Güvenlik Bilimleri Dergisi / Journal of Security Sciences

Jandarma ve Sahil Güvenlik Akademisi
Güvenlik Bilimleri Enstitüsü
Güvenlik Bilimleri Dergisi, Kasım 2021,Cilt:10, Sayı:2, 447-466
doi:10.28956/gbd.1028677

Gendarmerie and Coast Guard Academy
Institute of Security Sciences
Journal of Security Sciences, November 2021,Volume:10, Issue:2, 447-466
doi:10.28956/gbd.1028677

Makale Türü ve Başlığı / Article Type and Title

Araştırma / Review Article

Investigating hhe Periodic Structure of Traffic Accidents in Turkey

Türkiye'deki Trafik Kazalarının Periyodik Yapısının Araştırılması

Yazar(lar) / Writer(s)

1-Yılmaz AKDİ. Prof., Ankara University, Ankara/TURKEY, yilmaz.akdi@ankara.edu.tr, ORCİD.:0000-0003-0188-0970
2-Yunus Emre KARAMANOĞLU. Ph.D., Gendarmerie and Coast Guard Academy, Ankara/TURKEY, eyunus@bilkent.edu.tr, ORCİD.: 0000-0001-9711-6867
3-Kamil Demirberk ÜNLÜ Ph.D., Atilim University, Ankara/TURKEY, demirberk.unlu@atilim.edu.tr, ORCİD.:0000-0002-2393-6691
4-Cem BAŞ Ph.D., Turkish Statistical Institute, Ankara/TURKEY, cem.bas@tuik.gov.tr, ORCİD.:0000-0003-4488-2562

Bilgilendirme / Acknowledgement:

-Yazarlar aşağıdaki bilgilendirmeleri yapmaktadırlar:
-Makalemizde etik kurulu izni ve/veya yasal/özel izin alınmasını gerektiren bir durum yoktur.
-Bu makalede araştırma ve yayın etiğine uyulmuştur.

Bu makale Turnitin tarafından kontrol edilmiştir.
This article was checked by Turnitin.

Makale Geliş Tarihi / First Received : 05.05.2021
Makale Kabul Tarihi / Accepted: 02.09.2021

Atıf Bilgisi / Citation:

Akdi Y., Karamanoğlu Y.E., Ünlü K.D., ve Baş C. (2021). Investigating the Periodic Structure of Traffic Accidents in Turkey. *Güvenlik Bilimleri Dergisi, 10(2), ss.447-466, doi:10.28956/gbd.1028677*

INVESTIGATING THE PERIODIC STRUCTURE OF TRAFFIC ACCIDENTS IN TURKEY

Abstract

In this study, a time-series analysis is applied to the daily accidents between the periods January 1, 2019 and December 31, 2019 of Turkey. The most important feature of the data used in the study is the official daily traffic accident records kept by the law enforcement units. Regarding these data, the most appropriate time series model is determined and it is examined whether there are periodic components in traffic accidents or not. It is observed that the data is first-order integrated. In this case, the difference of the series from the first order should be taken in terms of statistical conclusion. When the graphs of the series are examined, the stationarity of the series has been also tested with the periodogram-based unit root test proposed by Akdi and Dickey (1998) with the assumption that possible periodicity can be found in the series, and it has been observed that the series is stationary at the 10% significance level. According to the results 33, 36.5 and 73-day periods are significant in the number of daily traffic accidents in 2019. It has been demonstrated that the 73-day period corresponds to the period between Ramadan Feast and Feast of Sacrifice (there is a 70-day interval between the two religious holidays).

Keywords: Time Series Analysis, Daily Traffic Accidents, Periodicity, Turkey

TÜRKİYE'DEKİ TRAFİK KAZALARININ PERİYODİK YAPISININ ARAŞTIRILMASI

Öz

Bu çalışmada, Türkiye'de 2019 yılında meydana gelen günlük trafik kazaları verilerine zaman serisi analizi uygulanmıştır. Çalışmada kullanılan verilerin en önemli özelliği kolluk birimleri tarafından günlük olarak tutulan resmi trafik kazası kayıtları olmasıdır. Bu verilerle ilgili olarak en uygun zaman serisi modeli belirlenmiş ve trafik kazalarında periyodik bileşenlerin olup olmadığı incelenmiştir. Verilerin birinci dereceden entegre olduğu görülmektedir. Bu durumda serinin birinci dereceden farkı istatistiksel sonuç açısından alınmıştır. Serinin grafikleri incelendiğinde, serilerde olası periyodikliğin bulunabileceği varsayımı ile Akdi ve Dickey (1998) tarafından önerilen periodogram temelli birim kök testi ile serinin durağanlığı da test edilmiş ve serinin % 10 anlamlılık düzeyinde durağan olduğu görülmüştür. Elde edilen sonuçlara göre 2019 yılında günlük trafik kaza sayılarında 33, 36.5 ve 73 günlük dönemlerin önemli olduğu tespit edilmiştir. 73 günlük sürenin Ramazan Bayramı ile Kurban Bayramı arasındaki döneme denk geldiği (iki dini bayram arasında 70 günlük bir ara vardır) gösterilmiştir.

Anahtar Kelimeler: Zaman Serisi Analizi, Günlük Trafik Kazaları, Dönemsellik, Türkiye

INTRODUCTION

Time series analysis, which has uses in all fields of science, has an important application area in statistics and econometrics. A time series is a sequence of measurements observed over periodic time intervals. Monthly product quantities exported from a factory, weekly number of accidents occurring on a highway, hourly water level in a lake, annual import and export amounts of a country, annual investment and gross national product incomes, annual unemployment rates, and monthly precipitation amounts in a city can be given as an example for time series (Akdi, 2012). For example, if there are annual export volumes from the past years, a good forecast for the next year's export volumes is important during the budget preparation phase (Akdi, 2012).

Data about traffic accidents is an important example of time series in which the modeling is crucial to understand the underlying reasons in accidents and protect the people. It is also important to reveal the periodicity in data based on time series. A periodicity is a pattern in a time series that occurs at regular time intervals. More precisely, the time series is said cyclical, if the time intervals at which the pattern repeats itself can't be precisely defined and is not constant. The traffic accidents in Turkey are among the major issues on the national agenda. Considering the traffic accidents, annual casualties, and the number of injured people, it stands out as an important problem area where many citizens lost their lives (Ali and Tayfour, 2012). Traffic accidents, which are one of the most important problems of many countries, cause thousands of deaths and injuries every year, in addition to high amounts of financial losses. It is important to prevent loss of life in traffic and to ensure traffic safety (Theofilatos, 2016). In order to achieve this goal, it is necessary to carry out analyzes with correct methods and to take measures to prevent traffic accidents.

In this study; the traffic accidents occurred in Turkey and the data kept by law enforcement units in 2019 have been used.

In this study we utilize the methodology of Okkaoglu et al. (2020) in which periodogram based time series analysis is offered to identify the periodic structure of the times series. This methodology has many implementations such as weather-related time series (Akdi and Ünlü, 2021), electricity markets (Akdi et al., 2020c), air pollutions (Akdi et al., 2020a) and finance (Akdi et al., 2020b).

By revealing the periodicity of traffic accidents, the goals listed below can be achieved;1) Prevention of loss of life and property in traffic accidents, 2) providing a safer road transport for drivers and pedestrians, 3) determining the periods in which traffic accidents increase and decrease, 4) helping the state units working more effeectively to prevent traffic accidents, 5) preventing financial losses occurring in the country's economy.

Putting forth the periodic structure of traffic accidents in Turkey is also expected to contribute to the literature in this area.

The organization of the paper is as follows; a literature review is given in the first section. The second section is devoted to time series model for the traffic accident data respectively; the third section summarized the methodology; the fourth section contains empirical evidence and finally the last section concludes the study.

1. LITERATURE

There are many studies in the literature on modeling traffic accidents. In these studies, traffic accidents are modeled by time series analysis.

Abdel-Aty and Radwan (2000) employed the negative binomial model to study accident data of Central Florida for a three-year period. Their findings showed that negative binomial model is better than the Poisson regression. Also, it is shown that age and gender are the important explanatory variables. Chin and Quddus (2003) proposed zero-inflated-count models to investigate the traffic accidents of Singapore at signalized intersections. Their results showed that the proposed methodology is appropriate for the traffic accidents.

Chung et al. (2014) used Markov process to reveal the relation between traffic accidents and weather conditions.

Al-Harbi (2012) studied the relations between meteorological conditions and the traffic accidents in Kuwait. Lognormal and normal distributions with their cumulative functions were used to model the effect of meteorological impact. The results showed that temperature had the highest impact among the investigated weather variables.

Ali and Tayfour (2012) investigated the traffic accidents of Sudan by using regression and Artificial Neural Networks (ANN). The results showed that ANN had good performance on predicting the number of car accidents; they also showed that the reasons behind the car accidents were driver behavior, vehicle fleet and road defects.

Singh et al. (2020) used deep neural networks to predict the traffic accidents of India by using 16 explanatory variables. The proposed model is compared with gene expression programming and random effect negative binomial model. It is shown that the proposed methodology outperforms the others.

Al-Turaiki et al. (2016) modeled the traffic accidents in Saudi Arabia using classifications algorithms. The results implied that the age of the car was the most significant reason for the accidents.

Bayata et al. (2011) modeled the monthly traffic accidents of Turkey by utilizing deep learning. Artificial neural networks used to forecast the data. It is shown that the proposed model is powerful to forecast the number of accidents. Moreover, it is revealed that the main factors of the accidents are environmental effects, car features, driver's behavior and traffic characteristics.

Taamned et al. (2017) analyzed the number of traffic accidents of United Arab Emirates by data-mining techniques. The findings showed that male involved more than the females in car accidents and frequently car accidents occur on steep slopes.

Gao et al. (2018) developed an automated algorithm which uses associations rules to investigate the accidents influence factors of Shanghai for the period between April, 2014 and June, 2014. The findings showed that the proposed methodology can be used to prevent car accidents.

Theofilatos et al. (2016) studied the road accidents of Greece for the periods between 2008 and 2011 by employing rare event logit model. The statistical testing shows that the proposed model fits the data well and revealed the negative relation between logarithm of the speed in the accident area and the accident occurrence.

Chubukov et al. (2017) revealed the relations between number of deaths in the road accidents and socio-economic factors of Russia by using statistical and probabilistic models.

Kennet (2021) employed regression techniques to predict traffic accidents in Nigeria. The features of the model were mechanical fault, reckless driving and over-loading. Their findings implied that the proposed model with the features had good performance to predict car accidents with adjusted R-square value of 76.7%.

Alkan et al. (2021) predicted the potential car accidents injuries and their economic impacts in United States. They demonstrated the economic consequences and effects of traffic accidents.

2. DATA

This study investigates the number of daily traffic accidents between the periods January 1, 2019 and December 31, 2019 of Turkey. The analyzed data consisted of the official daily traffic accident records kept by the National Police and the Gendarmerie General Command at the level of police and gendarmerie stations.

When the time series is examined, it can be easily seen that the daily averages of accidents on weekends are higher than the other days. Especially the number of traffic accidents that occur on Sundays shows a significant difference compared to other days. Average number of accidents occurring by days are given in Table 1. In addition, the monthly total number of accidents is also given in Table 2. In Turkey people usually spend their holidays in the summer thus the number of accidents occurring during the summer seems to be quite high compared to other months.

Table 1. Daily average number of accidents

Day	**Monday**	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Sunday**
Average	70.25	68.09	68.25	67.19	73.04	77.56	***88.19***

Table 2. Monthly number of accidents

Month	**January**	**February**	**March**	**April**	**May**	**June**
Total	1404	1251	1622	1768	2236	3230
Month	**July**	**August**	**September**	**October**	**November**	**December**
Total	3124	3503	2612	2354	2014	1586

Figure 1. shows the number of daily accidents, injuries and deaths in 2019 with autocorrelation function (ACF) and partial autocorrelation functions (PACF). Since the number of injuries and deaths are directly related to the number of accidents that occurred, only the number of accidents was considered in the study.

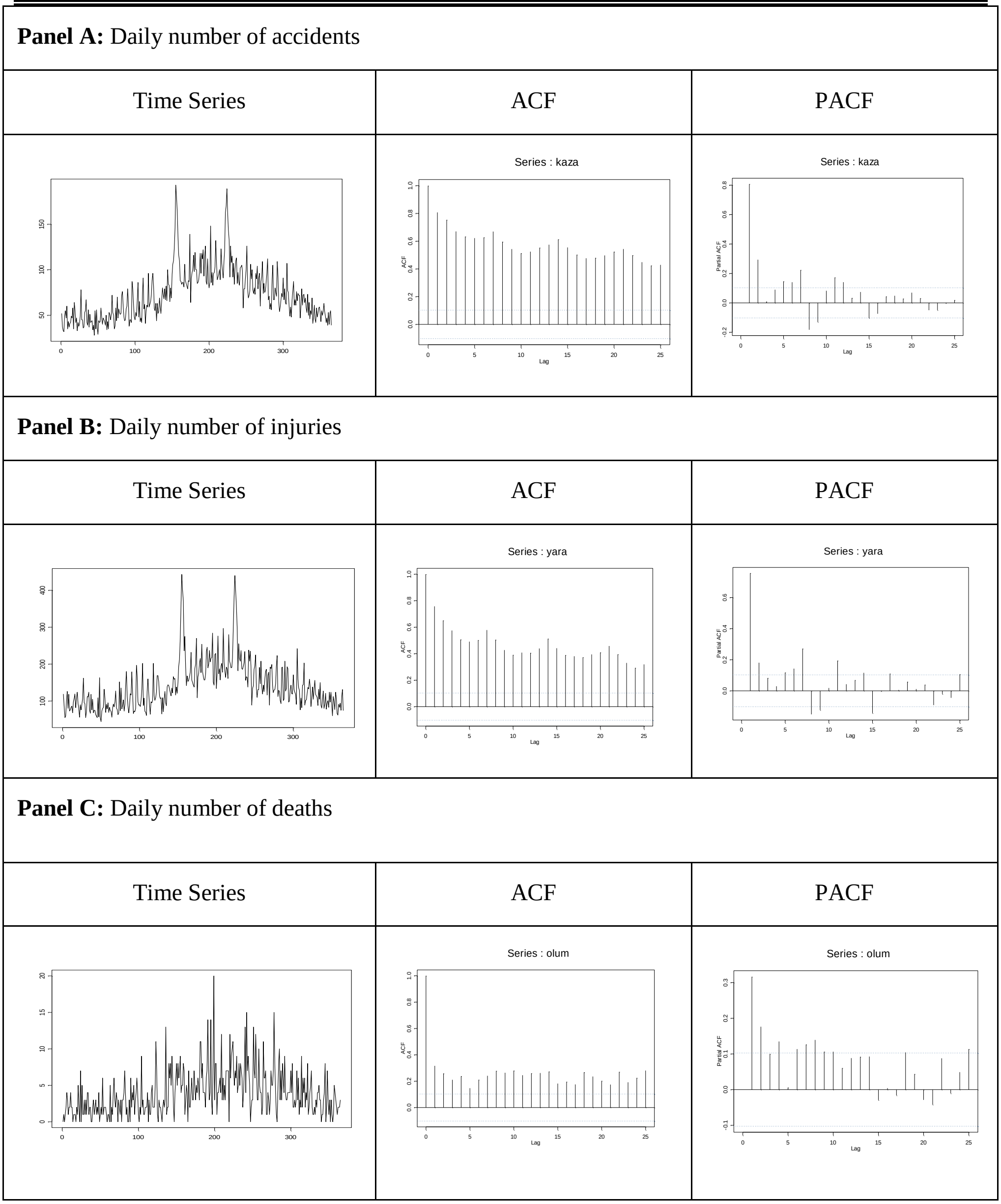

Figure 1. The number of daily accidents, injuries and deaths in 2019 with ACF and PACF

In order to fit an appropriate time series model to the daily accident numbers Y_t, by using ACF and PACF, different time series models were taken into consideration and the value of the smallest Akaike Information Criteria (AIC) statistic is used to determine the model.

The values of the AIC statistics calculated for these models and the estimation of the variance of the white noise series are given in Table 3.

Table 3. Variance estimates of white noise series with AIC statistics values

Lag	1	2	3	4	5	6
AIC	3101.49	3070.17	3072.12	3070.63	3064.24	3058.04
$\hat{\sigma}_n^2$	285.40	261.22	261.91	260.13	254.93	249.96
Lag	7	8	(7)	(1,7)	(1,7,8)	(1,2,7,8)
AIC	3038.60	**3029.15**	3265.50	3050.53	3043.89	3023.89
$\hat{\sigma}_n^2$	236.37	**229.71**	447.30	247.54	242.41	228.87

Considering different models, the model that gives the smallest AIC statistics is to be the most suitable model. Let $e_t \sim WN(0,\sigma^2)$ then the best model is given in Equation 1.

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + \ldots + \alpha_8 Y_{t-8} + e_t \, , t = 1, 2, \ldots, n \qquad (1)$$

Parameter estimations of the Model 1 is given in Table 4.

Table 4. Parameter estimations of Model 1.

Parameter	Estimate	Error	t-value	Pr > |t|	Lag
α_0	58.11551	9.85759	5.90	<0.0001	0
α_1	0.53890	0.05223	10.32	<0.0001	1
α_2	0.25766	0.05700	4.52	<0.0001	2
α_3	-0.06183	0.05850	-1.06	**0.2913**	3
α_4	-0.00733	0.05866	-0.13	**0.9005**	4
α_5	-0.00394	0.05870	-0.07	**0.9465**	5
α_6	0.06911	0.05863	1.18	**0.2393**	6
α_7	0.32819	0.05716	5.74	<0.0001	7
α_8	-0.17171	0.05239	-3.38	0.0008	8

Considering Table 4, it is seen that some parameters are not significant. When these parameters are removed from the model, the resulting model is as follows:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_7 Y_{t-7} + \alpha_8 Y_{t-8} + e_t \, , t = 1, 2, \ldots, n \, . \qquad (2)$$

The parameter estimation of Model 2 is given in Table 5.

Table 5. Parameter estimations of Model 2.

Parameter	Estimate	Error	t-value	Pr > |t|	Lag
α_0	57.32268	9.99359	5.74	<0.0001	0
α_1	0.52579	0.05067	10.38	<0.0001	1
α_2	0.23227	0.04863	4.78	<0.0001	2
α_7	0.35071	0.04879	7.19	<0.0001	7
α_8	-0.16315	0.05087	-3.21	0.0015	8

The model given in Equation (2) is suitable for the data. The time series must be stationary for any statistical inference. The stationarity of the time series is investigated with the standard Augmented Dickey Fuller (ADF) (Dickey and Fuller, 1979) test and it is observed that it is not stationary at 5% significance level and even at 10% significance level. ADF test results are given in Table 6.

Table 6. Unit root test results of the investigated time series

Panel A: Level			
		t-Statistics	Probabilities
ADF test statistics:		-1.824066	0.3685
Critical values:	1%	-3.448675	
	5%	-2.869511	
	10%	-2.571085	
Panel B: First Difference			
		t-Statistics	Probabilities
ADF test statistics:		-9.152113	0.0000
Critical values:	1%	-3.448675	
	5%	-2.869511	
	10%	-2.571085	

When Table 6 is considered, it is observed that the time series is first-order integrated (the series itself is not stationary, but the first difference of the time series is stationary). In this case, the difference of the series should be taken from a stationary time series. When the graphs of the series are considered, it is seen that possible periodicity can be found in the series. The stationarity of the series has also been tested with the periodogram-based unit root test proposed by Akdi and Dickey (1998). It has been observed that the time series is stationary at 10% significance level.

3. METHOD

Periodograms are obtained by using trigonometric transforms of the series. Trigonometric functions initially come to mind when periodic functions are mentioned. Therefore, let e_t are random variables having the same distribution, and independent of each other with expected value 0 and variance σ^2, the model is considered for the data;

$$Y_t = \mu + R\cos(wt + \phi) + e_t,\ \ t = 1,2,\ldots,n \tag{3}$$

In this model μ, R, ϕ and w terms are referred as expected value, range, phase and frequency, respectively. All of these parameters need to be estimated.

w_k are the Fourier frequancies when $w_k = 2\pi k / n$. This model can be written as, with the properties of the cosine function $a = R\cos(\phi)$ and $b = R\sin(\phi)$,

$$Y_t = \mu + a cos(w_k t) + b sin(w_k t) + e_t,\ t = 1,2, \ldots, n \tag{4}$$

Least-squares estimators of parameters according to this model are;

$$\hat{\mu} = \bar{Y}_n, \quad a_k = \frac{2}{n}\sum_{t=1}^{n}(Y_t - \bar{Y}_n)\cos(w_k t) \text{ and } b_k = \frac{2}{n}\sum_{t=1}^{n}(Y_t - \bar{Y}_n)\sin(w_k t)$$

These calculated a_k and b_k values are the Fourier coefficients. The above property of trigonometric functions imply that Fourier frequencies are invariant with respect to the mean.

$$\sum_{t=1}^{n}\cos(w_k t) = \sum_{t=1}^{n}\sin(w_k t) = 0 \tag{5}$$

Periodogram ordinate of time series at frequency w_k is calculated as follow:

$$I_n(w_k) = \frac{n}{2}\left(a_k^2 + b_k^2\right) \tag{6}$$

On the other hand, when $f(w_k)$ has the spectral density function if $\{Y_1, Y_2, \ldots, Y_n\}$ time series are stationary with $n \to \infty$, the equation is $I_n(w_k) / f(w_k) \xrightarrow{D} \chi_2^2$. In this equation, "$\xrightarrow{D}$" refers to converge in distribution (Fuller, 1996, Wei, 2006).

Akdi and Dickey (1998) showed that, under the assumption that the series has unit root,

$$T_n(w_k) = \frac{2(1 - cos(w_k))}{\sigma_n^2} I_n(w_k) \tag{7}$$

When $n \to \infty$, the distribution of the test statistic described as;

$$T_n(w_k) \xrightarrow{D} Z_1^2 + 3Z_2^2 \quad n \to \infty \tag{8}$$

In equation (8) Z_1 and Z_2 are independent standard normal distributed random variables. Briefly, when $n \to \infty$ the distribution is explained as:

$$T_n(w_k) \xrightarrow{D} \chi_1^2 + 3\chi_1^2$$

In this case, distribution of the $T_n(w_k)$ statistic is obtained when the series is stationary and the series has unit root. Therefore, the test statistics given in Equation (7) can be used to test the stationarity of the series. The critical values of the distribution are given in the related article. Some of those are;

$P(T \leq 0.034818) = 0.01$, $P(T \leq 0.178496) = 0.05$, $P(T \leq 0.369089) = 0.10$.

If the value of the statistic $T_n(w_k)$ (for instance we say, $t_n(w_k)$) is less than the critical value, the null hypothesis that the series has unit root is rejected. Although the asymptotic distribution is valid for any constant k, k=1 is generally preferred.

There are several reasons to use periodogram-based analysis. Fuller (1996), Wei (2006), Brockwell and Davis (1987) and Akdi and Dickey (1998) note advantages of using periodogram-based analysis as:

(i) The periodograms are calculated by using trigonometric transformations without depending on any model specifications. Also, the method is invariant to the mean.

(ii)The fact that the critical values of the distribution do not depend on the sample size provides more efficient estimates for small samples.

(iii) There is no need for estimating any parameters other than the variance of the white noise series.

(iv) The analytical power function exists for the test since the normalized periodogram is asymptotically distributed as χ^2 with 2 degrees of freedom under the alternative hypothesis.

(v) If the data have periodic components, the results seem to be robust.

Because of these properties, periodogram-based analysis is utilized in this study.

Periodograms are also used to search possible periodicities in data. In order to search for possible periodic components in the data, we consider the harmonic regression given in Equation 4. If the null hypothesis of $H_0: a = b = 0$ is rejected, then the series contains periodic component. The hypothesis seems to be tested by the usual F test. However, since the frequency w_k is unknown the F test may not be appropriate (Wei, 2006, page 294-95). Therefore, a test statistic

$$V = I_n(w_{(1)}) \left[\sum_{k=1}^{m} I_n(w_k) \right]^{-1} \tag{9}$$

is defined to search for a possible periodic component. Here, $I_n(w_{(1)})$ is the largest periodogram value and m is the integer part of $n/2$ (that is $m = [n/2]$). Under the null hypothesis $H_0: a = b = 0$

$$P(V > c_\alpha) = \alpha \cong m(1 - c_\alpha)^{m-1} \tag{10}$$

(Wei, 2006). Here, c_α is the critical value for a corresponding significance level α. Using the equation in (10) the critical values can be calculated as

$$c_\alpha = 1 - (\alpha / m)^{1/(m-1)}. \tag{11}$$

If the value of V statistics is greater than the critical value ($V > c_\alpha$), the null hypothesis is rejected, and it is concluded that the series contains a periodic component. The method allows to search further periodicity. In order to do that a test statistic,

$$V_i = I_n(w_{(i)}) \left[\sum_{k=1}^{m} I_n(w_k) - \sum_{k=1}^{i-1} I_n(w_{(k)}) \right]^{-1} \tag{12}$$

is defined and if the value of V_i is greater than the critical values, it is concluded that the series contains a periodic component at the corresponding period. Here, $I_n(w_{(i)})$ denotes the i th largest periodogram value.

If we observe periodic components in the data (say p_1 and p_2) we consider a harmonic regression equation as

$$Y_t = \mu_1 + A_1 \cos\left(\frac{2\pi t}{p_1}\right) + B_1 \sin\left(\frac{2\pi t}{p_1}\right) + B_2 \sin\left(\frac{2\pi t}{p_2}\right) + B_2 \sin\left(\frac{2\pi t}{p_2}\right) + e_t \ , \ t = 1, 2, ..., n \ . \tag{13}$$

Equation 13 takes periodic fluctuations into account for making predictions and forecasts for a given time series. Since the series is stationary, the parameters can be estimated using the ordinary least squares and significance of the parameters can be tested by using usual *t* test.

4. EMPIRICAL EVIDENCE

Previously, it was observed that the time series model given in Equation 2 was the most appropriate model and it was observed that the time series was not stationary even at the 10% significance level. Considering the possible periodicity of the series and the advantages of the periodogram-based unit root test, the stationarity of the series was also tested with the periodogram-based unit root test. For that reason, the value of the test statistic given in Equation 7 are calculated as

$$I_n(w_1) = 182334.27 \text{ , } \sigma_n^2 = 228.87 \text{ and } t_n(w_1) = 0.23607.$$

Moreover, the critical values of $T_n(w_1)$ statistics are 0.0348, 0.178, 0.369 respectively for 1%,5% and 10% significant level. That is

$$P(T \le 0.034818) = 0.01 \text{ , } \quad P(T \le 0.178496) = 0.05 \text{ , } \quad P(T \le 0.369089) = 0.10$$

Since $t_n(w_1) < c_{0.10}$ the number of daily accidents in 2019 is stationary at 10% significance level. However, according to the ADF test results, the series is not stationary at the 10% significance level. Since the series is stationary according to periodograms based unit root test, possible periodicity can be investigated. Therefore, the value of the test statistic for the five largest periodogram values and the period corresponding to this value are given in Table 7.

Table 7. The largest 5 periodogram values and corresponding statistics

i	1	2	3	4	5
$I_n(w_{(i)})$	182334.27	10581.39	5993.05	5809.91	5498.48
Periods	365	7.00	33	36.5	73
V_i	**0.6077**	**0.0899**	**0.0559**	**0.0574**	**0.0577**

By using Equation 11 the critical values V_i are calculated and they are given in Table 8.

Table 8. Critical values of V_i test statisitics

α	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**
c_α	0.0527	0.0491	0.0469	0.0455	0.0443
α	**0.06**	**0.07**	**0.08**	**0.09**	**0.10**
c_α	0.0433	0.0425	0.0418	0.0412	0.0406

According to Table 8, $V_i > c_{0.01}$ for $i=1,2,...,5$ all periods are statistically significant. The periods are 365, 7, 33, 36.5 and 73-day. Since the 365-day period here corresponds to all of the data, it is not meaningful to consider this number as a period.

$$\begin{aligned} Y_t = \mu &+ A_1\cos\left(\frac{2\pi t}{365}\right) + B_1\sin\left(\frac{2\pi t}{365}\right) + A_2\cos\left(\frac{2\pi t}{7}\right) + B_2\sin\left(\frac{2\pi t}{7}\right) \\ &+ A_3\cos\left(\frac{2\pi t}{33}\right) + B_3\sin\left(\frac{2\pi t}{33}\right) + A_4\cos\left(\frac{2\pi t}{36.5}\right) + B_4\sin\left(\frac{2\pi t}{36.5}\right) \\ &+ A_5\cos\left(\frac{2\pi t}{73}\right) + B_5\sin\left(\frac{2\pi t}{73}\right) + e_t\ ,\ t=1,2,...,365 \end{aligned} \quad (14)$$

The parameter estimates for this model given in Equation 14 are given in Table 9.

Table 9. Parameter estimations of the model given in Equation 14

Variable	Estimate	Standard Error	t-Value	P>|t|
μ	73.15727	0.83210	87.92	<0.0001
A_1	-27.72732	1.17703	-23.56	<0.0001
B_1	-15.19917	1.17650	-12.92	<0.0001
A_2	2.62856	1.17689	2.23	0.0261
B_2	-7.679001	1.17619	-6.53	<0.0001
A_3	-0.34595	1.17785	-0.29	0.7691
B_3	-5.16119	1.18430	-4.36	<.0001
A_4	-0.73506	1.17791	-0.62	0.5330
B_4	5.19890	1.18539	4.39	<.0001
A_5	3.34674	1.17739	2.84	0.0047
B_5	4.28510	1.17671	3.64	0.0003

According to the results in Table 9, when non-significant parameters are removed from the model, we have

$$\begin{aligned} Y_t = \mu &+ A_1\cos\left(\frac{2\pi t}{365}\right) + B_1\sin\left(\frac{2\pi t}{365}\right) + A_2\cos\left(\frac{2\pi t}{7}\right) + B_2\sin\left(\frac{2\pi t}{7}\right) \\ &+ B_3\sin\left(\frac{2\pi t}{33}\right) + B_4\sin\left(\frac{2\pi t}{36.5}\right) + A_5\cos\left(\frac{2\pi t}{73}\right) + B_5\sin\left(\frac{2\pi t}{73}\right) + e_t\ ,\ t=1,2,...,365 \end{aligned} \quad (15)$$

The model given in Equation 15 is appropriate, the parameter estimates related to this model are given in Table 10.

Table 10. Parameter estimations of the model given in Equation 15

Variable	Estimate	Standard Error	t-Value	P>|t|
μ	73.14707	0.83014	88.11	<0.0001
A_1	-27.724792	1.17407	-23.63	<0.0001
B_1	-15.19825	1.17394	-12.95	<0.0001
A_2	2.62731	1.17433	2.24	0.0259
B_2	-7.68321	1.17361	-6.55	<0.0001
B_3	-5.18177	1.18128	-4.39	<0.0001
B_4	5.19479	1.18263	4.39	<0.0001
A_2	3.31974	1.17414	2.83	0.0050
B_5	4.29108	1.17411	3.65	0.0003

According to Table 10, it can be seen that the parameters of the model given in Equation 15 are statistically significant thus it can be used to model the number of daily accidents for 2019. The graph of the estimates obtained according to the model and the observations are given in Figure 2.

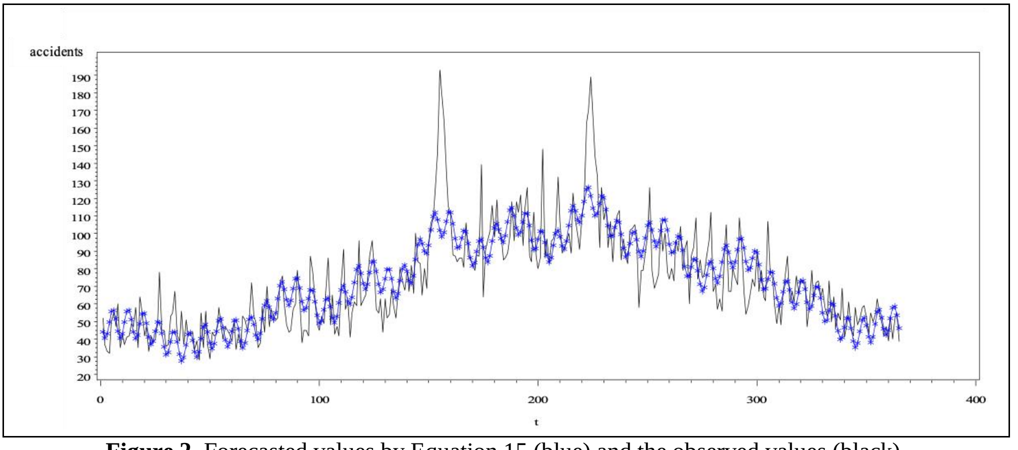

Figure 2. Forecasted values by Equation 15 (blue) and the observed values (black).

The proposed model aims to reveal the periodic structure in traffic accidents. Looking at Figure 2, the distance between the two peaks corresponds to a period of 70 days. This should also be related to the time passed between the two religious holidays. In the model, we obtained that the time between these two peaks is 73 days. Using the model given in Equation 15, the forecast values for January 2019 (31 days) were calculated and the predictions are given in Table 11.

Table11. The forecast values for January 2019 by using model 15

Date	Day	Forecasted Value	Date	Day	Forecasted Value
Jan 1, 2019	Tuesday	17.0846	Jan 16, 2019	Wednesday	22.5554
Jan 2, 2019	Wednesday	18.7143	Jan 17, 2019	Thursday	24.2805
Jan 3, 2019	Thursday	20.4455	Jan 18, 2019	Friday	24.5836
Jan 4, 2019	Friday	20.9559	Jan 19, 2019	Saturday	23.1491
Jan 5, 2019	Saturday	19.9154	Jan 20, 2019	Sunday	20.9119
Jan 6, 2019	Sunday	18.2327	Jan 21, 2019	Monday	19.3656
Jan 7, 2019	Monday	17.3652	Jan 22, 2019	Tuesday	19.4524
Jan 8, 2019	Tuesday	18.2109	Jan 23, 2019	Wednesday	20.8706
Jan 9, 2019	Wednesday	20.4192	Jan 24, 2019	Thursday	22.3196
Jan 10, 2019	Thursday	22.6384	Jan 25, 2019	Friday	22.4968
Jan 11, 2019	Friday	23.5162	Jan 26, 2019	Saturday	21.0951
Jan 12, 2019	Saturday	22.7004	Jan 27, 2019	Sunday	19.0481
Jan 13, 2019	Sunday	21.0862	Jan 28, 2019	Monday	17.8378
Jan 14, 2019	Monday	20.1275	Jan 29, 2019	Tuesday	18.3859
Jan 15, 2019	Tuesday	20.7288	Jan 30, 2019	Wednesday	20.3622
			Jan 31, 2019	Thursday	22.4314

One-year predictive values were calculated by harmonic regression and the daily averages and monthly averages are given in Table 12 and Table 13.

Table 12. Forecasted values for the one-year (January) period and actual day-average values

Day	Observed average	Forecasted average by Model 15
Monday	21.52	20.33
Tuesday	22.49	20.70
Wednesday	23.13	22.47
Thursday	24.19	24.24
Friday	23.79	24.69
Saturday	21.00	23.48
Sunday	18.63	21.54

Table 13. Monthly average forecasted values and actual day-average values

Month	Observed average	Forecasted average by Model 15
January	22.36	20.69
February	22.32	22.97
March	26.92	25.24
April	23.81	25.40
May	22.85	23.75
June	20.32	21.32
July	19.39	20.45
August	18.48	21.70
September	22.44	22.51
October	21.48	22.72
November	23.07	22.11
December	21.90	21.10

5. CONCLUSION

Traffic accidents are one of the most important issues in our country. Many studies are carried out to prevent and reduce accidents. In Turkey, 95.2% of passenger transport is carried out by roads. This rate is 89% in the USA and 79% in EU countries. Road usage rate in freight transportation is around 76.1% in Turkey. This rate is 69.5% in the USA and about 45% in EU countries. Making the most preferred transportation system safer is one of the most important duties of the state.

In this study, a time series analysis has been made on daily traffic accidents data occurred in specific time period 2019. The main purpose of the study is to reveal whether there is a periodicity in daily traffic accidents in Turkey. The amount of the traffic accident was examined and the periodicities were revealed. The most important feature of the data used in the study is that it is recorded by the law enforcement units. By determining the periodic structure of the traffic accidents, loss of life in car accidents can be prevented and transportation can be safer. The findings of the study can be used by state officials to form effective strategies which may reduce loss of life and property.

The proposed time-series model for the daily number of traffic accidents occurred in 2019 in Turkey is the most significant model and the time series has been shown to be stationary at the 10% significance level. Considering the possible periodicity of the series and the advantages of the periodogram-based unit root test, the stationarity of the series has been tested with the periodogram-based unit root test. The harmonic regression model was used by considering the periodicities. When the predicted values are analyzed, traffic accidents are low on Sunday, Monday and Tuesday; it has been determined to be high on Thursdays and Fridays. This result can be attributed to the increase in traffic towards the weekends. Similarly, the decrease can be attributed to the fact that drivers drive more rested at the beginning of the week.

As a result of the analysis, it has been observed that 365, 7, 33, 36.5 and 73-day periods are significant in the number of daily accidents in 2019. Since 365 days cycle corresponds to the whole data, 7 days cycles corresponds to weekly periods, 33 and 36.5 cycle corresponds to monthly periods, it is considered that these periods are not meaningful to consider as a period. However, the 73-day period is considered as the period between Ramadan Feast and Sacrifice Feast in Turkey (there are 70 days between two religious holidays). In this case, the periodicity in the traffic accidents in Turkey corresponds to 73 days period.

In Turkey, where road transportation is mostly preferred, many steps have been taken in recent years to ensure traffic safety. The most important of these are the construction of new roads and bridges, the expansion of existing roads, and the use of technology more on highways. In addition to all efforts, statistical studies are also important in this field where the human factor comes to the fore. It is considered that the results obtained will contribute to the units responsible for traffic safety and to the literature as well.

REFERENCES

Abdel-Aty, M. A., & Radwan, A. E. (2000). Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention*, 32(5), 633-642. https://doi.org/10.1016/S0001-4575(99)00094-9

Akdi, Y. (2012). *Zaman Serileri Analizi-Birim Kökler ve Kointegrasyon*, 3'nci Baskı, Gazi Kitapevi, Ankara.

Akdi, Y. and D. A. Dickey. (1998). Periodograms of unit root time series: distributions and test, *Communications in Statistics-Theory and Methods*, 27(1), 69-87. https://doi.org/10.1080/03610929808832651

Akdi, Y., Okkaoğlu, Y., Gölveren, E., and Yücel, M. E. (2020a) Estimation and forecasting of PM 10 air pollution in Ankara via time series and harmonic regressions. *International Journal of Environmental Science and Technology*, 1-14. https://doi.org/10.1007/s13762-020-02705-0

Akdi, Y., Varlik, S., and Berument, M. H. (2020b) Duration of Global Financial Cycles. *Physica A: Statistical Mechanics and its Applications*, 124331. https://doi.org/10.1016/j.physa.2020.124331

Akdi, Y., Gölveren, E., and Okkaoğlu, Y. (2020c) Daily electrical energy consumption: Periodicity, harmonic regression method and forecasting. *Energy*, 191, 116524. https://doi.org/10.1016/j.energy.2019.116524

Akdi, Y. and Ünlü, K.D. (2021). Periodicity in precipitation and temperature for monthly data of Turkey. *Theor Appl Climatol* 143, 957–968. https://doi.org/10.1007/s00704-020-03459-y

Al-Harbi, M., Yassin, M. F., and Shams, M. B. (2012). Stochastic modeling of the impact of meteorological conditions on road traffic accidents. *Stochastic Environmental Research and Risk Assessment*, 26(5), 739-750.

Alkan, G., Farrow, R., Liu, H., Moore, C., Ng, H. K. T., Stokes, L., and Zhong, Y. (2021). Predictive modeling of maximum injury severity and potential economic cost in a car accident based on the General Estimates System data. *Computational Statistics*, 1-15.

Al-Turaiki, I., Aloumi, M., Aloumi, N., & Alghamdi, K. (2016, November). Modeling traffic accidents in Saudi Arabia using classification techniques. In 2016 4th *Saudi International Conference on Information Technology* (Big Data Analysis) (KACSTIT) (pp. 1-5). IEEE.

Ali, G. A., and Tayfour, A. (2012). Characteristics and prediction of traffic accident casualties in Sudan using statistical modeling and artificial neural networks. *International journal of transportation science and technology*, 1(4), 305-317.

Bayata, H. F., Hattatoğlu, F., & Karslı, N. (2011). Modeling of monthly traffic accidents with the artificial neural network method. *International Journal of the Physical Sciences* Vol. 6(2), 244-254.

Brockwell, P. J., Davis, R. A., & Fienberg, S. E. (1991). *Time series: theory and methods: theory and methods*. Springer Science & Business Media.

Chin, H. C., & Quddus, M. A. (2003). Modeling count data with excess zeroes: An empirical application to traffic accidents. *Sociological methods & research*, 32(1), 90-116. https://doi.org/10.1177/0049124103253459

Chubukov, A., Kapitanov, V., Monina, O., Silyanov, V., & Brannolte, U. (2017). Simulation of regional mortality rate in road accidents. *Transportation Research Procedia*, 20, 112-124. https://doi.org/10.1016/j.trpro.2017.01.031

Chung, Y. S., Park, R. K., & Kim, J. M. (2014). Study on predictive modeling of incidence of traffic accidents caused by weather conditions. *Journal of the Korea Convergence Society*, 5(1), 9-15. https://doi.org/10.15207/JKCS.2014.5.1.009

Dickey, D. A., and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of The American Statistical Association*, 74(366a), 427-431. https://doi.org/10.1080/01621459.1979.10482531

Fuller, W. A. (1996). *Introduction to statistical time series.* 2nd ed. New York: John Wiley & Sons.

Gao, Z., Pan, R., Yu, R., & Wang, X. (2018). Research on automated modeling algorithm using association rules for traffic accidents. In 2018 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 127-132). IEEE. https://doi.org10.1109/BigComp.2018.00027

Kenneth, G. E. (2021). Statistical Application of Regression techniques in Modeling Road Accidents in Edo State, Nigeria. *Sch J. Phys. Math Stat*, 1, 14-18.

Okkaoğlu, Y., Akdi, Y., and Ünlü, K. D. (2020). Daily PM10, periodicity and harmonic regression model: The case of London. *Atmospheric Environment*, 117755. https://doi.org/10.1016/j.atmosenv.2020.117755

Singh, G., Pal, M., Yadav, Y., & Singla, T. (2020). Deep neural network-based predictive modeling of road accidents. *Neural Computing and Applications*, 1-10. https://doi.org/10.1007/s00521-019-04695-8

Taamneh, M., Alkheder, S., and Taamneh, S. (2017). Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates. *Journal of Transportation Safety & Security*, *9*(2), 146-166.

Theofilatos, A., Yannis, G., Kopelias, P., & Papadimitriou, F. (2016). Predicting road accidents: a rare-events modeling approach. *Transportation research procedia*, 14, 3399-3405. https://doi.org/10.1016/j.trpro.2016.05.293

Wei, W.W.S. (2006). Time Series Analysis: Univariate and Multivariate Methods, Second Edition, Pearson Addison Wesley, Boston