INVESTIGATING FATAL ROAD ACCIDENT DATA

Andri van Niekerk

Thesis presented in partial fulfillment of the requirements for the degree of Master of Civil Engineering at the University of Stellenbosch.

Prof. CJ Bester **March 2007**

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature: .....
Andri van Niekerk

Date: 28/02/2007

EXECUTIVE SUMMARY

This thesis concerns the investigation of four analyses techniques in terms of their utility and adequacy for analyzing fatal road accident data in South Africa.

PROBLEM DEFINITION

Road accident data are summarized annually in various forms, but the relationships between the different categorical variables are not determined. The study aimed to address this problem.

Road accident rates are published in order to compare year-to-year change in an accident rate. It was necessary to investigate a method to determine whether these year-to-year changes are statistically significant and whether there should necessarily be a reason for concern when an increase in accident rate is detected.

Multiple regression models also including qualitative variables were investigated in this study.

ACCIDENT DATA AND ANALYSIS TECHNIQUES

Road accident data were found available in the format of a MS Access database which could be manually investigated. Traffic and speed data were readily available from Mikros Traffic Monitoring (Pty) Ltd in the form of SANRAL's CTO Yearbooks and was found to be reliable and sufficiently detailed. Any road geometric data were omitted from the study due to insufficient detail available.

All data were found to show levels of poor data quality. Certain variables were thus omitted from the study e.g. the age group variable.

The fatal road accident database was analysed using Correspondence Analysis and Association Rules (for analyses of the categorical variables) and, the application of the Poisson distribution for chance variation analyses and Multiple Regression Analyses (for the continuous variables).

METHODOLOGY

Fatal road accident data were gathered by performing queries in the fatal road accident database. Traffic and Speed data were gathered by manually investigating the SANRAL CTO Yearbooks and manipulating the data to be integrated with the fatal road accident database. After all data manipulation was completed, the four analyses techniques mentioned above were applied using the software package *Statistica*.

FINDINGS

Correspondence Analysis and Association Rules were found to be adequate for analysing categorical road accident data variables with some data quality limitations and insufficient data sampling. The time period used for chance variation analysis was too short to deliver significant results. Three multiple regression models were created with one of the models being able to predict the number of fatalities per fatal accident with R^2 equal to approximately 40%.

CONCLUSIONS AND RECOMMENDATIONS

The following conclusions are drawn and recommendations are made based on the findings of this study:

- Detailed and quality road accident data for South Africa is unavailable. Better quality data are urgently needed for the purpose of analysis.
- Correspondence Analysis is found to be the most appropriate technique for road accident data analysis and should be applied on an annual basis.
- Association Rules Analysis results are influenced by small sample sizes and too many unknown variable categories. Larger sample sizes and exclusion of the unknown categories might improve the results.
- The analysis period for chance variation is too short and a longer period will provide more significant results.
- The multiple regression model predicting the number of fatalities per fatal accident is accepted in terms of utility and adequacy.

SAMEVATTING

Hierdie tesis bespreek die toepassing van vier verskillende analise tegnieke in terme van elkeen se geskiktheid om noodlottige padongelukke in Suid-Afrika te ondersoek.

PROBLEEM DEFINISIE

Padongeluk data word jaarliks opgesom en publiseer in verskillende vorme, maar die verwantskappe tussen kategoriese veranderlikes word nie direk bepaal nie. Die studie het probeer om hierdie probleem aan te spreek.

Padongeluk koerse word gepubliseer om verandering in ongeluk syfers waar te neem van jaar tot jaar. Dit was nodig om 'n metode te ondersoek om te bepaal wanneer enige verandering in ongelukskoerse statisties betekenisvol is en of daar noodwendig rede vir kommer behoort te wees indien 'n toename in ongelukskoers waargeneem is.

Veelvoudige regressie modelle wat ook kwalitatiewe veranderlikes insluit is ondersoek in hierdie studie.

ONGELUKSDATA EN ANALISE TEGNIEKE

Padongeluksdata was beskikbaar in 'n MS Access dokument wat met die hand ondersoek kon word. Verkeers- en Spoed data was beskikbaar van Mikros Traffic Monitoring (Edms.) Bpk. vanuit SANRAL se *CTO Yearbooks*. Die data was betroubaar en beskikbaar en in voldoende detail. Geometriese inligting van die betrokke padseksies is uitgesluit by die studie a.g.v. onvoldoende detail beskikbaar.

Alle data ingesamel het verskeie vlakke van lae data kwaliteit getoon. Sekere veranderlikes is daarom uitgesluit, bv. die ouderdomsveranderlike.

Die noodlottige padongeluk databasis is geanaliseer deur die gebruik van Ooreenkomsanalise en Assosiasie Reëls (vir die kategoriese veranderlikes) en die toepassing van die *Poisson* verspreiding vir ewekansige variasie en Veelvoudige Regressie Analise (vir die kontinue veranderlikes).

METODIEK

Noodlottige padongeluk data is ingesamel deur *queries* uit te voer in die MS Access databasis. Verkeers- en Spoed data is ingesamel deur die *CTO Yearbooks* van SANRAL met die hand te ondersoek en die data te integreer met die ongeluksdatabasis. Nadat alle relevante data met die ongeluksdatabasis geïntegreer is, is die vier bovermelde analise tegnieke uitgevoer m.b.v. die sagteware pakket *Statistica*.

BEVINDINGS

Ooreenkomsanalise en Assosiasie Reëls is die mees geskikte analise tegnieke vir kategoriese veranderlikes, alhoewel relatief lae data kwaliteit en onvoldoende steekproef trekking beperkings daar gestel het. Die analise periode wat gebruik is vir ewekansige variasie is te kort om statisties betekenisvolle resultate te lewer. Drie meervoudige regressie modelle is opgestel. Dit is bevind dat een van die modelle die aantal noodlottige gevalle per noodlottige padongeluk met 'n R^2-waarde van ongeveer 40% voorspel.

GEVOLGTREKKINGS EN AANBEVELINGS

Die volgende gevolgtrekkings en aanbevelings word gemaak volgens die bevindings van hierdie studie:

- Gedetailleerde en kwaliteit padongeluksdata vir Suid-Afrika is nie beskikbaar nie. Beter data kwaliteit word dringend benodig vir analise doeleindes.
- Ooreenkomsanalise is die mees geskikte analise tegniek vir padongeluk data analise en behoort jaarliks toegepas te word.
- Assosiasie Reëls resultate word grootliks beïnvloed deur klein steekproef groottes en te veel onbekende veranderlike kategorieë. Groter steekproewe en die uitsluit van onbekende kategorieë mag die resultate verbeter.
- Die analise tydperk vir ewekansige variasie analise is te kort en 'n langer tydperk sal meer betekenisvolle resultate lewer.
- Die meervoudige regressie model wat die aantal noodlottige gevalle per noodlottige ongeluk voorspel word aanvaar in terme van sy bruikbaarheid en geskiktheid.

ACKNOWLEDGEMENTS

Hereby, the author would like to express appreciation and thanks to the following individuals:

Mr and Mrs B van Niekerk

Mr and Mrs A Schrader

Mr N Kleynhans

Mr J van Niekerk

Prof. Christo Bester, Study leader, Department of Civil Engineering, University of Stellenbosch

Stellenbosch Traffic Department

Mr. Tjaart van Wijck

Mr. Neil Sadie

Mr. Carel Verwey

Dr. Martin Kidd, Department of Statistics, University of Stellenbosch

Ms. Cielie Karow, RTMC

TABLE OF CONTENTS

CHAPTER 4: FINDINGS AND DISCUSSION

LIST OF FIGURES

CHAPTER 2: ACCIDENT DATA AND ANALYSIS TECHNIQUES

CHAPTER 4: FINDINGS AND DISCUSSION

LIST OF TABLES

CHAPTER 2: ACCIDENT DATA AND ANALYSIS TECHNIQUES

CHAPTER 3: METHODOLOGY

CHAPTER 4: FINDINGS AND DISCUSSION

LIST OF APPENDICES
(Provided in Electronic Format on Enclosed Disc)

APPENDIX A

A1: Accident Types and Contributory Factors

A2: Example of CTO (Comprehensive Traffic Observation) Station Report

A3: Example of Road log Report

A4: Maps of Road sections along the N1, N2 and N7 in Western Cape Province and positions of CTO counting stations

APPENDIX B

B1: Correspondence Analysis Output

B2: Correspondence Analysis - Individual Interpretations for Each Variable Pair

APPENDIX C

C1: Association Rules Output - If(*Human Factor, Vehicle Factor, Road Factor*) Then(*Accident Type*)

C2: Association Rules Output – If(*Area Type, Vehicle Type*) Then(*Accident Type*)

C3: Association Rules Output – If(*Terrain Type, Vehicle Type*) Then(*Accident Type*)

C4: Association Rules Output – If(*Area Type, Terrain Type, Vehicle Type*) Then(*Accident Type*)

APPENDIX D

D1: Sample data for Multiple Regression Analyses

D2: Multiple Regression Analyses Output

CHAPTER 1

INTRODUCTION

1.1 Problem Definition

1.1.1 Relationships between Categorical Variables in Road Accident Data

Each year, road accident data is published and summarized in various forms (e.g. per province, per year, per day of the week etc.), including summaries of accident data for the current year as well as for previous years. Accident rates and raw accident frequencies are most commonly given, but the relationships between different road accident categorical variables are not necessarily given and it is sometimes too complex to interpret directly from cross tabulation tables, especially when a particular table is relatively large. This study aimed to address this problem.

1.1.2 Determination of Significance of Year-to-Year Accident Rate Change

Accident rates are calculated and also included in publications, showing the change in accident rates from year to year. An increase in accident rates are viewed as unfavourable, but it is not necessarily a significant increase worthy of concern. This study aimed to investigate a method to determine if and when changes in accident rates are indeed significant or not.

1.1.3 Accident Rate Prediction Models Including Only Continuous Variables

In some instances, accident rate prediction models are created with only continuous variables included as predictors. The study aimed to create prediction models to also include categorical variables (qualitative predictors).

1.2 Objective of Study

The objective of this study was to find a suitable road accident data source and analyse road accident data by applying four different analysis techniques in order to address the problems discussed in the section 1.1. These techniques are then discussed in terms of their adequacy, accuracy and utility.

1.3 Limitations and Scope of Investigation

Fatal road accident data are limited to data for the years 2002, 2003 and 2004. Although data quality obstacles were encountered, the data were nevertheless analysed. Additional data sources were used for additional data integrated with the accident database. This was time consuming and subject to careful data preparation. Data are limited to data from South Africa and where applicable, data from Western Cape Province was also analysed to compare the results with results obtained for the whole country.

1.4 Plan of Development

A chapter (Chapter 2) is provided containing general discussions on accident data with some accident data provided for South Africa for illustration. Data availability, data quality and data analysis are disussed as applicable to this study. Chapter 3 discusses the different methodologies used for data gathering and analysis and Chapter 4 comments on the study findings and results. In Chapter 5 conclusions are drawn based on the results in Chapter 4 and then finally Chapter 6 makes recommendations based on the conclusions drawn in Chapter 5.

CHAPTER 2

ACCIDENT DATA AND ANALYSIS TECHNIQUES

2.1 Data/Information Availability

2.1.1 Accident Data/Information Availability

In this section the availability of road accident data is discussed. Road accident statistics and accident rates are always readily available in various publications per province or for the whole country. It is most of the time not available for specific routes or specific points on roads where accidents occur, because usually the exact point or coordinates of where an accident occurred on a road is not available. This has proven to be mostly the case in South Africa.

What this section is especially referring to, is the details behind each accident and all (or most) of the factors involved in a particular accident. This kind of information is not so easy to keep record of, because of factors which include human factors such as under trained traffic officials or under trained accident investigators and other external factors (e.g. weather conditions).

The availability of road accident data and information has always been a problem in South Africa due to various reasons which can never exactly be pinpointed, but where accident data is available, it is either of poor quality (i.e. underreported) or outdated (data quality will be discussed in the sections to follow). Even though there are different factors (i.e. human errors and negligence etc.) affecting the availability and quality of data, it is the author's opinion, based on experience while doing this study, that there are too many different methods used for collecting data, organizing and reporting it. This becomes more problematic as the requirement for accident data becomes more detailed, because not only can the type and level of detail of available data vary per province, but it cannot always be pooled to be used collectively for the country as a whole, though collective statistics are available from different provincial administrations, if the correct procedures are followed to obtain it. Careful planning and time management is necessary to do this, because of the amount of time necessary to do the correspondence.

Arrive Alive and the RTMC (Road Traffic Management Corporation) has the task of researching and compiling accident statistics in South Africa. Arrive Alive's website states clearly that the importance of the data is not only for its statistical significance, but it is about how it affects the measures which need to be taken to reduce future accidents. It is important that these reports be available to the public, but it should be available to especially the road safety role players, such as the academics doing research in the field of road safety, or traffic engineers who work in the field.

The following quote from www.arrivealive.co.za states their vision:

"It is the vision of the arrive alive web site to be an effective information portal for Road Safety , and these Statistical reports should assist journalists and educators to create further awareness"

In South Africa the following procedure for collecting accident statistics is followed, according to Arrive Alive:

The National Fatal Accident Centre, National Department of Transport receives all the accidents that are recorded on a SAPS CAS system on a daily basis. This information contains the CAS number, date and time and the location of the accident divided in provinces. The following procedure then follows in the Centre:

- The SAPS report fatal accidents on an ongoing basis to the National Fatal Accident Information Centre.
- A printout will be made of all the reported accidents and will be verified against the accidents that were received from the SAPS head office.
- The Centre will then follow up accident information by phoning the specific police stations.
- If information cannot be obtained from certain stations, the area coordinators and provincial coordinators will be contacted telephonically and the unreported accidents will be faxed to them for further action from their side.
- At the end of each month all unreported accidents would be captured in the system. As the accident data is received from the SAPS the system will automatically take the accident off the unreported list.
- The system also picks up duplicates immediately and do not allow the capturer to input any duplicates.

- The follow up of unreported accidents is an ongoing process in the Centre.

Even though the above steps seem like a sound system for reporting and capturing accident data, there is too much negligence in the way the databases are managed. Data quality will be discussed in paragraphs to follow and no further attention will be given to data needs and issues in this section. The author is just emphasizing how with even the best (presumably) data gathering and managing methods, human error is still the main factor causing databases to be incomplete or outdated. Even if databases are up to date, it may not be complete and the data contained within it may be of such bad quality that it has no value (depending on and in terms of the purpose it is to be used for).

In the Road to Safety 2001-2005 Strategy the following is said on statistical output according to Arrive Alive (June 2006):

"Our statistical output problem, of course, is that the nearer to 100% crash coverage we get, the "worse" the crash and fatality rates become, as measured against previous years' (under-reported) statistics. In other words, as we get closer and closer to the goal of 100% reporting across the whole of South Africa, achieved reductions in crash and fatality rates will not reflect as positively as they should in year-on-year terms, because the base coverage is itself becoming more comprehensive each year."

It is the author's opinion that the problem of statistical output, as stated above, is not a problem which deserves major concern, due to its temporary nature. This problem will solve itself in time as statistical output will stabilise once the coverage of accidents has reached a consistent and constant level. 100% coverage will not, however, be achieved "over night" and in South Africa this matter needs immediate and constant attention.

For the purpose of this thesis a MS Access database was obtained from Arrive Alive containing a wide range of details on fatal road accidents which occurred during 2002, 2003 and 2004. Until recently the database was readily available in MS Access format and could easily be written on a compact disc and posted.

Due to certain technical difficulties with the MS Access database and reasons of which there is still uncertainty, the RTMC was considering the transfer of accident data to the NaTiS (National Traffic Information System). If this transfer was to happen the accident module of the system would not necessarily be available (or there would be a lower level of detailed information available regarding road accidents) and any further detailed data or information would have to be obtained through

Provincial Administration following the proper procedures. The plans for doing the transfer of data to NaTiS was not finalized at the time of print of this document and how accident data would be made available to the public and to the research fraternity was still being considered.

No other fatal road accident data source was available at the required level of detail. The database will be discussed in further detail in later sections when the various data gathering methods used in the study are discussed. Diagrams illustrating the relationships between different data tables within the database are provided below (fig 2.1.1(a) and 2.1.1(b)), to indicate the type of variables available within the database. These relationships will again be referred to in the following chapter in terms of the methodology for data gathering.

Table 2.1 provides the different accident types and vehicle types which feature in the database (refer to Appendix A1 for specific accident types and some contributory factors). Table 2.2 provides the different accident factors (human factors, vehicle factors and road factors) which feature in the database.

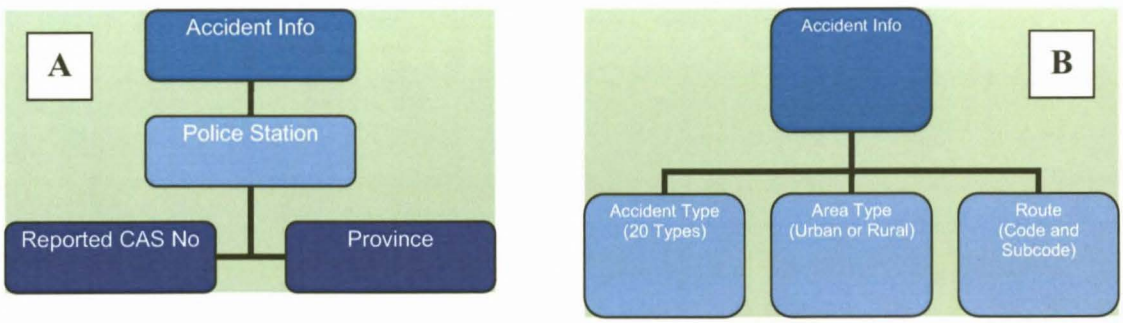

Fig. 2.1.1(a): Illustration of Data table Relationships within MS Access Database

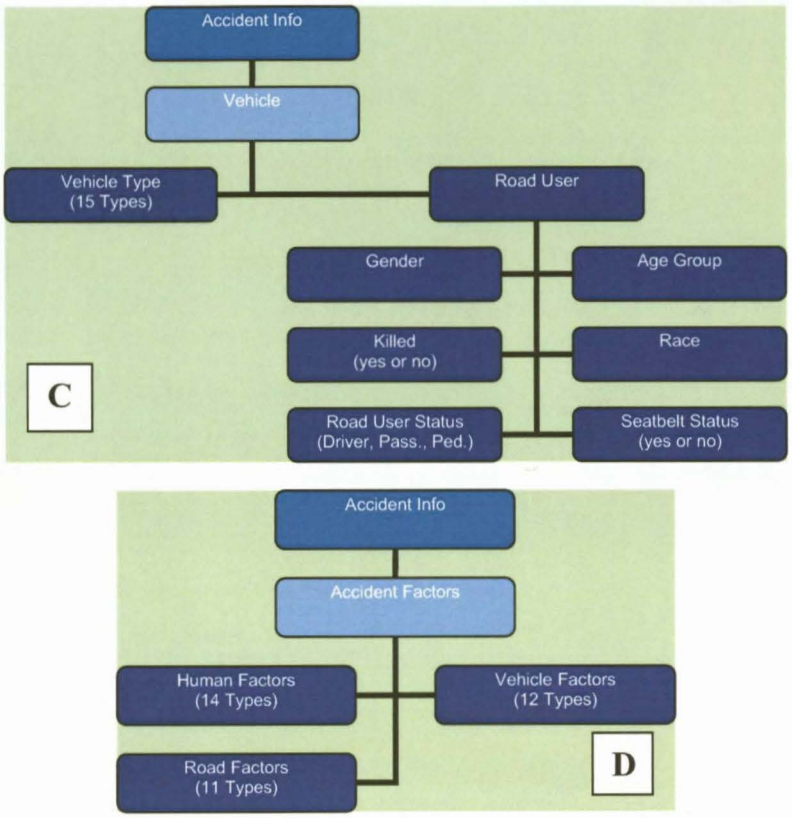

Fig. 2.1.1(b): Illustration of Data table Relationships within MS Access Database

Table 2.1: Table of Accident Types and Vehicle Types

ACCIDENT TYPES	VEHICLE TYPES
➢ Head-Rear end ➢ Sideswipe same direction ➢ Turn from wrong lane ➢ Head on ➢ Sideswipe opposite direction ➢ Turn in face of oncoming traffic ➢ Approach at angle ➢ Reversing ➢ Overturned ➢ Other ➢ Pedestrian ➢ Collision - Fixed object ➢ Hit and run ➢ Person fell off LDV or Truck ➢ Motorcycle ➢ Multiple vehicle ➢ Jack knife ➢ Cyclist ➢ Train ➢ Animal ➢ Unknown	➢ Sedan ➢ LDV / Bakkie ➢ Heavy vehicle ➢ Minibus ➢ Minibus Taxi ➢ Bus ➢ Motorcycle ➢ Bicycle ➢ Tractor ➢ Other ➢ Bus-train ➢ Panel van ➢ Articulated truck ➢ Caravan / trailer ➢ Animal drawn ➢ Unknown

Table 2.2: Table of Accident Factors

HUMAN FACTORS	VEHICLE FACTORS	ROAD FACTORS
➢ Pedestrian: Jay walking ➢ Speed too high for circumstances ➢ Overtook when unlawful/unsafe ➢ Turned in front of oncoming traffic ➢ Disregarded red traffic light / stop sign / yield sign ➢ Followed too closely ➢ Hit-and-run ➢ Intoxicated Driver: Use of liquor or drugs suspected ➢ Intoxicated Pedestrian: Use of liquor or drugs suspected ➢ Fatigue / Driver falling asleep ➢ Cell phone use / holding ➢ Other ➢ U-turn ➢ Intoxicated Cyclist: Use of liquor or drugs suspected ➢ Not known	➢ Overloading: Cargo / Passengers ➢ Brakes: Faulty ➢ Tyre burst prior to accident ➢ Smooth tyres ➢ Lights: Faulty, not switched on, blinding etc. ➢ Other ➢ Bicycle: No head lamp ➢ Lights: Dirty ➢ Chevrons: Dirty ➢ Chevrons: No reflective stripes ➢ Steering: Faulty ➢ Bicycle: No rear reflectors ➢ Unknown	➢ Poor visibility (Rain, mist, dust, smoke, dawn, dusk) ➢ Poor street lighting ➢ Sharp bend ➢ Blind rise / Corner ➢ Poor condition of road surface ➢ Road surface slippery / wet ➢ Traffic light / Road sign / Road marking defective ➢ Narrow road lane ➢ Road works ➢ Other ➢ Animals: Stray / Wild ➢ Unknown

Fatal road accident data was chosen for the purpose of this study, because literature showed that fatal road accident counts are generally thought of to be the most reliable of all (Hauer, E & Hakkert, AS, TRB 1185). The accuracy with which road safety is measured depends on the proportion of accidents reported and the accuracy with which this proportion is known. It is also important to note that not all accidents are reportable and not all reportable accidents are in fact reported. The probability of reporting also depends strongly on injury severity (Elvik, R & Borger Mysen, A, TRB 1665). Fatal injuries are believed to be almost completely reported in official road accident statistics.

2.1.2 Information on Vehicles Involved in Fatal Accidents

Included in the MS Access database are the vehicle registration numbers of all the vehicles involved in the fatal accidents. It was initiated to find more variables to include in the study regarding the involved vehicles i.e. colour, manufacturer, year, age etc. These variables were perceived as useful in terms of possible multiple regression models. An effort was made to obtain this data from the traffic department, but with no success. The author's request to obtain access to the information was denied.

2.1.3 Traffic Volume and Speed Data Availability

Other information which was found to be relatively available was road *traffic* and *speed* data. Mikros Traffic Monitoring (Pty) Ltd acts as a service provider for SANRAL by collecting traffic data for information purposes. The data is captured and published in the CTO Yearbook, which then serves as a compendium of traffic information obtained at CTO stations (electronic counting equipment primarily utilizing loop technology; enhanced on certain installations by axle or weigh-in-motion sensors) on the primary roads, highlighting the latest available traffic characteristics. It contains information on 324 permanent stations and 429 secondary stations (according to the CTO Yearbook 2002) and this information is accessible and can be made available from the CTO Data Bank on request from SANRAL or by contacting Mikros Traffic Monitoring (Pty) Ltd for the latest CTO Yearbook.

The CTO stations are placed on selected links of the national and primary road network and yield information such as Average Daily Traffic (ADT), Average Daily Truck traffic (ADTT) and estimated Average Daily 80 kN Equivalent Axles in the worst lane (ADE80). The Yearbook contains flow information in tabular as well as in graphical format. The directional traffic distribution during an average normal weekday, hourly flow variations during an average normal week and daily traffic variations are also provided through graphical representations.

Comprehensive Traffic Observations (CTO) started in South Africa in 1984 when a pilot study was conducted on the 600 km long national route N3 between Johannesburg and Durban. As a result of the success of this study the National Transport Commission (now the South African National Roads Agency Limited) (SANRAL) decided in June 1985 to expand the CTO network to traffic counting stations.

The CTO Yearbook of 2002 contained the following passage on the importance and availability of traffic data:

"Traffic data is perhaps the most important component of the information necessary for the planning, design and operation (including maintenance) of a road network. Preferable this information should be obtained by measuring existing traffic characteristics such as speed, volume and composition on a continuous basis."

By keeping the above statement in mind, traffic data should always remain well recorded, organized and be available for public use. As mentioned before, traffic data can be made available for public use by contacting SANRAL or Mikros Traffic Monitoring (Pty) Ltd for the latest CTO Yearbook.

Below are typical variables of which data are available from the CTO Yearbooks as given per CTO counting station report (see Appendix A2 for an example of such a report):

- Total number of vehicles
- ADT (Average Daily Traffic; veh/d)
- ADTT (Average Daily Truck Traffic)
- Percentage of Trucks
- Truck Split % (short:medium:long)
- Percentage of Night Traffic (20:00-06:00)
- Speed limit (km/h)
- Average Speed (km/h)
- Average Speed – Light Vehicles (km/h)
- Average Speed – Heavy Vehicles (km/h)
- Average Night Speed (km/h)
- 15th percentile speed
- 85th percentile speed
- Percentage Vehicles in excess of the speed limit
- Percentage vehicles in flows over 600 veh/h
- Highest volumes on the road and in northern and southern directions
- 15th highest volumes on the road and in northern and southern directions
- 30th highest volumes on the road and in northern and southern directions
- Total number of heavy vehicles
- Est. average number of axles per truck
- Est. truck mass (ton/truck)
- Est. average E80/truck
- Est. daily E80's on the road, in the northern and southern direction and northern and southern worst lanes

Another source, from which speed and traffic volume data can be obtained, is the website of the Provincial Administration of the Western Cape (see the references at the end of this document for the

reference link). This website contains road network information reports on traffic volumes and speed data on the different road types within the Western Cape and is very user-friendly. The user can perform queries by following the instructions on an easy-to-use input screen, after which the query is performed and the user is redirected to an output screen containing the query results.

A typical station data report contains hourly traffic flows and graphical representations of traffic distributions. The report can immediately be printed from the output screen. Obtain the report by providing the following input variables:

- Road Number (e.g. NR00101, MR00584, DR02306)
- Authority Type (e.g. specify "Districts Roads Engineer" or "District Municipality")
- Area (e.g. Paarl, Cape Winelands)
- Road Type (e.g. Main Road, Divisional Road, National Road, Trunk Road)

Link volume reports can be obtained by following the same procedure as stated above. A traffic distribution report per authority type can be directly obtained from the website without any variable specification. This report is only available in *.PDF* format.

After considering each data source, namely the CTO station data from SANRAL and the PAWC website, it was decided that the CTO Yearbooks from SANRAL were the most suited data source for the purpose of this study. The data obtained from CTO Yearbooks were easier to integrate with the MS Access accident database on fatal road accidents as discussed in previous paragraphs and paragraphs to follow. Even though the type of data contained within the station data reports and link volume reports from the PAWC proved to be very comprehensive and detailed, the data are outdated, but it is still available for use, depending on the purpose it is to be used for.

2.1.4 Geometric and Terrain Information Availability

A source which was found to be the most convenient for gathering geometric and terrain data was the website of the Provincial Administration of the Western Cape containing road network information reports. It is a convenient source, because any data or information downloaded is immediately in electronic format. The website is very user-friendly and can easily be used to perform queries (as mentioned before). These queries can be downloaded in report form and in a printable format such as

.PDF. Another option is to download the information to a MS Excel format so data are immediately available for manipulation or analysis.

Geometric and terrain information can be found from road logs and strip charts on relevant routes by following the appropriate links on the website. The variables included in a specific road log can be specified by the user. Road logs can therefore be custom made according to the user's needs. Strip charts are visual representations of the information contained within a road log. Each route is illustrated along its chain distances and all geometric and terrain information is included (from the content of the road log which had been customized by the user).

Below is an example from a typical strip chart (extract from NR00104 Touws River Bridge to Jct. Mun MR83 Laingsburg):

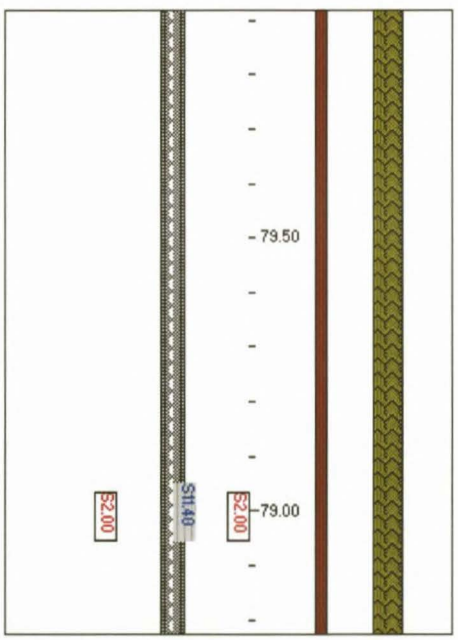

Fig. 2.1.2: Example of a typical strip chart (PAWC website)

As can be seen in Figure 2.1.2 above typical geometric information available is the total paved width and shoulder widths in metres. From these values the lane width can be calculated. In this case the lane width is 3.70m ((11.40 – 2*2.00)/2), which is a standard lane width used in South African practice. Other information available is the climate and terrain type. In the example above the terrain is flat (thin brown strip directly to the right of the road) and the area has a dry climate (the olive green strip on the left being the indicator). An example of a typical road log report can be found in Appendix A3.

2.2 Data Quality

In this section it is discussed what is considered to be the components of data quality. It will then be discussed how these theoretical principles apply to this study.

It is important to remember that there is a difference between data and information. Data is information in its early stages, while information is the processed data, presumably in usable form, ready to be used for whatever purpose it was intended. It is also relative, as one user's data could be another's information and vice versa. Data quality measures, therefore, are only prescribed if there exists a clear understanding and consideration of all intended uses of the data.

The following are components of data quality (O'Day, 1993):

- Completeness of coverage – the degree to which the data collection system contains all the cases defined by the data collection threshold.
- Consistency of coverage – whether the degree of reporting varies by jurisdiction, time, personal characteristics, weather, or other factors.
- Missing data – in addition to the problem of missing cases, there may be missing data elements for cases that are reported on.
- Consistency of Interpretation – whether the report elements are reported in the same manner in different jurisdictions, or by different reporting officers.
- The right data – having the right data elements for the purpose it is going to be used for.
- Appropriate level of detail – some data elements are needed in more detail than others for different purposes.
- Correct entry procedures – the correct way of entering the data in to a computer database.
- Freedom from response error – when something was measured, was it measured correctly?

Turner (December, 2002) prepared a white paper for the Office of Policy Federal Highway Administration, Washington DC, in which he defined data quality as follows:

"Data quality is the fitness of data for all purposes that require it. Measuring data quality requires an understanding of all intended purposes for that data."

After Turner viewed literature on data quality and recommended data quality measures, he finally recommended the following components:

- Accuracy - The measure or degree of agreement between a data value or set of values and a source assumed to be correct. Also, a qualitative assessment of freedom from error, with a high assessment corresponding to a small error.
- Completeness (also referred to as availability) - The degree to which data values are present in the attributes (e.g., volume and speed are attributes of traffic) that require them.
- Validity - The degree to which data values satisfy acceptance requirements of the validation criteria or fall within the respective domain of acceptable values.
- Timeliness - The degree to which data values or a set of values are provided at the time required or specified.
- Coverage - The degree to which data values in a sample accurately represent the whole of that which is to be measured.
- Accessibility (also referred to as usability) - The relative ease with which data can be retrieved and manipulated by data consumers to meet their needs.

According to Turner, there are several other data quality measures that could be appropriate to specific traffic data applications, but these six measures above are fundamental measures that should be universally considered for measuring data quality in traffic data applications.

The principles presented by Turner (2002) are very similar to the principles presented by O'Day (1993) on data quality. It was decided to investigate the MS Access database according to O'Day's principles. Only four principals were applicable:

- Completeness of coverage – The database used for this study is based only on fatal road accidents and accident rates and frequencies should not be mistaken as applicable to the total number of accidents for South Africa. It is outdated as the last accident which was recorded dates 23 July 2005.

- Missing data – It became evident that there was a number of missing data cases which exist in the database. The variable "Age Group" had to be omitted from the study due to a relatively large number of missing entries. When queries were performed, two queries with at least one variable in

common, were suppose to show the same number of items for a certain variable, but the queries failed to do so. The author had no choice, but to assume that this was due to missing data.

- Appropriate level of detail – A variable which had a very low level of detail was "Route Description". It was impossible to pinpoint the exact location of where accidents occurred, as the route description was indicated as e.g. "Cape Town – Goodwood" with no distances or exact coordinates. This jeopardizes a thorough investigation of a particular route in terms of past and future accidents.

- Correct entry procedure – The "Age Group" variable showed not only cases of missing data entries, but the data that were entered were not entered in a correct consistent manner. It seemed as if a coding system was applied in some instances and the correct numerical number for a particular age was entered for other cases. After investigating the matter it was confirmed by the RTMC that there occurred an error in the database which scrambled the "Age Group" variable. Time constraints did not allow for remedial action.

In the Road Safety Good Practice Guide of the UK (http://www.dft.gov.uk), in terms of monitoring programmes, the following examples are given of where data quality can be impaired through inadequate briefing:

- Collection in inappropriate weather conditions.
- Collection at inappropriate locations.
- Measuring both directions of travel (without identifying each measurement), when only one is required or both required separately; and
- not collecting sufficient, or even any, "before" data.

The UK Guide also states that experience has shown considerable variation in the quality of data collection even when collected against a prior written specification. The following guidelines are suggested:

- Where relevant, the same equipment and, preferably, personnel should be used for *before* and *after* monitoring to ensure the consistency of results.

- Experience also suggests that automatic equipment should be checked more than once a week to ensure it is continuing to operate correctly and has not been vandalised.
- In the case of attitude surveys, it is desirable that the commissioning agent attends the interviewer briefing meeting to maintain consistency of approach and hence quality of the data collected.
- Data that can be collected automatically must still be analysed consistently. Careful specification, briefing and supervision of the analysis will be essential to obtain reliable results.
- Back-up plans should be in place in case things do go wrong. It is recommended that a contingency element be included in the monitoring budget in case of such problems (e.g. vandalism, theft, bad weather).

Consistent application of above suggestions should ensure that data quality is continually enhanced. Even though the suggestions given above are ones which might already be used in South Africa, it is important to be reminded of them to ensure that they are constantly adhered to.

In an ETSC (European Transport Safety Council) briefing (2001) it was summarized how perceived and real responsibilities for road safety have changed over the last 30 years. The following principles or statements about sharing responsibility for road safety in terms of data collection and quality were deduced:

- The past will repeat itself if one chooses to ignore it and the same is true if you have inadequate data.
- Police data must not be considered as the only data source for obtaining adequate knowledge about accident specifics and the causes of injuries, and the epidemiology of road accidents.
- Police data by their very nature cannot provide the in-depth information necessary to evaluate highway design and behavioural causation issues, nor vehicle design and the biomechanics of injuries. A number of EU Member States have inadequate national data collection systems even for police data.
- Comparison studies of police and hospital records show gross under-reporting of several casualty classes (and it can also be assumed to be true for the RSA).
- Good data are fundamental to science-based strategies and their evaluation. This means that responsibilities should be shared between health and transport sectors and public/private partnerships should be developed in in-depth accident investigation.

In the *British Medical Journal, Vol 324* of 11 May 2002 two experts are asked their views on how road safety can be improved, one of whom is Richard Scurfield, leader of the transport sector of the World Bank's Transport and Urban Development Department. He was asked what the main obstacles are to promoting a scientific approach to road safety, to which he answered that the main obstacle is the poor quality of data in many countries and that it is essential to have reliable statistics for effective research and the development of well founded national road safety strategies.

It can be concluded that poor decisions are made when poor data quality exists. There are too many decisions made on inaccurate or incomplete data. Also, the impact that poor data quality can have is seldomly understood. Unless the time is taken to understand what data is available in current and planned applications, there might be too many databases to manage, which will inevitably have an impact on data quality.

High quality data will ensure that there exists only one view of a particular situation i.e. one view of what the real circumstances are like, particularly in a field like road safety. In the context of road safety very real and practical measures must be implemented to show a potential decrease in accident rates on a particular road section. The particular remedial action taken must be based on data which can make decision making possible within a relatively narrow confidence interval for success at a relatively high confidence level. Even though this is seldom achieved, it is certainly something to strive for, for as far as it is possible and allowed by the overall scope of a particular project.

2.3 Fatal Road Accident Statistics from Arrive Alive

This section gives some fatal road accident statistics as published by Arrive Alive, either on their website (http://www.arrivealive.co.za) or in other publications. These statistics are mostly summarized by province. The Western Cape Province is highlighted in discussions, as this study focused mostly on road sections along the national roads in the Western Cape Province.

In order to give some background as to how the fatal road accident rate and fatality rates changed over the course of 20 years, Figure 2.6 is provided based on fatal crash rates and fatality rates (fatal crashes and fatalities per 100 million veh-kms travelled) for the period 1983-2003 for South Africa. At the

time of compilation of this document, no further data after the year 2003 were available for inclusion in the historic dataset.

It can be seen from Figure 2.3.1 that from 1983 onwards there were on average a steady decrease in the fatal accident and fatality rates until 2000, where the rates suddenly increased relatively sharply over the course of two years. It seems that from 2002 onwards the fatal accident and fatality rates, although probably insignificant, both show a relatively small decrease.

In the following sections fatal accident and fatality rates per province and some road user information per province are provided to illustrate the current road safety situation of South Africa.

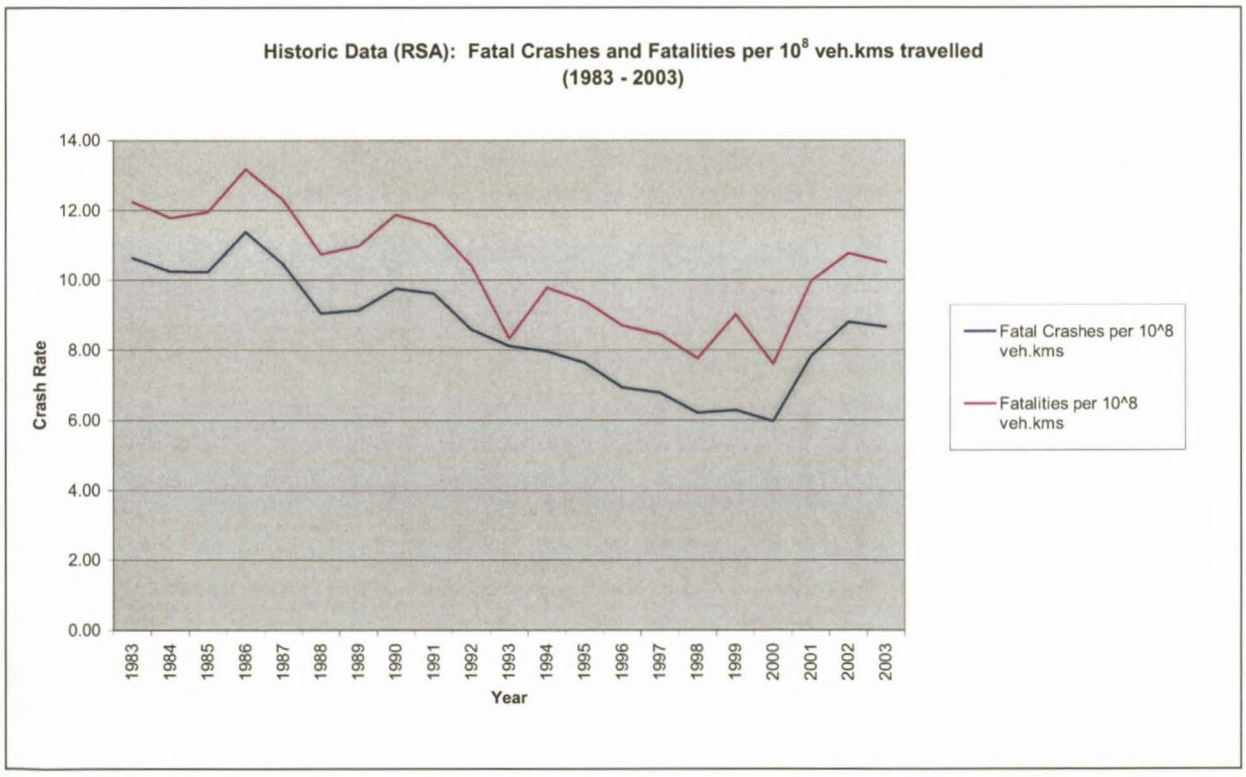

Fig. 2.3.1: Historic Data (RSA) - Fatal Accidents and Fatalities per 100 million veh-kms travelled 1983-2003

2.3.1 Fatality and Fatal Accident Rates and Frequencies per Province for South Africa

Figures 2.3.2 and 2.3.3 illustrate the fatal road accident and fatality rates (fatal accidents and fatalities per 100 million veh-kms travelled) per province for South Africa for the years 2002-2004. While observing these figures it can be seen that Kwazulu-Natal still have one of the highest fatal road

accident and fatality rates compared to the rest of the country with the Limpopo Province showing the highest fatal accident and fatality rates. The rest of the country shows fatal road accident and fatality rates of approximately the same order with minimal difference.

Figures 2.3.4 and 2.3.5 illustrate the number of fatal accidents and fatalities per province for South Africa for the years 2002-2004. It is clear that Gauteng and Kwazulu-Natal show the highest fatal road accident and fatality frequencies of all nine provinces with the Western Cape Province having the third highest frequencies in fatal road accidents and fatalities. The rest of the provinces show frequencies of approximately the same order with the Northern Cape having the least fatal road accidents and fatalities in the country.

It is interesting to note that even though Gauteng has the highest fatal accident and fatality frequencies, the accident and fatality rates for this province are the lowest in the whole country. The same effect is illustrated for the Western Cape. In spite of having the third highest fatal road accident and fatality frequencies in the country, the fatal road accident and fatality rates for this province are the second lowest in the country. Kwazulu-Natal remains top ranked in terms of fatal road accident and fatality frequencies as well as rates for this province.

The provinces which show an increase in fatal road accident and fatality rates are Eastern Cape, North West and Limpopo. Provinces showing an decrease in fatal road accident and fatality rates are Gauteng, Kwazulu-Natal, Western Cape, Free State, Mpumalanga and Northern Cape.

In terms of fatal road accident and fatality frequencies, the provinces showing an increase in these frequencies are Kwazulu-Natal, Eastern Cape, North West and Limpopo. Provinces showing an decrease in these frequencies are Free State and Western Cape, with Gauteng only showing a decrease in the number of fatalities. Provinces which has shown to have no significant change in fatal road accident and fatality frequencies in the course of the period 2002-2004 are the Northern Cape and Mpumalanga Provinces. Gauteng shows no significant change in the number of fatal road accidents.

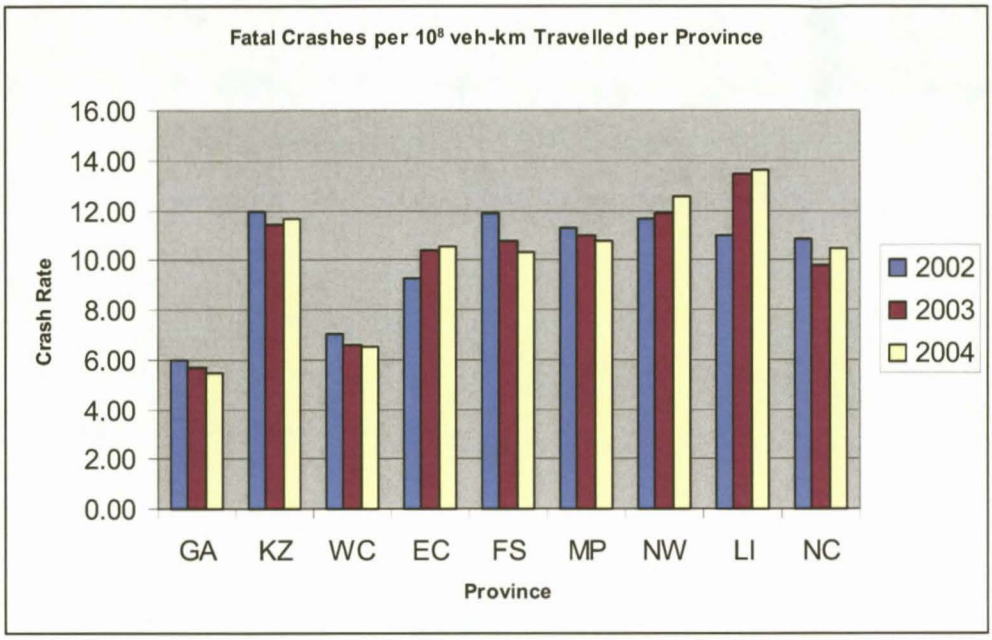

Fig. 2.3.2: Fatal Accidents per 100 million veh-km Travelled per Province 2002, 2003 and 2004

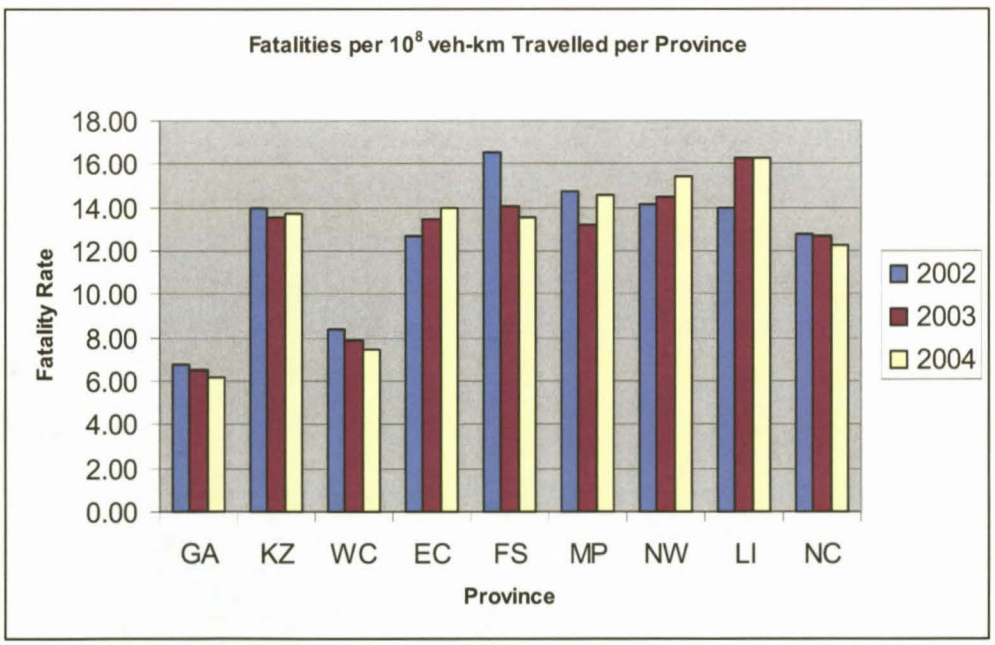

Fig. 2.3.3: Fatalities per 100 million veh-km Travelled per Province 2002, 2003 and 2004

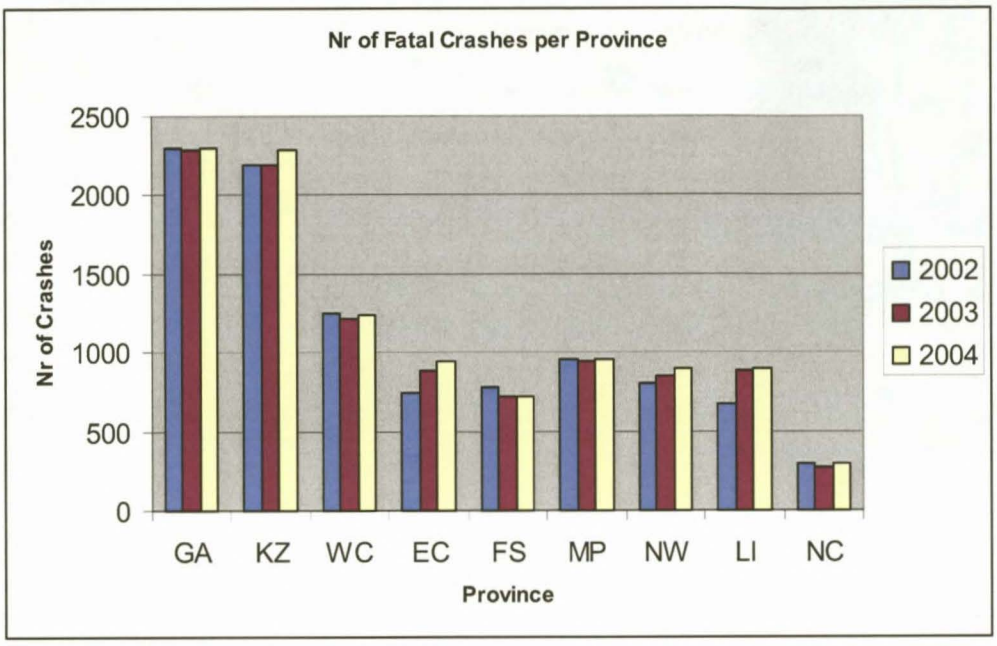

Fig. 2.3.4: Nr of Fatal Accidents per Province 2002, 2003 and 2004

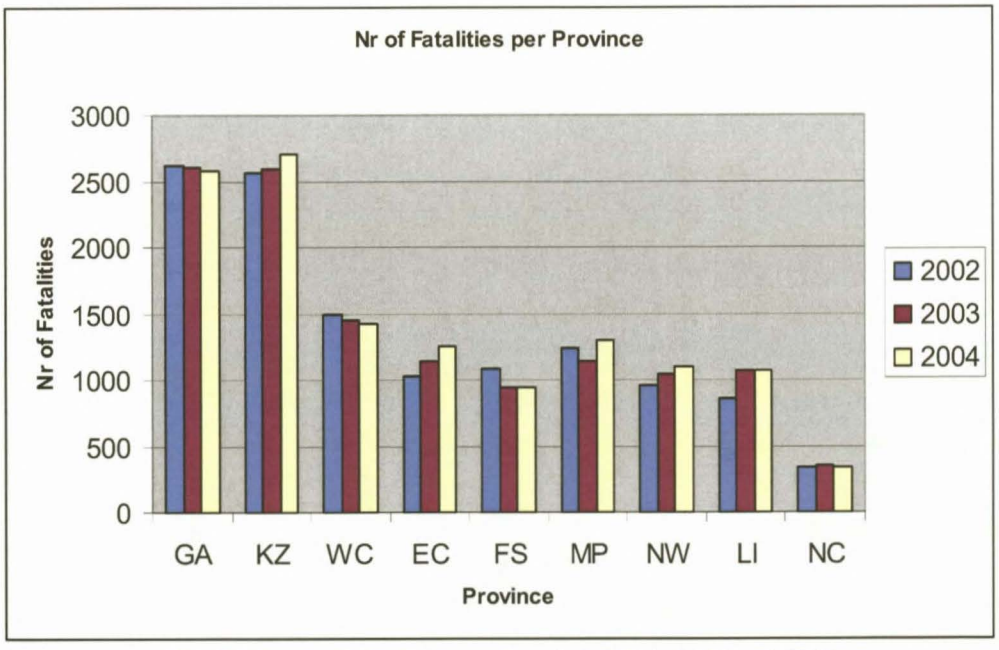

Fig. 2.3.5: Nr of Fatalities per Province 2002, 2003 and 2004

2.3.2 Million Veh-kms Travelled per Province for South Africa

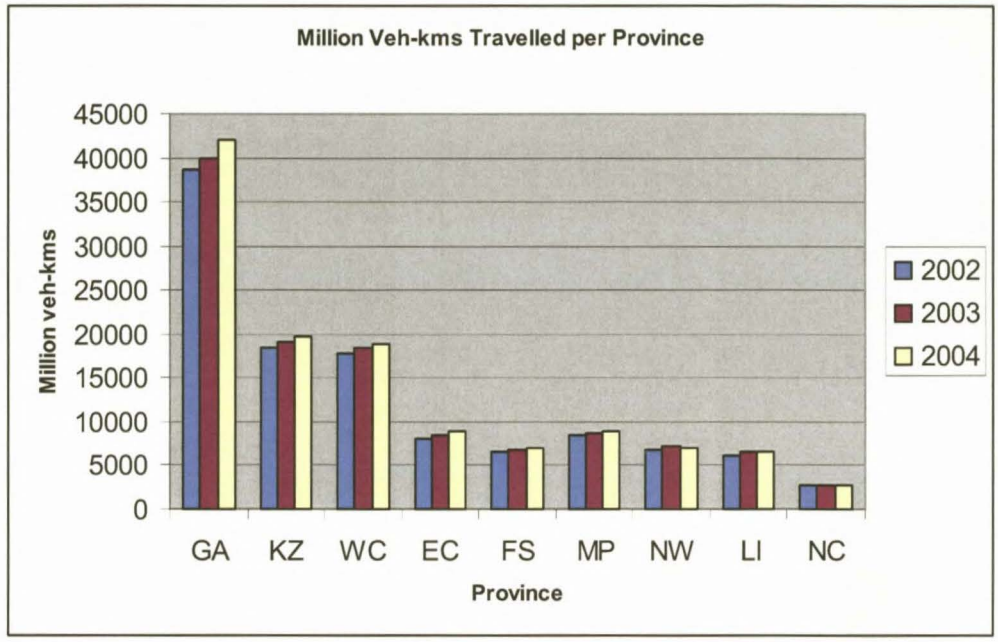

Fig. 2.3.6: Million Veh-kms Travelled per Province 2002, 2003 and 2004

Figure 2.3.6 illustrates one of the possible reasons for the fluctuations in fatal road accident and fatality rates per province in comparison with the fatal road accident and fatality frequencies per province. It was noted before that Gauteng shows the lowest fatal road accident and fatality rates in the country. This might be explained by observing from Figure 2.3.6 that this province has a higher level of exposure than the rest of the country. The same argument could be valid for the Western Cape Province. The Northern Cape Province shows relatively low fatal accident and fatality frequencies and from Figure 2.3.6 it also shows a relatively low exposure level than for the rest of the country, which could mainly be due to the lower traffic volumes on the roads in comparison with the rest of the country. The fatal road accident and fatality rates of this province is one of the highest in the country though.

All provinces show an increase in the level of exposure in terms of 100 million veh-kms travelled. The exception is the Northern Cape Province which seems to show no significant change in the course of the period 2002-2004.

2.3.3 Fatality and Fatal Accident Rates and Frequencies Data per Province for South Africa

All graphical representations given in paragraphs 2.3.1 and 2.3.2 are based on the data given below in Tables 2.3 to 2.5.

Table 2.3: Nr of Fatal Accidents and Fatalities per Province for South Africa 2002-2004

Province	Nr of Fatal Accidents			Nr of Fatalities		
	2002	2003	2004	2002	2003	2004
GA	2297	2284	2296	2621	2608	2574
KZ	2191	2189	2288	2567	2593	2705
WC	1253	1210	1240	1499	1455	1421
EC	747	886	947	1023	1144	1255
FS	778	725	718	1085	949	947
MP	950	949	959	1245	1144	1298
NW	797	850	891	964	1037	1095
LI	671	880	892	855	1066	1071
NC	289	273	292	340	353	344
RSA	9973	10246	10523	12198	12348	12709

Table 2.4: Fatal Accidents and Fatalities per 100 Million veh-kms Travelled per Province for South Africa 2002-2004

Province	Nr of Fatal Accidents/10^8 veh-km			Nr of Fatalities/10^8 veh-km		
	2002	2003	2004	2002	2003	2004
GA	5.95	5.71	5.47	6.79	6.52	6.13
KZ	11.93	11.43	11.63	13.98	13.54	13.75
WC	7.02	6.59	6.54	8.39	7.92	7.49
EC	9.24	10.39	10.56	12.66	13.41	13.99
FS	11.87	10.76	10.30	16.55	14.09	13.58
MP	11.26	10.97	10.74	14.76	13.22	14.53
NW	11.69	11.91	12.54	14.14	14.53	15.41
LI	10.98	13.47	13.59	13.99	16.32	16.32
NC	10.82	9.78	10.44	12.73	12.65	12.30
RSA	8.79	8.69	8.63	10.75	10.48	10.42

Table 2.5: Million Veh-kms Travelled per Province for South Africa 2002-2004

Province	Million Veh-kms Travelled		
	2002	2003	2004
GA	38600.48	39974.39	41968.16
KZ	18363.22	19150.32	19678.47
WC	17861.38	18370.88	18966.76
EC	8082.32	8528.07	8970.3
FS	6554.77	6737.33	6973.17
MP	8435.13	8651.27	8930.54
NW	6815.67	7138.41	7105.8
LI	6111.37	6532.99	6563.98
NC	2669.82	2791.15	2796.79
RSA	113494.15	117874.8	121953.98

2.3.4 Road User Data for Fatal Road Accidents per Province for South Africa

Statistical summaries given in this section are based on data in the MS Access database analysed for this study. Fatalities summarized by province and by different road user categories (i.e. race group, road user groups, seatbelt status etc.) were not readily available. Road user information which is available in publications is either not in electronic format or is not available in the form of data tables which can be manipulated and used as needed for research purposes (e.g. only graphical summaries are given).

Note that frequencies cannot be compared in the same way as was the case in the previous section when fatal road accident and fatality rates were compared by province. Here, the relevant rates are not calculated and this should be taken into account when inspecting the graphs which is to follow based on different road user categories as was involved with fatal road accidents.

i) Fatalities by Road User Group per Province

Figures 2.3.7, 2.3.8 and 2.3.9 represents driver, passenger and pedestrian fatalities for South Africa per province.

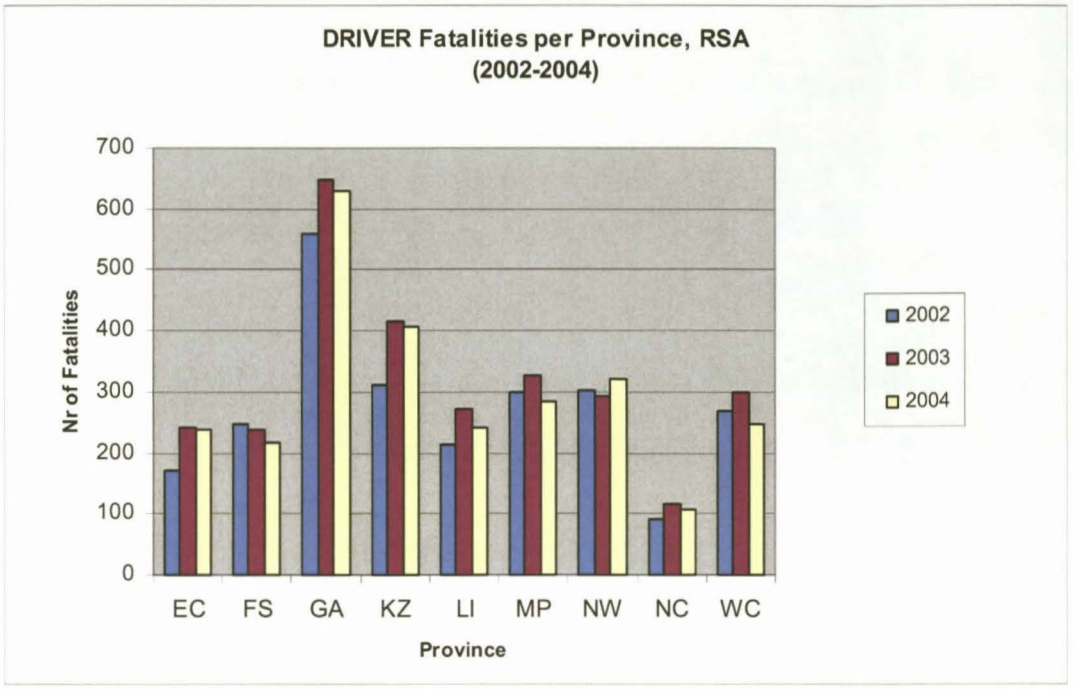

Fig. 2.3.7: Driver fatalities by Road User Group per Province, RSA, 2002-2004

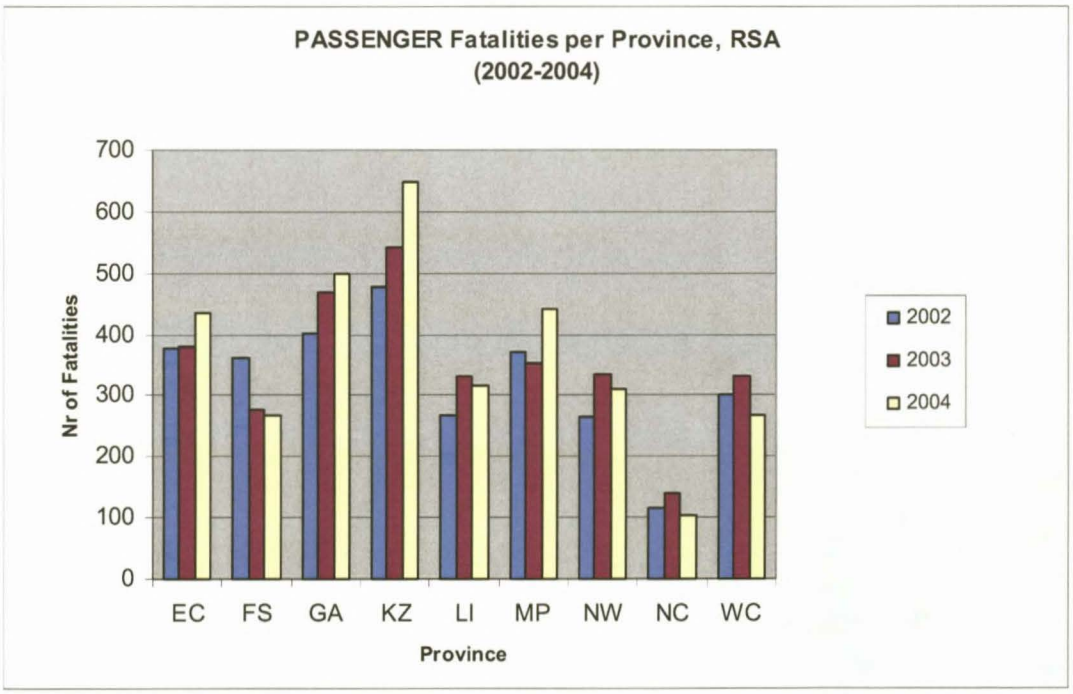

Fig. 2.3.8: Passenger fatalities by Road User Group per Province, RSA, 2002-2004

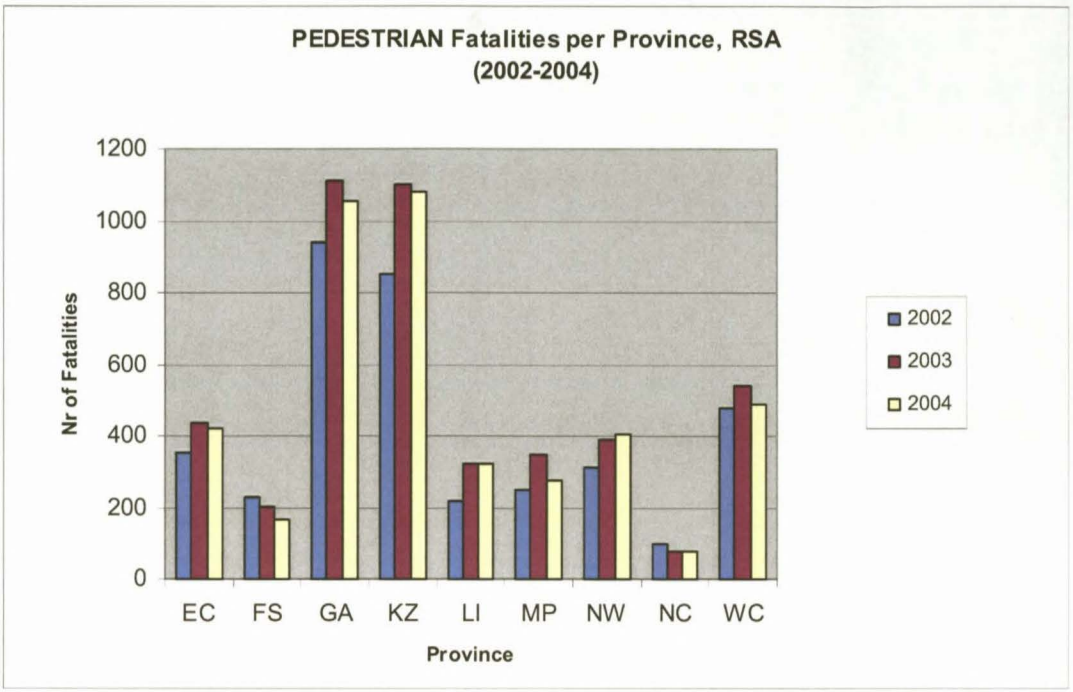

Fig. 2.3.9: Pedestrian fatalities by Road User Group per Province, RSA, 2002-2004

The graphs given above are based on the following data table i.e. Table 2.6.

Table 2.6: Fatalities per Road User Group per Province, RSA, 2002-2004 (Studied database)

	Driver			Passenger			Pedestrian			TOTAL		
	2002	2003	2004	2002	2003	2004	2002	2003	2004	2002	2003	2004
EC	170	241	238	378	379	435	351	437	421	899	1057	1094
FS	248	237	217	361	276	267	231	204	165	840	717	649
GA	558	649	631	401	468	499	938	1112	1055	1897	2229	2185
KZ	313	416	407	479	541	649	851	1102	1081	1643	2059	2137
LI	214	271	240	269	332	316	217	321	321	700	924	877
MP	301	327	283	372	354	440	248	348	276	921	1029	999
NW	304	294	322	265	335	309	311	391	404	880	1020	1035
NC	91	116	107	116	139	104	100	78	77	307	333	288
WC	270	301	247	300	333	268	477	539	489	1047	1173	1004
RSA	**2469**	**2852**	**2692**	**2941**	**3157**	**3287**	**3724**	**4532**	**4289**	**9134**	**10541**	**10268**

ii) Fatalities by Race Group per Province

Fig. 2.3.10 illustrates that Gauteng have the most white fatalities in the country with the Western Cape and North West provinces with approximately the second highest white fatality frequencies. The Eastern Cape and Mpumalanga is showing a decreasing tendency with the Free State showing no significant change. The rest of the country has an increasing tendency in white fatalities.

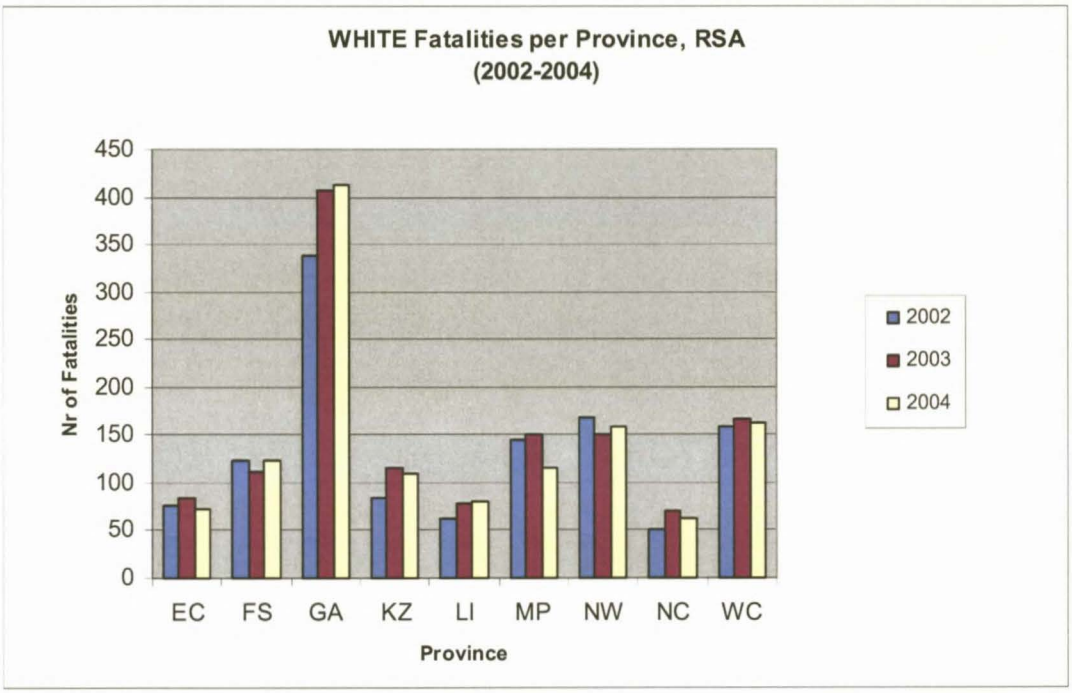

Fig. 2.3.10: White race group fatalities per Province, RSA, 2002-2004

The Western Cape has the highest number of coloured fatalities in the country by a significant amount more than the rest of the country, although this province is showing a decreasing tendency in the coloured fatality frequency (refer to Figure 2.3.11). The Northern Cape and Eastern Cape have approximately the second highest number of coloured fatality frequencies in the country and are also showing a decreasing tendency. The rest of the country is showing an overall decreasing tendency in coloured fatality frequency.

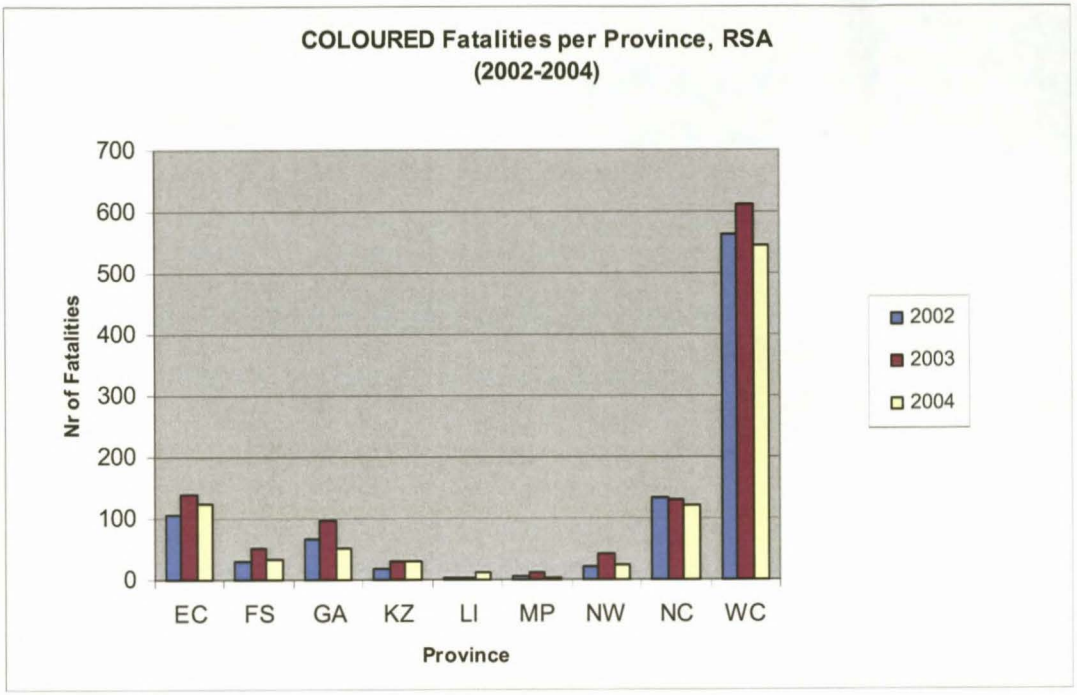

Fig. 2.3.11: Coloured race group fatalities per Province, RSA, 2002-2004

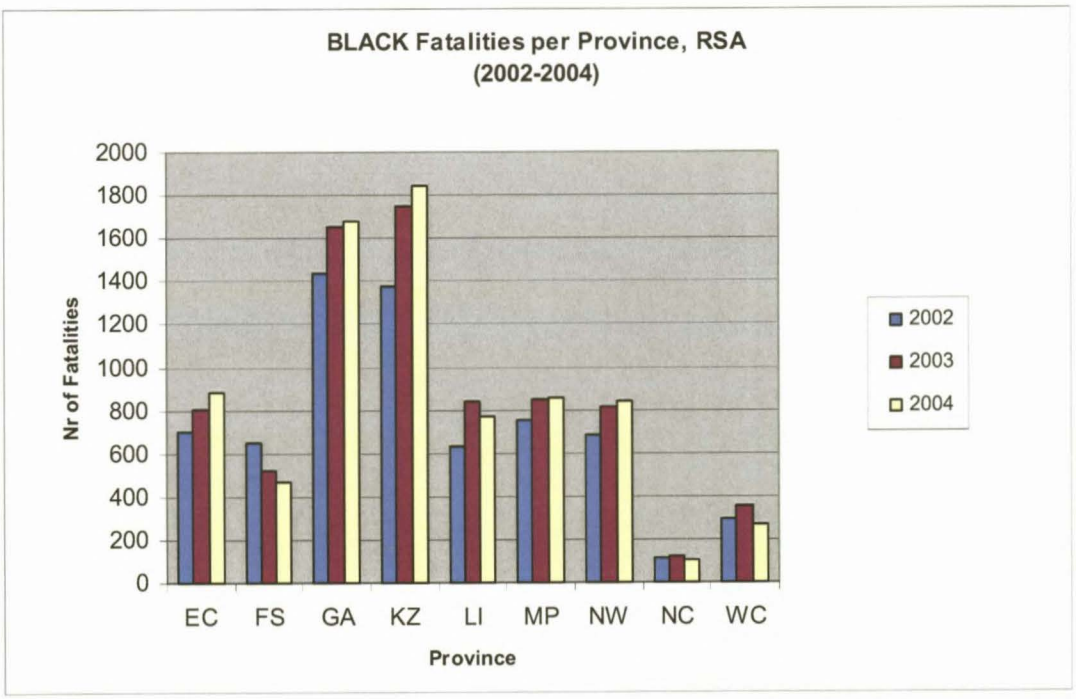

Fig. 2.3.12: Black race group fatalities per Province, RSA, 2002-2004

In Figure 2.3.12 it can be seen that Gauteng and Kwazulu-Natal have the highest black fatality frequencies in the country and are showing and increasing tendency in these frequencies. The Eastern

Cape, Limpopo, Mpumalanga and North West have approximately the same black fatality frequencies; also showing and increasing tendency in these frequencies. These provinces overall have the second highest number of black fatalities in the country. The Free State, Northern Cape and Western Cape Province have the lowest number of black fatalities in the country and are showing a decreasing tendency in these frequencies.

Table 2.7(a) contains the data on which Figures 2.3.10, 2.3.11 and 2.3.12 are based.

Table 2.7(a): Fatalities per Race Group per Province, RSA, 2002-2004

	White			Coloured			Black			TOTAL		
	2002	2003	2004	2002	2003	2004	2002	2003	2004	2002	2003	2004
EC	76	85	72	107	138	125	701	809	887	884	1032	1084
FS	123	111	124	29	51	32	650	518	464	802	680	620
GA	338	406	412	66	97	51	1433	1651	1677	1837	2154	2140
KZ	84	116	110	19	30	29	1373	1748	1844	1476	1894	1983
LI	62	78	81	2	3	11	631	838	768	695	919	860
MP	145	150	116	5	11	4	749	846	858	899	1007	978
NW	169	150	159	20	42	25	682	814	840	871	1006	1024
NC	50	70	63	133	129	121	111	121	103	294	320	287
WC	158	166	163	563	611	546	296	359	272	1017	1136	981
RSA	1205	1332	1300	944	1112	944	6626	7704	7713	8775	10148	9957

Kwazulu-Natal has a definite majority in Asian fatalities with Gauteng having the second highest number of Asian fatalities in the country, but with a significant amount less than Kwazulu-Natal (Fig. 2.3.13). Both these provinces have a decreasing tendency in terms of these frequencies. The rest of the country has a relatively small amount of Asian fatalities i.e. approximately 20 or less Asian fatalities per province, overall showing a decreasing tendency.

The fatalities involving unknown race groups are summarized in Fig. 2.3.14. It seems as if 2003 was a year of large amounts of unknown race group fatalities. By 2004 the amount of unknowns significantly decreased to almost the same level as in the year 2002. A possible reason for the decrease in the number of unknown race group fatalities could be due to better completion of the accident report form by traffic officials and accident investigators. It is unclear however why 2003 showed such a significant increase in the number of unknown race group fatalities.

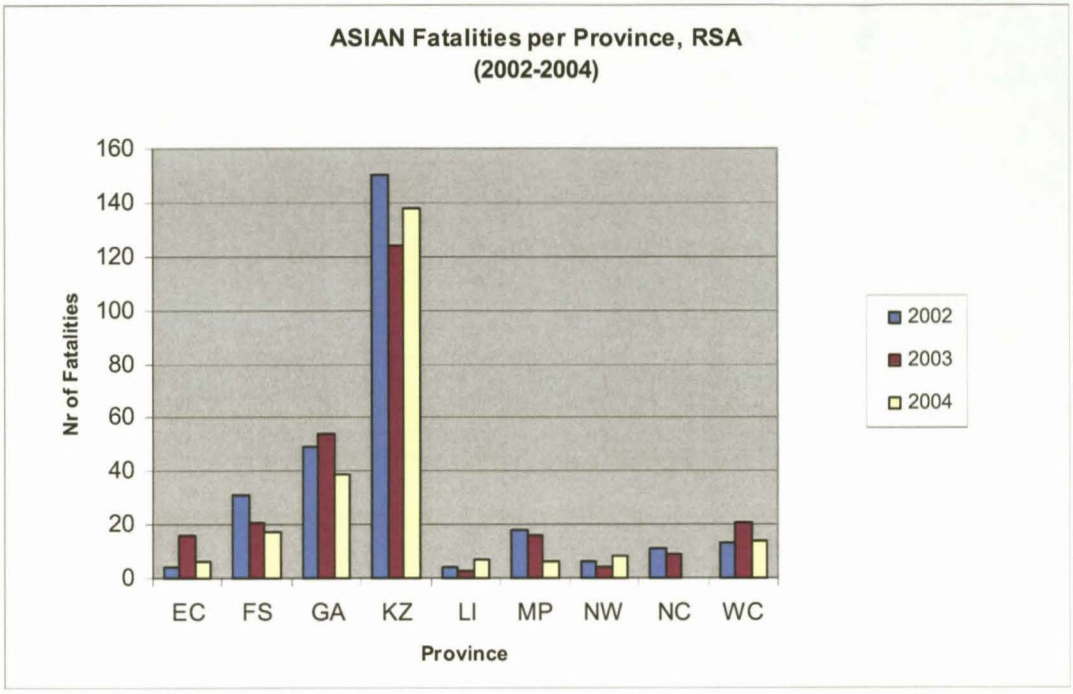

Fig. 2.3.13: Asian race group fatalities per Province, RSA, 2002-2004

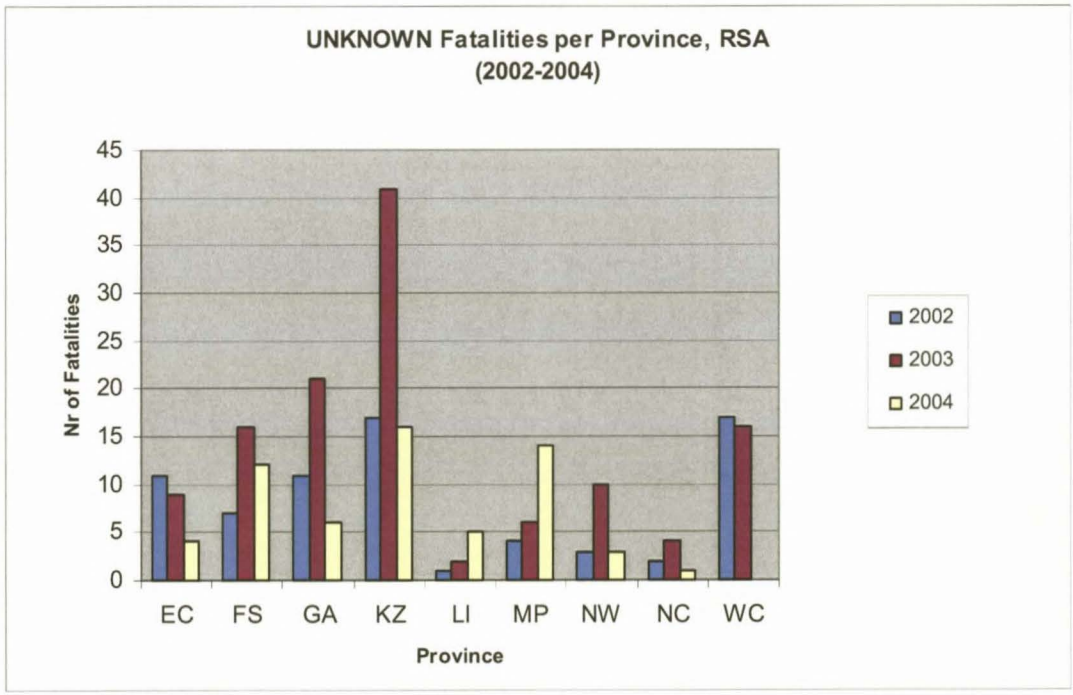

Fig. 2.3.14: Unknown race group fatalities per Province, RSA, 2002-2004

A graphical summary of the Foreign fatalities were not created due to the very small amount of foreigners who died on South African roads during the years 2002 to 2004. According to the database

there were only 6 Foreign fatalities which occurred during 2004 with 5 Foreign fatalities in the Limpopo Province and 1 Foreign fatality in Mpumalanga.

Table 2.7(b) contains the data on which Figures 2.3.13 and 2.3.14 are based.

Table 2.7(b): Fatalities per Race Group per Province, RSA, 2002-2004

	Asian			Foreigner			Unknown			TOTAL		
	2002	2003	2004	2002	2003	2004	2002	2003	2004	2002	2003	2004
EC	4	16	6	0	0	0	11	9	4	15	25	10
FS	31	21	17	0	0	0	7	16	12	38	37	29
GA	49	54	39	0	0	0	11	21	6	60	75	45
KZ	150	124	138	0	0	0	17	41	16	167	165	154
LI	4	3	7	0	0	5	1	2	5	5	5	17
MP	18	16	6	0	0	1	4	6	14	22	22	21
NW	6	4	8	0	0	0	3	10	3	9	14	11
NC	11	9		0	0	0	2	4	1	13	13	1
WC	13	21	14	0	0	0	17	16	9	30	37	23
RSA	286	268	235	0	0	6	73	125	70	359	393	311

iii) Seatbelt status (only fatality cases; Pedestrians excl.) per Province

Figure 2.3.15 illustrates the fatality cases for 2002-2004 where a seatbelt was worn, not worn or where the seatbelt status was unknown. There is a very small amount of cases where the seatbelt status was noted as "yes" relative to the amount of cases where "no" and "unknown" were noted. It is thus difficult to clearly see from this fatal road accident dataset what the seatbelt wearing behaviour of the accident victims truly was due to the very large amount of unknown cases. This could typically be the case where, for example, a large amount of passenger fatalities occurred and the seatbelt status of each victim could not necessarily be determined. Another factor (one of multiple factors) which has an influence is incomplete accident forms.

Table 2.8 contains the data on which Figure 2.3.15 is based.

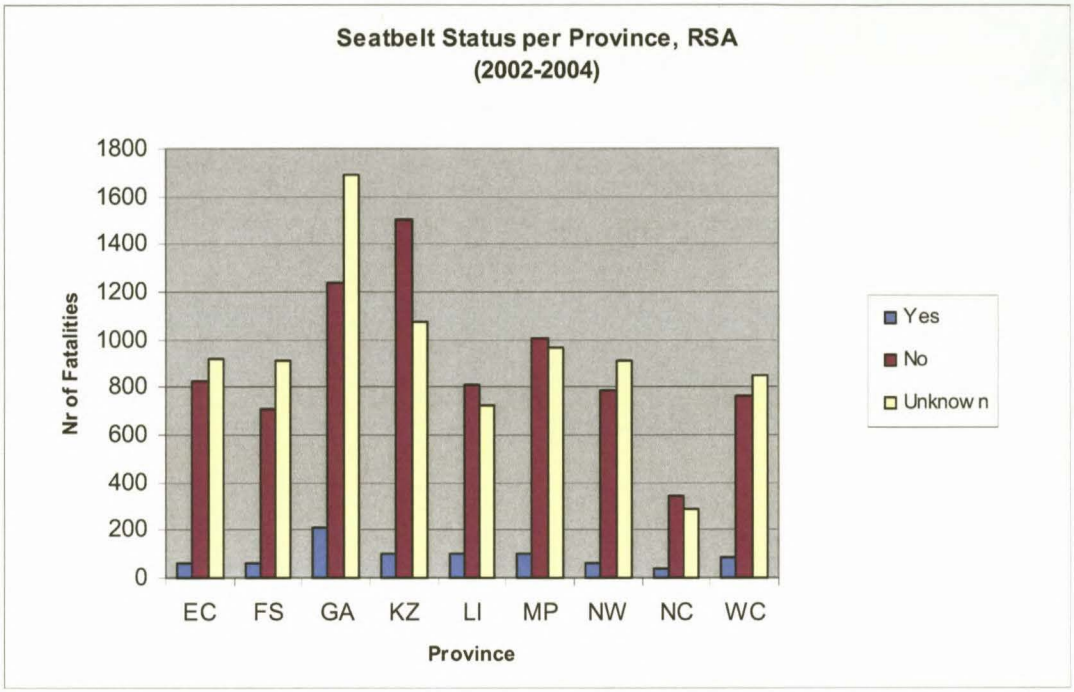

Fig. 2.3.15: Seatbelt Status (only fatality cases; excl. Pedestrians) per Province, RSA, 2002-2004

Table 2.8: Seatbelt Status (only fatality cases; excl. Pedestrians) per Province, RSA, 2002-2004

	Yes			Total	No			Total	Unknown			Total	TOTAL		
	2002	2003	2004		2002	2003	2004		2002	2003	2004		2002	2003	2004
EC	28	22	12	62	213	261	354	828	307	307	307	921	548	590	673
FS	29	19	15	63	237	231	243	711	341	341	227	909	607	591	485
GA	87	77	48	212	290	400	548	1238	580	580	531	1691	957	1057	1127
KZ	41	35	26	102	379	434	688	1501	369	369	341	1079	789	838	1055
LI	44	31	24	99	163	293	355	811	275	275	176	726	482	599	555
MP	36	37	30	103	262	276	470	1008	373	373	223	969	671	686	723
NW	28	12	20	60	219	233	332	784	318	318	278	914	565	563	630
NC	6	17	14	37	93	125	127	345	109	109	70	288	208	251	211
WC	44	18	26	88	204	273	283	760	322	322	205	849	570	613	514
RSA	**343**	**268**	**215**	**826**	**2060**	**2526**	**3400**	**7986**	**2994**	**2994**	**2358**	**8346**	**5397**	**5788**	**5973**

iv) Fatalities by Gender per Province

Figure 2.3.16 summarizes the fatalities for 2002 to 2004 by Gender. The frequency patterns are similar to those in the previous sections with Gauteng, Kwazulu-Natal, Western Cape and Eastern Cape having the most fatalities in the country. Male fatalities are higher than female fatalities in South Africa.

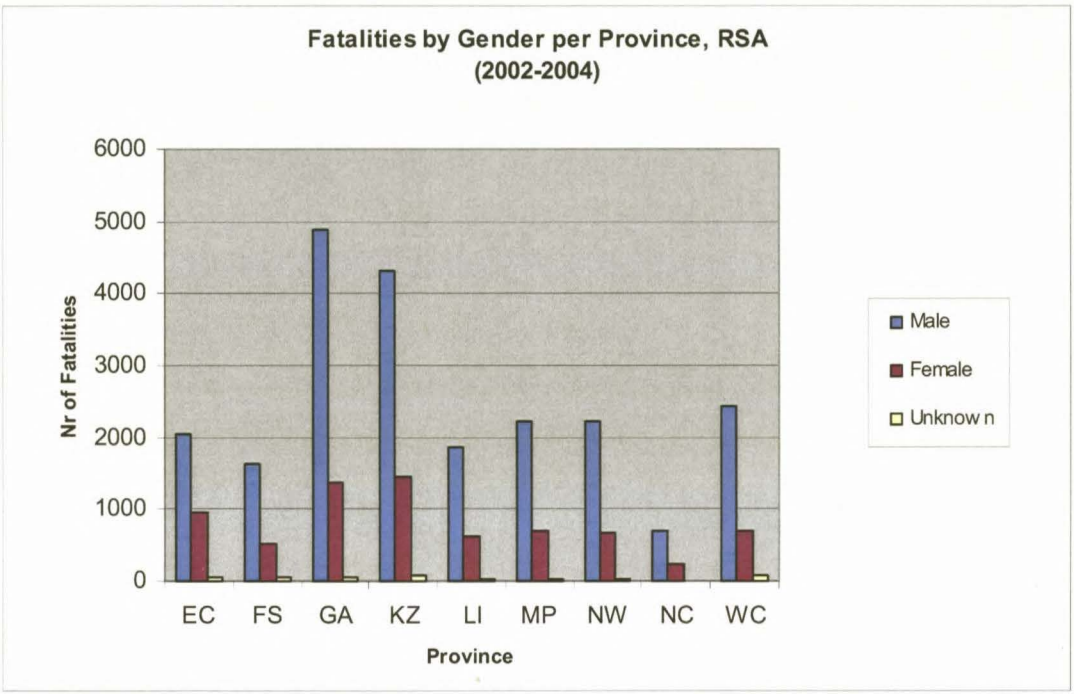

Fig. 2.3.16: Fatalities by Gender per Province, RSA, 2002-2004

Table 2.9 contains the data on which Figure 2.3.16 is based.

Table 2.9: Fatalities per Gender per Province, RSA, 2002-2004

	Male			Female			Unknown			TOTAL		
	2002	2003	2004	2002	2003	2004	2002	2003	2004	2002	2003	2004
EC	590	700	765	279	346	324	30	11	5	899	1057	1094
FS	605	544	482	206	157	154	29	16	13	840	717	649
GA	1478	1738	1677	406	468	489	13	23	19	1897	2229	2185
KZ	1184	1508	1621	436	504	499	23	47	17	1643	2059	2137
LI	543	690	630	155	231	230	2	3	17	700	924	877
MP	702	784	747	216	239	243	3	6	9	921	1029	999
NW	684	741	811	193	266	223	3	13	1	880	1020	1035
NC	234	245	217	71	84	69	2	4	2	307	333	288
WC	789	886	756	217	250	244	41	37	4	1047	1173	1004
RSA	6809	7836	7706	2179	2545	2475	146	160	87	9134	10541	10268

2.4 Data Analysis

In this section a theoretical discussion is given on the statistical techniques used to analyse the data for this study. It discusses the following techniques:

- Correspondence Analysis and Multiple Correspondence Analysis: Descriptive/exploratory techniques designed to analyze simple two-way and multi-way tables containing some measure of correspondence between the rows and columns.

- Association Rules: A data mining technique used to detect relationships or associations between specific values of categorical variables in large data sets.

- Chance Variation: The application of the Poisson distribution to fatal road accident rates in order to identify significant accident rate variation.

- Multiple Regression: A technique for describing and summarizing data consisting of observations on a dependent variable or response variable y and more than one independent variable. This leads to a best-fit prediction equation and a value of the coefficient of multiple determination R^2. A point prediction of y results from substituting specified values of the independent variables into the prediction equation.

2.4.1 Correspondence Analysis

i) General Methodology

Correspondence Analysis is a descriptive/exploratory technique designed to analyze simple two-way and multi-way tables containing some measure of correspondence between the rows and columns. It allows one to explore the structure of categorical variables included in the table. The most common kind of table of this type is the two-way frequency cross tabulation table. This technique is extremely useful for the purpose of this study, as the fatal accident database available contains variables with large amounts of categories. Cross tabulation of these variables provide large frequency tables which cannot be interpreted so easily through inspection alone.

The cross tabulation table of frequencies is first standardized, so that the relative frequencies across all cells sum to 1.0. The goal of a typical analysis is to represent the entries in the table of relative frequencies in terms of the distances between individual rows and/or columns in a low-dimensional space. The term distances will be used throughout the text, but it actually refers to the differences between the pattern of relative frequencies for the rows across the columns, and the columns across the rows, which are to be reproduced in a lower-dimensional solution.

Table 2.10: Frequency Table of South African Fatalities between December 2002 and August 2005 by Gender and Road User Type (from the Arrive Alive database used for this study)

	MALE	FEMALE	UNKNOWN	TOTAL
PEDESTRIAN	10805	3490	129	14424
DRIVER	8577	723	51	9351
PASSENGER	6506	4104	301	10911
TOTAL	25888	8317	481	**34686**

Table 2.11: Relative Frequency Table of South African Fatalities between December 2002 and August 2005 by Gender and Road User Type (from the Arrive Alive database used for this study)

	Male	Female	Unknown	Total
Pedestrian	0.311509	0.100617	0.003719	0.415845
Driver	0.247276	0.020844	0.001470	0.269590
Passenger	0.187568	0.118319	0.008678	0.314565
Total	0.746353	0.239780	0.013867	1.000000

Table 2.11 consists of 3 data rows and columns. The 3 column values in each row of the table are coordinates in a 3-dimensional space, and the (Euclidean) distances between the 3 row points could be computed in the 3-dimensional space. These distances between the points summarize all information about the similarities between the rows in the table.

Suppose that there exists a lower-dimensional space, in which the row points can be positioned in a manner that retains all, or almost all, of the information about the differences between the rows. All the information about the similarities between the rows (road user types in this case) could then be presented in a 1- or 2-dimensional graph. For small tables like the one provided above, this does not appear to be particularly useful, but the presentation and interpretation of very large tables are greatly simplified by using this method (e.g. representing 19 road accident type frequencies in a two-dimensional space).

The entries in the table can be reproduced from the row and column totals alone (or row and column *profiles* in the terminology of correspondence analysis; explained in paragraphs to follow), if the rows and columns in a table are completely independent of each other. According to the formula for computing the *Chi-square* statistic for two-way tables, the expected frequencies in a table, where the column and rows are independent of each other, are equal to the respective column total times the row total, divided by the grand total. Any deviations from the expected values (expected under the hypothesis of complete independence of the row and column variables) will contribute to the overall *Chi-square*. Thus, another way of looking at correspondence analysis is to consider it a method for decomposing the overall *Chi-square* statistic (or *Inertia* = *Chi-Square/Total N*) by identifying a small number of dimensions in which the deviations from the expected values can be represented.

Row points which appear to be relatively close to each other in the two-dimensional plot are similar with regard to their pattern of relative frequencies across the columns. This pattern becomes clear when a table of relative *row* frequencies are computed. Each entry r_{ij} (i = rows, j = columns) can be interpreted as the conditional probability that a case belongs to the j^{th} column, given its membership in the i^{th} row.

Looking at the table of relative *row* frequencies given below (i.e. row frequencies standardized, so that their sum in each row is equal to 100%; Table 2.12), it is not all that evident that there exists similarity, but it seems as if *Pedestrian* and *Passenger* show greater similarity between them, however minimal it may be. The distances between the points on the produced 2-dimensional plot (to follow) also confirms the small similarity and it can be calculated from the row coordinates (to be provided later in this section) that *Passenger* is about 7.08% closer to *Pedestrian* relative to the distance between *Driver* and *Pedestrian*.

Table 2.12: Table of Relative Row Frequencies (Row profile matrix) of Illustrated Example

	MALE (%)	FEMALE (%)	UNKNOWN (%)	TOTAL (%)
PEDESTRIAN	74.91	24.20	0.89	100
DRIVER	91.72	7.73	0.55	100
PASSENGER	59.63	37.61	2.76	100

After the 2-dimensional plot was produced (see Fig. 2.4.1) it was evident, along the most important first axis in the plot representing the first dimension (also to be explained later in this section), that the

row points have very large distances between them. *Driver* and *Pedestrian* is found on the left side of the origin (scale position 0) (*Pedestrian* is just left of the origin by a small amount, but this might be significant in terms of the interpretation of the results) and *Passenger* on the right side. The analysis not only distinguishes between row points by the distances between them, but also by their relative displacements from the origin. This is important, because great care needs to be taken when interpreting the results from a correspondence analysis, as will shortly be discussed.

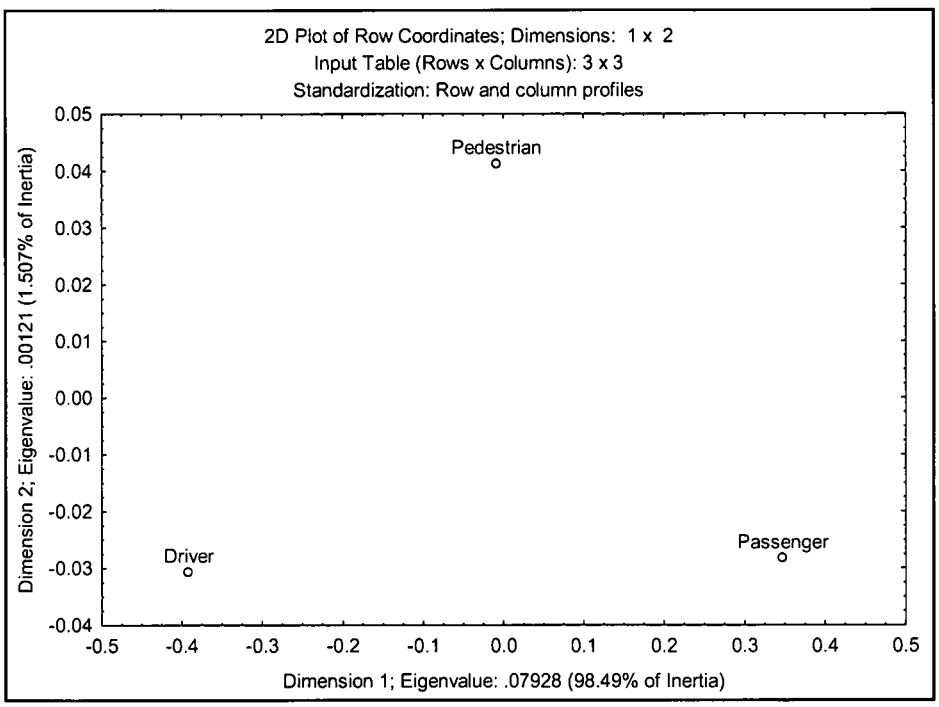

Fig. 2.4.1: Two-dimensional Plot of the Row Coordinates of the Illustrated Example

The table below contains the unstandardized coordinates of the results represented by the above 2-dimensional graph of the first two dimensions. The coordinates are calculated based on the row profile matrix (table of relative row frequencies). They are calculated to maximize the differences between the points with respect to the row profiles (row percentages).

Table 2.13: Unstandardized Row Coordinates resulting from a typical Correspondence Analysis

Row Name	Coordin. Dim.1	Coordin. Dim.2
Pedestrian	-0.008695	0.041262
Driver	-0.391614	-0.030656
Passenger	0.347117	-0.028275

The simple example provided above began with a discussion of the row-points in the table. However, one may rather be interested in the column totals. The column points may then be plotted in a small-dimensional space, which satisfactorily reproduces the similarity (and distances) between the relative frequencies for the columns, across the rows, in the table. This similar pattern across the rows will also become clear when a table of relative *column* frequencies are computed (see Table 2.14 below). Again, each entry r_{ij} (i = rows, j = columns) can be interpreted as the conditional probability that a case belongs to row i, given its membership in column j.

Table 2.14: Table of Relative Column Frequencies (Column Profile Matrix) of Illustrated Example

	MALE (%)	FEMALE (%)	UNKNOWN (%)
PEDESTRIAN	41.74	41.96	26.82
DRIVER	33.13	8.69	10.60
PASSENGER	25.13	49.34	62.58
TOTAL	100.00	100.00	100.00

See the figure below for the plot of column points for the first and second dimensions from the correspondence analysis results. A table containing the unstandardized column coordinates which resulted from the analysis is also provided. The coordinates are calculated based on the column profile matrix (table of relative column frequencies). They are calculated to maximize the differences between the points with respect to the column profiles (column percentages).

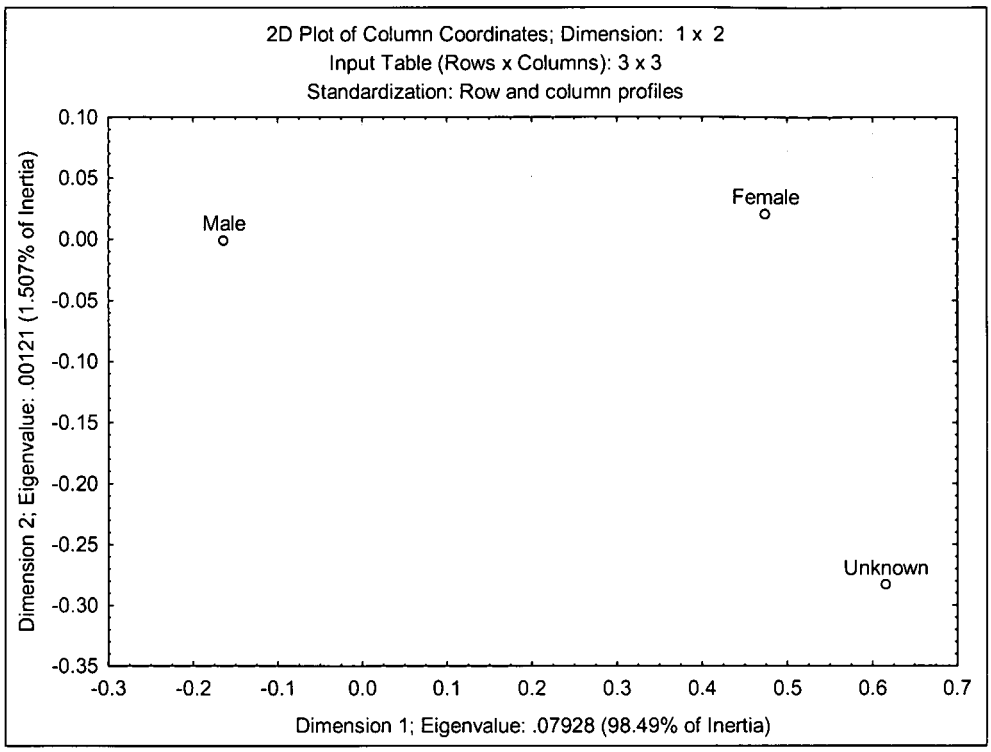

Fig. 2.4.2: Two-dimensional Plot of the Column Coordinates of the Illustrated Example

Table 2.15: Unstandardized Column Coordinates resulting from a typical Correspondence Analysis

Column Name	Coordin. Dim.1	Coordin. Dim.2
Male	-0.163871	-0.001166
Female	0.474465	0.020033
Unknown	0.615720	-0.283635

It appears that the first dimension distinguishes particularly between category *Male* and the others (see Fig. 2.4.2). Thus one can interpret the greater similarity of *Pedestrian* with *Passenger* with regard to their position on the first axis, as mostly deriving from the relatively large numbers of *Male* road users in these two groups of road user types.

It is customary to simultaneously plot the column points and the row points in a single graph, to summarize the information contained in a two-way table. What is important to remember, is that only the distances between row points and the distances between column points can be interpreted and not the distances between row and column points. The methods behind determining the coordinate system is beyond the scope of this discussion and no further detail will be given in this regard. For now it is

sufficient to take note that it is the way the coordinate system is determined, which dictates that the distances between row and column points cannot be interpreted.

It would not be appropriate to say that the category *Male* is similar to *Pedestrian* (the two points are relatively close to each other; see the plot of row and column points, Fig. 2.4.3). However, it is appropriate to make general statements about the nature of the dimensions, based on which side of the origin particular points fall. *Male* is the only column point on the left side of the origin for the first axis, and the road user types *Pedestrian* and *Driver* also fall onto that side of the first axis. It may be concluded that the first axis separates *Male* road users from the other road user categories, and that, say, *Driver* is different from, for example, *Passenger*, in that there are relatively more male drivers than passengers.

The figure below is a simultaneous plot of the row and column points from the discussed example.

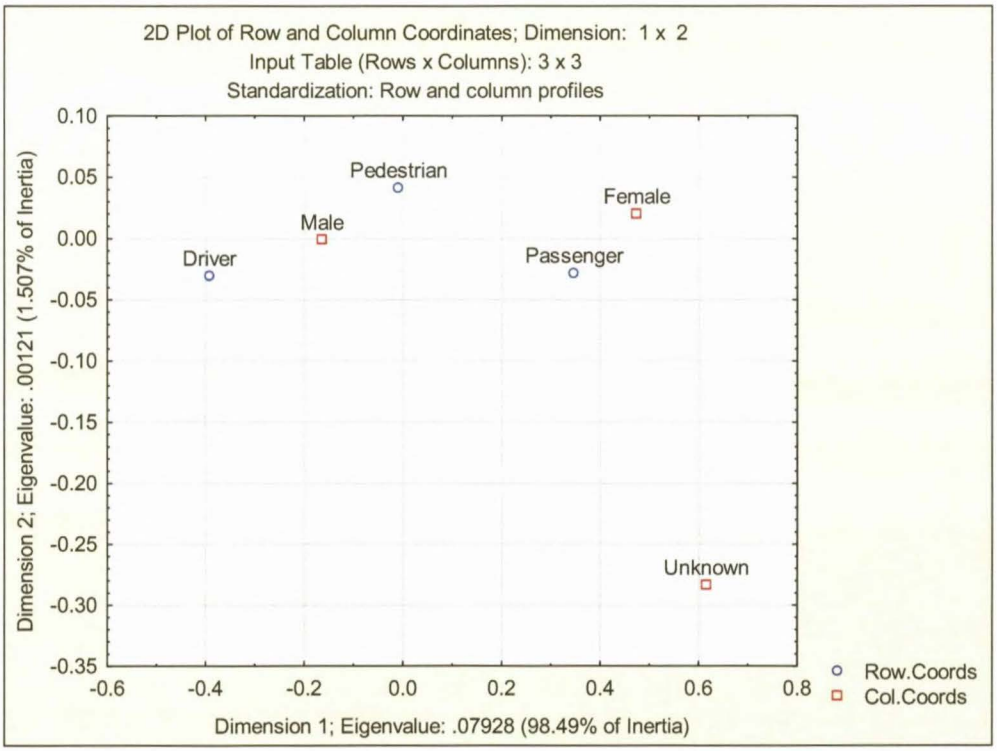

Fig. 2.4.3: Two-dimensional Plot of the Row and Column Coordinates of the Illustrated Example

ii) Correspondence Analysis Terminology

Before a discussion can follow on how the correspondence analysis output tables are interpreted, it is necessary that some terminology be explained:

- Row (Column) Mass: The row (column) totals of the relative frequency table (refer to Table 2.11).

- Inertia: The integral of *mass* times the squared distance to the centroid, as used in applied mathematics. *Inertia* is defined as the total Pearson *Chi-square* for the two-way table divided by the total sum (34686 in the example presented).

- Relative Inertia: Proportion of the contribution of a point to the overall inertia *(chi-square).*

- Relative Inertia for each dimension: The proportion of the contribution of a point to the overall inertia as contributed per dimension.

- Maximum number of dimensions (maximum extracted eigenvalues): There are only [no. of columns – 1] independent entries in each row and [no. of rows – 1] independent entries in each column of the table (once these entries are known, the rest can be filled in based on the knowledge of the column and row marginal totals). The maximum number of eigenvalues that can, be extracted from a two-way table is equal to the minimum of the number of columns minus 1, and the number of rows minus 1. If the maximum number of dimensions is extracted, all information contained in the table can be reproduced exactly. The dimensions are "extracted" so as to maximize the distances between the row or column points, and successive dimensions (which are independent of or orthogonal to each other) will "explain" less and less of the overall *Chi-square* value (and, thus, the *inertia*).

- Quality: A percentage value which is an evaluation of how properly at least most row and column points are represented by the solution. It's a measure of how satisfactory the distance of one point to the others is approximated.

- Cosine^2 : The quality representation of each point per dimension. The total of the quality values from all the dimensions must sum to the overall quality of a particular points representation by the analysis.

iii) Interpretation of Correspondence Analysis Output

After performing a correspondence analysis on the input frequency table provided at the start of this section, the following results were obtained for illustration and are given in the table shown below (Table 2.16). The table contains the *singular values, eigenvalues, and percentages of inertia explained, cumulative percentages*, and the contribution to the overall *Chi- square*.

Table 2.16: Results of a typical Correspondence Analysis (Eigenvalues and Inertia for all Dimensions)

Eigenvalues and Inertia for all Dimensions Input Table (Rows x Columns): 3 x 3 Total Inertia = .08049 Chi²=2791.9 df=4 p=0.0000					
Nr of Dim	**Singular Values**	**Eigen-Values**	**Perc. of Inertia**	**Cumulatv Percent**	**Chi Squares**
1	0.281564	0.079278	98.49318	98.4932	2749.842
2	0.034826	0.001213	1.50682	100.0000	42.069

When looking at the results, it appears that, with a single dimension, 98.49% of the *inertia* can be "explained", that is, the relative frequency values that can be reconstructed from a single dimension can reproduce 98.49% of the total *Chi-square* value (and, thus, of the *inertia*) for this two-way table. Two dimensions allow one to explain 100%. Two dimensions are thus sufficient to represent the overall *inertia* of the input frequency table.

In order to evaluate the *quality* of the respective chosen numbers of dimensions, a number of auxiliary statistics form part of the output as an aid. As explained before, the concern is that all, or at least most, points are properly represented by the respective solution and that their distances to other points can be approximated to a satisfactory degree. Shown below (Table 2.17) are all statistics reported for the row coordinates for the example, based on a one-dimensional solution only (i.e., only one dimension is used to reconstruct the patterns of relative frequencies across the columns).

Table 2.17: Reported Statistics on the Row Coordinates for the Illustrated Example based on a one-dim. solution

	Coordin. Dim.1	Mass	Quality	Relative Inertia	Inertia Dim.1	Cosine² Dim.1
Pedestrian	-0.008695	0.415845	0.042517	0.009187	0.000397	0.042517
Driver	-0.391614	0.269590	0.993909	0.516803	0.521514	0.993909
Passenger	0.347117	0.314565	0.993409	0.474010	0.478090	0.993409

The first numeric column shown in Table 2.17 contains the coordinates, as discussed in the previous paragraphs. The way these coordinates are interpreted depends on the specific coordinate system used (the way they are standardized) as mentioned above. The number of dimensions is chosen by the user (in this case only one dimension was chosen), and coordinate values will be shown for each dimension (i.e., there will be one column with coordinate values for each dimension).

The *Mass* column of Table 2.17 contains the row totals (since these are the row coordinates) for the table of relative frequencies (i.e., for the table where each entry is the respective *mass*, as discussed earlier in this section).

The *Quality* column of Table 2.17 contains information concerning the quality of representation of the respective row point in the coordinate system defined by the respective numbers of dimensions, chosen by the user. In this example, only one dimension was chosen, and the values in the *Quality* column pertain to the quality of representation in the one-dimensional space. The goal of the correspondence analysis is to reproduce the distances between points in a low-dimensional space. If you extracted (i.e., interpreted) the maximum number of dimensions, you could reconstruct all distances exactly.

A low quality means that the current number of dimensions does not well represent the respective row (or column). In Table 2.17, the quality for the first row (*Pedestrian*) is even less than 0.1, indicating that this row point is not well represented by the one-dimensional representation of the points.

The quality measure does not indicate whether or not, and to what extent, the respective point does in fact contribute to the overall inertia (*Chi- square* value). The *relative inertia* represents the proportion of the total inertia accounted for by the respective point, and it is independent of the number of dimensions chosen by the user. Note that a particular solution may represent a point very well (high *Quality*), but the same point may not contribute much to the overall inertia (e.g., a row point with a pattern of relative frequencies across the columns that is similar to the average pattern across all rows).

The *relative inertia* column in Table 2.17 contains the relative contributions of the respective (row) points to the inertia "accounted for" by the respective dimension. Thus, this value will be reported for each (row or column) point, for each dimension.

The *cosine*^{2} column contains the *quality* for each point, by dimension. The sum of the values in these columns across the dimensions is equal to the total *quality* value (as given in the section above on CA terminology). These values may also be interpreted as the "correlations" of the respective points with the respective dimension. The term *cosine*^{2} refers to the fact that this value is also the squared cosine value of the angle the point makes with the respective dimension.

It should be noted at this point that correspondence analysis is an exploratory technique. There is no statistical significance tests customarily applied to the results of a correspondence analysis; the primary purpose of the technique is to produce a simplified (low- dimensional) representation of the information in a large frequency table (or tables with similar measures of correspondence).

iv) Multiple Correspondence Analysis (MCA)

This technique is in fact an extension of the simple correspondence analysis to more than two variables at a time. An indicator (or design) matrix is used as input where the row and column categories are entered as columns and the individual cases are entered as rows. In the example discussed above, the indicator (or design) matrix would have the form as illustrated in Table 2.18.

Table 2.18: Example of Indicator (Design) Matrix for Illustrated Example on MCA

	Road User Type			Gender		
Case nr.	Pedestrian	Driver	Passenger	Male	Female	Unknown
1	1	0	0	1	0	0
2	1	0	0	0	1	0
3	1	0	0	0	0	1
4	0	1	0	1	0	0
5	0	1	0	0	1	0
. . . .						
34684						
34685						
34686						

The example discussed previously has only two variables, but will be used to illustrate Multiple Correspondence Analysis, because of its familiarity at this point of the discussion. It is clear that the design matrix as indicated above can easily be extended to more than two variables due to the table's particular form.

When the design table is analysed in the same way as for a regular two-way frequency table (as discussed in previous paragraphs), the output statistics are interpreted in exactly the same way as for the two-way frequency table, and the statistics are based on the overall inertia of the design matrix.

An application which is particularly useful when using Multiple Correspondence Analysis and the indicator (or design) matrix, is adding supplementary column points to the design matrix to perform the equivalent of Multiple Regression for categorical variables. This is sometimes referred to as *Predictive Mapping*.

If, for example, an additional variable is added to the existing two variables included in the indicator matrix given above, namely *Seatbelt status* ("yes" or "no" with *1* indicating "yes and *0* indicating "no"), the output statistic *quality* (as explained above) for these new columns after the analysis was performed, will give an indication of how well *Seatbelt status* can be "explained" as a function of *Gender* and *Road User Type*. *Predictive Mapping* was not applied in this study due to alternative methods used which will be discussed in paragraphs to follow as well as in the next chapter.

The particular application of Correspondence Analysis for the purpose of this study will be discussed in the following chapter.

2.4.2 Association Rules

i) General Methodology

The general methodology of *Association Rules* is discussed here in terms of its applicability to the study. Other details (beyond the scope of the application of this technique for the purpose of this study) on data requirements and conditions which are necessary and applicable to this technique are not

discussed here. For further information, the reader is referred to the software package *Statistica* and its *Help* documentation.

The goal of Association Rules is to detect relationships or associations between specific values of categorical variables in large data sets. It enables analysts to uncover hidden patterns, which cannot be found from simple inspection of frequency tables or large tables of raw data. Association rules, which identify items and the co-occurrences of different items that appear with the greatest (co-)frequencies, are derived. These rules are generally of the form *If X then (likely) Y* where *X* and *Y* can be single values, items, words, etc., or conjunctions of values, items, words, etc., or, *If "Body" then "Head"*, where *Body* and *Head* stand for simple codes or text values (items), or the conjunction of codes and text values.

Cross tabulation tables, and in particular Multiple Response tables can be used to analyze data. However, in cases when the number of different items (categories) in the data is very large (and not known ahead of time), and when the "factorial degree" of important association rules is not known ahead of time, then these tabulation facilities may be too cumbersome to use, or simply not applicable.

The algorithm will determine association rules without requiring the user to specify the number of distinct categories present in the data, or any prior knowledge regarding the maximum factorial degree or complexity of the important associations.

In order to overcome the problem of generating large amounts of association rules which do not necessarily have any meaning, the user can specify limits to the output statistics so association rules are retained which (1) have a *confidence value* that is greater than some user-defined minimum confidence value, (2) have a *support value* that is greater than some user-defined minimum support value, and (3) have a *correlation value* that is greater than some minimum correlation. The analyst should beware of setting these minimum limits too high to prevent missing any meaningful associations. Any user-defined value for the output statistic is dependent on the specific context of the data and can thus be a very subjective decision.

Unless the process stops because no further associations can be found that satisfy the minimum *support*, *confidence*, and *correlation conditions*, the process could continue to build very complex association rules (e.g., *if X1 and X2 .. and X20 then Y1 and Y2 ... and Y20*). To avoid excessive

complexity, additionally, the maximum number of items in the *Body* and *Head* of the association rules can be specified.

ii) Association Rules Terminology

The terminology relevant to association rules are given below for clarity. It is important to know these terms to comprehend how Association Rules are interpreted.

- Support Value: relative frequency of the *Body* or *Head* of the rule (for this study the support value of the *Body* is given i.e. the relative frequency of the "left side" of the rule)

- Confidence Value: The conditional probability of the *Head* of the association rule, given the *Body* of the association rule.

- Correlation Value: *Support* for *Body* and *Head*, divided by the square root of the product of the S*upport* for the *Body* and the *Support* for the *Head* (for this study the correlation values are provided, but not used for interpretation purposes).

- Lift Value: A value which measures how much better the rule is for prediction than a random guess. Lift values > 1 imply a useful rule. It is a ratio indicating how much more likely an item in the *Head* can be found in the *Body* subset, than in the whole population. It is the *Confidence* value divided by the *Support Value* for the *Head.*

iii) Interpretation and Understanding of Association Rules Output

The major statistics computed for the association rules are *Support*, *Confidence*, *Correlation* and *Lift.* These statistics can be summarized in a spreadsheet, as shown in Fig. 2.4.4 below. Fig. 2.4.4 is an example of statistical output from an analysis done on sample data relevant to this study.

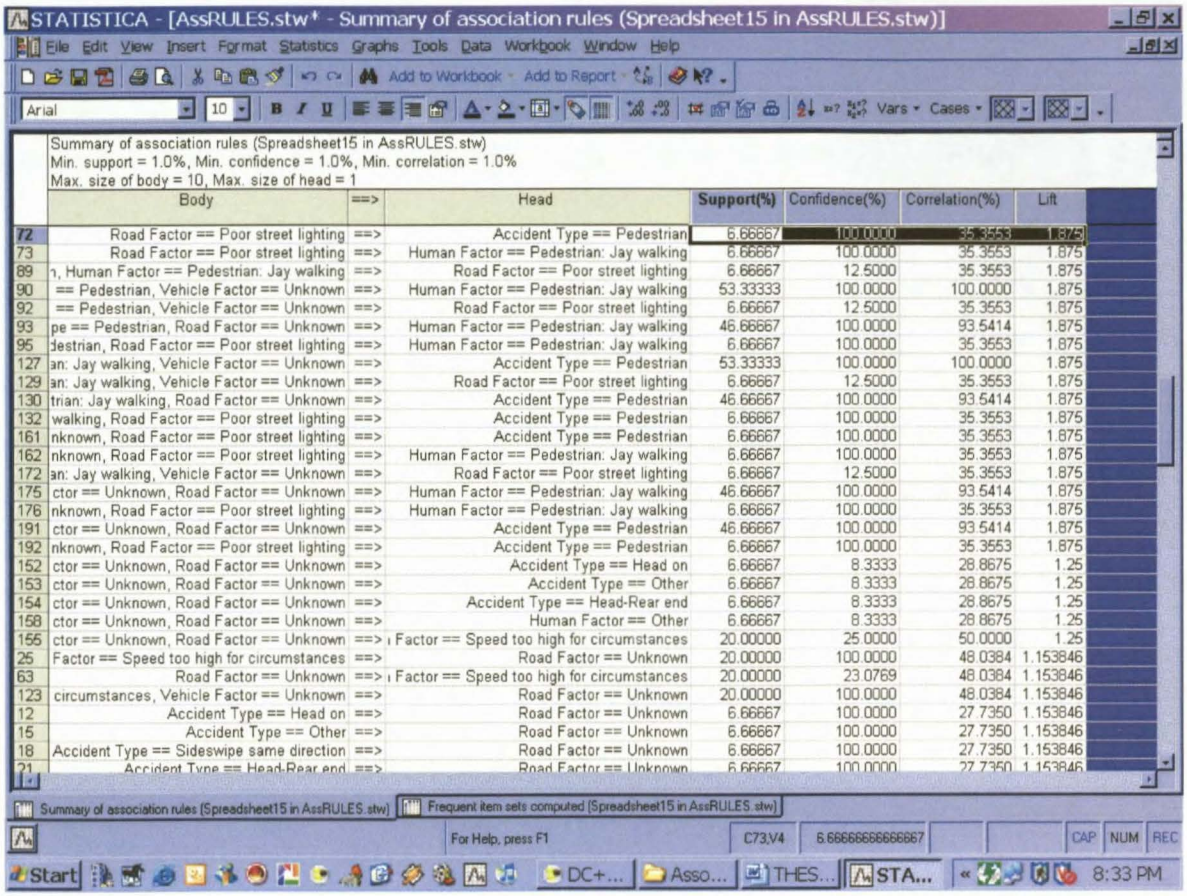

Fig. 2.4.4: Example of a Tabular Representation of Association Rules

The values for *support*, *confidence*, and *correlation* are expressed in percentage values as can be seen in the figure above. Observing the first rule visible (the highlighted row), it can be seen that there is only one item in the *Body*, namely "Road Factor = = Poor street lighting" and one item included in the *Head*, namely "Accident Type = = Pedestrian". This is interpreted as: "if poor street lighting exists, it is likely that a Pedestrian accident occurred or will occur".

The *Support* from these results suggests that poor street lighting featured only 6.67% of the time in the dataset, but the *Confidence* suggests that, given the road factor *poor street lighting*, the probability of a Pedestrian accident occurring is 100%. The *Lift* value is = 1.875 >1, indicating that this rule can most likely be used for prediction rather than only a random guess. It is 1.875 times more likely to find the accident type *Pedestrian* in the data subset for *poor street lighting* than in the whole dataset (used as input for the analysis).

When comparing the results of applying association rules to those from simple frequency or cross-tabulation tables, very high-frequency items are in some instances not part of any association rule. This can be seen from the illustrated example above. Even though *poor street lighting* featured only 6.67% of the time, an association was found between *poor street lighting* and *pedestrian* accidents. If only a cross tabulation table of road factors against accident types were inspected, this association may well have gone unnoticed.

The application of Association Rules in the context of this study is further discussed in Chapter 3.

2.4.3 Accident Rates, Exposure and Chance Variation

This section is included at this stage, to discuss accident rates and their application in identifying hazardous road locations (HRL's).

i) Hazardous road location (HRL) identification via Exposure Measures

Accident rates are calculated for the identification of hazardous sites, routes or areas (Ogden, K.W., 1996). Accident frequencies are normalized by some measure which is intended, directly or indirectly, to account for exposure. Accounting for exposure to risk of a road traffic accident is just one of the major theoretical and practical problems safety analysts have to face. "Exposure" is a relatively simple concept: the more a person is involved in traffic (e.g. the amount of distance travelled) the greater the chance that the person will be involved in an accident. Road users are involved in traffic in many different ways and this points to the need for a meaningful basis of assessing the relative safety of a system. This becomes evident when comparisons are made, for example, between different road user types or across different timeframes etc.

Different exposure measures include the following:

- Accidents per km: The longer the length of a road section, the more accidents to be expected.

- Accidents per 10^8 veh-km travelled: If traffic flow data is available, this is a better measure and is commonly used to define exposure. Typically the total traffic flow expressed as the average annual daily traffic (AADT) is used as a measure. The rate is expressed as the annual accidents per vehicle

kilometre (AADT x 365 x length of section), usually expressed as accidents per 10^8 vehicle km of travel.

- Accident severity: Accidents are sometimes stratified by severity, where the severity of an accident is based upon the most severe injury sustained by any person involved in the accident. This type of classification may be used to attempt to identify sites having a high number and/or high rate of serious accidents.

  Weighing each accident with a weight representing the average cost of accident in a severity category in which it falls, leads to fatal accidents dominating the identification procedure. Concentration on fatal accidents alone may lead to the selection of sites which do not in fact have a high accident risk. Also, circumstances which lead to fatal accidents may be very similar to those which produce injury accidents, the severity outcome being a matter of chance.

- Time period: This refers to the amount of historical accident data which should be used in order to identify a HRL. Factors which influence the choice of time period include (1) attempts to avoid environmental (e.g. traffic growth) and other trends affecting the results, (2) the use of annual accident count data to avoid cyclic or seasonal variations in accident occurrence, (3) computer storage and processing costs, and (4) changes in data base definitions introducing discontinuities in the data.

  Time periods rarely exceed 5 years in practice. 3 years are the most common time period, but 5 years are more suitable in terms of the viewpoint of statistical reliability.

In a recent study by Tarko and Kanodia (TRB 1897, 2004), safety performance functions based on the negative binomial distribution were used to predict the typical accident frequency at a location. The methods were proposed to rank locations and to evaluate the degree of hazard at an individual location without referring to other locations. These methods determine the evidence of hazard at all types of locations (intersections and segments), and they can use accident data from periods shorter than one year. Indices of accident frequency and cost were evaluated and found helpful for safety management that reduces the number of accidents and risk variability across a road network. For details on the safety performance functions and the accident and cost indices, please refer to the relevant literature.

ii) Chance Variation

Like any data analysis, accident data can be subjected to statistical analysis in order to distinguish between significant events and those occurring just through chance. It is important to be able to assess whether a certain number of accidents occurring on a site in a particular time period (e.g. one year) must be taken as "abnormally" high or whether the variations must be taken as mere chance variation. If it is assumed that the variation in accidents is random from year to year, the *Poisson* distribution may be used:

$$P(x) = \frac{m^x e^{-m}}{x!}$$

Where $P(x)$ = the probability of x occurrences of an event for which the expected number of occurrences is m.

If the accident history at a site for the last five years is 3, 2, 0, 3 and 7, there might be reason to be concerned, because of the sudden increase in accidents in the most recent year. It would be helpful to be able to assess how likely this result may have occurred by chance. Using the accident history to determine an estimate for m, it implies $m = 3$ accidents per year. If the estimate is determined over a longer period it may be assumed to be approximately the correct value if other influencing factors stay fixed. The following probabilities are calculated using the above formula:

x	P(x)
0	0.0498
1	0.1494
2	0.2240
3	0.2240
4	0.1680
5	0.1008
6	0.0504
7	0.0216

The probability of 7 accidents occurring in a year is 2.2% or about one in 46 and the probability of having more than 3 accidents in a year at a site is 1-{P(0) + P(1) + P(2) + P(3)} = 0.3528 or 35.3% (approximately 1 in 3).

This is a useful test when deciding whether a site is worthy of further investigation. It will give an indication as to whether apparently high accident occurrence is due to random variation. Though a

useful test, it is important to take into account the real world changes that might affect the "expected" accident frequency e.g. changes in traffic flow and other environmental factors.

2.4.4 Multiple Regression in terms of Road Safety

Making the appropriate choice of multiple regression models to be used, requires fine decision making in terms of the statistical significance of the final results as well as the context in which the technique is applied (e.g. a high R^2-value does not always indicate a good prediction model, depending on the context). What is important to be aware of and take into consideration, is that a prediction model should best fit the purpose of predicting the dependent variable in terms of the whole population of observations of the dependent variable, instead of just fitting a model to best suit a particular sample.

In terms of road safety and the variables which determine the occurrence of road accidents and depending on the chosen dependent variable, high R^2-values is unlikely, because of variables which are most of the time virtually impossible to monitor (e.g. variables based on human behaviour or incontrollable environmental factors). In this case the most *realistic* R^2-value is sought after, even if this value seems relatively low in terms of statistical significance. Instead of assuming that a relatively low R^2-value indicates a useless prediction model, it might indicate that there are in fact many other predictors which do provide useful information on the dependent variable, but which are impossible to include in the model (as explained above). This might also indicate that more effort should be made to establish methods to identify and monitor these "missing" variables.

The matter of low R^2-values in the road safety context will become clearer in Chapter 3, when the methodology of application of an appropriately selected multiple regression model for this study is given and discussed. Chapter 4 includes the discussion on the final results.

i) Multiple Regression Models

The multiple regression models available from Devore and Farnum (1999) are the following:

- General Additive Multiple Regression Model
- k^{th}-degree Polynomial Regression Model
- Multiple Regression Model with Interaction Predictors

- Multiple Regression Models with Qualitative Predictor Variables
- Nonlinear Multiple Regression Models

The models are given below with a short definition of each.

General Additive Multiple Regression Model

A dependent variable y is related to k predictor variables $x_1, x_2, \ldots, x_k$ given by the following equation:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + e$$

e is a random deviation which is assumed to be normally distributed with a mean value of 0 and a variance of σ^2 for any particular values of the predictor variables. All e's resulting from different observations is assumed to be independent of one another.

The β_i's are called the *population regression coefficients* and the function $\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + e$ is called the *population regression function*.

This type of model is widely used, because of its simplicity and will be considered first and foremost. If a *General Additive Multiple Regression* model can be fitted which provides prediction of a dependent variable at satisfactory significant levels (e.g. $p < 0.05$, or in some cases $p < 0.01$) and a satisfactory R^2-value within context, this model will be preferred to alternative models which are more complex (see the following paragraphs).

kth-degree Polynomial Regression Model

This model is a special case of the General Additive Multiple Regression model with $x_1 = x, x_2 = x^2, \ldots, x_k = x^k$:

$$y = \alpha + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k + e$$

Multiple Regression Model with Interaction Predictors

In the case where the mean change in the value of y associated with a 1-unit increase in one independent variable depends on the value of a second independent variable, there is *interaction* between these two variables. If a model has this property, a third predictor variable $x_3 = x_1x_2$ is included in addition to the predictors x_1 and x_2.

The general equation for such a model with two independent variables x_1 and x_2 is,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e \qquad \text{with} \qquad x_3 = x_1 x_2$$

More than one interaction predictor can be included where there are more than two independent variables available. Three independent variables could result in the following model with interaction predictors:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + e$$

Three-way interaction predictors can be included ($x_7 = x_1x_2x_3$), although this is rarely done in practice. A model which is frequently used with $k = 5$ and which is based in two independent variables x_1 and x_2 is the *Full Quadratic* or *Complete Second-order* model. The function has the following form:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + e$$

It is clear that a great number of models can be created from just a small number of independent variables.

Multiple Regression Models with Qualitative Predictor Variables

The models up to now have only included quantitive (numerical) predictor variables. Numerical coding makes it possible to also include qualitative (categorical) variables. Dichotomous variables (variables with just two possible categories e.g. *yes* or *no*, *male* or *female* etc.) are sometimes considered to be included. A *dummy* or *indicator* variable x is associated with any such variable and

has two possible values 0 and 1 which indicate which category is relevant for any particular observation.

Nonlinear Multiple Regression Models

An example of a *Nonlinear Multiple Regression Model* is the *Multiplicative Exponential Model*:

$$y = \beta_0 e^{\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k} \cdot \varepsilon$$

The model above is not intrinsically nonlinear as it can be linearized by a simple transformation by taking the natural logs on both sides, however by replacing the multiplication of the error factor by an addition sign for the error term will render an intrinsically nonlinear model.

ii) Model Utility

Model utility is measured according to the *coefficient of multiple determination* R^2. R^2 is the proportion of variation in the observed y values that can be attributed to (or explained by) the model relationship between y and the predictors. The closer R^2 is to 1, the more effectively the model has explained the variation in y by relating it to its predictors. Adding more predictors to a model will generally increase R^2 and cannot lead to a decrease in R^2 with every extra predictor. R^2 can be made very close to 1 when the number of predictors is close to the sample size.

High values of R^2 suggest that a model fit is useful, but in some instances this conclusion may depend on the context of the study as explained before. In general, how large should these values be before the conclusion is drawn that the model is useful?

A formal test procedure may indicate whether there exists usefulness in the relationship between y and *any* of k predictors included in the model. The *Model Utility F-Test* is used in Multiple Regression for this purpose and is based on F distributions. A large R^2 is no indication that the model will be judged useful by the *F-Test*. Software packages may include the F-statistic or simply provide the p-significance level. The p-value is the area under the corresponding F-curve to the right of the calculated F-value.

The null hypothesis (H_0) states that there exists no useful relationship between y and *any* of the k predictors. The null hypothesis is rejected if the p-value is less than or equal to the chosen significance level. Output from the multiple regression analyses performed for this study includes this p-value and will be directly interpreted according to the principles above. Calculation of F-statistics is thus not necessary and won't be discussed further.

iii) Checking Model Adequacy

Checks of model adequacy are based on the residuals and in particular various plots involving these or related quantities. The residuals are the differences between the observed and predicted y values. Each one of these residuals is subjected to randomness, before the data is even obtained. If the correct model has been fitted, the mean of any particular residual is zero. The amount of variability in any observed residual depends on the values of the predictors at which the corresponding observations are made. This makes it difficult to compare residuals. A remedy for this problem is to standardize the residuals. Standardized residuals will be provided by statistical software packages on request.

It is assumed that the random deviations (e) (included as part of the multiple regression models given above) are normally distributed. The normal probability plot of the standardized residuals is used to check this normality assumption. The expected normal value of the residual is plotted against the residual values. If the plot gives a reasonably linear relationship, the normality assumption is plausible. Fig. 2.4.5 is an example of a normal probability plot for standardized residuals. There appears to be a satisfactory linear relationship and the normality assumption is plausible. The normal probability plots of the analyses performed for this study are included as part of the output and will be interpreted according to the principles above.

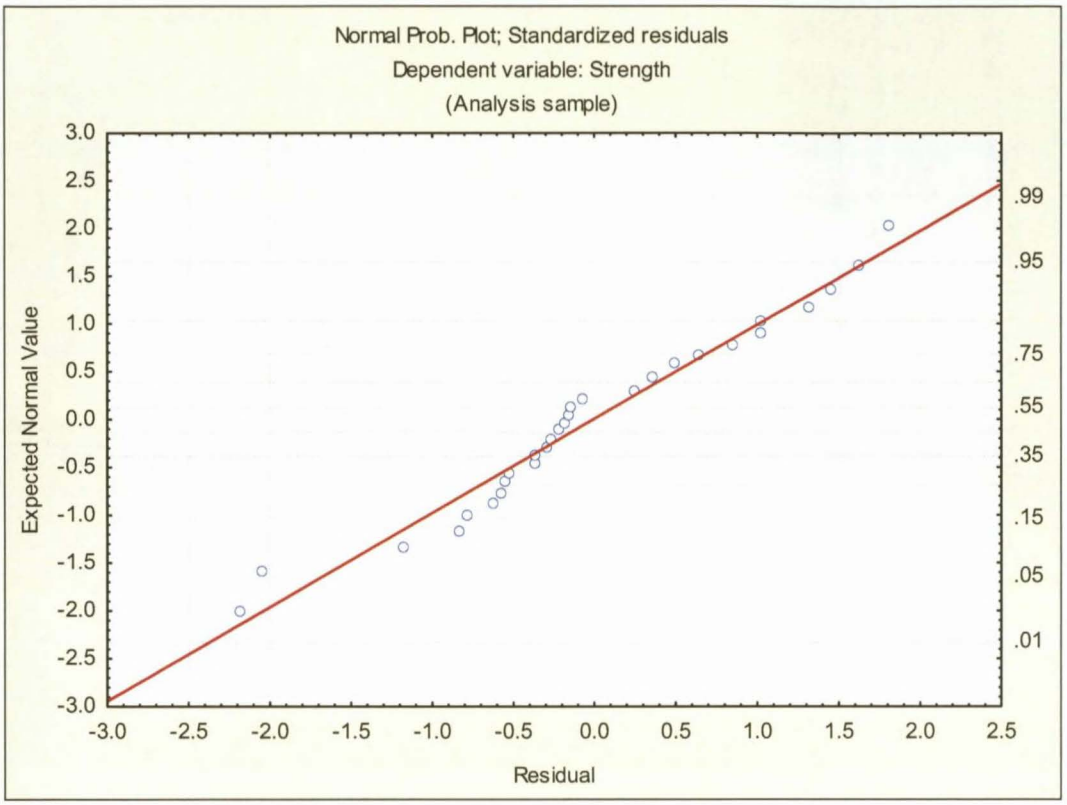

Fig. 2.4.5: Example of a Normal Probability Plot for Standardized Residuals

Another model assumption is that the variance σ^2 of a random deviation (e) is a constant and that it does not depend on the values of the predictors. A plot of the residuals versus the predicted values of the model can indicate whether this assumption is plausible. Ideally the points on the plot should appear randomly placed with no discernible pattern. Any marked deviation from randomness should indicate that remedial action is necessary. Fig. 2.4.6 is an example of the predicted values of a model plotted against the standardized residuals. There appear to be randomness in the points without any clear discernible pattern and the assumption stated above can be taken as plausible.

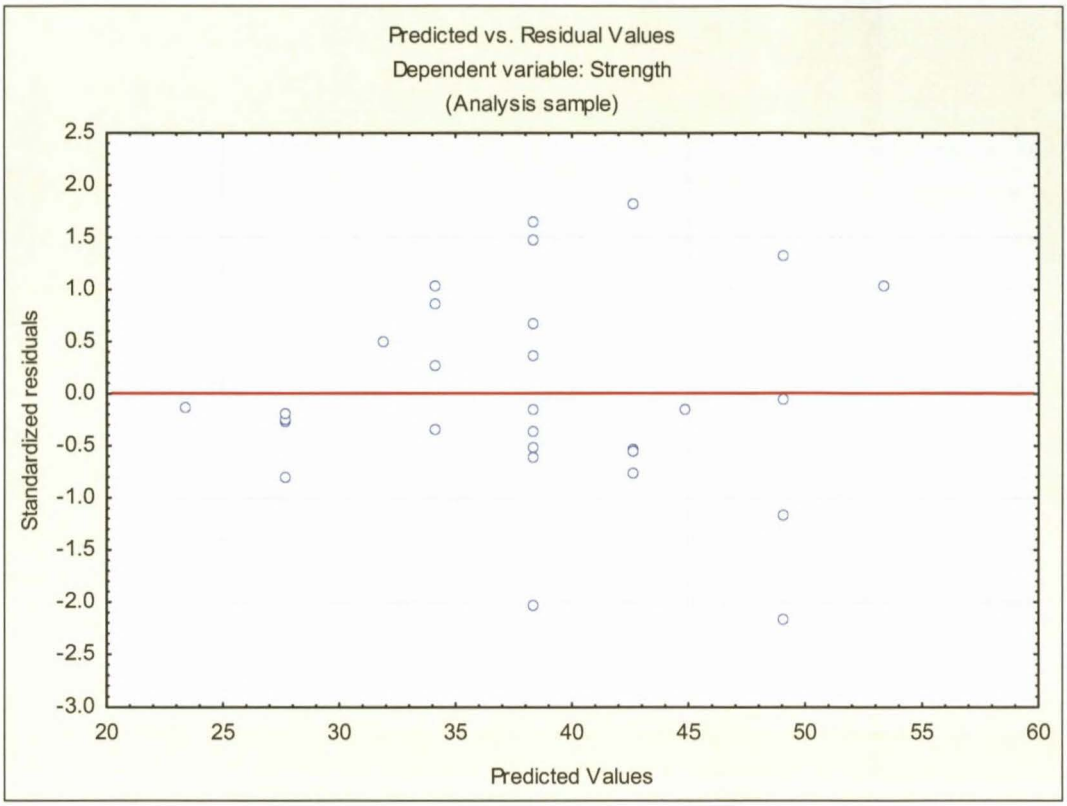

Fig. 2.4.6: The Predicted Values versus the Standardized Residual values for a Multiple Regression Model

iv) Model Selection

Frequently, some predictors are only weakly related to y or contain information that duplicates information provided by some of the other predictors. The issue is how to select a subset of predictors from the candidate pool to obtain an effective model. The most popular criteria used to select the most appropriate multiple regression model is the following:

- R^2-value: The highest value for R^2 is used as criteria with the minimum amount of predictors after which additional predictors don't significantly increase the R^2-value.

- Backward Elimination: First, all the predictors in the candidate pool is included in the model fit. The predictors are eliminated one by one until at some point all remaining predictors seem important.

- Forward Selection: Predictors are added from the candidate pool one by one until at a certain point none of the predictors not already added appears useful.

- Stepwise Regression: Predictors are added one by one with the option of deleting a predictor at some later stage that was added previously (the alternative is *backward stepwise regression* where all predictors from the candidate pool are included and are removed one by one with the option of adding a predictor at some later stage that was removed previously).

- Best Subsets: All possible models are fitted and one or more summary quantities from each fit are calculated and compared to identify the most satisfactory model.

The latter selection method namely *Best Subsets* is used for the purpose of this study and the methodology for model selection as applicable is discussed in Chapter 3.

CHAPTER 3

METHODOLOGY

3.1 Data Gathering Methodology

3.1.1 Fatal Accident Database: RTMC and Arrive Alive

i) Accident Database Preparation

In Chapter 2 it was already briefly described how the data for this study was obtained. An MS Access Database was received from Arrive Alive and the RTMC, containing a wide range of details on fatal accidents and fatalities for the years 2002, 2003 and 2004. The database was received in September 2005 after generous amounts of telephonic and email correspondence. The database was written on a CD and posted. It was sorted through to check for any significant errors and checked for data quality.

After becoming familiar with the database and MS Access as database software, the database was duplicated to overcome a few versatility problems which existed on the file received on disc (e.g. security locks on the function to create new queries). This was done by copying the existing data tables to a new MS Access file and recreating the existing queries by first setting the relevant relationships between the different data tables so the queries could be performed.

The received database was not versatile; it did not allow the ability to create new queries or tables and only allowed the user to perform some of the final queries which were already included in the database. There also existed some compatibility problems with the version of MS Access used by the author and the version of MS Access used by the original creator of the database. It was necessary to have a duplicated version to increase potential uses for the database and to increase its flexibility and versatility. The duplicated version allowed for several new queries.

The study focused on several different road sections along national roads N1, N2 and N7 within South Africa in the Western Cape Province (although all routes relevant to RSA and/or Western Cape

Province were included in application of certain statistical analysis techniques as will be seen later in this document). Geometric data, traffic data and terrain type variables relevant to the road sections under study needed to be integrated with the accident database. This was done by creating new tables in the duplicated version of the accident database for new data input. This extension of the database allowed for further queries which could easily be done once the proper relationships between the new and old data tables were set, which was a relatively easy task.

ii) Queries and Cross tabulations for Analysis

Queries were created in the MS Access accident database for the whole country's fatal accidents for South Africa as well as the Western Cape Province. These queries can be customized to include any road section as required. Significant categorical variable pairs were chosen in such a way that a clear understanding could be gained of how the variables relate to one another in terms of frequency of occurrence. Cross tabulation tables were created for summary purposes as well as for input for analysis.

The table below contains the various categorical variable pairs used for analysis.

Table 3.1: Variable pairs used for MS Access queries and cross tabulation

Road User Type vs. X	Type of Accident vs. X
➢ Road User Type vs. Gender ➢ Road User Type vs. Vehicle Type ➢ Road User Type vs. Race Group ➢ Road User Type vs. Seatbelt status	➢ Type of Accident vs. Area Type ➢ Type of Accident vs. Vehicle Type ➢ Type of Accident vs. Gender ➢ Type of Accident vs. Road User Type ➢ Type of Accident vs. Race Group ➢ Type of Accident vs. Human Factor ➢ Type of Accident vs. Road Factor ➢ Type of Accident vs. Vehicle Factor

In variable pairs where road user frequencies featured (first column of the table above), all involved road user frequencies were included in the queries (fatalities and non-fatalities), although, where it could provide significant results, the road user queries could easily be customized to include only fatalities or non-fatalities. Where the type of accident feature, all accident types were included as well as only the top 5 accident types occurring with the highest frequencies.

3.1.2 Traffic Data: Mikros Traffic Monitoring (Pty) Ltd (SANRAL CTO Yearbook)

In §2.1.1 it was explained what types of data and information are available and contained within the CTO Yearbook. In this section it will be explained how this data and information are captured and processed by Mikros Traffic Monitoring (Pty) Ltd. The author will then proceed to describe the methods used for capturing and recording the necessary traffic data from the various CTO Yearbooks used for the intended study.

i) Methodology for Traffic Data Capture used by Mikros Traffic Monitoring (Pty) Ltd

The traffic information is generally gathered by means of the Traffic Event Logger (TEL). The TEL receives signals from inductive loops installed beneath the road surface. The TEL records, with respect to every single vehicle, the exact time of departure (to the nearest tenth of a second), speed (to the nearest km/h), length (to the nearest tenth of a meter), the chassis height (low, medium, high) and the lane number.

The speed of each vehicle is calculated from the time at which the front of the vehicle arrives at the first and second loops, and the length of the vehicle is calculated from the time the second loop is occupied. Where traffic counting stations are equipped with loops only, the vehicle classification is deducted from the vehicle length and the chassis height. This allows a distinction to be made among five types of vehicles, viz, short light vehicles (cars, bakkies), long light vehicles (e.g. a car towing a caravan), short trucks (single-chassis units with two or three axles) medium trucks (articulated units consisting of a truck-tractor and semi-trailer) and long trucks (usually a truck-tractor plus a semi-trailer and full trailer combination).

The collected data are summarized every hour and stored in the TEL's memory. The data from permanent stations are, where possible, retrieved using a computer connected to the TEL by means of a telephone/modem. Where a traffic counting station is equipped with axle sensors in addition to loops, the vehicle classification is deducted from the vehicle length and the chassis height and axle spacing. This allows the additional distinction of vehicles into various class schemes, typically the Federal Highway Administration Scheme F13 or the SANRAL Vehicle Classification Scheme. Traffic stations, which have Weigh-in-Motion (WIM) sensors in addition to loops, provide the same vehicle distinction

capabilities as stations equipped with axle sensors. In this instance, the masses of axles are estimated from the dynamic masses by the WIM sensors, which are in turn added to the vehicle data recorded.

From the raw data, extensive information on traffic counts, speeds, headway's, flows, quality of service and estimated loading can be determined.

ii) Methodology for Traffic Data Capture from SANRAL's CTO Yearbooks

As discussed before, this study focuses on road sections along the N1, N2 and N7 in the Western Cape. Each of the national roads was divided into smaller road sections. In order to conduct the study, traffic data variables had to be captured based on the road sections as governed by the fatal road accident database. Therefore, one value for each traffic data variable per road section was necessary to make integration with the accident database possible.

The different road sections studied are given in the table below (as indicated according to the fatal road accident database):

Table 3.2: Table of Road Sections under Study along the N1, N2 and N7 (Western Cape)

N1	N2	N7
➢ Cape Town – Goodwood ➢ Goodwood – Bellville ➢ Bellville – Paarl ➢ Paarl – Worcester ➢ Worcester – Touwsriver ➢ Touwsriver – Laingsburg ➢ Laingsburg – Leeuw Gamka ➢ Leeuw Gamka – Beaufort West ➢ Beaufort West – Three Sisters	➢ Cape Town – Somerset West ➢ Somerset West – Strand ➢ Strand – Grabouw ➢ Grabouw – Caledon ➢ Caledon – Riviersonderend ➢ Riversonderend – Swellendam ➢ Swellendam – Riversdale ➢ Riversdale – Mosselbay ➢ Mosselbay – Hartenbos ➢ Hartenbos – George ➢ George – Sedgefield ➢ Sedgefield – Knysna ➢ Knysna – Plettenberg Bay	➢ Cape Town – Goodwood ➢ Goodwood – Malmesbury ➢ Malmesbury – Piketberg ➢ Piketberg – Citrusdal ➢ Citrusdal – Clanwilliam ➢ Clanwilliam – VanRhynsdorp ➢ VanRhynsdorp – Bitterfontein
*The road section Somerset West – Strand is omitted in any analyses performed for the purpose of this study due to an inappropriate level of detail regarding the specific route description		

For each road section, various traffic data variables were manually captured by browsing the SANRAL CTO Yearbooks and gathering the data from the CTO counting stations applicable to each road section. The CTO counting stations applicable to each road section were determined by inspecting a map of the

Western Cape indicating the positions of the CTO stations. Appendix A4 contains maps on each road section illustrating the relative positions of the applicable CTO counting stations. The yearbooks are in Adobe Acrobat format and all data and information had to be manually copied from all the relevant CTO station records to Excel spreadsheets to be processed.

Each road section was categorized according to the amount of CTO stations available and each route's complexity for data gathering. The relevant categories can be found in Table 3.3 below. Most CTO stations were located in urban areas and it was mostly these road sections located in urban areas which were complex to analyse for their traffic volume and speed data. Some rural road sections either had only one CTO station available from which data could directly be taken or had more than one station, but which proved to provide no challenge in gathering the necessary data (i.e. averages could be taken). Two road sections had no CTO counting station report available and for each the preceding road section's data were used.

Table 3.3: Road section categories for traffic and speed data gathering

one station	>one station (simple)	>one station (complex)	no information available
➢ Touwsriver – Laingsburg ➢ Leeuw Gamka – Beaufort West ➢ Beaufort West – Three Sisters ➢ Grabouw – Caledon ➢ Caledon – Riviersonderend ➢ Riversonderend – Swellendam ➢ Mosselbay – Hartenbos ➢ Knysna – Plettenberg Bay ➢ Citrusdal – Clanwilliam ➢ Clanwilliam – VanRhynsdorp ➢ VanRhynsdorp – Bitterfontein	➢ Paarl – Worcester ➢ Worcester – Touwsriver ➢ Beaufort West – Three Sisters ➢ Strand – Grabouw ➢ Swellendam – Riversdale ➢ Riversdale – Mosselbay ➢ Sedgefield – Knysna ➢ Malmesbury – Piketberg	➢ Cape Town – Goodwood (N1) ➢ Goodwood – Bellville ➢ Bellville – Paarl ➢ Paarl – Worcester ➢ Cape Town – Somerset West ➢ Hartenbos – George ➢ George – Sedgefield ➢ Cape Town – Goodwood (N7) ➢ Goodwood – Malmesbury	➢ Laingsburg – Leeuw Gamka ➢ Piketberg – Citrusdal

Some of the counting stations applicable were lacking traffic counts for one, and sometimes two, out of the three years included in the timeframe of the accident dataset. Some stations only had traffic counts available for, say, 2002. In the latter case, traffic counts were then adjusted by a growth factor of 3% to obtain estimates for 2003 and 2004. The same principle was applied wherever traffic counts did not exist for a particular counting station for a particular year. A traffic growth factor of 3% is a typical growth factor for traffic in South Africa and was assumed to be applicable for all types of traffic.

Processing the data from each CTO station was a cumbersome task, because in most cases more that one CTO station was available per road section. Averages (or weighted averages in some instances) had to be taken where appropriate, or trend lines had to be fitted to scatter plots created from the data along certain road sections to help with visualizing the variation of traffic data for a particular road section.

For each road section, special care needed to be taken of how the correct data value was calculated in order to get the most accurate result for that specific route. Even though the same basic method applied (as discussed above) for each road section, each dataset was inspected individually to identify possible limitations to the data or other obstacles which could have an influence on the final estimate. These obstacles then had to be addressed in the appropriate manner as dictated by the specific road section.

The processed data were used as input for the additional data tables created in the MS Access database, as discussed in previous paragraphs. The various traffic volume and speed variables captured by means of above discussed method are given in the table below:

Table 3.4: Traffic Volume and Speed Data Variables captured from SANRAL CTO Yearbooks

Traffic Volume Data Variables	Traffic Speed Data Variables
➢ Average Daily Traffic (ADT) (veh/d) ➢ Average Daily Truck Traffic (ADTT) (veh/d) ➢ Percentage Vehicles in flows over 600 veh/h (%) ➢ Percentage Vehicles over the speed limit (%)	➢ Average Speed (km/h) ➢ Average Night Speed (km/h) ➢ Average Light Vehicle Speed (km/h) ➢ Average Heavy Vehicle Speed (km/h)

See Appendix A2 for an example of a typical CTO station record as it is published in the SANRAL CTO Yearbook. Note from the station record that not all traffic variables recorded by such a station were captured for this study. Only a few significant variables were selected.

iii) General Limitations and Problems Encountered during Traffic Data Capture Process

The following obstacles and data limitations were encountered during the process of gathering traffic data for the purpose of this study:

- The exact chain distances (kilometre distances as specified by the Provincial Administration of the Western Cape Province) where CTO counting stations are located along the road sections are not indicated in a consistent manner, making the data gathering process complicated in terms of estimating an overall value for any traffic characteristic on a particular route. The chain distance was the only "weighing factor" available for calculating weighted averages.

- It should be clear from the methodology explanations in previous paragraphs that most of the estimates were made according to the author's insights and subjective opinion of what the "most accurate" estimate is supposed to be.

- The estimates were biased, due to unequal amounts of CTO counting stations available on location along each road section. This dictated the way how different types of counting stations were investigated.

- Different types of counting stations were pooled together and their traffic counts and speed data were used collectively to determine estimates for the traffic and speed variables for a particular road section. This has an influence on the final estimates, because each type of counting station gathers data for different time periods. Permanent counting stations have data available for at least a whole year at a time in comparison to Secondary stations which are not used on a continual basis throughout the whole of a year. The permanent stations' information can thus be considered more reliable in terms of determining yearly estimates.

3.1.3 Geometric and Terrain Information: SANRAL

An effort was made to gather geometric and terrain information for the road sections under study by inspecting a collection of road logs which were electronically downloaded from the PAWC website and their road network information reports. Road logs and strip charts could be downloaded for the N1 and N2 and none for the N7. Geometric and terrain information could thus not be obtained for the N7. Time constraints prohibited the search for another suitable data source for finding information on the N7 route.

Road numbers (e.g. NR00101) were used as input and queries were performed to obtain the road logs on the relevant road sections. Fortunately the road sections along the N1 and N2 in the Western Cape

on which road logs and strip charts were based, were very similar to the road sections as given in the fatal accident database used for this study. This made the inspection of these road logs a relative simple task to perform. Where overlapping of road sections occurred between those given in the road logs and those given by the fatal accident database, chain distances were inspected to subdivide the datasets appropriately to distinguish between each road section as given in the fatal road accident database.

Each dataset was inspected individually to gather information for each road section. Collected information was then integrated with the fatal road accident database by entering the information into additional data tables which were created for this purpose as explained in Chapter 2. Road section lengths were obtained from the *New Southern African Book of the Road, AA* and also entered into the fatal road accident database.

Effort was made to capture the following variables:

- Terrain type (i.e. Flat, Rolling or Mountainous)
- Lane widths
- Shoulder widths
- Number of lanes (Left and Right)

More on the results of this effort will be discussed in Chapter 4.

3.2 Data Analysis Methodology

The four analysis techniques applied to the available fatal road accident data (as discussed theoretically in Chapter 2) and motivations for using each technique are given below. The methodology for each technique as applied for this study is discussed in the paragraphs to follow.

- Correspondence Analysis: to provide visual representations of large two-way frequency tables, for the easy interpretation and understanding of the correspondences between categorical variables.

- Association Rules: to find the co-occurrence frequencies of particular combinations of categorical variables, hidden patterns between variables not necessarily occurring with the highest frequencies and to find the most meaningful association rules to possibly predict certain variable occurrences.
- Chance Variation: to determined whether statistically significant year-to-year changes in fatal accident rates occurred.
- Multiple Regression: to provide potential prediction models for the prediction of fatal road accident rates.

3.2.1 Correspondence Analysis

Correspondence analysis is an exploratory/descriptive technique which is useful for studying large two-way frequency tables. It is a useful technique for graphically representing the information contained in large two-way frequency tables. This provides a convenient visual interpretation of the correspondence between variables. In order to correctly interpret the graphical representations of the results, one needs to have a thorough understanding of the technique to prevent drawing the wrong conclusions from the analysis output.

Using *Statistica*, Correspondence Analysis was performed on the chosen variable pairs previously given in this chapter. Each variable pair's dataset was analysed for the period 1 Jan. 2002 – 31 Dec. 2004. Datasets featuring the variable *Accident Type* only included the accident types occurring with the 5 highest frequencies in the fatal accident database although the relative frequency tables containing all featuring accident types are provided as part of the analysis output in Appendix B1 of this document.

The accident database included all road user frequencies featuring in the accident database (fatal or not fatal). For the purpose of correspondence analysis, only fatalities were included where the road user type variable featured. Also, the analyses were based on fatal road accident data available for all road sections, as specified in the accident database, for the RSA and Western Cape Province, and not only the N1, N2 and N7 routes in the Western Cape Province.

The number of dimensions inspected for a particular solution was determined by using the *Percentage of Inertia* and *Cumulative Percentage* outputs as criteria (see Chapter 2). These outputs indicate how

much of the *overall inertia* (or total chi-square) of a particular frequency table is reproduced by a particular number of dimensions.

For each solution, the smallest number of dimensions were chosen which would provide a significant amount of reproduction of the frequency table's overall inertia, but which would satisfactorily simplify each solution. It was explained in Chapter 2 how the purpose of this technique is to simplify the interpretation of a large frequency table for easy investigation of the correspondence between categorical variables, by choosing the smallest amount of dimensions as possible for a solution.

The number of dimensions chosen for an analysis determined the number of graphical representations which could be created from the output. Where the number of dimensions equalled or exceeded three, one, or more than one, three-dimensional graph could be obtained. One, or more than one, one-dimensional or two-dimensional graph could also be created.

Different combinations of dimensions are used to produce a number of graphical representations. This is why the smallest number of dimensions is chosen for a solution; otherwise the output produces a bulk number of graphical representations which, in the end, defies the main goal of the analysis technique, namely to simplify the solution.

All the plots can be used for interpretation of the results, but not all the plots have reliable information. As the number of dimensions increases, the amount of overall inertia reproduced decreases and the plots representing the inertia reproduced by the higher dimensions become relatively less reliable. In general, the first dimension produces the largest percentage of the overall inertia of a frequency table. The graph which has this first dimension included as an axis in the display will thus be the most reliable in terms of the relative significance of the results.

Output generated for each variable pair from *Statistica* for the particular time period included the following:

- Row coordinates and contributions to inertia
- Column coordinates and contributions to inertia
- Eigenvalues and inertia for all dimensions
- Matrix of relative frequencies and Cross tabulation tables

- Plot(s) of row and column coordinates for dimension(s)

Relative row and column frequency tables were also created for each variable pair, in order to verify the accuracy with which the graphical representations were interpreted. It was explained in Chapter 2 how the frequency patterns between row and column points can be indicated by the relative row and column frequency tables respectively. These tables may thus be interpreted together with the plots on each variable pair, but should not (for practical reasons) necessarily be taken as the main source for output interpretation, depending on the size of each relative row/column frequency table.

The application of correspondence analysis by the described methods was not road section specific (as was mentioned before). The correspondence analyses were applied with the aim of establishing a general overview of the correspondences between various variables featuring in fatal road accidents for South Africa and separately for the Western Cape Province and to compare how the results for the Western Cape differ from the results for the whole country (if there exists any differences). Another goal was to determine whether sensible conclusions could be drawn from the graphical output alone. In Chapter 4 the results are discussed as was interpreted from the graphical output.

No multiple correspondence analyses were performed. It was not deemed necessary, because the correspondence between more than two variables could easily be interpreted from the "two-way" analysis outputs without the application of additional analyses of calculations, even though this does require some inspection of the output.

After inspecting and summarizing the analysis output on each variable pair, the different sets of summaries were compiled to create more general summary tables containing all the results on each variable pair for easy interpretation. Refer to Chapter 4 for these summary tables.

3.2.2 Association Rules

The methodology of Association Rules will now be explained in terms of how it was applied in this study. Different subsets of data were used for analysis using this technique, each obtained from queries performed in the MS Access fatal accident database. Detailed results can be found in Appendix C and

a discussion and interpretation of the results can be found in Chapter 4. The methodology for each set of association rules (as compiled for this study) will now be discussed under the relevant subheadings.

i) Accident Type vs. Accident Factors: N1, N2 and N7, Western Cape, RSA

Different accident types (see Table 2.1) for the total timeframe of the available MS Access database (1 December 2001 – 23 July 2005), were cross tabulated against the different accident factors which featured with each fatal accident (human, vehicle and road factors). This was done for each road section under study (see Table 3.2) along the national routes in the Western Cape Province, RSA (with the exception of the road section *Somerset West- Strand* as motivated in Chapter 2). The objective was to find the co-occurrences of accident factors for each accident type for each road section and to see whether meaningful association rules can be derived.

29 subsets with 4 variables each (3 accident factor variables and 1 accident type variable) were used as input for analysis using the association rules data mining technique. The pre-defined minimum values for the output statistics (as explained in Chapter 2) were as follow:

- Min. Support Value = 0.01
- Min. Confidence Value = 0.01
- Min. Correlation Value = 0.01
- Max. nr of items in Head = 1
- Max. nr of items in Body = 10

The specification of these minimum values was done by setting each of the values to a relatively small number so the final number of association rules generated through these parameters could be manually filtered for meaningful rules (if any). Parameters could be assigned small values with relative safety, because of each subset having only 4 variables, which would not lead to a large collection of combinations for each type of association rule.

The maximum number of items in the *Body* and *Head* were specified as 10 and 1 respectively. Having only 4 variables in each subset meant that the association rules generated would not be too complex. The maximum number of items in the *Body* could also have been set to 4. It was necessary, though, to specify the maximum number of items in the *Head* as 1, as the purpose was to create association rules

which would potentially find hidden patterns between different combinations of accident factors and one accident type at a time.

Finally, the results were filtered by excluding association rules which contained any accident factor category in the *Head*. After deleting these rules, the rest of the results (with only one accident type in the *Head* of each rule), were sorted according to the *Lift* value so the most meaningful association rules (the rules which could more likely be used for prediction purposes) were sorted in descending order (most meaningful to least meaningful) with the *Lift*-value indicating how much more likely an item contained in the *Head* can be found in the data subset of the item in the *Body* than in the whole dataset (see Chapter 2).

ii) Accident Type vs. Vehicle Type and Terrain Type: N1, N2 and N7, Western Cape, RSA

The different accident types (see Table 2.1) which featured in fatal accidents for the total timeframe of the available MS Access database (1 December 2001 – 23 July 2005), were cross tabulated in this case, against the different vehicle type and terrain type (flat, rolling or mountainous) categories which featured with each fatal road accident. This was done collectively for all the road sections under study along the N1, N2 and N7 national routes within the Western Cape Province, RSA. The objective was to find the confidence with which an accident type could possibly be predicted in terms of the co-occurrences of certain vehicle types and terrain types along a road section.

One data subset containing the three relevant variables was used as input and the minimum values for the output statistics (as explained in the previous paragraphs) were set to the same values as for the previous analysis on accident types and accident factors. The maximum number of items in the *Body* and *Head* were also set to the same values as explained before. The same output filtering procedure was also followed as before.

iii) Accident Type vs. Vehicle Type and Area Type: All routes, Western Cape and RSA

Accident types (Table 2.1) were cross tabulated in this case, against the different vehicle type and area type (urban or rural) categories which featured with each fatal road accident. This was done collectively for all the routes within the Western Cape Province as well as for the RSA. The objective

was to find the confidence with which an accident type could possibly be predicted in terms of the co-occurrences of certain vehicle types and the area type (urban or rural).

Two data subsets containing the three relevant variables each were used as input (i.e. one data subset for Western Cape Province and one for South Africa) and the minimum values for the output statistics (as explained in the previous paragraphs) were set to the same values as for the previous analysis on accident types and accident factors. The maximum number of items in the *Body* and *Head* were also set to the same values as explained before. The same output filtering procedure was followed as for the previous analysis.

iv) Accident Type vs. Vehicle Type, Area Type and Terrain Type: N1, N2 and N7, Western Cape, RSA

Accident types were cross tabulated against the different vehicle type, area type as well as terrain type categories which featured with each fatal road accident. The terrain type variable was captured only for the road sections along the N1, N2 and N7 within the Western Cape Province and so this analysis only included these road sections. The same goal is valid as for the previous association rules analyses i.e. to find the confidence with which an accident type can be predicted in terms of the co-occurrences of certain vehicle types, area types (urban or rural) and terrain types (flat, rolling or mountainous).

Four variables were included in the input dataset and the minimum values for the output statistics were set to the same values as discussed before. The maximum number of items in the *Body* and *Head* were also set to the same values as before. Output was filtered by using the same method as discussed previously.

3.2.3 Calculation of Fatal Accident and Fatality Rates

This section will describe the methodology for calculating the fatal accident and fatality rates for the road sections under study using accident frequencies from the MS Access fatal accident database investigated for this research.

Fatal accident and fatality rates were calculated for all the road sections along the national routes N1, N2 and N7 within the Western Cape Province, RSA. Rates were calculated for 2002, 2003 and 2004,

as well as for the period 2002-2004. The objective was to find accident rates for possible use in multiple regression models as the dependent variables. Multiple Regression is discussed in the section to follow.
The following rates were calculated:

- Fatalities per fatal accident
- Fatalities per 100 million veh-km's travelled
- Fatal accidents per 100 million veh-km's travelled

Average Daily Traffic (veh/d) and each road section's distance (in km), were used to calculate the number of vehicle-kilometres travelled. The values for ADT, route distance, number of fatalities and number of fatal accidents per road section were queried from the MS Access fatal accident database for each time period under consideration. The values of ADT for each road section for the year 2003 were used as point estimates for the period 2002-2004 when fatality and fatal accident rates were calculated for this total period of three years.

Accident rates which are not route specific were obtained from the RTMC and the Department of Transport and are provided in Chapter 2 by province. The calculated fatality and fatal accident rates per route as described above will be summarized in Chapter 4.

3.2.4 Calculation of Chance Variation in Accident Occurrence

It was discussed in Chapter 2 how the Poisson distribution can be applied to assess how likely a seemingly "abnormally" high accident frequency may have occurred by chance, assuming that accidents (and in this case fatal accidents) vary randomly from year to year. For this study, the Poisson distribution was applied to determine the chance variation of a certain fatal accident rate for each road section along the N1, N2 and N7 under study within the Western Cape Province.

The number of fatal accidents per 100 million veh-km travelled was evaluated for chance variation. The chance variation of the number of accidents occurring on a site can be determined without accounting for any exposure measures, because a site is not necessarily evaluated in comparison to other sites. For the purpose of this study different road sections are compared along the national routes of the Western Cape Province. The chance variation of accident frequencies for each road section is

therefore determined by accounting for any influence any exposure measures might have on the "expected" accident frequency so the final index on which the chance variation is determined is normalized and different road sections can be compared. If the number of fatal accidents per 100 million veh-km is evaluated, exposure measures such as the road section length and traffic volume is accounted for.

After the Poisson distribution was calculated for each particular road section, the results were plotted on a separate graph for each of the national routes in the Western Cape under study. The results and discussion of the results can be found in Chapter 4.

3.2.5 Multiple Regression Model Application

The *General Additive Multiple Regression Model with Qualitative Predictor Variables* was selected and fitted to the sample data (see Appendix D1). Variables included in the sample data were:

- Route distance (in km)
- ADT (Average Daily Traffic in veh/d)
- ADTT (Average Daily Truck Traffic in veh/d)
- Average Speed (km/h)
- Average Night Speed (km/h)
- Average Light Vehicle Speed (km/h)
- Average Heavy Vehicle Speed (km/h)
- Terrain Type (Flat, Rolling or Mountainous)

The variables given above are all the continuous variables gathered from the SANRAL Yearbooks and CTO counting station records as given in Table 3.4, with the exception of the last variable (Terrain Type), which was collected from road log reports from the PAWC website and which is a categorical variable. This variable was included in the dataset by assigning a code to each category for the purpose of including it in the Multiple Regression Analysis e.g. a value of 1 is assigned to the variable *F* ("Flat terrain") when a fatal accident rate is calculated for a flat terrain and a value of 0 otherwise. The same applies for *R* ("Rolling terrain") and *M* ("Mountainous terrain").

A model was created for each of the following dependent variables using *Statistica*:

- Fatalities per 100 million veh-kms (FRate = "Fatality Rate")
- Fatal Accidents per 100 million veh-kms (FARate = "Fatal Accident Rate")
- Fatalities per Fatal Accident (FFARate = "Fatalities per Fatal Accident Rate")

The dataset consists of 28 data points (one data point for each road section under study along the national routes within the Western Cape and as specified in the fatal road accident database). Due to this relatively small number of data points available for analysis and using the method of *Best subsets* for selecting an appropriate model, a maximum number of 5 predictors were specified for each model under consideration. The choice of 5 predictors were also made for the sake of having enough predictors for a model, but also to not have a too large number of predictors which would increase R^2 to a point where the fit would be a sample fit rather than a model fit. A correlation threshold of 0.7 was specified i.e. predictors were excluded where correlations of higher than 0.7 existed between these potential predictors.

A list of the output tables and plots generated appear below:

- Summary Statistics (Including R^2)
- Regression Summary for dependent variable (including coefficients and p-significance)
- Scatter plot: Breakdown of descriptive statistics
- Summary of best subsets
- Scatter plot: Normal probability plot of residuals
- Scatter plot: Predicted vs. Residual scores
- Correlation tables

Predictors were added step-by-step and after each addition the relevant R^2 value was calculated and plotted. This resulted in a scatter plot of R^2 versus the number of predictors (see Appendix D2). Investigating this plot would indicate the minimum number of predictors after which R^2 would not increase significantly. For these particular analyses the minimum number of predictors was found to be four. R^2 increased significantly with each addition of a predictor until the number of predictors was equal to four. The fifth predictor does not lead to a significant increase in R^2 and was thus omitted.

The summary of best subsets indicates all the possible models for different numbers and combinations of predictors and different values of R^2 for each model. As was explained above, the most appropriate

subsets appeared to be those containing only 4 predictors in the model. From these subsets the one with the highest R^2 was selected.

The models and output statistics are discussed in Chapter 4 in terms of *Model Utility* and *Model Adequacy* principles discussed in Chapter 2.

CHAPTER 4

FINDINGS AND DISCUSSION

4.1 Fatal Road Accident Data

4.1.1 Fatality and Fatal Accident Rates and Frequencies, N1, N2 and N7, Western Cape Province

The diagrams provided in this section illustrate the fatality and fatal road accident frequencies and rates along the national routes N1, N2 and N7 within the Western Cape Province for a period of three years (2002-2004). The methodology of calculating these rates was discussed in Chapter 3. Comments based on the results will be given in the paragraphs to follow.

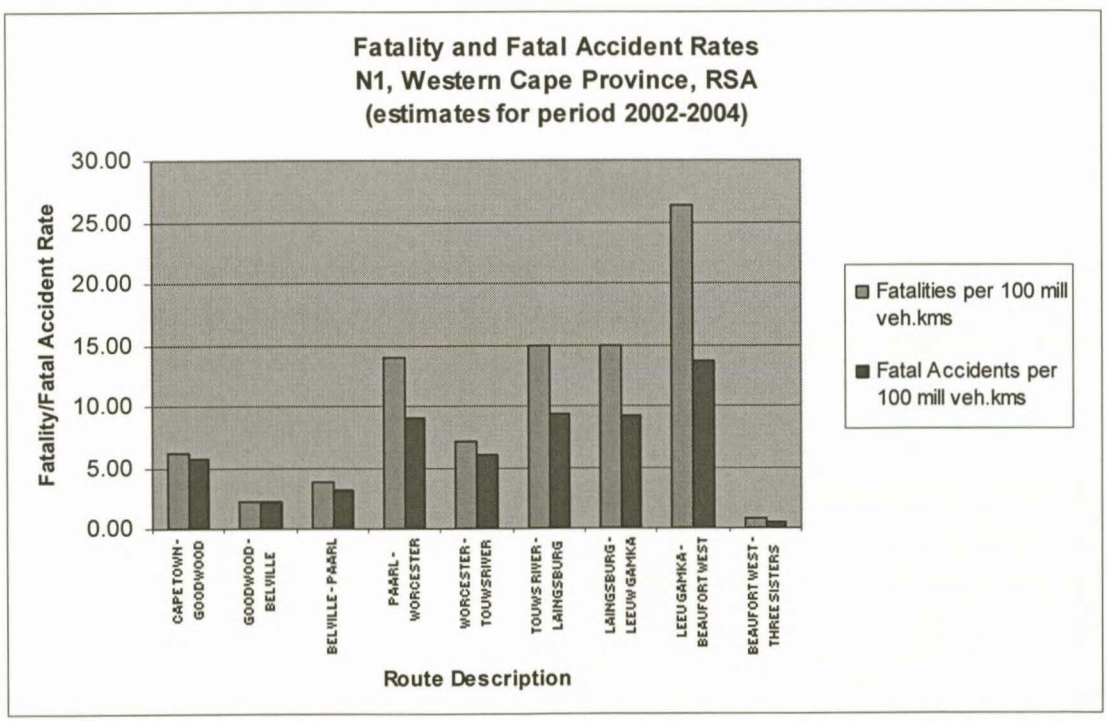

Fig. 4.1.1: Fatality and Fatal Accident Rates for N1, Western Cape Province 2002-2004

It can be seen from Fig. 4.1.1 that the road section between Leeuw Gamka and Beaufort West has the highest fatality and fatal road accident rate on the N1 in the Western Cape. Accident and fatality rates seem to have a tendency to increase as one moves farther away from Cape Town and the urban areas

and closer to more rural areas until Beaufort West is reached. Accident and fatality rates reach a peak value between Leeuw Gamka and Beaufort West after which the rates sharply decreases between Beaufort West and Three Sisters. This is possibly due to under/non reporting by police stations for the latter road section. The highest fatal road accident and fatality rates thus seem to be between Paarl and Beaufort West.

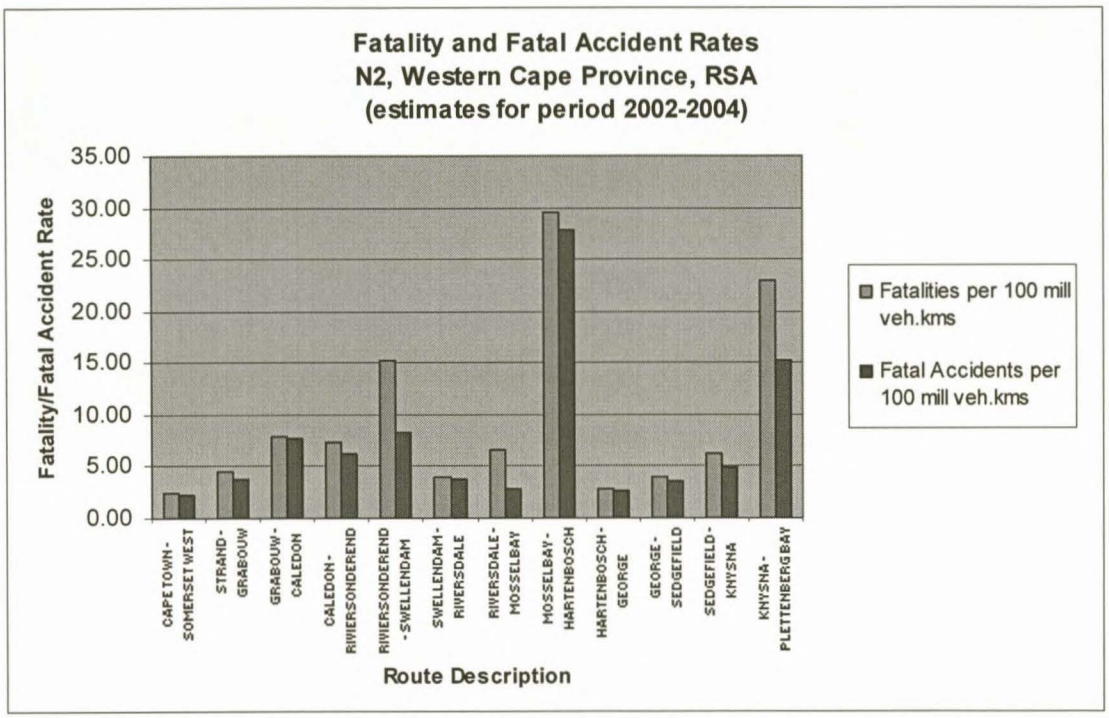

Fig. 4.1.2: Fatality and Fatal Accident Rates for N2, Western Cape Province 2002-2004

The fatal road accident and fatality rates along the N2 varies inconsistently along the route within the Western Cape. Accident and fatality rates show an overall tendency to increase from Cape Town up to Swellendam after which the rates sharply declines until Mosselbay is reached. The accident and fatality rates increases extremely sharply between Mosselbay and Hartenbos and then decreases at just as large a rate between Hartenbos and George. Rates steadily increase until Knysna and then very sharply increase again between Knysna and Plettenbergbay.

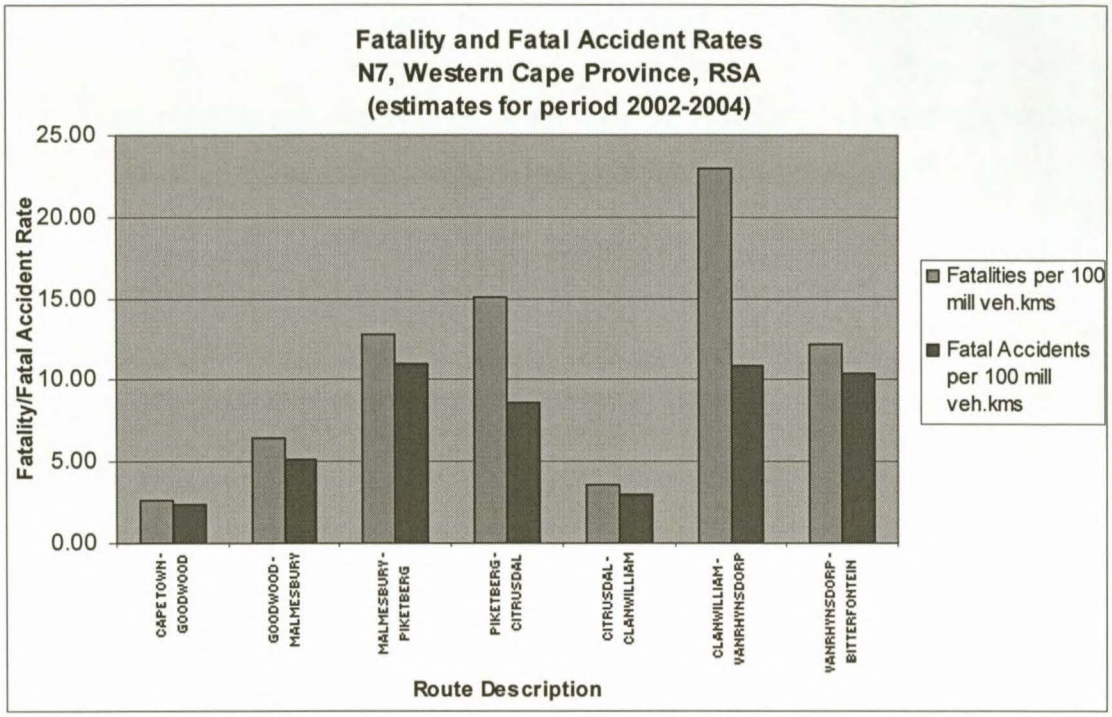

Fig. 4.1.3: Fatality and Fatal Accident Rates for N7, Western Cape Province 2002-2004

Fatal road accident and fatality rates along the N7 increases steadily from Cape Town until a peak in the rates is reached approximately between Malmesbury and Citrusdal. The rates decreases until relatively low fatal road accident and fatality rates are observed between Citrusdal and Clanwilliam (approximately the same rates as for the road section between Cape Town and Goodwood along the N7). There is a sharp increase between Clanwilliam and Vanrhynsdorp after which a decrease in rates are detected between Vanrhynsdorp and Bitterfontein which is approximately the same as the rates calculated for the road section between Malmesbury and Piketberg.

It is also concluded that fatal road accident and fatality rates are of approximately the same magnitude between the national roads N1, N2 and N7 in the Western Cape Province with the maximum rates varying between 20 and 30 fatal accidents/fatalities per 100 million veh-km. The minimum fatal accident/fatality rates vary between 0 and 5 per 100 million veh-km.

The reader is referred to Figures 4.1.4, 4.1.5 and 4.1.6 which are the graphs based on fatal road accident and fatality frequencies along the N1, N2 and N7 in the Western Cape Province as it was obtained from the database analysed for this study.

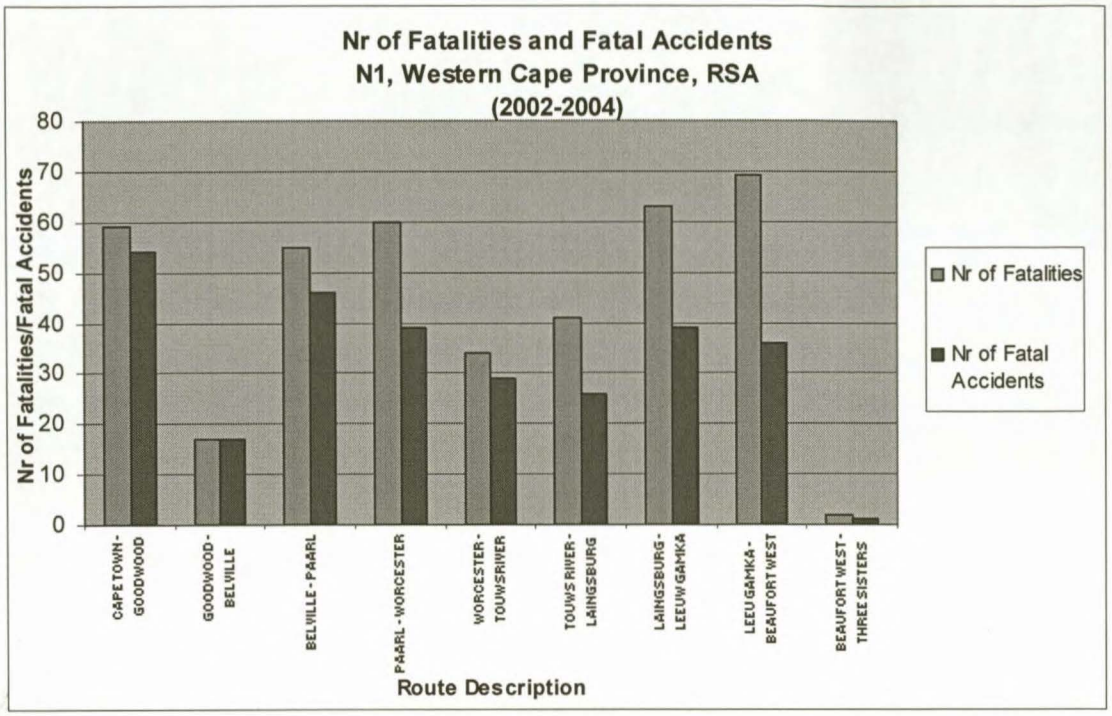

Fig. 4.1.4: Nr of Fatalities and Fatal Accidents for N1, Western Cape Province 2002-2004

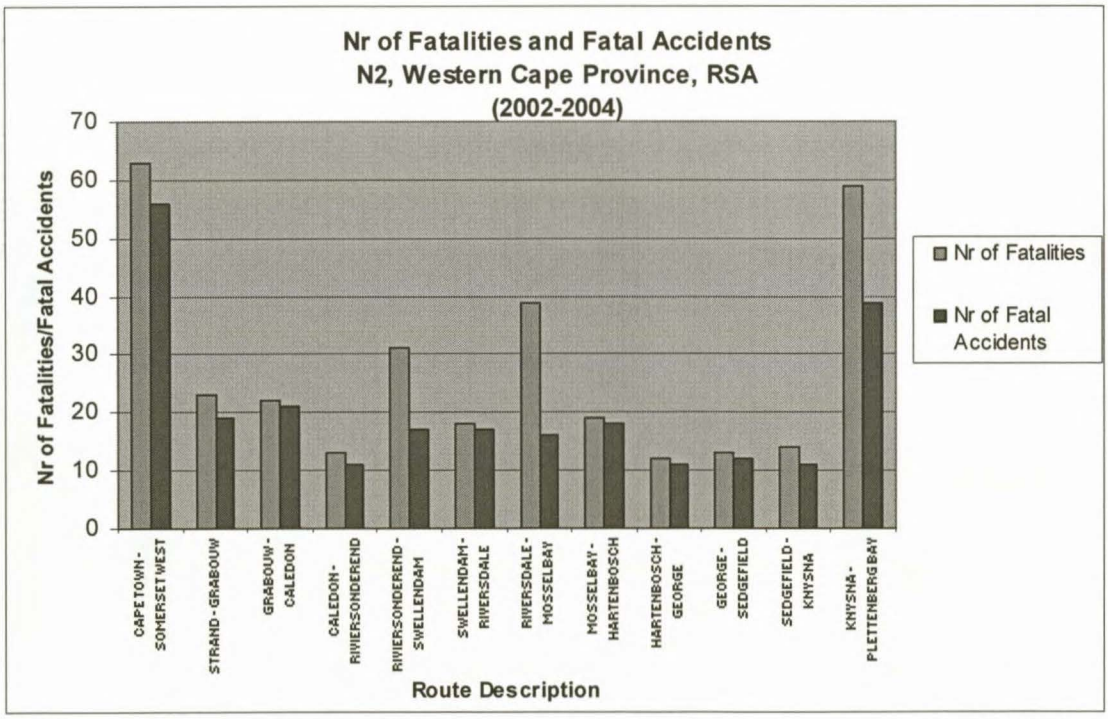

Fig. 4.1.5: Nr of Fatalities and Fatal Accidents for N2, Western Cape Province 2002-2004

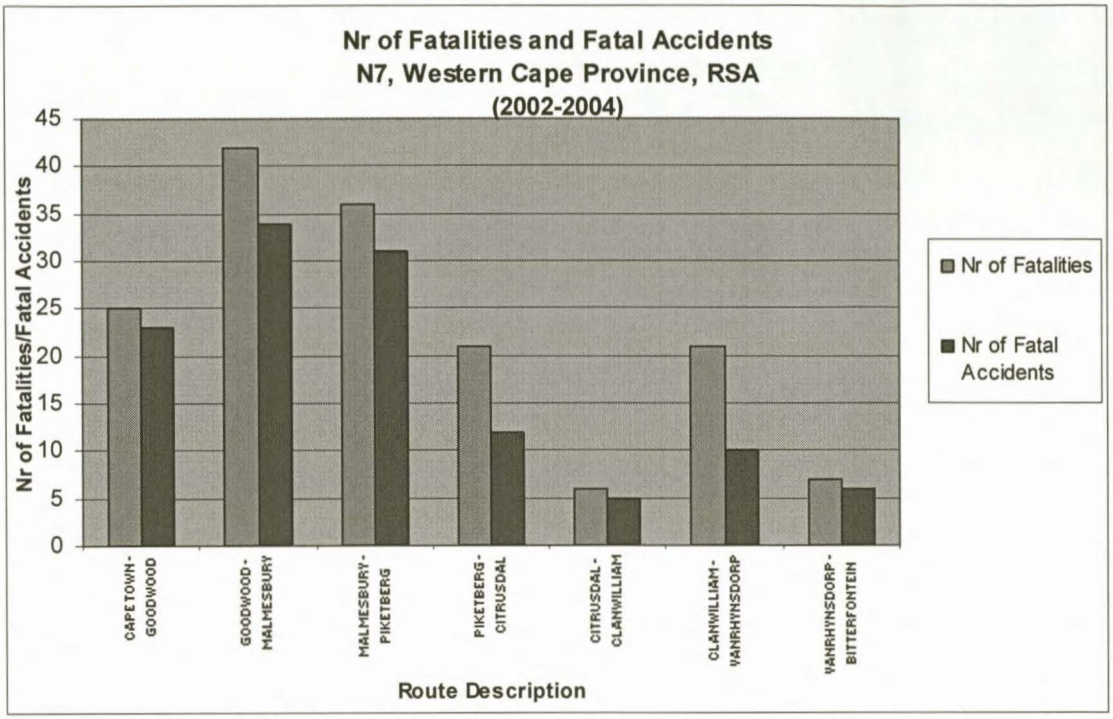

Fig. 4.1.6: Nr of Fatalities and Fatal Accidents for N7, Western Cape Province 2002-2004

It should be noted that the calculation of fatal road accident and fatality rates was based on traffic data which was manually obtained from SANRAL CTO Yearbooks as explained in Chapter 3. Any inconsistencies in any calculations done for each road section along the national roads within the Western Cape could be due to error in the subjective judgment which was used when the traffic data for each road section was obtained (the summarizing of CTO counting station data explained in Chapter 3). Under- or non reporting by police stations could also be suggested as a reason for the inconsistencies.

4.1.2 100 Million veh-km's Travelled, N1, N2 and N7, Western Cape Province

The following three figures (Figures 4.1.7, 4.1.8 and 4.1.9) illustrate the exposure measure along the national road sections in the Western Cape Province. It can be seen that the amount of veh-kms travelled on each national road section shows a tendency to decrease when progressing from Cape Town to more rural areas along the national roads. This corresponds to the increasing tendency of the fatal road accident and fatality rates along these roads due to the inversely proportional relationship between the exposure and the accident/fatality rates. The decreasing exposure measure along the roads can also be ascribed to decreasing traffic volumes along the roads (in an easterly direction from Cape Town onwards; northern direction in the case of the N7). The various road lengths also play a role.

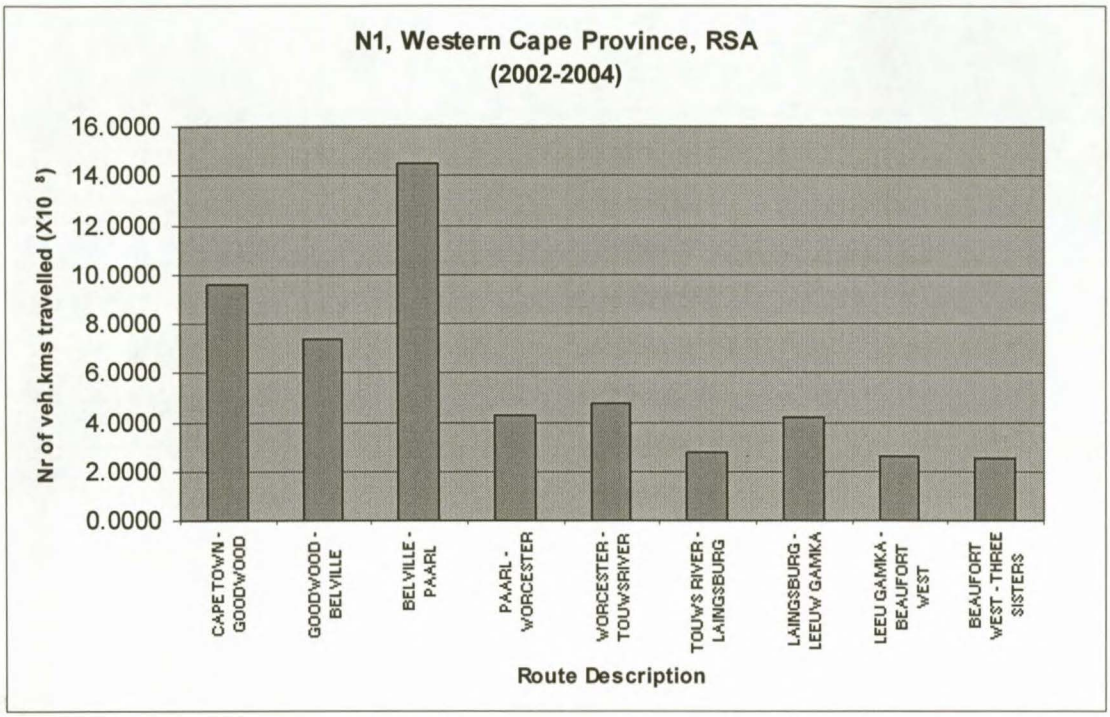

Fig. 4.1.7: 100 million Veh-km's travelled for N1, Western Cape Province 2002-2004

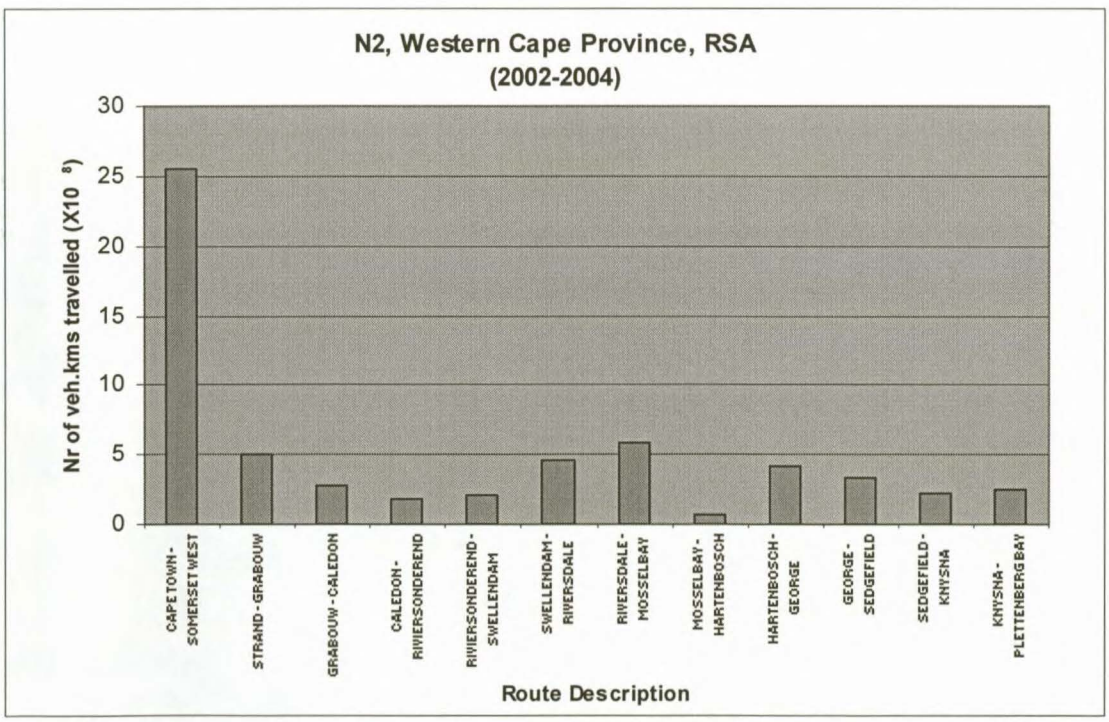

Fig. 4.1.8: 100 million Veh-km's travelled for N2, Western Cape Province 2002-2004

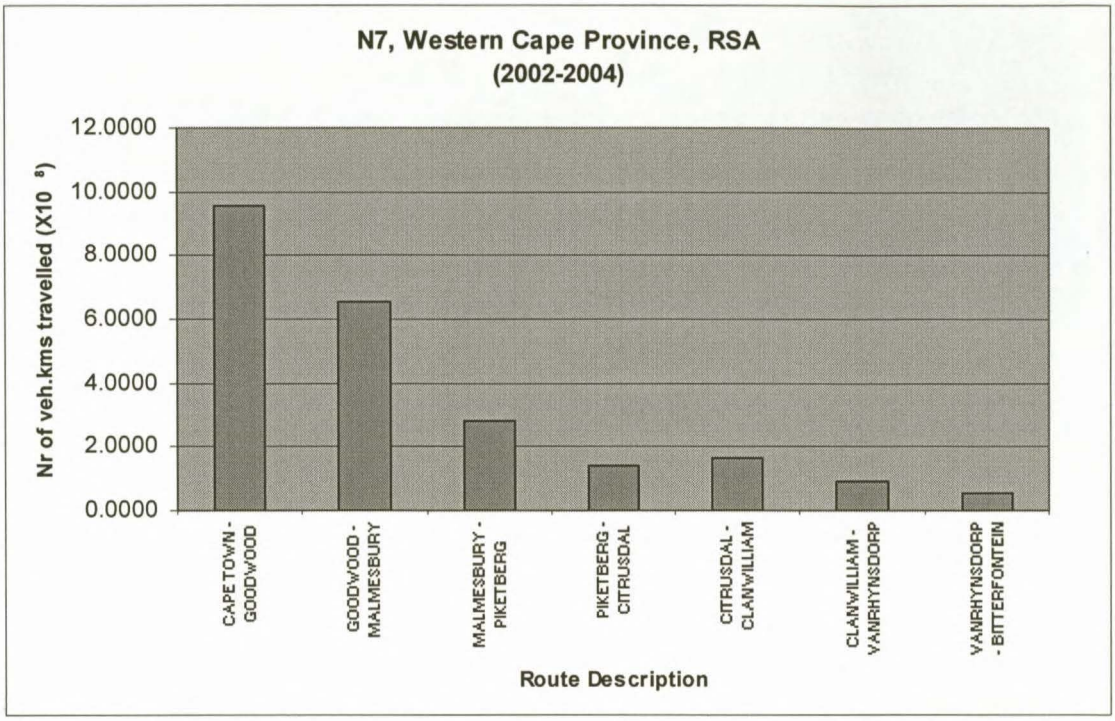

Fig. 4.1.9: 100 million Veh-km's travelled for N7, Western Cape Province 2002-2004

Table 4.1.1 contains all the data on which Figures 4.1.7, 4.1.8 and 4.1.9 are based.

4.1.3 Fatality and Fatal Crash Rates and Frequencies Data per National Road Section, Western Cape Province

See the table below for the fatality and fatal crash rates as calculated for each national road section in the Western Cape Province.

Table 4.1.1: Summary of Fatality and Fatal Crash Rates for the N1, N2 and N7, Western Cape Province, 2002-2004

		Route Description	ADT (veh/d)	Route Dist. (km)	No Fatalities	Fatal Crashes	Fatalities per Fatal Acc	100 million Veh.km's Travelled	Fatalities per 100 million veh.km's travelled	Fatal Crashes per 100 million veh.km's travelled
N1	W1	CAPE TOWN - GOODWOOD	87438	10	59	54	1.09	9.5745	6.16	5.64
N1	W2	GOODWOOD - BELVILLE	67500	10	17	17	1.00	7.3913	2.30	2.30
N1	W3	BELVILLE - PAARL	36693	36	55	46	1.20	14.4644	3.80	3.18
N1	W4	PAARL - WORCESTER	8928	44	60	39	1.54	4.3015	13.95	9.07
N1	W5	WORCESTER - TOUWSRIVER	5136	85	34	29	1.17	4.7803	7.11	6.07
N1	W6	TOUWS RIVER - LAINGSBURG	3111	81	41	26	1.58	2.7593	14.86	9.42
N1	W7	LAINGSBURG - LEEUW GAMKA	3111	124	63	39	1.62	4.2241	14.91	9.23
N1	W8	LEEU GAMKA - BEAUFORT WEST	3195	75	69	36	1.92	2.6239	26.30	13.72
N1	W9	BEAUFORT WEST - THREE SISTERS	2945	78	2	1	2.00	2.5153	0.80	0.40
N2	W1	CAPE TOWN - SOMERSET WEST	51802	45	63	56	1.13	25.5254	2.47	2.19
N2	W3	STRAND - GRABOUW	17085	27	23	19	1.21	5.0512	4.55	3.76
N2	W4	GRABOUW - CALEDON	6451	39	22	21	1.05	2.7549	7.99	7.62
N2	W5	CALEDON - RIVIERSONDEREND	2958	55	13	11	1.18	1.7815	7.30	6.17
N2	W6	RIVIERSONDEREND - SWELLENDAM	3450	54	31	17	1.82	2.0400	15.20	8.33
N2	W7	SWELLENDAM - RIVERSDALE	5015	84	18	17	1.06	4.6128	3.90	3.69
N2	W8	RIVERSDALE - MOSSELBAY	5937	90	39	16	2.44	5.8509	6.67	2.73
N2	W9	MOSSELBAY - HARTENBOSCH	8409	7	19	18	1.06	0.6445	29.48	27.93
N2	W10	HARTENBOSCH - GEORGE	8556	45	12	11	1.09	4.2160	2.85	2.61
N2	W11	GEORGE - SEDGEFIELD	8392	36	13	12	1.08	3.3081	3.93	3.63
N2	W12	SEDGEFIELD - KNYSNA	8227	25	14	11	1.27	2.2521	6.22	4.88
N2	W13	KNYSNA - PLETTENBERG BAY	6021	39	59	39	1.51	2.5713	22.95	15.17
N7	W1	CAPE TOWN - GOODWOOD	87438	10	25	23	1.09	9.5745	2.61	2.40
N7	W2	GOODWOOD - MALMESBURY	11499	52	42	34	1.24	6.5475	6.41	5.19
N7	W3	MALMESBURY - PIKETBERG	4150	62	36	31	1.16	2.8174	12.78	11.00
N7	W4	PIKETBERG - CITRUSDAL	2831	45	21	12	1.75	1.3950	15.05	8.60
N7	W5	CITRUSDAL - CLANWILLIAM	2831	54	6	5	1.20	1.6740	3.58	2.99
N7	W6	CLANWILLIAM - VANRHYNSDORP	1072	78	21	10	2.10	0.9156	22.94	10.92
N7	W7	VANRHYNSDORP - BITTERFONTEIN	782	67	7	6	1.17	0.5737	12.20	10.46
					884	**656**	**1.35**			

NOTE: 2003's ADT values used as estimates for all three years.

4.1.4 Road User Information, N1, N2 and N7, Western Cape Province

The following subsections summarize some road user information on each of the road sections along the N1, N2 and N7 for each year between 2002 and 2004. Fatalities are summarized by Road User Group, Race Group, Gender and Seatbelt status (excl. Pedestrians).

i) N1: Fatalities by Road User Group, N1, N2 and N7, Western Cape Province

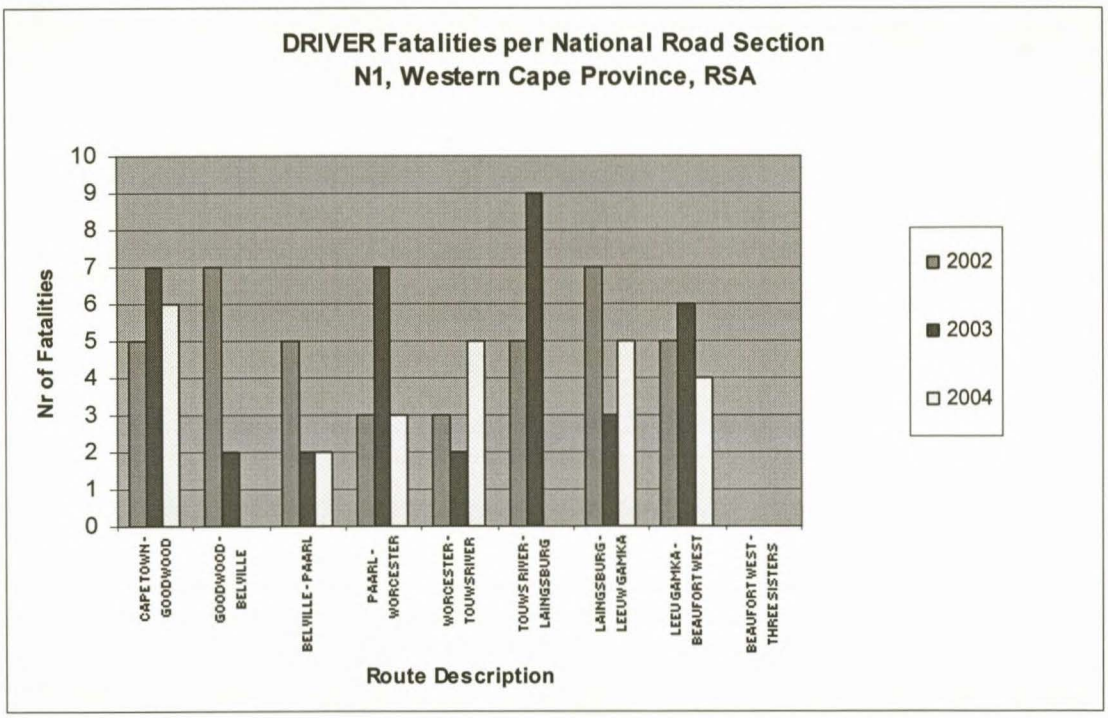

Fig. 4.1.10: Driver Fatalities per National Road Section, Western Cape Province, N1

The amount of driver fatalities along the N1 seems to vary randomly from year to year. This conclusion, of course, is only made based on three years' worth of data. No significant pattern can be discerned from Figure 4.1.10 alone though. The minimum number of driver fatalities occurred between Goodwood and Paarl and again between Worcester and Touwsriver in 2003 in the amount of 2 driver fatalities for each road section. The maximum number of driver fatalities occurred between 2002 and 2003 between Cape Town and Bellville, between Paarl and Worcester and then again between Touws river and Leeuw Gamka. The diagram illustrates and overall tendency of decrease in driver fatalities along the N1 from 2003 onwards.

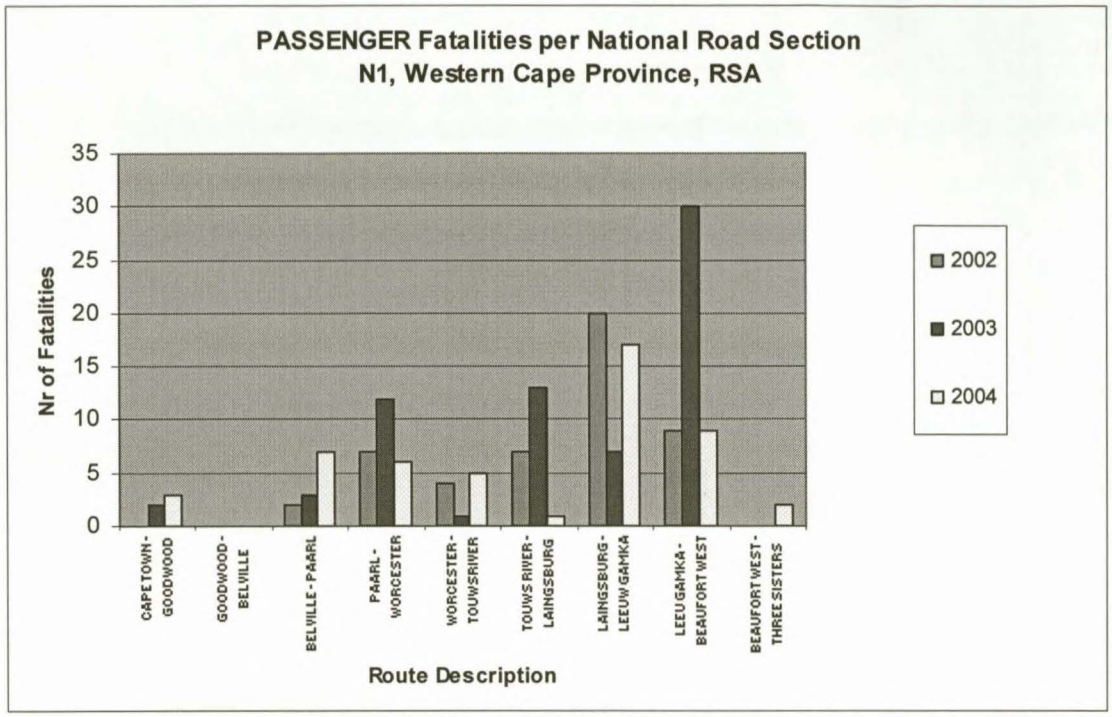

Fig. 4.1.11: Passenger Fatalities per National Road Section, Western Cape Province, N1

The passenger and pedestrian frequencies along the N1 tend to show different distributions along this national road. Passenger fatalities seem to increase in a easterly direction (towards more rural areas) with pedestrian fatalities occurring mostly in more urban areas between Cape Town and Paarl.

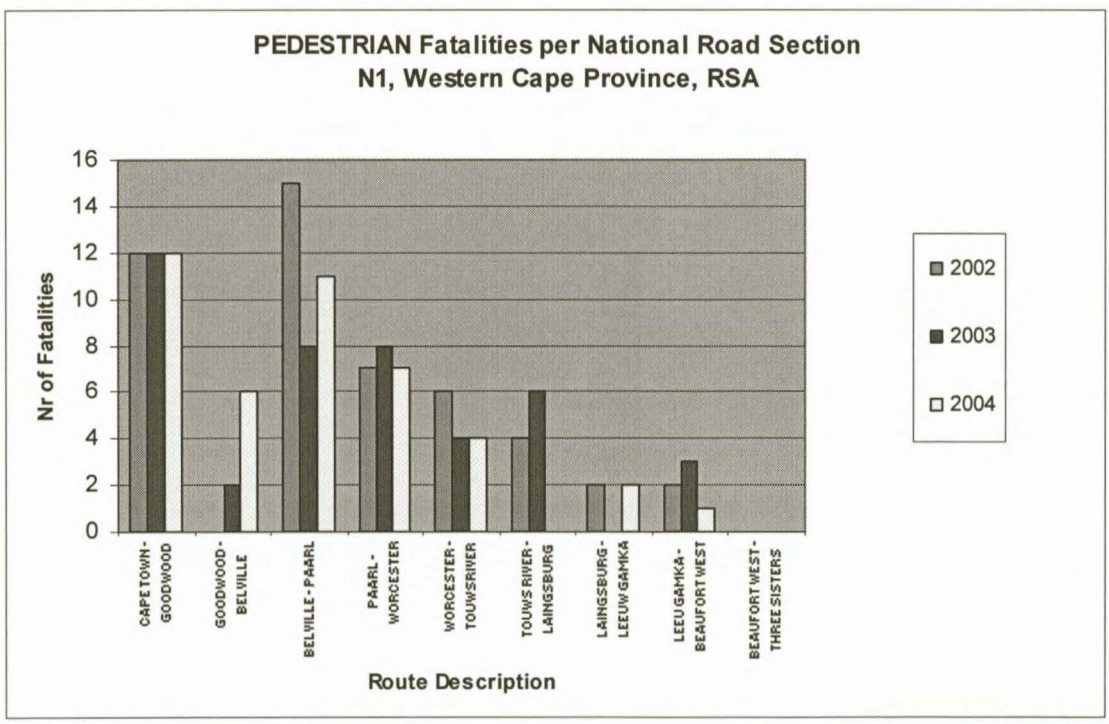

Fig. 4.1.12: Pedestrian Fatalities per National Road Section, Western Cape Province, N1

ii) N2: Fatalities by Road User Group, N1, N2 and N7, Western Cape Province

Figure 4.1.13 indicates that there is an overall increasing tendency in driver fatalities between 2002 and 2004. The driver fatality distribution is similar to the distribution along the N2 for the total amount of fatalities on this national road.

Passenger fatalities occurred in large amounts in 2003 along the N2, specifically between Riversdale and Mosselbay and then again between Knysna and Plettenbergbay. The fatality frequencies for latter road sections decreased rapidly from 2003 to 2004 though. It is again seen according to Figure 4.1.14 that passenger fatalities tend to occur more in rural areas (the same tendency was observed on the summaries for the N1).

Pedestrian fatalities along the N2 seemed to be mostly in the range of 0 to 5 pedestrian fatalities per road section along this national road for each year (2002-2004). The exception was the road section between Cape Town and Somerset West which showed a very rapid increase in pedestrian fatalities between the years 2002 and 2004 from between 0 and 5 pedestrian fatalities to approximately between 15 and 25 pedestrian fatalities.

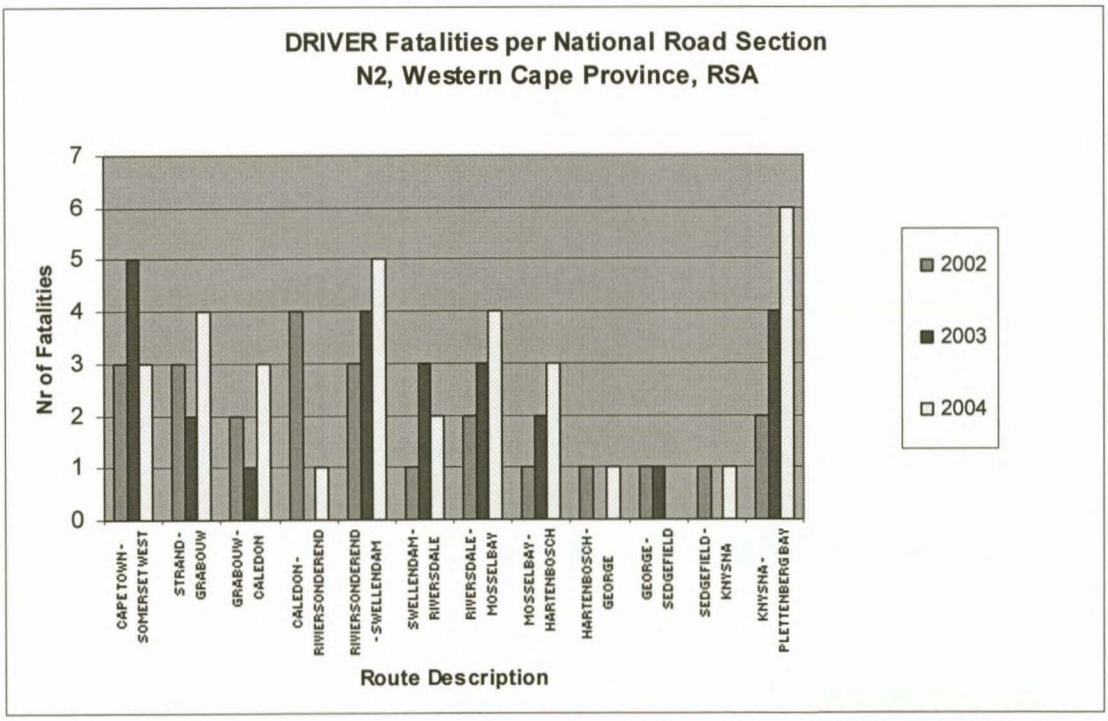

Fig. 4.1.13: Driver Fatalities per National Road Section, Western Cape Province, N2

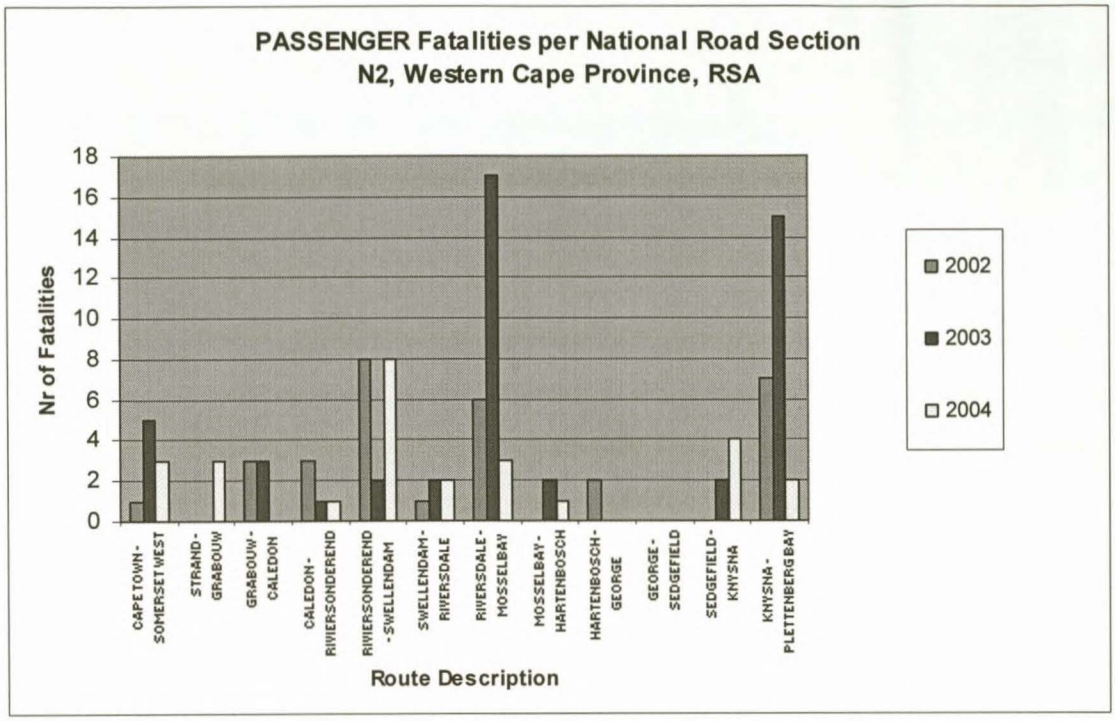

Fig. 4.1.14: Passenger Fatalities per National Road Section, Western Cape Province, N2

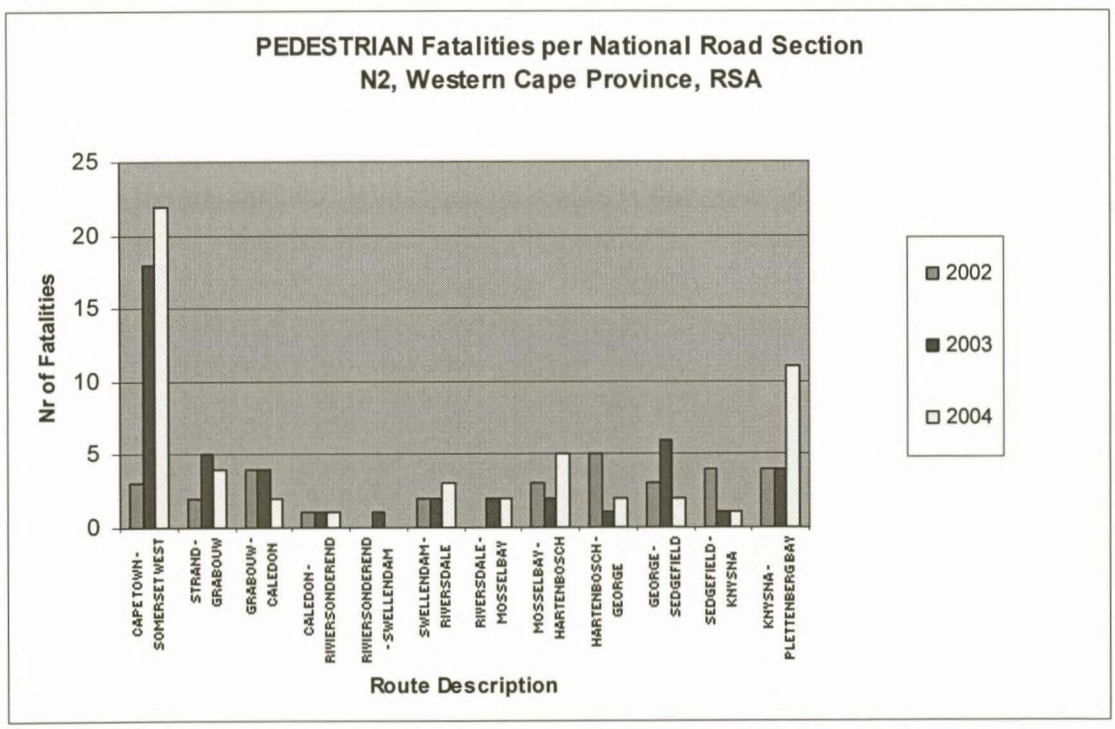

Fig. 4.1.15: Pedestrian Fatalities per National Road Section, Western Cape Province, N2

iii) N7: Fatalities by Road User Group, N1, N2 and N7, Western Cape Province

Figures 4.1.16, 4.1.17 and 4.1.18 suggest an overall increase in driver fatalities along the N7 with passenger and pedestrian fatalities showing a more decreasing tendency between the years 2002 and 2004. The exception being the road section between Goodwood and Malmesbury which actually showed an relatively rapid increase in pedestrian fatalities between 2002 and 2004.

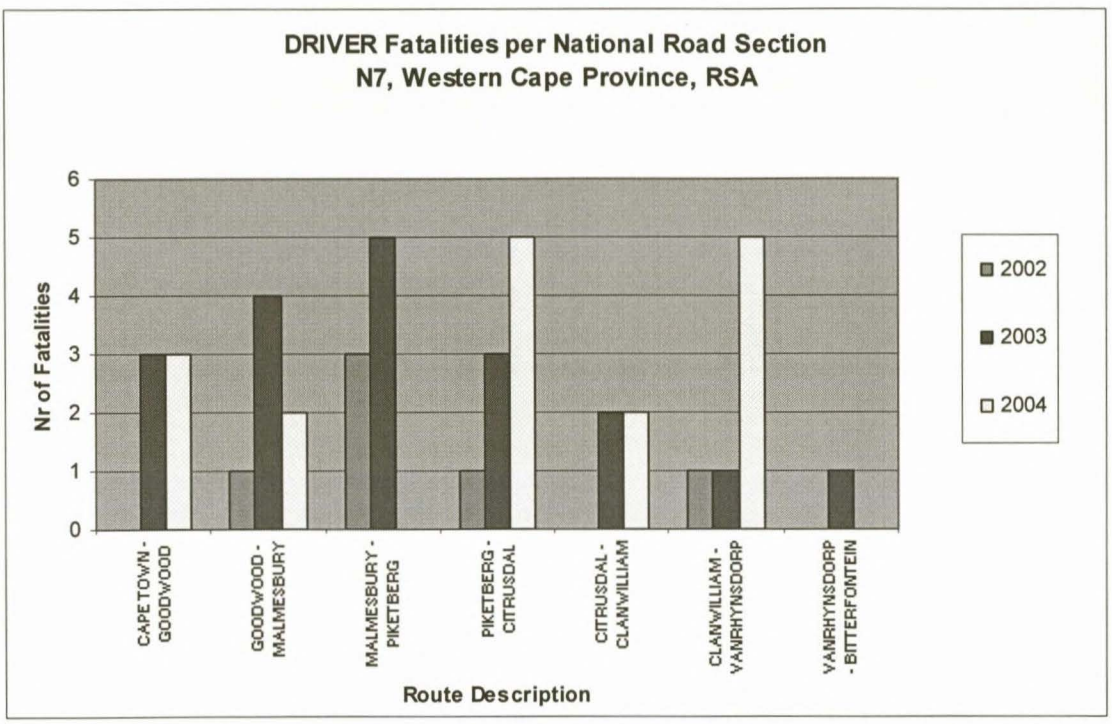

Fig. 4.1.16: Driver Fatalities per National Road Section, Western Cape Province, N7

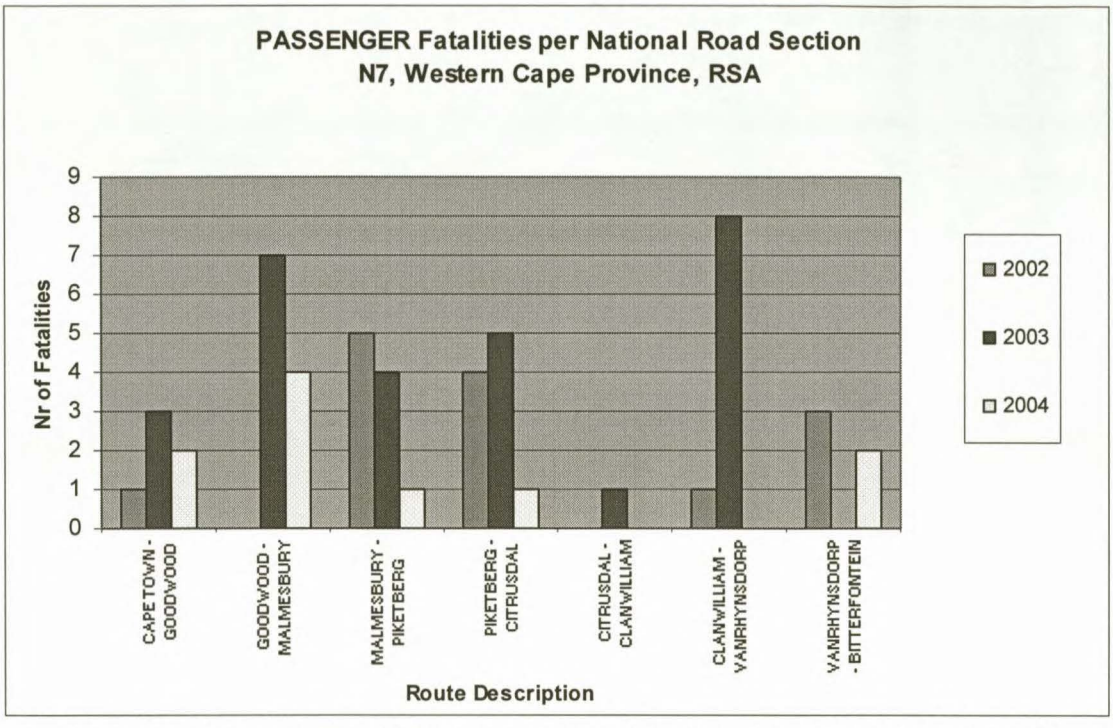

Fig. 4.1.17: Passenger Fatalities per National Road Section, Western Cape Province, N7

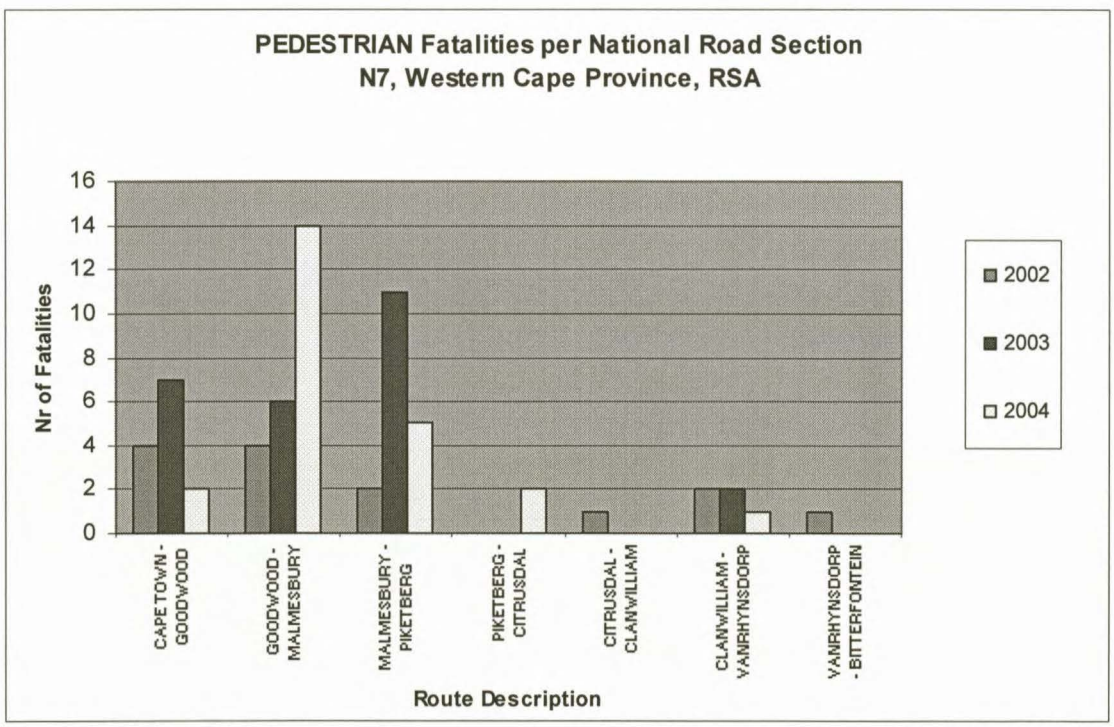

Fig. 4.1.18: Pedestrian Fatalities per National Road Section, Western Cape Province, N7

Table 4.1.2 contains all the data on which Figures 4.1.10 to 4.1.18 are based.

Table 4.1.2: Fatalities by Road User Group per National Road Section, Western Cape Province

			2002			Total	2003			Total	2004			Total	TOTAL
		Route Description	Driver	Passenger	Pedestrian		Driver	Passenger	Pedestrian		Driver	Passenger	Pedestrian		
N1	W1	CAPE TOWN - GOODWOOD	5	0	12	**17**	7	2	12	**21**	6	3	12	**21**	**59**
N1	W2	GOODWOOD - BELVILLE	7	0	0	**7**	2	0	2	**4**	0	0	6	**6**	**17**
N1	W3	BELVILLE - PAARL	5	2	15	**22**	2	3	8	**13**	2	7	11	**20**	**55**
N1	W4	PAARL - WORCESTER	3	7	7	**17**	7	12	8	**27**	3	6	7	**16**	**60**
N1	W5	WORCESTER - TOUWSRIVER	3	4	6	**13**	2	1	4	**7**	5	5	4	**14**	**34**
N1	W6	TOUWS RIVER - LAINGSBURG	5	7	4	**16**	9	13	6	**28**	0	1	0	**1**	**45**
N1	W7	LAINGSBURG - LEEUW GAMKA	7	20	2	**29**	3	7	0	**10**	5	17	2	**24**	**63**
N1	W8	LEEU GAMKA - BEAUFORT WEST	5	9	2	**16**	6	30	3	**39**	4	9	1	**14**	**69**
N1	W9	BEAUFORT WEST - THREE SISTERS	0	0	0	**0**	0	0	0	**0**	0	2	0	**2**	**2**
N2	W1	CAPE TOWN - SOMERSET WEST	3	1	3	**7**	5	5	18	**28**	3	3	22	**28**	**63**
N2	W3	STRAND - GRABOUW	3	0	2	**5**	2	0	5	**7**	4	3	4	**11**	**23**
N2	W4	GRABOUW - CALEDON	2	3	4	**9**	1	3	4	**8**	3	0	2	**5**	**22**
N2	W5	CALEDON - RIVIERSONDEREND	4	3	1	**8**	0	1	1	**2**	1	1	1	**3**	**13**
N2	W6	RIVIERSONDEREND - SWELLENDAM	3	8	0	**11**	4	2	1	**7**	5	8	0	**13**	**31**
N2	W7	SWELLENDAM - RIVERSDALE	1	1	2	**4**	3	2	2	**7**	2	2	3	**7**	**18**
N2	W8	RIVERSDALE - MOSSELBAY	2	6	0	**8**	3	17	2	**22**	4	3	2	**9**	**39**
N2	W9	MOSSELBAY - HARTENBOSCH	1	0	3	**4**	2	2	2	**6**	3	1	5	**9**	**19**
N2	W10	HARTENBOSCH - GEORGE	1	2	5	**8**	0	0	1	**1**	1	0	2	**3**	**12**
N2	W11	GEORGE - SEDGEFIELD	1	0	3	**4**	1	0	6	**7**	0	0	2	**2**	**13**
N2	W12	SEDGEFIELD - KNYSNA	1	0	4	**5**	0	2	1	**3**	1	4	1	**6**	**14**
N2	W13	KNYSNA - PLETTENBERG BAY	2	7	4	**13**	4	15	4	**23**	6	2	11	**19**	**55**
N7	W1	CAPE TOWN - GOODWOOD	0	1	4	**5**	3	3	7	**13**	3	2	2	**7**	**25**
N7	W2	GOODWOOD - MALMESBURY	1	0	4	**5**	4	7	6	**17**	2	4	14	**20**	**42**
N7	W3	MALMESBURY - PIKETBERG	3	5	2	**10**	5	4	11	**20**	0	1	5	**6**	**36**
N7	W4	PIKETBERG - CITRUSDAL	1	4	0	**5**	3	5	0	**8**	5	1	2	**8**	**21**
N7	W5	CITRUSDAL - CLANWILLIAM	0	0	1	**1**	2	1	0	**3**	2	0	0	**2**	**6**
N7	W6	CLANWILLIAM - VANRHYNSDORP	1	1	2	**4**	1	8	2	**11**	5	0	1	**6**	**21**
N7	W7	VANRHYNSDORP - BITTERFONTEIN	0	3	1	**4**	1	0	0	**1**	0	2	0	**2**	**7**
			70	**94**	**93**	**257**	**82**	**145**	**116**	**343**	**75**	**87**	**122**	**284**	**884**

iv) Fatalities by Race Group, N1, N2 and N7, Western Cape Province

Figures 4.1.19, 4.1.20 and 4.1.21 illustrate the distribution between the different race groups along the national roads N1, N2 and N7 within the Western Cape as was queried from the MS Access database.

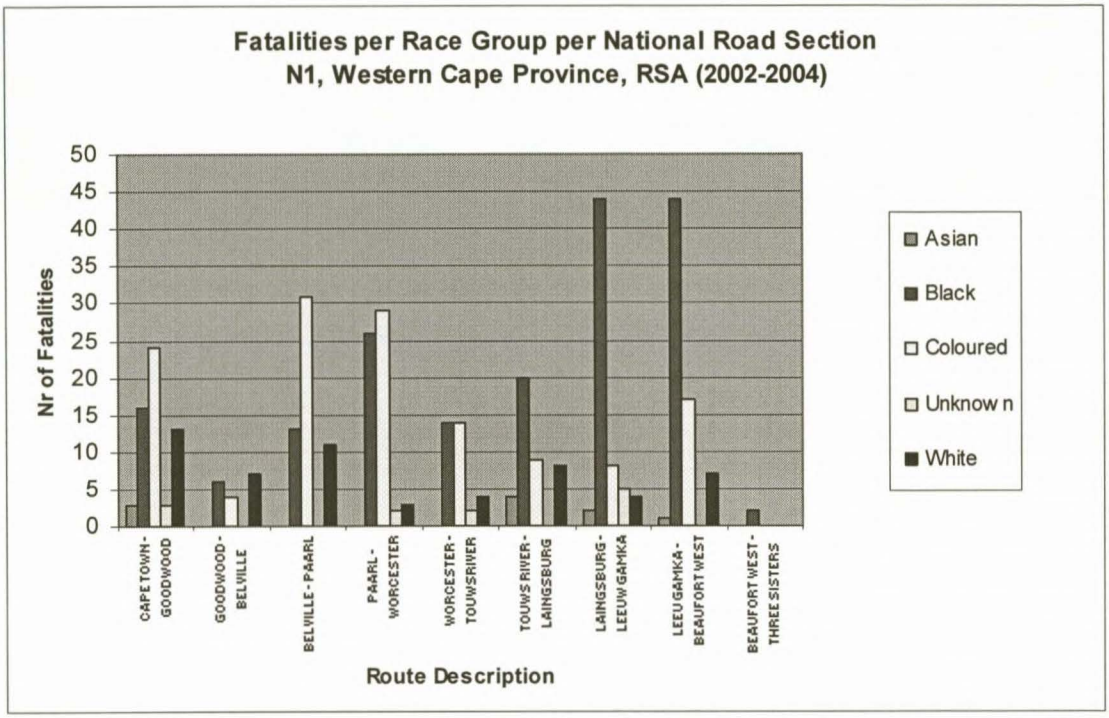

Fig. 4.1.19: Fatalities per Race Group per National Road Section, Western Cape Province N1, 2002-2004

Mostly black and coloured fatalities occur on the N1 with peak values between Bellville and Worcester and between Laingsburg and Beaufort West (see Figure 4.1.19 above).

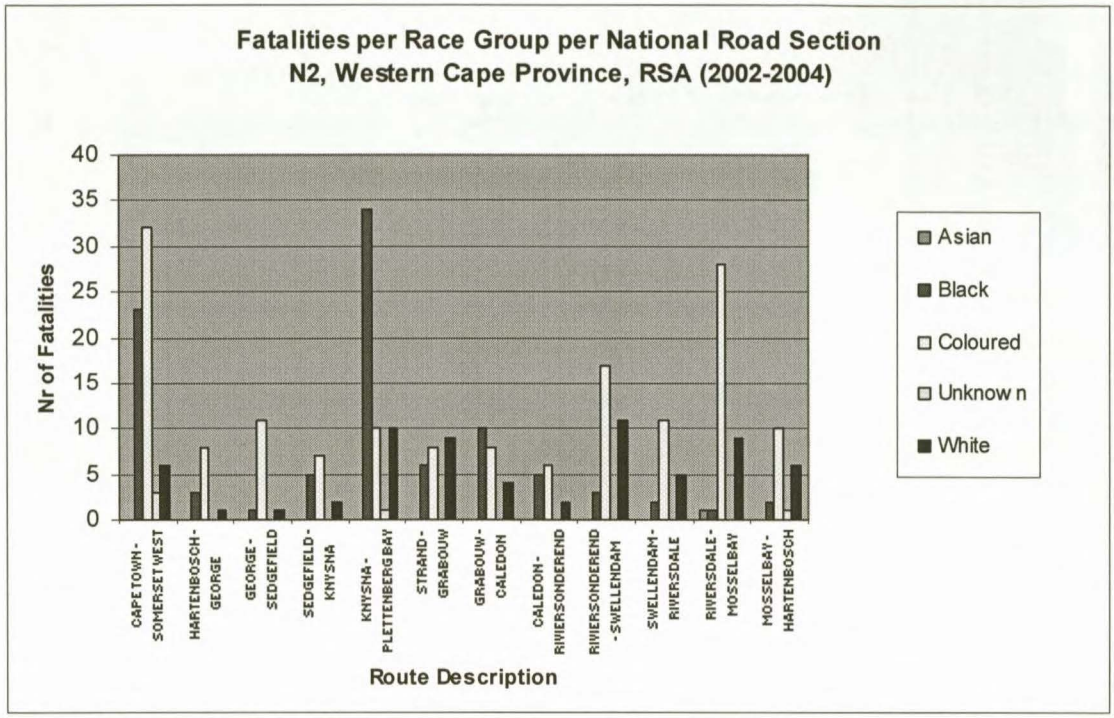

Fig. 4.1.20: Fatalities per Race Group per National Road Section, Western Cape Province N2, 2002-2004

On the N2 mostly coloured and black fatalities occur on the N2 with some peak values between Cape Town and Somerset West, between Riversdale and Mosselbay and between Knysna and Plettenbergbay.

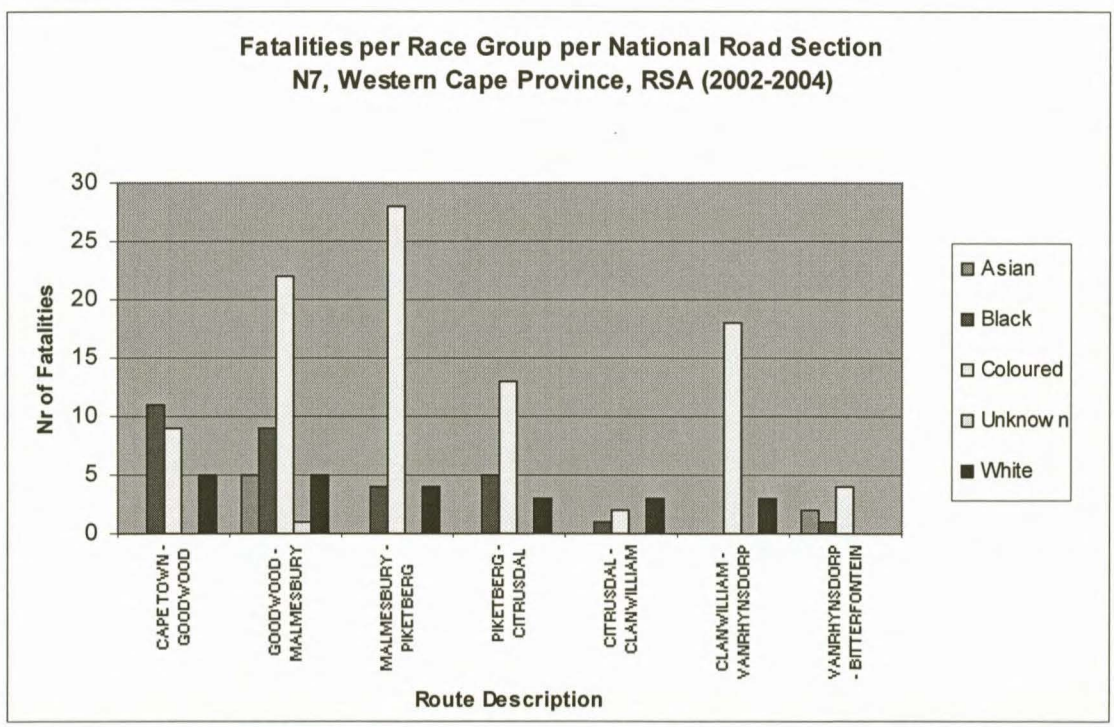

Fig. 4.1.21: Fatalities per Race Group per National Road Section, Western Cape Province N7, 2002-2004

Fig. 4.1.21 shows that coloured fatalities are a large percentage of the total number of fatalities on the N7 with most coloured fatalities occurring between Goodwood and Piketberg.

v) Seatbelt status (only fatality cases; Pedestrians excl.), N1, N2 and N7, Western Cape Province

Figures 4.1.22, 4.1.23 and 4.1.24 illustrate the seatbelt wearing behaviour of fatality cases along the national roads N1, N2 and N7 within the Western Cape Province. Pedestrian fatality cases are not included in the illustrations.

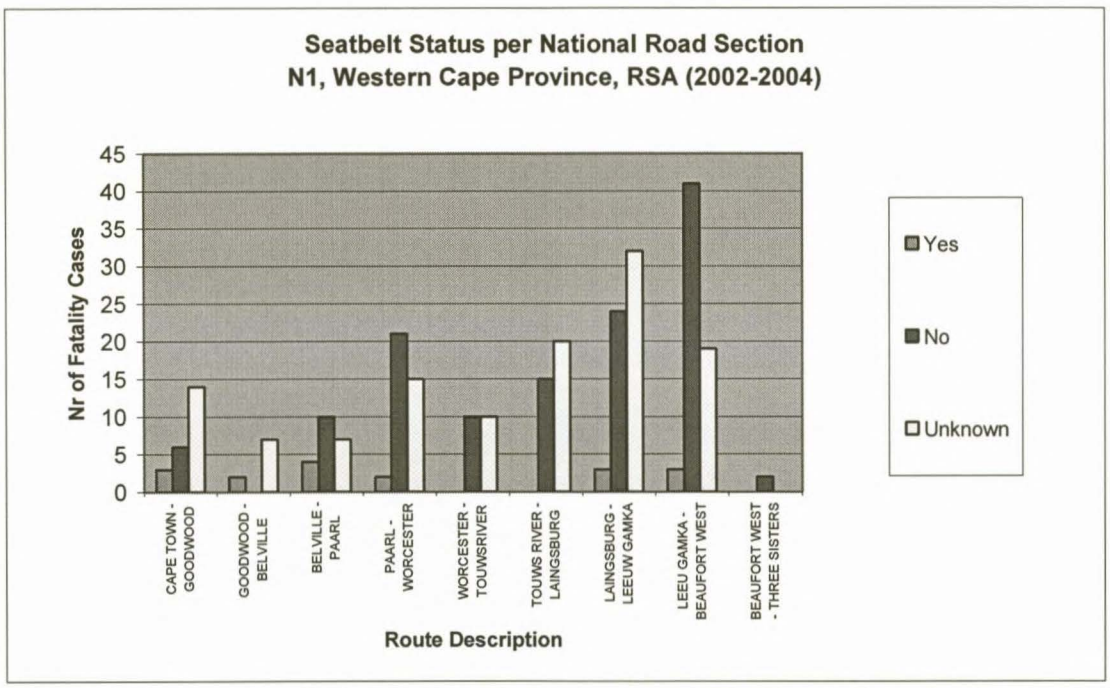

Fig. 4.1.22: Seatbelt status (only fatality cases; Pedestrians excl.), N1, Western Cape Province, 2002-2004

It is clear from all figures that the amount of unknown cases are a very large percentage of the total number of fatalities on the national roads (as is the case for the whole country). It was explained previously that this could be due to large amounts of passenger fatalities where the seatbelt status of each victim might be difficult to determine or monitor, especially in multiple fatalities when buses, minibuses or minibus taxi's are involved in fatal road accidents. It could also be due to poor completion of the accident report forms. Accident investigations mostly focus on how an accident happened or what the possible causes could have been, instead of also including an investigation into why a road accident has a particular level of severity in terms of the fatalities. Modern vehicles also

have airbags in which case the seatbelt status of a victim dying in such a vehicle would most likely be noted as "no".

According to these findings a very small percentage of road fatalities on the national roads wore their seatbelts prior to their death in fatal accidents thus resulting in a large group who did not wear any seatbelts when the fatal accidents occurred. Due to the large amount of unknown cases it is difficult to accurately determine how reliable these findings are and any studies based on the road behaviour of road users cannot be based on this data.

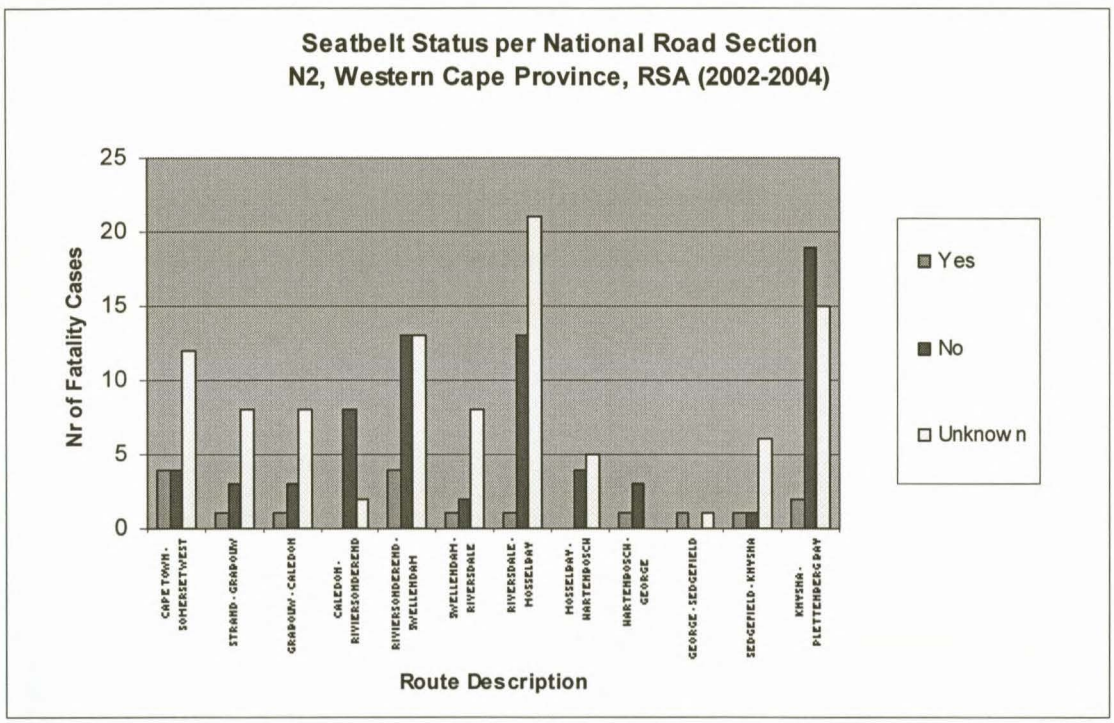

Fig. 4.1.23: Seatbelt status (only fatality cases; Pedestrians excl.), N2, Western Cape Province, 2002-2004

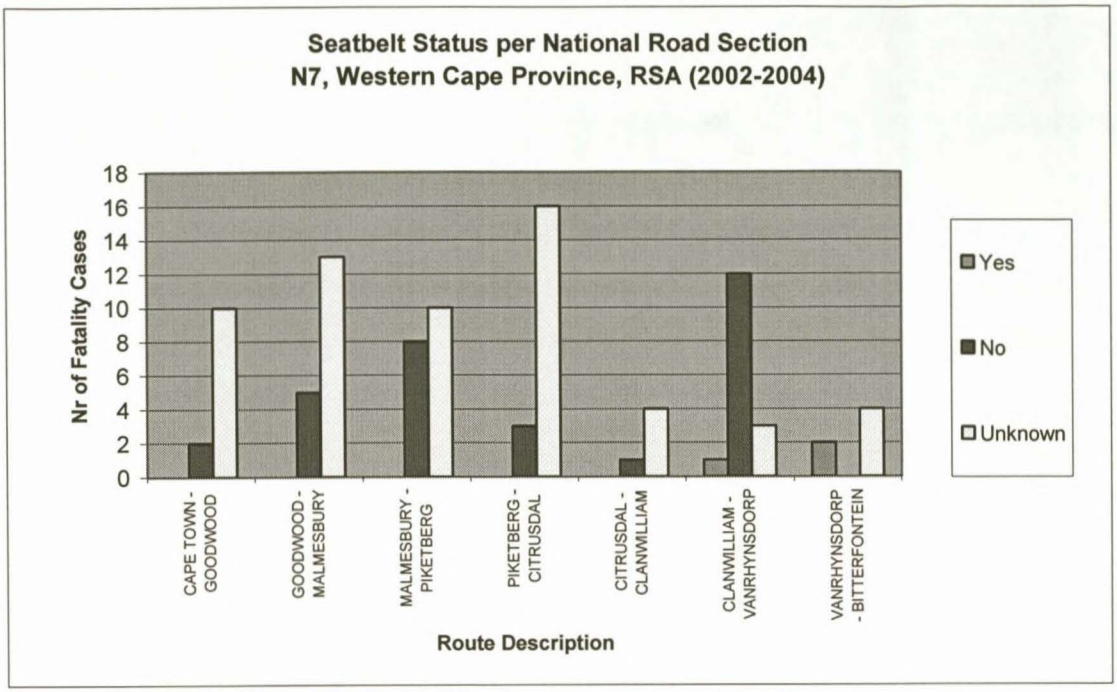

Fig. 4.1.24: Seatbelt status (only fatality cases; Pedestrians excl.), N7, Western Cape Province, 2002-2004

vi) Fatalities by Gender, N1, N2 and N7, Western Cape Province

Figures 4.1.25, 4.1.26 and 4.1.27 summarize fatalities by gender along the N1, N2 and N7 within the Western Cape Province.

From all figures it is clear that male fatalities are the majority. A very small amount of unknown cases are noted in each case with the most unknown gender fatality cases noted on the N1 between Laingsburg and Beaufort West. The small amount of unknown cases could again be due to the care taken in completing accident report forms by traffic officials or accident investigators.

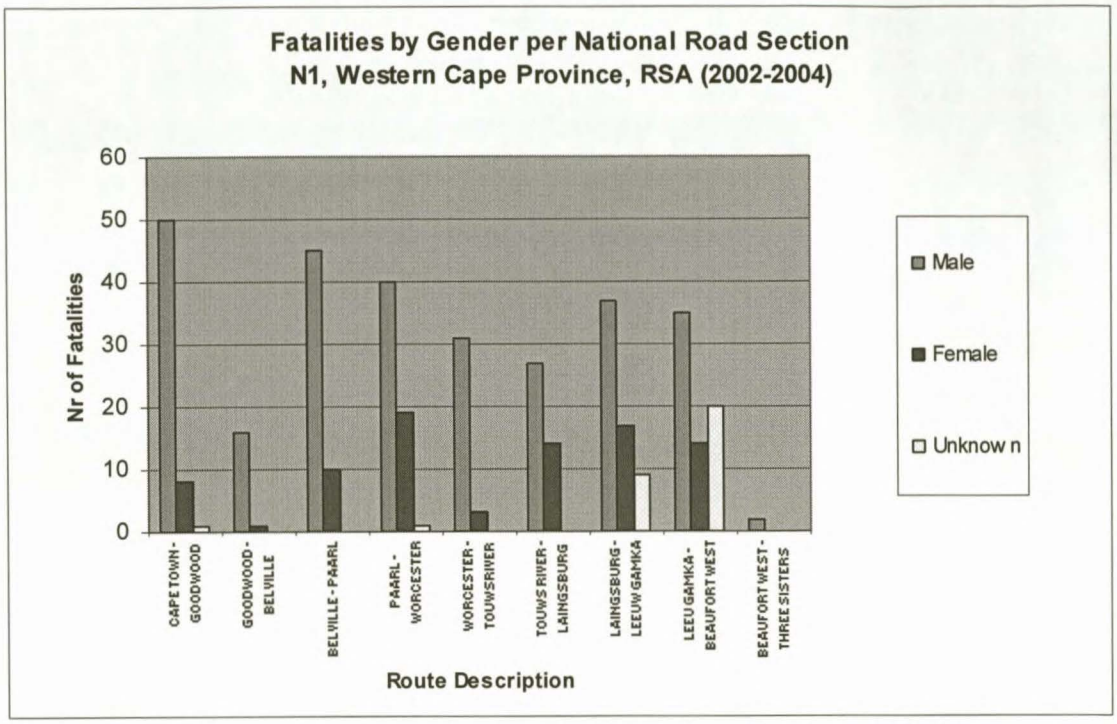

Fig. 4.1.25: Fatalities per Gender, N1, Western Cape Province, 2002-2004

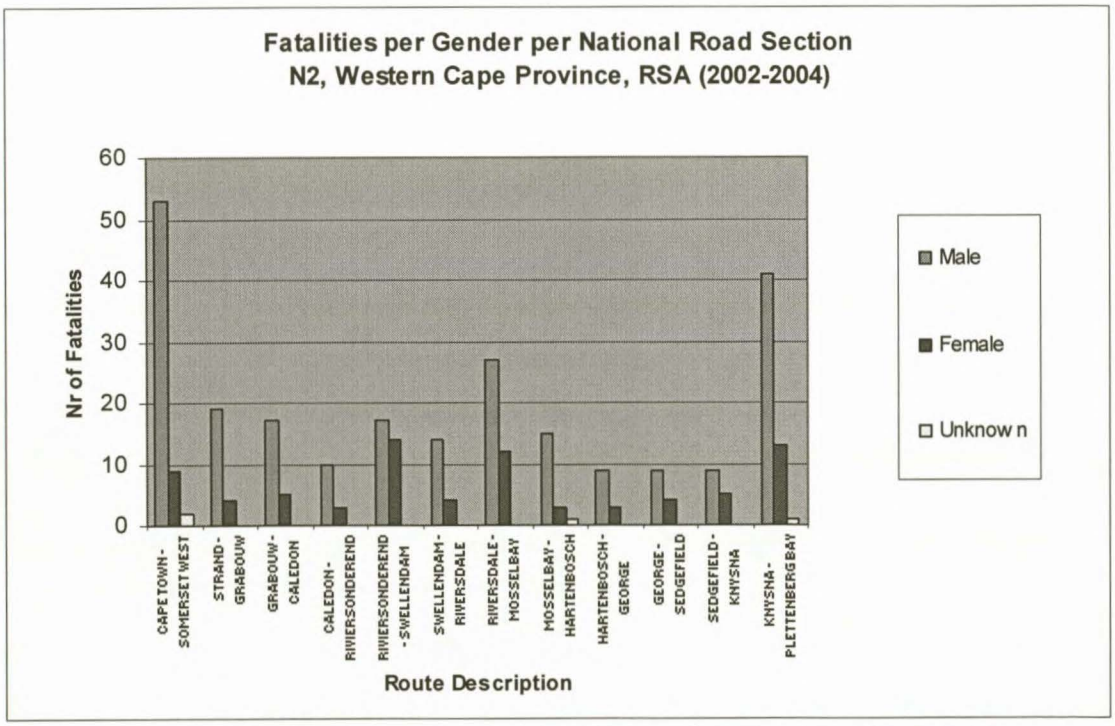

Fig. 4.1.26: Fatalities per Gender, N2, Western Cape Province, 2002-2004

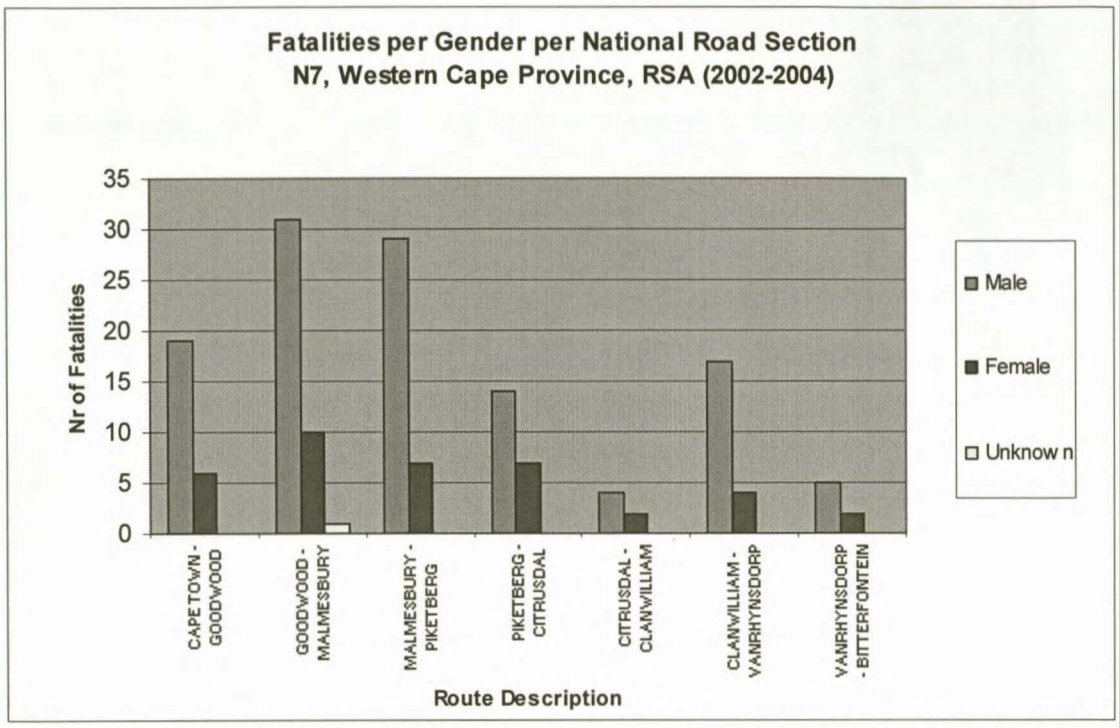

Fig. 4.1.27: Fatalities per Gender, N7, Western Cape Province, 2002-2004

4.2 Traffic and Speed Data

Chapter 2 and 3 described the relevant data sources and methodology for capturing traffic and speed data for the road sections under study along the N1, N2 and N7 within the Western Cape Province, South Africa. In this section the results are represented graphically, with data tables provided following the graphs.

Each graph represents the data gathered for the period 2002-2004 with the exception of the graphs on average speeds along the road sections, which only give the average speeds for 2003. The average speed values for 2003 can safely be used as point estimates for the whole period of 2002-2004, because of insignificant differences in average speeds in the course of three years.

The ADT (Average Daily Traffic in veh/d) and ADTT (Average Daily Truck Traffic in veh/d) estimates were very dependent on the method described in Chapter 3 on how traffic volumes and speed data variables were determined for each road section. The fatal road accident database mostly governed the manner in which data along a particular route was to be analysed i.e. it was not known at what exact location the accidents occurred; only road sections between different towns are known.

Point estimates then had to be determined for ADT, ADTT and the speed variables for these road sections as explained before.

4.2.1 Traffic and Speed Data: N1, Western Cape Province

As can be expected, the ADT point estimates for each road section along the N1 national road decrease gradually as is moved in an easterly direction along the N1, away from Cape Town and towards rural territory (see Fig. 4.2.1). The ADT also shows a steady growth from year to year.

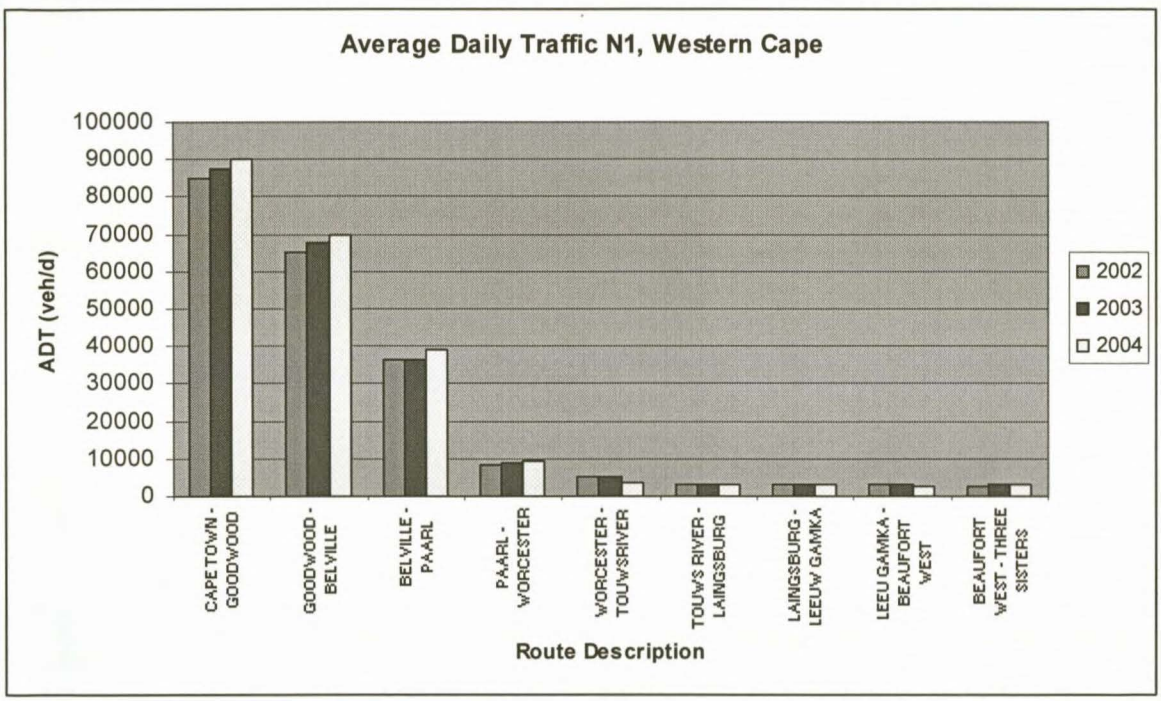

Fig. 4.2.1: Average Daily Traffic for N1, Western Cape Province 2002-2004

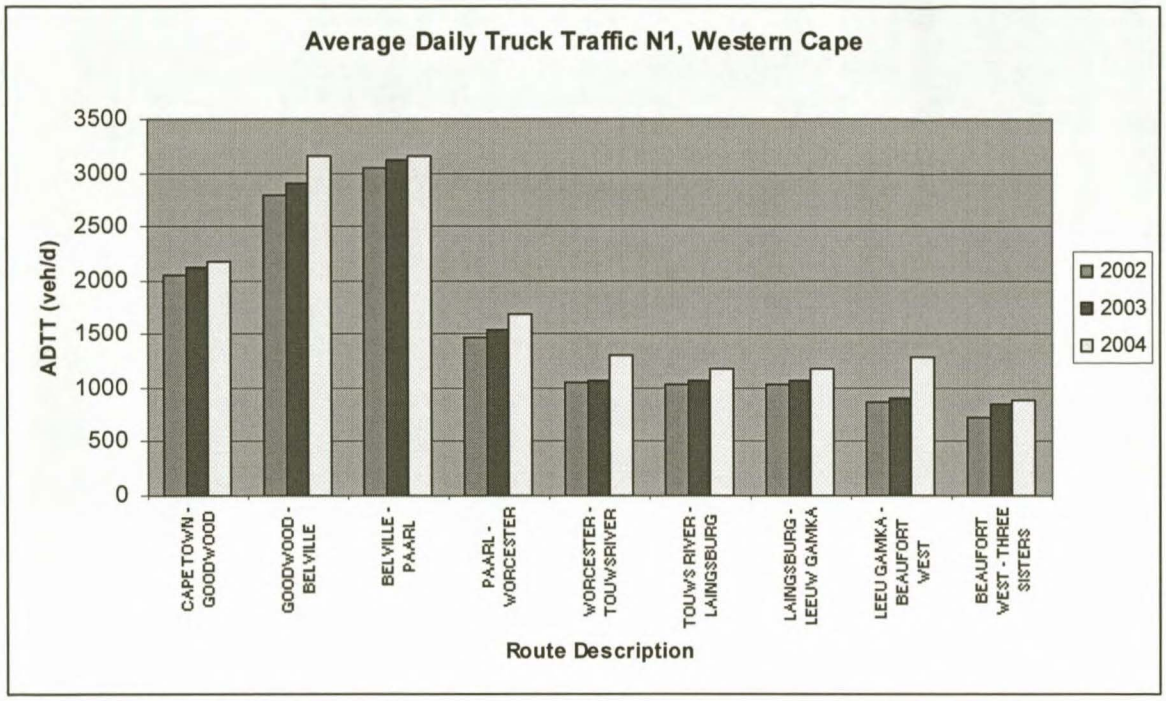

Fig. 4.2.2: Average Daily Truck Traffic for N1, Western Cape Province 2002-2004

ADTT point estimates along the N1 (Fig. 4.2.2) do not follow the same distribution as the ADT estimates, but are similar with regard to the decreasing pattern in an easterly direction towards rural areas. Year-to-year ADTT growth seems to be at a steady pace along this national road.

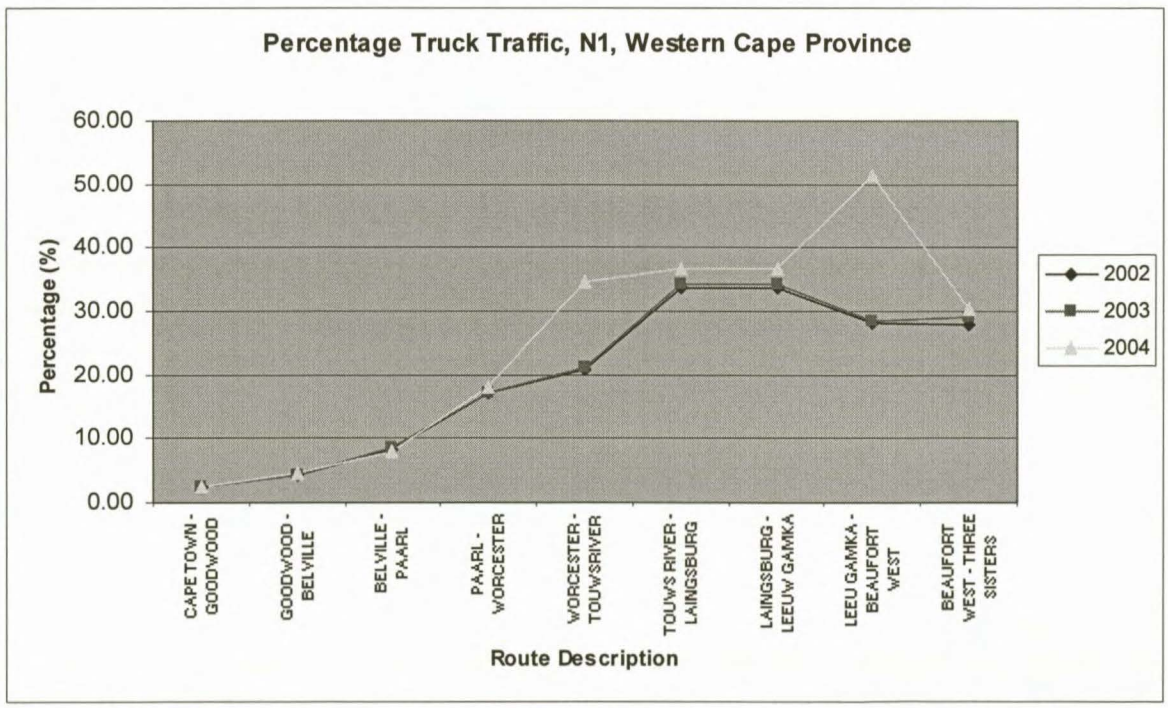

Fig. 4.2.3: Percentage Truck Traffic for N1, Western Cape Province 2002-2004

Although the traffic volumes along the N1 decreases in an easterly direction within the Western Cape Province, the percentage of truck traffic increases along the route as can be clearly seen from Fig. 4.2.3. It can also be observed that the increasing amount of truck traffic along the N1 (see Fig. 4.2.2) does not necessarily mean that the average speeds at which vehicles travel will be subdued.

When inspecting Figures 4.2.4 and 4.2.5, it is clear that the variation in average speeds along the N1 corresponds to the variation in the percentage vehicles that travel at average speeds of higher than the speed limit.

There is a direct proportional relationship between average speed and the percentage vehicles travelling at average speeds of higher than the speed limit i.e. the higher the average speeds on a road the higher the likelihood that the percentage of vehicles speeding is relatively high (as would be expected).

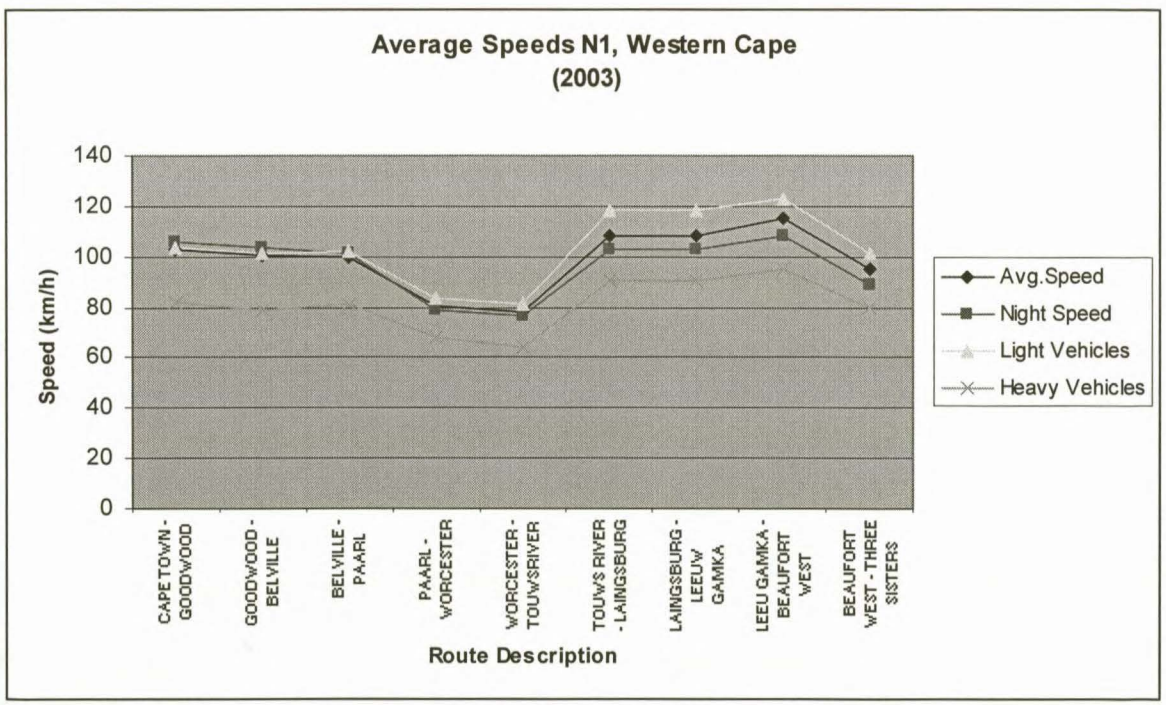

Fig. 4.2.4: Average Speeds for N1, Western Cape Province 2002-2004

Vehicles travel at fairly constant average speeds between Cape Town and Paarl (light vehicles travel at average speeds of between 100km/h and 110km/h and heavy vehicles at average speeds of approximately 80km/h) (Fig. 4.2.4). 10 to 20 percent of the total traffic volume was observed to speed between Cape Town and Paarl. The average speeds of light vehicles between Paarl and Touwsriver are approximately 80km/h and heavy vehicles between 60km/h and 70km/h. Approximately 5% of traffic

was observed to speed on this road section. Between Touwsriver and Beaufort West light vehicles travel at an approximate constant average speed of 120km/h while heavy vehicles travel at constant average speeds between 90km/h and 100km/h. 30 to 45 percent of the traffic was observed to speed between Touwsriver and Beaufort West.

Average night speeds were observed to vary between 100km/h and 110km/h between Cape Town and Paarl, remain constant at approximately 80km/h between Paarl and Touwsriver and again vary between 100km/h and 110km/h between Touwsriver and Beaufort West.

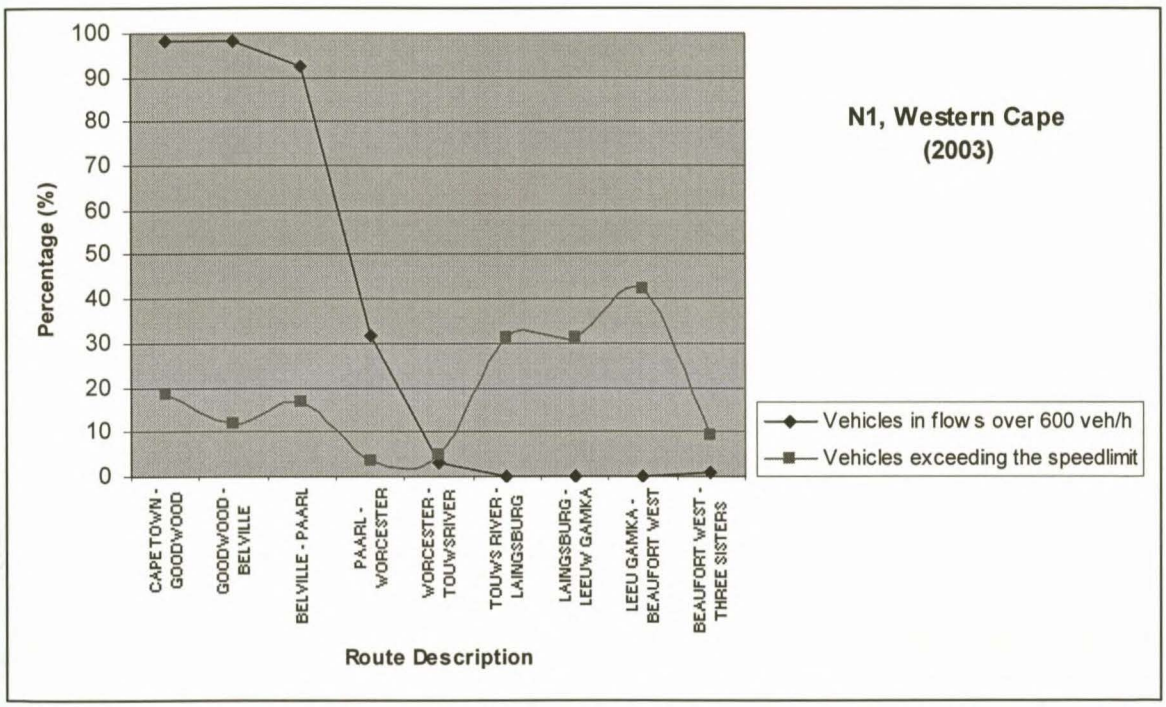

Fig. 4.2.5: Percentage Vehicles in flows over 600 veh/h and vehicles exceeding the speed limit for N1, Western Cape Province 2002-2004

As would be expected, the percentage of vehicles travelling in volumes over 600 veh/h (≡ ± 14400veh/d) decreases at a rate corresponding to the rate at which traffic volumes decrease in an easterly direction along the N1 within the Western Cape Province. Also, Figure 4.2.5 illustrates the tendency of more vehicles speeding as traffic volume decreases along the N1.

Table 4.2.1 contains the data on which Figures 4.2.1, 4.2.2, 4.2.3, 4.2.4 and 4.2.5 are based.

Table 4.2.1: Traffic and Speed Data per National Road Section for the N1, Western Cape Province

2002									
strRouteDescription	ADT2002 (veh/d)	ADTT2002 (veh/d)	%Trucks	Average Speed (km/h)	Avg Night Speed (km/h)	Speed L Veh (km/h)	Speed H Veh (km/h)	Over 600 veh/h (%)	Over Speed Lim (%)
CAPE TOWN - GOODWOOD	84891	2053	2.42	102.8	106.2	103.3	82.1	98.4	18.7
GOODWOOD - BELVILLE	65000	2800	4.31	100.5	103.9	101.4	78.7	97.95	12
BELVILLE - PAARL	36418	3050	8.37	102.3	102.7	104.6	80.6	92.2	20.7
PAARL - WORCESTER	8584	1475	17.18	84.4	80.7	88.3	64.9	23.7	5.4
WORCESTER - TOUWSRIVER	4987	1045	20.95	77.8	76.6	81.5	64.3	3.3	4.8
TOUWS RIVER - LAINGSBURG	3048	1028	33.73	109	102.7	118.2	90.8	0	31.6
LAINGSBURG - LEEUW GAMKA	3048	1028	33.73	109	102.7	118.2	90.8	0	31.6
LEEU GAMKA - BEAUFORT WEST	3102	878	28.30	115.2	108.6	123.1	95.2	0	42.5
BEAUFORT WEST - THREE SISTERS	2620	728	27.79	94.5	88.5	100.6	78.6	1.6	9.4

2003									
strRouteDescription	ADT2003 (veh/d)	ADTT2003 (veh/d)	%Trucks	Average Speed (km/h)	Avg Night Speed (km/h)	Speed L Veh (km/h)	Speed H Veh (km/h)	Over 600 veh/h (%)	Over Speed Lim (%)
CAPE TOWN - GOODWOOD	87438	2114	2.42	102.8	106.2	103.3	82.1	98.4	18.7
GOODWOOD - BELVILLE	67500	2900	4.30	100.35	104	101.25	78.65	98.05	12
BELVILLE - PAARL	36693	3111	8.48	100.1	101.1	102.1	81.2	92.4	16.9
PAARL - WORCESTER	8928	1544	17.29	80.5	78.6	83.6	67.7	31.6	3.7
WORCESTER - TOUWSRIVER	5136	1077	20.97	77.8	76.6	81.5	64.3	3.3	4.8
TOUWS RIVER - LAINGSBURG	3111	1066	34.27	108.6	102.5	118.1	90.6	0.2	31.1
LAINGSBURG - LEEUW GAMKA	3111	1066	34.27	108.6	102.5	118.1	90.6	0.2	31.1
LEEU GAMKA - BEAUFORT WEST	3195	905	28.33	115.2	108.6	123.1	95.2	0	42.5
BEAUFORT WEST - THREE SISTERS	2945	860	29.20	95.1	89.05	101.3	79.8	1	9.2

2004									
strRouteDescription	ADT2004 (veh/d)	ADTT2004 (veh/d)	%Trucks	Average Speed (km/h)	Avg Night Speed (km/h)	Speed L Veh (km/h)	Speed H Veh (km/h)	Over 600 veh/h (%)	Over Speed Lim (%)
CAPE TOWN - GOODWOOD	90061	2178	2.42	102.8	106.2	103.3	82.1	98.4	18.7
GOODWOOD - BELVILLE	70000	3150	4.50	100.2	103.75	101.1	79.1	98.1	12
BELVILLE - PAARL	39070	3155	8.08	100.3	101.8	102.1	81.5	93	16.3
PAARL - WORCESTER	9373	1688	18.01	81.1	78.9	84	68.5	34.4	3.2
WORCESTER - TOUWSRIVER	3798	1311	34.52	87.3	83.1	92.9	78.3	0	3.4
TOUWS RIVER - LAINGSBURG	3219	1180	36.66	107.2	101.6	116.8	90.5	0.2	28.2
LAINGSBURG - LEEUW GAMKA	3219	1180	36.66	107.2	101.6	116.8	90.5	0.2	28.2
LEEU GAMKA - BEAUFORT WEST	2490	1282	51.49	104.9	101.4	118.2	92.3	0	23.9
BEAUFORT WEST - THREE SISTERS	2943	890	30.24	94.5	88.7	100.7	79.7	1.4	8.75

4.2.2 Traffic and Speed Data: N2, Western Cape Province

ADT point estimates along the N2 national road (see Fig. 4.2.6) are mostly below 10000 vehicles per day per road section with the exception of the road section between Cape Town and Somerset West with a peak value for ADT of between 40000 and 55000 vehicles per day. Keep in mind that values given in the graphical representations are point estimates of the true circumstances and that the values cannot really be taken as absolute. Refer to Chapter 3 for the methodology on how the values for each road section were estimated. Traffic growth from year-to-year seems to occur at a steady pace along the road.

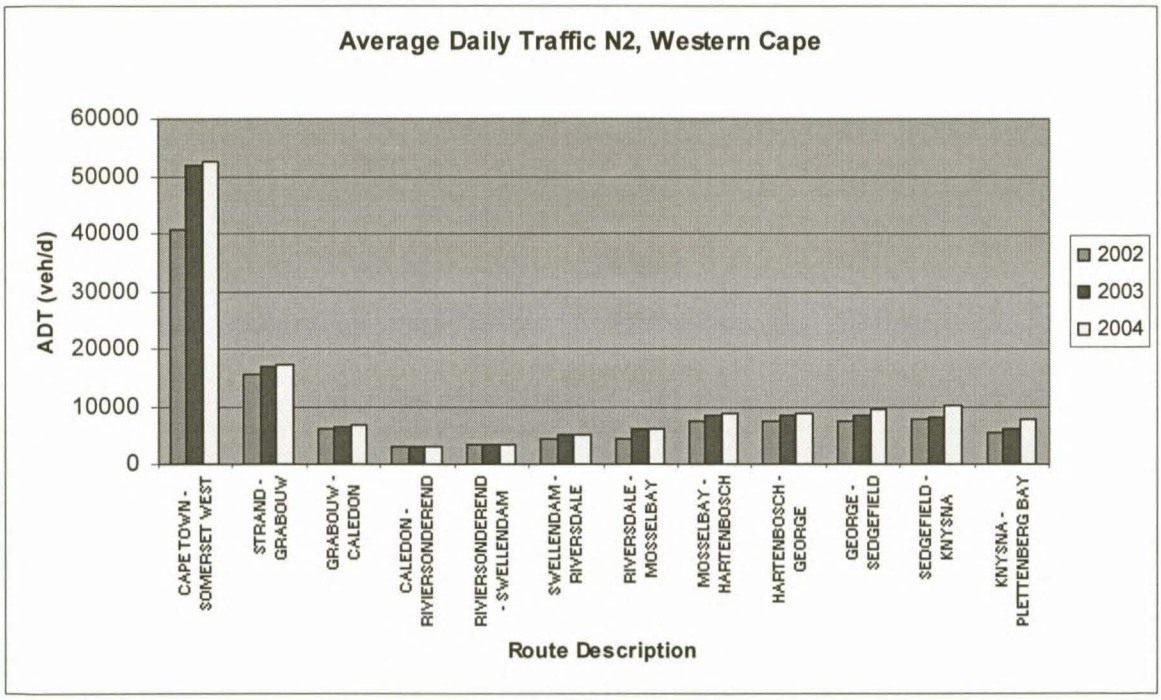

Fig. 4.2.6: Average Daily Traffic for N2, Western Cape Province 2002-2004

The ADTT distribution along the N2 follows a similar pattern than the ADT distribution along the road (see Fig. 4.2.7). Again, the point estimate for the road section between Cape Town and Somerset West could possibly be seen as an outlier with values varying approximately between 2000 and 2500 trucks per day or it could be interpreted that the rest of the point estimates for the N2 are in fact higher. Most road sections have ADTT estimates varying approximately between 500 and 1000 trucks per day.

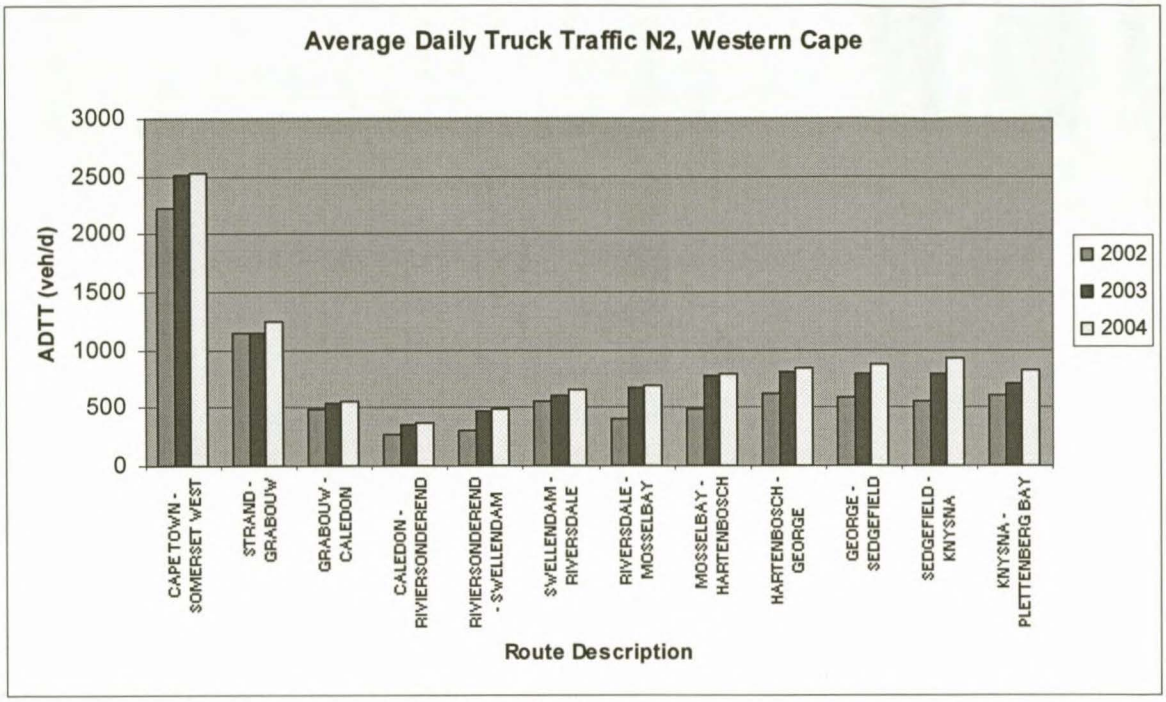

Fig. 4.2.7: Average Daily Truck Traffic for N2, Western Cape Province 2002-2004

The variation of the percentage truck traffic shows an increase in the relative amount of truck traffic from Cape Town up to Swellendam after which it decreases up to Hartenbos and then remains fairly constant until Knysna. As was discussed before the amount of truck traffic does not necessarily subdue the average speeds as can be seen from Fig. 4.2.9. The percentage vehicles speeding and the speed limit on any particular road section play a bigger role in determining what a particular average speed distribution along any road looks like.

Average speeds of light vehicles along the N2 increases gradually from an average value of 100km/h at Cape Town to approximately 120km/h between Caledon and Swellendam (Fig. 4.2.9). These average speeds then decrease again until a value of approximately 90km/h is reached between Mosselbay and Hartenbos. Light vehicle average speeds then remain at this approximate constant value for the remainder of the N2 up to Plettenbergbay.

Heavy vehicle average speeds follow the same distribution along the N2 than for light vehicle average speeds, except that the peak value between Caledon and Swellendam for heavy vehicles is approximately 90km/h (Fig. 4.2.9). These speeds then decrease until an approximate value of 80km/h is reached between Mosselbay and Hartenbos. Heavy vehicle average speeds then remain between 70km/h and 80 km/h up to Plettenbergbay.

Between Cape Town and Somerset West about 15% of vehicles were observed to speed. About 45% of vehicles were observed to speed between Caledon and Swellendam and about 10% between Mosselbay and Hartenbosch.

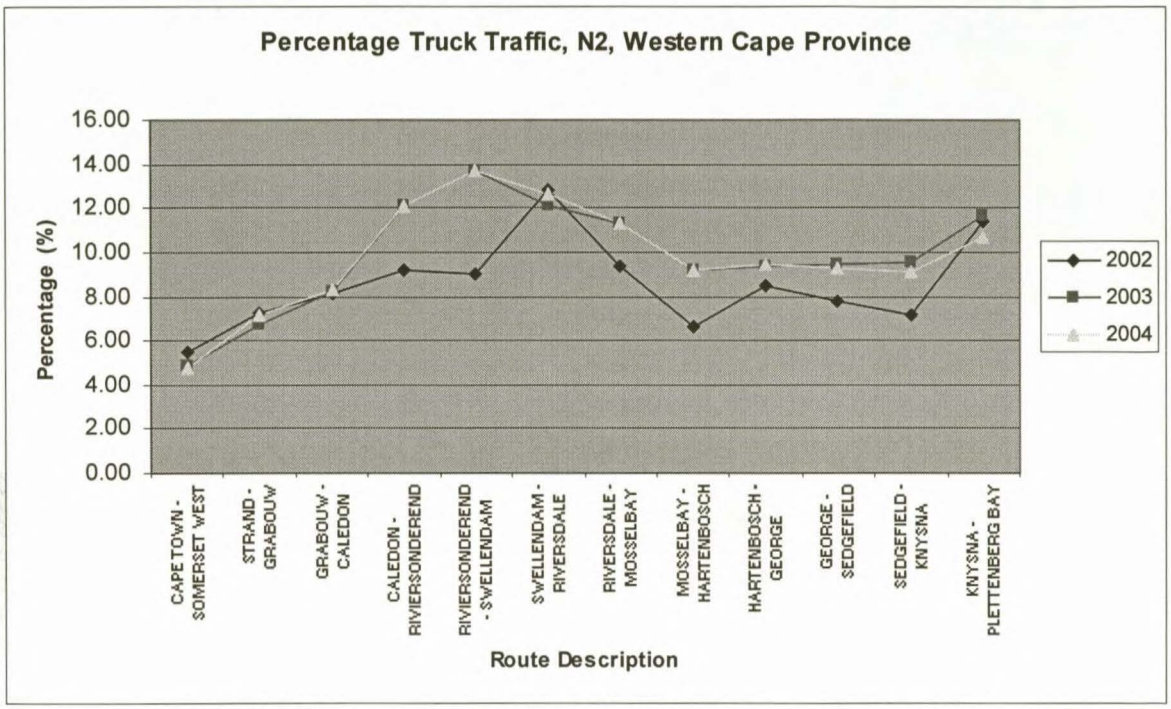

Fig. 4.2.8: Percentage Truck Traffic for N2, Western Cape Province 2002-2004

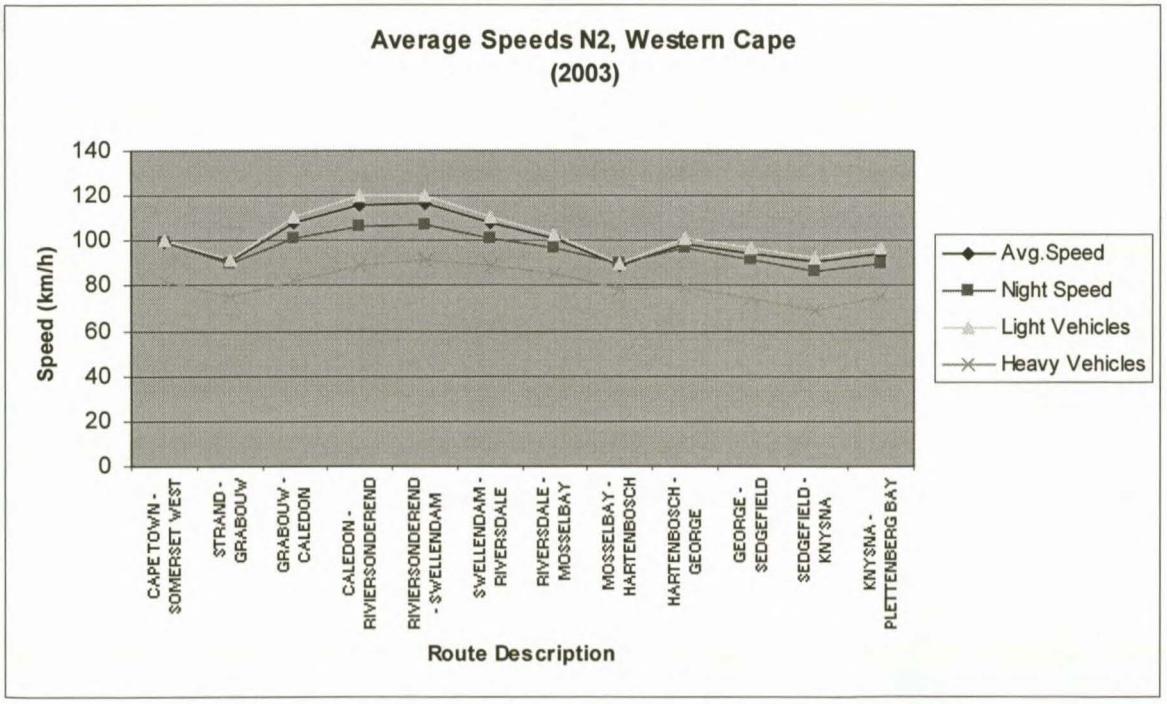

Fig. 4.2.9: Average Speeds for N2, Western Cape Province 2002-2004

Average night speeds between Cape Town and Swellendam varies between 100km/h and 110km/h and then gradually decreases until an approximately value of 90km/h is reached between Knysna and Plettenbergbay.

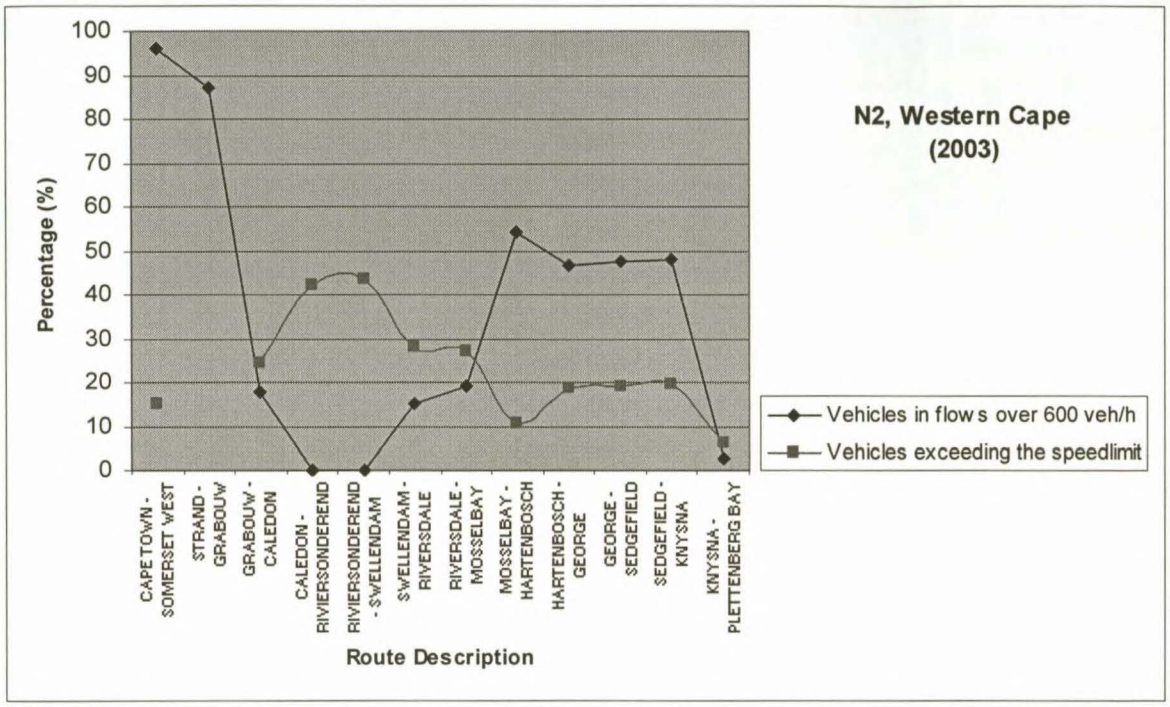

Fig. 4.2.10: Percentage Vehicles in flows over 600 veh/h and vehicles exceeding the speed limit for N2, Western Cape Province 2002-2004

The variation in the percentage vehicles speeding along the N2 corresponds to the distribution in average speeds along the road (Fig. 4.2.10). This was also observed for the N1 (previously discussed). The same relationship between the percentage vehicles travelling in flows higher than 600veh/h (≡ ± 14400 veh/d) and traffic volumes along the N2 were observed than for the N1. More vehicles travel at speeds higher than the speed limit as traffic volumes decrease.

Table 4.2.2 contains all the data on which Figures 4.2.6, 4.2.7, 4.2.8, 4.2.9 and 4.2.10 are based.

Table 4.2.2: Traffic and Speed Data per National Road Section for the N2, Western Cape Province

2002									
strRouteDescription	ADT2002 (veh/d)	ADTT2002 (veh/d)	%Trucks	Average Speed (km/h)	Avg Night Speed (km/h)	Speed L Veh (km/h)	Speed H Veh (km/h)	Over 600 veh/h (%)	Over Speed Lim (%)
CAPE TOWN - SOMERSET WEST	40586	2229	5.49	104.4	102.9	106.2	77.7	95.1	21.1
STRAND - GRABOUW	15746	1138	7.23	85.6	86.7	87.2	65.6	85.5	
GRABOUW - CALEDON	6028	491	8.15	108.1	101.1	110.4	81.7	15.4	24.5
CALEDON - RIVIERSONDEREND	2971	273	9.19	113.6	108.4	116.2	88	0	39
RIVIERSONDEREND - SWELLENDAM	3373	304	9.01	118	109.7	120.4	93.6	0	45.3
SWELLENDAM - RIVERSDALE	4289	550	12.82	108.8	101.6	111.7	89.25	10.8	29.85
RIVERSDALE - MOSSELBAY	4309	405	9.40	105.2	100.2	107.4	84.5	17.9	21.7
MOSSELBAY - HARTENBOSCH	7331	486	6.63	89.5	89.4	90.2	80.4	12.7	14.4
HARTENBOSCH - GEORGE	7296	616	8.44	99.9	97	101.6	80.9	21	19.2
GEORGE - SEDGEFIELD	7480	583	7.79	104.35	100.95	105.9	85.4	25.55	20.85
SEDGEFIELD - KNYSNA	7663	549	7.16	108.8	104.9	110.2	89.9	30.1	22.5
KNYSNA - PLETTENBERG BAY	5315	606	11.40	95.2	89	98	73.8	0	6.8

2003									
strRouteDescription	ADT2003 (veh/d)	ADTT2003 (veh/d)	%Trucks	Average Speed (km/h)	Avg Night Speed (km/h)	Speed L Veh (km/h)	Speed H Veh (km/h)	Over 600 veh/h (%)	Over Speed Lim (%)
CAPE TOWN - SOMERSET WEST	51802	2506	4.84	99.3	99.2	100.2	82.2	96.2	14.9
STRAND - GRABOUW	17085	1150	6.73	90.8	89.9	91.9	74.8	87	
GRABOUW - CALEDON	6451	533	8.26	108.1	101.3	110.4	82.2	17.7	24.3
CALEDON - RIVIERSONDEREND	2958	359	12.14	116	106.6	119.8	88.6	0	42.3
RIVIERSONDEREND - SWELLENDAM	3450	474	13.74	116.5	107.4	120.5	91.7	0	43.7
SWELLENDAM - RIVERSDALE	5015	608	12.12	107.75	101	110.3	89	15.2	27.8
RIVERSDALE - MOSSELBAY	5937	671	11.30	100.8	96.4	102.7	85.4	18.9	27.2
MOSSELBAY - HARTENBOSCH	8409	775	9.22	88.6	89.8	89.6	78.3	54.3	10.8
HARTENBOSCH - GEORGE	8556	803	9.39	98.6	96.7	100.7	79.6	46.8	18.8
GEORGE - SEDGEFIELD	8392	794	9.46	94.55	91.4	96.75	74.4	47.45	19.1
SEDGEFIELD - KNYSNA	8227	785	9.54	90.5	86.1	92.8	69.2	48.1	19.4
KNYSNA - PLETTENBERG BAY	6021	705	11.71	94.3	89.6	96.8	75.3	2.5	6.2

2004									
strRouteDescription	ADT2004 (veh/d)	ADTT2004 (veh/d)	%Trucks	Average Speed (km/h)	Avg Night Speed (km/h)	Speed L Veh (km/h)	Speed H Veh (km/h)	Over 600 veh/h (%)	Over Speed Lim (%)
CAPE TOWN - SOMERSET WEST	52613	2521	4.79	99.1	99.5	100	82	96.3	15
STRAND - GRABOUW	17415	1249	7.17	88.4	90.2	89.6	74.1	87.2	
GRABOUW - CALEDON	6636	549	8.27	107.5	102.3	109.7	83.8	19.3	22.8
CALEDON - RIVIERSONDEREND	3047	370	12.14	116	106.6	119.8	88.6	0	42.3
RIVIERSONDEREND - SWELLENDAM	3554	488	13.73	116.5	107.4	120.5	91.7	0	43.7
SWELLENDAM - RIVERSDALE	5205	659	12.66	107.8	101.15	110.4	89.2	15.7	27.85
RIVERSDALE - MOSSELBAY	6115	691	11.30	100.8	96.4	102.7	85.4	18.9	27.2
MOSSELBAY - HARTENBOSCH	8661	798	9.21	88.6	89.8	89.6	78.3	54.3	10.8
HARTENBOSCH - GEORGE	8871	837	9.44	98.6	96.9	100.7	79.6	48.3	18.8
GEORGE - SEDGEFIELD	9489	881	9.28	97.95	95.8	100.15	76.95	61.75	19.95
SEDGEFIELD - KNYSNA	10107	924	9.14	97.3	94.7	99.6	74.3	75.2	21.1
KNYSNA - PLETTENBERG BAY	7701	826	10.73	90.3	88.2	92.2	74.9	30.6	3.3

4.2.3 Traffic and Speed Data: N7, Western Cape Province

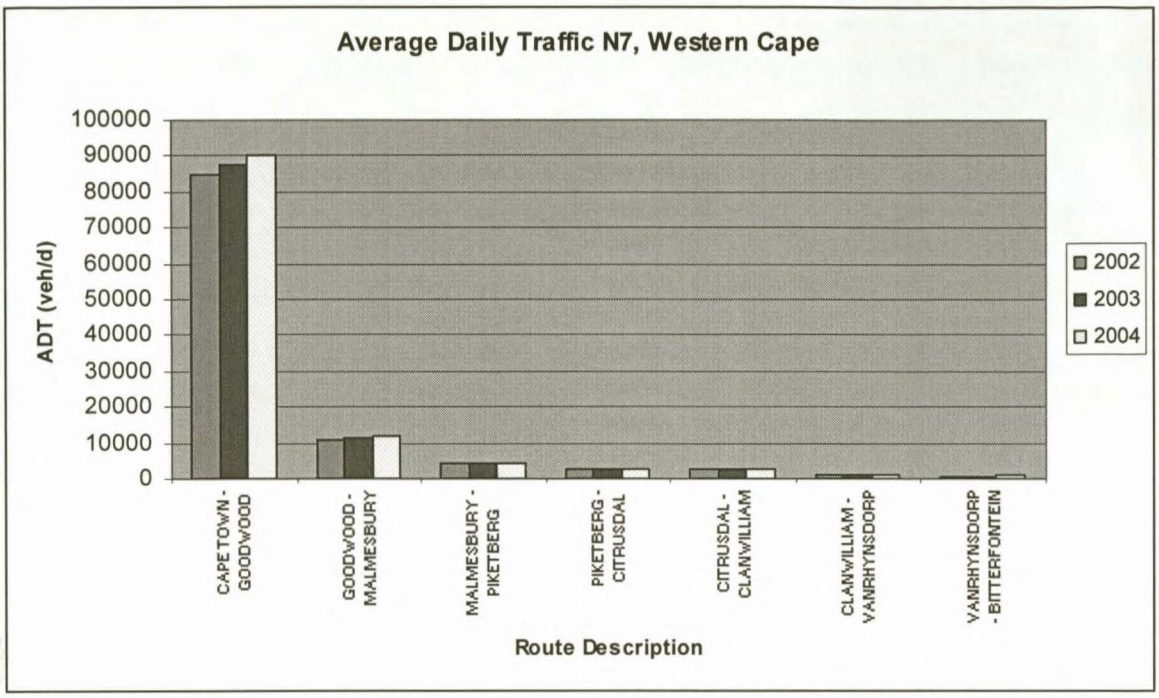

Fig. 4.2.11: Average Daily Traffic for N7, Western Cape Province 2002-2004

ADT point estimates are showing the expected tendency to decrease along the national road when moving in a northern direction along the N7 away from Cape Town into rural areas (Fig. 4.2.11).

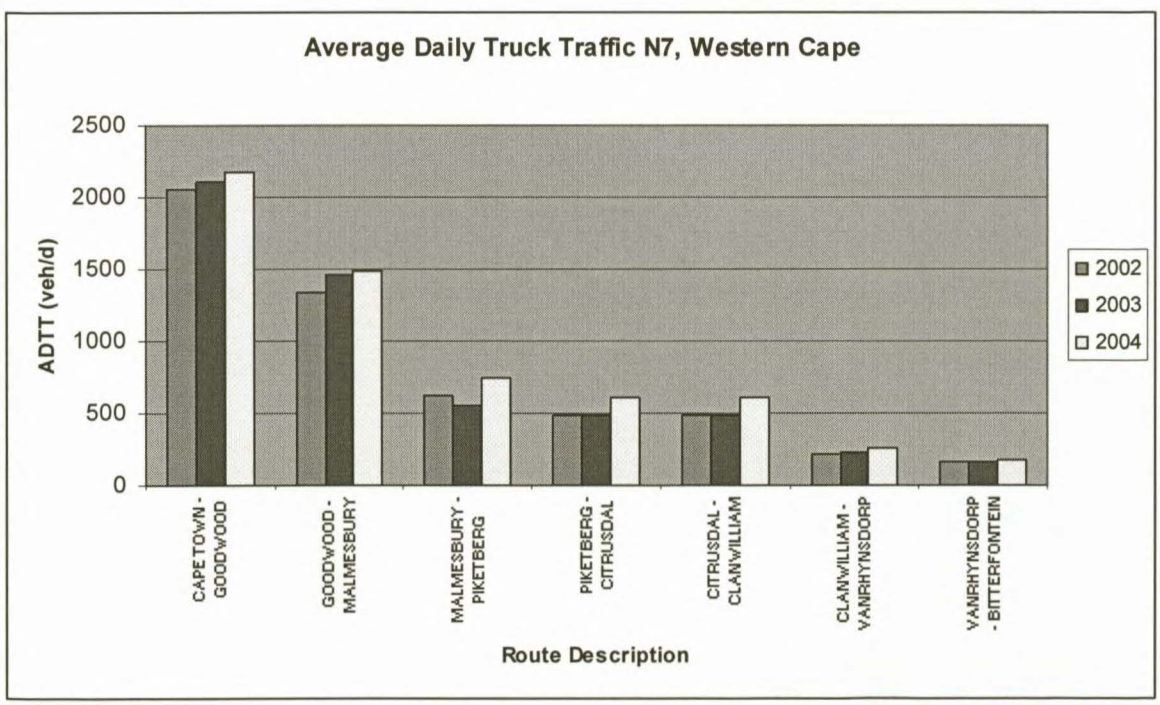

Fig. 4.2.12: Average Daily Truck Traffic for N7, Western Cape Province 2002-2004

A similar decreasing tendency for ADTT point estimates along the N7 is observed when moving away from Cape Town northwards into rural territory (Fig. 4.2.12).

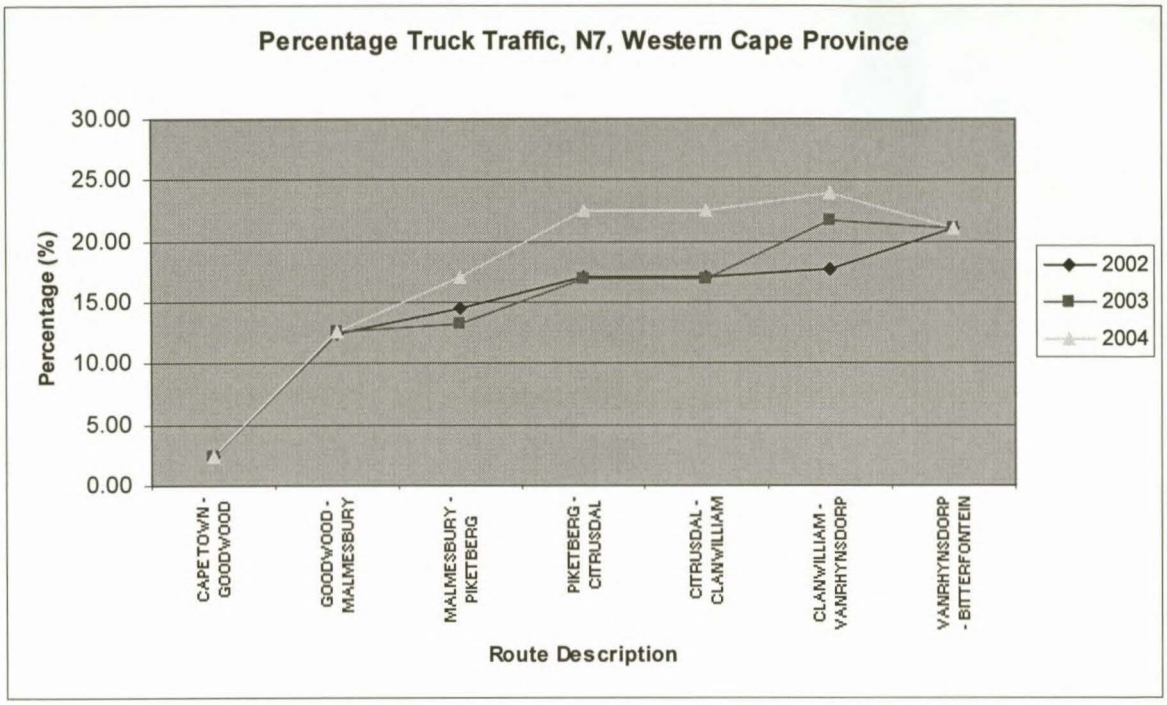

Fig. 4.2.13: Percentage Truck Traffic for N7, Western Cape Province 2002-2004

The same circumstances were observed for the N7 as for the N1 national road in terms of percentage truck traffic and the truck traffic volumes distribution along the road. From Fig. 4.2.12 it can be seen that truck traffic decreases along the N7, but there is a clear increase in the percentage truck traffic along the N7 (Fig. 4.2.13).

Average light vehicle speeds vary between 100km/h and 110km/h between Cape Town and Piketberg. Between Piketberg and Clanwilliam average light vehicle speeds remain constant at approximately 70km/h after which it increases to values between 110km/h and 120km/h between Clanwilliam and Bitterfontein.

Average heavy vehicle speeds vary between 80km/h and 90km/h between Cape Town and Piketberg. Between Piketberg and Clanwilliam average heavy vehicle speeds remain constant at approximately 45 km/h after which it increases to values between 85km/h and 90km/h.

Average night speeds remain constant at approximately 105km/h between Cape Town and Piketberg, at 65km/h between Piketberg and Clanwilliam and at 100km/h between Clanwilliam and Bitterfontein. Refer to Fig. 4.2.14 for the average speed distributions along the N7 within the Western Cape Province.

Between Cape Town and Piketberg between 20% and 25% of vehicles were observed to speed along the N7. About 10% of vehicles were observed to speed between Piketberg and Clanwilliam with between 30% and 35% of vehicles speeding between Clanwilliam and Bitterfontein. The reader is referred to Fig. 4.2.15.

The same observations are made for the N7 national road as before with the N1 and N2 in terms of the percentage vehicles speeding and the percentage vehicles in traffic flows over 600 veh/h (≡ ± 14400 veh/d) (Figures 4.2.14 and 4.2.15) and their relationships to the distribution of average speeds and traffic volumes along the road.

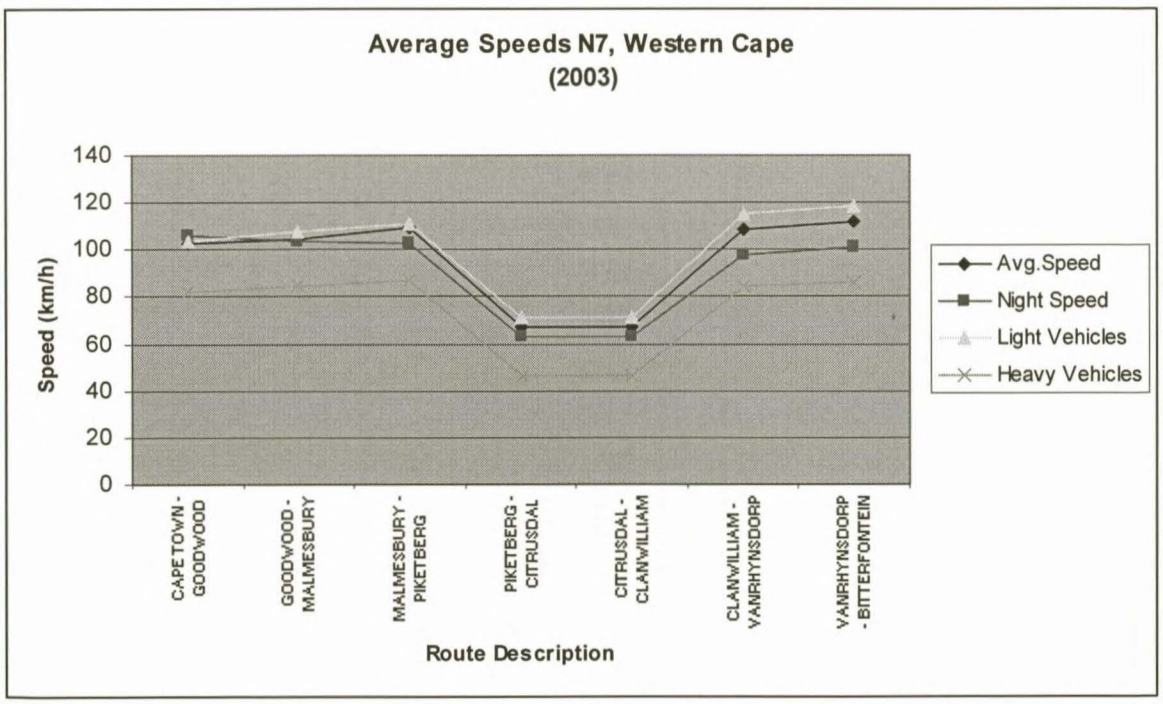

Fig. 4.2.14: Average Speeds for N7, Western Cape Province 2002-2004

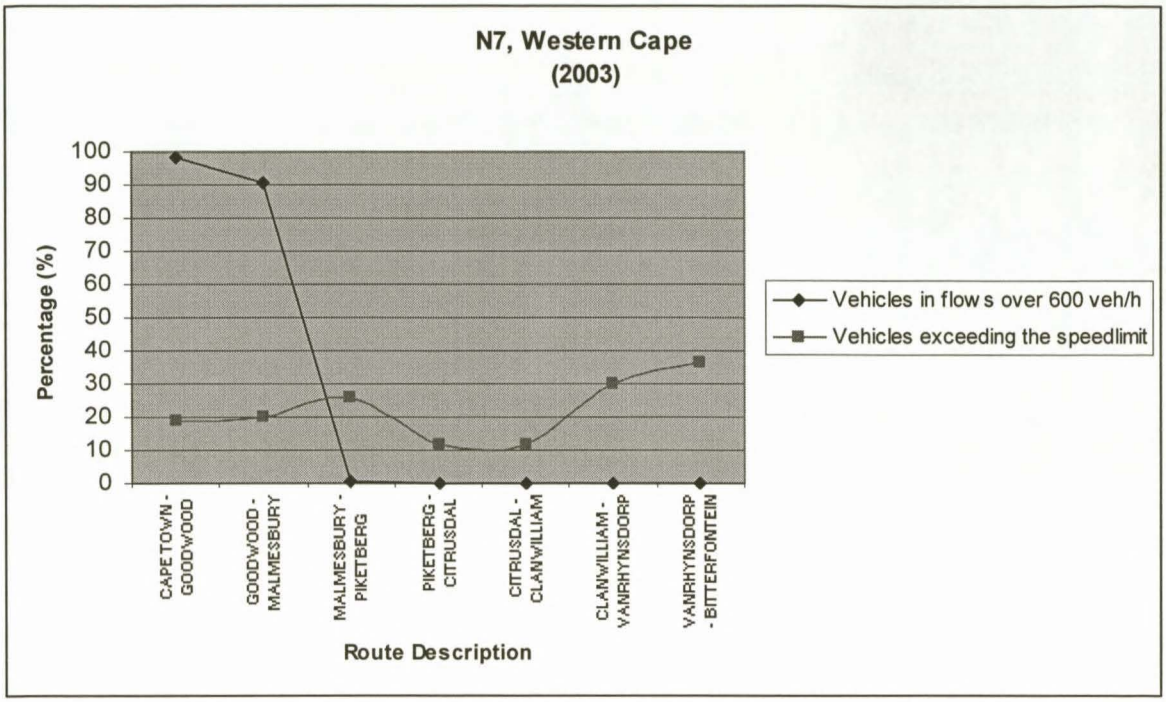

Fig. 4.2.15: Percentage Vehicles in flows over 600 veh/h and vehicles exceeding the speed limit for N7, Western Cape Province 2002-2004

Table 4.2.3 contains the data on which Figures 4.2.11, 4.2.12, 4.2.13, 4.2.14 and 4.2.15 are based.

Table 4.2.3: Traffic and Speed Data per National Road Section for the N7, Western Cape Province

2002									
strRouteDescription	ADT2002 (veh/d)	ADTT2002 (veh/d)	%Trucks	Average Speed (km/h)	Avg Night Speed (km/h)	Speed L Veh (km/h)	Speed H Veh (km/h)	Over 600 veh/h (%)	Over Speed Lim (%)
CAPE TOWN - GOODWOOD	84891	2053	2.42	102.8	106.2	103.3	82.1	98.4	18.7
GOODWOOD - MALMESBURY	10719	1342	12.52	105.7	102.9	108.8	84.6	93.5	22.9
MALMESBURY - PIKETBERG	4300	625	14.53	109	103	111	87	0.45	27.5
PIKETBERG - CITRUSDAL	2803	480	17.12	66.8	63	71.8	42.7	0	12.6
CITRUSDAL - CLANWILLIAM	2803	480	17.12	66.8	63	71.8	42.7	0	12.6
CLANWILLIAM - VANRHYNSDORP	1194	212	17.76	109.5	101.3	115.1	84.1	0	32.1
VANRHYNSDORP - BITTERFONTEIN	796	168	21.11	111.1	101	118	85.4	0	34.8

2003									
strRouteDescription	ADT2003 (veh/d)	ADTT2003 (veh/d)	%Trucks	Average Speed (km/h)	Avg Night Speed (km/h)	Speed L Veh (km/h)	Speed H Veh (km/h)	Over 600 veh/h (%)	Over Speed Lim (%)
CAPE TOWN - GOODWOOD	87438	2114	2.42	102.8	106.2	103.3	82.1	98.4	18.7
GOODWOOD - MALMESBURY	11499	1457	12.67	104.6	103.9	107.5	84.4	90.5	20
MALMESBURY - PIKETBERG	4150	550	13.25	109	103	111	87	0.55	26
PIKETBERG - CITRUSDAL	2831	481	16.99	67.1	62.9	71.5	46.4	0	12
CITRUSDAL - CLANWILLIAM	2831	481	16.99	67.1	62.9	71.5	46.4	0	12
CLANWILLIAM - VANRHYNSDORP	1072	232	21.64	108.4	97.8	115.2	84.1	0	30
VANRHYNSDORP - BITTERFONTEIN	782	165	21.10	111.8	101.2	118.8	85.8	0	36.2

2004									
strRouteDescription	ADT2004 (veh/d)	ADTT2004 (veh/d)	%Trucks	Average Speed (km/h)	Avg Night Speed (km/h)	Speed L Veh (km/h)	Speed H Veh (km/h)	Over 600 veh/h (%)	Over Speed Lim (%)
CAPE TOWN - GOODWOOD	90061	2178	2.42	102.8	106.2	103.3	82.1	98.4	18.7
GOODWOOD - MALMESBURY	11773	1482	12.59	103.3	101.9	105.9	84.4	90.1	16.6
MALMESBURY - PIKETBERG	4400	750	17.05	106	101	110	85	0.75	23
PIKETBERG - CITRUSDAL	2727	612	22.44	65.9	62.3	71.4	46.8	0	11.1
CITRUSDAL - CLANWILLIAM	2727	612	22.44	65.9	62.3	71.4	46.8	0	11.1
CLANWILLIAM - VANRHYNSDORP	1052	251	23.86	108.2	98.8	116.2	82.7	0	31.6
VANRHYNSDORP - BITTERFONTEIN	819	172	21.00	112.1	101.8	119	86.4	0	36.5

4.3 Geometric and Terrain Data

Geometric and Terrain information as gathered according to the methods described in Chapter 3 are discussed in this section as far as data was collected for the road sections under study along the N1, N2 and N7 of the Western Cape. No suitable source for geometric information could be found in time for the N7 for this research.

As information was gathered on only the national routes of the Western Cape Province, the results showed very similar values regarding lane and shoulder widths. These variables were thus not suitable for use in any regression analysis, because it is a case where there exists too little variation in the values. The number of lanes left and right of any of the road sections along the national roads was in some cases too diverse to determine average values for a particular road section and this variable was thus not used in any analysis.

The gathering of geometric information for the relevant road section was halted after it was determined that these variables would not provide any useful information regarding inclusion in a multiple regression analysis. The data is thus incomplete and will not be included in this document.

Terrain data was relatively easy to obtain, based on only three categories namely *Flat*, *Rolling* or *Mountainous*. For information on the different terrain types along the national routes, refer to Appendix D2.

The terrain variable could be included in a multiple regression analysis (discussed in sections to follow), but the geometric information was not suitable due to little variation in values and too few data points (as mentioned before). Due to focus on only the national roads in the Western Cape Province and the fact that no fatal accident featuring in the database could be traced to an exact location (exact coordinates and thus exact positions were not provided in the database), there were too few data points. If exact information on the road geometrics of a point where an accident occurred was available, more data points would have been possible to include in a regression analysis and then the geometric data variables would also have formed part of the collection of predictors.

4.4 Correspondence Analysis

Graphical representations of the analyses results are given in this section with cross tabulation tables of each variable pair's frequencies included in each analysis as queried from the MS Access database. The graphical representations were used as the main source for interpretations, although the relative row and column frequency tables created for each analysis could be used for verification in case of uncertainty. Where necessary, zoomed versions of the original graphs were provided where column and row points appeared to be too cluttered for easy inspection. These interpretations were very dependent on the "subjective eye" of the author. Therefore the reader is referred to the output statistics at the end of this document (Appendix B1) to make his/her own interpretation of the results, should he/she disagree with the results given in this document, bearing in mind all the principles described in Chapter 2.

The interpretation of the results was done according to the theory and application of the technique described in Chapter 2 and 3 respectively. Analysis output was interpreted for the period 2002-2004 for both the RSA and Western Cape Province. The results for the Western Cape Province and South Africa were then compared to observe any differences (if any) in correspondences between variables for each set of results. A two-dimensional solution was sufficient to investigate all cases where a solution with two or more than two dimensions was possible, but the *quality* of representation of the row and column points were not always consistent and in some instances led to inconclusive results. Interpretation results were nevertheless provided in spite of the latter. In some instances only a one-dimensional solution was possible and interpretation of the results was relatively simple, also the row and column points could be represented exactly with a one-dimensional solution.

The amount of *overall inertia* reproduced by each dimension is provided with comments on the *quality* of representation of the row and column points to give insight into the reliability of the interpretations from the graphs.

It was discussed previously how Correspondence Analysis is a useful technique when the correspondences between variables in large frequency tables need to be investigated. It is not particularly necessary to apply this technique to frequency tables with smaller dimensions, but it is still useful for providing a visual interpretation of the information contained in the frequency table. In this

study this technique was applied to relatively large as well as relatively small frequency tables for the sake of completion and demonstration.

The following types of variable pairs were investigated (as previously discussed):

- Type of accident vs. X
- Road User Type vs. X

where X denotes any particular variable with which the accident type or road user type variables were cross tabulated with.

Summary tables containing the results (the proposed "correspondences" between variables) as interpreted from the graphical representations appear following each set of graphical representations and analyses based on the given two types of variable pairs. Conclusions drawn from these results can be found in paragraph 4.4.3.

4.4.1 Correspondence between Accident Type and Variable X

This section provides the results of correspondence analyses performed on relative frequency tables cross tabulating the variable *Type of Accident* and the following categorical variables:

- Area Type
- Road Factor
- Vehicle Factor
- Vehicle Type
- Road User Type
- Gender
- Race Group
- Human Factor

The proposed correspondences as interpreted for each individual analysis on each variable pair can be found in Appendix B2. At the end of this section the final summary tables containing the proposed correspondences between accident types and all of the abovementioned variables will be provided.

i) Type of Accident vs. Area Type

The analysis extracted only one dimension for both datasets (South Africa and Western Cape Province) according to the definition for maximum number of dimensions as explained in Chapter 2. All the information (100%) contained in the relative frequency table could be reproduced by this one dimension. Only a one-dimensional plot could thus be created illustrating the correspondence between *Accident Type* and *Area Type*.

The row and column points could be represented exactly, which led to a *Quality* statistic equal to 1 for all row and column points.

Table 4.4.1(a): Cross tabulation – Type of Accident vs. Area Type, South Africa, 2002-2004

Accident Type	RURAL	URBAN	Total
Head-Rear end	1178	411	**1589**
Head on	1567	490	**2057**
Overturned	4428	835	**5263**
Pedestrian	5964	3793	**9757**
Hit and run	1426	971	**2397**
Total	**14563**	**6500**	**21063**

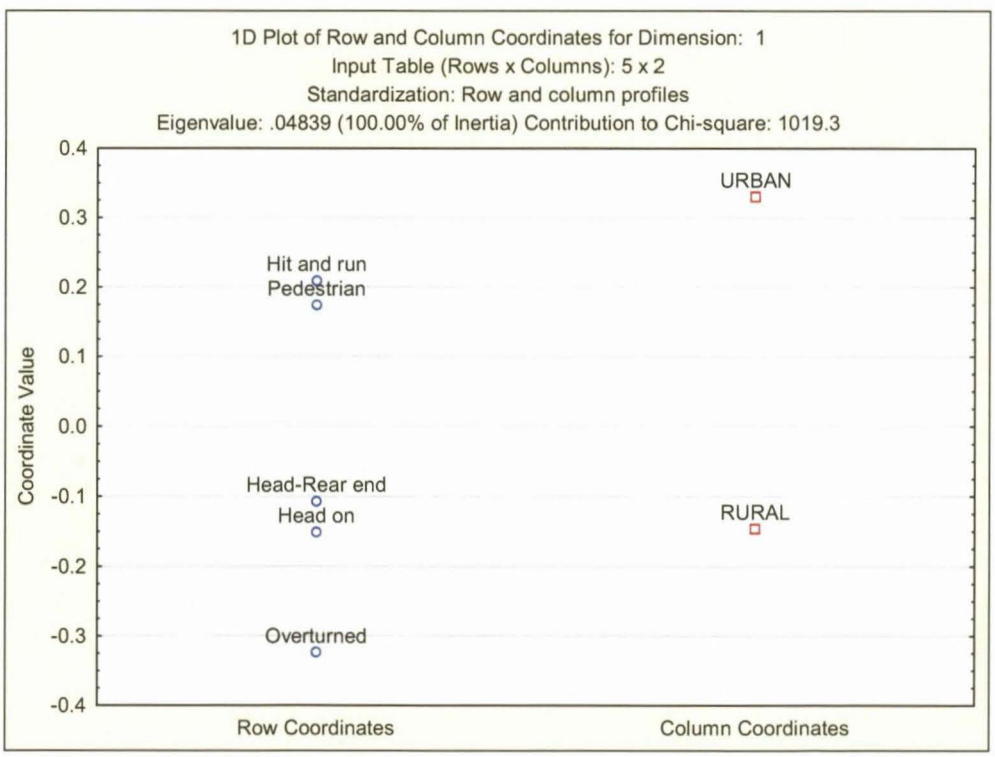

Fig. 4.4.1: One-dimensional solution – Correspondence between Accident Type and Area Type, South Africa

Table 4.4.1(b): Cross tabulation – Type of Accident vs. Area Type, Western Cape, 2002-2004

Accident Type	RURAL	URBAN	Total
Head-Rear end	101	55	**156**
Head on	148	59	**207**
Overturned	360	90	**450**
Pedestrian	679	513	**1192**
Collision - Fixed object	84	56	**140**
Hit and run	145	140	**285**
Total	**1517**	**913**	**2430**

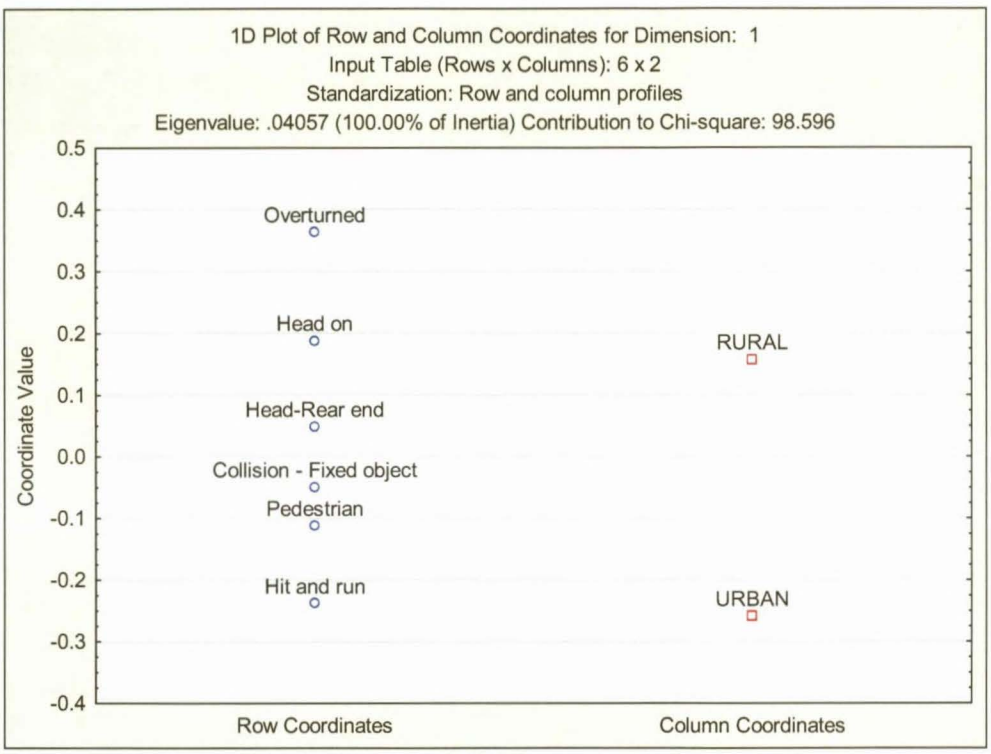

Fig. 4.4.2: One-dimensional solution – Correspondence between Accident Type and Area Type, Western Cape

ii) Type of Accident vs. Road Factor

The two-dimensional solution reproduced the *overall inertia* of the tables of relative frequencies in the following amounts:

For South Africa: 81.05% by Dimension 1 and 17.25%. by Dimension 2. For Western Cape Province: 49.95% by Dimension 1 and 21.61% by Dimension 2. The *Quality* statistic for every row and column point indicated a satisfactory representation of all column and row points. Refer to Appendix B1 for the output tables and statistics.

Table 4.4.2(a): Cross tabulation – Type of Accident vs. Road Factor, South Africa, 2002-2004

Road Factor	Head-Rear End	Head on	Overturned	Pedestrian	Hit and Run	Total
Poor visibility (Rain, mist, dust, smoke, dawn, dusk)	67	61	86	135	16	**365**
Poor street lighting	30	25	29	102	18	**204**
Sharp bend	13	55	231	19	1	**319**
Other	81	96	210	413	87	**887**
Road surface slippery / wet	20	27	60	29	2	**138**
Blind rise / Corner	9	24	31	15	0	**79**
Unknown	441	474	1323	2701	687	**5626**
Road works	5	8	9	15	3	**40**
Traffic light / Road sign / Road marking defective	3	3	8	3	1	**18**
Narrow road lane	7	11	19	13	0	**50**
Poor condition of road surface	3	14	118	14	1	**150**
Animals: Stray / Wild	3	4	19	4	0	**30**
Total	**682**	**802**	**2143**	**3463**	**816**	**7906**

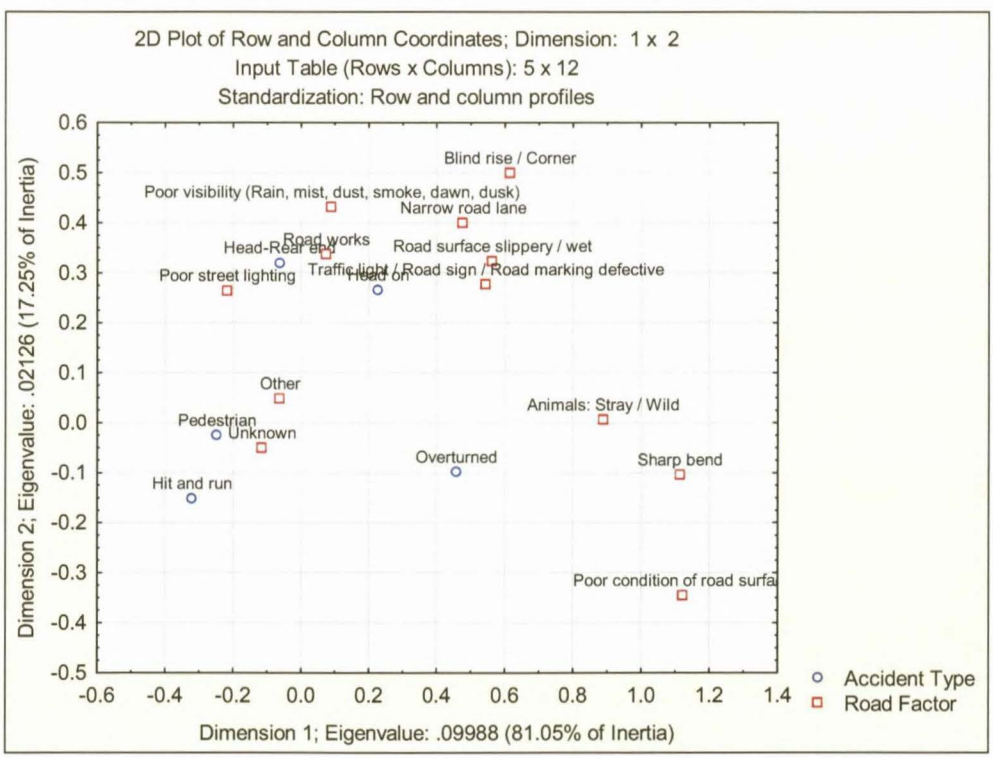

Fig. 4.4.3: Two-dimensional solution – Correspondence between Accident Type and Road Factor, RSA

Table 4.4.2(b): Cross tabulation – Type of Accident vs. Road Factor, Western Cape, 2002-2004

Road Factor	Head-Rear End	Head on	Overturned	Pedestrian	Collision - Fixed Object	Hit and Run	Total
Poor visibility (Rain, mist, dust, smoke, dawn, dusk)	7	3	6	9	2	0	**27**
Other	4	12	17	52	3	10	**98**
Road surface slippery / wet	5	1	6	2	2	1	**17**
Unknown	51	52	134	350	28	107	**722**
Road works	1	0	0	2	0	0	**3**
Blind rise / Corner	1	2	1	0	0	0	**4**
Sharp bend	1	4	11	2	6	1	**25**
Poor street lighting	1	3	3	10	1	2	**20**
Poor condition of road surface	0	0	5	1	0	0	**6**
Narrow road lane	0	0	1	2	0	0	**3**
Animals: Stray / Wild	0	0	0	1	0	0	**1**
Traffic light / Road sign / Road marking defective	0	0	0	0	1	0	**1**
Total	**71**	**77**	**184**	**431**	**43**	**121**	**927**

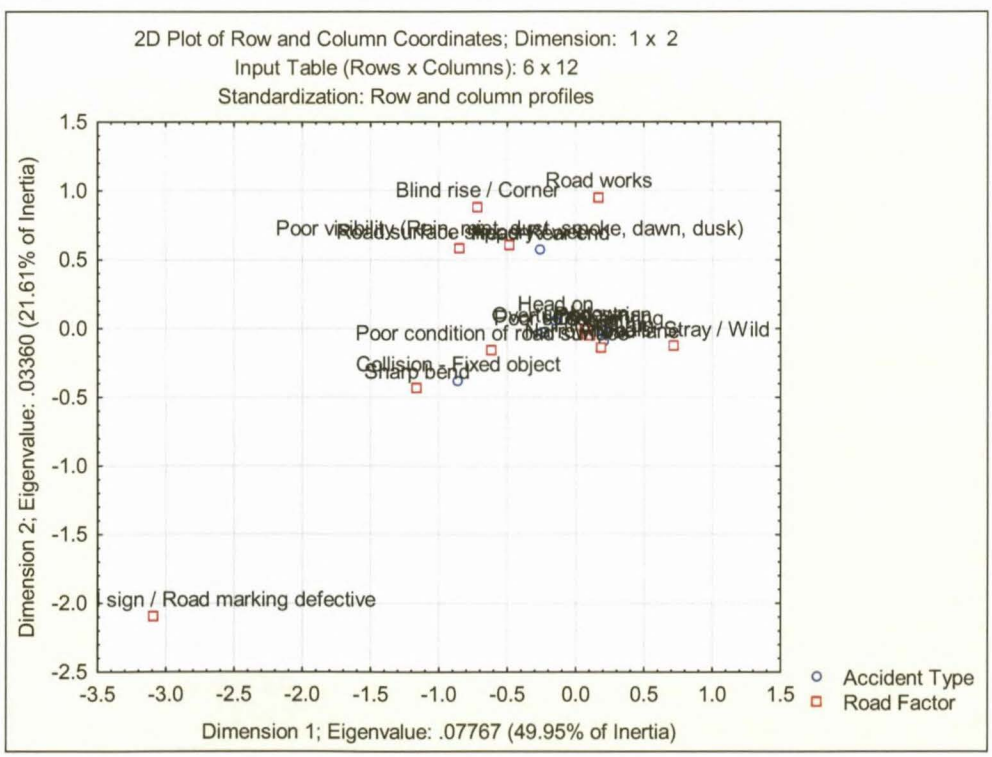

Fig. 4.4.4(a): Two-dimensional solution – Correspondence between Accident Type and Road Factor, Western Cape

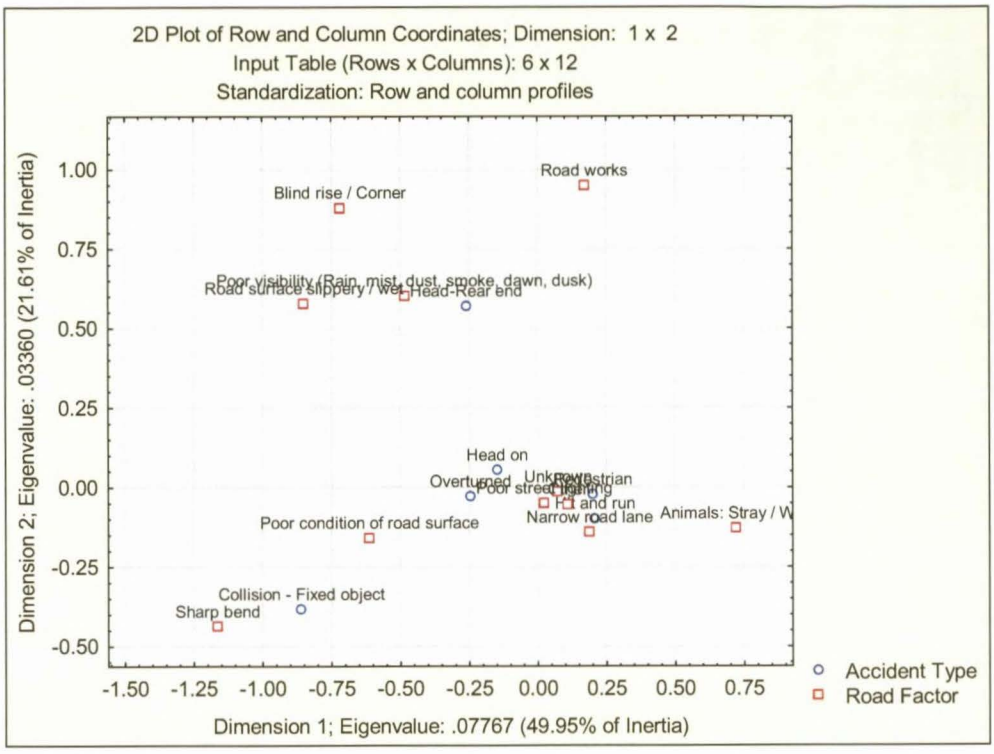

Fig. 4.4.4(b): Two-dimensional solution (zoomed) – Correspondence between Accident Type and Road Factor, Western Cape

iii) Type of Accident vs. Vehicle Factor

The two-dimensional solution reproduced the *overall inertia* of the tables of relative frequencies in the following amounts:

For South Africa: 89.02% by Dimension 1 and 8.33% by Dimension 2 (a total of 97.3%). For Western Cape Province: 82.30% by Dimension 1 and 8.94% by Dimension 2 (a total of 91.2%). Most row and column points were sufficiently represented as indicated by the *Quality* statistics for each row and column point. Refer to Appendix B1 for the output statistics and tables.

Table 4.4.3(a): Cross tabulation – Type of Accident vs. Vehicle Factor, South Africa, 2002-2004

Vehicle Factor	Head-Rear End	Head on	Overturned	Pedestrian	Hit and Run	Total
Chevrons: No reflective stripes	2	0	0	0	0	**2**
Unknown	485	556	1359	2814	753	**5967**
Other	74	71	118	433	45	**741**
Brakes: Faulty	24	15	121	44	1	**205**
Tyre burst prior to accident	17	53	717	5	0	**792**
Lights: Faulty, not switched on, blinding, etc	30	20	17	15	0	**82**
Overloading: Cargo / Passengers	11	11	61	11	0	**94**
Smooth tyres	1	3	14	9	0	**27**
Steering: Faulty	0	0	1	0	0	**1**
Total	**644**	**729**	**2408**	**3331**	**799**	**7911**

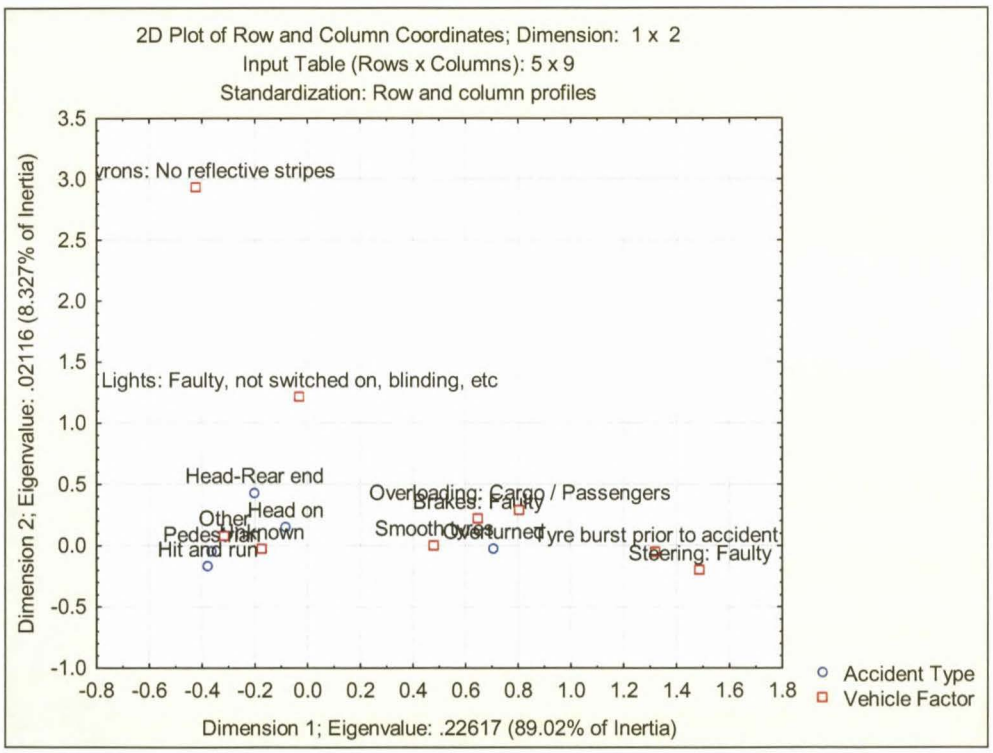

Fig. 4.4.5: Two-dimensional solution – Correspondence between Accident Type and Vehicle Factor, South Africa

Table 4.4.3(b): Cross tabulation – Type of Accident vs. Vehicle Factor, Western Cape, 2002-2004

Vehicle Factor	Head-Rear End	Head on	Overturned	Pedestrian	Collision - Fixed Object	Hit and Run	Total
Chevrons: No reflective stripes	1	0	0	0	0	0	**1**
Other	5	9	15	56	2	2	**89**
Lights: Faulty, not switched on, blinding, etc	2	2	2	1	0	0	**7**
Brakes: Faulty	3	2	12	2	0	0	**19**
Unknown	53	58	117	356	31	115	**730**
Tyre burst prior to accident	4	5	68	1	8	0	**86**
Overloading: Cargo / Passengers	0	0	4	2	0	0	**6**
Smooth tyres	0	0	1	0	0	0	**1**
Total	**68**	**76**	**219**	**418**	**41**	**117**	**939**

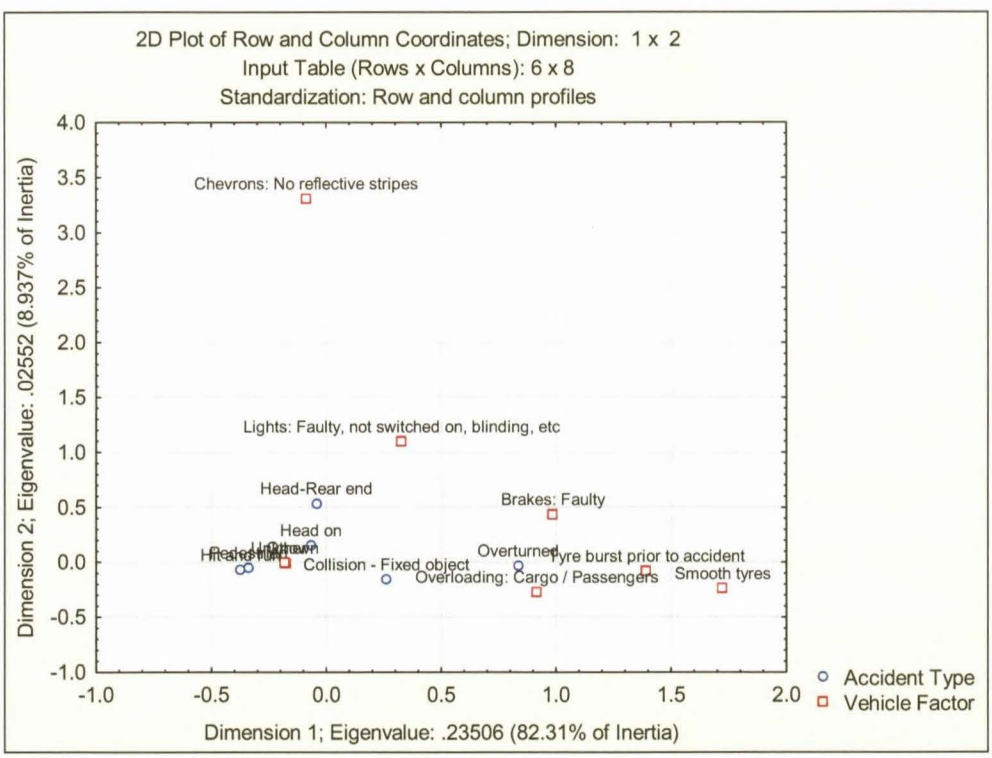

Fig. 4.4.6: Two-dimensional solution – Correspondence between Accident Type and Vehicle Factor, Western Cape

iv) Type of Accident vs. Vehicle Type

The two-dimensional solution reproduced the *overall inertia* of the tables of relative frequencies in the following amounts:

For South Africa: 91.03% by Dimension 1 and 7.23% by Dimension 2 (total of 98.30%). For Western Cape Province: 88.09% by Dimension 1 and 6.18% by Dimension 2 (total of 94.27%). Refer to Appendix B1 for the *Quality* statistic values for each row and column point. The *quality* of representation of each row and column point varies to such an extent that no "to-the-point" conclusion can be drawn. The reader is invited to use his/her own discretion in this regard, but an attempt was still made to interpret the results.

Table 4.4.4(a): Cross tabulation – Type of Accident vs. Vehicle Type, South Africa, 2002-2004

Vehicle Type	Head-Rear end	Head on	Overturned	Pedestrian	Hit and Run	Total
Heavy vehicle	665	582	375	796	5	**2423**
Sedan	1492	2160	2509	4958	37	**11156**
Tractor	44	8	105	64	0	**221**
LDV / Bakkie	632	970	1510	2193	10	**5315**
Bicycle	266	17	5	14	25	**327**
Other	20	16	7	47	1	**91**
Minibus	146	200	331	638	7	**1322**
Motorcycle	72	59	87	45	3	**266**
Minibus Taxi	161	203	355	832	7	**1558**
Bus	60	71	57	228	3	**419**
Unknown	32	23	23	104	2330	**2512**
Panel van	0	0	0	2	0	**2**
Total	**3590**	**4309**	**5364**	**9921**	**2428**	**25612**

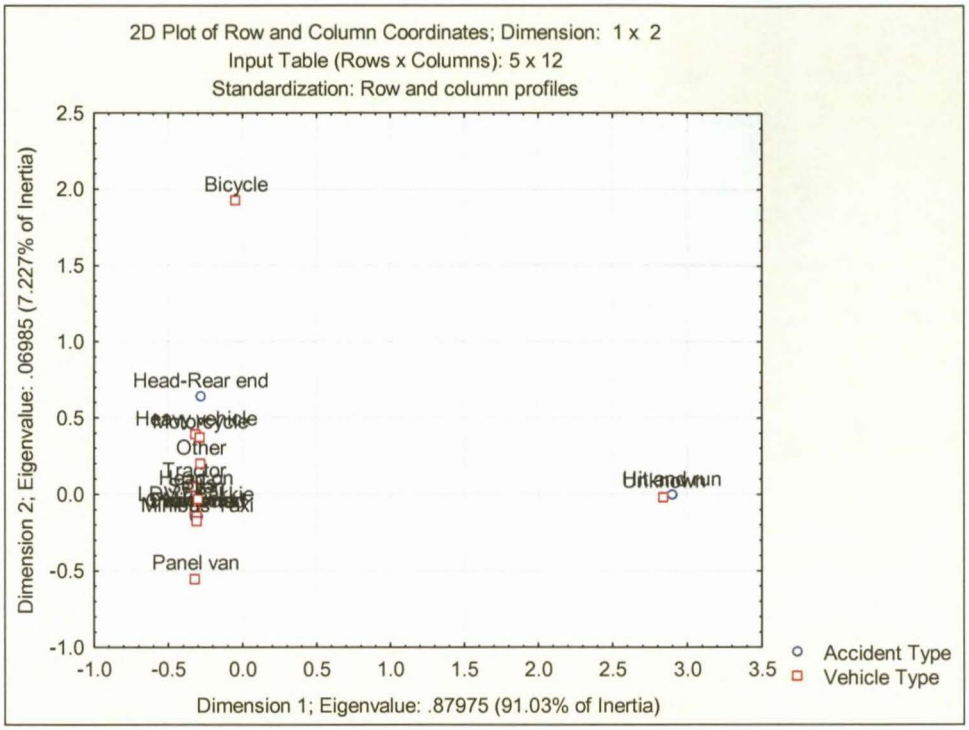

Fig. 4.4.7(a): Two-dimensional solution – Correspondence between Accident Type and Vehicle Type, South Africa

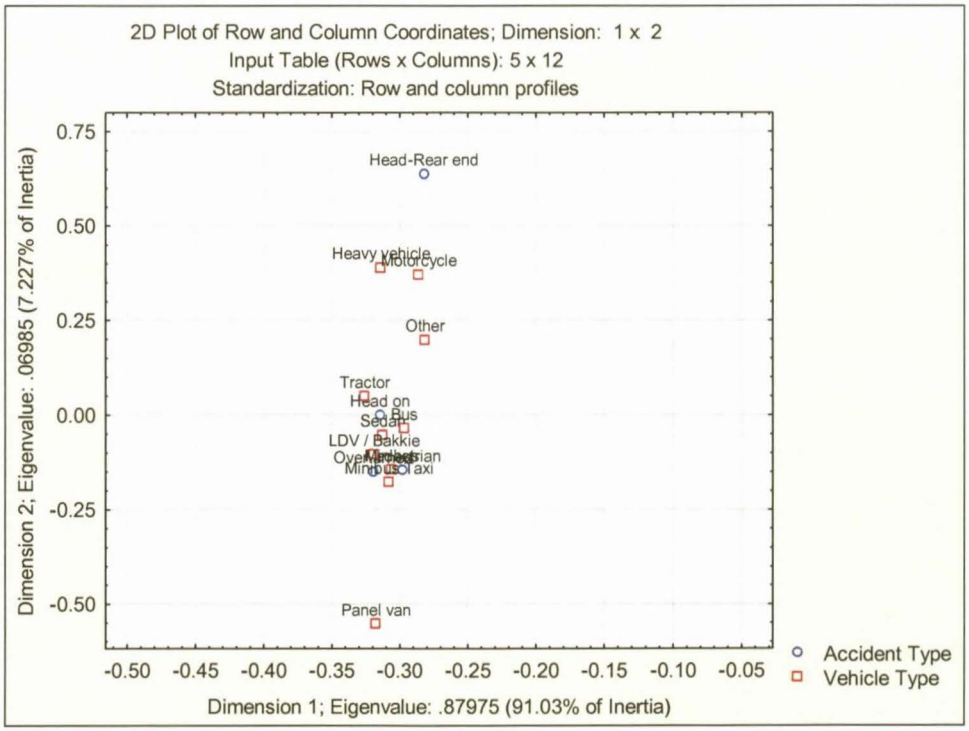

Fig. 4.4.7(b): Two-dimensional solution – Correspondence between Accident Type and Vehicle Type, South Africa

Table 4.4.4(b): Cross tabulation – Type of Accident vs. Vehicle Type, Western Cape, 2002-2004

Vehicle Type	Head-Rear End	Head on	Overturned	Pedestrian	Collision - Fixed Object	Hit and Run	Total
Tractor	5	1	7	3	1	0	**17**
Sedan	177	228	233	707	96	6	**1447**
Bicycle	23	0	1	2	0	4	**30**
LDV / Bakkie	52	101	106	255	13	0	**527**
Minibus	12	16	37	68	5	0	**138**
Heavy vehicle	62	48	29	102	6	0	**247**
Motorcycle	12	8	11	9	17	0	**57**
Minibus Taxi	8	12	26	33	0	1	**80**
Unknown	2	6	3	13	2	278	**304**
Other	1	2	0	4	0	0	**7**
Bus	2	13	3	16	0	0	**34**
Total	**356**	**435**	**456**	**1212**	**140**	**289**	**2888**

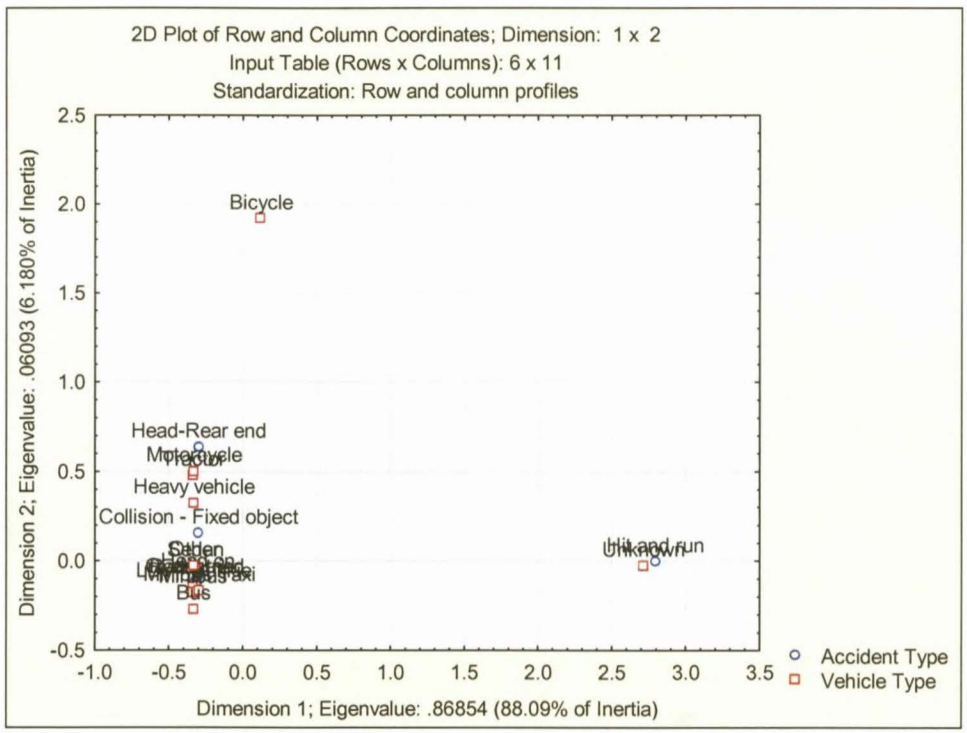

Fig. 4.4.8(a): Two-dimensional solution – Correspondence between Accident Type and Vehicle Type, Western Cape

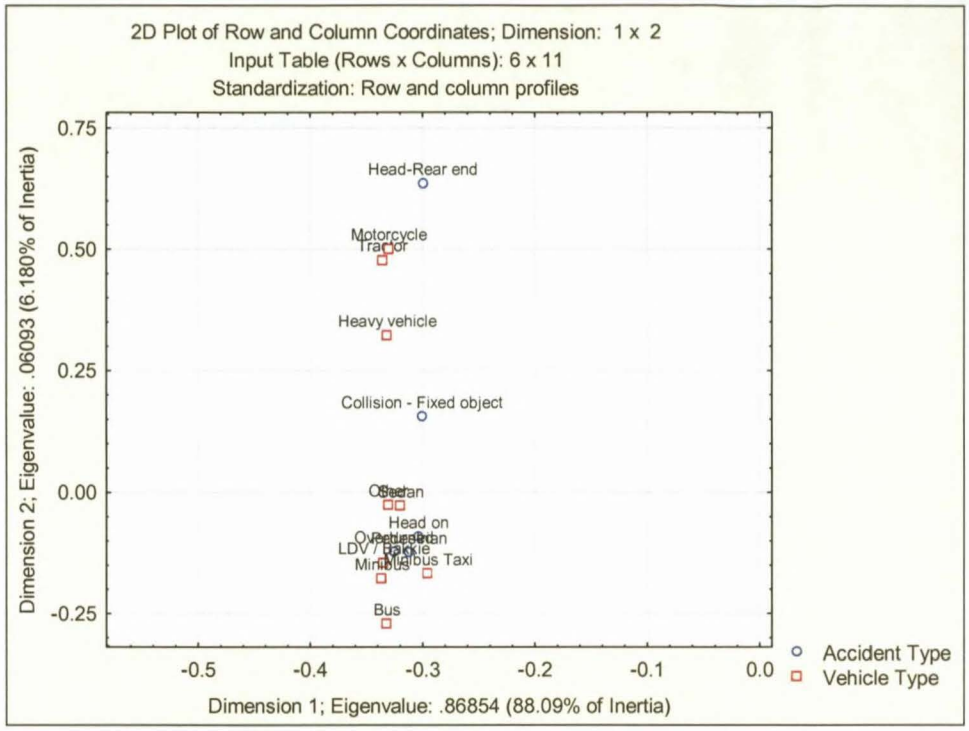

Fig. 4.4.8(b): Two-dimensional solution – Correspondence between Accident Type and Vehicle Type, Western Cape

v) Type of Accident vs. Road User Type (Fatalities)

The same correspondences were achieved for South Africa and the Western Cape Province. The two-dimensional solution reproduced 98.89% of the *overall inertia* via the first dimension and 1.11% via the second dimension for the results based on the South African data. 97.24% of the *overall inertia* was reproduced by Dimension 1 and 2.76% reproduced by Dimension 2 for the results based on the Western Cape Province data. Refer to Appendix B1 for the relevant output tables indicating the *quality* of representation of each row and column point for both sets of analyses. Only two dimensions were possible which leads to the two-dimensional solution reproducing a total of 100% of the *overall inertia* for the South African results and the Western Cape Province results.

Table 4.4.5(a): Cross tabulation – Type of Accident vs. Road User Type (Fatalities), South Africa, 2002-2004

Accident Type	DRIVER	PASSENGER	PEDESTRIAN	Total
Head-Rear end	1180	1003	28	**2211**
Head on	1750	1942	12	**3704**
Overturned	2766	4094	32	**6892**
Pedestrian	46	52	9937	**10035**
Hit and run	26	5	2386	**2417**
Total	**5768**	**7096**	**12395**	**25259**

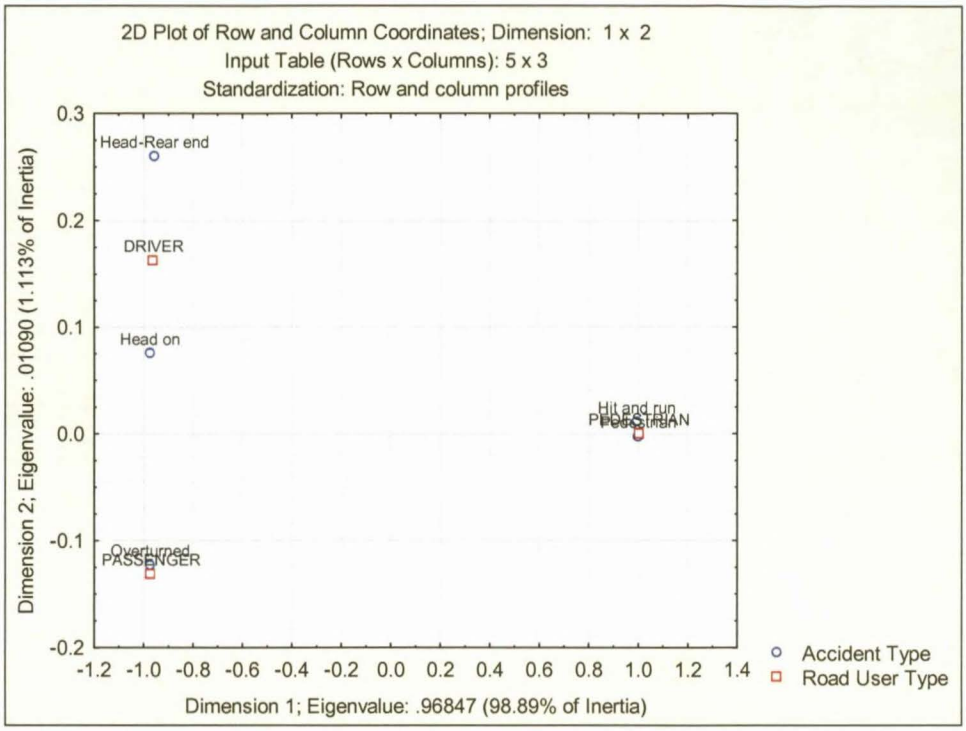

Fig. 4.4.9: Two-dimensional solution – Correspondence between Accident Type and Road User Type, South Africa

Table 4.4.5(b): Cross tabulation – Type of Accident vs. Road User Type (Fatalities), Western Cape, 2002-2004

Accident Type	DRIVER	PASSENGER	PEDESTRIAN	Total
Head-Rear end	124	95	3	**222**
Head on	189	176	0	**365**
Overturned	224	362	3	**589**
Pedestrian	2	5	1204	**1211**
Collision - Fixed object	95	69	5	**169**
Hit and run	4	0	282	**286**
Total	**638**	**707**	**1497**	**2842**

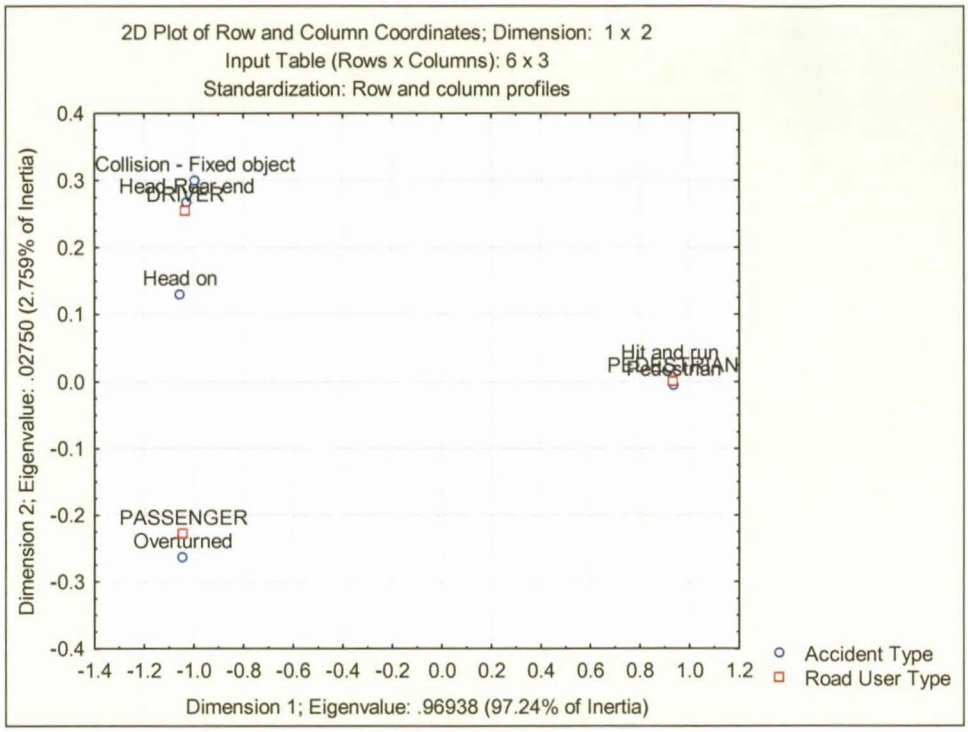

Fig. 4.4.10: Two-dimensional solution – Correspondence between Accident Type and Road User Type, Western Cape

vi) Type of Accident vs. Gender

Only two dimensions were possible for this analysis for both South Africa and Western Cape Province. The two-dimensional solution thus reproduces 100% of the *overall inertia* of the relative frequency tables. Refer to Appendix B1 for the relevant output statistics.

Table 4.4.6(a): Cross tabulation – Type of Accident vs. Gender, South Africa, 2002-2004

Accident Type	MALE	FEMALE	UNKNOWN	Total
Head-Rear end	1687	477	47	**2211**
Head on	2664	942	98	**3704**
Overturned	5098	1734	60	**6892**
Pedestrian	7372	2576	87	**10035**
Hit and run	1942	443	32	**2417**
Total	**18763**	**6172**	**324**	**25259**

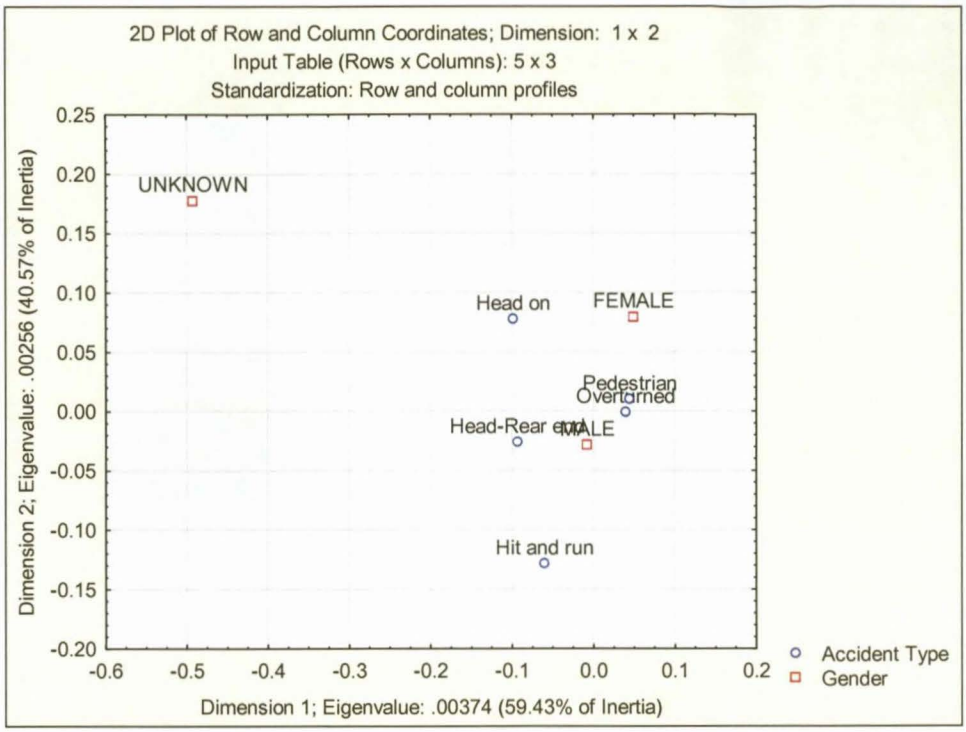

Fig. 4.4.11: Two-dimensional solution – Correspondence between Accident Type and Gender, South Africa

Table 4.4.6(b): Cross tabulation – Type of Accident vs. Gender, Western Cape, 2002-2004

Accident Type	MALE	FEMALE	UNKNOWN	Total
Head-Rear end	168	43	11	**222**
Head on	256	98	11	**365**
Overturned	434	139	16	**589**
Pedestrian	910	283	18	**1211**
Collision - Fixed object	139	28	2	**169**
Hit and run	237	46	3	**286**
Total	**2144**	**637**	**61**	**2842**

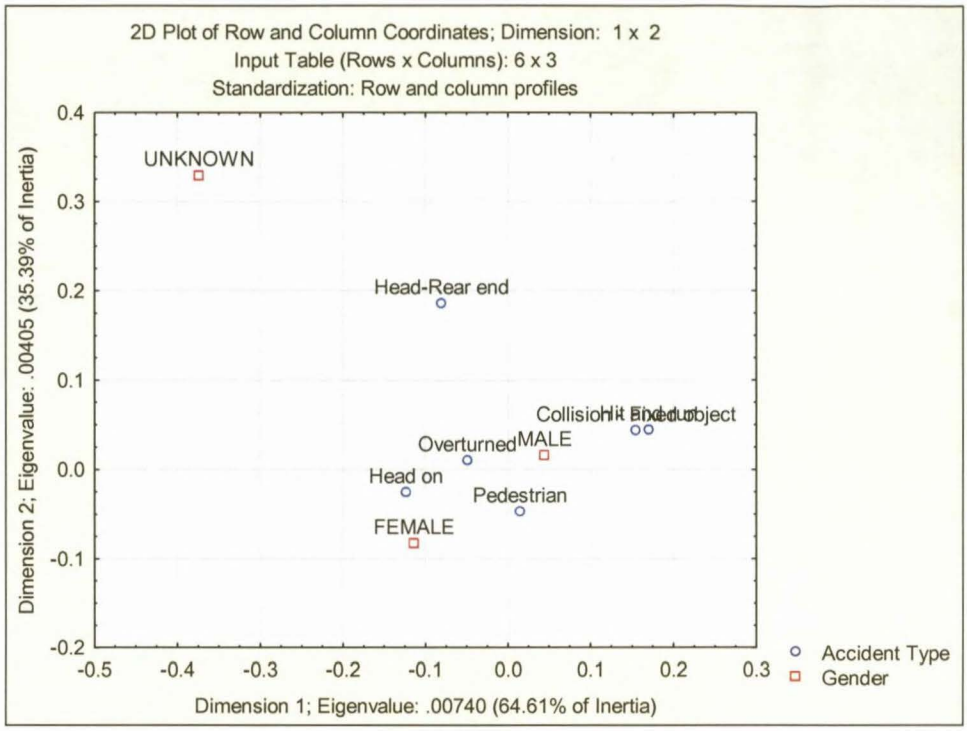

Fig. 4.4.12: Two-dimensional solution – Correspondence between Accident Type and Gender, Western Cape

vii) Type of Accident vs. Race Group

The results showed a satisfactory representation of the rows and column points for both datasets, South Africa as well as Western Cape Province. For South Africa, the amount of *overall inertia* reproduced by the first two dimensions is: 97.34% for Dimension 1 and 2.32% for Dimensions 2 (total of 99.66%). For Western Cape Province the amount of *overall inertia* reproduced by the first two dimensions is: 88.81% for Dimension 1 and 7.96% for Dimension 2 (total of 96.77%). Refer to Appendix B1 for the relevant output statistics.

Table 4.4.7(a): Cross tabulation – Type of Accident vs. Race Group, South Africa, 2002-2004

Population Group	Head-Rear End	Head on	Overturned	Pedestrian	Hit and Run	Total
BLACK	1475	2355	4787	8451	2076	**19144**
COLOURED	177	307	666	1147	238	**2535**
WHITE	456	820	1178	251	49	**2754**
ASIAN	87	153	219	103	19	**581**
UNKNOWN	16	69	41	83	35	**244**
FOREIGNER	0	0	1	0	0	**1**
Total	**2211**	**3704**	**6892**	**10035**	**2417**	**25259**

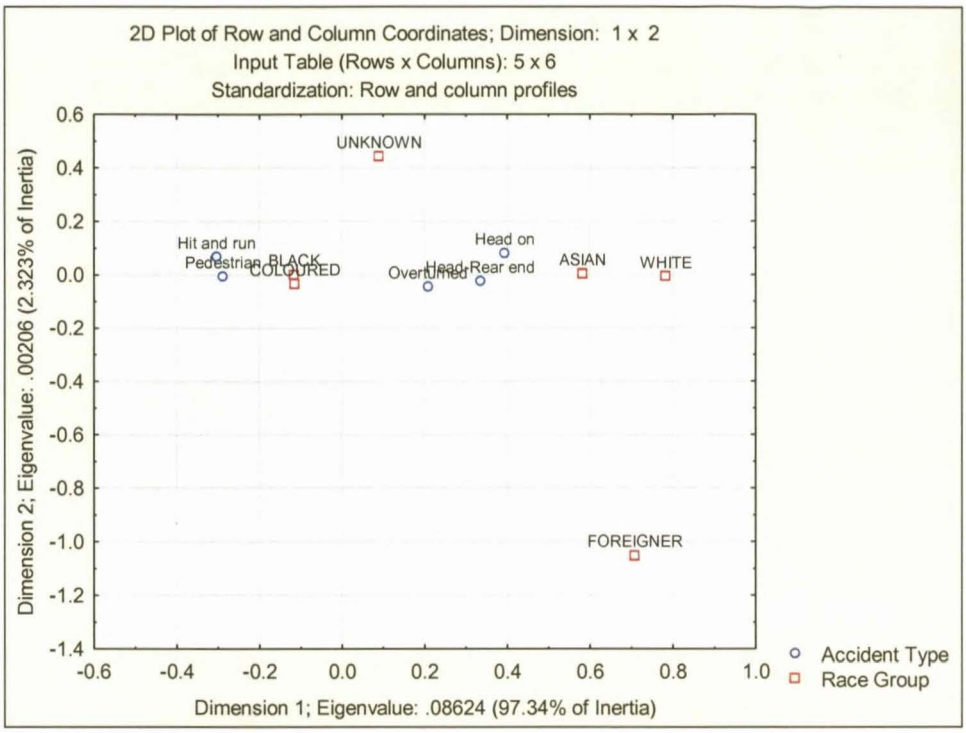

Fig. 4.4.13: Two-dimensional solution – Correspondence between Accident Type and Race Group, South Africa

Table 4.4.7(b): Cross tabulation – Type of Accident vs. Race Group, Western Cape, 2002-2004

Population Group	Head-Rear End	Head on	Overturned	Pedestrian	Collision - Fixed Object	Hit and Run	Total
COLOURED	100	160	275	778	80	150	**1543**
BLACK	61	76	185	368	24	122	**836**
WHITE	53	117	111	39	61	9	**390**
ASIAN	6	4	14	7	2	0	**33**
UNKNOWN	2	8	4	19	2	5	**40**
Total	**222**	**365**	**589**	**1211**	**169**	**286**	**2842**

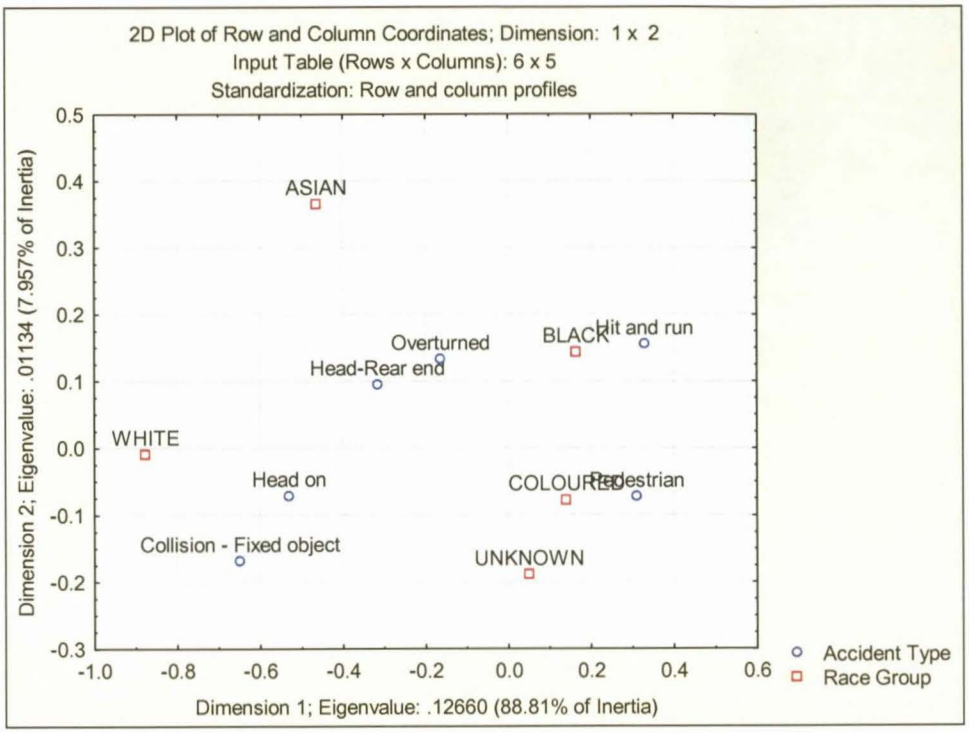

Fig. 4.4.14: Two-dimensional solution – Correspondence between Accident Type and Race Group, Western Cape Province

viii) Type of Accident vs. Human Factor

The same correspondences were found in the analysis output for South Africa and the Western Cape Province. The two-dimensional solution reproduced the *overall inertia* of the relative frequency tables as follow:

For South Africa: 45.67% via Dimension 1 and 35.63% via Dimension 2 (total of 81.30%). For Western Cape Province: 43.76% via Dimension 1 and 32.49% via Dimension 2 (total of 76.24%). The representativity of this solution is not as satisfactory in terms of the *Quality* statistic for each row and column point, but the results are nevertheless interpreted as far as possible. Please refer to Appendix B1 for the output tables for this analysis and relevant output statistics.

Table 4.4.8(a): Cross tabulation – Type of Accident vs. Human Factor, South Africa, 2002-2004

Human Factor	Head-Rear End	Head on	Overturned	Pedestrian	Hit and Run	Total
Speed too high for circumstances	885	735	3377	238	24	**5259**
Followed too closely	150	5	8	1	1	**165**
Other	35	16	121	47	12	**231**
Not Known	101	103	470	79	496	**1249**
Fatigue / Driver falling asleep	35	81	209	8	0	**333**
Overtook when unlawful / unsafe	64	560	53	34	0	**711**
Disregarded red traffic light / stop sign / yield sign	74	69	12	29	1	**185**
Intoxicated Driver: Use of liquor or drugs suspected	52	97	217	69	2	**437**
Turned in front of oncoming traffic	34	221	16	9	0	**280**
Cell phone use / holding	1	5	5	1	0	**12**
U-turn	3	1	0	0	0	**4**
Pedestrian: Jay walking	2	10	17	8959	352	**9340**
Intoxicated Pedestrian: Use of liquor or drugs suspected	1	0	0	240	16	**257**
Hit-and-run	1	0	3	2	1458	**1464**
Total	**1438**	**1903**	**4508**	**9716**	**2362**	**19927**

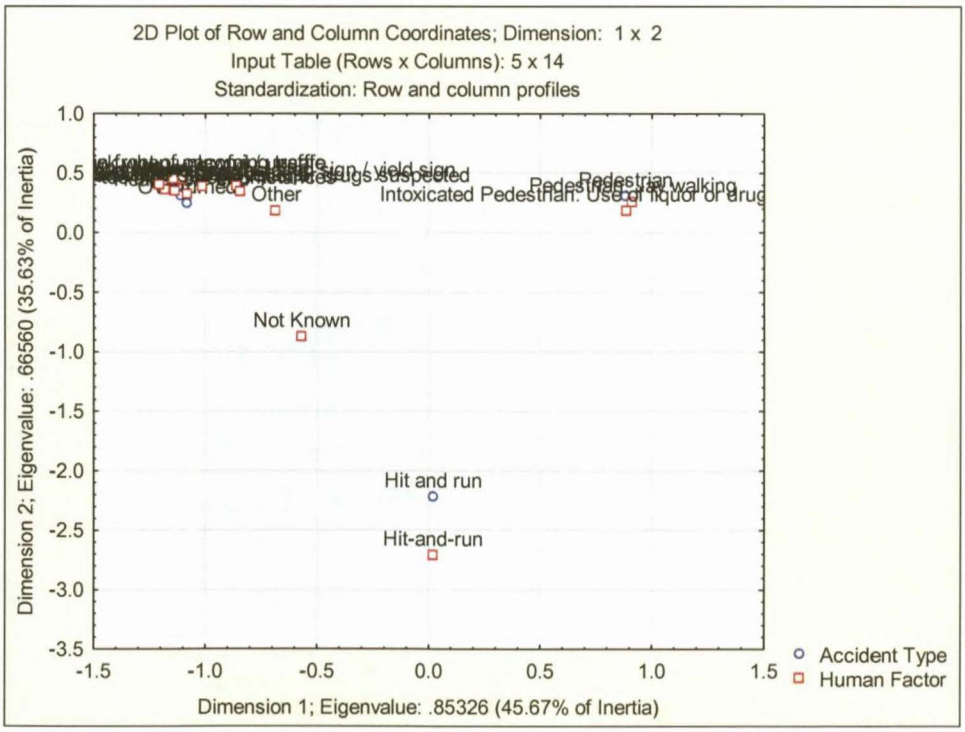

Fig. 4.4.15(a): Two-dimensional solution – Correspondence between Accident Type and Human Factor, South Africa

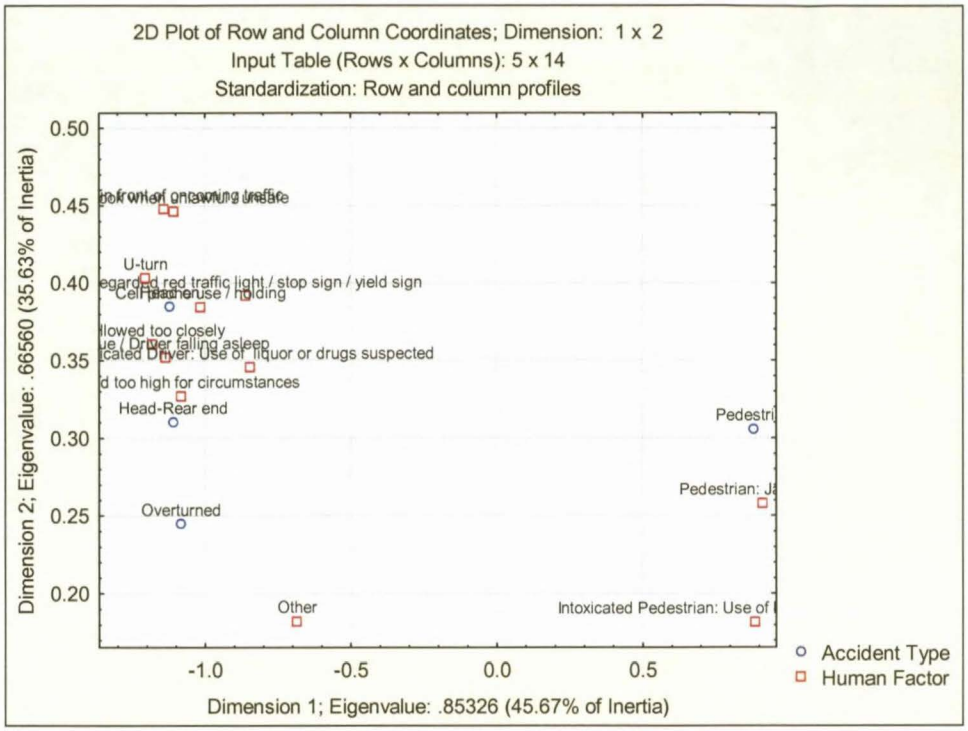

Fig. 4.4.15(b): Two-dimensional solution (zoomed) – Correspondence between Accident Type and Human Factor, South Africa

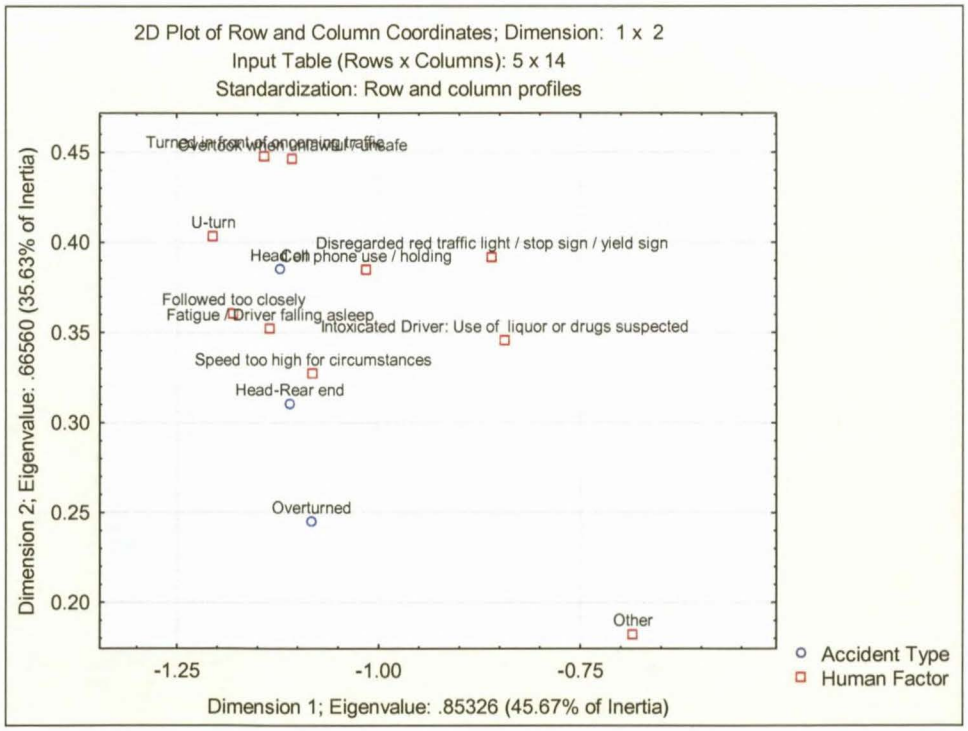

Fig. 4.4.15(c): Two-dimensional solution (zoomed) – Correspondence between Accident Type and Human Factor, South Africa

Table 4.4.8(b): Cross tabulation – Type of Accident vs. Human Factor, Western Cape, 2002-2004

Human Factor	Head-Rear End	Head on	Overturned	Pedestrian	Collision - Fixed Object	Hit and Run	Total
Speed too high for circumstances	91	61	291	21	91	0	**555**
Other	3	2	9	6	2	2	**24**
Overtook when unlawful / unsafe	6	68	5	4	1	0	**84**
Followed too closely	14	0	0	0	0	0	**14**
Fatigue / Driver falling asleep	2	7	21	1	12	0	**43**
Disregarded red traffic light / stop sign / yield sign	7	4	0	4	2	0	**17**
Turned in front of oncoming traffic	3	35	1	1	0	0	**40**
Intoxicated Driver: Use of liquor or drugs suspected	6	11	21	7	9	0	**54**
Not Known	9	8	40	13	7	74	**151**
Hit-and-run	1	0	1	0	0	156	**158**
Pedestrian: Jay walking	0	0	2	1092	0	45	**1139**
Intoxicated Pedestrian: Use of liquor or drugs suspected	0	0	0	43	0	5	**48**
Total	**142**	**196**	**391**	**1192**	**124**	**282**	**2327**

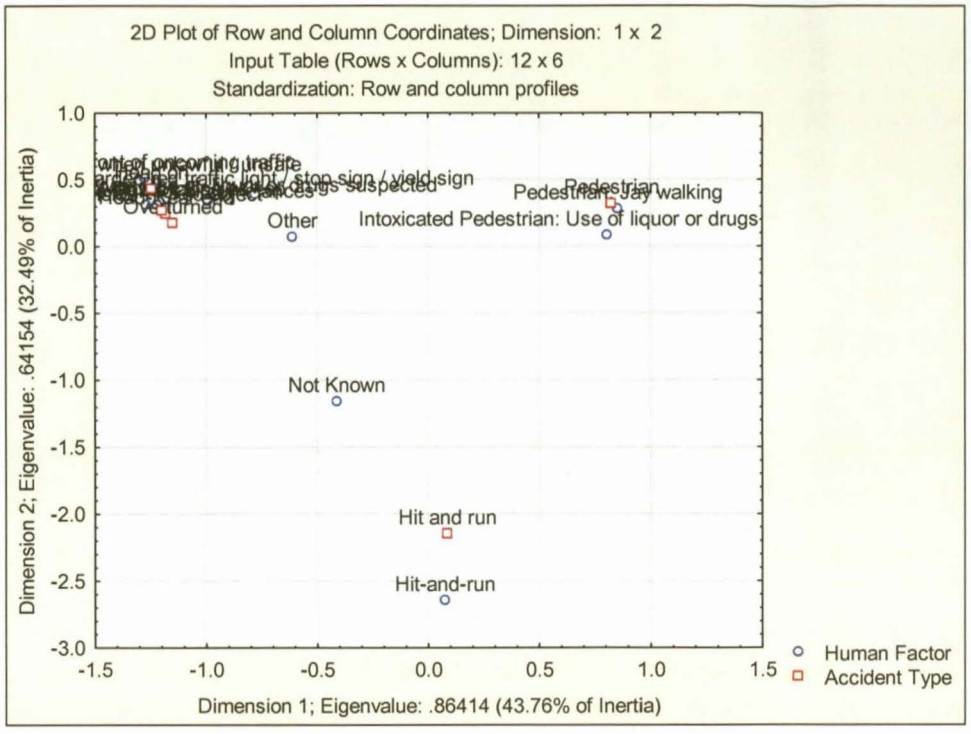

Fig. 4.4.16(a): Two-dimensional solution – Correspondence between Accident Type and Human Factor, Western Cape Province

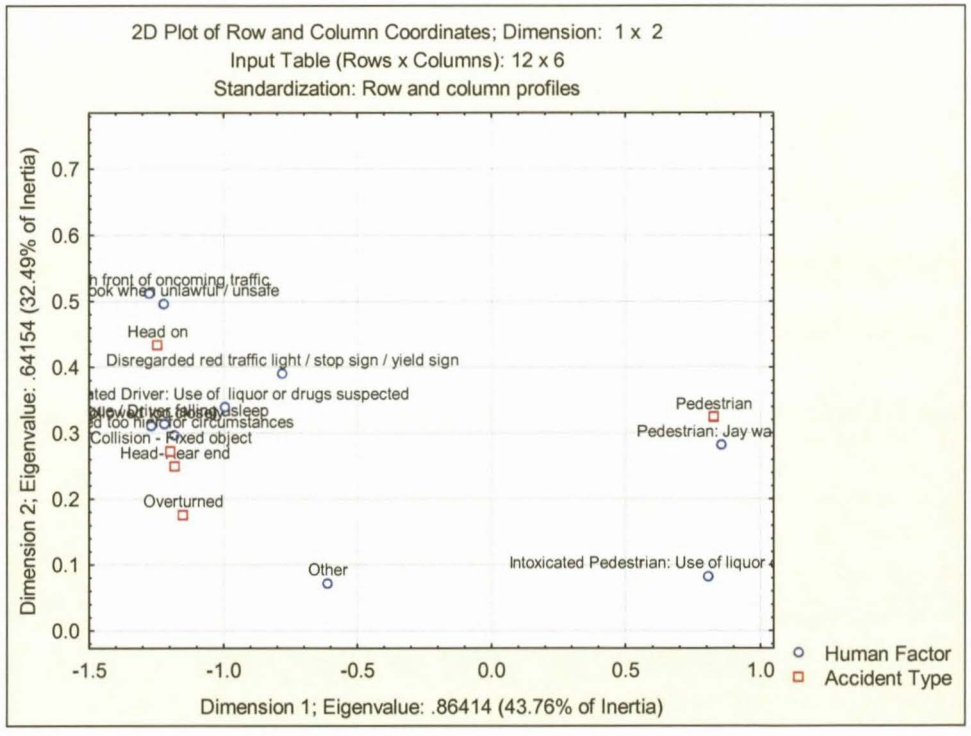

Fig. 4.4.16(b): Two-dimensional solution (zoomed) – Correspondence between Accident Type and Human Factor, Western Cape Province

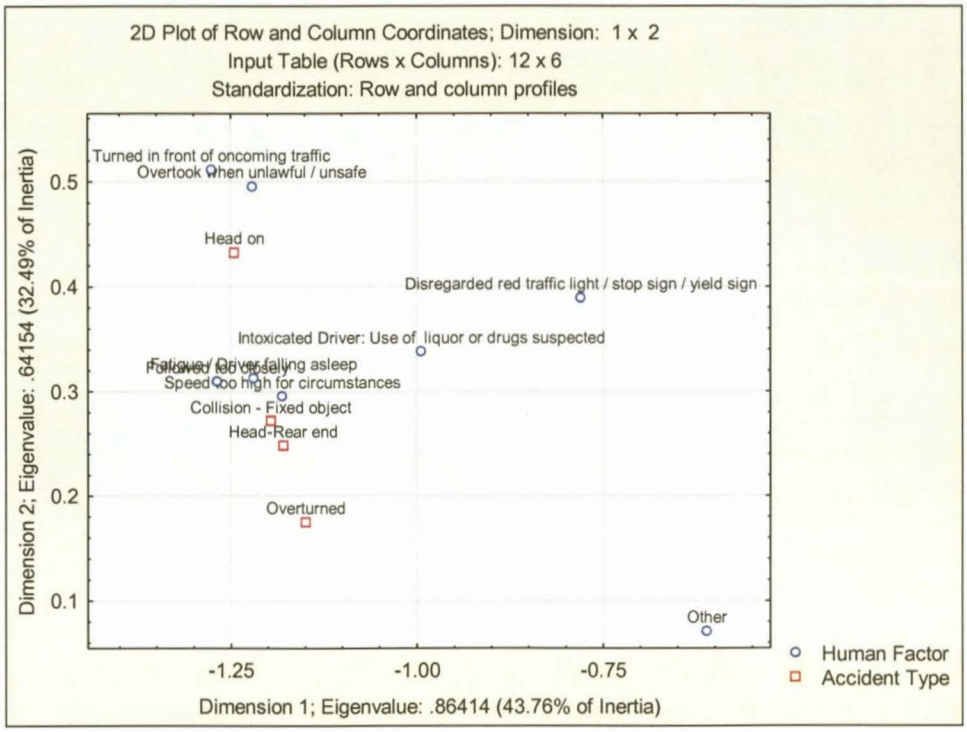

Fig. 4.4.16(c): Two-dimensional solution (zoomed) – Correspondence between Accident Type and Human Factor, Western Cape Province

ix) Results of Analysis: Type of Accident vs. Variable X

Table 4.4.9(a): Correspondences between variables for Type of Accident vs. X – South Africa

Type of Accident	Area Type	Road Factor	Vehicle Factor	Human Factor	Vehicle Type	Road User Type	Gender	Race Group
Overturned	➢ Rural	➢ Animals: Stray / Wild ➢ Sharp bend ➢ Poor condition of road surface	➢ Tyre burst prior to accident ➢ Smooth tyres ➢ Steering: Faulty ➢ Overloading: Cargo / Passengers ➢ Brakes: Faulty		➢ Minibus ➢ Minibus Taxi ➢ LDV / Bakkie ➢ Sedan ➢ Bus ➢ Tractor ➢ Other	➢ Passenger	➢ Male ➢ Female	➢ Asian ➢ White ➢ Foreigner
Hit and Run	➢ Urban	➢ Unknown	➢ Other ➢ Unknown	➢ Not known ➢ Hit-and-run	➢ Unknown	➢ Pedestrian	➢ Male	➢ Black ➢ Coloured ➢ Unknown
Pedestrian	➢ Urban	➢ Unknown	➢ Other ➢ Unknown	➢ Pedestrian: Jay walking ➢ Intoxicated Pedestrian: Use of liquor or drugs	➢ Minibus ➢ Minibus Taxi ➢ LDV / Bakkie ➢ Sedan ➢ Bus ➢ Tractor ➢ Other	➢ Pedestrian	➢ Male ➢ Female	➢ Black ➢ Coloured ➢ Unknown
Head on	➢ Rural	➢ Road works ➢ Road surface slippery / wet ➢ Traffic light / Road sign / Road marking defective ➢ Narrow road lane ➢ Blind rise / Corner ➢ Poor visibility (Rain, mist, dust, smoke, dawn, dusk)	➢ Chevrons: No reflective stripes ➢ Lights: Faulty, not switched on, blinding etc.	➢ Cell phone use / holding ➢ Disregarded red traffic light / stop sign / yield sign ➢ Turned in front of oncoming traffic ➢ Overtook when unlawful/unsafe	➢ Minibus ➢ Minibus Taxi ➢ LDV / Bakkie ➢ Sedan ➢ Bus ➢ Tractor ➢ Other	➢ Driver	➢ Female	➢ Asian ➢ White ➢ Foreigner
Head-Rear End	➢ Rural	➢ Poor street lighting ➢ Road works ➢ Poor visibility (Rain, mist, dust, smoke, dawn, dusk)	➢ Chevrons: No reflective stripes ➢ Lights: Faulty, not switched on, blinding etc.	➢ Fatigue / Driver falling asleep ➢ Followed too closely ➢ Intoxicated Driver: Use of liquor or drugs suspected ➢ Speed too high for circumstances	➢ Heavy Vehicle ➢ Motorcycle ➢ Other ➢ Bicycle	➢ Driver	➢ Male	➢ Asian ➢ White ➢ Foreigner

Table 4.4.9(b): Correspondences between variables for Type of Accident vs. X – Western Cape Province

Type of Accident	Area Type	Road Factor	Vehicle Factor	Human Factor	Vehicle Type	Road User Type	Gender	Race Group
Overturned	➢ Rural	➢ Traffic light / Road sign / Road marking defective ➢ Poor condition of road surface ➢ Sharp bend	➢ Tyre burst prior to accident ➢ Smooth tyres ➢ Overloading: Cargo / Passengers		➢ Sedan ➢ Other ➢ Minibus Taxi ➢ Minibus ➢ Bus ➢ LDV / Bakkie	➢ Passenger	➢ Male ➢ Female	➢ Asian
Hit and Run	➢ Urban	➢ Unknown ➢ Narrow road lane ➢ Other ➢ Poor street lighting ➢ Animals: Stray / Wild	➢ Other ➢ Unknown	➢ Not known ➢ Hit-and-run	➢ Unknown	➢ Pedestrian	➢ Male	➢ Black
Pedestrian	➢ Urban	➢ Unknown ➢ Narrow road lane ➢ Other ➢ Poor street lighting ➢ Animals: Stray / Wild	➢ Other ➢ Unknown	➢ Pedestrian: Jay walking ➢ Intoxicated Pedestrian: Use of liquor or drugs	➢ Sedan ➢ Other ➢ Minibus Taxi ➢ Minibus ➢ Bus ➢ LDV / Bakkie	➢ Pedestrian	➢ Male	➢ Coloured ➢ Unknown
Head on	➢ Rural	➢ Blind rise / Corner ➢ Poor visibility (Rain, mist, dust, smoke, dawn, dusk) ➢ Road surface slippery / wet ➢ Road works	➢ Chevrons: No reflective stripes ➢ Lights: Faulty, not switched on, blinding etc.	➢ Cell phone use / holding ➢ Disregarded red traffic light / stop sign / yield sign ➢ Turned in front of oncoming traffic ➢ Overtook when unlawful/unsafe	➢ Sedan ➢ Other ➢ Minibus Taxi ➢ Minibus ➢ Bus ➢ LDV / Bakkie	➢ Driver	➢ Female	➢ White
Head-Rear End	➢ Rural	➢ Blind rise / Corner ➢ Poor visibility (Rain, mist, dust, smoke, dawn, dusk) ➢ Road surface slippery / wet ➢ Road works	➢ Chevrons: No reflective stripes ➢ Lights: Faulty, not switched on, blinding etc.	➢ Fatigue / Driver falling asleep ➢ Followed too closely ➢ Intoxicated Driver: Use of liquor or drugs suspected ➢ Speed too high for circumstances	➢ Motorcycle ➢ Sedan ➢ Tractor ➢ Heavy Vehicle ➢ Other	➢ Driver	➢ Male ➢ Female	➢ Asian

4.4.2 Correspondence between Road User Type (Fatalities) and Variable *X*

This section provides the results of correspondence analyses performed on relative frequency tables cross tabulating the variable *Road User Type* and the following categorical variables:

- Gender
- Race Group
- Seatbelt Status
- Vehicle Type

The proposed correspondences as interpreted for each individual analysis on each variable pair can be found in Appendix B2. At the end of this section the final summarizing tables containing the proposed correspondences between road user type fatalities and all of the abovementioned variables will be provided.

i) Road User Status vs. Gender

The same results were obtained for South Africa and the Western Cape Provinces. The *overall inertia* of the relative frequency tables for South Africa and the Western Cape Province was reproduced as follow:

For South Africa: 99.06% by Dimension 1 and 0.94% by Dimension 2 (total of 100%). For Western Cape Province: 95.34% by Dimension 1 and 4.66% by Dimension 2 (total of 100%). The *quality* statistic showed satisfactory values in the representation of each column and row point. Please refer to Appendix B1 for the relevant output statistic tables.

Table 4.4.10(a): Cross tabulation – Road User Status (Fatalities) vs. Gender, South Africa, 2002-2004

Road User Status	FEMALE	MALE	UNKNOWN	Total
PEDESTRIAN	3069	9373	121	**12563**
DRIVER	610	7379	45	**8034**
PASSENGER	3534	5640	230	**9404**
Total	**7213**	**22392**	**396**	**30001**

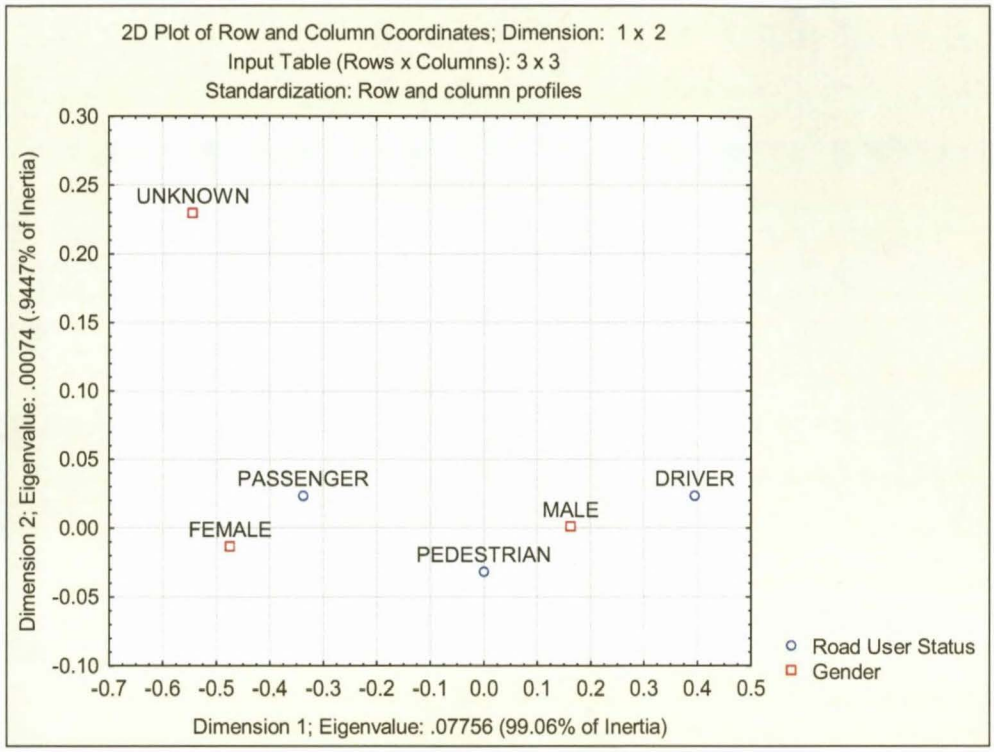

Fig. 4.4.17: Two-dimensional solution – Correspondence between Road User Status and Gender, South Africa

Table 4.4.10(b): Cross tabulation – Road User Status (Fatalities) vs. Gender, Western Cape, 2002-2004

Road User Status	FEMALE	MALE	UNKNOWN	Total
PEDESTRIAN	334	1152	21	**1507**
PASSENGER	302	544	55	**901**
DRIVER	76	737	6	**819**
Total	**712**	**2433**	**82**	**3227**

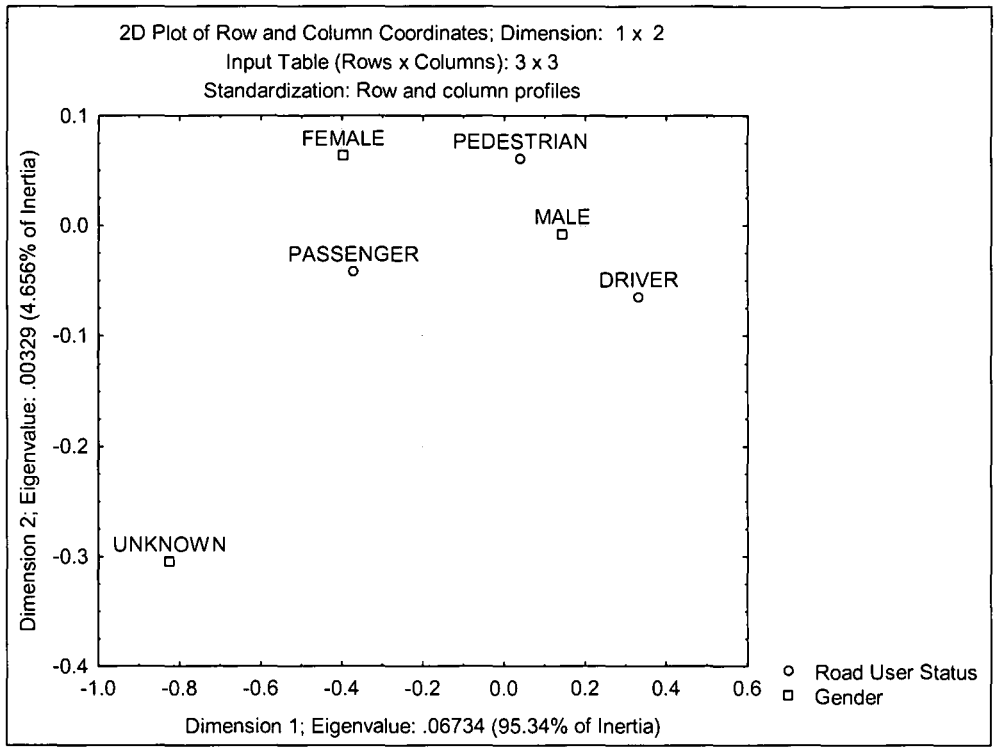

Fig. 4.4.18: Two-dimensional solution – Correspondence between Road User Status and Gender, Western Cape

ii) Road User Status vs. Race Group

Only two dimensions were possible. The reproduction of the *overall inertia* of the relative frequency tables for South Africa and Western Cape Province was as follow:

For South Africa: 98.92% by Dimension 1 and 1.08% by Dimension 2. For Western Cape Province: 95.69% by Dimension 1 and 4.31% by Dimension 2. The *quality* of representation of the row and column points is at a satisfactory degree. Please refer to Appendix B1 for the output statistics and output tables.

Table 4.4.11(a): Cross tabulation – Road User Status (Fatalities) vs. Race Group, South Africa, 2002-2004

Population Group	DRIVER	PEDESTRIAN	PASSENGER	Total
BLACK	4458	10648	6970	**22076**
WHITE	2484	279	1093	**3856**
COLOURED	672	1396	936	**3004**
ASIAN	359	124	306	**789**
UNKNOWN	61	116	93	**270**
FOREIGNER	0	0	6	**6**
Total	**8034**	**12563**	**9404**	**30001**

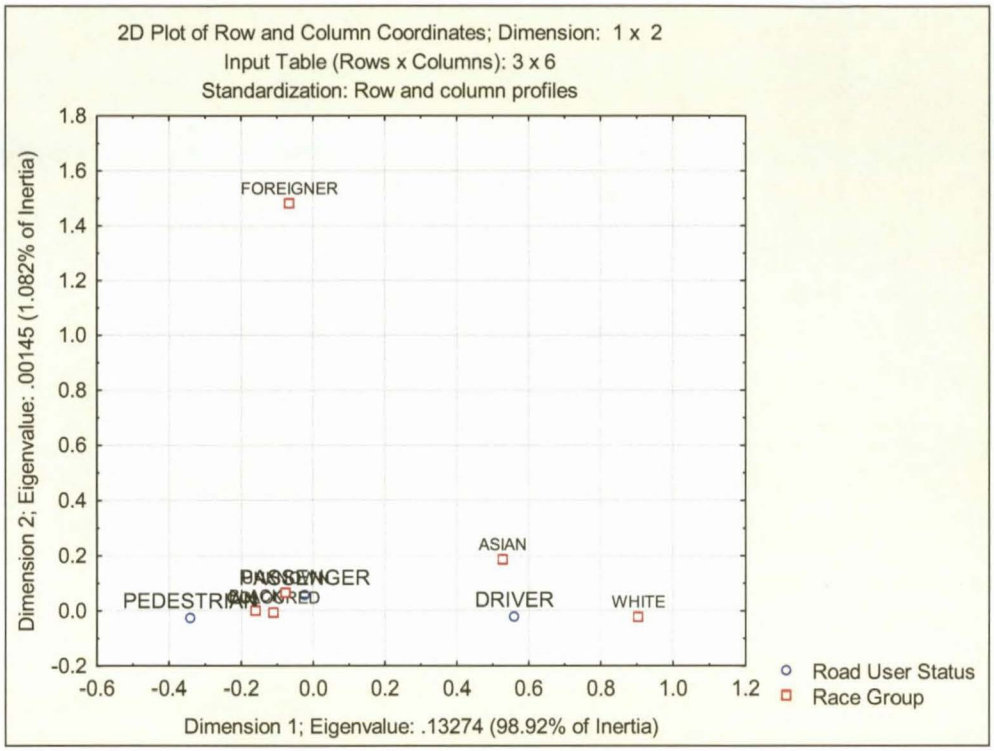

Fig. 4.4.19: Two-dimensional solution – Correspondence between Road User Status and Race Group, South Africa

Table 4.4.11(b): Cross tabulation – Road User Status (Fatalities) vs. Race Group, Western Cape, 2002-2004

Population Group	PEDESTRIAN	PASSENGER	DRIVER	Total
COLOURED	938	439	344	**1721**
BLACK	491	299	138	**928**
WHITE	47	130	311	**488**
UNKNOWN	24	10	8	**42**
ASIAN	7	23	18	**48**
Total	**1507**	**901**	**819**	**3227**

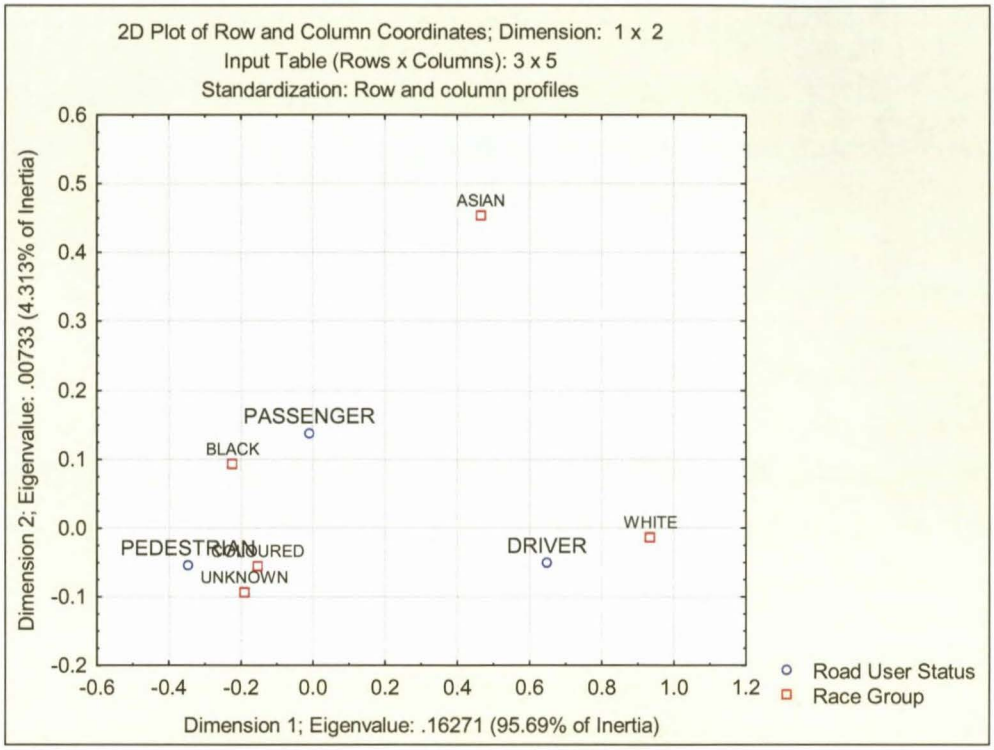

Fig. 4.4.20: Two-dimensional solution – Correspondence between Road User Status and Race Group, Western Cape

iii) Road User Status vs. Seatbelt status

The same results were obtained for South Africa and the Western Cape Province. Only one dimension was possible for the solution. 100% of the *overall inertia* could thus be reproduced for both the results for South Africa and Western Cape Province. Each row and column point could also be reproduced exactly. Refer to Appendix B1 for the output tables and statistics.

Table 4.4.12(a): Cross tabulation – Road User Status (Fatalities) vs. Seatbelt status, South Africa, 2002-2004

Seatbelt Status	DRIVER	PASSENGER	Total
UNKNOWN	4568	4022	**8590**
NO	2894	5099	**7993**
YES	571	256	**827**
Total	**8033**	**9377**	**17410**

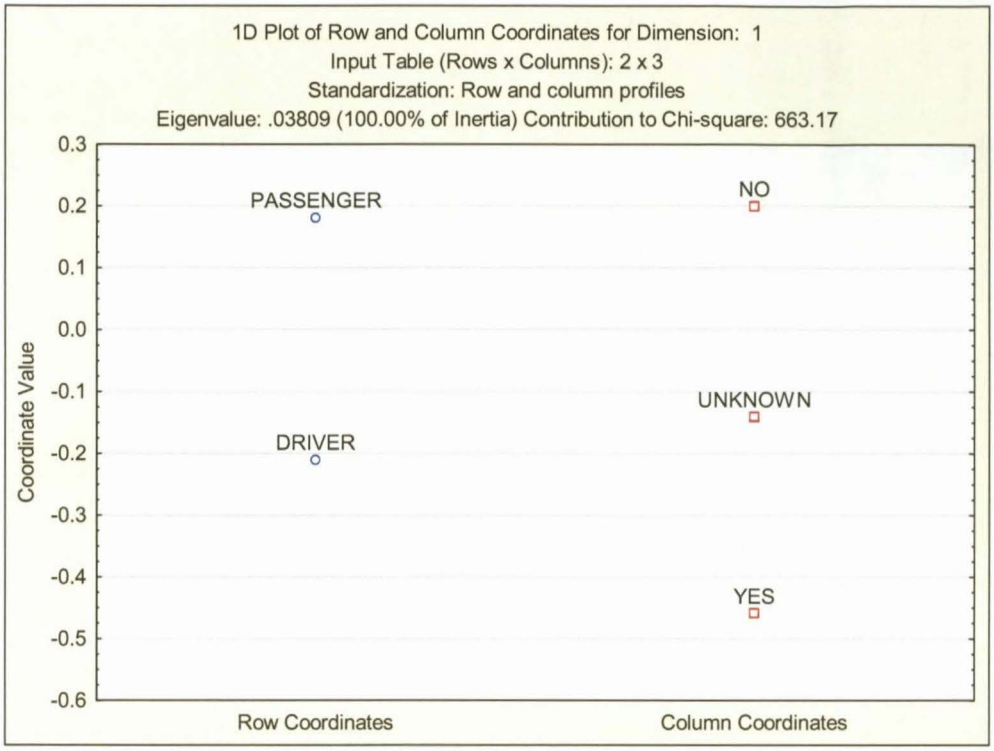

Fig. 4.4.21: One-dimensional solution – Correspondence between Road User Status and Seatbelt status, South Africa

Table 4.4.12(b): Cross tabulation – Road User Status (Fatalities) vs. Seatbelt status, Western Cape, 2002-2004

Seatbelt Status	PASSENGER	DRIVER	Total
NO	463	297	**760**
YES	30	58	**88**
UNKNOWN	406	464	**870**
Total	**899**	**819**	**1718**

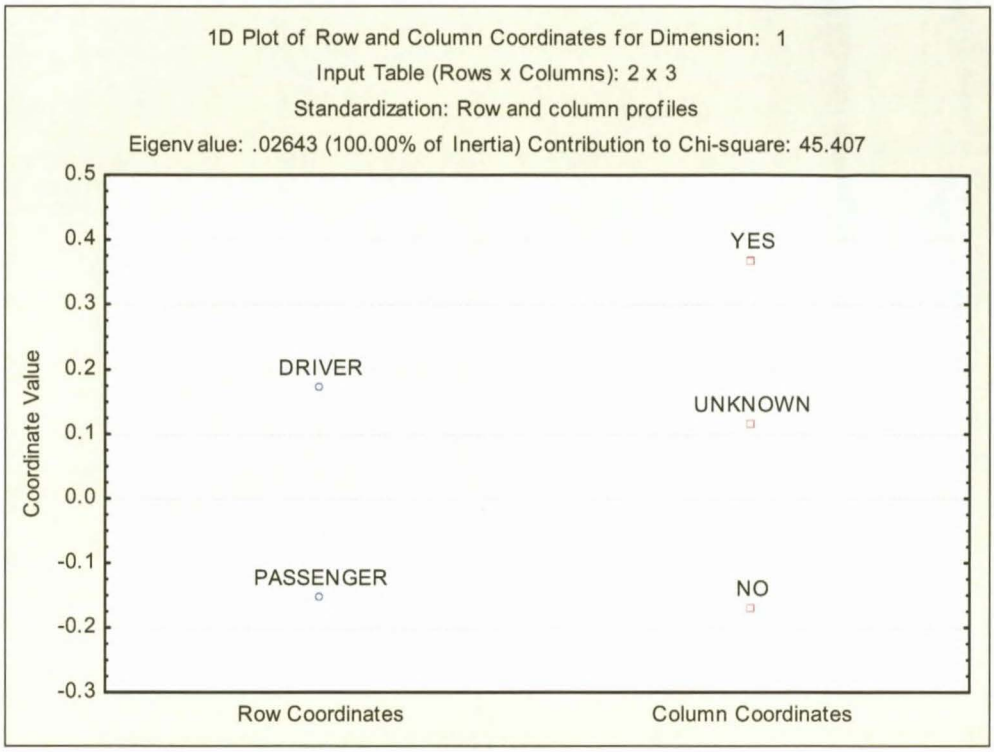

Fig. 4.4.22: One-dimensional solution – Correspondence between Road User Status and Seatbelt status, Western Cape

iv) Road User Status vs. Vehicle Type

The same correspondences were obtained for South Africa and the Western Cape Province. The *overall inertia* of the tables of relative frequency for South Africa and Western Cape Province was reproduced in the following amounts:

For South Africa: 67.65% by Dimension 1 and 32.35% by Dimension 2. For Western Cape Province: 63.79% by Dimension 1 and 36.21% by Dimension 2. The *quality* of representation is at a satisfactory degree. Please refer to the output tables and statistics in Appendix B1.

Table 4.4.13(a): Cross tabulation – Road User Status (Fatalities) vs. Vehicle Type, South Africa, 2002-2004

Vehicle Type	PEDESTRIAN	DRIVER	PASSENGER	Total
Sedan	5050	4091	3834	**12975**
LDV / Bakkie	2263	1567	2320	**6150**
Unknown	2436	20	57	**2513**
Minibus Taxi	869	263	1126	**2258**
Motorcycle	46	400	60	**506**
Minibus	646	255	785	**1686**
Bicycle	8	864	13	**885**
Heavy vehicle	874	420	623	**1917**
Tractor	73	96	195	**364**
Bus	246	34	364	**644**
Other	50	24	27	**101**
Panel van	2	0	0	**2**
Total	**12563**	**8034**	**9404**	**30001**

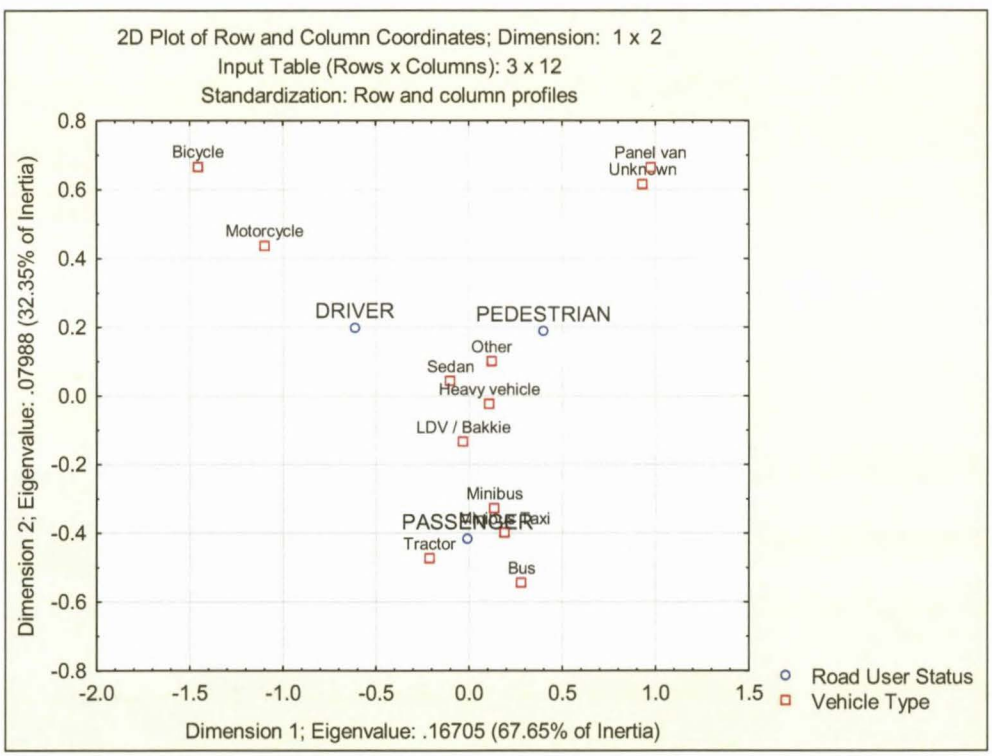

Fig. 4.4.23: Two-dimensional solution – Correspondence between Road User Status and Vehicle Type, South Africa

Table 4.4.13(b): Cross tabulation – Road User Status (Fatalities) vs. Vehicle Type, Western Cape, 2002-2004

Vehicle Type	PEDESTRIAN	PASSENGER	DRIVER	Total
Sedan	713	422	449	**1584**
LDV / Bakkie	260	190	124	**574**
Motorcycle	9	16	57	**82**
Minibus Taxi	38	63	13	**114**
Minibus	67	83	25	**175**
Tractor	4	24	5	**33**
Bicycle	1	1	96	**98**
Heavy vehicle	103	54	39	**196**
Unknown	292	5	4	**301**
Bus	16	42	4	**62**
Other	4	1	3	**8**
Total	**1507**	**901**	**819**	**3227**

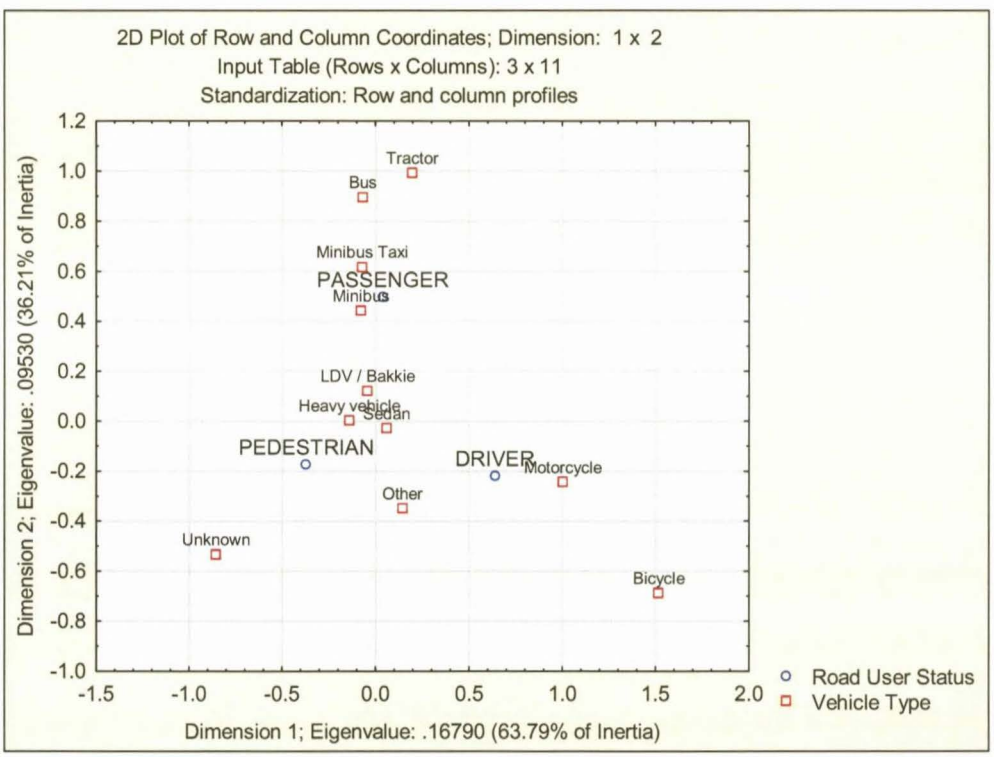

Fig. 4.4.24: Two-dimensional solution – Correspondence between Road User Status and Vehicle Type, Western Cape

v) *Results of Analysis: Road User Type (Fatalities) vs. Variable X*

Table 4.4.14(a): Correspondences between variables for Road User Status (Fatalities) vs. *X* - South Africa

Road User Status	Gender	Race Group	Seatbelt status	Vehicle Type
Driver	➢ Male	➢ White ➢ Asian	➢ Unknown ➢ Yes	➢ Motorcycle ➢ Bicycle ➢ Sedan ➢ Other ➢ Heavy Vehicle ➢ LDV / Bakkie
Passenger	➢ Female ➢ Unknown	➢ Black ➢ Coloured ➢ Unknown ➢ Foreigner	➢ No	➢ Minibus ➢ Minibus Taxi ➢ Tractor ➢ Bus ➢ Sedan ➢ Other ➢ Heavy Vehicle ➢ LDV / Bakkie
Pedestrian	➢ Male	➢ Black ➢ Coloured ➢ Unknown ➢ Foreigner	-	➢ Panel van ➢ Unknown ➢ Sedan ➢ Other ➢ Heavy Vehicle ➢ LDV / Bakkie

Table 4.4.14(b): Correspondences between variables for Road User Status vs. X - Western Cape Province

Road User Status	Gender	Race Group	Seatbelt status	Vehicle Type
Driver	➢ Male	➢ White ➢ Asian	➢ Unknown ➢ Yes	➢ Motorcycle ➢ Bicycle ➢ Sedan ➢ Other ➢ Heavy Vehicle ➢ LDV / Bakkie
Passenger	➢ Female ➢ Unknown	➢ Black	➢ No	➢ Minibus ➢ Minibus Taxi ➢ Tractor ➢ Bus ➢ Sedan ➢ Other ➢ Heavy Vehicle ➢ LDV / Bakkie
Pedestrian	➢ Male	➢ Coloured ➢ Unknown	-	➢ Unknown ➢ Sedan ➢ Other ➢ Heavy Vehicle ➢ LDV / Bakkie

4.4.3 Discussion of Correspondence Analysis Results

The results of the correspondence analyses are discussed in this section with regard to the final summary tables given at the end of sections 4.4.1 and 4.4.2 (Tables 4.4.9(a), 4.4.9(b), 4.4.14(a) and 4.4.14(b)).

Abovementioned tables provide the reader with convenient summaries which set out the results as was interpreted from the graphical output of the analyses. The tables are read from left to right across the columns with each row representing the correspondences relating to a particular accident type (Tables 4.4.9(a) and 4.4.9(b)) or road user type (Tables 4.4.14(a) and 4.4.14(b)). If reading table 4.4.9(a) from left to right across the columns for *Overturned* accidents, for example, it is interpreted that *Overturned* accidents correspond mostly to *Rural* area types, mostly to *Passenger* fatalities and also, mostly to *Asian, White* and/or *Foreigner* fatalities. Al other rows in the mentioned tables are interpreted in the same way.

Not all variables significantly correspond to one another, but it should be noted that just because a certain item (variable category) is not proposed to show a correspondence to a particular accident type, it does not mean that there is no correspondence e.g. X might correspond *more* to Y than Z, but it doesn't mean that no correspondence exists between X and Z. Where variables correspond significantly to more than one accident type, the particular variable category showing correspondence is repeated under each row representing the accident type with which correspondences were found.

The terminology of Correspondence Analysis was already explained in Chapter 2. In the preceding sections, comments on the *quality* of representation of each variable were given as well as the percentage *overall inertia* which was reproduced from each input table, according to the given amount of dimensions extracted by each of the analyses. This information was provided to give the reader some insight into the accuracy with which each analysis's output can be interpreted. The reader was also advised to consult the output statistics in Appendix B1 in case there is a difference of opinion on what the final results should have been. No further detailed discussions will be given with regard to the output statistics and what these represent as thorough explanations were provided in previous chapters. Only the tables of relative row and column frequencies were used to verify the results as given in Table 4.4.9(a) and (b) and Table 4.4.14(a) and (b).

This analysis technique was applied to observe whether sensible conclusions could be drawn from the results i.e. whether sensible conclusions could be drawn from the graphical output alone (instead of potentially using a lot of time and effort to manually inspect the frequency and relative frequency tables for correspondences i.e. how one variable "relate" to another). The results of the analyses will be discussed with regard to latter statement under the relevant subheadings.

i) Type of Accident vs. Variable X

Tables 4.4.9(a) and 4.4.9(b) contain the summaries of the results from the analyses performed between accident type variables and the eight variables with which the accident types were cross tabulated with as given in the tables. Cross tabulation tables used as input for the analyses were created based on fatal road accident datasets for South Africa as well as for the Western Cape Province alone (as explained before). The accident types given are the top five accident types which occurred with the highest frequencies in the fatal road accident database for the years 2002-2004 as explained in Chapter 3.

The results for South Africa and for the Western Cape Province were found to not show major differences. Differences were to be expected, though, as analysis of a dataset representing the data for a whole country should deliver results which should be able to provide a general overview of how variables correspond to one another in the whole country. Analysis of a dataset representing the data for one province within the country should deliver results which should be able to provide a general overview of how variables correspond to one another as applicable to that province.

When the conditional probabilities of each variable's occurrence were verified with the results as interpreted from the graphical output (given in Tables 4.4.9(a) and (b)), it was found that some categories were included in the summary tables which did not show a necessarily strong correspondence with the particular fatal accident type with which correspondence analyses were performed. Some variables were found to have strong correspondence with a particular accident type which was not included in the final summary tables. These problems are common when only the graphical output is interpreted. It is necessary that the interpretations from the graphical output be verified with the relative row/column frequencies of the relevant variables.

The arguments just mentioned above also apply to the method of interpretation of the results for the following section and won't be repeated under the following set of discussions.

ii) Road User Type (Fatalities) vs. Variable X

Tables 4.4.14(a) and 4.4.14(b) contain the summaries of the results from the analyses performed between road user type variables and the four variables with which the road user types were cross tabulated with as given in the tables. Cross tabulation tables used as input for the analyses were created based on fatal road accident datasets for South Africa as well as for the Western Cape Province (as explained before).

The results for South Africa and for the Western Cape Province were found to not show major differences, but the results for South Africa did indicate more possible correspondences than the results for Western Cape Province.

Variables included in this set of analyses (Summary Tables 4.4.14(a) and (b)) which were also included in the previous set of analyses (cross tabulation with *Type of Accident*) do not correspond to road user type variables in the same way as one would interpret from Tables 4.4.9(a) and (b). This set of analyses provided different sets of graphical output with different levels of *quality* of representation than the previous analyses and also with different sample sizes. Due to these differences the interpretations might deliver different results in cases where the two sets of analyses have some variables in common.

Tables 4.4.14(a) and (b) indicate that *Seatbelt status* is the only variable which these tables and Tables 4.4.9(a) and (b) do not have in common. The rest of the variables in Tables 4.4.14(a) and (b) are all the same variables included in the analyses results given in Tables 4.4.9(a) and (b).

A wide selection of vehicle types was included as part of the solution in Tables 4.4.14(a) and (b) even though not all of the included vehicles have equally strong correspondence with the particular road user type fatality. This is due to interpretation of the results from graphical output and not from frequency tables or conditional probabilities. Latter method would have narrowed the amount of vehicles down significantly, making the results more conclusive than the results provided in Tables 4.4.14(a) and (b) with regard to vehicle types.

Correspondence between road user type fatalities and respectively gender, race group and seatbelt status as interpreted from the graphical output were found to be relatively conclusive. Refer to the

previous set of discussions on correspondence between fatal accident types and different variables involved with regard to the arguments highlighted in that section.

4.5 Data Mining: Association Rules

The theoretical principles and methodology for the application of association rules for the purpose of this study was discussed in detail in Chapters 2 and 3 respectively. In this section the results are commented on.

4.5.1 Association Rules Results for Accident Factor Combinations for National Roads N1, N2 and N7, Western Cape Province

The first set of association rules created were between accident factors and accident types for each road section along the national roads N1, N2 and N7, Western Cape Province. The output delivered a large collection of association rules which had to be filtered for meaningful rules which could possibly be applied for the purpose of prediction (Chapter 3). It was found that it was rarely the case that all three accident factors were indicated for a particular fatal accident case. Sometimes only one accident factor was entered and otherwise only two. For the purpose of this analysis only the cases where al three accident factors were indeed entered, were investigated i.e. the output was filtered for those rules indicating three accident factors in the *body* of the rule (human, vehicle and road factors) and one accident type item in the *head* of the rule, as explained in Chapter 3. This was done, because the co-occurrences of accident factors with relation to different accident types were investigated i.e. combinations of accident factors and their relationship to different accident types were inspected.

After completion of this filtering process, the association rules with three accident factors in the *body* were summarized for each road section together with their respective *support*, *confidence* and *lift* values in easy-to-interpret tables which will be provided in sections to follow (Tables 4.5.1, 4.5.2 and 4.5.3). The association rules for each road section were also sorted in descending order of *lift* value, so the most meaningful rules (the rules most likely suitable for prediction purposes) were easy to interpret from the summaries.

The summary tables are read from left to right. For example, if the first association rule is observed for the road section Cape Town – Goodwood in Table 4.5.1, it can be interpreted that if the human factor is *turned in front of oncoming traffic*, the vehicle factor is unknown and the road factor is unknown it is likely that a *Head on* accident will occur. This particular combination of human, vehicle and road factors is found 1.64% of the time in the sample data (the *support* of the *body*). Given this combination of accident factors the theoretical probability of a *Head on* collision is indicated as 100%. The *lift* value is 61, indicating that the chance of observing the *head* item in the data subset for this accident factor combination is 61 times greater than the chance of finding this accident factor combination in the whole dataset. The *lift* value is relatively high, indicating a useful association rule for this road section. All the summary tables must be interpreted in this manner.

Relatively small *support* values are given for the *body* (ranging mostly between 1% and approximately 16%) which led to accident factor combination subsets to be poorly represented. Also, because of relatively small sample sizes for each road section under study (sample sizes also differed for each road section), the *confidence* values (being conditional probabilities) were found to be relatively large i.e. larger than what might be the case in reality.

It became evident through association rules that there is a large amount of unknown road factors and vehicle factors. This can clearly be observed in Tables 4.5.1, 4.5.2 and 4.5.3, but because combinations of accident factors are inspected for each road section, at least the particular human factors as applicable to the fatal road accidents for each road section can be observed from these tables.

The same human factors are applicable to most road sections and no significant difference in the accident factor categories included in the association rules can be distinguished, except that some minor differences were observed between each national road in terms of the human factors included. Also, not all the road factors/vehicle factors included in the association rules in the summary tables are unknown, although these large amounts of unknown factors did lead to association rules which are not as conclusive as one would hope for.

The results contained in the summary tables are very similar to the Correspondence Analysis results discussed earlier and no new information can necessarily be concluded from these results. Correspondences can be interpreted from these summary tables as well, but the difference is that combinations of accident factors can be observed for each road section under study, which was not

investigated with Correspondence Analysis. Also, association rules were found for accident types that do not necessarily occur with the highest frequencies in the fatal accident database.

Tables 4.5.1, 4.5.2 and 4.5.3 are given on the next few pages. A full set of the association rules for this analysis is provided in Appendix C.

Table 4.5.1: Association Rules Summary - Accident Factor Combinations, N1, Western Cape Province

Accident Factors (Body)			Accident Type (Head)	Support (%)	Confidence (%)	Lift
Human Factor	Vehicle Factor	Road Factor				
CAPE TOWN - GOODWOOD						
Turned in front of oncoming traffic	Unknown	Unknown	Head on	1.63934	100.0000	61.0000
Other	Unknown	Unknown	Other	1.63934	100.0000	61.0000
Turned in front of oncoming traffic	Brakes: Faulty	Unknown	Sideswipe same direction	1.63934	100.0000	61.0000
Not Known	Unknown	Road surface slippery / wet	Overturned	1.63934	100.0000	6.1000
Speed too high for circumstances	Unknown	Unknown	Overturned	3.27869	66.6667	4.0667
Speed too high for circumstances	Unknown	Unknown	Head-Rear End	1.63934	33.3333	4.0667
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	11.47541	100.0000	1.5641
Pedestrian: Jay walking	Unknown	Poor street lighting	Pedestrian	1.63934	100.0000	1.5641
GOODWOOD - BELLVILLE						
Speed too high for circumstances	Unknown	Unknown	Overturned	11.76471	100.0000	8.5000
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	11.76471	100.0000	2.1250
BELLVILLE - PAARL						
Speed too high for circumstances	Chevrons: No reflective stripes	Poor visibility (Rain, mist, dust, smoke, dawn, dusk)	Head-Rear End	1.92308	100.0000	26.0000
Speed too high for circumstances	Unknown	Unknown	Overturned	5.76923	100.0000	8.6667
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	11.53846	100.0000	1.5758
Other	Other	Unknown	Pedestrian	1.92308	100.0000	1.5758
PAARL - WORCESTER						
Not Known	Tyre burst prior to accident	Unknown	Overturned	2.22222	100.0000	7.5000
Speed too high for circumstances	Unknown	Unknown	Overturned	2.22222	100.0000	7.5000
Turned in front of oncoming traffic	Unknown	Unknown	Turn in face of oncoming traffic	2.22222	50.0000	7.5000
Overtook when unlawful / unsafe	Lights: Faulty, not switched on, blinding etc	Other	Head on	2.22222	100.0000	5.6250
Turned in front of oncoming traffic	Other	Other	Head on	2.22222	100.0000	5.6250
Turned in front of oncoming traffic	Unknown	Unknown	Head on	2.22222	50.0000	2.8125
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	15.55556	100.0000	1.8750
WORCESTER - TOUWSRIVIER						
Speed too high for circumstances	Unknown	Unknown	Overturned	3.12500	100.0000	3.5556
Speed too high for circumstances	Tyre burst prior to accident	Unknown	Overturned	3.12500	100.0000	3.5556
Pedestrian: Jay walking	Other	Other	Pedestrian	3.12500	100.0000	2.2857
TOUWSRIVIER - LAINGSBURG						
Speed too high for circumstances	Other	Other	Collision - Fixed Object	3.12500	100.0000	16.0000
Followed too closely	Unknown	Unknown	Head-Rear End	3.12500	100.0000	16.0000
Speed too high for circumstances	Tyre burst prior to accident	Unknown	Head-Rear End	3.12500	50.0000	8.0000
Not Known	Unknown	Unknown	Sideswipe same direction	3.12500	33.3333	10.6667
Not Known	Unknown	Unknown	Turn in face of oncoming traffic	3.12500	33.3333	10.6667
Overtook when unlawful / unsafe	Unknown	Unknown	Head on	3.12500	100.0000	6.4000
Fatigue / Driver falling asleep	Unknown	Unknown	Head on	3.12500	100.0000	6.4000
Pedestrian: Jay walking	Other	Other	Pedestrian	3.12500	100.0000	5.3333
Other	Unknown	Unknown	Pedestrian	3.12500	100.0000	5.3333
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	9.37500	100.0000	5.3333
Not Known	Tyre burst prior to accident	Unknown	Overturned	3.12500	100.0000	2.4615
Speed too high for circumstances	Unknown	Unknown	Overturned	9.37500	100.0000	2.4615
Speed too high for circumstances	Tyre burst prior to accident	Unknown	Overturned	3.12500	50.0000	1.2308
Not Known	Unknown	Unknown	Overturned	3.12500	33.3333	0.8205
LAINGSBURG - LEEUW GAMKA						
Intoxicated Pedestrian: Use of liquor or drugs suspected	Unknown	Unknown	Pedestrian	2.22222	100.0000	9.0000
Other	Tyre burst prior to accident	Other	Head on	2.22222	100.0000	7.5000
Fatigue / Driver falling asleep	Lights: Faulty, not switched on, blinding etc	Poor visibility (Rain, mist, dust, smoke, dawn, dusk)	Head-Rear End	2.22222	100.0000	7.5000
Speed too high for circumstances	Unknown	Unknown	Head-Rear End	4.44444	66.6667	5.0000
Not Known	Tyre burst prior to accident	Unknown	Overturned	8.88889	100.0000	2.1429
Speed too high for circumstances	Tyre burst prior to accident	Unknown	Overturned	2.22222	100.0000	2.1429
Speed too high for circumstances	Unknown	Unknown	Overturned	2.22222	33.3333	0.7143

LEEUW GAMKA - BEAUFORT WEST						
Turned in front of oncoming traffic	Unknown	Unknown	Turn in face of oncoming traffic	2.38095	100.0000	42.0000
Speed too high for circumstances	Unknown	Unknown	Sideswipe same direction	2.38095	33.3333	14.0000
Not Known	Unknown	Unknown	Sideswipe opposite direction	2.38095	50.0000	7.0000
Speed too high for circumstances	Unknown	Unknown	Head-Rear End	4.76190	66.6667	7.0000
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	9.52381	100.0000	7.0000
Not Known	Tyre burst prior to accident	Unknown	Overturned	4.76190	100.0000	2.3333
Speed too high for circumstances	Overloading: Cargo / Passengers	Other	Overturned	2.38095	100.0000	2.3333
Not Known	Unknown	Unknown	Overturned	2.38095	50.0000	1.1667
BEAUFORT WEST- THREE SISTERS						
(Missing data)						

Table 4.5.2: Association Rules Summary - Accident Factor Combinations, N2, Western Cape Province

Accident Factors (Body)			Accident Type (Head)	Support (%)	Confidence (%)	Lift
Human Factor	Vehicle Factor	Road Factor				
CAPE TOWN - SOMERSET - WEST						
Overtook when unlawful / unsafe	Unknown	Unknown	Sideswipe same direction	1.28205	100.0000	78.0000
Turned in front of oncoming traffic	Unknown	Unknown	Head on	1.28205	100.0000	39.0000
Speed too high for circumstances	Unknown	Unknown	Collision - Fixed Object	1.28205	50.0000	13.0000
Not known	Brakes: Faulty	Other	Overturned	1.28205	100.0000	9.7500
Speed too high for circumstances	Unknown	Unknown	Head-Rear End	1.28205	50.0000	6.5000
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	16.66667	92.8571	1.7245
Pedestrian: Jay walking	Unknown	Unknown	Hit and Run	1.28205	7.1429	0.3980
STRAND - GRABOUW						
Speed too high for circumstances	Unknown	Unknown	Sideswipe same direction	4.76190	100.0000	21.0000
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	9.52381	100.0000	2.1000
Pedestrian: Jay walking	Other	Other	Pedestrian	4.76190	100.0000	2.1000
Other	Other	Other	Pedestrian	4.76190	100.0000	2.1000
GRABOUW - CALEDON						
Not known	Unknown	Unknown	Other	4.54545	100.0000	11.0000
Other	Unknown	Unknown	Other	4.54545	100.0000	11.0000
Speed too high for circumstances	Unknown	Unknown	Head-Rear End	4.54545	100.0000	5.5000
Other	Other	Other	Head-Rear End	4.54545	100.0000	5.5000
Pedestrian: Jay walking	Other	Unknown	Pedestrian	9.09091	100.0000	3.1429
Pedestrian: Jay walking	Unknown	Other	Pedestrian	13.63636	75.0000	2.3571
Pedestrian: Jay walking	Unknown	Other	Hit and Run	4.54545	25.0000	1.8333
CALEDON - RIVIERSONDEREND						
Pedestrian: Jay walking	Unknown	Other	Pedestrian	9.09091	100.0000	3.6667
Speed too high for circumstances	Unknown	Other	Head on	9.09091	50.0000	1.8333
Speed too high for circumstances	Unknown	Other	Overturned	9.09091	50.0000	1.8333
RIVIERSONDEREND - SWELLENDAM						
Pedestrian: Jay walking	Unknown	Other	Pedestrian	5.26316	100.0000	19.0000
Turned in front of oncoming traffic	Unknown	Other	Sideswipe opposite direction	5.26316	50.0000	9.5000
Speed too high for circumstances	Overloading: Cargo / Passengers	Poor visibility (Rain, mist, dust, smoke, dawn, dusk)	Overturned	5.26316	100.0000	2.1111
Fatigue: Driver falling asleep	Unknown	Unknown	Overturned	10.52632	100.0000	2.1111
Speed too high for circumstances	Unknown	Unknown	Overturned	10.52632	66.6667	1.4074
Turned in front of oncoming traffic	Unknown	Unknown	Head on	5.26316	50.0000	1.9000
Speed too high for circumstances	Unknown	Unknown	Head on	5.26316	33.3333	1.2667
SWELLENDAM - RIVERSDALE						
Turned in front of oncoming traffic	Unknown	Unknown	Head-Rear End	4.54545	100.0000	22.0000
Speed too high for circumstances	Unknown	Unknown	Overturned	9.09091	100.0000	3.6667
Fatigue: Driver falling asleep	Unknown	Poor street lighting	Overturned	4.54545	100.0000	3.6667
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	13.63636	100.0000	3.6667
RIVERSDALE - MOSSELBAY						
Followed too closely	Unknown	Unknown	Other	5.26316	100.0000	19.0000
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	5.26316	100.0000	4.7500
Turned in front of oncoming traffic	Unknown	Unknown	Head on	5.26316	100.0000	3.8000
Speed too high for circumstances	Unknown	Unknown	Overturned	5.26316	100.0000	2.7143
MOSSELBAY - HARTENBOSCH						
Speed too high for circumstances	Unknown	Unknown	Head-Rear End	5.26316	100.0000	19.0000
Turned in front of oncoming traffic	Unknown	Unknown	Turn in face of oncoming traffic	5.26316	100.0000	19.0000
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	10.52632	100.0000	1.9000
HARTENBOSCH - GEORGE						
Intoxicated Driver: Use of liquor or drugs suspected	Other	Other	Pedestrian	6.25000	100.0000	2.6667
GEORGE - SEDGEFIELD						
Turned in front of oncoming traffic	Unknown	Unknown	Head on	6.66667	50.0000	7.5000

Turned in front of oncoming traffic	Unknown	Unknown	Sideswipe opposite direction	6.66667	50.0000	7.5000
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	46.66667	100.0000	1.3636
SEDGEFIELD - KNYSNA						
Overtook when unlawful / unsafe	Unknown	Unknown	Head on	5.88235	100.0000	5.6667
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	5.88235	100.0000	2.1250
KNYSNA - PLETTENBERGBAY						
Speed too high for circumstances	Unknown	Unknown	Sideswipe opposite direction	2.08333	100.0000	48.0000
Disregarded red traffic light / stop sign / yield sign	Unknown	Unknown	Approach at angle	2.08333	100.0000	24.0000
Overtook when unlawful / unsafe	Unknown	Unknown	Head-Rear End	2.08333	100.0000	24.0000
Speed too high for circumstances	Tyre burst prior to accident	Unknown	Overturned	2.08333	100.0000	16.0000
Overtook when unlawful / unsafe	Overloading: Cargo / Passengers	Poor visibility (Rain, mist, dust, smoke, dawn, dusk)	Turn in face of oncoming traffic	2.08333	100.0000	9.6000
Turned in front of oncoming traffic	Other	Other	Turn in face of oncoming traffic	2.08333	100.0000	9.6000
Not known	Unknown	Unknown	Hit and Run	2.08333	100.0000	8.0000
Other	Unknown	Unknown	Hit and Run	2.08333	100.0000	8.0000
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	4.16667	100.0000	2.8235

Table 4.5.3: Association Rules Summary - Accident Factor Combinations, N7, Western Cape Province

Accident Factors (Body)			Accident Type (Head)	Support (%)	Confidence (%)	Lift
Human Factor	Vehicle Factor	Road Factor				
CAPE TOWN - GOODWOOD						
Not known	Unknown	Poor street lighting	Head on	4.00000	100.0000	8.3333
Turned in front of oncoming traffic	Unknown	Unknown	Turn in face of oncoming traffic	8.00000	100.0000	5.0000
Turned in front of oncoming traffic	Unknown	Other	Turn in face of oncoming traffic	4.00000	100.0000	5.0000
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	16.00000	100.0000	1.9231
Pedestrian: Jay walking	Other	Other	Pedestrian	4.00000	100.0000	1.9231
GOODWOOD - MALMESBURY						
Speed too high for circumstances	Overloading: Cargo / Passengers	Sharp Bend	Sideswipe same direction	2.70270	100.0000	37.0000
Not known	Unknown	Unknown	Overturned	2.70270	50.0000	18.5000
Speed too high for circumstances	Unknown	Unknown	Collision - Fixed Object	2.70270	100.0000	12.3333
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	5.40541	100.0000	1.5417
Pedestrian: Jay walking	Other	Other	Pedestrian	5.40541	100.0000	1.5417
Other	Unknown	Unknown	Pedestrian	2.70270	100.0000	1.5417
Not known	Unknown	Unknown	Pedestrian	2.70270	50.0000	0.7708
MALMESBURY - PIKETBERG						
Speed too high for circumstances	Tyre burst prior to accident	Unknown	Collision - Fixed Object	3.12500	100.0000	16.0000
Turned in front of oncoming traffic	Unknown	Unknown	Head on	3.12500	100.0000	8.0000
Turned in front of oncoming traffic	Other	Other	Head on	3.12500	100.0000	8.0000
Speed too high for circumstances	Unknown	Unknown	Head on	3.12500	33.3333	2.6667
Speed too high for circumstances	Tyre burst prior to accident	Other	Overturned	3.12500	100.0000	5.3333
Speed too high for circumstances	Unknown	Unknown	Overturned	6.25000	66.6667	3.5556
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	15.62500	100.0000	2.0000
Pedestrian: Jay walking	Other	Other	Pedestrian	3.12500	100.0000	2.0000
PIKETBERG - CITRUSDAL						
Other	Tyre burst prior to accident	Unknown	Collision - Fixed Object	6.25000	100.0000	16.0000
Speed too high for circumstances	Unknown	Unknown	Head-Rear End	6.25000	100.0000	16.0000
Not known	Unknown	Sharp Bend	Head on	6.25000	100.0000	4.0000
Not known	Unknown	Unknown	Overturned	6.25000	100.0000	2.6667
Not known	Tyre burst prior to accident	Unknown	Overturned	6.25000	100.0000	2.6667
Fatigue / Driver falling asleep	Unknown	Unknown	Overturned	6.25000	100.0000	2.6667
Other	Unknown	Unknown	Overturned	6.25000	100.0000	2.6667
CITRUSDAL - CLANWILLIAM						
Speed too high for circumstances	Unknown	Unknown	Head on	12.50000	50.0000	2.0000
Speed too high for circumstances	Unknown	Unknown	Overturned	12.50000	50.0000	2.0000
CLANWILLIAM - VANRHYNSDORP						
Fatigue / Driver falling asleep	Unknown	Unknown	Head on	10.00000	100.0000	5.0000
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	10.00000	100.0000	2.0000
VANRHYNSDORP - BITTERFONTEIN						
Pedestrian: Jay walking	Unknown	Unknown	Pedestrian	16.66667	100.0000	6.0000

4.5.2 Association Rules Results for Non-Route Specific Analyses

i) Association Rules for Area Type, Vehicle Type and Accident Type: RSA and WC

The second set of association rules compiled were between the area types (urban or rural), vehicle types and accident types based on sample data for South Africa and sample data for Western Cape Province (i.e. for all roads and the total timeframe of the fatal accident database). Association rules could be created for both area types, urban and rural.

The relatively large collection of association rules was filtered for rules containing two items in the *body* (one area type item and one vehicle type item) and one accident type item in the *head*. These rules, together with their respective *support*, *confidence* and *lift* values, were summarized in Tables 4.5.4(a) and 4.5.4(b). For each combination of accident type and area type, the rules were sorted in descending order of *support* value i.e. the *support* value of the *body*, to illustrate the most representative combination of area type and vehicle type for the particular accident type (the *head* of the association rule). The *head* item is given prior to the *body* item(s) (when reading the tables from left to right), although any association rule is still interpreted as *if(body) then(head)*.

Support values were found to be relatively small (frequencies of combinations of variables are generally not large in the fatal accident database). The *confidence* values were found to be more accurate than those found in the analysis on accident factors (paragraph (i)) due to the larger sample sizes used for these analyses. It was mentioned that these analyses were not route specific, so a sufficient sample size was possible.

Support values for the association rules summarized in Tables 4.5.4(a) and 4.5.4(b) indicate that sedan vehicle types are the best represented by the association rules. Generally, sedans are a large percentage of the total vehicle population for South Africa and this result is to be expected. LDV's/Bakkies and heavy vehicles are well represented by the results of these analyses. Minibus and minibus taxi vehicle types are found in association rules with pedestrian and overturned accident type categories as *head* items respectively. No other vehicle types are included in the association rules as *body* items, because no other useful association rules (according to the principles in Chapter 2) could be created for the other vehicle types.

Lift values are the highest for rules with the accident type category hit and run as the *head* item. These values are approximately ten times higher than for the other association rules. High lift values for association rules based on this accident type is to be expected, because it is by definition true that the vehicle type involved in such an accident is unknown, both in rural and urban areas (see both Tables 4.5.4(a) and 4.5.4(b)).

Association rules were also found for accident types which did not necessarily appear in the fatal accident database with the highest of frequencies. Generally similar conclusions can be drawn from these analyses than what was concluded from the Correspondence Analyses previously done (and as explained before), but association rules provide information on the co-occurrences of certain variables and this leads to more detailed conclusions than the correspondence analysis results.

The most useful association rules from Tables 4.5.4(a) and 4.5.4(b) are highlighted according to the highest lift values for each accident type and area type.

Table 4.5.4(a): Association rules between Area Type, Vehicle Type and Accident Type for the Western Cape Province

Accident Type (Head)	Body Items		Support Body (%)	Confidence (%)	Lift
	(1) Area Type	(2) Vehicle Type			
Hit and Run	**Rural**	Unknown	4.06687	87.16578	10.5867
	Urban	Unknown	3.89222	91.22807	11.0801
Overturned	**Rural**	Sedan	5.51397	18.05556	1.2652
		LDV / Bakkie	2.81936	22.33202	1.5648
		Minibus	1.04790	33.60000	2.3543
	Urban	Sedan	1.84631	10.10929	0.7084
Head-Rear End	**Rural**	Sedan	2.96906	9.72222	1.0174
		Heavy Vehicle	1.24750	18.86792	1.9745
		LDV / Bakkie	1.04790	8.30040	0.8686
	Urban	Sedan	1.77146	9.69945	1.0150
Collision - Fixed Object	**Rural**	Sedan	1.47206	4.82026	1.2384
	Urban	Sedan	1.19760	6.55738	1.6847
Pedestrian	**Rural**	Sedan	11.55190	37.82680	1.0954
		LDV / Bakkie	4.36627	34.58498	1.0016
		Heavy Vehicle	1.79641	27.16981	0.7868
	Urban	Sedan	8.75749	47.95082	1.3886
		LDV / Bakkie	2.66966	52.45098	1.5190
		Heavy Vehicle	1.12275	46.39175	1.3435
Head on	**Rural**	Sedan	4.56587	14.95098	1.1285
		LDV / Bakkie	2.19561	17.39130	1.3127
		Heavy Vehicle	1.24750	18.86792	1.4242
	Urban	Sedan	2.44511	13.38798	1.0105
Turn in face of oncoming traffic	**Rural**	Sedan	1.67166	5.47386	1.1923
	Urban	-	-	-	-

Table 4.5.4(b): Associations between Area Type, Vehicle Type and Accident Type for the RSA

Accident Type (Head)	Body Items		Support Body (%)	Confidence (%)	Lift
	(1) Area Type	(2) Vehicle Type			
Hit and Run	**Rural**	Unknown	4.33430	89.17847	11.6661
	Urban	Unknown	3.01528	92.79661	12.1394
Overturned	**Rural**	Sedan	6.38304	22.17970	1.3031
		LDV / Bakkie	4.12777	27.20015	1.5981
		Heavy Vehicle	1.07118	13.65871	0.8025
	Urban	Sedan	1.63018	11.17825	0.6568
Head-Rear End	**Rural**	Sedan	3.22732	11.21424	1.0287
		Heavy Vehicle	1.69076	21.55899	1.9776
		LDV / Bakkie	1.53105	10.08891	0.9254
	Urban	Sedan	1.33554	9.15785	0.8400
Collision - Fixed Object	**Rural**	Sedan	1.10423	3.83695	1.2116
	Urban	-	-	-	-
Pedestrian	**Rural**	Sedan	9.35977	32.52320	1.0395
		LDV / Bakkie	4.57111	30.12158	0.9627
		Heavy Vehicle	1.68525	21.48876	0.6868
		Minibus Taxi	1.40713	38.50791	1.2308
		Minibus	1.16205	33.35968	1.0662
	Urban	Sedan	6.42985	44.08988	1.4092
		LDV / Bakkie	2.24701	46.20612	1.4768
		Minibus Taxi	1.11524	57.77461	1.8466
Head on	**Rural**	Sedan	5.03098	17.48158	1.2517
		LDV / Bakkie	2.51687	16.58501	1.1875
		Heavy Vehicle	1.62467	20.71629	1.4833
	Urban	Sedan	2.01570	13.82175	0.9896
Turn in face of oncoming traffic	**Rural**	Sedan	1.05741	3.67429	0.9586
	Urban	-	-	-	-

ii) Associations between Terrain Type, Vehicle Type and Accident Type: WC

The third set of association rules compiled were between the terrain types (flat, rolling or mountainous), vehicle types and accident types based on sample data for Western Cape Province (for the road sections along the national roads N1, N2 and N7 within Western Cape Province, because the terrain type variable were captured for these routes only). Association rules could be created for all three terrain types (flat, rolling and mountainous), but only in the case of pedestrian fatal accidents and head on fatal accidents. Other association rules could only be created for flat and rolling terrain. No relatively useful association rules could be created for the other accident types not included in the summary table. Useful association rules could only be created for the accident types occurring with the highest frequencies in the fatal accident database.

The filtering and summarizing methodology for this set of results are similar to that described in paragraph (i). The results are given in Table 4.5.5 and this table is interpreted in the same way as Tables 4.5.4(a) and 4.5.4(b). The same comments on the *Support*, *Confidence* and *Lift* values (also in terms of hit and run accidents and in terms of the accuracy with which these values could be calculated) are valid than was the case for the analysis results in paragraph (i).

Like before, sedan vehicle types and LDV's/Bakkies are also well represented according to the support values in Table 4.5.5. Minibus and minibus taxi vehicle types are found in association rules with overturned accidents as the *head* item. No other vehicle types are included in the association rules as *body* items, because (like before) no other useful association rules (according to the principles in Chapter 2) could be created for the other vehicle types. The most useful association rules from Table 4.5.5 are highlighted according to the highest lift values for each accident type and terrain type.

Table 4.5.5: Associations between Terrain Type, Vehicle Type and Accident Type for Western Cape

Accident Type (Head)	Body Items		Support Body (%)	Confidence (%)	Lift
	(1) Terrain Type	(2) Vehicle Type			
Hit and Run	**Flat**	Unknown	1.62369	89.4737	21.2907
	Rolling	Unknown	2.10124	84.6154	20.1346
Overturned	**Flat**	Sedan	2.67431	15.2174	0.9540
		LDV / Bakkie	1.43266	28.8462	1.8085
		Minibus	1.33715	48.2759	3.0266
		Minibus Taxi	1.24164	56.5217	3.5436
	Rolling	Sedan	2.96084	13.5371	0.8487
		LDV / Bakkie	1.14613	11.8812	0.7449
Head-Rear End	**Flat**	Sedan	2.19675	12.5000	1.1791
		Heavy Vehicle	1.81471	30.1587	2.8447
	Rolling	Sedan	2.00573	9.1703	0.8650
		LDV / Bakkie	1.33715	13.8614	1.3075
		Heavy Vehicle	1.14613	26.0870	2.4606
Collision - Fixed Object	**Rolling**	Sedan	1.05062	4.8035	1.4369
Pedestrian	**Flat**	Sedan	8.02292	45.6522	1.3935
		LDV / Bakkie	1.71920	34.6154	1.0566
	Rolling	Sedan	8.97803	41.0480	1.2530
		LDV / Bakkie	3.05635	31.6832	0.9671
		Heavy Vehicle	1.24164	28.2609	0.8627
	Mountainous	Sedan	2.10124	40.0000	1.2210
Head on	**Flat**	Sedan	2.10124	11.9565	0.7033
		Heavy Vehicle	1.43266	23.8095	1.4005
	Rolling	Sedan	3.91595	17.9039	1.0531
		LDV / Bakkie	2.10124	21.7822	1.2812
	Mountainous	Sedan	1.52818	29.0909	1.7111

iii) Associations between Area Type, Terrain Type, Vehicle Type and Accident Type: WC

The fourth set of association rules compiled were between the area types (urban or rural), terrain types (flat, rolling or mountainous), vehicle types and accident types based on sample data for Western Cape Province (for the road sections along the national roads N1, N2 and N7 within Western Cape Province, because the terrain type variable were captured for these routes only). Association rules could be created for all three terrain types (flat, rolling and mountainous), but only in the case of pedestrian fatal accidents and head on fatal accidents. Other association rules could only be created for flat and rolling terrain. No useful association rules could be created for urban areas in this analysis, only rural. Also, no relatively useful association rules could be created for the other accident types not included in the summary table. Useful association rules could only be created for the accident types occurring with the highest frequencies in the fatal accident database.

The filtering and summarizing methodology for this set of results are similar to that described in paragraphs (i). The results are given in Table 4.5.6 and this table is interpreted in the same way as the previous summary tables. The same comments on the *Support*, *Confidence* and *Lift* values (also in terms of hit and run accidents and in terms of the accuracy with which these values could be calculated) are valid than was the case for the analysis results in paragraph (i) (and (ii)).

Also like before, sedan vehicle types and LDV's/Bakkies are well represented according to the support values in Table 4.5.6. Minibus and minibus taxi vehicle types are found in association rules with overturned accidents as the *head* item. No other vehicle types are included in the association rules as *body* items, because (like before) no other useful association rules (according to the principles in Chapter 2) could be created for the other vehicle types.

The most useful association rules from Table 4.5.6 are highlighted according to the highest lift values for each accident type and terrain type.

Table 4.5.6: Associations between Area Type, Terrain Type, Vehicle Type and Accident Type for Western Cape

Accident Type (Head)	Body Items			Support Body (%)	Confidence (%)	Lift
	(1) Area Type	(2) Terrain Type	(3) Vehicle Type			
Hit and Run	**Rural**	**Rolling**	Unknown	2.10124	84.6154	20.1346
		Flat	Unknown	1.52818	88.8889	21.1515
	Urban	-	-	-	-	-
Overturned	**Rural**	**Flat**	Sedan	2.48329	14.7727	0.9262
			LDV / Bakkie	1.43266	28.8462	1.8085
			Minibus Taxi	1.24164	56.5217	3.5436
			Minibus	1.24164	46.4286	2.9108
		Rolling	Sedan	2.96084	13.5965	0.8524
			LDV / Bakkie	1.14613	11.8812	0.7449
	Urban	-	-	-	-	-
Head-Rear End	**Rural**	**Flat**	Sedan	2.19675	13.0682	1.2326
			Heavy Vehicle	1.81471	31.1475	2.9380
		Rolling	Sedan	2.00573	9.2105	0.8688
			LDV / Bakkie	1.33715	13.8614	1.3075
			Heavy Vehicle	1.14613	26.0870	2.4606
	Urban	-	-	-	-	-
Collision - Fixed Object	**Rural**	**Rolling**	Sedan	1.05062	4.8246	1.4432
	Urban	-	-	-	-	-
Pedestrian	**Rural**	**Flat**	Sedan	7.64088	45.4545	1.3875
			LDV / Bakkie	1.71920	34.6154	1.0566
		Rolling	Sedan	8.88252	40.7895	1.2451
			LDV / Bakkie	3.05635	31.6832	0.9671
			Heavy Vehicle	1.24164	28.2609	0.8627
		Mountainous	Sedan	2.10124	40.0000	1.2210
	Urban	-	-	-	-	-
Head on	**Rural**	**Flat**	Sedan	1.91022	11.3636	0.6684
			Heavy Vehicle	1.24164	21.3115	1.2535
		Rolling	Sedan	3.91595	17.9825	1.0577
			LDV / Bakkie	2.10124	21.7822	1.2812
		Mountainous	Sedan	1.52818	29.0909	1.7111
	Urban	-	-	-	-	-

4.6 Chance Variation of Fatal Accident Rates on National Road Sections

The goal of chance variation calculations in the context of this study was to investigate whether significant variation in fatal accident rates had occurred in the course of three years for all the road sections under study along the N1, N2 and N7 within the Western Cape Province. Table 4.6.3 contains the results for these calculations.

For each road section, the fatal accident rate is provided for each of the years 2002, 2003 and 2004 together with the probability of each road section having that particular fatal accident rate according to

the Poisson distribution (refer to Chapter 2 for the motivation for this method). Following Table 4.6.1, three graphs for each of the national routes N1, N2 and N7 within the Western Cape Province are given providing the Poisson distribution of fatal accident rates (fatal accidents per 100 million vehicle-km) for each road section along the national roads (Figures 4.6.1, 4.6.2 and 4.6.3). These distributions are based on the *expected* number of fatal accidents per 100 million vehicle-km per year for the three years 2002, 2003 and 2004 (the average annual number of fatal accidents per year over the three years for a particular road section).

Table 4.6.1: Chance Variation of Fatal Accident Rates for 2002-2004, South Africa (Poisson distribution)

			2002		2003		2004	
		Route Description	Fatal Accidents per 100 million veh-km's travelled	P(x)	Fatal Accidents per 100 million veh-km's travelled	P(x)	Fatal Accidents per 100 million veh-km's travelled	P(x)
N1	W1	CAPE TOWN - GOODWOOD	5.49	16.90%	5.64	16.90%	5.78	16.90%
N1	W2	GOODWOOD - BELLVILLE	2.95	26.52%	1.62	23.06%	2.35	26.52%
N1	W3	BELLVILLE - PAARL	4.18	17.72%	2.49	21.02%	2.73	21.02%
N1	W4	PAARL - WORCESTER	7.98	11.54%	11.16	9.84%	7.97	11.54%
N1	W5	WORCESTER - TOUWSRIVER	7.11	13.92%	4.39	13.09%	9.34	7.11%
N1	W6	TOUWS RIVER - LAINGSBURG	7.77	10.58%	19.57	0.21%	1.05	0.08%
N1	W7	LAINGSBURG - LEEUW GAMKA	13.05	5.56%	5.68	5.47%	8.92	12.81%
N1	W8	LEEU GAMKA - BEAUFORT WEST	14.13	10.57%	17.15	6.69%	13.20	10.78%
N1	W9	BEAUFORT WEST - THREE SISTERS	0.00	67.20%	0.00	67.20%	1.19	26.71%
N2	W1	CAPE TOWN - SOMERSET WEST	0.90	11.15%	2.70	26.83%	3.12	19.62%
N2	W10	HARTENBOSCH - GEORGE	5.84	7.42%	0.71	7.36%	2.06	25.05%
N2	W11	GEORGE - SEDGEFIELD	4.07	19.18%	5.44	13.91%	1.60	9.64%
N2	W12	SEDGEFIELD - KNYSNA	7.15	9.95%	2.66	9.02%	4.34	17.94%
N2	W13	KNYSNA - PLETTENBERG BAY	11.90	6.33%	12.83	8.01%	17.33	8.66%
N2	W3	STRAND - GRABOUW	3.22	20.62%	4.16	19.39%	4.08	19.39%
N2	W4	GRABOUW - CALEDON	10.49	8.93%	7.62	14.52%	5.29	10.49%
N2	W5	CALEDON - RIVIERSONDEREND	10.06	4.62%	3.37	8.17%	4.90	12.61%
N2	W6	RIVIERSONDEREND - SWELLENDAM	6.02	11.18%	10.29	10.70%	8.57	13.86%
N2	W7	SWELLENDAM - RIVERSDALE	3.04	20.93%	3.90	20.93%	4.39	19.28%
N2	W8	RIVERSDALE - MOSSELBAY	2.83	24.27%	2.56	24.27%	3.48	22.13%
N2	W9	MOSSELBAY - HARTENBOSCH	21.36	3.39%	27.93	7.54%	36.15	2.28%
N7	W1	CAPE TOWN - GOODWOOD	1.61	21.74%	3.45	20.91%	2.13	26.12%
N7	W2	GOODWOOD - MALMESBURY	2.46	7.49%	5.50	17.48%	7.61	11.22%
N7	W3	MALMESBURY - PIKETBERG	6.17	4.10%	20.23	0.46%	6.03	4.10%
N7	W4	PIKETBERG - CITRUSDAL	2.17	0.68%	10.75	11.23%	13.40	4.17%
N7	W5	CITRUSDAL - CLANWILLIAM	1.81	15.07%	3.58	22.40%	3.72	22.40%
N7	W6	CLANWILLIAM - VANRHYNSDORP	5.88	2.34%	9.83	11.01%	16.69	3.54%
N7	W7	VANRHYNSDORP - BITTERFONTEIN	15.41	4.30%	5.23	2.99%	9.99	11.84%

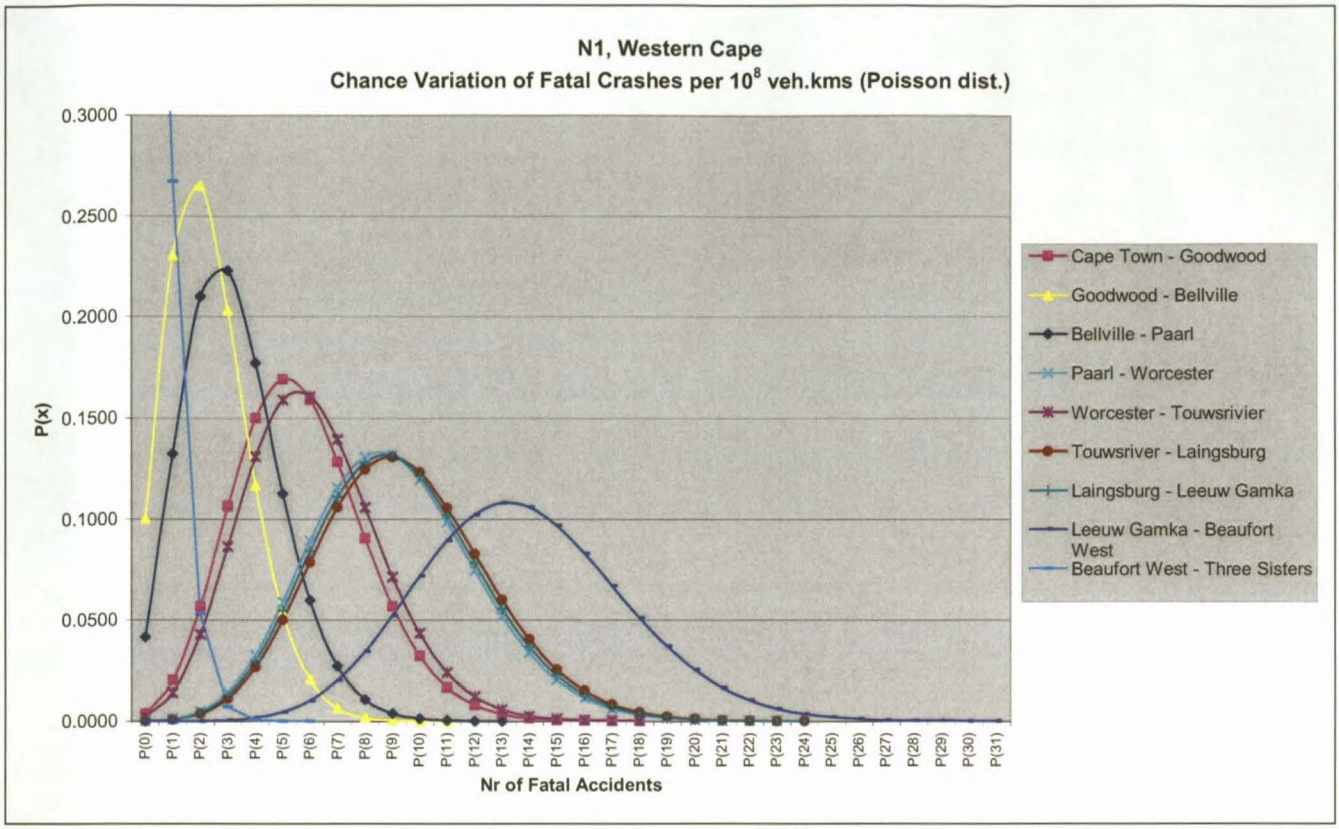

Fig. 4.6.1: Chance Variation of Fatal Accident Rates for the N1, Western Cape Province (Poisson dist.)

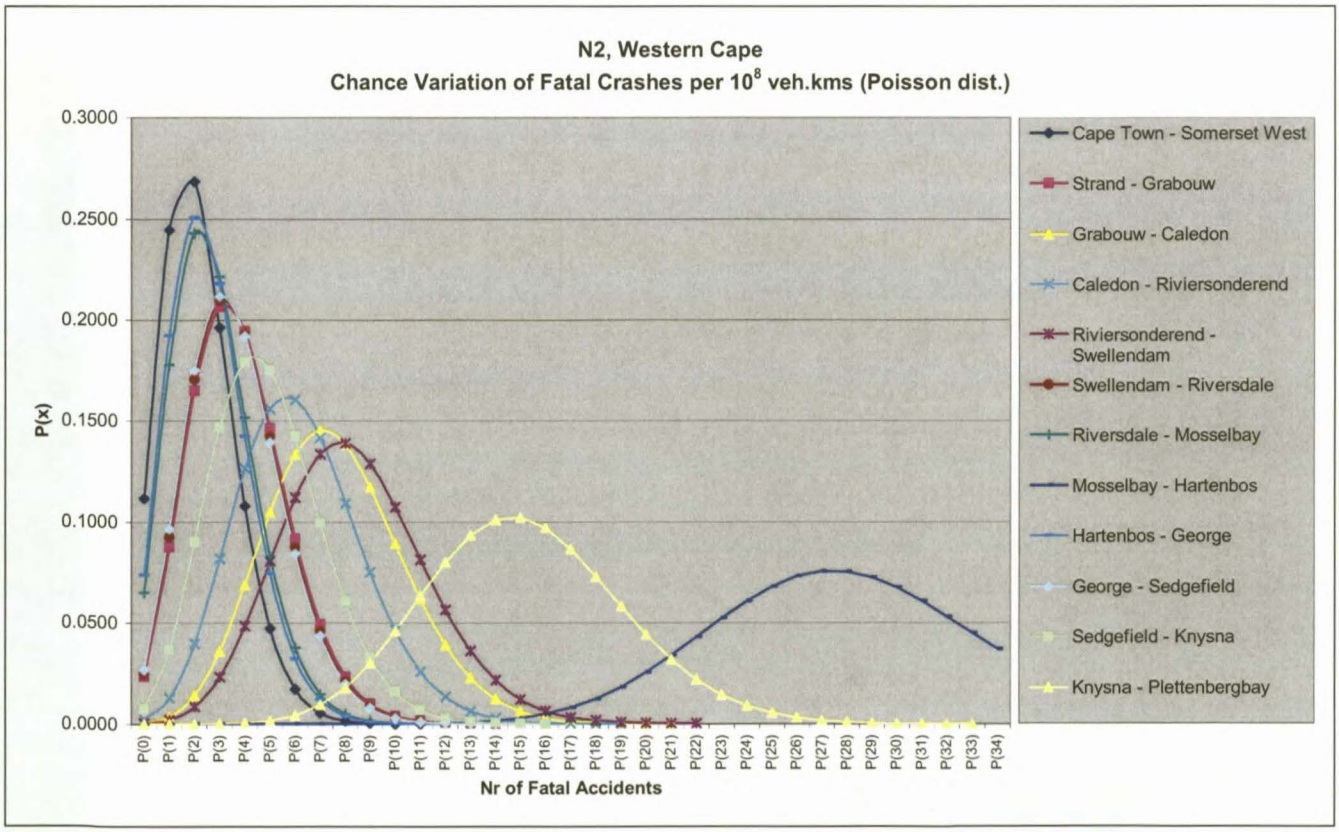

Fig. 4.6.2: Chance Variation of Fatal Accident Rates for the N2, Western Cape Province (Poisson dist.)

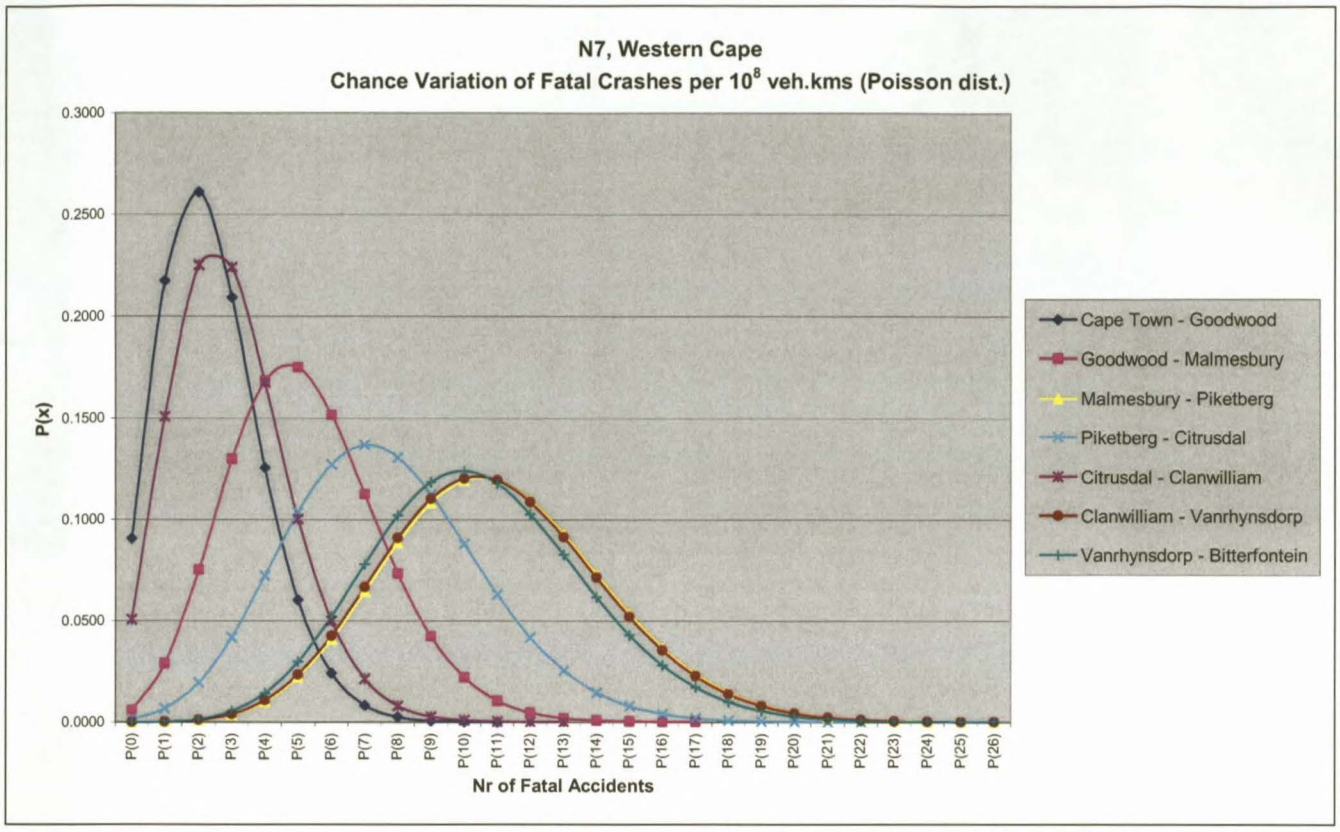

Fig. 4.6.3: Chance Variation of Fatal Accident Rates for the N7, Western Cape Province (Poisson dist.)

It is important to note that one of four possible scenarios (Scenarios A to D below) might illustrate the year-to-year chance variation of fatal accident rates for each road section as given in Table 4.6.1 (except for the scenario where no chance variation exists). The bell-shaped curves might also help to understand the chance variation scenarios:

A. An increase in the fatal accident rate as well as an increase in the probability for the increased fatal accident rate. This scenario indicates that the accident rate was most probably already below the expected rate, increased the next year, but the increased accident rate has a higher probability to exist. There is thus no reason for major concern, as the increased accident rate is just closer to what it is "expected" to be.

B. An increase in the fatal accident rate and a decrease in the probability for the increased fatal accident rate. This scenario indicates that the accident rate was most probably already above the expected rate, increased the next year with the increased accident rate having a lower probability to exist. It could be that this is in fact an "abnormal" increase in accident rate and that there should be some concern.

C. A decrease in the fatal accident rate and an increase in the probability for the decreased fatal accident rate. This scenario indicates that the accident rate was most probably already above the expected rate, decreased the next year with the decreased accident rate having a higher probability to exist. This indicates that there might be improvement and that there should not necessarily be reason for concern, but also, it should not necessarily be seen as an improvement due to induced remedial actions, because the increased probability might only be an indication that accident rates are returning to the "expected" rates and that is just chance variation.

D. A decrease in the fatal accident rate and a decrease in the probability for the decreased fatal accident rate. This scenario indicates that the accident rate was most probably already below the expected rate, decreased the next year with the decreased accident rate having a lower probability to exist. This should thus not be seen as a positive decrease in accident rates.

4.6.1 Chance Variation in Fatal Accident Rates: 2002-2003

The following table contains the national road sections along the N1, N2 and N7 within the Western Cape Province categorized according to Scenarios A to D (explained earlier) in terms of their fatal accident rate chance variation from the year 2002 to 2003. These results can be verified by inspecting each road section for its Poisson distribution plot provided in either Figure 4.6.1, 4.6.2 or 4.6.3.

Table 4.6.2: Chance Variation Categorization for Change in Fatal Accident Rates (fatal accidents per 100 mill veh-km) on National Road Sections from 2002 to 2003

Scenario A (increase in rate, increase in prob.)	Scenario B (increase in rate, decrease in prob.)	Scenario C (decrease in rate, increase in prob.)	Scenario D (decrease in rate, decrease in prob.)
➢ Cape Town – Goodwood (N1) ➢ Cape Town – Somerset West (N2) ➢ Knysna – Plettenbergbay (N2) ➢ Swellendam – Riversdale (N2) ➢ Mosselbay – Hartenbosch (N2) ➢ Goodwood – Malmesbury (N7) ➢ Piketberg – Citrusdal (N7) ➢ Citrusdal – Clanwilliam ➢ Clanwilliam – Vanrhynsdorp	➢ Paarl – Worcester (N1) ➢ Touwsriver – Laingsburg (N1) ➢ Leeuw Gamka – Beaufort West (N1) ➢ George – Sedgefield (N2) ➢ Sedgefield – Knysna (N2) ➢ Strand – Grabouw (N2) ➢ Riviersonderend – Swellendam (N2) ➢ Cape Town – Goodwood (N7) ➢ Malmesbury – Piketberg (N7)	➢ Bellville – Paarl (N1) ➢ Grabouw – Caledon (N2) ➢ Caledon – Riviersonderend (N2) ➢ Riversdale – Mosselbay (N2)	➢ Goodwood – Bellville (N1) ➢ Worcester – Touwsriver (N1) ➢ Laingsburg – Leeuw Gamka (N1) ➢ Hartenbosch – George (N2) ➢ Vanrhynsdorp – Bitterfontein (N7)
*Beaufort West – Three Sisters showed no chance variation between 2002 and 2003 *Road Section Somerset West – Strand omitted.			

4.6.2 Chance Variation in Fatal Accident Rates: 2003-2004

The following table contains the national road sections on the N1, N2 and N7 within the Western Cape Province categorized according to Scenarios A to D (explained earlier) in terms of their fatal accident rate chance variation from the year 2003 to 2004. These results can be verified by inspecting each road section for its Poisson distribution plot provided in either Figure 4.6.1, 4.6.2 or 4.6.3.

Table 4.6.3: Chance Variation Categorization for Change in Fatal Accident Rates (fatal accidents per 100 mill veh-km) on National Road Sections from 2003 to 2004

Scenario A (increase in rate, increase in prob.)	Scenario B (increase in rate, decrease in prob.)	Scenario C (decrease in rate, increase in prob.)	Scenario D (decrease in rate, decrease in prob.)
➢ Cape Town – Goodwood (N1) ➢ Goodwood – Bellville (N1) ➢ Bellville – Paarl (N1) ➢ Laingsburg – Leeuw Gamka (N1) ➢ Hartenbosch – George (N2) ➢ Sedgefield – Knysna (N2) ➢ Knysna – Plettenbergbay (N2) ➢ Caledon – Riversonderend (N2) ➢ Citrusdal – Clanwilliam (N7) ➢ Vanrhynsdorp – Bitterfontein (N7)	➢ Worcester – Touwsriver (N1) ➢ Beaufort West – Three Sisters (N1) ➢ Cape Town – Somerset West (N2) ➢ Swellendam – Riversdale (N2) ➢ Riversdale – Mosselbay (N2) ➢ Mosselbay – Hartenbosch (N2) ➢ Goodwood – Malmesbury (N7) ➢ Piketberg – Citrusdal (N7) ➢ Clanwilliam – Vanrhynsdorp (N7)	➢ Paarl – Worcester (N1) ➢ Leeuw Gamka – Beaufort West (N1) ➢ Strand – Grabouw (N2) ➢ Riviersonderend – Swellendam (N2) ➢ Cape Town – Goodwood (N7) ➢ Malmesbury – Piketberg (N7)	➢ Touwsriver – Laingsburg (N1) ➢ George – Sedgefield (N2) ➢ Grabouw – Caledon (N2)

*Road Section Somerset West – Strand was omitted due to uncertainty and inappropriate level of detail regarding the specific description of the route.

4.6.3 Change in Chance Variation Scenarios between 2002 and 2004

Table 4.6.4 is a diagram giving each national road section in the Western Cape Province and how each road section's chance variation scenario (as explained earlier) changed from the variation in fatal accident rates scenario from 2002 to 2003 to the variation in fatal accident rates scenario from 2003 to 2004. Most road sections were found to have a change from Scenario A to B and from Scenario B to C (see Table 4.6.4 for the road section descriptions). Only these scenario changes will be highlighted in the paragraphs which are to follow and other scenario changes can be interpreted using the same arguments.

The road sections which underwent a change from A to B firstly indicated an increase in fatal accident rate with an increase in probability for the greater fatal accident rate from 2002 to 2003 (this leads to no reason to necessarily be concerned for these road sections, because the greater fatal accident rate could also just be closer to the "expected" fatal accident rate for that road section) and then, after having another increase in fatal accident rate from 2003 to 2004, a decrease in the probability for the greater fatal accident rate took place. According to this latter scenario there is more reason for concern due to the possibility that the higher fatal accident rate is in fact an "abnormal" increase and that safety measures might have to be considered. A more optimistic view would be that fatal road accident rates might decrease again in the following year based on chance variation and that there should not be too much premature concern. In general, change from scenario A to scenario B will be viewed as a negative effect.

Road sections which underwent a change in chance variation scenario from B to C firstly indicated an increase in fatal accident rate with a decrease in probability for the greater fatal accident rate from 2002 to 2003 (the same argument is valid as explained in the last lines of the previous paragraph). The road sections then experienced a decrease in fatal accident rate (generally seen as a positive effect), but with an increase in probability for the smaller fatal accident rate from 2003 to 2004, suggesting a positive effect in that accident rates are most probably decreasing to be nearer to what is "expected". A change from scenario B to scenario C will thus be viewed as positive.

Three years is a relatively short period when chance variation in accident frequencies/rates is investigated, because only one change in the scenario can be observed. Historical data for a period of at least five years could provide more certainty to a particular road section's accident frequency/rate chance variation behaviour i.e. three changes in chance variation scenarios can be observed.

Table 4.6.4: Year-to-Year Change in Chance Variation Scenarios between 2002 and 2004 per National road section, Western Cape Province

$A \rightarrow A$ Cape Town – Goodwood (N1) Knysna – Plettenbergbay (N1) Citrusdal – Clanwilliam (N7)	
$A \rightarrow B$ Cape Town – Somerset West (N2) Swellendam – Riversdale (N2) Mosselbay – Hartenbosch (N2) Goodwood – Malmesbury (N7) Piketberg – Citrusdal (N7) Clanwilliam – Vanrhynsdorp (N7)	$B \rightarrow A$ Sedgefield – Knysna (N2)
$B \rightarrow C$ Paarl – Worcester (N1) Leeuw Gamka – Beaufort West (N1) Strand – Grabouw (N2) Riviersonderend – Swellendam (N2) Cape Town – Goodwood (N7) Malmesbury – Piketberg (N7)	$C \rightarrow B$ Riversdale – Mosselbay (N2)
$B \rightarrow D$ Touwsriver – Laingsburg (N1)	$D \rightarrow B$ Worcester – Touwsriver (N1)
$C \rightarrow A$ Bellville – Paarl (N1) Caledon – Riviersonderend (N2)	
$C \rightarrow D$ Grabouw – Caledon (N2)	
$D \rightarrow A$ Goodwood – Bellville (N1) Laingsburg – Leeuw Gamka (N1) Vanrhynsdorp – Bitterfontein (N7)	

4.7 Multiple Regression Models

The literature review for Multiple Regression Models explaining the theory for *model selection* and checking *model utility* and *adequacy* was provided in Chapter 2. Chapter 3 discussed the application of

the particular type of multiple regression model selected (*General Additive Multiple Regression Model with Qualitative Predictor Variables*), particularly in terms of the number of predictors included for each model and how the final number of predictors were decided upon. Three models were created with each model predicting a different dependent variable as explained before (refer to Chapter 3 for the methodology) and will be given below under the relevant subheadings. Each prediction model will be discussed in terms of each model's *utility* and *adequacy*.

4.7.1 Prediction Model for FRate: Fatalities per 100 million veh-km

i) Model Statistics and Equation: FRate

Tables 4.7.1 and 4.7.2 contain the model statistics and regression summary of the prediction model for *FRate* i.e. the dependent variable which is to be predicted (number of fatalities per 100 mill veh-km). The *utility* and *adequacy* of the model will be discussed based on the information in the tables below as well as the regression output in Appendix D2.

Table 4.7.1: Summary Statistics for Multiple Regression Analysis – FRate (Fatalities per 100 mill veh-km)

	Value
Multiple R	0.480143
Multiple R²	0.230538
Adjusted R²	0.096718
F(4,23)	1.722751
p	0.179303
Std.Err. of Estimate	7.477288

Table 4.7.2: Regression Summary for Multiple Regression Analysis – FRate (Fatalities per 100 mill veh-km)

	Beta	Std.Err. of Beta	B	Std.Err. of B	t(23)	p-level
Intercept			3.82244	10.46835	0.36514	0.718342
ADTT	-0.419246	0.186844	-0.00420	0.00187	-2.24383	0.034766
AvgHVehSpeed	0.288915	0.215383	0.19276	0.14370	1.34140	0.192886
F	-0.291912	0.300939	-4.82888	4.97822	-0.97000	0.342137
R	-0.314167	0.277129	-4.96974	4.38383	-1.13365	0.268620
ADT	Excluded					
AvgSpeed	Excluded					
AvgNightSpeed	Excluded					
AvgLVehSpeed	Excluded					
M	Excluded					

$$FRate = -0.00420ADTT + 0.19276HVehS - 4.82888F - 4.96974R + 3.82244 \quad (1)$$

Where,

$ADTT$	=	Average Daily Truck Traffic (veh/d)
$HVehS$	=	Average Heavy Vehicle Speed (km/h)
F	=	1, when *Flat* terrain
		0, otherwise
R	=	1, when *Rolling* terrain
		0, otherwise

ii) Model Utility: FRate

R^2 equals approximately 0.23, which means that approximately 23% of the dependent variable's variation can be "explained" by the relationship between the dependent variable (*FRate*) and the relevant predictors. This is not a particularly high value for R^2 and statistically it discredits the model in terms of its utility.

The *p*-significance value for the model is given as approximately 0.18. The null hypothesis states that there is no significant relationship between the dependent variable and any of the predictors. The hypothesis is rejected if the *p*-value is less than of equal to a chosen significance level. The most common significance levels used are 0.10, 0.05, 0.01 and 0.001. The *p*-significance for this model is higher than 0.10, indicating that the hypothesis should not be rejected at any significance level i.e. there is a significant probability that the prediction model does not provide an adequate relationship between the dependent variable and any of the predictors.

iii) Model Adequacy: FRate

Appendix D2 contains the normal probability plot of the expected normal value of residuals vs. the residual values. It was discussed in Chapter 2 how the random deviations (*e*) are assumed to be normally distributed. If a linear relationship is found on the normal probability plot, the normality assumption is plausible and the model can be accepted as adequate. On the normal probability plot for the model under discussion in this section, a few points deviated from the fitted line, but a reasonably

acceptable linear relationship was visible. The model can thus be accepted to be adequate based on this argument.

Another model assumption made was that the variance of a random deviation (e) is a constant. The plot of residuals vs. predicted values should illustrate points which are randomly placed with no discernible pattern. If this is indeed the case the assumption is plausible and the model can accepted as adequate based on this argument. The plot of predicted values vs. residuals for the model under discussion in this section appears to have its points placed at random with no discernible pattern (see Appendix D2). The model assumption of a constant variance for any random deviation thus is plausible and the model can be accepted as adequate.

4.7.2 Prediction Model for FARate: Fatal Accidents per 100 million veh-km

i) Model Statistics and Equation: FARate

Tables 4.7.3 and 4.7.4 contains the model statistics and regression summary of the prediction model for *FARate* i.e. the dependent variable which is to be predicted (number of fatal accidents per 100 mill veh-km). The *utility* and *adequacy* of the model will be discussed based on the information in the tables below as well as the regression output in Appendix D2.

Table 4.7.3: Summary Statistics for Multiple Regression Analysis – FARate (Fatal Accidents per 10^8 veh-km)

	Value
Multiple R	0.413624
Multiple R²	0.171085
Adjusted R²	0.026926
F(4,23)	1.186780
p	0.342966
Std.Err. of Estimate	5.485258

Table 4.7.4: Regression Summary for Multiple Regression Analysis – FARate (Fatal Accidents per 100 mill veh-km)

	Beta	Std.Err. of Beta	B	Std.Err. of B	t(23)	p-level
Intercept			3.08702	7.679470	0.40198	0.691407
ADTT	-0.346120	0.193928	-0.00245	0.001373	-1.78478	0.087494
AvgHVehSpeed	0.253168	0.223549	0.11938	0.105416	1.13249	0.269097
F	-0.316153	0.312349	-3.69645	3.651969	-1.01218	0.321983
R	-0.233297	0.287636	-2.60840	3.215934	-0.81109	0.425631
ADT	Excluded					
AvgSpeed	Excluded					
AvgNightSpeed	Excluded					
AvgLVehSpeed	Excluded					
M	Excluded					

$$FARate = -0.00245ADTT + 0.11938HVehS - 3.69645F - 2.60840R + 3.08702 \qquad (2)$$

Where,

$ADTT$	=	Average Daily Truck Traffic (veh/d)
$HVehS$	=	Average Heavy Vehicle Speed (km/h)
F	=	1, when *Flat* terrain
		0, otherwise
R	=	1, when *Rolling* terrain
		0, otherwise

ii) Model Utility: FARate

R^2 equals approximately 0.17, which means that 17% of the dependent variable's variation can be "explained" by the relationship between the dependent variable (*FARate*) and the relevant predictors. This is a relatively small value for R^2 (and smaller than R^2 of the previous model for *FRate*) and statistically it discredits the model in terms of its utility.

The *p*-significance value for the model is given as approximately 0.34. The null hypothesis states that there is no significant relationship between the dependent variable and any of the predictors. The hypothesis is rejected if the *p*-value is less than of equal to a chosen significance level. The most common significance levels used are 0.10, 0.05, 0.01 and 0.001. The *p*-significance for this model is significantly higher than 0.10, indicating that the hypothesis should not be rejected at any significance

level i.e. there is a significant probability that the prediction model does not provide an adequate relationship between the dependent variable and any of the predictors.

iii) Model Adequacy: FARate

The same arguments in terms of model adequacy are valid for the model under discussion in this section as was valid for the previous model. It was found that for this model the normality assumption for random deviations (e) and the constant variance assumption for any random deviation (e) were plausible and that the model was thus found to be adequate.

4.7.3 Prediction Model for FFARate: Fatalities per Fatal Accident

i) Model Statistics and Equation: FFARate

Tables 4.7.5 and 4.7.6 contains the model statistics and regression summary of the prediction model for *FFARate* i.e. the dependent variable which is to be predicted (number of fatalities per fatal accident). The *utility* and *adequacy* of the model will be discussed based on the information in the tables below as well as the regression output in Appendix D2.

Table 4.7.5: Summary Statistics for Multiple Regression Analysis – FFARate (Fatalities per Fatal Accident)

	Value
Multiple R	0.634028
Multiple R^2	0.401992
Adjusted R^2	0.297990
F(4,23)	3.865249
p	0.015209
Std.Err. of Estimate	0.320601

Table 4.7.6: Regression Summary for Multiple Regression Analysis – FFARate (Fatalities per Fatal Accident)

	Beta	Std.Err. of Beta	B	Std.Err. of B	t(23)	p-level
Intercept			1.085373	0.272963	3.97626	0.000597
ADT	-0.495462	0.171291	-0.000007	0.000003	-2.89251	0.008213
F	0.806179	0.323753	0.648616	0.260477	2.49010	0.020435
R	0.354876	0.329434	0.273030	0.253456	1.07723	0.292548
M	0.304935	0.245182	0.327435	0.263273	1.24371	0.226135
RouteDist	Excluded					
ADTT	Excluded					
AvgSpeed	Excluded					
AvgNightSpeed	Excluded					
AvgLVehSpeed	Excluded					
AvgHVehSpeed	Excluded					

$$FFARate = -0.000007ADT + 0.648616F + 0.273030R + 0.327435M + 1.085373 \quad (3)$$

Where,

ADT = Average Daily Traffic (veh/d)

F = 1, when *Flat* terrain
0, otherwise

R = 1, when *Rolling* terrain
0, otherwise

M = 1, when *Mountainous* terrain
0, otherwise

ii) Model Utility: FFARate

R^2 equals approximately 0.40, which means that approximately 40% of the dependent variable's variation can be "explained" by the relationship between the dependent variable (*FFARate*) and the relevant predictors. This value for R^2 is higher than those for the previous models, but is still not acceptable in terms of statistical theory. In the context of this study a higher R^2 is not likely expected.

The *p*-significance value for the model is given as approximately 0.015. The null hypothesis states that there is no significant relationship between the dependent variable and any of the predictors. The hypothesis is rejected if the *p*-value is less than of equal to a chosen significance level. The most common significance levels used are 0.10, 0.05, 0.01 and 0.001. The *p*-significance for this model is between 0.01 and 0.05, indicating that the hypothesis should not be rejected at significance levels 0.01

and 0.001 i.e. at those chosen significance levels there is a probability that the prediction model does not provide an adequate relationship between the dependent variable and any of the predictors. If a significance level of 0.05 is chosen, the hypothesis can be rejected i.e. the model is accepted (at a 0.05 significant level) to provide an adequate relationship between the dependent variables and the predictors.

iii) Model Adequacy: FFARate

The same arguments in terms of model adequacy are valid for the model under discussion in this section as was valid for the previous model. It was found that for this model the normality assumption for random deviations (e) and the constant variance assumption for any random deviation (e) were plausible and that the model was thus found to be adequate.

CHAPTER 5

CONCLUSIONS

In this chapter conclusions are drawn based on the findings of Chapter 4.

5.1 Detailed and Quality Road Accident Data for South Africa is Unavailable

If a road accident data module is created in NaTiS (as discussed in Chapter 2), road accident data will be even harder to obtain and will thus complicate future road accident data gathering. Road accident data are generally not available in the appropriate amount of detail or does not have the appropriate level of data quality to be suitable for road engineering purposes i.e. based on the findings for this study that the exact positions of the road accidents in the database are unknown. This had a major influence on all manual calculations done for the purpose of this study (e.g. traffic and speed estimates, fatality and fatal accident rates etc).

5.2 Correspondence Analysis is Found to be the Most Appropriate Analysis Technique for Road Accident Data in Spite of some Practical Limitations

Of the four analysis techniques investigated for the purpose of this study, it is concluded that Correspondence Analysis is found to be the most useful tool for analysing road accident data. In some instances, the visual output was found to be insufficient to interpret alone in which case the graphical representations need to be verified with the relative row/column frequencies of each input cross tabulation table (as discussed before). Sensible results could still be obtained when only the graphical representations were interpreted. Even though large amounts of unknown categories have influence on the reliability of a sample, it does not influence the interpretation of correspondence analysis results. It is concluded that this technique is theoretically and practically appropriate for application, especially for analysing large two-way frequency tables.

5.3 Small Sample Sizes and Unknown Variable Categories Have an Influence on Association Rules Analysis Results

Although theoretically adequate and useful to find hidden patterns between categorical variables and co-occurrence frequencies, Association Rules results are mainly influenced by sample size and large amounts of unknown variable categories. This compromises which association rules are found to be the most meaningful in terms of the *Lift* value and how accurate the output statistics are calculated (i.e. *confidence* values are calculated to be more inaccurate with decreasing sample size). Refer specifically to Tables 4.5.1 to 4.5.3.

5.4 A Relatively Short Analysis Period is Unsuitable when Determining Chance Variation in Accident Frequencies/Rates

Based on the findings of this study it is concluded that three years are too short for determining chance variation in accident frequencies/rates. It is not suitable for determining trends and a longer time period is necessary to obtain more conclusive results.

5.5 The General Additive Multiple Regression Model with Qualitative Predictors, Predicting Number of Fatalities per Fatal Accident, is Accepted in Terms of Utility and Adequacy

Statistical output for the multiple regression model predicting *FFARate* (number of fatalities per fatal accident) suggests a useful model in terms of its adequacy and utility. Refer to Section 4.7.3.

5.6 The Use of Too Few Data points Have an Influence on Multiple Regression Results

Based on the findings for Multiple Regression for the purpose of this study, it is concluded that too few data points (28 data points) compromised the results leading most probably to the rejection of two other prediction models created in addition to the accepted model as mentioned in section 5.5. above.

CHAPTER 6

RECOMMENDATIONS

This chapter makes recommendations based on the conclusions drawn in Chapter 5.

6.1 Better Quality Data should be Used When Applying any Analysis Technique to Road Accident Data

The unavailability of data in the appropriate amount of detail and the required level of quality influences the application of any data analysis technique. This is especially true for road accident data in South Africa and thus it is highly recommended that data quality improvement be considered as high priority. It is recommended that the relevant authorities make greater effort in improving South Africa's road accident data system in terms of reliability and quality, especially when it is necessary for road safety engineers to rely on historical road accident data for improving South Africa's roads.

6.2 Correspondence Analysis Should be Performed on Road Accident Data on an Annual Basis

It was concluded that Correspondence Analysis is a useful tool for investigating the correspondence between categorical variables in large two-way frequency tables. It is recommended that this technique be applied to road accident data on an annual basis to monitor any changes in correspondence between variables i.e. the results will provide information which cannot necessarily be interpreted directly from cross tabulation tables. Another example of a potential application is that road accident data based on road accidents with different levels of severity can be investigated and compared.

6.3 Graphical Output from Correspondence Analyses Should be Verified with Relative Row/Column Frequencies when *Quality* and Amount of *Overall Inertia* Representation are Inadequate

In cases where row/column points are poorly represented (their *Quality* statistics are low) and only a relatively small amount of the *overall inertia* of a cross tabulation table is represented by the relevant amount of extracted dimensions, the graphical output of Correspondence Analysis should be verified with the relative row/column frequencies and output statistics. This verification should also be done in

cases where graphical output is not visually sufficient for direct interpretation (even when zoomed in on) or where there is any doubt about results obtained from the visual representations only.

6.4 Larger Sample Sizes Should be Used when Applying Association Rules Analysis and Multiple Regression Analysis

It is recommended that larger sample sizes be used when Association Rules Analysis is performed to ensure the accuracy and reliability of the results (the output statistics). Multiple Regression Analysis results will also benefit from the use of more data points as input. The utility of the particular model will not necessarily improve from more data points, but it will be more reliable.

6.5 Unknown Variable Categories Should be Excluded When Applying Association Rules Analysis

Based on the significant influence which unknown variable categories have on association rules results, it is recommended that these unknown categories be excluded from any association rules analysis to ensure that the most meaningful association rules in terms of the *Lift* value be highlighted containing *Body* and *Head* category items which are known.

6.6 A Longer Time Period (i.e. five years) Must be Used When Determining Chance Variation of Accident Frequencies/Rates

Three years are not a sufficient time period to conclusively determine chance variation of accident frequencies/rates. It is recommended that a minimum of five years' worth of accident data be used for determining chance variation.

6.7 More Data points Must be Used for Multiple Regression Analyses

It is recommended that multiple regression analysis be performed on a sample containing more data points. In this study it was found that 28 points were insufficient. The sample used for multiple regression analysis in this study was based on datapoints representing the national roads in the Western Cape Province. It is recommended that more road types (i.e. provincial and other minor roads) be included to obtain more data points.

REFERENCES

AA. 1995. *New Southern African Book of the Road.* 2nd Edition. AA The Motorist Publications (Pty) Ltd.

DEVORE, J. & FARNUM, N. 1999. *Applied Statistics for Engineers and Scientists.* California State University. Duxbury Press.

ELVIK, R. & BORGER MYSEN, A. *Incomplete Accident Reporting: Meta-Analysis of Studies Made in 13 Countries.* Transportation Research Record 1665:133-140. National Research Council, Washington, DC.

EUROPEAN TRANSPORT SAFETY COUNCIL. 2001. *Sharing Responsibilities for Road Safety.* Briefing (electronic document).

HAUER, E. & HAKKERT, A.S. *Extent and Some Implications of Incomplete Accident Reporting.* Transportation Research Record 1185:1-10.

http://www.arrivealive.co.za

http://www.dft.gov.uk

http://www.transport.gov.za

MIKROS TRAFFIC MONITORING (PTY). *Comprehensive Traffic Observations: Yearbook.* 1998-2004. South African National Roads Agency Ltd (SANRAL).

O'DAY, J. 1993. *Accident Data Quality: A Synthesis of Highway Practice.* Transportation Research Board. NCHRP Synthesis 192:5.

OGDEN, K.W. 1996. *Safer Roads: A Guide to Road Safety Engineering.* Dept. of Civil Engineering. Monash University, Melbourne, Australia.

PROVINCIAL ADMINISTRATION OF WESTERN CAPE PROVINCE. http://rnis.wcape.gov.za/rnis/rnis_web_reports.main

STATSOFT, INC. 2004. STATISTICA (data analysis software system), version 7. www.statsoft.com.

TARKO, ANDREW P. & KANODIA, M. 2004. *Effective and Fair Identification of Hazardous Locations.* Transportation Research Record 1897:64-70. National Research Council, Washington, DC.

TURNER, S. 2002. *Defining and Measuring Traffic Data Quality: White Paper.* Traffic Data Quality Workshop. Federal Highway Administration, Washington, DC.

UNIVERSITY OF STELLENBOSCH. 2006. The Second Stellenbosch Data Mining Workshop. Conference Notes.

WORLD BANK. 2002. *Poor quality data are major obstacle to improving road safety, says World Bank.* British Medical Journal 324:1116, 11 May 2002.

BIBLIOGRAPHY

AUTOMOBILE ASSOCIATION OF SOUTH AFRICA. 1993. *Annual Traffic Safety Audit* 1991. Sponsored by Mutual & Federal Insurance Co Ltd.

BESTER, CHRISTO J. 2001. *Explaining National Road Fatalities*. Accident Analysis & Prevention 33:663-672.

BESTER, CHRISTO J. 2005. *Road Safety Engineering Course*. Notes. University of Stellenbosch (unpublished).

CHU, X., GUTTENPLAN, M. & BALTES, MICHAEL R. 2004. *Why People Cross Where They Do: The Role of Street Environment*. Transportation Research Record 1878:3-10. National Research Council, Washington, DC.

DEPARTMENT OF TRANSPORT, REPUBLIC OF SOUTH AFRICA. 2004. *Road Traffic and Fatal Crash Statistics: 1990-2003*. Arrive Alive.

ELVIK, R. 2004. *To What Extent is There Bias by Selection?: Selection for Road Safety Treatment in Norway*. Transportation Research Record 1897:200-205. National Research Council, Washington, DC.

GOOGLE EARTH. www.google.com

IVAN, JOHN N. 2004. *New Approach for Including Traffic Volumes in Crash Rate Analysis and Forecasting*. Transportation Research Record 1897:134-141. National Research Council, Washington, DC.

LORD, D. & PERSAUD, BHAGWANT N. *Accident Prediction Models With and Without Trend: Application of the Generalized Estimating*. Transportation Research Record 1717:102-10. National Research Council, Washington, DC.

LOTTER, H.J.S. & VAN NIEKERK, E. 2001. *Evaluation & Monitoring of Critical Road Traffic Safety Issues*. CSIR Transportek. Province of the Western Cape.

PAPACOSTAS, C.S. & PREVEDOUROS, P.D. 2001. *Transportation Engineering & Planning*. 3rd Edition. University of Hawaii at Manoa, Honolulu, Hawaii. Prentice Hall.

RAESIDE, R. & WHITE, D. 2004. *Predicting Casualty Numbers in Great Britain*. Transportation Research Record 1897:142-147. National Research Council, Washington, DC.

RIBBENS, H. *Pedestrian Facilities in South Africa: Research and Practice*. Transportation Research Record 1538:10-18. National Research Council, Washington, DC.

SARKAR, S. & ANDREAS, M. 2004. *Drivers' Perception of Pedestrians' Rights and Walking Environments*. Transportation Research Record 1878:75-82. National Research Council, Washington, DC.

VAN NIEKERK, A. 2005. *Road Safety Research Via The Internet: A Brief Comparison Of Two Developed Countries' Department Of Transport Websites Based On User friendliness And Available Road Safety Information*. Assignment for Road Safety Course. University of Stellenbosch. (unpublished).

WOUDBERG, M.L. 2005. *Free Flow Speed versus Road Width*. Final Year Thesis for B.Eng Degree. University of Stellenbosch (unpublished).