United States
Department of
Agriculture

Statistical
Research
Service

Estimates
and Statistical
Research
Divisions

SRS Staff Report
Number YRB-85-02

April 1985

# Simplifying Soybean Objective Yield Estimation

Robert Battaglia and Jack Nealon

SIMPLIFYING SOYBEAN OBJECTIVE YIELD ESTIMATION By Robert Battaglia and Jack Nealon, Statistical Research Division, Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C. 20250. April 1985. Staff Report No. YRB-85-02.

ABSTRACT

This report justifies the replacement of the operational soybean objective yield estimator with an alternative estimator that is easier to understand and use, less susceptible to plant-handling effects, and slightly more precise. Analysis indicates that in most cases yield estimates from the alternative estimator are not significantly different from the one now used.



CONTENTS

SUMMARY

This report recommends the replacement of the operational yield estimator in the soybean objective yield program with one that is computationally simpler, easier to understand, and permits simple verification of unusual yields by field office personnel. The alternative estimator is less susceptible to bias that might be introduced by repeated handling of plants in the samples and is slightly more precise in most instances. An analysis of soybean objective yield data from 1981 to 1984 for the 15 States in the program showed few significant differences between the operational and alternative estimators.

# SIMPLIFYING SOYBEAN OBJECTIVE YIELD ESTIMATION

Robert Battaglia and Jack Nealon 1/

## INTRODUCTION

Yield estimation from objective yield surveys is based on relatively simple and straightforward observations of the mature crop. Two units, which make up an objective yield sample, are laid out randomly in each selected field and harvested by the enumerator when the crop is mature. Yield per acre is then computed based on counts, measurements, and weights from the two units.

The statistical procedures used to forecast or estimate yield at maturity in objective yield surveys are sometimes unnecessarily complex. This complexity can affect the survey results by exposing the yield estimates to additional nonsampling errors. The operational yield estimator for the soybean objective yield program uses a complicated statistical procedure to estimate yield at maturity. This complexity has created three problems:

(1) Those involved in collecting the data, editing, and setting the yield estimate do not understand the operational estimator.

(2) The yield estimate for each sample is very difficult to compute manually. Data collected both in the field and in the lab must be weighed together based on the variability between the pod counts from the two units in each sample. Because of the computational complexity of this estimator, unusual yields cannot be easily verified by field office personnel.

(3) The operational estimator uses field counts from the two 6-inch sections in each unit. Research has shown this procedure to be susceptible to bias in some States due to repeated plant handling [2,3]. 2/

Methods Staff and Yield Research Branch desire to simplify objective yield procedures whenever possible, in order to

---

1/ The authors are mathematical statisticians with the Statistical Reporting Service, U.S. Department of Agriculture.
2/ Numbers in brackets refer to literature cited in the References at the end of the report.

minimize the effects of nonsampling errors and to make the procedures easier to understand. The purpose of this report is to justify the use of an alternative yield estimator in the soybean objective yield program. This estimator is much easier to understand, makes it easier to verify unusual yields, and is less susceptible to possible bias from handling the plants in the samples.

This report discusses the operational and alternative yield estimators and compares the yield estimates from these estimators using soybean objective yield data from 1981 to 1984. If the yield estimates are statistically the same for most comparisons, we will recommend the alternative estimator be used in future surveys. If the yield estimates are statistically different in many cases, we will recommend that research be conducted to uncover why the yield estimates differ between the two estimators before a decision is made whether to use the alternative estimator operationally.

## ESTIMATORS

This section describes the operational and alternative yield estimators. The only difference between the two estimators is the source of measurements on the number of pods with beans. Both estimators require row width measurements to compute the number of pods with beans per 18 square feet component of yield [4]. The current survey design and computer summary programs require the "reported" value from each sampled field to be "yield per acre" in bushels. Subsequently this value will be referred to as "yield". Yield at maturity for a soybean objective yield sample can be calculated using the following formula:

$$\begin{pmatrix}\text{Pods with beans}\\ \text{per 18 sq. ft.}\end{pmatrix} * \begin{pmatrix}\text{Weight of beans per}\\ \text{pod with beans}\end{pmatrix} * \begin{pmatrix}\text{Conversion}\\ \text{factor}\end{pmatrix}$$

where the number of pods with beans per 18 square feet is estimated from field (Form B) and/or lab (Form C) data, weight of beans per pod with beans is calculated using lab data, and the conversion factor converts the weight per 18 square feet to bushels per acre. The operational and alternative yield estimators differ only in how they estimate the number of pods with beans per 18 square feet. The operational estimator uses pod counts from both Forms B and C while the alternative estimator relies solely on Form C pod counts. A description of each yield estimator follows.

### Operational Estimator

The operational yield estimator requires that pods with beans per 18 square feet be estimated from Form B and Form C data. The Form B estimate of pods with beans per 18 square feet for a sample is based on the enumerator's plant counts in the 42-inch rows, the row width, and the number of pods with beans counted in the 6-inch sections [5]. In formula notation, the estimate of pods with beans per 18 square feet based on Form B data is:

$$\frac{(A1 * B1) + (A2 * B2)}{2}$$

where Ai is the number of plants per 18 square feet in unit i based on the counts in the two 42-inch rows in unit i, and Bi is the number of pods with beans per plant in unit i based on the counts in the two 6-inch sections in each row in unit i.

The Form C estimate of number of pods with beans is currently based on lab data from the 3-foot section of the first row of each unit [5]. In formula notation, the estimate of pods with beans per 18 square feet based on Form C data is:

$$\frac{P1 + (P1 * W2 / W1)}{2}$$

where P1 is the estimate of number of pods with beans per 18 square feet for unit 1 based on the lab count in the 3-foot section of row 1 in unit 1, and (P1 * W2 / W1) is a type of ratio estimate of pods with beans per 18 square feet for unit 2 derived by adjusting the pod estimate for unit 1 (P1) by the ratio of the weight of pods and beans in unit 2 (W2) to the weight of pods and beans in unit 1 (W1). This estimate assumes that the ratio of the number of pods with beans to total weight of beans and pods is the same for the two units and is used in lieu of actually counting the pods with beans for unit 2.

Next, the operational estimator weights together the Form B and C estimates of pods with beans for each sample based on their relative variances [4]. Relative variance is synonymous with the coefficient of variation squared. The relative variances are based on the variation between the pods with beans estimates from each of the two units in a sample. Since the pod estimate for unit 2 from Form C is derived, it is not really appropriate to derive a relative variance based on the Form C counts. In formula notation, the operational estimator for pods with beans per 18 square feet for a sample is:

$$P * \begin{pmatrix}\text{Pods with beans} \\ \text{per 18 sq. ft.} \\ \text{(Form B)}\end{pmatrix} + (1-P) * \begin{pmatrix}\text{Pods with beans} \\ \text{per 18 sq. ft.} \\ \text{(Form C)}\end{pmatrix}$$

where P is the relative variance of the pods-with-beans estimate from Form C divided by the sum of the relative variances of the pods-with-beans estimates from Forms B and C for the sample. That is, P is a weight between 0 and 1 that is assigned to the Form B estimate of pods with beans for the sample. The weight given to the pods-with-beans estimate from Form C is (1-P). The weights assigned to the Form B or C estimates differ from sample to sample depending upon the relative variances between the units in each sample. Overall, however, the average weights assigned to the Form B and C estimates are about .45 and .55, respectively. Therefore, the operational estimator relies very heavily on both the Form B and C estimates.

Finally, the operational yield estimate for each sample is derived by multiplying the weighted estimate of pods with beans per 18 square feet by the weight of beans per pod and then expanding this product to bushels per acre.

Alternative Estimator:

The alternative yield estimator, hereafter referred to as the Form C yield estimator, relies solely on Form C pod counts. To arrive at yield per acre, the Form C estimator simply takes the weight of beans from the 3-foot section in row 1 of each of the two units, adjusts this weight to the standard moisture content, and expands this product to bushels per acre. Therefore, the Form C estimator does not rely on the Form B counts and the complex weighting scheme to calculate yield at maturity for each sample. This results in an estimator that is much easier to understand and greatly simplifies the verification of unusual yields by field office personnel. Also, the Form C estimator is less susceptible to potential bias caused by handling the plants, since it does not rely on the field counts from the 6-inch sections. It should be mentioned, however, that even though the counts of pods with beans from Forms B and C would no longer be needed to estimate yield at maturity if the Form C estimator is adopted, these counts will still be needed in the objective yield program to make yield forecasts when the crop is not mature.

## ANALYSIS

This section compares the soybean yields at maturity from the operational and Form C estimators. Soybean objective yield data from 1981 to 1984 for the 15 States in the program was used in the analysis. Yield per acre at maturity was derived for each sample using the operational and Form C estimators. Table 1 shows the mean difference between the Form C and operational yields for each year when the data is combined over all 15 States. The mean difference is expressed in bushels per acre. The yield estimates were matched for each sample and a paired t-statistic was used to test if the mean differences were significantly different from zero. A twotailed alternative hypothesis was used. This test is the usual method employed by SRS. Problems with its application due to violations of simple random sample and normality assumptions are being addressed by Fecso [1].

For 3 of the 4 years, the Form C and operational estimates were virtually the same. In 1982, however, the Form C yield estimate was about one-half bushel lower than the operational estimate. This difference was significantly different from zero and represents a 1.6-percent decrease in yield. Even though the difference in 1982 was significantly different from zero it resulted in a yield estimate that was closer to the official Board figure. For the other years the mean difference accounted for less than 1 percent of final yield.

In 1981 and 1982, the operational yield estimate from the 15 States was slightly higher than the Form C estimate, while in 1983 and 1984 it was lower. In 1981 and 1982, the soybean objective yield survey started around October 1 in 9 of the 15 States while in 1983 and 1984 the survey started around September 1 in these states. Also, the number of counts made each month in the 6-inch sections was reduced from 10 to 6 in 1982. Therefore, the Form B counts were least susceptible to plant handling effects in 1982. This fact might possibly have caused the operational estimator to behave differently in 1982, thereby contributing to the significant difference.

Table 1: Mean difference between operational and alternative soybean yield estimates for 15 States combined, 1981-84.

| Year | Sample size | Mean difference 1/ (bushels) | Significance level |
|---|---|---|---|
| 1984 | 1568 | .12 | .46 |
| 1983 | 1568 | .18 | .22 |
| 1982 | 1406 | -.52 | .01 |
| 1981 | 1415 | -.21 | .24 |

1/ Form C yield minus the operational yield.

Table 2 shows the mean differences by State. The sample sizes for each State are listed in the Appendix. An asterisk in table 2 denotes that the paired t-statistic was significantly different from zero at the .05 significance level.

The Form C and operational yield estimates were not significantly different for 53 of the 60 comparisons. With multiple tests--60 tests in table 2--we expect some significant differences to be erroneously stated due to type 1 error. Assumption violations can also increase the number of tests that appear to be significant. Therefore, the seven significant differences shown are probably not cause for alarm, especially since the differences were not consistently positive or negative.

The main reason for showing the yield differences for each state is to look for patterns in the comparisons. In general, the Form C yield estimates were not higher or lower than the operational yield estimates. The Form C estimates were lower than the operational estimates, however, for all 4 years in Iowa and South Carolina.

Table 2: Mean difference between soybean yield estimates by State, 1981-84 1/.

| State | 1984 | 1983 | 1982 | 1981 |
|---|---|---|---|---|
| | | Bushels | | |
| Alabama | -.32 | .59 | -2.05 | .87 |
| Arkansas | 1.78 * | .32 | -.72 | -.31 |
| Georgia | -.01 | -.08 | .79 | .56 |
| Illinois | 1.24 * | .61 | -.52 | -.21 |
| Indiana | 1.40 | .52 | -1.17 | 1.00 |
| Iowa | -.59 | -.62 | -.69 | -1.57 * |
| Louisiana | -.29 | 1.82 | .78 | .40 |
| Minnesota | -.68 | 1.40 * | -.26 | -.09 |
| Mississippi | .65 | -1.42 * | -.70 | .55 |
| Missouri | -.11 | .45 | .57 | -.44 |
| Nebraska | -.46 | .58 | -1.29 | .55 |
| North Carolina | -1.04 | -.37 | 1.09 | -.40 |
| Ohio | -.37 | .18 | -.92 | -.84 |
| South Carolina | -1.10 | -1.24 * | -2.03 * | -2.06 |
| Tennessee | .60 | .04 | -1.11 | .17 |

1/ Asterisk denotes significantly different from zero at the .05 significance level.

The results in tables 1 and 2 indicate that replacing the operational estimator with the Form C estimator will not significantly change the level of the objective yield estimate at maturity in most instances. For cases where the level is changed considerably, we recommend that the Crop Reporting Board rely on the Form C estimate rather than the operational estimate.

The coefficients of variation (CV's) were compared between the two yield estimators in each state for the 1984 survey to determine if the Form C estimator was as precise as the operational estimator. The CV's are shown in table 3. In 13 of the 15 States, the Form C estimator was as precise or slightly more precise than the operational estimator. At the 15-State level, the CV's for the Form C and operational yield estimators were 1.21 and 1.24 percent, respectively.

Table 3: Coefficients of variation for operational and Form C estimators, 1984 survey.

| State | Coefficient of Variation<br>Operational estimator | <br>Form C estimator |
|---|---|---|
| | Percent | |
| Alabama | 6.3 | 6.1 |
| Arkansas | 4.2 | 4.1 |
| Georgia | 7.4 | 7.5 |
| Illinois | 3.7 | 3.6 |
| Indiana | 3.7 | 3.4 |
| Iowa | 2.7 | 2.7 |
| Louisiana | 4.6 | 4.5 |
| Minnesota | 3.5 | 3.3 |
| Mississippi | 4.4 | 4.4 |
| Missouri | 4.3 | 4.3 |
| Nebraska | 5.5 | 4.9 |
| North Carolina | 4.9 | 4.8 |
| Ohio | 3.9 | 3.6 |
| South Carolina | 7.0 | 6.2 |
| Tennessee | 5.0 | 5.1 |

RECOMMENDATION

We recommend adoption of the Form C yield estimator to estimate yield at maturity for future soybean objective yield surveys. This estimator is computationally simpler, easier to understand, simplifies the verification of unusual yields by field office personnel, is less susceptible to plant-handling effects and is slightly more precise.

REFERENCES

[1] Fecso, Ron, "Test Statistics for Objective Yield Survey Data," Memo to F. Vogel, U.S. Department of Agriculture, Statistical Reporting Service, April 22,1985.

[2] Nealon, Jack, "1984 Soybean Validation Study," Statistical Reporting Service, U.S. Department of Agriculture, 1985.

[3] Nelson, D.C., "Soybean Objective Yield Destructive Counting Study," Statistical Reporting Service, U.S. Department of Agriculture, 1980.

[4] U.S. Department of Agriculture, Statistical Reporting Service. "Forecasting and Estimation Models," Objective Yield Supervising and Editing Manual, Section 15D, 1984.

[5] __________ . Soybean Objective Yield Enumerator's Manual. 1984.

APPENDIX

Number of soybean objective yield samples harvested by enumerators, 1981-84.

| State | 1984 | 1983 | 1982 | 1981 |
|---|---|---|---|---|
| Alabama | 81 | 83 | 60 | 65 |
| Arkansas | 127 | 121 | 124 | 112 |
| Georgia | 67 | 87 | 61 | 54 |
| Illinois | 155 | 159 | 160 | 160 |
| Indiana | 107 | 96 | 105 | 94 |
| Iowa | 131 | 150 | 138 | 152 |
| Louisiana | 91 | 97 | 73 | 76 |
| Minnesota | 102 | 99 | 89 | 90 |
| Mississippi | 103 | 100 | 96 | 117 |
| Missouri | 142 | 141 | 135 | 127 |
| Nebraska | 83 | 77 | 56 | 55 |
| North Carolina | 92 | 86 | 63 | 72 |
| Ohio | 113 | 101 | 112 | 102 |
| South Carolina | 93 | 93 | 64 | 65 |
| Tennessee | 81 | 78 | 70 | 74 |