DETECTION AND ANALYSIS OF BLACK SPOTS WITH EVEN SMALL ACCIDENT FIGURES

S. Oppe

Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV

1. Accident black-spot techniques

Accident black spots are usually defined as road locations with (relatively) high accident potentials.
In order to detect such a hazardous location, we have to know the probability of an accident for a traffic situation of some kind (e.g. the crossing of a pedestrian or the encounter between two cars), or the mean number of accidents for some unit of time.
The comparison of the probability or the mean with some norm (absolute black spots) or with the probability or mean of other locations (relative black spots) may result in the detection of a black spot.
There are a lot of problems related to this definition. In order to define our sample space, we have to know what is and what is not an accident. Furthermore there are weighting problems if one is interested in loss resulting from accidents instead of accidents themselves (e.g. weighting with respect to severity). Although these problems are in general underestimated, we will not go into detail on this subject and concentrate on the general structure of black spot analysis.
In almost all known procedures, road locations are treated as isolated spots.
One tries to detect the black spots by estimating the expected number of future accidents at a specific location from the number of accidents that already have occurred at that location. For many locations, especially in built-up areas, the number of observed accidents is too small to give an accurate estimation of the accident potential. This leaves us with a very inaccurate ordering of locations with regard to accident risk. We know that the black spots on the average are placed higher on the list, but we cannot distinguish them sufficiently from the grey, or even white spots.

If one still uses this detection method, then the next problem is to find the causes of the supposed danger. Little information is given in the

small accident numbers and one is almost completely dependent on an ad-hoc analysis of the location, based on rather general theories only. This approach, in which locations are investigated as isolated spots, does not seem promising to us, especially not if the accident numbers are small.
An alternative procedure starts from the comparison between the road locations. The central question is: "what do accident black spots have in common and in which respects do they differ from safe locations?"
If we cannot relate accident figures to characteristics of the locations then treatment of black spots from general theories does not seem possible at all. Therefore we think that the analysis of black spots should start with the investigation of the relations between the characteristics of road locations and accidents for a group of locations that can be compared with each other.
Multiple linear regression analysis and canonical correlation analysis are often used to detect such relations.
In this case, however, there are a number of problems to be solved before these techniques can be applied. Several characteristics (such as the kind of road surface etc.) do not seem metric and some of the metric characteristics do not need to be linearly related to the probability of an accident. It seems not unreasonable to expect e.g. a curvilinear relation between the probability of an accident and the width of a road. Furthermore, reflection on the combined effect of characteristics suggests to use multiplicative models instead of models that are additive in the independent variables the probability of an accident at a location with characteristic A and B will be equal to the product of the probabilities for A and for B if both are independent. Experimental evidence supports these multiplicative models (see: Rasch, 1973; Oppe, 1979).
However, new techniques are developed that account for all these problems. The solution of the problem is related to the canonical analysis of contingency tables approach as described e.g. by Kendall & Stuart, 1969 Vol II, pp 568 vv.
Recently Goodman (1981) compares this model with the log-linear models. The difference between both methods is that in the canonical analysis approach, one is interested in the scaling of variables in order to maximize the correlation or dependency, where as in log-lineair analysis

one rescales the variables under the assumption of independency. Interaction, or association as Goodman calls it, can be investiated within the loglinear model if one adds further restrictions on the residuals with regard to the row and column position of these residuals. Under special restrictions of this kind, both models result in identical solutions.
The fundamental idea behind the canonical-analysis approach is, that the computation of the correlation coëfficient between the "non-linear" row and column variable makes sense after the proper rescaling of these variables. The analysis results in that scaling of both variables that maximizes this correlation coëfficient.
If we generalize this procedure to multiway tables, then we arrive at some kind of non-linear principal-components analysis: variables are rescaled in such a way that they are as "homogeneous" as possible (which means that there mean intercorrelation is maximal).
A second generalization is found if we add new rows from different row variables to the table and eventually new columns from other new column variables. We then have some kind of super canonical-analysis problem, that reduces, after rescaling, to multiple linear regression (if there is only one column variable) or to the classical canonical correlation analysis (if there are more than one column variables).
These analysis techniques and the related computer programmes (Homals for the generalized homogeneity analysis and Canals for the generalized Canonical analysis) are developed at the Department of Data theory of the Leyden State University. A full description is given in Gifi (1981).

We will describe how we used these techniques for the description of the relations between accident figures and the characteristics of road locations.

2. Blackspot data

SWOV started an extensive research project in one of the Dutch provencies, called Noord-Brabant. This research was financed by the Ministry of Transport and the Noord-Brabant Provincial Council. One of the investigations within the project was concerned with a description of the relations between many accident-, road- and traffic characteristics of almost all public roads outside built-up areas in that province. Data collection is done by the Provincial Public Works Department and the

regional department of "Rijkswaterstaat". The engineering office D.H.V. took care of all data handling necessary before starting the analyses of this data.
The roads were classified in single-lane and dual-lane roads and each class consisted of three sub-categories. Each road was divided in parts of 100 meters. Intersections were deleted in the first analysis. New studies, concerning the intersections and larger units (routes) take place at the moment. A full report of this study is found in SWOV (1981).
We use only some of the results, in order to demonstrate the usefulness of the relational techniques for black-spot analysis.
In Table 1 one will find the marginals, with regard to the total number of injury accidents for each group of road locations. Black-spot detection and analysis based on the accident figures of these locations as such do not seem practical at all.

We see that motorways have on the average the lowest number of accidents. The highest mean number of accidents (M) is found with dual-lane roads closed for slow traffic. If we correct for traffic flow then the single-lane roads will most likely turn out to be more dangerous.
As to the variance (V) we see that this measure exceeds the mean, except for the dual-lane roads closed for slow traffic. The z-values, standard normal values derived from the Poisson index of dispersion which is defined as

$$X^2 = \sum_{i=1}^{T} (X_i - M)^2/M$$

are significant, except for the one road category mentioned. This suggests that all other sets of roads are heterogeneous and an investigation with regard to differences in accident potential does make sense.
In a mixed Poisson distribution, an estimate of the variance in Poisson parameters is given by the difference between the variance and the mean of X (see last column in Table 1).
If we delete the locations without accidents and fit a truncated Poisson distribution to this data, then we find that not only the number of locations without accidents, but also the number of locations with 1, 4 and 5 accidents are systematicly underestimated, while the numbers for 2

and 3 accidents are overestimated. Therefore it is [illegible] the difference between locations with and without accidents that accounts for the variance in the Poisson parameters. Estimates with corresponding χ^2-values and df (ignoring the zero class) are given in Table 2. The estimates for the negative binomial distribution and the corresponding χ^2-values are also given in Table 2. Here the zeroclass has been included. These values show a reasonable fit. The χ^2-value is significant only for the category of roads with mixed traffic. This suggests that the distribution of X is indeed a mixed Poisson distribution of a type as found by Greenwood & Yule (1920). The distribution of Poisson parameters for the category of roads with mixed traffic is perhaps more complicated then that of the other road types.

3. Application of relational techniques for the analysis of road sections with mixed traffic

The major aim of the analyses that we have done first was to find relations between 26 road and traffic characteristics of the 3833 single-lane roads with mixed traffic and their observed number of accidents.
A list of these characteristics is found in the legenda of Figure 2. As can be seen from table 1, most of the locations do not have injury accidents within the 5-year period. We have accomplished a second analysis using only the 685 accident locations. Both Canals analyses are in fact "non-linear" multiple-regression analyses, because there was only one dependent variable: the total number of injury accidents. From Figure 1 we can see that both in the first and in the second analysis the scaling of the dependent variable is logarithmic. This is in agreement with the assumption of a multiplicative (log-linear) model: the model is linear in the independent variables with regard to the log-value of the accident numbers. Also the conclusion, drawn from the fit of the truncated Poisson distribution, that the difference in Poisson parameters is more complicated than between locations with and without accidents, is confirmed with this scaling. If there had been a clear distinction, then we should have found a dichotomous scale. The scale found here suggests a more continuous distribution of accident probabilities.
Here we will not discuss the solutions with respect to the independent variables. The main difference in both solutions was due to the influence

	Number of accidents (X)										
dual lane roads:	0	1	2	3	4	5^+	T	M	V	Z	V-M
1. motorways	873	139	21	8	1	-	1042	.201	.258	6.17	.057
2. other road for motor vehicles	207	49	15	5	8	-	284	.444	.776	8.15	.332
3. roads closed for slow traffic	68	40	12	1	-	-	121	.554	.495	-.75	-
single lane roads:											
1. roads for motor vehicles	345	50	23	4	9	-	431	.334	.636	11.18	.302
2. roads closed for slow traffic	1867	329	113	41	22	8	2380	.339	.600	22.89	.260
3. roads with mixed traffic	3148	424	167	56	27	11	3833	.284	.520	30.93	.236

Table 1. Accident figures for almost all non-reconstructed 100 meter sections (intersections excluded) of primary and secundary roads outside built-up areas in Noord-Brabant. Data are collected over the five year-period from 1971 through 1975.

	Truncated Poisson							Negative Binomial						
	0	1	2	3	4	5^+	χ^2 /df	0	1	2	3	4	5^+	χ^2 /df
D1	304.9	134.5	29.6	4.4	.5	-	6.08/2	873.5	136.1	25.8	5.2	1.4	-	1.83/2
D2	39.6	42.8	23.1	8.3	2.2	.5	15.45/3	204.9	50.7	17.5	6.6	2.6	1.7	4.01/3
D3	84.0	41.1	10.0	1.6	.2	-	.78/2	67.4	41.7	10.5	1.3	.1	-	.40/2
S1	40.6	46.2	26.3	10.0	2.8	.6	13.55/3	339.7	59.7	19.4	7.3	2.9	2.0	7.25/3
S2	306.2	301.3	148.3	48.6	12.0	2.4	30.13/4	1851.8	353.8	110.9	39.3	14.7	9.5	5.84/4
S3	391.2	395.9	200.3	67.6	17.1	3.5	26.86/4	3116.9	483.7	147.3	52.2	19.8	13.1	13.54/4

Table 2. Estimated values and the corresponding χ^2-values and df's for the truncated Poisson distribution and the negative binomial distribution for the data of table 1.

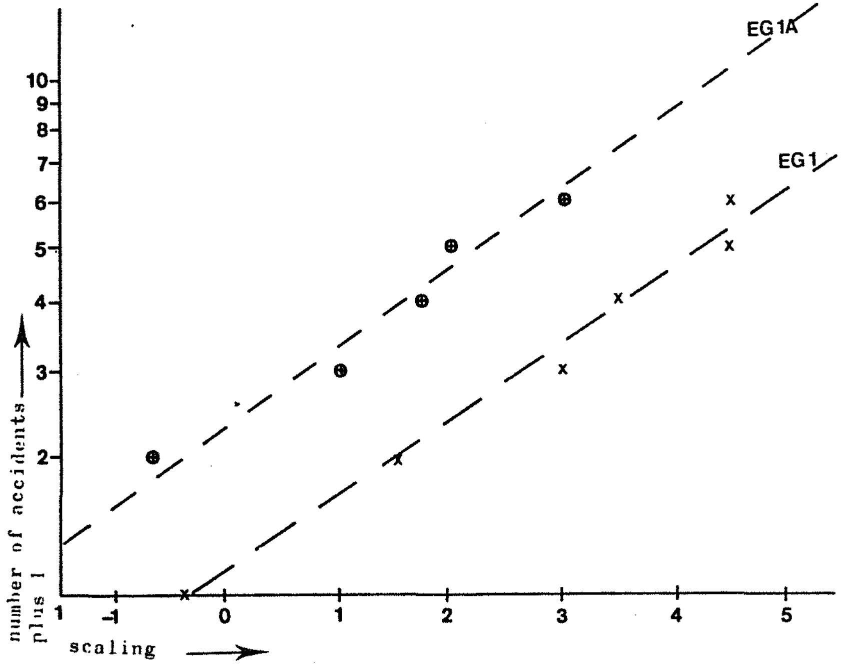


Fig 1. scale value versus number of accidents (on a log-scale) , For the total set of locations with mixed traffic (EG 1) and for the set of accident locations with mixed traffic (EG 1A).

of traffic volume on accident numbers. Traffic volume is an important variable in the analysis of all locations but not in the analysis of accident locations only.

Succeeding analyses were concerned with more than one dependent variable. In these analyses various types of accidents were investigated together with the total number of accidents. The total number of accidents was included in each analysis in order to find an explanation for the specific accident types in addition to the explanation of the total number of accidents. For these analyses we used only the 685 locations with accidents. Some analyses had more than two dependent variables. We will describe one of these non-linear canonical analyses here in order to explain the black spot method. We choose the analysis with the total number of accidents and the number of fatal accidents as dependent variables. This analysis has been done in order to investigate to what

extend the explanation of the most severe accidents differs from that of less severe accidents.

The first canonical axis corresponds almost completely to the total number of accidents. The canonical correlation after rescaling is $r_{c1} = .41$.

The second canonical axis corresponds primarily to the number of fatal accidents. The canonical correlation for this axis is $r_{c2} = .27$. In order to visualize relations between variables, we may represent variables graphicly by vectors in a space spanned by the locations (a space with 685 dimensions). The correlation between two variables is then represented by the cosine of the angle between the corresponding vectors. A correlation of 1 means a cosine of 1 and an angle of 0 degrees. A correlation of 0 means an angle of 90 degrees.

In figure 2 the projection of the independent variables on the plane through the dependent variables (in the space spanned by the locations) is given for the scaled variables.

Figure 3 shows us the scaling of the dependent variables and the most important independent variables for the explanation i.e. the variables with the largest projections. If we look at the canonical correlations, then at first glance these values seem to be low. Especially for a situation where 26 independent variables are used which are rescaled such that the canonical correlation is maximal. We did a bootstrap analysis to investigate the stability of the solution. This bootstrap analysis was done by taking samples (with replacement) from the 685 locations. In order to make comparisons with the original analysis, each sample existed again of 685 locations. We concluded that the results were more stable than expected. A plot of the mean bootstrap-analysis is given in figure 4. From this bootstrap study we estimated the canonical correlations for the population to be $r_{c1} = .35$ and $r_{c2} = .20$ for the first and second dimension.

Reflection on these figures learned that the the correlations may be that low primarily due to the low accident figures for each location and not because of the non-existence of relations between the accident probabilities and the characteristics of the locations. We cannot predict such small accident figures for locations accurately even if we know the real accident probabilities. This was in fact our initial problem. In

order to investigate to what extend this effect might influence our results, we did a Monte-Carlo study as follows.
The canonical scores for the locations that resulted from the analysis may be regarded as proportional to the logarithm of the probability of an accident at that location, because of the fact that the first canonical axis almost completely coincides with rescaled number of accidents. Therefore we transformed these values into real accident probabilities for the 685 locations and regarded these values as real population values. We then used these values as multinomial probabilities in an experiment in which we distributed 404 accidents over the 685 locations, according to the multinomial probabilities. We have chosen this number of accidents, because there are 1089 - 685 = 404 accidents that are freely distributed over the total set of accident locations.
Then we computed the correlation between the accident probabilities and the number of allocated accidents that resulted from the multinomial experiment. The mean correlation for 100 of these Monte-Carlo runs was r = .45. Using samples of 10 times as much accidents (4040 accidents), we found r = .84, this to give an indication of the increase of r with sample size. From the Monte-Carlo study we conclude that the maximum value to be expected for the canonical correlation of the first dimension is .45. The estimated population-value of r_{c1} = .35, resulting from the bootstrap study, seems rather high if we compare this value with the maximum of .45.
Therefore our conclusion is that, because we used the information of all locations together in our canonical analysis, we were able to predict the accident probability for each location a lot better from the road and traffic characteristics of the locations than it should have been possible using their individual accident number only.
Furthermore, this analysis gives us the relation between the danger and the road and traffic characteristics. This information can be used in order to take countermeasures.

4. Black-spot investigation based on non-metric canonical analysis

In the previous paragraphs, we found that the accident probabilites of locations differ especially for roads with mixed traffic. Furthermore we found that relational techniques for categorical data seem to be useful

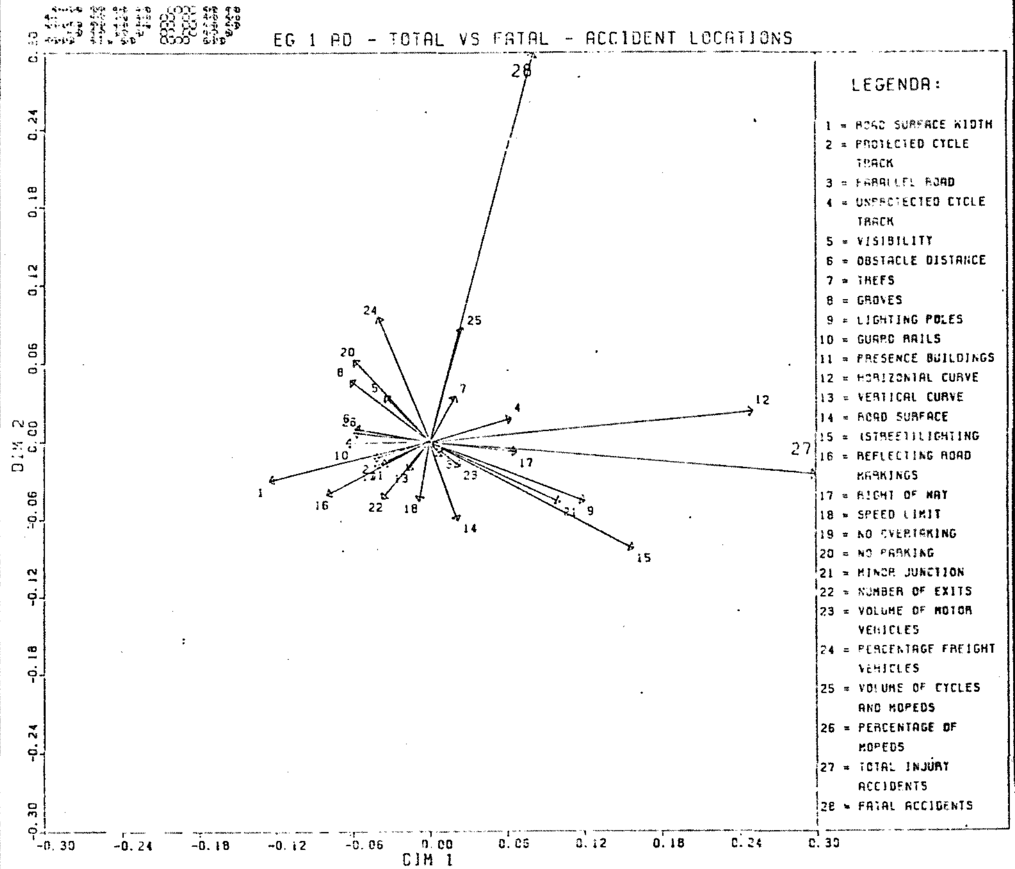


Figure 2

techniques to predict accident potential. We will give here a more explicite description of how these techniques can be used for black spot analysis.
In order to accomplish an analysis as described, we have to collect the relevant data for the investigation. The object of investigation may be

an intersection, a road section of some specific length as we used in the example, a pedestrian crossing in a small residential area, although the detailed comparison of complete areas will become increasingly difficult. For each object we have to measure the criterion value(s), e.g. the total number of injury accidents, accidents at daytime and nighttime, accidents with pedestrians involved etc.

Furthermore we must select the relevant characteristics of the units with regard to the explanation of our criterion. For black spot analysis, this will be primarily variables that are related to road characteristics or road conditions and controlling variables such as traffic volume, percentage of freight vehicles etc. This results in a "data matrix" consisting of n rows, corresponding to the n objects and m columns corresponding to the m characteristics. After the Canals analysis we get a new data matrix of rescaled variables. This rescaling is part of the solution that describes the relation between the criterion and the road and traffic characteristics. In addition, the solution results in an ordering of the characteristics with regard to the contribution of the independent variables to the explanation of safety. Finally we get an ordering of the locations with regard to unsafety.

In the example that has been described, we find a rescaling for each characteristic, and an ordering of objects and variables for each dimension. Table 3 shows this ordering for the first dimension. Only the five most important explaining variables are represented for the 25 most dangerous and 25 least dangerous locations. Table 4 shows us the same data for the second dimension. Figure 3 gave us the scaling of the five major independent variables for each dimension and the scaling of the dependent variables.

If we look e.g. at location 3 and 4 of Table 3, we see that these are two adjacent locations that are curved and have two and three minor crossings respectively. Furthermore the road is rather small at these points (< 6 m) and has orientation lighting. One location has one accident, the other has three accidents, including one fatal accident. Figure 5 gives us an idea of these locations.

A plot of the most dangerous locations on a map may suggest structural countermeasures. An analysis of and comparison with the least dangerous spots may also suggest countermeasures.

From Table 4 we see that if we want to concentrate on fatal accidents, countermeasures with regard to a high percentage of freight vehicles,

Table 3

location code	canonical score for set of road and traffic variables	canonical score for set of accident variables	total number of accidents + 1	number of lethal accidents + 1	horizontal curve	lighting	road surface width	lamppoles	number of minor junctions
10210716	3.699	1.735	4	1	-0.547	1.062	-0.806	11.180	3.486
10210405	3.197	0.845	3	1	1.929	1.062	0.146	5.336	-0.280
10208437	3.052	-0.756	2	1	1.929	1.062	-0.806	0.517	3.486
10208436	2.945	1.810	4	2	1.929	1.062	-0.806	0.517	4.203
10205126	2.859	0.920	3	2	1.929	1.062	0.146	-0.223	-0.280
10205117	2.720	0.845	3	1	1.929	1.062	-0.806	-0.223	-0.280
10201623	2.590	-0.756	2	1	1.929	1.062	0.146	5.336	3.486
10208404	2.572	1.735	4	1	1.929	1.062	-0.806	0.517	-0.280
10210418	2.493	1.810	5	2	1.929	1.062	-0.806	0.517	-0.280
10205672	2.420	0.845	3	1	1.929	1.062	-0.806	-0.223	-0.280
10210854	2.414	1.735	4	1	1.929	1.062	1.563	0.517	-0.280
10200417	2.296	1.735	4	1	-0.547	1.062	0.146	11.180	-0.280
10210415	2.250	0.845	3	1	1.929	1.062	-0.806	0.517	-0.280
10204427	2.236	0.845	3	1	1.929	1.062	-0.806	-0.223	-0.280
10205664	2.188	-0.756	2	1	1.929	1.062	-0.806	-0.223	-0.280
10204420	2.185	0.920	3	2	1.929	1.062	-0.806	-0.223	3.486
10210403	2.181	0.845	3	1	1.929	1.062	-0.806	0.517	-0.280
10201868	2.174	0.845	3	1	1.929	-0.942	0.146	-0.223	-0.280
10202556	2.139	0.845	3	1	1.929	1.062	-0.806	0.517	-0.280
10201019	2.090	0.920	3	2	-0.547	1.062	-0.806	5.336	-0.280
10200419	2.057	0.845	3	1	1.929	1.062	1.563	5.336	-0.280
10206377	2.041	2.922	6	3	1.929	-0.942	1.563	0.517	-0.280
10210428	2.032	0.920	3	2	1.929	1.062	-0.806	0.517	-0.280
10202010	2.011	-0.756	2	1	1.929	1.062	-0.806	-0.223	-0.280
10211227	2.008	1.735	4	1	1.929	1.062	-0.806	0.517	-0.280

25 locations with highest canonical scores on first dimension

10202430	-1.689	-0.756	2	1	1.929	1.062	1.563	-0.223	-0.280
10202418	-1.691	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10206117	-1.770	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10202428	-1.791	-0.756	2	1	1.005	-0.942	1.563	-0.223	-0.280
10210388	-1.804	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10210692	-1.816	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10130352	-1.827	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10210335	-1.858	-0.681	2	2	-0.547	1.062	1.563	0.517	-0.280
10209612	-1.876	-0.756	2	1	-0.547	1.062	1.563	0.517	-0.280
10208991	-1.927	-0.756	2	1	-0.547	1.062	-0.806	-0.223	-0.280
10209650	-1.955	0.845	3	1	-0.547	-0.942	0.146	-0.223	-0.280
10204492	-1.957	-0.756	2	1	-0.547	-0.942	-0.806	-0.223	-0.280
10202411	-2.116	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10202395	-2.124	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10211478	-2.223	-0.756	2	1	-0.547	1.062	1.563	0.517	-0.280
10130379	-2.227	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10208990	-2.230	-0.756	2	1	-0.547	1.062	-0.806	-0.223	-0.280
10208963	-2.230	-0.756	2	1	-0.547	1.062	-0.806	-0.223	-0.280
10208966	-2.230	-0.756	2	1	-0.547	1.062	-0.806	-0.223	-0.280
10209610	-2.264	-0.756	2	1	-0.547	1.062	1.563	-0.223	-0.280
10206253	-2.299	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10202406	-2.334	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10209641	-2.437	-0.756	2	1	-0.547	-0.942	0.146	0.517	-0.280
10209646	-2.606	-0.756	2	1	-0.547	-0.942	0.146	-0.223	-0.280
10202126	-2.667	-0.756	2	1	-0.547	1.062	1.563	-0.223	-0.280

25 locations with lowest canonical scores on first dimension

order of the locations with regard to the predicted accident potential
(first canonical dimension) together with information about the most relevant characteristics

81

Table 4

location code					percentage of freight-vehicles	moped and bicycle volume	lighting	par king prohibited	road surface
10204992	4.957	-0.014	2	1	2.693	0.816	-0.942	2.699	-0.056
10204994	4.957	-0.443	3	1	2.693	0.816	-0.942	2.699	-0.056
10204987	4.094	9.066	4	3	2.693	0.816	1.062	2.699	-0.056
10204985	4.063	9.304	3	3	2.693	0.816	1.062	2.699	-0.056
10204984	3.794	-0.014	2	1	2.693	0.816	1.062	2.699	-0.056
10208951	3.488	-0.014	2	1	2.693	0.816	1.062	-0.346	2.183
10204792	3.328	-0.014	2	1	2.693	0.816	-0.942	-0.346	-0.056
10201603	3.286	0.867	2	2	0.128	0.816	-0.942	-0.346	-0.056
10204788	3.195	0.438	3	2	2.693	0.816	-0.942	-0.346	-0.056
10204790	3.128	-0.443	3	1	2.693	0.816	1.062	-0.346	-0.056
10206395	2.961	-0.443	3	1	2.693	0.816	-0.942	-0.346	-2.055
10210925	2.911	-0.778	6	1	2.693	-1.134	-0.942	2.699	-0.056
10210930	2.844	-0.014	2	1	2.693	-1.134	-0.942	2.699	-0.056
10201606	2.733	9.304	3	3	0.128	0.816	-0.942	-0.346	-0.056
10210921	2.534	-0.014	2	1	2.693	-1.134	-0.942	2.699	-0.056
10210920	2.534	-0.014	2	1	2.693	-1.134	-0.942	2.699	-0.056
10211647	2.441	-0.443	3	1	0.128	1.314	1.062	2.699	-0.056
10206402	2.394	-0.014	2	1	2.693	0.816	1.062	-0.346	-2.055
10204942	2.332	0.867	2	2	0.128	0.816	-0.942	2.699	-0.056
10206387	2.256	-0.443	3	1	0.128	0.816	-0.942	-0.346	-2.055
10204934	2.245	0.867	2	2	0.128	0.816	-0.942	2.699	-0.056
10208004	2.199	-0.014	2	1	0.128	-1.134	-0.942	-0.346	-0.056
10204492	2.196	-0.014	2	1	0.128	-1.134	-0.942	-0.346	-0.056
10206380	2.157	-0.014	2	1	0.128	0.816	-0.942	-0.346	-2.055
10206385	2.157	-0.014	2	1	0.128	0.816	-0.942	-0.346	-2.055

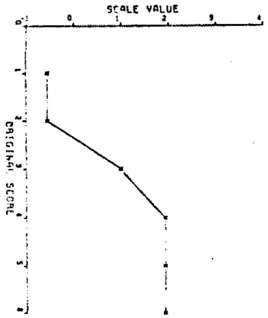
25 locations with highest canonical scores on second dimension

10202097	-1.623	-0.014	2	1	-0.416	-1.134	-0.942	-0.346	-0.056
10202556	-1.690	-0.443	3	1	-0.416	-1.134	1.062	-0.346	-0.055
10201651	-1.691	-0.014	2	1	0.128	-1.134	1.062	-0.346	-0.056
10208399	-1.692	-0.682	4	1	0.128	-1.134	1.062	-0.346	-0.056
10206174	-1.694	-0.682	4	1	-0.416	-1.134	1.062	-0.346	-0.056
10200417	-1.715	-0.682	4	1	0.128	0.816	1.062	-0.346	-0.055
10203438	-1.731	-0.443	3	1	0.128	0.816	1.062	-0.346	2.183
10202416	-1.760	-0.014	2	1	-0.416	-1.134	1.062	-0.346	2.183
10202363	-1.787	-0.443	3	1	-0.416	-1.134	1.062	-0.346	2.183
10201626	-1.805	-0.443	3	1	0.128	-1.134	1.062	-0.346	-0.056
10206167	-1.806	-0.443	3	1	-0.416	-1.134	-0.942	-0.346	-0.056
10206172	-1.806	-0.014	2	1	-0.416	-1.134	-0.942	-0.346	-0.056
10202096	-1.813	-0.443	3	1	-0.416	-1.134	-0.942	-0.346	-0.056
10211055	-1.843	-0.014	2	1	-0.416	-1.134	1.062	2.699	2.183
10206081	-1.868	-0.014	2	1	0.128	1.314	1.062	-0.346	-0.056
10203272	-1.887	-0.443	3	1	0.128	0.816	-0.942	-0.346	-2.055
10201583	-1.921	-0.014	2	1	0.128	-1.134	1.062	-0.346	-0.056
10202391	-1.924	-0.014	2	1	-0.416	-1.134	1.062	-0.346	2.183
10205111	-1.930	-0.014	2	1	-5.556	-2.129	1.062	-0.346	2.183
10202401	-2.009	-0.014	2	1	-0.416	-1.134	1.062	-0.346	2.183
10203441	-2.210	-0.014	2	1	0.128	0.816	1.062	-0.346	2.183
10206554	-2.269	-0.014	2	1	0.128	-1.134	1.062	-0.346	-0.056
10205106	-2.625	-0.443	3	1	-5.556	-2.129	1.062	-0.346	2.183
10201623	-2.845	-0.014	2	1	0.128	-1.134	1.062	-0.345	-0.056
10200221	-3.443	-0.014	2	1	-0.416	0.816	1.062	-0.346	-0.056

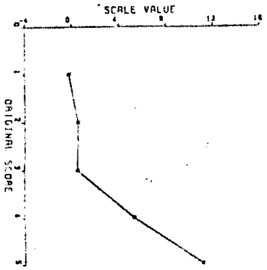
25 locations with lowest canonical scores on second dimension

order of the locations with regard to the predicted number of lethal accidents (second canonical dimension) with information about the most relevant characteristics

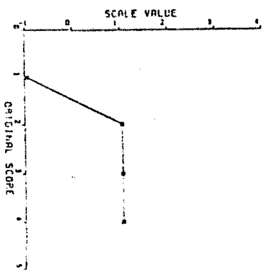
Figure 3a



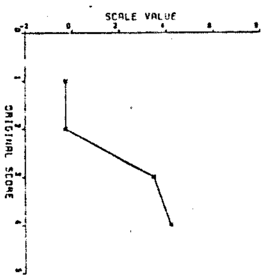
12. horizontal curve
cat. 1 (-0.547): no (C ≥ 1000 m)
cat. 2 (-0.547): 700 ≤ C < 1000 m
cat. 3 (1.005): 500 ≤ C < 700 m
cat. 4 (1.929): 400 ≤ C < 500 m
cat. 5 (1.929): 200 ≤ C < 400 m
cat. 6 (1.929): C < 200 m



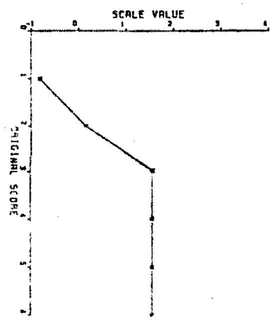
9. lampoles
number of section parts of lamppoles
cat. 1 (-0.223): 0
cat. 2 (0.517): 1
cat. 3 (0.517): 2
cat. 4 (5.336): 3
cat. 5 (11.180): 4



15. lighting
cat. 1 (-0.942): not present
cat. 2 (1.602): partially
cat. 3 (1.062): orientation lighting
cat. 4 (1.062): full lighting



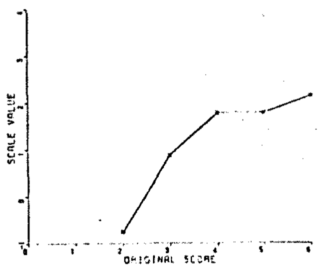
21. number of minor junctions
cat. 1 (-0.280): 0
cat. 2 (-0.280): 1
cat. 3 (3.486): 2
cat. 4 (4.203): 3



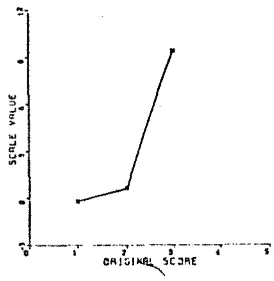
1. road surface width
cat. 1 (-0.806): W < 6.0 m
cat. 2 (0.146): 6.0 < W ≤ 6.5 m
cat. 3 (1.563): 6.5 < W ≤ 7.0 m
cat. 4 (1.563): 7.0 < W ≤ 7.5m
cat. 5 (1.563): 7.5 < W ≤ 8.0 m
cat. 6 (1.563): W > 8.0 m

Figure 3c

- 16 -



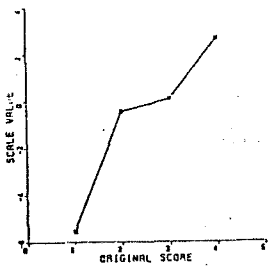
27. total number of accidents
cat. x: number of accidents plus 1



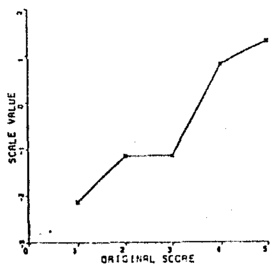
28. lethal accidents
cat. x: number of lethal accidents plus 1

Figure 3b

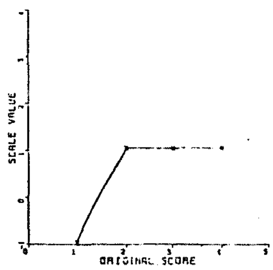
- 15 -



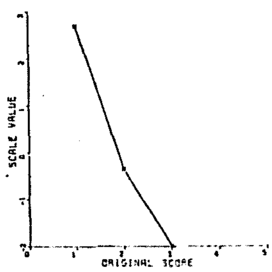
24. percentage of freight vehicles
cat. 1 (-5.556): null
cat. 2 (-0.416): 0 < P ≤ 10
cat. 3 (0.127): 10 < P ≤ 20
cat. 4 (2.693): P > 20



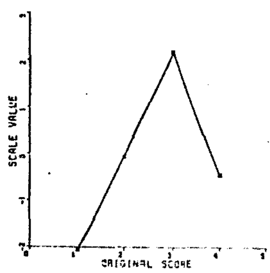
25. volume of cycles and mopeds
cat. 1 (-2.129): V = 0
cat. 2 (-1.134): 0 < V ≤ 117
cat. 3 (-1.134): 118 ≤ V ≤ 351
cat. 4 (0.816): 352 ≤ V ≤ 1053
cat. 5 (1.314): > 1054



15. lighting
cat. 1 (-0.942): not present
cat. 2 (1.062): partially
cat. 3 (1.062): oriëntation lighting
cat. 4 (1.062): full lighting

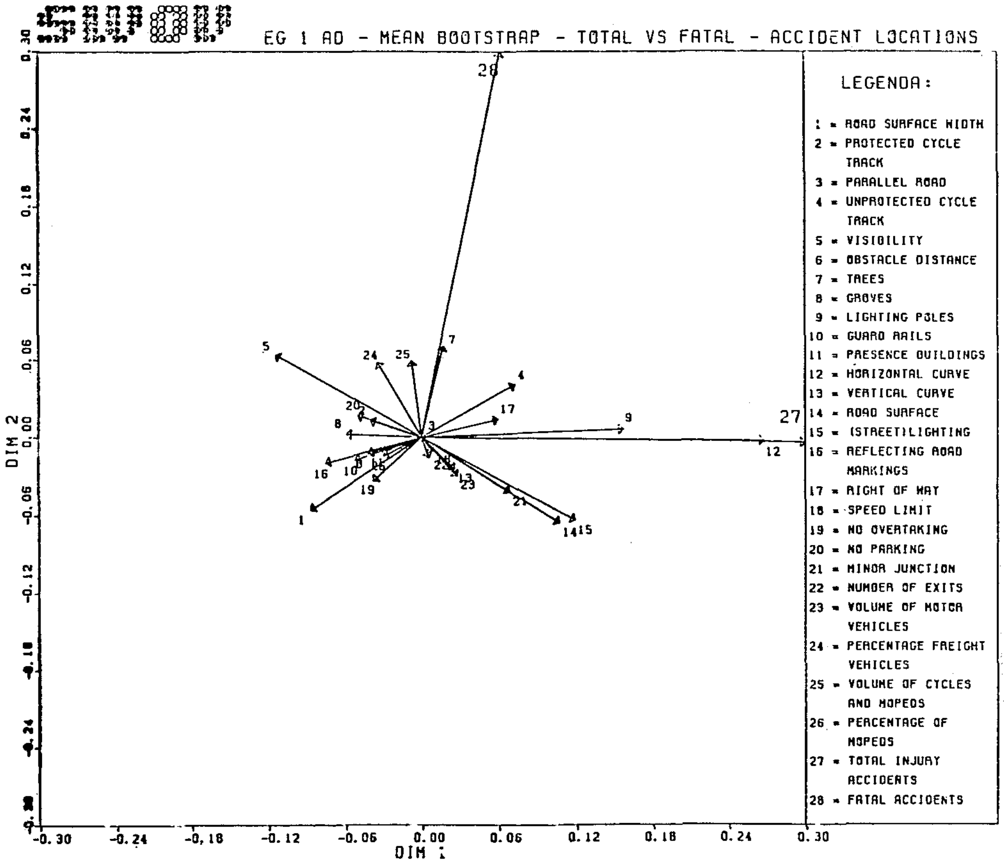


20. no parking
cat. 1 (2.699): parking allowed
cat. 2 (-0.346): no parking because of road type
cat. 3 (-1.984): no parking because of shield



14. road surface
cat. 1 (-2.055): concrete
cat. 2 (-0.056): asphalt
cat. 3 (2.183): pavingbricks
cat. 4 (-0.453): setts

Figure 4



together with high bicycle volumes seem to be urgent. The first five locations are almost adjacent. Two of the locations have two fatal accidents. Figure 6 gives us an idea of these locations. Structural measures instead of measures on the locations itself seem to be indicated here.

Fig. 5. Hazardous location of first dimension

This is just an example to show that this technique works and how it works.

Finally we will mention the advantage of this procedure for the evaluation of safety measures. A general problem in the evaluation of safety measures is the effect of the "regression-to-the-mean". This effect is due to the fact that if we divide the locations into two groups, one with high numbers of accidents in the past and the other with low accident numbers, then there will be a tendency for the mean accident number of the first group to decrease in time and for the mean accident number of the second group to increase, even if we do not change any location. This results from the fact that several locations in the first group have high accident numbers and several locations in the second group low accident numbers by chance. These effects can be very substantial and suggest accident reductions that are far too optimistic.

We may want to solve the problem by incorporating the accidents of all locations (including the locations that have not been treated) in the evaluation study or even estimate the regression-to-the-mean effect using

Fig. 6. Hazardous location of second dimension

the non-treated locations only. Here we do not have to deal with this problem, because we can estimate the expected number of accidents for a given location if there will be no treatment. Furthermore we can compute the accident reduction as a result of the countermeasures that have been taken, without referring directly to the number of accidents that occurred in the past on that location.

LITERATURE

1. Gifi, A. Non-linear multivariate analysis, Leiden, 1981.

2. Goodman, L.A. Association Models and Canonical Correlation in the analysis of Cross-Classifications having ordered categories, J.A.S.A., 76, 1981.

3. Greenwood, M. & Yule, G.U. An inquiry into the nature of frequency distributions ..., J.R. Statist. Soc., 83, 1920.

4. Kendall, M.G. & Stuart, A. The advanced theory of statistics, Vol II, London, 1967.

5. Oppe, S. The use of multiplicative models for the analysis of road safety data, A.A.P., 11. 1979.

6. Rasch, G. Two applications of the multiplicative Poisson models in road accidents statistics, In: Proc. of the 38th session of the ISI, Wien, 1973.

7. SWOV. De verkeersonveiligheid in de provincie Noord-Brabant IXA, IXB, IXC, SWOV, Voorburg, 1981.