Ecological models in electoral geography: problems in interpretation

David G. Rogers

ECOLOGICAL MODELS IN ELECTORAL GEOGRAPHY: PROBLEMS IN INTERPRETATION

David G. Rogers

Dissertation submitted in partial fulfilment
of the requirements of the M.Sc. Course in
Spatial Data Analysis in Geography, University
of Durham.

May 1980

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

I would like to thank the following people for their help during the preparation and writing of this dissertation:

Dr. David Rhind, my supervisor, who has been an invaluable help during my time at Durham.

Dr. Ian Evans, whose ideas almost always proved illuminating and who made me aware of previously unsuspected problems in quantitative geography.

The staff, postgraduates, and undergraduates of the Department of Geography for providing a congenial and stimulating academic environment in which to work.

The technical staff of the department, and the staff of the Computer Unit, for all their help over the past 18 months.

Margaret Bell and Kathryn Anderson for typing this manuscript, especially at such short notice.

The Social Science Research Council for providing a one-year studentship for the session 1978-79, and for providing the data used in this study.

The domestic staff of Collingwood College, for making me feel at home during my first year in Durham, and especially for looking after me so well when I was ill.

Many of my friends provided invaluable encouragement during the long gestation period of this dissertation. There are too many people to mention by name, but they know who they are. I am one of the few people who can claim the "dual nationality" of the Huts and Skylab and would like to thank the occupants of both, no matter how much they dislike being classed together. Many thanks are also due to friends from college and from Liverpool.

Finally, and most importantly, I would like to thank my family: once again they have provided incalculable support.

David Rogers

Durham, May 1980.

CHAPTER ONE

INTRODUCTION

Although elections have a long history in human affairs, the development of mass franchise has been very much a twentieth-century phenomenon. The electoral system based on mass franchise has become important for the formation of governments in many countries of the world since that time, so it is hardly surprising that a mass of literature has built up based on the study of elections. This study has by no means been confined to one particular discipline, and the ends to which study appear to have been aimed are as wide as the range of disciplines involved.

In common with developments in other fields, the origins of electoral geography lie in the early years of the century. Also in common with other fields, the diversity of approach has been manifest. Electoral geography in Britain has been largely ignored until the recent past, with many of the major developments taking place on the continent. However much more emphasis has been placed upon electoral studies recently, and this has led to the assertion that "quite simply, electoral patterns reflect and help produce the overall human geography of a region or state." (Taylor and Johnston, 1979 p.18). To test the veracity or otherwise of this statement requires the consideration of voting choices in relation to individual circumstance and regional characteristics. In geography the main emphasis has been on the latter, which has led to the development of studies based on aggregates of voters for which we have supposedly accurate socio-economic data. Unfortunately many studies have been carried out which do not discuss the drawbacks of using data of this kind : it is the aim of this dissertation to illustrate the problems associated with these analyses.

The structure of the dissertation is as follows : first, a brief overview of the development of this type of study in electoral geography is given (chapter 2). Chapter 3 discusses the statistical assumptions of the models used and the utilization of transformations of the data to help avoid these. The following chapter (chapter 4) demonstrates the reliance placed on ratios or proportions in this type of study, and the problems ensuing from their use : discussion in chapter four and the previous one utilize data from the 1966 sample census of Great Britain in order to demonstrate the points made. Chapter 5 is a review of ecological models in general, and a particular ecological analyses in electoral geography. The final chapter discusses the results obtained and relates them to earlier chapters.

C H A P T E R   T W O

THE BACKGROUND TO THE STUDY AND THE DATA USED

> "The problem of statistical inference in geography can effectively be resolved by a careful evaluation of the test procedures available for a given set of circumstances, together with a full specification of the assumptions made in collecting and manipulating the data".
>
> David Harvey (1969, p. 286)

The use of sophisticated statistical tools in most branches of human geography has burgeoned over the past 20 years, and electoral geography is no exception to this general rule. Since 1965, the use of models based on the General Linear Model has grown considerably especially since 1973 (e.g. Bartlett (1973); Rassmussen (1973); Miller et.al. (1974); Crewe and Payne (1976)). As Harvey notes, there are problems involved in the application of such models, and other workers have identified broad problems within the whole field of geography (e.g. Poole and O'Farrell (1971)), or directly connected to specific fields (e.g. Evans, Catterall and Rhind (1975); Cliff and Ord (1975); Harvey (1968) etc.). It was concluded that a similar review and consideration of the problems was required in electoral geography.

Blondel (1974, pp. 48-50) noted that there were few examples of ecological studies of elections in Britain. Since he wrote, this situation has altered somewhat (see chapter 5), but the level of use of this type of analysis is still a long way below comparable continental studies (Taylor and Johnston 1979). One of the reasons for this is the British method of releasing voting data for national elections: this is only done on constituency level, compared to the release of statistics for much smaller units elsewhere (Blondel 1964, p. 47). Another reason is that census variables which are compared with voting patterns were not published by constituency in the United Kingdom before 1966 (Taylor and Johnston 1979, p. 85). However, this should not be taken to mean that such analyses did not take place - but only that previous work was both sporadic and largely qualitative in character.

2.1. The origins of Electoral Geography

The generally recognised pioneer of electoral geography was Andre Siegfried (Busteed, 1975, p. 23; Alford, 1967, p. 70; Taylor and Johnston, 1979, p. 24), who studied voting patterns in France between 1871 - 1912 (Siegfried, 1913). Siegfried based his work on the comparison of a series of maps. The tradition of the use of map comparison has survived to the present day (e.g. Lewis, 1965; Kinnear, 1968). After he initiated this school of study on the continent, developments in the U.S.A. soon followed, but not in Britain (Busteed, 1975, p. 23). Paradoxically, one of the first studies that closely followed Siegfried's was an analysis of British elections by an American (Krehbiel, 1916). Krehbiel not only looked at maps of coalfields etc., but at actual census statistics in his analysis; he inherited a degree of environmental determinism from Siegfried in his work ("the principles just stated for the greater part embody environmental or natural influence on man in his political action", p. 424), but made an important observation - the marked tendency towards class voting in British elections:

> "when the laboring class is most numerous in a county constituency the chances are that it will incline to the Liberal or Labor (sic) party"

The importance of class voting in British elections has been identified by many writers (Alford 1967, Pelling 1967), and is held by Jennings to extend right back to 1832, if not earlier (Jennings, 1960 pp. 327-339). The overwhelming importance of class voting in the U.K. may have served to suppress the desire for ecological analysis (Taylor and Johnston, 1979, p. 207). However, in the same article, Krehbiel notes that:

> "the really surprising fact, however, is that so large a proportion of the industrial boroughs are for the Tories. This is notably true of Liverpool, but it is conspicuous in all the large manufacturing centres".
>
> (Krehbiel, 1961, p. 430)

Here, then, is an early recognition of "working class Tories" as defined by McKenzie and Silver (1967), who stated that "it is also difficult to think of working class Conservatives as a pathetic, ignorant, or alienated people" and declared that the voting choice of the "working-class Tories" cannot be seen as "political pathology". Indeed they pointed out that working-class Conservatives tend to be better informed on political matters than the Labour counterparts (McKenzie and Silver, 1967, p. 119).

This emphasis on the social context allied to the nature of the two major British parties from the mid-20th century onwards led to an emphasis on national rather than regional studies. Indeed, Britain was regarded by many as socially and politically homogeneous (Blondel, 1974). Hence studies of elections tended to concentrate on the operation of the electoral system, or on the view of elections as a national event deciding the nature and outlook of the government (Busteed, 1975). Therefore, electoral behaviour was studied in the national context (e.g. McCallum and Redman, 1951). However, even given this and allowing for the importance in social class in determining electoral choice, this does not mean that the spatial context can be ignored; as Johnston (1979, p. 179) states:

> "British parties have a strong class bias to their support and social classes . . . are spatially segregated".

2.2 Recent Developments

Since the middle nineteen-sixties, the emphases in electoral studies have become more broadly based. Miller et al., (1974, p. 384) state :

> "The formation of political partisanship is a social and historical process, not only, or even mainly, an individual response to personal, social or economic influences"

The importance of the social and historical perspective has been recognised by Lipset and Rokkan (1967). They see the development of the political parties as the representation of the outcome of conflicts, or "cleavages" in society: their views are very adequately summarised in Taylor and Johnston (1979, chapters 3 and 4). Lipset and Rokkan represent the conflicts by two axes which define the "conflict space". One of these orthoganal axes is a "functional" one, concerning conflicts based on different interest groups in the whole country: these are socio-economic conflicts and produce cleavage patterns based on the different group-roles in society (Taylor and Johnston, 1979, p. 122). The other axis is a "territorial" one, representing the conflicts between different regions in a country: thus it has a distinctive geographical flavour.

Within this framework Lipset and Rokkan identify four central cleavages (Lipset and Rokkan, 1967, p. 14), two of which they term products of the "National revolution" (subject v. dominant culture; church v. state), two as a product of the industrial revolution (landowners v. capitalists; workers v. employers). Taylor and Johnston (1979, p. 116) state that "modern party systems . . . derive from these four conflicts". There is an important distinction to be made between the policies of the parties and the political stance of the electorate who support them. Jennings (1960, pp. 288-291) has pointed out that most electors "vote not for a party policy but for a party image". However the parties' various positions on "critical" issues usually is of importance in the determination of party images, and thus political partnership is closely associated with party policy.

Taylor and Johnston (1979, p. 166) report work by Lijphart (1971) who derived three indices of the bases of voting analogous to three major Rokkan cleavages: Religion; urban/ rural; and class. Taylor and Johnston state that "the size of the indices within each country indicates the relative importance of Rokkan's cleavages in modern voting behaviour (1979, p. 166: see table 2.1). However, it must be borne in

Table 2.1: Cleavages and the social bases of voting, c. 1960

Country	Religion	Rural/urban	Class
A. Anglo-American Countries			
Britain	7	10	37
U.S.A.	16	11	20
Canada	22	-	8
Australia	14	-	33
B. Scandinavian Countries			
Sweden	16	-8	53
Norway	21	2	46
Denmark	-	-	44
Finland	-	-	59
C. Other European Countries			
France	59	11	15
Italy	51	12	19
West Germany	40	17	27
Netherlands	73	10	26
Belgium	72	7	25
Austria	54	22	31
Switzerland	59	-	26

Source: Taylor and Johnston (1979, p. 165) (from Lijphart, 1971) mind that, owing to problems of derivation of the indices they are unlikely to be mutually exclusive. Taylor and Johnston (1979, p. 167) state that "in all cases the rural/urban cleavage seems to be the least important". This is true for all bar one county in the table - Britain. The main class cleavage has already been detailed above. Rural/urban cleavage is seen by Lipset and Rokkan to develop from the territorial cleavages. Examples of the territorial cleavage in politics are manifold. An example from Britain was the demand of the overwhelming majority of Irish voters and M.P.'s to secede from the union with Britain in the 19th century through to 1921. In Ireland itself, this could be confused with a religious cleavage, but cleavage also took place within the politics of the rest of the U.K., with the relative positions of the home rule debate of the Liberals and the Conservatives after Gladstone had declared for Home Rule for Ireland in 1885. Further, this caused a split in the Liberal party itself, which divided into Liberals and Liberal Unionists (Lyons, 1973, pp. 293-294). A similar situation can also be seen today within the U.K. with the development of nationalist parties in Wales and Scotland (and, of course, Northern Ireland), but the cleavage also seems to have transcended the national scale with Britain's entry to the European Economic Community: the division of opinion in the Labour party over the issue seems not dissimilar to the position the Liberal party found itself in 100 years ago. Given these issues, it seems a little rash to subsume the "core - periphery" cleavage totally within a rural-urban one. However, despite these problems, the concept of cleavage can be seen as offering a useful framework for electoral analysis.

2.3 Testing the models: the data available

Miller et. al., (1974, p. 384) state :

> "We need to examine environmental influences explicitly and test for them empirically before inferring a socio-political relationship".

The data available for the type of testing that Miller et. al. call for are available in two forms: voting records and censuses; and survey research. Most previous analyses have utilised the aggregate analysis approach of using census data, chiefly due to the logistical and financial problems of obtaining accurate survey results. The review of survey methods and subsequent analyses is beyond the scope of this dissertation and have been reviewed in depth elsewhere. (Taylor and Johnston, 1979, pp. 92-102; see especially Butler and Stokes, 1969). The lack of suitable survey data led to concentration on census sources and thus to development in electoral geography of a "tradition of viewing election results in relation to the socio-economic and demographic features of the constituencies in which they occur". Early analyses were based on a cartographic approach, but there has been an increase in the use of statistical tests to examine the relationships between voting patterns and socio-economic factors (Busteed, 1975, p. 33). The development of this type of analysis has been reviewed by Busteed (1975, pp. 32-40) and by Taylor and Johnston (1979, pp. 72-102), and some of the models considered in greater detail in chapter 5.

2.4 The data used in this study

The census data used in this analysis are taken from the 1966 10% sample census of population, conducted in April 1966 by the Office for Population Censuses and Surveys (OPCS). Census information has been collected on the population of the U.K. at 10-yearly intervals since 1801, with the exception of 1941, the interruption being due to the Second World War. In addition to the decennial census, a 10% sample census was carried out in 1966: it is from this census that data for this analysis are taken. Since 1961, censuses have collected a wide range of socio-economic data which are of interest to the electoral geographer.

The accuracy of the data in studies of this nature must be an important consideration. OPCS attempt to check the accuracy of their data, especially post-enumeration surveys (Benjamin, 1970, ch. 13), but bias may be present. As the census is only a sample, this also introduces the possibility of imprecision. Miller et al. (1974, p. 391) have concluded that the sample was "sufficient to make the sampling error negligible in our analysis", but Benjamin reports a $1\frac{2}{3}$% shortfall in population enumeration in the 1966 census (Benjamin, 1970, p. 15), but concludes that to regard the shortfall as a "serious error" is "hardly justifiable" (1970, pp. 15-16).

The voting figures used are for English constituencies in the 1966 General Election, which took place on 31st March 1966. Thus the problems correlating the census data to the electoral data are minimal, such is the temporal proximity of the two events.

Voting returns and census data were provided by the Social Science Research Council archive. As noted above, earlier censuses are unsuitable for such analyses, data not being available at constituency levels and data on a constituency scale were not available from SSRC for the 1971 census. The variables supplied put a serious constraint on the nature of the analysis, some of which were rejected - mostly those measuring aspects of the male population (the corresponding variables relating to the total population were retained). This left a list of 27 putative predictor variables: these are listed and defined in appendix 1: see also list of acronyms (table 2.2). The constraints put on the analysis by restriction to the variables is obvious - no detailed demographic data, for instance, are available. The data are further discussed in chapters 3 and 4.

2.5 The 1966 General Election

The 1966 General Election closely followed that of October 1964 which had given the Labour party a slender absolute majority of four, which oscillated between one and five during the lifetime of the Parliament. The voting at the previous election had been 43.4% Conservative, 44.1% Labour and 11.2% Liberal. The 1966 election saw a 3.1% swing from the Conservatives to Labour, returning them with a 110 seat majority over the Conservatives in the House of Commons. The election took place on 31st March 1966. Only results in England were analysed in this study in order to remove the affects in the analyses of the Nationalists. The voting in

Table 2.2: Acronyms used for variables in the analysis

ACRONYM	VARIABLE
PROF	Professional Workers
EMPL	Managers and Employers
NONM	Non-manual workers
SKIL	Skilled workers
SEMI	Semi-skilled workers
UNSK	Unskilled workers
AGRI	Agricultural workers
MIN	Miners
MFG	Workers in manufacturing industries
TRANS	Workers in transport industries
DIST	Workers in distribution services
GOVT	Workers in local and national government
N CAR	Households without a car
OWNOCC	Households in owner-occupied accommodation
COUNCL	Households in council accommodation
PRIV	Households in privately rented accommodation
AMEN	Households with exclusive use of amenities
HDENS	Households in high density living
RM3	Small households (less than 4 rooms)
RM7	Large households (more than 6 rooms)
IRISH	Irish nationals (including Northern Irish)
NEWCOM	New commonwealth born
YOUNG	Young people (15-24)
OLD	Old people (65 or over)
UNEMMA	Unemployed males
INMIG	In-migration to area
WITHMIG	Migration within an area
PCC66	Conservative vote 1966
PCLAB66	Labour vote 1966
PCLIB66	Liberal vote 1966

Fuller details of the variables and their definitions are given in Appendix 1.

England was Conservative 42.8%; Labour 47.9%; and Liberal 8.6% (in seats where the Liberals stood, the mean Liberal vote was 16.3%).

CHAPTER THREE

DATA TRANSFORMATION

The question of data transformation in human geographical studies is one that has been subjected to much debate (cf. Johnston, 1978) but remains one that has yet to be satisfactorily resolved. Discussion as to the desirability and effects of the transformation of data in studies of a similar nature to this one are at one of two extremes. In electoral studies little or no account of the various transforms available for use has been made - in most cases, if transformation has been carried out, the transform used has not been reported in the published work. Results and analyses have been presented and discussed in terms of the original variables with no reference made towards the problems of the correct interpretation of the results obtained from the transformed variables or of the statistical problems encountered in the analysis of untransformed variables. In census studies the problems of transformation have been treated in greater depth (e.g. Evans et. al, 1975), although the problem of inference still remains.

3.1 Why Transform?

Attempts to erect models of explanation and prediction in electoral studies are dependent for testing upon the use of a range of statistical tests such as correlation and regression : most of these tests are members of a family of techniques based on the General Linear Model (Johnston, 1978; Mather, 1976 ch.2). The use of such "classical" (parametric) statistical tests rests upon a number of assumptions about the data used to test the model (Poole and O'Farrell, 1971; Hoyle, 1973; etc.). Poole and O'Farrell (1971 p.148) list six critical assumptions that are required to be satisfied for the derivation of the best linear unbiased estimators of the value of the dependent variable in the regression. They are :

(i) There is no measurement error present in the observations.

(ii) The relationship between the dependent variable (Y) and each of the dependent variables (X_i) is linear.

(iii) The mean of the residual values about the regression line for each value of X_i is zero.

(iv) The variances of the above distribution of residuals (U) around the regression line for each value of X_i are equal (this is known as the homoscedasticity assumption).

(v) The independent variables, X_i, are linearly independent of each other. The absence of such independence is known as multicollinearity.

(vi) The values of the residuals are serially, i.e. all values of U are independent of each other, and the residual for one value of X_i cannot be predicted by knowing the value of any other residual. This effect is known as the problem of autocorrelation.

Finally, Poole and O'Farrell identify a seventh assumption which is important if the regression model is to be used for inferential as well as for predictive purposes :

(vii) The overall frequency distributions for both the dependent and the independent variables are normal.

It is obvious from a cursory glance at the data that most of these assumptions have been violated. In the data set used in this analysis, five errors were found in the voting returns, and as the size of the data set prohibits any non-automable search for any less obvious errors, it is possible that many have gone undetected. Although actual measurement (as opposed to transliteration) error for the voting returns can be assumed to be nil, this is not necessarily the case for the census variables, where they would be far more important (Poole and O'Farrell, 1971, p.149): however the random sample of 10 per cent is sufficient to overcome serious problems of imprecision and bias (Miller et. al., 1974, p.391), but the recording of these variables may be subject to an unquantifiable error. Multicollinearity is obviously present in the data given the high correlations present between many of the "independent" variables (see Chapter six), and spatial autocorrelation is probably also present. Transformation can do little to help circumvent these problems, but is particularly important in ensuring linearity and normality in the data set.

As has been mentioned, any violation of these assumptions must render any results pursuant from such analyses open to question. Hence methods should be sought to ensure that the data -within physical and logistical constraints - meet the requirements of the model as closely as possible. A change of measurement scale used can be of help in modifying the data values if they do not fit the assumptions with the linearity (additivity) assumptions, if the relationships are curvilinear a linear model will not describe them adequately, hence a transformation to a linear relationship is desirable. This assumption can be tested visually by the inspection of scattergrams. The normality assumption can be tested by the inspection of certain descriptive statistics; namely the kurtosis and (especially) the skewness. Most transformations applied attempt to minimise skewness. The presence of non-normality in the data was considered by Poole and O'Farrell (1971 p.155) : "This assumption may frequently be relaxed. This is because such statistical inference procedures are not particularly sensitive to departures from normality", especially when samples are large (as is the case here). However Evans et . al. (1975 p.6) point to a considerable improvement in the quality of the correlation co-efficient after certain variable-specific transforms.

Inspection of the descriptive statistics for the untransformed variables (see Table3.1) shows that many of the variables used in the analysis were highly positively skewed. Such a departure fron normality was regarded as a serious violation of the normality assumption and it was decided that transformation was needed to bring the frequency distributions closer to normality.

3.2 Which Transform?

After the decision to transform has been made, the question then becomes which transform to adopt. The question lies between the application of a "blanket" transformation

Table 3.1: Mean, skewness, and kurtosis for the variables used in the analysis

Variable	Mean	Skewness	Kurtosis
PROF	4.486	1.340	2.264
EMPL	11.209	0.574	-0.146
NONM	17.817	0.557	-0.280
SKIL	39.458	-0.564	1.296
SEMI	18.261	0.725	1.357
UNSK	8.763	11.448	3.613
AGRI	3.003	2.363	5.833
MIN	2.007	4.248	20.463
MFG	44.101	0.083	-0.705
TRANS	6.713	1.478	2.874
DIST	37.782	0.586	0.122
GOVT	5.896	2.822	11.495
NOCAR	55.384	0.119	-0.805
OWNOCC	45.540	-0.701	0.276
COUNCL	25.620	1.134	1.926
PRIV	23.729	1.666	2.341
AMEN	71.444	-0.925	0.410
HDENS	5.639	1.943	4.650
RM3	11.143	2.749	8.329
RM7	11.416	0.438	-0.252
IRISH	2.005	2.583	8.452
NEWCOM	2.009	2.554	6.674
YOUNG	18.818	0.397	3.683
OLD	16.206	1.756	5.559
UNEMMA	0.955	3.288	20.587
INMIG	14.829	0.659	1.053
WITHMIG	15.735	0.334	0.048
PCC66	42.097	-0.661	0.204
PCLAB 66	49.272	0.038	-0.385
PCLIB 66	16.033	2.045	6.168

(i.e. the same transformation applied to every variable) or "variable-specific" transforms (i.e. choosing the transformation that minimises skewness to the greatest degree for each variable (Clark 1973)). This problem has been discussed in relation to grid square census data by Evans et. al. (1975), and further developed by Evans (1979). Evans et. al. (1975 pp.6-8) conclude that variable-specific transforms were desirable. Although there were substantial differences between the data analysed here and the data analysed by Evans et. al. (despite both being sets of census data : these differences will be expanded upon later) it was decided to test four different transforms : square root, logarithm, angular, and logit.

(i) The square root transform. This is simply replacing the original value by its square root :

$$X_1 = \sqrt{(X)} \qquad \text{(Hoyle, 1973, p.207)}$$

where X_1 = transformed value, X = original value (throughout)

(ii) The logarithm transforms given by

$$X_1 = \log_{10}(X) \qquad \text{(Hoyle, 1973 p.207)}$$

(Note that this transform was performed using $\log_{10}$ and not $\log_e$ as Hoyle suggests).

(iii) The angular transform, given by

$$X_1 = \sqrt{\frac{(X+W_1)}{(N+W_2)}} \qquad \text{(Hoyle, 1973 p.209)}$$

i.e. the angle whose sine is the square root of the original proportion. Note that in this analysis W_1 and W_2 have been set to zero; N equals the sample size (constant)

(iv) The logit transform, given by

$$X_1 = \log_e \left(\frac{X}{1 - X}\right) \qquad \text{(Hoyle, 1973 p.212)}$$

All of these transforms tend to aid re-expression to linearity (Box and Cox, 1964 p.212) and tend to lessen positive skew, although they may well worsen the skewness of variables with negative skew. Distinction has to be made between the one-bend transformations (including the log and square root) and two-bend transformations (including the arcsin and logit) : the names are derived from the appearance on a graph of the plot of the transformed values against the original. The basic difference between the two is that the one-bend transforms "compress" one end of the data values disproportionately, whereas in the two-bend compression takes place in the mid-range values (for a further discussion see Kruskall, 1968).

The effects of these transformations on skewness and kurtosis are given in tables 3.2 and 3.3. Transformation shows a great improvement towards normality, although some values still remain high. In most cases the choice for the best transform lies between the logarithm (log) and the logit transformation. Variables are best left untransformed (at least by these transforms) in four cases, square-root transformed in three, and angular transformed in one. In very few cases (e.g. SKIL) does the log or logit transform result in a markedly worse value than either the untransformed or otherwise transformed variable.

A decision now has to be taken as whether to apply a specific transform for each variable or a blanket log or

Table 3.2: Effects of transformation on skewness of variables used

	UNTRANS-FORMED	LOG	SQRT	ANGULAR	LOGIT
PROF	1.34	-0.21	0.58	0.62	-0.14
EMPL	0.57	-0.36	0.12	0.18	-0.25
NONM	0.56	-0.06	0.26	0.33	0.08
SKIL	-0.56	-1.67	-1.03	-0.78	-1.04
SEMI	0.73	-0.27	0.23	0.35	-0.05
UNSK	1.45	-0.09	0.68	0.76	0.06
AGRI	2.36	0.19	1.20	1.25	0.24
MIN	4.25	0.87	2.24	2.51	0.94
MFG	0.08	-0.57	-0.22	0.01	-0.07
TRANS	1.48	0.31	0.87	0.92	0.40
DIST	0.59	0.04	0.31	0.51	0.42
GOVT	2.82	0.61	1.56	1.68	0.76
NOCAR	0.12	-0.28	-0.08	0.21	0.33
OWNOCC	-0.70	-2.91	-1.50	-1.06	-1.79
COUNCL	1.13	-0.66	0.34	0.72	0.10
PRIV	1.67	0.55	1.15	1.41	1.08
AMEN	-0.93	-1.51	-1.20	-0.63	-0.24
HDENS	1.94	0.54	1.21	1.27	0.63
RM3	2.75	0.85	1.78	2.04	1.24
RM7	0.44	-0.51	-0.04	0.02	-0.40
IRISH	2.58	0.11	1.35	1.39	0.16
NEWCOM	2.55	0.17	1.47	1.51	0.22
YOUNG	0.40	-0.36	0.01	0.10	-0.19
OLD	1.76	0.76	1.23	1.35	0.97
UNEMMA	3.29	0.17	1.33	1.35	0.19
INMIG	0.66	-0.68	0.00	0.12	-0.43
WITHMIG	0.33	-1.59	-0.31	-1.20	-1.20
PCC66	-0.66	-1.05	-	0.79	0.69
PCLAB66	0.04	0.06	-	0.00	0.04
PCLIB66	2.05	0.79	-	3.42	1.04

Table 3.3: Effects of transformation on kurtosis of variables used

	UNTRANS-FORMED	LOG	SQRT	ANGULAR	LOGIT
PROF	2.26	-0.00	0.24	0.33	-0.01
EMPL	-0.15	-0.37	-0.57	-0.53	-0.43
NONM	-0.28	-0.46	-0.51	-0.46	-0.48
SKIL	1.30	6.72	3.11	2.15	3.45
SEMI	1.36	0.58	0.62	0.78	0.60
UNSK	3.61	0.78	1.14	1.38	0.80
AGRI	5.83	-1.21	0.67	0.87	-1.16
MIN	20.46	0.37	5.95	6.81	-0.16
MFG	-0.71	0.10	-0.50	-0.59	-0.43
TRANS	2.87	-0.09	0.83	0.99	0.03
DIST	0.12	-0.45	-0.30	0.06	0.01
GOVT	11.50	0.50	3.76	4.41	0.94
NOCAR	-0.81	-0.73	-0.84	-0.61	-0.29
OWNOCC	0.28	11.04	2.72	1.47	4.97
COUNCL	1.93	1.73	0.39	1.32	1.46
PRIV	2.34	0.04	0.84	1.68	1.06
AMEN	0.41	2.22	1.14	0.13	0.11
HDENS	4.65	0.20	1.66	1.87	0.37
RM3	8.33	0.44	3.27	4.64	1.66
RM7	-0.25	-0.49	-0.73	-0.68	-0.55
IRISH	8.45	0.25	2.26	2.41	0.29
NEWCOM	6.68	-0.31	1.91	2.06	0.26
YOUNG	3.69	2.99	3.00	3.16	3.01
OLD	5.56	1.59	3.05	3.62	2.27
UNEMMA	20.59	0.46	4.01	4.24	0.50
INMIG	1.05	0.17	-0.17	0.02	0.05
WITHMIG	0.05	7.25	0.44	0.33	4.96
PCC66	0.20	1.51	-	-0.83	-1.05
PCLAB66	-0.39	0.70	-	0.04	0.06
PCLIB66	6.17	1.51	-	1.40	0.79

logit transform. Evans et. al. (1975) and Evans (1979) argue for a variable-specific transformation but there are significant differences between the data set used by them and the one used in this study. For instance, original skewness levels are far less variable in this data set. This is almost certainly a scale factor : the smaller grid-squares show far much more internal homogeneity and therefore differences between them are highlighted compared to the more heterogeneous constituences. This increases the likelihood at an overall transform performing reasonably well over the whole set of variables. Also, the types of variables utilised in the two Evans analyses are divisible into three groups (Evans, 1979, p.148) : (1) absolute numbers; (2) ratios with no upper limit; (3) closed ratios. Evans concludes (1979, p.148):

> "all three types of variable have different frequency distribution and no single transform would be appropriate."

However this is not the case in this study : all the variables in the analysis here are closed ratios, so this argument does not apply.

With all transformed variables there is a problem with interpretation of the results presented in transformed format. The problem is discussed at some length by Hoyle (1973) Evans et. al. account this a "minor problem" (1975, p.6), and state (p.7) :

> "the objection to specific transformation is often rationalized as interpretability. How, the question is mockingly posed... can the cube root of one variable be related to the reciprocal or logarithm of another?"

This statement seems harsh, especially in its assessment of most qualms over interpretability as "rationalization". Although the article defends the use of adopting variable-specific transforms, the question of interpretability is not really answered. As is pointed out, relationships can be interpreted in the same "way", i.e. have the same sign (barring reciprocal transformations). There are severe problems, however, in the interpretation of multiple regression equations. These attempt to model the percentage vote for a particular party in terms of a constant plus differing contributions from the predictor variables used in the equation; analyses such as that of Crewe and Payne (1976) and Miller et. al. (1974) attempt to explain the percentage share of the vote in terms of the untransformed variables. These analyses are obviously complicated if, for instance, instead of explaining the percent Conservative vote in terms of the per cent middle class and the percent owner-occupiers in the constituency (Miller et. al., 1974 p.404), one was explaining the square root of the percentage Conservative vote in terms of the logit of the percent middle class and the log of the percent owner-occupiers. This is <u>not</u> to question that such a relationship can exist but rather (<u>pace</u> Evans et. al.) the correct analysis of the regression co-efficients when it does. It must be added, however, that in a blanket transformation there still remain problems of interpretation of co-efficients but the variables are all at least treated in an identical manner, so there are fewer problems of comparison.

It was thus decided to adopt a blanket transformation. Of the transforms tried, the choice obviously lies between a log and a logit transformation : there appears little to choose between them. The logit transform was chosen - the overall reduction in skewness and kurtosis is slightly better than for the logarithm transform, and it performs slightly better in cutting down the very high values. Also a logit transform is helpful in alleviating problems due to ratio correlation (Evans 1979 : see following chapter). It is noticeable, however, that even after such a powerful transform as the logit, skewness can still be high (e.g. OLD (0.97), PRIV (1.08) etc.), such is the nature of the data, but in general the reductions in skewness values are impressive. As kurtosis is usually related to the modulus of skewness (Evans et. al., 1975 p.2) a concomitant reduction of kurtosis values was noted.

3.3. The Effects of Transformation

Evans et. al. (1975 p.2) point out two ways in which transformation to reduce overall skewness produces an improvement in the correlation co-efficient. Transforms to reduce skewness generally cause a reduction in the curvilinearity of a relationship, and this tends to produce a higher correlation co-efficient. Conversely transformation decreases the influence of outliers on the statistic, and thus tends to produce a lower co-efficient. With these data, skewness values are not as high as with the grid-square data analysed by Evans et. al., and therefore one would expect there to be fewer outliers : inspection of scattergrams confirm this. Also, from scattergrams, few relationships are markedly curvilinear, but some of the relationships amongst the independent variables (especially those affected by ratio correlation) are greatly helped in this respect.

Transformation does produce marked effects on the resultant correlation structure (see correlation matrices in Appendix 3). In general, transformation improved the strength of the relationship, but there were many relationships whose correlation declined after transformation. Of the 351 correlations in the 27 x 27 correlation matrix of "independent" variables, 90 had their correlations improved by more than 0.05 (if significant); or had correlations previously not significant at the 0.05 level of significance which became so. Forty-nine correlations decreased by the same margin or became non-significant. As can be expected from a skewness-minimising transformation, the correlations most affected were those involving variables with initially high skewness levels, e.g. PROF, IRISH, NEWCOM, etc : for these variables many of their correlations with other variables changed markedly. In contrast, variables such as NOCAR which were fairly normally distributed before transformation had correlation co-efficients which tended to be fairly stable, except for those correlations with the highly skewed variables (e.g. IRISH, NEWCOM). In a few cases, as has been noted above, the transform adopted has actually made the skewness worse : this is especially so in the case of UNSK. From comparison of the two correlation matrices, it is noticeable that five of the correlations of UNSK with other variables are markedly lower (>0.05) after transform and only two were higher - the reverse behaviour to that observed with most of the other variables.

The relationships were then examined in greater detail, with the inspection of bivariate scattergrams(for all scattergrams discussed, see Appendix 5). The points made by Evans et. al. (1975) were in many cases seen to have been borne out. Reductions in curvilinearity do not seem to be too important, although some relationships are made more linear and therefore their correlation increases (e.g. PROF-UNSK; PROF-HDENS; EMPL-UNSK).

The other main effect of transformation is the reduction of importance of outliers. Evans et. al. (1975 p.2) state that outliers can results in a spuriously high correlation, "since great weight is given to a few outlying points". Scattergrams of the transformed variables give a "more realistic portrayal of the relationship", and the correlation co-efficient drops. This problem of outliers is obviously manifest in many of the relationships here (e.g. PRIV-IRISH, where the correlation co-efficient drops from 0.61 to 0.48; and RM3-NEWCOM, which drops from 0.72 to 0.63). There are cases where a previously significant relationship becomes non-significant after transformation - e.g. PROF-RM3 (0.12 dropping to 0.0 [not significant]). In this case and others like it, the initial weak but significant correlations are almost totally due to outliers : when the influence of these are removed, there is no correlation at all.

However, the effect of transformation on outliers can also act the other way. If the outliers are away from the main trend of the data, they can "pull" the regression line away and thus reduce the correlation considerably, as these will be far more large residuals. For instance, there is a marked contrast between the behaviour of the correlations of RM3 with INMIG and WITHMIG before and after transformation.

The correlations of RM3 with the two variables are 0.39 and 0.18 respectively. Inspection of the scattergrams (see appendix 5) shows that, in the case of RM3-INMIG, there are a large number of outliers on the main trend of the regression line. With transformation, the effect of these outliers is lessened and, as expected, the correlation drops, to 0.30. In the case of the RM3-WITHMIG relationship the initial correlation before transformation is low (only 0.18), and inspection of the scattergram shows a number of positive residual outliers which are pulling the regression line towards them. Again, after transformation the importance of these outliers decreases but this time this results in an increase in the correlation (see scattergram) to 0.30. Thus two widely differing correlations (0.39, $r^2 = 0.15$; and 0.18, $r^2 = 0.03$)- one of which may have been treated as important - have been, after transformation, determined to be of equal strength. Many other relationships show the presence of this effect, e.g. DIST-GOVT (0.27 to 0.45), GOVT-RM3 (0.33 to 0.44), EMPL-NOCAR (0.68 to 0.73) etc.

There are other cases where no relationship was detected by consideration of the untransformed data, and where a low but significant one exists afterwards. One example of this is IRISH-UNEMMA (0.03 to 0.15), where the distribution of residuals has masked the weak relationship existing in the rest of the points. Inspection of bivariate scattergrams shows that this is the reason in most cases for the development of significant relationships after transformation where none existed before.

A special variant of this type of relationship occurs when the majority of values for both variables are low (close to zero) with only a few high values of both variables. This is true of many pairings - for example AGRI-IRISH and especially MIN-NEWCOM, as can be seen from the relevant scattergrams. In some ways, these points can be regarded as outliers, but these are often quite a large number of them (c.100 out of 511). There is a measure of negative correlation built into these points (lying so close to the axes), and this may not reflect the relationship in the vast majority of constituencies which have very low percentages of both variables. Even if no relationship or even a weak positive relationship exiled between these variables it would be overwhelmed by the 100 or so points lying outside the main cluster. However, to counterbalance this there is obviously an element of real negative correlation between the variables MIN and NEWCOM as they "avoid" each other to a great extent (there are no constituencies with a nigh number of New Commonwealth immigrants and a large number of miners). So transformation will have two effects : it will make the relationships more linear (removing the constraint of having zero as an underbound) - this should result in an increase in correlation; in addition, it will reduce the effects of outliers, which will have a similar effect. Examination of the correlation matrix and scattergrams show that the relationship becomes far more linear and that correlation markedly improves (in this case from -0.22 to -0.56). This shows the true value of transformation : the relationships are much clearer and the statistics much more meaningful (as they are based on data which do not violate the assumptions of the model to as great a degree) after transformation has been carried out.

A major problem suggested by the above discussion involves the assumption of consistency of relationships over the whole country. Cox (1969a, p.99) pointed to the differing relationships that existed between voting variables and social variables in differing parts of the country (see Chapter 5), and the interrelationships between the census variables are likewise susceptible to the same problem. If differences are slight (either a difference in intercept or a difference in regression co-efficient), these differences will be masked by the overall trend of the data and the scatter about those lines obtained from the regression (although these putative differences could be tested by means of analysis of variance). However if these are marked differences these might be visible on a scattergram. In the data here, this effect could be present in a number of relationships (e.g. PRIV-OLD, SEMI-UNSK). Obviously transformation will have the same effect as it will on other residuals : it will decrease their importance (as, by definition, most of the points in question must fall at some distance from the general trend) and hence improve the correlation. There is an important point here - do we want to improve the correlations in this manner? Such an "improvement" may be regarded as a distortion of the data set. The data may fit the assumptions of the model more closely after a transformation, if this alters the nature of the relationships present then the value of such a transform is questionable. Inspection of scattergrams shows that there are a few relationships where there may be markedly different relationships present in the data, but obviously a degree of uncertainty when deviations are fine. Ideally a check should be carried out to see if the residuals from the main regression had any obvious spatial or functional links (i.e. a group of residuals may all be concentrated in one area - London or the home counties, for example; or else they may be all in inner city areas, or all in new towns etc.). The first, if present, implies spatial autocorrelation (almost certainly present in the data set) : the second that the absence of another important aspatial controlling variable is manifesting itself. Analysis of residuals was carried out by Crewe and Payne (1976) (see chapter 5).

Such, then, are the effects of the transformations on the individual correlations or on groups of correlations in general. The impact of such changes on any analysis must now be considered. First and foremost is the change in the correlation matrix itself : often this is a starting-point for many analyses, as the data are explored and basic relationships sought. Therefore anything which distorts the correlation matrix can have a fundamental and far-reaching far-reading effect on subsequent analyses : apart from the statistical problems that follow, it can result in the formulation of misconceptions in the analyst's mind which may affect his interpretation of later results. Any study that is itself based to a large extent on the analysis of correlations should take into account the effect that transformations can have.

Whilst transformation has no really dramatic results (no significant correlations change sign, for instance), some of the changes are important : e.g. the change of EMPL-HDENS (0.58 - 0.69) involves a change of 14% ($\Delta r^2 = 0.14$) in the amount of variance explained in the relationship just by considering those two variables before and after transformation. Such a change cannot be ignored, while the general structure of the matrix remains unchanged, the size of the correlation matrix tends to make such changes as do occur seem less important than they would do if the matrix contained fewer variables.

As with correlation, so with many other related techniques, the correlation merely shows the efficiency of the calculated regression line(s) in describing a relationship. Thus, whilst a change in correlation co-efficient does not ipso facto mean that there is a concomitant change in the regression parameters, there are two reasons (apart from the obvious change in measurement scale, which in itself is not important) to show that it is indicative of a change in them. The first case involves the methods used; that of the effect of modifying the two frequency distributions on the regression line. If outliers are reduced in importance, their effect on regression lines as well as on the correlation co-efficient will be less (they will exert less "pull") - therefore both the intercept and regression co-efficient may change. Thus the implied form and nature of a relationship described by such a regression will change also. Where curvilinearity is the main problem such a change is less likely (it will probably chiefly be the distribution and size of the residuals that is most affected), but there is a chance that values will change in this case also.

The second case is concerned with multiple regression. As has been noted, Poole and O'Farrell (1971) state that multicollinearity is held to exist when the "independent" variables are not linearly independent of each other. It is obvious there that there is considerable multicollinearity in the data set. All the variables in the 27 x 27 correlation matrix are controlling or predictor variables, and hence any change in the correlation (measuring the amount of co-variation) of any variables will alter the amount of collinearity. Thus transformation, although helping to meet the normality assumption, has a deleterious effect on the multicollinearity one, as, in most cases, correlations are improved rather than reduced by transformation.

Other workers on constituencies (e.g. Webber, 1978; Taylor and Johnston, 1979) include factor analytical and principal component methods to analyse the data. Consideration of the unrotated factor matrix shows considerable changes after transformation. Loadings on the first (most important) factor show many differences of more than ± 0.1 (see Appendix 4). On the whole, most high loadings (e.g. PROF, NONM, SEMI, DIST) are made higher, whilst most low loadings (e.g. TRANS, IRISH, NEWCOM) drop after transformation. There are exceptions to this general rule : OWNOCC, for instance, has a high loading which decreases after transformation - this may be due to the closure effect, to be discussed in the following chapter. Similar changes occur with the loadings on the other factors, and such polarisation of loadings is beneficial as it aids interpretation of factors. There are problems, however :

consider factor six. This factor may be considered in the analysis if the criterion for inclusion for factors is their having an eigenvalue of greater than one (this point is considered in greater detail in Chapter 6), and the untransformed variables' factor loadings would suggest an interpretation of the factor as a "mining" factor - mining has a loading of 0.62 on this factor, the next highest variable loading being 0.35. After transformation this clear interpretation is not possible, as the loading of mining on the factor drops and becomes similar to the loadings of a number of other variables.

Comparing factor loading matrices after rotation the changes were less obvious but there were still changes in structure. Again the point must be made here that in this 27 x 27 variable factor analysis, as with the correlation matrix, the overall structure is fairly stable owing to the large number of variables present. However, if the total number of variables is decreased changes may be far more important.

Thus the problem of transformation, little discussed in the relevant literature, can be seen to be a problem of some important in such analyses. Consideration should be taken of the frequency distributions and linearity of relationships before analysis takes place, and suitable transforms applied.

3.4 Weighting

In grid square studies of census date (e.g. Coulter, 1977; Evans, 1979) weighting of variable values in each grid square by total population has been carried out. This is necessary because of the huge variation in population values between grid squares, and an absence of weighting would give disproportionate influence to the less populous grid-squares at the expense of the higher density ones. In the case, the variation in population between constituencies is much smaller (the largest constituency is about $3\frac{1}{2}$ times the size of the smallest : most are roughly equivalent) and the problem does not exist to the same extent. It was therefore decided not to introduce complications by weighting for size of electorate or total population. A further defence of this is that constituencies, unlike grid squares, are "real", not arbitrary units and may have an effect on voting patterns (etc.) as such. However, it should be noted that the absence of such weighting can bias a study in favour of the smaller constituencies.

CHAPTER FOUR

PERCENTAGES AND RATIOS IN ELECTORAL STUDIES

4.1 Introduction

The use of percentages (or proportions) and ratios in electoral studies is so widespread to be almost universal: this is in common with many other branches of human geography. There are two main reasons for this (Evans and Jones, in press) :

(i) as a control for a major, dominant variable which would otherwise mask the relationships between other variables. In most branches of geography this is population, size etc. or some other related variable. This is exemplified in electoral geographical studies by the use of the percentage number of people voting for a particular party (or parties) rather than the total number of people voting for that party, thus controlling for variation in the size of electorate: this is particularly important when other data are provided in or only available in percentage or proportion form as in the census (although in many cases it would be possible to transform back to the original counts).

(ii) to create a new variable from the existing variables (e.g. persons/room) which is considered more interesting than the originals. With the variables used in this study none of the variables are found to meet this criterion.

As in the rest of human geography, use is often made of proportions and percentages with very little thought given to the problems arising from their use. Johnston (1978, p.368) lists two such problems :

(i) First, in regression, the range of numbers over which the regression can operate is restricted, which leads to difficulties in interpretation. For instance Miller et. al. (1974, p.398) give a regression equation predicting the percentage Conservative vote (expressed as a proportion of the combined Labour/Conservative vote) with the percentage of middle class workers in the constituency as the controlling variable. The equation they give is

$$2\text{-CON} = 15.9 + 0.931 \text{ MID},$$

where 2-CON is the percentage Conservative vote; MID is the percentage middle class. For a constituency with 91% of the electorate in the middle class, this equation would predict a Conservative vote of over 100%. This is obviously ludicrous (there were, in fact, six constituencies in England in 1966 which had over 85% middle class). Johnston (1978) suggests 3 methods to circumvent the problem.

a) the first is to state clearly the range of values over which the equation can be applied: this is known as the <u>domain</u> of the independent variable. In this case, the equation should be stated thus:

$$2\text{-CON} = 15.9 + 0.93 \text{ MID} \quad 0 \leq \text{MID} \leq 90.33,$$

which would limit interpretation of the equation to values of MID lying between the two logical extremes

The other two methods involve transformation of the data.

b) transforming the percentage values to an infinite ratio scale - e.g. the logit, the transform used here. The logit transform has been discussed at greater length in Chapter Three. The important factor to take into consideration here is that values of the dependent variables can only attain values of between 0 and 100, irrespective of the value of the independent variable :

c) an alternative transform can be sought which forces the regression line through X=Y=100 (where X and Y are the independent and dependent variables) and through the intersection of the mean values of X and Y. This regression is a type of logistic curve, the equation for calculating which is

$$Y = \frac{U}{1 + ae^{-bx}}$$

a = intercept, b = slope.

U = upper limit of values for Y.

(ii) The second major problem of proportions to which Johnston refers is that which he terms the independence problem. It is dealt with and amplified below under the section "ratio correlation".

(iii) Wrigley (1973) notes another problem arising from the use of percentages. Wrigley takes, as example, a study by O'Sullivan (1968) who uses percentage immigrant (of total county population) as the dependent variable. Wrigley points out that, whereas the percentage number of immigrants in a country is theoretically free to assume any value between 0 and 100, at the individual level being an immigrant or not is a once-and-for-all matter: i.e. people can either score 1 (i.e. an immigrant) or 0 (not an immigrant). The variable is therefore binomially distributed; in linear regression one of the assumptions of the model is that of constant variance of the error term (Poole and O'Farrell, 1971, p.148) and, as Wrigley points out (p.184) the error variance when using a binomially distributed case gives rise to constant variance of the error term only when all the proportions are the same. Thus the assumption is almost certain to be broken. In electoral studies the problem is similar. If one is interested in the differences between labour and conservative voting patterns, there is again a simple dichotomy - people either vote Labour or Conservative. This problem can be alleviated by transformation, and the appropriate transform is again the logit. The situation can and does get more complex when the dependent variable can assume more than two values (i.e. is polychotomous : as in this case Labour, Liberal, and Conservative) (Wrigley, 1973, p.186).

4.2 Ratio Correlation

The problem of ratio correlation has received scant attention in the electoral studies literature. The main problem stems from treating correlations obtained from ratios as if they were correlations obtained from the corresponding whole numbers. When correlating whole numbers or counts, the null expectation (i.e. the correlation that one would expect to exist between two variables that are uncorrelated) is zero: this is not always the case when correlations are calculated between variables measured in ratio or proportion form. This problem of ratio correlation is not a new one - it was first noted by Sir Karl Pearson in 1897 in the course of his contributions to the mathematical theory of evolution (Pearson, 1897), but subsequently the concept of "supurious" correlation has been largely ignored by social scientists, at least until very recently. Work in petrology by Chayes (1949, 1960, etc) and further expositions by Mosimann (1962) and Darroch (1969) have clarified matters to a certain extent. There are two main types of ratio correlation which concern us here :

(i) where two or more of the variables are expressed as a proportion of a total of which they themselves are a part: this is known as the closure effect.

(ii) where the denominators of two or more variables are identical : the common denominator effect.

The effects of these two different types of ratio correlation are unlike and must therefore be distinguished.

4.3 Closure Correlation

A simple example of closure correlation is given by Davis (1973). In this example he considers a profile along a geological section comprising two different rock types. As one goes from west to east both rock beds get thicker, and in strict proportion (see Figure 4.1), if a correlation co-efficient is calculated between unit A and B based on the actual thickness in feet or metres, the correlation co-efficient will be very high and positive (say 0.99 or 1.0): i.e. as rock A gets thicker, so does rock B and vice versa. If, however, A and B are expressed as percentages (or proportions: there is no difference here) of the total thickness at that point (i.e. as a percentage of A+B) and then correlated, the correlation co-efficient will be -1.0. As a general principle, if A/A+B is correlated with B/A+B, the correlation co-efficient must equal -1.0: if the thickness of A is 30%, B must equal 70% - if A increases to 35%, B must show a corresponding decrease to 65%.

In the case of electoral variables, this is shown most clearly by correlating the percentage Labour vote of the

Fig. 4.1: Rock thicknesses illustrating ratio correlation.

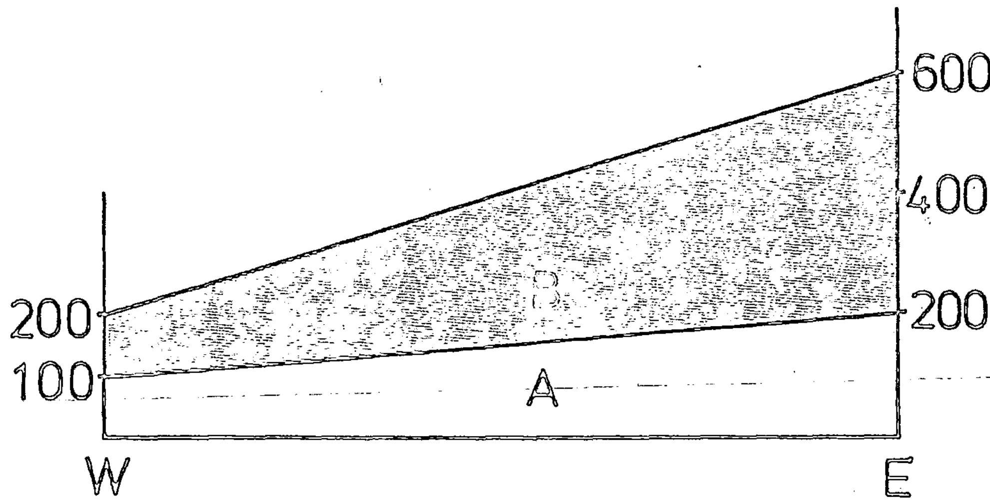
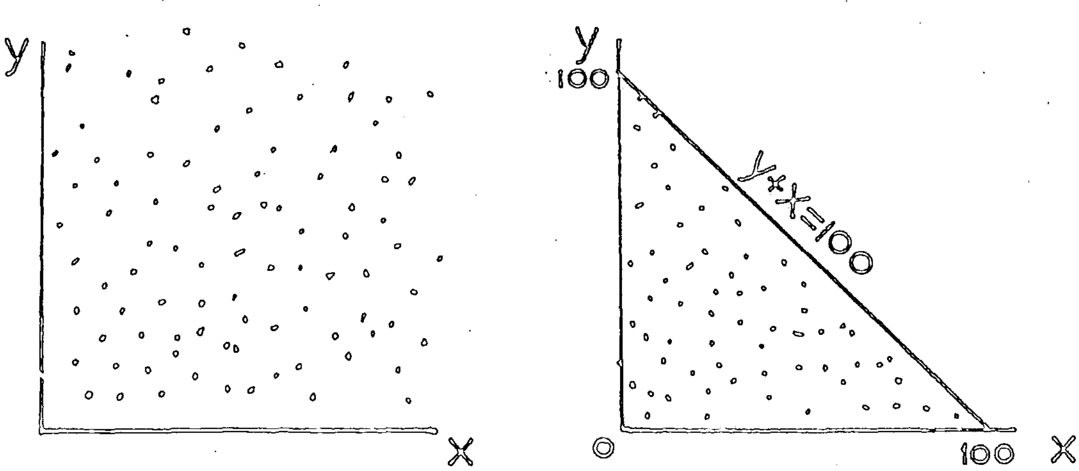


Fig. 4.2: Scattergrams illustrating ratio correlation.

A: absolute numbers B: percentages



combined Labour/Conservative vote with the percentage Conservative vote of the same: the correlation must again be -1.0.

In the two-variable case this is both obvious and simple (although it does have ramifications on subsequent analyses: see below). The problem still exists, however, with three or more variables of the same subset. Consider any two uncorrelated variables of this nature. In figure 4.2a, the scattergram is between X and Y measured in absolute units or counts, and it is apparent that no correlation exists. In figure 4.2b, the scattergram is between the same variables (X and Y) where they are measured as percentages of a fixed sum of which they themselves are a part. Thus no values can lie to the right of the line X=Y=100 (as X+Y cannot be greater than 100). Therefore, even if points are randomly distributed in the remaining space, there will still be an inbuilt negative correlation between the two previously uncorrelated variables: therefore the null (expected) correlation between the two uncorrelated variables is <u>not</u> zero.

Attempts have been made to allow for the effects of closure correlations on the null expectation: most use as their starting point Pearson's 1897 paper. The most simple case assumes that the means and variances of the variables in question are equal. If these are m variables, the expected closure correlation between each of them is given by the formula

$$r_{xy} = 1/(m-1)$$

(Evans, 1977, p.7; see also Chayes, 1960, p.4190). This would give closure correlations as in table 4.1. As can be seen, there is a rapid fall-off -

Table 4.1 : Simple closure correlations

No of vars	null r	null r^2	No. of vars	null r	null r^2
2	-1.0	-1.0	15	-0.071	-0.005
3	-0.5	-0.25	20	-0.053	-0.003
4	-0.333	-0.111	25	-0.042	-0.002
5	-0.25	-0.063	50	-0.020	0.001
10	-0.111	-0.012	100	-0.010	0.001

in the effect of closure correlation with larger numbers of variables.

This model is obviously highly unrealistic - very rarely do equal means and variances exist in real-world situations. An alternative model was derived by Mosimann (1962) which allows the means of variables to differ, but assumes that the variances are proportional to the means. Although this is not strictly true in reality, in most cases it is a reasonable approximation. Mosimann derived the following equation to determine null correlations:

$$r_{xy} = \frac{-Px\ Py}{(1-Px)}\ (1-Py) \quad \text{where } Px, Py = \text{proportions of X and Y}$$

(Mosimann, 1962 p.68 eqn. 18)

(A similar derivation was found independently by Darroch (1969, p.233 eqn. 7)). In this case, closure correlation is higher when both the proportions of the two variables are high (irrespective of the total number of variables in the group) and very low if the proportions of the two are low. The model is derived from a multinomial distribution of the <u>absolute</u> values, and a multivariate - beta distribution of the ratios (the beta distribution is a function of a variable measured on a continuous random scale taking on values of between 0 and 1 (Mood, Graybill, and Boes, 1974, p.115).

A different model has been developed by Chayes and Kruskal (1966; see also 1970). Chayes and Kruskal studied the effects not only of variation in mean proportions, but also of variance. It is a less restrictive null model than Mosimann's, but its application does give rise to problems (Evans and Jones, in press). First, an hypothetical set of uncorrelated open variables is involved: this does not always exist. Secondly, the equations used by Chayes and Kruskal (developed from Pearson) are only approximations which ignore high power deviations. An assumption is also made that the co-efficient of variation (the ratio of the standard deviation to the mean) of the numerator or denominator should not exceed 15: this assumption is broken by almost all census variables. Thirdly, the open matrix of uncorrelated variables invoked by Chayes and Kruskal is only one of many open matrices which would, on closure, yield the observed matrix (Evans and Jones, in press). These restrictions on the use of the Chayes-Kruskal model render it inappropriate for use at the moment and Mosimann's null model is adopted for use. In passing, however, one facet of the Chayes-Kruskal model should be noted: Chayes and Kruskal calculate some null expectations to be positive. This seems to conflict with the intuitive notion of variables competing for a finite space, but is possible when variances do not vary in proportion to means.

4.4 Closure correlation and the data set

In the data set used in the analysis there are seven groups of variables which could manifest the closure effect:

(1) the 'social' variables : PROF, EMPL, NONM, SKIL, SEMI, UNSK.

(2) 'occupation' variables : AGRI, MIN, MFG, TRANS, DIST, GOVT.

In each of these first two groups, the summed proportions total 100%.

(3) the housing type variables : PRIV, OWNOCC, COUNCL.

(4) the voting variables : LAB, CON, LIB.

these sum to almost 100% the remainder (apart from few exceptional cases) is 0.

(5) household size : RM3, RM7.

(6) age structure variables : OLD, YOUNG.

(7) immigrants : IRISH, NEWCOM.

In those latter three cases the proportions of each variable are fairly small, so that closure correlation is not a severe problem. Severe problems, however, exist when inspecting the correlations between members of the other groups. In view of this, null correlations for certain relationships were calculated using the Mosimann/Darroch model. These are first compared to the observed values and, later, more substantive effects noted.

(i) <u>OWNOCC, PRIV, COUNCL</u> This is a group of 3 variables in which high inbuilt closure correlation could therefore be expected. In England, the relative proportions of these three variables in 1966 were 0.455, 0.237, and 0.256 respectively. Mosimann null expectations were calculated and compared to the observed values as follows -

Table 4.2 : Mosimann null correlations for housing variables

variables	null r	observed r
OWNOCC-PRIV	-0.57	-0.62
OWNOCC-CONCL	-0.54	-0.55
COUNCL-PRIV	-0.33	-0.27

It can be seen that all three observed correlations can essentially be explained by the closure effect. Null expectations, far from being close to zero, are considerable, and therefore a rider must be put upon the interpretation of the correlation co-efficients. In this case, for instance, where there is a greater proportion of council property, there is "less room" for privately owned property etc., so there is no evidence that OWNOCC "avoids" COUNCL. Inspection of scattergrams of these three relationships show the effect of ratio correlation here quite well.

(ii) The PROF-UNSK group see correlation matrix (table 4.3).

It can be seen that the closure effect is less marked here, and that a few of the correlations do not need to be re-interpreted. However it is clear that one of the correlations (SEMI-SKIL) is almost totally due to closure, whilst some high correlations (e.g. SKIL-NONM, SKIL-UNSK) are considerably weakened by the effect.

(iii) The AGRI-GOVR group see correlation matrix (table 4.4)

Again, with more groups, some of the calculated null correlations are low. Some, however, are very important:

Table 4.3: Observed and Mosimann null correlations for the PROF - UNSK group

p = proportion of sum total of closed set.
observed correlations 0.80
expected correlations (-0.01)

p						
0.045	PROF					
0.122	EMPL	0.80 (-0.01)				
0.178	NONM	0.70 (-0.01)	0.56 (-0.17)			
0.395	SKIL	-0.76 (-0.17)	-0.78 (-0.28)	0.37) (0.37)		
0.183	SEMI	-0.64 (-0.01)	-0.53 (-0.14)	-0.78 (-0.20)	-0.37 (-0.37)	
0.088	UNSK	-0.64 (-0.00)	-0.71 (-0.01)	-0.46 (-0.14)	0.37 (-0.24)	0.32 (-0.14)
		PROF	EMPL	NONM	SKIL	SEMI

Table 4.4: Observed and Mosimann null correlations for the AGRI - GOVT group

p = proportion of sum total of closed set.
observed correlations -0.74
expected correlations (-0.69)

p						
0.03	AGRI					
0.02	MIN	0.02 (-0.00)				
0.44	MFG	-0.45 (-0.15)	-0.11 (-0.12)			
0.06	TRANS	-0.22 (-0.00)	-0.22 (-0.00)	-0.19 (-0.23)		
0.38	DIST	0.04 (-0.14)	-0.33 (-0.11)	-0.74 (-0.69)	0.21 (-0.21)	
0.06	GOVT	0.27 (-0.00)	-0.14 (-0.00)	-0.55 (-0.22)	-0.03 (-0.01)	0.27 (-0.19)
		AGRI	MIN	MFG	TRANS	DIST

the observed very high negative correlation between DIST and MFG, for instance, is almost totally explained by the closure effect (as is the low correlation between MFG and TRANS), whilst DIST-GOVT is weakened by closure.

(iv) The voting variables from the view point of the electoral geographer, perhaps the most important manifestation of ratio correlation is in the dependent (i.e. voting) variables. Table 4.5 (below) again gives the expected and observed correlations.

Table 4.5 : Mosimann null correlations for voting variables

variables	observed	expected
LABOUR - CONSERVATIVE	-0.793	-0.836
LABOUR - LIBERAL	-0.691	-0.302
CONSERVATIVE - LIBERAL	-0.127	-0.266

As has been already suggested, the correlation between the percentage Labour and percentage Conservative vote would be expected to be high and negative owing to closure: this is what we find. There would seem to be little correlation between Labour and Conservative vote once allowance is made for the closure effect - this lack of negative correlation probably reflects the effects of the variation of constituency size. The negative correlation between Labour and Liberal is enhanced by the closure effect (although even after consideration of it still remains quite strong), whilst the Liberal-Conservative correlation is weakened.

4.5 Effects of closure correlation analysis

This precis has considered the effects of closure correlation on the individual correlations of variables in these groups : their effect on the analysis as a whole needs to be considered further. This is done in two stages : first, the effects on relationships between the "independent" variables; secondly, the effect on the dependent variables and interaction between the two sets.

Effects of closure on the 'independent' variables

(i) The simple correlations are affected, as outlined above. Obviously this in itself will affect interpretation, but problems can also be carried through to associated analyses.

(ii) The correlation structure may be altered. For instance, the housing type variables OWNOCC, PRIV, and COUNCL have high correlations with each other which, we have seen, may be largely explicable by the closure effect. Therefore groups of variables (which may be used as multiple indicators of voting patterns) which may be inferred from the correlation matrix and corresponding correlation structure diagram may well be "spurious".

(iii) these "inbuilt" high correlations will have an effect on any other techniques based on the General Linear Model. As has been pointed out in a previous chapter (see also Poole and O'Farrell, 1971), one of the assumptions of the general linear model is that of low multicollinearity. Multicollinearity is held to exist when there are high correlations between the controlling (independent) variables. Multicollinearity is to be avoided whenever possible, and here it may be due, not to any substantive relationship between the controlling variables, but to the closure effect. As well as the statistical violation biasing the correlation matrix and affecting multiple regression analyses, Johnston (1978, p.262) states that the use of closed number sets can bias regression (and other techniques) by "recording the same thing more than once" : i.e. there is redundant information. This will be especially the case in factor analyses. Clearly the "spurious" high correlations in the correlation matrix will also affect factor analyses. The inclusion of highly intercorrelated variables will obviously affect the factor analyses and, if these high correlations are due in some degree due to closure, then there will be a concomitant effect on the interpretations put on the factors and their loadings. As Johnston (1978, p.278) points out, the inclusion of ratio variables in component analysis will affect not only the loadings of the other variable members of that closed set, but also the loadings of other variables loading on to the same components. In factor analysis, ratio variables will also affect the rotation carried out when the components have been erected. What Johnson omitted to point out in his review was that inclusion of ratio variables will also affect the magnitude of the factor strengths.

Enough problems exist in the independent variables therefore to make us chary of any analysis carried out using them. Further problems emerge when the dependent variables are also considered.

(ii) The dependent variables

As we have seen, the simple correlations between the voting variables are also affected by closure. This is not too important *per se* (we are not particularly interested in these correlations on their own) but the fact does have important implications for subsequent analyses. We have seen that, in a two-variable case where there are proportions of the common total, the correlation between these two must be -1.0: e.g. the correlation between Labour and Conservative (of the total Labour plus Conservative vote) is -1.0. More importantly, the correlations of these two variables with any third variable must be equal and opposite. As a general principle, if A=A/A+B and B=B/A+B, then the correlations with a third variable (c) will be as follows:

$$\text{if } r_{ac} = X, \text{ then } r_{bc} = -X$$

Thus, if the correlation of the Labour vote with variable A is 0.5, then the correlation of the Conservative vote with that same variable must be -0.5. Therefore certain observed correlations are not due solely to any process at work in reality but are, to a certain extent, modified by the methods used in recording and using the data. Although in England voting is not split up into a strict dichotomy of Labour and Conservative, the two are so important as to make this problem severe. The correlation between Labour and Conservative over the whole country is -0.79, so we must expect a certain degree of "mirror image" correlations present *no matter what* the true relationships are. There is a further complication in England : here we can effectively envisage the votes as being split three ways - apart from the odd constituency where an independent did well (none were elected), almost all the votes (over 98% in the country as a whole) will go to the Labour, Conservative, or Liberal candidates. So the country must be further divided into two groups - constituencies where the Liberals stood, and constituencies where they did not.

a) In constituencies where the Liberals stood (273 of them in England and Wales in 1966), the percentage vote for the Liberals will allow some freedom for the Labour and Conservative proportions to vary with at least some of the constraint removed, and thus allow some variation in correlation co-efficients between the two major parties votes and any social variables.

b) In constituencies where Liberals did not stand (238 in all at that time), the situation is, to all intents and purposes, the same as that outlined above : the inverse correlations will be obtained. It is true that one might expect to find different relationships to hold in constituencies where there is a wider choice for the electorate - but there must obviously be a question mark as to how much of these differences are affected by the closure constraint.

This particular problem will become less important in future in analyses of the 1974 and 1979 elections, owing to the decisions of the Liberal party to put up candidates in almost all English constituencies, but it must be borne in mind in these analyses. It should also be noted that there is a smaller but still finite effect of closure when comparing the correlations for the Liberal party with those for Labour and Conservative.

These relationships will also affect different levels of analysis based on the general linear model: for instance, Miller et. al. (1974 p.397) in discussing their multiple regression model, conclude that "the sum of Conservative plus

Liberal is more socially explicable than the Conservative vote alone" - i.e. is better explained by the variables used, to a certain extent this is probably a result of the closure constraint. In particular, factor analyses which include voting variables will be susceptible to this problem (for example Taylor and Johnston's (1979) extraction of the 'normal' vote by factor analysis). Before any remedies for dealing with closure correlation are suggested, another important type of ratio correlation will be considered.

4.6 "Common denominator" ratio correlation

Other types of ratio correlation can exist, and the one that is most important in this type of study is where the ratios have a common denominator : Chayes (1971, pp.13-14) showed that where the denominator of two ratios is the same, the null correlation should be positive. Hence it is of vital importance, when correlating ratios, to determine whether the two ratios are part of a closed set or merely have a common denominator, as the null expectations of the two are very different. It was this type of correlation which first encouraged Pearson to consider ratio correlations, and most of the subsequent work is derived from his 1897 paper. The effects of this type of ratio correlation is a contentious issue in the literature, however, and not as clear-cut as in the case of closure correlation. Pearson started off with a null model of uncorrelated absolute numbers and derived 'spurious' ratio correlations from them on closure (Pearson 1897). This was challenged by Yule (1910) who argued the reverse case: if one invoked uncorrelated ratios, then the absolute numbers derived from them would show 'spurious' relationships. Chayes (1949, p.241-2; and 1971, pp.13-14) and Atchely, Gaskins, and Anderson (1976) derived models to predict the expected null correlations in a similar way to the calculation of null closure correlations. However the calculations are all measures based only on approximations of Pearson's null estimate, and also suffer from the same problem as Chayes and Kruskal's (1966) model for closure correlation: "they break down with absolute numbers which have co-efficients of variation of about 0.15; these include almost all of the skewed distributions of human geography" (Evans and Jones, in press) Evans and Jones continue in a pessimistic manner :

> "The present situation is that we cannot estimate null values for correlations between ratios based on these variables. We cannot demonstrate that the inbuilt effect is important, or that it is unimportant. We suspect that it is quite important in many cases studied by geographers, even where numerators and denominators are highly correlated."

Many of the variables in this (and in other electoral studies) are of this kind: e.g. NOCAR. Other variables may have denominators which, if not identical, are very highly correlated (e.g. population, electorate, total economically active): indeed the denominators of all the variables in this study are very highly correlated. Thus a problem of ratio correlation of this form exists, but the exact extent of the influence is not known. This should again lead to circumspection when considering the results of such analyses.

4.7 Remedies

Evans (1977) suggests eight methods for coping with ratio correlation. Some of these methods are inappropriate for this type of analysis. Evans and Jones (in press) recommend three of these procedures.

(i) reframe the hypothesis in terms of absolute numbers and counts and correlate (etc) them. In the case of this study, this proved impossible (owing to the nature of the available data as supplied by the S.S.R.C. Survey Archive and the time constraint), but this approach could be used in other studies. As has been noted, the denominators in all the variables are highly correlated, and there are also very few variables that need to be expressed in ratio form (although if a variable such as persons/room, for example, was being used, this would not be the case). A far better reflection of the correlations between the independent variables would then be obtained, which would aid in the identification of multicollinearity, bad variable selection, factor interpretation etc. by removing the ratio correlation effect. Problems would, of course, occur when any attempt was made to include the dependent variables in a multiple regression equation: there is little point in predicting a Labour vote of 10,000 if the size of constituency is unknown. However the demand for this additional knowledge may be a small price to pay for the removal of uncertainties about the problems introduced by ratio correlation.

(ii) Partial correlations between absolute numbers controlling for the effect of population, electorate size etc. can be calculated for the ratios (Brown, Greenwood & Wood, 1914). This proved impossible to do in the time available, owing to the nature of the data used.

(iii) Data transformation can be carried out to minimise the effects of ratio correlation. Evans (1979, p.156) states that, for analyses using closed proportions, "only transformations such as angular and logit, symmetrical around 50% should be used". This redefines (and minimises) the space of "impossible" combinations, but does not remove it altogether. In regression, not only does the logit model help (as detailed early in the chapter) by ensuring that no "impossible" combinations can occur, but also helps in the solving of the binomial problem. Transformation is, of course, also needed in these data to avoid violating other assumptions favoured by Evans and Jones (in press). There are a few problems with the model (it is non-linear in form), but its advantages outweigh the disadvantages.

(iv) A fourth, more subjective method can be proposed which at least alleviates some problems of ratio correlation, especially the "double counting" that Johnston (1978) drew attention to. This comes at the variable selection stage for the final model. During this stage, variables are eliminated for a variety of reasons: multicollinearity, lack of explanation of the dependent variable etc. It is also possible to take into account possible ratio and especially closure correlations at this point. No two variables with high Mosimann null correlations (<u>irrespective</u> of the observed correlations) would be included in the final model. If more than one variable in any closed set needed to be included, it might be sensible to include variables with a smaller proportion of the total of that set, even though their correlations with the dependent variables may be slightly lower than others with correspondingly higher proportions.

Whenever possible, one of these approaches should be adopted (this list is by no means exhaustive - see for instance, Evans (1977)). The two approaches adopted in this dissertation are those of logit transformation and subsequent variable selection to rid the analysis of the worst effects of ratio correlation present in the data. In electoral geography and electoral studies as a whole, more thought needs to be given to the problems outlined above deriving from the use of ratios, instead of the 'blind' assumption that is usually made that ratios do have similarity to absolute numbers.

C H A P T E R   F I V E

ECOLOGICAL ANALYSIS AND ECOLOGICAL ANALYSES

5.1 Problems of Ecological Analysis

The problem of ecological correlation and its associate analyses is an old one and, although not as old as that of ratio correlation, has been much more widely discussed in the literature. The problem was first stated explicitly by Robinson in a seminal paper in 1950. Robinson pointed out that many previous workers had inferred from the 'ecological' correlation to the corresponding 'individual' correlation as if the two were identical. The individual correlation is the correlation based on statistical individuals : e.g. individual people. The ecological correlation is based on the attributes of a group of individuals : often geographic area.

Robinson took as his examples the correlations between colour and illiteracy. The correlation co-efficient between percent and negro and percent illiterate for the nine census Bureau divisions of the United States in 1930 was very high (0.946). This was calculated using the nine census divisions statistical individuals, plotting the percent illiterate in each region on the Y axis and percent negro on the X axis. However, when survey data relating colour to illiteracy in individuals was substituted for the regional data, the corresponding correlation coefficient was much lower, only 0.203. Thus conclusions drawn about individual behaviour from aggregate correlations would be misleading. Although the individual correlation confirms that there was a correlation between colour and illiteracy in the US in 1930, (i.e. more negroes tended to be illiterate than whites), this relationship was nowhere near as clear as might be inferred from simple inspection of the ecological correlation. Robinson also pointed out that ecological correlations differed with different groupings of areas: for instance when the colour-illiteracy correlation was calculated using states as the basis the correlation was 0.773. Hence the level of aggregation was also important in determining the strength of the correlation: this is known as the cross-level problem (Alker, 1969). Robinson demonstrated mathematically why ecological correlations should differ from the individual, and showed that two effects were important in the determination of the size of the correlation co-efficient at different levels (1950, p.356) as aggregation takes place.

(i) The average within area individual correlation increases in size because of increasing heterogeneity of the sub-areas : this tends to <u>decrease</u> the correlation co-efficient.

(ii) The correlation ratios decrease because of decrease in homogeneity in values (of the variables) within the sub-areas. This tends to <u>increase</u> the ecological correlation.

Robinson found (ii) far more important than (i), and therefore suggested that the value of the ecological correlation would tend to increase as "consolidation", i.e. aggregation, takes place (p.357).

Further work has supported and extended Robinson's original exposition of the problem Alker (1969) has summarised a wide range of these "ecological fallacies". Yule and Kendall

(1950) demonstrated (independently) Robinson's assertion about the increase in correlated yields of wheat and potato per unit area in England and Wales in 1936 at different levels, and the results are given in table 5.1.

Table 5.1 : Correlation between crop yields in England and Wales in 1936 at different levels of aggregation

Group	correlation co-efficient
48 counties	0.22
24 groups	0.30
12 "	0.58
6 "	0.76
3 "	0.99

Openshaw (1977) also considered this problem. He re-analysed data given by Cliff and Ord (1969), who correlated the number of milch cows per county with rainfall in Ireland and calculated (for 25 counties, excluding Dublin) a correlation co-efficient of 0.4051. Openshaw (p.466) calculated a series of correlation co-efficients using various 10-zone groupings, and obtained correlation co-efficient ranging between 0.0 and 0.996. Hence Openshaw showed that aggregation does not always increase the correlation co-efficients, but more importantly that the way in which we choose to group our data (or the way our data are grouped by enforced boundaries)can have a very important effect on the correlation co-efficients. He concluded that results obtained were "not independent of scale and aggregation effects implied in the choice of total boundaries" (1977, p.460). He also tested the correlation between different house types in Northumberland (based on a 100 x 100 grid mesh) at different scales of spatial resolution and concluded that the larger the grid size the larger the correlation, thus supporting Robinson. All this adds weight to Sawaki's (1973) warning against "searching for truth" at one level of analysis from statistics generated at another.

Hammond (1973) points to two sources of error in the interpretation of ecological data :

(i) when individuals are grouped into neighbourhoods on the basis of homogeneity on an independent variable, the ecological correlation will be higher, but the regression co-efficients will be unbiased estimates of the individual parameters.

(ii) when the independent variable has a contextual effect, or when individuals are grouped on the basis of their similarity on dependent variables, aggregation bias is present in the regression co-efficients and no inference can be drawn from the ecological to the individual.

So it can be seen that for a wide range of reasons electoral geographers should be circumspect of models derived from the aggregate data available to them. The effect of the "ecological problem" has been studied with particular reference to voting patterns (e.g. Shiveley (1967); Jones, (1972); Kousser (1975; Miller et. al. (1974); Crew and Payne (1976); etc.). Taylor and Johnston (1979) show an example of the problem. They quote an ecological correlation and regression between the percentage of rural form residents for the 10 counties of New Hampshire (1960 census), and the vote for Richard Nixon in the 1968 presidential election. The regression equation they obtained was -

$$Y = 43.4 + 3.53X \quad r_{xy} = 0.55$$ (Taylor and Johnston, 1979 p.83)

where Y = percentage vote for Nixon, x = percentage rural farm residents

This equation suggests that Nixon performed considerably better in areas with a larger proportion of rural farm residents. However, when the results were reanalysed on the nine census divisions of New Hampshire the regression became

$$Y = 45.63 + 0.24X \quad r_{xy} = 0.44$$

with a change in the b parameter (and therefore necessarily a change in the intercept, as the total percent of votes cast for Nixon in New Hampshire was, of course, the same in both analyses) and the correlation co-efficient was considerably smaller. When the results are calculated on the basis of the four census regions of New Hampshire, the equation becomes even more interesting :

$$Y = 51.79 - 0.10X \quad r_{xy} = -0.22$$

The b parameter (and therefore the correlation co-efficient) becomes negative. Thus the inference drawn from the county data would be that Nixon fared better in areas with more rural-farm residents, whereas the inference drawn from the regional data would be that he fared worse in the rural areas. As Taylor and Johnston point out, there is no question that one or other of these results is wrong: they are both right. They differ because they refer to different ecological aggregates. This is the "cross-level" fallacy of Alker (1969). The ecological fallacy drawn from these sets of data would be that, at the county level, that rural farm residents were more likely to vote for Nixon, whilst at the regional level, the inference drawn would be that they would be less likely to vote for him.

In addition to these two fallacies, Taylor and Johnston also point to the "individualistic fallacy", which involves making inferences from individuals to aggregates : the exact opposite of the ecological fallacy. This would not seem to be a problem given the nature of the electoral and census data used in these analyses, but if one is building models that require any cross level inference then it should be borne in mind.

5.2 The Need for Ecological Models

Shively (1967, p. 184) states that "the electorate is not a meaningful entity whose characteristics we need to study". If we accept this statement as it stands, the _raison d'etre_ for studies of this type is called into question. If we are interested only in developing unambiguous models of individual voting behaviour, shorn of the problems of the ecological fallacy, then we can accept Shively's criticism and we must rely on survey methods where the statistical individual is also a human individual. However, as Taylor (1977) has pointed out, as geographers we are interested in spatial pattern, and are thus interested in the behaviour of constituencies _per se_. Also, we have much data available of this aggregate, ecological type, and very little comparable survey data. The researcher then seems to be between the devil and the deep blue sea : to have data, and use it, though with serious reservations about interpretation, or not to have any suitable data at all. It is hardly surprising that many workers have opted for the first approach. Moreover it should be pointed out that, unlike most studies in human geography, the units that the data are collected upon are of substantive interest in themselves. In most cases, the data-collection units used in analyses are of no importance in themselves: they are merely an administrative and/or logistical convenience (e.g. Enumeration Districts, Grid Squares, etc.). This is not the case with constituencies. There are two important factors to bear in mind here:

(i) The definition of constituency boundaries in the U.K. Constituency boundaries are drawn up by the Boundary Commission. Between elections minor adjustments sometimes take place, with the occasional complete re-consideration of constituency boundaries (e.g. in 1947, 1954, and 1969 - although the effects of the 1969 changes were not felt until the 1974 elections). Although the main criterion for the definition of constituences is one of approximate population equality, there are other criteria which can (and frequently do) overrule this :

(a) The necessity of constituency boundaries to conform to existing local government boundaries, which are not fixed by the Boundary Commissioners (Taylor & Gugin, 1976).

(b) The 'accessibility' rule - this concerns constituencies in remote areas: this has led to the larger in size, predominantly rural constituencies being generally smaller in electorate than the urban ones (Busteed, 1975 p.6).

(c) Most importantly, Boundary Commissioners are careful to maintain the "community of interests", giving some unity to a constituency and ensuring that recognisable communities are not split between constituencies (Jennings, 1960 p.38). This tends to make the constituencies internally homogeneous, but, more importantly from the view point of the electoral geographer, this will tend to polarise the spatial groupings of characteristics which are hypothesised to affect voting patterns: and the effect may also be to a certain extent a sorting into constituencies on the independent variable. This if true, has several implications for ecological analyses as it violates the assumptions made in many ecological regression models (e.g. Goodman (1959), Blalock (1964) : this is pointed out by Shively (1967), and also by Hammond (1973)).

(ii) Cox (1969), amongst others, has noted the presence of a "neighbourhood effect" in voting patterns. The size of neighbourhood that produces this effect is unknown, and the data available on constituency level precludes any identification of a size smaller than this (although work has been done on local government elections e.g. Cox (1969), Rowley (1965 1971) where the basic unit is the ward. Crewe and Payne (1976) implicitly accept that the constituency as the unit most likely to show this effect (being unable to check smaller units), and their choise is intuitively reasonable: apart from other considerations people are likely to be aware of belonging to a particular constituency and conscious of the party which is represented by the sitting M.P. in their constituency, the candidates standing, and the party those candidates represent. This means that one has the apparently absurd position of having the variable "percentage voting Labour" as a productor variable in the regression to predict "percentage voting Labour" Crewe and Payne (1976, p.67) represent this by classifying a constituency according to its political complexion in 1966 (the election previous to the one that they studied), and is held to be substantively non-trivial for the above reason.

Finally, we do have the hope that predictive models built will illustrate some behavioural explanations of the pattern. While one should be wary of Grunfeld and Griliches' (1960) assertion that aggregation is a positive good, the above considerations justify our putting some faith in the results.

5.3 <u>Ecological Analyses in Electoral Geography</u>

Previous work on electoral geography has sought to overcome the problems. Norris, Hudson, and Rhind (1980, p.15) point out that many data sets are not amenable to Openshaw's (1977) method of re-zoning through aggregation of smaller units. This is true in the case of electoral data, and although groupings into larger units can (and does) take place, these tend to have some administrative function, or else grouping has been for substantive reasons (e.g. grouping into regions, urban areas), and Openshaw's method has not been used. However, other methods have been developed to deal with the problem.

One method, used by Crewe and Payne (1976), was developed by Goodman (1959) and applied to politics by Jones (1972), which involves the use of the proportion of each variable of the sum total or the denominator used. Jones (1952, p.250) starts with a simple case with two dichotomous variables; we have data for the proportion of white voters (P_w) in a series of electoral units, and also for the proportion nonwhite (P_{nw}), the proportion voting Democratic (P_d) and the proportion not voting Democratic (P_{nd}). We know that P_n+P_{nw} = 1.0; also that P_d+P_{nd} = 1.0. We wish to estimate the proportions of white voters voting Democratic (P_{wd}) and the proportion of non-white voters voting Democratic (P_{nwd}). It follows that $P_d=P_{wd}P_w+P_{nwd}p_{nw}$. Since $P_{nw}=1-P_w$, we can substitute and get $P_d=P_{wd}P_w+P_{nwd}(1-P_w)$. On expansion this becomes $P_d=P_{nwd}+(P_{wd}-P_{nwd})P_w$ (Jones, 1972 p.250). This is identical in form to the normal linear regression equation Y=a=bx, where $a=P_{nwd}$ $b=P_{wd}-P_{nwd}$. The proportion of non-whites voting democratic can be seen to be a+b. However this method makes the important assumption (noted above : see also Goodman (1959, pp.612-3) that there is no aggregation on the dependent variable, except indirectly through the independent variables. Goodman (1959) and (Jones 1973) offer methods of controlling for this, but Jones considers that the control devices are "not entirely satisfactory" (1972, p.255). There are other assumptions of the model (e.g. that the cell proportion will be randomly distributed about the mean value estimate for every unit (i.e. constituency) in the analysis: this is almost certainly an unrealistic assumption (Crewe and Payne, 1976 p.50). Another problem with the model is that it can only deal with simplistic dichotomies, although with sophistication these can be broken down into more categories. Crewe and Payne's model (1976, derived from Crewe and Payne, 1971) considers the prediction (and, by implication, explanation) of the Labour portion of the Labour/Conservative vote only, and they point out that the model of the Conservative of the same is, by definition, the model of (100% - the per Cent voting Labour). This leads to problems of ratio correlation outlined in the previous chapter. If the model is used simply for predictive purposes this is acceptable, but Crewe and Payne criticise other workers for portraying models as being wholly predictive and which "do not hint at intuitive, plausible explanations of the dependent variable" (1976, p.58), and claim that their model can be interpreted at the individual level. Again we encounter the problem of the straight-jacket: we have to assume that the behavioural processes which produce a Labour vote are equal and opposite to those which produce a Conservative vote. The problem is further demonstrated when extention beyond the aggregate level is made : as it only applies to the Labour and Conservative vote, it only applies to 64% of the electorate, as that was the combined total of votes for Labour and Conservative of those eligible to vote in 1970, the election analysed by Crewe and Payne. This figure dropped to 59% in the subsequent election, in February 1974.

Despite this problem, Crewe and Payne's analysis is extremely useful and of marked substantive interest. Their original (1971) paper utilized a simple bivariate model with a percentage Labour vote as the dependent variable and percentage manual workers as the predictor. They then analysed the residuals (differences between model prediction and reality) in order to identify other important variables. Five other measures were included in the analysis, all scored on a binary code after classification of the constituencies: for instance, if the constituency was as an agricultural seat (defined as having 3.5% of the male workforce employed in agriculture) it was given a code 1; if not agricultural, it was coded 0. The variables included are shown in table 5.2.

Although there are apparently 10 predictor variables in the subsequent regression equation, these in reality reduce to six - a constituency cannot be both 'fairly Labour' and 'very Labour', for instance -

Table 5.2 : The variables used in Crewe and Payne's (1976) analysis

Variable	Description
X_1	Percentage manual workers
X_2	Agricultural seat (3.5% employed in agriculture), or not
X_3	Mining sea (5% in mining), or not
	Minor party strength 1966 election
X_4	Minor party strong at expense of Conservative, or not
X_5	Minor party strong at expense of Labour, or not
	Partisanship 1966 election
X_6	Very Labour seat (Lab. 75% of 2-party vote), or not
X_7	Fairly Labour seat (Lab. 55-75%), or not
X_8	Fairly Conservative (Lab.25-45%), or not
X_9	Very Conservative (Lab. 25%), or not
	Nationalist
X_{10}	Nationlist stands, or not
a	the constant (0.307)
Y	Percentage Labour vote of combined Labour/Conservative vote.

The regression equation obtained was:

$$Y = 30.7 + 0.24x_1 - 0.045x_2 + 0.23x_6 + 0.096x_7 - 0.073x_8 - 0.162x_9 + 0.033x_{10} + 0.036x_3 + 0.023x_4 - 0.64x_5$$

The interpretation put upon the results by Crewe and Payne is illuminating : Social class is by far the most important predictor variable, and this is followed by agricultural employment. This suggests a secondary weak rural - urban cleavage.

Crewe and Payne then proceeded to analyse the subsequent terms in the regression; but these added successfully less and less to the explanation (see table 5.3), and it is doubtful that some of the later variables, although statistically significant, are adding much by way of explanation.

Table 5.3 : Change in predictive power of regression as variables are added

Variable	R^2	ΔR^2
X_1 (manual workers)	0.51	0.51
X_2 (agricultural)	0.67	0.16
X_6 (very Labour)	0.74	0.07
X_7 (fairly Labour)	0.81	0.07
X_8 (fairly Conservative)	0.84	0.03
X_9 (very Conservative)	0.87	0.03
X_{10} (Nationalist)	0.88	0.01
all others weaker than this		

The main conclusions drawn were that Labour benefitted more than the Conservatives from Nationalist intervention, and also from the possession of safe seats. The effects of the partnership variables is explained in terms of the "neighbourhood effect" discussed above. However, Crewe and Payne simply discuss this in terms of loyalty, which may not be the case. Butler and King (1966, p.284-286) show that voting tends to be higher in marginal seats than in 'safe' seats. Therefore a differing explanation of this phenomenon is not the loyalty of the Labour or Conservative voters in constituencies where they are in a majority, but of the differental turnout induced by the safety of the seat : it can be hypothesised that this is likely to be to the detriment of the party in the minority - their voters are more likely to become dissillusioned. This seems to be borne out by the figures.

Miller, Raab, and Britto (Miller et. al., 1974) adopted a different approach. They confined their analysis to using census (or census-derived) variables only as predictor variables, but instead of just analysing at constituency level they adopted a cross-level approach for the 1966 election, analysing at four levels : constituency, borough, county, and region. They found that, at different levels, the best predictor variables were not necessarily the same. They also found, surprisingly, that for the two main parties relationships tended to weaken at higher levels of grouping, which is the opposite result to that which we would usually expect from Robinson (1950) and Openshaw (1977). However, it can be seen from the regression co-efficients' behaviour at different levels that grouping has not been carried out on the dependent variable, at least not at these scales (Miller et. al., 1974, p.401). The fact that correlation decreases in size would tend to point that the first processes listed by Robinson (1950, p.306) was the more important; i.e. that as consolidation takes place, the average within-areas correlation increases in size because of heterogeneity. Thus, taken into account with the change in regression co-efficients at different levels, lead Miller et. al. to conclude that the social variables affect voting patterns at constituency level or below (1974, p.100).

Miller et. al. looked at both linear and quadratic models for explanation. For the main parties they found the expected pattern of results - social class was the main predictor (usually through the proportion of employers and managers). However, the second and additional variables in general added little to the R^2 value, and the model seems one dimensional. This is possibly due to the large amount of multicollinearity present in the data. More interesting is their extension of the analysis to include the Liberal vote. They find that, at the constituency level, the most important predictor is the percentage employed in agriculture, which suggests a predominantly urban/rural cleavage. They note that this may represent areas where the local Labour and Conservative parties are less well organised, and that in rural areas generally local government elections are less likely to be organised on party lines. At two of their levels of analysis, they find that the percentage of retired people correlates fairly highly with the Liberal vote : arguing down to the individual level, this could represent voters whose political opinions hardened before the Labour party became an important political power : on the other hand it could just represent the differing age structure in rural areas. Miller, Raab, and Britto's work is explicitly concerned with the prediction and the behaviour of constituencies : they are not concerned with explanation at the individual level. However, this is implied in much of their analysis (the 'neighbourhood' effect, for instance).

The two papers by Crewe and Payne (1976) and by Miller et.al. (1974) are the two main references in this field in the U.K. Other work has been carried out at a regional level (e.g. Cox 1969b, Rowley 1967, 1969) or of particular types of area (e.g. Roberts and Ramage 1965). Work on the national scale includes Barnett (1973) and Rasmussen (1973). Crewe and Payne (1976) criticise both these multiple regression models as being substantially uninteresting, and Taylor and Johnston (1979, p.212) state that Barnett and Rasmussen are merely producing "general structural measures of constituency type which relate to affluence and hence social class".

Abroad, as has been noted in Chapter 2, such studies have had a much longer history. Following Siegfied's (1913) work, statistical analyses developed in the USA (e.g. Gosnell, 1975, Ogden 1919, 1929 etc). On the continent voting patterns in Italy have been related to a whole series of variables (Capecchi and Galli (1969), and the religions class factor in each voting analysed by the same methods (Dogan (1969)), but such studies are beyond the scope of the work here.

In summary, then, ecological analyses can be seen to be useful, but the results must be viewed with an amount of circumspection. Langbien and Lichtman (1978) point to the importance of proper model specification in dealing with ecological correlation. They suggest that in order to obtain the correct deduction from the aggregation model, it should resemble the individual model as closely as possible. The models should be applicable to both the micro- and the macro-scale. This does however lead to problems of the individualistic fallacy noted above. This echoes the call of Crewe and Payne (1976 p.58) for aggregate models to "hint at intuitively plausible explanations of voting behaviour".

CHAPTER SIX

DISCUSSION

6.1 Relationships between the predictor variables

Correlations between the 27 census variables used in the analysis are given in appendix 3. The resulting correlation matrix shows very well the segregation of the population of the country to which Johnston (1979) refers. In considering the matrix,however, the problems outlined in the preceding chapters should be borne in mind - especially those of ecological analysis and ratio correlation.

Many of the correlations in the matrix are intuitively obvious : for instance the correlation co-efficient between PROF and NOCAR is 0.64. This simply indicates that the greater proportion of professional workers in a constituency the fewer households without a car there will tend to be : this also makes sense at the individual level. However, with other relationships this may not hold true : high correlations may be due to ratio correlation (e.g. EMPL-UNSK), aggregation effects (e.g. YOUNG-OLD), or through their relationship with a third variable (e.g. PROF-OWNOCC).

Unfortunately, the very size of the correlation matrix means that it is an unwieldy summary of the relationships, and other methods have to be sought to collapse the 27 variables into meaningful groups. Webber (1978) suggests two methods : the use of principal components analysis and the construction of a minimum spanning tree.

The minimum spanning tree for the 27 variables used in this analysis is shown in fig. 6.1. The construction of a minimum spanning tree involves two phases : first, each variable is linked to the variable with which it has the highest correlation. This results in the clustering of variables into a series of groups. The second stage involves the linking of a variable within each of the groups to the variable outside the group with which it has the highest correlation : thus the "islands" of variables, as Webber terms them, are joined up to form a single network. The strength of the correlation is depicted by a varying number of lines joining the variables : the more lines, the stronger the correlation.

The spanning tree constructed differed from that of Webber. The main difference is obviously in the variables used : Webber used 40, whereas only 27 were used in this analysis. There were other differences - Webber's analysis was for the whole of Great Britain, not just England; and was based on the 1971 census, not the 1966 sample census. Webber does not discuss transformation and, as has been demonstrated in chapter 3, transformation can have a marked effect on some of the correlations - these changes may alter the spanning tree : this point is expanded upon below.

In the spanning tree six groups can be distinguished. The core group comprises the social status variables EMPL, PROF, UNSK, together with UNEMMA (which could also be interpreted as a social status variable), RM7 and COUNCL. Three of the other five groups are linked to variables in

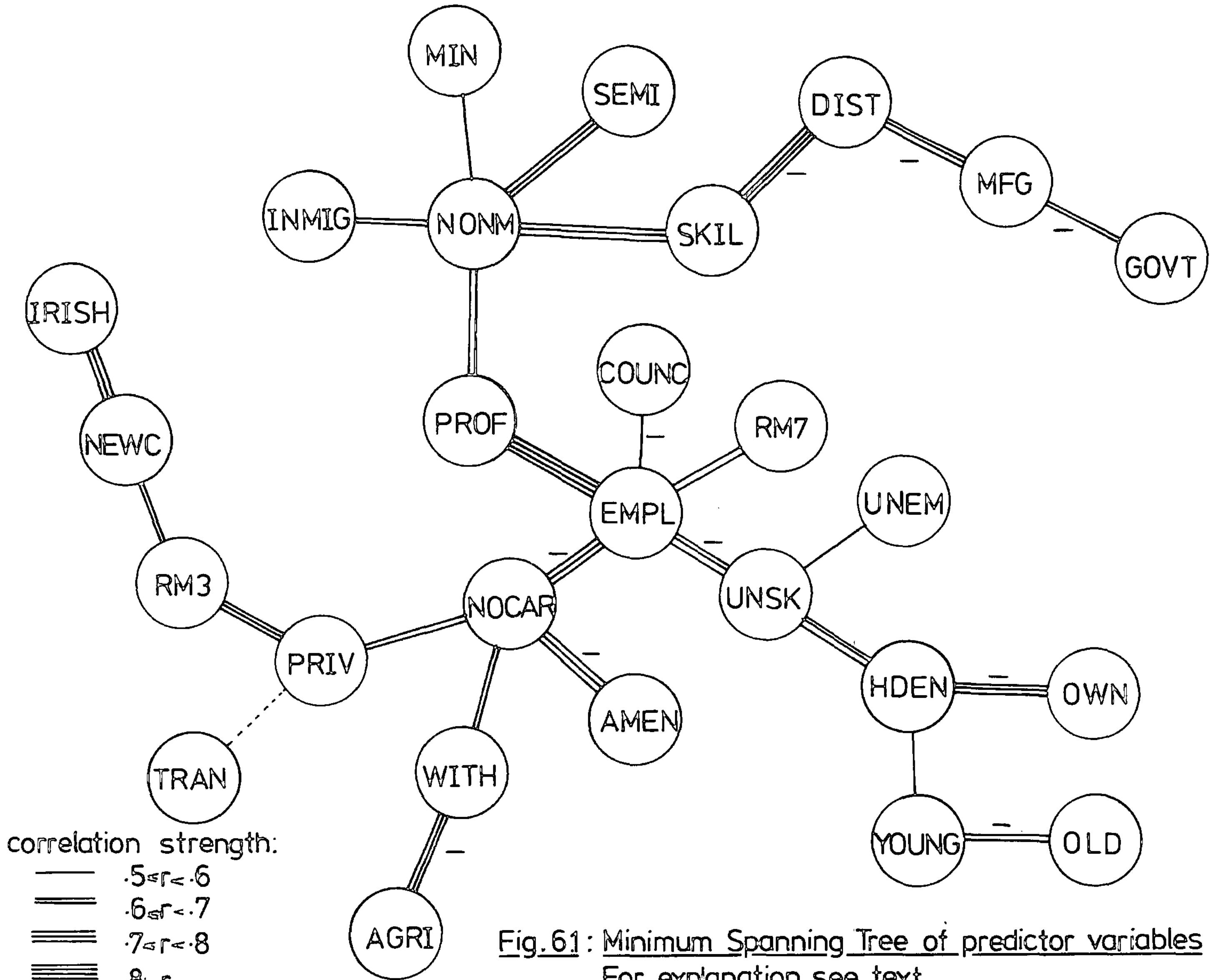


Fig. 6.1: Minimum Spanning Tree of predictor variables
For explanation see text

this central group. Closely related to this group is the group of NOCAR, AMEN, WITHMIG, and AGRI. NOCAR and AMEN can be interpreted as income surrogates, WITHMIG showing the high outmigration from such poor areas which tend to be rural. The third group of HDENS, OWNOCC, OLD and YOUNG which is connected to the main group through UNSK. This grouping is less easy to explain than the two previous ones. The negative link between HDENS and OWNOCC seems logical enough, as does the one between OLD and YOUNG. The two sub-groups are linked by YOUNG-HDENS. A fourth group to the top of the diagram is also difficult to explain : NONM, SEMI, MIN, and INMIG. Connected to this group through NONM is the fifth group of SKIL and the occupation variables DIST, MFG, and GOVT. The final group to the left of the diagram consists of IRISH, NEWCOM, RM3, and PRIV. This is again a fairly logical grouping between areas of high immigration and small households which tend to occur in areas of private rental.

So it can be seen that the minimum spanning tree is a fairly useful summary of the relationships. However it is a fairly flimsy structure, as can be demonstrated from the consideration of a few examples. Consider if the tree had been constructed using untransformed data. A few minor linkages would be changed (e.g. YOUNG would be joined to OWNOCC instead of HDENS), but there would also be a few major changes (NOCAR would be linked with UNSK, for example) which would alter the whole network structure. There are also problems with ratio correlation : the group of variables DIST-MFG-GOVT all show high negative correlations. As has been demonstrated in chapter three, an element of this high correlation may be due to closure and these variables may be linked with others if this effect could be better accounted for. Given these problems, plus the fact that the structure omits so many high correlations, one should be wary about placing too much emphasis on the technique.

Another method of simplifying the correlation matrix is by principal component and factor analytical methods. These techniques are based on the assumption that there is a high degree of multicollinearity existing in the data set and that the variables can be compressed into a smaller number of components or factors (a full discussion is given in Mather, 1976, chapters 4 and 5). The components are formed from a set of loadings for each variable which measure the correlation between that variable and the component. The first component is extracted in such a way as to maximise the amount of variance that it extracts from the original variance of the standardized variables, and subsequent components also attempt to extract the maximum amount of remaining variance, with the additional constraint that they are orthogonal to all other components (that is, they are unrelated to them). Factor analysis then rotates these axes in order to aid interpretation.

The problem of how many components to select is not an easy one and has been the subject of much debate (Davis, 1973 pp. 508-509). The conventional wisdom is to retain all factors with an eigenvalue of greater than one - i.e. which explains a greater amount of variance than any of the original standardised variables (the size of the eigenvalue gives the amount of the total variance explained by that component or factor : the sum of the eigenvalues must equal the number of variables in the analysis - in this case, 27). However, there are problems with this method : the decision to take the inclusion level as one is rather arbitrary and would lead to one interpreting a component with an eigenvalue of 1.01 but not one with 0.99. This seems not only arbitrary but also unreasonable. The further problems which ensue are those of interpretability of some of the factors : Webber notes this (1978, p.9).

There are alternative criteria for choosing the number of components to analyse, but these also involve a degree of subjective decision. Instead of searching for the eigenvalues of greater than one, the position of maximum break in the eigenvalues can be used. This means that when an eigenvalue is encountered that predicts proportionately much less than the previous one, it and subsequent components are dropped from the analysis. Another method is to look at the factor loadings : if there are no high loadings the factor is not interpretable. Again there are problems : a decision has to be made as to what constitutes a high loading. If a loading of 0.70 is adopted, this means that just under half of the variance of that variable is subsumed within the relevant component. If the loadings drop much lower than this then the proportion of the variance explained becomes fairly low.

The results of a principal components analysis carried out on the study data are given in Appendix Four. The first six components (those with eigenvalues of greater than one) accounted for 81.1% of the total variance in the data set. The eigenvalues for the first seven components are shown in table 6.1. It can be seen that the major break in the eigenvalues occurs between the second and third components. This would lead one to select a two-component solution. Consideration of the component loadings (see Appendix Four) also point towards a two-component solution. There are a considerable number of high loadings on the first two components, but none above 0.7 on components 3 to 6. Indeed component 3 has only one loading of over 0.6 (and that only just), while none of the other 3 components have loadings of such a magnitude.

Table 6.1 : Eigenvalues for first seven components

Component	Eigenvalue	Percent of variance explained
1	8.98	33.3
2	5.85	21.7
3	2.82	10.4
4	1.76	6.5
5	1.33	4.8
6	1.16	4.3
7	0.79	2.9

A two-component model explains 54.9% of the variance, so there is still 45.1% unexplained. The communalities (the amount of variance of each individual variable explained by the two components) must average 0.549, but there is a considerable variation within this. Some of the variables are very well explained e.g. EMPL (0.91), NOCAR (0.82), whilst some are not so well explained e.g. OLD (0.10), AGRI(0.43). The component loadings show high loadings on component 1 of the social class variables (PROF-UNSK); also DIST;NOCAR; the housing variables OWNOCC and COUNCL; AMEN; HDENS; RM3; and INMIG. Hence it is a component that correlates positively with measures of wealth and high social class, mobility etc. and negatively with measures of deprivation and low income etc. The second component has fewer high loadings : AGRI, PRIV, IRISH, NEWCOM, and RM3, and is also more difficult to interpret, although a broad similarity between this group and one of the groups at the minimum spanning tree must be noted.

The two principal component axes were then rotated to perform a factor analysis. The idea of the factor analysis is to redistribute the variance explained in such a way that loadings of variables on the factors are polarised so that the factors become more interpretable. The results of the 2-factor varimax solution is also given in Appendix Four. Rotation does help, especially for factor 2 - six loadings are above 0.7 compared with only three before.

Principal components analysis and factor analysis, being based on the correlation matrix, are subject to the problems of transformation and ratio correlation. Transformation does appear to have had a minor effect on the component loadings (compare the two 2-component models in

Appendix four) : the loading of PROF on component 1 changes from 0.76 to 0.85 when transformed data is used for instance. There are a few major changes : for example the loadings of INMIG and WITHMIG on component 2 are 0.54 and 0.11 respectively before transformation, and 0.38 and 0.30 after. Changes of this magnitude are important and further demonstrate the need for the consideration of transformation problems.

Ratio correlation will also be affecting the analyses, although the size of this effect is difficult to quantify. Any "spurious" correlations due to closure or the common denominator ratio correlation will distort the factor matrix : high negative correlations induced by closure will result in variables loading onto the same factor, for instance.

Such, then, is the structure of the independent variables and the problems associated with them Before any attempt is made to make a selection from these variables, the general relationships between the predictor variables and the voting variables will be considered.

6.2. Relationships between predictor and electoral variables

From the correlation matrix of census and electoral variables (see Appendix 3) few surprises are evident. The Conservative vote is positively correlated with variables such as PROF, EMPL, OWNOCC, RM7; and negatively correlated with UNSK, NOCAR, MDENS etc. These correlations are ecological ones but also make sense when reasoned down to the individual level. The correlations for Labour tended to be opposite in sign and similar in strength compared to the corresponding

Conservative ones: this is expected for reasons given in Chapter Four. However the correlations do tend to be slightly higher than those for the Conservative. One of the reasons for this is seen when the correlations with the Liberal vote are considered : these are, in general, considerably weaker tban those for the two main parties, but tend to be the same sign as the corresponding correlations with the Conservative vote. Hence Miller et. al.'s assertion that the combined Liberal-Conservative vote is more socially explicable than the Conservative vote alone. Liberal voting does not correlate very highly with any variables, but has moderate (but highly significant) correlations with AGRI, EMPL, NOCAR etc., and a surprisingly high correlation with RM7 : this does not seem to make any sense on the individual level.

6.3 Variable selection and analysis

It will be obvious from the discussions on this and preceding chapters that the number of predictor variables has to be severely pruned. In summary -

(i) No two variables with high simple correlations should be included in the analysis to avoid problems of multicollinearity. However this means that many variables of substantive interest will be omitted.

(ii) No two variables of the same closed set should be included to help avoid the problems of ratio correlation.

(iii) Following the advice of Crewe and Payne (1976) and Langbein and Lichtman (1978), only variables of "substantive interest" should be included. This means variables which can be interpreted at the individual level.

In order to choose variables the principal components and factor analyses are useful. Both of these suggest that there are two dimensions to the data set, and variables could be sought that have a high loading on one of the two factors.

The simple correlations suggest that, for both major parties that EMPL is the most important variable. It has the highest correlation and has a high loading on factor 1. Partial correlations between the Conservative vote and all the other variables controlling for EMPL were then calculated (see table 6.2). They are all, bar a few, very low indeed,

Table 6.2 : Partial correlations of variables with the Conservative vote controlling for EMPL

Variable	Corre-lation	Variable	Corre-lation	Variable	Corr-lation
PROF	0.15	DIST	0.03	RM7	0.33
NONM	0.08	GOUT	0.16	IRISH	0.15
SKIL	0.04	NOCAR	-0.10	NEWCOM	0.16
SEMI	-0.07	OWNOCC	0.18	YOUNG	0.13
UNSK	0.06	COUNCC	-0.17	OLD	0.08
AGRI	0.00	PRIV	-0.0	UNEMMA	-0.05
MIN	-0.33	AMEN	-0.02	INMIG	-0.07
MFG	0.08	HDENS	-0.09	WITHMIG	0.02
TRANS	-0.12	RM3	-0.09		

only MIN and RM7 exceeding ± 0.25. MIN indeed had a high loading on component 2, and therefore was chosen to be the next variable in a regression. The two variable regression caused the amount of variance explained to be increased from 62% to 66% - just over 4%. Any subsequent addition of variables added less than 2%.A summary of regressions can be seen in table 6.3. The results were similar for the Labour vote, with again EMPL and MIN being the first two variables input; the variance explained increasing from 76% (with one variable) to 78% (with both variables).

When the Liberal vote was considered, the variable with the highest simple correlation, RM7, was regarded as being substantively uninteresting. The two variables included were EMPL and AGRI (see table 6.3). The r^2 value for Liberal-AGRI was 0.15, increasing to 0.32 when EMPL was introduced. Subsquent variables added only a very small amount to the additional variance explained. It is interesting to note that the partial correlation between the Liberal vote and AGRI, controlling for EMPL is 0.45 : compare this with the low partial correlation between AGRI and the Conservative vote.

The results of multiple regressions and correlations are disheartening. They again emphasise the importance of social class and leave little room for any subsequent analyses. The major finding of previous workers are, to a great extent, borne out, but there are differences : this analysis did not pick up the rural/urban cleavage that Crewe and Payne's (1976) analysis did, for instance. Indeed, as has been shown, once the influence of the class variable EMPL had been accounted for, the correlation between Conservative voting and AGRI was almost nil. The implications of this, and of other problems, are discussed in the following concluding section.

Table 6.3: Selected results from regression analyses

Variables entered	Regression coefficients a	b	r^2	Δr^2
A. Conservative vote				
Step one: EMPL	1.10	0.67	0.56	-
Step two: EMPL	0.65	0.60	-	-
MIN	-	-0.06	0.61	0.05
B. Labour vote				
Step one: EMPL	-2.52	-1.15	0.70	-
Step two: EMPL	-3.13	-1.07	-	-
MIN	-	0.47	0.73	0.03
C. Liberal vote				
Step one: AGRI	-1.2	0.11	0.15	-
Step two: AGRI	0.30	0.09	-	-
EMPL	-	0.55	0.31	0.16

6.4 Conclusion

Any further detailed discussion of the results would be rather fruitless, as it would either duplicate that which has been written in previous analyses or else replicate material discussed in earlier chapters. The most worrying aspect of this and similar studies is the dimensionality of the model. The use of a census social class variable as the main predictor has meant that there is very little variation in voting patterns left over that can be described by other methods. A corollary of this is that when a slightly different indicator of class is used, slightly different variable definition is used the interpretation put upon the remaining terms may vary considerably : consider, for instance, the difference between the variable AGRI in this analysis and a similar variable in the analysis performed by Crewe and Payne (1976).

Class voting is important in British elections, but the inclusion of a census class variable appears to be an overgeneralisation of processes. As McKenzie and Silver (1967) amongst others have shown a considerable number of working class people vote Conservative, but these analyses, being based on an aggregate level, fail to pick these deviations up. At this level, class variables are so highly correlated with a whole series of other factors (as can be seen from the correlation matrix) that the chances of gaining any useful information beyond the trite generalisation that people in higher class areas tend to vote Conservative are minimal. Despite the usefulness of Crewe and Payne's analysis, it seems wrong to spend about half a page discussing the meaning of a variable that explains over half of the variance, and the rest of the article analysing the meaning of variables which add (at most) 16% to the variance explained.

In summary, then, a series of recommendations can be made :

(1) If possible, survey results relating to individuals should be used in preference to census data describing areas, although these can be used as a check that survey results are accurate.

(2) Data collected should, if at all possible, be analysed as whole numbers or counts in order to counteract the problems of ratio correlation : or if that is impossible, a more careful variable selection.

(3) In the case of data violating the assumptions of a statistical model, the use of transformations should be studied carefully, and report made of any transformation procedure used.

The author does <u>not</u> see any great future for this type of analysis in human geography. However constituency level analysis is useful in that it does provide an insight to the nature of constituencies, and provides another level of analysis for study, and as has been shown, multilevel analyses are most valuable. However any study which does not take into account the problems discussed in the above chapters runs the risk of producing results which are open to serious misinterpretation.

References and Bibliography

Alford, R.R. (1967) : Class voting in Anglo-American Political systems. In Lipset, S.M. and Rokkan, S. (eds.) q.v., 67-93.

Alker, H.R. (1969) : A typology of ecological fallacies. In Doyan, M. and Rokkan, S. (eds.) q.v., 69-86.

Atchley, W.R., Gaskins, C.T., and Anderson, D. (1976) : Statistical properties of ratios : I - empirical results. Systematic Zoology 25, 137-148.

Barnett, M.J. (1973) : Aggregate models of British voting behaviour. Political studies 21, 121-134.

Benjamin, B. (1970) : The Population Census . Social Science Research Council Review of Current Research, No. 7.

Blalock, H.M. (1964) : Causal Inferences in non-experimental research. Chapel Hill, North Carolina : University of North Carolina Press.

Blondel, J. (1974) : Voters, parties, and leaders : the social fabric of British politics. Revised edition. Harmondsworth : Penguin.

Brown, J.W., Greenwood, M., and Wood, F. (1914) : A study of index correlations". Journal of the Royal Statistical Society 77, 317-329.

Busteed, M.A. (1975) : Geography and voting behaviour. London : Oxford University Press.

Butler, D.E. and King A. (1966) : The British general election of 1966. London : Macmillan.

Butler, D.E. and Stokes, D. (1969) : Political change in Britain : forces shaping electoral choice. London : Macmillan.

Butler, J.C. (1974) : Analysis of correlations between percentages. Journal of Geological Education 74, 56-61.

Capecchi, V. and Galli, G. (1969) : Determinants of voting behaviour in Italy : a linear causal model of analysis. In Dogan, M. and Rokkan, S. (eds.), q.v. 235-283.

Chayes, F. (1949) : Ratio correlation in petrography. Journal of Geology 57, 239-254.

Chayes, F. (1960) : On correlation between variables of constant sum. Journal of Geophysical Research 65 (12), 4185-4193.

Chayes, F. (1971) : Ratio correlation : a manual for students of geochemistry and petrology. Chicago : University of Chicago Press.

Chayes, F. and Kruskal, W. (1966) : An approximate statistical test for correlations between proportions. Journal of Geology 74, 692-702

Chayes, F. and Kruskal, W. (1970): An approximate statistical test for correlations between proportions : Some corrections. Journal of Geology 78, 380.

Clark, D. (1973) : Normality, transformation and the principal components solution : an empirical note. Area 5 (2), 110-113.

~~Cliff, A.D. and Ord, K.: The problem of spatial autocorrelation.~~

Cliff, A.D. and Ord, K. (1975) : The comparison of means when samples consist of spatially autocorrelated observations. Environment and Planning A 7, 725-734.

Coulter, J. (1977) : The use of grid-square census data in a study of deprivation in the conurbations of the United Kingdom. Unpublished M.Sc. dissertation, University of Durham.

Cox, K.R. (1969a) : The voting decision in a spatial context. Progress in Geography 1, 81-117.

Cox, K.R. (1969b) : Voting in the London suburbs : a factor analysis and a causal model. In Dogan, M and Rokkan, S. (eds.) q.v. 343-369.

Crewe, I. and Payne, C. (1971) : Analysing the census data. In Butler, D.E. and Pinto-Duschinsky, M., The British general election of 1970. Macmillan : London, 416-436.

Crewe, I. and Payne, C. (1976) : Another game with nature : an ecological regression model of the British two-party vote ratio in 1970. British Journal of Political Science 6, 43-81.

Darroch, J.N. (1969) : Null correlation for proportions. Journal of the International Association for Mathematical Geology 1(2), 221-227.

Davis, J.C. (1973) : Statistics and data analysis in geology. New York : John Wiley.

Dogan, M. (1969) : A covariance analysis of French electoral data. In Dogan, M. and Rokkan, S. (eds., q.v. 285-298.

Dogan, M. and Rokkan, S. (eds.) (1969) : Quantitative ecological analysis in the social sciences. Cambridge, Massachussetts : The M.I.T. Press.

Evans, I.S. (1973) : Strategies for dealing with percentage and ratio correlation. Unpublished typescript, Durham University.

Evans, I.S. (1979) : Relationships between Great Britain census variables at the 2 km aggregate level. In Wrigley, N. (ed.), Statistical applications in the spatial sciences. London : Pion, 141-184.

Evans, I.S., Catterall, J., and Rhind, D.W. (1975) : Specific transformations are necessary. Census Research unit Working Paper 2, University of Durham.

Evans, I.S. and Jones, K.J. (in press) : Ratios. In Wrigley, N. and Bennet, R.J. (eds.), Quantitative geography in Britain : retrospect and prospect. London : Routledge and Kegan Paul, Ch. 3.3.

Goodman, L. (1959) : Some alternatives to ecological correlation. American Journal of Sociology 64, 610-625.

Grunfeld, Y. and Griliches, Z. (1960) : Is aggregation necessarily bad?. Review of Economics and statistics 42, 1-13.

Hammond, J.L., Jr. (1973) : Two sources of error in ecological correlations. American Sociological Review 38, 764-777

Harvey, D.W. (1968) : Pattern, process and the scale problem in geographical research. Transactions of the Institute of British Geographers 40, 81-95.

Harvey, D.W. (1969) : Explanation in Geography, London : Edward Arnold.

Haynes, R.M. (1978) : A note on dimensions and relationships in human geography. Geographical Analysis 10, 288-291.

Hoyle, M.H. (1973) : Transformations - an introduction and a bibliography. International Statistical Review 41 (2), 203-223.

Jennings, I. (1960) : Party Politics - Volume 1 : Appeal to the People. Cambridge : Cambridge University Press.

Johnston, R.J. (1978) : Multivariate statistical analysis in geography : a primer on the general linear model. London : Longmans.

Johnston, R.J. (1979) : Political, electoral, and spatial systems. Oxford : Oxford University Press.

Jones, E.T. (1972) : Ecological inference and electoral analysis. Journal of Interdisciplinary History 2, 249-269.

Kinnear, M. (1968) : The British voter : an atlas and survey since 1885. London : Batsford.

Kousser, J.M. (1973) : Ecological regression and the analysis of past politics. Journal of Interdisciplinary History 4, 237-262.

Krehbiel, E. (1916) : Geographical influences in British elections. Geographical Review 2, 412-432.

Kruskal, J.B. (1968) : Transformations of data. In Sills, D.L. (ed.) International Encyclopedia of the Social Sciences v. 15. New York : Macmillan, 182-193.

Langbein, L.I. and Lichtman, A.J. (1978) : Ecological inference. Sage University paper : Quantitative applications in the social sciences, 10.

Lewis, P.F. (1965) : Impact of negro migration on the electoral geography of Flint, Michigan, 1932-62 : a cartographic analysis. Annals of the Association of American Geographers 55, 1-25.

Lijphart, A. (1971) : Class voting and religious voting in European democracies. Acta Politica 6, 158-171.

Lipset, S.M. and Rokkan, S. (eds.) (1967a) : Party Systems and voter alignments : cross-national perspectives. New York : The Free Press.

Lipset, S.M. and Rokkan, S. (1967b) : Cleavage Structures, party systems, and voter alignments : an introduction. In Lipset, S.M. and Rokkan, S. (eds.) q.v., 1-64.

Lyons, F.S. (1973) : Ireland since the famine. Glasgow : Fontana.

McCallum, R.B. and Redman, A. (1951) : The British general election of 1950. London : Oxford University Press.

McKenzie, R.T. and Silver, A. (1967) : The delicate experiment : Industrialism, Conservation and working-class Tories in England. In Lipset, S.M. and Rokkan, S. (eds.), q.v. 115-125.

Mather, P.M. (1976) : Computational methods of multivariate analyses in physical geography. London : John Wiley.

Miller, W.L., Raab, G., and Britto, K. (1974) : Voting research and the population census 1918-1971 : Surrogate data for constituency analyses. Journal of the Royal Statistical Society A, 137, 384-411.

Mood, A.M., Graybill, F.A., and Boes, D.C.(1974). Introduction to the theory of statistics. 3rd edition. New York: McGraw-Hill.

Mosteller, F. and Tukey, J.W. (1977) : Data analysis and regression, Reading, Massachussetts : Addison-Wesley.

Mosimann, J.E. (1962) : On the compound multinomial distribution, the multivariate beta distribution, and correlations amongst proportions. Biometrika 49, 65-82.

Norris, P., Rhind, D.W., and Hudson, R.(1980) : Alchemy in action? regional databases, statistical analysis and cartography in a computer environment. Paper given at the Symposium for the Development of Joint Data Bases for regional Analysis of Computer Cartography, Turin, 17-22 March.

O'Sullivan, P.M. (1968) : Accessibility and the spatial structure of the Irish economy. Regional Studies 2, 195-206.

Openshaw, S. (1977) : A geographical solution to scale and aggregation models in region-building, partitioning and spatial modelling. Transactions of the Institute of British Geographers N.S. 2, 459-472.

Pearson, K. (1897) : Mathematical contributions to the theory of evolution - on a form a spurious correlations which may arise when indices are used in measurement of organs. Proceedings of the Royal Society A, 60, 489-502.

Pelliny, H. (1967) : The social geography of British elections 1885-1910. London : Macmillan.

Poole, M.A. and O'Farrell, P.N. (1971) : The assumptions of the linear regression model. Transactions of the Institute of British Geographers 52, 145-155.

Rassmussen, J. (1973) : The impact of constituency structural characteristics upon political preferences in Britain. Comparative Politics 5, 123-145.

Robinson, W.S. (1950) : Ecological correlation and the behaviour of individuals American Sociological Review 15, 351-357.

Rowley, G. (1965) : The Greater London Council elections of 1964 : some geographical considerations. Tisd. Econ. Soc. Geog. 56, 113-114.

Rowley, G. (1971) : The Greater London Council elections of 1964 and 1967 : a study in electoral geography. Transactions of the Institute of British Geographers 53, 117-131.

Sawici, D.D. (1973) : Studies of aggregated areal data : problems of statistical inference. Land Economics 49, 109-114.

Shively, W.P. (1967) : 'Ecological inference' : the use of aggregate data to study individuals. American Political Science Review, 63, 1183-1196.

Siegfried, A. (1913) : Tableau politique de la France de l'ouest. Paris : Aramand Colin.

Taylor, P.J. (1977) : Quantitative methods in geography Boston : Houghton Mifflin.

Taylor, P.J. and Gudgin, G. (1976) : The statistical basis of decision making in electoral districting. Environment and Planning A, 8, 43-58.

Taylor, P.J. and Johnston, R.J. (1979) : Geography of elections. Harmondsworth : Penguin.

**
Wrigley, N. (1973) : The use of percentages in geographical research. Area 5, 183-186.

Wrigley, N. (1976) : An introduction to the use of logit models in geography. Concepts and Techniques in Modern Geography 10.

Yule, G.U. (1910) : On the interpretation of correlations between indices or ratios. Journal of the Royal Statistical Society 13, 644-647.

Yule, G.U. and Kendall, M.G. (1950) : An introduction to the theory of Statistics, London : Charles Griffiths.

* Roberts, M.C. and Ramage, K.W. (1965) : The spatial variations in urban left-wing voting in England and Wales in 1951. Annals of the Association of American Geographers, 55, 161-178.

** Webber, R. (1978) Parliamentary Constituencies : a socio-economic classification. Office of Population Censuses and Surveys, occasional paper No.13

Appendix 1: Definition of Variables Used in the Analysis

All the data used in this analysis were supplied by the Social Science Research Council Survey Archive, taken from Study No. 661. Data were extracted from Census 1966: General and Parliamentary Tables (H.M.S.O.: London, 1969). Full definitions of the variables used are given in Census 1966: General and Parliamentary Constituency Tables and The Classification of Occupations 1966 (H.M.S.O.: London, 1967).

Acronym	Definition
(a) Social Class variables	
PROF	Percentage professional workers (i.e. economically active and retired men in socio-economic groups 3 and 4 - described as "professional workers" in the census - as a percentage of all economically active and retired men, excluding the armed forces and inadequately described occupations)
EMPL	Percentage managerial workers (i.e. socio-economic groups 1, 2, 13 - described as "employers and managers" in the census)
NONM	Percentage in routine non-manual work (i.e. socio-economic groups 5 and 6 - described as "non-manual workers")
SKIL	Percentage skilled workers (i.e. socio-economic groups 8, 9, 12 and 14 - described as "foremen, skilled manual workers, and own account workers (other than professional)")
SEMI	Percentage semi-skilled workers (i.e. socio-economic groups 7, 10 and 15 - described as "personal service workers, semi-skilled manual workers and agricultural workers")
UNSK	Percentage unskilled workers (i.e. socio-economic group 11 - described as unskilled manual workers")
(b) Industrial variables	
AGRI	Percentage labour force employed in agriculture (i.e. the percentage of economically active men and women (including those whose occupations are inadequately described) employed in agriculture and related industries (horticulture etc.))
MIN	Percentage labour force employed in mining (including quarrying; and tin and china-clay mining)

Acronym	Definition
MFG	Percentage labour force employed in manufacturing (includes construction, gas, electricity and water)
TRANS	Percentage labour force employed in transport
DIST	Percentage labour force employed in distribution and civilian services
GOVT	Percentage labour force employed in local and national government
(c) Cars	
NOCAR	Percentage of households without a car
(d) Housing	
OWNOCC	Percentage of households living in owner occupied accommodation (including accommodation which is being bought through a loan from a building society, bank, etc.)
COUNCL	Percentage of households living in accommodation rented from the local Council
PRIV	Percentage of households living in accommodation rented privately
AMEN	Percentage of households living in accommodation with exclusive use of inside toilet, bath and hot water tap
HDENS	Percentage of households consisting of one or more persons per room (high density living)
RM3	Percentage of households living in accommodation with three or fewer rooms
RM7	Percentage of households living in accommodation with seven or more rooms
(e) Immigrants	
IRISH	Percentage Irish (i.e. those born in Northern Ireland or the Republic of Ireland as a percentage of the total population in the constituency)
NEWCOM	Percentage from the New Commonwealth (i.e. those born in the Commonwealth, excluding New Zealand, Australia, and Canada, as a percentage of total population living in the constituency)

Acronym	Definition
(f) Age variables	
YOUNG	Percentage young people (i.e. men and women aged 15-24 as a percentage of all men and women aged 15 and over)
OLD	Percentage old people (i.e. men and women aged 65 or over as a percentage of all men and women aged 15 and over)
(g) Unemployment	
UNEMMA	Percentage unemployed (i.e. men not employed during the previous year as a percentage of all economically active men)
(h) Migration	
INMIG	Men and women aged 15 and over who have moved into the Local Authority area from outside during the previous 5 years, as a percentage of all men and women aged 15 and over.
WITHMIG	Men and women aged 15 and over who have moved house within the Local Authority area during the previous 5 years, as a percentage of all men and women aged 15 and over.
(i) Voting*	
PCC66	Percentage voting Conservative in the General Election of 1966, of the total numbers of electors who voted.
PCLAB66	Percentage who voted Labour, of the total number of electors who voted.
PCLIB66	Percentage voting Liberal, of the total number of electors who voted.

All the variables used in the analysis were recorded to decimal place, thus there may be an unquantifiable degree of rounding error present in the analysis.

* A note of the voting figures: to avoid confusion, the speaker's seat was not included. Also, in 1966, there were a few remaining "National Conservative and Liberal" candidates. As in constituencies where they stood, they were sometimes opposed by Liberal candidates but never by Conservative candidates, they have been considered Conservative candidates.

Appendix 2: Computing

All computing in this analysis utilized facilities provided by the Northumbrian Universities Multiple Access Computer Board (NUMAC). Analysis was carried out using the main I.B.M. 370/168 computer, under the Michigan Terminal System (MTS). The analysis was carried out using the package program SPSS (Statistical Package for the Social Sciences), further details of which can be found in Norman H. Nie et. al. (1975) SPSS: Statistical Package for the Social Sciences (New York: McGraw-Hill). The help of the University of Durham Computer Unit is gratefully acknowledged.

Appendix 3: Correlations between the variables

A3.1: Correlations between the predictor and voting variables

Variable	Conservative	Labour	Liberal
PROF	0.69	-0.69	-0.11**
EMPL	0.77	-0.86	0.47
NONM	0.47	-0.45	-0.11**
SKIL	-0.51	0.58	-0.16*
SEMI	-0.46	0.48	-0.08**
UNSK	-0.56	0.62	-0.31
AGRI	0.30	-0.43	0.40
MIN	-0.44	0.40	0.13**
MFG	-0.35	0.48	-0.33
TRANS	-0.17	0.15	-0.28
DIST	0.50	-0.58	0.20
GOVT	0.42	0.46	0.17*
NOCAR	-0.60	0.66	-0.36
OWNOCC	0.53	-0.53	0.29
COUNCL	-0.51	0.57	-0.17*
PRIV	-0.18	0.17	-0.26
AMEN	0.45	-0.47	0.30
HDENS	-0.57	0.61	-0.39
RM3	-0.14*	0.11**	0.24
RM7	0.66	-0.72	0.45
IRISH	0.08**	-0.05**	-0.34
NEWCOM	0.09**	-0.08**	-0.28
YOUNG	-0.29	0.46	-0.42
OLD	0.32	-0.45	0.35
UNEMMA	-0.35	0.32	0.06**
INMIG	0.37	-0.45	0.02**
WITHMIG	-0.37	0.46	0.22

* not significant at α = 0.01
** not significant at α = 0.001

A3.2: Correlations between the predictor variables

These correlations are given on computer output contained in the slip-case at the end of the dissertation. Pages 1A - 1C give the untransformed correlations, 2A - 2C give the transformed correlations.

It should be noted that, although the acronyms listed in table 2.2 are used, the prefix 'T' is used for the transformed variable.

Appendix 4: Results of Principal Component and Factor Analyses

A4.1 Untransformed variables: P.C.A. 2-component solution

Variable	Component 1	Component 2
PROF	0.76	0.33
EMPL	0.93	0.14
NONM	0.58	0.62
SKIL	-0.66	-0.53
SEMI	-0.51	-0.42
UNSK	-0.79	0.14
AGRI	0.34	-0.34
MIN	-0.25	-0.34
MIG	-0.52	-0.36
TRANS	-0.20	0.46
DIST	0.60	0.67
GOVT	0.38	0.10
NOCAR	-0.80	0.36
OWNOCC	0.70	-0.42
COUNCL	-0.58	-0.30
PRIV	-0.33	0.83
AMEN	0.68	-0.46
HDENS	-0.74	0.48
RM3	-0.23	0.87
RM7	0.68	0.11
IRISH	-0.21	0.73
NEWCOM	-0.24	0.74
YOUNG	-0.49	0.17
OLD	0.38	0.07
UNEMMA	-0.48	0.17
INMIG	0.55	0.54
WITHMIG	-0.60	0.12

For eigenvalues see Table 6.1

A4.2 Transformed variables: P.C.A. 2- and 6- component solution

	Component Loadings					
Variable	1	2	3	4	5	6
PROF	0.85	0.10	-0.27	0.07	0.11	-0.22
EMPL	0.94	-0.49	0.06	0.00	0.57	-0.09
NONM	0.67	0.57	-0.25	-0.03	0.25	0.12
SKIL	-0.70	-0.42	-0.16	-0.27	-0.17	0.09
SEMI	-0.61	-0.40	0.38	0.27	-0.24	-0.12
UNSK	-0.80	0.17	0.24	0.03	-0.13	0.31
AGRI	0.36	-0.60	0.48	0.24	-0.29	0.05
MIN	-0.40	-0.50	0.27	0.06	0.31	-0.48
MFG	-0.54	-0.24	-0.64	-0.35	0.16	0.03
TRANS	-0.06	0.55	0.10	0.09	0.22	0.59
DIST	0.66	0.58	0.27	0.04	0.21	0.07
GOVT	0.54	0.13	0.34	0.40	-0.05	0.20
NOCAR	-0.77	0.47	0.18	-0.23	0.15	-0.04
OWNOCC	0.61	-0.50	0.17	-0.42	-0.11	0.09
COUNCL	-0.61	-0.29	-0.19	0.38	0.28	0.21
PRIV	-0.22	0.81	0.29	-0.08	-0.14	-0.22
AMEN	0.63	-0.46	0.31	0.15	0.35	0.13
HDENS	-0.75	0.44	-0.01	0.29	-0.02	-0.12
RM3	-0.12	0.86	0.11	0.05	0.07	-0.23
RM7	0.69	0.05	0.28	-0.01	-0.26	0.23
IRISH	0.06	0.74	-0.36	0.01	-0.33	0.02
NEWCOM	0.07	0.76	-0.22	-0.02	-0.42	-0.03
YOUNG	-0.48	0.21	-0.36	0.51	-0.02	0.00
OLD	0.36	0.06	0.64	-0.54	-0.04	0.05
UNEMMA	-0.49	0.20	0.59	-0.16	0.30	0.02
INMIG	0.62	0.38	-0.17	0.07	0.02	-0.27
WITHMIG	-0.52	0.30	-0.23	-0.41	0.35	0.08

A4.3 Communalities for variables: 2- and 6- component P.C.A. and 2-factor varimax

Variable	2-component communality	6-component communality
PROF	0.73	0.88
EMPL	0.90	0.91
NONM	0.76	0.91
SKIL	0.65	0.80
SEMI	0.49	0.83
UNSK	0.64	0.83
AGRI	0.43	0.88
MIN	0.36	0.82
MFG	0.32	0.92
TRANS	0.25	0.74
DIST	0.78	0.90
GOVT	0.27	0.62
NOCAR	0.82	0.93
OWNOCC	0.58	0.84
COUNCL	0.41	0.76
PRIV	0.69	0.87
AMEN	0.58	0.88
HDENS	0.73	0.85
RM3	0.76	0.84
RM7	0.44	0.68
IRISH	0.49	0.79
NEWCOM	0.52	0.81
YOUNG	0.23	0.66
OLD	0.10	0.83
UNEMMA	0.25	0.75
INMIG	0.49	0.64
WITHMIG	0.32	0.72

A4.4 Transformed variables: 2- factor varimax solution

Variable	Factor 1	Factor 2
PROF	0.78	-0.35
EMPL	0.80	-0.54
NONM	0.86	0.14
SKIL	-0.80	0.00
SEMI	-0.71	-0.02
UNSK	-0.59	0.55
AGRI	0.00	-0.66
MIN	-0.56	-0.19
MFG	-0.56	0.07
TRANS	0.21	0.46
DIST	0.87	0.16
GOVT	0.49	-0.16
NOCAR	-0.41	0.81
OWNOCC	0.26	-0.72
COUNCL	-0.64	0.07
PRIV	0.22	0.80
AMEN	0.30	-0.70
HDENS	-0.41	0.75
RM3	-0.35	0.80
RM7	0.59	-0.30
IRISH	0.41	0.57
NEWCOM	0.43	0.58
YOUNG	-0.28	0.38
OLD	0.30	-0.12
UNEMMA	-0.28	0.40
INMIG	0.70	-0.00
WITHMIG	-0.28	0.49

Appendix 5: Scattergrams

All the scattergrams referred to individually are to be found in the slip-case at the end of the dissertation. It should be noted that although the acronyms listed in table 2.2 are used, the prefix 'T' is included in the transformed variables. The scattergrams included are:

Untransformed variables

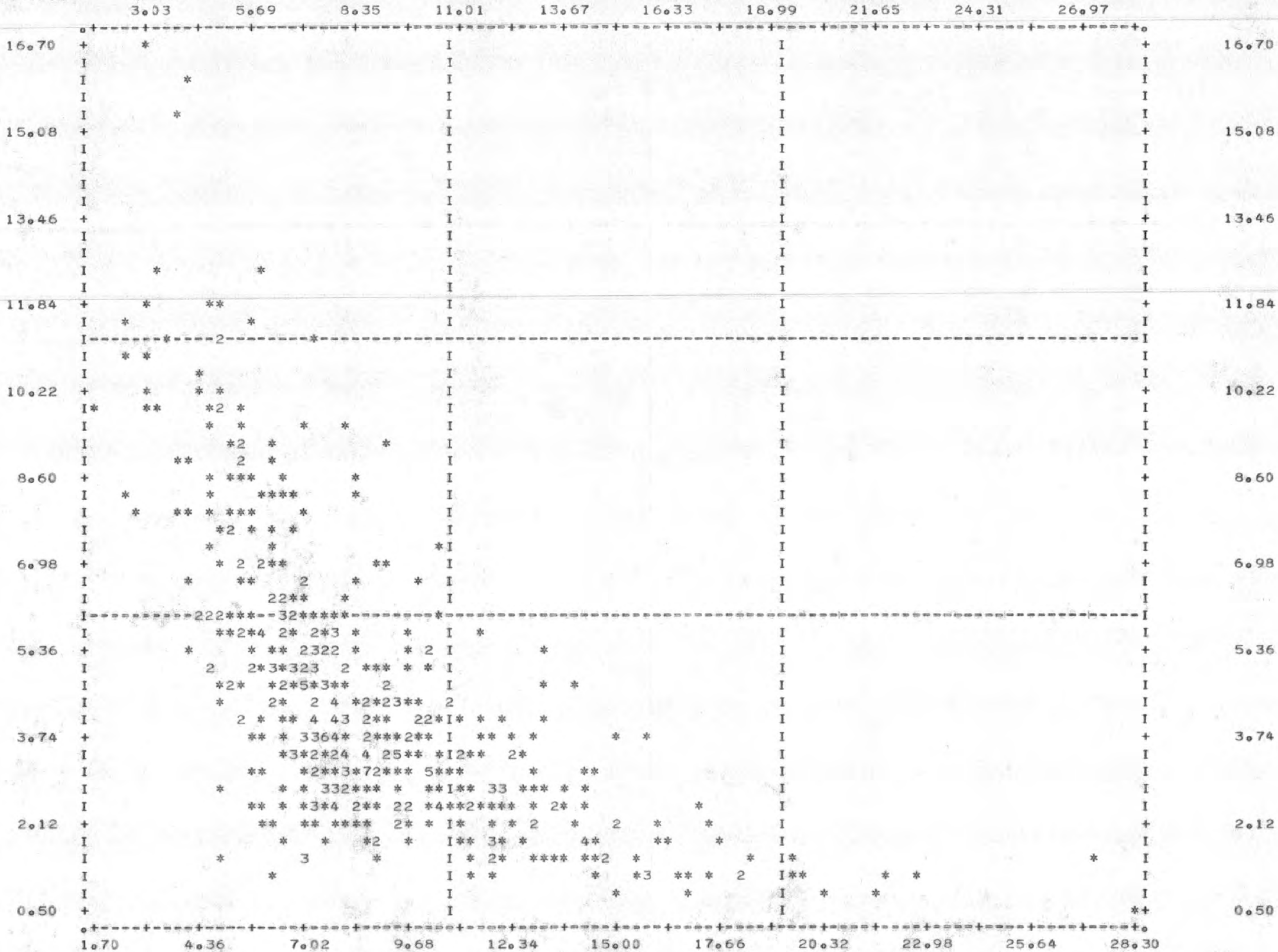
Number	Variables
3	PROF - UNSK
4	PROF - HDENS
5	PROF - RM3
6	EMPL - UNSK
7	EMPL - NOCAR
8	SEMI - UNSK
9	AGRI - IRISH
10	MIN - NEWCOM
11	DIST - GOVT
12	PRIV - OLD
13	RM3 - NEWCOM
14	RM3 - INMIG
15	RM3 - WITHMIG
16	IRISH - PRIV
17	IRISH - UNEMMA
18	PCC66 - PCLAB66
19	OWNOCC - PRIV
20	OWNOCC - COUNCL
21	PRIV - COUNCL
22	MFG - DIST

Transformed variables

Number	Variables
3T	TPROF - TUNSK
4T	TPROF - THDENS
5T	TPROF - TRM3
6T	TEMPL - TUNSK
7T	TEMPL - TNOCAR
8T	TSEMI - TUNSK
9T	TAGRI - TIRISH
10T	TMIN - TNEWCOM
11T	TDIST - TGOVT
12T	TPRIV - TOLD
13T	TRM3 - TNEWCOM
14T	TRM3 - TINMIG
15T	TRM3 - TWITHMIG
16T	TIRISH - TPRIV
17T	TIRISH - TUNEMMA

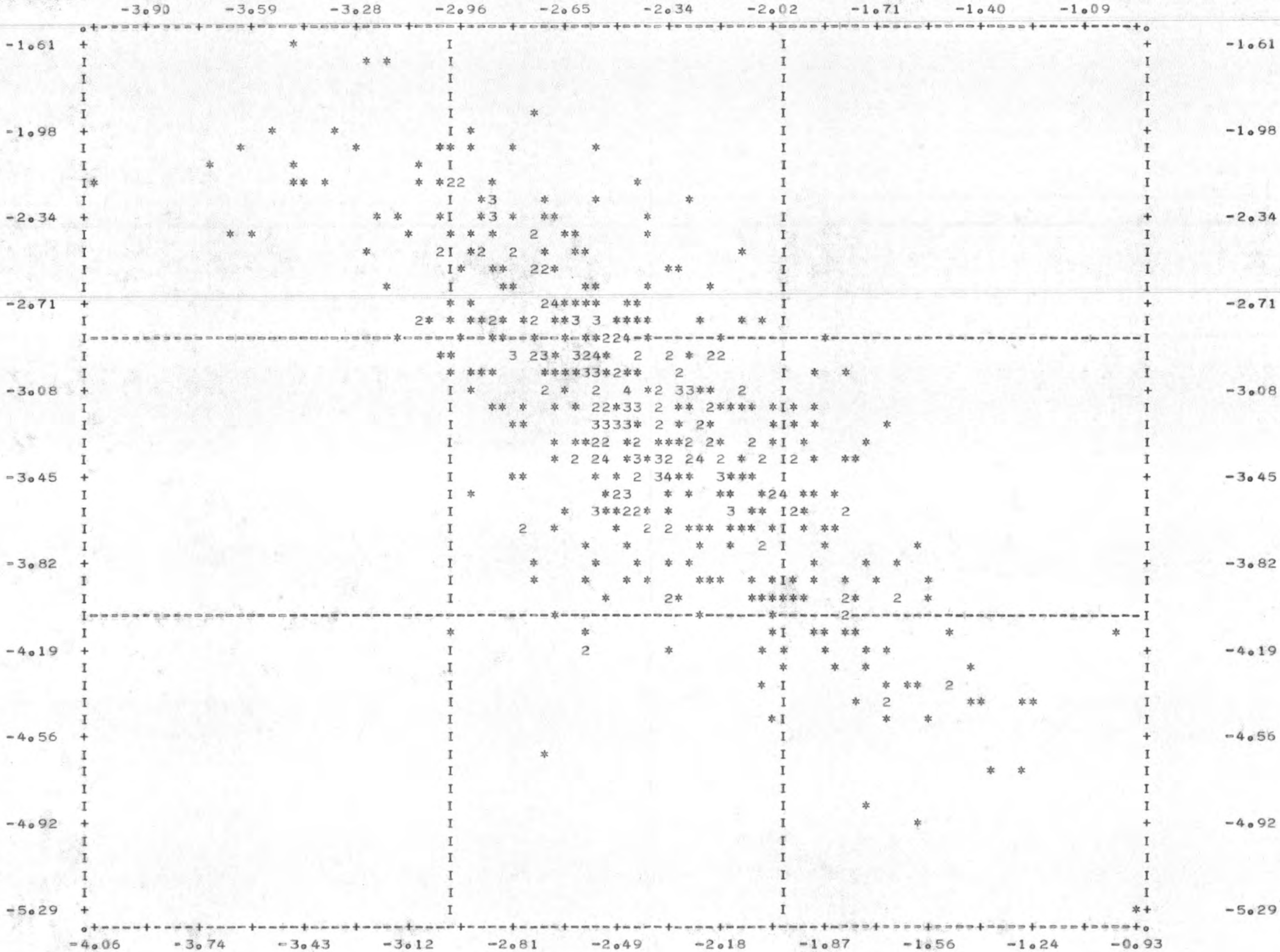
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) PROF (ACROSS) UNSK



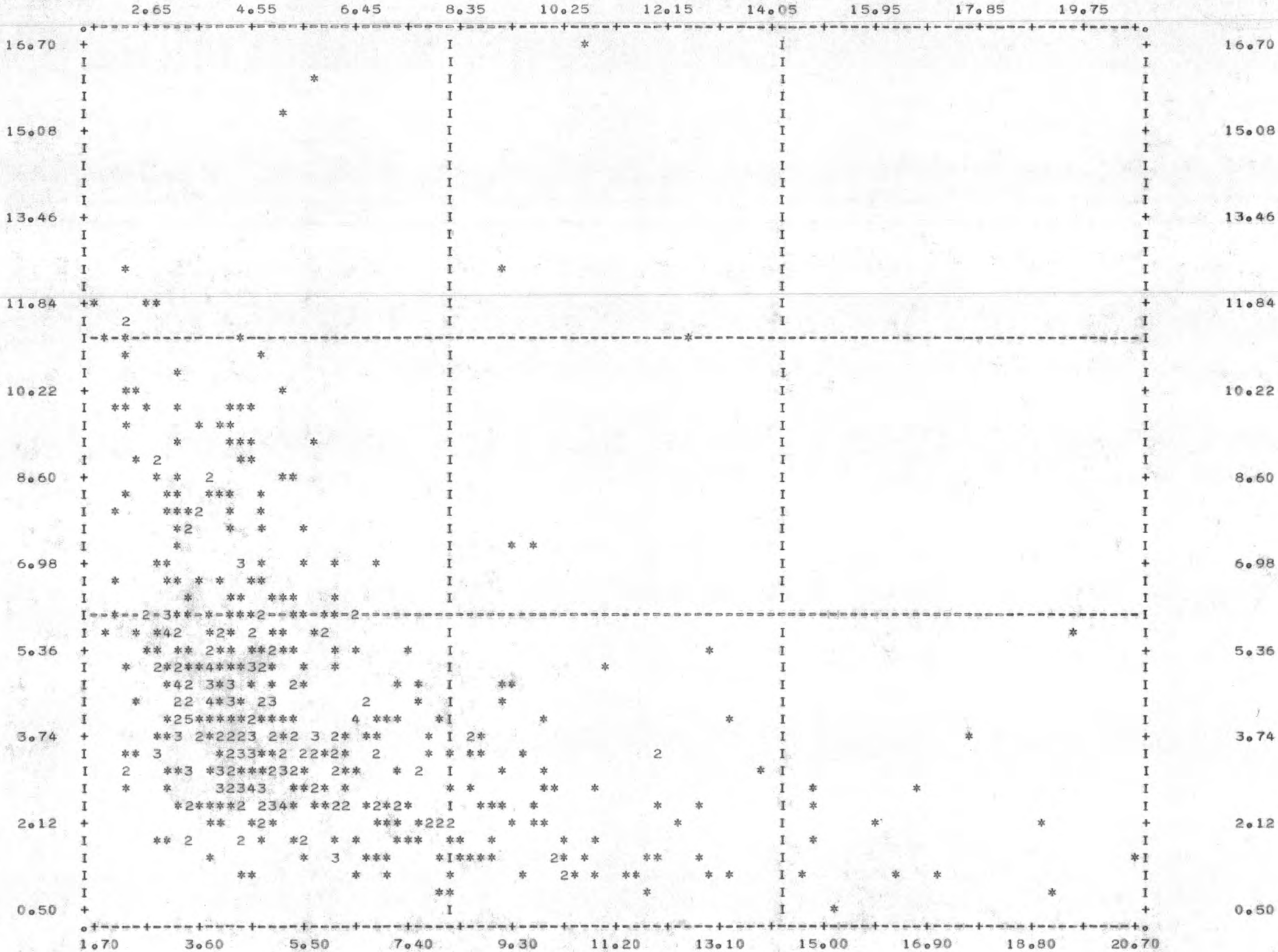
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) TPROF (ACROSS) TUNSK

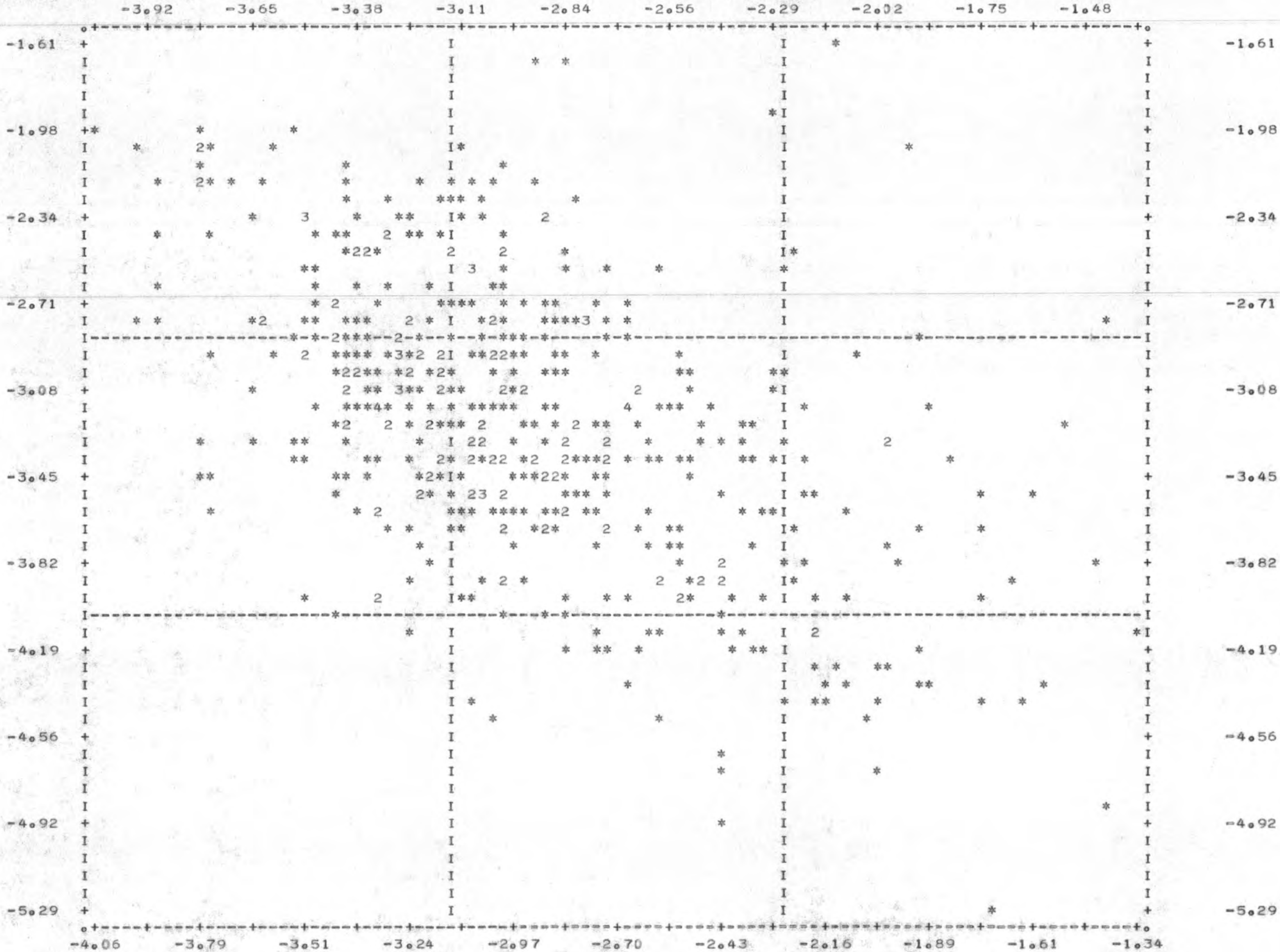


FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) PROF (ACROSS) HDENS

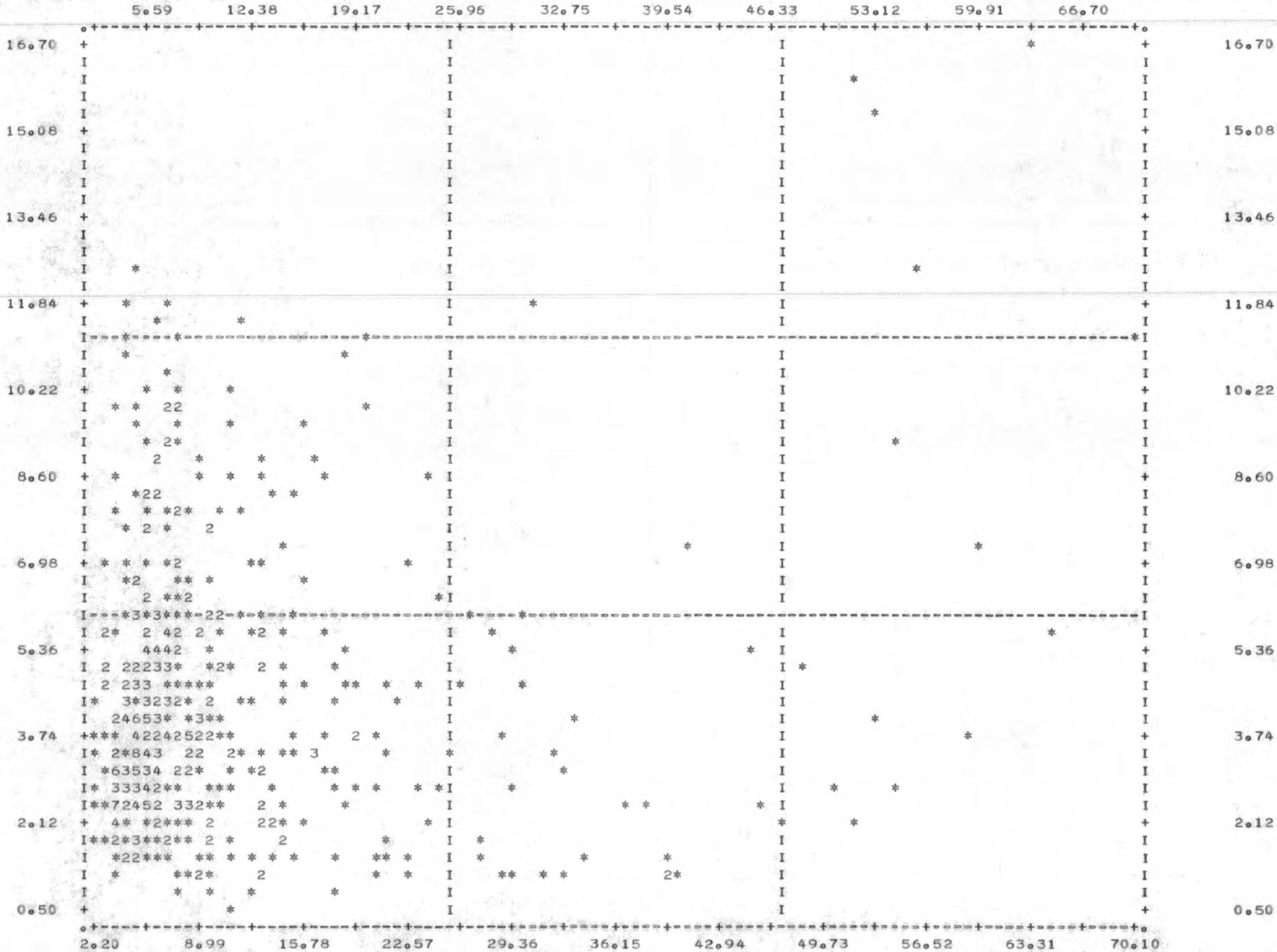


FILE TRANS (CREATION DATE = 02/25/80)
SCATTERGRAM OF (DOWN) TPROF (ACROSS) THDENS

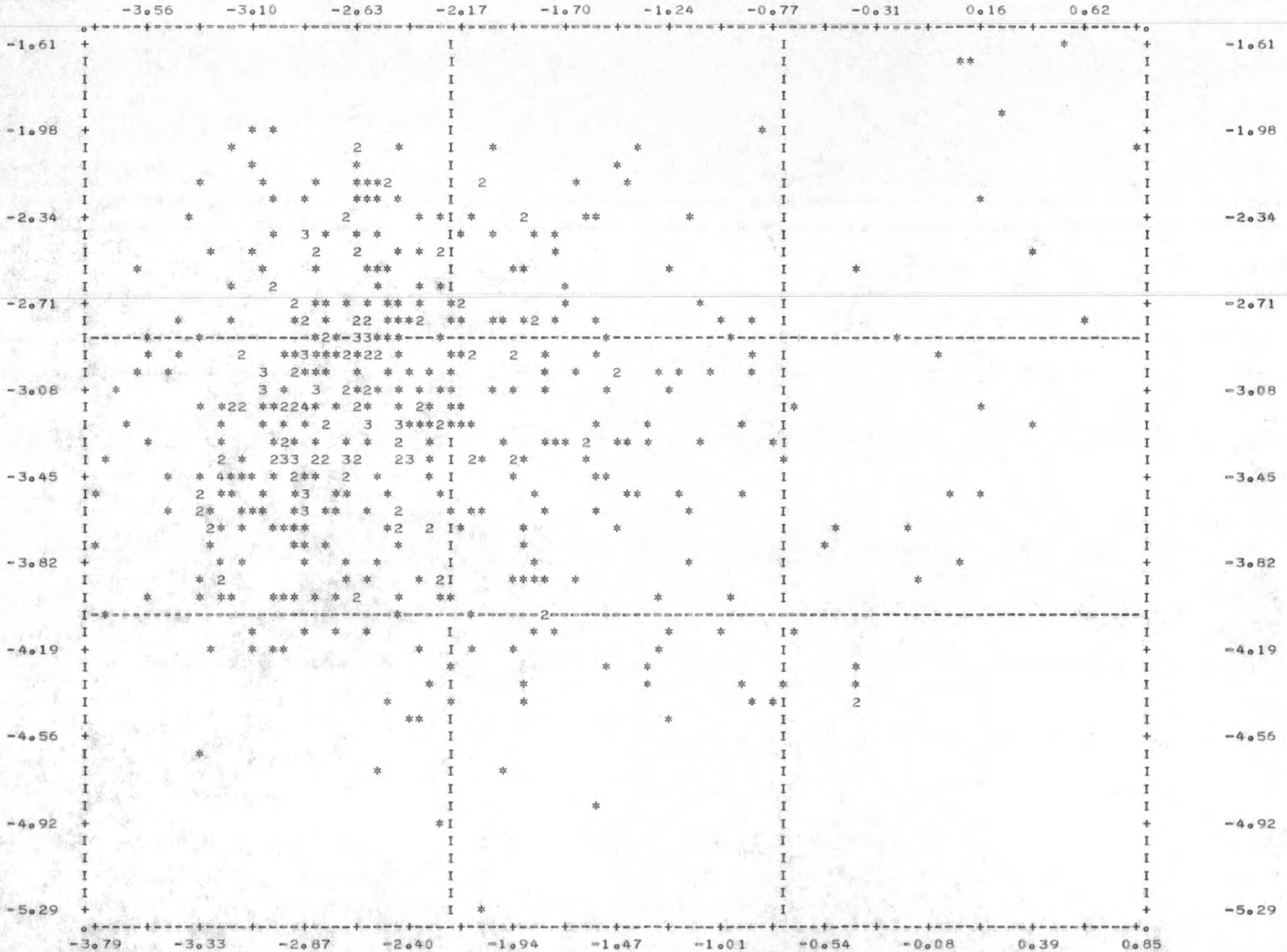


FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) PROF (ACROSS) RM3

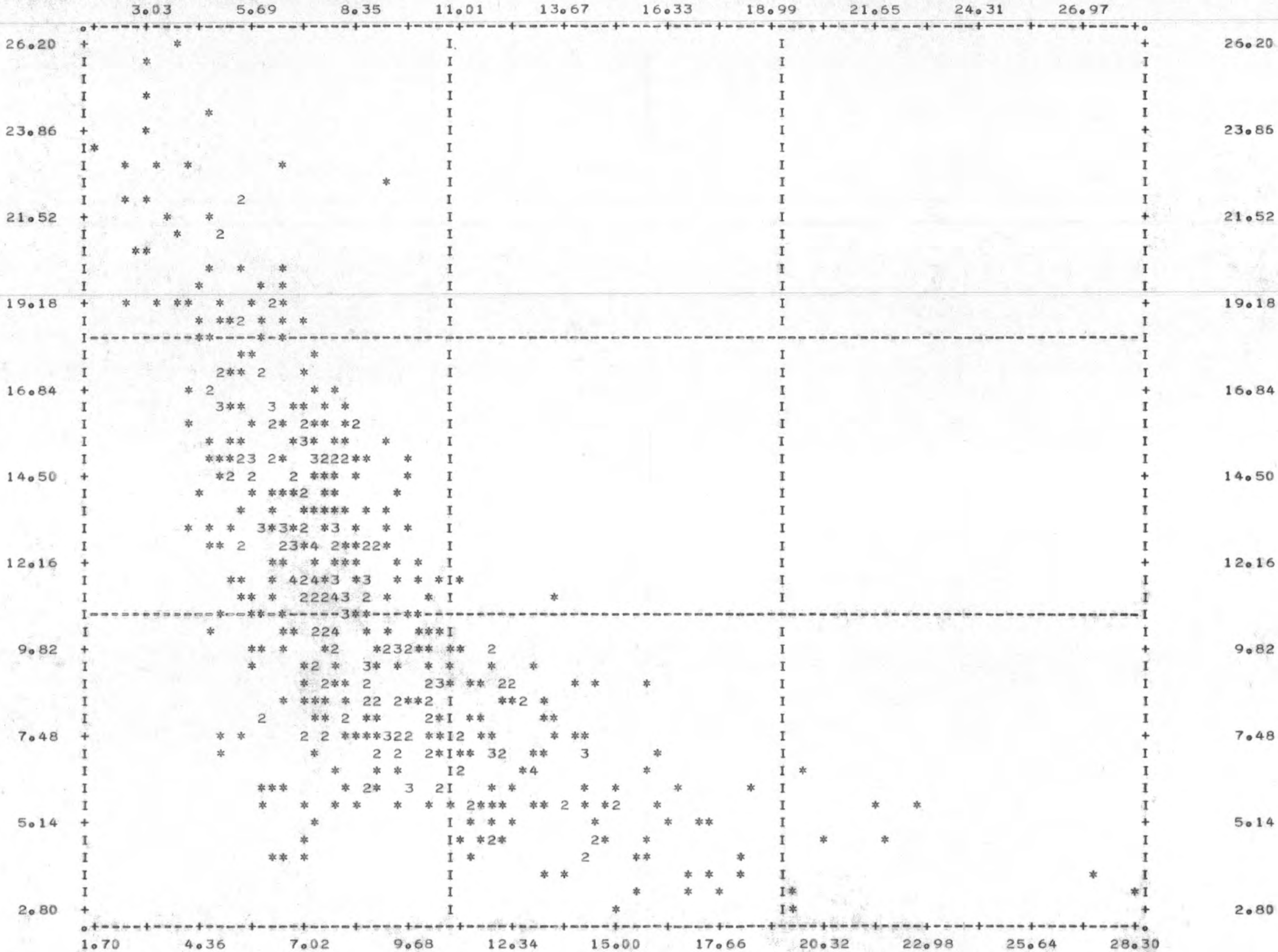


FILE TRANS (CREATION DATE = 02/25/80)
SCATTERGRAM OF (DOWN) TPROF (ACROSS) TRM3

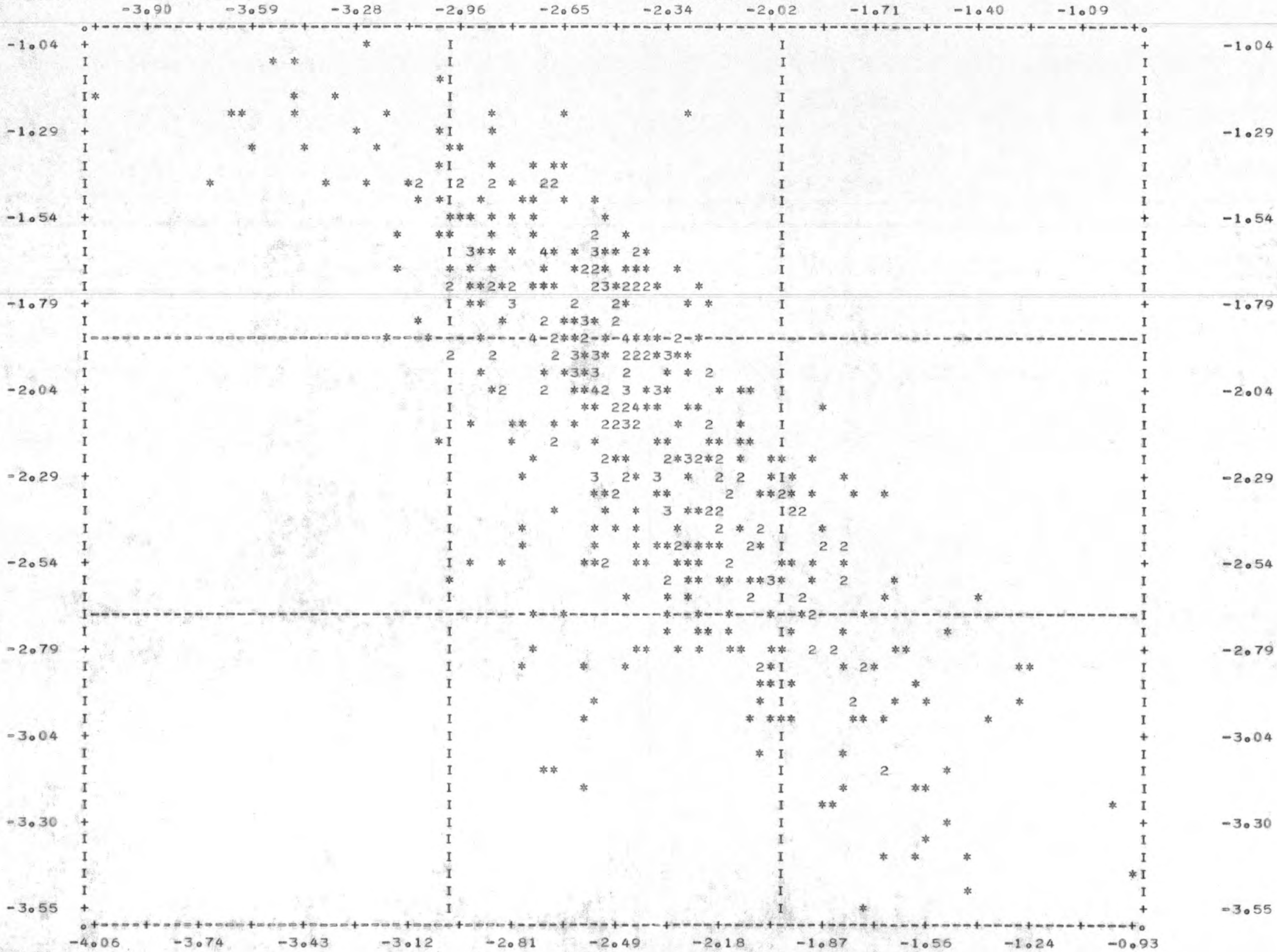


FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) EMPL (ACROSS) UNSK

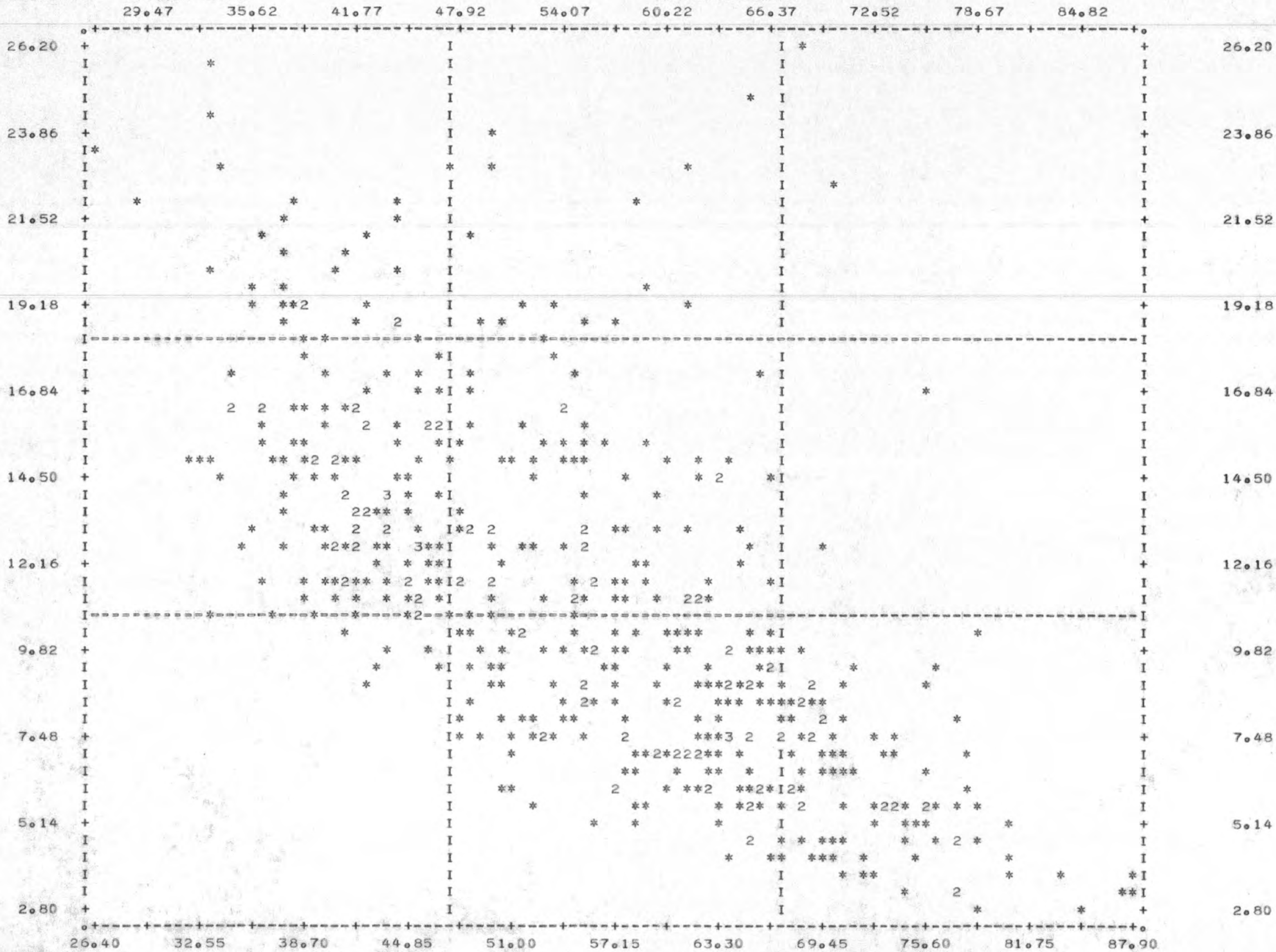


FILE TRANS (CREATION DATE = 02/25/80)
SCATTERGRAM OF (DOWN) TEMPL (ACROSS) TUNSK



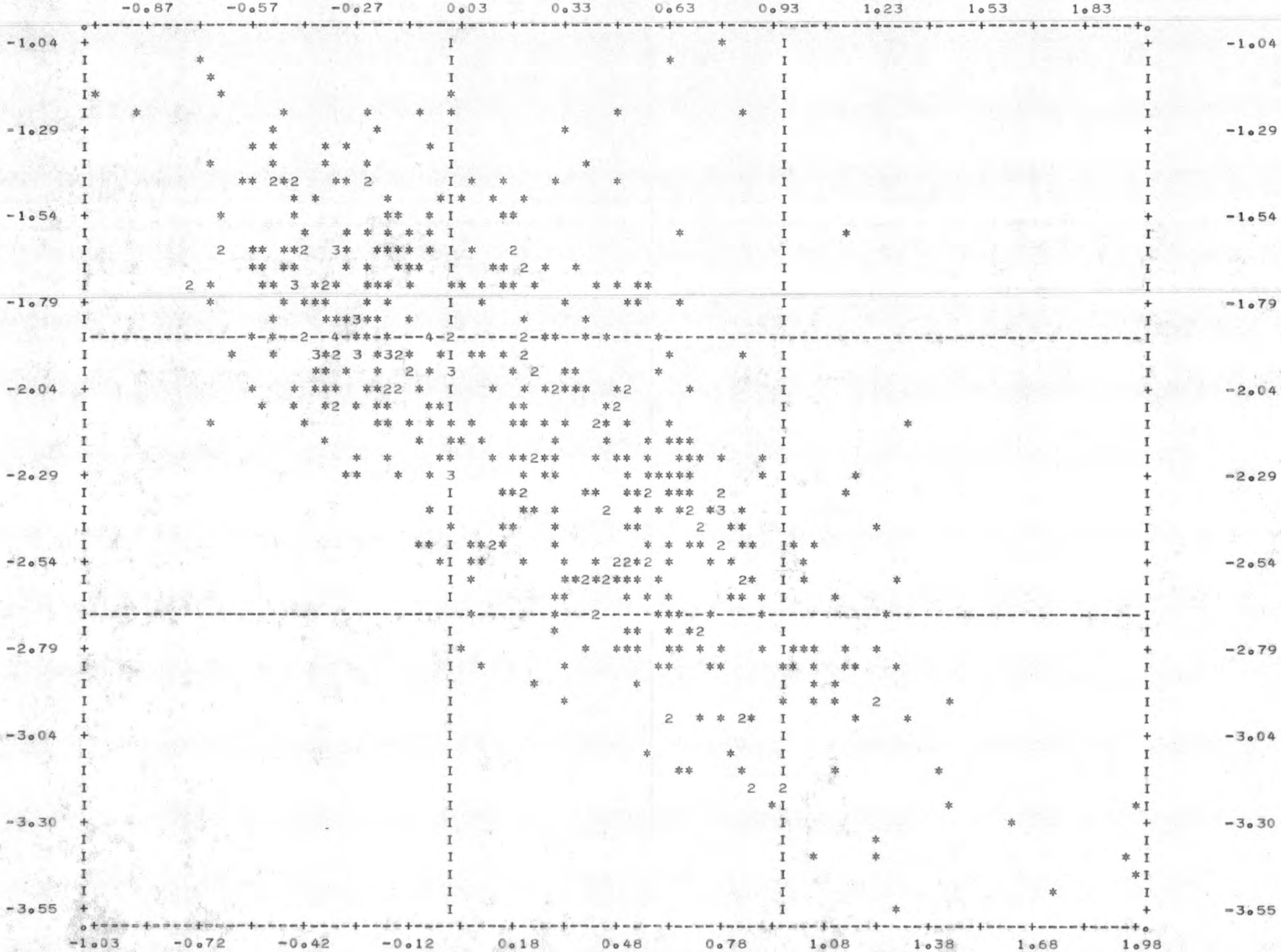
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) EMPL (ACROSS) NOCAR



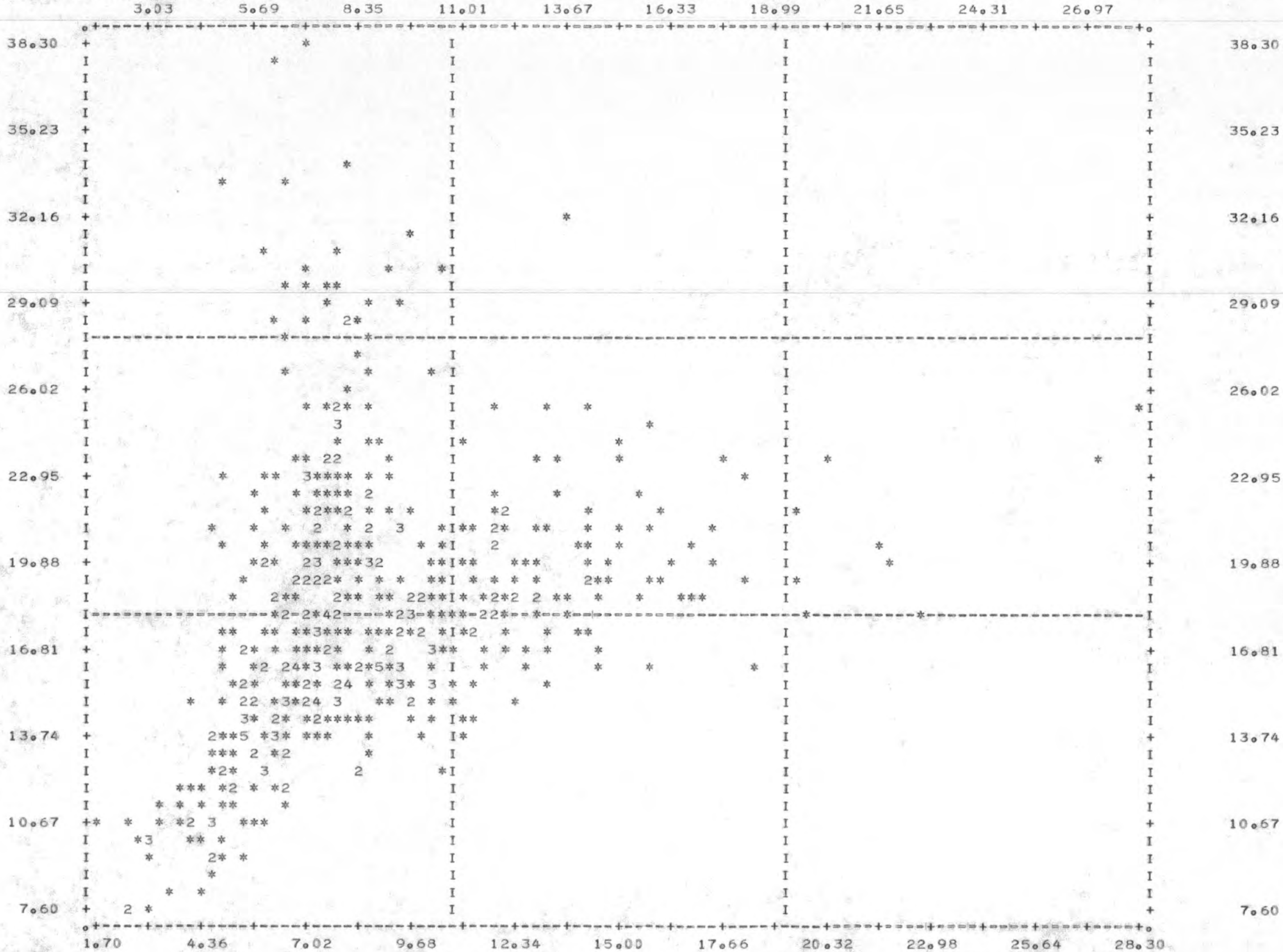
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) TEMPL (ACROSS) TNOCAR



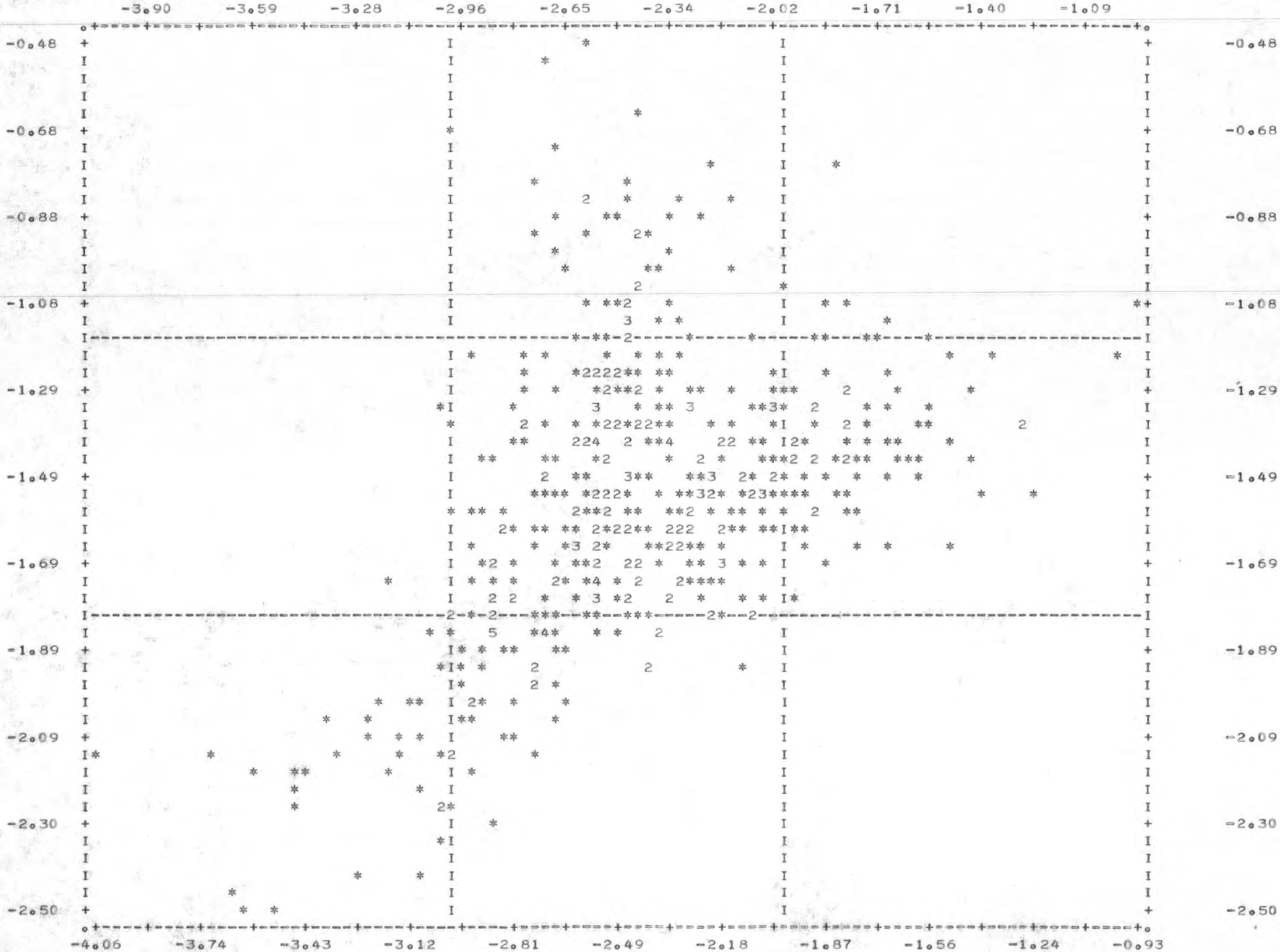
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) SEMI (ACROSS) UNSK



FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) TSEMI (ACROSS) TUNSK



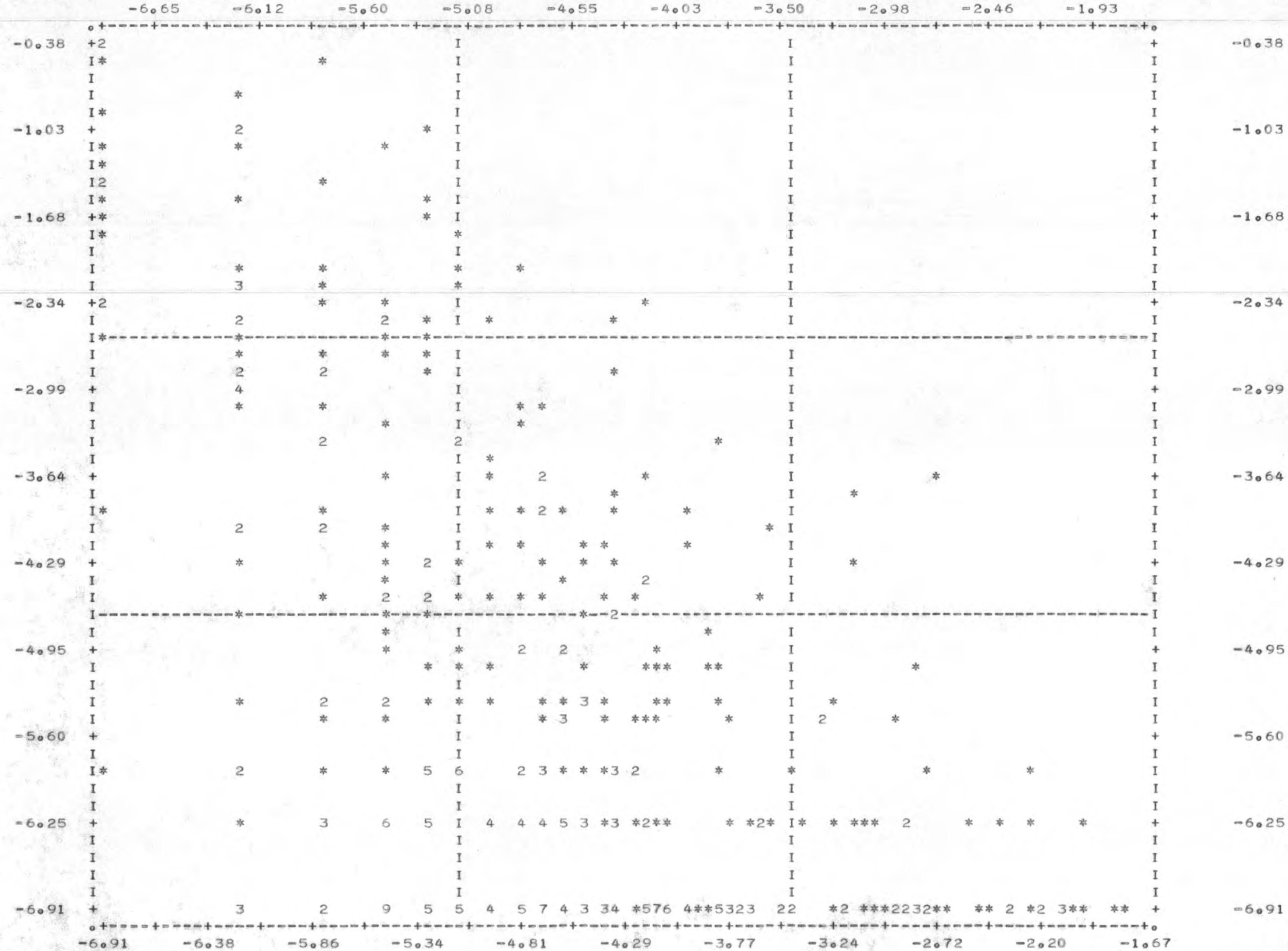
```
FILE   TRANS    (CREATION DATE = 02/25/80)
SCATTERGRAM OF    (DOWN) AGRI                                   (ACROSS) IRISH
               0.82      2.26      3.70      5.14      6.58      8.02      9.46     10.90     12.34     13.78
          o+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+o
     27.70 + *                                I                               I                                +      27.70
           I                                  I                               I                                I
           I                                  I                               I                                I
           I                                  I                               I                                I
           I                                  I                               I                                I
     24.93 +      *                           I                               I                                +      24.93
           I    *                             I                               I                                I
           I    *                             I                               I                                I
           I     *                            I                               I                                I
           I   2                              I                               I                                I
     22.16 +                                  I                               I                                +      22.16
           I   *                              I                               I                                I
           I   *                              I                               I                                I
           I   *                              I                               I                                I
           I    *                             I                               I                                I
     19.39 +                                  I                               I                                +      19.39
           I     *                            I                               I                                I
           I---*-----------------------------------------------------------------------------------------------I
           I   ***                            I                               I                                I
           I      *                           I                               I                                I
     16.62 +                                  I                               I                                +      16.62
           I        *                         I                               I                                I
           I    * *                           I                               I                                I
           I                                  I                               I                                I
           I                                  I                               I                                I
     13.85 +   * *2 2                         I                               I                                +      13.85
           I   ****                           I                               I                                I
           I  *  *  *                         I                               I                                I
           I     *2                           I                               I                                I
           I   * ****                         I                               I                                I
     11.08 +      *                           I                               I                                +      11.08
           I       ***                        I                               I                                I
           I   2 **  *                        I                               I                                I
           I--*-*-*-*-*------------------------------------------------------------------------------------------I
           I    *  **                         I                               I                                I
      8.31 +     *5   2*                      I                               I                                +       8.31
           I      *22                         I                               I                                I
           I      2  *                        I                               I                                I
           I    ** * ** 2                     I                               I                                I
           I  ** * 22 *                       I                               I                                I
      5.54 + **3   22  *3  *                  I                               I                                +       5.54
           I * * ***22                        I                               I                                I
           I  *  *22 2**                      I                               I                                I
           I   *  **3 *                       I                               I                                I
           I * *  *2 ****    * * *            I                               I                                I
      2.77 +   *2*  225 2 *                   I                               I                                +       2.77
           I* * *22*  2 * *2       *          I                               I                                I
           I *2*332*2*2 3  22      *          I                               I                                I
           I * 4**3 *3542 *2**      *         I                               I                                I
           I 439449395936798*323* * * *      *I         2                     I                                I
       0.0 + 4243563*295944229563254335*33**232* *2*3 2 2 *2 *2** * * *** * ***  2     *                      *+       0.0
          o+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+o
           0.10      1.54      2.98      4.42      5.86      7.30      8.74     10.18     11.62     13.06     14.50
```

```
FILE    TRANS     (CREATION DATE = 02/25/80)
SCATTERGRAM OF    (DOWN) TAGRI                                        (ACROSS) TIRISH
                -6.65     -6.14     -5.62     -5.11     -4.60     -4.08     -3.57     -3.06     -2.54     -2.03
        o+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+o
-0.96    +                     *            I                               I                                  +   -0.96
         I                                  I        *                      I                                  I
         I                               *  I*  2  *                        I                                  I
         I                               3  I                               I                                  I
         I                                  I*  *  *                        I                                  I
-1.55    +                                  I*  *  * *                      I                                  +   -1.55
         I                                  I   *      *   *                I                                  I
         I                                  I*     * * *   2                I                                  I
         I                           *   *  I   *  2   2    *               I                                  I
         I                                  I*     2 * * * *                I                                  I
-2.15    +                                  I*     * * * * *  2             I                                  +   -2.15
         I                           *   *  I   *      *   **  *            I                                  I
         I                                  I*     * 3 2 *     2  *         I                                  I
         I                                  I        * * 2 2                I                                  I
         I                                  I   *  * *   *    2*   **       I                                  I
-2.74    +                           2      I*     *     3 3    *           I                                  +   -2.74
         I              2                   I3           *  * *   **2   *   I                                  I
         I---------------------------*-------*-----2-*-*-2-2--3---*--------------------------------------------I
         I                                  I*         2 *      2           I                                  I
         I                                  I            *  3               I                                  I
-3.34    +                     *         *  I          * 2        * *      *I *                                +   -3.34
         I                                  I   *             3 3        *  I                                  I
         I                                  I*  *           2   *  ** *     I                                  I
         I                                  I   *  2   2       3    *   2   I  *                               I
         I*                                 I            *     *      *     I                                  I
-3.93    +                           *      I      *   *    *  *   *        I                                  +   -3.93
         I              *            2   *  I   2  2                   *    I  *                               I
         I                                  I          * *     *   *    3   I                                  I
         I                                  I*  2  *        * *     *       I                                  I
         I                     *            I                  *  *   **2   I                                  I
-4.53    +                                  I          *      ***   *    *  I                                  +   -4.53
         I                               3  I      * * *   *  2** 3*        I    *                             I
         I              *      *         *  I        2 *    *  *     *  *   I                                  I
         I-----------------------------------------------------------------------------------------------------I
         I                     2            I2           * ** 2        *    I                                  I
-5.12    +                           *   *  I*     * *   * *   *2  ***2 *   I                                  +   -5.12
         I                           *   2  I   2      *   *  *** 2  2 *   3I                                  I
         I                                  I                               I                                  I
         I                               2  I*  *  *   * * ** *** ****33* 2 I                *                 I
         I                                  I                               I                                  I
-5.72    +                                  I                               I                                  +   -5.72
         I                           *      I2  *  2 2 2   *2 **3  2 26*  *2I** * *   *     *                  I
         I                                  I                               I                                  I
         I                                  I                               I                                  I
         I                                  I   *  2   * *     22 4 **2 422*232** *2  2   *     **             I
-6.31    +                                  I                               I                                  +   -6.31
         I                                  I                               I                                  I
         I                                  I                               I                                  I
         I                                  I                               I                                  I
         I                                  I                               I                                  I
-6.91    +              *      2     2   *  I2  *  2 3 * *    **2  362   225* 4**22 3222  3***22 **2 *2 *    **+   -6.91
        o+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+o
       -6.91     -6.39     -5.88     -5.37     -4.85     -4.34     -3.83     -3.31     -2.80     -2.29     -1.77
```

```
FILE    TRANS     (CREATION DATE = 02/25/80)
SCATTERGRAM OF    (DOWN) MIN                                        (ACROSS) NEWCOM
                  0.79      2.37      3.95      5.53      7.11      8.69     10.27     11.85     13.43     15.01
            o+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+o
      40.70  + 2                                I                               I                                  +   40.70
             I                                  I                               I                                  I
             I                                  I                               I                                  I
             I *                                I                               I                                  I
             I                                  I                               I                                  I
      36.63  +  *                               I                               I                                  +   36.63
             I                                  I                               I                                  I
             I                                  I                               I                                  I
             I                                  I                               I                                  I
             I                                  I                               I                                  I
      32.56  +                                  I                               I                                  +   32.56
             I *                                I                               I                                  I
             I                                  I                               I                                  I
             I                                  I                               I                                  I
             I                                  I                               I                                  I
      28.49  +                                  I                               I                                  +   28.49
             I *                                I                               I                                  I
             I-*-----------------------------------------------------------------------------------------------------I
             I *                                I                               I                                  I
             I   *                              I                               I                                  I
      24.42  + * *                              I                               I                                  +   24.42
             I *                                I                               I                                  I
             I                                  I                               I                                  I
             I                                  I                               I                                  I
             I *                                I                               I                                  I
      20.35  +                                  I                               I                                  +   20.35
             I **                               I                               I                                  I
             I *                                I                               I                                  I
             I                                  I                               I                                  I
             I 2 *                              I                               I                                  I
      16.28  +                                  I                               I                                  +   16.28
             I * *                              I                               I                                  I
             I                                  I                               I                                  I
             I-*-----------------------------------------------------------------------------------------------------I
             I    *                             I                               I                                  I
      12.21  +*                                 I                               I                                  +   12.21
             I ** *                             I                               I                                  I
             I *  **                            I                               I                                  I
             I 2*                               I                               I                                  I
             I 2**                              I                               I                                  I
       8.14  +   2*    **                       I                               I                                  +    8.14
             I 2 *                              I                               I                                  I
             I 2 4                              I                               I                                  I
             I 23*    *                         I                               I                                  I
             I 5                                I                               I                                  I
       4.07  + *** **                           I                               I                                  +    4.07
             I  2 3*     *                      I                               I                                  I
             I   *2 2 ** *              *       I     *                         I                                  I
             I 43522423  *      *       *       I                               I                                  I
             I 23973654432 3* *      *          I*                              I                                  I
        0.0  + 8999999999949799 534323*2*4* 4 33423*22 ** ** *  2* 23 ** ***2*  I*2*   2 3 *    ***  *   *        **+    0.0
            o+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+o
             0.0       1.58      3.16      4.74      6.32      7.90      9.48     11.06     12.64     14.22     15.80
```

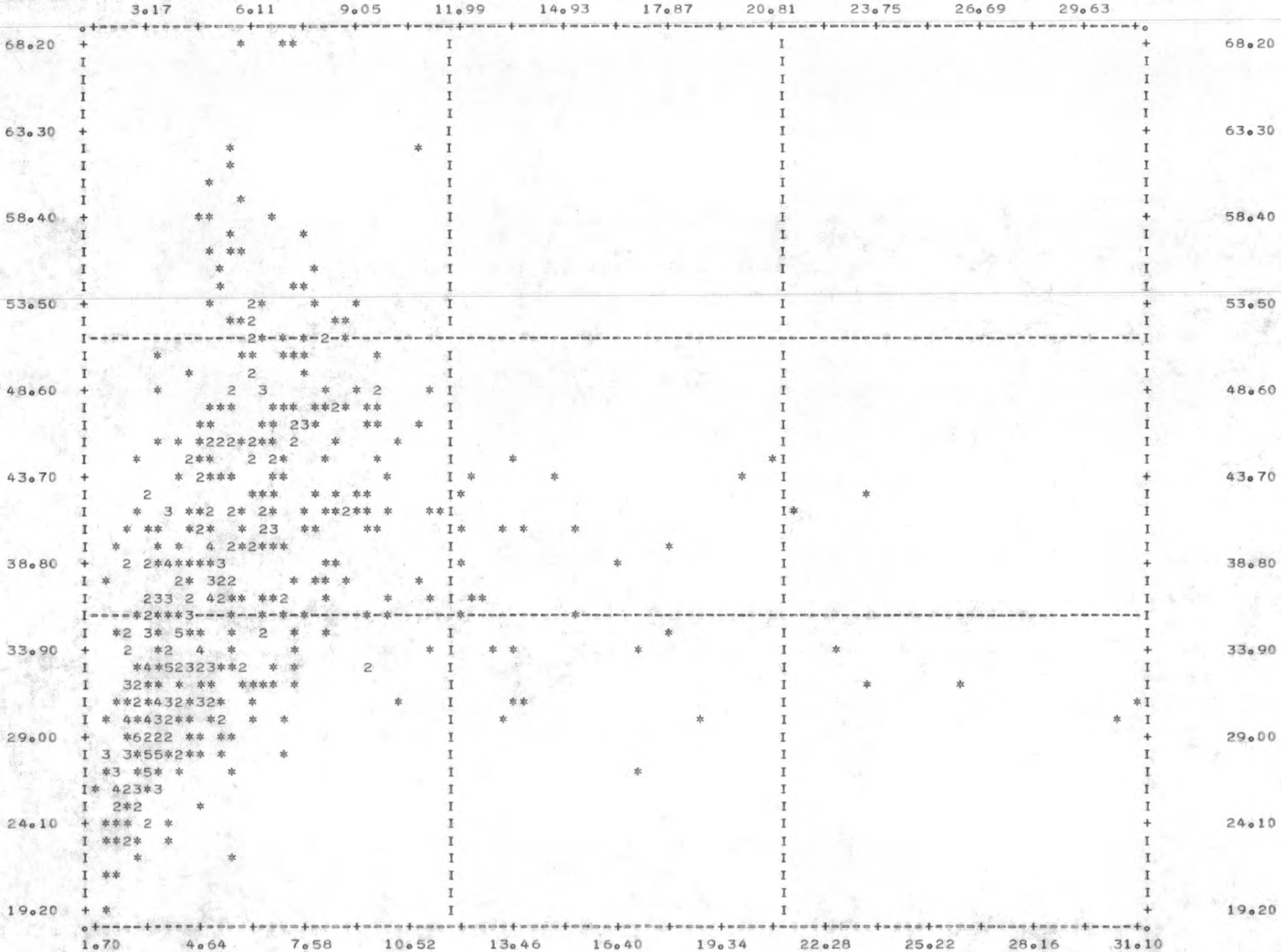
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) TMIN (ACROSS) TNEWCOM



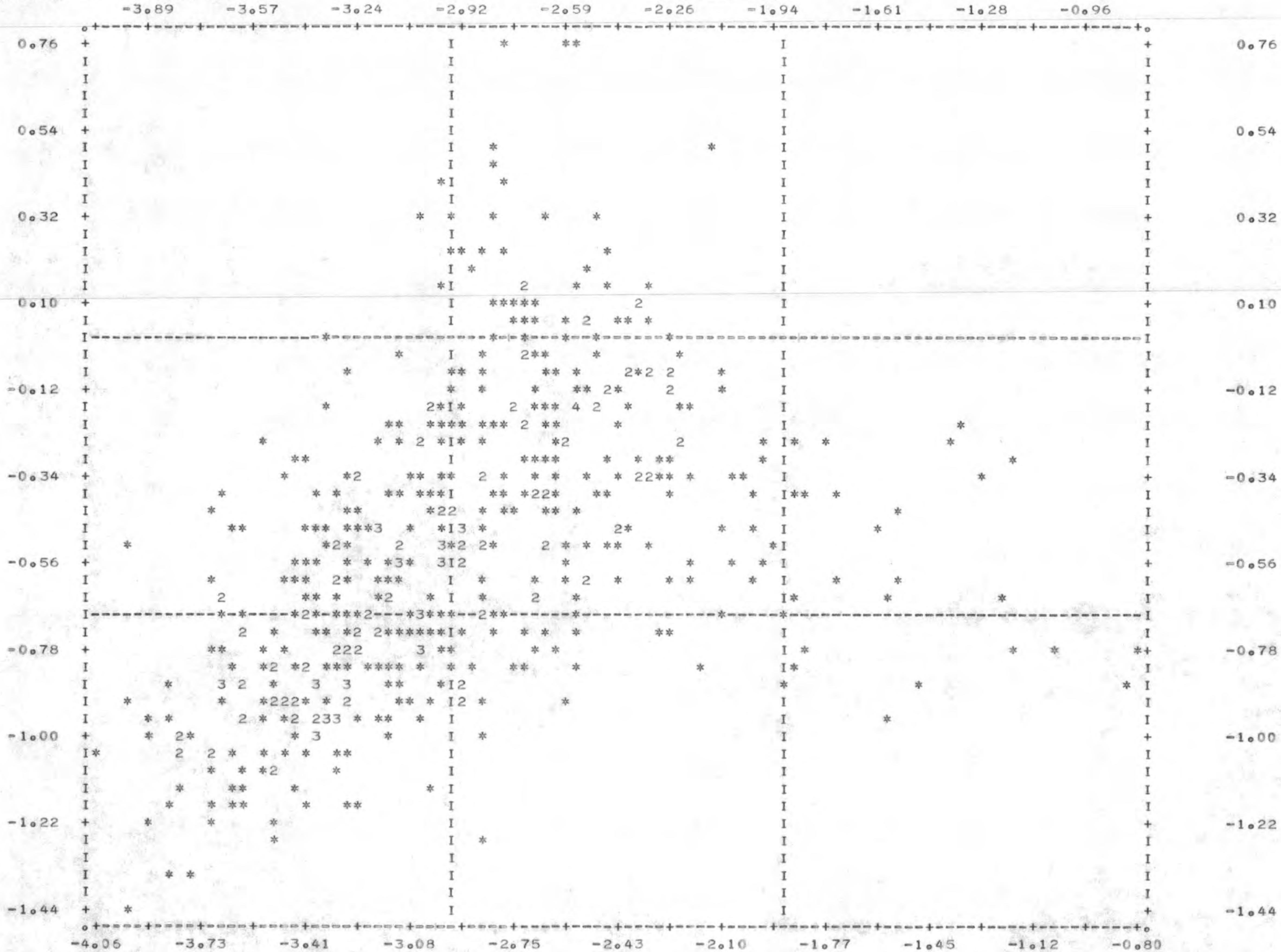
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) DIST (ACROSS) GOVT



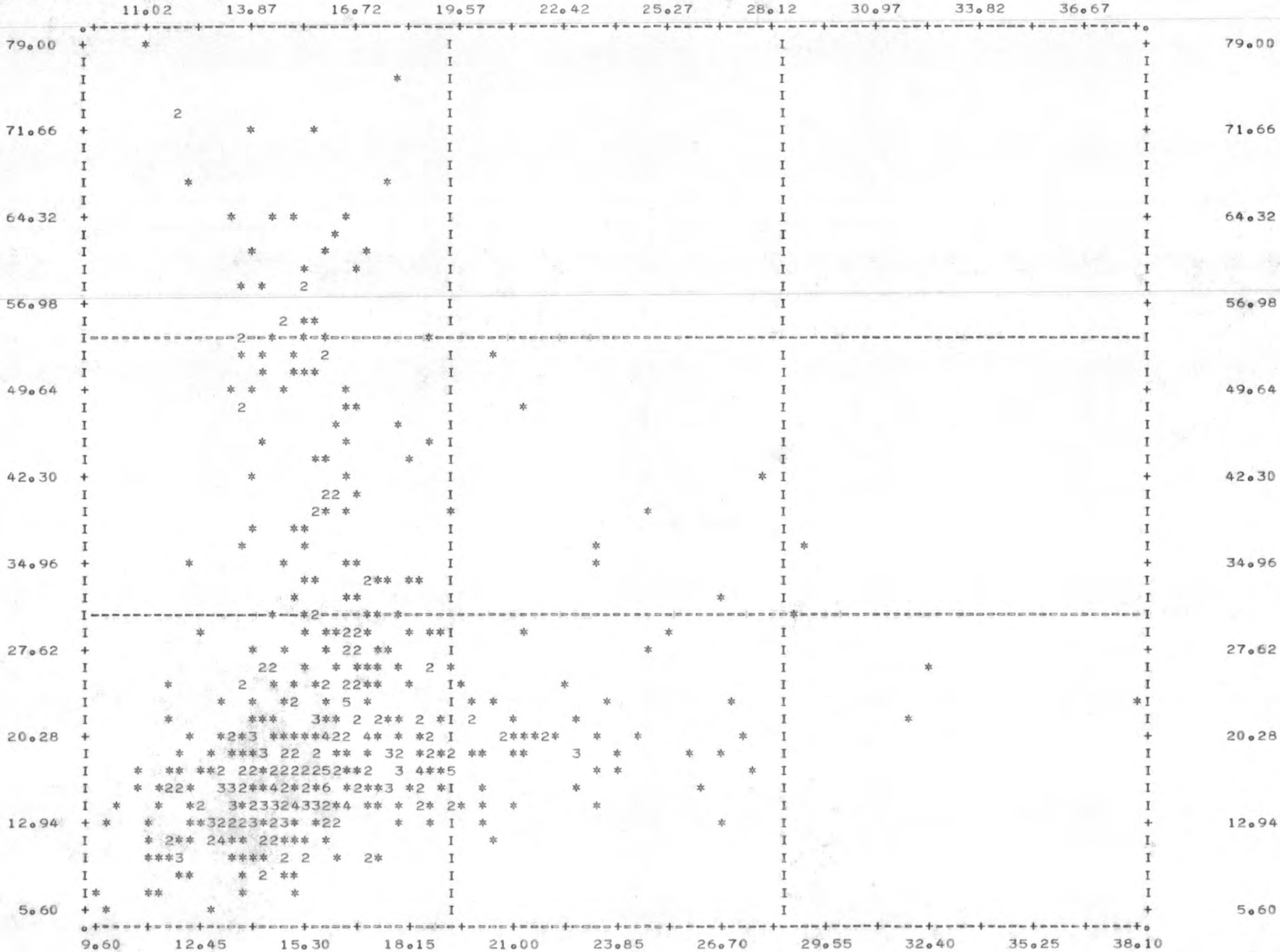
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) TDIST (ACROSS) TGOVT



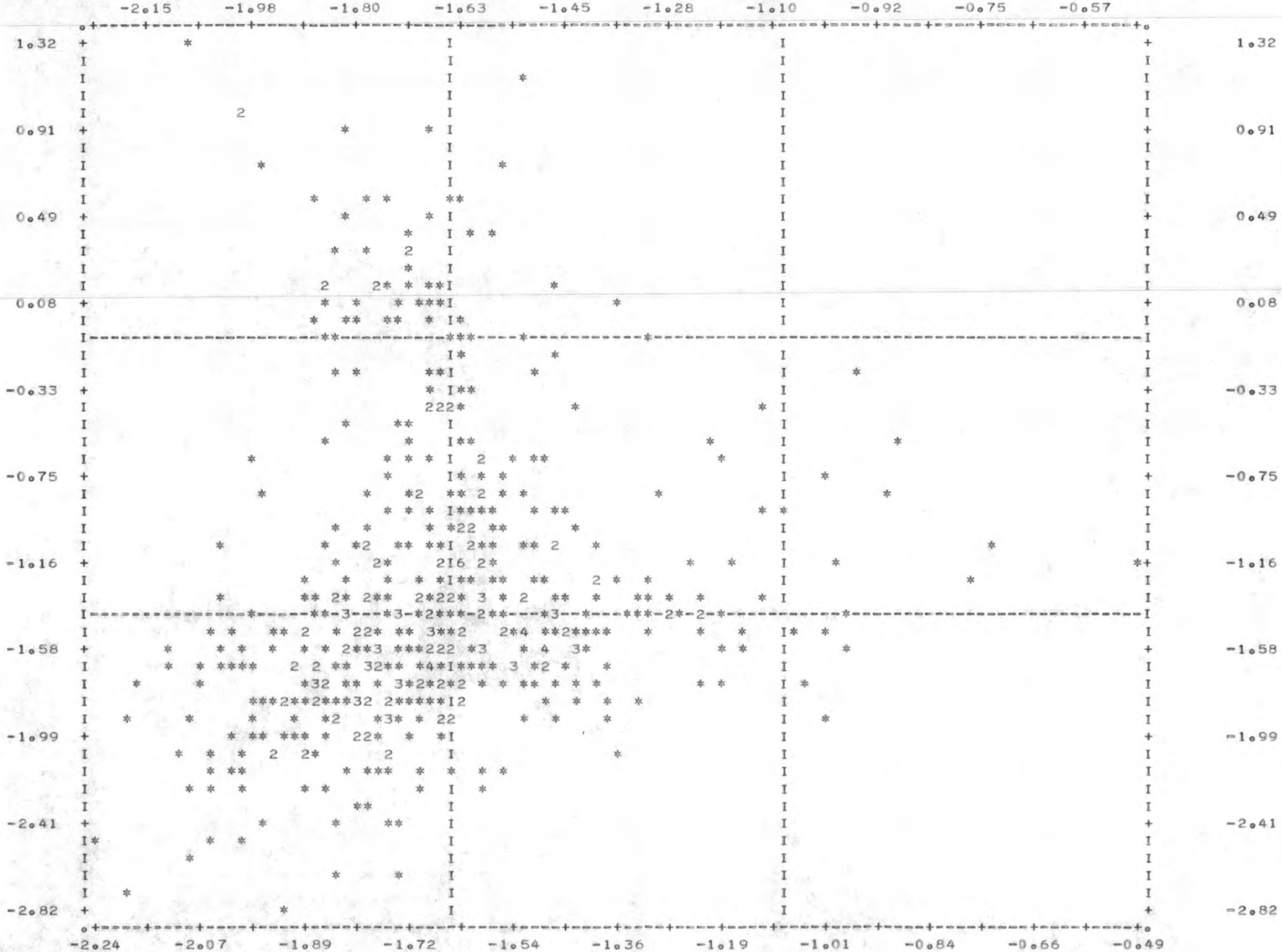
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) PRIV (ACROSS) OLD



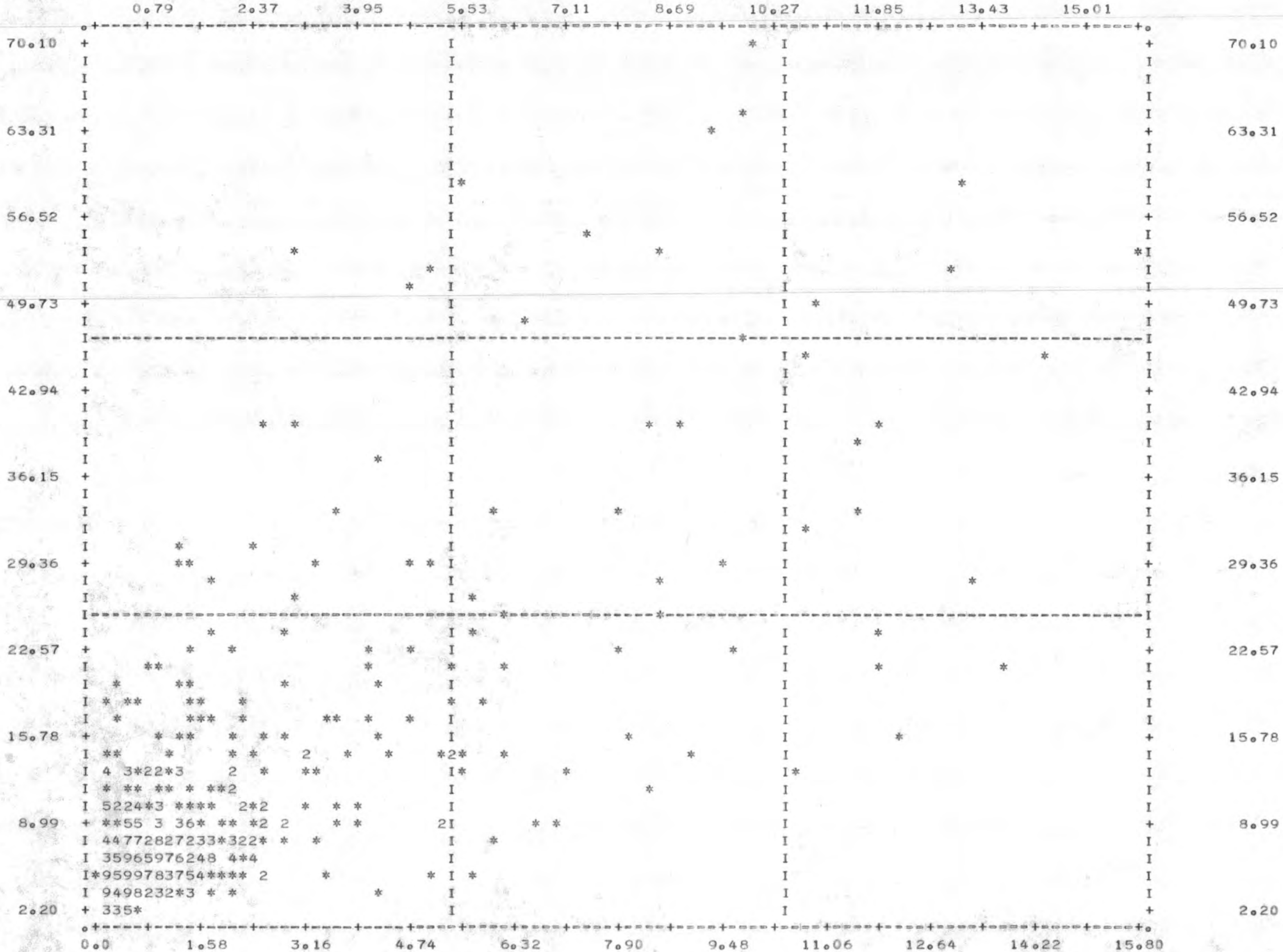
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) TPRIV (ACROSS) TOLD



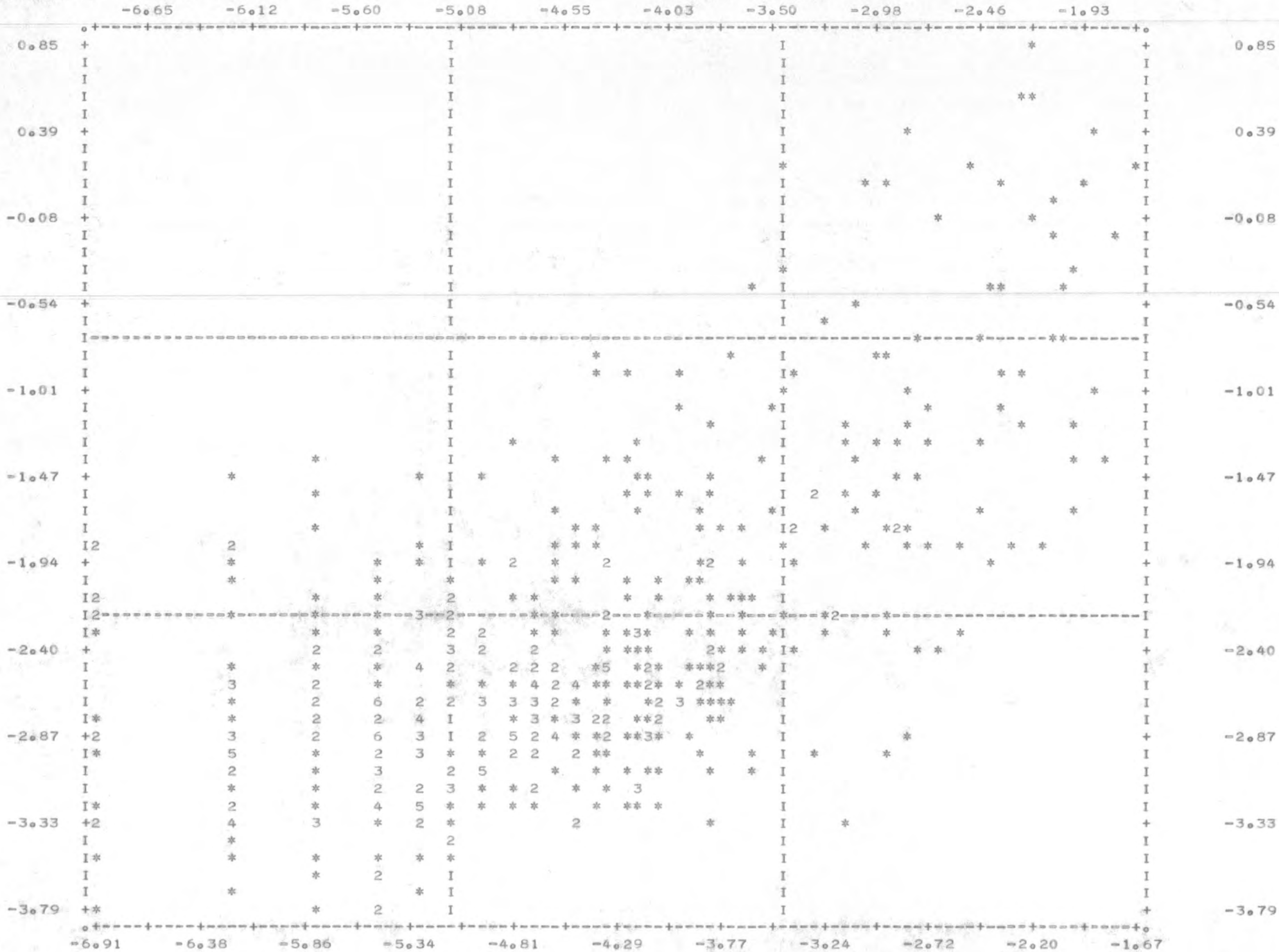
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) RM3 (ACROSS) NEWCOM



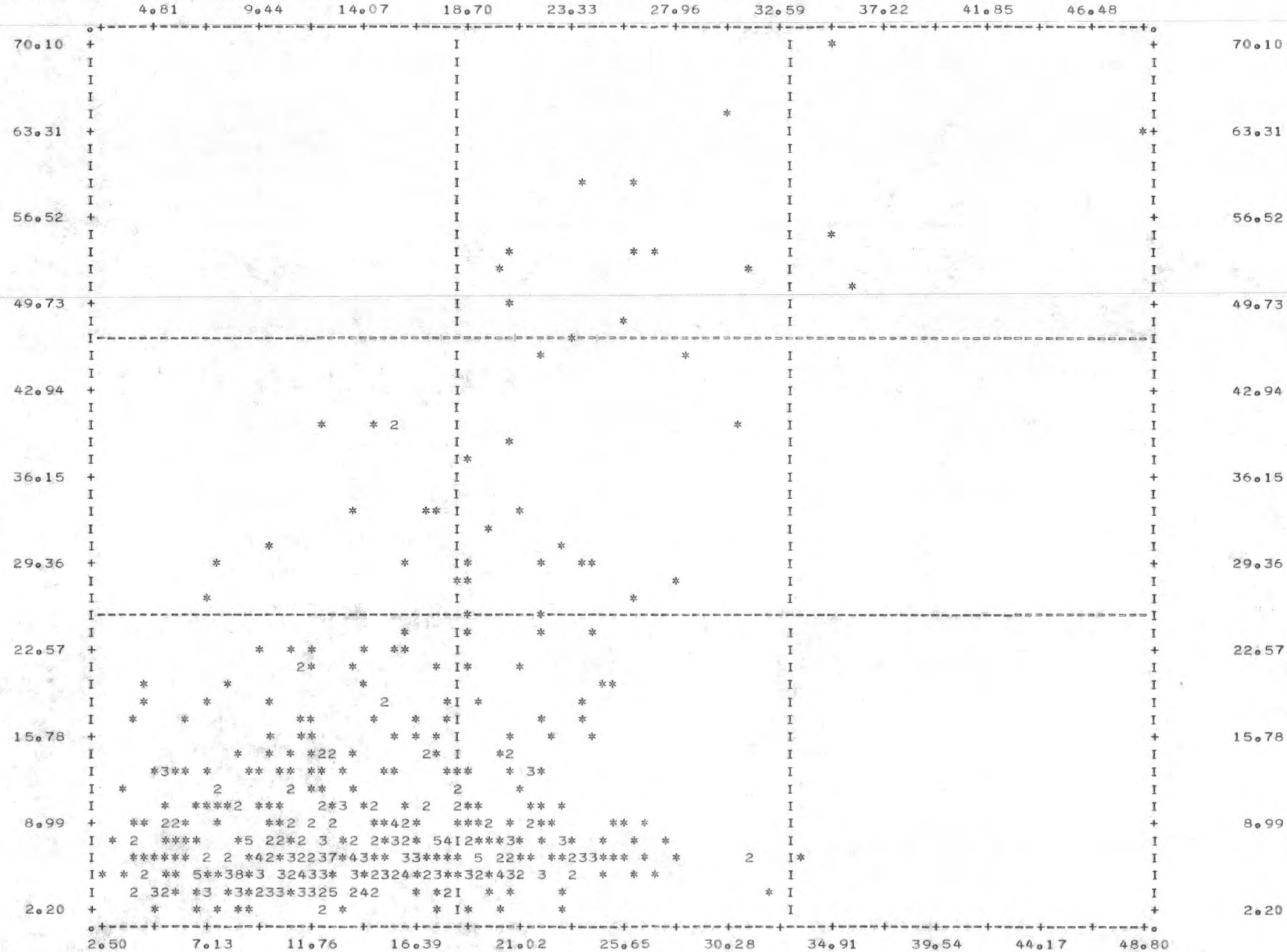
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) TRM3 (ACROSS) TNEWCOM

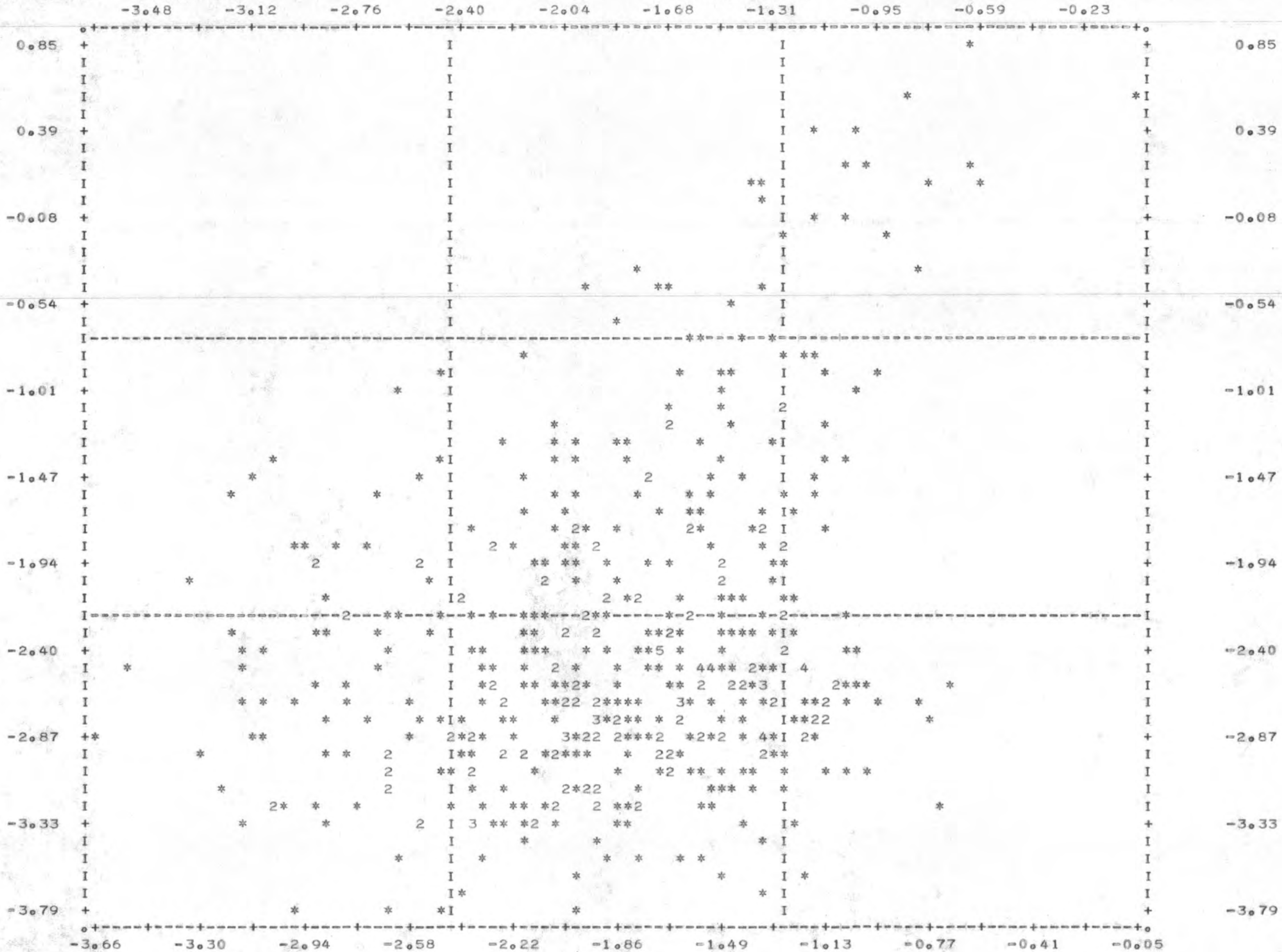


FILE TRANS (CREATION DATE = 02/25/80)

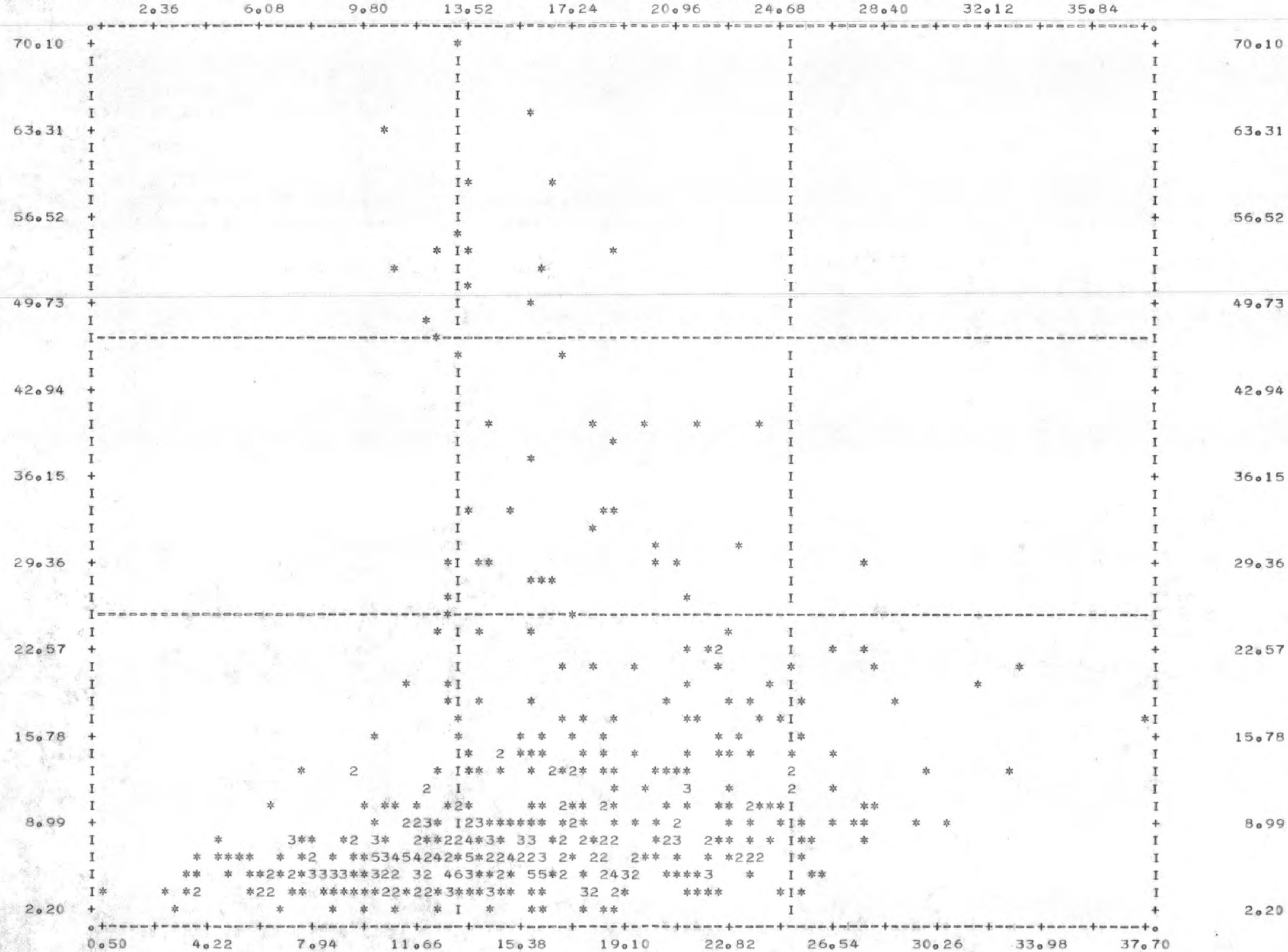
SCATTERGRAM OF (DOWN) RM3 (ACROSS) INMIG



FILE TRANS (CREATION DATE = 02/25/80)
SCATTERGRAM OF (DOWN) TRM3 (ACROSS) TINMIG

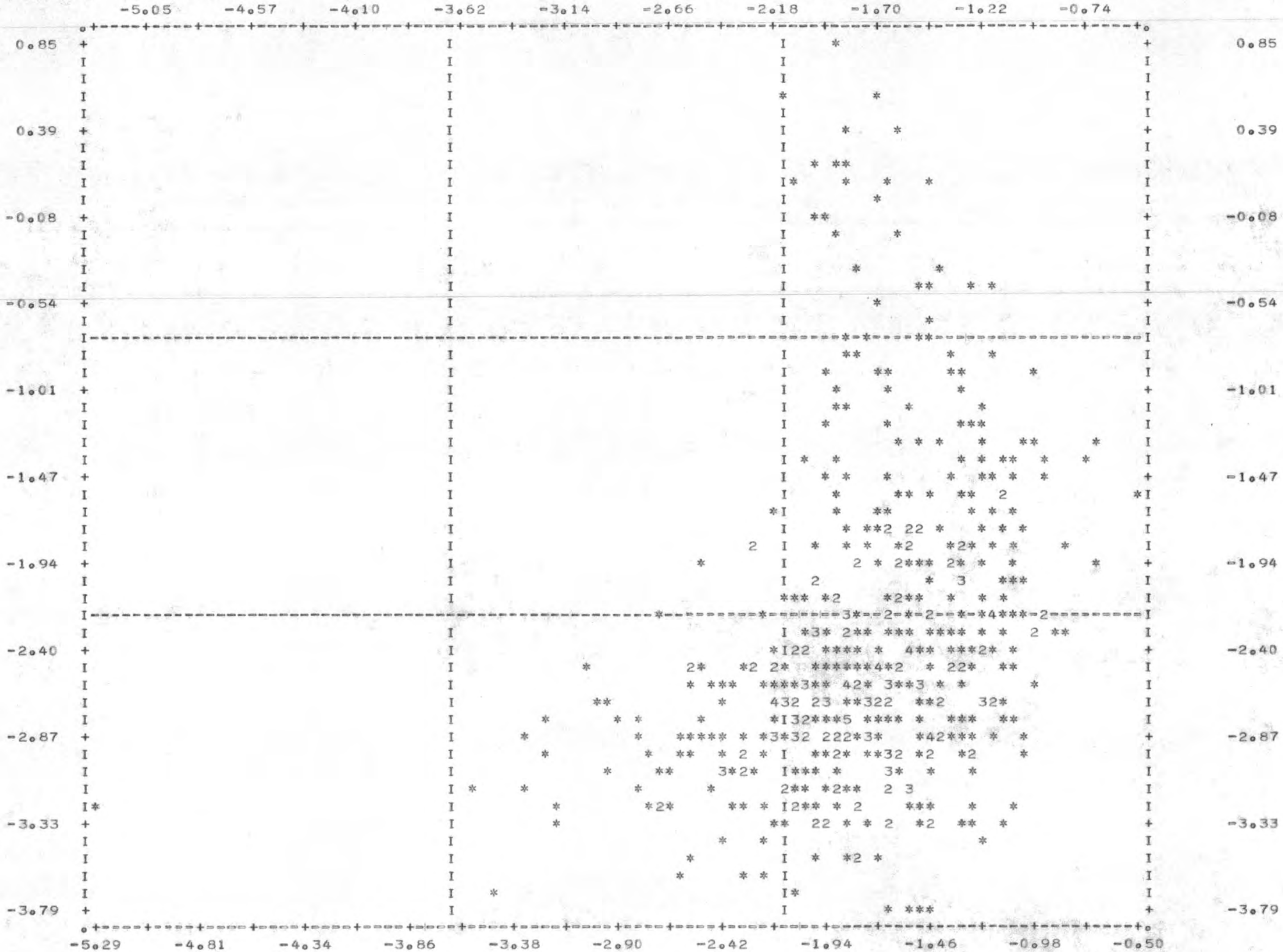


FILE TRANS (CREATION DATE = 02/25/80)
SCATTERGRAM OF (DOWN) RM3 (ACROSS) WITHMIG



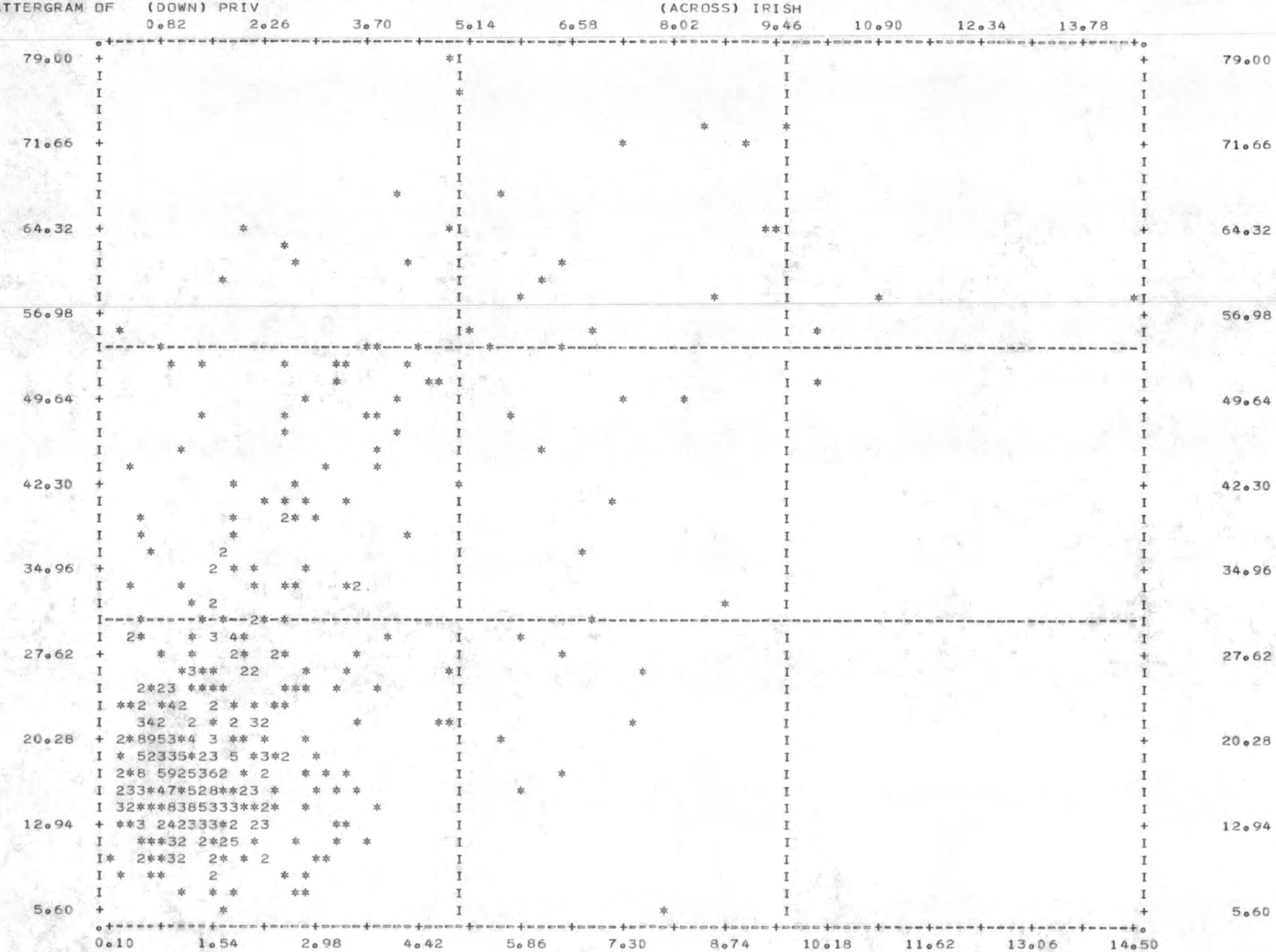
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) TRM3 (ACROSS) TWITHMIG



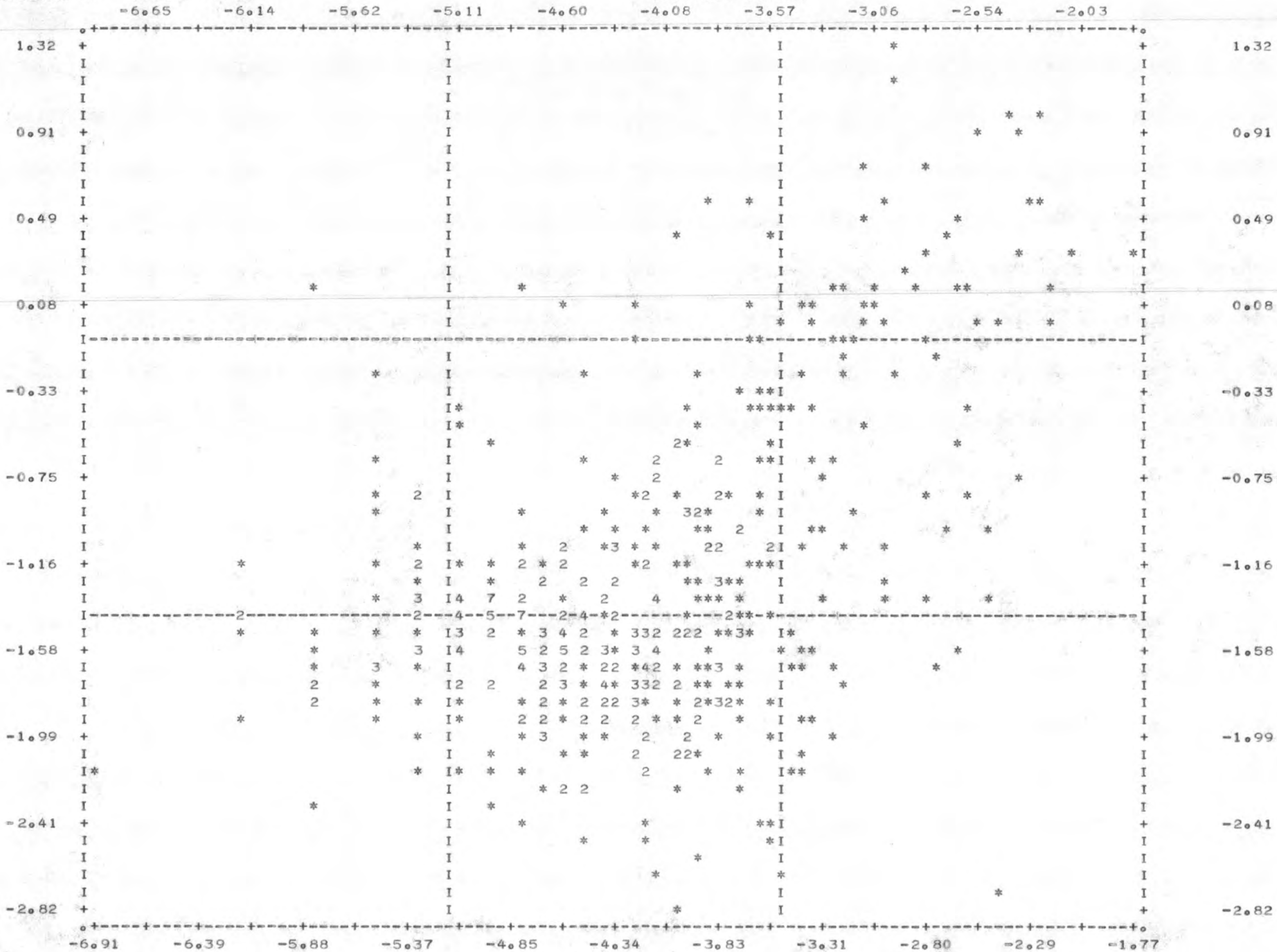
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) PRIV (ACROSS) IRISH



FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) TPRIV (ACROSS) TIRISH

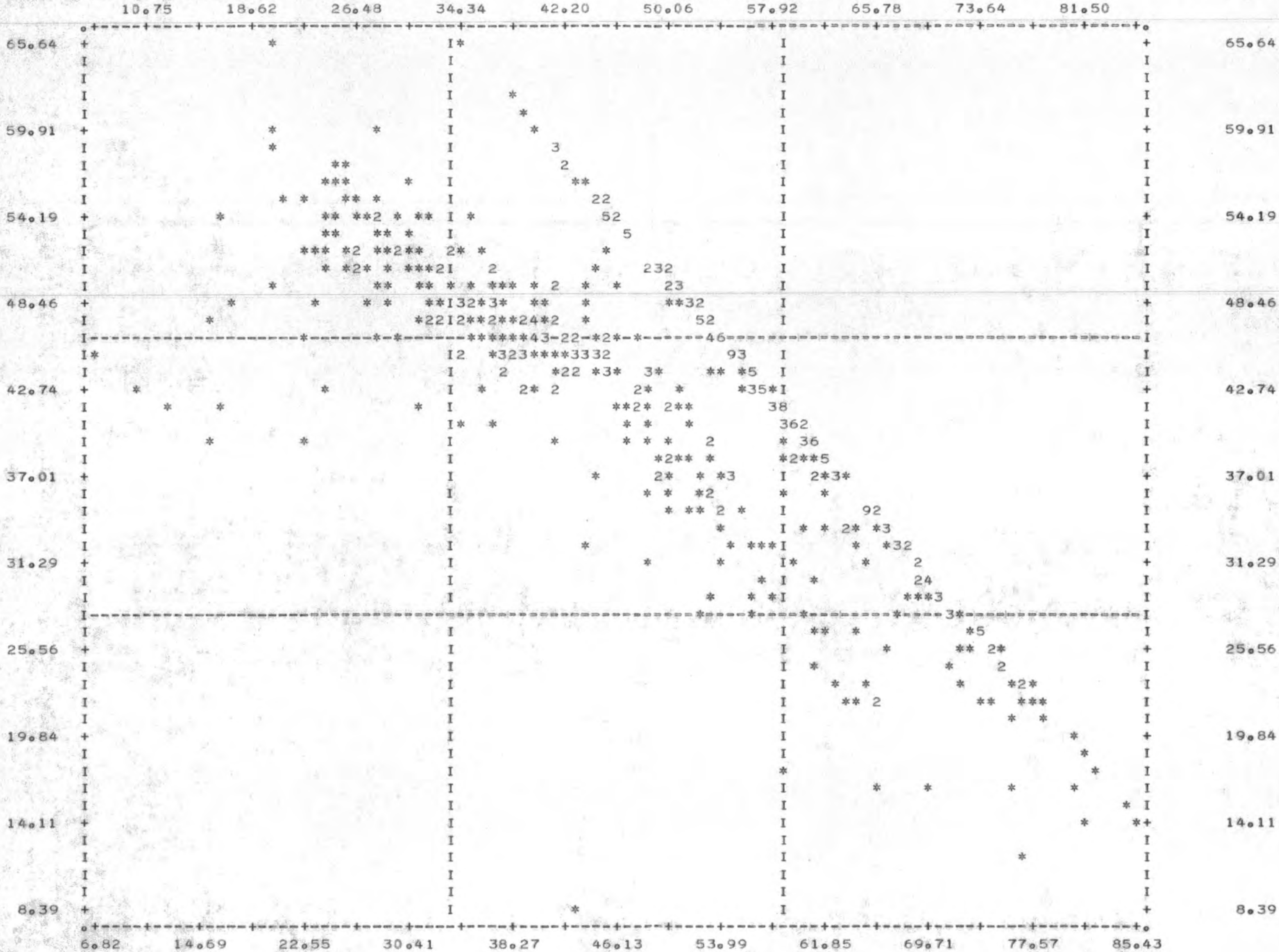


```
FILE    TRANS     (CREATION DATE = 02/25/80)
SCATTERGRAM OF    (DOWN) IRISH                                              (ACROSS) UNEMMA
              0.43      1.10      1.77      2.44      3.11      3.78      4.45      5.12      5.79      6.46
        o+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+-+o
   14.50 +            *                     I                               I                                  +   14.50
         I                                  I                               I                                  I
         I                                  I                               I                                  I
         I                                  I                               I                                  I
         I                                  I                               I                                  I
   13.06 +                                  I                               I                                  +   13.06
         I                                  I                               I                                  I
         I                                  I                               I                                  I
         I                                  I                               I                                  I
         I                                  I                               I                                  I
   11.62 +                                  I                               I                                  +   11.62
         I                                  I                               I                                  I
         I                                  I                               I                                  I
         I                *                 I                               I                                  I
         I                                  I                               I                                  I
   10.18 +                                 *I                               I                                  +   10.18
         I            *                     I                               I                                  I
         I---------------*------*------------------------------------------------------------------------------I
         I                *                 I                               I                                  I
         I                   *              I                               I                                  I
    8.74 +             *    *               I                               I                                  +    8.74
         I                           *      I                               I                                  I
         I                            *     I                               I                                  I
         I                     *            I                               I                                  I
         I      *           *               I                               I                                  I
    7.30 +      *      *                    I                               I                                  +    7.30
         I       *    *                     I                               I                                  I
         I               *      *           I                               I                                  I
         I         **    *                  I                               I*                                 I
         I       *              *           I                               I                                  I
    5.86 +       * **                       I                               I                                  +    5.86
         I       *  *  *       *            I                               I                                  I
         I                        *         I                               I                                  I
         I-------*-------**----------------------------------------------------------------------------------------I
         I*      * *   *                    I        *                      I                                  I
    4.42 +          * **    2               I                               I                                  +    4.42
         I             *  * *              *I                               I                                  I
         I      ** 2  *   *                 I                               I                                  I
         I   *2 *  2     *            *     I                               I                                  I
         I   2  22 ** *   * *   *           I                               I                                  I
    2.98 +   ** 4* **           *           I                               I                                 *+    2.98
         I    2 42 *2 3  3* *      *  *     I                               I                                  I
         I 2 22 43 23 ** *                  *                               I                                  I
         I * *7 72 25 32 2  2*  *           I     *                         I                                  I
         I   37 65 6* 34 3* ** 2* ** *      I                               I                                  I
    1.54 +   97 79 46 38 23  *  *  *  **  * I                 *             I                                  +    1.54
         I   *3 45 75 *  3*  *  *  ** *     I                               I                                  I
         I   43 97 95 87 5* 2* 42  *    *  *I                               I                                  I
         I    4 59 67 66 86 45 ** *  ** *   I                               I           *                      I
         I    2 2* 32 2* 42 32 2* 3  42  2  2     *                         I                                  I
    0.10 +                     *      *     I *  *2                         I                                  +    0.10
        o+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+-+o
         0.10      0.77      1.44      2.11      2.78      3.45      4.12      4.79      5.46      6.13      6.80
```

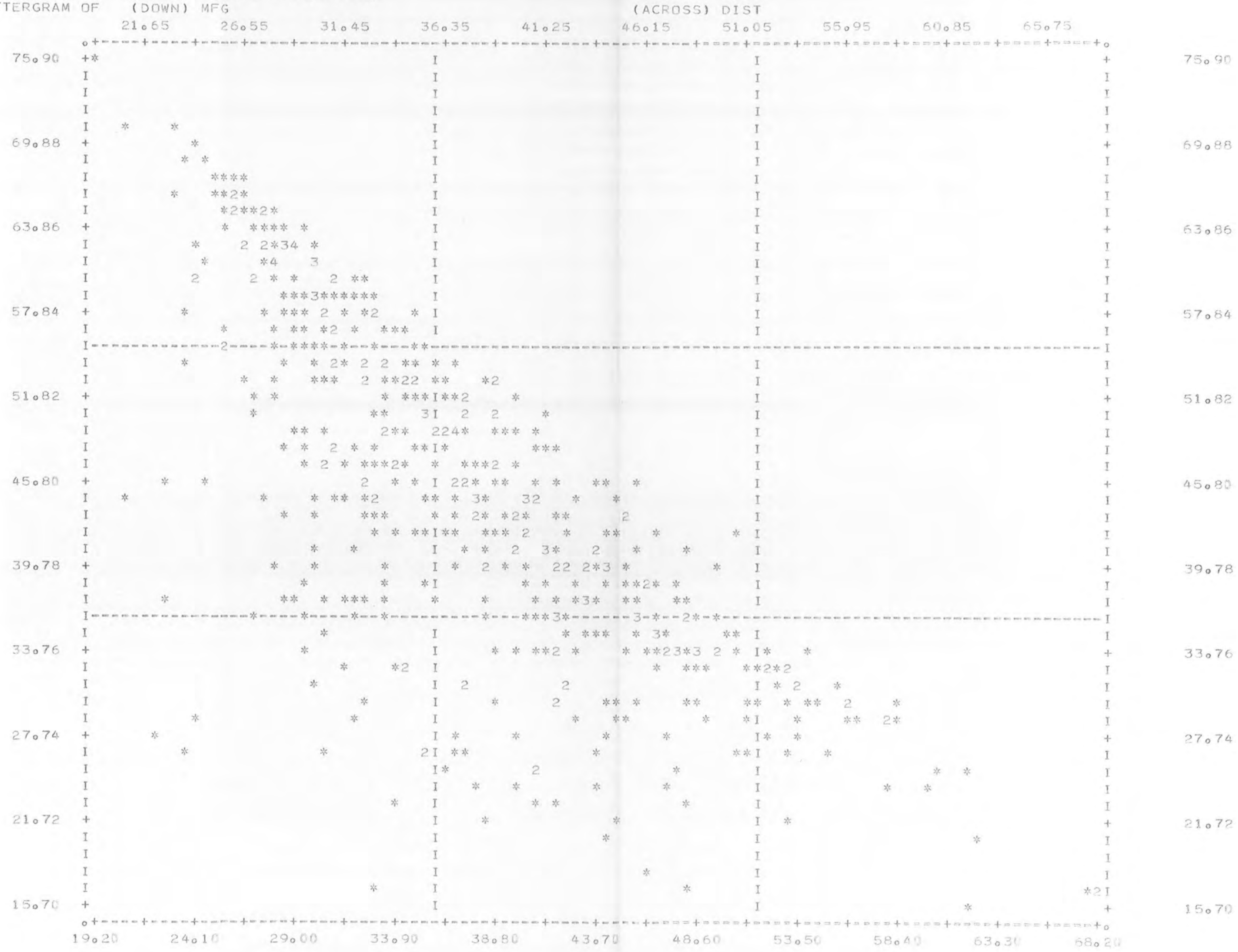
FILE TRANS (CREATION DATE = 02/25/80)
SCATTERGRAM OF (DOWN) TIRISH (ACROSS) TUNEMMA

```
             -6.69     -6.26     -5.83     -5.41     -4.98     -4.55     -4.12     -3.69     -3.26     -2.83
        o----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+-o
   -1.77 +                                  I                *              I                                  +   -1.77
         I                                  I                               I                                  I
         I                                  I                               I                                  I
         I                                  I                       *       I                                  I
         I                                  I                *              I       *                          I
   -2.29 +                                  I                     * *   *  *I                                  +   -2.29
         I                                  I                   *     *     I  **                              I
         I                                  I   2               *     *  *  I                                  I
         I                                  I       *        *    *        *I                                  I
         I                                  I       *   *  *      *        *I                       *          I
   -2.80 +                                  I       2   *  2    *        *  I                                  +   -2.80
         I                                  I                     * *       *                                  I
         I*                                 I       2   *       *     *     I             *                    I
         I                                  I              * *  2   * *     I       *                          I
         I                          *       I   *   *   2    *    * * *     I   *                              I
   -3.31 +                          *     2 I   2       2  * *      * *     I                                  +   -3.31
         I                          *       I   2   3   *  *               *I                                  I
         I--------------------------*-----2-----5---*---*----*----*--------*----------------------------------*I
         I                                * I   2   3   2  3 2  * 3 * *     I * *    *                         I
         I                2         2     2 I   4   *   *  2 *              I                                  I
   -3.83 +                *               6 I   5   *   *  2 3  2 2   *  * *I                                  +   -3.83
         I                          3     4 I   3   5   2  4 *  2     2    *I           *                      I
         I                          *     4 I   5   *   5    2  2 3 *   *2  *  **                              I
         I                          5     6 I   6   5   2  5 2  6 * *    *  *  **  *                           I
         I                          4     * I   *   5   2  * *  2 * 2   *   I                 *                I
   -4.34 +                                  I   4   2   4  *      2     **  I *                                +   -4.34
         I                          *     3 I       3   3  4 *    * *       *  *                               I
         I                          4     2 I   6   4   8  3 7  5 3   * *3 *I *  *  *                          I
         I                                  I                               I                                  I
         I                                * I   3   3   3  2 *  2 2 * *  * *I                                  I
   -4.85 +                                2 I   3   8   3  2 3  2 2 3 *     I                           *      +   -4.85
         I                                2 I       2   *  * 2  3 4 2 2 2   I   *                              I
         I                                  I   2   *   2  4 *  * 2 * * 3* **  * *                             I
         I-----------------------------------------------------------------------------------------------------I
         I                                * I   2       *  2 *  * 3 * 2 *  *2  *                               I
   -5.37 +                                  I                               I                                  +   -5.37
         I                                * I       *   2    *      *   *2  *  *     *  *                      I
         I                                  I                               I                                  I
         I                                  I                               I                                  I
         I                                  I                     *   *     I  22  2 *                         I
   -5.88 +                                  I                               I                                  +   -5.88
         I                                  I                               I                                  I
         I                                  I                               I                                  I
         I                                  I                               I   *     **2                      I
         I                                  I                               I                                  I
   -6.39 +                                  I                               I                                  +   -6.39
         I                                  I                               I                                  I
         I                                  I                               I                                  I
         I                                  I                               I                                  I
         I                                  I                               I                                  I
   -6.91 +                                  I                            *  I                                  +   -6.91
        o----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+-o
        -6.91     -6.48     -6.05     -5.62     -5.19     -4.76     -4.33     -3.90     -3.48     -3.05     -2.62
```

FILE TRANS (CREATION DATE = 02/25/80)
SCATTERGRAM OF (DOWN) PCC66 (ACROSS) PCLAB66



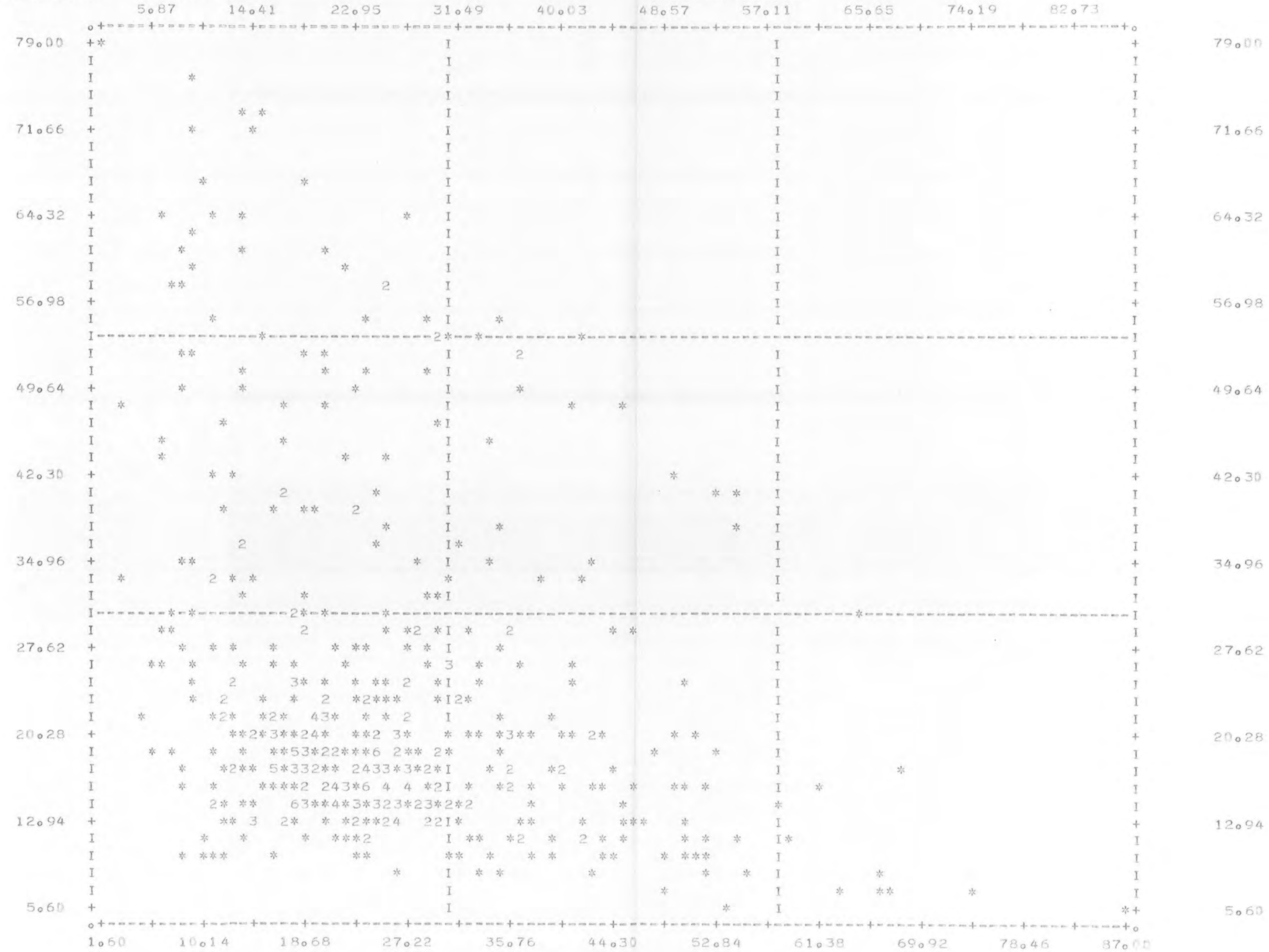
FILE TRANS (CREATION DATE = 02/25/80)
SCATTERGRAM OF (DOWN) MFG (ACROSS) DIST



SCATTERGRAMS OF UNTRANSFORMED VARIABLES 05/21/80 PAGE 7

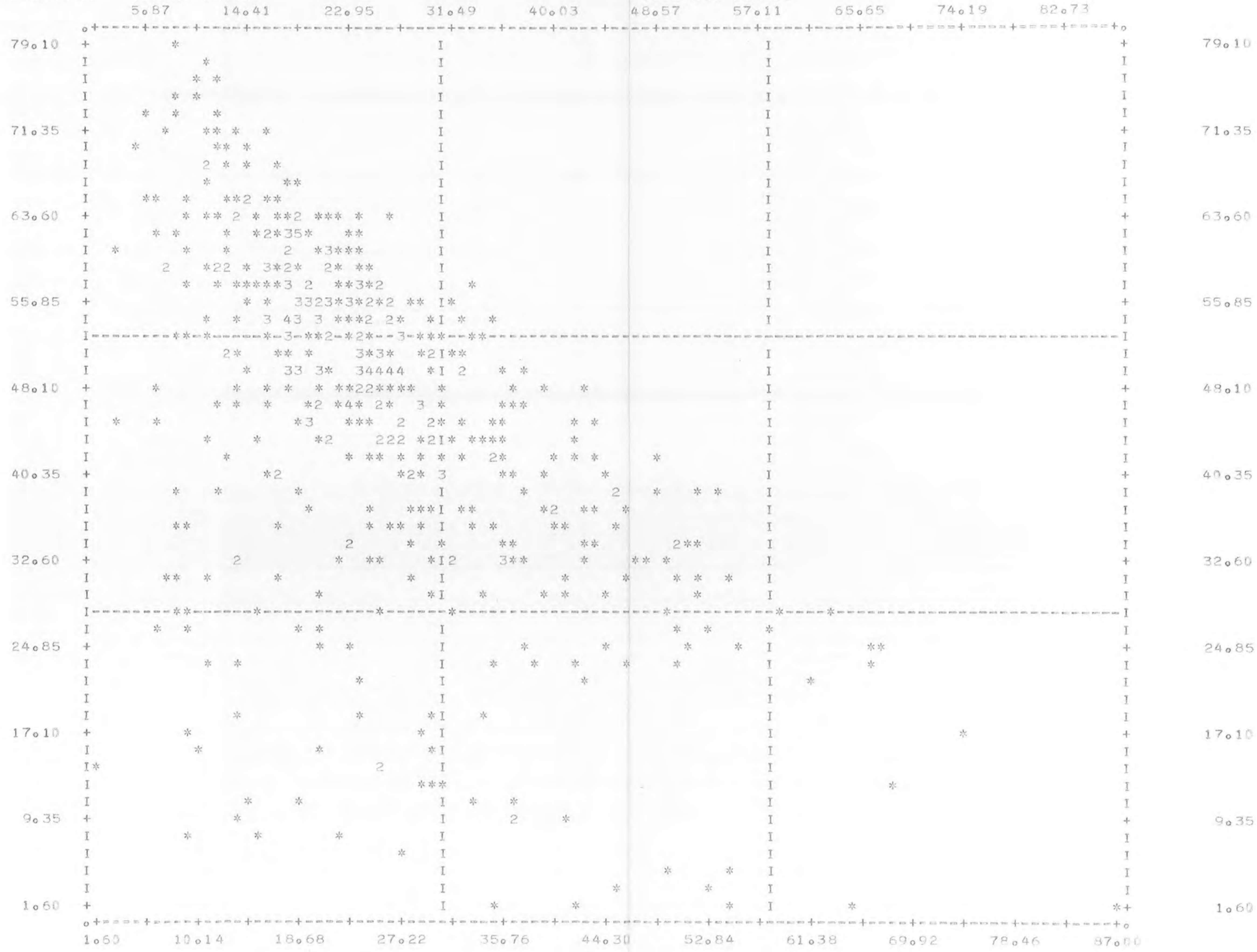
FILE TRANS (CREATION DATE = 02/25/80)

SCATTERGRAM OF (DOWN) PRIV (ACROSS) COUNCL



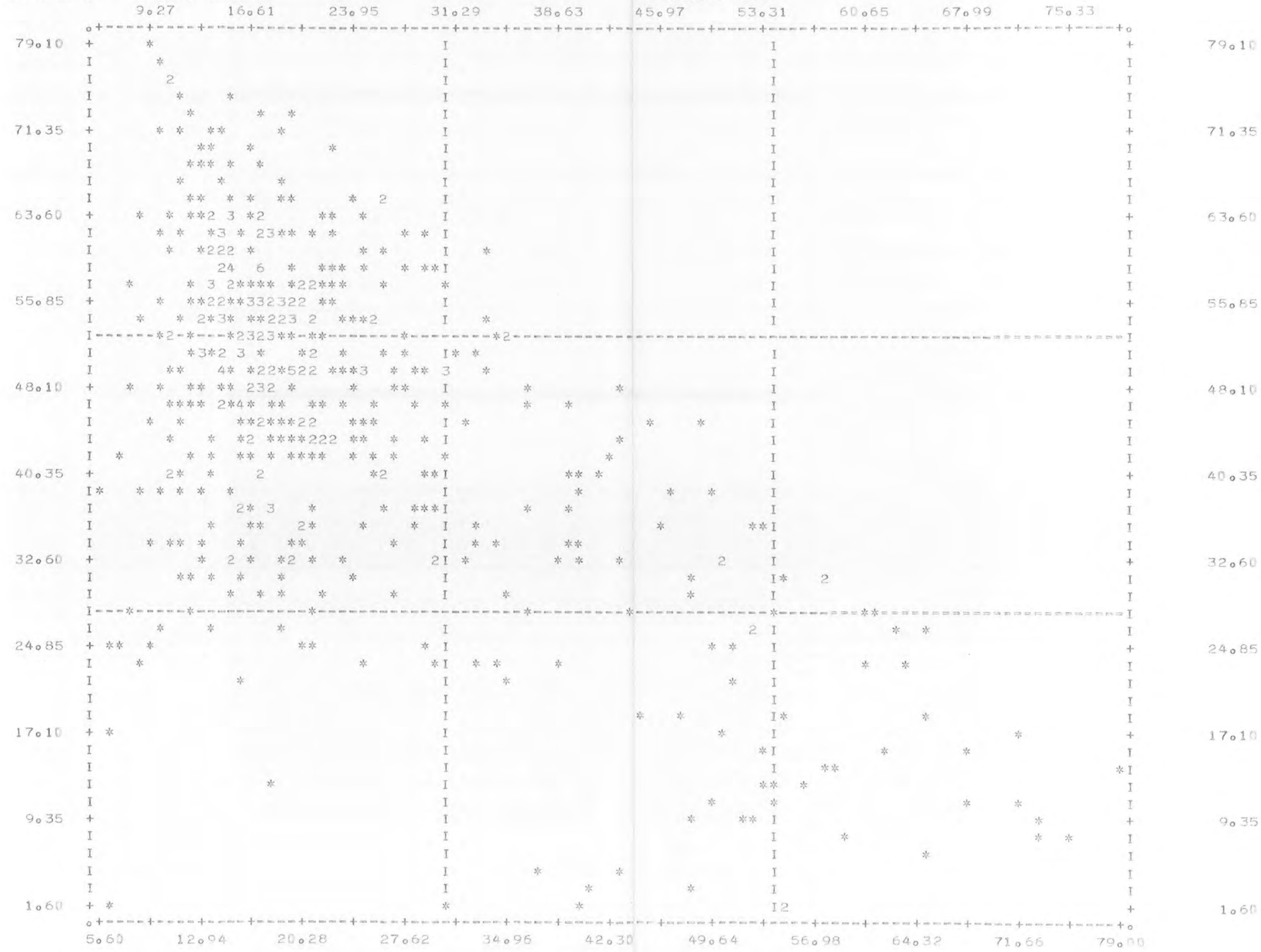
SCATTERGRAMS OF UNTRANSFORMED VARIABLES 05/21/80 PAGE 4

FILE` TRANS (CREATION DATE = 02/25/80)
SCATTERGRAM OF (DOWN) OWNOCC (ACROSS) COUNCL



SCATTERGRAMS OF UNTRANSFORMED VARIABLES 05/21/80 PAGE 2

FILE TRANS (CREATION DATE = 02/25/80)
SCATTERGRAM OF (DOWN) OWNOCC (ACROSS) PRIV



2A

PARAMETRIC CORRELATIONS ON LOGIT TRANSFORMED VARIABLES 05/13/80 PAGE 2

FILE TRANS (CREATION DATE = 02/25/80)

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TPROF	TEMPL	TNONM	TSKIL	TSEMI	TUNSK	TAGRI	TMIN	TMFG	TTRANS
TPROF	1.0000**	0.8230**	0.6860**	-0.6732**	-0.6572**	-0.7526**	0.0975	-0.3404**	-0.3247**	-0.1583**
TEMPL	0.8230**	1.0000**	0.5570**	-0.6962**	-0.5490**	-0.7720**	0.3924**	-0.3037**	-0.5258**	-0.1290*
TNONM	0.6860**	0.5570**	1.0000**	-0.6682**	-0.8133**	-0.4985**	-0.3049**	-0.5813**	-0.3599**	0.3747**
TSKIL	-0.6732**	-0.6962**	-0.6682**	1.0000**	0.4404**	0.4681**	-0.0715	0.3809**	0.6866**	-0.2088**
TSEMI	-0.6572**	-0.5490**	-0.8133**	0.4404**	1.0000**	0.4678**	0.3270**	0.5365**	0.1064	-0.2029**
TUNSK	-0.7526**	-0.7720**	-0.4985**	0.4681**	0.4678**	1.0000**	-0.1841**	0.1020	0.3050**	0.2844**
TAGRI	0.0975	0.3924**	-0.3049**	-0.0715	0.3270**	-0.1841**	1.0000**	0.1388*	-0.3809**	-0.3003**
TMIN	-0.3404**	-0.3037**	-0.5813**	0.3809**	0.5365**	0.1020	0.1388*	1.0000**	0.0474	-0.3287**
TMFG	-0.3247**	-0.5258**	-0.3599**	0.6866**	0.1064	0.3050**	-0.3809**	0.0474	1.0000**	-0.2384**
TTRANS	-0.1583**	-0.1290*	0.3747**	-0.2088**	-0.2029**	0.2844**	-0.3003**	-0.3287**	-0.2384**	1.0000**
TDIST	0.5850**	0.6316**	0.7451**	-0.8055**	-0.5624**	-0.3801**	-0.0102	-0.4611**	-0.7410**	0.3075**
TGOVT	0.3335**	0.4369**	0.3907**	-0.4483**	-0.2117**	-0.2317**	0.3540**	-0.2388**	-0.6880**	0.1110
TNOCAR	-0.6424**	-0.7295**	-0.2468**	0.3155**	0.2596**	0.7136**	-0.5970**	0.1634**	0.2428**	0.2893**
TOWNOCC	0.4628**	0.5498**	0.1826**	-0.0621	-0.2961**	-0.5699**	0.3487**	-0.1070	0.0478	-0.2933**
TCOUNCL	-0.4714**	-0.5483**	-0.4545**	0.4808**	0.4372**	0.4376**	-0.0895	0.3000**	0.3688**	-0.0264
TPRIV	-0.1587**	-0.2190**	0.1883**	-0.2007**	-0.0800	0.3222**	-0.4328**	-0.1571*	-0.2051**	0.3903**
TAMEN	0.5840**	0.6129**	0.3015**	-0.2776**	-0.3351**	-0.6578**	0.3066**	-0.0591	-0.1420*	-0.1878**
THDENS	-0.5292**	-0.6930**	-0.2882**	0.2513**	0.3853**	0.6559**	-0.4456**	0.1254	0.2209**	0.2302**
TRM3	0.0050	-0.0999	0.3613**	-0.3028**	-0.2021**	0.1817**	-0.5072**	-0.2564**	-0.2416**	0.3494**
TRM7	0.4812**	0.6572**	0.3511**	-0.5120**	-0.3047**	-0.3783**	0.3971**	-0.3556**	-0.4955**	0.0521
TIRISH	0.1790**	-0.0291	0.4621**	-0.2548**	-0.3347**	0.0111	-0.4920**	-0.5819**	0.0468	0.2968**
TNEWCOM	0.1322*	-0.0122	0.4035**	-0.2350**	-0.3045**	0.0431	-0.4243**	-0.5606**	-0.0325	0.2511**
TYOUNG	-0.2146**	-0.4741**	-0.1315*	0.1435*	0.1911**	0.3379**	-0.3876**	0.0289	0.2528**	0.1059
TOLD	0.0886	0.3624**	0.1247*	-0.1847**	-0.1296*	-0.1156*	0.2761**	-0.0909	-0.4203**	-0.0216
TUNEMMA	-0.4888**	-0.4204**	-0.2809**	0.1212*	0.3260**	0.5471**	-0.1302*	0.2686**	-0.1194*	0.2130**
TINMIG	0.6359**	0.5289**	0.6460**	-0.5340**	-0.5422**	-0.4730**	0.0025	-0.3095**	-0.3355**	0.0724
TWITHMIG	-0.3530**	-0.5049**	-0.0546	0.3376**	-0.0291	0.3791**	-0.6366**	0.0423	0.4049**	0.1704**

* - SIGNIF. LE .01 ** - SIGNIF. LE .001 (99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

2B

PARAMETRIC CORRELATIONS ON LOGIT TRANSFORMED VARIABLES 05/13/80 PAGE 3

FILE TRANS (CREATION DATE = 02/25/80)

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TDIST	TGOVT	TNOCAR	TOWNOCC	TCOUNCL	TPRIV	TAMEN	THDENS	TRM3	TRM7
TPROF	0.5850**	0.3335**	-0.6424**	0.4628**	-0.4714**	-0.1587**	0.5840**	-0.5292**	0.0050	0.4812**
TEMPL	0.6316**	0.4369**	-0.7295**	0.5498**	-0.5483**	-0.2190**	0.6129**	-0.6930**	-0.0999	0.6572**
TNONM	0.7451**	0.3907**	-0.2468**	0.1826**	-0.4545**	0.1883**	0.3015**	-0.2882**	0.3613**	0.3511**
TSKIL	-0.8055**	-0.4483**	0.3155**	-0.0621	0.4808**	-0.2007**	-0.2776**	0.2513**	-0.3028**	-0.5120**
TSEMI	-0.5624**	-0.2117**	0.2596**	-0.2961**	0.4372**	-0.0800	-0.3351**	0.3853**	-0.2021**	-0.3047**
TUNSK	-0.3801**	-0.2317**	0.7136**	-0.5699**	0.4376**	0.3222**	-0.6578**	0.6559**	0.1817**	-0.3783**
TAGRI	-0.0102	0.3540**	-0.5970**	0.3487**	-0.0895	-0.4328**	0.3066**	-0.4456**	-0.5072**	0.3971**
TMIN	-0.4611**	-0.2388**	0.1634**	-0.1070	0.3000**	-0.1571*	-0.0591	0.1254	-0.2564**	-0.3556**
TMFG	-0.7410**	-0.6880**	0.2428**	0.0478	0.3688**	-0.2051**	-0.1420*	0.2209**	-0.2416**	-0.4955**
TTRANS	0.3075**	0.1110	0.2893**	-0.2933**	-0.0264	0.3903**	-0.1878**	0.2302**	0.3494**	0.0521
TDIST	1.0000**	0.4519**	-0.1658**	0.0015	-0.5068**	0.3464**	0.1547**	-0.2302**	0.4704**	0.5246**
TGOVT	0.4519**	1.0000**	-0.3820**	0.1077	-0.2550**	-0.0148	0.2420**	-0.2817**	0.0790	0.4428**
TNOCAR	-0.1658**	-0.3820**	1.0000**	-0.6477**	0.2255**	0.6182**	-0.7512**	0.7016**	0.5041**	-0.5199**
TOWNOCC	0.0015	0.1077	-0.6477**	1.0000**	-0.4774**	-0.5613**	0.5630**	-0.7603**	-0.5977**	0.4074**
TCOUNCL	-0.5068**	-0.2550**	0.2255**	-0.4774**	1.0000**	-0.3489**	0.0247	0.4173**	-0.1347*	-0.5093**
TPRIV	0.3464**	-0.0148	0.6182**	-0.5613**	-0.3489**	1.0000**	-0.7318**	0.4764**	0.7431**	-0.0244
TAMEN	0.1547**	0.2420**	-0.7512**	0.5630**	0.0247	-0.7318**	1.0000**	-0.6060**	-0.4705**	0.2581**
THDENS	-0.2302**	-0.2817**	0.7016**	-0.7603**	0.4173**	0.4764**	-0.6060**	1.0000**	0.5363**	-0.5013**
TRM3	0.4704**	0.0790	0.5041**	-0.5977**	-0.1347*	0.7431**	-0.4705**	0.5363**	1.0000**	-0.1023
TRM7	0.5246**	0.4428**	-0.5199**	0.4074**	-0.5093**	-0.0244	0.2581**	-0.5013**	-0.1023	1.0000**
TIRISH	0.3195**	-0.0002	0.1870**	-0.2568**	-0.2235**	0.4749**	-0.2444**	0.3006**	0.5331**	0.0765
TNEWCOM	0.3330**	0.1199*	0.1736**	-0.2524**	-0.2675**	0.5304**	-0.3475**	0.2963**	0.6344**	0.0996
TYOUNG	-0.2508**	-0.1508**	0.2992**	-0.4477**	0.3624**	0.1911**	-0.2680**	0.5305**	0.1602**	-0.2879**
TOLD	0.4416**	0.2524**	-0.0165	0.2522**	-0.4344**	0.1322*	-0.0495	-0.4133**	0.0833	0.4246**
TUNEMMA	0.0247	-0.1398*	0.6501**	-0.4063**	0.1225*	0.3776**	-0.4711**	0.4285**	0.2664**	-0.1747**
TINMIG	0.5413**	0.3502**	-0.3410**	0.1769**	-0.4524**	0.1369*	0.2667**	-0.2258**	0.2999**	0.2770**
TWITHMIG	-0.1715**	-0.3919**	0.6156**	-0.2656**	0.2542**	0.2231**	-0.3207**	0.3790**	0.3006**	-0.3972**

* - SIGNIF. LE .01 ** - SIGNIF. LE .001 (99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

2C

PARAMETRIC CORRELATIONS ON LOGIT TRANSFORMED VARIABLES 05/13/80 PAGE 4

FILE TRANS (CREATION DATE = 02/25/80)

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TIRISH	TNEWCOM	TYOUNG	TOLD	TUNEMMA	TINMIG	TWITHMIG
TPROF	0.1790**	0.1322*	-0.2146**	0.0886	-0.4888**	0.6359**	-0.3530**
TEMPL	-0.0291	-0.0122	-0.4741**	0.3624**	-0.4204**	0.5289**	-0.5049**
TNONM	0.4621**	0.4035**	-0.1315*	0.1247*	-0.2809**	0.6460**	-0.0546
TSKIL	-0.2548**	-0.2350**	0.1435*	-0.1847**	0.1212*	-0.5340**	0.3376**
TSEMI	-0.3347**	-0.3045**	0.1911**	-0.1296*	0.3260**	-0.5422**	-0.0291
TUNSK	0.0111	0.0431	0.3379**	-0.1156*	0.5471**	-0.4730**	0.3791**
TAGRI	-0.4920**	-0.4243**	-0.3876**	0.2761**	-0.1302*	0.0025	-0.6366**
TMIN	-0.5819**	-0.5606**	0.0289	-0.0909	0.2686**	-0.3095**	0.0423
TMFG	0.0468	-0.0325	0.2528**	-0.4203**	-0.1194*	-0.3355**	0.4049**
TTRANS	0.2968**	0.2511**	0.1059	-0.0216	0.2130**	0.0724	0.1704**
TDIST	0.3195**	0.3330**	-0.2608**	0.4416**	0.0247	0.5413**	-0.1715**
TGOVT	-0.0002	0.1199*	-0.1508**	0.2524**	-0.1398*	0.3502**	-0.3919**
TNOCAR	0.1870**	0.1736**	0.2992**	-0.0165	0.6501**	-0.3410**	0.6156**
TOWNOCC	-0.2568**	-0.2524**	-0.4477**	0.2522**	-0.4063**	0.1769**	-0.2656**
TCOUNCL	-0.2235**	-0.2675**	0.3624**	-0.4344**	0.1225*	-0.4524**	0.2542**
TPRIV	0.4749**	0.5304**	0.1911**	0.1322*	0.3776**	0.1369*	0.2231**
TAMEN	-0.2444**	-0.3475**	-0.2680**	-0.0495	-0.4711**	0.2667**	-0.3207**
THDENS	0.3006**	0.2963**	0.5305**	-0.4133**	0.4285**	-0.2258**	0.3790**
TRM3	0.5331**	0.6344**	0.1602**	0.0833	0.2664**	0.2999**	0.3006**
TRM7	0.0765	0.0996	-0.2879**	0.4246**	-0.1747**	0.2770**	-0.3972**
TIRISH	1.0000**	0.7499**	0.2145**	-0.1289*	-0.1527**	0.3605**	0.1560**
TNEWCOM	0.7499**	1.0000**	0.1360*	0.0081	-0.1318*	0.3824**	0.1564**
TYOUNG	0.2145**	0.1360*	1.0000**	-0.6689**	0.0278	-0.2189**	0.2306**
TOLD	-0.1289*	0.0081	-0.6689**	1.0000**	0.2339**	0.0687	-0.0953
TUNEMMA	-0.1527**	-0.1318*	0.0278	0.2339**	1.0000**	-0.3138**	0.3309**
TINMIG	0.3605**	0.3824**	-0.2189**	0.0687	-0.3138**	1.0000**	-0.1853**
TWITHMIG	0.1560**	0.1564**	0.2306**	-0.0953	0.3309**	-0.1853**	1.0000**

* - SIGNIF. LE .01 ** - SIGNIF. LE .001 (99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

1A

PARAMETRIC CORRELATION ON UNTRANSFORMED VARIABLES 05/13/80 PAGE 2

FILE TRANS (CREATION DATE = 02/25/80)

- - - - - - - - - - - - - P E A R S O N   C O R R E L A T I O N   C O E F F I C I E N T S - - - - - - - - - - - - -

	PROF	EMPL	NONM	SKIL	SEMI	UNSK	AGRI	MIN	MFG	TRANS
PROF	1.0000**	0.7954**	0.6962**	-0.7560**	-0.6368**	-0.6419**	-0.0666	-0.2382**	-0.3498**	-0.1472**
EMPL	0.7954**	1.0000**	0.5588**	-0.7578**	-0.5266**	-0.7107**	0.3153**	-0.2591**	-0.5470**	-0.1684**
NONM	0.6962**	0.5588**	1.0000**	-0.6974**	-0.7824**	-0.4559**	-0.3265**	-0.3878**	-0.3919**	0.2878**
SKIL	-0.7560**	-0.7578**	-0.6974**	1.0000**	0.3680**	0.3658**	-0.1421*	0.2232**	0.6971**	-0.1771**
SEMI	-0.6368**	-0.5266**	-0.7824**	0.3680**	1.0000**	0.3249**	0.3947**	0.5662**	0.0575	-0.1562**
UNSK	-0.6419**	-0.7107**	-0.4559**	0.3658**	0.3249**	1.0000**	-0.1826**	-0.0596	0.2822**	0.4186**
AGRI	-0.0666	0.3153**	-0.3265**	-0.1421*	0.3947**	-0.1826**	1.0000**	-0.0169	-0.4461**	-0.2390**
MIN	-0.2382**	-0.2591**	-0.3878**	0.2232**	0.5662**	-0.0596	-0.0169	1.0000**	-0.1052	-0.2215**
MFG	-0.3498**	-0.5470**	-0.3919**	0.6971**	0.0575	0.2822**	-0.4461**	-0.1052	1.0000**	-0.1924**
TRANS	-0.1472**	-0.1684**	0.2878**	-0.1771**	-0.1562**	0.4186**	-0.2390**	-0.2215**	-0.1924**	1.0000**
DIST	0.6268**	0.6701**	0.7458**	-0.8077**	-0.5356**	-0.3293**	0.0359	-0.3331**	-0.7403**	0.2050**
GOVT	0.1725**	0.2634**	0.2485**	-0.2982**	-0.1368*	-0.1429*	0.2676**	-0.1412*	-0.5501**	-0.0255
NOCAR	-0.5441**	-0.6780**	-0.2248**	0.3466**	0.2315**	0.7119**	-0.4638**	0.1484**	0.2564**	0.3045**
OWNOCC	0.3570**	0.5365**	0.2113**	-0.1290*	-0.3367**	-0.6076**	0.1689**	-0.1535**	0.0212	-0.3105**
COUNCL	-0.4245**	-0.5344**	-0.4103**	0.4312**	0.3552**	0.3994**	-0.1670**	0.2307**	0.3532**	0.0466
PRIV	-0.0261	-0.1754**	0.2345**	-0.2020**	-0.0953	0.3781**	-0.2630**	-0.1378*	-0.1948**	0.3994**
AMEN	0.4689**	0.5711**	0.2381**	-0.2539**	-0.2614**	-0.6521**	0.1833**	0.0055	-0.1512**	-0.2692**
HDENS	-0.3862**	-0.5790**	-0.2041**	0.1729**	0.2831**	0.6628**	-0.2976**	0.0718	0.1581**	0.3126**
RM3	0.1198*	-0.0558	0.3463**	-0.3002**	-0.1640**	0.2131**	-0.2872**	-0.1369*	-0.2525**	0.2918**
RM7	0.4274**	0.6497**	0.3056**	-0.5338**	-0.2583**	-0.3567**	0.4678**	-0.3028**	-0.5392**	-0.0047
IRISH	0.1179*	-0.1159*	0.3102**	-0.1843**	-0.1652**	0.1511**	-0.3235**	-0.2584**	0.0192	0.2059**
NEWCOM	0.0299	-0.1507**	0.2551**	-0.1254*	-0.1215*	0.1816**	-0.2760**	-0.2187**	-0.0064	0.2322**
YOUNG	-0.1026	-0.4516**	-0.1032	0.1436*	0.1687**	0.3505**	-0.3003**	0.0605	0.2300**	0.1530**
OLD	0.0319	0.3737**	0.1217*	-0.2040**	-0.1346*	-0.1674**	0.3032**	-0.1507**	-0.4182**	-0.0661
UNEMMA	-0.3663**	-0.3467**	-0.2606**	0.0844	0.2833**	0.5730**	-0.0611	0.1809**	-0.0763	0.2151**
INMIG	0.6572**	0.5459**	0.6649**	-0.5924**	-0.5380**	-0.4379**	-0.1478**	-0.2203**	-0.3648**	0.0288
WITHMIG	-0.3600**	-0.5210**	-0.1422*	0.3896**	0.0205	0.4636**	-0.5623**	0.0042	0.4316**	0.1709**

* - SIGNIF. LE .01 ** - SIGNIF. LE .001 (99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

1B

PARAMETRIC CORRELATION ON UNTRANSFORMED VARIABLES 05/13/80 PAGE 3

FILE TRANS (CREATION DATE = 02/25/80)

- - - - - - - - - - - - - P E A R S O N   C O R R E L A T I O N   C O E F F I C I E N T S - - - - - - - - - - - - -

	DIST	GOVT	NOCAR	OWNOCC	COUNCL	PRIV	AMEN	HDENS	RM3	RM7
PROF	0.6268**	0.1725**	-0.5441**	0.3670**	-0.4245**	-0.0261	0.4689**	-0.3862**	0.1198*	0.4274**
EMPL	0.6701**	0.2634**	-0.6780**	0.5365**	-0.5344**	-0.1754**	0.5711**	-0.5790**	-0.0558	0.6497**
NONM	0.7458**	0.2485**	-0.2248**	0.2113**	-0.4103**	0.2345**	0.2381**	-0.2041**	0.3463**	0.3056**
SKIL	-0.8077**	-0.2982**	0.3466**	-0.1290*	0.4312**	-0.2020**	-0.2539**	0.1729**	-0.3002**	-0.5338**
SEMI	-0.5356**	-0.1368*	0.2315**	-0.3367**	0.3552**	-0.0953	-0.2614**	0.2831**	-0.1640**	-0.2583**
UNSK	-0.3293**	-0.1429*	0.7119**	-0.6076**	0.3994**	0.3781**	-0.6521**	0.6628**	0.2131**	-0.3567**
AGRI	0.0359	0.2676**	-0.4638**	0.1689**	-0.1670**	-0.2630**	0.1833**	-0.2976**	-0.2872**	0.4678**
MIN	-0.3331**	-0.1412*	0.1484**	-0.1535**	0.2307**	-0.1378*	0.0055	0.0718	-0.1369*	-0.3028**
MFG	-0.7403**	-0.5501**	0.2564**	0.0212	0.3532**	-0.1948**	-0.1512**	0.1581**	-0.2525**	-0.5392**
TRANS	0.2050**	-0.0255	0.3045**	-0.3105**	0.0466	0.3994**	-0.2692**	0.3126**	0.2918**	-0.0047
DIST	1.0000**	0.2672**	-0.1657**	0.0483	-0.4580**	0.3638**	0.1229*	-0.1271*	0.4583**	0.5281**
GOVT	0.2672**	1.0000**	-0.3027**	0.1075	-0.2071**	-0.0476	0.1814**	-0.2007**	0.0128	0.3313**
NOCAR	-0.1657**	-0.3027**	1.0000**	-0.6545**	0.2642**	0.6141**	-0.7513**	0.6789**	0.4858**	-0.4949**
OWNOCC	0.0483	0.1075	-0.6545**	1.0000**	-0.5488**	-0.6152**	0.6027**	-0.7710**	-0.6018**	0.3762**
COUNCL	-0.4580**	-0.2071**	0.2642**	-0.5488**	1.0000**	-0.2726**	0.0320	0.3783**	-0.0870	-0.5068**
PRIV	0.3638**	-0.0476	0.6141**	-0.6152**	-0.2726**	1.0000**	-0.7579**	0.5697**	0.8100**	-0.0476
AMEN	0.1229*	0.1814**	-0.7513**	0.6027**	0.0320	-0.7579**	1.0000**	-0.6503**	-0.5310**	0.2587**
HDENS	-0.1271*	-0.2007**	0.6789**	-0.7710**	0.3783**	0.5697**	-0.6503**	1.0000**	0.6264**	-0.4379**
RM3	0.4583**	0.0128	0.4858**	-0.6018**	-0.0870	0.8100**	-0.5310**	0.6264**	1.0000**	-0.1491**
RM7	0.5281**	0.3313**	-0.4949**	0.3762**	-0.5068**	-0.0476	0.2587**	-0.4379**	-0.1491**	1.0000**
IRISH	0.2596**	-0.0784	0.3184**	-0.3679**	-0.1287*	0.6061**	-0.4494**	0.5544**	0.6727**	-0.0732
NEWCOM	0.2195**	-0.0372	0.3482**	-0.3735**	-0.1786**	0.6511**	-0.5589**	0.5860**	0.7168**	-0.0593
YOUNG	-0.2263**	-0.0395	0.2989**	-0.5098**	0.3901**	0.2440**	-0.2741**	0.4644**	0.2102**	-0.3077**
OLD	0.4501**	0.1668**	-0.0183	0.2741**	-0.4248**	0.0533	0.0301	-0.3604**	0.0011	0.4409**
UNEMMA	0.0108	-0.1428*	0.5844**	-0.4201**	0.1493**	0.3523**	-0.4225**	0.4887**	0.2236**	-0.1441*
INMIG	0.5874**	0.2433**	-0.3151**	0.1725**	-0.4268**	0.2164**	0.2103**	-0.1006	0.3942**	0.2290**
WITHMIG	-0.2106**	-0.2922**	0.6545**	-0.3122**	0.3319**	0.2131**	-0.3707**	0.3936**	0.1781**	-0.4063**

* - SIGNIF. LE .01 ** - SIGNIF. LE .001 (99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

1C

PARAMETRIC CORRELATION ON UNTRANSFORMED VARIABLES 05/13/80 PAGE 4

FILE TRANS (CREATION DATE = 02/25/80)

- - - - - - - - - - - - - P E A R S O N   C O R R E L A T I O N   C O E F F I C I E N T S - - - - - - - - - - - - -

	IRISH	NEWCOM	YOUNG	OLD	UNEMMA	INMIG	WITHMIG
PROF	0.1179*	0.0299	-0.1026	0.0319	-0.3663**	0.6672**	-0.3600**
EMPL	-0.1159*	-0.1507**	-0.4516**	0.3737**	-0.3467**	0.5459**	-0.5210**
NONM	0.3102**	0.2551**	-0.1032	0.1217*	-0.2606**	0.6649**	-0.1422*
SKIL	-0.1843**	-0.1254*	0.1436*	-0.2040**	0.0844	-0.5924**	0.3896**
SEMI	-0.1652**	-0.1215*	0.1687**	-0.1346*	0.2833**	-0.5380**	0.0205
UNSK	0.1511**	0.1816**	0.3505**	-0.1674**	0.5730**	-0.4379**	0.4636**
AGRI	-0.3235**	-0.2760**	-0.3003**	0.3032**	-0.0611	-0.1478**	-0.5623**
MIN	-0.2584**	-0.2187**	0.0605	-0.1507**	0.1809**	-0.2203**	0.0042
MFG	0.0192	-0.0064	0.2300**	-0.4182**	-0.0763	-0.3648**	0.4316**
TRANS	0.2059**	0.2322**	0.1530**	-0.0661	0.2151**	0.0288	0.1709**
DIST	0.2596**	0.2195**	-0.2263**	0.4501**	0.0108	0.5874**	-0.2106**
GOVT	-0.0784	-0.0372	-0.0395	0.1668**	-0.1428*	0.2433**	-0.2922**
NOCAR	0.3184**	0.3482**	0.2989**	-0.0183	0.5844**	-0.3151**	0.6545**
OWNOCC	-0.3679**	-0.3735**	-0.5098**	0.2741**	-0.4201**	0.1725**	-0.3122**
COUNCL	-0.1287*	-0.1786**	0.3901**	-0.4248**	0.1493**	-0.4268**	0.3319**
PRIV	0.6061**	0.6511**	0.2440**	0.0533	0.3523**	0.2164**	0.2131**
AMEN	-0.4494**	-0.5589**	-0.2741**	0.0301	-0.4225**	0.2103**	-0.3707**
HDENS	0.5544**	0.5860**	0.4644**	-0.3604**	0.4887**	-0.1006	0.3936**
RM3	0.6727**	0.7168**	0.2102**	0.0011	0.2236**	0.3942**	0.1781**
RM7	-0.0732	-0.0593	-0.3077**	0.4409**	-0.1441*	0.2290**	-0.4063**
IRISH	1.0000**	0.7446**	0.2573**	-0.1739**	0.0338	0.3225**	0.1577**
NEWCOM	0.7446**	1.0000**	0.1932**	-0.0938	0.0185	0.3317**	0.1400*
YOUNG	0.2573**	0.1932**	1.0000**	-0.6503**	0.0745	-0.1530**	0.2633**
OLD	-0.1739**	-0.0938	-0.6503**	1.0000**	0.1256*	0.0714	-0.0968
UNEMMA	0.0338	0.0185	0.0745	0.1256*	1.0000**	-0.2753**	0.3568**
INMIG	0.3225**	0.3317**	-0.1530**	0.0714	-0.2753**	1.0000**	-0.3292**
WITHMIG	0.1577**	0.1400*	0.2633**	-0.0968	0.3568**	-0.3292**	1.0000**

* - SIGNIF. LE .01 ** - SIGNIF. LE .001 (99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)