

Determinants of the Algerian Tourism Demand (1991-2023): Regularization approaches using LASSO, RIDGE, and Elastic Net

BOUHALI Ihssane*
High National School of
Statistics and Applied
Economics, Algeria
Bouhali.ihssane@gmail.com

DJEMACI Brahim
M'Hamed Bougara University
of Bomeredes, Algeria
Brahim.djemaci@gmail.com

submitted:28/05/2025

Accepted:25/06/2025

published:30/06/2025

Abstract:

The tourism industry can play an important role in the development of the Algerian economy and the promotion of its population's welfare. The development of such an industry needs first of all the analysis of its demand to figure out the determinants of tourism demand, which help in the design of supply strategy and the making of policies to attract tourists. Tourism demand is generally measured by the number of tourist arrivals and explained by several variables; so, for each country, there is a specific model of tourism demand. The issue of multicollinearity sometimes involves a high correlation between variables; that's why a regularization method can be applied, such as LASSO, RIDGE, and Elastic Net, to solve this problem. An application of these three methods on the Algerian data (1991-2023) aims to evaluate their performance in handling multicollinearity. The results reveal that RIDGE regression is the best-performing model compared to LASSO and Elastic Net. The results of the Ridge regularization method demonstrate that population, GDP, foreign investment, and bed capacity have a positive significant effect on the Algerian tourism demand, except the exchange rate, which has a negative influence.

Keywords: tourism demand, Algeria, regularization methods, determinants.

JEL Classification Codes: C52, C55, Z32

correspondent author *

Introduction:

Tourism plays a significant role and has an effect on various economies and the global world; it has become one of the important sectors that grows faster and takes a major place in designing strategies and policies, not only for the development of this sector but also for the development of the whole economic system, because the success of such a sector, tourism, is highly related to the success and the development of other economic sectors: transport, agriculture, industry, ICT, handicrafts, energy, etc.

Tourism is the only sector that can be defined by its demand, which reflects the evolution of this sector by either the number of arrivals or “tourism expenditures.” Countries should really invest in studying this side of the sector in order to have a clear idea of how it can attract more tourists to visit it.

Algeria has over two million Km² and can supply the national and international tourist a large and varied tourism product, but its policies and economic strategies put the tourism in a neglected position since the independence until 1970, when the government paid attention to the importance of this sector, where some complexes and hotels had been constructed. Unfortunately, these interests and projects were interrupted and stopped because of security instabilities. In reality, and with the start of the 90s, some policies and measures were taken that were not acting positively for the tourism development. The beginning of the 2000s was like the light for Algerian tourism, which is trying to recreate a place in the economy. In 2007, the administrations and offices of tourism changed the vision of developing this sector by launching a new strategy, “SDAT,” where there are objectives for 2015 and 2025, but they have been extended to 2030.

Research problem

This research paper aims to address the following problematic question:

How can the regularization method explain and identify the determinants of the tourism demand in Algeria?

Hypotheses

The research relies on three hypotheses:

- The regularization methods provide a better fit for modeling Algerian tourism demand with the presence of the multicollinearity problem.
- Elastic Net regression is an effective modeling technique for tourism demand, as it combines the LASSO and RIDGE regressions.
- Income per capita and bed capacity have a positive influence on the Algerian tourism demand, whereas exchange rate has a negative effect.

Aims of this study

The aim of this paper is to identify, among many factors collected from the literature, which influence and determine the Algerian tourism demand.

Add to this, a comparison of the regularization methods aims to identify the best regression fit model under a multicollinearity problem.

Study structure

The first section is allocated to the review of literature, where we summarize some of the studies about tourism demand. The next section is about methodology, where the different variables are described with the database sample and the methodology used in modeling tourism demand. In the last section, print out the principal results of the study.

Literature review:

Christine Lim, in her paper “An econometric classification and review of international tourism demand models” (1997), evaluated 100 published empirical studies of tourism demand modeling, in which most studies used annual data from national tourism offices and international organizations (IMF, WTO, etc.). She also analyzed these models in order to select the most appropriate factors that influence tourism demand, such as relative prices, transport costs, and income.... This will help in modeling and forecasting international tourism demand.

Haiyan Song and Gang Li, in “Tourism Demand Modeling and Forecasting: A Review of Recent Research,” pointed out the diversity of methods used in analyzing and forecasting tourism demand, including time series methods, econometric approaches, and newer techniques that integrate both quantitative and qualitative variables.

Also, Brian H. Archer, in his article “Forecasting Demand,” described the most useful methods for both quantitative variables (such as time series and econometric methods) and qualitative factors, whose analysis

depends on informed judgment (such as morphological analysis and the Delphi technique).

Peter Johnson and John Ashworth, "Modeling tourism demand: A summary review" (2006), indicated the importance of econometric methods in modeling demand for tourism. They also highlighted the variables used in previous studies, where number of arrivals or tourism expenditure are considered as dependent variables, which are explained by income per capita, relative exchange rates, relative prices, and relative cost of travel.

These independent variables were also mentioned in the Witt & Martin paper "Econometric Models for Forecasting International Tourism Demand" (they estimated the model of international tourism demand in Germany using ordinary least squares (OLS)).

The use of number of arrivals as a dependent variable, representing tourism demand, was also mentioned in Sevgin Akis's paper (A Compact Econometric Model of Tourism Demand for Turkey, 1998). This study of a regression model taking the period 1980-1993 of 18 countries indicated a positive relationship between tourism demand and income and a negative relationship between number of arrivals and prices.

In order to understand the major factors that influence tourism demand in Spain, Pilar, Gronzalez, and Paz Moral, in their paper titled "An Analysis of the International Tourism Demand in Spain" (1995), found that the explanatory variables are prices, income, and population, on which the indicator of tourism demand (tourism expenditures or number of arrivals) depends.

For the same aim, Nikolaos Dritskis and Spiros Athanasiadis used an econometric approach to model tourism demand in Greece, which is measured by the number of arrivals, depending on population, income, average total cost for a 10-day stay in Greece and other competitive Mediterranean countries, exchange rate, investment, and political stability. This analysis found that competitive destinations, exchange rate, and investment have a big influence on tourism demand.

A survey done by Alan Sauron (1978) indicated that the economic variables used in the determination of tourism demand are real income, related prices, transport costs, and advertising. Alastaire Alinsato and Gnimagon Fabrice M. (2018) used the Generalized Method of Moments (GMM) to analyze international tourism demand in Africa. The study indicated that political stability, exchange rate, ecotourism, and the quality of airport infrastructure can affect the choice of destination in Africa.

In the case of Morocco, Bouzahzah Mohamed and ElMeyari Younesse (2012) aimed to explain Moroccan tourism demand, quantified by the number of arrivals, using a Vector Error Correction Model. The results have shown that revenue, bed capacity, and exchange rate all have a positive influence on tourism demand.

For instance, Dekkiche Djamel and Daoudi AbdElfatah (2019) employed tourism receipts as the dependent variable in a time series analysis to examine the determinants of Algerian tourism demand, which has a positive relationship with tourism spending and trade openness, however, the exchange rate has a negative influence on it.

A recent paper authored by Braham Mahmoud Salih and Ahmed Suleiman Yassin Al-Saffar (2021) highlights the relevance of econometric models in identifying and analyzing the determinants of tourism demand, based on a literature review of multiple studies.

Firstly. Methodology:

The increase in tourist movements in recent years depends on multiple variables (qualitative or quantitative); in this article, we focus on the quantitative ones in order to determine the factors that influence tourism demand.

Variables and Data:

In this section, we will define the quantitative variables of our model.

1.1. Dependent variable:

In most previous studies, the number of arrivals is the variable that indicates and reflects tourism demand, which represents the number of tourists (resident or not) who enter a country and practice any form of tourism.

Independent variables:

The main variables that explain the number of arrivals are:

- “Income per capita”: is the most suggested variable in almost all studies of tourism demand. That is because when income per capita increases, the need for tourism increases, and the population demands to visit more places.

- “Price”: the use of price of products may reflect an important influence on tourism demand.

- “investment”: in this study, we have actually two variables of investment:

- “Bed capacity”: This indicates the supply of the country.

➤ “Foreign investment”: to indicate the other type of investment (infrastructure, handcraft).

-“Exchange rate: According to the literature review, the “exchange rate may have a significant effect on tourists; that is to say, the foreign currency influences the tourists.

-“Population”: the number of people in a country could also create some new opportunities to develop the tourism demand.

All the variables used in this study were supplied by the World Bank database, except for bed capacity data, which was provided by the Algerian Ministry of Tourism. The sample of the study contains 32 observations from 1991 to 2023. The area of this study is Algeria.

2. Data modeling:

After the selection of variables and sample of the study, the following model was obtained:

$$Arr = f(IPC, PR, FI, BC, ER, PO)$$

With:

-Arr: number of tourist arrivals

-IPC: income per capita

-PR: price

-FI: foreign investment

-BC: bed capacity

-ER: exchange rate

-PO: population

The regularization methods were employed to estimate this function of tourism demand for Algeria, where multicollinearity seemed to be a problem to solve.

2.1. Multicollinearity:

In a regression model where the dependent variable is explained by multiple independent variables, a relationship between these variables can be a problem if it is highly significant, as it can lead to incorrect model assumptions and estimations.

In this paper, we use VIF (Variance Inflation Factor) to select any multicollinearity between our data variables. VIF values are calculated using the following equation:

$$VIF = 1/(1-R_i^2)$$

We say that multicollinearity exists if $VIF \geq 5$.

2.2. Regularization method:

Regularization is a set of methods aimed at avoiding overfitting of a model. Three regression models were applied to our data:

2.2.1 RIDGE regression:

Ridge regression, also known as ‘‘L2 regularization’’, aims to reduce the risk of error due to overfitting of a model by adding a penalty term on the sum of the squared coefficients to the model function. This helps make the model more stable, especially in the presence of multicollinearity issues.

Performing the RIDGE regression with the optimal λ value using the formula:

$$\min \beta_{ridge} \{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \beta_j^2 \}$$

2.2.2. LASSO regression: Least Absolute Shrinkage and Selection Operator:

Lasso regression, or ‘‘L1 regularization’’, serves not only to regularize the model but also to perform variable selection, where LASSO can shrink some of the coefficients to zero by adding a penalty based on the absolute value of the coefficients.

Performing the LASSO regression with the optimal λ value using the formula:

$$\min \beta_{lasso} \{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j| \}$$

2.2.3. Elastic Net Regression:

A combination of the two previous regularization techniques (LASSO and RIDGE) is known as Elastic Net regression. The strength of this method lies in the incorporation of the penalty terms L1 and L2, allowing it to perform relevant feature selection while also addressing multicollinearity issues.

Performing the Elastic Net regression with the optimal λ value using the formula:

$$\min_{\beta_{elasticnet}} \{ \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \}$$

Secondly. Results and discussion:

The econometric analysis of the variables that can influence Algerian tourism demand allows us to discuss the empirical results, which are summarized in the tables below. This part is divided into three sections: first, the descriptive analysis and tests of all the variables will be discussed, then, the different regressions will be estimated, and finally, the validation of each one will be tested in the third section.

1. Variable analysis and tests:

This section presents an overview of the statistical descriptive analysis, correlation, and multicollinearity test.

1.1 Descriptive analysis:

The following table summarizes the descriptive analysis of all data variables:

Table N°1: statistical descriptive analysis

Variables	Mean	Std, DEV	Observation
Arr	1480142.45	824682.66	The number of arrivals varies on average by ±824682 from the mean, which indicates that the tourist demand is not stable over this period of study.
BC	88886.36	2214.72	The statistics show a relative stability over the period for bed capacity.
FI	109466996.8	225638635.9	An extreme fluctuation in IDE.
PR	97.33	44.868	A high variation of standard deviation to the mean, that is to say that the price index is not stable over the period of study.
PO	35229466.2	6008562.91	A high standard deviation for a population can be explained by a significant growth of it.

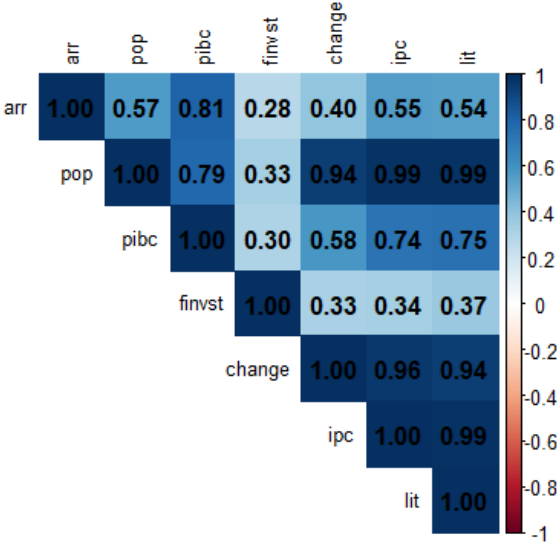
IPC	3511.89	1615.9	The GDP variable shows a high variation from year to year.
ER	78.78	32.20	Comparing the standard deviation to the mean, the exchange rate had a high fluctuation over the years.

Source: Prepared by the researcher.

1.2 Correlation Test:

To analyze the relationship between the variables of our data, we used RStudio to calculate the correlation coefficient.

Figure N°01: Correlation graph generated in Rstudio.



Source : prepared by the researcher.

The graph below shows that most of the variables are highly correlated since the correlation coefficient is superior to ± 0.5 and near ± 1 , except for foreign investments, which are lowly correlated with the others.

1.3 Multicollinearity test:

Multicollinearity tests aim to identify the existence of high correlation between the independent variables in a linear regression model. Multicollinearity can be tested using VIF (Variance Inflation Factor), we can say that a variable is correlated with the other independent ones if $VIF \geq 5$.

Table N°2: VIF's values.

Independent variables	VIF
Population	96.16
GDP	8.06
IDE	1.36
Exchange Rate	38.4
Price	114.82
Bed Capacity	68.8

Source : prepared by the researcher.

We can clearly see that the independent variables are highly correlated, with the exception of IDE (foreign investment). This shows a multicollinearity in data.

2. Data modeling using the regularization methods:

2.1 Determining the different lambda values:

The table displays the lambda values generated by Rstudio for the three-regularization model: LASSO, Ridge, and Elastic Net.

Table N°3: Lambda values

	Lambda (λ)	
	λ_{min}	λ_{1se}
LASSO	77192	342007
RIDGE	86711	1173245
Elastic Net	106410.4	429580.8

Source : prepared by the researcher.

These values are estimated using the “cv.glmnet” function, which performs K-fold cross-validation by default. We have two specific lambda values:

- Lambda min (λ_{min}): the value of lambda that minimizes the cross-validation error.
- Lambda 1se (λ_{1se}): largest value of lambda, which gives a more regularized model.

2.2 Model estimation:

The table resumes the different estimated coefficients of the LASSO, Ridge, and Elastic Net for lambda min and lambda 1se.

Table N°4: Estimated coefficient: LASSO, Ridge, and Elastic Net (Rstudio output)

Coefficient	LASSO		RIDGE		Elastic Net	
	$\lambda_{\min}$	λ_{1se}	$\lambda_{\min}$	λ_{1se}	$\lambda_{\min}$	λ_{1se}
B(intercept)	202828.2269	787284.5811	218770	346021.6	231818	710164.8841
PO	-	-	0,00174	0.010898	-	-
IPC	363.7107	197.2888	354.2	107.933	355.3151	219.2484
FI	-	-	0.00022	0.000164	0.000004518	-
EC	-	-	-3818.607	42.395	-	-
PR	-	-	2526.848	1360.393	-	-
BC	-	-	0.151	2.104884	-	-

Source : prepared by the researcher.

Based on the estimation presented in this table, the regularization regression model is as follows:

- LASSO regression:
 - LASSO regression with $\lambda_{\min}$:

Arr = 202828.22+363.7107*IPC

- LASSO regression with λ_{1se} :

Arr =78728.5811+197.288*IPC

- RIDGE regression:
 - RIDGE regression with $\lambda_{\min}$:

$$\text{Arr} = 218770 + 0.00174 * \text{PO} + 354.2 * \text{IPC} + 0.00022 * \text{FI} + (-)3818.607 * \text{ER} + 2526.88 * \text{PR} + 0.151 * \text{BC}$$

- RIDGE regression with $\lambda_{1\text{se}}$:

$$\text{Arr} = 346021.6 + 0.010898 * \text{PO} + 107.933 * \text{IPC} + 0.00164 * \text{FI} + 42.395 * \text{ER} + 1360.393 * \text{PR} + 2.104884 * \text{BC}$$

- Elastic Net Regression:
 - Elastic Net regression with $\lambda_{\min}$:

$$\text{Arr} = 231818 + 355.3151 * \text{IPC} + 0.000004518 * \text{FI}$$

- Elastic Net with $\lambda_{1\text{se}}$:

$$\text{Arr} = 710164.8841 + 219.2484 * \text{IPC}$$

From these equations we can conclude:

- LASSO regularization: From the two equations, we observe that only IPC (PIB per capita) explains and determines the number of tourist arrivals. A positive estimated coefficient explains a positive relationship between two of the variables.
- RIDGE regression: We conclude that all the variables influence positively the number of arrivals in the case of $\lambda_{1\text{se}}$; however, if $\lambda_{\min}$, a negative influence is observed between arrival and exchange rate.
- Elastic Net regression: The two equations show a positive influence between the dependent variable “arrival” and the independent variables. We have a positive influence of IPC and FI on the number of arrivals in the case of $\lambda_{\min}$. However, if $\lambda_{1\text{se}}$, only IPC has a positive impact on it.

2.3 Model validation:

A validation of a regression model can be assessed using the R^2 on the testing data. A higher value of R^2 , affords a better ability for the model to explain the dependent variable.

Table N°5: R^2 values of the regularization model

	R^2	
	λ_{min}	λ_{1se}
LASSO	0.4011833	0.4011833
RIDGE	0.6499775	0.4011833
Elastic Net	0.640135	0.509433

Source : prepared by the researcher

From the table, we can conclude:

- LASSO regression: R^2 is the same in the cases (λ_{min} and λ_{1se}); that is to say that increasing regularization did not impact the explanatory power.
- A significant drop in the R^2 from 64.9% to 40.1% for RIDGE regression, which means a loss in the fit model with λ_{1se} .
- For Elastic Net regularization, we also observe a drop of R^2 from 64% to 50.9%, but it retains more explanatory power under λ_{1se} than LASSO and RIDGE.

The results presented in table N°5 indicate that RIDGE regression using the optimal regularization coefficient (λ_{min}) provides the best fit for modeling tourism demand; it achieves the highest determination coefficient R^2 (64.9%).

This model is closely followed by the Elastic Net under λ_{min} , also with a R^2 of 64%, while a LASSO regression reflects a limited performance with an R^2 of 40%.

The Ridge model is retained to estimate tourism demand in Algeria; it shows a coefficient of determination equal to 65%, which remains a relevant rate in the presence of multicollinearity among economic variables.

The following interpretation can be drawn:

- Population has a positive effect on tourism demand, which can be explained not only by the increase in the young population, which creates a large base of potential consumers, thereby boosting domestic tourism demand, but also by a strong international demand limited to the Algerian diaspora.
- The positive link between prices and tourism demand may reflect an increase in the purchasing power of the population.
- Foreign investment contributes to the development and the modernization of infrastructure, thereby enhancing the attractiveness of Algeria as a destination.
- An increase in supply strongly correlated with higher tourism demand. Bed capacity is a crucial lever since it facilitates hosting tourists.
- Unexpected negative influence between tourism demand and exchange rate: In our case “Algeria,” this can be explained by the instability of the currency market and the significant gap between the official exchange rate and the parallel rate market, which is widely used by economic agents.

Conclusion:

The importance of tourism resides in its capacity to generate income and create job opportunities quickly. It has the power to promote population welfare.

This paper aims to compare the regularization methods (LASSO, RIDGE, and Elastic Net) to determine the model that best explains the Algerian tourism demand.

The analysis of the three techniques based on our sample data shows that a RIDGE regression provides the best-fit model. It can be concluded that all of population, GDP, bed capacity, foreign investment, and prices have a positive impact on the Algerian tourism demand, except for exchange rate, which has a negative impact on it.

The results of this study lead us to verify the hypotheses:

- The presence of multicollinearity in the study data is effectively addressed by the regularization method.
- The coefficient of determination shows that RIDGE regression is more effective than LASSO and Elastic Net.

- Effectively, bed capacity and income per capita have a positive influence on the tourism demand, whereas the exchange rate has a negative effect.

Recommendations:

- Algerian tourism authorities can clearly use the findings of our study in developing innovative and diversified supply strategies.
- The modernization of bed-capacity infrastructure.
- Build a coherent Algerian tourism image.

Future outlook:

- Integration of additional dimensions such as health to adjust policies and strategies in a post-pandemic world.
- Application of advanced machine learning techniques that deal with both quantitative and qualitative variables simultaneously.
- Collaboration with other scientific fields will significantly enrich the understanding of the fluctuation in the tourism demand.

List of references:

1. Alan Sauran (1978), "Economic determinants of tourist demand: A survey", *The Tourist Review*, No 33, pp: 2 – 4.
2. Arief Rahman Nur, Andi Kresna Jaya & Siswanto Siswanto (2024), "Comparative analysis of Ridge, LASSO and Elastic Net Regularization approaches in handling multicollinearity for infant mortality Data in South Slawesi", *Jurnal Matematika, statistika DANKOMPUTASI*, Vol 20, No 2.
3. Brian H. Archer (1976), "Forecasting demand in tourism", University of Wales Press, pp: 14-16.
4. Christin Lim (1997), "An econometric classification and review of international tourism demand models", *Tourism Economics*, No 1, pp: 69-81.
5. Christin Lim (1997), "Review of international tourism demand models", *Annals of tourism research*, No 4, pp: 835-849.
6. Egon Smeral (1988), "Tourism demand, Economic Theory and Econometrics: An integrated Approach", *Journal of travel research*, No26, pp: 38-43.
7. Geoffrey I. Crouch (1995), "A META-ANALYSIS OF TOURISM DEMAND", *Annals of Tourism Research*, No 1, pp: 103-118.
8. Haiyan Song & Gang Li (2008), "Tourism demand Modelling and Forecasting – A Review of recent research", *tourism management*, No 29, pp: 203-220.

9. Haiyan Song & Stephen F. Witt (2012), "Tourism demand modeling and forecasting: modern econometric approaches", Routledge, Taylor and Francis group, New York USA.
10. Muzaffer Uysal & John L. Crompton (1984), "Determinants of demand for international tourist flows to Turkey", *Tourism management*, No 5, pp: 288-297.
11. Nikolaos Dritsakakis & Spiros Athanasiadis (2000), "An econometric model of tourist demand: the Case of Greece", *Journal of Hospitality and Leisure*, No4, pp: 83-90.
12. Peter Johanson & John Ashworth (1990), "Modelling tourism demand: A summary review" *leisure studies*, No 2, pp: 145-161.
13. Pilar Gonszlez & Paz Moral (1995), "An analysis of the international tourism demand in Spain", *International Journal of Forecasting*, No 11, pp: 233-251.
14. Sevgin Akis (1998), "A compact econometric model of tourism demand for Turkey", *Tourism Management*, No 1, pp: 99-102.
15. Shady I Altelbany (2021), "Evaluation of Ridge, Elastic Net and LASSO regression methods in precedence of multicollinearity problem: A simulation Study", *Journal of Applied Economics and business studies*.
16. Stephen F. Witt & Christine A. Martin (1987), "Econometric models for forecasting international tourism demand", *Journal of travel research*, No 25, pp: 23-30.
17. Tadayuki Hara (2008), "Quantitative tourism industry analysis", Butterworth-Heinemann, Elsevier, Oxford UK.
18. Wasiaturrahma & Hilda Rohmawati (2021), "Multicollinearity in tourism demand model: Evidence from Indonesia", *Economics Development Analysis Journal*.
19. Alastaire Alinsato & Gnimagnon Fabrice M (2018), "Analyse de la demande internationale du tourisme en Afrique", *WTO chair Programms*.
20. Mohammed Bouzahzah & Younesse El Menyari (2012), "Les Déterminants de la demande touristique. Le cas d Maroc", *Munich Personal RePEc Archive*, No39029, www.mpra.ub.uni-muenchen.de.
21. برهم محمود صالح واحمد سليمان ياسين الصفار (2021)، "الطلب على السياحة ونمذجته"، *مجلة العلوم الإنسانية لجامعة زاخو*، المجلد 9، العدد 4، ص 951-928.
22. دقيش جمال ودواودي عبد الفتاح (2019)، "دراسة قياسية لمحددات الطلب السياحي في الجزائر خلال الفترة 1995-2016"، *مجلة الاجتهاد للدراسات القانونية والاقتصادية*، المجلد 8، العدد 3.
23. عميش سميرة (2016)، "مصادر الطلب السياحي الأجنبي على الجزائر خلال الفترة الزمنية 2000-2013"، *مجلة العلوم الاقتصادية والتسيير والعلوم التجارية*، المجلد 10، العدد 16، ص 347-334.
24. www.datacamp.com/tutorial/tutorial-ridge-lasso-elastic-net.
25. www.ibm.com/fr-fr/think/topics/ridge-regression.