ARE E-BIKE CYCLISTS IN RURAL AREAS WILLING TO CYCLE LONGER DISTANCES TO LESS DIVERSE GOALS?

Analysing the data of the Bicycle Stimulation Program inNorth Brabant, the Netherlands.

Utrecht University

Master thesis
Didde Keck
Supervisor: Simon Scheider
Department: Geosciences
Master: Urban Geography
Student number: 6034322
Email: d.keck@students.uu.nl

Preface

With this preface I would like to welcome you to my thesis on e-bike usage in North Brabant. The dataset on e-bike usage was provided by the University Utrecht and was one of the subjects to choose from to write your thesis on. The thesis is part of the Master's program and a requirement to fulfil the Master Urban Geography at the University Utrecht (UU). I started this research in February and ended in April 2018.

The main objective to choose this subject was to learn more on Python scripting and to gain more information on the differences between rural and urban e-bike riders. As someone who is born and raised in Amsterdam, cycling has always been a part of my life since I was a little girl. This didn't seem to be the case for my friends who grew up in more rural areas, as soon as they could get their drivers' license, they would. This difference intrigued me and when I got the opportunity to work with this e-bike data I knew quickly that I wanted to research if there is an actual difference.

The audience for this thesis is in general for everyone who is interested in e-bike usage and the differences between urban and rural areas, but mainly for policy makers who want to make their country, city or village more e-bike proof. The aim is to see if e-bike users cycle differently in urban and rural areas and whether this difference lays in the trip length and/or the purpose in trips. The thesis also contributes to the debate about a more environmental friendly world and what the role of bicycles and e-bike is in this.

I would like to take this opportunity to firstly thank my supervisor Simon Scheider, not only for his honest feedback but also for the practical and emotional support. Secondly, I would like to thank Lian and Lisanne, both students of the master, with whom I discussed the different angles and research questions I could choose. Thirdly, I would like to thank Jens who helped me answering questions about the Python scripts. I also would like to thank the second reader for taking the time to review my thesis and Joost Kruijff for getting background information on the dataset. At last I would like to thank my family, boyfriend and friends for always supporting me and helping me wherever they could. Without all of you this thesis could not been made. I hope you'll enjoy reading my thesis.

Didde Keck
Amsterdam, 27-12-2017

Content

1. Introduction

Nowadays there are more and more policies that encourage cycling and discourage cars. One example is the New Urbanism in the USA or the Compact City Policy in Europe. They *"aim at reducing car use and travel distances because high density and mixed-use neighbourhoods are believed to be associated with shorter trips and more non-motorized trips*" (van Acker et al., 2010). These policies come up because the auto-dependent city harms the environment and decreases natural resources rapidly, congestion is even becoming such a huge problem that it costs 12 trillion dollars per year globally (Rybarczy & Wu, 2010). Cycling, however, is considered a healthy, environmental friendly and cheap mode of transport. There are two options to encourage bicycle use either improve the attractiveness of cycling or make the competitive mode – in this case motorized vehicles – less attractive (Rietveld & Daniel, 2004). During this research, the focus will be mainly on the first option.

One way to make cycling more attractive is to encourage the use of e-bikes. The disadvantages of cycling – greater physical effort, weather dependent, being more slowly than motorized vehicles outside high density areas – are becoming less of a problem with an e-bike (Heinen et al., 2010). To see whether this is actually true, it is necessary to have an insight how the e-bike is used in daily life. This can be achieved by looking at the Bicycle Stimulation Program Brabant (BSP). The BSP gathered data through 581 participants. These participants used the app B-Riders to follow their cycling trips and the purpose of these trips. The program followed the participants from September 2013 until October 2014. The participants were paid to keep track of their trips and signed in for the program themselves.

The Netherlands is an interesting case study because it is the leading country in the industrialized world in terms of bicycle usage. On average 27% of all trips are made by bicycle (Martens, 2004). It is unknown whether this is also the case with e-bikes, but together with Germany, the Netherlands is the leading market in the EU concerning sales (Fishman & Cherry, 2016). Even though there is lots of information on how to improve e-bike (and bicycle) usage, there is not much research done on how people cycle if they cycle and if location has an influence on the length, duration or purpose of the trips they make.

1.1 Research questions

So, to provide insight into the difference of citizens of cities and villages and their ways of travelling I came up with the following question: *What are the major differences in e-bike usage, regarding length and purpose of the trip, between citizens from rural areas compared to urban areas?*

The sub questions to answer the main question are:
- Are e-bikers living in rural areas willing to cycle longer in time and kilometres than e-bikers living in urban areas?
- For which purposes do e-bikers use their bicycle, and how does this differ between rural and urban areas?

1.2 Relevance

There are a couple of reasons why this research is relevant. The first one is that researchers and policies focused mainly on car use or used methods that are fitted for motorized vehicles on bicycling and walking. Now this focus is shifting to an interest in more sustainable ways of transportation (Broach, Dill & Gliebe, 2012). However, there is still a lack of data on cycling behaviour and especially on e-bike behaviour. "*When cycling is included [in a travel mode model], a typical practice has been to assume that cyclists choose the minimum-distance path between origins and destinations using a fixed travel speed (Larsen and El-Geneidy, 2011). This approach ignores network features, such as slope, traffic volumes, and the presence of on and off-street bikeways, and does not differentiate between bicycle trip purposes* (Broach, Dill & Gliebe, 2012). This is where my research can contribute. To give more insight in the behaviour of e-bike cyclists and how far they are willing to cycle to get to different purposes.

Another reason is that the outcome can contribute to better fitted policies to make e-bike usage more attractive, which is a step forward to a more sustainable world and healthier citizens. *In fact, as electric bicycles [will] become more prevalent, they might change traffic dynamics as the proportion of road users travelling by different modes changes, giving rise to unforeseen traffic situations and road user interactions* (Dozza et al., 2016). It is not only important to get more insight to avoid accidents but also to change infrastructure that is profitable for all cyclists. If there is a difference between e-bike cyclists in urban and rural areas, this should be taken into account when new infrastructure is build. For example, if the outcome of the research is that rural living cyclists are willing to cycle more than urban living cyclists, it can be lucrative to build more cycle highways in rural areas to encourage e-bike cycling.

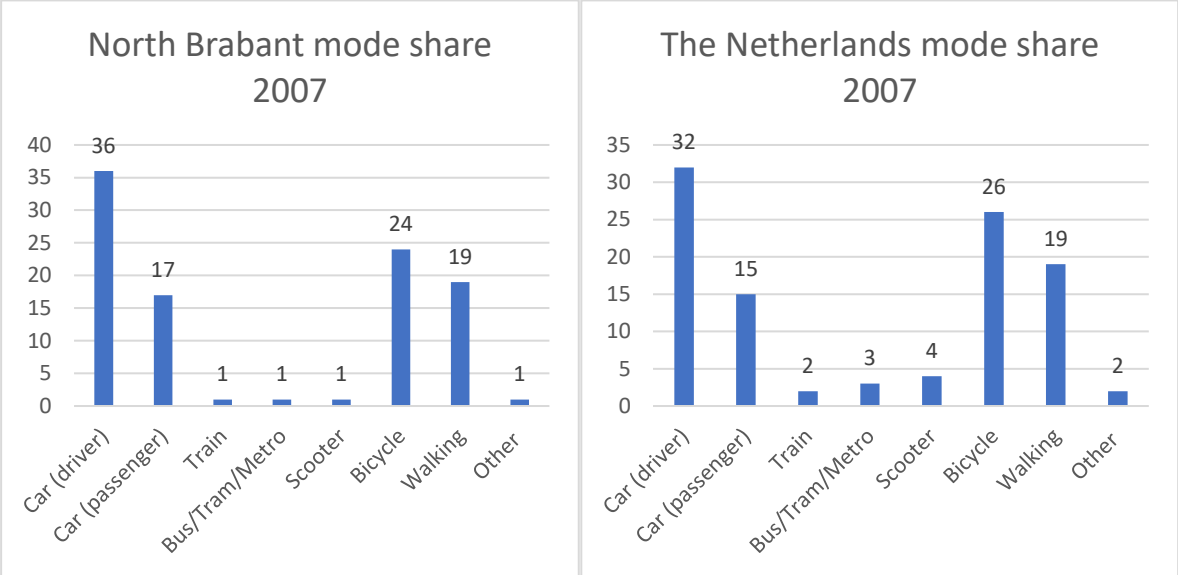
The last reason is that this research adds a new layer by using GPS tracking instead of surveys or interviews. GPS tracking can quantify the actual cycling behaviour instead of the cycling behaviour people remember they did. GPS is not yet being used a lot in cycling behaviour research because analysing the data is still consuming too much time. The software to process data is nowadays insufficient to quickly process the data (Carlson et al., 2014; Spek at al., 2009). The outcome of analysed data obtained by GPS tracking (actual behaviour) can be used to explain general patterns over larger samples, this is not possible with obtained surveys (stated behaviour). So, the relevance of my research concerning GPS tracking is twofold. One, research analysing GPS tracking is rare so this research contributes to a better understanding of analysing GPS tracking. And second, the outcome can be generalized to similar kinds of bicycle stimulation programs and thus can be used to improve cycle infrastructure and a better understanding of e-bike cycling behaviour.

1.3 Introduction to the region

In 2009, the region Brabant decided to create a vision and action plan for 2020 to encourage the citizens of Brabant to cycle more, called "Fiets in de Versnelling" (Bicycle in Acceleration) (Provincie Noord-Brabant, 2009). The municipality believed that cycling can be a part of the solution of the problems Brabant faces. These problems are air pollution, insufficient accessibility to and from the city centres and the increased change of obesity because people are moving less.

When Brabant is compared with the rest of the Netherlands, it is shown that citizens of Brabant cycle less than citizens in other parts of the Netherlands. As can be seen in both tables, the average share of cycling is in the region Brabant 24% and in the Netherlands in general 26%. The car is on the other hand 5% more used in Brabant than average if car driving and passengers are combined.

Figure 1



Source: Provincie Noord-Brabant, 2009

Some of the reasons for this difference are culture, facilities for bicycles and facilities for cars and the difference in urbanism and demographics throughout the country. Especially the facilities for bicycles are in Brabant less available than in the rest of the Netherlands on average, whereas there are more car facilities in Brabant when it is compared to the average of the rest of the Netherlands.

The last couple of years, different parties in Brabant have worked separately on an improved bicycle vision and in 2008 they decided to combine this vision and have a regional bicycle plan that should be implemented in 2020. The vision has three pillars;

1. Improving comfort and ease of cycling

The first priority within this pillar is to improve the bicycle network, both commuting as recreative connections. Not only with improving the roads but also with creating so-called cycle highways and priority traffic lights for cyclists. The second priority is the improvement of bicycle parking. In Brabant there is a big shortage on parking both in city centres as next to stations and the parking that is available is not safe, according to the locals. This results in using the bicycle less and seeing the car as the best alternative. So, the plan is not only to make more parking facilities for bicycles but also to make them easy to use, safe and with a good price-quality balance.

2. Seducing different groups to use the bicycle

There are different groups that can play an important role in improving the bicycle mode share. The first group are the commuters. Commuters are cycling less in Brabant compared to the Netherlands as a whole, but if they use the bicycle to commute the distances are relatively long. Commuters are an important group, because they mainly use the car and if they would cycle more air pollution and traffic jams will be reduced. Some examples to stimulate this group are more facilities when they arrive at work (showers, parking space), and more cycle highways. The second group is the youth. Cycling doesn't have a positive image among the youth and parents set a bad example by dropping their kids off at school or friends with the car. Nowadays you can also start driving lessons at 17 so there is only a short time period where kids need a bicycle to go from place A to B. If cycling would become cool again and get a positive image, the youth would probably use the bicycle more and longer, which has a positive affect later in life. The third group are the immigrants. Immigrants are inclined to use a car or public transport more often because they are not used to cycle. So, if municipalities and schools are pro-active in teaching immigrants to cycle, with (free) lessons and education, immigrants might become more quickly part of the community they live in.

3. Collaborating with each other

Different parties on different scales are working on plans to improve bicycle usage. Not only the municipality but also the government, Europe and companies, all from a different perspective and with different goals. A priority is to combine these plans and visions and collaborate with each other to share knowledge and innovations to improve the bicycle usage and promote it on different scale levels. Although this vision should be implemented in three years, the last report was from 2009 and it hasn't been updated since.

In the following chapters, I will first look at the different theories provided by previous researches. Then I will explain the methodology to conduct this research, analyse the data and answer the research question in the conclusion.

2. Theoretical framework

This chapter will explain three concepts, the first one is the difference between urban and rural space, the second one is cycling behaviour and the third one is GPS tracking. I'll end with my hypotheses, whom are based on the literature.

2.1 Rural versus urban space

To see whether there is a difference in bicycle usage between urban and rural space, it is necessary to conceptualize urban and rural space. Urban areas are often described as areas with a high population density and rural areas are thus areas with a low population density (Irwin et al., 2009). As I will describe in this part, the dichotomy between the two concepts is not as clear as it was or is possibly not even present anymore.

First, the concept of rural space will be explained, then the concept of urban space will be explained and at last I will discuss if it is still possible to explain both concepts on their own or if urban and rural space are so intertwined in this globalizing world that it is not possible anymore to point out completely rural or completely urban spaces.

2.1.1 The definition of rural space

There are a lot of different methods to describe rural space. The simplest explanation is that it is a rural space if there are agriculture activities, but it includes more than that (Madsen & Adriansen, 2004). In 1977 Paul Cloke developed an "*index of rurality for England and Wales", the different definitions were for a specific use, like employment, housing conditions, population or migration and "therefore not a general measure of rurality*" (Halfacree, 1993). Another way to describe rural space is to focus on the difference between socio-cultural characteristics of people who live in urban or rural space (Halfacree, 1993). An example of this is that people perceive "friendliness, neighbourliness and a sense of community" as characteristics of a rural space as oppose to anonymity in urban space (Mahon, 2007).

Halfacree (2004) proposed a "*network of rurality*" by applying the work of Jones (1995) and Lefebvre (1991) and made a "four-fold model of rurality:

- Rural locality, as characterized by distinct spatial practices, articulated in abstraction mainly through academic discourses.
- Formal representations of the rural.
- Everyday lives of the rural
- Lay discourses of the rural, i.e. social representations, that comprised all of the means of intentional and incidental communication used and encountered by people in their everyday lives" (Halfacree, 2004 in Mahon, 2007).

This model was tested in Ireland to "*establish respondents' perceptions and understandings of the place they resided in*" (Mahon, 2007). The most frequent answers were physical aspects such as green areas or open fields. Farming was also a largely responded answer, although the perception that farmers had to live of the land was not shared. The social aspects of living in rural areas were less prominent appointed, but the combination of feeling safe and the acquaintance with neighbours was seen as a feature of rural space (Mahon, 2007).

Rural space was long seen as a servant of the urban space and in modern time as a declining space due to the large migration to cities. Although this migration is still happening, rural areas renewed themselves and thereby revitalized the rural space. Agriculture is not only used for production but also to give an experience of the rural landscape to urban and foreign visitors (Galani-Moutafi, 2013).

"*Under the logic of hypermodern consumption, new forms of symbolic production arise and farmers are turned into living embodiments of collective natural and cultural histories*" (Heatherington, 2011). The rural represents an "idealization of the rural and a nostalgia for a simpler way of life" (Galani-Moutafi, 2013).

Thus, "*whilst one may wish to be wary of distinguishing between the 'real' and the 'unreal' rural, there is a growing realization in the literature that the quest for any single, all-embracing definition of the rural is neither desirable nor feasible*" (Halfacree, 1993).

2.1.2 The definition of urban space

According to Halfacree (1993) urbanism is "*characterized as being dynamic, unstable, mobile in stratification and impersonal, with contacts being determined by one's precise situation at the time (work, home, leisure)*". Although this still holds, there are many more versions of urban space, such as suburban, exurban and urban-rural space. Suburban is defined as "*areas immediately around cities that are densely settled; traditionally residential, but many modern suburbs include office, retail and commercial clusters*". Exurban is defined as "*low-density areas outside urban areas, but with a high degree of economic and social dependence on proximate urban and suburban areas*". Urban-rural space is defined as "*areas whose landscapes appear largely rural, but are substantially economically and socially tied to urban areas; includes exurban and amenity based rural areas*" (Irwin et al., 2009).

The same respondents that described the rural characteristics in the previous chapter, identified urban space also mainly by the physical features. Over one third of the respondents answered that they associated urban space with shops, business, employment and services like entertainment. However, "*built-up areas, lack of space, traffic/parking, crowds, congestion and the notion of a lack of privacy were [also] mentioned by a significant number of respondents*" (Mahon, 2007). The social components that were named were mostly negative, like concerns about crime, lack of a sense of community and unfriendliness, but this can possibly be a result of stereotyping the urban space (Mahon, 2007).

2.1.3 Urban and rural space or urban-rural space?

As I have shown above, it is not that easy to pinpoint an exact definition of both urban and rural space. Urban and rural areas are interdependent and making a division between the two would be an oversimplification (Bosworth & Venhorst, 2015). In this globalizing world it is not necessary anymore to go to the cities for work, shopping or leisure. Due to the internet and easy commuting, it is possible to live in a rural area and work and shop in an urban area or even online (Irwin et al., 2009; Bosworth & Venhorst, 2015). "*Increasingly, we observe places that are "rural" based on their location and landscape form, but nonetheless partially "urban" in their higher-order economic functioning and composition. We refer to such regions as urban-rural space to emphasize the fact that urban and rural are no longer distinct geographic entities, but rather end points of an economic and geographic continuum*" (Irwin et al., 2009).

This interdependence of urban and rural space has not only implications for new researches but also for development policies. Infrastructure, revenues, costs and public service also need to be more intertwined to benefit all the regions that are connected (Irwin et al., 2009). Researchers cannot use the simple division between urban and rural, so when a research is focused on the rural, it is necessary to include the urban effect as well, and vice versa. This is necessary to create the whole picture instead of researching isolated areas (Mahon, 2007). However, there is one way to make a distinction between urban and rural and that is by looking at the build environment. Thus, make the distinction between villages and cities. This divide was used for this thesis and will be explained further in chapter 3.

2.2 Cycling behaviour

Travel behaviour is a combination of rational economic benefits and routine according to different researches (van Acker et al., 2010). There are a lot of different aspects that influence the use of a bicycle and the purpose of the trips people make on their bicycles (Heinen et al., 2010). The purpose of trips also varies per country. In Western Europe, cycling is for example part of the daily commute, whereas in the United States it is more a form of exercise and for recreational purposes (Pucher & Buehler, 2010).

Van Acker et al. (2010) divides these aspects into three different categories namely, spatial, socioeconomic and individual factors. For this thesis only the first category is relevant, because I focus on the influence on trip length and purpose. Spatial factors are the size of the city and the infrastructure for cyclists. The other two factors will be shortly introduced to give an overview of all the factors that influence cycling behaviour in general.

Cycling behaviour can be about 'normal' bicycles and e-bikes, so I'll differentiate between the two, by first explaining how spatial factors influence 'normal' bicycles and then I'll focus on e-bikes.

2.2.1 The influence of different factors on cycling with a 'normal' bicycle

2.2.1.1 Spatial factor: City size

"Distance, either commuting distance or the distance between activities, is almost always taken into consideration when investigating an individual's choice to cycle or to use other transport modes" (Heinen et al., 2010). This is also in relation with the size of the city. The literature differentiates between rural areas (villages), small-/middle-sized cities and large/metropolitan cities. Cycling is done the least in rural areas because distances are far and public transport less available compared to cities. This makes it easier to use the car. Another reason cars have the biggest mode share is that there is more space in rural areas and so there is more space to park cars, which is an extra incentive to use it.

Small- and medium-sized cities are best for cycling and reach a maximum in bicycle use (Rietveld & Daniel, 2004). There are different reasons why these cities are perfect for cycling. The first one is the geographic size, this "*may be naturally more supportive of cycling or at least more easily modified*" (Pucher & Buehler, 2010). The smaller size means that facilities are likely within bicycle reach. Another effect of small- and medium sized cities, is that there is less traffic than in urban centres due to lower population and work opportunities, which makes it attractive for cyclists to take part in the traffic. Standalone cities have often a target area, like a university, major employer or town centre, which makes it interesting for companies to invest in bicycle infrastructure. For small cities embedded in metropolitan areas this is less often an advantage (Pucher & Buehler, 2010). The second one is the social characteristic of the city. People who are socially connected with each other, are (directly and indirectly) influenced by each other. So, the more residents who cycle the more they influence others in their surroundings to do the same (Marsden & Friedkin, 1993; Pucher & Buehler, 2010). In both Europe and the United States there are examples where small cities are bicycle-friendly and exceeded their larger counterparts in terms of bicycle share. One of these cities is Houten in the Netherlands. City planners designed the city in a way that cyclists "*have direct and easy access to the city centre, whereas the cars have to make substantial detours via a ring road*" (Rietveld & Daniel, 2004). Another example is Davis in the United States. Where trough "*a comprehensive program of infrastructure investments and promotional programs*" a culture of cycling is encouraged. Although this helped the growing popularity of cycling increase, it was not the only reason why it was successful in Davis. Davis has favourable land-use patterns, a strong commercial district so destinations are in cycling distance and it has a strong public transportation

service which promotes cycling "*as a means of travel to and from the train station"* (Pucher & Buehler, 2010). Even though these strategies are used in most Western cities, the higher level of cycling can be mainly seen in Europe due to important other factors, such as: historically compact cities with a defined core, national policies who are supportive towards cycling, implementation of deterrents to driving, cycling integrated into transportation planning and transportation integrated into land-use planning (Pucher & Buehler, 2010).

Large cities or metropolitan regions have a great share in bicycle use but compared to small- and medium-sized cities it falls down a bit. Although it might be expected that due to the high density of facilities and destinations, cycling is encouraged in large cities, there are also other factors that eliminate this effect (Rietveld & Daniel, 2004; Pucher & Buehler, 2010; Heinen et al., 2010). One of these is that the higher density and larger population causes more traffic and limited space on roads, which can frighten people to take part in traffic on a bicycle (Pucher & Buehler, 2010). Although this can be also an encouragement to use a bicycle. Due to congestion and traffic jams in cities, cars have a lower average speed and therefore it can be quicker to get around by bicycle (Rietveld & Daniel, 2004). Another factor is bicycle theft and damage to parked bicycles, this is higher in large cities and thus a discouragement for people to use their bicycle (Rietveld & Daniel, 2004). The last factor is the competition with public transport. In large cities public transport is usually well organised and places within and outside the city can be easily reached (Heinen et all., 2010; Rietveld & Daniel, 2004). So, competition with public transport and the fear of bicycle theft and the dense traffic makes it more difficult to cycle.

2.2.1.2 Spatial factor: Bicycle infrastructure

The importance of bicycle infrastructure depends on how confident someone is on their bicycle and if the purpose of the trip is recreational or not. Inexperienced cyclists might feel safer with separate bike lanes and traffic lights than experienced cyclists who feel comfortable on the road regardless, and recreational cyclists can choose different routes and are less dependent on the quickest route as opposed to cyclists on a commute that need to go from A to B as fast as possible (Heinen et al., 2010). There are different forms of infrastructure regarding to cycling. The obvious one is bicycle paths, but traffic lights and right of way at crossings are also important (Rietveld & Daniel, 2004; Heinen et al., 2010). How and where the paths are situated (separate lanes, marked sections on roads, adjacent to parking space) is important in different situations. In urban areas cyclists are more used to adjacent parking than in rural areas so this is considered to be less of a problem (Stinson & Bhat, 2003; Heinen et al., 2010). In urban areas cyclists also tend to avoid traffic lights and especially experienced cyclists find them annoying (Stinson & Bhat, 2003). In general, *continuous bicycle infrastructure and roads without parking* are preferred, although it is not clear if these conditions actually increase the cycling frequency (Heinen et al., 2010).

In the Netherlands, however, the amount of bicycle infrastructure at least helped increasing the share of bicycle mode. This is not a surprise when 16% of the total road network are bicycle paths. The Netherlands was the first country with a national bicycle policy, this policy gave subsidies to municipalities of urban areas up to 80% of the construction costs and in rural areas up to 50% (Rietveld & Daniel, 2004). The Netherlands invested in bicycle paths and lanes since the 1970s as a good alternative for motorized transport because of the oil crisis and the negative impacts of car use, and *since then the bicycle network more than doubled in length: from 9282 km in 1978 to 18.948 km in 1996* (Martens, 2004; Pucher & Dijkstra, 2000). Although the Netherlands is ahead of most countries if you look at bicycle infrastructure, the rest of Europe is not that far behind, especially when it is compared to the United States or Australia. In Europe, *cyclists can reach virtually any destination by bike without riding on roads with heavy car traffic volumes and high travel speeds*

(Pucher & Buehler, 2010). Another regulation which is more common in Europe than overseas is traffic calming areas, where cars need to adjust their speed limit to give way for cyclists and pedestrians (Pucher & Buehler, 2010). All these rules and regulations increase the safety of cycling, which might increase the bicycle mode share. It can be said that the relationship between bicycle infrastructure and the number of cyclists goes both ways, when there are enough bicycle paths there will be more cyclists who use it but when there are more cyclists it also becomes a higher priority to build more bicycle infrastructure (Rietveld & Daniel, 2004).

2.2.1.3 Other factors

There are three other classes of factors that influence bicycle behaviour, namely socioeconomic factors, individual factor and car use. Socioeconomic factors are income, age and gender and they all have both a positive and a negative influence on bicycle usage. A higher income enables a person to spend more money on a bicycle which increases the chances that he/she uses it more but also that that person can spend money on a car or on public transport which decreases the use of a bicycle (Witlox & Tindemans, 2004). Increase in age is related to an increase in active transport (bicycling and walking) because from 65 and older, people are more likely to be retired and have more time to use slower modes of transport. The negative influence on the other hand is that there is more danger in using a bicycle because the chance of injuries is higher (Scheepers et al., 2013). Gender related to cycling is country specific, in countries where people are used to cycling -like the Netherlands- women use the bicycle as much as men, but in countries where there are low cycling rates men tend to cycle more than women (Garrard et al., 2008).

The individual factor is the risk of injury by using a bicycle. This risk can be a real or perceived risk. Real risks are automobile traffic, driver behaviour, weather and personal security. Perceived risks are the things that might happen according to the person itself, whether this is based on existing threats or not (Rybarczy & Wu, 2010). Attitudes towards cycling and the rate of real and perceived risks are important factors to explain bicycle usage. It doesn't matter how many bicycle lanes you build or how safe the roads are, if people don't want to take the bicycle it is difficult to change that (Heinen et al., 2010).

"*Increasing car use in cities led to environmental pollution, roadway congestion, and a sharp rise in traffic injuries and fatalities. Those harmful impacts of car use provoked a dramatic reversal of the transportation policies of most German, Dutch, and Danish cities*" (Pucher & Buehler, 2010). To tackle this problem, city governments could either adapt the city to the car or restrict car use and promoting other options like cycling, public transportation and walking. Especially the focus on (e-)bicycle infrastructure helped to boost cycling in these countries (Pucher & Buehler, 2010). The restriction in car use is visible in "*sales taxes on fuel and new car purchases, import tariffs, registration fees, license fees, driver training fees, and parking fees*" (Pucher & Buehler, 2010). All these taxes and fees are considerably higher in Europe compared to other Western countries like the United States, Canada and Australia. The result is that the costs to own and use a car in Europe are two to three times higher than in those other countries. Since it is more expensive to own a car, people are looking for a cheaper alternative which is cycling. Yet the correlation between car ownership and cycling goes both ways; less cars per household encourage more cycling and when people cycle more the need to own a car may become less of a priority (Heinen et al., 2010).

2.2.2 The influence of different factors on cycling with an e-bike

"Giving the sensation of cycling with a tail wind or slightly downhill, the e-bike is quicker, it enables longer trips over hilly routes and it is an alternative for people who for various reasons are averse to bicycling. Compared to local public transport and rush-hour driving, the e-bike offers competitive travel speeds. Clearly, it has the potential to replace many car and public transport trips, all to the benefit of the environment, public health and other motorists" (Fyhri & Fearnley, 2015). Although there are many advantages for using e-bikes and sales are increasing all over the world, there is not a lot of research that shows the effects on e-bike usage and the effect they have on motorized modes of travel (Rose, 2011; Fyhri & Fearnley, 2015).

2.2.2.1 Sales

The sales on e-bikes has grown rapidly in the past decade, over 150 million have been sold worldwide and that is the "*most rapid uptake of alternative fuelled vehicles in the history of motorisation"* (Fishman & Cherry, 2016; Jamerson & Benjamin, 2013). The main consumer group are the elderly although other groups, especially in Asia, are discovering the e-bike as their preferred mode of transport more and more (Fietsberaad, 2013; Fyhri & Fearnley, 2015). In the EU the two-leading e-bike markets are Germany and the Netherlands. They account respectively for 44% and 21% of all sales (Fishman & Cherry, 2016). What is remarkable about this is that Denmark does not have a spot in the top five of highest e-bike sales in Europe, even though it has the second highest bike share (18%) in Europe (Pucher & Buehler, 2010). While it is a fact that e-bike sales are growing, the reason why is still unknown. Two possible reasons are the higher fuel prices, which makes owning a car more expensive and city regulations that are trying to keep the car more and more out of the city centre (Rose, 2011).

2.2.2.2 Spatial factor: City size

The general belief of cycling patterns is that people in rural areas rely more on the car and less on the bicycle because facilities are further apart. In urban areas this is the other way, due to congestion, high parking prices and close by facilities the bicycle becomes a better alternative than the car (Harms et al., 2014). However, it seems that if you look at e-bikes this trend is inverted. The "*increases in travel distances and poor public transportation service are likely to stimulate a mode shift from bike or bus to e-bike and thus increase the ownership of e-bikes […]"* (Zhang et al., 2013). In urban areas the availability of public transportation and the mixed development has a negative influence on the ownership of e-bikes (Zhang et al., 2013). The reason why there is a difference in bicycle and e-bike usage was not been researched (yet).

2.2.2.3 Spatial factor: Bicycle infrastructure

As I explained before, there are some factors that decrease the likelihood that people use a normal bicycle for their daily commute. Topography, distance, physique and weather are some of these factors (Heinen et al., 2010), and e-bikes can make these factors less of a constraint and even make "*cycling fun again"* (Popovich et al., 2014; Fishman & Cherry, 2016). Although this higher increase in e-bikes is a positive outcome for environmental and health issues, it creates new problems as well. E-bikes come, for example, in different forms and can range in speed. Most infrastructure is built for either cars/motors or normal bicycles, so where does the e-bike fit in this? It is important to think how infrastructure can be adjusted to more cyclists, whether they are on a normal bicycle or e-bike (Rose, 2012; Harms et al., 2014). Perhaps "*a more fundamental reconsideration of the allocation of street space for the different urban transportation modes (cyclists, pedestrians, cars and public transport) might be needed"* (Harms et al., 2014).

2.2.2.4 Other factors

The socioeconomic and individual factors that influence normal bicycle usage have also an effect on e-bike usage. One of these factors is age. You could argue that e-bikes are mainly used by the elderly, due to declining physical abilities. Although there is a reluctance among the young to buy e-bikes, this is not due to the image of an e-bike but more to the costs of it (Rose, 2012; Fyhri & Fearnley, 2015). Another factor is gender. In countries where cycling is not a big part of the daily commute, men tend to cycle more than women. When you look at e-bikes, this effect becomes less. Women tend to use an e-bike more often than a normal bicycle, so if the number of men that use an e-bike is equal to the number of men that use a normal bicycle, the number of people that use an e-bike is higher than the people that use a normal bicycle (Fyhri & Fearnley, 2015). The last factor is income. Research shows "*that e-bikes are less a transitional mode between a bike and automobile than an affordable, higher-quality mobility option to public transport*" (Zhang et al., 2013). Thus e-bikes are mainly used by low- and middle-income households.

Using an e-bike can also change the perception of safety and thus increase the likeability people use this healthy alternative instead of their car. "*In a North American survey of e-bike owners 60% feel safer riding an e-bike and 42% said the e-bike had assisted in avoiding crashes*" (MacArthur et al., 2014). The speed of an e-bike and thus keeping up with traffic was the main reason why they felt safer (Fishman & Cherry, 2016). A project in Canada, presented the same conclusions, where respondents felt safer on an e-bike than on a normal bicycle due to the fact that they can react quicker in traffic. This study "*noted that 83% of respondents felt as safe on an e-bike as on a conventional bicycle, with 95% feeling that they had complete control when the motor was running*" (Rose, 2012). This is a positive and somewhat surprising effect of e-bikes on safety. It can be expected that people who do not feel safe on a bicycle, find an e-bike even more frightening because of the increased speed, but according to the researches the speed is exactly the reason people feel safer.

2.3 GPS measurements on cycling behaviour

"*The availability of so-called geopositioning devices such as GPS (Global Positioning System) devices has grown enormously in the last decade and is still increasing. More and more people own a navigation system such as a TomTom, a GPS for orientation for outdoor uses, biking and geo-caching or a mobile phone or other handheld communication device with built-in GPS. These devices are mainly used for orientation (determining where you are), navigation (determining where to go) and communication (exchanging information with others or accessing information services). But the devices can also be used for tracking, i.e. saving a travelled route into a track log. This ability makes the technology useful to collect spatial-temporal data and thus as 'sensors' for observing and measuring activities of people*" (Spek et al., 2009).

One major implication of cyclist behaviour research is that models are based on cars and the assumption that everyone acts on their economic instincts: minimise travel time and the smallest costs. However, it is known that cyclists choose their routes for various other reasons, such as safety and cycle facilities. Route measurements through GPS tracking would help to solve this problem and get more insight in how cycles behave (Ehrgott et al, 2012).

According to Spek et al., (2009) there are in fact three different perspectives where GPS contributes to urban research. The first one is visualisation. In the three stages of processing, analysing and communicating it has value. In the first stage it is important because it gives "*manual validation of [the] data*". In the analysing stage it helps as a tool to analyse and in the communication stage it is important for both experts as the public. This is because it makes statements and numbers more

readable by looking at maps or other visual data. The second one is accuracy. Usually post-hoc mapping and dairies made by participants are used for registering routes in research, however these methods are biased because there are based on the memories of the participants. GPS can avoid this bias and, in the future, maybe even visualise real-time behaviour. The last one is validation, GPS can also assist other urban research methods, which improves the research in itself.

However, GPS is not perfect yet. There are some major implications. The first one, is that the reception is not ideal in urban settings, especially when reflections of other buildings can confuse the signal. The second one, is the improvement of the software to increase the speed of data processing. Nowadays a determination of research is the limitation of the scale, if software to process the data would be upgraded this would be less of a concern (Carlson et al., 2014; Spek at al., 2009). The third one relates to the second one, GPS data consists usually of large amounts of data and to grasp that and deal with it without losing critical information is difficult. In combination with slow software, researchers rather not work with this data.

Despite these limitations, "*in this new era of advanced technology, the extensive use of cyberspace has enabled the provision of real-time information with sophisticated geographic information systems. This can be a means to induce changes in travel behaviour*" (Ehrgott et al, 2012).

2.4 Hypotheses

Derived from the literature, there are two hypotheses that I want to test. *The first hypothesis is that people who live in cities make more frequent trips but these trips are shorter than the trips made by people who live in villages.* According to the literature, cyclists in cities are more inclined to cycle because facilities are close by and it is most of the times quicker due to congestion and traffic jams (Pucher & Buehler, 2010; Rietveld & Daniel, 2004). When facilities are close by, it is less of a hassle to make a trip per purpose instead of combining varies purposes within one trip, whereas in villages it takes longer to do a trip per purpose because facilities are widespread.

The second hypothesis is that the purpose of people who cycle in villages is mostly work-home related and that the purpose of people who cycle in cities varies more. Cities offer more recreational facilities, like restaurants, theatre, cinemas, which makes it easier for people in cities to go there regularly. Another reason is that city municipalities try to ban cars from the inner cities so when people go out, they are more inclined to use their bicycle (Pucher & Buehler, 2010). Thus, people who live in cities have a more variety of purposes than people who live in villages.

3. Methodology

This chapter starts with the background information of the data sample I used. Then it explains the steps to prepare the data for analysing and then it explains the different methods I used to analyse the data. It concludes with the possible problems and risks for this research.

3.1 The Bicycle Stimulation Program

The Bicycle Stimulation Program Brabant (BSP) gathered data about e-bikes and how participants cycle over a period of 1,5 years. They followed 581 participants from September 2013 until October 2014. These participants used the app B-Riders to follow their cycling trips and the purpose of these trips. The participants were paid to keep track of their trips and signed in for the program themselves (Feng & Timmermans, n.d.).

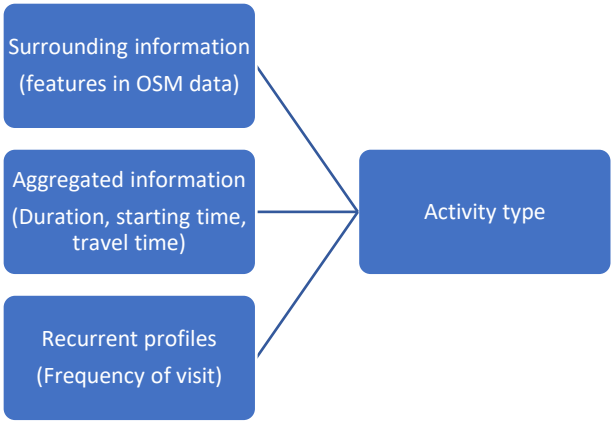
There were a couple of boxes you had to tick to be allegeable for this study. The first one was that your work should be further away than 3 kilometres of your home. The second one was that, before the program, you used your car 50% or more for commuting, the third was that you work in the region North Brabant and the last was that you were between the age 18 and 65 (Feng & Timmermans, n.d.).

The data was saved in a personal csv file for each participant. Although they collected the data from September 2013 until October 2014, for my thesis I only used the data from January, July and September 2014 in order to make the data size not too big and manageable for the time of the thesis. The home locations are given in X and Y coordinates in the dataset. It is necessary to prepare the data using Python because not all information provided by the dataset is useful for this research. In the csv files not only, the X and Y coordinates of the routes are collected but also the person id number, the track id number, the date and time, the accuracy in meters, the speed in km/hour, the heading in degrees, the modality where 2 means a 'normal' bicycle and more than 2 an e-bike, the place of origin, the place of destination and the purpose of the trip. The different purposes were modelled by researchers Feng and Timmermans of the TU Delft, which I reuse in my thesis.

They argue in their article that: *Comparing with existing works to detect activity type, […] we attempt to introduce additional information that is commonly available in large-scale aggregate GPS panel data, based on repeated multiple observations. Assuming that people's daily activities are based on scripts, we argue that repetitive patterns in GPS panel data can be successfully detect a certain type of activity using frequency and regularity information. To represent such a concept, the heat map of activity locations for an individual during one month indicates that the two main activities have a higher frequency to happen at the same or similar location than other activities.*

Specifically, the frequency of activity locations can be generated through matching the activity locations with the zonal data. Here, the frequency defined also compensates the deficiency of the percent of areas. In case that people frequently go to a place for leisure where the area is mostly covered by shops, using the spatial variable only may lead to inaccuracy, because there might be a higher probability for shopping rather than leisure activity. By having the frequency of activity locations within zones, if people often go to the same place for leisure activity, the inherent pattern will be captured by the frequency in the sense that frequency in that zone will be higher than those in other zones. Of course, one cannot rule out the case that multiple frequent activities happen within the same zone, but it can be easily overcome through the use of fine-grained zones or grids (Feng & Timmermans, n.d.).

The model they created is the following:



From this model they derived ten different purposes: home, paid work, non-daily groceries, daily groceries, social, recreation, spare time, services, study and unknown.

For the csv file to be included in my research I decided to have three conditions. The first one was that the home coordinates were in North-Brabant, the second that (almost) all trips were made by e-bike and the last one that the person made at least 10 trips in those three months. I did this because I believe that people who cycled less than 10 trips in three months is not representative and thus can create a bias in the outcome of the analysis. The same goes for people who didn't have their residence in North-Brabant because the outcome is supposed to say something about e-bike cyclists in North-Brabant, so people who reside somewhere else creates a bias in the outcome. After removing all files that didn't met these conditions, I had 538 files (and thus participants) that were useful for my research.

3.2 Data preparation

The dataset was so large that it was necessary to first make a selection of the data I needed and prepare it so I could use it in SPSS. The steps that I made are described below.

3.2.1 Determine whether home is in an urban or rural area

After I looked at all the csv files, I chose one home route from each file. The X and Y coordinates of each file went one by one through a calculator on the website https://www.gps-coordinaten.nl/, this gave the precise address of each contestant. Due to possible privacy issues, I only wrote down the person ID and the name of the place they lived in an excel sheet. To determine whether it was a city or a village, I wrote down all the places with city rights according to the book "*Repertorium van de stadsrechten in Nederland*" by Joost Cox (Cox, n.d.) and matched them with the places I had. The places that did not match were double checked to see if they were actually laying in North Brabant and if they were, they were placed under village. Even though I explained in the literature chapter that it is almost impossible to see urban and rural as two detached areas, it is however possible to determine if a place is a city or a village. In the Netherlands a place can only call themselves a city if it has city rights. So, for this research to make the division between urban and rural, I looked at the division between city and village.

3.2.2 Determine the length, time and purpose of each trip

Python scripting was used to determine the different aspects of each trip. First, we converted the text files into a string so Python can work with the csv files. Then I decided which information I needed for my thesis. I wanted a duration table with the total duration of each trip per person, the length in metres of each trip per person, number of trips per person and the number of trips of each purpose per person. To get the duration table, we had to make a script that first grouped each trip with the same track ID and then subtract the starting time off the finished time. The length of the trips was more difficult to calculate. First, we had to incorporate spatial references because the csv files only provided X and Y coordinates but without spatial reference these are useless for calculating the length of trips. Then we grouped the trips again and add the coordinates from point zero of each trip, which Python converted to metres. The number of trips was a count of all the trips with a different track ID. The number of trips per purpose was also a count of the grouped trips and then a value count. We did this for both cities and villages separately. This created two long lists with the information I needed, but it was not easy readable. So, I wrote another script where I put all the information in columns per person and converted it to an excel file. Both these scripts can be found in the appendix.

3.2.3 Determine average length of trip and most important purpose for urban and rural areas separately.

With the excel sheet, I had all the information I needed. For every part of information, I wanted to research, I made a graph. To make the graphs more readable I included the median and the 75% quartile. In the analysis chapter I will elaborate more on the graphs and the differences between the e-bike cycle patterns of people who live in cities and villages.

3.3 Analysis methods

For the analysis I used four different methods. To test the first hypothesis, I used two t-tests and three ordinal logistic regression models. To test the second hypothesis, I used a crosstabulation table and Pearson Chi-Square and a binary logistic regression model. Why and how I used these tests, will now be explained.

3.3.1 T-tests

The data consists of two samples, the urban sample and the rural sample. To compare these two and look for potential differences it is possible to use T-tests. There are six assumptions that need to be met before a T-test can be done. The first assumption is that the dependent variable is continuous, for the variables trip frequency (number), duration, duration first hour and length this holds. The variable purposes however is not a continuous variable so the T-test will not be done for that variable. Thus, only the first hypothesis will be tested with the t-tests. The second assumption is that the independent variable is categorical, this holds because the variable location has two categories. The third assumption is that all the variables need to have cases, which also holds. The last assumption is that there is a normal distribution of the dependent variable. However, with large data this doesn't have a large influence on the accuracy of the p value (Bryman, 2008; Field, 2009).

3.3.2 Crosstabulation table and Pearson Chi-Square

To test whether the distribution over the variable purposes differs between urban and rural, I had to choose a different method because it cannot be tested by a t-test due to the fact that it is a categorical variable. This is why I chose to make a crosstabulation table and did a Pearson Chi-Square.

There are three assumptions to do a Chi-Square. The first is that both variables need to be an ordinal or nominal variable, this is met because both purposes and location are categorical variables and thus nominal. The second is that both variables should consist of at least two independent categories, this is also met because the variable location has two categories, villages and cities and the variable purposes has ten categories. The last one is that all "*expected frequencies should be greater than 5*" (Field, 2009). If this last assumption is not met, it is advised to gather more data or exclude this data. If it would be included it can "*fail to detect a genuine effect*" (Field, 2009). The data was already gathered, so the first option was not possible, that is why I excluded the purpose unknown for this model. After that adjustment, all assumptions were met.

3.3.3 Ordinal logistic regression

As an addition to the T-tests and the Chi-Square I wanted to measure the influence of rural and urban areas on the behaviour variables. The variables trip frequency (number), duration, duration first hour and length are all ordinal variables, so I could do an ordinal logistic regression for these variables.

Before I could conduct an ordinal logistic regression, I had to check if the following three assumptions were met. The first assumption is that the dependent variable is ordinal. As I said before, this is correct for the four variables above. The second assumption is that the independent variable is either continuous, ordinal or categorical. In this case the independent variable is location which is a categorical variable, with two categories, villages and cities. The third assumption is that there is no multicollinearity. This can occur when two independent variables are highly correlated. For this model I only use one independent variable, so this is not possible which means that the assumption is met (Field, 2009).

3.3.4 Binary logistic regression

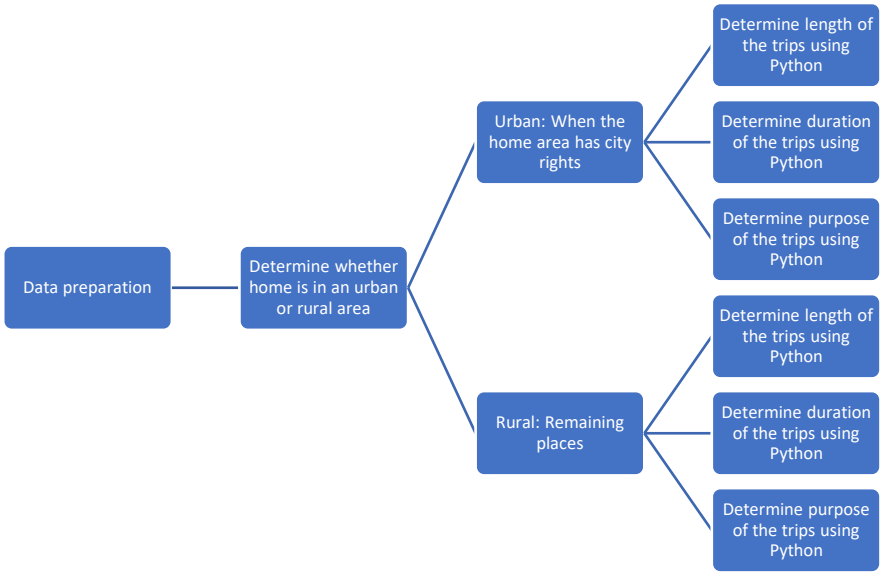
To measure the influence of urban and rural areas on the purpose behaviour I chose to do a binary logistic regression with SPSS. The purpose variable is a categorical variable so it was not possible to do an ordinal logistic regression as with the other variables. I chose to separate all the purposes and do a binary regression for each of them. The data became a 2x2 table with the purpose divided in either true or false and the independent variable (location) in either village or city.

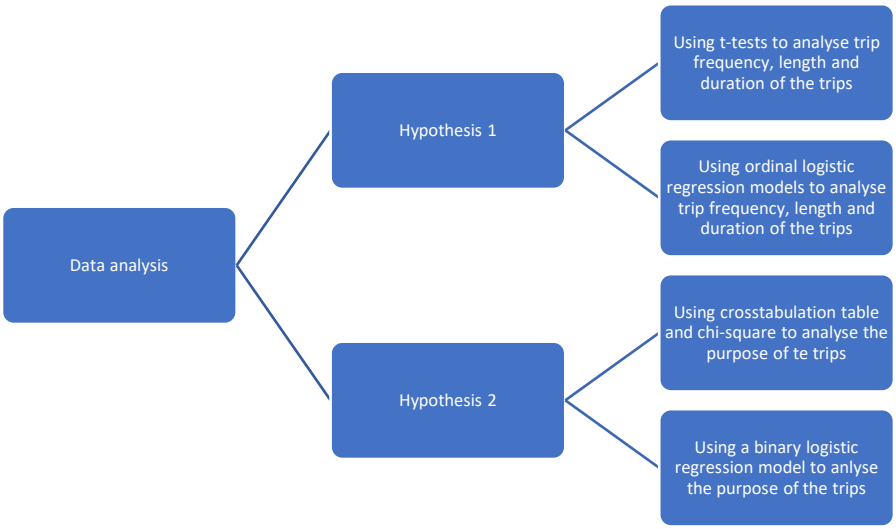
Before a binary logistic regression can be done it is necessary to check if the data fulfils a couple of assumptions. The first assumption is that the dependent variable is a binary variable. The second assumption is that the independent variable(s) are categorical or continuous. In this case are the variables length, duration, duration first hour and trip frequency continuous variables and the variable purpose is a categorical one. The third assumption is that the categories of the dependent variable should be exhaustive and exclusive (Bryman, 2008; Field, 2009). The fourth and last assumption is that "*there is a linear relationship between any continuous predictors and the logit of the outcome variable*" (Field, 2009). All these assumptions are fulfilled for this data.

The results of the regression analyses will be explained in the analysis chapter and the output of SPSS can be found in the appendix.

3.3.5 Flow chart

To make the different steps easier to read and to summarize them, I decided to make two flow charts. The first one shows the steps I did to prepare the date and the second one shows the steps I did to analyse the data.





3.4 Problems and risks

There are two main limitations within this research. The first one concerns the B-riders' dataset and the second one concerns the binary logistic regression analysis. The limitation of the B-riders' dataset is that the conclusions cannot be generalized for the whole province of North-Brabant or the Netherlands. The people who participated, where paid and applied for it themselves. They were not randomly chosen by the organization. The conclusions, if they are statistically significant, are, thus, only generalizable for similar kinds of bicycle stimulation programs.

The limitation of the binary logistic regression analysis is that if the variables are not correct, the assumptions of the model can be weakened and even lead to errors. This can easily be resolved by knowing which variables have a definite influence on location (concerning e-bikes) but because this is a unique research there isn't previous research to compare the variables with. Thus, when reading the results, the possibility of a weakened model should be taking into account.

4. Analysis

This chapter will show the different outcomes of the purposes, durations and lengths of the trips. It is divided between citizens of cities and citizens of villages in North Brabant, explained in the previous chapter. The first part will look at the relative differences and make a comparison between the cities and the villages. The second part will analyse the outcome of the t-tests and cross tabulation and the third part will analyse the outcome of the logistic regressions.

In total there were 160 citizens spread over 20 cities in North Brabant who volunteered to participate. These 160 citizens cycled 11.594 trips in three months which means that they cycled 72,46 trips on average per person over these three months. The 376 citizens spread over 118 villages in North Brabant cycled 26.376 trips in the same time span, which is 70,15 trips on average per person. The average length of the trips of people who live in cities is 8,08 km and in villages 9,3 km, which is in line with my hypothesis. The most important purpose for both people in cities and villages is home. This information is the basis for the analysis that now will be unfolded.

4.1 Descriptive statistics of trips in cities and in villages

This part will look at the relative data to make a comparison between the people who live in the cities and the people who live in the villages. The first comparison is about the purpose of the trips. As can be seen in the graphs on this page and the next page, there are not a lot of differences, around 70% of the trips are to go either to a paid job or to go to their home. This is not surprising, because the age group was 18-65 and you had to work so people spend most of their trips between home and work. There is a difference visible for groceries, daily and non-daily, social, recreation and spare time. People in the cities made more trips with this purpose, but whether this holds statistically will be checked in the next part.

Figure 2

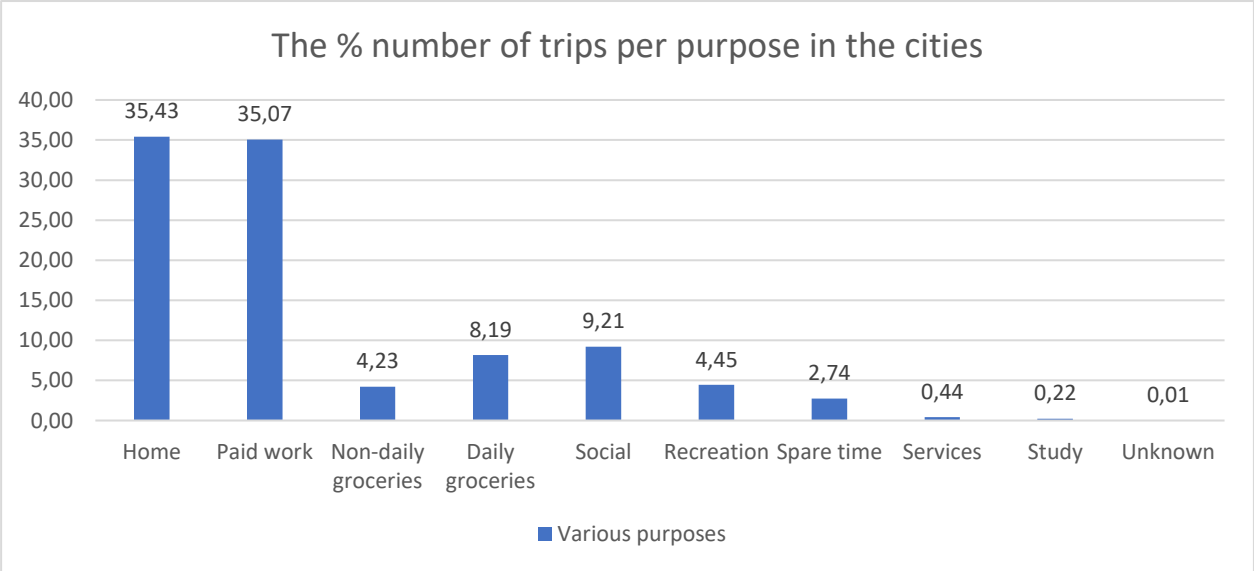


Figure 3

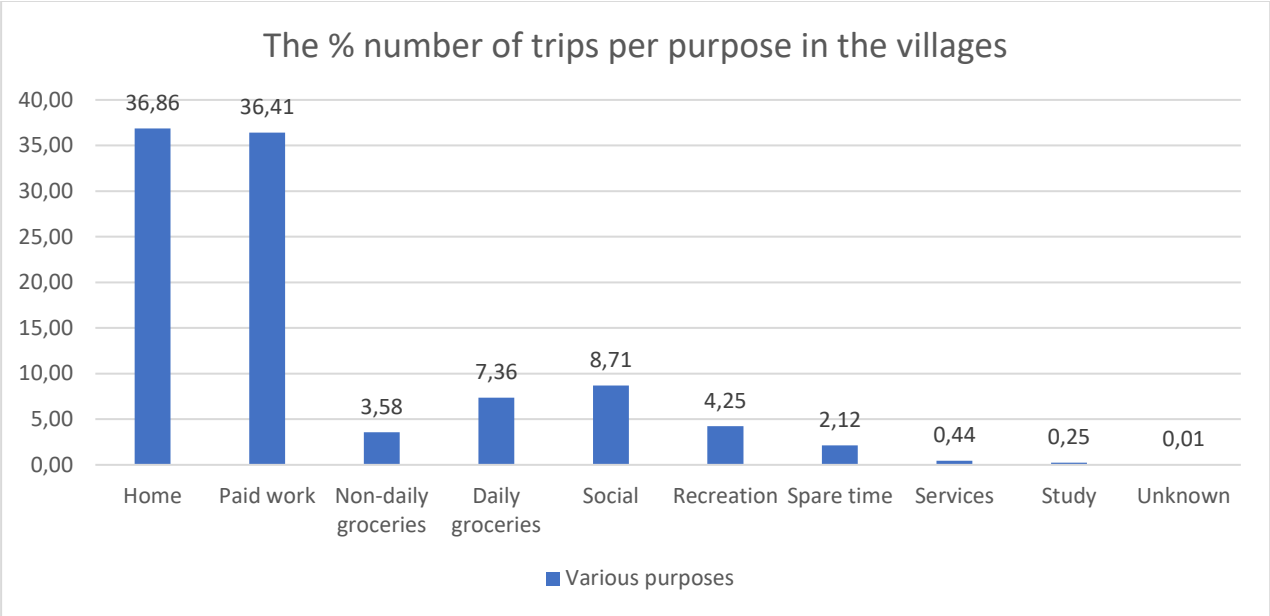


Figure 3 and 4 show the difference in trip frequency distribution. The two categories with the biggest share, both in the cities and in the villages, are the categories 31-60 and 61-90 trips. If the two graphs are compared, it shows that the share of the trip frequency is higher in the villages for the categories 31-60 and 61-90 than in the cities, but for the other categories this is vice versa. Which is interesting because the hypothesis is that people in cities make more trips. But if the median is calculated, it can be seen that the median in the cities is 65,5 trips and in the villages it is 45 trips. As is the 75% quartile, which is 100,25 trips in the cities versus 95 trips in the villages. This can be seen in the boxplot in figure 5. This is in line with the hypothesis. A possible explanation is that facilities in cities are close by so people use different trips for different purposes instead of combining the different purposes into one trip.

Figure 4

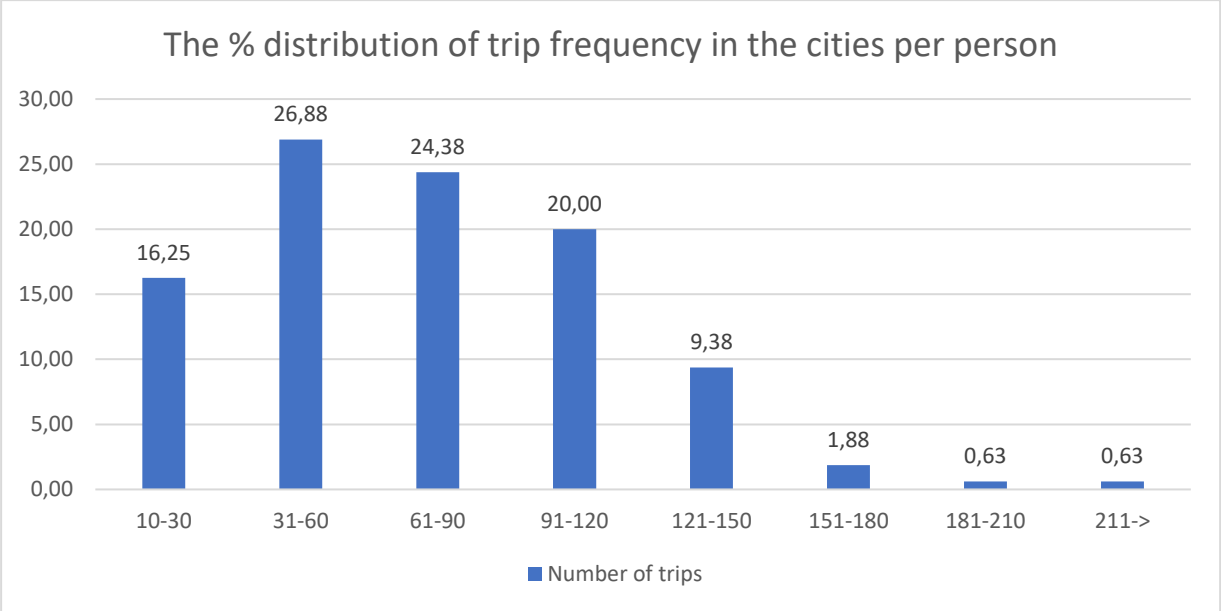


Figure 5

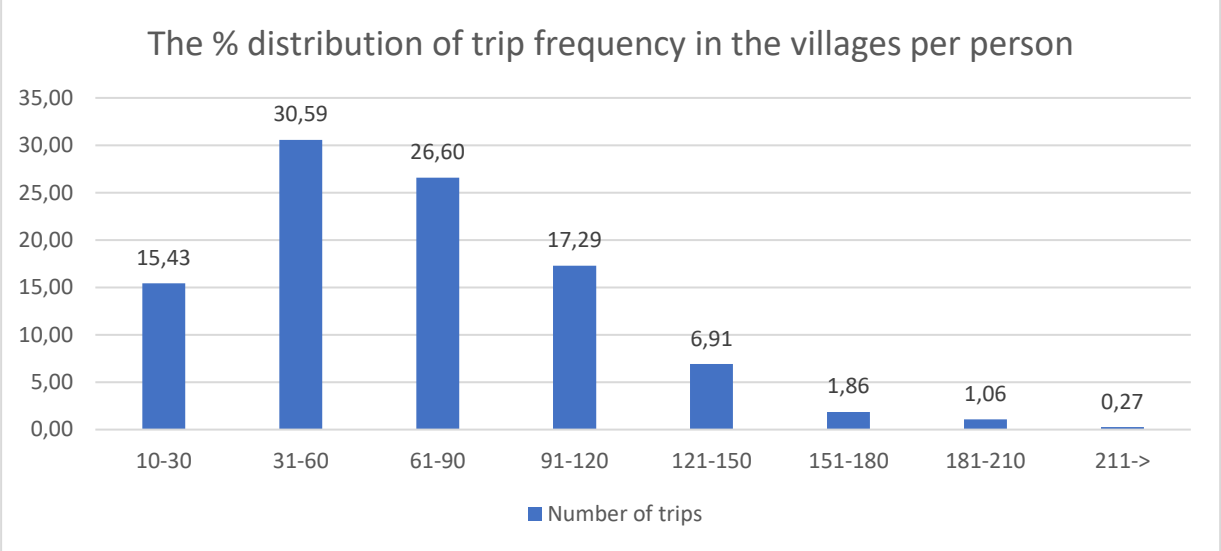


Figure 6: The percentiles of trip frequency

		City	Village
N	Valid	11594	26376
	Missing	0	0
Quartiles	25	42,75	40,00
	Median	65,50	45,00
	50	100,25	95,00

The other hypothesis was that people in cities cycle shorter trips, due to the proximity of facilities, than people in villages. As you can see in the two graphs below, both groups cycle maximum an hour per trip. But within that hour there are quite some differences. If you look at the first two graphs, it is interesting to see that after the hour, the number of trips diminishes quickly with one person who made a trip that lasted between 4,5 and 5 hours. If you look at the graph from the citizens in the villages you can see the same pattern but it diminishes less quickly and there is one person who made a trip that lasted between 5,5 and 6 hours. The median for the cities is 00:21:41 minutes per trip and for the villages it is 00:25:05 minutes per trip. This is also in line with the hypothesis that people in cities cycle a shorter duration per trip. The same goes for the 75% quartile, which is 00:35:41 minutes in the cities and 00:37:02 minutes in the villages. The differences aren't as big as compared to the number of trips but the t-test and regression analysis have to confirm if both outcomes are still relevant.

Figure 7

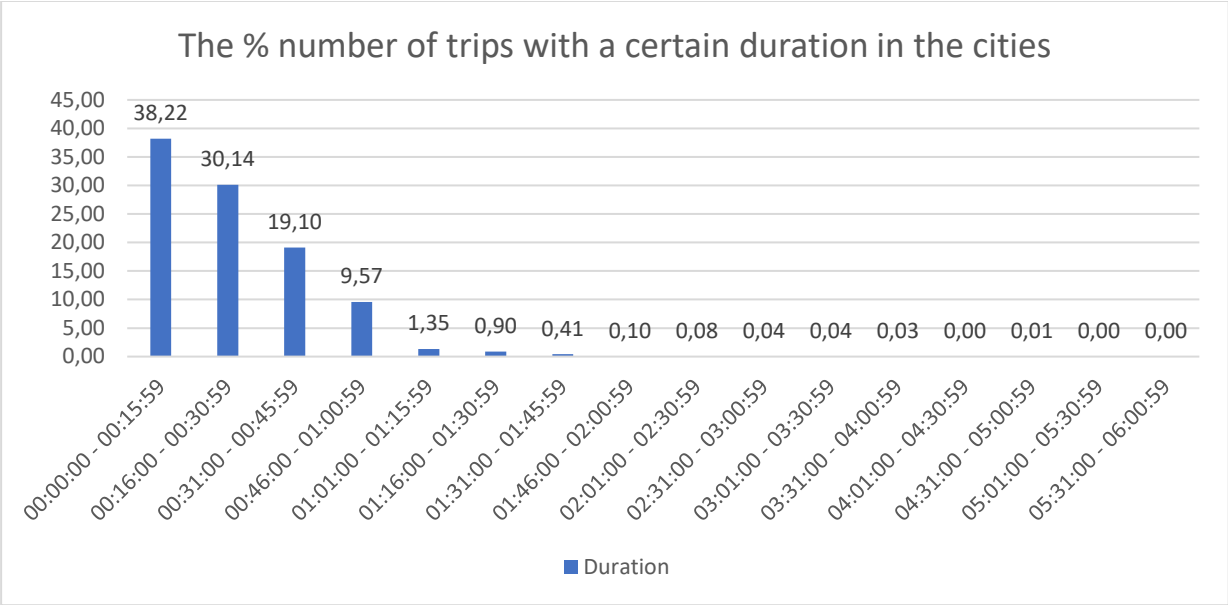


Figure 8

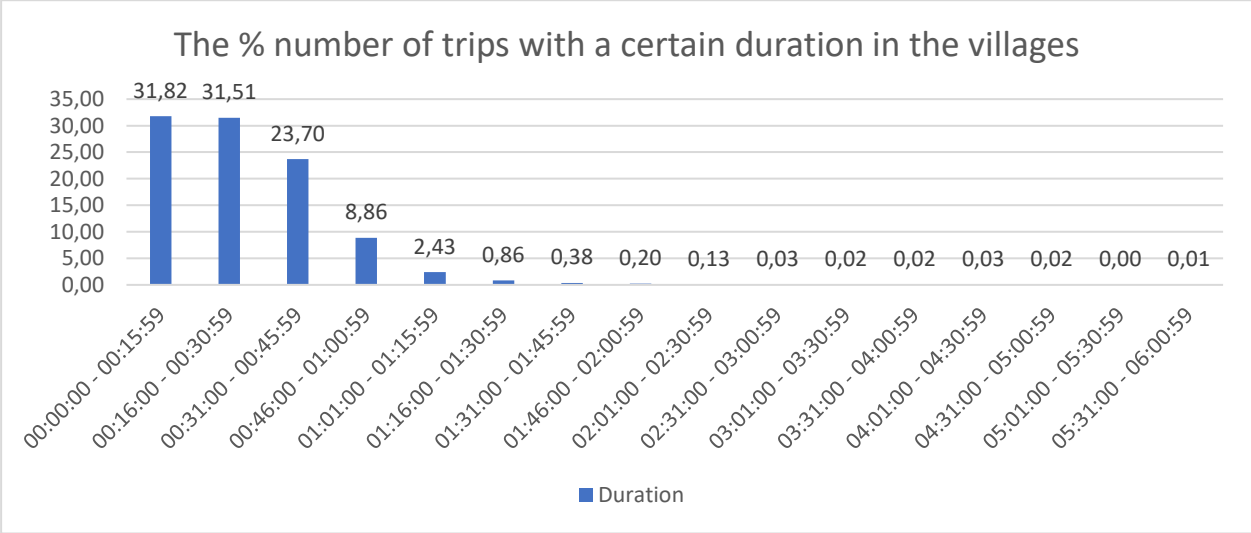


Figure 9: The percentiles of the duration of the trips

		City	Village
N	Valid	11594	26376
	Missing	0	0
Quartiles	25	00:10:01	00:12:35
	Median	00:21:41	00:25:05
	50	00:35:41	00:37:02

There is a big difference between citizens in the cities and the villages when only the first hour is accounted for. In the cities, most trips are made within 6 and 11 minutes and the second one is even shorter, namely 0 till 6 minutes. Whereas in the villages, most trips took between 21 and 26 minutes and the second highest were 11 till 15 minutes which is also longer than in the cities. You can also see a nice declining slope after 11 minutes in figure 9 (with the exception of the trips that took between 21 and 26 minutes). Where in figure 10, the declining slope starts only at 21 till 26 minutes. Before that the differences per 5 minutes are larger. The reason for this is unknown based on the data, but one possibility is that people either live really close by their friends or families or that they have to cycle further to other facilities or social activities that are not close by. Another possibility is that work in villages is further away or even in the next city or village and that people for that reason have to cycle longer. This could be an example of a qualitative follow-up research question. The median in the cities is still 00:21:06 minutes per trip and in the villages it is 00:24:08 minutes. This is a small difference compared to the duration in total, but that can be explained because almost all trips have a duration within an hour. The 75% quartile is 00:33:58 minutes in the cities and 00:35:20 in the villages. This again can be an indicator that people in villages make longer trips than people in cities but the t-tests and the logistic regression analysis has to reveal if it is statistically significant.

Figure 10

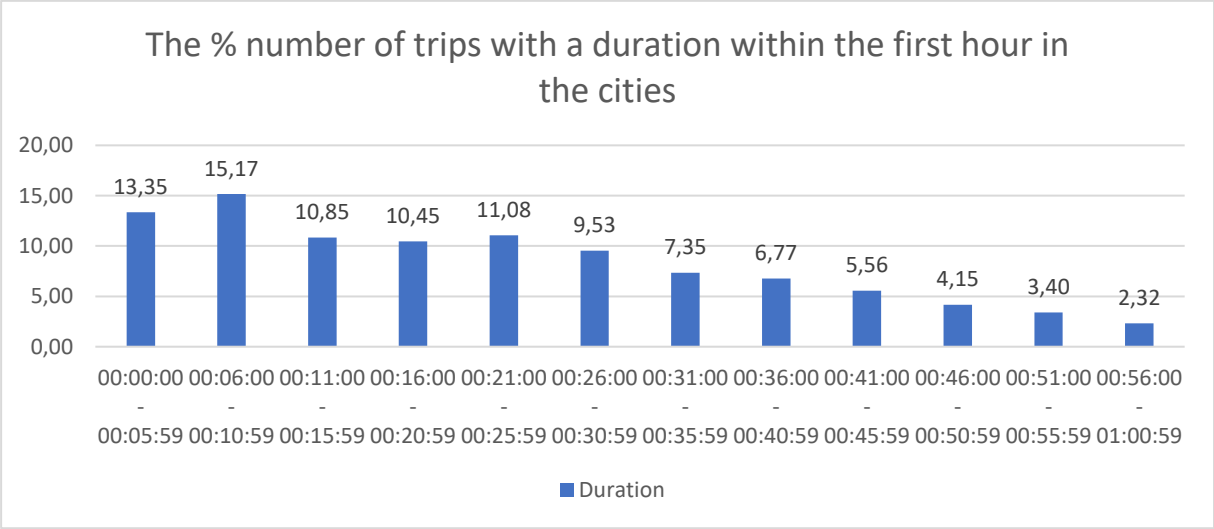


Figure 11

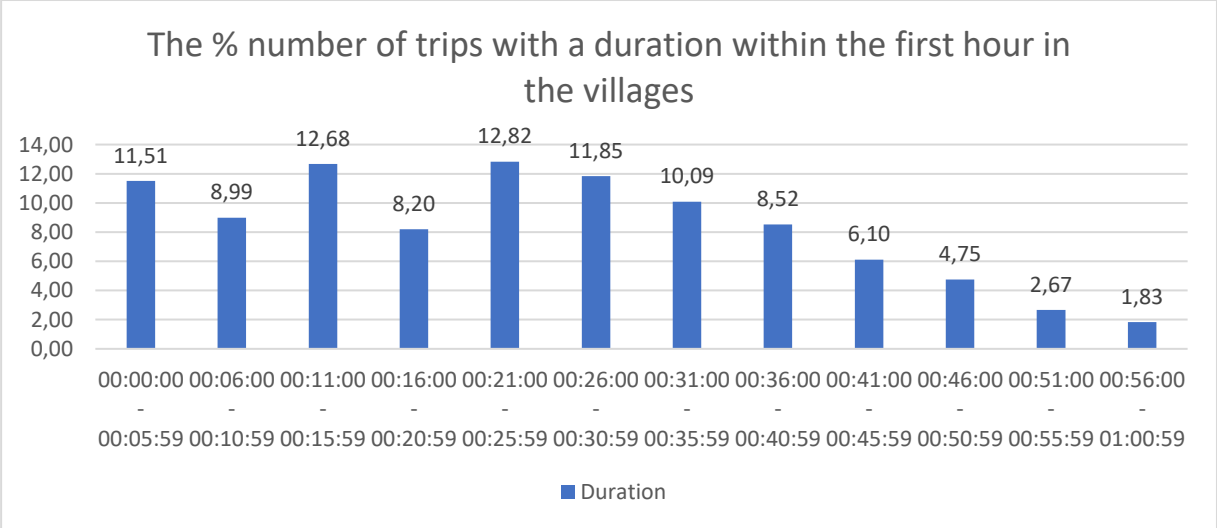


Figure 12: The percentiles of the subsample of trips less than an hour

		City	Village
N	Valid	11594	26376
	Missing	0	0
Quartiles	25	00:09:49	00:11:59
	Median	00:21:06	00:24:08
	50	00:33:58	00:35:20

At figure 12 and 13 you see the same difference as we saw earlier. In the cities the biggest part of the trips is between 1 and 7,5 km long, whereas in the villages it is between 7,5 and 15 km. Two other distances where there is a difference are the 0 till 1 and 15 till 30 distances. In the villages the share of trips that are between 15 and 30 km is higher compared to the cities. And vice versa in regards to the 0 till 1 km trips. This is in line with the literature and hypothesis 1 that people in cities cycle more shorter trips than people in villages. The median and 75% quartile also reflect this. The median in cities is 6,91 km per trip and in villages it is 8,58 km per trip, whereas the 75% quartile respectfully 11,68 km (cities) and 13,26 km (villages) are.

Figure 13

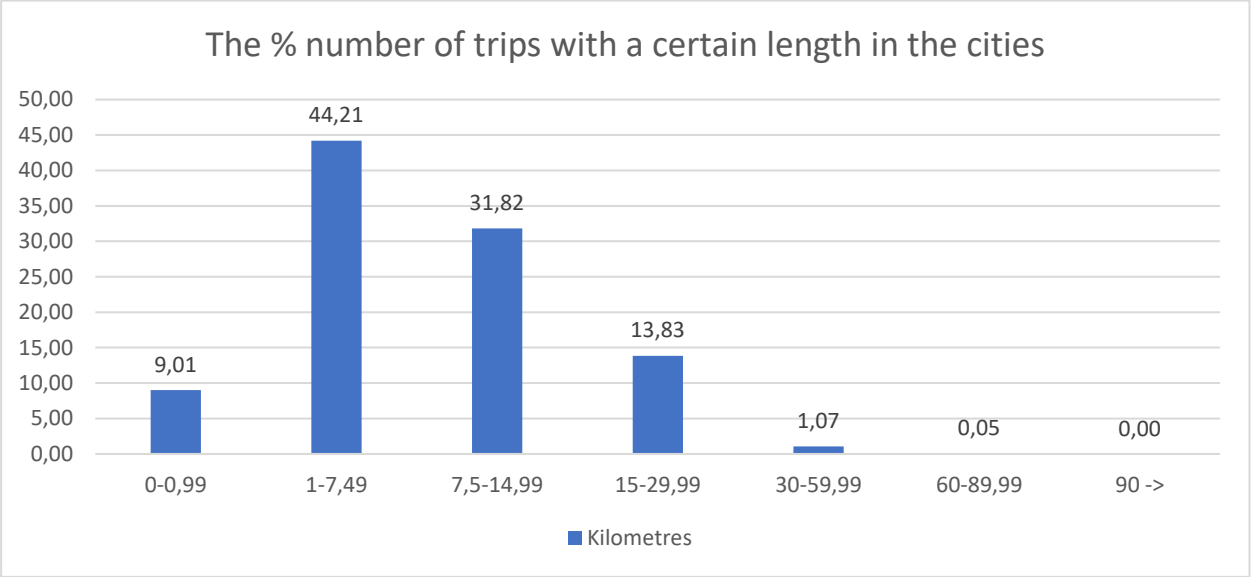


Figure 14

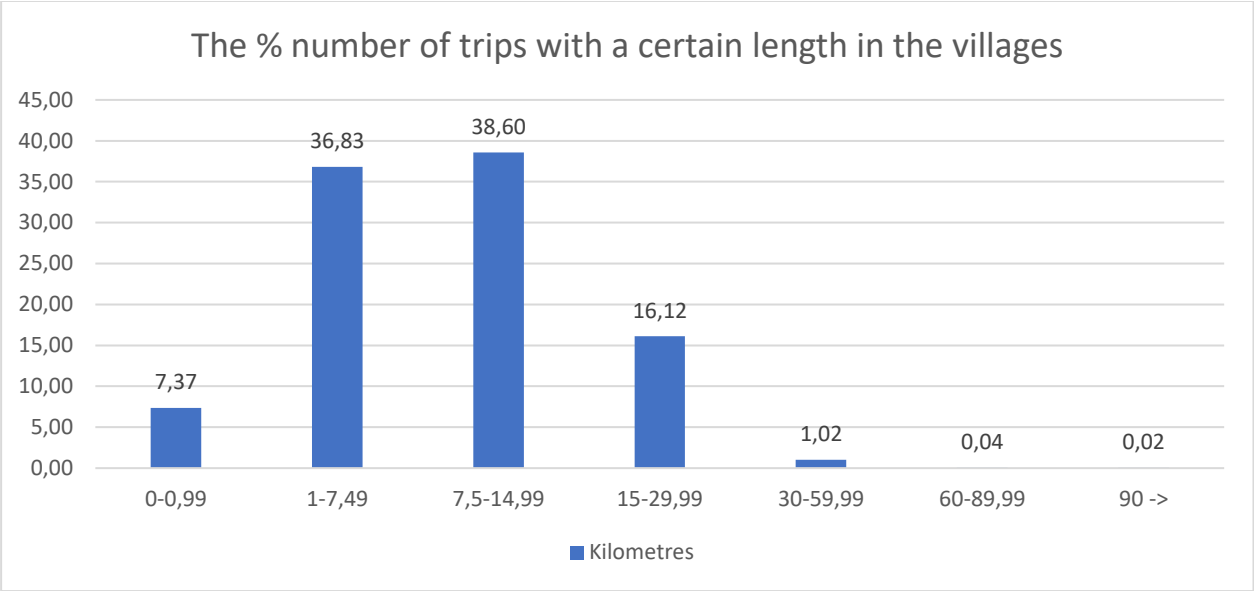


Figure 15: The percentiles of the length of the trips

		City	Village
N	Valid	11594	26376
	Missing	0	0
Quartiles	25	2,58	3,86
	Median	6,91	8,58
	50	11,68	13,26

4.2 The outcome of the t-tests and cross tabulation

To research the two hypotheses, I first did two different t-tests, one for the frequency of the trips (number of trips) and one for the variables duration, duration first hour and length. To analyse the variable purpose, I used a crosstabulation table with the Pearson Chi-Square.

4.2.1 Do people in cities cycle more frequent, but shorter trips than people in villages?

The first hypothesis is that people who live in cities make more frequently trips but these trips are shorter than the trips made by people who live in villages. This hypothesis consists of two parts. The first part regards the trip frequency and the second part the length and duration of the trip. To research this I did two different t-tests.

The first model shows the relationship of the independent variable (location) on the dependent variable (trip frequency). This was not statistically significant (p=0,477), so there is no direct relation between trip frequency and location.

Figure 16: T-test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Number	Equal variances assumed	0,506	0,477	-0,772	534	0,440	-0,09800	0,12700	-0,34600	0,15100
	Equal variances not assumed			-0,757	287,372	0,449	-0,09800	0,12900	-0,35200	0,15600

The second model shows the relationship of the independent variable location on the dependent variable duration, the subsample of trips of less than an hour and length. All variables are significant as can be seen in the table below. So, there is a direct relationship between location and duration, location and duration in the first hour and location and length. The t-value for duration is 9,631 which means that the longer the trips are the more chance there is that the trip was cycled by someone who lives in a village. The t-value for duration in the first hour is 8,675 and thus, again, means that the if the duration is longer, but stays within the first hour the chances that the trip was cycled by someone who lives in a village are higher than if that person lives in a city. The t-value of the variable length is also positive (13,143) so the same goes for this variable.

Figure 17: T-test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Duration	Equal variances assumed	11,207	0,001	9,557	37988	0,000	0,13194	0,01381	0,10488	0,15900
	Equal variances not assumed			9,631	22608,845	0,000	0,13194	0,01370	0,10509	0,15879
Dur_hour	Equal variances assumed	19,084	0,000	8,807	36556	0,000	0,29656	0,03367	0,23056	0,36256
	Equal variances not assumed			8,675	20885,358	0,000	0,29656	0,03418	0,22955	0,36356
Length	Equal variances assumed	9,998	0,002	13,189	37988	0,000	0,12862	0,00975	0,10951	0,14773
	Equal variances not assumed			13,143	22017,084	0,000	0,12862	0,00979	0,10944	0,14780

This means that the first hypothesis is partly true. The first part, people who live in cities make more frequently trips, does not hold because the model is not significant. The second part, people who live in cities make shorter trips than people who live in villages, however does hold. All three variables are significant and they have a positive relationship. This makes sense because facilities, in general, are further apart in villages.

4.2.2 Do people in cities use cycling for different kind of purposes than people in villages?

The second hypothesis is that the purpose of people who cycle in villages is mostly work-home related and that the purpose of people who cycle in cities varies more. For this hypothesis I tested the relationship of the dependent variable purpose on the independent variable location with a crosstabulation table and the Chi-Square. The crosstabulation table shows that 69,4% of the total trips are made in the villages and 30,6% in the cities. The biggest different is at the purpose study, where 72% of the trips are made in the villages and 28% in the cities.

Before looking at the statistics it is important to check whether the chi-square assumption is met. The assumption is that in a crosstabulation table, all expected frequencies should be greater than 5 (Field, 2009). One purpose does not meet this assumption, namely the purpose unknown (city) has an expected count of 1,2. So I decided to exclude this purpose for this model, because when this assumption is not met, it is advised to collect more data. This research, however, uses an already existing dataset so this option was not feasible. If I would include the purpose unknown, the result can fail to detect a genuine effect (Field, 2009). Now the lowest expected count is 28,4 (study in city), which is higher than 5 and thus the analyse can be continued.

As can be seen in the table below, the Chi-Square is significant thus there is a relation between the purposes of the trip and the location of the respondents. The crosstabulation table in figure 18 shows the percentages of each purpose per location and within all purposes combined. The second hypothesis holds; however, it is not possible to see if there is a difference in the varies purposes and if some purposes are perhaps more important than others. This can hopefully be resolved, with the binary logistic regression in the next chapter.

Figure 18: Chi-square test of the variable purpose

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	42,538[a]	9	0,000	.[b]	
Likelihood Ratio	41,850	9	0,000	.[b]	
Fisher's Exact Test	.[b]			.[b]	
Linear-by-Linear Association	22,672	1	0,000	.[b]	.[b]
N of Valid Cases	37991				

a. 2 cells (10,0%) have expected count less than 5. The minimum expected count is 1,22.

b. Cannot be computed because there is insufficient memory.

Figure 19: Cross-tabulation table of the variable purpose

			Location		Total
			Village	City	
Purpose	Home	Count	9723	4114	13837
		Expected Count	9606,6	4230,4	13837,0
		% within Purpose	70,3%	29,7%	100,0%
		% within Location	36,9%	35,4%	36,4%
		% of Total	25,6%	10,8%	36,4%
		Standardized Residual	1,2	-1,8	
	Paid work	Count	9604	4072	13676
		Expected Count	9494,8	4181,2	13676,0
		% within Purpose	70,2%	29,8%	100,0%
		% within Location	36,4%	35,1%	36,0%
		% of Total	25,3%	10,7%	36,0%
		Standardized Residual	1,1	-1,7	
	Non-daily groceries	Count	944	492	1436
		Expected Count	997,0	439,0	1436,0
		% within Purpose	65,7%	34,3%	100,0%
		% within Location	3,6%	4,2%	3,8%
		% of Total	2,5%	1,3%	3,8%
		Standardized Residual	-1,7	2,5	
	Daily groceries	Count	1940	951	2891
		Expected Count	2007,1	883,9	2891,0
		% within Purpose	67,1%	32,9%	100,0%
		% within Location	7,4%	8,2%	7,6%
		% of Total	5,1%	2,5%	7,6%
		Standardized Residual	-1,5	2,3	
	Social	Count	2298	1072	3370
		Expected Count	2339,7	1030,3	3370,0
		% within Purpose	68,2%	31,8%	100,0%
		% within Location	8,7%	9,2%	8,9%
		% of Total	6,0%	2,8%	8,9%
		Standardized Residual	-0,9	1,3	

			Location		Total
			Village	City	
Purpose	Recreation	Count	1121	517	1638
		Expected Count	1137,2	500,8	1638,0
		% within Purpose	68,4%	31,6%	100,0%
		% within Location	4,3%	4,5%	4,3%
		% of Total	3,0%	1,4%	4,3%
		Standardized Residual	-0,5	0,7	
	Spare time	Count	559	319	878
		Expected Count	609,6	268,4	878,0
		% within Purpose	63,7%	36,3%	100,0%
		% within Location	2,1%	2,7%	2,3%
		% of Total	1,5%	0,8%	2,3%
		Standardized Residual	-2,0	3,1	
	Services	Count	117	51	168
		Expected Count	116,6	51,4	168,0
		% within Purpose	69,6%	30,4%	100,0%
		% within Location	0,4%	0,4%	0,4%
		% of Total	0,3%	0,1%	0,4%
		Standardized Residual	0,0	-0,1	
	Study	Count	67	26	93
		Expected Count	64,6	28,4	93,0
		% within Purpose	72,0%	28,0%	100,0%
		% within Location	0,3%	0,2%	0,2%
		% of Total	0,2%	0,1%	0,2%
		Standardized Residual	0,3	-0,5	
	Unknown	Count	3	1	4
		Expected Count	2,8	1,2	4,0
		% within Purpose	75,0%	25,0%	100,0%
		% within Location	0,0%	0,0%	0,0%
		% of Total	0,0%	0,0%	0,0%
		Standardized Residual	0,1	-0,2	
Total		Count	26376	11615	37991
		Expected Count	26376,0	11615,0	37991,0
		% within Purpose	69,4%	30,6%	100,0%
		% within Location	100,0%	100,0%	100,0%
		% of Total	69,4%	30,6%	100,0%

4.3 The outcome of the ordinal logistic regressions and the binary logistic regression.

In this chapter, I used three ordinal logistic regression analyses and a binary logistic regression to try to get a deeper analysis of the two hypotheses. The settings for these analyses were equal. The method was the enter procedure, so all variables were assessed on the same time. The cut value was 0,5, the alpha level 0,05 and the level of removal 0,10. The Hosmer and Lemeshow goodness-of-fit test and Nagelkerke R-square were used to assess the fit of the model. The outcome of all the analyses can be found in the appendix.

4.3.1 Do people in cities cycle more frequent, but shorter trips than people in villages?

The first hypothesis is that people who live in cities make more frequently trips but these trips are shorter than the trips made by people who live in villages. This hypothesis consists of two parts. The first part regards the trip frequency and the second part the length and duration of the trip.

The first model shows the trip frequency. As in the previous chapter, this model is also not significant (p=0,434). So, there is no direct relationship between trip frequency and the location of the cyclists.

Figure 20: Logistic regression table of the location regressed over trip frequency

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Number = 1]	-1,777	0,169	110,693	1	0,000	-2,108	-1,446
	[Number = 2]	-0,288	0,147	3,870	1	0,049	-0,576	-0,001
	[Number = 3]	0,807	0,151	28,747	1	0,000	0,512	1,102
	[Number = 4]	2,018	0,181	124,806	1	0,000	1,664	2,372
	[Number = 5]	3,328	0,272	149,801	1	0,000	2,795	3,861
	[Number = 6]	4,235	0,397	113,559	1	0,000	3,456	5,013
	[Number = 7]	5,497	0,718	58,670	1	0,000	4,090	6,903
Location	Village	-0,132	0,168	0,613	1	0,434	-0,461	0,198
	City	0[a]			0			

Link function: Logit.

a. This parameter is set to zero because it is redundant.

The second model consists of the variable duration. This model is significant (p=0,000) so there is a direct relationship between duration and the location of the cyclists. The relation is positive (estimate is 0,230) which means that the longer the duration of the trips is, the higher the chance is that the cycler lives in a village.

Figure 21: Logistic regression table of the location regressed over duration of the trips

		Estimate	Std. Error	Wald	df	Sig.	Interval	
							Lower Bound	Upper Bound
Threshold	[Duration = 1,00]	-0,512	0,018	847,808	1	0,000	-0,547	-0,478
	[Duration = 2,00]	0,777	0,018	1893,695	1	0,000	0,742	0,812
	[Duration = 3,00]	2,082	0,021	9693,930	1	0,000	2,041	2,124
	[Duration = 4,00]	3,405	0,031	12328,602	1	0,000	3,345	3,465
	[Duration = 5,00]	4,241	0,043	9887,170	1	0,000	4,158	4,325
	[Duration = 6,00]	4,989	0,060	7024,993	1	0,000	4,872	5,105
	[Duration = 7,00]	5,669	0,082	4771,818	1	0,000	5,508	5,830
	[Duration = 8,00]	6,231	0,108	3345,004	1	0,000	6,020	6,442
	[Duration = 9,00]	6,903	0,150	2120,911	1	0,000	6,609	7,196
	[Duration = 10,00]	7,213	0,175	1703,308	1	0,000	6,870	7,556
	[Duration = 11,00]	7,574	0,209	1312,213	1	0,000	7,164	7,984
	[Duration = 12,00]	8,002	0,259	957,001	1	0,000	7,495	8,509
	[Duration = 13,00]	8,631	0,354	594,769	1	0,000	7,937	9,324
	[Duration = 14,00]	9,612	0,578	276,954	1	0,000	8,480	10,744
	[Duration = 15,00]	10,017	0,707	200,592	1	0,000	8,631	11,403
Location	Village	0,230	0,020	129,821	1	0,000	0,191	0,270
	City	0[a]			0			

Link function: Logit.

a. This parameter is set to zero because it is redundant.

The third model analyses the variable duration in the first hour and this variable is also significant (p=0,000). This relationship is also positive, thus the longer the duration of the trips is, within the first hour, the higher the chance that the cycler lives in a village. This is consistent with the t-tests in the previous chapter.

Figure 22: Logistic regression table of the location regressed over the subsample of trips in the first hour

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Dur_hour = 1,00]	-1,844	0,021	7868,260	1	0,000	-1,884	-1,803
	[Dur_hour = 2,00]	-0,980	0,018	2942,555	1	0,000	-1,015	-0,945
	[Dur_hour = 3,00]	-0,470	0,017	730,445	1	0,000	-0,504	-0,436
	[Dur_hour = 4,00]	-0,036	0,017	4,317	1	0,038	-0,069	-0,002
	[Dur_hour = 5,00]	0,398	0,017	528,065	1	0,000	0,364	0,432
	[Dur_hour = 6,00]	0,875	0,018	2413,876	1	0,000	0,840	0,910
	[Dur_hour = 7,00]	1,339	0,019	5142,747	1	0,000	1,302	1,375
	[Dur_hour = 8,00]	1,857	0,020	8438,265	1	0,000	1,818	1,897
	[Dur_hour = 9,00]	2,414	0,023	11181,829	1	0,000	2,369	2,459
	[Dur_hour = 10,00]	3,121	0,028	12294,008	1	0,000	3,066	3,176
	[Dur_hour = 11,00]	4,051	0,040	10171,722	1	0,000	3,972	4,130
Location	Village	0,207	0,020	109,684	1	0,000	0,168	0,246
	City	0[a]			0			

Link function: Logit.

a. This parameter is set to zero because it is redundant.

The last model consists of the variable length and shows also a significant relationship (p=0,000). This is also a positive relationship so the longer the length, the higher the chance that the cycler lives in a village.

Figure 23: Logistic regression table of the location regressed over the length of the trips

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Length = 1,00]	-2,261	0,023	9384,294	1	0,000	-2,307	-2,215
	[Length = 2,00]	0,085	0,018	23,483	1	0,000	0,051	0,120
	[Length = 3,00]	1,834	0,020	8136,299	1	0,000	1,794	1,874
	[Length = 4,00]	4,717	0,052	8381,551	1	0,000	4,616	4,818
	[Length = 5,00]	7,625	0,209	1329,370	1	0,000	7,215	8,035
	[Length = 6,00]	8,969	0,409	481,929	1	0,000	8,168	9,770
Location	Village	0,298	0,021	209,684	1	0,000	0,258	0,338
	City	0[a]			0			

Link function: Logit.

a. This parameter is set to zero because it is redundant.

Thus, the outcomes are the same as in the previous chapter. The first part of the hypothesis does not hold, trip frequency is not significant so it is not possible to say that people who live in cities cycle more often. However, the second part of the hypothesis does hold. Duration, duration in the first hour and length are all significant and show a positive relation so the longer the trip (both in time and kilometres) the higher the chances that that person lives in a village. So, if the hypothesis would be changed to *people who live in villages cycle longer than people who live in cities*, the hypothesis would hold.

4.3.2 Do people in cities use cycling for different kind of purposes than people in villages?

The second hypothesis is that the purpose of people who cycle in villages is mostly work-home related and that the purpose of people who cycle in cities varies more. For this hypothesis I tested the relationship of the dependent variable purpose on the independent variable location. The variable purpose is a categorical variable, with ten different categories. I decided to separate each category and made it a dummy variable by either assigning a yes if the trip does have that purpose or a no if the trip does not have that purpose. The independent variable is also a categorical variable, so for this model it is necessary to do an extra step in SPSS, namely to put the variable in the categorical covariates box so SPSS can make it a dummy variable (Field, 2009). Now every purpose will be assessed individually.

The first model shows a predicted value of 63,6% and the overall model was significant (p=0,000). If the purpose home is entered in the model, it is still significant (p=0,007) and the odds that a trip to home is made by someone who lives in a village is 1,065 times greater than by someone who lives in the city as can be seen in the table below. Thus, the purpose home significantly contributes to explaining whether someone lives in a village or a city.

Figure 24: Binary logistic regression table of the location regressed over the purpose home

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Location(1)	0,063	0,023	7,254	1	0,007	1,065	1,017	1,114
	Constant	-0,601	0,019	958,503	1	0,000	0,548		

a. Variable(s) entered on step 1: Location.

The second model shows a predicted value of 64,0% and the overall model was significant (p=0,000). If the purpose paid work is entered in the model, it is still significant (p=0,011) and the odds that a trip to their paid job is made by someone who lives in a village is 1,061 times greater than by someone who lives in the city as can be seen in the table below. Thus, the purpose paid work significantly contributes to explaining whether someone lives in a village or a city. The outcomes are almost the same as the purpose home, which is not strange because most trips where made either to their job or back home.

Figure 25: Binary logistic regression table of the location regressed over the purpose paid work

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Location(1)	0,059	0,023	6,414	1	0,011	1,061	1,013	1,110
	Constant	-0,616	0,019	1005,029	1	0,000	0,540		

a. Variable(s) entered on step 1: Location.

The third model shows a predicted value of 96,2% and the overall model was significant (p=0,000). If the purpose non-daily groceries is entered in the model, it is still significant (p=0,002) but the relationship is negative. So, the odds that a trip is used to do non-daily groceries by someone who lives in a village is smaller than by someone who lives in the city as can be seen in the table below. Thus, the purpose non-daily groceries significantly contributes to explaining whether someone lives in a village or a city.

Figure 26: Binary logistic regression table of the location regressed over the purpose non-daily groceries

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Location(1)	-0,175	0,057	9,545	1	0,002	0,839	0,751	0,938
	Constant	-3,118	0,046	4581,432	1	0,000	0,044		

a. Variable(s) entered on step 1: Location.

The fourth model shows a predicted value of 92,4% and the overall model was significant (p=0,000). If the purpose daily groceries is entered in the model, it is still significant (p=0,05) and the relationship is also negative. So, the odds that a trip is used to do daily groceries by someone who lives in a village is smaller than by someone who lives in the city as can be seen in the table below. Thus, the purpose daily groceries significantly contributes to explaining whether someone lives in a village or a city. It is interesting to see that doing groceries have a negative relationship to living in a village. This can be explained because there are more stores in cities usually so people spend more trips for shopping (both daily and non-daily groceries).

Figure 27: Binary logistic regression table of the location regressed over the purpose daily groceries

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Location(1)	-0,116	0,041	7,942	1	0,005	0,890	0,821	0,965
	Constant	-2,417	0,034	5101,243	1	0,000	0,089		

a. Variable(s) entered on step 1: Location.

The models where the predicted values were between 91,1% and 100% and the overall model was significant (p=0,000) but where not significant anymore when the variable were added were, the purpose social, recreation, services, study and unknown. Respectively models 5, 6, 8, 9 and 10. So there is no relationship between these individual variables and the location of the cyclists.

The seventh model shows a predicted value of 97,7% and the overall model was significant (p=0,000). If the purpose spare time is entered in the model, it is still significant (p=0,000) and this relationship is also negative. So, the odds that a trip is made to do something in their spare time by someone who lives in a village is smaller than by someone who lives in the city as can be seen in the table below. Thus, the purpose spare time significantly contributes to explaining whether someone lives in a village or a city. It is not surprising that the relationship between spare time and the location (village) is negative because cities have usually more entertainment like, cinemas, restaurants, theatres etc than villages so people who live in cities can more easily use them.

Figure 28: Binary logistic regression table of the location regressed over the purpose spare time

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Location(1)	-0,266	0,071	13,969	1	0,000	0,767	0,667	0,881
	Constant	-3,567	0,057	3947,349	1	0,000	0,028		

a. Variable(s) entered on step 1: Location.

So, of all the purposes only half are significant and have a direct relationship with the location of the cyclists, these are home, paid work, non-daily groceries, daily groceries and spare time. The purposes home and paid work were the only one with a positive relationship, this means that people who cycle to either home or work have a higher chance of living in a village in North-Brabant. Non-daily groceries, daily groceries and spare time have a negative relationship, thus there was a higher chance that people who cycled a trip with this purpose are living in a city in North-Brabant. One possible reason behind this can be that facilities are close-by in cities so people are more inclined to separate purposes per trip and that people in villages combine their trips more. Thus, the hypothesis is true. It would be interesting to do a follow-up research to find out if people in villages indeed combine their trips more than people in cities.

5. Conclusion

This thesis aimed to research the influence of location on the trip length, duration and purpose of e-bike users in North Brabant. It became apparent that researches on (e-)bicycle usage and the location of the user was limited; especially the question whether there is a difference in usage between urban and rural cyclists. This chapter will discuss the results of the previous chapter and link it to the literature.

The data was derived from the Bicycle Stimulation Program Brabant (BSP). They followed 581 participants from September 2013 until October 2014. These participants used the app B-Riders to follow their cycling trips and the purpose of these trips. The participants were paid to keep track of their trips and signed in for the program themselves (Timmermans & Feng, n.d.). Of these 581 participants, 538 were allegeable for my research and used in this thesis.

One of the reasons why some participants didn't partake in my thesis was that their home location was not in Brabant. Due to the fact that location the most important factor was for my research, I first had to research what the general difference of rural and urban space was. Rural space was long seen as a servant of the urban space and in modern time as a declining space due to the large migration to cities. Although this migration is still happening, rural areas renewed themselves and thereby revitalized the rural space. Agriculture, for example, is not only used for production but also to give an experience of the rural landscape to urban and foreign visitors (Galani-Moutafi, 2013). The rural represents an "*idealization of the rural and a nostalgia for a simpler way of life*" (Galani-Moutafi, 2013). On the other hand, urbanism is, according to Halfacree (1993), "*characterized as being dynamic, unstable, mobile in stratification and impersonal, with contacts being determined by one's precise situation at the time (work, home, leisure)*". Although this still holds, there are many other versions of urban space. As for versions where rural and urban are so much intertwined that it is not clear where urban begins and rural ends. That is why it is in this age almost impossible to make a clear distinction between the two. One way to still do that is to look up which cities have city rights and consider all the other places as villages. This method was used for this thesis.

5.1 The different frequencies of the trips

The first hypothesis *people who live in cities make more frequently trips but these trips are shorter than the trips made by people who live in villages* was based on researches by Heinen et al. (2010), Rietveld & Daniel (2004) and Pucher & Buehler (2010). They stated that people in cities cycle more for a couple of reasons, firstly because facilities are close by so it is easier to make a couple of different trips per day instead of combining it within one trip. Secondly, there is more traffic in cities which causes congestion so it can be quicker to use a bicycle. Thirdly, there are more policies in city centres that prohibits cars or make it expensive to park so cycling is easier and cheaper. Fourthly, in rural areas there is more space for car parking and thus an extra incentive to use the car. Small- and medium-sized cities (which most cities in Brabant are) are the best for bicycle use due to their geographic size which "*may be naturally more supportive of cycling or at least more easily modified*" (Pucher & Buehler, 2010). Larger cities or metropolitan areas do have the advantages as smaller cities with regard to the close by facilities, the car congestion and the policies against cars but bicycles also compete with public transport and the fear of bicycle theft which makes not only bicycle use less attractive but it also has a negative effect on the ownership of e-bikes (Zhang et al., 2013). So, the target areas for planners and city councils for promoting e-bikes are rural areas, because car use is still the major transportation use, and small- and medium-sized cities, because inhabitants of these cities are already positive towards 'normal' cycling.

This pattern was also seen in the outcome of the t-tests and logistic regression model. Both analyses showed that the first hypothesis was partly true. The t-tests showed that the first part, people who live in cities make more frequently trips, did not hold because the model was not significant. The second part, people who live in cities make shorter trips than people who live in villages, however did hold. All three variables were significant and they had a positive relationship. This makes sense because facilities, in general, are further apart in villages so they have to cycle longer to get there. The logistic regression model showed the same, the first part of the hypothesis did not hold, trip frequency was not significant so it is not possible to say that people who live in cities cycle more often. However, the second part of the hypothesis did hold. Duration, duration in the first hour and length were all significant and showed a positive relation, so the longer the trip (both in time and kilometres) the higher the chances that that person lives in a village. So, after analysing the data it can be concluded that if the hypothesis would be changed to *e-cyclists who live in villages cycle longer than e-cyclists who live in cities*, the hypothesis would hold in both models and corresponds with previous researches.

5.2 The different purposes of the trips

The second hypothesis *the purpose of people who cycle in villages is mostly work-home related and that the purpose of people who cycle in cities varies more* was mainly based on the researches of Pucher & Buehler (2010) and Rietveld & Daniel (2004). They stated that people in cities have more different options for recreational use, like restaurants, theatre, cinemas, which makes it easier for people in cities to go there regularly and that these facilities are more within bicycle reach in cities. Another reason is that city municipalities try to ban cars from the inner cities so when people go out, they are more inclined to use their bicycle. In graphs 1 and 2 this pattern, although not overwhelmingly clear, was also shown. There was a difference visible for groceries, daily and non-daily, social, recreation and spare time. People in the cities made more trips with this purpose.

For the second hypothesis I used a crosstabulation table and a binary logistic regression per purpose. In the first model, the Chi-Square was significant thus there is a relation between the purposes of the trip and the location of the respondents. This means that the second hypothesis holds, however it is not possible to see if there is a difference in the varies purposes and if some purposes are perhaps more important than others. This is why I also did a binary logistic regression. The outcome of that model was more in depth than the crosstabulation table. Of all the purposes only, half were significant and had a direct relationship with the location of the cyclists, these were home, paid work, non-daily groceries, daily groceries and spare time. The purposes home and paid work were the only one with a positive relationship, this means that people who cycle to either home or work have a higher chance of living in a village in North-Brabant. Non-daily groceries, daily groceries and spare time have a negative relationship, thus there was a higher chance that people who cycled a trip with this purpose are living in a city in North-Brabant. One possible reason behind this can be that facilities are close-by in cities so people are more inclined to separate purposes per trip and that people in villages combine their trips more. Thus, the hypothesis is true in both models and corresponds with previous researches.

Thus, the GPS study conducted in Brabant is comparable to other, stated behaviour, researches and can be used to support other researches with the same subject.

5.3 Future researches

Both hypotheses were significant, although the first hypothesis should be slightly adjusted, which means that my research can be used to support previous researches. This is an asset because there are not (yet) many researches that use GPS data and the previous researches used in the theoretical framework were all based on stated behaviour (obtained by surveys).

There are three angles that makes the outcome of this research interesting for next researches. The first one is off course, the roll of e-bikes. E-bikes are increasing in popularity, especially in countries in Asia and this can *change traffic dynamics* (Dozza et al., 2016). An interesting question would be how the infrastructure can be changed that even more people are willing to cycle more or further from their homes. Another interesting question would be to compare e-bike usage with normal bicycle usage and see if there are any differences. Or to combine GPS data with surveys to discover more about the reasoning of why people make the decision they make concerning cycling. And last to do this study with an unbiased GPS track dataset.

The second angle is to measure urbanity differently. In my thesis I saw urban and rural areas, on purpose, as independent spaces that do not overlap, but this is not true in the real world. If my research could be expanded to a research where the intertwining between them is also taken into account and the different levels of urbanity, then it would give a more in-depth and realistic view of how e-bike cycles behave. However, it doesn't only have implications for new researches but also for development policies. Infrastructure, revenues, costs and public service also need to be more intertwined to benefit all the regions that are connected (Irwin et al., 2009).

The third angle is to improve GPS data and to use my thesis as an example of how to analyse GPS data. GPS will never be perfect and there are some major implications. The first one, is that the reception is not ideal in urban settings, especially when people walk in-and-out of buildings, reception can get lost or reflections of other buildings can confuse the signal. And the second one, is the improvement of the software to increase the speed of data processing. Nowadays a determination of research is the limitation of the scale, if software to process the data would be upgraded this would be less of a concern (Carlson et al., 2014; Spek at al., 2009). However, if reception and process software will be improved more, researches can use more real-time data and have a more realistic view of processes.

5.4 Reflection on the research

This research was a lot harder than I expected beforehand. Specially to learn Python and work with it to get the parts of data I needed was much more difficult. I appreciate the help of my supervisor, Simon Scheider, for helping me with this, because otherwise this thesis wouldn't be here.

Another part where I struggled with was my previous experience with my bachelor thesis. It took me 4 different theses to finally get one that was sufficient to pass and this was my biggest fear to happen this time as well. This held me back to get the help I needed and to just do it and write.

The subject, however, was super interesting. I did not know a lot about cycling let alone of e-bikes and the insights I got from working on this thesis were really interesting and I learned a lot. So, in the end I am content with my choice for this thesis subject and to work, for the first time, with such a large GPS data sample and to learn (the beginnings) of coding with Python.

6. Literature

Acker, V. van, Wee, B. van, Witlox, F. (2010). When transport geography meets social psychology: toward a conceptual model of travel behaviour. *Transport Reviews,* 30 (2), 219–240.

Bikeriders. (N.D.) *Bikeprint: Een innovatieve schakel tussen onderzoek en fietsbeleid*, http://www.bikeprint.nl/ (first visite 05-02-2017).

Bosworth, G. & Venhorst, V. (2015) Economic linkages between urban and rural regions – what's in it for the rural? *Paper prepared for the Regional Studies Association Annual Conference 2015,* 1-16.

Bryman, A. (2008), *Social research methods*. University press, Oxford.

Buehler, R., & Pucher, J. (2012). Big City Cycling in Europe, North America, and Australia. *Pucher, J, Buehler, R. eds*, 287-318.

Carlson, J.A. et al. (2015). Association between neighborhood walkability and GPS-measured walking, bicycling and vehicle time in adolescents. *Health & Place,* 32, 1-7.

Cox, J. (N.D.). *Repertorium Van De Stadsrechten In Nederland,* VNG-uitgeverij: Den Haag.

Dozza, M., Piccinini, G. F. B., & Werneke, J. (2016). Using naturalistic data to asses e-cyclist behaviour. *Transportation research,* 41(F), 217-226.

Ehrgott, M., Wang, J.Y.T., Raith, A., & Houtte, C. van. (2012). A bi-objective cyclist route choice model. *Transportation Research Part A,* 46, 652-663.

Feng, T. & Timmermans, H. J. P. (N.D.) Using Recurrent Spatio-Temporal Profiles in GPS Panel Data for Enhancing Imputation of Activity Type. *Urban planning group of the Eindhoven University of Technology,* 1-12.

Field, A. (2009). *Discovering statistics using SPSS,* Sage: London.

Fietsberaad. (2013). *Feiten over de elektrische fiets,* Utrecht: Fietsberaad.

Fishman, E., & Cherry, C. (2016). E-bikes in the Mainstream: Reviewing a Decade of Research, *Transport Reviews,* 36(1), 72-91.

Fyhri, A., & Fearnley, N. (2015). Effects of e-bikes on bicycle use and mode share. *Transportation Research Part D,* 36, 45-52.

Galani-Moutafi, V. (2013). Rural space (re)produced – Practices, performances and visions: A case study from an Aegan island. *Journal of Rural Studies,* 32, 103-113.

Garrard, J., Rose, G. & Lo, S. K. (2008). Promoting transportation cycling for women: the role of bicycle infrastructure. *Preventive Medicine,* 46(1), 55–59.

GPS coördinaten (N.D.) *Google Maps GPS Coördinaten,* https://www.gps-coordinaten.nl/ (first visited 16-04-2017).

Halfacree, K.H. (1993). Locality and Social Representation: Space, Discourse and Alternative Definitions of the Rural. *Journal of Rural Studies,* 9(1), 23-37.

Halfacree, K.H. (2004). Rethinking 'rurality'. In: Champion, T., Hugo, G. (Eds.), New Forms of Urbanization; Beyond the Urban-Rural Dichotomy. Ashgate, Aldershot, pp. 285–304.

Harms, L., Bertolini, L., & te Brömmelstroet, M. (2014). Spatial and social variations in cycling patterns in a mature cycling country exploring differences and trends. *Journal of Transport & Health,* 1, 232-242.

Heatherington, T. (2011). Remaking rural landscapes in twenty-first century Europe. *Anthropological Journal of European Cultures,* 20(1), 1-9.

Heinen, E., Van Wee, B., & Maat, K. (2010). Commuting by bicycle: An overview of the literature. *Transport Reviews*, 30(1), 59-96.

Hendriksen, I., Engbers, L., Schrijver, J., Gijlswijk, R. V., Weltevreden, J., & Wilting, J. (2008). *Elektrisch Fietsen: marktonderzoek en verkenning toekomstmogelijkheden* (No. KvL/B&G 2008.067). TNO.

Irwin, E. G., Bell, K. P., Bockstael, N., Newburn, D. A., Partridge, M. D., & Wu, J. (2009). The Economics of Urban-Rural Space. *Annual Review of Resource Economics,* 1(4), 35-59.

Jacobsen, P. L. (2003). Safety in numbers: more walkers and bicyclists, safer walking and bicycling, *Injury Prevention,* 9, 205-209.

Jamerson, F. E., & Benjamin, E. (2013). Electric bikes worldwide reports. In: Fishman, E., & Cherry, C. (2016). E-bikes in the Mainstream: Reviewing a Decade of Research, *Transport Reviews,* 36(1), 72-91.

Keijer, M. J. N., & Rietveld, P. (2000). How do people get to the railway station? The Dutch experience. *Transportation Planning and Technology*, 23(3), 215-235.

Krizek, K. J., & Johnson, P. J. (2006). Proximity to trails and retail: Effects on urban cycling and walking. *Journal of the American Planning Association,* 72(1), 33-42.

Litman, T. (2007) *Land Use Impacts on Transport: How Land Use Factors Affect Travel Behaviour* Victoria: Victoria Transport Institute.

MacArthur, J., Dill, J., & Person, M. (2014). Electric Bikes in North America: Results from an online survey. *Transportation Research Record: Journal of the Transportation Research Board*, 2468, 123–130.

Madsen, L. M., & Adriansen, H. K. (2004). Understanding the use of rural space: the need for multi-methods. *Journal of Rural Studies,* 20(4), 485-497.

Mahon, M. (2007). New populations; shifting expectations: The changing experience of Irish rural space and place. *Journal of Rural Studies,* 23, 345-356.

Marsden, P. V., & Friedkin, N. E. (1993) Network studies of social influence, *Sociological Methods & Research,* 22(1), 127-151. In: Pucher, J., & Buehler, R. (2012). *City cycling*. MIT Press: Cambridge, Mass.

Martens, K. (2004). The bicycle as a fendering mode: experiences from three European countries. *Transportation Research Part*, 9, 281-294.

Popovich, N., Gordon, E., Shao, Z., Xing, Y., Wang, Y., & Handy, S. (2014). Experiences of electric bicycle users in the Sacramento, California area. *Travel Behaviour and Society,* 1, 37-44.

Provincie Noord-Brabant. (2017) *De bevolkings- en woningbehoefteprognose Noord-Brabant: Actualisering 2017,* http://bevolkingsprognose.brabant.nl/ (first visited 20-06-2017).

Provincie Noord-Brabant (N.D.) *Feiten en cijfers bevolking en wonen,* http://www.brabant.nl/dossiers/dossiers-op-thema/bouwen-en-wonen/feiten-en-cijfers-wonen.aspx (first visite 3-07-2017).

Provincie Noord-Brabant. (2009). *Fiets in de Versnelling: Visie 2020 met actieprogramma,* Noord-Brabant: Provincie Noord-Brabant.

Pucher, J., & Buehler, R. (2012). *City cycling*. MIT Press: Cambridge, Mass.

Rietveld, P., & Daniel, V. (2004). Determinants of bicycle use: do municipal policies matter? *Transportation Research Part,* 38, 531-550.

Rose, G. E-Bikes and Urban Transportation: Emerging Issues and Unresolved Questions. *Transportation,* 39(1), 81–96.

Rybarczy, G., & Wu, C. (2010). Bicycle facility planning using GIS and multi-criteria decision analysis. *Applied Geography,* 30, 282-293.

Scheepers E., Wendel-Vos W., van Kempen E., Panis L. I., Maas J., et al. (2013). Personal and Environmental Characteristics Associated with Choice of Active Transport Modes versus Car Use for Different Trip Purposes of Trips up to 7.5 Kilometres in The Netherlands*. PLoS ONE, 8(9).*

Scheider, S., (N.D.) *Spatio-temporal behaviour of cyclists in Brabant: GIMA Modul 2 case study manual*, Utrecht: Utrecht University.

Spek, S. van der, Schaick, J. van, Bois, P. de, & Haan, R. de. (2009) Sensing Human Activity: GPS Tracking. *Sensors,* 9, 3033-3055.

Steinbach, R., Green, J., Datta, J., & Edwards, P. (2011). Cycling and the city: A case study of how gendered, ethnic and class identities can shape healthy transport choices. *Social Science & Medicine,* 72, 1123-1130.

Stinson, M. A. & Bhat, C. R. (2003) *An Analysis of Commuter Bicyclist Route Choice Using Stated Preference Survey,* Washington DC: Transportation Research Board.

Witlox, F. & Tindemans, H. (2004). Evaluating bicycle-car transport mode competitiveness in an urban environment: an activity-based approach, *World Transport Policy and Practice,* 10(4), 32–42.

Zhang, Y., Li, Y., Yang, X., Liu, Q., & Li, C. (2013). Built environment and household electric bike ownership: Insights from Zhongshan Metropolitan Area, China, *Transportation Research Record: Journal of the Transportation Research Board,* 102-111.

7. Appendix

7.1 Python scripts

These are the python scripts that were used for this thesis:

```python
#-------------------------------------------------------------------------------
# Name:        RuralUrbanCounter
# Purpose:
#
# Author:      simon
#
# Created:     10/07/2017
# Copyright:   (c) simon 2017
# Licence:     <your licence>
#-------------------------------------------------------------------------------

import numpy as np
import pandas as pd
from pandas import Series
import arcpy
import csv
import glob
import json

def file_to_str(fn, dt, sepp, headerl=None, lterminator='\n'):
    #dateparser = lambda dates: [pd.datetime.strptime(d, '%Y-%m-%d %H:%M:%S') for d in dates]
    """
    Loads the content of a text file into a DataFrame, where datetimes in columns dt are converted to timestamps
    @return a string
    """
    content = pd.read_csv(fn, index_col=False, lineterminator='\n', header = headerl, parse_dates=dt, dayfirst=True, sep=sepp)
    content = content.drop_duplicates()
    #print content[0:10]
    #with open(fn, 'r') as f:
    #    content=f.read()
    return content


def main():
    personstats = {}
    i = 0
    out = "C:/Users/didde/Documents/Didde/universiteit/Master/Scriptie/joinedtracks/Villages/stats.json"
    with open(out, 'w') as fp:
        for f in glob.glob("C:/Users/didde/Documents/Didde/universiteit/Master/Scriptie/joinedtracks/Villages/*.csv"):
            l = file_to_str(f, [2], ',', headerl = 0)
            person = str(l['person'].iloc[0])
            numberoftracks = l.track.value_counts().size
            #print numberoftracks
            durationtable = l.groupby(['track'])['datetime'].agg(['min','max'])
            durationtable['difference']= durationtable['max']-durationtable['min']
            durationtabled = {k:{str(k2):str(v2) for k2,v2 in v.items()} for k,v in durationtable.to_dict().items()}
##            for row in durationtable.itertuples():
##                durationtabled[row.index]= row.difference
                #break
            #print durationtable
            #print durationtable.difference.value_counts()
            #This extracts and counts the purposes of a track
            purposestats = (l.groupby('track')['purto'].agg('max')).value_counts()
            #print(purposestats)
            purposestatsd = {}
            for k in purposestats
                purposestatsd[str
            #Getting lengths
            tracklength ={}
            oldpoint = {}
            oldtrack = l['track'].iloc[0]
            dist = 0
        for row in l.itertuples():
            spatial_reference = arcpy.SpatialReference(4326)
            pnt_geometry = arcpy.PointGeometry(arcpy.Point(row.X, row.Y), spatial_reference)
            point = pnt_geometry.projectAs(arcpy.SpatialReference(28992))
            #print point.firstPoint.X
            #print point.firstPoint.Y
            if oldpoint != {} and row.track == oldtrack:
                dist += point.distanceTo(oldpoint)
                #print (str(row.track)+ ':' +str(dist))
                tracklength[str(row.track)]=str(dist)
            else:
                dist = 0
            oldtrack = row.track
            oldpoint = point
        #print  tracklength
        personstats[person]={'tracklength':tracklength, 'numberoftracks': numberoftracks, 'purposestats': purposestatsd, 'durationtable': durationtabled}
        #i+= 1
        #if i == 1:
        #print personstats
        print "person number: "+str(person)

    json.dump(personstats,fp)


# for row in durationtable.itertuples():
#  print(row.max -row.min)


 #for row in l.itertuples():
 # track_id = row[2]
 # print(track_id)
```

```
class str
Defined in module __builtin__
```

```
Local variable oldpoint of function main
Defined in module ruralurbancounter (61)
Type: dict
```

```
#-------------------------------------------------------------------------------
# Name:       module1
# Purpose:
#
# Author:      Didde
#
# Created:     03-09-2017
# Copyright:   (c) Didde 2017
# Licence:     <your licence>
#-------------------------------------------------------------------------------

import json
import pandas as pd
from datetime import datetime, timedelta

with open ("joinedtracks/Villages/stats.json", "rb") as f:
    stats = json.load(f)

tracks = [ ]
for person, cycle_stats in stats.items ( ):
    durations = cycle_stats['durationtable']['difference']
    track_lengths = cycle_stats['tracklength']
    for track, duration in durations.items():
        duration = datetime.strptime(duration, "%w days %H:%M:%S").time()
        try:
            track_length = track_lengths[track]
        except KeyError:
            continue
        else:
            tracks.append((track, person, duration, track_length))

df = pd.DataFrame(tracks, columns=["track", "person", "duration","track_length"]).set_index("track")
print df.head()
df.to_excel("test.xlsx")
```

7.2 Excel sheet

PURPOSE

Cities	**Amount**	**Percentage**	**Median**	**75 quartile**
Home	4108	35,43	4,34	8,96
Paid work	4066	35,07	4,34	8,96
Non-daily groceries	491	4,23	4,34	8,96
Daily groceries	949	8,19	4,34	8,96
Social	1068	9,21	4,34	8,96
Recreation	516	4,45	4,34	8,96
Spare time	318	2,74	4,34	8,96
Services	51	0,44	4,34	8,96
Study	26	0,22	4,34	8,96
Unknown	1	0,01	4,34	8,96
Total	11594	100		

Villages	**Amount**	**Percentage**	**Median**	**75 quartile**
Home	9723	36,86	3,91	8,37
Paid work	9604	36,41	3,91	8,37
Non-daily groceries	944	3,58	3,91	8,37
Daily groceries	1940	7,36	3,91	8,37
Social	2298	8,71	3,91	8,37
Recreation	1121	4,25	3,91	8,37
Spare time	559	2,12	3,91	8,37
Services	117	0,44	3,91	8,37
Study	67	0,25	3,91	8,37
Unknown	3	0,01	3,91	8,37
Total	26376	100		

LENGTH IN KM

Cities	Amount	Percentage	Median	75 quartile
0-0,99	1045	9,01	9,01	22,83
1-7,49	5126	44,21	9,01	22,83
7,5-14,99	3689	31,82	9,01	22,83
15-29,99	1604	13,83	9,01	22,83
30-59,99	124	1,07	9,01	22,83
60-89,99	6	0,05	9,01	22,83
90 ->	0	0,00	9,01	22,83
Total	11594	100		

Villages	Amount	Percentage	Median	75 quartile
0-0,99	1945	7,37	7,37	26,47
1-7,49	9713	36,83	7,37	26,47
7,5-14,99	10180	38,60	7,37	26,47
15-29,99	4251	16,12	7,37	26,47
30-59,99	270	1,02	7,37	26,47
60-89,99	11	0,04	7,37	26,47
90 ->	6	0,02	7,37	26,47
Total	26376	100		

NUMBER OF TRIPS

Cities	Amount	Percentage	Median	75 quartile
10-30	26	16,25	12,81	21,09
31-60	43	26,88	12,81	21,09
61-90	39	24,38	12,81	21,09
91-120	32	20,00	12,81	21,09
121-150	15	9,38	12,81	21,09
151-180	3	1,88	12,81	21,09
181-210	1	0,63	12,81	21,09
211->	1	0,63	12,81	21,09
Total	160	100		

Villages	Amount	Percentage	Median	75 quartile
10-30	58	15,43	11,17	19,61
31-60	115	30,59	11,17	19,61
61-90	100	26,60	11,17	19,61
91-120	65	17,29	11,17	19,61
121-150	26	6,91	11,17	19,61
151-180	7	1,86	11,17	19,61
181-210	4	1,06	11,17	19,61
211->	1	0,27	11,17	19,61
Total	376	100		

DURATION

Cities	Amount	Percentage	Median	75 quartile
00:00:00 - 00:15:59	4431	38,22	0,09	3,40
00:16:00 - 00:30:59	3494	30,14	0,09	3,40
00:31:00 - 00:45:59	2215	19,10	0,09	3,40
00:46:00 - 01:00:59	1110	9,57	0,09	3,40
01:01:00 - 01:15:59	156	1,35	0,09	3,40
01:16:00 - 01:30:59	104	0,90	0,09	3,40
01:31:00 - 01:45:59	48	0,41	0,09	3,40
01:46:00 - 02:00:59	12	0,10	0,09	3,40
02:01:00 - 02:30:59	9	0,08	0,09	3,40
02:31:00 - 03:00:59	5	0,04	0,09	3,40
03:01:00 - 03:30:59	5	0,04	0,09	3,40
03:31:00 - 04:00:59	4	0,03	0,09	3,40
04:01:00 - 04:30:59	0	0,00	0,09	3,40
04:31:00 - 05:00:59	1	0,01	0,09	3,40
05:01:00 - 05:30:59	0	0,00	0,09	3,40
05:31:00 - 06:00:59	0	0,00	0,09	3,40
Total	11594	100		

Villages	Amount	Percentage	Median	75 quartile
00:00:00 - 00:15:59	8392	31,82	0,17	4,04
00:16:00 - 00:30:59	8311	31,51	0,17	4,04
00:31:00 - 00:45:59	6250	23,70	0,17	4,04
00:46:00 - 01:00:59	2337	8,86	0,17	4,04
01:01:00 - 01:15:59	641	2,43	0,17	4,04
01:16:00 - 01:30:59	226	0,86	0,17	4,04
01:31:00 - 01:45:59	101	0,38	0,17	4,04
01:46:00 - 02:00:59	54	0,20	0,17	4,04
02:01:00 - 02:30:59	34	0,13	0,17	4,04
02:31:00 - 03:00:59	7	0,03	0,17	4,04
03:01:00 - 03:30:59	5	0,02	0,17	4,04
03:31:00 - 04:00:59	4	0,02	0,17	4,04
04:01:00 - 04:30:59	7	0,03	0,17	4,04
04:31:00 - 05:00:59	4	0,02	0,17	4,04
05:01:00 - 05:30:59	1	0,00	0,17	4,04
05:31:00 - 06:00:59	2	0,01	0,17	4,04
Total	26376	100		

DURATION FIRST HOUR

Cities	Amount	Percentage	Median	75 quartile
00:00:00 - 00:05:59	1502	13,35	8,44	10,91
00:06:00 - 00:10:59	1707	15,17	8,44	10,91
00:11:00 - 00:15:59	1221	10,85	8,44	10,91
00:16:00 - 00:20:59	1176	10,45	8,44	10,91
00:21:00 - 00:25:59	1246	11,08	8,44	10,91
00:26:00 - 00:30:59	1072	9,53	8,44	10,91
00:31:00 - 00:35:59	827	7,35	8,44	10,91
00:36:00 - 00:40:59	762	6,77	8,44	10,91
00:41:00 - 00:45:59	626	5,56	8,44	10,91
00:46:00 - 00:50:59	467	4,15	8,44	10,91
00:51:00 - 00:55:59	382	3,40	8,44	10,91
00:56:00 - 01:00:59	261	2,32	8,44	10,91
Total	11249	100		

Villages	Amount	Percentage	Median	75 quartile
00:00:00 - 00:05:59	2911	11,51	8,76	11,59
00:06:00 - 00:10:59	2274	8,99	8,76	11,59
00:11:00 - 00:15:59	3206	12,68	8,76	11,59
00:16:00 - 00:20:59	2073	8,20	8,76	11,59
00:21:00 - 00:25:59	3242	12,82	8,76	11,59
00:26:00 - 00:30:59	2996	11,85	8,76	11,59
00:31:00 - 00:35:59	2552	10,09	8,76	11,59
00:36:00 - 00:40:59	2155	8,52	8,76	11,59
00:41:00 - 00:45:59	1543	6,10	8,76	11,59
00:46:00 - 00:50:59	1200	4,75	8,76	11,59
00:51:00 - 00:55:59	675	2,67	8,76	11,59
00:56:00 - 01:00:59	462	1,83	8,76	11,59
Total	25289	100		

7.3 Output t-tests and cross tabulation

T-Test

[DataSet3] \\soliscom.uu.nl\uu\Users\F124497\Scriptie\Number of tracks goed.sav

Group Statistics

	Location	N	Mean	Std. Deviation	Std. Error Mean
Number	Village	376	2,81	1,321	,068
	City	160	2,91	1,386	,110

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Number	Equal variances assumed	,506	,477	-,772	534	,440	-,098	,127	-,346	,151
	Equal variances not assumed			-,757	287,372	,449	-,098	,129	-,352	,156

T-Test

[DataSet2] \\soliscom.uu.nl\uu\Users\F124497\Scriptie\Analyse goed.sav

Group Statistics

	Location	N	Mean	Std. Deviation	Std. Error Mean
Duration	Village	26375	2,2534	1,24728	,00768
	City	11615	2,1215	1,22258	,01134
Dur_hour	Village	25288	5,2295	2,93622	,01846
	City	11270	4,9329	3,05418	,02877
Length	Village	26375	2,6681	,87326	,00538
	City	11615	2,5395	,88126	,00818

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Duration	Equal variances assumed	11,207	,001	9,557	37988	,000	,13194	,01381	,10488	,15900
	Equal variances not assumed			9,631	22608,845	,000	,13194	,01370	,10509	,15879
Dur_hour	Equal variances assumed	19,084	,000	8,807	36556	,000	,29656	,03367	,23056	,36256
	Equal variances not assumed			8,675	20885,358	,000	,29656	,03418	,22955	,36356
Length	Equal variances assumed	9,998	,002	13,189	37988	,000	,12862	,00975	,10951	,14773
	Equal variances not assumed			13,143	22017,084	,000	,12862	,00979	,10944	,14780

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Purpose * Location	37991	100,0%	0	0,0%	37991	100,0%

Purpose * Location Crosstabulation

			Location		
			Villages	Cities	Total
Purpose	Home	Count	9723	4114	13837
		Expected Count	9606,5	4230,5	13837,0
		% within Purpose	70,3%	29,7%	100,0%
		% of Total	25,6%	10,8%	36,4%
	Paid work	Count	9604	4072	13676
		Expected Count	9494,8	4181,2	13676,0
		% within Purpose	70,2%	29,8%	100,0%
		% of Total	25,3%	10,7%	36,0%
	Non-daily groceries	Count	944	492	1436
		Expected Count	997,0	439,0	1436,0
		% within Purpose	65,7%	34,3%	100,0%
		% of Total	2,5%	1,3%	3,8%
	Daily groceries	Count	1940	951	2891
		Expected Count	2007,1	883,9	2891,0
		% within Purpose	67,1%	32,9%	100,0%
		% of Total	5,1%	2,5%	7,6%
	Social	Count	2298	1072	3370
		Expected Count	2339,7	1030,3	3370,0
		% within Purpose	68,2%	31,8%	100,0%
		% of Total	6,0%	2,8%	8,9%

	Recreation	Count	1121	517	1638
		Expected Count	1137,2	500,8	1638,0
		% within Purpose	68,4%	31,6%	100,0%
		% of Total	3,0%	1,4%	4,3%
	Spare time	Count	559	319	878
		Expected Count	609,6	268,4	878,0
		% within Purpose	63,7%	36,3%	100,0%
		% of Total	1,5%	0,8%	2,3%
	Services	Count	117	51	168
		Expected Count	116,6	51,4	168,0
		% within Purpose	69,6%	30,4%	100,0%
		% of Total	0,3%	0,1%	0,4%
	Study	Count	67	26	93
		Expected Count	64,6	28,4	93,0
		% within Purpose	72,0%	28,0%	100,0%
		% of Total	0,2%	0,1%	0,2%
Total		Count	26373	11614	37987
		Expected Count	26373,0	11614,0	37987,0
		% within Purpose	69,4%	30,6%	100,0%
		% of Total	69,4%	30,6%	100,0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	42,479[a]	8	,000	.[b]	
Likelihood Ratio	41,789	8	,000	.[b]	
Fisher's Exact Test	.[b]			.[b]	
Linear-by-Linear Association	22,826	1	,000	.[b]	.[b]
N of Valid Cases	37987				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 28,43.

b. Cannot be computed because there is insufficient memory.

7.4 Output logistic regression analysis

7.4.1 Number of tracks

PLUM - Ordinal Regression

```
[DataSet2] \\soliscom.uu.nl\uu\Users\F124497\Scriptie\Number of tracks goed.sav
```

Case Processing Summary

		N	Marginal Percentage
Number	10-30	84	15,7%
	31-60	158	29,5%
	61-90	139	25,9%
	91-120	97	18,1%
	121-150	41	7,6%
	151-180	10	1,9%
	181-210	5	0,9%
	211->	2	0,4%
Location	Village	376	70,1%
	City	160	29,9%
Valid		536	100,0%
Missing		0	
Total		536	

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	60,213			
Final	59,609	,604	1	,437

Link function: Logit.

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	2,099	6	,910
Deviance	2,102	6	,910

Link function: Logit.

Pseudo R-Square

Cox and Snell	,001
Nagelkerke	,001
McFadden	,000

Link function: Logit.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Number = 1]	-1,777	,169	110,693	1	,000	-2,108	-1,446
	[Number = 2]	-,288	,147	3,870	1	,049	-,576	-,001
	[Number = 3]	,807	,151	28,747	1	,000	,512	1,102
	[Number = 4]	2,018	,181	124,806	1	,000	1,664	2,372
	[Number = 5]	3,328	,272	149,801	1	,000	2,795	3,861
	[Number = 6]	4,235	,397	113,559	1	,000	3,456	5,013
	[Number = 7]	5,497	,718	58,670	1	,000	4,090	6,903
Location	[Location=0]	-,132	,168	,613	1	,434	-,461	,198
	[Location=1]	0[a]	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

7.4.2 Duration

PLUM - Ordinal Regression

Warnings

There are 3 (9,4%) cells (i.e., dependent variable levels by observed combinations of predictor variable values) with zero frequencies.

Case Processing Summary

		N	Marginal Percentage
Duration	00:00:00 - 00:15:59	12850	33,8%
	00:16:00 - 00:30:59	11803	31,1%
	00:31:00 - 00:45:59	8469	22,3%
	00:46:00 - 01:00:59	3436	9,0%
	01:01:00 - 01:15:59	798	2,1%
	01:16:00 - 01:30:59	331	0,9%
	01:31:00 - 01:45:59	149	0,4%
	01:46:00 - 02:00:59	66	0,2%
	02:01:00 - 02:30:59	43	0,1%
	02:31:00 - 03:00:59	12	0,0%
	03:01:00 - 03:30:59	10	0,0%
	03:31:00 - 04:00:59	8	0,0%
	04:01:00 - 04:30:59	7	0,0%
	04:31:00 - 05:00:59	5	0,0%
	05:01:00 - 05:30:59	1	0,0%
	05:31:00 - 06:00:59	2	0,0%
Location	Village	26375	69,4%
	City	11615	30,6%
Valid		37990	100,0%
Missing		0	
Total		37990	

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	399,971			
Final	270,802	129,168	1	,000

Link function: Logit.

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	117,885	14	,000
Deviance	118,315	14	,000

Link function: Logit.

Pseudo R-Square

Cox and Snell	,003
Nagelkerke	,004
McFadden	,001

Link function: Logit.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Threshold	[Duration = 1,00]	-,512	,018	847,808	1	,000	-,547	-,478
	[Duration = 2,00]	,777	,018	1893,695	1	,000	,742	,812
	[Duration = 3,00]	2,082	,021	9693,930	1	,000	2,041	2,124
	[Duration = 4,00]	3,405	,031	12328,602	1	,000	3,345	3,465
	[Duration = 5,00]	4,241	,043	9887,170	1	,000	4,158	4,325
	[Duration = 6,00]	4,989	,060	7024,993	1	,000	4,872	5,105
	[Duration = 7,00]	5,669	,082	4771,818	1	,000	5,508	5,830
	[Duration = 8,00]	6,231	,108	3345,004	1	,000	6,020	6,442
	[Duration = 9,00]	6,903	,150	2120,911	1	,000	6,609	7,196
	[Duration = 10,00]	7,213	,175	1703,308	1	,000	6,870	7,556
	[Duration = 11,00]	7,574	,209	1312,213	1	,000	7,164	7,984
	[Duration = 12,00]	8,002	,259	957,001	1	,000	7,495	8,509
	[Duration = 13,00]	8,631	,354	594,769	1	,000	7,937	9,324
	[Duration = 14,00]	9,612	,578	276,954	1	,000	8,480	10,744
	[Duration = 15,00]	10,017	,707	200,592	1	,000	8,631	11,403
Location	[Location=0]	,230	,020	129,821	1	,000	,191	,270
	[Location=1]	0[a]	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

7.4.3 Duration in the first hour

PLUM - Ordinal Regression

Case Processing Summary

		N	Marginal Percentage
Dur_hour	00:00:00 - 00:05:59	4424	12,1%
	00:06:00 - 00:10:59	4563	12,5%
	00:11:00 - 00:15:59	3863	10,6%
	00:16:00 - 00:20:59	3791	10,4%
	00:21:00 - 00:25:59	3935	10,8%
	00:26:00 - 00:30:59	4077	11,2%
	00:31:00 - 00:35:59	3384	9,3%
	00:36:00 - 00:40:59	2918	8,0%
	00:41:00 - 00:45:59	2167	5,9%
	00:46:00 - 00:50:59	1657	4,5%
	00:51:00 - 00:55:59	1056	2,9%
	00:56:00 - 01:00:59	723	2,0%
Location	Village	25288	69,2%
	City	11270	30,8%
Valid		36558	100,0%
Missing		1432	
Total		37990	

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	475,741			
Final	367,334	108,407	1	,000

Link function: Logit.

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	179,545	10	,000
Deviance	175,678	10	,000

Link function: Logit.

Pseudo R-Square

Cox and Snell	,003
Nagelkerke	,003
McFadden	,001

Link function: Logit.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Dur_hour = 1,00]	-1,844	,021	7868,260	1	,000	-1,884	-1,803
	[Dur_hour = 2,00]	-,980	,018	2942,555	1	,000	-1,015	-,945
	[Dur_hour = 3,00]	-,470	,017	730,445	1	,000	-,504	-,436
	[Dur_hour = 4,00]	-,036	,017	4,317	1	,038	-,069	-,002
	[Dur_hour = 5,00]	,398	,017	528,065	1	,000	,364	,432
	[Dur_hour = 6,00]	,875	,018	2413,876	1	,000	,840	,910
	[Dur_hour = 7,00]	1,339	,019	5142,747	1	,000	1,302	1,375
	[Dur_hour = 8,00]	1,857	,020	8438,265	1	,000	1,818	1,897
	[Dur_hour = 9,00]	2,414	,023	11181,829	1	,000	2,369	2,459
	[Dur_hour = 10,00]	3,121	,028	12294,008	1	,000	3,066	3,176
	[Dur_hour = 11,00]	4,051	,040	10171,722	1	,000	3,972	4,130
Location	[Location=0]	,207	,020	109,684	1	,000	,168	,246
	[Location=1]	0[a]	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

7.4.4 Length

PLUM - Ordinal Regression

Warnings

There are 1 (7,1%) cells (i.e., dependent variable levels by observed combinations of predictor variable values) with zero frequencies.

Case Processing Summary

		N	Marginal Percentage
Length	0-0,99	2992	7,9%
	1-7,49	14845	39,1%
	7,5-14,99	13873	36,5%
	15-29,99	5863	15,4%
	30-59,99	394	1,0%
	60-89,99	17	0,0%
	90 ->	6	0,0%
Location	Village	26375	69,4%
	City	11615	30,6%
Valid		37990	100,0%
Missing		0	
Total		37990	

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	355,308			
Final	145,998	209,309	1	,000

Link function: Logit.

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	62,447	5	,000
Deviance	63,415	5	,000

Link function: Logit.

Pseudo R-Square

Cox and Snell	,005
Nagelkerke	,006
McFadden	,002

Link function: Logit.

Parameter Estimates

							95% Confidence Interval	
		Estimate	Std. Error	Wald	df	Sig.	Lower Bound	Upper Bound
Threshold	[Length = 1,00]	-2,261	,023	9384,294	1	,000	-2,307	-2,215
	[Length = 2,00]	,085	,018	23,483	1	,000	,051	,120
	[Length = 3,00]	1,834	,020	8136,299	1	,000	1,794	1,874
	[Length = 4,00]	4,717	,052	8381,551	1	,000	4,616	4,818
	[Length = 5,00]	7,625	,209	1329,370	1	,000	7,215	8,035
	[Length = 6,00]	8,969	,409	481,929	1	,000	8,168	9,770
Location	[Location=0]	,298	,021	209,684	1	,000	,258	,338
	[Location=1]	0[a]	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

7.4.5 Purpose: Home

```
LOGISTIC REGRESSION VARIABLES Home
  /METHOD=ENTER Location
  /CONTRAST (Location)=Indicator
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=GOODFIT SUMMARY CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

```
[DataSet1] \\soliscom.uu.nl\uu\Users\F124497\Scriptie\Purpose binary logistic.sav
```

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	37991	100,0
	Missing Cases	0	,0
	Total	37991	100,0
Unselected Cases		0	,0
Total		37991	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
Location	Villages	26376	1,000
	Cities	11615	,000

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted Home No	Predicted Home Yes	Percentage Correct
Step 0	Home	No	24154	0	100,0
		Yes	13837	0	,0
	Overall Percentage				63,6

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-,557	,011	2730,377	1	,000	,573

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Location(1)	7,255	1	,007
	Overall Statistics		7,255	1	,007

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	7,271	1	,007
	Block	7,271	1	,007
	Model	7,271	1	,007

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	49822,214[a]	,000	,000

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	,000	0	.

Contingency Table for Hosmer and Lemeshow Test

		Home = No		Home = Yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	7501	7501,000	4114	4114,000	11615
	2	16653	16653,000	9723	9723,000	26376

Classification Table[a]

	Observed		Predicted Home No	Predicted Home Yes	Percentage Correct
Step 1	Home	No	24154	0	100,0
		Yes	13837	0	,0
	Overall Percentage				63,6

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	Location(1)	,063	,023	7,254	1	,007	1,065	1,017	1,114
	Constant	-,601	,019	958,503	1	,000	,548		

a. Variable(s) entered on step 1: Location.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

   32000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                    Y                                                               I
R  24000 +                                    Y                                                               +
E        I                                    Y                                                               I
Q        I                                    Y                                                               I
U        I                                    Y                                                               I
E  16000 +                                    N                                                               +
N        I                                    N                                                               I
C        I                                   YN                                                               I
Y        I                                   YN                                                               I
    8000 +                                   NN                                                               +
         I                                   NN                                                               I
         I                                   NN                                                               I
         I                                   NN                                                               I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        ,1        ,2        ,3        ,4        ,5        ,6        ,7        ,8        ,9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for Yes
          The Cut Value is ,50
          Symbols: N - No
                   Y - Yes
          Each Symbol Represents 2000 Cases.
```

7.4.6 Purpose: Paid work

```
LOGISTIC REGRESSION VARIABLES Paidwork
  /METHOD=ENTER Location
  /CONTRAST (Location)=Indicator
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=GOODFIT SUMMARY CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	37991	100,0
	Missing Cases	0	,0
	Total	37991	100,0
Unselected Cases		0	,0
Total		37991	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
Location	Villages	26376	1,000
	Cities	11615	,000

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted: Paid work No	Predicted: Paid work Yes	Percentage Correct
Step 0	Paid work	No	24315	0	100,0
		Yes	13676	0	,0
	Overall Percentage				64,0

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-,575	,011	2898,473	1	,000	,562

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Location(1)	6,415	1	,011
	Overall Statistics		6,415	1	,011

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	6,429	1	,011
	Block	6,429	1	,011
	Model	6,429	1	,011

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	49640,718[a]	,000	,000

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	,000	0	.

Contingency Table for Hosmer and Lemeshow Test

		Paid work = No		Paid work = Yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	7543	7543,000	4072	4072,000	11615
	2	16772	16772,000	9604	9604,000	26376

Classification Table[a]

	Observed		Predicted		
			Paid work		Percentage Correct
			No	Yes	
Step 1	Paid work	No	24315	0	100,0
		Yes	13676	0	,0
	Overall Percentage				64,0

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Location(1)	,059	,023	6,414	1	,011	1,061	1,013	1,110
	Constant	-,616	,019	1005,029	1	,000	,540		

a. Variable(s) entered on step 1: Location.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

   32000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                    Y                                                               I
R  24000 +                                    Y                                                               +
E        I                                    Y                                                               I
Q        I                                    Y                                                               I
U        I                                    Y                                                               I
E  16000 +                                    N                                                               +
N        I                                    N                                                               I
C        I                                   YN                                                               I
Y        I                                   YN                                                               I
    8000 +                                   NN                                                               +
         I                                   NN                                                               I
         I                                   NN                                                               I
         I                                   NN                                                               I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       ,1        ,2        ,3        ,4        ,5        ,6        ,7        ,8        ,9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for Yes
          The Cut Value is ,50
          Symbols: N - No
                   Y - Yes
          Each Symbol Represents 2000 Cases.
```

7.4.7 Purpose: Non-daily groceries

```
LOGISTIC REGRESSION VARIABLES Nondailygroc
  /METHOD=ENTER Location
  /CONTRAST (Location)=Indicator
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=GOODFIT SUMMARY CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	37991	100,0
	Missing Cases	0	,0
	Total	37991	100,0
Unselected Cases		0	,0
Total		37991	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
Location	Villages	26376	1,000
	Cities	11615	,000

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted: Non-daily groc No	Predicted: Non-daily groc Yes	Percentage Correct
Step 0	Non-daily groc	No	36555	0	100,0
		Yes	1436	0	,0
	Overall Percentage				96,2

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-3,237	,027	14477,522	1	,000	,039

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Location(1)	9,567	1	,002
	Overall Statistics		9,567	1	,002

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	9,376	1	,002
	Block	9,376	1	,002
	Model	9,376	1	,002

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	12214,849[a]	,000	,001

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	,000	0	.

Contingency Table for Hosmer and Lemeshow Test

		Non-daily groc = No		Non-daily groc = Yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	25432	25432,000	944	944,000	26376
	2	11123	11123,000	492	492,000	11615

Classification Table[a]

	Observed		Predicted		
			Non-daily groc		Percentage Correct
			No	Yes	
Step 1	Non-daily groc	No	36555	0	100,0
		Yes	1436	0	,0
	Overall Percentage				96,2

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Location(1)	-,175	,057	9,545	1	,002	,839	,751	,938
	Constant	-3,118	,046	4581,432	1	,000	,044		

a. Variable(s) entered on step 1: Location.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

   32000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I   N                                                                                                I
R  24000 +   N                                                                                                +
E        I   N                                                                                                I
Q        I   N                                                                                                I
U        I   N                                                                                                I
E  16000 +   N                                                                                                +
N        I   N                                                                                                I
C        I   NN                                                                                               I
Y        I   NN                                                                                               I
    8000 +   NN                                                                                               +
         I   NN                                                                                               I
         I   NN                                                                                               I
         I   NN                                                                                               I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       ,1        ,2        ,3        ,4        ,5        ,6        ,7        ,8        ,9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

         Predicted Probability is of Membership for Yes
         The Cut Value is ,50
         Symbols: N - No
                  Y - Yes
         Each Symbol Represents 2000 Cases.
```

7.4.8 Purpose: Daily groceries

```
LOGISTIC REGRESSION VARIABLES Dailygroc
  /METHOD=ENTER Location
  /CONTRAST (Location)=Indicator
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=GOODFIT SUMMARY CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	37991	100,0
	Missing Cases	0	,0
	Total	37991	100,0
Unselected Cases		0	,0
Total		37991	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
Location	Villages	26376	1,000
	Cities	11615	,000

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted: Daily groc: No	Predicted: Daily groc: Yes	Percentage Correct
Step 0	Daily groc	No	35100	0	100,0
		Yes	2891	0	,0
	Overall Percentage				92,4

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-2,497	,019	16648,379	1	,000	,082

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Location(1)	7,950	1	,005
	Overall Statistics		7,950	1	,005

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	7,851	1	,005
	Block	7,851	1	,005
	Model	7,851	1	,005

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	20441,318[a]	,000	,000

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	,000	0	.

Contingency Table for Hosmer and Lemeshow Test

		Daily groc = No		Daily groc = Yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	24436	24436,000	1940	1940,000	26376
	2	10664	10664,000	951	951,000	11615

Classification Table[a]

	Observed		Predicted		
			Daily groc		Percentage Correct
			No	Yes	
Step 1	Daily groc	No	35100	0	100,0
		Yes	2891	0	,0
	Overall Percentage				92,4

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Location(1)	-,116	,041	7,942	1	,005	,890	,821	,965
	Constant	-2,417	,034	5101,243	1	,000	,089		

a. Variable(s) entered on step 1: Location.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

   32000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I       Y                                                                                            I
R  24000 +       N                                                                                            +
E        I       N                                                                                            I
Q        I       N                                                                                            I
U        I       N                                                                                            I
E  16000 +       N                                                                                            +
N        I       N                                                                                            I
C        I       NY                                                                                           I
Y        I       NN                                                                                           I
    8000 +       NN                                                                                           +
         I       NN                                                                                           I
         I       NN                                                                                           I
         I       NN                                                                                           I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       ,1        ,2        ,3        ,4        ,5        ,6        ,7        ,8        ,9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

         Predicted Probability is of Membership for Yes
         The Cut Value is ,50
         Symbols: N - No
                  Y - Yes
         Each Symbol Represents 2000 Cases.
```

7.4.9 Purpose: Social

```
LOGISTIC REGRESSION VARIABLES Social
  /METHOD=ENTER Location
  /CONTRAST (Location)=Indicator
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=GOODFIT SUMMARY CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	37991	100,0
	Missing Cases	0	,0
	Total	37991	100,0
Unselected Cases		0	,0
Total		37991	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
Location	Villages	26376	1,000
	Cities	11615	,000

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted: Social: No	Predicted: Social: Yes	Percentage Correct
Step 0	Social	No	34621	0	100,0
		Yes	3370	0	,0
	Overall Percentage				91,1

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-2,330	,018	16666,024	1	,000	,097

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Location(1)	2,666	1	,103
	Overall Statistics		2,666	1	,103

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	2,648	1	,104
	Block	2,648	1	,104
	Model	2,648	1	,104

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	22756,384[a]	,000	,000

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	,000	0	.

Contingency Table for Hosmer and Lemeshow Test

		Social = No		Social = Yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	24078	24078,000	2298	2298,000	26376
	2	10543	10543,000	1072	1072,000	11615

Classification Table[a]

	Observed		Predicted Social No	Predicted Social Yes	Percentage Correct
Step 1	Social	No	34621	0	100,0
		Yes	3370	0	,0
	Overall Percentage				91,1

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	Location(1)	-,063	,039	2,665	1	,103	,939	,870	1,013
	Constant	-2,286	,032	5084,730	1	,000	,102		

a. Variable(s) entered on step 1: Location.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

   32000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I        Y                                                                                           I
R  24000 +        N                                                                                           +
E        I        N                                                                                           I
Q        I        N                                                                                           I
U        I        N                                                                                           I
E  16000 +        N                                                                                           +
N        I        N                                                                                           I
C        I        NY                                                                                          I
Y        I        NN                                                                                          I
    8000 +        NN                                                                                          +
         I        NN                                                                                          I
         I        NN                                                                                          I
         I        NN                                                                                          I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       ,1        ,2        ,3        ,4        ,5        ,6        ,7        ,8        ,9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

         Predicted Probability is of Membership for Yes
         The Cut Value is ,50
         Symbols: N - No
                  Y - Yes
         Each Symbol Represents 2000 Cases.
```

7.4.10 Purpose: Recreation

```
LOGISTIC REGRESSION VARIABLES Recreation
  /METHOD=ENTER Location
  /CONTRAST (Location)=Indicator
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=GOODFIT SUMMARY CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	37991	100,0
	Missing Cases	0	,0
	Total	37991	100,0
Unselected Cases		0	,0
Total		37991	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
Location	Villages	26376	1,000
	Cities	11615	,000

Block 0: Beginning Block

Classification Table[a,b]

Observed			Predicted Recreation No	Predicted Recreation Yes	Percentage Correct
Step 0	Recreation	No	36353	0	100,0
		Yes	1638	0	,0
	Overall Percentage				95,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-3,100	,025	15060,555	1	,000	,045

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Location(1)	,790	1	,374
	Overall Statistics		,790	1	,374

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	,786	1	,375
	Block	,786	1	,375
	Model	,786	1	,375

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	13502,883[a]	,000	,000

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	,000	0	.

Contingency Table for Hosmer and Lemeshow Test

		Recreation = No		Recreation = Yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	25255	25255,000	1121	1121,000	26376
	2	11098	11098,000	517	517,000	11615

Classification Table[a]

	Observed		Predicted Recreation No	Predicted Recreation Yes	Percentage Correct
Step 1	Recreation	No	36353	0	100,0
		Yes	1638	0	,0
	Overall Percentage				95,7

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	Location(1)	-,048	,054	,790	1	,374	,953	,857	1,060
	Constant	-3,066	,045	4645,105	1	,000	,047		

a. Variable(s) entered on step 1: Location.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

   40000 +                                                                                                    +
         I    N                                                                                               I
         I    N                                                                                               I
F        I    N                                                                                               I
R  30000 +    N                                                                                               +
E        I    N                                                                                               I
Q        I    N                                                                                               I
U        I    N                                                                                               I
E  20000 +    N                                                                                               +
N        I    N                                                                                               I
C        I    N                                                                                               I
Y        I    N                                                                                               I
   10000 +    N                                                                                               +
         I    N                                                                                               I
         I    N                                                                                               I
         I    N                                                                                               I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       ,1        ,2        ,3        ,4        ,5        ,6        ,7        ,8        ,9        1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for Yes
          The Cut Value is ,50
          Symbols: N - No
                   Y - Yes
          Each Symbol Represents 2500 Cases.
```

7.4.11 Purpose: Spare time

```
LOGISTIC REGRESSION VARIABLES Sparetime
  /METHOD=ENTER Location
  /CONTRAST (Location)=Indicator
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=GOODFIT SUMMARY CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	37991	100,0
	Missing Cases	0	,0
	Total	37991	100,0
Unselected Cases		0	,0
Total		37991	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
Location	Villages	26376	1,000
	Cities	11615	,000

Block 0: Beginning Block

Classification Table[a,b]

Observed			Predicted: Spare time No	Predicted: Spare time Yes	Percentage Correct
Step 0	Spare time	No	37113	0	100,0
		Yes	878	0	,0
	Overall Percentage				97,7

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-3,744	,034	12023,452	1	,000	,024

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Location(1)	14,046	1	,000
	Overall Statistics		14,046	1	,000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	13,620	1	,000
	Block	13,620	1	,000
	Model	13,620	1	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	8337,587[a]	,000	,002

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	,000	0	.

Contingency Table for Hosmer and Lemeshow Test

		Spare time = No		Spare time = Yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	25817	25817,000	559	559,000	26376
	2	11296	11296,000	319	319,000	11615

Classification Table[a]

	Observed		Predicted Spare time No	Predicted Spare time Yes	Percentage Correct
Step 1	Spare time	No	37113	0	100,0
		Yes	878	0	,0
	Overall Percentage				97,7

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	Location(1)	-,266	,071	13,969	1	,000	,767	,667	,881
	Constant	-3,567	,057	3947,349	1	,000	,028		

a. Variable(s) entered on step 1: Location.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

   40000 +                                                                                                    +
         I  N                                                                                                 I
         I  N                                                                                                 I
F        I  N                                                                                                 I
R  30000 +  N                                                                                                 +
E        I  N                                                                                                 I
Q        I  N                                                                                                 I
U        I  N                                                                                                 I
E  20000 +  N                                                                                                 +
N        I  N                                                                                                 I
C        I  N                                                                                                 I
Y        I  N                                                                                                 I
   10000 +  N                                                                                                 +
         I  N                                                                                                 I
         I  N                                                                                                 I
         I  N                                                                                                 I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       ,1        ,2        ,3        ,4        ,5        ,6        ,7        ,8        ,9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for Yes
          The Cut Value is ,50
          Symbols: N - No
                   Y - Yes
          Each Symbol Represents 2500 Cases.
```

7.4.12 Purpose: Services

```
LOGISTIC REGRESSION VARIABLES Services
  /METHOD=ENTER Location
  /CONTRAST (Location)=Indicator
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=GOODFIT SUMMARY CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	37991	100,0
	Missing Cases	0	,0
	Total	37991	100,0
Unselected Cases		0	,0
Total		37991	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
Location	Villages	26376	1,000
	Cities	11615	,000

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted: Services: No	Predicted: Services: Yes	Percentage Correct
Step 0	Services	No	37823	0	100,0
		Yes	168	0	,0
	Overall Percentage				99,6

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-5,417	,077	4907,446	1	,000	,004

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Location(1)	,004	1	,951
	Overall Statistics		,004	1	,951

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	,004	1	,951
	Block	,004	1	,951
	Model	,004	1	,951

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	2156,756[a]	,000	,000

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	,000	0	.

Contingency Table for Hosmer and Lemeshow Test

		Services = No		Services = Yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	11564	11564,000	51	51,000	11615
	2	26259	26259,000	117	117,000	26376

Classification Table[a]

	Observed		Predicted Services No	Predicted Services Yes	Percentage Correct
Step 1	Services	No	37823	0	100,0
		Yes	168	0	,0
	Overall Percentage				99,6

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	Location(1)	,010	,168	,004	1	,951	1,010	,727	1,405
	Constant	-5,424	,140	1493,725	1	,000	,004		

a. Variable(s) entered on step 1: Location.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

   40000 +                                                                                                    +
         IN                                                                                                   I
         IN                                                                                                   I
F        IN                                                                                                   I
R  30000 +N                                                                                                   +
E        IN                                                                                                   I
Q        IN                                                                                                   I
U        IN                                                                                                   I
E  20000 +N                                                                                                   +
N        IN                                                                                                   I
C        IN                                                                                                   I
Y        IN                                                                                                   I
   10000 +N                                                                                                   +
         IN                                                                                                   I
         IN                                                                                                   I
         IN                                                                                                   I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        ,1        ,2        ,3        ,4        ,5        ,6        ,7        ,8        ,9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for Yes
          The Cut Value is ,50
          Symbols: N - No
                   Y - Yes
          Each Symbol Represents 2500 Cases.
```

7.4.13 Purpose: Study

```
LOGISTIC REGRESSION VARIABLES Study
  /METHOD=ENTER Location
  /CONTRAST (Location)=Indicator
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=GOODFIT SUMMARY CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	37991	100,0
	Missing Cases	0	,0
	Total	37991	100,0
Unselected Cases		0	,0
Total		37991	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
Location	Villages	26376	1,000
	Cities	11615	,000

Block 0: Beginning Block

Classification Table[a,b]

Observed			Predicted: Study: No	Predicted: Study: Yes	Percentage Correct
Step 0	Study	No	37898	0	100,0
		Yes	93	0	,0
	Overall Percentage				99,8

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-6,010	,104	3351,007	1	,000	,002

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Location(1)	,301	1	,584
	Overall Statistics		,301	1	,584

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	,306	1	,580
	Block	,306	1	,580
	Model	,306	1	,580

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1303,792[a]	,000	,000

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	,000	0	.

Contingency Table for Hosmer and Lemeshow Test

		Study = No		Study = Yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	11589	11589,000	26	26,000	11615
	2	26309	26309,000	67	67,000	26376

Classification Table[a]

	Observed		Predicted Study No	Predicted Study Yes	Percentage Correct
Step 1	Study	No	37898	0	100,0
		Yes	93	0	,0
	Overall Percentage				99,8

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Location(1)	,127	,231	,300	1	,584	1,135	,721	1,786
	Constant	-6,100	,196	965,204	1	,000	,002		

a. Variable(s) entered on step 1: Location.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

   40000 +                                                                                                    +
         IN                                                                                                   I
         IN                                                                                                   I
F        IN                                                                                                   I
R  30000 +N                                                                                                   +
E        IN                                                                                                   I
Q        IN                                                                                                   I
U        IN                                                                                                   I
E  20000 +N                                                                                                   +
N        IN                                                                                                   I
C        IN                                                                                                   I
Y        IN                                                                                                   I
   10000 +N                                                                                                   +
         IN                                                                                                   I
         IN                                                                                                   I
         IN                                                                                                   I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        ,1        ,2        ,3        ,4        ,5        ,6        ,7        ,8        ,9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

         Predicted Probability is of Membership for Yes
         The Cut Value is ,50
         Symbols: N - No
                  Y - Yes
         Each Symbol Represents 2500 Cases.
```

7.4.14 Purpose: Unknown

```
LOGISTIC REGRESSION VARIABLES Unknown
  /METHOD=ENTER Location
  /CONTRAST (Location)=Indicator
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=GOODFIT SUMMARY CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	37991	100,0
	Missing Cases	0	,0
	Total	37991	100,0
Unselected Cases		0	,0
Total		37991	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
Location	Villages	26376	1,000
	Cities	11615	,000

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted: Unknown: No	Predicted: Unknown: Yes	Percentage Correct
Step 0	Unknown	No	37987	0	100,0
		Yes	4	0	,0
	Overall Percentage				100,0

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-9,159	,500	335,492	1	,000	,000

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Location(1)	,059	1	,809
	Overall Statistics		,059	1	,809

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	,061	1	,805
	Block	,061	1	,805
	Model	,061	1	,805

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	81,209[a]	,000	,001

a. Estimation terminated at iteration number 12 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	,000	0	.

Contingency Table for Hosmer and Lemeshow Test

		Unknown = No		Unknown = Yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	11614	11614,000	1	1,000	11615
	2	26373	26373,000	3	3,000	26376

Classification Table[a]

	Observed		Predicted		
			Unknown		Percentage Correct
			No	Yes	
Step 1	Unknown	No	37987	0	100,0
		Yes	4	0	,0
	Overall Percentage				100,0

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Location(1)	,278	1,155	,058	1	,809	1,321	,137	12,702
	Constant	-9,360	1,000	87,601	1	,000	,000		

a. Variable(s) entered on step 1: Location.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

   40000 +                                                                                                    +
         IN                                                                                                   I
         IN                                                                                                   I
F        IN                                                                                                   I
R  30000 +N                                                                                                   +
E        IN                                                                                                   I
Q        IN                                                                                                   I
U        IN                                                                                                   I
E  20000 +N                                                                                                   +
N        IN                                                                                                   I
C        IN                                                                                                   I
Y        IN                                                                                                   I
   10000 +N                                                                                                   +
         IN                                                                                                   I
         IN                                                                                                   I
         IN                                                                                                   I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       ,1        ,2        ,3        ,4        ,5        ,6        ,7        ,8        ,9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for Yes
          The Cut Value is ,50
          Symbols: N - No
                   Y - Yes
          Each Symbol Represents 2500 Cases.
```