



# Article The Severity of Traffic Crashes in Italy: An Explorative Analysis among Different Driving Circumstances

# Laura Eboli and Carmen Forciniti \*D

Department of Civil Engineering, University of Calabria, 87036 Rende, Italy; laura.eboli@unical.it \* Correspondence: carmen.forciniti@unical.it

Received: 4 January 2020; Accepted: 20 January 2020; Published: 23 January 2020



**Abstract:** Analyzing traffic accidents is very important due to their direct impact on the social environment. In the literature, many studies focus on the different aspects that influence traffic accidents, such as human, vehicle, road and environment risk factors. In this paper, we propose a methodology for testing the relationship between road, external environment, driver and vehicle characteristics, and certain circumstances that lead to the traffic crashes. Particularly, we elaborate on logistic regression models for evaluating how these different characteristics impact on crash severity, considering the combination of traffic circumstances that caused the crash. In each combination, a vehicle proceeded regularly, whereas the other vehicle did an incorrect maneuver (the vehicle proceeded: with distracted driving; without maintaining the safety distance; with speeding; by maneuvering to join the circulation flow; against the flow). The present work analyzes data related to road crashes which occurred in Italy during 2016 involving two vehicles. The results show that the variables significantly influencing crash severity are different depending on the combinations of circumstances that cause the crash.

Keywords: road safety; traffic accident circumstances; crash severity

# 1. Introduction

Improving road safety and reducing the number of road crashes represents an important strategy towards the achievement of a sustainable world. The costs due to road accidents can be referred to the accident victim (social) and to the road accident (economic) [1]. For example, in 2017, road accidents cost about EUR 20 billion in Italy, corresponding to 1.1% of the national Gross Domestic Product (GDP) [2]. Motor-vehicle crashes are unfortunately very common in the major parts of the world, and the majority of these road crashes are caused by human error. The driver is one of the main actors who contributes to the occurrence of a traffic accident. His or her behavior is responsible for numerous traffic accidents due to the high level of risk taken during driving, and also by erroneous risk perception. According to the annual road safety statistics published by the European Commission, over 90% of traffic accidents were caused by human errors in 2016; high-speed and driving under the influence of alcohol or drugs were the three causes accounting for about 80% of traffic accidents and deaths in Europe. In addition, the non-use of seat belts contributes to increasing the number of injuries and victims [3].

In 2016, 175,791 road accidents occurred in Italy resulting in death or injury: there were 3283 deaths (within 30 days) and 249,175 injured [4].

Given the importance of this issue due to the enormous losses to society, scientific research has continually been drawn to the topic. With the present paper, we want to give a useful contribution to the literature by presenting an analysis of the two-vehicle crashes which occurred on Italian roads in 2016. More specifically, we investigate on how the characteristics of road, external environment, driver and vehicle influence crash severity occurring under certain circumstances. In fact, when a crash

occurs, the police or the other authorities in charge take care of detecting the presumed circumstances of the crash. These circumstances represent situations of driving in which vehicles were at the moment of the crash. In several cases, crashes involved two vehicles, where the driver of one vehicle was proceeding regularly, whereas the driver of the other vehicle made a traffic violation or behaved incorrectly, causing the crash. The aim of the paper is just to investigate on the several aspects occurring in a traffic crash that influence its severity, by differentiating the crashes according to the circumstances. In this regard, we assume that there could be differences between a crash that happened because the driver was exceeding speed limits, and a crash caused by distracted driving. Not many studies have demonstrated these evidences, and for this reason, we decided to contribute to the literature by proposing this kind of analysis. In addition, a large portion of the studies—as we will show in the successive section—analyze data collected from experimental surveys addressed to samples of drivers, and often drivers' perceptions about crash risk; instead, we analyze data regarding real accidents referred to the whole population. This is the added value of the paper as a contribution to the literature. Since the data are recorded by police and by other authorities in charge, it is official data collected using the same procedure on the whole national territory.

In our study, we only considered the crashes that happened under five different combinations of circumstances, where the first vehicle was always under the same circumstance, which is "proceeded regularly", whereas the second vehicle could present five different circumstances always represented by an incorrect driving behavior: distracted driving; not respecting safety distance; speeding; maneuvering to get into the traffic flow; going against the traffic flow. The first stage of the analysis was directed to test the independence between the variables presumably influencing crash severity and each of the combinations of circumstances through a Chi-Square statistic. The results of this analysis were used for developing logistic regression models for each combination of circumstances that cause the road accident.

In the following section, we propose a literature review of the studies investigating on accident severity. Then, we provide a description of the data supporting our work, and we present the methodology. Section 3 regards the results of the statistical analysis of the data and of the regression models. Finally, we discuss the results and we report the main conclusions of the work.

## 2. Literature Review

The aim of this study is to investigate on the factors influencing accident severity in order to reduce it. Most of the literature studies dealing with traffic accidents focused on accident severity in terms of human fatalities and injuries. It is well-known that the main cause of accidents and crashes is human error [5–7]. Over-speeding, rash driving, violation of rules, failure to understand signs, fatigue, and alcohol are examples of common human behaviors which result in traffic accidents. Various authors adopted different kinds of mathematical models for identifying the influence of the various factors on the severity of an accident. For example, Al-Ghamdi [8] used logistic regression to analyze data collected from a sample of traffic accidents and found that accident location and cause are significantly associated with a fatal accident. de Oña et al. [9] used Latent Class clustering and Bayesian networks to identify key factors affecting accident severity. They found that the combined use of both techniques revealed further information that would not have been obtained without a prior segmentation of the data; accident type and sight distance were identified as the variables that best identified accidents with killed or seriously injured in all the analyzed cases. Dissanayake [10] adopted a logistic regression for comparing the factors affecting severity between young and older drivers involved in single vehicle crashes; speeding and non-usage of a restraint device were the two most important factors leading towards increased crash severity.

Other studies were instead oriented towards the evaluation of drivers' accident risk perception, with the aim to identify the key factors that affect risk perception and the behavior of drivers. As an example, Cardamone et al. [11] investigated the influence of the attitudes and behavior of

3 of 19

Italian drivers on the perceived risk of accidents using the Stated Preferences (SP). Among the various findings, the authors discovered that the respect/violation of the driving rules (e.g., violation of the safety distance or speed limits) and correct/incorrect driver behavior (e.g., driver's distraction) affect the level of accident risk perceptions. de Oña et al. [12] proposed a similar work for comparing driving behavior and risk perceptions of Italian and Spanish drivers. They found that both Italian and Spanish drivers consider driving in an altered psychophysical state and violating the overtaking rules to be the most risky behaviors. Machado et al. [13] investigated the differences in perceptions among drivers. As an example, they found that drivers of increased experience tend to consider certain risky driving behaviors such as "not respecting the speed limits" and "distracted driver" to be less dangerous. They found also that certain risky driving behaviors are considered more dangerous by men than by women and vice versa.

Particularly interesting are a series of studies investigating the traffic violation or circumstances that cause traffic accidents. Among others, Renner and Anderle [14] found a relationship between certain characteristics of personality and incurring traffic violations in a sample of young drivers. As reported by Ayuso et al. [15], there is no empirical analysis of the impact of traffic violations on the severity of an accident. In this regard, they found that some traffic violations cause a significant increase in the probability of suffering a severe accident, compared to situations where there are no traffic violations. Penmetsa and Pulugurtha [16] developed a multinomial logit model for examining the relation between different traffic violations and driver injury severity. The results showed that exceeding the speed limits is more likely to result in severe injury compared to disregarding traffic signals. However, going the wrong way is more likely to result in severe injury to other drivers when compared to any other traffic violation. These findings evidence the necessity to educate drivers about the adoption of safe driving behaviors. Traffic violations can be influenced by personal characteristics. As an example, drivers are more willing to limit the speed as age increases, while this willingness diminishes as the drivers' education level and income increase [17]. In addition, male drivers tend to drive faster than females [18]. Laapotti et al. [19] concluded that some driving conditions increase the risk of certain types of accidents among certain driver groups, but not among all drivers. For example, slippery road conditions were overrepresented in young male drivers' minor (self-reported) accidents, but not in their fatal accidents. For young female drivers, slippery road conditions seem to increase the propensity of fatal accidents. Zhang et al. [20] analyzed the risk factors, grouped in human, vehicle, road and environmental risk factors, related to traffic violations. They concluded that traffic violations represent one of the major risks threatening road safety. In order to reduce traffic violation rate, and consequently accident fatality rate, measures of traffic regulations and legislation have to be carried out. The authors highlighted that these measures, such as road safety programs, should consider different vehicle types or driver groups with respect to the various human, vehicle, road and environment risk factors.

Traditional statistical modeling does not allow parameter estimates to vary across observations. If some parameters do vary across observations and the model is estimated as if they were fixed, the resulting parameter estimates will be biased and possible erroneous inferences could be drawn [21].

Many studies in the literature deal with the issue of heterogeneity in crashes severity. Kim et al. [22] developed a mixed logit model of driver-injury severity in single-vehicle crashes in California, especially considering the heterogeneous effects of age and gender.

For considering both time-varying and time-constant unobserved heterogeneity in an ordered discrete-response probability model, the study of Xiong et al. [23] proposed a Markov switching random parameters structure (which accounts for heterogeneity across observations).

In some situations, it is impossible to have access to all of the data that could potentially determine the likelihood of a road accident or its resulting injury severity. The absence of such important data can potentially present serious specification problems for traditional statistical analyses, that can lead to biased and inconsistent parameter estimates, erroneous inferences and erroneous accident predictions. The problem of the unobserved heterogeneity frequently appears in the context of accident data and analysis. Various statistical approaches are available to address this unobserved heterogeneity [24].

To select the methodology that best accounts for unobserved heterogeneity between crash outcomes, Adanu and Jones [25] developed a latent class (LC) logit model and random parameters logit (RPL) model. Comparison of model fit statistics shows that the LC logit model outperformed the RPL model, as an alternative to the traditional multinomial logit (MNL) model.

In order to account for the unobserved heterogeneity, crash injury severity which occurred on a Greek urban motorway was explored by utilizing finite mixture logit models. Results indicate that a number of traffic parameters such as truck proportion, average flow and standard deviation of occupancy, as well as other risk factors, such as accident type and engine size have a significant effect on the injury severity outcome of vehicle occupants [26].

As demonstrated in the case of single-vehicle crashes occurred in Washington State from 2010 to 2016 [27], the latent class mixed logit model is able to interpret both within- and across- class unobserved heterogeneity and temporal variation, when assessing the effects of significant contributing factors to driver injury severity.

Hou et al. [28] indicated that the mixed logit model with heterogeneity-in-means can provide a superior goodness-of-fit and offer more insights into the factors behind driver injury severities. They involved elements of driver characteristics, environmental factors, roadway attributes and crash characteristics in the analysis. A heterogeneity-in-means mixed logit model was developed as an alternative to the frequently used multinomial logit model and mixed logit model, to fully account for unobserved heterogeneity, particularly the heterogeneity resulting from driver characteristics.

After a deep analysis of the current literature, we retain that investigating on the circumstances causing traffic accidents is an issue of relevant importance. However, many researchers have analyzed how the traffic violations determine crash severity, without considering the factors related to the road, external environment, vehicle and driver characteristics that influence such traffic violations. In this respect, our paper represents a useful contribution to the literature review in terms of an analysis of data concerning real-life, two-vehicle crashes with a view toward the whole Italian population. More specifically, we analyze how the various aspects characterizing a crash influence crash severity as a function of traffic violations (or circumstances) due to the human errors.

#### 3. Materials and Methods

## 3.1. Data

The work focuses on the data of traffic accidents that occurred in Italy during 2016 and collected by means of a survey conducted by the Italian Institute of Statistics (Istat) [4].

The survey, in a similar way to what happens in other European countries, was carried out in all of the Italian regions with the collaboration of public bodies with local expertise (traffic police, local or municipal police, and other bodies), that have the possibility to collect the various elements characterizing traffic accidents. The police, or the other authorities in charge that take over the crash scene, were asked to fill a paper questionnaire, regarding the features of the crash. All traffic accidents involving stationary or moving vehicles and injured persons were recorded, excluding claims from which injured persons did not occur in areas open to public traffic, and claims where no vehicles were involved. The data were then collected by Istat and each record of the database is a crash with injured or dead persons, where information on violations of the road rules is included as a support of the traditional indicators on road accidents. The crashes with injured persons included in the database correspond to crashes in which injured persons were hospitalized.

The data that referred to crashes between two vehicles were selected from the whole database. The dataset includes information about aspects of the crash regarding road, external environment, driver and vehicle involved in the crash, and accidents characteristics. In addition, the dataset provides the circumstances that caused the crash. From a statistical descriptive analysis, we selected a certain number of most frequent combinations of circumstances. More specifically, we included the crashes that occurred when vehicle A proceeded regularly whereas vehicle B proceeded with an incorrect driving behavior in the analysis. Particularly, five circumstances were considered, namely: (1) distracted driving; (2) not respecting safety distance; (3) speeding; (4) maneuvering to join the traffic flow; (5) going against the traffic flow (Table 1). Therefore, the final dataset consists of 13,629 records. The crashes that occurred when both the vehicles proceeded with incorrect driving behavior are few in the database and are not included in the analysis.

Circumstances	Vehicle A	Vehicle B	No.	Percentages
C1	regularly	distracted driving	3612	26.5
C2	regularly	not respecting safety distance	4374	32.1
C3	regularly	speeding	2347	17.2
C4	regularly	maneuvering to join the traffic flow	2264	16.6
C5	regularly	going against the traffic flow	1032	7.6
	Total		13,629	100.0

Table 1. Circumstances that cause road accidents.

The dataset includes only categorical (or qualitative) variables, that are not numerical, but describe data that fits into categories. The descriptive statistics analyses are reported in Table 2. Crashes mainly occurred within built-up areas, and on straight roads. The roads have mainly two-way roadways, dry surfaces, and vertical and horizontal signposting. The crashes are almost evenly distributed among the days of the week and the seasons, whereas they mostly occurred during afternoon and in sunny days. Most of the crashes were detected in northern Italy. It is appropriate to distinguish between north, central and south Italy to take into account the different socio-economic and infrastructural characteristics. The drivers who mainly caused road accidents were males (74.2%). This can be explained considering that, as reported in [29], males are more involved in accidents than females because they have different approaches to risk taking while driving, but also because men make more trips and/or travel longer distances. Most drivers were men aged between 30 and 44 years (about 30%), and held the car license, issued after 2006. In about 53% of cases, the driver was unharmed. Cars were the vehicles most involved in crashes, although the percentage of motorcycles is high, as accidents occur on urban roads; most of the vehicles were registered after 2006. Data regarding driver and vehicle characteristics are referred to with vehicle B, which proceeded with incorrect driving behavior and caused the road accident. The majority are non-fatal crashes, with an injured person. In certain crashes (0.4%), only dead persons are recorded.

Items	Variables	Percentages
	Location	Built-up area (68.2%); extra urban area (25.8%); highway (6.0%)
	Street	One-way roadway (13.9%); two-way roadway (70.3%); two roadways (13.9%); more than two roadways (1.8%)
Road characteristics	Stretch	Straight road (80.1%); Curve (16.6%); Bump, bottleneck (0.8%); Slope (1.8%); Lit tunnel (0.6%); Unlit tunnel (0.1%)
	Road surface	Dry (82.5%); Wet (16.7%); Slippery (0.5%); Icy (0.3%); Snowy (0.1%)
	Signposting	Absent (7.5%); vertical (7.9%); horizontal (12.5%); vertical and horizontal (71.7%); provisional, construction site (0.4%)
	Day of the week	Monday (15.3%); Tuesday (15.2%); Wednesday (15.9%); Thursday (14.9%); Friday (16.5%); Saturday (13.0%); Sunday (9.2%)
External environment characteristics	Hour	Morning [6:00–12:59] (32.8%); afternoon [13:00–18:59] (42.1%); evening [19:00–22:59] (18.2%); night [23:00–5:59] (6.9%)
	Weather conditions	Sunny (83.8%); fog (1.2%); rain (9.8%); snow (0.2%); strong wind (0.1%); other (4.9%)
	Season	Winter (21.6%); spring (27.3%); summer (26.3%); autumn (24.8%)
	Region	North (46.6%); Centre (34.7%); South (18.7%)
	Age	<18 (1.2%); 18–29 (24.1%); 30–44 (29.9%); 45–54 (20.2%); 55–64 (12.2%); >64 (12.3%)
	Gender	Male (74.2%); Female (25.8%)
Driver characteristics	License	Motorcycle license (5.5%); Car license (85.9%); Heavy vehicle license (8.5%); Driving license not required (0.2%)
	License issue year	Before 1976 (2.0%); (1976–1986] (5.8%); (1986,1996] (9.2%); (1996,2006] (14.0%); (2006,2016] (69.0%)
	Driver's injury	Unharmed (53.3%); Injured (45.9%); Dead within 24 h (0.7%), dead within 30 days (0.1%)
Vehicle characteristics	Vehicle	Car (74.2%); Bus (0.6%); Heavy vehicle (7.8%); Motorcycle (16.0%); Other vehicles (1.4%)
	Year of vehicle registration	Before 1986 (0.5%); (1986,1996] (3.9%); (1996,2006] (30.9%); (2006,2016] (64.6%)
Accident severity	Dead within 24 h	0(98.8%); 1(1.1%); 2(0.1) 0(99.7%) + 1(0.3%)
Accident severity	Injured persons	0 (22.7%), 1 (0.5%) 0 (0.4%), 1 (63.9%); 2 (24.4%); 3 (7.4%); 4(2.7%); 5(0.8%); >5 (0.5%)

# Table 2. Dataset descriptive statistics.

The variables were also analysed considering the combinations of circumstances. On the basis of the literature and of previous studies, we chose the variables regarding road, external environment, driver and vehicle characteristics that we retain could influence the circumstances that cause the crash between two vehicles.

The Tables 3–6 show the categories belonging to each variable and the percentages corresponding to the five combinations of circumstances. Observing the percentages reported in these tables, some differences of the same category can be highlighted among the circumstances.

Variables	Categories	C1 [%]	C2 [%]	C3 [%]	C4 [%]	C5 [%]
	Within built-up area	70.1	65.6	66.0	79.5	52.3
Location	In extra urban area	21.3	25.0	30.5	19.7	47.6
	Highway	8.6	9.4	3.5	0.8	0.1
	One-way roadway	15.2	16.6	11.0	13.1	6.6
Stroot	Two-way roadway	66.5	64.4	72.0	78.7	86.5
Sheet	Two roadways	16.3	17.0	14.5	6.9	6.2
	More than two roadways	2.0	2.0	2.4	1.3	0.7
	Straight road	80.8	85.6	71.4	86.3	60.0
	Curve	15.8	11.3	24.9	10.8	36.1
Stratab	Bump, bottleneck	0.7	0.7	1.0	0.8	1.2
Stretch	Slope	2.2	1.4	1.9	2.0	1.9
	Lit tunnel	0.5	0.8	0.6	0.1	0.8
	Unlit tunnel	0.1	0.1	0.2	0.0	0.0
	Dry	83.6	85.1	74.6	84.4	81.5
	Wet	15.6	14.3	23.6	15.0	18.2
Surface	Slippery	0.4	0.4	0.9	0.4	0.1
	Icy	0.3	0.1	0.6	0.2	0.1
	Snowy	0.1	0.1	0.3	0.0	0.1
	Absent	6.4	6.1	7.9	10.8	9.1
	Vertical	6.9	7.9	7.6	8.0	11.2
Signposting	Horizontal	11.6	13.1	12.0	13.2	12.7
	Vertical and horizontal	74.8	72.3	72.0	67.8	66.4
	Provisional, construction site	0.3	0.6	0.5	0.3	0.6

Table 3. Road variables and circumstances.

Table 4. External environment variables and circumstances.

Variables	Categories	C1 [%]	C2 [%]	C3 [%]	C4 [%]	C5 [%]
	Monday	15.0	16.8	15.1	13.3	14.5
	Tuesday	15.8	15.0	14.7	15.9	13.2
Davistika	Wednesday	15.6	15.9	15.6	17.7	13.6
Day of the	Thursday	15.6	14.8	15.4	14.4	12.7
WEEK	Friday	16.2	16.9	15.8	17.1	16.8
	Saturday	13.0	11.8	13.7	13.1	16.4
	Sunday	8.9	8.8	9.7	8.5	12.9
	Morning [6:00-12:59]	42.0	40.9	44.4	41.7	43.5
Harr	Afternoon [13:00-18:59]	18.4	18.4	16.8	19.3	17.1
Hour	Evening [19:00-22:59]	32.6	34.3	32.0	33.4	27.9
	Night [23:00-5:59]	7.0	6.5	6.9	5.6	11.5
	Sunny	84.3	85.3	79.3	85.1	83.9
	Fog	1.2	0.8	1.1	1.5	1.8
Weather	Rain	9.1	8.7	14.1	9.1	9.4
conditions	Snow	0.2	0.1	0.3	0.0	0.1
	Strong wind	0.0	0.1	0.2	0.0	0.1
	Other	5.1	5.0	4.9	4.3	4.7
	Winter	21.2	21.5	24.0	21.2	18.3
Casaan	Spring	28.4	26.8	25.5	28.5	27.3
Season	Summer	24.6	28.1	24.8	26.1	28.5
	Autumn	25.8	23.7	25.7	24.1	25.9
	North	50.7	44.8	48.6	45.1	38.8
Region	Centre	34.4	34.9	30.9	36.4	39.3
	South	14.9	20.3	20.5	18.5	21.9

Variables	Categories	C1 [%]	C2 [%]	C3 [%]	C4 [%]	C5 [%]
	<18	1.0	1.5	1.5	0.9	0.5
	18–29	24.3	23.8	28.3	18.0	28.8
Δσο	30–44	29.9	30.7	31.7	27.4	28.2
Age	45–54	21.2	21.0	19.2	19.9	16.8
	55-64	11.8	12.3	11.2	14.0	11.4
	>64	11.7	10.7	8.1	19.8	14.3
Gender	Male	73.5	74.3	75.9	72.3	77.0
	Female	26.5	25.7	24.1	27.7	23.0
	Motorcycle license	5.7	6.2	5.5	3.7	3.7
Driver	Car license	85.2	84.2	85.6	89.6	89.6
license	Heavy vehicle license	8.9	9.6	8.7	6.6	6.6
	Driving license not required	0.3	0.1	0.2	0.1	0.1
	Before 1976	2.1	1.8	1.7	2.4	1.6
Liconso issuo	(1976–1986]	5.5	5.3	5.8	6.8	6.0
License issue	(1986–1996]	9.6	8.8	8.8	10.2	8.2
year	(1996–2006]	13.7	14.4	13.8	15.2	11.7
	(2006–2016]	69.0	69.7	69.8	65.4	72.5

Table 5. Driver variables and circumstances.

Table 6. Vehicle variables and circums	stances.
--	----------

Variables	Categories	C1 [%]	C2 [%]	C3 [%]	C4 [%]	C5 [%]
	Car	71.7	71.7	70.8	84.5	78.7
Valatiala	Heavy vehicle	8.9	9.4	8.4	6.8	6.1
venicle	Motorcycle	1.6	1.6	1.3	1.1	0.4
	Other vehicles	17.8	17.4	19.5	7.6	14.8
	Before 1986	0.5	0.6	0.5	0.6	0.2
Year of vehicle	(1986–1996]	3.2	3.9	4.0	4.7	4.9
registration	(1996–2006]	29.7	30.9	33.1	30.1	31.9
	(2006–2016]	66.6	64.6	62.4	64.6	63.0

From Table 3, we can observe that for circumstances from 1 to 4, most of the crashes happened within built-up areas, while for circumstance 5, which occurs when one vehicle proceeds against the traffic flow, the distribution of the crashes according the location is evenly spread between built-up areas and extra urban areas, and never on the highway. Almost all the crashes that occurred in circumstance 5 happened on a two-way roadway. This is also the most frequent type of street for the other circumstances, but to a lesser degree of severity. Over 80% of crashes happened on a straight road for circumstances 1, 2 and 4; while for circumstance 5 there is a relevant percentage (36.1%) of road accidents which occurred in a curve, which goes to about 25% for the circumstance 3 (speeding). The majority of the crashes happened on a dry surface, amounting to three quarters of all the circumstances. Furthermore, regarding signposting, there are no relevant differences of percentage between the five circumstances; a percentage varying from 66% to 75% happened on a part of road with vertical and horizontal signposting.

From Table 4, the most common day of the week for a crash to occur is Friday for all the circumstances, with the exception of the case of one vehicle maneuvering to join the traffic flow, which occurred mostly on Wednesday. Over 40% of the crashes happened in the morning and about 30% in the evening. The fifth circumstance is more frequent in the night, where one vehicle proceeded against the flow (11.5%).

Regarding weather conditions, we can say that most of the crashes happened on sunny days for all the circumstances, and that circumstance 3 (one vehicle speeding) is at its most frequent when it is raining. In the spring, accidents in circumstances 1 and 4 mostly occurred (when one vehicle

drives while distracted, or by maneuvering to get into the traffic flow), while in the summer crashes in circumstance 2 and 5 (not respecting safety distance or going against the traffic flow). Almost half of the crashes happened in Northern Italy; this percentage is slightly higher for circumstance 1, and lower for circumstance 5. Percentages from about 30% to 40% can be observed for the central regions. Finally, about 20% of the crashes happened in the South of Italy; this percentage decreases to about 15% for circumstance 1.

If we observe the characteristics of the driver (Table 5), we can note that there is a relevant percentage (about 20%) of crashes involving drivers older than 64 years for circumstance 4 (one vehicle maneuvering to join the traffic flow) in comparison to the other circumstances. There are no relevant differences concerning gender and the type of driving license.

Finally, from Table 6 we can observe that crashes occurred in circumstance 4 and 5 (one vehicle maneuvering to join the traffic flow or going against the traffic flow) involved cars more than the other circumstances.

These variations in the percentages could be casual, or a sign of the different weights that the categories of the same variable have on the occurrence of certain circumstances that lead to a road accident. This could be better investigated using the following proposed methodology.

## 3.2. Methodology

The variables that influence the severity of a crash can be multiple, as reported in many papers cited in the literature review. On the basis of the literature and of the available data, we have selected a list of variables that could impact on crash severity, and grouped these into four items, as road, external environment, driver, and vehicle. Unlike the studies already present, we aim to analyze the possible relationship between these variables and the circumstances that cause the crash.

This objective can be pursued by conducting the chi-square test for independence between two categorical variables, *X* and *Y*. The null hypothesis  $H_0$  to be rejected and the alternative hypothesis  $H_1$  are the following:

#### $H_0$ . X and Y are independent (there are no relationships between the two variables);

#### $H_1$ . X and Y are dependent (there is a relationship between the two variables).

The Chi-Square Test is based on the next formula:

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$
(1)

where  $f_0$  is the observed frequency and  $f_e$  the expected frequency.

The decision rule consists of rejecting H<sub>0</sub> if the observed value of the Chi-Square statistic is greater than the critical value  $\chi^2_U$  of the distribution  $\chi^2$  with  $(r - 1) \times (c - 1)$  degrees of freedom, with *r* and *c* the number of categories of each variable.

By using this statistic, we tested the independence considering each selected variable and each combination of circumstances—the variables for which the hypothesis of independence with the combination of circumstances was not verified were included in the Logistic regression model as independent variables. The dependent variable corresponds to accident severity.

The Logistic regression model is a regression method applied when the dependent variable Y is dichotomous. Therefore, the method of the regression function estimation joins the probability that the response takes the value of 1 with a set of explanatory or independent variables [30]. The explanatory variables  $X_k$  can be dichotomous, nominal, ordinal or quantitative.

The probability of *Y* occurring can be written as:

$$P(Y) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}$$
(2)

where:  $\beta_0$  is the intercept at *y*-axis,  $x_n$  are the independent variables and  $\beta_n$  are the coefficients of the independent variables.

While in linear regression, the estimation of the *Y* variable has a value between  $-\infty$  and  $+\infty$ , in the logistic regression the estimation of *Y* varies between 0 and 1, and assumes the meaning of probability that *Y* takes the value of 1 as a function of the explanatory variables  $X_k$ . When the variables are of a qualitative nature, the  $\beta_k$  coefficients are estimated by the Maximum Likelihood method, which provides the parameter values for maximizing the probability of observing the experimental data set.

Five different models were calibrated, taking into account the five combinations of presumed circumstances that cause the crash. The independent variables can be treated as categorical (or qualitative) variables that are not numerical, but describe data that fits into categories. The dependent variable describes the crash severity by considering—for each crash—the number of dead persons within 24 h, the number of dead persons within 30 days, and the number of injured persons. We recoded the variable into two levels: 0 for crashes where only one injured person is recorded (not much serious crashes), 1 for crashes where there are more injured or dead persons (serious and fatal crashes) [8,31].

## 4. Results

The Chi-Square Test for independence was performed to check the independence between each variable and each combination of circumstances. We chose to reject the hypothesis of independence when the  $\chi^2$  statistic was greater than the critical value with a 99% significance level.

The results of the statistic test related to circumstance 1 are reported in Table 7. In a few cases, the hypothesis of independence can be rejected. In particular, circumstance 1 (Vehicle A proceeded regularly, and vehicle B proceeded driving with distraction) depends on the variables "Location", "Street", "Signposting", "Region", "Vehicle". The other variables and the results for circumstance 1 are independent.

Items	Variables	Chi-Square Test	df	<i>p</i> -Value
	Location	95.514	2	0.0000
	Street	36.816	3	0.0000
Road	Stretch	6.684	5	0.2452
	Surface	4.542	4	0.3376
	Signposting	26.305	4	0.0000
	Day of the week	4.394	6	0.6235
To to see 1	Hour	0.247	3	0.970
External	Weather conditions	8.333	6	0.2147
environment	Season	9.356	3	0.0249
	Region	215.349	19	0.0000
	Age	5.68	5	0.3387
Driver	Gender	1.433	1	0.2314
	Driver license	6.303	3	0.0978
	License issue year	2.626	4	0.6223
Vehicle	Vehicle	16.965	3	0.0007
	Year of vehicle registration	12.72	5	0.0262

Table 7. Chi-Square Test for independence (Circumstance 1: vehicle A regularly-vehicle B distracted driving).

In Table 8, we reported the results of the statistic test related to circumstance 2 (Vehicle A proceeded regularly, and vehicle B proceeded not respecting safety distance). In this case, the hypothesis of independence can be rejected for more variables with regard to circumstance 2. Specifically, circumstance 2 depends on all the variables related to road, and for other kind of variables, such as "Region", "Age", "Driving license", and "Vehicle".

Items	Variables	<b>Chi-Square Test</b>	df	<i>p</i> -Value
	Location	128.053	2	0.0000
	Street	110.225	3	0.0000
Road	Stretch	144.976	5	0.0000
	Surface	31.509	4	0.0000
	Signposting	22.809	4	0.0001
	Day of the week	19.39	6	0.0036
Entranal	Hour	8.699	3	0.0340
External	Weather conditions	18.676	6	0.0047
environment	Season	11.965	3	0.0075
	Region	92.164	19	0.0000
	Age	21.194	5	0.0007
Duiton	Gender	0.004	1	0.9519
Diivei	Driver license	18.375	3	0.0004
	License issue year	5.345	4	0.2537
Vahiala	Vehicle	21.701	3	0.0001
venicie	Year of vehicle registration	5.643	5	0.3425

**Table 8.** Chi-Square Test for independence (Circumstance 2: vehicle A regularly–vehicle B not respecting safety distance).

Table 9 shows the results of the statistic test related to the circumstance 3; where Vehicle A proceeded regularly, and vehicle B proceeded with speeding. Specifically, the variables that can be considered as having influence on circumstance 3 belong, for the most part, to road ("Location", "Street", "Stretch", "Surface") and external environment ("Weather conditions", "Season", "Region") characteristics. The occurrence of a crash caused by speeding is also dependent on driver age and the type of vehicle.

Table 9.	Chi-Square	Test for independence	(Circumstance 3:	vehicle A regul	arly–vehicle l	B speeding).
----------	------------	-----------------------	------------------	-----------------	----------------	--------------

Items	Variables	Chi-Square Test	df	<i>p</i> -Value
	Location	57.868	2	0.0000
	Street	23.982	3	0.0000
Road	Stretch	146.494	5	0.0000
	Surface	136.77	4	0.0000
	Signposting	1.666	4	0.7969
	Day of the week	3.742	6	0.7115
External	Hour	7.069	3	0.0700
	Weather conditions	73.327	6	0.0000
environment	Season	14.658	3	0.0021
	Region	141.405	19	0.0000
	Age	71.091	5	0.0000
Duiman	Gender	4.296	1	0.0382
Driver	Driver license	0.209	3	0.9761
	License issue year	1.635	4	0.8025
Valatela	Vehicle	26.215	3	0.0000
venicie	Year of vehicle registration	7.324	5	0.1977

Table 10 reports the results concerning circumstance 4; where Vehicle A proceeded regularly, and vehicle B proceeded by maneuvering to get into the traffic flow. In this case, the variables for which the hypothesis of independence is rejected mostly belong to road ("Location", "Street", "Stretch",

"Signposting") and driver ("Age", "Driver license", and "License issue year") characteristics. The other variables influencing the occurrence of circumstance 4 are "Hour", "Region", and "Vehicle".

Items	Variables	Chi-Square Test	df	<i>p</i> -Value
	Location	211.213	2	0.0000
	Street	126.911	3	0.0000
Road	Stretch	79.442	5	0.0000
	Surface	10.032	4	0.0399
	Signposting	49.299	4	0.0000
	Day of the week	16.172	6	0.0129
E. G. and	Hour	8.917	3	0.0300
External	Weather conditions	12.058	6	0.0607
environment	Season	2.158	3	0.5402
	Region	71.331	19	0.0000
	Age	181.887	5	0.0000
Duiner	Gender	5.223	1	0.0223
Driver	Driver license	32.511	3	0.0000
	License issue year	18.864	4	0.0008
Vahiala	Vehicle	170.326	3	0.0000
venicie	Year of vehicle registration	5.366	5	0.3728

**Table 10.** Chi-Square Test for independence (Circumstance 4: vehicle A regularly–vehicle B maneuvering to get into the traffic flow).

Finally, in Table 11, the results concerning circumstance 5 are shown. In this circumstance, Vehicle A proceeded regularly, and vehicle B proceeded going against the traffic flow. Furthermore, in this case, the variables for which the hypothesis of independence is rejected mostly belong to road characteristics ("Location", "Street", "Stretch", "Signposting"). Concerning environmental factors, there is a dependence regarding the day of the week and the hour of the accident, in addition to the region. The other variables influencing the occurrence of circumstance 5 are "Age" and "Vehicle".

Items	Variables	<b>Chi-Square Test</b>	df	<i>p</i> -Value
	Location	312.128	2	0.0000
	Street	140.591	3	0.0000
Road	Stretch	315.867	5	0.0000
	Surface	6.134	4	0.1894
	Signposting	25.072	4	0.0000
	Day of the week	36.756	6	0.0000
E. G. and I	Hour	43.936	3	0.0000
External	Weather conditions	11.088	6	0.0857
environment	Season	7.983	3	0.0464
	Region	84.178	19	0.0000
	Age	26.64	5	0.0001
	Gender	4.611	1	0.0318
Driver	Driver license	7.153	3	0.0672
	License issue year	8.417	4	0.0774
37.1.1.1.	Vehicle	18.982	3	0.0003
Vehicle	Year of vehicle registration	6.329	5	0.2755

**Table 11.** Chi-Square Test for independence (Circumstance 5: vehicle A regularly–vehicle B going against the traffic flow).

After the statistical analysis, logistic regression models were performed including the variables from which the circumstances are not independent. Each variable is divided into the respective categories. In this way, each category can be considered as a variable in the model. The reference case is based on the last category of the variables. Exp (B) is the probability that a fatal crash occurs, compared to the reference case. The significance of the variables was tested by using the Wald statistic. Not all the variables produced significant results from the calibration. In the following tables, the results of the final models are shown.

The first model (Table 12) considers the crashes that occur when vehicle A proceeded regularly and vehicle B proceeded with distracted driving (circumstances 1). The independent variables are location, street, region and vehicle. The reference case corresponds to highway (location); more than two roadways (street); south (region); other vehicles (vehicle). The Cox and Snell R Square and Nagelkerke R Square have low values, if compared to R Square values typically encountered with good linear regression models. However, certain authors [32] have argued that low R Square values in logistic regression are the norm and, thus, this indicator is not recommended as a measure of model goodness. The independent variables included into the model explicate more than 67% of the dependent variables.

Variables	В	Wald	Sig.	Exp(B)
location (within built-up area)	-0.965	44.692	0.000	0.381
street (two-way roadway)	0.629	4.364	0.037	1.877
region (north)	-0.502	20.710	0.000	0.591
region (centre)	-0.429	14.526	0.000	0.651
vehicle (car)	0.780	52.391	0.000	2.182
vehicle (heavy vehicle)	0.322	3.992	0.046	1.379
constant	-1.509	2.949	0.086	0.221
No. of observations 3612				
Cox and Snell R Square		0.076		
Nagelkerke R Square		0.105		
Overall predicted percentage		67.3		

Table 12. Model results (Circumstance 1: vehicle A regularly-vehicle B distracted driving).

In relation to the location, the probability of having a serious crash decreases by about 62% when the road accident occurs within a built-up area, compared to crashes that occur on highway. The probability increases by about 90% on a street with a two-way roadway. Compared with the south region, the probability of having a serious crash decreases in the northern and central regions. Considering the vehicle, the probability that a serious crash occurs increases in the case of car and heavy vehicles compared to other vehicles.

Model 2 (Table 13) considers the accidents that occur when vehicle A proceeds regularly, and vehicle B proceeds without respecting safety distances (circumstance 2). The independent variables are location, day of the week, region, age and vehicle. The reference case corresponds to highway (location); Sunday (day of the week); south (region); >64 (age); other vehicles (vehicle). Even if the R-Square values are low, the overall predicted percentage is above 67%.

Variables	В	Wald	Sig.	Exp(B)
location (within built-up area)	-0.774	33.883	0.000	0.475
location (in extra urban area)	-0.341	6.091	0.014	0.711
day (Monday)	-0.538	16.011	0.000	0.584
day (Tuesday)	-0.612	19.540	0.000	0.542
day (Wednesday)	-0.639	21.572	0.000	0.528
day (Thursday)	-0.567	16.671	0.000	0.567
day (Friday)	-0.554	19.936	0.000	0.575
day (Saturday)	-0.255	3.210	0.073	0.775
region (north)	-0.452	24.560	0.000	0.636
region (centre)	-0.579	37.051	0.000	0.560
age (18–29)	0.282	5.317	0.021	1.326
age (30–44)	0.208	3.065	0.080	1.231
vehicle (car)	1.113	98.240	0.000	3.045
vehicle (truck)	0.939	37.987	0.000	2.557
vehicle (motorcycle)	0.769	7.300	0.007	2.157
constant	1.504	0.000	1.000	4.499
No. of observations		4374		
Cox and Snell R Square		0.087		
Nagelkerke R Square		0.120		
Overall predicted percentage		67.8		

Table 13. Model results (Circumstance 2: vehicle A regularly-vehicle B not respecting safety distance).

The probability that a serious crash occurs decreases if the crash happens within a built-up area or in extra urban area compared to the highway. A serious crash is more likely on Sundays than on the other days of the week. Even in circumstances where the vehicle proceeded without respecting safety distance, the probability of having a serious crash decreases in the northern and central regions compared with the southern regions. Crashes are more likely among drivers aged between 18 and 29 than between 30 and 44 years. If the vehicle that causes the crash is a car, a heavy vehicle or a motorcycle, then serious crashes have a greater probability to happen, in comparison with other vehicles.

In model 3 (Table 14), crashes that occur when vehicle A proceeds regularly and vehicle B proceeds while speeding (circumstances 3) are included. The independent variables are location, stretch, weather conditions, region, and vehicle. The reference case corresponds to highway (location); unlit tunnel (stretch); weather (other); south (region); other vehicles (vehicle). The overall predicted percentage is above 67%.

Variables	В	Wald	Sig.	Exp(B)
location (within built-up area)	-0.781	9.271	0.002	0.458
stretch (straight road)	-2.397	3.975	0.046	0.091
stretch (curve)	-2.104	3.048	0.081	0.122
stretch (slope)	-2.502	4.009	0.045	0.082
stretch (lit tunnel)	-2.202	2.792	0.095	0.111
weather (sunny)	0.365	2.716	0.099	1.441
weather (rain)	0.445	3.184	0.074	1.560
region (north)	-0.625	27.327	0.000	0.535
region (centre)	-0.691	28.440	0.000	0.501
vehicle (car)	1.359	90.945	0.000	3.891
vehicle (heavy vehicle)	0.513	5.820	0.016	1.671
vehicle (motorcycle)	0.951	5.365	0.021	2.589
constant	1.195	0.384	0.536	3.302
No. of observations	5	2347		
Cox and Snell R Square		0.134		
Nagelkerke R Square		0.181		
Overall predicted percentage		67.6		
weather (sunny) weather (rain) region (north) region (centre) vehicle (car) vehicle (heavy vehicle) vehicle (motorcycle) constant No. of observations Cox and Snell R Squa Nagelkerke R Squar Overall predicted percer	0.365 0.445 -0.625 -0.691 1.359 0.513 0.951 1.195 source reentage	2.716 3.184 27.327 28.440 90.945 5.820 5.365 0.384 2347 0.134 0.181 67.6	0.099 0.074 0.000 0.000 0.000 0.016 0.021 0.536	1.441 1.560 0.535 0.501 3.891 1.671 2.589 3.302

Table 14. Model results (Circumstance 3: vehicle A regularly-vehicle B speeding).

If the crash happens within a built-up area, the probability of having a serious crash decreases compared to the reference case. A serious crash is less likely if it happens on a straight road, curve, slope, or lit tunnel, compared to in unlit tunnels. On rainy days, serious crashes are more likely compared to on sunny days and in other weather conditions. On the basis of the results, in the northern and central regions, the probability of having a serious crash decreases compared with the southern regions. Furthermore, in these circumstances, the vehicle with the highest probability of causing a serious road accident is the car, followed by the motorcycle and then the heavy vehicles, compared to the case of other vehicles.

Model 4 (Table 15) refers to the crashes that occurred when vehicle A proceeded regularly, whereas vehicle B proceeded by maneuvering to join the traffic flow (circumstance 4). The model was developed considering location, hour, region and license issue year as independent variables. The reference case corresponds to within built-up area (location); night (hour); south (region); and before 1976 (license issue year). The overall predicted percentage is 73%.

**Table 15.** Model results (Circumstance 4: vehicle A regularly–vehicle B maneuvering to get into the traffic flow).

Variables	В	Wald	Sig.	Exp(B)
location (within built-up area)	-0.994	3.873	0.049	0.370
hour (morning)	-1.102	27.800	0.000	0.332
hour (afternoon)	-0.946	21.582	0.000	0.388
hour (evening)	-0.552	6.578	0.010	0.576
region (north)	-0.478	12.416	0.000	0.620
region (centre)	-0.622	19.865	0.000	0.537
license issue year (before 1976)	0.580	3.657	0.056	1.786
constant	23.585	0.000	1.000	
No. of observations	No. of observations 2264			
Cox and Snell R Squar	Cox and Snell R Square 0.064			
Nagelkerke R Square		0.093		
Overall predicted percentage		73.0		

Observing the values of Exp(B), within the built-up area, the probability that a serious crash occurs in these circumstances decreases compared to the reference case. The probability of having a serious crash decreases during the other hours of the day compared to at night. In addition, the probability decreases in the northern and central regions compared to the reference case represented by the southern region.

Circumstance 5, where vehicle A proceeded regularly, and vehicle B proceeded against the traffic flow are considered in model 5 (Table 16). The independent variables included in the model are street, day, hour, region, and vehicle. The reference case corresponds to more than two roadways (street); Sunday (day); night (hour); south (region); other vehicles (vehicle). The overall predicted percentage is above 66%.

The probability of having a serious crash increases if the street has two roadways, compared to the reference case of more than two roadways. The probability of having a serious crash decreases if it happens on a Tuesday morning or afternoon, in the central region, compared to the reference case. If the vehicle that causes the crash is a car, the serious crashes have a greater probability of happening compared to the case of other vehicles.

Variables	В	Wald	Sig.	Exp(B)
street (two roadways)	2.066	3.314	0.069	7.891
day (Tuesday)	-0.467	2.964	0.085	0.627
hour (morning)	-0.542	4.517	0.034	0.582
hour (afternoon)	-0.453	3.526	0.060	0.636
region (centre)	-0.355	3.607	0.058	0.701
vehicle (car)	1.016	26.248	0.000	2.761
constant	21.987	0.000	1.000	
No. of observations		1032		
Cox and Snell R Square		0.103		
Nagelkerke R Square		0.139		
Overall predicted percentage		66.3		

Table 16. Model results (Circumstance 5: vehicle A regularly-vehicle B going against the traffic flow).

#### 5. Discussion

From the comparison of the five models, interesting findings can be highlighted concerning crash severity in relation to the different analyzed circumstances. First, we can note that the sole variable included in all the models is "region", and it has the same meaning for all the five circumstances. More specifically, we registered that the probability of having a serious crash decreases in the northern and central regions, compared with the southern regions independently from the circumstances. In other words, serious crashes are more likely in the South of Italy, although it is interested by the lowest number of crashes as regards the rest of Italy.

The variable linked to location and vehicle resulted as significant only for the first three circumstances. Concerning the location of the crashes, we can conclude that the probability of having a serious crash decreases when it occurs in other areas that are not highways. In other words, serious crashes are more likely on the highways, although the major part of crashes happen within built-up or extra-urban areas. The not correct behavior of the drivers has a more negative impact on the highways, where drivers run at a higher speed, than the other roads. On the other hand, regarding the type of vehicle, we observed that the probability that a serious crash occurs increases in the case of car and heavy vehicles, compared to other vehicles; when crash circumstance is the non-respect of safety distance, also motorcycles represent a category increasing the probability of having a serious crash.

The hour of the crash has a meaning for the last two analyzed circumstances: where one vehicle proceeded by maneuvering to get into the flow or going against the flow. More specifically, the probability of having a serious crash decreases during the other hours of the day compared to the night. In other words, serious crashes in these two circumstances are more likely during the night. Furthermore, for this variable, the category for which crashes were more serious is the category where the lowest percentage of crashes was registered; in fact, only about 6% of crashes happened during the night in circumstance 4, and about 11% in the fifth circumstance. Surely, the darkness of the night represents a factor that increases the possibility of the driver incurring a road accident.

Finally, in circumstance 3 (where one vehicle proceeded with speeding) the variables related to stretch, surface, and weather have an influence on crash severity. Specifically, a serious crash is less likely if it happens in straight road, curve, slope, or lit tunnel, compared to the unlit tunnel, while it is more likely on slippery surfaces and in rainy days. This proves that participating in dangerous behavior such as speeding is riskier in critical conditions of the weather and road surface. Furthermore, Laapotti et al. [19] experimented the influence of slippery surface on accidents severity; more specifically, they discovered that for the specific category of young female drivers, slippery road conditions seemed to increase the propensity of fatal accidents.

## 6. Conclusions

The present work aimed to analyze traffic crash severity as a function of different circumstances that occur when an accident happens. The circumstances represent traffic violations or human errors made while driving. Several factors related to road, external environment, driver, and vehicle characteristics were considered as variables that can play a role influencing the severity of crashes happened in certain circumstances. This kind of analysis allows us to investigate how the influence on crash severity of the various aspects intervening in a crash differs as a function of the human errors causing the crash. More specifically, five combinations of circumstances (or traffic violations, or human errors) were selected from the database. In each combination, one vehicle proceeded regularly, whereas the other vehicle did an incorrect maneuver (the vehicle proceeded: with distracted driving; without respecting the safety distance; with speeding; by maneuvering to join the circulation flow; by going against the flow). The analysis was conducted considering only crashes happened between two vehicles on Italian roads during 2016.

Five logistic regression models were calibrated, each for one combination of circumstances. The dependent variable of the models was represented by a dichotomous variable explaining crash severity between less serious crashes and serious and fatal accidents. Not all the investigated variables related to road, external environment, driver, and vehicle characteristics were significant in all the models; however, interesting findings resulted from the comparisons.

The results highlighted the variables that mainly affected crash severity in the different circumstances that lead to the crash. The variables mostly influencing severity relate to the region of Italy (divided in North, Centre, and South), the location of the crash, and the type of vehicle. Only for certain circumstances did other factors—such as weather conditions, the time of the crash, or road characteristics such as stretch or surface—influence the severity. Regardless, the variables concerning the characteristics of the drivers resulted as not significant in the differentiation of the accidents according to the circumstances. However, we can conclude that there is a certain difference in the influence of some crash aspects on crash severity depending on driver behavior. For example, we demonstrated that some human errors (represented by the analyzed circumstances) entail more severity if the crash happens in certain weather conditions or on certain road surfaces, and also depending on the type of road or the region.

Definitively, the findings from the proposed models can be useful for characterizing the various types of accidents and understand which characteristics have to be considered and improved for reducing the severity of traffic accidents and increasing the levels of safety on the road. What we registered from our analysis is that there are regions in Italy where particular attention should be addressed in order to improve the conditions of the roads and reduce the severity of the crashes. We also verified that the night is more dangerous than the other hours of the day in terms of crash severity; this evidence suggests that we should take remedial actions for reducing this phenomenon, for example, improving the illumination of the roads. Furthermore, the use of ITS (Intelligent Transport Systems) oriented to improve vehicle performances and assist the drivers while driving represents a strategy to reduce crash severity and improve road safety.

Currently, a limitation of the present work is the issue of unobserved heterogeneity. In the analysis, we neglected the unobserved heterogeneity in the factors that influence crash severity. The main reason is that this work is an exploratory analysis of the impact that the characteristics of road, external environment, driver and vehicle have on crash severity in different driving circumstances. The proposed models aim to evaluate the influence of each variable on crash severity, taking into account the circumstances that cause the road accident. After this wide analysis of data, the future development of the work will be addressed towards the issue of unobserved heterogeneity related to crash severity, developing random-parameters models.

**Author Contributions:** L.E. and C.F. jointly worked to the conceptualization of the research idea and to the implementation of the methodology; L.E. worked on the analysis of the literature review on the treated subject, while C.F. worked on the data elaboration and model calibration; L.E. and C.F. jointly worked on the writing-original

draft preparation, writing-review and editing of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

# References

- 1. Russo, F.; Comi, A. From the analysis of European accident data to safety assessment for planning: The role of good vehicles in urban area. *Eur. Transp. Res. Rev.* **2017**, *9*, 9. [CrossRef]
- 2. ACI. La Statistica ACI-Istat-Sintesi Dello Studio 2017; ACI: Rome, Italy, 2018.
- 3. European Commission. *Annual Accident Report;* European Commission, Directorate General for Transport: Bruxelles, Belgium, 2018.
- 4. Istat. *Rilevazione Degli Incidenti Stradali Con Lesioni a Persone—Periodo di Riferimento: Anno 2016;* Istat: Rome, Italy, 2017.
- 5. Cardamone, A.S.; Eboli, L.; Forciniti, C.; Mazzulla, G. Willingness to use mobile application for smartphone for improving road safety. *Int. J. Inj. Control Saf. Promot.* **2016**, *23*, 155–169. [CrossRef] [PubMed]
- 6. Cardamone, A.S.; Eboli, L.; Forciniti, C.; Mazzulla, G. How usual behaviour can affect perceived drivers' psychological state while driving. *Transport* **2017**, *32*, 13–22. [CrossRef]
- 7. Eboli, L.; Mazzulla, G.; Pungillo, G. How to define the accident risk level of car drivers by combining objective and subjective measures of driving style. *Transp. Res. Part F Traffic Psychol. Behav.* **2017**, *49*, 29–38. [CrossRef]
- 8. Al-Ghamdi, A.S. Using logistic regression to estimate the influence of accident factors on accident severity. *Accid. Anal. Prev.* **2002**, *34*, 729–741. [CrossRef]
- 9. De Oña, J.; López, G.; Mujalli, R.; Calvo, F. Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian networks. *Accid. Anal. Prev.* **2013**, *51*, 1–10. [CrossRef]
- 10. Dissanayake, S. Comparison of severity affecting factors between young and older drivers involved in single vehicle crashes. *IATSS Res.* **2014**, *28*, 48–54. [CrossRef]
- 11. Cardamone, A.S.; Eboli, L.; Mazzulla, G. Drivers' road accident risk perception. A comparison between face-to-face interview and web-based survey, Advances in Transportation Studies. *Int. J.* **2014**, *33*, 59–72.
- De Oña, J.; de Oña, R.; Eboli, L.; Forciniti, C.; Mazzulla, G. How to identify the key factors that affect driver perception of accident risk. A comparison between Italian and Spanish driver behavior. *Accid. Anal. Prev.* 2014, 73, 225–235. [CrossRef]
- 13. Machado-Leon, J.L.; de Oña, J.; de Oña, R.; Eboli, L.; Mazzulla, G. Socio-economic and driving experience factors affecting drivers' perceptions of traffic crash risk. *Transp. Res. Part F Traffic Psychol. Behav.* **2016**, *37*, 41–51. [CrossRef]
- 14. Renner, W.; Anderle, F.G. Venturesomeness and extraversion as correlates of juvenile drivers' traffic violations. *Accid. Anal. Prev.* **2000**, *32*, 673–678. [CrossRef]
- 15. Ayuso, M.; Guillén, M.; Alcañiz, M. The impact of traffic violations on the estimated cost of traffic accidents with victims. *Accid. Anal. Prev.* **2010**, *42*, 709–717. [CrossRef] [PubMed]
- 16. Penmetsa, P.; Pulugurtha, S.S. Risk drivers pose to themselves and other drivers by violating traffic rules. *Traffic Inj. Prev.* **2017**, *18*, 63–69. [CrossRef] [PubMed]
- 17. Shinar, D.; Schechtman, E.; Compton, R. Self-reports of safe driving behaviour in relationship to sex, age, education and income in the US adult driving population. *Accid. Anal. Prev.* **2001**, *33*, 111–116. [CrossRef]
- 18. Fosgerau, M. Speed and income. J. Transp. Econ. Policy 2005, 39, 225–240.
- 19. Laapotti, S.; Keskinen, E.; Hatakka, M.; Hernetkoski, K.; Katila, A.; Peräaho, M.; Salo, I. Driving circumstances and accidents among novice drivers. *Traffic Inj. Prev.* **2006**, *7*, 232–237. [CrossRef]
- 20. Zhang, G.; Yau, K.K.W.; Chen, G. Risk factors associated with traffic violations and accident severity in China. *Accid. Anal. Prev.* **2013**, *59*, 18–25. [CrossRef]
- 21. Lord, D.; Mannering, F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transp. Res. Part A* **2010**, *44*, 291–305. [CrossRef]
- 22. Kim, J.; Ulfarsson, G.F.; Kim, S.; Shankar, V.N. Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. *Accid. Anal. Prev.* **2013**, *50*, 1073–1081. [CrossRef]

- 23. Xiong, Y.; Tobias, J.L.; Mannering, F.L. The analysis of vehicle crash injury-severity data: A Markov switching approach with road-segment heterogeneity. *Transp. Rese. Part B Methodol.* **2014**, *67*, 109–128. [CrossRef]
- 24. Mannering, F.L.; Shankar, V.; Bhat, C.R. Unobserved heterogeneity and the statistical analysis of highway accident data. *Anal. Methods Accid. Res.* **2016**, *11*, 1–16. [CrossRef]
- 25. Adanu, E.K.; Jones, S. Effects of Human-Centered Factors on Crash Injury Severities. J. Adv. Transp. 2017, 1208170. [CrossRef]
- 26. Theofilatos, A.; Yannis, G. Exploring Crash Injury Severity on Urban Motorways by Applying Finite Mixture Models. *Transp. Res. Procedia* **2019**, *41*, 480–487. [CrossRef]
- Yu, H.; Li, Z.; Zhang, G.; Liu, P. A latent class approach for driver injury severity analysis in highway single vehicle crash considering unobserved heterogeneity and temporal influence. *Anal. Methods Accid. Res.* 2019, 24, 100110. [CrossRef]
- 28. Hou, Q.; Huo, X.; Leng, L.; Cheng, Y. Examination of driver injury severity in freeway single-vehicle crashes using a mixed logit model with heterogeneity-in-means. *Phys. A Stat. Mech. Appl.* **2019**, 531, 121760. [CrossRef]
- 29. ITF. *Road Safety in European Cities: Performance Indicators and Governance Solutions;* International Transport Forum Policy Papers, No. 67; OECD Publishing: Paris, France, 2019.
- 30. Fabbris, L. Statistica Multivariata. Analisi Esplorativa Dei Dati; McGraw-Hill Libri Italia: Milan, Italy, 1997.
- 31. Yannis, G.; Papadimitriou, E.; Dupont, E.; Martensen, H. Estimation of Fatality and Injury Risk by Means of In-Depth Fatal Accident Investigation Data. *Traffic Inj. Prev.* **2010**, *11*, 492–502. [CrossRef]
- 32. Hosmer, D.W.; Lemeshow, S. Applied Logistic Regression; Wiley: New York, NY, USA, 2013.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).