International Journal of Data and Network Science 7 (2023). 91–96

Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Relationship between speeding and severe road accidents on the Peruvian Pan American highway

Anderson Edwin Antialon Macias[a*], Deiby Luis Medina Corillocla[a], Marcia Yesenia Jeremias Porras[a] and Johan James Hinostroza Yucra[a]

[a]*Universidad Continental, Peru*

C H R O N I C L E

Article history:
Received: May 20, 2022
Received in revised format: September 26, 2022
Accepted: November 30, 2022
Available online: December 2 2022

Keywords:
Speeding
Road accidents highways
Logit

A B S T R A C T

Road accidents are an essential issue for every country in the world because their consequences are devastating for the people involved in all senses. Speeding has been considered one of the main reasons for these tragedies. Therefore, the current research analyzed the relationship between speeding and severe accidents on the Pan-American highway in Peru. Publicly available data kept in a Peruvian official agency was employed. Methodologically, the logit regression was harnessed since the data were categorical. After the analysis, the research found that it was riskier to have severe accidents than speeding in urban areas crossing the highway and driving or being inside a truck.

1. Introduction

Road accidents have substantial economic, social, and human costs. In the whole world, it is among the ten leading death causes, with more than 3000 deaths daily (World Bank, 2017). Therefore, it is imperative to analyze the causes of road crashes to prevent them. Nonetheless, the simple act of driving is dangerous; it demands that drivers put all their senses on the road (Fisa et al., 2022). Drive defensive behavior is essential for the security of everyone on the road because most accidents are because of human factors (Ho & Gee, 2008). Those negligent factors are driving under the influence of alcohol, speeding, not respecting the spotlights, and the inexistent use of the signal lights (Topol & Drahotský, 2017). Studies like Stanojević et al. (2013) found that these factors are related to most car accidents. It is known that people tend to lie, which might bias the information they get (Lee, 2012). Fortunately, the Peruvian government has made available important information about the occurrence of car accidents on its roads, which facilitates statistical analysis.

Data kept in the Peruvian *Portal de Datos Abiertos* (Superintendencia de Transporte Terrestre de Personas, 2021) is essential for a country with more than four hundred daily accidents (Instituto Nacional de Estadística e Informática, 2016). Lima, the most populated Peruvian city, leads the national car accident ranking. About half of Peru's accidents happen in this city and its surroundings. The high number of cars in its downtown generates the most accidents. Although most accidents happen around or in the main Peruvian towns, the most deadly occur on the roads (Instituto Nacional de Estadística e Informática, 2016).

Peru has a road system far below its needs, generating heavy traffic. Also, how they are designed makes those high-speed roads cross avenues at the level in cities. Those avenues are the place of schools, hospitals, or residential areas. Besides, drivers of informal means of transport do not have any respect for the transit norms. Therefore, it generates a dangerous

* Corresponding author.
E-mail address: 72907474@continental.edu.pe (A. E. Antialon Macias)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2022.12.002

environment for everyone on or around the road (Commission for Latin America and the Caribbean, 2020). It is understandable, then, that the second-place car accidents happen on the streets, but accidents that occur here are the deadliest.

Consequently, how certain is it that speed violation is a cause of car accidents on Peruvian roads? This research aims to understand the relationship between Peru's two most important roads. The Northern Pan-American and the Southern Pan-American.

2. Literature review

Around the world, the speed limit is a debatable issue. On one side, the inhabitants of places where the roads cross at the level claim that the speed limits should be the same as the cities (Richards, 2010). On the other hand, drivers believe the speed limit should be the same as stipulated in the Peruvian norm. Here the speed limit on highways is as follows: 80 kilometers per hour for trucks, 90 kilometers per hour for public transport, and 100 kilometers per hour for little cars. Meanwhile, the speed limit in cities is as follows: 30 kilometers per hour in commercial zones, hospitals, schools, and streets, and 50 kilometers per hour in residential areas and avenues. The penalty fee for speeding is about 250 dollars and half the full sanction for suspending the driver's license. The effectiveness of these measures for preventing car accidents on roads is controversial. According to the Instituto Nacional de Estadística e Informática (2016), road accidents have increased over the years. Hence, this situation makes lawmakers punish speeding more heavily, believing it would reduce car accidents because it is thought to have a direct relationship.

Empirically, Malyshkina & Mannering (2008)analyzed the effect of speed limits on roadway safety in Indiana. Then, he found that the speed limit ranges did not significantly affect car accident severity. Moreover, it was found that sometimes road speed limits were associated with higher accident severity. Alonso et al. (2013) studied the reasons for Spanish drivers' speed behaviorally, employing the preconceived idea that more speed means accidents. The interviewed people claimed that they speeded because they felt that speed limits were low, and it was probable to get caught. Also, they stated that the penalty received when getting caught speeding changed their habits in a short time, but they still believe that speeding is a habit for them.

The idea that more velocity is riskier for car accidents is common around the globe. For instance, Kloeden et al. (1997) found that on roads where the speed limit was 60 kilometers per hour, drivers who surpassed this limit had more chances of being involved in accidents than those who did not. Hence, governments are expected to employ fines for speed exceeding to lower the accident rate. This trend is widespread around the world. An example is a study by Gerald et al. (2022), who found that this measure is believed to be the panacea for reducing car accidents. Is reducing speed limits to prevent accidents enough? Silvano (2016)analyzed the impact of the posted speed limits on roads. Those speed limits reduced the speed limit to ten kilometers per hour. Although it might be thought that lowering the speed limit would change the driver's mind, the study found that it was not the most relevant for changing the driver's attitude. Often, drivers violate traffic norms for a wide variety of reasons. Those reasons are not necessarily deliberate. For instance, biological needs like sleeping or eating can change the driver's attitude and ability (Factor, 2014).

Therefore, it is not necessarily true that fines can reduce by itself the incidence of car accidents. Indeed, they might not work as expected. For instance, Goldenbeld et al. (2011) found that drivers who constantly violate traffic norms had more chances of participating in an accident than those receiving only one fine per year. Hence, it appears the question of how much control there has to be on roads for problematic drivers to reduce accidents. Here, Yannis et al. (2007) encountered that roads with vigilance had lower car accidents than roads that were not supervised. A common control is the usage of speedometers placed in strategic places on the road. In Peru, those machines are put in zones on the Pan-American highway. Also, members of the Sutran, a Peruvian road complementary authority, measure the roads' velocity. It might be expected that this sort of control should harm the occurrence of car accidents on heavy traffic roads like the Pan American. Also, an unexpected benefit of these controls is the information gathered about speeding in the controlled zones. Then, what is the relationship between speeding and road accidents on the Peruvian Pan-American highway?

3. Methodology

The current research employed a transversal dataset that registered car accidents along the Pan-American highway in 2020. Also, another dataset that contained information about speeding information was used. Both were provided by SUTRAN and are available in the Peruvian Open Data Portal. The research chose the Pan-American Highway since it is the country's main road. It crosses the country from its north limit with Ecuador and arrives at the frontier with Chile. In the center, it is placed Lima, which serves as the division of the Peruvian road. Hence, the road from Tumbes, the frontier with Ecuador, to Lima is known as the North Pan-American Highway. Meanwhile, the road from Lima to Tacna, the border with Chile, is known as the South Pan-American road.

The first database contained 67 776 registers from January 1st, 2020, and December 31st, 2020. It is necessary to add that due to the global emergency, the data recorded had a gap between March and September. This gap does not mean the trucks stopped their activities, but the cars do. Also, during this period, speeding was not recorded.

The second database included road accidents on North and South Pan American highways. In total, there were 2904 accidents. Due to those accidents, there were 1 376 deaths and 10 671 injured. Similar to the first report, this one does have a gap due to the global emergency for the Covid 19.

3.1 Variables

There were two dependent variables for this research. The first one was the severity of the road accidents on the North Pan American Highway, and the other one was the road accidents on the South Pan American Highway. The severity of road accidents mixed both death and injured people. The data recorded were classified according to the day of the occurrence. These variables were dichotomized. Hence, days that did not report accidents were assigned 0, while days with road accidents were written with the number 1.

The independent variables were classified according to the speeding recorded. Hence, a variable recorded the number of speeding below 80 kilometers per hour. Another variable recorded the number of speeding above 80 kilometers per hour, another variable recorded the speeding number above 90 kilometers per hour, and a final variable counted the number of speeding cases regarding a limit of 100 kilometers per hour. These four variables were assigned a number from 0 to 2 depending on the occurrences and the day they were recorded. Hence, on days when there were not any road accidents, it was assigned to 0. On days where the number of casualties was below the mean, the number assigned was 1, while on the days' recorded road accidents above the total average received the number 2. This method was taken from Factor (2014).

3.2 Method of analysis

As written previously, the dependent variables employed were dichotomic, while the independent variables were multicategory. Therefore, it was necessary to use logistic regressions because they can test these variables better than alternatives like ordinary least square regressions (Fox, 2013).

3.3 Logistic Regression

The logistic regression assumes that k is the group of independent observations *y1, ..., yk,* and its i-th element can be managed as the realization of the Yi variable with a binomial distribution (Rodriguez, 2007). Hence, $Y_i \sim B(n_i, \pi_i)$. n_i is the binomial denominator, while π_i is the probabilistic one. Besides, the binomial denominator has individual data for every i defining the model's stochastic structure.

The probability logit is a linear predictors function represented by $logit(\pi_i) = x_i'\beta_j$. Here, x_i' is a covariance vector and β is the logit regression coefficient vector. It is necessary to state that this equation is the model structure definition. In a logit regression, the coefficients are logit values, then, β_j is the probability logit change in function of a shift in the j-th predictor while other predictors remain equal (Rodriguez, 2007). After the previous equation is taken into the exponential, the i-th units are given by $\frac{\pi_i}{1-\pi_i} = \exp\{x_i'\beta\}$. This equation defines a multiplicative probabilistic model. Therefore, when the j-th is changed, it is possible to multiply its probabilities with $\exp\{\beta_j\}$, then $\exp\{x_i'\beta\}\exp\{\beta_j\}$. In consequence, the $\exp\{x_i'\beta\}$ represent the probability ratios of the variables.

Finally, when the probability of π_i is calculated, it is necessary to employ derivates. Then, $\frac{d\pi_i}{dx_{ij}} = \beta_j\pi_i(1-\pi_i)$. This final equation represents that the j-th predictor of π_i depends on both β_j and the probability value can provide information about the predictor's probability (Rodriguez, 2007).

Additionally, it is necessary to execute additional analyses like the goodness of fit, sensibility, and specificity to ensure the model's good adaptation (Rodriguez, 2007).

3. Results

Table 1 shows the results of the survey.

Table 1
Descriptive statistics

Independent variables	Range	Mean	Std. Dev.	Dependent variables	Range	Mean	Std. Dev.
Exceed 35 km/h	0-2	64.34	71.57	North Pan American accidents	0-1	0.92	1.15
Exceed 80 km/h	0-2	45.52	36.49				
Exceed 90 km/h	0-2	48.83	39.13				
Exceed 100 km/h	0-2	76.23	67.33				
Exceed 35 km/h	0-2	21.75	42.59	South Pan American accidents	0-1	0.99	1.26
Exceed 80 km/h	0-2	1.35	3.14				
Exceed 90 km/h	0-2	0.51	1.51				
Exceed 100 km/h	0-2	49.09	52.81				

Table 1 shows the variable employed for the current regression. Each dependent variable had four independent variables classified according to the number of speeding they reported. Dataset was ordered by the day of the occurrence. Hence, the data was organized from January 1^{st} to December 31^{st} of 2020. The classification was from 0 to 2. Zero was for the inexistence of reported speeding. Number one was assigned for the days where the report quantity was below the annual average, while number two was for the days that surpassed this average. The dependent variables were classified according to the existence of injured or dead people. One lasts for the days that had injured or dead people due to road accidents, and 0 for days that did not have any casualties.

Table 2
Logit regression for North Pan American highway

Regressors	Odds Ratio	Robust Standard Error	z	p>z	95% confidence interval		chi2	p>chi2
Exceed 35 km/h	1.12	0.31	0.39	0.70	0.65	1.93		
Exceed 80 km/h*	2.45	1.25	1.74	0.08	0.90	6.68		
Exceed 90 km/h	0.82	0.53	-0.31	0.76	0.24	2.87		
Exceed 100 km/h	0.58	0.28	-1.12	0.26	0.22	1.52		
Constant	2.166134	1.06	1.58	0.12	0.83	5.67		
Chi2							25.21	0.00
Pseudo R2	0.09							
GOF							1.72	0.67
Sensitivity	4.51%							
Specificity	95.00%							
Correctly	29.83%							

*significant at 10%

Table 2 shows the logit regression for the first dependent variable. Here, it is possible to see that only the variable that counted the number of speed limits exceeding 80 kilometers per hour and below 90 kilometers per hour was significant for the model. Hence, it had a positive relationship with the dependent variable. Also, Table 2 shows the odds ratio for the independent variables. Therefore, it is possible to state that one additional unit of speeding larger than 80 kilometers per hour and lower than 90 kilometers per hour might increase the number of accidents by 0.27.

Table 3
Logit regression for South Pan American highway

Regressors	Odds Ratio	Robust Standard Error	z	p>z	95% confidence interval		chi2	p>chi2
Exceed 35 km/h**	2.50	1.02	2.23	0.03	1.12	5.58		
Exceed 80 km/h*	0.27	0.18	-2.00	0.05	0.07	0.97		
Exceed 90 km/h	1.10	0.81	0.13	0.90	0.26	4.66		
Exceed 100 km/h	0.90	0.20	-0.47	0.64	0.58	1.40		
Constant	3.16885	1.19	3.08	0.00	1.52	6.60		
Chi2							8.78	0.07
Pseudo R2	0.04							
GOF							6.56	0.16
Sensitivity	60.43%							
Specificity	50.00%							
Correctly	58.01%							

** significant at 5%, *significant at 10%

The chi2 test suggested that the model did not have a total explicative value. The reason is that all the other variables did not have statistical significance. However, the goodness of fit test showed no difference between the expected and observed values, demonstrating the model's predictive power. Additionally, the sensitivity value was 4.51%, the specificity was 95%, and the correctly classified data was only 29.83%. Fig. 1 helps us understand the previous sentence since it portrays the optime point to get as much sensitivity and specificity as possible.

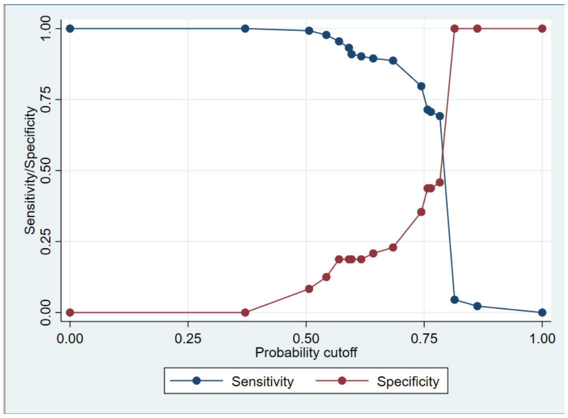

Fig. 1. Cut off value for North Pan American road accidents

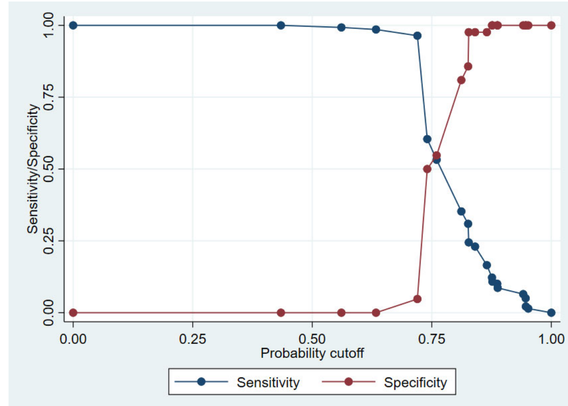

Fig. 2. the optimal value where it was possible to maximize sensitivity and specificity

Table 3 shows the logit regression for the South American highway. As stated before, it employed the same quantity of variables as the previous dependent variable. Therefore, it was observed that speeding beyond 35 kilometers per hour and exceeding 80 kilometers per hour but below 90 kilometers per hour had significant relationships with the dependent variable. In the odds ratio, it can be established that an additional unit of speeding larger than 35 kilometers per hour and shorter than 80 kilometers per hour might increase the number of severe accidents by 2.50. Likely, an additional unit of speeding larger than 80 kilometers per hour and shorter than 90 kilometers per hour might increase the number of severe accidents by 0.27.

The chi2 test showed that the regression model did not have a solid explicative value. The main reason for the previous regression is that the other independent variables did not have statistical significance. Nonetheless, according to the goodness of fit test, the regression had specific predictive power. Hence, it had a sensitivity of 60.43%, a specificity of 50%, and a percentage of correctly classified data of 58.01%. Also, Fig. 2. shows the optimal value where it was possible to maximize sensitivity and specificity.

4. Discussion

The current regression showed that not all speeding was related to severe road accidents on the Pan American highway. Indeed, only surpassing the limit of 35 kilometers per hour as long as it is below 90 kilometers per hour had a significant positive relationship with severe accidents. Therefore, it is surprising that speeding beyond 90 kilometers per hour had no meaningful relationship with road accidents. The possible explanation for this situation might be as follows. The speed limit of 35 kilometers per hour is principally for urban areas. As stated before, the highways in Peru cross at level cities and places with markets, hospitals, schools, and houses. Therefore, to stop accidents, the Peruvian set a speed limit of 35 kilometers per hour in those areas. Then, it can be stated that most accidents happen in those areas—at least the deadliest, which is the reason for this study. Also, in the case of the 80 kilometers per hour speed limit variable, it is necessary to keep in mind that this limit is set only for trucks in places where public transport has a limit of 90 kilometers per hour and cars of 100 kilometers per hour. Therefore, it is possible to claim that exceeding the speed limits is more dangerous for trucks than other transport means. Of course, due to the inconvenient test results to ensure that the model is highly explicative, it cannot be stated that any evidence was not found that surpassing the velocity beyond 90 kilometers per hour increases the risk of severe accidents.

Somewhat the analysis results match the results of Malyshkina & Mannering (2008)and Kloeden et al. (1997) since both found that surpassing certain speed limits can be riskier than other habits at the time of having a car accident. Besides, it is interesting that the speed limit records are easier to get for trucks and public transport than other means of transportation since the Peruvian regulatory agency obligates them to carry a GPS inside their units. However, it is not compulsory for cars, which might generate missing data. The only way to catch them violating the speed limit is through speedometers placed in a bunch of zones on the road. Hence, if the study did not consider this issue, it would be possible to state that it is not adequate for road controls, at least when they detect speeding in places where the limit is 90 kilometers per hour, as stated by Goldenbeld et al. (2011). However, if we considered this, then this research would match the results of Yannis et al. (2007)).

5. Conclusion

After the analysis, the current research could claim that surpassing the speed limit is riskier under two factors: the place of the norm violation and the unit involved. Hence, exceeding the speed limit in urban areas is more dangerous than in plain areas. Also, it is riskier for trucks to surpass these limits than for other transport means. However, the research recognizes that data might not fit well with the regression model, which can make some of these results less complete than desired. As a disclaimer, this research employed all data available by the SUTRAN.

Nonetheless, it seems that other data, such as gender, age, or education level, could be necessary to get better results. It will be desirable for further analysis that this data becomes available. Also, the practical implications of this research are to recommend that all vehicles should have a GPS to track the velocity as is done in trucks and public service units. Besides, it is necessary to avoid highways as the Pan Americans do not cross at level cities or urban areas. Then, it is essential to invest in overpasses to prevent accidents. When it is not possible, it might be helpful to install intelligent spotlights and conscientization programs on the effect of car accidents.

References

Alonso Plá, F. M., Esteban Martínez, C., Calatayud Miñana, C., & Sanmartín, J. (2013). Speed and Road Accidents: Behaviors, Motives, and Assessment of the Effectiveness of Penalties for Speeding. *American Journal of Applied Psychology*, *1*(3), 58–64. https://doi.org/10.12691/ajap-1-3-5

Commission for Latin America and the Caribbean. (2020). *Un análisis de la siniestralidad vial y su relación con la niñez: el caso de Uruguay* (Issue 2007).

Factor, R. (2014). The effect of traffic tickets on road traffic crashes. *Accident Analysis and Prevention*, *64*, 86–91. https://doi.org/10.1016/j.aap.2013.11.010

Fisa, R., Musukuma, M., Sampa, M., Musonda, P., & Young, T. (2022). Effects of interventions for preventing road traffic crashes: an overview of systematic reviews. *BMC Public Health*, *22*(1), 513. https://doi.org/10.1186/s12889-021-12253-y

Fox, J. (2013). Applied Regression Analyis and Generated Linear Models 3rd edition. In *Journal of Chemical Information and Modeling* (Third edit, Vol. 53, Issue 9). SAGE Publications.

Gerald, A., Mbue, I. N., Merlin, A. Z., & Achiri, A. T. (2022). A Systematic Review of Excessive Speed Monitoring and Control System for Accident Prevention on Cameroon Highways. *World Journal of Engineering and Technology*, *10*(01), 28–39. https://doi.org/10.4236/wjet.2022.101002

Goldenbeld, C., Reurings, M. C. B., van Norden, Y., & Stipdonk, H. L. (2011). *Relatie tussen verkeersovertredingen en verkeersongevallen*. https://swov.nl/sites/default/files/publicaties/rapport/r-2011-19.pdf

Ho, R., & Gee, R. Y. (2008). Young men driving dangerously: Development of the Motives for Dangerous Driving Scale (MDDS). *Australian Journal of Psychology*, *60*(2), 91–100. https://doi.org/https://doi.org/10.1080/00049530701452095

Instituto Nacional de Estadística e Informática. (2016). *Analisis de los accidentes de tránsitos ocurridos en el año 2016: Vol. III*.

Kloeden, C. N., McLean, A. J., Moore, V. M., & Ponte, G. (1997). TRAVELLING SPEED AND THE Volume 1 - Findings by. *NHMRC Road Accident Research Unit, The University of Adelaide*, *1*(November), 1–44.

Lee, D. N. (2012). Do Traffic Tickets Reduce Motor Vehicle Accidents? Evidence from a Natural Experiment. *SSRN Electronic Journal*, *December 2011*. https://doi.org/10.2139/ssrn.1939586

Malyshkina, N. V., & Mannering, F. L. (2008). *Analysis of the Effect of Speed Limit Increases on Accident-Injury Severities*. *July*. http://arxiv.org/abs/0806.1354

Richards, D. C. (2010). Relationship between Speed and Risk of Fatal Injury: Pedestrians and Car Occupants. *Road Safety Web Publication*, *16*, 1–41. http://www.dft.gov.uk/pgr/roadsafety/research/rsrr/theme5/researchreport16/doc/rswp16.doc

Rodriguez, G. (2007). Logit Models for Binary Data. In *Lecture Notes on Generalized Linear Models*. Princenton University. https://doi.org/10.1002/ird.2208

Silvano, A. (2016). *Advancing traffic safety: an evaluation of speed limits, vehicle-bicycle interactions, and I2V systems*. *16–008*, 1 file. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-195640%0Ahttps://trid.trb.org/view/1459969

Stanojević, P., Jovanović, D., & Lajunen, T. (2013). Influence of traffic enforcement on the attitudes and behavior of drivers. *Accident; Analysis and Prevention*, *52*, 29–38. https://doi.org/10.1016/j.aap.2012.12.019

Superintendencia de Transporte Terrestre de Personas. (2021). *Portal de Datos Abiertos*. https://www.datosabiertos.gob.pe/

Topol, L., & Drahotský, I. (2017). Dangerous Driver'S Behavior. *Acta Polytechnica CTU Proceedings*, *12*, 112. https://doi.org/10.14311/app.2017.12.0112

World Bank. (2017). *The high toll of traffic injuries: unacceptable and preventable*.

Yannis, G., Papadimitriou, E., & Antoniou, C. (2007). Multilevel modelling for the regional effect of enforcement on road accidents. *Accident Analysis & Prevention*, *39*(4), 818–825. https://doi.org/https://doi.org/10.1016/j.aap.2006.12.004