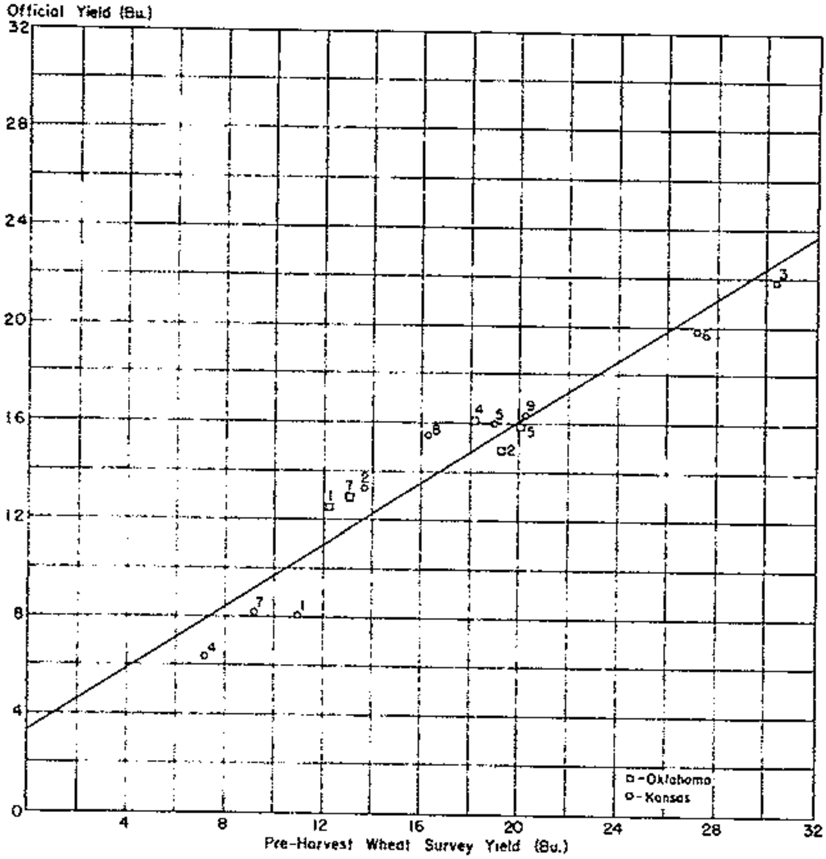


FIGURE 32. Yield of wheat per harvested acre in Kansas and Oklahoma, 1940. Regression of official yield on yield indicated by preharvest survey, by crop reporting districts.

significant might be the best equations for forecasting. The a constant in each equation, however, must be taken into consideration. These constants do not show a consistent change from one kind of regression to another but in any 1 year the larger the b the smaller the a. This must be true as all regressions for any 1 year for each State pass through the intersection of the means of x and y. Since both the a's and b's affect the forecast, a balance between these two must be considered. From a comparison of the different a's and b's in table 23, it is difficult to draw conclusions as to which type of regression would give the best forecast. A more direct way to determine this is to substitute in each of the 1939 regressions the weighted mean number of heads indicated by the sample in 1940. The "forecast" thus obtained can then be compared with the actual yield obtained from the 1940 sample. This comparison would furnish a basis for evaluating the three sets of regressions given in table 23.

In table 24 the 1940 forecast yields based upon the three different kinds of regressions, that is, total, within district, and within county variety groups, are compared with the 1940 actual yields. A comparison is also made between actual sample yields in 1939 and the "forecasts" made for 1939 from the 1940 regression equation and the weighted mean number of heads in 1939 taken from the sample.

TABLE 24.—*Kansas and Oklahoma wheat. Yields from regression equation based on number of heads compared with actual sample yields*

1939 CROP[1]

Regression used in forecasting	Kansas			Oklahoma[2]		
	Actual yield from sample	Yield from regression	Deviation from actual	Actual yield from sample	Yield from regression	Deviation from actual
	Bushels	*Bushels*	*Bushels*	*Bushels*	*Bushels*	*Bushels*
Total	15.0	21.8	6.8	16.3	19.3	3.0
Within district	15.0	20.9	5.9	16.3	19.4	3.1
Within county variety groups	15.0	20.2	5.2	16.3	18.3	2.0

1940 CROP[1]

Regression used in forecasting	Actual yield from sample (Kansas)	Yield from regression (Kansas)	Deviation from actual (Kansas)	Actual yield from sample (Oklahoma)	Yield from regression (Oklahoma)	Deviation from actual (Oklahoma)
Total	15.6	10.8	−4.8	16.7	12.6	−4.1
Within district	15.6	11.2	−4.4	16.7	13.1	−3.6
Within county variety groups	15.6	11.1	−4.5	16.7	12.4	−4.3

[1] Yield forecasts for 1939 were based on regressions from 1940 survey data. Yield forecasts for 1940 were based on 1939 regression equations.
[2] Crop reporting districts 1, 2, 4, and 7 were sampled in both 1939 and 1940.

For Kansas the forecasts based upon the total regression which includes geographic and varietal variation were less accurate than forecasts from either of the other two regressions. For Oklahoma the 1940 regression within county variety groups gave the best "forecast" of the 1939 yield. The best forecast of the 1940 yield, however, was secured from the 1939 within district regression.

It should be noted that the difference between the accuracy of the forecast yields based upon the three kinds of regressions was relatively small compared with the large error resulting from the year to year change in the *a* constant and the *b* coefficient.

The forecasts for 1940 based upon the 1939 regressions were in all three cases below the actual, while the 1939 forecasts based upon the 1940 regressions were above the actual yields, indicating that the regression line was at a higher level in 1940 than in 1939. This might imply that other variables are influencing yield and therefore need to be included in the regression equation. For example, it is believed that available soil moisture would be a useful variable to use in an equation for forecasting wheat yield in the western Great Plains area.

In computing the 1940 forecasts in table 24, the regression within district was based upon an average of individual regressions for each

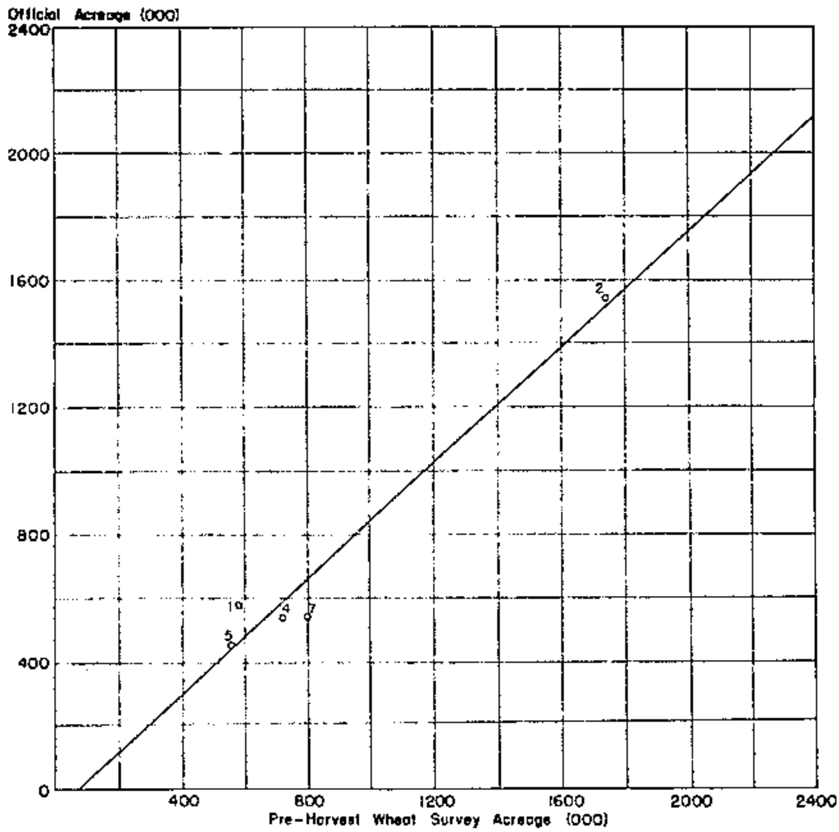


FIGURE 33. Acreage of wheat harvested in Oklahoma, 1940. Regression of official acreage on acreage indicated by preharvest survey, by crop reporting districts.

of the nine districts, and each district regression was weighted according to the number of samples in the districts in 1939. Inasmuch as the number of samples in each district was not proportional to the acreages in the districts, there would be errors in the weights based on the number of samples. The assumption was also made that the district weights did not change from year to year, which was not the case. Therefore, the forecast might be improved by using the separate district regressions and weighting the forecasted yields for each district by the estimated acreages as shown by the sample for the year for which the forecast is made.

For the State of Kansas the district regressions are shown in figures 39 and 40.

In an attempt to determine whether it would be worthwhile to use the individual regressions rather than the regression within districts, a direct substitution of mean number of heads in each district was made in the regressions.

The results are shown in table 25.

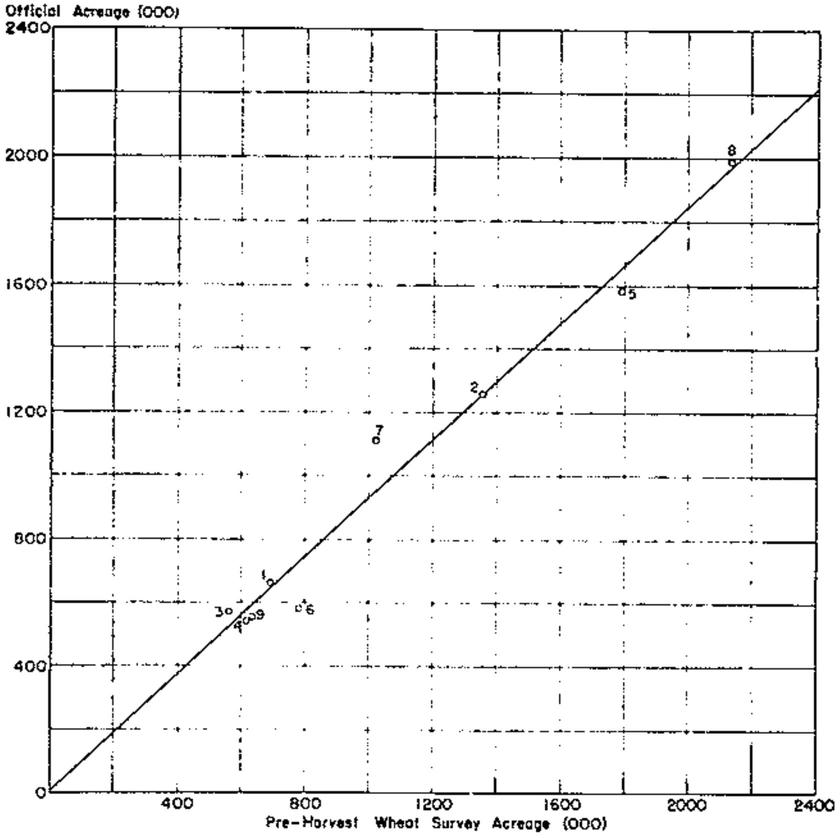


FIGURE 34. Acreage of wheat harvested in Kansas, 1940. Regression of official acreage on acreage indicated by preharvest survey, by crop reporting districts.

In comparing the results of district forecasts obtained by substituting the mean number of heads for each district in both the individual district regression and the State regression with district variation removed, it is noted that in 1939 the individual district regressions give better results in the western and central districts of Kansas with the exception of district 5. The State regression gives better results for the eastern districts and district 5. In 1940 this situation is just the opposite in most districts. In both years, however, the weighted State average of the districts is more accurate when the State regression is used. For Oklahoma three out of the four districts in 1939 show better forecasts from the individual district regression but for 1940 the opposite is true. The weighted State regression in 1939 is closer to the actual yield when using the individual regression whereas in 1940 both give the same results.

TABLE 25.—*Kansas and Oklahoma wheat: Yields from district regression equations based on number of heads compared with actual sample yield*

State and district	1939 crop [1]			1940 crop [1]		
	Actual yield from sample	Yield from individual district regression	Yield from State regression with district variation removed	Actual yield from sample	Yield from individual district regression	Yield from State regression with district variation removed
	Bushels	*Bushels*	*Bushels*	*Bushels*	*Bushels*	*Bushels*
Kansas:						
1	7.6	8.8	16.1	10.1	8.6	12.9
4	5.6	9.4	15.1	8.5	5.3	10.3
7	6.8	9.7	16.1	8.7	5.8	10.0
2	12.7	17.1	20.0	12.5	9.3	10.3
5	14.5	23.1	21.0	16.6	10.8	11.0
8	19.5	22.6	24.2	14.3	11.9	11.0
3	21.3	40.0	25.4	32.4	17.3	14.7
6	20.3	34.9	23.0	27.3	15.4	12.4
9	23.0	29.7	23.3	21.2	15.6	10.9
Weighted State average	15.0	22.0	20.9	15.6	10.7	11.2
Oklahoma: [2]						
1	8.2	14.2	16.7	11.3	6.7	10.6
2	23.0	24.2	21.8	19.2	15.5	12.6
4	17.0	19.4	20.4	17.9	15.6	15.9
7	7.8	11.3	15.4	12.8	9.2	13.7
Weighted State average	16.3	19.1	19.4	16.7	13.1	13.1

[1] Yield forecasts for 1939 were based on regressions from 1940 survey data. Yield forecasts for 1940 were based on 1939 regression equations.
[2] Crop reporting districts 1, 2, 4, and 7 were sampled in both 1939 and 1940.

Since the results are not consistent by districts, no definite evidence is presented that would indicate that one type of regression is superior to the other.

From the graphs it can be seen that in both 1939 and 1940 the regression lines in the three eastern crop reporting districts of Kansas have a steeper slope than those for the three central districts, which in turn have a steeper slope than those for the three western districts. An average regression for each third of the State was obtained as shown in figure 41.

Attention is directed to the fact that in 1940 the regressions for all three sections have a steeper slope and are at a higher level than those for 1939. Since a higher yield is indicated for 1940 than for 1939 for the same number of heads, this might be due to either a larger number of kernels per head or greater weight per kernel. This same observation can be made for eastern Kansas compared with central and western Kansas. To improve the forecasts of yield it might be necessary to predict the number and weight of kernels per head. If it were found that depth of soil moisture earlier in the season or height of plant was correlated with number or weight per kernel, a prediction of these two variables might be made. This prediction might be used in a forecasting equation for yield to give a better

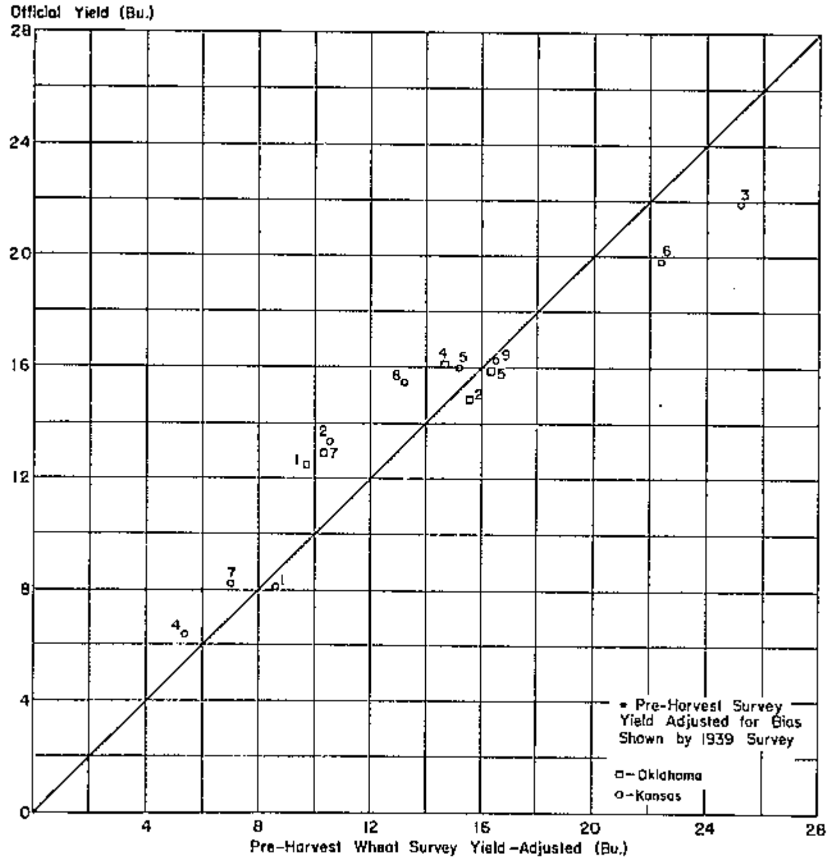


FIGURE 35.—Yield of wheat per harvested acre in Kansas and Oklahoma, 1940. Yields indicated by preharvest survey were adjusted for bias found in yields from the 1939 preharvest survey and then plotted against the official yields, by crop reporting districts.

forecast. However, it would be necessary to test out this hypothesis by obtaining the depth of soil moisture earlier in the season and correlating it with number and weight of kernels.

So far only the effect of number of heads on yield has been discussed in any detail. In table 22, a considerable increase in R^2_2 over r^2_1 due to the addition of another variable, plant height, was shown. To make a forecast of yield based on these two variables, height and number of heads, it would be necessary to use a regression equation such as $y=\bar{y}+b_1(h-\bar{h})+b_2(n-\bar{n})=a+b_1h+b_2n$ obtained by a previous investigation. In the equation, h is the height, n is the number of heads, b_1 is the regression coefficient of yield on height independent of number of heads, b_2 is the regression coefficient of yield on number of heads independent of height, and a is the constant denoting the level of the regression, all obtained in the previous investigation. The weighted mean height and number of heads obtained in the

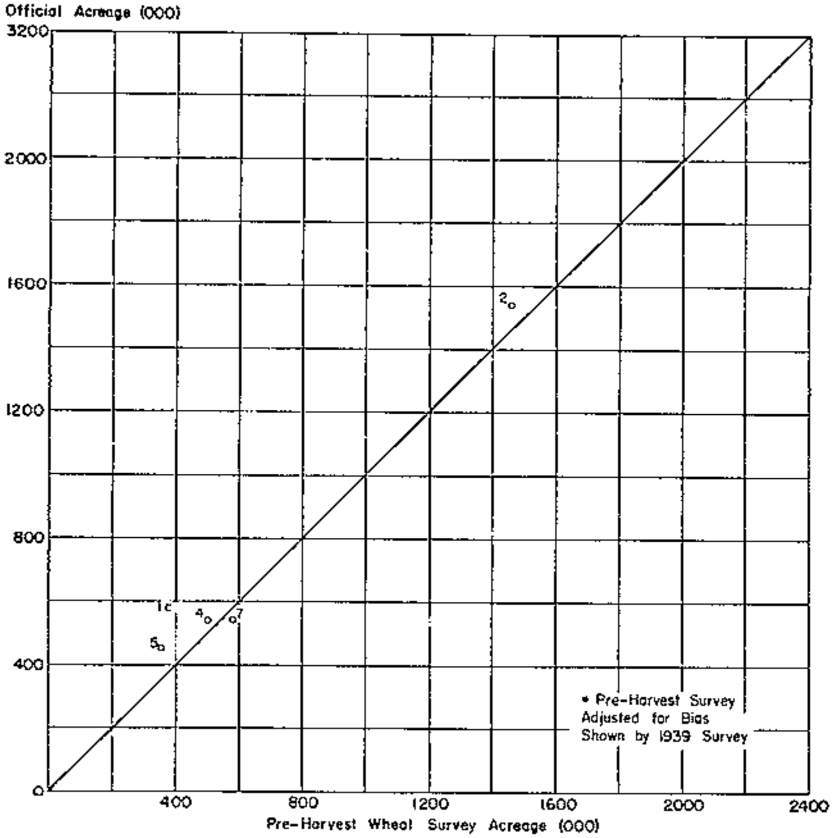


FIGURE 36.—Acreage of wheat harvested in Oklahoma, 1940. Acreages indicated by preharvest survey were adjusted for bias found in acreages from the 1939 preharvest survey and then plotted against the official acreages, by crop reporting districts.

year for which the forecast is to be made would be substituted in the above equation for h and n to get the forecast yield.

By using the same type of regressions discussed in relation to table 23, forecasted yields were obtained and are shown in table 26.

Table 26 does not give much evidence in favor of one type of regression equation over the others. Although the total State regression without any variation removed seems to give the best results for Kansas in both years, the regression with the variation among county variety groups removed differs from the total State regression by only 0.2 bushel for 1939 whereas the regression with district variation removed differs from the total State regression by only

0.2 bushel for 1940. In Oklahoma the regression with county variety group variation removed gave the best by as much as 0.6 bushel over either of the other two in 1939 while in 1940 it again was the best, but by only 0.2 bushel.

TABLE 26.—*Kansas and Oklahoma wheat: Yields from regression equations based on number of heads and height of plants compared with actual sample yield*

1939 CROP[1]

Regression used in forecasting	Kansas			Oklahoma		
	Actual yield from sample	Yield from regression	Deviation from actual	Actual yield from sample	Yield from regression	Deviation from actual
	Bushels	*Bushels*	*Bushels*	*Bushels*	*Bushels*	*Bushels*
Total	15.0	16.1	1.1	16.3	18.4	2.1
Within district	15.0	16.9	1.9	16.3	18.5	2.2
Within county variety groups	15.0	16.3	1.3	16.3	17.8	1.5
1940 CROP[2]						
Total	15.6	14.7	-.9	16.7	14.1	-2.6
Within district	15.6	14.5	-1.1	16.7	14.1	-2.6
Within county variety groups	15.6	13.0	-2.6	16.7	14.3	-2.4

[1] Yield forecasts for 1939 were based on regressions from 1940 survey data.
[2] Yield forecasts for 1940 were based on regressions from 1939 survey data.

By comparing the deviations shown in table 26 with those in table 24, however, it can be seen that considerable improvement was made in the forecast yield obtained from regression equations with two independent variables, height and number of heads, over that obtained from regression equations with number of heads alone. The use of height in addition to number of heads in the forecasting equation definitely appears to be worthwhile.

It must be remembered that the observations on which this study was made were taken very close to harvest so that if it were decided to take samples earlier for forecasting purposes the equation might not give the same results. Observations on height taken as soon as counts on number of heads could be made might very probably be different from those observed in this study. The results would probably be somewhat affected, but whether they would be better or worse can only be decided after actually taking samples at the earlier date and noting the effect.

It was found that the incorporation of a third independent variable, length of head, in the forecasting equation contributed little, if at all, to the accuracy of the forecasts.

Obviously, if a method of forecasting is to be of much value, it would be necessary to be able to reduce or anticipate any bias found in the forecasts. If the regression technique for forecasting yields from plant counts taken prior to harvest is to be most useful in making forecasts of wheat production, it will also be necessary to obtain reliable indications of the acreage at the time the plant observations are secured.

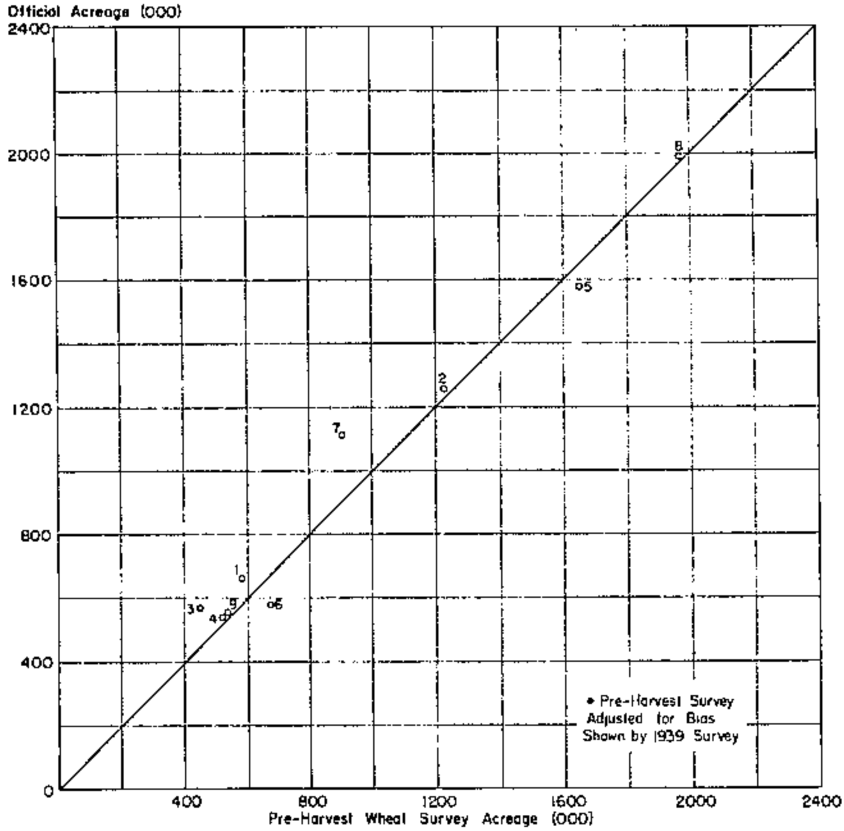


FIGURE 37.- Acreage of wheat harvested in Kansas, 1940. Acreages indicated by preharvest survey were adjusted for bias found in acreages from the 1939 preharvest survey and then plotted against the official acreages, by crop reporting districts.

ACCURACY OF WHEAT PRODUCTION FORECASTS BASED ON PREHARVEST SURVEY DATA

When the data collected at or just prior to harvest in 1939 and 1940 are used in an attempt to learn how accurate production forecasts might be, if field sampling of wheat were conducted about a month before harvest, it is necessary to make three assumptions:

First, that the number of heads that would be counted at flowering time would be the same as the number of heads that would be counted shortly before harvest.

Second, that the acreage metered as wheat for harvest would be the same for a survey made a month before harvest as it would be if the survey were made at harvest.

Third, that height of plant would remain relatively unchanged during the 4 weeks immediately preceding harvest.

The number of heads counted a month before harvest would, of course, exceed the number counted at harvest by the number that

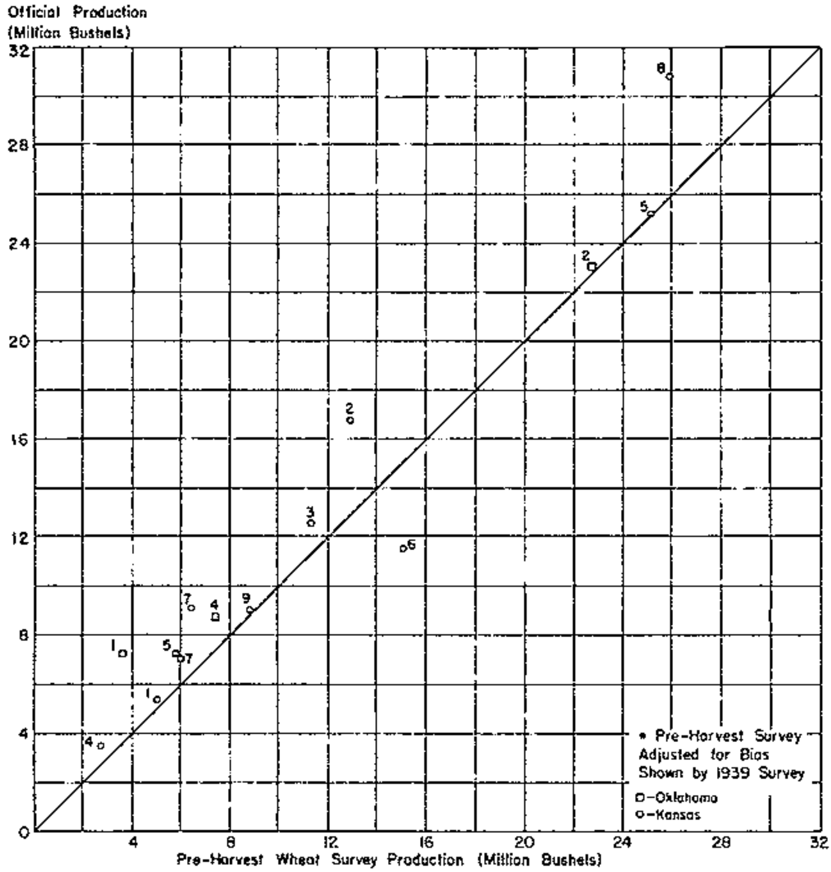


FIGURE 38.—Production of wheat in Kansas and Oklahoma, 1940. Production figures obtained from preharvest survey acreages and yields, both adjusted for bias found in 1939, were plotted against the official estimates of production, by crop reporting districts.

might be destroyed in the interim by grasshoppers, army worms, hail, smut, etc.

The metering of wheat acreage for harvest would also contain more error if it were done a month before harvest than if it were done just prior to harvest; at least, this would be expected to be true in many years in areas with light rainfall, such as western Kansas. Losses from such causes as hail would also increase the error in acreage estimates obtained at the early date.

Height of wheat was found by Florell and Faulkner (*4*) to increase rapidly during heading and to continue until flowering was com-

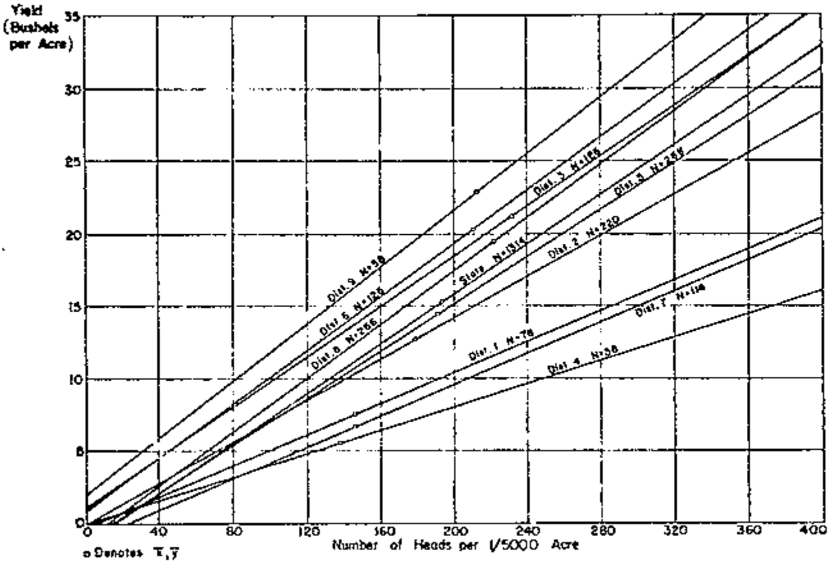


FIGURE 39.—Yield of wheat in Kansas, 1939. Regression of yield on number of heads, by crop reporting districts.

pleted but to increase little or none thereafter. The experiments on which these conclusions were based were conducted in the Northwest. The same results might not be found in the Great Plains area. If the growth pattern in the Great Plains is similar to that in the Northwest, it would seem that samples to forecast yield could be obtained immediately after flowering was completed.

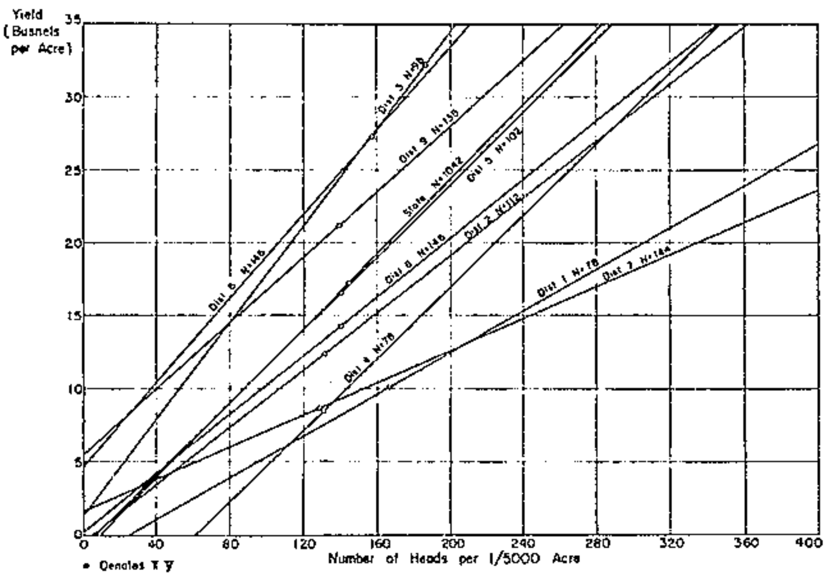


FIGURE 40.—Yield of wheat in Kansas, 1940. Regression of yield on number of heads, by crop reporting districts.

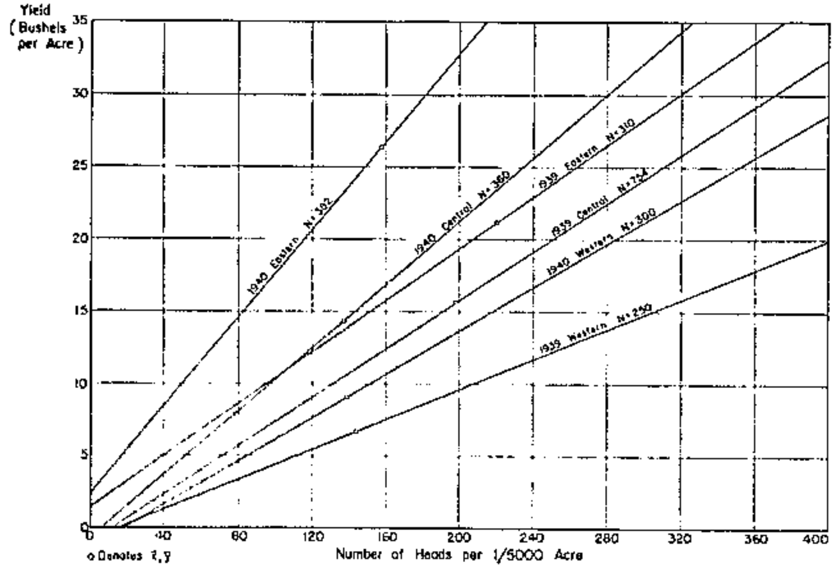


FIGURE 41. Yield of wheat in Kansas, 1939 and 1940. Regression of yield on number of heads, by thirds of the State.

Production "forecasts" from the preharvest surveys are compared with the official figures for Kansas in table 27. The yield forecasts for 1940 were computed from regression equations based on data obtained from the 1939 preharvest field sampling while the 1939 yield "forecasts" were obtained through use of equations based on data from field sampling conducted in 1940.

TABLE 27. *Kansas wheat production, official estimates and indications from survey, 1939 and 1940*

Official estimates and indications from survey	1939		1940	
	Million bushels		*Million bushels*	
Official estimates:				
May 1	146		65	
June 1	106		76	
July 1	111		89	
Aug. 1	116		101	
December	112		124	
	(1)	(2)	(1)	(2)
Indications from regressions, preharvest survey (all samples):				
Combined on State basis	174	138	72	100
With district variation removed	168	143	75	99
With county variety group variation removed	164	139	74	88
Indication from district regressions	169	147	71	94

[1] Based on number of heads.
[2] Based on number of heads and plant height.

In table 27 preharvest survey forecasts of wheat production based on number of heads would be more nearly comparable with the official estimates as of May 1 and June 1 than on other dates because the number of heads could ordinarily be counted throughout Kansas between those two dates. On the other hand, if plant height is to be used along with number of heads to forecast the crop, the data on which a forecast could be based would be gathered at somewhat later dates. Therefore, the forecasts in table 27 based on the two independent variables, number of heads and height, would be more nearly comparable with the June 1 and July 1 official forecasts than any others.

The production figures from the preharvest surveys were obtained by applying a yield, forecast from sample data collected at harvest, to the acreage for harvest from crop meter indications also secured at harvest. Both the yield and the acreage obtained for 1939 were adjusted for the difference between the 1940 sample data and the 1940 official estimates. Likewise, the 1940 indicated acreage and yield were adjusted for the difference between the 1939 sample data and the 1939 official estimates of yield and acreage. This was done in an attempt to allow for bias in the survey indications. After this adjustment, the adjusted acreages were then multiplied by the adjusted yields to obtain the survey production figures shown in table 27.

It can be seen from the table that the survey "forecasts" of the 1939 crop are substantially larger than the official forecasts and also larger than the official estimate in December. It would appear, therefore, that the official forecasts in 1939 were better than forecasts from the preharvest survey would have been. In 1940, the forecasts based on number of heads were lower than the official forecasts made at or prior to harvest, but those based on number of heads and plant height were closer to the final production than the official forecasts.

From the limited amount of concrete evidence available, it would seem that sampling of wheat fields a month before harvest to forecast the size of the crop might have some utility as a supplement to regular crop forecasting methods.

After preharvest survey data have been collected for several years with increased crop meter coverage of the areas sampled and samples obtained from a greater number of fields than in 1939 and 1940, this method of crop forecasting might prove to be much better than is indicated now.

CONCLUSIONS

1. After detailed statistical investigation of the preharvest wheat survey data, it appears that the route sampling of a crop of wheat at or prior to harvest furnishes a useful method for estimating quality and production.

2. As a supplement to present methods of estimating wheat production, the preharvest wheat survey might be expected to promote timeliness, geographic detail, and objectivity in estimating both production and quality of the crop at harvest.

3. The survey method of route sampling will probably be most useful in unusual years when abnormal development of the wheat crop creates production conditions not readily detected by usual crop reporting methods.

4. Relatively small sampling errors were found in the estimates of protein and test weight as compared with the sampling errors found in the yield and acreage estimates.

5. Slight bias is believed to occur in the estimates of protein and test weight because some fields were sampled too long before harvest—that is when the wheat kernel might not yet have indicated its quality as actually shown at harvest.

6. The estimates of yield and acreage were found to contain systematic errors or bias as well as random sampling errors.

7. The random component of error can be reduced sufficiently through improved sampling technique to render this source of error relatively unimportant in the estimates for both quality and production.

8. The systematic error or bias in the yield and acreage estimates presents a more serious problem in that the total amount of bias differs between years and between States in a given year.

9. In order to obtain more accurate estimates from the survey, it will be necessary further to investigate the sources of bias found and establish a more definite basis for removing this component of error from the final estimates.

10. In order to have obtained a maximum amount of information per dollar spent in estimating a production, the sampling interval should have been increased and the total mileage driven should have been increased in each district.

11. Of several attributes of yield examined, number of heads and plant height appear to be most useful in forecasting the yield of wheat from plant observations taken 3 or 4 weeks prior to harvest.

12. The regression equation of yield on number of heads and plant height was not the same in each year. Thus, the usefulness of such an equation for forecasting yield is seriously limited by the failure of the equation found one year to apply to the other year. If it were possible to formulate a specific equation to fit each of various crop conditions as they occur each year, the forecasting of yield from stand and height of grain could be more accurately performed.

LITERATURE CITED

(1) Alsberg, C. L.
1926. PROTEIN CONTENT, A NEGLECTED FACTOR IN WHEAT GRADES. Wheat Studies, Food Research Institute, Stanford, California. 2: No. 4.

(2) Cochran, W. G.
1939. THE USE OF ANALYSES OF VARIANCE IN ENUMERATION BY SAMPLING. Jour. Amer. Stat. Assoc. 34: 492–510.

(3) Coleman, D. A., and Fellows, H. C.
1936. HANDBOOK OF INSTRUCTIONS FOR THE OPERATION OF THE TAG HEPPENSTALL MOISTURE METER. Rev. U. S. Dept. Agr. Bur. Agr. Econ. Washington, D. C.

(4) Florell, V. H., and Faulkner, Boyd.
1931. GROWTH OF WHEAT DURING THE HEADING PERIOD. Jour. Amer. Soc. Agron.

(5) King, A. J., and Jebe, E. H.
1940. AN EXPERIMENT IN PRE-HARVEST SAMPLING OF WHEAT FIELDS. Iowa Agr. Expt. Sta. Res. Bul. 273.

(6) Sherwood, R. C.
1927. SURVEYING THE NEW WHEAT CROP. Jour. Cereal Chem. 4: 395.

(7) Yates, F., and Zacopanay, I.
1935. THE ESTIMATION OF THE EFFICIENCY OF SAMPLING WITH SPECIAL REFERENCE TO SAMPLING FOR YIELD IN CEREAL EXPERIMENTS. Jour. Agr. Sci. 25: 545–577.

APPENDIX

UNITED STATES DEPARTMENT OF AGRICULTURE

Bureau of Agricultural Economics

Special Release KANSAS CITY, MO., *June 24, 1939.*

PREHARVEST WHEAT SURVEY

Test weights ranging from 53 to 62 pounds per bushel and protein content of from 9 to 18 percent, are shown on the first two Kansas districts completed in the preharvest wheat survey being conducted by the United States Department of Agriculture in cooperation with State Agricultural Colleges. Considerable variation in protein content is evident between fields within counties, as well as between samples within a field for the same variety.

In District No. 8, South Central Kansas, comprising the counties of Sumner, Harper, Barber, Comanche, Kiowa, Pratt, Kingman, Sedgwick, Harvey, Reno, Stafford, Edwards, and Pawnee, 134 fields sampled between June 10 and 20, showed an average test weight of approximately 58 pounds per bushel. Forty percent of the fields tested between 59 and 62 pounds, and the remaining 60 percent ranged down to 54 pounds per bushel. Test weights averaged highest in eastern and northern counties of this district. Protein of 57 percent of the fields ranged from 13 to 18 percent. The remaining 43 percent ranged from 9 to 12.75 percent protein (proteins calculated on a 13.5 percent moisture basis). The highest proteins came from the western counties in this district. The grain from 69 percent of the fields sampled was classed as Dark Hard, 18 percent as Hard, 10 percent as Yellow Hard, 2 percent as Soft and 1 percent as Mixed. Several fields had a small mixture of rye.

In District No. 9, Southeastern Kansas, 29 fields were sampled in 11 of the 14 counties on June 15 and 16. Over 75 percent of these fields had a test weight between 57 and 61 pounds per bushel. The remaining 25 percent ranged down to 53 pounds. Approximately 41 percent of the fields showed protein ranging between 13 and 18 percent, with the remaining 59 percent between 9 and 12.75 percent protein. The grain from 28 percent of these fields was classed as Dark Hard, 10 percent as Yellow Hard, 7 percent as Mixed, and 55 percent as Soft. With the exception of three western counties--Butler, Cowley, and Chautauqua—the wheat raised in District No. 9 is predominantly Soft. Yields averaged fairly high over the District as a whole.

Completion of sampling in the western four counties—Alfalfa, Woods, Major, and Woodward—of District No. 2, North Central Oklahoma, showed test weights from 53 to 60 pounds and protein from 9 to 18 percent. Eighty percent of the samples from 50 fields obtained in the four counties ranged from 56 to 60 pounds per bushel. The remaining 20 percent ranged down to 53 pounds. Forty-eight percent of the samples showed a protein content from 13 to 18 percent, and the remaining 52 percent ranged from 9 to 12.75 percent. The grain from 42 percent of the fields was classed as Dark Hard, 26 percent as Hard, 22 percent as Yellow Hard, and 10 percent as Soft. There were no samples from fields in these four counties having hard and soft wheat mixtures, and varietal mixtures in the fields sampled were not as prevalent as in the eastern half of the district. The samples classed as Yellow Hard tested below average in protein. Several fields were badly lodged.

Over 800 samples have been received to date at the Manhattan, Kansas, Laboratory. The field crews have progressed as far northward as Central Kansas. Appreciable loss from hail damage is evident in some fields, and combine harvesting in some areas has been delayed by local rains.

Distributed by Branch Market News Office, Bureau of Agricultural Economics, 1513 Genesee Street, Kansas City, Mo. Don E. Rogers, local representative.

UNITED STATES DEPARTMENT OF AGRICULTURE

AGRICULTURAL MARKETING SERVICE

Special Release No. 6 KANSAS CITY, MO., *June 26, 1940.*

PREHARVEST WHEAT SURVEY

Wheat samples obtained by preharvest wheat survey crews of the Agricultural Marketing Service in Garfield, Grant, Kay, and Noble Counties in North Central Oklahoma, show an *average protein content of 13.2 percent*, with 67 percent of the field samples falling between 11.7 and 14.7 percent protein. The *test weight averaged 57.9 pounds per bushel* with 67 percent of the samples falling between 56.0 and 59.8 pounds per bushel. In these counties in 1939, the protein content averaged 12.0 percent and the test weight averaged 58.7 pounds per bushel. In 1939, 67 percent of the fields fell between 10.3 percent and 13.7 percent protein, and 57.1 to 60.3 pounds test weight.

In District 4, West Central Oklahoma, the ripening of the crop is progressing slowly. Custer and Blaine Counties have been sampled by field crews.

Garfield, Grant, Kay, and Noble Counties were sampled June 17 to 20. Custer and Blaine Counties were sampled June 19 to 22. Details by counties are shown below.

OKLAHOMA

District and county	Number of 1940 samples	Average protein		Average test weight per bushel		Percent in each class				
		1940	1939	1940	1939	Dark Hard	Hard	Yellow Hard	Soft	Mixed
District 2:										
Grant	44	14.6	11.0	56.3	59.2	95	5	0	0	0
Garfield	60	13.6	12.2	58.2	59.1	73	0	0	17	10
Kay	41	12.0	12.4	58.2	57.8	41	9	0	36	14
Noble	58	12.8	13.1	58.5	57.9	62	8	0	10	20
District 4:										
Blaine	48	13.5	13.3	58.9	59.5	92	2	0	4	2
Custer	64	13.2	14.1	59.5	58.7	91	5	1	3	0

The samples indicated that stem rust was prevalent in these counties. Estimates of the percentage of the total culm surface covered by pustules was made by the field crews at the time of sampling. For the individual counties the average percent of total culm surface covered by pustules is Garfield 5 percent, Grant 8 percent, Kay 4 percent, Noble 7 percent, Custer 1.5 percent, and Blaine 16 percent.

Distributed by Branch Market News Office, Agricultural Marketing Service, 325 United States Court House, Kansas City, Mo. Don E. Rogers, local representative.

U. S. DEPARTMENT OF AGRICULTURE
Agricultural Marketing Service

KANSAS STATE BOARD OF AGRICULTURE
Division of Statistics

OFFICE OF THE AGRICULTURAL STATISTICIAN

207 Federal Building, Topeka, Kans.

JULY 26, 1940.

KANSAS 1940 PREHARVEST WHEAT SURVEY [1]

SUMMARY REPORT

An average protein content of 13.0 percent and an average test weight of 58.2 pounds per bushel were shown by 2,206 samples of wheat taken from fields in Kansas this season just prior to harvest. In 1939 the protein content averaged 14.0 percent and the average test weight was 57.0 pounds per bushel. From the table of district averages for the 2 years it can be seen that the differences between the 2 years are not uniform by areas. In the eastern and southern parts of the State the protein averaged lower in 1940, while in the western and north central parts protein content was about the same. The test weight in the southwest and west central districts averaged approximately 4 pounds per bushel lower in 1940. In the eastern two-thirds of Kansas the test weight was higher in 1940, ranging from 4.5 pounds per bushel higher in the north central to 0.6 pounds per bushel higher in the south central counties.

The 1940 samples show that Blackhull and Turkey varieties continue to predominate in western Kansas. In the central third of the State the 1940 samples show a larger proportion of the crop grown to Tenmarq than in 1939. Tenmarq, Blackhull, and Turkey constitute approximately 85 percent of the wheat produced in the central third of the State. Kawvale, the leading single variety in eastern Kansas, constitutes about one-third of the total production in this section of the State. While the proportion of Kawvale has increased in the east central and southeast districts, the 1940 samples show that for the northeastern district the proportion of Tenmarq exceeds that of Kawvale.

The attached maps and tables summarize the quality and variety information from the 2,206 samples of wheat obtained by 8,089 miles of travel on the preharvest wheat survey in Kansas. The object of the survey has been to supply wheat growers, dealers, millers, and others at harvest and prior to marketing, with information as to the relative quality of the wheat crop grown in different localities. This report summarizes information released as the harvest progressed.

PREHARVEST WHEAT SURVEY PROCEDURE

The samples have been obtained by four field crews, each crew consisting of two men to a car. The crews started sampling in Labette County, Kans., on June 20, and continued sampling as they moved north with the harvest. The routes driven were well scattered over each county. Each car was equipped with a crop meter, which was used to measure the total miles driven and the frontages of the wheat fields bordering the roads driven. A fixed rate of sampling was maintained in each county, such as every 2 miles of wheat as indicated by the crop meter, so it was possible to distribute the sampling according to wheat acreage. Since the total land in a county is known, and the crop meter gives an indication of the ratio of wheat to total land, it was possible to obtain an indication of the wheat acreage by multiplying the total land by the wheat ratio obtained by the crop meter. Two samples, $\frac{1}{5000}$ of an acre in size, were selected at random from each field sampled. These samples gave a basis for calculating the yield per acre, the percent protein, test weight per bushel, as well as other factors.

As soon as the samples were obtained, they were placed in envelopes and mailed to a central laboratory where counts, measurements, weighings, and chemical tests were made. To obtain a more accurate indication of the dis-

[1] The 1940 preharvest wheat survey in Kansas was made possible through the combined efforts of the Agricultural Marketing Service, the Kansas State Board of Agriculture, and the Kansas State College.

tribution of varieties in different areas, extra samples were taken between the fields sampled for both quality and variety.

The percent protein for each sample was tested according to the Kjeldahl-Gunning-Arnold procedure, using a one-gram sample, the nitrogen content being converted to protein by means of the factor 5.7. All protein results were reported on the basis of 13.5 percent moisture (Tag-Heppenstall) in the wheat.

A wide spread in the dates of seeding last Fall, unfavorable weather conditions during the Fall and early Winter, a wet, cool Spring, coupled with black stem rust in some fields resulted in great variability in yield per acre, percent protein content, and test weight per bushel between counties, between fields, and within fields. The large number of samples obtained should, however, insure reliable indications. In a few counties the fields did not ripen at the same time and a few samples obtained were slightly immature, which may possibly result in the samples indicating a low test weight. Any bias due to immaturity is confined to a few fields, because the number of immature fields sampled was too small to affect the quality indications materially.

Kansas—District and State summary of average protein content and average test weight in 1940 and 1939 samples

District	Number of samples		Average protein content		Average test weight	
	1940 [1]	1939	1940	1939	1940	1939
Northwest	159	78	16. 7	16. 7	54. 9	55. 4
West Central	169	64	16. 6	16. 8	52. 4	56. 2
Southwest	298	110	16. 4	16. 5	54. 0	58. 3
North Central	234	232	14. 7	14. 2	60. 1	55. 6
Central	220	272	12. 7	14. 8	57. 9	56. 5
South Central	312	268	12. 0	13. 0	58. 5	58. 0
Northeast	208	126	11. 8	13. 1	60. 6	56. 0
East Central	316	126	11. 8	13. 5	58. 6	56. 2
Southeast	290	58	11. 5	12. 7	58. 4	57. 2
State	2, 206	1, 334	13. 0	14. 0	58. 2	57. 0

[1] Does not include 973 additional samples obtained to assist in determining the distribution of varieties in the different areas.

KANSAS PREHARVEST WHEAT SUMMARY—1940

Percent protein and average test weight per bushel of wheat as sampled—county and district summary 1940

County and district	Route miles driven	Number of samples	Protein: Percent average	Protein: 67 percent [1] range, Lower	Protein: 67 percent [1] range, Upper	Test weight: Pounds per bushel average	Test weight: 67 percent [1] range, Lower	Test weight: 67 percent [1] range, Upper
Cheyenne	70	30	16.0	14.4	17.6	57.8	54.8	60.8
Decatur	66	15	17.9	16.7	19.1	55.3	54.4	56.2
Graham	66	24	17.2	16.2	18.2	50.9	48.9	52.9
Norton	66	12	17.5	15.9	19.1	51.9	47.0	56.8
Rawlins	70	12	17.9	16.8	19.0	53.5	51.0	56.0
Sheridan	58	18	17.9	17.3	18.5	51.3	49.6	53.0
Sherman	69	28	17.4	16.0	18.8	54.6	52.7	56.5
Thomas	77	20	16.7	15.3	18.1	52.1	48.0	56.2
Northwest	542	159	[2] 16.7			[2] 54.9		
Gove	70	21	18.4	17.7	19.1	50.6	49.2	52.0
Greeley	58	12	17.9	16.2	19.6	54.4	51.8	57.0
Lane	83	28	16.4	15.5	17.3	50.6	46.5	54.7
Logan	110	10	17.5	16.8	18.2	50.1	48.0	52.2
Ness	94	26	16.4	15.5	17.3	51.9	41.9	61.9
Scott	77	20	16.4	14.6	18.2	53.9	48.3	59.5
Trego	60	24	17.2	16.0	18.4	51.4	49.2	53.6
Wallace	65	6	17.7	16.0	19.4	55.4	52.3	58.5
Wichita	63	22	16.2	14.8	17.6	55.4	49.7	61.1
West Central	680	169	[2] 16.6			[2] 52.4		
Clark	62	8	15.1	14.2	16.0	55.5	51.7	59.3
Finney	96	18	15.4	12.8	18.0	55.6	50.2	61.0
Ford	93	40	16.0	15.0	17.0	50.6	46.0	55.2
Grant	48	22	17.3	16.7	17.9	54.0	50.2	57.8
Gray	72	30	16.9	16.1	17.7	52.5	48.5	56.5
Hamilton	83	16	17.3	16.3	18.3	53.7	50.0	57.4
Haskell	53	28	16.8	15.3	18.3	54.1	50.2	58.0
Hodgeman	65	18	16.2	15.2	17.2	48.6	43.9	53.3
Kearny	74	8	17.0	15.9	18.1	55.4	52.2	58.6
Meade	67	22	16.0	15.0	17.0	55.3	53.3	57.3
Morton	45	4	[3]			[3]		
Seward	56	28	16.7	15.1	18.3	54.9	53.3	56.5
Stanton	45	36	17.0	16.2	17.8	54.2	51.3	57.1
Stevens	17	20	17.1	16.3	17.9	56.6	54.0	59.2
Southwest	906	298	[2] 16.4			[2] 54.0		
Clay	75	16	13.6	11.7	15.5	61.2	60.0	62.4
Cloud	85	20	14.8	12.4	17.2	60.7	58.4	63.0
Jewell	113	28	15.5	13.3	17.7	57.8	55.4	60.2
Mitchell	72	26	15.0	13.8	17.2	59.9	57.5	62.3
Osborne	81	28	17.0	15.6	18.1	56.3	53.3	59.3
Ottawa	111	28	14.2	11.9	16.5	61.3	60.3	62.3
Phillips	76	12	17.6	16.1	19.1	51.4	49.6	53.2
Republic	81	18	14.2	11.5	16.9	61.3	59.7	62.9
Rooks	63	22	16.5	15.1	17.6	51.2	52.0	56.4
Smith	108	16	17.2	15.4	19.0	54.3	49.9	58.7
Washington	103	20	13.9	11.6	16.2	61.1	59.0	63.2

See footnotes at end of table.

KANSAS PREHARVEST WHEAT SUMMARY—1940—Continued

Percent protein and average test weight per bushel of wheat as sampled—county and district summary 1940—Continued

County and district	Route miles driven	Number of samples	Protein: Percent average	Protein: 67 percent [1] range: Lower	Protein: 67 percent [1] range: Upper	Test weight: Pounds per bushel average	Test weight: 67 percent [1] range: Lower	Test weight: 67 percent [1] range: Upper
North Central	968	234	[2] 14.7			[2] 60.1		
Barton	96	24	15.8	13.4	18.2	56.5	53.2	59.8
Dickinson	91	18	12.5	11.0	14.0	59.3	55.5	63.1
Ellis	105	26	16.6	15.0	18.2	52.2	48.2	56.2
Ellsworth	105	16	15.7	14.1	17.3	56.6	51.9	61.6
Lincoln	72	16	18.0	16.8	19.2	56.2	53.3	59.1
McPherson	116	26	10.4	9.0	11.8	57.8	54.7	60.9
Marion	86	18	10.8	8.8	12.8	58.5	54.5	62.5
Rice	105	22	12.9	10.8	15.0	57.0	53.1	60.9
Rush	62	10	17.3	16.2	18.4	58.0	55.7	60.3
Russell	105	18	15.7	14.9	16.5	58.8	56.2	61.4
Saline	91	26	14.5	12.4	16.6	59.8	58.0	61.6
Central	1,034	220	[2] 12.7			[2] 57.9		
Barber	105	42	14.1	12.5	15.7	57.7	54.6	60.8
Comanche	73	22	14.7	13.5	15.9	51.5	47.6	55.4
Edwards	76	12	14.8	13.3	16.3	50.7	47.8	53.6
Harper	102	46	13.0	10.9	15.1	59.1	56.2	62.0
Harvey	85	18	11.0	9.8	12.2	61.1	59.9	62.3
Kingman	101	22	12.0	9.9	14.1	57.0	53.3	60.7
Kiowa	73	18	15.3	13.6	17.0	50.8	47.3	54.3
Pawnee	76	12	15.4	12.6	18.2	53.8	48.2	59.4
Pratt	93	18	12.8	11.5	14.1	52.0	45.7	58.3
Reno	109	22	11.7	9.6	13.8	59.8	55.9	63.7
Sedgwick	134	24	10.2	9.4	11.0	59.6	57.7	61.5
Stafford	73	16	12.6	10.3	14.9	57.5	53.7	61.3
Sumner	153	40	12.7	11.2	14.2	58.1	54.3	61.9
South Central	1,253	312	[2] 12.0			[2] 58.5		
Atchison	60	16	12.0	10.3	13.7	63.4	61.8	65.0
Brown	92	18	12.8	11.0	14.6	60.6	57.9	63.3
Doniphan	75	8	12.3	9.5	15.1	60.0	55.5	64.5
Jackson	81	20	13.6	11.8	15.4	60.2	57.6	62.8
Jefferson	52	22	11.1	8.8	13.4	57.4	52.1	62.7
Leavenworth	49	22	10.6	8.9	12.3	59.5	56.1	62.9
Marshall	98	18	12.8	10.7	14.9	62.3	60.6	64.0
Nemaha	91	20	12.8	10.3	15.3	61.4	58.9	63.9
Pottawatomie	126	30	11.4	9.6	13.2	60.1	58.0	62.2
Riley	61	30	11.5	9.9	13.1	60.7	59.3	62.1
Wyandotte	30	4	[3]			[3]		

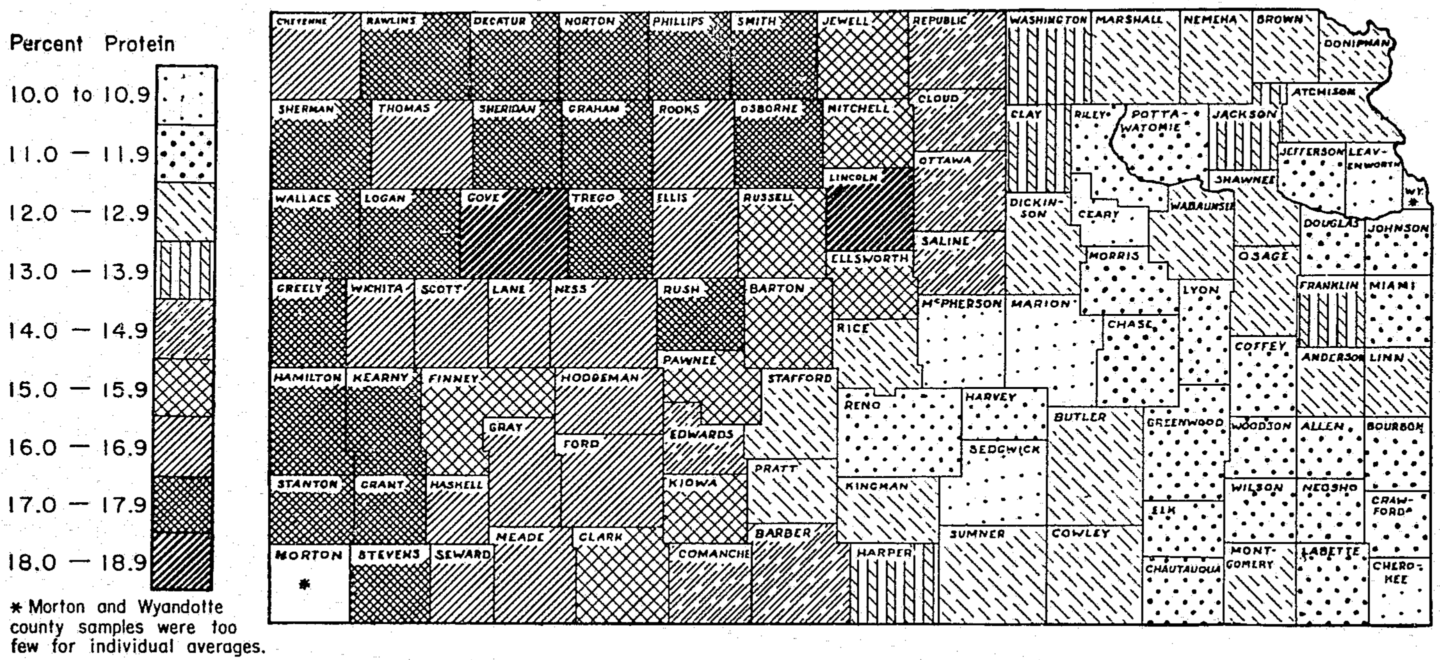
See footnotes at end of table.

KANSAS PREHARVEST WHEAT SUMMARY—1940—Continued

Percent protein and average test weight per bushel of wheat as sampled—county and district summary 1940—Continued

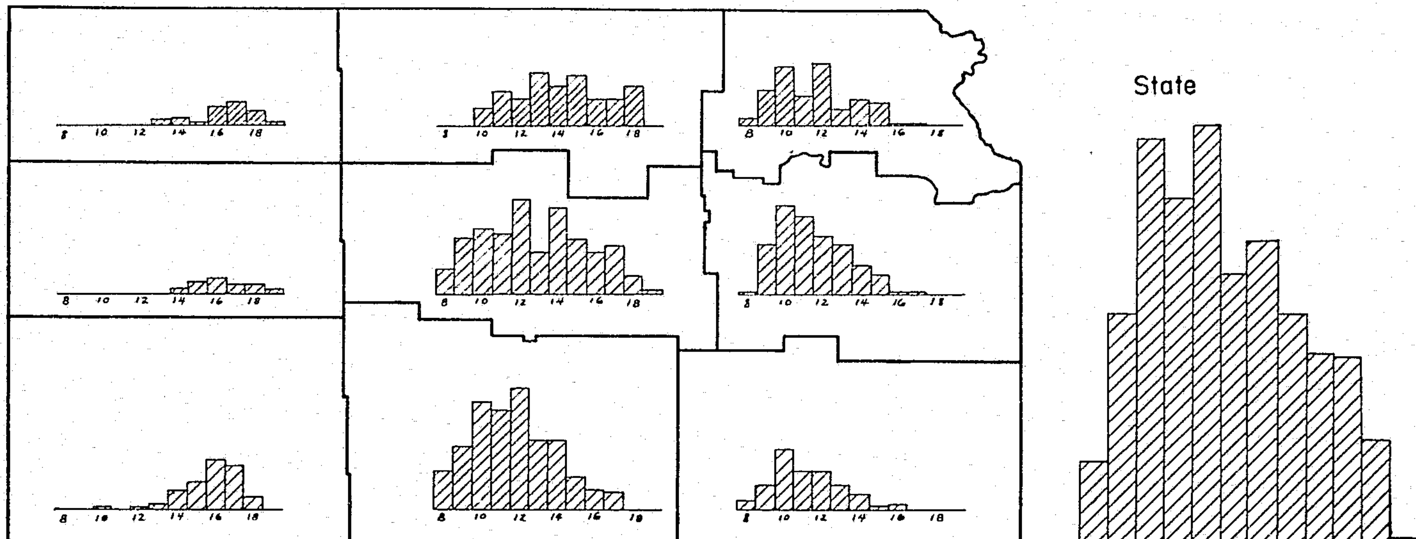
County and district	Route miles driven	Number of samples	Protein: Percent average	Protein: 67 percent [1] range, Lower	Protein: 67 percent [1] range, Upper	Test weight: Pounds per bushel average	Test weight: 67 percent [1] range, Lower	Test weight: 67 percent [1] range, Upper
Northeast	815	208	[2] 11.8	------	------	[2] 60.6	------	------
Anderson	52	24	12.4	10.5	14.3	58.4	55.4	61.4
Chase	71	36	11.4	10.0	12.8	58.6	53.9	63.3
Coffey	44	26	11.2	9.1	13.3	58.7	55.6	61.8
Douglas	56	18	11.1	9.5	12.7	56.7	50.2	63.2
Franklin	66	26	13.0	10.6	15.4	53.9	47.9	59.9
Geary	36	24	10.7	9.7	11.7	61.1	60.2	62.0
Johnson	51	20	11.5	10.4	12.6	60.4	58.1	62.7
Linn	65	26	12.3	10.5	14.1	58.6	56.0	61.2
Lyon	62	22	11.6	10.0	13.2	57.8	53.1	62.5
Miami	61	22	11.4	9.9	12.9	60.4	58.1	62.7
Morris	48	14	11.7	9.7	13.7	60.6	59.4	61.8
Osage	71	18	12.3	11.0	13.6	54.7	49.3	60.1
Shawnee	77	24	12.9	10.6	15.2	58.4	56.5	60.3
Wabaunsee	57	16	12.4	10.4	14.4	60.2	58.3	62.1
East Central	817	316	[2] 11.8	------	------	[2] 58.6	------	------
Allen	61	14	11.2	9.4	13.0	60.1	57.5	62.7
Bourbon	72	12	11.7	10.2	13.2	59.5	55.8	63.2
Butler	122	22	12.1	10.9	13.3	60.1	56.3	63.9
Chautauqua	62	14	11.3	10.3	12.3	56.2	51.7	60.7
Cherokee	76	28	10.0	9.0	11.0	58.6	57.6	59.6
Cowley	94	34	12.0	9.7	14.3	59.4	57.1	61.7
Crawford	65	18	11.7	9.9	13.5	59.3	57.2	61.4
Elk	65	18	11.2	9.7	12.7	59.4	57.2	61.6
Greenwood	94	16	11.5	10.0	13.0	58.4	55.9	60.9
Labette	91	20	11.3	10.9	11.7	57.4	55.3	59.5
Montgomery	64	22	12.7	11.0	14.4	55.9	51.7	60.1
Neosho	67	20	11.0	10.3	11.7	58.3	55.8	60.8
Wilson	64	22	11.9	10.4	13.4	56.4	51.7	61.1
Woodson	77	30	11.8	9.5	14.1	57.0	52.5	61.5
Southeast	1,074	290	[2] 11.5	------	------	[2] 58.4	------	------
State	8,089	2,206	[2] 13.0	------	------	[2] 58.2	------	------

[1] In the sample approximately 67 percent of the fields fall between the upper and lower range as shown in the table.
[2] District average is from county averages weighted by production.
[3] Too few fields for reliable county average.



PROTEIN CONTENT OF WHEAT, KANSAS 1940

As indicated by preharvest survey conducted by Agricultural Marketing Service of the United States Department of Agriculture.

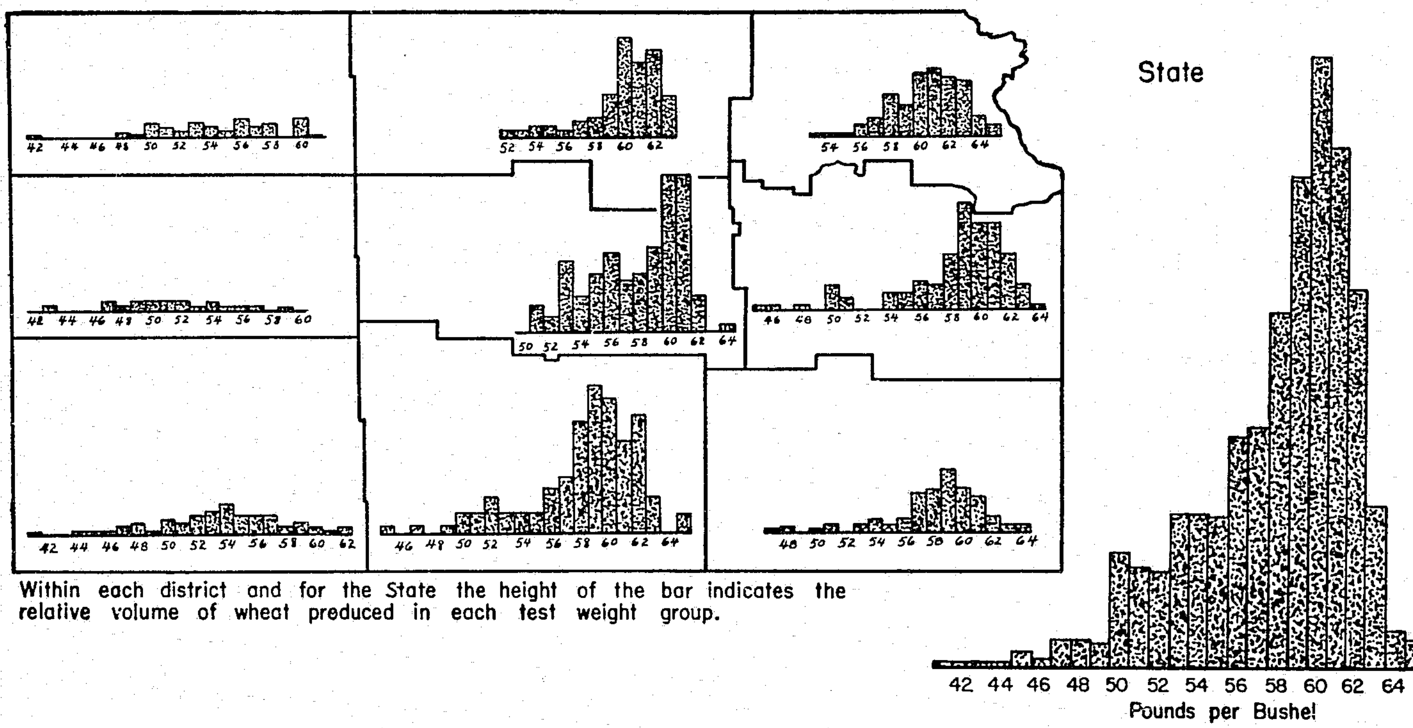


Within each district and for the State the height of the bar indicates the relative volume of wheat produced in each protein group.

Percent Protein (Range)	8.0 8.9	9.0 9.9	10.0 10.9	11.0 11.9	12.0 12.9	13.0 13.9	14.0 14.9	15.0 15.9	16.0 16.9	17.0 17.9	18.0 18.9	19.0 19.9
Percent of Total	3.0	8.3	14.5	12.4	15.0	9.7	10.9	8.3	6.9	6.8	3.8	0.4

WHEAT PROTEIN CONTENT DISTRIBUTION, KANSAS 1940

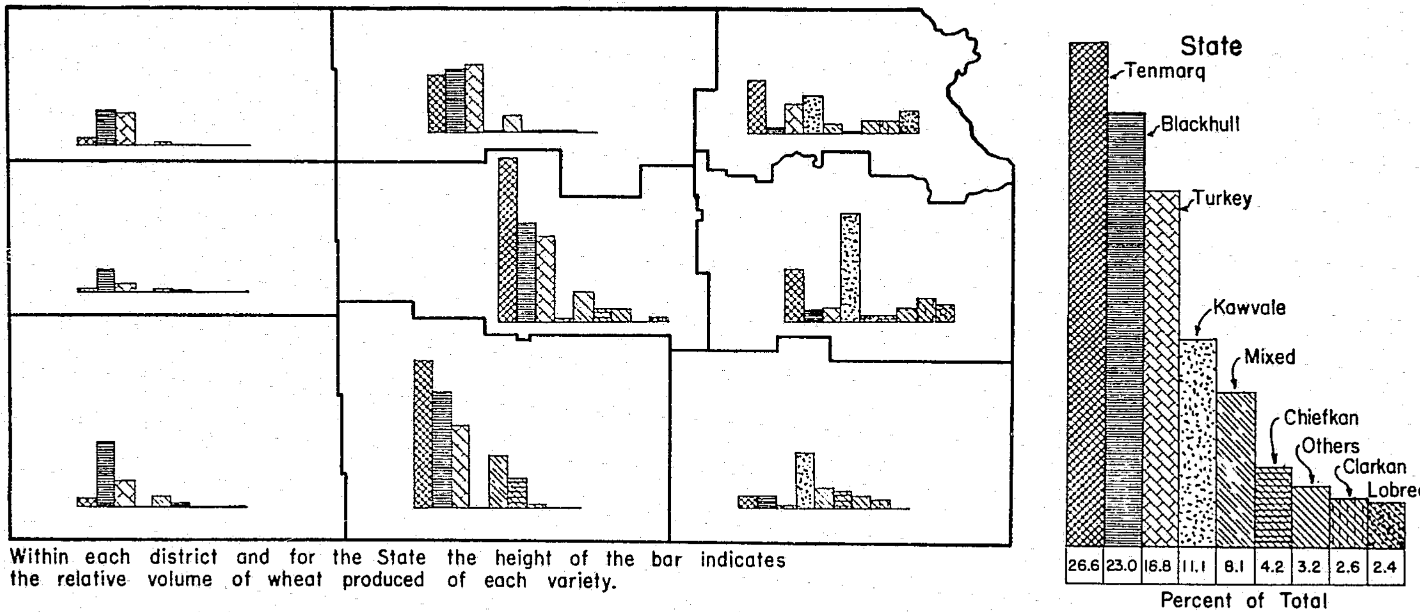
As indicated by preharvest survey conducted by Agricultural Marketing Service of the United States Department of Agriculture.



Within each district and for the State the height of the bar indicates the relative volume of wheat produced in each test weight group.

WHEAT TEST-WEIGHT DISTRIBUTION, KANSAS 1940

As indicated by preharvest survey conducted by Agricultural Marketing Service of the United States Department of Agriculture.



Within each district and for the State the height of the bar indicates the relative volume of wheat produced of each variety.

WHEAT VARIETY DISTRIBUTION, KANSAS 1940

As indicated by preharvest survey conducted by Agricultural Marketing Service of the United States Department of Agriculture.

U. S. GOVERNMENT PRINTING OFFICE: 1942

END