United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Applications
Division

SRB Research Report
Number SRB-90-01

October 1990

A STATISTICAL EDIT FOR LIVESTOCK SLAUGHTER DATA

Cathy Mazur

A STATISTICAL EDIT FOR LIVESTOCK SLAUGHTER DATA, by Cathy Mazur, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250, 1990, Research Report Number SRB 90-01.

ABSTRACT

The main goal of this research project was to create a statistical edit for livestock slaughter data, by utilizing a plant's historic data to define edit limits for that plant. Classical and simple robust estimators were considered in the analysis, but a more complex robust estimator known as Tukey's biweight was selected to calculate edit limits. There are several other potential agency data series that would benefit from this or similar statistical edit techniques.

KEYWORDS

Livestock Slaughter Data, Tukey's Biweight, Robust Estimation, Outlier, Inlier, Double Root Residual.

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

ACKNOWLEDGEMENTS

The author would like to express her thanks to several people who contributed to the success of this project. Richard Allen, Robert Tortora and George Hanuschak initiated the project. Bernie McCullough (who was the subject matter specialist in Livestock Section) contributed in its development by working with me on all phases of the project. Ben Klugh (formerly, the Head of Estimates Research Section) encouraged and directed me during the research. George Patton (Processing Section) helped me access the proper data and obtain cost figures. Mitch Graham (formerly in the Support Systems Development Section) is working on the implementation of the statistical edit. Charles Perry helped me with several of the graphs. Lastly, the author would like to thank other members of Survey Research Branch and Livestock Branch for their support.

Table of Contents

Table of Contents

Tables

Graphs

SUMMARY

A census of federally inspected livestock slaughter plants occurs each week using a one page mail questionnaire. U. S. Department of Agriculture livestock slaughter inspectors in the plants report daily numbers of head killed, and weekly weight totals. Several limitations in the old edit prompted this research project. The main goal of this project was to create a statistical edit, unique for each plant, by utilizing a plant's historic data to define edit limits for that plant.

Using 64 weeks of data in five states from 116 plants, classical and simple robust estimators were considered in the initial analysis. This analysis consisted of 4 measures of location and 4 measures of spread over 4 time periods. Since none of these measures seemed appropriate, more complex robust estimators were investigated. Tukey's biweight was considered in the subsequent analysis, because of its many interesting properties. The biweight was selected to calculate edit limits, as it worked well in varying data distributions. The resulting statistical edit is now fully operational.

This report covers cases where outliers (that is, values which are far from the main group of data) and inliers (for the sake of this paper, suspicious values in the middle of the data) are found. In both, a question is raised on whether the value comes from the same population as the remaining values, or whether some measurement or reporting error occurred. For example, outliers may be extremely high or low reported weights, whereas inliers may be a series of weights which are not extreme, but which do not change, or change very little over time. The report also includes general and cost saving features of the new edit. Lastly, some additional methodologic research is recommended.

With regards to the statistical edit technique, there are several other data series which are potential candidates, as they collect data from the same units over time. These are Poultry Slaughter, Turkey Hatchery, Manufactured Dairy Products, Peanut Stocks, Off-Farm Grain Stocks, several data series in the monthly Eggs Chickens and Turkeys report, Cold Storage, Cattle on Feed, and large farm extreme operators in probability based surveys.

INTRODUCTION

A cooperative program currently exists between the Food Safety and Inspection Service (FSIS), National Agricultural Statistics Service (NASS), and Agricultural Marketing Service (AMS) for collecting livestock slaughter data in federally inspected plants. The joint effort involves data collection, editing, summarization, and public dissemination of data. In addition to being published, data on the number of head are currently used by NASS as check data for the Quarterly Agriculture Survey (QAS). Current livestock estimates are validated by adding births and subtracting deaths from the previous survey's livestock figure in a balance sheet approach.

In the old edit system, data were entered on a personal computer using a software package called KeyEntry III and uploaded at the end of the day to a leased mainframe. The data were then edited using a Generalized Edit System (a parameter driven program run in batch mode). The results of the edit were available several hours later at a higher cost, or the next morning at a lower cost. Analysts pored over printouts in order to resolve errors, and corrections had to be rekeyed on the personal computer. An outlier for head data was a value which differed more than a given percent from the plant's previous 3 week average (calculated using positive and zero kill days), and an outlier for weight data was a value which was outside some predetermined weight range for each class of livestock.

One problem with this edit was that some head data values were incorrectly identified as outliers during holiday weeks. A reason for this is that the old edit did not take the zero kill days into account. Therefore, plants which did not kill the same number of days each week were not being edited reasonably. For example, a plant which normally slaughters Monday through Friday, but misses Thursday and Friday due to a Thanksgiving holiday, would probably show a change from the previous 3 week average. A second problem is that plants slaughtering specialty weight animals were incorrectly flagged as errors. For example, a plant which normally slaughters veal calves would tend to report lower weights than a plant slaughtering normal weight calves. The reason for this is that the same edit limits were used for all plants. These as well as other problems compelled Livestock Branch (who runs the survey in NASS) to request improved editing techniques for Livestock Slaughter data.

Consequently, a research project was initiated to develop specifications for a statistical edit for Livestock Slaughter data, by utilizing each plant's historic data. The problem with head data during holidays could be solved by basing the edit only on positive kill days using a robust estimator, so that the values reported, and not a holiday would cause the error. Plants which slaughter 5 or 6 days a week provide enough positive data in a few weeks to calculate a daily average, but plants which only slaughter 1 or 2 days a week would not supply enough data. Therefore, more weeks of data must be used in these plants. A way to handle the problem with specialty weight plants is to use a statistical approach, by editing each plant based on that plant's historic data. In addition, plants with a lower coefficient of variation in head counts and weights would be edited more accurately, checks on a plant's weekly slaughter pattern (days of the week with positive head counts) would be made, and the edit would take place interactively on the personal computer.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

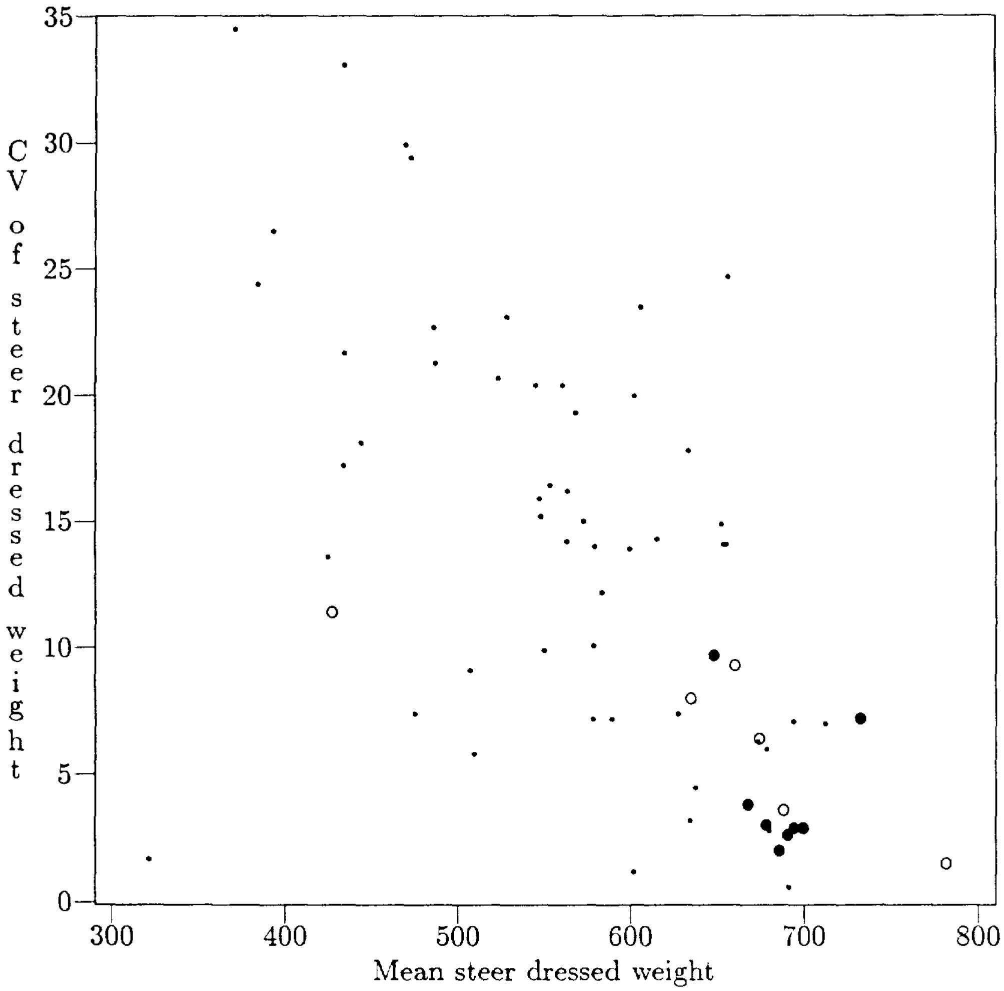
TABLE 1

1987 PLANT SUMMARY BY SIZE (NUMBER OF HEAD IN PLANT) & SPECIES

Species	Size	% Plants	% Head	Base of %
Cattle	1 - 1,000	68.9	0.8	
	1,000 - 9,999	15.4	2.0	
	10,000 - 49,999	7.1	6.5	# Plants = 1,317
	50,000 - 99,999	3.3	9.3	# Head = 34,004,000
	100,000 - 249,999	2.8	17.8	
	250,000 - 499,999	1.1	14.0	
	500,000+	**1.4**	**49.6**	
Calves	1 - 100	69.8	1.1	
	100 - 999	9.9	1.1	# Plants = 686
	1,000 - 9,999	12.1	6.9	# Head = 2,644,000
	10,000+	**8.2**	**90.9**	
Hogs	1 - 1,000	68.6	0.4	
	1,000 - 9,999	16.6	0.8	
	10,000 - 99,999	7.4	3.9	# Plants = 1,182
	100,000 - 249,999	1.1	3.1	# Head = 78,127,000
	250,000 - 499,999	1.1	7.0	
	500,000 - 999,999	2.1	16.1	
	1,000,000 - 1,499,999	1.1	12.9	
	1,500,000	**2.0**	**55.8**	
Sheep	1 - 100	72.1	0.4	
	100 - 999	18.9	1.0	# Plants = 906
	1,000 - 9,999	6.6	3.4	# Head = 5,002,000
	10,000+	**2.4**	**95.2**	

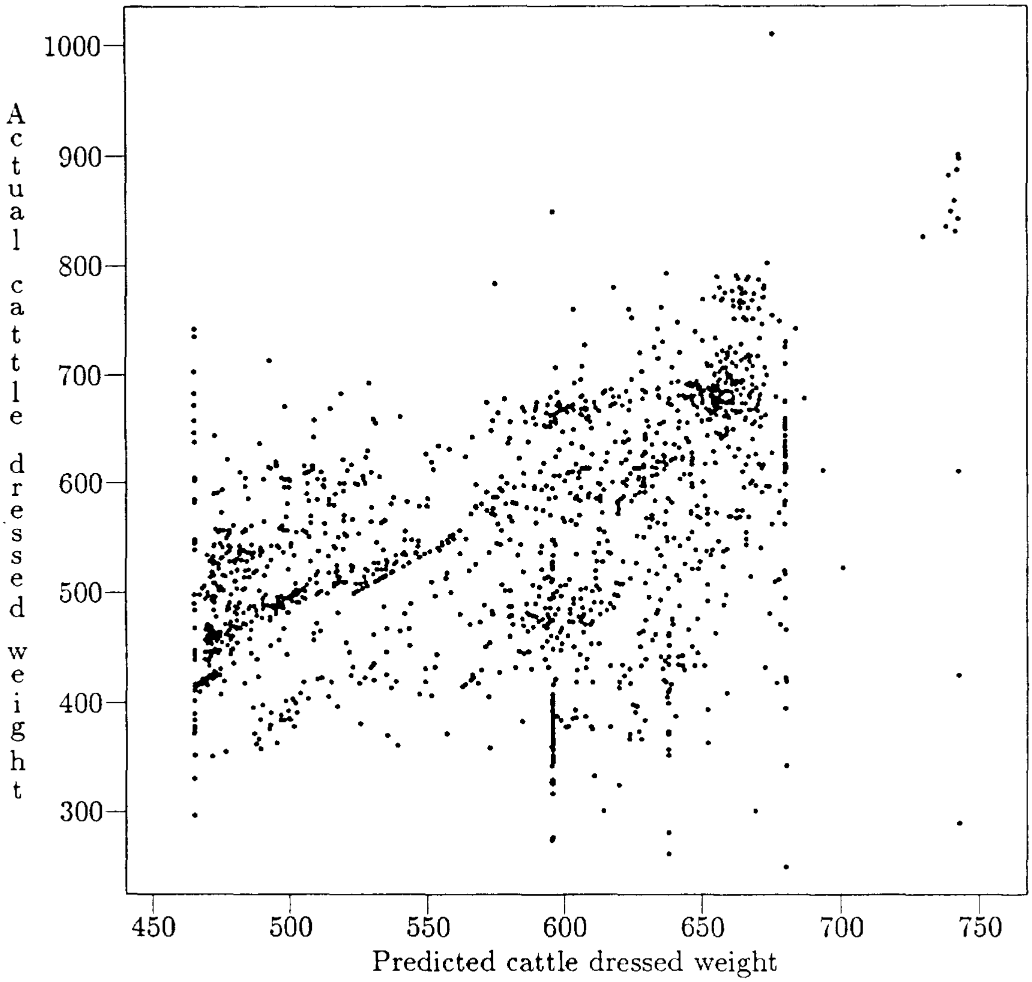
Graph 1. CV of Steer Dressed Weight *vs.* Mean Steer Dressed Weight

(For Small(.), Medium(○) and Large(●) Plants)



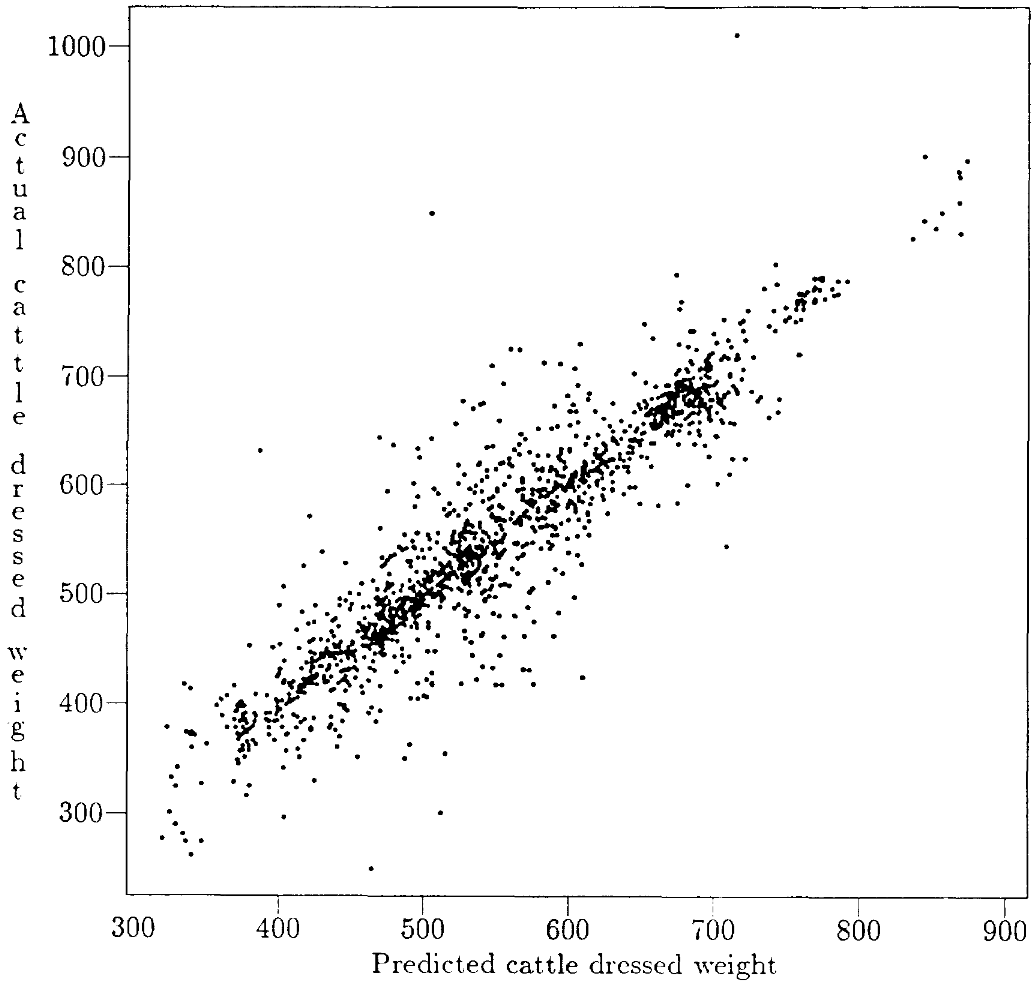
Graph 2. Actual *vs.* Predicted Cattle Dressed Weight

(Using First Half Universal Means)



Graph 3. Actual *vs.* Predicted Cattle Dressed Weight

(Using Biweight Means)



GENERAL METHODOLOGY

The purpose of the livestock edit is to determine whether plant data is reasonable (that is, to check for reporting errors and keying mistakes). Errors can be in the form of outliers (extreme values) as well as inliers (suspicious values in the middle of the data which do not change, or change very little over time). The manner in which these will be identified will be discussed below.

1. Identification of Outliers

The first step in constructing a statistical edit was to determine which statistical estimator to use to define the edit limits. The goal was to choose a measure of center and spread that would quickly stabilize to new levels when true changes did occur in the data, or return to old levels in the presence of outliers.

As to time series models, a few data sets were analyzed using a time series analysis package. However, the resulting model only incorporated the plant mean. The exponentially weighted moving average (EWMA) was also considered, but a robust method was desired, and the data did not seem correlated enough to support the exponentially weighted method. A plot of weekly average steer dressed weights by size group showed no obvious trend, and a visual examination of many time series plots for steer dressed weights showed many plants with no trend, and the plants with a trend were large ones. Lastly, as the edit would be unique within a plant, there was less concern that one plant's increase was due to some universal increase. Since only 64 weeks of data were available, research on time series models and any seasonality effect was postponed.

Robust measures were considered, as they are resistant to outliers; whereas, the standard statistical method (mean and standard deviation) works best only in the Normal distribution and is affected by outliers. A robust method is one which is insensitive to underlying assumptions, or in simpler terms, one which is best in a broad range of situations, rather than one particular situation.

In the **initial analysis**, classical types and simple robust measures were examined. This analysis included 4 measures of center and 4 measures of spread over 4 time periods (6, 10, 13, and 26 weeks). Weight data were used to evaluate these measures, where the definitions follow.

Measure of Center

a) Mean - sum all values (X_i) and divide by n (the number of values).

$$\overline{X} = \sum_{i=1}^{n} X_i / n$$

b) Median, $M = X_{[m]}$, where

$$X_{[m]} = \begin{cases} X_{[(n+1)/2]} & \text{if n is odd} \\ (X_{[n/2]} + X_{[(n/2)+1]})/2 & \text{if n is even} \end{cases}$$

$X_{[k]}$ = the k^{th} order statistic, i.e., $X_{[1]}$ is the smallest, and $X_{[n]}$ is the largest of the n observations.

c) Trimean, T1 = (Q1 +2M +Q3)/4 , where Q1 and Q3 are the 25th and 75th percentiles. If m is noninteger, then drop the decimal part, and keep only the integer part of m.

$$Q1 = \begin{cases} X_{[(m+1)/2]} & \text{if m is odd} \\ (X_{[m/2]} + X_{[(m/2)+1]})/2 & \text{if m is even} \end{cases}$$

$$Q3 = \begin{cases} X_{[n-((m-1)/2)]} & \text{if m is odd} \\ (X_{[n-(m/2)]} + X_{[n-(m/2)+1]})/2 & \text{if m is even} \end{cases}$$

d) 20% Trimmed Mean (T2) - the lowest n*0.20 values and the highest n*0.20 values are dropped, then T2 is the mean of the center n*0.6 values.

Measure of Spread

e) Standard Deviation (SD) - sum the squares of the deviations of each value from the mean, and divide by n-1 (one less than the number of values).

$$SD = \sqrt{\Sigma(X_i - \overline{X})^2/(n-1)}$$

f) Inter-Quartile Range, IQ = Q3 - Q1.

g) Median Absolute Difference (MAD) - tranform each value by subtracting the median (M) and taking the absolute values. Then obtain the new median of the transformed values.
$MAD = \text{median } \{|X_i - M|\}$

h) 20% Trimmed Standard Deviation (TSD) is the standard deviation of the center n*0.60 values.

In the analysis, the 4 measures of center and spread were calculated using **all** sample cases for the 4 time periods, using the appropriate number of data values prior to the current week. For example, using a 6 week time period, X_{t-6} through X_{t-1} were used in calculating the statistical measures of variable X for time period t. The performance of these estimators in a variety of data situations were observed. Hoaglin (1983, pgs. 325-332) also compares these measures and provides several statistical results (such as the variance and efficiency of the estimators).

Several conclusions were made with regards to the measures of center. When outliers were present, the mean changed considerably, as all values (reasonable and unreasonable) were included. The trimean was dropped early in the analysis, as the mean, median, and 20% trimmed mean seemed sufficient. The median and 20% trimmed mean were inadequate as good values were being excluded (e.g. the upper and lower 20% in the trimmed mean, and all but the center values in the median).

To summarize the results obtained when all sample data was considered, a representative data set from one mid-sized steer plant is used. This data set, shown in Table 2, consists of average steer dressed weights over time with several outliers. The first value is labeled as week 1, but it really represents one week in a long time series. Therefore, the 13 values prior to that week with positive data were used to calculate the corresponding measures of center. A visual comparison of these measures of center is shown in Graphs 4a (with the Actual DW) and 4b (without the Actual DW), where outliers (outside the biweight prediction interval in Graph 7) are represented by an '*'. The actual data (solid line) is the average steer dressed weight for a week. The measures of center are close, but the mean does tend to lag a bit.

As to the measures of spread, the standard deviation is also greatly affected by outliers, as it includes reasonable and unreasonable values. The 20% TSD, the IQ range, and the MAD (although robust) are also inadequate due to the exclusion of good data. The 20% TSD excludes the upper and lower 20%, the IQ range includes only the 25th and 75th percentiles, and the MAD only looks at 50% of the data.

Note in Table 2, the standard deviation increases drastically due to the outliers in weeks 6 and 7. In fact, these outliers may cause the system to miss the outliers in weeks 15 and 16 (by being inside the Lsd-Usd prediction interval in Graph 7), since the prior 13 values are used. A visual comparison of the measures of spread are displayed in Graph 5. The IQ range and MAD were normalized by dividing by the corresponding value for the "standard" normal distribution (1.349 and 0.6745) to enable comparison with the SD (represented as SIQ and SMAD). Although the SIQ, SMAD, and TSD are not nearly as affected by the outliers, the concern is that they may underestimate the measure of spread.

These measures of center and spread can be characterized in several ways. The following table lists some of these - the number of values used to calculate the estimate, the weights assigned, whether the estimate is affected by inliers (i.e., changes in the middle of the distribution) or outliers, and whether reasonable data are being excluded.

Estimate	# Values	Weights	Affected by Inliers	Affected by Outliers	Excludes Good Data	Comments
Mean	N	1/N	No	Yes	No	Affected by outliers
Median	1 or 2	1.0 or 0.5	Yes	No	Yes	Susceptible to grouping/rounding
Trimean	3	0.5 or 0.25	Yes	No	Yes	
20% T2	0.6*N	1/(0.6*N)	No	No	Yes	
SD	N	1/N	No	Yes	No	Affected by outliers
IQ Range	2	0.5	No	No	Yes	May underestimate the measure of spread; 25% breakdown bound
MAD	2	0.5	Yes	No	Yes	May underestimate the measure of spread; problem with clusters
20% TSD	0.6*N	1/(0.6*N)	No	No	Yes	May underestimate the measure of spread

As to the number of values used in the calculations, the 6 and 10 week time periods provided unstable measures of spread. The 26 week period required too much data (half a year), and took longer to detect changes. In Table 3, the mean and standard deviation are calculated using the four time periods for our one example. The 6 week SD ranges from 19 to 153, and the outlier at week 6

TABLE 2

CALCULATION OF DIFFERENT MEASURES OF CENTER AND SPREAD USING INITIAL MEASURES FOR STEER DW (13 weeks)

Wk.	AvDw	Mn	Md	TrMn	SD	SIQ	SMAD	TSD
1	628	656	657	655	26.52	14.08	19.27	7.42
2	732	653	652	651	26.74	15.57	16.31	8.41
3	684	654	652	651	29.53	15.57	16.31	8.41
4	623	659	657	655	28.20	14.08	17.79	7.42
5	638	657	657	654	29.68	17.05	19.27	9.09
6	**332**	656	657	652	30.14	18.53	25.20	10.84
7	**787**	632	657	649	94.94	25.95	28.17	14.18
8	660	643	659	654	104.30	35.58	28.17	16.67
9	659	643	659	654	104.29	35.58	40.03	16.62
10	659	642	659	652	103.95	25.95	37.06	14.32
11	668	641	659	652	103.88	23.72	37.06	13.93
12	651	642	659	653	104.04	29.65	37.06	14.97
13	644	644	659	656	103.83	22.24	31.13	10.80
14	654	643	659	654	103.76	22.24	31.13	11.59
15	**852**	645	659	657	103.69	17.79	22.24	8.89
16	**852**	655	659	657	116.58	17.79	22.24	8.89
17	651	668	659	662	128.77	17.79	22.24	21.66
18	645	670	659	663	128.19	12.60	13.34	20.96
19	667	670	659	663	128.06	12.60	13.34	20.96
20	644	696	659	666	78.39	12.60	11.86	20.65
21	652	685	659	657	74.50	11.86	11.86	5.92

Notes:

1. AvDw refers to unweighted average dressed weight of Steers. See page 3 for a discussion of average dressed weight, and Appendix 2 for a discussion of the weighted approach.

2. Mn, Md, and TrMn refer to the Mean, Median, and 20% Trimmed Mean, respectively.

3. SD is the Standard Deviation, and TSD is the 20% Trimmed Standard Deviation.

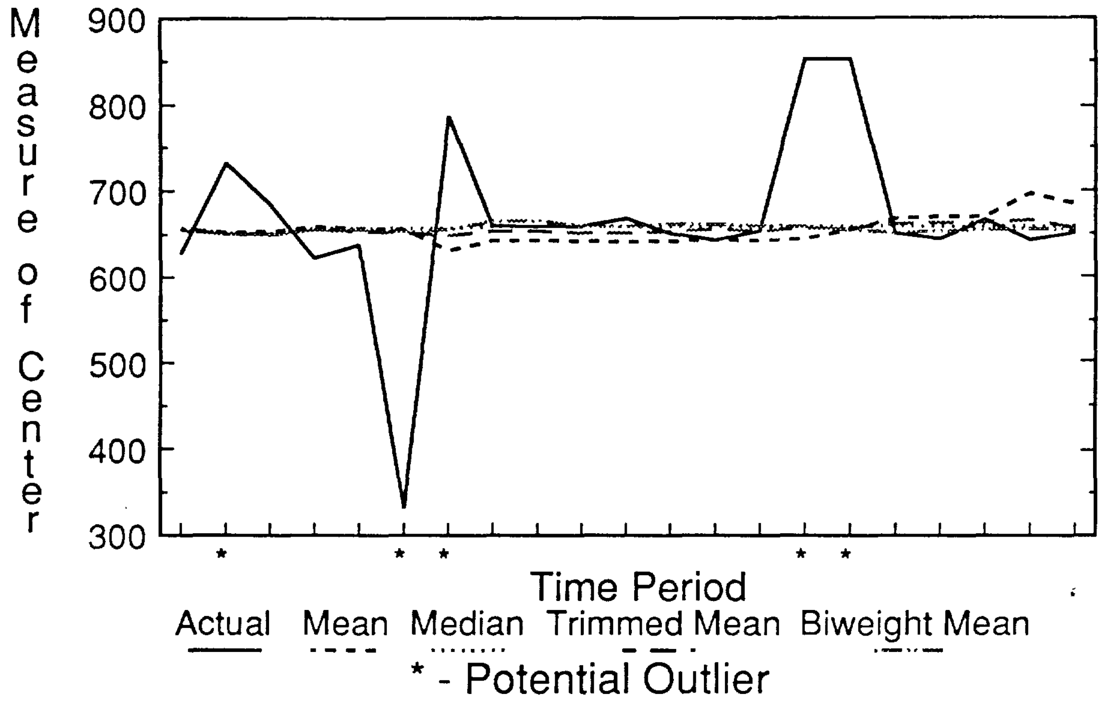
4. SIQ is the Standardized Interquartile Range (IQ/1.349), and SMAD is the Standardized Median Absolute Difference (MAD/0.6745). See page 10.

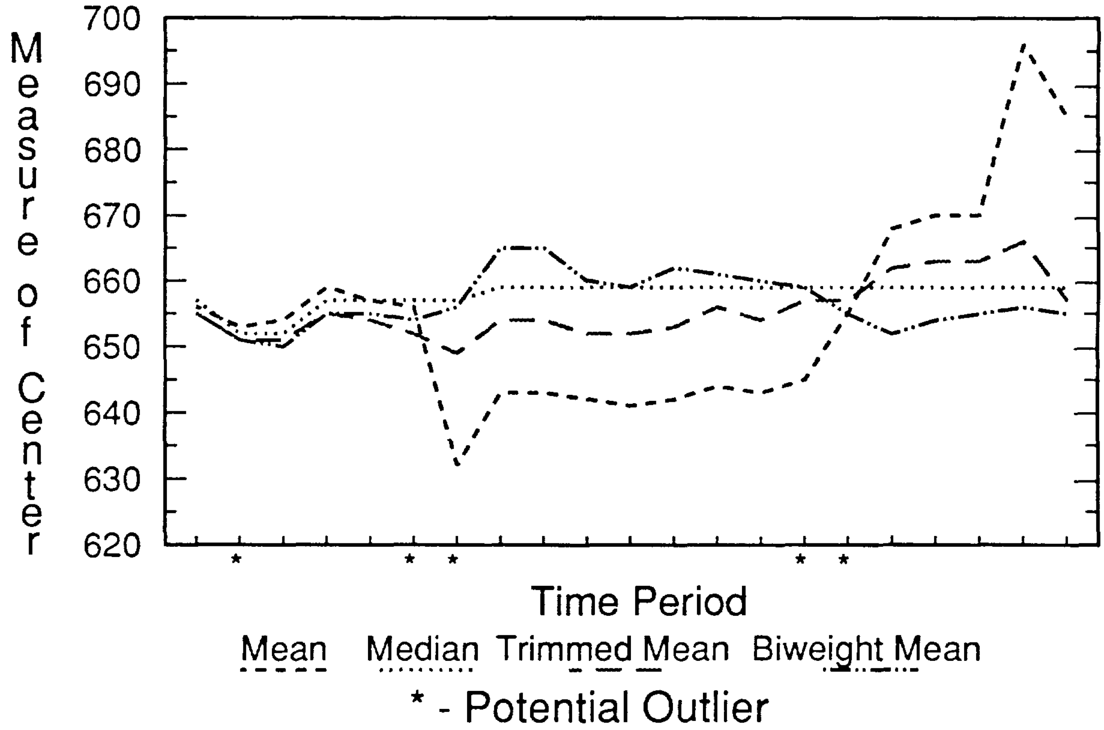
5. The measures of center and spread are calculated by using the previous 13 week's positive average dressed weights. For example, the values of AvDw for Week 1 through 13 are used to calculate the measures shown for Week 14.

6. The old edit for steer uses 250-900 pounds for plants with less than 100 head, 375-800 for 100-500 head, and 500-800 for over 500 head.

GRAPHS 4a and 4b

PLOT OF 4 MEASURES OF CENTER FOR STEER DW OVER TIME (13 weeks)
- with and without the Actual DW





GRAPH 5

PLOT OF 5 MEASURES OF SPREAD FOR STEER DW OVER TIME (13 weeks)

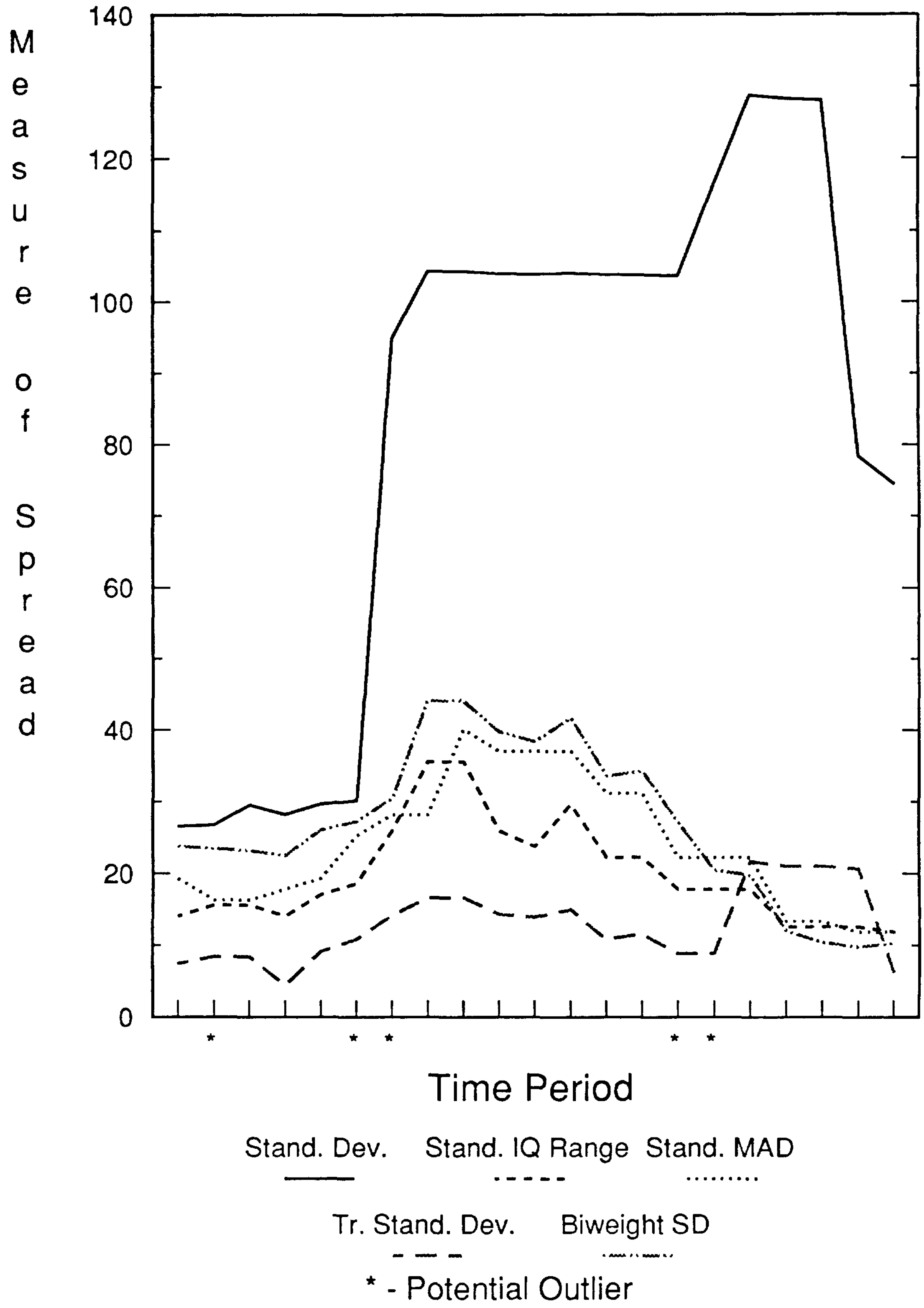


TABLE 3

CALCULATION OF THE MEAN AND STANDARD DEVIATION USING DIFFERENT TIME PERIODS (6, 10, 13, 26 weeks) FOR STEER DW

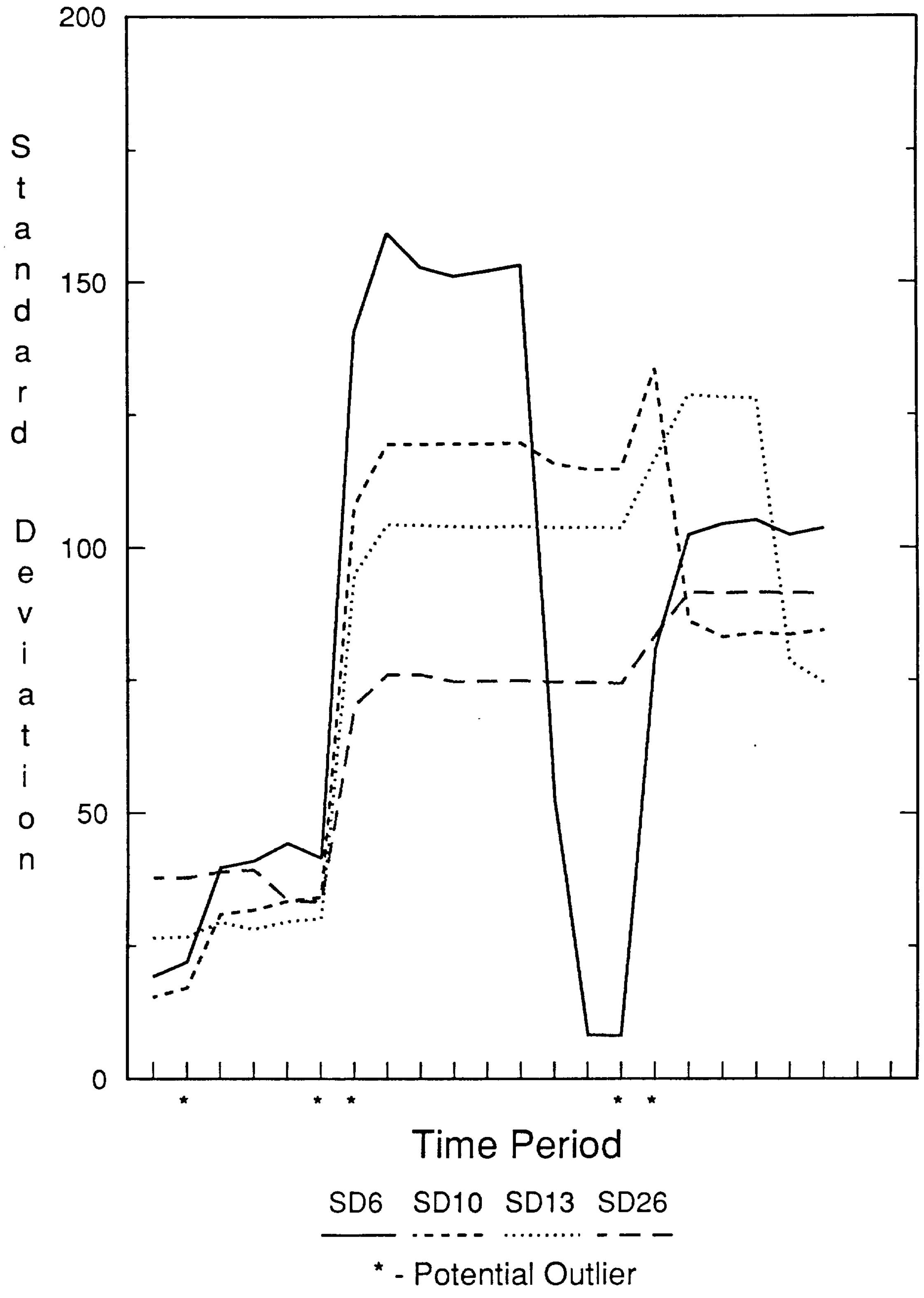
Wk	AvDw	Mn6	Mn10	Mn13	Mn26	Sd6	Sd10	Sd13	Sd26
1	628	656	652	656	644	19.25	15.47	26.52	37.81
2	732	650	650	653	644	22.00	17.14	26.74	37.75
3	684	660	658	654	650	39.77	30.99	29.53	38.93
4	623	663	662	659	652	41.02	31.81	28.20	39.26
5	638	657	660	657	647	44.26	33.52	29.68	33.70
6	**332**	660	658	656	648	41.58	34.24	30.14	33.30
7	**787**	606	624	632	637	140.59	107.81	94.94	70.14
8	660	633	636	643	642	159.28	119.37	104.30	75.97
9	659	621	636	643	643	152.88	119.39	104.29	76.02
10	659	616	640	642	647	151.14	119.42	103.95	74.70
11	668	622	640	641	647	152.16	119.46	103.88	74.74
12	651	628	644	642	648	153.26	119.67	104.04	74.84
13	644	681	636	644	649	52.37	115.75	103.83	74.58
14	654	657	632	643	650	8.28	114.59	103.76	74.49
15	**852**	656	635	645	649	8.18	114.74	103.69	74.28
16	**852**	688	657	655	654	80.74	133.71	116.58	83.32
17	651	720	709	668	663	102.42	86.07	128.77	91.44
18	645	717	695	670	664	104.36	83.00	128.19	91.39
19	667	716	694	670	663	105.15	83.83	128.06	91.43
20	644	720	694	696	664	102.37	83.50	78.39	91.40
21	652	718	693	685	664	103.74	84.34	74.50	91.36

Notes:

1. See notes 1, 5, and 6 from Table 2.
2. Mn6 through Mn26 refer to the Mean calculated using 6, 10, 13, and 26 weeks respectively.
3. Sd6 through Sd26 refer to the Standard Deviation calculated using 6, 10, 13, and 26 weeks respectively.

GRAPH 6

PLOT OF STANDARD DEVIATION (STEER DW) vs. TIME PERIOD



affects the SD for 6 weeks. The 10 and 13 week calculations peak at subsequently lower values, but the effect of the outlier is felt over more weeks. The 26 week SD is much more constant, with gradual (but minimal) increases due to the outliers. Graph 6 displays these trends.

The shortcomings discussed above were a motivation to do a literature search to find other estimators of center and spread. The mean and standard deviation are good estimators, but they are largely affected by outliers (i.e., they are not robust). The other estimators are not as affected by outliers, but they exclude good data, and thus may underestimate the measure of spread. The set of statistical measures from the first analysis are L estimators, or linear combinations of order statistics. One characteristic of these estimators is that the same weights are used for all data sets, that is, the weight is independent of the data set. For example, the median of any data set (where n is odd) is calculated by giving the center value a weight of one, and all other values a weight of zero. In the next analysis, we chose an estimate from the class of W and M estimators called Tukey's biweight. This class of estimators differs from L estimates in that weights differ for different data sets, that is, the weight is dependent on the data set.

In the **second analysis**, Tukey's biweight was calculated on a subset of sample cases using a 13 week time period (chosen as the best time period from the first analysis). Head data was used as well as weight data to evaluate this measure (using the number of whole weeks so that at least 13 positive values occurred).

The biweight mean (BiMn) and biweight standard deviation (BiSd) incorporate unequal weights, where reasonable values are given weights close to 1, and unreasonable values (outliers) are given very small weights or are excluded altogether (by giving a zero weight). The BiMn has the advantage of the mean if the data is normal (all good values are included), but has the advantage of the median if outliers are present (it excludes them).

In order to calculate the BiMn and BiSd, the weight which each X_i will receive must be calculated. The weight is a function of U_i, a standard distance measure. Therefore, the first step is to calculate U_i. U_i represents the measure of distance each X_i value is from the measure of center (M), in terms of some multiple "c" of the measure of spread (S). U_i is very small close to M, and gets larger the further away you get. By selecting M, S, and c, the user determines the point beyond which values are so far away that they are excluded from the calculation of BiMn and BiSd. This is the rejection point ("cS"), beyond which U_i is greater than 1 or less than -1.

$$U_i = \frac{X_i - M}{cS}$$

For this research, the median was chosen as the measure of center, and both the IQ range and the MAD were considered as measures of spread (S). The parameter "c" represents the number of measures of spread a value must be away from the measure of center before the value (X_i) is excluded entirely. Reasonable values of c are between 6 and 12. In this analysis, 6 and 9 were used.

$$U_i = \frac{X_i - M}{c * IQ} \quad \text{or} \quad \frac{X_i - M}{c * MAD}$$

Because of the way the MAD and the IQ range are calculated, a direct comparison can not be made with the SD. By normalizing MAD and the IQ range (dividing by 0.6745 and 1.349 respectively), the various rejection points tried can be put in the same units (number of SDs). To make this comparison, the c parameter must be multiplied by the same fraction. (See Hoaglin (1983, pp. 368))

6	MAD	≈	4.05	SD
9	MAD	≈	6.07	SD
6	IQ	≈	8.09	SD
9	IQ	≈	12.14	SD

The second step is to calculate the weight each X_i value receives as a function of the U_i. Values near the measure of center get the largest weight (close to 1). Any value which is too far away (beyond the rejection point) receives a zero weight.

$$\begin{cases} (1-U_i^2)^2 & \text{if } |U_i| \leq 1 \\ 0 & \text{if } |U_i| > 1 \end{cases}$$

The BiMn and BiSd are given by the formulas below, where values having a U_i greater than the absolute value of 1 are excluded from both calculations. One problem with the BiSd is that it is possible for a value to have a negative term in the denominator. (See Hoaglin (1983, pp. 397-8))

$$\text{BiMn} = \frac{\Sigma[X_i * (1 - U_i^2)^2]}{\Sigma(1 - U_i^2)^2}$$

$$\text{BiSd} = \frac{\sqrt{n * \Sigma[(X_i - M)^2 * (1 - U_i^2)^4]}}{|\Sigma[(1 - U_i^2) * (1 - (5 * U_i^2))]|}$$

Table 4 lists the BiMn and BiSd for the same data set used in Tables 2 and 3 (for comparison). Calculations using c=6 and c=9, in combination with IQ and MAD are shown. Notice that the BiSd is not drastically affected by the outliers. Graphs 4a, 4b and 5 (pages 12-13) display the BiMn and BiSD with the IQ range and c=6. In Graph 5, BiSd is larger than SIQ, SMAD, and TSD because the last three may underestimate the measure of spread. Hoaglin (1983,pgs. 390-394, 414) compares these measures and provides several statistical results (such as the variance and efficiency of the estimators).

The biweight has many interesting properties. It is flexible, yet computationally simple. The biweight has an iterative form (which falls in the class of M estimators), but a single step form is also available (which falls in the class of W estimators). In the iterative form, U_i is first calculated using the median as the measure of center (M) and the MAD or IQ range as the measure of spread (S) to calculate the BiMn and BiSd. In the second step, the U_i is recalculated using the BiMn from

TABLE 4

CALCULATION OF THE MEASURE OF CENTER AND MEASURE OF SPREAD USING THE BIWEIGHT - BiMn, BiSd (c=6,9 using IQ and MAD for 13 weeks)

Wk	AvDw	BiMn				BiSd			
		6d	9d	6q	9q	6d	9d	6q	9q
1	628	654	655	655	656	22.40	23.83	23.76	24.67
2	732	649	650	651	652	20.49	22.65	23.46	24.60
3	684	649	649	650	652	20.49	21.79	23.15	25.67
4	623	654	655	655	657	19.85	22.16	22.44	24.63
5	638	653	654	655	656	24.19	25.52	26.05	27.26
6	**332**	652	654	654	655	26.81	27.29	27.25	28.08
7	**787**	653	655	656	656	29.86	30.18	30.49	30.92
8	660	655	659	665	662	33.27	37.85	44.18	54.65
9	659	658	663	665	662	37.22	42.34	44.18	54.65
10	659	655	661	660	664	35.16	40.73	39.77	44.53
11	668	655	661	659	663	35.09	40.70	38.43	43.66
12	651	656	661	662	664	35.43	40.88	41.73	48.20
13	644	657	662	661	666	26.71	34.51	33.60	40.50
14	654	656	661	660	665	27.72	35.16	34.30	40.91
15	**852**	656	658	659	665	21.33	25.71	27.15	35.35
16	**852**	655	654	655	660	18.90	19.68	20.41	31.33
17	651	652	651	652	659	17.60	18.91	19.76	36.24
18	645	654	654	654	655	12.17	12.08	12.05	14.14
19	667	655	655	655	656	10.73	10.46	10.38	12.39
20	644	656	656	656	657	10.31	9.93	9.80	11.49
21	652	655	655	655	655	11.00	10.45	10.28	10.16

Notes:

1. See notes 1, 5, and 6 from Table 2.
2. BiMn and BiSd are the Biweight Mean and Biweight Standard Deviation using the equations on pages 17.
3. The subheadings 6d, 9d, 6q, and 9q refer to calculations using parameters c=6 or c=9, and the MAD (d) or IQ Range (q). See the equation for U_i and normalized distances on pages 16 and 17.

the first step as the measure of center and BiSd from the first step as the measure of spread to recalculate the new BiMn and BiSd. Also, the biweight takes into account the grouping and rounding effect. This is important since some plants report weights which fit a bimodal distribution (due to rounding of weights to the nearest 25 or 50 pounds for example) rather than a bell shaped distribution. In general, grouping and rounding refers to how changes in values near the center of the distribution can affect the estimator (e.g. the median). An example of this is duscussed below. The only assumptions for the biweight are that the distibution is symmetric about the center, and that the percent of outliers is less than 50 percent (see the discussion of the breakdown bound in Hoaglin (1983, pp. 357-8)). In symmetric distributions, the measures of center almost coincide (e.g. mean, median, BiMn). In skewed (or nonsymmetric) distributions, the measures of center differ. In this case, a bias must be considered, since the mean estimate of a target value and the target value do not correspond, due to systematic errors (Hoaglin (1983, pp. 287-9)).

The analysis identified two interesting cases. The first occurred in a data set (real data) with two similar size clusters. The second occurred in a data set (made up data, but could occur) with greater than 25% outliers. In both cases, the BiSd was much larger when the IQ range was used, than when the MAD was used. Table 5 contains examples of these two problems showing how the weight, BiMn and BiSd vary when IQ and MAD are used, and by choice of c parameter. The first data set consists of 13 values forming 2 clusters. The first cluster contains the median, which forces the U_i values for the second cluster to exceed the cutoff value of 1 and be ignored in the calculation of BiMn and BiSd (receive a weight of 0). In the second data set, the BiSd using the IQ Range is larger, as more than 25% outliers exist.

These two cases show how the value of BiSd varies by whether the IQ range or the MAD is used. Hoaglin prefers to use the MAD since the breakdown bound is higher (50%) than the IQ range (25%). The goal was to set up reasonable edit limits for livestock slaughter. From the subject matter point of view, cases like the first one (plants which report rounded weights) are a concern. Therefore, we decided to use the biweight with c=6 and IQ as the measure of spread. To account for the IQ range's lower breakdown bound, a test will be done for cases where the proportion of outliers is greater than 25 percent. If one is found, the MAD rather than the IQ range will be used.

Once the appropriate estimates were decided on, procedures for obtaining specific edit limits for livestock slaughter could be determined. These edit limits will provide a range, within which reasonable data values are expected. Any values outside this range will be flagged. This range will be formed by calculating the biweight prediction interval. For the livestock edit, the calculation of the prediction interval will require that the Coefficient of Variance be at least 1%. If not, the BiSd will be set to 1% of the BiMn. This will be done not for statistical reasons, but to set up "reasonable" edit limits for these plants (several cases were found where the BiSd was about 1 pound, and the BiMn was 650 or so pounds).

In calculating the biweight prediction interval, the t distribution with 0.7*(n-1) degrees of freedom and a 5% level of significance is recommended. However, the t parameter must be multiplied by an additional factor (given below) to account for sample sizes (number of values which go in to calculating the biweight) less than 20 . This factor was calculated by interpolation, using T_{bi} and $t_{.7(n-1)}$ for sample sizes 10 and 20 (see Hoaglin (1983, p. 423)). Therefore, the lower and upper

TABLE 5

CALCULATION OF THE BIWEIGHT'S INITIAL WEIGHT
(c=6,9 using IQand MAD for 13 weeks)

Wk	AvDw	Weight 6d	9d	6q	9q
Clusters					
1	8	0	0	0.971	0.987
2	25	0.945	0.976	0.999	0.999
3	25	0.945	0.976	0.999	0.999
4	25	0.945	0.976	0.999	0.999
5	26	1	1	1	1
6	26	1	1	1	1
7	26	1	1	1	1
8	26	1	1	1	1
9	42	0	0	0.977	0.995
10	50	0	0	0.949	0.977
11	50	0	0	0.949	0.977
12	52	0	0	0.941	0.973
13	52	0	0	0.941	0.973
BiMn		25.6	25.6	33.0	33.2
BiSd		0.91	0.90	15.61	15.50
>25% Outliers					
1	50	0	0	0.025	0.393
2	60	0	0	0.191	0.563
3	70	0	0.156	0.436	0.720
4	100	0.945	0.976	0.986	0.994
5	102	0.980	0.991	0.995	0.998
6	103	0.991	0.996	0.998	0.999
7	105	1	1	1	1
8	107	0.991	0.996	0.998	0.999
9	108	0.980	0.991	0.995	0.998
10	110	0.945	0.976	0.986	0.994
11	140	0	0.156	0.436	0.720
12	150	0	0	0.191	0.563
13	160	0	0	0.025	0.393
BiMn		105.0	105.0	105.0	105.0
BiSd		4.63	8.03	28.37	36.49

bounds of the biweight prediction interval are calculated as:

$$[BiMn-(t_{0.7(n-1)})(factor)(BiSd) \text{ , } BiMn+(t_{0.7(n-1)})(factor)(BiSd)].$$

The factor and the degrees of freedom used in finding the t parameter are in the following table.

n	factor	$df_{0.7(n-1)}$	n	factor	$df_{0.7(n-1)}$
13	1.071	8.4	17	1.044	11.2
14	1.068	9.1	18	1.036	11.9
15	1.063	9.8	19	1.023	12.6
16	1.055	10.5	20	1.009	13.3

As a further comparison, a prediction interval was calculated using 3 methods - the current method (see page 2), a prediction interval using the mean and standard deviation, and a prediction interval using BiMn and BiSd. For the current method, the same prediction interval was used for each plant, whereas in the other two methods, the prediction interval was based on each plant's past 13 positive weeks data. The problem with the current method is that specialty weight plants and holidays are not accounted for, and although the mean and standard deviation use each plant's historic data, the range is affected by outliers. As an example, Graph 7 displays the 3 prediction intervals calculated using the data from Table 2 and 4 for weight data using the unweighted approach. Two values were outside the limits using the current method [Lold,Uold], 2 values were outside the limits using the mean and standard deviation [Lsd,Usd], and 5 values were outside the limits using the biweight [Lbi,Ubi]. However, what is important is the type of outliers, rather than the number of outliers detected, and the idea that the biweight can be adjusted by the user.

Since each value for weight is an average (total weight divided by the number of animals for a week), a weighted approach will be used (see Appendix 2) to account for the different numbers of animals each week, rather than regarding the 13 averages as equal. For example, a plant which slaughters 500 steers for each of 12 weeks, and 50 for the 13th week might have a different average weight because of the fewer number of animals. Another way of looking at this is that each animal is assumed to have the average weight for that week. The biweight is then calculated on the total number of animals for the 13 weeks. Additional discussions of the biweight and its properties are found in Hoaglin (1983, ch. 9-12).

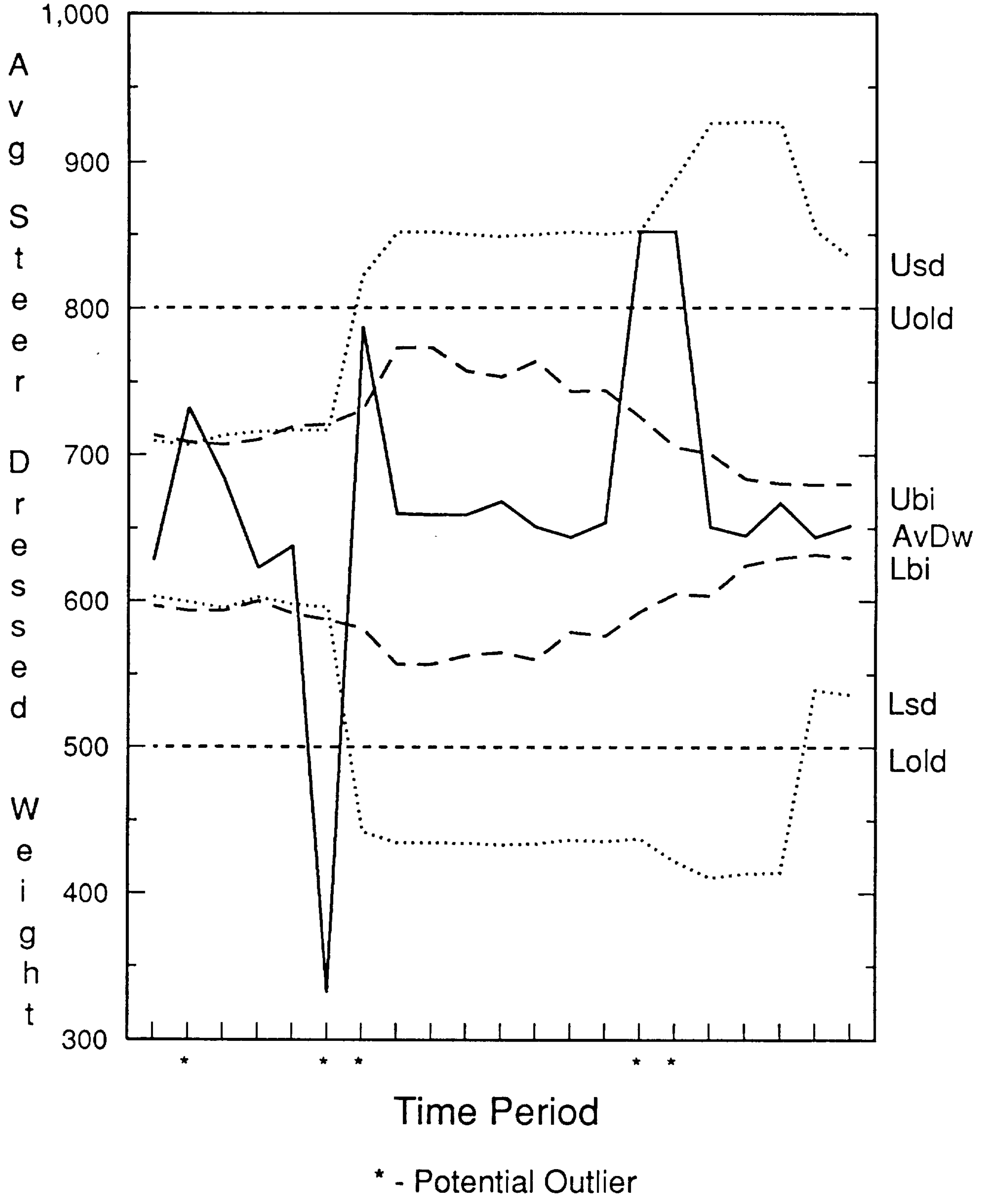
2. Identification of Inliers

An investigation of data from the previous analysis, showed that the weights for some plants do not change much over time. A few explanations for this are plants with the same imputed value over time, plants without the proper scales that report the same average weight over time, and plants with a low coefficient of variation. The goal was to determine a method for identifying these inliers in a distribution.

The Double Root Residual (DRR) measures how close the estimated and the observed values are each week. Typically, one expects a certain amount of variability between the estimated and observed values. By keeping track of this difference over time, inliers can be identified. However, the

Graph 7

PREDICTION INTERVALS on AVERAGE STEER DRESSED WEIGHT
(using current method, mean/standard deviation, and biweight)



DRR is a measure which assumes a Poisson distribution, that is, it considers the number of "successes" over a given time interval. Since weights assume a Normal distribution, a standardized residual should be used to compare estimated and predicted values. However, because of practical purposes (the amount of computations required) and because the DRR method works in an approximate way, the DRR will be used. That is, the DRR will be calculated each week for dressed weights and live weights, and the sum over time stored in SDDR (where 'w' is the week).

$$DRR_w = \sqrt{(2+(4*obs.))} - \sqrt{(1+(4*pred.))}$$

$$SDDR = \sum_w |DRR_w|$$

As an example, week 1 in Table 4 has an actual value (AvDw) of 628, and a predicted value (BiMn using 6q) of 655. Therefore DRR_1 is $\sqrt{2+(4*628)} - \sqrt{1+(4*655)}$, which equals negative 1.056.

A plant is flagged as an inlier when the SDRR is below a certain value, that is, the biweight (pred.) too closely predicts the observed value (obs.) over time. See Hoaglin (1981, pp. 265-6). For Livestock Slaughter, DW and LW will be checked in this way. Table 6 lists the SDRR for 15 plants, summed over 3 time periods. Plants 2 and 10 are highlighted, but investigation showed only minimal variability. Plants 13 and 15, however, are cases with the identical average DW over time (which questions their validity).

3. Use of Outlier/Inlier Detection Techniques in the New Edit System

Procedures have been written to incorporate the outlier and inlier detection techniques described above into the Livestock Slaughter Edit. These are described in detail in Appendix 3, and a summary is given below. Also, a discussion of results in practice is given on page 26.

i. Daily Head Data
In this section, the species daily values and the weekly totals will be edited. Also, holidays will be accounted for, and patterns in daily head kill will be checked.

ii. Average Dressed Weights
The species dressed weights will be edited, and inliers will be checked.

iii. Live Weights
Live weights for calves will be checked by comparing the current week's dressed weight with the historical proportion of DW to LW. However, live weight for cattle, hogs and sheep will be checked by using a regression equation for each class (where the class live weight is the dependent variable, the species dressed weights are the independent variables).

TABLE 6

CALCULATION OF THE DOUBLE ROOT RESIDUAL (summed over 3 time periods)

Plant	1st 15 wks	2nd 15 wks	30 wks	Comments
1	43.9	30.7	74.6	
2	**5.0**	10.9	15.9	Low variability.
3	27.3	5.1	32.4	
4	12.3	14.8	27.1	
5	57.3	48.4	105.7	
6	6.8	7.9	14.7	
7	30.2	35.1	65.3	
8	20.9	23.1	44.0	
9	6.8	7.7	14.5	
10	**3.2**	8.9	12.1	Low variability.
11	66.0	14.5	80.5	
12	19.9	31.2	51.1	
13	11.1	**4.2**	15.3	All the same value.
14	36.2	58.4	94.6	
15	**4.1**	**3.5**	**7.6**	All the same value.

Notes:

1. Plants where the sum of the DRR over 15 weeks is less than 5, or the sum of the DRR over 30 weeks is less than 10 will be highlighted. These values were chosen as rough critical values (obtained by looking at several cases). In future the F distribution should be considered for this critical value.

FEATURES OF THE SYSTEM

Up to this point the discussion has been fairly general, on which statistical estimators to use to determine the edit limits. With a decision to use the biweight prediction interval to flag outliers, and the DRR to flag inliers, we can now specify how the new edit system will work.

First, the edit system must perform validation edits, or within record checks. These include identification code checks, checks that certain rows and columns sum to the appropriate totals, checks that the number of head in the head section corresponds to the number of head in the weight section, and that dressed weight is less than live weight (or %DW/LW is between 0 and 1). Secondly, statistical edits representing between-record checks must be done (in our case, using historical data within plant across time) to validate the reported number of head and total weight. Details of these edits are provided in Appendix 3, however, the general features of the system are provided below.

1. **Stratification**

   Slaughter plants will be stratified to allow editing and imputation of plants with insufficient historical data, and to set up reasonable edit limits for very small plants. A biweight mean and standard deviation will be calculated for the strata, where the strata will be based on size (the number of animals) for each class. The same strata definitions will exist for each state. If necessary, strata will be collapsed by state.

   a) A prediction interval based on the stratum biweight will be used to edit plants with not enough data to calculate their own biweight (<13 values in the last year). It can also be used for new, changed or seasonal plants.

   b) A prediction interval based on the stratum biweight will be used to edit small plants (e.g. <20 animals per day).

   c) Missing weight data will be imputed using the plant's biweight if sufficient historical data exists, otherwise the stratum biweight will be used (plants with <13 values in the last year).

2. **Journal**

   The journal file will contain a record of all changes made to the data during the edit. This audit trail will allow the user to determine the effect of the edit on the summary, and identify the types of errors made. Further comments are given in results in practice on page 26.

3. **Master ID File**

   This file can be used to identify plants which are closed for some reason (strike, holiday or other), but it can also be used to verify id codes and protect against duplication.

4. **Missing Analysis Routine**

   This routine enables the user to determine the number of plants not yet reported for a week, and the effect on the summary.

5. **User Interaction**

   The user is able to set the necessary parameters, and the strata definitions.

6. **Interactive Microcomputer-based Edit**

   The integrated system uses DBase III+ on the PC (compiled in multiuser Foxbase 2.1) to enter, edit, and summarize the data. One reason for using a database package was the ability to update (correct) records at any time. Currently, updates are only done once per year due to the high cost of processing a sequential file on the mainframe. A modular program allows changes or other data series to be incorporated.

COSTS

The new edit system will result in substantial cost savings for NASS. The yearly leasing cost of processing and storing data on the mainframe (federally inspected plants) will be exchanged for microcomputer equipment which will be purchased initially (network), but require only maintenance charges thereafter. Equipment purchases will be low, as several PCs are already available. The non-federally inspected plants' records (used in the summary) will be on the mainframe, but will be downloaded to the PC for summary. Roughly speaking, a 75% savings will result the first year, and an 81% savings in the following years (compared to what it would have cost on the current edit). Note that costs are for federal and non-federally inspected plants together, as separate costs could not be obtained.

RESULTS IN PRACTICE

As of September 1990, all livestock slaughter data is being entered and edited on the PC based system. Data is being uploaded to the mainframe to cross check with the old generalized edit and perform the necessary summaries. When the PC summary system is completed, data will no longer be uploaded to the main frame.

Several features of the PC based edit are still being worked on. This includes the DRR calculation which is currently being debugged, and the inlier test which is based on the DRR calculation. Also, the journal is not hooked up at present. There was some concern about it slowing down the system. Lastly, the live weight edit check for each of the four classes is being done by inflating the edited DW for a species, by a predetermined value of %DW/LW for the species and summing across the class (rather then using a regression equation as on page 32). The edit range is then the predicted value plus or minus 10% of the predicted value.

Generally, the livestock staff is very happy with the new edit system, as it makes their jobs much easier and takes less time. Also, management has now been made aware of the substantial proportion of imputation that is being done with weight data.

FUTURE RESEARCH

1. Other data series

   A similar approach could be considered for several other data series which collect data in a manner similar to Livestock Slaughter. That is, they collect data over a long time from the same units. In addition, most of them are using the generalized edit system, and would save costs by limiting data processing and storage on the mainframe. These are Poultry Slaughter, Cold Storage, Peanut Stocks, Manufactured Dairy Products, Off Farm Grain Stocks, Turkey Hatchery, Extreme Operators, and Cattle on Feed.

2. Graphics System

   A system to plot historic time series (with confidence intervals) would help the specialists recognize outliers, and visualize patterns in the data.

3. Seasonality

   Research should be done to see if seasonality exists (e.g. with the large plants), and if it could be incorporated into the biweight. The reason for this, is that some values may be identified as "outliers" using the proposed method (and vice versa), which would not be if a seasonal factor were incorporated.

4. Individual plant holidays

   A facility to identify holiday patterns by plant could be added. For example, identifying which holidays a plant takes, whether or not the day is made up, and vacations.

REFERENCES:

1. Andrews, et al., (1972): *Robust Estimates of Location: Survey and Advances.* Princeton University Press, N.J.

2. Du Mond, C., and Lenth, R., (1986): *A Robust Confidence Interval.* Abstract from the ASA Proceedings of the Statistical Computing Section, pp. 139-143.

3. Hoaglin, D., and Velleman, P.F., (1981): *Applications, Basics, and Computing of Exploratory Data Analysis.* Duxbury Press, MA., pp. 265-6.

4. Hoaglin, D., Mosteller, F., and Tukey, J., editors, (1983): *Understanding Robust and Exploratory Data Analysis.* John Wiley & Sons, N.Y.

(Note the Reference to 1981 Personal Communication by John Tukey on pg. 387.)

5. Pierzchala, M., (1988): *A Review of the State of the Art in Automated Data Editing and Imputation.* NASS Staff Report No. SRB-88-10.

APPENDIX 1

QUESTIONNAIRE

NASS:AMS:FSIS
Form LS-149

UNITED STATES DEPARTMENT OF AGRICULTURE

Form Approved
OMB Number 0535-0005
Expiration Date 12/31/89

WEEKLY LIVESTOCK SLAUGHTER REPORT

Plant________________ State________________

Inspector________________ Establishment No.________________

1-2 STATE	3-8 NUMERIC	9-12 ALPHA
13-14 MONTH	15-16 DAY	17-18 YEAR
19-20 MONTH	21-22 DAY	23-24 YEAR

Response to this form is voluntary and not required by law. However, cooperation is very important in order to fulfill responsibilities mandated by the Meat Inspection Act and to provide statistical information to maintain an orderly flow of red meat throughout the livestock industry.

INSTRUCTIONS: *Include all species slaughtered in each plant including custom slaughter. Complete a separate Form LS-149 each week for each Federally inspected plant. See the back of this form for detailed instructions.*

WEEK ENDING ____/____/____
Mo. Day Yr.

	NUMBER HEAD SLAUGHERED DAILY (Including Post-Mortem Condemnations)					
Species and Class	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Steers	101	201	301	401	501	601
Heifers	102	202	302	402	502	602
Cows: Dairy	105	205	305	405	505	605
All Other	106	206	306	406	506	606
Bulls & Stags	104	204	304	404	504	604
CATTLE-Total 010	100	200	300	400	500	600
CALVES-Total 011	110	210	310	410	510	610
Barrows & Gilts	121	221	321	421	521	621
Sows	122	222	322	422	522	622
Stags & Boars	123	223	323	423	523	623
HOGS-Total 020	120	220	320	420	520	620
Mature Sheep	131	231	331	431	531	631
Lambs & Yearlings	132	232	332	432	532	632
SHEEP-Total 030	130	230	330	430	530	630
GOATS-Total →	140	240	340	440	540	640
EQUINE-Total →	150	250	350	450	550	650

WEEK ENDING ____/____/____
Mo. Day. Yr.

	WEEKLY TOTAL HEAD, TOTAL LIVE WEIGHT AND TOTAL DRESSED WEIGHT (Exclude Post-Motem Condemnations)		
Species and Class	Number of Head	Total Live Weight	Total Dressed Weight
Steers	701		801
Heifers	702		802
Total Cows	703		803
Bulls & Stags	704		804
CATTLE-Total →	700	705	800
CALVES-Total →	711	716	811
Barrows & Gilts	721		821
Sows	722		822
Stags & Boars	723		823
HOGS-Total →	720	725	820
Mature Sheep	731		831
Lambs & Yearlings	732		832
SHEEP-Total →	730	735	830

Form Number LS-149
Rev. December 1986

NATIONAL AGRICULTURAL STATISTICS SERVICE COPY

APPENDIX 2

Weighted Biweight Method

1. The calculation of BiMn and BiSd must take into account the number of animals contributing to the average dressed weight or %DW/LW each week. This is done whether a single plant or a stratum is involved.

2. A "weighted" median is determined by replacing n with the total number of animals for all 13 weeks. If $Freq_i$ represents the number of animals each week, the $\Sigma Freq_i$ represents the total number of animals for all 13 weeks. Then, cumulate these frequencies (with records in sorted order) and choose that value that contains the $(\Sigma Freq_i + 1)/2^{th}$ number.

3. Calculate the "weighted" IQ range or MAD by finding the appropriate values (Q1, Q3, or median) using the cumulated frequencies.

4. The weight for each value is calculated the same, but the weighted median and IQ range (or MAD) is used.

5. Incorporate $Freq_i$ into the BiMn and BiSd equations as follows:

$$BiMn = \frac{\Sigma[X_i * Freq_i * (1 - U_i^2)^2]}{\Sigma\,[Freq_i * (1 - U_i^2)^2]}$$

$$BiSd = \frac{\sqrt{\Sigma Freq_i * \Sigma[(X_i - M)^2 * Freq_i * (1 - U_i^2)^4]}}{|\,\Sigma[Freq_i * (1 - U_i^2) * (1 - (5 * U_i^2))]\,|}$$

where M represents the weighted median from above.

APPENDIX 3

Detailed Procedures

1. Head Data

The number of head slaughtered daily (Monday through Saturday) is reported for each class at the top of the questionnaire. Class week totals and species daily totals are calculated.

a. Daily and Weekly Tests

Each class with positive slaughter will be edited on a daily basis. The editing range will be for expected slaughter on a given positive kill day. The edit limits will be calculated using the biweight method by considering the previous number of weeks which contain at least 13 positive values. That is, 13 weeks will be used for plants which slaughter one day per week, and 3 weeks will be used for plants which report 6 days per week (and the entire 18 values will be used).

Class week totals will also be edited. The editing range will be for expected week total slaughter. The edit limits will be calculated using the biweight method by considering the last 13 positive slaughter weeks.

Both of these tests are necessary. During holiday weeks for example, a plant which misses a day (due to the holiday) but does not make it up elsewhere will pass the daily test. However, a plant which misses the holiday, but makes it up by slaughtering twice as many on a different day (or slaughtering on a Saturday) will pass the week test.

b. Patterns and Holidays

A pattern test will be done to check plants with a consistent slaughter pattern. A plant will be considered to have a consistent pattern if 3 patterns cover over 90% of the weeks over the last year after excluding weeks with no slaughter, and accounting for changes due to holidays. This test will check, for example, to see if a plant which typically slaughters on Monday, Wednesday and Friday is sticking to this pattern. A file will contain the ids of plants which follow a consistent pattern. The specific patterns will be stored as a base 10 number, and interpreted as strings of 1s and 0s (for positive and zero slaughter days)

A plant with a consistent pattern will need to account for holidays. A holiday file will contain the dates of holidays, such as New Year's Day, Memorial Day, and Thanksgiving Day. During the week that contains Memorial Day, a plant which typically slaughters Monday, Wednesday and Friday will not report an error, if no animals are slaughtered on that Monday.

2. Weight Data

Week totals for the number of head, dressed weight, and live weight are reported at the bottom of the questionnaire.

a. Average Dressed Weights

Dressed weight for each class will be edited based on the average dressed weight per animal (that is, total dressed weight / total number head after accounting for condemnations). The editing range will be calculated using a weighted biweight approach (see Appendix 2), using the last 13 positive unimputed week's data.

b. LW (Calves)

Live weight for calves will be edited differently than the other species, since calves are not broken out by class. Since the dressed weight to live weight ratio is fairly consistent and LW = DW/(%DW/LW), the edit will be based on this ratio. The editing range will be calculated from the last 13 week's %DW/LW using the weighted biweight method (see Appendix 3). The current week's ratio can then be compared with the historic ratio. See the discussion on page 26 for results in practice.

c. LW (Cattle, Hogs, and Sheep)

Cattle, hogs, and sheep are broken into several classes each, however live weight is only reported for the species. Since each class has its own percent dressed weight to live weight, and the same classes are not slaughtered each week, the biweight approach used for calves will not work. However, a regression equation will work well.

$$\hat{Y} = b_o + \Sigma b_i X_i$$

Note, Y represents the dependent variable (species live weight), b_i represents the parameters or constants (inverse of the ratio of dressed weight to live weight for each class), b_o is the Y intercept, and X_i represents the independent variables (dressed weight for each class). Note that the ratio of dressed weight to live weight for each class is not available. This ratio is estimated using a regression. Once the parameters have been determined (for each class), the total species live weight for a plant can be approximated by multiplying the total class dressed weight for the current week by the appropriate parameter, and summing over all classes. The standard deviation on species live weight is then as follows.

$$SD(\hat{Y}) = [s^2 + (s^2/n) + \Sigma(X_{ok} - \overline{X}_k)^2 * VAR(b_k) + 2\Sigma(X_{oj} - \overline{X}_j)(X_{ok} - \overline{X}_k) * COV(b_j, b_k)]$$

The edit limits would then be:

$$\hat{Y} - t_{n-p} * SD(\hat{Y}) \, , \, \hat{Y} + t_{n-p} * SD(\hat{Y})$$

The letter "p" refers to the number of regression parameters. See the discussion on page 26 for results in practice.

☆U.S. Government Printing Office : 1990 - 282-964/21003