

STATISTICAL RESEARCH DIVISION REPORT SERIES
Statistical Research Report Number:
CENSUS/SRD/RR-85/22

Administrative Data Study for the
1982 Economic Censuses

by

Geneva A. Burns
Statistical Research Division
U.S. Bureau of the Census
Rm. 3552, F.O.B. #3
Washington, D.C. 20233

(301) - 763 - 7200

With the assistance of:

Linda Moore
Business Division

Pamela McKee
Industry Division

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Recommended by: Nash Monsour
Report Completed: March 4, 1986
Report Issued: March 12, 1986

1. Executive Summary

The purpose of this study was to investigate the differences between the administrative record data and Census data for establishments which reported to the Internal Revenue Service (IRS) and responded to the 1982 Censuses of Retail Trade, Wholesale Trade, Selected Services, Construction Industries, or Manufactures. In this study, rather than applying statistical sampling for selection of cases, entire Standard Industrial Classifications (SIC's) by geographic areas were selected. The selected SIC's were ones that were thought to present problems when administrative data were used in lieu of census data. Thus, data sets for each SIC were treated as the universe and results from each SIC are applicable to only that selected SIC.

This report covers the results obtained thus far from examining two SIC's from wholesale trade, three SIC's from selected services, one SIC from construction, and one SIC from manufactures. The examination of these SIC's showed the following information existent in these data sets.

1. Patterns of differences between administrative record data and census data were not found. Differences in the data are sporadic and often appear to be due to processing or transcription problems.
2. The direction of the differences between administrative and census data shows that the administrative figures are lower than the census figures before extensive editing has been performed but larger than census data after all editing has been performed on both kinds of data.

If patterns of differences between the administrative and census data had been detected, the next phase of the project would have been to consider alternative methods of adjusting the administrative record data. Based on the results obtained, the second phase of the project was unwarranted for these SIC's.

2. Detailed Findings from this Study

The findings include documentation of the examination of data based on two SIC's from wholesale trade, three SIC's from selected services, one SIC from construction, and one SIC from manufactures. Section 2.1 contains tables showing totals for census and administrative data and differences between these totals. Section 2.2 discusses outliers in the data and Section 2.3 contains a discussion on the graphs attached to this report. In Section 2.4, there is a description of additional scattergrams, histograms, and tables included in the Administrative Data Study Reference Memorandum. Section 2.5 discusses additional analysis of data while Section 2.6 contains a brief discussion of a related part of this study.

Most of the results discussed in this report are for SIC 50 for wholesale trade (durable goods) and SIC 51 for wholesale trade (nondurable goods) in Nebraska. The universe for SIC 50 is 618 establishments and for SIC 51 is 651 establishments. Cases which had a blank or zero for an item were declared missing for analysis purposes. This declaration of missing cases is mentioned throughout this section.

2.1 Tables Showing Statistics From Data

Tables 1-8 show totals, means, and standard deviations from SIC 50 and SIC 51 in Nebraska for (1) unedited administrative versus original census data, (2) edited administrative versus original census data, (3) unedited administrative versus tabulated census data, (4) edited administrative versus tabulated census data. Differences in number of cases for tables showing the same kinds of data are explained by the methods used for declaring missing cases in a computer run. A case is declared missing from the universe if the administrative and/or census figure for an item is blank or zero.

Totals of differences were obtained by subtracting the total census figures for all establishments from the total administrative figures for all establishments. The difference shows the direction of the difference, i.e., a negative difference indicates the total administrative data were smaller than the total census data and a positive difference shows the total administrative data were larger than the total census data. Totals for the absolute difference for each item were computed by subtracting the census figure from the administrative figure for each establishment and summing the absolute value of these differences. These totals show the magnitude of the differences between census and administrative data without consideration of the direction of these differences. Figures showing means and standard deviations were calculated by using the totals mentioned above in the basic formulae for computing means and standard deviations of a population.

**Table 1. Totals for Unedited Administrative versus Original Census Data
Nebraska - SIC 50 - Single Units**

<u>Item</u>	<u>Number of Cases</u>	<u>Total</u> ¹	<u>Population Mean</u>	<u>Population Standard Deviation</u> ²
Employment				
Administrative	569	8,797	15.460	20.952
Census	569	9,995	17.566	47.457
Difference	569	-1,198	-2.105	43.072
Absolute Difference	569	2,014	3.540	42.978
First Quarter Payroll				
Administrative	543	36,298	66.847	104.092
Census	543	91,435	168.389	1,515.413
Difference	543	-55,137	-101.541	1,514.011
Absolute Difference	543	60,113	110.705	1,513.369
Annual Payroll				
Administrative	572	149,592	261.524	391.456
Census	572	985,635	1,723.138	19,841.179
Difference	572	-836,043	-1,461.614	19,823.800
Absolute Difference	572	848,901	1,484.093	19,822.129
Receipts				
Administrative	460	913,262	1,985.352	2,268.347
Census	460	12,369,062	26,889.265	391,420.164
Difference	460	-11,455,800	-24,903.913	391,144.889
Absolute Difference	460	11,529,622	25,064.396	391,134.638

¹ These numbers have been rounded to thousands for first quarter payroll, annual payroll, and receipts.

² These standard deviations were derived by taking the square root of the squares of the differences between individual observations from the population mean divided by the number of cases.

Table 2. Totals for Edited Administrative versus Original Census Data
Nebraska - SIC 50 - Single Units

<u>Item</u>	<u>Number of Cases</u>	<u>Total</u> ¹	<u>Population Mean</u>	<u>Population Standard Deviation</u> ²
Employment				
Administrative	402	5,103	12.694	11.248
Census	402	5,323	13.241	17.376
Difference	402	-220	-.547	13.465
Absolute Difference	402	670	1.667	13.373
First Quarter Payroll				
Administrative	379	19,372	51.113	47.995
Census	379	75,795	199.987	1,809.765
Difference	379	-56,423	-148.873	1,810.067
Absolute Difference	379	59,127	156.008	1,809.466
Annual Payroll				
Administrative	397	81,784	206.005	174.665
Census	397	886,589	2,233.222	23,725.572
Difference	397	-804,805	-2,027.217	23,715.740
Absolute Difference	397	814,043	2,050.486	23,713.739
Receipts				
Administrative	393	758,409	1,929.794	2,962.381
Census	393	734,178	1,868.137	6,998.431
Difference	393	24,231	61.656	7,323.582
Absolute Difference	393	373,255	949.758	7,261.997

¹ These numbers have been rounded to thousands for first quarter payroll, annual payroll, and receipts.

² These standard deviations were derived by taking the square root of the squares of the differences between individual observations from the population mean divided by the number of cases.

Table 3. Totals for Unedited Administrative versus Tabulated Census Data
Nebraska - SIC 50 - Single Units

<u>Item</u>	<u>Number of Cases</u>	<u>Total</u> ¹	<u>Population Mean</u>	<u>Population Standard Deviation</u> ²
Employment				
Administrative	543	7,810	14.383	18.942
Census	543	7,749	14.271	18.696
Difference	543	61	.112	2.713
Absolute Difference	543	583	1.074	2.488
First Quarter Payroll				
Administrative	543	32,790	60.387	90.647
Census	543	31,808	58.578	90.040
Difference	543	982	1.808	19.415
Absolute Difference	543	2,960	5.451	18.722
Annual Payroll				
Administrative	547	130,023	237.702	339.538
Census	547	126,975	232.130	335.485
Difference	547	3,048	5.572	48.555
Absolute Difference	547	7,856	14.362	46.715
Receipts				
Administrative	447	847,863	1,896.785	1,982.713
Census	447	895,872	2,004.188	2,276.998
Difference	447	-48,009	-107.403	1,206.992
Absolute Difference	447	111,331	249.063	1,185.888

¹ These numbers have been rounded to thousands for first quarter payroll, annual payroll, and receipts.

² These standard deviations were derived by taking the square root of the squares of the differences between individual observations from the population mean divided by the number of cases.

**Table 4. Totals for Edited Administrative versus Tabulated Census Data
Nebraska - SIC 50 - Single Units**

<u>Item</u>	<u>Number of Cases</u>	<u>Total</u> ¹	<u>Population Mean</u>	<u>Population Standard Deviation</u> ²
Employment				
Administrative	420	5,294	12.605	11.229
Census	420	5,248	12.495	11.258
Difference	420	46	.110	2.294
Absolute Difference	420	404	.962	2.086
First Quarter Payroll				
Administrative	420	21,060	50.143	46.911
Census	420	20,680	49.238	47.178
Difference	420	380	.905	23.059
Absolute Difference	420	2,466	5.871	22.317
Annual Payroll				
Administrative	420	85,370	203.262	173.488
Census	420	82,801	197.145	163.399
Difference	420	2,569	6.117	52.427
Absolute Difference	420	7,163	17.055	49.951
Receipts				
Administrative	421	789,796	1,876.000	2,889.036
Census	421	642,947	1,527.190	1,573.446
Difference	421	146,849	348.810	2,301.060
Absolute Difference	421	230,127	546.620	2,262.245

¹ These numbers have been rounded to thousands for first quarter payroll, annual payroll, and receipts.

² These standard deviations were derived by taking the square root of the squares of the differences between individual observations from the population mean divided by the number of cases.

Table 5. Totals for Unedited Administrative versus Original Census Data
Nebraska - SIC 51 - Single Units

<u>Item</u>	<u>Number of Cases</u>	<u>Total</u> ¹	<u>Population Mean</u>	<u>Population Standard Deviation</u> ²
Employment				
Administrative	597	8,182	13.705	20.184
Census	597	8,252	13.822	19.835
Difference	597	-70	-.117	7.963
Absolute Difference	597	1,174	1.966	7.718
First Quarter Payroll				
Administrative	561	28,885	51.488	109.282
Census	561	193,471	344.868	4,314.838
Difference	561	-164,586	-293.380	4,312.359
Absolute Difference	561	166,664	297.084	4,312.106
Annual Payroll				
Administrative	606	128,185	211.526	464.912
Census	606	1,766,785	2,915.487	31,897.372
Difference	606	-1,638,600	-2,703.960	31,884.773
Absolute Difference	606	1,646,698	2,717.323	31,883.637
Receipts				
Administrative	380	1,690,435	4,448.513	13,407.637
Census	380	2,102,357	5,532.518	14,384.123
Difference	380	-411,922	-1,084.005	6,966.046
Absolute Difference	380	634,158	1,668.837	6,849.315

¹ These numbers have been rounded to thousands for first quarter payroll, annual payroll, and receipts.

² These standard deviations were derived by taking the square root of the squares of the differences between individual observations from the population mean divided by the number of cases.

**Table 6. Totals for Edited Administrative versus Original Census Data
Nebraska - SIC 51 - Single Units**

<u>Item</u>	<u>Number of Cases</u>	<u>Total¹</u>	<u>Population Mean</u>	<u>Population Standard Deviation²</u>
Employment				
Administrative	458	4,567	9.972	9.694
Census	458	4,655	10.164	10.579
Difference	458	-88	-.192	6.014
Absolute Difference	458	662	1.445	5.841
First Quarter Payroll				
Administrative	431	16,306	37.833	42.845
Census	431	72,535	168.295	2,719.883
Difference	431	-56,229	-130.462	2,718.070
Absolute Difference	431	58,127	134.865	2,717.856
Annual Payroll				
Administrative	459	71,169	155.052	163.861
Census	459	731,224	1,593.081	22,427.199
Difference	459	-660,055	-1,438.028	22,423.363
Absolute Difference	459	667,141	1,453.466	22,422.367
Receipts				
Administrative	453	1,176,563	2,597.269	3,320.803
Census	453	1,433,781	3,165.808	4,783.229
Difference	453	-257,548	-568.539	3,966.065
Absolute Difference	453	506,264	1,117.581	3,847.586

¹ These numbers have been rounded to thousands for first quarter payroll, annual payroll, and receipts.

² These standard deviations were derived by taking the square root of the squares of the differences between individual observations from the population mean divided by the number of cases.

**Table 7. Totals for Unedited Administrative versus Tabulated Census Data
Nebraska - SIC 51 - Single Units**

<u>Item</u>	<u>Number of Cases</u>	<u>Total¹</u>	<u>Population Mean</u>	<u>Population Standard Deviation²</u>
Employment				
Administrative	562	7,145	12.714	18.396
Census	562	7,077	12.584	17.012
Difference	562	73	.130	4.766
Absolute Difference	562	829	1.475	4.537
First Quarter Payroll				
Administrative	562	25,200	44.840	74.812
Census	562	24,979	44.447	67.686
Difference	562	221	.393	27.220
Absolute Difference	562	2,921	5.198	26.722
Annual Payroll				
Administrative	569	106,743	187.598	300.173
Census	569	103,810	182.443	276.350
Difference	569	2,933	5.155	86.509
Absolute Difference	569	8,855	15.562	85.254
Receipts				
Administrative	371	1,250,772	3,371.353	5,318.534
Census	371	1,691,718	4,559.887	7,244.588
Difference	371	-440,946	-1,188.534	6,268.085
Absolute Difference	371	631,044	1,700.927	6,148.850

¹ These numbers have been rounded to thousands for first quarter payroll, annual payroll, and receipts.

² These standard deviations were derived by taking the square root of the squares of the differences between individual observations from the population mean divided by the number of cases.

**Table 8. Totals for Edited Administrative versus Tabulated Census Data
Nebraska - SIC 51 - Single Units**

<u>Item</u>	<u>Number of Cases</u>	<u>Total¹</u>	<u>Population Mean</u>	<u>Population Standard Deviation²</u>
Employment				
Administrative	478	4,780	10.000	9.775
Census	478	4,777	9.994	9.392
Difference	478	3	.006	2.723
Absolute Difference	478	549	1.149	2.469
First Quarter Payroll				
Administrative	478	17,472	36.552	41.876
Census	478	17,332	36.259	42.697
Difference	478	140	.293	13.126
Absolute Difference	478	1,746	3.653	12.611
Annual Payroll				
Administrative	478	73,447	153.655	164.833
Census	478	72,457	151.584	165.198
Difference	478	990	2.071	37.424
Absolute Difference	478	6,044	12.644	35.284
Receipts				
Administrative	479	1,284,985	2,682.641	3,662.129
Census	479	1,420,583	2,965.727	3,195.405
Difference	479	-135,598	-283.086	2,594.374
Absolute Difference	479	451,352	942.280	2,433.726

¹ These numbers have been rounded to thousands for first quarter payroll, annual payroll, and receipts.

² These standard deviations were derived by taking the square root of the squares of the differences between individual observations from the population mean divided by the number of cases.

Tables 1 and 5 show that when unedited administrative are compared with original census data, the differences for all four items (employment, first quarter payroll, annual payroll, and receipts) are negative. In other words, the totals for administrative data are smaller than census totals. Tables 2 and 6 show that this still holds true, except for receipts in SIC 50, after only the administrative data have been edited. However, in Tables 3 and 7, this situation is reversed, i.e. administrative totals are larger than census totals, except for receipts in SIC's 50 and 51, after only census data have been edited. This remains true, as shown in Tables 4 and 8, except for receipts in SIC 51, after the administrative and census data have both been edited. This reversal in the magnitude of the differences suggests that the percentage of change in the data due to editing is greater for one type of data than for the other. In this case, both types of data are decreased by editing but the percentage of decrease is greater for census than for administrative data.

2.2 Outliers

In determining whether differences between census and administrative data fit any particular patterns or whether adjustment of administrative data is warranted, outliers in the data were taken into consideration. The approach taken for this study was to examine scatterplots of the data for patterns of outliers. Although patterns of outliers were not found, some outliers were sporadic throughout the data. These outlier cases were omitted from the data set and regressions were run on the remaining cases to examine the potential for predicting census data based on administrative data.

Tables 9a and 9b show these regression statistics for all cases and nonoutliers for unedited administrative versus original census data and for edited administrative versus tabulated census data for employment, annual

payroll, and receipts in Nebraska for single units in SIC 50. The arbitrary definition used for outliers was any case where the absolute difference between the administrative and census figure was greater than seventy-five percent of the census number.

**Table 9a. Regression Statistics for All Cases and Nonoutliers
Unedited Administrative versus Original Census Data
Nebraska - SIC 50 - Single Units**

	<u>Employment</u>	<u>Annual Payroll</u>	<u>Receipts</u>
All Cases			
Number of cases	569	571	460
r (correlation)	.42039	.05416	.12421
Slope	.95222	2.74627	21.43337
Intercept	2.84421	1,006.68037	-15,663.51832
Nonoutliers			
Number of cases	555	551	438
r (correlation)	.99609	.99396	.99174
Slope	.99106	.99636	.98543
Intercept	.24636	-1.82460	30.64871

**Table 9b. Regression Statistics for All Cases and Nonoutliers
Edited Administrative versus Tabulated Census Data
Nebraska - SIC 50 - Single Units**

	<u>Employment</u>	<u>Annual Payroll</u>	<u>Receipts</u>
All Cases			
Number of cases	420	420	421
r (correlation)	.97918	.95332	.60797
Slope	.98174	.89788	.33112
Intercept	.12069	14.64107	906.01179
Nonoutliers			
Number of cases	415	408	378
r (correlation)	.98779	.96782	.96492
Slope	.99811	.93068	.97959
Intercept	.07168	11.51060	90.84732

In comparing the regression statistics for all cases with those for nonoutliers, the correlation increased as was expected when the outliers were omitted from the data. However, the comparison between unedited administrative versus original census data and edited administrative versus tabulated census for nonoutliers shows a slight decrease in the correlation after editing. For example, the correlation for nonoutliers from unedited data is .99609 for employment compared with .98779 from edited data for employment. This decrease in correlations after editing may be related to the reversal of the magnitude of the differences between administrative and census totals mentioned in Section 2.1 of this report. However, to determine the cause and effect of editing on the correlations, the editing procedures for both administrative and census data would need to be examined. This examination of editing procedures is outside the scope of this study.

2.3 Pictorial Displays of Data

Attached to this report in Appendix A is a sample of the scattergrams produced for analyzing the data. These scattergrams were used in the analyses to visually display the administrative versus census data so that differences between the two kinds of data could be easily detected. The statistics accompanying the scattergrams were reviewed as a method for determining the effect of differences between the data. Also, these statistics can be used as a measure to determine the feasibility of predicting census data based on administrative data. The following details are necessary for interpretation of these graphs:

- (1) First quarter payroll, annual payroll, and receipts figures are rounded to thousands.

- (2) An asterisk on the scattergrams indicates one plotted point and numbers 2-8 show that 2-8 cases fell on one point. A 9 indicates that 9 or more cases fell on the same point.
- (3) Missing values are defined as cases where the administrative or census values, or both, were zero or blank. These cases are omitted from graphs and from statistics.
- (4) Excluded values are defined as those cases where the values for the two variables displayed were outside the specified range. Excluded values are mutually exclusive of missing values and are not included in the statistics. Scattergrams were produced that allowed for no excluded values but the points on the graphs were concentrated too heavily in the lower left-hand corner for viewing purposes. Also, since these excluded values generally are outliers, the statistics produced when all values were included were meaningless.

Figures 1-8 in Appendix A show unedited administrative versus original census data by employment, first quarter payroll, annual payroll and receipts, and edited administrative versus tabulated census data by these same items for SIC 50 in Nebraska. The correlations between the administrative and census data from these graphs are summarized below. Also, correlations for the basic items are shown for unedited administrative versus tabulated census data and edited administrative versus original census data.

**Table 10. Summary of Correlations for the Four Kinds of Data
by Employment, First Quarter Payroll, Annual Payroll,
and Receipts for SIC 50 in Nebraska**

<u>Item</u>	Unedited Administrative versus <u>Original Census</u>	Edited Administrative versus <u>Original Census</u>	Unedited Administrative versus <u>Tabulated Census</u>	Edited Administrative versus <u>Tabulated Census</u>
Employment	.96209	.97809	.97865	.97918
First Quarter Payroll	.97180	.86428	.93292	.87989
Annual Payroll	.97230	.95195	.96934	.95332
Receipts	.96325	.73162	.95453	.75245

This shows that except for employment, the correlations between administrative and census data decreased after complex editing was performed on only the administrative data, only the census data, and on both the administrative and census data. This indicates that editing had an unexpected effect on the correlations but as mentioned in Section 2.2, the cause is outside the scope of this study.

Corresponding correlations are shown in Table 11 for SIC 51 in Nebraska.

**Table 11. Summary of Correlations for the Four Kinds of Data by
Employment, First Quarter Payroll, Annual Payroll,
and Receipts for SIC 51 in Nebraska**

<u>Item</u>	Unedited Administrative versus <u>Original Census</u>	Edited Administrative versus <u>Original Census</u>	Unedited Administrative versus <u>Tabulated Census</u>	Edited Administrative versus <u>Tabulated Census</u>
Employment	.83337	.82748	.96090	.96043
First Quarter Payroll	.92994	.95945	.95269	.95163
Annual Payroll	.98167	.96620	.98388	.97413
Receipts	.72213	.74901	.64040	.78553

These correlations fluctuated during the different editing phases, but in comparing unedited administrative versus original census data with edited administrative versus tabulated census data, the correlations for every item except annual payroll increased after both administrative and census data were edited.

The unedited administrative versus original census data for receipts in SIC 51 have a correlation of .72213 which is low in comparison to the corresponding correlation (.96325) for receipts in SIC 50. The scattergram (not shown here) for SIC 51 shows many cases with census receipts much higher than administrative receipts. The raw data indicate processing errors for some of these cases but no pattern of differences appears evident. After editing by both types of data, the correlation for receipts in SIC 51 increased slightly to .78553 and some cases still show high census receipts with low administrative receipts.

Table 12 shows correlations from SIC 50 in Nebraska broken down by counties 55, 79, and 109 for employment and annual payroll by unedited administrative versus original census data and by edited administrative versus tabulated census data.

Table 12. Summary of Correlations for Unedited Administrative versus Original Census Data and Edited Administrative versus Tabulated Census Data by Employment and Annual Payroll for SIC 50 in Nebraska by County Code

<u>Item</u>	Unedited Administrative versus <u>Original Census</u>			Edited Administrative versus <u>Tabulated Census</u>		
	County <u>55</u>	County <u>79</u>	County <u>109</u>	County <u>55</u>	County <u>79</u>	County <u>109</u>
Employment	.96443	.99304	.98869	.99240	.99149	.99507
Annual Payroll	.97445	.99627	.99272	.95165	.99832	.99118

The summary of correlations shows some increases and some decreases in the correlations after editing but all changes are slight which implies that when the data were grouped by county for this SIC, editing had little effect on the correlations.

Table 13 shows correlations for unedited administrative versus original census data and edited administrative versus tabulated census for employment and annual payroll for SIC 50 in Nebraska broken down by TO (type of operation) codes 00, 10, and 11.

Table 13. Summary of Correlations for Unedited Administrative versus Original Census Data and Edited Administrative versus Tabulated Census Data by Employment and Annual Payroll for SIC 50 in Nebraska by TO-Code

<u>Item</u>	Unedited Administrative versus <u>Original Census</u>			Edited Administrative versus <u>Tabulated Census</u>		
	<u>TO</u>	<u>TO</u>	<u>TO</u>	<u>TO</u>	<u>TO</u>	<u>TO</u>
	<u>Code 00</u>	<u>Code 10</u>	<u>Code 11</u>	<u>Code 00</u>	<u>Code 10</u>	<u>Code 11</u>
Employment	.89453	.97470	.97571	.98353	.99361	.97773
Annual Payroll	.98461	.99775	.96843	.99187	.99457	.94606

As with the breakdown by counties, these breakdowns by TO codes show slight increases and decreases in the correlations after complex editing which implies that with grouping by TO, editing had little effect on the correlations.

Table 14 shows correlations for manufactures SIC 2335 in New York by unedited administrative versus original census data for employment, annual payroll, and receipts. SIC 2335 was broken down by county codes for employment.

Table 14. Summary of Correlations for Unedited Administrative versus Original Census Data for SIC 2335 for Employment, Annual Payroll, and Receipts in New York by County Code

<u>Item</u>	<u>All</u> <u>Counties</u>	<u>County</u> <u>005</u>	<u>County</u> <u>061</u>	<u>County</u> <u>047</u>	<u>County</u> <u>081</u>	<u>County</u> <u>103</u>
Employment	.88094	.66340	.80247	.89150	.94674	.88223
Annual Payroll	.97143					
Receipts	.94714					

The correlation for employment in county 005 is particularly low and appears to be due to outliers in the data. Edited administrative and tabulated census data for manufacturing SIC's were unavailable for analysis at the time this report was produced.

2.4 Additional Displays of Data

Many additional scattergrams and histograms were produced and reviewed that are not included with this report in order to keep this report at a manageable size. These have been compiled in the Administrative Data Study Reference Memorandum and copies will be made available upon request by interested persons. This memorandum consists of scattergrams for the four kinds of data by employment, first quarter payroll, annual payroll, and receipts for SIC's 50, 51, 72, and 73. Breakdowns by 3-and 4-digit SIC's, county codes, TO codes, and size of employment class are depicted in the reference memorandum. Also, tables which follow the same format as Tables 1-8 are available for SIC's 72 and 73 in Maryland for personal and business services, respectively. Breakdowns by county codes, TO codes, 3-digit SIC's, and 4-digit SIC's are shown for SIC's 50 and 51. For unedited administrative versus original census data, tables showing totals are available for SIC 1622, which is bridge, tunnel, and elevated highway construction for the entire U.S.; for SIC 2421 in Kentucky for sawmills and planing mills; and for 2335 in New York, which is women's dresses.

2.5 Additional Analysis of Data

Time did not permit the analysis of all the data provided to SRD by the subject-matter divisions. The majority of time allotted for this study was spent on the time-consuming tasks of copying the data sets required for this study from tapes and reformatting these data into files that could be handled easily by the Statistical Package for the Social Sciences (SPSS). Thus far,

all of the data except edited administrative and tabulated census for construction and manufactures have been reformatted. In other words, the majority of these data are now available for the analysis stage.

Single-unit establishments have been the focus of this analysis and it was thought that the same type of data review would be used for multi-unit establishments and companies in the study. Most of the SPSS programs written for single-unit data could be adapted easily for multiunits and companies.

If further study is conducted, one of two approaches for further analysis could be taken. These recommended approaches are as follows:

1. Continue the analysis until sufficient data have been examined to ascertain whether patterns of differences exist. If patterns are not found, the analysis should be discontinued. If patterns are found, then determine whether further analyses are necessary for consideration of adjustment procedures.
2. Continue the analysis until all data collected for the study have been examined. If patterns of differences are found, then adjustment procedures should be considered.

2.6 Part II of this study

Results from Part II of this study will be documented in a separate report. For this part of the study, a subset of the establishments included in Part I was chosen with the goal of examining photocopies of census questionnaires and photocopies of the associated IRS records (1065,1120,941) for the same establishments. There is the possibility that the paper documents will show handwritten notes or scratchwork which might shed light on discrepancies. Keying errors might also have had some impact.

3. Sample Design and Selection

Statistical sampling was not applied in this study since it was determined that selection of specific SIC's was appropriate. Selection of cases was based on choosing SIC's where problems with administrative data were suspected to exist. Staff members from the economic areas supplied Statistical Research Division (SRD) with a list of SIC's that presented difficulties when using administrative data in lieu of census data. The SIC's, along with their geographic and trade areas, that were selected are as follows:

	<u>SIC</u>	<u>Geographic Area</u>	<u>Trade Area</u>
(1)	52 - 59	Nebraska	Retail
(2)	422, 4722, 7XXX, 80 - 806, 81, 823, 824, 829, 83, 84, 861, 862, 864, 869, 89	Nebraska	Services
(3)	50 - 51	Nebraska	Wholesale
(4)	72	Maryland	Services
(5)	73	Maryland	Services
(6)	554	Virginia	Retail
(7)	5147	United States	Retail
(8)	7372	United States	Services
(9)	5983	United States	Retail
(10)	2335	New York	Manufactures
(11)	2421	Kentucky	Manufactures
(12)	3679	United States	Manufactures
(13)	1521	Texas	Construction
(14)	1622	United States	Construction
(15)	1721	California	Construction

These selected SIC's contain approximately 50,000 cases, and include both single-unit and multi-unit establishments. Although data were collected for all areas, the data analyses were performed only on two SIC's from wholesale trade, three SIC's from selected services, one SIC from construction, and one SIC from manufactures.

4. Processing of Data

This study was divided into two parts that are related but involved different research procedures. For Part I, which is the basis for this report, comparisons were made between computerized data files containing administrative and economic censuses response data sets while Part II involved examining photocopies of census and IRS records. This report focused on Part I of the study and results from Part II will be made available in a separate document.

The original census data as keyed and the unedited administrative data for all trade areas were supplied by Economic Surveys Division (ESD) on computer tapes. Business Division (BUS), Construction Statistics Division (CSD), and Industry Division (IND) supplied computer tapes with the edited administrative and tabulated data when these data became available. Since cases in this study include only respondents to both the IRS and Census, normally the administrative data for these cases would not be included in the editing process. Although these cases were respondents to the census, they had to be treated as nonrespondents for this study and receive the same type of treatment as cases that were delinquent to the census would receive so that administrative data would be included in the editing process.

After the four kinds of data were received, records were formed for each Census File Number (CFN) in the study. These records contained identifying information and the basic items, i.e., employment, first quarter payroll, annual payroll, and receipts, for the four kinds of data. This merging of data was completed for the retail, wholesale, and services cases. For construction and manufacturing, only the unedited administrative and original census data sets have been merged.

Figure 1. Scattergram of Unedited Administrative Versus Original Census Data
Employment for Single Units in SIC 50 in Nebraska

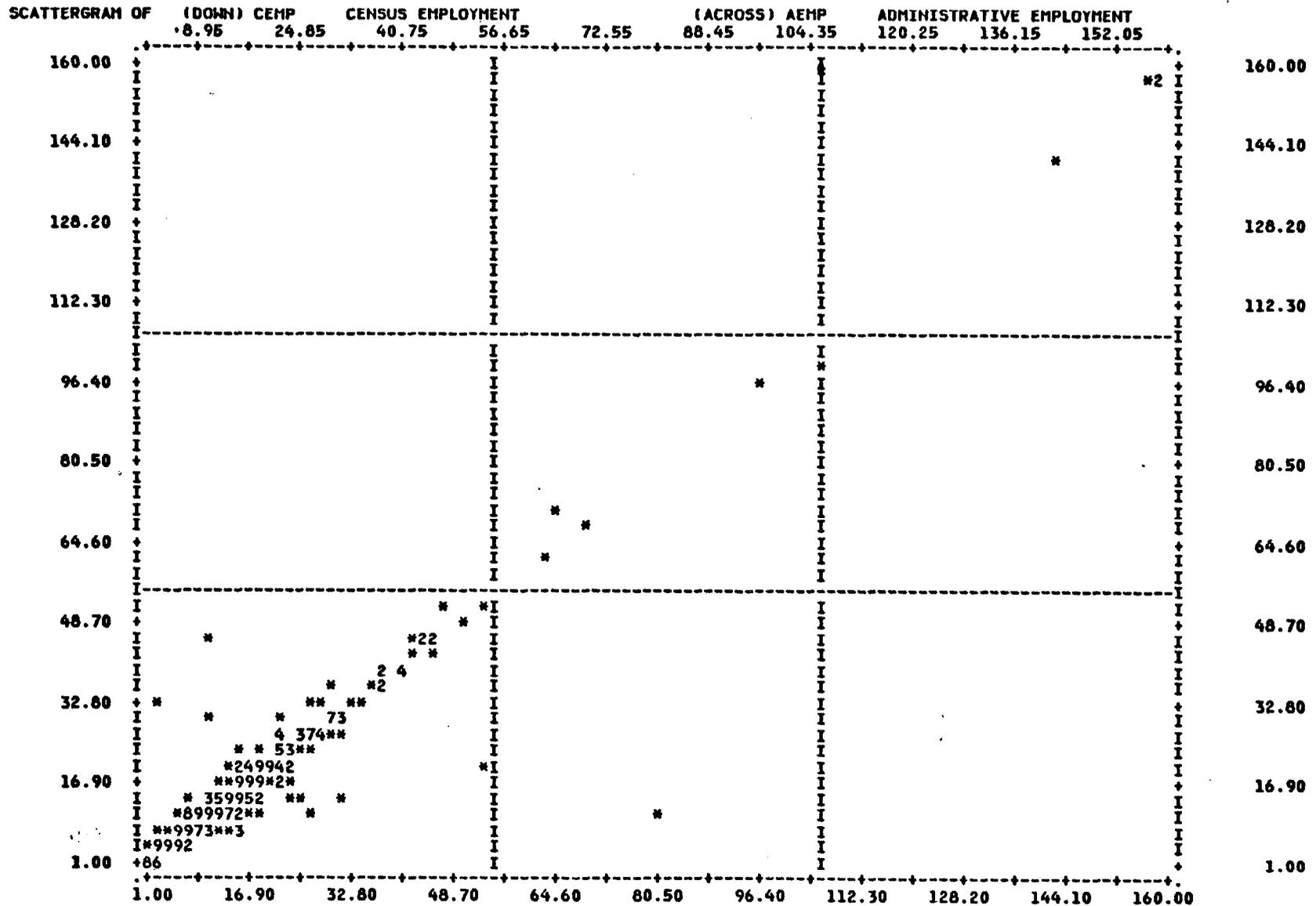
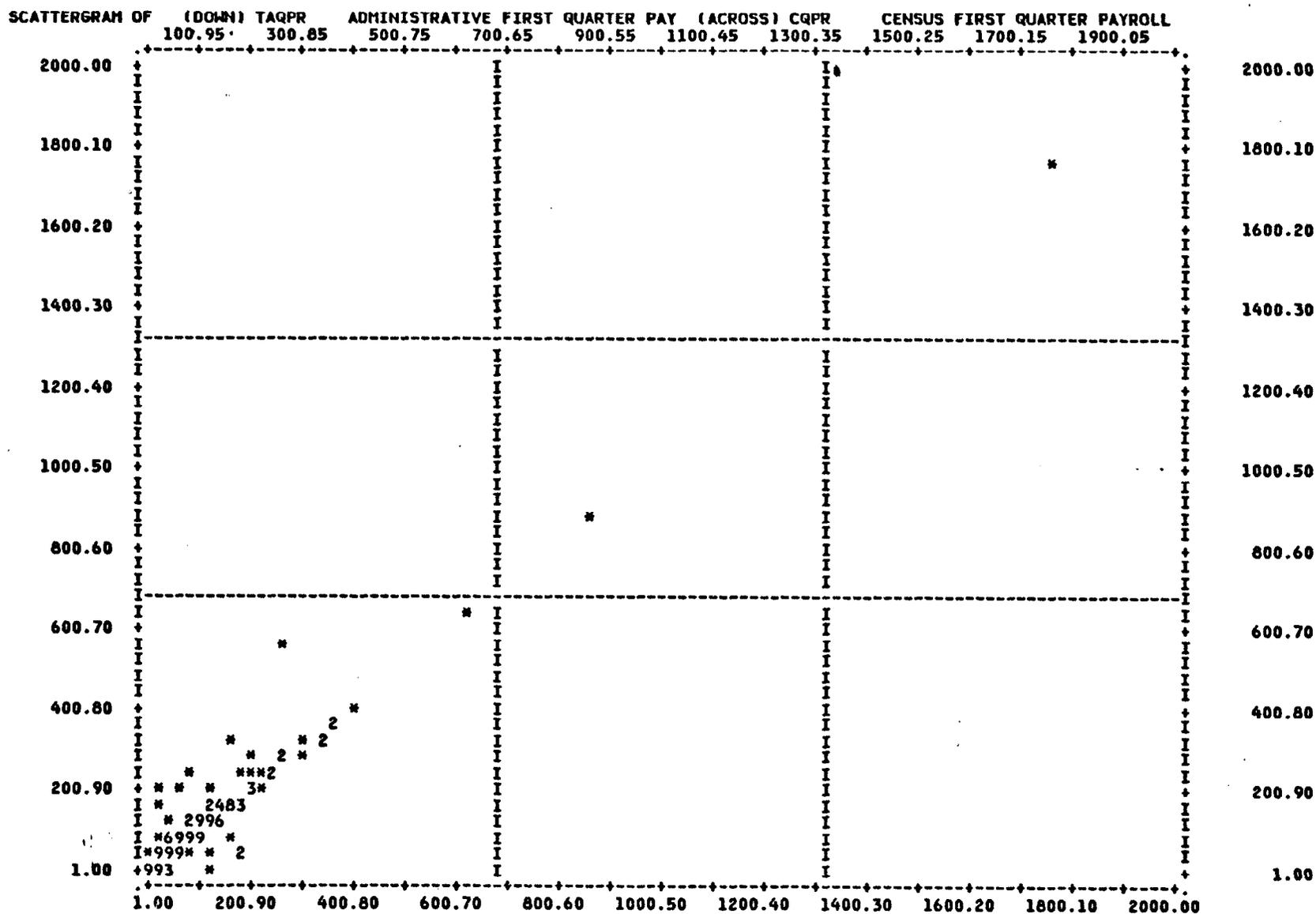


Figure 2. Scattergram of Unedited Administrative Versus Original Census Data
First Quarter Payroll for Single Units in SIC 50 in Nebraska

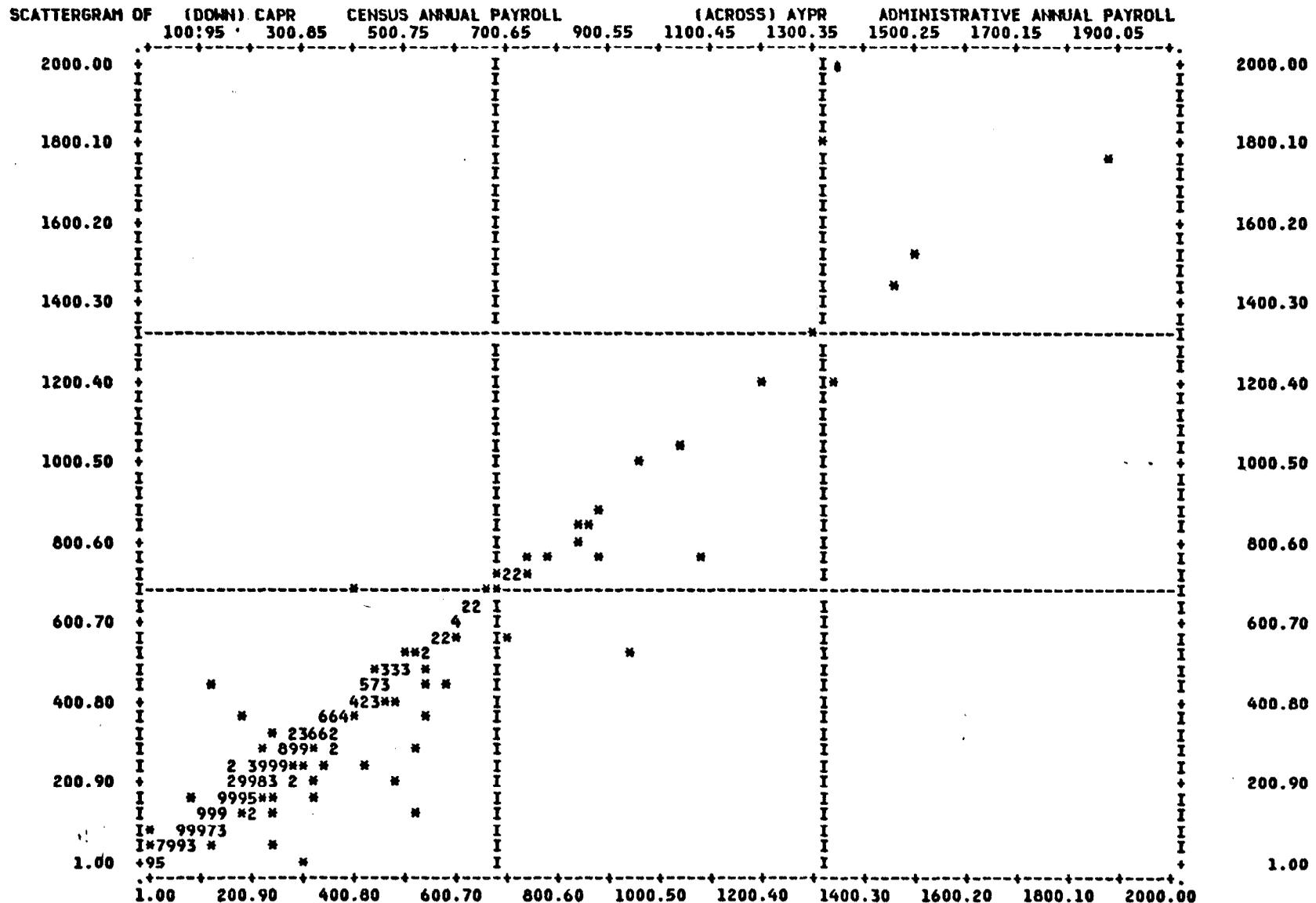


STATISTICS..

CORRELATION (R)-	.97180	R SQUARED -	.94439	SIGNIFICANCE -	.00000
STD ERR OF EST -	24.68482	INTERCEPT (A) -	2.59759	SLOPE (B) -	1.00098
PLOTTED VALUES -	537	EXCLUDED VALUES-	3	MISSING VALUES -	78

***** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3. Scattergram of Unedited Administrative Versus Original Census Data Annual Payroll for Single Units in SIC 50 in Nebraska

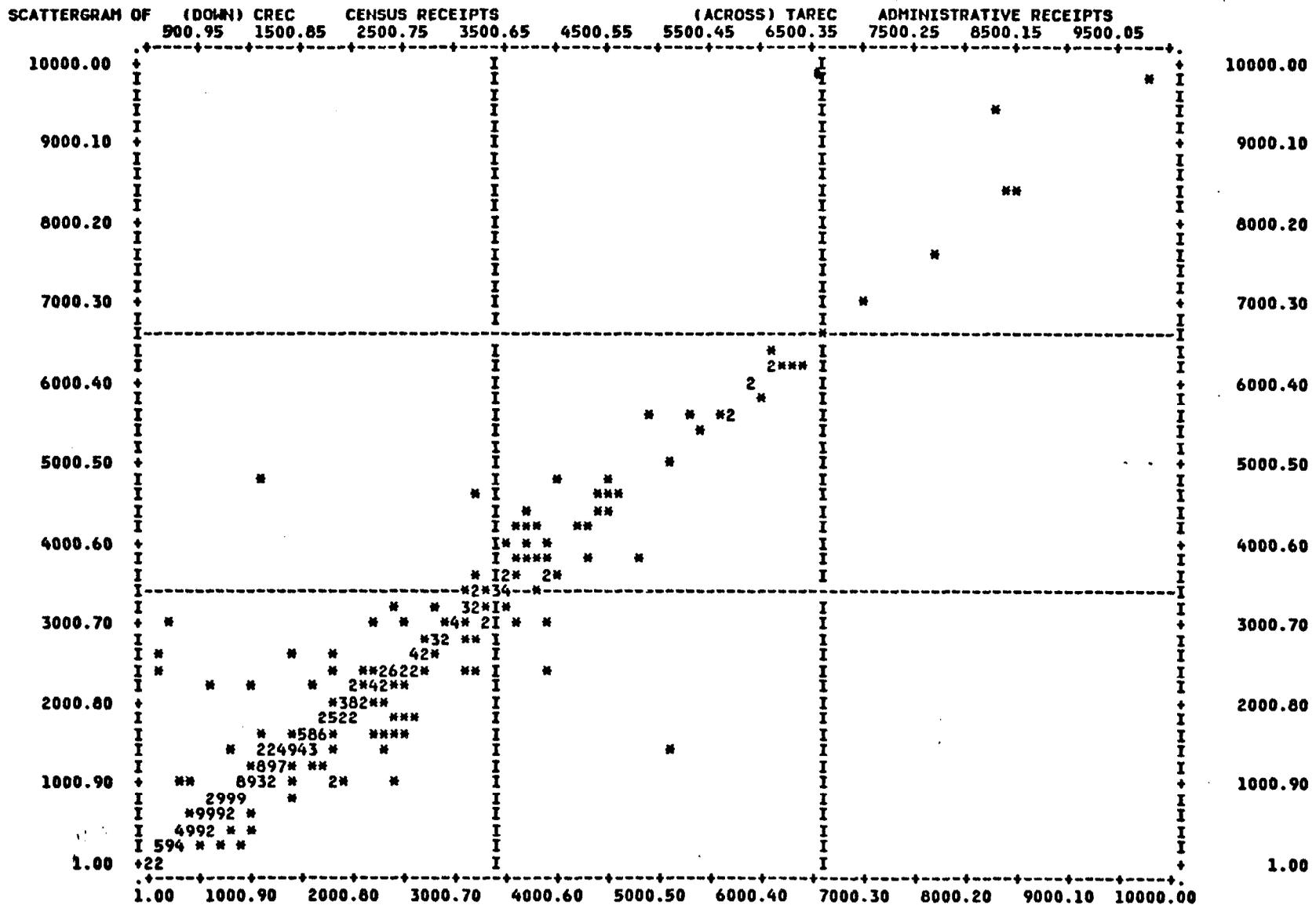


STATISTICS..

CORRELATION (R)-	.97230	R SQUARED	-	.94537	SIGNIFICANCE	-	.00000
STD ERR OF EST -	51.18929	INTERCEPT (A) -		2.47091	SLOPE (B)	-	.96243
PLOTTED VALUES -	563	EXCLUDED VALUES-		8	MISSING VALUES -		47

'*****' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 4. Scattergram of Unedited Administrative Versus Original Census Data Receipts for Single Units in SIC 50 in Nebraska

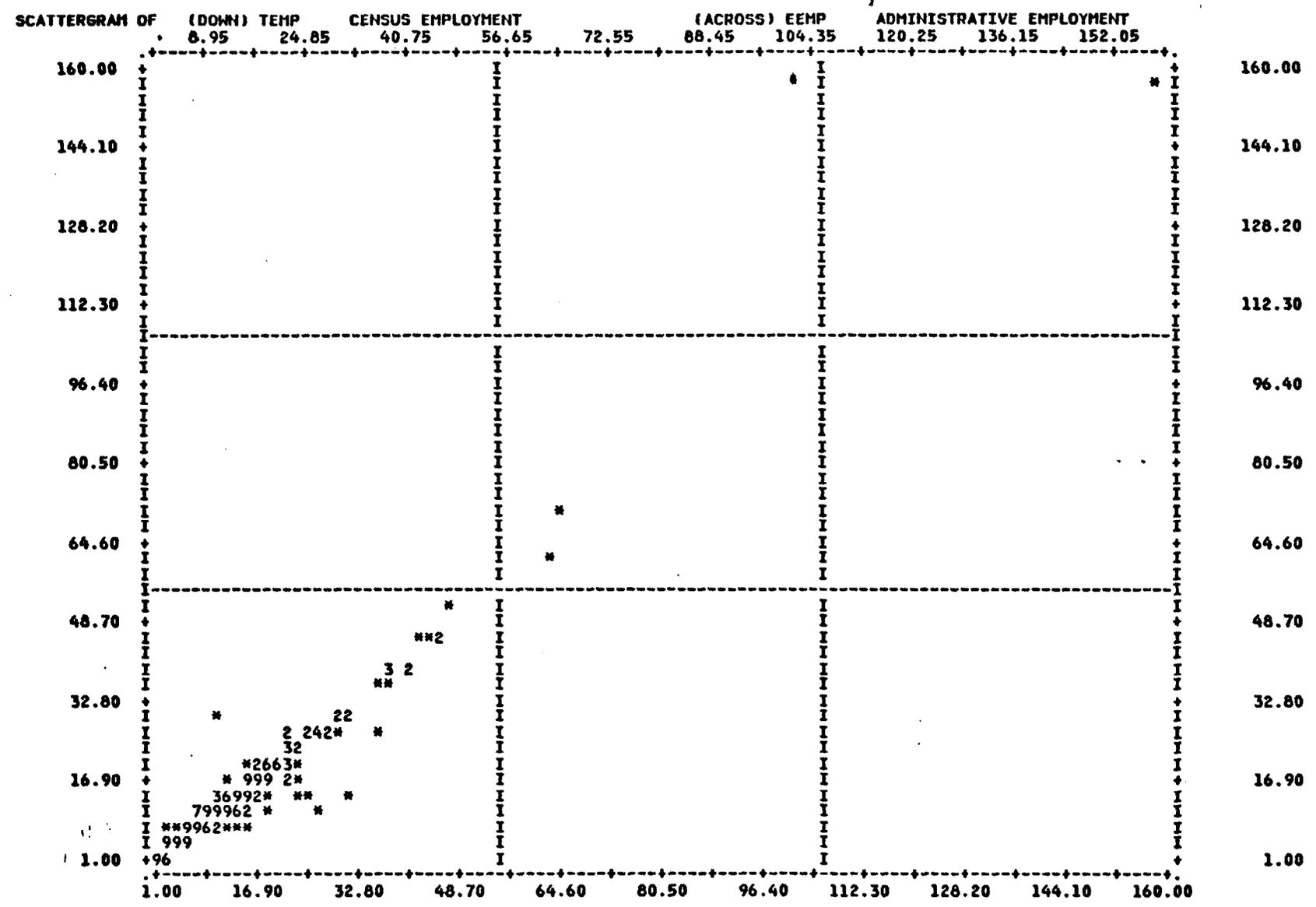


STATISTICS..

CORRELATION (R)-	.96325	R SQUARED	-	.92785	SIGNIFICANCE	-	.00000
STD ERR OF EST -	426.08069	INTERCEPT (A) -		59.33923	SLOPE (B)	-	.97234
PLOTTED VALUES -	450	EXCLUDED VALUES-		10	MISSING VALUES -		158

'*****' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 5. Scattergram of Edited Administrative Versus Tabulated Census Data Employment for Single Units in SIC 50 in Nebraska

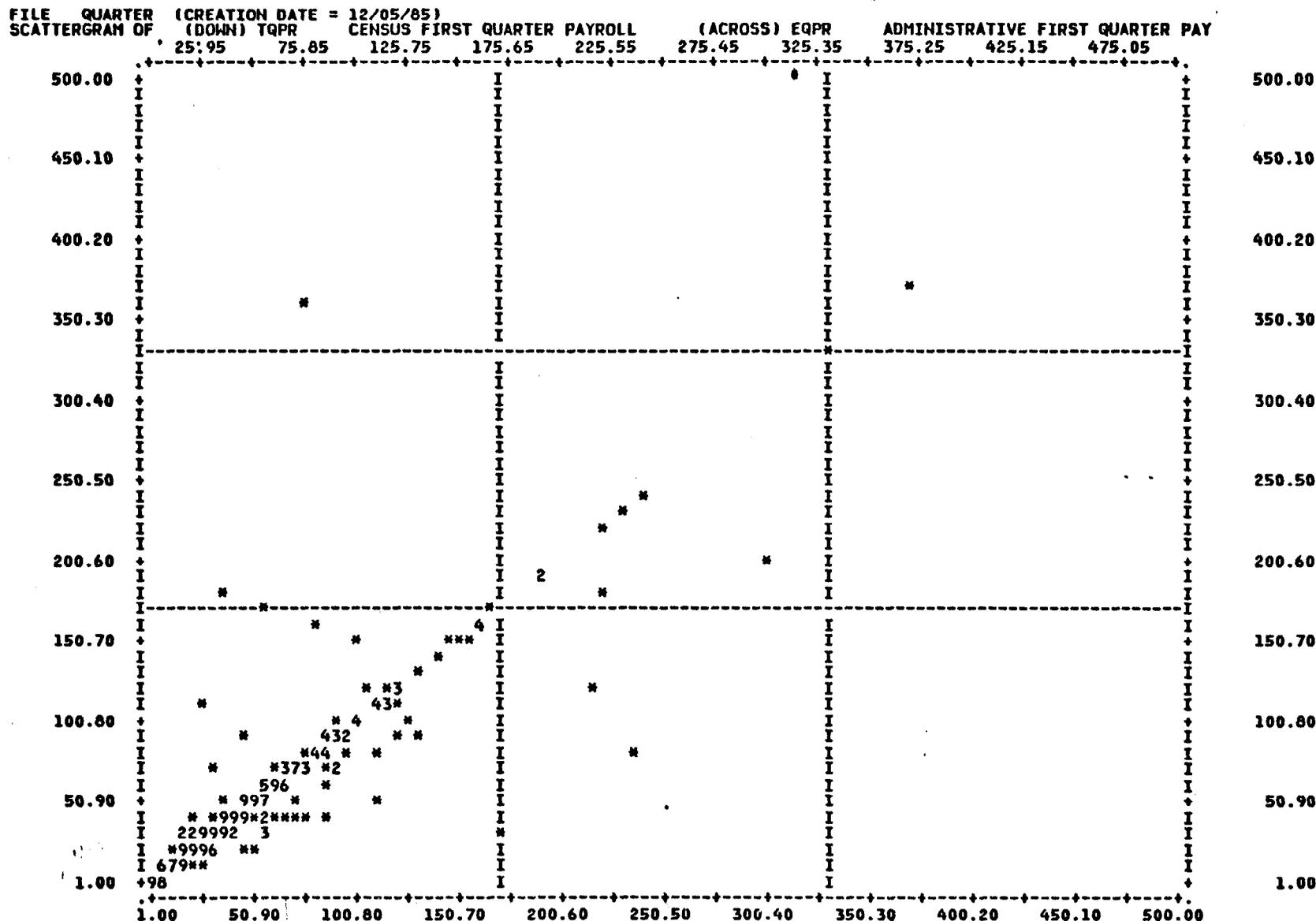


STATISTICS..

CORRELATION (R)-	.97918	R SQUARED	-	.95880	SIGNIFICANCE	-	.00000
STD ERR OF EST -	2.29066	INTERCEPT (A) -		.12069	SLOPE (B)	-	.98174
PLOTTED VALUES -	420	EXCLUDED VALUES-		0	MISSING VALUES -		198

'*****' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 6. Scattergram of Edited Administrative Versus Tabulated Census Data
First Quarter Payroll for Single Units in SIC 50 in Nebraska

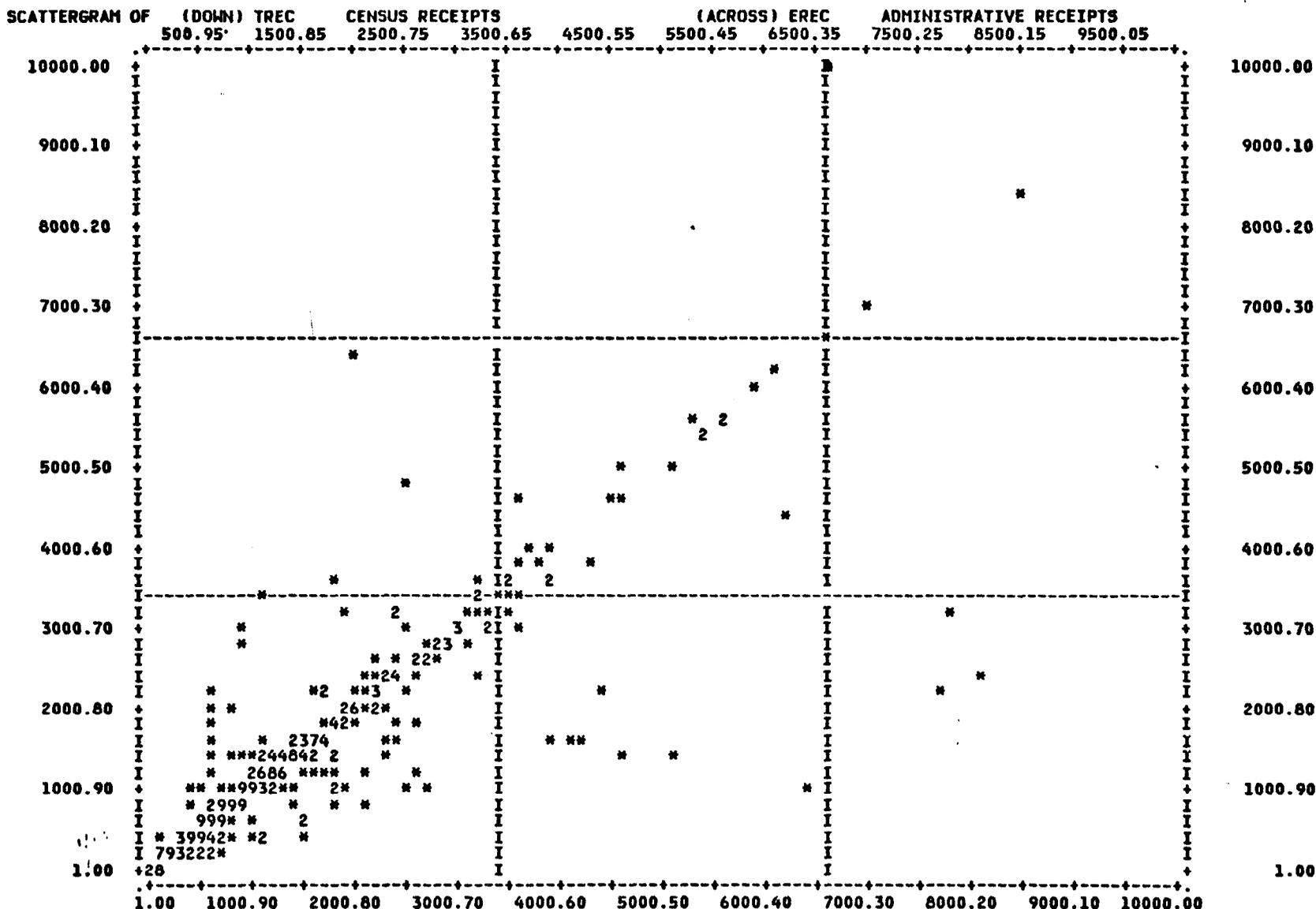


STATISTICS..

CORRELATION (R)-	.87989	R SQUARED	-	.77421	SIGNIFICANCE	-	.00000
STD ERR OF EST -	22.47116	INTERCEPT (A) -	-	4.86692	SLOPE (B)	-	.88490
PLOTTED VALUES -	420	EXCLUDED VALUES-	0	MISSING VALUES -	198		

'*****' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 8. Scattergram of Edited Administrative Versus Tabulated Census Data Receipts for Single Units in SIC 50 in Nebraska



'*****' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.