

Title

The data-driven pattern for healthy behaviors of car drivers based on daily records of Traffic Count data from 2018 to 2019 near airports. A Functional Data Analysis.

Mohammad Fayaz¹, Alireza Abadi^{2,3*}, Soheila Khodakarim, Mohammadreza Hoseini⁵, Alireza Razzaghi⁶

¹ PhD Student of Biostatistics, Department of Biostatistics, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

² Professor of Biostatistics, Department of Community Medicine, Faculty of medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

³ Social Determinants of Health Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

⁴ Associate Professor, School of Allied Medical Sciences, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran

⁵ Master Student of Transportation, Tehran North Branch, Islamic Azad University, Tehran, Iran

⁶ Road Traffic Injury Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

* Corresponding Author: Dr. Alireza Abadi (alirezaabadi@gmail.com)

Professor of Biostatistics, Department of Community Medicine, Faculty of medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Social Determinants of Health Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Koodakyar Ave., Daneshju Blvd., Velenjak, Tehran, IRAN.

Abstract

The road traffic injuries risk factors such as driving offenses and average speed are concerns for health organizations to reduce the number of injuries. Without any comprehensive view of each road, one cannot decide about the effective policy. In this manner, the data-driven policy will help to improve and assess the decisions. The count data near the road of two airports is surveyed for investigating the time-varying speed zones.

The descriptive statistics, ANOVA, and functional data analysis were used. The hourly data of traffic counts for four different locations at the entrance of the two airports, international and domestics, were collected for one the year 2018 to 2019.

The hourly pattern of driving offenses for each road was assessed and the to and from airport roads had different peaks (<0.05). The hour, weekdays, type of airport, direction and their interactions were statistically significant (<0.05) for the chance of driving offenses. The speed average during the day was statistically different (<0.5) by the number of different types of vehicles.

The traffic count data is a great resource for decision making in safe driving subjects such as driving offenses. With functional data analysis, we can analyze them to get the most of the characteristics of this data. The airports are public places with high traffic demand in all countries that yields the different pattern of traffic transportation, therefore we extract the factors that affect the driving offenses. Finally, we conclude that conducting a time-varying speed zone near the airports seems vital.

Keywords

Driving Offenses, Speed Zone, Airports, Functional Data Analysis, Data-Driven Policy,

Introduction

The road traffic deaths which reach 1.35 million each year globally, are the 8th leading cause of death of all ages (2.5% of total deaths) and high-risk driving is one of the most important causes of death and major injuries in the developing countries. (1) The association between vehicle crashing risk and speed traveling seems obvious. (2, 3) The speeding, overtaking from the wrong side, and rapid changing of driving lanes are high risk driving behaviors and they count as the driving offenses which are the most causes of road traffic crashed (RTC) in Tehran. (4) The prevalence of a few risky driving behaviors in Iran was estimated nationally with surveys as follows: 0.5 (0.4 - 0.6) for drink driving and 3.5 (3.2 – 3.8) for being an occupant in a car with a drunk driver. (5) The other risky behavior is the lack of attention, getting trapped in the car, listening to the music, fatigue and sleepiness, duration and distance, and negligence of seatbelt while driving. (6) On the other hand, safety driving is relating to demographic and health status (7, 8). The traffic laws modifications, police controls enhancement, transport infrastructure improvement, and education for drivers were mention as policies for reducing fatal road traffic injuries in Iran. (6, 9) The penalty for risky behavior such as using a cell phone while driving is protective. (10) Therefore, identification of the main risk factors for RTC such as situations, places, behaviors, etc. helps us to control, manage, and prevent it and set health policies.

Many meta-analyses showed that the maintenance of the speed cameras is effective in speed management, reducing fatal and serious injury crashes(11, 12) and safe system (13). There are different speed camera programs in different countries, for example in Iran, the number of speed cameras on rural roads improves the safety of the roads. (14, 15) While speed cameras record many frames at a time interval of data subjects and use big data and AI for automatic decision making (16), their privacy and GDPR is under discussion. (17)

The functional data analysis is a field in statistics which deals with high dimensional data with underlying function, for example, we can consider the speed camera hourly records as a function. (18) The hourly records are high dimensional and using multivariate statistical methods does not cover the underlying functions in these datasets. Therefore, we use functional canonical correlation to study the relationships between two functions and we also use the functional regression to estimate the effects of covariates on the whole functions. (19, 20) Other methods such as multi-level mixed models were used to model the traffic count data. (21)

We can mention controlling the road speed, the hourly travel patterns, and decreasing traffic congestion as a data-driven policy for the sustainable smart city from analyzing the official statistics (22), computer vision (23), and speed cameras data. (24, 25). The implementation of the data-driven policy has three phases include predictive and problem definition, design and experimentation, evaluation, and implementation. (26)

There is some speed limitation in special areas such as school, playground zones (27, 28), residential areas (29), rural (30), and urban areas (31) in many countries (32) for protecting children, cyclists, etc. from traffic collisions. In this study, we first show the temporal pattern of speeding offenses in the airport zone, both domestic and international. They are changing in the special hours. Airports are one of the places which people often want to reach or get out very quickly and therefore we see the occurrence of speeding offenses near these areas. Secondly, the average speed of different types of vehicles is studied as a behavioral factor in airport zones. Many passengers want to reach international airports in a short time, therefore, it seems a high risk driving includes high speed driving appears in the ways to the airport.

Material and Methods

Population and Sample

According to the Iranian Airports Holding Company, Iran has 51 airports. We choose two airports in this study. The International Imam Khomeini Airport, OIIE (IKA), is located outside the capital city of Tehran and has 8,647 domestic and 8,843,585 international passengers. And it is only for international flights. The Isfahan International Airport (Esfahan Shahid Beheshti International), OIFM (IFN), is located

near the Isfahan, one of the most important cities in Iran. And it has 2,258,806 domestic and 525,810 international passengers. And the major flights in this airport are domestic flights. (33)

Dataset

To analyses the pattern of traffic near the airports, the nearest traffic camera beside the highways to and from these airports is selected (fig 1). There is two traffic camera, one is near the *to* the airport and the other is near the *from* airport roads. Then we collect the hourly traffic data to reveal: first is the difference between patterns of to/from airport, second is the difference pattern between an international and domestic airport and third is the difference traffic patterns between weekdays.



Imam Khomeini International Airport		Isfahan Shahid Beheshti International Airport	
			
Code Name	Address (Lat/Lng)	Code Name	Address (Lat/Lng)
To the airport: 113603	(35.400174,51.187047)	To the airport: 213114	(32.718855,51.849681)
From the airport: 113653	(35.399151,51.186704)	From the airport: 213164	(32.718855,51.849681)

Figure 1 – The information about the counter stations.

The traffic data set for Traffic Counts with its characteristics such as latitude, longitude, number of lines, and capacity is available from the Ministry of Roads and Urban Development website. (34) There are 15 columns in each dataset including the road code (the six digits code for go/return), the road name, start time of capturing, end time of capturing, functional time of Traffic Counts (60 minutes in a usual capturing), Total Vehicles, total vehicles type 1 to type 5 (type 1 is cars and pick-ups, type 2 is trucks, small trucks, and minibusses, type 3 is Ordinary lorries less than 4m and three axles, type 4 is the bus, and type 5 is Trailers and freighters higher than three axles), Average speed (calculated from all type of vehicles), number of speeding offenses (the number of the vehicle in a time interval on the road which has higher speed than speed limits) and number of failing to follow a vehicle at a safe distance (at least less than two seconds distances with the previous vehicle. These datasets are available hourly and daily for each month with precision at least between 90% to 95%.

Data Quality and Data Cleaning

The hourly dataset for each Traffic Count is captured for one year, from 2018-10-23 to 2019-10-22. The data for 2019-07-01 to 2019-07-31 is not available. There is total 31,282 observation (15,714 for Tehran and 15,568 for Isfahan), but about 9% of the observation is excluded because the counter station is capturing less than 60 minutes in an hour and only 29,502 observation (15,307 for Tehran and 14,195 for Isfahan) is used in this analysis. There are also some missing values among datasets then we compute them with the stations-weekdays-hours specific mean.

Indices Defining

We created two new indices: First is total vehicles per minute by dividing the total vehicles to functional time of counter stations which its duration is 60 minutes. Therefore, the records with different functional time against 60 minutes are omitted. Second is the total driving offenses per minute by summing all two type offenses and dividing them to 60 minutes.

Then we calculate the occurrence probability of any offenses per minute, $P(x)$, by dividing total driving offenses per minute to total vehicles. Then we transform the probability to the logit function as $\log(P(x)/(1-P(X)))$. Its interpretation is simple and the boundaries vary from negative infinity to positive

infinity. Therefore, the odds ratio of the occurrence of an offense per minute is calculated for each hour in each station in each day of the week.

The analysis data structure consists of a functional offenses logit during 0-23 hour, the weekday's replication, and the counter stations.

Statistical Data Analysis

Descriptive Statistics

We calculate the descriptive statistics including the mean and standard deviation of total vehicles and total offenses per hour for stations and directions separately.

Full Factorial Analysis of Variance

We use a three-way analysis of variance (ANOVA) to investigate the effects of City, Direct, and Days on the chance of offenses. The `anova()` function from R (35) is used. The adjusted stations-weekdays-hours mean is used for imputation.

Functional Data Analysis

The functional data analysis (FDA) is dealing with high-dimensional data that are recorded on continuous domains such as time, space, etc. The hourly pattern of the chance of offenses is functional data. Therefore, we use the B-Spline basis function to estimate the functions. The number of basis and regularization parameter is obtained with Generalized Cross-Validation methods. (36)

Functional Canonical Correlation

We calculate the relationship between two random functions X and Y with the sample functional canonical correlation ($U = \langle \mathbf{u}, X \rangle = \int_{I_1} u^T(s)X(s)ds$ and $V = \langle \mathbf{v}, Y \rangle = \int_{I_2} v^T(t)Y(t)dt$, are functional canonical variables, respectively). In the following equation, $\mathbf{u}(s) = \Phi_1(s)\mathbf{u}$ and $\mathbf{v}(t) = \Phi_2(t)\mathbf{v}$, where $\mathbf{u} \in \mathbb{R}^{K_1+p}$ and $\mathbf{v} \in \mathbb{R}^{K_2+q}$, \mathbf{u} and \mathbf{v} are weight functions and are chosen to maximize the sample correlation coefficient. (37, 38)

Function-on-Scalar Regression

The function-on-Scalar regression (FOSR) is a class of FDA methods when the response is the function and the covariates are vectors. We use fully Bayesian FOSR (39), the formula is in the supplement part 1.1.1, for high dimensional data because of computational issues, regularization and choosing the factors, and calculating the importance of each factor.

Generalized Additive Models

The Generalized Additive Models (GAM) are used for considering non-linear covariate effects with cubic regression spline. (40)

Results

The Descriptive Statistics

The descriptive statistics including the mean and standard deviation of total vehicles and total offenses per hour, airports, and direction is presented in table 1.

Table – 1 - The mean (SD) of Total Vehicles and Total Offences per hour by airports and directions

Hours	Mean (SD) of Total Vehicles (By Hour)				Mean (SD) of Total Offences (By Hour)			
	To IKA	From IKA	To IKA	From IKA	To IKA	From IKA	To IKA	From IKA
0	1114.72 (234.02)	769.6 (162.87)	272.18 (70.11)	347.98 (140.54)	120.93 (75.27)	109.85 (53.94)	41.97 (22.18)	56.2 (42.26)
1	940.73 (233.54)	718.2 (147.53)	191.63 (53.3)	251.12 (103.61)	100.43 (63.14)	98.93 (57.07)	26.52 (16.04)	37.02 (28.9)

2	760.88 (175.86)	646.71 (147.44)	153.58 (43.31)	174.21 (61.44)	80.65 (48.2)	87.54 (54.88)	19.76 (10.82)	20.64 (14.68)
3	732.48 (130.63)	561.01 (143.79)	135.82 (39.55)	159.79 (75.92)	81.11 (46.14)	75.96 (50.85)	18.69 (11.28)	19.47 (16.04)
4	764.47 (135.63)	664.74 (152.64)	197.01 (63.47)	150.45 (49.73)	90.48 (48.16)	87.6 (34.89)	31.32 (18.17)	16.07 (16.27)
5	731.12 (127.22)	774.66 (162.66)	347.87 (101.01)	174.56 (54.65)	108.48 (55.97)	119.67 (49.41)	57.08 (32.81)	17.82 (16.49)
6	779.72 (137.09)	1132.57 (356.2)	855.21 (282.83)	397.59 (125.36)	159.9 (79.03)	307.75 (153.02)	242.96 (150.39)	53.73 (30.49)
7	1154.38 (216.89)	1571.31 (464.97)	1597.14 (492.26)	665.17 (209.42)	246.3 (95.35)	561.83 (249.35)	617.76 (324.49)	114.15 (61.93)
8	1185.15 (273.83)	1513.67 (341.42)	1156.13 (280.87)	738.3 (224.65)	288.85 (113.84)	505.1 (215.99)	366.39 (170.96)	137.96 (71.33)
9	1088.45 (229.44)	1209.47 (273.24)	1066.9 (177.1)	680.17 (179.45)	258.13 (103.71)	388.91 (171.77)	297.46 (105.36)	119.61 (55.21)
10	1059.61 (167.15)	1067.78 (187.47)	1150.45 (169.75)	830.7 (205.68)	246.31 (93.61)	320.47 (112.42)	325.02 (101.05)	170.22 (78.37)
11	1078.32 (140.73)	1045.19 (157.45)	1046.37 (158.05)	925.61 (184.63)	256.37 (91.78)	293.67 (109.32)	276.02 (78.51)	200.28 (80.8)
12	1128.4 (142.57)	971.79 (135.86)	954.26 (154.11)	891.66 (181.84)	285.24 (98.2)	259.95 (92.52)	238.01 (74.27)	198.18 (74.7)
13	1181.31 (180.49)	874.98 (129.74)	930.05 (175.21)	913.68 (211.05)	303.07 (112.97)	230.87 (89.66)	227.87 (70.37)	206.34 (86.71)
14	1282.29 (216.51)	956.27 (163.76)	887.88 (205.11)	923.85 (231.92)	321.22 (120.17)	242.24 (92.05)	211.8 (77.48)	212.88 (89.64)
15	1413.07 (215.46)	1196.63 (175.78)	943.16 (248.29)	850.48 (230.8)	344.91 (148.81)	283.87 (98.73)	230.07 (103.54)	185.55 (80.43)
16	1516.02 (204.47)	1195.87 (175.2)	937.04 (198.69)	1092.73 (307.6)	365.52 (163.05)	289.37 (106.4)	229.12 (96.22)	285.72 (129.59)
17	1761.67 (205.92)	1165.87 (166.83)	935.03 (207.66)	1274.91 (362.25)	419.18 (189.83)	306.28 (108.38)	235.22 (125.21)	358.11 (162.05)
18	1915.51 (215.9)	1152.91 (188.61)	883.37 (228.84)	1149.54 (253.11)	403.69 (204.58)	254.95 (94.7)	223.68 (155.03)	297.39 (127.48)
19	1872.83 (219.84)	1148.33 (208.53)	739.59 (161.96)	993.47 (254.34)	370.55 (190.11)	222.95 (81.63)	168.11 (98.62)	234.17 (117.65)
20	1499.5 (244.98)	925.73 (229.04)	610.47 (142.62)	772.99 (263.55)	204.88 (117.09)	147.9 (79.19)	125.86 (62.93)	162.71 (129.95)
21	1218.82 (218.76)	759.88 (191.26)	580.24 (149.19)	570.4 (177.53)	148.81 (77.38)	116.76 (64.35)	116.3 (60.04)	96.88 (58.16)
22	1100.94 (203.76)	732.94 (179.31)	453.42 (121.25)	561.26 (175.56)	132.16 (69.97)	109.69 (54.69)	83.61 (47.24)	95.26 (55.29)
23	1156 (208.8)	800.07 (214.19)	339.79 (87)	414.96 (158.85)	130.67 (76.39)	113.54 (57.97)	54.54 (30.47)	67.92 (49.55)
AV								
G	1184.17	980.06	717.10	662.58	227.08	229.9	184.03	140.3
SD	387.75	337.34	428.53	389.01	155.75	169.3	175.17	124.84
Total	9,034,055	7,524,896	5,137,256	4,658,663	1,732,418	1,765,173	1,318,413	986,472

The “To IKA” road has the highest total vehicle transport rate at 18:00 with 1915.51 (215.19) and it has also other peak busy hours at 8:00 with 1185.15 (273.83). The total vehicle traffic rate on average is 1184.17 with 387.75 standard deviations. In this road, the total offenses have two peaks at 17:00 and 8:00 with 419.18 (189.83) and 288.85 (113.84) respectively. And on average, the total offenses are 227.08 with 155.75 standard deviations. Totally in this period with considering data cleaning, there are 9,034,055 total vehicles with 1,732,418 total offenses. The other directions and airports have different patterns.

There are two types of driving offenses: speeding and failing to follow a vehicle at a safe distance. According to our analysis, the speeding offenses probability for an international airport in the hour/direction/weekday is higher than the domestic airport. In contrast, the probability of failing to follow a vehicle at a safe distance for the domestic airport in each hour/direction/weekday is higher than an international airport. (Supplement Part 3)

The Full Factorial ANOVA

The full-factorial-three-way ANOVA was used to assess the effects of city, directions, hours, and all combination effects with aov() function in R. The dependent variable is the average of offenses occurrence chance in an hour. (Table 2)

Table 2 – The Analysis of Variance table

	Df	Sum of Squares	Mean sq	F-Value	P-Value
City	1	2	2.2	4.844	0.03*

Direct	1	1	0.8	1.668	0.20
Days	6	11	1.8	4.038	0.00*
City:Direct	1	443	442.8	970.663	<0.00*
City:Days	6	3	0.5	1.201	0.30
Direct:Days	6	13	2.2	4.778	<0.00*
City:Direct:Days	6	3	0.5	1.034	0.40
Residuals	29351	13388	0.5		

According to this table, the main effects of City and Weekdays and the interaction effects City: Direct and Direct: Days are statistically significant (P-Value < 0.05). Therefore, the adjusted average by City, Direct, Days, and month was used for imputing missing values in the dataset. Then the adjusted average by city, direct, and days was used for imputing the remaining missing values.

The Functional Canonical Correlation

The pattern of the offenses chance per hour into the airport against the from the airport is investigated by functional canonical correlation. Therefore, we transform the data to B-Spline with 26 number of basis with order 6 by fda packages functions. The generalized cross-validation (GCV) method was used to find the smoothing parameter, lambda. Then we have six pairs for studying the relationship, we only describe the two most important of them here. (Figure 2)

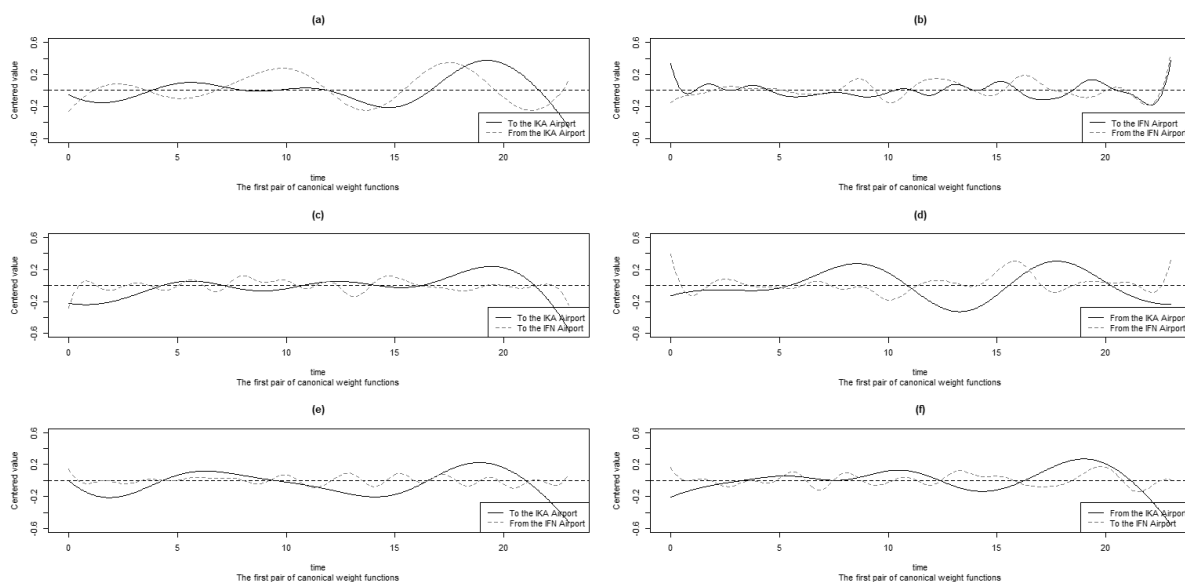


Figure 2 – The Canonical weight functions in six situations.

The canonical correlation coefficient is estimated at 0.7 and a pair of canonical weight functions are plotted in six situations. We observe that the weight pattern of “To the IKA airport” against “From the IKA airport” is different. In the “To the IKA airport”, there is a bump from 19 to 20 but in the “From the IKA airport” there are two bumps, one is near 10 and the other is near 18 and 19.

But we observe that the weight pattern of “To the IFN airport” against “From the IFN airport” have slightly different.

The chance of offenses per hour pattern is different both in directions and airport City. To investigate these factors, we use Bayesian function-on-scalar regression.

The Function-on-Scalar Regression

To understand the effect of the airport city, directions, and weekend to the chance of the driving offenses, we use Bayesian Function-On-Scalar regression with fofsr R package. We also like to study

the interaction effects of these factors. Therefore, we use a 23 basis function to model the response curve. It took about 16 minutes to run on a regular laptop.

All variables are all significant (with a p-value of less than 0.05). And the most important variable is the interaction between station and direction, after that, the station is in the second and in the third is the direction and the interaction between station and weekend. (Table 3)

Table 3 – The global Bayesian P-Value with DSS model selection for the chance of driving offences

	Variable Names	Global Bayesian P-Value	Suggested Model by DSS					
			1	2	3	4	5	6
Intercept	Intercept	<0.000	Yes	Yes	Yes	Yes	Yes	Yes
Main Effects	Station: IFN	<0.000	No	Yes	Yes	Yes	Yes	Yes
	Direction: From	<0.000	No	No	Yes	Yes	Yes	Yes
	Weekend: is	<0.000	No	No	No	No	Yes	Yes
Interaction Effect	Station: IFN and Direction: From	<0.000	Yes	Yes	Yes	Yes	Yes	Yes
	Station: IFN and Weekend: is	<0.000	No	No	Yes	Yes	Yes	Yes
	Direction: From and Weekend: is	0.001	No	No	No	No	No	Yes
	Station: IFN and Direction: From and Weekend:is	<0.000	No	No	No	Yes	Yes	Yes

The estimated coefficient functions and 95% simultaneous credible bands for the proposed regression model are presented in the supplement. The station effect, with IKA as a baseline, has a positive effect from 6:00 to 10:00 and from 20:00 to 24:00 which are the busy times in the roads. The direction effect, with To as a baseline, has the largest effect between 6:00 to 12:00. The pattern is changed in the weekend, with non-weekend as a baseline, from 7:00 to 19:00 has positive effect and in non-weekend has negative effect with the most negative effect at 12:00.

There is three estimated two-way interaction functional effect. The first one is the interaction between the Station and Direction, which is the most important effect. The interaction between station and weekend has a negative effect from 5:00 to 17:00 and in the other hours have a positive effect. The interaction effect between Direction and weekend has a negative effect from 5:00 to 10:00.

And finally, the three-way interaction effect between station, direction, and the weekend is almost positive from 1:00 to 12:00, with the highest at 7:00, and at the remaining hours are negative, with the biggest value at 17:00.

The Behavioral Factor

The Descriptive Statistics

The average and standard deviation per hour for speed are calculated for all types of vehicles. The total vehicles with the maximum number of them from type 1 to type 5 per hour are presented in table 4.

Table 4: The descriptive statistics for average speed (Standard Deviations) and total number of vehicle types (Maximum number of vehicles) per hour.

Hour	AVG Speed (SD)	Type 1 (Max)	Type 2 (Max)	Type 3 (Max)	Type 4 (Max)	Type 5 (Max)
0	88 (4.9)	696,999 (2,212)	26,050 (125)	15,347 (69)	6,786 (43)	30,298 (94)
1	87.4 (5.4)	590,726 (1,809)	21,785 (86)	13,218 (65)	5,750 (42)	23,083 (77)
2	87 (5.3)	485,820 (1,254)	17,814 (73)	10,684 (43)	5,107 (24)	19,260 (65)
3	87.6 (5.6)	447,255 (1,329)	16,756 (59)	8,479 (49)	4,798 (31)	15,667 (59)
4	88.1 (5.8)	508,121 (1,127)	20,682 (94)	8,334 (36)	4,984 (29)	13,973 (54)
5	88.4 (6.4)	561,290 (1,209)	35,503 (178)	9,388 (36)	6,911 (33)	15,038 (58)
6	92.1 (7)	804,477 (1,744)	71,017 (175)	18,431 (74)	10,918 (79)	20,851 (110)
7	92.9 (7.6)	1,272,667 (2,282)	140,788 (646)	40,800 (109)	22,384 (102)	41,557 (166)
8	94 (8.3)	1,148,646 (2,085)	127,899 (400)	43,456 (117)	23,487 (109)	52,267 (190)
9	93.7 (8.3)	997,090 (2,070)	102,950 (345)	40,075 (131)	23,652 (91)	55,762 (148)
10	93.3 (8.5)	996,792 (2,215)	98,269 (350)	42,652 (184)	22,228 (118)	61,937 (204)
11	92.9 (8.5)	1,004,407 (1,960)	97,184 (332)	45,059 (231)	23,109 (130)	66,211 (241)
12	92.8 (8.1)	959,089 (1,601)	97,494 (318)	44,397 (202)	22,037 (120)	65,593 (258)
13	92.9 (7.7)	937,819 (1,571)	108,962 (303)	41,488 (150)	20,998 (109)	62,528 (203)
14	92.8 (7.4)	954,706 (2,110)	128,211 (400)	40,569 (157)	21,326 (136)	60,393 (220)
15	92.3 (7.1)	1,046,477 (1,963)	142,414 (552)	45,633 (148)	24,336 (102)	66,272 (204)
16	92.3 (6.7)	1,159,476 (2,147)	144,090 (664)	51,216 (298)	24,989 (137)	67,344 (233)
17	92 (6.9)	1,306,804 (2,204)	123,523 (539)	43,895 (239)	22,635 (162)	64,183 (227)
18	90.3 (5.9)	1,349,407 (2,547)	91,137 (493)	40,353 (193)	19,623 (124)	62,647 (161)
19	89.9 (5.7)	1,287,400 (2,322)	69,465 (257)	36,845 (128)	16,047 (126)	58,482 (176)
20	88.9 (4.9)	1,045,859 (2,312)	42,589 (190)	25,713 (152)	11,561 (116)	49,774 (136)
21	88.6 (4.7)	853,277 (1,949)	35,963 (176)	20,870 (145)	9,512 (69)	45,655 (133)
22	88.5 (4.8)	776,457 (2,053)	34,606 (254)	19,847 (95)	9,046 (62)	44,061 (128)
23	88.2 (4.9)	745,475 (2,041)	27,906 (225)	16,798 (93)	7,244 (79)	36,322 (122)

The average speed is the highest at 8:00 with average speed 94 and a standard deviation of 8.3. The total number of vehicles for type 1,2,3,4 and 5 has two peaks at 8:00 and 20:00, at 8:00 and 17:00, at 11:00 and 17:00, at 8:00 and 16:00, and finally at 11:00 and 16:00, respectively. (The further analysis is done in part 2 of the supplement)

Two Type of Driving Offenses

There are two types of driving offenses, and we study each of them with GAM to estimate the nonlinear effects with cubic regression splines and linear effects with Restricted Maximum Likelihood (REML). The penalty based model selection is used by setting `select = TRUE` in the `gam()` function. The AIC, BIC, adjusted R-square, and percent of deviance explained were used to select the best model. The degree of cubic regression spline was determined with an effective degree of freedom.

Speeding

The probability of speeding offenses is near zero under 90 Km/h average speed for all airports and directions, except in the international airport and “from” direction. On the other hand, the estimated coefficients are not zero for 90-120 Km/h average speed in all hours, supplementary part 3.1.

Failing to follow a vehicle at a safe distance

Only the domestic airport has positive effects in failing to follow a vehicle at a safe distance, supplementary part 3.2.

Discussion

The average hourly traffic volume near the airport in 2018-2019 is compared with 4 different countries. The count traffic data of the Cork and Shanon airports in Ireland, the Heathrow and Gatwick airports in the UK, the Hartsfield–Jackson Atlanta International Airport in the USA, IKA, and IFN airports in Iran are gathered and average on all traffic counts in airport zones. (41-43) The figure in part 4 of the supplementary states that there are two peaks, first is between 5 to 10 am and the second is between 15 to 20 and the same pattern is observed among 4 countries. But this pattern is different based on the road direction, airport type (international vs domestic), and weekdays. The full list of traffic sites with their information is available in the supporting file. The chance of traffic offenses will increase with increasing the average hourly volume.

The rates of the road traffic death per 100,000 population in the world are 18.2%, with the highest rate in Africa equals to 26.6 and the lowest rate in Europe equals 9.3 and in Eastern Mediterranean is 18%. According to the Global Status Report on Road Safety, about 132 countries have funded national road safety goals, targets, and strategies. And 109 countries have also targets for the reduction of road traffic deaths. There are 12 global road safety performance targets for 2020 and 2030. Among them, target number 6 states that “By 2030, halve the proportion of vehicles traveling over the posted speed limit and achieve a reduction in speed-related injuries and fatalities”. The key risk factors to prevent road traffic deaths are speed, drink-driving, motorcycle helmet use, use of seat-belts, and child restraint systems. Only less than 10 country has laws for all of these risk factors. The number of countries with speed law is 46, with Drink-driving law is 45, with Seat-belt law is 105, and with Helmet law is 49 and with Child restrains law is 33. The speed management decreases the number of fatalities, serious injuries, and death in traveling vehicles. There are three best practices criteria to assess legislation on speed laws including the existence of a national speed limit law, speed limits not exceeding 50 km/h in the urban area, and the power of modification speed limits by local authorities. According to the Safe road users, Iran has a national speed limit, National drink-driving, National motorcycle helmet, National seat-belt, National law on mobile phone use while driving, and National drug-driving laws. The National speed limit law stated that the maximum urban speed limit is 60 km/h (which is higher than 50 km/h), the maximum rural speed limit is 95 km/h, and the maximum motorway speed limit is 120 km/h. (1, 44)

This study shows that the chance of driving offenses in airport roads is not the same during all day and has some different peaks, and this pattern is different for the international and domestic airport, and at

the end, the direction to the airport against from the airport is also different with some different peaks. Therefore, the chance of the driving offenses in the roads to the airports vary according to temporal effects, the time of day and day of the week, and type of Airport effect, domestic or international and direction effect, to the airport or from the airport.

In addition to the mentioned effects in the driving offenses, we also study the behavioral effect, the average speed, and the vehicle types. The hourly pattern of the average speed does not differ widely because we consider the average, but each type of vehicle has different behaviors. Finally, the other behavioral risk factors such as a driver's status, skills, etc. are not considered in this study.

This study suggests making data-driven policies under modification speed limits by local authorities for managing, controlling, and preventing driving offenses, especially in a specific area like airports. Meanwhile, the privacy of data subjects, vehicles, and driver behaviors, must be considered while their data are used, GDPR, etc. In the near future, the need for such policies is vital with wide uses of self-control and automatic vehicles, which communicate with each other and police station (45) and intelligent speed adaptation (ISA) (46)

Conclusions

The data-driven policy under modification speed limits by local authorities will help for controlling, managing, and preventing driving offenses as a high-risk behavioral health index. The time of day, day of the week, international or domestic airport, to the airport or from the airport and type of vehicles are factors that influence this policy.

Appendices

The supplementary includes two files: the further analysis is produced in the document file and the other is R codes with the main dataset for reproducing the result.

References

1. Organization WH. Global status report on road safety 2018 (2018). Geneva (Switzerland): WHO. 2019.
2. McLean J. Reflections on speed control from a public health perspective. *Journal of the Australasian College of Road Safety*. 2012;23(3):51.
3. Farmer CM. Relationship of traffic fatality rates to maximum state speed limits. *Traffic injury prevention*. 2017;18(4):375-80.
4. Shams M, Rahimi-Movaghar V. Risky driving behaviors in Tehran, Iran. *Traffic injury prevention*. 2009;10(1):91-4.
5. Fathollahi S, Saeedi Moghaddam S, Rezaei N, Jafari A, Peykari N, Haghshenas R, et al. Prevalence of behavioural risk factors for road-traffic injuries among the Iranian population: findings from STEPs 2016. *International journal of epidemiology*. 2019;48(4):1187-96.
6. Sadeghi-Bazargani H, Ayubi E, Azami-Aghdash S, Abedi L, Zemestani A, Amanati L, et al. Epidemiological patterns of road traffic crashes during the last two decades in Iran: a review of the literature from 1996 to 2014. *Archives of trauma research*. 2016;5(3).
7. Shinar D, Schechtman E, Compton R. Self-reports of safe driving behaviors in relationship to sex, age, education and income in the US adult driving population. *Accident Analysis & Prevention*. 2001;33(1):111-6.
8. Shinar D, Schechtman E, Compton R. Trends in safe driving behaviors and in relation to trends in health maintenance behaviors in the USA: 1985–1995. *Accident Analysis & Prevention*. 1999;31(5):497-503.
9. Khorasani-Zavareh D, Mohammadi R, Khankeh HR, Laflamme L, Bikmoradi A, Haglund BJ. The requirements and challenges in preventing of road traffic injury in Iran. A qualitative study. *BMC public health*. 2009;9(1):486.

10. Khadem-Rezaiyan M, Moallem SR, Vakili V. High-risk behaviors while driving: A population-based study from Iran. *Traffic injury prevention*. 2017;18(3):257-61.
11. Pilkington P, Kinra S. Effectiveness of speed cameras in preventing road traffic collisions and related casualties: systematic review. *Bmj*. 2005;330(7487):331-4.
12. Wilson C, Willis C, Hendrikz JK, Le Brocq R, Bellamy N. Speed cameras for the prevention of road traffic injuries and deaths. *Cochrane database of systematic reviews*. 2010(11).
13. Doecke SD, Kloeden CN, Dutschke JK, Baldock MR. Safe speed limits for a safe system: The relationship between speed limit and fatal crash rate for different crash types. *Traffic injury prevention*. 2018;19(4):404-8.
14. Delaney A, Ward H, Cameron M, Williams AF. Controversies and speed cameras: lessons learnt internationally. *Journal of public health policy*. 2005;26(4):404-15.
15. Besharati MM, Kashani AT, Li Z, Washington S, Prato CG. A bivariate random effects spatial model of traffic fatalities and injuries across provinces of Iran. *Accident Analysis & Prevention*. 2020;136:105394.
16. Chen T, Sze N, Saxena S, Pinjari AR, Bhat CR, Bai L. Evaluation of penalty and enforcement strategies to combat speeding offences among professional drivers: a Hong Kong stated preference experiment. *Accident Analysis & Prevention*. 2020;135:105366.
17. Mitrou L. Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR)'Artificial Intelligence-Proof'? *Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR)'Artificial Intelligence-Proof*. 2018.
18. Ramsay JO. Functional data analysis. *Encyclopedia of Statistical Sciences*. 2004;4.
19. Horváth L, Kokoszka P. Inference for functional data with applications: Springer Science & Business Media; 2012.
20. Kokoszka P, Reimherr M. Introduction to functional data analysis: Chapman and Hall/CRC; 2017.
21. Kwayu KM, Kwigizile V, Oh J-S. Assessing the safety impacts of raising the speed limit on Michigan freeways using the multilevel mixed-effects negative binomial model. *Traffic injury prevention*. 2020:1-6.
22. Radermacher WJ. Governing-by-the-numbers/statistical governance reflections on the future of official statistics in a digital and globalised society. *Statistical Journal of the IAOS*. (Preprint):1-18.
23. Ibrahim MR, Haworth J, Cheng T. Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities*. 2020;96:102481.
24. Bibri SE, editor *Advances in smart sustainable urbanism: data-driven and data-intensive scientific approaches to wicked problems*. Proceedings of the 4th International Conference on Smart City Applications; 2019.
25. Bibri SE. The anatomy of the data-driven smart sustainable city: instrumentation, datafication, computerization and related applications. *Journal of Big Data*. 2019;6(1):59.
26. van Veenstra AF, Kotterink B, editors. *Data-driven policy making: The policy lab approach*. International Conference on Electronic Participation; 2017: Springer.
27. Kattan L, Tay R, Acharjee S. Managing speed at school and playground zones. *Accident Analysis & Prevention*. 2011;43(5):1887-91.
28. Gregory B, Irwin JD, Faulks IJ, Chekaluk E. Speeding in school zones: violation or lapse in prospective memory? *Journal of experimental psychology: applied*. 2014;20(3):191.
29. Fildes B, Lawrence B, Oxley J. Low speed limits in residential areas in Melbourne, Australia. *Traffic injury prevention*. 2019;20(sup2):S155-S7.
30. Patterson TL, Frith WJ, Povey LJ, Keall MD. The effect of increasing rural interstate speed limits in the United States. *Traffic Injury Prevention*. 2002;3(4):316-20.
31. Isaksson-Hellman I, Töreki J. The effect of speed limit reductions in urban areas on cyclists' injuries in collisions with cars. *Traffic injury prevention*. 2019;20(sup3):39-44.
32. Vadeby A, Forsman Å. Evaluation of new speed limits in Sweden: A sample survey. *Traffic injury prevention*. 2014;15(8):778-85.

33. Iranian Airports Holding Company (Feb 2019).
34. Islamic Republic of Iran Road Maintenance & Transportation Organization.
35. Team RC. R: A Language and Environment for Statistical Computing.
36. Febrero Bande M, Oviedo de la Fuente M. Statistical computing in functional data analysis: The R package fda. usc. 2012.
37. Hooker JORaHWaSGaG. fda: Functional Data Analysis. 2018.
38. Górecki T, Krzyśko M, Wołyński W. Independence test and canonical correlation analysis based on the alignment between kernel matrices for multivariate functional data. *Artificial Intelligence Review*. 2020;53(1):475-99.
39. Kowal DR, Bourgeois DC. Bayesian Function-on-Scalars Regression for High-Dimensional Data. *Journal of Computational and Graphical Statistics*. 2020(just-accepted):1-26.
40. Harezlak J, Ruppert D, Wand MP. Semiparametric regression with R: Springer; 2018.
41. Ireland TII Traffic Data Site Website.
42. UK Road traffic statistics website.
43. The USA ,the Georgia Department of Transportation's Traffic Analysis and Data Application Website.
44. The Vision Zero Chicago 2017.
45. Teoh ER, Kidd DG. Rage against the machine? Google's self-driving cars versus human drivers. *Journal of safety research*. 2017;63:57-60.
46. Matsuo K, Sugihara M, Yamazaki M, Mimura Y, Yang J, Kanno K, et al. Hierarchical Bayesian modeling to evaluate the impacts of intelligent speed adaptation considering individuals' usual speeding tendencies: A correlated random parameters approach. *Analytic Methods in Accident Research*. 2020:100125.