United States
Department of
Agriculture

National
Agricultural
Statistics
Service

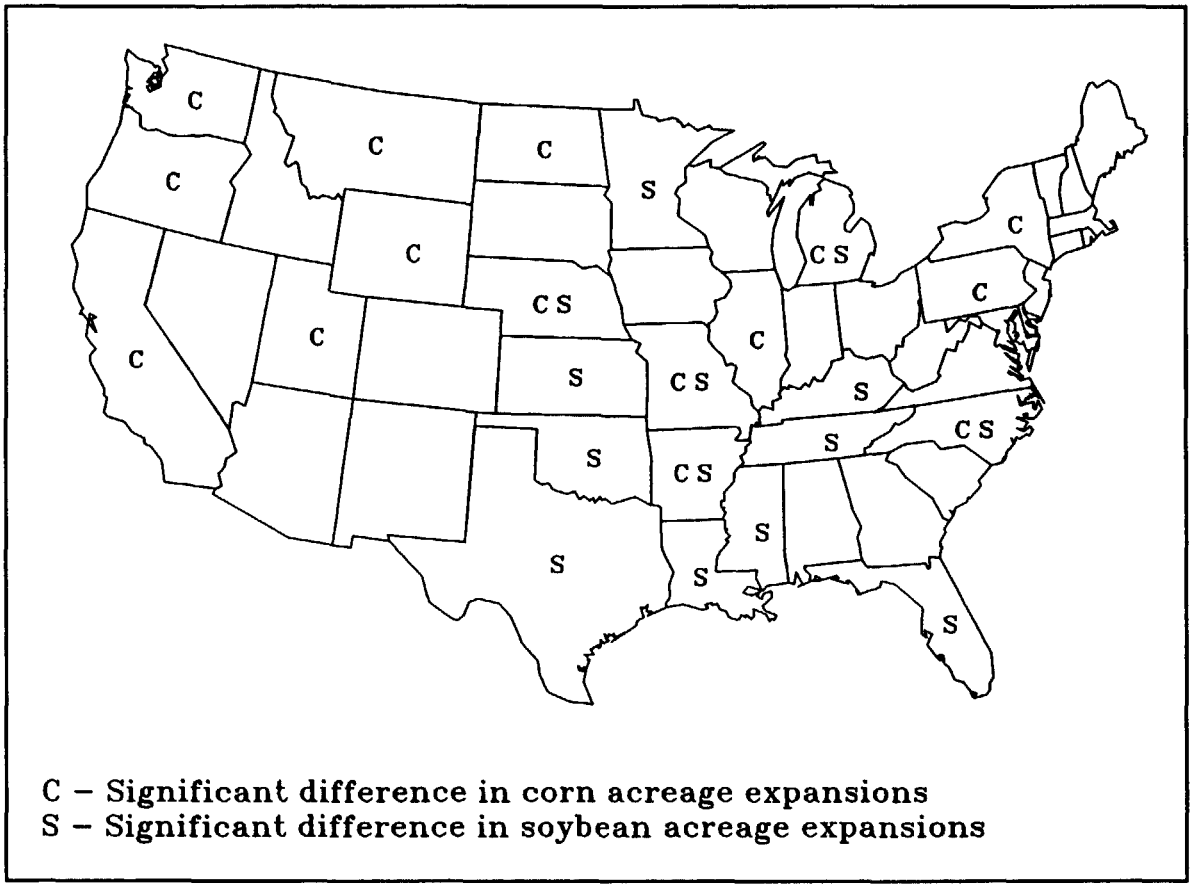
Research Division

SRB Research Report
Number SRB-93-02

January 1993

EVALUATION OF HISTORICAL DATA USE IN THE 1992 AUGUST YIELD SURVEY

Jeffrey T. Bailey



C – Significant difference in corn acreage expansions
S – Significant difference in soybean acreage expansions

EVALUATION OF HISTORICAL DATA USE IN THE 1992 AUGUST YIELD SURVEY, by Jeffrey T. Bailey. Research Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250. January 1993. Research Report No. SRB-93-02.

ABSTRACT

The National Agricultural Statistics Service (NASS) is evaluating the use of historical data in its agricultural surveys to improve data quality and reduce respondent burden. Historical data are any data previously collected for an operation that could be used to guide or check a current response. Previous research has shown (O'Connor and Mergerson, 1992) that the use of historical data as a real-time editing tool in Computer Assisted Telephone Interviewing (CATI) with the 1988 March Agricultural Survey produced estimates closer to the "truth" for grain stocks. Assuming that historical data used in the same manner will result in responses closer to the "truth" for other commodities and surveys, we need only to determine if it affects final response significantly. This research analyzed data from the 1992 August Yield Survey. The differences between the responses prior to and after a historical data check were examined to quantify the direct impact of historical data use. The results showed there was a statistically significant difference between the initial and final August responses for both corn harvested and soybean planted acres in about 1/2 of the states. Soybean planted acres were significantly different at the U.S. level. States offset each other for corn harvested acres so it was not significant at the U.S. level. The final response for soybean planted acres was much closer to the June estimated acres indicating a reduced response bias.

Key Words: Historical Data, Response Check, CATI

This paper was prepared for limited distribution to the research community outside the U.S. Department of Agriculture.

ACKNOWLEDGEMENTS

The author would like to thank Dan Capstick of the Computer Assisted Survey Section for coding the CATI instrument to collect the data and Frank Benford of the Statistical Methods Branch for modifying the computer edit to accommodate the additional items required for this analysis. The help of Mark Harris also of the Statistical Methods Branch was very beneficial in obtaining information and survey data. Thanks to Dale Atkinson for this assistance throughout the entire project.

TABLE OF CONTENTS

SUMMARY

The use of historical data in the 1992 August Yield Survey had a statistically significant impact upon the estimates. For corn harvested acres, the estimate of the total from responses prior to and after the historical data check were significantly different in 15 states. Similarly, for soybean planted acres, 15 states had a significant difference between their initial and final responses. At the U.S. level, corn harvested acres were not significantly different since many states offset one another, but soybean planted acres were significantly different. A comparison of the ratios of the initial response to June reported acres and the final response to June reported acres showed similar results.

Computer Assisted Telephone Interviewing (CATI) was used for the data collection in this study. The CATI instrument facilitates using historical data and easily captures both the responses prior to and after the historical data check. This made it possible to see how often and by how much respondents changed their answers when they were given their previous responses. When corn or soybeans was reported, 17 percent of the responses were outside the prescribed limit and were reviewed by the enumerators with the respondent. Thirty percent of those reviewed changed their August answer for corn and/or soybean acreage. This might represent a lower bound since respondents are reluctant to change a current response. Those reviewed that did not change their answer had a smaller difference between their initial August response and June response than those who did change their answer.

These results show that using historical data as an editing tool during CATI has a substantial impact upon the data, but have we reduced total response error? The previous research that used historical data as a response check (O'Connor and Mergerson, 1992) showed that not only were the estimates significantly different but they were closer to the "truth." For this study, the final responses are closer to the June estimates than the initial responses, evidence that we have reduced response errors.

Partially offsetting the benefit of reduced response errors are the cost, increased interviewer time, increased workload of handling historical data, risk of bias, and possibly increased respondent burden, although the on-line editing could actually reduce respondent burden if it saves a call-back. However, considering the significant impact that real-time editing with historical data has on the survey estimates, its use is warranted.

INTRODUCTION

The National Agricultural Statistics Service (NASS) is investigating the use of previously reported data during later agricultural survey interviews. This research evaluates the real-time editing use of historical data in the 1992 August Yield Survey, one of a series of monthly surveys conducted from May to November each year. While the primary purpose of these monthly surveys is to estimate yields, indications of acreages are also obtained from ratios calculated from the current survey to the base survey. The base survey is the March Agricultural Survey or June Agricultural Survey depending upon the crop. The sample consists of list frame operations that reported the crops of interest during the base survey.

For the August Yield Survey, row crops are subsampled from the June survey and small grains are subsampled from March. The acres reported in the August Yield Survey are compared to the reported acres in the base survey for each crop. For this evaluation of historical data we looked at corn harvested acres and soybean planted acres only. As row crops, corn and soybeans are subsampled from the June Agricultural Survey.

The majority of the data for the August Yield Survey was collected by Computer Assisted Telephone Interviewing (CATI). CATI simplifies comparisons with previously reported data. The CATI instrument edits the responses entered by the enumerator and routes the flow of questions that appear on the screen. When a response is outside a prescribed range from the previously reported data, it will present a response review screen for the enumerator to verify the response with the respondent. With CATI we are able to store the responses both before and after their answer is reviewed by the enumerator with the respondent. This enabled us to analyze the direct effect of using historical data by comparing the two responses.

BACKGROUND

Total survey error is commonly divided into sampling error and nonsampling error. Sampling error is that part which is due to the fact that only a subset of the population is observed rather than the entire population. Nonsampling errors are all other errors, some of which are due to response, coverage, nonresponse and processing. Response errors are one type of nonsampling error of particular concern in this study. Response errors have been defined simply as the "difference between reported and true value" (Sanchez-Crespo, 1975). Some response errors are caused by respondents when they guess or don't know the answer, misunderstand the question or deliberately give false information. Other response errors are caused by the enumerators recording the incorrect value, misunderstanding the answer or failing to ask the question correctly.

One way NASS is trying to help control response errors is to use historical data. Since

NASS makes repeated contacts of many operations during the year, previously reported data are available to use in an effort to improve the current response. There are different ways to use historical data. It can be used directly by telling the respondent their previous answer prior to obtaining a current survey response, or as a response check after their response. Depending on how it is used, care must be taken to ensure that bias is not introduced.

Previous research was done in the collection of acreage data in the 1985 California Fall Acreage and Production Survey (Pafford, 1986). In this study the historical data were analyzed at four different levels of usage. Treatment 1 (control) did not use historical data at all. Treatment 2 (probe) provided the previous data on the CATI screen for the enumerators to probe the respondent when they thought an answer deviated too much. Treatment 3 (upfront) had the previous response worded into the question. Treatment 4 (not asked) did not ask for a current survey response when historical data were available. As expected the upfront use of historical data gave expansions closest to the historical data and the largest differences were between the control and the upfront use. From this study it was recommended that when historical data are used for probing, it needs to be done consistently. The enumerators should all know when to probe based on some limits and probe in the same manner. This can be done with CATI routing the flow to a response review screen only when the response is outside a prescribed limit.

In the 1986 April ISP Grain Stocks Survey (Pafford, 1988) a split sample was used with one group not using historical data and the second with direct use. The historical data were the January grain stocks data which are normally higher than April grain stocks since the grains are consumed and sold between those months. Since January stocks are higher it was found, as expected, that the group with the direct use of historical data was significantly higher. They found when farmers are given the previous number, they tend to answer the current question with "the same as before," often resulting in a biased response. Therefore, it was recommended that historical data should not be given directly to the respondent, but should be used only as an editing tool to check unusual responses.

More recent research was done on the use of historical data in CATI grain stocks enumeration (Mergerson & O'Connor, 1992). In this research historical data were used as an edit check and not given directly to the respondent. The study was done in conjunction with a reinterview project where the "truth" was obtained by reinterviewing respondents face to face. The results indicated that when historical data were used the responses were closer to the "truth" and were significantly different from their original responses.

With these findings NASS began using historical data as a response check during CATI interviews. Using historical data in CATI has several advantages. 1) It provides consistency of historical data usage among enumerators. 2) It is used as necessary, having the enumerator review the response with the respondent only when it is outside

some prescribed limit, thus not over burdening the respondent. 3) It doesn't bias the response since it is provided after the original response.

The impact of historical data is still being discussed. How much should we use it? Is it worth the effort to incorporate the historical data into the CATI instrument? Does it have practical significance upon the data? Using a large scale study will help to answer these questions and with NASS moving toward more telephone data collection with CATI we will have the ability to collect and study the issue further. CATI provides a means to easily do response checks during the interview and store an individual's responses both before and after the check. Previous research used a split sample to measure the effect of historic data, but with CATI we can assess the "direct" impact upon an individual's response. That is, did a respondent change an answer and if so by how much, when asked to verify the relationship to the historical number.

METHODS

The 1992 August Yield Survey was conducted in 48 states with a sample size of 29,214. The sample sizes varied between states depending upon the crops grown in the state, the relative importance of the state's production of the crop to the U. S. total and the number of positive responses from the base survey. If there were not enough positive reports for a particular crop, extra replications were added to obtain sufficient sample sizes. All states had CATI available for the data collection.

Historical data were used as an editing tool to prompt the review of responses for acres reported that were outside the prescribed range. The range for this survey was a modified 25 percent limit in which (reported August acres + 100) divided by (June acres + 100) was compared to lower and upper limits of 0.75 and 1.25, respectively. This range was used for ease of programming and simplicity. The previous study (Mergerson & O'Connor, 1992) with grain stocks had different ranges that varied by the size of the response. The CATI instrument for this survey did the calculation and controlled the flow of questions that were asked during the interview. Although the comparisons were done for all crops during the survey, only data for corn and soybean acres were captured and studied.

The following example explains the sequence of the CATI session using corn harvested acres that were collected in August.

1. The respondent reports his corn harvested acres and it is entered into the CATI instrument by the enumerator.
2. After the CATI instrument compares this value against the state's specified upper acreage limit for corn and routes the session to a response review screen if needed, the current response is saved as a separate variable that can not be changed.

3. The CATI instrument compares the response to data collected in June and if the response falls outside the modified 25 percent limit the response review screen appears (See sample below).

4. The enumerator verifies that the response is correct or makes changes.

5. The initial and final responses are output from CATI.

The sequence described above is important to this study. Since the CATI instrument retains the initial response after the upper acreage limit check and before the historical data check, any difference between the initial and final values should be indicative of the impact of the historical data edit. The initial response is retained in a separate variable that can not be changed by the enumerator. Some states ask irrigated and nonirrigated acres for corn and soybeans or double and single crop soybeans. In these states the check is made only on the combined total acres after both acres and yield are collected for the parts.

Sample Response Review Screen in CATI Instrument:

```
CASEID: 00001

>e531<
     VERIFY CHANGE IN CORN ACRES FROM JUNE TILL NOW!

     Our records show that on JUNE 1, this operation had
     xxx acres of corn intended for harvest.

     I now record xxx acres of corn for harvest.

     Do I have this recorded correctly?
     ----------------------------------------------------

     JUNE ACRES        ..................... xxx
     CURRENT ACRES     ..................... xxx     Ctrl-F1

     <1> YES, (NO CHANGES NEEDED FOR CURRENT ACRES) (specify)
     ===>
```

Upon completion of an interview with CATI, the data are edited with a mainframe batch computer edit. The error print from the Survey Processing System (SPS) edit is then reviewed by statisticians in the state offices and the final data values from CATI can be updated. Since it is possible for the result from CATI to be changed during this process and the final number that was output from CATI was not saved as a separate variable, we were not totally able to isolate CATI changes from subsequent SPS editing changes. For purposes of this study, we will assume that if the initial response was outside the modified 25 percent limit and reviewed during the CATI session, the change was made at that time. If the initial response was within the 25 percent limit and not reviewed during the CATI session, then the change was made in the editing process. Based on these assumptions we will be using only those changes attributable to the CATI session for this study.

To assess the effect of historical data two different tests were done. First, the difference between the expansion totals for the initial and final response were tested with a stratified univariate t-test (See Appendix B). To obtain the expansion totals and to make this test, several pieces of information were needed that were not available on the data file. Since only positive reports from June were sampled, the effective sample size should also represent the zero reports in June. For this analysis, the expansion factor was obtained from the summary data file and the population counts were obtained from another file. Dividing the population counts by the expansion factor calculated during the summary gave us the usable sample size. The crop acreage estimates of the total cover less than 50 percent of the U. S. acreage for corn and soybeans since many large strata were not sampled for this survey.

The major acreage indications from this survey are not the direct expansions, but the ratios of August reported to June reported acreage. These ratios were computed as a ratio of the expansions calculated for testing the differences in expansions. The test compared the ratios of the initial responses to June to the ratios of the final responses to June. (Appendix C).

RESULTS

All the offices except Alabama, Arizona, New Jersey, New England, and South Carolina used CATI for data collection on a total of 23,859 cases (See Table 1, appendix page A-1). These cases resulted in 20,260 usable reports, of which 12,737 were sampled from June and had historical data for both corn and soybeans to use as a check during the interview.

A total of 6,547 of those interviewed with CATI had positive **corn harvested acres** reported in June or August (See Table 2, appendix page A-2). There were 16.6 percent of these cases where the response was outside the modified 25 percent limit and reviewed by the enumerator with the respondent. Of those reviewed, 28.4 percent changed their initial response. The net average change at the U.S. level was a gain of

2 acres. Several states had large differences, but at the U. S. level individual state differences were offsetting. The absolute average difference between the initial August response and the June response was 52 acres (23 percent) smaller at the U. S. level for those that did **not** change their answer than for those who did change their answer.

A total of 5,164 of those interviewed with CATI had positive **soybean planted acres** reported in June or August (See Tables 3, appendix page A-3). There were 18.2 percent of these cases where the response was outside the modified 25 percent limit and reviewed by the enumerator with the respondent. Of those reviewed, 31.1 percent changed their initial response. The net average change at the U. S. level was a gain of 169 acres. **This is not the change from June but the average change made in their initial August response during the August interview.** In Pennsylvania, one respondent reported 0 acres of soybeans and changed to 1,700 acres after the response check. Nebraska had two operations that initially reported 0 acres of corn and changed to 720 and 750 acres after the response check. The absolute average difference between the initial August response and the June response was 82 acres (28 percent) smaller at the U. S. level for those that did **not** change their answer than for those who did change their answer.

Maps of the number of times an answer was changed indicate that the states with larger corn and soybean acreage made more changes, but on a percentage basis there was no pattern (Figures 1 and 3 appendix, pages A-4 and A-5). Maps of the average change made revealed no regional patterns across the U.S. (See Figures 2 and 4, appendix pages A-4 and A-5). States that asked irrigated soybeans had larger changes made to nonirrigated acres. The difference in other soybean producing states that only asked nonirrigated acres were similar, so no particular type of state seemed to cause significance at the U.S. level. It appears that respondents know their irrigated acres better, which is not surprising since farmers know how many acres their irrigation equipment covers.

A test was done to determine if the expanded totals were significantly different between initial and final responses. The test results indicated significance in 15 states at the 5 percent level for corn, but none at the U. S. level (See Table 4, appendix page A-5). A Bonferroni p-value was calculated for all these tests to account for the number of tests being done. For soybeans, 15 states and the U.S. total showed significance at the 5 percent level (See Table 5, appendix page A-6).

The important acreage indications for this survey are the ratios of August reported to June reported acreage. The expansions for the ratio estimates were calculated the same as in the test for difference in the direct expansions. The ratio estimates calculated are combined ratio estimates at the state level across all agricultural statistics districts (ASD). The operational monthly Ag Yield summary prints both these combined ratio estimates and the ASD ratios weighted by the respective crop acres in each ASD. When we compare the difference in ratios using the initial and final August responses to June reported acreage, there were 15 states significantly different for corn (See Table 6,

appendix page A-7) while the U.S. level was not significantly different. For soybeans, 17 states and the U.S. are significant at the 5 percent level (See Table 7, appendix page A-7).

DISCUSSION AND RECOMMENDATIONS

The use of historical data can have a large impact on survey results. If historical data are properly used, reduced response errors can result. The question is, "Do the benefits outweigh the costs?" There are several actual and potential "costs" involved: 1) Handling and storing the historical data requires time and storage space, 2) Programming CATI requires labor to write and keep the code updated, 3) There is increased time of interview and possibly extra response burden, and 4) There is a chance of the introduction of bias. Some of the benefits are: 1) The response errors may be reduced, 2) Time is saved in the field offices that would otherwise have been spent checking reports and recalling respondents, and 3) Data collected by phone may be improved. With the expanded use of phone data collection that lacks the personal face-to-face contact, additional checks are needed to attempt to maintain data quality.

I recommend that NASS continue to expand the use of historical data as a real-time editing tool in CATI. This is especially useful in NASS' agricultural surveys where repeated contacts are made. The historical data used should be recent data that are highly related to the current item. As the use of historical data is expanded the impact should be monitored and its benefits evaluated. The Research Division plans to work with Estimates Division in this type of impact monitoring.

In future research with CATI data collection, CATI should output not only the initial and final response, but output the final value twice, one that can be edited and one that can not be edited in the SPS edit. This would allow us to isolate changes made during the CATI session from those made during the subsequent SPS edit. It would also be interesting to determine how many of the large differences reviewed in CATI would have been found during the editing process. No doubt some of these might have been found, but there are others that would never have been discovered.

We have assumed that the response bias has been reduced since we are using historical data in a manner that has been shown to reduce response bias. For specific operations it is obvious that the corrected number is more accurate when the respondent has forgotten or omitted a response, but we can not be certain that our overall bias is reduced. From other experiences, respondents are 10 times more likely to correct a previous response over a current response when one would not expect either number to be more accurate than the other. This fact puts considerable pressure on the reconciliation procedure used to determine a "proxy to truth." Considerable effort and attention is required in planning and implementing this activity, whether the reconciliation is part of a reinterview survey or is accomplished with additional probing questions in the regular survey contact.

BIBLIOGRAPHY

Cochran, William G. (1977). Sampling Techniques, Third Edition, Wiley & Sons, New York.

Mergerson, James W. and O'Connor, Terry P. (1992). "The Effect of Using Historic Data in CATI Grain Stocks Enumeration in the March 1988 Agricultural Survey." United States Department of Agriculture, National Agricultural Statistics Service, Research and Applications Division, NASS SRB Research Report Number SRB-92-01.

Pafford, Brad (1986). "Response Errors in Nass Surveys: The Effect of Using Previous Survey Data in the 1985 California fall Acreage and Production Survey." United States Department of Agriculture, National Agricultural Statistics Service, Statistical Research Division, NASS Staff Report Number SF&SRB 86-99.

Pafford, Bradley V. (1988). "The Influence of Using Previous Survey Data in the 1986 April ISP Grain Stocks Survey." United States Department of Agriculture, National Agricultural Statistics Service, Research and Applications Division, NASS Research Report Number SRB-88-01.

Sanchez-Crespo, J. L. (1975). "Notes on the Accuracy, Precision and Reliability of Statistical Data", Bulletin of International Statistical Institute, Warsaw.

Table 1: August 1992 Ag Yield Survey Contacts Completed on CATI

State	Total Sample Size	Total Cases Attempted on CATI		Total Usable CATI Cases		Usable CATI Cases with Matching June	
		No.	Percent	No.	Percent	No.	Percent
AL	249	-	-	-	-	-	-
AZ	158	-	-	-	-	-	-
AR	536	535	99.8	434	81.1	434	100.0
CA	681	681	100.0	516	75.8	377	73.1
CO	667	555	83.2	442	79.6	260	58.8
CT	64	-	-	-	-	-	-
DE	69	69	100.0	43	62.3	43	100.0
FL	252	174	69.0	158	90.8	158	100.0
GA	467	382	81.8	302	79.1	302	100.0
ID	674	672	99.7	547	81.4	289	52.8
IL	1,271	736	57.9	659	89.5	420	63.7
IN	1,060	587	55.4	508	86.5	307	60.4
IA	1,245	1,245	100.0	972	78.1	533	54.8
KS	1,540	1,096	71.2	972	88.7	436	44.9
KY	760	703	92.5	565	80.4	465	82.3
LA	349	176	50.4	166	94.3	166	100.0
ME	91	-	-	-	-	-	-
MD	231	231	100.0	170	73.6	139	81.8
MA	104	-	-	-	-	-	-
MI	868	689	79.4	526	76.3	298	56.7
MN	1,326	1,326	100.0	985	74.3	543	55.1
MS	453	453	100.0	369	81.5	369	100.0
MO	1,160	691	59.6	624	90.3	405	64.9
MT	867	865	99.8	796	92.0	336	42.2
NE	1,265	1,016	80.3	879	86.5	539	61.3
NV	75	51	68.0	46	90.2	46	100.0
NH	72	-	-	-	-	-	-
NJ	141	-	-	-	-	-	-
NM	285	246	86.3	210	85.4	210	100.0
NY	465	383	82.4	249	65.0	168	67.5
NC	790	750	94.9	655	87.3	479	73.1
ND	1,220	1,082	88.7	995	92.0	398	40.0
OH	1,091	1,084	99.4	1002	92.4	518	51.7
OK	990	958	96.8	883	92.2	463	52.4
OR	412	412	100.0	351	85.2	230	65.5
PA	681	681	100.0	552	81.1	342	62.0
RI	12	-	-	-	-	-	-
SC	199	-	-	-	-	-	-
SD	1,228	1,221	99.4	1079	88.4	521	48.3
TN	656	602	91.8	545	90.5	545	100.0
TX	1,247	803	64.4	739	92.0	455	61.6
UT	294	294	100.0	270	91.8	185	68.5
VT	131	-	-	-	-	-	-
VA	442	442	100.0	323	73.1	286	88.5
WA	575	447	77.7	387	86.6	200	51.7
WV	250	250	100.0	212	84.8	212	100.0
WI	1,235	1,023	82.8	931	91.0	516	55.4
WY	316	248	78.5	198	79.8	144	72.7
US	29,214	23,859	81.7	20,260	84.9	12,737	62.9

Table 2: Historical Data Use in August 1992 Ag Yield Survey, Corn Harvested Acres

State	Total CATI Interviews With Historical Data	Reports With Positive Corn Acres Reported	Cases Where August Response Was Outside Historical Edit Limits 1/		Cases Where August Acres Were Changed When Given Historical Data		Average Acres Changed From Initial to Final Response
			No.	Percent	No.	Percent	
AR	434	50	14	28.0	8	57.1	158
CA	377	44	26	59.1	10	38.5	36
CO	260	114	23	20.2	0	0.0	-
DE	43	35	5	14.3	0	0.0	-
FL	158	90	11	12.2	4	36.4	426
GA	302	174	24	13.8	9	37.5	-206
ID	289	29	11	37.9	3	27.3	50
IL	420	392	45	11.5	26	57.8	-53
IN	307	282	39	13.8	7	17.9	-39
IA	533	514	76	14.8	42	55.3	123
KS	436	150	30	20.0	0	0.0	-
KY	465	282	37	13.1	11	29.7	41
LA	165	37	9	24.3	1	11.1	-190
MD	139	98	20	20.4	0	0.0	-
MI	298	247	50	20.2	14	28.0	-53
MN	543	436	59	13.5	22	37.3	-33
MS	369	99	20	20.2	6	30.0	-73
MO	405	216	35	16.2	12	34.3	-71
MT	336	30	13	43.3	5	38.5	-55
NE	539	447	53	11.9	4	7.5	510
NM	210	43	10	23.3	1	10.0	-140
NY	168	98	24	24.5	7	29.2	-291
NC	479	266	36	13.5	20	55.6	-117
ND	398	133	39	29.3	8	20.5	-34
OH	518	465	50	10.8	13	26.0	-3
OK	463	54	18	33.3	0	0.0	-
OR	230	25	12	48.0	5	41.7	162
PA	342	276	33	12.0	10	30.3	-10
SD	521	392	81	20.7	19	23.5	-71
TN	545	220	39	17.7	16	41.0	73
TX	455	90	23	25.6	0	0.0	-
UT	185	29	7	24.1	2	28.6	-46
VA	286	130	26	20.0	2	7.7	-57
WA	200	28	9	32.1	3	33.3	200
WV	212	64	4	6.2	3	75.0	0
WI	516	449	70	15.6	14	20.0	27
WY	144	19	7	36.8	2	28.6	-125
Total	12,690	6,547	1,088	16.6	309	28.4	2

1/ Error screen appears for cases when (reported acres +100) divided by (June acres + 100) is greater than 1.25 or less than 0.75.

Table 3: Historical Data Use in August 1992 Ag Yield Survey, Soybean Planted Acres

State	Total CATI Interviews With Historical Data	Reports With Positive Soybean Acres Reported	Cases Where August Response Was Outside Historical Edit Limits 1/		Cases Where August Acres Were Changed When Given Historical Data		Average Acres Changed From Initial to Final Response
			No.	Percent	No.	Percent	
AR	434	343	114	33.2	71	62.3	302
DE	43	37	6	16.2	1	16.7	92
FL	158	58	13	22.4	2	15.4	63
GA	302	117	29	24.8	6	20.7	31
IL	420	385	42	10.9	12	28.6	39
IN	307	262	44	16.8	8	18.2	-25
IA	533	449	70	15.6	33	47.1	106
KS	436	180	53	29.4	13	24.5	147
KY	465	207	33	15.9	11	33.3	376
LA	165	78	29	37.2	3	10.3	46
MD	139	74	13	17.6	1	7.7	193
MI	298	144	24	16.7	6	25.0	79
MN	543	384	43	11.2	13	30.2	121
MS	369	199	38	19.1	4	10.5	44
MO	405	284	53	18.7	16	30.2	-16
NE	539	289	55	19.0	23	41.8	78
NC	479	292	46	15.8	11	23.9	61
ND	398	74	7	9.5	1	14.3	-70
OH	518	398	46	11.6	7	15.2	75
OK	463	59	24	40.7	6	25.0	245
PA	342	112	4	3.6	1	25.0	1700
SD	521	262	47	17.9	12	25.5	312
TN	545	170	54	31.8	16	29.6	170
TX	455	49	28	57.1	7	25.0	148
VA	286	95	15	15.8	5	33.3	339
WI	516	163	8	4.9	3	37.5	8
Total	10,080	5,164	938	18.2	292	31.1	169

1/ Error screen appears for cases when (reported acres +100) divided by (June acres + 100) is greater than 1.25 or less than 0.75.

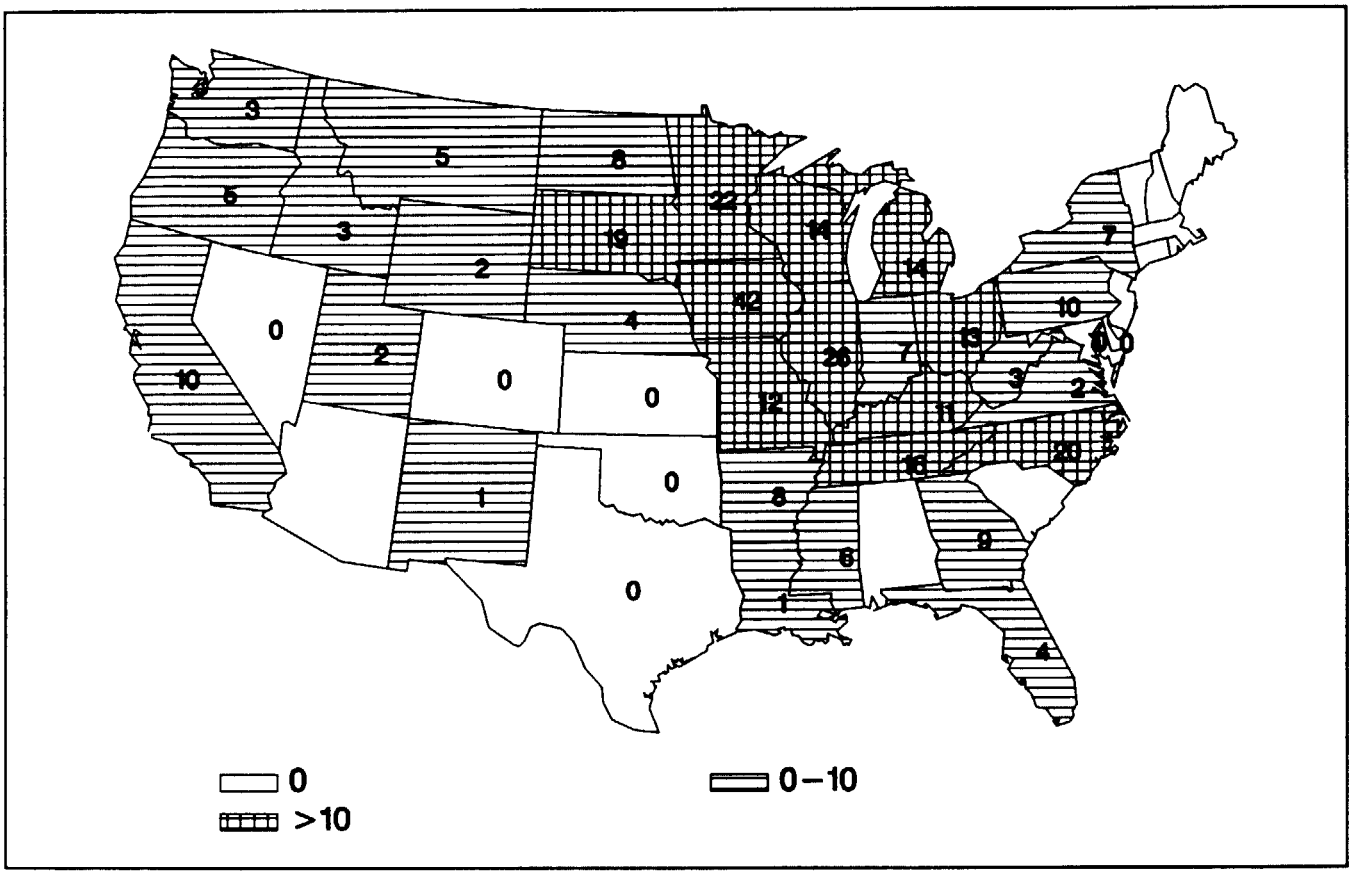


Figure 1: Number of Times Answer was Changed, Corn Harvested Acres

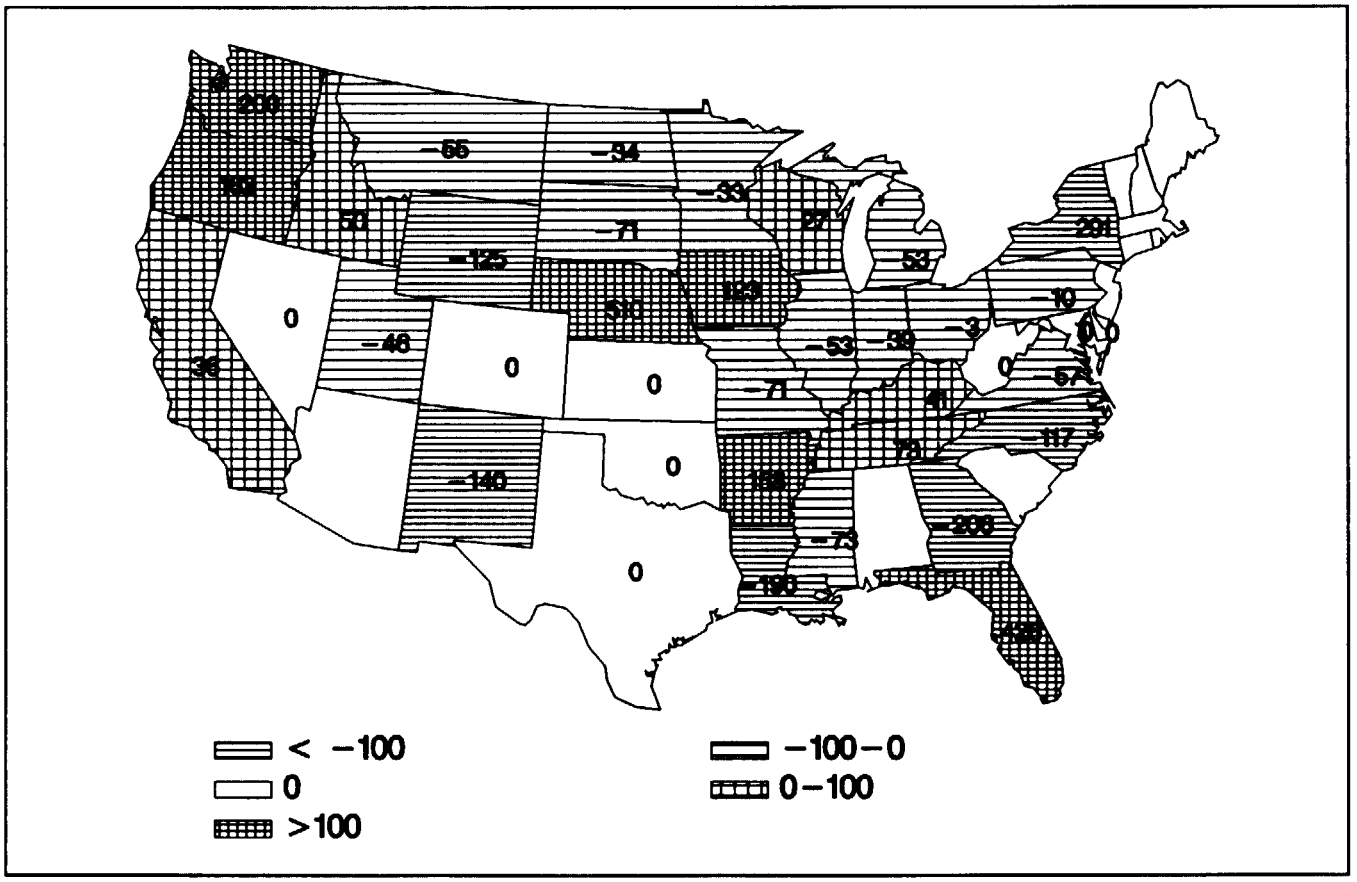


Figure 2: Average Change Made in CATI, Corn Harvested Acres

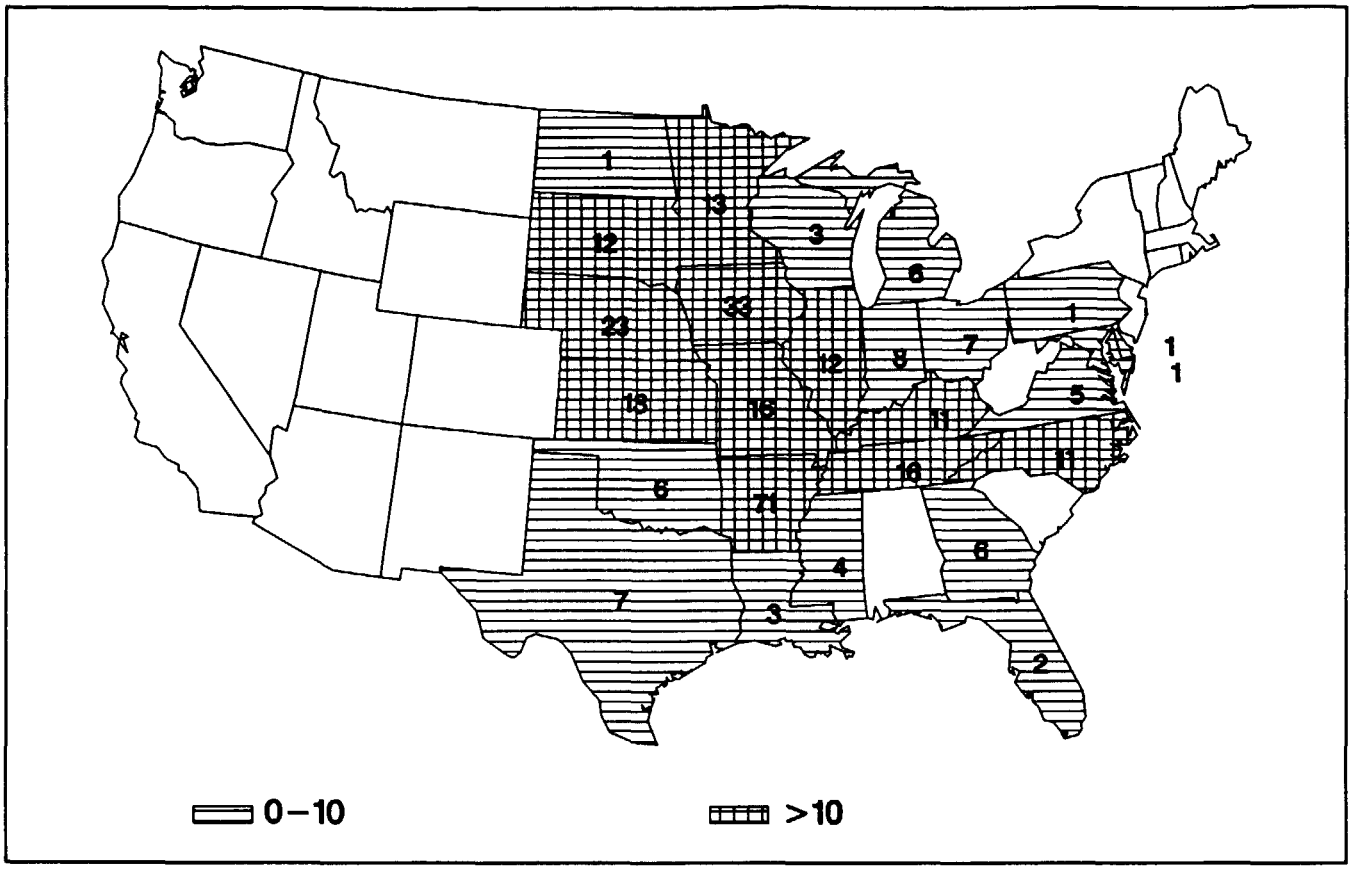


Figure 3: Number of Times Answer was Changed, Soybean Planted Acres

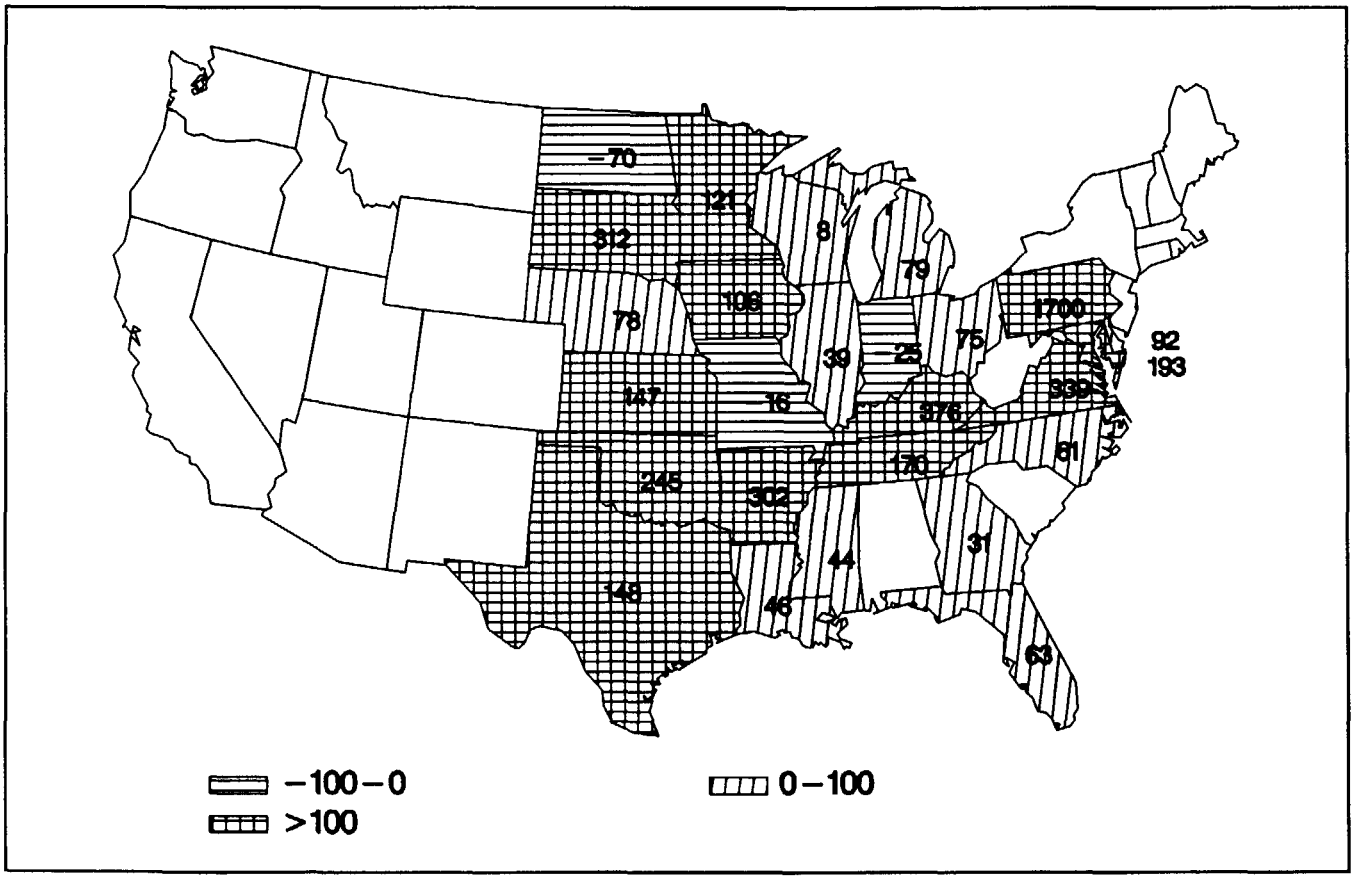


Figure 4: Average Change Made in CATI, Soybean Planted Acres

Table 4: Difference Between Initial and Final Responses, Corn Harvested Acres

State	August 1 Expansions 1/		Expanded Difference Final-Initial Response	Ratio Final / Initial Response	P-Value for Testing Differences	
	Using Final Edited Value	Using Initial Response				
AR	41,264	35,307	5,957	1.17	0.0002	*
CA	36,474	37,812	-1,338	0.96	0.0000	*
CO	215,850	215,850	0	1.00	.	
DE	38,435	38,435	0	1.00	.	
FL	61,654	43,836	17,818	1.41	0.0443	
GA	495,145	524,782	-29,637	0.94	0.0048	
ID	25,655	30,297	-4,642	0.85	0.2486	
IL	5,749,044	5,778,956	-29,912	0.99	0.0001	*
IN	4,343,184	4,339,240	3,944	1.00	0.9010	
IA	9,696,206	9,373,278	322,928	1.03	0.0062	
KS	521,637	521,637	0	1.00	.	
KY	485,111	475,292	9,819	1.02	0.0025	
LA	147,089	149,094	-2,005	0.99	0.0488	
MD	196,638	196,638	0	1.00	.	
MI	1,407,132	1,444,865	-37,733	0.97	0.0000	*
MN	1,805,859	1,809,625	-3,766	1.00	0.1654	
MS	77,474	88,095	-10,621	0.88	0.0039	
MO	811,638	826,885	-15,247	0.98	0.0000	*
MT	14,483	16,754	-2,271	0.86	0.0004	*
NE	2,402,344	2,372,182	30,162	1.01	0.0000	*
NM	35,315	35,949	-634	0.98	0.1568	
NY	156,725	178,720	-21,995	0.88	0.0000	*
NC	336,870	367,151	-30,281	0.92	0.0000	*
ND	444,373	450,161	-5,788	0.99	0.0003	*
OH	1,044,174	1,055,627	-11,453	0.99	0.0569	
OK	49,266	49,266	0	1.00	.	
OR	9,957	7,464	2,493	1.33	0.0001	*
PA	319,296	324,950	-5,654	0.98	0.0001	*
SD	2,824,105	2,864,175	-40,070	0.99	0.1972	
TN	284,064	281,216	2,848	1.01	0.0975	
TX	580,089	580,089	0	1.00	.	
UT	7,473	9,037	-1,564	0.83	0.0000	*
VA	126,121	125,826	295	1.00	0.4606	
WA	27,719	24,691	3,028	1.12	0.0000	*
WV	18,812	18,984	-172	0.99	0.1509	
WI	1,589,732	1,596,067	-6,335	1.00	0.2779	
WY	38,459	40,036	-1,577	0.96	0.0002	*
Total	36,464,866	36,328,269	136,597	1.00	0.2860	

* Significant at .05 level [.05/37=.00135=Bonferroni p]

1/ The expansions do not cover the entire population since some strata were excluded from the survey.

Table 5: Difference Between Initial and Final Responses, Soybean Planted Acres

State	August 1 Expansions1/		Expanded Difference Final-Initial Response	Ratio Final / Initial Response	P-Value for Testing Differences
	Using Final Edited Value	Using Initial Response			
AR	937,684	830,400	107,284	1.13	0.0000 *
DE	63,315	62,515	800	1.01	0.0005 *
FL	32,745	32,104	641	1.02	0.0000 *
GA	460,875	451,516	9,359	1.02	0.0815
IL	5,001,402	4,994,517	6,885	1.00	0.3968
IN	3,408,788	3,404,302	4,486	1.00	0.4675
IA	6,527,635	6,309,155	218,480	1.03	0.0037
KS	554,807	518,750	36,057	1.07	0.0000 *
KY	461,773	426,624	35,149	1.08	0.0000 *
LA	546,580	541,956	4,624	1.01	0.0003 *
MD	146,175	142,657	3,518	1.02	0.0036
MI	652,788	635,865	16,923	1.03	0.0003 *
MN	1,517,350	1,489,917	27,433	1.02	0.0000 *
MS	392,222	391,386	836	1.00	0.0000 *
MO	1,440,559	1,463,607	-23,048	0.98	0.0000 *
NE	711,306	682,245	29,061	1.04	0.0000 *
NC	446,431	442,663	3,768	1.01	0.0001 *
ND	373,014	375,487	-2,473	0.99	0.0028
OH	1,057,414	1,054,950	2,464	1.00	0.2904
OK	119,641	105,886	13,755	1.13	0.0000 *
PA	122,879	86,849	36,030	1.41	0.0054
SD	1,943,740	1,848,851	94,889	1.05	0.0033
TN	373,749	352,491	21,258	1.06	0.0000 *
TX	152,057	147,205	4,852	1.03	0.0000 *
VA	183,679	169,607	14,072	1.08	0.0631
WI	338,992	341,330	-2,338	0.99	0.1545
Total	27,967,600	27,302,835	664,765	1.02	0.0000 *

* Significant at .05 level [.05/27=.00185=Bonferroni p]

1/ The expansions do not cover the entire population since some strata were excluded from the survey.

Table 6: Difference in Ratios of August Reported to June Reported, Corn Harvested Acres

State	August 1 Expansions 1/			Ratio of Final / June Value	Ratio of Initial / June Value	P - Value of Test in Difference
	Using Final Edited Value	Using Initial Response	Using June Value			
AR	41,264	35,307	40,858	1.01	0.86	0.0000 *
CA	36,474	37,812	44,182	0.83	0.86	0.0000 *
CO	215,850	215,850	224,964	0.96	0.96	.
DE	38,435	38,435	40,000	0.96	0.96	.
FL	61,654	43,836	72,079	0.86	0.61	0.0333
GA	495,145	524,782	522,938	0.95	1.00	0.0054
ID	25,655	30,297	33,480	0.77	0.90	0.2575
IL	5,749,044	5,778,956	5,919,711	0.97	0.98	0.0000 *
IN	4,343,184	4,339,240	4,403,741	0.99	0.99	0.8181
IA	9,696,206	9,373,278	9,852,155	0.98	0.95	0.0025
KS	521,637	521,637	523,022	1.00	1.00	.
KY	485,111	475,292	480,532	1.01	0.99	0.0018
LA	147,089	149,094	147,449	1.00	1.01	0.0494
MD	196,638	196,638	184,144	1.07	1.07	.
MI	1,407,132	1,444,865	1,420,378	0.99	1.02	0.0000 *
MN	1,805,859	1,809,625	1,857,520	0.97	0.97	0.1365
MS	77,474	88,095	80,410	0.96	1.10	0.0039
MO	811,638	826,885	787,617	1.03	1.05	0.0000 *
MT	14,483	16,754	15,387	0.94	1.09	0.0000 *
NE	2,402,344	2,372,182	2,374,854	1.01	1.00	0.0000 *
NM	35,315	35,949	45,672	0.77	0.79	0.1584
NY	156,725	178,720	161,123	0.97	1.11	0.0000 *
NC	336,870	367,151	340,914	0.99	1.08	0.0000 *
ND	444,373	450,161	455,274	0.98	0.99	0.0005 *
OH	1,044,174	1,055,627	1,047,493	1.00	1.01	0.0544
OK	49,266	49,266	49,020	1.01	1.01	.
OR	9,957	7,464	12,544	0.79	0.59	0.0000 *
PA	319,296	324,950	326,400	0.98	1.00	0.0001 *
SD	2,824,105	2,864,175	2,914,799	0.97	0.98	0.2006
TN	284,064	281,216	294,628	0.96	0.95	0.0937
TX	580,089	580,089	515,280	1.13	1.13	.
UT	7,473	9,037	7,828	0.95	1.15	0.0000 *
VA	126,121	125,826	129,614	0.97	0.97	0.4607
WA	27,719	24,691	26,630	1.04	0.93	0.0000 *
WV	18,812	18,984	19,448	0.97	0.98	0.1634
WI	1,589,732	1,596,067	1,633,679	0.97	0.98	0.1884
WY	38,459	40,036	33,950	1.13	1.18	0.0002 *
Total	36,464,866	36,328,269	37,039,717	0.98	0.98	0.2306

* Significant at .05 level [.05/37=.00135=Bonferroni p]

1/ The expansions do not cover the entire population since some strata were excluded from the survey.

Table 7: Difference in Ratios of August Reported to June Reported, Soybean Planted Acres

State	August 1 Expansions 1/			Ratio of Final / June Value	Ratio of Initial / June Value	P - Value of Test in Difference
	Using Final Edited Value	Using Initial Response	Using June Value			
AR	937,684	830,400	913,869	1.03	0.91	0.0000 *
DE	63,315	62,515	62,296	1.02	1.00	0.0006 *
FL	32,745	32,104	40,260	0.81	0.80	0.0000 *
GA	460,875	451,516	499,634	0.92	0.90	0.0793
IL	5,001,402	4,994,517	5,148,706	0.97	0.97	0.1105
IN	3,408,788	3,404,302	3,551,393	0.96	0.96	0.2337
IA	6,527,635	6,309,155	6,585,805	0.99	0.96	0.0003 *
KS	554,807	518,750	569,309	0.97	0.91	0.0000 *
KY	461,773	426,624	445,550	1.04	0.96	0.0000 *
LA	546,580	541,956	590,730	0.93	0.92	0.0003 *
MD	146,175	142,657	170,242	0.86	0.84	0.0026
MI	652,788	635,865	748,575	0.87	0.85	0.0002 *
MN	1,517,350	1,489,917	1,557,137	0.97	0.96	0.0000 *
MS	392,222	391,386	430,115	0.91	0.91	0.0000 *
MO	1,440,559	1,463,607	1,444,951	1.00	1.01	0.0000 *
NE	711,306	682,245	735,868	0.97	0.93	0.0000 *
NC	446,431	442,663	460,509	0.97	0.96	0.0000 *
ND	373,014	375,487	383,894	0.97	0.98	0.0029
OH	1,057,414	1,054,950	1,047,216	1.01	1.01	0.2382
OK	119,641	105,886	133,284	0.90	0.79	0.0000 *
PA	122,879	86,849	127,337	0.96	0.68	0.0002 *
SD	1,943,740	1,848,851	1,975,710	0.98	0.94	0.0029
TN	373,749	352,491	356,382	1.05	0.99	0.0000 *
TX	152,057	147,205	88,375	1.72	1.67	0.0000 *
VA	183,679	169,607	166,064	1.11	1.02	0.0304
WI	338,992	341,330	325,974	1.04	1.05	0.0083
Total	27,967,600	27,302,835	28,559,185	0.98	0.96	0.0000 *

* Significant at .05 level [.05/27=.00185=Bonferroni p]

1/ The expansions do not cover the entire population since some strata were excluded from the survey.

Appendix B: Statistical Test for the Difference in Responses (Paired T-test)

Let x = Initial Response
y = Final Response
d = x - y = Difference

H_O: $D = 0$
H_A: $D \neq 0$

$$\text{if } z > Z_{\frac{\alpha}{2}} \text{ then reject } H_O$$

$$z = \frac{\hat{d} - 0}{\sqrt{Var\ \hat{d}}}$$

$$\hat{d} = \sum_{h=1}^{N} N_h \bar{d}_h \qquad \bar{d}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} d_{hi}$$

$$Var(\hat{d}) = \sum_{h=1}^{L} N_h (N_h - n_h) \frac{s_h^2}{n_h} \qquad s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (d_{hi} - \bar{d}_h)^2$$

Appendix C: Formula for the Variance of the Difference between Ratios

The variance of the difference in ratios was calculated from the formula for a combined ratio estimate as described in Cochran (1977).

$$V(\hat{R}_d) = \frac{\sum_{h=1}^{L} \left(\frac{N_h(N_h-n_h)}{n_h(n_h-1)} \sum_{i=1}^{n_h} (u_{hi}-\bar{u}_h)^2 \right)}{\left(\sum_{h=1}^{L} N_h \bar{x}_h \right)^2}$$

where

$\hat{R}_d = (\hat{R} - \hat{R}') = Difference\ between\ Ratios$

$$\hat{R} = \frac{\hat{Y}}{\hat{X}} \qquad \hat{R}' = \frac{\hat{Y}'}{\hat{X}}$$

$$\hat{Y} = \sum_{h=1}^{L} N_h \bar{y}_h \qquad \bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$$

$$\hat{Y}' = \sum_{h=1}^{L} N_h \bar{y}'_h \qquad \bar{y}'_h = \frac{\sum_{i=1}^{n_h} y'_{hi}}{n_h}$$

$$\hat{X} = \sum_{h=1}^{L} N_h \bar{x}_h \qquad \bar{x}_h = \frac{\sum_{i=1}^{n_h} x_{hi}}{n_h}$$

$$u_{hi} = y_{hi} - \hat{R}_d x_{hi}$$

y_{hi} = initial response for the i^{th} observation in stratum h

y'_{hi} = final response for the i^{th} observation in stratum h

x_{hi} = June value for the i^{th} observation in stratum h