Hendricks

UNITED STATES DEPARTMENT OF AGRICULTURE
AGRICULTURAL MARKETING SERVICE

A REPORT ON THE SAMPLING OF CORN FIELDS

Walter A. Hendricks,
Associate Agricultural Statistician

Results of an analysis of data on the sampling of corn fields, obtained by the Agricultural Adjustment Administration[1/], are given in this report. Individual fields were sampled by selecting 4-hill blocks in a systematic manner which may be described briefly as follows:

The sampler proceeded along the side of the field to a point between 24 and 25 rows from one corner, at which point he entered the field to a point between 24 and 25 rows from the side of the field and selected a 4-hill block of which two hills were on row 24 and two were on row 25. Similar samples were taken by continuing across the field and taking a 4-hill block on every succeeding twenty-fourth and twenty-fifth row until the opposite side of the field was reached. The sampler then proceeded down the side of the field for the same distance and began taking samples while crossing the field in the opposite direction. This was continued until the entire field was sampled.

The four hills in each block were considered to be two samples of two hills each and the ears of corn on each 2-hill sample within a block were weighed as a unit, all weights being expressed in pounds.

The data for five of the fields sampled in Illinois were used in the present study. An analysis of variance was obtained to determine the extent of the average variability between blocks in the same field and between 2-hill samples in the same block. The data yielded 491 degrees of freedom for estimating the variance between blocks and 496 degrees of freedom for estimating the variance between 2-hill samples within blocks. The average variance between blocks in the same field was 0.74118 while the variance between 2-hill samples within blocks was only 0.43024. The average weight of the corn on a 2-hill sample was 3.04 pounds, from which it is evident that the coefficient of variability of the weight of the corn on 2-hill samples within blocks was about 21.6 percent.

From such an analysis, considerable information regarding the relative efficiencies of different methods of sampling can be obtained. Strictly speaking, such deductions are predicated on the assumption of a random distribution of blocks within fields. The fact that the blocks in the present sampling

---

1/ These data were obtained in 1938 under the supervision of Mr. C. J. Otten of the Economics and Research Section of the North Central Division in cooperation with State statisticians of the Agricultural Marketing Service, State Agricultural Conservation Committees, and County Committees for Illinois, Indiana, and Iowa; and statisticians of the Statistical Laboratory at Iowa State College.

study were taken in systematic fashion is disturbing from the point of view of statistical theory, but its effect on the numerical results is probably not very great because differences among the various parts of any given field may have been distirbuted at random.

The data can be applied to determine the extent to which the variability of the means of any given number of 2-hill samples is affected by increasing the number of 2-hill samples in a block and reducing the number of blocks or vice versa. In order to solve this problem it is necessary to think of the variability of the block means within a field as consisting of two components. The first component represents the actual block-to-block differences and the second represents an additional effect produced by the sampling errors in the various estimates of the individual block means. Since each block mean is the mean of two 2-hill samples, the variance of each of these means, considered as estimates of the corresponding hypothetical true means is .43024/2. The variance of the observed block means about their grand mean is .74118/2. The difference between these two figures, or 0.15547, may be regarded as the variance of the true block means which represents the actual block-to-block variation after the effects of the sampling errors in the estimates of the individual block means have been removed. If a number of blocks are chosen at random and there are N 2-hill samples in each block, the variance of the observed block means about their grand mean would be 0.15547 + .43024/N. If there are M blocks, the variance of the mean of the M individual block means in random samples of M blocks would be .15547/M + .43024/MN. MN represents the total number of 2-hill samples involved. Therefore, the above expression gives the variance of the mean of MN 2-hill samples distributed in M blocks each containing N samples. If MN is held constant at a value, K, the expression may be written in the form .15547N/K + .43024/K.

By means of this formula it is possible to compute the variance of the mean of any number of 2-hill samples distributed in any given number of blocks. For example, if one is interested in the mean of 100 2-hill samples, the variance is 0.0015547N + 0.0043024 in which N is the number of 2-hill samples in a block. The following figures show how the variance of the mean of 100 2-hill samples changes when these samples are grouped in blocks of different sizes:

No. of 2-hill samples in a block	No. of blocks	Variance of mean of 100 2-hill samples
1	100	0.00586
2	50	.00741
4	25	.01052
5	20	.01208
10	10	.01984

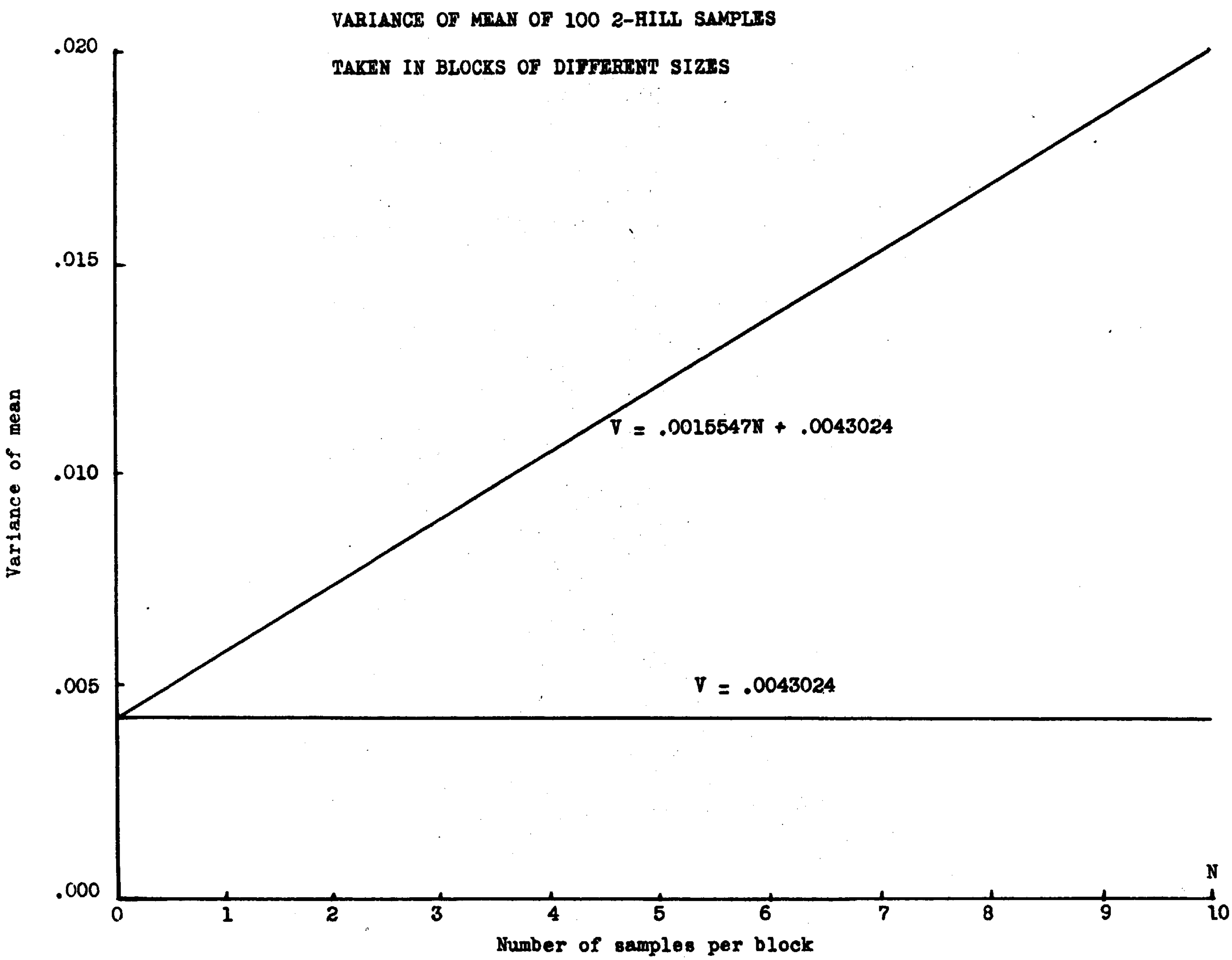
The relation is shown graphically in the attached chart. The horizontal line in this chart shows the value that the variance of the mean of 100 2-hill samples would assume for all values of N if there were no differences between true means of the blocks.

These results were derived by considering the 2-hill sample as the sampling unit. Similar results can be obtained even though the total number of hills in a block is not an exact multiple of 2. A block containing three hills may be treated mathematically as a block containing one and one-half 2-hill samples when the fundamental data are expressed in terms of a 2-hill sampling unit. For example, consider the following problem:

It is desired to compute the extent to which the variance in the mean of 40 blocks would be increased by reducing the number of hills per block from 4 to 3. The variance of the mean of 40 4-hill blocks is .15547/40 + .43024/80 or 0.009265, since the number of 2-hill samples per block is equal to 2. If there were only 3 hills in a block, the number of 2-hill samples per block would be equal to 1-1/2 and the variance of the mean of 40 such blocks would be .15547/40 + .43024/60 or 0.011058.

The elimination of one hill from each block increases the variance of the mean of the 40 blocks 19.35 percent. Eliminating one hill from each block is equivalent to reducing the total number of hills from 160 to 120. It is interesting to see how the variance of the mean would be affected if the total number of hills were reduced from 160 to 120 by eliminating 10 whole blocks and retaining 4 hills in each of the remaining 30 blocks. In this case there would be two 2-hill samples in each block and the variance of the mean of the 30 blocks would be .15547/30 + .43024/60 or 0.012353. Reducing the total number of hills from 160 to 120 by this method increases the variance of the mean 33.33 percent. It is evident that, if one wishes to reduce the size of the total sample taken from a field, the concomitant loss of precision will be greater when whole blocks are eliminated than when an equivalent number of hills are eliminated from each block.

The results presented above show that samples of hills of corn, taken for the purpose of estimating yield per acre, should be grouped in blocks as small as practical considerations will permit. From the point of view of statistical efficiency, maximum efficiency is attained with the smallest number of hills per block, but the additional labor involved in increasing the total number of blocks must be taken into account to decide whether or not the increased precision of results is worth the price of the additional time and labor required to attain it.



VARIANCE OF MEAN OF 100 2-HILL SAMPLES
TAKEN IN BLOCKS OF DIFFERENT SIZES
.020
.015
.010
.005
.000
Variance of mean
V = .0015547N + .0043024
V = .0043024
N
0
1
2
3
4
5
6
7
8
9
10
Number of samples per block