Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Development
Division

RDD Research Report
Number RDD-00-03

December 2000

Overview and Evaluation of AGGIES, an Automated Edit and Imputation System

Kara R. Perritt and Todd A. Todaro

OVERVIEW AND EVALUATION OF AGGIES, AN AUTOMATED EDIT AND IMPUTATION SYSTEM, by Kara R. Perritt and Todd A. Todaro, New Technology Research Section, Research and Development Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250-2000, March 2000, NASS Research Report, Report No. RDD-00-03.

ABSTRACT

The National Agricultural Statistics Service (NASS) currently uses several editing tools to ensure that data reported by respondents are consistent and complete. Such tools include manual editing, interactive micro level editing, batch micro level editing, and interactive macro level editing. By using all of these tools, even the most complex editing schemes can be managed and aggregate level impact can be evaluated. However, since these tools are not totally integrated, maintenance is costly, redundancy is apparent, and editing is not always performed in a consistent manner.

This paper discusses the continued evaluation of the AGricultural Generalized Imputation and Edit System (AGGIES) as a possible core tool in NASS's complete editing strategy. AGGIES is appealing in that it is an automated system that provides statistically consistent results in the edit and imputation process, it is written in a language, SAS, that makes for easy integration with tools currently being used, it can be applied to any number of surveys and censuses, and it minimizes the need for a complete manual review of the data at the micro level.

KEY WORDS

Accuracy indices; Complete edit strategy; Data editing; Error localization; Generalized automated edit and imputation system; Imputation estimators; Macro editing; Micro editing.

ACKNOWLEDGMENTS

The authors would like to thank Roberta Pense for continuing to support, encourage and guide the development and evaluation of the system; Arnie Wilcox for his immense help in the evaluation of the system; Brett Franklin and the California SSO, Christel Hoffman and the Colorado SSO, Carol Pittman and the Texas SSO, and Cindy Adamson and the Wyoming SSO for willingness to evaluate the system and provide feedback; Leonard Hart for modifying the Blaise instrument; and Anne Peterson for reviewing drafts of this report. Additional thanks to management, Carol House, Dale Atkinson, and George Hanuschak, for their continued support of this project.

TABLE OF CONTENTS

SUMMARY

Data collected from producers through surveys and censuses conducted by the National Agricultural Statistics Service (NASS) are summarized and published to provide information about the nation's agriculture. Prior to summarization, however, these data are edited using a variety of tools to ensure completeness and consistency. Currently, NASS uses Blaise for data collection and interactive editing, the Survey Processing System (SPS) for computer edit checks, the Interactive Data Analysis System (IDAS) for a macro level review of the data, and the Ag Complex edit (used exclusively for the Census of Agriculture) for edit and imputation (Pense, 1997; Todaro, 1999b).

Many other tools, not being used by NASS, exist and more are under development. In order to make certain that the agency's data processing procedures are as efficient as possible, yet still maintain data quality, all editing tools need to be considered. One tool under development is the AGricultural Generalized Imputation and Edit System (AGGIES) which is an automated edit and imputation system. The system offers the potential to improve the efficiency of the data processing procedures.

AGGIES is comprised of several modules to facilitate maintenance of the system. The modules are as follows: edit specification, check edits, edit summary, outlier detection, error localization and imputation. Each of these modules is thoroughly described in an earlier report by Todaro (1999a); however, since that publication, several enhancements have been made to the system. These modifications and updates include an updated data set selection screen allowing for selection of multiple data sets to be edited, increasing the number of variables allowed to be used in constructing an edit, adding a description option for each edit, redesigning the process for forming data and edit groups, increasing the output information for the check edits and edit summary reports, allowing for error localization and imputation only after a certain percentage of records have been edited, adding an interactive module to allow for on-screen updates with an automatic error check, and integrating IDAS and AGGIES to combine micro and macro level checks into one modular system.

Several evaluations of the system have been completed using data from various surveys. A preliminary study using hog data from Iowa's September 1996 Hog Report showed promising results (Todaro, 1999a) leading to more extensive evaluations.

Following that preliminary evaluation, data from California, Colorado, Texas and Wyoming's 1999 January Sheep Report were run through AGGIES using procedures that might be used in production: minimal hand editing was done on paper questionnaires and administrative pre-edit checks were completed by Blaise for list frame records in the non-extreme operator stratum (extreme operator and area frame records were subjected to the usual, complete interactive Blaise edit). After the data were run through AGGIES, its output file was compared to the clean data file from the production survey at an aggregate level. At this level, ten of the eighty total variables across all states exceeded a 5 percent tolerance. This outcome demonstrates that data edited and imputed by AGGIES resulted in a data set similar to that currently produced by NASS. To further study the results, Manzari and Della Rocca's (1999) accuracy indices were calculated. The indices evaluated the system's editing and imputation capability based on the number of detected, undetected and introduced errors. These

indices established that AGGIES performed well on January's data with only three variables having the overall editing and imputation accuracy index less than 90 percent, where 100 percent is maximum accuracy.

January's study led the way for a July sheep study that took place in the same four states immediately following the production summary. The evaluation procedures were similar to January's with some exceptions as follows: states were given specific editing guidelines for minimal editing, all data were run through AGGIES with edit and data groups assigned to separate extreme operators from other records, and after the AGGIES run, the records were reviewed by each respective state's sheep statistician in an interactive IDAS/AGGIES setting. As with January's data, the aggregate level statistics from the AGGIES and production outputs were compared and accuracy indices calculated. The results were comparable to January's results given the heavier use of minus ones (indicating missing data) and the availability of interactive IDAS editing. At the aggregate level, twelve of the forty-four total variables across all states exceeded a tolerance of 5 percent and two of these variables had an overall editing and imputation accuracy index under 90 percent.

Overall, evaluations showed that for commodity data editing, AGGIES generally did no worse than the current processing system. Using AGGIES provides the following potential benefits:

1) provides statistically consistent results
2) is written in an agency supported language, SAS, which simplifies integration with currently used tools
3) conserves resources in the development and maintenance of a single system
4) minimizes the need for a complete manual review at the micro level

However, there remain several issues to address when considering AGGIES as a potential editing tool. The following recommendations are made:

1) address functional issues of AGGIES based on feedback from the July project
2) evaluate AGGIES using the 1997 Census of Agriculture data
3) port AGGIES to the mainframe to evaluate computational power and speed
4) evaluate AGGIES on crop/stock data to give a more complete picture of the capabilities of the system

1. INTRODUCTION

The National Agricultural Statistics Service (NASS) is charged with collecting, summarizing and publishing information about agriculture in the United States. To accomplish this task, NASS uses a variety of surveys along with the Census of Agriculture to obtain information from producers. Once data are collected, they are edited to ensure their accuracy and completeness. Accurate data are important for making inferences about the underlying population characteristics, for improving the accuracy of estimates and for designing future surveys.

Over the past several decades, the development of editing tools has progressed in an effort to improve both the editing process and data quality. Examples of these tools include macro editing, selective editing and statistical editing. These different types of editing tools can be thought to have a complementary effect. For example, macro editing techniques can be used to selectively identify suspicious data having a large impact at the aggregate level. Then, by utilizing a drill down capability, corrective actions can be taken at the micro level. Finally, data having a minor impact on aggregate levels could be edited using a micro editing system to ensure consistency within the record. Ultimately, the goal is to come up with an appropriate mix of edit tools to form a complete edit strategy that will maintain data quality and at the same time, improve the efficiency of the edit and imputation process (De Jong, 1996).

AGGIES, the AGricultural Generalized Imputation and Edit System, is one such editing tool that is being developed. It is an automated edit and imputation system for use in editing data for completeness and consistency prior to summarization. When problems with data records are encountered by the system, it automatically determines which values to change and imputes for those values such that, after processing, all records are complete and consistent.

Currently, NASS uses the following editing tools: Blaise, the Survey Processing System (SPS), the Interactive Data Analysis System (IDAS), and the Ag Complex Edit (Pense, 1997; Todaro, 1999b). Blaise, SPS and IDAS are used for editing survey data but generally require editor intervention to correct data, i.e., machine imputation is seldom done. The Ag Complex Edit does allow for machine imputation but it can only be used for the Census of Agriculture. AGGIES, on the other hand, does both editing and imputation without intervention and can be used for surveys and, theoretically, for censuses. Also, since AGGIES and IDAS are written in SAS, the systems have been integrated and could form an enlarged core system that performs micro and macro editing functions.

Section 2 of this paper gives an overview of AGGIES that includes a discussion of the enhancements done since Todaro's last report on a preliminary evaluation of the system (1999a). After which, a section with results from current evaluations is presented and future evaluation plans are outlined.

2. AGGIES OVERVIEW

2.1 DESCRIPTION OF THE SYSTEM

Much of the methodology for AGGIES is based on the Generalized Edit and Imputation System (GEIS) developed at Statistics Canada (Cotton, 1993). AGGIES was written in the SAS programming language and, with its

object oriented features, allows the user to easily run the system and make selections using the mouse to point and click. It is designed to edit non-negative, continuous values and requires the edits to be of linear form (linear inequalities or linear equalities). The system comprises a number of modules, each performing a separate function. This section provides a general overview of the AGGIES modules as they currently exist. For a more thorough description, refer to Todaro (1999a) and Appendix 1 of this document.

2.1.1 EDIT SPECIFICATION

The edits are entered into the system in the edit specification module (Figure A3 in Appendix 1) as linear edits and specify pass conditions, i.e., records satisfying the edit pass the edit. Thus, collectively, the set of edits describe an acceptable region. A data record that lies within this acceptable region satisfies all edits simultaneously; otherwise one or more edits are violated. An edit is specified by typing in the edit identifier and coefficients, and selecting the variables and an inequality or equality sign from selection lists. A maximum of twenty variables can contribute to an edit. Error checking features ensure that all coefficients are numeric, variables are selected at most once, and all components forming the edit are entered.

An edit may be modified by selecting the 'Modify Edit' icon on the utility screen (Figure A4 in Appendix 1) which displays a list of edit identifiers corresponding to all edits that have been entered into the system. Upon the selection of an edit identifier, a screen similar to the edit specification screen appears with the edit information filled in for the corresponding edit. To delete an edit, the 'Delete Edit' icon on the utility screen is selected, followed by the selection of an edit identifier corresponding to the edit to be deleted.

Edit and data groups may be formed by selecting the 'Form Groups' icon on the utility screen. An edit group is a subset of edits that are applied to a collection of data records called a data group. Each edit group is created by selecting the edit identifiers corresponding to the edits forming the edit group. A data group is created by forming a SAS subsetting condition that describes the data records belonging to the data group. Any number of edit and data group pairs may be formed. AGGIES will process all of the groups in a single run.

2.1.2 CHECK EDITS

Selecting the 'Check Edits' icon on the utility screen checks for logical consistency of the entire edit set, redundant edits and hidden equality edits. An edit set is logically inconsistent if no data record can satisfy all edits simultaneously; otherwise it is logically consistent. A redundant edit is an edit that is implied by two or more other edits in the edit set. A hidden equality edit is an equality edit not contained in the edit set, but rather, implied by two or more inequality edits in the edit set. The output of this module displays a message if the edit set is logically inconsistent, identifies any edits that are redundant, lists any edits that imply a hidden equality, and shows the range of values for every variable involved in at least one edit. It is noted that simply deleting all redundant edits may result in a subset of edits that describe a different acceptable region than the acceptable region described by the originally specified set of edits. Thus, if there are any redundant edits, the edits should be examined

closely to identify a set of edits free of any redundant edits.

2.1.3 EDIT SUMMARY

Selecting the 'Edit Summary' icon on the utility screen displays summary information from applying the edits to the data records. Counts of the number of records satisfying all edits and failing at least one edit are displayed in the first section of the output. The second section of the output displays for each edit, including positivity edits (since the values are required to be non-negative), the number of records satisfying and failing each edit. The edit summary module also "manages" the data flow by requiring data groups to have a specified cumulative frequency before allowing error localization and imputation to be executed (see Appendix 3.3).

2.1.4 OUTLIER DETECTION

Outliers can be detected by selecting the 'Outlier Detection' icon on the utility screen. This module identifies univariate outliers utilizing the Hidiroglou-Berthelot method (Cotton, 1993) using current data. Since it has been observed that a large number of outliers may result, only those outlying records that are also involved in a failed edit are displayed.

2.1.5 ERROR LOCALIZATION

For those data records failing one or more edits, selecting the 'Error Localization' icon on the utility screen identifies the fewest values to change per record so that after imputation, all of the data records can satisfy all of the edits simultaneously. An option allows for the specification of variable reliability weights, with the default weights equal to one. If weights other than one are specified, then the fewest weighted values are changed per record rather than the fewest values. Thus, all things being equal, the higher the weight for a variable, the less likely the variable value will be changed. The methodology underlying this module is based on Chernikova's algorithm (Schiopu-Kratina and Kovar, 1989). The output of this module consists of two parts. The number of times each value was identified to be changed is displayed in the first part. The second part displays for each record having at least one value identified to be changed, the originally reported record followed by the error-localized record. The distinguishing feature of the error-localized record is the placement of the value '-1' for those values identified to be changed.

2.1.6 IMPUTATION

Prior to the imputation of values, several input options are available. The first allows for the selection of the order in which the variables are imputed for all of the data records. Second, since the data records are processed sequentially, previously imputed values may either be selected to be included or excluded when imputing for values in the current data record. Third, for each variable, the selection of up to six imputation estimators (see Appendix 5) and their order of application may be made. If more than one imputation estimator is selected for a particular variable, imputation is attempted using the estimators in the selected order. The value of the first imputation estimator that will result in the data record satisfying all edits is imputed. If no imputation estimator is selected for a particular variable, or if none of the selected imputation estimators will result in the record satisfying all edits, then the set of values that will result in the record satisfying all edits

simultaneously is calculated and the midpoint of this set is imputed. This default midpoint imputation method, borrowed from the Structured Programs for Economic Editing and Referrals (SPEER) system, guarantees that each data record will satisfy all edits after imputation (Todaro, 1997).

The imputation output consists of two parts. Imputation counts by imputation estimator by variable are displayed in the first part. The second part displays for each data record having at least one value imputed, the originally reported record followed by the imputed record.

2.2 PRINCIPAL ENHANCEMENTS

Several modifications were made to AGGIES, some of which were made to accommodate testing for the July 1999 Sheep Survey for the four states, California, Colorado, Texas and Wyoming. This section will summarize two principal enhancements: the addition of an interactive editing screen and the integration of AGGIES with IDAS. These two enhancements, and numerous other modifications, are described in detail in Appendix 1.

2.2.1 INTERACTIVE EDITING SCREEN

An interactive screen, shown in Figure 1, has been added to AGGIES that allows the batch edited and imputed values to be interactively updated. It is noted that this screen has been customized for the July 1999 Sheep Survey. A generalized interactive edit has not been developed for AGGIES.

This screen displays two forms. The form on the right-hand side displays the current AGGIES batch-edited data and can be interactively modified. The form on the left-hand side displays information that may be useful for editing the data, such as originally reported data or historical data. If, in the process of interactively editing the values on the right-hand side form, one of the edits is violated, those cells containing values that are involved in at least one failed edit are highlighted in yellow.

The radio box beneath the left-hand side form provides for the selection of three options. The first, and default option, displays the originally reported values in the left-hand side form. When the reported values are displayed and there are differences in the values of the variables between the two forms, the differing values are displayed in red which can expedite the interactive editing process. The second and third options display the previous January and July values, respectively, for the data record, if available. These data are provided to aid in interactively editing data records that look suspicious or in reviewing changes made by AGGIES that appear suspect.

Changes made to the right-hand side form may be submitted by clicking on the 'Update' push button located to the bottom right of the screen. A comment facility is available by clicking on the 'Comments' push button located to the left of the 'Update' push button. When clicked, a screen is displayed whereby comments may be entered regarding interactive changes made. These comments can be accessed later through the use of IDAS.

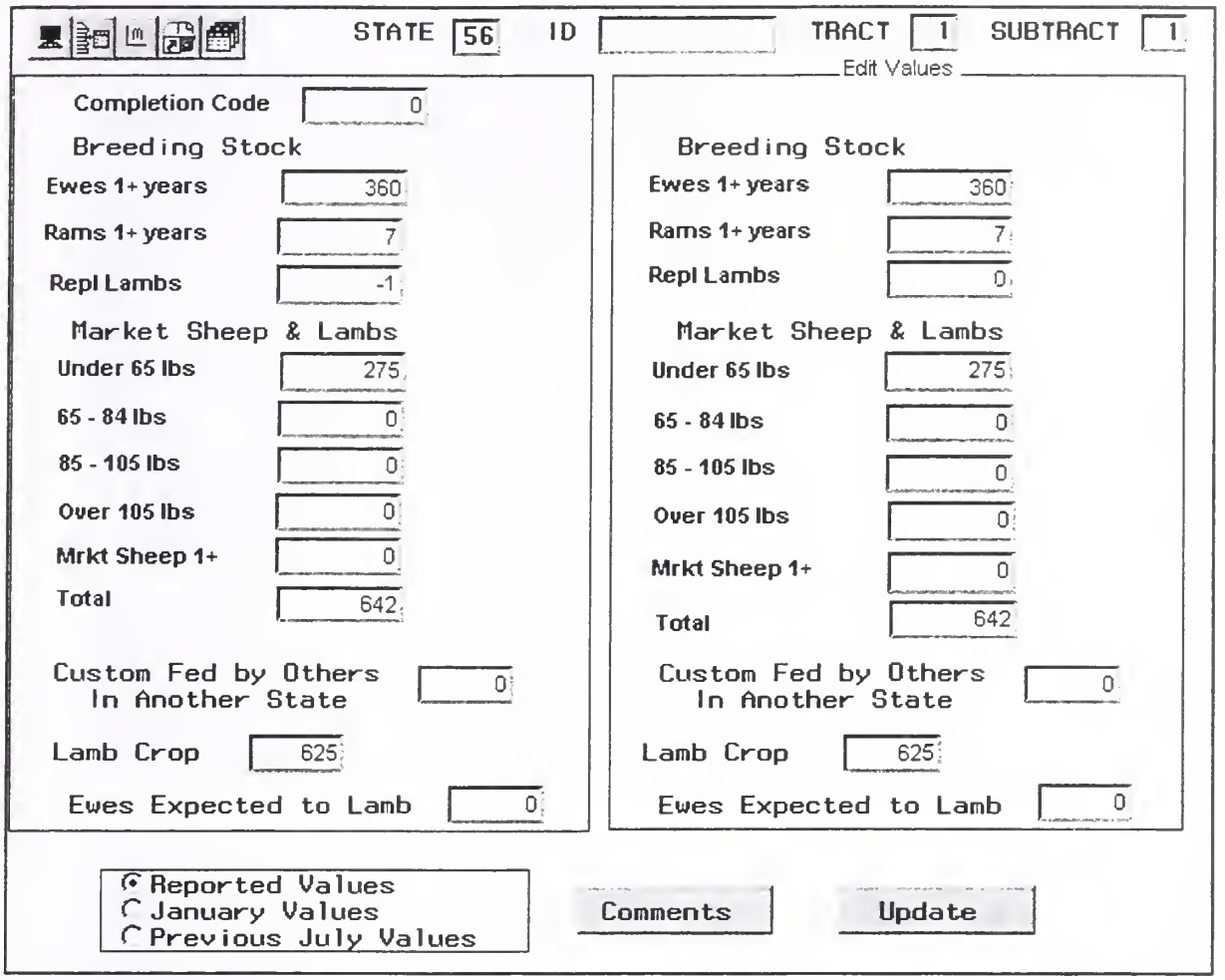


Figure 1. Interactive Screen

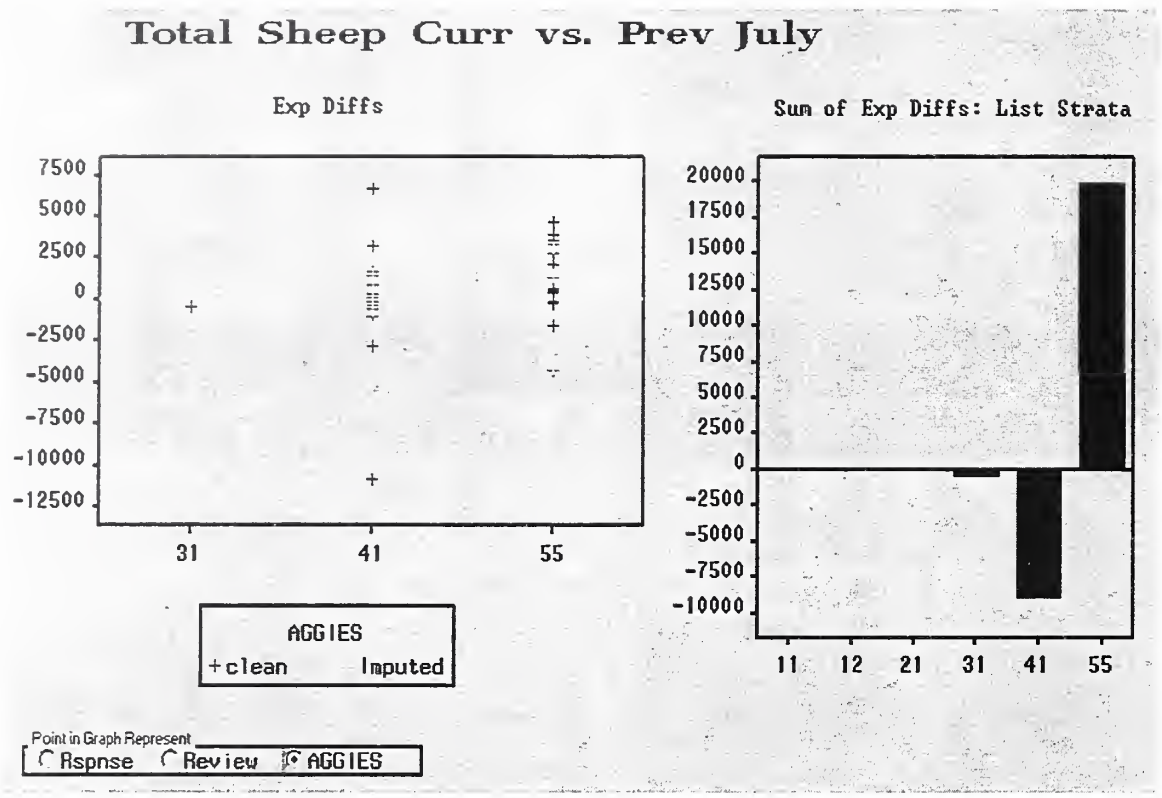


Figure 2. Scatter Plot

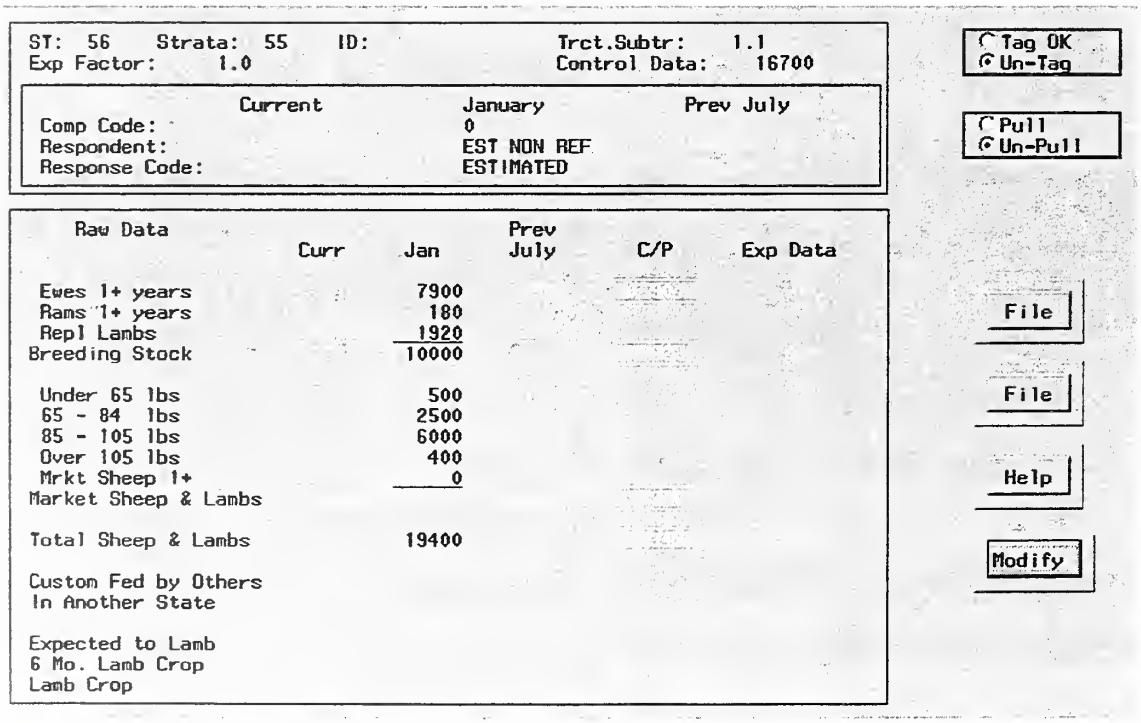


Figure 3. Drill Down Screen

2.2.2 MODIFICATIONS TO IDAS

For the July 1999 Sheep Survey, attempts were made to integrate AGGIES and IDAS. This integration occurred in two places in IDAS. The first distinguishes data records that passed all edits from those that AGGIES imputed due to one or more failed edits. Figure 2 displays a scatter plot by stratum obtained by selecting from the IDAS main menu – Daily Data Analysis, Analysis Tables, Curr vs. Prev, Total Sheep (or any other available selection). The clean values correspond to data records that passed all edits, while the imputed values failed at least one edit.

From the scatter plot screen, the drill down feature of IDAS is used to bring up the screen shown in Figure 3. It is on this screen that the second integration of AGGIES and IDAS took place. The top 'File' push button, on the far right side of the screen, was modified to provide access to the same comment file used in AGGIES. Also, the bottom push button, labeled 'Modify', was added to the screen to allow editors to modify data on a particular record. When 'Modify' is clicked, the interactive screen (Figure 1) appears and the editor can modify the data. The AGGIES edits would be interactively invoked. If interactive editing created no errors, the editor would then return to the IDAS screens and review other records. However, the IDAS set-up would need to be rerun to see the changes reflected in IDAS.

3. APPLICATION

Evaluation of AGGIES has been completed on the following three data sets: September 1996 Hog Report (Iowa), January 1999 Sheep Report (California, Colorado, Texas and Wyoming) and July 1999 Sheep Report (California, Colorado, Texas and Wyoming). Each of these studies will be discussed in turn. Following these discussions, results will be summarized and feedback from users presented.

3.1 SEPTEMBER 1996 HOG SURVEY - IA

The first data used to evaluate AGGIES were from the September 1996 Iowa Quarterly Hog Report survey. For this evaluation, aggregate statistics from AGGIES were compared with those from the current Blaise/SPS/IDAS editing system which was treated as "truth". The results, published by Todaro (1999a), were encouraging; however, since it was a one state, one survey study, a more complete evaluation of the system was needed.

3.2 JANUARY 1999 SHEEP SURVEY - CA, CO, TX AND WY

The next evaluation used the January 1999 Sheep Report survey data for California, Colorado, Texas and Wyoming. The following gives the basic evaluation procedures used and is succeeded by the results.

Prior to the survey period, several sources were used to establish the following input parameters: edits, reliability weights, imputation order and imputation estimators. The Sheep Editing and Analysis Team report (Anderson et al., 1998) and advice from sheep commodity experts were used to specify edits identical to the critical edits used during survey production. Reliability weights, imputation order and imputation estimators were mainly developed under the direction of sheep commodity experts. Once developed,

historical Sheep Report data were used to fine-tune all these input parameters. See Appendix 2.1 for the finalized input parameters. Next, the Blaise interactive edit (IE) was modified to allow for missing values (-1's) in every cell, to only flag administrative coding errors, and to calculate weighting adjustments. As the survey commenced, the four states involved were instructed to do minimal manual editing on paper questionnaires but to otherwise process the data as usual, i.e., use the Blaise/SPS/IDAS editing system. After the survey was completed, the post-Blaise but pre-SPS/IDAS data were made available to Research Division's staff and from there, were run through AGGIES. In other words, prior to the AGGIES run, the only edits done on these data were the administrative checks from the modified Blaise IE. Directly from the AGGIES output, with no statistician review, expanded aggregate statistics were compared to those calculated from the survey production data (also known in NASS as clean data or the D4 data) that had, during the live production, gone through the current, complete Blaise/SPS/IDAS editing process. Following this AGGIES to survey production comparison, data for the state with the largest percentage of records with errors, Wyoming, were run through AGGIES three times to assess repeatability of the results. Variability between runs can occur when the error localization module encounters multiple solutions. Finally, to complete the evaluation, editing and imputation accuracy indices were calculated for each state's data.

Before discussing any specific results, a few comments regarding procedures should be mentioned. First, no review of the AGGIES imputed data file was completed before the summary because the researchers would have had to make subjective decisions that may or may not concur with those made by state office statisticians. Second, since it often happens that one or two extremely dirty records take up the majority of error localization time, a 10-second time limit was imposed on the module. In other words, the computer was allotted 10 CPU seconds to error localize each individual record. This optimizes the process by allowing the computer to clean up the majority of records in a minimal amount of time. If the time limit was exceeded for a particular record, AGGIES stopped processing that record and went on to the next one. For this comparison, the data for any record that exceeded this 10-second limit were replaced with the survey production data since it was assumed that human intervention would have had to occur. Likewise, data from any record identified as an outlier with respect to the 'total sheep and lambs' variable were replaced with the survey production data. Third, area frame records and records classified in the extreme operator (EO) stratum were processed through the unmodified Blaise IE prior to going through AGGIES. Therefore, AGGIES did not edit or impute any of these records as all critical errors were updated during the Blaise IE. Fourth, due to data processing difficulties in the State offices, not all records were in the post-Blaise data files that were made available to Research Division's staff for this project. Data for these missing records were obtained directly from the survey production data. Finally, the Colorado office had different operating procedures which caused the effect of AGGIES to be masked. Thus, results from Colorado only appear in the Appendices.

Table 1. Overview of Evaluation Results

State	Total Records	AGGIES Edited Records[1/]	Records with Errors (% of Edited)	Outlying Records[2/]	Error Localization Time[3/] (min)	Records Exceeding 10-Second Limit	Variables[4/] Exceeding 5% Difference[5/] (% of 20)
CA	555	370	69 (19%)	2	2	0	5 (25%)
TX	2,400	1,540	99 (6%)	3	4	2	3 (15%)
WY[6/]	906	749	184 (25%)	0, 0, 0	12, 12, 12	15, 0, 0	2, 2, 2 (10%)

1/ Excludes area frame, extreme operator, and missing records
2/ Number of outlying records based on total sheep
3/ 400-MHz Pentium computer
4/ Out of 20 variables common to all states
5/ Absolute percent difference between expanded AGGIES output data and expanded survey production data
6/ All three runs listed in run order

3.2.1 EDIT AND IMPUTATION COUNTS

Table 1 displays an overview of the evaluation results for California, Texas, and Wyoming. Included are the total number of records processed, the number of records that AGGIES actually edited (excludes missing, area frame and extreme operator records), the number of records failing at least one AGGIES edit, the number of records identified as outliers, the total time required for error localization, the number of records exceeding the imposed 10-second time limit, and the number of variables, out of 20 variables common to all states, with an absolute percent difference greater than five. This percent difference compared the expanded AGGIES output data to the expanded survey production data.

This table shows that although Texas had the largest total number of records (1,540) edited by AGGIES, Wyoming had the most records failing at least one edit (184). This likely accounts for the higher error localization time required for the Wyoming data (12 minutes). Also note that, due to heavy, local area network (LAN) traffic, the first Wyoming run had 15 time limit exceeded records, while all other runs had none. Finally, the last column shows that most of the variables at the expanded aggregate level fell within five percent of the current procedure's expanded aggregate with California having the largest number (5) outside that range.

The next table, Table 2, lists the percent of valid zeros and the count of missing values, by state, for each variable. This indicates that data as reported by the respondents were sparse yet fairly complete, i.e., in the reported data, there were many valid zeros and relatively few missing values (-1's). The table is sorted by variable in the order that the variables appear in the questionnaire (see Appendix 2.4 for a condensed copy of the questionnaire).

From the table note that Wyoming used minus ones (missing data) for more variables and generally at a greater frequency by variable than the other states, especially for wool production and wool price. Because AGGIES has non-negativity constraints, these missing values are errors and are the reason for the higher error rate for Wyoming seen in Table 1.

Table 2. Percent of Valid Zeros and Count of Missing Values in the Reported Data, By State

Variable	Percent of Valid Zeros			Count of Missing Values (-1's)		
	CA (555)	TX (2400)	WY (906)	CA (555)	TX (2400)	WY (906)
Ewes for Breeding	52	68	37	0	0	1
Rams for Breeding	54	70	45	0	0	1
Replacement Lambs for Breeding	70	83	54	0	2	1
Market Lambs Under 65 lbs.	71	89	97	0	1	2
Market Lambs 65 to 84 lbs.	93	95	95	1	1	2
Market Lambs 85 to 105 lbs.	93	97	91	1	3	2
Market Lambs Over 105 lbs.	94	99	94	0	1	2
Market Sheep	95	98	96	0	2	1
Total Sheep and Lambs	48	66	35	0	0	1
Out of State Sheep	99	NA	100	0	NA	0
Lamb Crop	55	70	33	4	10	46
Breeding Animals Shorn	56	70	34	2	6	42
Wool from Breeding Animals	58	71	35	31	25	118
Market Animals Shorn	85	94	93	7	10	20
Wool from Market Animals	87	95	94	10	7	9
Average Wool Price	63	79	42	53	73	251
Average Ewe Value	58	72	47	31	34	62
Average Ram Value	60	74	54	30	37	38
Average Replacement Lamb Value	72	84	62	21	30	44
Average Market Lamb Value	73	88	85	20	29	24
Average Market Sheep Value	92	97	96	6	4	5

Number in parenthesis ( ) is the total number of records

3.2.2 COMPARISON OF EXPANDED DATA

Wyoming's data were run through AGGIES three times to assess the variability between runs that can exist when error localization encounters multiple solutions. Table 3 indicates the variability at the expanded level between these three runs by displaying the AGGIES expanded total for each run and the standard deviation between the runs. The table is sorted by the standard deviation in descending order.

Notice that, for the first four variables listed, run 2 and run 3 have identical expanded totals so the only contribution to variability for these variables is from run 1. The fifteen records that exceeded the error localization time limit in the first run caused its results to differ from the results of the two other runs. Reported data from these fifteen reports were available in runs 2 and 3, but not in run 1, when the system calculated the imputation estimators used to impute missing data. Thus, the imputation estimators, and the number of records contributing to these estimators, for

Table 3. Variability in the Expanded Total for the Three Wyoming Runs

Variable	AGGIES Expanded Total			Standard Deviation
	Run 1	Run 2	Run 3	
Wool from Breeding Animals	4,198,744	4,201,843	4,201,843	1,789
Wool from Market Animals	493,216	494,612	494,612	806
Breeding Animals Shorn	446,837	448,106	448,106	733
Market Animals Shorn	121,925	122,273	122,273	201
Total Sheep and Lambs	574,431	574,415	574,244	104
Market Lambs 65 to 84 lbs.	16,505	16,502	16,331	100
Market Lambs 85 to 105 lbs.	73,393	73,381	73,381	7
Lamb Crop	394,351	394,346	394,346	3
Rams for Breeding	12,215	12,214	12,214	1
Ewes for Breeding	363,222	363,222	363,222	0
Replacement Lambs for Breeding	74,480	74,480	74,480	0
Market Lambs Under 65 lbs.	2,578	2,578	2,578	0
Market Lambs Over 105 lbs.	29,419	29,419	29,419	0
Market Sheep	2,618	2,618	2,618	0
Out of State Sheep	3,929	3,929	3,929	0
Average Wool Price	0.76	0.76	0.76	0
Average Ewe Value	88	88	88	0
Average Ram Value	283	283	283	0
Average Replacement Lamb Value	79	79	79	0
Average Market Lamb Value	70	70	70	0
Average Market Sheep Value	38	38	38	0

runs 2 and 3 were identical but were different from run 1. Increasing the time limit or processing on a faster computer would likely rectify this situation involving time limit exceeded records. The remaining run-to-run variation, due to multiple solutions, is negligible.

Table 4 shows for each survey variable by state, the percent difference between the two expanded totals, i.e., one calculated from the AGGIES output data and the other calculated from the survey production data (see Appendix 2.2 for expanded totals), and the count of imputations done by AGGIES. The table is sorted by variable in the order of appearance in the questionnaire.

Note that a difference between AGGIES and survey production expanded data can exist even without any AGGIES imputations. California's 'market sheep' variable, 36.72% difference and no AGGIES imputations, is an example. Expanded totals are different because during production, the reported 'market sheep' value for one record was updated but in AGGIES, its reported value was not imputed. Other scenarios exist, but this is the most common cause for expanded differences without AGGIES imputations.

Table 4. Percent Difference and Number of AGGIES Imputations, by State

Variable	Percent Difference[1/]			Number of AGGIES Imputations		
	CA	TX	WY[2/]	CA	TX	WY[3/]
Ewes for Breeding	-1.49	-0.49	0.59	0	0	0,0,0
Rams for Breeding	0.41	-0.02	0.50	0	0	1,1,1
Replacement Lambs for Breeding	-0.03	-0.04	0.41	0	2	1,1,1
Market Lambs Under 65 lbs.	-0.21	0.00	0.00	0	1	1,1,1
Market Lambs 65 to 84 lbs.	1.41	0.28	0.56	3	5	2,4,3
Market Lambs 85 to 105 lbs.	0.47	5.16	0.01	2	4	2,1,1
Market Lambs Over 105 lbs.	0.00	0.00	0.03	0	1	1,1,1
Market Sheep	36.72	17.38	50.55	0	2	0,0,0
Total Sheep and Lambs	0.34	0.02	0.61	10	18	9,8,9
Out of State Sheep	0.00	NA	0.00	0	NA	0,0,0
Lamb Crop	-1.22	-1.00	0.23	4	10	40,40,40
Breeding Animals Shorn	-2.54	-0.99	0.93	2	5	37,37,37
Wool from Breeding Animals	-2.52	-0.75	-0.58	40	48	130,130,130
Market Animals Shorn	-18.20	0.55	0.29	7	9	18,18,18
Wool from Market Animals	-22.07	-3.76	0.11	20	26	24,24,24
Average Wool Price	6.67	0.00	-1.30	0	19	75,84,84
Average Ewe Value	-1.02	0.00	0.00	1	2	4,4,4
Average Ram Value	0.82	19.02	0.71	0	0	1,1,1
Average Replacement Lamb Value	15.38	2.98	0.00	1	0	3,3,3
Average Market Lamb Value	1.39	0.00	0.00	0	0	0,0,0
Average Market Sheep Value	0.00	0.00	-33.33	0	0	0,0,0

1/ Between expanded AGGIES output data and expanded survey production data
2/ Average expanded totals of the three AGGIES runs used in calculating percent difference
3/ All three runs listed in run order

For Table 4, any absolute percent difference greater than five percent was analyzed. In five of the ten cases (California. Texas and Wyoming's 'market sheep', Texas's 'market lambs 85 to 105 lbs.' and Wyoming's 'average market sheep value'), the difference was due to a single report where AGGIES changed one variable but during production a different variable was changed. The only way to exactly duplicate what was done during production for these reports would be to lose one of the main attractive features of AGGIES - generality. A more effective approach, not affecting generality, would be to review the AGGIES imputed file using IDAS. Two other cases with an absolute percent difference greater than five percent were variables 'wool from market animals' and 'market animals shorn' for California. The explanation for the difference between AGGIES totals and the survey production totals involves several records. For each record, the reported data value for both these variables was empty. Since this is not an error in AGGIES, it did not update either variable; whereas, during production, both variables were updated with positive values. Doing a comparison between the current and the previously reported wool

production data in IDAS may indicate when to edit in missing wool production data. The last three cases with an absolute percent difference greater than five percent were the following average value variables: California's 'average replacement lamb value' and 'average wool price' and Texas's 'average ram value'. Several records attributed to the difference between AGGIES and the production survey for each of these variables with no single explanation. Identifying current survey editing and/or imputation procedures and using the IDAS editing tool are two approaches that may lead to remedying the inconsistencies between the AGGIES values and the current survey values for these three variables.

3.2.3 EDITING AND IMPUTATION ACCURACY INDICES

To complete this evaluation of AGGIES, Manzari and Della Rocca's (1999) accuracy indices were calculated for each variable by state. These indices evaluate the quality of the AGGIES editing and imputation procedures based on the number of detected, undetected and introduced errors. All indices range from 0% (no accuracy) to 100% (maximum accuracy). The indices are divided into three groups of three: the first three indices assess the quality of editing, the next three assess the quality of imputation and the final three assess the overall quality of both editing and imputation. Table 5 describes each index, grouped by the quality it assesses (see Appendix 4 for formulas and further details regarding these indices).

For any quality index that assesses imputation, I4 through I9, a value was classified as correctly imputed if the AGGIES imputed value was exclusively within a certain percent of the survey production value. To arrive at this threshold, the coefficient of variation (CV) for each variable from the production survey summary was reviewed. The AGGIES imputed values would only be required to be as precise as the CV's indicated. The review lead to a five percent cut-off value for all

Table 5. Accuracy Indices (Manzari and Della Rocca, 1999)

Assessing ...	Index
Editing Quality	I1: fraction of unmodified data correctly handled
	I2: fraction of modified data correctly handled
	I3: fraction of total data correctly handled
Imputation Quality	I4: fraction of changed, unmodified data whose value is correctly imputed
	I5: fraction of changed, modified data whose value is correctly imputed
	I6: fraction of changed total data whose value is correctly imputed
Overall Editing and Imputation Quality	I7: fraction of unmodified data whose value is correctly imputed
	I8: fraction of modified data whose value is correctly imputed
	I9: fraction of total data whose value is correctly imputed

Where:

modified = survey production data that does not equal the reported data
unmodified = survey production data that equals the reported data
changed = AGGIES output data that does not equal the reported data
not changed = AGGIES output data that equals the reported data

variables for all states. That is, an imputed value was classified as correctly imputed if it was exclusively within five percent of the survey production value.

Indices I1 through I3 are indicators of the editing quality. Specifically, index I1 indicates whether or not the system introduces new errors in the data. The I2 index measures the ability of the system to detect errors in the data. Finally, index I3 gives an indication of overall performance of the error localization algorithm, which is the editing algorithm. Table 6 displays the overall editing accuracy index I3 for each state (see Appendix 2.3 for a listing of all indices). The table is sorted by variable in the order that the variables appear in the questionnaire.

Note that all variables have a high I3 value indicating that, overall, the error localization algorithm was able to detect the same errors in the data as the current survey production system and it introduced few new errors.

Table 6. Editing Accuracy Index I3 for Each State

Variable	I3		
	CA	TX	WY[1/]
Ewes for Breeding	99	100	99
Rams for Breeding	100	100	99
Replacement Lambs for Breeding	100	100	100
Market Lambs Under 65 lbs.	99	100	100
Market Lambs 65 to 84 lbs.	100	100	100
Market Lambs 85 to 105 lbs.	100	100	100
Market Lambs Over 105 lbs.	100	100	100
Market Sheep	100	100	100
Total Sheep and Lambs	99	100	99
Out of State Sheep	100	NA	100
Lamb Crop	99	100	100
Breeding Animals Shorn	99	100	99
Wool from Breeding Animals	99	100	99
Market Animals Shorn	99	100	100
Wool from Market Animals	98	100	99
Average Wool Price	99	100	100
Average Ewe Value	99	100	100
Average Ram Value	99	99	100
Average Replacement Lamb Value	100	100	99
Average Market Lamb Value	91	99	99
Average Market Sheep Value	100	100	99

1/ Averaged over the three AGGIES runs

Indices I4 through I6 are indicators of the imputation quality. The lack of unmodified data changed by AGGIES somewhat diminishes the ability of indices I4 and I6 to assess imputation quality. However, index I5, which measures the effectiveness of AGGIES to impute values within five percent of the modified survey production values, can give an indication of how well the system imputes. Table 7 displays the index I5 for each state. The table is sorted by variable in the order of questionnaire appearance.

The two variables that consistently had low I5 values across all states were the variables 'wool from market animals' and 'wool from breeding animals'. Current survey imputation procedures were compared to the AGGIES imputation procedures in order to analyze these low I5 values. An unpublished document by the Sheep Editing and Analysis Team (Anderson et al., 1998) noted that historically statisticians used the average fleece weight, which varies by state, to impute missing wool production. By doing so, they noted that the natural distribution of reported data was lost. The team recommended that for January 1999 (the survey used for this project), a 3-year state average fleece weight be used instead. However, they did caution

Table 7. Imputation Accuracy Index I5 for Each State

Variable	I5[2/]		
	California	Texas	Wyoming[1/]
Ewes for Breeding	100	100	100
Rams for Breeding	100	100	87
Replacement Lambs for Breeding	100	100	100
Market Lambs Under 65 lbs.	100	100	100
Market Lambs 65 to 84 lbs.	100	100	81
Market Lambs 85 to 105 lbs.	100	100	100
Market Lambs Over 105 lbs.	-	100	100
Market Sheep	100	100	100
Total Sheep and Lambs	91	100	91
Out of State Sheep	-	NA	-
Lamb Crop	67	45	87
Breeding Animals Shorn	60	57	53
Wool from Breeding Animals	50	27	16
Market Animals Shorn	22	80	47
Wool from Market Animals	31	11	13
Average Wool Price	96	79	67
Average Ewe Value	100	95	94
Average Ram Value	100	100	98
Average Replacement Lamb Value	97	100	95
Average Market Lamb Value	100	100	100
Average Market Sheep Value	100	100	100

1/ Averaged over the three AGGIES runs

2/ A dash (-) indicates the index could not be computed because calculations would have resulted in division by zero

that this 3-year average may be adversely affected by the imputation done in previous years. Also, the 3-year average does not take into account existing industry practices. AGGIES, in contrast to using constant averages for its imputation, used current ratios and auxiliary trends to estimate wool production. This may explain the inconsistency between the AGGIES imputed values and the survey production values for these variables.

Indices I7 through I9 indicate the quality of both editing and imputation. Indices I7 and I8 assess the quality for unmodified and modified data, respectively, while I9 evaluates overall editing and imputation quality. Table 8 displays the overall editing and imputation accuracy index I9 for each state. The table is sorted by variable in the order that the variables appear in the questionnaire.

Only one variable had an I9 value under 90%: Wyoming's 'wool from breeding animals'. As alluded to previously in the I5 index discussion, this low index value is due to poor imputation accuracy. On the whole, however, the system was proficient in treating the data as compared to the current survey production system.

Table 8. Editing and Imputation Accuracy Index I9 for Each State

Variable	I9		
	California	Texas	Wyoming[1/]
Ewes for Breeding	99	100	99
Rams for Breeding	100	100	99
Replacement Lambs for Breeding	100	100	100
Market Lambs Under 65 lbs.	99	100	100
Market Lambs 65 to 84 lbs.	100	100	100
Market Lambs 85 to 105 lbs.	100	100	100
Market Lambs Over 105 lbs.	100	100	100
Market Sheep	100	100	100
Total Sheep and Lambs	99	100	99
Out of State Sheep	100	NA	100
Lamb Crop	99	99	99
Breeding Animals Shorn	98	100	97
Wool from Breeding Animals	94	98	86
Market Animals Shorn	98	100	99
Wool from Market Animals	96	99	98
Average Wool Price	98	99	91
Average Ewe Value	99	100	99
Average Ram Value	99	99	99
Average Replacement Lamb Value	99	100	99
Average Market Lamb Value	91	99	99
Average Market Sheep Value	100	100	99

1/ Averaged over the three AGGIES runs

3.2.4 ROBUSTNESS OF WEIGHTS ON ERROR LOCALIZATION

In order to evaluate the effect that the reliability weights (see Appendix 2.1) had on the error localization algorithm, Wyoming's weights were changed and their data were run three additional times through AGGIES, i.e., a total of six runs were completed on Wyoming's data. For each of these additional runs, every variable was assigned the default reliability weight of one. The expanded totals from these three runs were compared to both the results from the previous three runs and the survey production expanded totals. This comparison showed that the expanded totals from the three runs using the default weights were virtually identical to the expanded totals from the original three runs. This is not necessarily surprising since there were few edits and there were very few errors in the data aside from those errors due to missing data. AGGIES, therefore, identified few variables to be imputed during error localization, limiting the need for reliability weights for these data.

3.2.5 EVALUATION OF PREDATOR LOSS DATA

The predator loss evaluation for Colorado, Texas and Wyoming had to be completed through separate AGGIES runs because the number of variables contained in that section (57 variables), coupled with the heavy use of minus ones (missing data), caused more time limit exceeded records in error localization than was acceptable. Dividing the variables into three mutually exclusive groups and running each group separately through AGGIES greatly reduced the number of time limit exceeded records. Two other factors complicated the evaluation of this section. First, statisticians were allowed to edit in a minus one for the total and leave the breakdown, the parts of the sum, blank. For these records, AGGIES always imputed in a zero for the minus one since the parts of the sum were all zero. This caused the AGGIES aggregate for each variable to be under-expanded. Second, it was acceptable to have minus ones for every breakdown part and have a positive entry in the total. AGGIES usually imputed a positive value for each breakdown variable when in actuality, the data are very sparse and most variables should have been imputed with a zero. This caused the AGGIES aggregate for several variables to be over-expanded. The confounding of these two factors greatly hindered the analysis of this section and conclusive evidence cannot be established. In order to study this section in the future, clear editing guidelines and specifications for the statisticians and Blaise IE need to be instituted.

3.2.6 SUMMARY OF JANUARY 1999 SHEEP PROJECT

When viewed on the whole, the indices and expanded total comparison indicate that AGGIES was proficient in treating the January 1999 sheep data. The imputation performed on some of the variables could perhaps be improved but, overall, the results were very promising and led to the more comprehensive project done in July 1999.

3.3 JULY 1999 SHEEP SURVEY - CA, CO, TX AND WY

The AGGIES evaluation using the July 1999 Sheep Report data was completed for the same states involved in the January 1999 project: California, Colorado, Texas and Wyoming. The July evaluation took place

immediately after the completion of the operational processing of the survey and used the same general procedures as the January evaluation. A brief overview of July's procedures follows.

In an effort to improve the system's imputation, a teleconference with the four states involved was conducted to get input on the AGGIES parameters. It was decided that since July's questionnaire is basically a shorter version of the January questionnaire (see Appendices 2.4 and 3.6), the edits, i.e., the critical edits used during operational processing, were to remain unchanged from January except that edits not applicable would be deleted. A consensus from the teleconference was reached to arrive at the following parameters: reliability weights, imputation order and imputation estimators (see Appendix 3.1).

Just prior to data collection, the states were given some very specific editing guidelines (see Appendix 3.2) for minimally editing questionnaires. These guidelines encouraged the use of minus ones for individual missing values. Sectional completion codes were used to calculate summary weights to account for complete non-response. However, since non-response completion codes are not allowed on EO (extreme operator) reports, minus ones were coded into every individual cell for all non-responding EO's.

As the survey commenced, the data were processed two ways: one for the operational results and one for the AGGIES evaluation. The following is a brief description of these processes (refer to the flowchart in Appendix 3.3 for more details).

After data collection, both by paper questionnaires and computer-aided telephone interviews, all data, including EO's, were processed through the modified Blaise IE which, as before, accepted minus one as valid for any cell. States then read the data out of Blaise and processed it as usual, i.e., through the SPS edit, IDAS review and SAS summary. Immediately following this production summary, NASS Headquarter staff traveled to each of the four states to demonstrate how data might be processed through AGGIES. Data read out of Blaise, which will be referred to as the reported data since only administrative edit checks had been completed, were run through AGGIES in batches. During these AGGIES runs, the following two data/edit groups were defined based on the non-EO and sampled EO definitions in the Survey Administration Manual for Agricultural Surveys, AgSAM (1999): strata less than 38 and strata greater than or equal to 38. All edits were applied to both groups; however, imputation was done within groups. After AGGIES imputation, records outlying with respect to the 'total sheep and lambs' variable and records that exceeded an imposed 10-second (CPU) error localization time limit were interactively reviewed and updated as needed in the interactive AGGIES module. Following this review, the state office statistician in charge of the sheep survey examined the data at the macro level using IDAS. Updates to the micro level data were made as necessary in the IDAS/AGGIES interactive module. After a thorough IDAS review, the data were run through the same SAS summary used on the production data and the results were compared to the production run. As before, editing and imputation accuracy indices were calculated for each state's data to complete the evaluation.

The major procedural changes from January include the following: heavy use of minus ones for missing data as directed by the specific editing guidelines, no estimation for non-response in the EO stratum by statisticians during manual editing, i.e., all EO's were run through AGGIES for edit and imputation, formation of data/edit groups in AGGIES and use of IDAS to review AGGIES output. Comparing the results from July back to those from January seems natural; however, the procedural changes above, and the fact that July's sample size was an average 80% less than January's, makes such a comparison less than ideal.

3.3.1 EDIT AND IMPUTATION COUNTS

Table 9 presents an overview of the results from the July data. Again, because Colorado's operating procedures masked the effects of AGGIES, results from that state will only appear in the Appendices. Specifically, Table 9 shows total number of records processed, the number of records failing at least one AGGIES edit, the number of records in each of the two data/edit groups, the number of records identified as outliers, the number of records exceeding the imposed 10-second time limit, and the number of variables, out of eleven variables common to all states, with an absolute percent difference greater than five.

This table shows that Wyoming had the largest percent of records failing at least one edit (24%). The total number of records is roughly split equally between the two data/edit groups for each state except for Wyoming. Few outlying records were encountered and no records exceeded the time limit. The last column shows that for California and Wyoming almost half of the variables fell outside the five percent variance range. Each of these variables and the two from Texas were analyzed and will be discussed later in this report.

The next table, Table 10, displays the percent of valid zeros and the count of missing values (-1's) for each variable to give an indication of how sparse and incomplete the data were as reported by the respondent. The table is sorted by variable in the order that the variables appear in the questionnaire.

Table 10 shows that for comparable variables, July's data are as sparse as January's data with similar valid zero percentages. However, the number of missing values increased for comparable variables, despite the decreases in

Table 9. Overview of Evaluation Results

State	Total Records	Records with Errors (% of Total)	Non-EO (Strata<38)	Sampled EO (Strata≥38)	Outlying Records[1/]	Records Exceeding 10-Second Limit	Variables[2/] Exceeding 5% Difference[3/] (% of 11)
CA	141	16 (11%)	68	73	0	0	5 (45%)
TX	521	32 (6%)	285	236	1	0	2 (18%)
WY	139	33 (24%)	47	92	1	0	5 (45%)

1/ Number of outlying records based on total sheep
2/ Out of 11 variables common to all states
3/ Absolute percent difference between expanded AGGIES output data and expanded survey production data

Table 10. Percent of Valid Zeros and Count of Missing Values in the Reported Data, By State

Variable	Percent of Valid Zeros			Count of Missing Values (-1's)		
	CA (141)	TX (521)	WY (139)	CA (141)	TX (521)	WY (139)
Ewes for Breeding	39	55	21	11	16	17
Rams for Breeding	42	57	24	11	15	17
Replacement Lambs for Breeding	58	72	40	11	19	21
Market Lambs Under 65 lbs.	73	70	29	11	20	20
Market Lambs 65 to 84 lbs.	70	88	80	12	15	13
Market Lambs 85 to 105 lbs.	71	97	90	11	7	2
Market Lambs Over 105 lbs.	84	97	94	12	6	3
Market Sheep	89	96	89	11	4	2
Total Sheep and Lambs	38	53	19	12	19	20
Out of State Sheep	89	NA	96	10	NA	0
Lamb Crop	43	57	23	15	20	22
Ewes Expected to Lamb	52	83	91	13	17	3

Number in parenthesis ( ) is the total number of records

sample sizes. The increase in missing data is largely due to the EO stratum. In January, the statisticians were required to estimate for EO non-response; in July, AGGIES did most of the imputation for non-response.

3.3.2 COMPARISON OF EXPANDED DATA

Table 11 displays for each survey variable, the percent difference between the AGGIES output expanded total and the survey production survey expanded total (see Appendix 3.4 for specific expanded totals). Also, the number of imputations done by AGGIES is shown for each variable. The table is sorted by variable by the order of appearance in the questionnaire.

All thirteen of the cases with an absolute percent difference greater than five percent were analyzed. In three cases, California's 'out of state sheep' and 'lamb crop' and Wyoming's 'ewes expected to lamb', the difference was mainly due to a single report. Because IDAS does not have graphics devoted specifically to these variables, even the most thorough review would not have discovered a problem with these reports. A possible solution is to develop new IDAS graphics for these particular variables. The nine weight-group breakdown variables that were out of the five percent range had anywhere from one to five records causing the discrepancy between the AGGIES and the survey production value. Every one of these reports was classified as a sampled EO (strata 55 and 41) and all the originally reported data values were missing (-1). This demonstrates the need for an extensive review of AGGIES imputations on non-response records. Another option would be to improve

Table 11. Percent Difference and Number of AGGIES Imputations, by State

Variable	Percent Difference[1/]			Number of AGGIES Imputations		
	CA	TX	WY	CA	TX	WY
Ewes for Breeding	-4.84	0.28	2.85	11	16	17
Rams for Breeding	-3.43	0.97	1.40	11	15	17
Replacement Lambs for Breeding	-2.45	1.14	12.21	11	19	21
Market Lambs Under 65 lbs.	3.02	-1.40	-9.24	11	22	21
Market Lambs 65 to 84 lbs.	-6.59	2.29	-9.84	12	15	13
Market Lambs 85 to 105 lbs.	-14.39	-2.23	42.16	11	7	2
Market Lambs Over 105 lbs.	24.56	-11.95	-0.25	12	6	3
Market Sheep	17.48	3.61	-1.13	11	4	2
Total Sheep and Lambs	-3.84	-0.03	-0.42	12	21	27
Out of State Sheep	-22.87	NA	-0.61	10	NA	0
Lamb Crop	-6.45	3.14	0.31	15	20	22
Ewes Expected to Lamb	0.90	9.92	83.31	14	18	3

1/ Between expanded AGGIES output data and expanded survey production data

imputation for these variables by using alternate imputation schemes, i.e., different imputation order and/or imputation estimator combinations. Also, imputation improvement may be found by investigating additional imputation estimators, such as the raking ratio estimator for variables, such as these, that are constrained by a balance edit. Texas's 'ewes expected to lamb' is the last variable out of the five percent range. Four reports were the main contributors to the difference. In all four of these reports, the reported value was missing (-1). As a result, AGGIES imputed a positive number; however, in the survey production data, the value was set to zero. The placement of a minus one for this variable needs careful consideration in order to avoid a positive AGGIES imputation occurring when a zero is wanted, i.e., only put a minus one in this cell when a positive number is definitely wanted. Also, as mentioned above, the lack of an IDAS graphic for this variable reduced the chance of correcting the AGGIES imputation.

3.3.3 EDITING AND IMPUTATION ACCURACY INDICES

Editing and imputation accuracy indices were calculated for July's data in a similar manner to calculations done for January's data.

The editing and imputation accuracy indices will be presented as they were previously for the January data. Indices I3, I5 and I9 follow in respective tables with values ranging from 0% (no accuracy) to 100% (maximum accuracy). All tables are sorted by variable in the order that the variables appear in the questionnaire. For a complete listing of all indices, I1 through I9, see Appendix 3.5.

The high values shown in Table 12 for index I3, which gives an indication of overall editing accuracy, suggests that AGGIES performed as well in editing these data as the operational processing system. In general, the editing algorithm was able to detect errors and very few new errors were introduced.

Table 13 shows the I5 index along with the threshold percent (determined from the CV's) used to classify a value as correctly imputed. Recall, an imputed value was classified as correctly imputed if it was exclusively within the threshold percent of the survey production value.

To reiterate a point already made, comparing these results to the results from January can be misleading. Namely, the heavier use of minus ones led to an increase in imputation. This increase in imputation combined with the decrease in sample size causes a seeming decline in imputation performance. Another factor affecting July's I5 index for all variables is that the EO's were run through AGGIES. The I5 index compares these AGGIES imputed values to the survey production values. In the survey production data, the non-responding EO's were estimated, generally using data from the previous survey.

In many cases, these historical data, often themselves estimated, were simply pulled forward. Therefore, the accepted practice of treating the survey production data as the "truth", as the I5 index calculation does, may not be totally reasonable. Since the EO's were not run through AGGIES in January, such a dilemma did not exist.

Two options exist for improving the I5 values even with heavy minus one use and non-response EO imputation through AGGIES. First, do a more detailed review of the imputation for non-response reports using current-to-previous IDAS graphs. A second, probably more effective solution, especially for EO non-response, would be to use a "seed" value for a variable, say 'ewes for breeding', to start the imputation off at an appropriate level.

Table 14 indicates that Wyoming had the only variables with an I9 index value less than 90%: namely, 'replacement lambs for breeding' and 'market lambs under 65 lbs.'.

Table 12. Editing Accuracy Index I3 for Each State

Variable	I3		
	California	Texas	Wyoming
Ewes for Breeding	100	100	100
Rams for Breeding	100	100	100
Replacement Lambs for Breeding	100	100	99
Market Lambs Under 65 lbs.	100	100	98
Market Lambs 65 to 84 lbs.	100	100	99
Market Lambs 85 to 105 lbs.	100	100	100
Market Lambs Over 105 lbs.	100	100	100
Market Sheep	100	100	100
Total Sheep and Lambs	100	99	96
Out of State Sheep	100	NA	100
Lamb Crop	99	100	100
Ewes Expected to Lamb	99	100	100

Table 13. Imputation Accuracy Index I5 for Each State

Variable	Threshold Percent	I5[1/]		
		California	Texas	Wyoming
Ewes for Breeding	10	58	76	94
Rams for Breeding	10	36	20	71
Replacement Lambs for Breeding	20	50	21	18
Market Lambs Under 65 lbs.	20	42	19	43
Market Lambs 65 to 84 lbs.	20	46	40	50
Market Lambs 85 to 105 lbs.	20	64	100	50
Market Lambs Over 105 lbs.	20	67	67	100
Market Sheep	20	36	75	100
Total Sheep and Lambs	10	31	45	70
Out of State Sheep	25	80	NA	-
Lamb Crop	10	38	24	79
Ewes Expected to Lamb	25	47	32	40

1/ A dash (-) indicates the index could not be computed because calculations would have resulted in division by zero

Low I5 values in addition to I3 values less than 100% contributed to these less than optimal I9 values. Nevertheless, the overall editing and imputation accuracy index, I9, gives evidence that AGGIES performed as well in processing these data as the current system.

3.3.4 SUMMARY OF THE JULY 1999 SHEEP PROJECT

Both the indices and the expanded total comparison indicated that AGGIES handled the July data appropriately. However, two more important accomplishments were achieved during this evaluation. First, since this project actually took place on location at the state offices, data management and flow could be assessed at the local level. No major problems were encountered. Second, with the states' sheep survey statisticians getting a first hand "look and feel" of AGGIES, excellent user feedback was provided to the researchers and developers of the system. Positive comments about AGGIES included the following: minus ones were allowed for every commodity cell, estimation of complete non-response was unnecessary even for EO's, the elimination of heavy manual editing saved time, and the integration of AGGIES and IDAS allowed for reviews and analyses to be completed in one system. Dislikes of the system included the following comments: imputation used only one historical data file, there were too many screens and choices to make, there was not a 're-impute' option in the interactive module, and the interactive module did not include the operation name, operator name, county or other identification information except for the ID. The final feedback comments dealt with general issues that need to be addressed such as user friendliness of the system, imputation order and estimators by state/region, availability of the system for use on State run surveys, and documentation for State use.

Table 14. Editing and Imputation Accuracy Index I9 for Each State

Variable	Threshold Percent	I9		
		California	Texas	Wyoming
Ewes for Breeding	10	96	99	99
Rams for Breeding	10	95	98	96
Replacement Lambs for Breeding	20	96	97	86
Market Lambs Under 65 lbs.	20	95	97	89
Market Lambs 65 to 84 lbs.	20	95	98	94
Market Lambs 85 to 105 lbs.	20	97	100	99
Market Lambs Over 105 lbs.	20	97	100	100
Market Sheep	20	95	100	100
Total Sheep and Lambs	10	94	97	90
Out of State Sheep	25	99	NA	100
Lamb Crop	10	92	97	96
Ewes Expected to Lamb	25	94	97	98

4. CONCLUSIONS AND RECOMMENDATIONS

Evaluations thus far have shown that the commodity data editing and imputation performed by AGGIES results in a data set similar to the one produced by NASS. At the very least, AGGIES does no worse at editing and imputing these data than the current data processing system.

Using AGGIES offers NASS several potential benefits:

1) The system provides statistically consistent results in the editing and imputation process. Results are nearly repeatable because the system is automated with the computer making all editing and imputation decisions.

2) The system is programmed in SAS, an agency supported language, thus, integration and implementation as a core data processing system is simplified.

3) The system can easily be applied to any number of surveys and, theoretically, censuses. Resources can be conserved in the development and maintenance of a single edit system.

4) The system minimizes the need to do a complete manual review at the micro level. This allows time for a more thorough review at the macro level which, in turn, can add to data quality.

However, there remain issues to address when considering AGGIES as a potential editing tool:

1) AGGIES will not perform all editing functions. It is designed for continuous, non-negative data. Editing of completion codes and data adjustment factors must be performed outside of the system.

2) A plan as to how AGGIES could be implemented in NASS's data processing to form a complete edit and imputation

strategy needs to be finalized and system integration details need to be settled. The integration of AGGIES with NASS's IDAS is already well underway. Results from the July 1999 Sheep Report study will provide insight into completing this integration and should lay the groundwork for AGGIES implementation plans and additional integrations.

The following recommendations are made:

1) Address AGGIES feedback comments and issues received from the state offices. Specifically, allow for more than one historical file, add more identification information to the interactive screens, and research alternative imputation order/estimator schemes. User friendliness and documentation should both remain as ongoing issues.

2) Research AGGIES using data from the 1997 Census of Agriculture and compare results to the final Census numbers.

3) Port AGGIES to the mainframe for all future evaluations in order to evaluate computational power, speed, and to simulate an operational client-server environment.

4) Evaluate AGGIES on crop/stock data. One major obstacle in evaluating these data is the sectional use of a categorical completion code. This code indicates whether the section is complete, either with positive data or valid zeros, or whether it should contain positive commodity data where actual inventories are unknown. For the livestock surveys, the summary weights are based on these completion codes, so item imputation is not affected by their use. However, on crop/stock surveys, item imputation is done based on this code. Since, AGGIES does not handle categorical data, pre-processing must be done. The extent of this pre-processing must be determined in order for evaluations of the use of AGGIES with these data to proceed.

5. REFERENCES

Cotton, C. (1993), "Functional Description of the Generalized Edit and Imputation System," Statistics Canada Technical Report.

De Jong, W. A. M. (1996), "Designing a Complete Edit Strategy; Combining Techniques," Working Paper No. 29, Conference of European Statisticians, Work Session on Statistical Data Editing, United Nations Statistical Commission and Economic Commission for Europe.

Manzari, A. and Della Rocca, G. (1999), "A Generalised System Based On a Simulation Approach to Test the Quality of Editing and Imputation Procedures," Working Paper No. 13, Conference of European Statisticians, Work Session on Statistical Data Editing, United Nations Statistical Commission and Economic Commission for Europe.

Pense, R. (1997), "Editing Strategies at the National Agricultural Statistics Service," Working Paper No. 31, Conference of European Statisticians, Work Session on Statistical Data Editing, United Nations Statistical Commission and Economic Commission for Europe.

Schiopu-Kratina, I. and Kovar, J. G. (1989), "Use of Chernikova's Algorithm in the Generalized Edit and Imputation System," Statistics Canada, Methodology Branch Working Paper No. BSMD-89-001E.

Survey Administration Manual for Agricultural Surveys, AgSAM (1999), unpublished documentation, National Agricultural Statistics Service, USDA, Washington, D.C.

Todaro, T. A. (1997), "Evaluation of the SPEER Automatic Edit and Imputation System," National Agricultural Statistics Service, USDA, Washington, D.C., RD Research Report No. RD-97-04.

Todaro, T. A. (1999a), "Evaluation of the AGGIES Automated Edit and Imputation System," National Agricultural Statistics Service, USDA, Washington, D.C., RD Research Report No. RD-99-01.

Todaro, T. A. (1999b), "Overview and Evaluation of the AGGIES Automated Edit and Imputation System," Working Paper No. 19, Conference of European Statisticians, Work Session on Statistical Data Editing, United Nations Statistical Commission and Economic Commission for Europe.

APPENDIX 1 - SYSTEM ENHANCEMENTS

Several modifications have been made to AGGIES since the preliminary evaluation of the system described by Todaro (1999a). This section, detailing these enhancements, will be organized by the order of their execution within the system: initializing the system, edit specification, edit/data group formation, check edits, edit summary, outlier detection, error localization, imputation, interactive editing, and IDAS integration.

INITIATING THE SYSTEM

The system is initiated by running the set-up program 'aggies.sas' which displays a 'Start' icon. Clicking on 'Start' displays the following screen, shown in Figure A1.

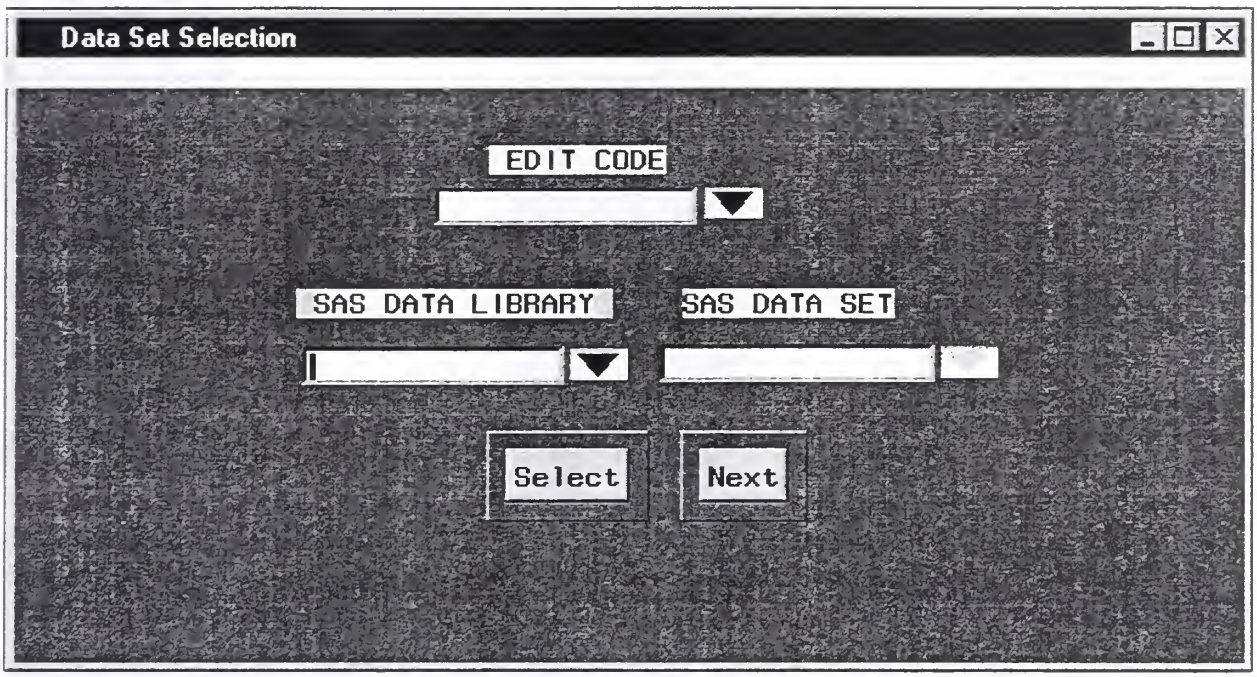


Figure A1. Selection of SAS Data Sets

The value for the edit code, a required entry, should be a SAS name that will be uniquely associated with a set of edits, edit descriptions and edit/data groups. A previously specified edit code can be selected from a list of all existing edit codes which is displayed by clicking on the control object (solid down arrow) located directly below the 'EDIT CODE' text label. When an existing edit code is selected, AGGIES retrieves the associated set of edits, edit descriptions, edit groups and data groups.

The file (SAS data set) to edit is chosen by selecting first a SAS data library and then a SAS data set. Pop-down menus used for these selections eliminate data entry errors. As an additional

enhancement, multiple SAS data sets can be chosen, provided the content of the SAS data sets is the same, i.e., same sort order, name, number and type of variables. These multiple SAS data sets are concatenated to form a single SAS data set to be edited. The selection of multiple SAS data set is accomplished by clicking on the 'Select' push button for each SAS data set displayed in the region below the 'SAS DATA SET' text label.

Once the SAS data set(s) has been chosen for editing, clicking on the 'Next' push button displays the following screen, shown in Figure A2.

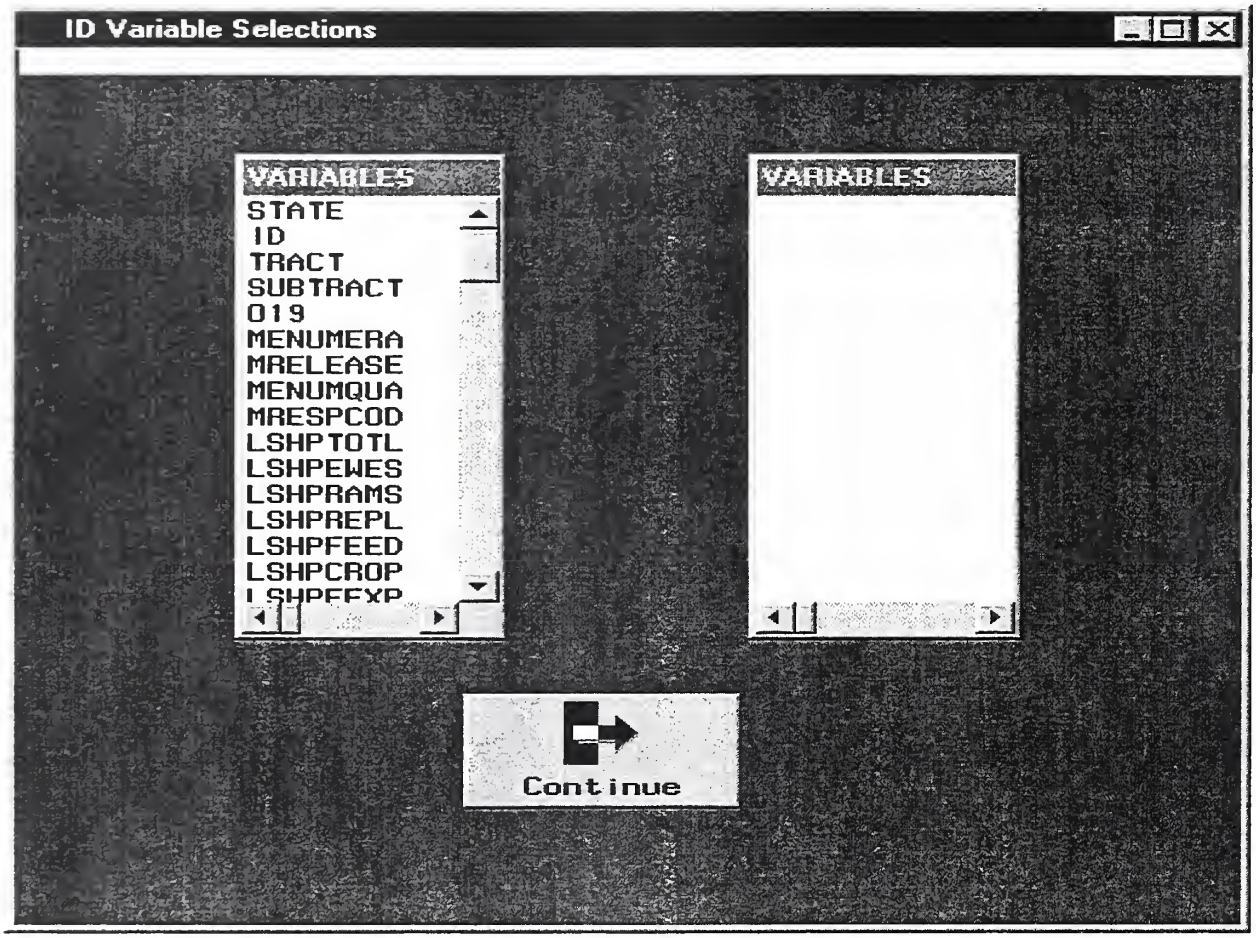


Figure A2. Selection of Identification Variables

This screen allows for the selection of up to five record identification variables whose values should uniquely identify each data record on the SAS data set(s) selected for editing. The left listbox in Figure A2 shows the variables in the order they appear on the SAS data set(s). An identification variable is selected by clicking on one of the variables in the left listbox, which moves the variable from the left listbox to the right listbox. Once selected, an identification variable can be de-selected by clicking on it in the right listbox.

After the selection of identification variables, the 'Continue' push button invokes AGGIES to search for edits, edit descriptions, edit groups and data groups that have been previously entered and associated with the edit code value. If no association is found, the edit specification screen is displayed as shown in Figure A3; otherwise the utility screen in Figure A4 is displayed.

SPECIFYING EDITS IN THE AGGIES

The most significant modification to this module was increasing the number of variables that may be used to construct an edit from ten to twenty. Another modification was the addition of a 'Description' push button that allows for entering a description of up to 200 characters describing the associated linear edit. The final modification was reversing the functions performed by the push buttons 'Continue' and 'Submit Edit'. The 'Continue' push button adds the edit and clears the screen at which time another edit may be entered.

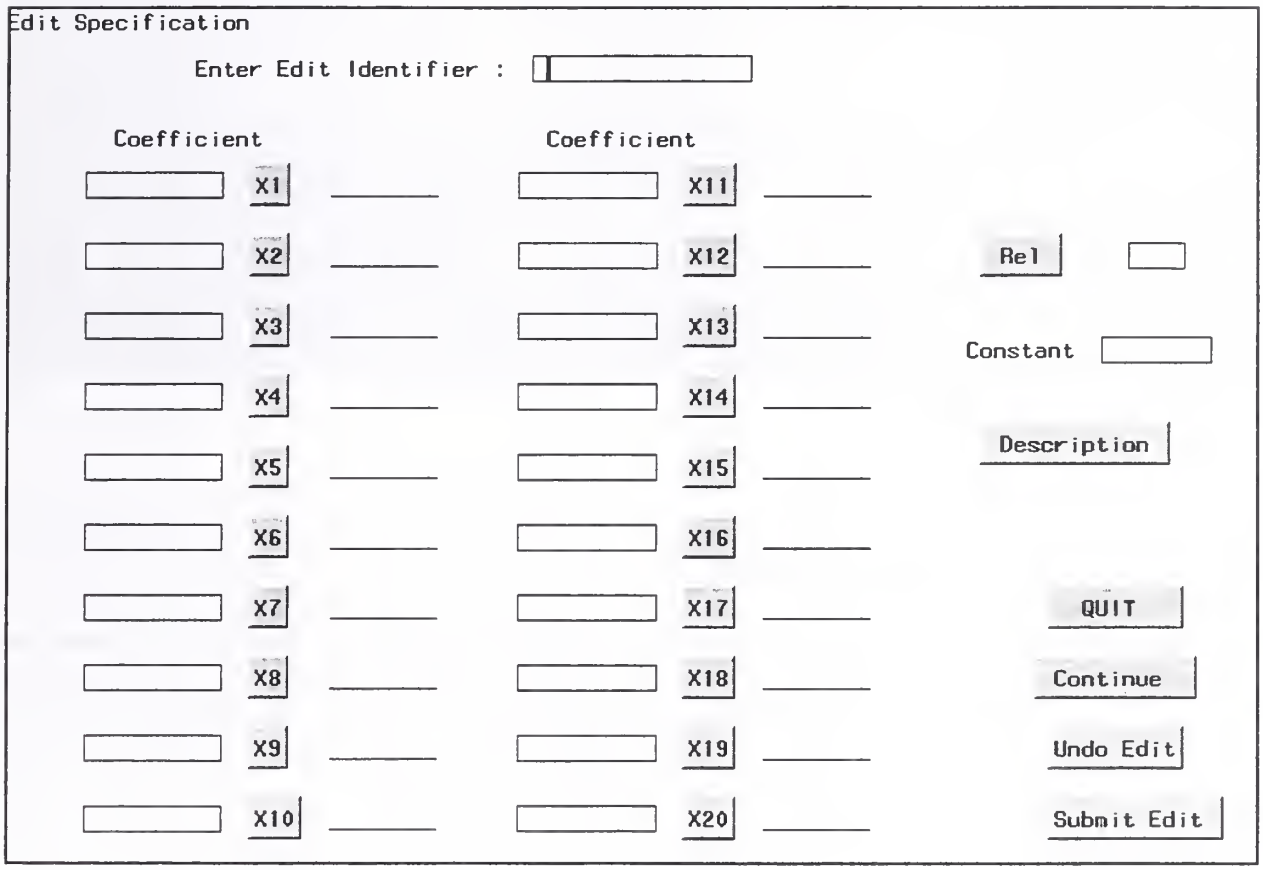


Figure A3. Edit Specification Screen

Clicking on the 'Submit Edit' push button displays the utility screen in Figure A4. With the addition of edit descriptions, modifications were made to the functions performed by the 'View All Edits' and 'Modify Edit' icons. Clicking on the 'View All Edits' icon allows for the display of edit descriptions

or the display of the linear edits. Selecting to view the edit descriptions displays the output in two columns: edit identifiers and edit descriptions. A 'Description' push button was added to the screen to modify edit descriptions which are displayed when the 'Modify Edit' icon is clicked.

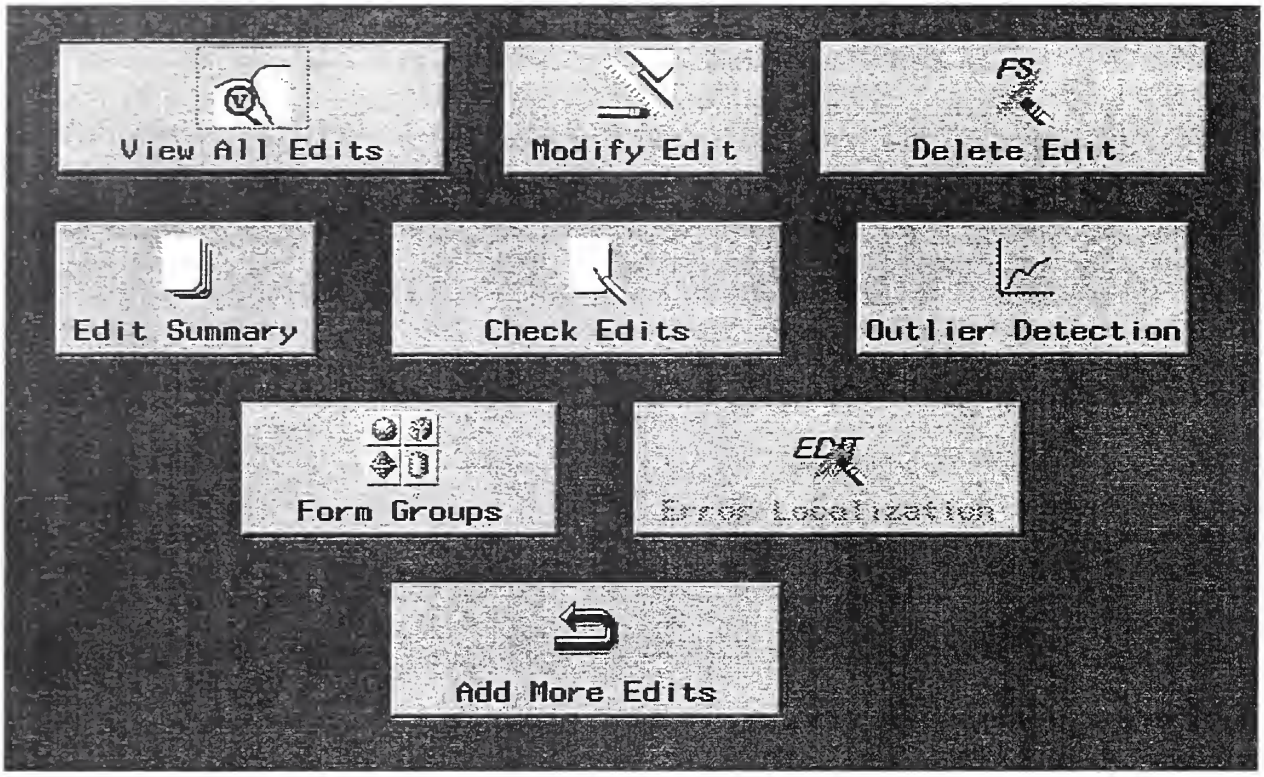


Figure A4. Utility Screen

FORMATION OF EDIT/DATA GROUPS

Data groups define subsets of the data, to which sets of edits (edit groups) are applied. The formation of edit and data groups has been re-designed so that the data groups are formed first and any number of data groups can be formed prior to forming the edit groups, i.e., prior to identifying which edits are valid for the subset of data. The data groups are formed by clicking on the 'Form Groups' icon on the Utility Screen, which displays the following figure, Figure A5.

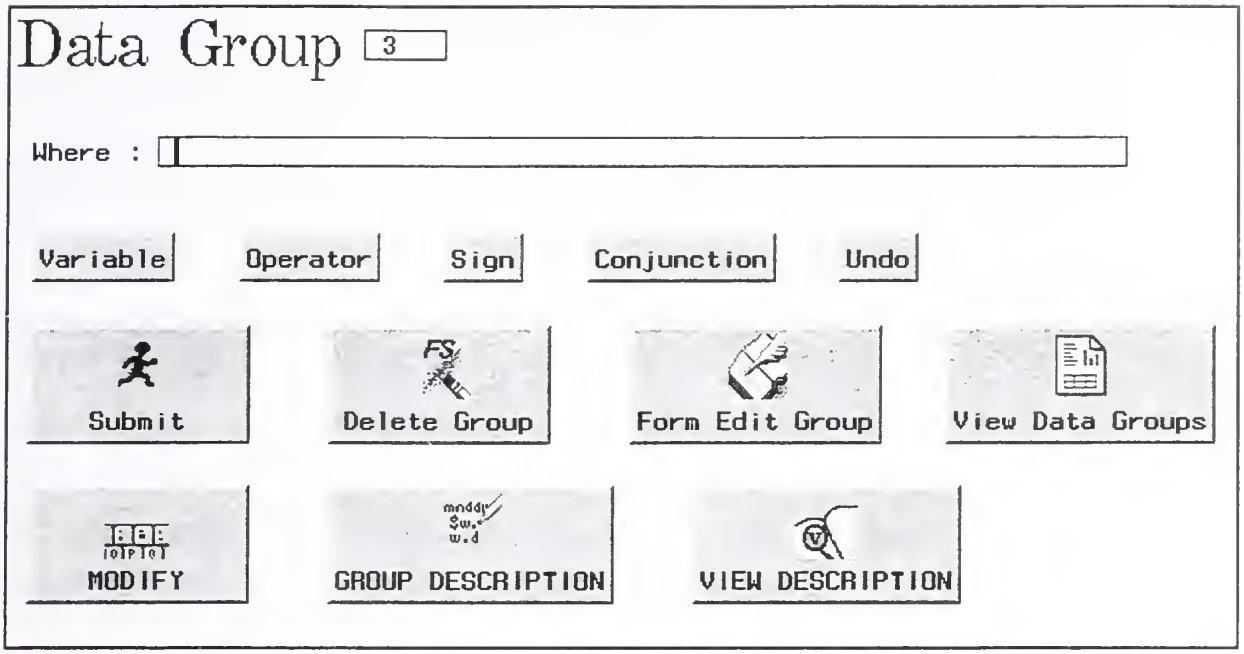


Figure A5. Data Group Screen

The data group number currently being formed is displayed in the text entry field to the right of the text label 'Data Group'. Data groups are numbered sequentially beginning with one. Any or all of the data groups can be formed prior to the formation of the associated edit groups. A data group is described by forming a valid SAS expression (SAS subsetting condition) describing the records that form the group in the text entry field to the right of the 'Where :' text label. New features include modifying data group expressions, forming group descriptions and viewing group descriptions.

Clicking on the 'Modify' icon displays a list of numbers corresponding to all data groups that have been submitted. Selecting a number from the list displays that data group's information, i.e., group number and its subsetting condition. After modifying the subsetting condition, the information can be submitted by clicking on the 'Submit' icon. This clears the subsetting condition expression and the next unused group number is displayed to the right of the 'Data Group' text label.

Group descriptions may be entered by clicking on the 'Group Description' icon and typing in a description of up to 200 characters. Clicking on the 'View Description' icon displays two columns, group number and group description. It is noted that a group description can be modified by clicking on the 'Modify' icon, selecting the group number from the displayed list of group numbers, clicking on the 'Group Description' icon and modifying the description.

After one or more data groups have been formed, the associated edit groups may be formed by clicking on the 'Form Edit Group' icon which displays the following screen, shown in Figure A6.

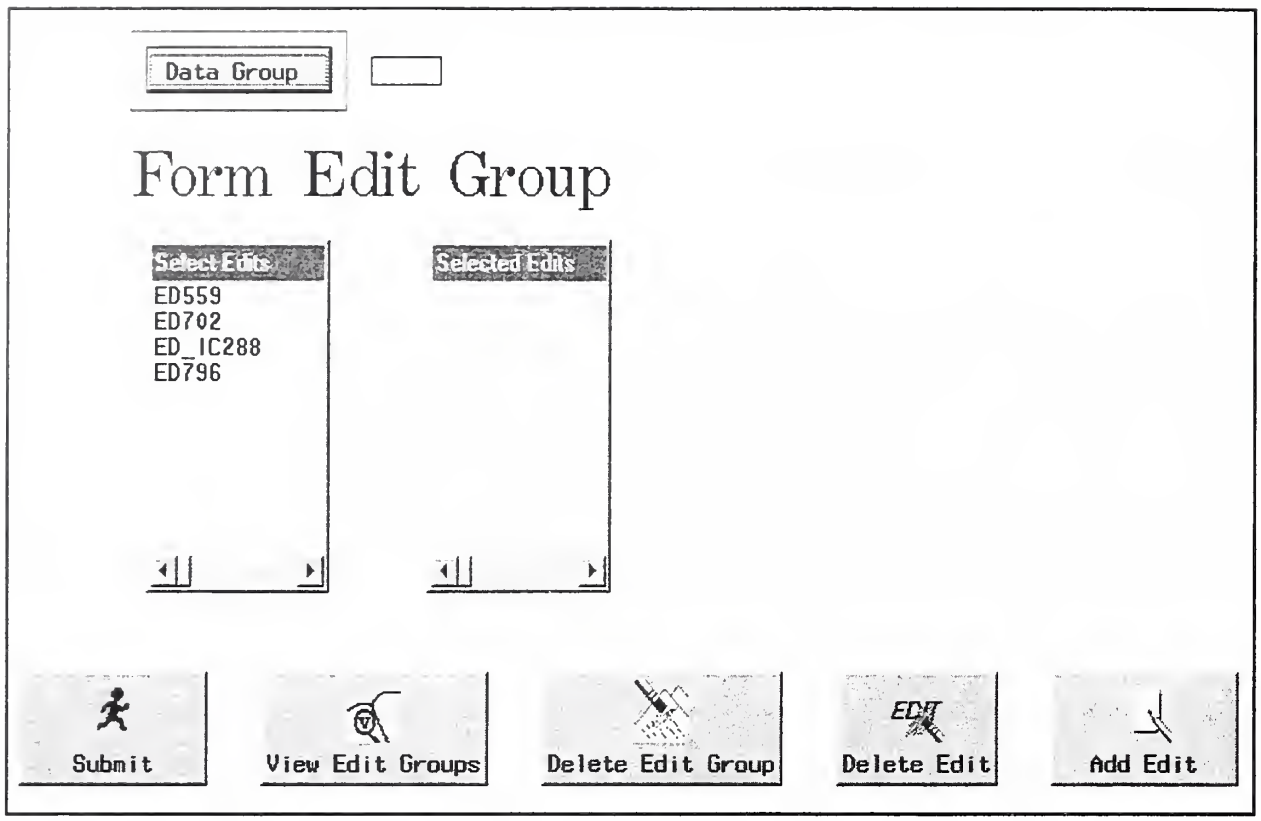


Figure A6. Edit Group Screen

Clicking on the 'Data Group' push button displays the list of data group numbers associated with the data groups that have been formed. The selected data group number is displayed in the text entry to the right of the text label 'Data Group'. The edit group formed is linked with this data group. Edits comprising an edit group are selected by clicking on the associated edit identifiers in the left listbox (Select Edits) which moves them to the right listbox (Selected Edits). Edits may be de-selected by clicking on the associated edit identifiers in the right listbox which moves them back to the left listbox. After an edit group has been submitted by clicking on the 'Submit' icon, an edit can be added to an edit group by clicking on the 'Add Edit' icon and selecting the edit group number and edit identifier.

CHECK EDITS

The output of the check edits module was modified to display the data group's mathematical expression next to the group number for each group formed.

EDIT SUMMARY

The edit summary module was modified to "manage" the data flow by requiring data groups to have

a specified cumulative frequency before allowing error localization and imputation to be executed (see Appendix 3.3). If a group did not have enough records for processing, the data records were permanently stored in SAS data sets and would be retrieved by AGGIES after exiting and re-entering a SAS session. When the specified percentage of data records was accumulated for a particular group, those data records for the group satisfying all edits were sent to a file accessible by IDAS for analysis, while those data records failing one or more edits were processed through error localization and imputation in AGGIES before being appended to the file accessible by IDAS.

The output of edit summary was expanded to display the data group's mathematical expression next to the group number for each group formed, to display summary statistics listing the cumulative percentage of records processed in each group, and to indicate the total number of data records contributing to each group. The number of data records tabulated as passing and failing each edit for each group are cumulated until the group has enough records for processing. When a group has enough data records for processing, a subsequent SAS session for the group will tabulate the number of records passing and failing each edit only for that current SAS session. Since the user-specified edits in the output are represented by the associated edit identifiers, an icon was added that displays the edit descriptions when clicked.

To compute the cumulative percentage of records processed for the groups and the total number of data records contributing to the groups, AGGIES accessed the sample master file of the form 'Smpmstr.smplXX', where XX is the state FIPS code specified in the set-up program, 'aggies.sas'. However, requiring this sample master file to exist for all applications of AGGIES would severely limit its generality. Therefore, if the sample master file does not exist, AGGIES will process the SAS data set selected for editing without requiring a certain percentage of records in each group.

OUTLIER DETECTION

The only modification made to the outlier detection module was to save the outliers to a file that can be accessed in the interactive module. This allows the analyst to override any changes made by AGGIES for those outlying data records.

ERROR LOCALIZATION

Variable weights were hard coded for the variables in the July 1999 Sheep Survey to reflect the perceived reliability of the variable values. Since, for the July 1999 Sheep Survey, a specified percentage of records was required before processing data records, the 'Error Localization' icon was grayed as unavailable until this percentage was met for at least one group. Once the required percentage was met for at least one group, the 'Error Localization' icon remained ungrayed. The output of this module was modified to display the data group's mathematical expression next to the group number for each group formed.

IMPUTATION

The selection of imputation estimators was enhanced to allow for selecting up to three current ratio estimators and up to three auxiliary trend estimators for each variable requiring imputation. This allows for the repeated use of these imputation estimators while using different auxiliary variables.

The output of the imputation module was modified to display the data group's mathematical expression next to the group number for each group formed. Additionally, an 'Interactive' icon was added to the output which, when clicked, displays the interactive screen as shown in Figure A7.

INTERACTIVE EDITING SCREEN

An interactive screen, shown in Figure A7, has been added to AGGIES that allows the batch-edited values to be interactively edited. It is noted that this screen has been customized for the July 1999 Sheep Survey. A generalized interactive edit has not yet been developed for AGGIES.

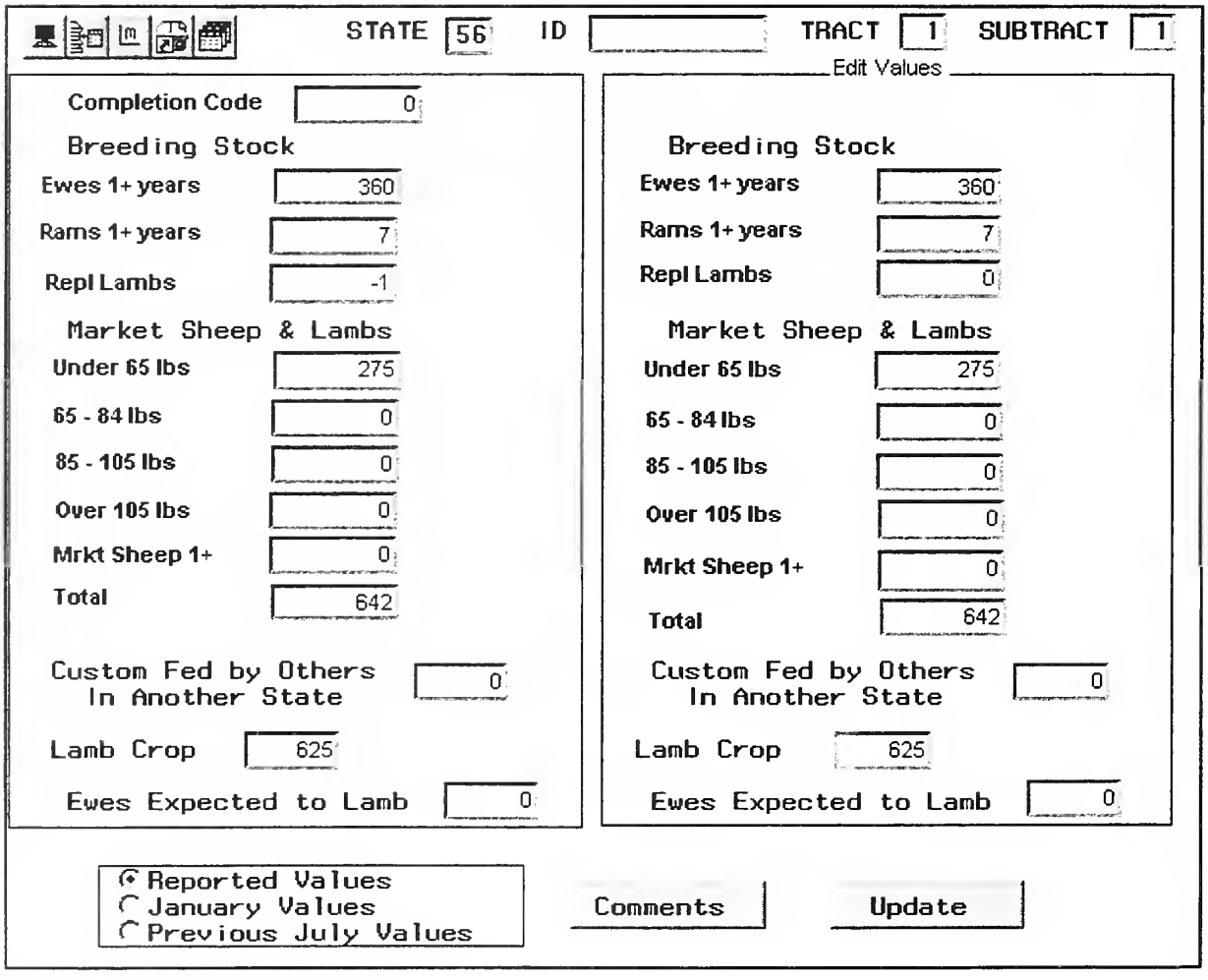


Figure A7. Interactive Screen

This screen displays two forms. The form on the right-hand side displays the current AGGIES batch-edited data that can be interactively modified. The form on the left-hand side displays information that may be useful for editing the data, such as originally reported data or historical data. When, in the process of interactively editing the values on the right-hand side form, one of the edits is violated, those cells containing values that are involved in at least one failed edit are highlighted in yellow.

Above the two forms, four identification variables and their values are displayed for the current data record. If data for a particular record are wanted, its identification values can be typed into the text entries to display its data values in the two forms. However, if incorrect identification values are entered, a message to this effect is displayed and all cells in the two forms are grayed.

The radio box beneath the left-hand side form provides for the selection of three options. The first, and default option, displays the originally reported values in the left-hand side form. When the reported values are displayed and there are differences in the values of the variables between the two forms, the differing values are displayed in red which can expedite the interactive editing process. The second and third options display the previous January and July values, respectively, for the data record, if available. These data are provided to aid in interactively editing those data records that look suspicious or to review changes made by AGGIES that appear suspect.

The toolbar located above the left form contains five icons. When the cursor is placed on any one of the icons, a short description of the icon's function is displayed. Clicking on the first icon displays the edit descriptions of edits that have been violated while interactively editing the current record. A listing of all data records failing one or more edits is displayed when the second icon is clicked. Thus, a review and possible update of all of the data records for which AGGIES made changes can be done. The third icon will display those data records which were classified as outliers in the outlier detection module, EO records and records that AGGIES failed to correct during error localization due to exceeding an imposed time limit. The fourth icon, when clicked, copies all values from the displayed left form to the right form. When clicking the fifth and final icon, the AGGIES batch-edited values are restored. This is a convenient option when changes have been made to the right form and it is desired to undo all of the changes made.

Changes made to the right-hand side form may be submitted by clicking on the 'Update' push button located to the bottom right of the screen. A comment facility is available by clicking on the 'Comments' push button located to the left of the 'Update' push button. When clicked, a screen is displayed whereby comments may be entered regarding interactive changes made. These comments can be accessed later through the use of IDAS.

MODIFICATIONS TO IDAS

For the July 1999 Sheep Survey, attempts were made to integrate AGGIES and IDAS. This integration occurred in two places in IDAS. The first distinguishes data records that passed all edits from those that AGGIES imputed due to one or more failed edits. Figure A8 displays a scatter plot obtained by selecting from the IDAS main menu – Daily Data Analysis, Analysis Tables, Curr vs.

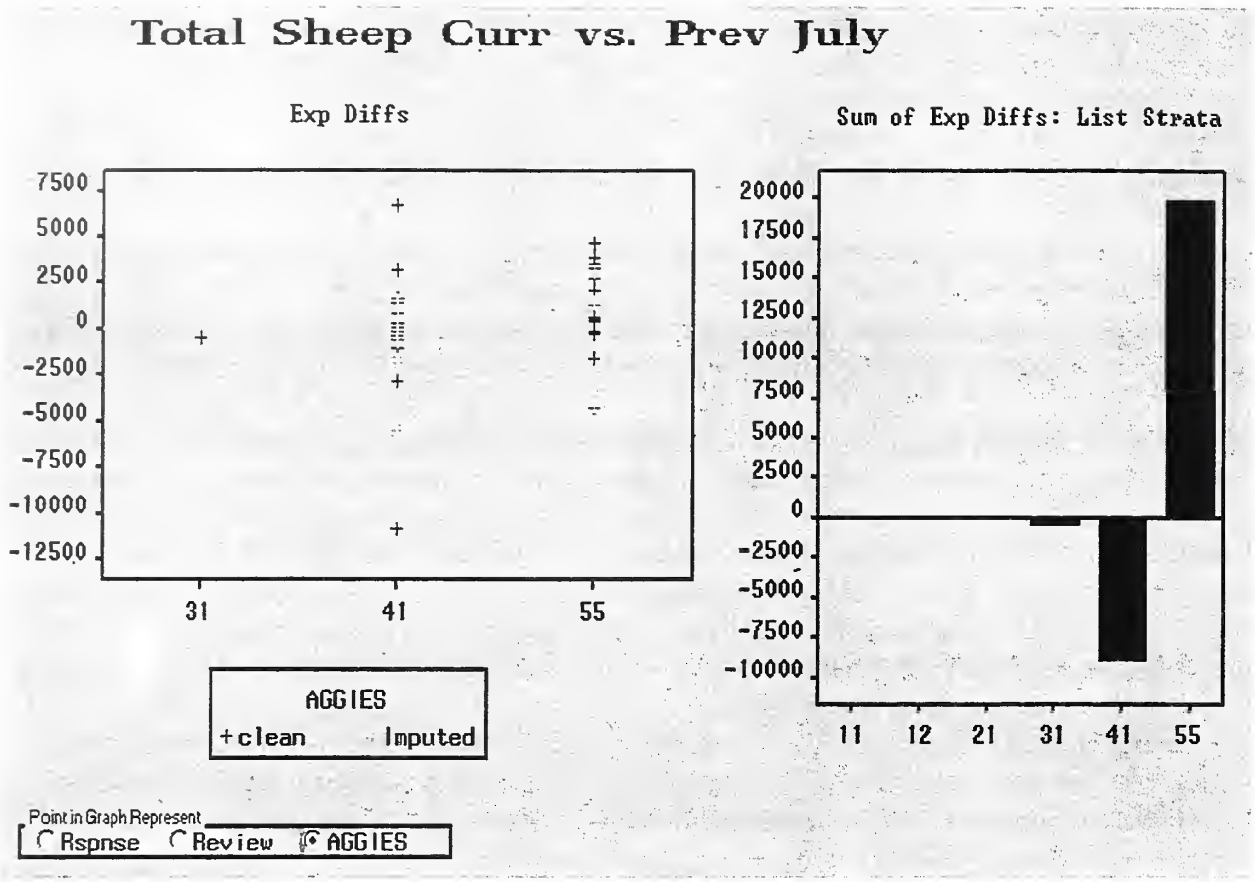


Figure A8. Scatter Plot

Prev, Total Sheep (or any other available selection). The clean values correspond to data records that passed all edits, while the imputed values failed at least one edit.

From the scatter plot screen, the drill down feature of IDAS is used to bring up the screen shown in Figure A9. It is on this screen that the second integration of AGGIES and IDAS took place. The top 'File' push button, on the far right side of the screen below the word 'Comment', was modified to provide access to the same comment file used in AGGIES. Also, the bottom push button, labeled 'Modify', was added to the screen to allow editors to modify data on a particular record. When 'Modify' is clicked, the interactive screen (Figure A7) appears and the editor can modify the data. The AGGIES edits would be interactively invoked. If interactive editing created no errors, the editor would then return to the IDAS screens and review other records. However, the IDAS set-up would need to be rerun to see the changes reflected in IDAS.

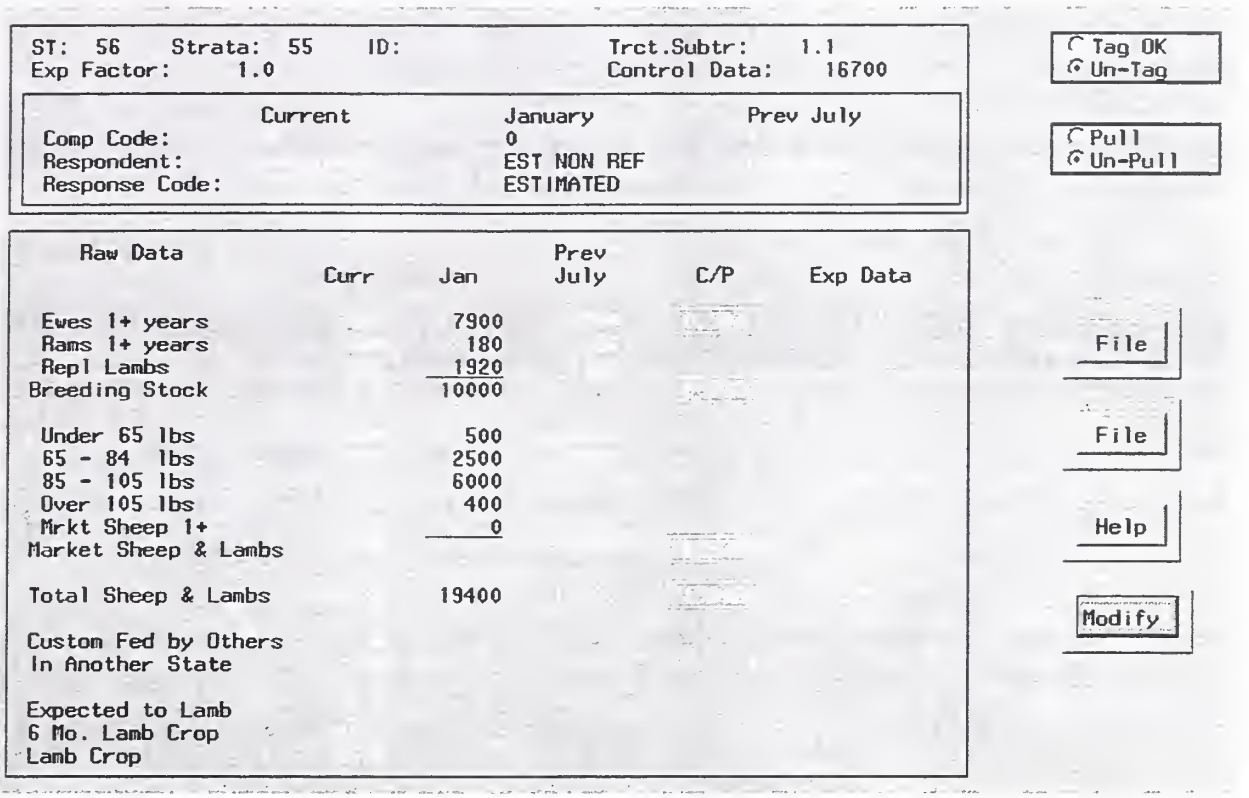


Figure A9. Drill Down Screen

APPENDIX 2 - JANUARY 1999 SHEEP PROJECT DETAILS

APPENDIX 2.1 - JANUARY AGGIES PARAMETER INPUTS

The following table shows the edits, along with descriptions, used to define an acceptable record in AGGIES. Descriptions for the SAS variable names can be found in Appendix 2.4.

Linear Edit Using SAS Variable Names	Description
lshpewes + lshprams + lshprepl + lshpu065 lshp6584 + lshp8505 + lshpo105 + lshpfeed - lshptotl = 0	Sum of ewes, rams, market lambs, replacement lambs and market sheep must equal total sheep and lambs
lshpewes - (0.001) lshpvewe >= 0	If ewe value is positive, then ewe inventory must be positive
lshprams - (0.0002) lshpvram >= 0	If ram value is positive, then ram inventory must be positive
lshprepl - (0.001) lshpvlmb >= 0	If replacement lamb value is positive, then replacement lamb inventory must be positive
lshpu065 + lshp6584 + lshp8505 + lshpo105 - (0.001) lsoflval >= 0	If market lamb value is positive, then market lamb inventory must be positive
lshpfeed - (0.001) lsofsval >= 0	If market sheep value is positive, then market sheep inventory must be positive
lshpwool + lshpmkwl - (0.1) lshppric >= 0	If wool price is positive, then wool production must be positive
lshpotst - lshptotl <= 0 (CA, CO, WY)	Number of head in another state must be less than total on hand
lshpwool - (0.00001) lshpshrn >= 0	If number of breeding head shorn is positive, then wool production for breeding head must be positive
lshpmkwl - (0.00001) lshpmksh >= 0	If number of market head shorn is positive, then wool production for market head must be positive
lshpvewe <= 999	Ewe value must be less than or equal to $999
lshpvram <= 5000	Ram value must be less than or equal to $5000
lshpvlmb <= 999	Replacement lamb value must be less than or equal to $999
lsoflval <= 999	Market lamb value must be less than or equal to $999
lsofsval <= 999	Market sheep value must be less than or equal to $999
lshpwool - (25) lshpshrn <= 0	Breeding wool production per head must be less than or equal to 25 pounds/head
lshpmkwl - (25) lshpmksh <= 0	Market wool production per head must be less than or equal to 25 pounds/head
lshpcrop <= 999999	Positivity edit such that missing lamb crop values are imputed

The table below displays other parameters used for each variable in AGGIES. Specifically shown are the reliability weights, imputation order and imputation estimators which used the January 1998 Sheep Report as the historical data source. Formulas for the different imputation estimators can be found in Appendix 5.

Variable	Weights	Order	Imputation Estimators
Ewes for Breeding	4	1	Auxiliary trend with rams for breeding Previous value
Rams for Breeding	5	2	Auxiliary trend with ewes for breeding Previous value
Replacement Lambs for Breeding	4	3	Auxiliary trend with ewes for breeding Previous value
Market Lambs Under 65 lbs.	2	4	Auxiliary trend with total sheep and lambs Previous value
Market Lambs 65 to 84 lbs.	1	5	Auxiliary trend with total sheep and lambs Previous value
Market Lambs 85 to 105 lbs.	1	6	Auxiliary trend with total sheep and lambs Previous value
Market Lambs Over 105 lbs.	2	7	Auxiliary trend with total sheep and lambs Previous value
Market Sheep	3	8	Auxiliary trend with total sheep and lambs Previous value
Total Sheep and Lambs	1	9	Difference trend Previous value
Out of State Sheep (CA, CO, WY)	3	10	Auxiliary trend with total sheep and lambs Previous value
Lamb Crop	1	11	Current ratio with ewes for breeding
Breeding Animals Shorn	2	12	Current ratio with wool from breeding animals Auxiliary trend with ewes for breeding Current ratio with ewes for breeding
Wool from Breeding Animals	1	13	Current ratio with breeding animals shorn
Market Animals Shorn	2	14	Current ratio with wool from market animals Auxiliary trend with total sheep and lambs Current ratio with total sheep and lambs
Wool from Market Animals	1	15	Current ratio with market animals shorn
Average Wool Price	1	16	Current mean
Average Ewe Value	1	17	Current mean
Average Ram Value	1	18	Current mean
Average Replacement Lamb Value	1	19	Current mean
Average Market Lamb Value	1	20	Current mean
Average Market Sheep Value	1	21	Current mean

APPENDIX 2.2 - JANUARY EXPANDED TOTALS

The following two tables display the expanded totals from both the AGGIES output file and the survey production file. Data for California and Colorado are shown in the first table followed by the second table showing Texas and Wyoming's data. Tables are sorted by variable in the order that the variables appear in the questionnaire.

Variable	California		Colorado	
	AGGIES Expanded Total	Survey Expanded Total[1/]	AGGIES Expanded Total	Survey Expanded Total
Ewes for Breeding	321,018	325,880	480,249	480,249
Rams for Breeding	14,824	14,764	11,998	11,998
Replacement Lambs for Breeding	45,597	45,609	78,584	78,584
Market Lambs Under 65 lbs.	213,881	214,326	7,383	7,373
Market Lambs 65 to 84 lbs.	79,375	78,274	5,159	5,021
Market Lambs 85 to 105 lbs.	67,960	67,643	51,135	51,260
Market Lambs Over 105 lbs.	28,137	28,137	152,108	152,108
Market Sheep	24,323	17,791	461	461
Total Sheep and Lambs	795,115	792,425	787,069	787,053
Out of State Sheep	23,037	23,037	26,529	26,529
Lamb Crop	286,230	289,773	764,924	765,536
Breeding Animals Shorn	330,253	338,863	533,329	533,073
Wool from Breeding Animals	2,758,588	2,829,874	5,173,817	5,176,730
Market Animals Shorn	123,053	150,430	203,091	203,341
Wool from Market Animals	484,885	622,187	980,930	982,945
Average Wool Price	0.64	0.60	0.53	0.53
Average Ewe Value	97	98	100	100
Average Ram Value	247	245	355	345
Average Replacement Lamb Value	105	91	88	88
Average Market Lamb Value	73	72	85	81
Average Market Sheep Value	85	85	54	54

1/ Reweighted estimator

Variable	Texas		Wyoming	
	AGGIES Expanded Total	Survey Expanded Total[1/]	AGGIES Expanded Total[2/]	Survey Expanded Total
Ewes for Breeding	885,371	889,686	363,222	361,095
Rams for Breeding	42,145	42,155	12,214	12,153
Replacement Lambs for Breeding	113,353	113,397	74,480	74,178
Market Lambs Under 65 lbs.	111,502	111,502	2,578	2,578
Market Lambs 65 to 84 lbs.	63,429	63,249	16,446	16,355
Market Lambs 85 to 105 lbs.	57,156	54,353	73,385	73,381
Market Lambs Over 105 lbs.	41,435	41,435	29,419	29,411
Market Sheep	11,359	9,677	2,618	1,739
Total Sheep and Lambs	1,325,750	1,325,453	574,363	570,890
Out of State Sheep	NA	NA	3,929	3,929
Lamb Crop	781,588	789,444	394,348	393,433
Breeding Animals Shorn	976,077	985,863	447,683	443,548
Wool from Breeding Animals	7,216,410	7,271,166	4,200,810	4,225,283
Market Animals Shorn	206,640	205,519	122,157	121,809
Wool from Market Animals	924,488	960,584	494,147	493,623
Average Wool Price	0.70	0.70	0.76	0.77
Average Ewe Value	69	69	88	88
Average Ram Value	219	184	283	281
Average Replacement Lamb Value	69	67	79	79
Average Market Lamb Value	65	65	70	70
Average Market Sheep Value	62	62	38	57

1/ Reweighted estimator
2/ Total is averaged over the three AGGIES runs

APPENDIX 2.3 - JANUARY ACCURACY INDICES

The following four tables, one for each state, show the editing and imputation accuracy indices, I1 through I9. All indices range from 0% (no accuracy) to 100% (maximum accuracy). Appendix 4 details the calculations of these indices. Tables are sorted by variable in the order that the variables appear in the questionnaire. Note: a dash (-) in the tables indicates that particular index could not be computed because the calculations would have resulted in division by zero.

California

Variable	I1	I2	I3	I4	I5	I6	I7	I8	I9
Ewes for Breeding	100	45	99	-	100	100	100	45	99
Rams for Breeding	100	50	100	-	100	100	100	50	100
Replacement Lambs for Breeding	100	67	100	-	100	100	100	67	100
Market Lambs Under 65 lbs.	100	50	99	-	100	100	100	50	99
Market Lambs 65 to 84 lbs.	100	100	100	0	100	33	100	100	100
Market Lambs 85 to 105 lbs.	100	100	100	0	100	67	100	100	100
Market Lambs Over 105 lbs.	100	-	100	-	-	-	100	-	100
Market Sheep	100	67	100	-	100	100	100	67	100
Total Sheep and Lambs	99	92	99	0	91	71	99	83	99
Out of State Sheep	100	-	100	-	-	-	100	-	100
Lamb Crop	100	60	99	-	67	67	100	40	99
Breeding Animals Shorn	100	42	99	-	60	60	100	25	98
Wool from Breeding Animals	100	85	99	-	50	50	100	43	94
Market Animals Shorn	100	69	99	-	22	22	100	15	98
Wool from Market Animals	99	70	98	0	31	24	99	22	96
Average Wool Price	100	87	99	-	96	96	100	84	98
Average Ewe Value	100	98	99	0	100	95	100	98	99
Average Ram Value	100	92	99	-	100	100	100	92	99
Average Replacement Lamb Value	100	95	100	-	97	97	100	92	99
Average Market Lamb Value	91	86	91	0	100	35	91	86	91
Average Market Sheep Value	100	100	100	-	100	100	100	100	100

Colorado

Variable	I1	I2	I3	I4	I5	I6	I7	I8	I9
Ewes for Breeding	100	100	100	-	100	100	100	100	100
Rams for Breeding	100	100	100	-	100	100	100	100	100
Replacement Lambs for Breeding	100	100	100	-	100	100	100	100	100
Market Lambs Under 65 lbs.	100	100	100	0	100	50	100	100	100
Market Lambs 65 to 84 lbs.	100	67	99	0	50	25	100	33	99
Market Lambs 85 to 105 lbs.	100	80	100	-	100	100	100	80	100
Market Lambs Over 105 lbs.	100	100	100	-	100	100	100	100	100
Market Sheep	100	100	100	-	100	100	100	100	100
Total Sheep and Lambs	100	95	99	-	100	100	100	95	99
Out of State Sheep	100	-	100	-	-	-	100	-	100
Lamb Crop	100	83	99	-	100	100	100	83	99
Breeding Animals Shorn	100	88	100	-	100	100	100	88	100
Wool from Breeding Animals	100	98	99	0	20	19	100	19	88
Market Animals Shorn	100	67	100	-	100	100	100	67	100
Wool from Market Animals	100	83	99	-	67	67	100	56	99
Average Wool Price	100	11	73	-	85	85	100	9	73
Average Ewe Value	100	95	100	-	62	62	100	59	98
Average Ram Value	100	87	99	-	65	65	100	57	98
Average Replacement Lamb Value	100	97	100	-	89	89	100	86	99
Average Market Lamb Value	98	95	98	0	94	59	98	89	98
Average Market Sheep Value	100	100	100	-	100	100	100	100	100

Texas

Variable	I1	I2	I3	I4	I5	I6	I7	I8	I9
Ewes for Breeding	100	33	100	-	100	100	100	33	100
Rams for Breeding	100	33	100	-	100	100	100	33	100
Replacement Lambs for Breeding	100	75	100	-	100	100	100	75	100
Market Lambs Under 65 lbs.	100	100	100	-	100	100	100	100	100
Market Lambs 65 to 84 lbs.	100	100	100	0	100	50	100	100	100
Market Lambs 85 to 105 lbs.	100	100	100	0	100	80	100	100	100
Market Lambs Over 105 lbs.	100	100	100	-	100	100	100	100	100
Market Sheep	100	75	100	-	100	100	100	75	100
Total Sheep and Lambs	100	88	100	17	100	76	100	88	100
Lamb Crop	100	52	100	-	45	45	100	24	99
Breeding Animals Shorn	100	54	100	-	57	57	100	31	100
Wool from Breeding Animals	100	87	100	-	27	27	100	24	98
Market Animals Shorn	100	71	100	-	80	80	100	57	100
Wool from Market Animals	100	95	100	0	11	9	100	10	99
Average Wool Price	100	100	100	0	79	77	100	79	99
Average Ewe Value	100	98	100	-	95	95	100	93	100
Average Ram Value	100	78	99	0	100	98	100	78	99
Average Replacement Lamb Value	100	94	100	-	100	100	100	94	100
Average Market Lamb Value	99	98	99	0	100	79	99	98	99
Average Market Sheep Value	100	100	100	-	100	100	100	100	100

Wyoming[1/]

Variable	I1	I2	I3	I4	I5	I6	I7	I8	I9
Ewes for Breeding	100	50	99	-	100	100	100	50	99
Rams for Breeding	100	50	99	-	87	87	100	43	99
Replacement Lambs for Breeding	100	56	100	-	100	100	100	56	100
Market Lambs Under 65 lbs.	100	100	100	-	100	100	100	100	100
Market Lambs 65 to 84 lbs.	100	75	100	75	81	74	100	58	100
Market Lambs 85 to 105 lbs.	100	100	100	0	100	89	100	100	100
Market Lambs Over 105 lbs.	100	67	100	-	100	100	100	67	100
Market Sheep	100	40	100	-	100	100	100	40	100
Total Sheep and Lambs	100	65	99	41	91	78	100	59	99
Out of State Sheep	100	-	100	-	-	-	100	-	100
Lamb Crop	100	97	100	-	87	87	100	84	99
Breeding Animals Shorn	100	91	99	-	53	53	100	48	97
Wool from Breeding Animals	100	94	99	0	16	16	100	15	86
Market Animals Shorn	100	100	100	-	47	47	100	47	99
Wool from Market Animals	99	77	99	0	13	9	99	10	98
Average Wool Price	100	100	100	-	67	67	100	67	91
Average Ewe Value	100	100	100	0	94	93	100	94	99
Average Ram Value	100	94	100	0	98	96	100	91	99
Average Replacement Lamb Value	100	93	99	0	95	93	100	89	99
Average Market Lamb Value	100	98	99	0	100	92	100	98	99
Average Market Sheep Value	100	85	99	-	100	100	100	85	99

1/ Indices averaged over the three AGGIES runs

APPENDIX 2.4 - JANUARY SHEEP REPORT QUESTIONNAIRE

The following January 1999 Sheep questionnaire is a condensed version that displays all variables used in the AGGIES evaluation. Variables that are state specific are identified as such. Each cell has a number and a set of letters indicating the key item codes and SAS variable names, respectively.

2.	a. **Ewes** 1 year old and older?	+	281 lshpewes
	b. **Rams** 1 year old and older?	+	282 lshprams
	c. **Replacement Lambs** under 1 year old?	+	285 lshprepl
4.	a. (1) **Market Lambs** under 65 pounds?	+	836 lshpu065
	(2) **Market Lambs** 65 to 84 pounds?	+	837 lshp6584
	(3) **Market Lambs** 85 to 105 pounds?	+	838 lshp8505
	(4) **Market Lambs** Over 105 pounds?	+	839 lshpo105
	b. **Market Sheep** 1 year old and older?	+	287 lshpfeed
5.	Then the **Total Sheep and Lambs** owned or custom fed by this operation on January 1 was:	=	280 lshptotl
6.	(CA, CO, and WY) How many **head were in another State**?		385 lshpotst
7.	How many **Lambs Dropped** during1998 were or will be Marked, Docked, or Branded?		288 lshpcrop
11.	How many **sheep and lambs for breeding were shorn** in 1998? **Head**		274 lshpshrn
	11a. How many **pounds of wool were shorn** from these sheep and lambs for breeding in 1998? **Pounds**		275 lshpwool
12.	How many **sheep and lambs for market were shorn** in 1998? **Head**		276 lshpmksh
	12a. How many **pounds of wool were shorn** from these sheep and lambs for market in 1998? **Pounds**		277 lshpmkwl
13.	**Average price received per pound** for wool sold in 1998? **Dollars and cents**		296 lshppric
14.	14a. **Average value** per head for breeding ewes 1 year old and older?	$	680 lshpvewe
	14b. **Average value** per head for breeding rams 1 year old and older?	$	681 lshpvram
	14c. **Average value** per head for breeding lambs under 1 year old?	$	679 lshpvlmb
	14d. **Average value** per head for market lambs under 1 year old?	$	845 lsoflval
	14e. **Average value** per head for market sheep 1 year old and older?	$	846 lsofsval

(CO and WY)

16. Predator Causes:		LAMB deaths before being marked, docked, or branded	LAMB deaths after being marked, docked, or branded	SHEEP deaths
Bears	+	163 lslmlbbr	953 lslmlabr	042 lshplpbr
Bobcats or lynx	+	089 lslmlbbc	952 lslmlabc	041 lshplpbc
Coyotes	+	087 lslmlbcy	950 lslmlacy	038 lshplpcy
Dogs	+	086 lslmlbdg	689 lslmladg	037 lshplpdg
Mountain lions	+	164 lslmlbml	954 lslmlaml	980 lshplpml
Fox	+	085 lslmlbfx	688 lslmlafx	036 lshplpfx
Wolves	+	084 lslmlbwv	687 lslmlawv	039 lshplpwv
Eagles	+	088 lslmlbeg	951 lslmlaeg	040 lshplpeg
Other predators [*specify*]	+	165 lslmlboa	955 lslmlaoa	049 lshplpoa
Unknown predators	+	168 lslmlbmd	960 lslmlard	060 lshplnod
17. Non-predator Causes:				
Disease	+	171 lslmlbdo	963 lslmlado	063 lshplndo
Weather related causes	+	166 lslmlbmw	956 lslmlarw	050 lshplnow
Lambing problems	+	390 lslmlbmc		053 lshplnoc
Old age	+			055 lshplnag
Being on their back	+	392 lslmlbmb	959 lslmlarb	054 lshplnob
Poisoning	+	389 lslmlbmp	958 lslmlarp	052 lshplnop
Theft	+	394 lslmlbmt	024 lslmlart	056 lshplnot
Other non-predator causes	+	685 lslmlbmo	027 lslmlaro	057 lshplnoo
Unknown non-predator causes	+	686 lslmlbuk	032 lslmlauk	058 lshplouk
18. *[Add lamb and sheep deaths by cause in each column.]*	=	690 lslmlbot	028 lslmlaot	059 lshplotl

(TX)

		LAMBS	SHEEP
12.	12a. How many lambs and sheep were killed by predators? . . . . . +	035 lslmlapd	981 lshplbpd
	12b. How many lambs and sheep died or were lost from disease or other **known** causes? . . . . . +	027 lslmlaro	057 lshplnoo
	12c. How many lambs and sheep died or were lost from **unknown** causes? . . . . . +	032 lslmlauk	058 lshplouk
13.	[*Add lamb and sheep deaths by cause in each column.*] . . . . . =	028 lslmlaot	059 lshplotl

APPENDIX 3 - JULY 1999 SHEEP PROJECT DETAILS

APPENDIX 3.1 - JULY AGGIES PARAMETER INPUTS

The following table shows the edits, along with descriptions, used to define an acceptable record in AGGIES. Descriptions for the SAS variable names can be found in Appendix 3.6.

Linear Edit Using SAS Variable Names	Description
lshpewes + lshprams + lshprepl + lshpu065 lshp6584 + lshp8505 + lshpo105 + lshpfeed - lshptotl = 0	Sum of ewes, rams, market lambs, replacement lambs and market sheep must equal total sheep and lambs
lshpotst - lshptotl <= 0 (CA, CO, WY)	Number of head in another state must be less than total on hand
lshpcrop <= 999999	Positivity edit such that missing lamb crop values are imputed
lshpeexp - lshpewes <= 0	Ewes expected to lamb must be less than ewe inventory on hand

The table below displays other parameters used for each variable in AGGIES. Specifically shown are the reliability weights, imputation order and imputation estimators which used the January 1999 Sheep Report as the historical data source. Formulas for the different imputation estimators can be found in Appendix 5.

Variable	Weights	Order	Imputation Estimators
Ewes for Breeding	4	1	Auxiliary trend with rams for breeding Current ratio with rams for breeding Difference trend Previous value Current mean
Rams for Breeding	5	2	Current ratio with ewes for breeding
Replacement Lambs for Breeding	4	3	Current ratio with ewes for breeding
Market Lambs Under 65 lbs.	1	9	Current ratio with total sheep and lambs Auxiliary trend with total sheep and lambs
Market Lambs 65 to 84 lbs.	1	8	Current ratio with total sheep and lambs Auxiliary trend with total sheep and lambs
Market Lambs 85 to 105 lbs.	2	7	Current ratio with total sheep and lambs Auxiliary trend with total sheep and lambs
Market Lambs Over 105 lbs.	2	6	Current ratio with total sheep and lambs Auxiliary trend with total sheep and lambs
Market Sheep	3	5	Current ratio with total sheep and lambs Auxiliary trend with total sheep and lambs
Total Sheep and Lambs	1	4	Difference trend Previous value Current mean
Out of State Sheep (CA, CO, WY)	3	10	Auxiliary trend with total sheep and lambs Previous value Current ratio with total sheep and lambs
Lamb Crop	1	12	Current ratio with ewes for breeding
Ewes Expected to Lamb	1	11	Current ratio with ewes for breeding

APPENDIX 3.2 - JULY MEMO ON MANUAL EDITING GUIDELINES

Plans are to follow the flow as outlined on the flowchart given out at the National Conference. Pre-survey processing, the Blaise setup and data collection are as usual except hand editing should be basically limited to coding (cells 921, 930, 941, 924-928, 291, 099, 101, 910, 098, 100, 987, and 789), legibility checks, updating sheep data based on enumerator notes, changing DK's to -1, putting in -1 for missing sheep data you know should be positive (perhaps enumerator notes tell you, or you know from personal experience with the operation), and following the AgSAM instructions when dealing with sheep in another state. Blaise interviewing, interactive editing, Blaise data management, SPS edits and updates, IDAS reviews, etc., should all proceed as usual.

The following are a few guidelines for editing any paper questionnaires for the July 1999 Sheep Survey.

First, try to send all questionnaires through Blaise. I know that toward the end of the survey, the pressure is on and getting those last inaccessibles through Blaise is not a high priority. If some don't make it through, please let us know when we get out there.

Blaise interactive edit has been modified to flag coding and DAF problems but to let other errors, like the sum of the parts unequal to the total, go through without an error. The Blaise coding checks (i.e., face page, partner page, completion code, and back page) will be identical to those in the past. That is:

921 = 1-5, 8-13 (for EO's, must not be coded 11 and 12)
930 is for out-of-business coding
941 is substitution coding
925-928 is for partners
924 is partner operating status
101 = 1-6 (for 5 and 6, 910 =4)
910 = 1-9 (for EO's, must not be coded 6-9)

Also, Blaise should accept minus one (-1) as valid for any cell in section 1 regardless of EO/non-EO status. Use your best judgement on whether to use minus ones or the completion code box for partial non-response. I would treat it like the stocks page on the crops/stocks questionnaire. That is:

> If the respondent has inventory for a particular cell, but the amount is not known, enter (-1) in the cell. Data will be imputed for items coded (-1) only, rather than for the entire sheep section. Leave the completion code blank. Do this for EO's and non-EO's.
>
> If the respondent has sheep but break-outs and amounts are not known: For non-EO's, code the completion code 1. For EO's, put in (-1) for every cell and leave the completion code blank.

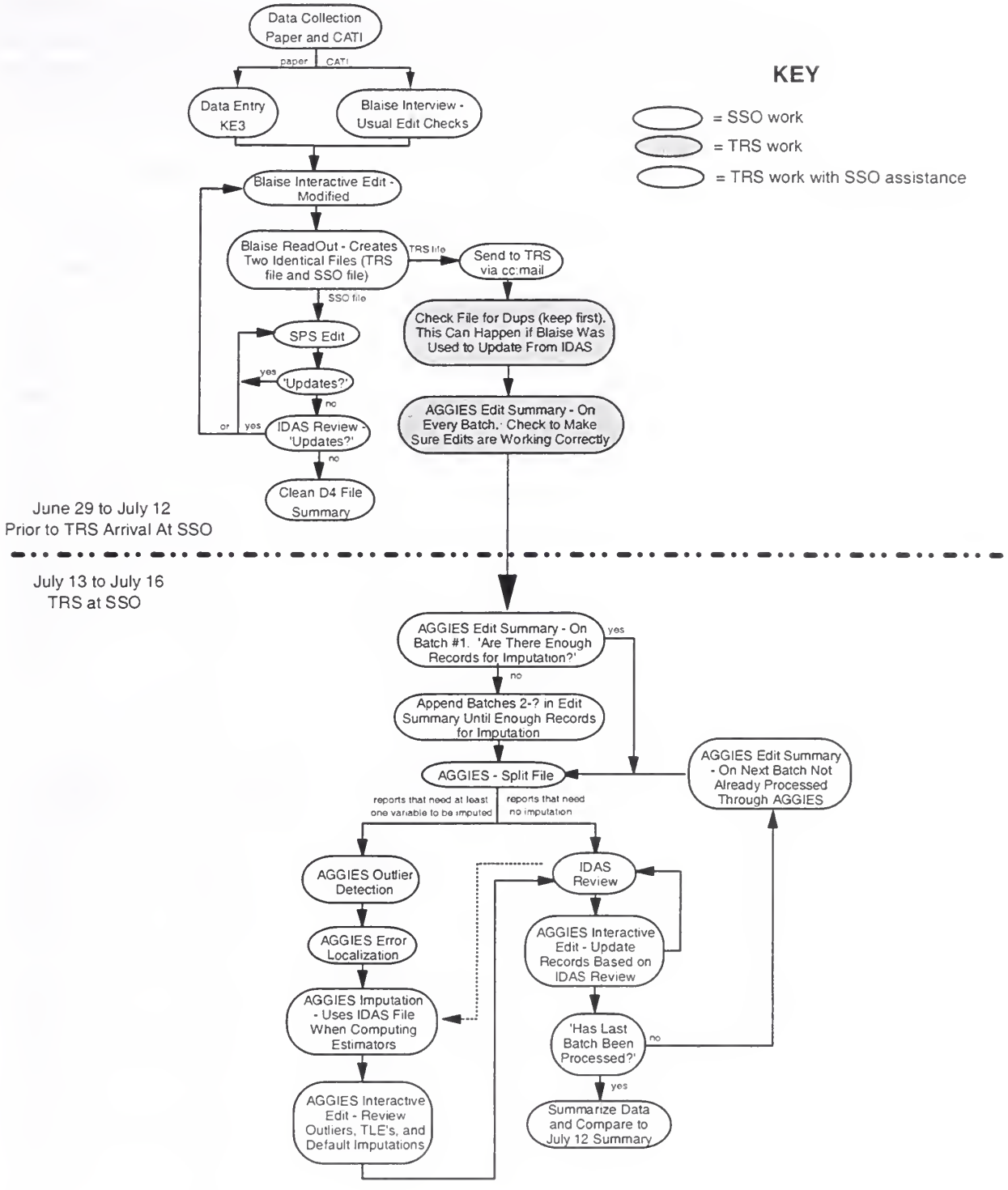
If you don't know whether the operation has sheep: For non-EO's, code the completion code 2. For EO's, put (-1) in every cell and leave the completion code blank.

Remember, the SPS edit has not changed. You will have to clean up all the minus ones and other errors that Blaise allowed and the EO's will have to be estimated for the SPS edit like always. To make updating the SPS edit easier, you may want to enter these estimated values in the questionnaire margin prior to key entry. However, please make sure that key punch is instructed not to key in those values; they'll be keyed later as updates on the SPS edit.

When you're correcting errors from the SPS edit or IDAS reviews, make corrections where you normally would, i.e., either as updates on the SPS edit, through Blaise as a reconverted case, or a combination thereof. It makes no difference to this research project.

During the survey period (from now through July 11), use the usual sheep survey project code. For the week we're out there (July 12 through July 16), use project code 505, New Technology Research, for any work relating to sheep and/or AGGIES.

APPENDIX 3.3 - JULY PROJECT DATA FLOW



APPENDIX 3.4 - JULY EXPANDED TOTALS

The following two tables display the expanded totals, which include the area non-overlap estimate (NOL), from both the AGGIES output file and the survey production file. Data for California and Colorado are shown in the first table followed by the second table showing Texas and Wyoming's data. Tables are sorted by variable in the order that the variables appear in the questionnaire.

Variable	California		Colorado	
	AGGIES Expanded Total	Survey Expanded Total[1/]	AGGIES Expanded Total	Survey Expanded Total
Ewes for Breeding	335,795	352,874	175,366	175,366
Rams for Breeding	14,586	15,104	5,218	5,218
Replacement Lambs for Breeding	32,414	33,229	22,016	22,016
Market Lambs Under 65 lbs.	27,386	26,583	121,456	123,913
Market Lambs 65 to 84 lbs.	43,176	46,221	59,485	59,485
Market Lambs 85 to 105 lbs.	36,586	42,736	39,611	39,611
Market Lambs Over 105 lbs.	25,508	20,479	93,831	93,831
Market Sheep	6,258	5,327	457	457
Total Sheep and Lambs	521,707	542,553	517,439	519,896
Out of State Sheep	39,345	51,014	12,089	12,089
Lamb Crop	159,425	170,416	192,578	192,920
Ewes Expected to Lamb	224,057	222,051	10,421	10,421

Variable	Texas		Wyoming	
	AGGIES Expanded Total	Survey Expanded Total[1/]	AGGIES Expanded Total	Survey Expanded Total
Ewes for Breeding	898,162	895,661	393,528	382,633
Rams for Breeding	43,921	43,498	13,414	13,229
Replacement Lambs for Breeding	135,632	134,106	79,493	70,843
Market Lambs Under 65 lbs.	359,770	364,868	260,225	286,722
Market Lambs 65 to 84 lbs.	113,082	110,551	20,549	22,791
Market Lambs 85 to 105 lbs.	37,504	38,360	19,425	13,664
Market Lambs Over 105 lbs.	14,649	16,637	8,621	8,643
Market Sheep	15,005	14,482	12,476	12,618
Total Sheep and Lambs	1,617,725	1,618,162	807,730	811,143
Out of State Sheep	NA	NA	7,600	7,647
Lamb Crop	731,232	708,977	363,561	362,452
Ewes Expected to Lamb	209,084	190,211	4,898	2,672

1/ Reweighted estimator

APPENDIX 3.5 - JULY ACCURACY INDICES

The following four tables, one for each state, show the editing and imputation accuracy indices, I1 through I9. All indices range from 0% (no accuracy) to 100% (maximum accuracy). See Appendix 4 for details on index calculations. Tables are sorted by variable in the order that the variables appear in the questionnaire. Note: a dash (-) in the tables indicates that particular index could not be computed because the calculations would have resulted in division by zero.

California

Variable	I1	I2	I3	I4	I5	I6	I7	I8	I9
Ewes for Breeding	100	100	100	-	58	58	100	58	96
Rams for Breeding	100	100	100	-	36	36	100	36	95
Replacement Lambs for Breeding	100	100	100	-	50	50	100	50	96
Market Lambs Under 65 lbs.	100	100	100	-	42	42	100	42	95
Market Lambs 65 to 84 lbs.	100	100	100	-	46	46	100	46	95
Market Lambs 85 to 105 lbs.	100	100	100	-	64	64	100	64	97
Market Lambs Over 105 lbs.	100	100	100	-	67	67	100	67	97
Market Sheep	100	100	100	-	36	36	100	36	95
Total Sheep and Lambs	100	100	100	-	31	31	100	31	94
Out of State Sheep	100	100	100	-	80	80	100	80	99
Lamb Crop	99	100	99	0	38	35	99	38	92
Ewes Expected to Lamb	100	94	99	-	47	47	100	44	94

Colorado

Variable	I1	I2	I3	I4	I5	I6	I7	I8	I9
Ewes for Breeding	100	-	100	-	-	-	100	-	100
Rams for Breeding	100	-	100	-	-	-	100	-	100
Replacement Lambs for Breeding	100	-	100	-	-	-	100	-	100
Market Lambs Under 65 lbs.	99	-	99	0	-	0	99	-	99
Market Lambs 65 to 84 lbs.	100	-	100	-	-	-	100	-	100
Market Lambs 85 to 105 lbs.	100	-	100	-	-	-	100	-	100
Market Lambs Over 105 lbs.	100	-	100	-	-	-	100	-	100
Market Sheep	100	-	100	-	-	-	100	-	100
Total Sheep and Lambs	100	83	99	-	100	100	100	83	99
Out of State Sheep	100	-	100	-	-	-	100	-	100
Lamb Crop	100	67	99	-	100	100	100	67	99
Ewes Expected to Lamb	100	-	100	-	-	-	100	-	100

Texas

Variable	I1	I2	I3	I4	I5	I6	I7	I8	I9
Ewes for Breeding	100	100	100	-	76	76	100	76	99
Rams for Breeding	100	100	100	-	20	20	100	20	98
Replacement Lambs for Breeding	100	100	100	-	21	21	100	21	97
Market Lambs Under 65 lbs.	100	100	100	50	19	22	100	19	97
Market Lambs 65 to 84 lbs.	100	100	100	-	40	40	100	40	98
Market Lambs 85 to 105 lbs.	100	100	100	-	100	100	100	100	100
Market Lambs Over 105 lbs.	100	100	100	-	67	67	100	67	100
Market Sheep	100	100	100	-	75	75	100	75	100
Total Sheep and Lambs	100	87	99	-	45	45	100	39	97
Lamb Crop	100	100	100	0	24	22	100	24	97
Ewes Expected to Lamb	100	100	100	0	32	30	100	32	97

Wyoming

Variable	I1	I2	I3	I4	I5	I6	I7	I8	I9
Ewes for Breeding	100	100	100	-	94	94	100	94	99
Rams for Breeding	100	100	100	-	71	71	100	71	96
Replacement Lambs for Breeding	100	96	99	-	18	18	100	17	86
Market Lambs Under 65 lbs.	98	95	98	0	43	39	98	41	89
Market Lambs 65 to 84 lbs.	100	88	99	-	50	50	100	44	94
Market Lambs 85 to 105 lbs.	100	100	100	-	50	50	100	50	99
Market Lambs Over 105 lbs.	100	100	100	-	100	100	100	100	100
Market Sheep	100	100	100	-	100	100	100	100	100
Total Sheep and Lambs	99	88	96	0	70	68	99	62	90
Out of State Sheep	100	-	100	-	-	-	100	-	100
Lamb Crop	100	100	100	-	79	79	100	79	96
Ewes Expected to Lamb	100	100	100	-	40	40	100	40	98

APPENDIX 3.6 - JULY SHEEP REPORT QUESTIONNAIRE

The following July 1999 questionnaire is a condensed version that displays all variables used in the AGGIES evaluation. Variables that are state specific are identified as such. Each cell has a number and a set of letters indicating the key item codes and SAS variable names, respectively.

2.	a. **Ewes** 1 year old and older?	+	281 lshpewes
	b. **Rams** 1 year old and older?	+	282 lshprams
	c. **Replacement Lambs** under 1 year old?	+	285 lshprepl
3.	a. (1) **Market Lambs** under 65 pounds?	+	836 lshpu065
	(2) **Market Lambs** 65 to 84 pounds?	+	837 lshp6584
	(3) **Market Lambs** 85 to 105 pounds?	+	838 lshp8505
	(4) **Market Lambs** Over 105 pounds?	+	839 lshpo105
	b. **Market Sheep** 1 year old and older?	+	287 lshpfeed
4.	Then the **Total Sheep and Lambs** on hand July 1 was:	=	280 lshptotl
5.	(CA, CO, and WY) How many **head were in another State**?		385 lshpotst
6.	How many **Lambs Dropped** from January 1, 1999 through June 30, 1999 were or will be Marked, Dropped, or Branded?		288 lshpcrop
7.	Of the Ewes on the total acres operated on July 1, how many are **Expected to Lamb** between July 1 and December 31, 1999?		289 lshpeexp

APPENDIX 4 - EDITING AND IMPUTATION ACCURACY INDICES DETAILS

For each variable that was used to evaluate AGGIES, the below contingency table was produced (Manzari and Della Rocca, 1999).

		AGGIES Output Data	
		Changed	**Not Changed**
Survey Production Data	**Modified**	$a = a_s + a_f$	b
	Unmodified	$c = c_s + c_f$	d

Where:

modified = survey production data that does not equal the reported data
unmodified = survey production data that equals the reported data
changed = AGGIES output data that does not equal the reported data
not changed = AGGIES output data that equals the reported data
a = number of modified data identified to be changed in AGGIES
a_s = number of modified data identified to be changed by AGGIES and imputation was successful
a_f = number of modified data identified to be changed by AGGIES and imputation failed
b = number of modified data identified not to be changed by AGGIES
c = number of unmodified data identified to be changed by AGGIES
c_s = number of unmodified data identified to be changed by AGGIES and imputation was successful
c_f = number of unmodified data identified to be changed by AGGIES and imputation failed
d = number of unmodified data identified not to be changed by AGGIES

Using the counts from the contingency table, nine accuracy indices were calculated for each variable. The following table supplies the formula for each index (Manzari and Della Rocca, 1999).

Assessing ...	**Index**	**Calculation**
Editing Quality	I1	$d/(c+d)$
	I2	$a/(a+b)$
	I3	$(a+d)/(a+b+c+d)$
Imputation Quality	I4	c_s/c
	I5	a_s/a
	I6	$(a_s+c_s)/(a+c)$
Overall Editing and Imputation Quality	I7	$(c_s+d)/(c+d)$
	I8	$a_s/(a+b)$
	I9	$(a_s+c_s+d)/(a+b+c+d)$

APPENDIX 5 - IMPUTATION ESTIMATOR OPTIONS AND FORMULAS

The following table displays all the imputation estimator options currently available in AGGIES and supplies a formula for each.

Imputation Estimator	Formula	Where...
current mean	$Y_{it} = \overline{Y}_t$	Y = variable to be imputed X = auxiliary variable i = the unit/report t = current survey period $(t - 1)$ = historical survey period
current ratio	$Y_{it} = \frac{\overline{Y}_t}{\overline{X}_t} X_{it}$	
previous value	$Y_{it} = Y_{i(t-1)}$	
previous mean	$Y_{it} = \overline{Y}_{(t-1)}$	
auxiliary trend	$Y_{it} = \frac{X_{it}}{X_{i(t-1)}} Y_{i(t-1)}$	
difference trend	$Y_{it} = \frac{\overline{Y}_t}{\overline{Y}_{(t-1)}} Y_{i(t-1)}$	