RESEARCH TECHNIQUES

Splitting Background Variables: Aid-Analysis Applied to Migration and Literacy Research

Sune Åkerman, Hans Norman, Egil Johansson
Universities of Uppsala and Umea

Introduction

Scholars who are dealing with more complicated social science research often face the problem that more crude analyses of causal inferences based upon one or a few explanatory variables are unsatisfactory. For that reason it is now more common to try some sort of multivariate analysis. The most frequently used type of such an analysis is multiple regression.

It is apparent, however, that multiple regression analysis has some inherent weaknesses. Using this technique we are forced to work with linear correlations and to accept that the model of explanation has a purely additive shape. An even more serious drawback is the necessity to limit the variables which we are using to those representing a certain level of measurement. We are thus restricted to variables on the level of the interval scale. This means that the researcher has to accept an irritating omission of seemingly very essential predictors which must cause a distortion of our results.

The AID-method

A new technique has, however, during the last decade, been developed, the so-called AID-analysis, which means Automatic Interaction Detector. It has been elaborated and tested at the Institute for Political Research of Ann Arbor, USA. The technique has been presented by John A. Sonquist and James N. Morgan.[1] The technique is basically intended to help to orientate

[1] Sonquist 1970.
A similar technique, the Multiple Classification Analysis (MCA), has also been developed at the same institute. (Andrews-Morgan-Sonquist-Klem, 1973).

the scholar in a vast and complicated mass of data. But at the same time it gives us a possibility to trace the interaction between quite a few variables.

The great advantage of the new technique is that we can also use variables on the level of the nominal scale in our analysis. We can use this technique regarding both the predictors and the dependent variable (criterion variable) under study. A population under study is split into boxes consisting of groups of individuals etc. on the criterion variable. The opposite development is registered on the other half of the so-called tree-output. (Several such outputs are presented further on in our research report). The technique is also called data-split.[2]

It is obvious that the AID-analysis can be used as an alternative to multiple regression analysis. In contrast to this technique, AID-variables on the nominal scale do not cause problems (e.g. civil status, sex, geographic position). Nor will there be assumptions about linearity or normality.

The AID-method is a step-by-step analysis which means that the population will be split successively into mutually exclusive sub-populations. The whole population will be treated in a first stage and further on one after another of the resultant sub-populations. Each predictor has to be divided into classes (at least two). When a population is split, two sub-populations emerge according to the predictor that effects the split, thereby separating one or more classes of the predictor into one sub-population and the others into the second sub-population. The number of possible splits is normally large depending on the number of the predictors and the amount of the classes within them. The classifying of the predictors is performed before the running of the computer (Cf. Appendix).

Out of all possible cleavages on a certain level the computer programme will choose the one that explains most of the variance in the dependent variable. This can be expressed in the following way:

Out of N possible splits the difference in the dependent variable, Y, is defined in the two sub-populations as

$$\lambda_i = \frac{(\bar{Y}' - \bar{Y}'')}{\sigma\,(\bar{Y}' - \bar{Y}'')} \quad i = 1, 2, \ldots N$$

where $\bar{Y}'$ and $\bar{Y}''$ are mean values of the dependent variable within the sub-populations concerned. The output which will yield the biggest λ-value will perform the cleavage. The performance of the cleavage means that the variances in the dependent variable will be " explained " as well as the emergence of a stronger homogeneity as regards the criterion variable.

---

[2] This presentation of the AID-technique is mainly based on the evaluation done by G. Eklund in his introduction to the method (unpublished paper 1974, Department of Statistics, University of Stockholm).

When dealing with dichotomized variables the classification problem is easy to solve. These variables can only take two values, 1 or 0. The split can consequently be made in only one way. If every predictor is divided into several classes a population will be divided in many different ways according to the predictor. Hence we have to consider two types of predictors - *free* and *monotonous*. The last type will cover a natural ranking scale (e.g. age, income). Only those classes which are naturally kept together will be combined. We are not allowed to make groupings like *young* and *old* people into one sub-group and *middle-aged* into another. A uniform predictor with M classes can be divided into M-1 ways. A free predictor does, on the contrary, not have a natural ranking order between the classes. We can accept a cleavage into two sub-populations consisting of any possible combination of classes. (e.g. married + unmarried against divorced + widows/widowers or married + widows/widowers against divorced and unmarried).

There will be more combination options this time. A free predictor of M classes can be split into $(2^{m-1} - 1)$ different combinations. This is very important. There is a much bigger risk that free predictors with many classes will yield splitting artefacts than in the case of free predictors with few classes or uniform ones. The reason for this is that the number of possible splits is different in the case of free and uniform predictors. This is true especially if the analysis includes uniform and/or free predictors with a small amount of classes as well.

A key issue for the performer of an AID-analysis is whether an *interplay* between different predictors can be discerned. Even if the computer programme claims to be able to trace such interplays it is often rather difficult to prove the existence of a true amplifying force. By interplay we mean that several predictors have a *common effect* on the dependent variable, not just an additive effect.[3]

---

[3] G. EKLUND, *op. cit.*, makes the following estimation in a comparison between the multiple regression and the AID technique:

Advantages of AID

1) When the criterion variable is dichotomous.
2) When there is an interaction between the predictors.
3) Yields higher explanatory value (R^2) than multiple regression.
4) The model specification of the AID-analysis applies only to the free or monotonous predictors.
5) The method is easy to understand even for non-statisticians.
6) Easy to present the out-put by means of tree-diagrams.
7) Losses of single observations can easily be handled.

Disadvantages of AID

1) Analyses of significance are difficult to apply, i.e. to limit and control false splits which appear.
2) R^2 too high.

It has been suggested that splits effected by the same predictor on different sub-populations at the same cleavage level would enable us to isolate an interplay situation. (e.g. the " explanatory power " differs consistently between the splits). Boxes on the same level of cleavage will, however, not be split very often by the same predictor.

This method can thus not always be used. We are left in other cases with the possibility to register from the computer lists if interplay-patterns can be found in the ranking of (BSS)-values. This will be demonstrated further on in our study.

Three applications

With the purpose of avoiding a too technical report, we have preferred to present the AID-method applied to three empirical studies. One study covers on an aggregated level three counties in central Sweden during the last decades of the XIXth century. Our units of observation have here been made up of whole parishes in an attempt to shed light on the background variables of emigration.[4]

Another application of the method can be found in an article in SEHR 1974: 1, " The Background Variables of Population Mobility: An Attempt at Automatic Interaction Analysis ", where the internal and external migration of a small area in southern Sweden in the 1880's is investigated.[5]

The method is finally illustrated in a study of reading tests from church examination records, *i.e.* research into the growth of the ability to read in a Swedish parish during the 1690's.[6]

The local variations of emigration

1. *Introductory Perspective*

Emigration from Sweden is characterized by a high degree of geographical variation. This applies both to different parts of the country and within various counties. The Lake Mälar Valley district — for example the counties of Uppland, Södermanland, and Västmanland — stands out as a region with low emigration. On the other hand, emigration to America was high in large

---

3) The stage-wise procedure.
4) The correlations may be linear and lack interaction.
5) The analysis may yield spurious interactions.
6) More expensive to run than multiple regression on the computer.

[4] This part of report is from Norman 1974.
[5] Akerman-Cassel-Johansson 1974, pp. 32-60.
[6] Johansson, 1973.

parts of southern and western Sweden. The county of Värmland was one of the regions which claimed the strongest stream of emigration.[7] (Cf. maps, Figure 6 and 7).

Observing one county (Örebro), we find that the strength of emigration also varied to a large extent within the different administrative areas. Emigration was, as a rule, low in the flat-land parishes. It was, however, considerably higher in the southern and western forest districts and particularly strong within the Bergslag (iron mill) area in the northwestern sector of this county. (Cf. maps, Figure 1).

The causes of the varying extent of emigration within different regions have been much discussed. A number of explanatory variables have been pointed out, such as the economic conditions in trade and industry, the geographical situation, and the intensity and pattern of internal population turnover.

Contrasts in emigration intensity have thus been found between the rural parishes of the county of Västernorrland depending on whether these areas had predominantly agricultural or industrial economies. Emigration was most intensive in the industrialized areas. The findings of other regional studies confirm this picture: as an economic group, the industrial sector had higher emigration frequencies than the agricultural one. But there are also cases where pure agrarian areas had very high emigration. Certain farming districts in Dalsland, for example, are among the areas in Sweden which were most affected by emigration.[8]

The intensity of emigration has also been related to the proximity to larger urban centers of certain areas — i. e. the effect of urban influence fields. Close proximity to a city has, in several cases, led to low emigration figures. In his study of the surrounding area of Gothenburg, E. De Geer has demonstrated the force of attraction exerted by the sphere of commerce and the dominance of a bigger place upon the internal migration patterns of adjacent parishes. This force of attraction has subsequently curbed the emigration from these areas. Similar conditions have been found by S. Carlsson with respect to Småland and by L.-G. Tedebrand with respect to Västernorrland.[9]

The persecution of religious groups which dissented from the state church is commonly advanced as a reason for emigration. S. Carlsson has studied a possible, regional connection between religious separatism and emigration to America during the 1860's by comparing the varying intensity of emigration from Småland with the spread of prayer chapels in one of its counties and

---

[7] *Emigrationsutredningen* (The Report of the Swedish Commission on Emigration), Appendix V, Tab. 77.

[8] Tedebrand, 1972, p. 147; Ebbeson, 1966, p. 61; Norman, 1967, p. 111; Rondahl, 1972, p. 140; Noreen, 1967, p. 26.

[9] De Geer, 1959, p. 194. A study which is now in progress of the Helsingfors and St. Petersburg areas yields a similar line of development. Carlsson, 1966-1967, p. 55; Tedebrand, 1972, p. 145.

the general extent of separatists and Low Church congregations in Småland. He concludes that it is uncertain whether religious factors were relevant in emigration as a mass phenomenon.[10]

Migration researchers have also often treated the issue of a possible connection between internal migration and emigration. Several factors can be mentioned here: 1) the significance of the general population mobility in different regions for the extent of emigration; 2) the role played by the structure of migrational movements, *i. e.* their various distances. S. Åkerman has analyzed research findings regarding internal migration and emigration and advanced conceivable lines of interpretation covering these problems.[11]

2. *Underlying Issues and Study Variables*

The following discussion will examine how some of the above-mentioned factors can be of help in explaining the local variations of emigration from Sweden. Three factors are singled out for analysis: the significance of economic conditions relative to whether the parishes were predominantly agrarian or industrial; the geographical situation of the study areas with respect to urban influence; and the connection between internal migration and emigration.

The study is first carried out in the form of a stage-analysis of the county of Örebro using a smaller number (4) of variables. Further on in the study these are expanded to nine, which have been processed in an AID-program. The following variables are tested in the initial study of the county of Örebro:

A. Economic Factors

1. The size of the arable acreage as a percentage of the area of the parishes.[12]
2. The level of industrialization of the parishes, calculated *a*) on the property ther than agricultural as a percentage of the entire tax value and *b*) on the reported frequency of industrial workers.

B. Geographical Situation

3. The effect of the urban influence of Örebro upon the surrounding areas of migration. The immediate *Hinterland* of Örebro is here equivalent to all parishes within a radius of 20 km from the city.[13]

---

[10] CARLSSON, 1967, p. 118.

[11] AKERMAN, 1971, *passim*.

[12] The calculations of the percentage size of arable acreage are based on the reports contained in *Jordbruksräkningen* 1913-1920 (The Census of Agriculture, 1913-1920).

[13] Parishes have been assigned to this category in cases where more than half of their surface area has been judged to lie within a 20 km radius from Örebro. (Cf. map of the county of Örebro, Figure 1).

4. Location within the Bergslag district in the county's northwest sector. The purpose is here to isolate the iron-producing regions in this part of the county of Örebro.

The study includes the 57 rural parishes of the county of Örebro. The parishes in Table 1 are ranked according to their emigration frequencies. (Mean values 1861-1905).

3. *Economic and Geographic Factors*

It is necessary first of all to determine whether there is a connection between the extent of emigration in the parishes of the county of Örebro, the size of arable acreage, and the level of industrialization. The parishes have for this purpose been ranked in order of priority *a*) on the basis of their relative share of the total emigration between 1861 and 1905, as calculated in per mille of the annual mean population and *b*) on the basis of the percentage share of the cultivated land and the level of industrialization. The probable connection has then been calculated with the aid of Spearman's rank correlation method.

When emigration intensity is related to the cultivated land, a coefficient of −0.43 is obtained. This indicates the existence of a slight negative connection which implies that emigration was less intensive in more developed agricultural areas. A similar estimate regarding the extent of emigration and the level of industrialization gives a coefficient of +0.35. This indicates that more industrialized parishes had higher emigration intensities.

Our evaluation is, however, complicated since the variables to some extent cover each other. Many of the typical agricultural parishes are situated in the flat land of the province of Närke and lie within a 20 km radius from Örebro. Several parishes with a high level of industrialization are similarly situated in the (peripheral) Bergslag region. This is apparent from Table 1. A system of symbols in separate columns marks the 19 parishes ($^1/_3$) which had the highest/lowest figures for cultivated land and industrialization level, the 24 parishes located within a 20 km radius from Örebro, and the 10 parishes in the Bergslag region, in the north-west part of the county.

The parish grouping in this table confirms the results which were indicated by the correlation coefficients. Emigration was thus not particularly intensive in parishes with the largest share of farmland. They appear in the lower level of the scale, and only two reach the upper $^1/_3$. The parishes with less developed agriculture show the opposite tendency. Only two of these are in the lowest part of the emigration scale. In like measure, slightly industrialized parishes had low emigration figures, those with the highest rate of industrialization were characterized by high emigration. The 24 parishes in the vicinity of Örebro

TABLE 1

THE EFFECT OF ARABLE ACREAGE, INDUSTRIALIZATION LEVEL AND GEOGRAPHICAL LOCATION ON EMIGRATION INTENSITY IN THE PARISHES OF THE COUNTY OF ÖREBRO, 1861-1905, EMIGRATION EXPRESSED PER MILLE OF THE ANNUAL MEAN POPULATION, MEAN VALUES FOR THE YEARS 1861-1905

Nr	Parish	‰ emigr. average per year	‰ cultivated land of total area		Degree of industrialisation		Parishes within 20 km from Örebro	Parishes within the Bergslag district
			19 of higest value	19 of lowest value	19 of higest value	19 of lowest value		
1	KARLSKOGA	7.83				△		▬
2	LJUSNARSBERG	5.96				△		▬
3	BO	5.83				△		
4	HAMMAR	4.89				△		
5	NORA IF	4.91				△		▬
6	GRYTHYTTAN	4.80				△		▬
7	SVENNEVAD	4.77						
8	HÄLLEFORS	4.22				△		▬
9	VIKER	4.16				△		▬
10	RAMUNDEBODA	4.09				△		
11	SKÖLLERSTA	4.01			▲			
12	HALLSBERG	3.98				△		
13	LERBÄCK	3.90						
14	HACKVAD	3.65	■		▲			
15	KVISTBRO	3.58				△		
16	LÄNNÄS	3.47			▲			
17	KUMLA	3.43	■			△	●	
18	NYSUND	3.42			▲			
19	JÄRNBOÅS	3.37				△		▬
20	NORRBYÅS	3.27	■				●	
21	EKER	3.22			▲		●	
22	EDSBERG	3.18	■					
23	EKEBY	3.02	■		▲		●	
24	RAMSBERG	3.01						▬
25	LINDESBERG IF	3.00						▬
26	ÖDEBY	2.97					●	
27	HJULSJÖ	2.96				△		▬
28	VIBY	2.85						
29	TÅGERÅSA	2.79						
30	ÅNSTA	2.78	■			△	●	
31	GÄLLERSTA	2.76	■				●	
32	VINTROSA	2.73	■		▲		●	
33	ASKER	2.68				△		
34	FELLINGSBRO	2.68						
35	MOSJÖ	2.61	■		▲		●	
36	KRÄCKLINGE	2.61	■		▲		●	
37	SKAGERSHULT	2.58						
38	LÅNGBRO	2.56	■			△	●	
39	SNAVLUNDA	2.56						
40	HOFSTA	2.52	■		▲		●	
41	ST. MELLÖSA	2.48	■		▲		●	
42	HIDINGE	2.46				△		
43	HARDEMO	2.39	■		▲			
44	ALMBY	2.38					●	
45	GÖTLUNDA	2.23			▲			
46	KIL	2.20					●	
47	KNISTA	2.19						
48	TYSSLINGE	2.19			▲		●	
49	TÄBY	2.19	■		▲		●	
50	RINKABY	2.11	■		▲		●	
51	GRÄVE	2.01	■		▲		●	
52	GLANSHAMMAR	1.72	■		▲		●	
53	AXBERG	1.71					●	
54	ERVALLA	1.49	■				●	
55	NÄSBY	1.40				△	●	
56	LILLKYRKA	1.31			▲		●	
57	ASKERSUND IF	1.30						

Source: *Summariska folkmängsredogörelser*, SCB (Condensed Population Reports, The National Swedish Bureau of Statistics); *Jordbruksräkningen 1913-1920* (The Census of Agriculture, 1913-1920); *Emigrationsutredningen* (The Reports of the Swedish Commission on Emigration), V, Tab. 76; *Population Movements*, 1941, p. 507; Thomas, 1941, p. 212; *Primärmaterial från fabriker* (Primary Material from Factories), The National Archives.

(within 20 km) are Skewed towards low emigration, the Bergslag region on the other hand had parishes with a high degree of emigration.

To sum up, we have noticed the emergence of two groups of parishes. One of them consists of area with a large field acreage, not very developed industry, and a close proximity to the urban centre of Örebro. This group had a low degree of emigration. The other group with its smaller proportion of field acreage, more developed industry, and peripheral location had emigration frequencies far above the average.

4. *Multivariate Analysis with the AID - Method*

The AID-analysis will be used to test the conclusions above. It facilitates the use of a larger set of variables.

Variables in the analysis

The statistics on emigration intensity and internal migration intensity have been tabulated from the Condensed Population Reports and represent mean values of annual statistics for the decade 1881-1890 per mille of the average population.[14] The number of variables used previously has been increased in two ways: first, by a subdivision and further specification of existing variables and, secondly, by the addition of new variables.

In the previous approach a crude subdivision of the variables was applied. This time, however, the parishes have been classified as *agricultural*, *mixed*, and *industrial* ones on the basis of the level of industrialization. The size of the units of cultivation is also treated. Finally a variable called " the tradition of emigration " is introduced, since several studies show that an early acceptance of emigration in an area is usually a prerequisite for subsequent high emigration.[15]

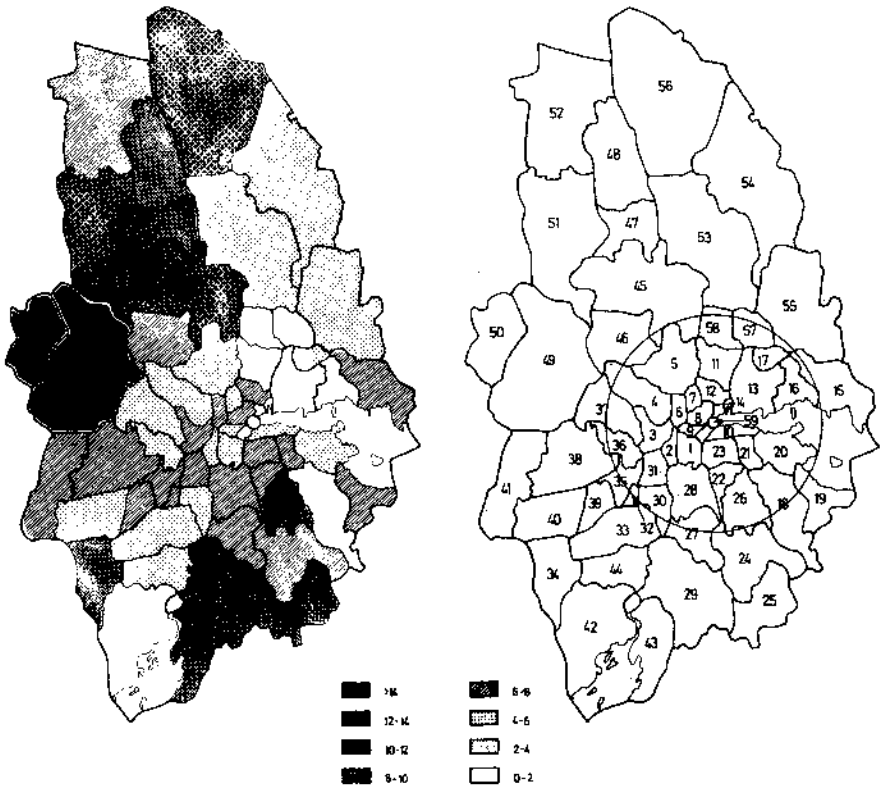
The variables:

1. Parish classes: Agricultural, Mixed, and Industrial.
2. The size of the cultivation units in the parishes:
   a) 0.25-2 field hectares;
   b) 2-10 field hectares;
   c) 10-30 field hectares;
   d) 30-100 field hectares.

   as % of cultivation units above 0.25 hectares

[14] It must be strongly stressed that the analysis is based on observations of individuals from which calculations of mean values in administrative units have been made. Every parish comprises quite a few individuals (in 1880 from 430-11,150 in the county of Örebro) which gives a high degree of stability to the results. This fact has not been observed by Gullberg-Odén in their very detailed criticism of our AID-runs. (Gullberg-Odén 1976, pp. 1-27, 31 f.).

[15] For comments on the variables and the classification for AID-analysis, see Appendix.



Örebro län

1 Mosjö
2 Täby
3 Vintrosa
4 Tysslinge
5 Kil
6 Gräve
7 Eker
8 Längbro
9 Ånsta
10 Almby
11 Axberg
12 Hovsta
13 Glanshammar
14 Rinkaby
15 Götlunda
16 Lillkyrka
17 Ödeby
18 Asker
19 Lännäs
20 Stora Mellösa
21 Norrbyås
22 Ekeby
23 Gällersta
24 Svennevad
25 Bo
26 Sköllersta
27 Hallsberg
28 Kumla
29 Lerbäck
30 Hardemo
31 Kräcklinge
32 Hackvad
33 Viby
34 Ramundeboda
35 Edsberg
36 Knista
37 Hidinge
38 Kvistbro
39 Tångeråsa
40 Skagershult
41 Nysund
42 Askerunds landsf.
43 Hammar
44 Snavlunda
45 Nora landsf.
46 Viker
47 Järnboås
48 Hjulsjö
49 Karlskoga
50 Bjurtjärn
51 Grythyttan
52 Hällefors
53 Lindesbergs landsf.
54 Ramsberg
55 Fellingsbro
56 Ljusnarsberg
57 Näsby
58 Ervalla
59 Örebro stad

Figure 1 - Emigration Frequencies from Örebro län 1881-1890 (parish level).

3. Number of farms over 100 field hectares.
4. Share of crofters as % of cultivation units over 0.25 field hectares.
5. Area of the parish.
6. Field acreage as % of the total area of the parish.
7. Distance from the parishes to the urban centre (urban influence field).
8. Internal migration intensity of the parish.
9. Tradition of emigration of the parish.

Out-put

The result of the AID-analysis for the county of Örebro is presented in Figure 2a. The first cleavage is, in line with the previous observations, effected by the variable expressing shares of field acreage. This means that a rough subdivision occurred between the farming parishes of the flat land, on one hand — with high field acreage and low emigration — and, on the other, the group of mixed farming, foresting and mining, with higher emigration figures.

If a follow-up study is made of the 21 parishes with a high degree of emigration (compare upper branch of tree output) a pattern emerges which conforms to the findings in the previous study. Field acreage is still of great importance, once again effecting cleavage on the third level. There is, however, another variable of great importance, namely the tradition of emigration — *i. e.* an early start of emigration.[16]

It effected cleavage on both the second and the fourth levels. The maximum BSS in Group 13 reveals moreover that the emigration tradition once again is in a position to effect the next cleavage (not visible in Figure 2a).

The highest emigration rates are represented by five parishes with a strong tradition of emigration in Group 5. These parishes are Bjurtjärn with 14.7 ‰, Karlskoga (14.5 ‰), Ljusnarsberg (11.6 ‰), Nora Rural (11.4 ‰), and Hällefors (7.7 ‰). Bjurtjärn and Karlskoga had, out of these five parishes, the highest tradition of emigration according to the scale used in this study. The importance of field acreage and emigration tradition should, however, not be overestimated. By looking at the maximum BSS, one can get an idea as to which variable has a strong explanatory value. At the initial stage, field acreage, the tradition of emigration, migration intensity, and distance from the county urban centre rank as the first four in the same order. Only two of them have effected cleavages of the parish groups in this instance, which can be explained by the fact that we face a strong multicollinear situation; which is schematically expressed in Figure 2b illustrating the position of the four variables with the strongest explanatory values which more or less cover each other. When one variable splits, the others lose their position in the newly-formed groups since they operate in the same direction as the variable which effected cleavage.

[16] See Appendix.

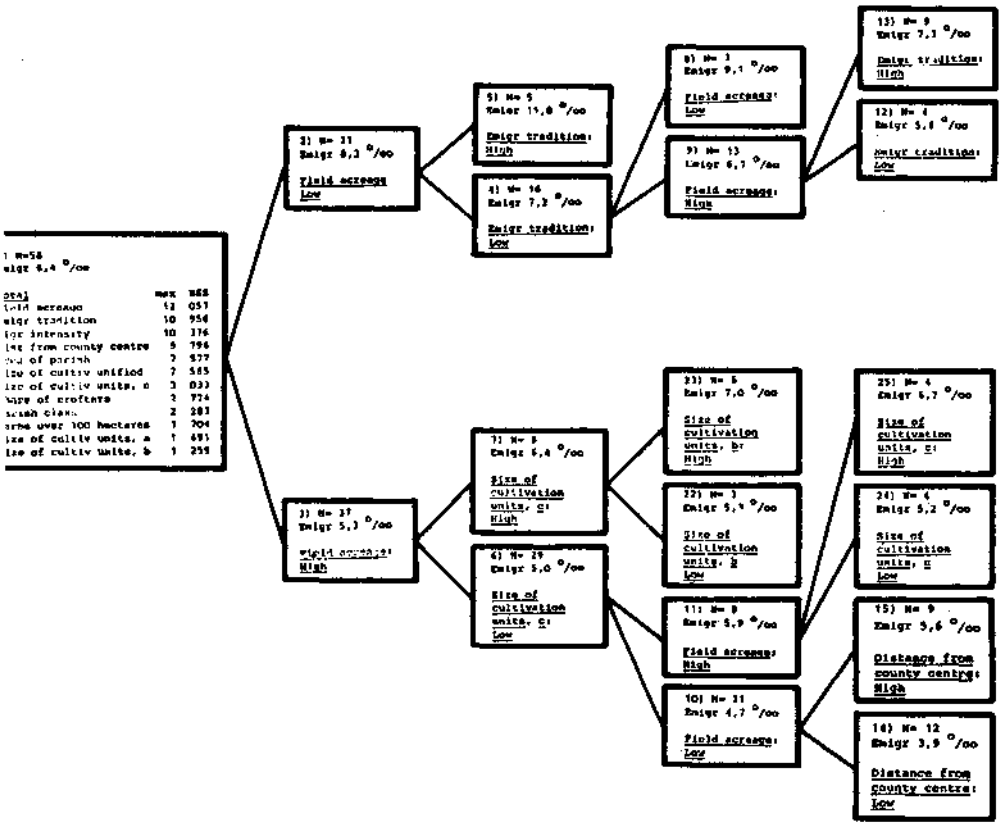
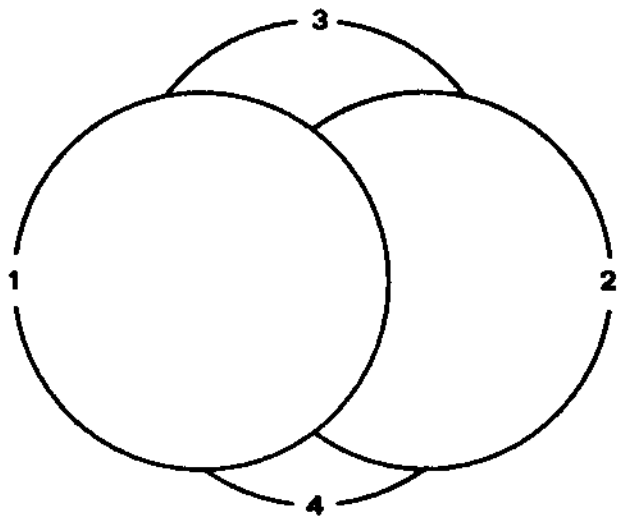


Figure 2 a – AID-Analysis Using Gross Emigration from the Parishes in Örebro län as the Criterion (N = 58). Emigration Expressed Per Mille of the Annual Mean Population. Mean Figures for the Decade 1881-1890.

Sources: D.S. THOMAS (1941), pp. 507 ff.; *Sveriges Officiella Statistik*, Serie Jordbruk och boskapsskötsel, 1913-1920; Condensed Population Reports, The Central Bureau of Statistics, Stockholm; NORMAN (1974), p. 125.

The lower the rates of emigration, the less important the possibility of each variable to demonstrate its explanatory power which is demonstrated in the lower branch of the tree-output, leaving a scope for random results. The size of the cultivation units creates several successive cleavages. No clear pattern can be observed in this development. High percentage shares of cultivation units of both 2-10 and 10-30 hectares result in high emigration. A group



1 Field acreage
2 Emigration tradition
3 Migration intensity
4 Distance from the county centre

Figure 2 b - Theoretical Relationship between Variables in an AID-analysis. (Examplified from Figure 2 a).

with high field acreage similarly claims the highest degree of emigration in this case. The emigration tradition has moreover played a subordinate role among this particular category of parishes in the county of Örebro. This is only natural since these parishes had a late start in their emigration. As shown by the size of the maximum BSS, the tradition of emigration here stands at a consistently low level.

The variable of the urban influence (distance from the county centre) is as important here as in the prior study. The urban influence thus tends to curb the emigration rates, demonstrated by the parishes close to Örebro. The maximum BSS in Group 15 shows, moreover, that field acreage again is in a position to effect the next cleavage. The result of this cleavage matched the anticipated course of development: a low share of field acreage again gave the highest degree of emigration.

The findings of this AID-analysis fall into line with those obtained in the stage-analysis of the county of Örebro, which also gave two main parish groups.[17] The greatest emigration came from the group of parishes farthest from the centre of the county with a low share of field acreage and an established tradition of emigration. This parish category also had a wide surface area which tended to give a low level of registered internal migration.[18] The lowest emigration came from parishes with the opposite character (close to the county centre, a high share of field acreage, a late acceptance of the emigration alternative, a higher level of internal migration). The findings from the county of Örebro were also tested on an extended base, comprising the counties of Värmland and Västmanland as well as the county of Örebro. The AID-analysis was thereby tried on 214 parishes. In this combined analysis the variable of migration intensity dominated as a cleavage factor. There are in this case two circumstances influencing the connection between migration and emigration. In the county of Värmland a low internal migration intensity, primarily in the parishes along the Norwegian border, was combined with a high emigration. From these border parishes a strong unregistered seasonal migration occurred, however, at the same time.[19] As for the counties of Örebro and Västmanland, prominent emigration was prevalent in the peripheral parishes with large areas and a low registered migration intensity. Emigration was, on the other hand, lower and the internal migration higher in plain farming districts where the parishes were small and fell within the urban influence fields of Örebro and Västerås.

The three counties differ in many respects, offering contrasts in their geography, natural resources, and internal migration patterns, as well as in their emigration intensity. This study has nevertheless revealed several consistent features. The effect of the urban influence field stands out in all three counties besides that of the internal population mobility. The economic structure of the counties was also important: more industrialized parishes had the highest degree of emigration. This is best illustrated in the counties of Västmanland and Örebro.

The variables covering the size of cultivation units again failed to provide a distinct pattern. A high share of crofters seems to be associated with a low level of emigration: this is probably to be regarded as a coincidence.[20] Like

---

[17] The fact that the two different methods used here almost give the same result seems to confirm the value of the AID-technique. This interpretation has, however, been questioned by K. AGREN in a review of this study. (*Historisk Tidskrift* 1975).

[18] Cf. NORMAN 1974, pp. 113, 141.

[19] Cf. NORMAN 1974, p. 130, note 8.

[20] The crofters and their families have often been regarded as a group which was very prone to emigrate. Obviously the general situation of the crofters has varied considerably yielding different emigration patterns. NOREEN 1968, pp. 102-107.

other categories of property owners, crofters ought to have had a low inclination to emigrate. There is, however, nothing pointing to the fact that their large Numbers of children would have been less inclined to emigrate than other categories. One variable, which consistently gave a clear and uniform result in all three cases, was the tradition of emigration.

The comparison of the results from the AID-analysis with the findings of the prior stage analysis of the county of Örebro, has as a conclusion yielded a rather uniform pattern. The subsequent, broader analysis of the three counties has shown that certain uniform features can be observed, despite regional differences. Thus, the registered internal migration proves to vary inversely to the emigration rates. Industrialized parishes tend, moreover, to show a high level of emigration. The effect of the urban influence fields on the internal migration patterns of the parishes appears, furthermore, to have curbed emigration from these areas. Finally, a strong tradition of emigration — *i. e.* an early start of the mass movement from a region — has usually spurred the emigration later on.

The background variables of emigration

The investigation above dealt with parishes as units of study. The next step will be to make observations on the level of the individuals. Such a study has been performed for a parish in western Sweden (Köinge) during the period 1878-1888. Our main topic for research has in this case been *emigration*, and the choice of the investigation area was dictated by that. The main part of western Sweden had a high emigration rate and it culminated in the 1880's. The parish that has been investigated had an extremely high degree of emigration (24.8 ‰ net-emigration) as compared to the national total (6.0 ‰).

Three analyses

Population movements form the criterion in the present study. The persons studied are recorded either as having moved away or emigrated during the currency of the church examination register (1878-88) or are still domiciled at the expiry of this register in 1888. Thus, the criterion takes no account of how individuals were entered in the register. This factor presents itself instead as one of the explanatory variables. The criterion can be changed in numerous ways. Three variations, leading to three different analyses, are presented below. Only persons 15 years and older are involved in the examination.

TABLE 2

THREE VARIANTS OF THE CRITERION

	Person	Criterion values I	II	III
Emigrated 1878–88	130	1	1	1
Out-migrated (within Sweden) 1878–88	234	1	0	—
Still domiciled 1888	408	0	—	0
Total number of persons	772			
Mean criterion value (%)		48	35	24

The table reveals how the criterion in each analysis is composed. In the first one, all persons who migrated are contrasted with persons who did not migrate. In the second, persons who migrated within Sweden are compared with persons who emigrated abroad. And in the third, emigrants are put against persons still domiciled. The criterion values 1 or 0 group the individuals in different ways. The mean criterion value in each analysis is shown in the last line of the table in the form of percentages.

First analysis: persons out-migrated or emigrated compared with persons still domiciled

The criterion in the first cleavage comprises all persons (364) who out-migrated or emigrated in the period 1878–88 as against persons (408) still domiciled in 1888. The criterion value 1 has been assigned to the former group and 0 to the latter. The mean criterion value of the former combined group is thus $\bar{y} = 0.475$, which means 48% mobile persons.

The explanatory variables are village, occupational group, size of family, year of birth, county of birth, sex, civil status, number of migrations, in-registration in the church examination register, and reading grades awarded in this register. (The coding and frequency distribution of these variables can be seen in Åkerman et alii 1971, pp. 45–54). These variables are thus competing in the splitting process. The pattern of the cleavage is represented in a tree diagram (Figure 3).

The number of each new box indicates the sequence in which the cleavage proceeded. The sequence gives information about the capacity of the explanatory variables to perform the cleavage. The tree diagram thus shows how the proportion of persons that out-migrated and emigrated changes in each new box which is illustrated in Figure 4.

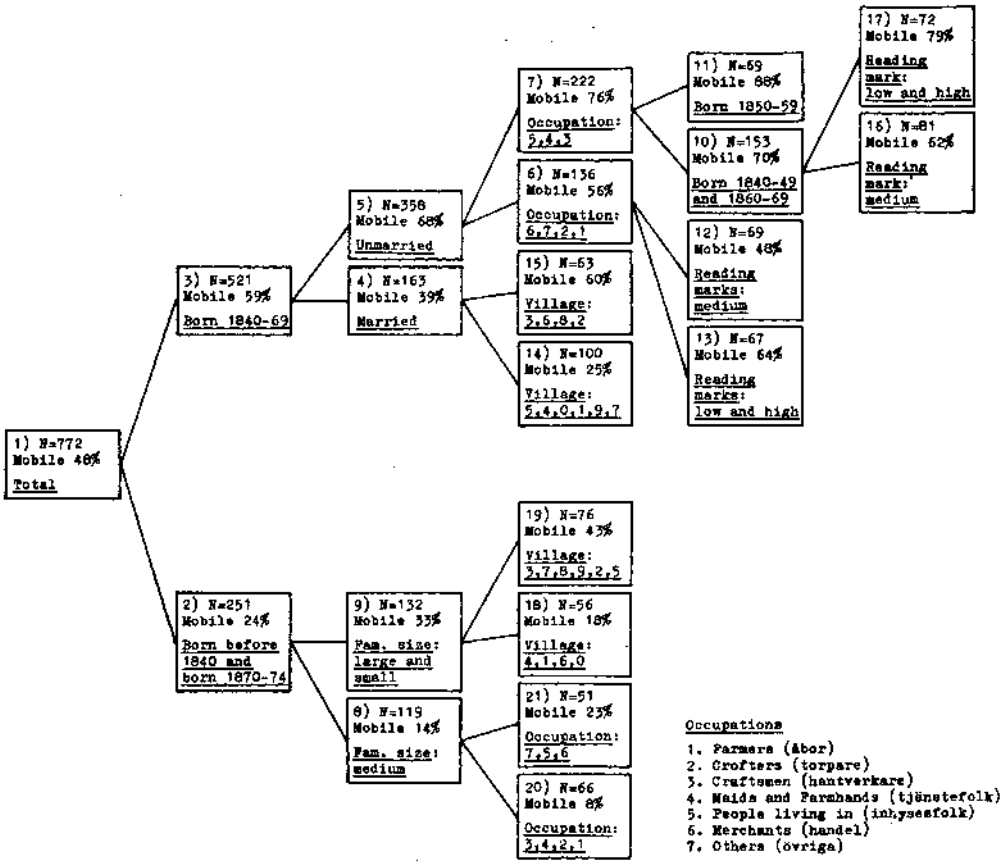


Figure 3 - *Data Cleavage I*

Persons Still-domiciled and Mobile (emigrated and internal migrants) Köinge 1878-88 (N=772).

Criterion

Mobile (emigrated and internal migrants) = 1
Still-domiciled 1888 = 0.

In the original group we find that 48% of the individuals belonged to the mobile category. Distribution by age produces, however, a more typical migrant group (group 3) and a more typical still-domiciled group (group 2). Almost 60% of the younger people, born in 1840-69, are migrants. Most of the older persons, born in 1830 and earlier, and the youngest, born in 1870-74 are on the other hand still domiciled; only 24% are mobile.

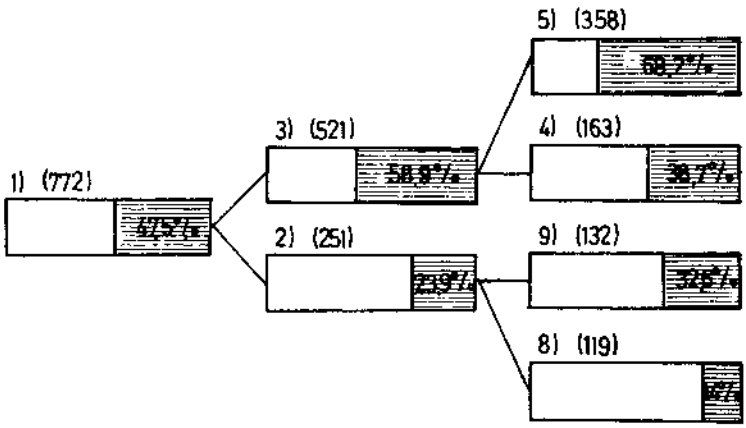


Figure 4 - Percentage share of migrants on different cleavage levels. (data split of Fig. 3).

So the next step for these two new groups is to be split. Next in turn was the younger group (group 3). Civil status is the operative factor here. The unmarried are mobile (68%). Married people are to a greater extent immobile, only 39% migrants.

If we look at how these predominantly single persons (group 5) distribute themselves later on, the first thing we observe is that occupation is an operative factor. Servants and "*inhysesfolk*" [21] (group 7) are for instance more mobile than other occupational groups (group 6). When the former group (group 7) has once again been subdivided into two age-groups (group 10 & 11), we have a further and final cleavage on the basis of reading ability (groups 12, 13, 16 & 17). At both these cleavages the migrants reveal a reading-mark profile distinctly different from the others. Hence, the lowest and highest reading-mark values are to be found in group 13 & 17. Groups 12 and 16, on the other hand, show more or less average reading-marks.

Another pattern of cleavage will be presented before further conclusions are drawn. Attention will this time be focused on *migrants*.

Second analysis: persons emigrated compared with persons out-migrated

The criterion of the next cleavage is based upon the two groups of persons emigrated (130) and persons out-migrated within Sweden (234). The former are assigned the criterion value 1 and the latter the value 0. The mean criterion

[21] Persons living in a household without being members of the family.

value is thus 0.354, i. e. 35% emigrants. This cleavage can be regarded as an amplification of the one above.

The explanatory variables are the same as in the previous cleavage. These explanatory variables produce the pattern shown in Figure 5. A difference can at once be observed compared with the last result. The two highly effective personal variables, year of birth and civil status, do not occur in the first part of the diagram. Most of the migrants dealt with are young and single. Cleavage begins instead with the social and geographical factors: occupation, rural settlement area, size of family (groups 2-7, 12-15). Two distinct areas of emigration in the parish of Köinge can be observed. Emigration occurred thus with a clear geographical concentration.

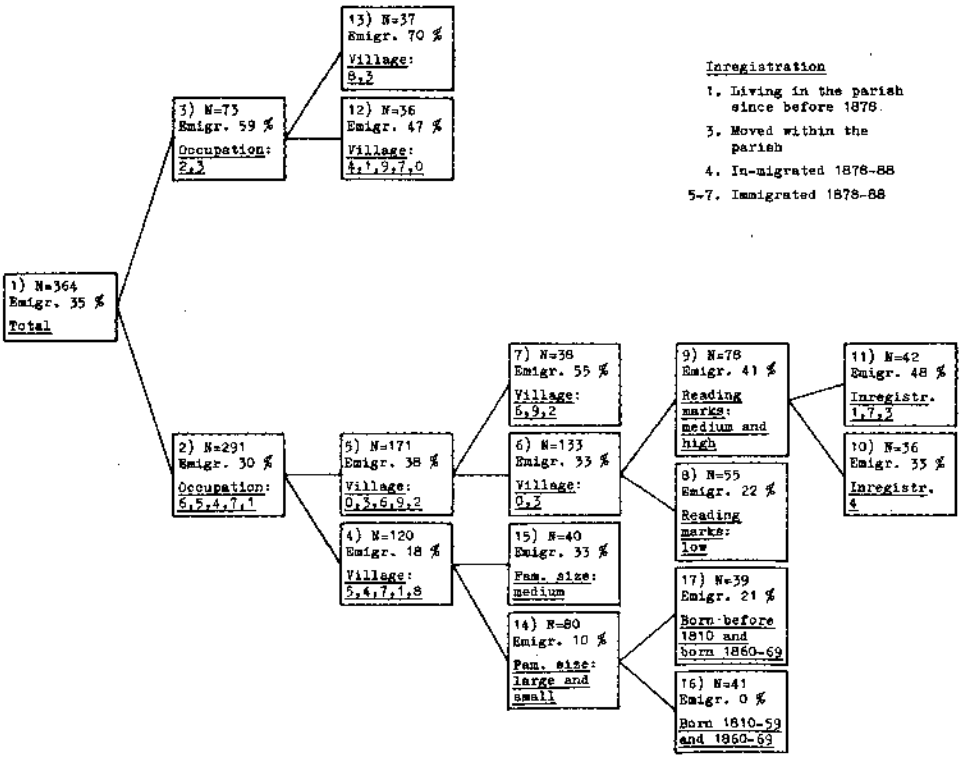


Figure 5 - *Data Cleavage II*
Persons Moved Away and Emigrated. Köinge 1878-88 (N = 364).

Criterion

Emigrated = 1 (N = 130)
Internally Out-migrated = 0 (N = 234).

When 133 persons are distributed according to reading ability, the result is again different from that of the former cleavage. Middle-range and higher marks cluster in a group with a large proportion of emigrants (group 9), while migrants within Sweden (group 8) show low marks. Some comment is needed here. Migrants within Sweden sometimes had lower marks recorded in the church examination registers. Their mobility made them less well known to the marker. So far the result in group 8 is as one would expect. When the group with a larger proportion of emigrants (group 9) shows higher marks this gives us, however, a hint that emigrants deviate from migrants in their reading ability. When we add to this the observation from the previous cleavage that persons who outmigrated or emigrated obtained marks that were sometimes lower and sometimes higher than those of the still-domiciled, a composite explanation can be arrived at. The argument above can be shown schematically in a figure (Figure 6).

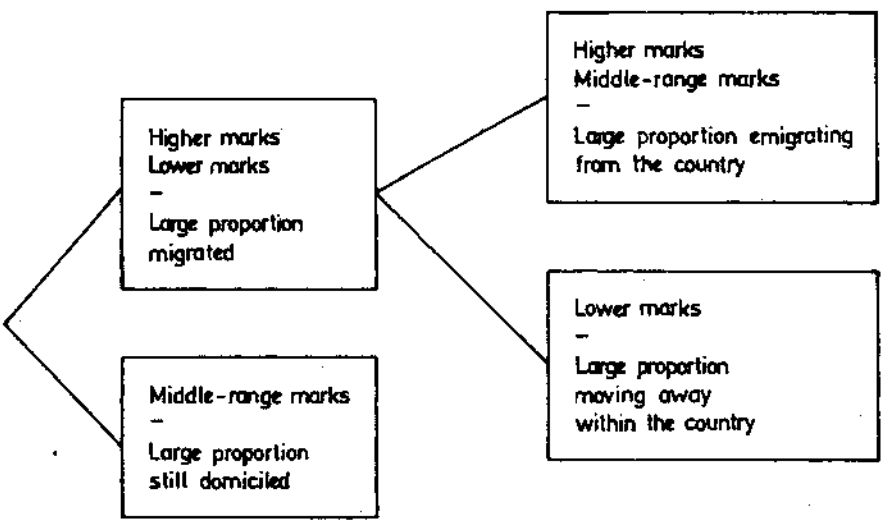


Figure 6 - Schematical representation of the reading results.

There are certain defects in this argument. Its greatest weakness is that the result is based upon much reduced and selected groups. It provides, nevertheless, an idea of the place of reading ability in the intricate web of relationships. A circumstance that seems to increase the force of the result is that the ranks of those attaining the highest marks probably already had been thinned out by earlier emigration. This makes it the more remarkable that these marks were so well represented among the emigrants. The results should also be

compared with the values obtained by presenting the total population in a cross-table:

TABLE 3

READING MARKS FOR ALL INDIVIDUALS (N = 772)

	Readings marks in Church examination records 1878–88 *				
	Low %	Medium %	High %	Total %	Total N
Emigrated	35	51	14	100	(130)
Out-migrated	38	51	11	100	(234)
Still domiciled	28	61	11	100	(408)
Total	32	57	11	100	(772)

* Low = d-cd-e, medium = bc-c, high = ba-ab-a-A.

The table confirms the split pattern. Some of the emigrants have lower (35%) and some have higher marks (14%) compared with the rest. Persons who out-migrated occupy a notably low position on the marking scale with 38% achieving low marks. The still-domiciled, on the other hand, tend to cluster around the middle-range of marks (61%). Notice that the effects of the split have not been used here. The groups are instead presented in their entirety. If age is kept constant and only one cohort (those born 1860-69) is focussed, the contours of the picture are reinforced:

TABLE 4

READING MARKS OF THE COHORT BORN 1860-69 (N = 240)

	Low %	Medium %	High %	Total %	Total N
Emigrated	30	43	27	100	(49)
Out-migrated	34	50	16	100	(100)
Still domiciled	23	59	18	100	(91)
Total	29	52	19	100	(240)

It will be observed that younger persons attain higher marks. This may have something to do with their being near the age of confirmation, when the marking was most detailed and differentiated. At higher ages the marking is more clustered around the middle range. There is also the fact that the whole system of marking began to weaken in the latter part of the XIXth century. But apart from this difference, this table displays the same result as the previous one, only more sharply outlined. Some emigrants have low marks and some

have high ones. Those who migrated within Sweden have a disitnctly lower distribution of marks. The still-domiciled are again clustered around the middle range of marks.

The results obtained thus suggest that the marking profile of emigrants in the church examination records is of a quite special character. Since this is not a random sample of large areas no far-reaching generalizations can be made: the result applies to the individuals dealt with.

Third analysis: emigrants compared with still-domiciled

Emigrants (130) are now compared with still-domiciled (408) in a final cleavage pattern. Using the same criterion as before (emigrants 1 and still-domiciled 0) provides a mean criterion value for the first undestributed group of 0.24, *i. e.* 24% emigrants (Figure 7).

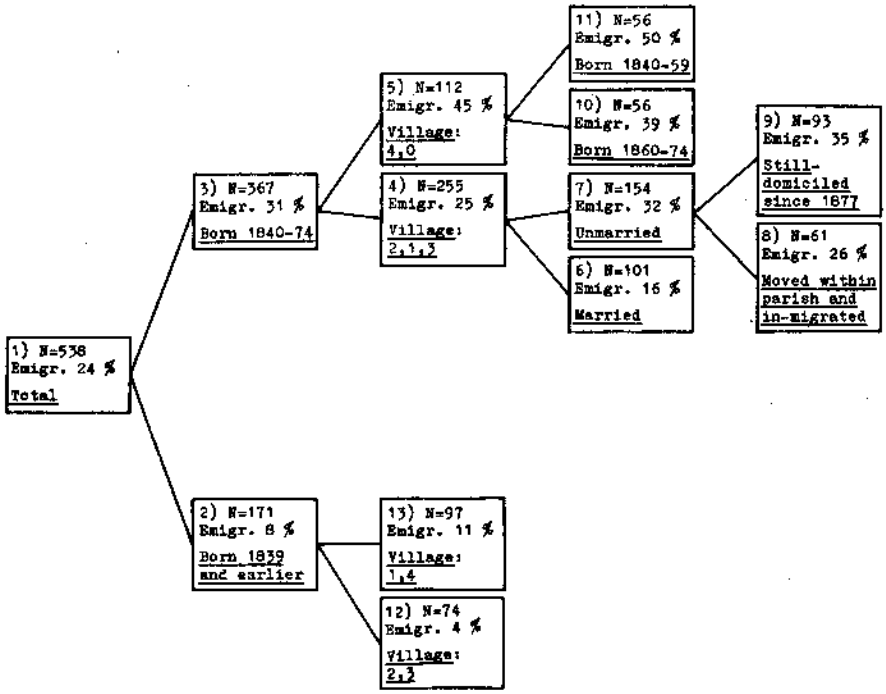


Figure 7 - *Data Cleavage III*
Still-domiciled and Emigrants. Köinge 1878-88 (N = 538).

Criterion
Emigrants = 1
Still-domiciled = 0.

The explanatory variables are the rural settlement area, occupation, year of birth, civil status, in-registration in the church examination record and reading marks in this record. This cleavage repeats the first one in large measure. The age factors that are crucial to emigration emerge first of all. Emigrants are recruited from the young ones (group 3). Older people make up the still-domiciled (group 2). The geographical factor is decisive (groups 4-5 and 12-13) in the next split and further on civil status and then age once again (groups 6-7 and 10-11). The type of in-registration finally becomes the critical factor (groups 8-9). Recent inmigrants were less prone to emigrate.

The pattern largely confirms the first cleavage. The explanatory capacity of the variables can be detected by closer observation of the ranking of the competing variables, (Figure 8). The age factor (BSS=6.25) is in the first group followed fairly closely by rural settlement area (BSS = 4.80), civil status (BSS = 4.72), and occupation (BSS = 4.18). After cleavage by age, the explanatory capability of the latter factor declines in groups 2 and 3. The explanatory capacity of the rural settlement area increases instead in groups 2 and 3, where this variable comes first in the new ranking. When the geographical factor too has exhausted its explanatory capacity, civil status dominates in group 4 and the type of inregistration in group 7. The latter variable had low values in earlier boxes but is thus left as the strongest in the not very extensive group 7.

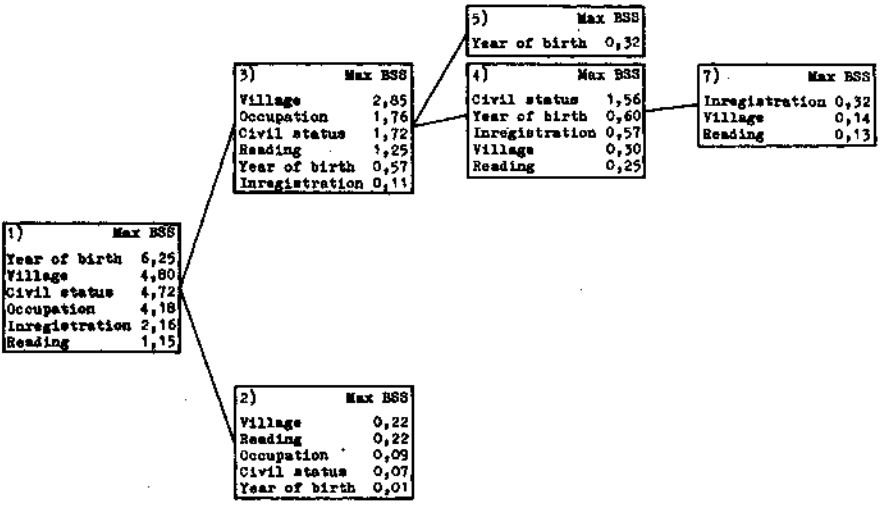


Figure 8 - *Data Cleavage* III b
Still-domiciled and Emigrants. Köinge 1878-88 (N = 538).

Order of precedence of explanatory variables according to Max BSS.

This way of evaluating the ranking of the variables also sheds light on the position of reading marks in the final data cleavage. In group 2, reading mark competes with the rural settlement area, both having the same max BSS. In group 3, reading ability is also high up in the ranking. It occupies a less prominent position in group 4, but is still important in group 7.

This more in-depth interpretation of the data cleavage makes our analysis more penetrating. It allows us to compare various cleavage patterns more easily. It also elucidates how one variable can be covered by others obviously measuring approximately the same thing.

The growth of literacy in a swedish parish in the late 1600's [22]

In the oldest church examination register of the parish of Tuna in the county of Västernorrland (cf. map) 1688-1691, the villages are divided into six small districts. In addition to names and ages of all the individuals there are special columns with notes about reading ability and memorization. The notes about reading ("kan läsa i bok") are in plain language as follows: "inte" — can *not* read, "lite" — a *little*, "begynt" — has *begun* to read, and "kan" — *can* read. The knowledge of the Biblical and the Apostolical words of the Catechism ("Simpliciter partes Catechismi") is noted in these columns. Then follow separate columns with notes on each of the Explanations by Luther of every part of the Catechism ("Decalogus, Symbolum, Oratio Dominica, Baptismus, Sacra Caena cum explicatione"). Then follow two columns, the last one referring to the "Hustavla," *i. e.* the Biblical words of the political, social and religious order of the Catechism. Reading ability is compared with catechetical knowledge in Table 5.

Table 5

READING ABILITY IN TUNA 1688-1691
COMPARED TO CATECHISM KNOWLEDGE (N = 397)

Columns about catechism	No note	Reading marks "Can not"	"A little" "Has begun"	"Can"	N
6-8	1	12	27	74	114
5	1	21	35	52	109
1-4	3	30	37	43	113
No note	14	20	20	7	61
Total	19	83	119	176	397

[22] For recent literacy-research, see e. g. Graff (1974), Johansson (1973), Lockridge (1974), Schofield (1972).

397 persons are noted in the districts 2-6. (The first district is omitted because there are some pages missing in the beginning of the record). 19 persons have not got notes in the column for reading – 83 are marked " can *not* read " while 119 " has *begun* " or can read " a *little* " and 176 persons " *can* read ". The connection with the Catechism is obvious. Those who have got a high score in memorization also have high reading marks. 61 persons have no notes in Catechism compared to 19 for reading. This is typical. Notes about reading are often given earlier in childhood than notes about memorization. But it is important to observe that " no note " is not equal to " can not." " No note " is more like " has not been examined " and typical for the youngest (and oldest) persons. 12% of the children younger than 11 years and 9% of the persons over 60 years of age have no note in the column for reading. The persons without notes have been exluded in the following tables.

A feasible way of analyzing the reading marks is to compare the marks with age (Table 6).

TABLE 6

READING ABILITY IN TUNA 1688-1691 COMPARED WITH AGE (N = 370)

Age	Reading marks			Total %	N
	" Can not " %	" A little " " Has begun " %	" Can " %		
Age > 60	47	8	45	100	(38)
51-60	46	13	41	100	(32)
41-50	37	21	42	100	(38)
31-40	21	37	42	100	(43)
26-30	21	32	47	100	(38)
21-25	14	23	63	100	(58)
16-20	16	27	57	100	(37)
11-15	2	47	51	100	(49)
< 11	5	70	25	100	(37)
Total	22	31	47	100	(370)

There is a notable difference between the age groups. In the oldest groups almost half of the registered persons are noted as *not* reading and some of them as " a *little*." " A *little* " belongs to the older groups and the note " has *begun* " belongs to the younger ones. Almost everybody has got reading marks in the youngest group. The middle column, " has *begun* to read," is very frequent among the young ones. This is very plausible in the ages when one learns to

read. The youngest children in this register are from six to eleven years of age. The same typical differences between ages has been found in other examinations. This demonstrates how the reading campaign in Sweden began in the late 1600's.

The age distribution of reading indicates differences between the sexes. (Table 7).

TABLE 7

READING ABILITY IN TUNA 1688-1691 COMPARED WITH AGE AND SEX

Age Sex	Reading marks " A little " " Can not " %	 " Has begun " %	 " Can " %	 Total %	 N
> 50					
Male	34	10	56	100	(38)
Female	49	16	35	100	(70)
21-50					
Male	10	25	65	100	(57)
Female	23	33	44	100	(82)
< 21					
Male	1	52	47	100	(58)
Female	12	45	43	100	(65)
Total	22	31	47	100	(370)

The differences between the sexes are most striking in the oldest group. They are less appearant in the younger age groups although still quite clear.

This typical pattern of literacy, sex and age groups, can be analyzed in many ways. We will now look at an interaction analysis where a lot of variables are involved at the same time in a computer programme, the AID-analysis. The proportion of notes for reading: " a *little*," " has *begun* to read " and " *can* read " is taken as the dependent variable. Catechism, sex, age, district, family position, and occupation are used as explanatory variables. The first three are already presented in the previous tables. A geographical factor will be covered by the variable of districts. We are most closely examining the distance from the parish church. Social factors and realities are mirrored by the two variables family position (children, parents, relatives, others in the household), and occupation (servants and farmers' families).

The data are in each step divided into two new sub-groups in a one-way analysis of variance according to the highest BSS (between sums of squares). The numbers of individuals (N), the percentage of reading marks (%) and the classifications according to the splitting variable are given in each group as was done earlier. (Figure 9).

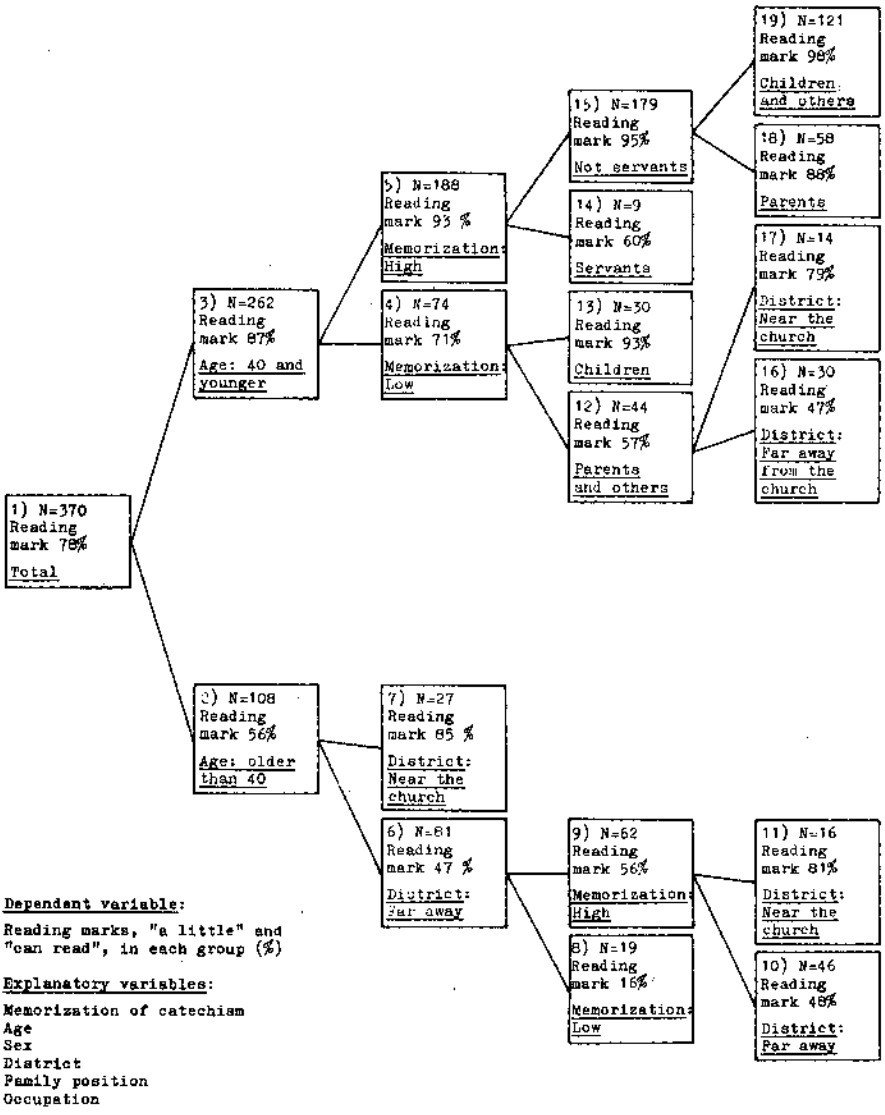


Figure 9 - *The Background Variables of Literacy (Tuna Parish 1688-1691).*

The total literacy rate was 78 per cent. The first division is carried out by age. Among people younger than 40 years the reading ability was high (87%) while older ages only reached slightly more than 50%. Memorization and social indicators perform the splitting of the youngest individuals on the next levels. The older people, on the other hand, were divided according to district i.e. if living near the parish church or more far away. The main finding has been the changing level of literacy from one age cohort to the next. At least three generations can be noticed (compare Figure 10). This means that the alphabetization campaign could improve the situation for adults (even the oldest ones) — important in the short run — but that the real target was the small children. It is remarkable that *sex* has been unable to perform any cleavage. This may be due to the shape of this variable. Since it is dichotomized it can only obtain two values. Hence such a variable, as was mentioned earlier, can have difficulties in competing with more specified ones. The shaded out-puts show, however, that *sex* is very close to a split several times.

The results for Tuna are consistent with what we have found in other parishes. A sample of parishes for the whole country has been processed. It is quite evident that Sweden already had a literate population around the

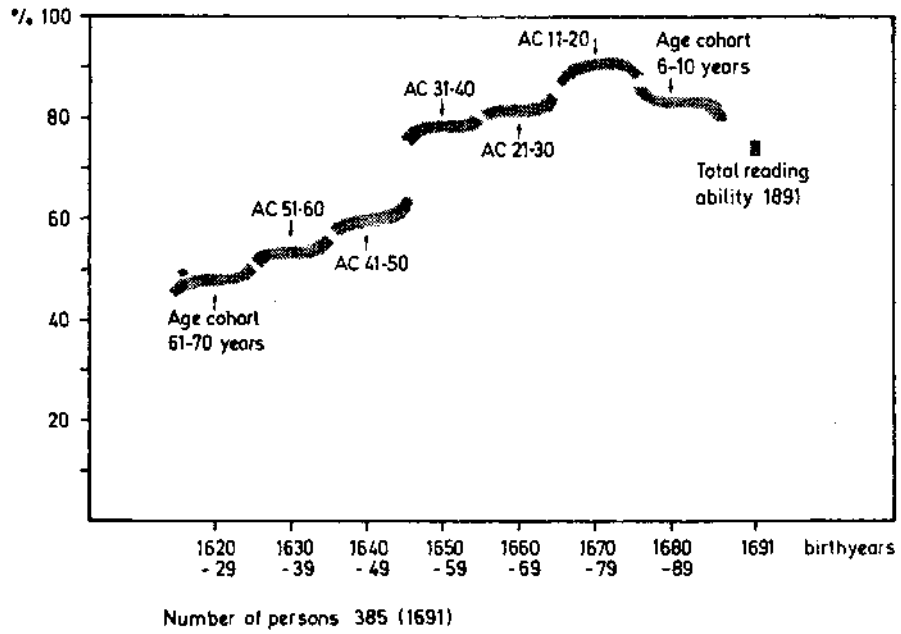


Figure 10 - Result of the Alphabetization Campaign in Tuna Parish (Cohort-wise Increase of Reading Ability Mirrored in the Literacy Rates of 1691).

turn of the century (1700), at least if we talk about an elementary reading ability. The Lutheran Church in cooperation with the different households and families were responsible for this development. The idea was that parents gave their children and other household members instruction. It was not, however, until the XIXth century that writing was made more general. This time the agent was the compulsory school.

AID-ANALYSIS. A SEARCH STRATEGY AND AN INTERPRETATION DEVICE

Some experiments have recently been carried out by historians and educationalists to use the so-called AID-analysis and at the same time try to evaluate it as an empirical method. Two investigations dealing with migration and one study of literacy have been presented here.

Our experience is that the method can be applied successfully to these research topics. The structure obtained has been shown to be consistent with earlier research of a more traditional kind, especially concerning emigration, and it also seem to be of some consequence and logic.

The conclusions do not, of course, exclude the fact that the AID-analysis, like other similar statistical techniques has certain general weaknesses. At the same time it creates some interpretation problems of its own. We have in our delineation pointed out some of these difficulties. Substantial difficulties thus appear when we try to create a good set of variables. If by chance an important variable drops out, depending on minimal or fragmentary information a whole tree-output may be curtailed. This is, however, a weakness that applies less to the AID-analysis than to other comparable statistical techniques which do not treat variables on the level of the nominal scale. We may at the same time add that we have the opportunity to experiment with a set of variables by cutting out one or more important variables from our analysis. Such an experiment can shed light on the different patterns of interaction and their variations according to the number of variables used.

This method thus offers an unusual span since it enables us to use an extensive set of variables. This is rather easy to perform by way of computer processing. This may even, paradoxically enough, cause problems. We may face such an abundance of variables and such complex information, that problems of interpretation appear. In such situations it will be valuable to construct what we call a "shading output," presenting the variables that compete with the one that has performed the split on every level (= ranking of maximum BSS). This procedure reveals part of the complicated interplay between different variables and we find it possible to register both the ranking of the variables and to trace the interplay between them after each split (clustering of the variables).

The persons behind the AID-analysis have stressed that the method can primarily be used to trace structures. Thus the method helps us to reveal the topography of a landscape. This means that we also can use it to generate hypotheses. We must at the same time stress that this method has a pedagogical advantage. It helps us to master a rich and sometimes bewildering empiricism. One of our migration studies especially sheds light on this potential of the method. We have in this case been able to reduce the multitude of variables to a smaller group of the most important ones.

At the same time that we state that AID-analysis can function as a search strategy we have also more and more become aware of how important it is that the scholar from the very beginning has a good knowledge of his material and preferably also has had a possibility to test it with other statistical methods since an application of the AID-method may produce pure artefacts.

This is a consequence of defects in the variables, which we have to work with and it can also be a result of the technique of grouping the classes. AID-analysis is, however, not only a search strategy. Its capacity to take care of a great variety of factors creates an opportunity to make a synthesis. It can therefore also function as an instrument to test hypotheses. (This concept is used in its broader meaning, consequently not with a statistical terminology). We can thus trace an interaction between a great many variables telling us about actual causal inferences.

Statistical research is still working on this level to find out which type of interactional pattern a specific data-split represents. There are different interpretations: one can find additive, cumulative and substitutive outputs. The empirically inclined researchers have every reason to catch up with this work of development. All these intriguing problems are far from being solved. The impression of our research is that causal connections cannot easily be captured in this type of simplified interpretation. Some of our data-splits seem to indicate more complicated structures. This fact raises the question if it is not necessary to plan and carry out more empirical investigations than has been done up to now by the use of AID-analyses. That would be an important complement to the current rather intense statistical work of development.

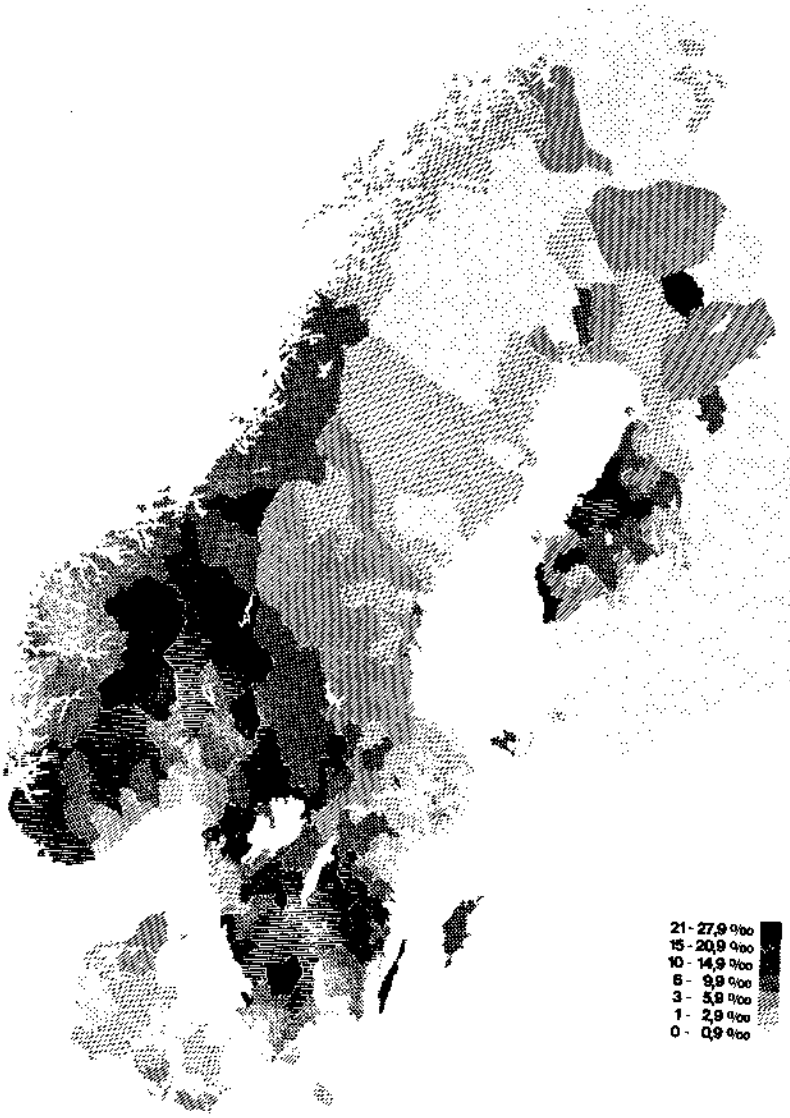


Figure 11 - Gross Emigration Rates for Sweden, Norway, Denmark and Finland 1885-1889.

Source: Material from the *Nordic Emigration Project.*

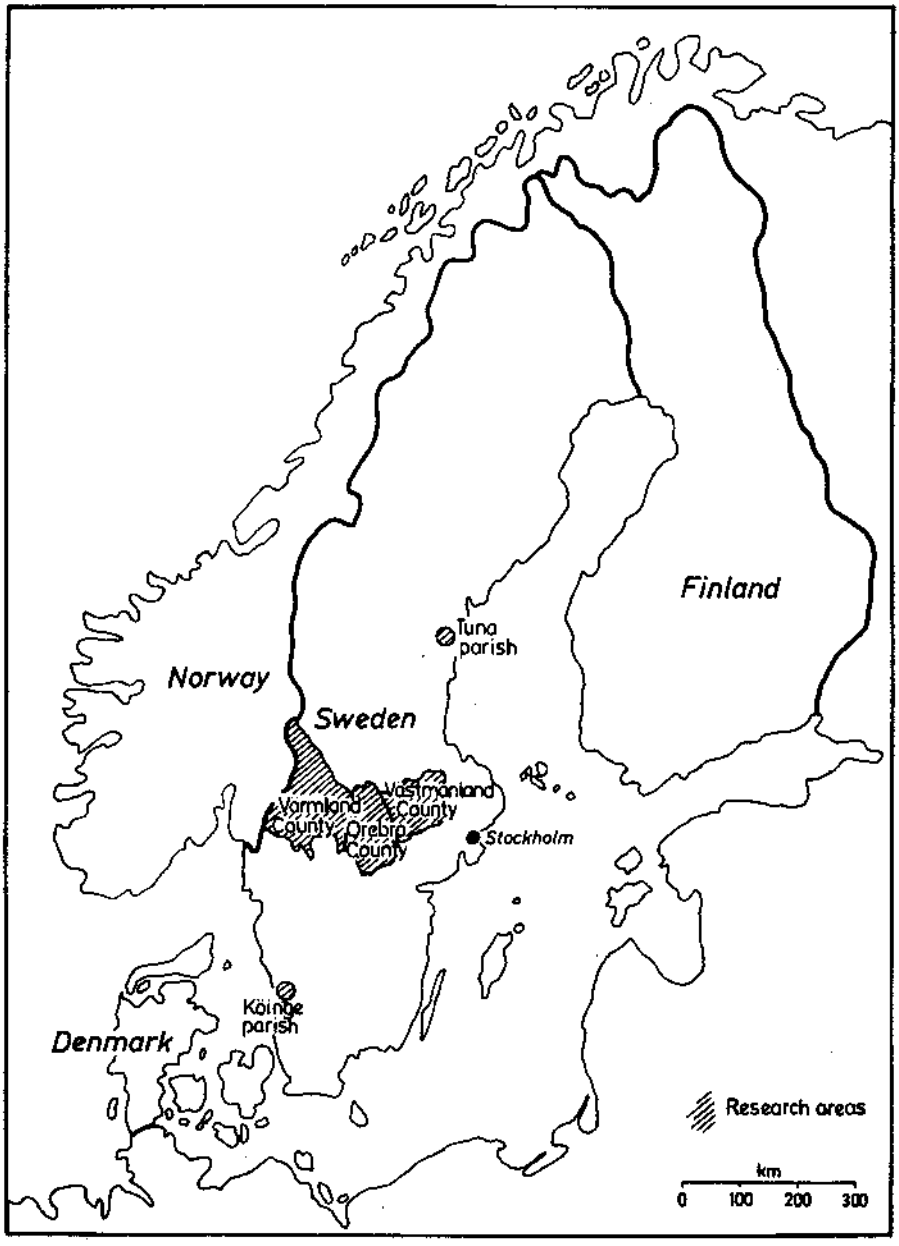


Figure 12 - Research Areas.

Appendix

Commentary on the Variables and Classification for AID-analysis.

The subdivision of parish classes into agricultural, mixed, and industrial parishes is based on the same system used by G. Sundbärg in the *Report of the Swedish Commission on Emigration*, where the basic starting point is the tax value of property other than agricultural. This parish classification has been further developed by D. S. Thomas and B. Carlsson, who have primarily utilized economic data on parish trade and industry but who have also taken into account the population mobility in the respective areas.[23] The subdivision of parishes in the county of Örebro has also been complemented with figures on the reported number of industrial workers.

The statistics for variables 2-3 have been taken from *Sweden's Official Statistics, The Census of Agriculture 1912-1920.*[24]

The measurement of the tradition of emigration has been based upon figures for the emigration intensity of the parishes during the 1850s and 1860s. The highest classification value for this variable was received by parishes where the emigration totalled 20 $^{0}/_{00}$ or higher of the mean population during any year of the 1850s. The other values are based on the emigration intensity of the parishes during the 1860s, according to the following classification. (Variable 9).

---

[23] Population Movements, 1941, p. 507; THOMAS, 1941, p. 212.

[24] The classification governing the proportion of cultivated acreage out of the total farming units has been drawn up in consultation with G. ENEQUIST, Department of Geography, Uppsala.

The classification and coding of parishes was performed as follows:

Variable 1	Code
Agricultural Parishes	0
Mixed Parishes	1
Industrial Parishes	2

Variables 2, 4 and 6

According to % figures	Code
0 - 5.0	0
5.1 - 10.0	1
10.1 - 15.0	2
15.1 - 20.0	3
20.1 - 25.0	4
30.1 - 40.0	6
40.1 - 50.0	7
50.1 - 60.0	8
60.1 - 70.0	9
70.1 - 85.0	&
85.1 - 100.0	—

Variable 3

According to absolute figures

Variable 5

According to the number of hectares	Code
0 - 2.000	0
2.001 - 5.000	1
5.001 - 10.000	2
10.001 - 25.000	3
25.001 - 50.000	4
50.001 - 100.000	5
100.001 - 250.000	6
250.001 - 500.000	7
500.001 - 1.000.000	8
1.000.001 - w	9

Variable 7

According to the number of kilometres	Code
0 - 10	0
11 - 20	1
21 - 30	2
31 - 40	3
41 - 50	4
51 - 60	5
61 - 70	6
71 - 80	7
81 - 90	8
91 - w	9

Variable 8

Mean values per year per mille of the mean population.

Variable 9

Highest figures for emigration during any year per mille of the mean population.

Number ‰	Code
1860s	
0 - 5.0	0
5.1 - 10.0	1
10.1 - 15.0	2
15.1 - 20.0	3
20.1 - 25.0	4
25.1 - 30.0	5
30.1 - 40.0	6
40.1 - 50.0	7
50.1 - w	8
1850s	
20 - w	9

References. Sources and Literature

The AID-method

ANDREW, F.M., MORGAN, J.N., SONQUIST, J.A., KLEM, L., *Multiple Classification Analysis*, Ann Arbor, Michigan 1973 (Andrew-Morgan-Sonquist-Klem 1973).

GAVATIN-AVÉN, A., *Signifikansproblemet i AID-analys*. Opubl. licentiatavhandling, Statistiska inst., Stockholms universitat 1973.

GULBERG, B., ODÉN B., *AID Analysis and Migration History*. Scandinavian Economic History Review XXIV: 1, 1976, pp. 1-27, 31 f. (Gullberg-Odén 1976).

SONQUIST, J.A., *Multivariate Model Building. The Validation of a Search Strategy*. University of Michigan, Ann Arbor, Michigan 1970. (Sonquist 1970).

SONQUIST, J.A., BAKER, E.L., MORGAN, J.N., *Searching for Structure*. The University of Michigan, Ann Arbor, Michigan 1971.

Population mobility and literacy

CARLSSON, S., *Emigration från Småland och Öland 1861-1930. Social och regional fördelning*. (HLFÅ 1966-1967). (Carlsson, 1966-1967).

CARLSSON, S., *Frikyrklighet och emigration*. Ett bidrag. (Kyrka, folk, stat. Festskrift till Sven Kjöllerström. Lund 1967). (Carlsson, 1967).

DE GEER, E., *Emigrationen i Västsverige i slutet av 1800- talet*. (Ymer 1959). (De Geer 1959).

EBBESON, U., *Emigration från en bruksbygd i Östergötland*. Utvandringen till Amerika från Risinge socken med Finspångs styckebruk. (Opubl. lic. avh., Hist. inst., Uppsala 1968). (Ebbeson 1968).

Emigrationsutredningen. Report of the Swedish Commission on Emigration. Stocholm 1908-1913. (EU).

GRAFF, H., *An Introduction to the Study of Literacy in Sweden*. The Canadian Social Project, Report number 5. The Ontario Institute for Studies in Education, Toronto 1973-1974 (Gratt 1974).

HEDMAN, I., och KINDENBERG, U., *Arbetsvandringar från Fryksdalen under senare delen av 1800-talet och början av 1900-talet*. (Opubl. uppsats, Institutet för folklivsforskning, Stockholm 1973). (Hedman-Kindenberg 1973).

HVIDT, K., *Flugten til Amerika eller drivekraefter i masseudvandringen fra Danmark 1868-1914*. Odense 1971. (Hvidt 1971).

JOHANSSON, E., *The history of literacy in Sweden in comparison with some other countries*. Educational Reports Nr 12, Dept of Education, Umeå University, 1977. (Johansson 1977).

JOHANSSON, E., *En studie med kvantitativa metoder av folkundervisningen i Bygdea socken 1845-1873*. (Opubl. doktorsavh., Pedagogiska inst. Umeå 1972). (Johansson 1972).

JOHANSSON, E., *Literacy in the Swedish Church Examination Registers.* Umeå 1973. (Johansson 1973).

Jordbruk och boskapsskötsel 1913-1920. Stockolm 1916-1923. (The Census of Agriculture 1913-1920).

LITHELL, U.-B., *Arbetsvandringar och befolkningsrörelser från Lekvattnets församling 1861-1914.* (Opubl. uppsats, Hist. inst., Uppsala 1971). (Lithell 1971).

LOCKRIDGE, K., *Literacy in Colonial New England*, New York 1974. (Lockridge 1974).

NOREEN, P., *Emigrationen från Sundals härad i Dalsland 1860-1895. Emigrationens orsaker och förlopp.* (Opubl. lic. avh., Hist. inst., Uppsala 1967). (Noreen 1967).

NORMAN, H., *Emigration från bygder med uppväxanda industrier i Örebro län 1861-1905. Med särskild hänsyn till Karlskoga och Kumla socknar.* (Opubl. lic. avh., Hist. inst., Uppsala 1967). (Norman 1967).

NORMAN, H., *Från Bergslagen till Nordamerika.* Studier i migrationsmönster, social rörlighet och demografisk struktur med utgångspunkt från Örebro län 1851-1915. Uppsala 1974. (Norman 1974).

Primärmaterialet från fabriker 1906-1915. Örebro län. (Primary Material from Factories).

Population Movements and Industrialization. Swedish Counties 1895-1930. By the Staff of the Institute for Social Sciences. Stockholm University. Vol. II. (Stockholm Economic Studies 10: 2 1941). (Population Movements 1941).

RONDAHL, B., *Emigration, folkomflyttning och säsongarbete i ett sågverksdistrikt i södra Hälsingland 1865-1910.* (Studia Historica Upsaliensia nr 40. Uppsala 1972). (Rondahl 1972).

ROSANDER, G., *Herrarbete. Dalfolkets säsongvisa arbetsvandringar i jämförande belysning.* (Skrifter utgivna genom landsmåls- och folkminnesarkivet i Uppsala. Ser. B: 13. Uppsala 1967). (Rosander 1967).

SCHOFIELD, R., *The Measurement of Literacy in Pre-Industrial England.* Past and Present, 56 (Schofield 1972).

Summariska redogörelser för folkmängden. (Condensed Population Reports).

TEDEBRAND, L.-G., *Västernorrland och Nordamerika 1875-1913. Utvandring och återinvandring.* (Studia Historica Upsaliensia nr 42. Uppsala 1972). (Tederbrand 1972).

THOMAS, D.S., *Social and Economic Aspects of Swedish Population Movements 1750-1933.* (Stockholm Economic Studies 10: 1. New York 1941). (Thomas 1941).

AKERMAN, S., *Intern befolkningsomflyttning och emigration.* (Emigrationen fra Norden indtil I. Verdenskrig. Rapporter til Det Nordiske Historikermöde i Köpenhavn 1971, 9-12 August. Köpenhamn 1971). (Akerman 1971).

AKERMAN, S., CASSEL, P.G. and JOHANSSON, E., *Background variables of Population Mobility: An Attempt at Automatic Interaction Detector Analysis.* A Preliminary Research Report (Scandinavian Economic History Review Vol. XXII: 1 1974, pp. 32-60). (Akerman-Cassel-Johansson 1974). (Swedish version with Appendix, 1971).