United States Department of Agriculture

National Agricultural Statistics Service

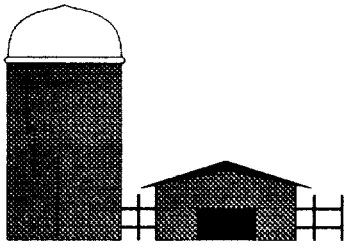
Research and Applications Division

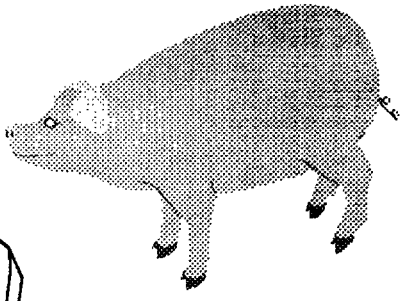
SRB Research Report Number SRB-91-10

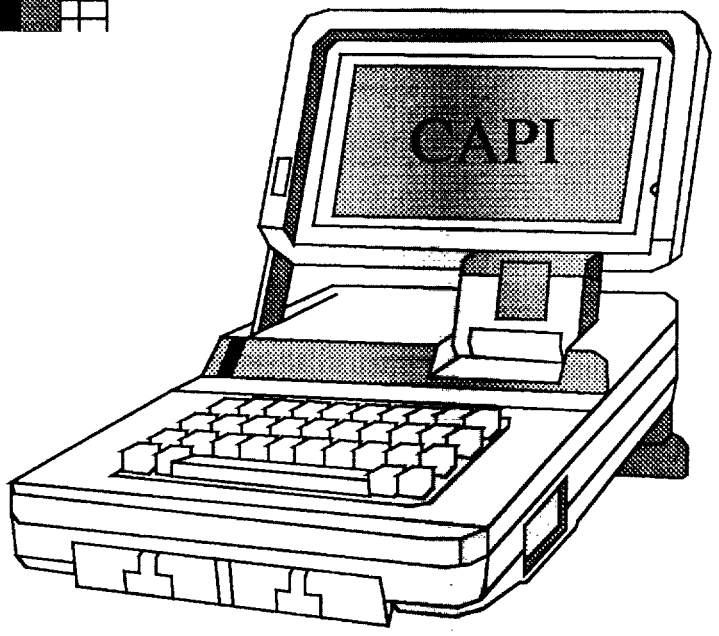
November 1991

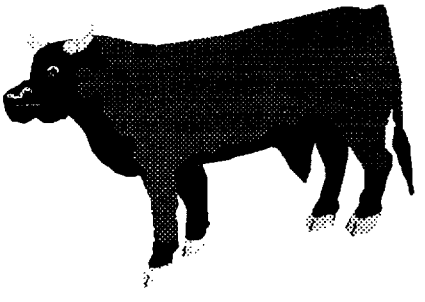
MOBILE COMPUTERIZED DATA ENTRY AND EDIT FOR THE LIVESTOCK PRICES RECEIVED SURVEY

Bruce Eklund









MOBILE COMPUTERIZED DATA ENTRY AND EDIT FOR THE LIVESTOCK PRICES RECEIVED SURVEY, by Bruce Eklund, Research and Applications Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250. November, 1991. NASS Research Report No. SRB 91-10.

ABSTRACT

The initial pilot research project for the National Agricultural Statistics Service (NASS) mobile computerized data entry and edit (McDEE) began in portions of Indiana and Ohio in May,1988 for a monthly survey of prices paid to farmers for livestock. Laptop computers replaced paper questionnaires and enumerators entered and edited data interactively. Post-collection batch edit checks indicate that McDEE has reduced error rates. Consequently, in addition to the improvement in data quality, statisticians spend less time editing the data. Also, McDEE has very substantially reduced office key punching. It is recommended that this McDEE application be made operational in two States and then be evaluated and expanded from that point of reference.

KEY WORDS

Computer assisted personal interview, CAPI, interactive edit, McDEE, survey instrument.

ACKNOWLEDGEMENTS

The author would like to acknowledge the major contributions of Mark Harris, Robert Hale, and Martin Holko to the project. Mark wrote the original survey instrument. Robert enhanced the instrument and he and Martin implemented the pilot project. Rita Hohenbrink later helped support the project.

Jamie Hodges, Doug Boline, and the Indiana State Statistical Office as a whole have been helpful in support of the project. The office has used McDEE successfully and has expanded it to use computers to collect most of the data for the Livestock Prices survey.

Carol Bodkyn willingly accepted the challenge of being one of the first NASS enumerators to collect data with McDEE. She has displayed competence, a good attitude, and perseverance in her work.

The author appreciates the encouragement and management support of George Hanuschak, Brian Carney, and Jim Cotter throughout the project. Stan Mason contributed to the project with time-saving hardware and software support. Virginia McBride reviewed this article carefully and with aptitude. Amy Curkendall prepared tables and used her skill with desk top publishing software to enhance this document.

TABLE OF CONTENTS

SUMMARY

In1988 the Indiana and Ohio State Statistical Offices (SSOs) were selected for pilot McDEE tests for a portion of their Livestock Prices Survey. Initial feedback varied widely. The Indiana SSO strongly supported continued use of the technology. The Ohio SSO and the Economic Statistics Branch believed that the project was premature and wished to delay survey work until there was further developments in technology and software.

The Indiana SSO has expanded the use of the technology for their Livestock Prices survey. The computer assisted data entry and edit survey instrument was written with CASES software. The survey does not consist of a personal interview. The enumerator samples transactions of livestock sales at a firm and enters the data into a laptop computer instead of a paper questionnaire. The software makes validation checks as the enumerator enters data. The data are later sent from the enumerator's home to the SSO by telecommunications. Enumerators send a backup copy by mail on a diskette.

Both the data collected electronically and the data collected by paper are subject to a batch edit check at the SSO. The data collected by computer had lower error rates. McDEE has reduced the work load for office keypunchers and reduced time required for statisticians to review flagged errors from the batch edits.

The Survey Research Branch recommends that McDEE for livestock prices be transferred from research and be used full scale and operationally in Indiana and at least one other SSO and be supported by operational headquarters units. In addition, other pilot level research studies should be conducted on surveys such as the Farm Cost and Return Survey, the Agricultural Survey, objective yield surveys, enumerator time sheets and quality control checks in these SSOs.

INTRODUCTION

NASS first used computer assisted telephone interviews (CATI) to collect and edit survey data in 1981. NASS now uses CATI for dozens of surveys and is expanding use to some 42 SSOs. Since several of NASS's major surveys include extensive personal enumeration, a natural extension of CATI is McDEE.

The intent of this report is to relate early NASS McDEE experience to the potential of McDEE at NASS, to enable other McDEE developers to learn from this experience including the mistakes, and to map a strategy for NASS McDEE implementation.

METHODS

Cost/Benefit Ratios

The implementation of McDEE and computer assisted personal interviews (CAPI) on a full scale operational basis would entail large capital expenditures for hardware and a well planned survey management and processing system. Three major cost/benefit factors made it prudent to first develop McDEE (and CAPI) at NASS on a small scale.

- The price of hardware for McDEE (and CAPI) has continually dropped while hardware capability has increased.

- Any initial mistakes made on a small scale are less costly in monetary terms, in the possible inadvertent alteration of data, and in creating enumerator apprehension.

- The more McDEE (and CAPI) survey applications available, the lower the cost/benefit ratio. The strategy is to first develop several applications on a small scale, then purchase hardware to use on a larger scale when the survey program of the Agency fully supports a cost effective implementation.

Moreover, before NASS expands McDEE use to a large scale, it should include the support and input of numerous operational survey specialists. Hence, the philosophy has been to research and test several diverse McDEE survey applications, including livestock prices, on a small scale.

Statistical Evaluation of McDEE

Does McDEE increase data quality? Does McDEE reduce the overall processing time for a survey? How does McDEE affect interview time, enumerator-respondent communication, and enumerator morale? Does McDEE effect the respondent's perception of data confidentiality pledges? How does McDEE effect total survey costs? If McDEE produces different survey estimates than paper and pencil (P&P) surveys, which is closer to "the truth?" How should one adjust time series P&P or McDEE data? Although much has been learned about these issues, it is unreasonable to expect

to answer all of these questions from pilot McDEE projects. One must first determine if McDEE is feasible for NASS, and if so, how NASS should best implement McDEE. Then NASS can better address the aforementioned issues. If one attempts to measure differences in data in the early stages of McDEE, one will confound differences between McDEE data and data collected by other methods with differences between what McDEE *is* and what McDEE *should* be. The purpose of these initial projects is to determine feasibility and to learn how to best develop McDEE.

Hardware

In late 1987 the Survey Research Branch (SRB) purchased three Zenith ZWL-183 laptop personal computers for the livestock prices project. Each has a 10 megabyte hard drive, a floppy drive, 80C88 CPU (XT), an internal 1200/2400 baud modem, and a liquid crystal display screen. Each weighs about 15 pounds.

Much lighter, more advanced, and less expensive laptops were soon available. With limited budgets, it is unwise to try to always have the most advanced hardware. Instead, one should focus on developing *software* for surveys and learn from practical field experience. With the development of McDEE capabilities, NASS will be best prepared to make larger scale hardware purchases shortly before large scale use.

Software

The survey instrument (data collection tool) for the livestock prices was coded using CASES software. NASS began work with the Computer-assisted Survey Methods (CSM) Program of the University of California at Berkeley to collect survey data by CATI with CASES software in 1981. NASS has benefited from a long and close working relationship with CSM whose staff teach CASES programing as well as support utilities and specific CATI instruments. CSM has also created survey management applications, often with NASS needs in mind. The NASS CATI Section supports many surveys and is expanding CATI to about 42 of the 45 NASS SSOs.

The Survey

NASS samples livestock sold for slaughter sales transactions between farmers and agribusiness. Data include number of head, total animal weight, and total price received for cattle, hogs, and sheep. Enumerators visit firms to sample the previous month's slaughter data. They mail completed paper questionnaires to the SSO where keypunchers input the data and statisticians edit the data via a batch edit.

McDEE

Although the Prices survey was chosen as a pilot project for CAPIat NASS, it is not truly a CAPI application. Although enumerators use laptop computers and a survey instrument to collect and edit data, there are no personal interviews. McDEE is a more accurate acronym for this survey, and CAPI is a specific type of McDEE. NASS has several potential McDEE applications that would also be CAPI applications (objective yield surveys are an exception to this.) McDEE for the

Livestock Prices survey has many CAPI attributes. Field enumerators collect and edit survey data from a remote site. They send data to a central location by telecommunications and/or by mail on a diskette. Experience gleaned from McDEE will help NASS develop CAPI.

The Instruments

The survey instruments were coded with CASES software. There is an instrument for each of the three animal types: cattle, hogs, and sheep. Enumerators access the instruments from a computer screen menu. They verify that the month and that the case (interview) code that they entered matches the name of the firm. Enumerators enter the number of transactions for the firm and the instrument generates a random start for sampling. They can convert dressed weights to live weights automatically. After each record, total head, total animal weight, and dollars received are displayed for verification. McDEE invokes data validation range checks for average animal weight and price per head after the enumerator enters and hopefully verifies each transaction (one data record). The enumerator must respond to McDEE validation checks to continue data collection. See Appendix A for illustrations and further description of the instrument.

Training

One day was allocated for training enumerators in both the Indiana and Ohio SSOs. The enumerators had no previous computer experience. They had more difficulty learning to change answers than any other facet of McDEE. Whether the enumerators or the instrument caught an error, the enumerators had to learn to use commands to move and change answers.

The trainers spent the day after training with the enumerators in the field, collecting operational data. Most data collection went smoothly. The trainers/programmers discovered, however, that the enumerator must be able to exit and then reenter a case to add new data. One firm required that the enumerator drive a long distance between two auction barns to complete the case.

One enumerator collected data inside a car for one case. The computer used power from the car's battery through an adapter for the cigarette lighter. This was functional but awkward as the enumerator handled the computer and the paper receipts. The following day enumerators collected data alone. One enumerator crammed and jammed a diskette incorrectly into the laptop, highlighting the importance of thorough training for both McDEE and for general laptop use.

The field work gave the trainers ideas of how to improve the instrument, data transmission, and data backup. The experience reinforced the concept that developers should pretest the survey instrument to appreciate the enumerators' perspective. Developers made software improvements as a result of the field work.

Length of Data Collection Sessions

Although this project was designed as a feasibility study and not for statistical comparison, McDEE interview length was measured. The computer's internal clock could have been used to measure "interview" starting and ending times. This method would underestimate collection time for McDEE in comparison to P&P because it would exclude time spent to set up the computer and other necessary activities. Instead, both McDEE and P&P enumerators were given forms and

explicit instructions to record times of arrival and departure at each firm. Several months into the project, the author observed an enumerator use McDEE without following the instructions. The enumerator did not enter the time upon leaving and returning to the car. At the end of the work day the enumerator would estimate the time spent at each firm.

It is likely that error introduced by enumerator estimation of time is greater than the difference in time between McDEE and P&P. Thus, the data comparison is not presented here. However, enumerator comments and plotting of the data indicated that the length of the McDEE survey approached that of P&P survey length after the second month. A better method may be to measure survey time with the computer and add an estimate of additional time to start the McDEE survey.

This experience points to a McDEE research issue. One must be especially careful about assumptions when testing in a somewhat uncontrolled environment. It is easy to deceive one self and make erroneous assumptions with limited direct observation of data collection.

Error Rates

Both McDEE and P&P data are subject to a batch edit executed at the SSO. The McDEE data are also subject to the aforementioned checks at the time of collection. NASS uses its own batch edit software, the Generalized Edit (GE). NASS defines GE edit checks as either critical or non-critical. The statistician must address all critical errors before the data summary. Critical errors pertain to problems such as missing items that are necessary in order to use the data record or data with *unreasonable* values. Non-critical errors are merely flags for suspicious data; the statistician in the SSO may investigate these suspicious data values but need not change them. Non-critical errors usually include data or data relationships that are in *suspicious* ranges.

Both critical and non-critical error rates were tracked for McDEE and P&P data. Each month, the GE output included data for the number and type of errors for each firm but not total records sampled. Hence, statisticians in the SSO ran a separate summary of tickets sampled by firm by month with a frequency procedure written in SAS, a statistical software package. Error counts came from the monthly electronic capture of GE output in Headquarters. They were input into a SAS summary program to counts error types and source (P&P or McDEE).

RESULTS

This was intended to be a temporary pilot level research project. After five months, critiques were solicited from both SSOs and the Economic Statistics Branch. The Ohio SSO and the Economic Statistics Branch wished to discontinue the project because of initial frustration in field use. The McDEE concept was termed "good," but not ready for field use. The Indiana SSO requested to continue and expand the project. They appreciated the lightened load for key punchers and for statisticians. McDEE collected data required less time to address data problems flagged by the office batch edit. The California SSO has also started to use McDEE for the survey.

Enumerator comments were favorable and their suggestions for further improvements were valid: they wanted more training, laptops were too heavy, and screen visibility was inadequate. (The computer industry has been responding to the latter two concerns.) The developers improved the instrument during the early days of McDEE based on the enumerator comments and based on a better understanding of the survey.

The following tables compare error counts and error sources for P&P and McDEE for the same period. All data are subject to an office, batch edit before summary. McDEE data are first subject to the critical level edit checks as the enumerator enters each record. The error types coded 700 or greater are critical; the others are non-critical.

Livestock Prices Received Error Counts

Error Code	P&P Error Count	P&P Error Rate (%)	McDEE Error Count	McDEE Error Rate (%)
		-- Non Critical Errors --		
2	3	0.03	0	-
4	151	1.34	114	0.97
5	252	2.24	139	1.18
6	1	0.01	0	-
99	2	0.02	0	-
101	1	0.01	0	-
141	7	0.06	0	-
151	3	0.03	0	-
		-- Critical Errors --		
700	3	0.03	0	-
701	0	-	1	0.01
704	15	0.13	0	-
705	19	0.17	1	0.01

	For P&P	
n ~ 11274	non-critical error pct. = 3.74%	critical error pct. =.33%
	For McDEE	
n ~ 11790	non-critical error pct. = 2.15%	critical error pct. =.02%

McDEE virtually eliminated critical errors. McDEE flags potential critical errors that exceed reasonable ranges in the field. The other major source of critical errors, missing data, is unlikely to occur because the instrument requires one to answer key questions to continue data collection. McDEE does not flag non-critical "errors". However, the instrument prompts enumerators to review each record (data from one ticket) before continuing. Also, non-critical errors may be valid data so one would not expect as a pronounced difference as with critical errors.

The difference between McDEE and P&P error rates may have been slightly greater. Originally P&P and McDEE data were distinguishable by the firm associated with the data, which indicated the enumerator, which indicated the mode of data collection. Occasionally, however, a mishap would cause P&P to replace McDEE. For instance, a laptop was once temporarily out of commission after an observing employee spilled coffee on the keyboard. One would need to keep track of such occurrences to avoid overestimating errors collected via McDEE. A survey code was later added to distinguish collection mode automatically.

The following from the Prices Received Manual describe the errors flagged during the batch edit.

ERROR CODE	DESCRIPTION
2	Number of head not reported.
4	Average weight outside limits.
5	Price per pound outside limits.
6	Data in All Hogs is equal to the sum of barrows and gilts and sows live weight and/or dollars.
101	Cow average live weight less than 500 pounds.
141	Steer &Heifer average live weight less than 500 pounds.
151	Calf average live weight greater than 500 pounds.
700	Live weight and/or dollars missing.
701	Percent from producers or sub-sampling interval missing.
703	Percent from producer and/or sub-sampling interval with ID other than page "00", line "00".
704	Unreasonable average live weight.
705	Unreasonable average price per pound.

See Appendix B for critical and non-critical edit limits.

CONCLUSIONS

The purposes of the pilot project was for NASS to gain experience and to determine technical feasibility. The answer to the question of technical feasibility is an emphatic yes. Cost and management feasibility remain as future challenges for McDEE or CAPI full scale implementations in NASS.

Non-computer specialists can program survey instruments for NASS surveys to collect and edit data interactively in the field. Enumerators have shown the aptitude to learn to use computers and to learn to use the survey instruments to collect field data and conduct interviews. Most of the enumerators had no previous computer experience. Some were wary at first but felt at ease with practice. However, the SSOs selected most of the enumerators for the pilot projects because they were capable. It remains to be seen if the entire field enumerator work force can and is willing to use computers to collect data. Those who have learned so far have learned remarkably well.

Enumerator training was (and will be) a learning experience for trainers. Early experience shows that enumerators should have ample time for hands on training. In general, allowing enumerators to use the computer and survey instrument is a better use of time then having them passively listen to lectures or view demonstrations. Individual attention is important while the enumerator is first using the computer. Developers/trainers should provide enumerators documentation to use in training and to keep for reference.

The data can be securely transmitted to the office and then combined and summarized with data collected by other methods. NASS should not, however, rely solely upon telecommunications but incorporate backup procedures, preferably diskette mailing.

Interview time was not measured precisely. It appears that collection time approaches P&P collection time after the enumerator becomes familiar with McDEE. Time differences will likely vary by survey. For instance, McDEE will help save time for surveys for which it can replace enumerators' arithmetic calculations.

McDEE has shown the ability to reduce nonsampling errors.

The development of McDEE to its full potential will require input from experts who are responsible for each aspect of the survey. Those responsible for survey design, methodology, administration, data editing, and summary should experience McDEE in relation to their work and then offer suggestions to strengthen McDEE. After they have contributed to McDEE development, NASS will be best equipped to make hardware and software choices for a large scale.

The Survey Research Branch should focus on:

1) New McDEE and CAPI applications and enhancements.
2) Research, such as the effect of McDEE and CAPI on nonsampling error.

The Survey Research Branch should also work cooperatively with those who use McDEE. However, McDEE will not develop as well if support is solely from within the Survey Research Branch. Too much expertise will be lost if McDEE development is confined to the Survey Research Branch.

RECOMMENDATIONS

Large scale implementation of McDEE will require fairly large capital expenditures. NASS should first have a wide range of operational experience using McDEE on a small scale. This will prepare NASS to make the best software and hardware decisions to move McDEE Agency-wide when it is cost beneficial to do so.

The Survey Research Branch recommends that McDEE for livestock prices be transferred from research and be used full scale and operationally in Indiana and at least one other SSO and be supported by operational headquarters units. In addition, other pilot level research studies should be conducted on surveys such as the Farm Cost and Return Survey, the Agricultural Survey, objective yield surveys, enumerator time sheets and quality control checks in these SSOs.

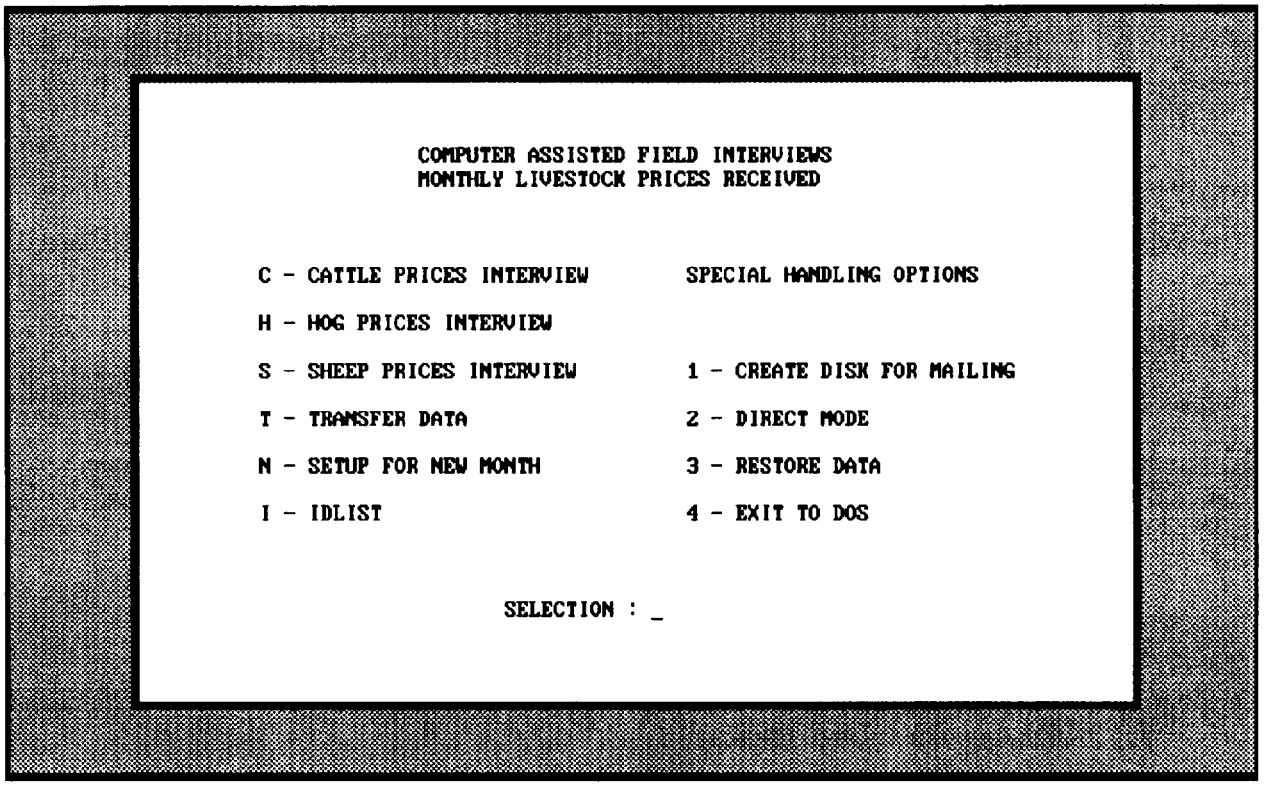
Selected SSOs should use, and Headquarters should support, multiple McDEE applications, including the livestock prices received survey. It is essential that operational personnel now contribute for optimal McDEE development. The Survey Research Branch should assist this effort but should focus on new survey applications and research issues. As NASS develops more applications it can then use each computer to collect more data, and the cost/benefits ratio will continue to improve.

Appendix A: Livestock Prices Received Software

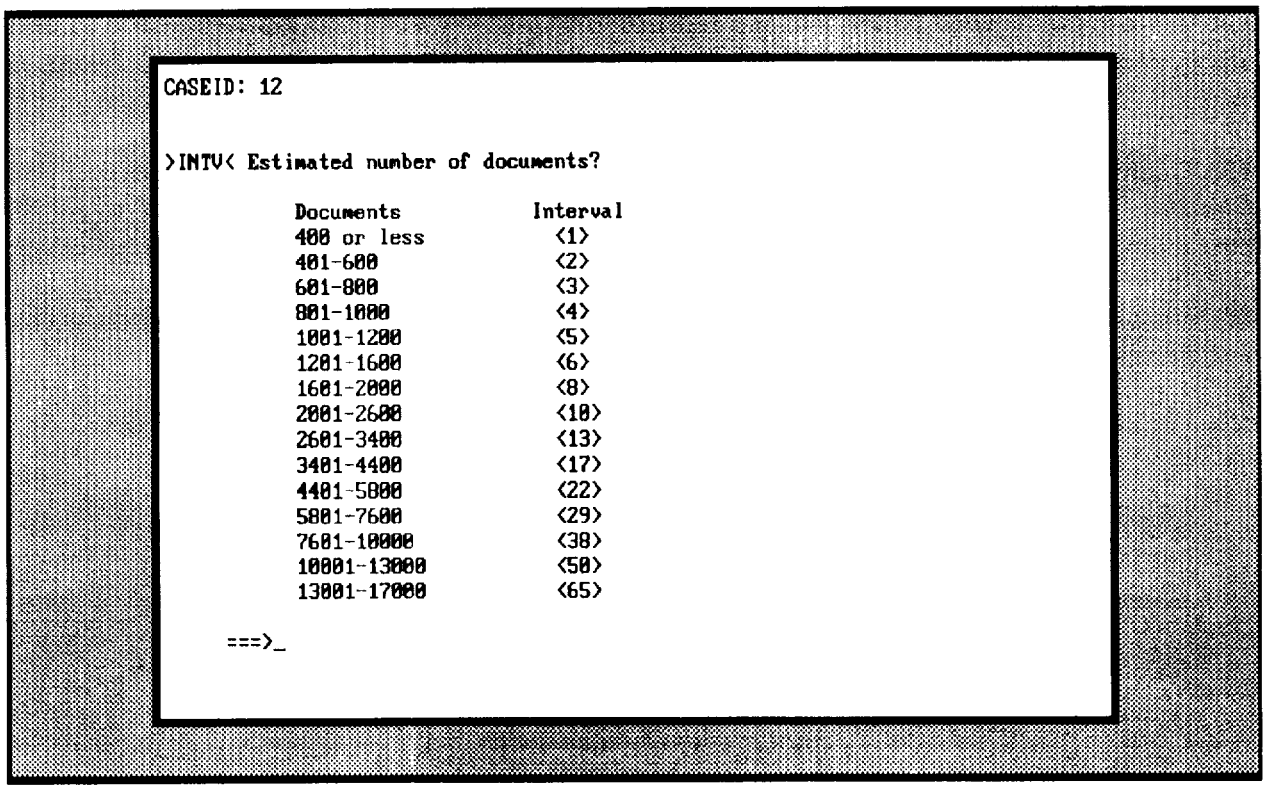
The menu appears when the enumerator turns on the computer and after the completion of each menu selection. The enumerator selects an instrument from three animal types: cattle, hogs, or sheep.

The "Transfer Data" option is for automated telecommunications. This option also copies data onto diskettes. "Setup for New Month" executes CASES commands used to prepare the survey for the next month. "IDlist," for enumerator reference, displays and matches firm names and case codes.

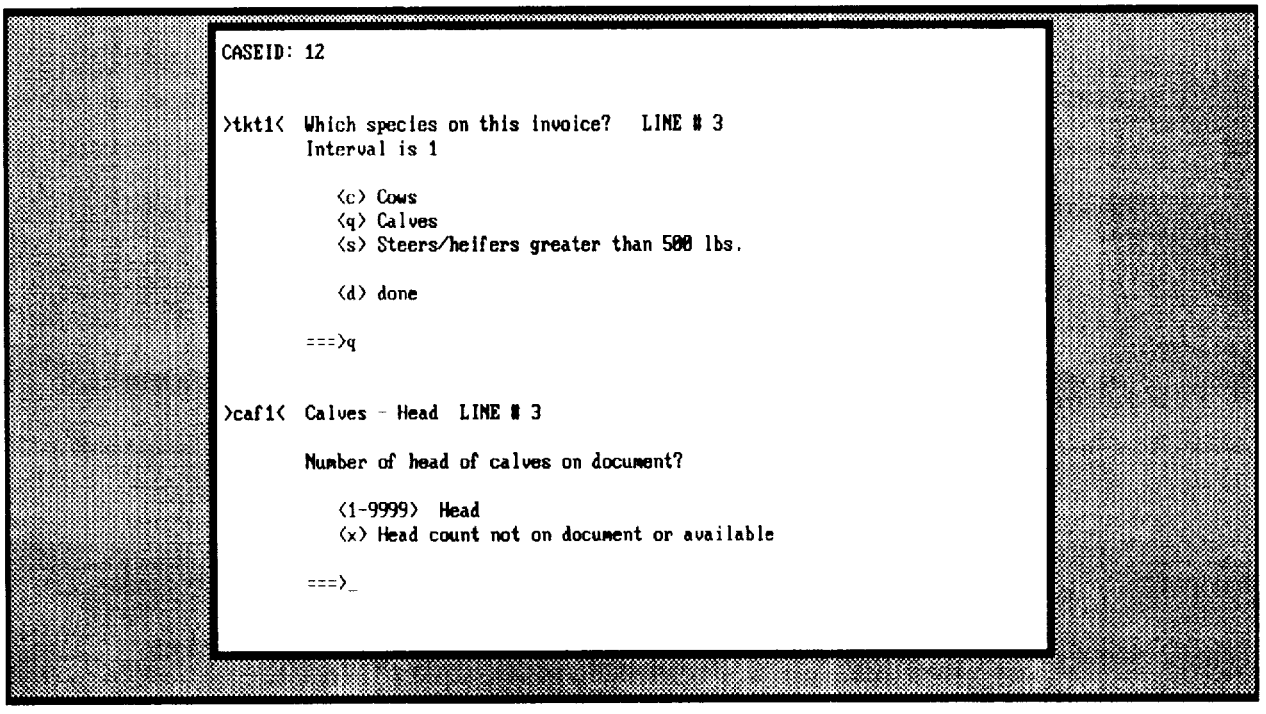
Additional options perform diskette backups while bypassing telecommunications, remote access into the enumerator's computer, the reentering of closed cases, and entry into DOS.



The enumerator counts or estimates the total number of livestock transactions for the month, enters the corresponding interval, and is then given a random starting number.



The four questions on the following two screens are for data collected for one transaction (record): animal type, number of head, total weight, and total price.



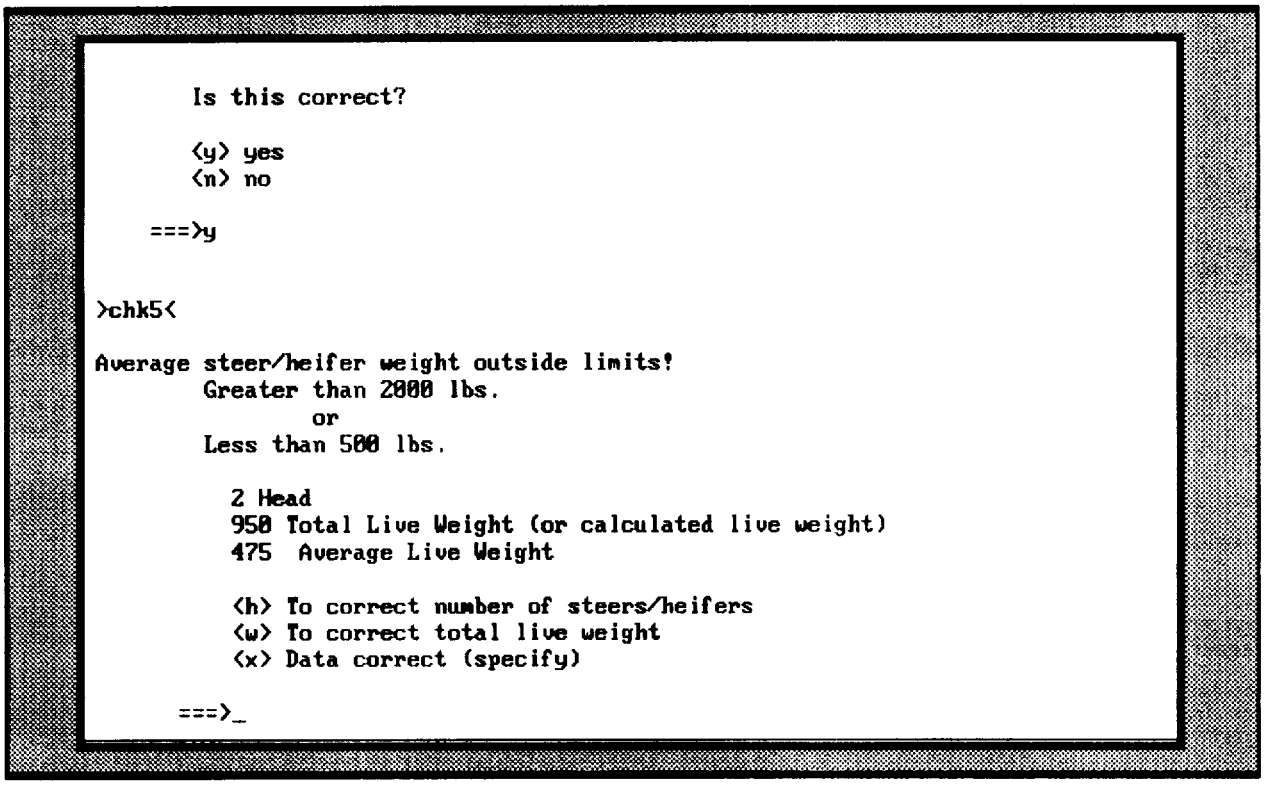
```
        Number of head of calves on document?

           <1-9999>  Head
           <x> Head count not on document or available

        ===>3


>caf2<  Calves - Total Live Weight   LINE # 3

         Total live weight on document

           <1-500000> Pounds
           <x> To enter Carcass or Dressed Weight

        ===>1125


>caf3<  Calves - Gross to Producers   LINE # 3

        Total dollars received for the 3 head, 1,125 pounds.

           <1-999999999> Dollars

        ===>844_
```

As the enumerator enters each record but before edit checks are invoked, the instrument asks the enumerator to verify the data. These data are valid.

```
CASEID: 12


>prt2<
       You have entered
       3 calves
       1,125 total pounds
       844 total dollars

       Is this correct?

       <y> yes
       <n> no

     ===>y_
```

The enumerator should inspect the data to catch errors that may not be caught by the automated edit. Thus average animal weight and average price per pound are checked *after* the enumerator verifies that the data are correct.

In another example, where average animal weight was outside the error limits, the instrument gives the options to move and change the number of head or the total live weight. If the enumerator still feels that the data are correct then he or she must include a comment or explanation. In this instance it is likely that the animal type was incorrect. Perhaps calves were classified as heifers. If so, the enumerator would move backward and correct the animal type rather than selecting one of the choices shown: h, w, or x.



Appendix B: Edit Limits for the Livestock Prices Received Survey

The following are the critical and non-critical edit parameters for the GE batch edit for the Prices Survey. NASS subjects all the Prices data, irrespective of collection mode, to the GE edit before summary. McDEE data are also subject to the critical level edit checks as the enumerator enters each record.

Average Animal Weight (per transaction)

	Lower Edit Limits (lbs./animal)		Upper Edit Limits (lbs./animal)	
	McDEE & GE Batch Critical	GE Batch Non-Critical	GE Batch Non-Critical	McDEE & GE Batch Critical
Cows	500	600	2000	2500
Steers & Heifers	500	500	1600	2000
Calves	20	50	500	500
Sows	125	180	850	1500
Barrows & Guilts	20	150	395	450
All Hogs	20	150	850	1500
Lambs	10	20	150	200
Sheep	40	60	225	300

Average Price

	Lower Edit Limits ($/lbs.)		Upper Edit Limits ($/lbs.)	
	McDEE & GE Batch Critical	GE Batch Non-Critical	GE Batch Non-Critical	McDEE & GE Batch Critical
Cows	.10	.25	.70	1.50
Steers & Heifers	.20	.40	1.00	1.50
Calves	.10	.40	1.75	3.00
Sows	.05	.25	.65	1.00
Barrows & Guilts	.10	.35	.75	1.00
All Hogs	.10	.25	.75	1.00
Lambs	.05	.35	1.05	2.00
Sheep	.03	.08	.40	3.00

* U.S. G.P.O.:1991-311-404:40095/NASS