A bus demand model for Low-Density Territories in Continental Portugal

HARLEY LARGO, PAULO J. G. RIBEIRO
Department of Civil Engineering
University of Minho
Campus de Azurém, 4800-058
GUIMARÃES, PORTUGAL
harleylargoamado94@hotmail.com, pauloribeiro@civil.uminho.pt

Abstract: Continental Portugal has 278 municipalities with 164 being classified as low density territories (LDT), according to criteria mainly centered in population density and per capita income. LDT's are characterized as territories with economic and labor problems, which also have suffered a significant reduction in resident population. Thus, in the last decade, studies have been developed in Portugal for these territories to enhance the quality of life and living conditions. For several reasons, public transport represents an important mode of transport to guarantee cohesion and equity among different groups of population. Thus, it is important to characterize the demand patterns of public transport in LDT, in order to better plan and promote its use, especially for bus services. Therefore, this paper presents a model to estimate the demand of a bus transportation in low-density areas of Portugal. The mathematical model used to estimate the demand was the multiple linear regression (MLR) models, which is a function of the most relevant and influential socioeconomic and demographic variables for LDT. The MLR model were developed with the statistical tool SPSS (Statistical Package for the Social Sciences). It is important to highlight that were created three groups of Portuguese municipalities according to the population density to create adjusted demand model for bus services in LDT. The bus demand MLR model presented a low level of adjustment, probably due to the amount of data used to estimate each model. Results shown that the model that have a better adjustment to estimate the number of bus trips was achieved for the group of a population density lower than 50 inhabitants/Km^2 that was supported by two variables: illiterate people and the number of unemployed. Thus, future works must estimate bus trips through another estimation approach for transport demand in low-density territories.

Key-Words: Low-density territories; demand model; public transport; bus service; multiple linear regression; SPSS.

1 Introduction

The transport system is dynamic and integrated by a series of interrelated elements, which aims to provide a service that allows the efficient, economic and safe movement of people and goods. One of the subsystem is the bus system, which has undergone through significant changes, namely in terms of supply, demand, management, and operation and more recently in relation to the transportation authority's responsible for this sector.

One of the biggest problems with the transportation system, which has been increasing over time, is caused by high levels of demand (users) in relation to the transportation supply. This problem is not different in the LDT's, where according to Fernandes [2] the problem of mobility and equity in the population's access to goods and services in these territories is a catalyst for aggravating inequalities and the phenomena of social exclusion.

The concept of Low-Density Territories varies according to the guidelines of the public policies of a region or country. Some indicators that allow to define a low-density territory are the population density, the per capita GDP, the index of dependence of the elderly and the index of aging of the population, the fertility rate and the variation in population [3].

In order to estimate the demand for transport in LDT, it is necessary to know what are the main problems in these territories, especially in demographic, economic and social domains. According to Domingues [4], the LDT's present accessibility problems related with the difficulty of access to employment, services and other types of assets by the resident population. In addition, Litman [5] states that demographic, geographic and economic factors can affect travel demand and have interactive effects on people's travel behavior. Some examples of these factors are the number of people (residents, employees and visitors), employment, education, age/life cycle, lifestyles, among other cultural aspects.

According to [6]–[8] the multiple linear regression method is one of the most used demand methods since its general use presents values more similar to the actual behavior. Thus, it is important to determine which variables should be taken into account for the estimation of transport demand.

The definition of the variables is one of the most important works in the development of a demand model for bus transportation. Juan de Dios Ortuzar [9], shown that the exogenous variables are the "cornerstone" of the transportation demand model since they have socioeconomic and demographic information of the area to which the model will be applied. Curtis & Perkins [10] studied the impact of demographic variables on travel behavior and found some relevant relationships between mode choice and variables such as age, gender, vehicle ownership, structure of residential population and income. Dieleman [11] concluded that households with higher wages have higher levels of ownership and use of the motor vehicle (car). In addition, families with children are more likely to use the car regularly, and age groups and sex are, in fact, one of the factors that affect travel behavior and patterns.

2 Methodology

2.1 Definition of variables

According to literature research carried out by Amado [12], the variables that have influence to characterize the dynamic travel behavior in low-density territories, in mainland Portuguese municipalities are the following:

- People between the ages of 15 and 24 and 55 or over;
- Number of women's trips;
- Number of unemployed;
- Number of persons without driving license;
- Number of vehicles per dwelling;
- Average income per family (monthly);
- Number of trips for purchases or social purposes;
- Travel time;
- Level of education;
- Walking time;
- Number of dwellings with one person.

2.2 Data collection

The data collection for the municipalities classified as low-density territories (population) was very dependent on the data available and presented in the 2011 Portuguese Statistics (INE). The Census 2011 were the main source of information for the characterization of the population and territories, especially in social, economic and environmental domains.

2.3 Definition of the groups of municipalities

In this study, it was assumed that there could be a more homogenous behavior for territories with similar population levels, i.e., for territories with the same levels of population density.

Thus, in order to estimate the demand for public transportation by bus, 3 groups of population densities were proposed according to the histogram analysis (Figure 1), resulting in the following classes:

1. $0 \leq$ Pop. density ≤ 50 hab/Km^2
2. 50 hab/Km$^2 <$ Pop. density ≤ 100 hab/Km^2
3. Population density > 100 hab/Km^2

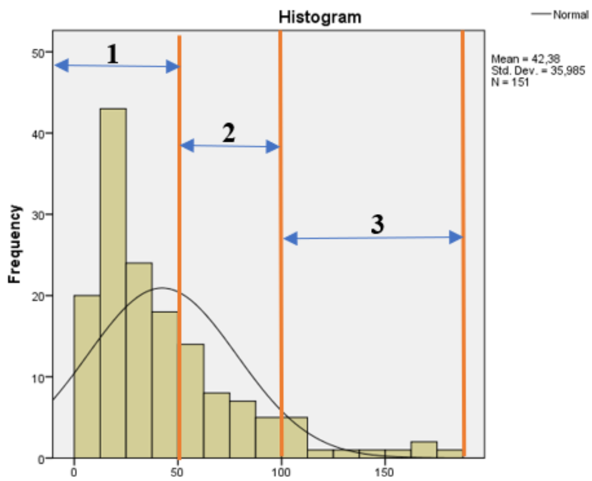

Fig. 1 – Definition of groups of LDTs

The selection of the sample of LDTs to define the bus demand model for each of the group was determined based on the total number of municipalities defined by CIC Portugal 2020 [1]. In addition, an outlier analysis was applied to establish which municipalities can be assumed as extreme values and consequently being removed from the analysis.

On the other hand, to define the bus demand model for LDT, 85% of municipalities were selected from each of the groups, using the random method, while the others (15%) were used in the validation process for the proposed model.

2.4 Multiple Linear Regression (MLR) model

When there are several exogenous variables, the model is called the Multiple Linear Regression (MLR) model, and it is generically represented by:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_j X_{ji} + \cdots + \beta_k X_{ki} + \varepsilon_i \qquad (1)$$

$i = 1,2,\ldots,n$

- n: sample size;
- k: Number of observable exogenous variables added to the constant, where X's and Y are observable variables
- εi: Non-observable and random exogenous variable, which includes all influences in Y that are not explained by X's;
- βk: are parameters of the model, that is, quantities that always assume the same value in it.

2.4.1 Model estimation

The parameters / variables to be used in the model based on the information obtained with the information resulting from point 2.2 are presented, i.e., the equation of the model for each group is going to determine the variables that are significant to be used in the bus demand model for LDTs, being defined as:

$$\begin{aligned} \text{VGA} = {} & \hat{B}_o + \hat{B}_1(Dsgd) + \hat{B}_2 PdC + \hat{B}_3 DP \\ & + \hat{B}_4 IM + \hat{B}_5 GM + \hat{B}_6 SnE \\ & + \hat{B}_7 ES + \hat{B}_8 IP \end{aligned} \qquad (2)$$

VGA: Number of trips by bus;
Dsgd: Number of unemployed;
PdC: Purchasing power per capita;
DP: Population density;
IM: Number of women over 15 years of age;
IP: Number of people between the ages of 15 and 24 and over 50;
GM: Average monthly income;
SnE: Number of illiterate people;
ES: Number of people with higher education degree.

To estimate the model, it was used the SPSS software. According to Pestana and Gageiro [14], the selection of the significant variables for the model it will be the stepwise method, since it allows solving multicollinearity problems. The stepwise consists of entering the explanatory variable that presents the highest coefficient of correlation with the independent variable (number of trips by bus). Then, the partial correlation coefficients are calculated for all the variables that are not part of the first regression. Thus, the next variable that enters is the one with the highest partial correlation coefficient. The new regression (equation) is estimated and is analyzed whether one of the two independent variables should be excluded from the model. At the end, if both variables have significant *t-values*, new correlation coefficients are calculated for the variables that did not enter. The process ends, when it reaches the situation where no variables should be added to the equation [15].

A model with many explanatory variables will be more difficult to be interpreted since the effects of some of the variables will be small, so it will be better to determine a model with fewer variables, but with stronger explanation capacity, i.e., that has a smaller predictive error when compared with a model with all variables [9].

2.5.2 Statistical diagnosis of the MLR model

The statistical models are based on assumptions that allow a better approach to the problems, including MLR. In this way, the assumptions considered in such analysis will be presented, taking into account only the most relevant aspects of MLR [13], such as:

- Linearity;
- Independence;
- Homoscedasticity;
- Normality.

This basic MLR model conditions will be analyzed for each of the estimated models.

2.5.3 Validation of the model for each TBD group

The validation of the bus demand model for low-density territories consists in verifying that the models fulfill their function. In other words, that the model provides reasonable and acceptable results to forecast the demand behavior of bus usage, i.e., verify if the modelled values are similar to real values. In order to carry out this validation, 15% of the total municipalities were randomly selected for each group.

For each municipality of the three LDT groups, the corresponding model will estimate the number of trips in bus (estimated Y) through the equation (2) of

Table 1 – Model summary for Group 1

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					
					R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	0,746^a	0,556	0,549	62,481	0,556	73,903	1	59	0	
2	0,790^b	0,625	0,612	57,93	0,069	10,633	1	58	0,002	1,878

a. Predictors: (Constant), Number of illiterate people

b. Predictors: (Constant), Number of illiterate people, Number of unemployed

c. Dependent Variable: Number of trips by bus

the MLR model, and then compared with the actual value of the number of trips in bus, which will allow to determine the percentage of error of each estimation. Based on this calculation, the frequency of errors will be determined for the validation data in each group of LDT, in order to evaluate if the models are relevant for the estimation of bus trips in low density territories.

3 Case study

The 59% of the municipalities in Continental Portugal are considered low-density territories [1]. These municipalities are located in the majority inner country. As already mentioned in the previous sections, municipalities are clustered into three groups, depending on their population density. The number of municipalities for each group is shown in table 1.

Table 2 – Definition of the groups

ID TEAM	Population Density (PD)	Number of municipalities
1	$0 \leq PD \leq 50$	80
2	$50 < PD \leq 100$	49
3	$PD > 100$	22

To estimate the MLR for each group of municipalities were used the nine variables presented in the equation (2), according the work of Amado [12]. MLR must avoid multicollinearity, especially among explanatory variables. Thus, it was determined the intensity of association between all the variables using the statistical measures R of Pearson and Rho of Spearman, being classified as shown in table 2.

Table 3 – Bivariate correlation analysis

	Density	# of people between the ages of 15 and 24 and over 50	unemployedo	# of women over 15 years of age	# of illiterate people	# of people with higher education	Average monthly gain	Purchasing power
Density								
Number of people between the ages of 15 and 24 and over 50	Moderate ,428							
unemployedo	Moderate ,484	Strong ,887						
Number of women over 15 years of age	Moderate ,485	Very strong ,994	Very strong ,911					
Number of illiterate people	Moderate ,414	Very strong ,983	Very strong ,901	Very strong ,983				
Number of people with higher education	Moderate ,488	Strong ,933	Strong ,893	Muito forte ,948	Muito forte ,903			
Average monthly gain	very weak ,031	very weak ,180	Weak ,297	Weak ,203	very weak ,158	Fraca ,344		
Purchasing power	very weak ,03	Weak ,0236	Weak ,327	Weak ,257	Weak ,212	Moderate ,405	Moderate ,613	
Number of trips by bus	moderate ,579	Strong ,836	Strong ,857	Strong ,887	Strong ,875	Strong ,813	very weak ,110	very weak ,073

The variables that will be taken into account in the estimation of the three demand models were the ones that showed a stronger relation with the dependent variable (number of bus trips). From the results of table 2 it can be concluded that economic variables were not used in model estimation due to the weak relationship with the dependent variable.

3.1 Estimation of MLR model for Group 1 (0≤PD≤50 inhabitants/Km^2)

The number of municipalities with a population density of less than 50 inhabitants/ Km^2 is 80 (53% of the total number of municipalities considered as low density territories). Table 3 shows the summary of the MLR model showing that 63% of the behavior of the number of bus trips is explained by MLR model, in addition to the adjusted coefficient (Adjusted R Square), which determines the quality of the adjustment, presents a value of 0.61, allowing to consider that this MLR model presents a moderate level of adjustment.

Table 4 – Model Coefficients for Group 1

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficienst Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part	Colinearity Statistics Tolerance	Colinearity Statistics VIF
1	(Constant)	-6,34	28,102		-0,226	0,822					
	Number of illiterate people	0,165	0,019	0,746	8,597	0,000	0,746	0,746	0,746	1,000	1,000
2	(Constant)	-28,291	26,911		-1,051	0,297					
	Number of illiterate people	0,121	0,022	0,547	5,427	0,000	0,746	0,58	0,436	0,636	1,572
	Number of unemployed	0,316	0,97	0,329	3,261	0,002	0,659	0,394	0,262	0,636	1,572

a. Dependent Variable: Number of trips by bus

Table 5 - Model summary for Group 2

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics R Square Change	Change Statistics F Change	Change Statistics df1	Change Statistics df2	Change Statistics Sig. F Change	Durbin-Watson
1	0,557[a]	0,311	0,29	157,922	0,311	15,33	1	34	0,000	
2	0,658[b]	0,432	0,398	145,469	0,122	7,071	1	33	0,012	1,800

a. Predictors: (Constant), Number of women over 15 years of age

b. Predictors: (Constant), Number of women over 15 years of age, Purchasing power per capita

c. Dependent Variable: Number of trips by bus

Table 6 - Model Coefficients for Group 2

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficienst Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part	Colinearity Statistics Tolerance	Colinearity Statistics VIF
1	(Constant)	53,172	120,735		0,44	0,662					
	Number of women over 15 years of age	0,81	0,021	0,557	3,915	0,000	0,557	0,557	0,557	1,000	1,000
2	(Constant)	472,659	193,019		2,449	0,02					
	Number of women over 15 years of age	0,92	0,019	0,634	4,723	0,000	0,557	0,635	0,619	0,954	1,048
	Purchasing power per capita	-6,519	2,452	-0,357	-2,659	0,012	-0,221	-0,420	-0,349	0,954	1,048

a. Dependent Variable: Number of trips by bus

In Table 4, it is possible to analyze the significance value for the *t* test, which is less than 0.05 for both variables, having a significant explanatory power in the estimation of number of trips by bus. However, for the value of the constant the significance of the *t* test is considerably higher than 0.05, so the use of this model for inference purposes must be cautious.

The estimated model for municipalities with a population density of less than 50 inhabitants/ Km^2 is:

$$VGA = -28{,}291 + 0{,}121 * SnE + 0{,}316 * Dsgd \quad (3)$$

Where:

VGA - Number of trips by bus;

SnE - Number of illiterate people, with SnE varying between [645; 2200];

Dsgd - Number of unemployed, with Dsdg varying between [91;494].

3.2 Estimation of MLR model for Group 2 (50<PD≤100 inhabitants/Km²)

The number of municipalities in this category is 49, this is equivalent to 32% of total municipalities. Table 5 shows that 43% of the behavior of the dependent variable (number of trips by bus) is taken into account by MLR model for the Group 2, which has an Adjuted R Square of 0.40, that is, there is an adjustment that is not very strong.

Table 6 shows that the explanatory variables are different from the MLR model found for Group 1, i.e., women over 15 years-old have a greater influence on the prediction or estimation of the dependent variable, followed by the purchasing power of the resident population. The significance value for the statistical inference test (test t) is less than 0.05 for both variables, so, the null hypothesis is rejected, which means that the variables have explanatory capacity for the dependent variable and coefficients values are not null.

Table 7 - Model summary for Group 3

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	0,812[a]	0,66	0,639	192,634	0,66	31,046	1	16	0,000	
2	0,873[b]	0,763	0,731	166,116	0,103	6,516	1	15	0,022	2,341

a. Predictors: (Constant), Log_Number of unemployed persons
b. Predictors: (Constant), Log_Number of unemployed persons, Purchasing power per capita
c. Dependent Variable: Number of trips by bus

Table 8 - Model Coefficients for Group

		Unstandardized Coefficients		Standardized Coefficienst			Correlations			Colinearity Statistics	
Model		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-4939,93	1096,14		-4,507	0,000					
	Log_Number of unemployed persons	1947,775	349,571	0,812	5,572	0,000	0,812	0,812	0,812	1,000	1,000
2	(Constant)	-4817,07	946,466		-5,09	0,000					
	Log_Number of unemployed persons	2110,213	308,092	0,88	6,849	0,000	0,812	0,87	0,861	0,957	1,045
	Purchasing power per capita	-8,169	3,2	-0,328	-2,553	0,022	-0,146	-0,550	-0,32	0,957	1,045

a. Dependent Variable: Number of trips by bus

3

The estimated model for the municipalities of Group 2 with a population density between 50 and 100 inhabitants/ Km^2 is given by:

$$\mathrm{VGA} = 472{,}66 + 0{,}92 * \mathrm{IM} - 6{,}519 * PdC \quad (4)$$

Where:

VGA: Number of trips by bus;

IM: Number of women over 15 years of age, with IM varying between [4148; 8052];

PdC: Purchasing power per capita, with PdC varying between [58,25; 95,96].

3.3 Estimation of MLR model for group 3 (PD>100 inhabitants/Km^2)

The number of municipalities in this category is 22, this is equivalent to 15% of total municipalities. Table 7 shows that 76% of the behavior of the dependent variable (number of trips by bus) is taken into account by model 2, with an Adjusted R Square of 0.73, which allows to be considered as a strong adjustment. This adjustment value tends to be influenced by the sample size that is considered small, since according to [16] it is less than 30.

In Table 8 it is possible to analyze the influence of each variable in the model for group 3 and to know the effect on the prediction or estimation of the dependent variable. Thus, as for Group 2, the variables are different from those in Group 1 and 2, and in this case it was necessary to use the transformation of the variable "number of unemployed" through the use of a logarithm function. It should also be noted that the level and importance of the variables in the MLR model is very different, since the "number of unemployed" represent about 3 times more weight than the "purchasing power per capita". Conversely, there are variables with an opposite sign, that is, the number of trips increases with the number of unemployed, and decreases with the increase in purchasing power.

The estimated model for the municipalities of Group 3, with a population density greater than 100 inhabitants/ Km^2, is given by:

$$VGA = -4817{,}07 + 2110{,}213 \ \mathrm{Log(Dsgd)} - 8{,}169 * PdC \quad (5)$$

Where:

VGA: Number of trips by bus;
Dsgd: Number of unemployed persons, with Dsdg varying between [711; 2368];
PdC: Purchasing power per capita, with PdC varying between [56,08;105,70]

Table 11 - Statistical hypothesis of a MLR model

		GROUP 1		GROUP 2		GROUP 3	
		k = 2	n = 68	k = 2	n = 36	k = 2	n = 18
	Condition	Value / Range	state	Value / Range	state	Value / Range	state
Linearity	The independent variables present a linear relationship with the dependent variable	SnE: R=0,75 ; Dsgd:R=0,66	Two variables present a linear relationship	IM: R=0,56 ; PdC: R= 0,22	Only one variable presents moderate relation	Dsgd: R=0,81 ; PdC: R=0,14	Only one variable has a linear relationship
Homoscedasticity	The graphs: Residual Studentized Vs. Standardized Predicted; Standardized Residual Vs. Unstanded Predicted do not have increasing or descending tendencies	-	Meets condition	-	Meets condition	-	Meets condition
Independence	The value of the Darwin-Watson test must be within the non-rejection region of the null hypothesis	[1,668;2,332]	D-W = 1,878 meets	[1,587;2,413]	D-W = 1,800 meets	[1,535;2,465]	D-W = 2,341 meets
Normality	Residues exhibit normal behavior, ie the significance of the Kolmogorov-Smirnov or Shapiro-Wilk test is greater than 0,05	K-S: sig > 0,05	sig = 0,041 Does not comply	S-W: sig > 0,05	sig = 0,104 meets	S-W: sig > 0,05	sig = 0,2 meets
Multicollinearity	Independent variables are not correlated with each other. The Variance Inflation Factor (VIF)	VIF < 5	VIF -SnE = 1,5 ; VIF - Dsgd = 1,5	VIF < 5	VIF -IM = 1,048 ; VIF -PdC = 1,048	VIF < 5	VIF - Dsg = 1,045 ; VIF - PdC = 1,045

3.4 Diagnostics of MLR models

In this section, it was verified if the previously presented models fulfill all the statistical hypothesis of a MLR model, in order to make sure that the models can be used to estimate the bus demand in the Low Density Territories.

Thus it will be analyzed the hypotheses of homoscedasticity, independence and normality of the residual random variables and the hypothesis of multicollinearity between the independent variables.

The result for each of these hypotheses is presented in table 11, checking if the hypothesis fulfilled the condition that has to be analyzed and the value for each group. The red background shown that the hypothesis is not accomplished, and conversely the green shown that the condition is positively checked. According with the results presented in the table it can be stated that the supporting hypothesis for the MLR model are verified.

4 Validation of demand models

Validation consisted in comparing the estimated (modelled) with the real (CENSUS 2011) number of bus trips, for each group of municipalities. The estimated values resulted of the application of the MLR model defined in the equations of the sections 3.1, 3.2 and 3.3. From the differences found an error (%) was calculated for each of the municipality.

In Table 9 are presented the number of municipalities used in the validation process by interval of error. The interval of error was defined in classes with an amplitude of 5%. It is important to enhance that the number of municipalities for each group corresponds to 15% of the population of each group, resulting in 22 municipalities.

From Table 9 it can be concluded that the model that presents a better adjustment is the model of Group 1 which has 92% of the data with error less than 50%. However, it must be noted that the number of municipalities (sample size) in this model can have a strong influence in the quality of the adjustment, since this is the biggest group of the three.

The model for the group 2 presents 86% of the data with an error of less than 35% and the model for group 3 presents 67% of the municipalities with an error of less than 50%. Yet, some caution must be taken in interpreting the results of these two models since the number of municipalities is small, especially for the group 3.

Table 9 – Frequencies by interval of error for validation sample (15% = 22 municipalities)

		GROUP 1 (15% *80 =12)		GROUP 2 (15%*49=7)		GROUP 3 (15% *22 =3)	
	Error Intervals (%)	# Mun	% Acum	# Mun	% Acum	# Mun	% Acum
Validation data	0 - 5	1	8%	0	0%	0	0%
	5 - 10	4	42%	3	43%	0	0%
	10 - 15	2	58%	2	71%	0	0%
	15 - 20	1	67%	0	71%	0	0%
	20 - 25	1	75%	0	71%	1	33%
	25 - 30	0	75%	0	71%	0	33%
	30 - 35	1	83%	1	86%	0	33%
	35 - 40	0	83%	0	86%	0	33%
	40 - 45	0	83%	0	86%	0	33%
	45 - 50	1	92%	0	86%	1	67%
	50 - 55	0	92%	0	86%	0	67%
	55 - 60	1	100%	0	86%	0	67%
	60 - 65	0	100%	0	86%	0	67%
	65 - 70	0	100%	0	86%	0	67%
	70 - 75	0	100%	0	86%	0	67%
	75 - 80	0	100%	0	86%	0	67%
	80 - 85	0	100%	0	86%	0	67%
	85 - 90	0	100%	0	86%	0	67%
	90 - 95	0	100%	1	100%	0	67%
	95 - 100	0	100%	0	100%	0	67%
	>100	0	100%	0	100%	1	100%

6 Results

Figure 2 presents the distribution of the municipalities by the percentage of error and by the corresponding group. From the figure it can be analyzed that the majority of the results (75% of the municipalities) present an error of less than 35%, thus it can be concluded that the estimation of bus trips for each group has a good prediction quality. However, some results show an error greater than 100%, and an isolated analysis is necessary to identify the possible cause of the error.

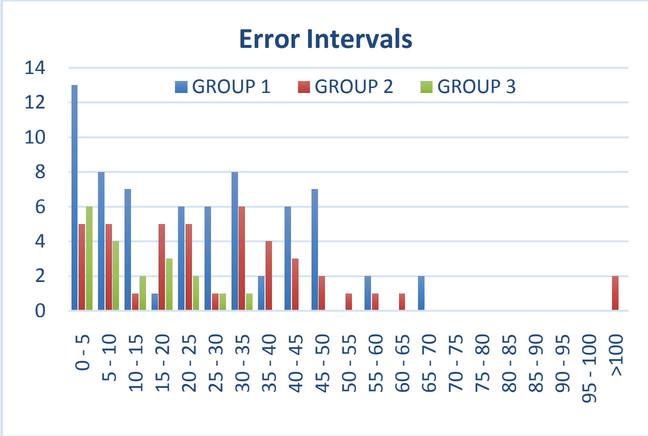

Fig. 2 – Distribution of the number of municipalities by interval of error

Table 10 allows a more specific analysis of the distribution of errors for each group. From this results, 63% of the municipalities (three groups), corresponds to 81 municipalities, have a percentage error of less than 30%. This table also allows to determine the critical line which corresponds to a percentage of error of less than 50%, which corresponds to about 93% of the municipalities of the TBDs (equivalent to 120 municipalities).

Table 10 - Frequencies by interval of error for modelled sample (85% = 129 municipalities)

		GROUP 1 (85% *80 =68)		GROUP 2 (85%*49=42)		GROUP 3 (85% *22 =19)	
	Error Intervals (%)	# Mun	% Acum	# Mun	% Acum	# Mun	% Acum
Results of municipalities by groups	0 - 5	13	19%	5	12%	6	32%
	5 - 10	8	31%	5	24%	4	53%
	10 - 15	7	41%	1	26%	2	63%
	15 - 20	1	43%	5	38%	3	79%
	20 - 25	6	51%	5	50%	2	89%
	25 - 30	6	60%	1	52%	1	95%
	30 - 35	8	72%	6	67%	1	100%
	35 - 40	2	75%	4	76%	0	100%
	40 - 45	6	84%	3	83%	0	100%
	45 - 50	7	94%	2	88%	0	100%
	50 - 55	0	94%	1	90%	0	100%
	55 - 60	2	97%	1	93%	0	100%
	60 - 65	0	97%	1	95%	0	100%
	65 - 70	2	100%	0	95%	0	100%
	70 - 75	0	100%	0	95%	0	100%
	75 - 80	0	100%	0	95%	0	100%
	80 - 85	0	100%	0	95%	0	100%

	85 - 90	0	100%	0	95%	0	100%
	90 - 95	0	100%	0	95%	0	100%
	95 - 100	0	100%	0	95%	0	100%
	>100	0	100%	2	100%	0	100%

7 Conclusions

To define a Bus demand model for Low-density territories in Continental Portugal this work used a Multiple Linear Regression (MLR) model based on information of the Census 2011. In order to get a better territorial adjustment were defined three groups of territories based on the distribution of the population density: Group 1 – a density lower than 50 inhabitants/Km^2; Group 2 - a density between 50 and 100 inhabitants/Km^2; and, Group 3 – a density higher than 100 inhabitants/Km^2. Thus, three MLR models were generated and analyzed, one for each group.

The results of the MLR model to estimate a Bus demand model through the estimation of the number of bus trips allowed to identify the following explanatory variables: Group 1 - Number of trips by bus, Number of illiterate people, and the Number of unemployed; Group 2 - Number of women over 15 years old, Purchasing power per capita; and, Group 3 - Number of unemployed, and Purchasing power per capita.

The explanation capacity, or estimation reliability for the three models can be analyzed through the adjusted R squared of each model, showing Group 1, 2 and 3 has 63%, 43%, 26% explanatory capacity, respectively. This results demonstrate a low reliability and estimation capacity to use a segmented MLR model in the estimation of Bus trips through statistical data available in Census. However, assuming an acceptable estimation error of 50% for modelled values in relation to real, the results shown that around 95% of municipalities are within that interval. It should be noted that for a very accurate analysis the MLR model are not convenient, but may be useful to estimate demand values for BUS usage in low density territories in very early phases of the bus network planning.

On the other hand, this work results that a more comprehensive MLR model can be analyzed, since it can increase the size of the sample / population and improve the results of the adjustment and explanatory capacity of the models. Or, in a more disruptive way, use other statistical techniques such as generalized models.

References:

[1] M. C. Almeida, "Alteração da deliberação relativa à classificação de territórios de baixa densidade para aplicação de medidas de diferenciação positiva dos territórios," Lisboa, 2015.

[2] A. Fernandes, J. F. Sousa and M. Fonseca, "A Problemática da Mobilidade em Espaço Rural e Áreas de Baixa Densidade Urbana: o caso dos concelhos de Mértola e Ourique," in Anais do I Congresso de Desenvolvimento Regional de Cabo Verde, 2009, pp. 2590–2617.

[3] M. G. P. M. Tsukamoto, "Governança Multiníveis em Territórios de Baixa Densidade: As Comunidades Intermunicipais do Alto Alentejo e da Beira Baixa." 2017.

[4] A. E. D. Domingues, "Transporte Público em Territórios de Baixa Densidade-O Caso de Melgaço–Alto Minho." 2009.

[5] T. Litman, Understanding transport demands and elasticities. Victoria Transport Policy Inst., 2017.

[6] V. F. R. Teixeira, "Um modelo de procura de transporte público rodoviário de passageiros: aplicação a Portugal e à região do Algarve." 2002.

[7] B. (Colombia). A. M. S. de T. y Transporte and W. F. C. Triana, Manual de planeación y diseño para la administración del tránsito y el transporte. Alcaldía Mayor, 2005.

[8] J. de Dios Ortuzar and L. G. Willumsen, Modelling transport. John Wiley & Sons, 2011.

[9] J. de Dios Ortúzar, Modelos de demanda de transporte. Ediciones UC, 2012.

[10] C. Curtis and T. Perkins, "Travel Behaviour: A review of recent literature," Perth, WA Urbanet, Curtin Univ. Technol., 2006.

[11] F. M. Dieleman, M. Dijst, and G. Burghouwt, "Urban form and travel behaviour: micro-level household attributes and residential context," Urban Stud., vol. 39, no. 3, pp. 507–527, 2002.

[12] H. A. Amado, "Modelo de procura de transporte público em autocarro para territórios de baixa densidade," University of Minho, 2018.

[13] J. A. Cavada Herrera, "Formulación y análisis de modelos de demanda agregada de validaciones y viajes de Transantiago," 2014.

[14] M. Pestana and J. Gageiro, Descobrindo a regressão: com a complementaridade SPSS. 2005.

[15] S. C. A. Rodrigues, "Modelo de regressão linear e suas aplicações." Universidade da Beira Interior, 2012.
[16] M. H. Pestana and J. N. Gageiro, "Análise de dados para ciências sociais: a complementaridade do SPSS," 2008.