ACCIDENT TYPOLOGY

R-90-48

C.M. Gundy

Leidschendam, 1990
SWOV Institute for Road Safety Research, The Netherlands

CONTENTS

1. Introduction

2. Methods
2.1. Data
2.2. Analysis procedure
2.3. Provisos

3. Results: Auto-bicycle accidents
3.1. Description of the data
3.2. Description of the analysis
3.3. Cluster analysis
3.4. The clusters

4. Results: Auto-moped accidents
4.1. Description of the data
4.2. Description of the analysis
4.3. Cluster analysis
4.4. The clusters

5. Results: Auto-pedestrian accidents
5.1. Description of the data
5.2. Description of the analysis
5.3. Cluster analysis
5.4. The clusters

6. Results: Auto-auto accidents
6.1. Description of the data
6.2. Description of the analysis
6.3. Cluster analysis
6.4. The clysters

7. Results: Auto-object
7.1. Description of the data (single cars and another object)
7.2. Description of the analysis
7.3. Cluster analysis
7.4. The clusters

8. Results: Auto-No object accidents
8.1. Description of the data (single cars only)
8.2. Description of the analysis
8.3. Cluster analysis
8.4. The clusters

9. Discussions and conclusions
9.1. Procedural evaluation
9.2. Methodological evaluation
9.3. Substantial evaluation

Bibliography

Tables 1-6

Appendices I-III

INTRODUCTION

The original (in 1983) purpose of this study was quite straightforward. This author intended studying behaviour in traffic situations and needed a method for determining which behaviour in which traffic situations could be relevant for traffic safety. A number of methods were available, such as top-down theory formation, and an examination of empirical research literature, and it was, of course, the intention to utilize them. Nevertheless, it was felt that a primary input to this problem would involve the consideration of actual traffic accidents in the Netherlands. Therefore, in order to study behaviour in traffic safety relevant situations (instead of traffic behaviour in general), one should at least consider the circumstances during which traffic accidents occur and whether one could distinguish between different types of accidents. If it were possible to distinguish between accident types then it might be possible to identify behavioral scenarios for further research.

This is tantamount to asking whether there was a system for classifying accidents in general, and Dutch traffic accidents in particular. A quick perusal of the literature determined that there had been some theoretical (e.g., the work of Rasmussen and his colleagues) as well as empirical work (e.g., Treat et al., 1977) done in the area of accident classification, but the applicability of this work to the Dutch (traffic) situation had not been determined. Furthermore, the SWOV Institute for Road Safety Research had access to the national traffic accident database, and it was felt that examination of this data would be a preliminary and useful step.

Two problems became quickly apparent. In the first place, the database, while containing a great deal of information such as time, place, weather, number of injured, etc., was relatively impoverished with respect to behaviorally relevant data. In the second place, meaning in general and classification in particular is context-sensitive to a large extent. To be more explicit; with few exceptions individual traffic safety researchers are concerned with two types of accidents: those they are interested in and those that are used as a control. That is, depending upon one's field of study, accidents may be classified as drinking and driving accidents, young child mid-block dart-out accidents, overtaking accidents, failure to yield right-of-way accidents, etc. While these distinctions are useful, they are not necessarily conducive to generating an overarching and systematic typology, such as available in the physical and life sciences.

Slowly, then, this researcher's interest shifted from finding a quick and dirty method for generating research scenarios to understanding structural and semantic aspects of classification as applied to the traffic safety problem. While it is interesting to find a methodology for pigeonholing accidents, it is even more important to understand the (certainly disparate) ways in which traffic safety researchers, policy makers, police, road administrators, and road users' understand and classify traffic accidents, road characteristics, behaviour, and traffic situations.

The value of this last statement is slowly becoming more widely recognized in the traffic safety research world. Nevertheless, this report does not cover this topic in any detail (yet see Gundy, in preparation). Rather, it describes the results of an attempt to systematically pigeonhole accidents using a data analytic technique. (By this last point is implied that such a classifica-tion procedure will be semantically weak, with the exception of the seman-tics involved with the original data collection and coding.)

Despite it's semantic weakness (other shortcomings will be discussed in a later section), such a procedure and it's attending outcome could obtain useful consequences. For example, if the classification results are aesthetically pleasing (?), then they could be used for the originally intended purpose. If accepted on a wider scale, then the classification scheme could be used to identify and trace the development of different accident types over the years. In addition, the possible identification of heretofore unrecognized accident types might lead to new ideas concerning safety measures. On the other hand, if the results are not pleasing and/or are not widely acceptable, one is then lead to the question of identifying specific shortcomings which may then be amenable to correction and/or further research. Finally, regardless of the final evaluation, the results could function as a first seed for crystallizing a conceptual integration in traffic safety research and application.

2. METHODS

2.1. Data

As previously mentioned (see Gundy, 1989), the data set analyzed consisted of slightly more than 45,000 personal injury accidents registered by the police in the Netherlands during 1982. This data was collected, coded, and placed on a magnetic medium by the Road Accident Records Office (VOR), and was made available to the SWOV Institute for Road Safety Research. This file was then reformatted to a more suitable form for data analysis. (For the interested reader, CBS (1984) presents 90 pages of selected cross-tables for the entire 1982 accident file.)

It was decided that, in order to avoid trivial structural dependencies, it would be necessary to divide accidents into sub-groups prior to subsequent analysis. (By trivial structural dependencies, one can imagine discovering that trees are hardly ever charged with drinking and driving even though they may be frequently involved in alcohol accidents; babies hardly ever drive cars; parked cars are only rarely found implementing an overtaking manoeuvre; etc.). Noting that type of vehicle is an important determinant of traffic behaviour (i.e., it determines which traffic laws are applicable, where the vehicle is likely to be on the road, the age of the road user, etc.), it was decided to partition accidents on the basis of the type of vehicle(s) involved. Furthermore, to keep things simple, it was decided to initially investigate only those accidents involving less than three vehicles. Finally, since budgetary and practical limitations prevented analyses of all possible combinations of vehicle pairs, it was decided to begin analysis with only five groups of vehicle pairs. The final selection of these groups was determined by two factors:

1. absolute size of the group (the larger the better, for statistical reasons), and
2. at least one of the vehicles had to be an automobile.

The five resulting groups included car-bicycle, car-moped, car-pedestrian, car-car and single car collisions. These five mutually exclusive accident groups accounted for about 54% of all registered injury accidents in 1982.

The following variables were used, if applicable:

- time of day, day of the week, and month of the year in which the accident occurred;
- the sex and age of the road users' involved (but not their passengers), and whether there was any indication of alcohol use;
- the population of the local jurisdiction in which the accident occurred, whether the accident was inside or outside of built up areas, and the province in which the accident occurred;
- the applicable speed limit, the administrative unit responsible for the road section where the accident occurred, the presence of unusual and/or temporary road conditions, and whether the road section involved was a straight road section, a curve, or an intersection;
- the presence of street lighting, natural light conditions, weather and pavement conditions, and the type of road surface;
- the location on the road (or next to it) of each road user prior to the accident, their manoeuvres, the place on the vehicle where initial contact was made, and possible traffic violations and/or accidents 'causes' as determined by the police.

No use was made of the CBS manoeuvre variable in these analyses, since it seemed to be a summary of a number of other accident characteristics, and would bias the analysis as discussed in a following section. Nevertheless, it was cross-tabulated with the derived clusters for the purpose of illustration, and a description of the CBS manoeuvre codes is given in Appendix I.

2.2. Analysis procedure

The analysis procedure consisted of a number of steps.

First of all, the data for each conflict type (accident pair) was separately submitted to homogeneity analysis. This was done with the help of HOMALS, a computer program developed by the University of Leyden (Gifi, 1981). HOMALS, among other things, can be viewed as a principal component analysis of a transformed indicator matrix, and was used to derive numerical scores (which is required for the following steps) from otherwise mostly nominal and ordinal variables. HOMALS is notable for it's conceptual and algorithmic simplicity. I.e., anyone with computer software that can solve eigen problems can, with some pre- and post-processing, generate equivalent solutions. Another simplifying advantage of HOMALS is that it's solutions are nested (i.e., one obtains the same eigen vectors regardless of the chosen dimensionality).

Secondly, the by HOMALS derived object scores were, after rescaling such that the variance of each dimension was set equal to its eigenvalue, submitted to the SAS clustering procedure FASTCLUS to achieve a 'quick and dirty' reduction in the number of data points (SAS User's Guide, 1985). This only reason for this intermediate step, which reduced the number of points to about 7% of the original number, was to achieve an appreciable reduction in required computer time. This reduced number of data points was then submitted to a SAS hierarchical centroid clustering technique (SAS, 1985).
The choice of the centroid method itself was somewhat arbitrary, in the sense that budgetary limitations did not allow extensive testing and comparison of techniques. Nevertheless, some preliminary testing indicated that the method derived adequate results. Furthermore, the method has been found to be less sensitive to outliers, is not parametric, and is not biased towards clusters of certain sizes or dispersions (SAS, 1985; see also Everitt, 1980). Since this study was completely exploratory, and that problems with outliers were quite likely, it was felt that a relatively robust and non-biased method was indicated. (Other choices could, of course, be defended, as every technique could be shown to be inadequate for some specific data set.)
About 2% of the data points were culled in order to avoid clustering outliers. This technique assigned each original accident to it's own unique cluster. The actual choice of the number of clusters selected depended on the results of a number of statistical tests (i.e., the Cubic Clustering Criterion, Pseudo F, and Pseudo T**2: see SAS, 1985), and a requirement that there not be more than ten clusters per conflict type. This last criterium was chosen in order to avoid too much complexity. A final criterium required that each cluster include at least 5% of the total number of accidents in that conflict type or at least 100 observations. This was done to attempt to avoid descriptions of statistically unreliable clusters.

Third of all, the original HOMALS scores were used to predict the derived cluster numbers. This was done by means of discriminant analysis (see

SAS, 1985). This step was primarily done to assess the extent that the clusters overlapped each other, and to assess the relative loss of information due to the FASTCLUS step.

Finally, all of the original (recoded) variable values were crosstabulated with the derived cluster scores. While the use of the previous techniques or of another technique (such as log-linear analysis) would be adequate for this task, it was felt that cross-tables would be a form a presentation easily understandable by most researchers.

2.3. Provisos

The results of an analysis are obviously influenced by the characteristics of the units analyzed. These units are sampled from a population and the generalizability of analysis results is limited by the representativeness of the sample. Since it well known (e.g., Maas & Harris, 1984) that registered injury traffic accidents are a biased sample of all injury traffic accidents, accidents with less severe injuries being more likely to be excluded than accidents with severe injuries, it would be quite dangerous to generalize the results found here to traffic accidents in general. Futhermore, it would be reckless to generalize results to a subset of the analyzed accidents, such as fatal accidents, unless one is willing to assume that the only difference between a fatal accident and an injury accident is the outcome. (If one wanted to describe only fatal accidents, then one should analyze only fatal accidents. Due to the relatively small numbers, one would then after to aggregate over a number of years). In our view, unless one is willing to make a number of strong assumptions, the results of the following analyses are in principle only generalizable to Dutch accident files sampled in more or less the same way as the 1982 VOR file. Since the entire VOR injury accident file is routinely used and it's statistics are compared from year to year, it is felt that the assumption of stationary sampling bias and the use of the entire file (instead of some portion of it) are not completely wide of the mark. Of course, anyone who completely disagrees is welcome to implement these analyses on his or her sample of interest.

A second point of great consequence for analysis is the selection of variables used in the analysis. The primary limiting factor here is the present structure of police accident reports and the VOR coding protocol. The present police-VOR collection protocol is probably based on a number of theoretical, historical and pragmatical reasons, all of which influence the analysis and its interpretation. While it would not be very difficult to find many researchers who feel that the recorded accident data is clearly insufficient for extensive analysis (e.g., Oude Egberink et al., 1988), it is also true that one cannot analyze data that one doesn't have. It is felt that, even though the present data set may be inadequate for a number of reasons, pragmatic arguments (i.e., the data is already standardly analyzed as is) are sufficient to justify preliminary analysis.

A third point concerns the lack of standardization of accident characteristics with respect to the background population. While such an endeavour may have been possible for some variables, e.g., sex and age distributions of road users, it would have been completely out of the question for others, e.g., relative frequency of various vehicle manoeuvres. This problem has been solved, in this case, by simply declaring this aspect to be outside the scope of the present study.

A fourth point concerns the 'distortion' introduced by the present author. Of course recoding variables and interpreting results introduces possibilities for distortion, but this is an inescapable part of doing research. Even so, it was intended to make this form of distortion as transparent as possible.

Nevertheless, two forms of systematic bias were introduced. In the first case, a number of variables were a priori excluded from the analysis. For example, the possession of a driver's license and the nationality of the car were not analyzed, which one may either characterize as an oversight or as a blunder. Injury severity, characteristics of the injured, and the location of the vehicle after it had come to a still stand were also not analyzed. The argument used here was that this study was concerned with accident causes and traffic behaviour and not with accident outcomes, which is a clear bias of the author. Another variable, the direction of vehicle movement with respect to the direction of the road, was also excluded because it's coding was rather difficult to decipher. These choices could quite conceivably have some consequences for the results of the analysis.

Another type of selection was made as a consequence of analysis results.

The variable 'CBS manoeuvre' was explicitly excluded from the analysis because it was felt that it only served to summarize a number of other variables, and then in a rather inconsistent manner. It was felt that including this variable tended to create an artificial semblance of structure in the data set. A second correlational lynch-pin, street-lighting, was also adapted. It was originally coded as street-light absent, street-light present but not burning, and street-light present and burning. This coding correlated well with time of day, and atmospheric lighting conditions, and tended to identify a dimension which stated that if it was dark and a street light was present then it would be burning. While this may be generally true in the Netherlands, it was not felt to be very informative. Furthermore, it had, in one case, the consequence that almost all accidents were grouped into the same (very large) cluster. It was decided to recode the street-light variable to indicate whether a street light was present or not. This decision had the general consequence that the difference between day and night became less important in the resulting classifications than would otherwise be the case.

These last two decisions resulted in a discrepancy between the present description of car-pedestrian accidents and that mentioned in Gundy (1989). Namely, the original distinction between daytime young-child mid-block dart-outs, and nighttime adult mid-block dart-outs has been glossed over.

A final form of researcher bias may be mentioned: the selection the dimensions retained for further analysis. The twin criteria of simplicity and consistency were chosen as the primary considerations for retaining a dimension. Each analysis generated exactly six dimensions, all of which consistently contained more information than random vectors. Unfortunately, inclusion of the dimensions with smaller eigen values often lead to undesirable clustering results. This was due to the fact that 'weaker' dimensions often described small and compact groups of unusual accidents, and the clustering algorithm then distinguished between these small groups and the 'rest' of the accidents. Since it was not the primary intention of this study to classify outliers, and since limited time did not permit a systematic investigation of the interpretability and classification results of varying numbers of selected dimensions, it was decided to only retain the first three dimensions of the HOMALS solutions. This was occasionally troublesome because, now and then, the fourth dimension of a solution may have been equally or even more interesting than the third. With more time and experience with multiple analyses and data sets, more pleasing and universal guidelines for consistent inter-analysis choice of dimensionality would be possible.

For those discouraged by the aforementioned qualifications and short-comings, one should make the general remark that doing research implies making choices, and that making choices imposes limitations. It is felt that the choices made here are not only defensible, but also explicit and essentially pragmatic.

3. RESULTS: AUTO-BICYCLE ACCIDENTS

6333 accidents between cars and bicyclists were selected from the 1982 VOR accident file. The selection criteria required that 2 and only 2 'colliding' accident 'objects' were involved, both of which were nonparked vehicles. Of course, one of those vehicles was a bicycle and the other was an automobile.

In the following paragraph a short summary of the marginal frequencies of the accident characteristics will be given.

3.1. Description of the data

Collisions between cars and bicyclists are primarily harmful for the bicyclist. 181 of the 6333 (2.9%) registered accidents resulted in death for the bicyclist, 2066 (32.6%) required that (at least one) bicyclist be taken to the hospital, and 3831 (60.55%) resulted in some registered injury which did not require a trip to the hospital. None of the automobile occupants were killed, 12 (0.2%) were taken to the hospital, and some light injury was observed in 49 (0.7%) of these accidents. Only one person, i.e., the bicyclist, was injured or killed in 6202 (97.9%) in these accidents.

These accidents tend to occur primarily during weekdays (83%) mainly between 7 o'clock in the morning and 8 o'clock in the evening (93%). They seem to occur slightly more often during the late spring and fall months, and slightly less often during the first four months of the year. 81% occur during daylight, while 80% occur in the vicinity of non-lit street lights. Only about 3% occur with no street light in the vicinity. 80% occur during dry weather, and 13% occur during rain. Other extreme weather conditions are only rarely noted. The road surface, which consists of bitumen in 74% of the time and clinkers 25% of the time, is also dry 74% of the time and wet 25% of the time.

85% of these accidents occur inside built-up areas, where the maximum speed is 50 km/hour in 84% of the cases and 80 km/hour in 13% of the cases. 86% of these accidents occur on roads administered by the city government, the rest being evenly divided between provincial and federal authorities. 32% of these cases occur in cities with more than 100,000 inhabitants, 43% in cities with between 20,000 and 100,000 inhabitants, and the other 25% occur in municipalities with less than 20,000 citizens. 56% occur in the three most populous provinces of North Holland, South Holland, and North Brabant.

79% of the car drivers are men, and 31% of the drivers are 25 years of age or less. 56% are between 25 and 56 years of age. 96% have a drivers' license, and alcohol use was determined in 2.5% of the cases. Interestingly enough, 3% of the cases have a missing value for alcohol use. 1.5% of bicyclists had been known to have consumed alcohol, and in 3% of the cases there is also a missing value for alcohol use. 56% of the bicyclists are men, and 40% of the bicyclists are less than 18 years old. 25% are between 18 and 40 years of age, and 18% are between 40 and 65 years. The remaining 15% are 65 or older.

29% of the accidents occur on a straight road section, while 68% occur at an intersection. 8% occur in the vicinity of a pedestrian crossing, and 5% occur near an road exit. In 84% of the cases, there is no unusual traffic situation noted. The vehicle manoeuvres being executed are, in order of frequency,:

- both vehicles are driving straight ahead, yet cross each other at right angles at an intersection (28%);
- one vehicle is turning left and is struck from behind (12%);
- one vehicle is turning left at a crossing and is struck on the left side by the other vehicle on the other road (11%);
- one vehicle crosses the road (yet not at a crossing) and is struck by the other vehicle (6%);
- rear end collision (4%);
- both vehicles are driving in the same direction on the same road and they collide on their flanks (4%);

  etc.

93% of the drivers are on the right hand side of the road, either driving straight ahead (71%), turning left (14%) or right (7%), or are nearly standing still (5%). 66% of the bicyclists are on the right hand side of the road and 22% are on a bicycle path. 59% are riding straight ahead, 26% are turning left, and 6% are crossing the road. 41% of the bicyclists are struck from the left side, 32% from the front, and 21% from the right side. 6% are struck from behind. 39% of the drivers are struck directly in front, 23% on the right front bumper, 18% on the left front bumper, and 15% on either flank. 5% are struck from behind.

39% are the drivers are charged with some infraction of the traffic regulations. 22% are accused of giving no right of way, and 4% are accused of driving too far to the right. On the other hand, bicyclists are blamed in 68% of the cases. The main accusations are: not giving right of way (47%), riding through a stop sign or light (4%), not riding far enough to the right (4%), and suddenly crossing the road (3%).

3.2. Description of the analysis

Homogeneity analysis

6333 observations, each consisting of observations on 30 variables, were submitted to a HOMALS analysis after recoding. This recoding was done to reduce the number of infrequently used categories. (See the previous section for a short summary of the marginal frequency distribution of the variables).

The first six dimensions of the HOMALS analysis 'explained' respectively 12.6%, 9.7%, 8.5%, 8.1%, 7.7%, and 7.3% of the total variance. For the reasons mentioned previously, only the first three dimensions were used for further analysis.

Variables having a discrimination measure greater than 0.20 on at least one of the first three dimensions were:
- built-up area
- maximum speed
- road situation
- unusual road situation
- street lights
- road surface condition
- the driver's manoeuvre
- what the driver was blamed for
- the manoeuvre of the bicyclist
- what the bicyclist was blamed for
- the age of the bicyclist.

Variables which had a loading equal to or greater than 0.20 on at least one of the last three dimensions, which did not load on one of the first three include the location of the driver on the road and the location of the bicyclist on the road.

The object score plot for the first two dimensions revealed a central 'mountain range' with one large peak and perhaps one subsidiary one. The 'highlands' had a few isolated peaks of no general importance. There also appeared to be a small group of perhaps 50 outliers surrounded on three sides by the 'lowlands'. There appear to be no clearly discernible clusters.

3.3. <u>Cluster analysis</u>

The first three dimensions of the HOMALS results were selected for further analysis. They were then re-scaled such that the variance for each dimension was set equal to the percentage variance 'explained' by that dimension divided by the variance 'explained' by the first dimension.

The construction of clusters proceeded in a number of steps.

First of all, the matrix of (re-scaled) matrix of 6333 observations times 3 values per observation was submitted to the SAS procedure FASTCLUS, which reduced the 6333 observations to 450 clusters. This was a fixed and arbitrary percentage of the original number of observations and was only implemented in order to reduce the computational requirements of the following cluster procedure.

Secondly, these 450 clusters were then submitted to the SAS procedure CLUSTER, which used the centroid algorithm. 75 clusters with a total number of 128 observations were then trimmed away. Visual inspection of the cubic clustering criterion, the pseudo F, and the pseudo T squared criteria indicated that about 8 clusters would be acceptable.

Third of all, the 8 clusters were then selected and all clusters with less than 100 observations or less than 5% of the total number of observations were eliminated. This procedure resulted in 4 clusters with a total of 5837 observations. 496 observations (8% or the total) were scrapped. The largest cluster contained 3058 observations, the second contained 1707 observations, and the last two clusters contained respecttively 615 and 457 observations. Thus, of all objects finally clustered, respectively 52%, 29%, 11% and 8% were divided into these four clusters. This result does not differ dramatically from the clusters found in other accident vehicle pairs, and is not surprising given the two dimensional HOMALS object score plots, which indicated a central concentration of data points.

Finally, the derived cluster numbers were merged with their original data values and their HOMALS object scores. Visual inspections of object score plots indicated that clusters did not overlap in the original three dimensional HOMALS space. Discriminant analysis confirmed this impression. Objects were split into two groups, and the HOMALS scores were used to discriminate between the clusters. In the first group, the discriminant function weights were estimated, and the second group was used to confirm these weights. Unequal prior probabilities and variance-covariance matrices were assumed.

The analysis on the first group indicated that about 98% of the observations in the largest group were correctly clustered. About 90% of the second and third largest clusters were also correctly classified. Only about 81% of the smallest group was correctly classified. Very similar results were obtained for the analysis on the second, confirmatory, group. The poorer results for the smallest cluster reflect, among other things, the difficulty of estimating a variance-covariance matrix on the basis of a limited number of observations. These results indicated that the object clustering was adequate.

3.4. The clusters

The derived clusters were bound, as previously mentioned, to their original data values. These accident objects were cross-tabulated: cluster scores were crossed with all of the original variables. Inspection of these cross-tables will indicate the 'meaning' of each derived cluster and determine it's usefulness for the practitioner. It is to the interpretation of the found clusters that we will now turn.

The chi^2-value for each of the cross-tables for the analyzed variables in found in Table 1. (Also included are the values for the CBS manoeuvre and death and injury variables, which were not included in the analysis). As can be seen from this table, almost all interactions are significant, some chi^2-values being quite extreme. This result confirms our expectation that the HOMALS and Cluster analysis have successfully found groups of accidents objects which are discriminable on the basis of their original characteristics.

In principle, we could (and perhaps should) carefully examine each and every cross-table with a significant chi^2. As previously mentioned, this approach would cost a great deal of effort and may tend to obscure the characterization of the clusters. For that reason, we will attempt to describe the clusters in turn in terms of their salient characteristics. To ensure some degree of reliability, we will only describe those cluster characteristics which have a chi^2 reliable at the 5% level. (Possible exceptions will be explicitly noted in the text.) As such, it must be emphasized that the following descriptions

Nevertheless, to aid the reader while describing the following clusters, a selection of some of the relevant cross-tabulations is found in Appendix II. It should be emphasized that the marginal frequencies in these tables are not identical to those described in the previous section, due to the exclusion of a number of data points. Again, the characteristics mentioned for each cluster are only relevant with respect to deviations from expected values, and not with respect to relative frequencies.

Cluster AF1

Cluster AB1 is by far the largest cluster, having 3058 observations. These accidents bear no clear relation to the day of the week and no clear relation to month of the year even though they are over-represented during January. They are clearly under-represented between 9 o'clock in the evening and 9 o'clock in the morning, and somewhat over-represented during the afternoon between 12 and 5 o'clock.

They are clearly over-represented inside built-up areas, with a maximum speed of 50 km per hour and with the local authorities responsible for the road. There is no clear relation with the province, even though Limburg is under-represented, yet are over-represented in medium size communities with between 20,000 and 100,000 inhabitants and under-represented in communities with less than 10,000 inhabitants.

These accidents are over-represented during daylight conditions, and tend to occur less frequently during dusk or darkness. The small chance of a auto-bicycle accident occurring at a location without street lighting is even smaller than expected for this accident type. It is a bit less likely to be raining during one of these accidents, yet there is little relation between either the condition of the road surface or the type of road surface.

These accidents are somewhat more likely than expected to occur in the vicinity of a pedestrian or bicycle crossing and a bit less likely to occur in a curve. There is no relation with temporary road conditions.

The driver of the car is less likely than expected to have alcohol use detected, is of no special sex, and is less likely to be older than 65 years of age. Drivers between 30 and 40 are over-represented. Unfortunately, no clear age trend is detectible. The bicyclist, whose alcohol use is less likely to be categorized as 'unknown', is somewhat more likely than expected to be male, and is clearly more likely to be less than 14 years of age and less likely to be between 18 and 65. Bicyclists older than 65 are again over-represented.

The car driver seems to be driving more often than expected on the right hand of the road, is more likely to be normally driving straight ahead, and is more likely to be struck on the middle or right side of the front of his automobile. He is much more likely to be exonerated of any blame. The bicyclist is more likely to be coming from a road exit or entrance, or from the sidewalk, and is somewhat less likely to be on a bicycle path. He is more likely to be turning left or right or to be crossing the road, and is struck from behind more often than expected. The bicyclist is more likely to be blamed for the accident, primarily accused of 'suddenly crossing the road' or not lending right-of-way.

The combined CBS manoeuvre variable (which was not used during the analysis due to redundancies: see Appendix I) indicates that accidents manoeuvres 322, 400 type accidents with the exception of type 411, and type 511 are clearly over-represented. A 322 represents two vehicles driving in the same direction when the first vehicle (presumably in this case the bicyclist) turns left and is hit on the flank. A 511 represents two vehicles driving straight ahead on crossing roads or on an exit/ entrance road, yet intersect each other at right angles. The 400 series indicates two vehicles driving in the opposite direction on the same road (or crossing straight across that road(?)) with one or both vehicles making a turn. The exclusion of type 411 excludes the not uncommon situations where one vehicle makes a left-hand turn and is struck by a vehicle from the opposite direction.

This type of accident is not particularly deadly (for the bicyclist), yet there is a slightly higher chance that he or she will be taken to the hospital.

<u>Summary</u>

This day-time type of accident tends to occur at a pedestrian or bicycle crossing inside built-up areas, yet is not especially urban in nature. A somewhat middle-aged driver is likely to be just driving straight ahead, when either a very young or very old bicyclist suddenly crosses the road or turns into the path of the automobile.

<u>Cluster AB2</u>

Cluster AB2 is much smaller than the first cluster and consists of 1707 accidents.

These accidents tend to occur somewhat less during the weekend, and are yet very clearly a winter type of accidents, being over-represented between November and January and under-represented between April through September. They are clearly not afternoon accidents, being over-represented between 8 o'clock in the evening and 9 o'clock in the morning, and under-represented during the afternoon.

They are also more likely than expected to occur inside built-up areas, with a speed limit of 50 km per hour, and with local authorities being responsible for the road. They tend to occur much more often than expected in cities with more than 50,000 inhabitants and much less often than expected in cities with less than 20,000 inhabitants. North Holland (Amsterdam?) is over-represented, and a few less densely populated provinces, Friesland, Drenthe, and Gelderland, are under-represented.

Of course, they are more likely than expected to occur during dusk or dark, and are more probably than expected to be in the vicinity of a street light. It is more likely than expected to be raining, and the road surface is more likely to be wet. The road surface is also less likely to be made of bitumen and more likely to be made of 'clinkers'.

These accidents tend to occur much more often than expected at intersections or traffic circles, and much less likely on straight road sections. There is no connection with temporary road situations, while the road situation is more likely to be unremarkable.

The driver is clearly much more likely to have been drinking than can be expected, is of no exceptional sex, and is slightly more likely to be older than 65, and less likely to be in his thirties. Doubts about the alcohol use of the bicyclists are more likely to be registered, and women are clearly over-represented. The bicyclist is clearly more likely to be between 18 and 65, and less likely to be younger than 15 or older than 65.

The driver's location on the road is unremarkable, while there is a slight tendency to be more often on a road with multiple lanes. It is much more probable that he is turning left or right or accelerating from a standstill and is more likely to be struck on the left front of the vehicle, on ones of the flanks, or on the rear of the vehicle. He is additionally quite often blamed for failure to yield right-of-way, or driving incorrectly on a curve. The bicyclist, on the other hand, is more likely to be on a bicycle path, is much more likely to driving riding straight ahead, or in some cases, is braking or even standing still. He is much more likely to be hit in front and is very likely to be exonerated from any blame.

The CBS manoeuvre variable (see Appendix I) indicates that these accidents over-involve two vehicles travelling on the same road, either in the same or opposite direction, while one of them turns and is struck by the other vehicle. They also over-involve accidents where one vehicle is turning onto a road and is struck by a vehicle on that road.

This type of accident is clearly less deadly than expected, and there are also less hospital injured than is expected. Lightly injured bicyclists are over-represented.

Summary

This type of accident tends to occur during the winter months, during the evening or night, and during poor weather conditions. The accident location tends more often to be an intersection in an urban area. Alcohol seems to play a role with the driver and possibly with the cyclist, who is neither very young or elderly. The driver is either turning or accelerating from a stand still, when he strikes the bicyclist who is most likely to be innocently riding straight ahead on a bicycle path.

Cluster AB3

This group of accidents includes 615 cases.

The weekend is somewhat over-represented in this cluster, and is much less likely to be a winter type of accident. July and September are over-represented. These accidents have a tendency to occur more frequently in the afternoon, yet this tendency is statistically rather weak.

They are extremely likely to occur outside of built-up areas and on roads with a maximum speed of 80 km/hour. Provincial and federal authorities are more likely to be responsible for the road. Cities with a population of less than 20,000 inhabitants are very clearly over-represented, and cities with 50,000 or more inhabitants are under-represented. 'Rural' provinces, such as Friesland, Drenthe, Gelderland, Zeeland, North Brabant, and Limburg, are clearly over-represented, and more 'urban' provinces, i.e., North and South Holland, are clearly under-represented.

These accidents tend to occur somewhat less frequently during darkness, and street lighting is more likely to be absent. They tend to occur somewhat less often during rain and the road surface is more likely to be dry. The road surface is much more likely to consist of bitumen.

These accidents tend to occur more frequently on straight road sections or on curves, while intersections are under-represented. There is no relation with temporary road conditions, yet is somewhat more likely that a road entrance/exit or a pedestrian/bicycle crossing is present.

There seems little to be said about possible alcohol use by the driver nor by the bicyclist. One could make a case that the driver is a bit more likely to be a man, yet this relation is rather weak. There is no relation with driver age. One the other hand, the bicyclist is more likely to be male, and between the ages of 6 and 14, or older than 65. Bicyclists between 18 and 65 are registered relatively much less often.

The location of the car has no special descriptive value in this cluster, yet it is much more likely to just be driving straight ahead, without any unusual manoeuvres. It is more likely to be struck in the middle front of the vehicle and less likely in other places. The driver is quite likely to be absolved of any blame. The bicyclist is somewhat more likely to be on a road exit or entrance or on the 'other bicycle path' category, and more likely to be performing some manoeuvre, such as turning, crossing the road, changing lanes, etc. He is also more likely to be struck from behind or the right flank, and likely to be blamed for failure to yield right of way.

The CBS manoeuvre series (see Appendix I) indicates that a number of manoeuvres are found to be somewhat more likely in this cluster:
- both vehicles are driving on the same road, but in opposite directions, without one of the vehicles turning;
- both vehicles are driving on the same road and in the same direction The first vehicle is turning left and is hit on the flank by the following vehicle;
- both vehicles are on the same road, but travelling in opposite directions, while one of them is turning;
- one of the vehicles is turning left onto a road and is struck by a vehicle coming from the right on that road.

Accidents in this cluster are clearly more deadly than can be expected and there are more injured taken to the hospital.

Summary

This rather serious type of accident tends to be more of a summer, good weather, and daylight accident on a rural, high-speed, straight road section or curve. There is an increased likelihood that there is a road exit/entrance or a pedestrian/bicycle crossing in the neighbourhood. The either very young or very old, male bicyclist is most likely to be crossing the road, turning, or changing lanes when he is struck by the driver, who is 'innocently' just driving straight ahead.

Cluster AB4

This cluster is the smallest of the four and contains 457 observations.

It occurs on no special day of the week (relative to the marginal frequencies), and tends to be over-represented during the summer months (May to August) as opposed to the winter months (November to January). This cluster seems to distinguish between early morning (between 0 and 9 o'clock) and later morning (between 9 and 12 o'clock) accidents, the latter being over-represented. Other hours are neither over nor under-represented.

They are somewhat less likely to occur inside built-up areas, and less frequently on roads with a 80 km per hour speed limit. (Roads with other speed limits are neither under- nor over-represented). The road authorities are less likely to be provincial or federal. The situation is not very clear with respect to number of inhabitants: small towns with between 20 and 50 thousand inhabitants are somewhat under-represented, yet there is no clear trend. Neither is there any obvious relation with province.

These accidents are clearly more likely to occur during daylight hours rather than during dusk or dark. There is no relation with the presence of street lighting. It is more likely to be dry weather, and the road surface is more likely to be dry than expected. The road surface itself is more likely than expected to consist of 'clinkers' as opposed to bitumen.

Accidents in this cluster are much more likely than expected to occur on straight road sections (or perhaps curves) as opposed to intersections. They are clearly more likely than expected to occur by a road exit/ entrance or by a parking place, and less likely in a 'normal' situation. The relation with temporary conditions is not worth mentioning.

The driver of the car may have a slightly higher chance of having been drinking, yet this relation is weak. Neither the driver's sex nor age play any role in this cluster. The chance of the bicyclist having consumed alcohol is unremarkable, as is his/her sex. Bicyclists less than 14 years old are under-represented, and those between 31 and 65 seem to be over-represented.

Drivers on the left hand side of the road or by a road exit/entrance or by a parallel road are over-represented, while those driving on the right hand side of the road are under-represented. Either braking or some 'other' manoeuvre are over-represented, while driving straight ahead or turning left are under-represented. They seem more likely to be struck from behind or on the right flank, and less often in the front. These drivers seem to be much more likely to be blamed for the accidents, reasons being: driving too much to the right, not yielding right-of-way, following too closely, and 'other' violations. The bicyclist distinguishes himself in this cluster by being on the right hand side of the road as opposed to anywhere else, and is clearly just riding straight ahead. (There may be a heightened chance that he may being changing lanes or performing some 'other' manoeuvre.) He is more likely to be struck either in front or on the left side as opposed to other locations. He is more likely to be exonerated from any blame or may be accused of some violation 'other' than not yielding right-of-way, suddenly crossing, or ignoring traffic signs.

The CBS manoeuvre variable (see Appendix I) clearly indicates that the 100 and 200 series are over-represented. These series describe two vehicles travelling on the same road either in the same or opposite direction. Neither vehicle is turning, yet some manoeuvre, such as changing lanes or braking, may be indicated.

For this cluster, it is unclear whether it is less deadly than expected or not, yet there are obviously less injuries requiring being taken to the hospital, and clearly more apparently light injuries.

Summary

This accident also seems to be a summer, good weather accident during daylight conditions at a straight road section inside a built-up area. It does not seem to be an especially urban type of accident. A road exit/ entrance or a parking lot is more likely to be in the vicinity. The driver is more likely to be on the left hand side of the road, at a road exit/entrance or on a parallel road, and is apparently executing some 'unusual' manoeuvre such as braking, when he strikes the middle-aged cyclist who is innocently riding straight ahead on the right hand side of the road. The driver is often blamed for failure to yield right-of-way, following too closely, or driving too much to the right, while the cyclist is mainly exonerated, even though there is a chance that he is also executing some manoeuvre which may have led to the accident. It appears that either the bicyclist runs into a suddenly braking car or is side-swiped.

4. RESULTS: AUTO-MOPED ACCIDENTS

6263 accidents between cars and mopeds were selected from the 1982 VOR accident file. The selection criteria required that 2 and only 2 'colliding' accident 'objects' were involved, both of which were non-parked vehicles. Of course, one of those vehicles was a moped and the other was an automobile.

In the following paragraph a short summary of the marginal frequencies of the accident characteristics will be given.

4.1. Description of the data

Collisions between cars and mopeds are primarily harmful for the moped rider, even though they seem to be somewhat less serious than car-bicyclist collisions. 47 of the 6263 (0.8%) registered accidents resulted in death for the moped rider, 1739 (27.8%) required that (at least one) moped rider be taken to the hospital, and 4269 (68.1%) resulted in some registered injury which did not require a trip to the hospital. None of the automobile occupants were killed, 6 (0.1%) were taken to the hospital, and some light injury was observed in 50 (0.8%) of these accidents. Only one person was injured or killed in 5827 (93.0%) in these accidents. This is somewhat less often than in the case of car-bicyclist collisions, which may be due to the probability that moped accidents seem to involve passengers injuries more often.

These accidents also tend to occur primarily during weekdays (78%) mainly between 7 o'clock in the morning and 8 o'clock in the evening (90%). They seem to occur more often during the late spring and fall months, and less often during the winter months. 80% occur during daylight, while 79% occur in the vicinity of non-lit street lights. Only about 2% occur with no street light in the vicinity. 88% occur during dry weather, and 11% occur during rain. Other extreme weather conditions are only rarely noted. The road surface, which consists of bitumen in 65% of the time and clinkers 33% of the time, is also dry 76% of the time and wet 23% of the time.

84% of these accidents occur inside built-up areas, where the maximum speed is 50 km/hour in 83% of the cases and 80 km/hour in 14% of the cases. 84% of these accidents occur on roads administered by the city government, the rest being evenly divided between provincial and federal authorities. 31% of these cases occur in cities with more than 100,000 inhabitants, 43% in cities with between 20,000 and 100,000 inhabitants, and the other 27% occur in municipalities with less than 20,000 citizens. 55% occur in the three most populous provinces of North Holland, South Holland, and North Brabant.

80% of the car drivers are men, and 29% of the drivers are 25 years of age or less. 57% are between 25 and 56 years of age. 97% have a drivers' license, and alcohol use was determined in 2.5% of the cases. Interestingly enough, 2% of the cases have a missing value for alcohol use. 1.8% of moped users had been known to have consumed alcohol, and in 2% of the cases there is also a missing value for alcohol use. 80% of the moped users are men, which is more frequent than by bicyclists, and 56% of the moped users are less than 18 years old. 35% are between 18 and 40 years of age, and 7% are between 40 and 65 years. The remaining 2% are 65 or older. The age distribution for accident involved moped users is clearly different from that of bicyclists.

27% of the accidents occur on a straight road section, while 70% occur at an intersection. 4% occur in the vicinity of a pedestrian crossing, and 10% occur near a road exit. In 82% of the cases, there is no unusual traffic situation noted. The vehicle manoeuvres being executed are, in order of frequency,:

- both vehicles are driving straight ahead on two perpendicular intersecting roads (27%);
- both vehicles are driving on the same road, but in opposite directions, and one vehicle turns left (12%);
- the two vehicles are driving on two perpendicular roads; one vehicle turns left onto the other road and is struck by the second vehicle which comes from the left side (11%);
- both vehicles are travelling in the same direction on the same road. the front vehicle turns right and is struck on the flank by the following second vehicle (10%);

  etc.

82% of the drivers are on the right hand side of the road, either driving straight ahead (47%), turning left (22%) or right (19%), or are nearly standing still (6%). Turning accidents appear to be more frequent than in the case of bicycle-car accidents. 61% of the moped users are on the right hand side of the road and 26% are on a bicycle path. 80% are riding straight ahead, 11% are turning left, and 3% are turning right, which again reveals a different manoeuvre pattern than for bicyclists. 23% of the moped users are struck from the left side, 59% from the front, and 34% from the right side. Only 1% are struck from behind. 23% of the drivers' are struck directly in front, 22% on the right front bumper, 21% on the left front bumper, and 24% on either flank. 10% are struck from behind.

55% of the drivers are charged with some infraction of the traffic regulations. 54% are accused of not yielding right of way! Moped users are blamed in 53% of the cases. The main accusations are: not yielding right of way (32%), driving through a stop sign or light (3%), not riding far enough to the right (3%), and driving incorrectly through a curve (3%).

There are clear differences between car-bicycle and car-moped accidents, which mainly involve the type of manoeuvres and the type of collisions involved. Otherwise, the similarities, are somewhat surprising.

4.2. <u>Description of the analysis</u>

<u>Homogeneity analysis</u>
6263 observations, each consisting of observations on 30 variables, were submitted to a HOMALS analysis after recoding. This recoding was done to reduce the number of infrequently used categories. (See the previous section for a short summary of the marginal frequency distribution of the variables.)

The first six dimensions of the HOMALS analysis 'explained' respectively 11.8%, 10.0%, 8.8%, 7.9%, 7.2%, and 7.0% of the total variance. For the reasons mentioned previously, only the first three dimensions were used for further analysis.

Variables having a discrimination measure greater than 0.20 on at least one of the first three dimensions were:

- built-up area
- maximum speed
- road administrator
- road situation
- unusual road situation
- city size
- the driver's location on the road
- the driver's manoeuvre
- what the driver was blamed for
- the moped rider's location on the road
- the manoeuvre of the moped rider
- what the bicyclist was blamed for.

Variables which had a loading equal to or greater than 0.20 on at least one of the last three dimensions, which did not load on one of the first three include:
- time of day
- light situation
- condition of the road surface
- where the car was struck.

Inspection of the two dimensional object score scatter plot revealed a 'mountain range' with two equally high peaks which are less 'pointed' than in the car-bicyclist case. Both peaks gradually descend into two parallel and separate highlands, yet quickly become lowlands in other directions. It should not be difficult to obtain 2 or 3 clearly defined clusters.

It should be re-emphasized (see previous chapters) that the original analysis of car-moped accidents yielded one large cluster with more than 75% of the accidents, and for this reason, the street-light variable was recoded. This had the consequence of reducing the relative importance of the distinction between day and night, which only reappears in a higher and non-used dimension. (See above.)

4.3. Cluster analysis

The first three dimensions of the HOMALS results were selected for further analysis. They were then re-scaled such that the variance for each dimension was set equal to the percentage variance 'explained' by that dimension divided by the variance 'explained' by the first dimension.

The construction of clusters proceeded in a number of steps.

First of all, the matrix of (re-scaled) matrix of 6263 observations times 3 values per observation was submitted to the SAS procedure FASTCLUS, which reduced the 6263 observations to 450 clusters. This was a fixed and arbitrary percentage of the original number of observations and was only implemented in order to reduce the computational requirements of the following cluster procedure.

Secondly, these 450 clusters were then submitted to the SAS procedure CLUSTER, which used the centroid algorithm. 69 clusters with a total number of 127 observations were then trimmed away. Visual inspection of the cubic clustering criterion, the pseudo F, and the pseudo T squared criteria indicated that about 9 or 10 clusters would be acceptable.

Third of all, the 9 (or 10) clusters were then selected and all clusters with less than 100 observations or less than 5% of the total number of observations were eliminated. This procedure resulted in 5 clusters with a total of 6054 observations. 209 observations (3% of the total) were scrapped. The largest cluster contained 2241 observations, the second contained 2234 observations, and the last three clusters contained respectively 623, 596 and 360 observations. Thus, of all objects finally clustered, respectively 37%,37%,10%,10%, and 6% were divided into these five clusters. This result differs from the clusters found in other accident vehicle pairs in that there are two large equally sized clusters, instead of the more common finding of only one very large cluster. This result is more aesthetically pleasing.

Finally, the derived cluster numbers were merged with their original data values and their HOMALS object scores. Visual inspections of object score plots indicated that clusters did not overlap in the original three dimensional HOMALS space. Discriminant analysis confirmed this impression. Objects were split into two groups, and the HOMALS scores were used to discriminate between the clusters. In the first group, the discriminant function weights were estimated, and the second group was used to confirm these weights. Unequal prior probabilities and variance-covariance matrices were assumed.

Discriminant analysis revealed that between 92% and 97% of the accidents in each cluster were correctly classified. Very similar results were obtained for the analysis on the second, confirmatory, group. These results indicated that the object clustering was adequate.

4.4. The clusters

The derived clusters were bound, as previously mentioned, to their original data values. These accident objects were cross-tabulated: cluster scores were crossed with all of the original variables. Inspection of these cross-tables will indicate the 'meaning' of each derived cluster and determine it's usefulness for the practitioner. It is to the interpretation of the found clusters that we will now turn.

The chi^2-value for each of the cross-tables for the analyzed variables in found in Table 2. (Also included are the values for the CBS manoeuvre and death and injury variables, which were not included in the analysis).
As can be seen from this table, almost all interactions are significant, some chi^2-values being quite extreme. This result confirms our expectation that the HOMALS and Cluster analysis have successfully found groups of accident objects which are discriminable on the basis of their original characteristics.

In principle, we could (and perhaps should) carefully examine each and every cross-table with a significant chi^2. As previously mentioned, this approach would cost a great deal of effort and may tend to obscure the characterization of the clusters. For that reason, we will attempt to describe the clusters in turn in terms of their salient characteristics. To ensure some degree of reliability, we will only describe those cluster characteristics which have a chi^2 reliable at the 5% level. (Possible exceptions will be explicitly noted in the text.)

Nevertheless, to aid the reader while describing the following clusters, a selection of some of the relevant cross-tabulations is found in Appendix

II. It should be emphasized that the marginal frequencies in these tables are not identical to those described in the previous section, due to the exclusion of a number of data points. Again, the characteristics mentioned for each cluster are only relevant with respect to deviations from expected values, and not with respect to relative frequencies.

Cluster AM1

This cluster contains 2241 accidents, or about 37% of the total number of car-moped accidents classified.

The accidents are under-represented on Sunday, yet it is not clear when they may be over-represented. They are over-represented during November and December and are under-represented during May. (There is an indication that January could also be considered a winter month, and that the summer months could be grouped with May.) They are under-represented during the evening and night, primarily between 6 and 7 o'clock in the evening, between 9 and 10 in the evening, and between midnight and 7 o'clock in the morning.

They are clearly over-represented inside built-up areas, and on roads with a maximum speed of 50 km per hour. They are clearly under-represented on 80 km per hour roads. The roads on which the accidents occur are somewhat less likely to be administered by provincial or railroad authorities. The provinces of Friesland, Drenthe, and Zeeland are under-represented, and South Holland is over-represented. It would seem that Gelderland and Flevoland could be grouped with the former provinces, and Utrecht with the latter. Somewhat surprisingly, North Holland is neither under nor over-represented in this cluster. This type of accidents tends to occur in medium and large cities with a population of 50,000 or more and tends to occur less often in small towns with less than 20,000 inhabitants.

These accidents have a weak tendency to occur a bit more often during dusk, and usually occur in the vicinity of a street light. They don't have any tendency to occur in certain types of weather, nor is the road surface conditions of any importance. The type of road surface doesn't seem to play any role here either.

These accidents occur more often at intersections than could be expected, and are less likely to occur in the vicinity of a parking lot or a (pedestrian) crossing. There is no relation with temporary road conditions.

This cluster has no clear relation with the driver's (possible) alcohol use. The driver's sex plays no role, yet he/she is more likely to be 50 years or older and less likely to be 26 to 30 years of age. The moped driver is less likely to have alcohol use detected, is somewhat more likely to be a woman, and is less likely to be 16 years old or less or older than 65. Drivers of 19 (through) 30 to 40 years of age are over-represented.

The driver is more likely to be on the right hand side of a multiple lane road, and is less likely to be on the left side of the road, by an exit/entrance or by 'other' situations. He is more likely to be turning left or right, making a U-turn, or accelerating out of a stand still. All other manoeuvres are less likely. He is most likely to be struck on the right front, flank, or rear, and less likely to be struck elsewhere. He is quite likely to be blamed for not yielding the right-of-way or for neglecting a stop sign or light. He is only rarely exonerated for any wrong-doing. The moped rider, on the other hand, is more likely to be on a bicycle path and less likely to be elsewhere. He/she is most likely to be just driving straight ahead, without any special manoeuvre, and is more likely to be struck head-on. The moped is more likely to be exonerated from any blame, but if he is blamed, he is more likely to be blamed for an incorrect passing manoeuvre.

The CBS manoeuvre variable (see Appendix I) indicates that accidents types 311, 312, 411, and 421 are over-represented. In the first two cases, this indicates an accident where one vehicle is turning right and is struck either on the flank or from behind by a following vehicle. The third and fourth type refer to accidents where two vehicles are driving in opposite directions on the same road and one vehicle is either turning left or right and is struck by the other vehicle coming from the opposite direction.

This type of accident is relatively less deadly for the moped rider, and is less likely to cause an injury requiring being taken to the hospital. Lighter injuries are over-represented.

Summary
This type of accident seems to be generally describable as a daytime, (winter), urban-intersection accident, with an older driver either turning or accelerating out of a stand still, when he strikes an older female moped rider who is riding straight ahead on a bicycle path.

Cluster AM2

This cluster is approximately the same size as the previous one, containing 2234 accidents.

These accidents are more likely to occur on Sunday, and have absolutely no relation to the month of the year. They are clearly more likely to occur between 6 and 11 o'clock in the evening, and between midnight and 7 in the morning. The morning and afternoon rush hours may be under-represented.

These accidents tend to occur more frequently inside built-up areas, and on roads with a 50 km per hour speed limit. The city is more likely than other authorities to be the road administrator. North Holland and only North Holland (see the previous cluster) is over-represented in this case, and only Utrecht is under-represented. This type of accidents tends to be over-represented in larger cities with more than 100,000 inhabitants. Very small towns with less than 10,000 inhabitants are under-represented.

These accidents tend to occur less often than expected during dusk, and there is a weak tendency to occur more often during darkness. They tend to occur more often in the vicinity of a street light. Neither the weather nor the condition of the road surface play any sort of role in this type of accident. The road itself is more likely to consist of 'clinkers'.

Accidents on a normal intersection (as distinguished from a T or Y intersection) or a curve are over-represented. It is less likely that the accident occurs in some unusual situation. Temporary abnormalities play no role.

Neither the drivers' alcohol use nor sex have any unusual role in this cluster. Nevertheless, the driver has a lessened chance of being 50 or older, and a heightened chance of being between 26 and 30 years of age. (See previous cluster AM1). The moped rider is more likely to be male, 16 years of age or less, or older than 65. Intermediate ages tend to be under-represented even though not every age group reaches significance. Alcohol use by the moped rider plays no role in this cluster.

The driver tends to be driving more often on the right hand side of the road (without multiple lanes) and is more likely to be just driving straight ahead. He is more likely to be struck on the middle or left front of the vehicle or on the left flank. Other contact places are under-represented. He is much more likely to be exonerated from all blame, or may be blamed for some 'other' violation. The moped rider is most likely to be riding on the right or left side of the road (without multiple lanes), and is more likely to be braking, turning left or right, or crossing the road. He is more likely than expected to be struck on either flank, and is more likely to be blamed for not yielding the right-of-way, incorrectly driving through a curve, neglecting a stop sign or light, or some 'other' violation.

With reference to the CBS manoeuvre variable (see Appendix I), this cluster is clearly the opposite of the previously mentioned cluster: every manoeuvre over-represented here is under-represented there, and vice versa. Accident types 211, 511 (and the rest of the 500 series), 621, and 641 are all over-represented. The first refers to a frontal collision between two vehicles travelling in the opposite direction on the same road, where neither one is turning. The second refers to two vehicles crossing each other path at right angles while driving through an intersection, where neither vehicle is turning. The last two types refer to an accident wherein one vehicle is turning onto a road and is struck by a vehicle travelling on that road and coming from the direction into which the turn is being made.

This accident cluster is not particularly deadly, yet there is a larger chance that the moped rider will be taken to the hospital.

Summary

This type of accident is more of a night-time, urban intersection type of accident, where the driver is mainly just driving straight ahead when he strikes a very young or old moped rider who crosses his path and fails to yield right of way.

Cluster AM3

This cluster is appreciably smaller than the previous two and contains 623 observations.

This type of accident is over-represented on Sundays, and underrepresented on Thursdays. These accidents are more likely to occur during the summer months (June till August) and less likely during the winter. They have a tendency to occur somewhat more frequently than expected between 11 in the evening and 7 in the morning. (It is not clear when they do not occur.)

This type of accident is clearly over-represented outside built-up areas, and then on roads with a 80 km per hour speed limit. The provincial or federal authorities are more likely to be responsible for the road. Friesland, Overijssel, Gelderland, and Zeeland show a statistically significant over-representation, and North and South Holland are under-represented. As could be expected small towns with less than 20,000 inhabitants are over-represented, and larger towns with 50,000 or more inhabitants are under-represented.

Light conditions play no role here, yet it is more likely (than expected) that there is no street light in the vicinity. It is somewhat less likely to be raining, and the road surface is more likely to be dry. The road surface itself is most likely to consist of bitumen.

These accidents tend to be over-represented on a straight road section or a curve and under-represented on intersections. Unusual temporary conditions play no role here, yet the probability is higher that these accidents occur by a road crossing or an exit/entrance.

The driver of the automobile is more likely to have consumed some alcohol, this cluster apparently representing the alcohol cluster for car-moped accidents (of which there is no great number). The driver is more likely to be male, but of no distinctive age. The moped rider is also somewhat more likely to have been drinking, but is of no special sex. In any case, he is more likely to be 15 or younger, or older than 65.

The car is more likely to be on the right hand side of the road (without multiple lanes), and just driving straight ahead. Most likely, he will be struck directly head-on. It is most likely than the driver will be exonerated, although there is a heightened chance that he'll be ticketed for driving incorrectly through a curve. On the other hand, the moped rider is more likely to be riding on the left side of the road, on some 'other' type of bicycle path, to be on a road exit/entrance, a parallel road, or somewhere else (?). He is more likely to be turning, crossing the road, or doing something else (?), and is more likely to be struck on either flank or from behind. He is also more likely to be blamed for not yielding the right-of-way, riding incorrectly through a curve, or doing something 'else'.

A number of CBS manoeuvres are over-represented.

This type of accident appears to be quite serious, more likely (than expected) causing death or a serious injury. Light injuries are relatively infrequent.

<u>Summary</u>

This relatively serious type of summer accident tends to occur on straight sections or curves on a rural 80 km per hour road, when a male driver (who may have been drinking), driving straight ahead, drives straight into a very young or old moped rider who is either turning or crossing the road. 'Unpredictable' moped manoeuvres and high speeds may play an important role in this type of accident.

Cluster AM4

This cluster is about the same size as the previous containing 596 accidents.

There is no clear relation with day of the week or month of the year. However, the early morning hours between midnight and 7 are under-represented, and the hours between 2 and 3 (and possibly 4) in the afternoon are over-represented.

This type of accident occurs primarily inside built-up areas, with a maximum speed of 50 km per hour also being over-represented. City authorities are most likely to be responsible for the road on which the accident occurs. Gelderland is under- and North Brabant is over-represented (?). There is no relation with the population of the city in which the accidents occurs.

These accidents tend to be over-represented during daylight conditions and are under-represented during conditions of darkness. They are also less likely to occur with no street lighting in the vicinity, even though this chance is quite small to begin with. There is no special relation with weather conditions, nor the condition of the road surface. The road surface itself is more likely to consist of 'clinkers' and less likely to consist of bitumen.

These accidents are clearly over-represented on straight road sections and only rarely occur at intersections. They are also more likely to occur in the vicinity of a road exit/entrance, a parking lot, or some 'other unusual' road feature. There is no relation with temporary situational characteristics.

The driver is more likely to have his possible alcohol use registered as 'unknown' (?), and is of no special sex. He is less likely to be 20 or 24 years of age. Neither the moped rider's registered alcohol use nor his sex plays a role in this cluster. However, he is less likely to be older than 65 years.

The driver is more likely than expected to be on the left side of the road, on a road exit/entrance, or 'someplace else'. He is more likely to be braking, standing still, making a U-turn, or doing 'something else'. He is more likely to be struck from behind or on the left flank, and less likely to be struck in the middle or right front of his vehicle, or on his right flank. He is more likely to be charged with not yielding the right-of-way, merging incorrectly, or some 'other' infraction. The moped rider is less likely to be on a bicycle path, and more likely to be on the right side of the road or somewhere 'else'. He is more likely to be just driving straight ahead or changing lanes, and less likely to be turning. He is more likely to be struck frontally, and less likely to be struck on the left flank. The moped rider is less likely to be charged with a right-of-way infraction, and is either more likely to be exonerated or charged with not keeping sufficient distance, a passing violation, or some 'other' violation.

The CBS manoeuvre variable reveals a grab-bag of manoeuvres: The entire 100 series, the 200 series with the exception of 211, the 300 series with the exception of 312 and 322, and the 500 series with the exception of 511. The 100 series represents collisions between two vehicles travelling in the same direction on the same road. Collisions may be bumper to bumper, during a passing manoeuvre, a change of lanes, or during a merging merging manoeuvres. The 200 series refers to two vehicles travelling in opposite directions on the same road. The collision occurs during a lane change or a merging manoeuvre. The 300 series refers to collisions between two vehicles travelling in the same direction on the same road, while one or both of the vehicles is turning. In this cluster, it seems less likely that the lead vehicle would be struck on the right flank. The 500 series refers to two vehicles crossing each others path either at an intersection or a road exit/entrance. Neither vehicle is turning, yet one of them is either accelerating or braking.

This type of accident tends to be somewhat less serious, being less likely to require a trip to the hospital.

<u>Summary</u>

This daytime straight road section (or road exit/entrance or parking lot) type of accident is more likely to occur inside a built-up area, yet doesn't seem to be a typically 'urban' type of accident. The driver may be braking, making a U-turn, or some type of manoeuvre other than driving straight ahead or turning, when he hits the moped rider, who (while not on a bicycle path) is driving straight ahead or changing lanes. This accident type seems to include a great variety of manoeuvres, and may indicate a complex group of behaviors on low-volume roads where there is uncertainty about expected manoeuvres.

Cluster AM5

This cluster is the smallest in this conflict partner group, and consists of 360 observations.

It appears less likely to occur during the weekend, and somewhat less likely to occur during the evening between 8 and 10 o'clock or during the afternoon between 1 and 2. It does appear to be over-represented between 4 and 5 in the afternoon. The month of year plays no special role.

These accidents are clearly more likely to occur outside built-up areas on roads with a maximum speed limit of 70 or 80 km per hour. The road autorities are more likely to be provincial or federal. The provinces of North and South Holland and Utrecht are under-represented, and Gelderland and Zeeland (?) are over-represented. Small towns with less than 20,000 inhabitants are over-represented.

Lighting conditions play no apparent role, and neither does the presence of street lighting. Weather conditions play no role, and the condition of the road surface is unremarkable. The road surface itself is more likely to consist of bitumen or some 'other' material.

These accidents are over-represented at T or Y junctions as opposed to 'normal' intersections which are actually under-represented. Road exits/ entrances also tend to be over-represented. Temporary conditions are not relevant here.

It is a bit less likely that the possible alcohol use of the driver is 'unknown', and his sex and age is unremarkable, except that he is more likely to be older than 65. The moped riders' alcohol use is unremarkable, as is his sex. Nevertheless, he is clearly less likely to be 17 or younger, and more likely to be 18 or between 21 and 30 years of age. (The age categories of 19 and 20 years yield unclear results.)

The driver is more likely than expected to be on the right side of a multiple lane road, or by a road exit/entrance, and is most likely turning left or right, or accelerating from a stand still. He is also more likely to be struck on the right front side, and is much more likely than expected to be blamed for failure to yield right of way. The moped rider, on the other hand, is clearly more likely to be on a bicycle path (or on a parallel road), and is clearly doing nothing else than just riding straight ahead. He is more likely to be struck frontally as opposed to somewhere else, and is clearly exonerated from all blame.

The CBS manoeuvre variable (see Appendix I) reveals an over-representation of cases 312, 421, and 621. The first case refers to an accident where two vehicles are travelling in the same direction on the same road, and the first vehicle turns right and is struck on the right side by the following vehicle. The second case refers to two vehicles riding in opposite directions on the same road while one vehicle makes a right hand turn and is struck by the vehicle coming from the opposite direction. The third case refers to two vehicles driving on two intersecting roads (or a road exit/entrance) where one vehicle turns right and is struck by a vehicle coming from that direction.

This type of accident doesn't appear to be any more or less serious than average.

Summary

This type of accident tends to happen on high-speed rural roads at a T or Y intersection when the driver, turning onto the other road, fails to yield right-of-way and strikes the (somewhat older) moped rider, who is riding straight ahead on a bicycle path.

5. RESULTS: AUTO-PEDESTRIAN ACCIDENTS

2900 accidents between cars and pedestrians were selected from the 1982 VOR accident file. The selection criteria required that 2 and only 2 'colliding' accident 'objects' were involved, both of which were non-parked vehicles. Of course, one of those vehicles was pedestrian and the other was an automobile.

In the following paragraph a short summary of the marginal frequencies of the accident characteristics will be given.

5.1. Description of the data

Collisions between cars and pedestrian are apparently only harmful for the pedestrian, and are relatively quite serious in comparison to the first two types of conflict partner pairs mentioned previously. 155 of the 2900 (5.3%) registered accidents resulted in death for the pedestrian, 1223 (42.2%) required that the pedestrian be taken to the hospital, and (only!) 1395 (48.1%) resulted in some registered injury which did not require a trip to the hospital. None of the automobile occupants were killed, none (0.0%) were taken to the hospital, and some light injury was observed in 5 (0.2%) of these accidents. Only one person, i.e., the pedestrian, was injured or killed in almost all of these accidents.

These accidents also tend to occur primarily during weekdays (77%), mainly between 7 o'clock in the morning and 8 o'clock in the evening (86%). The winter months, particularly between November and March, seem to be problematical and there seems to be a light reprieve in the number of accidents during July and August. 74% occur during daylight, and 74% occur in the vicinity of non-lit street lights. Only about 3% occur with no street light in the vicinity. 86% occur during dry weather, and 12% occur during rain. Other extreme weather conditions are only rarely noted. The road surface, which consists of bitumen in 70% of the time and clinkers 28% of the time, is also dry 74% of the time and wet 24% of the time.

91% of these accidents occur inside built-up areas, where the maximum speed is 50 km/hour in 90% of the cases and 80 km/hour in 8% of the cases. 91% of these accidents occur on roads administered by the city government, the rest being evenly divided between provincial and federal authorities. 49% of these cases occur in cities with more than 100,000 inhabitants, 32% in cities with between 20,000 and 100,000 inhabitants, and the other 19% occur in municipalities with less than 20,000 citizens. 66% occur in the three most populous provinces of North Holland, South Holland, and North Brabant. This type of accident is clearly of a more urban variety than those involving bicycles or mopeds.

80% of the car drivers are men, and 30% of the drivers are 25 years of age or less. 57% are between 25 and 56 years of age. 95% have a drivers' license, while about 1% had no valid drivers' license. Alcohol use was determined in 3.7% of these cases. Interestingly enough, 4% of the cases have a missing value for alcohol use, which is more frequent than seen in the cases till now. 4.1% of pedestrians had been known to have consumed alcohol, and in 10% (!!) of the cases there is also a missing value for alcohol use. 60% of the pedestrians are men, and 54% of the pedestrians are less than 18 years old. 17% are between 18 and 40 years of age, and 13% are between 40 and 65 years. The remaining 16% are 65 or older. The age distribution for accident involved pedestrians is clearly different from other types of accidents involving 'slow traffic', younger people being clearly involved more often.

75% (!!) of the accidents occur on a straight road section, while only 24% occur at an intersection. 17% occur in the vicinity of a (pedestrian) crossing, 2% occur near a road exit, 3% by a bus or trolley stop, and 2% by a parking lot. In 74% of the cases, there is no unusual traffic situation noted. The vehicle manoeuvres being executed are, in order of frequency,:

- the pedestrian 'suddenly' crosses the road and is struck by a car (40%);
- the pedestrian crosses the road after emerging from behind an object and is struck (23%);
- the pedestrian is crossing the road on a pedestrian-crossing or zebra (16%);
- the pedestrian is walking along the road and is struck (5%);
- the pedestrian crosses the road in 'some other' way (4%);
- the pedestrian is in the vicinity of a bus or trolley stop when he is struck (3%); etc.

94% of the drivers are on the right hand side of the road (which may have multiple lanes), either driving straight ahead (85%), turning left (4%) or right (2%). Turning accidents are apparently somewhat rare. 62% of the pedestrians depart from the sidewalk, 9% are on the road itself, and 12% are using a pedestrian crossing. 88% are crossing the road, and 7% are apparently walking along the road. 52% of the drivers are struck directly in front, 23% on the right front bumper, 16% on the left front bumper, and 7% on either flank. 2% are struck from behind.

Only 22% of the drivers are charged with some infraction of the traffic regulations. 11% are accused of not yielding right of way. Pedestrians are blamed in 85% (!) of the cases. The main accusations are: being careless while crossing the road (72%), being careless while walking along the road (4%), and ignoring a stop light or some other traffic sign (4%).

5.2. Description of the analysis

Homogeneity analysis

2900 observations, each consisting of observations on 29 variables, were submitted to a HOMALS analysis after recoding. This recoding was done to reduce the number of infrequently used categories. (See the previous section for a short summary of the marginal frequency distribution of the variables.)

The first six dimensions of the HOMALS analysis 'explained' respectively 15.0%, 13.8%, 9.6%, 8.6%, 7.2%, and 6.8% of the total variance. For the reasons mentioned previously, only the first three dimensions were used for further analysis.

Variables having a discrimination measure greater than 0.20 on at least one of the first three dimensions were:
- time of day;
- built-up area
- maximum speed limit
- road administrator
- road situation
- unusual road characteristics

- presence street lights;
- city size;
- the driver's manoeuvre;
- whether the driver was ticketed, and for which traffic violation;
- where the pedestrian started from;
- the pedestrian's manoeuvre;
- whether the pedestrian was blamed by the police and the reason for that;
- the age of the pedestrian.

Variables which had a loading equal to or greater than 0.20 on at least one of the last three dimensions, which did not load on one of the first three are: the light situation and the condition of the road surface;

Inspection of the two dimensional object score scatter plot revealed one central and very steep 'peak' with some gently rolling 'foothills'. It would appear that, in at least two dimensions, one large cluster should dominate.

It should be re-emphasized (see previous chapters) that the original analysis of car-moped accidents yielded one large cluster with more than 75% of the accidents, and for this reason, the street-light variable was recoded. This had the consequence of reducing the relative importance of the distinction between day and night, which only reappears in a higher and non-used dimension. (See above.) This results, in this case, to a loss of discrimination between daytime and nighttime mid-block dart-outs. This chapter could be compared with the preliminary results reported by Gundy (1989).

5.3. <u>Cluster analysis</u>

The first three dimensions of the HOMALS results were selected for further analysis. They were then re-scaled such that the variance for each dimension was set equal to the percentage variance 'explained' by that dimension divided by the variance 'explained' by the first dimension.

The construction of clusters proceeded in a number of steps.

First of all, the matrix of (re-scaled) matrix of 2900 observations times 3 values per observation was submitted to the SAS procedure FASTCLUS, which reduced the 2900 observations to 200 clusters. This was a fixed and arbitrary percentage of the original number of observations and was only implemented in order to reduce the computational requirements of the following cluster procedure.

Secondly, these 200 clusters were then submitted to the SAS procedure CLUSTER, which used the centroid algorithm. 39 clusters with a total number of 67 observations were then trimmed away. Visual inspection of the cubic clustering criterion, the pseudo F, and the pseudo T squared criteria indicated that about 5 clusters would be acceptable.

Third of all, the 5 clusters were then selected and all clusters with less than 100 observations or less than 5% of the total number of observations were eliminated. This procedure resulted in 4 clusters with a total of 2689 observations. Due to a programming oversight, 100 observations were incorrectly eliminated from further analysis. A total of 211 observations were eliminated, which is 7% of the total number of observations. The largest cluster contained 1933 observations, the second contained 450 observations, and the last two clusters contained respectively 154 and 152 observations. Thus, of all objects finally clustered, respectively 72%, 17%, 6%, and 6% were divided into these four clusters. This result is quite unpleasing in the sense that it doesn't seem to make very much sense to do a cluster analysis, only to group (almost) all observations into one large cluster. As an exercise, the interested reader could compare the results in the following section with those reported by Gundy (1989), which were derived from the same data set.

Finally, the derived cluster numbers were merged with their original data values and their HOMALS object scores. Visual inspections of object score plots indicated that clusters did not overlap in the original three dimensional HOMALS space. Discriminant analysis confirmed this impression. Objects were split into two groups, and the HOMALS scores were used to discriminate between the clusters. In the first group, the discriminant function weights were estimated, and the second group was used to confirm these weights. Unequal prior probabilities and variance-covariance matrices were assumed.

Discriminant analysis revealed that between 96% and 99% of the accidents in each cluster were correctly classified. Very similar results were obtained for the analysis on the second, confirmatory, group. These results indicated that the object clustering was quite adequate.

5.4. The clusters

The derived clusters were bound, as previously mentioned, to their original data values. These accident objects were cross-tabulated: cluster scores were crossed with all of the original variables. Inspection of these cross-tables will indicate the 'meaning' of each derived cluster and determine it's usefulness for the practitioner. It is to the interpretation of the found clusters that we will now turn.

The chi^2-value for each of the cross-tables for the analyzed variables is found in Table 3. (Also included are the values for the CBS manoeuvre and death and injury variables, which were not included in the analysis).
As can be seen from this table, almost all interactions are significant, some chi^2-values being quite extreme. This result confirms our expectation that the HOMALS and Cluster analysis have successfully found groups of accident objects which are discriminable on the basis of their original characteristics.

In principle, we could (and perhaps should) carefully examine each and every cross-table with a significant chi^2. As previously mentioned, this approach would cost a great deal of effort and may tend to obscure the characterization of the clusters. For that reason, we will attempt to describe the clusters in turn in terms of their salient characteristics. To ensure some degree of reliability, we will only describe those cluster characteristics which have a chi^2 reliable at the 5% level. (Possible exceptions will be explicitly noted in the text.)

Nevertheless, to aid the reader while describing the following clusters, a selection of some of the relevant cross-tabulations is found in Appendix II. It should be emphasized that the marginal frequencies in these tables are not identical to those described in the previous section, due to the exclusion of a number of data points. Again, the characteristics mentioned for each cluster are only relevant with respect to deviations from expected values, and not with respect to relative frequencies.

Cluster AP1

This cluster is by far the largest of all clusters found in this conflict partner type. It contains 1933 observations, or about 70% of the total number of objects clustered.

This cluster has no relation with day of the week, or month of the year, except that December is under-represented. They are under-represented between midnight and 9 in the morning.

They are clearly over-represented inside built-up areas, and on roads with a 50 km per hour speed limit. The federal or railway authorities are less likely to be responsible for the roads in this cluster. There is no relation with the province, nor the size of the city.

These accidents are over-represented during daylight conditions, and less often in the dark. Street lights are likely to be in the vicinity. It is less likely to be raining, and the road is less likely to be wet. The road surface itself is more likely to consist of 'clinkers', and bitumen is under-represented.

These accidents are more likely to occur on straight road sections, and the road situation is less likely to be unusual. The relation with temporary road situations is not significant.

The driver is less likely to have been detected as having exceeded the legal limit for alcohol use and his possible alcohol use data is less likely to be missing. Neither does his sex nor age play any special role. Alcohol use by the pedestrian is unremarkable in this cluster, yet he is more likely to be male and less than 10 years of age. Pedestrians 21 or older are clearly under-represented.

The driver is more likely to be just driving straight ahead on the right side of the road. Strangely enough, he is less likely to be struck from behind, yet no category is over-represented. The driver is more likely to be exonerated from any infraction. The pedestrian is clearly more likely to be coming from the sidewalk, and is crossing the road. He is clearly more likely to be blamed for crossing the road incautiously, perhaps from behind an object.

The CBS manoeuvre variable is quite clear: the pedestrian is either 'suddenly' crossing the road, or crossing from behind an object.

In this cluster, the pedestrian is less likely to be killed, yet other than that, there is no indication that injuries received are more or less serious than could be expected.

Summary

This dry weather, daytime, type of accident tends to occur on straight road sections inside a built-up area, yet is not especially urban in nature. The driver, who is less likely to have been drinking, is driving straight ahead on the right hand side of the road when he strikes a young, male, child who 'suddenly' crosses the street from the sidewalk, and possibly from behind an object. This less than lethal type of accident seems to represent the young child mid-block dart-out (even though it may include adult mid-block dart-outs).

Cluster AP2

This cluster is much smaller than the previous one and consists of 450 observations.

The weekend (primarily Sunday) is over-represented in this cluster and Thursday appears to be under-represented. The winter months, November through January are over-represented. June is under-represented. These accidents are over-represented between 9 and 10 in the morning.

They are clearly over-represented inside built-up areas, and not on 80 km per hour roads. The local authorities are more likely to administer the road in question. South Holland is over-represented, while the provinces of Gelderland, Zeeland, North Brabant, and Limburg are all under-represented. Larger cities, with populations of more than 100,000 inhabitants, are over-represented. Smaller towns, with less than 20,000 inhabitants are under-represented.

These accidents are over-represented during darkness (?), and the (small) chance of not being in the vicinity of a street light is even smaller in this cluster. It is more likely to be raining during these accidents, and the road surface is more likely to be wet. The road surface itself is more likely than expected to be constructed of bitumen.

These accidents tend to occur more frequently at intersections and on or in the vicinity of a pedestrian crossing. There are generally no unusual temporary conditions.

This accident type has no clear relation with the driver's alcohol use nor his sex. The relationship with his age in not clear, although drivers in their forties are under-represented. The pedestrian is less likely to be charged with breaking the drinking and driving laws (?), is clearly more likely to be female, and is clearly more likely to be 21 or older. Children less than 11 years are clearly under-represented.

The driver is more likely than expected to be on the right side of a multi-lane road and less likely than expected to be on the right side of a single lane road. He is most likely turning left or right or accelerating from a stand still, when he is struck in the front center of his vehicle. He is more likely than expected to be charged with failure to yield right-of-way or not obeying a traffic light or sign. The pedestrian, on the other hand, is most likely to be on a pedestrian crossing, and is more likely to be crossing the road as opposed to other manoeuvres. He/she is either more likely to be exonerated from a violation or to be charged with neglecting to obey a traffic light or sign.
The CBS manoeuvre variable is very clear that this cluster concerns pedestrians using a pedestrian crossing.

All in all, this cluster may be slightly less dangerous than other clusters in that the pedestrian is less likely to be taken to the hospital.

<u>Summary</u>

This type of accident tends to occur during the winter months during weekend mornings. Neither weather conditions nor (natural) lighting conditions are optimal. The location tends to be an urban intersection or pedestrian crossing on a multi-lane road. The driver, who is turning or accelerating from a stand still then strikes the adult female pedestrian who is likely to be crossing the road on a pedestrian crossing. The driver is likely to be charged with failure to yield right-of-way or neglecting to obey a traffic light or sign, even though it may happen that the pedestrian may be crossing against the traffic light.

Cluster AP3

This cluster is relatively small, containing only 154 observations. Due to the small numbers, it is quite likely that it will be increasingly difficult to describe the statistically reliable characteristics of this cluster.

This cluster has no clear relation with day of the week, and is over-represented during the month of June and between 2 and 3 o'clock in the afternoon. (One is inclined to think that this relation is probably a fluke).

Nevertheless, accidents in this cluster clearly occur outside built-up areas, on roads with an 80 km per hour speed limit, and where provincial and federal authorities are more likely to be responsible for the roads in question. They are over-represented in the provinces of Friesland and Drenthe and under-represented in the more urban provinces of North and South Holland. Cities with more than 50,000 inhabitants are underrepresented and towns with less than 20,000 inhabitants are under-represented.

There is no clear relation with light conditions, yet it is more probable that there will no be street lighting in the vicinity. There is also no significant relation with weather conditions, yet the road surface is less likely to be wet. The road surface is more likely to consist of bitumen.

They are less likely to occur at a 'normal' intersection, yet it is not clear where they are over-represented. They are less likely to occur in the neighborhood of a pedestrian crossing, and perhaps a bit more likely to occur in the vicinity of a road exit/entrance. Temporary road conditions are not considered, due to small numbers.

It is difficult to conclude anything about the drivers' possible alcohol use, or his sex. He is somewhat more likely to be 23 years of age and less likely to be between 26 and 30, yet one may also suspect that this is a statistical fluke. The possible alcohol use of the pedestrian plays no clear role, and his sex plays no role in this cluster. He is statistically more likely to be between 30 and 50 years of age, less likely to be younger than 8.

The position of the driver on the road is not significant, and the role of his manoeuvre is unclear. He is slightly more likely to be struck on the left front side of his vehicle, and is most likely to be exonerated from any wrong doing. On the other hand, the pedestrian is clearly more likely to be on the road itself, the shoulder of the road, or 'someplace else', and is less likely to be on a pedestrian crossing or on the sidewalk. He is more likely to be crossing the road (direction unknown), and is most likely to be blamed for carelessly crossing the road. The CBS manoeuvre variable only confirms these findings.

This cluster clearly represents an especially serious type of accidents: the pedestrian is much more likely to be fatally injured or be taken to the hospital. Light injuries are under-represented.

<u>Summary</u>
This serious type of accident tends to occur on a rural, high-speed road. The mainly middle-aged pedestrian is more likely to be on the shoulder of the road or crossing in mid-block, without the benefit of a pedestrian crossing, when he is struck.

<u>Cluster AP4</u>

This last cluster of car-pedestrian accidents is approximately the same size as the previous one, consisting of 152 observations.

This accidents are more likely than expected to occur on Sunday, and less likely to occur on Wednesday. There is no apparent relation with month of the year, yet they are more likely to occur between 8 in the evening and 9 o'clock in the morning. The lunch hour is under-represented.

These accidents have a slight tendency to occur more frequently inside built-up areas than is expected, and apparently do not tend to occur on 80 km per hour roads. There is no special relation with the road authority.

They appear to have nothing to do with either the Province in which the accident occurs or the size of the city.

These accidents tend to occur more frequently in the dark, and there is no relation with the presence of street lighting. There is also no relation with weather conditions or the condition of the road surface. The road surface itself is more likely to consist of 'clinkers'.

These accidents are somewhat less likely to occur at a 'normal' intersection, or on a pedestrian crossing. They do appear to be over-represented in the vicinity of a parking lot. Nothing can be said about temporary road conditions.

The driver is more likely to have consumed some alcohol, to be arrested for drinking and driving, or to have his possible alcohol use registered as 'unknown'. This appears to be a more or less typical 'alcohol' accident (even though the majority of drivers were recorded as not having been drinking). The sex of the driver plays no role in this cluster, and there is no apparent relation with his age. The pedestrian in this case is also more likely to have gotten in trouble due to his possible alcohol use, even though nothing had been noted for the majority of cases. The sex of the pedestrian plays no role, yet he is more likely to be between 11 and 65 years of age. Children less than 11 years of age are clearly under-represented. (Unfortunately in this case, ages between 11 and 20 were all merged into one category, which limits our ability to determine precisely when this turn-around occurs).

The driver is less likely to be on the right side of the road, and somewhat more likely to be 'elsewhere'. He is also less likely to be driving straight ahead, and is more likely to be doing something 'else'. (Small numbers, and their effect on a reliable chi^2, make it difficult to determine which categories are reliably over-represented.) The car is less likely to be struck in the front center, and is more likely to be struck 'elsewhere'. Nevertheless, he is less likely to be exonerated, and is more often blamed for driving too far to the right or 'some other' violation. (In this case, the statistics are reliable. The problem is that the category 'other' is over-represented). The pedestrian is clearly not on a pedestrian crossing nor on the sidewalk. He appears primarily to be walking along or standing still on the road itself, or have the (mysterious) code 'not applicable'. He is either more likely to be exonerated or to be charged with to be acting carelessly on the road.

The CBS manoeuvre variable indicates that this cluster refers to accidents with pedestrians either walking along the road, standing or playing on the road, or doing something 'else'. He is clearly not crossing the road.

This type of accident causes less injures requiring a trip to the hospital, and more less-serious injuries.

<u>Summary</u>
This weekend, nighttime accident tends to occur inside built-up areas, even though the location is not especially 'urban' in nature. The driver as well as the adult pedestrian may have been drinking. The driver seems to be at some 'unusual' location and is implementing some 'unusual' manoeuvre when he strikes the pedestrian who is clearly not crossing but walking along or standing on the road. Small numbers, as well as the complications of alcohol and darkness, make this type of accident difficult to interpret.

6. RESULTS: AUTO-AUTO ACCIDENTS

4655 accidents between two cars were selected from the 1982 VOR accident file. The selection criteria required that 2 and only 2 'colliding' accident 'objects' were involved, both of which were non-parked vehicles. Of course, both of those vehicles were automobiles.

In the following paragraph a short summary of the marginal frequencies of the accident characteristics will be given.

6.1. Description of the data

Collisions between two cars, as registered in this file, seem to be somewhat less deadly than collision between cars and pedestrians, in the sense that the chance of a fatal accident or injury seems to be somewhat less. 117 of the 4655 (2.5%) registered accidents resulted in death for at least one person. 1037 (29.2%) required that someone be taken to the hospital, and 3524 (75.7%) resulted in some registered injury which did not require a trip to the hospital. One crucial difference, however, is that since automobiles can carry more passengers, multiple deaths and severe injuries are more frequent even though uncommon: 62 accidents (1.3%) had more than four victims. One accident involved 5 fatalities, and one accident required that 9 people had to be taken to the hospital.

These accidents also tend to occur primarily during weekdays (67%), yet occur more frequently during the daytime, with 76% occurring between 7 o'clock in the morning and 8 o'clock in the evening. There is a slight tendency for these accidents to occur more frequently during the last three months of the year, and somewhat less often from July to October. (February appeared to be an unexplainably relatively accident free month during 1982). 70% occur during daylight, and 66% occur in the vicinity of non-lit street lights. Only about 7% occur with no street light in the vicinity. 80% occur during dry weather, and 17% occur during rain. Other extreme weather conditions are only rarely noted. The road surface, which consists of bitumen in 82% of the time and clinkers 16% of the time, is also dry 63% of the time and wet 35% of the time.

64% of these accidents occur inside built-up areas, where the maximum speed is 50 km/hour in 63% of the cases, 80 km/hour in 27% of the cases, and 100 km/hour in 5% of the cases. 69% of these accidents occur on roads administered by the city government, the rest being evenly divided between provincial and federal authorities. 32% of these cases occur in cities with more than 100,000 inhabitants, 37% in cities with between 20,000 and 100,000 inhabitants, and the other 16% occur in municipalities with less than 20,000 citizens. 72% occur in the three most populous provinces of North Holland, South Holland, and North Brabant.

23% of these accidents occur on a straight road section, 51% at a 'normal' intersection, 20% at a T or Y intersection, and 6% in a curve. 4% occur in the vicinity of a road exit/entrance, and 1% in the vicinity of a parking place. Other unusual road characteristics or temporary conditions are infrequent. The vehicle manoeuvre being executed are, in order of frequency:

- two vehicles crossing each other's path at right angles at an intersection without turning (34%);
- two cars are travelling on the same road in opposite direction while one vehicle attempts to turn left and is struck by the vehicle coming from the opposite direction (14%);
- bumper to bumper collisions (13%);
- two vehicles crossing each other's path at right angles at an intersection, while one vehicle attempts to turn left and is struck by the vehicle coming from the left (12%);
- two vehicles are traveling on the same road in opposite directions and collide frontally without a lane change or any special manoeuvre (11%);

  etc.

It should be pointed out here that, in other sections, a distinction could be made in vehicle manoeuvre and road user characteristics on the basis of the type of vehicle. This is not the case in two-car collisions. However, following the VOR's defaults, the first vehicle mentioned is the vehicle whose driver is primarily blamed for the accident, and the second vehicle's driver is mostly exonerated or only partially to blame.

78% of the first vehicle's drivers are men, as are 80% of the second vehicles. 29% of the first vehicles' and 25% of the second vehicles' drivers are less than 25 years of age, 54% and 64% respectively are between 25 and 56 years of age, and 18% and 11% are older. 89% of the first and 95% of the second vehicles' driver have not had any registered alcohol use. It should be mentioned here that, in contrast to all other analyses, the category 'alcohol use unknown' was incorrectly treated as missing instead of be assigned it's own category.

74% and 80% of these two vehicles drivers' are on the right side of a single lane road, 14% and 17% are on the right side of a multiple lane road, 3% and 1% are on the left side of a single lane road, and 5% and 1% are on a road exit/entrance. Concerning the first vehicle, 61% are driving straight ahead, 6% are standing still (yet are not parked), braking, or accelerating from a stand still. 24% are turning left, and 4% are turning right. 3% are changing lanes, and 2% are making a U-turn. 80% of the second vehicles' drivers' are driving straight ahead, and 11% are braking, standing still (yet are not parked), or accelerating from a stand still. 1% are turning right and 6% are turning left.

With respect to the first vehicle, 44% are struck in the center front of the vehicle, 12% and 11% to the center left and center right respectively, 16% and 13% on the right and left flanks respectively. Only 4% are struck from behind. 46% of the second vehicles are struck on the center front, and 14% and 7% on the left and right front respectively. 5% and 13% are struck on the left and right flank respectively, and 14% are struck from behind.

As mentioned previously, the first vehicle is (mainly) the vehicle who receives the primary blame for the accident. Indeed, only 3% of these drivers are exonerated. 61% are blamed for not yielding the right-of-way, 11% are blamed for not keeping sufficient distance, 4% are blamed for ignoring a stop sign or traffic light, 3% drove incorrectly through a curve, 3% didn't keep sufficiently to the right, and 3% were on the wrong side of the road. On the other hand, the second vehicle was accounted no blame in 92% of the cases. Failure to yield right-of-way was cited in 3% of these accidents.

6.2. Description of the analysis

Homogeneity analysis

4655 observations, each consisting of observations on 30 variables, were submitted to a HOMALS analysis after recoding. This recoding was done to reduce the number of infrequently used categories. (See the previous section for a short summary of the marginal frequency distribution of the variables.)

The first six dimensions of the HOMALS analysis 'explained' respectively 12.7%, 10.9%, 8.9%, 8.2%, 7.7%, and 6.9% of the total variance. For the reasons mentioned previously, only the first three dimensions were used for further analysis.

Variables having a discrimination measure greater than 0.20 on at least one of the first three dimensions were:

- time of day;
- built-up area;
- maximum speed limit;
- road administrator;
- road situation;
- presence street lights;
- city size;
- first drivers' position on the road;
- the blame attached to the first vehicle;
- the manoeuvre of the second vehicle;
- the place where the second vehicle was struck.

Variables which had a loading equal to or greater than 0.20 on at least one of the last three dimensions, which did not load on one of the first three are:

- unusual road situation;
- weather;
- condition of the road surface;
- the manoeuvre of the first vehicle,
- the place were the first vehicle was struck;
- the second drivers' position on the road.

Inspection of the two dimensional object score scatter plot revealed one central 'peak' connected via a ridge to a second, much lower and flatter hill top. As one moves perpendicularly to this ridge one quickly reaches some gently rolling foothills. It appears that, in the two dimensional plot, one large cluster, and a number of secondary ones could be found.

6.3. Cluster analysis

The first three dimensions of the HOMALS results were selected for further analysis. They were then re-scaled such that the variance for each dimension was set equal to the percentage variance 'explained' by that dimension divided by the variance 'explained' by the first dimension. The construction of clusters proceeded in a number of steps.

First of all, the matrix of (re-scaled) matrix of 4655 observations times 3 values per observation was submitted to the SAS procedure FASTCLUS, which reduced the 4655 observations to 300 clusters. This was a fixed and arbitrary percentage of the original number of observations and was only implemented in order to reduce the computational requirements of the following cluster procedure.

Secondly, these 300 clusters were then submitted to the SAS procedure CLUSTER, which used the centroid algorithm. 54 clusters with a total number of 97 observations were then trimmed away. Visual inspection of the cubic clustering criterion, the pseudo F, and the pseudo T squared criteria indicated that about 6 clusters would be acceptable.

Third of all, the 6 clusters were then selected and all clusters with less than 100 observations or less than 5% of the total number of observations were eliminated. This procedure resulted in 5 clusters with a total of 4371 observations. Due to a programming oversight, 103 observations were incorrectly eliminated from further analysis. A total of 387 observations were eliminated, which is 8% of the total number of observations. The largest cluster contained 2357 observations, the second contained 783 observations, and the last three clusters contained respectively 550, 298 and 280 observations. Thus, of all objects finally clustered, respectively 54%, 18%, 13%, 7% and 6% were divided into these five clusters. This distribution of cluster sizes is, of course, more pleasing than the auto-pedestrian typology.

Finally, the derived cluster numbers were merged with their original data values and their HOMALS object scores. Visual inspections of object score plots indicated that clusters did not overlap in the original three dimensional HOMALS space. Discriminant analysis confirmed this impression. Objects were split into two groups, and the HOMALS scores were used to discriminate between the clusters. In the first group, the discriminant function weights were estimated, and the second group was used to confirm these weights. Unequal prior probabilities and variance-covariance matrices were assumed.

Discriminant analysis revealed that between 88% and 97% of the accidents in each cluster were correctly classified. Very similar results were obtained for the analysis on the second, confirmatory, group, with the exception of the smallest cluster in which only 80% of the observations were correctly classified. These results indicated that the object clustering does have some room for improvement, yet this situation may have been unduly exacerbated due to a programming oversight, which coded the incorrectly eliminated sub-groups mentioned above into a cluster and included it in this analysis.

6.4. <u>The clusters</u>

The derived clusters were bound, as previously mentioned, to their original data values. These accident objects were cross-tabulated: cluster cluster scores were crossed with all of the original variables. Inspection of these cross-tables will indicate the 'meaning' of each derived cluster and determine it's usefulness for the practitioner. It is to the interpretation of the found clusters that we will now turn.

The chi^2-value for each of the cross-tables for the analyzed variables in found in Table 4. (Also included are the values for the CBS manoeuvre and death and injury variables, which were not included in the analysis). As can be seen from this table, almost all interactions are significant, some chi^2-values being quite extreme. This result confirms our expectation that the HOMALS and Cluster analysis have successfully found groups of accident objects which are discriminable on the basis of their original characteristics.

In principle, we could (and perhaps should) carefully examine each and every cross-table with a significant chi^2. As previously mentioned, this approach would cost a great deal of effort and may tend to obscure the characterization of the clusters. For that reason, we will attempt to describe the clusters in turn in terms of their salient characteristics. To ensure some degree of reliability, we will only describe those cluster characteristics which have a chi^2 reliable at the 5% level. (Possible exceptions will be explicitly noted in the text).

Nevertheless, to aid the reader while describing the following clusters, a selection of some of the relevant cross-tabulations is found in Appendix II. It should be emphasized that the marginal frequencies in these tables are not identical to those described in the previous section, due to the exclusion of a number of data points. Again, the characteristics mentioned for each cluster are only relevant with respect to deviations from expected values, and not with respect to relative frequencies.

Cluster AA1

This first cluster is the largest in this group, containing 2357 observations.

The accidents in this cluster are under-represented on Sunday, and have no relation with the month of the year. They are also over-represented between 5 and 6 o'clock in the afternoon.

The accidents in this cluster are clearly a phenomenon occurring inside built-up areas, on 50 km/hour roads where city officials are most likely responsible. These accidents are over-represented in South Holland, and there is a weak tendency that they are also over-represented in North Holland and Utrecht also. They are under-represented in the provinces of Overijssel, Gelderland, Zeeland, and Flevoland. They occur more frequently than expected in larger towns and cities with more than 50,000 inhabitants and are under-represented in places with less than 20,000 inhabitants.

There is no relation with lighting conditions, and they are likely to occur in the vicinity of a street lamp. There is less likelihood of 'unusual' weather conditions, and a (non-significant) tendency that it's more likely to be raining. There is less chance of some 'unusual' road surface condition. The road surface itself has a greater likelihood to consist of 'clinkers' and less likelihood to consist of bitumen or some 'other' substance.

These accidents are more likely than expect to occur at an intersection and less likely to occur on a straight road section or a curve. There is also less likelihood that some unusual road condition is present. Temporary conditions play no role in this cluster.

Both drivers are less likely to have had alcohol use registered. The sexes of both drivers are unremarkable as are their ages.

The driver of the first vehicle is more likely to be driving on the right side of a two-lane road, and less likely to be on the left side of the road or exit/entrance lane. There is a better chance than expected that he is turning left when he is struck on the right front or flank. He is more often accused of ignoring a stop light or sign or failing to yield right-of-way. The location of the second vehicle is unremarkable, and there is a larger chance the he is either driving straight ahead or turning left himself. He is more likely to be struck on the left or right front or on the left flank, and is (even) more likely to be exonerated of any blame.

The CBS manoeuvre variable (see Appendix I) indicates that accidents numbers 411, 511, and 641 are over-represented. 411 refers to accidents with two vehicles travelling in opposite direction on the same road when one vehicle attempts to turn left and is struck by the other approaching vehicle. 511 refers to accidents where two vehicles cross paths at right angles at an intersection. Neither vehicle is turning. 641 refers to accidents where one vehicle is turning left onto a road and is struck by a vehicle approaching from the left on that same road.

This type of accident is less deadly that expected, and there is a smaller chance that the injuries received would require a trip to the hospital.

<u>Summary</u>
This cluster of accidents seems to represent a run-of-the-mill urban intersection accident where the first vehicle is turning and fails to yield the right of way or to obey a traffic control device.

Cluster AA2

This cluster is a great deal smaller than the first and contains 783 accidents.

These accidents are less likely to occur on Friday, and between 9 o'clock in the evening and 7 o'clock in the evening. They are over-represented between 10 and 11 o'clock in the morning, 2 and 4 in the afternoon, and between 5 and 6 in the afternoon. They are under-represented between October and December (and perhaps January). They are over-represented during the other months of the year, yet only the results for April are statistically reliable.

These accidents tend to occur outside built-up areas, on 70, 80, and 100 km/hour roads, where either the province or the federal government is responsible. More rural provinces, i.e., Friesland, Overijssel, Gelderland, Limburg, Flevoland, and perhaps Zeeland, are over-represented, and North and South Holland under-represented. As could be expected cities with less than 50,000 inhabitants are also over-represented, and larger cities are under-represented.

These accidents tend to occur more often than expected during daylight, and it is less likely that a street light is in the vicinity. The weather is more likely to be dry, as is the road surface. The road surface itself is more likely to be bitumen or some surface other than 'clinkers'.

These accidents also tend to occur at an intersection instead of a straight road section or a curve. Neither unusual structural or temporary conditions play a role here.

It is unlikely that either driver had been drinking, and neither drivers' sex plays a role here. The driver of the first vehicle is less likely to be younger than 30 or 40, and more likely to be 50 or older. The driver of the second vehicle is also more likely to be between 50 and 65, and less likely be to 20 or between 26 and 30. (There appears to be no clear under- or over-representation for other age groups).

The driver of the first vehicle is somewhat more likely to be on a road exit or entrance, and less likely to be on a multiple-lane road or on the left side of the road. He is more likely than expected to be turning left or accelerating out of a stand still, and to be struck on the left or right flank, or on the right rear of his vehicle. He is most likely to be blamed for failure to yield right-of-way. The driver of the second vehicle is more likely to be on the right side of a multi-lane road, and is more likely to be just driving straight ahead, when he is struck head-on. He is most likely to be exonerated from any blame.

Accidents type 411 and 511 are over-represented in CBS manoeuvre variable. 411 refers to two vehicles travelling in opposite directions on the same road, when one vehicle turns left and is struck by the other oncoming vehicle. 511 refers to two vehicles crossing each others path at right angle at an intersection or a road exit/entrance. Neither vehicle is turning.

This type of accident is particularly deadly, and serious injuries, requiring a trip to the hospital, are more likely. Light injuries are less likely than expected.

<u>Summary</u>
This cluster describes day-time accidents at an intersection of a rural, high-speed road, where there seems to be some inequality between the arms of the intersection. The first vehicle is likely to be turning or crossing, when he fails to yield right of way to the second vehicle. This type of accidents tends to be rather serious, and tends to involve two older drivers.

<u>Cluster AA3</u>

This cluster contains (only) 550 cases.

This type of accident has no clear relation with day of the week, yet tends to be over-represented during December and under-represented in February and March(?). It tends to be over-represented between 5 and 6 o'clock in the afternoon and between 9 and 10 in the evening. It is under-represented between 10 and 11 in the morning.

This type of accident clearly occurs primarily outside built-up areas, on 80 and 100 km/hour roads. Road authorities, other than city officials are responsible for the road. The provinces of Friesland, Zeeland, Gelderland, and Overijssel are over-represented, and North and South Holland are under-represented. Smaller towns, with less than 20,000 inhabitants are over-represented.

Dusk and darkness are more common in this cluster than could be expected, and it is quite likely that there are no street lights in the vicinity. Rain is less likely, but 'other' unusual weather conditions are more likely. The surface of the road is a bit more likely to be in some unusual condition. The road surface itself is most likely to be bitumen, or some 'other' substance.

These accidents are more likely to occur on a curve or a straight road section, and are more likely than expected to occur in the vicinity of a road exit/entrance or some 'other' situation. Temporary road situations play no role.

The driver of the first vehicle is more likely to have been drinking, yet there is no relation with the alcohol use of the second driver. The first driver is more likely to be male, the second driver of no special sex. The first driver is less likely to be older than 50 years of age, and more likely to be either 18, 21, or between 26 and 30. (Other younger age groups all point in the correct direction, even though the individual categories are not significant). The driver of the second vehicle is of no special age.

The driver of the first vehicle is more likely to be on the left side of the road, on a road exit/entrance, or some 'other' location. He is more likely to be braking, changing lanes, making a U-turn, or some 'other' unusual manoeuvre. He is more likely to be struck on the left or center front, or from straight behind. He is most likely to be blamed for skidding, driving incorrectly through a curve, or some other 'error'. The driver of the second vehicle is less likely to be on one of the lanes of a multiple lane road, and more likely to just be on the right side of the road or somewhere 'else'. He is more likely to be just driving straight ahead instead of any other manoeuvre, and is most likely to be

struck in the front center of his vehicle. In this case, however, the second driver also has a higher likelihood of being blamed for 'something'.
The CBS manoeuvre 211, and the rest of the 200 and 300 series, and the 100 series with the exception of 121 and 122 are over-represented. The 200 series, which is the most important in this case, represents collisions between two vehicles travelling in opposite directions on the same road. Neither vehicle is turning. 211, which is heavily over-represented, refers specifically to frontal collisions without a lane change. The 100 and 300 series refers to two vehicles driving in the same direction on the same road with or without one of the vehicles turning. 121 and 122 refer to bumper to bumper collisions with a braking or stationary vehicle.

As could be expected, these collisions are especially fatal, also often requiring hospitalization. Light injuries are less common.

Summary

This very serious group of late afternoon and evening accidents tends to occur on high speed rural roads on a straight road section, curve, or a road exit/entrance. The driver of the first vehicle is more likely to be a young male, who may have been drinking, and is more likely to be on the wrong side of the road or some other 'unusual' place, and is either braking, changing lanes, making a U-turn or some other 'unusual' manoeuvre, when he is struck frontally (even though bumper-bumper collisions sometimes occur in this category) by the other vehicle. Apparently, the driver of the first vehicle left his lane for some reason, resulting mainly in a head-on collision. (Skidding and winter months are also implicated).

Cluster AA4

This is rather small, containing only 298 observations.

This cluster is over-represented on Sundays, and under-represented on Tuesday through Thursday. It is more likely than expected to occur during December or February, and less likely in July. (March through June are also under-represented, even though none of the categories for the individual months is statistically reliable). This accident tends to occur during the night, primarily between 9 in the evening and 7 in the morning. Morning and afternoon hours are also clearly under-represented.

This is an inside built-up areas accident, occurring primarily on 50 km/hour roads, with city officials administrating. North Holland is over-represented, while South Holland has a non-significant tendency to also be over-represented. Gelderland, Overijssel, and Drenthe are all under-represented. This cluster of accidents tend to occur in cities with a population of more than 100,000 and is relatively less common in cities with less than 50,000 inhabitants.

This type of accidents tends to occur during darkness, and there is almost always a street light in the vicinity. It is more likely to be raining, and the road is more likely to be wet. The road surface itself is more likely to consist of 'clinkers' than could be expected.

These accidents are more likely to occur on a straight road section or a curve, with some unusual road situation a bit more likely. Temporary road conditions remain irrelevant in this cluster.

There is a larger chance that the first driver had been drinking, with about 1/3 of all (first vehicle) drivers arrested for drinking and driving falling in this cluster. The second driver is also more likely to have been drinking, with 60% of all (second vehicle) drivers arrested for drinking and driving also falling in this cluster. The first driver is more likely to be male, while the sex of the second driver is unremarkable. The first driver, generally speaking, is more likely to be 30 or less, and less likely to be older than 30. The driver of the second vehicle is more likely to be in his twenties, and less likely to be older than 40.

The first driver is a bit more likely to be on the left side of the road, or some unusual place 'elsewhere'. He is more likely to be braking, turning right (but not left!), changing lanes, or doing something unusual. He is more likely to be struck on the left front or center of his vehicle, or from straight behind. He is more often blamed for incorrectly driving through a curve, skidding, or some 'other' infraction. Surprisingly, he is also relatively often not blamed at all. The second driver is slightly more likely to be 'somewhere else' than on the right side of the road, and has an increased chance of turning either left or right. He is more likely than expected to be struck on the left front, the right side, or the left rear of his vehicle (?). In contrast to the first driver, the second driver has a higher chance of being blamed for some 'error'.

The CBS manoeuvre (see Appendix I) is more likely to be a 211 or another member of the 200 series, or one of the 600 series with the exception of 641. We've already seen that 211 represents a frontal collision, while the 200 series refers in general to two vehicles driving in opposite directions on the same road. The 600 series in general refers to one vehicle turning on to another road and being struck by a vehicle on that road. Variations in the 600 series describe in which direction the one vehicle turned, and whether he was struck frontally or from behind.

This cluster seems to be relatively safe, in terms of a lessened chance of a fatal injury. Light injuries are relatively likely, and serious injuries are unremarkable.

Summary

These urban, nighttime type of accident tends to occur on straight road section or curves, and tends to occur somewhat more often during the winter months in the rain. Both drivers, of whom the first of which is more likely to be male, are more likely to be in their twenties and have quite probably been accused of drinking and driving. There is an increased chance that one or the other driver is somewhere 'else' than in their own lane, and there is an increased likelihood that one or both drivers are either turning, braking, or doing something 'else'. Skidding is also implicated. Either the alcohol and/or the rain created a somewhat unpredictable situation, which fortunately doesn't often lead to fatalities.

Cluster AA5

This cluster is about the same size as the previous one, containing 280 cases.

This accident is less likely to be a Sunday accident, yet is not especially likely for any other day of the week. Month of the year is not relevant in this case. It is less likely to occur after midnight, and more likely between 1 and 2 in the afternoon.

This accident tends to occur inside built-up areas, and on 50 km/hour roads administered by city officials. North Holland and South Holland are over-represented, as are cities with more than 100,000 inhabitants. Cities with less than 50,000 inhabitants are under-represented.

There is more likely to be daylight, and, of course, there is more likely a street light in the vicinity. There is no special weather effect in this cluster, and the condition of the road surface is unremarkable. The road surface is more likely to be bitumen instead of clinkers.

These accidents are more likely to occur on a straight road section, with some 'unusual' road situation a bit more likely. Temporary road situations might play a role here, but the small numbers plague the possibility of a statistically reliable statement.

Little extra can be said about the first drivers' alcohol use, and it may be the case that the second driver is less likely to have been (discovered) drinking. The first driver has a better chance of being female, and is of no clear age. The second driver is of no special sex, yet is more likely to be between 50 and 65 years of age.

The first driver is more likely to be in one of the lanes of a multiple-lane road, and is either driving straight ahead or braking. He is most likely struck in the center front of his vehicle, and is most likely blamed for not keeping sufficient distance. The second driver is somewhat more likely to be in one of the lanes on a multiple-lane road and is quite likely braking or even standing still. He is extremely likely to be struck from behind and is almost always exonerated.

CBS manoeuvres (see Appendix I) 111, 121 and 122 are all over-represented. It isn't really necessary to expostulate on a description of bumper to bumper collisions.

This type of accident is fatal less often than expected, and is less likely to require hospitalization. Light injuries (of only one person) are more likely.

Summary

This small cluster describes an afternoon, urban, straight road section of a multi-lane-road type of accident. The somewhat older driver of the second vehicle is either braking or standing still, when he is struck from behind by the first driver, who has a better chance of being a female. Perhaps this situation can be summarized as an arterial, urban, (non-intersection), bumper-to-bumper type of accident.

7. RESULTS: AUTO-OBJECT ACCIDENTS

Single vehicle, in this case automobile, accidents are somewhat more problematical than the previously discussed accident groups.
In the first place, they require a slightly different choice of variables to be analyzed. For example, it would seem rather ludicrous to analyze the age, sex, possible alcohol use, and manoeuvre of inanimate objects with which an automobile may collide.
Secondly, preliminary studies revealed that a large distinction was made between accidents wherein cars collide with an object and accidents wherein cars collide with nothing, e.g., slips, running off the road, etc. This result was deemed an artefact, noting than if a collision object did not exist, then all variables describing that (non-existing) object contained the same (lack of) information. For this reason, all further analyses were made on two, separate, a priori sub-groups: single vehicle car accidents with, and single vehicle car accidents without a collision object. (Preliminary analyses were made including the CBS-manoeuvre variable. This variable was subsequently removed from the analysis, and it was not investigated whether this artificial division into two sub-groups would be alleviated or not).
Third of all, care has to be taken in eliminating irrelevant accidents from the analyses. For instance, originally all single vehicle accidents involving a moving car were analyzed. It was later discovered that the data set included two non-colliding vehicles, e.g., bicyclists who slipped and fell in successfully avoiding a collision with a car. These cases were eliminated from further analysis. Therefore, the following analyses have been implemented on two sub-groups of accidents: accidents involving one moving vehicle (an automobile) who either does not collide nor is involved with any other object, or who collides with an object which is not another moving vehicle.

In the following paragraphs a short summary of the marginal frequencies of the accident characteristics for the first group, collisions between a car and another object, will be given, followed by a description of the data analysis and a description of the results.
Data description and analyses for the second group, single car accidents involving no other object, will be separately discussed in Chapter 8.

7.1. Description of the data (Single cars and another object)

The 3162 collisions between a car and another object, as registered in this file, seems to be especially serious. 221 of the 3162 (7.0%) registered accidents resulted in death for at least one person. 1331 (42.1%) required that someone be taken to the hospital, and 1765 (55.8%) resulted in some registered injury which did not require a trip to the hospital.

These accidents also tend to occur primarily during weekdays (59%) yet Saturdays and Sundays are clearly over-represented. They, surprisingly, occur more frequently during the evening and early morning hours, with (only) 48% occurring between 7 o'clock in the morning and 8 o'clock in the evening. There may be a slight tendency for these accidents to occur more frequently during December and January. Only 41% occur during daylight, and 77% occur in the vicinity of street lights. 23% occur with no street light in the vicinity. 81% occur during dry weather, and 13% occur during rain. Other extreme weather conditions are only rarely noted. The road surface, which consists of bitumen in 82% of the time and clinkers 15% of the time, is also dry 59% of the time and wet 32% of the time. "Other" conditions occur about 8% of the time.

(Only) 47% of these accidents occur inside built-up areas, where the maximum speed is 50 km/hour in 46% of the cases, 80 km/hour in 42% of the cases (!), and 100 km/hour in 8% of the cases. 69% of these accidents occur on roads administered by the city government, the rest being evenly divided between provincial and federal authorities. 23% of these cases occur in cities with more than 100,000 inhabitants, 34% in cities with between 20,000 and 100,000 inhabitants, and the other 42% (!) occur in municipalities with less than 20,000 citizens. (Only) 51% occur in the three most populous provinces of North Holland, South Holland, and North Brabant.

54% (!) of these accidents occur on a straight road section, only 14% at some sort of intersection, and 33% (!) in a curve. Unusual road situations are again rather infrequent with only 1% occur in the vicinity of a road exit/entrance, 1% in the vicinity of a parking place, and 2% in the vicinity of a bridge or tunnel. Other unusual road characteristics or temporary conditions are quite infrequent. The vehicle manoeuvres being executed are, in order of frequency:

- a single car strikes an object on the side of the road (57%);
- a single car strikes a streetlight on the side of the road (20%);
- a single vehicle strikes from behind a parked vehicle on the same road (5%);

  etc.

This variable is not terribly enlightening seeing that it, in this case, describes the object struck rather than the manoeuvre being executed.

82% of the drivers are men. 41% of the drivers are less than 25 years of age, 46% are between 25 and 50 years of age, and 12% are older. Only 65% of the drivers had not had any registered alcohol use, 21% (!) were charged with drinking and driving.

84% of the drivers were on the right side of a two lane-road, and 12% on the right side of a multiple lane road, with other starting positions being rather infrequent. In 87% of the cases, the driver was driving straight ahead, in 4%, 4%, and 2% of the cases, he was, respectively braking, or turning left or right. His car was initially struck on the front left in 6% of the cases, on the center front in 70% of the cases, and on the right front in 9% of the cases. In 5% of the cases, he was struck on the right flank, and on the right flank in 5% of the cases. In 24% of the cases, the driver 'lost control of the wheel'; in 18% of the cases, he was driving too far to the right. 17% of the time the driver went into a slip, and 15% of the time he drove incorrectly through a curve.

Trees were most frequently struck (44% of the time), followed by light masts (22%), and fixed objects (24%). Parked vehicles were struck about 8% of the time. The struck object was located on the shoulder of the road about 57% of the time, on the sidewalk 28% of the time, the remainder of objects being on the road itself, on a safety island, or somewhere 'else'.

7.2. Description of the analysis

Homogeneity analysis

3162 observations, each consisting of observations on 25 variables, were submitted to a HOMALS analysis after recoding. This recoding was done to reduce the number of infrequently used categories. (See the previous section for a short summary of the marginal frequency distribution of the variables).

The first six dimensions of the HOMALS analysis 'explained' respectively 17.4%, 12.1%, 10.3%, 10.0%, 8.7%, and 6.8% of the total variance. For the reasons mentioned previously, only the first three dimensions were used for further analysis.

Variables having a discrimination measure greater than 0.20 on at least one of the first three dimensions were:
- built-up area;
- maximum speed limit;
- road administrator;
- presence of street lights;
- city size;
- first drivers' position on the road;
- the blame attached to the driver;
- the type of object collided with;
- the object's location on the road.

Variables with a loading equal to or greater than 0.20 on at least one of the last three dimensions with no loading on one of the first three are:
- month of the year;
- time of day;
- road situation;
- light conditions;
- condition of the road surface;
- alcohol use.

Inspection of the two dimensional object score scatter plot revealed two, almost equally large and well separated clusters, as well as a more diffuse, smaller group of accidents. At least two large clusters should be easily obtainable.

7.3. Cluster analysis

The first three dimensions of the HOMALS results were selected for further analysis. They were then re-scaled such that the variance for each dimension was set equal to the percentage variance 'explained' by that dimension divided by the variance 'explained' by the first dimension.

The construction of clusters proceeded in a number of steps.

First of all, the matrix of (re-scaled) matrix of 3162 observations times 3 values per observation was submitted to the SAS procedure FASTCLUS, which reduced the 3162 observations to 200 clusters. This was a fixed and arbitrary percentage of the original number of observations and was only implemented in order to reduce the computational requirements of the following cluster procedure.

Secondly, these 200 clusters were then submitted to the SAS procedure CLUSTER, which used the centroid algorithm. 30 clusters with a total number of 64 observations were then trimmed away. Visual inspection of the cubic clustering criterion, the pseudo F, and the pseudo T squared criteria indicated that about 5 clusters would be acceptable.

Third of all, the 5 clusters were then selected and all clusters with less than 100 observations or less than 5% of the total number of observations were eliminated. This procedure resulted in 3 clusters with a total of 2959 observations. A total of 203 observations were eliminated, which is 6% of the total number of observations. The largest cluster contained 1328 observations, the second contained 1325 observations, and the last cluster contained 306 observations. Thus, of all objects finally clustered, respectively 45%, 45% and 10% were divided into these three clusters. This distribution of cluster sizes is, of course, again more pleasing than the auto-pedestrian typology.

Finally, the derived cluster numbers were merged with their original data values and their HOMALS object scores. Visual inspections of object score plots indicated that clusters did not overlap in the original three dimensional HOMALS space. Discriminant analysis confirmed this impression. Objects were split into two groups, and the HOMALS scores were used to discriminate between the clusters. In the first group, the discriminant function weights were estimated, and the second group was used to confirm these weights. Unequal prior probabilities and variance-covariance matrices were assumed.

Discriminant analysis revealed that between 86% and 96% of the accidents in each cluster were correctly classified. Very similar results were obtained for the analysis on the second, confirmatory, group. The accidents in the two larger clusters were more accurately classified than those in the smaller one.

7.4. The clusters

The derived clusters were bound, as previously mentioned, to their original data values. These accident objects were cross-tabulated: cluster scores were crossed with all of the original variables. Inspection of these cross-tables will indicate the 'meaning' of each derived cluster and determine it's usefulness for the practitioner. It is to the interpretation of the found clusters that we will now turn.

The chi^2-value for each of the cross-tables for the analyzed variables in found in Table 5. (Also included are the values for the CBS manoeuvre and death and injury variables, which were not included in the analysis). As can be seen from this table, almost all interactions are significant, some chi^2-values being quite extreme. This result confirms our expectation that the HOMALS and Cluster analysis have successfully found groups of accident objects which are discriminable on the basis of their original characteristics.

In principle, we could (and perhaps should) carefully examine each and every cross-table with a significant chi^2. As previously mentioned, this approach would cost a great deal of effort and may tend to obscure the characterization of the clusters. For that reason, we will attempt to describe the clusters in turn in terms of their salient characteristics. To ensure some degree of reliability, we will only describe those cluster characteristics which have a chi^2 reliable at the 5% level. (Possible exceptions will be explicitly noted in the text).

Nevertheless, to aid the reader while describing the following clusters, a selection of some of the relevant cross-tabulations is found in Appendix II. It should be emphasized that the marginal frequencies in these tables are not identical to those described in the previous section, due to the exclusion of a number of data points. Again, the characteristics mentioned for each cluster are only relevant with respect to deviations from expected values, and not with respect to relative frequencies.

Cluster A01

This cluster, which contains 1328 accidents, is only negligibly larger than the second cluster.

The accidents in this cluster bear no clear relation to day of the week, yet are under-represented during the winter months from November till January. They are less likely to occur between midnight and 5 o'clock in the morning, and are over-represented between 3 in the afternoon till 8 o'clock in the evening.

They are clearly over-represented outside built-up areas, and are clearly over-represented on 80 km/hour roads, roads with other speed limits being under-represented. The road authorities are more likely than expected to be provincial. Rural provinces, such as Groningen, Friesland, Drenthe, Overijssel, Gelderland, Zeeland, and North Brabant, are over-represented, and the more urban provinces of North and South Holland are under-represented. As could be expected, cities with less than 20,000 inhabitants are over-represented and cities with more than 50,000 inhabitants are under-represented.

These accidents are more likely than expected to occur during daylight, and less likely during darkness. Street lights have a larger chance of being absent. Misty weather is more likely than expected, and rain is under-represented. The road surface is more likely than expected to be dry than otherwise. The road surface is more likely to be bitumen or some other substance than 'clinkers'.

Curves are more likely than expected, intersections less likely. It is also a bit less likely that this type of accident occurs in the vicinity of a bridge or viaduct. It is also somewhat less likely to occur under unusual temporary circumstances.

The driver is more likely than expected to have no alcohol consumption detected, and is less likely to be arrested for drinking and driving. His sex plays no role by this type of accident, the role of his age being unclear, even though 19 year olds appear to be over-represented.

The driver is more likely to be on the right hand side of a single lane road than on a multiple lane road, and is somewhat more likely to be on the left hand side of the road, the chance of which nevertheless remains quite small. He is more likely than expected to be just driving straight ahead instead of implementing some special manoeuvre. He is somewhat less likely to be struck on the left front side of his vehicle, yet it is unclear where he is more likely to be struck. He is more likely than expected to be blamed for driving too far to the right or driving incorrectly through a curve.

The object with he collides is most likely to be a tree located on the shoulder of the road.
As previously noted, the CBS manoeuvre is not terribly enlightening in this case.

This type of accident is more likely to lead to a fatality or injuries requiring transport to the hospital. Light injuries are relatively less common.

<u>Summary</u>
This rather serious non-winter, afternoon type of accident tends to occur on a curve of a high-speed, single-lane rural road. The driver, who is just driving straight ahead, apparently just drives off the road and runs into a tree located on the shoulder. Mist may have something to do with this.

<u>Cluster A02</u>

This cluster, containing 1325 cases, is for all practical purposes equal in size to the previous cluster.

These accidents bear no clear relation to day of the week, and are over-represented during the winter months, as opposed to the previous cluster. They tend to be more of a night-time type of accident, being over-represented generally between 9 o'clock in the evening till 5 o'clock in the morning, even though not every hour is significantly over-represented. They are under-represented between 8 and 10 in the morning, and between 1 in the afternoon and 8 o'clock in the evening, even though not every hour is significantly under-represented.

These accidents are over-represented inside built-up areas, where the maximum speed limit is 50 km/hour, and where cities authorities are responsible for road administration. The urban provinces of North and South Holland are over-represented, while the provinces of Friesland, Overijssel, Gelderland, North Brabant, and Flevoland are under-represented. (Other 'rural' provinces also tend to be under-represented, albeit not significantly). As may be expected, cities with a population of more than 50,000 are over-represented, cities with less than 20,000 are under-represented.

These accidents are more likely than expected to occur during conditions of darkness, and in the vicinity of a street light. Mist is somewhat less likely than expected. The road surface is somewhat less likely to be dry, and less likely to consist of bitumen.

The accidents are more likely than expected to occur at an intersection (even though 75% of the accidents in this cluster do not occur at an intersection). There is apparently no special unusual permanent or temporary road situation.

There is a relatively large chance that the driver was arrested for drinking and driving, and neither his sex nor his age plays any role in this cluster. As in the previous cluster, the driver is more likely to be driving on the right side of a single lane road, and has an increased chance of either turning left or right. There is also an increased chance that he is struck on the right front of his vehicle, and less chance that he is struck on the right flank or overturns. There is no special blame attached to his behaviour.

He is also more likely to strike a street light and less likely to strike a tree. This object is also more likely to be located on the sidewalk or on a safety island.

This type of accident is less likely to result in a death or an injury requiring transfer to the hospital, and more likely to result in light injuries.

<u>Summary</u>
This winter evening type of accident tends to occur on urban roads, and intersections are over-represented. The driver is more likely to have been drinking, and, while turning, is likely to strike a lantern pole located on a safety island or sidewalk.

<u>Cluster A03</u>

This last cluster is relatively much smaller in comparison with the previous two, containing 306 cases.

There is no clear relation with day of the week nor month of the year. There is some indication that this is more of a day-time type of accident: it is over-represented between 9 and 10 in the morning and 5 and 6 in the afternoon and is under-represented between 11 and 12 at night. Other hour categories are not significant, yet tend to support this tendency.

These accidents are more likely to occur outside built-up areas, and on road with maximum speed limits of either 70 or 100 km/hour. Roads with speed limits of 50 or 80 km/hour are under-represented. Federal authorities are more likely to be responsible for the road in question. North and South Holland, and Utrecht are over-represented, while Groningen, Friesland, Overijssel, and North Brabant are under-represented. Cities with more than 100,000 inhabitants are over-represented, while towns with between 10,000 and 20,000 inhabitants(?) are under-represented.

Daylight conditions are over-represented and darkness is under-represented. There is no special relationship with the presence of street lighting. There is also no clear relation with either weather conditions or the condition of the road surface. The road surface itself is more likely to consist of bitumen and less likely to consist of clinkers or 'other' substances.

These accidents are more likely to occur on a straight road section and less likely to occur in a curve. The (slight) chance that these accidents occur in the vicinity of a bridge or tunnel is increased. The relation with temporary road conditions is unclear.

These type of accidents are less likely to be registered as alcohol involved, and the sex of the driver plays no role. He is less likely to be 19 or 20 years of age and more likely to be between 26 and 65.

The driver is clearly more likely to be on the right side of a multi-lane road or on a road exit/entrance. The (slight) chance that he is braking, changing lanes, or implementing 'other' manoeuvres is increased, and the (slight) chance that he is turning left is decreased. He is more likely to overturn or be struck on the left front of his vehicle, and is less likely to be struck on the right or center front of his vehicle. He is more likely to be blamed for not keeping far enough to the right hand side of the road, for skidding, for being tired, falling asleep or becoming ill, or some 'other' reason. He is less likely to be blamed for driving too far to the right, driving too fast, or driving incorrectly through a curve.

He is more likely to strike a safety rail or some other fixed object, and less likely to strike a tree or a street light. The struck object is clearly more likely to be located on the median of the road or somewhere 'else'. Sidewalks are under-represented.

These accidents are not unusual in the probability of involving a fatality, yet there is less chance of requiring transport to the hospital.

Summary

This type of accident occur typically during the daytime on 100 km per hour urban area roads outside built-up areas. They also tend to occur on straight road sections or near a road exit/entrance. The driver, who is more likely to be between 26 and 65 years old, is braking, changing lanes, merging, or implementing some other manoeuvre, when he strikes a safety rail or some fixed object. He may have fallen asleep, become ill, skidded, or for some reason didn't keep far enough to the right.

8. RESULTS: AUTO-NO OBJECT ACCIDENTS

8.1. Description of the data (single cars only)

The 851 collisions involving only one car and no other object, as registered in this file, also seems to be especially serious. 56 of the 851 (6.6%) registered accidents resulted in death for at least one person. 359 (42.2%) required that someone be taken to the hospital, and 469 (55.1%) resulted in some registered injury which did not require a trip to the hospital.

These accidents also tend to occur primarily during weekdays (60%), yet Saturdays and Sundays are clearly over-represented. They, surprisingly, occur more frequently during the evening and early morning hours, with (only) 50% occurring between 7 o'clock in the morning and 7 o'clock in the evening. There is no clear seasonal pattern to these accidents, with the exception that more than 11% occur during the month of July. 51% occur during daylight, and 53% (see previous section referring to collisions between one car and an object!) occur in the vicinity of street lights. 47% occur with no street light in the vicinity. 84% occur during dry weather, and 9% occur during rain. Other extreme weather conditions are only rarely noted. The road surface, which consists of bitumen in 93% of the time and clinkers 4% of the time (!), is also dry 64% of the time and wet 25% of the time. "Other" conditions occur about 11% of the time.

(Only) 13% (!) of these accidents occur inside built-up areas, which is extremely few. The maximum speed is 50 km/hour in 14% of the cases, 80 km/hour in 60% of the cases (!), and 100 km/hour in 23% of the cases. These single car accidents again appear to be quite different from the previously studied group, i.e, collisions between one car and an object. 47% of these accidents occur on roads administered by the city government, the rest being more or less evenly divided between provincial and federal authorities. Only 9% of these cases occur in cities with more than 100,000 inhabitants, 31% in cities with between 20,000 and 100,000 inhabitants, and the other 60% (!) occur in municipalities with less than 20,000 citizens. (Only) 44% occur in the three most populous provinces of North Holland, South Holland, and North Brabant.

54% (!) of these accidents occur on a straight road section, only 7% at some sort of intersection, and 40% (!) in a curve. Unusual road situations are again rather infrequent, being registered only about 2% of the time. Other unusual road characteristics or temporary conditions are quite infrequent. The vehicle manoeuvres being executed are, in order of frequency:

- a single car drives off the road into the water (54%);
- a single car drives off a straight road section (16%);
- a single vehicle skids, yet remains on the road (13%);
- a single vehicle drives off the road in the vicinity of a curve (13%); etc.

This variable is not terribly enlightening in this case either.

67% of the drivers are men. 43% of the drivers are less than 25 years of age, 46% are between 25 and 50 years of age, and 11% are older. Only 73% of the drivers had not had any registered alcohol use, 14% (!) were charged with drinking and driving.

77% of the drivers were on the right side of a two lane-road, and 15% on the right side of a multiple lane road, with other starting positions being rather infrequent. In 89% of the cases, the driver was driving straight ahead, in 4% and 4% of the cases, he was, respectively braking, or turning. In 32% of the cases, the car overturned. In 28% of the cases, the driver 'lost control of the wheel'; in 14% of the cases, he was driving too far to the right. 18% of the time the driver went into a skid, and 17% of the time he drove incorrectly through a curve.

It should be noted that there are some striking differences between the marginal frequencies in this case (single cars only) and the previous one (single cars and an object).

8.2. Description of the analysis

Homogeneity analysis

851 observations, each consisting of observations on 22 variables, were submitted to a HOMALS analysis after recoding. This recoding was done to reduce the number of infrequently used categories. (See the previous section for a short summary of the marginal frequency distribution of the variables).

The first six dimensions of the HOMALS analysis 'explained' respectively 14.8%, 12.6%, 12.2%, 10.2%, 9.1%, and 7.6% of the total variance. For the reasons mentioned previously, only the first three dimensions were used for further analysis.

Variables having a discrimination measure greater than 0.20 on at least one of the first three dimensions were:

- month of the year;
- time of day;
- built-up area;
- speed limit;
- road administrator;
- road situation;
- light conditions;
- condition of the road surface;
- city size;
- the location of the car on the road;
- the 'blame' for the accident;

Variables which had a loading equal to or greater than 0.20 on at least one of the last three dimensions, which did not load on one of the first three are: weather conditions and province.

Inspection of the two dimensional object score scatter plot revealed a large, rather diffuse cluster of points.

8.3. Cluster analysis

The first three dimensions of the HOMALS results were selected for futher analysis. They were then re-scaled such that the variance for each dimen- dimension was set equal to the percentage variance 'explained' by that sion divided by the variance 'explained' by the first dimension.

The construction of clusters proceeded in a number of steps.

First of all, the (re-scaled) matrix of 851 observations times 3 values per observation was submitted to the SAS procedure FASTCLUS, which reduced the 851 observations to 60 clusters. This was a fixed and arbitrar percentage of the original number of observations and was only implemented in order to reduce the computational requirements of the following cluster procedure.

Secondly, these 60 clusters were then submitted to the SAS procedure CLUSTER, which used the centroid algorithm. 10 clusters with a total number of 22 observations were then trimmed away. Visual inspection of the cubic clustering criterion, the pseudo F, and the pseudo T squared criteria indicated that about 4 clusters would be acceptable.

Third of all, the 4 clusters were then selected and all clusters with less than 50 observations or less than 5% of the total number of observations were eliminated. (Notice that our requirement of at least 100 observations has been relaxed in this case due to the rather small number of observations.) This procedure resulted in 4 clusters with a total of 829 observations. A total of 22 observations were eliminated, which is 3% of the total number of observations. The largest cluster contained 474 observations, the second contained 223 observations, and the last two clusters contained 64 and 58 observations respectively. Thus, of all objects finally clustered, respectively 57%, 27%, 8% and 7% were divided into these four clusters.

Finally, the derived cluster numbers were merged with their original data values and their HOMALS object scores. Visual inspections of object score plots indicated that clusters did not overlap in the original three dimensional HOMALS space. Discriminant analysis confirmed this impression. Objects were split into two groups, and the HOMALS scores were used to discriminate between the clusters. In the first group, the discriminant function weights were estimated, and the second group was used to confirm these weights. Unequal prior probabilities and variance-covariance matrices were assumed.

Discriminant analysis revealed that between 97% and 100% of the accidents in each cluster were correctly classified. Similar results were obtained for the analysis on the second, confirmatory, group, where correct classifications varied from about 89% to 97%.

8.4 The clusters

The derived clusters were bound, as previously mentioned, to their original data values. These accident objects were cross-tabulated: cluster scores were crossed with all of the original variables. Inspection of these cross-tables will indicate the 'meaning' of each derived cluster and determine it's usefulness for the practitioner. It is to the interpretation of the found clusters that we will now turn.

The chi^2-value for each of the cross-tables for the analyzed variables in found in Table 6. (Also included are the values for the CBS manoeuvre and death and injury variables, which were not included in the analysis). As can be seen from this table, almost all interactions are significant, some chi^2-values being quite extreme. This result confirms our expectation that the HOMALS and Cluster analysis have successfully found groups of accident objects which are discriminable on the basis of their original characteristics.

In principle, we could (and perhaps should) carefully examine each and every cross-table with a significant chi^2. As previously mentioned, this approach would cost a great deal of effort and may tend to obscure the characterization of the clusters. For that reason, we will attempt to describe the clusters in turn in terms of their salient characteristics. To ensure some degree of reliability, we will only describe those cluster characteristics which have a chi^2 reliable at the 5% level. (Possible exceptions will be explicitly noted in the text).

Cluster AN1

This first cluster is by far the largest, containing 474 accidents.

There is no convincing special relation with day of the week. November is over-represented and January under-represented, but altogether there is no really clear pattern over the months of the year. Nevertheless, this is clearly a nighttime sort of accident: the hours between 7 in the evening and 7 in the morning are clearly over-represented, and all other times of day are clearly under-represented.

These accidents are over-represented outside of built-up areas. (It should be noted that almost all of these single car accidents occur outside of built up areas). 80 km/hour roads are also over-represented, and roads with other speed limits (including 100 km/hour) are under-represented. Roads administered by city officials are over-represented, and those administered by national officials are under-represented. These accidents are over-represented in Zeeland (?), and tend to occur more often than expected in villages with less than 5,000 inhabitants. Large cities are under-represented.

These accidents occur more frequently than expected during darkness, yet there is no special connection with the presence of street lighting. There is no relation with weather conditions, yet there is a lessened chance that the road is covered with snow or ice. Wet surfaces are over-represented. The road surface itself consists of no special type of material.

These accidents are more likely than expected to occur in the vicinity of a curve, and less likely on a straight road section. Temporary circumstances play no role.

The driver is more likely to have been drinking or to have been arrested for drinking and driving. He is more likely to be a man, although his age plays no clear role.

The driver is more likely than expected to be on the right hand side of a two lane road, and is less likely to be changing lanes or making some 'other' type of manoeuvre. There is a higher chance that the car will overturn in this type of accident. The driver is more likely to be blamed for driving too fast or driving incorrectly though a curve.
There is also a larger chance that the car will end up in the ditch (CBS manoeuvre.)

This type of accident is not especially fatal, nor is there any unusual probability of having being taken to the hospital.

<u>Summary</u>
This nighttime accident tends to occur on curves on a 80 km per hour

road outside built-up areas in a rural environment. The road surface may be wet. The driver is more likely to be male and alcohol may be implicated. He appears to be driving on the right side of a two lane road, implementing no special manoeuvre, when he drives off the road and overturns or runs into a ditch. High speed may also be implicated.

Cluster AN2

This cluster is much smaller than the first, and contains 223 cases.

Wednesdays appear to be over-represented (?), yet there is no clear pattern for days of the week in relation to this cluster. December and January and under-represented, and July is over-represented. This is clearly a daytime accident, the hours between 10 o'clock in the morning and 7 o'clock in the evening are over-represented, and the hours between 7 in the evening and 7 in the morning are under-represented.

These accidents occur almost entirely outside of built-up areas, and tend to occur on 100 km/hour roads as opposed to roads with other speed limits. Federal authorities are clearly more likely than expected to be responible for the road; other authorities are under-represented. Gelderland, Utrecht, and Limburg are over-represented, and Groningen is under-represented (?). There is no apparent relation with city size.

These accidents are quite likely to occur during daylight conditions, and there is a slightly lessened chance that they occur in the vicinity of street lights. The slight chance of mist is even less in this cluster. The road surface is very likely dry, and the small chance that the road surface itself consists of 'clinkers' is even less in this case.

These accidents are more likely to occur on a straight road section, and quite unlikely to occur in a curve or at an intersection. We won't even mention temporary road situations.

Alcohol involvement in these accidents in much less likely than expected, and the driver is more likely than expected to be a female of unspecified age.

The driver is more likely to be on the right hand side of a multiple lane road, on the left hand side of a two-lane road, or on a road exit/ entrance. There is a slightly heightened chance that the driver is changing lanes. The vehicle is more likely to overturn, and the driver is more likely to be blamed for loosing control of the wheel or some 'other' error.

The CBS manoeuvre variable indicates in this case that the driver just ran off of the road.

Fortunately, these accidents aren't especially fatal nor is there any special chance of having to be taken to the hospital.

Summary

These day-time, fair weather accidents tend to occur on multi-lane 100 km per hour roads outside built-up areas, yet are not clearly rural in nature. The vehicle may also have been at a road exits/entrances or on the wrong side of the road. These accidents tend to occur on straight road sections, when the driver, who is more likely to be female and less likely to have been drinking, apparently is merging or changes lanes and 'looses control', runs off the road, and possible overturns.

Cluster AN3

This cluster is very small and contains only 68 cases. One could arguably choose to ignore it altogether, due to the statistical problems involved with such small numbers. We will nevertheless present the results, leaving it to the reader to decide whether the following description is informative or not.

There is no apparent relation with day of the week, and no clear relation with month of the year even though february is over-represented. These accidents are over-represented between 10 in the morning and noon.

They tend to occur more frequently inside built-up areas, in contradiction to this entire group of single car accidents. They occur primarily on 50 km/hour roads, administered by city officials. North and South Holland are over-represented. Cities with more than 50,000 inhabitants are also over-represented, and cities with less than 20,000 inhabitants are underrepresented.

There is no apparent relation with light conditions, yet it is more likely that street lights are located in the vicinity. There is also no relation with weather conditions, nor the condition of the road surface. Comments about the surface itself can not be made due to statistical problems.

These accidents have a heightened chance of occurring at an intersection, and no statement can be made with respect to temporary road conditions.

There is no relation with alcohol use, sex, or age of the driver.

It is unclear if there is any relation with location of the vehicle on the road while there is a slightly lesser chance of the accident occurring on a multi-lane road. Statistical problems hinder statements about the vehicles manoeuvre. The vehicle itself has no special chance of overturning, and some 'other' type of error is more likely than expected to have occurred.

The CBS manoeuvre variable indicates that the vehicle had a higher chance of skidding.

These accidents are not especially fatal, yet there is fortunately a smaller chance that transportation to the hospital would have to be arranged. Lighter injuries are more likely than expected.

<u>Summary</u>

These accidents are clearly urban and tend to occur inside built-up areas, in contradiction to the previous two categories, and tends to be less serious in nature. Otherwise, statistical problems makes interpretation difficult.

<u>Cluster AN4</u>

This cluster is also rather small, containing only 64 cases, and is also subject to statistical problems concerning the interpretation.

Thursday is over-represented and Sunday is under-represented. The winter months, December, January, and February, are all over-represented, and (almost) all other months are significantly under-represented. (In fact, with only the exception of one case, no accident in this cluster occurred between April and November). This accident also tends to primarily occur between 7 and 10 o'clock in the morning. The hours between noon and 7 in the morning are all under-represented.

There is no relation with built-up areas, maximum speed limit, or road adminstration. There is also no clear relation with the province in which the accident occurs, even though South Holland is under-represented. City size is also apparently irrelevant.

These accidents are more likely than expected to occur during daylight as opposed to during the dark, and there is no relationship with the presence of street lighting. Statistical problems make it difficult to say anything anything about weather conditions. However, the road surface is not dry, and is unlikely to even be wet. 'Other' conditions predominate, and a check of the original variables reveals that 'other' most likely means snow or ice in this case. The nature of the road surface itself is irrelevant.

The road situation is irrelevant, and temporary road situations are equally irrelevant.

It is less likely that the driver had been drinking, and the driver is more likely to be a woman. Age groups between 26 and 40 years of age are over-represented.

The cars' position on the road is not relevant, nor is the manoeuvre being implemented. Cars in this cluster are neither more nor less likely to overturn than the average single car-no object accident. The driver is very likely to have gone into a skid, which is clearly over-represented All other accident 'causes' are under-represented.
Oddly enough, the CBS manoeuvre variable sheds no special insight into this cluster of accidents, even though there is a special category for single vehicle skidding accidents. The reason for this discrepancy has not been resolved.

This type of accidents does not distinguish itself in terms of fatalities or injuries.

Summary
This winter month, morning accident has no other special characteristics except that the road surface is likely to be covered with snow or ice. The driver, who has not been drinking and is more likely to be a middle-aged woman, apparently skidded on a slippery surface.

9. DISCUSSIONS AND CONCLUSIONS

An evaluation of the results presented in the previous sections, both positive and negative, can be grouped under a number of headings: procedural, methodological, and substantive. However, before that one delves into these considerations, a preliminary remark should be made. Namely, the primary goal of this study was to provide a number of traffic accident scenarios, by means of data-analytic clustering of traffic accident records. As such, this goal has been achieved (see Appendix III for a summary), producing scenarios that are sometimes rather obvious, occasionally somewhat vague and, every now and then, even quite surprising. (It should be noted that the results mentioned in Appendix III are a list of salient accident characteristics which are found to occur more often than would be expected with respect to the marginal frequency distribution. It would be incorrect for the reader to assume that the characteristics listed there occur in most or even many of the accidents within a certain accident cluster).

9.1. Procedural evaluation

During the course of this study, with increasing experience in interpreting the results, culling options, making choices, and automating the procedure, it became increasingly easy to run an analysis from initial data reduction to the production of cross tabulations. With only two or three decision points where the researcher is required to intervene, requiring no more than 30 minutes of time, it is entirely feasible to make an entire analysis run within several hours, assuming that the computer is not over loaded. With sufficient computer capacity and an intelligent interleaving of activities, an experienced analyst could easily produce ten complete analyses per week. Noting that a new VOR accident file is only produced yearly, rather extensive analyses of this kind can be easily produced within a few months of receipt of a new yearly file.

Unfortunately, however, there are four rather large bottlenecks limiting such an optimistic view (assuming that we would want to continue such analyses).

The first bottleneck has to do with recoding the original file into a consistent form usable by the present statistical software. The SWOV SIDO project (Kars, 1990) has made great strides in making the original VOR file directly readable by the currently available software (SAS), however this is not sufficient. First of all, HOMALS, which presently has a rather poor user interface in any case, is not well integrated into the SAS system at the SWOV. This makes interpretation a tedious procedure of cross referencing variable and category numbers with code books. In the present application, one was often forced to compute two or three cross references before one could determine the meaning of a specific category code. In addition to being time-consuming, this extra complication is not conducive to accuracy. Secondly and perhaps more importantly, categories are difficult to code consistently across analyses. This has primarily to do with the fact that the distribution of characteristics differs over conflict types. For example, the distribution of ages of pedestrians, moped riders, riders, and car drivers differ dramatically as does their location on the and their manoeuvres. This implies that coding consistently over analyses leads to small numbers in many categories, which encourages outliers to dominate solutions and requires more computer time. Coding compactly per analysis to avoid this problem then often leads to non-uniform codings over analyses, which gives rise to the problems mentioned in the previous point.

The second bottleneck concerns the problem of striving for procedural consistency over analyses. It occurred regularly that a problem was detected on, for example, the analysis of the third conflict type, and to remain consistent, it's correction had also to be applied on the preceding two analyses. This phenomenon resulted in having to make several unplanned passes through preceding analyses. This required time which had to be debited against other activities, such as interpretation and write-up. Some agreement between researchers concerning which effects should be viewed as substantial and which as artifactual could help to alleviate this problem.

A third bottleneck concerns interpretation. Since each analysis involved about 30 variables and generated about 4 accident types, it is clear that several days can be invested in considering the outcome of an analysis that only takes a few hours to run. There is no clear way to circumvent this problem.

A fourth problem involves presentation. This researcher would have been satisfied with describing the HOMALS dimensions (spatially), and then describing the relation between these dimensions and the discriminant functions which separated the found clusters. Alternatively, a logit analysis of the relation between the original variables and the clusters would also have been satisfactory. Nevertheless, it was felt that describing the clusters by means of a number of cross tabulations would make the result more accessible for many. Unfortunately, with six analyses and about 30 variables per analysis, this would mean presenting about 180 cross tables, which can't be easily viewed as a compact solution to the problem. (It is interesting to point out that the results presented here are actually a condensation of 2610 original cross tabulations!) The present 'solution' of describing all of the accident types verbally (and presenting a selection of cross tables for illustration) can make the results more accessible to others, yet can hardly be described as a rigorous style of presentation. This problem has to be solved if there is to be any hope of a wide spread use of this kind of multivariate technology for this type of problem.

9.2. Methodological evaluation

Assuming that the problems mentioned in the previous section (e.g., consistency, choice of dimensionality, determining something is substantial or artifactual, manner of presentation, etc.) didn't exist or were solved, one could still concern oneself with whether the data reduction methodology was satisfactory or whether it could (and should) be improved.

In the first place, it is clear that the procedure followed (first optimal scaling via HOMALS and then explicitly clustering via a clustering technique) is a large improvement on previous applications where HOMALS object score plots were only 'eye-balled' in order to detect distinct clusters of objects (accidents). (Actually, most applications only utilize the copious presentation of cross tabulations). As such, the methodology presented here (or some improvement thereof) should be utilized when appropriate.

A second point concerns the choice of HOMALS (as a scaling technique) and hierarchal Centroid cluster analysis (as a clustering technique) in the place of other similar but alternative techniques. (This author recently learned of the existence of the GROUPALS algorithm (Van Buren & Heiser, 1989) which could accomplish these two steps simultaneously). It would be surprising, if one looked long and hard enough, if the present choice of algorithms and parameters turned out to be optimal (in the sense of cluster size and intepretability). Even so, such a search could easily demand a multiple of the resources invested in this initial application, unless one is lucky or more able.

More interesting is the possibility of using other algorithms which offer greater possibilities in deriving groups of objects. For example, the technique used here, unless otherwise explicitly determined, only considers relations between pairs of variables. It is easily conceivable that the really interesting clusters can only be discovered if one also consider the higher order interactions between variables. But how does one know which interactions are interesting?
This problem is hardly trivial due to the combinatorial explosion.

Another example is that, in the previous analyses, it occurred repeatedly that the distinction between day and night and between inside and outside built-up areas played an important role. While these two distinctions may be quite important theoretically, it is also possible that they achieved such a pervasive role because variables pertaining to these distinctions were sampled more often than we would wish. For illustration, infrastructural characteristics are important for describing an accident, but they are also rather permanent and easily measurable. Now, if we measured 20 infrastructural variables and one behavioral variable, then we shouldn't be surprised that infrastructural aspects dominate an analysis. In order to alleviate this 'artificial' dominance of some groups of variables, one could proceed to partition variables into a priori groups and allow each group a potentially equal 'vote' in the final classification. Of course, such a partitioning requires some kind of priori argumentation.

These two aspects have a tendency to oppose each other: modularity decreases interactions while considering higher-order interactions increases interactions.

There are a number of algorithms and combinations of algorithms which are capable of addressing one or both of these two aspects: one thinks of OVERALS (Verdegaal, 1986), genetic learning (Holland et al., 1986), modular competitive learning (Murre et al., 1989), etc. Use of these algorithms introduces an extra piece of subjectiveness, namely choosing an appro- priate architecture and loss function, yet this is unavoidable. Furthermore, to avoid committing the resources demanded of a systematic search of solution spaces, one would have to make a number of educated guesses. Nevertheless, it would be quite interesting to note whether different procedures could illuminate different and interesting aspects of the data set in question.

9.3. Substantial evaluation

Finally, one could ask whether the previous exercise achieved any new and/or useful results. This question isn't easy to answer.

A Kuhnian answer could be that the results are useful if colleagues agree that they are useful. A pragmatic corollary could be that the results are useful if colleagues actually use the results.
Aside from other considerations, at least five factors could inhibit dissemination and use of these findings. These factors could be designated as:

- classical conceptions of categorization;
- context-sensitivity;
- theoretical congruence;
- the amount of experience with the phenomenon in question, and
- presentation technology.

While the present report is not the appropriate place for a detailed discussion of these four points, we will nevertheless elucidate them in passing.

• Classical conceptions of categorization encompass the idea that categories or concepts consists of defining as well as incidental features. Once we know which characteristics define something, then we 'understand' what that something is. Unfortunately for this viewpoint, modern philosophers and psychologists are more likely to conclude that concepts are polythemic and probabilistic. Lay people also seem likely to believe that

defining characteristics exist, only no one seems to be able to point them out, as Socrates enjoyed demonstrating.
The typology derived in this study is typically polythemic and probabilisstic.

• Many modern theorists point out that typologies and taxonomies are mainly useful in terms of the goals which they serve (or the context in which they are placed). (Thus, if one is interested in alcohol-related accidents, the universe of accidents can be more or less divided into two major categories). This position certainly contains a great deal of truth, and is certainly supported by some psychological research. Even so, a number of widely accepted, non-goal directed taxonomies comes easily to mind, e.g., the phylogenetic tree. Theory driven taxonomies can be immensely powerful, yet this presupposes the existence of a powerful theory. One could easily suspect no such general theory exists for the traffic safety problem. Objections to a more or less data-driven typology, such as the one presented here, and the dearth of theory driven alternatives, could serve to make this more (painfully) clear.

• Acceptance of a new scheme, theory, typology, or whatever is facilitated by congruence with what is known (implicitly or explicitly) by the research community. It would be interesting to know which shortcomings of the typology presented above, aside from those already presented, would be pointed out by other researchers, as well as whether researchers are capable of agreeing on any classificatory strategy whatsoever.

• Direct experience with many individual cases of a phenomenon may facilitate the ability to identify classes. Most researchers using cross tables, on the other hand, rarely ever see a complete accident description.

• Poor presentation leads to poor retention and understanding. The present presentation certainly leaves a great deal to be desired, and it not clear which presentation form would be more easily understandable by researchers and policy makers.

In addition to the preceding, rather general, arguments one could ask whether analyses was able to achieve some interesting results.

In the first place, a number of useful, albeit not necessarily surprising, distinctions were found. E.g., contrasts between urban and rural roads, young and old road users, daytime and nighttime accidents, and inter-

section and non-intersection (and their corresponding manoeuvres) accidents were quite often found. In fact, it would have been rather surprising ex post facto (noting the composition of the variables included in the analyses) if the analyses had not highlighted the importance of infrastructure, diurnal and seasonal patterns, demographics, and patterns of mobility.
Unfortunately, a number of possible distinctions turned out to be rather fuzzy, and could only be (rather incompletely) inferred. The type and function of the road itself is a case in point. (Other researchers are also well aware of this point (OECD, 1988) In addition, the relative dearth of behavioral variables is a cause for concern.

Secondly, a number of more or less well-known accident scenarios were also detected by this analysis. A case in point is the young child mid-block dart-out. (Unfortunately, this scenario was somewhat blurred due to recoding of the street-light variable, which originally functioned as a lynch-pin emphasizing the day-night distinction. The de-emphasis of this distinction lead some merging of the young child and the nighttime adult mid-block dart-out mentioned in a preliminary study (Gundy, 1989).

Third of all, a number of surprises occurred. Namely, the archetypal alcohol accident is the young male, single moving automobile, rural, nighttime, run-off-the-road-in-a-curve type of accident in which a tree is involved. While the marginal frequencies indicate that single-moving-automobile type of accident indeed frequently involves alcohol use, the situation is clearly more complex. When the vehicle collides with another object, alcohol use is clearly over-represented in a cluster involving a more urban, bad weather, type of accident. When the vehicle doesn't strike anything (such as driving into a ditch), the more classical stereotype of the alcohol-involved accident is obtained (with the exception that no tree is involved). The classical rural, runoff-the-road on a curve into a tree type of accident tends to occur more frequently during the daytime, and is alcohol-under-involved!
It is unknown whether this unexpected finding is due to:
- the probabilistic character of the analysis method (i.e., we could have exactly defined the archetypal alcohol accident and seen whether alcohol use was indeed over- or even frequently involved);
- an artifact of police bias in reporting alcohol involvement;

or

- an actual reflection of Dutch population density and infrastructural characteristics.

This last possibility is especially intriguing.

A fourth point should be made. Namely, it was rather disappointing to find that almost all of the preceding analyses generated a large, dominant first cluster which often included more than 50% of the accidents within it's analyzed conflict type. This could force us to the conclusion that the variety of the analyzed variables was not sufficient to achieve a more subtle and interesting differentiation. Another conclusion could be that accidents, for the most part, are actually quite similar at this level of analysis. These alternatives are not contradictory.
Besides using other analysis techniques, another option is possible. Namely, the clustering technique used is a hierarchal technique, and the clusters described here are at a certain, high level in that hierarchy. One could conceivably describe the more numerous clusters lower in that hierarchy. That was not done in the present analysis due to the increasing burden of description (and problems with even smaller numbers). Nevertheless, one could choose to pursue the possibility of making finer distinctions within the largest branch of that hierarchal tree.

In conclusion, this study achieves what it was originally intended to do: produce a quick and dirty distinction between accident types that may be serve as a reference guide to orient further, in situ, research. Furthermore, it tends to emphasize not only more or less self-evident distinctions, e.g., that different types of manoeuvres are made on intersections as opposed to straight road sections, but also more general abstractions. For example, one could consider age-relevant differences in mobility patterns or the urban-rural distinction, which is much broader than only that of inside versus outside built up areas.
Of course, it is an open question whether or not these results are generalizable over the years, over other countries, or over other accident files containing other variables or accidents sampled in another way. Even so, a number of patterns detected here could be said to reflect (some) structural and enduring characteristics of the Dutch road system as well as road user behaviour. As such, these results are clearly neither chimerical nor irrelevant.

Whether these findings are sufficiently complete, clear, and imperative to enjoy wide-spread acceptance, is another question entirely. In it's most strict form, this author believes that the answer to this question is negative. E.g., the VOR accident file does not include many relevant accident variables, such as seat belt use, and therefore the present analyses necessarily neglect many important issues. It is nevertheless felt that, the obvious shortcomings aside, the major value of this study could be to stimulate discussion concerning major theoretical and structural agreements and discrepancies in the ways that researchers and policy makers think about traffic accidents. In other words, a confrontation between a (more or less) purely data-driven accident typology and the often (more or less) implicit theories about traffic accidents could lead to an improved <u>theory-driven</u> typology. While the determination and reconciliation of the results of such a confrontation is beyond the scope of this study, it is felt that such an exercise is to everyone's benefit.

BIBLIOGRAPHY

CBS (1984). Statistiek van de verkeersongevallen op de openbare weg 1982. Staatsuitgeverij, 's-Gravenhage.

Everitt, B.S. (1980). Cluster analysis. Second Edition. Heineman Books Ltd., London.

Gifi, A. (1981). Nonlinear multivariate analysis. Departement of Data Theory, University of Leyden.

Gundy, C.M. (1989). Towards a multivariate, similarity-based traffic accident typology. In: Mühlrad, N. (ed.) (1989). Proceedings of the Second European workshop on Recent developments in road safety research, 25-27 January 1989, Paris. Actes INRETS No. 17, pp. 93-100. Institut National de Recherche sur les Transports et leur Securité, Joinville.

Gundy, C.M. (In preparation). A bi-directional constraint satisfaction model of implicit inference and classification.

Holland, J.; Holyoak, K.; Nisbett, R. & Thagard, P. (1986). Induction : Processes of inference, learning, and discovery. MIT Press / Bradford Books.

Kars, V. (1990). SIDO-TVIS; Gebruikershandleiding Snelle informatie & documentatie voor onderzoekers, Tweede VOR informatie-systeem. SWOV, Leidschendam (Not published).

Maas, M. & Harris, S. (1984). Police recording of road accident in-patients; Investigation into the completeness, representativity, and reliability of police records of hospitalized traffic victims, Accid. Anal. & Prev. 16 (3): 167-184.

Murre, J.M.; Phaf, R.H. & Wolters, G. (1989). Calm networks; A modular approach to supervised and unsupervised learning. Proceedings of the 1989 Joint International Conference on Neural networks I: 649-656.

OECD (Scientific Expert Group T8; Chairman: S. Harris, SWOV) (1988). Framework for consistent traffic an accident statistical bases. SWOV, Leidschendam.

Oude Egberink, H.J.H.; Van der Linden, L.H.K.; Van der Molen, H.H. & Lourens, P.F. (1988). Menselijke factoren bij ongevallen op de A-28; Deel 1: Globale analyse. Report VK 88-05. Verkeerskundig Studie Centrum (VSC), Groningen State University.

SAS Institute Inc. (1985). SAS User's Guide: Statistics. Version 5 Edition. Cary, North Carolina.

Treat, J.R.; Tumbas, N.S.; McDonald, S.T.; Shinar, D.; Hume, R.D.; Mayer, R.E.; Stansifer, R.L. & Castellan, N.J. (1977). Tri-level study of the causes of traffic accidents: Final report. Vol.I: Causal factor tabulations and assessments; Vol.II: Special analyses. Report No. DOT-HS-034-3-535-77-TAC. Institute for Research in Public Safety, Indiana University.

Van Buren, S. & Heiser, W.J. (1989). Clustering N objects into K groups under optimal scaling of variables. Psyschometrika 54 (4): 699-706.

Verdegaal, R. (1986). OVERALS. Departement of Data Theory, Leyden State University.

TABLES 1 TO 6

Table 1. Chi^2-values for the cross-tabulations of cluster numbers versus their original variables for auto-bicycle accidents (N = 6333)

Table 2. Chi^2-values for the cross-tabulations of cluster numbers versus their original variables for auto-moped accidents (N = 6263)

Table 3. Chi^2-values for the cross-tabulations of cluster numbers versus their original variables for auto-pedestrian accidents (N = 2900)

Table 4. Chi^2-values for the cross-tabulations of cluster numbers versus their original variables for auto-auto accidents (N = 4655)

Table 5. Chi^2-values for the cross-tabulations of cluster numbers versus their original variables for single auto-object accidents (N = 3162)

Table 6. Chi^2-values for the cross-tabulations of cluster numbers versus their original variables for single auto-no object accidents (N = 851)

Variable Name	Chi^2	D.f.	Prob	#Missing
Day of the week	72.6	18	0.000	496
Month	294.7	33	0.000	496
Hour of the day	362.9	51	0.000	525
Built-up area	4476.0	3	0.000	496
Maximum speed	4597.6	9	0.000	515
Road authority	864.5	9	0.000	496
Type of road section	1190.5	12	0.000	500
Unusual road situation	689.2	12	0.000	496
Temporary road situation	6.0	3	0.110	496
Light conditions	353.3	6	0.000	498
Street light present	450.3	3	0.000	528
Weather	167.5	6	0.000	539
Condition road surface	188.3	6	0.000	542
Type road surface	203.9	9	0.000	539
CBS manoeuvre	2197.5	42	0.000	496
Province	307.2	33	0.000	496
City size	1019.5	15	0.000	496
Number of dead	130.6	3	0.000	496
Number of hospital injured	265.8*	6	0.000	496
Number of non-hospital injured	304.0	6	0.000	496
Number of unknown injured	16.0*	6	0.014	496
Driver's alcohol use	99.3	9	0.000	496
Driver's position	1050.6*	27	0.000	552
Driver's manoeuvre	3184.0	15	0.000	498
Location contact on car	672.3	21	0.000	501
Driver's blame	5219.7	24	0.000	519
Driver's sex	7.0	3	0.072	538
Driver's age	62.5	36	0.004	577
Bicyclist's alcohol use	41.1*	9	0.000	496
Bicyclist's position	501.7	21	0.000	750??
Bicyclist's manoeuvre	1574.9	21	0.000	498
Location contact on bicycle	659.6	9	0.000	499
Bicyclist's blame	4158.2	18	0.000	496
Bicyclist's sex	122.2	3	0.000	496
Bicyclist's age	1266.1	48	0.000	509

* More than 25% of the table cells have an expected value less than five. These chi^2-values are suspect.
? A coding error incorrectly eliminated a number of cases.

Table 1. Chi^2-values for the cross-tabulations of cluster numbers versus their original variables for auto-bicycle accidents (N = 6333)

Variable Name	Chi^2	D.f.	Prob	#Missing
Day of the week	80.3	24	0.000	209
Month	114.7	44	0.000	209
Hour of the day	190.3	68	0.000	246
Built-up area	4395.4	4	0.000	209
Maximum speed	4440.9	12	0.000	237
Road authority	1209.4	12	0.000	209
Type of road section	1900.0	16	0.000	212
Unusual road situation	1528.5	16	0.000	209
Temporary road situation	4.6	4	0.330	209
Light conditions	37.7	8	0.000	211
Street light present	469.0	4	0.000	224
Weather	43.0	8	0.000	250
Condition road surface	17.7	4	0.001	257
Type road surface	385.4	12	0.000	252
CBS manoeuvre	3471.0	64	0.000	209
Province	270.2	44	0.000	209
City size	1029.0	20	0.000	209
Number of dead	160.7*	4	0.000	209
Number of hospital injured	199.1	8	0.000	209
Number of non-hospital injured	187.0	8	0.000	209
Number of unknown injured	10.5	8	0.232	209
Driver's alcohol use	57.5	12	0.000	209
Driver's position	1948.8	16	0.000	256
Driver's manoeuvre	4659.8	32	0.000	209
Location contact on car	1486.5	32	0.000	213
Driver's blame	5741.3	24	0.000	210
Driver's sex	18.4	4	0.001	229
Driver's age	138.2	48	0.000	245
Moped rider's alcohol use	54.7	12	0.000	209
Moped rider's position	2120.5	32	0.000	285
Moped rider's manoeuvre	1380.0	24	0.000	211
Location contact on moped rider	401.8	12	0.000	211
Moped rider's blame	5469.6	32	0.000	209
Moped rider's sex	22.3	4	0.000	212
Moped rider's age	208.0	44	0.000	217

* More than 25% of the table cells have an expected value less than five. These chi^2-values are suspect.

Table 2. Chi^2-values for the cross-tabulations of cluster numbers versus their original variables for auto-moped accidents (N = 6263)

Variable Name	Chi^2	D.f.	Prob	#Missing
Day of the week	61.1	18	0.000	211
Month	93.1	33	0.000	211
Hour of the day	253.5	45	0.000	238
Built-up area	2359.8	3	0.000	211
Maximum speed	2298.0*	9	0.000	223
Road authority	467.4	9	0.000	211
Type of road section	741.4	12	0.000	213
Unusual road situation	1832.1*	15	0.000	211
Temporary road situation	14.1*	3	0.003	211
Light conditions	101.7	6	0.000	212
Street light present	251.5	3	0.000	227
Weather	51.3*	6	0.000	243
Condition road surface	95.1*	6	0.000	242
Type road surface	140.4*	6	0.000	246
CBS manoeuvre	3438.4	21	0.000	211
Province	135.8*	33	0.000	211
City size	362.9	15	0.000	211
Number of dead	52.0	3	0.000	211
Number of hospital injured	39.9	3	0.000	211
Number of non-hospital injured	59.1*	6	0.000	211
Number of unknown injured	2.6	3	0.464	211
Driver's alcohol use	5179.7*	9	0.000	211
Driver's position	229.8*	9	0.000	234
Driver's manoeuvre	587.6*	15	0.000	214
Location contact on car	347.5*	18	0.000	220
Driver's blame	2570.2*	18	0.000	213
Driver's sex	3.46	3	0.326	234
Driver's age	62.3	36	0.004	259
Pedestrian's alcohol use	45.2	6	0.000	211
Pedestrian's position	3458.8	21	0.000	229
Pedestrian's manoeuvre	1719.3	15	0.000	211
Location contact on pedestrian	not used			
Pedestrian's blame	3034.7	15	0.000	213
Pedestrian's sex	43.2	3	0.000	212
Pedestrian's age	463.1	33	0.000	232

* More than 25% of the table cells have an expected value less than five. These chi^2-values are suspect.

<u>Table 3</u>. Chi^2-values for the cross-tabulations of cluster numbers versus their original variables for auto-pedestrian accidents (N = 2900).

Variable Name	Chi^2	D.f.	Prob	#Missing
Day of the week	116.8	24	0.000	387
Month	117.4	44	0.000	387
Hour of the day	655.4	68	0.000	405
Built-up area	3326.8	4	0.000	387
Maximum speed	3308.4	12	0.000	399
Road authority	1292.2	12	0.000	387
Type of road section	1552.7	16	0.000	391
Unusual road situation	100.6	8	0.000	387
Temporary road stiuation	12.3	4	0.015	387
Light conditions	546.8	8	0.000	288
Street light present	1327.0	4	0.000	398
Weather	202.1	8	0.000	404
Condition road surface	194.3	8	0.000	403
Type road surface	291.2	8	0.000	402
CBS manoeuvre	5058.7*	55	0.000	387
Province	537.0	44	0.000	387
City size	1337.1	20	0.000	387
Number of dead	148.8*	8	0.000	387
Number of hospital injured	276.7*	28(!)	0.000	387
Number of non-hospital injured	129.4	12	0.000	387
Number of unknown injured	8.8	8	0.360	387
First driver's alcohol use	495.2	8	0.000	466
First driver's position	719.4	20	0.000	456
First driver's manoeuvre	807.8	28	0.000	387
Location contact on first car	625.8	28	0.000	496
First driver's blame	5375.7	28	0.000	388
First driver's sex	35.3	4	0.000	394
First driver's age	302.8	48	0.000	405
Second driver's alcohol use	420.3*	12	0.000	480
Second driver's position	235.8	8	0.000	413
Second driver's manoeuvre	3199.5	16	0.000	432
Location contact on second car	3848.0	28	0.000	392
Second driver's blame	338.0	4	0.000	387
Second driver's sex	6.9	4	0.142	390
Second driver's age	133.3	48	0.000	396

* More than 25% of the table cells have an expected value less than five. These chi^2-values are suspect.

Table 4. Chi^2-values for the cross-tabulations of cluster numbers versus their original variables for auto-auto accidents (N = 4655).

Variable Name	Chi^2	D.f.	Prob	#Missing
Day of the week	21.4	12	0.044	203
Month	52.2	22	0.000	203
Hour of the day	201.6	46	0.000	232
Built-up area	2321.1	2	0.000	203
Maximum speed	3868.4	6	0.000	211
Road authority	1229.9	6	0.000	203
Type of road section	311.2	6	0.000	208
Unusual road situation	36.8	8	0.000	203
Temporary road situation	12.5	2	0.002	203
Light conditions	124.6	4	0.000	205
Street light present	725.6	2	0.000	216
Weather	41.9	6	0.000	231
Condition road surface	24.2	4	0.000	226
Type road surface	281.9	4	0.000	218
CBS manoeuvre	1475.5	16	0.000	203
Province	418.7	22	0.000	203
City size	728.7	10	0.000	203
Number of dead	81.5	4	0.000	203
Number of hospital injured	93.9	4	0.000	203
Number of non-hospital injured	104.1	4	0.000	203
Number of unknown injured	37.9	4	0.000	203
Driver's alcohol use	108.7	6	0.000	203
Driver's position	1726.5*	12	0.000	208
Driver's manoeuvre	287.1	10	0.000	214
Location contact on car	159.8	14	0.000	209
Driver's blame	122.2	14	0.000	210
Driver's sex	3.99	2	0.136	205
Driver's age	59.7	26	0.004	211
Type of object	1031.5*	14	0.000	203
Location of object	2845.1	18	0.000	208

* More than 25% of the table cells have an expected value less than five. These chi^2-values are suspect.

Table 5. Chi^2-values for the cross-tabulations of cluster numbers versus their original variables for single auto-object accidents (N = 3162)

Variable Name	Chi^2	D.f.	Prob	#Missing
Day of the week	68.0	18	0.000	22
Month	234.3	33	0.000	22
Hour of the day	396.6	15	0.000	29
Built-up area	508.8	3	0.000	22
Maximum speed	760.5	9	0.000	23
Road authority	291.4	9	0.000	22
Type of road section	179.6	6	0.000	22
Unusual road situation	not used			
Temporary road situation	51.0*	3	0.000	22
Light conditions	129.5	6	0.000	22
Street light present	57.7	3	0.000	27
Weather	60.8*	9	0.000	33
Condition road surface	563.3	6	0.000	28
Type road surface	25.0*	6	0.000	28
CBS manoeuvre	130.4	12	0.000	27
Province	112.6*	33	0.000	22
City size	144.5	15	0.000	22
Number of dead	8.1*	3	0.044	22
Number of hospital injured	25.0	6	0.000	22
Number of non-hospital injured	17.6	6	0.007	22
Number of unknown injured	4.3*	6	0.633	22
Driver's alcohol use	81.9	9	0.000	22
Driver's position	315.6*	12	0.000	22
Driver's manoeuvre	107.3*	12	0.000	23
Location contact on car	37.7	3	0.000	22
Driver's blame	336.9	15	0.000	22
Driver's sex	73.8	3	0.000	24
Driver's age	44.6	24	0.007	25
Type of object	not applicable			
Location of object	not applicable			

* More than 25% of the table cells have an expected value less than five. These chi^2-values are suspect.

Table 6. Chi^2-values for the cross-tabulations of cluster numbers versus their original variables for single auto-no object accidents (N = 851).

APPENDICES I - III

Appendix I. Description of the accident manoeuvres (Omschrijving van de ongevallenmanoeuvres) (Source: CBS).

Appendix II. Examples of cluster features.

Appendix III. Summary of the results.

APPENDIX I. DESCRIPTION OF THE ACCIDENT MANOEUVRES (SOURCE: CBS)

Ongevallenmanoeuvres

Omschrijving van de ongevallenmanoeuvres

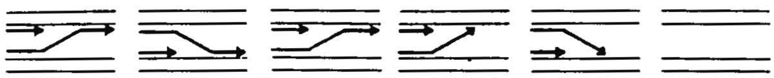
Hoofd-groep	Manoeu-vre	Omschrijving
1		*Botsingen tussen voertuigen op dezelfde weg in dezelfde richting zonder afslaan*
	111	Kop/staart botsing in vol verkeer
	121	Kop/staart botsing met remmend voertuig
	122	Kop/staart botsing met stilstaand of voorgesorteerd voertuig
	131	Botsing bij links inhalen
	132	Botsing bij rechts inhalen
	133	Botsing bij dubbel inhalen
	134	Schampbotsing
	141	Botsing bij veranderen van rijstrook naar links
	142	Botsing bij veranderen van rijstrook naar rechts
	151	Botsing met voertuig, dat optrekt van rechter wegzijde
	161	Botsing met invoegend voertuig bij doorgaand verkeer
	162	Botsing met uitvoegend voertuig bij doorgaand verkeer
	199	Overige botsingen binnen hoofdgroep 1
2		*Botsingen tussen voertuigen op dezelfde weg in tegengestelde richting zonder afslaan*
	211	Frontale botsing zonder rijstrookverandering
	212	Frontale botsing met rijstrookverandering van een voertuig
	213	Frontale botsing met rijstrookverandering van beide voertuigen
	221	Schampbotsing
	231	Botsing bij invoegen vanuit stilstand
	241	Botsing bij invoegen van doorgaand verkeer
	299	Overige botsingen binnen hoofdgroep 2
3		*Botsingen tussen voertuigen op dezelfde weg in dezelfde richting met afslaan*
	311	Rechtsafslaand voertuig wordt van achteren aangereden
	312	Rechtsafslaand voertuig wordt opzij aangereden
	313	Botsing van twee rechtsafslaande voertuigen
	321	Linksafslaand voertuig wordt van achteren aangereden
	322	Linksafslaand voertuig wordt opzij aangereden
	323	Botsing van twee linksafslaande voertuigen
	331	Botsing van kerende voertuigen met voertuig uit dezelfde richting
	399	Overige botsingen binnen hoofdgroep 3
4		*Botsingen tussen voertuigen op dezelfde weg in tegengestelde richting met afslaan*
	411	Botsing van linksafslaand met rechtdoorgaand voertuig
	421	Botsing van rechtsafslaand met rechtdoorgaand voertuig
	431	Botsing van twee afslaande voertuigen
	441	Botsing bij omkeren voor tegemoetkomend voertuig
	451	Botsing met dwars overstekend voertuig
	499	Overige botsingen binnen hoofdgroep 4
5		*Botsingen tussen voertuigen op kruisende wegen of uitrit, zonder afslaan*
	511	Botsing van twee rechtdoorgaande voertuigen
	512	Botsing van rechtdoorgaand voertuig met remmend voertuig
	513	Botsing van rechtdoorgaand voertuig met stilstaand voertuig
	521	Botsing van rechtdoorgaande voertuigen met rijstrookverandering van een of beide voertuigen
	531	Botsing met trein op bewaakte of onbewaakte overweg
	599	Overige botsingen binnen hoofdgroep 5
6		*Botsingen tussen voertuigen op kruisende wegen of uitrit met afslaan*
	611	Botsing bij afslaan naar rechts voor van links komend voertuig
	621	Botsing bij afslaan naar rechts voor van rechts komend voertuig
	631	Botsing bij afslaan naar links voor van rechts komend voertuig
	641	Botsing bij afslaan naar links voor van links komend voertuig
	651	Botsing bij afslaan naar rechts voor van rechts komend voertuig, dat links afslaat
	699	Overige botsingen binnen hoofdgroep 6
7		*Botsingen tussen een rijdend voertuig en een geparkeerd voertuig*
	711	Botsing met een op dezelfde weg geparkeerd voertuig, van achteren
	721	Botsing met een op dezelfde weg geparkeerd voertuig, van voren
	731	Botsing met een op linker kruisende weg geparkeerd voertuig
	732	Botsing met een op rechter kruisende weg geparkeerd voertuig
	741	Botsing met geparkeerd voertuig, waarvan een portier wordt geopend (alle situaties)
	799	Overige botsingen binnen hoofdgroep 7
8		*Botsingen tussen een rijdend voertuig en een voetganger*
	811	Op een voetgangersoversteekplaats, zebra
	812	Op een voetgangersoversteekplaats, geen zebra
	813	Bij een bushalte of tramhalte
	821	Met voetganger, die achter een object vandaan komt
	822	Met voetganger, die plotseling oversteekt
	823	Met voetganger, die op andere wijze oversteekt
	831	Met voetganger, die stilstaat op de weg
	832	Met voetganger, die speelt op de weg
	841	Met voetganger, die op de rijweg loopt
	851	Met voetganger, die op het fietspad loopt
	861	Met voetganger op trottoir
	871	Met voetganger op bewaakte of onbewaakte overweg door trein
	899	Overige botsingen binnen hoofdgroep 8
9		*Botsingen tussen een rijdend voertuig en een voorwerp of dier*
	911	Met boom, hek, huis enz. links of rechts van de weg
	912	Met lichtmast of lantaarnpaal, links of rechts van de weg
	913	Met verkeersbord, links of rechts van de weg
	921	Met verkeerszuil of verkeersbord op vluchtheuvel
	931	Met vangrail of bermbeveiliging, links, rechts of midden van de weg
	941	Met overstekend dier
	951	Met los voorwerp op of langs de weg
	952	Met ander ongeval
	990	Overige botsingen
0		*Eenzijdige verkeersongevallen*
	011	Voertuig blijft op de weg, slippen
	021	Voertuig raakt van de weg af op een rechte weg
	022	Voertuig raakt van de weg af in of na een bocht
	031	Voertuig raakt van de weg af op kruising, bij uitrit enz
	041	Voertuig raakt van de weg af in het water
	042	Voertuig raakt van de weg af in greppel of sloot
	099	Overige eenzijdige ongevallen

Ongevallenmanoeuvres

1. botsingen tussen voertuigen op dezelfde weg in dezelfde richting zonder afslaan

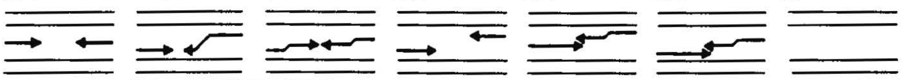
111 121 122 131 132 133 134

141 142 151 161 162 199



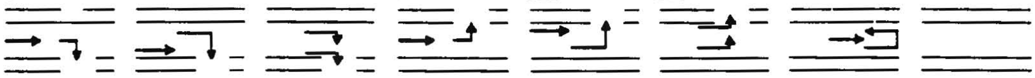
2. botsingen tussen voertuigen op dezelfde weg in tegengestelde richting zonder afslaan

211 212 213 221 231 241 299



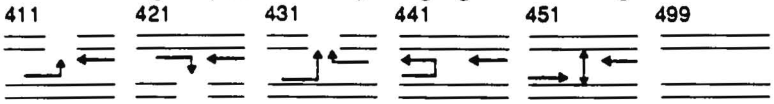
3. botsingen tussen voertuigen op dezelfde weg in dezelfde richting met afslaan

311 312 313 321 322 323 331 399



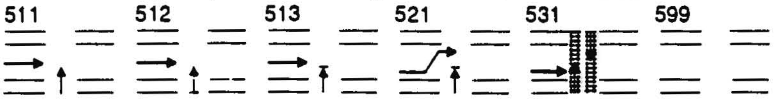
4. botsingen tussen voertuigen op dezelfde weg in tegengestelde richting met afslaan

411 421 431 441 451 499



5. botsingen tussen voertuigen op kruisende wegen of uitrit zonder afslaan

511 512 513 521 531 599

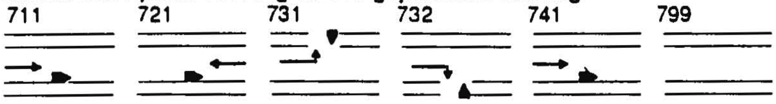


6. botsingen tussen voertuigen op kruisende wegen of uitrit met afslaan

611 621 631 641 651 699

7. botsingen tussen een rijdend voertuig en een geparkeerd voertuig

711 721 731 732 741 799



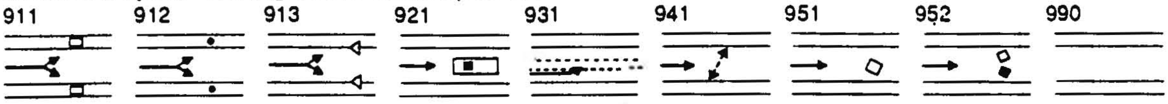
8. botsingen tussen een rijdend voertuig en een voetganger

811 812 813 821 822 823 831 832

841 851 861 871 899

9. botsingen tussen een rijdend voertuig en een voorwerp of dier

911 912 913 921 931 941 951 952 990



0. eenzijdige verkeersongevallen

011 021 022 031 041 042 099

voor beschrijving zie blz. 8.

APPENDIX II. EXAMPLES OF CLUSTER FEATURES

AUTO-BICYCLE ACCIDENTS 1982

TABLE OF CLUSTER BY MAXSNELH

CLUSTER(CLUSTERNR)   MAXSNELH(MAX.SNELHEID)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	=< 50 km	70 km	80-90 km	100 km	TOTAL
CLUSTER 1	69	19	511	11	610
	541.4	10.3	57.1	1.2	
	-472.4	8.7	453.9	9.8	
	412.223	7.4088	3604.86	84.0681	
	1.19	0.33	8.78	0.19	10.48
	11.31	3.11	83.77	1.80	
	1.34	19.39	93.76	100.00	
	1.34	19.39	93.76	100.00	10.48
CLUSTER 2	1679	21	5	0	1705
	1513.3	28.7	159.7	3.2	
	165.7	-7.7	-154.7	-3.2	
	18.1339	2.07492	149.872	3.22362	
	28.86	0.36	0.09	0.00	29.31
	98.48	1.23	0.29	0.00	
	32.51	21.43	0.92	0.00	
	33.85	40.82	94.68	100.00	39.79
CLUSTER 3	2983	54	13	0	3050
	2707.2	51.4	285.7	5.8	
	275.8	2.6	-272.7	-5.8	
	28.1052	0.13412	260.3	5.76659	
	51.27	0.93	0.22	0.00	52.42
	97.80	1.77	0.43	0.00	
	57.77	55.10	2.39	0.00	
	91.62	95.92	97.06	100.00	92.21
CLUSTER 4	433	4	16	0	453
	402.1	7.6	42.4	0.9	
	30.9	-3.6	-26.4	-0.9	
	2.37801	1.72732	16.4675	0.85648	
	7.44	0.07	0.28	0.00	7.79
	95.58	0.88	3.53	0.00	
	8.38	4.08	2.94	0.00	
	100.00	100.00	100.00	100.00	100.00
TOTAL	5164	98	545	11	5818
	88.76	1.68	9.37	0.19	100.00

FREQUENCY MISSING = 515

STATISTICS FOR TABLE OF CLUSTER BY MAXSNELH

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	9	4597.600	0.000
LIKELIHOOD RATIO CHI-SQUARE	9	2860.475	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	2039.623	0.000
PHI		0.889	
CONTINGENCY COEFFICIENT		0.664	
CRAMER'S V		0.513	

EFFECTIVE SAMPLE SIZE = 5818
FREQUENCY MISSING = 515

AUTO-BICYCLE ACCIDENTS 1982

TABLE OF CLUSTER BY WEGSITUA

CLUSTER(CLUSTERNR) WEGSITUA(WEGSITUATIE)

FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	rechte weg	kruising	T/Y kruising	verkeersplein	bocht	TOTAL
CLUSTER 1	266 170.7 95.3 53.207 4.56 43.25 16.43 16.43	169 240.6 -79.6 25.4953 2.90 27.48 7.17 7.17	162 180.6 -18.6 1.91746 2.78 26.34 9.46 9.46	0 4.5 -4.5 4.53369 0.00 0.00 0.00 0.00	18 10.5 7.5 5.27341 0.31 2.93 18.00 18.00	615 10.54 10.54
CLUSTER 2	123 473.2 -350.2 259.207 2.11 7.21 7.60 24.03	930 689.2 240.8 84.093 15.94 54.55 39.44 46.61	599 500.7 98.3 19.2927 10.27 35.13 34.97 44.42	27 12.6 14.4 16.5688 0.46 1.58 62.79 62.79	26 29.2 -3.2 .356975 0.45 1.52 26.00 44.00	1705 29.23 39.77
CLUSTER 3	863 848.5 14.5 .247794 14.80 28.23 53.30 77.33	1233 1235.8 -2.8 .006332 21.14 40.33 52.29 98.90	915 897.8 17.2 0.33102 15.69 29.93 53.42 97.84	16 22.5 -6.5 1.89548 0.27 0.52 37.21 100.00	30 52.4 -22.4 9.58143 0.51 0.98 30.00 74.00	3057 52.41 92.18
CLUSTER 4	367 126.6 240.4 456.74 6.29 80.48 22.67 100.00	26 184.3 -158.3 136.006 0.45 5.70 1.10 100.00	37 133.9 -96.9 70.1392 0.63 8.11 2.16 100.00	0 3.4 -3.4 3.36156 0.00 0.00 0.00 100.00	26 7.8 18.2 42.2893 0.45 5.70 26.00 100.00	456 7.82 100.00
TOTAL	1619 27.76	2358 40.43	1713 29.37	43 0.74	100 1.71	5833 100.00

FREQUENCY MISSING = 500

STATISTICS FOR TABLE OF CLUSTER BY WEGSITUA

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	12	1190.542	0.000
LIKELIHOOD RATIO CHI-SQUARE	12	1226.118	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	115.792	0.000
PHI		0.452	
CONTINGENCY COEFFICIENT		0.412	
CRAMER'S V		0.261	

EFFECTIVE SAMPLE SIZE = 5833
FREQUENCY MISSING = 500

AUTO-BICYCLE ACCIDENTS 1982

TABLE OF CLUSTER BY LICHTGES

CLUSTER(CLUSTERNR) LICHTGES(LICHTGESTELDH)

CLUSTER	FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	daglicht	duisternis	schemer	TOTAL
CLUSTER 1	FREQUENCY	547	53	15	615
	EXPECTED	513.1	81.7	20.2	
	DEVIATION	33.9	-28.7	-5.2	
	CELL CHI2	2.24251	10.0725	1.35503	
	PERCENT	9.37	0.91	0.26	10.54
	ROW PCT	88.94	8.62	2.44	
	COL PCT	11.24	6.84	7.81	
	CUM COL%	11.24	6.84	7.81	10.54
CLUSTER 2	FREQUENCY	1187	422	98	1707
	EXPECTED	1424.1	226.7	56.2	
	DEVIATION	-237.1	195.3	41.8	
	CELL CHI2	39.4778	168.194	31.1537	
	PERCENT	20.34	7.23	1.68	29.25
	ROW PCT	69.54	24.72	5.74	
	COL PCT	24.38	54.45	51.04	
	CUM COL%	35.62	61.29	58.85	39.79
CLUSTER 3	FREQUENCY	2697	285	74	3056
	EXPECTED	2549.5	405.9	100.6	
	DEVIATION	147.5	-120.9	-26.6	
	CELL CHI2	8.52792	36.0086	7.01383	
	PERCENT	46.22	4.88	1.27	52.37
	ROW PCT	88.25	9.33	2.42	
	COL PCT	55.40	36.77	38.54	
	CUM COL%	91.02	98.06	97.40	92.17
CLUSTER 4	FREQUENCY	437	15	5	457
	EXPECTED	381.3	60.7	15.0	
	DEVIATION	55.7	-45.7	-10.0	
	CELL CHI2	8.14787	34.4052	6.70004	
	PERCENT	7.49	0.26	0.09	7.83
	ROW PCT	95.62	3.28	1.09	
	COL PCT	8.98	1.94	2.60	
	CUM COL%	100.00	100.00	100.00	100.00
TOTAL		4868	775	192	5835
		83.43	13.28	3.29	100.00

FREQUENCY MISSING = 498

STATISTICS FOR TABLE OF CLUSTER BY LICHTGES

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	6	353.299	0.000
LIKELIHOOD RATIO CHI-SQUARE	6	338.943	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	88.344	0.000
PHI		0.246	
CONTINGENCY COEFFICIENT		0.239	
CRAMER'S V		0.174	

EFFECTIVE SAMPLE SIZE = 5835
FREQUENCY MISSING = 498

<u>AUTO-BICYCLE ACCIDENTS 1982</u>

TABLE OF CLUSTER BY GEMEKLAS

CLUSTER(CLUSTERNR)   GEMEKLAS(GEM.KLASSE)

FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	> 100.000	50 - 100	20 - 50	10 - 20	5 - 10	< 5.000	TOTAL
CLUSTER 1	17	35	167	198	127	71	615
	206.0	114.0	157.3	86.5	37.5	13.7	
	-189.0	-79.0	9.7	111.5	89.5	57.3	
	173.396	54.7475	.597396	143.715	213.513	239.731	
	0.29	0.60	2.86	3.39	2.18	1.22	10.54
	2.76	5.69	27.15	32.20	20.65	11.54	
	0.87	3.23	11.19	24.12	35.67	54.62	
	0.87	3.23	11.19	24.12	35.67	54.62	10.54
CLUSTER 2	748	363	396	156	34	10	1707
	571.7	316.4	436.6	240.1	104.1	38.0	
	176.3	46.6	-40.6	-84.1	-70.1	-28.0	
	54.3461	6.85536	3.779	29.4561	47.2139	20.6492	
	12.81	6.22	6.78	2.67	0.58	0.17	29.24
	43.82	21.27	23.20	9.14	1.99	0.59	
	38.26	33.55	26.52	19.00	9.55	7.69	
	39.13	36.78	37.71	43.12	45.22	62.31	39.78
CLUSTER 3	1015	603	835	403	164	38	3058
	1024.2	566.9	782.2	430.1	186.5	68.1	
	-9.2	36.1	52.8	-27.1	-22.5	-30.1	
	[illegible]	2.30423	3.56667	1.71013	2.71632	13.3089	
	17.39	10.33	14.31	6.90	2.81	0.65	52.39
	33.19	19.72	27.31	13.18	5.36	1.24	
	51.92	55.73	55.93	49.09	46.07	29.23	
	91.05	92.51	93.64	92.20	91.29	91.54	92.17
CLUSTER 4	175	81	95	64	31	11	457
	153.1	84.7	116.9	64.3	27.9	10.2	
	21.9	-3.7	-21.9	-0.3	3.1	0.8	
	[illegible]	.162804	4.10016	.001212	0.35092	.066358	
	3.00	1.39	1.63	1.10	0.53	0.19	7.83
	38.29	17.72	20.79	14.00	6.78	2.41	
	8.95	7.49	6.36	7.80	8.71	8.46	
	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TOTAL	1955	1082	1493	821	356	130	5837
	33.49	18.54	25.58	14.07	6.10	2.23	100.00

FREQUENCY MISSING = 496

STATISTICS FOR TABLE OF CLUSTER BY GEMEKLAS

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	15	1019.503	0.000
LIKELIHOOD RATIO CHI-SQUARE	15	945.142	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	206.541	0.000
PHI		0.418	
CONTINGENCY COEFFICIENT		0.386	
CRAMER'S V		0.241	

EFFECTIVE SAMPLE SIZE = 5837
FREQUENCY MISSING = 496

<u>AUTO-BICYCLE ACCIDENTS 1982</u>

TABLE OF CLUSTER BY VOR3MANE

CLUSTER(CLUSTERNR)   VOR3MANE(VOR3MAN1)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	rudem/lo pen	vanuit s tilstand	afremmen	afslaan rechts	afslaan links	overig	TOTAL
CLUSTER 1	588	2	16	0	2	7	615
	436.1	16.1	14.5	42.1	90.1	16.0	
	151.9	-14.1	1.5	-42.1	-88.1	-9.0	
	52.8772	12.374	.145553	42.054	86.1601	5.07913	
	10.08	0.03	0.27	0.00	0.03	0.12	10.54
	95.61	0.33	2.60	0.00	0.33	1.14	
	14.21	1.31	11.59	0.00	0.23	4.61	
	14.21	1.31	11.59	0.00	0.23	4.61	10.54
CLUSTER 2	517	106	15	319	714	35	1706
	1209.8	44.7	40.3	116.7	250.0	44.4	
	-692.8	61.3	-25.3	202.3	464.0	-9.4	
	396.771	83.9115	15.9241	350.966	861.331	2.00556	
	8.86	1.82	0.26	5.47	12.24	0.60	29.24
	30.30	6.21	0.88	18.70	41.85	2.05	
	12.49	69.28	10.87	79.95	83.51	23.03	
	26.70	70.59	22.46	79.95	83.74	27.63	39.78
CLUSTER 3	2791	27	84	43	95	17	3057
	2167.9	80.2	72.3	209.0	447.9	79.6	
	623.1	-53.2	11.7	-166.0	-352.9	-62.6	
	179.073	35.2524	1.89363	131.884	278.089	49.263	
	47.83	0.46	1.44	0.74	1.63	0.29	52.39
	91.30	0.88	2.75	1.41	3.11	0.56	
	67.45	17.65	60.87	10.78	11.11	11.18	
	94.15	88.24	83.33	90.73	94.85	38.82	92.17
CLUSTER 4	242	18	23	37	44	93	457
	324.1	12.0	10.8	31.2	67.0	11.9	
	-82.1	6.0	12.2	5.8	-23.0	81.1	
	20.793	3.02125	13.7524	1.05805	7.87506	552.424	
	4.15	0.31	0.39	0.63	0.75	1.59	7.83
	52.95	3.94	5.03	8.10	9.63	20.35	
	5.85	11.76	16.67	9.27	5.15	61.18	
	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TOTAL	4138	153	138	399	855	152	5835
	70.92	2.62	2.37	6.84	14.65	2.60	100.00

FREQUENCY MISSING = 498

STATISTICS FOR TABLE OF CLUSTER BY VOR3MANE

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	15	3183.978	0.000
LIKELIHOOD RATIO CHI-SQUARE	15	2876.479	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	101.195	0.000
PHI		0.739	
CONTINGENCY COEFFICIENT		0.594	
CRAMER'S V		0.426	

EFFECTIVE SAMPLE SIZE = 5835
FREQUENCY MISSING = 498

AUTO-BICYCLE ACCIDENTS 1982

TABLE OF CLUSTER BY VORGMANT

CLUSTER(CLUSTERNR)   VORGMANT(VORGMAN2)

CLUSTER	FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	rydenvl[illegible] pen	afremmen/ stilstaan	afslaan rechts	afslaan links	oversteken	rijbaan wisselen	oversteken dwars	overig	TOTAL
CLUSTER 1	FREQUENCY	236	10	35	235	21	15	47	[illegible]	615
	EXPECTED	348.[illegible]	14.8	22.0	171.1	6.7	7.5	35.0	[illegible]	
	DEVIATION	-112.0	-4.8	13.0	63.9	14.3	7.5	12.0	[illegible]	
	CELL CHI2	36.[illegible]	1.5[illegible]	7.6[illegible]	23.8994	30.[illegible]	7.55[illegible]	4.12048	3.74[illegible]	
	PERCENT	4.04	0.17	0.60	4.03	0.36	0.26	0.81	0.27	10.54
	ROW PCT	38.37	1.63	5.69	38.21	3.41	2.44	7.64	2.60	
	COL PCT	[illegible]	7.14	16.75	14.48	32.81	21.13	14.16	17.[illegible]	
	CUM COL%	[illegible]	7.14	16.75	14.49	32.81	21.13	14.16	17.[illegible]	10.54
CLUSTER 2	FREQUENCY	[illegible]	78	21	99	0	2	8	14	1795
	EXPECTED	985.4	40.9	61.1	474.5	18.7	20.8	97.1	27.5	
	DEVIATION	[illegible]	37.1	-40.1	-375.5	-18.7	-19.[illegible]	-89.1	-1[illegible]	
	CELL CHI2	291.7[illegible]	[illegible]	26.323	295.762	18.7119	16.951[illegible]	81.7274	5.51[illegible]	
	PERCENT	[illegible]	1.30	0.36	1.68	0.00	0.03	0.14	0.[illegible]4	29.24
	ROW PCT	[illegible]	4.35	1.[illegible]	5.74	0.[illegible]	0.12	0.47	0.[illegible]	
	COL PCT	45.0[illegible]	54.[illegible]	10.05	6.04	0.00	2.82	2.41	14.[illegible]	
	CUM COL%	52.1[illegible]	61.4[illegible]	26.79	20.52	32.81	23.94	16.57	31.[illegible]1	39.78
CLUSTER 3	FREQUENCY	1184	44	142	1277	43	39	277	51	3057
	EXPECTED	17[illegible]	73.3	109.5	850.3	33.5	37.2	173.9	49.2	
	DEVIATION	-5[illegible]	-29.3	32.5	426.7	9.5	1.8	103.1	1.8	
	CELL CHI2	17[illegible]	11.7421	9.6484	214.126	2.6746	.08735[illegible]	61.0676	.062[illegible]8	
	PERCENT	20.[illegible]	0.75	2.43	21.89	0.74	0.67	4.75	0.37	52.39
	ROW PCT	38.73	1.44	4.65	41.77	1.41	1.29	9.06	1.67	
	COL PCT	35.85	31.43	67.94	78.68	67.19	54.93	83.43	54.26	
	CUM COL%	[illegible]	92.86	94.74	99.2[illegible]	100.00	78.8[illegible]	100.00	86.[illegible]	92.17
CLUSTER 4	FREQUENCY	39[illegible]	10	11	13	0	15	0	13	457
	EXPECTED	258.[illegible]	11.0	16.4	127.1	5.0	5.6	26.0	7.4	
	DEVIATION	13[illegible].4	-1.0	-5.4	-114.1	-5.0	9.4	-26.0	5.6	
	CELL CHI2	7[illegible]	.0849[illegible]	1.76101	102.444	5.0[illegible]251	16.0229	26.0024	4.31[illegible]	
	PERCENT	6.7[illegible]	0.17	0.19	0.22	0.00	0.26	0.00	0.22	7.83
	ROW PCT	[illegible].43	2.19	2.41	2.84	0.00	3.28	0.00	2.84	
	COL PCT	[illegible]	7.14	5.26	0.80	0.00	21.1[illegible]	0.00	13.83	
	CUM COL%	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TOTAL		3338	140	209	1623	64	71	332	94	5835
		[illegible]	2.40	3.58	27.81	1.10	1.22	5.69	1.61	100.00

FREQUENCY MISSING = 498

STATISTICS FOR TABLE OF CLUSTER BY VORGMANT

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	21	1574.866	0.000
LIKELIHOOD RATIO CHI-SQUARE	21	1823.972	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	34.758	0.000
PHI		0.520	
CONTINGENCY COEFFICIENT		0.461	
CRAMER'S V		0.300	

EFFECTIVE SAMPLE SIZE = 5835
FREQUENCY MISSING = 498

AUTO-MOPED ACCIDENTS 1982

TABLE OF CLUSTER BY MAXSNELH

CLUSTER(CLUSTERNR)   MAXSNELH(MAX.SNELHEID)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	=< 50 km	70 km	80-90 km	100 km	TOTAL
CLUSTER 1	2204 1902.9 301.1 47.6559 36.57 98.83 42.86 42.86	21 45.9 -24.9 13.4982 0.35 0.94 16.94 16.94	5 276.1 -271.1 266.158 0.08 0.22 0.67 0.67	0 5.2 -5.2 5.18088 0.00 0.00 0.00 0.00	2230 37.01 37.01
CLUSTER 2	2186 1908.0 278.0 40.5102 36.28 97.76 42.51 85.38	44 46.0 -2.0 .087919 0.73 1.97 35.48 52.42	6 276.8 -270.8 264.94 0.10 0.27 0.80 1.47	0 5.2 -5.2 5.19482 0.00 0.00 0.00 0.00	2236 37.11 74.11
CLUSTER 3	576 503.4 72.6 10.4554 9.56 97.63 11.20 96.58	1 12.1 -11.1 10.2231 0.02 0.17 0.81 53.23	13 73.0 -60.0 49.354 0.22 2.20 1.74 3.22	0 1.4 -1.4 1.37073 0.00 0.00 0.00 0.00	590 9.79 83.90
CLUSTER 4	128 525.6 -397.6 300.834 2.12 20.78 2.49 99.07	23 12.7 10.3 8.40901 0.38 3.73 18.55 71.77	453 76.3 376.7 1861.21 7.52 73.54 60.72 63.94	12 1.4 10.6 78.0508 0.20 1.95 85.71 85.71	616 10.22 94.13
CLUSTER 5	48 302.1 -254.1 213.696 0.80 13.56 0.93 100.00	35 7.3 27.7 105.451 0.58 9.89 28.23 100.00	269 43.8 225.2 1155.99 4.46 75.99 36.06 100.00	2 0.8 1.2 1.68804 0.03 0.56 14.29 100.00	354 5.87 100.00
TOTAL	5142 85.33	124 2.06	746 12.38	14 0.23	6026 100.00

FREQUENCY MISSING = 237

STATISTICS FOR TABLE OF CLUSTER BY MAXSNELH

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	12	4440.931	0.000
LIKELIHOOD RATIO CHI-SQUARE	12	3469.817	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	3132.898	0.000
PHI		0.858	
CONTINGENCY COEFFICIENT		0.651	
CRAMER'S V		0.496	

EFFECTIVE SAMPLE SIZE = 6026
FREQUENCY MISSING = 237

AUTO-MOPED ACCIDENTS 1982

TABLE OF CLUSTER BY WEGSITUA

CLUSTER(CLUSTERNR)   WEGSITUA(WEGSITUATIE)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	rechte weg	kruising	T/Y kruising	verkeersplein	bocht	TOTAL
CLUSTER 1	299	1167	660	3	103	2232
	574.0	905.9	672.8	3.0	76.4	
	-275.0	261.1	-12.8	0.0	26.6	
	131.717	75.2338	.243864	8164E-7	9.29811	
	4.94	19.29	10.91	0.05	1.70	36.89
	13.40	52.28	29.57	0.13	4.61	
	19.22	47.52	36.18	37.50	49.76	
	19.22	47.52	36.18	37.50	49.76	36.89
CLUSTER 2	405	974	849	5	8	2241
	576.3	909.6	675.5	3.0	76.7	
	-171.3	64.4	173.5	2.0	-68.7	
	50.9011	4.56181	44.5501	1.40073	61.4977	
	6.69	16.10	14.03	0.08	0.13	37.04
	18.07	43.46	37.88	0.22	0.36	
	26.03	39.66	46.55	62.50	3.86	
	45.24	87.17	82.73	100.00	53.62	73.92
CLUSTER 3	534	17	28	0	16	595
	153.0	241.5	179.4	0.8	20.4	
	381.0	-224.5	-151.4	-0.8	-4.4	
	948.733	208.697	127.727	.786647	.931566	
	8.82	0.28	0.46	0.00	0.26	9.83
	89.75	2.86	4.71	0.00	2.69	
	34.32	0.69	1.54	0.00	7.73	
	79.56	87.87	84.27	100.00	61.35	83.75
CLUSTER 4	212	195	138	0	78	623
	160.2	252.9	187.8	0.8	21.3	
	51.8	-57.9	-49.8	-0.8	56.7	
	16.7471	13.2418	13.2038	.823666	150.781	
	3.50	3.22	2.28	0.00	1.29	10.30
	34.03	31.30	22.15	0.00	12.52	
	13.62	7.94	7.57	0.00	37.68	
	93.19	95.81	91.83	100.00	99.03	94.05
CLUSTER 5	106	103	149	0	2	360
	92.6	146.1	108.5	0.5	12.3	
	13.4	-43.1	40.5	-0.5	-10.3	
	1.94744	12.7237	15.1019	.475954	8.64012	
	1.75	1.70	2.46	0.00	0.03	5.95
	29.44	28.61	41.39	0.00	0.56	
	6.81	4.19	8.17	0.00	0.97	
	100.00	100.00	100.00	100.00	100.00	100.00
TOTAL	1556	2456	1824	8	207	6051
	25.71	40.59	30.14	0,13	3.42	100.00

FREQUENCY MISSING = 212

STATISTICS FOR TABLE OF CLUSTER BY WEGSITUA

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	16	1899.967	0.000
LIKELIHOOD RATIO CHI-SQUARE	16	1753.959	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	82.307	0.000
PHI		0.560	
CONTINGENCY COEFFICIENT		0.489	
CRAMER'S V		0.280	

EFFECTIVE SAMPLE SIZE = 6051
FREQUENCY MISSING = 212

AUTO-MOPED ACCIDENTS 1982

TABLE OF CLUSTER BY LICHTGES

CLUSTER(CLUSTERNR)   LICHTGES(LICHTGESTELDH)

CLUSTER	FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	daglicht	duisternis	schemer	TOTAL
CLUSTER 1	FREQUENCY	1762	417	53	2232
	EXPECTED	1780.2	381.3	70.4	
	DEVIATION	-18.2	35.7	-17.4	
	CELL CHI2	.185394	3.33406	4.31856	
	PERCENT	29.11	6.89	0.88	36.88
	ROW PCT	78.94	18.68	2.37	
	COL PCT	36.50	40.33	27.75	
	CUM COL%	36.50	40.33	27.75	36.88
CLUSTER 2	FREQUENCY	1755	399	87	2241
	EXPECTED	1787.4	382.9	70.7	
	DEVIATION	-32.4	16.1	16.3	
	CELL CHI2	.587087	.678623	3.74487	
	PERCENT	29.00	6.59	1.44	37.03
	ROW PCT	78.31	17.80	3.88	
	COL PCT	36.36	38.59	45.55	
	CUM COL%	72.86	78.92	73.30	73.91
CLUSTER 3	FREQUENCY	520	59	17	596
	EXPECTED	475.4	101.8	18.8	
	DEVIATION	44.6	-42.8	-1.8	
	CELL CHI2	4.19161	18.0132	.174104	
	PERCENT	8.59	0.97	0.28	9.85
	ROW PCT	87.25	9.90	2.85	
	COL PCT	10.77	5.71	8.90	
	CUM COL%	83.63	84.62	82.20	83.76
CLUSTER 4	FREQUENCY	498	106	19	623
	EXPECTED	496.9	106.4	19.7	
	DEVIATION	1.1	-0.4	-0.7	
	CELL CHI2	.002448	.001829	.022273	
	PERCENT	8.23	1.75	0.31	10.29
	ROW PCT	79.94	17.01	3.05	
	COL PCT	10.32	10.25	9.95	
	CUM COL%	93.95	94.87	92.15	94.05
CLUSTER 5	FREQUENCY	292	53	15	360
	EXPECTED	287.1	61.5	11.4	
	DEVIATION	4.9	-8.5	3.6	
	CELL CHI2	.082548	1.17658	1.1652	
	PERCENT	4.82	0.88	0.25	5.95
	ROW PCT	81.11	14.72	4.17	
	COL PCT	6.05	5.13	7.85	
	CUM COL%	100.00	100.00	100.00	100.00
TOTAL		4827	1034	191	6052
		79.76	17.09	3.16	100.00

FREQUENCY MISSING = 211

STATISTICS FOR TABLE OF CLUSTER BY LICHTGES

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	8	37.679	0.000
LIKELIHOOD RATIO CHI-SQUARE	8	40.801	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	1.442	0.230
PHI		0.079	
CONTINGENCY COEFFICIENT		0.079	
CRAMER'S V		0.056	

EFFECTIVE SAMPLE SIZE = 6052
FREQUENCY MISSING = 211

<u>AUTO-MOPED ACCIDENTS 1982</u>

TABLE OF CLUSTER BY GEMEKLAS

CLUSTER(CLUSTERNR) GEMEKLAS(GEM.KLASSE)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	> 100.000	50 - 100	20 - 50	10 - 20	5 - 10	< 5.000	TOTAL
CLUSTER 1	773 695.5 33.4 9.76432 12.95 34.83 41.03 41.27	442 412.6 29.4 2.13146 7.30 19.39 39.53 39.53	533 548.4 -15.4 .429913 8.80 23.86 35.87 35.87	309 333.6 -24.6 1.81219 5.10 13.83 34.18 34.18	129 180.1 -51.1 14.4879 2.13 5.77 26.43 26.43	43 63.8 -20.8 6.30255 0.71 1.92 24.86 24.86	2234 36.90 36.90
CLUSTER 2	973 597.8 175.2 44.0565 14.42 38.75 46.31 87.59	470 413.8 56.2 7.51575 7.76 20.97 42.04 81.57	544 550.1 -6.1 0.06699 8.99 24.27 36.61 72.48	231 334.6 -103.6 32.0939 3.82 10.31 25.55 59.73	101 180.6 -79.6 35.113 1.67 4.51 20.70 47.13	22 64.0 -42.0 27.597 0.36 0.98 12.72 37.57	2241 37.02 73.92
CLUSTER 3	190 185.5 4.1 .105601 3.14 31.88 10.08 97.67	122 110.1 11.9 1.29439 2.02 20.47 10.91 92.49	154 146.3 7.7 .406052 2.54 25.84 10.36 82.84	72 89.0 -17.0 3.24594 1.19 12.08 7.96 67.70	45 48.0 -3.0 .192653 0.74 7.55 9.22 56.35	13 17.0 -4.0 .954242 0.21 2.18 7.51 45.09	596 9.84 83.76
CLUSTER 4	22 194.2 -172.2 152.175 0.36 3.53 [illegible] 98.83	40 115.1 -75.1 48.9572 0.66 6.42 3.58 95.06	166 152.9 13.1 1.11879 2.74 26.65 11.17 94.01	176 93.0 83.0 74.0028 2.91 28.25 19.47 87.17	148 50.2 97.8 190.381 2.44 23.76 30.33 86.68	71 17.8 53.2 158.958 1.17 11.40 41.04 86.13	623 10.29 94.05
CLUSTER 5	22 113.1 -91.1 73.4[illegible] 0.36 6.11 1.17 100.00	44 66.5 -22.5 7.60248 0.73 12.22 3.94 100.00	89 88.4 0.6 .004567 1.47 24.72 5.99 100.00	116 53.8 62.2 72.0715 1.92 32.22 12.83 100.00	65 29.0 36.0 44.6139 1.07 18.06 13.32 100.00	24 10.3 13.7 18.2783 0.40 6.67 13.87 100.00	360 5.95 100.00
TOTAL	1885 31.14	1118 18.47	1486 24.55	904 14.93	488 8.06	173 2.86	6054 100.00

FREQUENCY MISSING = 209

STATISTICS FOR TABLE OF CLUSTER BY GEMEKLAS

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	20	1028.976	0.000
LIKELIHOOD RATIO CHI-SQUARE	20	1012.022	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	597.308	0.000
PHI		0.412	
CONTINGENCY COEFFICIENT		0.381	
CRAMER'S V		0.206	

EFFECTIVE SAMPLE SIZE = 6054
FREQUENCY MISSING = 209

AUTO-MOPED ACCIDENTS 1982

TABLE OF CLUSTER BY VORGMANE

CLUSTER(CLUSTERNR) VORGMANE VORGMAN1)

FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%

CLUSTER		rudenV/in pen	varuit s tilstand	afremmen	stilstaan	afslaan rechts	afslaan links	keren li nksom	overig	99	TOTAL
CLUSTER 1	FREQUENCY	1795	18	45	36	49	269	3	18	0	2234
	EXPECTED	1054.6	5[illegible]	52.4	33.9	422.9	512.9	36.2	68.6	0.7	
	DEVIATION	741.4	-33.7	-7.4	2.1	-373.9	-243.9	-33.2	-50.6	-0.7	
	CELL CHI2	521.145	21.9[illegible]	1.04497	.123894	330.565	116.002	30.4121	37.35[illegible]	.738[illegible]	
	PERCENT	29.65	0.30	0.74	0.59	0.81	4.44	0.05	0.30	0.00	36.90
	ROW PCT	80.39	0.81	2.01	1.61	2.19	12.04	0.13	0.81	0.00	
	COL PCT	[illegible]	[illegible]	[illegible]	39.13	4.28	19.35	3.06	9.58	0.00	
	CUM COL%	62.84	12.86	[illegible]	39.13	4.28	19.35	3.06	9.58	0.00	36.90
CLUSTER 2	FREQUENCY	361	57	14	10	849	857	71	11	1	2241
	EXPECTED	1057.9	51.8	[illegible]	34.1	424.2	514.5	36.3	68.9	0.7	
	DEVIATION	[illegible]	[illegible]	[illegible]	-24.1	424.8	342.5	34.7	-57.9	0.3	
	CELL CHI2	459.135	4.44[illegible]	23.[illegible]	16.9919	425.362	227.94	33.237	48.6087	.091[illegible]	
	PERCENT	5.96	1.11	0.23	0.17	14.02	14.16	1.17	0.18	0.02	37.02
	ROW PCT	16.1[illegible]	[illegible]	0.62	0.45	37.88	38.24	3.17	0.49	0.04	
	COL PCT	12.63	41.85	9.[illegible]	10.87	74.08	61.65	72.45	5.91	50.00	
	CUM COL%	75.47	61.[illegible]	41.[illegible]	50.00	78.36	81.01	75.51	15.59	50.00	73.92
CLUSTER 3	FREQUENCY	75	38	74	38	60	136	21	153	1	596
	EXPECTED	281.4	13.8	14.3	9.1	112.8	136.8	9.6	18.3	0.2	
	DEVIATION	-206.4	[illegible]	[illegible]	28.9	-52.8	-0.8	11.4	134.7	0.8	
	CELL CHI2	151.354	42.[illegible]	25[illegible]	92.4892	24.7297	.005178	13.3576	990.709	3.27[illegible]	
	PERCENT	1.24	[illegible]	[illegible]	0.63	0.99	2.25	0.35	2.53	0.02	9.84
	ROW PCT	12.58	[illegible]	12.42	6.38	10.07	22.82	3.52	25.67	0.17	
	COL PCT	2.62	[illegible]	[illegible]	41.30	5.24	9.78	21.43	82.26	50.00	
	CUM COL%	78.10	[illegible]	[illegible]	91.30	83.60	90.79	96.94	97.85	100.00	83.76
CLUSTER 4	FREQUENCY	568	0	9	7	10	25	0	4	0	623
	EXPECTED	294.1	14.4	14.6	9.5	117.9	143.0	10.1	19.1	0.2	
	DEVIATION	273.9	-14.4	[illegible]	-2.5	-107.9	-118.0	-10.1	-15.1	-0.2	
	CELL CHI2	255.064	14.407	2.1[illegible]	.643082	98.7796	97.4103	10.0849	11.9766	.205814	
	PERCENT	9.38	0.00	0.15	0.12	0.17	0.41	0.00	0.07	0.00	10.29
	ROW PCT	91.17	0.00	1.44	1.12	1.61	4.01	0.00	0.64	0.00	
	COL PCT	19.87	[illegible]	[illegible]	7.61	0.87	1.80	0.00	2.15	0.00	
	CUM COL%	97.97	87.35	1[illegible]	98.91	84.47	92.59	96.94	100.00	100.00	94.05
CLUSTER 5	FREQUENCY	59	17	0	1	178	103	3	0	0	360
	EXPECTED	170.0	8.3	8.4	5.5	68.1	82.7	5.8	11.1	0.1	
	DEVIATION	-112.0	[illegible]	-8.4	-4.5	109.9	20.3	-2.8	-11.1	-0.1	
	CELL CHI2	73.744[illegible]	9.07[illegible]	[illegible]	3.6[illegible]	177.085	5.00719	1.37194	11.06[illegible]	0.11893	
	PERCENT	0.96	0.28	[illegible]	0.02	2.94	1.70	0.05	0.00	0.00	5.95
	ROW PCT	16.11	4.[illegible]	[illegible]	0.28	49.44	28.61	0.83	0.00	0.00	
	COL PCT	2.03	12.14	[illegible]	1.09	15.53	7.41	3.06	0.00	0.00	
	CUM COL%	100.00	[illegible]	[illegible]	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TOTAL		2859	140	145	92	1146	1390	98	186	2	6054
		47.2[illegible]	2.31	[illegible]	1.52	18.93	22.96	1.62	3.07	0.03	100.00

FREQUENCY MISSING = 209

STATISTICS FOR TABLE OF CLUSTER BY VORGMANE

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	32	4659.834	0.000
LIKELIHOOD RATIO CHI-SQUARE	32	4255.394	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	142.886	0.000
PHI		0.877	
CONTINGENCY COEFFICIENT		0.659	
CRAMER'S V		0.439	

EFFECTIVE SAMPLE SIZE = 6054
FREQUENCY MISSING = 209

AUTO-MOPED ACCIDENTS 1982

TABLE OF CLUSTER BY VORDMANT

CLUSTER(CLUSTERNR) VORDMANT(VORDMAN2)

FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	rydenrlc per	afremmen st.ista	afslaan rechts	afslaan links	rijbaan wisselen	oversteken	overig	TOTAL
CLUSTER 1	1459	97	153	432	23	44	27	2234
	1782.3	56.8	74.6	237.0	29.2	27.3	25.8	
	-323.3	40.2	78.4	195.0	-6.2	16.7	1.2	
	59.[illegible]	28.[illegible]	82.[illegible]	160.48	1.30189	10.1903	0.052[illegible]	
	24.[illegible]	1.60	2.53	7.14	0.38	0.73	0.45	36.91
	65.26	4.34	6.85	19.34	1.03	1.97	1.21	
	[illegible]	[illegible]	75.74	67.29	29.11	59.46	38.57	
	30.13	[illegible]	75.74	67.29	29.11	59.46	38.57	36.91
CLUSTER 2	[illegible]	24	5	22	23	2	3	2240
	17[illegible].1	57.0	74.8	237.6	29.2	27.4	25.9	
	[illegible]	-33.0	-69.8	-215.5	-6.2	-25.4	-22.9	
	[illegible]	19.[illegible]	65.[illegible]97	195.65[illegible]	1.33163	23.5353	20.25[illegible]	
	35.71	0.40	0.08	0.36	0.38	0.03	0.05	37.01
	[illegible]	[illegible]	0.22	0.98	1.03	0.09	0.13	
	44.[illegible]3	15.58	2.48	3.43	29.11	2.70	4.29	
	74.[illegible]1	[illegible]	78.22	70.72	58.23	62.16	42.86	73.93
CLUSTER 3	532	15	5	8	27	0	7	595
	47[illegible]	15.[illegible]	19.9	63.1	7.8	7.3	6.9	
	5[illegible]	-0.[illegible]	-14.9	-55.1	19.2	-7.3	0.1	
	6.39[illegible]	.0[illegible]6[illegible]98	11.1184	48.132	47.6273	7.27528	.00202[illegible]	
	8.7[illegible]	0.25	0.08	0.13	0.45	0.00	0.12	9.83
	89.41	2.5[illegible]	0.84	1.34	4.54	0.00	1.18	
	[illegible]	[illegible]	2.48	1.25	34.18	0.00	10.00	
	6[illegible].[illegible]2	88.96	80.69	71.96	92.41	62.16	52.86	83.76
CLUSTER 4	320	17	39	180	6	28	33	623
	497.3	15.9	20.8	66.1	8.1	7.6	7.2	
	-177.3	1.1	18.2	113.9	-2.1	20.4	25.8	
	63.2171	.082957	15.9398	196.342	.559116	54.5366	92.33[illegible]4	
	5.29	0.28	0.64	2.97	0.10	0.46	0.55	10.29
	51.36	2.73	6.26	28.89	0.96	4.49	5.30	
	[illegible]	11.04	19.31	28.04	7.59	37.84	47.14	
	92.55	100.00	100.00	100.00	100.00	100.00	100.00	94.05
CLUSTER 5	360	0	0	0	0	0	0	360
	287.4	9.2	12.0	38.2	4.7	4.4	4.2	
	72.6	-9.2	-12.0	-38.2	-4.7	-4.4	-4.2	
	18.3[illegible]	9.15961	12.0159	38.189	4.69927	4.40185	4.16391	
	5.95	0.00	0.00	0.00	0.00	0.00	0.00	5.95
	100.00	0.00	0.00	0.00	0.00	0.00	0.00	
	7.45	0.00	0.00	0.00	0.00	0.00	0.00	
	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TOTAL	4631	154	202	642	79	74	70	6052
	79.52	2.54	3.34	10.61	1.31	1.22	1.16	100.00

FREQUENCY MISSING = 211

STATISTICS FOR TABLE OF CLUSTER BY VORDMANT

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	24	1389.041	0.000
LIKELIHOOD RATIO CHI-SQUARE	24	1504.530	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	10.403	0.001
PHI		0.478	
CONTINGENCY COEFFICIENT		0.431	
CRAMER'S V		0.239	

EFFECTIVE SAMPLE SIZE = 6052
FREQUENCY MISSING = 211

AUTO-PEDESTRIAN ACCIDENTS 1982

TABLE OF CLUSTER BY MAXSNELH

CLUSTER(CLUSTERNR)   MAXSNELH(MAX.SNELHEID)

CLUSTER	FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	=< 50 km	70 km	80-90 km	100 km	TOTAL
CLUSTER 1	FREQUENCY	1899	19	6	0	1924
	EXPECTED	1800.4	20.8	99.2	3.6	
	DEVIATION	98.6	-1.8	-93.2	-3.6	
	CELL CHI2	5.40202	.162919	87.5456	3.59357	
	PERCENT	70.94	0.71	0.22	0.00	71.87
	ROW PCT	98.70	0.99	0.31	0.00	
	COL PCT	75.81	65.52	4.35	0.00	
	CUM COL%	75.81	65.52	4.35	0.00	71.87
CLUSTER 2	FREQUENCY	448	2	0	0	450
	EXPECTED	421.1	4.9	23.2	0.8	
	DEVIATION	26.9	-2.9	-23.2	-0.8	
	CELL CHI2	1.72009	1.6954	23.1976	.840493	
	PERCENT	16.74	0.07	0.00	0.00	16.81
	ROW PCT	99.56	0.44	0.00	0.00	
	COL PCT	17.88	6.90	0.00	0.00	
	CUM COL%	93.69	72.41	4.35	0.00	88.68
CLUSTER 3	FREQUENCY	8	8	131	5	152
	EXPECTED	142.2	1.6	7.8	0.3	
	DEVIATION	-134.2	6.4	123.2	4.7	
	CELL CHI2	126.694	24.5141	1935.96	78.3431	
	PERCENT	0.30	0.30	4.89	0.19	5.68
	ROW PCT	5.26	5.26	86.18	3.29	
	COL PCT	0.32	27.59	94.93	100.00	
	CUM COL%	94.01	100.00	99.28	100.00	94.36
CLUSTER 4	FREQUENCY	150	0	1	0	151
	EXPECTED	141.3	1.6	7.8	0.3	
	DEVIATION	8.7	-1.6	-6.8	-0.3	
	CELL CHI2	.535911	1.63579	5.91255	.282032	
	PERCENT	5.60	0.00	0.04	0.00	5.64
	ROW PCT	99.34	0.00	0.66	0.00	
	COL PCT	5.99	0.00	0.72	0.00	
	CUM COL%	100.00	100.00	100.00	100.00	100.00
TOTAL		2505	29	138	5	2677
		93.57	1.08	5.16	0.19	100.00

FREQUENCY MISSING = 223

STATISTICS FOR TABLE OF CLUSTER BY MAXSNELH

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	9	2298.023	0.000
LIKELIHOOD RATIO CHI-SQUARE	9	976.991	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	457.323	0.000
PHI		0.927	
CONTINGENCY COEFFICIENT		0.680	
CRAMER'S V		0.535	

EFFECTIVE SAMPLE SIZE = 2677
FREQUENCY MISSING = 223
WARNING: 43% OF THE CELLS HAVE EXPECTED COUNTS LESS THAN 5. CHI-SQUARE MAY NOT BE A VALID TEST.

AUTO-PEDESTRIAN ACCIDENTS 1982

TABLE OF CLUSTER BY WEGSITUA

CLUSTER(CLUSTERNR) WEGSITUA(WEGSITUATIE)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	rechte weg	kruising	T/Y kruising	verkeersplein	bocht	TOTAL
CLUSTER 1	1620 1443.1 176.9 21.6933 60.29 83.85 80.72 80.72	145 300.5 -155.5 80.5046 5.40 7.51 34.69 34.69	140 158.2 -18.2 2.0903 5.21 7.25 63.64 63.64	2 4.3 -2.3 1.2413 0.07 0.10 33.33 33.33	25 25.9 -0.9 .030233 0.93 1.29 69.44 69.44	1932 71.90 71.90
CLUSTER 2	135 336.1 -201.1 120.34 5.02 30.00 6.73 87.44	249 70.0 179.0 457.685 9.27 55.33 59.57 94.26	61 36.8 24.2 15.8373 2.27 13.56 27.73 91.36	4 1.0 3.0 8.9278 0.15 0.89 66.67 100.00	1 6.0 -5.0 4.19489 0.04 0.22 2.78 72.22	450 16.75 88.65
CLUSTER 3	124 114.3 9.7 .826685 4.61 81.05 6.18 93.62	12 23.8 -11.8 5.85136 0.45 7.84 2.87 97.13	10 12.5 -2.5 .509751 0.37 6.54 4.55 95.91	0 0.3 -0.3 .341645 0.00 0.00 0.00 100.00	7 2.0 5.0 11.9538 0.26 4.58 19.44 91.67	153 5.69 94.34
CLUSTER 4	128 113.5 14.5 1.84328 4.76 84.21 6.38 100.00	12 23.6 -11.6 5.7356 0.45 7.89 2.87 100.00	9 12.4 -3.4 .953689 0.33 5.92 4.09 100.00	0 0.3 -0.3 .339412 0.00 0.00 0.00 100.00	3 2.0 1.0 0.45588 0.11 1.97 8.33 100.00	152 5.66 100.00
TOTAL	2007 74.69	418 15.56	220 8.19	6 0.22	36 1.34	2687 100.00

FREQUENCY MISSING = 213

STATISTICS FOR TABLE OF CLUSTER BY WEGSITUA

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	12	741.356	0.000
LIKELIHOOD RATIO CHI-SQUARE	12	606.343	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	27.677	0.000
PHI		0.525	
CONTINGENCY COEFFICIENT		0.465	
CRAMER'S V		0.303	

EFFECTIVE SAMPLE SIZE = 2687
FREQUENCY MISSING = 213
WARNING: 30% OF THE CELLS HAVE EXPECTED COUNTS LESS THAN 5. CHI-SQUARE MAY NOT BE A VALID TEST.

AUTO-PEDESTRIAN ACCIDENTS 1982

TABLE OF CLUSTER BY LICHTGES

CLUSTER(CLUSTERNR)   LICHTGES(LICHTGESTELDH)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	daglicht	duisternis	schemer	TOTAL
CLUSTER 1	1552	318	63	1933
	1468.4	394.1	70.5	
	83.6	-76.1	-7.5	
	4.75405	14.6874	.792634	
	57.74	11.83	2.34	71.91
	80.29	16.45	3.26	
	76.00	58.03	64.29	
	76.00	58.03	64.29	71.91
CLUSTER 2	291	138	21	450
	341.9	91.7	16.4	
	-50.9	46.3	4.6	
	7.56465	23.3253	1.28625	
	10.83	5.13	0.78	16.74
	64.67	30.67	4.67	
	14.25	25.18	21.43	
	90.25	83.21	85.71	88.65
CLUSTER 3	118	27	8	153
	116.2	31.2	5.6	
	1.8	-4.2	2.4	
	.026957	.563368	1.05151	
	4.39	1.00	0.30	5.69
	77.12	17.65	5.23	
	5.78	4.93	8.16	
	96.03	88.14	93.88	94.35
CLUSTER 4	81	65	6	152
	115.5	31.0	5.5	
	-34.5	34.0	0.5	
	10.2901	37.3308	.037907	
	3.01	2.42	0.22	5.65
	53.29	42.76	3.95	
	3.97	11.86	6.12	
	100.00	100.00	100.00	100.00
TOTAL	2042	548	98	2688
	75.97	20.39	3.65	100.00

FREQUENCY MISSING = 212

STATISTICS FOR TABLE OF CLUSTER BY LICHTGES

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	6	101.711	0.000
LIKELIHOOD RATIO CHI-SQUARE	6	91.903	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	50.710	0.000
PHI		0.195	
CONTINGENCY COEFFICIENT		0.191	
CRAMER'S V		0.138	

EFFECTIVE SAMPLE SIZE = 2688
FREQUENCY MISSING = 212

AUTO-PEDESTRIAN ACCIDENTS 1982

TABLE OF CLUSTER BY GEMEKLAS

CLUSTER(CLUSTERNR) GEMEKLAS(GEM.KLASSE)

FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%

CLUSTER		> 100.000	50 - 100	20 - 50	10 - 20	5 - 10	< 5.000	TOTAL
CLUSTER 1	FREQUENCY	974	293	367	177	87	35	1933
	EXPECTED	971.9	281.1	358.0	190.5	95.6	35.9	
	DEVIATION	2.1	11.9	9.0	-13.5	-8.6	-0.9	
	CELL CHI2	.004575	.506182	.226787	0.95621	.774957	.024727	
	PERCENT	36.22	10.90	13.65	6.58	3.24	1.30	71.89
	ROW PCT	50.39	15.16	18.99	9.16	4.50	1.81	
	COL PCT	72.04	74.94	73.69	66.79	65.41	70.00	
	CUM COL%	72.04	74.94	73.69	66.79	65.41	70.00	71.89
CLUSTER 2	FREQUENCY	297	63	70	17	3	0	450
	EXPECTED	226.3	65.4	83.3	44.3	22.3	8.4	
	DEVIATION	70.7	-2.4	-13.3	-27.3	-19.3	-8.4	
	CELL CHI2	22.1203	.090484	2.13516	16.8641	16.6617	8.36742	
	PERCENT	11.04	2.34	2.60	0.63	0.11	0.00	16.73
	ROW PCT	66.00	14.00	15.56	3.78	0.67	0.00	
	COL PCT	21.97	16.11	14.06	6.42	2.26	.0.00	
	CUM COL%	94.01	91.05	87.75	73.21	67.67	70.00	88.62
CLUSTER 3	FREQUENCY	12	11	31	51	38	11	154
	EXPECTED	77.4	22.4	28.5	15.2	7.6	2.9	
	DEVIATION	-65.4	-11.4	2.5	35.8	30.4	8.1	
	CELL CHI2	55.2893	5.79626	.215536	84.5584	121.194	23.1192	
	PERCENT	0.45	0.41	1.15	1.90	1.41	0.41	5.73
	ROW PCT	7.79	7.14	20.13	33.12	24.68	7.14	
	COL PCT	0.89	2.81	6.22	19.25	28.57	22.00	
	CUM COL%	94.90	93.86	93.98	92.45	96.24	92.00	94.35
CLUSTER 4	FREQUENCY	69	24	30	20	5	4	152
	EXPECTED	76.4	22.1	28.2	15.0	7.5	2.8	
	DEVIATION	-7.4	1.9	1.8	5.0	-2.5	1.2	
	CELL CHI2	.721175	.163008	.121548	1.68262	.843373	.487382	
	PERCENT	2.57	0.89	1.12	0.74	0.19	0.15	5.65
	ROW PCT	45.39	15.79	19.74	13.16	3.29	2.63	
	COL PCT	5.10	6.14	6.02	7.55	3.76	8.00	
	CUM COL%	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TOTAL		1352	391	498	265	133	50	2689
		50.28	14.54	18.52	9.85	4.95	1.86	100.00

FREQUENCY MISSING = 211

STATISTICS FOR TABLE OF CLUSTER BY GEMEKLAS

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	15	362.924	0.000
LIKELIHOOD RATIO CHI-SQUARE	15	314.425	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	34.942	0.000
PHI		0.367	
CONTINGENCY COEFFICIENT		0.345	
CRAMER'S V		0.212	

EFFECTIVE SAMPLE SIZE = 2689
FREQUENCY MISSING = 211

<u>AUTO-PEDESTRIAN ACCIDENTS 1982</u>

TABLE OF CLUSTER BY VOR3MANE

CLUSTER(CLUSTERNR) VOR3MANE(VOR3MAN1)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	ryder/lo pen	vanuit s tilstand	afremmen / tilsta	afslaan rechts	afslaan links	overig	TOTAL
CLUSTER 1	1795	10	101	6	11	9	1932
	1691.0	32.4	91.3	28.1	71.9	17.3	
	104.0	-22.4	9.7	-22.1	-60.9	-8.3	
	6.39121	15.4573	1.01958	17.3354	51.6107	3.955	
	66.83	0.37	3.76	0.22	0.41	0.34	71.93
	92.91	0.52	5.23	0.31	0.57	0.47	
	76.35	22.22	79.53	15.38	11.00	37.50	
	76.35	22.22	79.53	15.38	11.00	37.50	71.93
CLUSTER 2	310	18	13	29	76	3	449
	393.0	7.5	21.2	6.5	16.7	4.0	
	-83.0	10.5	-8.2	22.5	59.3	-1.0	
	17.5294	14.5941	3.17025	77.5198	210.247	.255232	
	11.54	0.67	0.48	1.08	2.83	0.11	16.72
	69.04	4.01	2.90	6.46	16.93	0.67	
	13.19	40.00	10.24	74.36	76.00	12.50	
	89.54	62.22	89.76	89.74	87.00	50.00	88.64
CLUSTER 3	144	0	8	0	0	2	154
	134.8	2.6	7.3	2.2	5.7	1.4	
	9.2	-2.6	0.7	-2.2	-5.7	0.6	
	.628582	2.58004	.070906	2.23604	5.73343	0.28295	
	5.36	0.00	0.30	0.00	0.00	0.07	5.73
	93.51	0.00	5.19	0.00	0.00	1.30	
	6.13	0.00	6.30	0.00	0.00	8.33	
	95.66	62.22	96.06	89.74	87.00	58.33	94.38
CLUSTER 4	102	17	5	4	13	10	151
	132.2	2.5	7.1	2.2	5.6	1.3	
	-30.2	14.5	-2.1	1.8	7.4	8.7	
	6.89566	82.7688	.641203	1.49015	9.6836	55.4662	
	3.80	0.63	0.19	0.15	0.48	0.37	5.62
	67.55	11.26	3.31	2.65	8.61	6.62	
	4.34	37.78	3.94	10.26	13.00	41.67	
	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TOTAL	2351	45	127	39	100	24	2686
	87.53	1.68	4.73	1.45	3.72	0.89	100.00

FREQUENCY MISSING = 214

STATISTICS FOR TABLE OF CLUSTER BY VOR3MANE

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	15	587.573	0.000
LIKELIHOOD RATIO CHI-SQUARE	15	423.064	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	130.337	0.000
PHI		0.468	
CONTINGENCY COEFFICIENT		0.424	
CRAMER'S V		0.270	

EFFECTIVE SAMPLE SIZE = 2686
FREQUENCY MISSING = 214
WARNING: 29% OF THE CELLS HAVE EXPECTED COUNTS LESS THAN 5. CHI-SQUARE MAY NOT BE A VALID TEST.

AUTO-PEDESTRIAN ACCIDENTS 1982

TABLE OF CLUSTER BY VOR3MANT

CLUSTER(CLUSTERNR) VOR3MANT(VOR3MAN2)

FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	ryden/lopen	stilstaan	oversteken	overst. r/l	overst. l/r	~~afstaan~~ n.v.t.	TOTAL
CLUSTER 1	35	3	212	853	821	9	1933
	97.8	21.6	210.6	809.4	763.4	30.2	
	-62.8	-18.6	1.4	43.6	57.6	-21.2	
	40.2944	15.983	.008984	2.34526	4.34234	14.8747	
	1.30	0.11	7.88	31.72	30.53	0.33	71.89
	1.81	0.16	10.97	44.13	42.47	0.47	
	25.74	10.00	72.35	75.75	77.31	21.43	
	25.74	10.00	72.35	75.75	77.31	21.43	71.89
CLUSTER 2	12	1	39	214	183	1	450
	22.8	5.0	49.0	188.4	177.7	7.0	
	-10.8	-4.0	-10.0	25.6	5.3	-6.0	
	5.08545	3.21964	2.05296	3.46859	.156622	5.17091	
	0.45	0.04	1.45	7.96	6.81	0.04	16.73
	2.67	0.22	8.67	47.56	40.67	0.22	
	8.82	3.33	13.31	19.01	17.23	2.38	
	34.56	13.33	85.67	94.76	94.54	23.81	88.62
CLUSTER 3	3	0	38	56	55	2	154
	7.8	1.7	16.8	64.5	60.8	2.4	
	-4.8	-1.7	21.2	-8.5	-5.8	-0.4	
	2.94428	1.71811	26.8339	1.11682	.557133	.068311	
	0.11	0.00	1.41	2.08	2.05	0.07	5.73
	1.95	0.00	24.68	36.36	35.71	1.30	
	2.21	0.00	12.97	4.97	5.18	4.76	
	36.76	13.33	98.63	99.73	99.72	28.57	94.35
CLUSTER 4	86	26	4	3	3	30	152
	7.7	1.7	16.6	63.6	60.0	2.4	
	78.3	24.3	-12.6	-60.6	-57.0	27.6	
	797.754	348.328	9.52834	57.7903	54.1812	321.462	
	3.20	0.97	0.15	0.11	0.11	1.12	5.65
	56.58	17.11	2.63	1.97	1.97	19.74	
	63.24	86.67	1.37	0.27	0.28	71.43	
	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TOTAL	136	30	293	1126	1062	42	2689
	5.06	1.12	10.90	41.87	39.49	1.56	100.00

FREQUENCY MISSING = 211

STATISTICS FOR TABLE OF CLUSTER BY VOR3MANT

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	15	1719.287	0.000
LIKELIHOOD RATIO CHI-SQUARE	15	821.484	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	320.216	0.000
PHI		0.800	
CONTINGENCY COEFFICIENT		0.625	
CRAMER'S V		0.462	

EFFECTIVE SAMPLE SIZE = 2689
FREQUENCY MISSING = 211

<u>AUTO-AUTO ACCIDENTS 1982</u>

TABLE OF CLUSTER BY MAXSNELH

CLUSTER(CLUSTERNR)   MAXSNELH(MAX.SNELHEID)

FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	=< 50 km	70 km	80-90 km	100 km	TOTAL
cluster 1	280	10	8	0	298
	193.8	14.2	78.5	6.5	
	81.2	-4.2	-70.5	-6.5	
	33.1826	1.24922	63.3065	6.51175	
	6.58	0.23	0.19	0.00	7.00
	93.96	3.36	2.68	0.00	
	9.86	4.93	0.71	0.00	
	9.86	4.93	0.71	0.00	7.00
cluster 2	250	15	14	0	279
	186.1	13.3	73.5	6.1	
	63.9	1.7	-59.5	-6.1	
	21.9335	.215241	48.1538	6.09657	
	5.87	0.35	0.33	0.00	6.56
	89.61	5.38	5.02	0.00	
	8.81	7.39	1.25	0.00	
	18.67	12.32	1.96	0.00	13.56
cluster 3	2213	83	55	0	2351
	1568.3	112.1	619.2	51.4	
	644.7	-29.1	-564.2	-51.4	
	255.07	7.57054	514.122	51.3729	
	52.00	1.95	1.29	0.00	55.24
	94.13	3.53	2.34	0.00	
	77.95	40.89	4.91	0.00	
	96.62	53.20	6.87	0.00	68.80
cluster 4	58	81	586	56	781
	521.0	37.3	205.7	17.1	
	-463.0	43.7	380.3	38.9	
	411.43	51.3781	703.032	88.8229	
	1.36	1.90	13.77	1.32	18.35
	7.43	10.37	75.03	7.17	
	2.04	39.90	52.27	60.22	
	98.66	93.10	59.14	60.22	87.15
cluster 5	38	14	458	37	547
	364.9	26.1	144.1	12.0	
	-326.9	-12.1	313.9	25.0	
	292.838	5.60278	684.003	52.4869	
	0.89	0.33	10.76	0.87	12.85
	6.95	2.56	83.73	6.76	
	1.34	6.90	40.86	39.78	
	100.00	100.00	100.00	100.00	100.00
TOTAL	2839	203	1121	93	4256
	66.71	4.77	26.34	2.19	100.00

FREQUENCY MISSING = 399

STATISTICS FOR TABLE OF CLUSTER BY MAXSNELH

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	12	3308.378	0.000
LIKELIHOOD RATIO CHI-SQUARE	12	3645.820	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	2050.673	0.000
PHI		0.882	
CONTINGENCY COEFFICIENT		0.661	
CRAMER'S V		0.509	

EFFECTIVE SAMPLE SIZE = 4256
FREQUENCY MISSING = 399

<u>AUTO-AUTO ACCIDENTS 1982</u>

TABLE OF CLUSTER BY WEGSITUA

CLUSTER(CLUSTERNR) WEGSITUA(WEGSITUATIE)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	rechte weg	kruising	T/Y kruising	verkeersplein	bocht	TOTAL
cluster 1	106	100	65	0	27	298
	57.5	161.3	60.5	0.9	17.8	
	48.5	-61.3	4.5	-0.9	9.2	
	40.8671	23.2964	.342063	.908537	4.72741	
	2.49	2.35	1.52	0.00	0.63	6.99
	35.57	33.56	21.81	0.00	9.06	
	12.88	4.33	7.51	0.00	10.59	
	12.88	4.33	7.51	0.00	10.59	6.99
cluster 2	132	75	63	7	3	280
	54.0	151.6	56.8	0.9	16.7	
	78.0	-76.6	6.2	6.1	-13.7	
	112.452	38.6719	.676501	44.2537	11.2823	
	3.10	1.76	1.48	0.16	0.07	6.57
	47.14	26.79	22.50	2.50	1.07	
	16.04	3.25	7.28	53.85	1.18	
	28.92	7.58	14.80	53.85	11.76	13.56
cluster 3	234	1550	499	6	65	2354
	454.3	1274.2	477.5	7.2	140.8	
	[illegible]	275.8	21.5	-1.2	-75.8	
	[illegible]	59.7144	.964826	.192972	40.7884	
	5.49	36.35	11.70	0.14	1.52	55.21
	9.94	65.85	21.20	0.25	2.76	
	28.43	67.16	57.69	46.15	25.49	
	57.35	74.74	72.49	100.00	37.25	68.76
cluster 4	76	529	174	0	3	782
	150.9	423.3	158.6	2.4	46.8	
	-74.9	105.7	15.4	-2.4	-43.8	
	37.203	26.4063	1.49772	2.38415	40.9584	
	1.78	12.41	4.08	0.00	0.07	18.34
	9.72	67.65	22.25	0.00	0.38	
	9.23	22.92	20.12	0.00	1.18	
	66.59	97.66	92.60	100.00	38.43	87.10
cluster 5	275	54	64	0	157	550
	106.2	297.7	111.6	1.7	32.9	
	168.8	-243.7	-47.6	-1.7	124.1	
	268.55	199.497	20.2848	1.67683	468.292	
	6.45	1.27	1.50	0.00	3.63	12.90
	50.00	9.82	11.64	0.00	28.55	
	33.41	2.34	7.40	0.00	61.57	
	100.00	100.00	100.00	100.00	100.00	100.00
TOTAL	823	2308	865	13	255	4264
	19.30	54.13	20.29	0.30	5.98	100.00

FREQUENCY MISSING = 391

STATISTICS FOR TABLE OF CLUSTER BY WEGSITUA

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	16	1552.742	0.000
LIKELIHOOD RATIO CHI-SQUARE	16	1372.594	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	41.040	0.000
PHI		0.603	
CONTINGENCY COEFFICIENT		0.517	
CRAMER'S V		0.302	

EFFECTIVE SAMPLE SIZE = 4264
FREQUENCY MISSING = 391

AUTO-AUTO ACCIDENTS 1982

TABLE OF CLUSTER BY LICHTGES

CLUSTER(CLUSTERNR)   LICHTGES(LICHTGESTELDH)

FREQUENCY
EXPECTED
DEVIATION
CELL CHI2
PERCENT
ROW PCT
COL PCT
CUM COL%

	daglicht	duisternis	schemer	TOTAL
cluster 1	61	226	11	298
	208.9	80.3	8.8	
	-147.9	145.7	2.2	
	104.7	264.268	.550211	
	1.43	5.30	0.26	6.98
	20.47	75.84	3.69	
	2.04	19.65	8.73	
	2.04	19.65	8.73	6.98
cluster 2	218	56	5	279
	195.6	75.2	8.2	
	22.4	-19.2	-3.2	
	2.57297	4.89916	1.27308	
	5.11	1.31	0.12	6.54
	78.14	20.07	1.79	
	7.29	4.87	3.97	
	9.33	24.52	12.70	13.52
cluster 3	1664	634	59	2357
	1652.2	635.2	69.6	
	11.8	-1.2	-10.6	
	.084782	.002403	1.61429	
	39.00	14.86	1.38	55.24
	70.60	26.90	2.50	
	55.63	55.13	46.83	
	64.96	79.65	59.52	68.76
cluster 4	700	59	24	783
	548.9	211.0	23.1	
	151.1	-152.0	0.9	
	41.6243	109.522	.033405	
	16.40	1.38	0.56	18.35
	89.40	7.54	3.07	
	23.40	5.13	19.05	
	88.37	84.78	78.57	87.11
cluster 5	348	175	27	550
	385.5	148.2	16.2	
	-37.5	26.8	10.8	
	3.65313	4.83436	7.12754	
	8.16	4.10	0.63	12.89
	63.27	31.82	4.91	
	11.63	15.22	21.43	
	100.00	100.00	100.00	100.00
TOTAL	2991	1150	126	4267
	70.10	26.95	2.95	100.00

FREQUENCY MISSING = 388

STATISTICS FOR TABLE OF CLUSTER BY LICHTGES

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	8	546.759	0.000
LIKELIHOOD RATIO CHI-SQUARE	8	539.738	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	91.334	0.000
PHI		0.358	
CONTINGENCY COEFFICIENT		0.337	
CRAMER'S V		0.253	

EFFECTIVE SAMPLE SIZE = 4267
FREQUENCY MISSING = 388

AUTO-AUTO ACCIDENTS 1982

TABLE OF CLUSTER BY GEMEKLAS

CLUSTER(CLUSTERNR)   GEMEKLAS(GEM.KLASSE)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	> 100.000	50 - 100	20 - 50	10 - 20	5 - 10	< 5.000	TOTAL
cluster 1	177 95.9 81.1 68.6671 4.15 59.40 12.89 12.89	41 42.2 -1.2 .036533 0.96 13.76 6.78 6.78	43 68.4 -25.4 9.40536 1.01 14.43 4.39 4.39	21 49.7 -28.7 16.5841 0.49 7.05 2.95 2.95	12 29.0 -17.0 9.94571 0.28 4.03 2.89 2.89	4 12.8 -8.8 6.09264 0.09 1.34 2.17 2.17	298 6.98 6.98
cluster 2	144 90.1 53.9 32.2332 3.37 51.43 10.49 23.39	51 39.7 11.3 3.22241 1.19 18.21 8.43 15.21	44 64.2 -20.2 6.36998 1.03 15.71 4.49 8.89	25 46.7 -21.7 10.0907 0.59 8.93 3.51 6.46	14 27.2 -13.2 6.4249 0.33 5.00 3.37 6.27	2 12.1 -10.1 8.40259 0.05 0.71 1.09 3.26	280 6.56 13.54
cluster 3	987 758.2 228.8 69.0179 23.13 41.88 71.89 95.27	419 334.1 84.9 21.5682 9.82 17.78 69.26 84.46	559 540.7 18.3 .622668 13.10 23.72 57.10 65.99	264 393.2 -129.2 42.4541 6.19 11.20 37.08 43.54	107 229.2 -122.2 65.1391 2.51 4.54 25.78 32.05	21 101.6 -80.6 63.9538 0.49 0.89 11.41 14.67	2357 55.22 68.77
cluster 4	38 251.9 -213.9 181.621 0.89 4.85 2.77 98.03	60 111.0 -51.0 23.427 1.41 7.66 9.92 94.38	225 179.6 45.4 11.4732 5.27 28.74 22.98 88.97	216 130.6 85.4 55.9048 5.06 27.59 30.34 73.88	169 76.1 92.9 113.271 3.96 21.58 40.72 72.77	75 33.8 41.2 50.3918 1.76 9.58 40.76 55.43	723 18.35 87.11
cluster 5	27 176.9 -149.9 127.053 0.63 4.91 1.97 100.00	34 78.0 -44.0 24.7913 0.80 6.18 5.62 100.00	108 126.2 -18.2 2.61397 2.53 19.64 11.03 100.00	186 91.8 94.2 96.8101 4.36 33.82 26.12 100.00	113 53.5 59.5 66.2443 2.65 20.55 27.23 100.00	82 23.7 58.3 143.289 1.92 14.91 44.57 100.00	550 12.89 100.00
TOTAL	1373 32.17	605 14.18	979 22.94	712 16.68	415 9.72	184 4.31	4268 100.00

FREQUENCY MISSING = 387

STATISTICS FOR TABLE OF CLUSTER BY GEMEKLAS

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	20	1337.071	0.000
LIKELIHOOD RATIO CHI-SQUARE	20	1420.676	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	955.745	0.000
PHI		0.560	
CONTINGENCY COEFFICIENT		0.488	
CRAMER'S V		0.280	

EFFECTIVE SAMPLE SIZE = 4268
FREQUENCY MISSING = 387

AUTO-AUTO ACCIDENTS 1982

TABLE OF CLUSTER BY VORMANE

CLUSTER(CLUSTERNR)   VORMANE (VORMANI)

CLUSTER	FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	ryder/l[illegible] pen	vanuit s tilstand	afremmen	afslaan rechts	afslaan links	wisseler rijbaan	keren [illegible] nksom	over[illegible]	TOTAL
cluster 1	FREQUENCY	19[illegible]	[illegible]	2[illegible]	23	23	12	4	24	[illegible]
	EXPECTED	176.7	[illegible]	6.[illegible]	11.[illegible]	7[illegible]	[illegible]	6.4	[illegible]	
	DEVIATION	14.3	-[illegible]	13.4	11.[illegible]	-54.5	[illegible]	-2.4	1[illegible]	
	CELL CHI2	1.154[illegible]	5.05[illegible]	3[illegible]	10.6[illegible]	38.[illegible]	7.[illegible]	.87197[illegible]	[illegible].3[illegible]	
	PERCENT	4.4[illegible]	0.[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.[illegible]	[illegible]	6.[illegible]3
	ROW PCT	64.09	0.6[illegible]	6.[illegible]	7.72	[illegible].38	4.03	1.34	8.[illegible]	
	COL PCT	7.55	1.45	21.[illegible]	13.61	2.01	15.19	4.40	34.[illegible]	
	CUM COL%	7.55	1.45	21.[illegible]	13.61	2.01	15.19	4.40	34.[illegible]	[illegible]
cluster 2	FREQUENCY	249	[illegible]	14	4	1	1	0	[illegible]	[illegible]
	EXPECTED	165.0	9.1	6.2	11.[illegible]	71.9	5.2	6.0	[illegible]	
	DEVIATION	83.0	-[illegible]	7.[illegible]	-7.1	-70.9	-4.2	-6.0	-[illegible]	
	CELL CHI2	41.[illegible]	[illegible]	[illegible]	4.5[illegible]	69.9154	3.[illegible]	5.77031	1.41[illegible]	
	PERCENT	5.83	0.[illegible]	0.33	0.09	0.02	0.02	0.00	0.[illegible]	6.56
	ROW PCT	88.[illegible]	[illegible]	5.00	1.43	0.36	0.36	0.00	0.71	
	COL PCT	9.84	[illegible]	14.74	2.37	0.09	1.27	0.00	2.[illegible]0	
	CUM COL%	17.38	[illegible]	35.79	15.98	2.10	16.46	4.40	37.[illegible]8	13.54
cluster 3	FREQUENCY	1350	64	24	102	732	13	47	25	2357
	EXPECTED	139[illegible].7	76.[illegible]	52.5	93.3	605.3	43.6	50.3	38.1	
	DEVIATION	-4[illegible]	[illegible]	-28.[illegible]	8.[illegible]	126.7	-30.6	-3.3	-13.1	
	CELL CHI2	1.53077	1.95[illegible]	15.1427	.805384	26.5366	21.5014	.210786	4.50[illegible]	
	PERCENT	31.63	1.50	0.56	2.39	17.15	0.30	1.10	0.59	55.22
	ROW PCT	57.28	2.[illegible]	1.02	4.33	31.06	0.55	1.99	1.06	
	COL PCT	53.34	46.[illegible]	25.26	60.36	66.79	16.46	51.65	36.23	
	CUM COL%	70.72	54.[illegible]	61.05	76.33	68.89	32.91	56.04	73.[illegible]1	68.77
cluster 4	FREQUENCY	411	60	1	13	278	0	18	2	783
	EXPECTED	464.3	25.3	17.4	31.0	201.1	14.5	16.7	13.7	
	DEVIATION	-53.3	34.7	-16.4	-18.0	76.9	-14.5	1.3	-11.7	
	CELL CHI2	6.12578	47.5128	15.4859	10.4553	29.4334	14.4932	.102056	9.974[illegible]1	
	PERCENT	9.63	1.41	0.02	0.30	6.51	0.00	0.42	0.05	18.35
	ROW PCT	52.49	7.66	0.13	1.66	35.50	0.00	2.30	0.26	
	COL PCT	16.24	43.48	1.05	7.69	25.36	0.00	19.78	2.90	
	CUM COL%	86.96	97.83	62.11	64.02	94.25	32.91	75.82	76.81	87.11
cluster 5	FREQUENCY	330	3	36	27	63	53	22	16	550
	EXPECTED	326.2	17.8	12.2	21.8	141.2	10.2	11.7	8.9	
	DEVIATION	3.8	-14.8	23.8	5.2	-78.2	42.8	10.3	7.1	
	CELL CHI2	.0452[illegible]	12.29[illegible]	46.105	1.23[illegible]	43.3388	180.1[illegible]	8.[illegible]	5.68[illegible]	
	PERCENT	7.73	0.07	0.84	0.63	1.48	1.24	0.52	0.37	12.89
	ROW PCT	60.00	0.55	6.55	4.91	11.45	9.64	4.00	2.91	
	COL PCT	13.04	2.17	37.89	15.98	5.75	67.09	24.18	23.19	
	CUM COL%	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TOTAL		2531	138	95	169	1096	79	91	59	4268
		59.30	3.23	2.23	3.96	25.68	1.85	2.13	1.[illegible]2	100.00

FREQUENCY MISSING = 387

STATISTICS FOR TABLE OF CLUSTER BY VORMANE

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	28	807.810	0.000
LIKELIHOOD RATIO CHI-SQUARE	28	776.722	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	22.307	0.000
PHI		0.435	
CONTINGENCY COEFFICIENT		0.399	
CRAMER'S V		0.218	

EFFECTIVE SAMPLE SIZE = 4268
FREQUENCY MISSING = 387

AUTO-AUTO ACCIDENTS 1982

TABLE OF CLUSTER BY VORMANT

CLUSTER(CLUSTERNR) VORMANT(VORMAN2)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	rijden/lo[illegible]pen	afremmen	stilstaan	afslaan rechts	afslaan links	TOTAL
cluster 1	2[illegible]0 24[illegible].4 -27.9 3.1[illegible]88 5.21 75.60 5.11 6.11	4 12.2 -8.2 5.5086 0.09 1.37 2.26 2.26	3 7.9 -4.9 3.06019 0.07 1.03 2.61 2.61	13 3.3 9.7 28.402 0.31 4.47 27.08 27.08	51 19.6 31.4 50.0802 1.21 17.53 17.89 17.89	291 6.89 6.89
cluster 2	[illegible]3 229.2 -[illegible] [illegible] [illegible].54 [illegible] [illegible].64 [illegible]	140 11.3 128.7 1469.88 3.32 52.04 79.10 81.36	101 7.3 93.7 1197.38 2.39 37.55 87.83 90.43	1 3.1 -2.1 1.[illegible] 0.02 0.37 2.08 29.17	4 18.2 -14.2 11.0355 0.09 1.49 1.40 19.30	26[illegible] 6.37 13.26
cluster 3	2092 199[illegible].2 [illegible] 4.[illegible] 4[illegible].54 8[illegible].29 [illegible].14 [illegible]	22 98.2 -7[illegible] 59.1315 0.52 0.94 12.43 93.79	10 63.8 -5[illegible].8 45.3715 0.24 0.43 8.70 99.13	29 26.6 2.4 .210681 0.69 1.24 60.42 89.58	190 158.1 31.9 6.42612 4.50 8.11 66.67 85.96	2343 55.48 68.74
cluster 4	758 662.0 96.0 13.92 [illegible] 97.55 21.07 85.[illegible]	4 32.6 -28.6 25.058 0.09 0.51 2.26 96.05	1 21.2 -20.2 19.2[illegible]64 0.02 0.13 0.87 100.00	0 8.8 -8.8 8.83164 0.00 0.00 0.00 89.58	14 52.4 -38.4 28.1756 0.33 1.80 4.91 90.88	777 18.40 87.14
cluster 5	[illegible]05 462.6 42.4 [illegible] 11.96 93.00 14.04 100.00	7 22.8 -15.8 10.9119 0.17 1.29 3.95 100.00	0 14.3 -14.3 14.73[illegible] 0.00 0.00 0.00 100.00	5 6.2 -1.2 .222522 0.12 0.92 10.42 100.00	26 36.6 -10.6 3.0726[illegible] 0.62 4.79 9.12 100.00	543 12.86 100.00
TOTAL	3598 85.20	177 4.19	115 2.72	48 1.14	285 6.75	4223 100.00

FREQUENCY MISSING = 432

STATISTICS FOR TABLE OF CLUSTER BY VORMANT

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	16	3199.500	0.000
LIKELIHOOD RATIO CHI-SQUARE	16	1546.188	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	189.122	0.000
PHI		0.870	
CONTINGENCY COEFFICIENT		0.657	
CRAMER'S V		0.435	

EFFECTIVE SAMPLE SIZE = 4223
FREQUENCY MISSING = 432

AUTO-OBJECT ACCIDENTS 1982

TABLE OF CLUSTER BY MAXSNELH

CLUSTER(CLUSTERNR) MAXSNELH(MAX.SNELHEID)

CLUSTER	FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	=< 50 km	70 km	80-90 km	100 km	TOTAL
cluster 1	FREQUENCY	50	24	1218	31	1323
	EXPECTED	588.2	51.6	585.5	97.7	
	DEVIATION	-538.2	-27.6	632.5	-66.7	
	CELL CHI2	492.45	14.7292	683.242	45.5671	
	PERCENT	1.69	0.81	41.27	1.05	44.83
	ROW PCT	3.78	1.81	92.06	2.34	
	COL PCT	3.81	20.87	93.26	14.22	
	CUM COL%	3.81	20.87	93.26	14.22	44.83
cluster 2	FREQUENCY	50	39	29	187	305
	EXPECTED	135.6	11.9	135.0	22.5	
	DEVIATION	-85.6	27.1	-106.0	164.5	
	CELL CHI2	54.0379	61.8536	83.2119	1200.55	
	PERCENT	1.69	1.32	0.98	6.34	10.34
	ROW PCT	16.39	12.79	9.51	61.31	
	COL PCT	3.81	33.91	2.22	85.78	
	CUM COL%	7.62	54.78	95.48	100.00	55.17
cluster 3	FREQUENCY	1212	52	59	0	1323
	EXPECTED	588.2	51.6	585.5	97.7	
	DEVIATION	623.8	0.4	-526.5	-97.7	
	CELL CHI2	661.557	.003805	473.455	97.7343	
	PERCENT	41.07	1.76	2.00	0.00	44.83
	ROW PCT	91.61	3.93	4.46	0.00	
	COL PCT	92.38	45.22	4.52	0.00	
	CUM COL%	100.00	100.00	100.00	100.00	100.00
TOTAL		1312	115	1306	218	2951
		44.46	3.90	44.26	7.39	100.00

FREQUENCY MISSING = 211

STATISTICS FOR TABLE OF CLUSTER BY MAXSNELH

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	6	3868.388	0.000
LIKELIHOOD RATIO CHI-SQUARE	6	3607.650	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	1847.237	0.000
PHI		1.145	
CONTINGENCY COEFFICIENT		0.753	
CRAMER'S V		0.810	

EFFECTIVE SAMPLE SIZE = 2951
FREQUENCY MISSING = 211

AUTO-OBJECT ACCIDENTS 1982

TABLE OF CLUSTER BY WEGSITUA

CLUSTER(CLUSTERNR)   WEGSITUA(WEGSITUATIE)

CLUSTER	FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	rechte weg	kruising	T/Y kruising	bocht	TOTAL
cluster 1	FREQUENCY	707	9	37	574	1327
	EXPECTED	681.5	71.4	117.7	456.4	
	DEVIATION	25.5	-62.4	-80.7	117.6	
	CELL CHI2	.956521	54.5602	55.3277	30.2966	
	PERCENT	23.93	0.30	1.25	19.43	44.92
	ROW PCT	53.28	0.68	2.79	43.26	
	COL PCT	46.61	5.66	14.12	56.50	
	CUM COL%	46.61	5.66	14.12	56.50	44.92
cluster 2	FREQUENCY	204	13	28	58	303
	EXPECTED	155.6	16.3	26.9	104.2	
	DEVIATION	48.4	-3.3	1.1	-46.2	
	CELL CHI2	15.0529	.671403	.047173	20.4937	
	PERCENT	6.91	0.44	0.95	1.96	10.26
	ROW PCT	67.33	4.29	9.24	19.14	
	COL PCT	13.45	8.18	10.69	5.71	
	CUM COL%	60.05	13.84	24.81	62.20	55.18
cluster 3	FREQUENCY	606	137	197	384	1324
	EXPECTED	679.9	71.3	117.4	455.4	
	DEVIATION	-73.9	65.7	79.6	-71.4	
	CELL CHI2	8.03818	60.6349	53.9164	11.1879	
	PERCENT	20.51	4.64	6.67	13.00	44.82
	ROW PCT	45.77	10.35	14.88	29.00	
	COL PCT	39.95	86.16	75.19	37.80	
	CUM COL%	100.00	100.00	100.00	100.00	100.00
TOTAL		1517	159	262	1016	2954
		51.35	5.38	8.87	34.39	100.00

FREQUENCY MISSING = 208

STATISTICS FOR TABLE OF CLUSTER BY WEGSITUA

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	6	311.184	0.000
LIKELIHOOD RATIO CHI-SQUARE	6	343.655	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	2.761	0.097
PHI		0.325	
CONTINGENCY COEFFICIENT		0.309	
CRAMER'S V		0.230	

EFFECTIVE SAMPLE SIZE = 2954
FREQUENCY MISSING = 208

AUTO-OBJECT ACCIDENTS 1982

TABLE OF CLUSTER BY LICHTGES

CLUSTER(CLUSTERNR) LICHTGES(LICHTGESTELDH)

FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	daglicht	duisternis	schemer	TOTAL
cluster 1	639 539.0 100.0 18.566 21.61 48.15 53.21 53.21	642 744.5 -102.5 14.1124 21.71 48.38 38.70 38.70	46 43.5 2.5 .140123 1.56 3.47 47.42 47.42	1327 44.88 44.88
cluster 2	162 124.3 37.7 11.446 5.48 52.94 13.49 66.69	128 171.7 -43.7 11.1128 4.33 41.83 7.72 46.41	16 10.0 6.0 3.54128 0.54 5.23 16.49 63.92	306 10.35 55.22
cluster 3	400 537.7 -137.7 35.2856 13.53 30.21 33.31 100.00	889 742.8 146.2 28.7673 30.06 67.15 53.59 100.00	35 43.4 -8.4 1.63696 1.18 2.64 36.08 100.00	1324 44.78 100.00
TOTAL	1201 40.62	1659 56.10	97 3.28	2957 100.00

FREQUENCY MISSING = 205

STATISTICS FOR TABLE OF CLUSTER BY LICHTGES

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	124.608	0.000
LIKELIHOOD RATIO CHI-SQUARE	4	125.646	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	64.795	0.000
PHI		0.205	
CONTINGENCY COEFFICIENT		0.201	
CRAMER'S V		0.145	

EFFECTIVE SAMPLE SIZE = 2957
FREQUENCY MISSING = 205

AUTO-OBJECT ACCIDENTS 1982

TABLE OF CLUSTER BY GEMEKLAS

CLUSTER(CLUSTERNR) GEMEKLAS(GEM.KLASSE)

CLUSTER	Statistic	> 100.000	50 - 100	20 - 50	10 - 20	5 - 10	< 5.000	TOTAL
cluster 1	FREQUENCY	55	105	279	418	337	134	1328
	EXPECTED	304.3	163.8	281.8	280.9	213.6	83.5	
	DEVIATION	-249.3	-58.8	-2.8	137.1	123.4	50.5	
	CELL CHI2	204.228	21.1148	0.02875	66.8555	71.247	30.5784	
	PERCENT	1.86	3.55	9.43	14.13	11.39	4.53	44.88
	ROW PCT	4.14	7.91	21.01	31.48	25.38	10.09	
	COL PCT	8.11	28.77	44.43	66.77	70.80	72.04	
	CUM COL%	8.11	28.77	44.43	66.77	70.80	72.04	44.88
cluster 2	FREQUENCY	106	39	60	47	36	18	306
	EXPECTED	70.1	37.7	64.9	64.7	49.2	19.2	
	DEVIATION	35.9	1.3	-4.9	-17.7	-13.2	-1.2	
	CELL CHI2	18.367	0.04167	.376308	4.85956	3.55296	.079279	
	PERCENT	3.58	1.32	2.03	1.59	1.22	0.61	10.34
	ROW PCT	34.64	12.75	19.61	15.36	11.76	5.88	
	COL PCT	15.63	10.68	9.55	7.51	7.56	9.68	
	CUM COL%	23.75	39.45	53.98	74.28	78.36	81.72	55.22
cluster 3	FREQUENCY	517	221	289	161	103	34	1325
	EXPECTED	303.6	163.4	281.2	280.3	213.1	83.3	
	DEVIATION	213.4	57.6	7.8	-119.3	-110.1	-49.3	
	CELL CHI2	150	20.2697	.215804	50.7855	56.9197	29.1678	
	PERCENT	17.47	7.47	9.77	5.44	3.48	1.15	44.78
	ROW PCT	39.02	16.68	21.81	12.15	7.77	2.57	
	COL PCT	76.25	60.55	46.02	25.72	21.64	18.28	
	CUM COL%	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TOTAL		678	365	628	626	476	186	2959
		22.91	12.34	21.22	21.16	16.09	6.29	100.00

FREQUENCY MISSING = 203

STATISTICS FOR TABLE OF CLUSTER BY GEMEKLAS

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	10	728.688	0.000
LIKELIHOOD RATIO CHI-SQUARE	10	814.940	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	657.990	0.000
PHI		0.496	
CONTINGENCY COEFFICIENT		0.445	
CRAMER'S V		0.351	

EFFECTIVE SAMPLE SIZE = 2959
FREQUENCY MISSING = 203

AUTO-OBJECT ACCIDENTS 1982

TABLE OF CLUSTER BY VOR3MANE

CLUSTER(CLUSTERNR) VOR3MANE(VOR3MAN1)

CLUSTER	FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	ryden/lopen	afremmen/stoppen	afslaan rechts	afslaan links	wisselen rybaan	overig	TOTAL
cluster 1	FREQUENCY	1260	42	4	5	13	3	1327
	EXPECTED	1162.4	50.8	31.9	52.6	16.6	12.6	
	DEVIATION	97.6	-8.8	-27.9	-47.6	-3.6	-9.6	
	CELL CHI2	8.19179	1.52872	24.4283	43.0875	.795505	7.30578	
	PERCENT	42.70	1.42	0.14	0.17	0.44	0.10	44.97
	ROW PCT	94.95	3.17	0.30	0.38	0.98	0.23	
	COL PCT	48.74	37.17	5.63	4.27	35.14	10.71	
	CUM COL%	48.74	37.17	5.63	4.27	35.14	10.71	44.97
cluster 2	FREQUENCY	242	22	6	4	17	14	305
	EXPECTED	267.2	11.7	7.3	12.1	3.8	2.9	
	DEVIATION	-25.2	10.3	-1.3	-8.1	13.2	11.1	
	CELL CHI2	2.37164	9.12067	.244032	5.41564	45.3969	42.6218	
	PERCENT	8.20	0.75	0.20	0.14	0.58	0.47	10.34
	ROW PCT	79.34	7.21	1.97	1.31	5.57	4.59	
	COL PCT	9.36	19.47	8.45	3.42	45.95	50.00	
	CUM COL%	58.10	56.64	14.08	7.69	81.08	60.71	55.30
cluster 3	FREQUENCY	1083	49	61	108	7	11	1319
	EXPECTED	1155.4	50.5	31.7	52.3	16.5	12.5	
	DEVIATION	-72.4	-1.5	29.3	55.7	-9.5	-1.5	
	CELL CHI2	4.53797	.044982	26.9881	59.3369	5.5007	.183416	
	PERCENT	36.70	1.66	2.07	3.66	0.24	0.37	44.70
	ROW PCT	82.11	3.71	4.62	8.19	0.53	0.83	
	COL PCT	41.90	43.36	85.92	92.31	18.92	39.29	
	CUM COL%	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TOTAL		2585	113	71	117	37	28	2951
		87.60	3.83	2.41	3.96	1.25	0.95	100.00

FREQUENCY MISSING = 211

STATISTICS FOR TABLE OF CLUSTER BY VOR3MANE

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	10	287.100	0.000
LIKELIHOOD RATIO CHI-SQUARE	10	273.870	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	90.722	0.000
PHI		0.312	
CONTINGENCY COEFFICIENT		0.298	
CRAMER'S V		0.221	

EFFECTIVE SAMPLE SIZE = 2951
FREQUENCY MISSING = 211

AUTO-OBJECT ACCIDENTS 1982

TABLE OF CLUSTER BY ALCOHOLE

CLUSTER(CLUSTERNR)   ALCOHOLE(ALCOHOL1)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	geen	wel	art. 26 WVW	onbekend	TOTAL
cluster 1	915	145	211	57	1328
	849.1	150.3	276.5	52.1	
	65.9	-5.3	-65.5	4.9	
	5.10975	.190239	15.5	.468594	
	30.92	4.90	7.13	1.93	44.88
	68.90	10.92	15.89	4.29	
	48.36	43.28	34.25	49.14	
	48.36	43.28	34.25	49.14	44.88
cluster 2	234	24	29	19	306
	195.7	34.6	63.7	12.0	
	38.3	-10.6	-34.7	7.0	
	7.51367	3.26998	18.9046	4.08945	
	7.91	0.81	0.98	0.64	10.34
	76.47	7.84	9.48	6.21	
	12.37	7.16	4.71	16.38	
	60.73	50.45	38.96	65.52	55.22
cluster 3	743	166	376	40	1325
	847.2	150.0	275.8	51.9	
	-104.2	16.0	100.2	-11.9	
	12.8187	1.70477	36.3721	2.74609	
	25.11	5.61	12.71	1.35	44.78
	56.08	12.53	28.38	3.02	
	39.27	49.55	61.04	34.48	
	100.00	100.00	100.00	100.00	100.00
TOTAL	1892	335	616	116	2959
	63.94	11.32	20.82	3.92	100.00

FREQUENCY MISSING = 203

STATISTICS FOR TABLE OF CLUSTER BY ALCOHOLE

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	6	108.688	0.000
LIKELIHOOD RATIO CHI-SQUARE	6	111.126	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	39.041	0.000
PHI		0.192	
CONTINGENCY COEFFICIENT		0.188	
CRAMER'S V		0.136	

EFFECTIVE SAMPLE SIZE = 2959
FREQUENCY MISSING = 203

AUTO-NO OBJECT ACCIDENTS 1982

TABLE OF CLUSTER BY MAXSNELH

CLUSTER(CLUSTERNR)   MAXSNELH(MAX.SNELHEID)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	=< 50 km	70 km	80-90 km	100 km	TOTAL
cluster 1	15 71.3 -56.3 44.4501 1.76 2.97 12.50 12.50	8 13.1 -5.1 1.96709 0.94 1.58 36.36 36.36	418 303.0 115.0 43.6469 49.18 82.77 81.96 81.96	64 117.6 -53.6 24.4548 7.53 12.67 32.32 32.32	505 59.41 59.41
cluster 2	1 19.1 -18.1 17.1113 0.12 0.74 0.83 13.33	0 3.5 -3.5 3.49412 0.00 0.00 0.00 36.36	18 81.0 -63.0 49 2.12 13.33 3.53 85.49	116 31.4 84.6 227.341 13.65 85.93 58.59 90.91	135 15.88 75.29
cluster 3	100 19.5 80.5 332.767 11.76 72.46 83.33 96.67	13 3.6 9.4 24.8873 1.53 9.42 59.09 95.45	20 82.8 -62.8 47.6309 2.35 14.49 3.92 89.41	5 32.1 -27.1 22.9236 0.59 3.62 2.53 93.43	138 16.24 91.53
cluster 4	4 10.2 -6.2 3.73678 0.47 5.56 3.33 100.00	1 1.9 -0.9 .400145 0.12 1.39 4.55 100.00	54 43.2 10.8 2.7 6.35 75.00 10.59 100.00	13 16.8 -3.8 .848224 1.53 18.06 6.57 100.00	72 8.47 100.00
TOTAL	120 14.12	22 2.59	510 60.00	198 23.29	850 100.00

FREQUENCY MISSING = 1

STATISTICS FOR TABLE OF CLUSTER BY MAXSNELH

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	9	847.361	0.000
LIKELIHOOD RATIO CHI-SQUARE	9	673.289	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	82.511	0.000
PHI		0.998	
CONTINGENCY COEFFICIENT		0.707	
CRAMER'S V		0.576	

EFFECTIVE SAMPLE SIZE = 850
FREQUENCY MISSING = 1

AUTO-NO OBJECT ACCIDENTS 1982

TABLE OF CLUSTER BY WEGSITUA

CLUSTER(CLUSTERNR)   WEGSITUA(WEGSITUATIE)

FREQUENCY / EXPECTED / DEVIATION / CELL CHI2 / PERCENT / ROW PCT / COL PCT / CUM COL%	rechte weg	kruising T/Y	bocht	TOTAL
cluster 1	249 266.4 -17.4 1.14222 29.26 49.31 55.46 55.46	10 35.0 -25.0 17.8679 1.18 1.98 16.95 16.95	246 203.5 42.5 8.85615 28.91 48.71 71.72 71.72	505 59.34 59.34
cluster 2	112 71.2 40.8 23.3385 13.16 82.96 24.94 80.40	7 9.4 -2.4 .594256 0.82 5.19 11.86 28.81	16 54.4 -38.4 27.1173 1.88 11.85 4.66 76.38	135 15.86 75.21
cluster 3	51 72.8 -21.8 6.53353 5.99 36.96 11.36 91.76	39 9.6 29.4 90.5421 4.58 28.26 66.10 94.92	48 55.6 -7.6 1.04436 5.64 34.78 13.99 90.38	138 16.22 91.42
cluster 4	37 38.5 -1.5 0.05966 4.35 50.68 8.24 100.00	3 5.1 -2.1 .839373 0.35 4.11 5.08 100.00	33 29.4 3.6 .434853 3.88 45.21 9.62 100.00	73 8.58 100.00
TOTAL	449 52.76	59 6.93	343 40.31	851 100.00

STATISTICS FOR TABLE OF CLUSTER BY WEGSITUA

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	6	178.371	0.000
LIKELIHOOD RATIO CHI-SQUARE	6	153.181	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	1.570	0.210
PHI		0.458	
CONTINGENCY COEFFICIENT		0.416	
CRAMER'S V		0.324	

SAMPLE SIZE = 851

AUTO-NO OBJECT ACCIDENTS 1982

TABLE OF CLUSTER BY LICHTGES

CLUSTER(CLUSTERNR)   LICHTGES(LICHTGESTELDH)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	daglicht	duisternis	schemer	TOTAL
cluster 1	232 257.5 -25.5 2.53354 27.26 45.94 53.46 53.46	252 224.9 27.1 3.26076 29.61 49.90 66.49 66.49	21 22.5 -1.5 .106533 2.47 4.16 55.26 55.26	505 59.34 59.34
cluster 2	101 68.8 32.2 15.0145 11.87 74.81 23.27 76.73	24 60.1 -36.1 21.7037 2.82 17.78 6.33 72.82	10 6.0 4.0 2.6169 1.18 7.41 26.32 81.58	135 15.86 75.21
cluster 3	62 70.4 -8.4 .997425 7.29 44.93 14.29 91.01	74 61.5 12.5 2.53884 8.70 53.62 19.53 92.35	2 6.2 -4.2 2.81128 0.24 1.45 5.26 86.84	138 16.22 91.42
cluster 4	39 37.2 1.8 .084233 4.58 53.42 8.99 100.00	29 32.5 -3.5 .379201 3.41 39.73 7.65 100.00	5 3.3 1.7 .929125 0.59 6.85 13.16 100.00	73 8.58 100.00
TOTAL	434 51.00	379 44.54	38 4.47	851 100.00

STATISTICS FOR TABLE OF CLUSTER BY LICHTGES

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	6	52.999	0.000
LIKELIHOOD RATIO CHI-SQUARE	6	57.860	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	1.325	0.250
PHI		0.250	
CONTINGENCY COEFFICIENT		0.242	
CRAMER'S V		0.176	

SAMPLE SIZE = 851

AUTO-NO OBJECT ACCIDENTS 1982

TABLE OF CLUSTER BY GEMEKLAS

CLUSTER(CLUSTERNR)   GEMEKLAS(GEM.KLASSE)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	> 100.000	50 - 100	20 - 50	10 - 20	5 - 10	< 5.000	TOTAL
cluster 1	11 45.7 -34.[illegible] 25.[illegible] 1.29 2.1[illegible] 14.29 14.[illegible]	22 3[illegible].8 -1[illegible].8 7.[illegible] 2.59 4.36 3[illegible].84 32.84	111 115.9 -[illegible].9 .2[illegible] 13.04 21.99 5[illegible].35 5[illegible].35	169 148.9 20.1 2.[illegible]4 1[illegible].86 33.47 67.33 67.33	115 98.5 17.[illegible] 3.1[illegible] 13.6[illegible] 2[illegible].9[illegible] 69.88 69.88	7[illegible] [illegible] [illegible] 7.8[illegible] 8.[illegible] 15.[illegible] 81.7[illegible] 81.7[illegible]	5[illegible]5 59.[illegible]4 59.[illegible]4
cluster 2	[illegible] [illegible] [illegible] 9.5[illegible] [illegible] 17.04 29.87 44.1[illegible]	15 1[illegible].5 [illegible].4 1.79783 1.76 11.11 22.39 55.22	32 3[illegible] [illegible] .017929 3.7[illegible] 23.70 16.24 72.59	38 3[illegible].8 -1.8 .9[illegible]93 4.47 28.15 15.14 8[illegible].47	22 26.3 -4.3 .713198 2.59 16.30 13.25 83.13	5 14.[illegible] -9.[illegible] 6.44778 0.5[illegible] 3.70 5.38 87.10	[illegible] 15.[illegible] 75.[illegible]
cluster 3	41 [illegible] [illegible] 65.[illegible] 4.[illegible] [illegible] [illegible] 97.4[illegible]	24 1[illegible].9 13.1 1[illegible].[illegible] [illegible].82 1[illegible] [illegible] 91.04	32 31.9 0.1 91[illegible] 3.76 2[illegible].1[illegible] [illegible] [illegible].83	24 40.7 -16.7 [illegible].[illegible]41 2.82 17.39 [illegible] 9[illegible].03	13 26.9 -13.9 7.19[illegible] 1.53 9.42 7.8[illegible] 90.96	4 15.1 -11.1 8.142[illegible] 0.47 2.[illegible] 4.[illegible] 91.40	[illegible]8 16.[illegible]2 91.[illegible]2
cluster 4	[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] 1[illegible]	6 5.7 [illegible] .011106 0.71 8.[illegible] 8.96 100.00	22 1[illegible].9 [illegible] 1.5[illegible] 2.5[illegible] [illegible] 1[illegible].17 1[illegible]	20 21.5 [illegible] .10[illegible] 2.35 2[illegible].[illegible] 7.97 100.00	15 14.2 0.8 .04[illegible] 1.76 2[illegible] 9.04 100.00	6 8.0 [illegible] 6249[illegible]E-[illegible] 0.[illegible]4 10.[illegible] 8.[illegible] 100.00	[illegible]3 8.[illegible]5 100.00
TOTAL	77 [illegible]	67 7.87	197 2[illegible].15	251 29.49	166 19.5[illegible]	9[illegible] 10.[illegible]	851 100.00

STATISTICS FOR TABLE OF CLUSTER BY GEMEKLAS

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	15	174.904	0.000
LIKELIHOOD RATIO CHI-SQUARE	15	165.795	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	65.926	0.000
PHI		0.453	
CONTINGENCY COEFFICIENT		0.413	
CRAMER'S V		0.262	

SAMPLE SIZE = 851

AUTO-NO OBJECT ACCIDENTS 1982

TABLE OF CLUSTER BY VOR3MANE

CLUSTER(CLUSTERNR)   VOR3MANE(VOR3MAN1)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	ryden/lopen	afremmen/stoppen	afslaan	wisselen baan	5	TOTAL
cluster 1	479 450.9 28.1 1.75484 5[illegible].42 94.85 63.19 63.19	10 18.4 -8.4 3.86253 1.18 1.98 32.26 32.26	3 18.4 -15.4 12.9274 0.35 0.59 9.68 9.68	11 10.1 0.9 0.078 1.30 2.18 64.71 64.71	2 7.1 -5.1 3.69821 0.24 0.40 16.67 16.67	505 59.48 59.48
cluster 2	114 120.5 -6.5 .353782 13.43 84.44 15.04 78.23	8 4.9 3.1 1.91284 0.94 5.93 25.81 58.06	0 4.9 -4.9 4.92933 0.00 0.00 0.00 9.68	5 2.7 2.3 1.95155 0.59 3.70 29.41 94.12	8 1.9 6.1 19.4489 0.94 5.93 66.67 83.33	135 15.90 75.38
cluster 3	98 121.4 -23.4 4.51534 11.54 72.06 12.93 91.16	10 5.0 5.0 5.10341 1.18 7.35 32.26 90.32	26 5.0 21.0 89.0958 3.06 19.12 83.87 93.55	0 2.7 -2.7 2.7232 0.00 0.00 0.00 94.12	2 1.9 0.1 .003144 0.24 1.47 16.67 100.00	136 16.02 91.40
cluster 4	67 65.2 1.8 .051074 7.89 91.78 8.84 100.00	3 2.7 0.3 0.04198 0.35 4.11 9.68 100.00	2 2.7 -0.7 .166152 0.24 2.74 6.45 100.00	1 1.5 -0.5 .145845 0.12 1.37 5.88 100.00	0 1.0 -1.0 1.0319 0.00 0.00 0.00 100.00	73 8.60 100.00
TOTAL	758 89.28	31 3.65	31 3.65	17 2.00	12 1.41	849 100.00

FREQUENCY MISSING = 2

STATISTICS FOR TABLE OF CLUSTER BY VOR3MANE

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	12	153.798	0.000
LIKELIHOOD RATIO CHI-SQUARE	12	116.478	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	14.785	0.000
PHI		0.426	
CONTINGENCY COEFFICIENT		0.392	
CRAMER'S V		0.246	

EFFECTIVE SAMPLE SIZE = 849
FREQUENCY MISSING = 2
WARNING: 60% OF THE CELLS HAVE EXPECTED COUNTS LESS THAN 5. CHI-SQUARE MAY NOT BE A VALID TEST.

AUTO-NO OBJECT ACCIDENTS 1982

TABLE OF CLUSTER BY ALCOHOLE

CLUSTER(CLUSTERNR)   ALCOHOLE(ALCOHOL1)

FREQUENCY EXPECTED DEVIATION CELL CHI2 PERCENT ROW PCT COL PCT CUM COL%	geen	wel	art. 26 WVW	onbekend	TOTAL
cluster 1	340	63	83	19	505
	367.9	46.3	71.2	19.6	
	-27.9	16.7	11.8	-0.6	
	2.11875	6.03486	1.95191	.017347	
	39.95	7.40	9.75	2.23	59.34
	67.33	12.48	16.44	3.76	
	54.84	80.77	69.17	57.58	
	54.84	80.77	69.17	57.58	59.34
cluster 2	123	2	5	5	135
	98.4	12.4	19.0	5.2	
	24.6	-10.4	-14.0	-0.2	
	6.17541	8.69694	10.3497	.010551	
	14.45	0.24	0.59	0.59	15.86
	91.11	1.48	3.70	3.70	
	19.84	2.56	4.17	15.15	
	74.68	83.33	73.33	72.73	75.21
cluster 3	91	12	29	6	138
	100.5	12.6	19.5	5.4	
	-9.5	-0.6	9.5	0.6	
	.905325	.033264	4.67752	.078624	
	10.69	1.41	3.41	0.71	16.22
	65.94	8.70	21.01	4.35	
	14.68	15.38	24.17	18.18	
	89.35	98.72	97.50	90.91	91.42
cluster 4	66	1	3	3	73
	53.2	6.7	10.3	2.8	
	12.8	-5.7	-7.3	0.2	
	3.08807	4.84041	5.16809	.010115	
	7.76	0.12	0.35	0.35	8.58
	90.41	1.37	4.11	4.11	
	10.65	1.28	2.50	9.09	
	100.00	100.00	100.00	100.00	100.00
TOTAL	620	78	120	33	851
	72.86	9.17	14.10	3.88	100.00

STATISTICS FOR TABLE OF CLUSTER BY ALCOHOLE

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	9	54.157	0.000
LIKELIHOOD RATIO CHI-SQUARE	9	66.099	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	4.957	0.026
PHI		0.252	
CONTINGENCY COEFFICIENT		0.245	
CRAMER'S V		0.146	

SAMPLE SIZE = 851

APPENDIX III. SUMMARY OF THE RESULTS

Auto-Bicycle Accidents (N=6333, with 496 cases deleted)

AB1 (N=3058)

This day-time type of accident tends tends to occur at a pedestrian or bicycle crossing inside built-up areas, yet is not especially urban in nature. A somewhat middle-aged driver is likely to be just driving straight ahead, when either a very young or very old bicyclist suddenly crosses the road or turns into the path of the automobile.

AB2 (N=1707)

This type of accident tends to occur during the winter months, during the evening or night, and during poor weather conditions. The accident location tends more often to be an intersection in an urban area. Alcohol seems to play a role with the driver and possibly with the cyclist, who is neither very young or elderly. The driver is either turning or accelerating from a stand still, when he strikes the bicyclist who is most likely to be innocently riding straight ahead on a bicycle path.

AB3 (N=615)

This rather serious type of accident tends to be more of a summer, good weather, and daylight accident on a rural, high-speed, straight road section or curve. There is an increased likelihood that there is a road exit/entrance or a pedestrian/bicycle crossing in the neighborhoood. The either very young or very old, male bicyclist is most likely to be crossing the road, turning, or changing lanes when he is struck by the driver, who is 'innocently'just driving straight ahead.

AB4 (N=457)

This accident also seems to be a summer, good weather accident during daylight conditions at a straight road section inside a built-up area. It does not seem to be an especially urban type of accident. A road exit/entrance or a parking lot is more likely to be in the vicinity. The driver is more likely to be on the left hand side of the road, at a road exit/entrance or on a parallel road, and is apparently executing some 'unusual' manoeuvre such as braking, when he strikes the middle-aged cyclist who is innocently riding straight ahead on the right hand side of the road. The driver is often blamed for failure to yield right-of-way, following too closely, or driving too much to the right, while the cyclist is mainly exonerated, even though there is a chance that he is also executing some manoeuvre which may have led to the accident. It appears that either the bicylist runs into a suddenly braking car or is side-swiped.

Auto-Moped Accidents (N=6263, with 209 cases deleted)

AM1 (N=2241)

This type of accident seems to be generally describable as a daytime, (winter), urban-intersection accident, with an older driver either turning or accelerating out of a stand still, when he strikes an older female moped rider who is riding straight ahead on a bicycle path.

AM2 (N=2234)

This type of accident is more of a night-time, urban intersection type of accident, where the driver is mainly just driving straight ahead when he strikes a very young or old moped rider who crosses his path and fails to yield right of way.

AM3 (N=623)

This relatively serious type of summer accident tends to occur on straight sections or curves on a rural 80 km per hour road, when a male driver (who may have been drinking), driving straight ahead, drives straight into a very young or old moped rider who is either turning or crossing the road. 'Unpredictable' moped manoeuvres and high speeds may play an important role in this type of accident.

AM4 (N=596)

This daytime straight road section (or road exit/entrance or parking lot) type of accident is more likely to occur inside a built-up area, yet doesn't seem to be a typically 'urban' type of accident. The driver may be braking, making a U-turn, or some type of manoeuvre other than driving straight ahead or turning, when he hits the moped rider, who (while not on a bicycle path) is driving straight ahead or changing lanes. This accident accident type seems to include a great variety of manoeuvres, and may indicate a complex group of behaviors on low-volume roads where there is uncertainty about expected manoeuvres.

AM5 (N=360)

This type of accident tends to happen on high-speed rural roads at a T or Y intersection when the driver, turning onto the other road, fails to yield right-of-way and strikes the (somewhat older) moped rider, who is riding straight ahead on a bicycle path.

Auto-Pedestrian Accidents (N=2900, with 211 cases deleted)

AP1 (N=1933)

This dry weather, daytime, type of accident tends to occur on straight road sections inside a built-up area, yet is not especially urban in nature. The driver, who is less likely to have been drinking, is driving straight ahead on the right hand side of the road when he strikes a young, male, child who 'suddenly' crosses the street from the sidewalk, and possibly from behind an object. This less than lethal type of accident seems to represent the young child mid-block dart-out (even though it may include adult mid-block dart-outs).

AP2 (N=450)

This type of accident tends to occur during the winter months during weekend mornings. Neither weather conditions nor (natural) lighting conditions are optimal. The location tends to be an urban intersection or pedestrian crossing on a multi-lane road. The driver, who is turning or accelerating from a stand still then strikes the adult female pedestrian who is likely to be crossing the road on a pedestrian crossing. The driver is likely to be charged with failure to yield right-of-way or neglecting to obey a traffic light or sign, even though it may happen that the pedestrian may be crossing against the traffic light.

AP3 (N=154)

This serious type of accident tends to occur on a rural, high-speed road. The mainly middle-aged pedestrian is more likely to be on the shoulder of the road or crossing in mid-block, without the benefit of a pedestrian crossing, when he is struck.

AP4 (N=152)

This weekend, nighttime accident tends to occur inside built-up areas, even though the location is not especially 'urban' in nature. The driver as well as the adult pedestrian may have been drinking. The driver seems to be at some 'unusual' location and is implementing some 'unusual' manoeuvre when he strikes the pedestrian who is clearly not crossing but walking along or standing on the road. Small numbers, as well as the complications of alcohol and darkness, make this type of accident difficult to interpret.

Auto-Auto Accidents (N=4655, with 387 cases deleted)

AA1 (N=2357

This cluster of accidents seems to represent a run-of-the-mill urban intersection accident where the first vehicle is turning and fails to yield the right of way or to obey a traffic control device.

AA2 (N=783)

This cluster describes day-time accidents at an intersection of a rural, high-speed road, where there seems to be some inequality between the arms of the intersection. The first vehicle is likely to be turning or crossing, when he fails to yield right of way to the second vehicle. This type of accidents tends to be rather serious, and tends to involve two older drivers.

AA3 (N=550)

This very serious group of late afternoon and evening accidents tends to occur on high speed rural roads on a straight road section, curve, or a road exit/entrance. The driver of the first vehicle is more likely to be a young male, who may have been drinking, and is more likely to be on the wrong side of the road or some other 'unsual' place, and is either braking, changing lanes, making a U-turn or some other 'unusual' manoeuvre, when he is struck frontally (even though bumper-bumper collisions sometimes occur in this category) by the other vehicle. Apparently, the driver of the first vehicle left his lane for some reason, resulting mainly in a head-on collision. (Skidding and winter months are also implicated).

AA4 (N=298)

These urban, nighttime type of accident tends to occur on straight road section or curves, and tends to occur somewhat more often during the winter months in the rain. Both drivers, of whom the first of which is more likely to be male, are more likely to be in their twenties and have quite probably been accused of drinking and driving. There is an increased chance that one or the other driver is somewhere 'else' than in their own lane, and there is an increased likelihood that one or both drivers are either turning, braking, or doing something 'else'. Skidding is also implicated. Either the alcohol and/or the rain created a somewhat unpredictable situation, which fortunately doesn't often lead to fatalities.

AA5 (N=280)

This small cluster describes an afternoon, urban, straight road section of a multi-lane-road type of accident. The somewhat older driver of the second vehicle is either braking or standing still, when he is struck from behind by the first driver, who has a better chance of being a female. Perhaps this situation can be summarized as an arterial, urban, non-intersection), bumper-to-bumper type of accident.

Auto-Object Accidents (N=3162, with 203 cases deleted)

A01 (N=1328)

This rather serious non-winter, afternoon type of accident tends to occur on a curve of a high-speed, single-lane rural road. The driver, who is just driving straight ahead, apparently just drives off the road and runs into a tree located on the shoulder. Mist may have something to do with this.

A02 (N=1325)

This winter evening type of accident tends to occur on urban roads, and intersections are over-represented. The driver is more likely to have been drinking, and, while turning, is likely to strike a lantern pole located on a safety island or sidewalk.

A03 (N=306)

This type of accident occur typically during the daytime on 100 km per hour urban area roads outside built-up areas. They also tend to occur on straight road sections or near a road exit/entrance. The driver, who is more likely to be between 26 and 65 years old, is braking, changing lanes, merging, or implementing some other manouevre, when he strikes a safety rail or some fixed object. He may have fallen asleep, become ill, skidded, or for some reason didn't keep far enough to the right.

Auto-No object Accidents (N=851, with 22 cases deleted)

AN1 (N=474)

This night-time accident tends to occur on curves on a 80 km per hour road outside built-up areas in a rural environment. The road surface may be wet. The driver is more likely to be male and alcohol may be implicated. He appears to be driving on the right side of a two lane road, implementing no special manoeuvre, when he drives off the road and overturns or runs into a ditch. High speed may also be implicated.

AN2 (N=223)

These day-time, fair weather accidents tend to occur on multi-lane 100 km per hour roads outside built-up areas, yet are not clearly rural in nature. The vehicle may also have been at a road exits/entrances or on the wrong side of the road. These accidents tend to occur on straight road sections, when the driver, who is more likely to be female and less likely to have been drinking, apparently is merging or changes lanes and 'loses control', runs off the road, and possible overturns.

AN3 (N=68)

These accidents are clearly urban and tend to to occur inside built-up areas, in contradiction to the previous two categories, and tends to be less serious in nature. Otherwise, statistical problems makes interpretation difficult.

AN4 (N=64)

This winter month, morning accident has no other special characteristics except that the road surface is likely to be covered with snow or ice. The driver, who has not been drinking and is more likely to be a middle-aged woman, apparently skidded on a slippery surface.