**Anbar Journal of Engineering Science (AJES )** **Vol. 12 , No. 2 (2021) , pp.** 222 ~ 228
P-ISSN: 1997-9428; E-ISSN: 2705-7440 DOI: https://doi.org/10.37649/aengs.2021.171190

Unviersty of Anbar

**Anbar Journal Of Engineering Science©**

**journal homepage: http://www.uoanbar.edu.iq/Evaluate/**

# Modal Split Model Using Multiple Linear Regression Analysis

**Omaima A. Yousif, Adil N. Abed, Hamid A. Awad**

*Department of Civil Engineering, College of Engineering, University Of Anbar. Ramadi, Anbar, Iraq*

**PAPER INFO**

***Paper history:***
Received 8/09/2021
Revised 14/10/2021
Accepted 26/10/2021

***Keywords:***
Modal split models; traffic analysis zones (TAZ); multiple linear regression



**ABSTRACT**

Several modal split models have been created around the world to forecast which mode of transportation will be selected by the trip - maker from among a variety of available modes of transportation. This modeling is essential from a planning standpoint, as transportation systems typically receive significant investment. In this study, the main purpose was to develop a mode choice model using multiple linear regressions for Ramadi city in Iraq. The study area was divided into traffic analysis zones (TAZ) to facilitate data collection. The data was collected through a home interview of the trip makers in their home units through a questionnaire designed for this purpose. The result showed that the most influential factors on the mode choice for the general trips model using multiple linear regressions are car ownership, age, and trip cost. This model gave a good correlation coefficient of 0.829 meaning that the independent variables explain 82.9 of variance in the dependent variable (type of mode), which will help transport planners in developing policies and solutions for future.

## 1. Introduction

Transportation planning plays a critical role to the long-term vision of a state, region, or community. It evaluates a wide range of potential strategies. various points of view, as well as the participation of transportation-related agencies and organizations in the evaluation process. Basic element of transportation planning consists of several processes that begin with identifying the regional vision and goals, then alternate and evaluate the main strategies, and finally develop a system performance [1]. Travel demand relates to the movement of people, vehicles, and goods on a particular part of the transportation system and travel demand arise mainly from the need of people for work, study, leisure, social communication, and so on. Travel demand and traffic forecasting depend on accurate predictions of travelers' decisions. Long-term factors such as commercial and residential distributions and typical infrastructure directly influence travel decisions because they change slowly over time. Short-term factors such as highway traffic factors also influence traveler decisions and choices [2]. Four kinds of decisions are made in a typical individual's daily decision-making process: trip generating, trip distribution, mode choice, and route assignment, which are the primary stages in the travel demand and are critical in transportation planning. The analyzes of modal split (mode choice analyzes) are the third step the four-step transportation forecasting model after trip generation and trip distribution but before the assignment of routes. Mode choice is one of the most important steps in the transport planning process, it relates to the type of mode that travelers use to reach their destination and has an important effect on the decision-making process. Mode of choice models examine the human decision-

* Corresponding author. : Omaima A. Yousif ; : omaima.yousif@uoanbar.edu.iq ; +9647813458932

making behavior directly and, as a result, continue to draw researchers into the decision-making process of commuters [3]. Multiple linear regression (MLR), often known as multiple regression, is a statistical technique that predicts the result of a response variable by combining numerous explanatory variables. Multiple linear regression attempts to represent the linear relationship between explanatory (independent) and response (dependent) variables.to build a model and analyze the factors affecting mode choice in Ramadi city using multiple linear regression analysis. Many researchers have developed mode choice prediction models worldwide, Khudhair and Alsadik, (2020): developed a mode choice model for students at Al-Mustansiriya University, Baghdad University, and Karkh University using a multinomial logit model. The study found that the cost of travel, travel time, gender, comfort, and safety are the most influencing factors on the choice of mode of transport, which will help the concerned authorities to try to improve public transport services and contribute to sustainability by transferring students traveling from private transport to public transport [4]**.** Elharoun etal.,(2018): studied a behavioral model for choosing the mode choice for Mansoura city in Egypt using the multinomial logit model**.** The results of the research showed that the most influential factors on mode choice are travel time, total cost, car ownership, driving license and monthly income. The results gave good prediction accuracy, which will help in developing policies and solutions to reduce overcrowding in Mansoura city, which is considered one of the most crowded cities in Egypt[5].Goel and Sinah, (2014): developed a model to analyze the mode of trans choice behavior for work trips in Delhi urban areas. To do this Adaptive Neural fuzzy inference system was used. The results showed that the size of the fleet and the level of service were found to have a significant impact on the mode of choice for work trips [6]. Al-Ani (2010): developed a model for the factors affecting the mode choice in Ramadi city using the technique of statistical analysis by adopting discriminant analysis. The results show that cost and travel time are the important factors affecting mode choice [7]. Chalumuri et al., (2009): developed a mode choice model in Visakhapatnam and Negpur cities in India using artificial neural networks and comparing it with the multinomial logit model The results of the analysis showed that the artificial intelligence method gives better prediction accuracy than the logarithm method [8].

## 2. Methodology

### 2.1: Study Area

The study area is Ramadi city, which is the political and administrative center of Anbar Governorate and is located in the southeast. The city's basic design area is estimated to be around 143335 km$^2$, and it comprises of thirty residential communities spread across the city's basic design map. For the purpose of study and data collection, it was divided into traffic analysis zones as shown in Figure 1, which illustrate the location map and traffic zones of Ramadi City. A city is located between latitudes (33$^o$-23$^o$) and (27$^o$-33$^o$) north and longitudes (12$^o$-43$^o$) and (20$^o$-43$^o$) east. Ramadi city is located 46 kilometers east of Fallujah, 160 kilometers east of Haditha, 270 kilometers north of Al-Qaim, and 60 kilometers away from Heet, 300 kilometers west of Al-Rutba [9]. The importance of the location of Ramadi lies in its association with various cities of the governorates, cities and the governorate itself, and neighboring countries with tiled roads and the city is the only passage for oil tankers to cross towards Syria and Jordan.

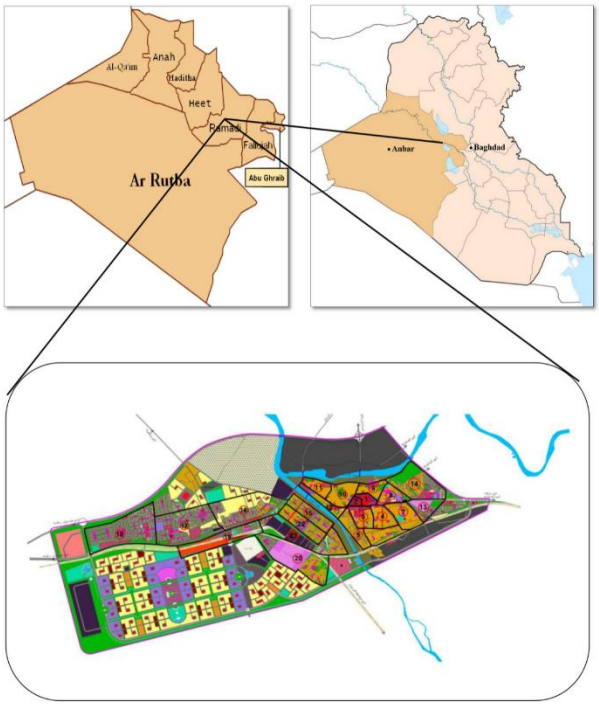
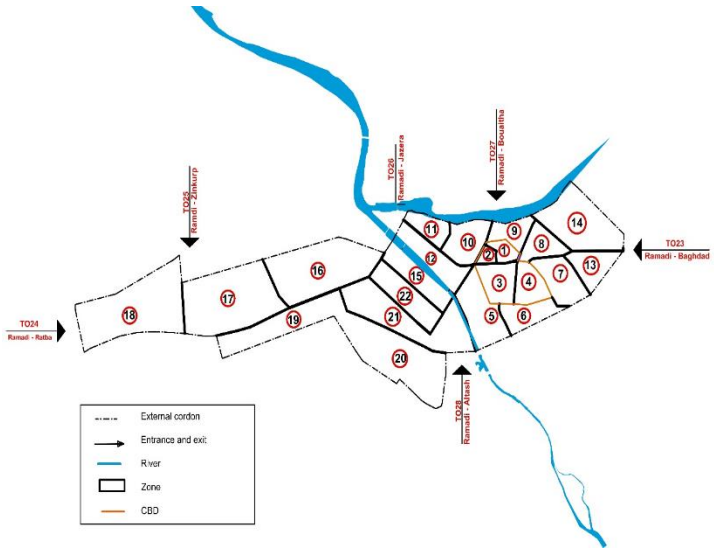


**Figure 1:** Location Map of Study Area (Ramadi City)

**Figure2**: Map of Zones of Study Area

### 2.2: Study Area Boundary and Traffic Zones

The study area was divided into traffic analysis zones to obtain data on future travel demand and to enable us to link information about travel activity, travel, and transportation between the study area zones (TAZ). As shown in Figure 2, the total number of traffic zones was 28 traffic zones, 22 of which were internal traffic zones and 6 of which were external traffic zones. Because zone 19 and zone 21 represent the expansion area and industrial zone, respectively, and there are no households in these zones, the research was confined to 26 zones. The imaginary line representing the boundary of study area is termed as Cordon Line and the survey done inside the area covered by cordon line to study travel pattern to large extent is known as Cordon Line Survey. The area inside cordon line is studied extensively whereas the area outside cordon line is studied in a lesser degree of detail.




## 2.3 Data Collection

The purpose of the field survey for this study is to obtain data by selecting a random sample of housing units to represent the biggest total number of people with certain characteristics. In this study, the field survey was conducted by conducting direct home interviews with family members in their homes within Ramadi's traffic zones. Data can be collected in a variety of ways, the surveys being one of the most essential and widely used. The household interview questionnaires are distinguished by a high response rate (70-80%) and the accuracy of the data provided Figure 3 shows the survey form used in the data collection process. The sample size was determined at random, then calculated using the appropriate statistical process before being distributed to the household. The size of the sample to be interviewed is determined by the overall population, the level of accuracy necessary, and the population density in the study area. The following is the statistical formula for calculating sample size, which can be found in most statistical textbooks [10]

$$n= X^2 \times N \times P(1-P) \div \{ME^2\} \times (N-1) + \{X^2 \times P \times (1-P)\} \quad \ldots\ldots\ldots(1)$$

Where n is sample size; N is Population size; p is Population proportion (equal to 0.5); $X^2$ is the Chi-square at 1 degree of freedom for the specified confidence level. This computation is based on a normal Gaussian distribution with more than 30 samples, according to the assumption. And according to margin of error (2.5%) 95 % confidence level, and 50 percent response distribution, the population size in this study is (73215) households. As a result, the sample size is 1505 samples. According to the variables that were adopted in building the model, the questionnaire includes a question about the following items; first, Characteristics of trip makers such as Age, Gender, Income per household, car ownership per household, and Family size. Second Trip characteristics such as: Trip purpose (work, shopping, education, recreation, other), Trip origin and destination, and the beginning and end time of trip. Third; mode Characteristics such as: Type of mode used, the cost of trip, waiting time, and time inside mode.


استمارة مسح لدراسة النقل لمدينة الرمادي ٢٠٢١

جامعة الانبار/ كلية الهندسة ..القسم المدني /الدراسات العليا

إن هذه الاستمارة جزء من متطلبات رسالة الماجستير وهدف الرسالة تطوير النقل في مدينة الرمادي راجين تعاونكم معانا

عنوان السكن

عدد افراد الاسرة

معدل دخل الأسرة الشهري

ملكية الاسرة للمركبة (السيارة): ١- لاتملك سيارة ٢- تملك١سيارة ٣-تملك سيارتين فأكثر

عدد الرحلات التي تقوم بها الاسرة خلال اليوم ( كم شخص يخرج من المنزل

يرجى ملئ الجدول مع مراعاة ما يلي:

- يوضع الرقم الخاص بوسيلة النقل في الحقل المخصص له
- يوضع رقم الغرض من الرحلة في المكان المخصص له

| وسيلة النقل | | | | الغرض من الرحلة | | | |
|---|---|---|---|---|---|---|---|
| ١-سائق سيارة خاصة ٢-راكب سيارة خاصة ٣-راكب سيارة حكومية ٤-راكب تاكسي ٥-راكب باص خاص(تشمل باصات المدارس والمستاريكس ) ٦-راكب باص عام ٧-راكب دراجة نارية ٨-راكب دراجة هوائية ٩- مشيا ١٠-اخرى | | | | ١-رحلة عمل ٢ -رحلة تسوق ٣-رحلة تعليم ٤-رحلة ترفيهية ٥- رحلة اخرى | | | |
| عمر الشخص الذي يقوم بالرحلة خلال اليوم ( اذكر اي شخص يقوم برحلة من افراد اسرتك) | جنسه 0 ذكر 1 انثى | وقت الخروج من المنزل | وقت الوصول للمكان المطلوب | نوع وسيلة النقل التي تستخدمها | الغرض من الرحلة التي تقوم بها | كلفة الرحلة (دينار عراقي) | عنوان بداية الرحلة | عنوان نهاية الرحلة |
| ١ - | | | | | | | | |
| ٢- | | | | | | | | |
| ٣- | | | | | | | | |
| ٤- | | | | | | | | |
| ٥- | | | | | | | | |
| ٦- | | | | | | | | |

إذا استخدمت التكسي او الباص: ما زمن الانتظار لوسيلة النقل دقيقة

ما الزمن المستغرق داخل وسيلة النقل دقيقة

**Figure 3:** Survey form used in the data collection process

## 2.4 Modal Split modelling Using Multiple linear Regression

The multiple linear regression (MLR) technique is applied by using SPSS statistical package to construct mathematical relationships between the type of mode (dependent variable) and the social-economic features (independent variables) as shown in table 1. The stepwise method is the most commonly used to create basic forecasting regression models for each independent parameter. Stepwise regression selects variables in a sequential manner. One by one, independent variables are added or removed based on their statistical significance. Stepwise either adds or subtracts the most significant variable. When the method is finished, it outputs a single regression model [11].

**Table 1: Dependent and independent variable used in regression analyzing**

| Independent Variables | NO |
|---|---|
| $X_1$= Age | 1 |
| $X_2$= Gender | 2 |
| $X_3$= Income | 3 |
| $X_4$= Car ownership | 4 |
| $X_5$= Beginning time | 5 |
| $X_6$= End time | 6 |
| $X_7$= Trip cost | 7 |
| $X_8$= Trip origin | 8 |
| $X_9$=Trip destination | 9 |
| $X_{10}$= Waiting time | 10 |
| $X_{11}$= Time inside mode | 11 |
| $X_{12}$= Family size | 12 |
| **Dependent Variable** | |
| Y= Type of mod | |

## 3.Results

A preliminary analysis of data collected from a field survey has been carried out prior to the development of a model to better understand the travel behavior of commuters in the Ramadi city. The data analyzed could help in decision-making in the future as a result of model development. The figures that follow will summarize the conclusions of the data collected from Ramadi's various zones. From Figure 4, it could be concluded that the percent private car seen to be the most common transportation mode, contributing for 32.2% (as a driver and passenger) of the overall transportation system supply. Taxies contribute for around 24.6% percent of total transportation modes. private bus (includes school buses and starex) contribute 12.2%, while government and public transportation each contribute 2.07% and 1.5% respectively. The most interesting aspect is that walking accounts about 14% of the overall transportation system.

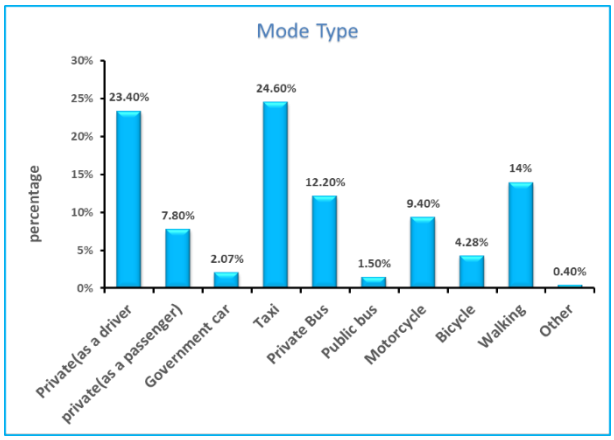



**Figure 4:** Mode Type in Ramadi City

The gender distribution of travelers may be seen in Figure 5, Males represent for 69.9% of respondents, while females represent for 30.1%.

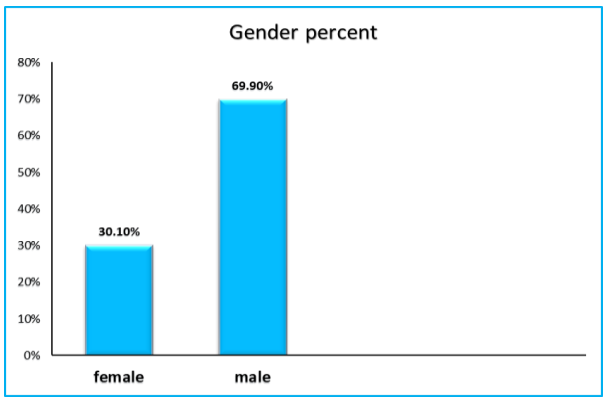



**Figure 5:** Age Percent in Ramadi City

Figure 6 shows that work trips are the most of all trips, followed by educational, shopping, recreation, and other trips.

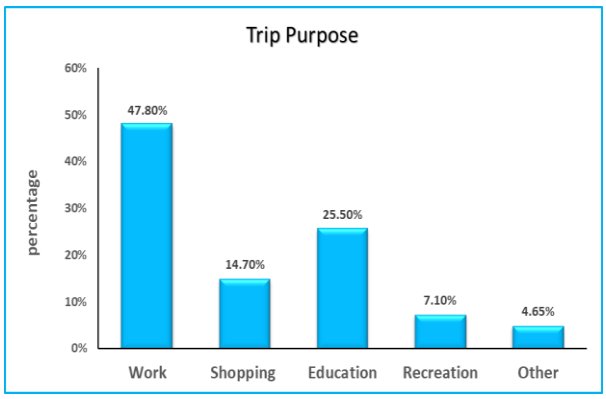



**Figure 6**: Kinds of Trip in Ramadi City

Figure 7 shows the percentage of trips attracted to each zones of Ramadi city, it is clear that zone No.1, which represents the Qattaneh zone, is the most attractive zone for trips, due to the concentration of commercial uses and private services in it, at a rate of 17.8%, and then comes zone No. 20, which represents the university district (University of Anbar), which attracts people for the purpose of study, at a rate of 10.8%. Then comes zone No.3, which represents Al-Andalus district, which is characterized by containing health service centers. It also includes Al-Maustawda Street, which is a commercial service street that people go to to meet their needs and desires

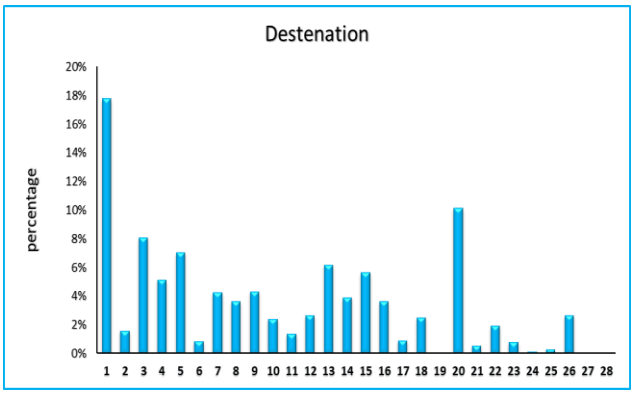



**Figure 7:** Percent of Trips Attracted to Ramadi Zones

Modal split model based on multiple linear regression is proposed to estimate the mode choice behavior of individuals by looking at the variables affecting the mode choice, namely age, gender, family income, car ownership, family size, travel cost, start and end time of the trip, trip origin and destination. The data collected from the home interview was used as inputs to the model. Iteration was conducted by decreasing the independent variables with lower t-values. The results showed that the most influential factors mode choice is the car ownership, the age of the person making the trip, the cost of the trip, followed by the the origin of the trip, trip time, waiting time, time spent inside the mode, family size, gender and trip destination. Table 2 shows the coefficients of the regression model, which help in determining the equation of the regression line between the parameters, which is as follows:

$$Y(\text{Forecasted})=8.02-0.059X_1+0.27X_2-1.52X_4+0.01X_5-0.031X_8-0.015X_9-0.218X_{10}+0.031X_{11}+0.056X_{12} \quad \ldots\ldots\ldots\ldots\ldots (2)$$

Two factors were excluded from the model, namely the monthly household income and the end time of trip, these variables are statistically non-significant and so removed from the model. Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100) as shown in Table 3

**Table 2: The coefficients of the regression modal split model**

| Model | Unstandardized Coefficients | | Standardized Coefficients | t | sig |
|---|---|---|---|---|---|
| | B | std error | Beta | | |
| (Constant) | 8.020 | .250 | | 32.128 | .000 |
| CAROWNER | -1.52- | .069 | -.361- | -22.01- | .000 |
| AGE | -.059- | .003 | -.275- | -16.95- | .000 |
| COST | .000 | .000 | -.242- | -11.92- | .000 |
| ORIGON | -.031- | .007 | -.072- | -4.289- | .000 |
| BEGINNINGTIME | .001 | .000 | .058 | 3.509 | .000 |
| WATINGTIME | -.218- | .030 | -.177- | -7.235- | .000 |
| TIMEINSIDEMODE | .031 | .005 | .174 | 6.440 | .000 |
| FAMILYSIZE | .056 | .022 | .040 | 2.510 | .012 |
| GENDER | .270 | .107 | .040 | 2.51( | |
| DESTINATION | -.015- | .007 | -.035- | -2.13( | |

**Table 3: Excluded Variables from multiple linear regression**

| Model | | Excluded Variables | | | | |
|---|---|---|---|---|---|---|
| | | Beta In | T | Sig. | Partial Correlation | Co linearity Statistics Tolerance |
| 1 | Income/HH | -.012-[k] | -.606- | .545 | -.011- | .656 |
| | End time of trip | .017[k] | .694 | .488 | .013 | .381 |

The regression results for the estimated general modal split model are summarized in Table 4. The correlation coefficient between the dependent variable and the independent variables and its value 0.837 as well as the square of the correlation coefficient and its value 0.829 and the standard error of estimation 0.195. Thus, the independent variables explain 82.9% of the variance of the dependent variable**.**

**Table 4: Regression results for the general modal split model**

| Model | R | R Square | Adjusted R square | std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .83 | .829 | .828 | 0.195 |

The regression standardized residual is shown in Figure 7. The residuals represent the difference between the predicted values and the expected values, and it is clear from the figure that it follows a normal distribution.

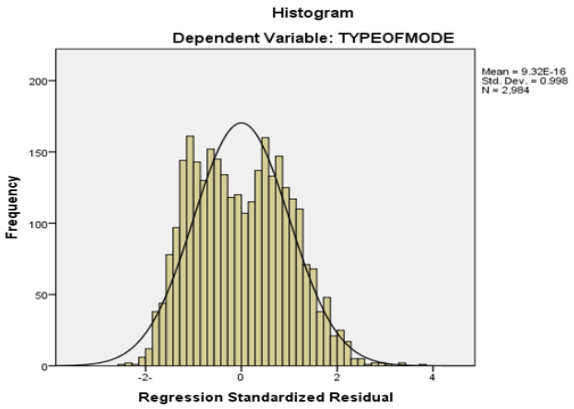



**Figure 8:** The regression standardized residual for general modal split model

## Conclusion

1-The results of the preliminary analysis of the data showed that the percentage of using private cars in the city reached 31.2%, while taxis account for 21.6%, and public transportation constitutes a very small percentage and is almost non-existent. This results in air pollution and congestion, especially at peak times,

so the city needs to develop policies and solutions to provide comfortable and affordable public transportation.
2-Most of the trips go to the Central Business District (CBD) of Ramadi and to the university district zone represented by the University of Anbar (zone 1 and 20 respectively), as a result, while developing the city's transportation system, these zones should be well considered through the use of bus stops, waiting cabinets, vehicle parks, suitable roadways, and traffic flow.
3-Due to the high proportion of students and workers in families, the number of work and education trips formed the highest percentage of total daily trips.
4-The most influential factors on the mode choice for the general trips model using multiple linear regression are car ownership, age and trip cost. This model gave a good correlation coefficient of 0.829 meaning that the independent variables explain 82.9 of variance in the dependent variable (type of mode). Which will help transport planners in developing policies and solutions, the most important of which is improving public transport and providing comfortable and affordable buses, especially in areas that attract trips, which will help reduce congestion, especially at peak times.

## References


1-WRIGHT, P. H. & PAQUETTE, R. J. 1987. Highway engineering.

2-MANNERING, F. L., WASHBURN, S. S. & KILARESKI, W. 2004. 'Principles of Highway Engineering and Traffic Analysis', John Viley & Sons. Inc, USA.

3-MINAL & SEKHAR, C. R. 2014. Mode Choice Analysis: The Data, the Models and Future Ahead. International Journal for Traffic and Transport Engineering, 4, 269-285.

4-KHUDHAIR, H. A. & ALSADIK, S. M. Development of Mode Choice Models for Undergraduate Students in Baghdad City. AWAM International Conference on Civil Engineering, 2019. Springer, 1443-1458.

5-ELHAROUN, M., SHAHDAH, U. E. & EL-BADAWY, S. M. 2018. Developing a Mode Choice Model for Mansoura City in Egypt. International Journal for Traffic & Transport Engineering, 8, 528-542.

6- GOEL, S. & SINHA, A. K. An improved modeling of mode-choice behavior in urban area using adaptive neural fuzzy inference system. 2014 International Conference on Computing for Sustainable Global Development (INDIACom), 2014. IEEE, 286-290.

7- Al_Ani, N. A,2010. Developing A Model of Mode Choice in Ramadi city. MSc Thesis, University of Anbar, Iraq.

8- CHALUMURI, R. S., ERRAMPALLI, M., BUJANGAN, K. D. & SUBAMAY, G. Applications of neural networks in mode choice modelling for second order metropolitan cities of India. Proceedings of the Eastern Asia Society for Transportation Studies Vol. 7 (The 8th International Conference of Eastern Asia Society for Transportation Studies, 2009), 2009. Eastern Asia Society for Transportation Studies, 134-134.

9-Aleisawy, Y. E. S, 2019. Traffic efficiency in Ramadi city. MSc Thesis, University of Anbar, Iraq.

10- BLUMAN, A. G. 2012. Elementary statistics: A step by step approach, McGraw-Hill.

11- BONNEL, P. Estimation of aggregated modal split mode. AET. European Transport Conference 2003–ETC 2003, 8-10 october 2003, Strasbourg, 2003. AET, 23