A COMPARISON OF ALTERNATIVE WINTER WHEAT OBJECTIVE YIELD UNIT SIZES

Research and Development Branch
Research Division
Statistical Reporting Service

September 1973

CONTENTS

ACKNOWLEDGEMENTS

I want to thank the following persons for their assistance and guidance in this research: Richard D. Allen, Harold F. Huddleston, Wendell W. Wilson, and Ronald A. Wood. I would also like to thank Mary Ann Higgs for her expert typing of this report.

SUMMARY

Two alternate rigid steel frames of different lengths (14.4 inches and 21.6 inches) were compared to the 26.1 inch. The comparisons were based on counts of number of stalks, emerged heads, heads in late boot, and stalks greater than ten inches tall within each count unit. Simultaneous paired T^2 indicated that there was no significant differences between the 26.1 inch frame and the two alternate frame sizes for any of the variables observed (stalks, emerged heads, etc.). The sample was based on a subsample of field units drawn from the December 1971 and June 1972 Enumerative Surveys in Kansas and Washington. In addition, there was no significant difference between the 26.1 inch frame and the two alternate frame sizes for threshing percent and gross yield. The 21.6 inch frame took approximately 20 percent less time than the 26.1 inch frame to make counts within the unit. The 14.4 inch unit saved 40 percent when compared to the 26.1 inch frame.

The analysis for a single season in these two States indicates even smaller plots size can be used than is presently employed. An optimum allocation analysis is given in the appendices. The optimum allocation analysis is presented for two of the independent variables (stalks, emerged heads) because of their importance in the Wheat Forecasting Model(s). For these two variables either 1 or 2 is the "best" allocation for units per field. But since we desire an estimate of within field variance, we suggest 2 units per field. The allocation of rows per unit was more variable (between 1 and 5) based on just frequency two would be the "best" allocation. Appendix II presents an analysis of the optimum frame length. The 14.4 inch frame was found to be the optimum. Appendix III investigates the possible biases in the three plot sizes. Based on our data, no significant biases were found for estimating number of stalks or number of emerged heads.

A Comparison of Alternative Winter Wheat Objective Yield Unit Sizes

By

Chapman P. Gleason

INTRODUCTION

Results of an optimum plot size study conducted by the Research and Development Branch [10] during 1971 indicated that a count unit shorter than the standard 26.1 inches would be more efficient. The study indicated that the optimum number of rows per unit should be three, the number included in the present objective yield plot.

To investigate the validity of the shorter unit and to test any problems which would be encountered with a smaller unit, two new frames were developed for testing during 1972. One of these frames measured 21.6 inches, the other 14.4 inches. These particular exact lengths were chosen since they give 5 row equivalent measurements of 3.0 and 2.0 feet, respectively, if a plot is established in broadcast wheat. The standard frame of 26.1 inches provides a 5 row equivalent of 3.3 feet.

The new frames differ in construction from the existing frame in operational use. Each new frame is a single piece of steel made from 3/8" steel stock with an arm at each end approximately four inches long. The present frame has removable arms about 27 inches long which are bolted to a steel back piece. These long arms can be bent from the perpendicular, and can bias our estimate if they are not frequently checked and measured. This is one of the disadvantages of the present frame construction.

Thirty-seven fields were selected in the State of Washington and forty fields were selected in Kansas from the regular objective yield survey fields. State supervisors selected at least five enumerators in each State to work on the special frame size study. The enumerators were selected based on their past performance of high quality work and to represent all major winter wheat producing areas of the State. Enumerator workloads were determined by the State supervisor based on the additional work required and the enumerator's responsibilities.

FIELD PROCEDURES

In order to avoid affecting results in the regular objective yield program, the current 26.1 inch frame unit was laid out first in every unit. On the first visit, the regular unit and clip areas were laid out and marked and all usual counts made. A buffer zone was established between the 26.1 frame clip area and the "new" frames. The 21.6 unit was laid out and all counts made before laying out the 14.4 unit. Figure 1 depicts the unit 1 layout.

The new frames must be laid down in each row separately to determine the endpoints of the row. The present frame is slid across all three rows at the same time. This is the only difference in field procedures between the two types of frames.

Times were recorded separately for the three units on each visit. Data on time required per unit allows evaluation of possible savings by use of smaller units.

Each of the three units was visited during monthly survey periods. At harvest, heads were clipped for each size of unit and sent to the regional laboratories. The regular laboratory form (C-2) was used by the regional laboratory for each unit.

Figure 1.--Unit layout, winter wheat, Kansas and Washington, 1972

Count area 26.1"	Clip area 26.1"	Clip area 26.1"	Buffer zone 26.1"	Special count area 21.6"	Buffer zone 26.1"	Special count area 14.4"
Regular objective yield unit				The two special units		

THE ANALYSIS

Analysis Assumptions

Each of the frame sizes was assumed to be a fixed treatment applied to the population of wheat fields in Kansas and Washington. All counts and weights were converted to per acre equivalents in order to properly compare treatment (frame size) effects. The present 26.1 inch frame was considered to be the standard or control and results of each of the other two were compared with the control, see Dunnett [4] for a discussion.

The units observed were assumed to be a random sample from a population of all possible units for calculation of means, variances and covariances. The observations were paired comparisons, since the three different units were laid out in the same field within approximately 0, 8-2/3 and, 12-2/3 feet of a random starting point. It is immaterial whether a new frame gave exactly the same expanded results for the kth unit as the control but whether the new frame estimates the same population parameters as the control.

The Analysis of Monthly Counts

As many as four different correlated variables are measured on each unit for each survey visit. The use of multiple paired t tests would increase the overall Type I error to an unacceptable .18 on individual t tests done at the .05 level, when four characteristics are measured. That is, if four univariate paired t tests were performed with $\alpha = .05$, the probability is .18 that you reject at least one of the four hypotheses even if all hypotheses are true. A test which does not have this drawback is the paired T^2 test. Kramer [6] has a discussion of the T^2 test. The T^2 statistic is a multivariate generalization of Student's t statistic.

Let $\mu^{(1)}$, $\mu^{(2)}$, and $\mu^{(3)}$ be the population mean vectors for the 26.1, 21.6 and the 14.4 inch units, respectively, for various plant counts. Thus, for example,

$$\mu^{(1)} = \begin{bmatrix} \text{mean no. stalks/acre, 26.1 inch unit} \\ \text{mean no. stalks>10''/acre, 26.1 inch unit} \\ \text{mean no. emerged heads/acre, 26.1 inch unit} \\ \text{mean no. head in late boot/acre, 26 inch unit} \end{bmatrix}$$

Tables 1 and 2 present the sample means, which are unbiased minimum variance estimates of the population means, for each frame size by month.

The hypothesis we wish to test is H_0: $\mu^{(1)} = \mu^{(2)}$, $\mu^{(1)} = \mu^{(3)}$ against H_A: $\mu^{(1)} \neq \mu^{(2)}$ and/or $\mu^{(1)} \neq \mu^{(3)}$. The hypothesis H_0 specifies that the vector of population means for the 26.1 inch frame is equal to the vector of means for the 21.6 inch frame and the vector of population means for the 26.1 inch frame is equal to the vector of means for the 14.4 inch frame.

The above hypothesis implies that we want to test two hypotheses, say H_{01}: $\mu^{(1)} = \mu^{(2)}$ and H_{02}: $\mu^{(1)} = \mu^{(3)}$, but reach separate conclusions about each. Hence, a simultaneous multivariate procedure must be employed to test H_0. Kramer [6] has a chapter on simultaneous multivariate procedures. The calculations needed to compute this will be presented.

Table 1.--Expanded means per acre by unit size, Winter Wheat, Kansas, 1972

Month	Characteristic	Unit size		
		14.4 inches	21.6 inches	26.1 inches
May	No. stalks	3,431,000	3,102,000	3,385,420
	No. stalks > 10"	702,219	629,678	586,230
	No. emerged heads	107,847	90,143	79,693
	No. heads in late boot	291,319	320,965	313,750
June	No. stalks > 10"	2,001,981	1,902,216	1,991,724
	No. emerged heads	1,949,453	1,843,811	1,915,434
	No. heads in late boot	43,228	50,466	69,297
July	No. emerged heads	1,946,200	1,850,060	1,933,410
	No. heads in late boot	0	140	0

Table 2.--Expanded means per acre by unit size, Winter Wheat, Washington, 1972

Month	Characteristic	Unit size		
		14.4 inches	21.6 inches	26.1 inches
May	No. of stalks	3,152,888	3,279,292	3,192,920
June	No. stalks > 10"	1,185,528	1,248,890	1,299,202
	No. emerged heads	241,198	256,685	267,491
	No. heads in late boot	122,175	141,295	143,386
July	No. emerged heads	1,652,324	1,713,390	1,722,450
	No. heads in late boot	57,722	29,766	48,729
August	No. emerged heads	1,685,638	1,703,211	1,746,340
	No. heads in late boot	6,619	5,215	5,055

Let $Y_1^{(1)}, Y_2^{(1)}, \ldots . Y_N^{(1)}$ be the sample of random p x 1 vectors observed on the 26.1 inch unit. And let $Y_1^{(2)}, Y_2^{(2)}, \ldots . Y_N^{(2)}$ be the sample of random vectors observed on the 21.6 inch unit. For example, the vector of observed variables on the 26.1 inch unit for the May survey period in Kansas is the collection of vectors of the form

$$Y_i^{(1)} = \begin{bmatrix} y_{1i}^{(1)} \\ y_{2i}^{(1)} \\ y_{3i}^{(1)} \\ y_{4i}^{(1)} \end{bmatrix} = \begin{bmatrix} \text{No. stalks, } i^{th} \text{ unit} \\ \text{No. stalks > 10'', } i^{th} \text{ unit} \\ \text{No. emerged heads, } i^{th} \text{ unit} \\ \text{No. heads in late boot, } i^{th} \text{ unit} \end{bmatrix} i=1,2,\ldots N.$$

Now let,

$$D_i = \begin{bmatrix} d_{1i} \\ d_{2i} \\ \vdots \\ d_{pi} \end{bmatrix} = \begin{bmatrix} y_{1i}^{(1)} - y_{1i}^{(2)} \\ y_{2i}^{(1)} - y_{2i}^{(2)} \\ \vdots \\ y_{pi}^{(1)} - y_{pi}^{(2)} \end{bmatrix} = \begin{bmatrix} y_{1i}^{(1)} \\ y_{2i}^{(1)} \\ \vdots \\ y_{pi}^{(1)} \end{bmatrix} - \begin{bmatrix} y_{1i}^{(2)} \\ y_{2i}^{(2)} \\ \vdots \\ y_{pi}^{(1)} \end{bmatrix} = Y_i^{(1)} - Y_i^{(2)}$$

So D_i is a vector whose rows are the difference of the 21.6 inch unit from the 26.1 inch unit for each particular variable observed.

Assume that D_i has a distribution which is normal with mean δ and covariance matrix Σ. Then $\delta = E(D_i) = E(Y_i^{(1)} - Y_i^{(2)} = \mu^{(1)} - \mu^{(2)}$. Hence, testing H_{01} is equivalent to testing H_{01}^*: $\delta = 0$. The test statistic used to test H_{01}^* is $T^2 = N \bar{D}' S^{-1} \bar{D}$, where

$$\bar{D} = \sum_{i=1}^{N} D_i/N = \sum_{i=1}^{N} \begin{bmatrix} d_{1i} \\ d_{2i} \\ \vdots \\ d_{pi} \end{bmatrix} /N = \begin{bmatrix} \sum_{i=1}^{N} d_{1i}/N \\ \sum_{i=1}^{N} d_{2i}/N \\ \vdots \\ \sum_{i=1}^{N} d_{pi}/N \end{bmatrix} = \begin{bmatrix} \bar{d}_1 \\ \bar{d}_2 \\ \vdots \\ \bar{d}_p \end{bmatrix} =$$

$$\begin{bmatrix} \bar{y}_1^{(1)} - \bar{y}_1^{(2)} \\ \bar{y}_2^{(1)} - \bar{y}_2^{(2)} \\ \vdots \\ \bar{y}_p^{(1)} - \bar{y}_p^{(2)} \end{bmatrix}$$

and,

$$S = \sum_{i=1}^{N} (D_i-\bar{D})\ (D_i-\bar{D})'\ /N-1 = (\sum_{i=1}^{N} D_i D_i' - N\bar{D}\bar{D}')/N-1 =$$

$$\begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & & s_{pp} \end{bmatrix}$$

and,

$$s_{kk} = \frac{SS_{kk}}{N-1} = \frac{\sum_{i=1}^{N} d_{ki}^2 - N\bar{d}_k}{N-1}, \quad k = 1,2,\ldots,p.$$

$$s_{kj} = \frac{SS_{kj}}{N-1} = \frac{\sum_{i=1}^{N} d_{ki}\ d_{ji} - N\bar{d}_k\bar{d}_j}{N-1}, \quad k \neq j = 1,2,\ldots,p.$$

The statistic T^2 is distributed as Hotelling's T^2 with N-1 degrees of freedom (proved in Anderson [1]). A similar development can be made to test H_{02}. As pointed out earlier since we want to test H_{01} and H_{02} jointly (i.e. Test H_0) but reach separate conclusions on each hypothesis we must employ a simultaneous procedure. Thus, individual T^2 tests done at the .025 level on H_{01} and H_{02} will yield a test of H_0 at a significant level bounded by .05 (i.e., $0<\alpha\leq .05$).

Tables 3 and 4 present the characteristics counted for each unit, the T^2 tests for H_{01} and H_{02}, and the critical value of Hotelling's T^2 distribution with proper number of degrees of freedom.

Table 3.--Variables observed and T^2's for test of H_{01} and H_{02} by month, Winter Wheat, Kansas 1972

Month	Vector of p observed variables	Sample size=N	T^2 for test of H_{01}	T^2 for test of H_{02}	Critical value of $T^2_{p,N-1,.025}$ 1/
May	No. stalks No. stalks > 10" No. emerged heads No. heads in late boot	42	7.941	4.079	$T^2_{4,41,.025}=13.582$
June	No. stalks > 10" No. emerged heads No. heads in late boot	80	5.012	2/	$T^2_{3,79,.025}=10.131$
June	No. stalks > 10" No. emerged heads	80		3.601	$T^2_{2,79,.025}=7.838$
July	No. emerged heads No. heads in late boot	80	4.305	3/	$T^2_{2,79,.025}=7.838$
July	No. emerged heads	80		.059	$T^2_{1,79,.025}=5.221$

1/ These values were approximated using linear interpolation.

2/ For the hypothesis H_{02} the June data yielded a negative T^2 using the variables no. stalks>10", no. emerged heads, no. heads in late boot. Thus, only no. stalks>10" and emerged heads were used.

3/ No. of heads in late boot was zero for both the 26.1 and 14.4 inch frames in July. Hence, the covariance matrix was not invertible. No. emerged heads was used.

Table 4.--Variables observed and T^2 for test of H_{01} and H_{02} by month, Winter Wheat, Washington 1972

Month	Vector of p observed variables	Sample size=N	T^2 for test of H_{01}	T^2 for test of H_{02}	Critical value of $T^2_{p,N-1,.025}$ 1/
May	[No. stalks]	74	.726	.092	$T^2_{1,73,.025}$= 5.238
June	[No. stalks > 10" No. emerged heads No. heads in late boot]	72	1.185	3.519	$T^2_{3,71,.025}$=10.22
July	[No. emerged heads No. heads in late boot]	70	2.675	4.025	$T^2_{2,69,.025}$= 7.908
August	[No. emerged heads]	72	.943	1.977	$T^2_{4,71,.025}$= 7.892

1/ These values were approximated using linear interpolation.

None of the computed T^2 values exceeds the critical value of the T^2 distribution with the proper degrees of freedom for any month. Thus, we conclude that the expanded monthly counts per acre made on either of the two smaller frames does not differ significantly from those made on the 26.1 inch unit.

The Analysis of Harvest Counts and Measurements

All heads in each unit within a field were clipped when the unit was in the hard dough or ripe stage of maturity. Heads were sent to the regional laboratories where they are counted and weighed for each field. The two samples from a field are then combined, threshed, weighed (after threshing), and then tested for moisture content. Fields may be combined to test for moisture content, if sufficient grain is not available. Average gross yield is weight of mature heads adjusted for threshing percent and moisture content. Table 5 presents the average gross yield in bushels per acre and the standard error of the mean, for each frame size and State.

Determinations of moisture content often required that grain from more than one field to be combined, thus confounding moisture content determination of individual fields. Since moisture and adjustments are applied to weight of mature heads, at the field level, gross yield is also confounded.

It was felt that the smaller sized unit would have a different threshing percent since an increased number of samples was needed to determine the threshing percent. To investigate this, the mean threshing percent was computed for each frame size. Table 6 represents the mean threshing percent for each frame size by State.

Table 5.--Average expanded gross yield, in bushels per acre by unit size and State, Winter Wheat, 1972

Unit size	Kansas		Washington	
	Mean gross yield	S. E. of mean	Mean gross yield	S. E. of mean
14.4 inch.....	34.766	5.497	57.029	9.375
21.6 inch.....	33.141	5.240	56.352	9.264
26.1 inch.....	34.904	5.519	60.703	9.979

Table 6.--Average threshing percent by unit size and State, Winter Wheat, 1972

Unit size	Kansas		Washington	
	Mean threshing %	S. E. of mean	Mean threshing %	S. E. of mean
14.4 inch.....	67.900	10.736	73.728	12.121
21.6 inch.....	67.277	10.637	73.737	12.122
26.1 inch.....	66.296	10.482	73.454	12.076

To determine if there was any significant differences between the mean gross yield and threshing percent for the two new frames and the standard 26.1 inch frame, a paired T^2, test was used again to test H_0: $\mu^{(1)} = \mu^{(2)}$,

$$\mu^{(1)} = \mu^{(3)} \text{ where } \mu^{(i)} = \begin{bmatrix} \text{mean gross yield for the } i^{th} \text{ frame} \\ \text{mean threshing percent for the } i^{th} \text{ frame} \end{bmatrix} \quad i = 1,2,3.$$

Table 7 presents the T^2 test for H_{01}: $\mu^{(1)} = \mu^{(2)}$ and H_{02}: $\mu^{(1)} = \mu^{(3)}$ by State.

Table 7.--Variables and T^2 for test of H_{01} and H_{02} by State, Winter Wheat, 1972

State	Vector of p variables	Sample size=N	T^2 for test of H_{01}	T^2 for test of H_{02}	Critical value of 1/ $T^2_{p,N-1,.025}$
Kansas	[Gross yield, threshing %]	40	6.023	1.848	$T^2_{2,39,.025} = 8.512$
Washington	[Gross yield, threshing %]	37	6.709	2.991	$T^2_{2,36,.025} = 8.452$

1/ These values were approximated using linear interpolation.

Since the T^2's are not significant for Kansas or Washington, we accept the hypothesis H_0 in each instance. Thus, for each State, there are no significant differences between the smaller size units and the 26.1 inch unit for gross yield or threshing percent.

Time Per Unit Analysis

The total time needed to count the observed variables was recorded for each unit. The average time per unit for each frame size was computed. Tables 8 and 9 present these averages for each survey period and State.

The data in Tables 8 and 9 indicate time savings of three to five minutes per unit for the 21.6 inch unit and six to eight minutes for the 14.4 inch unit when compared with the 26.1 inch frame. Another method of comparison is that count times for the 21.6 inch unit average about 20 percent less than for the 26.1 inch unit. The 14.4 inch unit's mean count time is approximately 40 percent less than the 26.1 inch unit.

The within unit time savings for Kansas for the entire season would amount to about 110 hours for the 21.6 inch unit and 205 hours for the 14.4 inch unit. This would be only a small portion of the savings, however, since the time saved by use of a smaller clip area might be nearly as great as the time saved within the unit counting. Spending less time per field should enable enumerator to work an extra field on some days which would mean that some expensive trips to pick up only one field on another day might be avoided. The smaller sample size for laboratory work should shorten the time and cost of this portion of the survey.

Table 8.--Average time in minutes and standard error per unit by length of unit and survey, Winter Wheat, Kansas, 1972

Survey period	26.1 inch unit		21.6 inch unit		14.4 inch unit	
	Mean	S. E. of mean	Mean	S. E. of mean	Mean	S. E. of mean
May 1	24.69	2.320	20.93	1.781	16.90	1.442
June 1	12.41	0.760	9.52	0.564	7.71	0.418
July 1	20.72	1.041	16.30	0.780	12.68	0.689

Table 9.--Average time in minutes and standard error per unit by length of unit and survey, Winter Wheat, Washington, 1972

Survey period	26.1 inch unit		21.6 inch unit		14.4 inch unit	
	Mean	S. E. of mean	Mean	S. E. of mean	Mean	S. E. of mean
May 1.......	17.57	1.149	14.40	1.000	10.10	0.648
June 1......	16.34	1.187	11.51	0.783	8.43	0.556
July 1......	9.69	0.725	7.09	0.561	5.66	0.578
August 1....	14.36	0.646	12.03	0.609	8.07	0.381

Appendix I: Optimum Allocation of Rows and Units

The current objective yield winter wheat sampling design allocations using the 26.1 inch frame are: two 26.1 inch units per field, and three rows per unit. Since there would be no time savings with the adoption of a smaller size unit if, for example, to obtain accurate estimates at a fixed cost three units per field and four rows per unit were needed with the smaller frame size(s). The investigation of the optimum allocation for units per field and rows per unit for each frame size was computed for a fixed cost. Cochran [2] discusses Optimum Allocation for fixed costs.

The variables on which the analysis was performed was the number of stalks for the May 1 survey in Kansas and Washington and the number of emerged heads for the July 1 survey in Kansas and the August 1 survey in Washington.

For three-stage sampling the optimum solution is to minimize the product of variance and cost with respect to the sample size n_j at each stage, with the cost fixed. The estimated variance

$$\hat{\sigma}^2_{\ell} = \frac{\hat{\sigma}^2_{1_\ell}}{n_1} + \frac{\hat{\sigma}^2_{2_\ell}}{n_1 n_2} + \frac{\hat{\sigma}^2_{3_\ell}}{n_1 n_2 n_3}$$

where

$\hat{\sigma}^2_{1_\ell}$ = the estimated between fields variance component

$\hat{\sigma}^2_{2_\ell}$ = the estimated between units variance component

$\hat{\sigma}^2_{3_\ell}$ = the estimated between rows variance component, and

ℓ = 14.4, 21.6, 26.1, (length of plot in inches).

The cost $C_\ell = n_1 \; C_1 + n_1(n_2-1)C_2 + n_1 \; n_2 \; n_3 \; C_{3\ell}$

where C_1 = average cost between fields

C_2 = average cost between units within fields,

C_{3_ℓ} = average cost per row, for the ℓ - length frame.

The cost is in terms of time not dollars.

But C_1, the average cost per field, is difficult to estimate, since no accurate records were kept. Hence, estimating C_1 will be subjective. However reasonable assumptions for C_1 are 60 and 45 minutes. Both of these assumptions will be used in the optimum allocation analysis.

The cost between fields, in terms of the time, is the time spent traveling between the last unit of a particular field to the first unit of the next field visited. The average cost between units (C_2) was computed as the time spent traveling from unit 1 to unit 2. C_{3_ℓ} is the average within unit count time divided by three, for the ℓ - length frame. Tables 12 and 13 present the costs at each level of sampling by frame size for each State.

Taking the partial derivatives of the variance x cost function with respect to n_j, setting the partials equal to zero and solving for n_j yields the optimum allocations. The optimum allocation for units within fields is,

$$\text{opt}(n_2) = \sqrt{\frac{(C_1-C_2)}{C_2} \frac{\hat{\sigma}^2_{2_\ell}}{\hat{\sigma}^2_{1_\ell}}}$$

The optimum allocation for rows within units is

$$\text{opt}(n_3) = \sqrt{\frac{C_2}{C_3} \frac{\hat{\sigma}^2_{3_\ell}}{\hat{\sigma}^2_{2_\ell}}}$$

Tables 14 thru 17 presents the optimum allocations of units and rows by State, survey period, and C_1 equal to 60 and 45 minutes. The variable components are presented in Appendix II.

The optimum allocation of units and rows are real numbers not necessarily integers which have minimum estimated sampling variance for a fixed cost. However, practical considerations dictate that the allocations be integer valued. The problem now is how to select these integer valued allocations given the optimum allocations.

Consider the lattice in Figure 2 for May 1 number of stalks in Kansas with the 26.1 inch frame. The optimum number of units per field and rows per unit is 1.46 and 1.19 respectively. Write these as an ordered pair (1.46,1.19).

Several criteria can be used to select integer valued allocations, among them are:

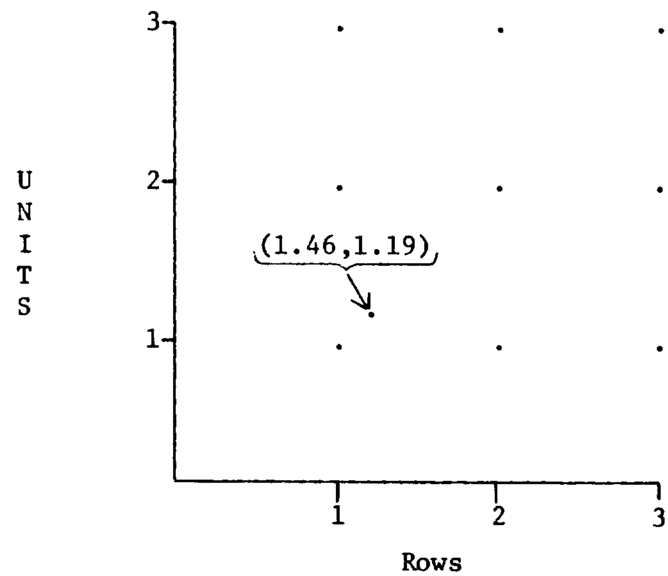
(1) round each component up (down) to integer values,

(2) compute the Euclidean distance between any integer valued ordered pair and the optimum and select the ordered pair with minimum distance,

(3) select the integer valued allocation with minimum estimated sampling variance.

For the ordered pair (1.46, 1.19) the three criteria above would yield (2,2), (2,2), (2,1) respectively.

Figure 2.--Graph of optimum allocation for May 1 number of stalks, 26.1 inch frame, Kansas, Winter Wheat, 1972



Tables 18 to 21 present for each frame size the integer valued allocations which have minimum variance over all sample designs with integer valued allocations by State, survey period, and time between fields. The number of fields, the highest level of sampling, is not entered in the table since it depends on the survey budget allocations to the individual State or more generally a fixed budget or a fixed level of precision.

It should be noted here that the optimum allocation was done on two of the independent variables (no. stalk, emerged head) in the winter wheat forecasting model and not the dependent variable, number of heads at harvest. However, Cochran [2, P. 118] in discussing the problem(s) of allocation with more than one variable in stratified sampling indicates that allocations may differ relatively little when correlations among the variables are high. If this relationship holds for the two-stage cluster sampling then we can expect the optima for number of emerged heads to differ relatively little for number of heads at harvest. However, one must consider all variables and especially the dependent variables (weight per head and number of heads) in the forecasting model(s) before any definite conclusions can be reached on the optimum allocations. These conclusions cannot be reached by this study since the necessary data collection was not included in the design.

Table 12.--Costs in minutes at each level of sampling by frame size and month, Washington, Winter Wheat, 1972

Month	Level of cost	Frame size		
		14.4 inches	21.6 inches	26.1 inches
May	Units (C_2)	8.14	8.14	8.14
	Rows (C_3)	3.37	4.80	5.86
August	Units (C_2)	6.61	6.61	6.61
	Rows (C_3)	2.69	4.01	4.79

Table 13.--Costs in minutes at each level of sampling by frame size and month, Kansas, Winter Wheat, 1972

Month	Level of cost	Frame size: 14.4 inches	21.6 inches	26.1 inches
May	Units (C_2)	7.00	7.00	7.00
	Rows (C_3)	5.63	6.98	8.23
July	Units (C_2)	6.06	6.06	6.06
	Rows (C_3)	4.23	5.43	6.91

Table 14.--Optimum allocations of units and rows C_1 = 60 minutes, Winter Wheat, Washington, 1972

Survey period and variable	Allocations					
	14.4 inch frame		21.6 inch frame		26.1 inch frame	
	Units	Rows	Units	Rows	Units	Rows
May 1, number of stalks.......	1.863	1.652	1.029	1.919	1.391	1.220
August 1, number of emerged heads..............	1.318	4.045	0.879	3.574	1.411	1.624

Table 15.--Optimum allocation of units and rows C_1 = 45 minutes, Winter Wheat, Washington, 1972

Survey period and variable	Allocations					
	14.4 inch frame		21.6 inch frame		26.1 inch frame	
	Units	Rows	Units	Rows	Units	Rows
May 1, number of stalks.......	1.571	1.652	0.868	1.919	1.173	1.220
August 1, number of emerged heads..............	1.117	4.045	0.746	3.574	1.196	1.624

Table 16.--Optimum allocation of units and rows C_1 = 60 minutes fields, Winter Wheat, Kansas, 1972

Survey period and variable	Allocations					
	14.4 inch frame		21.6 inch frame		26.1 inch frame	
	Units	Rows	Units	Rows	Units	Rows
May 1, number of stalks.......	1.519	1.376	1/	1/	1.723	1.192
July 1, number of emerged heads...............	2.113	1.747	0.691	5.617	1.718	1.675

1/ No optimum allocation was computed since a negative variance component was estimated.

Table 17.--Optimum allocation of units and rows C_1 = 45 minutes fields, Winter Wheat, Kansas, 1972

Survey period and variable	Allocations					
	14.4 inch frame		21.6 inch frame		26.1 inch frame	
	Units	Rows	Units	Rows	Units	Rows
May 1, number of stalks.......	1.286	1.376	1/	1/	1.459	1.192
July 1, number of emerged heads...............	1.795	1.747	0.587	5.617	1.459	1.675

1/ No optimum allocation was computed since a negative variance component was estimated.

Table 18.--Integer valued allocations which have minimum estimated variance by survey period and frame size, C_1 = 60 minutes, Winter Wheat, Washington, 1972

Survey period and variable	Allocations					
	14.4 inch frame		21.6 inch frame		26.1 inch frame	
	Units	Rows	Units	Rows	Units	Rows
May 1, number of stalks.......	2	2	1	2	2	1
August 1, number of emerged heads...............	1	5	1	3	1	2

Table 19.--Integer valued allocations which have minimum estimated variance by survey period and frame size, C_1 = 45 minutes, Winter Wheat, Washington, 1972

Survey period and variable	Allocations					
	14.4 inch frame		21.6 inch frame		26.1 inch frame	
	Units	Rows	Units	Rows	Units	Rows
May 1, number of stalks.......	2	2	1	2	1	1
August 1, number of emerged heads..............	1	5	1	3	1	2

Table 20.--Integer valued allocations which have minimum estimated variance by survey period and frame size, C_1 = 60 minutes, Winter Wheat, Kansas, 1972

Survey period and variable	Allocations					
	14.4 inch frame		21.6 inch frame		26.1 inch frame	
	Units	Rows	Units	Rows	Units	Rows
May 1, number of stalks.......	2	1	2	1	2	1
July 1, number of emerged heads..............	2	2	1	5	2	2

Table 21.--Integer valued allocations which have minimum estimated variance by survey period and frame size, C_1 = 45 minutes, Winter Wheat, Kansas, 1972

Survey period and variable	Allocations					
	14.4 inch frame		21.6 inch frame		26.1 inch frame	
	Units	Rows	Units	Rows	Units	Rows
May 1, number of stalks.......	1	2	2	1	2	1
July 1, number of emerged heads..............	2	2	1	5	1	2

Appendix II: The Optimum of the Three Frame Sizes

In Appendix I, we investigated the allocation of the levels of sampling for each frame size by considering the costs (C_i) fixed and letting n_1,n_2, and n_3 vary in the variance x cost function. In this investigation of the optimum frame size we will fix the allocation at two units per field and three rows per unit and determine the minimum value of the cost x variance function for the alternative frame sizes. The allocation for n_1, the number of fields, is a fixed positive integer. The frame size with minimum cost x variance is the optimum frame size. The analysis will be presented for the same survey periods and variables as for the optimum allocation analysis.

Consider again the general cost function $C_\ell = n_1C_1 + n_1(n_2-1)C_2 + n_1n_2n_3C_{3\ell}$ where C_1,C_2,$C_{3\ell}$ were defined in the previous section. If we let $n_2 = 2$ and $n_3 = 3$, then $C_\ell = n_1(C_1+C_2+6C_{3\ell})$. C_ℓ is then the total cost in terms of time for enumerations of the ℓ^{th} frame size for a particular month and State when the allocations are two units per field and three rows per unit. As before since we have no accurate estimates of C_1, 60 and 45 minutes were used. For these two assumed values of C_1 a separate estimate of the total time C_ℓ was computed. Table 22 presents the total time as a constant times n_1, the number of fields. The reason for this will be apparent later. This can be used to estimate the total time saved on n_1 fields between two frame sizes for a particular month and State. For example, an estimate of total time saved by the 14.4 inch frame over the 26.1 inch frame for 100 samples in Kansas in May with $C_1 = 60$ is 100 (103.30 - 88.36) = 1494 minutes.

Table 22.--Total cost (C_ℓ) in minutes for each unit size by State and month for C_1 equal to 60 and 45 minutes, Winter Wheat, 1972

State	Month	Frame size					
		14.4 inches		21.6 inches		26.1 inches	
		$C_1 = 60$	$C_1 = 45$	$C_1 = 60$	$C_1 = 45$	$C_1 = 60$	$C_1 = 45$
Washington	May	88.36	73.36	96.94	81.94	103.20	88.30
Washington	August	82.75	67.75	90.67	75.67	95.35	80.35
Kansas	May	100.78	85.78	108.88	93.88	116.38	101.38
Kansas	August	91.44	76.44	98.64	83.64	107.52	92.52

Now consider the estimated variance for the ℓ^{th} frame size

$$\hat{\sigma}^2_\ell = \frac{\hat{\sigma}^2_{1_\ell}}{n_1} + \frac{\hat{\sigma}^2_{2_\ell}}{n_1 n_2} + \frac{\hat{\sigma}^2_{3_\ell}}{n_1 n_2 n_3}$$ where

the estimated variance components $\hat{\sigma}^2_{i\ell}$, i = 1,2,3 were defined in the previous section. With $n_2 = 2$ and $n_3 = 3$, $\hat{\sigma}^2_\ell$ becomes $\hat{\sigma}^2_\ell = \frac{1}{n_1} (\hat{\sigma}^2_{1_\ell} + \hat{\sigma}^2_{2_\ell}/2 + \hat{\sigma}_{3_\ell}/6)$.

The estimate of the variance components for each frame size, month, and State are given in the nested analyses of variance presented in Tables 25 thru 36. The estimate of $\hat{\sigma}^2_\ell$ is presented in Table 23 as a constant divided by n_1. Thus to estimate $\hat{\sigma}^2_{14.4}$ based on a sample of size 100 in May from Kansas for the variable number of stalks, the estimate is 1543.73/100 = 15.44.

Table 23.--Estimated variance for each unit size and month, Winter Wheat, 1972

State	Month	Variable	Frame size		
			14.4 inches	21.6 inches	26.1 inches
Washington	May	No. of stalks	$1327.27/n_1$	$3549.57/n_1$	$4814.73/n_1$
Washington	August	No. emerged heads	$215.02/n_1$	$616.21/n_1$	$886.57/n_1$
Kansas	May	No. of stalks	$1543.73/n_1$	$2232.82/n_1$	$2466.28/n_1$
Kansas	August	No. emerged heads	$242.44/n_1$	$370.62/n_1$	$562.64/n_1$

As stated earlier for fixed $n_2 = 2$ and $n_3 = 3$ the optimum frame size among the 26.1 and 14.4 inch frames is the frame size with minimum cost x variance. The cost x variance for the ℓ^{th} frame size is

$$C_\ell \times \hat{\sigma}^2_\ell = n_1 \ (C_1 + C_2 + 6C_3) \times \frac{1}{n_1} \ (\hat{\sigma}^2_{1_\ell} + \hat{\sigma}^2_{2_{\ell/2}} + \hat{\sigma}^2_{2_{\ell/6}})$$

$$= (C_1 + C_2 + 6C_3) \ (\hat{\sigma}^2_{1_\ell} + \hat{\sigma}^2_{2_{\ell/2}} + \hat{\sigma}^2_{3_{\ell/6}})$$

As we can see by our last equality the cost x variance does not depend on the number of fields sampled but on the costs and variance associated with each level of sampling. This is the reason n_1 was left arbitrary in the analysis and in the tables.

Table 24 presents the cost x variance for each frame size by State and month for C_1 equal to 60 or 45 minutes. In every case the minimum cost x variance was for the 14.4 inch frame. Hence, the optimum frame size among the 26.1, 21.6 and the 14.4 inch frame is the 14.4 inch frame.

Table 24.--Cost x variance for each frame size by State and month for C_1 equal to 60 and 45 minutes, Winter Wheat, 1972

State	Month	Variable	14.4 inches		21.6 inches		26.1 inches	
			C_1 = 60	C_1 = 45	C_1 = 60	C_1 = 45	C_1 = 60	C_1 = 45
Washington	May	No. of stalks	117,278	97,369	344,095	290,852	497,362	425,141
Washington	August	No. emerged heads	17,793	14,568	55,872	46,629	84,534	71,236
Kansas	May	No. of stalks	155,577	132,421	243,109	209,617	287,026	250,031
Kansas	August	No. emerged heads	22,169	18,532	36,558	30,999	60,495	52,055

Table 25.--Nested analysis of variance for number of stalks, 14.4 inch frame, May, Washington, 1972

Variance source	D. F.	Sum of squares	Mean squares	Estimated variance components
Total..........	233	479,948	2,059.86	2,085.09
Field..........	38	302,619	7,963.66	965.49
Unit...........	39	84,658	2,170.72	525.55
Rows...........	156	92,671	594.05	594.05

Table 26.--Nested analysis of variance for number of stalks, 21.6 inch frame, May, Washington, 1972

Variance source	D. F.	Sum of squares	Mean squares	Estimated variance components
Total..........	233	1,087,876	4,669.00	4,740.06
Fields.........	38	809,322	21,297.95	3,105.36
Units..........	39	103,965	2,665.77	515.54
Rows...........	156	174,589	1,119.16	1,119.16

Table 27.--Nested analysis of variance for number of stalks, 26.1 inch frame, May, Washington, 1972

Variance source	D. F.	Sum of squares	Mean squares	Estimated variance components
Total..........	221	1,414,964	6,402.55	6,503.84
Fields.........	36	1,039,981	28,888.37	3,992.09
Units..........	37	182,626	4,935.83	1,212.04
Rows...........	148	192,357	1,299.71	1,299.71

Table 28.--Nested analysis of variance for number of emerged heads, 14.4 inch frame, August, Washington, 1972

Variance source	D. F.	Sum of squares	Mean squares	Estimated variance components
Total..........	221	89,864	406.63	410.61
Fields.........	36	46,444	1,290.13	161.98
Units..........	37	11,775	318.25	34.81
Rows...........	148	31,644	213.82	213.82

Table 29.--Nested analysis of variance for number of emerged heads, 21.6 inch frame, August, Washington, 1972

Variance source	D. F.	Sum of squares	Mean squares	Estimated variance components
Total	221	210,818	953.93	966.28
Fields	36	133,052	3,695.90	525.79
Units	37	20,023	541.18	50.34
Rows	148	57,742	390.15	390.15

Table 30.--Nested analysis of variance for number of emerged heads, 26.1 inch frame, August, Washington, 1972

Variance source	D. F.	Sum of squares	Mean squares	Estimated variance components
Total	215	268,351	1,248.15	1,267.00
Fields	35	186,179	5,319.41	737.79
Units	36	32,137	892.70	181.75
Rows	144	50,034	347.46	347.46

Table 31.--Nested analysis of variance for number of stalks, 14.4 inch frame, May, Kansas, 1972

Variance source	D. F.	Sum of squares	Mean squares	Estimated variance components
Total	125	270,534	2,164.28	2,220.61
Fields	20	185,248	9,262.41	1,255.28
Units	21	36,345	1,730.72	382.70
Rows	84	48,941	582.63	582.63

Table 32.--Nested analysis of variance for number of stalks, 21.6 inch frame, May, Kansas, 1972

Variance source	D. F.	Sum of squares	Mean squares	Estimated variance component
Total..........	125	428,592	3,428.74	3,522.34
Fields.........	20	267,937	13,396.89	1,983.61
Units..........	21	31,400	1,495.26	-14.49
Rows...........	84	129,254	1,538.74	1,538.74

Table 33.--Nested analysis of variance for number of stalks, 26.1 inch frame, May, Kansas, 1972

Variance source	D. F.	Sum of squares	Mean squares	Estimated variance component
Total..........	125	472,608	3,780.86	3,868.30
Fields.........	20	295,953	14,797.66	1,889.43
Units..........	21	72,683	3,461.09	741.11
Rows...........	84	103,972	1,237.76	1,237.76

Table 34.--Nested analysis of variance for number of emerged heads, 14.4 inch frame, July, Kansas, 1972

Variance source	D. F.	Sum of squares	Mean squares	Estimated variance component
Total..........	239	103,219	431.88	436.14
Fields.........	39	56,730	1,454.61	169.63
Units..........	40	17,474	436.85	85.17
Rows...........	160	29,015	181.35	181.34

Table 35.--Nested analysis of variance for number of emerged heads, 21.6 inch frame, July, Kansas, 1972

Variance source	D. F.	Sum of squares	Mean squares	Estimated variance component
Total..........	239	176,523	738.59	744.77
Fields.........	39	86,726	2,223.75	289.57
Units..........	40	19,452	486.31	15.55
Rows...........	160	70,344	439.65	439.65

Table 36.--Nested analysis of variance for number of emerged heads, 26.1 inch frame, July, Kansas, 1972

Variance source	D. F.	Sum of squares	Mean squares	Estimated variance component
Total..........	239	237,242	992.64	1,002.57
Fields.........	39	131,659	337.58	419.05
Units..........	40	34,461	861.54	139.00
Rows...........	160	71,123	444.52	444.52

Appendix III: Bias in Plots Sizes

This section presents an analysis of the possible bias in plot sizes. A linear relationship of stalk counts on plot length with zero intercept should exist in the absence of any bias. If the intercept is significantly different from zero, then one or more of the plot sizes are biased, since a plot length of zero inches should estimate zero stalks. Prior to testing the intercept, one must test the assumption that the regression of stalk counts on plot length is linear.

For each month and for each State, suppose we have n_1, n_2, n_3 samples using the 14.4, 21.6 and 26.1 inch frame, respectively. Let Y_{iu}, i=1, 2, 3, u=1, 2,...n_i denote the u^{th} measurement on the i^{th} frame. (The first is the 14.4 inch, etc.). Y_{iu} is either the number of stalks or the number of emerged heads depending on the month and State. Consider the linear model

(1) $Y_{iu} = \beta_0 + \beta_1 X_{iu} + \varepsilon_{iu}$, where for all u

$$X_{iu} = \begin{cases} 14.4 & \text{if } i = 1 \\ 21.6 & \text{if } i = 2 \\ 26.1 & \text{if } i = 3, \text{ and} \end{cases} \qquad \varepsilon_{iu} \sim (0, \sigma^2)$$

A least squares fit of this model using all $n = n_1 + n_2 + n_3$ observations will produce a linear regression of the number of stalks (or emerged heads) on the plot length.

The assumption that the regression of emerged heads on plot size is linear can be tested. The technique is to partition the residual sums of squares from fitting the total n observations into two components, the sums of squares due to lack of fit and the sums of squares due to pure error. The pure error SS divided by its degrees of freedom estimates σ^2 (the error variance) and the lack of fit SS divided by its degrees of freedom estimates σ^2 if the model (1) is correct. The ratio of the two mean squares has the F distribution if the model is correct. The model suffers from lack of fit if the computed F value is significant at a specified level of significance. The necessary calculation and a theoretical summary of the testing the adequacy of a linear model can be found in Draper and Smith [3, Section 1.5].

The model (1) was fitted using the regression procedure of the Statistical Analysis System (SAS). The pure error sums of squares was computed using the procedure MEANS in SAS. Tables 37 thru 40 presents the analysis of variance tables for testing the lack of fit of the linear model (1) for each month and State.

Table 37.--ANOVA for test of lack of fit of regression of number of stalks on length of frame, May, Washington, 1972

Source	D. F.	Sums of squares	Mean square	F	Critical $F_{.05}$ value
Regression..........	1	385363.01			
Error...............	664	2850839.99	4293.434		
Lack of fit........	1	3305.23	3305.225	0.770	3.854
Pure error........	663	2847534.76	4294.924		
Corrected total.....	665	3236703.00			

Table 38.--ANOVA for test of lack of fit of regression of number of emerged heads on length of frame, August, Washington, 1972

Source	D. F.	Sums of squares	Mean square	F	Critical $F_{.05}$ value
Regression..........	1	125160.18			
Error...............	646	558396.14	864.390		
Lack of fit........	1	21.63	21.633	0.025	3.854
Pure error........	645	558374.50	865.697		
Corrected total.....	647	683556.32			

Table 39.--ANOVA for test of lack of fit of the regression of number of stalks on length of frame, Kansas, May, 1972

Source	D. F.	Sums of squares	Mean square	F	Critical $F_{.05}$ value
Regression..........	1	187825.12			
Error...............	376	1179142.54	3136.017		
Lack of fit.......	1	7407.66	7407.664	2.371	3.894
Pure error........	375	1171734.88	3124.636		
Corrected total.....	377	1366967.66			

Table 40.--ANOVA for test of lack of fit of the regression of number of emerged heads on length of frame, Kansas, July, 1972

Source	D. F.	Sums of squares	Mean square	F	Critical $F_{.05}$ value
Regression..........	1	124460.65			
Error...............	718	518141.28	721.645		
Lack of fit.......	1	1158.06	1158.059	1.606	3.853
Pure error........	717	516983.22	721.036		
Corrected total.....	719	642601.93			

Since none of the computed F values exceeded the critical value of the F distribution, there is no lack of fit in any of the models. Some words of caution are in order here. Our range of experience is only from 14.4 to 26.1 and the results indicate that a linear regression model is governing the data in this range.

Since we have repeat observations on each unit size, we have only three data points, the means at each unit size. The deviations of the variable(s) about the mean is pure error and this was used in the above analysis to test lack of fit. We know from the lack of fit analysis presented above that the regression of stalk counts on plot size is linear. Given this, is the intercept zero? If the intercept is not zero, then the plots are biased. To test this the mean number of stalks (or emerged heads) was regressed on the plot size. Table 41 presents the regressions for number of stalks and number of emerged heads.

Table 41.--Estimated regression parameters and their t values for the regression on plot size by State and month, Winter Wheat, 1972

State	Month	Variable	β_0		β_1	
			Estimate	T-test	Estimate	T-test
Washington	May	No. of stalks	-2.331	-0.237	4.992	10.798*
Washington	August	No. emerged heads	-1.817	-2.255	2.884	76.062*
Kansas	May	No. of stalks	3.055	0.156	4.626	5.035
Kansas	July	No. emerged heads	1.030	0.184	2.728	10.367**

* Indicates significant at $\alpha = .05$, with 1 d.f.
** Indicates significant at $\alpha = .10$, with 1 d.f.

Since none of the β_0 parameters were significantly different from zero, based on t-tests computed from the data, there is no reason to believe that the frames are biased.

References

[1] Anderson, T. W., An Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons, 1958.

[2] Cochran, William G., Sampling Techniques, New York: John Wiley and Sons, 1953.

[3] Draper, Norman and Smith, Harry, Applied Regression Analysis, New York: John Wiley and Sons, 1966.

[4] Dunnett, C. W., "A multiple comparison procedure for comparing several treatments with a control," Journal of the American Statistical Association, Vol. 50, pp. 1096-1121.

[5] Homeyer, P. G., Black, C. A., "Sampling relicated field experiments on oats for yield determinations." Soil Science Proceedings, Vol. 11, pp. 341-344.

[6] Kramer, Clyde Y., Methods of Multivariate Analysis. Ann Arbor, Michigan: Edwards Brothers, Inc., 1972.

[7] Siotani, Minoru, "The extreme value of generalized distances and its applications," Bull. Inter. Stat. Institute, Vol. 38, pp. 391-9.

[8] Sukhatme, P. V., "The problem of plot size in large-scale yield surveys," Journal of the American Statistical Association., Vol. 42, pp. 297-310.

[9] Wilson, Wendell W. and Von Steen, Donald H., Texas vegetables remote sensing study, 1969. Determination of optimum plot size and shape for estimation of carrot yield. Research and Development Branch, Standards and Research Division, Statistical Reporting Service, U. S. Department of Agriculture, Washington, D. C.

[10] Wood, Ronald A., Optimum plot size for winter wheat. Research and Development Branch, Standards and Research Division, Statistical Reporting Service, U. S. Department of Agriculture, Washington, D. C.