JORGE TIAGO BASTOS

# ANÁLISE ESTRATÉGICA DA SEGURANÇA VIÁRIA NO BRASIL:

# PESQUISA DE ÍNDICES E INDICADORES

Tese de Doutorado apresentada ao Departamento de Engenharia de Transportes da Escola de Engenharia de São Carlos, da Universidade de São Paulo e ao *Transportation Research Institute* da Hasselt University como parte dos requisitos para a obtenção do título de Doutor em Ciências (Programa de Pós-graduação em Engenharia de Transportes, área de concentração: Planejamento e Operação de Transportes) e *Doctor of Transportation Sciences*, respectivamente.

**Orientadores: Prof. Antonio Clóvis Pinto Ferraz**

**Prof[a]. Elke Hermans**

**Coorientador: Prof. Tom Brijs**

**São Carlos**
**Dezembro**
**2014**

Bastos, Jorge Tiago
B327a Análise estratégica da segurança viária no Brasil: pesquisa de índices e indicadores / Jorge Tiago Bastos; orientador Antonio Clóvis Pinto Ferraz; coorientadora Elke Hermans. São Carlos, 2014.

Tese (Doutorado) - Programa de Pós-Graduação em Engenharia de Transportes e Área de Concentração em Planejamento e Operação de Sistemas de Transporte -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2014.

1. Diagnóstico da segurança viária. 2. Metas de mortes no trânsito. 3. Clusters. 4. Indicadores de desempenho da segurança. 5. Exposição. 6. Data envelopment analysis. 7. Bootstrapping. 8. Estados brasileiros. I. Título.

## FOLHA DE JULGAMENTO

Candidato: Engenheiro **JORGE TIAGO BASTOS**

Título da tese: "Análise estratégica da segurança viária no Brasil: pesquisa de índices e indicadores"

Data da defesa: 09/12/2014

| **Comissão Julgadora:** | **Resultado:** |
|---|---|
| Prof. Titular **Antonio Clóvis Pinto Ferraz** **(Orientador)** (Escola de Engenharia de São Carlos/USP) | Aprovado |
| Profa. Dra. **Elke Hermans** (Universiteit Hasselt/Bélgica) | Approved |
| Profa. Dra. **Ana Paula Camargo Larocca** (Escola de Engenharia de São Carlos/USP) | Aprovado |
| Prof. Dr. **Tom Brijs** (Universiteit Hasselt/Bélgica) | Approved |
| Prof. Dr. **Archimedes Azevedo Raia Júnior** (Universidade Federal de São Carlos/UFSCar) | APROVADO |

Coordenador do Programa de Pós-Graduação em Engenharia de Transportes:
Prof. Associado **Paulo César Lima Segantine**

Presidente da Comissão de Pós-Graduação:
Prof. Associado **Paulo César Lima Segantine**

**JORGE TIAGO BASTOS**

# ROAD SAFETY STRATEGIC ANALYSIS IN BRAZIL: INDICATOR AND INDEX RESEARCH

Doctorate thesis presented to the Department of Transportation Engineering of the School of Engineering of São Carlos, from the University of São Paulo, and to the Transportation Research Institute of Hasselt University as part of the requirements for obtaining the degree of Doctor in Sciences and Doctor in Transportation Sciences, respectively.

**Promoters: Prof. Antonio Clóvis Pinto Ferraz**

**Prof. Elke Hermans**

**Co-promoter: Prof. Tom Brijs**

**São Carlos**
**December**
**2014**

Bastos, Jorge Tiago
B327r Road safety strategic analysis in Brazil: indicator and index research / Jorge Tiago Bastos; orientador Antonio Clóvis Pinto Ferraz; coorientadora Elke Hermans. São Carlos, 2014.

Tese (Doutorado) - Programa de Pós-Graduação em Engenharia de Transportes e Área de Concentração em Planejamento e Operação de Sistemas de Transporte -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2014.

1. Road safety diagnosis. 2. Traffic fatality targets. 3. Clusters. 4. Safety performance indicators. 5. Exposure. 6. Data envelopment analysis. 7. Bootstrapping. 8. Brazilian states. I. Título.

2014 | School for Transportation Sciences

DOCTORAL DISSERTATION

# Road safety strategic analysis in Brazil: indicator and index research

Doctoral dissertation submitted to obtain the degree of doctor of Transportation Sciences, to be defended by

**Jorge Tiago Bastos**

Promoters: Prof. Dr Elke Hermans | UHasselt
Prof. Dr Antonio Clóvis Pinto Ferraz | EESC-USP
Co-promoter: Prof. Dr Tom Brijs | UHasselt

This research has the support of

OBSERVATÓRIO
NACIONAL DE SEGURANÇA VIÁRIA

# DEDICATION

*To my nephews,*

*Caio, Isadora and Lara*

*for a safer tomorrow.*

# AKNOWLEDGEMENTS

At the end of this stage, I would like to acknowledge many people that contributed in the preparation of this joint doctorate thesis along these years.

First of all, my special thanks to my promoters, Dr. Elke Hermans and Dr. Coca Ferraz, for the trustworthy and for sharing their knowledge.

Next, my extensive gratitude to the members of my doctorate commission: Dr. Yongjun Shen, who dedicated special attention and support; Dr. Tom Brijs, my co-promoter, who was decisive for this joint work initiative; and Dr. Geert Wets, for his wise recommendations.

My sincere appreciation further goes to Dr. Heitor Vieria, MSc. Magaly Romão, and Dr. Bárbara Bezerra, for always motivating my career and me.

I would like to thank the National Road Safety Observatory (ONSV) team, especially Mr. José Aurélio Ramalho and Mr. Maximiliano Hahn Dalla Porta, for their motivating belief in this research potential. Moreover, I thank all sorts of data providers, including the collaboration of some state DETRANs.

My gratitude also to Dr. Antonio Nelson Rodrigues da Silva, coordinator of this joint degree agreement.

I thank my colleagues at USP and UHasselt, for their kind company and friendship. Additionally, I am also grateful for my colleagues of the Urban Mobility Division at Fetranspor.

It is impossible to imagine going through this period without my friends, especially those who were more closely involved in helping me and supporting me emotionally: Helder Fornari, Izabela Cardoso, Liesbeth Oeyen, Madalena Ribeiro, Adonais Sens, Rafael Godoi, Marcelo Mancini and many others.

I also would like to thank both USP and UHasselt staff, for their committed work in making this joint doctorate possible. In addition, my gratitude to Nadine Smeyers and Heloísa Belo for the “last minute” arrangements for the defense.

My acknowledgement to CNPq – *Conselho Nacional de Desenvolvimento Científico e Tecnológico* and the Science without Borders Program, for the financial support that enabled this fruitful experience.

Last but not least, my profound gratitude to my parents and family, whose always support me in my decisions and for being the source of infinite love.

Jorge Tiago Bastos

December 9th, 2014

# ABSTRACT

BASTOS, J. T. ***Road safety strategic analysis in Brazil: indicator and index research***. São Carlos, 2014, 290 p. Doctorate thesis – Double degree between *Escola de Engenharia de São Carlos (Universidade de São Paulo)* and *Instituut voor Mobiliteit (Universiteit Hasselt)*

The intense economic growth that Brazil has experienced in recent decades and its consequent explosive motorization process have resulted in an undesirable impact: the continuously increasing trend in traffic fatality numbers. This study presents a research on indicators and indexes with the objective of delivering both overall and disaggregated evidence about the road safety performance and targets in fatality reduction in Brazil at the state level taking the exposure into account. The intention is to support road safety strategic analysis in the country and to contribute to improve this critical scene. The methodological structure of this thesis consists of the following three main parts: (I) diagnosing the road safety situation at the state level using final outcome related information, in particular traffic fatality risk data; (II) setting a target number of traffic fatalities based on the relationship between the exposure level and the number of traffic fatalities in each state; and (III) suggesting domains for improvements based on the research of safety performance indicators representing three domains (road user, environment and vehicle) throughout the states. From a benchmarking point of view, we divided the Brazilian states into three separate clusters in order to provide more realistic state performance comparisons. After a data collection and indicators selection step, Data Envelopment Analysis (DEA) was the method used for executing the different steps, with the application of four different types of models specially developed for the identified research purposes. In addition, by means of bootstrapping the DEA scores we measured the sensitivity of the results to possible variations in the input data, for example concerning data quality and availability. As a result, we provided a road safety diagnosis per state as well as traffic fatality targets according to different perspectives: the entire group of road users (motorized and non-motorized ones), motor vehicle occupants, and finally a disaggregated performance evaluation by running four separate DEA models (for motorcycle, car, truck and bus). Moreover, the SPI research including a hierarchy of 27 safety performance indicators expressed the state's relative performance on the main road safety domains. Lastly, state profiles compiling all this information summarized the "per state" findings.

**Keywords:** Road safety diagnosis. Traffic fatality targets. Clusters. Safety performance indicators. Exposure. Data envelopment analysis. Bootstrapping. Brazilian states.

# RESUMO


BASTOS, J. T. ***Análise estratégica da segurança viária no Brasil: pesquisa de índices e indicadores***. São Carlos, 2014, 290 p. Tese de doutorado – Duplo diploma entre *Escola de Engenharia de São Carlos (Universidade de São Paulo)* e *Instituut voor Mobiliteit (Universiteit Hasselt)*

O intense crescimento econômico que o Brasil tem experimentado nas últimas décadas e seu consequente explosivo processo de motorização resultaram em um impacto indesejado: a tendência contínua do aumento do número de mortes no trânsito. Este estudo apresenta uma pesquisa acerca de índices e indicadores com o objetivo de fornecer evidências gerais e desagregadas sobre o desempenho da segurança viária e metas de redução no número de mortes no Brasil no âmbito estadual, levando a exposição em consideração. A intenção é embasar uma análise estratégica da segurança viária no país e contribuir para melhorar este cenário crítico. A estrutura metodológica desta tese consiste das seguintes três partes principais: (I) diagnóstico da situação da segurança viária no nível estadual utilizando informações relacionadas ao resultado final, em particular dados de risco de morte no trânsito; (II) estabelecer uma meta para o número de mortes no trânsito para cada estado; e (III) sugerir domínios para melhorias baseado em pesquisa de indicadores de desempenho da segurança viária voltada a três domínios (usuário da via, ambiente e veículo). Sob a ótica do benchmarking, dividiram-se os estados brasileiros em três *clusters* para proporcionar comparações mais realistas dos desempenhos estaduais. Após uma etapa de coleta e seleção de indicadores, utilizou-se o método de *Data Envelopment Analyis* (DEA) para executar as diferentes etapas, com a aplicação de quatro tipos distintos de modelos especialmente desenvolvidos para os propósitos da pesquisa. Além disso, por meio de *bootstrapping* dos escores obtidos com a DEA, mediu-se a sensibilidade dos resultados a possíveis variações nos dados de entrada, no que diz respeito a, por exemplo, qualidade e disponibilidade dos dados. Como resultado, propicia-se, a partir de diferentes perspectivas, um diagnóstico da segurança viária por estado, assim como metas no número de mortes: para todo o grupo de usuários (motorizados e não-motorizados), ocupantes de veículos motorizados, e finalmente uma avaliação desagregada por meio de quatro modelos separados (para motocicletas, automóveis, caminhões e ônibus). Adicionalmente, a pesquisa de indicadores de desempenho da segurança considerando a hierarquia de 27 indicadores expressou os desempenhos relativos dos estados nos principais domínios da segurança viária. Por fim, perfis estaduais compilando todas estas informações resumem os resultados para os estados.



**Keywords:** Diagnóstico da segurança viária. Metas de mortes no trânsito. Clusters. Indicadores de desempenho da segurança. Exposição. Data envelopment analysis. Bootstrapping. Estados brasileiros.

# SAMENVATTING

BASTOS, J. T. ***Strategische verkeersveiligheidanalyse in Brazilië: indicatoren en indexen onderzoek***. São Carlos, 2014, 290 p. Doctorate thesis – Double degree tussen *Escola de Engenharia de São Carlos (Universidade de São Paulo)* een *Instituut voor Mobiliteit (Universiteit Hasselt)*

De sterke economische groei die Brazilië de afgelopen decennia kende en het explosieve motorisatieproces dat daaruit voortvloeide, hadden een ongewenste impact: de continue stijging van het aantal verkeersdoden. Deze studie behandelt onderzoek naar indicatoren en indexen met als doel inzicht te verkrijgen in zowel de algemene als de gedesaggregeerde verkeersveiligheidsprestaties en de doelstellingen voor de verlaging van het aantal verkeersdoden in Brazilië op het niveau van de deelstaten, rekening houdend met de blootstelling. Wij beogen hierbij om de strategische verkeersveiligheidsanalyse in het land te ondersteunen en een bijdrage te leveren aan de verbetering van deze ernstige toestand. De methodologische structuur van deze thesis bestaat uit de volgende drie hoofdonderdelen: (I) de diagnose van de verkeersveiligheidssituatie op het niveau van de deelstaten op basis van informatie die gerelateerd is aan uiteindelijke resultaten, meer bepaald gegevens met betrekking tot het dodelijk risico; (II) het bepalen van een doelstelling voor het aantal verkeersdoden op basis van de relatie tussen het blootstellingsniveau en het aantal verkeersdoden voor elke deelstaat; en (III) het voorstellen van verbeterpunten op basis van het onderzoek in verband met verkeersveiligheidsprestatie-indicatoren voor drie domeinen (weggebruiker, omgeving en voertuig) voor alle deelstaten. Vanuit een benchmarking oogpunt hebben we de deelstaten van Brazilië opgedeeld in drie afzonderlijke clusters om hierdoor meer realistische vergelijkingen van de prestaties van de deelstaten mogelijk te maken. Na een eerste stap van gegevensverzameling en indicatorenselectie, werd de "Data Envelopment Analysis" (DEA) methode gebruikt voor het uitvoeren van de verschillende stappen, waarbij vier soorten modellen toegepast werden die speciaal ontwikkeld werden voor de geïdentificeerde onderzoeksdoeleinden. Bovendien konden we, door middel van het "bootstrappen" van de DEA scores, de sensitiviteit van de resultaten meten met betrekking tot mogelijke variaties bij de gegevens, bijvoorbeeld met betrekking tot de kwaliteit en beschikbaarheid van de gegevens. Op die manier verkregen wij een verkeersveiligheidsdiagnose per deelstaat alsook doelstellingen voor verkeersdoden voor verschillende categorieën: de volledige groep van weggebruikers (gemotoriseerd en niet-gemotoriseerd), inzittenden van gemotoriseerde voertuigen, en tot slot een gedesaggregeerde prestatie-evaluatie door het uitvoeren van vier afzonderlijke DEA-modellen (voor motor, auto, vrachtwagen en bus). Bovendien geeft het onderzoek op basis van een hiërarchie van 27 verkeersveiligheidsprestatie-indicatoren, de relatieve prestatie van de deelstaat weer met betrekking tot de belangrijkste verkeersveiligheidsdomeinen. Tot slot vatten de deelstaatprofielen, waarin alle informatie verzameld wordt, de bevindingen per deelstaat samen.

**Keywords:** Verkeersveiligheidsdiagnose. Doelstelling verkeersdoden. Clusters. Verkeersveiligheidsprestatie-indicatoren. Blootstelling. Data envelopment analysis. Bootstrappen. Braziliaanse staten.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ANP | *Agência Nacional do Petróleo Gás Natural e Biocombustíveis* |
| BEB | Brazilian Energy Balance |
| CCR | Charnes, Cooper, and Rhodes |
| CES | Cross-efficiency Score |
| CIS | Cross-index Score |
| CNT | *Confederação Nacional dos Transportes* |
| CONASS | *Conselho Nacional dos Secretários de Saúde* |
| DPRF | *Departamento de Polícia Rodoviária Estadual* |
| DEA | Data Envelopment Analysis |
| DENATRAN | *Departamento Nacional de Trânsito* |
| DETRAN | *Departamento Estadual de Trânsito* |
| DMU | Decision-Making Unit |
| FR | Fatality Rate |
| FSBE | Fuel Sale Based Estimative |
| GDP | Gross Domestic Product |
| IBGE | *Instituto Brasileiro de Geografia e Estatística* |
| INMETRO | *Instituto Nacional de Metrologia, Qualidade e Tecnologia* |
| METDFS | Method for Estimating the Traveled Distance using Fuel Sales |
| ML DEA-CI | Multiple Layer Data Envelopment Analysis Composite Indicator |
| ML DEA-ES | Multiple Layer Data Envelopment Analysis Efficiency Score |
| MR | Mortality Rate |
| OECD | Organization for Economic Co-operation and Development |

| | |
|---|---|
| OES | Optimum Efficiency Score |
| OIS | Optimum Index Score |
| ONSV | *Observatório Nacional de Segurança Viária* |
| PENSE | *Pesquisa Nacional de Saúde do Escolar* |
| R | R Core Team |
| SUN | Sweden, United Kingdom & the Netherlands |
| SPI | Safety Performance Indicator |
| VKT | Vehicle-Kilometer Traveled |
| VIGITEL | *Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telfônico* |
| VIVA | *Vilância de Violência e Acidentes* |
| WHO | World Health Organization |

| | |
|---|---|
| BR-27 | Set of 27 Brazilian States |
| AC | Acre |
| AL | Alagoas |
| AP | Amapá |
| AM | Amazonas |
| BA | Bahia |
| CE | Ceará |
| DF | Distrito Federal |
| ES | Espírito Santo |
| GO | Goiás |
| MA | Maranhão |
| MT | Mato Grosso |

| | |
|---|---|
| MS | Mato Grosso do Sul |
| MG | Minas Gerais |
| PA | Pará |
| PB | Paraíba |
| PE | Pernambuco |
| PI | Piauí |
| PR | Paraná |
| RJ | Rio de Janeiro |
| RN | Rio Grande do Norte |
| RS | Rio Grande do Sul |
| RO | Rondônia |
| RR | Roraima |
| SC | Santa Catarina |
| SP | São Paulo |
| SE | Sergipe |
| TO | Tocantins |

| | |
|---|---|
| EU-27 | Set of 27 European Union countries |
| AT | Austria |
| BE | Belgium |
| BG | Bulgaria |
| CY | Cyprus |
| CZ | Czech Republic |
| DK | Denmark |
| EE | Estonia |
| FI | Finland |

| | |
|---|---|
| FR | France |
| DE | Germany |
| EL | Greece |
| HU | Hungary |
| IE | Ireland |
| IT | Italy |
| LV | Latvia |
| LT | Lithuania |
| LU | Luxembourg |
| MT | Malta |
| NL | The Netherlands |
| RO | Romania |
| PL | Poland |
| PT | Portugal |
| SK | Slovakia |
| SI | Slovenia |
| ES | Spain |
| SE | Sweden |
| UK | United Kingdom |

# 1. INTRODUCTION

Road “unsafety” is a worldwide problem that affects millions of people with its measurable and immeasurable consequences of tragic statistics in injured and killed individuals. Over the last recent years, if we assume the evolution of the number of traffic fatalities as manifesting the road safety changes, many countries managed to improve their situation in the global perspective. That is the case for several high and middle-income nations and an exception for a small group of low-income nations. Unfortunately, Brazil is not included in any of these cases. The number of traffic fatalities in the country presents a growing tendency in the same period and the situation will probably continue to deteriorate if adequate effective measures are not taken into action.

A clear policy of encouraging people to purchase their first vehicle remains in Brazil (while there is an inadequate environment in terms of roads infrastructure, users behavior and many other related aspects) and it overrides the concern on road safety with unacceptable levels of casualties. In addition, there is no institution with governmental support responsible exclusively for road safety monitoring and guidelines formulation at the national level and on a regular basis. These and many other unfavorable aspects constitute a barrier for a systematic and scientific road safety related monitoring process and practical research in the country.

In order to deliver some contribution for the improvement of this critical situation, this research intends to execute a road safety indicator and index research, better supporting strategic analysis with respect to the road safety issue in Brazil and its member states[1]. The continental dimension of the country and the contrasting figures between its states require a disaggregated approach, since nationwide recommendations probably also generate diverse effects depending on the state’s scenery. Moreover, the majority of the results are presented in terms of clusters of states with the purpose of offering a more feasible basis for comparisons and knowledge transference between the set of states.

[1] In this research, the 26 Brazilian states and the Federal District (*Distrito Federal – DF*) are referred as states, although DF is actually defined as a federation unit.

The methodological core of this research lies in the application of four different Data Envelopment Analysis (DEA) models specially developed for the investigation of the road safety performance of large geographical areas, such as countries and states. Therefore, the application of DEA models produces the needed values for the quantitative and qualitative evaluation of the Brazilian states. In addition to several other methodological contributions, the presentation of a sensitivity analysis technique for testing the influence of uncertainties in the obtained DEA results is equally important.

This manuscript addresses the individual performance of Brazilian states through the analysis of information on different road safety outlooks: exposure to the risk, traffic fatality related indicators, set targets on traffic fatalities, and safety performance indicators. The combined investigation of these perspectives is beneficial in the construction of a more solid, maintainable and fruitful approach of road safety assessment, since it explores the essential aspects of diagnosing, setting attainable targets and recommending the most adequate lines of action for potential improvements.

It is always challenging to gather sufficiently reliable, complete and available data for road safety research. In Brazil and other developing nations this challenge is even bigger. This picture frequently pushes researches to the dilemma of “doing nothing”, with the argument that there is not enough available information, or “working under some limited assumptions”. This thesis describes a research in which this judgment is a recurrent situation due to the Brazilian unfavorable background scenery on data availability. It requires the adoption of alternative strategies, which might not be the ideal ones, but represent the best possible to capture from the real world situation throughout the existing information.

This thesis aims at executing road safety indicator and index research in order to better support strategic road safety analysis in Brazil and its various states, based on applications of data envelopment analysis techniques which were specially developed for road safety research purposes. The originality of this study lies in the consideration of road safety as a multidimensional concept in contrast to the single aspect evaluation generally performed; in the detailed state investigation; internal benchmarking process based on state comparisons, this is, inside the same national perspective (not comparing different countries); and in applying this for the first time to a developing country. The methodological structure (related to the results we search for) of the thesis consists of three main parts stated as follows:

I. Diagnosing the road safety situation at the state level using final outcome related information, this is, traffic fatality risk data;
II. Setting the target number of traffic fatalities based on the relationship between the exposure level and the number of traffic fatalities in each state;
III. Suggesting domains for improvements based on the research of safety performance indicators throughout the states.

The development of several steps that constitute this work is built around the seeking for answers to the following eight research questions:

**RQ1:** Since there is an unfavorable data availability scenario, is the quantity and quality of Brazilian data sufficient to execute this kind of analysis?

**RQ2:** Given the contrasting situation between the different Brazilian states, which are the most adequate clusters of states for conducting the research?

**RQ3:** Since the application of a DEA model for the mentioned purposes is inedited in Brazil (and also in many other non-European countries/developing countries), how to assess the applied models' performance in the country?

**RQ4:** How to test the sensitivity of the DEA model results, which manifest the "per state" road safety performance?

**RQ5:** Which are the best and the underperforming states with respect to the road safety?

**RQ6:** Based on a benchmarking exercise, which is the target number of fatalities for each state?

**RQ7:** How can the safety performance indicators research offer guidelines for improving road safety?

**RQ8:** Given the precedence of the application of a similar methodology for the European Union member countries, what can they learn from the Brazilian experience on DEA road safety research?

The next sections provide background knowledge about the Brazilian situation from an international perspective on road safety, enabling the comparative analysis of the country and its states in relation to the rest of the world and nationwide. Afterwards, an analysis of the evolution in traffic fatalities and related indicators (e.g. motorization rate) is carried out in order to provide some insight on how the Brazilian situation tends to develop over the upcoming years.

## 1.1. THE INTERNATIONAL PERSPECTIVE

The analysis of the road safety situation in an international (worldwide) perspective offers important evidences on how the country is managing its traffic fatalities problem in comparison to different sceneries. However, such type of comparison is usually subjected to significant dissimilarities in road safety related databases, which tend to e.g. present variations in underreporting levels over the countries. In order to overcome this difficulty and allow cross-country comparisons to be made, in the most recent Global Status Report on Road Safety elaborated by the World Health Organization (WHO, 2013) a regression model was used to estimate the number of traffic fatalities in 78 countries without an eligible traffic fatalities registration system. The base-year for the comparison is 2010, although there is already available Brazilian data for 2012, which information will be included in the upcoming chapters.

In terms of absolute number of fatalities, with 43,869 traffic fatalities in 2010 Brazil endures the 4th position in the world according to the WHO report; only China (275,983), India (231,027) and Nigeria (53,339) present higher values. Indonesia (42,434) completes the top-five list. The total number of fatalities in these five countries corresponds to about 50% of the estimated total number of traffic fatalities in the world for the same year.

For cross-country comparisons, we computed the fatality rate[2] (fatalities per 10,000 registered vehicles) using the information provided in the WHO report. Then, we classified the fatality rates into five groups intended to express the road safety performance of the countries:

- Best performing – from 0 to 2.5 fatalities/10,000 vehicles;
- Good performing – from 2.5 to 5.0 fatalities/10,000 vehicles;
- Average performing – from 5.0 to 7.5 fatalities/10,000 vehicles;
- Bad performing – from 7.5 to 10.0 fatalities/10,000 vehicles;
- Worst performing – more than 10.0 fatalities/10,000 vehicles.

[2] We discuss the different fatality related indicators characteristics in Chapter 2.

The worldwide situation is expressed in Figure 1 (according to 2010 data), where Brazil is classified as an average performing country, with similar levels of fatality rates as e.g. Uruguay, Chile and Mexico in Latin America; Liberia in Africa; Russia and Turkey in Europe; and Indonesia and Vietnam in Asia.

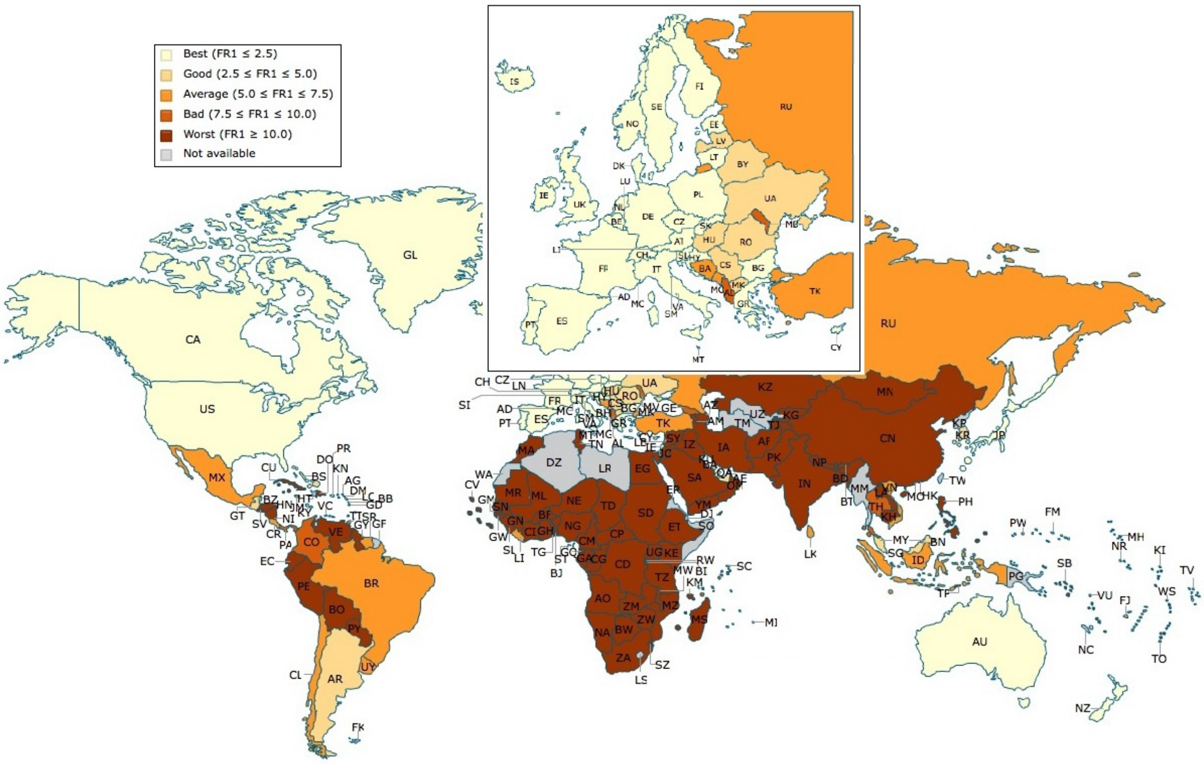



**Figure 1.1** – Worldwide road safety performance according to the fatality rate.

Based on WHO (2013)

The per-state disaggregated picture presented in Figure 1.2 shows very contrasting situations among the 27 Brazilian states. São Paulo (SP) and Rio Grande do Sul (RS) exhibit a better than average situation, presenting similar fatality rates as some Eastern European countries (e.g. Romania, Hungary) and Argentina. They are followed by six states that present performances corresponding to the national average and could be compared with e.g. Russia, Mexico and Chile. On the other hand, a group of nine Brazilian states presents below average bad road safety performances, with similar levels as e.g. Albania, Thailand and Colombia; and even nine states considered worst performing, comparable to India, China (and most Asian countries), Bolivia and some other Latin American low performing countries, and most African nations.

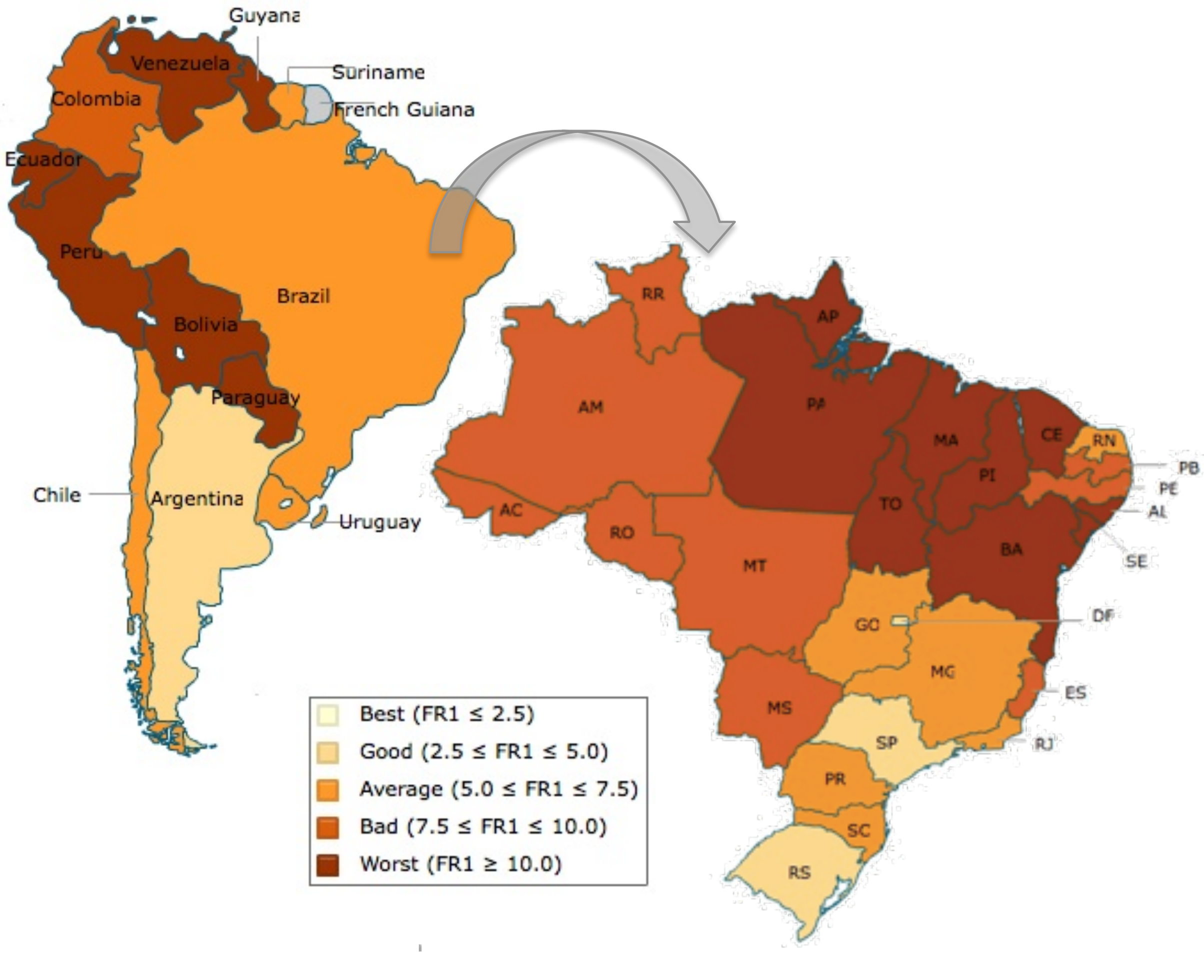



**Figure 1.2** – South American and Brazilian road safety performance according to the fatality rate

Based on WHO (2013)

## 1.2. NATIONAL OVERVIEW ON ROAD SAFETY

This section comprises a synthesis of the main perspectives regarding the traffic fatality issue in Brazil. It contains some state comparisons revealing very contrasting situations. In general, Brazil has a much lower road safety level than most developed countries and, although the Federal Government has set targets for traffic fatalities reduction, the scene evolution points in the opposite direction as there is a strong tendency of increases in the number of traffic fatalities to expect in the country for the upcoming years. The structure and data presented in this chapter has been inspired on the *Road Safety Vademecum* (European Commission, 2014) and the National Road Safety Observatory – ONSV (2014). The traffic fatality data was obtained from the Ministry of Health (*Ministério da Saúde*, 2014).

### 1.2.1 TARGETS AND REAL NUMBERS

According to the most recent statistics for 2012, 44,582 people died in traffic accidents in Brazil. It represents a 4% increase in relation to the year 2011. The number of traffic fatalities per 100,000 inhabitants (or the mortality rate) is 22.98[3,4]; the number of traffic fatalities per 10,000 vehicles rate (or the fatality rate) is 5.96[5]. At the beginning of the Decade of Action for Road Safety proclaimed by the United Nations as the period 2011-2020, the Brazilian Government set the ambitious target of reducing by half the predicted number of traffic fatalities for 2020[6]. It implies to decrease from 42,542 traffic fatalities in 2010 to 35,156[7] in 2020.

However, the fact is that in the first years of the decade the traffic fatalities number has been increasing, a tendency that is taking place since the year 2000 (with an isolated interruption from 2008 to 2009). This increasing trend resurged after the significant impact of the "New Brazilian Traffic Code" implementation in 1998; the result was a 20% lower number of traffic fatalities in 2010 compared to 1997 (before the new legislation's implementation). Supposing that from 2010 onwards efficacious actions towards road safety have been put into practice and some contribution to the target for 2020 already took place, in the years 2011 and 2012 a total of 4,891 lives could have been saved in the country.

The graph in Figure 1.1 shows the evolution in the number of fatalities since 2000 (the beginning of the decade) until 2012 (the most recent year with available information). Furthermore, we present the projection of the annual reductions necessary to reach the target set for 2020 (in the assumption of a constant decrease over time). It is evident that the Brazilian situation is evolving in the opposite direction in comparison to the ambitious target, because the number of traffic fatalities continues to grow year after year. It decreases the probability that impactful actions will result in the desired effect given the shorter remaining time period until the end of the Decade of Action for Road Safety.

---

[3] The global rate is 18.0 traffic fatalities per 100,000 inhabitants (WHO, 2013)
[4] While in United Kingdom, one of the lowest rates in the world, it is 2.8 traffic fatalities per 100,000 inhabitants in the same year (IRTAD, 2014);
[5] While in United Kingdom it is 0.5 traffic fatalities per 10,000 vehicles (IRTAD, 2014).
[6] Although setting this target in the national plan for accident reduction and road safety, it is still waiting for the presidential signature (MDT, 2014).
[7] Value computed by the author based on the prediction of 70,311 traffic fatalities in 2020, which was computed based on the annual traffic fatalities growth rate in the period 2010-2012.

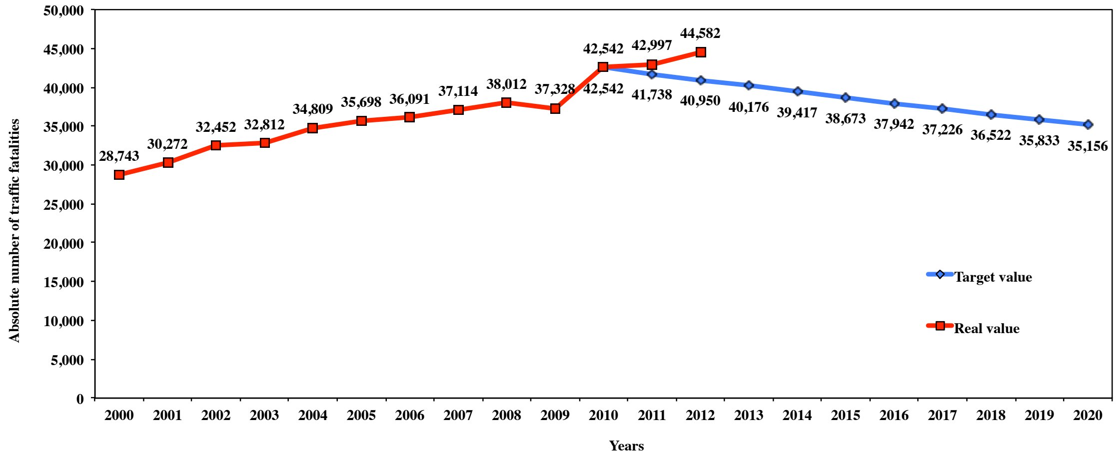



**Figure 1.3 –** Traffic fatalities in Brazil – targets and real numbers

### 1.2.2 BRAZIL AND ITS STATES

The vast Brazilian dimension and its regional contrasts in socioeconomic terms require the disaggregation of traffic fatality related rates according to each state, as presented in Figure 1.4.

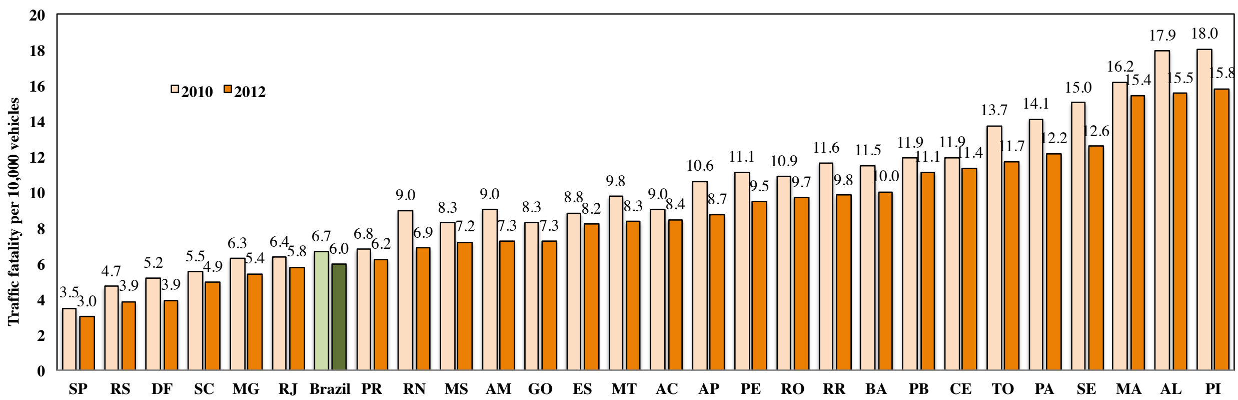



**Figure 1.4** – Fatality rate (traffic fatalities per 10,000 vehicles) in Brazil and its states

The national average is 5.96 traffic fatalities per 10,000 vehicles in 2012. The states with the most favorable performances considering the fatality rate are São Paulo – SP (3.01), Rio Grande do Sul – RS (3.87) and the Federal District – DF (3.93). Instead, in the group of states with the most unfavorable performances regarding the fatality rate there is Maranhão – MA

(15.43), Alagoas – AL (15.53) and Piauí (15.81), all sustaining very similar undesirable levels. In a comparison between the best (SP) and the worst (PI) state with respect to the number of fatalities per 10,000 vehicles the difference is more than 400%, or in other words, the fatality rate in the last state of the rank is almost 5 times the one of the first state of the rank.

Every state managed to reduce its fatality rate from 2010 to 2012, obviously reflecting the decrease in the national rate. It is important to stress, however, that the reduction of the fatality rate along the years is an expected tendency in places facing a booming motorization process. In order to obtain, apart from the reduction in this rate, also a decrease in the absolute number of traffic fatalities the fatality rate should be decreased to an even large extent.

A decline in the absolute number of fatalities happened in only 1/3 of the Brazilian states in the period 2010-2012. Figure 1.5 exhibits the average growth/reduction in the absolute traffic fatality number for each state when comparing between two periods: 2000-2010 and 2010-2012, before respectively after the Decade of Action for Road Safety proclaimed by the United Nations.

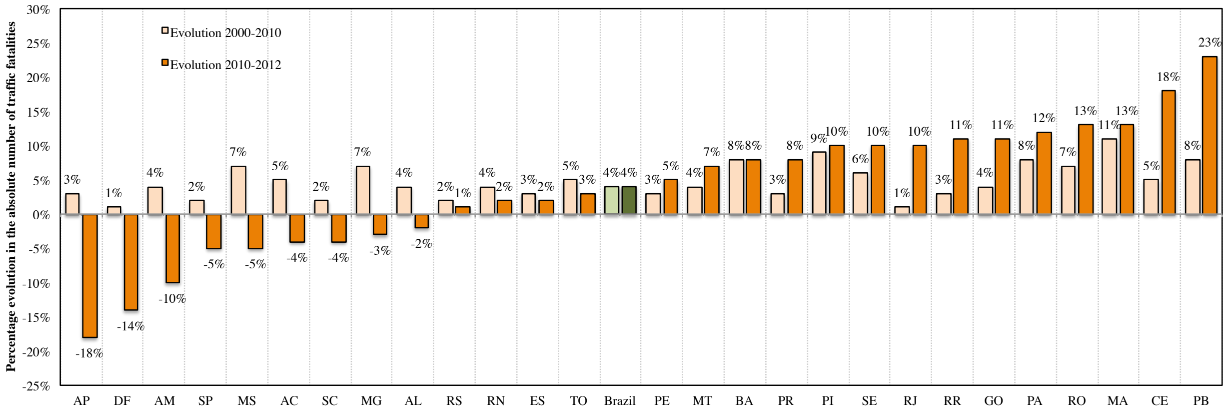



**Figure 1.5** – Evolution in the absolute number of traffic fatalities in Brazil comparing the years 2000-2010 and 2010-2012

While Brazil presented a 4% increase per year in the absolute number of traffic fatalities between 2010 and 2012, some states reached substantial reductions, e.g. Amapá (- 18% p.a.), Distrito Federal (-14% p.a.) and Amazonas (-10% p.a.). Contrarily, Paraíba (+23% p.a.) and

Ceará (+18% p.a.) sustained the highest annual growth rates in the absolute number of traffic fatalities during the most recent years.

### 1.2.3 CHARACTERISTICS OF TRAFFIC FATALITIES

Another strategic point of view to analyze the (national) trend in the traffic fatalities is by considering different user types, disaggregated according to the transportation mode. Since exposure plays an important role in the distribution of traffic fatalities between different transportation modes, the analysis in Table 1.1 is limited to show the percentage differences between the number of traffic fatalities by user type in the period 2010 to 2012. We notice a reduction in the number of pedestrian fatalities (-11%) and cyclists (-1%, less significant), while the fatalities in motor vehicles substantially increased, as shown by the bus occupant fatalities, which increased by 21% in the period, followed by motorcycle occupants (+15%).

**Table 1.1 –** Evolution of traffic fatalities by user type in the period 2010-2012

| Road user | Evolution 2010-2012 | Absolute number in 2012 |
|---|---|---|
| Pedestrian | -11% | 8,819 |
| Cyclist | -1% | 1,492 |
| Motorcycle occupant | 15% | 12,544 |
| Car occupant | 12% | 10,525 |
| Truck occupant | 11% | 863 |
| Bus occupant | 21% | 193 |
| All specified users | 5%[8] | 34,436 |

With respect to the traffic fatality rates by age group (considering the population corresponding to each age group interval), the only age group presenting a clear decrease in the rate in the period 2010-2012 is the age group under 15 years of age (see Table 1.2). The rate remained more of less the same for the group of 20-24 y.o. whereas there was an increase in the number of fatalities per population of that age group for the remaining age groups, mainly the 15-19 and 50-64 y.o. and age groups, with increases higher than 8%.

[8] This value differs from the 4% value we presented in Figure 1.5, because in Table 1.1 we did not consider the traffic fatalities in which the transportation mode of the victim was unspecified.

**Table 1.2 –** Evolution of traffic fatality rate by age group in the period 2010-2012

| Road user age group | Evolution 2010-2012 | Absolute number in 2012 |
|---|---|---|
| < 15 y.o. | -1.23% | 1,852 |
| 15-19 y.o. | 8.99% | 3,708 |
| 20 - 24 y.o. | -0.02% | 5,891 |
| 25 - 49 y.o. | 5.80% | 21,139 |
| 50 - 64 y.o. | 8.10% | 7,062 |
| > 64 y.o. | 1.63% | 4,626 |

Regarding the absolute number of traffic fatalities by age group and gender, the most commonly user profile as a fatal victim are men between 20 and 24 y.o. (see Figure 1.6). In addition, the concentration of traffic fatalities in the age group 20-59 y.o. is far more noticeable for males than for females, which present a more equilibrated distribution of fatalities throughout the age groups.

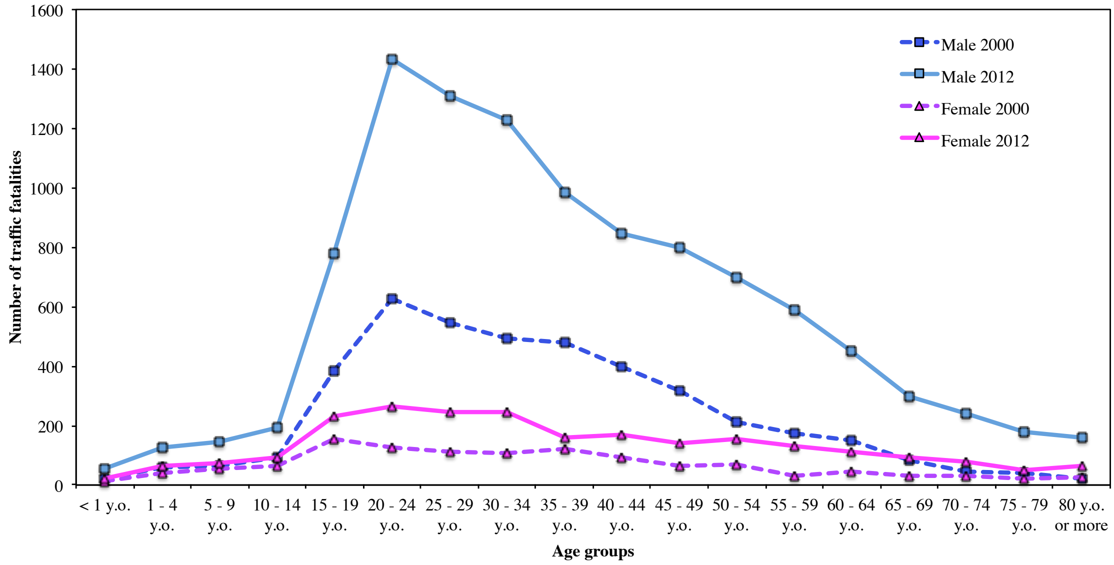



**Figure 1.6 –** Number of traffic fatalities by age group and gender in the years 2000 and 2012

## 1.3. ROAD SAFETY EVOLUTION

The relation between road safety and the level of socioeconomic development of a place is a concern in the transportation engineering field. The obstacles come up in the attempt to quantify such relationship, since it requires the adoption of some parameters intended to express both the road safety and socioeconomic development levels. For road safety, the traditional indicators of fatalities per inhabitants, per vehicles or per travelled kilometers are usually assumed. Nonetheless, whatever is the chosen socioeconomic parameter, it will only reveal one part of the story and the conclusions drawn from it have also a limited coverage taking into account the elevated complexity degree of this relationship.

The difficulty in establishing this association is in part attributed to the multidimensional nature of the road safety problem, which is influenced in different intensity levels by a great extend of aspects, like road users' education and culture, vehicle fleet and road system conditions, quality of the health system etc., which all, in the essence, are socioeconomic parameters. Additionally, it is important to consider that the elasticity of these relations is probably not constant over time, nor for diverse geographical areas (e.g. countries). So, although the relation between all sorts of parameters point in the same direction, this is, in theory, better socioeconomic indicators also imply a better road safety situation, the form of this relation is subjected to the different forms this relationship may assume given a certain combination of circumstances.

Also not helpful is the scarceness of available parameters to quantify and support these assumptions. In this sense, probably one of the most commonly available information (in national and international research), capable to play this role are the GDP *per capita* and the motorization rate. Therefore, the fundamental premise is that societies with higher GDP *per capita* and/or motorization rate are also more prone to offer better education, safer vehicles and roads and a higher health care quality, consequently decreasing both the risk of an accident and injury/fatality to occur. Based on this contextualization, in the following paragraphs such relationships are explored and discussed.

We constructed time series based on information on four basic variables (population, motor vehicle fleet, traffic fatalities and GDP) and their respective association into the following rates:

- Mortality rate – fatalities per 100 thousand inhabitants;
- Fatality rate – fatalities per 10 thousand motor vehicles;
- Motorization rate – motor vehicles per 100 inhabitants;
- GDP *per capita* – GDP per inhabitant.

We collected these parameters for five countries with available data for long period in time:

- Brazil (available time series from 1970 to 2012);

- United States (available time series from 1900 to 2012) – selected because of its great data availability and also because of the similarities with respect to the roadway development model, as the one being reproduced in Brazil over the last decades;
- United Kingdom (available time series from 1970 to 2012) – selected for being one of the best-performing countries in road safety in the European Union (forming the SUN group, together with Sweden and the Netherlands);
- Belgium (available time series from 1970 to 2012) – selected for being an average performing country with respect to road safety in Europe;
- Poland (available time series from 1970 to 2012) – selected for being a low performing country regarding road safety in the European Union and also because of its recent and fast motorization increase process (rather comparable to the Brazilian process).

In Figure 1.7 the evolution of these indicators over time is graphically represented and the data sources are provided in Table 1.3.

**Table 1.3 –** Data sources for Figure 1.7

| Country | Population | Vehicle fleet | GDP | Fatalities |
|---|---|---|---|---|
| USA | Census Bureau (2012) | FHWA (1997)<br>FHWA (2011) | OECD (2013) | FHWA (1997)<br>Census Bureau (2011) |
| UK<br>Belgium<br>Poland | OECD (2013) | European<br>Commision (2013) | | OECD (2013) |
| Brazil | | DENATRAN (1998)<br>DENATRAN (2013) | *Banco Central do Brasil* (2013) | *Ministério da Saúde* (2014)<br>ABRASPE (1983) |

From the graphs analysis some comments are derived as follows regarding the general evolution of the indicators:

- All selected countries are managing to reduce their absolute number of fatalities since the last two decades, except Brazil;
- The population growth is more pronounced in Brazil and USA;
- The motorization is increasing in all countries, although this process is faster in Brazil and Poland, the countries with the most recent development process;
- The GDP *per capita* is increasing in all selected countries;

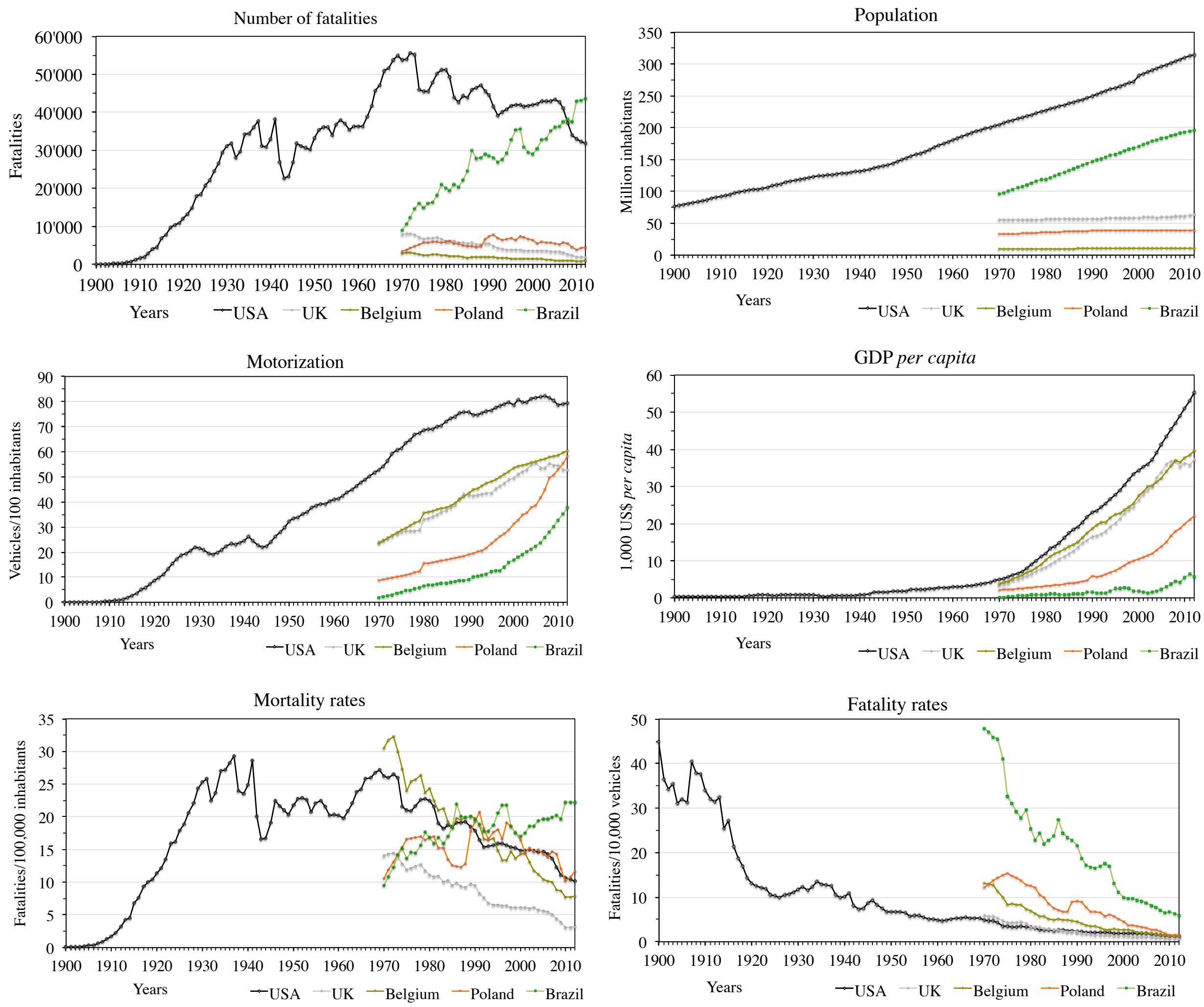



**Figure 1.7** – Evolution of indicators for chosen countries

- All selected countries are managing to reduce their mortality rates since the last two decades, except Brazil;
- All selected countries are managing to reduce their fatality rates since the last three decades;
- There is an apparent link between the motorization growth and the GDP *per capita* evolution over time, indicating a high degree of correlation between these two parameters (see Figure 1.7) – for this reason, the subsequential analyses in this section are performed considering exclusively the motorization rate as the socioeconomic parameter.

Through the observation of the evolution of these parameters over time in the USA, which presents the longest time series on the theme, it can be seen that the country was characterized by a period in which the number of fatalities raised from 1900 to 1972, with a peak of 55,600 fatalities. In the same period, the fatality rates presented a constant decreasing trend, accompanied by the fast increase in motorization. However, regarding the mortality rate, until 1937 it presented an increasing pattern, when, after reaching the peak of almost 30 fatalities/100,000 inhabitants, a slight decreasing process began and until 1969 it was oscilating over an average about 22 fatalities/100,000 inhabitants. From this year ahead, the mortality rate showed a declining tendency, reaching about 10 fatalities/100,000 inhabitants in 2012 (a value 60% smaller than the one for 1969). Thus, from the North American evolution, it might be deduced that:

- The mortality rate behavior is more sensible to changes in the absolute number of fatalities (rate's numerator), since it generally presents higher variations than the population;
- The fatality rate behavior is more sensible to changes in the vehicle fleet (rate's denominator), since it generally presents higher variations than the number of fatalities;

For the selected European countries, although the availability of data is from 1970 onwards, 1972 is also the year with the peak number of fatalities in UK and Belgium, being respectivelly equal to 8,135 and 3,130 fatalities. Since this record year, both countries are managing to reduce both the absolute number of fatalities and the mortality rate. The declining trend in fatality rates is verified since 1970 and probably has already started much earlier, as demosntrated for the USA. In Poland, there was a first 6,107 fatalities peak about 10 years later, in 1981, and then, with the harsh economic model changes in Eastern Europe, there was a second peak of 7,901 fatalities in 1991. A similar behavior is detected for Polish mortality rates over this period, while the fatality rates maintained a general reduction tendency (although passing through a considerable increase in 1991).

According to Brazilian data, the South American country seems to be passing through a stage that USA, UK and Belgium have experienced before 1970, with an increasing and almost unbroken trend in the absolute number of fatalities, accompanied by a slighter but still rising trend in the mortality rate – although the fatality rate is continuously decreasing. The observation of the quite similar North American and European experiences may provide some suggestions of what is going to happen in Brazil in future years.

The mentioned indicators evolution experimented by USA, UK, Belgium and more recently by Poland show that at earlier motorization stages, the prioritary interest was the economic development and the road safety issue was not so evident yet. In this type of policy, there is strong encouragement for people purchasing individual vehicles and the mobility tremendously increases; as a consequence, of course, the population is more exposed to the risk of accidents. Year after year, the negative impacts of this "boom" in mobility are evidenced by the growing number of traffic accidents and fatalities.

Besides the emotional pain and suffer caused by accidents, the increasing economic costs produced by this phenomenon start to take an important part of the public budget until the moment that the impact of accidents and fatalities become unacceptable. By this time, reactive measures towards road safety are gradually taken and their positive impact is expressed in the reduction of accident/fatality risk in the upcoming years. This process is described by Jørgensen (1996) through the characterization of three main stages in motorization development illustrated by the "S-shaped" curve in Figure 1.8.

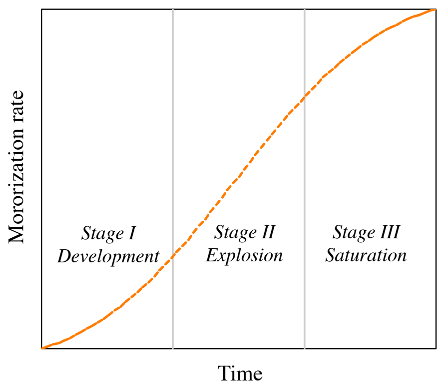



**Figure 1.8** – Stages of motorization development

Based on Jørgensen (1996)

- Stage I – Development (in general until 30 vehicles/100 inhabitants) – motorization growth is slower and since the economic growth is the priority, road safety is still a small concern;
- Stage II – Explosion (in general from 30 up to 50 vehicles/100 inhabitants) – motorization growth is accelerated and the road safety concern is emergent, since the impact of traffic accidents becomes more evident;

- Stage III – Saturation (in general from 50 vehicles/100 inhabitants) – motorization growth is slowed down again and road safety is the main concern.

Due to the great data availability in the USA, we can observe all these three stages from 1900, when motorization began in the country; up to 2012, when a certain degree of stabilization is found in the motorization evolution; what might be understood as a saturation stage. The association of these motorization stages with the road safety concern is also verified in the North American data through the mortality rate evolution, which in the beginning of the 20th century was fastly increasing, then from the 40's it became more stable, and, finally from the 70's a more clear decay process initiated. Therefore, the USA data provide a quite complete picture on how motorization and road safety are linked thoughout different development phases of a nation. In other words, the mentioned motorization stages express the changes over time in road safety policies and even in society's priority on managing the tenuous tradeoff between mobility and safety.

Despite the suggested intervals in motorization rates between the different stages, this boundary is obviously not fixed and its interpretation is reserved to overall character analysis. The transition from a road safety concern into real action and progress towards road safety (with the reduction of the absolute number of fatalities and its associated per inhabitant and per vehicle rates) may take some considerable time to materialize. It implies the possibility that two countries with similar motorization levels actually exihibt notably different road safety situations, since one has already left the "explosion stage" and reached the "saturation stage", while the other, regardless the similar motorization level, is still in the "explosion stage" (and the some logic applies to the transition between the "development" and the "explosion stages").

One example in this sense is observed by means of the comparison between the motorization rates and the mortality rates of the United Kingdom, Belgium and Poland: regarding the motorization level the order is, from highest to lowest, Belgium (60 vehicles/100 inhabitant), Poland (58 vehicles/100 inhabitant) and UK (52 vehicles/100 inhabitant); while regarding the fatality rate, from lower to higher, the United Kingdom (0.60 fatalities/10,000 vehicle), Belgium (1.30 fatalities/10,000 vehicle) and Poland (2.00 fatalities/10,000 vehicle). This alternance on the rank shows that Poland, despite its relatively high motorization level, still holds inferior standards on road safety.

Such differences between the motorization rank and the fatalities per vehicle rate might be attributed to the presence of other factors influencing the elasticity between the income level and the demand for individual vehicles, e.g. the offer and quality of a transit systems, as stated by Train (1986). In other words, there is a tendency that in really developed regions there exist transportation planning policies that offer good-quality public transportation, reducing the population's need of purchasing a car. It implies that the motorization will probably go up to a certain level, from which, there is a stabilization or even a slight decrease in motorization over time (similarly to what is verified in UK since 2008 – see Figure 1.7).

Even though the existence of this confounding behavior into the same motorization stage, the motorization level is still useful for the inference of the road safety level in comparisons in which there are more remarkable differences among the member countries (large enough differences to put them in distinct stages). In addition, the absence of a single ideal fitting socioeconomic parameter in the study of its relationship with road safety forces specialists to deal with a certain level of generalization in this kind of study.

The observation of the fatality numbers in the selected European countries and mainly in the USA, which exhibits the most extensive data series, clearly shows that at a certain point, the absolute fatality number starts to decrease. Since in Brazil the peak in fatalities is still unknown, due to the continuous increasing trend, the same inflection point identification is not possible. However, alternatively, such prediction might be done assuming that the past indicators evolution in the USA may also occur in Brasil. Supposing that this is the case, there are two ways on indirectely predicting when fatalities may start to decrease in Brazil:

- When Brazil reaches the same motorization level as the USA in 1972;
- When Brazil reaches the same fatality rate as the USA in 1972.

Figures 1.9 and 1.10 illustrate the predictions, in which the annual growth in the period from 1999 to 2012 was extrapolated for the upcoming years, creating a prediction until 2020. This prediction indicates that, based on the North American road safety evolution, Brazil will only manage to reduce its yearly number of traffic fatalities by 2017, according to the "fatality rate prediction" – when a peak around 51,000 fatalities may be reached; or in 2018, according to the "motorization rate prediction" – when a peak around 52,000 fatalities may be reached.

Whatever the indicator adopted on for the prediction, a substantial average delay of 45.5 years is detected in the Brazilian road safety evolution in comparison to the USA.

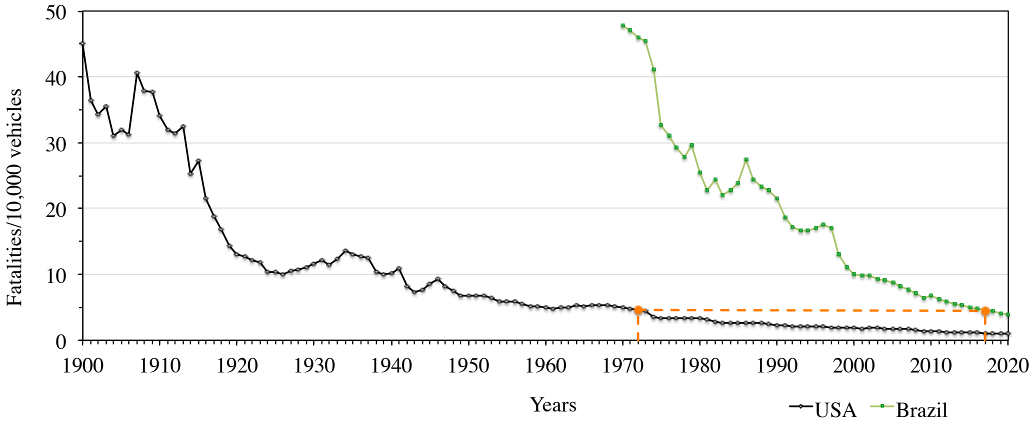



**Figure 1.9** – Prediction of fatalities inflection point for Brazil according to the fatality rate evolution

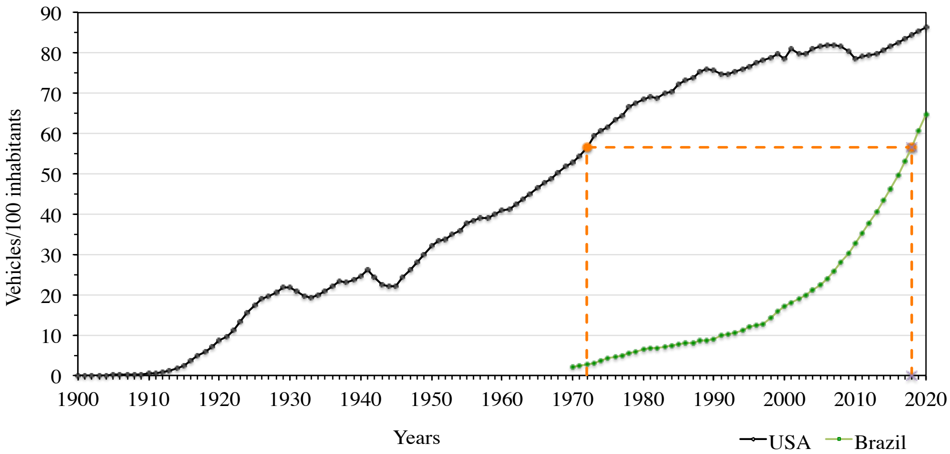



**Figure 1.10** – Prediction of fatalities inflection point for Brazil according to the motorization evolution

Indeed the background in the USA in 1972 is substantially different from the current one in Brazil regarding many aspects, such as:

- available technologies for both crash and injury prevention – mainly in the vehicle industry and roadway design;
- media means diffusion – which plays a role in informing the society about the problem dimensions, creating a "public pressure" on the theme;
- advances in medicine – enabling a better quality post-crash treatment;
- amount of knowledge produced on the theme – where Brazil could benefit from past experience of more developed countries on managing the road safety problem.

Nevertheless the existence of an apparent more favorable scenery for Brazil to tackle this problem nowadays, there are also significative throwbacks, which, in some extent, avoid the country to benefit from the available international experience in handling the road safety question. In this context, the political scene, related to an almost traditionally deficient planning strategy, seems to be an important negative factor. The motor vehicle fleet composition in Brazil also makes the national context quite dissimilar with a considerably high share of vulnerable two-wheeled vehicles. Moreover, there are also historical facts that certainly influenced the evolution of motorization in the USA, like the occurrence of two World Wars after the beginning of motorization in the country; in Brazil, for instance, although the first vehicle arrived in the country at the end of the 19$^{th}$ century, the massive motorization process actually started after World War II (Governo do Estado de São Paulo, 2013).

The application of the same kind of prediction is not possible when considering the European data, since the motorization and mortality rate levels of UK and Belgium (and actually for the entire set of EU countries) in 1972 have already been achieved in Brazil. In some extent, this might be interpreted that the road safety development models in Brazil and the USA are more similar if compared to the European ones.

## 1.4 MANUSCRIPT STRUCTURE

Having provided insight in the road safety problem in Brazil, from an international and national perspective as well as a time perspective, this introductory chapter closes with a section on the manuscript's structure. Figure 1.11 provides this strucure. In addition, in Appendix A, there are state profiles offering a useful summarized visualization of relevant road safety knowledge built up during this thesis.

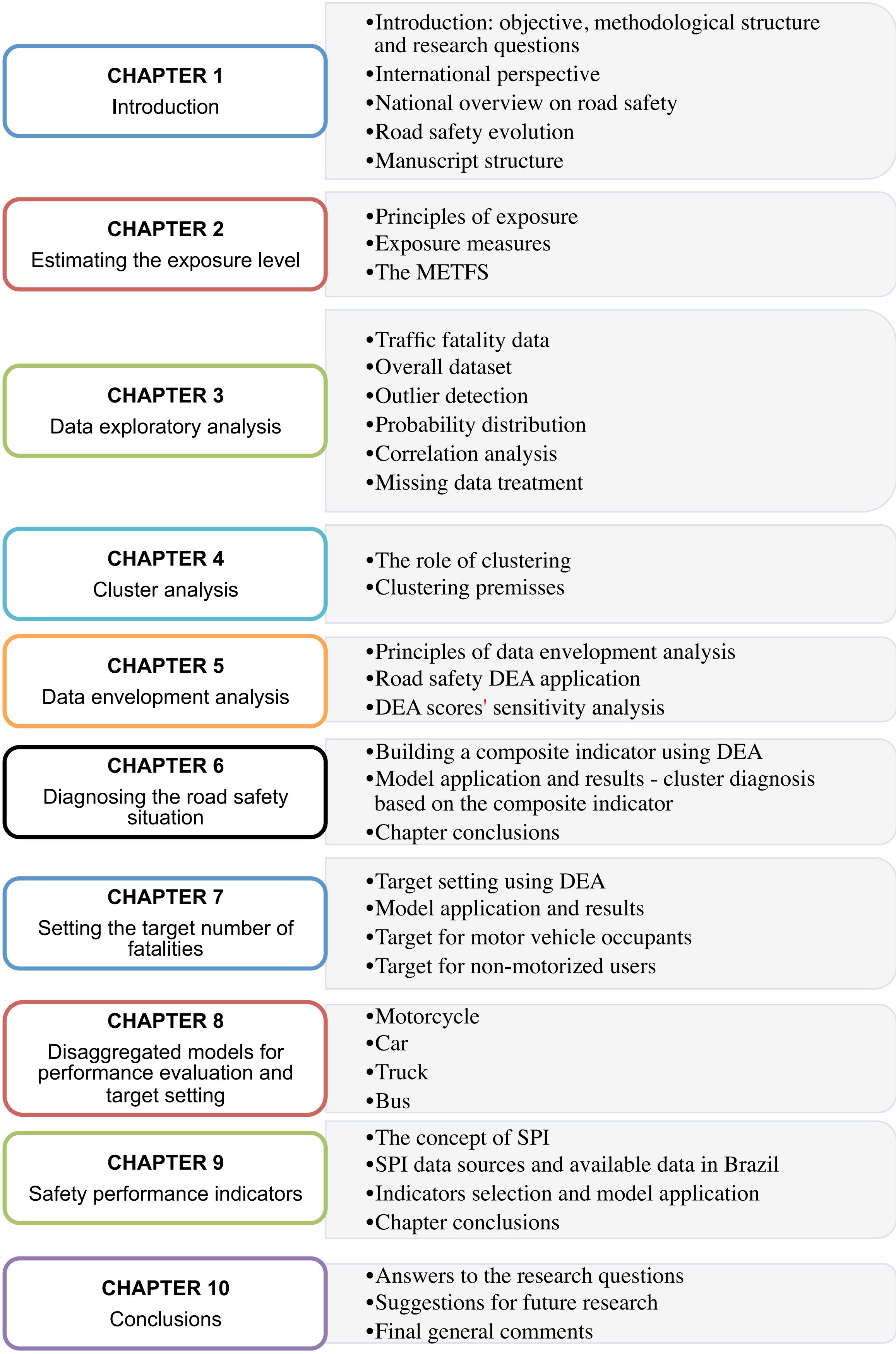



**Figure 1.11** – Manuscript structure

# 2. ESTIMATING THE EXPOSURE LEVEL

Exposure is the fundamental element for traffic accidents to occur; thus, its adequate estimate is as important as the availability of reliable traffic fatality information. For this reason, this chapter is dedicated to discuss the principles of exposure within the road safety framework and exposure data in its first section. In the second section, we describe the Method for Estimating the Traveled Distance using Fuel Sales (METDFS), which results are extensively applied in this research.

## 2.1. PRINCIPLES OF EXPOSURE

There is no single definition for exposure that is applied in every area, since it depends on the nature of the theme under investigation (Evans, 2004). In transportation sciences, exposure is the amount of transport in a given transportation system (Thomas, 2005; Hermans, Wets & Van den Bossche, 2006). A more specific definition from the road safety perspective, according to Elvik *et al.* (2009), is as follows: exposure expresses the amount of activity in which there is the possibility of an accident to occur.

Although there is clear a relationship between exposure and accidents - that is, the second follows from the existence of the first - this relationship is complex and depends on other aspects incorporated or inherent to this exposure and the influence on the amount of accidents occurs indirectly (by increasing or decreasing the risk of accidents). There are several examples of these aspects, such as: the vehicle type, the age of the driver as well as the gender and the personality, the environment, etc. (Evans, 2004). In more general words, in addition to the exposure, accidents occurrence depends on the pair legislation/enforcement and on a large range of factors associated to physical components of the traffic system: the human behavior, the vehicle fleet characteristics and the environment/road (e.g. Ferraz, Raia Júnior & Bezerra, 2008).

In a study about detecting risk mechanisms that, in addition to exposure, affect the probability of accidents occurrence, Elvik (2006) proposed four “laws” of accident causation:

- The universal law of learning – the ability of detecting and controlling hazards increases as the amount of travel increases;
- The law of the unpredictable – the rarer an event is, the higher is its effect on the accident rate;
- The law of complexity – the larger the number of information the road user has to process by unit of time, the larger the probability of committing a mistake;
- The law of cognitive capacity – the closer the driver is from his own cognitive capacity, the higher the risk of accidents is.

## 2.2. EXPOSURE MEASURES

For practical purposes of national or subnational road safety evaluations based on the comparison of different geographical units (i.e. countries, states or cities) or on the comparison of the evolution of a specific place over time, the most commonly used indicators are formed by a risk ratio, i.e. the ratio between a certain type of road safety outcome and an exposure measure. Here, the outcome is considered as the undesirable consequence of the phenomenon under investigation – the traffic fatality.

Since traffic fatality statistics present higher reliability levels compared to e.g. traffic injury or accident data (Connor et al., 2007), it is selected as the outcome measure on which the analyses will be based in this thesis. As the ratio denominator, four exposure measures are most frequently considered (Elvik & Vaa, 2004; Elvik et al., 2009; Jørgensen, 1996):

- Number of inhabitants (I) – it assumes that the whole population is exposed to the risk of being killed in a traffic accident;
- Number of motor vehicles (II) – it assumes that the risk of a traffic fatality is related to the registered number of motor vehicles;
- Number of vehicle kilometers traveled (III) – the same assumption as the previous indicator, however now considering the distance traveled by each vehicle, that is, the amount of traffic (given in vehicle x km);

- Number of passenger kilometers traveled (IV) – the same assumption as the previous indicator, however now considering the vehicle occupancy ratio (given in passenger x km).

In summary, each listed exposure measure consists of a complementation to the previous one, providing some extra information and leading to a more precise exposure definition (Evans, 2004). The choice of which indicator to use depends on the main purpose of the evaluation and on the availability of data. The indicator formed by exposure measure (I) is used to measure the negative impact of traffic accidents on society, enabling the comparison of traffic accidents with other causes of death, such as diseases and homicides (OECD, 2011a). Some authors even named this indicator as "health risk" (Elvik & Vaa, 2004; Elvik et al., 2009; Jørgensen, 1996), probably due to its relevance in public health policy planning; although here it is going to be referred to as "mortality rate" (MR), as e.g. is the case by the Institute for Road Safety Research (2013).

In quite general terms, indicators formed by exposure measures (II), (III) and (IV) provide the same information in three different levels of refinement regarding the risk exposure parameter definition; the last one being the most precise on quantifying the amount of exposure, but also the most rarely available. For a more detailed investigation, they can all be disaggregated to produce risk information for different transportation modes. This second group of indicators is referred to in literature as "traffic risk" (Elvik & Vaa, 2004; Elvik et al., 2009; Jørgensen, 2006), or "fatality rate" (FR), a denomination adopted in this study and also in e.g. Hermans (2009).

Whichever traffic fatality risk indicator is adopted, it is important to interpret it under the delimitated perspective provided by the information in the denominator; yet in practice they are all used to carry out road safety evaluations without proper appraisal regarding this topic, especially in developing countries. Also in developing countries, the availability of exposure measures (III) and (IV) is scarce.

In some high-income countries, the vehicle kilometers traveled – VKT are available with regards to the road-motorized vehicles. Two forms of traffic-related information produce VKT measurements: odometer reading (in periodic vehicle inspections) or using data from a network of vehicle counting stations. Two other non-traffic related sources might also provide the VKT in two cases: in travel surveys or in estimates using automotive fuel sales data (Bureau of Infrastructure, Transport and Regional Economics, 2011; Leduc, 2008).

In developing nations like Brazil, where both traffic-related and the first non-traffic related mechanisms do not exist or are poorly available (*Ministério da Ciência e Tecnologia*, 2006), the usage of fuel sales data is currently the only remaining alternative that enables the attempt of estimating the VKT.

## 2.3. THE METDFS

The circumstances mentioned in the end of the previous section motivated the development of a model for estimating the VKT by road-motorized vehicles in Brazil based on fuel sales and vehicle fleet characterization, which we describe in this section and is referred to as METDFS (Method for Estimating the Traveled Distance using Fuel Sales). Its first conception was published in 2010 in the paper in Portuguese "*Uma estimativa da mortalidade no trânsito considerando exposição ao risco no Brasil*" from Bastos *et al*. (2010), after which the methodology has been improved until its current configuration. More information about previous works on the method can be found in: Bastos (2011), Bastos *et al*. (2011), Bastos *et al*. (2012) & Ferraz *et al*. (2012).

It is important to state that the fuel-sale-based estimate of the VKT (referred here as FSBE-VKT) provides an aggregate picture in terms of spatial distribution of the estimated VKT; being based exclusively on fuel sales volumes, it is not possible to define the traveled distance environment, for example, an urban or rural environment. It is also difficult to assure inside which municipality the consumption of the sold fuel occurred (transformation of the fuel into VKT), since there exist inter-municipal trips.

Hence, the consumption of the fuel sold in a municipality is not mandatory to happen inside its boundaries (although this is a plausible hypothesis). Therefore, the reliability of the FSBE-VKT application tends to increase as the spatial boundaries of the analysis decreases. It limits the application of the FSBE-VKT to the analysis of relevant road transportation issues at broader levels, e.g. at the national level. In some specific cases, such as in vast countries like Brazil, this aggregation degree might be transferred to a state level (since most of Brazilian states are comparable to countries in terms of territory dimensions and population).

Analogously to the idea exposed in the previous paragraph, it is important to know whether the consumption of the fuel sold in a certain country actually is used for traveling within that nation's borders. In some European countries, as Belgium and the Netherlands for example, uncertainties due to the highly integrated roadway network connecting the neighboring countries might significantly influence the accuracy of the FSBE-VKT. On the other hand, in Portugal, for example, which has a concentration of population and activities near the coastline (OECD, 2011b) and one single border with Spain, this kind of estimate is more suitable. The Portuguese experience in the FSBE-VKT consists of using its own data on vehicle fleet and fuel sales and mathematical models fitted to existing data from other countries on fuel consumption, vehicle fleet and traffic volume (Cardoso, 2005).

However, the ideal scenery regarding this issue occurs in Australia (since this country is an island), where there is no interference of other countries on the FSBE-VKT. The Australian approach estimates the VKT quarterly by vehicle and fuel type from the state and territory fuel sales data for each of the eight states and territories in the country (Bureau of Infrastructure, Transport and Regional Economics, 2011; Hossain and Gargett, 2011).

Despite the fact that Brazil presents an extensive international borderline with other Latin-American countries, there is a strong concentration of population and economic activities near the coastline (Baer, 2008; *Instituto Brasileiro de Geografia e Estatística*, 2000; Yunes, 1972). As a consequence, there is also a heavy concentration of the urbanization and roadway connectivity alongside the same area. It constitutes a beneficial characteristic that enhances the trustworthiness of the FSBE-VKT.

Although there might exist trips using international fuel in Brazil and vice-versa, it is probably not significant in comparison to the national values. Regarding the "per state" disaggregation, the same assertion becomes less convincing. However, a compensatory effect is expected, since there are no apparent reasons (e.g. substantially different fuel prices between neighboring states) for not prevailing an overall tendency of equilibrium; this is, the VKT in a state A using the fuel sold in a state B is expected to be equivalent to the VKT in state B using fuel sold in state A. In conclusion, this suspicion is an inherent aspect of the FSBE-VKT and transportation analysts must take it into account instead of prohibiting its application.

The VKT are a fundamental element in the causation chain of an accident involving a road-motorized vehicle; it means that without exposure (or zero VKT) there is no risk of this type of accident to occur. Therefore, the VKT are basic “exposure to the risk” measures in road safety research, and its adequate estimate is crucial for assessing and comparing the risk level between different places. Based on that, the METDFS is an indirect method to measure the mobility level of motorized vehicles (expressed by the VKT), since it depends on auxiliary information, such as fuel consumption rates and vehicle fleet characterization, for instance. The next sections contain a description of the METDFS’ steps, as well as the explanations about some methodological details.

### 2.3.1 STRUCTURE AND CALCULATION OF THE METDFS

The method elaborated for the FSBE-VKT consists of the development of a model with official input information about the vehicle fleet characterization and the automotive fuel sales in the period 2010-2012 (the latest year for which all the needed information is available). Figure 2.1 contains a scheme of the model steps.

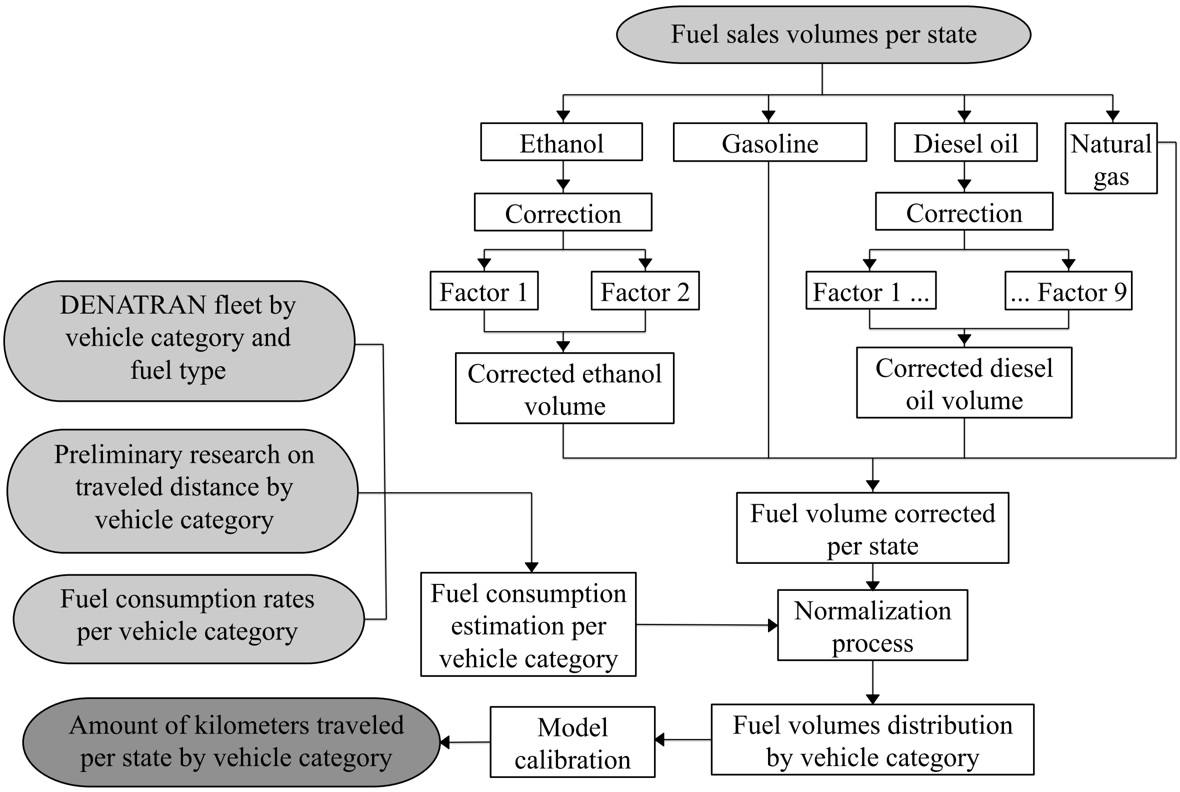



**Figure 2.1** – METDFS’ information scheme

Vehicle fleet characterization includes data about vehicle categories[1] (motorcycle, car, truck and bus) and fuel type used by each category according to the National Department for Traffic (DENATRAN, 2013), as well as different fuel consumption rates given in km/l according to the National Institute of Metrology, Standardization, and Industrial Quality (INMETRO, 2010). The automotive fuel sales include four different fuel types used in Brazil: ethanol, diesel oil, gasoline and natural gas collected from the National Agency of Petroleum, Natural Gas and Biofuels (ANP, 2012).

Previously to the VKT calculation, we correct the ethanol and diesel oil volumes due to the fact that these fuels are also used for other activities than roadway transportation (Ministério de Minas e Energia, 2012) (see Section 1.1.2). The next step consists of the attribution of each fuel type volume used to each vehicle category (i.e. the volume of gasoline used by motorcycles or cars, the volume of diesel oil used by buses or trucks, etc.).

In order to increase the accuracy of the estimates, we divide each vehicle category into a number of subcategories manifesting similar fuel consumption rates, as follows: motorcycles – according to their engine capacity into ≤ 115 CC, >115 CC and ≤ 250 CC, > 250 CC and ≤ 500 CC, and ≥ 500 CC; cars – according to thirty-two subcategories, such as sedan, hatchback, pick-up, station wagon, utility, minivan, etc.; trucks – according to their mass into light, medium and heavy; and buses – according to their size / type of use into microbus, urban bus and coach bus.

Equation 2.1 represents the computation of the fuel volumes consumed by each subcategory (in which the sum corresponds to the fuel volume consumed by each respective category):

$$V_{j_W klmn} = PVKT_k . F_{j_W klmn} / C_{j_W klmn} \tag{2.1}$$

In which: $V_{j_W klmn}$ – Estimate of the consumed fuel volume by vehicle subcategory "$j_w$", of

[1] Vehicle fleet information is presented by DENATRAN according to 21 classes: car, pick-up, station-wagon, commercial vehicle, truck, tractor truck, bus, minibus, motorcycle, moped, motor pedal cycle, three-wheeled motor vehicle, ATV cycle, tram, frame-platform, trailer, semi-trailer, side-car, bulldozer, wheel tractor and other. Some vehicle classes must be ignored in order to maintain the coherence between the effect (number of fatalities) and the cause (measure of exposure considered, in this case, the distance traveled). Such classes should be excluded because they are not capable to travel by itself (in other words, they need to be attached to some motorized unit); or because they are not considered as road transport vehicles: tram, frame-platform, trailer, semi-trailer, side-car, bulldozer, wheel tractor and other (non specified classes).

vehicle category "*k*", using fuel "*l*", in year "*m*", in state "*n*" [l]; $PVKT_k$ – Previous data[2] on annual VKT data for the vehicle class "*k*" [km]; $F_{j_w klmn}$ – Fleet of vehicle subcategory "$j_w$", from vehicle category "*k*", using fuel "*l*", in year "*m*", in state "*n*"; $C_{j_w klmn}$ – Fuel consumption rate of vehicle subcategory "$j_w$", of vehicle category "*k*", using fuel "*l*", in year "*m*" [km/l]; *w* – Number of subcategories.

Then, we repeat the process of Equation 2.1 for each vehicle category using the same fuel type. Equation 2.2 provides the estimate of total consumed volume for each fuel type:

$$TV_{lmn} = \sum_{i=1}^{a} V_{j_w klmn} \tag{2.2}$$

In which: $TV_{lmn}$ – Estimate of the total consumed fuel volume by vehicles using fuel *"l"*, in year *"m"*, in state *"n"* [l]; $V_{j_w klmn}$ – Estimate of the consumed fuel volume by vehicle subcategory *"$j_w$"*, of vehicle category *"k"*, using fuel *"l"*, in year *"m"*, in state *"n"* [l]; *a* – Number of subcategories of vehicles using the same fuel type, $a \leq 4$ (depending on the vehicle category).

Equation 2.3 describes the normalization process to compute the share of each fuel type attributed to each vehicle category:

$$S_{j_w klmn} = V_{j_w klmn} / TV_{lmn} \tag{2.3}$$

In which: $S_{j_w klmn}$ – Share of the fuel type *"l"* consumed by vehicle subcategory *"j"*, of vehicle category *"k"*, in year *"m"*, in state *"n"* [l]; $V_{j_w klmn}$ – Estimate of the consumed fuel volume by vehicle subcategory *"j"*, of vehicle category *"k"*, using fuel *"l"*, in

[2] Based on Borba (2008).

year *"m"*, in state *"n"* [l]; $TV_{lmn}$ – Estimate of the total consumed fuel volume by vehicles using fuel *"l"*, in year "m", in state *"n"* [l].

Equation 2.4 provides the VKT of all vehicles using the same fuel type in each state:

$$VKT_{klmn} = \sum_{i=1}^{w} F_{j_W klmn} \cdot C_{j_W klmn} \cdot S_{j_W klmn} V_{lmn} \tag{2.4}$$

In which: – Vehicle-kilometers traveled by vehicle category "k", using fuel "l", in year "m", in state $VKT_{klmn}$ "n" [km]; $F_{j_W klmn}$ – Fleet of vehicle subcategory *"$j_w$"*, from vehicle category *"k"*, using fuel *"l"*, in year *"m"*, in state *"n"*; $C_{j_W klmn}$ – Fuel consumption rate of vehicle subcategory *"$j_w$"*, of vehicle category *"k"*, using fuel *"l"*, in year *"m"* [km/l]; $S_{j_W klmn}$ – Share used by vehicle subcategory *"$j_w$"*, of vehicle category *"k"*, of fuel *"l"*, in year *"m"*, in state *"n"* [%]; $V_{lmn}$ – Fuel sales volume *"l"*, in year *"m"*, in state *"n"* [l]; *w* – Number of vehicle subcategories, $w \leq 32$ (depending on the vehicle category).

Equation 2.5 enables the calculation of the VKT by each vehicle category using all fuel types for each state:

$$VKT_{kmn} = \sum_{i=1}^{l} VKT_{klmn} \tag{2.5}$$

In which: $VKT_{kmn}$ – VKT by the vehicle category *"k"*, in year *"m"*, in state *"n"* [km]; $VKT_{klmn}$ – VKT by vehicle category *"k"*, using fuel *"l"*, in year *"m"*, in state *"n"* [km]; *l* – Number of fuel types used by each category, $l \leq 4$ (depending on the vehicle category).

### 2.3.2 METHODOLOGICAL DETAILS OF THE METDFS

The application of this kind of model requires the handling of certain specificities regarding the fuel market in Brazil. The first one relates to the usage of ethanol and diesel oil for other activities except road transportation, such as non-energetic consumption, railway and/or waterway transportation, as well as in agriculture machinery. The information regarding the non-automotive use of fuels is available in the yearly Brazilian energy balance (BEN – *Balanço Energético Nacional*) and we subtracted the corresponding values from the total fuel sales volume. This step corresponds to the computation of the correction factors mentioned in Figure 2.1.

Another aspect that required some special procedure is the popularization of bi-fuel (or flex-fuel) vehicles since 2003; this is, vehicles that can use both gasoline and/or ethanol as fuel. In order to adequately consider this effect, we divided the flex-fuel vehicle fleet into two shares: one supposed to exclusively use ethanol and the other supposed to exclusively use gasoline. The calculation of each share was based on the price of each fuel type in each state. For example, the ethanol tends to present a more attractive price in the producing states, inducing a higher share of the flex-fuel vehicles to use ethanol. On the other hand, the ethanol price tends to be high in non-producing states and states that are geographically distant from the production zones, inhibiting its selection; in these cases, flex-fuel vehicles are more prone to use gasoline.

The drivers use a widespread rule in the country to decide which fuel type to use: if the ethanol price corresponds to more than 70% of the gasoline price, they choose for gasoline (Losekann & Vilela, 2010a). This value is associated to the different energetic efficiencies of both fuels (in other words, with the same fuel volume, it is possible to travel a longer distance with gasoline than with ethanol). The yearly BEN reports provide time series data on ethanol and gasoline prices, enabling the control and treatment of this aspect.

A third aspect considered was the adaptation of vehicles for using natural gas, an alternative fuel that was introduced in some states from 1996 onwards (Cavalcanti, 2004). Since vehicles are not originally prepared to work on this kind of fuel, the drivers who want to use this cheaper fuel type have to adapt their vehicle’s engine to operate using natural gas.

Consequently, there is the need of subtracting these converted vehicles from their original fuel type fleet and consider them as natural gas propelled vehicles.

A fourth methodological detail corresponds to the calibration process mentioned in the scheme of Figure 2.1. This procedure was necessary since it is possible that the estimated fuel volumes computed through Equation 2.1 result in much higher or lower amounts than the official *VNA* value provided by ANP, even after applying the corrections discussed above. This discrepancy is represented by the *f* factor presented in Equation 2.6.

$$f_{lmn} = TV_{lmn} / VNA_{lmn} \tag{2.6}$$

In which: $f_{lmn}$ – Adjustment factor for fuel *"l"*, in year "m", in state *"n"* [l]; $TV_{lmn}$ – Estimate of the total consumed fuel volume by vehicles using fuel *"l"*, in year "m", in state *"n"* [l]; $VNA_{lmn}$ – Official value for the total consumed fuel volume by vehicles using fuel *"l"*, in year "m", in state *"n"* [l].

We considered acceptable *f* values between 0.70 and 1.30 (i.e., an estimated value differing between – 30% and + 30% of the official value). For other *f* values out of this range, we adopted the following correction: if $0,60 \leq f < 0,70$, we assume $f = 0,70$; if $0,50 \leq f < 0,60$, we assume $f = 0,60$; if $f < 0,50$, we assume $f = 0,50$; if $1,30 < f \leq 1,40$, we assume $f = 1,30$; if $1,40 < f \leq 1,50$, we assume $f = 1,40$; if $f > 1,50$, we assume $f = 1,50$.

These correction factors had to be mainly applied in international-bordering states; e.g. in Amapá (AP), where the official fuel volume was forced to reduction probably due to the influence of the French Guiana frontier (the favorable currency in this place might motivate drivers to buy fuel in Brazil because of cheaper prices, making the official sales of AP an overestimation). The opposite situation occurred in Roraima - RR, in which the border with Venezuela probably stimulated Brazilian drivers to buy fuel in the neighboring country, known for extremely low petrol-derived fuel prices; it resulted in an underestimated fuel sales for RR.

The use of a previous estimate of the VKT computed for Brazil was an alternative for guiding the division of the fuel volumes consumed by each vehicle category. These values were not directly used because they did not present the necessary disaggregation level needed for this research. Therefore, we simply adopted them as referential knowledge for the METDFS in the step of distributing the different fuel types among the different vehicle categories.

### 2.3.3 RESULTS OF THE METDFS

Figure 2.2 presents the total values we estimated for the VKT in Brazil, as well as the "per vehicle type" values in the period 2000-2012. Considering that these estimates are until 2012 and that motorization continued to grow in Brazil in the last years, the country will probably have reached the notable value of 1 trillion VKT by now. Cars correspond to the highest share in the VKT estimates, and both cars and motorcycles presented meaningful increases in the VKT in recent years.

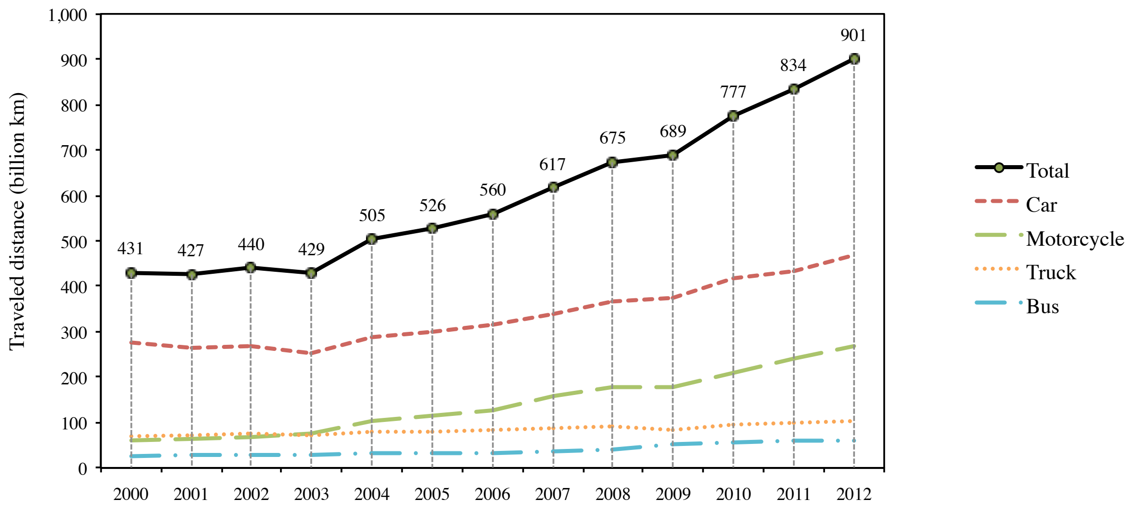



**Figure 2.2** – Yearly evolution of the VKT in Brazil by vehicle category in the period 2000-2012 according to the METDFS

Since this research focuses on the period 2010-2012, we present the "per state" values in Appendix B.

# 3. DATA EXPLORATORY ANALYSIS

Traffic fatality data is the main information used in this research, although a larger variety of data is considered. However, it is common that the available information regarding traffic fatalities is not presented in the most adequate manner to be used for road safety analysis, since its design is oriented for more general purposes. For this reason, we need to perform some initial procedures in order to create a more adequate data set ready to be used for the research purposes. This chapter also deals with an exploratory analysis of the data used in the upcoming steps of clustering the Brazilian states and in the research on outcome related indicators (3 traffic fatality related rates) as well as safety performance indicators – SPIs (beginning from an initial set of 36 indicators classified into 5 categories – road user, road infrastructure, health system, vehicle fleet and background) in the country. It includes investigations on outlier identification, probability distribution, correlation analysis and missing data treatment. The analyses performed in this chapter are helpful for a better comprehension of the dataset on which we base the application of the proposed methods.

## 3.1. TRAFFIC FATALITY DATA

The Ministry of Health publishes traffic fatality information according to the World Health Organization (WHO) recommendations, using the International Classification of Diseases in its 10$^{th}$ revision (ICD–10). This classification includes traffic fatality numbers under the denomination of fatalities caused by transportation accidents, which are divided in several subordinated groups, which can be seen in Table 3.1.

Each group contains specific categories, as Tables 3.2 to 3.10 show. The groups V90-V94, V95-V97 and V98-V99 are not detailed since are not related to our research scope on road safety.

**Table 3.1** – Transportation accident codes according to ICD–10

| Code | Specification |
|---|---|
| V01-V09 | Pedestrian injured in transport accident |
| V10-V19 | Pedal cyclist injured in transport accident |
| V20-V29 | Motorcycle rider injured in transport accident |
| V30-V39 | Occupant of three-wheeled motor vehicle injured in transport accident |
| V40-V49 | Car occupant injured in transport accident |
| V50-V59 | Occupant of pick-up truck or van injured in transport accident |
| V60-V69 | Occupant of heavy transport vehicle injured in transport accident |
| V70-V79 | Bus occupant injured in transport accident |
| V80-V89 | Other land transport accidents |
| V90-V94 | Water transport accidents |
| V95-V97 | Air and space transport accidents |
| V98-V99 | Other and unspecified transport accidents |

**Table 3.2** – Pedestrian fatalities codes

| Code | Specification |
|---|---|
| V01 | Pedestrian injured in collision with pedal cycle |
| V02 | Pedestrian injured in collision with two- or three-wheeled motor vehicle |
| V03 | Pedestrian injured in collision with car, pick-up truck or van |
| V04 | Pedestrian injured in collision with heavy transport vehicle or bus |
| V05 | Pedestrian injured in collision with railway train or railway vehicle |
| V06 | Pedestrian injured in collision with other nonmotor vehicle |
| V09 | Pedestrian injured in other and unspecified transport accidents |

**Table 3.3** – Cyclist fatalities codes

| Code | Specification |
|---|---|
| V10 | Pedal cyclist injured in collision with pedestrian or animal |
| V11 | Pedal cyclist injured in collision with other pedal cycle |
| V12 | Pedal cyclist injured in collision with two- or three-wheeled motor vehicle |
| V13 | Pedal cyclist injured in collision with car, pick-up truck or van |
| V14 | Pedal cyclist injured in collision with heavy transport vehicle or bus |
| V15 | Pedal cyclist injured in collision with railway train or railway vehicle |
| V16 | Pedal cyclist injured in collision with other nonmotor vehicle |
| V17 | Pedal cyclist injured in collision with fixed or stationary object |
| V18 | Pedal cyclist injured in noncollision transport accident |
| V19 | Pedal cyclist injured in other and unspecified transport accidents |

**Table 3.4 –** Motorcycle rider fatalities codes

| Code | Specification |
|---|---|
| V20 | Motorcycle rider injured in collision with pedestrian or animal |
| V21 | Motorcycle rider injured in collision with pedal cycle |
| V22 | Motorcycle rider injured in collision with two- or three-wheeled motor vehicle |
| V23 | Motorcycle rider injured in collision with car, pick-up truck or van |
| V24 | Motorcycle rider injured in collision with heavy transport vehicle or bus |
| V25 | Motorcycle rider injured in collision with railway train or railway vehicle |
| V26 | Motorcycle rider injured in collision with other nonmotor vehicle |
| V27 | Motorcycle rider injured in collision with fixed or stationary object |
| V28 | Motorcycle rider injured in noncollision transport accident |
| V29 | Motorcycle rider injured in other and unspecified transport accidents |

**Table 3.5 –** Occupant of three-wheeled motor vehicle fatalities codes

| Code | Specification |
|---|---|
| V30 | Occupant of three-wheeled motor vehicle injured in collision with pedestrian or animal |
| V31 | Occupant of three-wheeled motor vehicle injured in collision with pedal cycle |
| V32 | Occupant of three-wheeled motor vehicle injured in collision with two- or three-wheeled motor vehicle |
| V33 | Occupant of three-wheeled motor vehicle injured in collision with car, pick-up truck or van |
| V34 | Occupant of three-wheeled motor vehicle injured in collision with heavy transport vehicle or bus |
| V35 | Occupant of three-wheeled motor vehicle injured in collision with railway train or railway vehicle |
| V36 | Occupant of three-wheeled motor vehicle injured in collision with other nonmotor vehicle |
| V37 | Occupant of three-wheeled motor vehicle injured in collision with fixed or stationary object |
| V38 | Occupant of three-wheeled motor vehicle injured in noncollision transport accident |
| V39 | Occupant of three-wheeled motor vehicle injured in other and unspecified transport accidents |

**Table 3.6 –** Car occupant fatalities codes

| Code | Specification |
|---|---|
| V40 | Car occupant injured in collision with pedestrian or animal |
| V41 | Car occupant injured in collision with pedal cycle |
| V42 | Car occupant injured in collision with two- or three-wheeled motor vehicle |
| V43 | Car occupant injured in collision with car, pick-up truck or van |
| V44 | Car occupant injured in collision with heavy transport vehicle or bus |
| V45 | Car occupant injured in collision with railway train or railway vehicle |
| V46 | Car occupant injured in collision with other nonmotor vehicle |
| V47 | Car occupant injured in collision with fixed or stationary object |
| V48 | Car occupant injured in noncollision transport accident |
| V49 | Car occupant injured in other and unspecified transport accidents |

**Table 3.7 –** Occupant of pick-up truck or van fatalities codes

| Code | Specification |
|---|---|
| V50 | Occupant of pick-up truck or van injured in collision with pedestrian or animal |
| V51 | Occupant of pick-up truck or van injured in collision with pedal cycle |
| V52 | Occupant of pick-up truck or van injured in collision with two- or three-wheeled motor vehicle |
| V53 | Occupant of pick-up truck or van injured in collision with car, pick-up truck or van |
| V54 | Occupant of pick-up truck or van injured in collision with heavy transport vehicle or bus |
| V55 | Occupant of pick-up truck or van injured in collision with railway train or railway vehicle |
| V56 | Occupant of pick-up truck or van injured in collision with other nonmotor vehicle |
| V57 | Occupant of pick-up truck or van injured in collision with fixed or stationary object |
| V58 | Occupant of pick-up truck or van injured in noncollision transport accident |
| V59 | Occupant of pick-up truck or van injured in other and unspecified transport accidents |

**Table 3.8 –** Occupant of heavy transport vehicle fatalities codes

| Code | Specification |
|---|---|
| V60 | Occupant of heavy transport vehicle injured in collision with pedestrian or animal |
| V61 | Occupant of heavy transport vehicle injured in collision with pedal cycle |
| V62 | Occupant of heavy transport vehicle injured in collision with two- or three-wheeled motor vehicle |
| V63 | Occupant of heavy transport vehicle injured in collision with car, pick-up truck or van |
| V64 | Occupant of heavy transport vehicle injured in collision with heavy transport vehicle or bus |
| V65 | Occupant of heavy transport vehicle injured in collision with railway train or railway vehicle |
| V66 | Occupant of heavy transport vehicle injured in collision with other nonmotor vehicle |
| V67 | Occupant of heavy transport vehicle injured in collision with fixed or stationary object |
| V68 | Occupant of heavy transport vehicle injured in noncollision transport accident |
| V69 | Occupant of heavy transport vehicle injured in other and unspecified transport accidents |

**Table 3.9–** Bus occupant fatalities codes

| Code | Specification |
|---|---|
| V70 | Bus occupant injured in collision with pedestrian or animal |
| V71 | Bus occupant injured in collision with pedal cycle |
| V72 | Bus occupant injured in collision with two- or three-wheeled motor vehicle |
| V73 | Bus occupant injured in collision with car, pick-up truck or van |
| V74 | Bus occupant injured in collision with heavy transport vehicle or bus |
| V75 | Bus occupant injured in collision with railway train or railway vehicle |
| V76 | Bus occupant injured in collision with other nonmotor vehicle |
| V77 | Bus occupant injured in collision with fixed or stationary object |
| V78 | Bus occupant injured in noncollision transport accident |
| V79 | Bus occupant injured in other and unspecified transport accidents |

**Table 3.10 –** Other traffic fatalities codes

| Code | Specification |
|---|---|
| V80 | Animal-rider or occupant of animal-drawn vehicle injured in transport accident |
| V81 | Occupant of railway train or railway vehicle injured in transport accident |
| V82 | Occupant of streetcar injured in transport accident |
| V83 | Occupant of special vehicle mainly used on industrial premises injured in transport accident |
| V84 | Occupant of special vehicle mainly used in agriculture injured in transport accident |
| V85 | Occupant of special construction vehicle injured in transport accident |
| V86 | Occupant of special all-terrain or other motor vehicle designed primarily for off-road use, injured in transport accident |
| V87 | Traffic accident of specified type but victim's mode of transport unknown |
| V88 | Nontraffic accident of specified type but victim's mode of transport unknown |
| V89 | Motor- or nonmotor-vehicle accident, type of vehicle unspecified |

The final outcome indicators are risks consisting out of the number of fatalities (nominator) and an exposure measure (denominator). In order to have a good coherence between the two, some slight changes in the nominator are desirable. When the exposure measure is the population (resulting in the mortality rate), we excluded the categories presented in Table 3.11, because they do not represent transportation modes victims on which we focus in this research.

**Table 3.11 –** Fatality data excluded for mortality rate composition

| Code | Specification |
|---|---|
| V81 | Occupant of railway train or railway vehicle injured in transport accident |
| V82 | Occupant of streetcar injured in transport accident |
| V83 | Occupant of special vehicle mainly used on industrial premises injured in transport accident |
| V84 | Occupant of special vehicle mainly used in agriculture injured in transport accident |
| V85 | Occupant of special construction vehicle injured in transport accident |
| V86 | Occupant of special all-terrain or other motor vehicle designed primarily for off-road use, injured in transport accident |
| V88 | Nontraffic accident of specified type but victim's mode of transport unknown |

Nevertheless, when the exposure measure relates to the road transport motor vehicle fleet or the road transport motor vehicle kilometers traveled (VKT), resulting in a fatality rate (FR), fatality categories that have no relation to these vehicles, will not be considered. These categories and their respective codes are listed in Table 3.12.

**Table 3.12 –** Fatality data excluded for fatality rate composition

| Code | Specification |
|---|---|
| V01 | Pedestrian injured in collision with pedal cycle |
| V05 | Pedestrian injured in collision with railway train or railway vehicle |
| V06 | Pedestrian injured in collision with other nonmotor vehicle |
| V10 | Pedal cyclist injured in collision with pedestrian or animal |
| V11 | Pedal cyclist injured in collision with other pedal cycle |
| V15 | Pedal cyclist injured in collision with railway train or railway vehicle |
| V16 | Pedal cyclist injured in collision with other nonmotor vehicle |
| V17 | Pedal cyclist injured in collision with fixed or stationary object |
| V81 | Occupant of railway train or railway vehicle injured in transport accident |
| V82 | Occupant of streetcar injured in transport accident |
| V83 | Occupant of special vehicle mainly used on industrial premises injured in transport accident |
| V84 | Occupant of special vehicle mainly used in agriculture injured in transport accident |
| V85 | Occupant of special construction vehicle injured in transport accident |
| V86 | Occupant of special all-terrain or other motor vehicle designed primarily for off-road use, injured in transport accident |
| V88 | Nontraffic accident of specified type but victim's mode of transport unknown |

For disaggregated analysis in which we consider the main different vehicle categories (motorcycle, car, truck and bus) in relation to whatever exposure data (e.g. in the analysis we perform in Chapters 6 to 8), it is essential to consider the share of traffic fatalities classified as "V-89 Motor- or nonmotor-vehicle accident". The existence of high and substantially distinct shares of this type of classification affects the state comparisons for a certain vehicle category, since a low number of fatalities for occupants of this vehicle type may be mostly related to an underreporting problem and it should not be confounded with a better road safety situation.

Actually, as a general tendency, those states with an inferior road safety performance tend to present less reliable databases due to higher underreporting levels. The graph in Figure 3.1 contains the average percentages of V-89 fatalities for each Brazilian state in the period 2010-2012.

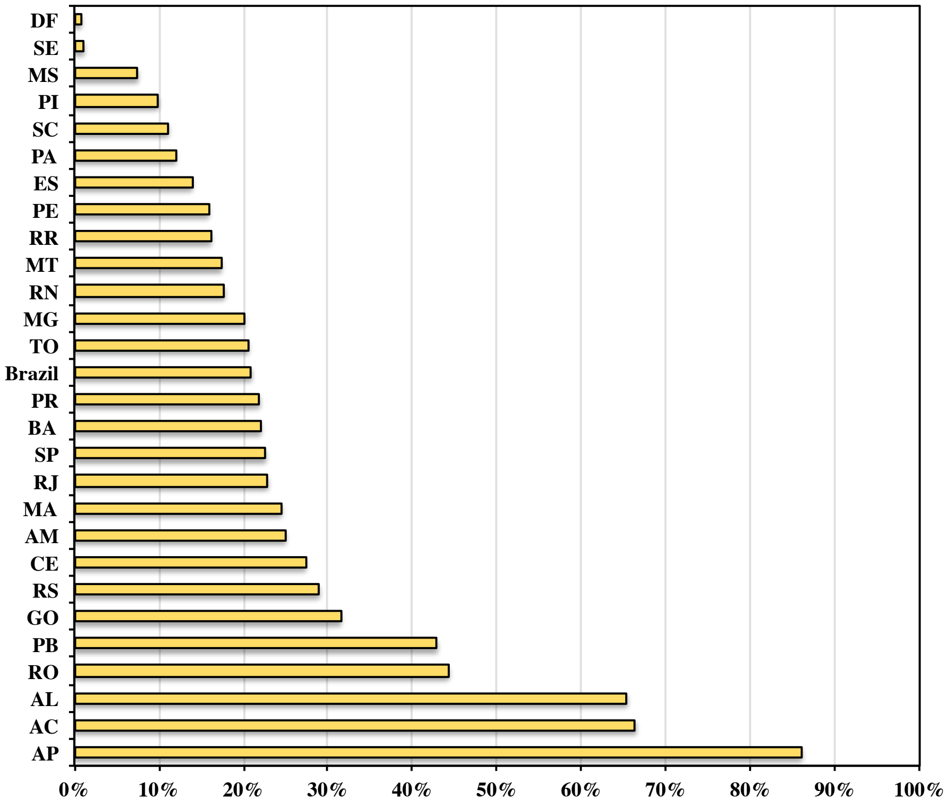



**Figure 3.1 –** Share of traffic fatalities classified as V89

The selected procedure to tackle this deficiency in the database is to distribute the traffic fatalities classified in the V89 code among the other identified types of fatalities according to their original distribution. For instance, if in the sum of motorcycle, car, truck and bus occupant fatalities the respective shares are 20%, 30%, 5% and 1% (the pedestrian and cyclist fatalities complete the 100% value, with e.g. 35% and 9%, respectively), it means that 20% of "V89 fatalities" will be considered as motorcycle occupant fatalities, 30% as car occupants fatalities, 5% as truck occupants fatalities and 1% as car occupants fatalities. This procedure allows a more fair comparison of traffic fatality data of different states on a supposed common basis of "reportability" and it will be particularly useful for the investigations we describe in Chapers 6 to 8.

## 3.2. OVERALL DATASET

This research is fundamentally based on traffic fatality data for the set of 27 Brazilian states. Therefore, it is important to statistically evaluate the three most common road safety risk indicators that one can derive from it:

- MR – mortality rate (fatalities per 100 thousand inhabitants);
- FR1 – fatality rate (fatalities per 10 thousand motor vehicles);
- FR2 – fatality rate (fatalities per billion vehicle kilometers traveled);

Table 3.13 contains information on these indicators for which the average of the annual values in the period 2010-2012 is considered. The complete data tables for the Brazilian states regarding each indicator are available in Appendix C.

**Table 3.13** – Outcome based road safety indicators

| Denomination | Indicator | Min | Max | Unit | Data source |
|---|---|---|---|---|---|
| Mortality rate (health risk) | Traffic fatalities *per capita* | 13.52 | 39.29 | Fat./$10^5$ inh. | DATASUS/ IBGE |
| Fatality rate (fleet risk) | Traffic fatalities per vehicle | 3.40 | 18.30 | Fat./$10^4$ veh. | DATASUS/DENATRAN |
| Fatality rate (traffic risk) | Traffic fatalities per traveled distance | 30.45 | 117.07 | Fat./$10^9$ km | DATASUS/Author's research[1] |

In addition, we constituted a set of information on different road safety domains using various sources. We identified a set of 36 indicators (in 5 categories) in an attempt to collect the key available information for road safety performance investigation at the "per state" level. The description of these indicators is presented in Tables 3.14 to 3.18, which also contain minimum and maximum columns corresponding to the average of the annual values in the period 2010-2012.

The consideration of penalty data requires some key assumptions. Firstly, enforcement actions are based on the hope to detect risk behaviors in a drivers' sample, from which the characteristics can be generalized to the entire drivers' population. Following this simple logic, if a certain risk behavior presents a higher incidence in comparison to others, more efforts should be put in trying to prevent such attitude. Thus, the penalty rates are useful to identify the most common risk behavior in a certain state.

In the particular case of Brazil, it is important to stress that a penalty rate (per population or per vehicle, for example) should not be used to infer if one population has a riskier behavior in comparison to another (i.e., considering the population of two states). The reason for that relies on the fact that the size of these rates is primarily affected by the enforcement level (and focus); i.e., the higher the enforcement level, the higher the penalty rates. All penalty rates

[1] Values available in Appendix B.

(*per capita* or per vehicle) present negative correlations with the fatality rates. So, if any state presents a higher penalty rate, it should not be interpreted as a worse behavior of drivers, but rather as a superior enforcement level there. So, they are not ideal SPIs.

Although we don't use the background indicators in the SPI research, we analyze them because they are used in the clustering procedure we describe in Chapter 4.

**Table 3.14** – Road user behavior related indicators

| Domain | Indicators | Min | Max | Unit | Data source |
|---|---|---|---|---|---|
| Alcohol | 1) Share of people who drink and drive in the last 30 days | 0.60 | 3.40 | % | *Ministério da Saúde* (2011a) |
| | 2) Share of people who drink and drive | 3.85 | 13.80 | % | |
| | 3) Share of 9th year students who declared to be passenger in a vehicle with a driver who drink | 18.40 | 31.50 | % | *Ministério da Saúde* (2012) |
| | 4) Share of people involved in accident who declared to drink before | 0.08 | 0.29 | % | *Ministério da Saúde* (2011b) |
| Cell phone | 5) Cell phone related infractions per capita* | 0.08 | 20.37 | Inf./$10^5$ inh. | DETRAN |
| | 6) Cell phone related infractions per vehicle* | 0.42 | 40.67 | Inf./ $10^5$ veh. | DETRAN |
| Seatbelt | 7) Share of people involved in accident wearing seatbelt on federal highways | 93.56 | 99.25 | % | *Ministério dos Transportes* (2013a) |
| | 8) Share of 9th year students who declared not to wear seatbelt in a car in the last 30 days (of those who were passenger in a car in this period) | 6.70 | 23.35 | % | *Ministério da Saúde* (2012) |
| | 9) Share of people involved in car accidents wearing seatbelt | 42.90 | 86.00 | % | *Ministério da Saúde* (2011b) |
| Helmet | 10) Share of people involved in accident wearing helmet on federal highways | 80.57 | 100.00 | % | *Ministério dos Transportes* (2013a) |
| | 11) Share of 9th year students who declared to wear helmet on a motorcycle in the last 30 days (of those who were passenger on a motorcycle in this period) | 66.86 | 97.92 | % | *Ministério da Saúde* (2012b) |
| | 12) Share of people involved in motorcycle accidents wearing helmet | 0.00 | 76.27 | % | *Ministério da Saúde* (2011b) |
| Speeding | 13) Speeding related infractions per capita* | 3.75 | 401.16 | Inf./inh. | DETRAN |
| | 14) Speeding related infractions per vehicle* | 10.16 | 801.00 | Inf./veh. | DETRAN |

*Incomplete datasets

**Table 3.15** – Road infrastructure related indicators

| Topic | | | Inidcator | Min | Max | Unit | Data source |
|---|---|---|---|---|---|---|---|
| Signing | Road markings | Central | 15) Share of highway length with central road markings | 69.17 | 99.43 | % | CNT (2011) |
| | | | 16) Share of highway length with clearly visible central road markings | 33.69 | 95.14 | % | |
| | | Lateral | 17) Share of highway length with lateral road markings | 43.77 | 95.40 | % | CNT (2011) |
| | | | 18) Share of highway length with clearly visible lateral road markings | 31.87 | 91.96 | % | |
| | Vertical signs | - | 19) Share of highway length with vertical signs | 29.43 | 86.16 | % | CNT (2011) |
| | | - | 20) Share of highway length with clearly visible vertical signs | 34.12 | 99.19 | % | |
| | | - | 21) Share of highway length with clearly legible vertical signs | 24.33 | 91.80 | % | |
| Roadside | Central division | - | 22) Share of multilane highways in relation to the paved highways | 0.00 | 17.06 | % | *Ministério dos Transportes* (2013b) |
| | Barriers | - | 23) Share of highway length equipped with adequate barriers | 1.00 | 73.10 | % | CNT (2011) |
| | Shoulders | - | 24) Share of highway length with adequate shoulders | 0.01 | 79.97 | % | |

**Table 3.16** – Health system related indicators

| Topic | Indicator | Min | Max | Unit | Data source |
|---|---|---|---|---|---|
| Health professionals | 25) Number of medical doctors *per capita* | 0.53 | 3.62 | Doc./$10^3$ inh. | *Ministério da Saúde* (2011c) |
| | 26) Number of health professionals *per capita* | 4.64 | 16.85 | Prof./$10^3$ inh. | |
| Health expenditure | 27) Health expenditure *per capita* | 444.02 | 957.46 | R$/inh. | *Ministério da Saúde* (2011c) |
| | 28) Number of hospital beds *per capita* | 4.64 | 16.85 | Beds/$10^3$ inh. | |

**Table 3.17** – Vehicle fleet related Indicators

| Topic | Indicators | Min | Max | Unit | Data source |
|---|---|---|---|---|---|
| Fleet composition | 29) Share of motorcycles in the total fleet | 11.25 | 56.41 | % | DENATRAN (2013) |
| | 30) Share of trucks in the total fleet | 1.76 | 5.92 | % | |
| Age of the fleet | 31) Share of 10-year vehicles or older | 22.00 | 56.27 | % | DENATRAN (2013) |
| | 32) Share of 5-year vehicles or newer | 25.83 | 56.34 | % | |

**Table 3.18** – Background indicators

| Topic | Indicator | Min | Max | Unit | Data source |
|---|---|---|---|---|---|
| Gross Domestic Product | 33) GDP *per capita* | 6,888.60 | 58,489.46 | R$/inh. | *Ministério da Saúde* (2011d) |
| Motorization level | 34) Motorized vehicles *per capita* | 12.91 | 54.39 | Veh./$10^2$ inh. | DENTRAN (2013) |
| Highway density | 35) Paved highways length per area | 1.31 | 517.32 | km/$10^3$ km$^2$ | *Ministério dos Transportes* (2013b) |
| Urbanization level | 36) Share of urban population | 78.97 | 99.19 | % | *Ministério da Saúde* (2013) |

## 3.3. OUTLIER DETECTION

Previously to running any analysis using a dataset, the step of detecting outlier observations is needed. An outlier is an extremely unrepresentative data point, which lies far away from the main body of the data (Petruccelli, Nandram, & Chen, 1999). In the scope of this research, the presence of values outside an expected range may be explainable due to very specific circumstance in a particular state or may be attributed to mistakes in data compilation or even to the consideration of different premises in data collection according to each state.

For the data collected on a national basis, in which data processing is performed in accordance with standardized procedures, e.g. the information provided by the Ministry of Health, the chance of finding (unintended) outliers is limited. On the other hand, for information collected by state level organizations/institutions, the chances of having aberrant observations increases, since data collection procedures may vary from one place to another (e.g. traffic infraction numbers provided by the DETRAN's).

We carried out a univariate outlier detection process by means of the z-score test. The z-score method establishes rule-of-thumb limits outside of which a measurement is considered to be an outlier. In general, observations with z-scores greater than 3 in absolute value are considered outliers (McClave & Sincich, 2003). Conceptually, the z-score is a measure of a certain data point deviation from the mean in terms of standard deviation units and can be computed as follows (Raktoe & Hubert, 1979):

$$z_s = \frac{(z - \bar{z})}{sd} \tag{1.1}$$

In which:

$$z_s = z - score \qquad z = data\ point \qquad \bar{z} = mean \qquad sd = standard\ deviation$$

From the whole data set tested (36 indicators), only 12 data points (equivalent to 1.23% of all observations) presented $z_s > 3.0$, as listed in Table 3.

**Table 3.19** – Outlier observations

| Parameter | State | $z_S$ |
|---|---|---|
| Share of people who drink and drive in the last 30 days | SP | -4,106 |
| Cell phone related infractions per capita | SP | 3,241 |
| Share of 9th year students who declared not to wear seatbelt in a car in the last 30 days (of those who were passenger in a car in this period) | RJ | -3,169 |
| Speeding related infractions per capita | SP | 3,093 |
| Speeding related infractions per vehicle | SP | 3,002 |
| Share of highway length with clearly visible vertical signs | PA | -3,073 |
| Share of multilane highways in relation to the paved highways | DF | 3,296 |
| Share of highway length equipped with adequate barriers | RJ | 3,053 |
| Share of motorcycles in the total fleet | DF | -3,489 |
| Share of trucks in the total fleet | DF | -3,612 |
| GDP *per capita* | DF | 3,997 |
| Paved highway length per area | DF | 4,503 |

Although technically speaking they are all outlier observations, their occurrence is to a large extent justifiable by the particular characteristics of these states. For example, São Paulo is by far the most developed state in Brazil, and also has the lowest fatality rates (both per vehicle and per kilometers traveled). Thus, it is not surprising that a higher enforcement level is found in comparison to the other federation units. The Distrito Federal, a relatively small area where construction dates from the 60's decade, and which contains the Brazilian capital city, presents some peculiar characteristics that help to explain the obtained values:

- This place houses the federal government headquarters and also many other federal public institutions, where the incomes are above the Brazilian average;
- The relative importance of the capital city numbers, where more services (including health services) are present, increases when the interior area is small, as is the case in the Distrito Federal;
- Many radial (and multilane) motorways have their origin in the political center of the country, causing a higher concentration closer to the capital Brasilia.

For the other states containing outlier observations, RJ and PA, although there are not so obvious explanations, the z-scores are not much higher than the limit. Hence, we decided to keep these data in the upcoming analyses, since they are part of the data set and their values, although out of range, are reasonable. Moreover, their deletion would generate the need to impute new values, which could also be contested by other reasons.

Next, we check the correlation that each indicator has with the outcome indicators as it helps in selecting a good safety performance indicator set. However, before proceeding to the correlation analysis, some investigation on the probability distribution followed by each indicator set of values is needed.

## 3.4. PROBABILITY DISTRIBUTION

Since the assumption of normality is an aspect present in many statistical procedures applied on the dataset, an effective test of whether the assumption holds or not is required. After many effectiveness studies on normality tests, the use of the Wilk-Shapiro test in everyday practice is recommended – preferably to the equally known and commonly used Kolmogorov-Smirnov or Chi-squared tests (Thode Jr., 2002).

The values for each indicator were submitted to the Wilk-Shapiro normality test performed in the software R® (R Core Team, 2014). The test statistic is obtained by the ratio between the square of an appropriate linear combination of the sample order statistics and the usual symmetric estimate of variance (Shapiro & Wilk, 1965). The test results are available in Table 3.20 and Table 3.21 respectively for the safety performance and outcome indicators.

The p-value determines whether the null hypothesis (i.e. the population from which the sample originates presents a normal distribution) or the alternative hypothesis (i.e. the population from which the sample originates does not present a normal distribution) applies. We adopted the significance level of $\alpha = 0.05$, therefore:

- If p-value $\geq 0.05$, the indicator presents a normal distribution (underlined values);
- If p-value $< 0.05$, the indicator does not present a normal distribution.

Now that we know the probability distribution of each indicator, appropriate correlation methods can be applied in order to verify the relationship between the SPIs and the road safety outcomes. We describe this procedure in the following item.

**Table 3.20** – Results of the Wilk-Shapiro normality test for the SPIs

| Indicator | Abbreviation | p-value |
|---|---|---|
| Share of people who drink and drive in the last 30 days | A1 | 0.00001 |
| Share of people who drink and drive | A2 | 0.14132 |
| Share of 9th year students who declared to be passenger in a vehicle with a driver who drink | A3 | 0.75638 |
| Share of people involved in accident who declared to drink before | A4 | 0.29137 |
| Cell phone related infractions per capita | CPPC | 0.00119 |
| Cell phone related infractions per vehicle | CPPV | 0.18500 |
| Share of people involved in accident wearing seatbelt in federal highways | H1 | 0.00020 |
| Share of 9th year students who declared not to wear seatbelt in a car in the last 30 days (of those who were passenger in a car in this period) | H2 | 0.00252 |
| Share of people involved in car accidents wearing seatbelt | H3 | 0.13011 |
| Share of people involved in accident wearing helmet in federal highways | SB1 | 0.00004 |
| Share of 9th year students who declared to wear helmet on a motorcycle in the last 30 days (of those who were passenger on a motorcycle in this period) | SB2 | 0.00084 |
| Share of people involved in motorcycle accidents wearing helmet | SB3 | 0.99889 |
| Speeding related infractions per capita* | S1 | 0.00013 |
| Speeding related infractions per vehicle* | S2 | 0.00059 |
| Share of highway length with central road markings | PCRM | 0.03906 |
| Share of highway length with clearly visible central road markings | VCRM | 0.01048 |
| Share of highway length with lateral road markings | PLRM | 0.23904 |
| Share of highway length with clearly visible lateral road markings | VLRM | 0.01101 |
| Share of highway length with vertical signs | PVS | 0.50557 |
| Share of highway length with clearly visible vertical signs | VVS | 0.01418 |
| Share of highway length with clearly legible vertical signs | LVS | 0.45557 |
| Share of multilane highways in relation to the paved highways | SMH | 0.00005 |
| Share of highway length equipped with adequate barriers | AB | 0.00003 |
| Share of highway length with adequate shoulders | PCS | 0.61133 |
| Number of hospital beds per capita | HBPC | 0.93691 |
| Number of medical doctors per capita | DPC | 0.00034 |
| Number of health professionals per capita | HPPC | 0.00382 |
| Health expenditure per capita | HEPC | 0.63965 |
| Share of motorcycles in the total fleet | SMF | 0.00007 |
| Share of trucks in the total fleet | STF | 0.00097 |
| Share of 10-year vehicles or older | VTYOM | 0.06243 |
| Share of 5-year vehicles or newer | VFYOL | 0.36612 |
| GDP per capita | GDPC | 0.00002 |
| Motorized vehicles per capita | MOTR | 0.03720 |
| Paved highways length per area | HD | 0.00000 |
| Share of urban population | UP | 0.87092 |

**Table 3.21** – Results of the Wilk-Shapiro normality test for the outcome indicators

| Outcome indicator | Abbreviation | p-value |
|---|---|---|
| Mortality rate | MR | 0.081589 |
| Fatality rate – per vehicle | FR1 | 0.002097 |
| Fatality rate – per km traveled | FR2 | 0.220797 |

## 3.5. CORRELATION ANALYSIS

In addition to data availability, the SPIs choice is based on their correlation with the final outcomes. Two correlation methods were applied depending on the data probability distribution: Pearson, for data following a normal distribution; and Spearman, for data not following a normal distribution.

The Pearson product moment coefficient of correlation is a measure of the strength of the linear relationship between two variables (McClave & Sincich, 2003). A coefficient value near or equal to *0* implies little or no linear relationship. In contrast, the closer the coefficient comes to 1 or -1, the stronger the linear relationship between the two variables. And if $r = 1$ or $r = -1$, all the sample points fall exactly on the least squares line. Positive values imply a positive linear relationship; that is, if one increases, the other also increases. Negative values imply a negative linear relationship; that is, if one increases, the other decreases. This correlation coefficient is computed according to Equation 1.2:

$$r_p = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} \tag{1.2}$$

In which:

$$SS_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

$$SS_{xx} = \sum(x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$SS_{yy} = \sum(y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$\boldsymbol{r_p = Pearson's\ correlation\ coefficient}$ $\quad$ $\boldsymbol{x, y = variables\ being\ tested}$

$\boldsymbol{SS = Sum\ of\ squares}$ $\quad$ $\boldsymbol{n = sample\ size}$

In the nonparametric statistics, Spearman's rank correlation coefficient provides a measure of correlation between two ranks (McClave & Sincich, 2003). Analogously, perfect positive correlation between the pair of ranks is characterized by a Spearman's correlation coefficient equal to 1; on the other hand, when the ranks indicate perfect disagreement, there is perfect negative correlation equal to -1. The nearer the coefficient is to 0, the lower the correlation is. Its computation is according to Equation 1.3.

$$r_s = \frac{SS_{uv}}{\sqrt{SS_{uu}SS_{vv}}} \tag{1.3}$$

In which:

$$SS_{uv} = \sum(u_i - \bar{u})(v_i - \bar{v}) = \sum u_i v_i - \frac{(\sum u_i)(\sum v_i)}{n}$$

$$SS_{uu} = \sum(u_i - \bar{u})^2 = \sum u_i^2 - \frac{(\sum u_i)^2}{n}$$

$$SS_{vv} = \sum(v_i - \bar{v})^2 = \sum v_i^2 - \frac{(\sum v_i)^2}{n}$$

$r_s = Spearman's\ correlation\ coefficient \quad SS = Sum\ of\ squares \quad n = sample\ size$

$u_i, v_i = rank\ of\ the\ i^{th}\ observation\ in\ samples\ 1\ and\ 2, respectively$

In order to test if the found correlations also apply for the population (which correlation is denoted by $\rho$), the statistical significance should be tested by conducting the following hypothesis test:

- $H_0$: $\rho = 0$ (no population correlation between variables/ranks)
- $H_a$: $\rho \neq 0$ (population correlation between variables/ranks)

Using "t" and "s" statistics, respectively for Pearson and Spearman coefficients, the presented hypotheses are tested for a confidence interval of 95% with:

- p-value $\leq 0.05$, the variables/ranks correlation is statistically significant;
- p-value $> 0.05$, the variables/ranks correlation is not statistically significant.

The correlations between each SPI and the three final outcome indicators are listed in Table 3.22 (in which the p-values are also mentioned).

**Table 3.22** – Correlations between the SPIs and the outcome indicators

| Indicators | Outcome indicators | | | | | |
|---|---|---|---|---|---|---|
| | **MR** | **p-value** | **FR1** | **p-value** | **FR2** | **p-value** |
| **A1** | 0.41 | 0.05263 | 0.21 | 0.00128 | 0.10 | <u>0.00837</u> |
| **A2** | 0.72 | <u>0.00001</u> | 0.02 | 0.73687 | 0.12 | 0.72782 |
| **A3** | 0.69 | <u>0.00142</u> | -0.18 | 0.81047 | 0.00 | 0.64411 |
| **A4** | -0.01 | 0.76645 | 0.61 | <u>0.01518</u> | 0.55 | <u>0.00119</u> |
| **CPPC** | 0.08 | 0.19553 | 0.76 | <u>0.00001</u> | 0.56 | <u>0.00029</u> |
| **CPPV** | 0.51 | 0.15495 | 0.41 | <u>0.01687</u> | 0.53 | 0.23882 |
| **H1** | -0.21 | 0.38338 | 0.68 | <u>0.00745</u> | 0.57 | <u>0.00582</u> |
| **H2** | -0.52 | <u>0.00578</u> | 0.06 | 0.38848 | 0.10 | 0.57725 |
| **H3** | -0.14 | 0.27764 | 0.49 | 0.08653 | 0.44 | <u>0.00970</u> |
| **SB1** | 0.16 | 0.97059 | 0.12 | 0.72600 | -0.06 | 0.81857 |
| **SB2** | 0.23 | 0.07728 | -0.46 | 0.12170 | -0.20 | 0.55882 |
| **SB3** | -0.50 | <u>0.03039</u> | 0.35 | 0.24600 | 0.16 | 0.19818 |
| **S1** | 0.30 | 0.12331 | 0.84 | <u>0.00026</u> | 0.66 | <u>0.00049</u> |
| **S2** | 0.59 | 0.05191 | 0.71 | <u>0.00103</u> | 0.54 | <u>0.00174</u> |
| **PCRM** | 0.20 | 0.41316 | 0.57 | <u>0.00322</u> | 0.50 | <u>0.01607</u> |
| **VCRM** | 0.18 | 0.77655 | 0.07 | 0.23662 | -0.15 | 0.76286 |
| **PLRM** | 0.16 | 0.43640 | 0.44 | <u>0.01468</u> | 0.37 | 0.06052 |
| **VLRM** | 0.19 | 0.91594 | 0.04 | 0.26746 | -0.17 | 0.73783 |
| **PVS** | 0.05 | 0.70436 | 0.67 | <u>0.00003</u> | 0.56 | <u>0.00990</u> |
| **VVS** | -0.06 | 0.53020 | 0.46 | <u>0.02911</u> | 0.30 | 0.15581 |
| **LVS** | -0.05 | 0.99636 | 0.22 | 0.15860 | 0.18 | 0.55478 |
| **SMH** | 0.17 | 0.31279 | 0.63 | <u>0.00001</u> | 0.41 | <u>0.00079</u> |
| **AB** | -0.02 | 0.80954 | 0.30 | 0.11991 | 0.07 | 0.28500 |
| **PCS** | -0.02 | 0.72716 | 0.39 | <u>0.00371</u> | 0.32 | 0.26269 |
| **HBPC** | -0.14 | 0.49087 | 0.52 | <u>0.01775</u> | 0.23 | 0.21403 |
| **DPC** | 0.01 | 0.48290 | 0.76 | <u>0.00000</u> | 0.49 | <u>0.00080</u> |
| **HPPC** | 0.05 | 0.22323 | 0.61 | <u>0.00010</u> | 0.68 | <u>0.00004</u> |
| **HEPC** | -0.05 | 0.81813 | 0.44 | 0.08023 | 0.45 | <u>0.00682</u> |
| **SMF** | 0.31 | 0.19977 | 0.77 | <u>0.00001</u> | 0.63 | <u>0.00028</u> |
| **STF** | 0.48 | 0.11394 | -0.13 | 0.41238 | 0.00 | 0.35026 |
| **VTYOM** | 0.08 | 0.97693 | -0.75 | <u>0.00016</u> | -0.46 | <u>0.01790</u> |
| **VFYOL** | 0.04 | 0.72946 | -0.77 | <u>0.00001</u> | -0.57 | <u>0.00682</u> |
| **GDPC** | -0.02 | 0.74435 | 0.86 | <u>0.00001</u> | 0.77 | <u>0.00010</u> |
| **MOTR** | -0.37 | 0.09677 | 0.79 | <u>0.00000</u> | 0.58 | <u>0.00039</u> |
| **HD** | 0.02 | 0.77165 | 0.35 | <u>0.01306</u> | 0.06 | 0.08338 |
| **UP** | 0.14 | 0.72487 | 0.80 | <u>0.00001</u> | 0.75 | <u>0.00006</u> |

Furthermore, it is desirable that each indicator in Table 3.22 presents an as low as possible correlation with each other, in a belief that different dimensions of the problem are covered and its overall investigation will be more complete. However, the existence of some degree of correlation is difficult to avoid, since all indicators are related to the same problem in such a level that new singular associations become more improbable.

## 3.6. MISSING DATA TREATMENT

As in many data-driven analysis studies, there is the problem of missing data for some of the indicators (especially the SPIs). The idea of obtaining a complete data set is valuable for enabling SPI research that is capable to cover the main strategic road safety aspects such as: the road user (more specifically behavior with respect to alcohol, cell-phone, protective systems and speeding), the road infrastructure, the health system and the vehicle fleet. In total, these aspects decompose into 32 SPIs – presented in the previous Table 3.22 (except for the last four background indicators).

Although we have a complete dataset regarding background indicators (such as GDP per capita, motorization rate, etc.), we do not consider them for the SPI research (subject of Chapter 9), because they represent themes in which it is difficult or at least unreasonable to achieve changes in order to produce benefits for road safety. In other words, background indicators do not provide a realistic direction for action towards road safety – we cannot suggest that a state should increase its GDP in order to improve road safety. Actually, background indicators actuate in a more fundamental level and are more useful to take into consideration when clustering the Brazilian states (subject of Chapter 4).

Figure 3.2 shows the missing data situation in the set of 32 safety performance indicators. The "missingness" level of 12.50% for the variables means that 12.5% of the SPIs contain missing data; the "missingness" level of 55.66% for the cases means that 55.56% of the states present missing data; and the "missingness" level of 5.32% for the values means that 5.32% of the observations (data points) are missing.

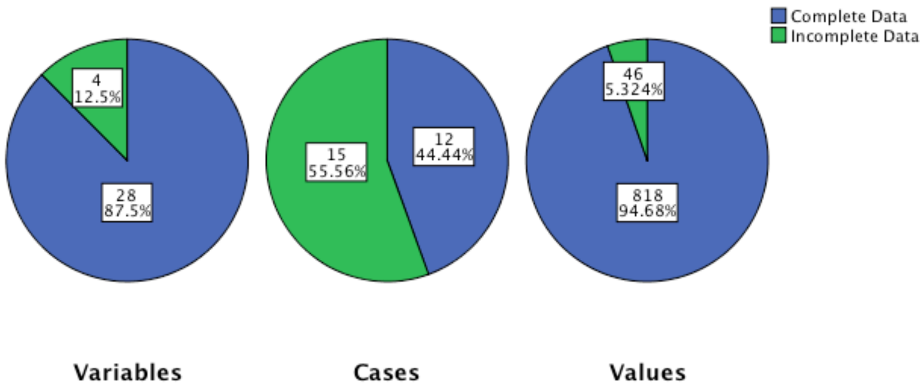



**Figure 3.2** – Overall summary of missing SPI values (extracted from SPSS)

Figure 3.3 contains the "missingness map" of the indicators in which the white boxes represent the missing data. The states that present a higher share of missing data are AC, AM, AL, BA, PB, MG, DF and MS (12.50%), followed by AP, CE, PI, GO and MT (6.25%).

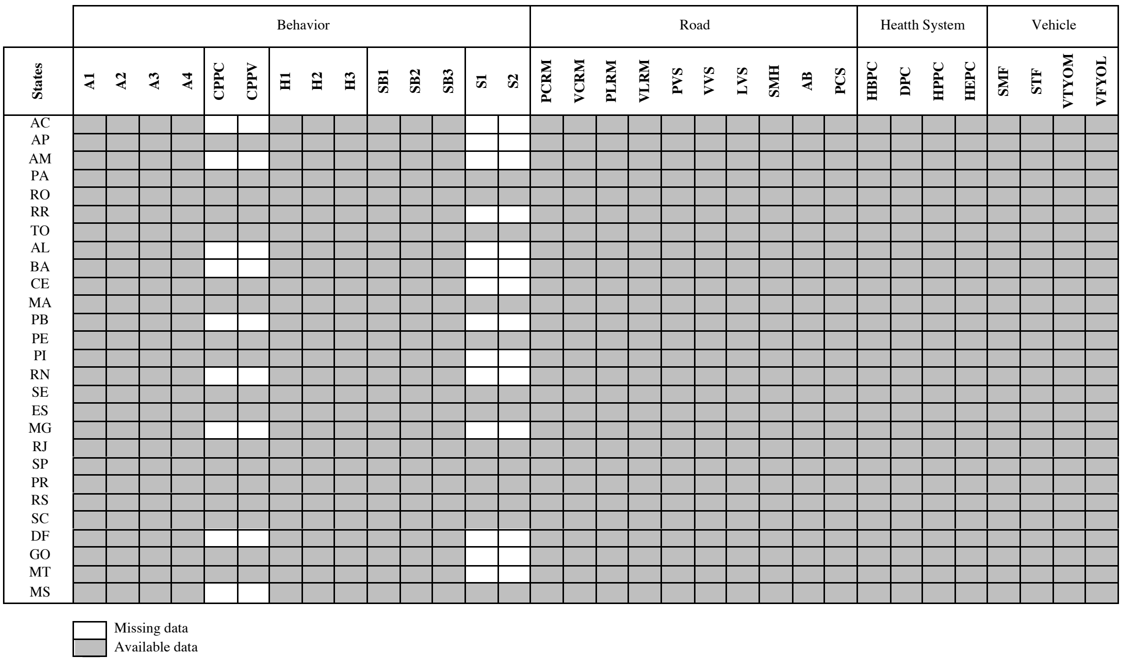



**Figure 3.3** – "Missingness map" for the set of selected indicators

According to the "missingness map", the missing information is concentrated in the behavior related group, more specifically in the indicators manifesting the cell-phone usage and speeding. On the other hand, complete data sets are available concerning the three remaining road safety aspects (road, health system and vehicle).

The concentration of missing data in the behavioral domain is probably attributed to the fact that this information is not collected in neither a systematic nor unified way; i.e. , it is information that is only available at the state level through the DETRANs. It implies that in the set of 27 Brazilian states, some of them publish this information on a regular basis (through their DETRAN's websites) and others do not. Additionally, the availability of information might vary according to the type of penalty. In order to compile a larger number of information regarding traffic penalties over the Brazilian states, an extensive research was conducted based on the DETRAN's websites. In the cases where states did not publish this information, electronic mailing was used to request it. Although this procedure was helpful to obtain more data, there is still some lack of information for a group of states; in other words, there is a certain amount of missing data.

In order to deal with missing data there are several procedures available in the specialized literature; they might be generalized into three main strategies: "using as it is", deletion and imputation (Kabak & Ruan, 2010). The first and the most arbitrary of them, "using as it is", is not an alternative for this research since the optimization models applied require the principle of positivity (with no zero or negative values).

The deletion strategy, which consists of excluding the units (states) with incomplete information, would cause the deletion of 15 of the 27 Brazilian states, introducing a substantial bias to the national level evaluation. It would also be an alternative to exclude this domain from the set of analyzed road safety indicators, since the "missingness" concentrates in the behavioral domain. However, due to the behavioral domain's relatively high importance in the road safety context (Wierwille et al., 2002), neglecting it is certainly not a plausible alternative, because the research conclusions would be of very low value without considering this key domain.

Since there is the belief of not excluding any state neither any indicators domain from the analysis, the following step is to artificially impute the missing information; in other words, properly substituting the missing information by computed values and completing the data set This is currently the most common strategy to deal with missing data. In literature, imputation methods are classified in two categories: single imputation and multiple imputation. In single imputation, the missing values are replaced based on the available information for the given variable itself, being the missing values substituted by its mean, median, mode or by the result of a regression or expectation-maximization model. However, these techniques ignore the uncertainty of missing data prediction (Rubin, 1987; Wilmots et al., 2011).

In the multiple imputation procedure, originally proposed by Rubin (1987), the missing values are replaced by a set of $m>1$ plausible random values representing the uncertainty about this prediction; this process results in valid statistical inferences that properly reflect the uncertainty due to missing values (Yuan, 2014). In this process, the available values are used as predictors of the values to be imputed. We produced $m=5$ imputed data sets using the software SPSS IBM Corp®. Next, we averaged the 5 imputed values in order to create a single complete data set, which is going to be applied in Chapters 4 and 9 and is available in Appendix D.

# 4. CLUSTER ANALYSIS

This chapter presents the procedures of the clustering process we carried out in order to group the Brazilian states in a number of clusters according to their background[1] information related to road safety. The expected benefit of this process is to compose a more feasible basis for knowledge transference from one (similar) state to another, improving the national road safety situation by means of a benchmarking exercise. The cluster division established in the current chapter is maintained during the entire research, enabling the different parts of the research to be compared.

## 4.1. THE ROLE OF CLUSTERING

The objective of a cluster analysis is to discover a categorical structure that fits a certain set of observations. In other words, the intention is to sort the observations into groups such that the degree of "natural association" is high among members of the same group and, conversely, low for members from different groups (Anderberg, 1973). In hierarchical clustering, the data are not partitioned into classes in a single step; instead they are first separated into a few broad classes, each of which is further split into minor classes and so on, until final classes are generated. The researcher needs to decide at which stage of the segmentation it is (more) convenient to stop (Everitt, 1980).

The advantage of clustering within the road safety framework is to gain insight in rather homogeneous or similar stats and next, to offer a more feasible basis for the transference of good experiences from best to underperforming states. Consequently, in this study the clustering procedure is centered on the premise that efficacious actions to improve road safety in a certain state are more likely to generate desirable results in states with a similar background. This is particularly important in a country with continental dimensions and very contrasting backgrounds in the different regions. These contrasts refer to many aspects, from which the most obvious concern socioeconomic issues.

[1] We use the term "background" in this sentence with a more general meaning, this is, not exclusively referring to the named "background indicators".

In addition to the large geographical separation between some Brazilian states, strong disparity levels concerning the mentioned background aspects may make the transference of good experiences from best to underperforming states difficult or, at least, result in the reduction of the possibility of success on this transference. It would probably not be so realistic to suggest that the worst-performing state of Brazil should adopt the best-performing state as a benchmark (a state that should be considered as an example state to learn from).

The gap throughout the set of Brazilian states in terms of road safety performance is substantially related to the socioeconomic background of the states. In this context, national measures towards road safety are applied in distinct state environments, each one offering a specific background for implementation. It suggests that the same accident countermeasure tends to produce more distinct effects in states with very contrasting realities, e.g. an educational campaign might generate very positive results at one place but not that much at another place. Analogously, a successful measure implemented at a place, when reproduced at another place due to its successful prognostic, might deliver frustrating results if the background for application is very dissimilar and was not taken into account during this experience and expertise transference process.

In Brazil, in general terms, there exists a so-called geo-economic division in macro regions, based on historical characteristics and economic integration, as explained hereafter (Geiger, 1969):

- Amazon – characterized by large areas with low population density, it contains isolated areas with more intense human occupation (predominantly in the states capital cities);
- Northeast – characterized by the contrast between the more urbanized and touristic coastal areas and the interior with low population density due to the semi-arid climate;
- Center-South – constituted by the Southern, Southeastern and Center-Western most integrated regions; this is the most urbanized and developed area of the country.

In parallel, there was also a regional division, coinciding with the states borderlines, established by the IBGE (Brazilian Institute of Geography and Statistics) one year later in 1970 (IBGE, 1970), which is the most commonly used nowadays. These five geographical regions of Brazil are illustrated in Figure 4.1, and the dashed black lines are a representation of the geo-economic division in macro regions described in the previous paragraph.

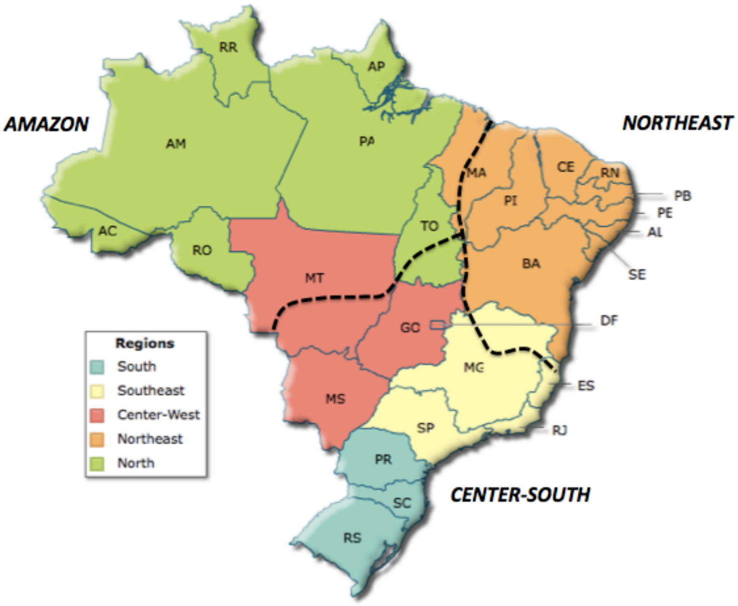



**Figure 4.1** – BR-27 regional and geo-economic division (dashed black line)

Adopting this division into 5 geographical regions, the southern and southeastern states present more favorable indicators related to most of the socioeconomic aspects in comparison with the northeastern and northern states, which usually present the lowest ones. The center-western states tend to present values closer to the national average (except the Federal District).

## 4.2. CLUSTERING PREMISSES

Previously to the clustering process itself, the main decision is to establish the aspects to base the clusters on. In this step, there is the need to find a cluster division suitable for the three main parts of this research: diagnosing the road safety situation, setting the target number of fatalities and suggesting guidelines to improve road safety (through the safety performance indicators research). The parameters used for clustering the states must present a relationship with the problem under investigation and must substantially influence the road safety situation of a place. In other words, variation in the chosen parameter should produce to some extent a change in the road safety situation. Moreover, it is desirable that the chosen parameters are available for every Brazilian state.

Based on these considerations, the selected aspects manifest some of the most relevant factors capable to distinguish contrasting backgrounds concerning the road safety situation: road user behavior, road infrastructure, health system, vehicle fleet characteristics, background (socioeconomic, demographical and geographical), modal distribution of the exposure and outcome based road safety indicators. A set of indicators composes each topic. The consideration of this kind of items for clustering is beneficial to increase the similarities inside each cluster – in a contribution to further establishing more attainable benchmarks.

We selected a set of 44 parameters (3 outcome related indicators plus the 36 presented in Chapter 3 plus 4 additional parameters presented in Table 4.1 plus one geographical proximity related parameter[2]) in the attempt to reach an adequate cluster configuration for Brazil. The data used represent the average of the annual values in the period 2010-2012. The complete data tables regarding each parameter per Brazilian state are available in Appendix C.

**Table 4.1** – Modal distribution of the exposure related parameters used for clustering BR-27

| 2nd level topic | Parameter | Min | Max | Unit | Data source |
|---|---|---|---|---|---|
| Motorcycle exposure | Share of motorcycle traveled kilometers in the total exposure | 13.35 | 55.05 | % | Author's research* |
| Car exposure | Share of car traveled kilometers in the total exposure | 30.20 | 73.92 | % | Author's research* |
| Truck exposure | Share of truck traveled kilometers in the total exposure | 5.28 | 25.06 | % | Author's research* |
| Bus exposure | Share of bus traveled kilometers in the total exposure | 3.48 | 11.62 | % | Author's research* |

* See Chapter 2 and Appendix B

After defining the variables to be considered and the data structure (this is, which data is linked to the same topic, e.g., two alcohol consumption indicators), some data processing occurred. First, by means of multiple imputation (see Section 3.6) we obtained a complete dataset (see Appendix D). Secondly, we normalized the data to render them in the same direction. The "distance to a reference method" is the procedure for normalization that we applied. This procedure is described in the Handbook of composite indicators (OECD, 2008). For this thesis' practical purpose, it means that each indicator value was normalized through

[2] Regional parameter: 0 for Northeastern states; 0.25 for Northern states; 0.50 for Center-Western states; 0.75 for Southeastern states; and 1.00 for Southern states.

dividing its value by the maximum indicator value[3]. Therefore, the state with the highest value will have a normalized indicator value equal to 1.00, and all the others will have lower values.

Thirdly, given the detailed structure of the indicators we averaged the normalized indicator values per category in order to merge the values into 23 more general categories. After that, we carried out the normalization procedure out again. These steps were necessary to balance the influence of the parameters in the clusters construction process, e.g. since the signing aspect contains many individual parameters, it would be unreasonable to directly compare these specific parameters with more general ones representing other aspects (e.g. GDP).

The analysis of dendrograms enables the constitution of the clusters. The dendrogram shown in Figure 4.2 was constructed using the software R® (R Core Team, 2014).

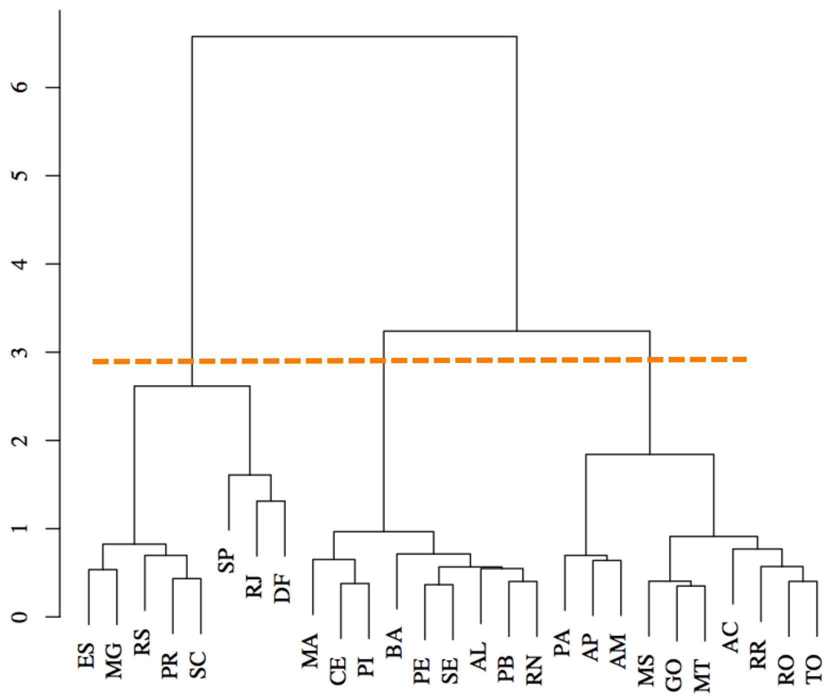



**Figure 4.2** – Hierarchical clusters dendrogram for BR-27

[3] When higher values represent a worse road safety situation, we divide the minimum value by the indicator value.

Dendrograms are two-dimensional diagrams represented by a inverted tree structure including the illustration of the junctions or divisions. We established these divisions according to the Ward's method, in which clusters are joined through the principle of maximizing the likelihood at each hierarchical level based on the Euclidean distance (Everitt, 1980).

According to the presented dendrogram, there are a few options for defining the clusters division. Here, from a benchmarking point of view, we decided to have clusters with a similar number of states, resulting in the following three clusters (see map in Figure 4.3):

- Cluster 1 – ES, MG, RJ, SP, PR, RS, SC and DF;
- Cluster 2 – AC, AP, AM, AP, PA, RO, RR, TO, GO, MS and MT;
- Cluster 3 – AL, BA, CE, MA, PB, PE, PI, RN and SE.

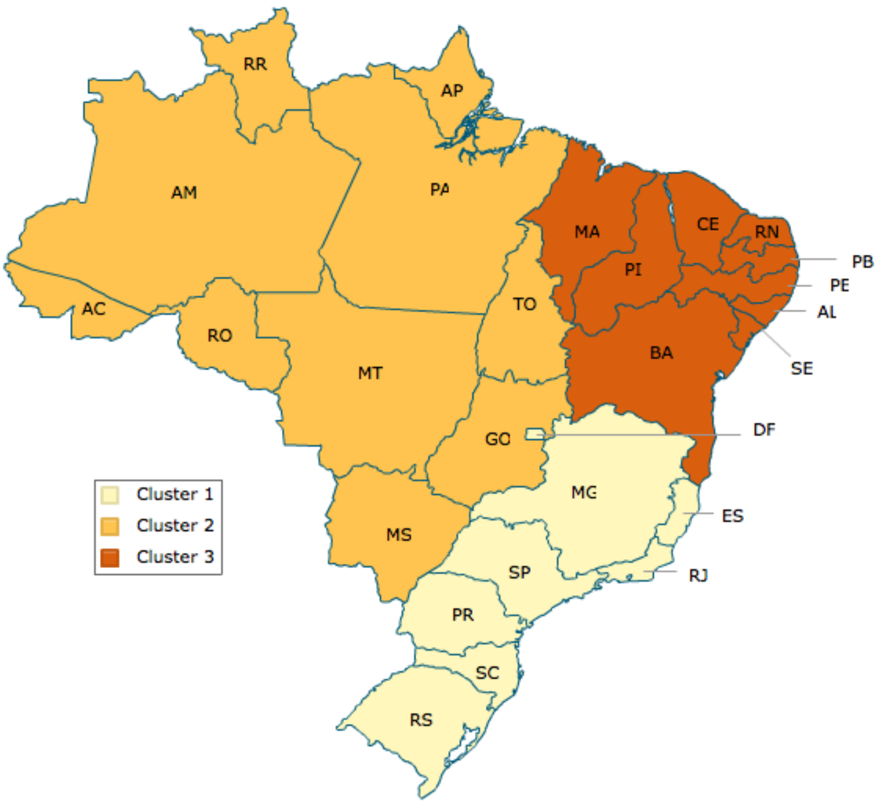



**Figure 4.3** –BR-27 clusters map

Southern and Southeastern regions plus the Distrito Federal (DF) form Cluster 1. The states containing the highest development levels are grouped in this cluster (see Table 4.2). All values have already been normalized to be interpreted in the same way; i.e. the higher the value, the better in terms of road safety.

**Table 4.2** – Normalized values of the parameters used for clustering the states

| Grouped parameter | Average normalized value | | |
|---|---|---|---|
| | **Cluster 1** | **Cluster 2** | **Cluster 3** |
| Alcohol | 0.80500 | 0.70423 | 0.64582 |
| Cell-phone | 0.46677 | 0.26404 | 0.24407 |
| Helmet | 0.89225 | 0.92597 | 0.83363 |
| Seatbelt | 0.83620 | 0.86344 | 0.78757 |
| Speeding | 0.34162 | 0.18698 | 0.20678 |
| Signing | 0.87624 | 0.63168 | 0.80007 |
| Roadside | 0.62754 | 0.21661 | 0.33726 |
| Health professionals | 0.67245 | 0.40631 | 0.33156 |
| Health resources | 0.89361 | 0.82264 | 0.75116 |
| Fleet composition | 0.53543 | 0.35151 | 0.38119 |
| Age of the fleet | 0.53013 | 0.73423 | 0.74482 |
| GDP per capita | 0.47940 | 0.25070 | 0.15794 |
| Motorization | 0.79944 | 0.51312 | 0.34761 |
| Mortality Rate | 0.58524 | 0.54551 | 0.56633 |
| Fatality Rate 1 | 0.63497 | 0.33745 | 0.26858 |
| Fatality Rate 2 | 0.66864 | 0.52144 | 0.36413 |
| Highway density | 0.28258 | 0.02279 | 0.11689 |
| Urban population | 0.91684 | 0.83060 | 0.76089 |
| Region | 0.81250 | 0.32500 | 0.00000 |
| Automobile exposure | 0.78036 | 0.54969 | 0.56992 |
| Motorcycle exposure | 0.58994 | 0.41444 | 0.31904 |
| Truck exposure | 0.55583 | 0.26893 | 0.69340 |
| Bus exposure | 0.61230 | 0.50451 | 0.60069 |

Central-Western and the Northern states constitute Cluster 2. Both regions are characterized by lower demographic densities. In a very summarized description, the Central-Western region has intense agriculture and livestock activities; while the Northern region is strongly influenced by the presence of the Amazon Rainforest.

Cluster 3 is equivalent to the Northeastern region, where the states are mostly characterized by the contrast between the coastal zone (where most part of the population and economic activities are concentrated) and the interior (with historical social problems related to the semi-arid climate).

# 5. DATA ENVELOPMENT ANALYSIS

This chapter introduces the Data Envelopment Analysis (DEA) technique, which constitutes the methodological core of this research. Firstly, we present the general principles of DEA and then the original model translation into the road safety perspective. Lastly, there is the introduction of the procedure to analyze the sensitivity of the DEA scores, the bootstrapping. For this issue, we developed a method applicable for every step of this research, enabling the sensitivity assessment along many of the research steps.

## 5.1. PRINCIPLES OF DEA

Data envelopment analysis or DEA is a term used to designate a “data oriented” approach in which mathematical programming methods are applied to handle large numbers of variables and relations (multiple input and multiple output – MIMO problems); that way, it has become an attractive tool to deal with complex problems. The first concepts on this theme emerged in 1957 for measuring the productive efficiency of industries (Farrell, 1957), although its most widely known basic form, the CCR model, was proposed by Charnes, Cooper and Rhodes in 1978 on the evaluation of production processes through exploring the relation between the amount of inputs and outputs. The entities under study, responsible to convert inputs into outputs, are named decision-making units – DMUs (Cooper et al., 2011, 2000).

The efficiency of a DMU (i.e. a firm, state or country) usually varies between 0 and 1, being that the closer to 0, the more inefficient the DMU is (bad performing) and the closer to 1, the more efficient (good performing) the DMU is. In other words, a DMU with a score equal to 1 is capable to optimally/efficiently convert all its inputs into outputs (output/input ratio is equal to one); in contrast, an inefficient DMU does not succeed in optimally/efficiently converting all its inputs into outputs and its output/input ratio does not reach one. Table 5.1 and Figure 5.1 contain a numerical and a graphical example, respectively, about the basic concept of the DEA technique, inspired on Shen (2012).

Suppose five DMUs – A, B, C, D and E (which can be five different companies producing the same product) each one responsible to transform a certain amount of inputs (e.g. raw material) into outputs (e.g. final product). The DMU capable to produce more outputs with the same amount of input is then the most efficient one of the set. According to the example in Table 5.1, DMU D is the most efficient one, scoring 1.00. The rest of the DMUs are underperforming units, and their efficiency can be expressed in relation to the most efficient DMU; this is, DMU A presents 40% of the efficiency of DMU D (being the worst performing of the five units), B presents 50%, C presents 56% and E presents 75%.

**Table 5.1** – Numerical example to illustrate the basic concept of DEA

| DMU | Input | Output | Efficiency score |
|---|---|---|---|
| A | 5.00 | 2.00 | 0.40 |
| B | 8.00 | 4.00 | 0.50 |
| C | 9.00 | 5.00 | 0.56 |
| D | 6.00 | 6.00 | 1.00 |
| E | 4.00 | 3.00 | 0.75 |

Now the input (horizontal axis) and output values (vertical axis) of the five DMUs are plotted in Figure 5.1.

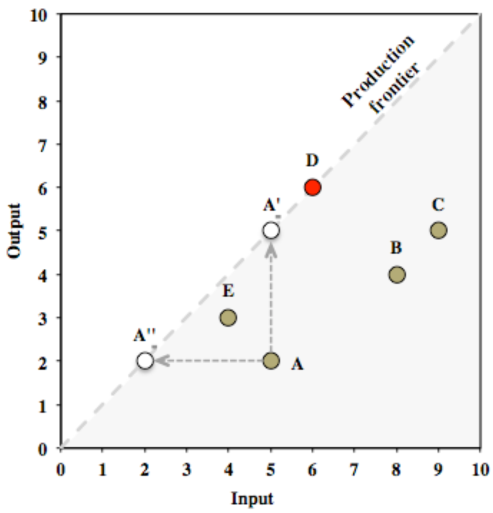



**Figure 5.1** – Graphical representation of the efficiency production frontier

The slope of the line connecting each point to the origin indicates the efficiency score of the corresponding DMU; and the line connecting DMU D to the origin represents the efficiency production frontier, since no other DMU can be more efficient than DMU D. The area between the production frontier and the horizontal axis is called production possibility set, and all the DMUs are "enveloped" within this set, hence the term data envelopment analysis.

The underperforming DMUs might achieve the best attainable performance (similar to DMU D) if they are able to increase their efficiency. For example, DMU A could be as efficient as DMU D by increasing its output value maintaining its input amount unchanged (A' point) or by decreasing its input amount without changing the output value (A" point). Therefore, the production frontier constructed based on DMU D's is valuable within a benchmarking process, which is an important tool aimed at promoting the transference of good practices from high to low performing DMUs, as it can be deduced from which DMUs a particular underperforming DMU can learn.

For simple examples as the one presented above, this graphical representation is a useful visual description of the fundamental mechanisms of DEA, yet difficult to reproduce when more parameters are considered in the investigation of the DMUs efficiency. The usual approach is to employ a linear programming formulation for DEA to estimate the efficient production frontier and compute the relative, optimum efficiency of each DMU. Equation 5.1 presents an input-oriented model, in which the outputs must be maximized in order to obtain the highest possible efficiency of each DMU under investigation.

$$OES_s = \max \sum_{i=1}^{p} w_{O_i} O_{i,s} \tag{5.1}$$

$$\text{subject to} \quad \sum_{j=1}^{q} w_{I_j} I_{j,s} = 1,$$

$$\sum_{i=1}^{p} w_{O_i} O_{i,r} - \sum_{j=1}^{q} w_{I_j} I_{j,r} \leq 0 \qquad r = 1,\ldots,n$$

$$w_{O_i} \geq 0 \quad i = 1,\ldots,p \qquad w_{I_j} \geq 0 \quad j = 1,\ldots,q \quad s = 1,\ldots,n$$

In which: $OES_s$ – Optimum efficiency score of the DMU under investigation; $O_{i,s}$ – *i-th* output of the *s-th* DMU; $w_{O_i}$ – Weight attributed to the output $O_i$; $I_{j,s}$ – *j-th* input of the *s-th* DMU; $w_{I_i}$ – Weight attributed to the input $I_j$; $n$ – total number of DMUs; $p$ – total number of outputs; $q$ – total number of inputs.

Additionally, we have to make a distinction between the concept of optimum efficiency score (OES) and cross- efficiency score (CES). The first identifies the highest possible score for a DMU under the imposed restrictions; however, the flexibility in selecting the most favorable weights for each DMU forbids the comparison of the efficiency score between various DMUs on a common basis. Therefore, it is recommended to use the CES, which is the average value of the product between each DMU's input and output and not only its own attributed weights, but also all the other DMUs' optimum weights (Doyle and Green, 1994; Sexton et al., 1986). Equation 5.2 enables the computation of the CES value (used for a direct comparison).

$$CES_s = \left(1/n\right)\sum_{r=1}^{n}\left[\sum_{i=1}^{p} w_{O_{i,r}} O_{i,s} / \sum_{j=1}^{q} w_{I_{j,r}} I_{j,s}\right] \tag{5.2}$$

In which: $CES_s$ – Cross efficiency score of DMUs under study

The attempt of translating a practical road safety problem into a mathematical model requires to orientate the model so that reasonable results are delivered. Mathematical models itself are not able to identify and adequately treat external factors that might influence the model results and their interpretation. The original model is quite general, allowing the free allocation of weights to reach an optimum solution. Without the appropriate procedures to render the DEA model adequate for the reality of the issue under investigation, there is a chance of obtaining completely useless results and drawing wrong conclusions. In other words, for the application of DEA in specific fields (e.g. road safety), it is plausible that researchers insert adequate conditions to fit the model to their field's needs.

## 5.2. ROAD SAFETY DEA APPLICATION

DEA models are applicable to a variety of fields inside and outside the road safety scope. Until today, many additional models have been developed to support a variety of practical problems encountered since the introduction of the technique (e.g. Adler et al., 2002), including the adaptation of the DEA principles presented in the previous section to the road safety field research. In the road safety framework, this technique has been applied for input-output data sets (e.g. Hermans et al., 2009; Shen et al., 2011) and composite indicator research (e.g. Bax et al., 2012; Shen et al., 2010); both types of these DEA model extensions will be applied in this thesis and described in its corresponding upcoming chapters. Their use on non-European data is innovative, also implying some methodological challenges that needed to be met.

We adapted the example presented in Section 5.1 with the intention of illustrating the applicability of DEA technique in the road safety field. Suppose the same five DMUs – A, B, C, D and E, which now, instead of five different companies producing the same product, represent five different geographical units (cities, states or countries) where a certain amount of exposure (this input is manifested for example as the amount of kilometers traveled) results in road safety output (e.g. traffic fatalities) – see Table 5.2.

**Table 5.2** – Numerical example to illustrate the concept of DEA applied for road safety research

| DMU | Exposure | Traffic fatalities | Efficiency score (road safety risk*) |
|---|---|---|---|
| A | 0.80 | 2.00 | 2.50 |
| B | 2.00 | 4.00 | 2.00 |
| C | 2.78 | 5.00 | 1.80 |
| D | 6.00 | 6.00 | 1.00 |
| E | 2.25 | 3.00 | 1.33 |

*The lower the better

The score indicating the best-performer(s) still presents a value equal to 1 (as in the original model); this is, they succeed best in minimizing their sustained risk level. According to the information in Table 5.2, DMU D is most efficient, scoring 1.00. However, the goal in this case of road safety is not to produce more outputs, since traffic fatalities are undesirable outcomes; therefore, the DMU capable to produce less outputs (traffic fatalities) with the same amount of input (exposure) is the most efficient one. This efficiency state implies a minimization of the traffic fatality risk level, and instead of an efficient production frontier, there now is an efficient road safety frontier.

All the other DMUs are underperforming units, and their efficiency can still be expressed in relation to the most efficient DMU, but now their inefficiency is shown by values higher than 1, indicating they did not succeed in minimizing their traffic fatality risk. In particular, DMU A presents 40% (1.00/2.50) of the efficiency of DMU D, B presents 50% (1.00/2.00), C presents 56% (1.00/1.80) and E presents 75% (1.00/1.33). Analogously, we can also say that 2.50 times more risk than DMU D, DMU B 2.00 times, DMU C 1.80 times and DMU E 1.33 times.

The values representing the exposure  (vertical axis) and the traffic fatalities (horizontal axis) of the five DMUs are plotted in Figure 5.2.

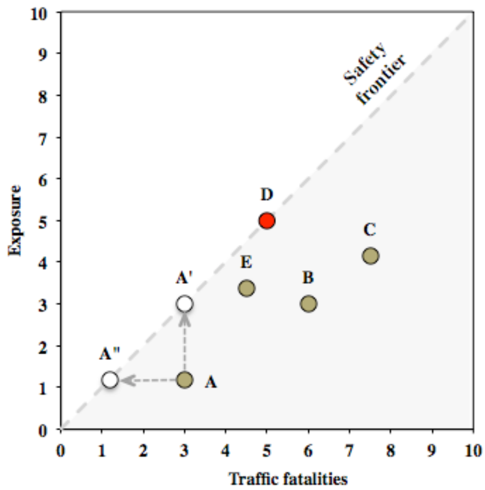



**Figure 5.2** – Graphical representation of the safety frontier

The slope of the line connecting each point to the origin indicates the road safety frontier of the corresponding DMU; and the line connecting DMU D to the origin represents the overall road safety frontier, since no other DMU obtains a better road safety level than DMU D does. All DMUs are now enveloped in a “road safety possibility set” defined by the area between the safety frontier and the horizontal axis.

Analogously to the original model, the underperforming DMUs can achieve the best attainable performance (similar as DMU D) if they are able to decrease their risk score. For example, DMU A could be as safe as DMU D by increasing its input value (exposure) preserving its output amount (traffic fatalities) unchanged (A’ point); or by decreasing its output value (traffic fatalities) maintaining its input amount (exposure) unchanged (A” point). Hence, the safety frontier constructed based on DMU D’s performance is a useful tool in terms of a benchmarking process, in which benchmark DMUs can be identified, the size of inefficiency quantified and after more detailed analysis, directions for improvement in terms of good practice road safety measures can be learned from a better-performing DMU.

## 5.3. DEA SCORES SENSITIVITY ANALYSIS

DEA's attractiveness of no functional form requirements for input information contributed to its dissemination and application in various fields; however, there is still some discussion about the statistical properties of the produced scores. The obtained optimum scores might represent a very particular combination of parameters (e.g. indicators) and weights that may be unlikely to be found in realistic circumstances (requiring the imposition of weight restrictions).

In this context, the general criticism relies on the deterministic nature of DEA estimators, centered on the argument that the technique does not account for uncertainties in the data, being merely a point estimator from which no statistical inference can be derived. When running a DEA model, a production frontier based on the provided data is constructed and is used as a reference to compute the set of scores; thus, it is not difficult to conclude that the configuration of such a frontier is susceptible to variations in the information (Löthgren, 1998; Simar and Wilson, 1998; Walden, 2006).

In addition to these issues, the sensitivity analysis has also the purpose of simulating different data sceneries, since the exposure and fatality data we use is subjected to some level of uncertainty. That is actually the main motivation for testing the sensitivity.

The suggested procedure to tackle these inherent vulnerabilities and to test the sensitivity of the obtained DEA scores to sampling variations is to bootstrap the DEA estimators (Aliev and Ebadi, 2012; Simar and Wilson, 2000). The key idea of bootstrapping, firstly addressed by Efron in 1979, is to resample from the original data to create replicate datasets, which mimic the original unknown sampling distribution of the estimates of interest; in other words, an empirical distribution is artificially constructed. The principle is as follows: given a specified random variable $\theta$ $(y,f)$, depending on both $y$ and the unknown distribution $f$, estimate the sampling distribution of $\theta$ on the basis of the observed data $y$ (Chernick, 1999; Davison and Hinkley, 2003; Efron, 1992). The steps to bootstrap the DEA scores are summarized in the flowchart of Figure 5.3.

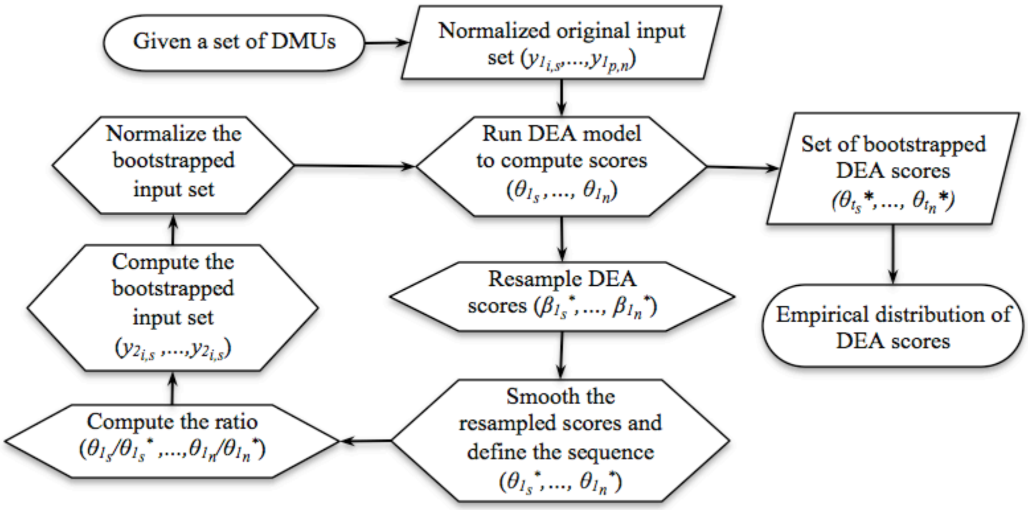



**Figure 5.3** – Flowchart containing the steps of the bootstrapping procedure

Given a set of $n$ DMUs and their corresponding $p$ parameter values $(y_{1i,s},\ldots,y_{1p,n})$, the first step is to normalize these values. Then, the DEA model is run in order to compute the original efficiency scores $(\theta_{1s},\ldots,\theta_{1n})$. These scores are resampled with replacement according to the empirical distribution, generating the set $\beta_{1s}^*,\ldots,\beta_{1n}^*$ and, afterwards, they should be smoothed, forming the set $\theta_{1s}^*,\ldots,\theta_{1n}^*$. Smoothening is the designation of the process of incorporating the properties of the original set of efficiency scores into the resampled set, through the application of reflection methods as suggested in Simar and Wilson (1998) and Walden (2006), represented by the application of Equations 5.3 to 5.6.

$$h = \left[\frac{4}{(p+q+2)}\right]^{[1/(p+q+4)]} * N^{[-1/(p+q+4)]} \tag{5.3}$$

$$\tilde{\theta}_i^* = \begin{cases} \beta_i^* + h.\varepsilon_i^* & \text{if } \beta_i^* + h.\varepsilon_i^* \geq 1 \\ 2 - \beta_i^* - h.\varepsilon_i^* & \text{otherwise} \end{cases} \tag{5.4}$$

$$\hat{\sigma}_{\theta}^2 = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(\hat{\theta}_i - \hat{\bar{\theta}}\right)^2 \tag{5.5}$$

$$\theta_i^* = \bar{\beta}^* + \frac{\left(\tilde{\theta}_i^* - \bar{\beta}^*\right)}{\sqrt{\left(1 + \dfrac{h^2}{\hat{\sigma}_{\hat{\theta}}^2}\right)}} \tag{5.6}$$

In which: $h$ is the smoothing parameter defined by the number of indicators $p$, the number of outputs $q$ and the number of DMUs $N$ presenting the set of original optimum scores $(\hat{\theta}_{1s},...,\hat{\theta}_{1n})$; $\varepsilon_i^*$ is a random number drawn from a standard normal distribution; $\tilde{\theta}_i^*$ is the smoothed sampled score $(\tilde{\theta}_{1s}^*,...,\tilde{\theta}_{1n}^*)$; $\hat{\sigma}_\theta^2$ is the plug-in estimator of the variance of the original scores $(\hat{\theta}_{1s},...,\hat{\theta}_{1n})$; $\beta_i^*$ is the average of the resampled score $(\beta_{1s}^*,...,\beta_{1n}^*)$; and $\theta_i^*$ is the variance corrected bootstrapped score $(\theta_{1s}^*,...,\theta_{1n}^*)$.

The next step is to compute the ratio between the original and the bootstrapped scores $(\theta_{1s}/\theta_{1s}^*,\ldots, \theta_{1n}/\theta_{1n}^*)$ and then multiply it by the original input set, producing the new input set $y_{2i,s}, \ldots, y_{2p,n}$, which is normalized prior to computing the second set of efficiency scores. By running the DEA model again, the first loop is done. This should be repeated $t$ times until an adequate set of bootstrapped scores $\theta_{ts}^*, \ldots, \theta_{tn}^*$ is obtained. In this research, we adopted $t=50$.

The steps of the bootstrapping procedure that we applied in this research are performed in different softwares, since there was no complete package capable to satisfy the research needs. Lingo® is the software that runs the DEA models, importing the DMU data from an Excel® spreadsheet; then Lingo® exports the model results to another Excel® spreadsheet, which is read by R® in order to generate the bootstrapped new input[1] data set read by Lingo® in the next iteration. Figure 5.4 is a representation of this iterative process including the use of different softwares.

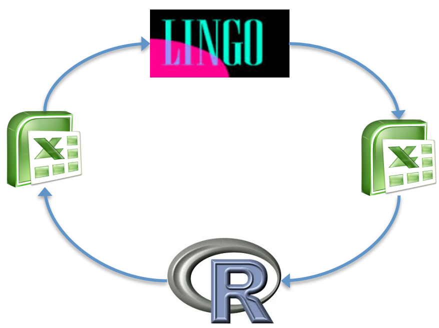



**Figure 5.4** – Software interface for the execution of the bootstrapping procedure

A complete bootstrapping iteration consists of the sequential steps described in the hexagonal forms and it produces a new input set intended to resemble the original distribution each time

[1] We use the word "input" here in a broader sense than when we mention this term in the input-output relationship.

it is executed; some degree of deviation from the seed values is simulated in order to express the uncertainties associated to the input data and to the probability of reaching the optimum computed score. As a consequence, the variety of created input sets induce the DEA algorithms to conceive different production frontiers, which is valuable to test the sensitivity of the results to changes in the model boundaries.

# 6. DIAGNOSING THE ROAD SAFETY SITUATION

In order to contribute to road safety diagnosis in Brazil, this chapter presents a research based on two main indicators available per state: the mortality rate (represented by fatalities *per capita*) and the fatality rate (represented by two sub-indicators, i.e., fatalities per vehicle and fatalities per vehicle kilometer traveled - VKT). The approach we describe here consists of aggregating these indicators into a composite indicator or index through a multiple layer data envelopment analysis composite indicator model (ML DEA-CI), which looks for the optimum combination of indicators' weights for each decision-making unit, in this case 27 Brazilian states. The index score represents the road safety outcome performance, based on which a ranking of states can be made. Since such a model has never been applied for road safety evaluation in Brazil, it was necessary to calibrate its parameters based on the experience of more consolidated European Union research in ranking its member countries using DEA techniques. Secondly, cluster analysis provided more realistic performance comparisons. Finally, we measured the sensitivity of the results through a bootstrapping method application.

## 6.1. BUILDING A COMPOSITE INDICATOR USING DEA

Decision makers must carefully consider both theoretical recommendations and data availability restrictions when performing a road safety evaluation based on a mortality or fatality rate; otherwise erroneous or at least inaccurate conclusions may arise. Subsequently, research into strategies capable of delivering an overall picture of the situation through the combination of indicators (computed in relation to different exposure measures), rather than focusing on just one indicator, is a valuable task. In this sense, the DEA technique is a promising alternative tool to manage the indicator choice dilemma.

The translation of the original DEA model (described in Chapter 5) to a composite indicator model implies that the latter typically focuses on the achievements without taking into account the input-side (Cherchye et al., 2006). In the road safety framework, this technique has been applied in, for example, Bax et al. (2012) and Shen et al. (2010). Equation 6.1 presents a converted form of the original DEA based composite indicator model as already applied by Cherchye et al. (2006), in which each particular indicator is referred to as an output to be minimized, because in road safety research fatalities are undesirable outputs. An optimum index score equal to 1 indicates the best-performer(s) (as in the original model). The difference now is that underperforming DMUs present a score larger than 1, since they did not succeed in minimizing their undesired outcome values to the best possible extent.

$$
\begin{aligned}
OIS_s &= \min \sum_{i=1}^{p} w_{i,s} y_{i,s} \\
\text{subject to} \quad & \sum_{i=1}^{p} w_{i,s} y_{i,r} \geq 1, \quad r = 1,\ldots,n \\
& w_{i,s} \geq 0, \quad i = 1,\ldots,p
\end{aligned}
\tag{6.1}
$$

$OIS_s$ is the optimum index (or composite indicator – CI) score of the *s-th* DMU, $y_{i,s}$ is the *i-th* indicator of the *s-th* DMU, $w_{i,s}$ is the weight attributed to indicator $i$, $n$ is the total number of DMUs and $p$ the total number of indicators. Thus, given the 3 indicators that are considered in this study, which are… the model estimates only 3 parameters (weights); however, since the model is run for each of the 27 Brazilian states, we have totally $n*p = 27*3 = 81$ weights that need to be estimated

By simply treating all the indicators as belonging to the same layer, the information about the hierarchical structure of the indicators is obviously ignored, which further leads to weak discriminating power and unrealistic weight allocations (Shen et al., 2011). The need to express a hierarchy in the set of selected indicators motivated the development of the multiple-layer DEA-based composite indicator model (ML DEA-CI) by Shen et al. (2013a). By solving Equation 6.2, the composite indicator based on a $K$-layered hierarchy of $p$ indicators can be calculated for each state $s$, where $u_{fK}$ is the weight given to the *f-th* category in the *K-th* layer and $w_{fK}^{(K)}$ denotes the non-negative internal weights associated with the indicators of the *f-th* category in the *K-th* layer; the sum of all $w_{fK}^{(K)}$ within a particular category is equal to one.

$$
OIS_s = \min \sum_{f_{K_1}=1}^{p^{(K)}} u_{f_K} \Big( \sum_{f_{K-1} \in A_K^{(K)}} w_{K-1}^{(K-1)} \Big( \ldots \sum_{f_k \in A_{k+1}^{(k+1)}} w_k^{(k)} \Big( \ldots \sum_{f_{k-1} \in A_{f_3}^{(3)}} w_2^{(2)} \Big( \sum_{f_1 \in A_{f_2}^{(2)}} w_{f_1}^{(1)} y_{f_1 s} \Big)\Big)\Big)\Big)
\tag{6.2}
$$

The total number of weights estimated in Equation 6.2 is equal to the number of parameters estimated in Equation 6.1 (one weight for each parameter). The weights corresponding to the secondary layers of the hierarchical structure are actually a result of the combination of each indicator's weight.

Finally, there is still the need of distinguishing the optimum index score (OIS) and the cross-index score (CIS). The OIS identifies the best-performing DMUs (those with a score equal to one) among all other assessed DMUs; however, the flexibility in selecting the most favorable weights for each DMU forbids the comparison of all the DMUs on a common basis. The CIS is obtained from the OIS, using the average value of the product between each indicator of a certain DMU and not only its own attributed weights, but also all the other DMUs' weights (Doyle and Green, 1994; Sexton et al., 1986). Equation 6.3 enables the computation of the CIS for a direct comparison.

$$CIS_s = \left(1/n\right)\sum_{r=1}^{n}\sum_{i=1}^{p}(w_{i,r}y_{i,s}) \qquad s = 1,\ldots,n \tag{6.3}$$

$CIS_s$ is the cross-index score of the s-th DMU, $y_{i,s}$ is the *i-th* indicator of the *s-th* DMU for which *CIS* is to be computed, $w_{i,r}$ is the weight of the *i-th* indicator of the *r-th* DMU, *n* is the total number of DMUs and *p* is the total number of indicators.

Although it seems that the DEA model is complicated and complex because of the number of parameters (weights) it estimates (in this case *3*27 = 81*), actually the total number of parameters results from the estimation of individual parameters for each DMU, instead of considering the same weights distribution for the entire set of DMUs. Therefore, it has the advantage of taking into account the individual characteristics of each DMU.

Hence, it means that the DEA results include different perspectives in the performance evaluation. Furthermore, the possibility of taking a hierarchical structure of indicators into account is very appealing.

### 6.1.1 MODEL CALIBRATION

Despite the fact that an identical model has not been applied to European data yet, several other similar models have already been explored; for more information see e.g. Shen (2012) and Shen et al. (2013b). The more mature experience in this study field related to the set of European Union countries, named here as EU-27, might be used as a reference to evaluate and better investigate the first application on Brazilian data.

The 27 member states of the EU are: Austria (AT), Belgium (BE), Bulgaria (BG), Cyprus (CY), Czech Republic (CZ), Denmark (DK), Estonia (EE), Finland (FI), France (FR), Germany (DE), Greece (EL), Hungary (HU), Ireland (IE), Italy (IT), Latvia (LV), Lithuania (LT), Luxembourg (LU), Malta (MT), the Netherlands (NL), Romania (RO), Poland (PL), Portugal (PT), Slovakia (SK), Slovenia (SI), Spain (ES), Sweden (SE) and the United Kingdom (UK).

The intention of this calibration process is to compare the DEA model results for the set of EU countries and the DEA results for the set of 27 Brazilian states. The data for the calibration process corresponds to two equivalent data sets collected for EU-27 and BR-27, in which the three considered indicators are: mortality rate, represented by fatalities per 100 thousand inhabitants (MR); and two fatality rates, represented by fatalities per 10 thousand vehicles (FR1) and fatalities per billion passenger kilometers traveled (FR2). The calibration is based on the average values for the period 2009-2011.

In order to facilitate the visualization of the collected data, Figures 6.1 and 6.2 contain a radar chart using the normalized values of the three indicators respectively for EU-27 and BR-27. The exact values of both data sets are available in Appendix E. In the radar charts, the distinct patterns in the indicators relationship are very clear.

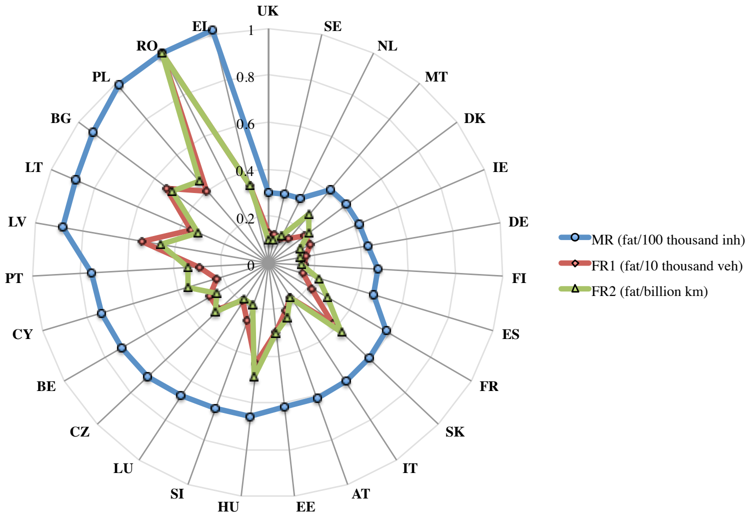



**Figure 6.1** – Radar chart for the normalized values of the EU-27 indicators

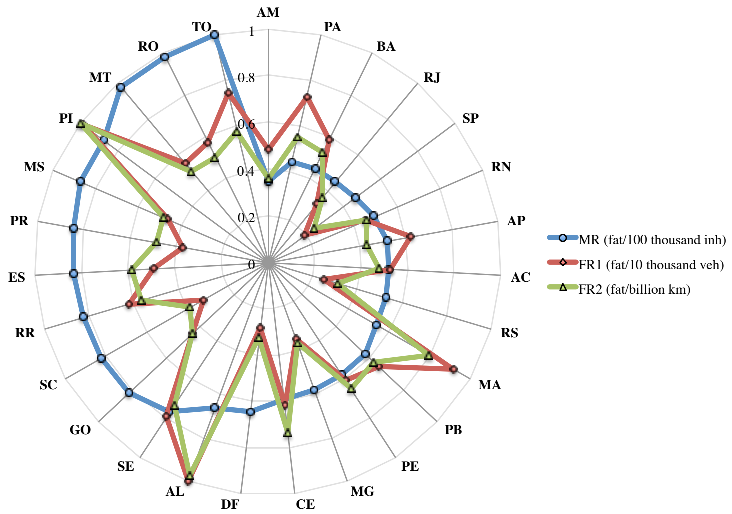



**Figure 6.2** – Radar chart for the normalized values of the BR-27 indicators

The MR and FR1 values for EU-27 are directly comparable with the Brazilian data. Contrarily, the FR2 value is not directly comparable since it has passenger kilometers in its denominator; i.e. occupancy rates were considered in the indicator construction for the EU-27; for Brazil, this information is unknown. In spite of this differentiation between both data sets, the assumption that the ranking positions are preserved is considered.

In the EU-27, the Spearman's rank correlation coefficient at the 95% confidence level (an adequate correlation measure for not normally distributed datasets) between the three ranks is quite high: 0.84 between MR and FR1; 0.79 between MR and FR2; and 0.91 between FR1 and FR2. Just as an example to confirm such a relationship, the first three countries according to the MR value also occupy the same range of positions in the FR1 and FR2 rankings.

In the BR-27, in a diverging tendency from the results based on the EU, according to Spearman's rank correlation coefficient, there is no statistically significant correlation between MR and FR1 or FR2 considering the 95% confidence level. Between FR1 and FR2 there is a correlation of 0.87.

An initial form of the ML DEA-CI model was applied to both data sets, the Brazilian and the European. Figure 6.3 presents the inserted hierarchical structure.

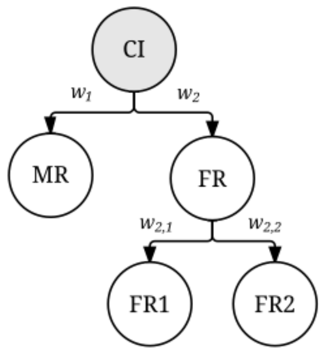



**Figure 6.3** – Hierarchical structure used to compute the (FO) composite indicator (CI)

This hierarchy implementation expresses the more similar nature of FR1 and FR2, since they bring the same information in different refinement levels regarding the risk exposure parameter. In Figure 6.3, the only weights estimated in the ML DEA-CI model are $w_1$, $w_2$ and $w_{2,1}$. Weight $w_{2,2}$, shown in Figure 5, is not estimated but is the result of $[w_{2,2} = (1 - w_{2,1})]$. The computation of the CI results from $[(w_1)(MR) + (w_{2,1})(FR1) + (w_{2,2})(FR2)]$.

The attempt to translate a practical road safety problem into a mathematical model requires the adaption of the model so that reasonable results are delivered; this is done through the insertion of the following commented weight restrictions. In order to avoid a unilateral weight distribution (this is, the index value being influenced exclusively by MR or FR), but still granting adequate flexibility to the model, the shares $w_1$*$MR$ and $w_2$*$FR$ were limited in the model definition to vary between 10 and 90% of the index value. Likewise, to control the weight distribution on FR and avoid an exaggerated weight attribution to FR1 or FR2, the weights $w_{2,1}$ and $w_{2,2}$ were allowed to vary within a 20% maximum range; this is, the weights should vary between 40% and 60% of the $w_2$ value.

The adoption of such a procedure relies on the assumption that there is no reason to attribute much more weight to any FR indicator, since FR1 and FR2 are supposed to show a high degree of association; i.e., theoretically they should show quite similar values and the composed indicator FR should tend towards the average of FR1 and FR2. Before running the model, the data provided in Tables 1 and 2 were normalized using the ‘distance to a reference’ method, in which each indicator’s values are divided by the maximum indicator value of the 27 states, based on recommendations of the OECD (2008).

Following the calibration process (explained hereafter), there was the insertion of additional weight restrictions between the shares $w_1*MR$ and $w_2*FR$, being that $w_2*FR$ must be larger than or equal to $1.5*w_1*MR$. The intention of this intervention is to reduce the importance attributed to a somewhat biased indicator (the MR). The 1.5 factor was established after testing various values: 1.0, 1.5 and 2.0 (differences in cross-index scores using these three values were minimal).

Two selected parameters intending to quantify the socioeconomic development level of a DMU guided the calibration stage: the GDP *per capita* (GDPC) and the motorization rate (MOTR). The GDPC is a key indicator which usage is consolidated in the economic field and its relation to road safety was already shown in e.g. Yannis et al. (2013). The motorization rate, defined as the ratio between the number of motor vehicles and the number of inhabitants, also reveals a lot about the social and economic development level of a country or state.

Mainly in developing countries, where the offer of a satisfactory public transport system is more unusual, the individual vehicle is an extremely desired consumption good, since it is seen as a social inclusion instrument. It is not a coincidence that the Brazilian “economic boom” is accompanied by an explosion in motorization in recent years. To support this process in Brazil, there is a clear governmental policy which encourages people to purchase their first vehicle by reducing the industrialized product taxes and smoothening credit access. The ratio between the amount of vehicles and the population expresses to which extent people are succeeding in satisfying this desire. For these reasons and also for the facility of obtaining it, the MOTR was taken as the second socioeconomic indicator to base the ranking evaluation on.

Table 6.1 exhibits a "before x after calibration" comparison for BR-27 and EU-27 in terms of the Pearson correlation coefficients between the CIS and two selected socioeconomic parameters (accompanied by their $p_{values}$ as an indication of the statistical significance level of the computed correlations considering a confidence interval of 95%, a correlation is considered statistically significant if its $p_{value} \leq 0.05$).

Before computing the correlation coefficient, we submitted the data sets to the Shapiro Wilk normality test and, considering a confidence interval of 95%, they are all considered as normally distributed, supporting the application of Pearson's correlation test. An outlier detection process preceded the computation of these values and aberrant values ($z_{score} > 3.0$) were found for the GDP *per capita* of DF (Distrito Federal) and LU (Luxembourg), and for the CIS of RO (Romania).

**Table 6.1** – Correlation coefficients between CIS, MOTR and GDPC

| Socioeconomic parameters | Before calibration | | After calibration | |
|---|---|---|---|---|
| | CIS BR-27 | CIS EU-27 | CIS BR-27 | CIS EU-27 |
| MOTR BR-27 | Corr = - 0.0201<br>$p_{value}$ = 0.9209 | - | Corr = - 0.5183<br>$p_{value}$ = 0.0056 | - |
| MOTR EU-27 | - | Corr = - 0.4869<br>$p_{value}$ = 0.0100 | - | Corr = - 0.5229[c]<br>$p_{value}$ = 0.0061[c] |
| GDPC BR-27 | Corr = - 0.3119[a]<br>$p_{value}$ = 0.1209[a] | - | Corr = - 0.7000[a]<br>$p_{value} = 6.88.10^{-5}$[a] | - |
| GDPC EU-27 | - | Corr = - 0.7557[b]<br>$p_{value} = 8.05.10^{-6}$[b] | - | Corr = - 0.7963[d]<br>$p_{value} = 1.93.10^{-6}$[d] |

[a]Excluding DF [b]Excluding LU [c]Excluding RO [d]Excluding LU and RO

The evidences of correlation between the CISs and the chosen available socioeconomic indicators are considerably stronger for the EU-27 than for BR-27 before calibration. These differences may be attributed to the distinguished effects in MR, FR1 and FR2 produced by contrasting motorization levels: 59 vehicles / 100 inhabitants (standard deviation equal to 14 vehicles / 100 inhabitants) in EU-27; and 38 vehicles / 100 inhabitants (standard deviation equal to 14 vehicles / 100 inhabitants) in BR-27.

The explanation for such comment is as follows. When the GDP *per capita* and, consequently, the motorization level are significantly low, as is the case for the rapidly increasing, but still moderate motorization levels of most Brazilian states (at least when compared to high-income road safety best-performing countries), a substantial bias can emerge on the "mortality x fatality rates" analysis: the fatalities per inhabitant rate (MR) tends to be directly proportional to the motorization rate, while the fatalities per vehicle or per vehicle kilometer traveled rate (FR1 or FR2) tends to be inversely proportional to the motorization rate.

As a consequence, in this motorization (and actually mobility development) stage, the adoption of MR as a road safety parameter to rank the road safety situation of countries/states provides very different rank configurations compared to the ones obtained based on FR1 or FR2. Therefore, quite low correlations between the cross-index score (CIS) and the listed socioeconomic parameters are expected, since depending on the indicator's weight combination resultant from the ML DEA-CI model, some states' CIS values might be predominantly influenced by MR, generating a biased rank regarding the correlation with socioeconomic parameters.

Theoretically, road safety becomes a more evident problem as motorization increases, until its impact reaches unacceptable levels and both society and government effectively react by implementing measures to improve road safety. At a certain stage, a breakdown point is reached and all three indicators (MR, FR1 and FR2) then tend to be inversely proportional to the motorization rate. Such a development stage is currently being experienced by most EU-27 countries, which have managed to reduce their MR, FR1 and FR2 in the last years. Thus, fairly high correlations between CIS and socioeconomic parameters are expected in EU-27, because whatever the predominant indicator forming the composite indicator is, they all point in the same direction regarding their association with the selected socioeconomic parameters.

Elvik and Vaa (2004), Jørgensen (2006) and Yannis et al. (2011) already stressed this switching behavior in the relation between mortality and fatality rates along different motorization stages. In practice, these particularities of the relationship between MR, FR1 and FR2 in relatively low-motorized countries demand that the analysis of results derived from a composite indicator constructed using both mortality and fatality rates must be carefully carried out in order not to fall in any pitfall offered by these indicators. In this context, the motorization rate was identified as a potential confounding parameter, affecting the relationship between mortality and fatality rates.

Consequently, the impact of the motorization level should be carefully controlled (through the mentioned weight restrictions of MR) and beforehand investigated in this type of study and conclusions must be drawn keeping in mind the socioeconomic context in which the decision-making units are situated; otherwise, the performance of some DMUs may be mistakenly overestimated.

After the calibration procedure, the main improvements with regard to the level of correlation were attained with the motorization rate (MOTR), which increased tremendously. Therefore, the produced effects on both sets of DMUs are desirable, and it justifies the inclusion of the described weight restrictions in the used DEA model. The comparison of values before and after the model calibration shows that the correlation between CIS and MOTR increased 25.79 times for BR-27, i.e. now there is some correlation, which previously did not exist.

Complementarily, the correlation between CIS and GDPC more than doubled for the Brazilian data. Concerning the European data, higher correlations were also verified, respectively 7 and 5% higher for the pairs CIS x MOTR and CIS x GDPC. The smaller effect of this intervention on the EU-27 data set reinforces the idea presented previously in this section that the pointed bias produced by MR on the relation between MR and FR1 is considerably lower in highly motorized countries.

If the weight restrictions between MR and FR were not included, this would lead to a road safety ranking (expressed by the CIS values) that is quite different from the socio-economic reality of the compared states, because of the bias transferred from the MR to the CIS, as demonstrated by the low correlation coefficient values shown in Table 6.1.

Although socio-economic issues do not completely explain the road safety problem of a place, at least some extent of association between them is expected. The optimum index scores (OIS) and the cross-index scores (CIS) are presented in Table 6.2, where SP proves to be the best-performing DMU. The influence of the model intervention (i.e. the insertion of the additional weight restrictions) on the EU-27 data set was very low; for this reason, no significant changes occurred in the European ranking.

**Table 6.2** – OIS and CIS for the BR-27 data set after model calibration

| State | OIS | CIS |
|---|---|---|
| SP | 1.00 | 1.00 |
| RS | 1.22 | 1.22 |
| RJ | 1.33 | 1.34 |
| DF | 1.37 | 1.38 |
| AM | 1.41 | 1.43 |
| MG | 1.47 | 1.48 |
| SC | 1.61 | 1.65 |
| RN | 1.67 | 1.69 |
| AC | 1.80 | 1.82 |
| AP | 1.82 | 1.85 |
| GO | 1.88 | 1.89 |
| PR | 1.88 | 1.90 |
| BA | 1.88 | 1.91 |
| MS | 2.02 | 2.03 |
| PA | 2.08 | 2.12 |
| ES | 2.19 | 2.20 |
| PE | 2.22 | 2.25 |
| MT | 2.24 | 2.25 |
| PB | 2.22 | 2.26 |
| RO | 2.27 | 2.28 |
| RR | 2.30 | 2.32 |
| CE | 2.39 | 2.42 |
| TO | 2.59 | 2.61 |
| SE | 2.71 | 2.75 |
| MA | 2.77 | 2.83 |
| AL | 3.27 | 3.33 |
| PI | 3.47 | 3.52 |

The degree of influence related to the MR share of the composite indicator may be assessed through the comparison of the three shares that together constitute the index value, which are $w_1*MR$, $w_{2,1}*FR1$ and $w_{2,2}*FR2$. The preliminary results showed that the share of $w_1*MR$ in the index value was on average much greater, indicating that the bias of the MR was strongly transferred to the index values. In the calibrated model, through the additional weight restriction, the degree of influence related to the MR share of the composite indicator was reduced, as well as the bias of the MR in the ranking. The average shares in the composite indicator values before and after calibration are indicated by the following items (accompanied by the lower and upper values):

- $w_1*MR$ = 68% before calibration (from 14 up to 87%) and 26% after calibration (from 14 up to 38%);

- $w_{2,1}$*FR1 = 16% before calibration (from 7 up to 32%) and 31% after calibration (from 22 up to 39%);
- $w_{2,2}$*FR2 = 16% before calibration (from 5 up to 54%) and 43% after calibration (from 30 up to 54%).

The substantial changes in the consideration of the most important indicator (given by the mentioned shares) that occurred due to the calibration show the need to properly investigate the nature of the information used to feed the DEA model and its corresponding translation into appropriate weight restrictions. The greater share initially attributed to MR has no theoretical meaning, since it was obtained by simply running the DEA model without any appraisal regarding the relationship between the problem under investigation, the available indicators and the selected technique to deal with it.

The result in terms of the states' ranking after the calibration is generally in line with expectations. In addition, based on available data investigation, one of the explanations for the fact that AM, which is expected to show an inferior performance, still ranks 5$^{th}$, may be the relatively low highway density of this sparsely populated state, where freight and passenger transport are extensively made by waterway connections.

Given the fact that accidents on highways represent an important share of all accidents, the migration of this kind of exposure to other non-road transportation modes is probably a factor capable of reducing the overall fatality number. Statistics from 2011 reveal that AM has a paved highway network density of 1.23 km/$10^3$ km$^2$, the lowest value among the Brazilian states; while the well performing states such as SP, RS, RJ and DF have paved highway density values equal to respectively 134.99 km/$10^3$ km$^2$ (the third highest among the BR-27), 331.82 km/$10^3$ km$^2$ (the highest among the BR-27), 169.50 km/$10^3$ km$^2$ (the second highest among the BR-27) and 44.13 km/$10^3$ km$^2$ (*Ministério dos Transportes*, 2013b).

Ideally, these comparisons should be made, taking into account the traffic volumes on the paved highway network in each state; however, this information is not available at this level of disaggregation. In conclusion, the fact is that it may not be suitable to only compare all BR-27 states in one single analysis. To overcome this complication, the following section presents an additional computation and investigation of the index score results carried out by means of cluster analysis.

## 6.2. MODEL APPLICATION AND RESULTS – CLUSTER DIAGNOSIS BASED ON THE CI

The need to carefully investigate the DEA model application for BR-27 and to adjust the model to the situation at hand was shown in the calibration section. After this process, the next step consists of dividing the states into the three clusters configuration mentioned in Chapter 4, in order to offer a more feasible basis for the transference of good experiences from best to underperforming states.

Then, the ML DEA-CI model presented in Section 6.1 was run for Clusters 1, 2 and 3, and the composite indicator scores were computed. We run the DEA models using the software Lingo®, developed by Lindo Systems. Furthermore, for testing the sensitivity of the DEA scores to variations in the model inputs (the three fatality related indicators), the bootstrapping procedure described in Section 5.3 was applied and artificial data sets were produced. In the bootstrapping procedure, the steps described in Figure 5.3 were applied multiple times[1] for each defined cluster. The model script is available in Appendix F.

By means of the bootstrapping process, we can assess the robustness of the computed DEA scores and contribute to the validation of the obtained rankings. The bootstrapping technique has been exclusively applied to the clustered DMUs since these provide a more feasible basis for performance comparison, whereas the analysis of all states together is most suitable for the calibration process.

Tables 6.3, 6.4 and 6.5 contain the mortality rate (MR) and the two fatality rates (fatalities per vehicle – FR1 and fatalities per vehicle kilometer traveled – FR2) used to obtain the CI for the states allocated to Clusters 1, 2 and 3, respectively. The data correspond to the average values in the period 2010-2012 and are ordered according to FR2.

Tables 6.5, 6.6 and 6.7 show the *OIS* (Optimum Index Score), the *CIS* (Cross-index Score), the average CIS value ($\overline{CIS}$), the standard deviation ($\sigma_{CIS}$), the median CIS value ($\widetilde{CIS}$), the CIS 1st quartile value ($Q_{1_{CIS}}$) and the CES 3rd quartile value ($Q_{3_{CIS}}$) for the entire set of bootstrapped scores. The values are ordered according to the ($\widetilde{CIS}$).

[1] In this research, we produced 50 bootstrapped data sets.

**Table 6.3** – Mortality and fatality rates for Cluster 1 states

| State | MR (fat/$10^5$ inh) | | FR1 (fat/$10^4$ veh) | | FR2 (fat/$10^9$ km) | |
|---|---|---|---|---|---|---|
| SP | 17.17 | ($1^{st}$) | 3.40 | ($1^{st}$) | 30.45 | ($1^{st}$) |
| RS | 19.57 | ($3^{rd}$) | 4.38 | ($2^{nd}$) | 38.68 | ($2^{nd}$) |
| DF | 23.39 | ($5^{th}$) | 4.83 | ($3^{rd}$) | 40.58 | ($3^{rd}$) |
| SC | 30.22 | ($6^{th}$) | 5.54 | ($4^{th}$) | 47.01 | ($4^{th}$) |
| MG | 22.45 | ($4^{th}$) | 6.16 | ($5^{th}$) | 47.96 | ($5^{th}$) |
| RJ | 17.97 | ($2^{nd}$) | 6.18 | ($6^{th}$) | 48.86 | ($6^{th}$) |
| PR | 32.85 | ($8^{th}$) | 6.64 | ($7^{th}$) | 60.55 | ($7^{th}$) |
| ES | 32.76 | ($7^{th}$) | 9.12 | ($8^{th}$) | 75.47 | ($8^{th}$) |

**Table 6.4** – Mortality and fatality rates for Cluster 2 states

| State | MR (fat/$10^5$ inh) | | FR1 (fat/$10^4$ veh) | | FR2 (fat/$10^9$ km) | |
|---|---|---|---|---|---|---|
| AM | 13.52 | ($1^{st}$) | 8.81 | ($3^{rd}$) | 45.75 | ($1^{st}$) |
| AP | 19.21 | ($3^{rd}$) | 10.97 | ($7^{th}$) | 50.61 | ($2^{nd}$) |
| GO | 31.71 | ($6^{th}$) | 7.93 | ($1^{st}$) | 56.03 | ($3^{rd}$) |
| AC | 20.45 | ($4^{th}$) | 9.73 | ($5^{th}$) | 56.22 | ($4^{th}$) |
| MS | 32.74 | ($7^{th}$) | 8.29 | ($2^{nd}$) | 56.65 | ($5^{th}$) |
| MT | 36.01 | ($8^{th}$) | 9.38 | ($4^{th}$) | 57.05 | ($6^{th}$) |
| RO | 39.19 | ($10^{th}$) | 10.62 | ($6^{th}$) | 64.85 | ($7^{th}$) |
| PA | 18.29 | ($2^{nd}$) | 13.85 | ($10^{th}$) | 68.23 | ($8^{th}$) |
| RR | 31.00 | ($5^{th}$) | 11.04 | ($8^{th}$) | 68.74 | ($9^{th}$) |
| TO | 38.00 | ($9^{th}$) | 13.32 | ($9^{th}$) | 70.34 | ($10^{th}$) |

**Table 6.5** – Mortality and fatality rates for Cluster 3 states

| State | MR (fat/$10^5$ inh) | | FR1 (fat/$10^4$ veh) | | FR2 (fat/$10^9$ km) | |
|---|---|---|---|---|---|---|
| RN | 19.21 | ($2^{nd}$) | 8.18 | ($1^{st}$) | 55.28 | ($1^{st}$) |
| BA | 19.07 | ($1^{st}$) | 11.25 | ($3^{rd}$) | 68.80 | ($2^{nd}$) |
| PB | 22.82 | ($5^{th}$) | 11.69 | ($4^{th}$) | 75.74 | ($3^{rd}$) |
| PE | 22.39 | ($4^{th}$) | 10.79 | ($2^{nd}$) | 77.12 | ($4^{th}$) |
| SE | 29.81 | ($8^{th}$) | 14.21 | ($6^{th}$) | 91.99 | ($5^{th}$) |
| CE | 25.46 | ($6^{th}$) | 12.10 | ($5^{th}$) | 92.54 | ($6^{th}$) |
| MA | 22.22 | ($3^{rd}$) | 17.11 | ($7^{th}$) | 99.16 | ($7^{th}$) |
| PI | 35.39 | ($9^{th}$) | 17.96 | ($8^{th}$) | 117.01 | ($8^{th}$) |
| AL | 26.32 | ($7^{th}$) | 18.30 | ($9^{th}$) | 117.07 | ($9^{th}$) |

**Table 6.6** – OIS and CIS for Cluster 1 states

| State | OIS | CIS | $\overline{CIS}$ | $\sigma_{CIS}$ | $\widetilde{CIS}$ | $Q_{1_{CIS}}$ | $Q_{3_{CIS}}$ |
|---|---|---|---|---|---|---|---|
| SP | 1.000 | 1.000 | 1.014 | 0.070 | 1.000 | 1.000 | 1.000 |
| RS | 1.226 | 1.233 | 1.636 | 0.432 | 1.617 | 1.357 | 1.835 |
| DF | 1.365 | 1.367 | 1.766 | 0.511 | 1.659 | 1.357 | 2.043 |
| RJ | 1.446 | 1.485 | 2.241 | 0.765 | 2.170 | 1.598 | 2.777 |
| MG | 1.532 | 1.555 | 2.442 | 0.707 | 2.330 | 1.888 | 2.906 |
| SC | 1.605 | 1.637 | 2.729 | 0.885 | 2.702 | 2.030 | 3.220 |
| PR | 1.947 | 1.955 | 3.916 | 1.418 | 3.755 | 2.829 | 4.536 |
| ES | 2.313 | 2.352 | 5.816 | 1.583 | 5.705 | 4.598 | 6.803 |

**Table 6.7** – OIS and CIS for Cluster 2 states

| State | OIS | CIS | $\overline{CIS}$ | $\sigma_{CIS}$ | $\widetilde{CIS}$ | $Q_{1_{CIS}}$ | $Q_{3_{CIS}}$ |
|---|---|---|---|---|---|---|---|
| AM | 1.000 | 1.000 | 1.001 | 0.006 | 1.000 | 1.000 | 1.000 |
| AP | 1.190 | 1.213 | 1.413 | 0.228 | 1.352 | 1.255 | 1.539 |
| GO | 1.179 | 1.205 | 1.411 | 0.197 | 1.394 | 1.260 | 1.522 |
| AC | 1.195 | 1.204 | 1.428 | 0.218 | 1.428 | 1.255 | 1.591 |
| MS | 1.214 | 1.239 | 1.509 | 0.215 | 1.485 | 1.355 | 1.643 |
| MT | 1.310 | 1.333 | 1.759 | 0.240 | 1.726 | 1.569 | 1.927 |
| RR | 1.460 | 1.479 | 2.159 | 0.299 | 2.126 | 1.999 | 2.330 |
| PA | 1.484 | 1.514 | 2.207 | 0.305 | 2.161 | 2.022 | 2.440 |
| RO | 1.472 | 1.497 | 2.194 | 0.318 | 2.194 | 1.976 | 2.417 |
| TO | 1.667 | 1.683 | 2.754 | 0.472 | 2.799 | 2.443 | 3.049 |

**Table 6.8** – OIS and CIS for Cluster 3 states

| State | OIS | CIS | $\overline{CIS}$ | $\sigma_{CIS}$ | $\widetilde{CIS}$ | $Q_{1_{CIS}}$ | $Q_{3_{CIS}}$ |
|---|---|---|---|---|---|---|---|
| RN | 1.000 | 1.000 | 1.007 | 0.034 | 1.000 | 1.000 | 1.000 |
| BA | 1.174 | 1.186 | 1.505 | 0.358 | 1.468 | 1.244 | 1.694 |
| PB | 1.311 | 1.318 | 1.784 | 0.443 | 1.717 | 1.448 | 2.039 |
| PE | 1.276 | 1.289 | 1.789 | 0.423 | 1.819 | 1.422 | 1.979 |
| CE | 1.466 | 1.491 | 2.316 | 0.584 | 2.179 | 1.930 | 2.733 |
| MA | 1.608 | 1.637 | 2.920 | 0.748 | 2.824 | 2.353 | 3.296 |
| SE | 1.636 | 1.642 | 2.944 | 0.634 | 2.861 | 2.568 | 3.317 |
| AL | 1.846 | 1.871 | 3.585 | 0.857 | 3.619 | 2.974 | 4.181 |
| PI | 2.025 | 2.035 | 4.453 | 1.180 | 4.276 | 3.651 | 5.089 |

In terms of the benchmarking exercise, the leading state in each table is considered as the best-performing reference in every cluster, i.e. other states can learn from its achievement and improve their own road safety condition. The boxplot diagrams in Figures 6.4, 6.5 and 6.6 provide an overview of the distribution of all computed CISs through the indication of the minimum, 1st quartile, median, 3rd quartile and maximum CIS's values for Clusters 1, 2 and 3, respectively. The dots represent outlier observations.

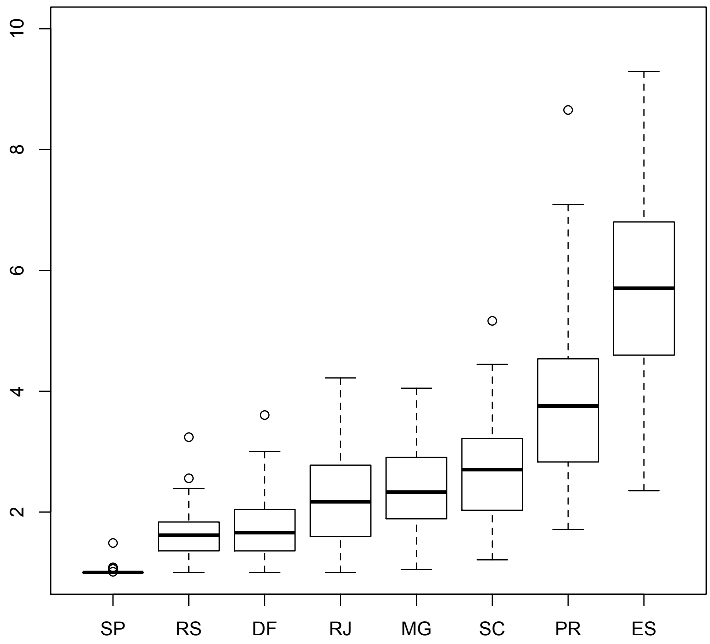



**Figure 6.4** – Distribution of Cluster 1 bootstrapped CISs

In Cluster 1, SP is the absolute best-performing state, occupying the first position in the cluster (and actually among all Brazilian states). One position lower, RS is also capable of reaching an index score equal to 1.00 and possibly occupying the first position, as well as DF and RJ. MG occupies the next position, with a very similar situation as RJ, although RJ performs slightly better. SC holds an unfavorable position between the formerly cited states and the two worst performing states in Cluster 1: PR and ES. The fact that PR and ES do not perform well in both rankings (the national and the clustered), but still show characteristics that allowed their classification in Cluster 1, is an indication of how intensely PR and ES are performing below expectations with respect to road safety risk.

In Cluster 2, which incorporates the Northern region and most part of the Center-Western region of the country, AM appears as the absolute best-performing DMU, since it was the best-performing state in every iteration. GO, AC and AP are consecutive well-performing states, where AC performs somewhat better than GO and AP. MS is quite close on the 5th position. In the second half, MT presents an average performance in Cluster 2. RR, PA and RO show practically the same performance, since their bootstrapped scores distribution presents similar parameters (minimum value, 1st quartile, median, 3rd quartile and maximum value). TO, the most recently established Brazilian state (since it was formerly part of GO) is the worst-performing state. The bootstrapped procedure enabled a clearer view on the low performing states RR, PA, RO and TO, since the original CIS of these states were actually classified as outliers amongst the whole set of generated scores for these states, as indicated by the dots in Figure 6.5. This is, the optimum scores computed using the original scores resulted in an uncommon value if compare to the rest of bootstrapped scores.

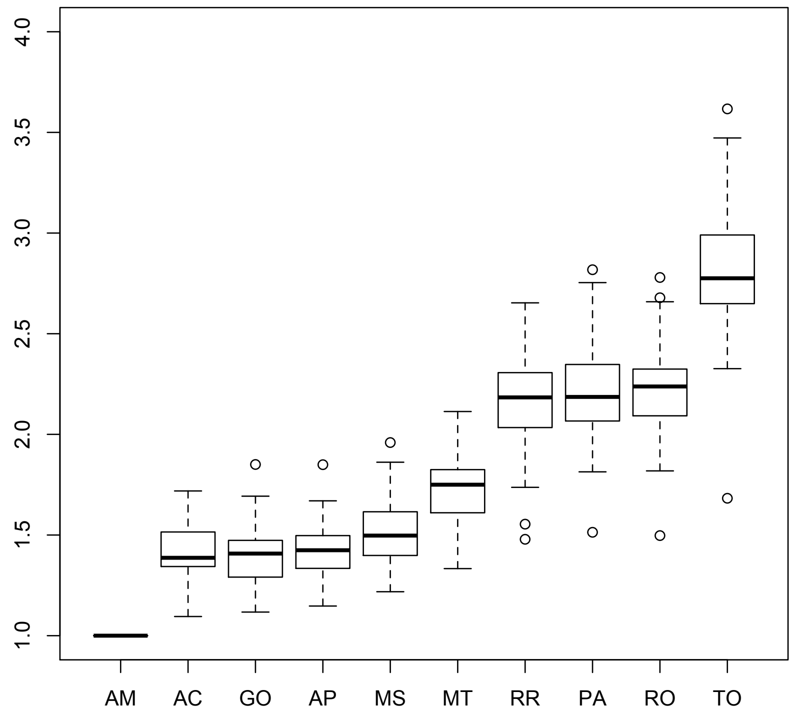



**Figure 6.5** – Distribution of Cluster 2 bootstrapped CISs

In Cluster 3, consisting of the entire set of nine Northeastern states, RN appears as the best-performer, even though in very few cases disproved by BA, PB and PE. BA also labels as a well-performing state in this region, as well as PB and PE. A second group of states led by CE and completed by MA and SE, constitute the average performers in the cluster. Further down, AL and PI appear as a third, below-average performing group. AL and PI hold the least favorable positions, not only within their cluster, but also when compared to the rest of Brazil.

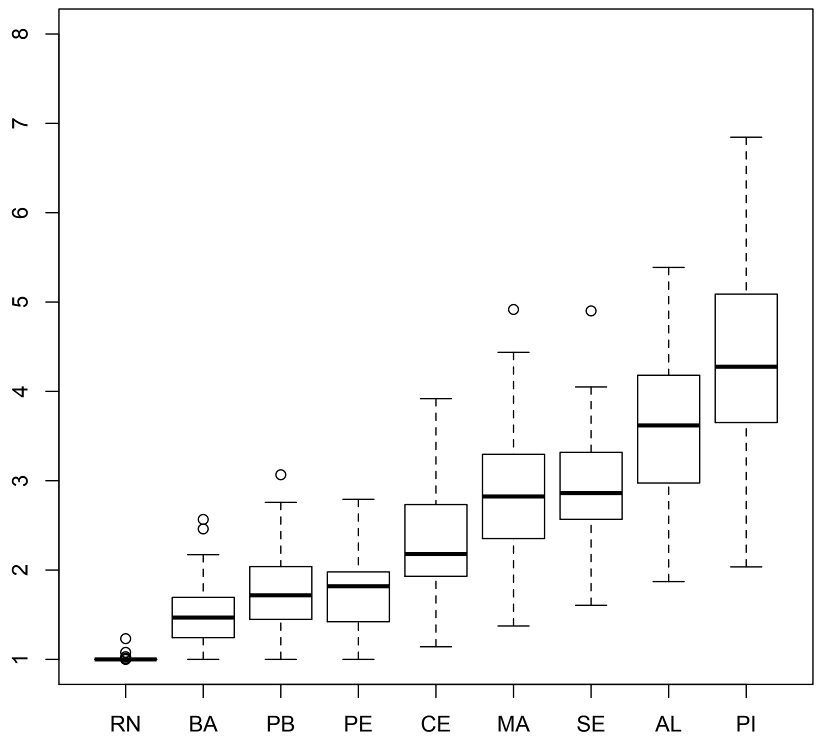



**Figure 6.6** – Distribution of Cluster 3 bootstrapped CISs

## 6.3. CHAPTER CONCLUSIONS

The DEA model computes the optimum solution for every state considering its respective cluster; thus, there is no excuse that the performance of a particular state was underestimated due to an unfair weight allocation process. This results in the OIS value, which is the CI used to identify the best-performing (safer) state; and the performance of the underperforming states is measured as a share of the best-performing state score.

DEA possesses some features that make its use very attractive for the topic of this manuscript. Firstly, obtaining a rank is not the only purpose of DEA; it can be seen as a benchmarking tool, since the performance of the DMUs is evaluated in relation to the best-performing ones. It suggests that underperforming DMUs can learn from the best-performers. In terms of road safety, it is particularly valuable, because states or countries that do not show good road safety results can learn from successful initiatives taken in best-performing states/countries.

Thus, according to the CI value, SP, AM and RN are the best-performing states in their respective cluster, and should be a reference for the other states' improvements in road safety. Nevertheless, in order to have a common basis to compare different states' performances, we compute the CIS value, in which each state's score is computed not only based on its own weights, but also using the weights attributed to all other states' indicators. It enables the ranking of the states according to their road safety performance, from best to worst performing. Even though there are clusters to avoid very contrasting comparisons (which could be unrealistic), taking into account the $\widetilde{CIS}$ value (this is, after the sensitivity analysis), the worst performing state in Cluster 1 presents only 18% of the safety level found in SP; in Cluster 2 TO presents only 36% of the safety level found in AM; and, in Cluster 3, PI presents only 23% of the safety level found in RN.

In Appendix A, state profiles are presented, including a summary of this and upcoming chapters' results.

# 7. SETTING THE TARGET NUMBER OF FATALITIES

This chapter presents the procedure for setting the target number of traffic fatalities in Brazilian states, considering the same cluster division as in the previous research steps. The approach to set the target number of traffic fatalities involves the application of an input-output DEA model, which differs from the model applied in Chapter 6 because now, instead of a composite indicator, the result is an efficiency score representing the overall traffic fatality risk. This efficiency is measured according to the ability of each state to transform its exposure level into as few fatalities as possible. Based on benchmarking principles, it is possible to derive targets on the number of traffic fatalities from the performance comparisons. Although we again obtain a performance measure through the efficiency score (and the ranking results based on this measure are actually quite similar to the results of the composite indicator model), the focus here is to suggest the target number of traffic fatalities. Regarding the sensitivity analysis, we apply again the bootstrapping procedure; that way, states can define more or less ambitious targets within a suggested range.

## 7.1. TARGET SETTING USING DEA

Although there is a general desire of reducing the number of fatalities in Brazil, the availability of a concrete and scientifically defined target, deduced from the definition of an ‘acceptable’ number of traffic fatalities, is an extremely important initial step for road safety strategic planning. The main bodies involved in the road safety issue (e.g. government, society, non-governmental organizations, automobile industry or any other stakeholder) should interpret it as a common goal.

However, so essential as the target definition is its feasibility; in other words, it is not effective to propose very ambitious targets if they are not expected to be attained in a realistic prognostic. This misconception is even more likely to happen when there is a common sense of transferring successful approaches towards road safety from high-income countries to Brazil, where the results might not be so promising.

In order to avoid frustration, which is discouraging for future actions, it seems reasonable to focus on better practice examples inserted into a similar context or reality. In the case of Brazil, for instance, it corresponds to comparing the performance of states grouped into the same cluster. This procedure enables a realistic benchmarking process, since an efficacious action towards road safety in a certain state (i.e. a state level campaign or a specific enforcement policy) is more likely to generate desirable results in a state containing a similar background. The assumption enables the states (DMUs) from the same clusters to have their performances directly compared and to compute a target number of fatalities for the underperforming DMUs in the cluster.

This target is the hypothetical number of fatalities the underperforming DMU would sustain if it were able to perform as good as the best-performing DMU. The concept of setting the target number of fatalities based on benchmarking the road safety performance has already been applied for European Union countries, however using different input information to feed the DEA-model (Shen, 2012). A numerical example explains this notion in the next paragraph.

When the *OES* value is equal to 1.00, it means the corresponding DMU succeeds in minimizing its traffic fatality risk level. For this reason, this DMU serves as a benchmarking reference for other DMUs in that cluster. In the benchmarking process, there is the idea that all the underperforming DMUs are capable to reach the performance level achieved by the leading DMU (in a theoretical premise that the aptitude in controlling the number of traffic fatalities could be transferred from one DMU to another). Thus, the *OES* of underperforming DMUs expresses an inefficiency rate in relation to the DMU holding a score equal to 1. For example, consider two DMUs: “A” presenting an *OES=1.00* and “B” an *OES=1.20*. In this situation, we can say that DMU B presents 83% of DMU A’s performance (1.00/1.20= 0.83). This concept is applicable to the input/output viewpoint, since it means that DMU B needs to decrease its output’s value by 17% and maintain its inputs unchanged in order to be as efficient as DMU A; in other words, DMU B’s target is to reduce its absolute number of fatalities by 17%.

Equation 7.1 presents an input-oriented converted form of the original CCR model (presented in Chapter 5, in which we introduced the DEA technique), in which the outputs to be minimized refer to the overall number of traffic fatalities in each state; while the respective amount of exposure remains unchanged; in other words, a decrease in the traffic fatality risk level. The exposure here is defined in terms of population, vehicle fleet and amount of kilometers traveled, which are the denominators of the most traditionally used traffic fatality related indicators (Chapters 1 and 2 adequately cover this theme). The scores indicating the best-performers present a value equal to 1 (as in the original model); that is, they succeed in minimizing their sustained risk level. On the other hand, underperforming states present a score higher than 1, since they did not succeed in minimizing their fatalities (output) most efficiently .

$$OES_s = \min \sum_{i=1}^{p} w_{O_i} O_{i,s} \tag{7.1}$$

$$subject \quad to \quad \sum_{j=1}^{q} w_{I_j} I_{j,s} = 1,$$

$$\sum_{i=1}^{p} w_{O_i} O_{i,r} - \sum_{j=1}^{q} w_{I_j} I_{j,r} \geq 0 \qquad r = 1,\ldots,n$$

$$w_{O_i} \geq 0 \quad i = 1,\ldots,p \qquad w_{I_j} \geq 0 \quad j = 1,\ldots,q \quad s = 1,\ldots,n$$

In which: $OES_s$ – Optimum efficiency score of the *s-th* DMU; $O_{i,s}$ – *i-th* output of the *s-th* DMU; $w_{O_i}$ – Weight attributed to the output $O_i$; $I_{j,s}$ – *j-th* input of the *s-th* DMU; $w_{I_i}$ – Weight attributed to the input $I_j$; $n$ – total number of DMUs; $p$ – total number of outputs; $q$ – total number of inputs.

The solution of Equation 7.2 provides for each state *s* the efficiency score indicator based on a *K*-layered hierarchy of *p* inputs can be calculated for each state *s*, where $u_{fK}$ is the weight given to the *f-th* category in the *K-th* layer and $w_{fK}^{(K)}$ denotes the non negative internal weights associated with the indicators of the *f-th* category in the *K-th* layer; the sum of all $w_{fK}^{(K)}$ within a particular category is equal to one. This model is named as ML DEA-ES (Multiple Layer DEA Efficiency Score). As we already stated for the ML DEA-CI model, considering all the indicators in the same layer leads to weak discriminating power and unrealistic weight allocations (Shen et al., 2011). This hierarchical model is important in order to consider the associations between the values we are using, e.g. the vehicle fleet and the VKT are firstly "merged" before composing it with the population value (which is less directly related).

$$OES_s = \min \sum_{f_{K_1}=1}^{p^{(K)}} u_{f_K} \left( \sum_{f_{K-1} \in A_K^{(K)}} w_{K-1}^{(K-1)} \left( \ldots \sum_{f_k \in A_{k+1}^{(k+1)}} w_k^{(k)} \left( \ldots \sum_{f_{k-1} \in A_{f_3}^{(3)}} w_2^{(2)} \left( \sum_{f_1 \in A_{f_2}^{(2)}} w_{f_1}^{(1)} O_{f_1 s} \right) \right) \right) \right) \tag{7.2}$$

The DEA model computes the optimum solution for every DMU; thus, there is no excuse that the efficiency performance of a particular DMU was underestimated due to an unfair weight allocation process. This results in the OES value. Nevertheless, with the same idea we exposed in Chapter 6 and obtaining a common basis to compare different DMUs' performances, we compute the CES value, according to Equation 5.3 (since now we have an input-output model again).

However, it is still necessary to bootstrap the DEA estimators in order to tackle the inherent vulnerabilities of the DEA model (already discussed in Section 5.3) and to test the sensitivity of the DEA scores to sampling variations. It is mainly expressed in the results throughout the intervals in the target number of fatalities instead of a single fixed target.

## 7.2. MODEL APPLICATION AND RESULTS

The input and output values, corresponding respectively to the three exposure measures (number of inhabitants, number of motorized vehicles and the VKT) and the number of traffic fatalities, are available in Tables 7.1 to 7.3 for Clusters 1, 2 and 3. The data correspond to the average values in the period 2010-2012 and are decreasingly ordered according to the number of traffic fatalities.

**Table 7.1**– Input (exposure) and output (traffic fatalities) for Cluster 1 states

| State | Input | | | Output |
|---|---|---|---|---|
| | Population (million inhabitants) | Fleet (million vehicles) | VKT (billion km) | Traffic fatalities |
| SP | 41.580 | 20.824 | 232.141 | 7,055 |
| MG | 19.726 | 7.155 | 91.680 | 4,390 |
| PR | 10.510 | 5.175 | 56.785 | 3,428 |
| RJ | 16.113 | 4.613 | 58.263 | 2,842 |
| RS | 10.733 | 4.800 | 54.294 | 2,092 |
| SC | 6.317 | 3.437 | 40.490 | 1,900 |
| ES | 3.546 | 1.277 | 15.451 | 1,160 |
| DF | 2.607 | 1.268 | 15.078 | 609 |

**Table 7.2** – Input (exposure) and output (traffic fatalities) for Cluster 2 states

| State | Input | | | Output |
|---|---|---|---|---|
| | Population (million inhabitants) | Fleet (million vehicles) | VKT (billion km) | Traffic fatalities |
| GO | 6.080 | 2.436 | 34.435 | 1,923 |
| PA | 7.700 | 1.018 | 20.661 | 1,401 |
| MT | 3.075 | 1.186 | 19.553 | 1,106 |
| MS | 2.477 | 0.981 | 14.380 | 810 |
| RO | 1.576 | 0.581 | 9.464 | 614 |
| TO | 1.401 | 0.401 | 7.588 | 532 |
| AM | 3.537 | 0.545 | 10.447 | 477 |
| AC | 0.746 | 0.158 | 2.731 | 153 |
| RR | 0.460 | 0.130 | 2.070 | 142 |
| AP | 0.684 | 0.120 | 2.626 | 131 |

**Table 7.3** – Input (exposure) and output (traffic fatalities) for Cluster 3 states

| State | Input | | | Output |
|---|---|---|---|---|
| | Population (million inhabitants) | Fleet (million vehicles) | VKT (billion km) | Traffic fatalities |
| BA | 14.098 | 2.400 | 39.159 | 2,685 |
| CE | 8.528 | 1.790 | 23.373 | 2,160 |
| PE | 8.864 | 1.847 | 25.821 | 1,979 |
| MA | 6.643 | 0.859 | 14.807 | 1,463 |
| PI | 3.140 | 0.621 | 9.522 | 1,108 |
| PB | 3.791 | 0.739 | 11.372 | 860 |
| AL | 3.143 | 0.454 | 7.099 | 825 |
| SE | 2.090 | 0.441 | 6.792 | 622 |
| RN | 3.198 | 0.757 | 11.218 | 613 |

The DEA model application (of which the script is available in Appendix G) resulted in the $OES$ value for each state, in which those that reached the value of 1.00 are the best-performers. The $CES$ value enables the comparison of the states' efficiency scores on a common basis. For the bootstrapping procedure, we applied the steps described in the previous section multiple times (exactly 50) for Clusters 1, 2 and 3. Tables 7.4 to 7.6 exhibit the $OES$, the $CES$, the average CES value ($\overline{CES}$), the standard deviation ($\sigma_{CES}$), the median CES value ($\widetilde{CES}$), the CES 1st quartile value ($Q_{1_{CES}}$) and the CES 3rd quartile value ($Q_{3_{CES}}$) for the entire set of bootstrapped scores. The values are ordered according to the ($\widetilde{CES}$).

**Table 7.4** – OES and CES for Cluster 1 states (all users)

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| SP | 1.000 | 1.000 | 1.020 | 0.074 | 1.000 | 1.000 | 1.000 |
| RS | 1.236 | 1.246 | 1.562 | 0.510 | 1.501 | 1.174 | 1.819 |
| DF | 1.365 | 1.370 | 1.868 | 0.657 | 1.735 | 1.400 | 2.205 |
| RJ | 1.426 | 1.490 | 2.274 | 0.842 | 2.245 | 1.657 | 2.629 |
| MG | 1.542 | 1.578 | 2.443 | 0.730 | 2.431 | 1.972 | 2.804 |
| SC | 1.597 | 1.621 | 2.808 | 0.995 | 2.823 | 2.163 | 3.257 |
| PR | 1.955 | 1.961 | 4.002 | 1.531 | 3.852 | 3.047 | 4.630 |
| ES | 2.331 | 2.388 | 5.739 | 2.282 | 5.436 | 4.260 | 7.073 |

**Table 7.5** – OES and CES for Cluster 2 states (all users)

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| AM | 1.000 | 1.000 | 1.000 | 0.002 | 1.000 | 1.000 | 1.000 |
| GO | 1.166 | 1.205 | 1.423 | 0.187 | 1.424 | 1.318 | 1.551 |
| AP | 1.186 | 1.207 | 1.458 | 0.218 | 1.452 | 1.284 | 1.602 |
| MS | 1.202 | 1.238 | 1.509 | 0.182 | 1.503 | 1.391 | 1.643 |
| AC | 1.220 | 1.233 | 1.554 | 0.205 | 1.518 | 1.410 | 1.726 |
| MT | 1.295 | 1.323 | 1.762 | 0.244 | 1.786 | 1.600 | 1.923 |
| PA | 1.448 | 1.482 | 2.194 | 0.328 | 2.184 | 1.987 | 2.369 |
| RO | 1.467 | 1.499 | 2.236 | 0.311 | 2.187 | 2.023 | 2.373 |
| RR | 1.487 | 1.517 | 2.330 | 0.329 | 2.287 | 2.119 | 2.602 |
| TO | 1.684 | 1.700 | 2.916 | 0.444 | 2.828 | 2.679 | 3.171 |

**Table 7.6** – OES and CES for Cluster 3 states (all users)

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| RN | 1.000 | 1.000 | 1.011 | 0.038 | 1.000 | 1.000 | 1.000 |
| BA | 1.198 | 1.212 | 1.506 | 0.386 | 1.471 | 1.191 | 1.796 |
| PE | 1.294 | 1.304 | 1.721 | 0.423 | 1.770 | 1.353 | 1.987 |
| PB | 1.335 | 1.343 | 1.908 | 0.454 | 1.908 | 1.657 | 2.161 |
| CE | 1.485 | 1.506 | 2.263 | 0.563 | 2.387 | 1.815 | 2.661 |
| MA | 1.613 | 1.647 | 2.841 | 0.716 | 2.691 | 2.408 | 3.213 |
| SE | 1.658 | 1.665 | 2.950 | 0.683 | 2.865 | 2.493 | 3.218 |
| AL | 1.863 | 1.891 | 3.603 | 0.839 | 3.426 | 2.982 | 4.167 |
| PI | 2.058 | 2.069 | 4.518 | 1.137 | 4.465 | 3.759 | 5.034 |

The boxplot diagrams in Figures 7.1 to 7.3 provide insight in the distribution of all computed CESs through the indication of the minimum, 1st quartile, median, 3rd quartile and maximum CES's values for Clusters 1, 2 and 3, respectively. The dots represent outlier observations.

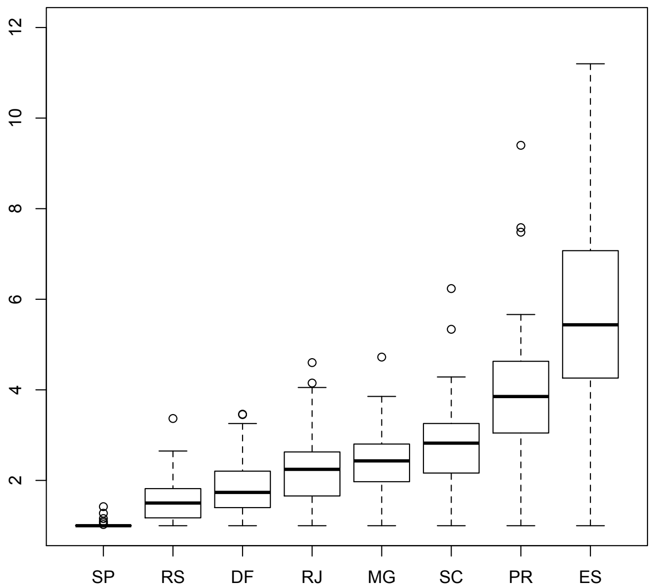



**Figure 7.1** – Distribution of Cluster 1 bootstrapped CESs (all users)

.

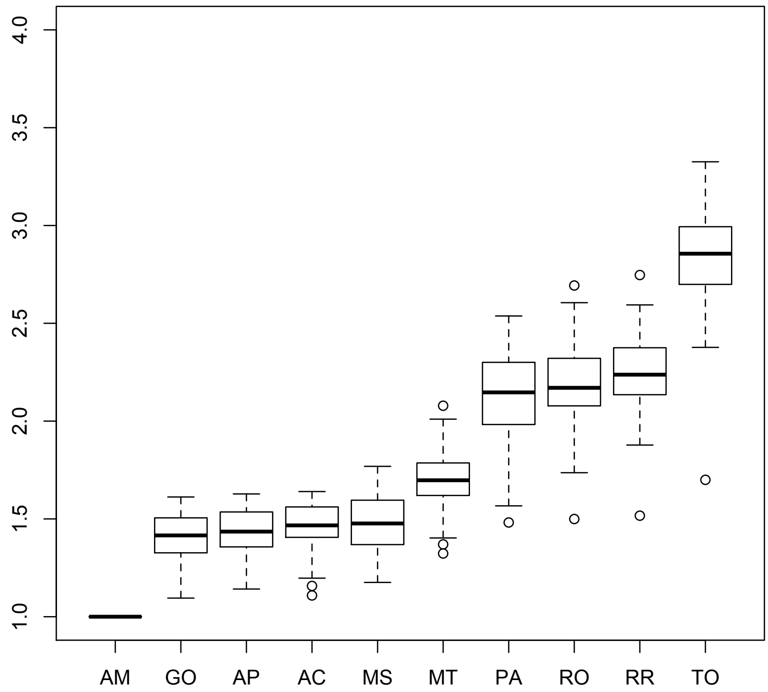



**Figure 7.2** – Distribution of Cluster 2 bootstrapped CESs (all users)

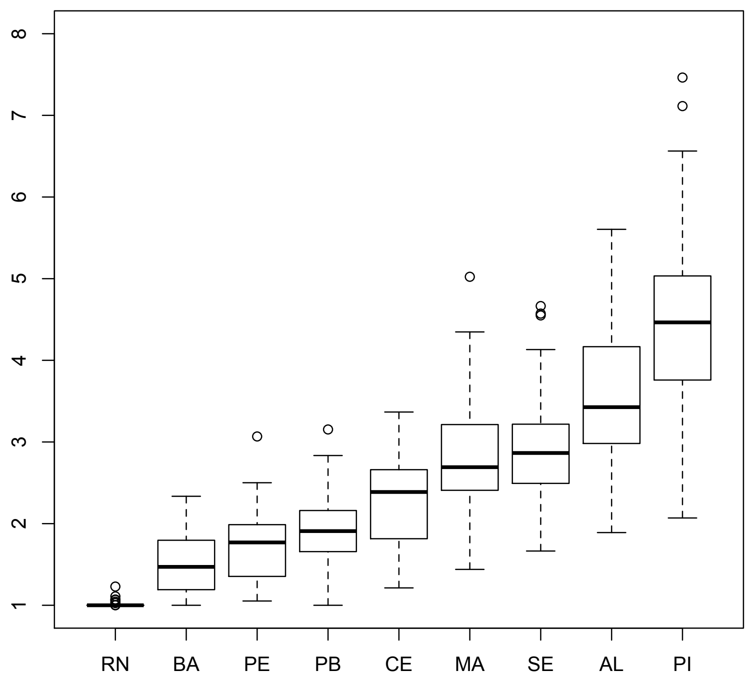



**Figure 7.3** – Distribution of Cluster 3 bootstrapped CESs (all users)

The results obtained with the ML DEA-CI (subject of Chapter 6) and the ML DEA-ES (subject of the current chapter) models are quite similar in terms of the performance ranking of the states previously provided with the intention of aggregating the most commonly used fatality related rates into a single value). Therefore, the intention of applying the ML DEA-ES model lies in the computation of the target number of fatalities in every state, based on the benchmarking concept, the leading state in the cluster each time we run the model is the reference of best-performance. Actually, each cluster might have more than one best-performing state in a given iteration; however, we present and discuss the average and median values of all iterations in this paper.

Here follows an example on how we derived the traffic fatality target number. Since the OES represents a share of the best-performing state performance and in the DEA model we assume that traffic fatalities have to be minimized while the exposure level is kept constant, the only way that an underperforming state can reach the benchmark state performance is by decreasing its number of traffic fatalities to a level in which it would result in the same performance as its benchmark (best-performing) state. In Cluster 1, RS has a total number of traffic fatalities equal to 2,092 – its target number is, therefore, 2,092/1.492 = 1,402.

Then, separately for Clusters 1, 2 and 3, Tables 7.7 to 7.9 exhibit the current number of traffic fatalities and the range of target values, resultant from the benchmarking process. In order to provide more and less ambitious targets, the tables also contain a range of target values computed throughout the addition and subtraction of the standard deviation from the average target value.

Best-performing states as, for instance, SP, AM and RN do not present a target, since this value would be equal to its current number of fatalities. It happens because they present a 100% performance in their respective cluster, considering the median score value, and also considering both quartiles (1$^{st}$ and 3$^{rd}$) – so no targets can be derived using these values.

Considering the values higher than the 3$^{rd}$ quartile, for example the "average score ± standard deviation" it is possible to derive a target, although it represents less probable performance combinations. The reason for this is that these targets result from the few bootstrapping iterations in which they were not considered as the best-performing DMU (i.e., they did not reach the score equal to 1.00).

**Table 7.7** – Traffic fatalities current number, target and target range for Cluster 1 states (all users)

| State | Number of fatalities | | |
|---|---|---|---|
| | Current | Target | Target range |
| SP | 7,055 | 7,055 | - |
| RS | 2,092 | 1,402 | [1,159; 1,769] |
| DF | 609 | 349 | [270; 434] |
| RJ | 2,842 | 1,322 | [1,129; 1,784] |
| MG | 4,390 | 1,847 | [1,602; 2,192] |
| SC | 1,900 | 683 | [593; 892] |
| PR | 3,428 | 893 | [745; 1,116] |
| ES | 1,160 | 218 | [168; 275] |

**Table 7.8** – Traffic fatalities current number, target and target range for Cluster 2 states (all users)

| State | Number of fatalities | | |
|---|---|---|---|
| | Current | Target | Target range |
| AM | 477 | 477 | - |
| GO | 1,923 | 1,388 | [1,274; 1,499] |
| AP | 131 | 92 | [83; 104] |
| AC | 153 | 553 | [506; 598] |
| MS | 810 | 101 | [89; 109] |
| MT | 1,106 | 632 | [586; 705] |
| PA | 1,401 | 660 | [605; 722] |
| RO | 614 | 287 | [264; 310] |
| RR | 142 | 63 | [56; 68] |
| TO | 532 | 190 | [169; 200] |

**Table 7.9** – Traffic fatalities current number, target and target range for Cluster 3 states (all users)

| State | Number of fatalities | | |
|---|---|---|---|
| | Current | Target | Target range |
| RN | 613 | 613 | - |
| BA | 2,685 | 1,835 | [1,512; 2,280] |
| PE | 1,979 | 1,127 | [1,004; 1,476] |
| PB | 860 | 453 | [401; 523] |
| CE | 2,160 | 918 | [823; 1,205] |
| MA | 1,463 | 555 | [467; 623] |
| SE | 622 | 218 | [194; 251] |
| AL | 825 | 244 | [201; 281] |
| PI | 1,108 | 250 | [221; 296] |

In total, the traffic fatality target number we presented (24,426 traffic fatalities, ranging from 22,267 to 27,857) is actually more ambitious than the target set for the end of the Decade of Action for Road Safety (35,156 traffic fatalities) – considering, of course, an equal time horizon for achieving both targets. Based exclusively on this comparison, it could be possible to suggest that the time horizon for the target we set in this research is after 2020. Moreover, 24,426 traffic fatalities correspond to the fact that every state is as good/efficient as the currently (2010-2012) best performers in each cluster.

Clusters containing states with more contrasting performances present more ambitious targets for those underperforming states. This is the case for Cluster 3, in which the states are suggested to reduce their overall number of fatalities by [(6,213/12,316)*100] = 50% [44%; 61%]. Then, there is Cluster 1, which is suggested to reduce its traffic fatalities by 41% [54%; 66%]. Lastly, Cluster 2 is recommended to decrease its traffic fatalities by 39% [56%; 66%].

Finally, setting a challenging yet achievable quantitative target is by no means easy in practice, since it needs to present a balance between the level of ambitious and the link to reality. Targets should be ambitious enough in order to render and motivate all the stakeholders to come together and share their responsibility in achieving common safety goals. Meanwhile, targets should also be realistic so as to keep and strengthen this motivation during the whole target period. In doing so, many factors have to be taken into account, such as the economic status of each state, the level of ambition and commitment, the potential of different measures, the available resources, the transportation infrastructure, characteristics of the drivers population and so on.

In Appendix A, state profiles include a summary of this, the previous, and upcoming chapters results.

As a complement, the next section presents the procedure for setting the target number of motor vehicle traffic fatalities in Brazilian states; now, the idea is to develop a more specific target setting process.

## 7.3. TARGET FOR MOTOR VEHICLE OCCUPANTS

In this section, we suggest a target for the set of fatalities of motor vehicle occupants, based on the exposure level of four categories of motorized vehicles (motorcycle, car, truck and bus) and, of course, on the corresponding (current) number of fatalities. The model results include an efficiency score manifesting the motor vehicle occupants traffic fatality risk, which offers the opportunity to derive targets on the corresponding number of the traffic fatalities. In order to ensure the possibility of comparison between the previous DEA model's results and the current one, we maintained the same cluster division. Regarding the sensitivity analysis, we again apply the bootstrapping procedure (see Chapter 5) – enabling the states to define more or less ambitious targets inside a determined range.

### 7.3.1 TARGET SETTING FOR MOTOR VEHICLES OCCUPANTS

We apply the model presented in Equation 7.1 (of Section 7.1), which consists of an input-oriented converted form of the original CCR model. The outputs to be minimized refer to the number of traffic fatalities for each of the four vehicle categories in each state given a certain amount of exposure (VKT) (i.e., we aim for a decrease in the traffic fatality risk level). Since we now deal with a more specific users population if compared to the previous chapter – different motor vehicle occupants – we adopted an also more precise exposure measure – different motor vehicle exposures.

The score indicating the best-performer(s) presents a value equal to 1 (as in the original model); they succeed in minimizing their sustained risk level most sufficiently. On the other hand, underperforming DMUs present a score higher than 1, since they did not succeed in minimizing their weighted fatalities (output) to the best possible extent. The methodology for estimating the VKT values per mode is available in Chapter 2.

The choice of a DEA technique to investigate the relationship between the estimated VKT and the number of traffic fatalities is also based on the complex nature of this association, since the VKT in a certain vehicle category is not only a determinant for the fatalities of the corresponding vehicle category occupants, but also for all other more vulnerable vehicle category occupants. For example, car exposure might generate car occupant fatalities and motorcycle occupant fatalities (in case of a "motorcycle x car" collision). In literature, this mutual influence is called "net effect" (Anderson, 2008).

Figure 7.4 illustrates this net effect for the four considered vehicle categories in Brazil in the period 2010-2012. The set of fatalities considered for the figure's percentages computation corresponds exclusively to fatal victims of motor vehicle collisions (this is, excluding those involving no motorized users).

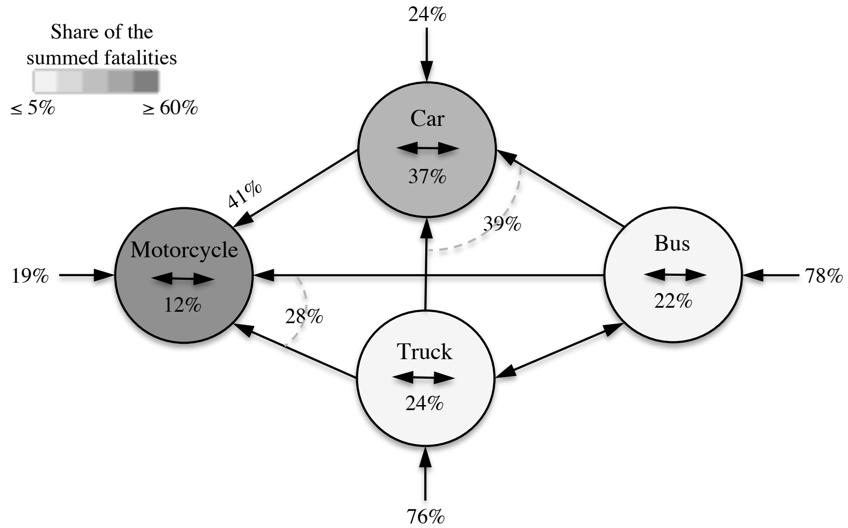



**Figure 7.4** – Net effect for the four considered motor vehicle categories in Brazil

The arrows connecting each category indicate the pair "victim x second party" in a collision (the arrow points in the direction of the victim – the most vulnerable party in the collision). The arrows pointing in both directions inside the circles indicate the collision involving two vehicles of the same category (e.g. "car x car"). The non-connected arrows arriving in the circles represent the single-vehicle accidents. The percentages represent the average shares of each pair for the traffic fatalities in the aforementioned period. The sum of the values of the arriving and internal arrows is equal to 100% (consisting of the total fatalities of the occupants of the corresponding vehicle category). Using car occupant fatalities as an example, 24% result from single-car accidents (e.g. run-off-the road), 39% from "car x truck or bus" and 37% from "car x car" collisions.

Additionally, the available WHO classification (CID-10) does not differentiate between trucks and buses when they are the "second party" in a collision, so for the mentioned 39% of "car x truck or bus" collisions we don't know if the second party involved was a truck or a bus. The color-scale of the circles represents the share of each vehicle category in the sum of the four considered categories (e.g. motorcycle occupant fatalities correspond to almost 60%

of the sum of the four considered categories of fatalities). The construction of Figure 7.4 requires some premises based on the Brazilian data set: trucks, buses and car occupants are not supposed to die in a collision with a motorcycle; truck and bus occupants are not supposed to die in collision with a car.

As Figure 7.4 illustrates, the influence of different road users is not equally distributed in determining the occurrence of a traffic fatality. Due to the disaggregation level of the data source, it is possible to compute the fatality shares mentioned in Figure 7.4 (in spite of a certain level of unspecified collisions). The same cannot be done for the interacting exposures; for instance, it is unmanageable to assess the exposure of motorcycle occupants to a "motorcycle x truck" collisions, since it is not simply a function of summing the exposure values of motorcycles and trucks. It depends on how intense and how direct the interaction of these two exposures is. Hence, the quantification of this mutual influence is an extremely complex and imprecise task due to the large amount of uncontrollable factors influencing this association.

The most obvious and simple way to associate the estimated VKT and the number of traffic fatalities is to compute the ratio between the sum of all types of traffic fatalities and all types of exposures, constituting the so-called fatality rate (FR) in road safety literature. The FR is useful for international or even inter-state comparisons and for over-time monitoring of the overall road safety situation. However, presenting this single number provides an aggregated picture of the situation and does not offer any indication on how to manage the problem (Stipdonk and Berends, 2008).

Thus, the comprehension on how the different exposure types interact and result in different types of traffic fatalities is valuable knowledge for road safety macro-level investigations (in which geographical areas are considered – e.g. provinces, states or countries). In this context, DEA represents a mathematical programming methodology to obtain empirical estimates of the relations between multiple inputs (different types of exposure) and multiple outputs (different types of fatalities) by constructing an efficiency production frontier to assess the so-called relative efficiency of a homogeneous set of DMUs, or states in this case (Shen, 2012). In addition, the advantage that DEA models do not require a priori knowledge about the form of the relation between inputs and outputs is another reason justifying its application for the phenomenon under investigation.

In an association with microeconomic production theory, a firm's input and output combinations are depicted using a production function, also known as efficient production frontier, which indicates the maximum quantity of outputs that can be obtained from a given combination of inputs (Seiford and Thrall, 1990). Using a logic adapted to road safety, the interest is to compute the minimum quantity of outputs (fatalities) that can result from a given combination of exposures (in different vehicle categories).

The decomposition of the exposure and traffic fatalities into different vehicle categories and the attribution of weights to each decomposed unit by the DEA optimization process is a useful strategy to assess the different possibilities of combination that might result from this relationship. In addition to a road safety ranking, the practical outcome of this approach is to indicate the most vulnerable or important vehicle category to focus efforts towards road safety on. In this process, the optimum weight combination resulting in a minimum number of fatalities given a certain level of exposure is computed for each state. The state(s) that is most capable to minimize its fatalities is the most efficient one or, in other words, best-performing regarding road safety.

Figure 7.5 contains a graphical representation of the optimization problem stated in the previous paragraphs.

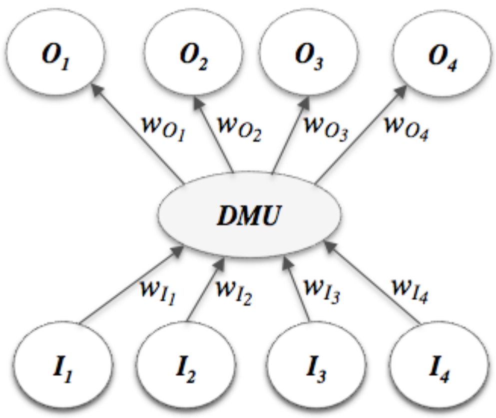



**Figure 7.5** – Structure of inputs and outputs of the DEA motor vehicle efficiency model

In the DEA framework, each DMU (state) is responsible to convert its inputs (VKT by the four different vehicle categories – motorcycle – $I_1$, car – $I_2$, truck – $I_3$ and bus – $I_4$) into outputs (fatalities of occupants of the respective four vehicle categories $O_1$, $O_2$, $O_3$ and $O_4$). The input values are available in the already mentioned Appendix B and the output values are available in Appendix H.

In the optimization model, each input and output is multiplied by a corresponding weight ($w_{I1}$,…, $w_{I4}$ for the inputs and $w_{O1}$,…, $w_{O4}$ for the outputs) which is allowed to vary in order to reach an optimum solution; i.e., minimizing the ratio between weighted outputs (fatalities) and weighted inputs (VKT).

In order to guide the weight attribution process, avoiding unreasonable weight distributions as well as a "black box" effect, we used the shares of each vehicle category in the total exposure and in the traffic fatalities as referential values for weight restriction intervals, which consist of a range of two standard deviations (SD) above and below the average (AVG) share for each cluster, resulting in an upper (UL) and a lower limit (LL), respectively. This flexibility degree allows the model to search for the optimum solution without detaching from reality. Although the referred shares do not correspond directly to the importance of each vehicle category, they constitute valuable hints at the beginning of the optimization process we carry out in this chapter.

**Table 7.10** – Weight restriction limits based on the shares of the VKT and fatalities for Cluster 1

| States (DMU) | Shares of the VKT | | | | Shares of the fatalities | | | |
|---|---|---|---|---|---|---|---|---|
| | Motorcycle | Car | Truck | Bus | Motorcycle | Car | Truck | Bus |
| ES | 32.70% | 49.51% | 8.88% | 8.91% | 46.75% | 50.84% | 2.33% | 0.09% |
| MG | 29.62% | 46.53% | 15.74% | 8.11% | 31.21% | 63.49% | 4.44% | 0.85% |
| RJ | 18.66% | 66.04% | 5.28% | 10.02% | 49.64% | 44.50% | 4.55% | 1.31% |
| SP | 23.32% | 61.40% | 8.58% | 6.70% | 52.27% | 42.76% | 3.86% | 1.11% |
| PR | 23.43% | 54.81% | 15.43% | 6.33% | 43.78% | 51.20% | 4.29% | 0.73% |
| RS | 24.75% | 55.18% | 13.20% | 6.87% | 40.51% | 53.13% | 5.48% | 0.87% |
| SC | 29.19% | 53.95% | 12.50% | 4.36% | 46.87% | 48.06% | 4.04% | 1.03% |
| DF | 13.36% | 73.93% | 7.17% | 5.54% | 37.22% | 60.43% | 1.72% | 0.63% |
| AVG | 24.38% | 57.67% | 10.85% | 7.10% | 43.53% | 51.80% | 3.84% | 0.83% |
| SD | 6.28% | 9.00% | 3.91% | 1.84% | 6.93% | 7.20% | 1.23% | 0.37% |
| 2x SD | 12.56% | 18.00% | 7.81% | 3.67% | 13.85% | 14.41% | 2.46% | 0.74% |
| LL | 11.82% | 39.67% | 3.04% | 3.43% | 29.68% | 37.39% | 1.38% | 0.09% |
| UL | 36.94% | 75.67% | 18.66% | 10.77% | 57.38% | 66.21% | 6.30% | 1.56% |

Table 7.10 contains one numerical example of the weight restriction computation based on data for Cluster 1.

The UL and LL values correspond to weight restrictions we inserted in the DEA model. For example, the share $I_1*w_{I1}$ (motorcycle VKT x weight) is limited to vary between 11.82% and 36.94% of the weighted sum ($I_1*w_{I1} + I_2*w_{I2} + I_3*w_{I3} + I_4*w_{I4}$), and the share $O_1*w_{O1}$ (motorcycle occupant fatalities x weight) is limited to vary between 29.68% and 57.38% of the weighted sum ($O_1*w_{O1} + O_2*w_{O2} + O_3*w_{O3} + O_4*w_{O4}$).

### 7.3.2 MODEL APPLICATION AND RESULTS

The input and output values, relating respectively to the four exposure measures (number of kilometers traveled by motorcycles, cars, trucks and buses) and the corresponding number of traffic fatalities are available in Tables 7.11 to 7.13 for Clusters 1, 2 and 3. The data represent the average values in the period 2010-2012 and are ordered according to the overall number of traffic fatalities.

**Table 7.11**– Disaggregated input (exposure) and output (traffic fatalities) for Cluster 1 states

| State | Input – VKT (billion km) | | | | Output – Traffic fatalities | | | |
|---|---|---|---|---|---|---|---|---|
| | Motorcycle | Car | Truck | Bus | Motorcycle | Cat | Truck | Bus |
| SP | 54.142 | 142.542 | 19.910 | 15.547 | 2220 | 1816 | 164 | 47 |
| MG | 27.159 | 42.657 | 14.426 | 7.437 | 1026 | 2086 | 146 | 28 |
| PR | 13.304 | 31.125 | 8.764 | 3.593 | 1045 | 1222 | 102 | 17 |
| SC | 11.818 | 21.845 | 5.063 | 1.764 | 666 | 682 | 57 | 15 |
| RS | 13.439 | 29.959 | 7.168 | 3.728 | 565 | 741 | 76 | 12 |
| RJ | 10.871 | 38.476 | 3.078 | 5.837 | 655 | 587 | 60 | 17 |
| ES | 5.052 | 7.650 | 1.371 | 1.377 | 412 | 448 | 21 | 1 |
| DF | 2.015 | 11.147 | 1.081 | 0.835 | 138 | 225 | 6 | 2 |

**Table 7.12** – Disaggregated input (exposure) and output (traffic fatalities) for Cluster 2 states

| State | Input – VKT (billion km) | | | | Output – Traffic fatalities | | | |
|---|---|---|---|---|---|---|---|---|
| | Motorcycle | Car | Truck | Bus | Motorcycle | Cat | Truck | Bus |
| GO | 10.08 | 17.26 | 5.50 | 1.60 | 696 | 636 | 66 | 17 |
| MT | 6.84 | 7.14 | 4.69 | 0.88 | 457 | 359 | 61 | 2 |
| PA | 7.19 | 7.17 | 4.50 | 1.80 | 503 | 164 | 20 | 9 |
| MS | 4.28 | 6.28 | 3.08 | 0.74 | 301 | 263 | 41 | 4 |
| RO | 3.83 | 2.86 | 2.29 | 0.48 | 286 | 163 | 23 | 4 |
| TO | 2.70 | 2.53 | 1.90 | 0.46 | 214 | 203 | 14 | 5 |
| AM | 2.43 | 5.49 | 1.60 | 0.92 | 145 | 74 | 7 | 2 |
| RR | 0.81 | 0.85 | 0.31 | 0.10 | 76 | 30 | 0 | 0 |
| AC | 1.05 | 1.03 | 0.55 | 0.10 | 67 | 29 | 1 | 1 |
| AP | 0.70 | 1.20 | 0.53 | 0.19 | 33 | 17 | 0 | 2 |

**Table 7.13** – Disaggregated input (exposure) and output (traffic fatalities) for Cluster 3 states

| State | Input – VKT (billion km) | | | | Output – Traffic fatalities | | | |
|---|---|---|---|---|---|---|---|---|
| | Motorcycle | Car | Truck | Bus | Motorcycle | Cat | Truck | Bus |
| BA | 12.948 | 17.416 | 4.256 | 4.539 | 669 | 1233 | 43 | 16 |
| CE | 11.379 | 9.043 | 1.645 | 1.305 | 1034 | 374 | 12 | 8 |
| PE | 9.787 | 11.692 | 2.461 | 1.881 | 904 | 429 | 35 | 8 |
| MA | 8.176 | 4.771 | 1.055 | 0.805 | 738 | 289 | 34 | 10 |
| PI | 5.156 | 3.020 | 0.768 | 0.578 | 677 | 166 | 21 | 2 |
| PB | 5.005 | 5.056 | 0.720 | 0.591 | 476 | 169 | 11 | 3 |
| SE | 2.606 | 3.107 | 0.530 | 0.549 | 338 | 120 | 23 | 1 |
| RN | 4.524 | 5.263 | 0.799 | 0.632 | 319 | 156 | 4 | 1 |
| AL | 2.616 | 3.471 | 0.459 | 0.552 | 356 | 61 | 5 | 1 |

The DEA model application (of which the script is available in Appendix I) resulted in the $OES$ value for each state, in which those that reached the value of 1.00 are the best-performers. The $CES$ value enables the comparison of the states' scores on a common basis. For the bootstrapping procedure, we again applied the steps described in Chapter 5 multiple times ($t$ = 50) for Clusters 1, 2 and 3. Tables 7.14 to 7.16 exhibit the $OES$, the $CES$, the average CES value ($\overline{CES}$), the standard deviation ($\sigma_{CES}$), the median CES value ($\widetilde{CES}$), the CES 1st quartile value ($Q_{1_{CES}}$) and the CES 3rd quartile value ($Q_{3_{CES}}$) for the entire set of bootstrapped scores. The values are ordered according to the $\widetilde{CES}$.

**Table 7.14** – OES and CES for Cluster 1 states (motor vehicle)

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| SP | 1.000 | 1.000 | 1.237 | 0.435 | 1.000 | 1.000 | 1.280 |
| DF | 1.276 | 1.323 | 2.199 | 1.281 | 1.719 | 1.016 | 3.046 |
| RJ | 1.212 | 1.247 | 1.981 | 0.910 | 1.785 | 1.254 | 2.531 |
| RS | 1.341 | 1.356 | 2.477 | 1.311 | 2.201 | 1.428 | 3.108 |
| MG | 1.790 | 1.834 | 4.532 | 2.495 | 4.014 | 2.697 | 5.968 |
| SC | 1.844 | 1.869 | 4.945 | 2.797 | 4.343 | 2.730 | 6.556 |
| PR | 2.202 | 2.228 | 6.593 | 3.824 | 5.773 | 4.325 | 8.660 |
| ES | 2.970 | 3.030 | 12.204 | 6.625 | 9.944 | 7.329 | 15.817 |

**Table 7.15** – OES and CES for Cluster 2 states (motor vehicle)

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| AP | 1.000 | 1.000 | 1.151 | 0.314 | 1.000 | 1.000 | 1.182 |
| AM | 1.055 | 1.088 | 1.396 | 0.555 | 1.109 | 1.000 | 1.661 |
| PA | 1.594 | 1.660 | 3.322 | 1.467 | 3.097 | 2.293 | 4.245 |
| AC | 1.721 | 1.767 | 3.836 | 1.653 | 3.604 | 2.599 | 4.849 |
| GO | 1.880 | 1.953 | 4.348 | 1.622 | 4.112 | 3.243 | 4.793 |
| MS | 1.909 | 1.974 | 4.633 | 1.913 | 4.149 | 3.307 | 5.599 |
| MT | 1.999 | 2.084 | 4.886 | 2.030 | 4.587 | 3.694 | 5.731 |
| RO | 2.285 | 2.369 | 6.071 | 2.487 | 5.550 | 4.678 | 7.129 |
| TO | 2.611 | 2.728 | 8.694 | 3.893 | 7.439 | 6.212 | 10.502 |
| RR | 2.478 | 2.556 | 8.276 | 2.911 | 7.576 | 6.472 | 10.645 |

**Table 7.16** – OES and CES for Cluster 3 states (motor vehicle)

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| RN | 1.000 | 1.000 | 1.062 | 0.134 | 1.000 | 1.000 | 1.057 |
| BA | 1.000 | 1.054 | 1.148 | 0.210 | 1.054 | 1.000 | 1.242 |
| PE | 1.220 | 1.234 | 1.701 | 0.308 | 1.727 | 1.512 | 1.892 |
| PB | 1.359 | 1.373 | 2.129 | 0.491 | 2.069 | 1.785 | 2.364 |
| AL | 1.441 | 1.460 | 2.237 | 0.583 | 2.203 | 1.865 | 2.712 |
| CE | 1.428 | 1.462 | 2.256 | 0.633 | 2.223 | 1.848 | 2.724 |
| MA | 1.625 | 1.703 | 2.973 | 0.658 | 2.994 | 2.541 | 3.288 |
| SE | 1.621 | 1.648 | 3.062 | 0.734 | 3.093 | 2.527 | 3.507 |
| PI | 2.124 | 2.214 | 5.203 | 1.233 | 5.200 | 4.247 | 5.941 |

The boxplot diagrams in Figures 7.6 to 7.8 provide an overview of the distribution of all computed CESs through the indication of the minimum, 1st quartile, median, 3rd quartile and maximum CES's values for Clusters 1, 2 and 3, respectively. The dots represent outlier observations.

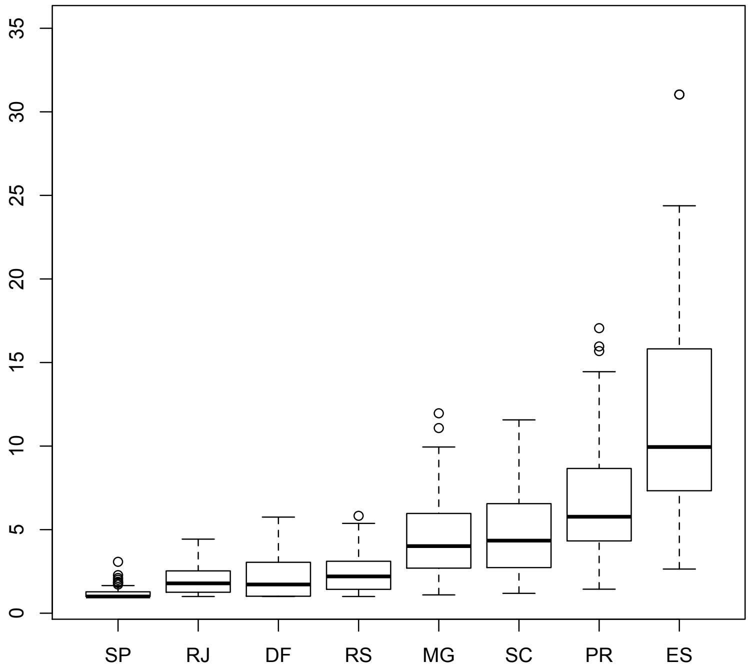



**Figure 7.6** – Distribution of Cluster 1 bootstrapped CESs (motor vehicle)

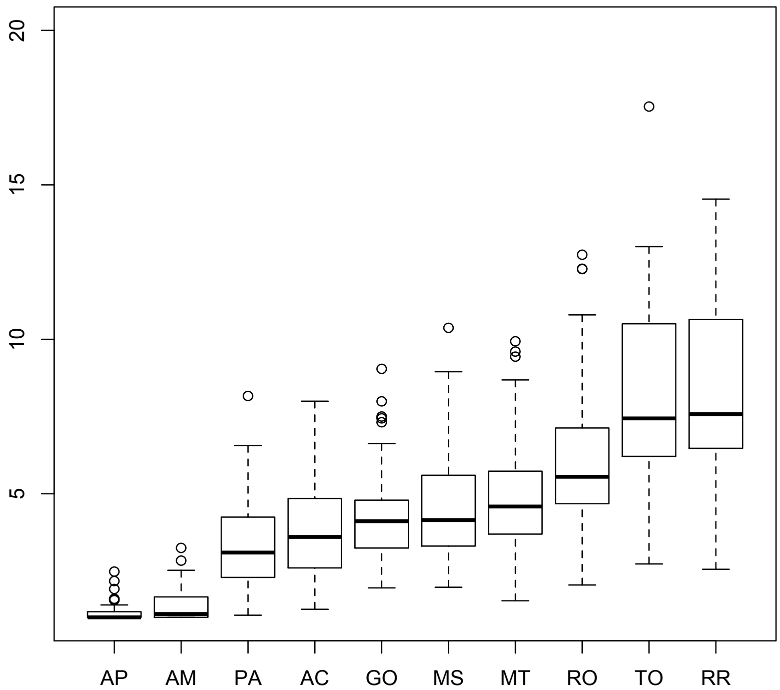



**Figure 7.7** – Distribution of Cluster 2 bootstrapped CESs (motor vehicle)

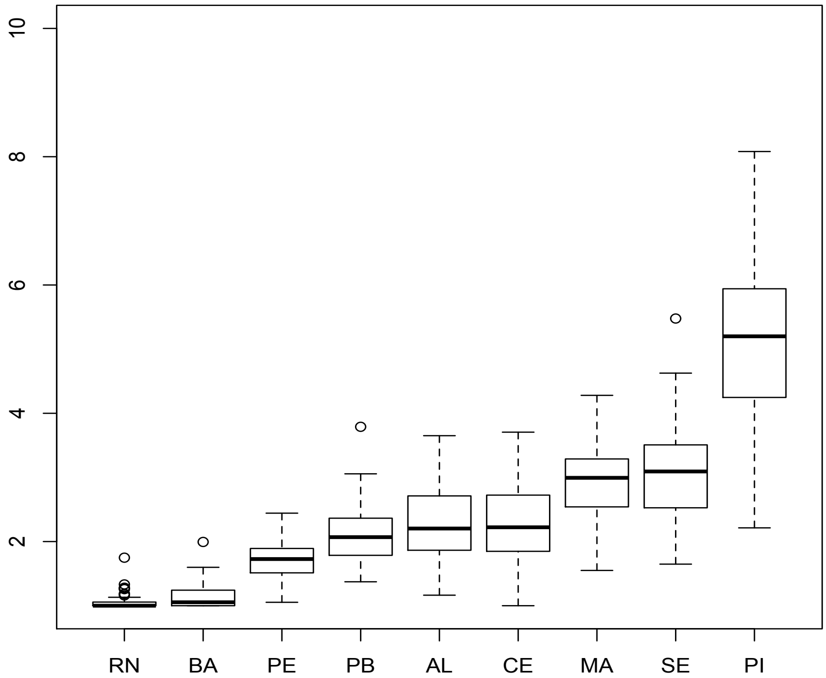


**Figure 7.8** – Distribution of Cluster 3 bootstrapped CESs (motor vehicle)

In Cluster 1, the best-performing state is SP, since it presents the *CES* closest to 1.00. RJ, DF and RS constitute the second-best-performing group of states. MG, SC and PR show relatively poor performances. PR and BA represent the one-but-last performing group. ES sustains the worst position.

In Cluster 2, AP and AM are clearly the best-performing states. PA, AC, GO and MS represent a second group with inferior performances in relation to the leading DMUs. MT and RO present even worse performances. Finally, TO and RR constitute the worst performing pair.

In Cluster 3, RN and BA form the best-performing group. PE, PB, AL and CE represent an average performance group. MA, SE and remarkably PI, show extremely poor performances.

Now, individually for Clusters 1, 2 and 3, Tables 7.17 to 7.19 exhibit the current number of traffic fatalities and the target values resulting from the benchmarking process inside the respective clusters. The added value now consists of delivering more specific targets, based on the median values, concerning the number of fatalities of motor vehicle occupants. In order to provide more and less ambitious targets, the tables also contain a range of target values varying between the 1st quartile and the 3rd quartile.

**Table 7.17** – Traffic fatalities current number, target and target range for Cluster 1 states (motor vehicle)

| State | Number of fatalities | | |
|---|---|---|---|
| | Current | Target | Target range |
| SP | 4247 | 4247 | [3,387; 4,247] |
| DF | 372 | 224 | [127; 224] |
| RJ | 1320 | 760 | [530; 1,073] |
| RS | 1394 | 640 | [455; 990] |
| MG | 3286 | 839 | [565; 1,249] |
| SC | 1420 | 331 | [220; 528] |
| PR | 2387 | 421 | [280; 559] |
| ES | 880 | 90 | [57; 123] |

**Table 7.18** – Traffic fatalities current number, target and target range for Cluster 2 states (motor vehicle)

| State | Number of fatalities | | |
|---|---|---|---|
| | Current | Target | Target range |
| AP | 53 | 53 | [46; -] |
| AM | 228 | 212 | [144; -] |
| PA | 697 | 234 | [171; 314] |
| AC | 98 | 28 | [21; 39] |
| GO | 1415 | 357 | [305; 452] |
| MS | 608 | 152 | [113; 191] |
| MT | 880 | 200 | [160; 248] |
| RO | 476 | 88 | [68; 104] |
| TO | 436 | 61 | [43; 73] |
| RR | 106 | 15 | [10; 17] |

**Table 7.19** – Traffic fatalities current number, target and target range for Cluster 3 states (motor vehicle)

| State | Number of fatalities | | |
|---|---|---|---|
| | Current | Target | Target range |
| RN | 481 | 481 | [462; -] |
| BA | 1961 | 1961 | - |
| PE | 1377 | 806 | [734; 920] |
| PB | 659 | 321 | [282; 372] |
| AL | 422 | 200 | [162; 234] |
| CE | 1428 | 656 | [537; 791] |
| MA | 1071 | 372 | [331; 441] |
| SE | 481 | 158 | [140; 193] |
| PI | 866 | 174 | [150; 208] |

Best-performing states as, for instance, SP, DF, AP, RN and BA do not present a less ambitious target, since this value would be equal to its current number of fatalities; the more ambitious target for these states results from the few bootstrapping iterations in which they were not considered as the best-performing DMU within the cluster.

In the next section, we briefly show a possibility for combining two DEA model results (in terms of traffic fatality target numbers) in order to obtain an indication of the possible targets for non-motorized users. Since each DEA model creates a specific efficiency frontier (as we explained in Chapter 5), and consequently, a specific reference for performance evaluation, they are not supposed to present combination properties. However, the intention here is mainly to provide some basic insight in the non-motorized users situation, instead of simply ignoring it in this research scope.

## 7.4. TARGET FOR NON-MOTORIZED USERS

Pedestrians consist of 27.87% of traffic fatalities in Brazil assuming the average values in the period 2010-2012; cyclists correspond to 4.46%. Together, as non-motorized users, pedestrians and cyclists represent a 32.33% share of the traffic fatalities in the country – thus, a considerable part of the road safety problem that should not be neglected. Figure 7.9 shows the "per state" share of non-motorized users in the total of specified transportation mode fatalities. In some states, non-motorized users account for more than 50% of the specified traffic fatalities.

Examining the probability of and degree to which pedestrians are exposed to "pedestrian x vehicle" fatal accident risk has important implications for formulating effective road safety measures, but this piece of information is often very difficult and costly to collect (Lam, Yao and Loo, 2014). In the United States, for example, although there is no consensus on how to ultimately define the metric and the best methods to implement the metric for pedestrian exposure data collection, pedestrian exposure research has been conducted for decades (Kennedy, 2008). Unfortunately, the same is not true in the Brazilian case.

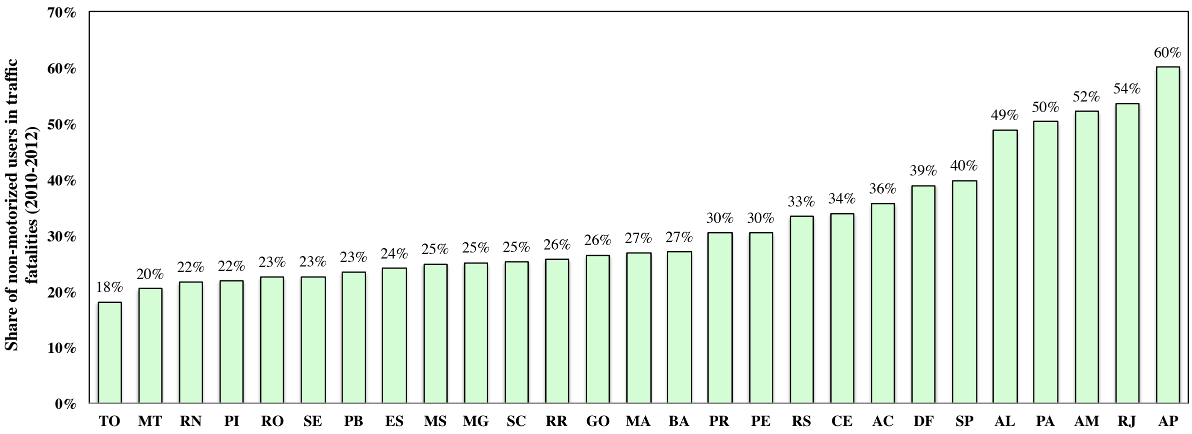



**Figure 7.9** – Share of non-motorized users in traffic fatalities

### 7.4.1 DEA MODEL RESULTS COMBINATION

In this manuscript, we describe the application of six input-output (exposure-fatality) DEA efficiency score models for the investigation of road safety concerning the six corresponding perspectives they provide: all road users (in an aggregated form using the total number of traffic fatalities – Chapter 6), motorized users (dealing with the traffic fatalities corresponding to motor vehicle occupants – Chapter 7) and then we built four models intending to assess the individual performance of each motor vehicle type – motorcycle, car, truck and bus (Chapter 8). Every DEA model used in this research generated one experiment, with the suggestion of a benchmark states(s) in terms of road safety; based on this, the DEA model measured the relative performance of the states allocated to the same cluster.

From the results of each different model that we applied, after a sensitivity analysis procedure (bootstrapping), we derived targets for the number of traffic fatalities in the corresponding six specification levels: for all users, motor vehicle occupants, motorcycle occupants, car occupants, truck occupants and bus occupants. These targets are not precise information, but will probably vary within a range of possibilities: from less ambitious, passing to an intermediary value, up to a more ambitious target. This range creation is possible thanks to the sensitivity analysis step, which enabled the exploration of different possible sceneries on the theme under investigation.

Due to limitations in exposure data availability for pedestrians and cyclists, these users were not directly/individually addressed within the scope of this research. Nevertheless, throughout the combination of the results of two of the models we applied, it is possible to obtain at least some insight in the non-motorized users situation. In other words, since we have a general model considering the entire set of traffic fatalities (including non-motorized users) – developed in Chapter 6 and a more specific model considering only the motorized vehicle occupant fatalities (not including non-motorized users) – developed in Chapter 6, by exclusion, it is possible to get some indication about possible non-motorized users' targets.

In practical terms, the subtraction of the motor vehicle related traffic fatality targets set in Chapter 7 from the general traffic fatality targets of Chapter 6, in theory, would result in the target number of traffic fatalities for non-motorized users. Tables 7.20 to 7.22 contain these values for Cluster 1 to 3, respectively. The computations are mostly based on the median of the bootstrapped values for the sets of targets. In the cases in which the combination of the model results did not provide a realistic target through the median value comparison, we extended the target suggestion to the quartile range or to the standard deviation comparison.

**Table 7.20** – Process of target setting for non-motorized users – Cluster 1

| State | Number of traffic fatalities | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | | | Motor vehicle occupants | | | Non-motorized users | | |
| | Current | % reduction | Target | Current | % reduction | Target | Current | % reduction | Target |
| ES | 1,160 | -81% | 218 | 880 | -90% | 90 | 279 | -54% | 128 |
| MG | 4,390 | -58% | 1,847 | 3,286 | -74% | 839 | 1,104 | -9% | 1,009 |
| RJ | 2,842 | -53% | 1,322 | 1,320 | -42% | 760 | 1,522 | -63% | 562 |
| SP | 7,055 | 0% | - | 4,247 | 0% | - | 2,808 | 0% | - |
| PR | 3,428 | -74% | 893 | 2,387 | -82% | 421 | 1,042 | -55% | 471 |
| RS | 2,092 | -33% | 1,402 | 1,394 | -54% | 640 | 698 | -76% | 169* |
| SC | 1,900 | -64% | 683 | 1,420 | -77% | 331 | 480 | -27% | 352 |
| DF | 609 | -43% | 349 | 372 | -40% | 224 | 237 | -47% | 126 |

*Target value based on the $Q_1$ values

**Table 7.21** – Process of target setting for non-motorized users – Cluster 2

| State | Number of traffic fatalities | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | | | Motor vehicle occupants | | | Non-motorized users | | |
| | Current | % reduction | Target | Current | % reduction | Target | Current | % reduction | Target |
| AC | 153 | -34% | 101 | 98 | -71% | 28 | 55 | -7% | 51* |
| AP | 131 | -30% | 92 | 53 | 0% | - | 78 | -50% | 39 |
| AM | 477 | 0% | - | 228 | -7% | 212 | 248 | -3% | 240** |
| PA | 1,401 | -53% | 660 | 697 | -66% | 234 | 704 | -39% | 426 |
| RO | 614 | -53% | 287 | 476 | -82% | 88 | 138 | -24% | 105** |
| RR | 142 | -56% | 63 | 106 | -86% | 15 | 36 | -11% | 32** |
| TO | 532 | -64% | 190 | 436 | -86% | 61 | 96 | -26% | 71** |
| GO | 1,923 | -28% | 1,388 | 1,415 | -75% | 357 | 509 | - | - |
| MT | 1,106 | -43% | 632 | 880 | -77% | 200 | 226 | - | - |
| MS | 810 | -32% | 553 | 608 | -75% | 152 | 201 | - | - |

*Target value based on the $Q_1$ values **Target value based on the σ values

**Table 7.22** – Process of target setting for non-motorized users – Cluster 3

| State | Number of traffic fatalities | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | | | Motor vehicle occupants | | | Non-motorized users | | |
| | Current | % reduction | Target | Current | % reduction | Target | Current | % reduction | Target |
| AL | 825 | -70% | 244 | 422 | -53% | 200 | 403 | -89% | 44 |
| BA | 2,685 | -32% | 1,835 | 1,961 | 0% | - | 724 | -17% | 599** |
| CE | 2,160 | -58% | 918 | 1,428 | -54% | 656 | 731 | -64% | 262 |
| MA | 1,463 | -62% | 555 | 1,071 | -65% | 372 | 392 | -53% | 183 |
| PB | 860 | -47% | 453 | 659 | -51% | 321 | 201 | -34% | 133 |
| PE | 1,979 | -43% | 1,127 | 1,377 | -41% | 806 | 602 | -47% | 321 |
| PI | 1,108 | -77% | 250 | 866 | -80% | 174 | 242 | -69% | 76 |
| RN | 613 | 0% | - | 481 | 0% | - | 132 | -32% | 90** |
| SE | 622 | -65% | 218 | 481 | -67% | 158 | 141 | -57% | 60 |

**Target value based on the σ values

The compiled information about targets for the number of traffic fatalities (overall and per category) is available in the state profiles in Appendix A.

# 8. DISAGGREGATED MODELS FOR PERFORMANCE EVALUATION AND TARGET SETTING

In the previous chapters, we carried out different approaches for evaluating the road safety situation and setting the target number of traffic fatalities for the Brazilian states allocated to their respective clusters. Firstly, we introduced a more general model, which computed an overall traffic fatality target. Then, considering disaggregated exposure measures we presented an innovative DEA model application, which provided a focus on the evaluation of motor vehicle occupants fatalities (risk analysis and target setting). In this chapter, we consider the same four vehicle categories and their respective exposure and fatalities data (see Appendix B and Appendix H) again, however now in individual models (as we indicate in the structures presented in Figure 8.1). Appendix J contains one example of the model scripts for car occupants related evaluation.

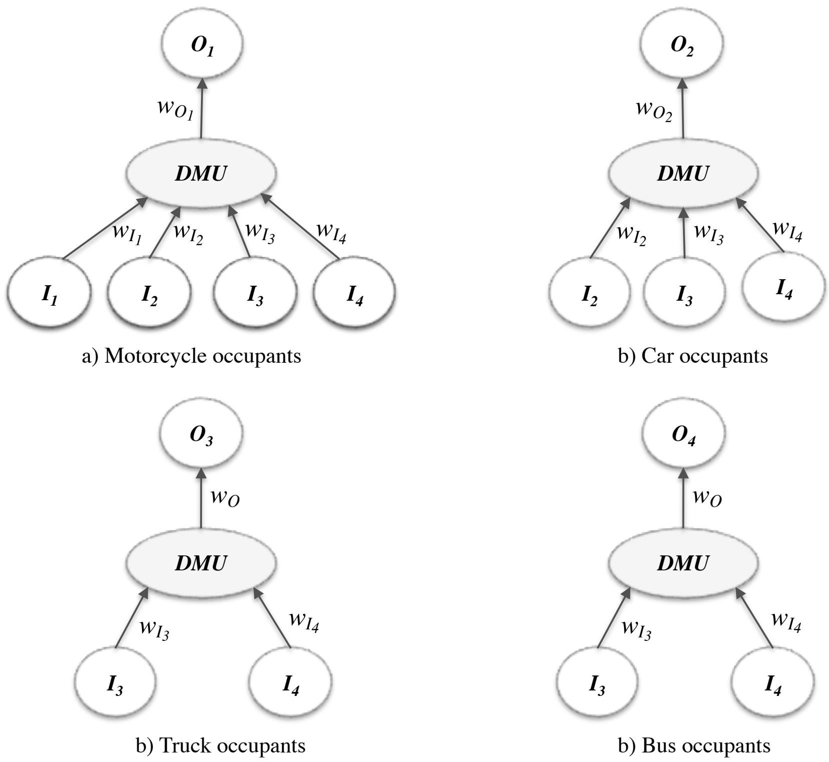



**Figure 8.1** – Model structures for assessing the disaggregated road safety risk level

Thus, the main goal of this procedure is to deliver more detailed risk evaluations and traffic fatality targets. The method contains the same procedures as applied in the previous models: DEA-based efficiency evaluation and sensitivity analysis with bootstrapping.

Figure 8.1 shows the structure of inputs (VKT by the four different motor vehicle categories – motorcycle – $I_1$, car – $I_2$, truck – $I_3$ and bus – $I_4$) and outputs (fatalities of occupants of the respective four categories $O_1$, $O_2$, $O_3$ and $O_4$). The corresponding weights for the inputs are $w_{I1},\dots, w_{I4}$ and for the outputs $w_{O1},\dots, w_{O4}$.

## 8.1. MOTORCYCLE

In terms of road safety, within the group of motorized vehicles, the motorcycle occupant is most vulnerable in relation to any other vehicle (a car, a truck or a bus). To capture the motorcyclist's fatality risk, not only its own VKT has to be considered but also that of the other motor vehicle categories.

### 8.1.1 RISK EVALUATION

Tables 1.1 to 1.3 exhibit the $OES$, the $CES$, the average CES value ($\overline{CES}$), the CES standard deviation ($\sigma_{CES}$), the median CES value ($\widetilde{CES}$), the CES 1st quartile value ($Q_{1_{CES}}$) and the CES 3rd quartile value ($Q_{3_{CES}}$) for the entire set of 50 bootstrapped scores. The values are ordered according to the $\widetilde{CES}$.

**Table 8.1** – Motorcycle related OES and CES for Cluster 1 states

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| SP | 1.000 | 1.009 | 1.203 | 0.331 | 1.006 | 1.000 | 1.237 |
| MG | 1.000 | 1.007 | 1.379 | 0.426 | 1.200 | 1.000 | 1.688 |
| RS | 1.036 | 1.053 | 1.489 | 0.500 | 1.338 | 1.061 | 1.830 |
| RJ | 1.288 | 1.352 | 2.342 | 0.896 | 2.121 | 1.547 | 3.021 |
| DF | 1.227 | 1.314 | 2.367 | 0.769 | 2.252 | 1.816 | 2.751 |
| SC | 1.477 | 1.503 | 3.195 | 1.201 | 3.129 | 2.341 | 4.198 |
| PR | 1.871 | 1.922 | 5.106 | 1.767 | 5.224 | 3.613 | 6.033 |
| ES | 2.191 | 2.261 | 6.870 | 2.175 | 6.821 | 5.228 | 8.454 |

**Table 8.2** – Motorcycle related OES and CES for Cluster 2 states

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| AP | 1.000 | 1.000 | 1.012 | 0.046 | 1.000 | 1.000 | 1.000 |
| AM | 1.132 | 1.209 | 1.461 | 0.363 | 1.363 | 1.207 | 1.627 |
| GO | 1.503 | 1.525 | 2.510 | 0.595 | 2.370 | 2.152 | 2.906 |
| MT | 1.498 | 1.533 | 2.490 | 0.582 | 2.416 | 2.088 | 2.870 |
| AC | 1.497 | 1.530 | 2.550 | 0.673 | 2.558 | 1.984 | 3.062 |
| MS | 1.521 | 1.533 | 2.578 | 0.694 | 2.598 | 2.070 | 2.965 |
| PA | 1.567 | 1.603 | 2.781 | 0.743 | 2.663 | 2.320 | 3.070 |
| RO | 1.731 | 1.789 | 3.479 | 0.812 | 3.351 | 2.887 | 4.105 |
| TO | 1.768 | 1.816 | 3.530 | 0.833 | 3.626 | 2.944 | 4.012 |
| RR | 2.214 | 2.269 | 5.672 | 1.418 | 5.547 | 4.790 | 6.206 |

**Table 8.3** – Motorcycle related OES and CES for Cluster 3 states

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| BA | 1.000 | 1.000 | 1.035 | 0.135 | 1.000 | 1.000 | 1.000 |
| RN | 1.431 | 1.441 | 2.256 | 1.234 | 1.898 | 1.424 | 2.767 |
| PE | 1.849 | 1.858 | 3.963 | 1.976 | 3.476 | 2.676 | 4.662 |
| MA | 1.950 | 1.986 | 4.551 | 2.315 | 3.788 | 2.716 | 5.664 |
| PB | 1.967 | 1.987 | 4.305 | 2.235 | 3.858 | 3.019 | 5.191 |
| CE | 1.917 | 1.944 | 4.611 | 2.812 | 3.937 | 2.391 | 5.917 |
| SE | 2.602 | 2.616 | 8.922 | 4.975 | 7.223 | 5.665 | 10.746 |
| PI | 2.826 | 2.877 | 9.283 | 5.958 | 7.702 | 5.538 | 11.266 |
| AL | 2.699 | 2.710 | 9.192 | 4.923 | 9.058 | 5.036 | 11.729 |

The boxplot diagrams in Figures 8.2 to 8.4 provide an overview of the distribution of all computed CESs through the indication of the minimum, 1$^{st}$ quartile, median, 3$^{rd}$ quartile and maximum CES's values for Clusters 1, 2 and 3, respectively. The dots represent outlier observations.

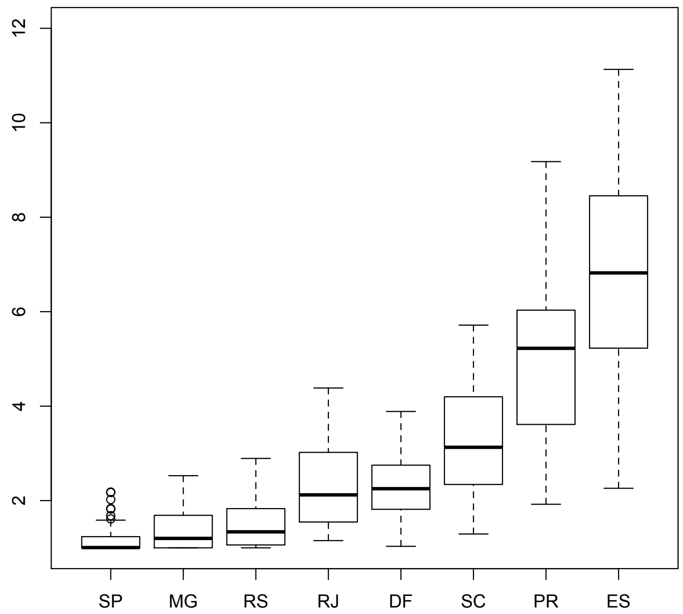



**Figure 8.2** – Distribution of Cluster 1 bootstrapped CESs (motorcycle)

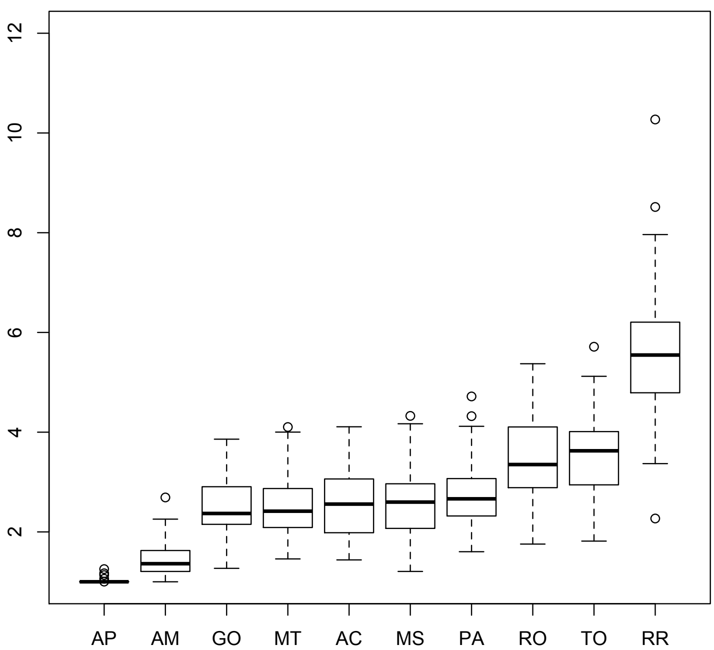



**Figure 8.3** – Distribution of Cluster 2 bootstrapped CESs (motorcycle)

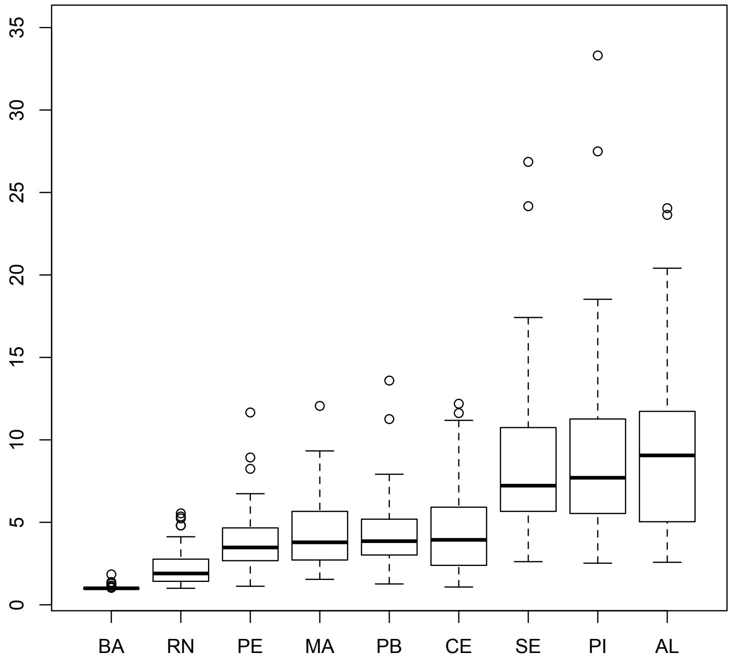



**Figure 8.4** – Distribution of Cluster 3 bootstrapped CESs (motorcycle)

### 8.1.2 TARGET SETTING

Tables 8.4 to 8.6 exhibit the $OES$, the $CES$, the average CES value ($\overline{CES}$), the standard deviation ($\sigma_{CES}$), the median CES value ($\widetilde{CES}$), the CES 1st quartile value ($Q_{1_{CES}}$) and the CES 3rd quartile value ($Q_{3_{CES}}$) for the entire set of bootstrapped scores. The values are ordered according to the $\widetilde{CES}$.

**Table 8.4** – Motorcycle occupants: traffic fatalities current number, target and target range for Cluster 1

| State | Number of fatalities – motorcycle occupants | | | |
|---|---|---|---|---|
| | Current | % needed reduction | Target | Target range |
| SP | 2,220 | 0% | - | [1,828; -] |
| MG | 1,026 | -15% | 876 | [625; -] |
| RS | 565 | -25% | 426 | [313; 537] |
| RJ | 655 | -50% | 326 | [230; 448] |
| DF | 138 | -52% | 67 | [54; 84] |
| SC | 666 | -67% | 216 | [161; 289] |
| PR | 1,045 | -81% | 202 | [177; 293] |
| ES | 412 | -85% | 62 | [50; 80] |

This paragraph shows by means of an example how to interpret the target tables we present. SP is the benchmark regarding reduction of motorcycle occupants fatality risk. If RJ wants to become as efficient as SP within Cluster 1 with respect to motorcycle occupant fatality risk, it needs to decrease its number of car occupant fatalities by 15%. RS needs to decrease it by 25%, and so on for the rest of the states, until ES, which needs to reduce its motorcycle occupants fatalities by 85%, all assuming no change in exposure.

**Table 8.5** – Motorcycle occupants: traffic fatalities current number, target and target range for Cluster 2

| State | Number of fatalities – motorcycle occupants | | | |
|---|---|---|---|---|
| | Current | % needed reduction | Target | Target range |
| AP | 33 | 0% | 33 | - |
| AM | 145 | -22% | 114 | [95; 128] |
| GO | 696 | -57% | 298 | [243; 328] |
| MT | 457 | -58% | 194 | [163; 225] |
| AC | 67 | -60% | 27 | [22; 35] |
| MS | 301 | -61% | 117 | [102; 146] |
| PA | 503 | -62% | 193 | [167; 222] |
| RO | 286 | -69% | 88 | [72; 102] |
| TO | 214 | -72% | 61 | [55; 75] |
| RR | 76 | -82% | 14 | [12; 16] |

**Table 8.6** – Motorcycle occupants: traffic fatalities current number, target and target range for Cluster 3

| State | Number of fatalities – motorcycle occupants | | | |
|---|---|---|---|---|
| | Current | % needed reduction | Target | Target range |
| BA | 669 | 0% | 669 | - |
| RN | 319 | -47% | 169 | [116; 226] |
| PE | 904 | -71% | 261 | [197; 341] |
| MA | 738 | -73% | 198 | [133; 278] |
| PB | 476 | -74% | 125 | [93; 159] |
| CE | 1,034 | -74% | 266 | [177; 439] |
| SE | 338 | -86% | 47 | [32; 60] |
| PI | 677 | -86% | 92 | [62; 126] |
| AL | 356 | -89% | 40 | [30; 71] |

As in the previous analysis, best performing states as SP, MG, AP, and BA do not present a less ambitious target, since this value would be equal to its current number of motorcycle occupant fatalities; the more ambitious target for these states results from the few bootstrapping iterations in which they were not considered as the best-performing DMU.

## 8.2. CAR

Cars are vulnerable in relation to other cars, trucks and buses. Therefore, the number of car occupant fatalities is put against not only cars VKT, but also trucks and buses VKT.

### 8.2.1 RISK EVALUATION

Tables 1.7 to 1.9 exhibit the $OES$, the $CES$, the average CES value ($\overline{CES}$), the standard deviation ($\sigma_{CES}$), the median CES value ($\widetilde{CES}$), the CES 1st quartile value ($Q_{1_{CES}}$) and the CES 3rd quartile value ($Q_{3_{CES}}$) for the entire set of bootstrapped scores. The values are ordered according to the ($\widetilde{CES}$).

**Table 8.7** – Car related OES and CES for Cluster 1 states

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| SP | 1.000 | 1.000 | 1.903 | 1.600 | 1.000 | 1.000 | 2.290 |
| RJ | 1.186 | 1.207 | 2.870 | 3.075 | 1.512 | 1.000 | 2.988 |
| DF | 1.599 | 1.620 | 6.124 | 8.249 | 2.652 | 1.836 | 8.342 |
| RS | 1.850 | 1.875 | 9.266 | 10.121 | 4.930 | 2.864 | 11.537 |
| SC | 2.366 | 2.401 | 15.534 | 18.164 | 6.287 | 4.060 | 20.627 |
| PR | 2.894 | 2.946 | 23.427 | 25.244 | 11.966 | 8.483 | 30.581 |
| MG | 3.464 | 3.557 | 33.973 | 33.576 | 20.824 | 10.135 | 48.038 |
| ES | 4.395 | 4.452 | 60.272 | 63.611 | 35.653 | 19.455 | 69.542 |

**Table 8.8** – Car related OES and CES for Cluster 2 states

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| AP | 1.000 | 1.000 | 2.560 | 2.990 | 1.203 | 1.000 | 2.756 |
| AM | 1.000 | 1.014 | 2.927 | 3.259 | 1.659 | 1.000 | 2.485 |
| PA | 1.503 | 1.535 | 6.661 | 6.164 | 3.936 | 2.223 | 7.700 |
| AC | 1.957 | 1.984 | 10.968 | 15.103 | 5.058 | 3.725 | 8.793 |
| GO | 2.750 | 2.809 | 21.155 | 24.947 | 11.639 | 7.390 | 24.957 |
| MS | 2.984 | 3.007 | 23.226 | 23.671 | 14.589 | 7.104 | 33.110 |
| RR | 2.580 | 2.613 | 24.632 | 26.318 | 16.220 | 7.207 | 36.110 |
| MT | 3.380 | 3.435 | 34.295 | 33.797 | 20.726 | 11.683 | 43.581 |
| RO | 3.617 | 3.692 | 37.958 | 37.493 | 20.997 | 13.158 | 58.370 |
| TO | 5.159 | 5.261 | 85.259 | 83.727 | 54.076 | 27.208 | 132.803 |

**Table 8.9** – Car related OES and CES for Cluster 3 states

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| AL | 1.000 | 1.000 | 1.027 | 0.087 | 1.000 | 1.000 | 1.000 |
| RN | 1.698 | 1.708 | 3.634 | 2.325 | 3.380 | 2.193 | 4.139 |
| PB | 1.920 | 1.933 | 4.347 | 3.077 | 3.708 | 2.036 | 5.881 |
| PE | 2.040 | 2.044 | 5.161 | 3.317 | 4.448 | 2.811 | 6.407 |
| SE | 2.167 | 2.172 | 5.642 | 3.660 | 4.724 | 3.191 | 6.864 |
| CE | 2.332 | 2.339 | 6.567 | 4.597 | 5.959 | 3.398 | 7.863 |
| PI | 2.971 | 2.987 | 11.464 | 8.182 | 9.372 | 5.450 | 13.853 |
| MA | 3.341 | 3.351 | 14.009 | 11.625 | 10.436 | 5.107 | 19.721 |
| BA | 3.761 | 3.798 | 18.871 | 16.011 | 14.503 | 9.307 | 20.545 |

The boxplots diagrams in Figures 8.5 to 1.7 provide an overview of the distribution of all computed CISs through the indication of the minimum, 1st quartile, median, 3rd quartile and maximum CIS's values for Clusters 1, 2 and 3, respectively. The dots represent the outlier observations.

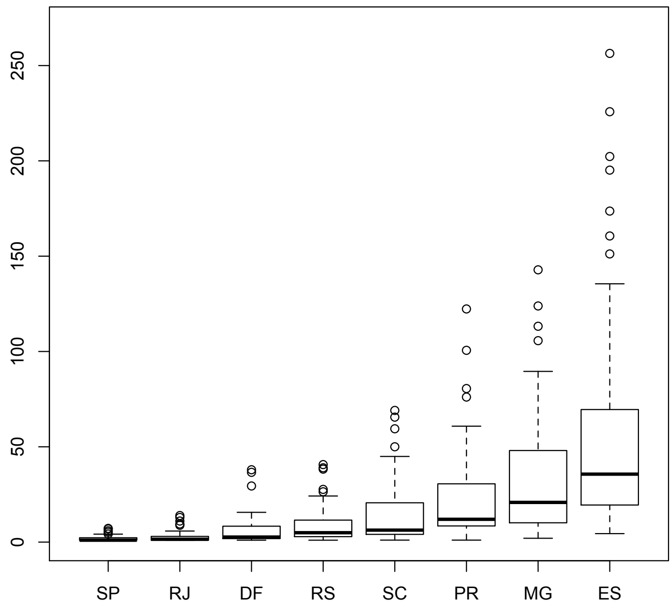



**Figure 8.5** – Distribution of Cluster 1 bootstrapped CESs (car)

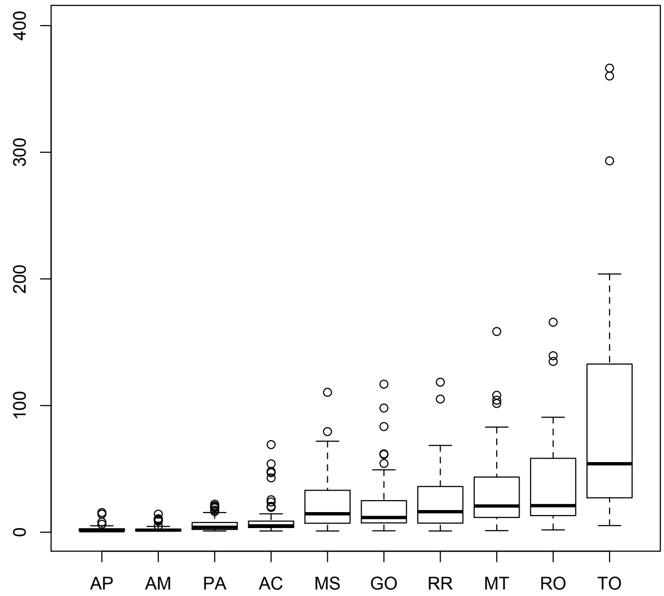



**Figure 8.6** – Distribution of Cluster 2 bootstrapped CESs (car)

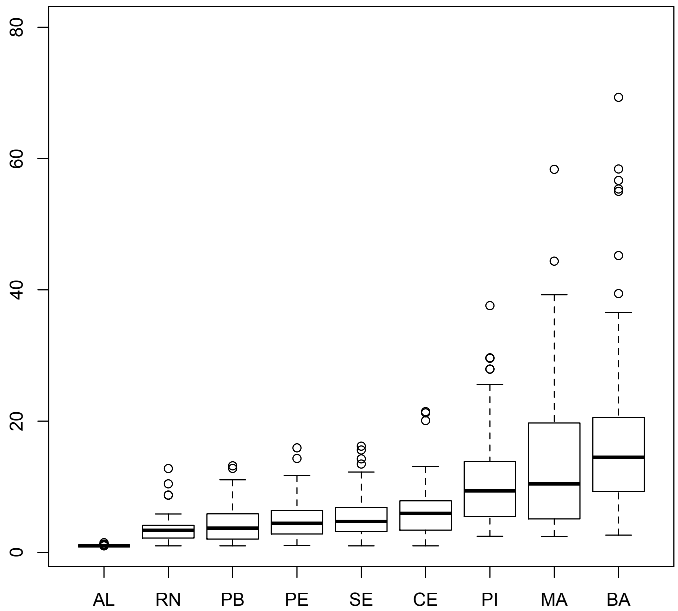



**Figure 8.7** – Distribution of Cluster 3 bootstrapped CESs (car)

### 8.2.2 TARGET SETTING

Tables 1.10 to 1.12 exhibit the *OES*, the *CES*, the average CES value ($\overline{CES}$), the standard deviation ($\sigma_{CES}$), the median CES value ($\widetilde{CES}$), the CES 1st quartile value ($Q_{1_{CES}}$) and the CES 3rd quartile value ($Q_{3_{CES}}$) for the entire set of bootstrapped scores. The values are ordered according to the ($\widetilde{CES}$).

**Table 8.10** – Car occupants: traffic fatalities current number, target and target range for Cluster 1

| State | Number of fatalities – car occupants | | | |
|---|---|---|---|---|
| | **Current** | **% needed reduction** | **Target** | **Target range** |
| SP | 1,816 | 0% | - | [800; -] |
| RJ | 587 | -33% | 396 | [200; -] |
| DF | 225 | -61% | 87 | [27; 123] |
| RS | 741 | -80% | 152 | [65; 262] |
| SC | 682 | -84% | 109 | [33; 170] |
| PR | 1,222 | -92% | 103 | [40; 146] |
| MG | 2,086 | -95% | 103 | [45; 210] |
| ES | 448 | -97% | 13 | [7; 23] |

**Table 8.11** – Car occupants: traffic fatalities current number, target and target range for Cluster 2

| State | Number of fatalities – car occupants | | | |
|---|---|---|---|---|
| | **Current** | **% needed reduction** | **Target** | **Target range** |
| AP | 17 | -15% | 14 | [6; -] |
| AM | 74 | -38% | 46 | [31; -] |
| PA | 164 | -74% | 43 | [22; 76] |
| AC | 29 | -80% | 6 | [3; 8] |
| GO | 636 | -94% | 56 | [26; 88] |
| MS | 263 | -91% | 18 | [8; 37] |
| RR | 30 | -93% | 2 | [1; 4] |
| MT | 359 | -95% | 18 | [8; 32] |
| RO | 163 | -95% | 8 | [3; 13] |
| TO | 203 | -98% | 4 | [2; 8] |

**Table 8.12** – Car occupants: traffic fatalities current number, target and target range for Cluster 3

| State | Number of fatalities – car occupants | | | |
|---|---|---|---|---|
| | **Current** | **% needed reduction** | **Target** | **Target range** |
| AL | 61 | 0% | 61 | - |
| RN | 156 | -70% | 47 | [38; 72] |
| PB | 169 | -73% | 46 | [29; 84] |
| PE | 429 | -77% | 97 | [67; 153] |
| SE | 120 | -79% | 25 | [18; 38] |
| CE | 374 | -83% | 63 | [48; 111] |
| PI | 166 | -89% | 18 | [12; 31] |
| MA | 289 | -90% | 28 | [15; 57] |
| BA | 1233 | -93% | 86 | [61; 134] |

As in the previous analysis, best performing states as SP, RJ, AP, AM and AL do not present a less ambitious target, since this value would be equal to its current number of car occupants fatalities; the more ambitious target for these states results from the few bootstrapping iterations in which they were not considered the best performing DMU.

## 8.3. TRUCK

Trucks only are vulnerable in relation to other trucks or buses. Thus, the number of truck occupant fatalities is put against trucks VKT and buses VKT. Two states do not have any registered truck occupant fatality in the period, AP and RR; for this reason, we excluded these DMUs and we carried out the analysis with the 8 remaining states in cluster 2.

### 8.3.1 RISK EVALUATION

Tables 1.13 to 1.15 exhibit the $OES$, the $CES$, the average CES value ($\overline{CES}$), the standard deviation ($\sigma_{CES}$), the median CES value ($\widetilde{CES}$), the CES 1st quartile value ($Q_{1_{CES}}$) and the CES 3rd quartile value ($Q_{3_{CES}}$) for the entire set of bootstrapped scores. The values are ordered according to the ($\widetilde{CES}$).

**Table 8.13** – Truck related OES and CES for Cluster 1 states

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| DF | 1.000 | 1.000 | 1.001 | 0.007 | 1.000 | 1.000 | 1.000 |
| SP | 1.391 | 1.393 | 2.082 | 0.709 | 1.819 | 1.603 | 2.467 |
| MG | 1.726 | 1.811 | 3.425 | 1.099 | 3.364 | 2.766 | 3.972 |
| RS | 1.819 | 1.907 | 4.086 | 1.656 | 3.659 | 3.117 | 4.595 |
| SC | 1.941 | 2.111 | 4.571 | 1.642 | 4.260 | 3.535 | 5.022 |
| PR | 1.998 | 2.144 | 4.977 | 2.157 | 4.338 | 3.527 | 6.168 |
| ES | 2.331 | 2.424 | 6.015 | 2.430 | 5.415 | 4.543 | 7.132 |
| RJ | 2.324 | 2.762 | 7.222 | 2.812 | 6.732 | 5.220 | 8.643 |

**Table 8.14** – Truck related OES and CES for Cluster 2 states

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| AC | 1.000 | 1.000 | 2.151 | 2.376 | 1.000 | 1.000 | 2.408 |
| AM | 2.007 | 2.007 | 26.897 | 47.670 | 4.186 | 2.026 | 25.065 |
| PA | 2.235 | 2.235 | 25.112 | 46.432 | 7.302 | 2.583 | 24.932 |
| RO | 5.543 | 5.543 | 174.231 | 318.805 | 31.551 | 8.724 | 197.893 |
| TO | 4.019 | 4.019 | 104.586 | 158.363 | 34.963 | 6.212 | 133.089 |
| GO | 6.268 | 6.268 | 251.314 | 364.094 | 62.247 | 23.372 | 316.024 |
| MS | 7.084 | 7.084 | 336.734 | 484.193 | 80.708 | 23.532 | 467.795 |
| MT | 7.111 | 7.111 | 273.865 | 393.554 | 114.931 | 43.786 | 316.993 |

**Table 8.15** – Truck related OES and CES for Cluster 3 states

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| RN | 1.000 | 1.000 | 6.164 | 10.754 | 1.312 | 1.000 | 3.933 |
| CE | 1.587 | 1.587 | 13.933 | 26.958 | 3.437 | 1.000 | 9.951 |
| BA | 1.998 | 2.034 | 21.236 | 38.945 | 5.756 | 2.403 | 18.061 |
| AL | 1.965 | 2.017 | 31.048 | 53.731 | 6.931 | 2.250 | 24.132 |
| PB | 3.157 | 3.163 | 60.340 | 100.307 | 15.137 | 3.783 | 63.892 |
| PE | 3.124 | 3.132 | 72.254 | 107.040 | 25.710 | 10.823 | 84.756 |
| PI | 6.176 | 6.198 | 304.286 | 558.956 | 58.363 | 18.700 | 261.618 |
| MA | 7.198 | 7.217 | 345.292 | 555.298 | 74.120 | 15.207 | 500.220 |
| SE | 8.483 | 8.621 | 548.843 | 679.879 | 151.453 | 69.534 | 925.156 |

The boxplots diagrams in Figures 8.8 to 8.10 provide an overview of the distribution of all computed CISs through the indication of the minimum, 1$^{st}$ quartile, median, 3$^{rd}$ quartile and maximum CIS's values for Clusters 1, 2 and 3, respectively. The dots represent the outlier observations.

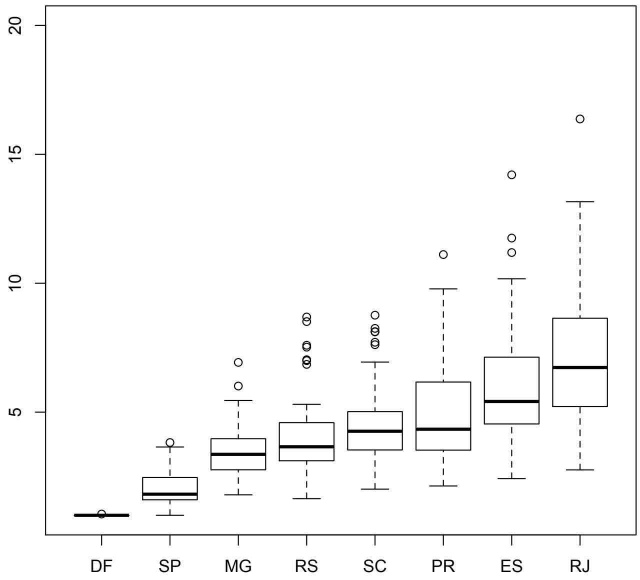



**Figure 8.8** – Distribution of Cluster 1 bootstrapped CESs (truck)

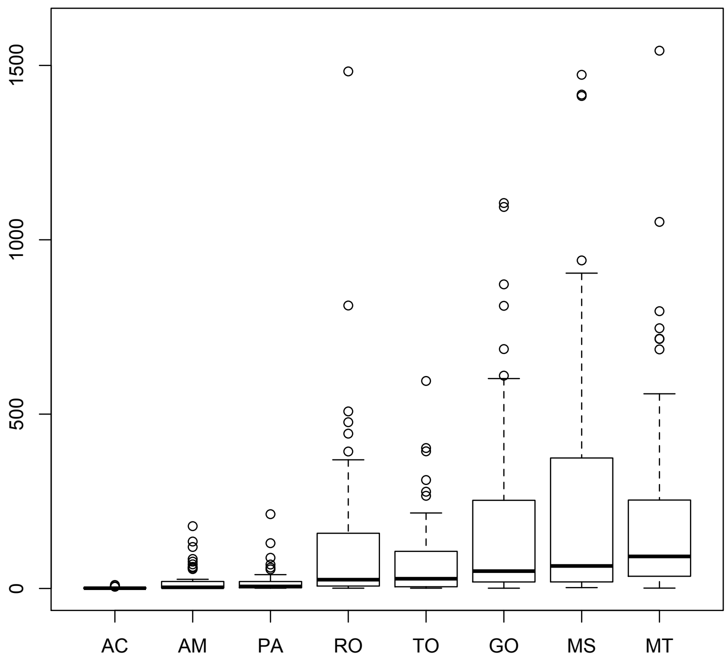



**Figure 8.9** – Distribution of Cluster 2 bootstrapped CESs (truck)

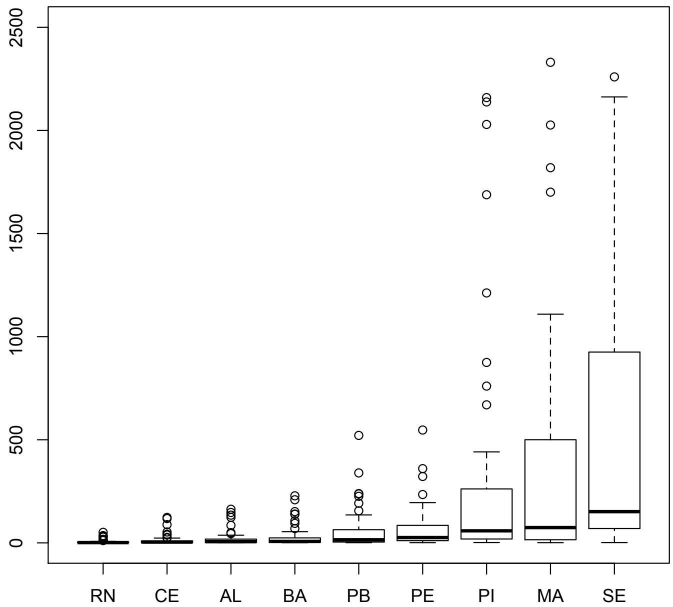



**Figure 8.10** – Distribution of Cluster 3 bootstrapped CESs (truck)

### 8.3.2 TARGET SETTING

Tables 1.16 to 1.18 exhibit the $OES$, the $CES$, the average CES value ($\overline{CES}$), the standard deviation ($\sigma_{CES}$), the median CES value ($\widetilde{CES}$), the CES 1st quartile value ($Q_{1_{CES}}$) and the CES 3rd quartile value ($Q_{3_{CES}}$) for the entire set of bootstrapped scores. The values are ordered according to the ($\widetilde{CES}$).

**Table 8.16** – Truck occupants: traffic fatalities current number, target and target range for Cluster 1

| State | Number of fatalities – truck occupants | | | |
|---|---|---|---|---|
| | Current | % needed reduction | Target | Target range |
| DF | 6 | 0% | 6 | - |
| SP | 164 | -45% | 90 | [67, 103] |
| MG | 146 | -69% | 45 | [38, 55] |
| RS | 76 | -72% | 22 | [17, 26] |
| SC | 57 | -75% | 14 | [12, 17] |
| PR | 102 | -75% | 25 | [18, 31] |
| ES | 21 | -81% | 4 | [3, 5] |
| RJ | 60 | -82% | 11 | [9, 14] |

**Table 8.17** – Truck occupants: traffic fatalities current number, target and target range for Cluster 2

| State | Number of fatalities – truck occupants | | | |
|---|---|---|---|---|
| | Current | % needed reduction | Target | Target range |
| AC | 1 | 0% | 1 | [0, -] |
| AM | 7 | -76% | 2 | [0, 4] |
| PA | 20 | -86% | 3 | [1, 8] |
| RO | 23 | -97% | 1 | [0, 3] |
| TO | 14 | -97% | 0 | [0, 2] |
| GO | 66 | -98% | 1 | [0, 3] |
| MS | 41 | -99% | 1 | [0, 2] |
| MT | 61 | -99% | 1 | [0, 1] |

**Table 8.18** – Truck occupants: traffic fatalities current number, target and target range for Cluster 3

| State | Number of fatalities – truck occupants | | | |
|---|---|---|---|---|
| | Current | % needed reduction | Target | Target range |
| RN | 4 | -23% | 3 | [1, -] |
| CE | 12 | -71% | 3 | [1, -] |
| BA | 43 | -82% | 8 | [2, 19] |
| AL | 5 | -85% | 1 | [0, 2] |
| PB | 11 | -93% | 1 | [0, 3] |
| PE | 35 | -96% | 1 | [0, 3] |
| PI | 21 | -98% | 0 | [0, 1] |
| MA | 34 | -99% | 0 | [0, 2] |
| SE | 23 | -99% | 0 | [0, -] |

Given the contrasting situations inside Clusters 2 and 3, very ambitious and maybe unrealistic targets were set. Best-performing states as SP, RJ, AP, AM and AL do not present a less ambitious target, since this value would be equal to its current number of truck occupants fatalities; the more ambitious target for these states results from the few bootstrapping iterations in which they were not considered the best performing DMU.

## 8.4. BUS

Buses are only vulnerable in relation to other buses or trucks. Thus, the number of bus occupants fatalities depends not only of bus VKT, but also of trucks VKT.

### 8.4.1 RISK EVALUATION

Tables 8.13 to 8.15 exhibit the $OES$, the $CES$, the average CES value ($\overline{CES}$), the standard deviation ($\sigma_{CES}$), the median CES value ($\widetilde{CES}$), the CES 1st quartile value ($Q_{1_{CES}}$) and the CES 3rd quartile value ($Q_{3_{CES}}$) for the entire set of bootstrapped scores. The values are ordered according to the ($\widetilde{CES}$).

**Table 8.19** – Bus related OES and CES for Cluster 1 states

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| ES | 1.000 | 1.000 | 1.099 | 0.281 | 1.000 | 1.000 | 1.000 |
| MG | 5.370 | 5.550 | 157.904 | 279.301 | 33.034 | 14.303 | 134.436 |
| RJ | 5.530 | 5.875 | 196.126 | 383.815 | 35.548 | 16.784 | 180.665 |
| SP | 5.000 | 5.053 | 94.981 | 205.154 | 25.178 | 6.903 | 99.385 |
| PR | 6.304 | 6.612 | 186.453 | 406.257 | 46.095 | 16.488 | 183.120 |
| RS | 4.692 | 4.847 | 195.463 | 307.606 | 22.791 | 6.029 | 306.947 |
| SC | 9.928 | 10.535 | 591.227 | 910.588 | 176.694 | 81.922 | 778.172 |
| DF | 4.648 | 4.701 | 117.376 | 217.602 | 40.624 | 9.113 | 94.000 |

**Table 8.20** – Bus related OES and CES for Cluster 2 states

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| AC | 3.576 | 3.581 | 36.401 | 58.622 | 14.692 | 7.905 | 36.889 |
| AP | 6.585 | 6.606 | 147.662 | 192.872 | 60.638 | 33.407 | 204.717 |
| AM | 1.567 | 1.576 | 9.916 | 11.502 | 4.641 | 1.557 | 14.856 |
| PA | 2.692 | 2.702 | 29.738 | 45.177 | 9.807 | 4.778 | 35.277 |
| RO | 2.920 | 2.921 | 41.127 | 60.045 | 15.640 | 5.888 | 61.858 |
| RR | 1.972 | 1.978 | 14.008 | 20.713 | 6.726 | 3.682 | 16.203 |
| TO | 4.665 | 4.671 | 91.665 | 148.116 | 37.667 | 12.797 | 92.399 |
| GO | 5.042 | 5.052 | 78.777 | 145.902 | 28.444 | 13.364 | 64.501 |
| MT | 1.000 | 1.000 | 2.140 | 1.964 | 1.000 | 1.000 | 2.738 |
| MS | 2.272 | 2.275 | 22.114 | 29.709 | 5.935 | 2.451 | 31.579 |

**Table 8.21** – Bus related OES and CES for Cluster 3 states

| State | OES | CES | $\overline{CES}$ | $\sigma_{CES}$ | $\widetilde{CES}$ | $Q_{1_{CES}}$ | $Q_{3_{CES}}$ |
|---|---|---|---|---|---|---|---|
| AL | 1.475 | 1.498 | 18.100 | 32.109 | 5.600 | 1.114 | 18.520 |
| BA | 2.862 | 2.870 | 64.898 | 82.060 | 19.774 | 8.256 | 84.703 |
| CE | 4.584 | 4.623 | 181.202 | 293.477 | 77.481 | 14.032 | 229.210 |
| MA | 8.970 | 9.058 | 756.086 | 1151.596 | 249.271 | 97.717 | 904.690 |
| PB | 3.635 | 3.661 | 139.128 | 211.419 | 59.452 | 13.105 | 195.477 |
| PE | 3.149 | 3.179 | 79.785 | 116.694 | 31.094 | 5.233 | 116.750 |
| PI | 2.261 | 2.284 | 27.205 | 43.373 | 5.182 | 2.422 | 37.114 |
| RN | 1.392 | 1.404 | 22.719 | 31.589 | 10.277 | 2.078 | 26.564 |
| SE | 1.000 | 1.000 | 5.563 | 8.262 | 1.764 | 1.000 | 4.816 |

The boxplots diagrams in Figures 8.11 to 1.13 provide an overview of the distribution of all computed CESs through the indication of the minimum, 1st quartile, median, 3rd quartile and maximum CES's values for Clusters 1, 2 and 3, respectively. The dots represent the outlier observations.

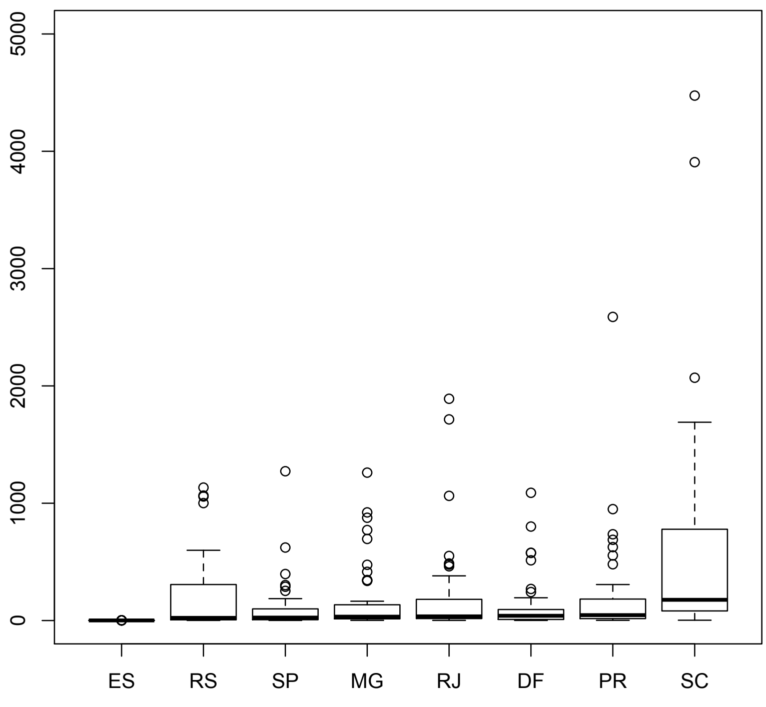



**Figure 8.11** – Distribution of Cluster 1 bootstrapped CESs (bus)

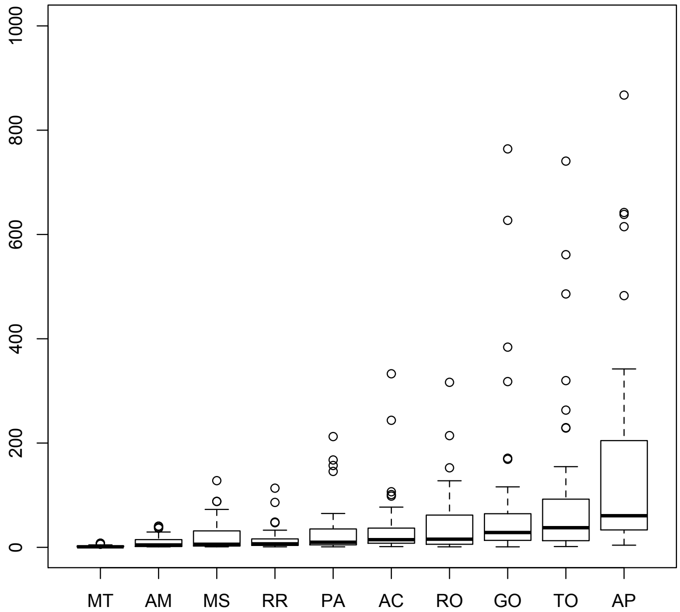



**Figure 8.12** – Distribution of Cluster 2 bootstrapped CESs (bus)

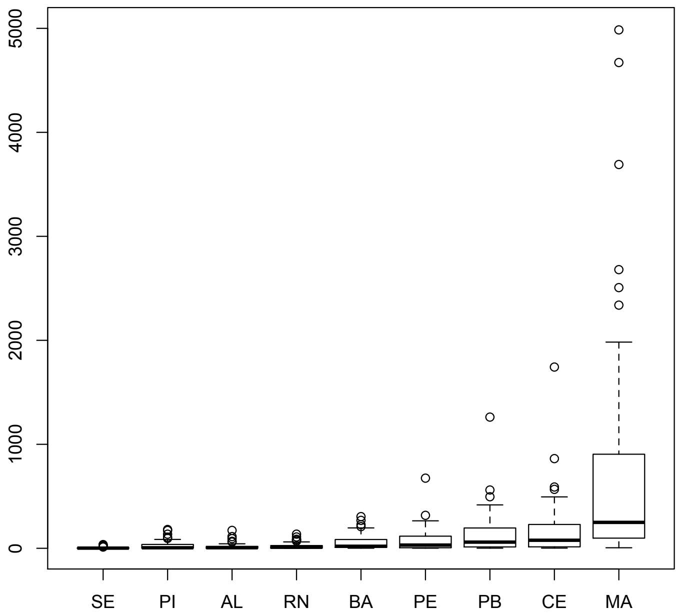



**Figure 8.13** – Distribution of Cluster 3 bootstrapped CESs (bus)

### 8.4.2 TARGET SETTING

Tables 1.16 to 1.18 exhibit the $OES$, the $CES$, the average CES value ($\overline{CES}$), the standard deviation ($\sigma_{CES}$), the median CES value ($\widetilde{CES}$), the CES 1st quartile value ($Q_{1_{CES}}$) and the CES 3rd quartile value ($Q_{3_{CES}}$) for the entire set of bootstrapped scores. The values are ordered according to the $\widetilde{CES}$.

**Table 8.22** – Bus occupants: traffic fatalities current number, target and target range for Cluster 1

| State | Number of fatalities | | | |
|---|---|---|---|---|
| | Current | % needed reduction | Target | Target range |
| ES | 1 | 0% | 1 | - |
| RS | 12 | -96% | 1 | [0; 2] |
| SP | 47 | -96% | 2 | [0;7] |
| MG | 28 | -97% | 1 | [0; 2] |
| RJ | 17 | -97% | 1 | [0; 1] |
| DF | 2 | -98% | 0 | [0; 0] |
| PR | 17 | -98% | 0 | [0; 1] |
| SC | 15 | -99% | 0 | [0; 0] |

**Table 8.23** – Bus occupants: traffic fatalities current number, target and target range for Cluster 2

| State | Number of fatalities | | | |
|---|---|---|---|---|
| | Current | % needed reduction | Target | Target range |
| MT | 2 | 0% | 2 | [1; -] |
| AM | 2 | -78% | 0 | [0; 1] |
| MS | 4 | -83% | 1 | [0; 0] |
| RR | 0 | -85% | 0 | [0; 2] |
| PA | 9 | -90% | 1 | [0; 0] |
| AC | 1 | -93% | 0 | [0; 1] |
| RO | 4 | -94% | 0 | [0; 0] |
| GO | 17 | -96% | 1 | [0; 1] |
| TO | 5 | -97% | 0 | [0; 1] |
| AP | 2 | -98% | 0 | [0; 2] |

**Table 8.24** – Bus occupants: traffic fatalities current number, target and target range for Cluster 3

| State | Number of fatalities | | | |
|---|---|---|---|---|
| | Current | % needed reduction | Target | Target range |
| SE | 1 | -43% | 0 | [0; -] |
| PI | 2 | -81% | 0 | [0; 1] |
| AL | 1 | -82% | 0 | [0; 1] |
| RN | 1 | -90% | 0 | [0; 1] |
| BA | 16 | -95% | 1 | [0; 2] |
| PE | 8 | -97% | 0 | [0; 2] |
| PB | 3 | -98% | 0 | [0; 0] |
| CE | 8 | -99% | 0 | [0; 1] |
| MA | 10 | -100% | 0 | [0; 0] |

Due to the contrasting situations inside Clusters 2 and 3 and their small absolute numbers, again very ambitious and maybe unrealistic targets were set. As in the previous analysis, best performing states as ES, MT and SE do not present a less ambitious target, since this value would be equal to its current number of bus occupants fatalities; the more ambitious target for these states results from the few bootstrapping iterations in which they were not considered the best performing DMU.

## 8.5. CHAPTER CONCLUSIONS

The context for the risk evaluation and consequently the target setting process may substantially vary according to the vehicle category we consider. For example, regarding motorcycle occupants, MG presents the 2$^{nd}$ lowest risk in comparison to the rest of Cluster 1; when we consider car occupants, MG presents the 2$^{nd}$ highest risk. These meaningful changes in risk evaluation occur for other states as well, stressing the importance of a proper disaggregation level in road safety analysis.

The percentage reductions shown in the tables indicate how contrasting the road safety level is between the vehicle category occupants in the different states. Therefore, in this chapter we provide more specific target suggestions for the states, thereby giving additional insights. Every evaluation we performed has its own applicability for decision makers depending on the investigation perspective: if it is more broad, we suggest a more general diagnosis and target setting; if it is more specific, we are also able to offer a more detailed, disaggregated diagnosis and target setting.

# 9. SAFETY PERFORMANCE INDICATORS

The monitoring of the road safety level in Brazil and particularly in its member states traditionally focuses on the available traffic fatality rates. However, this approach does not offer practical suggestions on how to improve road safety. In this context, the usage of Safety Performance Indicators (SPIs) is a growing and promising tendency. In this chapter, we start from the set of SPIs selected and described in Chapter 3 as good indicators of the operational conditions of the road traffic system in each of the 27 Brazilian states and representing a set of hierarchically structured SPIs related to the three road safety (risk) domains (road user, environment and vehicle). This chapter aims to describe this innovative SPI research in Brazil, using data envelopment analysis to aggregate the SPIs into a composite indicator, as well as to show the potential for supporting future improvements on the theme. The results consist of graphical representations expressing the overall performance (based on the composite safety performance indicator) and disaggregated performance (detailed per road safety domain) for all Brazilian states adequately divided into comparable clusters.

## 9.1. THE CONCEPT OF SPI

From the strategic road safety planning point of view, the traditionally available traffic fatality rates, such as fatalities per inhabitant or per registered vehicle, are the type of information most frequently used for monitoring the road safety level throughout the country and particularly in its various member states. Although very useful for a primary diagnosis and global view on the situation, the availability of fatality rates does not provide any suggestion on how to tackle the situation and improve road safety (Golob et al., 2004). In this context, the usage of Safety Performance Indicators (SPIs) in road safety research is continuously growing to complement and provide extra information in addition to the traditional outcome-based diagnosis using traffic fatalities or injuries (Shen, 2012).

SPIs are measures reflecting the operational conditions of the road traffic system, which influence the system's safety performance. The intention is to prevent the occurrence of problems at an early stage (before these problems result in accidents); in other words, an SPI serves as an assisting tool in assessing the current safety conditions, monitoring the evolution, measuring impacts of various safety interventions, making comparisons, and other purposes (e.g. Vis et al., 2005; Hakkert and Gitelman, 2007). Some examples of SPIs are: the seatbelt wearing rate, the average speed on a particular road, the percentage of drivers who drink and drive and how often they do that, the average rescue time after an accident, etc. In summary, an SPI should manifest real operational conditions and/or the power of remedial post accident measures.

In the pyramid framework that symbolizes the variety of elements that contribute to the road safety situation (in the end expressed by the social costs of killed and injured people), SPIs act as intermediate outcomes, one layer below the final outcomes, as shown in Figure 9.1, based on the adaptation of Koornstra et al. (2002) from LTSA (2000).

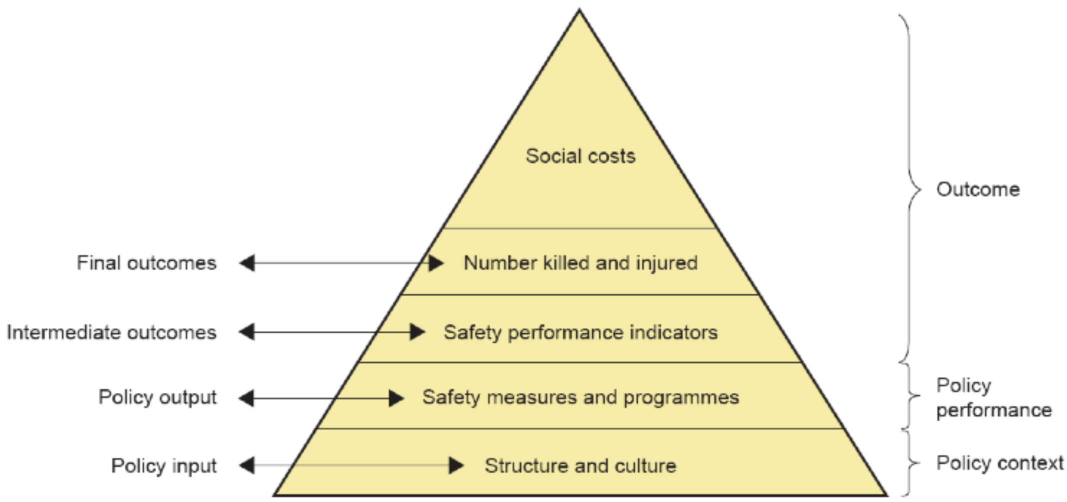



**Figure 9.1** – Target hierarchy for road safety (Koornstra et al., 2002)

SPIs are usually formulated with respect to the triad "road user – road – vehicle", which is the standard decomposition of a road safety problem (WHO, 2004). In a broader interpretation of the road related aspect, the more recent literature labels it as an environment related domain with the inclusion of trauma management aspects as well (e.g. Shen, 2012).

Some national and international research tried to quantify the share in which each domain influences the occurrence of an accident; in this sense, Table 9.1 shows a comparison between a North-American, a European and a Brazilian[1] study on the theme.

**Table 9.1** – Comparison of the shares of each road safety domain in accident causation

| Road safety domain | U.S.A (Wierwille et al., 2002) | Europe (Rumar, 1985) | Brazil (Scaringella et al., 2012) |
|---|---|---|---|
| Road user | 65% | 57% | 44% |
| Road/Environment | 2% | 3% | 0% |
| Vehicle | 2% | 2% | 0% |
| Road user + Road/Environment | 25% | 27% | 29% |
| Road user + Vehicle | 5% | 6% | 19% |
| Road/Environment + Vehicle | 0% | 1% | 0% |
| Road user + Road/Environment Road + Vehicle | 1% | 3% | 8% |
| Overall Road user | 96% | 93% | 100% |
| Overall Road/Environment | 26% | 34% | 37% |
| Overall Vehicle | 6% | 12% | 27% |

Table 9.1 illustrates that the North-American and European results present more similarities, while the Brazilian results indicate a higher overall contribution of the road user. The inferior standards of roads and trauma management in Brazil might explain the higher value attributed to the overall road/environment domain. Complementarily, the more unfavorable economic background in Brazil has resulted in an older vehicle fleet, poorer vehicle maintenance and inferior standards on vehicle technology. In addition, although the share of the road user is remarkable in all the three cases, it should not be used as an excuse to transfer all the responsibility of an accident to the road user, since appropriate strategies in the scope of the two remaining domains will contribute to decrease the chances of human error (Ferraz et al., 2012). Moreover, due to the complexity of finding a general reliable quantification of the share of each domain in accident causation, Table 9.1 should be considered as an indication of the possible reality.

Despite the level of complexity of a road safety explanatory framework, due to the demand for translating this knowledge into objective guidelines for road safety, it is apparently more valuable to invest in key indicators more closely related to the problem under investigation. In general, according to the European literature on SPIs, the following aspects by road safety domain are usually covered in research on the theme (Morsink et al., 2005; Hakkert and Gitelman, 2007; Vis and Eksler, 2008; Wegman et al., 2008; Hermans, 2009; Shen et al., 2011): Road user: drink-and-drive, speeding, and protective systems wearing rates (helmet

[1] Although more careful investigations concerning this theme are still needed in Brazil.

and seatbelt); Road/environment: road network, road design, quality of medical treatment, arrival time of emergency services at the place of the accident; Vehicle: fleet composition, fleet age and crashworthiness.

It is however challenging for researchers to formulate good SPIs and obtain reliable real-world data. In a worldwide perspective, the data availability in Europe can be considered as very good, but this relatively favorable situation does not guarantee the entire availability of the 'ideal' indicators, and the use of 'proxy' indicators is therefore already necessary as an alternative to enable the research on SPIs. An additional data problem is that, besides not being the ideal indicators, some of them are not available for the entire set of entities under evaluation (usually geographical areas of countries or states). It requires the usage of missing data imputation techniques in order to guarantee that all the entities can be taken into account in the analysis. Shen (2012) and Bax et al. (2012), for instance, applied multiple imputation techniques to complete the dataset for the investigation of SPIs in European countries.

In Brazil and other developing nations this challenge is even bigger, since a data collection process oriented towards SPIs investigation rarely exists. This frequently pushes researches to the dilemma of "doing nothing" (with the argument that there is not enough information available) or "working under some limited assumptions". This chapter describes SPI research in which this judgment had to be made due to the Brazilian unfavorable background scenery on data availability. It requires the adoption of alternative indicators, which are not the ideal ones, but represent satisfying indicators for which data exist, in most cases collected for different purposes (not necessarily for the transportation field) e.g. for public health studies or roadway infrastructure inventory. Therefore, the idea is to use this information as proxy variables, manifesting the desirable but unavailable data to a satisfying degree and under certain assumptions. Moreover, this research also intends to demonstrate the potentialities of SPI research in terms of delivering a valuable indication on how to effectively invest efforts towards road safety, and it might encourage the collection of more adequate indicators in the near future.

A great amount of possible SPIs exist, yet having a selected group of strategic SPIs is more appropriate for gaining valuable road safety interpretations and insights. Moreover, in order to extract an overall idea of the road safety operational performance in a certain region under investigation, it is also valuable to aggregate the diversity of SPIs into a single index, or composite indicator – CI, which intends to represent the intermediate outcomes of the target hierarchy for road safety presented in Figure 9.1. The construction of the CI is based on a data envelopment analysis procedure, in which for each Brazilian state (the decision making units) the most favorable combination of SPIs and corresponding weights is determined. The creation of such a safety performance CI offers crucial input information for decision makers for designing general policies for the nation, as well as specific ones adapted to the context of each member state or region.

## 9.2. SPI DATA SOURCES AND AVAILABLE DATA IN BRAZIL

The search for adequate data for SPIs research is a challenging issue, since there is quite a large gap between the theoretically ideal information, the information that is possible to monitor and collect considering the available collection methods and budget and, lastly, the available information. It makes it (almost) impossible to obtain a complete set of ideal SPIs for the road safety domains mentioned in Section 9.1.

Therefore, the alternative solution is to work with proxy indicators, which are intended to manifest the real situation regarding each road safety aspect under assessment. The next sections describe the available Brazilian sources for the collection of this kind of knowledge on a nationwide basis for all states, using the classification into road user, environmental (road plus health system) and vehicle related information. They all refer to information for the period 2010-2012. Additionally, it introduces some necessary assumptions for the appropriate interpretation of the available data.

### 9.2.1 ROAD USER RELATED INDICATORS

In terms of road user behavioral indicators in traffic, there are few researches at the national level that provide applicable information.

Yet, there is the 1st National Adolescent School-based Health Survey (PeNSE), conducted by means of a partnership between the Ministry of Health and the Brazilian Institute of Geography and Statistics (IBGE). The study investigated different risk and protection factors of the 9th grade[2] scholar health in 26 Brazilian capitals and the Federal District (Penna, 2010). A second example is the Emergency Department Injury Surveillance System (*Sistema de Vigilância de Violências e Acidentes*, VIVA), a new national surveillance system to supplement existing data sources implemented by Brazil's Ministry of Health. In this survey, there is the collection of data on injury-related cases seen in hospital emergency departments (Gawryszewski et al., 2008). Although the main focus of these researches is not road safety, there are some valuable questions concerning drink-and-drive habits and protective systems usage.

The Ministry of Transportation provides information related to protective systems, however for very specific conditions, such as the wearing rate of the seatbelt or helmet federal in highway accidents (*Ministério dos Transportes*, 2013a).

The State Traffic Departments (DETRANs), an entity of the National Traffic System, control the statistics on traffic infractions in its respective state. In general, the infractions data published by a DETRAN refers to the penalties applied by all institutions with this power: the police (federal or state); the DETRAN itself; the State Highways Department (DER); and also the traffic sector of the municipalities. Due to the absence of data concerning the road user behavior in traffic, the usage of these data under limited assumptions was the only alternative to enable the consideration of other important road user aspects beside alcohol and protective systems in the Brazilian SPI research. Therefore, we gathered data on infraction numbers related to the use of a cell-phone while driving and speeding.

The consideration of infraction data requires some key assumptions. Firstly, enforcement actions are based on the hope of detecting risk behaviors in a drivers sample from which the characteristics can be generalized to the entire drivers population. Following this simple logic, if a certain risk behavior presents a higher incidence in comparison to others, more efforts should be put in trying to prevent such attitude – it leads to the preliminary idea that the penalty rates are useful to identify the most common risk behavior in a certain state.

In the particular case of Brazil, it is important to stress that the infraction rates (e.g. per *capita* or per vehicle) should not be used to infer if a population has a higher risk behavior in comparison to another (i.e., considering the population of two states). The reason for that

[2] At this grade, the students' age is about 14 years.

relies on the fact that the value of these rates is fundamentally affected by the enforcement level; i.e. the higher the enforcement level, the larger the penalty rates. This assumption seems quite reasonable, since only utopic enforcement levels would be capable to detect all infracting actions (at least considering the Brazilian background scenery). As a consequence, all the infraction rates (*per capita* or per vehicle) present negative correlations with the traffic fatality rate of a state. So, if any state presents a high infraction rate, it does not necessarily mean that its drivers behave worse than others, but that the enforcement level is superior there, and, therefore, the road safety situation is better. According to the perspective provided in Figure 9.1, the enforcement level is considered as a policy output, which is actually part of the background for the safety performance in a place.

In a promising initiative, the National Road Safety Observatory (ONSV) performed a pilot research on random enforcement in a medium-size city of the state of São Paulo. The research consisted of observing the incidence of six common infracting behaviors on intersections: turning without signaling, non-use of seatbelt and helmet (by drivers and passengers), cell-phone usage, driving with the arm hanging out of the window and stop sign disobedience. The preliminary general results indicate a relation of one registered penalty for each 11.8 thousand real infracting actions (ONSV, 2013). Unfortunately, the limited scenario of application does not enable the use of the random enforcement research in this national level study.

### 9.2.2 ENVIRONMENT RELATED INDICATORS

The National Confederation of Transports (CNT) periodically publishes highway related information in the form of annual reports intending to evaluate the state of the Brazilian paved highways according to perceptible aspects for the users and the delivered road safety level. It is the largest study performed on highway infrastructure in the country, with the evaluation of about 90,000 km of paved highways located throughout the 27 federation units (CNT, 2011). This survey evaluates aspects such as the existence and type of central division, geometric design, pavement maintenance, signing, and roadside conditions.

Regarding data on health system related indicators, the source is the Ministry of Health, through its Information System on Health Public Budget (SIOPS) and the National Health Funding (FNS), for the information concerning the health expenditure and the Professional Councils (CONPROF) for the information concerning the number of health professionals (*Ministério da Saúde*, 2011c). The Unified Health System (SUS), created in 1988 under the subordination of the Ministry of Health, has been responsible for structuring and consolidating the public health system (CONASS, 2006). The existence of this kind of system was vital for uniform patterns on data collection implementation through the 27 federation units. Although it might be susceptible to failures in data collection, this system provides the most reliable information regarding public health issues in the country. The health expenditure value corresponds to the sum of federal, state and municipal values. The number of health professionals which are possibly involved in assisting traffic accident victims is available per state according to the following disaggregation: doctors, nurses and nursery technicians and auxiliaries.

### 9.2.3 VEHICLE RELATED INDICATORS

The National Traffic Department (DENATRAN), linked to the Ministry of Cities, is an entity of the National Traffic System (SNT) and the main official provider of motor vehicle fleet data. Although there are aspects indicating a certain superposition[3] level in DENATRAN's fleet registers – Losekann and Vilela (2010b) address an adequate discussion about the issue – it still is an attractive data source due to the disaggregation level of the information it provides: per state, per vehicle class and per vehicle age. Furthermore, some other aspects should prevail in order not to ignore this source: it is an official governmental entity; the information is easily available and constantly updated; and these possible distortions become less important when the analysis is comparative between the states. The available information consists of characteristics of the vehicle fleet, regarding its composition (being possible to identify the shares of cars, motorcycles, trucks and buses) and the age of the vehicle fleet.

## 9.3. INDICATORS SELECTION AND MODEL APPLICATION

### 9.3.1 INDICATORS SELECTION

Due to the limited options for the selection of indicators for the case of Brazil, it would not be a reasonable approach to straightforwardly base the choice of indicators on a single criterion. For this reason, in order to guide the process of choosing the most adequate SPIs we tried to maintain a good balance regarding the following aspects: correlation with the traffic fatality related rates, theoretical relationship with the road safety problem, objectivity of the indicator,

[3] In case of vehicle register state transfer, for example, it may account as a vehicle in registers of both states.

outlier detection and the level of missing observations. More details on these parameters are given in the following paragraphs. Intending to use similar data periods, we used the average values in the period 2010-2012, with exception of the infraction related indicators, for which we used the average in a larger period, 2009-2013 (to decrease the number of missing observations).

Since the SPIs research consists of analyzing factors capable to influence the occurrence of traffic accidents and consequently traffic fatalities, it is desirable that the chosen indicators present a high negative correlation with the undesired final outcomes(e.g. the traffic fatality or injury related rate). In the Brazilian scenario, the only reliable final outcome information available at the state level is the number of traffic fatalities, from which it is possible to compute the traffic fatality rates per *capita*, per vehicle fleet and per kilometer traveled. The measure of correlation is the Pearson product-moment or Spearman rank correlation coefficient (depending on the distribution of each indicator), and its statistical significance is tested at the 95% confidence level. Hence, there is a preference for indicators having a high statistically significant negative correlation coefficient with (at least one of) the undesired outcome measures.

Although high correlations are desirable, this criterion might sometimes be tricky, since there might exist confounding variables in this relationship and indicators highly theoretically associated with the road safety situation might present low correlation coefficients with the traffic fatality rates. One clear example of this situation occurs with the age of the vehicle fleet indicators. According to a straightforward interpretation of the correlation coefficient – the older the vehicle fleet in a state, the lower the traffic fatality rate – and, thus, the better the road safety situation. Conversely, in the theoretical relationship between the age of the fleet and road safety, keeping all other variables unchanged, an older vehicle fleet relates to a lower level of road safety (Page and Rackliff, 2006). In that case, the motorization stage of a state might be a confounding variable in the relationship between the age of the vehicle fleet and the traffic fatality related rate, given the fact that more economically developed states initiated their massive motorization process earlier, resulting in older vehicles in the current fleet. In conclusion, it means that if there is a clear theoretical relationship between one safety performance indicator and the road safety outcome situation, this indicator might be considered in spite of possibly presenting a low correlation coefficient.

Another important aspect to consider in the choice of SPIs is the level of objectiveness of the indicator; i.e., whether the indicator provides a clear and attainable suggestion of what should be done to improve the road safety situation and how changeable it is. It is possible to have an indicator that presents both a theoretical connection and a high correlation coefficient with the road safety situation, but its meaning does not reflect a reasonable suggestion on how to tackle the problem. The Gross Domestic Product *per capita* (GDPC) is a clear example of a theoretically related and highly correlated indicator, although it can be considered as non-to less-changeable (and for this reason it might not even be considered as an SPI). In other words, it is a quite generic guideline to suggest that in order to improve its road safety situation one state should increase its GDPC. Even though GDPC increases will probably lead to better road conditions, higher quality education and so on, decision makers will probably not put efforts in increasing the GDPC based on an expectation of improving road safety.

After considering the aforementioned aspects, we decided to use a set of 27 SPIs. Tables 9.2 to 9.5 contain the description of the chosen set of SPIs, representing the traditional distinction between road user, environment and vehicle related SPIs, as well as their corresponding subdivisions in subordinated domains.

**Table 9.2** – Road user behavior related SPIs

| Domain | Parameter | Min | Max | Unit | Data source |
|---|---|---|---|---|---|
| Alcohol | 1) Share of 9th year students who declared to be passenger in a vehicle with a driver who drunk | 18.40 | 31.50 | % | PENSE |
| | 2) Share of people involved in accident who declared to drink before | 0.08 | 0.29 | % | VIVA |
| Cell phone | 3) Cell phone related infractions per capita* | 0.08 | 20.37 | Inf./$10^5$ inh. | DETRAN/IBGE |
| | 4) Cell phone related infractions per vehicle* | 0.42 | 40.67 | Inf./ $10^5$ veh. | DETRAN/DENATRAN |
| Seatbelt | 5) Share of 9th year students who declared not to wear seatbelt in a car in the last 30 days (of those who were passenger in a car in this period) | 6.70 | 23.35 | % | PENSE |
| | 6) Share of people involved in car accidents wearing seatbelt | 42.90 | 86.00 | % | VIVA |
| Helmet | 7) Share of people involved in accident wearing helmet on federal highways | 80.57 | 100.00 | % | DPRF |
| | 8) Share of people involved in motorcycle accidents wearing helmet | 0.00 | 76.27 | % | VIVA |
| Speeding | 9) Speeding related infractions per capita* | 3.75 | 401.16 | Inf./inh. | DETRAN/IBGE |
| | 10) Speeding related infractions per vehicle* | 10.16 | 801.00 | Inf./veh. | DETRAN/DENATRAN |

*Imputed datasets

**Table 9.3** – Road infrastructure related SPIs (environment domain)

| 2nd level topic | 3rd level topic | 4th level topic | Parameter | Min | Max | Unit | Data source |
|---|---|---|---|---|---|---|---|
| Central division | - | - | 11) Share of multilane highways in relation to the paved highways | 0.00 | 17.06 | % | DNIT |
| Signing | Road markings | Central | 12) Share of highway length with central road markings | 69.17 | 99.43 | % | CNT |
| | | | 13) Share of highway length with clearly visible central road markings | 33.69 | 95.14 | % | |
| | | Lateral | 14) Share of highway length with lateral road markings | 43.77 | 95.40 | % | CNT |
| | | | 15) Share of highway length with clearly visible lateral road markings | 31.87 | 91.96 | % | |
| | Vertical signs | - | 16) Share of highway length with vertical signs | 29.43 | 86.16 | % | CNT |
| | | - | 17) Share of highway length with clearly visible vertical signs | 34.12 | 99.19 | % | |
| | | - | 18) Share of highway length with clearly legible vertical signs | 24.33 | 91.80 | % | |
| Roadside | Barriers | - | 19) Share of highway length equipped with adequate barriers | 1.00 | 73.10 | % | CNT |
| | Shoulders | - | 20) Share of highway length with adequate shoulders | 0.01 | 79.97 | % | |

**Table 9.4** – Health system related SPIs (environment domain)

| Domain | Parameter | Min | Max | Unit | Data source |
|---|---|---|---|---|---|
| Health professionals | 21) Number of medical doctors per capita | 0.53 | 3.62 | Doc./$10^3$ inh. | DATASUS/IBGE |
| | 22) Number of health professionals per capita | 4.64 | 16.85 | Prof./$10^3$ inh. | |
| Health expenditure | 23) Health expenditure per capita | 444.02 | 957.46 | R$/inh. | DATASUS/IBGE |

**Table 9.5** – Vehicle fleet related SPIs

| Domain | Parameter | Min | Max | Unit | Data source |
|---|---|---|---|---|---|
| Fleet composition | 24) Share of motorcycles in the total fleet | 11.25 | 56.41 | % | DENATRAN |
| | 25) Share of trucks in the total fleet | 1.76 | 5.92 | % | |
| Age of the fleet | 26) Share of 10-year vehicles or older | 22.00 | 56.27 | % | DENATRAN |
| | 27) Share of 5-year vehicles or newer | 25.83 | 56.34 | % | |

Figure 9.2 illustrates the hierarchical structure of the SPIs for the computation of the composite indicator (CI) for each Brazilian state. The numbers in Figure 9.2 correspond to the identification of each indicator according to Tables 9.2 to 9.5.

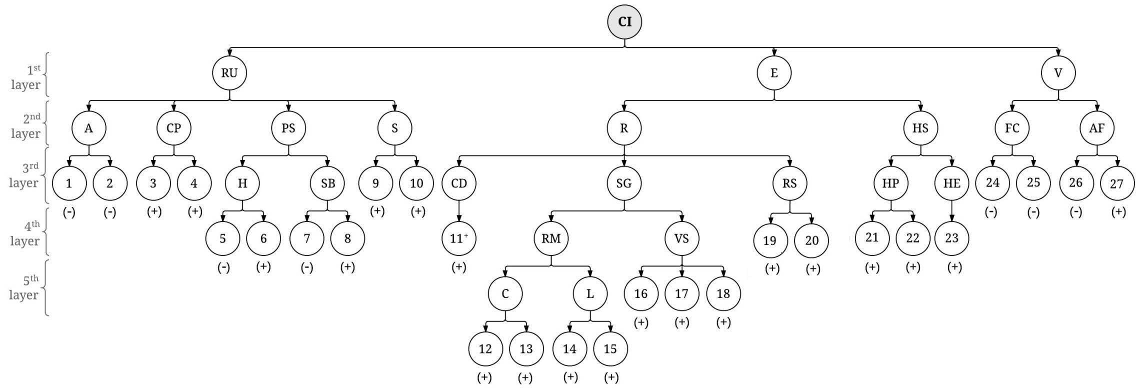



**Figure 9.2** – Hierarchical structure of the 27 SPIs used to compute the CI by means of the ML DEA-CI model

The hierarchical structure manifests the already mentioned traditional distinction between the road user (RU), the environment (E) and the vehicle (V) domain of SPIs. In the first domain (1st to 10th indicator), road user behavior decomposes into alcohol (A), cell-phone (CP), protective systems (PS) and speeding (S). Protective systems decompose into seatbelt (SB) and helmet (H). In the second domain, environmental SPIs decompose into road (R) and health system (HS). Road (11th to 20ht indicator), decomposes into central division (CD), signing (SG) and roadside (RS). Signing decomposes into road markings (RM) and vertical signs (VS). Road markings decompose into central (C) and lateral (L). Health system (21st to 23rd indicator) decomposes into health professionals (HP) and health expenditure (HE). In the third domain, vehicle related SPIs (24th to 27th indicator), decompose into fleet composition (FC) and age of the fleet (AF).

The (+) or (-) signals below each SPI indicate the direction of the normalization, i.e., (+) means the higher the SPI value, the better the road safety situation (positive correlation); (-) means the higher the SPI value, the worse the road safety situation (negative correlation).

To provide insight in the degree of mutual relationship that exists between the chosen indicators (SPIs plus a fatality rate) throughout all the 27 states, Figure 9.3 shows the Co-plot graphical representation of the set of SPIs.

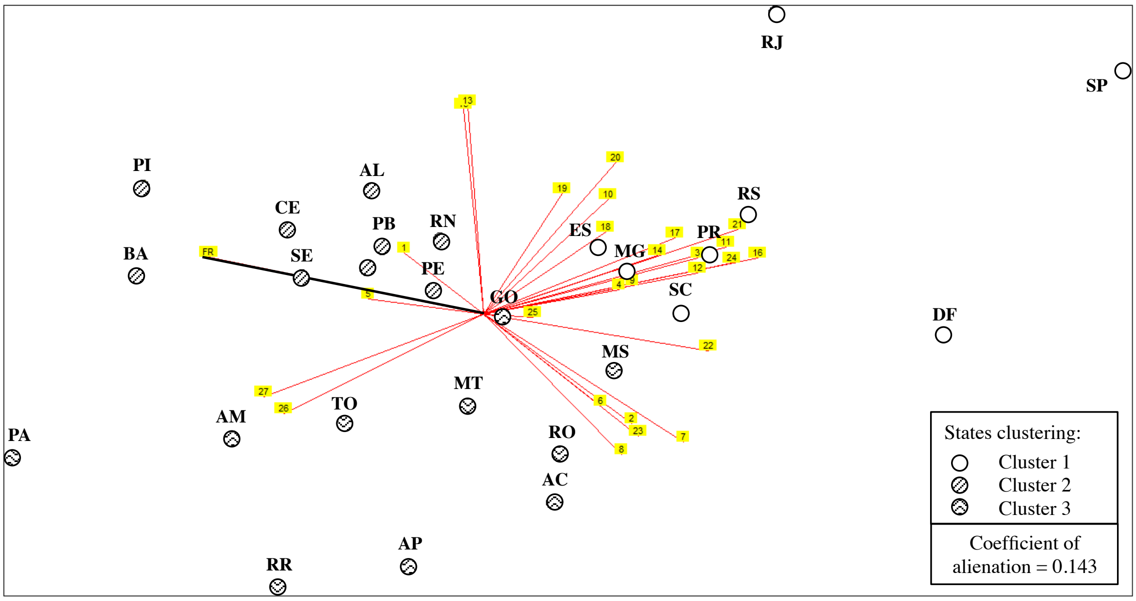



**Figure 9.3** – Co-plot graphical representation of the SPIs and traffic fatality related indicator (FR)

The construction of Figure 9.3 is based on the Co-plot method for representing multiple criteria in a two-dimensional space in which the location of each decision making unit (i.e. state) is determined by all criteria (SPIs) simultaneously. Talby and Raveh (1999) originally developed this technique and the software to build this visual representation. Some guidelines are necessary for the interpretation of Co-plot figures (Raveh, 2000; Adler and Raveh, 2008):

- Similar states are located closely to each other (e.g., PB is very similar to BA);

- Each arrow represents one SPI (1 to 27), and the cosines of the angles between the arrows are approximately proportional to the correlation between the SPIs. Therefore, a $0^o$ angle separating two arrows implies an extreme positive correlation between the corresponding SPIs, orthogonal arrows indicate slightly correlated indicators and a $180^o$ angle separating two arrows implies an extreme negative correlation between the respective SPIs (e.g., the pair of indicators 26 and 27 – both related to the age of the fleet –present a very high positive correlation while the pair of indicators 1 – related to alcohol – and 7 – related to seatbelt – present a vey high negative correlation);

- The closeness of the state's position and the direction that a certain arrow points indicates how good the state is with respect to that specific indicator, so the closer to the border of the figure, the better the performance (e.g.,, DF is a specialist in indicator 22, i.e., …).

Given that it is quite difficult to compile all this information in a single graphic representation, the software also provides two measures of goodness-of-fit. The first is the coefficient of alienation, which expresses the overall goodness-of-fit of the Co-plot graph, and its exhibited value in the graph is considered acceptable. Complementarily, each variable also has an individual measure of goodness-of-fit, which corresponds to how well it fits into the representation delivered by the Co-plot graph. Graphically, the length of the arrow denotes the indicators individual goodness-of-fit: the longer the arrow, the higher the goodness-of-fit.

Besides the values of the SPIs for the 27 Brazilian states we included one extra arrow, with the intention to show the correlation with an important final outcome indicator, namely the fatality rate – FR (traffic fatalities per registered vehicle). Moreover, the three different shading patterns for the circles representing each Brazilian state manifest the cluster division of the states as described in Chapter 4. The inclusion of the clusters representation in Figure 9.3 shows the correspondence between the identified similar states through the Co-plot on the one hand and the hierarchical clusters of Figure 9.2 on the other hand.

Figure 9.3 provides some important insights in the data set:

- Most of the SPIs present a positive correlation with the fatality rate indicator (FR) after adequately normalized;

- SPIs related to the age of the vehicle fleet (26 and 27) present a relatively strong negative correlation with FR, probably due to the confounding effect of different evolution patterns in the purchasing power of different Brazilian states for motorized vehicles;

- Depending on the position of the circles representing each state, it is possible to clearly identify three groups of states with similar road safety performance, which very much resembles the adopted cluster division.

### 9.3.2 MODEL APPLICATION

Equation 1.1 presents ML DEA-CI, in which the inputs to be maximized refer to the SPI values of each state. Therefore, all SPIs point in the "safe" direction; i.e., the larger the SPI value, the better the road safety situation. The score indicating the best performer(s) presents a value equal to 1 (as in the original model), since it succeeds in maximizing most efficiently its sustained road safety level; and underperforming DMUs present a score lower than 1, since they did not succeed in sufficiently maximizing their weighted inputs (or SPIs).

$$OIS_s = \max \sum_{j=1}^{q} w_{SPI_j} SPI_{j,s} \tag{1.1}$$

$$\begin{aligned} \textit{subject to} \quad & \sum_{j=1}^{q} w_{SPI_j} SPI_{j,r} \leq 1 \qquad r = 1,\dots,n \\ & w_{SPI_j} \geq 0 \quad j = 1,\dots,q \quad s = 1,\dots,n \end{aligned}$$

- $OIS_s$ – Optimum index score of the *s-th* DMU;
- $SPI_{j,s}$ – *j-th* SPI of the *s-th* DMU;
- $w_{SPI_j}$ – Weight attributed to $SPI_j$;
- $n$ – total number of DMUs;
- $q$ – total number of SPIs;

By solving Equation 1.2, the composite indicator based on a *K*-layered hierarchy of *q* SPIs can be calculated for each state *s*, where $u_{fK}$ is the weight given to the *f-th* category in the *K-th* layer and $w_{fK}^{(K)}$ denotes the non-negative internal weights associated with the SPIs of the *f-th* category in the *K-th* layer; the sum of all $w_{fK}^{(K)}$ within a particular category is equal to one.

$$OIS_s = \max \sum_{f_{K_1}=1}^{q^{(K)}} u_{f_K} \Big( \sum_{f_{K-1} \in A_K^{(K)}} w_{K-1}^{(K-1)} \Big( \dots \sum_{f_k \in A_{k+1}^{(k+1)}} w_k^{(k)} \Big( \dots \sum_{f_{k-1} \in A_{f_3}^{(3)}} w_2^{(2)} \Big( \sum_{f_1 \in A_{f_2}^{(2)}} w_{f_1}^{(1)} SPI_{f_1 s} \Big)\Big)\Big)\Big) \tag{1.2}$$

Additionally, we have to specify the concepts of optimum index score (OIS) and cross-index score (CIS). Equation 1.3 enables the computation of the CIS value (used for a direct state comparison).

$$CIS_s = \left(1/n\right)\sum_{r=1}^{n}\sum_{j=1}^{q} w_{SPI_{j,r}} w_{SPI_{j,r}} \tag{1.3}$$

- $CIS_s$ – Cross efficiency score of the *s-th* DMU.

In order to guide the weight attribution process, avoiding unreasonable weight distributions as well as a "black box" effect, we inserted weight restrictions reflecting the findings introduced in Table 9.1, which in spite of the differences in magnitude suggest that road user related indicators are more important than environment indicators, which are more important than vehicle indicators. Therefore, the shares attributed to the main domains in the total CI value should be: $90\%(CI) \geq Share_{RU} > Share_E > Share_V \geq 5\%(CI)$.

Due to the same reasons, the weights attributed to indicators belonging to the same domain or subdomain can only vary within a range from 0.6 to 1.4 of their average weights; for example, alcohol (A) is divided in two SPIs, so the average weight of the indicators "1" and "2" is equal to 0.5 and thereby should lie between 0.3 and 0.7. Lastly, to avoid the concentration of high shares of the CI in a single SPI, we also introduced minimum and maximum shares for each indicator, accordingly: $0.01\%(CI) \leq Share_{1,\ldots,27} \leq 15\%(CI)$. The model's script is available in Appendix K.

### 9.3.3 SPI RESEARCH RESULTS

Firstly, according to the main goal of this research, we describe the general results of creating a road safety performance index for the entire set of states. Table 9.6 shows the OIS values and the CIS (which dictates the ranking of states from best to worst performing) for each state. In addition, the table contains the shares of the CIS value attributed to each road safety domain (which are similar to the shares computed in relation to the OIS). The ranking of the states presents reasonable statistically significant correlation coefficients at the 95% confidence level with the outcome related rank based on the traffic fatalities per registered vehicle (equal to – 0.58).

Regarding the shares comparison, due to the inserted weight restrictions, the order of the relative importance of the road user, environment and vehicle related SPIs is the same in every state ($Share_{RU} > Share_E > Share_V$). However, in the comparison of a particular domain share between different states, the interpretation should obey the following logic: the smaller the share, the worse the performance in that domain.

**Table 9.6** – Rank of states according to the computed CIS, OIS and shares attributed to each road safety domain

| Position | State | OIS | CIS | Share of the CIS attributed to each domain | | |
|---|---|---|---|---|---|---|
| | | | | Human (road user) | Environment | Vehicle |
| 1st | SP | 1.000 | 0.997 | 64.07% | 28.81% | 7.12% |
| 2nd | DF | 1.000 | 0.910 | 33.34% | 33.33% | 33.33% |
| 3rd | RJ | 1.000 | 0.898 | 43.42% | 43.09% | 13.49% |
| 4th | AP | 0.946 | 0.856 | 49.66% | 29.28% | 21.06% |
| 5th | AC | 0.908 | 0.841 | 48.06% | 33.76% | 18.18% |
| 6th | ES | 0.862 | 0.816 | 49.90% | 34.07% | 16.03% |
| 7th | RS | 0.867 | 0.806 | 45.80% | 35.66% | 18.54% |
| 8th | MS | 0.859 | 0.805 | 52.66% | 34.29% | 13.05% |
| 9th | SC | 0.844 | 0.804 | 71.78% | 19.93% | 8.29% |
| 10th | RR | 0.863 | 0.798 | 52.39% | 30.70% | 16.91% |
| 11th | PR | 0.834 | 0.791 | 65.42% | 20.45% | 14.13% |
| 12th | AM | 0.853 | 0.777 | 51.73% | 30.13% | 18.14% |
| 13th | TO | 0.843 | 0.774 | 45.64% | 37.54% | 16.82% |
| 14th | MG | 0.804 | 0.773 | 51.39% | 34.09% | 14.52% |
| 15th | RN | 0.820 | 0.763 | 53.80% | 30.03% | 16.17% |
| 16th | RO | 0.822 | 0.761 | 45.70% | 36.67% | 17.63% |
| 17th | MT | 0.780 | 0.729 | 50.29% | 33.12% | 16.59% |
| 18th | AL | 0.775 | 0.728 | 49.10% | 32.07% | 18.83% |
| 19th | BA | 0.759 | 0.706 | 58.80% | 25.00% | 16.20% |
| 20th | PB | 0.757 | 0.705 | 45.38% | 33.78% | 20.84% |
| 21st | CE | 0.758 | 0.700 | 46.02% | 34.29% | 19.69% |
| 22nd | GO | 0.733 | 0.698 | 70.44% | 19.36% | 10.20% |
| 23rd | PE | 0.746 | 0.694 | 46.82% | 33.78% | 19.40% |
| 24th | PI | 0.762 | 0.694 | 46.86% | 30.73% | 22.41% |
| 25th | MA | 0.756 | 0.690 | 47.22% | 26.39% | 26.39% |
| 26th | PA | 0.743 | 0.675 | 74.25% | 13.33% | 12.42% |
| 27th | SE | 0.723 | 0.670 | 42.47% | 37.35% | 20.18% |

The Co-plot graphs (Figures 9.4 to 9.6) present a more visually appealing representation of the results, exhibiting the overall and two disaggregated forms of the performance of the states in relation to their respective clusters. The disaggregation is according to the first layer of the hierarchical representation of Figure 9.2, splitting the CI value into human (road user), environment (road and health system), and vehicle performance.

The gray scale of circles representing each state corresponds to the overall performance (the CI value). The position in the rank is shown as well, from best to worst performing. The Co-plot graph enables decision makers to check on which particular aspects a state is doing well and those on which it is deficient and should improve. Nevertheless, it is important to stress that the construction of the Co-plot representation does not consider the corresponding shares attributed to each domain/subdomain in the CI value; for this, it is necessary to consult the donut graphs in the individual state profiles of Appendix A. The thickness of the arrows manifests the share of the domain or subdomain in the CI value.

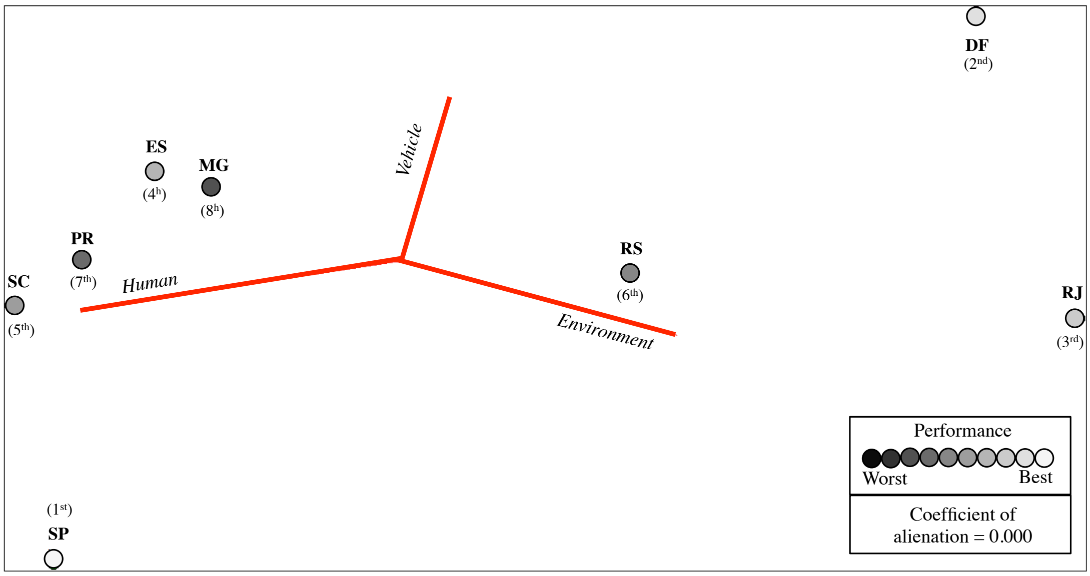



**Figure 9.4** – Graphical representation of the 1st hierarchical layer and overall performance in Cluster 1

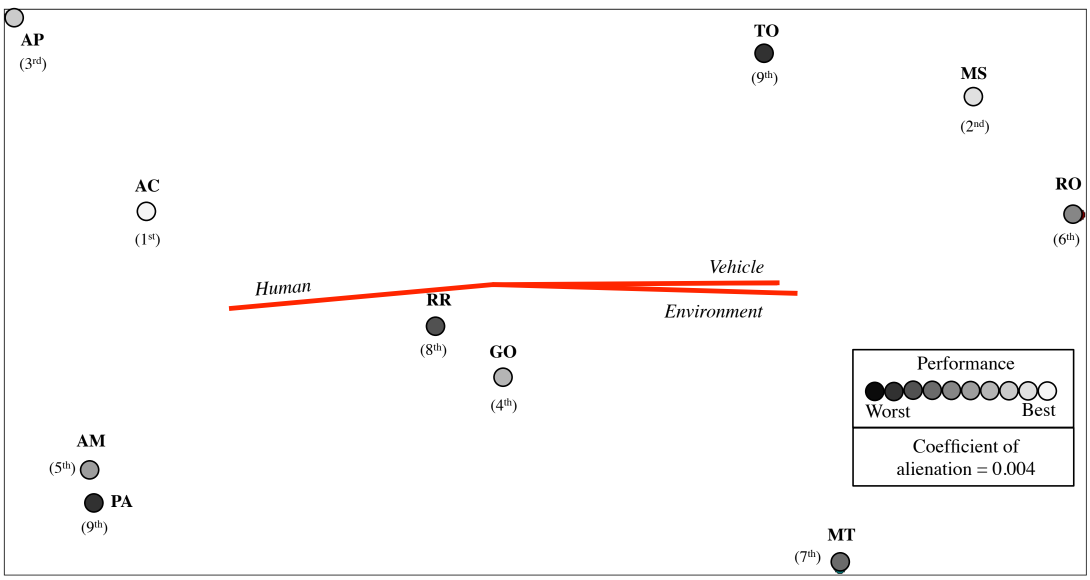



**Figure 9.5** – Graphical representation of the 1st hierarchical layer and overall performance in Cluster 2

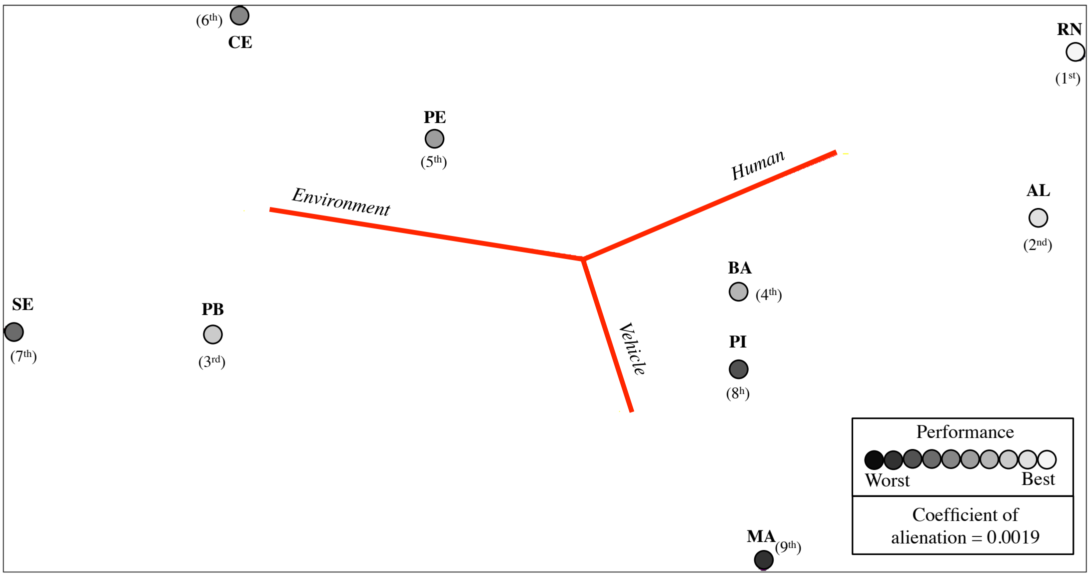



**Figure 9.6** – Graphical representation of the 1st hierarchical layer and overall performance in Cluster 3

The graphical representation of the 1st hierarchical layer in Figures 9.4 to 9.6 shows a regional insight in the information presented in Table 9.6. Although the analysis at the national level is important, focusing on the regional contexts can provide a more realistic basis for comparing performances and, in a next step, encourage the exchange of good practices based on policy documents.

## 9.4. CHAPTER CONCLUSIONS

Tables 9.7 to 9.9 summarize the recommendations for the states with respect to the priority road safety (sub) domain to which road safety efforts should be devoted. The rank order is according to the share of the corresponding domain in the CI value. The tables should be read as follows: alcohol related initiatives are more needed in e.g. DF, RJ and SC and less required in MG, ES, RS. In Appendix A, the priority domains for action in each state are underlined at the bottom right of the state profile page.

**Table 9.7** – Priority domains for actions towards improving road safety for Cluster 1 states

| Priority suggestion | Human | | | | | Environment | | Vehicle | |
|---|---|---|---|---|---|---|---|---|---|
| | Alcohol | Cell-phone | Seatbelt | Helmet | Speed | Road | Health system | Fleet composition | Age of the fleet |
| 1st | DF | RJ | RJ | RS | PR | ES | SP | SP | RJ |
| 2nd | RJ | DF | ES | SP | RS | MG | PR | ES | DF |
| 3rd | SC | SC | DF | MG | MG | SC | SC | SC | SP |
| 4th | SP | MG | SP | DF | RJ | PR | DF | MG | RS |
| 5th | PR | RS | RS | RJ | DF | RS | ES | PR | PR |
| 6th | MG | ES | PR | PR | SC | DF | MG | RS | SC |
| 7th | ES | PR | MG | ES | ES | SP | RJ | RJ | MG |
| 8th | RS | SP | SC | SC | SP | RJ | RS | DF | ES |

**Table 9.8** – Priority domains for actions towards improving road safety for Cluster 2 states

| Priority suggestion | Human | | | | | Environment | | Vehicle | |
|---|---|---|---|---|---|---|---|---|---|
| | Alcohol | Cell-phone | Seatbelt | Helmet | Speed | Road | Health system | Fleet composition | Age of the fleet |
| 1st | TO | AP | PA | MT | RO | RR | AP | RO | AM |
| 2nd | RR | RR | AM | AC | TO | AM | PA | MT | MS |
| 3rd | GO | GO | RO | GO | PA | AP | GO | TO | RR |
| 4th | MS | MT | RR | MS | MS | AC | AC | AC | GO |
| 5th | AM | TO | MS | PA | AM | PA | AM | PA | MT |
| 6th | MT | RO | TO | RO | AP | MS | MT | AP | RO |
| 7th | AC | AC | AP | TO | MT | TO | RO | MS | AP |
| 8th | RO | MS | GO | AM | RR | GO | TO | GO | TO |
| 9th | PA | AM | AC | AP | AC | MT | RR | RR | AC |
| 10th | AP | PA | MT | RR | GO | RO | MS | AM | PA |

**Table 9.9** – Priority domains for actions towards improving road safety for Cluster 3 states

| Priority suggestion | Human | | | | | Environment | | Vehicle | |
|---|---|---|---|---|---|---|---|---|---|
| | Alcohol | Cell-phone | Seatbelt | Helmet | Speed | Road | Health system | Fleet composition | Age of the fleet |
| 1st | BA | PI | PI | PE | SE | BA | MA | CE | RN |
| 2nd | PI | CE | CE | PB | MA | PI | AL | PI | PE |
| 3rd | CE | MA | PB | SE | PE | RN | PB | MA | PB |
| 4th | RN | SE | AL | AL | PB | AL | CE | AL | CE |
| 5th | SE | PB | BA | MA | RN | MA | PE | RN | BA |
| 6th | AL | PE | RN | RN | BA | PE | SE | PE | SE |
| 7th | PB | AL | SE | BA | AL | PB | RN | SE | AL |
| 8th | PE | BA | MA | CE | CE | SE | PI | BA | PI |
| 9th | MA | RN | PE | PI | PI | CE | BA | PB | MA |

# 10. CONCLUSIONS

In this manuscript, we described the procedures designed for executing strategic road safety indicator and index research for Brazilian states, creating results that can be used to support strategic road safety analysis. Three stages were the cornerstone for achieving the research objectives: diagnosing the road safety situation using final outcome related indicators; setting the target number of traffic fatalities; and research concerning safety performance indicators (all at the state level). We conclude this research by first answering the research questions proposed in Chapter 1, then we present suggestions for future research on the theme and finally, we end with some broad reflections about the theme.

## 10.1. ANSWERS TO RESEARCH QUESTIONS

In order to clearly state the main conclusions of the research, we answer the research questions one by one, as follows:

***RQ1:*** *Since there is an unfavorable data availability scenario, is the quantity and quality of Brazilian data sufficient to execute this kind of analysis?*

Information concerning the diagnosis step (final outcome based evaluation) is more easily available than that used for more specific guidance towards practical actions, as in the SPI research. The dilemma of “working under limited conditions” was frequent in this study, mainly in the SPI research. Regarding the exposure data, the availability of own, previous research was an important initial step for further improvements in this estimation. In spite of all obstacles, it was possible to perform the type of research we aimed for and show the potential of this research methodology, although extrapolating its results into real world practical actions should still be done carefully due to the limitations surrounding some parts of the research.

***RQ2:*** *Given to the contrasting situation through the different Brazilian states, which are the most adequate clusters of states for conducting the research?*

In Chapter 4, we described the process of clustering the Brazilian states into similar context groups, including the necessary premises and the specification of the parameters (related to different dimensions of the road safety problem) on which this process was based. In essence, the defined clusters fundamentally represented the socioeconomic division of the country in three main parts, namely Cluster 1 – Southern/Southeastern region (plus the federal district), Cluster 2 – Center-West/ North (Amazon) and, finally, Cluster 3 – the Northeastern region. It supports the idea that a group of states with a similar context provides more realistic comparisons and creates a more favorable environment for knowledge and good practice transfer. Therefore, the clustering process includes a variety of aspects (i.e., outcome related indicators, exposure characterization, road user behavior, road infrastructure, health system, vehicle fleet, socio-economic parameters and geographical proximity). These topics capture the main overall differentiation in road safety related characteristics across the country (and even within each of the 3 clusters, we still find very clear differences in performance).

***RQ3:*** *Since the application of data envelopment model for the mentioned purposes is inedited in Brazil (and also in many other non-European countries/developing countries), how to assess the applied models' performance in the country?*

In Chapter 6, we explained a calibration process used to assess the performance of the MLCI-DEA model application for Brazilian states (BR-27) based on the more consolidated European Union experience in ranking its member countries (EU-27) according to DEA scores. The search for better correlations between the DEA composite indicator scores and socioeconomic parameters was the principle guiding this calibration. It expressed the need for a special weight restriction in order to control the effects of a confounding variable, the mortality rate (MR), in the composite indicator value in the case of a less-motorized country. Moreover, the attempt of applying different models for different performance related issues enabled at some extent the comparison of the distinct models results. Consequently, we have another perspective for the evaluation of the models adequacy for investigating the road safety problem in its various components in Brazil.

***RQ4:*** *How to test the sensitivity of the data envelopment model results, which manifest the "per state" road safety situation regarding the considered dimensions?*

Bootstrapping the DEA scores was the chosen technique to test the sensitivity of the results to possible variations in the input data, which might be subjected to the influence of unfavorable aspects with respect to data quality and availability. For this reason, we performed a sensitivity analysis for every DEA experiment we conducted. It showed how susceptible the road safety related data is to uncertainties and to which extent it affects the research results. We described this technique, for the first time applied to road safety DEA results in a developing country, in Chapter 5. Generally, the median or average values of the bootstrapping results provided quite similar rankings in comparison to the original ranking based on the DEA efficiency/index scores (both OES/OIS and CES/CIS). The advantage now is that we by means of boxplot graphs revealed some knowledge on how likely each rank configuration could be obtained.

**RQ5:** Which are the best and the underperforming states with respect to the road safety?

We deliver this information in different perspectives. Chapters 6 and 7 brought general insight in road safety performance; this is, for the entire group of road users (motorized and non-motorized ones). It is the base ranking of the 27 states. Chapter 7 showed a second and more detailed perspective with an analysis focused on motor vehicle occupants using exposure and fatality data, and, although this information is not that appealing itself for practical use, it was helpful for deriving some insight about non-motorized fatal victims (as described in the last sections of Chapter 8). Then, Chapter 8 also contained a disaggregated performance evaluation by running four separate DEA models for each motor vehicle type (motorcycle, car, truck and bus). Finally, in Appendix A, the state profiles are presented; compiling the different parts of information and offering a useful summarized visualization of relevant road safety knowledge built up during this thesis.

***RQ6:*** *Based on a benchmarking exercise, which is the target number of fatalities for each state?*

Each input-output DEA efficiency model we constructed provided a target number of traffic fatalities for specific road users or set of road users of underperforming states as a result of a benchmarking exercise. We suggested traffic fatality targets in Chapters 7 and 8. Even though it is valuable for decision-makers in every state to have a target for traffic fatality reduction in mind, it is essential to emphasize that it does not consist of a precise or fixed value. Therefore, we suggested a range for the target number of traffic fatalities in each state, which was actually a result of the sensitivity analysis we carried out. Concerning the attainability and realism of the indicated targets, it depends on the time horizon decision makers would estipulate. One apparently unrealistic target might be possible to achieve if a longer time period is dedicated for that. Certainly, the commitment level of the bodies involved in the road safety improvement process is also influential in accelerating or impairing traffic fatality reductions at a reasonable time horizon. Appendix A provides compiled information about the various target specifications we provided in this thesis through the state profiles.

***RQ7:*** *How can the safety performance indicators research offer guidelines for improving road safety?*

The SPI research was probably the biggest challenge of this study, since traditionally in Brazil there is no systematic indicators collection process that could support this research purpose. In spite of the difficulties, we succeeded in gathering a set of 27 SPIs capable to manifest the situation throughout different road safety domains – given the assumption that the collected information from time to time served as proxy variables of the risk situation at hand. In other words, although the ideal set of safety performance indicators is not currently available, the ones that we have are supposed to represent the unknown ones to a satisfactory level. The hierarchical structure of SPIs we mentioned in Chapter 9 allows the analysis of distinct layers of information to which a specific weight was allocated by running the MLCI-DEA maximization model. Those indicators representing a large share of the road safety performance index score are supposed to indicate a favorable situation for that state; on the other hand, indicators associated to small shares of the composite indicator value represent an unfavorable performance for that particular domain, subdomain or indicator. The Co-plot graphs help to visualize the situation of specific road safety domains for states constituting a cluster. In addition, in Section 9.4 we compare the domains' performances between states. In summary, each state should invest in improving the situation of those indicators to which a relatively small weight was assigned if it aims to be as good as those best-performing states.

***RQ8:*** *Given the precedence of the application of a similar methodology for the European Union member countries, what can they learn from the Brazilian experience on DEA road safety research?*

In relation to the previous studies focused on the European Union context, we added some steps that contributed to improve the existing method. The main example in this sense is bootstrapping the DEA scores as a sensitivity analysis procedure. This step would also be valuable for investigating the road safety performance of European countries, since data uncertainty issues are also present in the European context. Clustering DMUs (countries) is not something new in EU investigations; however we believe that keeping the same cluster configuration across different research perspectives is valuable for comparison purposes. A second point is the indication of possible bias in the traffic fatality related indicators. The lessons learned from this process may also enhance the reliability of the European DEA estimators. More precisely, it would be especially valuable when extending the cross-country evaluations to other parts of Europe, which are at a quite distinguished motorization level, as the recently developing Eastern countries, that would probably require a distinct model calibration. Additionally, regarding the methodology we presented for the VKT estimates, it could be a reference for some European countries where this kind of data is still unknown.

## 10.2. SUGGESTIONS FOR FUTURE RESEARCH

During the development of this research, we identified a variety of aspects that, if already properly addressed in previous research, could be beneficial for enhancing the quality of the current study or should be included in future investigations in Brazil. They are divided into data availability issues and methodological challenges for the future.

The main data availability issues are:

- Availability of pedestrian and cyclist exposure data;
- Availability of extra base-information to improve the quality of the METDFS, increasing the reliability of the exposure (VKT) estimation (e.g. regular vehicle inspections with odometer reading);
- Knowledge on the distribution of this exposure according to the environment, age of the driver, weather conditions, etc.;
- Occupancy rates per vehicle category in the states;

- Precision of the traffic fatality classification in the CID-10 codes (with a minor share of traffic fatalities for which the transportation modes involved are unspecified);
- Separation of trucks and buses as second part in the CID-10 "death cause" specification code would improve the precision of investigations about the mutual influence of different exposure types in fatality causation;
- Knowledge about the distribution of the traffic fatalities between rural and urban environments (since there are substantially different patterns for accidents contributory factors in these two environments);
- Availability of more adequate SPIs concerning all behavioral aspects;
- Availability of SPIs related to the urban road system;
- Availability of SPIs related to the emergency response to traffic accidents;
- Availability of SPIs related to the presence of safety devices in the vehicle fleet (e.g. airbag or brakes with anti-blocking system);
- Availability of SPIs related to pedestrian and/or other non-motorized users behavior;
- Availability of a complete dataset regarding traffic infractions;

The key methodological challenges for the future are:

- Strategies for combining the DEA models' results should be improved, so it would become possible to obtain more consistent conclusions from that (e.g. in the non-motorized traffic fatality targets estimates);
- In spite of the cluster division we applied, there are still some clusters with relatively contrasting situations;
- Development of an even more automated process for bootstrapping the DEA scores for the specific models we deal with;
- The fact that best-performing states are considered as benchmarks implies few or none suggestions for improvements in these places. However, it is also true that these states labeled as best-performing do not exhibit the "state of the art" situation with respect to road safety;
- In line with the previous comment, the suggestion of less ambitious benchmarks would probably be helpful for making the performance evaluation and target setting process even more realistic. In other words, if a state is not able to reach the best-

performing state's level, it could learn from the performance of other states that present simply a better performance – and not necessarily the best one.

## 10.3. FINAL GENERAL COMMENTS

Brazil has experienced important transformations during the last decades and it is probably going to face even more changes in the upcoming years. Regarding the transportation area, there is still a lot of infrastructure to be built (motorways, urban road infrastructure, etc.) and road safety must be incorporated as a priority in such projects.

In this context, one of the most impactful topics is the motorization growth and its impacts on the mobility of the population. Although there are contrasting backgrounds throughout the states, in general, the country has not yet reached the saturation level on which road safety becomes a priority concept in transportation projects. As a result, traffic fatalities continue to be an accepted "price" for the development, although isolated actions are gradually implemented. This shift in priorities is fundamental for achieving better results in the road safety field and hopefully succeeding in reaching the fatality reduction target set at the beginning of the Decade of Action for Road Safety.

From the conceptual road safety perspective, more important than providing safer conditions for the exposure in road transportation modes is to transfer these users to safer public transportation systems (rail based systems, for example) or even to stimulate trips by means of road public transportation systems, as buses, which offer much safer travel conditions to the users. This transference is particularly important because it acts preventively in the root of the road safety problem – the exposure to the risk dimension.

In addition, non-motorized users have to become a priority in urban policies in a more practical manner. Pedestrians and cyclists are the most vulnerable road users and account for a substantial share in the total number of traffic fatalities in Brazil. These users even tend to be more endangered given the focus and priority for individual motorized transportation means. Therefore, decision-makers must act towards putting sustainable mobility alternatives at the first place in their political agenda.

More data oriented rather than pure political decisions are vital to address the road safety issue in Brazil. The creation and consolidation of a sound road safety related data oriented collection and monitoring system is essential for enabling future investigations, for elaborating adequate strategic planning, for establishing priority actions, for measuring the impact of implemented measures and finally, for rethinking and improving this process – in a virtuous cycle.

# REFERENCES

ABRASPE (1983). *A condição do pedestre no transporte urbano*. Associação Brasileira de Pedestres.

Adler, N., Friedman, L., & Sinuany-Stern, Z. (2002). Review of ranking methods in the data envelopment analysis context. *European Journal of Operational Research*, *140*(2), 249–265.

Adler, N., & Raveh, A. (2008). Presenting DEA graphically. *Omega*, *36*(5), 715–729. doi:10.1016/j.omega.2006.02.006

Aliev, F., & Ebadi, S. (2012). Ranking efficient DMUs by bootstrapping method (pp. 1–3). Presented at the Problems of Cybernetics and Informatics (PCI), 2012 IV International Conference. doi:10.1109/ICPCI.2012.6486425

Anderberg, M. R. (1973). The broad view of cluster analysis. In *Cluster analysis for applications* (pp. 1–9). London: Academic Press Inc. Ltd.

Anderson, M. (2008). Safety for whom? The effects of light trucks on traffic fatalities. *Journal of Health Economics*, *27*(4), 973–989. doi:10.1016/j.jhealeco.2008.02.001

ANP (2012). *Statistics on fuel sales 2000- 2012*. Brasília: Agência Nacional do Petróleo Gás Natural e Biocombustíveis. Retrieved from http://www.anp.gov.br/?pg=23307&m=&t1=&t2=&t3=&t4=&ar=&ps=&cachebust=1276622345913

Baer, W. (2008). Introduction. In *The Brazilian economy: growth and development* (6th ed., pp. 2–9). Lynne Rienner.

Banco Central do Brasil (2013). *Indicadores econômicos*.

Bastos, J. T., Ferraz, A. C. P., Vieira, H., & Bezerra, B. S. (2010). Uma estimativa da mortalidade no trânsito considerando exposição ao risco no Brasil. Presented at the XXIV Congresso de Pesquisa e Ensino em Transportes – ANPET, Salvador, Brazil.

Bastos, J. T. (2011). *Geografia da mortalidade no trânsito no Brasil* (Master). School of Engineering of São Carlos, University of São Paulo. Retrieved from http://www.teses.usp.br/teses/disponiveis/18/18144/tde-14032011-112111/

Bastos, J. T., Ferraz, A. C. P., Vieira, H., & Bezerra, B. S. (2011). Road safety development in Brazil in terms of fatalities and mobility data. Presented at the XV Congreso Chileno de Inginiería de Transporte, Santiago, Chile.

Bastos, J. T., Ferraz, A. C. P., Vieira, H., & Bezerra, B. S. (2012). Geography of traffic fatalities in Brazil. *Revista ANTT*, *4*, 249–265.

Bax, C., Wesemann, P., Gitelman, P., Shen, Y., Goldenbeld, C., Hermans, E., … Aarts, L. (2012). *Developing a Road Safety Index* (p. 185). Retrieved from http://www.dacota-project.eu/Deliverables/DaCoTA_D4.9_developing%20a%20RSI%20deliverable.pdf

Borba, B. S. M. C. (2008). *Metodologia de regionalização do mercado de combustíveis automotivos no Brasil* (Me). Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Bureau of Infrastructure, Transport and Regional Economics (2011). *Road vehicle-kilometres travelled: estimation from state and territory fuel sales*. Canberra, Australia. Retrieved from http://www.bitre.gov.au/publications/2011/files/report_124.pdf

Cardoso, J. L. (2005). The use of international data on fuel sales and vehicle fleet for the estimation of yearly national traffic volumes. *Accident Analysis & Prevention*, *37*(1), 207–215. doi:10.1016/j.aap.2003.12.005

Cavalcanti, M. C. B. (2004). Ascenção do gás natural no mercado de combustíveis automotivos no Brasil. Presented at the 3° Congresso Brasileiro de P&D em Petróleo e Gás, Salvador, Brasil. Retrieved from http://www.portalabpg.org.br/PDPetro/3/trabalhos/IBP0294_05.pdf

Census Bureau (2011). *Statistical Abstracts*. United States: Administration and Customer Services Division. Retrieved from http://www.census.gov/prod/www/statistical_abstract.html

Census Bureau (2012). *The 2012 Statistical Abstract*. United States. Retrieved from http://www.census.gov/compendia/statab/cats/population.html

Cherchye, L., Moesen, W., Rogge, N., Puyenbroeck, T. V., Saisana, M., Saltelli, A., … Tarantola, S. (2006). *Creating Composite Indicators with DEA and Robustness Analysis: the case of the Technology Achievement Index* (Public Economics Working Paper Series). Katholieke Universiteit Leuven, Centrum voor Economische Studiën, Working Group Public Economics. Retrieved from http://www.econ.kuleuven.ac.be/eng/ew/discussionpapers/Dps06/Dps0603.pdf

Chernick, M. R. (1999). What is bootstrapping? In *Bootstrap methods: a practitioner's guide* (pp. 1–22). John Wiley & Sons, Inc.

CNT (2011). *Pesquisa CNT de rodovias 2011*. Brasília: Confederação Nacional dos Trasnportes - CNT/ SEST-SENAT. Retrieved from http://pesquisarodovias.cnt.org.br/Downloads/Edicoes//2011/Apresentação/Apresentaao_Pesquisa_CNT_Rodovias_2011.pdf

CONASS. (2006). Breve história do SUS. In *SUS: avanços e desafios* (pp. 21–28). Brasília: Conselho Nacional de Secretários de Saúde.

Connor, J., Langley, J., & Cryer, C. (2007). *International comparisons of injury* (p. 56). New Zealand. Retrieved from http://www.cdc.gov/nchs/data/ice/Int_comp_Combined_9_07_2_%20w.pdf

Cooper, W. W., Lawrence M. Seiford, & Joe Zhu. (2011). Data envelopment analysis: history, models and interpretations. In *Handbook on Data Envelopment Analysis* (Vol. 164, pp. 1–39). Springer. Retrieved from http://www.springer.com/business+%26+management/operations+research/book/978-1-4419-6150-1

Cooper, W. W., Seiford, L. M., & Tone, K. (2000). The basic CCR model. In *Data Envelopment Analysis: A Comprehensive Text With Models, Applications, References and Dea-Solver Software* (pp. 21–40). Springer.

Davison, A. C., & Hinkley, D. V. (2003). The basics bootstraps. In *Bootstrap methods and their application* (2nd Revised edition., pp. 11–69). Press Syndicate of the University of Cambridge.

DENATRAN (1998). *Índice de violência de trânsito no Brasil*. Departamento Nacional de trânsito.

DENATRAN. (2013). Vehicle fleet. Departamento Nacional de trânsito. Retrieved January 12, 2014, from http://www.denatran.gov.br/frota.htm

DETRAN. (n.a.). *State level statistics provided by DETRANs*.

Doyle, J., & Green, R. (1994). Efficiency and Cross-Efficiency in DEA: Derivations, Meanings and Uses. *The Journal of the Operational Research Society*, *45*(5), 567.

Efron, B. (1992). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, *7*, 569–593.

Elvik, R., Høye, A., Vaa, T., & Sørensen, M. (2009). Factors contributing to road accidents. In *The hanbook of road safety measures* (2nd Revised edition., pp. 35–80). Emerald Group Publishing Limited.

Elvik, R., & Vaa, T. (2004). Factors contributing to road accidents. In *The handbook of road safety measures* (pp. 29–79). London: Elsevier.

Elvik, R. (2006). Laws of accident causation. *Accident Analysis & Prevention*, *38(4)*, 742–7.

European Commission (2013). *Transport database*. Retrieved from http://epp.eurostat.ec.europa.eu/portal/page/portal/transport/data/database

European Commission (2014). *Road Safety Vademecum*. Retrieved from http://ec.europa.eu/transport/road_safety/pdf/vademecum_2014.pdf

Evans, L. (2004). *Traffic Safety*. Science Serving Society.

Everitt, B. (1980). Chapters 1 and 3. In *Cluster analysis*. United Kingdom: British Library C.I.P. Data.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, *120*, 253–290.

Ferraz, A. C. P., Raia Jr., A. A., Bezerra, B. S., Bastos, J. T., & Silva, K. C. R. (2012). *Segurança Viária*. Suprema Gráfica e Editora.

Ferraz, A. C. P., Raia Jr., A. A., Bezerra, B. S. (2008). *Segurança no trânsito*. Grupo Gráfico São Francisco.

FHWA (1997). *Highway statistics summary to 1995*. United States: Federal Highway Administration.

FHWA (2011). *Highway statistics*. United States: Federal Highway Administration. Retrieved from http://www.fhwa.dot.gov/policy/ohpi/hss/index.htm

Gawryszewski, V. P., Silva, M. M. A., Malta, D. C., Kegler, S. R., Mercy, J. A., Mascarenhas, M. D. M., & Neto, O. L. M. (2008). Violence-related injury in emergency departments in Brazil. *Revista Panamericana de Salud Pública*, *24*(6).

Geiger, P. P. (1969). Regionalização. *Revista brasileira de Geografia*, *31*, 5–25.

Golob, T. F., Recker, W. W., & Alvarez, V. M. (2004). Freeway safety as a function of traffic flow. *Accident Analysis and Prevention*, *36*, 933–946.

Governo do Estado de São Paulo (2013). *Indústria automobilística*. Retrieved from http://www.saopaulo.sp.gov.br/conhecasp/historia_republica-industria-automobilistica

Hakkert, S., & Gitelman, V. (2007). *Road Safety Performance Indicators: Manual*.

Hermans, E., Wets, G. & Van den Bossche, F. (2006). Describing the Evolution in the Number of Highway Deaths by a Decomposition in Exposure, Accident Risk and Fatal risk. *Transportation Research Record*, *3590*, 1–8.

Hermans, E. (2009). *A methodology for developing a composite road safety performance index for cross-country comparison* (PhD). Transportation Research Institute, Hasselt University, Diepenbeek.

Hermans, E., Brijs, T., Wets, G., & Vanhoof, K. (2009). Benchmarking road safety: lessons to learn from a data envelopment analysis. *Accident Analysis and Prevention*, *41*(1), 174–182.

Hossain, A., & Gargett, D. (2011). Road vehicle-kilometres travelled estimated from state/territory fuel sales. Presented at the Australasian Transport Research Forum, Adelaide, Australia. Retrieved from http://www.patrec.org/atrf.aspx

INMETRO (2010). *Programa Brasileiro de Etiquetagem Veicular* (Conpet - Programa Nacional de Racionalização do Uso Dos Derivados do Petróleo e do Gás Natural). Brasília: Instituto Nacional de Metrologia, Normalização e Qualidade Industrial.

Institute for Road Safety Research – SWOV (2013). SWOV Fact sheet. Retrieved October 27, 2013, from https://www.swov.nl/rapport/Factsheets/UK/FS_Risk.pdf

Instituto Brasileiro de Geografia e Estatística (1970). Evolução político administrativa. Retrieved September 16, 2012, from http://www.ibge.gov.br/home/geociencias/cartogramas/evolucao.html

Instituto Brasileiro de Geografia e Estatística (2000). Population Distribution Map - 2000. Retrieved January 15, 2014, from http://www.ibge.gov.br/english/geociencias/geografia/mapas_doc1.shtm

IRTAD (2014). *Road Safety Annual Report 2014*. International Traffic Safety Data and Analysis Group. Retrieved from http://www.internationaltransportforum.org/pub/pdf/14IrtadReport.pdf

Jørgensen, N. O. (1996). Traffic safety. In B. Thagesen, *Highway and Traffic Engineering in Developing Countries* (pp. 121–138). Taylor & Francis.

Kabak, O., & Ruan, D. (2010). A cumulative belief degree-based approach for missing values in nuclear safeguards evaluation. *IEEE Transactions on Knowledge and Data Engineering*. Retrieved from http://doi.ieeecomputersociety.org/10.1109/TKDE.2010.60

Kennedy, J. F. (2008). *Estimating pedestrian volumes and crashes at urban signalized intersections* (MSc.). Faculty of the Virginia Polytechnic Institute and State University, Falls Church.

Koornstra, M., Lynam, D., Nilsson, G., Noordzij, P., Petterson, H.-E., Wegman, F., & Wouters, P. (2002). *SUNflower: a comparative study of the development of road safety in Sweden, the United Kingdom, and the Netherlands* (p. 128). Leidschendam: SWOV.

Lam, W. W. Y., Yao, S., & Loo, B. P. Y. (2014). Pedestrian exposure measures: A time-space framework. *Travel Behaviour and Society*, *1*(1), 22–30.

Leduc, G. (2008). *Road traffic data: collection methods and applications* (Working Papers on Energy, Transport and Climate Change). Luxembourg: European Commission. Retrieved from http://ftp.jrc.es/EURdoc/JRC47967.TN.pdf

Losekann, L., & Vilela, T. (2010a). Estimação da frota brasileira de automóveis flex e a nova dinâmica do consumo de etanol no Brasil a partir de 2003. Retrieved January 15, 2014, from http://infopetro.wordpress.com/2010/07/26/estimacao-da-frota-brasileira-de-automoveis-flex-e-a-nova-dinamica-do-consumo-de-etanol-no-brasil-a-partir-de-2003/

Losekann, L., & Vilela, T. (2010b). Frota brasileira de veículos leves: difusão dos flexíveis e do GNV. Retrieved January 15, 2014, from http://infopetro.wordpress.com/2010/04/19/frota-brasileira-de-veiculos-leves-difusao-dos-flexiveis-e-do-gnv/

Löthgren, M. (1998). How to Bootstrap DEA Estimators: A Monte Carlo Comparison. Retrieved October 29, 2013, from http://swopec.hhs.se/hastef/abs/hastef0223.htm

LTSA (2000). *Road safety 2010: a consultation document* (p. 102). Wellington: Land Transport Safety Authority. National Road Safety Commitee of New Zealand. Retrieved from www.ltsa.govt.nz

McClave, J. T., & Sincich, T. (2003). Chapters 11 and 14. In *Statistics* (9th ed.). New Jersey, USA: Prentice Hall, Inc.

MDT (2014). No último dia, Conselho aprova resoluções sobre Década de Segurança Viária e sobre liberação de recursos do FUNSET. *Movimento Nacional Pelo Direito Ao Transporte Público de Qualidade Para Todos*. Retrieved from http://www.mdt.org.br/portal/SitePages/not.aspx?AspXPage=g%5F99865BD2C31541249AA6D0E94F1879AD:%2540codigo%3Dimdt14079702

Ministério da Ciência e Tecnologia (2006). *Emissões de gases de efeito estufa por fontes móveis no setor enegético*. Brasília, Brazil.

Ministério da Saúde (2011a). *VIGITEL - Vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico*. Retrieved from http://tabnet.datasus.gov.br/cgi/dh.exe?vigitel/vigitel10.def

Ministério da Saúde (2011b). *VIVA - Vigilância de Violência e Acidentes*. Retrieved from http://tabnet.datasus.gov.br/cgi/deftohtm.exe?viva/2011/viva11p.def

Ministério da Saúde (2011c). Indicadores de recursos. Retrieved September 19, 2013, from http://tabnet.datasus.gov.br/cgi/idb2011/matriz.htm#recur

Ministério da Saúde (2011b). Indicadores socioeconômicos. Retrieved September 19, 2013, from http://tabnet.datasus.gov.br/cgi/deftohtm.exe?idb2012/b03.def

Ministério da Saúde (2012). *Pesquisa Nacional de Saúde Escolar*. Retrieved from http://www.ibge.gov.br/home/estatistica/populacao/pense/2012/

Ministério da Saúde (2013). Óbitos por causas externas. Retrieved July 15, 2012, from http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/ext10uf.def

Ministério da Saúde (2014). Censos (1980, 1991, 2000 e 2010), Contagem (1996) e projeções intercensitárias (1981 a 2012), segundo faixa etária, sexo e situação de domicílio. Retrieved July 10, 2014, from http://tabnet.datasus.gov.br/cgi/deftohtm.exe?ibge/cnv/popuf.def

Ministério de Minas e Energia (2012). *Balanço Energético Nacional de 2000 a 2011*. Brasília: Agência Nacional do Petróleo Gás Natural e.

Ministério dos Transportes (2013a). Estatísticas de acidentes nas rodovias federais. Retrieved July 15, 2013, from http://www.dnit.gov.br/rodovias/operacoes-rodoviarias/estatisticas-de-acidentes

Ministério dos Transportes (2013b). Rede rodoviária - totais gerais. Retrieved July 15, 2013, from http://www.dnit.gov.br/planejamento-e-pesquisa/planejamento/planejamento-rodoviario

Morsink, P., Oppe, S., Reurings, M., & Wegman, F. (2005). *SUNflower+6: Development and application of a footprint methodology for the SUNflower+6 countries* (p. 98). Leidschendam: SWOV.

OECD (2008). Handbook on constructing composite indicators: methodology and user guide. Retrieved October 29, 2013, from http://www.oecd.org/els/soc/handbookonconstructingcompositeindicatorsmethodologyanduserguide.htm

OECD (2011a). Road fatalities. In *OECD Factbook*. Organisation for Economic Co-operation and Development. Retrieved from http://www.oecd-ilibrary.org/content/chapter/factbook-2011-57-en

OECD (2011b). *OECD Environmental performance reviews: Portugal 2011*. OECD Publishing. Retrieved from http://dx.doi.org/10.1787/10/1787/9789264097896-en

OECD (2013). *Key Transport Statistics*. Organisation for Economic Co-operation and Development. Retrieved from http://www.internationaltransportforum.org/statistics/shortterm/QCountryNotes.pdf

ONSV (2013). *Fiscalização Aleatória* (p. 26). Indaiatuba: Observatório Nacional de Segurança Viária.

ONSV (2014). *Segurança viária no Brasil (2000-2012): Estatísticas, tendências e desafios*. Observatório Nacional de Segurança Viária. Retrieved from http://onsv.org.br/portal_d4dos/#/articles/1

Page, M., & Rackliff, L. (2006). Deriving and Validating a Road Safety Performance Indicator for Vehicle Fleet Passive Safety. *Annual Proceedings / Association for the Advancement of Automotive Medicine*, *50*, 317–332.

Penna, G. (2010). National Adolescent School-based Health Survey (PeNSE). *Ciência & Saúde Coletiva*, *15*(2).

Petruccelli, J. D., Nandram, B., & Chen, M. (1999). Chapter 2 and 6. In *Applied statistics for engineers and scientistis*. New Jersey, USA: Prentice Hall, Inc.

R Core Team (2014). R: A Language and Environment for Statistical Computing (Version 3.03). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Raktoe, B. L., & Hubert, J. J. (1979). Chapter 6 - Random variables and three popular probability distributions. In *Basic applied statistics* (Vol. 27, pp. 96–114). New York, USA: Marcel Dekker, Inc.

Raveh, A. (2000). Co-plot: A graphic display method for geometrical representations of MCDM. *European Journal of Operational Research*, *125*(3), 670–678. doi:10.1016/S0377-2217(99)00276-3

Rubin, D. B. (1987). *Multiple Imputation for Non-response in Surveys*. New York, USA: ohn Wiley & Sons.

Rumar, K. (1985). The Role of Perceptual and Cognitive Filters in Observed Behavior. In L. Evans & R. C. Schwing (Eds.), *Human Behavior and Traffic Safety* (pp. 151–170). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4613-2173-6_8

Scaringella, R. S., Filho, A. M., Joogi Mori, & Filho, M. P. S. (2012). *Investigação de Causas de Acidentes de Trânsito – Estudo de Amostra de Acidentes* (p. 176). São Paulo: Scaringella Trânsito.

Seiford, L. M., & Thrall, R. M. (1990). Recent developments in DEA: The mathematical programming approach to frontier analysis. *Journal of Econometrics*, *46*(1–2), 7–38. doi:10.1016/0304-4076(90)90045-U

Sexton, T. R., Silkman, R. H., & Hogan, A. J. (1986). Data envelopment analysis: Critique and extensions. *New Directions for Program Evaluation*, *1986*(32), 73–105. doi:10.1002/ev.1441

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika Trust Stable*, *52*, 591–611.

Shen, Y. (2012). *Inter-national benchmarking of road safety performance and development using indicators and indexes: data envelopment analysis based approaches* (PhD). Transportation Research Institute, Hasselt University, Diepenbeek.

Shen, Y., Hermans, E., Brijs, T., & Wets, G. (2013a). Data Envelopment Analysis for Composite Indicators: A Multiple Layer Model. *Social Indicators Research*, *114*(2), 739–756. doi:10.1007/s11205-012-0171-0

Shen, Y., Hermans, E., Bao, Q., Brijs, T., & Wets, G. (2013b). Road safety development in Europe: a decade of changes (2001-2010). *Accident Analysis and Prevention*, *60*, 85–94.

Shen, Y., Hermans, E., Ruan, D., Wets, G., Brijs, T., & Vanhoof, K. (2010). Evaluationg trauma managment performance in Europe: a multiple-layer data envelopment analysis model. *Transportation Research Record*, *2148*, 739–756.

Shen, Y., Hermans, E., Ruan, D., Wets, G., Brijs, T., & Vanhoof, K. (2011). A generalized multiple layer data envelopment analysis model for hierarchical structure assessment: a case study in road safety performance evaluation. *Expert Systems with Applications*, *38*(12), 15262–15272. doi:10.1016/j.eswa.2011.05.073

Simar, L., & Wilson, P. W. (1998). Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. Retrieved October 29, 2013, from http://pubsonline.informs.org/doi/abs/10.1287/mnsc.44.1.49?journalCode=mnsc

Simar, L., & Wilson, P. W. (2000). A General Methodology for Bootstrapping in Nonparametric Frontier Models. *Journal of Applied Statistics*, *27*, 779–802.

Stipdonk, H., & Berends, E. (2008). Distinguishing traffic modes in analysing road safety development. *Accident Analysis and Prevention*, *40*, 1383–1393.

Talby, D., & Raveh, A. (1999). Visual Co-Plot (Version 5.6). Retrieved from www.talby.com/coplot

Thode Jr., H. C. (2002). Chapters 1 and 6. In *Testing for normality*. Basel, Switzerland: Marcel Dekker, Inc.

Thomas, P. (2005). *State of the Art Report on Risk and Exposure Data* (SafetyNet Deliverable 2.1.). NTUA - National Technical University of Athens. Retrieved from http://www.dacota-project.eu/Links/erso/safetynet/fixed/WP2/Deliverable%20wp%202.1%20state%20of%20the%20art.pdf

Train, K. (1986). Qualitative choice analysis. Cambrige, MA.: The MIT Press, 1986.

Vis, M. A., & Eksler, V. (2008). *Road Safety Performance Indicators: Updated Country Comparisons*.

Vis, M. A., Hafen, K., Lerner, M., Allenbach, R., Verbeke, T., Eksler, V., Cardoso, J. (2005). Building the European Road Safety Observatory. SafetyNet. Deliverable D3. 1 State of the art report on road safety performance indicators. *(2005)*. Retrieved from http://discovery.ucl.ac.uk/98231/

Walden, J. B. (2006). Estimating vessel efficiency using a bootstrapped data envelopment analysis model. *Marine Resource Economics*, *21*(The MRE Foundation), 181–192.

Wegman, F., J. Commandeur, E. Doveh, V. Eksler, V. Gitelman, S. Hakkert, S. Oppe. (2008). SUNflowerNext: Towards a composite road safety performance index. Deliverable D6.16 of the EU FP6 project SafetyNet. Retrieved October 29, 2013, from http://www.swov.nl/rapport/sunflower/sunflowernext.pdf

Wierwille, W. W., Hanowski, R. J., Hankey, J. M., Kieliszewski, C. A., Lee, S. E., Medina, A., … Dingus, T. A. (2002). Identification and evaluation of driver errors: overview and recommendations. Retrieved from http://trid.trb.org/view.aspx?id=724268

Wilmots, B., Shen, Y., Hermans, E., & Ruan, D. (2011). Missing data treatment – Overview of possible solutions. *Diepenbeek: Policy Research Centre Mobility and Public Works, Track Traffic Safety*, (RA-MOW-2011- 002), 1–35.

WHO (2004). WHO | *World report on road traffic injury prevention* (p. 217). Geneva: World Health Organization. Retrieved from http://whqlibdoc.who.int/publications/2004/9241562609.pdf

WHO (2013). WHO | Global status report on road safety. World Health Organization. Retrieved October 29, 2013, from http://www.who.int/violence_injury_prevention/road_safety_status/2013/report/en/index.html

Yannis, G., Antoniou, C., Papadimitriou, E., & Katsochis, D. (2011). When may road fatalities start to decrease? *Journal of Safety Research*, *42*, 17–25.

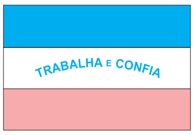
Yannis, G., Papadimitriou, E., & Folla, K. (2013). Effect of GDP changes on road traffic fatalities. *Safety Science*, *63*, 42–49.

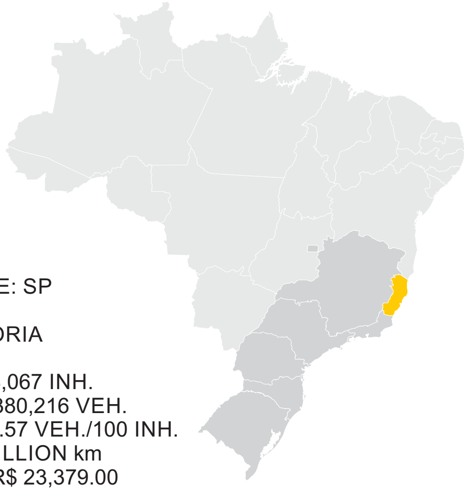
Yuan, Y. C. (2014). *Multiple Imputation for Missing Data: Concepts and New Development*. Rockville, USA: SAS Institute.

Yunes, J. (1972). The population of Brazil. *Revista de Saúde Pública*, *6*, 393–404.

# APPENDIX A

# ESPÍRITO SANTO - ES



CLUSTER: 1
BENCHMARK STATE: SP

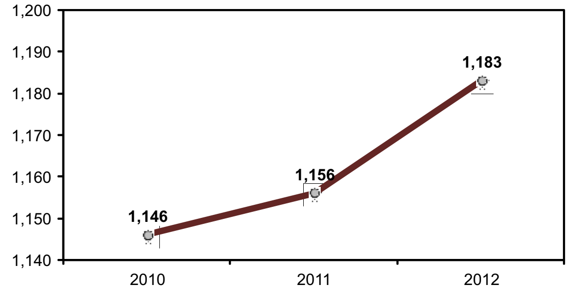
CAPITAL CITY: VITÓRIA

POPULATION: 3,578,067 INH.
VEHICLE FLEET: 1,380,216 VEH.
MOTORIZATION: 38.57 VEH./100 INH.
TOTAL VKT: 17.05 BILLION km
GDP *PER CAPITA*: R$ 23,379.00

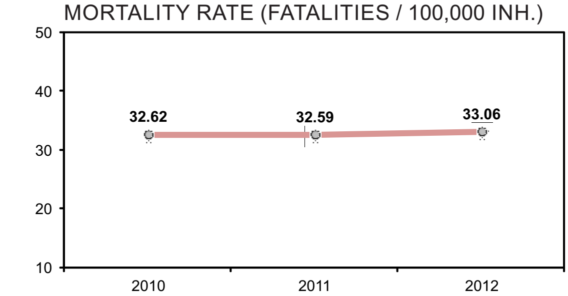
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **18% (8th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES: CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **41.9%** | 1,160 | 218 | 168 - 275 |
|  | **44.2%** | 412 | 62 | 50 - 80 |
|  | **22.5%** | 448 | 13 | 7 - 23 |
|  | **41.3%** | 21 | 4 | 3 - 5 |
|  | **100.0%** | 1 | n.a | n.a |
|  | - | 279 | 128 | 45 - 219 |

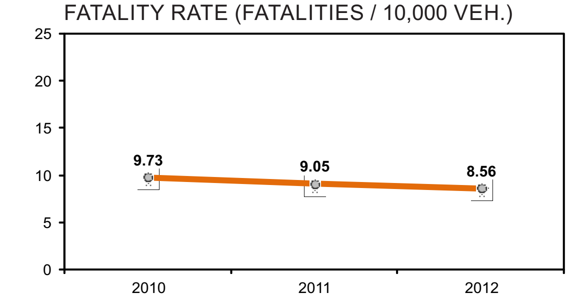
TRAFFIC FATALITY ABSOLUTE NUMBER



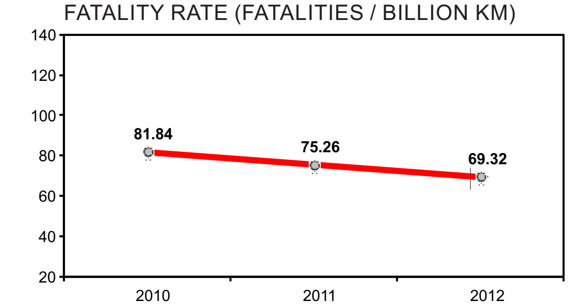

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

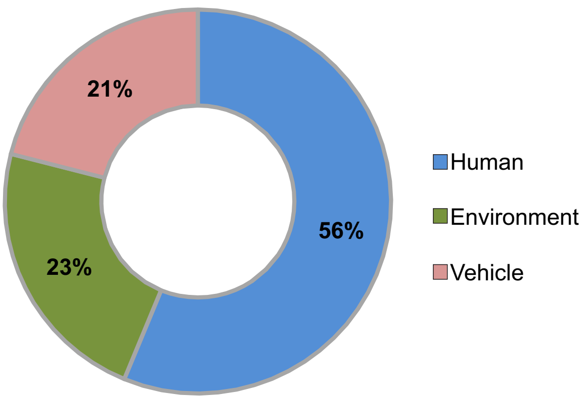



FATALITY RATE (FATALITIES / BILLION KM)

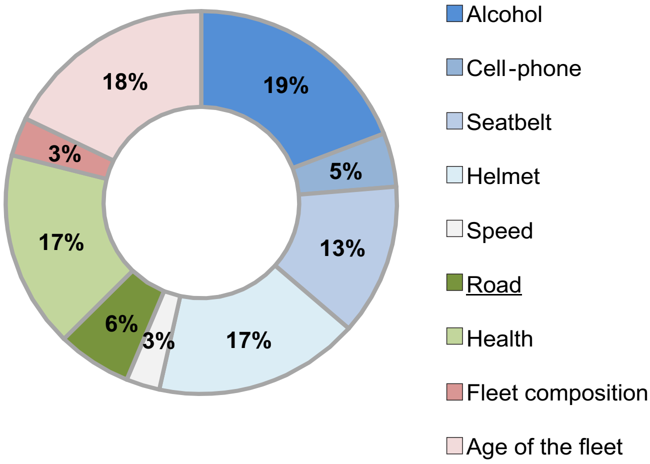



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **77% (4th)**

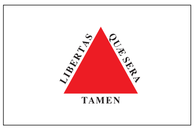
1st layer




2nd layer



# MINAS GERAIS - MG




CLUSTER: 1
BENCHMARK STATE: SP

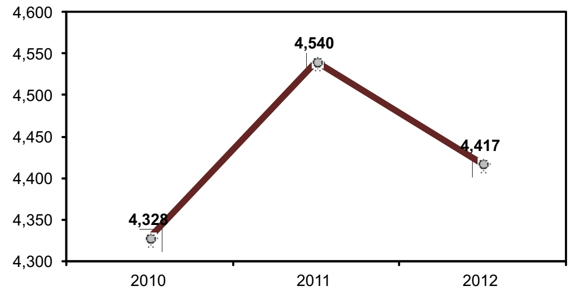
CAPITAL CITY: BELO HORIZONTE

POPULATION: 19,855,332 INH.
VEHICLE FLEET: 7,787,098 VEH.
MOTORIZATION: 39.22 VEH./100 INH.
TOTAL VKT: 97.01 BILLION km
GDP *PER CAPITA*: R$ 17,932.00

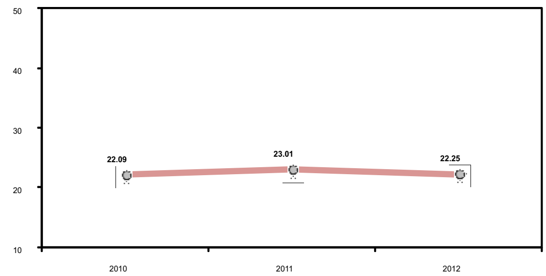
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **43% (5th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES | | |
|---|---|---|---|---|
| | | CURRENT | TARGET | RANGE |
| OVER ALL | **63.4%** | 4,390 | 1,847 | 1,602 - 2,192 |
| | **99.3%** | 1,026 | 876 | 625 -1,026 |
| | **28.1%** | 2,086 | 103 | 45 - 210 |
| | **55.2%** | 146 | 45 | 38 - 55 |
| | **18.0%** | 28 | 1 | 0 - 2 |
| | - | 1,104 | 1,009 | 353 - n.a |

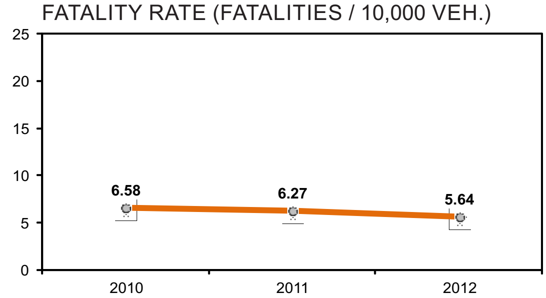
TRAFFIC FATALITY ABSOLUTE NUMBER



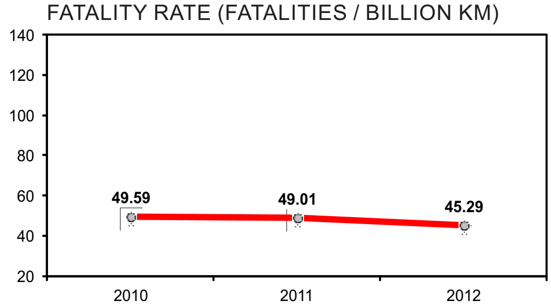

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

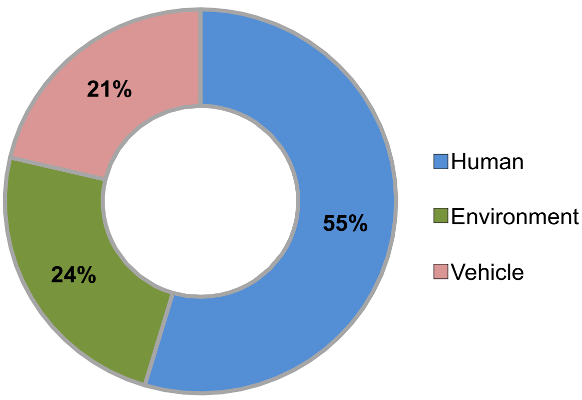



FATALITY RATE (FATALITIES / BILLION KM)

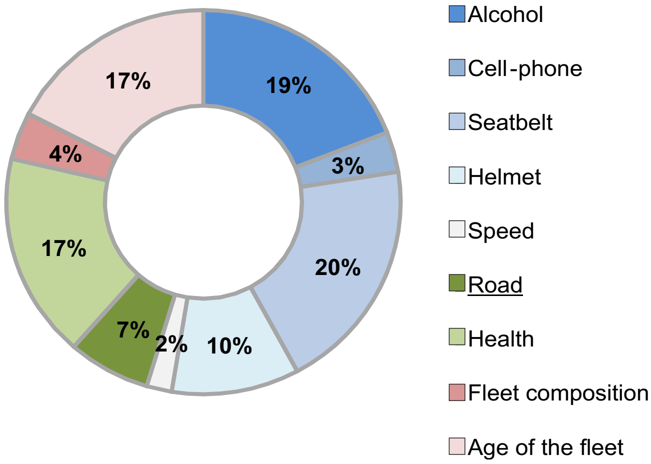



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **67% (8th)**

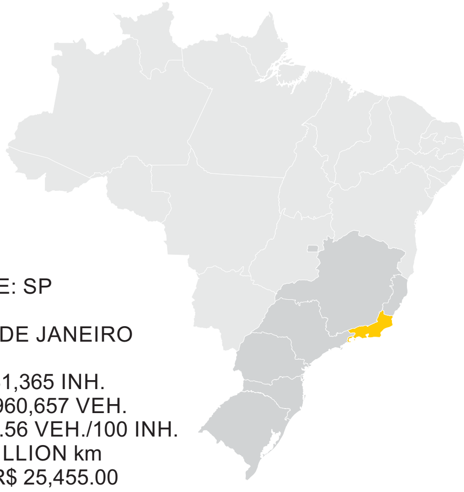
1st layer




2nd layer



# RIO DE JANEIRO - RJ



CLUSTER: 1
BENCHMARK STATE: SP

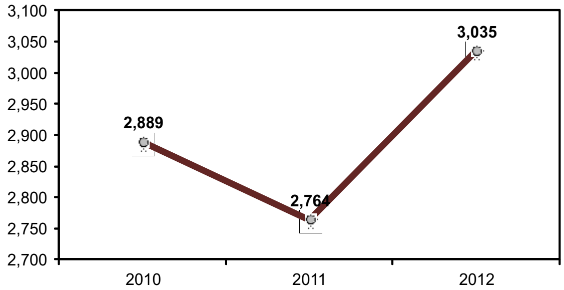
CAPITAL CITY: RIO DE JANEIRO

POPULATION: 16,231,365 INH.
VEHICLE FLEET: 4,960,657 VEH.
MOTORIZATION: 30.56 VEH./100 INH.
TOTAL VKT: 61.39 BILLION km
GDP *PER CAPITA*: R$ 25,455.00

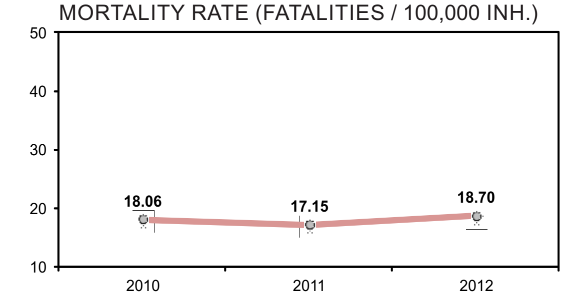
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **46% (4th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES: CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **67%** | 2,842 | 1,322 | 1,129 - 1,784 |
| | **74.0%** | 655 | 326 | 230 - 448 |
| | **82.9%** | 587 | 396 | 200 - n.a |
| | **36.2%** | 60 | 11 | 9 - 14 |
| | **17.0%** | 17 | 1 | 0 - 1 |
| | - | 1,522 | 562 | 56 - 1,254 |

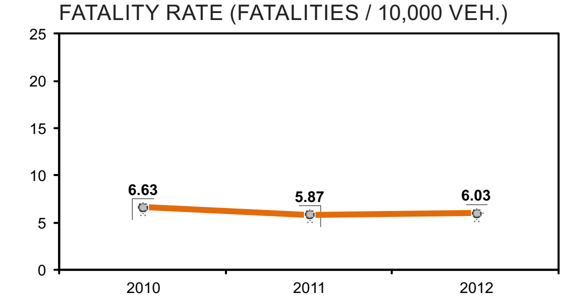
TRAFFIC FATALITY ABSOLUTE NUMBER



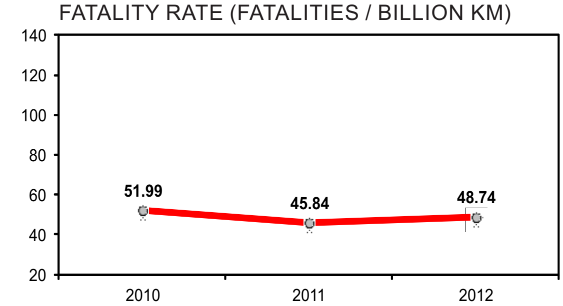

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

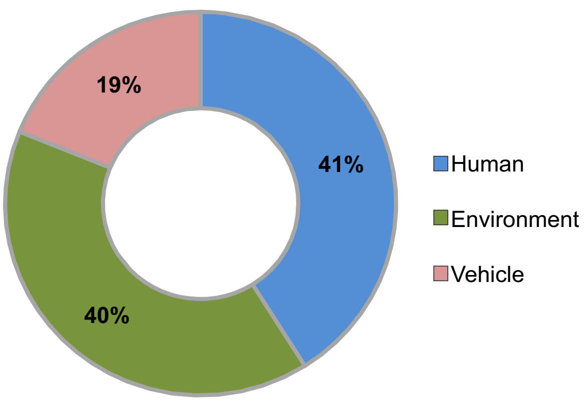



FATALITY RATE (FATALITIES / BILLION KM)

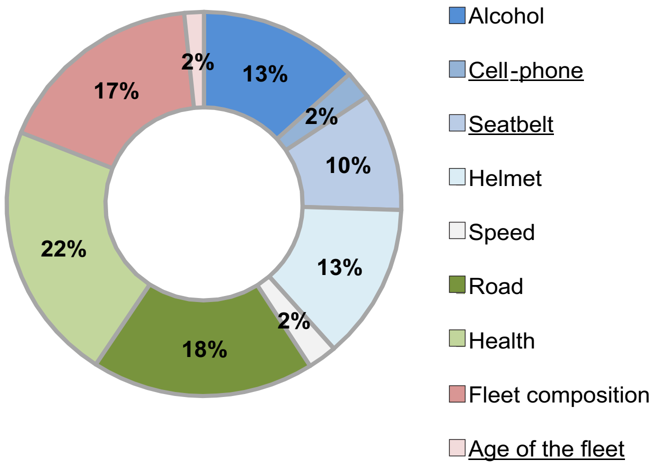



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **91% (3rd)**

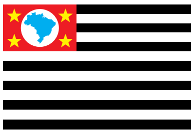
1st layer




2nd layer



# SÃO PAULO - SP



CLUSTER: 1
BENCHMARK STATE: -

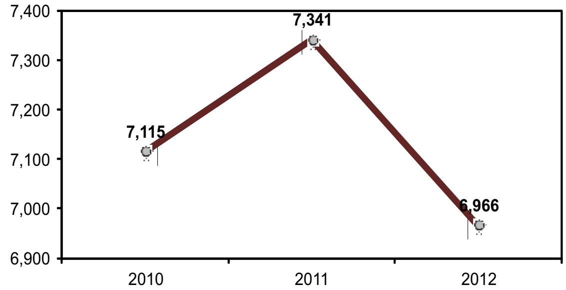
CAPITAL CITY: SÃO PAULO

POPULATION: 41,901,219 INH.
VEHICLE FLEET: 22,198,115 VEH.
MOTORIZATION: 52.98 VEH./100 INH.
TOTAL VKT: 242.50 BILLION km
GDP *PER CAPITA*: R$ 30,243.00

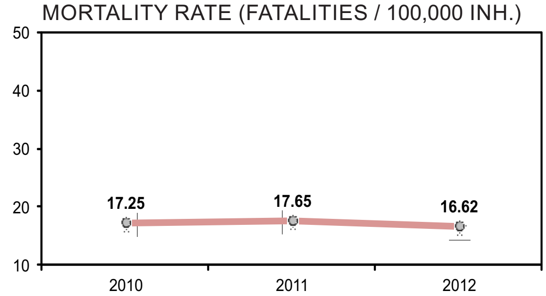
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **100% (1st)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES | | |
|---|---|---|---|---|
| | | CURRENT | TARGET | RANGE |
| OVER ALL | **100.0%** | 7,055 | n.a | n.a |
|  | **99.1%** | 2,220 | n.a | 1,828 - n.a |
|  | **100.0%** | 1,816 | n.a | 800 - n.a |
|  | **71.8%** | 164 | 90 | 67 - 103 |
|  | **19.8%** | 47 | 2 | 0 - 7 |
|  | - | 2,808 | n.a | 2,021 - n.a |

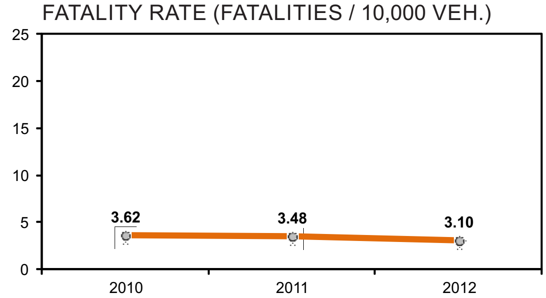
TRAFFIC FATALITY ABSOLUTE NUMBER



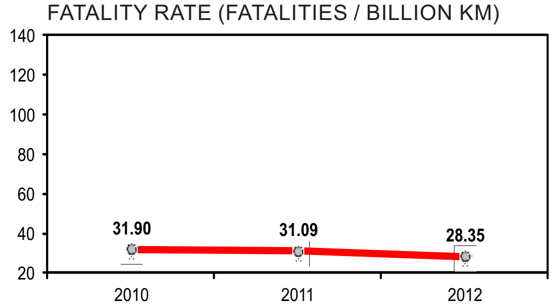

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

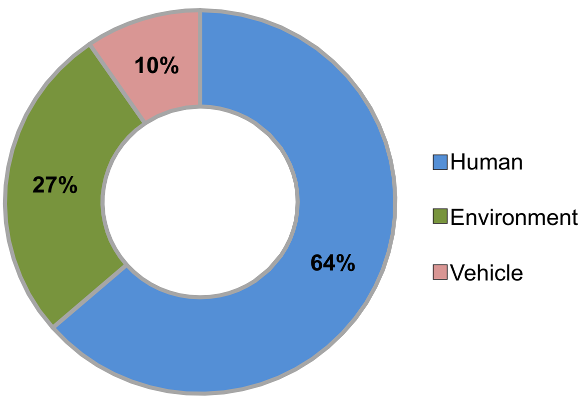



FATALITY RATE (FATALITIES / BILLION KM)

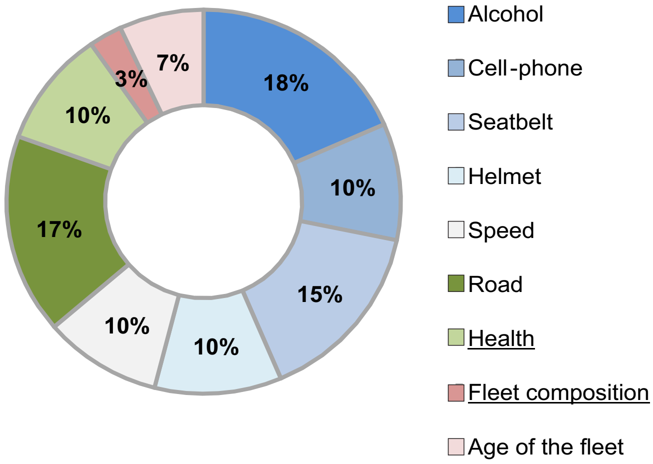



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **100% (1st)**

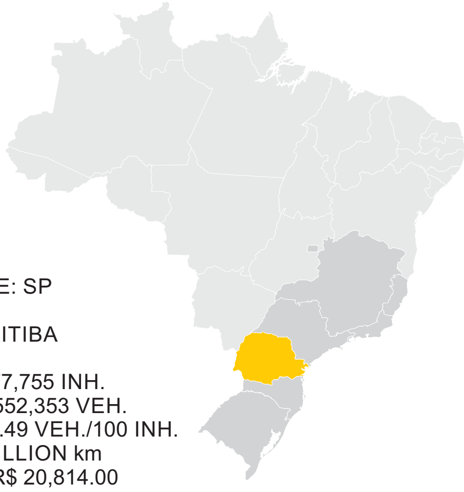
1st layer




2nd layer



# PARANÁ- PR



CLUSTER: 1
BENCHMARK STATE: SP

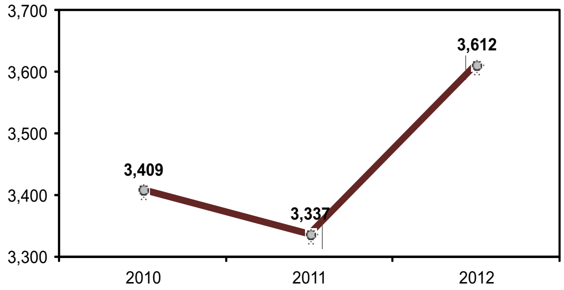
CAPITAL CITY: CURITIBA

POPULATION: 10,577,755 INH.
VEHICLE FLEET: 5,552,353 VEH.
MOTORIZATION: 52.49 VEH./100 INH.
TOTAL VKT: 61.99 BILLION km
GDP *PER CAPITA*: R$ 20,814.00

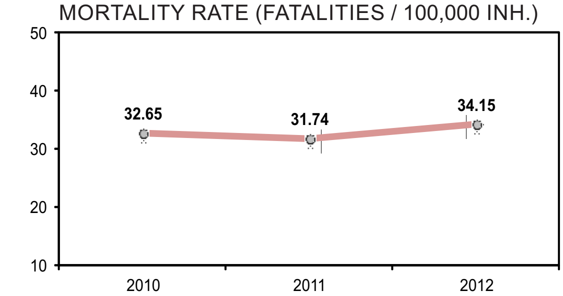
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **27% (7th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **51.0%** | 3,428 | 893 | 745 - 1,116 |
| | **52.0%** | 1,045 | 202 | 177 - 293 |
| | **33.9%** | 1,222 | 103 | 40 - 146 |
| | **46.6%** | 102 | 25 | 18 - 31 |
| | **15.1%** | 17 | 0 | 0 - 1 |
| | - | 1,042 | 471 | 186 - 836 |

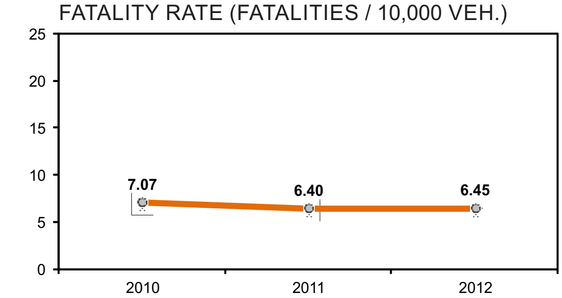
TRAFFIC FATALITY ABSOLUTE NUMBER



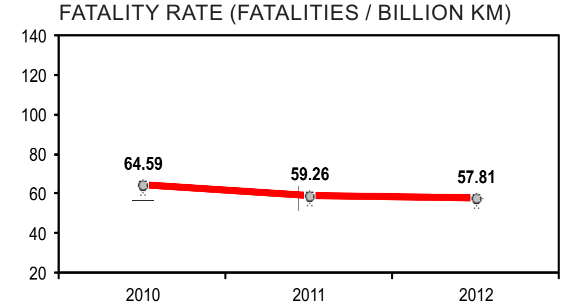

MORTALITY RATE (FATALITIES / 100,000 INH.)



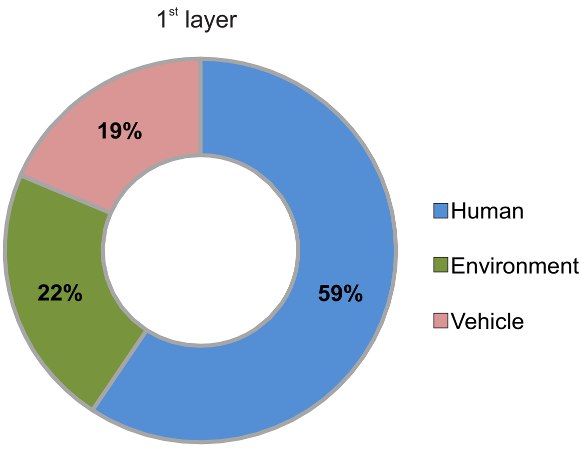

FATALITY RATE (FATALITIES / 10,000 VEH.)

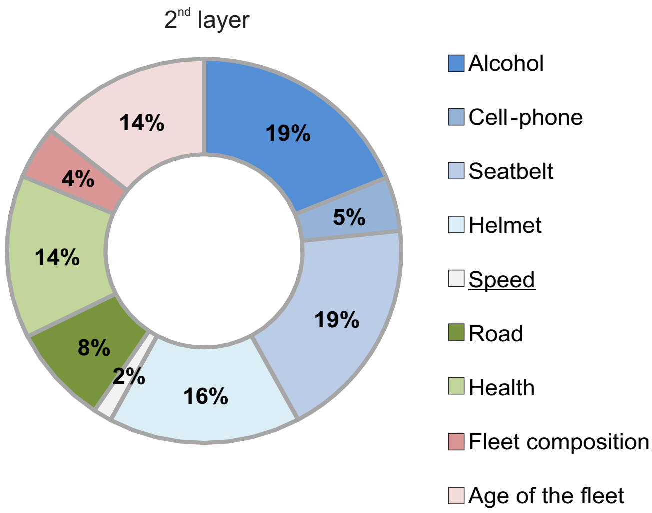



FATALITY RATE (FATALITIES / BILLION KM)

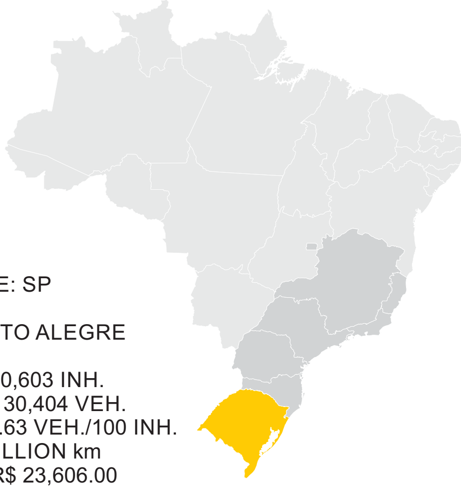



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **72% (7th)**



# RIO GRANDE DO SUL- RS



CLUSTER: 1
BENCHMARK STATE: SP

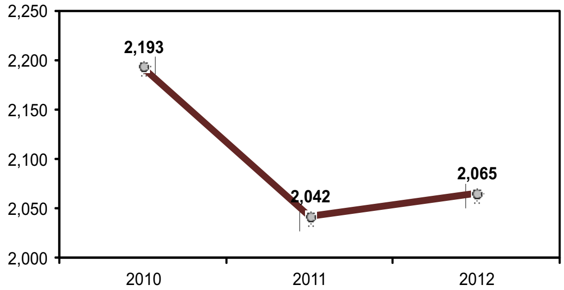
CAPITAL CITY: PORTO ALEGRE

POPULATION: 10,770,603 INH.
VEHICLE FLEET: 5,130,404 VEH.
MOTORIZATION: 47.63 VEH./100 INH.
TOTAL VKT: 57.74 BILLION km
GDP *PER CAPITA*: R$ 23,606.00

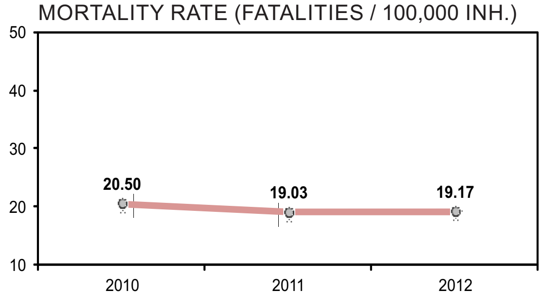
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **62% (2nd)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES: CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **80.3%** | 2,092 | 1,402 | 1,159 - 1,769 |
|  | **94.9%** | 565 | 426 | 313 - 537 |
|  | **53.3%** | 741 | 152 | 65 - 262 |
|  | **52.4%** | 76 | 22 | 17 - 26 |
|  | **20.6%** | 12 | 1 | 0 - 2 |
|  | - | 698 | n.a | 169 - n.a |

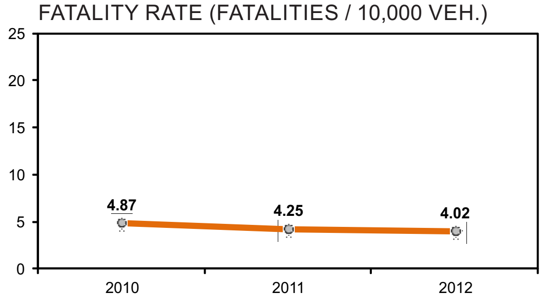
TRAFFIC FATALITY ABSOLUTE NUMBER



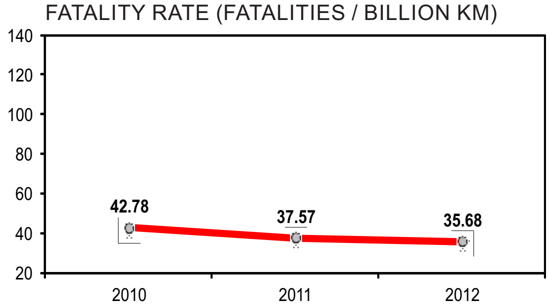

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

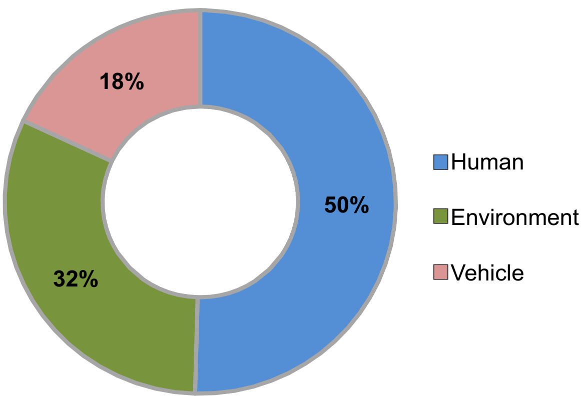



FATALITY RATE (FATALITIES / BILLION KM)

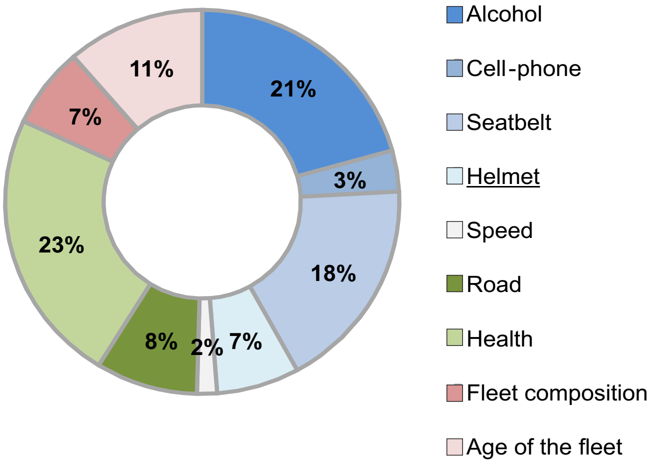



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **73% (6th)**

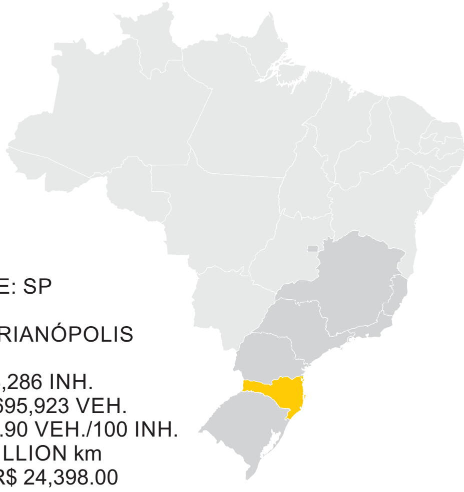
1st layer




2nd layer



# SANTA CATARINA- SC



CLUSTER: 1
BENCHMARK STATE: SP

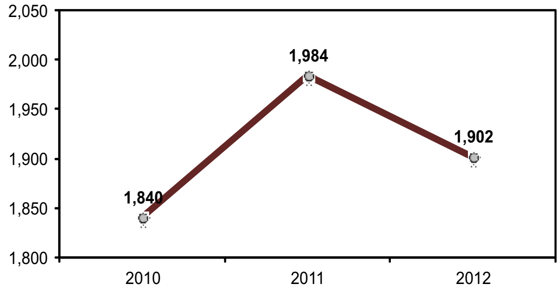
CAPITAL CITY: FLORIANÓPOLIS

POPULATION: 6,383,286 INH.
VEHICLE FLEET: 3,695,923 VEH.
MOTORIZATION: 57.90 VEH./100 INH.
TOTAL VKT: 43.03 BILLION km
GDP *PER CAPITA*: R$ 24,398.00

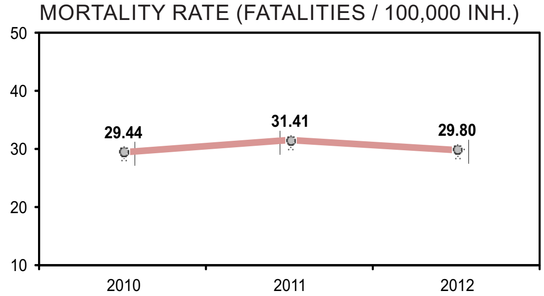
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **37% (6th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **61.7%** | 1,900 | 683 | 593 - 892 |
|  | **66.5%** | 666 | 216 | 161 - 289 |
|  | **41.7%** | 682 | 109 | 33 - 170 |
|  | **47.4%** | 57 | 14 | 12 - 17 |
|  | **9.5%** | 15 | 0 | n.a |
|  | **-** | 480 | 352 | 65 - n.a |

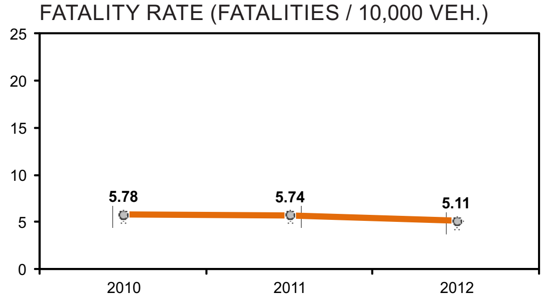
TRAFFIC FATALITY ABSOLUTE NUMBER



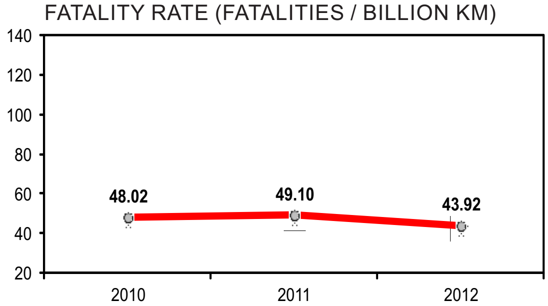

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

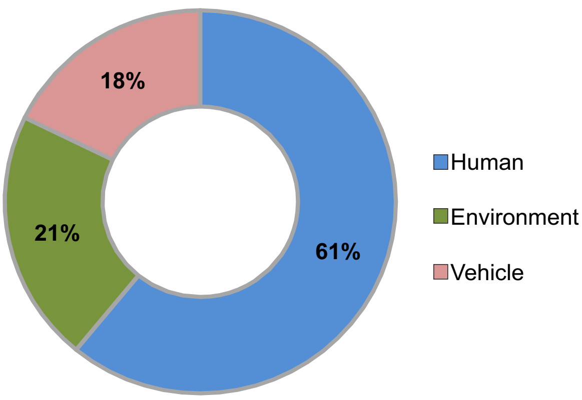



FATALITY RATE (FATALITIES / BILLION KM)

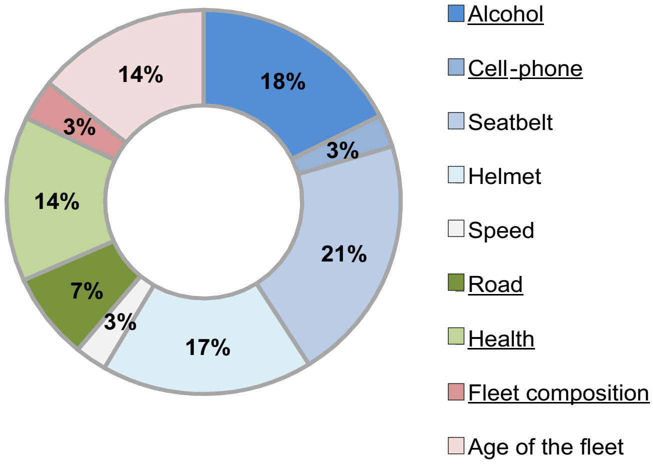



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **74% (5th)**

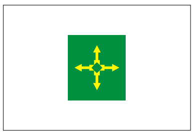
1st layer

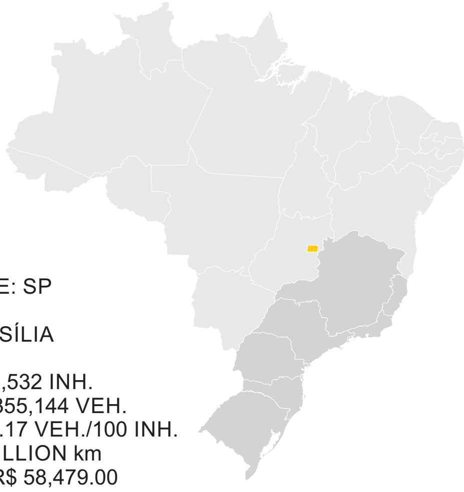



2nd layer



# DISTRITO FEDERAL- DF





CLUSTER: 1
BENCHMARK STATE: SP

CAPITAL CITY: BRASÍLIA

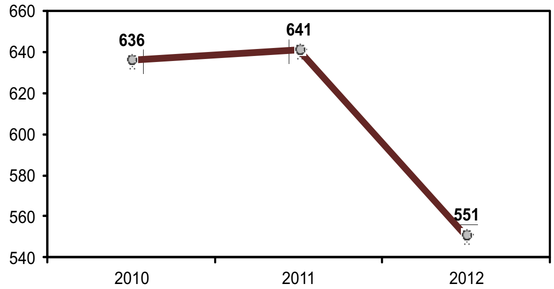
POPULATION: 2,648,532 INH.
VEHICLE FLEET: 1,355,144 VEH.
MOTORIZATION: 51.17 VEH./100 INH.
TOTAL VKT: 15.91 BILLION km
GDP *PER CAPITA*: R$ 58,479.00

ROAD SAFETY PERFORMANCE (INDEX-SCORE): **60% (3rd)**

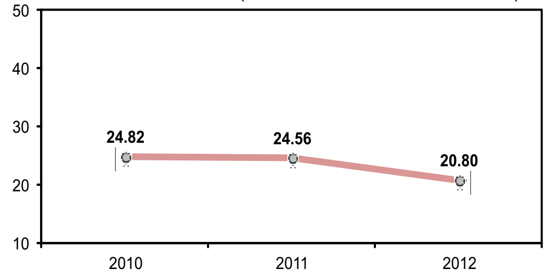
NUMBER OF TRAFFIC FATALITIES

| | SCORE | CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **73.0%** | 609 | 349 | 270 - 434 |
| | **76.1%** | 138 | 67 | 54 - 84 |
| | **61.7%** | 225 | 87 | 27 - 123 |
| | **100.0%** | 6 | n.a | n.a |
| | **21.3%** | 2 | 0 | n.a |
| | - | 237 | 126 | n.a |

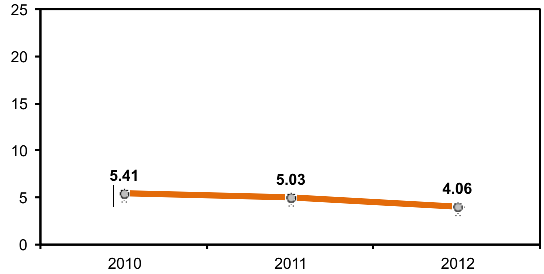
TRAFFIC FATALITY ABSOLUTE NUMBER



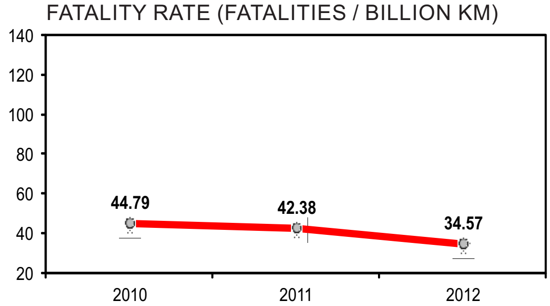

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

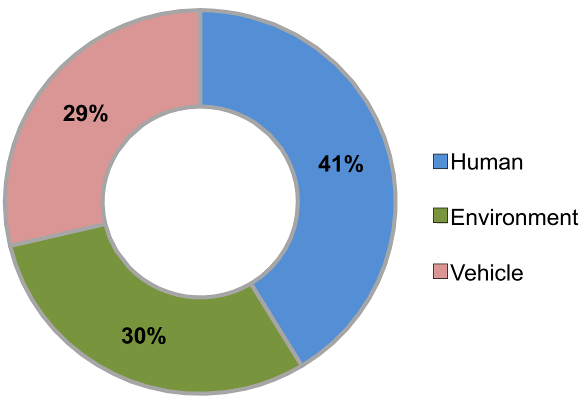



FATALITY RATE (FATALITIES / BILLION KM)

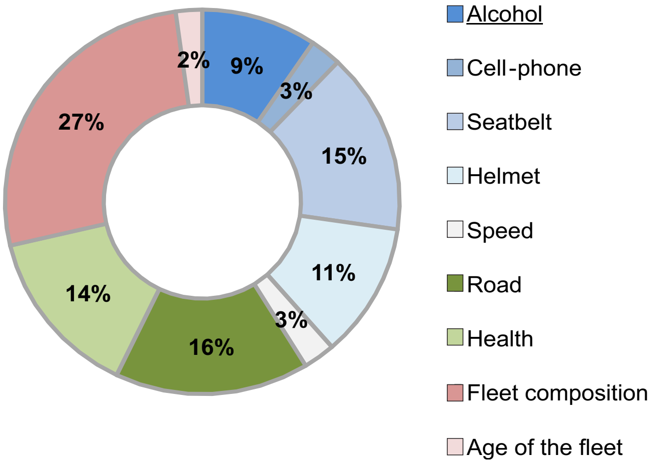



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **97% (2nd)**

1st layer




2nd layer



# ACRE- AC

CLUSTER: 2
BENCHMARK STATE: AM

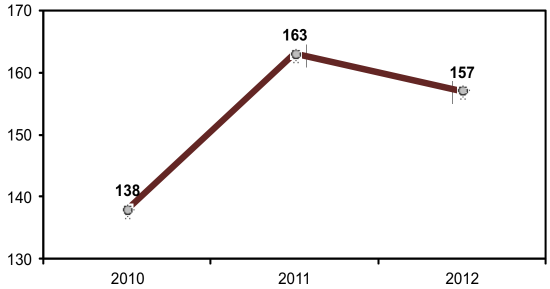
CAPITAL CITY: RIO BRANCO

POPULATION: 758,786 INH.
VEHICLE FLEET: 176,687 VEH.
MOTORIZATION: 23.29 VEH./100 INH.
TOTAL VKT: 3.10 BILLION km
GDP *PER CAPITA*: R$ 11,567.00

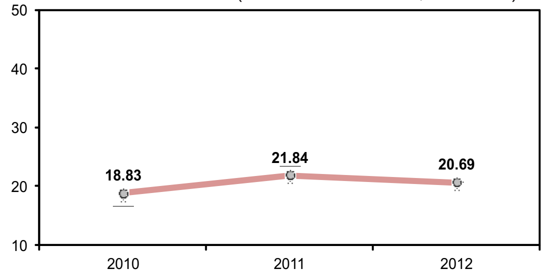
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **70% (4th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES: CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **81.1%** | 153 | 101 | 89 - 109 |
| | **65.4%** | 67 | 27 | 22 - 35 |
| | **50.4%** | 29 | 6 | 3 - 8 |
| | **100.0%** | 1 | n.a | 0 - n.a |
| | **27.9%** | 1 | 0 | 0 - n.a |
| | **-** | 55 | 74 | 51 - n.a |

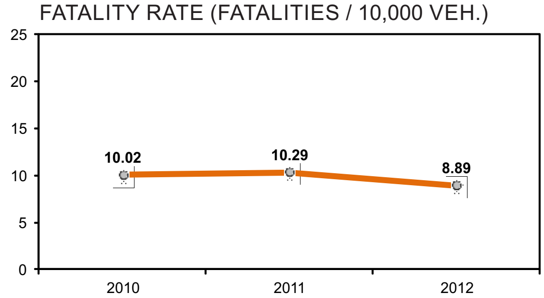
TRAFFIC FATALITY ABSOLUTE NUMBER



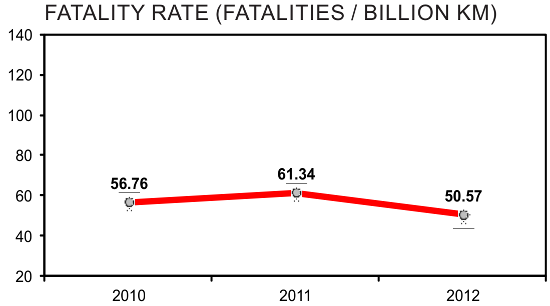

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

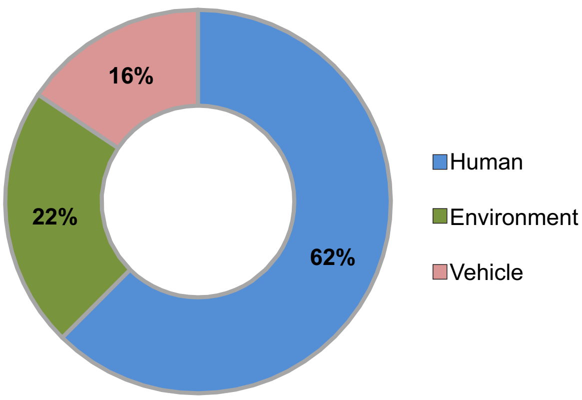



FATALITY RATE (FATALITIES / BILLION KM)

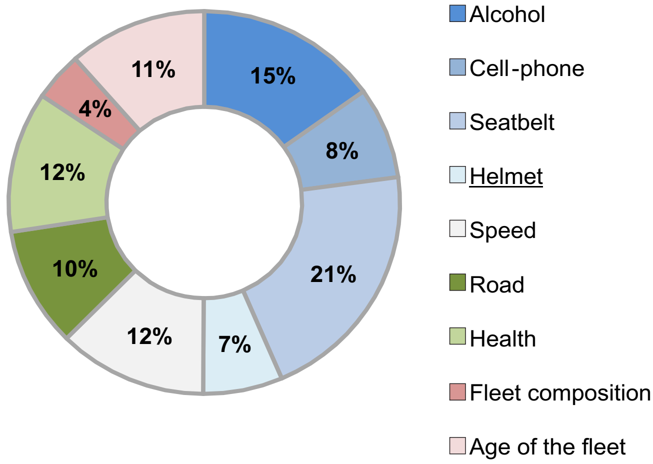



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **100% (1st)**

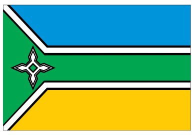
1st layer




2nd layer



# AMAPÁ - AP



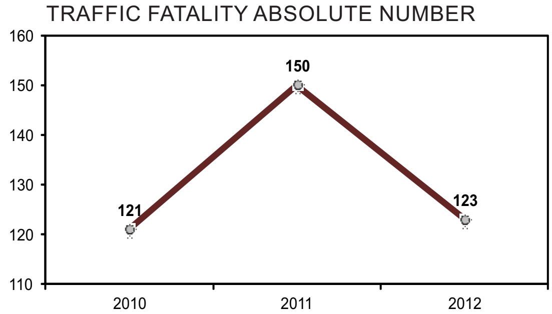
CLUSTER: 2
BENCHMARK STATE: AM

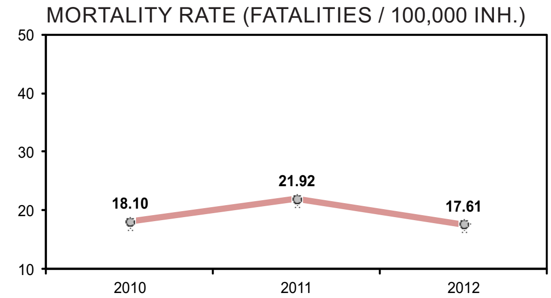
CAPITAL CITY: MACAPÁ

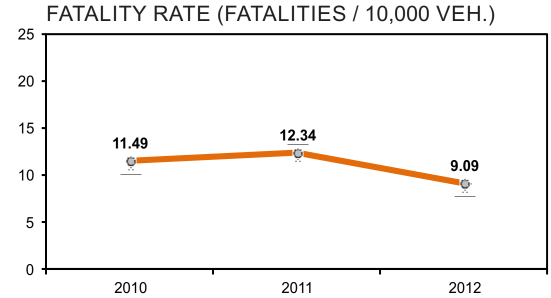
POPULATION: 698,602 INH.
VEHICLE FLEET: 134,148 VEH.
MOTORIZATION: 19.20 VEH./100 INH.
TOTAL VKT: 3.04 BILLION km
GDP *PER CAPITA*: R$ 12,361.00

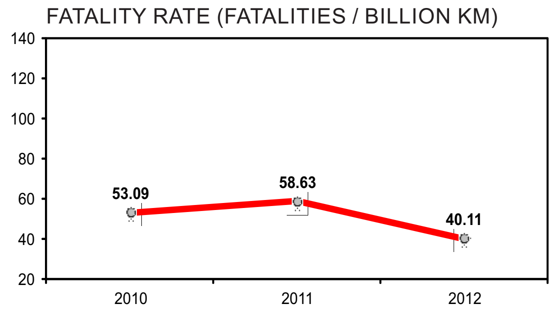
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **74% (2nd)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES | | |
|---|---|---|---|---|
| | | CURRENT | TARGET | RANGE |
| OVER ALL | **82.9%** | 131 | 92 | 83 - 104 |
| Motorcycle | **100.0%** | 33 | n.a. | n.a. |
| Car | **100.0%** | 17 | 14 | 6 - n.a. |
| Truck | - | 0 | - | - |
| Bus | **15.1%** | 2 | 0 | n.a. |
| Pedestrian/Cyclist | - | 78 | 39 | 30 - 58 |

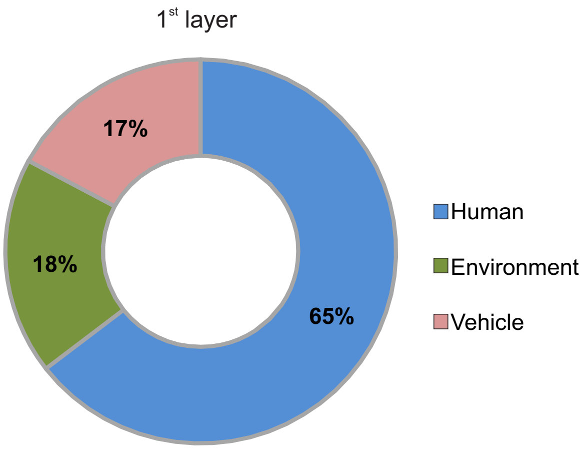


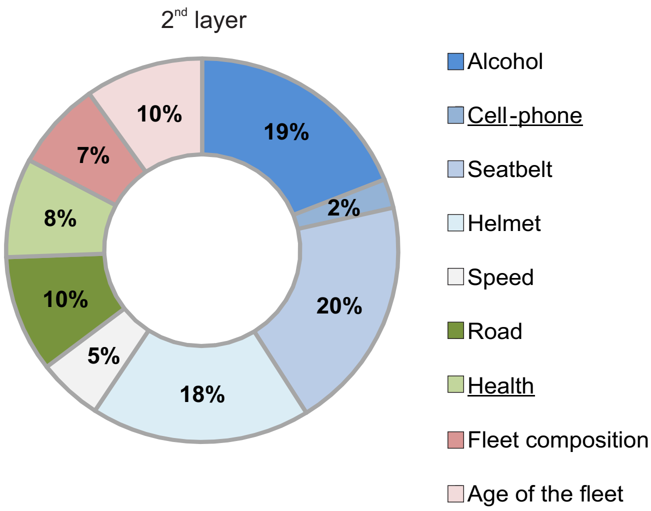



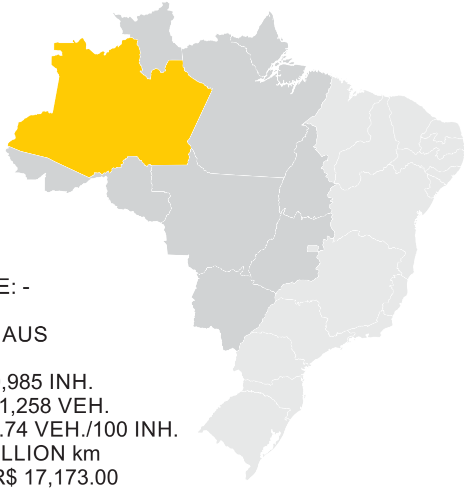




SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **96% (3rd)**



# AMAZONAS - AM



CLUSTER: 2
BENCHMARK STATE: -

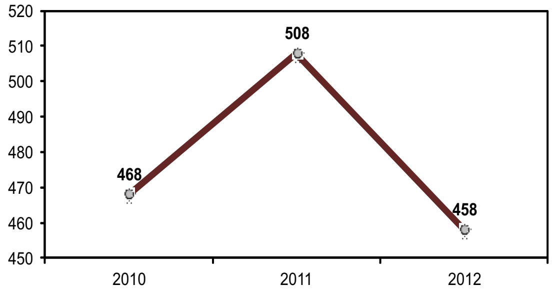
CAPITAL CITY: MANAUS

POPULATION: 3,590,985 INH.
VEHICLE FLEET: 601,258 VEH.
MOTORIZATION: 16.74 VEH./100 INH.
TOTAL VKT: 11.07 BILLION km
GDP *PER CAPITA*: R$ 17,173.00

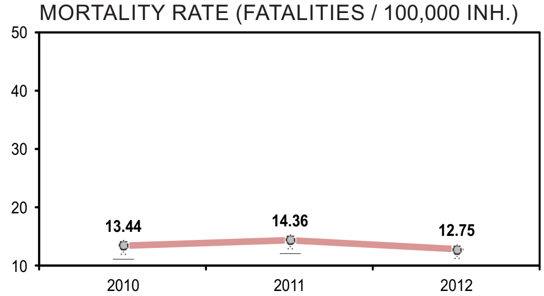
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **100% (1st)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES | | |
|---|---|---|---|---|
| | | CURRENT | TARGET | RANGE |
| OVER ALL | **100.0%** | 477 | n.a. | n.a. |
|  | **82.7%** | 145 | 114 | 95 - 128 |
|  | **98.6%** | 74 | 46 | 31 - n.a. |
|  | **49.8%** | 7 | 2 | 0 - 4 |
|  | **63.5%** | 2 | 0 | 0 - 1 |
|  | - | 248 | 240 | n.a. |

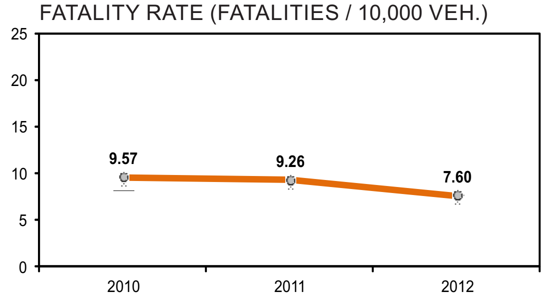
TRAFFIC FATALITY ABSOLUTE NUMBER



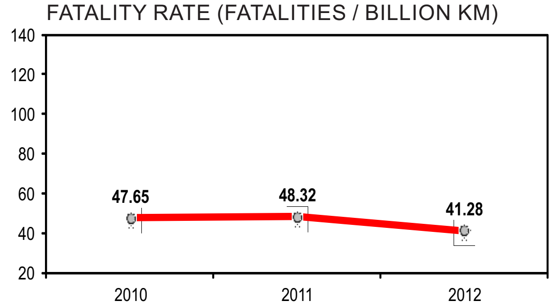

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

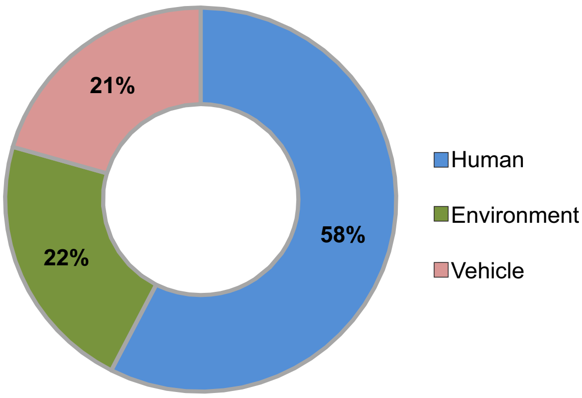



FATALITY RATE (FATALITIES / BILLION KM)

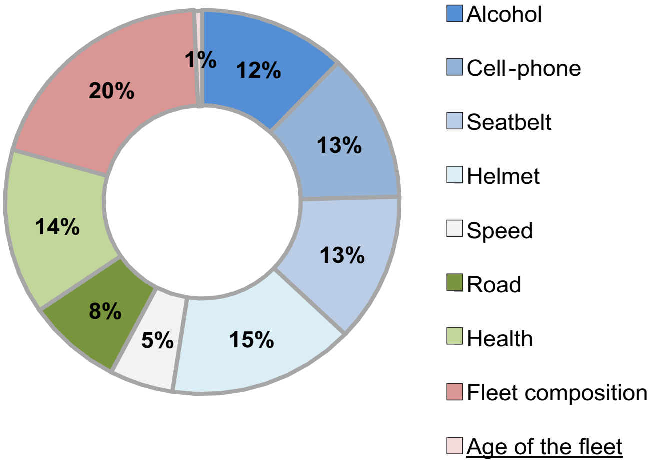



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **93% (5th)**

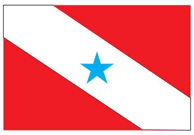
1st layer

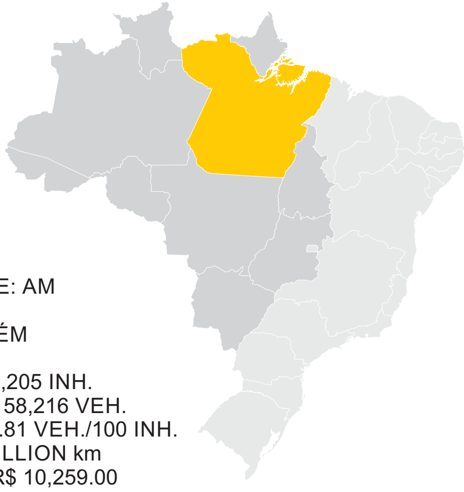



2nd layer



# PARÁ - PA





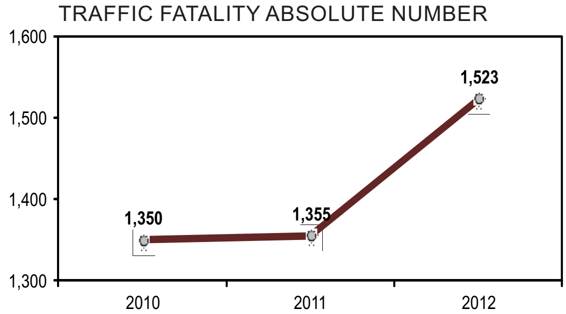
CLUSTER: 2
BENCHMARK STATE: AM

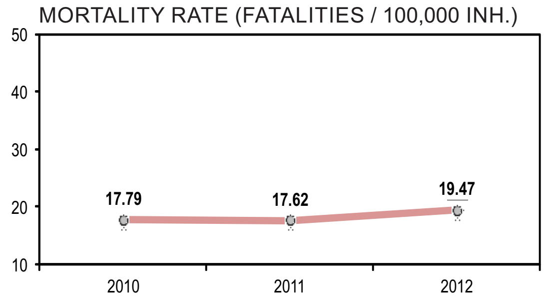
CAPITAL CITY: BELÉM

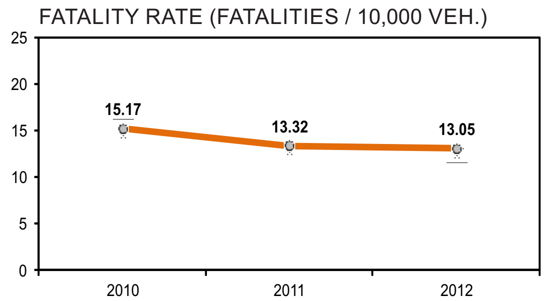
POPULATION: 7,822,205 INH.
VEHICLE FLEET: 1,158,216 VEH.
MOTORIZATION: 14.81 VEH./100 INH.
TOTAL VKT: 23.58 BILLION km
GDP *PER CAPITA*: R$ 10,259.00

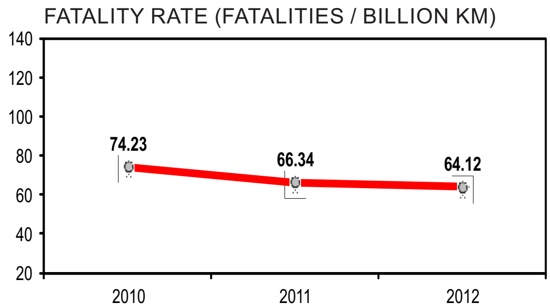
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **46% (8th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES: CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **67.5%** | 1,401 | 660 | 605 - 722 |
| | **62.4%** | 503 | 193 | 167 - 222 |
| | **65.2%** | 164 | 43 | 22 - 76 |
| | **44.7%** | 20 | 3 | 1 - 8 |
| | **37.0%** | 9 | 1 | 0 - n.a. |
| | - | 704 | 426 | 291 - 551 |

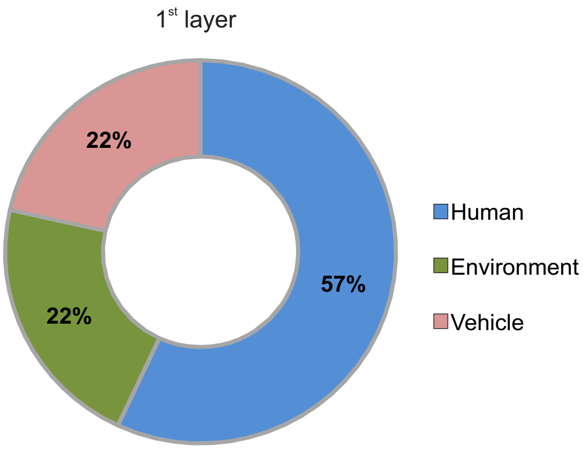


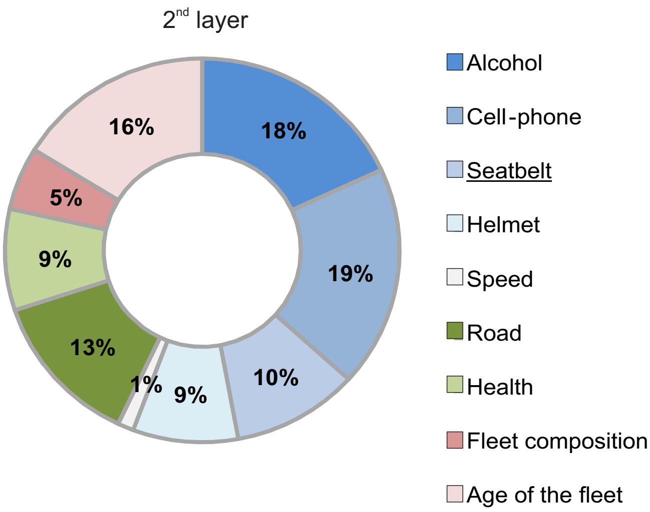



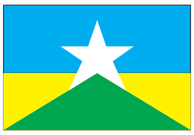



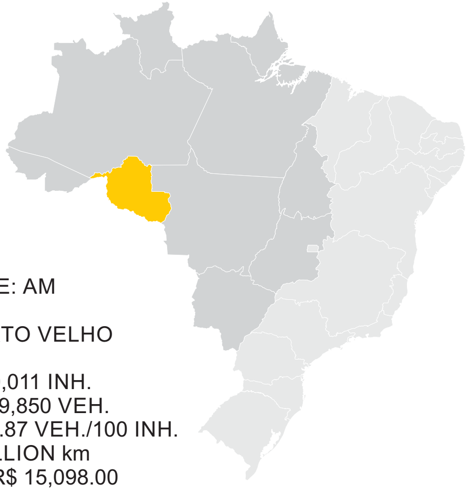

SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **72% (10th)**



# RONDÔNIA - RO





CLUSTER: 2
BENCHMARK STATE: AM

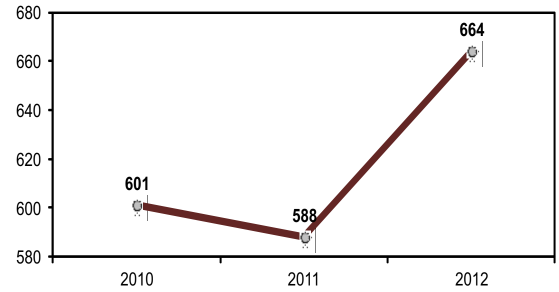
CAPITAL CITY: PORTO VELHO

POPULATION: 1,590,011 INH.
VEHICLE FLEET: 649,850 VEH.
MOTORIZATION: 40.87 VEH./100 INH.
TOTAL VKT: 9.99 BILLION km
GDP *PER CAPITA*: R$ 15,098.00

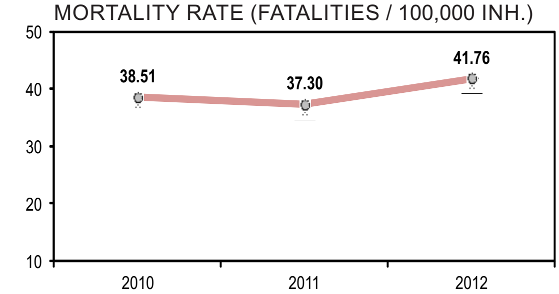
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **46% (9th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES | | |
|---|---|---|---|---|
| | | CURRENT | TARGET | RANGE |
| OVER ALL | **66.7%** | 614 | 287 | 264 - 310 |
|  | **55.9%** | 286 | 88 | 72 - 102 |
|  | **27.1%** | 163 | 8 | 3 - 13 |
|  | 18.0% | 23 | 1 | 0 - 3 |
|  | 34.2% | 4 | 0 | n.a. |
|  | - | 138 | .n.a | n.a. |

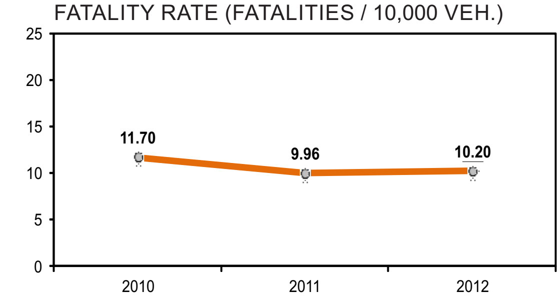
TRAFFIC FATALITY ABSOLUTE NUMBER



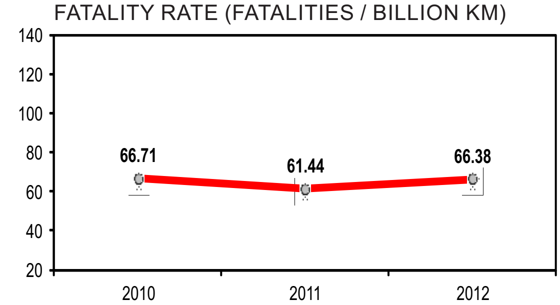

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

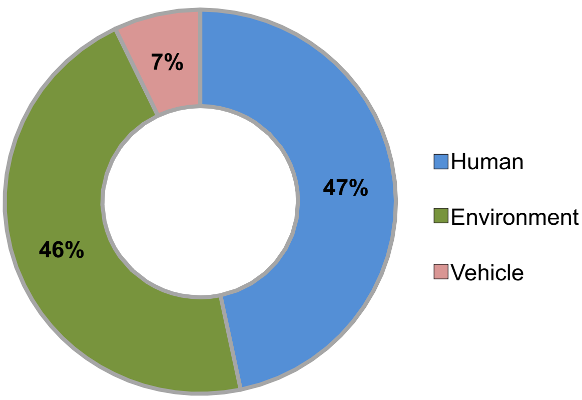



FATALITY RATE (FATALITIES / BILLION KM)

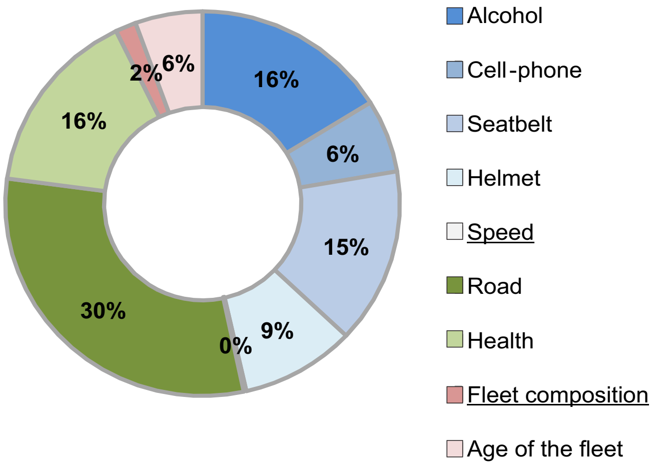



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **89% (6th)**

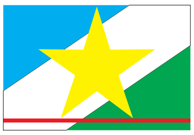
1st layer




2nd layer



# RORAIMA - RR



CLUSTER: 2
BENCHMARK STATE: AM

CAPITAL CITY: BOA VISTA

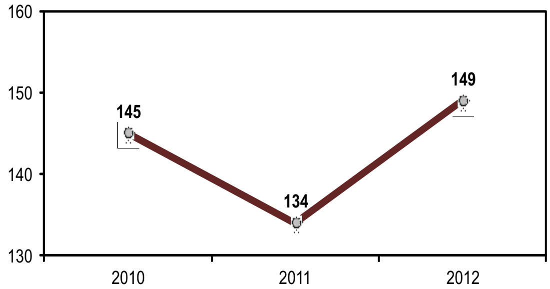
POPULATION: 469,524 INH.
VEHICLE FLEET: 143,858 VEH.
MOTORIZATION: 30.64 VEH./100 INH.
TOTAL VKT: 2.15 BILLION km
GDP *PER CAPITA*: R$ 14,052.00

ROAD SAFETY PERFORMANCE (INDEX-SCORE): **47% (7th)**

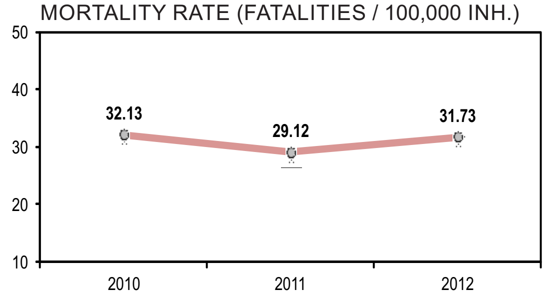
NUMBER OF TRAFFIC FATALITIES

| | SCORE | CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **65.9%** | 142 | 63 | 56 - 68 |
|  | **44.1%** | 76 | 14 | 12 - 16 |
|  | **38.3%** | 30 | 2 | 1 - 4 |
|  | - | 0 | - | - |
|  | **50.6%** | 0 | n.a. | n.a. |
|  | - | 36 | . 32 | n.a. |

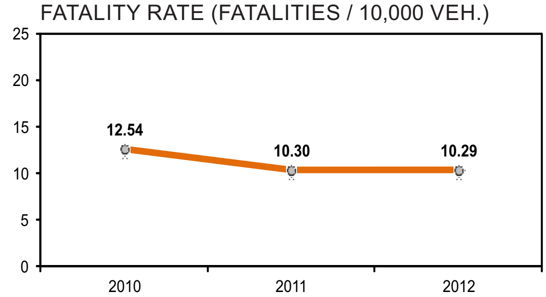
TRAFFIC FATALITY ABSOLUTE NUMBER



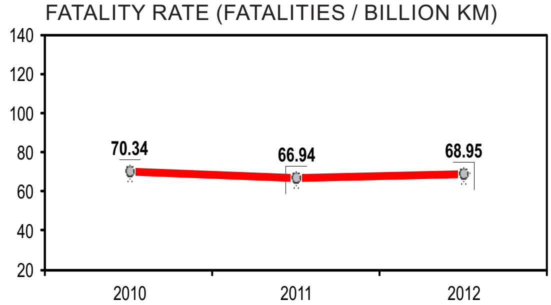

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

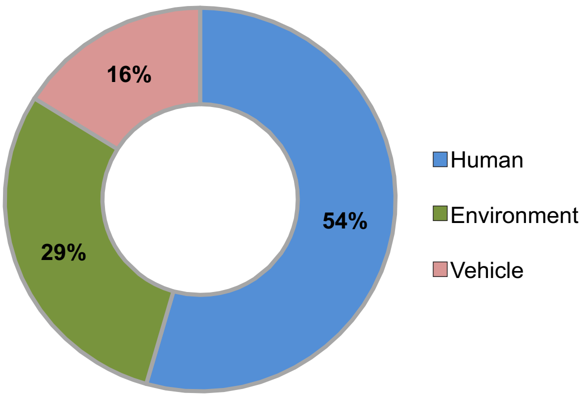



FATALITY RATE (FATALITIES / BILLION KM)

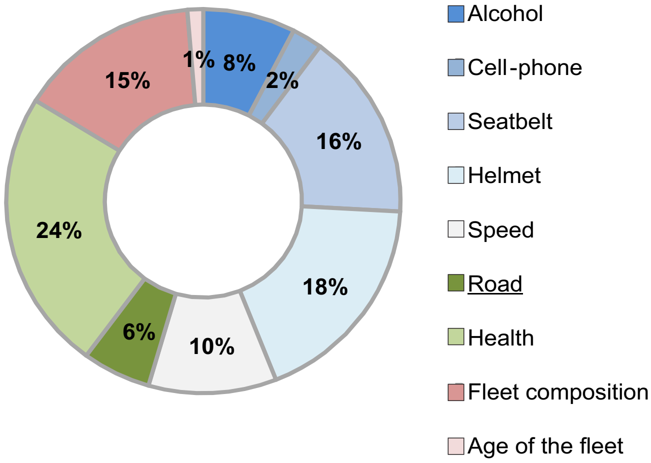



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **87% (8th)**

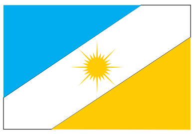
1st layer

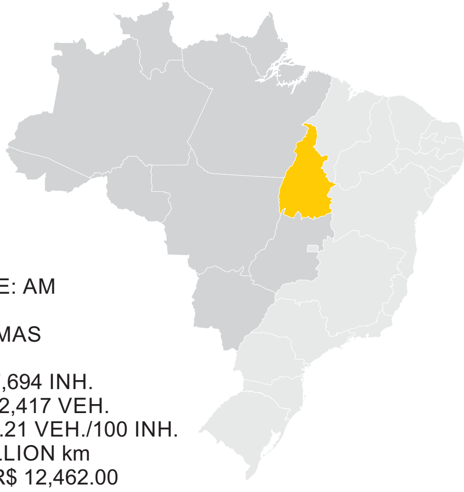



2nd layer



# TOCANTINS - TO





CLUSTER: 2
BENCHMARK STATE: AM

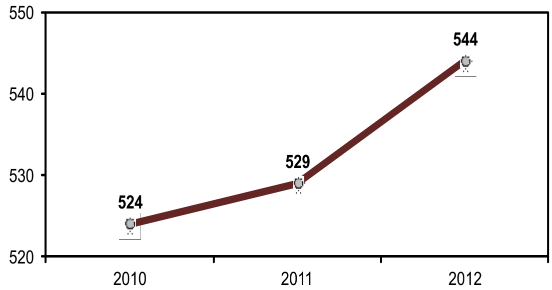
CAPITAL CITY: PALMAS

POPULATION: 1,417,694 INH.
VEHICLE FLEET: 442,417 VEH.
MOTORIZATION: 31.21 VEH./100 INH.
TOTAL VKT: 8.33 BILLION km
GDP *PER CAPITA*: R$ 12,462.00

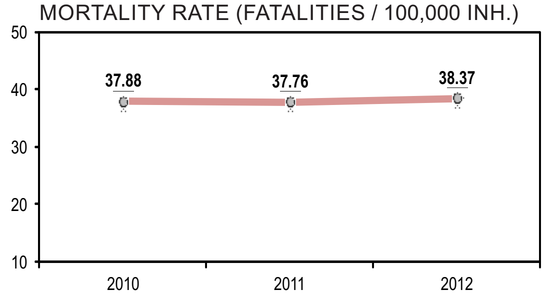
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **36% (10th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES | | |
|---|---|---|---|---|
| | | CURRENT | TARGET | RANGE |
| OVER ALL | **58.8%** | 532 | 190 | 169 - 200 |
|  | **55.1%** | 214 | 61 | 55 - 75 |
|  | **19.0%** | 203 | 4 | 2 - 8 |
|  | **24.9%** | 14 | 0 | 0 - 2 |
|  | **21.4%** | 5 | 0 | 0 - 1 |
|  | - | 96 | 71 | n.a. |

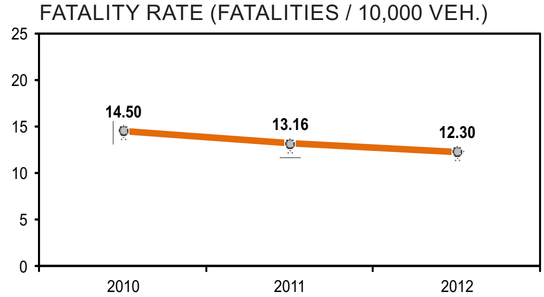
TRAFFIC FATALITY ABSOLUTE NUMBER



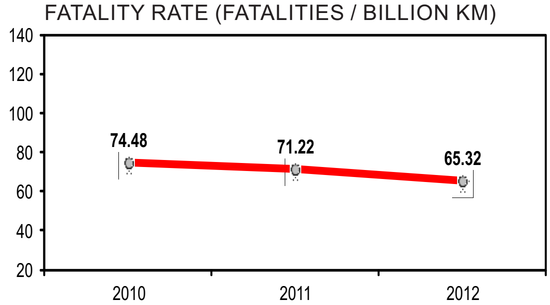

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

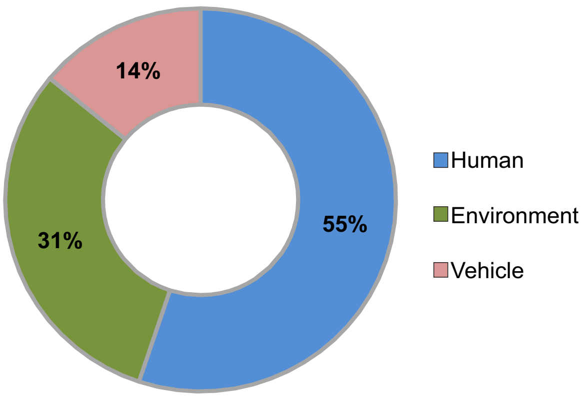



FATALITY RATE (FATALITIES / BILLION KM)

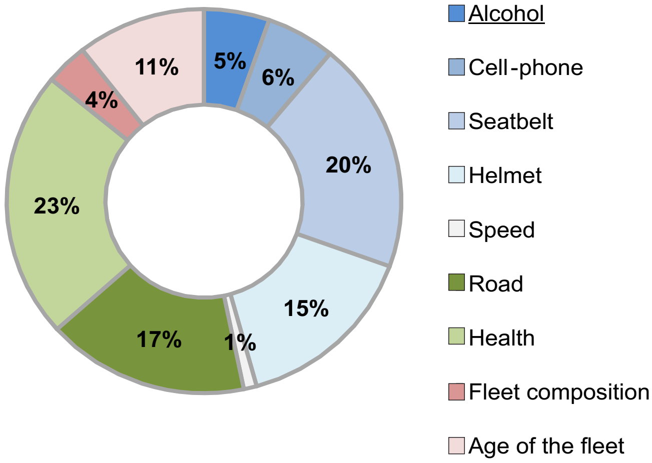



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **79% (9th)**

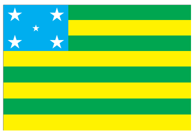
1st layer




2nd layer



# GOIÁS - GO



CLUSTER: 2
BENCHMARK STATE: AM

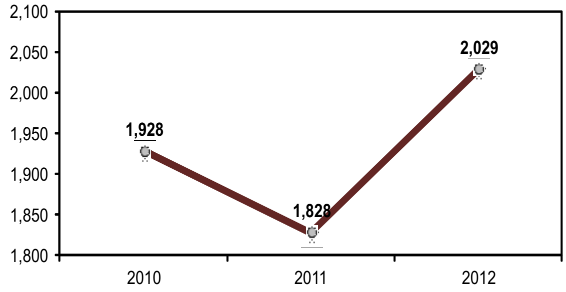
CAPITAL CITY: GOIÂNIA

POPULATION: 6,154,996 INH.
VEHICLE FLEET: 2,664,209 VEH.
MOTORIZATION: 43.29 VEH./100 INH.
TOTAL VKT: 37.77 BILLION km
GDP *PER CAPITA*: R$ 16,252.00

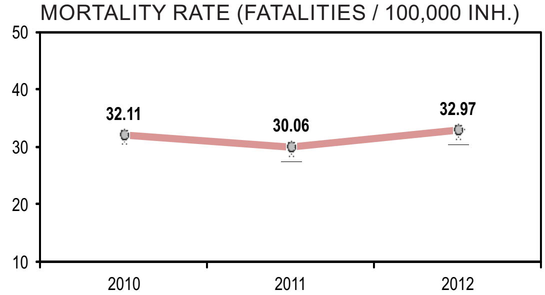
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **72% (3rd)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **83.0%** | 1,923 | 1,388 | 1,274 - 1,499 |
| | **65.6%** | 696 | 298 | 243 - 328 |
| | **35.6%** | 636 | 56 | 26 - 88 |
| | **16.0%** | 66 | 1 | 0 - 3 |
| | **19.8%** | 17 | 1 | 0 - n.a. |
| | - | 509 | n.a. | n.a. |

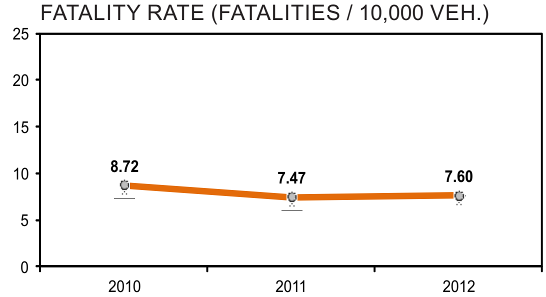
TRAFFIC FATALITY ABSOLUTE NUMBER



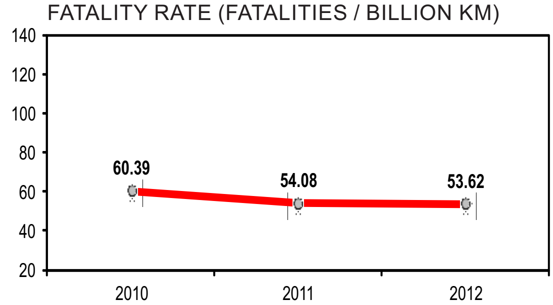

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

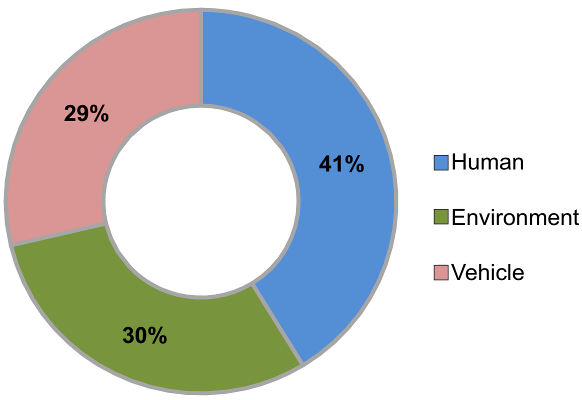



FATALITY RATE (FATALITIES / BILLION KM)

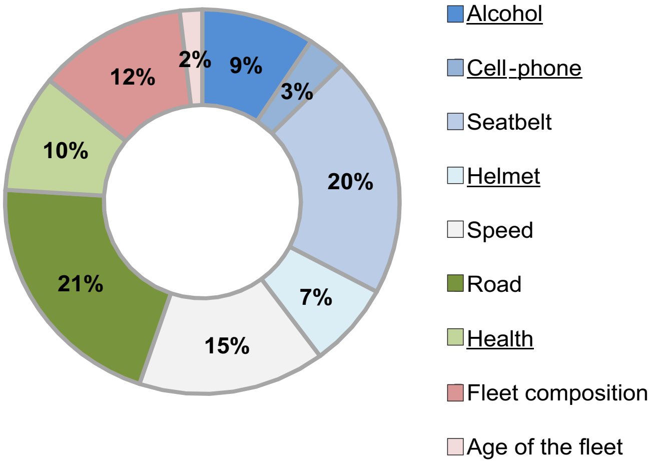



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **94% (4th)**

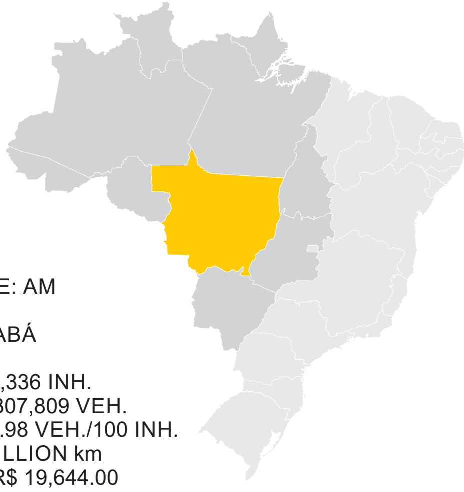
1st layer




2nd layer



# MATO GROSSO - MT



CLUSTER: 2
BENCHMARK STATE: AM

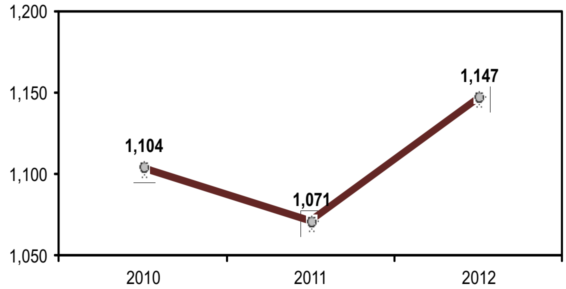
CAPITAL CITY: CUIABÁ

POPULATION: 3,115,336 INH.
VEHICLE FLEET: 1,307,809 VEH.
MOTORIZATION: 41.98 VEH./100 INH.
TOTAL VKT: 22.35 BILLION km
GDP *PER CAPITA*: R$ 19,644.00

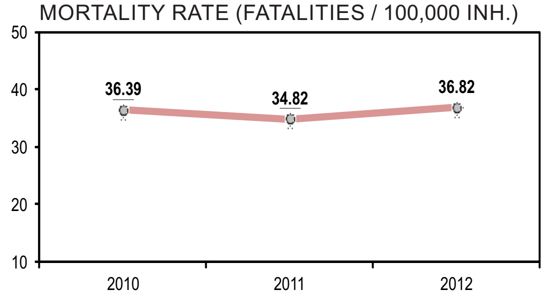
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **58% (6th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES - CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **75.6%** | 1,106 | 632 | 586 - 705 |
|  | **65.2%** | 457 | 194 | 163 - 225 |
|  | **29.1%** | 359 | 18 | 8 - 32 |
|  | **14.1%** | 61 | 1 | 0 - n.a. |
|  | **100.0%** | 2 | n.a. | 1 - n.a. |
|  | - | 226 | n.a | n.a. |

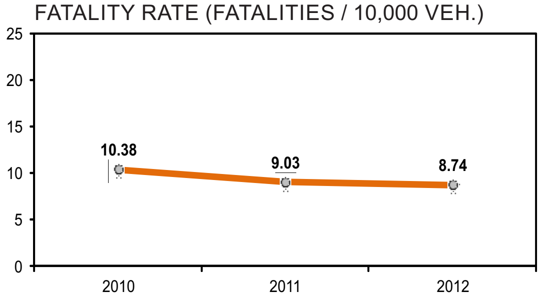
TRAFFIC FATALITY ABSOLUTE NUMBER



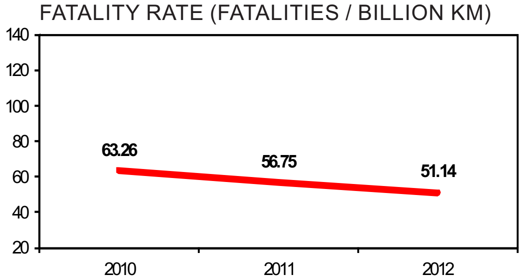

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

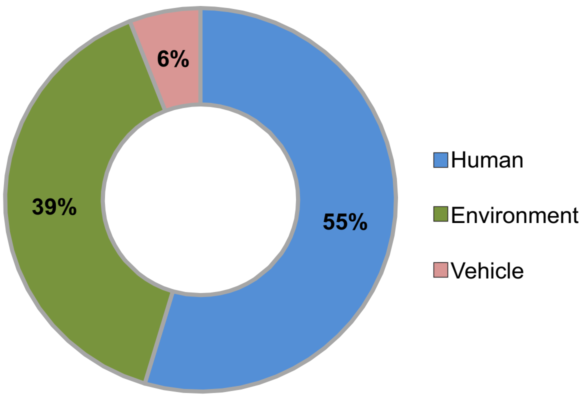



FATALITY RATE (FATALITIES / BILLION KM)

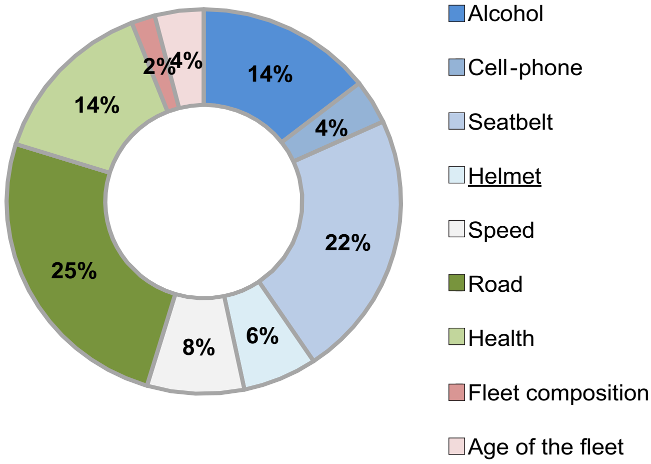



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **88% (7th)**

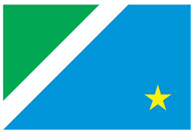
1st layer




2nd layer



# MATO GROSSO DO SUL - MS



CLUSTER: 2
BENCHMARK STATE: AM

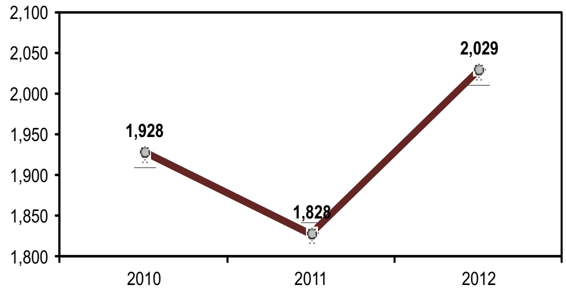
CAPITAL CITY: CAMPO GRANDE

POPULATION: 2,505,088 INH.
VEHICLE FLEET: 1,070,094 VEH.
MOTORIZATION: 42.72 VEH./100 INH.
TOTAL VKT: 15.92 BILLION km
GDP *PER CAPITA*: R$ 17,766.00

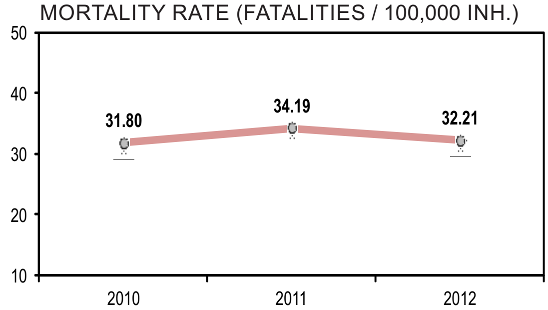
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **67% (5th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **80.8%** | 810 | 553 | 506 - 598 |
|  | **65.2%** | 301 | 117 | 102 - 146 |
|  | **33.3%** | 263 | 18 | 8 - 37 |
|  | **14.1%** | 41 | 1 | 0 - 2 |
|  | **44.0%** | 4 | 1 | 0 - n.a. |
|  | - | 201 | n.a. | n.a. |

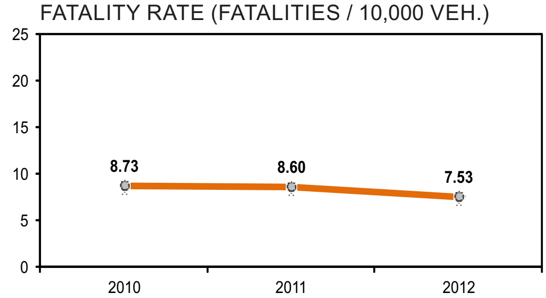
TRAFFIC FATALITY ABSOLUTE NUMBER



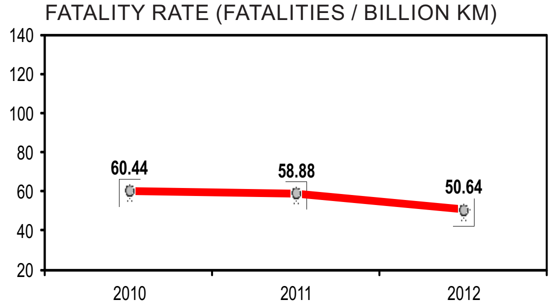

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

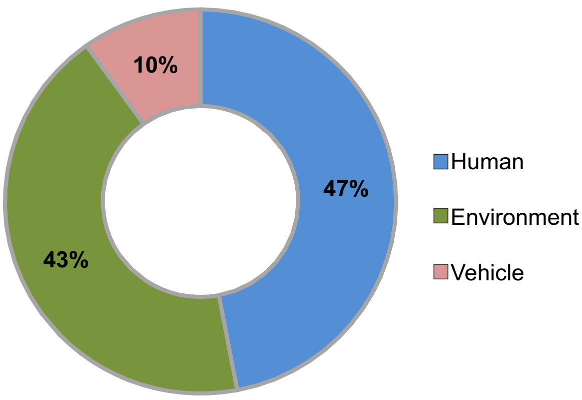



FATALITY RATE (FATALITIES / BILLION KM)

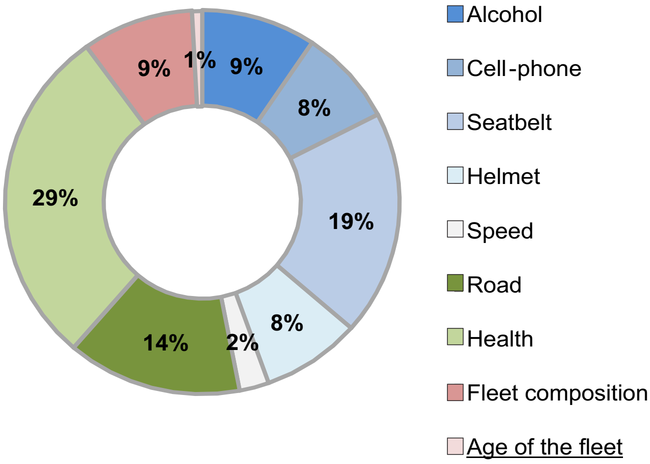



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **98% (2nd)**

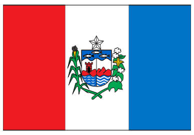
1st layer




2nd layer



# ALAGOAS - AL



CLUSTER: 3
BENCHMARK STATE: RN

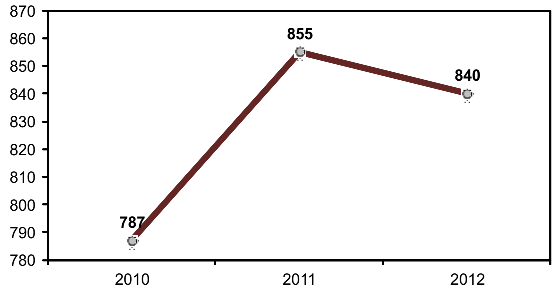
CAPITAL CITY: MACEIÓ

POPULATION: 3,165,472 INH.
VEHICLE FLEET: 510,293 VEH.
MOTORIZATION: 16.12 VEH./100 INH.
TOTAL VKT: 7.96 BILLION km
GDP *PER CAPITA*: R$ 7,874.00

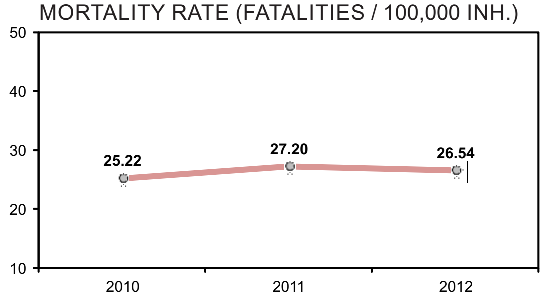
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **28% (8th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **52.9%** | 825 | 244 | 201 - 281 |
| | **36.9%** | 356 | 40 | 30 - 71 |
| | **100.0%** | 61 | n.a. | n.a. |
| | **49.6%** | 5 | 1 | 0 - 2 |
| | **66.8%** | 1 | 0 | n.a. |
| | - | 403 | 44 | n.a. - 119 |

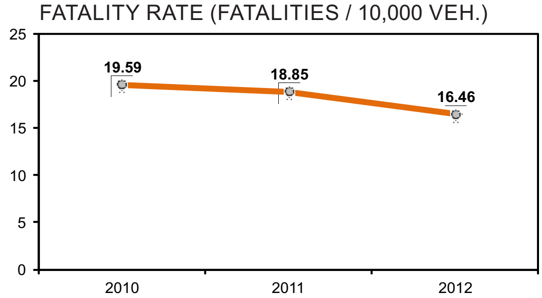
TRAFFIC FATALITY ABSOLUTE NUMBER



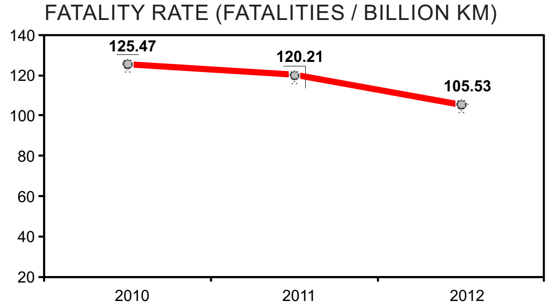

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

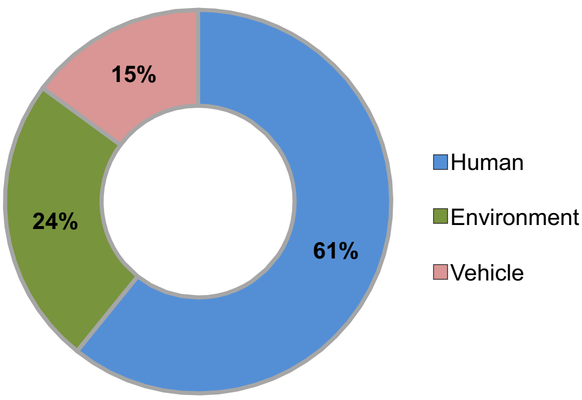



FATALITY RATE (FATALITIES / BILLION KM)

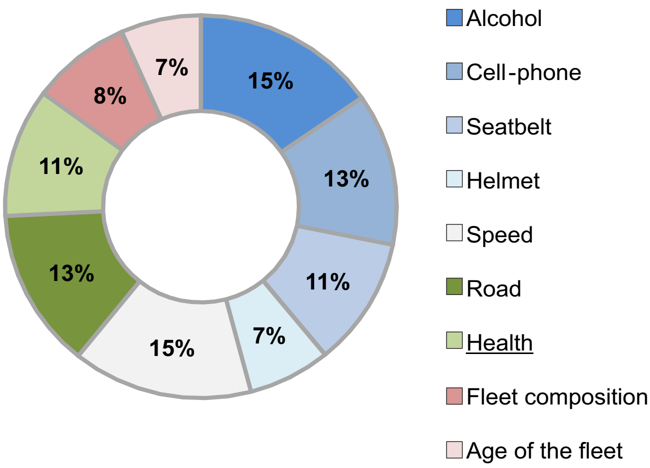



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **99% (2nd)**

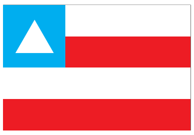
1st layer

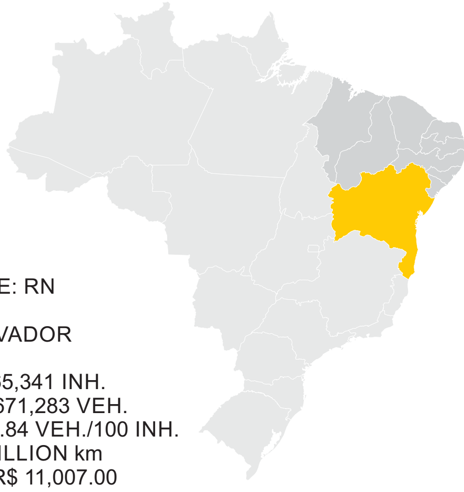



2nd layer



# BAHIA - BA





CLUSTER: 3
BENCHMARK STATE: RN

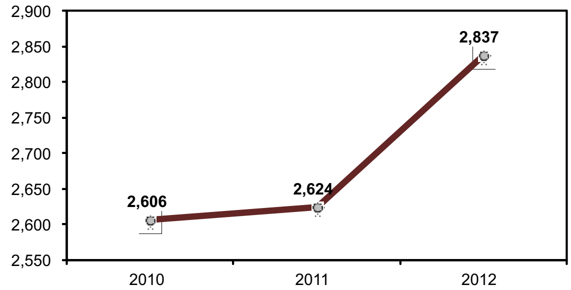
CAPITAL CITY: SALVADOR

POPULATION: 14,165,341 INH.
VEHICLE FLEET: 2,671,283 VEH.
MOTORIZATION: 18.84 VEH./100 INH.
TOTAL VKT: 43.12 BILLION km
GDP *PER CAPITA*: R$ 11,007.00

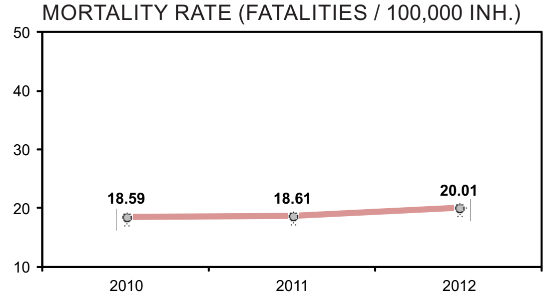
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **68% (2nd)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | 82.5% | 2,685 | 1,835 | 1,512 - 2,280 |
|  | 100.0% | 669 | n.a. | n.a. |
|  | 26.3% | 1,233 | 86 | 61 - 134 |
|  | 49.2% | 43 | 8 | 2 - 19 |
|  | 34.8% | 16 | 1 | 0 - 2 |
|  | - | 724 | 599 | n.a. |

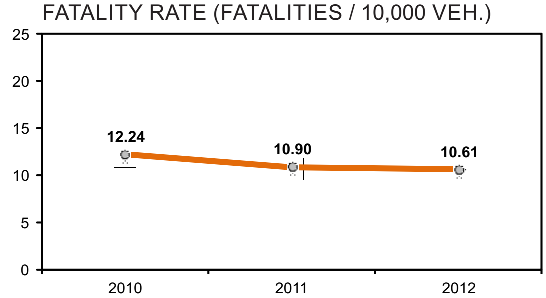
TRAFFIC FATALITY ABSOLUTE NUMBER



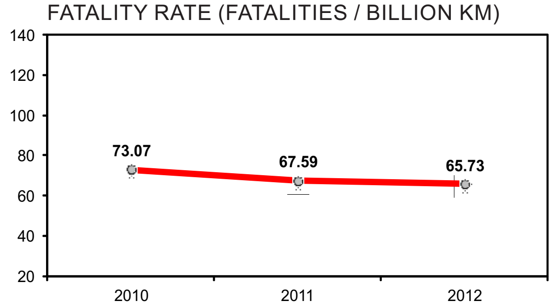

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

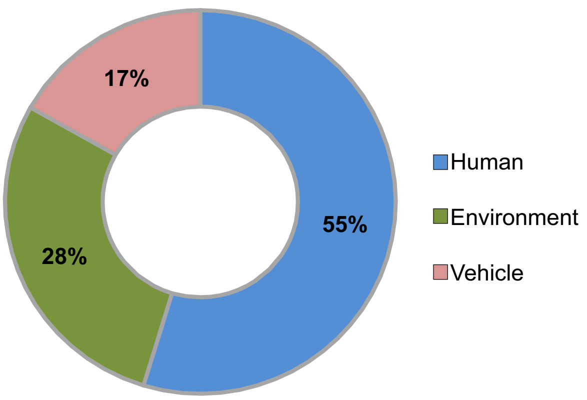



FATALITY RATE (FATALITIES / BILLION KM)

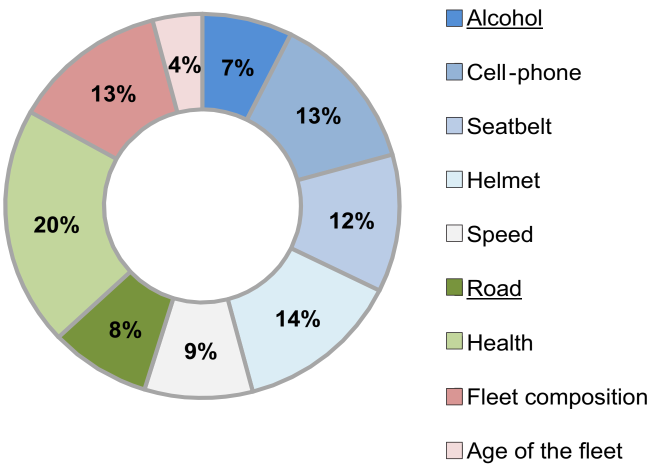



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **93% (4th)**

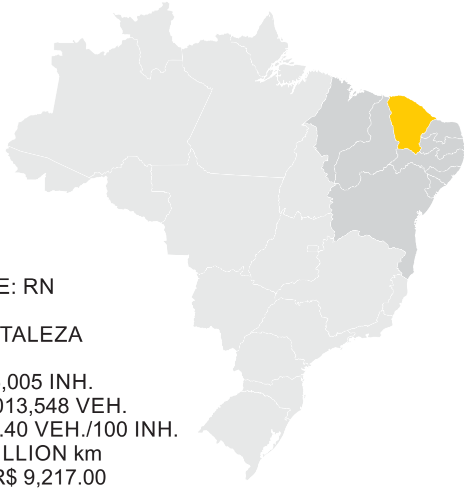
1st layer




2nd layer



# CEARÁ - CE



CLUSTER: 3
BENCHMARK STATE: RN

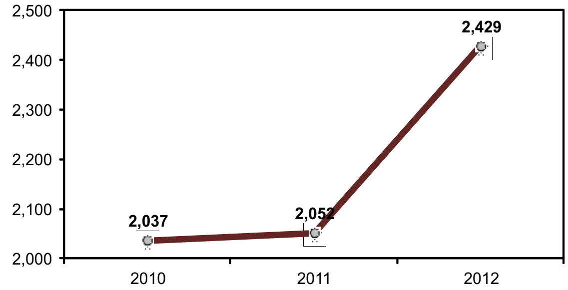
CAPITAL CITY: FORTALEZA

POPULATION: 8,606,005 INH.
VEHICLE FLEET: 2,013,548 VEH.
MOTORIZATION: 23.40 VEH./100 INH.
TOTAL VKT: 26.31 BILLION km
GDP *PER CAPITA*: R$ 9,217.00

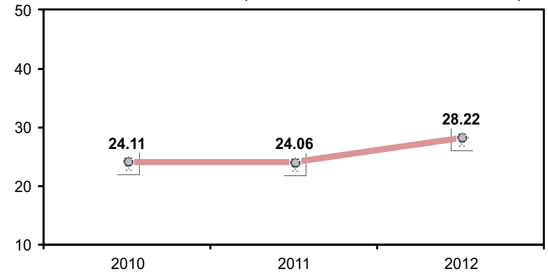
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **46% (5th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES: CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **66.4%** | 2,160 | 918 | 823 - 1,205 |
| | **51.4%** | 1,034 | 266 | 177 - 439 |
| | **42.8%** | 374 | 63 | 48 - 111 |
| | **63.0%** | 12 | 3 | 1 - n.a. |
| | **21.6%** | 8 | 0 | 0 - 1 |
| | - | 731 | 262 | 32 - 668 |

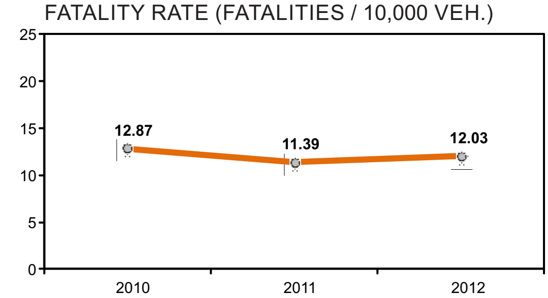
TRAFFIC FATALITY ABSOLUTE NUMBER



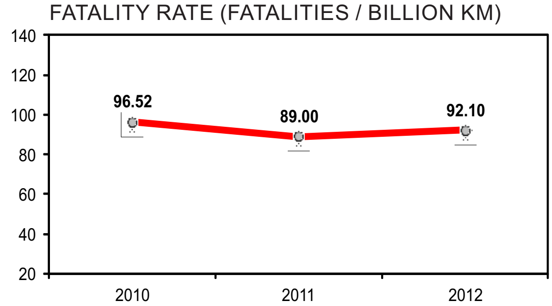

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

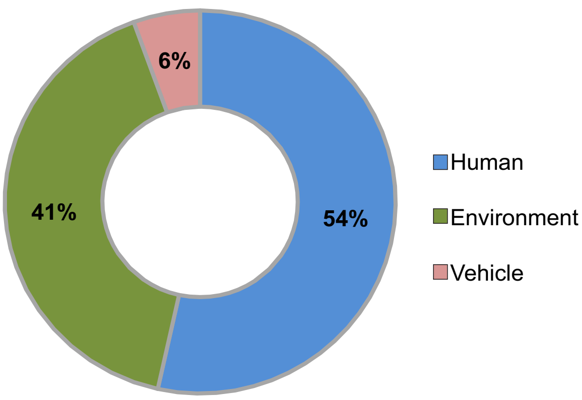



FATALITY RATE (FATALITIES / BILLION KM)

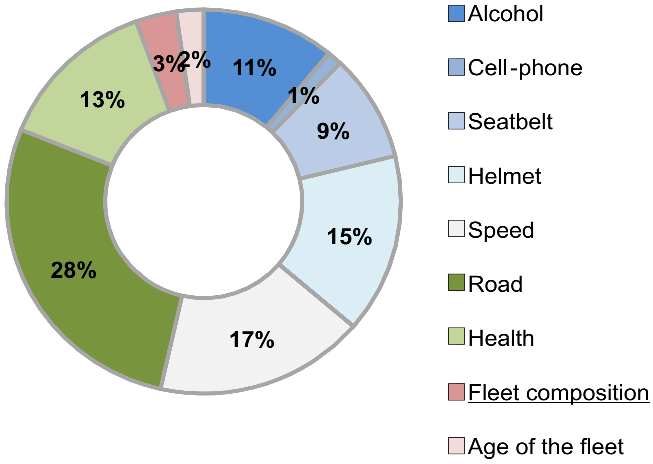



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **87% (6th)**

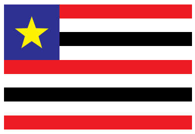
1st layer

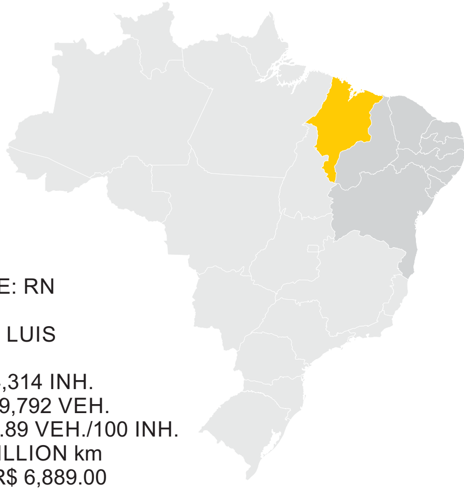



2nd layer



# MARANHÃO - MA





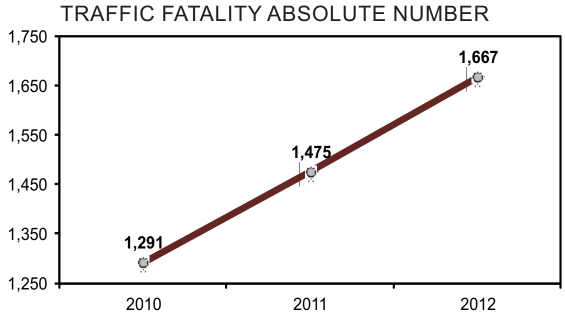
CLUSTER: 3
BENCHMARK STATE: RN

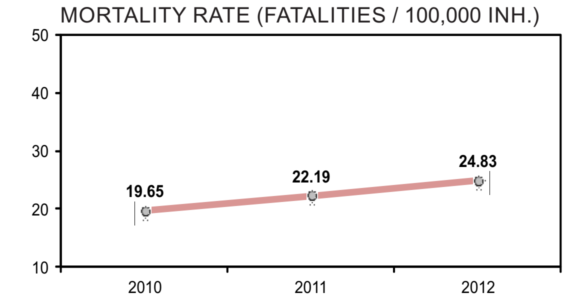
CAPITAL CITY: SÃO LUIS

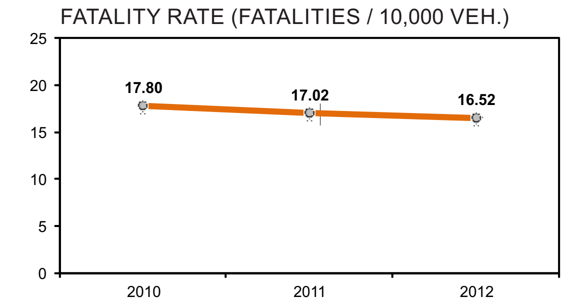
POPULATION: 6,714,314 INH.
VEHICLE FLEET: 999,792 VEH.
MOTORIZATION: 14.89 VEH./100 INH.
TOTAL VKT: 17.25 BILLION km
GDP *PER CAPITA*: R$ 6,889.00

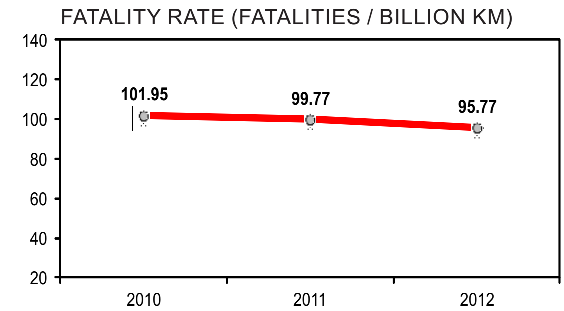
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **35% (6th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES - CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **60.7%** | 1,463 | 555 | 467 - 623 |
| Motorcycle | **50.3%** | 738 | 198 | 133 - 278 |
| Car | **29.8%** | 289 | 28 | 15 - 57 |
| Truck | **13.9%** | 34 | 0 | 0 - 2 |
| Bus | **11.0%** | 10 | 0 | n.a. |
| Pedestrian / Cyclist | - | 392 | 183 | 26 - 292 |

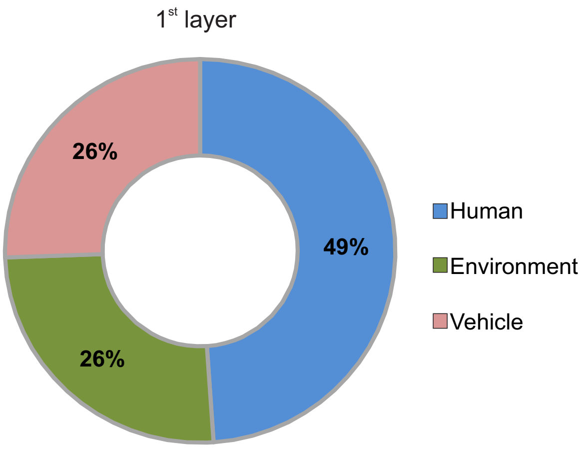


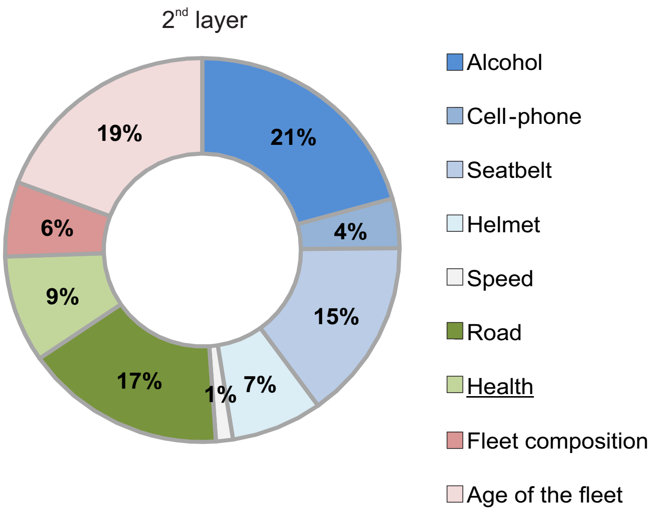




SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **67% (9th)**



# PARAÍBA - PB

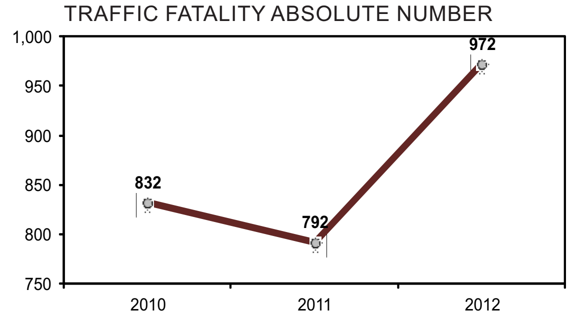

CLUSTER: 3
BENCHMARK STATE: RN

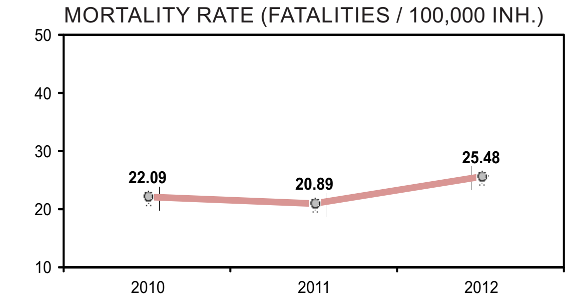
CAPITAL CITY: JOÃO PESSOA

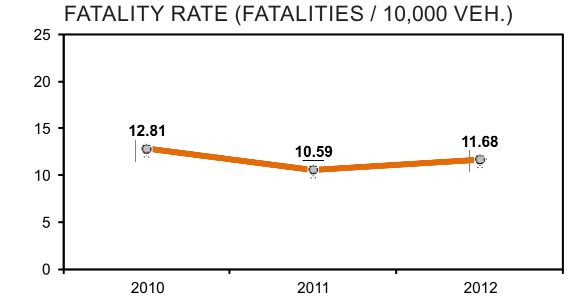
POPULATION: 3,815,171 INH.
VEHICLE FLEET: 830,385 VEH.
MOTORIZATION: 21.77 VEH./100 INH.
TOTAL VKT: 12.57 BILLION km
GDP *PER CAPITA*: R$ 8,481.00

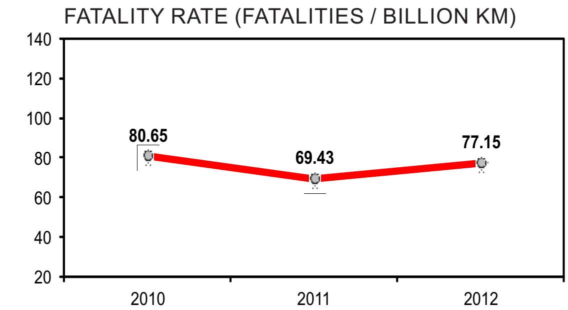
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **58% (3rd)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES: CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **74.5%** | 860 | 453 | 401 - 523 |
| Motorcycle | **50.3%** | 476 | 125 | 93 - 159 |
| Car | **51.7%** | 169 | 46 | 29 - 84 |
| Truck | **31.6%** | 11 | 1 | 0 - 3 |
| Bus | **27.3%** | 3 | 0 | n.a. |
| Pedestrian / Cyclist | - | 201 | 133 | 28 - n.a. |

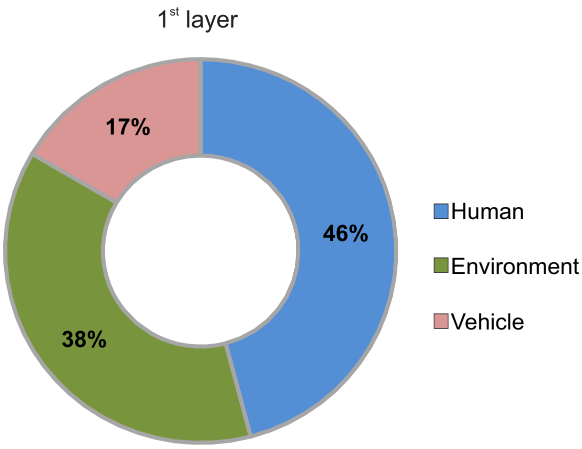


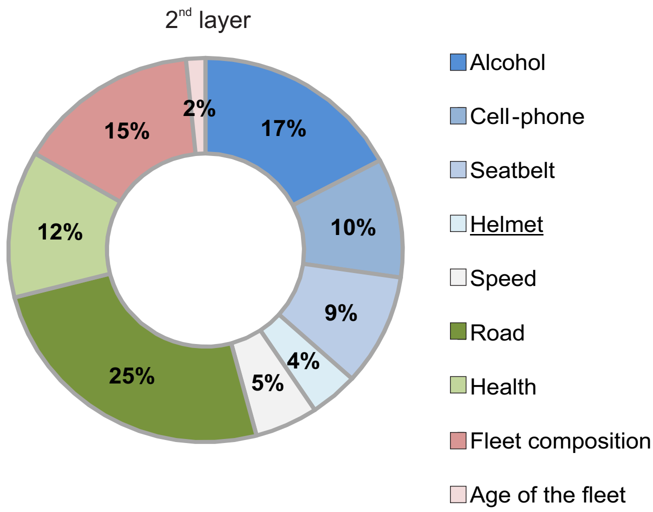



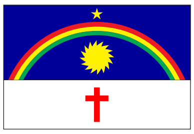



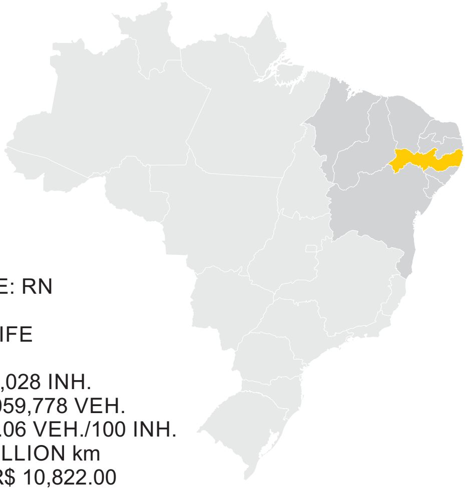

SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **94% (3rd)**



# PERNAMBUCO - PE





CLUSTER: 3
BENCHMARK STATE: RN

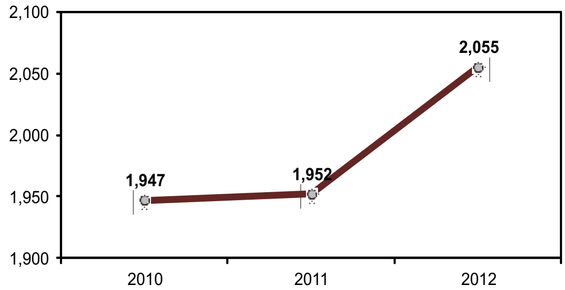
CAPITAL CITY: RECIFE

POPULATION: 8,931,028 INH.
VEHICLE FLEET: 2,059,778 VEH.
MOTORIZATION: 23.06 VEH./100 INH.
TOTAL VKT: 28.88 BILLION km
GDP *PER CAPITA*: R$ 10,822.00

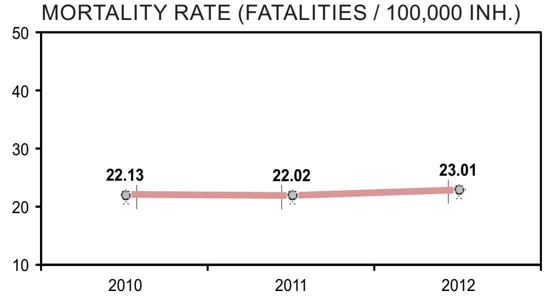
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **55% (4th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES | | |
|---|---|---|---|---|
| | | CURRENT | TARGET | RANGE |
| OVER ALL | **76.7%** | 1,979 | 1,127 | 1,004 - 1,476 |
| | **53.8%** | 904 | 261 | 197 - 341 |
| | **48.9%** | 429 | 97 | 67 - 153 |
| | **31.9%** | 35 | 1 | 0 - 3 |
| | **31.5%** | 8 | 0 | 0 - 2 |
| | - | 602 | 321 | 84 - n.a. |

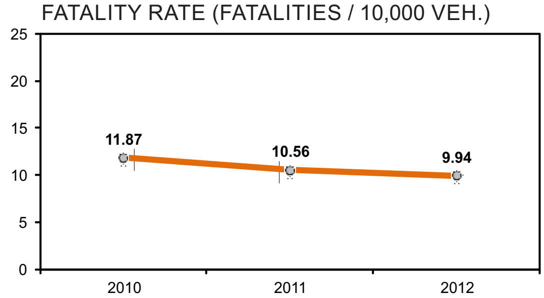
TRAFFIC FATALITY ABSOLUTE NUMBER



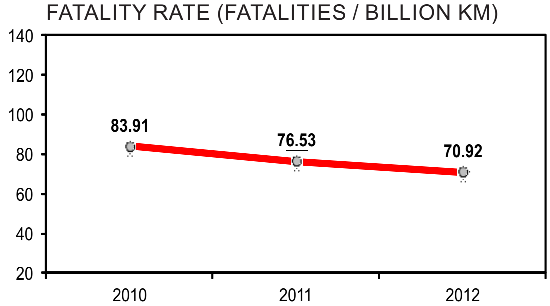

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

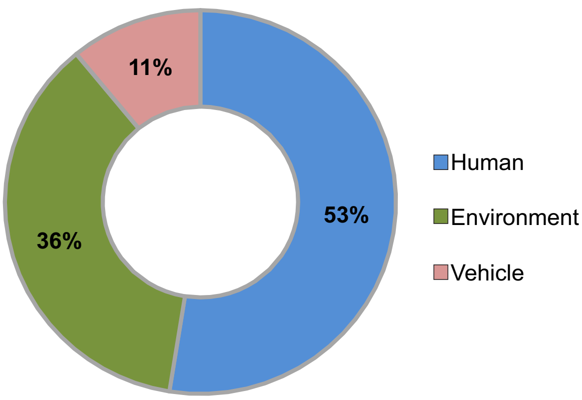



FATALITY RATE (FATALITIES / BILLION KM)

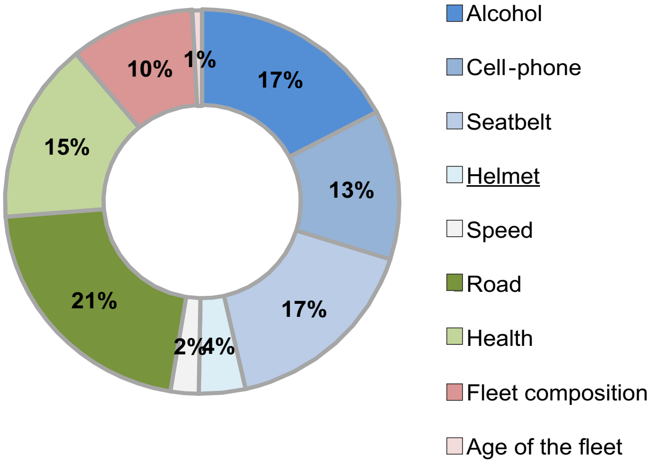



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **92% (5th)**

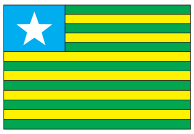
1st layer

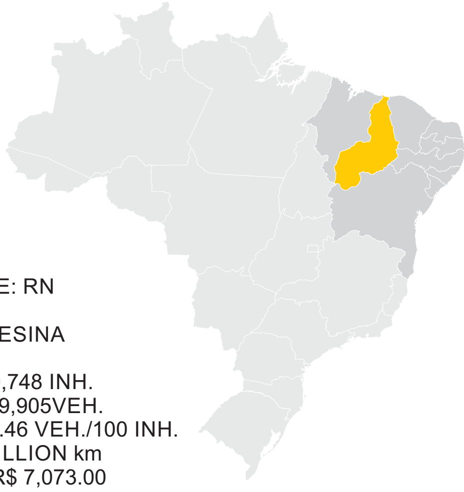



2nd layer



# PIAUÍ - PI





CLUSTER: 3
BENCHMARK STATE: RN

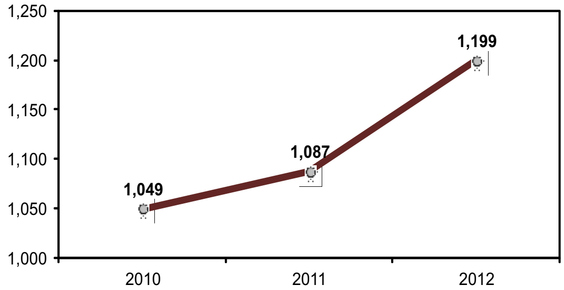
CAPITAL CITY: TERESINA

POPULATION: 3,160,748 INH.
VEHICLE FLEET: 709,905VEH.
MOTORIZATION: 22.46 VEH./100 INH.
TOTAL VKT: 10.97 BILLION km
GDP *PER CAPITA*: R$ 7,073.00

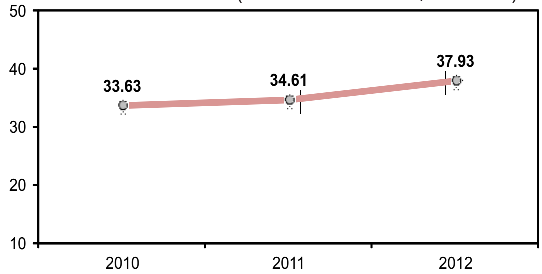
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **23% (9th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES | | |
|---|---|---|---|---|
| | | CURRENT | TARGET | RANGE |
| OVER ALL | **48.3%** | 1,108 | 250 | 221 - 296 |
| | **34.8%** | 677 | 92 | 62 - 126 |
| | **33.5%** | 166 | 18 | 12 - 31 |
| | **16.1%** | 21 | 0 | 0 - 1 |
| | **43.8%** | 2 | 0 | 0 - 1 |
| | - | 242 | 76 | 14 - 146 |

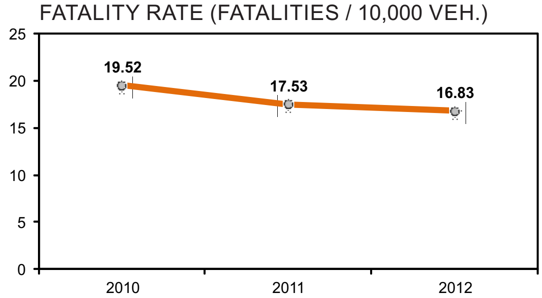
TRAFFIC FATALITY ABSOLUTE NUMBER



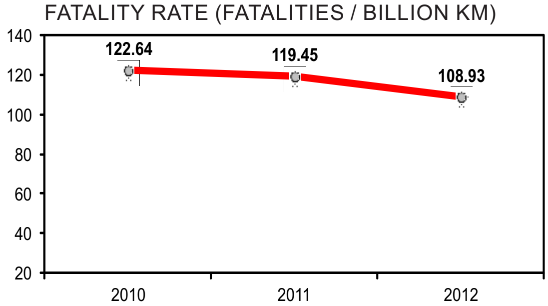

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

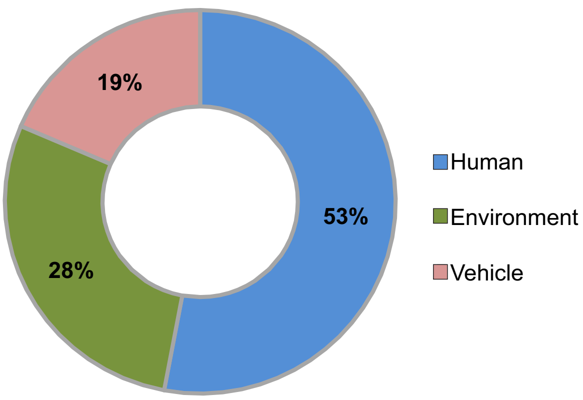



FATALITY RATE (FATALITIES / BILLION KM)

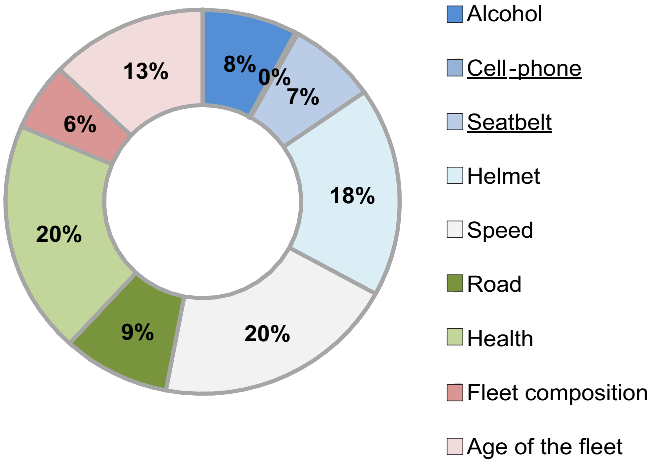



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **78% (8th)**

1st layer




2nd layer



# RIO GRANDE DO NORTE- RN

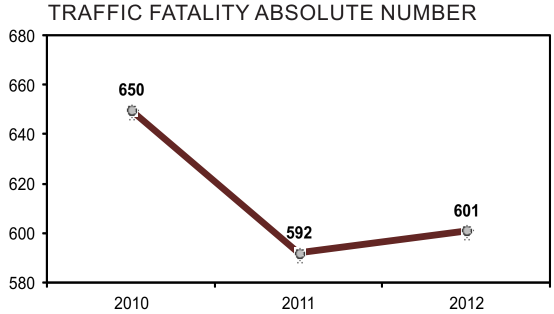
CLUSTER: 3
BENCHMARK STATE: -

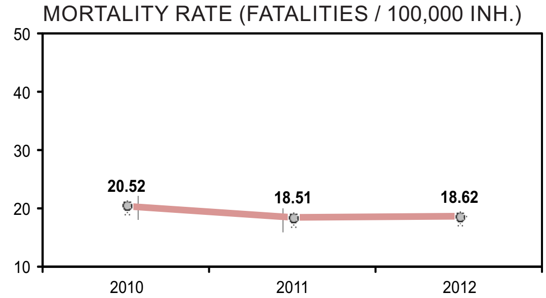
CAPITAL CITY: NATAL

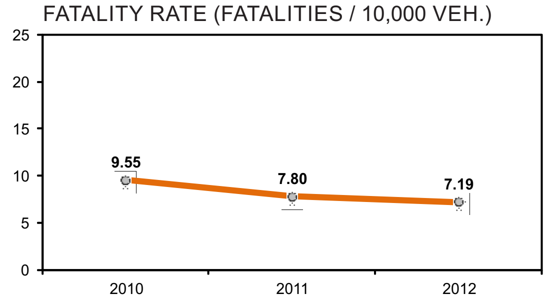
POPULATION: 3,228,198 INH.
VEHICLE FLEET: 834,229 VEH.
MOTORIZATION: 25.84 VEH./100 INH.
TOTAL VKT: 12.55 BILLION km
GDP *PER CAPITA*: R$ 10,208.00

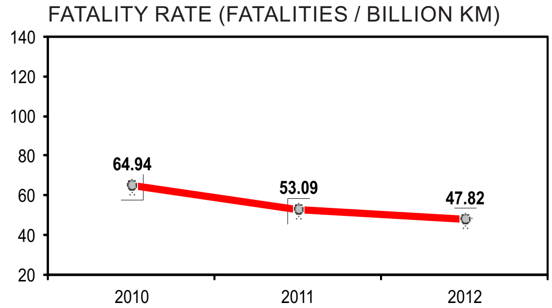
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **100% (1st)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES | | |
|---|---|---|---|---|
| | | CURRENT | TARGET | RANGE |
| OVER ALL | **100.0%** | 613 | n.a. | n.a. |
|  | **69.4%** | 319 | 169 | 116 - 226 |
|  | **58.5%** | 156 | 47 | 38 - 72 |
|  | **100.0%** | 4 | 3 | 1 - n.a. |
|  | **71.2%** | 1 | 0 | n.a. |
|  | - | 132 | 90 | n.a. |

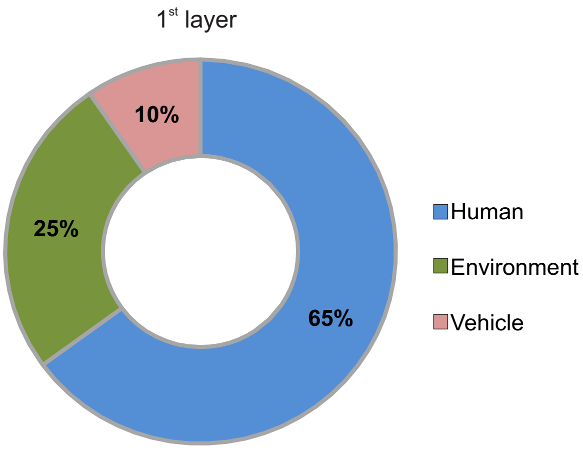


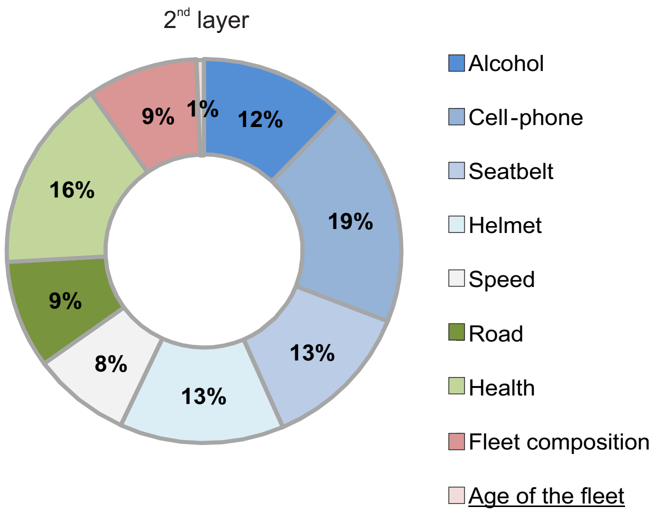



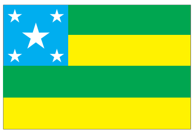




SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **100% (1st)**



# SERGIPE- SE



CLUSTER: 3
BENCHMARK STATE: -

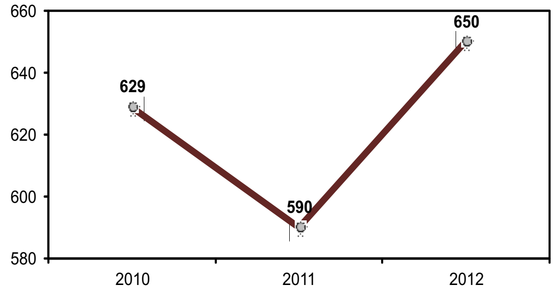
CAPITAL CITY: ARACAJÚ

POPULATION: 2,110,867 INH.
VEHICLE FLEET: 491,860 VEH.
MOTORIZATION: 23.30 VEH./100 INH.
TOTAL VKT: 7.24 BILLION km
GDP *PER CAPITA*: R$ 11,572.00

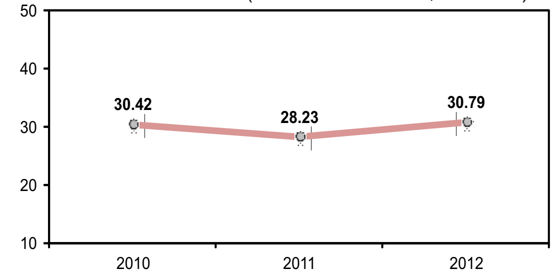
ROAD SAFETY PERFORMANCE (INDEX-SCORE): **35% (7th)**

| | SCORE | NUMBER OF TRAFFIC FATALITIES CURRENT | TARGET | RANGE |
|---|---|---|---|---|
| OVER ALL | **60.1%** | 622 | 218 | 194 - 251 |
|  | **38.2%** | 338 | 47 | 32 - 60 |
|  | **46.0%** | 120 | 25 | 18 - 38 |
|  | **11.6%** | 23 | 0 | 0 - 0 |
|  | **100.0%** | 1 | 0 | n.a. |
|  | - | 141 | 60 | 1 - 111 |

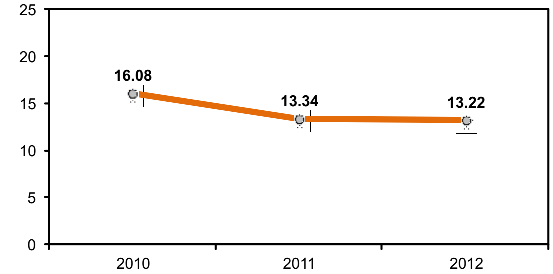
TRAFFIC FATALITY ABSOLUTE NUMBER



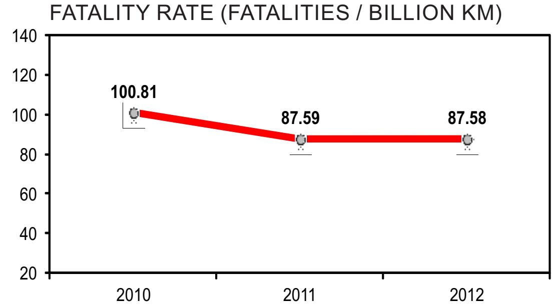

MORTALITY RATE (FATALITIES / 100,000 INH.)




FATALITY RATE (FATALITIES / 10,000 VEH.)

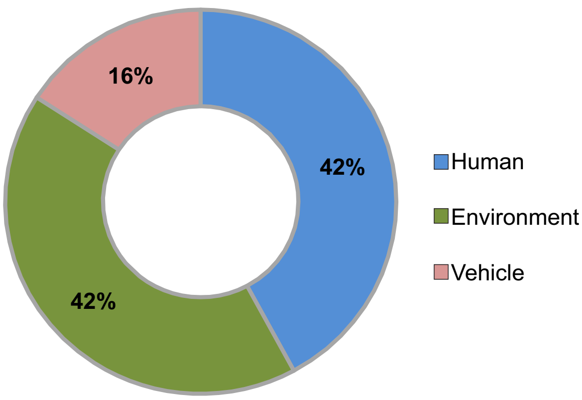



FATALITY RATE (FATALITIES / BILLION KM)

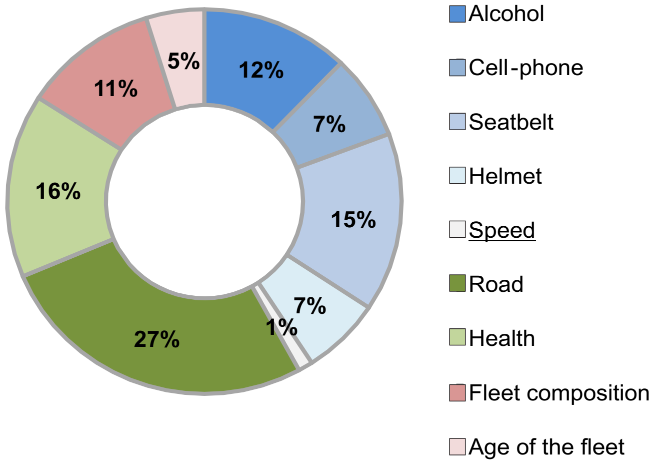



SAFETY PERFORMANCE INDICATORS
SPI INDEX-SCORE: **79% (7th)**

1st layer




2nd layer



# APPENDIX B

**Table B.1-** VKT "per state" values - average 2010-2012 (km)

| States | Total | Motorcycle | Car | Truck | Bus |
|---|---|---|---|---|---|
| **BR-27** | **837,227,334,584** | **239,779,383,333** | **440,223,555,802** | **98,689,912,095** | **58,534,483,354** |
| ES | **15,450,911,860** | 5,052,414,536 | 7,650,415,176 | 1,371,436,680 | 1,376,645,468 |
| MG | **91,679,607,806** | 27,159,024,948 | 42,657,407,900 | 14,426,051,982 | 7,437,122,976 |
| RJ | **58,262,731,833** | 10,871,340,650 | 38,476,329,376 | 3,078,241,487 | 5,836,820,320 |
| SP | **232,141,462,577** | 54,141,500,713 | 142,542,255,111 | 19,910,367,166 | 15,547,339,586 |
| PR | **56,784,897,843** | 13,303,699,104 | 31,124,553,167 | 8,763,979,120 | 3,592,666,453 |
| RS | **54,293,502,527** | 13,438,776,785 | 29,958,967,812 | 7,167,805,410 | 3,727,952,521 |
| SC | **40,490,316,243** | 11,817,988,871 | 21,845,414,048 | 5,063,209,485 | 1,763,703,839 |
| DF | **15,078,414,492** | 2,014,917,003 | 11,147,161,724 | 1,080,893,927 | 835,441,838 |
| AC | **7,098,727,439** | 2,616,481,421 | 3,471,220,269 | 458,634,199 | 552,391,550 |
| AP | **39,158,583,114** | 12,948,374,297 | 17,415,556,197 | 4,255,533,492 | 4,539,119,128 |
| AM | **23,372,656,126** | 11,378,736,883 | 9,043,110,940 | 1,645,438,269 | 1,305,370,035 |
| PA | **14,806,839,094** | 8,175,847,513 | 4,770,967,492 | 1,055,161,344 | 804,862,745 |
| RO | **11,371,758,873** | 5,004,756,177 | 5,055,690,266 | 720,431,784 | 590,880,647 |
| RR | **25,821,454,983** | 9,787,415,388 | 11,691,857,267 | 2,461,044,461 | 1,881,137,868 |
| TO | **9,522,120,026** | 5,156,026,013 | 3,019,751,282 | 768,420,381 | 577,922,350 |
| GO | **11,217,717,386** | 4,524,413,406 | 5,262,621,356 | 798,563,546 | 632,119,078 |
| MT | **6,792,353,221** | 2,605,802,732 | 3,107,093,286 | 530,082,276 | 549,374,926 |
| MS | **7,098,727,439** | 2,616,481,421 | 3,471,220,269 | 458,634,199 | 552,391,550 |
| AL | **2,731,082,775** | 1,050,993,953 | 1,033,149,222 | 551,126,362 | 95,813,237 |
| BA | **2,626,421,458** | 700,434,180 | 1,203,982,742 | 530,667,816 | 191,336,719 |
| CE | **10,447,052,306** | 2,434,764,337 | 5,494,898,842 | 1,600,324,375 | 917,064,752 |
| MA | **20,661,142,316** | 7,193,729,408 | 7,169,646,721 | 4,500,369,076 | 1,797,397,111 |
| PB | **9,463,961,033** | 3,830,256,804 | 2,861,435,571 | 2,292,836,476 | 479,432,183 |
| PE | **2,069,925,995** | 806,071,448 | 852,354,244 | 309,749,563 | 101,750,740 |
| PI | **7,587,770,834** | 2,699,512,387 | 2,529,224,989 | 1,897,306,606 | 461,726,853 |
| RN | **2,731,082,775** | 1,033,149,222 | 1,050,993,953 | 551,126,362 | 95,813,237 |
| SE | **2,626,421,458** | 1,203,982,742 | 700,434,180 | 530,667,816 | 191,336,719 |

**Table B.2-** VKT "per state" and "per vehicle" values - average 2010-2012 (km)

| States | Motorcycle | Car | Truck | Bus |
|---|---|---|---|---|
| **BR-27** | **9,361** | **13,141** | **36,283** | **75,104** |
| ES | 8,999 | 12,895 | 18,876 | 74,414 |
| MG | 8,489 | 13,384 | 45,360 | 80,034 |
| RJ | 10,010 | 14,564 | 23,594 | 81,405 |
| SP | 8,737 | 12,959 | 27,304 | 67,557 |
| PR | 8,007 | 11,653 | 29,966 | 71,855 |
| RS | 8,157 | 13,206 | 30,755 | 75,285 |
| SC | 8,802 | 13,198 | 30,183 | 69,544 |
| DF | 9,889 | 13,633 | 47,514 | 62,267 |
| AC | 13,601 | 12,390 | 90,289 | 107,745 |
| AP | 16,475 | 14,049 | 151,043 | 173,003 |
| AM | 14,942 | 13,787 | 80,796 | 91,830 |
| PA | 14,520 | 13,454 | 93,301 | 105,572 |
| RO | 12,134 | 11,072 | 80,783 | 96,937 |
| RR | 13,341 | 11,851 | 82,597 | 87,640 |
| TO | 14,228 | 12,483 | 83,607 | 94,925 |
| GO | 10,906 | 12,119 | 49,920 | 67,384 |
| MT | 11,720 | 12,281 | 64,357 | 85,470 |
| MS | 10,178 | 12,174 | 59,374 | 79,701 |
| AL | 12,355 | 15,004 | 24,911 | 55,921 |
| BA | 11,822 | 14,210 | 39,750 | 94,409 |
| CE | 9,892 | 12,374 | 27,725 | 66,834 |
| MA | 13,185 | 15,529 | 35,238 | 87,093 |
| PB | 12,161 | 14,959 | 29,407 | 71,966 |
| PE | 10,513 | 13,311 | 29,836 | 68,810 |
| PI | 11,389 | 13,840 | 38,132 | 97,937 |
| RN | 11,858 | 14,095 | 30,706 | 72,933 |
| SE | 11,940 | 14,374 | 28,876 | 73,212 |

# APPENDIX C

**Table C.1-** Outcome related data (normalized average values 2010-2012)

| State | MR | FR1 | FR2 |
|---|---|---|---|
| ES | 0.413 | 0.373 | 0.403 |
| MG | 0.602 | 0.552 | 0.635 |
| RJ | 0.752 | 0.550 | 0.623 |
| SP | 0.787 | 1.000 | 1.000 |
| PR | 0.412 | 0.512 | 0.503 |
| RS | 0.691 | 0.776 | 0.787 |
| SC | 0.447 | 0.613 | 0.648 |
| DF | 0.578 | 0.703 | 0.750 |
| AC | 0.661 | 0.349 | 0.542 |
| AP | 0.704 | 0.310 | 0.602 |
| AM | 1.000 | 0.386 | 0.665 |
| PA | 0.739 | 0.245 | 0.446 |
| RO | 0.345 | 0.320 | 0.469 |
| RR | 0.436 | 0.308 | 0.443 |
| TO | 0.356 | 0.255 | 0.433 |
| GO | 0.514 | 0.186 | 0.260 |
| MT | 0.709 | 0.302 | 0.443 |
| MS | 0.531 | 0.281 | 0.329 |
| AL | 0.608 | 0.199 | 0.307 |
| BA | 0.592 | 0.291 | 0.402 |
| CE | 0.604 | 0.315 | 0.395 |
| MA | 0.382 | 0.189 | 0.260 |
| PB | 0.704 | 0.416 | 0.551 |
| PE | 0.453 | 0.239 | 0.331 |
| PI | 0.413 | 0.373 | 0.403 |
| RN | 0.602 | 0.552 | 0.635 |
| SE | 0.752 | 0.550 | 0.623 |

MR: Traffic fatalities per capita

FR1: Traffic fatalities per vehicle

FR2: Traffic fatalities per traveled distance

**Table C.2-** Alcohol related data (normalized average values 2010-2012)

| State | A1 | A2 | A3 | A4 |
|---|---|---|---|---|
| ES | 0.545 | 0.570 | 0.730 | 0.750 |
| MG | 0.333 | 0.531 | 0.692 | 0.804 |
| RJ | 0.500 | 1.000 | 0.799 | 0.594 |
| SP | 1.000 | 0.626 | 0.803 | 0.661 |
| PR | 0.400 | 0.458 | 0.746 | 0.760 |
| RS | 0.333 | 0.720 | 0.833 | 0.842 |
| SC | 0.188 | 0.279 | 0.672 | 0.754 |
| DF | 0.353 | 0.435 | 0.670 | 0.518 |
| AC | 0.240 | 0.688 | 0.869 | 0.501 |
| AP | 0.261 | 0.494 | 0.672 | 1.000 |
| AM | 0.316 | 0.762 | 1.000 | 0.567 |
| PA | 0.375 | 0.963 | 0.884 | 0.616 |
| RO | 0.214 | 0.453 | 0.788 | 0.758 |
| RR | 0.261 | 0.438 | 0.702 | 0.547 |
| TO | 0.214 | 0.298 | 0.631 | 0.409 |
| GO | 0.231 | 0.352 | 0.624 | 0.609 |
| MT | 0.240 | 0.418 | 0.584 | 0.720 |
| MS | 0.353 | 0.391 | 0.658 | 0.661 |
| AL | 0.375 | 0.681 | 0.823 | 0.383 |
| BA | 0.462 | 0.636 | 0.761 | 0.275 |
| CE | 0.261 | 0.579 | 0.849 | 0.325 |
| MA | 0.300 | 0.550 | 0.837 | 0.418 |
| PB | 0.353 | 0.513 | 0.813 | 0.408 |
| PE | 0.286 | 0.748 | 0.743 | 0.421 |
| PI | 0.250 | 0.365 | 0.740 | 0.308 |
| RN | 0.231 | 0.513 | 0.833 | 0.311 |
| SE | 0.176 | 0.414 | 0.698 | 0.322 |

A1: Share of people who drink and drive in the last 30 days – Source

A2: Share of people who drink and drive

A3: Share of 9$^{th}$ year students who declared to be passenger in a vehicle with a driver who drink

A4: Share of people involved in accident who declared to drink before

**Table C.3-** Cell-phone and speeding related data (normalized average values 2010-2012)

| State | CPPC | CPPV | S1 | S2 |
|---|---|---|---|---|
| ES | 0.422 | 0.587 | 0.257 | 0.358 |
| MG | - | - | - | - |
| RJ | 0.216 | 0.377 | 0.217 | 0.379 |
| SP | 1.000 | 1.000 | 1.000 | 1.000 |
| PR | 0.502 | 0.510 | 0.167 | 0.170 |
| RS | 0.394 | 0.441 | 0.185 | 0.208 |
| SC | 0.308 | 0.284 | 0.298 | 0.274 |
| DF | - | - | - | - |
| AC | - | - | - | - |
| AP | 0.057 | 0.163 | - | - |
| AM | - | | - | - |
| PA | 0.209 | 0.791 | 0.032 | 0.120 |
| RO | 0.236 | 0.320 | 0.009 | 0.013 |
| RR | 0.101 | 0.179 | - | - |
| TO | 0.176 | 0.307 | 0.043 | 0.076 |
| GO | 0.114 | 0.142 | - | - |
| MT | 0.176 | 0.228 | - | - |
| MS | - | - | 0.070 | 0.089 |
| AL | - | - | - | - |
| BA | - | - | - | - |
| CE | 0.024 | 0.057 | - | - |
| MA | 0.068 | 0.264 | 0.027 | 0.106 |
| PB | - | | - | - |
| PE | 0.219 | 0.526 | 0.065 | 0.156 |
| PI | 0.004 | 0.010 | - | - |
| RN | - | - | - | - |
| SE | 0.137 | 0.324 | 0.035 | 0.082 |

CPPC: Cell phone related infractions per capita

CPPV: Cell phone related infractions per vehicle

S1: Speeding related infractions per capita

S2: Speeding related infractions per vehicle

**Table C.4-** Protective systems (helmet and seatbelt) related data (normalized average values 2010-2012)

| State | H1 | H2 | H3 | SB1 | SB2 | SB3 |
|---|---|---|---|---|---|---|
| ES | 0.989 | 0.910 | 0.643 | 0.997 | 0.431 | 0.738 |
| MG | 0.991 | 0.972 | 0.883 | 0.992 | 0.402 | 0.549 |
| RJ | 0.936 | 0.683 | 0.503 | 0.995 | 1.000 | 0.358 |
| SP | 0.992 | 0.798 | 0.859 | 0.998 | 0.439 | 0.569 |
| PR | 0.990 | 0.918 | 0.865 | 0.993 | 0.388 | 0.717 |
| RS | 0.994 | 0.959 | 0.790 | 0.945 | 0.407 | 0.485 |
| SC | 0.993 | 0.963 | 0.956 | 0.997 | 0.497 | 0.744 |
| DF | 0.990 | 0.918 | 0.879 | 0.991 | 0.389 | 0.668 |
| AC | 1.000 | 1.000 | 0.936 | 0.994 | 0.622 | 0.482 |
| AP | 0.984 | 0.975 | 0.806 | 0.961 | 0.917 | 0.578 |
| AM | 0.870 | 0.860 | 0.830 | 0.967 | 0.945 | 0.000 |
| PA | 0.912 | 0.814 | 0.499 | 0.996 | 0.542 | 0.397 |
| RO | 0.989 | 0.980 | 0.729 | 0.988 | 0.577 | 0.492 |
| RR | 0.974 | 0.976 | 0.971 | 0.950 | 0.545 | 1.000 |
| TO | 0.984 | 0.983 | 0.876 | 0.964 | 0.746 | 0.614 |
| GO | 0.993 | 0.976 | 0.850 | 0.990 | 0.370 | 0.598 |
| MT | 0.992 | 0.982 | 0.990 | 0.979 | 0.436 | 0.562 |
| MS | 0.996 | 0.999 | 1.000 | 0.991 | 0.481 | 0.859 |
| AL | 0.884 | 0.918 | 0.667 | 0.998 | 0.485 | 0.599 |
| BA | 0.943 | 0.822 | 0.733 | 0.986 | 0.657 | 0.428 |
| CE | 0.952 | 0.891 | 0.577 | 0.993 | 0.662 | 0.425 |
| MA | 0.910 | 0.939 | 0.788 | 0.985 | 0.671 | 0.215 |
| PB | 0.904 | 0.902 | 0.670 | 0.999 | 0.454 | 0.283 |
| PE | 0.969 | 0.892 | 0.776 | 0.996 | 0.419 | 0.195 |
| PI | 0.806 | 0.857 | 0.509 | 0.993 | 0.877 | 0.361 |
| RN | 0.927 | 0.955 | 0.849 | 1.000 | 0.442 | 0.807 |
| SE | 0.844 | 0.944 | 0.638 | 0.943 | 0.425 | 0.388 |

SB1: Share of people involved in accident wearing seatbelt on federal highways

SB2: Share of 9th year students who declared not to wear seatbelt in a car in the last 30 days (of those who were passenger in a car in this period)

SB3: Share of people involved in car accidents wearing seatbelt

H1: Share of people involved in accident wearing helmet on federal highways

H2: Share of 9th year students who declared to wear helmet on a motorcycle in the last 30 days (of those who were passenger on a motorcycle in this period)

H3: Share of people involved in motorcycle accidents wearing helmet

**Table C.5-** Signing related data (normalized average values 2010-2012)

| State | PCRM | VCRM | PLRM | VLRM | PVS | VVS | LVS |
|---|---|---|---|---|---|---|---|
| ES | 0.920 | 0.915 | 0.883 | 0.905 | 0.763 | 0.797 | 0.578 |
| MG | 0.955 | 0.906 | 0.895 | 0.892 | 0.771 | 0.709 | 0.729 |
| RJ | 0.937 | 0.947 | 0.924 | 0.906 | 0.888 | 0.878 | 0.684 |
| SP | 1.000 | 0.986 | 1.000 | 1.000 | 0.996 | 0.970 | 0.910 |
| PR | 0.965 | 0.772 | 0.951 | 0.775 | 0.808 | 0.950 | 0.722 |
| RS | 0.970 | 0.867 | 0.856 | 0.839 | 0.940 | 0.908 | 0.814 |
| SC | 0.936 | 0.668 | 0.848 | 0.619 | 1.000 | 0.938 | 0.639 |
| DF | 0.966 | 0.681 | 0.940 | 0.695 | 0.786 | 0.854 | 0.745 |
| AC | 0.918 | 0.318 | 0.888 | 0.344 | 0.870 | 1.000 | 0.927 |
| AP | 0.939 | 0.161 | 0.877 | 0.164 | 0.444 | 0.682 | 0.481 |
| AM | 0.801 | 0.389 | 0.675 | 0.443 | 0.452 | 0.639 | 0.528 |
| PA | 0.924 | 0.130 | 0.937 | 0.119 | 0.300 | 0.298 | 0.186 |
| RO | 0.932 | 0.312 | 0.930 | 0.329 | 0.829 | 0.965 | 1.000 |
| RR | 0.678 | 0.464 | 0.486 | 0.480 | 0.265 | 0.465 | 0.331 |
| TO | 0.832 | 0.662 | 0.716 | 0.602 | 0.398 | 0.620 | 0.560 |
| GO | 0.897 | 0.669 | 0.743 | 0.633 | 0.421 | 0.779 | 0.641 |
| MT | 0.819 | 0.531 | 0.624 | 0.521 | 0.593 | 0.855 | 0.906 |
| MS | 0.932 | 0.585 | 0.842 | 0.601 | 0.776 | 0.785 | 0.487 |
| AL | 0.982 | 1.000 | 0.843 | 0.942 | 0.655 | 0.757 | 0.780 |
| BA | 0.866 | 0.927 | 0.736 | 0.952 | 0.618 | 0.758 | 0.858 |
| CE | 0.893 | 0.837 | 0.822 | 0.828 | 0.506 | 0.605 | 0.704 |
| MA | 0.800 | 0.836 | 0.738 | 0.830 | 0.430 | 0.854 | 0.710 |
| PB | 0.892 | 0.882 | 0.828 | 0.927 | 0.636 | 0.878 | 0.904 |
| PE | 0.912 | 0.733 | 0.786 | 0.795 | 0.576 | 0.805 | 0.587 |
| PI | 0.791 | 0.917 | 0.780 | 0.923 | 0.469 | 0.840 | 0.702 |
| RN | 0.858 | 0.951 | 0.715 | 0.978 | 0.644 | 0.838 | 0.744 |
| SE | 0.778 | 0.807 | 0.688 | 0.877 | 0.539 | 0.795 | 0.559 |

PCRM: Share of highway length with central road markings

VCRM: Share of highway length with clearly visible central road markings

PLRM: Share of highway length with lateral road markings

VLRM: Share of highway length with clearly visible lateral road markings

PVS: Share of highway length with vertical signs

VVS: Share of highway length with clearly visible vertical signs

LVS: Share of highway length with clearly legible vertical signs

**Table C.6-** Central division related data (normalized average values 2010-2012)

| State | SMH | AB | PCS |
|---|---|---|---|
| ES | 0.204 | 0.423 | 0.611 |
| MG | 0.272 | 0.177 | 0.486 |
| RJ | 0.544 | 1.000 | 0.922 |
| SP | 0.808 | 0.690 | 1.000 |
| PR | 0.266 | 0.114 | 0.874 |
| RS | 0.219 | 0.179 | 0.754 |
| SC | 0.262 | 0.108 | 0.793 |
| DF | 1.000 | 0.073 | 0.764 |
| AC | 0.119 | 0.117 | 0.016 |
| AP | 0.000 | 0.048 | 0.747 |
| AM | 0.100 | 0.020 | 0.000 |
| PA | 0.080 | 0.077 | 0.212 |
| RO | 0.110 | 0.711 | 0.079 |
| RR | 0.000 | 0.014 | 0.142 |
| TO | 0.048 | 0.050 | 0.333 |
| GO | 0.250 | 0.088 | 0.617 |
| MT | 0.079 | 0.433 | 0.179 |
| MS | 0.130 | 0.102 | 0.511 |
| AL | 0.047 | 0.082 | 0.774 |
| BA | 0.063 | 0.162 | 0.478 |
| CE | 0.119 | 0.621 | 0.388 |
| MA | 0.026 | 0.072 | 0.348 |
| PB | 0.359 | 0.200 | 0.355 |
| PE | 0.279 | 0.215 | 0.497 |
| PI | 0.026 | 0.186 | 0.571 |
| RN | 0.162 | 0.220 | 0.338 |
| SE | 0.344 | 0.014 | 0.639 |

SMH: Share of multilane highways in relation to the paved highways

AB: Share of highway length equipped with adequate barriers

PCS: Share of highway length with adequate shoulders

**Table C.7-** Health System related data (normalized average values 2010-2012)

| State | DPC | HPPC | HEPC | HBPC |
|---|---|---|---|---|
| ES | 0.534 | 0.407 | 0.807 | 0.722 |
| MG | 0.502 | 0.545 | 0.706 | 0.746 |
| RJ | 0.972 | 1.000 | 0.822 | 1.000 |
| SP | 0.689 | 0.701 | 0.896 | 0.770 |
| PR | 0.544 | 0.416 | 0.690 | 0.911 |
| RS | 0.652 | 0.846 | 0.754 | 0.944 |
| SC | 0.462 | 0.490 | 0.752 | 0.809 |
| DF | 1.000 | 0.847 | 0.826 | 0.869 |
| AC | 0.255 | 0.528 | 0.999 | 0.669 |
| AP | 0.207 | 0.525 | 0.819 | 0.527 |
| AM | 0.294 | 0.500 | 0.817 | 0.563 |
| PA | 0.213 | 0.400 | 0.463 | 0.666 |
| RO | 0.284 | 0.563 | 0.762 | 0.843 |
| RR | 0.341 | 0.546 | 0.987 | 0.593 |
| TO | 0.273 | 0.558 | 1.000 | 0.578 |
| GO | 0.388 | 0.376 | 0.593 | 0.963 |
| MT | 0.315 | 0.435 | 0.723 | 0.720 |
| MS | 0.404 | 0.607 | 0.928 | 0.773 |
| AL | 0.322 | 0.304 | 0.576 | 0.661 |
| BA | 0.310 | 0.394 | 0.562 | 0.710 |
| CE | 0.293 | 0.330 | 0.605 | 0.738 |
| MA | 0.147 | 0.276 | 0.510 | 0.724 |
| PB | 0.328 | 0.405 | 0.630 | 0.830 |
| PE | 0.378 | 0.377 | 0.653 | 0.798 |
| PI | 0.257 | 0.343 | 0.601 | 0.864 |
| RN | 0.339 | 0.313 | 0.739 | 0.784 |
| SE | 0.358 | 0.411 | 0.712 | 0.616 |

DPC: Number of medical doctors per capita

HPPC: Number of health professionals per capita

HEPC: Health expenditure per capita

HBPC: Number of hospital beds per capita

**Table C.8-** Vehicle fleet related data (normalized average values 2010-2012)

| State | SMF | STF | VTYOM | VFYOL |
|---|---|---|---|---|
| ES | 0.387 | 0.319 | 0.546 | 0.697 |
| MG | 0.412 | 0.408 | 0.454 | 0.595 |
| RJ | 0.731 | 0.645 | 0.429 | 0.546 |
| SP | 0.577 | 0.513 | 0.430 | 0.537 |
| PR | 0.525 | 0.320 | 0.431 | 0.527 |
| RS | 0.545 | 0.371 | 0.391 | 0.458 |
| SC | 0.445 | 0.369 | 0.484 | 0.576 |
| DF | 1.000 | 1.000 | 0.604 | 0.778 |
| AC | 0.224 | 0.473 | 0.758 | 0.884 |
| AP | 0.293 | 0.618 | 0.985 | 0.956 |
| AM | 0.374 | 0.498 | 0.669 | 0.784 |
| PA | 0.234 | 0.386 | 0.804 | 0.878 |
| RO | 0.199 | 0.371 | 0.665 | 0.793 |
| RR | 0.225 | 0.628 | 0.630 | 0.756 |
| TO | 0.219 | 0.317 | 0.724 | 0.792 |
| GO | 0.343 | 0.403 | 0.498 | 0.642 |
| MT | 0.252 | 0.297 | 0.615 | 0.730 |
| MS | 0.328 | 0.345 | 0.505 | 0.616 |
| AL | 0.320 | 0.449 | 0.657 | 0.803 |
| BA | 0.318 | 0.412 | 0.641 | 0.786 |
| CE | 0.239 | 0.550 | 0.592 | 0.760 |
| MA | 0.201 | 0.536 | 1.000 | 1.000 |
| PB | 0.268 | 0.547 | 0.644 | 0.814 |
| PE | 0.305 | 0.410 | 0.593 | 0.776 |
| PI | 0.202 | 0.561 | 0.767 | 0.872 |
| RN | 0.283 | 0.529 | 0.598 | 0.736 |
| SE | 0.295 | 0.437 | 0.601 | 0.767 |

SMF: Share of motorcycles in the total fleet

STF: Share of trucks in the total fleet

VTYOM: Share of 10-year vehicles or older

VFYOL: Share of 5-year vehicles or newer

**Table C.9-** Background related data (normalized average values 2010-2012)

| State | SMF | STF | VTYOM | VFYOL |
|---|---|---|---|---|
| ES | 0.400 | 0.662 | 0.149 | 0.870 |
| MG | 0.307 | 0.667 | 0.085 | 0.889 |
| RJ | 0.435 | 0.526 | 0.328 | 1.000 |
| SP | 0.517 | 0.920 | 0.261 | 0.998 |
| PR | 0.356 | 0.905 | 0.197 | 0.890 |
| RS | 0.404 | 0.822 | 0.085 | 0.887 |
| SC | 0.417 | 1.000 | 0.156 | 0.877 |
| DF | 1.000 | 0.894 | 1.000 | 0.924 |
| AC | 0.198 | 0.388 | 0.018 | 0.762 |
| AP | 0.211 | 0.323 | 0.006 | 0.934 |
| AM | 0.294 | 0.283 | 0.003 | 0.825 |
| PA | 0.175 | 0.243 | 0.008 | 0.713 |
| RO | 0.258 | 0.678 | 0.022 | 0.773 |
| RR | 0.240 | 0.518 | 0.012 | 0.795 |
| TO | 0.213 | 0.526 | 0.045 | 0.822 |
| GO | 0.278 | 0.736 | 0.058 | 0.940 |
| MT | 0.336 | 0.708 | 0.019 | 0.852 |
| MS | 0.304 | 0.728 | 0.038 | 0.891 |
| AL | 0.135 | 0.265 | 0.172 | 0.770 |
| BA | 0.188 | 0.313 | 0.053 | 0.754 |
| CE | 0.158 | 0.386 | 0.113 | 0.784 |
| MA | 0.118 | 0.237 | 0.042 | 0.660 |
| PB | 0.145 | 0.358 | 0.128 | 0.783 |
| PE | 0.185 | 0.383 | 0.136 | 0.837 |
| PI | 0.121 | 0.364 | 0.050 | 0.686 |
| RN | 0.175 | 0.435 | 0.166 | 0.811 |
| SE | 0.198 | 0.388 | 0.191 | 0.764 |

GDPC: GDP per capita

MOTR: Motorized vehicles per capita

HD: Paved highways length per area

UP: Share of urban population

# APPENDIX D

**Table D.1-** Imputed values for the indicator CPPC

| State | Average | Imputation | | | | |
|---|---|---|---|---|---|---|
| | | **1** | **2** | **3** | **4** | **5** |
| AC | 0.337 | 0.402 | 0.515 | 0.004 | 0.369 | 0.397 |
| AM | 0.332 | 0.323 | 0.309 | 0.004 | 0.323 | 0.701 |
| AL | 0.401 | 0.676 | 0.449 | 0.427 | 0.219 | 0.233 |
| BA | 0.426 | 0.120 | 1.000 | 1.000 | 0.004 | 0.004 |
| PB | 0.221 | 1.000 | 0.004 | 0.004 | 0.004 | 0.093 |
| RN | 0.532 | 0.661 | 0.292 | 0.360 | 0.347 | 1.000 |
| MG | 0.384 | 0.868 | 1.000 | 0.044 | 0.004 | 0.004 |
| DF | 0.333 | 0.561 | 0.453 | 0.004 | 0.480 | 0.166 |
| MS | 0.263 | 0.242 | 0.004 | 0.363 | 0.270 | 0.437 |

**Table D.2-** Imputed values for the indicator CPPV

| State | Average | Imputation | | | | |
|---|---|---|---|---|---|---|
| | | **1** | **2** | **3** | **4** | **5** |
| AC | 0.331 | 0.547 | 0.010 | 0.437 | 0.493 | 0.167 |
| AM | 0.380 | 0.592 | 0.025 | 1.000 | 0.010 | 0.273 |
| AL | 0.219 | 0.449 | 0.477 | 0.036 | 0.121 | 0.010 |
| BA | 0.206 | 0.434 | 0.427 | 0.147 | 0.010 | 0.010 |
| PB | 0.380 | 0.010 | 0.363 | 1.000 | 0.248 | 0.281 |
| RN | 0.375 | 0.010 | 0.887 | 0.010 | 0.010 | 0.959 |
| MG | 0.395 | 0.662 | 0.274 | 0.735 | 0.098 | 0.206 |
| DF | 0.316 | 0.318 | 0.453 | 0.010 | 0.371 | 0.425 |
| MS | 0.439 | 0.412 | 0.521 | 0.423 | 0.829 | 0.010 |

**Table D.3-** Imputed values for the indicator S1

| State | Average | Imputation | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| AC | 0.713 | 0.499 | 1.000 | 0.066 | 1.000 | 1.000 |
| AM | 0.331 | 0.409 | 0.901 | 0.009 | 0.009 | 0.328 |
| RR | 0.195 | 0.009 | 0.009 | 0.318 | 0.009 | 0.628 |
| AL | 0.486 | 0.896 | 0.144 | 0.613 | 0.638 | 0.138 |
| BA | 0.273 | 0.233 | 0.760 | 0.156 | 0.208 | 0.009 |
| CE | 0.336 | 0.173 | 0.371 | 0.009 | 0.632 | 0.496 |
| PB | 0.088 | 0.091 | 0.151 | 0.009 | 0.009 | 0.180 |
| PI | 0.410 | 0.382 | 0.009 | 0.841 | 0.593 | 0.227 |
| RN | 0.080 | 0.360 | 0.009 | 0.009 | 0.009 | 0.009 |
| MG | 0.169 | 0.009 | 0.009 | 0.009 | 0.283 | 0.534 |
| DF | 0.393 | 0.736 | 0.009 | 0.453 | 0.447 | 0.321 |
| GO | 0.167 | 0.440 | 0.232 | 0.009 | 0.009 | 0.146 |
| MT | 0.161 | 0.009 | 0.755 | 0.009 | 0.024 | 0.009 |

**Table D.4-** Imputed values for the indicator S2

| State | Average | Imputation | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| AC | 0.233 | 0.143 | 0.013 | 0.267 | 0.666 | 0.075 |
| AM | 0.139 | 0.013 | 0.013 | 0.013 | 0.304 | 0.355 |
| RR | 0.251 | 0.013 | 0.360 | 0.013 | 0.013 | 0.857 |
| AL | 0.225 | 0.369 | 0.388 | 0.013 | 0.340 | 0.013 |
| BA | 0.210 | 0.013 | 0.013 | 1.000 | 0.013 | 0.013 |
| CE | 0.286 | 0.872 | 0.013 | 0.276 | 0.013 | 0.256 |
| PB | 0.254 | 0.013 | 0.095 | 0.147 | 0.013 | 1.000 |
| PI | 0.294 | 0.061 | 0.459 | 0.232 | 0.013 | 0.705 |
| RN | 0.309 | 0.170 | 0.359 | 0.763 | 0.013 | 0.238 |
| MG | 0.266 | 0.277 | 0.407 | 0.013 | 0.385 | 0.248 |
| DF | 0.125 | 0.013 | 0.013 | 0.013 | 0.575 | 0.013 |
| GO | 0.369 | 0.581 | 0.101 | 0.658 | 0.154 | 0.353 |
| MT | 0.234 | 0.864 | 0.167 | 0.116 | 0.013 | 0.013 |

# APPENDIX E

**Table E.1-** Mortality and fatality rates for BR-27 - ranked values (average 2009-2011)

| State | MR (fat/$10^5$ inh) | | FR1 (fat/$10^4$ veh) | | FR2 (fat/$10^9$ km) | |
|---|---|---|---|---|---|---|
| AM | 12.90 | (1st) | 8.55 | (11th) | 46.68 | (4th) |
| PA | 16.46 | (2nd) | 12.88 | (22nd) | 71.96 | (17th) |
| BA | 16.65 | (3rd) | 10.32 | (16th) | 68.94 | (16th) |
| RJ | 16.67 | (4th) | 5.85 | (5th) | 47.32 | (5th) |
| SP | 17.27 | (5th) | 3.49 | (1st) | 32.08 | (1st) |
| RN | 18.38 | (6th) | 8.04 | (9th) | 60.17 | (10th) |
| AP | 19.23 | (7th) | 11.03 | (20th) | 56.43 | (8th) |
| AC | 19.34 | (8th) | 9.33 | (13th) | 62.81 | (11th) |
| RS | 19.52 | (9th) | 4.45 | (2nd) | 41.02 | (2nd) |
| MA | 19.94 | (10th) | 16.21 | (25th) | 104.68 | (25th) |
| PB | 21.28 | (11th) | 11.53 | (21st) | 82.02 | (21st) |
| PE | 21.46 | (12th) | 10.78 | (17th) | 85.55 | (22nd) |
| MG | 21.70 | (13th) | 6.14 | (6th) | 48.61 | (6th) |
| CE | 22.06 | (14th) | 10.96 | (18th) | 96.46 | (24th) |
| DF | 23.94 | (15th) | 5.04 | (3rd) | 42.50 | (3rd) |
| AL | 24.73 | (16th) | 17.70 | (27th) | 128.19 | (26th) |
| SE | 28.43 | (17th) | 14.03 | (24th) | 95.89 | (23rd) |
| GO | 30.38 | (18th) | 7.79 | (8th) | 57.88 | (9th) |
| SC | 30.47 | (19th) | 5.66 | (4th) | 50.43 | (7th) |
| RR | 30.61 | (20th) | 10.98 | (19th) | 74.45 | (18th) |
| ES | 30.97 | (21st) | 8.65 | (12th) | 76.79 | (20th) |
| PR | 31.44 | (22nd) | 6.53 | (7th) | 63.60 | (12th) |
| MS | 32.33 | (23rd) | 8.22 | (10th) | 63.61 | (13th) |
| PI | 32.48 | (24th) | 17.48 | (26th) | 131.41 | (27th) |
| MT | 36.35 | (25th) | 9.74 | (14th) | 66.62 | (15th) |
| RO | 36.52 | (26th) | 10.13 | (15th) | 66.08 | (14th) |
| TO | 37.12 | (27th) | 13.14 | (23rd) | 75.66 | (19th) |
| BR-27 | 21.53 | | 6.50 | | 63.27 | |

**Table E.2-** Mortality and fatality rates for EU-27 - ranked values (average 2009-2011)

| Country | MR (fat/$10^5$ inh) | | FR1 (fat/$10^4$ veh) | | FR2 (fat/$10^9$ pass.km) | |
|---|---|---|---|---|---|---|
| UK | 3.33 | ($1^{st}$) | 0.60 | ($3^{rd}$) | 30.90 | ($2^{nd}$) |
| SE | 3.37 | ($2^{nd}$) | 0.58 | ($2^{nd}$) | 30.41 | ($1^{st}$) |
| NL | 3.47 | ($3^{rd}$) | 0.55 | ($1^{st}$) | 39.46 | ($3^{rd}$) |
| MT | 4.59 | ($4^{th}$) | 0.63 | ($4^{th}$) | 83.63 | ($14^{th}$) |
| DK | 4.68 | ($5^{th}$) | 0.91 | ($10^{th}$) | 67.27 | ($10^{th}$) |
| IE | 4.72 | ($6^{th}$) | 0.93 | ($11^{th}$) | 45.02 | ($6^{th}$) |
| DE | 4.81 | ($7^{th}$) | 0.77 | ($7^{th}$) | 42.59 | ($4^{th}$) |
| FI | 5.25 | ($8^{th}$) | 0.73 | ($5^{th}$) | 44.05 | ($5^{th}$) |
| ES | 5.26 | ($9^{th}$) | 0.74 | ($6^{th}$) | 68.97 | ($11^{th}$) |
| FR | 6.49 | ($10^{th}$) | 1.01 | ($12^{th}$) | 90.31 | ($15^{th}$) |
| SK | 6.62 | ($11^{th}$) | 1.78 | ($22^{nd}$) | 131.69 | ($22^{nd}$) |
| IT | 6.73 | ($12^{th}$) | 0.81 | ($8^{th}$) | 54.60 | ($7^{th}$) |
| AT | 6.80 | ($13^{th}$) | 1.02 | ($13^{th}$) | 76.15 | ($12^{th}$) |
| EE | 6.89 | ($14^{th}$) | 1.39 | ($19^{th}$) | 92.27 | ($16^{th}$) |
| HU | 7.33 | ($15^{th}$) | 2.03 | ($24^{th}$) | 150.05 | ($25^{th}$) |
| SI | 7.34 | ($16^{th}$) | 1.21 | ($15^{th}$) | 58.52 | ($9^{th}$) |
| LU | 7.5 | ($17^{th}$) | 0.90 | ($9^{th}$) | 55.13 | ($8^{th}$) |
| CZ | 7.87 | ($18^{th}$) | 1.36 | ($17^{th}$) | 94.05 | ($17^{th}$) |
| BE | 8.02 | ($19^{th}$) | 1.35 | ($16^{th}$) | 77.45 | ($13^{th}$) |
| CY | 8.23 | ($20^{th}$) | 1.07 | ($14^{th}$) | 108.95 | ($21^{st}$) |
| PT | 8.38 | ($21^{st}$) | 1.37 | ($18^{th}$) | 103.33 | ($19^{th}$) |
| LV | 9.89 | ($22^{nd}$) | 2.54 | ($26^{th}$) | 142.66 | ($24^{th}$) |
| LT | 9.93 | ($23^{rd}$) | 1.69 | ($21^{st}$) | 98.67 | ($18^{th}$) |
| BG | 10.36 | ($24^{th}$) | 2.52 | ($25^{th}$) | 156.55 | ($26^{th}$) |
| PL | 11.03 | ($25^{th}$) | 1.89 | ($23^{rd}$) | 139.28 | ($23^{rd}$) |
| RO | 11.17 | ($26^{th}$) | 4.70 | ($27^{th}$) | 306.65 | ($27^{th}$) |
| EL | 11.38 | ($27^{th}$) | 1.60 | ($20^{th}$) | 105.04 | ($20^{th}$) |
| EU-27 | 5.36 | | 0.87 | | 59.66 | |

# APPENDIX F

**Figure F.1-** Example of ML-DEA CI model script applied in Chapter 6

```
MODEL:
        ! Multiple Layer Data Envelopment Analysis based Composite Indicator;
        SETS:
        DMU/ES MG RJ SP PR RS SC DF/: !The decision making units;
      SCORE, W_1, W_2;! Each decision making unit has a score to be computed;
        FACTOR/MR FR1 FR2/; ! There is a set of factors;
        DXF(DMU, FACTOR): W, F; ! F1(I, J) = Jth factor of DMU I;
        ENDSETS
        DATA:
        WGTMIN=0.000001; !Min weight applied to every domain;
        WGTMIN1=0.1; !Min weight applied to every factor;
        BIGM=999999; !Biggest a weight can be;
        a=0.2;
        ENDDATA
        !------------------------------------------------------------;
        ! The Model;
        ! Try to make everyone's score as high as possible;
        Min = @SUM(DMU: SCORE);
        ! The LP for each DMU to get its score;
        @FOR(DMU(I):
                [CI] SCORE(I) = @SUM(FACTOR(J): F(I, J)* W(I, J));
                @FOR(DMU(K):
                        [LE] @SUM(FACTOR(J): F(K, J)* W(I, J))>=1;
                    [LE1]F(K,2)*W(I,2)+F(K,3)*W(I,3)>=1.50000*(F(K,1)*W(I,1));
                    [LE2]F(K,2)*W(I,2)+F(K,3)*W(I,3)<=9.00000*(F(K,1)*W(I,1));

);
             );
        @FOR (DXF(I, J):
        @BND (WGTMIN*WGTMIN1, W, BIGM);

        W(I, 1)=W_1(I);
        W(I, 2)+W(I, 3)=W_2(I);

        W_1(I)>=WGTMIN;
        W_2(I)>=WGTMIN;

        W(I, 2)>=1/2*(1-a)*W_2(I);
        W(I, 3)>=1/2*(1-a)*W_2(I);
        W(I, 2)<=1/2*(1+a)*W_2(I);
        W(I, 3)<=1/2*(1+a)*W_2(I);

        );
         DATA:
      F = @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_A\Cluster
1\INPUTN_1.XLS',INPUT);
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_A\Cluster
1\OUTPUT_1.XLS',SCORE)= SCORE;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_A\Cluster
1\OUTPUT_1.XLS',W_1)= W_1;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_A\Cluster
1\OUTPUT_1.XLS',W_2)= W_2;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_A\Cluster
1\OUTPUT_1.XLS',W)= W;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_A\Cluster
1\OUTPUT_1.XLS',F)= F;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_A\Cluster
1\OUTPUT_1.XLS',LE)= LE;
ENDDATA


END
```

**Figure F.2-** Example of R® script for bootstrapping applied in Chapter 6

```
#BOOTSTRAPPING PROCESS FOR CLUSTER 1
#CLUSTER 1 IS COMPOSED BY ES, MG, RJ, SP, PR, RS, SC, DF (ALWAYS IN THIS SEQUENCE)
#CAPITAL LETTERS UNDERLINE SIGNAL PLUS NUMBER AFTER THE NAME (I.E. SAMPLE_1) MEANS THAT THE
 VALUE WILL CHANGE IN EACH INTERTATION
#ONLY NUMBER OFTER THE NAME (I.E. thetaboot1) MEANS THAT IT IS AN INTERNAL PROCESS, AND THIS
 SEQUENCE WILL BE REPEATED IN EACH INTERATION

#REQUIRED PACKAGES
require(gdata)
require(WriteXLS)

#IMPORTING THE ORIGNAL INPUT DATA
INPUTN_1=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_A/Cluster
 1/INPUTN_1.XLS")
INPUTN_1=INPUTN_1[1:8,2:4]
INPUTN_1

#ORIGINAL OUTPUT SEPARATION
thetaoriginal=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_A/Cluster
 1/OUTPUT_1.XLS")
thetaoriginal=t(t(thetaoriginal[1:8,2]))
thetaoriginal
mean=mean(thetaoriginal[,1])
mean
N=8  #Number of observations in the sample = number of DMUs (respectively ES, MG, RJ, SP, PR,
 RS, SC, DF)
sigma=(1/N)*((((thetaoriginal[1,1])-mean)^2)+(((thetaoriginal[2,1])-
 mean)^2)+(((thetaoriginal[3,1])-mean)^2)+(((thetaoriginal[4,1])-
 mean)^2)+(((thetaoriginal[5,1])-mean)^2)+(((thetaoriginal[6,1])-
 mean)^2)+(((thetaoriginal[7,1])-mean)^2)+(((thetaoriginal[8,1])-mean)^2))
sigma  # Sigma deviation of the original scores

#IMPORTING PREVIOUS OUTPUT
OUTPUT_1=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_A/Cluster
 1/OUTPUT_1.XLS")
OUTPUT_1=t(t(OUTPUT_1[1:8,2]))
OUTPUT_1

#OBTAINING A SAMPLE FROM THE OUTPUT
SAMPLE_1=t(t(sample(OUTPUT_1[,1],8,replace=TRUE)))
SAMPLE_1
MEAN_1=mean(SAMPLE_1)
MEAN_1

#INSERTING PARAMETER TO THE SMOOTHING PARAMETER CALCULATION
h=0.7329972

#APPLICATION OF EQ 4.20 FROM SIMAR AND WILSON (1995)
EPSILON1_1=rnorm(1)
EPSILON1_1
RANDOMGENERATOR1_1=function(dmu,h=0.7329972,epsilon1_1=EPSILON1_1){    # h is computed by the
 application of eq 12 from Walden (2006)
theta=dmu+h*epsilon1_1
if(theta<1){     # condition present on eq 4.20 from Simar and Wilson (1995)
theta=2-dmu-h*epsilon1_1
}
theta
}
theta1=RANDOMGENERATOR1_1(dmu=SAMPLE_1[1,1])
```

```
EPSILON2_1=rnorm(1)
EPSILON2_1
RANDOMGENERATOR2_1=function(dmu,h=0.7329972,epsilon2_1=EPSILON2_1){
theta=dmu+h*epsilon2_1
if(theta<1){
theta=2-dmu-h*epsilon2_1
}
theta
}
theta2=RANDOMGENERATOR2_1(dmu=SAMPLE_1[2,1])

EPSILON3_1=rnorm(1)
EPSILON3_1
RANDOMGENERATOR3_1=function(dmu,h=0.7329972,epsilon3_1=EPSILON3_1){
theta=dmu+h*epsilon3_1
if(theta<1){
theta=2-dmu-h*epsilon3_1
}
theta
}
theta3=RANDOMGENERATOR3_1(dmu=SAMPLE_1[3,1])

EPSILON4_1=rnorm(1)
EPSILON4_1
RANDOMGENERATOR4_1=function(dmu,h=0.7329972,epsilon4_1=EPSILON4_1){
theta=dmu+h*epsilon4_1
if(theta<1){
theta=2-dmu-h*epsilon4_1
}
theta
}
theta4=RANDOMGENERATOR4_1(dmu=SAMPLE_1[4,1])

EPSILON5_1=rnorm(1)
EPSILON5_1
RANDOMGENERATOR5_1=function(dmu,h=0.7329972,epsilon5_1=EPSILON5_1){
theta=dmu+h*epsilon5_1
if(theta<1){
theta=2-dmu-h*epsilon5_1
}
theta
}
theta5=RANDOMGENERATOR5_1(dmu=SAMPLE_1[5,1])

EPSILON6_1=rnorm(1)
EPSILON6_1
RANDOMGENERATOR6_1=function(dmu,h=0.7329972,epsilon6_1=EPSILON6_1){
theta=dmu+h*epsilon6_1
if(theta<1){
theta=2-dmu-h*epsilon6_1
}
theta
}
theta6=RANDOMGENERATOR6_1(dmu=SAMPLE_1[6,1])

EPSILON7_1=rnorm(1)
EPSILON7_1
RANDOMGENERATOR7_1=function(dmu,h=0.7329972,epsilon7_1=EPSILON7_1){
theta=dmu+h*epsilon7_1
if(theta<1){
theta=2-dmu-h*epsilon7_1
}
theta
}
theta7=RANDOMGENERATOR7_1(dmu=SAMPLE_1[7,1])
```

```
EPSILON8_1=rnorm(1)
EPSILON8_1
RANDOMGENERATOR8_1=function(dmu,h=0.7329972,epsilon8_1=EPSILON8_1){
theta=dmu+h*epsilon8_1
if(theta<1){
theta=2-dmu-h*epsilon8_1
}
theta
}
theta8=RANDOMGENERATOR8_1(dmu=SAMPLE_1[8,1])

# COLLETCTING THETA FOR EACH OPTIMUM SCORE
THETASET_1=t(t(c(theta1,theta2,theta3,theta4,theta5,theta6,theta7,theta8)))
THETASET_1

#COMPUTATION OF THE SMOOTHED BOOTSTRAPPED SCORES (THETABOOT)
thetaboot1=MEAN_1+(theta1-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot2=MEAN_1+(theta2-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot3=MEAN_1+(theta3-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot4=MEAN_1+(theta4-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot5=MEAN_1+(theta5-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot6=MEAN_1+(theta6-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot7=MEAN_1+(theta7-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot8=MEAN_1+(theta8-MEAN_1)/(1+(h^2)/(sigma))^0.5
THETABOOTSET_1=t(t(c(thetaboot1,thetaboot2,thetaboot3,thetaboot4,thetaboot5,thetaboot6,thetabo
 ot7,thetaboot8)))
THETABOOTSET_1

#CALCULATING THE THETA RATIO FOR THE COMPUTATION OF THE NEW INPUT TO BE USED ON THE NEXT
 INTERATION
THETARATIO_1=thetaoriginal/THETABOOTSET_1
THETARATIO_1

#CALCULATING AND EXPORTING THE NEW INPUT MATRIX (TO BE USED ON THE NEXT INTERATION)
bootinputcol1=THETARATIO_1*(t(t(INPUTN_1[,1])))
bootinputcol2=THETARATIO_1*(t(t(INPUTN_1[,2])))
bootinputcol3=THETARATIO_1*(t(t(INPUTN_1[,3])))
INPUT_2=cbind(bootinputcol1,bootinputcol2,bootinputcol3)
INPUT_2=data.frame(INPUT_2)
INPUT_2
WriteXLS("INPUT_2",
 ExcelFileName="/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_A/Cluster 1/INPUT_2.xls")
```

# APPENDIX G

**Figure G.1-** Example of ML-DEA ES model script applied in Chapter 7

```
MODEL:
       ! Data Envelopment Analysis of Decision Maker Efficiency (ML Output oriented;
       SETS:
       DMU/ES MG RJ SP PR RS SC DF/: !The decision making units;
      SCORE, W_1, W_2;! Each decision making unit has a score to be computed;
       FACTOR/I1 I2 I3 O1/;
       ! There is a set of factors;
       DXF(DMU, FACTOR): W, F; ! F1( I. J) = Jth factor of DMU I;
       ENDSETS
       DATA:
       NINPUTS = 3; ! The first three NINPUTS factors are inputs;
       WGTMIN=0.000001; !Min weight applied to every factor;
       WGTMIN1=0.1; !Min weight applied to every factor;
       BIGM=999999; !Biggest a weight can be;
       a=0.2;
       ENDDATA
       !-----------------------------------------------------------;
       ! The Model;
       ! Try to make everyone's score as high as possible;
       Min = @SUM( DMU: SCORE);
       ! The LP for each DMU to get its score;
       @FOR(DMU(I):
              [RSS] SCORE(I) = @SUM( FACTOR(J)|J #GT# NINPUTS: F(I, J)* W(I, J));
              ! Sum of inputs(denominator) = 1;
              [BEP_IND] @SUM( FACTOR(J)| J #LE# NINPUTS: F(I, J)* W(I, J)) = 1;
              ! Using DMU I's weights, no DMU can score lower than 1,
              Note Numer/Denom >= 1 implies Numer >= Denom;
              @FOR(DMU(K):
                   [LE1] (F(K,2)*W(I,2)+F(K,3)*W(I,3))>=1.5*(F(K,1)*W(I,1));
                  [LE2] (F(K,2)*W(I,2)+F(K,3)*W(I,3))<=9.0*(F(K,1)*W(I,1));
                       [BEP_1] @SUM( FACTOR(J)| J #GT# NINPUTS: F(K, J) * W(I, J))
                      - @SUM( FACTOR(J)| J #LE# NINPUTS: F(K, J) * W(I, J)) >= 0
                      )
              );
       @FOR (DXF(I, J):
       @BND (WGTMIN*WGTMIN1, W, BIGM);

       W(I, 1)=W_1(I);
       W(I, 2)+W(I, 3)=W_2(I);

       W_1(I)>=WGTMIN;
       W_2(I)>=WGTMIN;

       W(I, 2)>=1/2*(1-a)*W_2(I);
       W(I, 3)>=1/2*(1-a)*W_2(I);
       W(I, 2)<=1/2*(1+a)*W_2(I);
       W(I, 3)<=1/2*(1+a)*W_2(I);

       );
        DATA:
      F = @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_B\Cluster
1\INPUTN_1.XLS',INPUT);
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_B\Cluster
1\OUTPUT_1.XLS',SCORE)= SCORE;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_B\Cluster
1\OUTPUT_1.XLS',W_1)= W_1;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_B\Cluster
1\OUTPUT_1.XLS',W_2)= W_2;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_B\Cluster
1\OUTPUT_1.XLS',W)= W;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_B\Cluster
1\OUTPUT_1.XLS',F)= F;

ENDDATA
       END
```

**Figure G.2-** Example of R® script for bootstrapping applied in Chapter 7

```
#BOOTSTRAPPING PROCESS FOR CLUSTER 1
#CLUSTER 1 IS COMPOSED BY ES, MG, RJ, SP, PR, RS, SC, DF (ALWAYS IN THIS SEQUENCE)
#CAPITAL LETTERS UNDERLINE SIGNAL PLUS NUMBER AFTER THE NAME (I.E. SAMPLE_1) MEANS THAT THE
 VALUE WILL CHANGE IN EACH INTERTATION
#ONLY NUMBER OFTER THE NAME (I.E. thetaboot1) MEANS THAT IT IS AN INTERNAL PROCESS, AND THIS
 SEQUENCE WILL BE REPEATED IN EACH INTERATION

#REQUIRED PACKAGES
require(gdata)
require(WriteXLS)

#IMPORTING THE ORIGNAL INPUT DATA
INPUTN_1=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_B/Cluster
 1/INPUTN_1.XLS")
INPUTN_1=INPUTN_1[1:8,2:4]
INPUTN_1

#ORIGINAL OUTPUT SEPARATION
thetaoriginal=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_B/Cluster
 1/OUTPUT_1.XLS")
thetaoriginal=t(t(thetaoriginal[1:8,2]))
thetaoriginal
mean=mean(thetaoriginal[,1])
mean
N=8  #Number of observations in the sample = number of DMUs (respectively ES, MG, RJ, SP, PR,
 RS, SC, DF)
sigma=(1/N)*((((thetaoriginal[1,1])-mean)^2)+(((thetaoriginal[2,1])-
 mean)^2)+(((thetaoriginal[3,1])-mean)^2)+(((thetaoriginal[4,1])-
 mean)^2)+(((thetaoriginal[5,1])-mean)^2)+(((thetaoriginal[6,1])-
 mean)^2)+(((thetaoriginal[7,1])-mean)^2)+(((thetaoriginal[8,1])-mean)^2))
sigma  # Sigma deviation of the original scores

#IMPORTING PREVIOUS OUTPUT
OUTPUT_1=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_B/Cluster
 1/OUTPUT_1.XLS")
OUTPUT_1=t(t(OUTPUT_1[1:8,2]))
OUTPUT_1

#OBTAINING A SAMPLE FROM THE OUTPUT
SAMPLE_1=t(t(sample(OUTPUT_1[,1],8,replace=TRUE)))
SAMPLE_1
MEAN_1=mean(SAMPLE_1)
MEAN_1

#INSERTING PARAMETER TO THE SMOOTHING PARAMETER CALCULATION
h=0.7329972

#APPLICATION OF EQ 4.20 FROM SIMAR AND WILSON (1995)
EPSILON1_1=rnorm(1)
EPSILON1_1
RANDOMGENERATOR1_1=function(dmu,h=0.7329972,epsilon1_1=EPSILON1_1){    # h is computed by the
 application of eq 12 from Walden (2006)
theta=dmu+h*epsilon1_1
if(theta<1){     # condition present on eq 4.20 from Simar and Wilson (1995)
theta=2-dmu-h*epsilon1_1
}
theta
}
theta1=RANDOMGENERATOR1_1(dmu=SAMPLE_1[1,1])
```

```
EPSILON2_1=rnorm(1)
EPSILON2_1
RANDOMGENERATOR2_1=function(dmu,h=0.7329972,epsilon2_1=EPSILON2_1){
theta=dmu+h*epsilon2_1
if(theta<1){
theta=2-dmu-h*epsilon2_1
}
theta
}
theta2=RANDOMGENERATOR2_1(dmu=SAMPLE_1[2,1])

EPSILON3_1=rnorm(1)
EPSILON3_1
RANDOMGENERATOR3_1=function(dmu,h=0.7329972,epsilon3_1=EPSILON3_1){
theta=dmu+h*epsilon3_1
if(theta<1){
theta=2-dmu-h*epsilon3_1
}
theta
}
theta3=RANDOMGENERATOR3_1(dmu=SAMPLE_1[3,1])

EPSILON4_1=rnorm(1)
EPSILON4_1
RANDOMGENERATOR4_1=function(dmu,h=0.7329972,epsilon4_1=EPSILON4_1){
theta=dmu+h*epsilon4_1
if(theta<1){
theta=2-dmu-h*epsilon4_1
}
theta
}
theta4=RANDOMGENERATOR4_1(dmu=SAMPLE_1[4,1])

EPSILON5_1=rnorm(1)
EPSILON5_1
RANDOMGENERATOR5_1=function(dmu,h=0.7329972,epsilon5_1=EPSILON5_1){
theta=dmu+h*epsilon5_1
if(theta<1){
theta=2-dmu-h*epsilon5_1
}
theta
}
theta5=RANDOMGENERATOR5_1(dmu=SAMPLE_1[5,1])

EPSILON6_1=rnorm(1)
EPSILON6_1
RANDOMGENERATOR6_1=function(dmu,h=0.7329972,epsilon6_1=EPSILON6_1){
theta=dmu+h*epsilon6_1
if(theta<1){
theta=2-dmu-h*epsilon6_1
}
theta
}
theta6=RANDOMGENERATOR6_1(dmu=SAMPLE_1[6,1])

EPSILON7_1=rnorm(1)
EPSILON7_1
RANDOMGENERATOR7_1=function(dmu,h=0.7329972,epsilon7_1=EPSILON7_1){
theta=dmu+h*epsilon7_1
if(theta<1){
theta=2-dmu-h*epsilon7_1
}
theta
}
theta7=RANDOMGENERATOR7_1(dmu=SAMPLE_1[7,1])
```

```
EPSILON8_1=rnorm(1)
EPSILON8_1
RANDOMGENERATOR8_1=function(dmu,h=0.7329972,epsilon8_1=EPSILON8_1){
theta=dmu+h*epsilon8_1
if(theta<1){
theta=2-dmu-h*epsilon8_1
}
theta
}
theta8=RANDOMGENERATOR8_1(dmu=SAMPLE_1[8,1])

# COLLETCTING THETA FOR EACH OPTIMUM SCORE
THETASET_1=t(t(c(theta1,theta2,theta3,theta4,theta5,theta6,theta7,theta8)))
THETASET_1

#COMPUTATION OF THE SMOOTHED BOOTSTRAPPED SCORES (THETABOOT)
thetaboot1=MEAN_1+(theta1-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot2=MEAN_1+(theta2-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot3=MEAN_1+(theta3-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot4=MEAN_1+(theta4-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot5=MEAN_1+(theta5-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot6=MEAN_1+(theta6-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot7=MEAN_1+(theta7-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot8=MEAN_1+(theta8-MEAN_1)/(1+(h^2)/(sigma))^0.5
THETABOOTSET_1=t(t(c(thetaboot1,thetaboot2,thetaboot3,thetaboot4,thetaboot5,thetaboot6,thetabo
 ot7,thetaboot8)))
THETABOOTSET_1

#CALCULATING THE THETA RATIO FOR THE COMPUTATION OF THE NEW INPUT TO BE USED ON THE NEXT
 INTERATION
THETARATIO_1=thetaoriginal/THETABOOTSET_1
THETARATIO_1

#CALCULATING AND EXPORTING THE NEW INPUT MATRIX (TO BE USED ON THE NEXT INTERATION)
bootinputcol1=(t(t(INPUTN_1[,1])))/THETARATIO_1
bootinputcol2=(t(t(INPUTN_1[,2])))/THETARATIO_1
bootinputcol3=(t(t(INPUTN_1[,3])))/THETARATIO_1
INPUT_2=cbind(bootinputcol1,bootinputcol2,bootinputcol3)
INPUT_2=data.frame(INPUT_2)
INPUT_2
WriteXLS("INPUT_2",
 ExcelFileName="/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_B/Cluster 1/INPUT_2.xls")
```

# APPENDIX H

**Table H.1-** Output information – traffic fatality numbers (average 2010-2012)

| States | Motorcycle occupants | Car occupants | Truck occupants | Bus occupants |
|---|---|---|---|---|
| ES | 411.595 | 447.604 | 20.522 | 0.774 |
| MG | 1,025.674 | 2,086.385 | 145.989 | 27.946 |
| RJ | 655.005 | 587.171 | 60.057 | 17.282 |
| SP | 2,219.599 | 1,816.075 | 164.080 | 46.941 |
| PR | 1,044.927 | 1,221.997 | 102.402 | 17.494 |
| RS | 564.873 | 740.809 | 76.474 | 12.198 |
| SC | 665.581 | 682.455 | 57.371 | 14.624 |
| DF | 138.470 | 224.846 | 6.386 | 2.353 |
| AC | 67.411 | 28.749 | 0.991 | 0.991 |
| AP | 33.345 | 16.673 | 0.000 | 2.382 |
| AM | 145.266 | 73.743 | 7.108 | 2.221 |
| PA | 503.240 | 164.339 | 20.448 | 8.709 |
| RO | 285.899 | 162.687 | 23.327 | 3.589 |
| RR | 75.540 | 29.818 | 0.000 | 0.398 |
| TO | 213.590 | 202.680 | 14.267 | 5.036 |
| GO | 695.943 | 635.595 | 66.188 | 17.034 |
| MT | 457.434 | 359.412 | 60.507 | 2.420 |
| MS | 300.517 | 263.043 | 40.718 | 3.964 |
| AL | 355.913 | 60.601 | 4.810 | 0.962 |
| BA | 669.112 | 1,232.620 | 43.182 | 15.819 |
| CE | 1,033.722 | 374.397 | 11.929 | 8.259 |
| MA | 738.067 | 288.694 | 34.431 | 10.153 |
| PB | 476.277 | 169.265 | 10.506 | 2.918 |
| PE | 904.301 | 429.166 | 34.872 | 8.322 |
| PI | 676.638 | 165.738 | 21.457 | 1.850 |
| RN | 319.262 | 156.390 | 3.646 | 1.215 |
| SE | 338.137 | 120.017 | 22.587 | 0.674 |

# APPENDIX I

**Figure I.1-** Example of ML-DEA ES model script applied in Chapter 7 – Motor vehicle occupants related

```
MODEL:
        ! Data Envelopment Analysis of Decision Maker Efficiency (ES Output oriented) ;
        SETS:
        DMU/ES MG RJ SP PR RS SC DF/: !The decision making units;
      SCORE;! Each decision making unit has a score to be computed;
        FACTOR/I1 I2 I3 I4 O1 O2 O3 O4/;
        ! There is a set of factors;
        DXF(DMU, FACTOR): W, F; ! F1( I. J) = Jth factor of DMU I;
        ENDSETS
        DATA:
        NINPUTS = 4; ! The first four NINPUTS factors are inputs;
        WGTMIN=0.000001; !Min weight applied to every factor;
        BIGM=999999; !Biggest a weight can be;
        ENDDATA
        !------------------------------------------------------------;
        ! The Model;
        ! Try to make everyone's score as high as possible;
        Min = @SUM( DMU: SCORE);
        ! The LP for each DMU to get its score;
        @FOR(DMU(I):
                [RSS] SCORE(I) = @SUM( FACTOR(J)|J #GT# NINPUTS: F(I, J)* W(I, J));
                ! Sum of inputs(denominator) = 1;
                [BEP_IND] @SUM( FACTOR(J)| J #LE# NINPUTS: F(I, J)* W(I, J)) = 1;
                ! Using DMU I's weights, no DMU can score lower than 1,
                Note Numer/Denom >= 1 implies Numer >= Denom;
                @FOR(DMU(K):
                    [LE1]
(F(K,1)*W(I,1))>=0.1182*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3))+(F(K,4)*W(I,4)));

                    [LE2]
(F(K,1)*W(I,1))<=0.3694*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3))+(F(K,4)*W(I,4)));

                    [LE3]
(F(K,2)*W(I,2))>=0.3967*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3))+(F(K,4)*W(I,4)));

                    [LE4]
(F(K,2)*W(I,2))<=0.7567*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3))+(F(K,4)*W(I,4)));
                    [LE5]
(F(K,3)*W(I,3))>=0.0304*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3))+(F(K,4)*W(I,4)));

                    [LE6]
(F(K,3)*W(I,3))<=0.1866*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3))+(F(K,4)*W(I,4)));
                    [LE7]
(F(K,4)*W(I,4))>=0.0343*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3))+(F(K,4)*W(I,4)));

                    [LE8]
(F(K,4)*W(I,4))<=0.1077*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3))+(F(K,4)*W(I,4)));
                    [LE9]
(F(K,5)*W(I,5))>=0.2968*((F(K,5)*W(I,5))+(F(K,6)*W(I,6))+(F(K,7)*W(I,7))+(F(K,8)*W(I,8)));

                    [LE10]
(F(K,5)*W(I,5))<=0.5738*((F(K,5)*W(I,5))+(F(K,6)*W(I,6))+(F(K,7)*W(I,7))+(F(K,8)*W(I,8)));

                    [LE11]
(F(K,6)*W(I,6))>=0.3739*((F(K,5)*W(I,5))+(F(K,6)*W(I,6))+(F(K,7)*W(I,7))+(F(K,8)*W(I,8)));

                    [LE12]
(F(K,6)*W(I,6))<=0.6621*((F(K,5)*W(I,5))+(F(K,6)*W(I,6))+(F(K,7)*W(I,7))+(F(K,8)*W(I,8)));
                    [LE13]
(F(K,7)*W(I,7))>=0.0138*((F(K,5)*W(I,5))+(F(K,6)*W(I,6))+(F(K,7)*W(I,7))+(F(K,8)*W(I,8)));
                    [LE14]
(F(K,7)*W(I,7))<=0.0630*((F(K,5)*W(I,5))+(F(K,6)*W(I,6))+(F(K,7)*W(I,7))+(F(K,8)*W(I,8)));

                    [LE15]
(F(K,8)*W(I,8))>=0.0009*((F(K,5)*W(I,5))+(F(K,6)*W(I,6))+(F(K,7)*W(I,7))+(F(K,8)*W(I,8)));
```

```
                 [LE16]
(F(K,8)*W(I,8))<=0.0156*((F(K,5)*W(I,5))+(F(K,6)*W(I,6))+(F(K,7)*W(I,7))+(F(K,8)*W(I,8)));

                 [BEP_1] @SUM( FACTOR(J)| J #GT# NINPUTS: F(K, J) * W(I, J))
                     - @SUM( FACTOR(J)| J #LE# NINPUTS: F(K, J) * W(I, J)) >= 0
                     )
              );

       @FOR (DXF(I, J):
       @BND (WGTMIN, W, BIGM);

       );
        DATA:
      F = @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_C\Cluster
1\INPUTN_1.XLS',INPUT);
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_C\Cluster
1\OUTPUT_1.XLS',SCORE)= SCORE;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_C\Cluster
1\OUTPUT_1.XLS',W)= W;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_C\Cluster
1\OUTPUT_1.XLS',F)= F;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_C\Cluster
1\OUTPUT_1.XLS',BEP_1)= BEP_1;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_C\Cluster
1\OUTPUT_1.XLS',DUALPR)= @DUAL( BEP_1);

ENDDATA
       END
```

**Figure I.2-** Example of R® script for bootstrapping applied in Chapter 7 – Motor vehicle occupants related

```
#BOOTSTRAPPING PROCESS FOR CLUSTER 1
#CLUSTER 1 IS COMPOSED BY ES, MG, RJ, SP, PR, RS, SC, DF (ALWAYS IN THIS SEQUENCE)
#CAPITAL LETTERS UNDERLINE SIGNAL PLUS NUMBER AFTER THE NAME (I.E. SAMPLE_1) MEANS THAT THE
 VALUE WILL CHANGE IN EACH INTERTATION
#ONLY NUMBER OFTER THE NAME (I.E. thetaboot1) MEANS THAT IT IS AN INTERNAL PROCESS, AND THIS
 SEQUENCE WILL BE REPEATED IN EACH INTERATION

#REQUIRED PACKAGES
require(gdata)
require(WriteXLS)

#IMPORTING THE ORIGNAL INPUT DATA
INPUTN_1=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_C/Cluster
 1/INPUTN_1.XLS")
INPUTN_1=INPUTN_1[1:8,2:5]
INPUTN_1

#ORIGINAL OUTPUT SEPARATION
thetaoriginal=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_C/Cluster
 1/OUTPUT_1.XLS")
thetaoriginal=t(t(thetaoriginal[1:8,2]))
thetaoriginal
mean=mean(thetaoriginal[,1])
mean
N=8  #Number of observations in the sample = number of DMUs (respectively ES, MG, RJ, SP, PR,
 RS, SC, DF)
sigma=(1/N)*((((thetaoriginal[1,1])-mean)^2)+(((thetaoriginal[2,1])-
 mean)^2)+(((thetaoriginal[3,1])-mean)^2)+(((thetaoriginal[4,1])-
 mean)^2)+(((thetaoriginal[5,1])-mean)^2)+(((thetaoriginal[6,1])-
 mean)^2)+(((thetaoriginal[7,1])-mean)^2)+(((thetaoriginal[8,1])-mean)^2))
sigma  # Sigma deviation of the original scores

#IMPORTING PREVIOUS OUTPUT
OUTPUT_1=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_C/Cluster
 1/OUTPUT_1.XLS")
OUTPUT_1=t(t(OUTPUT_1[1:8,2]))
OUTPUT_1

#OBTAINING A SAMPLE FROM THE OUTPUT
SAMPLE_1=t(t(sample(OUTPUT_1[,1],8,replace=TRUE)))
SAMPLE_1
MEAN_1=mean(SAMPLE_1)
MEAN_1

#INSERTING PARAMETER TO THE SMOOTHING PARAMETER CALCULATION
h=0.7790778

#APPLICATION OF EQ 4.20 FROM SIMAR AND WILSON (1995)
EPSILON1_1=rnorm(1)
EPSILON1_1
RANDOMGENERATOR1_1=function(dmu,h=0.7790778,epsilon1_1=EPSILON1_1){    # h is computed by the
 application of eq 12 from Walden (2006)
theta=dmu+h*epsilon1_1
if(theta<1){     # condition present on eq 4.20 from Simar and Wilson (1995)
theta=2-dmu-h*epsilon1_1
}
theta
}
theta1=RANDOMGENERATOR1_1(dmu=SAMPLE_1[1,1])

EPSILON2_1=rnorm(1)
EPSILON2_1
RANDOMGENERATOR2_1=function(dmu,h=0.7790778,epsilon2_1=EPSILON2_1){
```

```
theta=dmu+h*epsilon2_1
if(theta<1){
theta=2-dmu-h*epsilon2_1
}
theta
}
theta2=RANDOMGENERATOR2_1(dmu=SAMPLE_1[2,1])

EPSILON3_1=rnorm(1)
EPSILON3_1
RANDOMGENERATOR3_1=function(dmu,h=0.7790778,epsilon3_1=EPSILON3_1){
theta=dmu+h*epsilon3_1
if(theta<1){
theta=2-dmu-h*epsilon3_1
}
theta
}
theta3=RANDOMGENERATOR3_1(dmu=SAMPLE_1[3,1])

EPSILON4_1=rnorm(1)
EPSILON4_1
RANDOMGENERATOR4_1=function(dmu,h=0.7790778,epsilon4_1=EPSILON4_1){
theta=dmu+h*epsilon4_1
if(theta<1){
theta=2-dmu-h*epsilon4_1
}
theta
}
theta4=RANDOMGENERATOR4_1(dmu=SAMPLE_1[4,1])

EPSILON5_1=rnorm(1)
EPSILON5_1
RANDOMGENERATOR5_1=function(dmu,h=0.7790778,epsilon5_1=EPSILON5_1){
theta=dmu+h*epsilon5_1
if(theta<1){
theta=2-dmu-h*epsilon5_1
}
theta
}
theta5=RANDOMGENERATOR5_1(dmu=SAMPLE_1[5,1])

EPSILON6_1=rnorm(1)
EPSILON6_1
RANDOMGENERATOR6_1=function(dmu,h=0.7790778,epsilon6_1=EPSILON6_1){
theta=dmu+h*epsilon6_1
if(theta<1){
theta=2-dmu-h*epsilon6_1
}
theta
}
theta6=RANDOMGENERATOR6_1(dmu=SAMPLE_1[6,1])

EPSILON7_1=rnorm(1)
EPSILON7_1
RANDOMGENERATOR7_1=function(dmu,h=0.7790778,epsilon7_1=EPSILON7_1){
theta=dmu+h*epsilon7_1
if(theta<1){
theta=2-dmu-h*epsilon7_1
}
theta
}
theta7=RANDOMGENERATOR7_1(dmu=SAMPLE_1[7,1])
```

```
EPSILON8_1=rnorm(1)
EPSILON8_1
RANDOMGENERATOR8_1=function(dmu,h=0.7790778,epsilon8_1=EPSILON8_1){
theta=dmu+h*epsilon8_1
if(theta<1){
theta=2-dmu-h*epsilon8_1
}
theta
}
theta8=RANDOMGENERATOR8_1(dmu=SAMPLE_1[8,1])

# COLLETCTING THETA FOR EACH OPTIMUM SCORE
THETASET_1=t(t(c(theta1,theta2,theta3,theta4,theta5,theta6,theta7,theta8)))
THETASET_1

#COMPUTATION OF THE SMOOTHED BOOTSTRAPPED SCORES (THETABOOT)
thetaboot1=MEAN_1+(theta1-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot2=MEAN_1+(theta2-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot3=MEAN_1+(theta3-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot4=MEAN_1+(theta4-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot5=MEAN_1+(theta5-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot6=MEAN_1+(theta6-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot7=MEAN_1+(theta7-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot8=MEAN_1+(theta8-MEAN_1)/(1+(h^2)/(sigma))^0.5
THETABOOTSET_1=t(t(c(thetaboot1,thetaboot2,thetaboot3,thetaboot4,thetaboot5,thetaboot6,thetabo
 ot7,thetaboot8)))
THETABOOTSET_1

#CALCULATING THE THETA RATIO FOR THE COMPUTATION OF THE NEW INPUT TO BE USED ON THE NEXT
 INTERATION
THETARATIO_1=thetaoriginal/THETABOOTSET_1
THETARATIO_1

#CALCULATING AND EXPORTING THE NEW INPUT MATRIX (TO BE USED ON THE NEXT INTERATION)
bootinputcol1=(t(t(INPUTN_1[,1])))/THETARATIO_1
bootinputcol2=(t(t(INPUTN_1[,2])))/THETARATIO_1
bootinputcol3=(t(t(INPUTN_1[,3])))/THETARATIO_1
bootinputcol4=(t(t(INPUTN_1[,4])))/THETARATIO_1

INPUT_2=cbind(bootinputcol1,bootinputcol2,bootinputcol3,bootinputcol4)
INPUT_2=data.frame(INPUT_2)
INPUT_2
WriteXLS("INPUT_2",
 ExcelFileName="/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_C/Cluster 1/INPUT_2.xls")
```

# APPENDIX J

**Figure J.1-** Example of DEA ES model script applied in Chapter 8 – Disaggregated models (e.g. car)

```
MODEL:
        ! Data Envelopment Analysis of Decision Maker Efficiency (ES Output oriented) ;
        SETS:
        DMU/ES MG RJ SP PR RS SC DF/: !The decisionmaking units;
      SCORE;! Each decision making unit has a score to be computed;
        FACTOR/I1 I2 I3 O1/;
        ! There is a set of factors;
        DXF(DMU, FACTOR): W, F; ! F1( I. J) = Jth factor of DMU I;
        ENDSETS
        DATA:
        NINPUTS = 3; ! The first three NINPUTS factors are inputs;
        WGTMIN=0.000001; !Min weight applied to every factor;
        BIGM=999999; !Biggest a weight can be;
        ENDDATA
        !----------------------------------------------------------;
        ! The Model;
        ! Try to make everyone's score as high as possible;
        Min = @SUM( DMU: SCORE);
        ! The LP for each DMU to get its score;
        @FOR(DMU(I):
             [RSS] SCORE(I) = @SUM( FACTOR(J)|J #GT# NINPUTS: F(I, J)* W(I, J));
                ! Sum of inputs(denominator) = 1;
                [BEP_IND] @SUM( FACTOR(J)| J #LE# NINPUTS: F(I, J)* W(I, J)) = 1;
                ! Using DMU I's weights, no DMU can score lower than 1,
                Note Numer/Denom >= 1 implies Numer >= Denom;
                @FOR(DMU(K):
                    [LE1]
(F(K,1)*W(I,1))>=0.8160*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3)));

                    [LE2]
(F(K,1)*W(I,1))<=0.9900*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3)));

                    [LE3]
(F(K,2)*W(I,2))>=0.0100*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3)));

                    [LE4]
(F(K,2)*W(I,2))<=0.1350*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3)));
                    [LE5]
(F(K,3)*W(I,3))>=0.0100*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3)));

                    [LE6]
(F(K,3)*W(I,3))<=0.0716*((F(K,1)*W(I,1))+(F(K,2)*W(I,2))+(F(K,3)*W(I,3)));
                    [LE7]  (F(K,1)*W(I,1))>(F(K,2)*W(I,2));
                    [LE8]  (F(K,1)*W(I,1))>(F(K,3)*W(I,3));

                    [BEP_1] @SUM( FACTOR(J)| J #GT# NINPUTS: F(K, J) * W(I, J))
                        - @SUM( FACTOR(J)| J #LE# NINPUTS: F(K, J) * W(I, J)) >= 0
                        )
                );

        @FOR (DXF(I, J):
        @BND (WGTMIN, W, BIGM);

        );
         DATA:
      F = @OLE ('\\vmware-host\Shared
Folders\Documents\_PhD_UHasselt_MAC\__Model_D\CAR\Cluster 1\INPUTN_1.XLS',INPUT);
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_D\CAR\Cluster
1\OUTPUT_1.XLS',SCORE)= SCORE;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_D\CAR\Cluster
1\OUTPUT_1.XLS',W)= W;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_D\CAR\Cluster
1\OUTPUT_1.XLS',F)= F;
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_D\CAR\Cluster
1\OUTPUT_1.XLS',BEP_1)= BEP_1;
```

```
      @OLE ('\\vmware-host\Shared Folders\Documents\_PhD_UHasselt_MAC\__Model_D\CAR\Cluster
1\OUTPUT_1.XLS',DUALPR)= @DUAL( BEP_1);

ENDDATA
        END
```

**Figure J.2-** Example of R® script for bootstrapping applied in Chapter 8 – Disaggregated models (e.g. car)

```
#BOOTSTRAPPING PROCESS FOR CLUSTER 1
#CLUSTER 1 IS COMPOSED BY ES, MG, RJ, SP, PR, RS, SC, DF (ALWAYS IN THIS SEQUENCE)
#CAPITAL LETTERS UNDERLINE SIGNAL PLUS NUMBER AFTER THE NAME (I.E. SAMPLE_1) MEANS THAT THE
 VALUE WILL CHANGE IN EACH INTERTATION
#ONLY NUMBER OFTER THE NAME (I.E. thetaboot1) MEANS THAT IT IS AN INTERNAL PROCESS, AND THIS
 SEQUENCE WILL BE REPEATED IN EACH INTERATION

#REQUIRED PACKAGES
require(gdata)
require(WriteXLS)

#IMPORTING THE ORIGNAL INPUT DATA
INPUTN_1=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_D/CAR/Cluster
 1/INPUTN_1.XLS")
INPUTN_1=INPUTN_1[1:8,2:4]
INPUTN_1

#ORIGINAL OUTPUT SEPARATION
thetaoriginal=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_D/CAR/Cluster
 1/OUTPUT_1.XLS")
thetaoriginal=t(t(thetaoriginal[1:8,2]))
thetaoriginal
mean=mean(thetaoriginal[,1])
mean
N=8  #Number of observations in the sample = number of DMUs (respectively ES, MG, RJ, SP, PR,
 RS, SC, DF)
sigma=(1/N)*((((thetaoriginal[1,1])-mean)^2)+(((thetaoriginal[2,1])-
 mean)^2)+(((thetaoriginal[3,1])-mean)^2)+(((thetaoriginal[4,1])-
 mean)^2)+(((thetaoriginal[5,1])-mean)^2)+(((thetaoriginal[6,1])-
 mean)^2)+(((thetaoriginal[7,1])-mean)^2)+(((thetaoriginal[8,1])-mean)^2))
sigma  # Sigma deviation of the original scores

#IMPORTING PREVIOUS OUTPUT
OUTPUT_1=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_D/CAR/Cluster
 1/OUTPUT_1.XLS")
OUTPUT_1=t(t(OUTPUT_1[1:8,2]))
OUTPUT_1

#OBTAINING A SAMPLE FROM THE OUTPUT
SAMPLE_1=t(t(sample(OUTPUT_1[,1],8,replace=TRUE)))
SAMPLE_1
MEAN_1=mean(SAMPLE_1)
MEAN_1

#INSERTING PARAMETER TO THE SMOOTHING PARAMETER CALCULATION
h=0.7329972

#APPLICATION OF EQ 4.20 FROM SIMAR AND WILSON (1995)
EPSILON1_1=rnorm(1)
EPSILON1_1
RANDOMGENERATOR1_1=function(dmu,h=0.7329972,epsilon1_1=EPSILON1_1){    # h is computed by the
 application of eq 12 from Walden (2006)
theta=dmu+h*epsilon1_1
if(theta<1){     # condition present on eq 4.20 from Simar and Wilson (1995)
theta=2-dmu-h*epsilon1_1
}
theta
}
theta1=RANDOMGENERATOR1_1(dmu=SAMPLE_1[1,1])

EPSILON2_1=rnorm(1)
EPSILON2_1
RANDOMGENERATOR2_1=function(dmu,h=0.7329972,epsilon2_1=EPSILON2_1){
```

```
theta=dmu+h*epsilon2_1
if(theta<1){
theta=2-dmu-h*epsilon2_1
}
theta
}
theta2=RANDOMGENERATOR2_1(dmu=SAMPLE_1[2,1])

EPSILON3_1=rnorm(1)
EPSILON3_1
RANDOMGENERATOR3_1=function(dmu,h=0.7329972,epsilon3_1=EPSILON3_1){
theta=dmu+h*epsilon3_1
if(theta<1){
theta=2-dmu-h*epsilon3_1
}
theta
}
theta3=RANDOMGENERATOR3_1(dmu=SAMPLE_1[3,1])

EPSILON4_1=rnorm(1)
EPSILON4_1
RANDOMGENERATOR4_1=function(dmu,h=0.7329972,epsilon4_1=EPSILON4_1){
theta=dmu+h*epsilon4_1
if(theta<1){
theta=2-dmu-h*epsilon4_1
}
theta
}
theta4=RANDOMGENERATOR4_1(dmu=SAMPLE_1[4,1])

EPSILON5_1=rnorm(1)
EPSILON5_1
RANDOMGENERATOR5_1=function(dmu,h=0.7329972,epsilon5_1=EPSILON5_1){
theta=dmu+h*epsilon5_1
if(theta<1){
theta=2-dmu-h*epsilon5_1
}
theta
}
theta5=RANDOMGENERATOR5_1(dmu=SAMPLE_1[5,1])

EPSILON6_1=rnorm(1)
EPSILON6_1
RANDOMGENERATOR6_1=function(dmu,h=0.7329972,epsilon6_1=EPSILON6_1){
theta=dmu+h*epsilon6_1
if(theta<1){
theta=2-dmu-h*epsilon6_1
}
theta
}
theta6=RANDOMGENERATOR6_1(dmu=SAMPLE_1[6,1])

EPSILON7_1=rnorm(1)
EPSILON7_1
RANDOMGENERATOR7_1=function(dmu,h=0.7329972,epsilon7_1=EPSILON7_1){
theta=dmu+h*epsilon7_1
if(theta<1){
theta=2-dmu-h*epsilon7_1
}
theta
}
theta7=RANDOMGENERATOR7_1(dmu=SAMPLE_1[7,1])
```

```
EPSILON8_1=rnorm(1)
EPSILON8_1
RANDOMGENERATOR8_1=function(dmu,h=0.7329972,epsilon8_1=EPSILON8_1){
theta=dmu+h*epsilon8_1
if(theta<1){
theta=2-dmu-h*epsilon8_1
}
theta
}
theta8=RANDOMGENERATOR8_1(dmu=SAMPLE_1[8,1])

# COLLETCTING THETA FOR EACH OPTIMUM SCORE
THETASET_1=t(t(c(theta1,theta2,theta3,theta4,theta5,theta6,theta7,theta8)))
THETASET_1

#COMPUTATION OF THE SMOOTHED BOOTSTRAPPED SCORES (THETABOOT)
thetaboot1=MEAN_1+(theta1-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot2=MEAN_1+(theta2-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot3=MEAN_1+(theta3-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot4=MEAN_1+(theta4-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot5=MEAN_1+(theta5-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot6=MEAN_1+(theta6-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot7=MEAN_1+(theta7-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot8=MEAN_1+(theta8-MEAN_1)/(1+(h^2)/(sigma))^0.5
THETABOOTSET_1=t(t(c(thetaboot1,thetaboot2,thetaboot3,thetaboot4,thetaboot5,thetaboot6,thetabo
 ot7,thetaboot8)))
THETABOOTSET_1

#CALCULATING THE THETA RATIO FOR THE COMPUTATION OF THE NEW INPUT TO BE USED ON THE NEXT
 INTERATION
THETARATIO_1=thetaoriginal/THETABOOTSET_1
THETARATIO_1

#CALCULATING AND EXPORTING THE NEW INPUT MATRIX (TO BE USED ON THE NEXT INTERATION)
bootinputcol1=(t(t(INPUTN_1[,1])))/THETARATIO_1
bootinputcol2=(t(t(INPUTN_1[,2])))/THETARATIO_1
bootinputcol3=(t(t(INPUTN_1[,3])))/THETARATIO_1


INPUT_2=cbind(bootinputcol1,bootinputcol2,bootinputcol3)
INPUT_2=data.frame(INPUT_2)
INPUT_2
WriteXLS("INPUT_2",
 ExcelFileName="/Users/jtbastos/Documents/_PhD_UHasselt_MAC/__Model_D/CAR/Cluster
 1/INPUT_2.xls")
```

# APPENDIX K

**Figure K.1-** Example of ML-DEA CI model script applied in Chapter 9 – SPI research

```
MODEL:
        ! Multiple Layer Data Envelopment Analysis based Composte Indicator;
        SETS:
        DMU/ES MG RJ SP PR RS SC DF/: !The decisionmaking units;
      SCORE, W_1, W_2, W_3;! Each decision making unit has a score to be computed;
        FACTOR/I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 I13 I14 I15 I16 I17 I18  I19 I20 I21 I22
I23 I24 I25 I26 I27/; ! There is a set of factors;
        DXF(DMU, FACTOR): W, F; ! F1(I, J) = Jth factor of DMU I;
        ENDSETS
        DATA:
        WGTMIN=0.0001; !Min weight applied to every domain;
        WGTMIN1=0.1; !Min weight applied to every factor;
        BIGM=999999; !Biggest a weight can be;
      a=0.40; !General;
      b=0.40; !Human;
      c=0.40; !Cellphone;
      d=0.40; !Seat-belt;
      e=0.40; !Helmet;
      z=0.40; !Speeding;
      g=0.40; !Roadside;
      h=0.40; !Environment;

        ENDDATA
        !-----------------------------------------------------------;
        ! The Model;
        ! Try to make everyone's score as low as possible;
        Max = @SUM(DMU: SCORE);
        ! The LP for each DMU to get its score;
        @FOR(DMU(I):
                [CI] SCORE(I) = @SUM(FACTOR(J): F(I, J)* W(I, J));
                @FOR(DMU(K):
                    [LE] @SUM(FACTOR(J): F(K, J)* W(I, J))<=1;

[LE1](F(K,1)*W(I,1)+F(K,2)*W(I,2)+F(K,3)*W(I,3)+F(K,4)*W(I,4)+F(K,5)*W(I,5)+F(K,6)*W(I,6)+F(K,
7)*W(I,7)+F(K,8)*W(I,8)+F(K,9)*W(I,9)+F(K,10)*W(I,10))<=0.95*(@SUM(FACTOR(J): F(I, J)* W(I,
J)));

[LE2](F(K,1)*W(I,1)+F(K,2)*W(I,2)+F(K,3)*W(I,3)+F(K,4)*W(I,4)+F(K,5)*W(I,5)+F(K,6)*W(I,6)+F(K,
7)*W(I,7)+F(K,8)*W(I,8)+F(K,9)*W(I,9)+F(K,10)*W(I,10))>(F(K,11)*W(I,11)+F(K,12)*W(I,12)+F(K,13
)*W(I,13)+F(K,14)*W(I,14)+F(K,15)*W(I,15)+F(K,16)*W(I,16)+F(K,17)*W(I,17)+F(K,18)*W(I,18)+F(K,
19)*W(I,19)+F(K,20)*W(I,20)+F(K,21)*W(I,21)+F(K,22)*W(I,22)+F(K,23)*W(I,23));

[LE3](F(K,11)*W(I,11)+F(K,12)*W(I,12)+F(K,13)*W(I,13)+F(K,14)*W(I,14)+F(K,15)*W(I,15)+F(K,16)*
W(I,16)+F(K,17)*W(I,17)+F(K,18)*W(I,18)+F(K,19)*W(I,19)+F(K,20)*W(I,20)+F(K,21)*W(I,21)+F(K,22
)*W(I,22)+F(K,23)*W(I,23))>=(F(K,24)*W(I,24)+F(K,25)*W(I,25)+F(K,26)*W(I,26)+F(K,27)*W(I,27));

[LE4](F(K,24)*W(I,24)+F(K,25)*W(I,25)+F(K,26)*W(I,26)+F(K,27)*W(I,27))>=0.05*(@SUM(FACTOR(J):
F(I, J)* W(I, J)));

[LE5](F(K,21)*W(I,21)+F(K,22)*W(I,22)+F(K,23)*W(I,23))>=0.05*(@SUM(FACTOR(J): F(I, J)* W(I,
J)));

[LE6](F(K,11)*W(I,11)+F(K,12)*W(I,12)+F(K,13)*W(I,13)+F(K,14)*W(I,14)+F(K,15)*W(I,15)+F(K,16)*
W(I,16)+F(K,17)*W(I,17)+F(K,18)*W(I,18)+F(K,19)*W(I,19)+F(K,20)*W(I,20))>=0.05*(@SUM(FACTOR(J)
: F(I, J)* W(I, J)));
                    [LE7](F(K,1)*W(I,1))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                    [LE8](F(K,2)*W(I,2))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                    [LE9](F(K,3)*W(I,3))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                    [LE10](F(K,4)*W(I,4))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                    [LE11](F(K,5)*W(I,5))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                    [LE12](F(K,6)*W(I,6))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                    [LE13](F(K,7)*W(I,7))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                    [LE14](F(K,8)*W(I,8))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                    [LE15](F(K,9)*W(I,9))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                    [LE16](F(K,10)*W(I,10))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
```

```
                  [LE17](F(K,11)*W(I,11))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE18](F(K,12)*W(I,12))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE19](F(K,13)*W(I,13))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE20](F(K,14)*W(I,14))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE21](F(K,15)*W(I,15))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE22](F(K,16)*W(I,16))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE23](F(K,17)*W(I,17))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE24](F(K,18)*W(I,18))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE25](F(K,19)*W(I,19))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE26](F(K,20)*W(I,20))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE27](F(K,21)*W(I,21))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE28](F(K,22)*W(I,22))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE29](F(K,23)*W(I,23))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE30](F(K,24)*W(I,24))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE31](F(K,25)*W(I,25))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE32](F(K,1)*W(I,1))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE33](F(K,2)*W(I,2))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE34](F(K,3)*W(I,3))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE35](F(K,4)*W(I,4))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE36](F(K,5)*W(I,5))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE37](F(K,6)*W(I,6))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE38](F(K,7)*W(I,7))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE39](F(K,8)*W(I,8))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE40](F(K,9)*W(I,9))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE41](F(K,10)*W(I,10))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE42](F(K,11)*W(I,11))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE43](F(K,12)*W(I,12))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE44](F(K,13)*W(I,13))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE45](F(K,14)*W(I,14))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE46](F(K,15)*W(I,15))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE47](F(K,16)*W(I,16))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE48](F(K,17)*W(I,17))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE49](F(K,18)*W(I,18))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE50](F(K,19)*W(I,19))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE51](F(K,20)*W(I,20))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE52](F(K,21)*W(I,21))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE53](F(K,22)*W(I,22))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE54](F(K,23)*W(I,23))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE55](F(K,24)*W(I,24))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE56](F(K,25)*W(I,25))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE57](F(K,26)*W(I,26))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE58](F(K,27)*W(I,27))<=0.15*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE59](F(K,26)*W(I,26))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));
                  [LE60](F(K,27)*W(I,27))>=0.001*(@SUM(FACTOR(J): F(I, J)* W(I, J)));

               );
            );
       @FOR (DXF(I, J):
       @BND (WGTMIN*WGTMIN1*WGTMIN1*WGTMIN1*WGTMIN1, W, BIGM);


       W(I, 1)+W(I, 2)+W(I, 3)+W(I, 4)+W(I, 5)+W(I, 6)+W(I, 7)+W(I, 8)+W(I, 9)+W(I,
10)=W_1(I);
      W(I, 11)+W(I, 12)+W(I, 13)+W(I, 14)+W(I, 15)+W(I, 16)+W(I, 17)+W(I, 18)+W(I, 19)+W(I,
20)+W(I, 21)+W(I, 22)+W(I, 23)=W_2(I);
      W(I, 24)+W(I, 25)+W(I, 26)+W(I, 27)=W_3(I);

       W_1(I)>=WGTMIN;
       W_2(I)>=WGTMIN;
       W_3(I)>=WGTMIN;

      (W(I, 1)+W(I, 2))>=1/5*(1-b)*W_1(I);
      (W(I, 3)+W(I, 4))>=1/5*(1-b)*W_1(I);
      (W(I, 5)+W(I, 6))>=1/5*(1-b)*W_1(I);
      (W(I, 7)+W(I, 8))>=1/5*(1-b)*W_1(I);
      (W(I, 9)+W(I, 10))>=1/5*(1-b)*W_1(I);

       (W(I, 1)+W(I, 2))<=1/5*(1+b)*W_1(I);
       (W(I, 3)+W(I, 4))<=1/5*(1+b)*W_1(I);
```

**Figure K.2-** Example of R® script for bootstrapping applied in Chapter 9 – SPI research

```
#BOOTSTRAPPING PROCESS FOR CLUSTER 1
#CLUSTER 1 IS COMPOSED BY ES, MG, RJ, SP, PR, RS, SC, DF (ALWAYS IN THIS SEQUENCE)
#CAPITAL LETTERS UNDERLINE SIGNAL PLUS NUMBER AFTER THE NAME (I.E. SAMPLE_1) MEANS THAT THE
 VALUE WILL CHANGE IN EACH INTERTATION
#ONLY NUMBER OFTER THE NAME (I.E. thetaboot1) MEANS THAT IT IS AN INTERNAL PROCESS, AND THIS
 SEQUENCE WILL BE REPEATED IN EACH INTERATION

#REQUIRED PACKAGES
require(gdata)
require(WriteXLS)

#IMPORTING THE ORIGNAL INPUT DATA
INPUTN_1=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/_SPIs/Imputation_SPSS_CNT_Impro
 ved/Cluster 1/INPUTN_1.XLS")
INPUTN_1=INPUTN_1[1:8,2:28]
INPUTN_1

#ORIGINAL OUTPUT SEPARATION
thetaoriginal=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/_SPIs/Imputation_SPSS_CNT_
 Improved/Cluster 1/OUTPUT_1.XLS")
thetaoriginal=t(t(thetaoriginal[1:8,2]))
thetaoriginal
mean=mean(thetaoriginal[,1])
mean
N=8  #Number of observations in the sample = number of DMUs (respectively ES, MG, RJ, SP, PR,
 RS, SC, DF)
sigma=(1/N)*((((thetaoriginal[1,1])-mean)^2)+(((thetaoriginal[2,1])-
 mean)^2)+(((thetaoriginal[3,1])-mean)^2)+(((thetaoriginal[4,1])-
 mean)^2)+(((thetaoriginal[5,1])-mean)^2)+(((thetaoriginal[6,1])-
 mean)^2)+(((thetaoriginal[7,1])-mean)^2)+(((thetaoriginal[8,1])-mean)^2))
sigma  # Sigma deviation of the original scores

#IMPORTING PREVIOUS OUTPUT
OUTPUT_1=read.xls("/Users/jtbastos/Documents/_PhD_UHasselt_MAC/_SPIs/Imputation_SPSS_CNT_Impro
 ved/Cluster 1/OUTPUT_1.XLS")
OUTPUT_1=t(t(OUTPUT_1[1:8,2]))
OUTPUT_1

#OBTAINING A SAMPLE FROM THE OUTPUT
SAMPLE_1=t(t(sample(OUTPUT_1[,1],8,replace=TRUE)))
SAMPLE_1
MEAN_1=mean(SAMPLE_1)
MEAN_1

#INSERTING PARAMETER TO THE SMOOTHING PARAMETER CALCULATION
h=0.8798989

#APPLICATION OF EQ 4.20 FROM SIMAR AND WILSON (1995)
EPSILON1_1=rnorm(1)
EPSILON1_1
RANDOMGENERATOR1_1=function(dmu,h=0.8798989,epsilon1_1=EPSILON1_1){    # h is computed by the
 application of eq 12 from Walden (2006)
theta=dmu+h*epsilon1_1
if(theta<1){     # condition present on eq 4.20 from Simar and Wilson (1995)
theta=2-dmu-h*epsilon1_1
}
theta
}
theta1=RANDOMGENERATOR1_1(dmu=SAMPLE_1[1,1])

EPSILON2_1=rnorm(1)
EPSILON2_1
RANDOMGENERATOR2_1=function(dmu,h=0.8798989,epsilon2_1=EPSILON2_1){
```

```
theta=dmu+h*epsilon2_1
if(theta<1){
theta=2-dmu-h*epsilon2_1
}
theta
}
theta2=RANDOMGENERATOR2_1(dmu=SAMPLE_1[2,1])

EPSILON3_1=rnorm(1)
EPSILON3_1
RANDOMGENERATOR3_1=function(dmu,h=0.8798989,epsilon3_1=EPSILON3_1){
theta=dmu+h*epsilon3_1
if(theta<1){
theta=2-dmu-h*epsilon3_1
}
theta
}
theta3=RANDOMGENERATOR3_1(dmu=SAMPLE_1[3,1])

EPSILON4_1=rnorm(1)
EPSILON4_1
RANDOMGENERATOR4_1=function(dmu,h=0.8798989,epsilon4_1=EPSILON4_1){
theta=dmu+h*epsilon4_1
if(theta<1){
theta=2-dmu-h*epsilon4_1
}
theta
}
theta4=RANDOMGENERATOR4_1(dmu=SAMPLE_1[4,1])

EPSILON5_1=rnorm(1)
EPSILON5_1
RANDOMGENERATOR5_1=function(dmu,h=0.8798989,epsilon5_1=EPSILON5_1){
theta=dmu+h*epsilon5_1
if(theta<1){
theta=2-dmu-h*epsilon5_1
}
theta
}
theta5=RANDOMGENERATOR5_1(dmu=SAMPLE_1[5,1])

EPSILON6_1=rnorm(1)
EPSILON6_1
RANDOMGENERATOR6_1=function(dmu,h=0.8798989,epsilon6_1=EPSILON6_1){
theta=dmu+h*epsilon6_1
if(theta<1){
theta=2-dmu-h*epsilon6_1
}
theta
}
theta6=RANDOMGENERATOR6_1(dmu=SAMPLE_1[6,1])

EPSILON7_1=rnorm(1)
EPSILON7_1
RANDOMGENERATOR7_1=function(dmu,h=0.8798989,epsilon7_1=EPSILON7_1){
theta=dmu+h*epsilon7_1
if(theta<1){
theta=2-dmu-h*epsilon7_1
}
theta
}
theta7=RANDOMGENERATOR7_1(dmu=SAMPLE_1[7,1])
```

```
EPSILON8_1=rnorm(1)
EPSILON8_1
RANDOMGENERATOR8_1=function(dmu,h=0.8798989,epsilon8_1=EPSILON8_1){
theta=dmu+h*epsilon8_1
if(theta<1){
theta=2-dmu-h*epsilon8_1
}
theta
}
theta8=RANDOMGENERATOR8_1(dmu=SAMPLE_1[8,1])

# COLLETCTING THETA FOR EACH OPTIMUM SCORE
THETASET_1=t(t(c(theta1,theta2,theta3,theta4,theta5,theta6,theta7,theta8)))
THETASET_1

#COMPUTATION OF THE SMOOTHED BOOTSTRAPPED SCORES (THETABOOT)
thetaboot1=MEAN_1+(theta1-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot2=MEAN_1+(theta2-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot3=MEAN_1+(theta3-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot4=MEAN_1+(theta4-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot5=MEAN_1+(theta5-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot6=MEAN_1+(theta6-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot7=MEAN_1+(theta7-MEAN_1)/(1+(h^2)/(sigma))^0.5
thetaboot8=MEAN_1+(theta8-MEAN_1)/(1+(h^2)/(sigma))^0.5
THETABOOTSET_1=t(t(c(thetaboot1,thetaboot2,thetaboot3,thetaboot4,thetaboot5,thetaboot6,thetabo
 ot7,thetaboot8)))
THETABOOTSET_1

#CALCULATING THE THETA RATIO FOR THE COMPUTATION OF THE NEW INPUT TO BE USED ON THE NEXT
 INTERATION
THETARATIO_1=thetaoriginal/THETABOOTSET_1
THETARATIO_1

#CALCULATING AND EXPORTING THE NEW INPUT MATRIX (TO BE USED ON THE NEXT INTERATION)
bootinputcol1=THETARATIO_1*(t(t(INPUTN_1[,1])))
bootinputcol2=THETARATIO_1*(t(t(INPUTN_1[,2])))
bootinputcol3=THETARATIO_1*(t(t(INPUTN_1[,3])))
bootinputcol4=THETARATIO_1*(t(t(INPUTN_1[,4])))
bootinputcol5=THETARATIO_1*(t(t(INPUTN_1[,5])))
bootinputcol6=THETARATIO_1*(t(t(INPUTN_1[,6])))
bootinputcol7=THETARATIO_1*(t(t(INPUTN_1[,7])))
bootinputcol8=THETARATIO_1*(t(t(INPUTN_1[,8])))
bootinputcol9=THETARATIO_1*(t(t(INPUTN_1[,9])))
bootinputcol10=THETARATIO_1*(t(t(INPUTN_1[,10])))
bootinputcol11=THETARATIO_1*(t(t(INPUTN_1[,11])))
bootinputcol12=THETARATIO_1*(t(t(INPUTN_1[,12])))
bootinputcol13=THETARATIO_1*(t(t(INPUTN_1[,13])))
bootinputcol14=THETARATIO_1*(t(t(INPUTN_1[,14])))
bootinputcol15=THETARATIO_1*(t(t(INPUTN_1[,15])))
bootinputcol16=THETARATIO_1*(t(t(INPUTN_1[,16])))
bootinputcol17=THETARATIO_1*(t(t(INPUTN_1[,17])))
bootinputcol18=THETARATIO_1*(t(t(INPUTN_1[,18])))
bootinputcol19=THETARATIO_1*(t(t(INPUTN_1[,19])))
bootinputcol20=THETARATIO_1*(t(t(INPUTN_1[,20])))
bootinputcol21=THETARATIO_1*(t(t(INPUTN_1[,21])))
bootinputcol22=THETARATIO_1*(t(t(INPUTN_1[,22])))
bootinputcol23=THETARATIO_1*(t(t(INPUTN_1[,23])))
bootinputcol24=THETARATIO_1*(t(t(INPUTN_1[,24])))
bootinputcol25=THETARATIO_1*(t(t(INPUTN_1[,25])))
bootinputcol26=THETARATIO_1*(t(t(INPUTN_1[,26])))
bootinputcol27=THETARATIO_1*(t(t(INPUTN_1[,27])))

INPUT_2=cbind(bootinputcol1,bootinputcol2,bootinputcol3,bootinputcol4,bootinputcol5,bootinputc
 ol6,bootinputcol7,bootinputcol8,bootinputcol9,bootinputcol10,bootinputcol11,bootinputcol12,bo
```

```
otinputcol13,bootinputcol14,bootinputcol15,bootinputcol16,bootinputcol17,bootinputcol18,booti
nputcol19,bootinputcol20,bootinputcol21,bootinputcol22,bootinputcol23,bootinputcol24,bootinpu
tcol25,bootinputcol26,bootinputcol27)
INPUT_2=data.frame(INPUT_2)
INPUT_2
WriteXLS("INPUT_2",
 ExcelFileName="/Users/jtbastos/Documents/_PhD_UHasselt_MAC/_SPIs/Imputation_SPSS_CNT_Improved
/Cluster 1/INPUT_2.xls")
```