Master of Science Thesis:

Assessing the spatial relationship between public road infrastructure and the socio-economic indicators of urban poverty in southern Lima, Peru

Joahna Carvallo Salas

April, 2006

Master of Science Thesis: Assessing the spatial relationship between public road infrastructure and socio-economic indicators of urban poverty in Southern Lima, Peru

by

Joahna Carvallo Salas

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialization: Urban Infrastructure Management

Thesis Assessment Board

Name Examiner 1 (Chair): Dr. Richard Sliuzas
Name Examiner 2 (External): Dr. S.C.M. (Stan) Geertman
Supervisor: Dr. Michael McCall
Supervisor: Drs. Sherif Amer

INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
ENSCHEDE, THE NETHERLANDS

Disclaimer

This document describes work undertaken as part of a programme of study at the International Institute for Geo-information Science and Earth Observation. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

To my dear husband Javier and my lovely grandmother Eloisa.

Abstract

The socio-economic benefits caused by the provision of public road infrastructure (PRI) in poor areas may help reduce poverty levels. This benefit is measured in terms of the access gain provided by newly added road infrastructure and its incidence in the improvement of the indicators directly related with poverty. These indicators are also affected by the new services (i.e. electricity, sewerage, etc.) that paved roads ease to provide by improving the accessibility condition, necessary for their implementation. In this context, the present study aims to determine the spatial relationship between the PRI and a socio-economic (SE) classification in Villa El Salvador and Villa Maria del Triunfo, both located in southern Lima, Peru. The importance of the research points toward evidencing the degree of influence of PRI implementation towards the tackling of urban poverty through a strong positive influence on SE indicators in the area served by such. This is achieved in three steps: the first is to determine a SE index using Spatial Multi Criteria Evaluation (SMCE) based on variables from a housing and population census; the second is to create of network model for determining distances from the sampling units (the housing blocks) to the closest paved road; and the third is to find the association level between the previous two using statistical procedures. Histograms, scattergrams and Pearson's Product Moment (PPM) correlations test the variables, correlate them and find out the significance level of their relationships. Complementarily tri-dimensional visualization inside Geographic Information Systems is used to graphically represent these relationships. The results show there is a consistent and important correlation between the distances to paved roads and the socio-economic index in the two temporal scenarios, for 1984 and 1996. The correlation improves in time as the coefficients found are better for the latter scenario, suggesting an improvement in the SE condition due to the implementation of paved roads. In addition, the relationship found evidence the importance of being closer to a paved road in terms of development inside urban poverty areas. The analysis and the methodology used aim to provide planners with a clear view of the incidence of the physical presence of PRI on the SE level of urban poverty areas.

Acknowledgements

Firstly, I thank God for giving me the great opportunity to complete this research. He gave me strength, peace and maturity to reach this goal. In the same way, I thank my dear husband Javier who gave his indispensable love, which comes to me as a gift blessed by God. I am sure this experience in The Netherlands would not be as it is without his precious support. All the words of this world are not enough to express my gratefulness to God for allowing me to stay with Javier day by day.

Several people collaborated with this study by providing help in different ways. Among them, my dear professors Mr. Johan de Meijere and Mr. Arbind Tuladhar kindly provided me their wise advices as well as useful comments for this research. I would not have had better support than theirs. In a special way, I would like to acknowledge the willingness of Ms. Karin Pfeffer and Mr. Javier Martinez to answer my inquiries and provide useful information. I also must express thankfulness to my supervisor Mr. Michael Mc Call who invested some of his time on my research when he was not abroad.

During the fieldwork period in Lima, Mr. Hugo Soto, Director of Planning and Budget Department, Municipality of Villa El Salvador (MVES) allowed me to take part of his valuable time for interview. At the same time, I had the opportunity to carry out important interviews with Mr. Jaime Zea Usca, the Mayor of the MVES, Mr. Alex Cordoba, Director of Public Work Department, Mr. Rafael Zamora and Mr. Michel Azcueta, former mayor. All they helped me in the best possible way, not only with an opportune answer but with important documents. It is worth to remark the contact with these key informants helped the understanding of the development and formation process of the study area before becoming formal districts. I must also give a special recognition to Mr. Javier Cornejo Arana, Head of the Development Management, Autonomous Authority of the Electric Mass Transport System of Lima and Callao – Urban Train (AATE). He provided me all the support he could despite his overloaded agenda. Furthermore, honouring past times as good colleagues and better friends in the office, he provided this study with invaluable information from the AATE database.

Through the knowledge and experience of qualified professionals, valuable secondary data was identified and collected. Among them, Mr. Manuel Cox Ganoza and Mr. Enrique Cadenas from the company of Potable Water and Drainage Service of Lima (SEDAPAL) provided all the data available. In the same way, I must acknowledge Mr. Francisco Costa and Mr. Ricardo Lagunas from the National Institute of Statistic and Informatics (INEI). Their opinions were also useful to find the urban poverty indicators which measure the condition of the human settlements in the study area.

As a good Peruvians, my friends Pedro Vargas and Miller Carhuamaca give disinterested help with a constant positive attitude. They both took part of the observation of the study area in order to obtain information about its urban configuration and verify the secondary data.

Finally, I must thank my dear family in Lima for giving me support, assurance and keeping my spirit up despite some difficult times I had to go through while doing this research. I am sure they are proud of all the effort put in achieving this goal as I am of being one of them.

Table of contents

List of graphics

List of tables

List of photographs

List of acronyms

AATE	Autoridad Autónoma del Proyecto Especial Sistema de Transporte Rápido Masivo para Lima y Callao – Tren Urbano Autonomous Authority for the Special Project of Mass Rapid Transit for Lima and Callao - Urban Train
DEM	Digital Elevation Model
DESCO	Centro para el Desarrollo y la Promoción de Estudios (ONG) Centre of Development and Promotion Studies (NGO)
ESRI	Environmental Science Research Institute
GDP	Gross Domestic Product
GIS	Geographic Information System
ID	Identificator (Primary Key)
ILWIS	Integrated Land and Water Information System
INEI	Instituto Nacional de Estadística e Informática National Institute of Statistics and Informatics
ITC	International Institute for Geo-Information Science and Earth Observation
ONG	Organismo No-Gubernamental Non-Governmental Organization
PRI	Public Road Infrastructure
PEA	Población Económicamente Activa Economically Active Population
PPM	Pearson’s Product Moment
SEDAPAL	Servicio de Agua Potable y Alcantarillado de Lima Potable Water and Sewerage Service of Lima
SE	Socio-Economic
SMCE	Spatial Multi-Criteria Evaluation
TACH	Total Annual Cost per Household
TIN	Triangulated Irregular Network
UP	Urban Poverty
VES	Villa El Salvador
VMT	Villa Maria del Triunfo

Glossary

Adobe: Uncooked clay brick filled with straw fibbers, produced in low-income areas, made on the site by traditional methods. Adobes -of ancient use in Latin America- are left to dry by sunlight and are joined without a mortar. They come on a variety of non-standardized sizes.

Estera: A cheap thick-straw mat that is used for construction purposes in poor urban areas in Peru, where it has been the representation of land invasions since the 60's. Square sized, it has measures around 2 to 2.5 meters by side, being extremely light though easily degradable (See Photo 3).

P.E.A. Población Económicamente Activa or Economically Active Population is an index used by the Peruvian government to identify the group of individuals that are in the age of being productive in monetary terms. This applies to all persons that can carry out an economic activity –whether formal or informal- regardless of their age. Therefore, children of 5 years and above are included in this category.

Quincha: Wooden frame filled with cane and mud, finished with gypsum and exclusively used to build one-floor houses or second/third floors. Quincha is a cheap an ancient construction, though few modern populations still use it.

1. Introduction

1.1. Defining the main issues

One of the most important urban issues deals with the provision of adequate public infrastructure. This ensures the consistency of economic development towards improving the Socio-Economic (SE) condition within urban areas. Cotton and Franceys (1994) state that different infrastructure sectors have different effects on improving the quality of life and reducing poverty. Among these basic needs, the access to public road infrastructure (PRI) is necessary for the economic development as it has a direct influence over the areas it serves. A sustainable city's growth depends -among other factors- on the right access to all centres of attraction, productivity and economic interchange.

As evolving and dynamic entities, most cities in the Peruvian coast have experienced an accelerated growth. In the capital city, Lima, this growth has brought a huge migration phenomenon, having poor urban settlements as a consequence. These squatter areas lack property formalization and thus adequate basic infrastructure, especially those covering basic needs. The PRI provide the abovementioned access to services, facilities and others opportunities, without which the SE development of these settlements is extremely limited.

In addition, human settlements in Lima mainly born where there are neither services nor the facilities to provide it, especially in terms of PRI. There is an inadequate access to areas of interest (work and services) within the city fostered by the geography of these dwellings, mostly inaccessible hilly terrain, somehow unsuitable for housing.

In a more simple level, the same principle for the economic development of a city described above applies to areas of urban poverty. An adequate access to urban basic services may show better SE indicators in poor dwellings. For these emerging areas, tackling poverty is a matter of how much infrastructure is available, in physical and economic terms. As a result, the condition of PRI (existence or not; paved or unpaved) may determine the urban poverty. The PRI benefits imply access to services and employment opportunities, necessary for an adequate SE setup and development.

In the present study, poor areas within the southern part of Lima city, namely Villa el Salvador and Villa Maria del Triunfo, will be assessed using their SE indicators to determine a classification. This classification will then be correlated against the closeness –or farness- to adequate PRI that may connect them with the existing road network and thus the rest of the city. The SE classification will be determined by evaluating diverse indicators related to housing, population or employment figures. The study aims to measure the influence and effectiveness of PRI implementation, seeking to determine the exact range of its importance towards the economic development of depressed urban areas.

1.2. Justification of the research proposal

Along my expertise as a transport planner in Lima, I have been witness of a number of purportedly politically-driven policies towards the most needed areas in the city. During the early 90s, the Peruvian government conducted an extensive plan to provide such access to some of the poorest areas in the Lima's fringe. This policy was neither designed nor proposed by urban planners. It was mainly the resolution of population's claim, who demanded better roads among other services. For example, the provision of drinking water depended on trucks that carried it to the settlements. By improving the road condition, trucks could go further in the hilly areas and minimize the water cost, thus making it more accessible.

As a result, public institutions upgraded numerous roads and constructed some others in poor dwellings in the southern districts of Lima, lacking a designed plan for such. This policy was not even monitored after implemented, resulting on an awkward integration process of squatter areas to the consolidated road network of the city. The results of those policies, in terms of the real benefit for those populations, must be assessed to determine the real importance of PRI in the context of economic development. This will allow a correct redesign process in order to provide the planners with suitable procedures for helping tackling urban poverty (UP) trough the provision of adequate infrastructure.

Consequently, the study will evaluate the validity and implementation of PRI. The aim is to determine its incidence on SE indicators related to UP. This assessment will try to prove if there is a consistent correlation between being closer to paved roads and having a given SE level inside a poverty area. This means the evaluation of the real importance of the infrastructure in the economic development of the populations it serves.

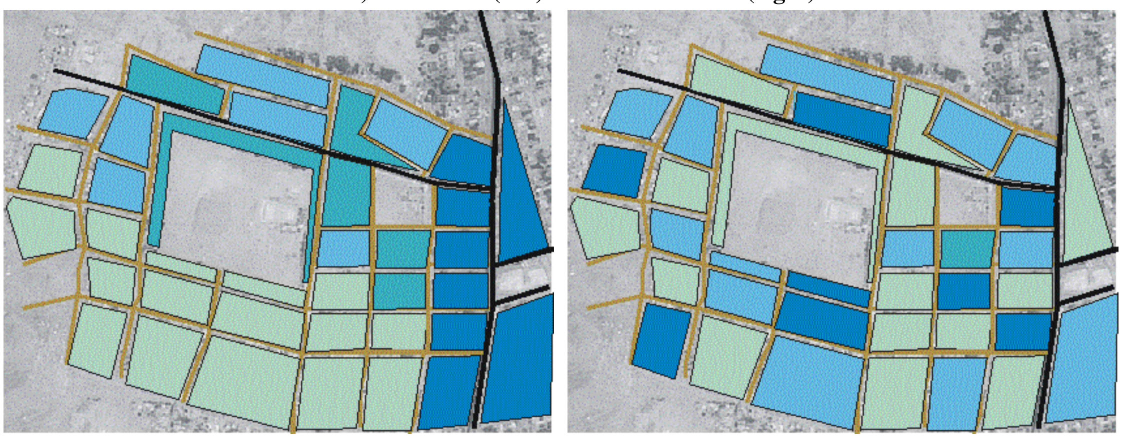
1.3. The research problem

Most of the professionals in charge of developing urban plans for poor areas are somehow conscious that the provision of public infrastructure such roads contribute with the SE development and growth of any urban area. One of the issues for the local authorities in Southern Lima is the lack of information and adequate studies that support the important role of PRI plays in human settlements. In addition, it is important to get to know how much such infrastructure influences the SE indicators or the living condition in those urban areas. Poverty maps are one of the currents tools used for such purpose, however unavailable in the present time. A first step towards this is an adequate identification of SE indicators and the factors that influence UP. The next natural step is to let know the authorities the exact degree of influence of newly added PRI on the improvement of SE layout of the area served. This may facilitate the allocation of funds and the measure of the return when dealing with infrastructure provision.

Several aspects of urban development may influence the SE level in a certain area. The lack of facilities and services are important issues, particularly basic needs such as water, electricity and transport. Among these factors, it is relevant the analysis of an adequate PRI in terms of better access to/from plots and to opportunities, facilities and services. The provision of services mainly starts with the PRI as it provides the adequate access for resources (machinery, materials) that ease the later implementation of other services. Thus, any improvement in UP may start with the right setting provided by an adequate connection to the existing road network.

The research problem is defined as proving an approach to determine the association between the level of PRI development and the SE classification of poverty areas within a city. This development is expressed in terms on the quality of the road surface, namely paved or unpaved. UP areas served by a particular road are to be assessed using common urban poverty indicators such as housing quality, access to services, employment, etc. Once defined, these areas and their indicators will be able to show correspondence with the better setting that may imply access by a paved road (see Graphic 1). Finding out the level of correspondence of the road with a SE pattern in the study area in order to better understand the exact range of their importance towards the socio economic development as well as their influence in the UP alleviation in depressed urban areas.

Graphic 1 : Representation of PRI (paved in black line, unpaved in grey) and SE classification (grey tones) correlated (left) and no correlated (right)



1.4. Research objectives

1.4.1. Main objective

- To determine if the PRI is related with the SE classification in the study area.

1.4.2. Sub objectives

- To produce a SE classification using indicators for the study area.
- To evidence the degree of relationship between distance to PRI and the classification of SE indicators.
- To evaluate the access to the rest of the city by existent PRI and its incidence in the SE classification.

1.5. Hypotheses

1.5.1. Main hypothesis:

- A SE classification is related to the distance from each sampling unit to the PRI in the study area.

1.5.2. Research hypothesis:

- The PRI configuration (road surface quality) has direct influence on the SE classification of the area it serves.
- The distance to PRI and SE classification has a negatively correlated relationship.

1.5.3. Methodology hypothesis:

- The PRI quality and the SE indicators are identifiable, measurable and can be analyzed in a spatial/statistical correlation.

1.6. Research questions

To produce a SE classification for the study area:

- Which indicators are used to determine the SE level of the area?
- How are these indicators combined to produce a SE classification?

To evidence the degree of relationship between distance to PRI and the classification of SE indicators:

- Is there any relationship between closeness to PRI and the location of higher SE classification?
- Does the PRI surface quality influence the spatial distribution of SE indicators? If so:
    - Is this influence positive or negative?
    - Is this relationship consistent?

To evaluate the access to the rest of the city by existent PRI and its incidence in the SE development:

- Is accessibility an issue in the SE classification of the study area?
- Is the access to opportunities and services given by PRI determinant to the SE classification of the study area?

1.7. Research methodology

The timeframe of the study is set in the past time. This is the time of full census data availability, as it is the main source for the analysis. The research methodology is outlined by three relevant steps: First, to determine a SE classification for each sampling (block) unit in the study area using census variables and indicators for a given year. Second is to make a network model to output distances from the sampling units to the paved roads and accessibility indicators using infrastructure available at that time. Third, is to demonstrate the relationships between these classification and network distances to the nearest paved road or the accessibility indexes of the sampling units. These three steps are described in detail in the following paragraphs.

In order to carry out the first step, a spatial multi-criteria evaluation (SMCE) will be used to produce a composite class map using variables and indicators. The variables to be used will be determined through interviews with key informants in the study area. The SMCE procedure creates a structured model of the data that can be easily modified and processed to produce multiple partial and final classifications. In order to be jointly evaluated, these maps will be standardized and weighted with the same values to represent equal importance of all variables in the SMCE. These quantitative data will be collected in the fieldwork from secondary sources (statistical housing and population census for the year 1993). At the same time, the output of the SMCE will represent an initial approach to poverty maps for the study area, having partial outputs for each SE indicators group in the criteria tree.

Then, the second step will start using high-resolution aerial photos from different time series. These will be used to extract road information (photo interpretation techniques) in terms on their physical condition, namely paved or unpaved. Using a base road layer taken from fieldwork, missing roads will be digitized and a network will be created to calculate distances as simulating walking trips from the sampling units (blocks) to the paved roads. This data will be compiled in a database to map the relationship between the two variables (SE classification and distance to the roads) and the influences in the spatial patterns that may be found.

Finally, all the maps produced will output tables that will be exported to database files in order to be statistically analysed. The analysis will comprise the use of histograms, scattergrams and Pearson's Product Moment (PPM) correlations. The procedure is designed to test the main hypotheses and generate answers to the questions defined in 1.4.

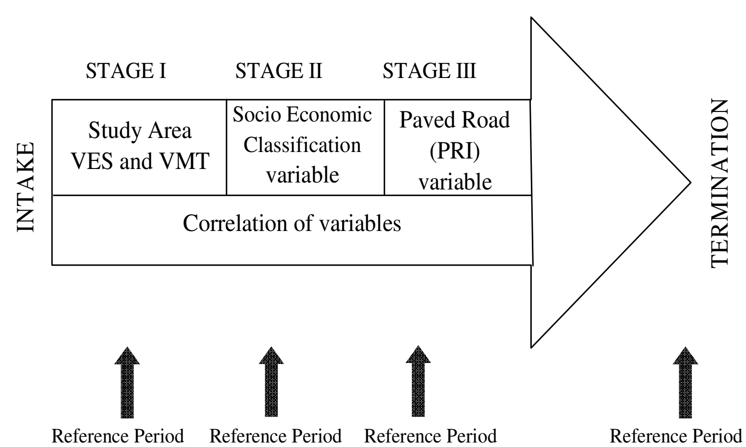
Complementarily, the use of primary data such as interviews, observation of the study area will support several issues of the research. First, they will help in determining the exact boundary of the study area. Second, they may guide the analysis by elucidating better indicators or improving the variable's outline. They can also give a flavour about the perceived influence of PRI in the view of key informants in the study area. All this may lead to tackle the challenge of demonstrating if the presence of adequate PRI has relation with SE patterns in the study area.

1.8. Study design

The study design selected for this research proposal is known as retrospective study. It was selected because the research will be in a particular situation that happened in the past (reference period: 1993). Hence, the areas being connected by PRI will be assessed to determine a correlation with SE indicators. This is done by comparing a SE classification in the study area with the closeness or farness of the sampling units to the implemented PRI. This type of study is defined as cross-sectional, with regard of both the study area and the time of the investigation, at one point in time, on one single study area (see Graphic 2 below). This design, being the most recurred in social sciences, is most appropriate for using secondary data of a certain past time as it is most likely to be found in developing countries.

In addition, the research proposal is classified as an experimental study by the nature of the investigation. This research will start from the cause, namely the setting of PRI, to determine a purported influence on the SE setup in a cause-effect relationship that transcend the time of the study, as this influence should have happen beforehand.

Graphic 2 : The cross-sectional design (adapted from Kumar (1999))



The overall research design comprises four issues, as to say a theoretical context, the literature review, the relationship of variables and the methodological approach. All this together with a technical framework (the analysis itself) will lead the research to conclusions and recommendations.

The overall schema of the study design is shown with more detail in Graphic 3 below. The strategy of investigation is conceived to obtain answers for the research questions. The research questions per sub-objective, the data requirement and technique proposed for answering each one is presented using a matrix, shown in Table 1 below.

Graphic 3 : Research design schema

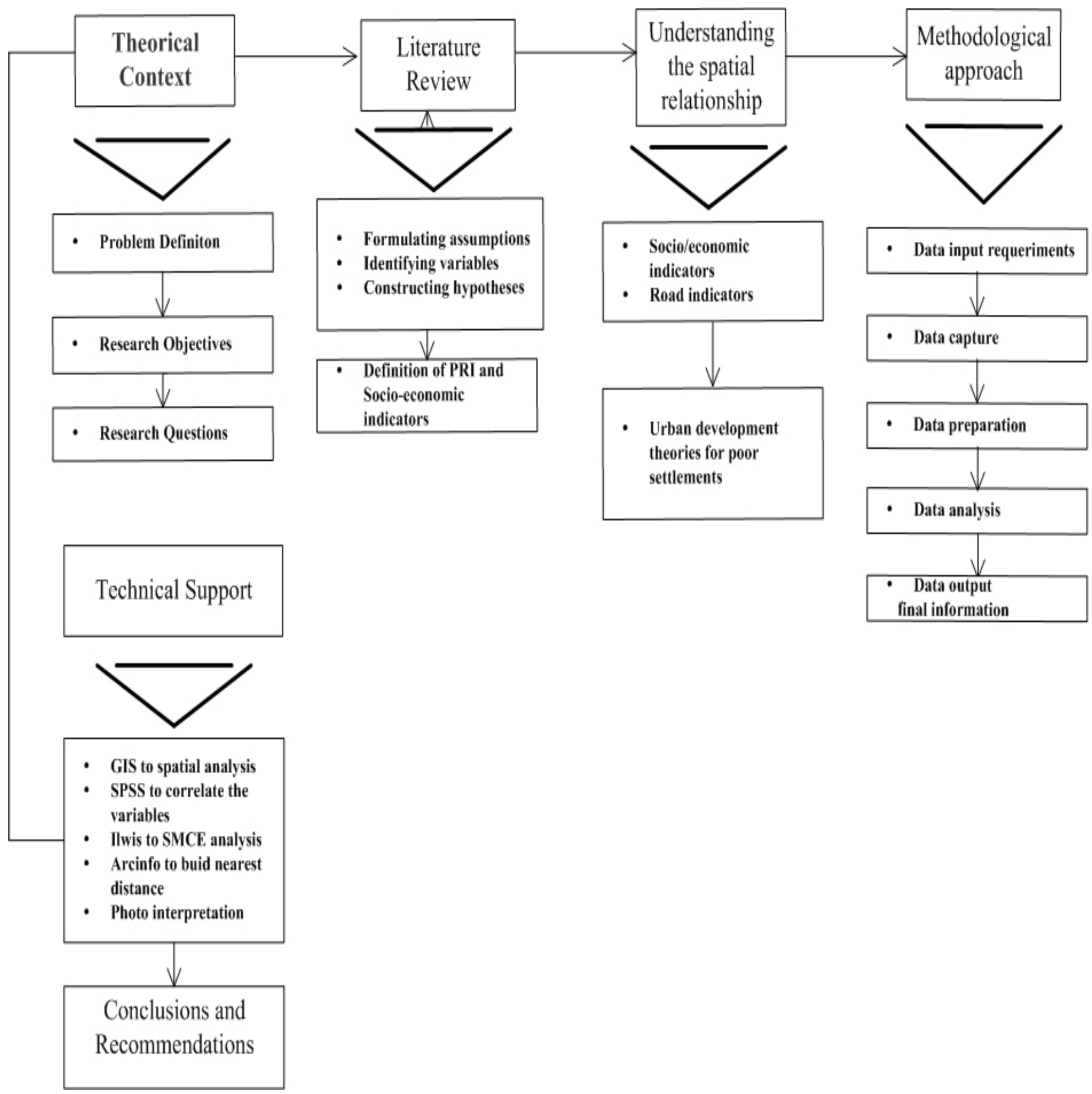


Table 1 : Matrix showing the research tools versus research questions

Research sub-objectives	Research questions	Data requirement	Techniques/tools
To produce a socio-economic classification for the study area	•Which indicators are used to determine the SE classification of the area? • How are these indicators combined to produce a socio-economic classification ?	• Socio-economic indicators (low income settlements). • Census data. • Cadastre maps.	• Literature review. • Interviews with local authorities, experts, local community leaders.
To evidence the degree of relationship between distance to PRI and the classification of socio-economic indicators	• Is there any relationship between closeness to PRI and the location of higher socio/economic classification? • Does PRI surface quality influence the spatial distribution of the SE indicators? If so: Is this influence positive or negative? Is this relationship consistent?	. Socio-economic indicators (low income settlements). • Census data. • Cadastre map. • Aerial photos and images. • Road network.	• Literature review. • Aerial photo/image interpretation. • Projecting and editing from map. • Georeferecing.
To evaluate the access to the rest of the city by existent PRI and its incidence in the socio-economic development	• Is accessibility an issue in SE classification of the study area? • Is the access to opportunities and services given by PRI determinant to the SE classification of the study area?	• Socio-economic indicators (populated and low income settlement). • Census data. • Cadastre map. • District's limits. • Road network.	• Literature review. • Projecting and editing from map. • Georeferecing.

1.9. Variables: Definitions and indicators

1.9.1. Urban poverty

The conventional definition of UP is based upon either per capita incomes or consumption. It is characterized by low incomes, lack of adequate health, and illiteracy. These are the result of processes that have their roots in economic, social, political, and cultural factors (Moser et. al., 1996).

On the other hand, May (2001) points out that the classic definition of UP sees it as “the inability to attain a minimal standard of living” measured in terms of basic consumption needs or the income required for satisfying them. In addition, May mentioned that UP can be considered in two dimensions: absolute and relative. In absolute terms, poor people are materially deprived but in relative sense, they are also deprived in relation to other social groups whose situation is less constraining. Another point of view concerns UP classification in different levels, namely experience, household levels and community. The latter comprises the poverty as the lack of infrastructure and remoteness (as well as instability and disunity), which are directly related with the proposed study.

1.9.1.1. Characteristics and components

The lack of basic roads, transportation and water infrastructure is seen as a defining characteristic of poverty. Therefore, good roads and services are essential to have access to employment, especially in poor urban areas. According to Guillen and Konings (1992) the components of UP are employment and income; housing; general public utilities and services (roads, drinking water, energy, sewerage and public transport) and education/health care. In addition, Guillen et al. affirm that one of the obvious characteristics of UP is that its inclusion in a single category is based on the spatial aspects of reproduction and collective consumption. This means that the identification of poor dwellings depends on the precarious condition of these areas rather than the situation of poor people in the labour process or their income categories. The study will consider this definition for delimiting the study area and its sampling units.

1.9.1.2. Indicators

The present study will use SE indicators which are defined as urban poverty indicators. Moser et al. (1996) mentions that there are two types of UP indicators: priority poverty indicators and key urban poverty indicators. These indicators facilitate analysis, monitoring and comparisons of urban poverty. The relevant key urban poverty indicators related to this research proposal are:

- Access to water and sanitation, the availability of water to urban poor as the percentage of poor households with various means of obtaining water supply. To hold a sewerage service as the percentage of poor households served by different types of provision for sewage disposal.
- Access to electricity supply, defined as the percentage of poor households with a household connection to the electricity network.
- Informal employment, percentage of the employed population, men and women, who have found a job opportunity in the informal sector. According to The Habitat Agenda (2004), the informal sector includes first, all unregistered commercial enterprises, and second, all non commercial enterprises that have no formal structure in terms of organization and operation.
- Population density, the ratio of people living in a specific area. Population densities are measured in terms of number of inhabitants per area unit, generally hectares (inhabit./Ha.), as it is approximately one 100-meter-side squared-shaped block, commonly used in the urban areas. Poor urban areas are characterized by a lack of space for the number on inhabitants and furthermore a bad quality of the overcrowded dwellings, resulting in high density rates.

1.9.2. Public infrastructure

There are numbers of definitions of public infrastructure. According to Brussel (2004), in the traditional sense, infrastructure refers to physical components such as utility networks, ports, bridges, railway, etc. It is also referred as a combination of physical infrastructure (all objects of the built environment) and social infrastructure (social services directed at improving social welfare). For the present study, the PRI definition will be based in a combination of physical and social issues, meaning the built asset itself and its social representation in terms of benefits for a certain community.

Public road infrastructure (PRI) plays an important role in the city. According Fann et al. (2003), an improved infrastructure helps create jobs and raise worker productivity. It saves time and human effort in transporting water, crops, wood, and other commodities. It also improves health and education by i.e. expanding access to schools, computers and lighting.

1.9.2.1. Indicators

Moser et al. (1996) affirm that PRI and the services that flow from it are crucial to the activities of households, producers and governments. As performance indicators vary significantly by transport mode and by focus (whether physical infrastructure or the services flowing from that infrastructure) highly specialized and carefully specified indicators are required. Among these indicators are:

- Roads, built (% of total roads), those surfaced with crushed stone (macadam) and hydrocarbon binder or bituminized agents, with concrete, pavement or with cobblestones, as a percentage of all the country's roads, measured in length.

- Physical condition, the quality of the construction materials used to build infrastructure in present time. In the case of roads, it refers to the surface quality provided. The quality ranges from simple compacted roads, to gravel, stone, macadam, concrete or pavement. This indicator is measured in terms of total length of paved roads as a percentage of the total length of roads in a given urban area (% of total roads).

- Infrastructure investment or the infrastructure costs as expenditures spent by different levels of government in order to provide infrastructure to the cities. These costs vary according to the quality of the infrastructure or the difficulty of construction. The indicator is sensitive to production costs, employment creation, access to markets and investment dependency, especially in transport. It measures the amount of monetary expend of public funds in the road construction, renovation, maintenance, etc. This indicator is a direct measure of the supply of infrastructure for land use's development.

1.9.3. Socio-Economic Classification

According to UN-HABITAT (1997), many indicators studies have concentrated on combining indicators to produce indices. However there is no a theory that indicates the exact quantity of indicators used to produce an index. (Openshaw, 1983) explains that, there is obviously no unique and 'correct' single set of variables to use. He also added that different people will use different variables for the same purpose, based upon experience of local conditions and/or intuition.

For this study, the SE classification is one of the most important components, which can be defined as an aggregate index with a range values from 0 to 1. In this classification, the 0 value represents the lower classification for all sampling units in the study area whilst the value 1 indicates the higher. In this sense, UN-HABITAT (1997) affirms that all methods for calculating such aggregate indices follow three important steps:

- Step 1, a set of variables are selected which are available for all cities under consideration, and which reflect aspects of the aggregate phenomenon to be studied. For this study, the SE classification (index) will use the indicators type of property, type of construction material (for houses), access to water, access to sewerage, access to electricity, population density and economically active population.

- Step 2, where the variables are transformed and normalised in some way. The transformations are intended to counter saturation effects or extreme values, while normalisation ensures that the scales of the different variables are similar. In this case, the variables will be normalised or standardized as a percentage along the range between the maximum (1) and minimum (0) values and intermediate values are spaced accordingly.

- Step 3, the different variables are standardized, weighted and they are added up to obtain an overall index.

The purpose of producing this overall index is to indicate and highlight some SE characteristics per sampling unit using variables and indicators taken from census data. In the same way, it is possible to

use this index to see the distribution of the range values on a final map and show the pattern of the socio economic classification in the study area.
The overall process to produce the SE classification as well as the analysis of the methodology used is explained in detail in the chapter 5 below.

1.10. Content of the thesis

Chapter 1 deals with a brief introduction, including a general outline of the topic, justification, problem statement and conceptual framework. The objectives, questions and hypotheses are stated with the assumption the limitations and shortcomings the proposal may have. In addition, the study design is described emphasizing the necessity of data and selection of an adequate study area. It also refers to the design of the study and the outline of the methodology to be used in the analysis. This chapter gives an initial idea about the goal and objectives of the research, the procedures for data collection and the proposed methodology for the analysis.

Chapter 2 shows the literature review, which includes theories behind the spatial relationship of PRI, its improvements and the important role that it plays in the development of urban poverty areas. This section will start from the cause, namely improvements in PRI, to establish the effects on UP in a cause-effect relationship. There is a critical analysis of theories for poverty alleviation in the literature reviewed. In addition, it includes general urban poverty policies applied in developing countries in the form of a case study.

Chapter 3 describes the background of the study area, showing the main aspects of the geography of Villa El Salvador and Villa Maria del Triunfo, in Lima, Peru. This description includes the administrative subdivisions, and the main geographical context of the study area. Some general aspects of the Lima in the early nineties, as well as the background and evolution of these urban poverty areas are includes in this review.

Chapter 4 explains the data collection process and the initial processing of the variables. Here, the gathering of the data, sources and procedures as well as the data quality are outlined. The chosen variables, namely SE variables and PRI variables and indicators are presented and described thoroughly. The chapter includes a full listing of all data collected, with quality assessment and sources. Data manipulation using appropriate techniques such as standardization is also described, for the corresponding variables.

Chapter 5 includes the research methodology and the analysis, which starts with the final processing of data variables to cope the study requirements. In addition, the application of GIS is exposed in the relevant aspects of the study. Final outputs are graphically presented and described, showing the achievements of spatial analysis and the correlation of the projected values in the study areas.

Chapter 6 discusses the results of the statistical analysis and contrasting them with indicators of the urban configuration in the study area. In addition, there is a comparison of results with relevant theory outlined in the literature review. Findings are remarked in a comparative analysis, showing the grade of relationship correspondence with the variables mentioned in Chapter 4.

Chapter 7 gives conclusions as well as all answers for the questions and hypotheses stated in the first chapter. General recommendations will be given for the central and local authority. In addition, there is an outline of other research topics inferred from this research.

1.11. Summary

The access to PRI has a direct influence over UP areas, where sustainable growth depends on the right access to all centres of attraction, productivity and economic interchange. The main indicators of PRI and UP were here introduced and their indicators defined. This section outlines the study in terms of the main issues behind its conception. It has given a brief introduction and stated the aim of the research. This is to find correlations between the PRI setting and a SE classification to be produced in the analysis out of secondary SE data. The main hypothesis refers to the SE classification related to the distance from sampling units to PRI, which develops several sub-objectives and referred questions to determine its validity in the study area. For such, the proposed methodology includes producing this classification and a network model to output the distances to PRI to be correlated together. The outline of all parts of the study was also presented in a brief description of each chapter of the current research.

2. Public road infrastructure and urban poverty

2.1. Introduction

The rapid urban growth has generated new population settlements which need an adequate public physical infrastructure to be provided by the current local authority. According to Wasike (2001) the development and maintenance of physical infrastructure are prerequisites for a continuous economic growth and poverty reduction. The infrastructure influences production costs, employment creation, access to markets and investments by stakeholders. In addition, access to public road infrastructure (PRI) is neuralgic for economic development, as it has a direct influence over the areas served by such. A sustainable growth of urban areas depends -among other factors- on the right access to all centres of attraction, productivity and economic interchange. PRI provides access to job opportunities and eases the implementation of basic facilities such as piped water supply, sewerage and electricity.

This paper focuses its analysis mainly on local roads with which the municipalities make investments which vary depending on the size of its jurisdiction. In this governmental level the municipality is responsible for providing roads to settlements, especially squatter areas. The poor areas with adequate PRI can have access to basic facilities that allow their development and economic. This concept is further analyzed in the present document. The real impacts caused by PRI on urban poverty will be explained later in a case study that analyzes the provision of infrastructure and the benefit to poor urban areas.

2.2. The definition and importance of public infrastructure

Depending on the economic, social and physical issues, there are hundreds of definitions of public infrastructure. According to The World Bank (1994) the combination the abovementioned issues can be classified as economic infrastructure. It includes services from public utilities such as electrical power, telecommunications, piped water supply, sanitation and sewerage, solid waste collection and disposal, and piped gas. Economic infrastructure also concerns municipal works, which include the provision of roads, drainages, dams, canals and transportation facilities (urban and inter urban transport, ports waterways, airports, etc.).

Another definition of public infrastructure is related with the physical and social aspects. Brussel (2004), explains that, in the traditional sense, infrastructure refers to physical components such as utility networks, ports bridges, railway and so on. According to this, infrastructure comes to be a combination of physical (all objects in the built environment) and social factors (services oriented to improve social welfare). Public infrastructure has a neuralgic and positive role in development. The World Bank (1994) finds a correlation that determines a point in percent increase in gross domestic product (percentage of GDP) for each point of percent increase in infrastructure stock across many countries (see Graphic 4). The more developed countries (Japan, Australia, Norway and Spain) have the highest levels in both indicators, whilst the less developed (Chad, Rwanda, Mali, Zambia and Bangladesh) have exactly the opposite.

As the countries develop and expand, the infrastructure must correspond with that growth increasing economic support and satisfying the crescent demand. In the same way, the stock of infrastructure (utilities, roads, communication, etc.) increase in relative way to such basic services as power, sanitation, transportation, etc. The different SE groups within a city have differentiated interest on a certain type of infrastructure development. Graphic 5 shows the corresponding infrastructure supply for different income levels. As seen there, the demand for infrastructure depends on the interest of the SE segment. Whilst the high-income groups are more interested in electrical power (to boost industry or simply power flat wide-screen televisions), more poor sectors focus on irrigation infrastructure to assure their subsistence by elementary agricultural means.

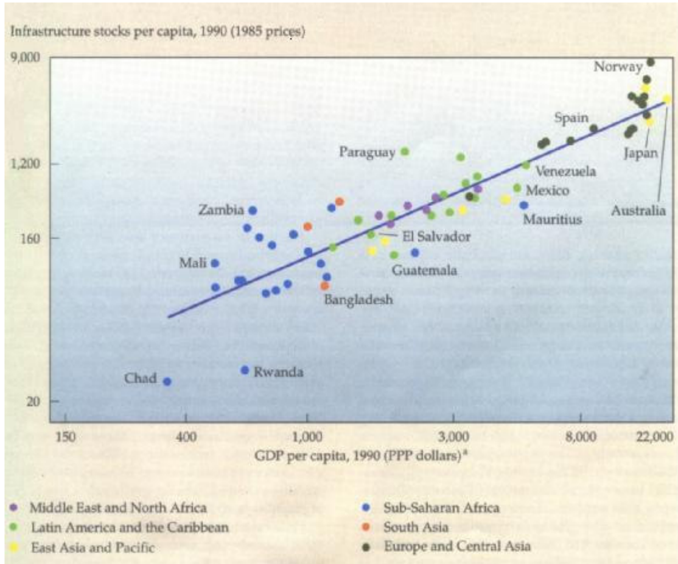
2.3. The concept of public road infrastructure

Roads in urban areas can be defined in three different approaches. First, in general terms road is a formed path or track suitable for use by all forms non guided vehicular transport. According to Wasike (2001) roads are divided into three levels: national or primary roads, departmental, provincial, regional or secondary roads and municipal, local and tertiary roads. The latter, applicable to the topic of this paper, seek to provide basic access to properties, opportunities and services in the whole city. The construction of a road implies factors such as main purpose (cargo transport, transit, private use, etc.) topography (the terrain configuration) and environment (exposure of the infrastructure to natural elements such as climate or soil condition). A road can be built in different stages, starting from the soil itself (a simple path) and lately passing to gravel, concrete or pavement. A good service quality depends of the adequate provision and maintenance of road infrastructure in terms of the physical condition of the road surface.

Second, as a geographical reality a road is a portion of land or space with defined boundaries. It has attributes such as width, length, inclination angle etc. and being spatially recognizable using projection coordinates. Kaufmann and Steudler (1998) define a land object as a piece of land in which homogeneous conditions exist within its outlines. This is a new concept given to define land parcel in the context of land administration and cadastre systems. However, if the road adopts this meaning we can see it in an abstract term. Therefore, a road can be manifested as rights to its adequate use and access defined by norms and laws. According to this, it will be useful if the roads could be registered as physical object in a road registration system for recognizing formalized rights of way. This information may be useful to know the total percentage of paved road measured in lengths in certain urban area managed by the authorities within their jurisdictions. In this way, they could better provide adequate roads, especially in squatter areas. The roads can be identified and the authorities in charge if they knew how many plots have access to them. With the road register they could cover the supply in a more effective way. In addition, with clear road boundaries the authorities may assure a right of way. This is important to formulate measures to avoid the invasion of the space reserved for future road construction (streets) or an inappropriate use of it by para-transit[1].

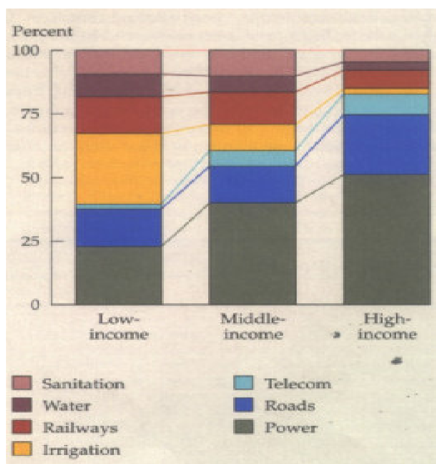
[1] Public transport service generally provided without an appropriate license. This sort of informal public transport is also characterized by inadequate transport units and poor service quality.

Graphic 4 : Correlation between infrastructure stocks and GDP per capita in countries worldwide



Source: The World Bank (1994)

Graphic 5 : Average percentage of infrastructure demand per income groups worldwide



Source: The World Bank (1994)

Third is the road as economic development. Here, it plays an important role in economic production, development and creation of benefits to the communities of different socio economic levels. The Asian Development Bank (2004) affirm that the infrastructure has a key role to play in streamlining product and factor market and extending opportunities. The Asian Development Bank added that urban infrastructure project in terms of financing, construction and maintenance is expected to directly improve the working and living condition of the urban poor. A combination of the geographical and economical points of view of roads can provide a more complex concept. In such terms, the physical condition of the road is an indicator that gives us information about the level of service provided to users and therefore their economical set up or potential. The combination of physical presence of a road serving a certain area and the economic setup of the population served is a matter of extreme importance in urban planning.

An adequate public road infrastructure (PRI) copes with population growth and helps improving environmental conditions and for long-term sustainable development. According to Fann et al. (2003) an improved PRI helps creating jobs and raising worker's productivity. It saves time and human effort in transporting water, crops, wood, and other commodities. It also improves health and education by expanding access to schools, computers and lighting. PRI is measurable using indicators such as paved roads (percentage of total road length) accounted as a total road network measured in length by kilometres (The World Bank (2003)). PRI provide access to areas of interest such as work and services as well as it eases economic development by fostering trade and goods exchange. Therefore, an adequate PRI helps the economic development by making the trade more dynamic, favouring access to services essential for community development (water, sewerage, electricity) and generating connections to employment centres. Inside the transport sector, (urban and inter-urban transport) the construction of adequate is essential to maintain the connection between areas. However, if the transport service is over dimensioned or operating under inadequate conditions, the impact of the fossil fuel emissions and noise in the environment would be greater.

Cotton and Franceys (1994) state that each infrastructure sector has a range of technology options which offer different levels of services. Consequently, the authorities can opt for selecting the cheaper technology in order to reduce costs. In the case of PRI there are a variety of ways in which roads can be provided to poor settlements. An option is the access to profiled earth road or path (it is adequate in certain conditions). The other is the access to surfaced road, ranging from gravelling to bituminous macadam or concrete pavements. Cotton clearly states that the first alternative is cheaper and it will be appropriate for most low-income settlements. However, this level of infrastructure provides an extremely limited gap for future economic growth. It is illogic to start a settlement from scratch and ask for elaborated infrastructure though such provision must be assured for the future. The best option is to carry out from time to time a deeper cost analysis of infrastructure to understand the benefit of implementing a new level of infrastructure (for instance, paving an earth road) as for the authorities as for dwellers.

The cost analysis involves the concept of life cycle cost related to the facility (road infrastructure service). Cotton and Franceys (1994) add that the infrastructure provided have to include or define an indicator called total annual cost per household (TACH) because not only it can reflect a sustainable provision of infrastructure for low income communities but also it can be represent an annual cost imputable to each plot served to recover the life cycle cost. Graphic 6 depicts TACH shown as a function of plot size. The high service costs (option 1) are almost twice as much as the lower service level (option 2). It also gives decision makers the opportunity of comparing the cost of different technologies (option 1 and 2) and services levels within household incomes. This analysis is important when is necessary to take decisions about savings in infrastructure cost and the desired level of service.

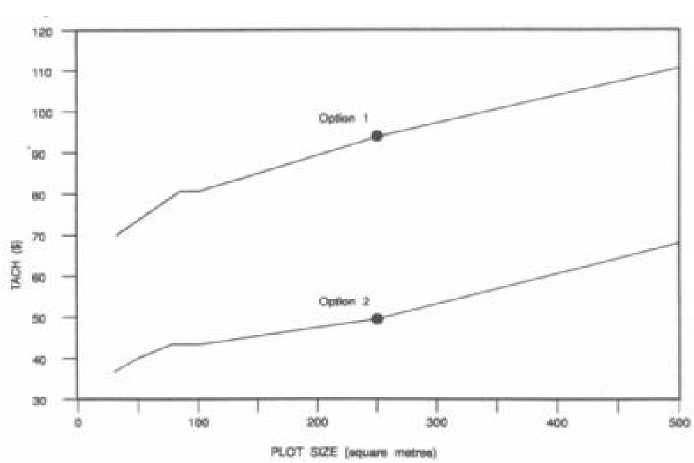
Unfortunately, there are poor settlements not properly integrated to the urban tissue because of the lack of adequate PRI. Squatter areas are mainly born in inaccessible places where there are neither services nor the basic facilities to provide it, in terms of road infrastructure. This lack of accessibility is also determined by the geography of the poor dwellings and squatter areas, which are mostly implemented over hilly terrain, not suitable for housing. Therefore, providing such basic roads is most of the times extremely costly and unachievable by the resources of the settlement itself. An adequate access to urban basic services may provide these poor dwellings a measurable improvement in their living conditions and thus a reduction of urban poverty. In terms of PRI suitability means good quality of service, and must be provided by authorities at the national or local level.

2.4. The benefits of physical infrastructure to poor areas

In the economic aspect, infrastructure always promotes benefits to population regardless of its SE level. In a more disaggregated level of analysis, Bond (1999) affirms that there are two different types of benefit of infrastructure for low-income people: direct and indirect economic benefits. The former include construction jobs, improvement in work productivity and the growth of small enterprises. The latter considers more time and resources for women, dramatic environmental benefits, public health benefits (provision of quality infrastructure, sufficient to enhance rather than endanger health) and the desegregation of urban society with respect to enhanced employment, educational and cultural opportunities. In both cases, PRI is direct related with economic benefit. For instance, the provision of drinking water to squatter areas depend on trucks that carry it there, which also implies that poor dwellers have to pay more for it. By improving the road condition, trucks could go further in the areas and minimize the cost of the water by facilitating its transport, which represent a direct benefit. Studies related to water and electricity sector, indicate that poor people that lack access to formal services are willing to pay extremely high prices for poor quality services. This is the paradox of urban transport infrastructure provision from different sectors (Brussel, 2004).

To provide infrastructure in terms on good quality of service is a challenge, for such it is important to recognize and define the most essential components in order to achieve this goal. These components are the public sector -represented by local/state authorities- and economic resources available. The public sector is directly involved in the performance of public infrastructure provision which include the public ownership and operation. It is the most common option that is practiced in many developing countries. According to Wu (1999), good performance stands out by a number of characteristics: efficiency of operations; adequate maintenance of existing capacity; financial efficiency good record of cost recovery; and responsiveness to user demand. It is important to add that the delivery of services is often better carried out when public organization are free from political interference and also they receive managerial and financial autonomy.

Graphic 6 : Correlation between total annual cost of infrastructure per household and plot size



Source: Cotton and Franceys (1994)

Different kinds of infrastructure have in common a set of activities, each with somewhat distinct institutional requirements. The provision of urban infrastructure is a process that involves financing, construction and maintenance activities. Each activity represents a high cost and with the tight-fisted budget of the public sector many governments in development countries have begun handing out responsibilities to the private sector. An option will be minimize investment cost on infrastructure using other alternative technology.

For instance, in Shanghai, China the predominant form to provide urban infrastructure is adopting public ownership and operation performance. In this way, since 1990, most urban residents have access to tap water; gas and public transport (see Table 1). In addition, the per capita paved urban road surface nearly tripled between 1990 and 1996, from 2.7 to 7.6 square meters. Such significant progress is the result of responsible operation of urban construction authorities and also it is attributable to a number of state working units as well as public agencies and other state organizations (Wu, 1999).

However, Cotton and Franceys (1994) affirm that the current modes of the provision of physical infrastructure to squatter areas are unlikely to meet the ever-increasing demand for services. This is because the levels of service provision to squatter areas (low income people) follow conventional method that are modelled on the high levels of services (costly technology) used in developed countries. Moreover, those implementing the infrastructure have a little knowledge or interest in communities living in these squatter areas. Therefore, it is necessary a radical rethinking to reduce cost for implementation of infrastructure and use of more appropriate technologies required. These rethinking can be supported by new policies oriented towards a participatory planning where the word of dwellers is also taken into consideration.

Table 2 : 1990 and 1996 access to urban infrastructure in Shanghai, China

	1990		1996	
Indicator	All cities	Shanghai	All cities	Shanghai
Paved road per capita (sq. m.)	2.7	2.3	7.6	4.5
Public transport (vehicles/10000 residents)	1.8	8.5	7.3	15.7
Access to tap water (percentage)	81	100	94.9	100
Per capita annual domestic water consumption (tons)	63.7	81.9	75.9	104.6
Access to gas (percentage)	8.1	46.1	73.2	91.5
Open space per capita (sq. m.)	--	--	5.3	1.9

Source: Wu (1999)

2.5. PRI for urban poverty reduction in Gujarat State, India

Despite the importance of the PRI provision in order to reduce the poverty in rural as well as in urban areas, there is scarce research in the latter. Most of the case studies found for this paper deal with highway implementation in remote rural areas or local roads in rural communities. However, a study made by The Asian Development Bank (2004) analyzed three case studies, one of which is developed in an urban area. The following paragraphs describe the study and discuss it thoroughly. Furthermore, the important role that private sector play in investment on infrastructure will be assessed.

The case study of Gujarat State, India is an example that evidences changes in urban poverty level if the households have access to infrastructure as road and energy through improved PRI. In addition, it shows how the households have been benefited from the investment on transport and energy. This study used intervention variables such as access to electricity, per capita transport and energy expenditure and road improvements/distance to. The situational factors are determined by social groups (castes), religion, education, access to credit and others. The analysis aims to reflect the effect of these situational factors that might vary across the districts.

Two main objectives are outlined for this case study. The first is to assess the impact of interventions in the transport (roads) and energy infrastructure at community, household and individual level. The second aims to identify the direct and indirect mechanisms through which this impact on poverty is produced. Among other techniques The Asian Development Bank (2004) conducted an econometric analysis, dividing households into four sub samples. Since all households have received access road improvement and electrification, the sub samples were based on less or more 0.5 Km from the road and connected or not to electricity. The different classes of variables (indicators) and the correlations found in the study are shown in Table 3.

In terms of PRI and poverty alleviation, the results show in Table 3 that road access was significant in a negative sense to the poverty status only in the Panchmahal district. This means that the higher access indicator generated a lower poverty measure. Distance from home to the improved road also supports no significant relationship to poverty status. These findings were clarified using a Probit[2] model to predict the probability of a household to become poorer.

Table 3 : Correlations of poverty levels in 4 districts of Gujarat State, India

Variable	Variable Signs and Significance			
	Jamnagar	Bharuch	Panchmahal	Kuchchh
Intervention variables				
Access to electricity			Negative**	Negative**
Access to improved road			Negative*	
Distance to improved road				
Per capita energy expenditure	Positive**	Positive**	Positive**	Positive**
Per capita transport expenditure	Negative*	Positive**	Positive**	Positive**
Situational factors				
Social group (caste)				
Religion				
Education (male)				Negative**
Education (female)				
Dependency ratio		Negative**	Negative**	Negative**
Per capita landholding	Positive**	Positive**		Positive*
Share of non-farm income			Positive**	Positive**
Access to credit				

*Results for relationships that were significant at $p < 0.10$
**Results for relationships that were significant at $p < 0.05$

Source: The Asian Development Bank (2004)

[2] Based on statistical probability by binary variables, also called probabilistic model.

Access to PRI by households has a positive effect on poverty reduction only for non-electrified houses. The level of poverty is higher (49%) in households that have both road access and energy than in households with electricity alone (46%). It establishes that poverty levels are higher in households closer to the roads even with no-electrified households in three of the four selected districts. These findings suggest that road access itself overcome the poverty levels in India. However, they show a tendency of poorer families to locate to closer to the improved roads in search of opportunities.

Finally, it is important to highlight the role of private investors in the development of infrastructure. In the past, infrastructure investments in India were mainly made by public sector, including the central and state governments because the lack of an appropriate regulatory framework to attract the participation of the private sector. This participation was crucial for the alleviation of poverty in urban areas through major investments formerly possible only by public funds. Gujarat has managed to reduce urban poverty since the past two and a half decades. From 1973-74 till 1999-2000, poverty ratios have decreased from 52.6% to 15.6% in urban areas. However, this decrease in urban poverty can not be solely explained through PRI provision.

2.6. Final remarks

Public Infrastructure has a complex meaning. It includes services from public utilities, municipal works and transport sector. Moreover the infrastructure play a neuralgic role in urban areas, especially in poor areas where it can provide a measure of development This is done in order to maintain the equilibrium between economic sectors, aiming to achieve the sustainable development.

This role in the development of the countries is only possible if public authorities provide the adequate infrastructure. To achieve this is necessary to take into account an important factor called infrastructure investment. This factor includes components such as the participation of public sector (state authorities) as well as the private sector. In addition, it is important to mention the use of more appropriate technologies to minimize the cost on infrastructure using alternative (and more economic) technologies to cover the infrastructure demand, especially in poor areas.

This is particularly important as there are several PRI levels. Choosing the appropriate require modern approaches, one of which deals with the registration of public roads in a cadastral system. This may help getting to know the exact level of demand as well as providing the adequate legal framework for PRI provision -such as right of ways- avoiding the illegal occupancy of reserved land. The provision of public infrastructure may help reducing the poverty; this is only possible if the service quality is good and adequate for development of urban areas though.

Finally, PRI may support the economic development of poor areas. Roads provide the vital access to services, employment and ease the implementation of basic utilities such as water, gas or electricity. This access is said to be crucial for sustainable development of new settlements, generally poor squatter areas within the cities. Therefore, PRI contributes to tackle urban poverty in those areas, being a decisive factor without such poor dwellings may remain the same. This conclusion is supported by the findings in the case study of Gujarat State, India, described in this paper. However, it is important to notice that more studies of public infrastructure role in reducing poverty must be carried out in the urban context. This may assure the consistency of the findings in the single case study for urban areas found.

2.7. Outline of the theoretical framework of the study

This section has demonstrated the importance of PRI in tackling poverty. The main concept of this section that provides a base for the study deals with the combination of the geographical and economical points of view of roads. As said before, the physical condition of the road is an indicator that gives us information about the level of service provided to users and therefore their SE set up. The combination of physical presence of a road serving a certain area and the economic setup of the population served is a matter of extreme importance in urban planning and the present will cover this issue.

As stated, more research of the incidence of PRI in poverty in urban contexts is needed. Therefore, the present study is developed in one of such areas. The southern part of Lima City, a modern example of a poverty area raised from a former human settlement was chosen. The aim is to get to know the level of relationship between urban poverty areas and road infrastructure. This case study may provide a better view of the existent relationship between PRI and the socio-economic layout of urban poverty areas.

2.8. Summary

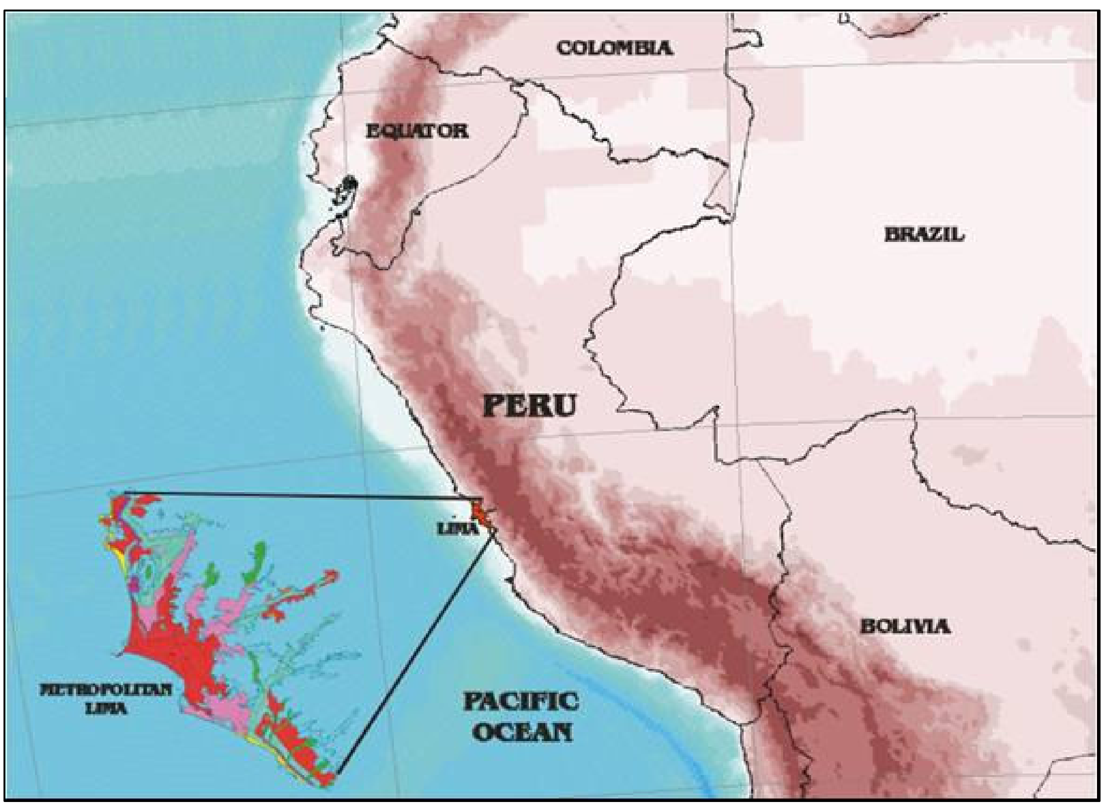
The socio economic benefits caused by the provision of road infrastructure in poor areas may help reduce poverty levels. This benefit is measured using SE indicators that are directly affected by the new services provided by such infrastructure. The present section describes the relationship between public road infrastructure and the reduction of poverty in urban areas. It defines public infrastructure based on a combination of social, geographical and economic issues. This means the built asset itself and its social representation in terms of benefits for certain low income areas, highlighting its neuralgic role in urban development. Therefore, roads may also be recognized as geographical objects, able to be registered in a cadastre. This provides means of infrastructure account for the authorities in charge. As presented in a case study in Gujarat State, India, the provision of road infrastructure may help reducing the urban poverty. It is indeed possible if the quality of the provided infrastructure is enough and adequate for development of urban areas. In order to achieve this is necessary to take into account the level of investment in infrastructure in terms of participation of public and private sectors and the use of appropriate technologies. The theoretical framework of the current study emerges out of this review. It refers to the physical condition of the PRI as an indicator for the level of service provided to users and therefore their socio-economical set up.

3. The Geography of Southern Lima

3.1. Location and administrative layout

The Republic of Peru is located in the central western region of South America. To the North, Peru borders with Ecuador and Colombia; to the South with Chile and to the East with Brazil and Bolivia, being the Pacific Ocean to the West (see Graphic 7 below). Peru is the third biggest country of South America with a territorial extension of 1,285,215 Km^2. It is divided in 24 departments and the constitutional province of Callao, having the department of Lima the biggest rate of population densities. Lima is divided into 10 provinces where the central province (with the same name) is located and where the metropolitan area is placed. According to INEI (2001), the capital city has a territorial extension of 2 811.65 Km^2 and is divided into 49 districts. Out of them, Villa El Salvador (VES) and Villa Maria del Triunfo (VMT) districts are located some 23 Km. to the South of Lima's downtown. VES has an area of 35.46 Km^2 whilst VMT has 70.57 Km^2. Both districts were parted of formation process of human settlements, both representing an example of urban growth based on fringe areas.

Graphic 7 : Location of the study area: Metropolitan Lima, capital city of the Republic of Peru



The urban space of the study area was born in May 1971 as consequence of a settlement promoted by the military government of General Velasco. In order to relocate population that had previously taken other land in the Eastern Lima, known as Pamplona, the area called 'Tablada de Lurin' in VMT was given to these informal settlers. The area was planned and the population was given a plot per family and organized into the modular structures so-called 'residential groups'. These groups were formed by 16 blocks and a total of 384 plots (See 0 below). All area needed for infrastructure provision, as well as blocks for parks and services were previewed. Starting from this pattern, the urban area of the current VMT and VES started to grow, covering all the suitable land, including the hillsides as high as possible.

Table 4 : Modular Structure of the Urban Zone of Villa El Salvador

Cadastral Division	Units
Plot	140 m^2
Block	24 plots
Residential Group	16 blocks
Sector*	22 and/or 23 residential groups

* complemented (21A, 22A, etc.)
Source: Municipalility of Villa El Salvador - MVES

3.2. Population

VES is the eight most populated district of Lima Metropolitan and most of the people there came from rural areas. According to the National Institute of Statistics and Informatics (INEI, 2001), VES had a projected population for 2005 of some 351 979 inhabitants (at a 4.8% increase per year and based on a measured population of around 254 641 in 1993 and 322 826 in 2001). This district has a population density of around 91 inhabit/Ha. approximately. VMT has around 263 554 inhabitants (half settlers are less than 18 years old) and its population grows at a rough 3.2% per year. This district has a population density of around 46.63 inhabit. /Ha. Because of the lack of commercial and industrial areas, most of the population has to travel to others districts for working purposes. The main activity inside this district is the commercialization of wood and the transport service.

3.3. Terrain configuration

Both districts has the same urban layout, it has grouped by different urban human settlement categories such as urbanizations, young town, housing associations, housing organizations and others, being the where most of the people live. VES and VMT are in the middle of the valley limited by the Rimac and Lurin rivers, being characterized by a lack of direct provision of water from both basins.

This predominantly dry climate is surrounded to the East by hills of the early Andes Mountains and the Pacific Ocean to the West. It also has an irregular topography with small dunes and sandy terrain. The main area is relatively flat, though with the presence of small terrain elevations in all its extension.

3.4. Urban Evolution

According to the Urban Program DESCO (2003), Lima city has had a predominant growth based on land invasions since the 50´s. As a result of such phenomenon, numerous poor human settlements appeared, being VES and VMT two of the most recurred dwellings for the poor migrants in recent decades.

According to the municipality of VMT, countless groups of families (called associations) organized invasions to take over sandy lands located some 20 km of the Lima-Atocongo highway. As a consequence of popular pressure, the government legally recognize the invasion and gave those areas for the birth of the association Pro-vivienda El Triunfo in 1952. This association changed its name for the Urbanization Villa Maria del Triunfo, which became the Villa Maria del Triunfo (VMT) district by a municipal ruling in 1961. By 1981, this district had 19 human settlements categorized as "Pueblos Jóvenes" (Young towns, (INEI, 1981)). One of the most important human settlements inside VMT was Villa el Salvador, which formally became a district in 1983.

Photo 1 : Terrain configuration of the urban area formed by VES and VMT

Retrieved from Google, July 2005

Photo 2 The hilly topography of the study area and the evident low density

One consequence of the invasion process was countless settlements over the most inaccessible areas, with people living in precarious huts (see Photo 3 below). There was a total lack of basic infrastructure, namely roads, water ands electricity or access to fossil fuels. This situation was maintained for several years, being the survival of those communities up to the dwellers themselves. For such purpose and sometimes long before the invasion, these inhabitants were self-organized to cope with the lack of authority and resources.

In this scenario, these urban poverty areas were in need of the necessary access to services and opportunities. Goods were to be purchased far away of the living areas and travelling implied long walks downhill and uphill over inexistent sidewalks. As poor dwellings, the motorization rate was almost null, thus roads were mainly unnecessary for private transport. There were no means of reaching those areas by public transport and lines in the closest roads were several minutes of walking away. Most of the inhabitants there experienced extremely long travel times (around 4 hours a day) to their daily destinies.

According to De Soto (1987), informal associations play a group of functions in order to protect and increase the value of the poor dwellings. One important function is to assign responsibilities among the settlers, organizing committees to obtain the provision of basic services and infrastructure. The first infrastructure there was provided by the same dwellers with some help of national authorities or private companies. The principle of self-provision, called ´auto-gestión´ (done for ourselves by ourselves) has its most notorious example in Villa El Salvador. The organized people started providing basic services such as sewers and unpaved roads. The latter were the first means of communication with the current road network and implied great slopes, sometimes unsuitable for heavy trucks such as those carrying water, fuel or construction materials. This situation lasted for decades and the capability of development and poverty alleviation was thus extremely limited.

Photo 3 : Land invaders in igloo-style huts at Villa El Salvador in the early 80´s

3.5. Summary

This chapter presents the study area and provides a brief description of its main characteristics as well as its urban evolution. It is composed by the districts of Villa El Salvador and Villa Maria del Triunfo is located in the southern part of Lima, Peru. Both districts have a similar urban layout, geographical and population characteristics. It is defined as an urban poverty area that started their formation as human settlements, consequence of an invasion process. For several years VES belonged to the VMT district, though later it became a district itself. Little attention has been given to both them by the government. The lack of adequate and enough urban infrastructure, particularly roads is evident. Consequently, both districts have constantly struggled to improve the level of poverty present there.

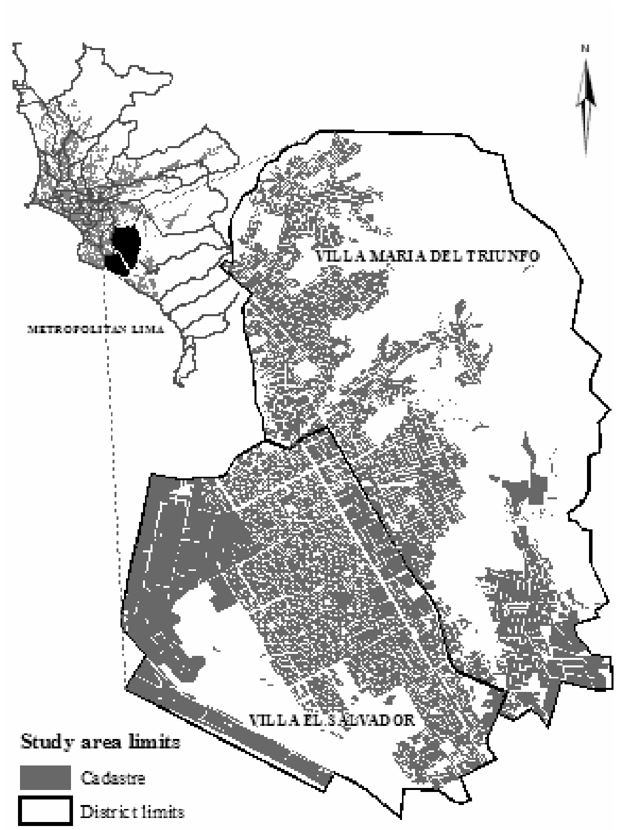
4. Data Collection and Processing

4.1. The Fieldwork

4.1.1. Selected Location

The study was planned to be conducted in the poor areas of the south part of Lima, Peru. Lima City is divided into 49 districts from which two districts were selected beforehand (See Graphic 8 below) based on their SE characteristics. These characteristics define them as predominantly urban poverty areas. Therefore, the urban areas of Villa El Salvador (VES) and Villa Maria del Triunfo (VMT) are to be defined as the study area.

Graphic 8 : Study area location inside Metropolitan Lima



4.1.2. Objectives

The fieldwork was carried out from September 21st until October 21st. Its main purpose was to gather secondary data related variables such as housing; population and road infrastructure for past years (see Table 5 below). In addition, complementary data related to the public road infrastructure, district's boundaries, land uses, transport network and cadastre maps were to be collected from local and

national institutions. These data will be used to perform the analysis, proving the main objective of the thesis and validate the main hypothesis.
Complementarily, primary data about past and current status of the urban areas and interviews with identified key informants was obtained. The observation of the study area was done in 3 weeks, including several study trips to each of the two districts to observe the urban configuration of the study area and verify the secondary data related to roads and the physical condition of housing areas.

The methodology for verifying the data was a sequence of tours in some urban areas selected in advance. Before going to the field, a plan defined the routes per day, starting each trip from the oldest urban areas. This activity was done sector by sector with the assistance of a transport expert, an urban planner and people who live in both districts.

4.1.3. Selected sampling units

The study area represents a vast population that has a certain degree of homogeneity in terms of SE characteristics. The objective of the data collection, defined in 4.1.2 above, implies the location of the sources, trying to get data in the more detailed level possible. As census or administrative zones in the study area are quite big (and therefore, holding aggregate data) they were discharged as possible sampling units. Consequently, the ideal was to find the data in the plot or block level. As censuses are conducted at the family or home level, this data had the priority as being suitable for the study. Finally, the data turned to be more comprehensive per blocks, as not only statistical but also geographical data could be found at that level. Therefore, the sampling unit for this analysis are the blocks aggregating data from housing units inside them. The study area has 5055 block units, from which 2826 are located in VMT and 2229 in VES.

4.1.4. Detail of activities conducted

The first approach to the data sources and key informants included determining variables, indicators and data availability for each one. In addition, interview appointments were set up at that point, trying to follow the time schedule as much as possible. Contacts with technical and professional staff with knowledge in road infrastructure (public works), transport network and urban planning were made. Four interviews were carried out at the Municipality of VES, with Mr. Jaime Zea Usca, the Mayor, Mr. Alex Cordoba, Director of Public Work Department, Mr. Hugo Soto, Director of Planning and Budget Department and Mr. Rafael Zamora. In addition, Mr. Michel Azcueta former Mayor of VES was interviewed. Their opinions were useful to find other factors that influenced the SE indicators of the human settlements in the study area. The information was also essential for improving the scope of the study. The contact with these key informants helped the understanding of the development and formation process of the study area before it formally became districts in Lima. In addition, they were very helpful in finding out the relationship between the provisions of PRI and the SE layout. This relationship will be spatially analyzed the using the data of socio- economic variables and the physical condition of roads that were collected during the fieldwork.

4.1.5. Data Collection

The data collection comprised the identification, location and request of data to the institution that holds it. It is important to mention that the 1984 and 1990 aerial photographs were provided as a personal favour from a private source. The cadastral, land-use and road network data was provided with a compromise of not providing it to a third party. The list of institutions that were main sources for data were:

- National Institute of Statistic and Informatics (INEI).
- Potable Water and Drainage Service of Lima (SEDAPAL).
- Autonomous Authority of the Electric Mass Transport System of Lima and Callao – Urban Train (AATE).
- Centre of Development and Promotion Studies (DESCO).
- Municipality of Villa El Salvador (MVES).

The data collection included SE indicators which were divided into two categories, as to say housing and population. These are reviewed in the following paragraphs and there is also a summary in Table 5 below.

4.1.5.1. Housing variables

This category includes all variables related to the houses and families that inhabited them in the study area. As the data collected is aggregated at block level, it represents a sum of all the housing units in that block. These variables comprise information about the house property (i.e. own, rented, with mortgage), the construction material divided per construction element (walls, roofs and floors) such as bricks, concrete, wood and the services they have access to (drinking water, drainage and electricity). All housing variables were collected from the INEI database. They were taken from the IX Population National and IV Housing Census 1993 database (in hardcopies).

The indicators that were collected under this item are:

- House Property, including the number of houses per block (the sampling unit) with a given types of ownership (Own house, rented, own with mortgage and invaded).
- House Material, showing the number of houses per block with a construction material for walls (brick, wood, estera, quincha or adobe), roofs (concrete, tiles, wood, estera, cane or palm leaves/straw) and floors (cement, tiles, polished wood, unpolished wood and bare soil).
- Access to water, which provides the number of houses per block with a certain type of water provision (from pipe inside/outside, from container, from dwell, from irrigation ditch and from a cistern truck).
- Access to sewerage, which indicate the number of houses per block with certain the type of sewerage (to pipe inside/outside, to dwell, to irrigation ditch and without).
- Access to electricity, that shows the number of houses per block with electricity connection (whether present or not).

4.1.5.2. Population variable

Data from the inhabitants of the study area was collected as well. This variable is related with the total amount of persons living in each block (population) and it include the total number of economically active persons (PEA, in Spanish). The PEA is related with the persons who participate in the labour force.

Population figures were collected in hardcopies from both the 1993 National Population and Housing Census performed by INEI and from the AATE database. The objective is to have a clearance about the data quality by taken it from two different sources. Finally, the PEA values were taken from AATE database, which was taken from 1993 INEI Census.

The indicators that were available for the study are:

- Population, total number of permanent inhabitants, aggregated per block.
- Economically Active Population (PEA), a parameter used by the Peruvian Government to identify the individuals that are in the capacity of generating some sort of income. This does not necessarily mean people with a profession but all kind of economic activities, whether formal or informal. Therefore, in this item there can be children from 5 years an up as they are able to perform or help to perform some sort of activity (predominantly informal, such as street vendors, car washers or shoe polishers). For this indicator there is the number of occupied and unoccupied individuals as well as the total PEA per sampling unit.

4.1.5.3. Road variables

The physical condition of the roads is to be taken from the 1984 and 1990 aerial photographs. These were scanned with high-resolution parameters, good enough to identify paved or unpaved roads. A transport network with attributes was collected from the AATE database. This was formed by paved roads in the study area and included road attributes such as lanes, speed, length, type (i.e. highway, local, connector, etc.). This variable is included as a GIS layers (paved roads) with road attributes such as length, width, speed, etc.

4.1.5.4. Urban Planning variables

The fieldwork also comprised collecting other extra information related to Municipal regulations about infrastructure implementation, cadastre maps and land use maps. All this data was collected at block level and included information related to each block such as area, perimeter, total number of housing units, cadastral code and predominant land use. Complementarily the limits of the districts that comprise the study area were collected in appropriate maps.

Three similar cadastres were taken from INEI and SEDAPAL, for 1993 and 2005 respectively and from AATE. The latter is compiled in yearly cadastre maps at block level for the years 1997, 1998, 2000, 2001, 2002 and 2004 in GIS files. There are also land use maps for the years 1997, 1998, 2000, 2001, 2002 and 2004. This information is geo-referenced in several projections and compiled in GIS maps. The primary key is a code IDMANZANA (or IDBLOCK in English) that is composed by the district code, the zone code and the block number, adding up 14 to 15 digits (some extensions of blocks have the same code with a letter at the end). This unique field is the identificator for all sampling units.

Documents called Municipal regulations that are belongs to the municipality of Villa El Salvador, were collected for the years 1995, 1996, 1997, 1998, 1999 and 2000 (in hardcopies). They include the information about the time of paved road implementation in the study area. All the data has been tested in Lima to assure consistency and data quality with good results.

4.2. Data Processing

4.2.1. Format conversions

The collected data came in different formats and some was to be used in special software only available at the institution holding it. That is the case of the AATE data which was originally in TransCAD software. The same institution transformed all the data into ESRI ArcGIS shapefiles out of the original format during fieldwork data collection. These shapefiles were successfully tested in the application, though they had to be projected and geo-referenced.

Other case refers to the 1993 Census data from INEI. The data was provided in Microsoft Excel format and included a lot of unnecessary information such as those used for presentation purposes. The tables had to be cleaned out of all that irrelevant information and exported to Data Base formats to be used inside GIS software. The result tables were double checked to assure data integrity and quality as well as to avoid missing values.

Table 5 : Detail of data collected per variable

	Variable/issue	Data Information/Details	Format	Data Source
Housing	Access to water and sewerage	1993 cadastre map and attribute table (block level)	GIS Files, database files, Hardcopies	SEDAPAL INEI
	Access to electricity	Spreadsheet table (block level)	Database files, hardcopies	The IX National Population and IV Housing Census 1993 (INEI)
	Housing material	Spreadsheet table (block level)	Database files, hardcopies	The IX National Population and IV Housing Census 1993 (INEI)
	Property type	Spreadsheet table (block level)	Database files, hardcopies	The IX National Population and IV Housing Census 1993 (INEI)
Population	Population	Spreadsheet table (block level)	Database files, hardcopies	INEI AATE (from 1993 INEI Census)
	Economically Active Population (PEA)	Spreadsheet table (block level)	Database files, hardcopies	AATE (from 1993 INEI Census)
Road	Road surface	1984 and 1990 aerial photos, 2002 IKONOS image	Scanned files and digital Image	Private photographic archive SEDAPAL
	Road network	1993 road network layer with attribute table	GIS files, database files	AATE
Urban Planning	Land use	1993 land use layer with attribute table	GIS files, database files	AATE
	District limits	1993 district's limits layer with attribute table	GIS files, database files	AATE
	Cadastre	1993 cadaster layer with attribute table	GIS files, database files	AATE
Topography	Contour levels	Contour lines layer with attribute table	GIS files, database files	AATE
	Elevation points	2005 drainage top levels map	CAD file	SEDAPAL

4.2.2. Setting up the variables

In order to have the data ready to perform the analysis, some transformations and complementary processing were needed. These processes are described in detail as they will have incidence in the quality of the posterior analysis. The processing applied to each group variable in described separately.

4.2.2.1. Housing variables

As explained in 4.1.5.1, the data related to housing variables came in total amounts per sampling unit or block. The value is the number of housing units for the indicator in that particular block (i.e. 30 houses -out of 100 total in the block- have brick walls). These values are amounts that can not be understood unless they represent a homogeneous index applied to all sampling units and able to be compared and analyzed. A given number of units for a determined indicator does not tell much about the block unless it is confronted with the total units of that block. For instance, if one block has 10 or 40 or 60 units of brick-walled houses, it is not relevant unless we contrast those values against the total number of units per block. Therefore, these indicators were converted to percentages of the total number of houses in the blocks. This is the result of dividing the indicator value by the total in the sampling unit (in the previous example, 30 divided by 1000, meaning 30% of houses in the block have brick walls).

4.2.2.2. Population variables

Data of population variables, such as PEA and total number of inhabitants as described in 4.1.5.2 is also in figures per sampling unit. In the case of PEA, these figures were converted following the same process as described in the previous point. Therefore, raw data was divided by the total number of dwellers in the block. In the case of the block's population itself, the density was the natural option, got by dividing the figures by each sampling unit area in hectares.

4.2.2.3. Road variables

Included in this variable are the aerial photographs and images as described in 4.1.5.3. Except the 2002 IKONOS image, the 1983 and 1990 photographs were unreferenced. Using the cadastre of INEI (which contained the most useful variables out of the 1993 Census and already in GIS format) all photographs for each year was geo-referenced. The results were quite acceptable despite the lack of a restitution process of the scanned photographs. Nevertheless, the sole purpose of these geo-referencing process was to determine the road surface for the two time periods, namely 1984 and 1990.

The road network layer as provided by AATE covered the full area of Metropolitan Lima and was geo-referenced using a different projection than the GIS data provided by SEDAPAL or INEI. As the latter has the most of the variables needed for the study, it was chosen as the main source. Therefore, the road data from AATE had to be projected and geo-referenced to the main data using ESRI ArcGIS procedures. These included the use of Define Projection and Project commands to standardize all to Peru Central Zone projection included in the mentioned software. The result was quite acceptable and the roads in the study area were given the information of the aerial photographs respect to their surface (whether paved or not). In addition, all the unpaved roads were digitized and added to the layer in order to finish setting up a complete road network.

4.3. Summary

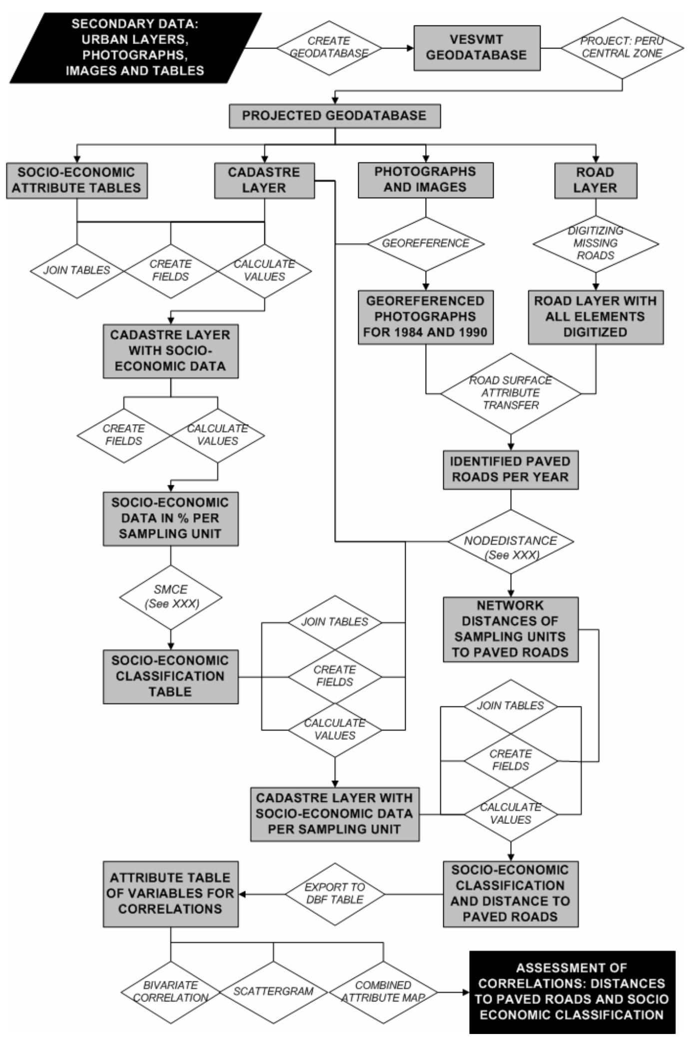
The fieldwork allowed collecting several SE variables and indicators that are going to be used in the final analysis. The data was tested and checked to assure usefulness and consistency. A full description of all the variables is included, emphasizing the level of detail of the collected data. Summary tables were also produced for all variables, indicating the type of data and the source that gave it. Finally, an initial processing of the data was performed. This has the objective of putting all data in the condition of being used in the final analysis. Format conversions, geo-referencing and projection were performed in most cases. The final data quality was assured and reliable data was provided for the final methodology application.

5. Analysis

5.1. Methology Application

As outlined in 1.7, the research methodology has three main parts. The first is to produce a SE classification based on the SE indicators obtained in the fieldwork. This SE output will later become the dependent variable. The second part deals with the elaboration of a network model in order to obtain distances from the sampling units to the paved roads. Thus, distance will become the independent variable against which the dependent will be analyzed. This last part will deal with the assessment of correlations in statistical procedures. The overall process of the methodology as seen in Graphic 9 below, is explained in the following paragraphs.

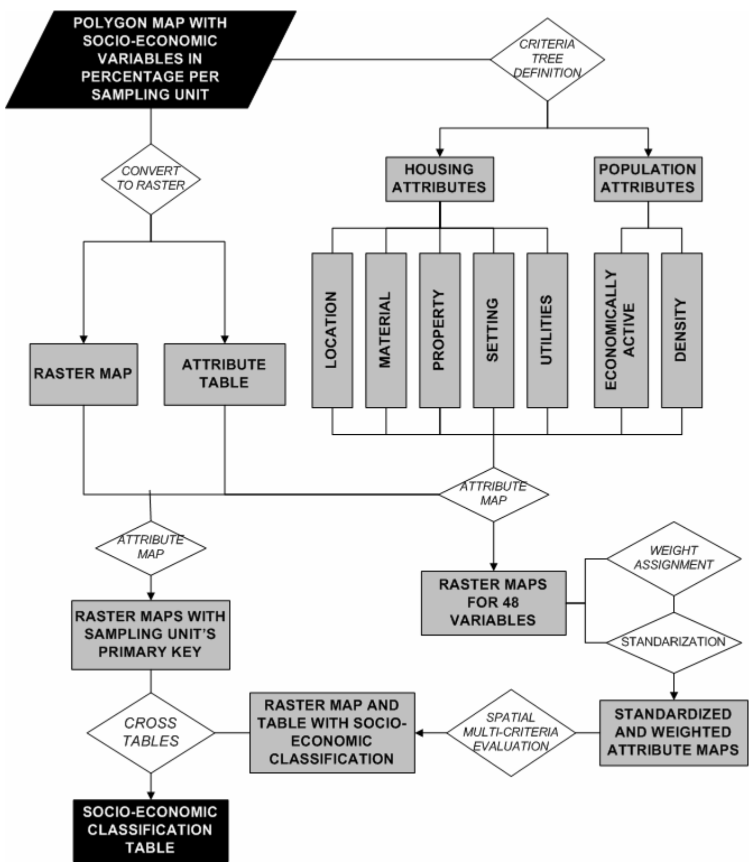
Graphic 9 : General flowchart of the analysis



5.1.1. Determining SE classification

The elaboration of a SE classification using socio economic indicators is one of the essential parts of the analysis for the present study. For this purpose, a Spatial Multi Criteria Evaluation (SMCE) model was used, aiming to obtain a classification based on standards, weights and cost-benefit analysis. Although the main use of SMCE is to determine best alternatives based on a set of options, it also can be used with one single option. This evaluation simulates a suitability analysis made to obtain the best SE classification for all sampling units (blocks). However, in this process, a full set of index (range value) for all sampling units was generated. The SE indicators are to be evaluated together based on the parameters specified for the model. The overall process of SMCE, shown in Graphic 10 below, uses the SE indicators collected in the fieldwork, grouped in two main streams or classes, namely housing total class and population total class.

Graphic 10 : Flowchart for the SE classification



5.1.1.1. Data preparation

The housing class has variables which refer to all SE indicators related to the building units in the study area whilst the population class contains variables related to the total number of inhabitants and job data (see details in Table 5 above). All them are in figures per block (sampling unit) and were divided by the total of houses per block to have a percentage instead of a raw number. In the case of the population, the indicators not only were divided by the total of population per block but also the density factor (habitants per hectare) was used. This way the values can be compared among sampling units regardless of their relative quantities.

As seen in Graphic 10 above, the processing of the maps started with the conversion to raster formats of the cadastre maps which include all sampling units in the study area. This map has an associated attribute table which has all the data from all SE indicators collected in fieldwork and generated for the analysis. The indicators had already been transformed to percentages of the total per sampling unit. The total number of housing units (housing variables), the total population (economically active population) and the area of each sampling unit were used for the corresponding case.

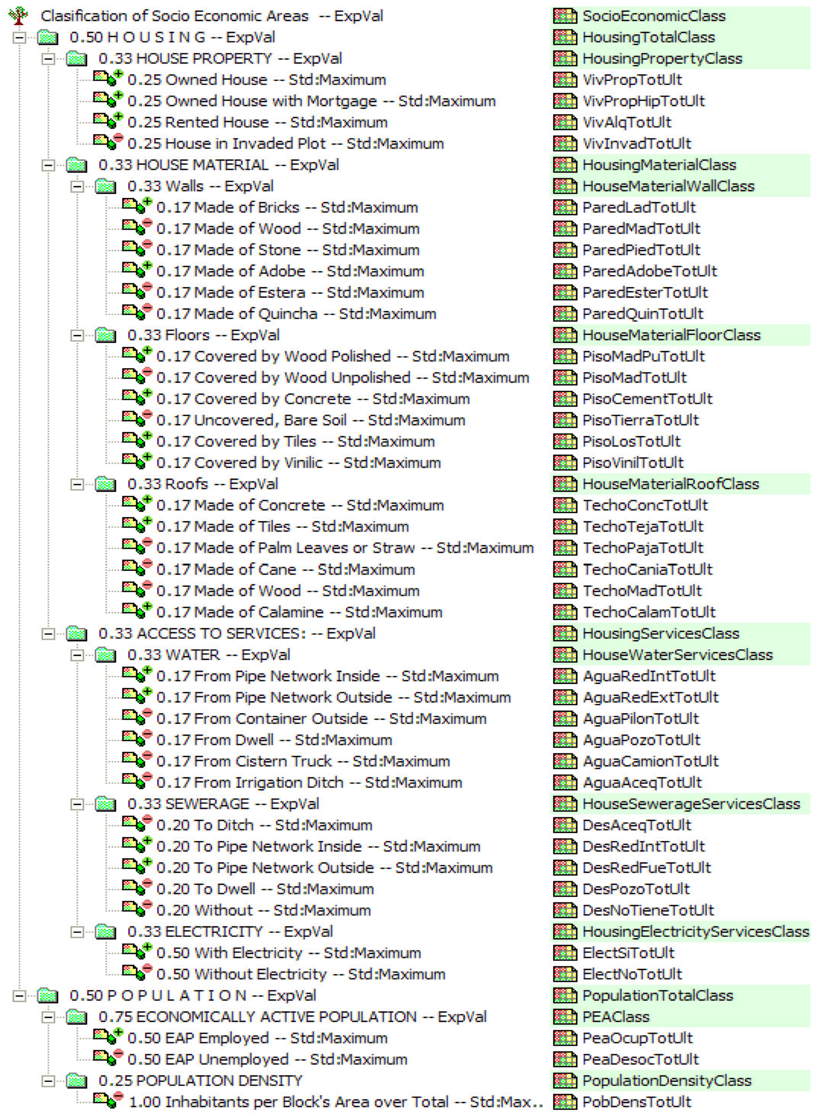
5.1.1.2. The SE classification through SMCE

To obtain a SE classification, the multi-criteria evaluation was the methodology chosen. The spatial variant, such as hold by ITC ILWIS was the tool selected for obtaining a SE classification. The SMCE allows generating a composite class map (as one option) by performing a problem definition and a multi-criteria evaluation that suits the amount of data available (38 indicators) and furthermore providing a flexible way of ordering the inputs, schematizing the classification and assigning importance to each one of the indicators. Moreover it allows having partial outputs for each class of variables and provides the final result in a raster map, ready to be understood. Having such flexibility and organization allowed to run more than one alternative with different weighting, standardization, etc. and giving the opportunity of straightforwardly making changes in the model. The 38 indicators are grouped into 7 variables (Housing Property, Housing Material, Access to Water, Access to Sewerage, Access to Electricity, Population Density and Economically Active Population) organized in two streams (Housing for the first 5 and Population for the remaining 2).

The first step on the SMCE is to set the goal for the selected methodology. As mentioned before, this will be to get SE indicators classification showing the full set of SE indexes as a raster map output with an associated attribute table. Out of this goal and as second step an organization of the SE indicators follows. The schema, shown in Graphic 11 below, is organized by two main classes of variables, namely housing and population. Both main classes will be part of the evaluation with equal importance. The housing class is subdivided in three branches or subclasses: property, which evaluates the ownership or occupancy of the property; material assess the type of construction material used, having three sub branches that include indicators for walls, roofs and floors respectively; and finally the access to services, which is divided into three sub branches that include indicators for assessing the access to basic services such as water, sewerage and electricity.

Each variable is to be produced from the base 1993 cadastre map with an associated attribute table. This map, loaded as a raster in ILWIS was used to produce the raster maps for each of the 38 indicators (see Graphic 11 below). The command Attribute Map is used, specifying the field (indicator) that is to be used. As an example, Graphic 12 below shows two of these raster inputs, access to internal-piped water and occupied economically active population, standardized in values from 0 to 1. Because of each factor (indicator) is part of the total percentage of each class or group of variables, the indicators are to be assigned the same weight as they have the same importance (see the equal values adding to 1 per branch level in Graphic 11 below). As this may be confusing from some points of view, it is important to clarify this issue.

Graphic 11 : SMCE tree with cost-benefit (+/- signs), weights (fractions) and standardization used (Std:)

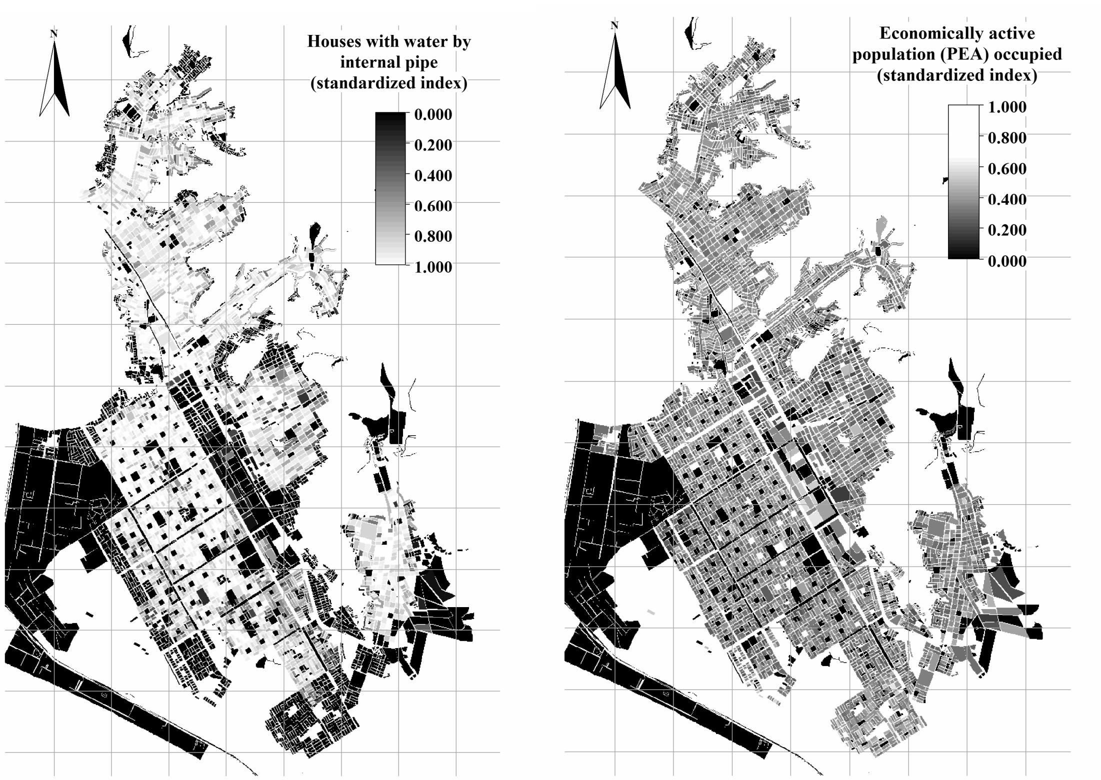


The evaluation that is to be carried out is not comparing i.e. the quality of housing materials, services or the layout of the property per se (keeping in mind that, for instance, having a wall of bricks is better than having one of adobe or an internal-piped sewer is better than a drainage to a dwell) but it is assessing which property (and therefore which tenant) has, i.e. which kind of material, service, number of jobless inhabitants, number of persons per area unit (density population), etc. The objective is to use these characteristics of the housing units to help determine (together with population indicators) the SE classification of the blocks units. It is therefore a quantitative assessment instead of a qualitative one. The number allocated for each indicator in each block is the one who contributes to the classification, not the common perception of what is better or preferable or what is not (as in the case of the brick and adobe walls).

However, there is one exception in the branch or subclass density in the Population class where weights were assigned different values (0.25 against 0.75 of PEA). This is an indicator of how crowded is a block unit and therefore much less important than PEA for the evaluation. In Lima, population density is considered an important indicator to measure the population concentration especially in poor urban areas. Although it does not necessarily determine a SE condition, it has been seen that higher densities are generally associated with poor living conditions.

The evaluation also considers cost-benefit factors as stated in the SMCE procedures in ILWIS. This implies that i.e., if a cost schema is used, the lower the index value, the better. This is applicable to indicators such as the number of houses with unsuitable construction materials in the sampling unit, where the lower indexes (closeness to or less presence in the block units, respectively) help to determine a higher SE classification. On the other hand, the benefit implies the higher value the better classification and is applicable to, for instance, houses with brick walls or concrete roofs. This means that the higher the number of houses with brick walls, the better SE classification index the sampling units get and vice versa. The cost-benefit criterion applies to all the 38 indicators.

Graphic 12 : Two examples of SMCE standardized rasters for the study area



Being all factors in each sub-branch weighted and with signs assigned, follows the loading of all individual raster (data) maps per variable. Names for each one of the branches' outputs are also necessary in this step, as well as an overall map name (called goal map) that will contain the final SE classification. This schema allows producing partial outputs (i.e. a housing material evaluation map or an overall population map). This is useful for a partial evaluation or for establishing classifications in a separate way, considering only the indicators in each branch or sub-branch, issue that will be raised lately in this chapter. At this point all the branches, sub-branches and factors (indicators) are weighted the latter have a cost-benefit schema assigned.

The final step previous to an SMCE run is to standardize the values. Almost all they were processed with the maximum standardization method as most values (already divided by the total value per sampling units) ranged from 0 to 1. The exceptions used the interval method as minimum and maximum values were more than 0 and less than 1 respectively (for more about this, please refer to the ILWIS SMCE user's manual). Setting this option the model is ready to be evaluated. The SMCE evaluation has a complex mathematical formula that uses the weights and signs of the cost-benefit to add up all standardized values. An overall generic formula for the ILWIS SMCE is shown as follows:

$$W_a(W_b(W_c(\ldots + W_n(V_1/V_{1Max}) + W_{1-n}(1 - V_2/V_{2Max})) + \ldots W_{1-c}(\ldots) + .W_{1-b}(\ldots).W_{1-a}(\ldots)) \quad (1)$$

where:
$W_{a,b,c...n}$ are the weights of the branch a, b, c or the factor n;
$W_{1-a,b,c...n}$ are the complementary weight per branch or group of factors (all adding to 1);
$V_{1,2}$ are the values of the factors 1,2 (indicators); and
$V_{1,2Max}$ are the maximum values of the factors 1,2 used for standardization.

Inside the formula, the setting of a benefit (the higher the value the better) and the setting of a cost (the lesser the better) for given factors V_X and V_Y respectively are:

$$W_X(V_X/V_{XMax}) \quad (2)$$

$$W_Y(1 - V_Y/V_{YMax}) \quad (3)$$

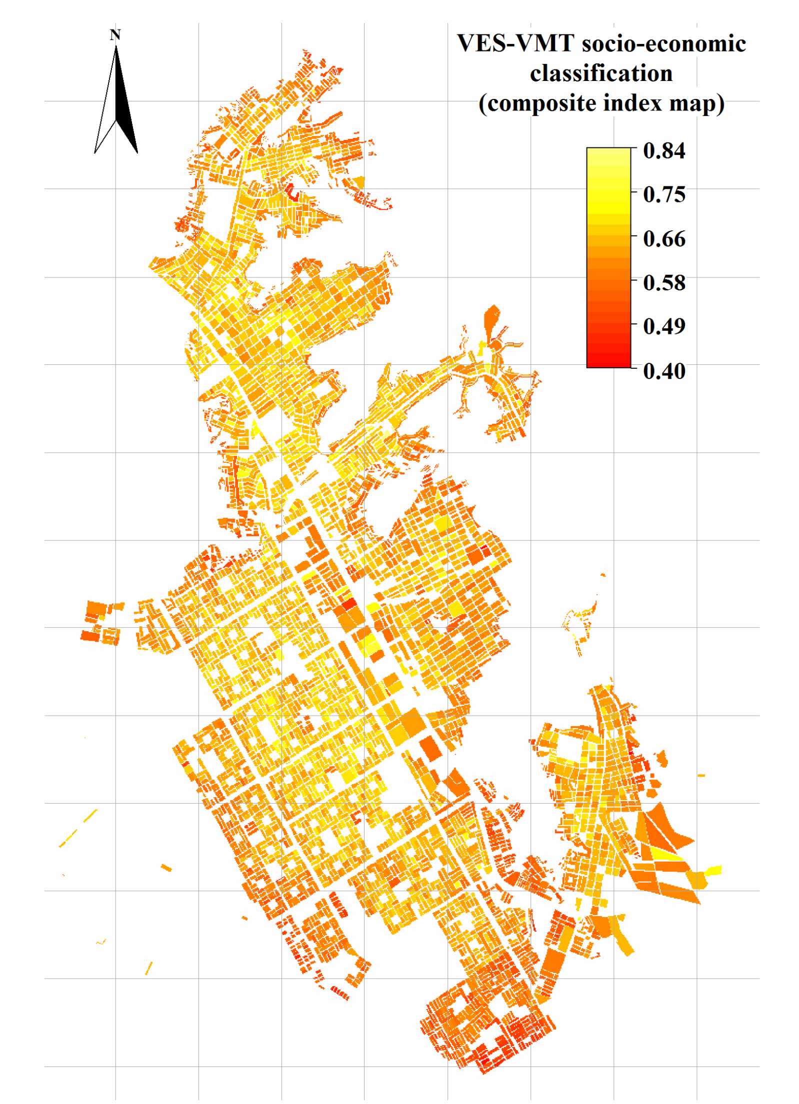
Where:
$W_{X,Y}$ are the weights of the factors x,y;
$V_{X,Y}$ are the values of the factors x,y (indicators); and
$V_{X,YMax}$ are the maximum values of the factors x,y used for standardization.

The run of the SMCE model yields a composite class map that has values at block level, as to say the SE classification, ranging from 0.37 (lowest) to 0.84 (highest). The final output map has these values in an attribute table and although it could be spatially joined with the original cadastre polygon map a more accurate procedure was used. Out of the same raster map imported to ILWIS, an Attribute Map was created containing the ID's of the blocks and the resulting table was joined to the SE index table by using the command CROSS. The final output of indexes with the sampling unit identificator was exported to ESRI ArcGIS for further representation procedures.

5.1.2. Calculating distances to paved roads

The second part of the analysis comprises the creation of a road network to be used for calculating distances and accessibility indexes. This network is based on the roads layer from AATE containing all paved roads for 1993 and includes all the unpaved roads digitized from the 1984 and 1990 aerial photographs and the cadastre map. The paved roads were categorized depending on the year of the photograph that was used to digitize them. Thus we have roads that were paved before 1984 (some 9 year-old or more) and before 1993 (base year of the analysis). The network distance was calculated for two scenarios, namely distances to paved roads in 1993 (distance to all paved roads) and to paved roads by 1984 (distance to 9+ year-old paved roads).

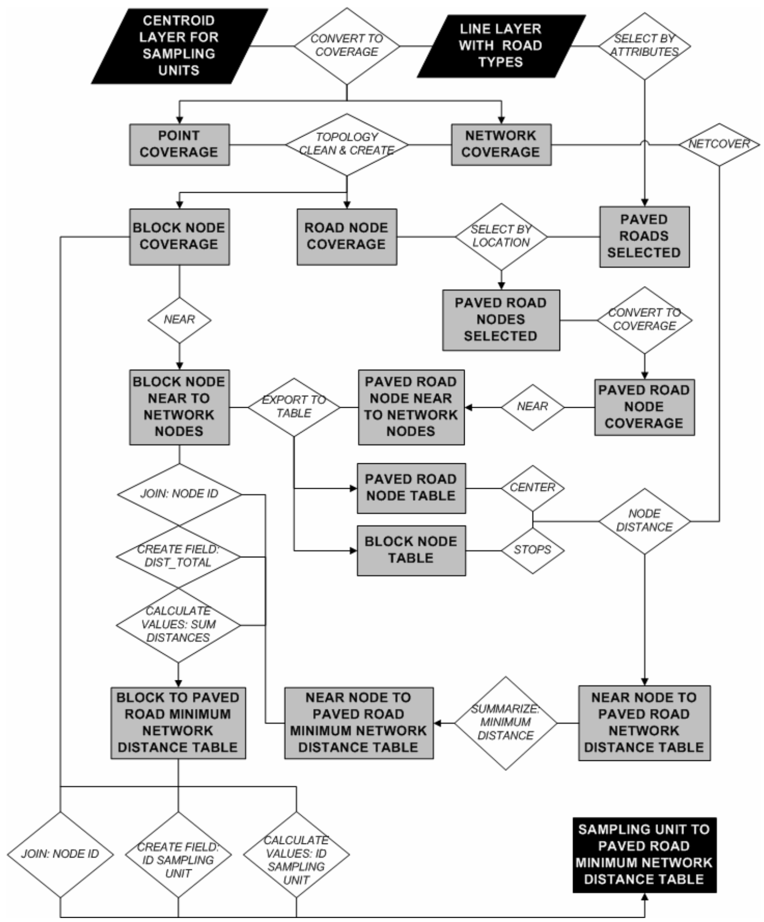
Graphic 13 : SMCE output map with the SE classification for the study area



The other inputs for this process are the centroids point layer created out of the cadastre layer and the paved roads nodes. The centroids are used as the trips origin as they represent all possible origins in each sampling unit, method commonly used in transport modelling. Each sampling unit has a unique centre point from which distances are to be calculated. The general process for each of the distance calculations is depicted in Graphic 14 below.

The road network is processed in ESRI ArcGIS to assure there are no dangling nodes and junctions are adequate. The conversion of the layer to coverage to be loaded in ESRI ARC ensures this topology consistency. Inside the coverage a node feature is created and the topology is finally cleaned. The network is ready to be used in the initial calculation. The centroids layer is also converted to point coverage as well as a spatial selection of the road-network nodes that connect with the paved roads. These points are going to be the destinations for the distances to be calculated from the centroids of the sampling units.

Graphic 14 : Flowchart for the calculation of network distances from sampling units to paved roads



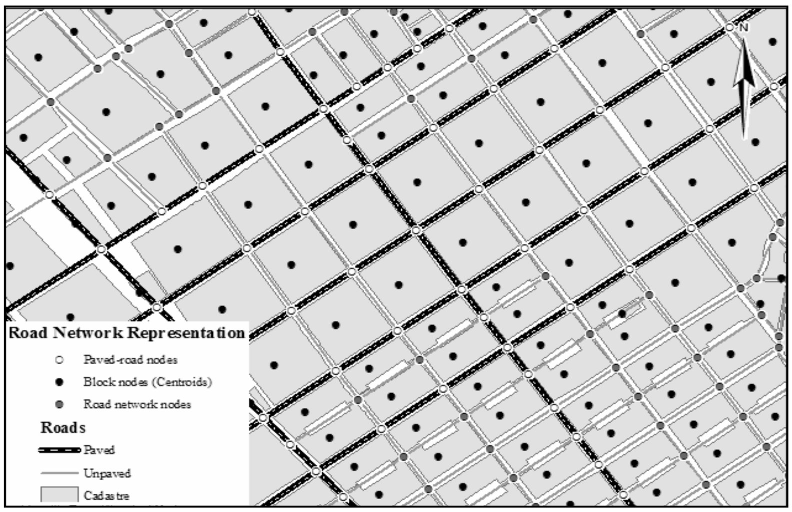
5.1.2.1. Finding the closest network nodes

The road network coverage, now with nodes and proved consistency is loaded in ESRI ARC and declared as the network for the calculus using the command with the same name. Then, the centroids coverage is invoked in the command NEAR to find out the nearest road-network node and the distance to it. The command is also applied to the paved-road nodes coverage as the closest road network node to them is also to be identified. This is done having in mind the next step, where an overall distance is to be calculated. The command that will be used (See NODEDISTANCE in 5.1.2.2) only calculates the distance from nodes in the same network coverage, thus the nodes that are closer to origins (sampling units) and destinations (paved-road nodes) have to be identified beforehand. The identification also includes the Euclidean distance to the nearest node, to be added to the final distance calculation. In the case of the paved-road nodes this distance is going to be zero for all them as these nodes were taken from a selection of the road network. The elements that take part in this calculation are shown in 0 below.

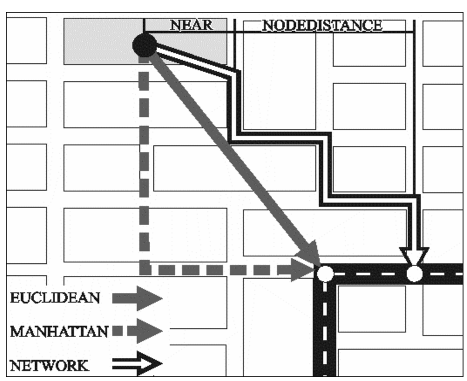
5.1.2.2. Network distances

The last part of the elaboration of the distance to paved roads variable is to determine the lengths. It will be a representation of the walking distance on unpaved roads to get from one sampling unit to the nearest paved road. Therefore, it is necessary to calculate network distances instead of using the conventional Euclidean or Manhattan methods. The former refers to the straight distance (as the crow flies) between two points, the shortest one. The Manhattan distance, more applicable to our analysis, uses the horizontal and vertical components of the Euclidean distance to give a more 'real' distance inside urban areas. However, neither the Euclidean nor the Manhattan methods represent reality in this context. The best setting for this calculation is to use the Network distance, as it gives the real length of the course from a given sampling unit to the nearest point (node) in a paved road. The schema showing the differences between the three options is shown in Graphic 16 below.

Graphic 15 : Road network representation for distance calculation (showing nodes and centroids)



Graphic 16 : Euclidean, Manhattan and Network distances from a sampling unit to a paved road

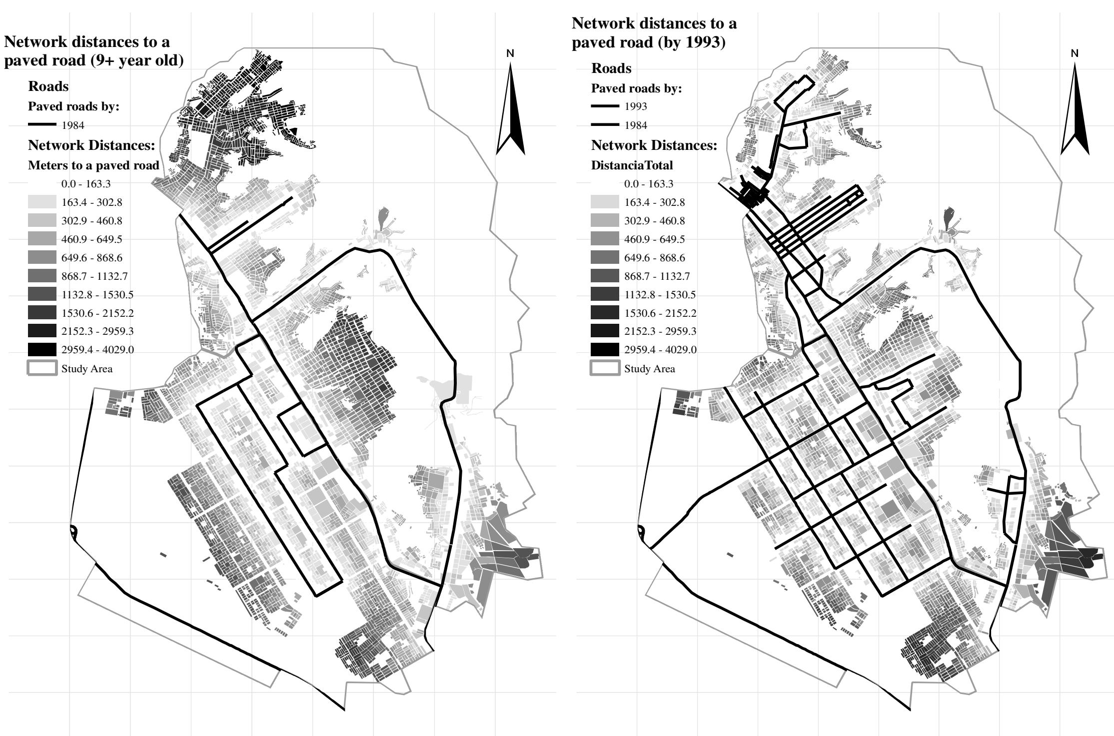


The calculation of the Network distance is possible inside ESRI ARC. The Arcplot extension is used for such purpose. As seen in Graphic 14 above, the command uses three sources, which are the network coverage and the two tables with the closest network nodes for sampling units and paved roads. These tables, output from the NEAR command in 5.1.2.1 above, have the identificator of the closest node in the road network coverage to the centroids in the sampling units and the intersections with the paved roads, together with the distances. These points will be used to set up the shortest path in order to calculate the minimum distances.

After declaring the network coverage to be used (the same as in 5.1.2.1 above) the two tables are loaded with the command NODES. The NODEDISTANCE command requires the previous specification of an impedance value that in this case was set as the current length of each road network link. This is done through the command IMPEDANCE though it could be left as the default is precisely the network links' length. Finally, the NODEDISTANCE command is invoked using the NODES option as the points of origin and destination of each short path were loaded by it. The distance can be set to be calculated in any of the three options described above, though the only suitable one is, as explained, the network method. An output table containing the distances for all sampling units is the final output.

The final step of this calculation is to join this results with the SE indexes output in 5.1.1.2 above. This is the previous step before the final analysis through statistical tools. The table with the distances for each sampling unit to the closest road network node as well as the table with the network distance from these nodes to the paved roads are to be included. Both distances, added up together will yield the final distance from sampling unit centroids to the paved roads in the road network. This is shown in Graphic 16 above, being the distances from sampling units to the nearest node labelled NEAR and the network distances as NODEDISTANCE according to the ESRI ARC command used for finding out each one. These two tables had to be joined with the centroids layer (holding the sampling units ID) and then the latter joined to the cadastre layer that holds the SE classification. A final database table is created with all these data, as well as a layer for representation purposes. This was done for each of the three scenarios depicted above (distances to all paved roads in 1993 and to roads paved before 1984). The final outputs (shown in Graphic 17 below) are now ready to be used in the statistical procedure.

Graphic 17 : Network distances to 9+ year-old (left) and by 1993 (right) paved roads in the study area



5.1.3. Assessing the dependency

The last part of the analysis comprises the use of statistical tools for assessing the association between independent and dependent variables. In this study the former are both the road network distance to the paved roads and the accessibility indexes whilst the latter are the SE indexes (and their partial outputs such as housing and population variables) generated by SMCE. The correlation shows if there is a level of correspondence between the dependent variables and the independents. As the independent variables are purely spatial (distance and access) the supposed dependency will have the same characteristics. Thus, we will have an assessment of the spatial correlation between the SE classification (index ratio) of each sampling unit with its closeness or farness to a paved road in 1993 (all paved roads) and 1984 (9+ year-old paved roads). The tools to be used as well as the procedures carried out will be thoroughly explained in the following paragraphs.

5.1.4. Statistical Correlations

The Person's Product Moment (PPM) is the statistical tool used to asses correlations between two variables. As described by Pacheco (2004) the PPM formula assumes a normal distribution in the data to be correlated. It can be evaluated through several tests, being the most common the comparisons between median and mean and some standard procedures as the Shapiro-Wilk Normality test. Histograms and distributions are used to visually inspect the normality of data. This process consists on assessing the tendency or skew ness of the data regarding a bell-shaped normal curve.

Both variables must have a normal distribution, which corresponds with one unique sampling group. A logarithmic conversion can be performed to eliminate the skew ness of the data. Provided all data has normal distribution, the next step was to test the correlation of dependent against independent variables. In that, PPM indexes the average association between two variables (Griffith and Layne, 1999). Its formula is expressed as follows:

$$r = \frac{\sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}}{\sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}} \sqrt{\sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n}}} \qquad (4)$$

And using the same values for x and y will lead to the autocorrelation formula, which is used to test the clustering pattern of a variable in the space. PPM result will output a coefficient that represents the sense and strength of the relationship. This number -between -1 and 1- shows a negative or positive correlation according to its sign. If it is positive, it means that both variables grow together: the greater value of one implies the raising of the other. If the coefficient is negative, the relation is expressed in the other way. To clarify this, take the example of a perfect correlation. In this case, the PPM coefficient for a variable with itself is +1. As all its values are compared with exactly the same the relation is perfectly positive (the growing of the variable is related to its values' raising) and completely strong as it has reached the maximum PPM value. As the strength of the relationship in a correlation is measured by PPM coefficient value, the closer to 1, the stronger the relation in the defined sense (positive or negative). However, this has only meaning in a linear way, as to say, when the regression formula of both variables is in the first degree (no quadratic formula, the graphic is a straight line).

However, having a low PPM coefficient does not necessarily mean the inexistence of a relationship; it may happen that this is expressed in a non-linear way. When evaluating this, the level of significance has equal amount of importance. Also known as p-value, it represents the probability of obtaining results as extreme as the one observed. In that case, the closer the value to 0, the lower the possibility of error. In PPM, higher p-values mean strong options of not having similar results, and thus weak options of proving a linear relationship.

In the example shown in Table 6 below, coefficients for Price versus Location and Area versus Price have a strong positive PPM correlation, as they are close and very close to +1 respectively. Location and Area variables have a lesser strength as their coefficient lower. The p-values for both reaffirm this. Between Price/Area and Price/Location the value shows the significance level is very small (less than 0.001) then the correlation is significant and the two variables are linearly related. In the case of Location/Area the p-value is higher, which means the linear relationship is not much consistent. In that case is better to explore the relation by plotting their values in scattergrams.

Table 6 : Pearson's Product Moment correlation example

		Area	Price	Location
Pearson Correlation	Area	1.000	**.960**	**.484**
	Price	**.960**	1.000	**.707**
	Location	**.484**	**.707**	1.000
Sig. (2-tailed)	Area	.	.000	.019
	Price	.000	.	.000
	Location	.019	.000	.
N	Area	24	23	23
	Price	23	23	22
	Location	23	22	23

Taken from SPSS Results' Coach

It is important to mention here that the number of observations play an important role in correlation and that mid-low coefficients are common and of high importance among social sciences (Griffith and Layne, 1999, Feng, 1996). Searching for the relationship of two variables with a few entries may yield a strong PPM coefficient, but it might be inconsistent and would also give each entry an extreme importance in the relationship. On the other hand, a considerable amount of points will always be more difficult to correlate. Thus, one of the most important issues in this is the p-value index, which makes the relationship significant. This means that the linearity, though low, is consistent and the values are corresponding with a slightly linear relationship. Consequently, in a large dataset such as the one used in the present study, a small PPM would turn significant if p-values are consistent with it. The principal purpose of the significance coefficient is to tell us how sure we can be when rejecting the null hypothesis of no-correlation.

The explained variance, defined by r^2, is used to show the percentage of the proportional reduction of error when one variable is used to predict the other. It is also possible to infer that r^2 percentage of the variance in the dependent variable rate is accounted for by the percent of the independent variable. The size of r^2 is determined by the spread of the actual observations around the linearity of the relationship. Thus, if all observations are on the line, the value would be 1; if they are randomly scattered it would approach 0 (Nachmias and Nachmias, 1997). However, if the latter happens, we can not rush to the conclusion that the variables are not related. Some complementary test (such as the test of significance) may help us determining whether the correlation exists or not. In the same way, it can be useful to prove if the relationship between two variables is significant or not.

5.1.5. Test of significance

The Test of Significance (*F*) or F distribution is a right-skewed distribution used most commonly in the analysis of variance. Likelihood that the obtained PPM correlation is due to sampling error is measured by *F*, based on the ratio of the explained r^2 to the unexplained $1\text{-}r^2$ variance. This significance test is defined by the following equation:

$$F = \frac{r^2}{1 - r^2}(n - 2) \qquad (5)$$

Where *n* stands the number of samples drawn. The degrees of freedom are inferred from it, being the amount of *n* samples minus 2. Resulting values are compared with the table of values for *F,* located using *n-2* and *n-1* as the degrees of freedom for the sampling group. Thus, the null hypotheses can be rejected when the value obtained is higher than the value in the table, for the selected probability range (0.05, 0.01 or 0.001, 95%, 99% or 99.9% respectively). Tables for *F* values are shown in the Annex 5.

Having 5055 sampling units in the present study, the minimum values required to reject the null hypotheses (no correlation and no significant relationship) are located in the ∞ degrees of freedom (5055-2=5053, out of the range of the table) and the column for 1 degree of freedom as we are testing two groups (2-1=1). Thus, minimum PPM required values calculated using (5) are summarized in Table 7 below.

Table 7 : Minimum PPM and *F* values required to reject the null hypothesis

Probability (*p*)	PPM (min.)	*F* required (min.)[f]
0.05	0.0553	3.8466
0.01	0.0726	6.6445
0.001	0.0926	10.8457

Taken from Blalock (1979)

5.1.6. Correlation Procedure

The data about SE classification and the distances to paved roads that was output in 5.1.1 and 5.1.2 above was joined in the sampling units table assigning each one the corresponding. This table was then exported to a database file format (DBF) in order to be opened in the statistical software SPSS. The entries were checked to ensure there were no null values as they will be exported like zero values in the DBF table. This may cause a wrong correlation assessment.

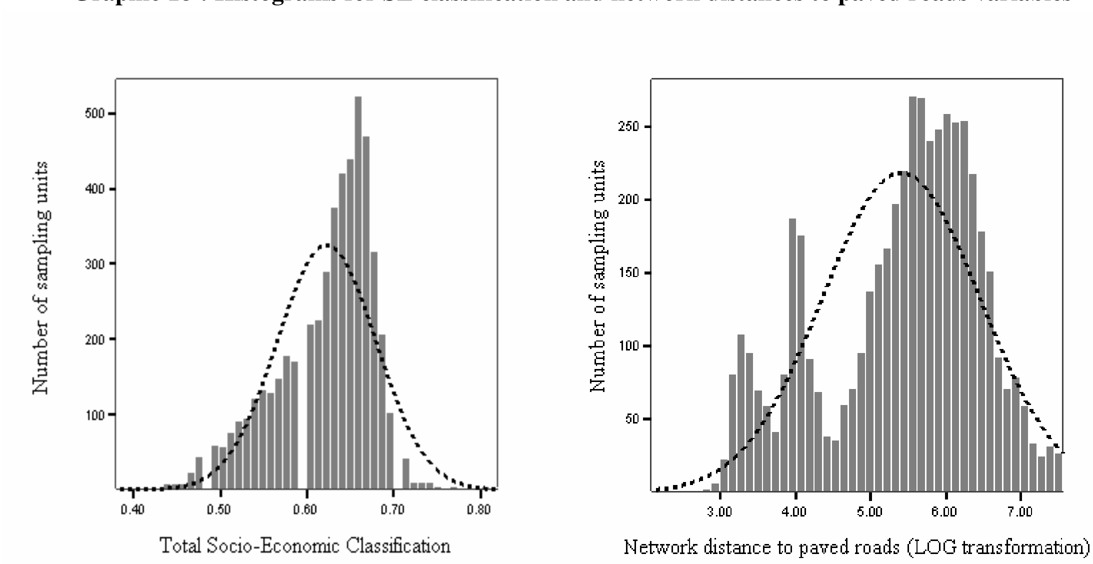
The SPSS Bivariate Correlations procedure computes PPM coefficient with their significance levels. PPM measures how variables or ratio data are related. Before calculating a correlation coefficient, data was screened for outliers (which can cause misleading results) and an evidence of a linear relationship. This was done by using the histograms and scattergrams. In the former, the normal distribution is sought. To evidence it in a skewed distribution (mean and median close to one side of the distribution), a logarithmic conversion of the values is often necessary. As said before, the PPM coefficient is a correct measure of linear association, situation where two variables can be perfectly related, but if the relationship is not linear, Pearson's correlation coefficient is not the appropriate procedure. Thus, scattergrams help to visualize the level of linearity between variables before proceeding.

The correlation analysis is divided in two sections. First, the correlations related to partial outputs (the branches that compose the SMCE evaluation) as to say Housing variables and Population variables are analysed. Finally the main correlations for the study, the SE classification (composed by the two partial outputs mentioned) is made. In both cases all steps mentioned at the beginning of this section (see 5.1.4 above) are applied.

5.1.6.1. Correlations of the final SE Classification

The histograms for the two variables to be correlated, the SE classification (dependent variable) and distances to a paved road (independent variable) are shown in Graphic 18 below. Although the histogram for the distance to all paved roads seems to be not so normally distributed, it is due to the distances may have been fairly homogeneous values in several cases. However, the overall assessment shown by the bell-shaped dotted line in the histogram defines a normal distribution. The same applies to the SE classification dependent variable and both can be correlated. The histograms for the remaining independent variables are in Annex 3.

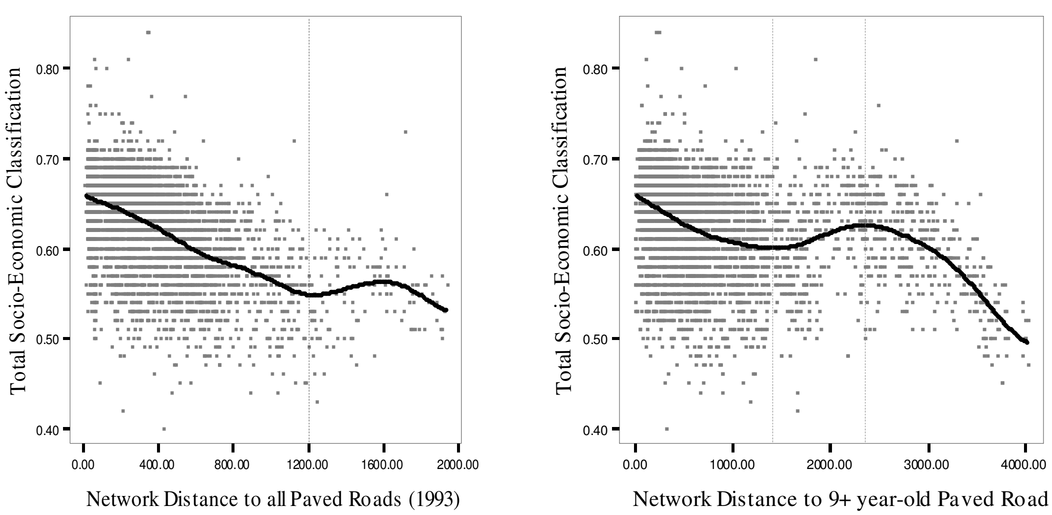
Graphic 18 : Histograms for SE classification and network distances to paved roads variables



Through the scattergram, shown in Graphic 19 below, the values that are out of the normal distribution of sampling units can be easily found and analyzed. These values can influence the correlations and therefore must be identified beforehand. An initial inspection of a correlation can also be performed by looking at the concentration of values in a certain shape, which may be preferably linear. The graphic bellow shows the relationship between the SE classification values and the network distance to paved roads for two different time period.

To the upper and lower parts of the scattergrams, some outliers can be seen (extremely high/low values outside the point cloud), though they do not greatly influence the final shape of the smoother fitted to the values (the black thick line). The scattergram for the 1993 scenario (to the left), evidences a negative correlation as shown by the smoother inclination from left to right, as to say lowering SE index values as the distance increases. This evidences that a shorter distance from the blocks (sampling units) to the paved roads corresponds to a higher SE index (classification) in 1993. In most of the graphic, the clustering of points to a linear pattern previews consistency in the correlation. The greater amount of point values are in this linear part of the scattergram. There is also a clear change in this shape in the 1200 meters distance to paved roads, where also the linearity of the correlation changes to an undefined pattern. Therefore, we can expect a consistent correlation, though the values in the not-linear part of the graphic as well as points very far from the smoother, point to a not high value (not close to -1 absolute). The second scattergram has slightly the same tendency, though the distances of the breaking points in the smother vary. For the distances to 9+ year-old roads the linearity is smooth roughly until the 1400-meter distance to a paved road. From there the linearity changes to a curved pattern, impossible to be correlated in a PPM analysis as explained before. Though there is no clear defined linearity, the expectation is about getting a high PPM value for the first scattergrams (SE classification against distances to all paved roads) and a much lower for the second (SE classification and distances to 9+ year-old paved roads). Although the tendency is somehow clear so far, the real importance of these correlations will be finally defined by the PPM coefficients.

Graphic 19 : Scattergrams for the SE classification and distances to paved roads, showing correlation



Finding a bivariate correlation is a very straightforward process in SPSS. The two fields that contain the dependent variable (SE classification) and the independent (distance to a paved road) are included for the analysis. The Pearson option is checked as what we are analysing is continuous and symmetrical (normally distributed) ratio data. The two-tailed option was also selected as we want to see the significance of the correlation at this level. This means that the test is to be applied in both sides of the distribution and that the level of confidence should be understood in such way. The correlations for the two variables are shown in Table 8 below.

These correlations confirm the initial tendency shown in the scattergrams. The two correlations are consistent and have the anticipated negative sense (the higher the distance to a paved road the lower the SE classification). As previewed the coefficient have different magnitude: the SE classification and the distance to all paved roads have the highest value (0.477) and the correlation with distances to 9+ year-old have a lower coefficient (0.263). The two correlations are significant at the 0.01 level in a two tail analysis (both sides of the probability distribution) meaning that there is a 98% confidence that these results are not produced by chance. If another random analysis is performed with the same sampling units there is only a 2% probability of not having the same results. This gives consistency to the results and supports the correlations found.

Table 8 : Results of PPM correlations for the SE classification dependent variable

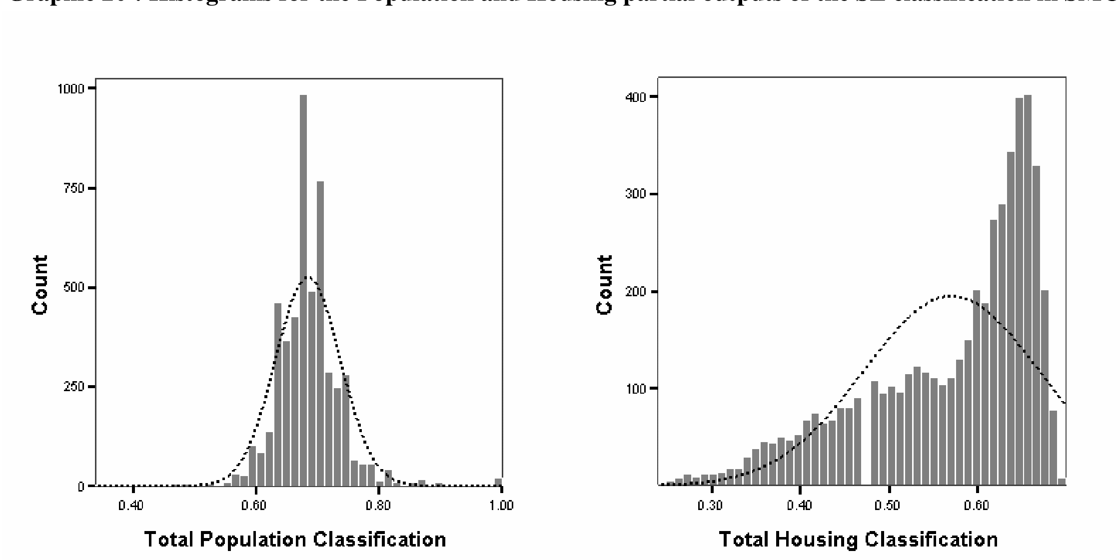
	Total socio-economic classification	Element description
Network distance to all paved roads (in 1993)	**-0.477****	PPM Coeficient
	.000	Significance (2-tailed)
	5055	Number of sampling units
Network distance to 9+ year-old paved roads	**-0.263****	PPM Coeficient
	.000	Significance (2-tailed)
	5055	Number of sampling units

**Correlation is significant at the 0.01 level (2-tailed).

5.1.6.2. Correlations of the partial SE evaluation

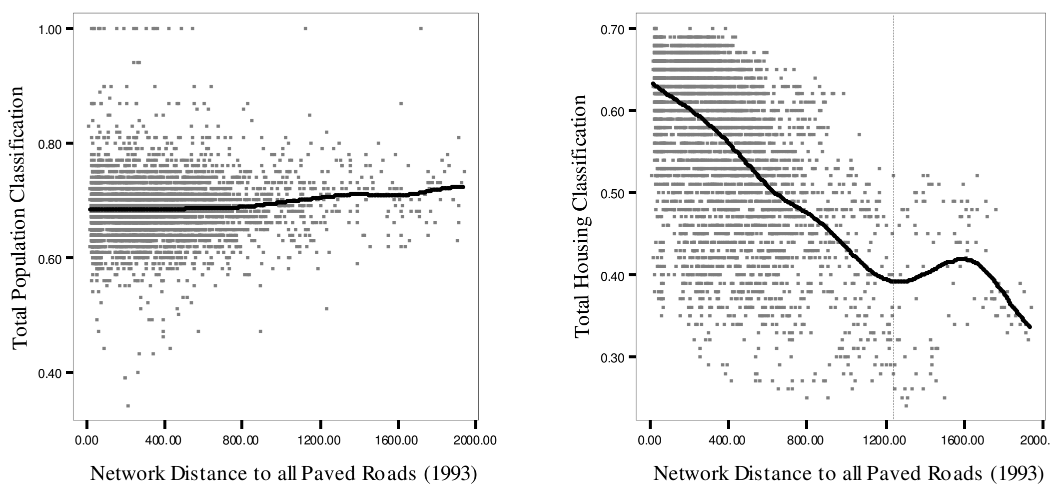
The SMCE procedure for finding out the SE classification yielded several partial outputs, being the most important the two main classifications or branches that support the SE classification, namely Housing and Population. The histograms for these two partial outputs are shown in Graphic 20 below. They evidences both composite variables are within the normal distribution, though the Housing variable is skewed to the right. Nevertheless, the presence of the normal curve and the absence of variations other than the higher values (around 0.65 index value) confirms the values come from one sampling population and is able to be correlated using PPM.

Graphic 20 : Histograms for the Population and Housing partial outputs of the SE classification in SMCE



For the two dependent variables, the scattergrams plot their values against the distances to the paved roads (the independent variables, shown in Graphic 21 below) and both them evidence opposite results. Whereas the scattergram of Population and distances to paved roads in 1993 show a clear lack of correspondence (evidenced by the almost horizontal line) the graphic for Housing and distances to paved roads present a clear correlation pattern. This correlation is negative and clearly linear until around the 1200 meter distance (dotted vertical line), from where the fitted smoother (thick black line) follows an irregular curved shape. Nevertheless, it is expected a high PPM correlation value for the latter.

Graphic 21 : Scattergrams of Population and Housing variables against distances to all paved roads



The results for the remaining independent variable distance to 9+ year-old paved roads are included in the Annex 4. For the Housing dependent variable the scattergram show a clear correlation pattern though the relationships seem to be much less linear than the Graphic 21 above. There are also two breaking distances in the smoother. The first one in the 1500 meters distance to paved roads approximately, where the linearity of the correlation changes to a undefined pattern though in the second breaking distance (around 2300 meters from the paved road), the linearity changes to a curved pattern as explained possible to be correlated in a PPM analysis. Therefore, we can expect a consistent correlation; it is expected negative correlations in both though the strength of the correlations seems to be dubious. For the Population dependent variable the same lack of correspondence is evident in both graphics.

The main part of the correlation process, the PPM calculation finishes defining the real sense, magnitude and signification of the correlations as outlined in the scattergrams. The final PPM results for the partial output variables are shown in Table 9 below.

Again, the initial characteristics of the correlations outlined by the scattergram analysis are confirmed by these results. The two correlations for the Housing variables are negative. Nevertheless, the first correlation (-0.528) is more consistent and also its coefficient is much higher than the second one (-0.280), both results represent a clear and consistent correlation between the variables. This means that the longer the distance to a paved road (in each scenario of the independent variable) the lower the

Housing classification of the sampling unit. It is worth to remark the very high coefficient achieved in the correlation of this variable with the distance to all paved roads. As explained in 5.1.4 above, having a coefficient of -0.528 for a social science study (where values in sampling units tend to be extremely discrete) is of great relevance, moreover if we realize that 5055 values were used in this correlation. High coefficients such as the one seen here are normally found in more exact sciences (such as natural sciences) or in a much lower number of sampling units. Finding a coefficient of 0.528 is clearly remarkable and strengths the importance of the result. For the Housing variables, the results are consistent at the 0.01 level in a two-tailed analysis.

Table 9 : Results of PPM correlations for Housing and Population dependent variables

	Total Housing classification	Total Population classification	Element description
Network distance to all paved roads (in 1993)	**-0.528****	0.077**	PPM Coeficient
	.000	.000	Significance (2-tailed)
	5055	5055	Number of sampling units
Network distance to 9+ year-old paved roads	**-0.280****	0.023	PPM Coeficient
	.000	0.096	Significance (2-tailed)
	5055	5055	Number of sampling units

**Correlation is significant at the 0.01 level (2-tailed).

In the case of the Population variable, the initial idea of a lack of correlation is confirmed. The PPM coefficients found are close to zero, meaning there is no correlation. There are also P-values equal zero; the exception is the correlation with the 9+ year-old paved roads. In this case the P-value is 0.096 that's means there are possibilities of not having a similar result as well as weak options of proving a linear relationship (the values of the two correlations are not consistent).

5.1.7. Testing the significance

Through the F test the significance of the PPM coefficients can also be tested. The coefficient of r^2 explained on 5.1.5 above will help to determine if the correlations are significant in terms of a possible sampling error. Although the PPM calculation in SPSS also yields a significance level (which was already explained) it is healthy for the study to double-check the result's consistency and assure its quality. The test is based on the assumption that there are two options regarding the significance of the PPM results: we have a valid hypothesis H_1 stating that the correlation is not 0 (correlation significant) and a null hypothesis H_0 where the correlation equals 0 (not significant). This corresponds to the two variables analysed, the SE classification and the distance to the paved roads. The applied formula (5) to the correlations (see 5.1.2 above) found for the SE classification and the distances to paved roads variables (r = -0.477 and r = -0.263) is shown as follows:

$$F = \frac{0.477^2}{1-0.477^2}(5055-2) = 1488.35 \qquad F = \frac{0.263^2}{1-0.263^2}(5055-2) = 375.48$$

And the results are far above the minimum values required for safely rejecting the null hypothesis, shown in Table 7, for all probabilities. Therefore, the F test allows rejecting the null hypothesis and therefore accepting the research hypothesis, strengthening the findings by proving consistency in the correlations. Finally, the testing of the significance also allows rejecting the null hypothesis for all consistent correlations for the partial outputs (Housing variables). The F values for the two consistent PPM values (0.528 and 0.280) are calculated as follows:

$$F = \frac{0.528^2}{1-0.528^2}(5055-2) = 1953.22 \qquad F = \frac{0.280^2}{1-0.280^2}(5055-2) = 429.86$$

Being all the results of the F test also far above the minimum value for safely rejecting the null hypothesis, as shown in Table 7. These results support the consistency of all the correlations found, validating the results obtained.

5.1.8. Observing the relationships

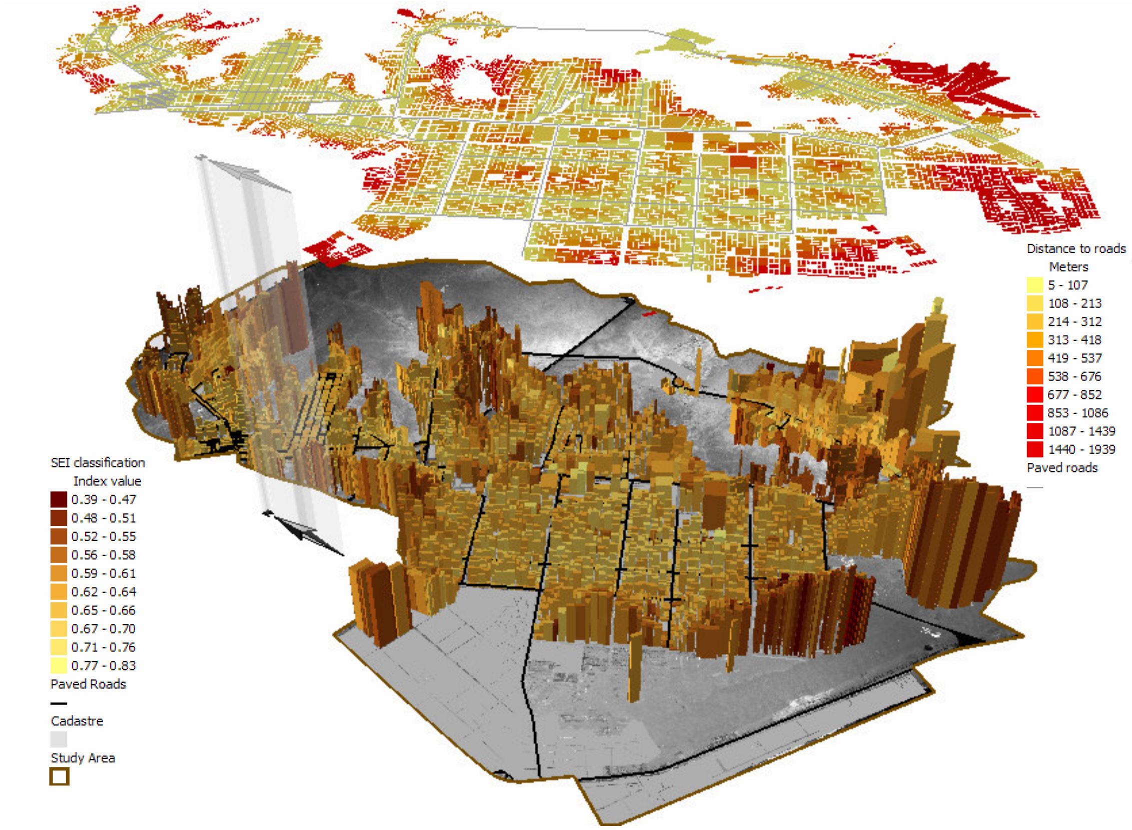
Other complementary way to verify the level of correlation is through a composite GIS map showing the two variables at the same time. This method for testing correlations was first introduced by Pacheco (2004). As in conventional GIS simbology only one variable can be represented in a map, the right tool is provided by ESRI Arc Scene. The software is mainly used to visualize three-dimensional maps such TINs or DEMs. However, it allows representing values as an extrusion (height) of the data for each polygon. Some of these graphics are shown and explained.

In the case of the SE classification and the distances to paved roads, one of the variables can be represented in a colour simbology whilst the other can appear using the extrusion option. Choosing which one is more suitable for one option is a matter of how dissimilar are the values. A simple statistical test showing the standard deviation will do. In this case, the distances to paved roads proved to be more suitable for the extrusions. The map for the distance to all paved roads in 1993 and the SE classification is shown in Graphic 22 below.

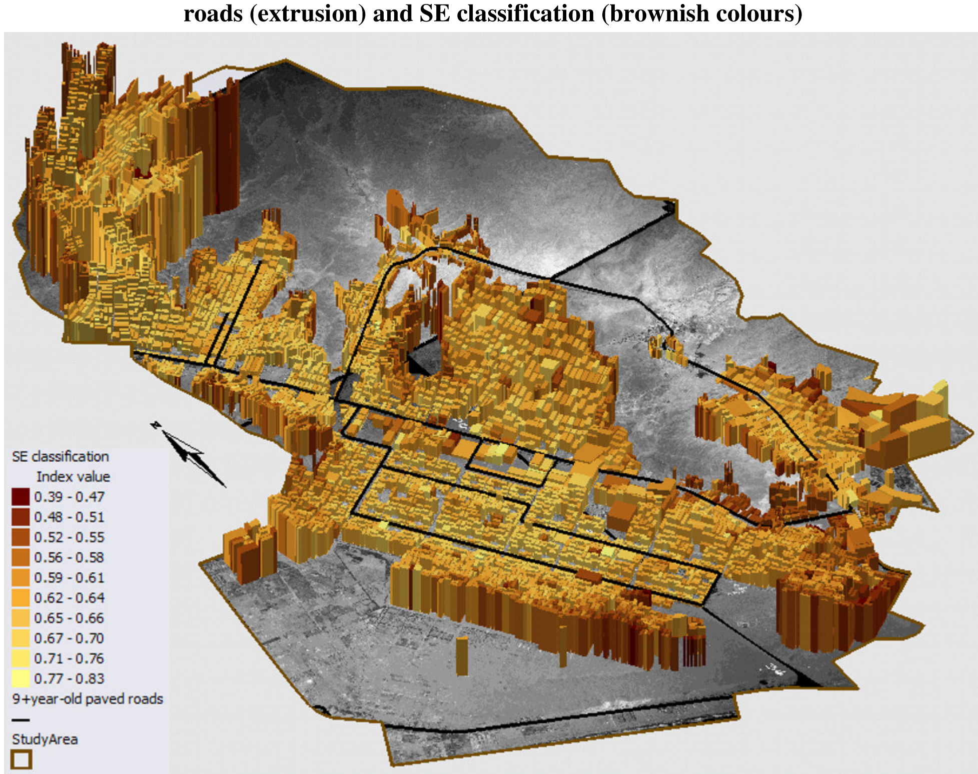
The graphic evidences a high level of association between high heights (the distance independent variable) and darker colours (the dependent variable SE index) and vice versa. This reflects the high PPM coefficient achieved for the correlation (0.477). A correlation is more than evident and the tool also helped to prove it. This is a reliable way of showing the relationship between variables and – contrary to the scattergrams- make them easy to understand for the majority of untrained observers.

The Graphic 23 below shows the correlation between distances to 9+ year-old paved roads and SE classification. The graphic evidence results quite similar to the previous correlation though the lesser PPM coefficient achieved is evident (see Table 8 above). In this case it is not clear whether higher extrusions heights have darker brownish colours assigned. This means that the longer distances to a 9+ year-old paved road is equally correspondent to higher and lower coefficients of the SE classification. The result is logic in the light of the lower coefficient (0.263) yielded by the PPM calculation.

Graphic 22 : Three-dimensional representation of the correlation between distances to all paved roads in 1993 (volume extrusion and upper yellow-red colour map) and the SE classification (brownish colours)

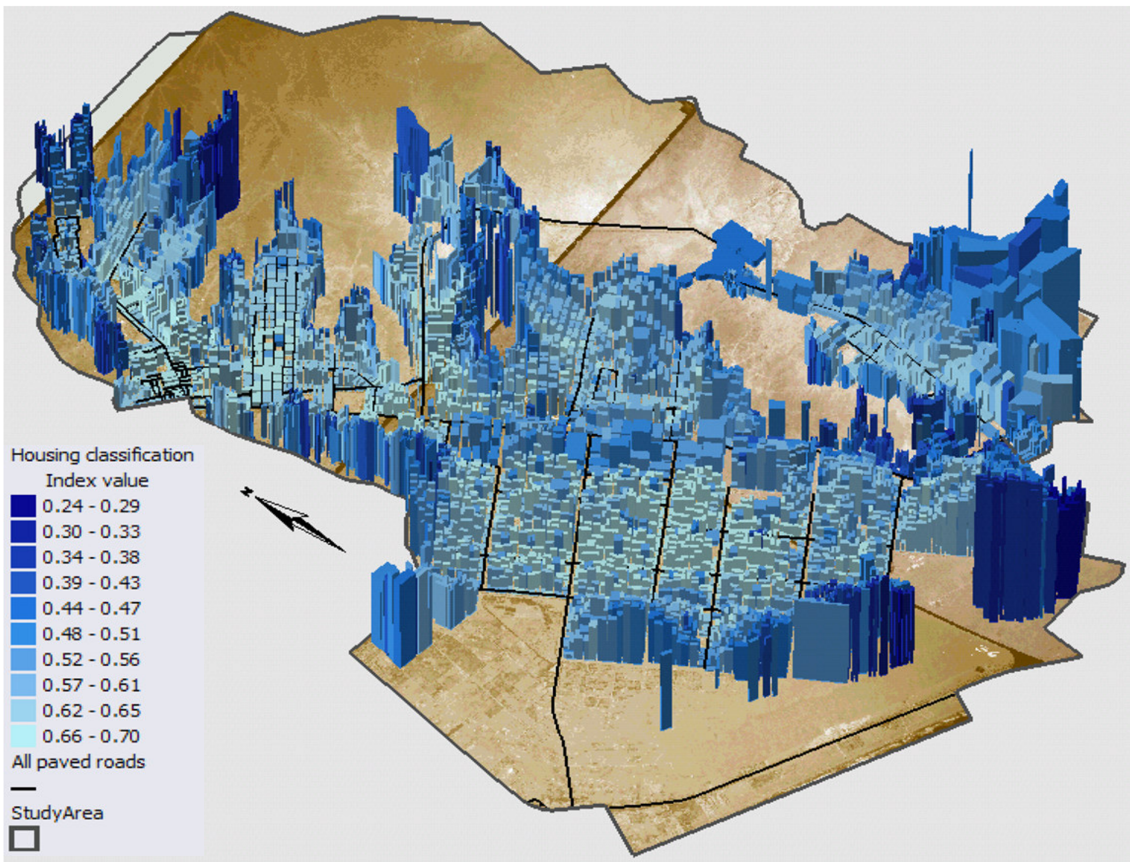


Graphic 23 : Three-dimensional representation of the correlation between distances to 9+ year-old paved roads (extrusion) and SE classification (brownish colours)



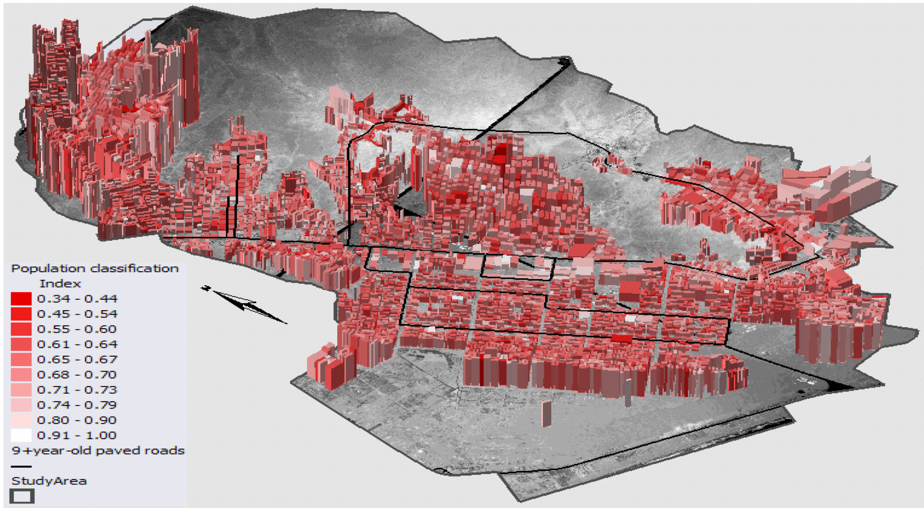
In the case of the partial output for the Housing variable, the correlation depicted in Graphic 24 below reflexes the high PPM coefficient yielded by the correlation analysis. The level of association between the representation of the distances to all paved roads by extrusion heights (distances) and the Housing classification (a partial output on the SE classification through the SMCE evaluation) showed in the graphic in bluish colours is evident. The majority of higher extrusion heights (longer distance to a paved road) correspond to a darker blue colour (lower Housing classification index value). It is difficult to see darker tones among the lower extrusions heights in the same schema, which would mean a lack of correlation. However, some higher extrusions do have a lighter colour and this is the portion of the sampling units that does not comply with the correlation. They account to lessen the correlation to the final 0.528 of the PPM result for these two variables.

Graphic 24 : Three-dimensional representation of the correlation between distances to all paved roads (extrusion) and Housing classification (bluish colours)



Finally, the correlation between the partial SE classification output for the Population classification and the distances to paved roads have in general a lack of correlation that is evident (see Graphic 24 above). The scatter of reddish colours (Population classification) all over the study area without any consistent association with higher or lower extrusion heights (distances to 9+ year-old paved roads) corresponds to the lower PPM coefficients yielded by the analysis for this correlation. It is impossible to devise a pattern in the graphic and therefore a correlation between the two variables represented there. The inexistence of the correlation in this case, depicted in the scattergrams and sustained by the PPM values is also supported by this graphic output.

Graphic 25 : Three-dimensional representation of the correlation between distances to 9+ year-old paved roads (extrusion heights) and Population classification (reddish colours)



5.2. Final evaluation of the analysis

The statistical correlation procedures confirmed the initial leads outlined by the scattergrams. PPM results show a clear tendency of the dependent variables (SE, Housing and Population classifications) to be associated with the independents (distances to all, 9+ year-old and 3+ year-old paved roads) in the majority of the correlations. The Table 10 and Table 11 below show the summary of results for all correlation procedure for each independent variable respectively. The higher PPM coefficients (0.593, 0.531 and even 0.470) of some correlations is a remarkable result for an analysis in social sciences. It has more importance regarding that 5033 sampling units were correlated. Having such a great coefficient with as much as the sampling used fosters the overall appreciation about the correlation found.

It is important to mention here that the number of observations play an important role in correlation and that mid-low coefficients are common and of high importance among social sciences (Griffith and Layne, 1999, Feng, 1996). Searching for the relationship of two variables with a few entries may yield a strong PPM coefficient, but it might be inconsistent and would also give each entry an extreme importance in the relationship. On the other hand, a considerable amount of values -such as the 5055 of this study- will always be more difficult to correlate. Thus, one of the most important issues in this is the p-value index, which makes the relationship significant. This means that the linearity, though low, is consistent and the values are corresponding with a slightly linear relationship. Consequently, in a large dataset such as the one used in the present study, a small PPM would turn significant if p-values are consistent with it. The principal purpose of the significance coefficient is to tell us how sure we can be when rejecting the null hypothesis of no-correlation. The test also supports the high correlation found and therefore prove the relationship with total confidence.

The lack of general correlation of the Population classification dependent variable is also noticeable. The only consistent correlation it has got, with the distances to 3+ year-old paved roads is positive, though it does not have a coefficient far from the zero value that defines a lack of association. The negative sign of the majority of the PPM coefficient also reflex the correlations' tendency, as both variables are inversely correlated. This means that as the distance to the paved roads get longer, the dependent variables' values tend to decrease in a linear way.

It is also evident a clear limit at the 1200 meter distance in this relationships with the SE classification variable. This limit represents the great part of the correlation result, where the values at higher distance show no correspondence with the correlation. Most of the correlations proved to be consistent at the 0.01 level in a two-tailed significance level, meaning that there is a 98% probability of randomly getting the same results (and therefore only 2% chance of not getting the result again). The complementary F test allowed rejecting the null hypothesis, strengthening the consistency of the results.

Table 10 : Summary of the correlation results for the distance to all roads independent variable

	Distance to all paved roads				Scattergram Output	Graphic Output	Relationship	Observations
	PPM	Significance/level	r²	F				
SE classification	-0.477	0.000/0.01	0.2275	1488.346	Linear correlation that shows consistency until around 1200 meter distance to paved roads	Values clearly associated between variables	Negative, consistent, null hypotheses rejected	High correlation coefficient for a Social Science study
Housing classification	-0.528	0.000/0.01	0.2788	1953.223	Linear correlation that shows consistency until around 1200 meter distance to paved roads	Values clearly associated between variables	Negative, consistent, null hypotheses rejected	High correlation coefficient for a Social Science study
Population classification	0.077	0.000/0.01	0.0059	30.13792	No Correlation, fitted smoother parallel to horizontal axis	Scattered values in the area, no association	Inexistent, eventhough null hypothesis is rejected	Almost zero PPM coefficient

Table 11 : Summary of the correlation results for the distance to 9+ year-old roads independent variable

	Distance to 9+ year-old paved roads				Scattergram Output	Graphic Output	Relationship	Observations
	PPM	Significance/level	r²	F				
SE classification	-0.263	0.000/0.01	0.0692	375.4827214	Linear correlation that shows consistency until around 1200 meter distance to paved roads	Values mainly associated between variables	Negative, consistent, null hypotheses rejected	Scattergram shows a curved pattern, only some linear tendency
Housing classification	-0.280	0.096/0.01	0.0784	429.8559028	Irregular shaped correlation (curved), shows some linearity	Values mainly associated between variables	Negative, consistent, null hypotheses rejected	Scattergram shows a mainly non linear relationship
Population classification	0.023	0.000/0.01	0.0005	2.674451785	No Correlation, fitted smoother parallel to horizontal axis	Scattered values in the area, no association	Inexistent, null hypothesis is not rejected	F value below the minimum required to reject the null hypothesis

5.3. Summary

This section has performed the analysis within a schema of thoroughly explaining each step taken. The use of an SMCE evaluation model yielded the SE classification for the study area based on the 48 variables collected in the fieldwork (see 4.1.5). The output produced the final SE classification and two partial outputs for the Housing and the Population branches of the SMCE evaluation. The three compose the group of dependent variables against which the independent were correlated. These were product of a network distance calculation for three scenarios: all paved roads, 3+ year-old paved roads and 9+ year-old paved roads. A network model was created and adequate procedures were followed to calculate and transfer the distances to a database table able to be used in the final correlation procedure. These correlations were performed using a stepwise schema. It started using histograms to test the required normality of each variable, followed by the production of scattergrams to initially see the correlations. These were followed by the PPM calculation where coefficients were output for each pair of independent and dependent variables. Although PPM results had their own significance test, a complementary F test was carried out to assure consistency of results by rejecting –or not- the null hypothesis. Finally, the correlations were mapped using three-dimensional graphics in order to visually inspect the likelihood of the results and made them understandable for the untrained viewers. Three tables with the summary of results were produced for each independent variable.

6. Discussion of Results

In this chapter, the results achieved in the previous will be analyzed to elucidate and interpret their true significance. This analysis use the qualitative data collected in the fieldwork in the form of interviews, study area characteristics (the so-called “ground truth”) and some literature related to the topic to contrast the magnitude of the results found in chapter 5 with the reality of the study area. The analysis comprises four points: first, the influence of the distance to PRI related to the SE classification will be assessed; second, the housing factors will be contrasted against the results with some partial correlations explained for such purpose; third, the geographical factors (slope and altitude of the sampling units) in the study area, which are relevant to the main objectives, are presented and compared; and fourth, a critical analysis of the benefits of the provision of adequate PRI, supported by theoretical issues. This is done in order to further develop the relationship between the SE classification and the PRI and clarify the findings.

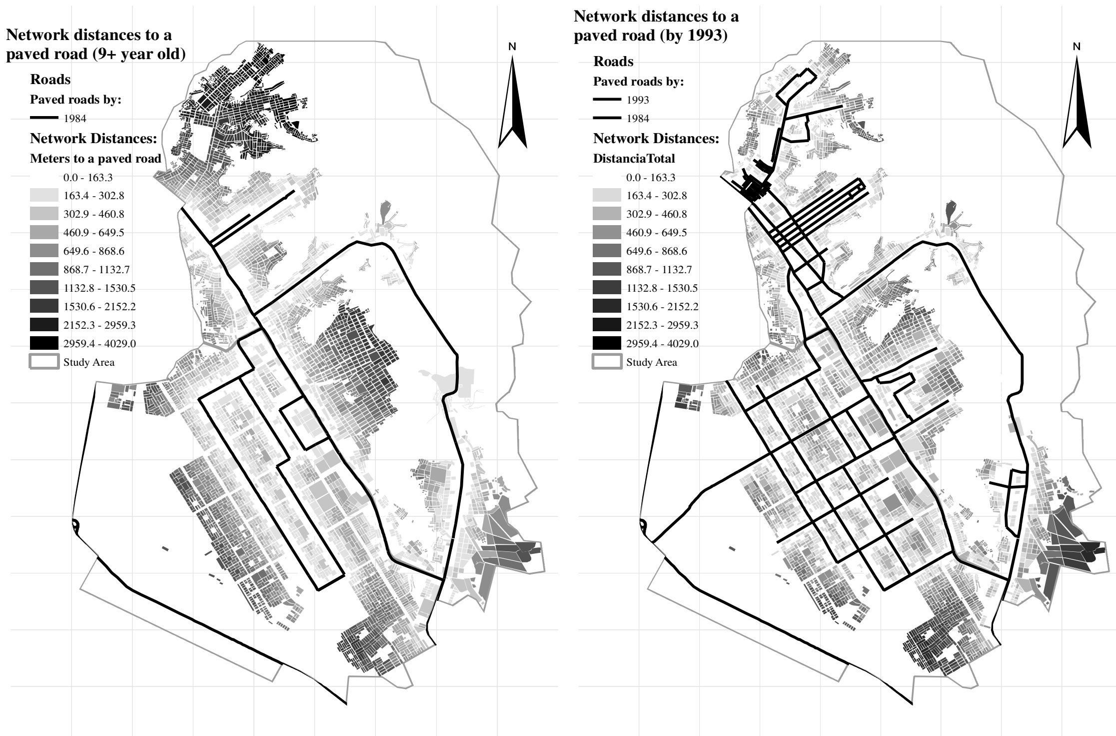
6.1. The importance of the distance to PRI

The data about the physical settings of the study area allow evidencing the differences between the two time-periods that comprise the present study. For the scenarios of 1984 and 1993 there are aerial photographs and road network data that give us a clear idea about the layout of both them. As an initial physical constraint, this layout may help to partially explain the meaning of the correlations found in the analysis performed in the previous chapter. They also may provide a better view about why these correlations have certain characteristics, such as the increase of the PPM coefficients for the 1984 to the 1993 scenarios.

As in an older timeframe, the study area’s road network of 1984 shows differences with the correspondent for 1993 (see Graphic 26 below). These differences are directly related with the quantity (total length) and location of paved roads in the two time periods. Looking at both layouts, it is evident that the 1984 scenario had much fewer paved roads. Having less paved roads, the blocks (sampling units) in 1984 are at a longer distance from /to a paved road in 1993, when more roads were paved and therefore more sampling units closer to these. Consequently, in 1984 the distances from the sampling units to a paved road are much longer than in year 1993. Even with some sort of SE development close to a paved road, the rest of the blocks are far from them and therefore expected to have lass the influence this study aims to prove (the correlation between being closer to a paved road and having a better SE index).

In this case, the distance factor could generate different living conditions for inhabitants, especially in urban poor areas where inhabitants have to invest more energy and time to get access to some basic services. This means, for instance, that the person who lives far away from a paved road, has to walk more time to access the nearest paved road in order to take public transport service and finally to arrive at the destination. If the same person lived near a paved road this time would be less. This may not only explain the high correlation found for the SE classification and the distances to paved roads, but also why the correlations are better in the 1993 scenario than in 1984.

Graphic 26 : Road network layout for the 1984 and 1993 scenarios



Having more paved roads covering more of the study area means less distance to them and there are more blocks (sampling units) close to them. Therefore much better correlations are expected if the correspondence of the two variables analyzed is adequate. However, it must be emphasized that having a better correlation does not depend on the number of units correlated but on the amount of them that are associated with their correspondent in the two variables. This means that it is more unlikely to have a better correlation when having more sampling units and if this happens the correlation is more consistent (and therefore hard to find). This is the case of the results found for the 1984 and 1993, with both correlations very high and furthermore consistent.

6.2. Explaining the importance of the Housing influence

The results show a greater influence of the Housing component than the Population correspondent in the SE classification. The following paragraphs will try to elucidate the extent of the housing and PRI relationship as well as provide an explanation for the magnitude of the relationship found between the Housing dependent variables with the independent represented by the distances to paved roads in the 1984 and 1993 scenarios.

6.2.1. The housing service factor in the study area

The results show the greatest influence in the final SE classification comes from the Housing stream. As said before, the correlation of this stream with the distances to paved roads, for the 1984 and 1993 scenarios are the highest achieved in the whole analysis process (see Table 1 above). This correspondence shows that it would be better to analyze the distances to paved roads in pure housing terms, rather than in an overall SE index as done in this study. However, that leads to another objectives and hypothesis different than the ones defined for the present study.

The relationship identified between the SE classification (or SE index) and the PRI (or Network distance to all paved roads) was found in others correlations which were calculated using others dependent variables like Access to services. This variable belongs to the Housing service classification and it was taken from the SE classification. The access to services is useful for assessing the connection to water and sewerage pipe network. The Housing service classification together with the network distance to all paved roads (in 1993) were correlated, proving to be more consistent and stronger than the main one.

The outlined characteristics are confirmed by these results. The two correlations are consistent and also have a negative sense. Nevertheless, the second correlation (-0.511) is stronger because its coefficient is higher than the main one (-0.477). Both results show that there is a clear relation between the respective variables. According to the key informants in the study area, this may mean that the lack of convenience for poor people who live far away from the paved road may be related to the SE classification. In addition, it is also appropriate to emphasize that the second correlation has the same tendency shown by the main correlation, that is to say: the greater the distance to a paved road the lower the SE classification. Consequently, the second correlation supports and strengthens this study; showing another way to prove the existence spatial relationships between the SE classification and Network distance to paved roads is direct related with an important development factor as the PRI.

The final explanation may refer to the close relationship between the level of development of the PRI and the quality of the housing. This is evident in the case study of Gujarat State in India, presented in the literature review (see 2.5 above). According to The Asian Development Bank (2004) the Gujarat case proved a close influence of PRI in the level of house services such as water drainage or electricity. The Asian Development Bank showed that those services were more available for the population that was closer to the roads than for the farther. As pointed out by the International Bank for Reconstruction and Development and the World Bank (1994) there is a clear cycle of PRI investments that lead to service provision due to the newly added access. The road infrastructure provides a sort of recognition of an urban area to be part of the city and furthermore being accessible trough its road network. It is undeniable that populations served by PRI are more eligible to receive private funding as their urban settlements are more identifiable with the city area than isolated new settlements. This financing often goes to the development of priority services such as water and drainage access. In addition, service and utility companies have a better layout for providing infrastructure for a given service once a base PRI setting is provided by the authorities. This also means the transport of heavy machinery and material is then possible and economic. These issues are further discussed in the next section.

6.2.2. The housing quality issue

At this point, it is necessary to add that there are other external factors that contribute to the socio economic development such as education, health, production, etc. But the way each one can improve the urban poor areas depends on the policies and decision making of the authorities in charge. Therefore, the external factors together with the provision of PRI could improve the socio-economic level in human settlements. One of these factors is related to the quality of the construction materials used by the population in the study area.

The benefits explained so far are related to the access to PRI in good condition meaning paved roads. In the study area, there are settlers who own properties which are built using good quality materials (i.e. houses with walls and roofs made of bricks and/or concrete) and moreover they have connection to piped water and sewerage supply**.** They may have those benefits due to the location of their properties. The settlers who live near or next to paved roads (i.e. avenues or important streets) have the opportunity to easily transport the materials and machinery for building such houses in less time and cost.

Table 12 : Results of PPM correlations for SE and Housing Service classifications dependent variables

	Total socio-economic classification	Housing service classification	Element description
Network distance to all paved roads (in 1993)	**-0.477****	**-0.511****	PPM Coeficient
	.000	.000	Significance (2-tailed)
	5055	5055	Number of sampling units
Network distance to 9+ year-old paved roads	-0.263**	-0.294**	PPM Coeficient
	.000	.000	Significance (2-tailed)
	5055	5055	Number of sampling units

**Correlation is significant at the 0.01 level (2-tailed).

Nevertheless, there is the possibility that some housing units were constructed before the provision of road infrastructure. In this case, this situation could be the result of an evolution process where each person improves the conditions of their properties little by little. That means, for instance, that they were converting their huts into more stable and strong houses, using better materials in a gradual way, maybe during decades. Although there is no proof to demonstrate this urban phenomenon, some study previously done in the study area could be useful to support and complement this idea. Some of these concepts are explained in the following paragraphs.

As mentioned before, the study area was a consequence of an invasion process. The inhabitants took empty land (regardless of public or private property) and each settler followed steps to satisfy the essential necessities of this informal acquisition of property. According to De Soto (1987), the invasion is an informal way to gain access to properties for living purpose. He also affirms that during (and after) the invasion process where the settler takes the land (which are previously divided into plots), informal comers do not take long time to appear. They seek to provide construction materials and all the staff required for building the first housing units. The same invaders are also organized to provide the necessary working force.

This is just an example that describes how poor people use creative solutions in order to survive in the middle of an isolated area. This reality reflects the constant struggle that people have to confront for improving their living conditions and socio economic level. It is obvious that construction materials and others can be transported more easily through a paved road. And this at its time may explain why the correlation of the Housing stream (composed of house quality, property and access to services) with the distance to paved roads is the highest achieved (see Table 1 above). Plots closer to paved roads are more prone to be easily built due to the access of construction materials and vice versa.

6.3. Benefits obtained through the PRI in good condition

The provision of adequate PRI plays an important role, producing benefits in any developing area. For instance, the PRI allows the physical connection among urban settlements, districts, etc. This connection allows poor areas to become more dynamic through the development of economic activities such as commerce and production. The relationship between PRI provision and increases in economic revenues was outlined in 2.3 in chapter 0. There, the correlation between PRI and the Gross Domestic Product increase is evidenced. The study area has experienced a crescent commercial activity that has been boosted by the incorporation of new roads in the late 80's and early 90's. This is the timeframe of the present study.

According to the key informants (see interview formats in Annex 1 and Annex 2 below), in the study area, commercial activity was generated when some families took advantage of the condition of the roads to convert their properties (or just one part of them) into stores, shops or places of business to increase the family earnings. In other words, the families build stores in their own houses because the properties are located near or in main avenues (paved roads) where the public and private vehicles move along. In this way, the transport moves people who become in the consumers, producing a commercial environment. In this environment, the commercial activity is complemented with the presence of suppliers which use the same paved roads to transport and provide their products. Therefore, the PRI makes the urban areas more dynamic and also produces benefits that not only can improve the quality of life but also decrease the poverty, especially in areas that are the result of an invasion process (illegal possession of land).

This situation is however influenced by the overall economic situation of the area where the activities and the provision of PRI take place. Regarding the situation of Lima in the late 80's, the economy had reached its worst point in the century, with inflation growing at a fast pace and the country completely isolated from the international funding stream due to foreign policies by the Peruvian government. Prices of basic goods rose almost daily and the recession was imminent. Although out of the scope of the study and this unable to be proved, the described situation may answer the low correlations of PRI and the population indicators.

6.4. The geography of the study area and the SE classification

The constant growth of human settlements in Villa Maria del Triunfo and Villa El Salvador districts is a typical characteristic of urban poverty areas. In this case, the accelerated urban sprawl is also a consequence of the invasion process. In both districts, the first inhabitants had the opportunity to select the most accessible places, located in flat areas and close to existing highways. The inaccessible and less adequate places to live (such as hilly areas) were populated after this core area was completely filled.

The geographic factors of the study area comprise the elevation and the slope. The former is defined as the height on which the block is located (relative to the closest paved road), whilst the latter is defined as the inclination of the access it (the inclination of the access to a paved road). Those two factors are accounted as different data types. For the study area, the elevation indicates the number of meters above sea level for each block or sampling unit whilst the slope factor indicates the inclination angle from the paved road which provides access to the sampling unit. This latter factor is normally used as a measure to design PRI in terms of the adequate surface (maximum slope) for heavy transport.

Due to the location and the geographic condition of the terrain, the most accessible areas were first occupied, whereas the remaining areas on the less accessible land were populated little by little. Therefore, the older settlers, better located and more able to have a better SE status by being established long ago, are supposed to show a better correspondence with higher SE classifications. Despite this plausible situation, the elevation has no correspondence on the socio economic classification. As explained before, we seek to find a pattern in the location of new settlers, later invaders with –according to De Soto (1987)- only the minimum living conditions, waiting for recognition and just establishing their survival network. As newly arrived to the study area, their socio-economic development would have been lower than the rest, settled long ago and with far more opportunities to improve their economic situation. However, the study did not find a negative correlation between the SE classification and the elevation of the sampling units (see Table 13 below). In the case of the Elevation factor, the lack of correlation is confirmed by the PPM result. The coefficient found is close to zero, meaning there is no correlation (see 5.1.4 above).

Table 13 : Results of PPM correlations between SE classification and the geographical factors

	Total socio-economic classification	Element description
Elevation (meters over sea level)	-0.055**	PPM Coeficient
	.000	Significance (2-tailed)
	5055	Number of sampling units
Slope from paved road (in %)	-0.053**	PPM Coeficient
	.000	Significance (2-tailed)
	5055	Number of sampling units

**Correlation is significant at the 0.01 level (2-tailed).

The same tendency is also shown by the slope factor. The slope is a measure closely related to the PRI and it is also direct result of the distance. As the correlations between the latter and the SE classification are consistent and appreciable (see Table 10 and Table 11 above) it was expected to find a good negative correlation between the slope and the SE classification. However, it is not the case and therefore the slope variable does not have relation with the SE classification. It is evident that the two mentioned geographic factors have no effect on the distribution of the range values of the SE classification of the study area.

The explanation of the lack of correlation in both cases ranges from issues with the data type to the layout of the study area. Whilst elevation is a constant value that changes in a more continuous way, the SE classification, as a result of a mathematical calculus, has completely discrete values. Therefore when correlating both types of values (continuous and discrete) a lack of association is foreseeable. This is also supported by the fact that the elevation where a block is located does not necessarily imply lack of or poor accessibility (which is not the case of the distance to a paved road that does determine easiness of access) and therefore does not mean lack of opportunities for development of improvement or any kind of disadvantageous situation.

In the case of the slope the value has the distance to paved roads as a component (remembering that the slope comes from an equation that defines the angle between the distance to a paved road and the elevation of the block) but it is influenced by the elevation to obtain the inclination. At the end it does not define access based on how far is the block from a paved road but just the impedance caused by the its inclination. In a simple sense it overrides the measure of distance, becoming a middle measure between both indicators and therefore has both influences. Nevertheless the no correlation found and the similarity of the coefficient with the elevation correlation (0.055 and 0.053) point towards the elimination of the distance factor and a closeness with purely elevation values. In this case the same explanation given for the elevation is applicable to the slope.

6.5. The dynamic of the population variables

Most of the results discussed so far, are related to correlations between housing variables and network distance to paved roads. However, there are other correlations that present interesting outcomes. They are dependent population variables included in the population classification, which belong to the SE classification (see the Graphic 11 above).

As seen in the Table 13 above, the two correlations show similar characteristics, both maintain a negative tendency, having PPM coefficients with low magnitude. In the case of the economically active population classification, the coefficients found are close to zero, meaning there is no correlation with the two independent variables. On the other hand, the coefficients obtained from the correlation between the Population density and the independent variables indicate that there is a weak relationship between them that shows significance.

Having such low coefficients (-0.210 and -0.168, though significant at the 0.01 level) could be attributed to the dynamic of the population density. The rate of changes of the population (may have considerable changes per weeks or months) is more constant than the changes due to the implementation of the infrastructure (changes only noticeable in months or years). Therefore the population density variations are measurable in time periods shorter than the allocated for the present study. Changes in population numbers within urban areas could be directly related to the constant growth as well as the shift of people from one area to another inside the city, due to work or change of dwelling. However, the increase of both coefficients from 1984 (0.168) to 1993 (0.210) may indicate that the population variables have certain influence over the socio-economic level in a progressive and continuous way. This may also be only the result of a normal growth of the population during the 9-year period between both scenarios.

Table 14 : Results of PPM correlations for each Population dependent variable

	Economically active population classification	Population density classification	Element description
Network distance to all paved roads (in 1993)	-0.003 .813 5055	-0.210** .000 5055	PPM Coeficient Significance (2-tailed) Number of sampling units
Network distance to 9+ year-old paved roads	-0.042** .003 5055	-0.168** .000 5055	PPM Coeficient Significance (2-tailed) Number of sampling units

**Correlation is significant at the 0.01 level (2-tailed).

Correlating infrastructure (such as PRI) with population variables may show lower correspondence because of the differences in the dynamics involved, though it can not be neglected there is some influence that was evidenced in the present study. Nevertheless, other areas with far more dramatic changes in shorter time periods may not present the same situation. It is clear so far that the housing variables have much more influence over the socio-economic level than the population variables. Due to the slower pace of change of the housing variables (i.e. the improvement of the physical condition of the houses or the implementation of housing services such as water or electricity), it could be more convenient to compare these variables directly with the PRI that have the same evolution characteristics. However, this does not mean that the population variables are less important for this study. The main hypothesis is to establish a correspondence between distances to paved roads and a SE classification that must include population variables per se. The SE classification may not depend purely on the characteristics of the housing units, which can be considered only a physical rather than a social analysis such the present.

6.6. Summary

This chapter has addressed the importance of the results obtained in the chapter 5 above in terms of several factors that may yield an explanation for them. Among them, the importance of the road infrastructure, the geographical configuration of the study area, and the SE classification. A critical analysis was carried out in order to provide an answer to the weak, or lack of correlations found in the population variables as well as in the geographical factors. The distance from the PRI to all sampling units is not only an important variable but also is a factor that can generate a negative social condition for people who live far away from a PRI or paved roads. This situation of distance, demands more cost, energy and time of the people, and the SE condition of poor people as measured by the SE classification, is affected. In the same way, there are other variables used to perform new correlations for proving the consistent relation between the SE classification and the PRI. However, most correlations showed results with low PPM (close to zero), meaning that the variables used are not related to the SE classification. The variables analyzed were: the two geographic factors and economically active population classification. The geographic factors refer to the terrain configuration where the sampling units are located. Both factors have no effect on the range values of the SE classification. Therefore, the socio-economic level in the study area does not depend on the geographical characteristics like the elevation and slope areas.

Nevertheless, another correlation done with the Housing Service classifications showed a high coefficient, proving that the correlations with housing variables are more relevant because they provide more consistency to the spatial relationship between the SE classification and the PRI.

7. Conclusions and Recommendations

The conclusions of this study are presented in this chapter. First, they are outlined from the results obtained and analyzed in the chapters 5 and 6 respectively. The recommendations contribute with guidelines for the decision making and policies that may lead towards poor urban areas development having some influence on major infrastructure projects, such as the provision of PRI. In that sense it is necessary a reference to the socio-economic indicators that play an important part in the study of human settlements. Finally, further research topics are inferred from the study and the spatial relationship between the Public Road Infrastructure (PRI) and the SE classification.

7.1. Conclusions

The results together with the issues established and discussed in the first, second and latter chapters, will be presented in order to draw conclusions. The results of the analysis will be used, making a distinction between main results (including the two main variables PRI and the SE classification) and others complementary results. In the same way, the objectives, hypotheses and questions are finally confronted and answered, explaining briefly each achievements across the methodology used in the present study.

7.1.1. Conclusions to the main objective

Main Objective: to determine if the PRI is related to the SE classification in the study area.
This study evidenced the relationship between the PRI and the Socio-economic level in the poorest area of Lima city. A high PPM coefficient of 0.477 proved that the relation between the mentioned variables is sufficiently strong to demonstrate that the PRI has influence over the socio-economic development of the urban areas it serves. In addition, this main correlation is consistent and significant at the 0.01 level in a two tail analysis (both sides of the probability distribution) meaning that there is a 98% confidence that these results are not produced by chance. If another random analysis is performed with the same sampling units there is only a 2% probability of not having the same results. This gives consistency to the results and supports the correlations found (see Table 10 and Table 11 in chapter 5).

7.1.2. Answers to the research questions

The schema of the study is designed to obtain answers for the research questions which are associated with proposed sub objectives. The research questions per sub-objective are taken from Table 1 in the chapter 1.

Sub-objective: To produce a SE classification for the study area

Which indicators are used to determine the SE level of the area?

How are these indicators combined to produce a SE classification?

The elaboration of the SE classification as well as the analysis of the socio-economic level or condition of poor urban areas required the use of key urban poverty indicators. These indicators facilitate the analysis, monitoring and comparisons of urban poverty in human settlements like Villa El Salvador and Villa Maria del Triunfo. The housing property, construction material, access to services, the population density and the economically active population were identified and used in the analysis.

Besides this, the development of PRI as the independent variable was also necessary. For such, other key urban poverty indicator related to the physical condition of the road was required. It refers to the paved roads as the surface quality provided.

Sub-objective: To assess the degree of relationship between distance to PRI and the classification of SE indicators:

> **Is there any relationship between closeness to PRI and the higher SE classification?**
> **Does the PRI surface quality influence the spatial distribution of SE indicators? If so: Is this influence positive or negative? Is this relationship consistent?**

The study proved a negative consistent relationship between being closer to a paved road and having a higher SE index based on the variables selected for such. The conclusion is supported by the high PPM coefficients (0.477) and found and the amount of sampling units used (5055). This means in a few words that the longer the distance to a paved road the lower the SE index of the sampling unit in the study area. The PRI quality in terms of paved or not shows a correspondence with the SE classification that may suggest the presence of an influence. However, the latter could not be proved by the present study as more data and a time series research schema would be needed.

Sub-objective: To evaluate the access to the rest of the city by existing PRI and its incidence in the SE development:

> **Is accessibility an issue in the SE classification of the study area?**
> **Is the access to opportunities and services given by PRI determinant to the SE classification of the study area?**

The final questions have an answer in the theoretical framework developed in chapter 2. In general terms, this study established that the presence of PRI means better accessibility and therefore better opportunities. This accessibility is accounted in terms of farness or closeness to the city's road network, namely the paved roads. It is a sort of economic infrastructure provided through municipal intervention which include the correspondent costs to invest in such infrastructure. This Public infrastructure and the access it provides has a neuralgic and positive role in the socio-economic development in poor urban areas which is generated by the benefits produced by the provision of adequate PRI denominated paved roads. For this study, the benefits refer to access to jobs, improvement productivity and the economic growth. In such cases, PRI is directly providing the right access for each one and the economic benefit. In the social aspect, an adequate PRI contributes to the access to all centres of attraction as well as the easy implementation of basic services such as piped water supply, sewerage and electricity, transport, etc. especially in inaccessible slope areas. A sustainable growth of poor urban areas depends -among other factors- on the access to productivity, economic interchange and services through the provision good quality of PRI (paved roads) which connect them with the existing road network and thus the rest of the City.

7.1.3. Achievement of the main hypothesis

Main hypothesis:

> **The SE classification is related to the distance from each sampling unit to the PRI in the study area.**

Through the consistency (allowing rejecting the null hypothesis) of the results found, the study can sustain the main hypothesis. The theoretical context the SE classification, defined as an overall index based on key urban poverty indicators has consistency in the methodology applied. This dependent variable has indeed a relation with the distance to the PRI in terms that the study has proved a higher

SE index corresponds with closeness to PRI in a limited way. The percentage of the values that account for that correspondence is defined by the r2 value of the PPM correlations, roughly the 25% of all values measured. With a total of 5055 sampling units the result is remarkable and consistent for a social science. Therefore, the distance is not only an important factor but also it plays as a link between the SE classification and the blocks in the study area from where adequate PRI must be provided by the authorities in order to contribute to the growth and development of poor urban areas.

7.1.4. Achievement of the research hypothesis

Research hypothesis:

> **The PRI configuration has direct influence on the SE classification of the area it serves.**
> **The distance to PRI and SE classification have a negatively correlated relationship.**

The PRI configuration refers to the physical condition of the road surface, different quality of the construction materials used to build infrastructure. The direct influence was verified through the statistical procedure used, therefore the hypothesis is validated. The use of two scenarios, for 1984 and 1993 allowed determining the influence of PRI on the SE classification. The two coefficients show a growing tendency from one scenario to the other in a sort of progression. The served area in both cases is also correspondent as the PRI in 1984 covered far leas area that the correspondent for 1993. Analysis performed of the PRI surface quality was required in order to reach this important research hypothesis.

7.1.5. Achievement of the methodology hypothesis

Methodology hypothesis:

> **The PRI quality and the SE indicators are identifiable, measurable and can be analyzed in a spatial/statistical correlation.**

The variables and indicators for the analysis were identified using the literature review and the information provided by primary data taken on the field. The indicators used were divided in two streams, namely the housing and population as census data includes both in the study area. For housing, the indicators of location, property, construction material and access to services were identified. For population the density and the economically active population were used. All them were able to be measured in terms of housing units with the indicator inside each sampling unit or block. The PRI variable was defined using aerial photographs and used the type of road surface as an indicator, measured as paved or unpaved. From them, distances were calculated to be used as the indicators for the independent variable. The methodology comprised all indicators inside SMCE and statistical approaches performed jointly or within a Geographic Information System. Therefore, the analysis was carried out in a spatial environment and the main analysis was performed using statistical tools. All correlations needed for probing relationships as well as for proving or rejecting the null hypothesis were achieved with the use of PPM as a statistical tool.

7.1.6. General conclusions

The study achieved the objectives and proved the hypothesis drawn in its design, correct as it smoothly allowed the performance of the research. The approach was therefore the adequate for the study carried out. The methodology used was effective for the purpose of fulfilling the research requirements. A SMCE model correctly determined a consistent SE classification for the study area. This procedure is identified as an approach for elaborating poverty maps in an effective way, using census and spatial data together. Using statistical procedures also evidenced great correspondence with the objectives, hypothesis and questions enunciated in the first chapter of this research. PPM procedure allowed finding the correlations that support this research and the answers to all inquiries placed before the analysis.

The research carried out duly completed the assessment of the spatial relationship (determined by the distance) between PRI and a SE classification based on key urban poverty indicators in the study area. These indicators, grouped in housing and population streams, were adequate for the research proposed, in terms of identification, measurability and collection of the data related. The complementary use of key informants enriched the study by adding the qualitative view over the quantitative results of the analysis performed.

7.2. Recommendations

An adequate identification of urban poverty indicators and the factors that influence it proved to be of importance. Some techniques applied to determine and visualize the spatial relationship between the SE classification and the PRI are suggested to planners in order to contribute with the development of human settlements. Therefore, some recommendations are given in the following statements.

The first recommendation is addressed to the authorities and the people in charge of infrastructure provision in the study area. The research showed the correspondence between roads infrastructure and a better setting of the socio-economic situation in terms of housing and population has a limited distance. This distance limit is situated in the range of 1200 meters, from which no housing unit may be situated farther. It is important to establish a plan for the provision of road infrastructure with that range in mind. All the dwellings in these urban poverty areas that are out of that range may be considered in priority for the provision of new roads if not connected to it as soon as possible.

As seen before, the SE classification is the result of the combination of urban poverty indicators. It can be used as a tool for the local authorities for getting to know the way to improve the socio economic conditions of the human settlements. For such purpose, the SE classification can be used to design poverty maps, with which urban planners and authorities in general may have a flavour about the living condition in a given area. In this way, experts can take decisions and develop other urban projects related with the provision of services or facilities such as schools or hospitals.

In the same way, the methodology proposed for this study may be useful to identify which areas are socio-economically benefited by the provision of adequate PRI. Among other reasons, this information is essential when the municipalities must make decisions and evaluate options for infrastructure implementation, particularly roads. Due to the local authorities having a limited budget to build paved roads in each urban area (the whole district), the urban planners can use the technique to identify benefited areas in different degrees, having the opportunity to visualize which areas need to be serviced as a priority.

7.2.1. Further research

These last paragraphs outline other possible research topics evidenced form the present study. Further research ideas are to be covered as they are considered as relevant as the research carried out.

It is important to conduct more research in the relationship between infrastructure in general and its influence on the socio-economic variables that characterize urban poverty. There has been identified a lack of studies in the topic, being the few existent done in the general terms of the influence. Coping urban poverty depend –among other factors- of the correct and adequate provision of infrastructure that foster opportunities and enhances economic potential where it may be inexistent.

The use of SMCE for processing variables in an ordered way had no associated literature that could be found at this time. It is recommended that this procedure may be applied as a methodology for processing huge amounts of data from large sets indicators. The ordered way it processes the information and the easy changeable settings may provide the right tool for carrying out further research in many scientific aspects. In this sense it is advisable to evaluate the possibility of producing poverty maps using the SMCE environment such as provided by ILWIS.

As mentioned before, during the statistical correlation procedures, the PPM coefficients showed a strong tendency for the dependent variables housing to be associated with the PRI as an independent variable, in the majority of the correlations. Therefore, it should be important to deepen the research in that sense, proposing studies that may better evidence the strong relationship found.

It may also be of importance to analyse other variables pertaining the socio-economic characteristics of urban poverty areas. Among others, the family composition, the access to education or health services may give light about the setup of those areas and thus more ways for improving their condition. In that sense it is important to mention the work carried out by worldwide institutions such as the International Bank for Reconstruction and Development or the World Bank in identifying such indicators.

7.3. Summary

This chapter summarizes the conclusions by reflecting back to the objectives, hypotheses and questions designed at the beginning of this study. All them are properly addressed through the results of the analysis and their interpretation, done in the two previous chapters. The study has proved the main objective and main hypothesis. For the main objective, this means that the PRI is related to the SE classification whilst for the main hypothesis the analysis established that people who live far from the paved road (PRI) have a lower SE condition based on key urban poverty indicators. In a few words, this can be expressed as the longer the distance of a block to a paved road the lower the SE classification it gets. In addition, a general conclusion was presented, defining the major achievements of the study in terms of correspondence among variables analyzed through the proposed methodology. It turned to be adequate for achieving the goals of the present study. The results are consistent with the initial objectives, hypothesis and questions enunciated at the beginning of this study. Based on the research performed together with the theoretical framework and defined techniques, there are recommendations and further research topics, aiming to improve the conditions in urban poverty areas. The proposed methodology allowed such task and proved to be efficient when assessing the influence of PRI over the SE classification of the study area.

References

The International Bank for Reconstruction and Development/The World Bank (1994). World Development Report - Infrastructure for Development. New York, The International Bank for Reconstruction and Development/The World Bank.

United Nations Centre for Human Settlements (Habitat) (1997). Urban Indicators Programme Phase One: 1994 - 1996. Nairobi, United Nations Centre for Human Settlements (Habitat).

The International Bank for Reconstruction and Development/The World Bank (2003). World Development Indicators. Washington, The International Bank for Reconstruction and Development/The World Bank.

Asian Development Bank - The Transport and Communications Division of the South Asian Department (2004). Assessing the Impact of Transport and Energy Infrastructure on Poverty Reduction Asian Development Bank - The Transport and Communications Division of the South Asian Department.

Blalock, H. M. (1979). Social Statistics. Singapore, McGraw-Hill International Editions.

Bond, P. (1999). Basic infrastructure for socio-economic development, environmental protection and geographical desegregation: South Africa's unmet challenge. Geoforum 30(1): 43-59.

Brussel, M. (2004). GIS in Urban Infrastructure. Enschede, ITC.

Cotton, A. and Franceys, R. (1994). Infrastructure for the urban poor. Policy and planning issues. Cities 11(1): 15-24.

De Soto, H. (1987). El Otro Sendero. Lima, Editorial La Oveja Negra Ltda.

DESCO (2003). Mapa social - parque metropolitano. Lima, DESCO programa urbano.

Fan, S. et al. (2003). Agricultural Research and Urban Poverty: The Case of China. World Development 31(4): 733-741.

Feng, H. M. (1996) Spatial Statistical Modelling of Explicit Determinants of Regional Fertility Rates: A Case Study of He-Nan Province, China. In Arlinghaus, S. L. (Ed.), Practical Handbook of Spatial Statistics, pp. USA, CRC Press.

Griffith, D. A. and Layne, L. J. (1999). A Casebook for Spatial Statistical Data Analysis. New York, Oxford, Oxford University Press.

Guillen, H. and Konings, K. (1992) Urban poverty and local policy responses: General aspects and the case of San Pedro Sula, Honduras. In Kleinrensink, J. (Ed.), Urban poverty alleviation in Latin America, Vol. 3, pp. 49-57. The Hague, Development Cooperation Information Department, Dutch Ministry of Foreign Affairs.

INEIOficina de Impresiones de la Secretaria General del INEI (1981). VIII Censo Nacional de Población y III Censo Nacional de Vivienda. Lima, Oficina de Impresiones de la Secretaria General del INEI.

INEI (2001). Almanaque de Lima y Callao 2001. Lima, Oficina de Impresiones de la Secretaria General del INEI.

Kaufmann, J. and Steudler, D. (1998). Cadastre 2014, A Vision for a Future Cadastral System. FIG,

May, H. (2001). Urban Poverty Alleviation. Urban Development 33(1): 56-78.

Moser, C. et al. UNDP/UNCHS (Habitat)/World Bank, Urban Management Program UMP (1996). Urban Poverty research sourcebook: Module I. Sub city level household survey Module II. Indicators of urban poverty (Working paper). Washington, UNDP/UNCHS (Habitat)/World Bank, Urban Management Program UMP.

Nachmias, C. F. and Nachmias, D. (1997). Research Methods in the Social Science. London, Arnold.

Openshaw, S. (1983) Multivariate analysis of census data: the classification data. In Rhind, D. (Ed.), A Census User's Handbook, pp. 243-263. London & New York, Methuen & Co.

Pacheco, J. F. (2004). Mass-Rapid Transit Systems and Urban Land: Spatial Interaction and Impacts. The case study of LRT-1, Metro Manila, Philippines. Master of Science thesis (unpublished). Enschede, International Institute for Geo-Information Science and Earth Observation (ITC) (available at http://www.itc.nl/library/Papers_2004/msc/upla/pacheco.pdf).

StatSoft (2004), Electronic Statistics Textbook. Retrieved from http://www.statsoft.com/textbook/stathome.html. in January 15, 2004 StatSoft, Inc.

Wasike, W. (2001), Road infrastructure policies in Kenya: historical trends and current challenges. Retrieved from http://www.kippra.org in July 15, 2005.

Wu, W. (1999). Reforming China's Institutional Environment for Urban Infrastructure Provision. Urban Studies 36(13): 2263-2282.

Annexes

Annex 1: Questionnaire A - Designed semi-structured interview for key informants

Institution Name:……………………………………………………………….

Address/Tel.: ………………………………………………………………………

Date:……………………

Questions to the respondent

What is your full name and position?

What is your experience in public administration? For how long?

Are you familiar with the formation process of human settlements in the 80´s?
If yes:
What was your position then? (professional, authority, community leader or other category)
For how long?
Did you have active participation in that process?
If no:
Do you have any idea about it? (go to question set alternative 2)

How those human settlements improved living condition in the past 20 years?
If so,
Which improvements: social organization, public infrastructure, titling, others?

---------------------END OF THE FIRST PART OF THE INTERVIEW-----------------------

ALTERNATIVE 1:

Based on your experience, which basic infrastructure were required for those poor settlements after their implementation?

How the majority of inhabitants of these human settlements had access to water?

Do you know if they implemented paths or other kind of road infrastructure themselves?
If so,
When and how?
Did they receive financial support from private sector and/or public sector and/or another organization?

After the implementation of human settlements in the 80´s, did they have access to paved roads, as well as stairways (for hilliest areas) and sidewalks?
If so,
When the mentioned infrastructure was provided?

Which are the human settlements that have reached great social-economic development? Where are they located? (Use the maps).

-----------------------------------END OF THE INTERVIEW-------------------------------------

ALTERNATIVE 2:

Are you acquainted with the formation process of human settlements?

Do you think that the urban area shown has changed over the past 20 years in terms of public infrastructure, services, social organization?

If so,

Which changes have you noticed?

Where are they located? (Use the maps).

What do you think that has influenced those changes:

Authorities' initiative?

Local impulse?

Central government backing?

Private sector backing?

International funding?

-----------------------------------END OF THE INTERVIEW-------------------------------------

Annex 2: Questionnaire B - Designed semi-structured interview for key informants

Institution Name:…………………………………………………………………….

Address/Tel.: …………………………………………………………………...

Date:……………………

Questions to the respondent

What is your full name and position? (relation to the urban/transport planning field)

What is your experience in the (previously named) field? For how long?

(a) What is your relationship with urban planning:

as a professional/authority?

Have you been involved in the provision of public road infrastructure in different sectors of this district or/and public work (pedestrian paths, paved road, stairs, etc.)? If so, For how long?; Which aspects of planning?

as a user?

Have you visited the sectors (areas)? If so, For how long?; How often? Which areas did you use to/usually go?

How do you describe the area in terms of development?

How have human settlements improved their socio-economic condition in the past 20 years?

If so,

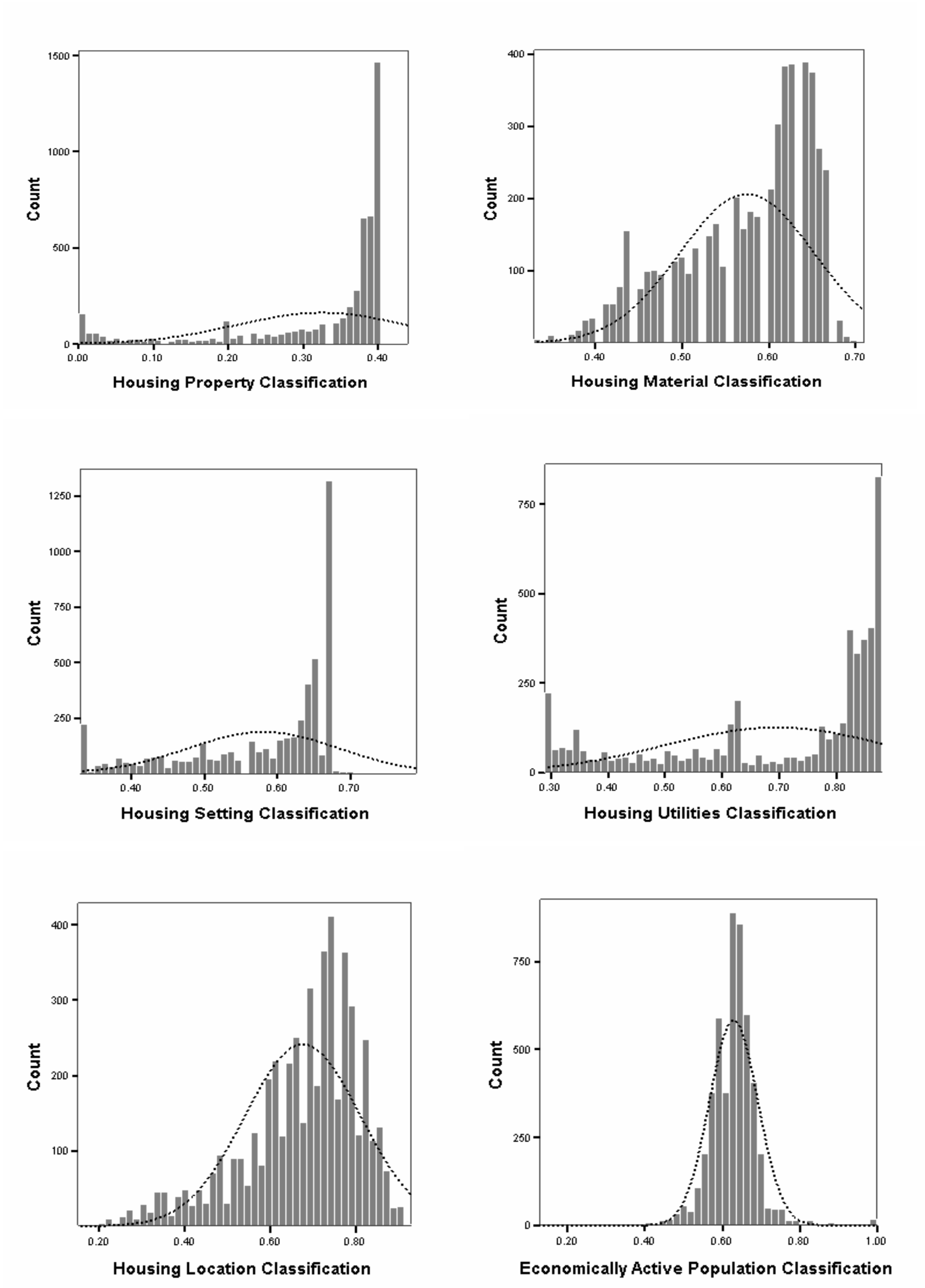
Which improvements: social organization, public infrastructure, titling, others?
Which changes have you noticed? Where are they located? (Use the maps)

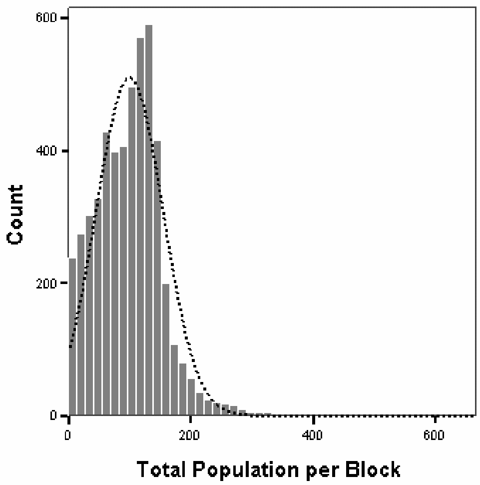
How can the socio-economic condition in poor urban areas be measured?

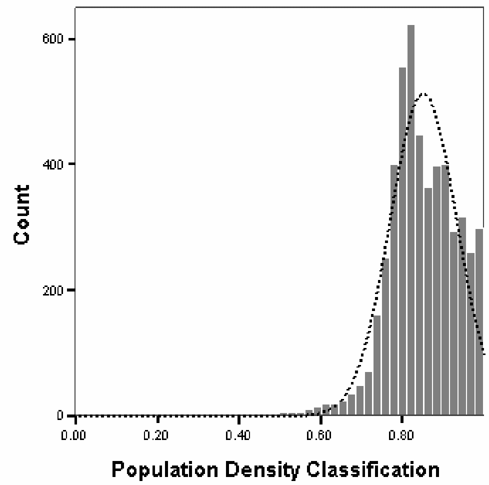
What do you think that has been influencing those changes?

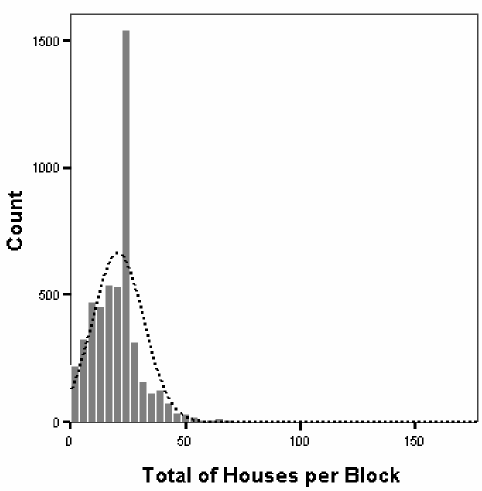
------------------------------------END OF THE INTERVIEW--------------------------------------

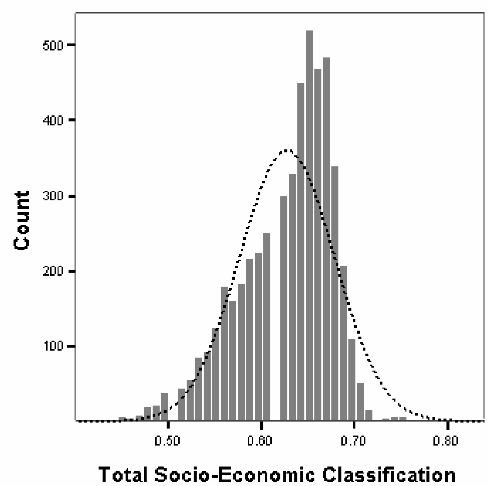
Annex 3: Rest of histograms for the variables used

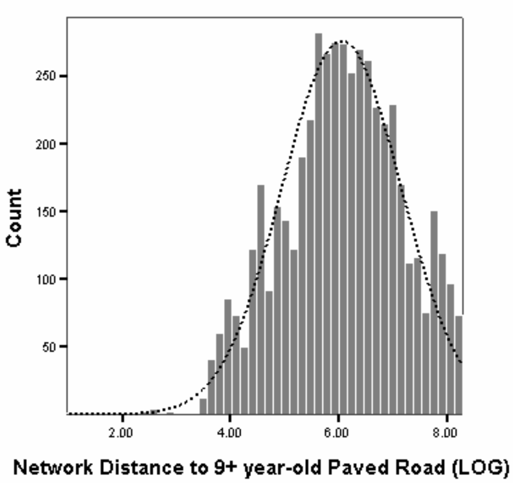



Count
600
400
200
0
0
200
400
600
Total Population per Block

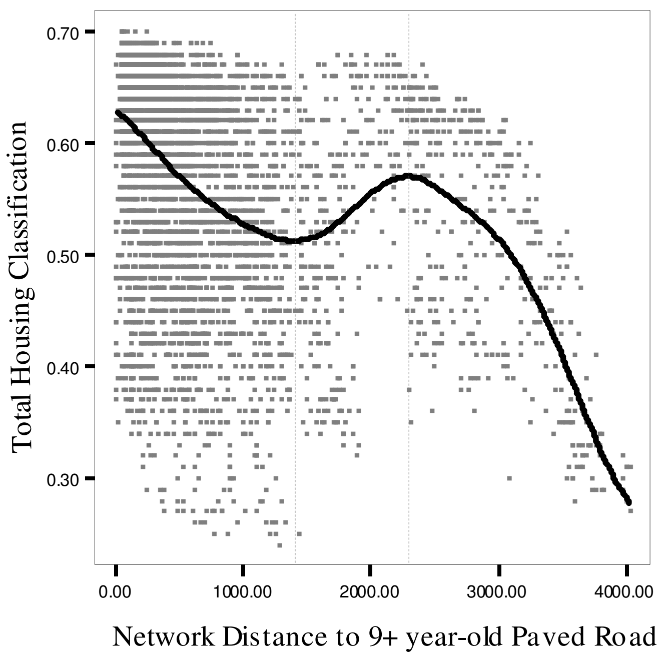

Count
600
400
200
0
0.00
0.20
0.40
0.60
0.80
Population Density Classification

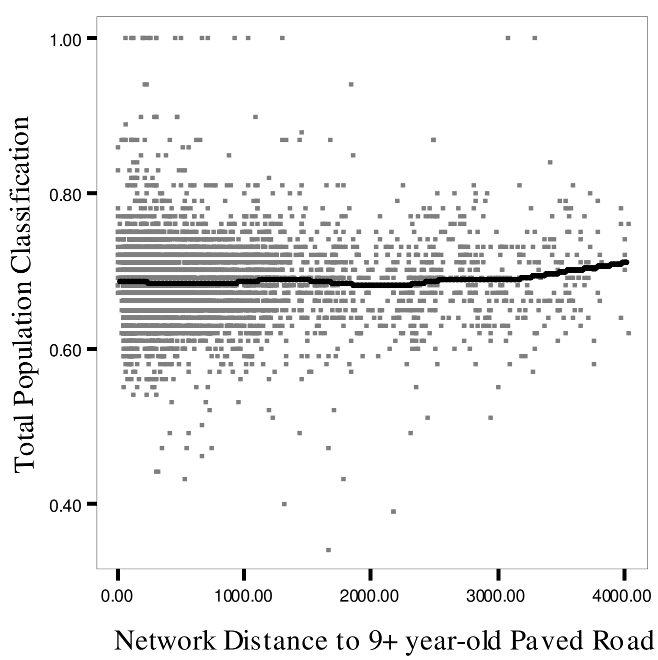

Count
1500
1000
500
0
0
50
100
150
Total of Houses per Block


Count
500
400
300
200
100
0.50
0.60
0.70
0.80
Total Socio-Economic Classification


Count
250
200
150
100
50
2.00
4.00
6.00
8.00
Network Distance to 9+ year-old Paved Road (LOG)

Annex 4 : Rest of scattergrams for correlations





Annex 5 : Tables of F Values for 0.05 and 0.01 probabilities

Table 1: probability α = 0.05

df2/df1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	INF
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
inf	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Table 2: probability α = 0.01

df2/df1	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	12.00	15.00	20.00	24.00	30.00	40.00	60.00	120.00	INF
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32	6157.29	6208.73	6234.63	6260.65	6286.78	6313.03	6339.39	6365.86
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.51	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.93	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.67	3.59	3.51	3.43	3.34	3.26	3.17
14	8.86	6.52	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.90	3.81	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.85	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.84	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.02	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.93	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.70	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.77	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.82	2.66	2.59	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.05	3.73	3.50	3.33	3.20	3.09	3.01	2.87	2.73	2.57	2.50	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.67	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.81
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.04	1.95	1.86	1.76	1.66	1.53	1.38
inf	6.64	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.19	2.04	1.88	1.79	1.70	1.59	1.47	1.33	1.00

Source: *Electronic Statistics Textbook*, StatSoft, Inc. (2004).