Social lag impact in Mexican municipalities from a demographic categorization in urban and rural areas.

Abstract

The present document acts as the link between a clear and latent social problem in Mexico being the severe social lag of its population and how it is related to demographic and urbanization aspects. More specifically, a mathematical and statistical analysis is performed to study the relation and impact of the population of the municipalities, categorized based on whether the municipality and the geography could be considered as an urban area, and its impacts on the social lag of the municipality populations. Thus, by identifying possible relations between these two aspects and the nature of this relation, interesting deductions and conclusions could be drawn on how remedial actions or long-term urbanization projects could generate a positive impact in decreasing the social lag of the less benefited populations. The dataset employed comes from a multidimensional poverty measure study by the CONEVAL in Mexico, using data from Intercensal Survey in (2015). These datasets provide the basic information of each of the 2457 municipalities in Mexico with their population and the social lag index. The method employed to carry out the statistical analysis was firstly to identify the relation between the variables and make the formulation of the main and statistical hypothesis being the equality of the means of the social lag index when the municipality is an urban area or not based on its total population. Then, a descriptive statistics analysis was performed to study the behavior of the variables, make the respective transformation, and identify the best distribution of the transformed variable based on a goodness of fit test. Then, the maximum likelihood estimators for the mean and its variance were estimated for the two samples and based on the distribution of this parameter the null hypothesis could be rejected or not with certain significance. To make a deeper analysis and generate a stronger conclusion, Bayesian analysis was implemented by using the prior and posterior distribution for the parameters, as well as a regression analysis to identify the significance of the regression coefficient. The results of the analysis states that the mean of the social lag index in municipalities is different depending on whether it is an urban area or not, more specifically there is a lower social lag index in urban areas. The contribution of this results stands in demonstrating that the population of the municipalities and whether they are considered urban areas has an impact on the social lag of their population where a deeper analysis might make more evident the reasons of this difference and be able to solve them correctly from specific policies.

Keywords: Social change, Demographic Factors, Quality of life, Urban Areas, Rural Areas.

1.Introduction

Poverty, in all its dimensions and meanings is without doubts a universal social problem that needs to be addressed. This is not an exception in Mexico since in the last decade it has been seen that nearly half of the population is living in multidimensional poverty and one tenth of the population is living in extreme poverty [1]. Multidimensional poverty because poverty is considered multidimensional phenomenon in the way it comprises aspects related to living conditions that threaten the dignity of people and limit their rights and freedoms [1].

This concept of multidimensional poverty and its respective measurement was first introduced by Mexico in an ambition to measure poverty beyond only economic resources but contemplating other social dimensions that should be addressed. This methodology was proposed by the CONEVAL (Council for the Evaluation of Social Development Policy) by linking two different perspectives, on one hand the economic wellbeing based on having the sufficient economic resources, and on the other hand the social rights seeking to satisfy all the social rights and needs such as health, education, food, social security, and others. [1].

This indicator of poverty developed by the CONEVAL consider income and six social dimensions as follows [1]:

- Income
- Education lag
- Access to health services
- Access to social security
- Access to food
- Housing quality and space

- Access to basic housing services
- Degree of social cohesion

Thus, a population in poverty would result in an income below the wellbeing threshold (LBE) and the deprivation of one or more of aspects like education, health services, access to social security, inadequate housing (Lack of basic housing services & inadequate quality of space) and access to food.

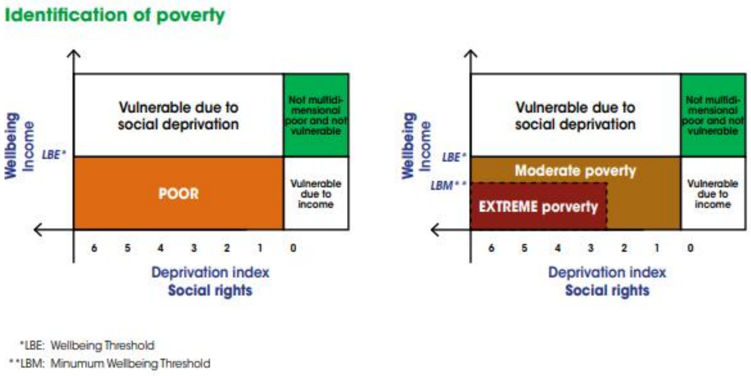

Figure 1: Identification of poverty considerations [1]

Previous studies on this index have shown that municipalities with the highest populations rates living in poverty were primarily rural, and those with the greatest number of people living in poverty were mostly urban [1]. The main problem of concern is whether the population of the municipalities, and more specifically, the categorization of this municipalities based on its population, having urban areas and non-urban areas is a factor in making a clear difference in the social lag of the municipalities, and based on that, what are the elements that makes this difference. This lack of identification of whether rural areas or urban areas concentrate more poverty gives the opportunity to try to identify the demographic aspects related to the poverty in Mexico. In this way the main problem here is whether the population and its magnitude influence the social lag, in which this problem involves other problems, since first identifying the nature of this relation and proving the difference in the social lag from municipality populations magnitudes would lead to identify the aspects of the type of municipalities benefited from less social lag and then try to replicate them in the other type of municipalities. In a more general and long-term perspective, the identification of social and demographic changes leading to better social and living conditions could be crucial in the considerations of long-term development plans to guarantee spaces with the required elements to assure a good quality of life for its inhabitants.

The contribution of this study resides in proving using statistical analysis the difference in the mean of the social lag in municipalities considered as urban areas and non-urban or rural areas and check on the direction and nature of this relation to use it as a guideline for future planification's considering the social welfare of the populations in a long term. Proving that the social lag of municipalities is lower in more populated municipalities could be a relief since the contemporary demographic tendency is the urbanization [2], thus a proposition that rural areas tend to have lower social lag would go against the natural demographic tendency and thus it would not be an encouraging result.

To carry out this study, first a review of the literature will be made to identify what previous works have already been made in this area of study and identify the gap amongst what has been studied so far and what is being researched in this project. Then, the main problem will be stated by identifying the problem, the motivations, the origins of the data and the main statistical hypothesis that are willed to be proved. Then, the analytical solution, the statistical procedure by means of different methods and the respective results and their analysis will be presented. Finally, a summary of the work and a conclusion proposing feature research directions from this work will be presented.

2.Literature Review

The main motivation from this study resides in the fact that the analysis proposed by the CONEVAL from the analysis of poverty in Mexico identified a lot of poor people living un urban settlements but most of the municipalities with high percentage of poor people are rural [1]. After research, many works have studied the relation between social lag or social change and quality of life, but not many have studied the relation between the urban life and the social lag. In fact, a deeper analysis is required to understand if there is more social lag in municipalities considered as urban areas or in rural areas based on the mean of the social lag in these two groups.

First, it is important to define and understand concepts related to population, territory, and its implications in a sustainable development. The population being defined as the number, growth, structure and spatial distribution of the persons and the demographic variables, a more bounded concept could be that population refers to spatial distribution and territorial mobility of people. On the same line, a territory is an area in which population live and move around [3]. Based on that, even if population concentration or urbanization is recommended for better social ends, it is important to consider the adversity of population concentrations since it puts excessive strain on the ecosystem occupied, saturates the infrastructure, collapses the social institutions and other non-beneficial aspects that could be counterproductive in the search for better quality of life and social welfare [3].

Based on this concept of population and territory, the link between those is the migration, which is a strategy used by communities, populations, and individuals to tackle adversity, build resources, and achieve social mobility to establish themselves in more comfortable and pleasant surroundings [3]. The tendency to migration, specially over the second half of the twentieth century has proven to go from rural to urban settlements, since around 70% of the populations lived in rural areas, which by 2010 this same percentage was of 20.4% in Latin America, and it is expected to be of 10% in the following decades [3].

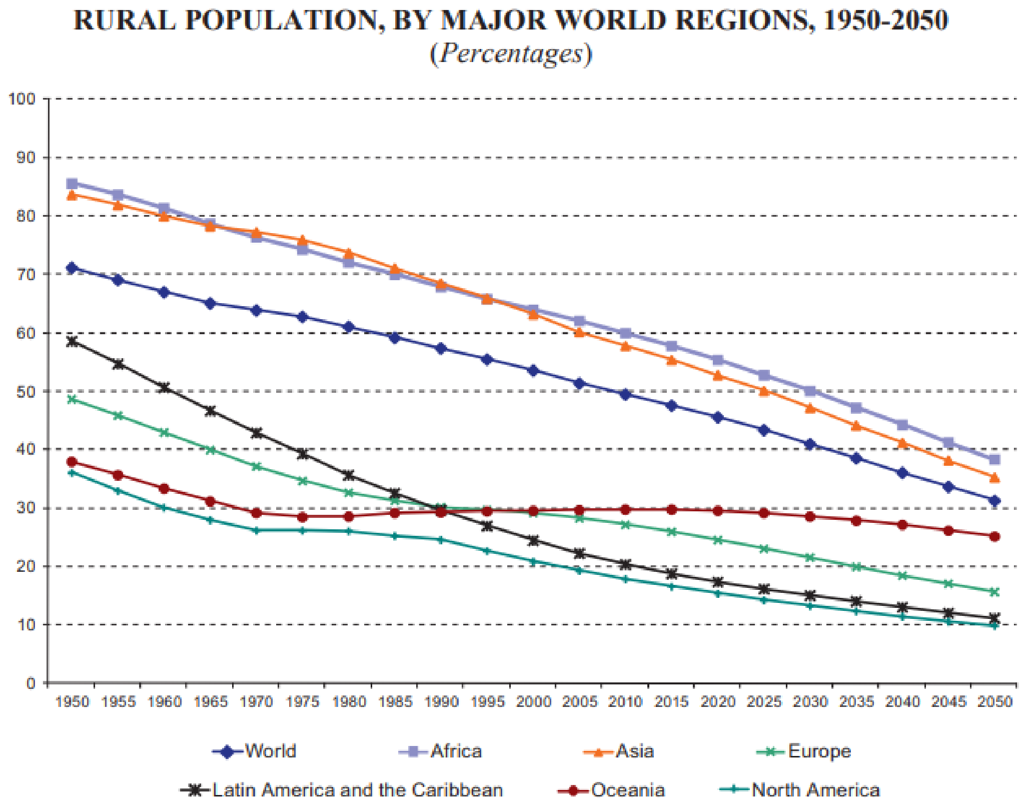

Figure 2: Rural Population by major world regions from 1950 to 2020 [3]

The expectancy of rural population has a negative tendency worldwide and in the case of Latin America it presents one of the lower weights overall [3].

This in fact, is an index that the population will move more and more to urban settlements, and they must be able to guarantee in capacity the settlement of all this migrant populations seeking for a better quality of life. The relation of these concepts and statistics with this research project is on the direction and trends of migrations that need to be considered when studying the direction and trends is social lag decrease from a specific migration type and verify if they are in accordance with each other.

Second, considerations of social and demographic change theory are important since either result from the analysis would result in certain considerations from this perspective. In this way, the urbanization is clearly related with technological innovations [16] since it involves a division of labor, more effective means of production allowing people to live in more dense areas having the required access to all its necessities [15]. This technological and technical changes also led to changes in the economic processes, capital accumulation, different ways to generate money and changing the whole rural perspective since technical knowledge leads to rising levels of production [17]. On the other hand, the social and economic change is not universal since innovation results only from the changes that are desired to be adopted in society, meaning that diffusion is not automatic, but selective, and adopted by people who are motivated by it. Following this, innovation in society tend to follow a pattern of diffusion from higher to lower status groups, based on its education, but also this change is the result of a clear goal directed large scale social planning [18]. The importance of mentioning this resides specially in the fact that a study of this nature based on the influence of demographic aspects on the quality of life and social welfare of the people involved has a final and long-term motivation which seeks to contribute in knowledge to society and influence in this large-scale social planning with positive results. This material changes being the technological and economic improvements or what is also called the material culture inevitably led to changes in the non-material culture which are the values, norms, laws, and social arrangements [19].

This non-material culture changes follows a longer process of implementation and lags the changes in the material culture. This lag is called the culture lag, which must be considered when proposing changes in long term, in this case in the demographic aspects, but also on how and how much time the society will take to assimilate them. This being the basis of all progress, to replaces the tools and the methods with more efficient ones, but always considering the implications and seeking for an equal change [19].

This being in direct relation with the intersection between the technological progress and the progress in the society, in this specific case, a demographic tendency for urbanization comes from improvements in technology but requires a social adaptation, so a tendency for urbanization or different demographic movements would require considering the social implications and the required times for these transformations to guarantee a good adaptability of the society to the environment.

On the other hand, territorial distribution of multidimensional poverty analysis identified that more than two thirds of the poor people in Mexico live in urban areas, but poverty rates are much higher in rural areas besides the clear existing inequality between the north and the south of Mexico, the south having much more social lag [1]. This source of extremely rural poverty, especially in the south of Mexico can be attributed to the conditions of geographical isolation with low accessibility levels and not any near economic and social infrastructure (12 % of population living in locations with low and very low accessibility levels). The current demographic situation in Mexico states that 74 metropolitan areas concentrate over 60% of the total population with areas that in theory are equipped with services, infrastructure, and more opportunities but on the other hand spatial segregation inside metropolitan areas has causes a disordered growth of cities which also impact on the social lag and the quality of living [2]. On the other hand, rural areas (less than 2500 people [14]) consist of 97% of the total localities of the country, and 90% of these localities are inhabited by less than 500 people. This implies that the demographic and territorial dynamics tends either to urban concentration or rural dispersion, both being unfavorable aspects for an overall social lag decrease [2]. Following this same line, five municipalities concentrate that same amount of population in poverty as 450 smaller municipalities [2].

Based on this proposition, multidimensional poverty corresponds to those who are deprived from at least one of the social dimensions mentioned in the introduction or having an income falling below the income poverty line (income needed to afford basic food baskets and non-basic food baskets of goods and services), following this guideline 43.6% of the population in 2016 lived in poverty [2].

On the other hand, the deprivation amongst population in poverty varies in its different parts depending on the size of localities and whether they are urban or rural communities. The main problems in rural localities are mostly related to lack of social security and lack of access to basic housing services, but for the urban localities, the main problem is regarding a lack of social security [2].

On the other hand, to determine the criteria for the partition of the data regarding the social lag index of municipalities based on the population of the municipalities, studying the urban and rural classifications was required to determine the limit of a population for a specific territory to be considered urban or not. An urbanized area is then considered to be a continuously built-up area with a population larger than 50 000 people, where an urban place outside an urbanized area is a place with at least 2 500 inhabitants [14].

Following a different line, and especially in the case of considering a result proposing a positive impact in the social lag in the cases of more populated municipalities, motivating thus to an urbanization, a study has found that a transition between the rural area from an agricultural employment to a formal sector employment is successful and firstly goes into non-agricultural wage work and entrepreneurship by informal sector paid jobs, and then as communities continue to experience gains in education, housing quality and poverty, the transitions goes now into a formal sector employment and non-agricultural self-employment. [5]. This proposition is important to this work since in the case of proving a that more urbanization is needed to reduce the social lag; it is also important to have theoretical works that proposes the accommodation in the work market of the increasing population migrating to urban areas.

Following this existing literature propositions and the nature of this research, it is important to confirm on the nature and direction of the relation between the social lag in municipalities and their population and whether their population are urban areas or not. The results of this research and the study of the actual demographic trends will allow to put in context the results of this project and give them a much better interpretation and application to contribute to the actual knowledge in this subject and be able to use these statistical demonstrations to some pragmatic ends.

3. Problem Definition and Formulation

The main motivation of this project, as it was already mentioned is on the understanding of the relation of demographics of municipalities in Mexico and whether they are considered urban areas or not and the social lag of the municipalities. Understandings from this possible relation could serve as a guideline for future long-term demographic planification and migration plans seeking to reduce the social lag of the Mexican population.

The dataset used to carry out this project proceeds from a multidimensional poverty measure study by the CONEVAL (Council for the Evaluation of Social Development Policy) using the data from the intercensal survey in 2015. This measurements on these dimensions take place every two years at state level and each five years at municipal level by the INEGI

(National Institute of Statistics and Geography) which provides this information to the CONEVAL.

The origins and motivations from this projects and dataset creations come from an ambition to measure poverty in Mexico based on its multidimensional nature and the fact of linking both the economic aspects, as well as the satisfaction of the main social rights [1]. Previous results have shown that it is important to make this analysis of poverty at municipal level since inside each state there may be very different conditions that could be left unrecognized if the analysis is made at state level [1].

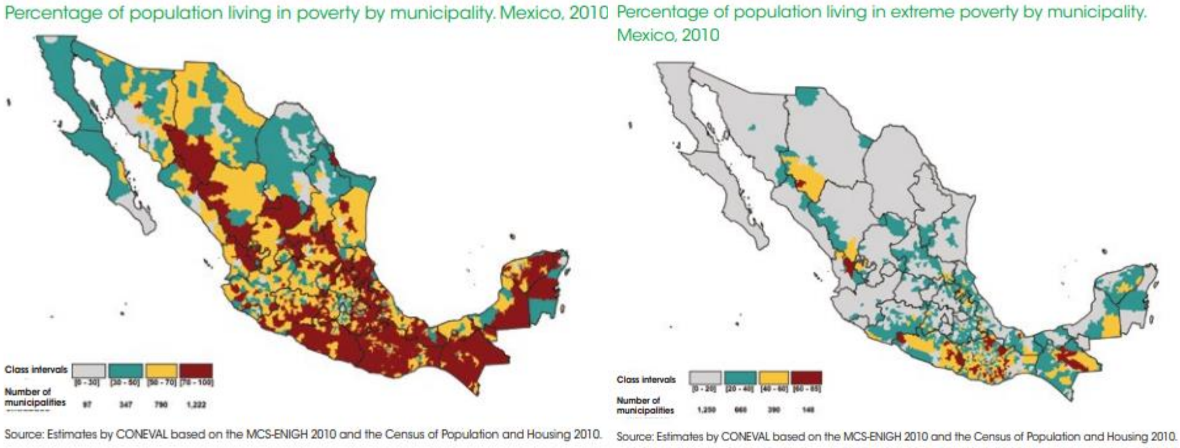

Figure 3: Poverty and extreme poverty in Mexican municipalities in 2010 [1]

Based on that, it was decided to use the dataset from the Intercensal Survey of 2015 which provides the social lag index, the variables from which the index is created and the population of each municipality, all for years 2000, 2005, 2010 and 2015. This index contemplates the variables based on access to education, access to health services, basic housing services and quality in the house spaces. [10]

More specifically and as it was already mentioned, the index is made from principal component analysis which is a statistical technique to reduce the dimensions of a set of variables by a linear combination of them. The specific variables contemplated to create the social lag index as percentage of population per municipality are [10]:

1.Percentage of illiterate 15 years and older population (0-100%)

2.Percentage of population aged 6-14 years old not attending school (0-100%)

3.Percentage of 15 years and older population with incomplete basic education (0-100%)

4.Percentage of population without access to basic health services

5.Percentage of households with dirt floors

6.Percentage of households without a toilet

7.Percentage of households without piped water from the public network

8.Percentage of households without drainage

9.Percentage of households without electricity

10. Percentage of households without a washing machine

11.Percentage of households without a refrigerator

The social lag index is calculated by means of component analysis from the weighted sum of 11 variables representing the percentage of population per municipality enjoying or not access to education, health and living condition facilities. The SLI is then categorized into 5 different categories (very low, low, medium, high, and very high social lag index). Considering that the dataset provides information for years 2000, 2005, 2010 and 2015, this study considers only the variables for year 2015 since they present the latest information, meaning that they could provide the more updated information and 929 observations representing the municipalities of 11 states in the central part of Mexico.

The main problem of concern in this research project is on identifying if there is a relation between the population of municipalities, more specifically if the municipalities are considered an urban area or not and if this has a significant effect in making a difference in the social lag of this municipalities. The identification of a significant difference in the social lag between these two categories of municipalities could be a first step in trying to identify more specific aspects based on the difference of these two types of demographic characterizations and try to replicate them in the less benefited categories seeking to reduce the social lag. Previous works has shown a lower social lag is present either in isolated rural communities or in the denser urban areas [1], meaning that the conclusion of the direction of the relation between the proposed variables could help clarify which type of demography of municipalities punishes more the social lag.

This problem will be addressed by analyzing the social lag in 2015 in 929 municipalities of 11 states in the central part of Mexico which in some way have similar geographic and cultural conditions by studying the distribution of the mean of the social lag index in the two different categories of municipality populations based on if it is an urban area or not.

Based on that, the main research question is whether the population of the municipalities and its type (Urban area or not) has an impact on the social lag index of the municipalities. Derived from this research question, a more mathematical and statistical hypothesis is formulated based on what is willed to be proved:

The mean of the social index of urban municipalities (Population bigger than 50 000 persons) is different than for the non-urban municipalities (Population lower than 50 000 persons).

H_0 : The mean of the social lag index of the municipalities with a population bigger than 50 000 persons is equal to the mean of the social lag index in municipalities with populations lower than 50 000.

H_a : The mean of the social lag index of the municipalities with a population bigger than 50 000 persons is different to the mean of the social lag index in municipalities with populations lower than 50 000.

$$\boldsymbol{H_0}: \boldsymbol{\mu_1} = \boldsymbol{\mu_2}$$

$$\boldsymbol{H_a}: \boldsymbol{\mu_1} \neq \boldsymbol{\mu_2}$$

Where:

$$\mu_1 = \mu_{SLI_{2015}|TP_{2015}>50\,000}$$

$$\mu_2 = \mu_{SLI_{2015}|TP_{2015}\leq 50\,000}$$

$$SLI_{2015} = Social\ Lag\ Index\ per\ municipality\ in\ 2015$$

$$TP_{2015} = Total\ population\ per\ municipality\ in\ 2015$$

The consideration to define the partition of the two data groups being a population bigger or lower than 50 000 persons is based on the definition of an urban area which is considered for populations bigger than 50 000 persons according to the urban and rural classifications [14]. Based on this hypothesis proposal, if the mean of the social lag index in the two different categories of municipalities based on its population is proven to be different based on a non-overlapping of the distribution of the mean of each category or an overlapping of less than a significance level considered, in this case $\alpha = 0.05$, a difference in the social lag between these two categories is proven. To validate the conclusion of the hypothesis, Bayesian methods will be used to calculate the prior and posterior distributions based on the prior and posterior hyperparameters of the distributions of the samples and regression analysis will also validate the relation between these variables.

On the other hand, if it cannot be proved with a level of significance that the mean of the social lag index from the two samples differ, then the null hypothesis is accepted and the considerations of the population of the municipalities having an effect in the social lag index of municipalities is proved wrong.

4. Analytic Solution

In this section the analytic solution to the problem will be presented by stating the general procedure followed to carry out a conclusion on the hypothesis involved and the main problem of this work. The procedure followed roughly consist of first presenting the data and its properties involved in the mathematical and statistical analysis, then how the data was cleaned and prepared for analysis, descriptive statistics of the data and the relations between them would make more evident the nature of the variables and their relations. Finally, having the data from the variables in their correct representation, the application of the statistical analysis techniques will be presented each with their results and interpretations.

4.1 Data and variables

As it was previously mentioned, the original dataset contains 2457 observations representing all the municipalities in Mexico and 64 variables, 4 representing the information of the state and municipality and all the other 60 variables represent 15 different variables expressed each in four different years (2000, 2005, 2010 and 2015). From these 15 variables, one variable represents the total population of the municipalities, three variable represents the social lag index, its category from very low to very high and the place occupied in the national context. The eleven remaining variables represents the variables on which the social lag index calculation is based representing the percentage of population per municipalities with the basic social rights on health, education and living conditions categories.

First, the data set was reduced from the 32 states representing 2457 municipalities to 11 states in the central part of Mexico, representing 969 municipalities. Some other variables were considered for different analysis involved in this project, but the main variables contemplated for analysis in this specific part of the project on which this research was made are only two:

- Social Lag Index per municipality in 2015 (SLI_{2015})
- Total population per municipality in 2015 (TP_{2015})

The original variable SLI_{2015} ranges from -1.62157 representing a low social lag index, to 3.71904 representing a high social lag index but for practical reasons and to be able to transform this variable correctly and explain it with a wider variety of distributions, it was shifted by 2 units to have its minimum over 0, so the new minimum is 0.37843 representing the lowest low social lag index and a maximum at 5.71904.

Table 1: Properties of the variables involved in the analysis.

Variable	Concept	Variable type	Units
SLI_{2015}	Social Lag Index per municipality in 2015	Numeric Float	Adimensional
TP_{2015}	Total population per municipality in 2015	Numeric Integer	Number of people

4.2 Data cleaning

The data cleaning process consisted as mentioned before in reducing the number of observations from 2457 to 969 firstly. For the case of the reduction of observations based on errors and “NA” fields, it was not required since the dataset in the variables contemplated did not present any lack of information. On the other hand, after analyzing the variable of total population TP_{2015} and making the required transformations on this variable, the best-found transformation still presented 4.127% of the data as outliers, so it was decided to remove the 40 outlier observations since the variable of total population is very important and was present in other analysis not presented in this document. The final number of observations kept is a total of 929, which is the final number of observations used for the whole analysis of the dataset.

4.3 Properties of variables and relations between them

To study the relation between the variables, some descriptive statistics as box plot, histograms and scatter plots were performed to check if a transformation was required or if it was required to remove outliers after transforming the data.

First, a scatter plot between the original variables (SLI_{2015} & TP_{2015}) was performed.

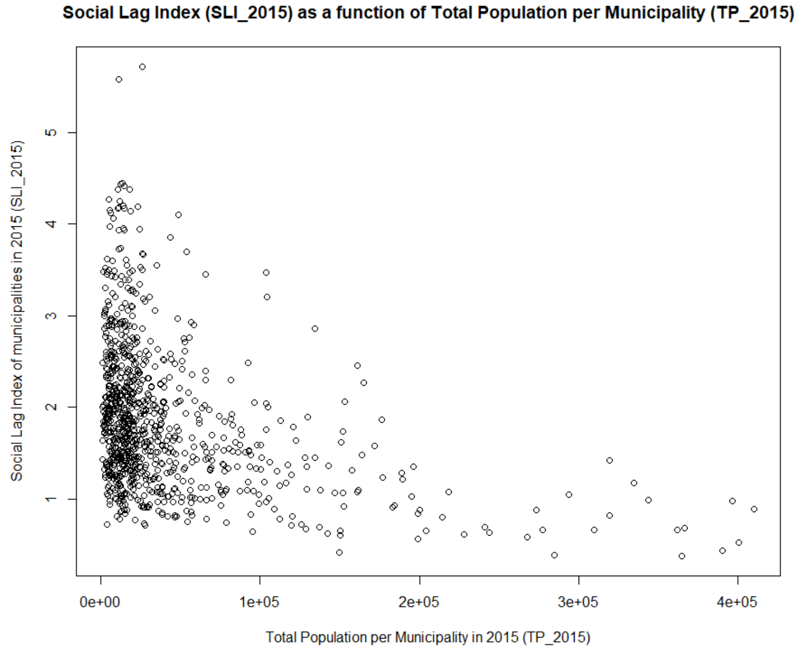

Figure 4: Social Lag Index as a function of Total Population per municipality in 2015

There is not a clear pattern showing the relation between the variables involved but the plot makes evident the required transformation on the variables.

The social lag index per municipality in 2015 SLI_{2015} behaves as follows:

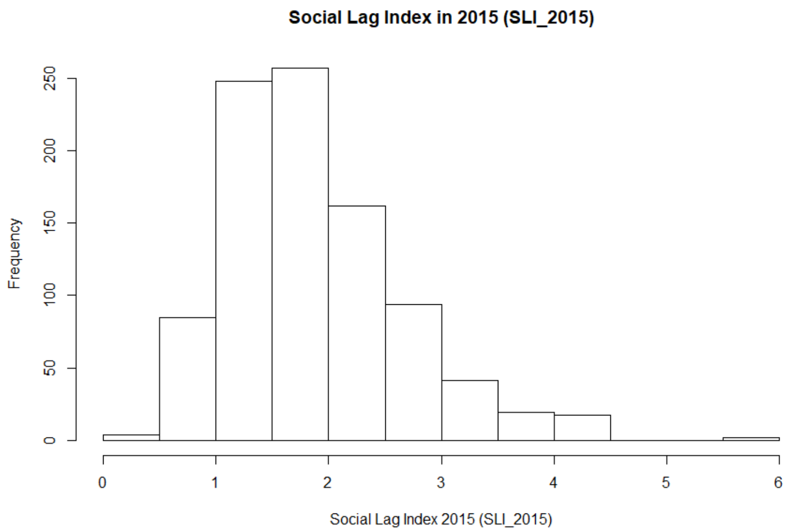

Figure 5: Histogram of variable representing the Social Lag Index per Municipality in 2015 (SLI_{2015})

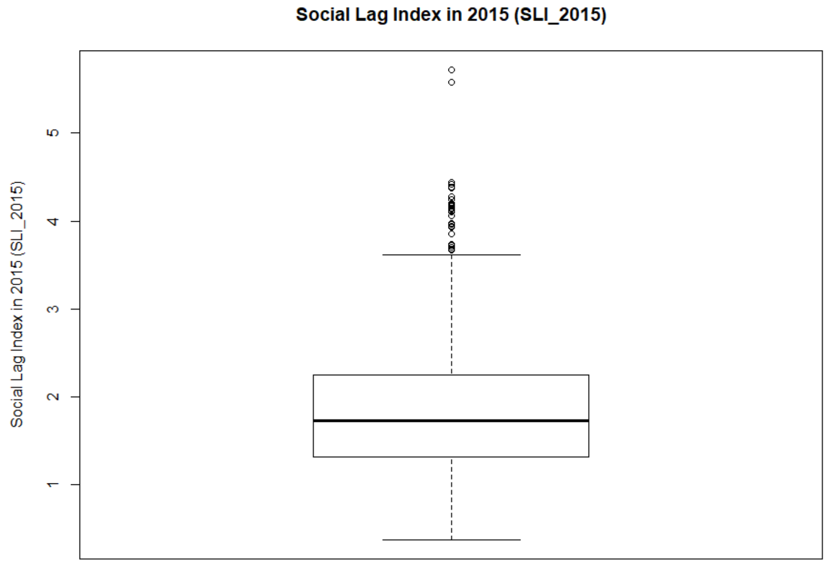

Figure 6: Box Plot of the Social Lag Index in 2015 (SLI_{2015})

From the histogram, the distribution of the original variable of SLI_{2015} presents a good distribution, near a normal distribution but a bit skewed to the left. From the boxplot, the presence of some outliers can be observed in the right tail, even if a central tendency can be observed from the distribution of the variables. The total number of outliers present is 30 out of 929 observations representing 3.229% of total observations, meaning that either a transformation of the variable or the removing of the outliers is required. It is better proposed to make a transformation of the variable.

Different transformation of the SLI_{2015} variable was checked and the one presenting the best proposal is the natural logarithm transformation.

$$SLI_{2015\,T} = \ln\left(SLI_{2015}\right)$$

The results from this transformation are the following:

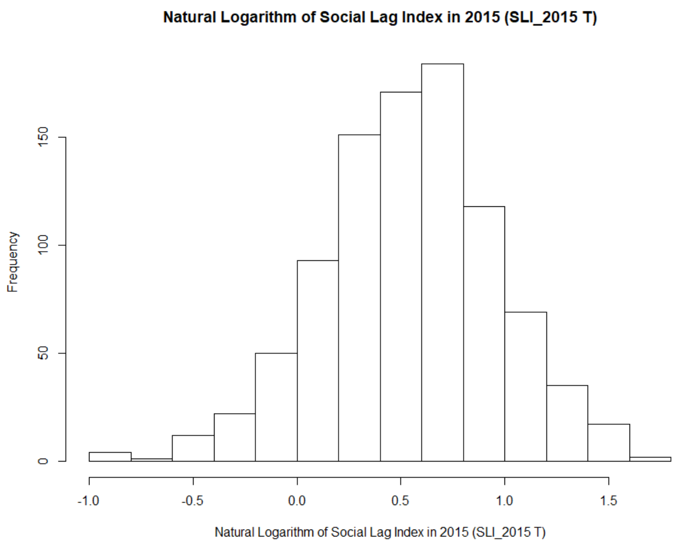

Figure 7: Histogram of natural logarithm transformation the Social Lag Index per Municipality in 2015 $(SLI_{2015\ T})$

Figure 8: Box Plot of the natural logarithm of the Social Lag Index in 2015 $(SLI_{2015\ T})$

From these plots, it is evident that the transformation on the variable creates a much better result since the variable seems to follow a normal distribution and the number of outliers is considerably reduced having in this case just 9 out of 929 outlying observations representing only 0.969% of the total observations. It is decided then to keep the natural logarithm transformation of the SLI_{2015} which will be represented by the variable $SLI_{2015\,T}$.

The same methodology is followed for checking on possible transformations for the variable representing the total population per municipality in 2015 (TP_{2015}).

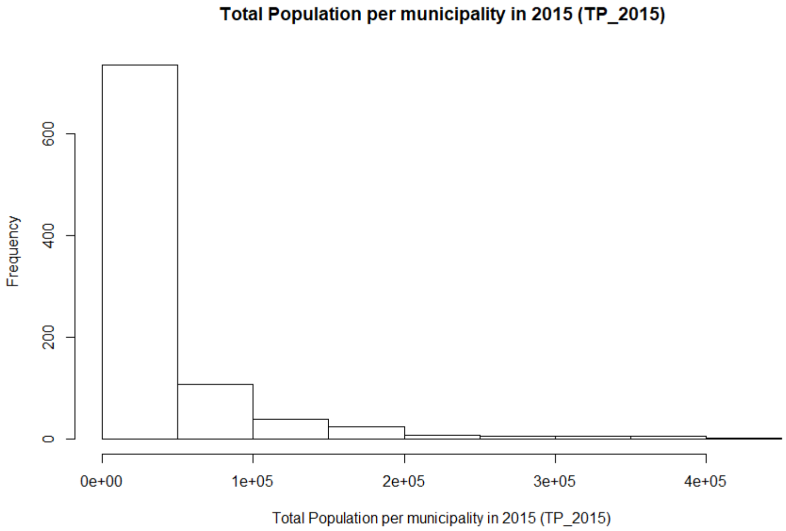

Figure 9: Histogram of Total Population per Municipality in 2015 (TP_{2015})

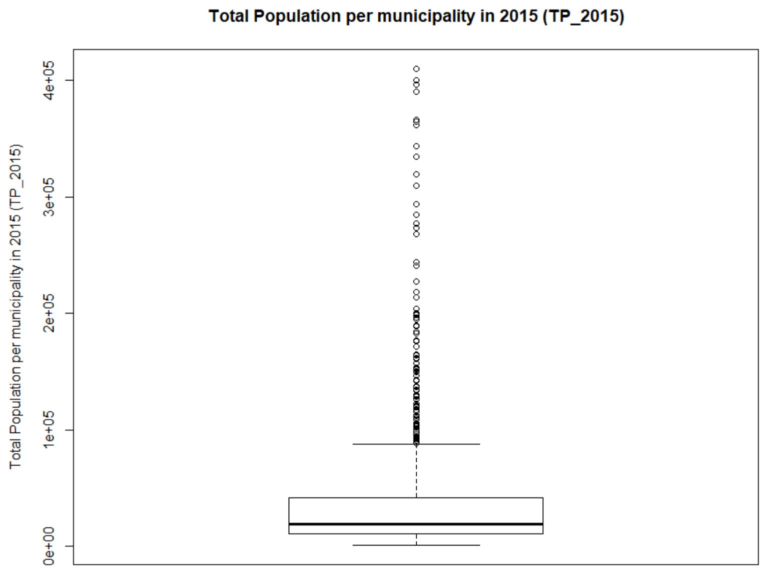

Figure 10: Box Plot of the Total Population per Municipality in 2015 (TP_{2015})

From the analysis of this plots, it is evident that the original variable representing the total population per municipality requires a transformation since the histogram shows a not very common or manageable distribution and on the other hand the number of outliers represented in the box plot is very considerable, with 103 out of 929 observations representing 11.087% of total observations, having very populated municipalities against most municipalities having lower populations.

Different transformation of the TP_{2015} variable was checked and the one presenting the best proposal is the natural logarithm transformation.

$$TP_{2015\,T} = \ln\left(TP_{2015}\right) \tag{1}$$

The results from this transformation are the following:

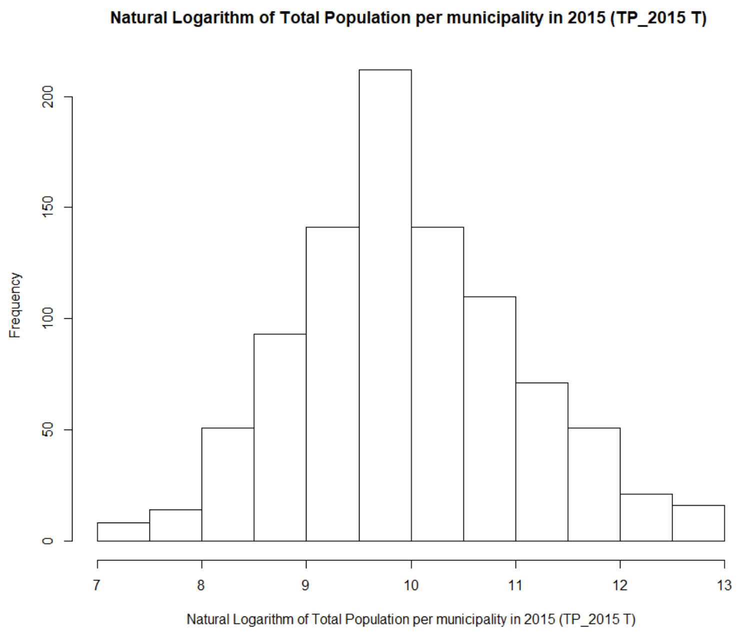

Figure 11: Histogram of Natural Logarithm of Total Population per Municipality in 2015 ($TP_{2015\,T}$)

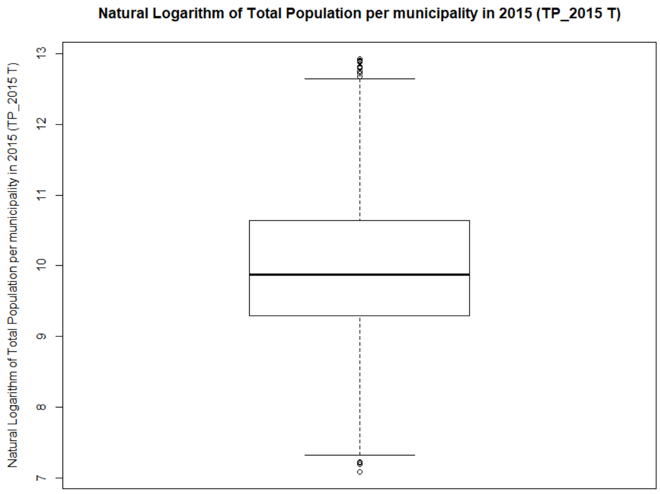

Figure 12: Box Plot of the Total Population per Municipality in 2015 (TP_{2015})

It is evident that a natural logarithm transformation on the variable representing the total population proposes a much better shape of the distribution of the variable and the box plot shows a much more central tendency and a drastic reduction in the number of outliers, having only 15 out of 929 outlying observations representing only 1.615% of the total observations. The natural logarithm transformation of the TP_{2015} is kept, which will be represented by the variable $TP_{2015\,T}$, having a scatter plot of both variables as follows:

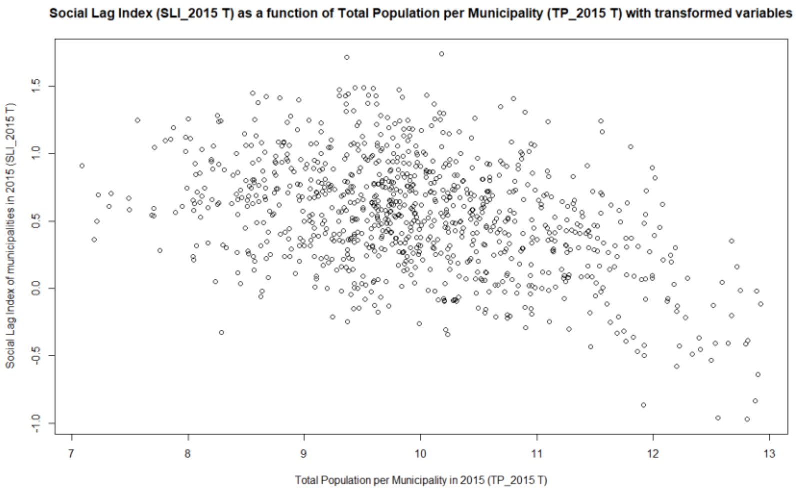

Figure 13: Scatter plot of the Social Lag Index as a function of Total Population per municipality in 2015 with transformed variables

The scatter plot with the transformed variables presents much better results and its interpretation provide more information.

The final transformations on the variables involved in the statistical procedures are presented in the following table:

Table 2 : Final transformation on the SLI_{2015} & TP_{2015} variables

Variable	Concept	Transformation	Name of transformed Variable
SLI_{2015}	Social Lag Index per municipality in 2015	ln (SLI_{2015})	$SLI_{2015\,T}$
TP_{2015}	Total population per municipality in 2015	ln (TP_{2015})	$TP_{2015\,T}$

There is a small tendency to have lower social lag in more populated municipalities, but still it is not clear the difference on the Social Lag based on the population of municipalities so a deeper statistical analysis will be made.

4.4 Data fitting in hypothesis evaluation and methodological approach overview

To link the variables presented in last section with the intended statistical development, the process will be to split the main variable $SLI_{2015\,T}$ involved in the hypothesis into its two groups based on the splitting criteria which is whether the total population of the municipality is greater or lower than 50 000 persons. It is important to mention that the $SLI_{2015\,T}$ will contemplate the natural logarithm transformation, but for the variable involved in the splitting criteria it is not required to use the transformed variable.

The two sample groups of the splitted variable are:

$$SLI_{2015\,T1} = SLI_{2015\,T} \mid TP_{2015} > 50000 \tag{2}$$

$$SLI_{2015\,T2} = SLI_{2015\,T} \mid TP_{2015} \leq 50000 \qquad (3)$$

By doing a goodness of fit simulation on the main variable it will be possible to determine the main distribution of the $SLI_{2015\,T}$. The distribution of the main variable will guide to the distribution of the partitioned variables and then the parameter for the mean of the distribution (μ) will be calculated with its respective distribution to determine whether there is a significant interception of the distribution of the means of each of the two populations in question.

The methodological approach to calculate the distribution of the mean parameter of the distribution of the main variable involved in the hypothesis will be first by a fisherian approach where the parameters are fixed and calculated by means of the maximization of the maximum likelihood function from the observed data to then approach another solution using the Bayesian method by calculating the prior and posterior hyperparameters based on the distribution of the variable and calculate the prior and posterior distributions considering the data evaluated. Finally, a regression analysis will be performed on the data to check on the significance of the regression coefficient from regressing the $SLI_{2015\,T}$ against the $TP_{2015\,T}$ and other regression parameter that could provide more information on the nature of the relationship between those variables. The different means and methods of arriving to a conclusion on the proposed hypothesis is mainly because the three approaches are different, when in the fisherian approach you get information by just analyzing the data in question and estimating the probability, in the Bayesian method you get a more precise and accurate distribution of the parameters since you consider the prior probability and the likelihood of the data. Finally, the regression approach is a much different one since it seeks to establish a relation between the involved variables in the analysis by looking on how the variability of the response variable is explained by the predictor variables, depending on the nature and magnitude of this relation it can be determined if one variable has significant effect in the variability of the other. All these approaches will be used to generate a much solid conclusion to the problem in question.

4.5 Variable distribution Identification

To calculate the parameters on which the hypothesis propositions is based it is required to know in advance the distribution of the $SLI_{2015\,T}$ to calculate it maximum likelihood estimators and the prior and posterior hyperparameters to obtain the prior and posterior distributions of the parameters to evaluate.

To determine the distribution of the $SLI_{2015\,T}$ the followed procedure is:

- Solving a continuous distribution fitting of the variable to determine the top distributions to fit the data based on some criteria as the log likelihood, AIC and the behavior of the QQ-Plot and PP-Plot.
- Goodness of fit simulation over 1000 runs calculating how many times each of the top 3 distributions was fitted to determine the most likely distribution.

For the case of the $SLI_{2015\,T}$ variable this procedure was applied obtaining the following results:

The continuous fitting of the $SLI_{2015\,T}$ variable proposes the following:

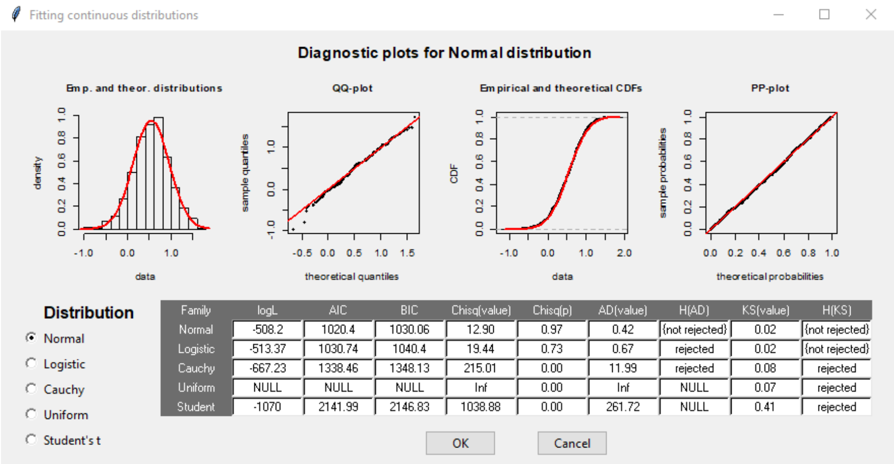

Figure 14: Diagnostics for possible distributions of $SLI_{2015\,T}$

A normal distribution is proposed as the best distribution for the $SLI_{2015\,T}$ variable since for this distribution the best log likelihood value is found with -508.2, the higher AIC value with 1020.4 and a QQ-Plot showing few presences of outliers and an almost perfectly fitted PP-Plot. The top 3 distributions to explain the $SLI_{2015\,T}$ variable is the normal, logistic and Cauchy distribution.

A simulation of the goodness of fit with replacement was performed to check on the frequency that each of the three distributions was the best candidate to explain the variable and the count of each was stored in a data frame as follows:

Table 3 1000 Goodness of fit simulation results for top 3 distributions explaining SLI_{2015}

Distribution	**Total**
Normal	911
Logistic	89
Cauchy	0

The distribution best explaining the transformed variable $SLI_{2015\,T}$ is the normal distribution with probability density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4}$$

Where:

$$\mu = Mean\ of\ the\ normal\ distribution$$

$$\sigma = Standard\ deviation\ of\ the\ normal\ distribution$$

In this case that the variable follows a normal distribution, the mean (μ) parameter is a parameter from the distribution, so by means of estimating the parameter with maximum likelihood estimators and the variability of the parameter using fisher information function it is possible to obtain the distribution of the parameters of each of the two samples.

4.6 Fisherian approach for mean estimation

The fisherian approach calculates the probability of seeing the data given the hypothesis. The maximum likelihood estimator (MLE) is used to estimate the parameters seeking to find the parameters which maximize the likelihood function consisting of the probability of having the data given that a specific parameter is chosen. The advantages of using MLE is that it is an automatic algorithm and it do not require further statistical inputs and that it enjoys excellent frequentist properties in large samples since the estimates tend to be nearly unbiased with few variances [9].

Having said that, the MLE is the value of the parameter maximizing the likelihood function.

-The likelihood function for a family of probability densities $f_\mu(x)$ is:

$$l_x(\mu) = \log f_\mu(x) \tag{5}$$

-The MLE is the value of the parameter μ in space Ω maximizing the likelihood function

$$\hat{\mu} = \arg\max l_{x_{\mu\in\Omega}}(\mu) \tag{6}$$

The procedure to obtain the MLE for the normal distribution is as follows, considering that the probability density function of the normal distribution is:

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{7}$$

$$\mu: Mean\ of\ the\ distribution$$

$$\sigma: Standard\ Deviation\ of\ the\ distribution$$

The likelihood function and the log likelihood function are presented on the following table and are the ones on which the calculation of the MLE and the variance of the parameters are performed.

Table 4 Likelihood and log-likelihood function of the normal distribution

Function	Equation
Likelihood Function	$\prod_{i=1}^{n} f(x_i\|\mu,\sigma) = \prod_{i=1}^{n}(\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu^2)}{s\sigma^2}}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-(\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2})}$
Log-Likelihood function	$\ln\left[\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\left(\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}\right)}\right] = -\left(\frac{n}{2}\right)\ln(2\pi\sigma^2) - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}$

MLE of μ

Equal the log likelihood function to 0 and solve for μ

$$\frac{\partial \ln(l_x)}{\partial \mu} = 0 \tag{10}$$

$$\frac{\partial \ln (l_x)}{\partial \mu} = \frac{1}{\sigma^2}(x_i - \mu) = 0$$

Since $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)}{n-1}$

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (11)$$

Applied to the data:

$$\hat{\mu}(SLI_{2015\,T}) = \frac{\sum_{i=1}^{929} \mathrm{SLI}_{2015\,T}}{929} = \frac{500.3538}{929} = 0.5386$$

$$\hat{\mu}\,(SLI_{2015\,T}) = 0.5385939$$

The same procedure was made to calculate the parameter of the mean for the partitions of the data.

-Partition 1

$$SLI_{2015\,T1} = SLI_{2015}\,|\,TP_{2015} > 50000$$

$$n_{T1} = 193$$

$$\hat{\mu}(SLI_{2015\,T1}) = \frac{\sum_{i=1}^{193} \mathrm{SLI}_{2015\,T1}}{193} = \frac{46.6141}{193} = 0.2415$$

$$\hat{\mu}(SLI_{2015\,T1}) = 0.2415236$$

-Partition 2

$$SLI_{2015\,T2} = SLI_{2015}\,|\,TP_{2015} \leq 50000$$

$$n_{T1} = 736$$

$$\mu\,(SLI_{2015\,T1}) = \frac{\sum_{i=1}^{736} \mathrm{SLI}_{2015\,T2}}{736} = \frac{453.7397}{736} = 0.6165$$

$$\hat{\mu}\ (SLI_{2015\,T2}) = 0.6164942$$

The following table summarizes the results for the MLE calculation of the mean μ of the $SLI_{2015\,T}$ from its two partitions.

Table Summary of results of the MLE calculation for the μ of the $SLI_{2015\,T}$

Variable	Details of variable	μ (Calculation)	μ (Software)
$SLI_{2015\,T}$	$\ln\,(SLI_{2015})$	$\mu = 0.5386$	$\mu = 0.5385939$
$SLI_{2015\,T1}$	$\ln(SLI_{2015})\,\|TP_{2015} > 50\,000$	$\mu = 0.2415236$	$\mu = 0.2415236$
$SLI_{2015\,T2}$	$\ln(SLI_{2015})\,\|TP_{2015} \leq 50\,000$	$\mu = 0.6165$	$\mu = 0.6164942$

First, it is important to remark that the results from the calculation and the results from the software following a numerical analysis, the same results were obtained meaning that the calculations were correctly performed. From the estimation of the mean from the two distributions, they have different magnitudes, still it would be important to define a distribution of the mean considering its variance to confirm on the means not being equal with certain level of significance. Thus, the fisher information function will be used to determine the Cramer's Roa Lower Bound and determine the variance of the parameters.

4.7 Fisherian approach for mean variance estimation

Since the knowing of the value of the parameter is not enough information to determine the distribution of the parameter, the inverse of the fisher information function gives the variance of the parameter. The fisher information function is obtained by solving the second partial derivate of the log likelihood function with respect to the parameter that is willed to be estimate its variability (μ).

$$I(\hat{\mu}) = -\frac{\partial^2 f(x|\mu,\sigma)}{\partial^2 \mu} \qquad (12)$$

$$\frac{\partial}{\partial \mu}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)\right) = \frac{n}{\sigma^2} \tag{13}$$

From previous section:

$$\sigma_{SLI_{2015\,T}} = 0.4181561$$

$$\sigma^2_{SLI_{2015\,T}} = (0.4181561)^2 = 0.1748545$$

$$I(\hat{\mu}) = \frac{929}{0.1748545} = 5312.99$$

Since the parameter $\mu \sim N(\hat{\mu}, \frac{1}{I(\hat{\mu})})$, the variance of the parameter is:

$$\sigma_{\hat{\mu}_{SLI_{2015\,T}}} = \frac{1}{I(\hat{\mu})} = \frac{1}{5312.99} = 0.0001882179$$

The same procedure was followed to obtain the variance of the mean of the sample partitions.

-Partition 1

$$\sigma_{\hat{\mu}_{SLI_{2015\,T1}}}$$

$$I\left(\hat{\mu}_{SLI_{2015\,T1}}\right) = \frac{n_{T1}}{\sigma^2_{SLI_{2015\,T1}}} = \frac{193}{0.1900621} = 1015.4576$$

$$n_{T1} = 193$$

$$\sigma^2_{SLI_{2015\,T1}} = 0.4359611^2 = 0.1900621$$

Since the parameter $\mu \sim N(\hat{\mu}, \frac{1}{I(\hat{\mu})})$, the variance of the parameter is:

$$\sigma_{\hat{\mu}_{SLI_{2015\,T1}}} = \frac{1}{I(\hat{\mu}_{SLI_{2015\,T1}})} = \frac{1}{1015.4576} = 0.0009847777$$

-Partition 2

$$\sigma_{\hat{\mu}_{SLI_{2015\,T2}}}$$

$$I(\hat{\mu}_{SLI_{2015\,T2}}) = \frac{n_{T2}}{\sigma^2_{SLI_{2015\,T2}}} = \frac{736}{0.1416563} = 5195.674$$

$$n_{T2} = 193$$

$$\sigma^2_{SLI_{2015\,T2}} = 0.3763726^2 = 0.1416563$$

Since the parameter $\mu \sim N(\hat{\mu}, \frac{1}{I(\hat{\mu})})$, the variance of the parameter is:

$$\sigma_{\hat{\mu}_{SLI_{2015\,T2}}} = \frac{1}{I(\hat{\mu}_{SLI_{2015\,T2}})} = \frac{1}{5195.674} = 0.0001924678$$

The following table summarizes the results for the estimation of the variance of the mean from the main variable and its two partitions.

Table 5 Results of the variance of the mean (μ) of the main variable and the two partitions

Variable	**Parameter calculated**	$\boldsymbol{\sigma_\mu}$ **(Calculation)**	$\boldsymbol{\sigma_\mu}$ **(Software)**
$SLI_{2015\,T}$	$\sigma_{\hat{\mu}_{SLI_{2015\,T}}}$	$\frac{1}{5312.99} = 0.0001882179$	$\frac{1}{5312.989} = 0.0001882180$
$SLI_{2015\,T1}$	$\sigma_{\hat{\mu}_{SLI_{2015\,T1}}}$	$\frac{1}{1015.4576} = 0.0009847777$	$\frac{1}{1015.458} = 0.0009847773$
$SLI_{2015\,T2}$	$\sigma_{\hat{\mu}_{SLI_{2015\,T2}}}$	$\frac{1}{5195.674} = 0.0001924678$	$\frac{1}{5195.672} = 0.0001924679$

The numeric and algebraic procedure proposes the same results in each case for standard deviation of the distribution of the estimated mean of the distribution following a normal distribution with mean 0 and standard deviation $\sigma_{\hat{\mu}_{SLI_{2015\,T2}}}$, $\sigma_{\hat{\mu}_{SLI_{2015\,T2}}}$ and

$\sigma_{\hat{\mu}_{SLI_{2015\,T2}}}$ respectively. The results for the distributions of the mean with their respective distribution plots are:

$$\mu_{SLI_{2015\,T}} \sim N\left(\hat{\mu}_{SLI_{2015\,T}}, \sigma_{\hat{\mu}_{SLI_{2015\,T}}}\right) \sim N\,(0.5385939\,, 0.0001882179)$$

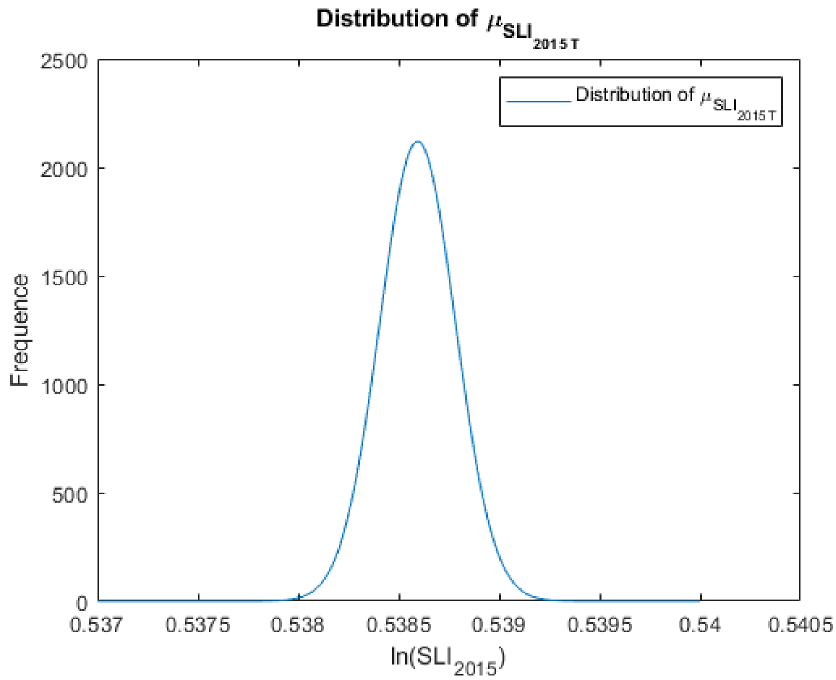

Figure 15: Distribution of the mean of the $SLI_{2015\,T}$ $(\mu_{SLI_{2015\,T}})$

$$\mu_{SLI_{2015\,T1}} \sim N\left(\hat{\mu}_{SLI_{2015\,T1}}, \sigma_{\hat{\mu}_{SLI_{2015\,T1}}}\right) \sim N\,(0.2415236\,, 0.0009847777)$$

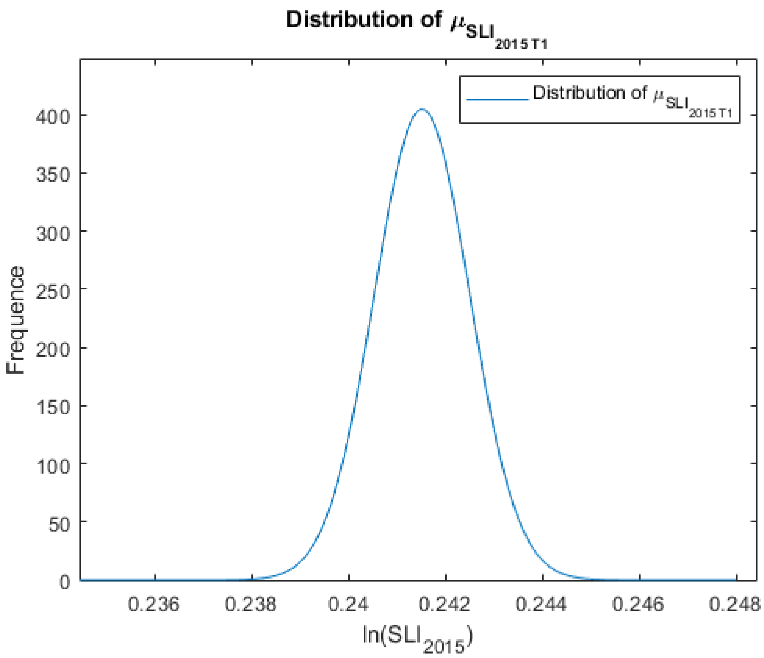

Figure 16: Distribution of the mean of the $SLI_{2015\,T1}$ $(\mu_{SLI_{2015\,T1}})$

$$\mu_{SLI_{2015\,T2}} \sim N\left(\hat{\mu}_{SLI_{2015\,T2}}, \sigma_{\hat{\mu}_{SLI_{2015\,T2}}}\right) \sim N\,(0.6164942\,, 0.0001924678)$$

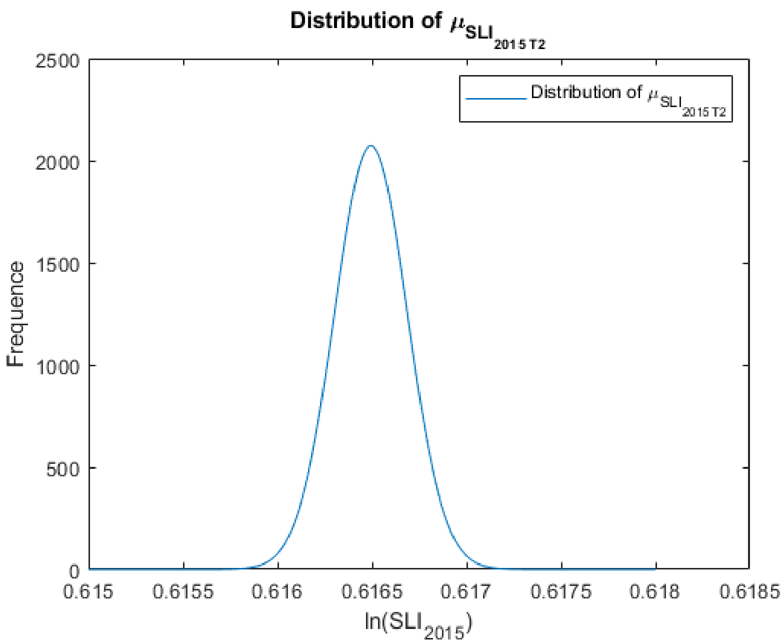

Figure 17: Distribution of the mean of the $SLI_{2015\,T2}$ $(\mu_{SLI_{2015\,T2}})$

Having the distribution of the mean of the variable $SLI_{2015\,T}$ and the two partitions, a combined plot would lead to a more visual representation to determine whether it is possible to reject with certain significance level that the means are equal between the two populations.

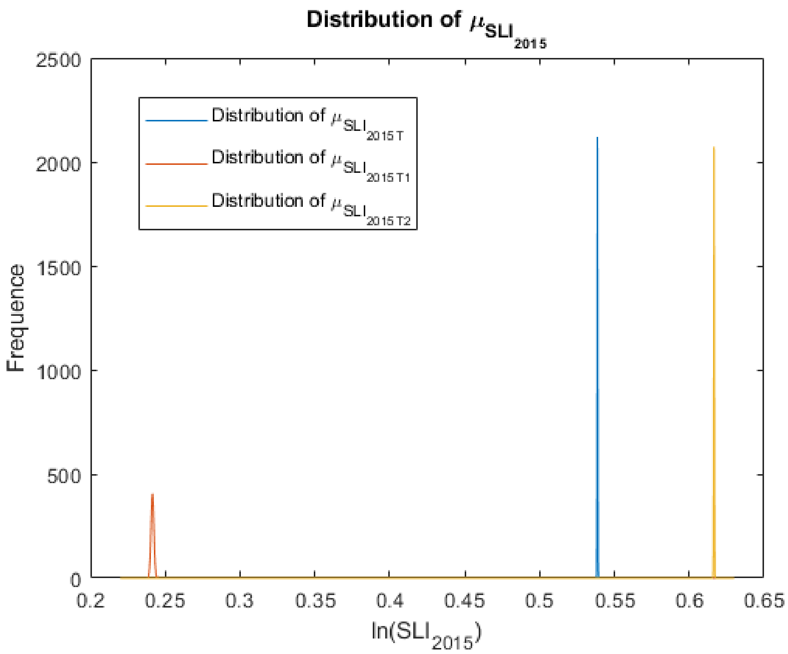

Figure 18: Plot combining the distribution of the mean of $SLI_{2015\,T}$ and the two partitions ($SLI_{2015\,T1}$ & $SLI_{2015\,T2}$)

From analyzing this plot, the distribution of the mean of the two partitions ($SLI_{2015\,T1}$ & $SLI_{2015\,T2}$) do not overlap, meaning that they are different. From the fisherian analysis perspective the null hypothesis which assumes that the two means are equal can be rejected since the distribution do not overlap and are considerably far away from each other.

Conclusion → $H_a: \mu_{SLI_{2015\,T1}} \neq \mu_{SLI_{2015\,T2}}$

4.8 Bayesian analysis approach for prior and posterior distributions

Bayesian analysis is a method of statistical inference allowing to combine prior information about a parameter from a population with evidence from the data in the sample to come up with a more robust probability prediction called the posterior probability distribution [20]. Bayes rule states the following:

$$g(\mu|x) = \frac{g(\mu) f_{\mu}(x)}{f(x)} \quad (14)$$

Where:

$$g(\mu|x) = Posterior\ density\ of\ \mu$$

$$g(\mu) = Prior\ density\ of\ \mu$$

$$f(x) = Marginal\ Density\ of\ x = \int_{\Omega} f_{\mu}(x) g(x) d\mu \quad (15)$$

$$f_{\mu}(x) = Likelihood\ of\ x\ given\ \mu$$

Based on that, the prior and posterior hyperparameters for the distribution of the mean of the normal distribution representing each partition of $SLI_{2015\,T}$ ($SLI_{2015\,T1}$ & $SLI_{2015\,T2}$) were calculated following a numerical procedure algorithm following this methodology:

- Create a vector of the value of $SLI_{2015\,T}$ for each partition.
- Do a 1000 simulation considering sampling with replacement of fitting the vector with a normal distribution and estimate the mean parameter μ to obtain a 1000 value vector of the estimation of the μ parameter.
- Fit the resulting vector of 1000 observations of the sampling means with a normal distribution to obtain the mean (μ) and the standard deviation (σ) of the prior distribution of the mean of each partition

- Calculate the posterior hyperparameters for considering a conjugate prior distribution being the normal distribution based on the following:

Table 6 Conjugate Prior Distributions and prior and posterior hyperparameters [13]

Likelihood	**Model Parameters**	**Conjugate Prior Distribution**	**Prior hyperparameters**	**Posterior Hyperparameters**
Normal with known variance σ^2	$\mu\ (Mean)$	Normal	μ_0, σ_0^2	$\frac{1}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^{n} x_i}{\sigma^2}\right), \left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)^{-1}$

- Create the prior and posterior probability density functions for the mean (μ) with the prior and posterior hyperparameters and plot the distributions
- Compare the behavior of the distribution of each parameter and state a conclusion.

The reason to use the likelihood as the normal with known variance is since the variance (σ^2) presents less variance than the mean (μ) based on the CRLB proposal so it should be fixed.

a) Prior hyperparameters calculation

-Partition 1: $SLI_{2015\,T1}$

The 1000 observations of the mean and the standard deviation (σ) of the distribution $SLI_{2015\,T1}$ from the simulation were fitted for a normal distribution considering this is the conjugate prior distribution and the mean of each distribution was obtained to get the prior hyperparameters.

$$\mu_{0_1} = 0.2404082$$

$$\sigma_{0_1} = 0.4342296$$

Then, the prior distribution being the normal distribution of the mean with mean μ_{0_1} and standard deviation σ_{o_1} is plotted.

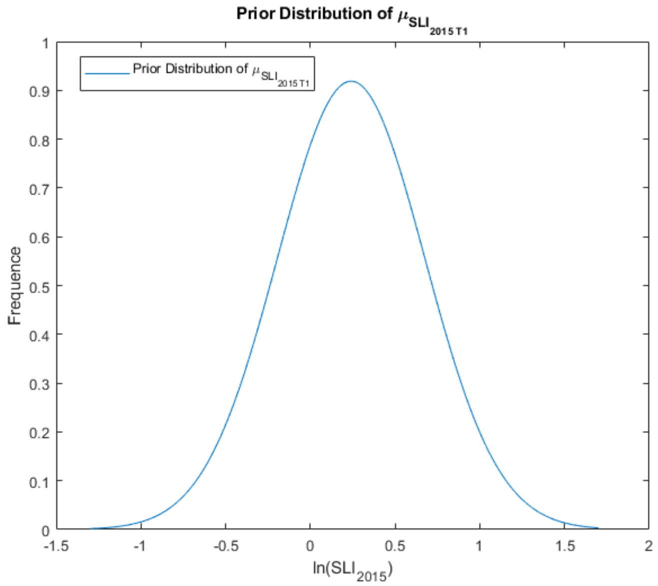

Figure 19: Prior distribution of the mean of $SLI_{2015\,T1}$

-Partition 2: $SLI_{2015\,T2}$

The same simulation is made for the data of $SLI_{2015\,T2}$

$$\mu_{0_2} = 0.6161996$$

$$\sigma_{0_2} = 0.3758478$$

Then, the prior distribution being the normal distribution of the mean with mean μ_{0_2} and standard deviation σ_{0_2} is plotted.

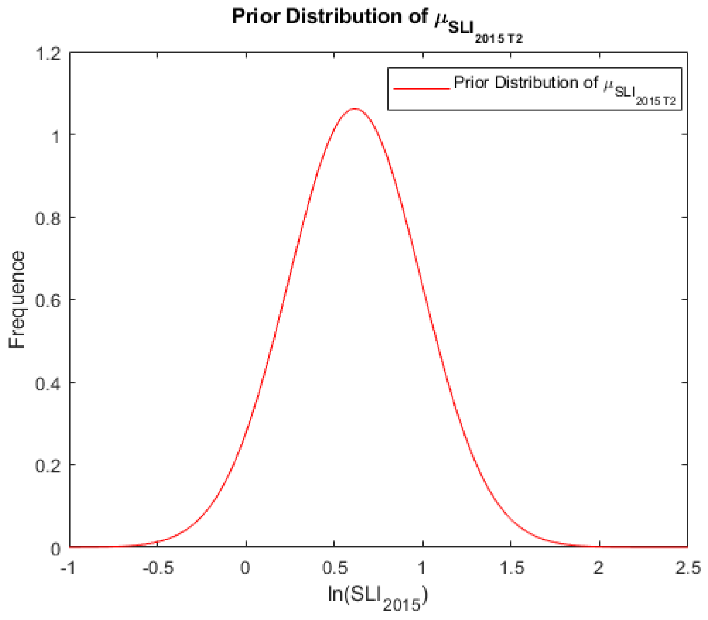

Figure 20: Prior distribution of the mean of $SLI_{2015\ T2}$

To compare both prior distributions, a mixed plot is prepared.

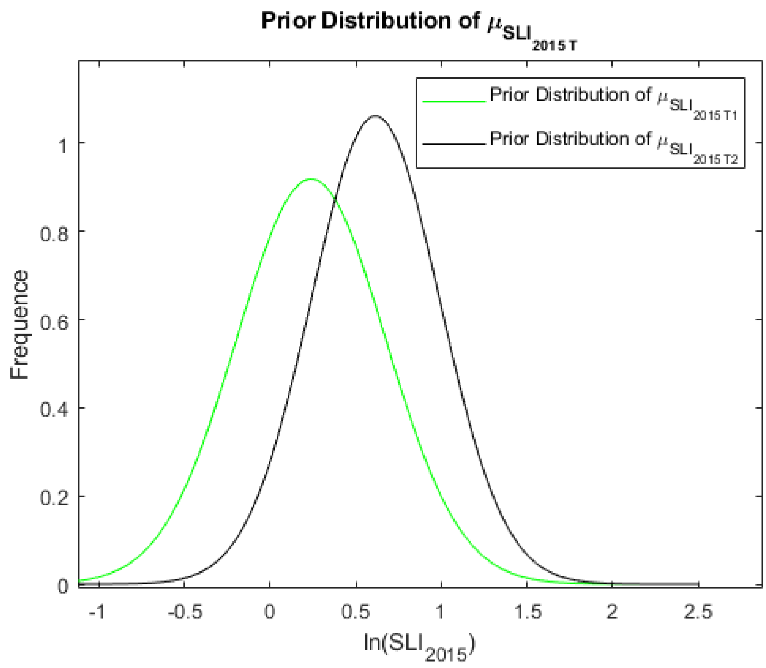

Figure 21: Combined plot of the Prior distributions of the mean of $SLI_{2015\ T1}$ & $SLI_{2015\ T2}$

b) Posterior hyperparameters calculation

Based on the prior hyperparameters calculated, the posterior hyperparameters are calculated.

-*Partition 1*: $SLI_{2015\,T1}$

-Calculation of the posterior hyperparameter of the mean μ_P

$$\mu_P = \frac{1}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^{n} x_i}{\sigma^2}\right) \tag{16}$$

For this partition:

$$n_1 = 193$$

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{193} SLI_{2015\,T1} = 46.61406$$

$$\sigma^2 = (0.4359611)^2 = 0.1900621$$

$$\mu_{P_1} = \frac{1}{\frac{1}{0.4339151^2}+\frac{193}{0.1900621}}\left(\frac{0.2387926}{0.4339151^2}+\frac{46.61406}{0.1900621}\right) = 0.2415094$$

-Calculation of the posterior hyperparameter of the standard deviation σ_P

$$\sigma_P^2 = \left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)^{-1}$$

$$\sigma_{P_1}^2 = \left(\frac{1}{0.4339151^2}+\frac{193}{0.1900621}\right)^{-1} = 0.0009796537$$

$$\sigma_{P_1} = \sqrt{\sigma_{P_1}^2} = 0.03129942$$

Then, the posterior distribution being the normal distribution of the mean with mean μ_{P_1} and standard deviation σ_{P_1} is plotted.

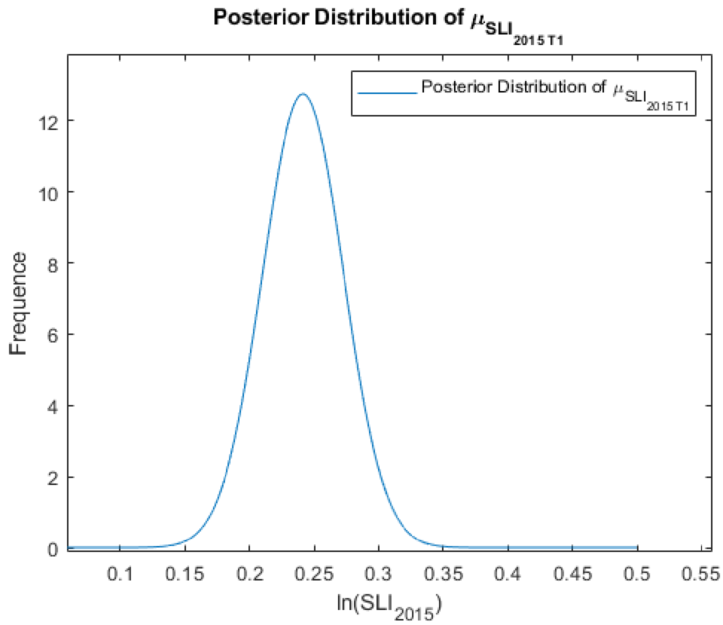

Figure 22: Posterior distribution of the mean of $SLI_{2015\,T1}$

-Partition 2: $SLI_{2015\,T2}$

The same procedure is followed but considering the second partition of the data $SLI_{2015\,T2}$

-Calculation of the posterior hyperparameter of the mean μ_P

$$\mu_P = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^{n} x_i}{\sigma^2} \right)$$

For this partition:

$$n_1 = 736$$

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{736} SLI_{2015\,T2} = 453.7397$$

$$\sigma^2 = (0.3763726)^2 = 0.1416563$$

$$\mu_{P_2} = \frac{1}{\frac{1}{0.3760189^2} + \frac{736}{0.1416563}} \left(\frac{0.6168067}{0.3760189^2} + \frac{453.7397}{0.1416563} \right) = 0.6164946$$

-Calculation of the posterior hyperparameter of the standard deviation σ_P

$$\sigma_P^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)^{-1}$$

$$\sigma_{P_2}^2 = \left(\frac{1}{0.3760189^2} + \frac{193}{0.1416563} \right)^{-1} = 0.0001922062$$

$$\sigma_{P_2} = \sqrt{\sigma_{P_2}^2} = 0.01386385$$

Then, the posterior distribution being the normal distribution of the mean with mean μ_{P_1} and standard deviation σ_{P_1} is plotted.

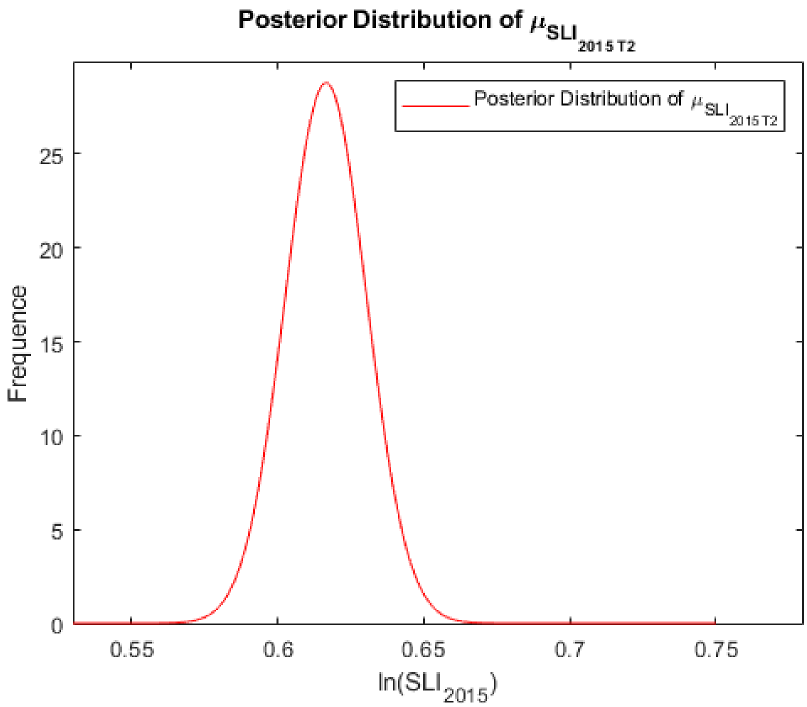

Figure 23: Posterior distribution of the mean of $SLI_{2015\,T2}$

The results from the prior and posterior hyperparameters calculation for the mean (μ) and standard deviation (σ) are presented in the following table:

Table 7 Presentation of results of the prior and posterior hyperparameters for the normal distribution

Variable	Prior μ	Prior σ	Posterior μ	Posterior σ
$SLI_{2015\ T1}$	$\mu_{0_1} = 0.2404082$	$\sigma_{0_1} = 0.4342296$	$\mu_{P_1} = 0.2415094$	$\sigma_{P_1} = 0.03129942$
$SLI_{2015\ T2}$	$\mu_{0_2} = 0.6161996$	$\sigma_{0_2} = 0.3758478$	$\mu_{P_2} = 0.6164946$	$\sigma_{P_2} = 0.01386385$

To compare both posterior distributions, a mixed plot is prepared.

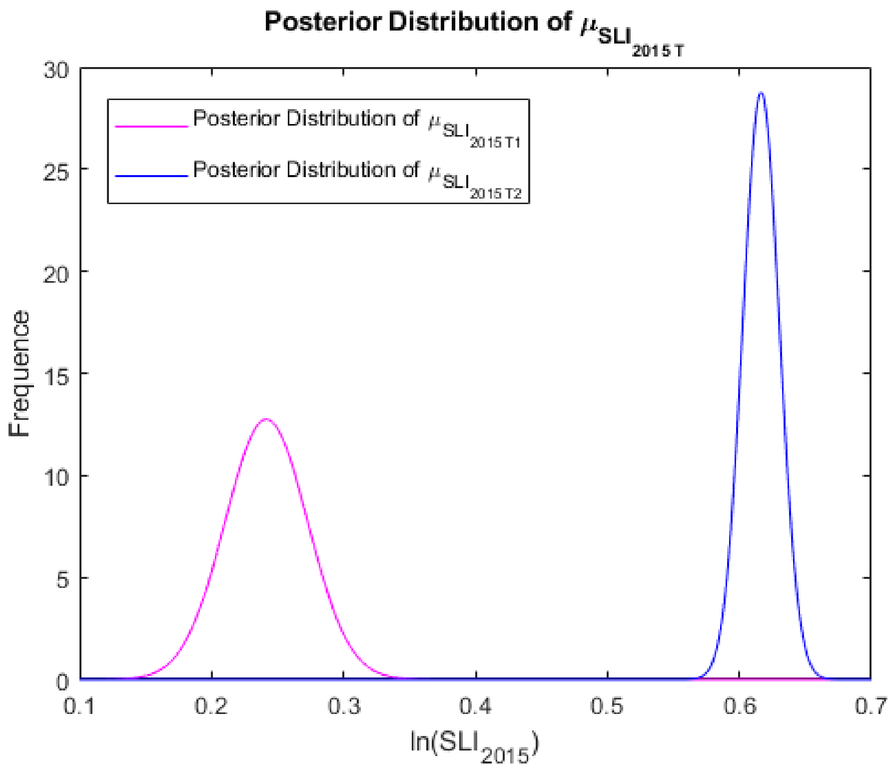

Figure 24: Combined plot of the Posterior distributions of the mean of $SLI_{2015\ T1}$ & $SLI_{2015\ T2}$

To conclude on the application of the Bayesian analysis, even if the prior distributions of the mean presented an overlapping meaning that the null hypothesis considering the equality of the means could not be rejected, the posterior distribution being more indicative of the real behavior of the parameter since it considers the prior distribution as well as the likelihood and the information given by the data. Analyzing the plot of the posterior distribution, if there is just a small overlapping, the means could be equal, but since in this case there is no overlapping at all, it can be concluded that the two means are not equal, or in other words, it can be rejected with certain level of significance that the two means are equal, meaning that they are not equal.

$$\textbf{\textit{Conclusion}} \rightarrow \; H_a: \mu_{SLI_{2015\,T1}} \neq \mu_{SLI_{2015\,T2}}$$

It is important to mention that both the fisherian and the Bayesian approach concluded on the rejection of the null hypothesis, considering the non-equality of the means of social lag index from urban and non-urban municipalities.

4.9 Regression Analysis Approach

The reason for a regression approach from this research comes from the fact that regression analysis allows to predict the value of a variable from certain independent variables [11]. In this specific case, the objective of performing regression analysis comes from identifying if there is a relation between the predictor variable, in this case the total population per municipality $TP_{2015\,T}$ and the dependent variable being the Social Lag Index per municipality $SLI_{2015\,T}$, by interpreting the regression coefficients of the linear model proposed and the level of significance of this regression coefficients it is possible to draw a conclusion on the main hypothesis that is transformed to its regression analysis form.

The regression approach that will be implemented follows 2 different criteria's for concluding on the proposed hypothesis, first regressing the response variable $SLI_{2015\,T}$ against the continuous predictor variable for the population $TP_{2015\,T}$ and checking if the regression coefficient has significance or not using a significance level of $\alpha = 0.05$ which corresponds to checking whether the regression coefficient is 0 or not. The second criteria is to use a categorical variable with two values depending on the partition of the TP_{2015} (Bigger or Lower than 50 000 persons) and regress the $SLI_{2015\,T}$ against this categorical variable, the value of the regression coefficient of the categorical variable will determine the change in the $\ln(SLI_{2015})$ from one category to the other, thus if this difference is considerable and different from 0, a difference in the social lag index from the two populations could be implied.

Having mentioned that, it is important to check on the normality of the response variable $SLI_{2015\,T}$ if a linear model is willed to be implemented [11], in other case, a generalized linear model with a specific link function based on the distribution of the response variable is required to be implemented to make the correction on the error terms based on a linear regression implementation. In this case, the response variable $SLI_{2015\,T}$ follows a normal distribution so a linear model could be directly applied, but either way the results will be checked using a Generalized Linear Model with a Gaussian family function corresponding to an identity link function.

Following this, the transformation of the hypothesis into its regression form for each of the proposal mentioned is the following:

-Proposal 1:

H_0 : The regression coefficient of regression $SLI_{2015\,T}$ against $TP_{2015\,T}$ is 0 with an $\alpha = 0.05$ level of significance.

H_a : The regression coefficient of regression $SLI_{2015\,T}$ against $TP_{2015\,T}$ is different from 0 with an $\alpha = 0.05$ level of significance.

$$H_0: \beta_{SLI_{2015\,T}\,|\,TP_{2015\,T}} = 0$$

$$H_a: \beta_{SLI_{2015\,T}\,|\,TP_{2015\,T}} \neq 0$$

Where:

$\beta_{SLI_{2015\,T}\,|\,TP_{2015\,T}}$=Regression Coefficient of $TP_{2015\,T}$ when regressing $SLI_{2015\,T}$ against $TP_{2015\,T}$

-Proposal 2:

H_0 : The p-value of the regression coefficient of regressing $SLI_{2015\,T}$ against the categorical variable of TP_{2015} $(TP_{2015\,c})$ is bigger than a level of significance $\alpha = 0.05$.

H_a : The p-value of the regression coefficient of regressing $SLI_{2015\,T}$ against the categorical variable of TP_{2015} $(TP_{2015\,c})$ is lower than a level of significance $\alpha = 0.05$.

$$H_0: p - value(\beta_{SLI_{2015\,T}\,|\,TP_{2015\,C}}) > \alpha$$

$$H_a: p - value(\beta_{SLI_{2015\,T}\,|\,TP_{2015\,C}}) \leq \alpha$$

Where:

Table 8 Explanation of the variables involved in the regression hypothesis.

Variable Representation	Variable Meaning	Details
$\beta_{SLI_{2015\,T}\,|\,TP_{2015\,C}}$	Regression Coefficient of $TP_{2015\,T}$ when regressing $SLI_{2015\,T}$ against $TP_{2015\,T}$	Level of significance employed to evaluate the significance of the regression coefficient. $\alpha = 0.05$
$TP_{2015\,C}$	Categorical variable of TP_{2015}	If $TP_{2015} > 50\,000 \rightarrow TP_{2015\,C} = 0$ If $TP_{2015} \leq 50\,000 \rightarrow TP_{2015\,C} = 1$

Before the implementation of the regression models, a 3D plot was made to study the impact in the social lag from the partition of the data based on the categorization made on the TP_{2015} variable.

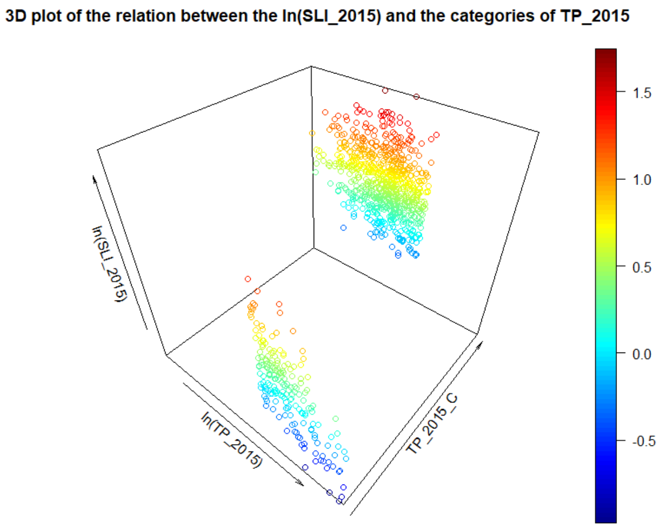

Figure 25: Plot of the relation between the ln(SLI_{2015}) and the categories of TP_{2015}

This plot is a clear indicator that there is a relation between the population of the municipalities with the social lag index since the z $scale$ of the graph represents the increment in the social lag and the background plane represents when $TP_{2015\,C} = 1$ meaning that the $TP_{2015} \leq 50\,000$. The red dots represent municipalities with higher social lag and since there is a much higher density of red plot when the municipalities have lower populations, more populated municipalities have a lower social lag. Still this relation will be analyzed from a regression analysis perspective, but this plot presents and hint of the results that will be found in future analysis.

The first proposal will be evaluated first with an implementation in the software to propose the linear regression model between $SLI_{2015\,T}$as the response variable and $TP_{2015\,T}$ as the predictor variable.

-The regression equation proposed is:

$$\ln(SLI_{2015}) = 2.1397 - 0.1606 \ln\,(TP_{2015}) \qquad (17)$$

-The sum of squares error explained by TP_{2015} is of 27.998 out of a total error of 162.44 representing 17.2%, meaning that 17.2% of the variance of $SLI_{2015\,T}$ is explained by $TP_{2015\,T}$.

$$R^2 = 0.1724$$

$$R_a^2 = 0.1715$$

-The significance of the regression coefficient is explained by the p-value representing the possibility that we have this data given that the regression coefficient is 0.

$$\ln(SLI_{2015}) = \beta_0 + \beta_1 \ln\,(TP_{2015\,T}) \qquad (18)$$

$$p - value_{\beta_0} = \ 2\,x\,10^{-16}$$

$$p - value_{\beta_1} = \ 2\,x\,10^{-16}$$

Both regression coefficients are very significative to the model since the p-value is much lower than the level of significance $\alpha = 0.05$.

-The standard error of β_1 is:

$$s\{\beta_1\} = 0.01156$$

The standard error of the regression coefficient is considerably small meaning that it does not have to much variance and is a good estimation of the real regression coefficient.

-Finally, the median of the residuals is the following:

$$\widetilde{\mu_e} = 0.00319$$

The median of the residuals is very close to 0, meaning that the distribution of the error is centered in 0.

To study a bit more the distribution of the residuals, a scatter plot of the residuals e_i against the predictor variable $TP_{2015\,T}$ and a scatter plot of $SLI_{2015\,T}$ against TP_{2015} are performed:

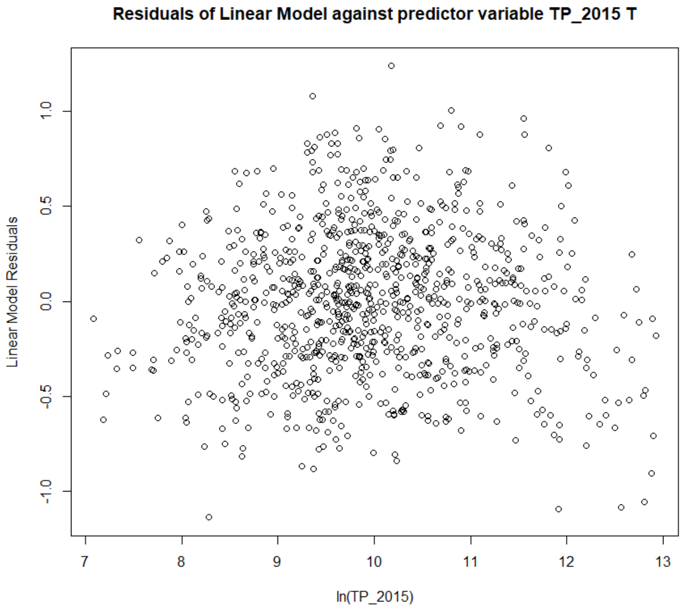

Figure 26: Scatter plot of the residuals of the linear model against the predictor variable $TP_{2015\,T}$

From this plot the residuals follow a random pattern centered and no apparent heteroscedasticity of the error terms is present meaning that a correct regression model was fitted to the data.

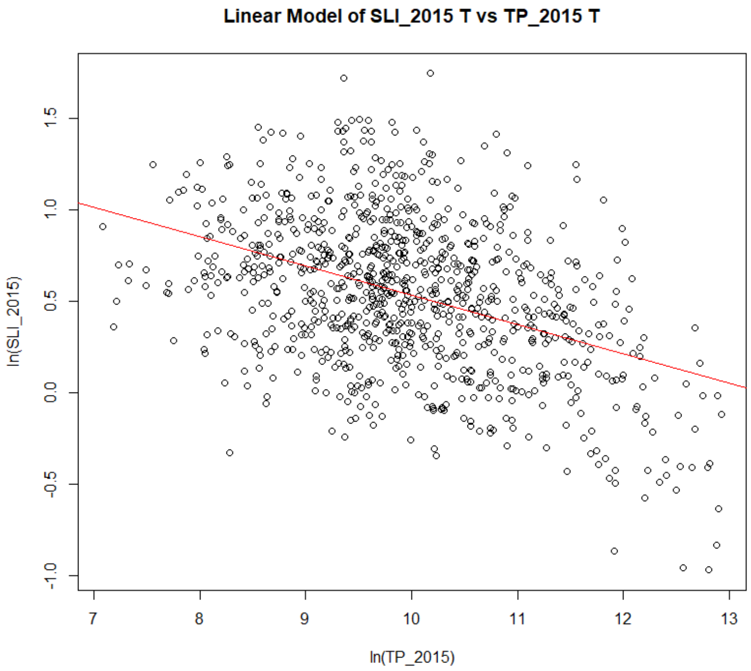

Figure 27: Scatter plot of $SLI_{2015\ T}$ against $TP_{2015\ T}$ with the regression line

From this plot, the regression equation line seems to adjust to the tendency of the data, and that this tendency follows an inverse relation, meaning that the social lag seems to be greater in municipalities with lower populations. The same procedure was applied using a Generalized Linear Model with the gaussian family using an identity link function obtaining the exact same results as previous.

Concluding on this first approach, a good regression model was proposed from analyzing the distribution of the error terms, and the standard error of the regression coefficient. On the other hand, the significance of the regression coefficient and its value not being 0 proposes a not so strong relation between the variables but still a significant one, meaning that it can be concluded with 95% of confidence that the regression coefficient is different from 0 and that there is an inverse linear relation between the social lag index and the population of the municipalities. Thus, the conclusion is to reject the null hypothesis that the regression coefficient is 0.

Since

$$\boldsymbol{H_a}: \boldsymbol{\beta_{SLI_{2015\,T}\,|\,TP_{2015\,T}}} \neq \mathbf{0}$$

➔ We conclude H_a.

The second proposal will be performed with a software implementation of the linear regression model between $SLI_{2015\,T}$ as the response variable and $TP_{2015\,C}$ as the predictor variable.

The obtained results are the following:

-The regression equation proposed is:

$$\ln(SLI_{2015}) = 0.2415 + 0.3750\, TP_{2015\,C} \tag{19}$$

The interpretation from this regression equation is slightly different from the previous since the predictor variable is categorical, in this case the intercept $\beta_0 = 0.2415$ is the mean response of the ln (SLI_{2015}) when the categorical variable is 0 ($TP_{2015} > 50\,000$), so when the predictor variable is 1 ($TP_{2015} \leq 50\,000$) the natural logarithm of the SLI_{2015} is incremented by $\beta_1 = 0.3750$. This means that when the population of the municipalities is lower than 50 000 people, the social lag index is bigger, meaning that the social lag index increments as the municipalities are less populated. This is in accordance with the previous proposition and the previous results found. In this case, the analysis of variance proposes that the error explained by $TP_{2015\,C}$ is of 21.499 out of a total error of 162.44 resulting in a 13.24% explanation of the variability of the $SLI_{2015\,T}$.

$$R^2 = 0.1323$$

$$R_a^2 = 0.1314$$

Still, the coefficient of determination of the regression model is not clearly interpretable when the predictor variable is a categorical variable such as this case.

-The p-value of the regression coefficients are:

Following:

$$\ln(SLI_{2015}) = \beta_0 + \beta_1 \ln\,(TP_{2015\,C}) \qquad (20)$$

$$p - value_{\beta_0} = \ 2\ x\ 10^{-16}$$

$$p - value_{\beta_1} = \ 2\ x\ 10^{-16}$$

Both regression coefficients are very significative to the model since the p-value is much lower than the level of significance $\alpha = 0.05$.

-The standard error of β_1 is:

$$s\{\beta_1\} = 0.03153$$

The standard error of the regression coefficient is considerably small meaning that it does not have to much variance and is a good estimation of the real regression coefficient.

-The median of the residuals is the following:

$$\widetilde{\mu_e} = 0.01357$$

The median of the residuals is close to 0, meaning that the distribution of the error is centered in 0.

A scatter plot of the residuals e_i against the categorical predictor variable $TP_{2015\,C}$ is performed to check on the distribution of the error terms for the 2 categories of the predictor variable.

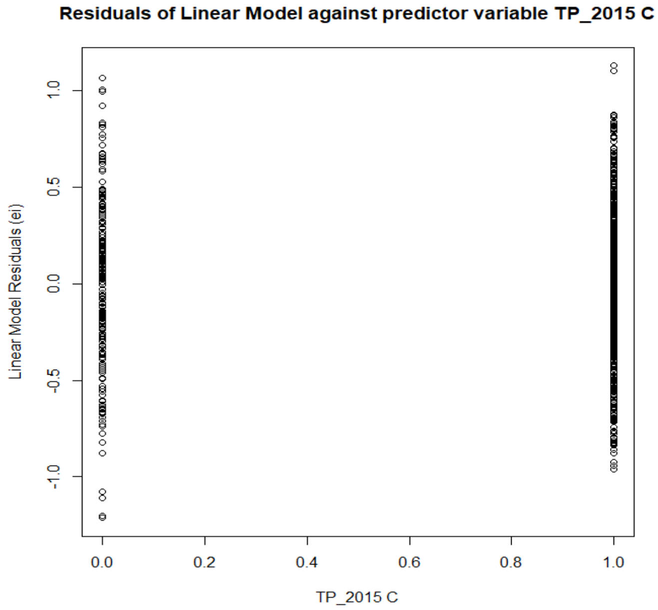

Figure 28: Residuals of Linear Model against categorical predictor variable $TP_{2015\ C}$

The interpretation of this residual plot is a bit different from conventional plot since the residuals only have two different values of the predictor variables, still it can be seen that in both values of the categorical predictor variable $TP_{2015\ C}$ the distribution of the error terms seems to follow a normal distribution around 0 which is confirmed by the quartiles of the residuals being the following:

$$\min e_i = -1.21325$$

$$\max e_i = 1.12731$$

$$Q_1\ e_i = -0.25810$$

$$Q_3\ e_i = 0.25099$$

Following the analysis of the residual plot and the quantiles of the error following a normal distribution implying that a good model was fitted.

In conclusion from this approach, a correct regression model was proposed since the error terms are correctly distributed and the standard error of the regression coefficient has low variability. On the other hand, the significance of the β_1 regression coefficient proposes that the categorical variable is significant to the model, meaning that there is a difference in the social lag from the two groups on which the partition was made. More precisely and as we already mentioned, there is an average increment of 0.3750 on the $\ln(SLI_{2015})$ when the municipalities have a population lower than 50 000 persons meaning that the population of the municipalities has an impact on the social lag.

As in previous approach, it can be concluded that the regression coefficient of the categorical predictor variable is very significant meaning that there is a difference in the social lag from the two groups considered based on the partition of the population of the municipalities. So, this second approach proposes again the rejection of the null hypothesis being the non-significance of the regression coefficient and proposes accepting the alternate hypothesis.

Thus, the conclusion is to reject the null hypothesis meaning the regression coefficient is significant to the model:

Since:

$$\boldsymbol{H_a: p - value(\beta_{SLI_{2015\,T}\,|\,TP_{2015\,C}}) \leq \alpha}$$

$$\boldsymbol{H_a: 2\ x\ 10^{-16} \leq 0.05}$$

➔ We conclude H_a.

From the conclusion of the two approaches on rejecting the null hypothesis, it can be seen from the regression analysis perspective that the total population has a significative effect in the social lag of the municipalities based on the significance of the regression coefficient and the previously analyzed results.

5. Conclusion and Feature Research Direction

In order to conclude on this research and based on the aim of the project being the proving and study of the relation between the population of Mexican municipalities and their social lag, and more specifically on the difference between municipalities considered urban areas and non-urban areas based on their population, the contributions of the study resided in proving by mathematical and statistical means the significant difference in the social lag between this two groups of municipalities by analyzing the CONEVAL database from the intercensal survey in 2015.

This procedure consisted first in studying the social and demographic context in Mexico as well as previous literature that have studied this type of relation and phenomenon before to enrich and contribute to this knowledge to hopefully end with positive pragmatic results. After that, the dataset was studied, cleaned, and properly edited to be able to process each of the variables in a statistical software. With all the information required and ready to be processed, descriptive statistics tools were used to understand the variables involved in the research and the relation between them to propose correct transformation of the variables to reduce the existence of outliers and be able to fit known distributions to them. Then a simulation of a goodness of fit test was applied to find the distribution best explaining the variable for future considerations, especially for the fisherian and Bayesian approach. With the correct identification of the distribution and the parameters involved, fisherian and Bayesian techniques were implemented to calculate the maximum likelihood estimators with their variance as well as the prior and posterior hyperparameters for the posterior distribution of parameters. Finally, a regression analysis technique was implemented to study the relation between the two variables if a deeper manner and complement the previous approaches.

Based on the results from each technique and on the comparison of results between the different techniques it is possible to determine the significant relation between the social lag and the population of municipalities since from each approach, the conclusion from the test hypothesis proposed that the mean of the social lag from municipalities considered urban areas was lower than municipalities considered non-urban areas. From the regression analysis, this direction of the relation was evidenced since an inverse linear relation was proposed for the variability of the social lag as the population of municipalities increased. Since all the different statistical techniques agreed on the same results, it is feasible to conclude on the impact of the municipality population on their social lag.

Practical implementations of this results could be handled from the government as important considerations for future public policies and for the development of the official federation diary which is updated yearly from focusing on sustainable and long-term demographic planification seeking to guarantee the positive aspects of urban areas in the reduction of the social lag in its populations. The limitations when trying to apply this knowledge as considerations for public policies development is that even if it is already demonstrated that the population and the urbanized aspect of the municipality has an impact in the social lag, it is still not clear what aspects of demographic density and urban areas conditions are the ones generating a reduction in the social lag. Having mentioned that, a future project research could focus on identifying by means of regression analysis and statistical inference techniques by using this same database and future databases following the same guideline from CONEVAL, which are the variables of the social lag that are more correlated with the demographic conditions of the municipalities; but also which are the critical variables in predicting a decrease in the social lag in order to focus on them from public policies and try to guarantee this conditions in rural areas, or contemplate this aspects in the planification of future urbanized areas.

6. References

[1] CONEVAL (Consejo Nacional de Evaluación de la política de Desarrollo Social). (2011). Multidimensional Measurement of poverty in Mexico: an economic wellbeing and social rights approach (p. 16). CONEVAL (Consejo Nacional de Evaluación de la política de Desarrollo Social).

[2] Ornelas, P. V. (2018). Rural poverty in Mexico: prevalence and challenges (p. 10). CONEVAL (Consejo Nacional de Evaluación de la política de Desarrollo Social). www.coneval.org.mx.mx

[3] United Nations. (2012). Population, territory, and sustainable development. Population, Territory and Sustainable Development, 53(9), 1689–1699.

[4] Vargas-Chanes, D., & Valdés-Cruz, S. (2019). A longitudinal study of social lag: regional inequalities of growth in Mexico 2000 to 2015. Journal of Chinese Sociology, 6(1). https://doi.org/10.1186/s40711-019-0100-6

[5] T. Wassink, J. (2018). Is Local Social Development Associated with Workforce Composition? A Municipal Analysis of Mexico, 1990–2015. Population Research and Policy Review, 37(6), 941–966. https://doi.org/10.1007/s11113-018-9490-4

[6] CONEVAL (Consejo Nacional de Evaluación de la política de Desarrollo Social). (2007). Los mapas de Pobreza en México: Anexo metodológico (Issue 160, p. 10). CONEVAL (Consejo Nacional de Evaluación de la política de Desarrollo Social). www.coneval.gob.mx

[7] Applied Linear Statistical Models (5th ed.). Michael H. Kutner, Christopher J. Nachtsheim, John Neter y William Li. 2005. McGraw-Hill.

[8] Sund, R. (2019). Computer Age Statistical Inference: Algorithms, Evidence, and Data Science Bradley Efron and Trevor Hastie Institute of Mathematical Statistics Monographs Cambridge University Press, 2016,(8th printing 2018), 475 pages, ISBN: 978-1-107-14989-2.

[9] F. Bartolucci, L. Scrucca. (2010). Point Estimation Methods with Applications to Item Response Theory Models. International Encyclopedia of Education , Third Edition, 366-373. 2009.

[10] CONEVAL (Consejo Nacional de Evaluación de la política de Desarrollo Social). (2016). Índice de Rezago Social 2015: Presentación de Resultados. In Índice de Rezago Social 2015: Presentación de Resultados (Issue 160, pp. 1–10). CONEVAL (Consejo Nacional de Evaluación de la política de Desarrollo Social). www.coneval.org.mx

[11] Deborah R. Abrahams . (2007). Introduction to Regression . 2021, de Princeton University Library Sitio web: https://dss.princeton.edu/online_help/analysis/regression_intro.htm

[12] Gopal K. Kanji . (2006). 100 Statistical Tests Third Edition. 2006: SAGE Publications Ltd.

[13] Murphy, K. P. (2007). Conjugate Bayesian Analysis of the Gaussian Distribution (Vol. 1, Issue 7, pp. 1–29). University of British Columbia. http://www-devel.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf%5Cnpapers2://publication/uuid/b216e20f-0d0e-4e56-b7b6-99e45f145814

[14] U.S. Department of Commerce . (1994). Geographic Areas Reference Manual. Suitland, Maryland, U.S.: Economics and Statistics Administration .

[15] Emile Durkeim. (1933). The Division of Labor in Society . Glencoe, Illinois, U.S.: The Free Press of Glencoe, Illinois.

[16] Carlo M. Cipolla. (1962). The Economic History of World Population. Harmondsworth, Middlesex, G.B.: Penguin Books .

[17] Douglass C. North . (1981). Structure and Change in the Economic History . New York, U.S.: Norton.

[18] Everett M. Rogers . (1995). Diffusion of Innovations . New York, U.S.: Free Press.

[19] William F. Ogburn. (1922). Social Change with respect to Culture and Original Nature. New York, U.S.: B.W. Huebsch, Inc.

[20] Press, S. James . (1989). Bayesian Statistics: Principles, Models, and Applications . Hoboken, New Jersey, U.S.: Wiley.

[21] Jonathan R. Olsen, Natalie Nicholls, Richard Mitchell, Are urban landscapes associated with reported life satisfaction and inequalities in life satisfaction at the city level? A cross-

sectional study of 66 European cities, Social Science & Medicine, Volume 226, 2019, Pages 263-274, ISSN 0277-9536, https://doi.org/10.1016/j.socscimed.2019.03.009. (https://www.sciencedirect.com/science/article/pii/S0277953619301510)

[22] Natalia R. Jones, Iain R. Lake, The combined impact of rural residence and socio-economic status on premature mortality, Health & Place, Volume 24, 2013, Pages 90-96, ISSN 1353-8292, https://doi.org/10.1016/j.healthplace.2013.08.010.

[23] M. Kaczmarek, J. Pacholska-Bogalska, W. Kwaśniewski, J. Kotarski, B. Halerz-Nowakowska, A. Goździcka-Józefiak, The association between socioeconomic status and health-related quality of life among Polish postmenopausal women from urban and rural communities, HOMO, Volume 68, Issue 1, 2017, Pages 42-50, ISSN 0018-442X, https://doi.org/10.1016/j.jchb.2016.11.004. (https://www.sciencedirect.com/science/article/pii/S0018442X16300713)

[24] Awang, Azahan & Jamaluddin, M.J. & Lukman, Z.M. & Aiyub, Kadaruddin & Arifin, Kadir. (2009). The quality of life in Malaysia's intermediate city: Urban dwellers perspective. European Journal of Social Sciences. 9. 161-167.

[25] Simon Appleton, Lina Song, Life Satisfaction in Urban China: Components and Determinants, World Development, Volume 36, Issue 11, 2008, Pages 2325-2340, ISSN 0305-750X, https://doi.org/10.1016/j.worlddev.2008.04.009. (https://www.sciencedirect.com/science/article/pii/S0305750X08001721)

[26] Chunping Han, Explaining the subjective well-being of urban and rural Chinese: Income, personal concerns, and societal evaluations, Social Science Research, Volume 49, 2015, Pages 179-190, ISSN 0049-089X, https://doi.org/10.1016/j.ssresearch.2014.08.006. (https://www.sciencedirect.com/science/article/pii/S0049089X14001616)

[27] Thomas, M. L. (1991). Social changes and problems emanating from industrialization in the ASEAN region. Regional Development Dialogue, 12(1), 1-20. Retrieved from www.scopus.com

[28] Etienne Audureau, Stéphane Rican, Joël Coste, From deindustrialization to individual health-related quality of life: Multilevel evidence of contextual predictors, mediators and modulators across French regions, 2003, Health & Place, Volume 22, 2013, Pages 140-152,

ISSN 1353-8292, https://doi.org/10.1016/j.healthplace.2013.03.004. (https://www.sciencedirect.com/science/article/pii/S1353829213000403)

[29] Yu, W. -., & Chiu, C. -. (2016). Growing pains: Changes in psychological well-being in urban china. Social Indicators Research, 129(3), 1349-1382. doi:10.1007/s11205-015-1166-

[30] Tononi, M., Pietta, A., & Bonati, S. (2017). Alternative spaces of urban sustainability: Results of a first integrative approach in the italian city of brescia. Geographical Journal, 183(2), 187-200. doi:10.1111/geoj.12207

[31] Parvaiz A. Bhat, Mifta ul Shafiq, Abaas A. Mir, Pervez Ahmed, Urban sprawl and its impact on landuse/land cover dynamics of Dehradun City, India, International Journal of Sustainable Built Environment, Volume 6, Issue 2, 2017, Pages 513-521, ISSN 2212-6090, https://0-doi-org.biblioteca-ils.tec.mx/10.1016/j.ijsbe.2017.10.003.

[32] Annie Harrison, Mel Hall, Annemarie Money, Julia Mueller, Hannah Waterson, Arpana Verma, Engaging older people to explore the age-friendliness of a rural community in Northern England: A photo-elicitation study, Journal of Aging Studies, Volume 58, 2021, 100936, ISSN 0890-4065, https://0-doi-org.biblioteca-ils.tec.mx/10.1016/j.jaging.2021.100936. (https://0-www-sciencedirect-com.biblioteca-ils.tec.mx/science/article/pii/S0890406521000268)

[33] V. Dimitra Pyrialakou, Konstantina Gkritza, Jon D. Fricker, Accessibility, mobility, and realized travel behavior: Assessing transport disadvantage from a policy perspective, Journal of Transport Geography, Volume 51, 2016, Pages 252-269, ISSN 0966-6923, https://0-doi-org.biblioteca-ils.tec.mx/10.1016/j.jtrangeo.2016.02.001. (https://0-www-sciencedirect-com.biblioteca-ils.tec.mx/science/article/pii/S0966692316000144)

[34] Sinha, B. R. K. (2015). A methodology of measuring human health. Human Geographies, 9(2), 225-240. doi:http://0-dx.doi.org.biblioteca-ils.tec.mx/10.5719/hgeo.2015.92.7

[35] Kotulla, T., Jon, M. D., Oust, A., & Beusker, E. (2019). What does it take to make the compact city liveable for wider groups? identifying key neighbourhood and dwelling features. Sustainability, 11(12) doi:http://0-dx.doi.org.biblioteca-ils.tec.mx/10.3390/su11123480

[36] Kumari, K., & Dhanda, B. (2020). Assessment of stress among working women in rural and urban areas. Indian Journal of Health and Wellbeing, 11(10-12), 596-599. Retrieved from https://0-search-proquest-com.biblioteca-ils.tec.mx/scholarly-journals/assessment-stress-among-working-women-rural-urban/docview/2493869038/se-2?accountid=11643

[37] Nikhar, H., Meshram, S., & Shinde, G. (2012). An anthropometric and hematological comparison of sickle cell disease children from rural and urban areas. Indian Journal of Human Genetics, 18(1), 40-42. doi:http://0-dx.doi.org.biblioteca-ils.tec.mx/10.4103/0971-6866.96643

[38] Nengsih, R., Juli Dwina, P. S., Maulida, Z., & Sijabat, F. N. (2015). The effectiveness of urban independent community empowerment program (PNPM) in tackling poverty in indonesia. Journal of Asian Scientific Research, 5(6), 320-327. doi:http://0-dx.doi.org.biblioteca-ils.tec.mx/10.18488/journal.2/2015.5.6/2.6.320.327

[39] Rajaei, S. A., & Mansourian, H. (2017). Urban growth and housing quality in iran. Social Indicators Research, 131(2), 587-605. doi:http://0-dx.doi.org.biblioteca-ils.tec.mx/10.1007/s11205-016-1260-2

[40] Ploch, L. A. (1977). The Reversal in Migration Patterns - Some Rural Development Consequences.

[41] Khanani, R.S., Adugbila, E.J., Martinez, J.A. et al. The Impact of Road Infrastructure Development Projects on Local Communities in Peri-Urban Areas: the Case of Kisumu, Kenya and Accra, Ghana. Int. Journal of Com. WB 4, 33–53 (2021). https://0-doi-org.biblioteca-ils.tec.mx/10.1007/s42413-020-00077-4

[42] Khorasani, M., Zarghamfard, M. Analyzing the Impacts of Spatial Factors on Livability of Peri-Urban Villages. Soc Indic Res 136, 693–717 (2018). https://0-doi-org.biblioteca-ils.tec.mx/10.1007/s11205-016-1546-4

[43] Glatz, C., Bodi-Fernandez, O. Individual social capital and subjective well-being in urban- and rural Austrian areas. Österreich Z Soziol 45, 139–163 (2020). https://0-doi-org.biblioteca-ils.tec.mx/10.1007/s11614-020-00399-9

[44] Omariba, D.W.R., Boyle, M.H. Rural–Urban Migration and Cross-National Variation in Infant Mortality in Less Developed Countries. Popul Res Policy Rev 29, 275–296 (2010). https://0-doi-org.biblioteca-ils.tec.mx/10.1007/s11113-009-9140-y

[45] Siddiqi, A. Small town growth and development policy in Pakistan. GeoJournal 30, 421–434 (1993). https://0-doi-org.biblioteca-ils.tec.mx/10.1007/BF00807223

[46] Kwon, TH. The Transformation of Korean Life: Demographic Trends and Their Social Implications. Social Indicators Research 62, 19–38 (2003). https://0-doi-org.biblioteca-ils.tec.mx/10.1023/A:1022628730152

[47] Konadu-Agyemang, K., Shabaya, J. What has Corruption Got to do with it? Understanding the Persistence of Rural-Urban and Inter-Regional Inequalities in Ghana and Zimbabwe. GeoJournal 62, 129–146 (2005).

[48] Mitrică, B., Şerban, P., Mocanu, I. et al. Social Development and Regional Disparities in the Rural Areas of Romania: Focus on the Social Disadvantaged Areas. Soc Indic Res 152, 67–89 (2020). https://0-doi-org.biblioteca-ils.tec.mx/10.1007/s11205-020-02415-7

[49] Hendriks, F. The post-industrialising city: political perspectives and cultural biases. GeoJournal 47, 425–432 (1999). https://0-doi-org.biblioteca-ils.tec.mx/10.1023/A:1007034010709

[50] Montgomery, M.R., Hewett, P.C. Urban poverty and health in developing countries: Household and neighborhood Effects. Demography 42, 397–425 (2005).

[51] Gaspar, J., Brito Henriques, E. & Vale, M. Economic restructuring, social re-composition and recent urban changes in Portugal. GeoJournal 46, 63 (1998). https://0-doi-org.biblioteca-ils.tec.mx/10.1023/A:1006987705815

[52] Bourgeois, R. What Future for Rural Areas? Seven plausible rural transformations. Development 58, 177–186 (2015). https://0-doi-org.biblioteca-ils.tec.mx/10.1057/s41301-016-0017-1

[53] Saharan, T., Pfeffer, K. & Baud, I. Urban Livelihoods in Slums of Chennai: Developing a Relational Understanding. Eur J Dev Res 30, 276–296 (2018). https://0-doi-org.biblioteca-ils.tec.mx/10.1057/s41287-017-0095-2