# CHANGING PREFERENCES IN MODAL CHOICE

*Are millennials really different?*


ABSTRACT

The presented study supplies evidence for a decrease in car use among the millennial cohort, based on logistic regression models. Controlling for relevant variables in transport literature, a negative effect on the probability of car use was observed. Density, as well as Residential Self Selection, were also found to influence the modal choice of millennials, lowering the probability of car use. Limited evidence was found for differences in the millennial cohort in 2016 and their same-aged counterparts in 2006.



Michael Leonhart

Master thesis Urban, Port and Transport Economics 2018-2019

Thesis supervisor: Dr. G. Mingardo

Second reader: Dr. M. Gerritse

# Contents

# 1. Introduction

In October 1994 the ANWB reported a record amount of 45 traffic jams across the Netherlands, with a total length of 300 kilometers. On the 30th October of 2018, a new record was reported, totaling 1135 kilometer of traffic jam (Verlaan, 2018). The amount of traffic thus has increased drastically over time, and it is expected to increase even further the coming 5 years by an expected 35% (Kennisinstituut Mobiliteit, 2018). In order to decrease congestion, the Dutch government plans to expand the infrastructure. Historical infrastructure increases have led to a 3 to 5 percent increase of cars on the road, implying more infrastructure leads to more traffic (Verlaan, 2018).

Traffic, and in particular congestion, has always been of interest to researchers of transport economics. The social costs related to congestion for the Netherlands were estimated between 2,7 and 3,8 milliard euro per year in 2017 (Versteegh, 2017). With the expected increase in traffic during the coming years, losses to society might increase even further. A field of interest in addressing possible solutions to the congestion problem is travel mode choice. Over the last decades, multiple papers have been written trying to address the modal split of travel options. For example, Heinen (2016) did a cross-sectional study on identity and commute mode choice.

The KiM recently reported that the amount of car use among young adults decreased. In particular among "Generation Y", also called Millennials, car use has declined over the past years (Jorritsma & Berveling, 2014), while the attitude of this group towards car use and ownership has remained constant over the last years (Kennisinstituut Mobiliteit, 2018). Jorritsma and Berveling (2014) argued that the reason for the decline in car use among millennials could be that a larger part of the young adults is a student, instead of already working. Financial motives may thus be of influence on car ownership. A large part of the millennials do state that they seek car ownership in the future (Jorritsma & Berveling, 2014). According to their report, millennials make less use of the car, make less or shorter trips for work, shopping and social contacts, but do undertake more trips for educational purposes compared to the same-aged group in 1995 (Jorritsma & Berveling, 2014). This decline in car use, as mentioned before, is expected to fade out, as millennials still seek car ownership in a later stage of their life.

Research did show that travel mode choice is driven by social identity (Murtagh, Gatersleben, & Uzzell, 2012). They conclude that the social identities of people should be included in addressing travel mode choice, and that social identities can be used to change the behavior of travelers (Murtagh, Gatersleben, & Uzzell, 2012). As described by Karakas, Manisaligil &Sarigollu (2015), millennials show some major distinctions from earlier generations, of which not all are positive. For example, millennials are characterized as impatient as they seek instant gratification, and have less tolerance for delays. Positively, millennials operate in a world that is never offline, boosting for example their multi-tasking and digital processing skills. Also, millennials have a "The sky is the limit" attitude. This, in effect, explains their constant demand for feedback, directly after delivering their product, and gaining recognition from others for their work (Karakas, Manisaligil, & Sarigollu, 2015). All these characteristics help build the "Millennial Identity".

As millennials and their follow-up generation are the generations that will most likely have to deal with the predicted increase in congestion, they are an interesting group for research. Understanding their preferences and identity, but also the changes in the economic and geographical environment millennials are dealing with, might provide governments with adequate policy options to motivate millennials to forego car use and ownership. Especially Millennials, with their constant demand for feedback (Karakas, Manisaligil, & Sarigollu, 2015), might be vulnerable for government policy.

There is no former research available that applies both the characteristics of millennials and the economic and infrastructural environment they live in addressing the modal choice of transportation. The presented thesis will explain the difference between millennials and former generations, the effect of the millennial generation on the modal choice of transportation.

Put formally, the following hypotheses are formulated.

*H1.* *There is a significant difference between millennials and former generations concerning the modal choice of transportation, controlling for other variables.*

*H2.* *There is a significant difference between same-aged individuals in 2006 and millennials in 2016 concerning the modal choice of transportation, controlling for other variables.*

*H3.* *There is a significant difference between millennials that live in high-density areas and millennials that do not live in high-density areas, controlling for other variables*

The characteristics of millennials and why they differ from earlier generations will be based on a literature study. The findings of this literature study will determine the factors that determine the modal choice of transportation. The effect of the millennial generation on modal choice of transportation will be explained using a logistic regression model, drawing on data in the OViN- and MON-databases that contain trip data from 2005-2016 for 30,000 to 50,000 randomly selected respondents per year. The hypotheses will be tested and used to formulate an answer to the research question of this research, which is formulated as

*RQ.* *Do millennials differ from former generations in terms of modal choice of transportation?*

# 2. Theoretical Framework

As the modal choice of transport, focused on millennials, is a relatively undiscovered topic, this section will discuss relevant literature on millennials apart from relevant literature on the modal choice of transportation. Heinen (2016) already proved that identity has an effect on travel mode choice, while Karakas, Manisaligil & Sarigollu (2015) provided literature to build the “Millennial identity”. These studies imply that combining literature on millennials with literature on transportation mode choice might help to understand which factors are determining transport mode choice and should thus be included in the empirical model as provided in chapter 4.

## 2.1 Millennials

The first part of this literature review will focus on the preferences and characteristics of millennials. Multiple studies will be reviewed and discussed, and where necessary the connection to transport economics will be explained. Section 2.2 will discuss relevant literature on the modal choice of transportation.

### 2.1.1 Identity

Both the studies by Murtagh, Gatersleben & Uzzell (2012) and Heinen (2016) discussed identity as a possible determinant of transportation mode choice. As mentioned in both studies, multiple definitions of identity are identified. For example, as mentioned by Murtagh, Gatersleben & Uzzell (2012), a distinction can be made between personal identity and social identity. Personal identity can for example relate to autonomy, while social identity relates to self-presentation. Social role theory provides the theoretical framework that is needed to define the concept of identity. Social role theory argues that identity is the internalisation of a socal role, also containing all inherent expactations and norms. Therefore, social role theory argues that individuals can have different and conflicting identities, based on the social environments they act in (Murtagh, Gatersleben, & Uzzell, 2012). Identity, and in particular identity theory, is used to examine and explain role choice behavior (Stryker, 2007). In addition, identity theory is better suited for situations in which actors have multiple viable alternatives  compared to situations where few or no options are available (Stryker, 2007). This means that identity, based on identity theory, is well suited in explaining and predicting travel mode choice, as in most cases multiple alternatives are available for individuals.

As mentioned, individuals might have multiple conflicting identities based on their social networks. Therefore, it might occur that the modal choice of transportation at a certain time is based on the environment or social network the individual is surrounded by at that moment. As mentioned by Jorritsma & Berveling (2014), the amount of millennials that own a car has declined over the past years, implying that an individual that identifies itself with the “millennial identitiy”, that is, an individual in a social environment of milennials, is more likely to refrain from car use, based on social role and identity theory. This same individual, however, might have a preference for car use in an environment where it is surrounded by for example older individuals. Travel behavior may be influenced by interplay of infrastructure, neighbourhood and social circumstances (Guell, Panter, Jones, & Ogilvie, 2012). Identity helps to form the social circumstances of an individual, and therefore it should be considered in research concerning the modal choice of transportation.

As it is now clear that identity can have a significant influence on decision making, in this case the modal choice of transportation, the “millennial identity” should be defined. It is noted that individuals are not homogenous, that is, differences between preferences, traits and personalities can be observed among groups. Therefore, the “Milennial Identity” will be a generalized concept based on former literature, in order to be able to include it in this research.

### 2.1.2 The "Millennial Identity"

Millennials are impatient, constantly looking for feedback and intolerant for delays (Karakas, Manisaligil, & Sarigollu, 2015). As mentioned before, these characteristics will not apply to all individuals that are seen as millennials. However, literature shows evidence that a large part of the millennial population does show these characteristics. This section will discuss multiple articles that define the characteristics of millennials and combine them in order to define the general "Millennial Identity".

Millennials, and more important, their characteristics and preferences that also apply to modal choice of transportation, are not dominantly present in available transport literature. Therefore, relevant insights in literature in different fields of interest are discussed and applied to modal choice of transportation. Most of these studies are conducted in the field of marketing and Human Resources (HR). An ambiguity that is easily found is the differences in definition of the millennials, as different studies define millennials to be born between different years. For this study, millennials are defined as the group born between 1982 and 2009, as proposed by Alexander & Sysko (2012). Evidence, albeit not completely unanimous, has been found that Millennials differ from the same age group in previous decades (Kultalahti & Viitala, 2015). Therefore, studying millennials as a group of interest might provide new insights in transport literature.

One of the aspects that formed the millennials, at least in their way of thinking, is the so-called phenomena of helicopter parents. As millennials grew up with their parents being child-centered and protective of their children, raising them in general with a "trophy for all" attitude, unrealistic expectations for the real world were given, as children were not raised to understand the concept of failing. This leads to the finding that millennials were impeded in their development of a sense of independence and responsibility (Alexander & Sysko, 2012). Relating this to the modal choice of transportation, as parents of millennials were often over-protective of their children, it can be argued that these children were often driven by their parents to school, social or sports activities and family. This, in turn, might influence their modal choice of transportation today. Past commuting behavior has a significant impact on future commuting behavior (Bamberg, Ajzen, & Schmidt, 2003).

In the theory of planned behavior, it is stated that individuals hold certain attitudes, subjective norms and intentions towards certain behavior. These aspects have been used to explain why individuals do choose for a healthy transport alternative, or what drives them to choose an unhealthy alternative. In addition to the theory of planned behavior, habit is a complementary psychological concept that assumes that individuals do not consciously make a decision every time they commute, but rather fall back on a routine (Guell, Panter, Jones, & Ogilvie, 2012). Following this reasoning, it can be argued that millennials tend to have a preference for car use, as the habit of using a car to commute was already promoted by their parents in early life. However, interventions make past behavior insignificant in predicting future behavior (Bamberg, Ajzen, & Schmidt, 2003). This means that it is possible to change the behavior regarding the modal choice of transportation of millennials.

Millennials are attentive and respectful, and have a desire to make the world a better place (Paulin, Ferguson, Jost, & Fallu, 2014). However, they are not loyal to a particular cause or organization (Paulin, Ferguson, Jost, & Fallu, 2014), as they are not loyal to any company they work for (Hubbard, 2013). This disloyalty can be derived from the fact that millennials are placing more emphasis on money and image, expressed in increased extrinsic behavior and materialism (Paulin, Ferguson, Jost, & Fallu, 2014), therefore first choosing what is best for themselves. Relating to the modal choice of transportation, this disloyalty might cause impediments in their desire to make the world a better place, as more sustainable alternatives to commute might not be the best option for the millennial itself, therefore leading to socially unfavorable decision-making. A clear example can be found in the impatience of millennials, as their tolerance of delays is low (Karakas, Manisaligil, & Sarigollu, 2015). Car use, which is the most environmentally unfriendly, is most often the fastest way of commuting and therefore leads to unfavorable transport behavior.

The "Millennial Identity" which applies to transport literature can thus be defined and summarized in three major traits. The first trait that is identified is impatience. Millennials are intolerant to delays and will therefore prefer the transport modality that has the lowest travel time combined with the lowest risk of delay. However, this tolerance might be relaxed when rewards or gratification is imposed. The second trait that can be identified is based on the desire to make the world a better place. The so-called "Sustainability" trait of millennials shows that they make decisions that are not only good for themselves, but also for their environment. It is noted though that millennials first make the decision that is best for themselves. The third trait can be defined as latent laziness, based on the habits and attitudes that were transferred from the parents to the millennials. As noted before, these traits are a generalization of the millennial cohort and will not apply to all millennials, but will be assumed to be inherent in the millennial cohort for the purpose of this research.

### 2.1.3 Changing Millennials' behavior

As mentioned by Stradling (2004), who conducted research about why the Scottish population did not make use of a car for their trips, the age group 25-39 was dominantly stating that environmental or health motives refrained them from car use. In addition, the population with a higher income and/or a degree level qualification also stated that environmental and health motives refrained them from car use (Stradling, 2004). This finding, already found in 2004, implies that the higher educated population is more likely to value sustainable and healthier alternatives. Jorritsma & Berveling (2014) stated that car use among millennials has declined over the past year, while millennials do state that they seek car ownership, and therefore use, in the future (Jorritsma & Berveling, 2014). At this point, government policy and marketing can make a big impact on the degree of car use and ownership the millennial population seeks in the future. As can be reasoned, based on the fact that car use has declined among millennials, it should be possible to change the attitude towards car use for them in the future as well. This statement is based on the fact that millennials manage to do without car use at this moment, so in theory it shoud be possible to do without car use in the future as well. It should be noted though, based on social role theory, that millennials also change identity, that is, they identify themselves with a different social group, and therefore their beliefs and preferences change as well. This means it is possible that car use and ownership is prefered by an individual that does not identify itself as a millennial anymore. Of course, this can also happen because the millennial reaches the next phase of his or her life.

Also important to note when trying to change the behavior of millennials is their quest for instant gratification for their actions (Karakas, Manisaligil, & Sarigollu, 2015), which allows government parties to effectively promote desired transport decisions by rewarding individuals when they choose these prefered modes of transport. An example can be found in OVmiles, a rewrad program fro the Rotterdam public transport company (RET), which reward a traveler with points for every trip by public transport in Rotterdam (RET, sd). Similar and more general programs are present as well, varying in scope and magnitude. These transportation demand management (TDM) programs do have disappointing results in terms of percentage of modal shift, emphasizing the need for futher understanding of the aspects that influence the modal choice of transportation (Ramezani, Pizzo, & Deakin, 2018)

## 2.2 Modal choice of transportation

The modal choice of transportation is a well-documented concept in transport literature. It should be noted that many of the studies used numerous aspects that influence modal choice of transportation, however, many of this studies do not incorporate all the aspects at once, leading to different and sometimes contrasting outcomes (Ramezani, Pizzo, & Deakin, 2018). This section will give an overview of the available literature and will use its outcomes to define relevant aspects that influence the modal choice of transportation.

### 2.2.1 Land use

The first variable that will be discussed is that of land use. Land use interacts strongly with the modal choice of transportation. This finding is based on extensive literature that has been formed over the last quarter century, which explain these interactions (Ramezani, Pizzo, & Deakin, 2018). Development that is oriented on car use will induce more demand for roads and parking, thus demanding further car-oriented development in the future. Compact urban environments, on the other hand, can induce more active travel, as well as more demand for public transport. Both transportation mode choice and land use do have strong environmental effects (Noth, Borning, & Waddel, 2003). Therefore, policymakers at multiple governmental levels in the USA used land use policies to reduce vehicle-miles travelled (VMT), and in effect greenhouse gas emissions (Cao, Mokhtarian, & Handy, 2009).

Individuals that prefer to walk and bike instead of using a car self-consciously choose to live in neighborhoods that are conducive to walking and cycling. This finding of self-selection does show the association between the built environment and travel behavior, although it does not directly imply causality (Cao, Mokhtarian, & Handy, 2009). The built environment contains factors like density, diversity and design (Ramezani, Pizzo, & Deakin, 2018), as well as distances to public transport connections (Van Wee, 2002). Relating to the modal choice of transportation, higher-density and mixed-use neighborhoods walk more and retain from car use, when compared to lower-density, single-use neighborhoods (Cao, Mokhtarian, & Handy, 2009). Keeping other factors constant, higher density implies that less distance has to be covered to reach facilities, thus diminishing VMT. This, in turn, will influence the modal choice of transportation, as individuals do value travel time negative (Van Wee, 2002). Mixed land use follows the same reasoning: if facilities are spread over town, on average facilities will be closer to dwellings compared to a town where all facilities are located in the center (Van Wee, 2002).

The location of the destination, and more important, its accessibility, are also clear determinants of transport mode choice. As firms decentralized into suburban areas, it was found that car use increased, although travel time declined (Meurs & Haaijer, 2001). This will most likely be the result of lower accessibility for trips undertaken by public transport, as it is likely that travel times are higher for trips by public transport in suburban areas. This is based on the reasoning that the public transport network will likely be less frequent or less convenient in lower density, suburban areas than in high density metropolitan areas. The relative attractiveness of public transport versus the car depends on both the spatial characteristics of the area, as well as on the demand and whether the supplier can provide the demanded need for public transport (Meurs & Haaijer, 2001). This assumption is further inclined by van Wee (2002), as this paper argues that building in higher density where also public transport (and in particular train station) is available, there will be an additional effect of replacing car use by public transport use (Van Wee, 2002).

As literature has shown, density is the main factor that shows the relation between land-use and transport mode choice. As addressed above, more urban form factors have been found over the year, of which diversity, design, mixed land use and access to public transport were already mentioned. The study by Aditjandra (2013) also identified settlement size, strategic development location, accessibility of key facilities, development site location, strategic transport network, job-housing balance, traffic demand management and parking (Aditjandra, 2013). These urban form factors, as well as socio-demographic factors were both included in the macro- and micromodel of Aditjandra (2013), while the micromodel also contained data on preferences and attitudes of individuals. The macromodel is based on conventional utility theory and therefore does not include attitudinal characteristics (Aditjandra, 2013). This means that attitudes and characteristics will be incorporated in the residue of a regression analysis.

It should be noted that the relation between urban forms and transportation mode choice is not undisputed for all urban forms. Different methodologies can yield different results in assessing the effect of built environment on travel behavior, predominantly because the measurement methods cannot measure the separable effects of the built environment on travel behavior. Therefore experimental evidence for the effects of urban form cannot be provided (Ramezani, Pizzo, & Deakin, 2018). As urban form is not the scope of this research, land-use will be represented by density, the amount of inhabitants and the location in the Netherlands. The location will be included on the national level (province).

Land use was found to be of significance when assessing travel behavior. One of the key concepts that is distinguished is the Residential Self Selection (RSS). The choice of residential location was proven to be determined by intended travel behavior. One of the most commonly named reasons to choose for a certain residential location is travel access (Guan & Wang, 2019). Although the magnitude and sign of the effect of RSS were inconclusive, evidence of over- or underestimation of land-use was found when a model did not control for RSS (Guan & Wang, 2019). This implies that individuals or households that share the same preferences for travel access or the same attitudes towards intended travel mode, might be clustered in the same neighborhoods. As this self-selection will occur among same-minded people, millennials will also tend to prefer living in certain neighborhoods. To control for this effect, an interaction term between millennials and density can be used to control for the effect of RSS.

### 2.2.2 Sociodemographic variables

Socio-demographic variables have always been present in research on transport mode choice. An extensive literature, for example, has been formed about the elderly cohort. Newbold et al. (2005) argue that the elderly cohort conducted roughly the same percentual amount of trips by car as younger cohorts in the United States. It was found as well that as population ages, the dependency on car use increased. Also, the proportion of elderly drivers has risen, and then use of public transport declined (Newbold, Scott, Spinney, Kanaroglou, & Páez, 2005). Böcker, van Amen & Helbich (2017) emphasize that the elderly population has the same dependence on car use as younger generations, but take less and sorter trips. Also, they argue that the elderly cohort might be too broad and that it is useful to distinguish between groups. In this case, they argue that the elderly cohort can be divided in a group between ages 65 and 75, and a group with ages 75 and above (Böcker, van Amen, & Helbich, 2017). The elderly cohort does differ from the millennial cohort in multiple ways. First of all, it can be reasoned that the elderly population is less suited for active travel modes than the millennials, as physical health is in general declining as population ages. The second difference can be attributed to the available level of public transport during the lives of  the now elderly cohort. The availability and convenience of public transport alternatives improved over the last decade, allowing younger cohorts to choose between a better alternative than the elderly cohort could in their younger years. This leads to certain attitudes and presumably the habit of using a car.

Despite the differences between the cohorts, some of the proposed socio-demographic variables are also relevant for this study. First of all, it was found that elderly females make fewer and shorter trips than elderly males, and that unemployed elderly make fewer and shorter trips than employed elderly (Böcker, van Amen, & Helbich, 2017). Limited evidence does find that the difference between males and females seems to converge, as the difference between total trip length in the USA declined from 11,6 minutes to 4,4 minutes between 1986 and 1998 (Newbold, Scott, Spinney, Kanaroglou, & Páez, 2005). More recent evidence found that the gender effect was only significant for bicycle use in the younger generations, while for the elderly the effect was also significant for walking and public transport use. This difference was attributed to an earlier finding which stated that elderly men were more likely to make use of a car (Böcker, van Amen, & Helbich, 2017). This finding is supported by a study of the Baby Boomer generation back in 2013. Also, it was stated that the baby boomer generation traveled more and farther than their same-age equivalents during their parents' elderly days (Siren & Haustein, 2013). The baby Boomers mentioned in the Danish study have now become the elderly cohort as described by Böcker, van Amen & Heiblich (2017). Siren & Haustein (2013) also argued that car driving still bears different emotional and cultural meanings for men and women, causing different qualitative and quantitative driving experiences for both genders. Most notable is the finding that woman tend to justify their use of the car with practical reasons, whereas men also give the pleasure of driving as a main factor of car use (Siren & Haustein, 2013). Based on these findings, gender is also expected to be of significance for the millennial cohort and will therefore be included as a controlling variable.

Unemployment was found to be of influence on the amount of trips undertaken by the elderly cohort (Böcker, van Amen, & Helbich, 2017). It can be reasoned that an individual that cannot afford a private car and will therefore make more use of other transport modes. Relating to unemployment, the effect of employment was further examined by Watkins (2016), who argues that income has an impact on the modal choice of transportation based on the level of employment. Individuals with a higher income can afford to live in more expensive neighborhoods, thus increasing their possibilities to relocate. This, in turn gives them more freedom in their mobility flexibility (Watkins, 2016). As commuting behavior is based on the residential location, which on its own is based on trade-offs between housing costs and commuting costs (Watkins, 2016), high-wage individuals have more possible residential locations they can afford, and therefore minimize their commuting costs. Preferences do influence the assessment of commuting costs. Commuting costs, in general, are seen as disutility. Individuals tend to minimize this disutility by either relocating or changing jobs (Kronenberg & Carree, 2012). As utility functions are unique for every individual, the magnitude of the relocation will differ between persons, based on their attitudes and preferences regarding commuting.

Based on this findings, income might influence both the length of trips as well as the modal choice of transportation. Also, as income was found to affect relocation and job-switching decisions, the distance from the residential location to the workplace, or in general, the location of the activity, might also be of interest.

Recent study on the effect of ethnicity on the modal choice of transportation in Australia found that many foreigners, consisting of overseas-born, recent migrants, and Australians with an Asian background, have more environmental sustainable preferences. They own less cars and make less use of cars than native Australians. This difference cannot be attributed to socio-economic disadvantages (Klocker, Toole, Tindale, & Kerr, 2015). These differences between natives and migrants was also found in the Netherlands. A study among different ethnicities in the Netherlands yielded different travel mode choices when compared to the native Dutch inhabitants. Second generation migrants were also found to make less use of bikes, which was attributed to cultural norms and perceived safety (Harms, sd). Both studies imply that ethnicity can have some effect on the modal choice of transportation. Also, based on Harms (sd), an interaction between ethnicity and gender might be a variable of interest, as women of Muslim migrants might not travel at all for religious reasons.

The last factor that influences the modal choice of transportation that will be discussed here is car ownership. Car ownership, that is, access to a private vehicle owned by the household, was argued to influence the habitual decision-making of individuals (Garcia-Sierra, Miralles-Guasch, Martínez-Melo, & Marquet, 2018). Car ownership is a cognitive choice made by an individual at a certain period in time, influencing future transportation mode choices by making a car accessible at all time. In addition to car ownership, possession of a drivers' license was also included in the model. Owning a drivers' license allows an individual to make use of a private car (if owned), borrow a car from friends or relatives or make use of a shared car (Garcia-Sierra, Miralles-Guasch, Martínez-Melo, & Marquet, 2018). These factors might very well influence the modal choice of transportation. Car ownership is often seen as a sunk cost. If the car is already there, it is perceived to be free to use. This does not hold, as cars need fuel to operate. As fuel costs are highly fluctuating, also these prices might influence the decision an individual takes, as the perceived benefits of car use should outweigh the costs. Therefore, fuel prices per year are considered as a relevant variable in predicting travel mode choice.

# 3. Data and Methodology

Now that relevant literature on modal choice has been reviewed, the next section will describe the dataset and methodology of this research. First, the history of the database will be discussed, followed by the adjustments and additions made in order to define relevant variables for this papers' research purpose. The last part of this section will specify the methodology used to test the hypotheses.

## 3.1 Dataset description

This section will discuss relevant information about the dataset and how the final dataset was constructed.

### 3.1.1 OViN and MON

In order to test the hypotheses stated in the introduction of this paper, data concerning the traveling patterns of individuals was needed. This data is supplied by the Dutch Central bureau of Statistics (CBS), using the outcomes of the "Onderzoek Verplaatsingen in Nederland" (OViN) survey (Centraal Bureau voor de Statistiek (CBS); Rijkswaterstaat (RWS), 2017). Each year, between 30000 and 50000 individuals are asked to record their travelling behavior of a selected day in the year. Personal information, for example gender, age, and household location, is included in the dataset, as well as information about income, education, the possession of a private car and drivers' license of individuals. Some land-use variables were also included, which were based on the household location of the respondent. Detailed information about the movements and trips of respondents was included, with the details going as far as describing every part of a movement ("verplaatsing"), and using this trip data to specify all modes of transport used to fulfill the movement (Sociaal Cultureel Planbureau, 2019). The datasets are available on DANS, a governmental institute that supplies permanent access to research (DANS, sd). From here, all data was collected. The OViN database is available from 2010 onwards, as it replaced the "Mobiliteitsonderzoek Nederland" (MON), which was conducted from 2004 until 2009. (Sociaal Cultureel Planbureau, 2019) Some variables were changed in terms of label and/or possible outcomes over the years, especially between 2009 and 2010. Adjustments in terms of recoding were made in order to obtain an uniform dataset. More details about the recoding efforts made, as well as the addition made to the basic database will be further discussed in section 3.1.2.

As mentioned, the database used in this research thus contains two sources of data. For years 2010 up to and including 2017, the database was drawn from OViN. This database is made up by respondents of a survey, which is rolled out among randomly selected Dutch individuals that are registered in the GBA (Basic administration). Respondents are selected in a two-way random sample. First, a stratified sample of parishes is drawn. These parishes are drawn with probabilities based on population size. Next, a single random sample is drawn from this sample of parishes. This implies that all individuals in a certain province have the same probability to be selected in the sample, and the amount of individuals selected per province is based on the population size of that province (Centraal Bureau voor de Statistiek; Rijkswaterstaat, 2017). Per year, the aim was to get 35000 respondents. Almost 55000 individuals were invited to participate each year, of which on average 52% did respond (Sociaal Cultureel Planbureau, 2019).

For years 2005-2009, MON was used as the source for the database. This gave rise to three distinguishable differences that should be noted. First, variables changed in terms of syntax or label. This will be more thoroughly discussed in section 3.1.2. Second, the sampling method for MON differs from OViN. As described, OViN uses a two way sampling method based on parishes, where in the end individuals from the GBA are selected. MON uses a similar source as the GBA, the DMdata consumerbase, to obtain a relevant target audience for the survey. From this audience, random samples were drawn each month to obtain the total dataset. (Sociaal Cultureel Planbureau, 2019) The third difference comes forth out of the second. The DMdata Consumerbase is a database at household level, while GBA is a database at individual level. This also requires some adjustments to make the OViN and MON comparable to each other.

### 3.1.2 Adjustments and modifications

As mentioned, the rough dataset was constructed by combining the MON datasets from years 2005 and 2009 and the OViN datasets from 2010 until 2017 to 1 large database of roughly 1.9 million observations. The distribution of the amount of observations per year is shown in Table 1. As the dataset contained sensitive information regarding household locations of respondents, the dataset was first modified to exclude these variables. These variables for example gave information on postal codes of both residential, business and school locations of respondents, which are irrelevant for this research. The land-use variables urbanity and size of municipality are still included in the data, and are based on the residential location of the respondent. Also, the residential province has been preserved, which can be used to distinguish respondents on their residential location at a high level. Provinces in the Netherlands are expected to differ from each other in terms of land-use, as for example “Zuid-Holland” is expected to be on average more densely populated if compared to for example “Limburg” or “Groningen”. Density was not included as a variable in the dataset. Urbanity is directly related to density and thus serves as the relevant variable of land-use, as well as municipality size.

| Year | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
|---|---|---|---|---|---|---|---|
| Observations | 221.986 | 187.589 | 180.656 | 162.933 | 116.267 | 145.499 | 134.200 |
| Year | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | |
| Observations | 138.491 | 135.762 | 136.062 | 115.987 | 114.348 | 115.161 | |

*Table 1. Number of observations per year*

As mentioned in the description of the dataset, recoding was needed to uniformize variables from MON and OViN. Most of these recoding efforts fixed unambiguity in the surveys, for example “Unknown” was coded as “0” in MON, while it was coded as “3” in OViN. These errors were fixed so there were no missing values left based on the difference between MON and OViN. Observations where unknown and missing variables existed for for example education, income and trip mode of transport were dropped from the dataset. Recoding for multiple variables was also necessary based on the age of respondents, as MON rendered respondents under 12 as irrelevant. In MON, respondents under 15 were excluded from some questions, as the answer “Respondent is younger than 15” was automatically supplied. In OViN, this age group was lowered to only include respondents under the age of 12. Therefore, for observations in OViN, the age group was also updated to ages below 15.

One of the most challenging differences between MON and OViN lies in their recording of income. The MON supplied only information about household income, while OViN only supplied information on personal income. Two solutions for this problem were explored, both with their own pros and cons. The first solution explored was to adjust the personal income based on the position in the household and the amount of individuals in the household. This asks for a well-educated weighing of household positions. As both personal income and household income were included in terms of ranges(e.g between 7500 and 15000), multiplication of income levels would be tedious and not accurate. The second solution was to include an income-proxy, based on the relative level of personal income or household income towards all respondents. Based on percentages, a new variable was generated, which took similar percentages of observations from both personal and household income. This way, no estimation of weights is needed.

The second solution was implemented to generate an unambiguous variable to measure income. The variable contains 5 categories, where similar percentages of the old variables were included by the number of observations, where the lowest incomes were put under “1” and the highest under “5”. As income is only a control variable, this way of measuring income was assumed as sufficient for its purpose. It is noted though, that the variable is not entirely reliable and should not be used to draw strong conclusions.

Fuel price was also included as a variable, drawing on publicly available history on fuel prices supplied by the Centraal Bureau voor de Statistiek (CBS, 2019). In order to correct for spatial differences, the fuel prices were corrected per province, based on fuel prices of the cheapest gas stations per province in June, 2018 (Omroep Flevoland, 2018). These prices represent the operating costs of a car and are expected to influence modal choice, as higher costs might exceed the marginal benefit of using a car, therefore possibly influencing modal choice

As mentioned before, the dataset contains in-depth information about multiple trips a person makes during the day. This causes a problem, as multiple observations per respondent are not independent, therefore not eligible for regression. A solution for this is formulated as follows. First, a random, normally distributed value was generated for all observations. The dataset was then sorted based on these values instead of respondent. This is necessary to tag one observation per respondent, as otherwise all first trips are selected, instead of a random trip. All observations that are not tagged, are dropped from the dataset, so now only one observation per respondent is preserved, and this observation is randomly chosen from all observations of each respondent. This way, independence of observations is achieved. It is noted that most respondents only documented one trip. This leads to the first, and thus sometimes the only trip of a respondent, being the majority in the dataset.

In the end, the final dataset consists of 317.879 observations, distributed over the period 2005-2017 as stated in Table 2. This means 16,68% of all observations were preserved after the adjustments made. Per year, there are some differences in the observations preserved. For example, the lowest percentage (12,2%) of observations was preserved in 2009, while the highest percentage (20,36%) was preserved of the observations of 2017. The final dataset contains the least observations from 2009, as 5,29% of the total dataset contain observations from 2009. The largest amount of observations in the final dataset is from 2005, as 9.48% of all observations are from 2005.

| Year | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
|---|---|---|---|---|---|---|---|
| Observations | 30.115 | 25.165 | 23.832 | 19.873 | 16.820 | 27.104 | 25.945 |
| Year | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | |
| Observations | 26.822 | 26.095 | 26.411 | 23.009 | 22.805 | 23.389 | |

*Table 2. Number of observations per year in the final dataset*

All variables of interest were checked on missing values and outliers. If such an observation was detected, it was dropped from the dataset. The final dataset was saved and no alterations were made beyond this moment.

## 3.2 Variable description

This section will discuss the dependent variable and the independent variables used in the model, as will be described in section 3.3. First, the dependent variable will be explored. Then, some variables of interest will be extensively discussed, and the section will end with a brief overview of the control variables added to the model. The descriptive statistics for all variables are given in Appendix A Table 1, which can be found in Appendix A. Most of the variables used in this study are categorical of nature, but are transformed into dummy-variables when used in the models. Therefore, the descriptive statistics of the dummy-variables are given.

### 3.2.1 Car use

The dependent variable in the model will be car use. Jorritsma & Berveling (2014) found that the level of car use has declined among millennials over the last years. However, the present dataset shows an opposite trend. In general, there is no clear trend visible, but among millennials the amount of car use is slightly increasing. See Appendix A Table 2 in Appendix A. One of the main factors that is also mentioned when addressing millennials in terms of modal choice is their choice of location. As an increasing number of millennials is studying instead of already working, more millennials choose to live close to their source of education, therefore coping better with alternative modes of travel instead of car use, as they have to travel less far in general (Jorritsma & Berveling, 2014). Therefore, car use is expected to decrease among millennials. When the drivers behind car use are well-understood, policy can be formulated to decrease car use, which will in turn have an effect on the problem of increasing congestion.

*Car use* was included as a binary variable. The value it takes was derived directly from the question of in both surveys that inquired about the mode of transport of each trip. This implies that only two possibilities are possible, as a respondent did or did not use a car for its trip. Among car users, passengers were also included, as they made use of a car to get to their destination, albeit indirectly. The distribution of this variable among its possible outcomes is almost equal. In total, slightly more respondents used a car. Among millennials however, there were more observations recorded where no car was used. This is summarized in Table 3.

| Caruse\Millennial | No | Yes | Total |
|---|---|---|---|
| No | 120.008 | 31.045 | 151.053 |
| Yes | 143.676 | 22.656 | 166.332 |
| Total | 263.684 | 53.701 | 317.385 |

*Table 3. Car use among millennials and non-millennials*

Choosing to commute by car but not driving might be a fundamentally different choice than choosing to use a car and drive it yourself. Therefore, a second car use variable was constructed, that excludes passengers from the car use variable. The “car driver” variable thus only consists of respondents who used a car to travel and drove it themselves. The possible difference between the effect of *millennial* on car use and car driver will be addressed to differentiate between direct and indirect car use.

### 3.2.2 Variables of interest

*Car use* was thus chosen as the dependent variable in this study. In order to formulate the regression equation used to test the hypotheses stated earlier in this paper, the variables of interest should be presented and discussed, as well as relevant control variables. These variables combined can then be used to formulate the models that are used to test the relationship between car use and the variables of interest.

#### *3.2.2.1 Millennial variables*

The first variable of interest that will be discussed is the variable that stated if the respondent classifies as a millennial. As described in the theoretical framework, the definition of Alexander & Sysko (2012) was used, stating someone classifies as being a millennial when its born between 1982 and 2009. As expected, in the earlier years of the dataset, a relatively small number of millennials is observed. The main reason for this is that in for example 2007, only respondents that were born between 1982 and 1992 classify as millennials. This holds because respondents under the age of 15 were dropped from the dataset, as they might be dependent on the modal choice of their parents or carekeeper. On the age of 16, Dutch inhabitants are eligible to drive a scooter, making them more independent in their mode of travel.

As shown in Table 3, a total amount of 166.332 respondents used a car for their preserved trip, compared to 151.053 that did not use the car. Among millennials, the distribution is the other way around, as 31.045 out of 53.701 millennials stated they did not use the car. Based on former literature, this finding was expected. In order to be able to derive conclusions from this finding, however, further statistical testing is needed. Controlling variables should be added to reduce Omitted Variable Bias (OVB).

As the difference between someone born in 1982 and 2009 is 27 years, an extra variable for millennials was constructed, dividing the millennials in half. This allows for a more in-depth analysis of millennials, as it is possible that some attitudes and values differ between those groups. For example, late millennials grew up during the beginning of publicly available internet. *Early Millennial* contains all individuals that were born between 1982 and 1995, while *late millennial* is specified as all individuals that were born between 1996 and 2009. The amount of observations per year that classify as millennials, and their distribution over the two defined subgroups can be found in table 4.

| Year | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Millennials | 1997 | 1874 | 2067 | 1868 | 1713 | 4837 | 5162 | 5369 | 5641 | 5736 | 5439 | 5853 | 6275 |
| Early millennials | 1997 | 1874 | 2067 | 1868 | 1713 | 4837 | 4695 | 4526 | 4363 | 4111 | 3797 | 3892 | 4019 |
| Late millennials | 0 | 0 | 0 | 0 | 0 | 0 | 467 | 843 | 1278 | 1625 | 1642 | 1961 | 2256 |

*Table 4. Millennials per year*

As can be seen in Table 3, a total amount of 166.635 respondents used a car for their preserved trip, compared to 151.244 that did not use the car. Among millennials, the distribution is the other way around, as 31.101 out of 53.831 millennials stated they did not use the car. Based on the reviewed literature by Jorritsma and Berveling (2014), this finding was expected. The theory of helicopter parents (Alexander & Sysko, 2012) and the strength of habit in modal choice (Guell, Panter, Jones, & Ogilvie, 2012), predict a higher level of car use among millennials, based on more psychological factors. These factors, like preferences, attitudes and habits, are hard to measure and not included in the dataset. Based on identity theory (Murtagh, Gatersleben, & Uzzell, 2012) and the “Millennial Identity” constructed in chapter 2, it is assumed that these factors are present in the group that is classified as millennials. Contrasting to the helicopter parents phenomena, millennials also have a desire to make the world a better place (Paulin, Ferguson, Jost, & Fallu, 2014). These conflicting attitudes among millennials make it hard to predict a possible effect of being a millennial on car use.

The millennial variables therefore are eligible to test the first hypotheses, which was formulated as follows:

*H1. There is a significant difference between millennials and former generations concerning the modal choice of transportation, controlling for other variables.*

The second hypotheses, however, tests the difference between millennials in 2016 and same-aged individuals in 2006, and was formally formulated as follows:

*H2. There is a significant difference between same-aged individuals in 2006 and millennials in 2016 concerning the modal choice of transportation, controlling for other variables.*

In order to test this hypothesis, a fake-millennial (milage2006) variable is thus needed. This variable was generated by including all individuals of ages 34 and younger, which is equal to the ages that are included in "millennial" in 2016. This way, keeping all other variables constant, the effects these variables have on car use can be compared.

The third and last hypothesis that was formulated to test the spatial effect between millennials and car use. As mentioned by Jorritsma and Berveling (2014), car use declined among millennials, as more millennials are studying instead of working and choose to relocate to the city where they get their education. The hypothesis was formulated as follows:

*H3. There is a significant difference between millennials that live in high-density areas and millennials that do not live in high-density areas, controlling for other variables.*

The interaction effect of density and millennial will be evaluated for this variable. Density was added to the dataset based on the urbanity level present. The urbanity variable was measured on the municipal level. Of each address supplied by the respondents, the density of addresses in an one square kilometer radius was measured. Based on all supplied address densities per municipality, the average was calculated and coded in five possible degrees of urbanity, with boundaries of 500, 1000, 1500, 2000 and 2500 addresses per square kilometer. Three categories for density were drawn out of these five categories for urbanity, dividing density in *low density* (less than 1000 addresses per square kilometer), *average density* (between 1000 and 1500 addresses per square kilometer) and *high density* (more than 1500 addresses per square kilometer). Based on the literature, a positive effect of higher density levels for millennials on car use is expected.

The effect of millennials in densely populated areas was thus already examined by the KiM, as they examined the effect of being a millennial on *car use* controlling for density. Following this finding, the next step to explore this relationship even further is to include an interaction effect of *density* and *millennial*. This allows for a more in-depth analysis of the behavior of millennials in terms of car use, as the effect of *millennial* on car use is isolated.

### 3.2.3 Controlling variables

As the effect of the variables of interest on car use can only be correctly addressed when the model is unbiased, following the Gauss-Markov assumptions. In effect, the preferred model should be "BLUE" (Best linear Unbiased Estimator). In order to formulate unbiased models that can be used to support the hypotheses, controlling variables should be added to improve the model and decrease the implied bias. Also, adding controlling variables decreases the OVB.

The first controlling variable that will be included is *density*. As the interaction effect of *millennial* and *density* is a variable of interest, density itself will be added as a controlling variable. *Density* is expected to have some explaining power in terms of car use, based on the research by Ramezani, Pizzo, & Deakin (2018). Density was not directly available in the dataset. Urbanity, which is based on density, however was availale in the dataset and therefore could be used to generate a variable for density. This variable takes on the value *Low density*, *Average density* or *High density* and is thus categorical in nature.

*Density* was thus based on the variable urbanity, which was already available in the dataset. As this variable was used to generate a variable for *density*, it is straightforward to observe a high level of correlation between both variables. For interpretation considerations, *density* was chosen as the main variable for land-use. Municipal size and province of residence were also included, to also control for the location of the city of residence of the respondent. This is based on the assumption that an average-density city in the edge-city (*Zuid-Holland, Noord Holland* and *Utrecht*) differs in terms of for example accessibility and availability of public transport from a city outside the edge-city, for example Zwolle (*Overijssel*) or Groningen (*Groningen*).

Next, controlling variables regarding trip length, trip purpose, trip day and travel costs were included. The natural logarithm of distance is added to represent trip length in the model. As can be reasoned, longer trip lengths decrease the viable transport mode options of the respondents. As shown in table 2 in Appendix A, the mean of trip lengths for car users is around four times the mean of trip lengths for non car users. Next, the purpose of the trip might also influence the modal choice of transportation. The variable *doel6* was generated to control for five main trip purposes, or "other" as the sixth possibilty. It can be argued that when more time is available for the individual, slower modes of transport are viable alternatives to car use, while when an individual is under time pressure, for example for showing up on time for work, these alternatives are not viable. This implies that the day of the weke also determines modal choice, as more individuals will have to work between Monday and Friday, while they can spend free time in the weekends. In those weekends, it is more likely that individuals undertake social actions, or go shopping. This reasoning proposes a possible effect of the day of the week on car use. Therefore, it will be included by incorporating a dummy variable for trips made in the weekend. As shown in Table 5 below, a relatively low number of work trips is undertaken in the weekend, while social trips are undertaken more in the weekend. As could also be expected, more trips for education purposes were undertaken during the week.

| | Work | Shopping | Services | Social | Education | Other |
|---|---|---|---|---|---|---|
| Monday | 18532 | 9544 | 2554 | 12871 | 2452 | 3200 |
| Tuesday | 19245 | 9756 | 2584 | 12574 | 2458 | 3127 |
| Wednesday | 17930 | 10463 | 2346 | 13463 | 2275 | 2990 |
| Thursday | 18210 | 10262 | 2385 | 12857 | 2283 | 3004 |
| Friday | 15373 | 13229 | 2053 | 13853 | 1729 | 2904 |
| Saturday | 3837 | 14872 | 737 | 17356 | 220 | 1561 |
| Sunday | 2066 | 2694 | 509 | 25606 | 93 | 1331 |

*Table 5. Trip purposes per weekday*

As mentioned by Garcia-Sierra, Miralles-Guasch, Martínez-Melo, & Marquet (2018), car ownership has a positive effect on car use, as the permanent availability makes it easier to choose for this mode of transportation. In this case, the original costs of purchasing a car are neglected, as they are seen as sunk costs. This leads to car owners using their car for more trips than necessary, as availability quite often leads to habituation. Multiple cars available in the household increase the availability even further. Therefore, the amount of household cars will also be added as a controlling variable.

An individual can only make use of a household car autonomuously when they are in possession of a drivers permit. If not, these individuals are dependent on other household members or friends that do own a drivers license when they want to make use of a car, decreasing their freedom in the modal choice of transportation. If a respondents was younger than eighteen years, it was put in a separate category, as this individual is too young to be able to obtain a drivers license. However, this extra category is necessary, as those young respondents might be more dependent on their parents or carekeepers than adults without a license are. Also, not having a license for them is not a choice they made, but rather a governmental policy that does not allow them to drive a car yet. Programs to allow underage individuals to obtain a drivers license are upcoming, but were not considered in this research.

As already mentioned, the costs already incurred for purchasing a car are seen as sunk costs when choosing a certain mode of transportation. However, a car uses fuel to operate, which implies direct costs for the car owner in terms of having to buy fuel and indiect costs for society in terms of pollution. Fuel prices are highly volatile and differ per location. Therefore, *fuel price* can be seen as a proxy for transport costs for individuals. As prices are highly volatile, the cheapest price recorded during each year was recorded for all years in the dataset. Also, as location was found to be important in determining fuel prices, the cheapest price per province was included. This way, the transport costs differ both per year and per province. In order to acknowledge similarities between years and provinces, cluster correction was applied to the model.

Last, the sociodemographics of respondents available in this dataset are considered as controlling variables. As it is impossible to explicitly incorporate habits and attitudes to the model, sociodemograhpics based on year of birth and societal position were added. Also, with the millennial dummy, most of the habits and attitudes of millennials are already assumed to be captured, following the "Millennial Identity", while not being a millennial assumes that a respondent does not hold the values and attitudes that are seen as inherent to millennials. In order to control as good as possible for these psychological factors, *age*, *gender*, *education*, *income* and *working hours* were added as controlling variables.

There was found some evidence in former literature that women were driving less often and less far in the past (Böcker, van Amen, & Helbich, 2017), as differences in emotional feelings between men and women were attached to driving (Siren & Haustein, 2013). This implies that women might be less reserved in using a car compared to men. Therefore, *gender* is added to the model to control for this possible effect. Considering age, Newbold et al. (2005) found evidence that dependence on car use increased as population aged. This means it is expected that car use increases when age increases. The variable age is added as a continuous variable to control for this effect.

Both *education* and *income* are expected to have an effect on car use. For *education*, higher education often implies more commitment to society and therefore the desire to make a sustainable choice regarding the modal choice of transportation. Higher income, is expected to increase the flexibility of individuals in their modal choice (Watkins, 2016). Education and income are often closely related, as higher-educated individuals tend to be able to fulfill higher paid jobs. In order to check for collinearity, the correlation matrix for education and the proposed incomeproxy was constructed. Returning a value of 0,3036 implies that both variables are not directly related, however they do have some relationship. Both variables were thus added to the model.

Working Hours, which also contains unemployment, was considered as a controlling variable, as Böcker, van Amen, & Helbich (2017) found that unemployment has a negative effect on car use. As can be expected, unemployment is likely to lead to lower (personal) income, which might disallow an individual to purchase or maintain car ownership. Also, the need for a car might be lower, as unemployment often means less trips are undertaken. Therefore, the level of employment in terms of unemployment, part-time or full-time work are also added as a controlling variable.

The last controlling variable that was included in the model is the year of observation. As the dataset consists of all years between 2005 and 2017, and technology and infrastructure changed over these years, there might be a timing effect on *car use*. As stated by Verlaan (2018), traffic jams are still increasing and therefore individuals look for alternatives for car use to reduce time spent in traffic.

Now that all variables were introduced and discussed, the next section will discuss the model and the analysis method used to validate or invalidate the hypotheses.

### 3.2.3 Excluded variables

Based on the former section, the model can be constructed. However, some of the variables described above do control for the effect of being millennial on car use, but are also expected to be influenced by the millennial variable. As millennials were defined as a distinct cohort with their own norms and values, it can easily be argued that a millennial, who in general is less likely to use a car based on these norms and values, also is less likely to seek possession of a car, as this person is not going to use it. In this case, car ownership is rather the mechanism that explains why a millennial is less likely to use a car, instead of a controlling variable that isolates the effect of being millennial on car use. The same reasoning holds for obtaining a driver’s license. If a millennial has an aversion for car use, it can be expected that this individual will also not get a driver’s license, as it is of no use for the individual. This, again, describes the mechanism of the effect instead of controlling for the effect.

Density can also be seen as a mechanism rather than a controlling variable when predicting the effect of being a millennial on *car use*. Following the theory of Residential Self Selection (RSS), millennials tend to relocate to the place where they work or study, which in most cases is in an urban environment. Therefore, it can be argued to exclude density from the model as well. However, as being millennial is related to density through the mechanism of RSS, the interaction effect between *millennial* and *density* isolates the effect of RSS and thus shows the unbiased effect of *millennial* on car use.

In the literature, ethnicity was also found to have an impact on car use (Klocker, Toole, Tindale, & Kerr, 2015). However, as information about ethnicity or country of birth was not available for the observations in the MON-database, this variable was also excluded for the years it was available for.

## 3.3 Methodology

This section will first discuss the process of selecting the best suited analysis method for this regression analysis. Next, the general model will be formulated, as well as the variations on this model that are used to test the hypotheses

### 3.3.1 Analysis Method

As the dependent variable, car use, was defined as a binary variable, three possibilities for regression analysis are present. The most basic out of the three is the Linear Probability Model (LPM), which draws on standard OLS mechanics (Wooldridge, 2014). LPM returns the response probability, which is linear in the parameters of all independent variables in the model. This means that a change in one of the independent variables means that the probability that the variable for car use is 1, which means a respondent used a car, will change as well (Wooldridge, 2014). This means that under LPM a probability is returned, which states how big the probability is that the respondent used a car. However, as parameters are assumed to be linear, some issues arise. The first issue is found in interpreting the parameters in the model. As this model assumes all parameters are linear, going from zero to one car per household will have the same effect as going from three to four cars in the household. As can be imagined, going from zero to one car is expected to have a bigger effect. The second limitation for LPM is, that given certain values in the dataset, the probability that the dependent variable is true might exceed “1” or be negative. As probabilities are always defined between 0 and 1, those values are not interpretable themselves. However, as the probabilities can be used to calculate the percent correctly predicted, the out of bound values predicted in the LPM can still be used. The percent correctly predicted is also referred to as a test for goodness of fit. A third issue is found in the Gauss-Markov assumptions, as the binary nature of the dependent variable implies that a LPM must contain heteroskedasticity. This, however, is easily fixable by using robust standard errors (Wooldridge, 2014).

LPM thus has some disadvantages. Still, it can be used as a method of analysis when these problems are addressed and if possible fixed. The main advantage, and therefore reason, to consider LPM as an analysis method in this research, is that the estimates are easily interpretable. As all parameters in LPM are linear, an increase in an independent variable means a linear increase in the probability of the variable of interest being equal to 1, keeping all other variables constant.

A second methodology that is possible is a logit regression. As mentioned, LPM assumes linearity in all its variables. This implies that the effect of a similar change in a particular variable will have the same effect for all levels of this variable. For example, under LPM, an one kilometer increase in trip distance will have the same effect on car use for a rise from 0,1 kilometer to 1,1 kilometer as it would have for an increase from 100 kilometer to 101 kilometer. In general, this effect will not be linear though, as it is likely that the increase at lower distances will have a bigger effect on car use, as other modes of transport like walking or biking are still feasible, while the effect on car use will be lower when the distance increases by the same amount at higher distances, as car use is already expected to be higher. Logit regression fits the model to the data that is used, thus ensuring the outcome of the model to be between 0 and 1 for all values of all variables. Logit regression, however, estimates coefficients as log-odds, which require some calculations to be interpreted.

As mentioned, a goodness of fit test for LPM would be the percent correctly predicted (PCP). This goodness of fit tests also applies for logit models (Wooldridge, 2014). The PCP is defined as the amount of predicted values that correspond with the values of the observations. As LPM and logit return values in terms of probabilities, predictions equal to or above 0,5 will be rounded up to 1. This means that the model predicts that an individual uses a car.

As non-linearity of variables is expected in this research, logit was chosen as the methodology of preference. The next section will discuss the calculation of probabilities out of log-odds.

### 3.3.2 Logistic regression

A standard logit model estimates the coefficients of the variables as log-odds. These coefficients are not relevant to interpret on their own, as they show the change in the logarithm of the odds that the dependent variable is one, for a one unit change in the variable of that particular coefficient. These log-odds can however be used to calculate the probability of  Formula 1 shows the formula of the logit regression mathematically.

$$\ln \frac{P(y=1|x_i)}{1-P(y=1|x_i)} = b_0 + b_1x_1 + b_2x_2 + .. + b_nx_n + u \quad for\ i = 1,2,..,n$$

*Where*

$P(y = 1|x_i) = the\ probability\ that\ y\ equals\ 1, given\ values\ for\ x_i$

$x_i = the\ value\ of\ variable$ i

$b_0 = the\ constant$

$b_i = the\ estimated\ coefficient\ of\ variable\ x$

$u = the\ error\ term$

$n = the\ number\ of\ independent\ variables\ in\ the\ model$

The logarithm can be eliminated by taking the exponent of both sides of the equation. By doing this, the formula now returns the odds of car use given the levels of all included variables $x_i$. Mathematically, the formula for the odds of car use is then formulated as:

$$\frac{P(y=1|x_i)}{1-P(y=1|x_i)} = e^{b_0+\Sigma(x_ib_i)+u}\ for\ all\ i.$$

This formula thus calculates the odds that the dependent variable has a positive outcome (y=1). For this research, this means the odds that a respondent used a car for its trip. The odds are formally formulated as “The odds of the dependent variable being true are $e^{b_0+\Sigma(x_ib_i)+u}$ to 1”. This also means that the odds of the dependent variable being false are 1 to $e^{b_0+\Sigma(x_ib_i)+u}$. As odds are still abstract to interpret, odds will be used to calculate the corresponding probability. Formula 3 shows the mathematics behind this calculation.

$$P(y=1|x_i) = \frac{e^{b_0+\Sigma(x_ib_i)+u}}{1+\ e^{b_0+\Sigma(x_ib_i)+u}}\ \ for\ all\ i$$

As the probability thus is an exponential function, the effect of a rise in $x_i$ will differ for different values of $x_i$, keeping all other variables constant. The marginal effect can be calculated by increasing a particular variable $x_j$ and keeping all other variables $x_i$ constant.

Now that the dataset, the variables and the methodology of preference were all discussed, the next section will present the models that will be used to test the hypotheses stated earlier in this research.

## 3.4 Model description

This section will give an overview of the models that will be used in order to test the formulated hypotheses.

### 3.4.1 The basic model

The basic model will consist only of the dependent variable, car use, and the variable of interest, which is the millennial variable. This model is likely to suffer from OVB, but can be used as a baseline for the advanced models. Model 1 is thus formulated as

$$car\ use = b_0 + b_{millenial} millennial + u$$

*Where:*

$$b_0 = the\ constant$$

$$b_{millenial} = the\ estimated\ coefficient\ of\ variable\ millenial$$

$$u = the\ error\ term$$

As mentioned, this model only describes the effect of being millennial on *car use*, without controlling for other variables. Therefore, no conclusions can be derived from the output of this model. It is only used to lay a foundation for the other models. As there is only one independent variable, no check on correlation was conducted.

### 3.4.2 The standard model

Model 2 will be estimated in order to test the first hypothesis, which was formulated as

*H1. There is a significant difference between millennials and former generations concerning the modal choice of transportation, controlling for other variables.*

The other variables mentioned which are referred to in this hypothesis were described in section 3.2.3. This model thus expands on Model 1, but adds the controlling variables. Model 2 is formally formulated as

$$\begin{aligned} car\ use = \ & b_0 + b_{millenial} millennial + b_{density} density + b_{gemgr} gemgr + b_{prov} prov \\ & + b_{lnkmafstr} lnkmafstr + b_{doel6} doel6 + b_{weekend} weekend \\ & + b_{brandstofprijs} brandstofprijs + b_{leeftijd} leeftijd + b_{geslacht} geslacht \\ & + b_{education} education + b_{incomeproxy} incomeproxy + b_{werkuren} werkuren \\ & + b_{year} jaar + u \end{aligned}$$

*Where:*

$$b_0 = the\ constant$$

$$b_i = the\ estimated\ coefficient\ of\ variable\ i \qquad i \in \{milennial, \dots, jaar\}$$

$$u = the\ error\ term$$

As discussed, both car ownership and possession of a driver's permit were excluded from the model as they are seen as a mechanism rather than a controlling variable.

As some controlling variables might have some relationship to each other, a correlation matrix was calculated to check for correlation between the controlling variables. Table 6 shows the correlation matrix for model 2.

| Correlation | millennial | density | gemgr | prov | Inkmafstr | doel6 | weekend | brandstofprijs | leeftijd | geslacht | education | incomeproxy | werkuren | jaar |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| millennial | 1 | | | | | | | | | | | | | |
| density | 0,0408 | 1 | | | | | | | | | | | | |
| gemgr | 0,0588 | 0,7076 | 1 | | | | | | | | | | | |
| prov | -0,006 | 0,1852 | 0,0501 | 1 | | | | | | | | | | |
| Inkmafstr | -0,0102 | -0,0699 | -0,0632 | -0,0242 | 1 | | | | | | | | | |
| doel6 | 0,0661 | -0,0121 | -0,0176 | 0,0097 | -0,1118 | 1 | | | | | | | | |
| weekend | 0,0007 | 0,0331 | 0,0155 | 0,0664 | 0,0042 | 0,1952 | 1 | | | | | | | |
| brandstofprijs | 0,1284 | -0,0203 | 0,0221 | -0,0455 | 0,0071 | 0,0315 | -0,0358 | 1 | | | | | | |
| leeftijd | -0,6511 | -0,0458 | -0,0676 | 0,0122 | -0,0648 | 0,0292 | 0,0036 | -0,008 | 1 | | | | | |
| geslacht | 0,0038 | 0,0135 | 0,014 | -0,0044 | -0,1479 | 0,0836 | 0,0017 | 0,037 | -0,0061 | 1 | | | | |
| education | -0,0185 | 0,0902 | 0,0904 | -0,0116 | 0,0892 | -0,0528 | 0,0014 | 0,0499 | -0,1576 | -0,039 | 1 | | | |
| incomeproxy | -0,053 | -0,0133 | -0,0359 | 0,0365 | 0,1446 | -0,0757 | 0,0043 | 0,0235 | -0,1039 | -0,1635 | 0,3036 | 1 | | |
| werkuren | -0,0268 | 0,0228 | 0,026 | 0,0078 | 0,1975 | -0,2932 | -0,0071 | -0,0691 | -0,3951 | -0,1927 | 0,3083 | 0,3324 | 1 | |
| jaar | 0,1819 | 0,0586 | 0,0781 | 0,012 | 0,0041 | 0,0443 | 0,0124 | 0,6104 | 0,0174 | 0,0407 | 0,0712 | 0,0946 | -0,0821 | 1 |

*Table 6. Correlation Matrix for Model 2*

As can be seen in Table 6, the correlation coefficients imply that there are no correlations between controlling variables for all instances except three. *Density* and *gemgr*(municipal size) are strongly correlated (altough no perfect correlation was found). As both are land use variables, and *density* was based on the amount of addresses in a one kilometer radius, this correlation can be explained. During analysis, this correlation should be taken into account when evaluating the effect on car use.

*Leeftijd* (age) and the millennial variable were also found to contain correlation. This can be explained by the fact that certain aged respondents are marked as "millennial", and therefore, the age of the respondent will be lower when they are a millennial, compared to not being amillennials.

The last noteworthy correlation was found between *brandstofprijs* (fuel price) and *jaar* (year of observation). As the fuel prices were included for each year in the dataset based on the CBS data on average fuel prices per year in the Netherlands, correlation is expected. Cluster correction will be used to correct for this correlation.

### 3.4.3 Models in a certain time period

Hypotheses 2 is somewhat more difficult to test, as it seeks to compare the effect of *millennial* on car use nowadays (2016) with the the effect of same-aged individuals in 2006. The basic model will thus be formulated in both cases, so a comparison of the magnitude and sign can be made. Then, the controlling variables as described in model 2 will be added to control for possible time-dependent changes in those controlling variables. These models will be used to test hypothesis 2, which was formulated as

*H2. There is a significant difference between same-aged individuals in 2006 and millennials in 2016 concerning the modal choice of transportation, controlling for other variables.*

The next subsections provide the models that will be used to evaluate this hypothesis.

#### *3.4.3.1 The model for 'Millennial-aged" individuals in 2006*

The basic model was thus adjusted, so that a comparison can be made between same-aged individuals in 2016 and 2006. The millennial variable was dropped from the model, while the fake-millennial variable *milage2006* was added instead. Model 1a tests the effect of being millennial-aged on car use in 2006 and is thus formulated as

$$car\ use = b_0 + b_{milage2006} milage2006 + u$$

*Where:*

$b_0 = the\ constant$

$b_{milage2006} = the\ estimated\ coefficient\ of\ variable\ milage2006$

$u = the\ error\ term$

$jaar = 2006$

The controlling variables also used in Model 2 will be added to control for other possible determinants of car use. Model 2a is thus defined as

$$\begin{aligned} car\ use = b_0 &+ b_{milage2006} milage2006 + b_{density} density + b_{gemgr} gemgr + b_{prov} prov \\ &+ b_{lnkmafstr} lnkmafstr + b_{doel6} doel6 + b_{weekend} weekend \\ &+ b_{brandstofprijs} brandstofprijs + b_{leeftijd} leeftijd + b_{geslacht} geslacht \\ &+ b_{education} education + b_{incomeproxy} incomeproxy + b_{werkuren} werkuren + u \end{aligned}$$

*Where:*

$b_0 = the\ constant$

$b_i = the\ estimated\ coefficient\ of\ variable\ i \qquad i \in \{milennial, \dots, werkuren\}$

$u = the\ error\ term$

$jaar = 2006$

Variable *jaar* was dropped from the model, as only entries from 2006 were evaluated in this model. The correlation matrix for this model was summarized in Appendix A Table 4 and can be found in Appendix A. The correlation coefficients are slightly different compared to Model 2, thus no inherent problems were found with Model 2a.

#### *3.4.3.2 The model for millennials in 2016*

For this model, Model 1 (The basic model) and Model 2 can be used, with only one alteration. As the only year of interest for this variable is 2016, *jaar* was dropped from the model, as only data of 2016 was used. Model 1b is thus formulated as

$$car\ use = b_0 + b_{millennial} millennial + u$$

*Where:*

$b_0 = the\ constant$

$b_{millennial} = the\ estimated\ coefficient\ of\ variable\ millennial$

$u = the\ error\ term$

$jaar = 2016$

Basically, Model 1b is equal to Model 1, except that only observations from 2016 are used. The same holds for Model 2b, but only *jaar* was dropped, as only observations from 2016 were included in this (sub)model. Model 2b is thus formulated as

$$
\begin{aligned}
car\ use = \ & b_0 + b_{millenial}millennial + b_{density}density + \ b_{gemgr}gemgr + b_{prov}prov \\
& + b_{lnkmafstr}lnkmafstr + b_{doel6}doel6 + b_{weekend}weekend \\
& + b_{brandstofprijs}brandstofprijs + b_{leeftijd}leeftijd + b_{geslacht}geslacht \\
& + b_{education}education + b_{incomeproxy}incomeproxy + b_{werkuren}werkuren + u
\end{aligned}
$$

*Where:*

$b_0 = the\ constant$

$b_i = the\ estimated\ coefficient\ of\ variable\ i \qquad i \in \{milennial, \dots, werkuren\}$

$u = the\ error\ term$

$jaar = 2016$

As for Model 2 and Model 2a, the correlation matrix for Model 2b was constructed. The correlation matrix for this model was summarized in Appendix A Table 5 and can be found in Appendix A. As with Model 2 and Model 2a, the same high correlations were found.

Model 1a and Model 1b can be used to check for a possible difference in behavior between millennials and same-aged respondents from 10 years ago. As these models do not control for other influences, no strong conclusions can be drawn from the outcomes. Model 2a and Model 2b do also correct for other factors and is therefore more suitable for analysis. It should also be noted that the year of observation, and therefore the year in which the observant chose to make use of a car or not, can also be of influence, as economic and/or geographic situation might differ.

### 3.4.4 The final model

The last hypothesis seeks to evaluate the difference between millennials in high-density areas and millennials that live in lower-density areas and was based on the earlier findings of Jorritsma & Berveling (2014). This hypothesis was formulated as

*H3. There is a significant difference between millennials that live in high-density areas and millennials that do not live in high-density areas, controlling for other variables.*

In order to isolate this effect, the interaction effect of *millennial* and *density* was added to the model. This required some alterations for *density,* as both *millennial* and *density* are categorical in nature, corresponding dummy variables were used for each possibility. This means that three dummy variables were constructed in order to construct the interaction term. Model 3 thus expands on model 2, taking into account all observations in the dataset. Model 3 was then formulated as

$$
\begin{aligned}
car\ use = \ & b_0 + b_{millennial}millennial + b_{middens}middens + b_{highdens}highdens \\
& + b_{mil\_middens}mil\_middens + b_{mil\_highdens}mil\_highdens + \ b_{gemgr}gemgr \\
& + b_{prov}prov + b_{lnkmafstr}lnkmafstr + b_{doel6}doel6 + b_{weekend}weekend \\
& + b_{brandstofprijs}brandstofprijs + b_{leeftijd}leeftijd + b_{geslacht}geslacht \\
& + b_{education}education + b_{incomeproxy}incomeproxy + b_{werkuren}werkuren \\
& + b_{year}jaar + u
\end{aligned}
$$

*Where:*

$b_0 = the\ constant$

$b_i = the\ estimated\ coefficient\ of\ variable\ i \qquad i \in \{milennial, \dots, jaar\}$

$u = the\ error\ term$

The effect of *density* was now captured by *middens* and *highdens* for respectively mid-density areas and high-density areas. Low density was thus taken as the reference level for density, and therefore the corresponding dummy is not included in the model. Not being a millennial was also taken as the reference category. *Millennial* thus shows the effect of being a millennial compared to not being a millennial, controlling for other variables. The interaction effect is captured by variables *mil_middens* for observations that are both millennial and living in a mid-density environment and *mil_highdens* for observations that are both millennial and living in a high-density environment. As both not being millennial and living in a low-density area are in the reference category, no dummy variable for this effect is present, as it is already present in the constant.

The correlation matrix for Model 3 was constructed and given below.

| Correlation Matrix | millennial | middens | highdens | mil_dmid | mil_dhigh | gemgr | prov | lnkmafstr | doel6 | weekend | brandstofprijs | leeftijd | geslacht | education | incomeproxy | werkuren | jaar |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| millennial | 1 | | | | | | | | | | | | | | | | |
| middens | -0,0198 | 1 | | | | | | | | | | | | | | | |
| highdens | 0,0447 | -0,432 | 1 | | | | | | | | | | | | | | |
| mil_dmid | 0,3943 | 0,3571 | -0,1543 | 1 | | | | | | | | | | | | | |
| mil_dhigh | 0,6575 | -0,1478 | 0,3422 | -0,0528 | 1 | | | | | | | | | | | | |
| gemgr | 0,0588 | -0,1579 | 0,7021 | -0,0527 | 0,2719 | 1 | | | | | | | | | | | |
| prov | -0,006 | 0,0319 | 0,1543 | 0,0099 | 0,0358 | 0,0501 | 1 | | | | | | | | | | |
| lnkmafstr | -0,0102 | 0,0005 | -0,0633 | -0,0017 | -0,037 | -0,0632 | -0,0242 | 1 | | | | | | | | | |
| doel6 | 0,0661 | 0,0019 | -0,0117 | 0,0277 | 0,0405 | -0,0176 | 0,0097 | -0,1118 | 1 | | | | | | | | |
| weekend | 0,0007 | 0,0016 | 0,0293 | -0,0011 | 0,0096 | 0,0155 | 0,0664 | 0,0042 | 0,1952 | 1 | | | | | | | |
| brandstofprijs | 0,1284 | -0,0139 | -0,0127 | 0,0474 | 0,083 | 0,0221 | -0,0455 | 0,0071 | 0,0315 | -0,0358 | 1 | | | | | | |
| leeftijd | -0,6511 | 0,0161 | -0,0479 | -0,2603 | -0,418 | -0,0676 | 0,0122 | -0,0648 | 0,0292 | 0,0036 | -0,008 | 1 | | | | | |
| geslacht | 0,0038 | 0 | 0,0122 | -0,0015 | 0,0109 | 0,014 | -0,0044 | -0,1479 | 0,0836 | 0,0017 | 0,037 | -0,0061 | 1 | | | | |
| education | -0,0185 | -0,0079 | 0,0846 | -0,0221 | 0,0239 | 0,0904 | -0,0116 | 0,0892 | -0,0528 | 0,0014 | 0,0499 | -0,1576 | -0,039 | 1 | | | |
| incomeproxy | -0,053 | 0,0192 | -0,0198 | -0,0021 | -0,0779 | -0,0359 | 0,0365 | 0,1446 | -0,0757 | 0,0043 | 0,0235 | -0,1039 | -0,1635 | 0,3036 | 1 | | |
| werkuren | -0,0268 | -0,0013 | 0,0211 | -0,0088 | -0,0229 | 0,026 | 0,0078 | 0,1975 | -0,2932 | -0,0071 | -0,0691 | -0,3951 | -0,1927 | 0,3083 | 0,3324 | 1 | |
| jaar | 0,1819 | -0,0333 | 0,0663 | 0,0582 | 0,152 | 0,0781 | 0,012 | 0,0041 | 0,0443 | 0,0124 | 0,6104 | 0,0174 | 0,0407 | 0,0712 | 0,0946 | -0,0821 | 1 |

*Table 7. Correlation matrix for Model 3*

Compared to Model 2, the same correlation coefficients (above 0,5) were highlighted. Noteworthy is the correlation between *density* and *gemgr* (municipality size) can be isolated to *highdens*, thus meaning that correlation between municipal size and density is stronger for high-density areas. As for Model 2, high correlation coefficients were found for *leeftijd*(age) and *millennial,* as wel as for *barndstofprijs*(fuel price) and *jaar*(year of observation). These correlations were discussed earlier and are not seen as a problem in this model as well.

One new noteworthy correlation was found, between *millennial* and *mil_dhigh*. This means that the initial effect of being a millennial is correlated with being a millennial living in a high density area. The interaction term is based on the main variable *millennial*, and therefore correlation was expected. As the interaction term is the variable of interest in this model, explaining power was chosen over the possible bias that might be present in this term, thus preserving both variables in the model.

The basic models needed to test the hypotheses of this research were now defined. Ne next section will discuss two more models that can be used for more extensive research. The first extensive model further defines the millennial variable in two groups, *early millennial* and *late millennial*, as the millennial cohort as a whole is rather broad. The second extensive model redefines the dependent variable.

### 3.4.5 Extensive models

This section provides two models that can be used to further analyze the effect of being a millennial on car use.

#### 3.4.5.1 Redefining millennials

Until now, the millennial cohort was seen as one group, defined as all individuals born between 1982 and 2009. As discussed in 3.2.2.1, the millennial cohort will be split into two groups. *Early_millennials* are born between 1982 and 1995, while *late_millennials* are born between 1996 and 2009. Model 4 evaluates the effects of those distinct groups of millennials on car use. Formally, Model 4 is formulated as

$$\begin{aligned} car\ use = \ & b_0 + b_{early\_millennial}early\_millennial \ + b_{late\_millennial}late\_millennial \\ & + b_{middens}middens + b_{highdens}highdens + b_{earlym\_dmid}earlym\_dmid \\ & + b_{earlym\_dhigh}earlym\_dhigh + b_{latem\_dmid}latem\_dmid \\ & + b_{latem\_dhigh}latem\_dhigh\ b_{gemgr}gemgr + b_{prov}prov + b_{lnkmafstr}lnkmafstr \\ & + b_{doel6}doel6 + b_{weekend}weekend + b_{brandstofprijs}brandstofprijs \\ & + b_{leeftijd}leeftijd + b_{geslacht}geslacht + b_{education}education \\ & + b_{incomeproxy}incomeproxy + b_{werkuren}werkuren + b_{jaar}jaar + u \end{aligned}$$

*Where:*

$$b_0 = the\ constant$$

$$b_i = the\ estimated\ coefficient\ of\ variable\ i \qquad\qquad i \in \{milennial, \dots, jaar\}$$

$$u = the\ error\ term$$

Model 4 thus replaces *millennial* with two sub-cohorts. The interaction effect with density was also added for both separate groups. As with Model 3, the reference category is a non-millennial, living in a low-density area. The correlation matrix for Model 4 was constructed and can be found in Appendix A Table 6. An interesting finding is that the correlation between *millennial* and *density* was further isolated, as only a strong correlation between early millennials and high density exists. For the same considerations as in Model 3, both variables were kept in the model.

#### 3.4.5.2 Redefining car use

Until now, the dependent variable has been *car use*, which was defined as all individuals that used a car for their recorded trip. In this definition, passengers were also included. Now, it might also be relevant to evaluate the effect of being millennial on active car use, or driving a car. As car use declined, attitudes towards car use remained stable for the millennial generation (Jorritsma & Berveling, 2014). As car driving is associated with freedom, redefining car use might give an additional insight in the choices of millennials regarding the modal choice of transportation. Therefore, a model was constructed that replaced *car use* with the dependent variable *cardriver* and was thus defined as all observants that drove a car for their recorded trip. Furthermore, Model 5 expands on Model 3, thus treating the millennial cohort as one group. Model 5 was the defined as

$$
\begin{aligned}
car\ driver = \ & b_0 + b_{millennial} millennial + b_{middens} middens + b_{highdens} highdens \\
& + b_{mil\_middens} mil\_middens + b_{mil\_highdens} mil\_highdens + \ b_{gemgr} gemgr \\
& + b_{prov} prov + b_{lnkmafstr} lnkmafstr + b_{doel6} doel6 + b_{weekend} weekend \\
& + b_{brandstofprijs} brandstofprijs + b_{leeftijd} leeftijd + b_{geslacht} geslacht \\
& + b_{education} education + b_{incomeproxy} incomeproxy + b_{werkuren} werkuren \\
& + b_{jaar} jaar + u
\end{aligned}
$$

*Where:*

$b_0 = the\ constant$

$b_i = the\ estimated\ coefficient\ of\ variable\ i \qquad i \in \{milennial, \dots, jaar\}$

$u = the\ error\ term$

Basically, Model 3 was refitted for *car driver* instead of *car use*. As the same variables were used, the correlation matrix for Model 5 is equal to the correlation matrix of Model 3.

# 4. Results

Now that all models have been formulated, this section will analyze the outcomes of the models. Chapter 5 will discuss these outcomes and test the hypotheses.

## 4.1 Model 1

Model 1, also referred to as the basic model, only considered the effect of *millennial* on *car use*. The logistic regression yielded the results as stated in Table 8.

| Coefficient | Value | Standard Error | Significance |
|---|---|---|---|
| **Constant** | 0,180749 | 0,0095718 | 0,000 |
| **Millennial** | -0,4936685 | 0,0039107 | 0,000 |
| **Observations** | 317.388 | | |
| **Pseudo R-Squared** | 0,0061 | | |
| **Log likelihood** | -218.275 | | |

*Table 8. Output of Model 1*

As shown in Table 6, all observations were included in Model 1. The intercept is significant at the 5% significance level and has a positive value of 0,180749 ceteris paribus. The log-odds of car use are thus 0.180749 when all other variables, in this case only *millennial*, are zero. As *millennial* is a categorial (dummy) variable, the intercept can be interpreted by itself. As stated in 3.3.2, the probability can be calculated from the log-odds by using

$$P(y = 1|x_i) = \frac{e^{b_0+\sum(x_i b_i)+u}}{1 + e^{b_0+\sum(x_i b_i)+u}} \quad for\ all\ i$$

The probability that an individual that is not millennial will use a car can thus be calculated by using the coefficients from Model 1 in the probability equation:

$$P(car\ use = 1|millennial = 0) = \frac{e^{(0{,}180749+(-0{,}4936685*0)}}{1 + e^{(0{,}180749+(-0{,}4936685*0)}}$$

Calculating the results, a value of 0,545 was found, which means that there is a 54,5% chance that an individual that is not a millennial makes use of a car.

The only independent variable in this model, *millennial*, has a negative effect on car use. It is significant at all relevant confidence levels and has a magnitude of -0.4936685 ceteris paribus. This means that the log-odds are 0,4936685 lower if the individual is a millennial, compared to not being a millennial. For an individual that classifies as a millennial, the probability equation is

$$P(car\ use = 1|millennial = 1) = \frac{e^{(0{,}180749+(-0{,}4936685*1)}}{1 + e^{(0{,}180749+(-0{,}4936685*1)}}$$

This yields a value of 0,4224, which means that a millennial has a 42.24% probability of using a car. This means that being millennial lowers the probability of using a car by 12.26%, ceteris paribus.

The percent correctly predicted will be calculated and serve as a Goodness of Fit-test. As shown in Table 9, for 119,961 observations the model predicted that the respondent used a car, while in truth no car was used.

| Prediction\car use | No | Yes | Total |
|---|---|---|---|
| **No** | 31.017 | 22.683 | 53700 |
| **Yes** | 119.961 | 143.727 | 263688 |
| **Total** | **150.978** | **166.410** | |

*Table 9. Comparison of predicted values and observations of car use for Model 1*

To calculate the PCP, all predicted values for car use were compared with the actual value of *car use*. The findings were summarized in table 10.

| | |
|---|---|
| Car use correctly predicted | 86% |
| Non-car use | 21% |
| Total correctly predicted | 55% |

*Table 10. PCP values for Model 1*

As shown, 86% of actual car users was correctly predicted. This figure suggests the model is stronger than it actually is. As is also shown in table xx, non-car use was only predicted right in 21% of the observations. Therefore, in total only 55% of the observations were predicted correctly, using the basic model.

As *millennial* is the only variable in the model, predictions are only based on the value of *millennial:* all millennials in the dataset were predicted not to use a car, as the model predicted the coefficient of *millennial* to have a negative effect on *car use*. Therefore, controlling variables are needed in order to correct for OVB.

## 4.2 Model 2

Model 1 thus shows a negative, significant effect of *millennial* on *car use*. However, as only one variable was included in the model, the predicted values were solely based being a millennial or not. OVB is thus likely to be present in Model 1. Model 2 was specified to include other factors that might be relevant when estimating *car use*. The logistic regression of Model 2 yielded the results as stated in Table 11. In order to control for the heteroskedasticity in *brandstofprijs*, cluster correction was applied.

| Coefficient | Value | Robust Standard Error | Significance | Coefficient | Value | Robust Standard Error | Significance |
|---|---|---|---|---|---|---|---|
| **Intercept** | -16,64218 | 0,0095718 | 0,011 | **Weekend** | 0,182982 | 0,0137961 | 0,000 |
| **Millennial** | -0,2623367 | 0,0258899 | 0,000 | **Fuel price** | 10,16642 | 4,787229 | 0,034 |
| **Low Density** | 0,000 | 0,000 | 0,000 | **Age** | 0,001104 | 0,001 | 0,050 |
| **Average Density** | -0,1840493 | 0,0166213 | 0,000 | **Gender** | 0,0610869 | 0,0110311 | 0,000 |
| **High Density** | -0,3112903 | 0,023094 | 0,000 | **No education** | 0 | 0 | 0,000 |
| **0-5000 Inhabitants** | 0 | 0 | 0,000 | **Basic or lower education** | 0,0635422 | 0,0596553 | 0,287 |
| **5000-10000 Inhabitants** | 0,1850703 | 0,129775 | 0,154 | **Lower advanced education** | 0,3071179 | 0,0603378 | 0,000 |
| **10000-20000 Inhabitants** | 0,1638594 | 0,124687 | 0,189 | **MBO, HAVO, VWO** | 0,440591 | 0,059677 | 0,000 |
| **20000-50000 Inhabitants** | 0,1703676 | 0,1272133 | 0,180 | **College or University** | 0,2474949 | 0,061261 | 0,000 |
| **50000-100000 Inhabitants** | 0,1740051 | 0,1274262 | 0,172 | **Other education** | 0,22932 | 0,0722017 | 0,001 |
| **100000-150000 Inhabitants** | 0,0642572 | 0,1262426 | 0,611 | **Lowest Income Group** | 0 | 0 | 0,000 |
| **150000-250000 inhabitants** | 0,0230861 | 0,1285124 | 0,857 | **Income Group 2** | 0,2326699 | 0,0163283 | 0,000 |
| **More than 250000 Inhabitants** | -0,3572431 | 0,1355678 | 0,008 | **Income Group 3** | 0,3087416 | 0,188245 | 0,000 |
| **Groningen** | 0 | 0 | 0,000 | **Income Group 4** | 0,3322579 | 0,0201116 | 0,000 |
| **Friesland** | 0,0687096 | 0,0245253 | 0,005 | **Income Group 5** | 0,4011631 | 0,0195907 | 0,000 |
| **Drenthe** | 0,3229762 | 0,0826909 | 0,000 | **No working hours** | 0 | 0 | 0,000 |
| **Overijssel** | 0,2802602 | 0,0769858 | 0,000 | **Less than 12 working hours** | 0,083317 | 0,027108 | 0,002 |
| **Flevoland** | -0,137622 | 0,0739108 | 0,852 | **12 to 30 working hours** | 0,4150621 | 0,0154463 | 0,000 |
| **Gelderland** | 0,3498433 | 0,11447631 | 0,002 | **30 or more working hours** | 0,5563194 | 0,0137386 | 0,000 |
| **Utrecht** | 0,1202816 | 0,0447631 | 0,007 | **2005** | 0 | 0 | 0,000 |
| **Noord-Holland** | 0,1313505 | 0,0876022 | 0,134 | **2006** | -0,2264085 | 0,1044615 | 0,030 |
| **Zuid-Holland** | 0,3787585 | 0,0902172 | 0,000 | **2007** | -0,6726089 | 0,3075315 | 0,029 |
| **Zeeland** | -0,0475017 | 0,0550605 | 0,388 | **2008** | -1,275816 | 0,6000111 | 0,033 |
| **Noord-Brabant** | 0,4173092 | 0,577752 | 0,000 | **2009** | -0,115338 | 0,0258427 | 0,000 |
| **Limburg** | 0,4268073 | 0,0321708 | 0,000 | **2010** | -1,685713 | 0,7300139 | 0,021 |
| **Inkmafstr** | 1,049833 | 0,0067197 | 0,000 | **2011** | -3,166596 | 1,391533 | 0,023 |
| **Work** | 0 | 0 | 0,000 | **2012** | -4,351617 | 1,968199 | 0,027 |
| **shopping** | 0,8704774 | 0,0177217 | 0,000 | **2013** | -4,111809 | 1,856733 | 0,027 |
| **Services** | 1,059901 | 0,0267522 | 0,000 | **2014** | -3,769768 | 1,655633 | 0,023 |
| **Social or Recrational** | 0,0514837 | 0,0212921 | 0,016 | **2015** | -2,313409 | 0,9948865 | 0,020 |
| **Education** | -1,132622 | 0,0395201 | 0,000 | **2016** | -1,50769 | 0,6052325 | 0,013 |
| **Other** | 1,33372 | 0,0260124 | 0,000 | **2017** | -2,298261 | 0,9657082 | 0,017 |

*Table 11. Estimated Coefficients, Robust Standard Errors and significances in Model 2*

The log likelihood of Model 2 was found to be -155,085.77. The pseudo R-squared was calculated as 0.2938. 317,388 observations were used in the dataset.

### 4.2.1 The intercept and variable of interest

The intercept in model 2 is significant at confidence levels of 90% and 95% and has a negative value of 16.64218. In this case, the intercept cannot be interpreted by itself, as not all other variables can be 0. Therefore, the intercept on its own has no relevant real-life meaning.

*Millennial* is significant at all relevant confidence levels (90%,95% and 99%) and has a negative value. It's magnitude is 0.2623367, which means that the log-odds will be 0.2623367 lower for millennials when compared to non-millennials, ceteris paribus.

### 4.2.2 Land use variables

*Density* was included as a categorical variable with three possible values: Low density, Average density and High density. Low density was chosen as the reference category, and therefore no effect on *car use* is observed. It is still possible that *Low density* has an effect on *car use*, but as each observation should be either *Low density*, *Average density* or *High density*, the value for *Low density* is included in the value of the intercept. The effects of *Average density* and *High density* thus show an effect relative to *Low density.*

*Average density* has a value of -0.1840493. It was found significant at all relevant confidence levels. *Average density* has a negative effect on *car use* with a magnitude of 0.1840493. This means that going from *Low density* to *Average density* lowers the log-odds of *car use* with 0.1840493, ceteris paribus.

*High density* has a value of -0.3112903. It was found significant at all relevant confidence levels (90%, 95% and 99%). *High density* has a negative effect on *car use* with a magnitude of 0.3112903. This means that going from *Low density* to *High density* lowers the log-odds of *car use* with 0.3112903, ceteris paribus.

*Low density, Average density* and *High density* together form the variable *density*. Therefore, the joint significance of these dummies was also tested, using a Chi-squared significance test. This test yields a value of 189,57 which supplies a p-value of 0,000. Therefore, joint significance is supported for *density*.

The next variable that will be described is Municipalitty size(*gemgr)*. This variable was also presented as a categorical variable, and therefore, corresponding dummies were constructed. The reference category was chosen to be *Less than 5000 inhabitants*. This means that if none of the dummies is true, this person lives in a low-populated municipality. Only one dummy can be true, and the effect on *car use* is the relative effect compared to *Less than 5000 Inhabitants*.

The second category that was distinguished and thus the first dummy to be included in this model is *5.000 to 10.000 Inhabitants*. It has a magnitude of 0.1850703 and has a positive effect on *car use*. However, *5.000 to 10.000 Inhabitants* was found to be not significant at either the 90%, 95% and 99% confidence levels. Therefore, the coefficient does not significantly differ from 0.

*10.000 to 20.000 Inhabitants* has an estimated coefficient of 0.1638594. This means that the log-odds of *car use* would be 0.1638594 higher if someone lives in a municipality with 10.000 to 20.000 inhabitants, ceteris paribus. However, the effect was found to be not significant, having a p-value of 0.189. Thus, *10.000 to 20.000 Inhabitants* has no significant on *car use* at either the 90%, 95% and 99% confidence interval. Similar results were found for *20.000 to 50.000* and *50.000 to 100.000 Inhabitants* categories: the magnitudes are 0.1703676 and 0.1740051 respectively, which would imply a positive effect on the log-odds of *car use* c.p. However, with p-values of 0.180 and 0.172 respectively, the effects on *car use* are not significant at all relevant confidence levels. Smaller effects were found for the categories *100.000 to 150.000 Inhabitants* and *150.000 to 250.000 Inhabitants*. With magnitudes of 0.0642572 and 0.0230861 respectively, a small increase in the lo-odds of *car use* is expected, ceteris paribus. However, as with almost all categories of *gemgr*, both *100.000 to 150.000 Inhabitants* and *150.000 to 250.000 Inhabitants* are not significant on all relevant confidence levels. The only category that does have a significant effect is *More than 250.000 Inhabitants*. The coefficient is estimated with a magnitude of 0.3572431 and has a negative effect on *car use* c.p. With a p-vale of 0.008, it is significant at all relevant confidence levels.

As only one of the categories of *gemgr* is significant in Model 2, the joint significance was tested using the Chi-squared test. This test returned a p-value of 0.00 and a corresponding Chi-squared value of 183.76. This implies that there is no evidence to assume that *gemgr* does not have an effect on *car use*, and thus that the dummy variables have a jointly significant effect on *car use*.

The province of Groningen was used as the reference category for *prov.* This means that the effect of a certain province on *car use* is a relative increase or decrease to the effect that *Groningen* has on *car use*. All significant categories of *prov* have a positive effect on *car use* with magnitudes between 0.0687096 (*Friesland*) and 0.4268073 (*Limburg*) c.p. The estimated coefficients for *Flevoland, Noord-Holland* and *Zeeland* were all found to be insignificant at all relevant confidence levels. This means that these coefficients are statistically not different than zero. The Chi-squared test for *prov* returns a value of 749.29, with a corresponding p-value of 0.00. Thus the dummies that are part of *prov* are jointly significant and therefore applicable in the model.

The variable for trip distance, *lnkmafstr* has a magnitude of 1.049833. As *lnkmafstr* is defined as a logarithmic variable, a 1 point increase in *lnkmafstr* will increase the log-odds of *car use* by 1.049833 c.p.. For example, a trip length of 2 kilometer will increase the log-odds of *car use* by 1.049833*ln(2)= 0.727689 c.p.. *Lnkmafstr* was found to be significant on all relevant confidence levels.

### 4.2.3 Trip-related variables

Trip purpose, *doel6*, was also transformed into relevant dummy variables to assess the effect on *car use*. *Work* was chosen as the reference category for trip purpose. All other categories of trip purpose were found to have a positive and significant effect on *car use,* except for *education*. The effect of *education* is also significant at all relevant confidence intervals, but was found to have a negative effect on car use. The effect of *doel6* on *car use* thus vary between -1.132622 c.p for *education* and 1.333732 c.p for *other*. The Wald-test yields a value of 7736,92 (0.000), which also implies the dummies for *doel6* are jointly significant.

Model 2 estimates the coefficient of *weekend* at 0.1829892, which implies that the log-odds of *car use* are 0.1829892 higher during weekends, c.p. The p-value of *weekend* is 0.000 which implies *weekend* has a positive effect on *car use* which is significant at all relevant confidence levels.

The fuel price was found to have a positive effect on *car use*. The sign of this effect is positive, which means that an increase in *fuel price* increases the log-odds for *car use* by 10.16642 c.p. The estimated coefficient has a p-value of 0.034 and is thus significant at the 90% and 95% confidence levels. In comparison with the other described variables, this effect seems rather big. However, as *Fuel price* can only assume values between 1.352 and 1.80192, the total effect of this variable is 0,44992*10.16642, which equals 4.574076. This is the difference between the lowest and largest possible value for *brandstofprijs*.

### 4.2.4 Socio-demographic variables

Model 2 estimated a positive effect on *car use* for both *age* and *gender*. For *age,* this effect is found to be significant at the 90% and 95% confidence level. The effect is thus positive and has a magnitude of 0.001104, which means that ageing one year will increase the log-odds of *car use* by 0.001104 c.p. For *gender,* the magnitude of the effect is 0.0610869, which means the log-odds of *car use* increase by 0.0610869 if the respondent is female, c.p. This effect is also found to be significant, as it’s p-value is 0.000, which means it is significant at all relevant confidence levels.

Education, as proposed by Watkins (2016), was also found to influence *car use*. *No education* was chosen as the reference category. All higher levels of *education* were found to have a positive effect on *car use*. Respondents that completed MBO (practice-oriented studies), HAVO or VWO (the two highest levels at middle school), were found to have a higher chance of using a car ceteris paribus. Noteworthy is that college or university-schooled respondents also tend to make more use of a car, but the effect on *car use* is roughly 0.20 points lower. Having followed basic or lower education was found to have an insignificant positive effect.

*Income* was found to have an increasing positive effect on *car use*. The effect is 0.2326699 c.p for respondents in *Income Group 2*, while the effect of respondents in *Income Group 5* is estimated as 0.4011631 c.p. All *income* variables are significant at all relevant confidence levels. Related to income are the hours worked by a respondent. *No working hours* was chosen as the reference category, applying to a respondent that has no work at all. Working more hours has a positive effect on *car use*, therefore increasing the chance of car use for respondents that work more.

Last, the year of observation was included, with *2005* chosen as the reference year. Model 2 estimates varying negative coefficients for the other *year* categories. This means that, when compared to *2005*, the chance to use a car is lower for each other category in *year,* c.p. All estimated coefficients are significant at the 90% and 95% confidence level. The largest negative effect was found in *2012*, as the log-odds of *car use* are 4.351617 lower for respondents in this year, c.p.

### 4.2.5 Goodness of fit

As for Model 1, the PCP was calculated for Model 2. Again, the amount of correctly predicted values for *car use* were included, as well as the predicted values for non-car use. The predicted values of Model 2 are shown against the values for *car use* in Table 12.

| **Prediction\car use** | **No** | **Yes** | **Total** |
|---|---|---|---|
| **No** | 113.515 | 35.782 | 149297 |
| **Yes** | 37.463 | 130.628 | 168091 |
| **Total** | **150.978** | **166.410** | |

*Table 12. Predictions and observed car use in Model 2*

When compared to Model 1, Model 2 predicts less accurate for car use: 78% of *car use* was predicted correctly if *car use* was true. However, Model 2 returns more accurate predictions for non-car use. The PCP for Model 2 is summarized in Table xx.

| | |
|---|---|
| Car use correctly predicted | 78% |
| Non-car use | 75% |
| Total correctly predicted | 77% |

*Table 13. PCP values for Model 2*

When compared to the PCP of Model 1, Model 2 predicts 8% less accurate on car use, and 54% more accurate on non-car use. In total, Model 2 thus predicted 77% of the observations correctly, against 55% of Model 1.

## 4.3 Year-specific models

This section will discuss Model 2a and Model 2b, which were formulated to compare millennials in 2016 with their same-aged equivalents in 2006.

### 4.3.1 Model 2a: 2006

Compared to Model 2, a few adjustments were made to formulate Model 2a. First, *millennial* was replaced by *milage2006*. Second, as only 2006 was used, *year* was removed from the model. Last, as *fuel price* in 2006 is solely dependent on *prov*, it was omitted from Model 2a. The estimated coefficients, standard errors and significance were summarized in Table 14.

| Coefficient | Value | Standard Error | Significance | Coefficient | Value | Standard Error | Significance |
|---|---|---|---|---|---|---|---|
| Intercept | -3,576926 | 0,4052561 | 0,000 | lnkmafstr | 1,121861 | 0,0153469 | 0,000 |
| Milage2006 | 0,2125769 | 0,0561105 | 0,000 | Work | 0 | 0 | 0,000 |
| Low Density | 0,000 | 0,000 | 0,000 | shopping | 0,9537412 | 0,0523567 | 0,000 |
| Average Density | -0,2005977 | 0,0518327 | 0,000 | Services | 1,060747 | 0,0908414 | 0,000 |
| High Density | -0,396713 | 0,0688427 | 0,000 | Social or Recrational | 0,1851969 | 0,0470045 | 0,000 |
| 0-5000 Inhabitants | 0 | 0 | 0,000 | Education | -1,162925 | 0,1350482 | 0,000 |
| 5000-10000 Inhabitants | 0,3021838 | 0,3706443 | 0,415 | Other | 1,359153 | 0,0741954 | 0,000 |
| 10000-20000 Inhabitants | 0,2536392 | 0,3603785 | 0,482 | Weekend | 0,1291449 | 0,0399866 | 0,001 |
| 20000-50000 Inhabitants | 0,3219002 | 0,3601427 | 0,371 | Age | 0,007889 | 0,0020094 | 0,000 |
| 50000-100000 Inhabitants | 0,3298999 | 0,3633979 | 0,364 | Gender | 0,0770888 | 0,037596 | 0,040 |
| 100000-150000 Inhabitants | 0,195306 | 0,3658251 | 0,593 | No education | 0 | 0 | 0,000 |
| 150000-250000 inhabitants | 0,336646 | 0,3717777 | 0,365 | Basic or lower education | 0 | 0 | 0,000 |
| More than 250000 Inhabitants | 0,0742148 | 0,3669274 | 0,840 | Lower advanced education | 0,2994477 | 0,0679536 | 0,000 |
| Groningen | 0 | 0 | 0,000 | MBO, HAVO, VWO | 0,3655417 | 0,0701342 | 0,000 |
| Friesland | 0,1106973 | 0,1223824 | 0,366 | College or University | 0,008719 | 0,0745796 | 0,907 |
| Drenthe | 0,3260479 | 0,1342431 | 0,015 | Other education | 0,2094537 | 0,0722017 | 0,383 |
| Overijssel | 0,1712713 | 0,1103042 | 0,120 | Lowest Income Group | 0 | 0 | 0,000 |
| Flevoland | -0,0544686 | 0,1514372 | 0,719 | Income Group 2 | 0,3424354 | 0,0639855 | 0,000 |
| Gelderland | 0,0633594 | 0,1019863 | 0,534 | Income Group 3 | 0,4618648 | 0,0687236 | 0,000 |
| Utrecht | 0,0886365 | 0,1113148 | 0,426 | Income Group 4 | 0,5509086 | 0,0750772 | 0,000 |
| Noord-Holland | 0,0449035 | 0,1045176 | 0,667 | Income Group 5 | 0,7249838 | 0,0809944 | 0,000 |
| Zuid-Holland | 0,2749915 | 0,1035225 | 0,008 | No working hours | 0 | 0 | 0,000 |
| Zeeland | 0,1239822 | 0,1377921 | 0,368 | Less than 12 working hours | 0,2968837 | 0,085196 | 0,000 |
| Noord-Brabant | 0,3782292 | 0,0997483 | 0,000 | 12 to 30 working hours | 0,5611737 | 0,0611436 | 0,000 |
| Limburg | 0,6176821 | 0,1104775 | 0,000 | 30 or more working hours | 0,6056469 | 0,0608914 | 0,000 |

*Table 14. Estimated coefficients, Standard Errors and Significance for Model 2a.*

Model 2a was found to have a log likelihood of -11875.728. The pseudo R-squared has a value of 0.3131 and 25,165 observations were used to construct Model 2a. As mentioned, only observations in 2006 were used to construct Model 2a.

#### *4.3.1.2 Variable of Interest and the Intercept*

As with Model 2, the *intercept* of Model 2a is significant at all relevant confidence levels. It can not be interpreted on its own, as it is impossible for all other variables to be zero. Thus, the intercept has no meaning on itself. As for *Milage2006*, the predicted coefficient is 0.2125769, which means the log-odds of *car use* are 0.2125769 higher for someone who had the "millennial age"(ages between 7 and 34) in 2006, ceteris paribus. This means a higher chance of car use is expected for an individual that had the "millennial age" in 2006, keeping all other variables constant.

#### *4.3.1.3 Land use variables*

The estimated coefficients of *density* are significant at all relevant confidence levels and have a negative effect on *car use*. As expected, the negative value for *High density* is higher than that of *Average density*. The estimated coefficients for *gemgr* were all found to be insignificant at all relevant confidence intervals. This means that there is no significant difference in the log-odds of *car use* between small municipalities and larger municipalities. A Wald-test strengthened this finding, as the Chi-squared value for all categories of gemgr is 41.73 which equals a significance level of 0.000. This means that, despite individually being insignificant, joint significance of *gemgr* on *car use* is observed. As for *prov, Zuid-Holland, Noord-Brabant* and *Limburg* were found to have a positive and significant effect on *car use* at all relevant confidence levels. The estimated coefficient for *Drenthe* was found to be significant, but only at the 90% and 95% confidence levels. A Wald-test for all categories of *prov* returned a Chi-squared value of 109.33 and a significance of 0.000, which means that the categories of *prov* are jointly significant when predicting *car use*.

#### *4.3.1.4 Trip-related variables*

As in Model 2, the effect of trip distance on *car use* was also estimated to be positive in Model 2a. This effect, as in Model 2, was found to be significant at all relevant confidence intervals. As for trip purpose, it was found that only *education* has a negative effect on *car use*, compared to *work*. All other variables were found to have a positive effect on *car use*. All categories of *doel6* were found to be significant at all relevant confidence levels. Also, it was found that there is a higher probability of car use if a trip was undertaken in the *weekend*.

#### *4.3.1.5 Socio-demographic variables*

Model 2a does not include *Basic or lower education*, as no respondent in this year was observed to have consumed this level of education as its highest. Also, *College or University* and *Other education* were found to have no significant effect on *car use*. *Education* was found to be significant, as the Wald-test showed joint significance. Further, all estimated coefficients for *Age, Income* and *working hours* are positive and significant at all relevant confidence levels.

### 4.3.2 Model 2b: 2016

Model 2b was defined similar to Model 2a, with the only difference being that *milage2006* replaced by *millennial* again. *Year* was excluded and *fuel price* was omitted for the same reasons as stated under 4.3.1. Model 2a was fitted only on observations in *2016*. The estimated coefficients, Standard Errors and significance are summarized in Table 15.

| Coefficient | Value | Standard Error | Significance |
|---|---|---|---|
| Intercept | -3,509491 | 0,7704658 | 0,000 |
| Millennial | -0,0694134 | 0,063757 | 0,276 |
| Low Density | 0,000 | 0,000 | 0,000 |
| Average Density | -0,148 | 0,530 | 0,005 |
| High Density | -0,2486704 | 0,0603012 | 0,000 |
| 0-5000 Inhabitants | 0 | 0 | 0,000 |
| 5000-10000 Inhabitants | 0,672015 | 0,7469175 | 0,368 |
| 10000-20000 Inhabitants | 0,4288459 | 0,728241 | 0,556 |
| 20000-50000 Inhabitants | 0,5238808 | 0,7268046 | 0,471 |
| 50000-100000 Inhabitants | 0,5342701 | 0,727508 | 0,463 |
| 100000-150000 Inhabitants | 0,4255961 | 0,7287395 | 0,559 |
| 150000-250000 inhabitants | 0,3496422 | 0,729508 | 0,632 |
| More than 250000 Inhabitants | -0,1013753 | 0,729391 | 0,889 |
| Groningen | 0 | 0 | 0,000 |
| Friesland | 0,2191929 | 0,1291467 | 0,090 |
| Drenthe | 0,2138245 | 0,1378539 | 0,121 |
| Overijssel | 0,1420274 | 0,1103042 | 0,220 |
| Flevoland | 0,2723079 | 0,1390275 | 0,050 |
| Gelderland | 0,1244555 | 0,1073417 | 0,246 |
| Utrecht | 0,0534295 | 0,1167019 | 0,426 |
| Noord-Holland | -0,0391223 | 0,1055985 | 0,711 |
| Zuid-Holland | 0,2447246 | 0,10721 | 0,022 |
| Zeeland | 0,0855529 | 0,1485526 | 0,565 |
| Noord-Brabant | 0,3022405 | 0,1046919 | 0,004 |
| Limburg | 0,5953282 | 0,116245 | 0,000 |

| Coefficient | Value | Standard Error | Significance |
|---|---|---|---|
| lnkmafstr | 1,015651 | 0,0153914 | 0,000 |
| Work | 0 | 0 | 0,000 |
| shopping | 0,9784674 | 0,0547236 | 0,000 |
| Services | 1,251125 | 0,0945574 | 0,000 |
| Social or Recrational | 0,0598836 | 0,0473614 | 0,206 |
| Education | -0,9526449 | 0,1015871 | 0,000 |
| Other | 1,577776 | 0,0831809 | 0,000 |
| Weekend | 0,204282 | 0,0394383 | 0,000 |
| Age | 0,0032935 | 0,0016985 | 0,052 |
| Gender | 0,1110392 | 0,0357366 | 0,002 |
| No education | 0 | 0 | 0,000 |
| Basic or lower education | -0,0776463 | 0,2052281 | 0,705 |
| Lower advanced education | 0,1345493 | 0,1962453 | 0,493 |
| MBO, HAVO, VWO | 0,2985743 | 0,195071 | 0,126 |
| College or University | 0,1297875 | 0,1958801 | 0,508 |
| Other education | 0,0349888 | 0,2435635 | 0,886 |
| Lowest Income Group | 0 | 0 | 0,000 |
| Income Group 2 | 0,307369 | 0,0679993 | 0,000 |
| Income Group 3 | 0,4606086 | 0,0675476 | 0,000 |
| Income Group 4 | 0,4653327 | 0,0680256 | 0,000 |
| Income Group 5 | 0,4245874 | 0,0634063 | 0,000 |
| No working hours | 0 | 0 | 0,000 |
| Less than 12 working hours | -0,021768 | 0,0955752 | 0,820 |
| 12 to 30 working hours | 0,4374641 | 0,0545822 | 0,000 |
| 30 or more working hours | 0,6027417 | 0,0479485 | 0,000 |

*Table 15. Estimated coefficients, standard errors and Significance for Model 2b*

The log likelihood of Model 2b is -11,317.995. It has a pseudo R-squared value of 0.2840. 22,806 observations were included in Model 2b, as only observations in 2016 were used.

As the same variables were included in Model 2a and Model 2b, a comparison of the models will be used to asses Hypothesis 2.

#### 4.3.2.2 Variables of interest and the Intercept

As for Model 2 and Model 2a, the intercept has a significant negative value. However, as under the earlier defined models, it has no inherent meaning, as not all other variables can be zero at the same time. Model 2b estimates the coefficient of *millennial* as -0.0694134, which means the being millennial would have a negative effect of 0.0694134 on the log-odds of *car use*, c.p. However, *millennial* is not significant in Model 2b, as it is not significant at the 90%, 95% and 99% confidence levels. This implies that the coefficient of *millennial* is not statistically different from 0.

#### 4.3.2.3 Land-use variables

The estimated coefficients for the *Density* categories have a negative sign and were found to be significant at the 90% and 95% confidence level for *Average density* and for all relevant confidence levels for *High density*. When compared to Model 2a, it is visible that the magnitudes of the coefficients are smaller for Model 2b than for Model 2a. Almost all categories of *gemgr* were found to have a positive effect on *car use*, except for the largest municipality size. However, all predicted coefficients were found to be not significant at all relevant confidence levels. Joint significance was however assumed, as a Wald-test on all categories of *gemgr* did not supply evidence that all coefficients equal zero.

#### 4.3.2.4 Trip information

The estimated coefficients of trip distance as well as *weekend* were both estimated to be positive. This implies that the log-odds of *car use* increase if the trip distance increases, or if the trip is undertaken in the weekend. Both coefficients are significant at all relevant confidence levels. *Social or recreational trips* are not statistically different from *work trips*, as the coefficient of *Social or Recreational trips* is not significant at all relevant confidence levels.

#### 4.3.2.5 Socio-Demographic variables

The most noteworthy for the socio-demographic variables is that all categories of *education* were found to be not significant when predicting *car use*. However, a Wald-test proved joint significance for the categories of *education*. The effect of *income* was estimated to be significant for all categories at all relevant confidence levels. The effect of the different categories of *income* is noteworthy, as the biggest magnitude is estimated at *Income Group 4*, where it is expected that the highest income group has the biggest effect on *car use.*

### 4.3.3 Comparing Model 2a and Model 2b

Both Model 2a and Model 2b thus returned different values for the estimated coefficients. This sections summarizes both models and shows the differences between the models.

Table 16 shows the coefficients and significance values for both models. The most notable difference can be found in the sign and significance of *milage2006* in Model 2a, when compared to *millennial* in Model 2b. As shown in Table 16., the coefficient of *milage2006* was estimated to be positive and significant at all relevant confidence intervals. Its 2016-counterpart, *millennial* in Model 2b, however, showed a negative sign. Although being estimated to be negative, it was also found to be insignificant at all relevant confidence levels. *Education* was also found to be significant in Model 2a, while estimated insignificant for all categories in Model 2b.

The goodness of fit of both models was calculated using the PCP. Appendix B Table 1 shows predictions of Model 2a when compared to the actual observations, while Appendix B Table 2 shows the PCP for Model 2a. Appendix B Table 3 and Appendix B Table 4 show the comparison between predictions and actual observations, and PCP respectively for Model 2b. These tables can be found in Appendix B.

| Coefficient | Value M2a | Significance M2a | Value M2b | Significance M2b |
|---|---|---|---|---|
| Intercept | -3,576926 | 0,000 | -3,509491 | 0,000 |
| Milage2006 | 0,2125769 | 0,000 | - | - |
| Millennial | - | - | -0,0694134 | 0,276 |
| Low Density | 0,000 | 0,000 | 0,000 | 0,000 |
| Average Density | -0,2005977 | 0,000 | -0,148 | 0,005 |
| High Density | -0,396713 | 0,000 | -0,2486704 | 0,000 |
| 0-5000 Inhabitants | 0 | 0,000 | 0 | 0,000 |
| 5000-10000 Inhabitants | 0,3021838 | 0,415 | 0,672015 | 0,368 |
| 10000-20000 Inhabitants | 0,2536392 | 0,482 | 0,4288459 | 0,556 |
| 20000-50000 Inhabitants | 0,3219002 | 0,371 | 0,5238808 | 0,471 |
| 50000-100000 Inhabitants | 0,3298999 | 0,364 | 0,5342701 | 0,463 |
| 100000-150000 Inhabitants | 0,195306 | 0,593 | 0,4255961 | 0,559 |
| 150000-250000 inhabitants | 0,336646 | 0,365 | 0,3496422 | 0,632 |
| More than 250000 Inhabitants | 0,0742148 | 0,840 | -0,1013753 | 0,889 |
| Groningen | 0 | 0,000 | 0 | 0,000 |
| Friesland | 0,1106973 | 0,366 | 0,2191929 | 0,090 |
| Drenthe | 0,3260479 | 0,015 | 0,2138245 | 0,121 |
| Overijssel | 0,1712713 | 0,120 | 0,1420274 | 0,220 |
| Flevoland | -0,0544686 | 0,719 | 0,2723079 | 0,050 |
| Gelderland | 0,0633594 | 0,534 | 0,1244555 | 0,246 |
| Utrecht | 0,0886365 | 0,426 | 0,0534295 | 0,426 |
| Noord-Holland | 0,0449035 | 0,667 | -0,0391223 | 0,711 |
| Zuid-Holland | 0,2749915 | 0,008 | 0,2447246 | 0,022 |
| Zeeland | 0,1239822 | 0,368 | 0,0855529 | 0,565 |
| Noord-Brabant | 0,3782292 | 0,000 | 0,3022405 | 0,004 |
| Limburg | 0,6176821 | 0,000 | 0,5953282 | 0,000 |
| Inkmafstr | 1,121861 | 0,000 | 1,015651 | 0,000 |
| Work | 0 | 0,000 | 0 | 0,000 |
| shopping | 0,9537412 | 0,000 | 0,9784674 | 0,000 |
| Services | 1,060747 | 0,000 | 1,251125 | 0,000 |
| Social or Recrational | 0,1851969 | 0,000 | 0,0598836 | 0,206 |
| Education | -1,162925 | 0,000 | -0,9526449 | 0,000 |
| Other | 1,359153 | 0,000 | 1,577776 | 0,000 |
| Weekend | 0,1291449 | 0,001 | 0,204282 | 0,000 |
| Age | 0,007889 | 0,000 | 0,0032935 | 0,052 |
| Gender | 0,0770888 | 0,040 | 0,1110392 | 0,002 |
| No education | 0 | 0,000 | 0 | 0,000 |
| Basic or lower education | 0 | 0,000 | -0,0776463 | 0,705 |
| Lower advanced education | 0,2994477 | 0,000 | 0,1345493 | 0,493 |
| MBO, HAVO, VWO | 0,3655417 | 0,000 | 0,2985743 | 0,126 |
| College or University | 0,008719 | 0,000 | 0,1297875 | 0,508 |
| Other education | 0,2094537 | 0,001 | 0,0349888 | 0,886 |
| Lowest Income Group | 0 | 0,000 | 0 | 0,000 |
| Income Group 2 | 0,3424354 | 0,000 | 0,307369 | 0,000 |
| Income Group 3 | 0,4618648 | 0,000 | 0,4606086 | 0,000 |
| Income Group 4 | 0,5509086 | 0,000 | 0,4653327 | 0,000 |
| Income Group 5 | 0,7249838 | 0,000 | 0,4245874 | 0,000 |
| No working hours | 0 | 0,000 | 0 | 0,000 |
| Less than 12 working hours | 0,2968837 | 0,000 | -0,021768 | 0,820 |
| 12 to 30 working hours | 0,5611737 | 0,000 | 0,4374641 | 0,000 |
| 30 or more working hours | 0,6056469 | 0,000 | 0,6027417 | 0,000 |

*Table 16. Comparison of Model 2a and Model 2b*

## 4.4 Model 3

Model 3 expands on Model 2, adding the interaction effect of *millennial* and *density*, to isolate the RSS from *millennial*. All other variables included in Model 2 were included in Model 3, and cluster correction was again applied to *brandstofprijs*. This leads to the coefficients, Robust Standard Errors and Significances shown in Table 17.

| Coefficient | Value | Robust Standard Error | Significance | Coefficient | Value | Robust Standard Error | Significance |
|---|---|---|---|---|---|---|---|
| **Intercept** | -16,32741 | 6,529882 | 0,012 | Weekend | 0,1828303 | 0,0138108 | 0,000 |
| **Millennial** | -0,2164384 | 0,0333372 | 0,000 | Fuel price | 9,931167 | 4,763065 | 0,037 |
| **Millenial and Average density** | 0,0026095 | 0,0322544 | 0,936 | Age | 0,0011612 | 0,0005673 | 0,041 |
| **Millenial and High density** | -0,0973445 | 0,0330055 | 0,003 | Gender | 0,0611177 | 0,0110524 | 0,000 |
| **Low Density** | 0,000 | 0,000 | 0,000 | No education | 0,000 | 0,000 | 0,000 |
| **Average Density** | -0,1840053 | 0,0171846 | 0,000 | Basic or lower education | 0,0643478 | 0,0596586 | 0,281 |
| **High Density** | -0,2961819 | 0,0237844 | 0,000 | Lower advanced education | 0,3089951 | 0,0603817 | 0,000 |
| **0-5000 Inhabitants** | 0 | 0 | 0,000 | MBO, HAVO, VWO | 0,4428963 | 0,0597154 | 0,000 |
| **5000-10000 Inhabitants** | 0,1846128 | 0,1298317 | 0,155 | College or University | 0,2504265 | 0,601129 | 0,000 |
| **10000-20000 Inhabitants** | 0,1629412 | 0,1248074 | 0,192 | Other education | 0,230877 | 0,0721589 | 0,001 |
| **20000-50000 Inhabitants** | 0,1691595 | 0,1272831 | 0,184 | Lowest Income Group | 0 | 0 | 0,000 |
| **50000-100000 Inhabitants** | 0,1740051 | 0,1274262 | 0,172 | Income Group 2 | 0,2308596 | 0,0163566 | 0,000 |
| **100000-150000 Inhabitants** | 0,0629266 | 0,1263127 | 0,618 | Income Group 3 | 0,6062507 | 0,0188372 | 0,000 |
| **150000-250000 inhabitants** | 0,023582 | 0,128621 | 0,855 | Income Group 4 | 0,3293883 | 0,02109 | 0,000 |
| **More than 250000 Inhabitants** | -0,357119 | 0,1356762 | 0,008 | Income Group 5 | 0,3974176 | 0,0195468 | 0,000 |
| **Groningen** | 0 | 0 | 0,000 | No working hours | 0 | 0 | 0,000 |
| **Friesland** | 0,0673094 | 0,0244686 | 0,006 | Less than 12 working hours | 0,0826848 | 0,0271135 | 0,002 |
| **Drenthe** | 0,3191772 | 0,0822725 | 0,000 | 12 to 30 working hours | 0,4160671 | 0,0155323 | 0,000 |
| **Overijssel** | 0,2754339 | 0,0766672 | 0,000 | 30 or more working hours | 0,5570966 | 0,0137691 | 0,000 |
| **Flevoland** | -0,0120801 | 0,0735524 | 0,870 | 2005 | 0 | 0 | 0,000 |
| **Gelderland** | 0,3431535 | 0,1139574 | 0,003 | 2006 | -0,2216603 | 0,104016 | 0,033 |
| **Utrecht** | 0,1184617 | 0,0445973 | 0,008 | 2007 | -0,6577977 | 0,3059739 | 0,032 |
| **Noord-Holland** | 0,1265015 | 0,0872023 | 0,147 | 2008 | -1,246208 | 0,5968817 | 0,037 |
| **Zuid-Holland** | 0,3736024 | 0,0899298 | 0,000 | 2009 | -0,1148701 | 0,0256135 | 0,000 |
| **Zeeland** | -0,0464609 | 0,0547534 | 0,396 | 2010 | -1,649745 | 0,7262838 | 0,023 |
| **Noord-Brabant** | 0,4135628 | 0,0575131 | 0,000 | 2011 | -3,09803 | 1,984471 | 0,025 |
| **Limburg** | 0,4261627 | 0,0320997 | 0,000 | 2012 | -4,254948 | 1,9582 | 0,030 |
| **Inkmafstr** | 1,049731 | 0,0067178 | 0,000 | 2013 | -4020725 | 1,84738 | 0,030 |
| **Work** | 0 | 0 | 0,000 | 2014 | -3,688119 | 1,647364 | 0,025 |
| **shopping** | 0,8706427 | 0,0177287 | 0,000 | 2015 | -2,263929 | 0,9897508 | 0,022 |
| **Services** | 1,059668 | 0,026728 | 0,000 | 2016 | -1,476867 | 0,6021085 | 0,014 |
| **Social or Recrational** | 0,051957 | 0,0213016 | 0,015 | 2017 | -2,2499 | 0,9608036 | 0,019 |
| **Education** | -1,139349 | 0,0395909 | 0,000 | | | | |
| **Other** | 1,333955 | 0,0260332 | 0,000 | | | | |

*Table 17. The Estimated coefficients, Robust Standard Errors and Significances of Model 3*

For Model 3, the log pseudolikelihood was given as -155,077.67. The pseudo R-squared of Model 3 is 0.2939 and 317,388 observations were included to construct Model 3.

### 4.4.1 Intercept and variables of interest

The *intercept* in Model 3 was estimated with a value of -16,32741 and a significance of 0.012. This means that the *intercept* is significant at the 90% and 95% confidence levels. As with Model 2, the intercept has no inherent meaning, as it is not possible that all other variables equal zero. This means that the intercept has no meaning of its own, but can be used in the model.

The estimated coefficient of *millennial* has a negative sign and a magnitude of -0.2164384. It is significant at all relevant confidence levels. This implies that being millennial will decrease the log-odds of *car use* by 0.2164384 ceteris paribus. The coefficients for the categories of *Density* were also estimated to be negative and significant at all relevant confidence levels. The magnitude of *High density*, as expected, exceeds the magnitude of *Average density*.

Model 3 also includes the interaction term of *Density* and *Millennial*. *Millennial and Low density* was used as the reference category, which means there is no extra effect if an individual is a millennial and lives in a low density neighborhood. The interaction term *Millennial and Average density* was estimated to be positive, thus decreasing the magnitude of the effect of *Millennial* and *Density*. However, with a value of 0.890, the coefficient is not significant at all relevant confidence levels, which implies the coefficient is not statistically different from zero. The interaction term *Millennial and High Density* is estimated to be negative. Also, it is significant at all relevant confidence levels.

This means that if an individual is a millennial and lives in a high density neighborhood, an additional effect of -0.0973445 on *car use* is found, ceteris paribus.

### 4.4.2 Land use variables

As *density* is a variable of interest in Model 3, it was already interpreted under section 4.4.1. Municipality size was found to have no significant effect on *car use,* except for the largest municipalities. For *250.000 or more Inhabitants*, a negative effect was estimated in Model 3 with a Significance of 0.008. This means the effect is significant at the 90% and 95% confidence levels. The magnitude is -0.357719, which means the log-odds of *car use* are 0.357719 lower for someone that lives in the largest municipalities, ceteris paribus. For all other categories, a positive coefficient was estimated. However, as already mentioned, all of these coefficients were found to be not significant at all relevant confidence levels. A Wald-test proved that all categories of *gemgr* are jointly significant when predicting *car use*.

The provinces of *prov* were all found to be significant at all relevant confidence levels, except for the coefficients of *Noord-Holland, Zeeland* and *Flevoland*. All significant coefficients are estimated to be positive, with means the log-odds of *car use* increase if an individual lives outside of *Groningen*.

### 4.4.3 Trip-related variables

Trip distance, trip purpose and trips in the weekend were all estimated to have a positive effect on the log-odds on *car use.* All coefficients of *lnkmafstr, weekend* and *doel6* were found to be significant at all relevant confidence levels, except for trip purpose *Social or Recreational*, which was found to be significant at the 90% and 95% confidence interval. Furthermore, of all trip-related variables, only trip purpose *education* was found to have a negative effect on the log-odds of *car use.*

### 4.4.4 Socio-demographic variables

Both *age and gender* were found to have a significant and positive effect on the log-odds of *car use*. The estimated effect of *age* is significant at the 90% and 95% confidence level, while gender is significant at all relevant confidence levels. This means that ageing increases the log-odds of *car use*, and the log-odds of *car use* are higher for women than they are for men.

The level of education of individuals was also found to have a significant effect on the log-odds of *car use*, except fot he *Basic or Lower education* category. The magnitudes of the effects vary between 0.0643478 for *Basic or Lower education* to 0.4428963 for *MBO, HAVO or VWO*.

For *income*, all categories were estimated to have a positive effect on the log-odds of *car use*. All coefficients are significant at the 90%, 95% and 99% confidence levels. Magnitudes vary between 0.2308596 for *Income Group 2* and 0.6062507 For *Income Group 3.* Noteworthy is that the coefficients for *Income Group 4* and *Income Group 5* were estimated between those values, while it was expected that the highest *Income Group* has the largest effect on the log-odds of *car use*.

The amount worked by an individual was also found to have a positive effect on *car use*, as more working hours increases the log-odds of *car use*. A relatively small increase of 0.0826848 ceteris paribus is expected for an individual that works less than 12 hours, while an increase of 0.5570966 ceteris paribus is expected for an individual that works more than 30 hours a week. All estimated coefficients of *Working hours* are significant at all relevant confidence levels.

Last, *year* was found to have a significant effect on *car use* on the 90% and 95% confidence levels, except for *2009*, which is significant at all relevant confidence intervals. All coefficients are estimated to be negative, which means the log-odds are lower for each year compared to *2005*.

### 4.4.5 Goodness of fit

As for Model 1 and Model 2, the PCP was calculated as a measure of Goodness of fit for Model 3. The predicted values are given and compared to the actual value in Table 18.

| Prediction\car use | No | Yes | Total |
|---|---|---|---|
| **No** | 113.453 | 35.773 | 149226 |
| **Yes** | 37.525 | 130.937 | 168462 |
| **Total** | **150.978** | **166.710** | |

*Table 18. Predictions and observations of car use in Model 3*

The table shows that out of 149,226 predictions for non-car use, 113,453 of them were correct. A total of 150,978 non-car users were found in the data, implying that 37,525 observations were wrongly predicted to be car users. This amounts to a PCP of non-car use of 75,15%. The percentage for correctly predicted car uses was calculated the same way and given in Table 19.

| | |
|---|---|
| Car use correctly predicted | 78,54% |
| Non-car use | 75,15% |
| Total correctly predicted | 76,93% |

*Table 19. PCP values for Model 3*

A total of 76,93% of all observations was thus calculated correctly using Model 3. This is an increase of 0,01% when compared to Model 2.

The former sections gave an overview of Model 1, Model 2, Model 2a, Model 2b and Model 3. These models were formulated as the required models to find an answer on the formulated hypotheses. The next section will briefly discuss the proposed extensions as discusses in section 3.4.5.

## 4.5 Extensive Models

The results of the extensive models will be given in this section. 4.5.1 will deal with the division of *millennial* into *early millennial* and *late millennial*. Section 4.5.2 will describe the results of the model where *car driver* is used as the dependent variable instead of *car use*.

### 4.5.3 Model 4: Millennial redefined

Model 4 extends on Model 3, including a second *millennial* category. This division was based on the notion that *millennial* now includes all individuals who are born between 1982 and 2009. The new variable *early millennial* consists of all individuals in the dataset born between 1982 and 1995, while *late millennial* consists of all individuals in the dataset born between 1995 and 2009. The interaction effect of *millennial* and *density* was thus replaced by an interaction effect between *early millennial* and *density*, as well as *late millennial* and *density.* The estimated model with its estimated coefficients, Robust Standard Errors and Significance are found in Table 20 below.

| Coefficient | Value | Robust Standard Error | Significance | Coefficient | Value | Robus Standard Error | Significance |
|---|---|---|---|---|---|---|---|
| **Intercept** | -16,6516 | 6,493673 | 0,010 | **Weekend** | 0,187165 | 0,0138402 | 0,000 |
| **Early Millennial** | -0,1336373 | 0,0371811 | 0,000 | **Fuel price** | 10,30559 | 4,73675 | 0,030 |
| **Late Millennial** | -1,111958 | 0,0756338 | 0,000 | **Age** | -0,00131 | 0,000529 | 0,013 |
| **Early Millennial and Average Density** | 0,0002568 | 0,0365104 | 0,994 | **Gender** | 0,052582 | 0,0107113 | 0,000 |
| **Early Millennial and High Density** | -0,1650151 | 0,037531 | 0,000 | **No education** | 0 | 0 | 0,000 |
| **Late Millenial and Average density** | -0,385542 | 0,0884307 | 0,663 | **Basic or lower education** | 0,068686 | 0,0599123 | 0,252 |
| **Late Millenial and High density** | 0,1265382 | 0,0835799 | 0,130 | **Lower advanced education** | 0,264459 | 0,0615626 | 0,000 |
| **Low Density** | 0,000 | 0,000 | 0,000 | **MBO, HAVO, VWO** | 0,372716 | 0,0611212 | 0,000 |
| **Average Density** | -0,1829992 | 0,017194 | 0,000 | **College or University** | 0,167809 | 0,0616214 | 0,006 |
| **High Density** | -0,2955056 | 0,2372 | 0,000 | **Other education** | 0,165209 | 0,0735563 | 0,025 |
| **0-5000 Inhabitants** | 0 | 0 | 0,000 | **Lowest Income Group** | 0 | 0 | 0,000 |
| **5000-10000 Inhabitants** | 0,1827867 | 0,1320694 | 0,166 | **Income Group 2** | 0,261224 | 0,0164664 | 0,000 |
| **10000-20000 Inhabitants** | 0,1627232 | 0,1273776 | 0,201 | **Income Group 3** | 0,347891 | 0,0197049 | 0,000 |
| **20000-50000 Inhabitants** | 0,1706322 | 0,1298002 | 0,189 | **Income Group 4** | 0,383077 | 0,0212857 | 0,000 |
| **50000-100000 Inhabitants** | 0,1746031 | 0,1287542 | 0,179 | **Income Group 5** | 0,468129 | 0,0214344 | 0,000 |
| **100000-150000 Inhabitants** | 0,0650124 | 0,1287542 | 0,614 | **No working hours** | 0 | 0 | 0,000 |
| **150000-250000 inhabitants** | 0,0252825 | 0,1310271 | 0,847 | **Less than 12 working hours** | 0,071993 | 0,0270851 | 0,008 |
| **More than 250000 Inhabitants** | -0,3579992 | 0,1381082 | 0,010 | **12 to 30 working hours** | 0,361197 | 0,0152832 | 0,000 |
| **Groningen** | 0 | 0 | 0,000 | **30 or more working hours** | 0,482162 | 0,0138845 | 0,000 |
| **Friesland** | 0,0690314 | 0,0245283 | 0,005 | **2005** | 0 | 0 | 0,000 |
| **Drenthe** | 0,3271453 | 0,0818577 | 0,000 | **2006** | -0,22961 | 0,1035009 | 0,027 |
| **Overijssel** | 0,2791452 | 0,0761739 | 0,000 | **2007** | -0,6805 | 0,3044266 | 0,025 |
| **Flevoland** | -0,0204679 | 0,0728023 | 0,779 | **2008** | -1,29162 | 0,593786 | 0,030 |
| **Gelderland** | 0,3509833 | 0,1134226 | 0,002 | **2009** | -0,11315 | 0,0256481 | 0,000 |
| **Utrecht** | 0,1209858 | 0,0443005 | 0,006 | **2010** | -1,72003 | 0,722537 | 0,017 |
| **Noord-Holland** | 0,1325379 | 0,0870384 | 0,128 | **2011** | -3,20719 | 1,376967 | 0,020 |
| **Zuid-Holland** | 0,3787976 | 0,0894803 | 0,000 | **2012** | -4,39666 | 1,947238 | 0,024 |
| **Zeeland** | -0,0493693 | 0,0544834 | 0,365 | **2013** | -4,14065 | 1,837213 | 0,024 |
| **Noord-Brabant** | 0,4153358 | 0,0571938 | 0,000 | **2014** | -3,78353 | 1,638045 | 0,019 |
| **Limburg** | 0,4236607 | 0,0315459 | 0,000 | **2015** | -2,30008 | 0,9843038 | 0,019 |
| **lnkmafstr** | 1,049874 | 0,0066865 | 0,000 | **2016** | -1,47539 | 0,5987737 | 0,014 |
| **Work** | 0 | 0 | 0,000 | **2017** | -2,26878 | 0,9555314 | 0,018 |
| **shopping** | 0,8630825 | 0,0176115 | 0,000 | | | | |
| **Services** | 1,05645 | 0,026807 | 0,000 | | | | |
| **Social or Recrational** | 0,0500432 | 0,0213196 | 0,019 | | | | |
| **Education** | -1,035414 | 0,040499 | 0,000 | | | | |
| **Other** | 1,318304 | 0,0256272 | 0,000 | | | | |

*Table 20. Estimated coefficients, standard errors and Significance for Model 4*

The log pseudolikelihood of Model 4 equals 167,687.08 and was reached after Iteration 4. The pseudo R-squared has a value of 0.2240. 317,388 observations were included in Model 4.

As Model 4 is an extension of Model 3, similarities in Significance and estimated coefficients is expected. Therefore, the variables of interest will be discussed thoroughly. For other variables, only major or important difference compared to Model 3 will be discussed in the next sections.

#### *4.5.3.1 Variables of interest*

Both variables that were derived from *millennial* are found to be significant at all relevant confidence levels. A clear difference between the magnitude of the predicted coefficient of *early millennial* and *late millennial* is observed. Both variables were found to have a negative effect on *car use*. For *early millennial*, the log-odds of *car use* are 0.1336373 lower when compared to a non-millennial, ceteris paribus. *Late millennial* has a magnitude of 1.111958, which means that the log-odds of *car use* are 1.111958 lower for someone born between 1995 and 2009, when compared to a non-millennial.

As in Model 3, the interaction effect of *density* with the millennial variables was included. The interaction effects of *late millennial* with *density* was found to be not significant at any relevant confidence interval. For *early millennial,* the interaction effect with *High density* was found to be negative and significant at all relevant confidence levels. This implies that the log odds for *car use,* next to the effects of *early millennial* and *High density*, are 0.165051 lower for individuals that are early millennials and live in a high-density neighborhood c.p.. In this case, the interaction effect thus lowers the log-odds of *car use* even further. *Average density and early millennial* is shown to be not significant at all relevant confidence levels.

#### 4.5.3.2 Major and important differences

As can be expected, the estimated coefficients of Model 3 differ from those estimated in Model 4. However, no coefficient was estimated with an opposing sign or a large difference in magnitude. There are some variables that reach a higher level of significance in Model 4, for example *Friesland* was significant at the 90% confidence level in Model 3, while it is also significant at the 95% confidence level in Model 4. Further, no important or major differences were observed

#### 4.5.3.3 Goodness of fit

In order to see if Model 4 is better suited to predict *car use*, the PCP was calculated as a measure of Goodness of Fit. The predicted values using Model 4 were compared to the observed values of *car use* in the dataset and yielded the results as given in Table 21.

| Prediction\car use | No | Yes | Total |
|---|---|---|---|
| **No** | 113.607 | 35.632 | 149239 |
| **Yes** | 37.371 | 130.778 | 168149 |
| **Total** | **150.978** | **166.410** | |

*Table 21. Predictions and observations of car use in Model 4*

As shown in Table xx, Model 4 predicted *car use* for 168,149 observations. Out of these predictions, 130,778 predictions were correct, and 37,371 were predicted wrong. The dataset contains a total of 166,410 observations where respondents stated they used a car. This means that 130,778 out of the 166,410 predictions were right, equaling a PCP of 78.59%. For non-car use, 113,607 predictions out of 150,978 observations were right, which equals a PCP of 75.25%. Combining all correctly predicted values of Model 4, a total PCP of 77% was observed. The PCP values for Model 4 are thus summarized in Table xx.

| | |
|---|---|
| Car use correctly predicted | 78,59% |
| Non-car use | 75,25% |
| Total correctly predicted | 77,00% |

*Table 22. PCP values for Model 4*

### 4.5.4 Model 5: car driver

For Model 5, Model 3 was adjusted to obtain a regression on *car driver* instead of *car use*. In Model 5, the dependent variable was thus replaced by *car driver*, which was formulated as all car users that did actually drive a vehicle themselves, thus excluding passengers. Furthermore, the same variables as in Model 3 were used, which leads to the estimated coefficients, Robust Standard Errors and Significances as show in Table 23.

| Coefficient | Value | Robust Standard Error | Significance | Coefficient | Value | Robust Standard Error | Significance |
|---|---|---|---|---|---|---|---|
| Intercept | -22,33634 | 6,177013 | 0,000 | Weekend | -0,19024 | 0,0131016 | 0,000 |
| Millennial | -0,4116236 | 0,031232 | 0,000 | Fuel price | 14,94176 | 4,496837 | 0,001 |
| Millennial and Average density | -0,0042115 | 0,0293214 | 0,886 | Age | 0,001096 | 0,0005581 | 0,050 |
| Millennial and High density | -0,082729 | 0,0293768 | 0,005 | Gender | -0,53334 | 0,012405 | 0,000 |
| Low Density | 0,000 | 0,000 | 0,000 | No education | 0 | 0 | 0,000 |
| Average Density | -0,1709717 | 0,0179814 | 0,000 | Basic or lower education | 0,081084 | 0,0676931 | 0,231 |
| High Density | -0,2913609 | 0,0238177 | 0,000 | Lower advanced education | 0,501216 | 0,0691791 | 0,000 |
| 0-5000 Inhabitants | 0 | 0 | 0,000 | MBO, HAVO, VWO | 0,754486 | 0,0666605 | 0,000 |
| 5000-10000 Inhabitants | 0,1920515 | 0,1470139 | 0,191 | College or University | 0,647823 | 0,0662383 | 0,000 |
| 10000-20000 Inhabitants | 0,141026 | 0,1392855 | 0,311 | Other education | 0,554482 | 0,0786775 | 0,000 |
| 20000-50000 Inhabitants | 0,1480461 | 0,140283 | 0,291 | Lowest Income Group | 0 | 0 | 0,000 |
| 50000-100000 Inhabitants | 0,1429331 | 0,1415926 | 0,313 | Income Group 2 | 0,195155 | 0,020952 | 0,000 |
| 100000-150000 Inhabitants | 0,0566874 | 0,141243 | 0,688 | Income Group 3 | 0,272459 | 0,0278596 | 0,000 |
| 150000-250000 inhabitants | 0,0388873 | 0,1426443 | 0,847 | Income Group 4 | 0,291294 | 0,0295881 | 0,000 |
| More than 250000 Inhabitants | -0,3068119 | 0,1492905 | 0,040 | Income Group 5 | 0,367857 | 0,0314963 | 0,000 |
| Groningen | 0 | 0 | 0,000 | No working hours | 0 | 0 | 0,000 |
| Friesland | 0,104999 | 0,0273973 | 0,000 | Less than 12 working hours | 0,139815 | 0,0258204 | 0,000 |
| Drenthe | 0,4126917 | 0,0822961 | 0,000 | 12 to 30 working hours | 0,568 | 0,0167138 | 0,000 |
| Overijssel | 0,3835146 | 0,0743721 | 0,000 | 30 or more working hours | 0,633032 | 0,0186541 | 0,000 |
| Flevoland | -0,0863102 | 0,0695498 | 0,215 | 2005 | 0 | 0 | 0,000 |
| Gelderland | 0,4901065 | 0,1099796 | 0,000 | 2006 | -0,31271 | 0,1004301 | 0,002 |
| Utrecht | 0,2114025 | 0,0433798 | 0,000 | 2007 | -0,91328 | 0,2958755 | 0,002 |
| Noord-Holland | 0,2671652 | 0,0850222 | 0,002 | 2008 | -1,81544 | 0,5666173 | 0,001 |
| Zuid-Holland | 0,4774798 | 0,0860527 | 0,000 | 2009 | -0,04004 | 0,0244935 | 0,102 |
| Zeeland | -0,0558979 | 0,0491321 | 0,255 | 2010 | -2,3566 | 0,68881 | 0,001 |
| Noord-Brabant | 0,454244 | 0,056793 | 0,000 | 2011 | -4,39804 | 1,310926 | 0,001 |
| Limburg | 0,3604278 | 0,0273711 | 0,000 | 2012 | -6,15837 | 1,848882 | 0,001 |
| lnkmafstr | 0,6854287 | 0,0053864 | 0,000 | 2013 | -5,81294 | 1,745599 | 0,001 |
| Work | 0 | 0 | 0,000 | 2014 | -5,24652 | 1,558299 | 0,001 |
| shopping | 0,3623585 | 0,0139371 | 0,000 | 2015 | -3,12785 | 0,9372483 | 0,001 |
| Services | 0,3307972 | 0,0241438 | 0,000 | 2016 | -1,924 | 0,5714196 | 0,001 |
| Social or Recrational | 0,4806513 | 0,0177656 | 0,000 | 2017 | -3,04979 | 0,9100725 | 0,001 |
| Education | -0,9689547 | 0,0351552 | 0,000 | | | | |
| Other | 0,8967654 | 0,0236356 | 0,000 | | | | |

*Table 23. Estimated coefficients, Robust Standard Errors and Significance for Model 5*

The log pseudolikelihood of Model 5 is -167,687.08 and the pseudo R-squared is 0.2240. Model 5 consists of 317,388 observations.

#### 4.5.2.1 Intercept and variables of interest

The intercept has a value of -22.33634 and is significant at all relevant confidence levels. This means the intercept has a relevant meaning when predicting car use. However, as not all other variables can be zero at the same time, it has no interpretable effect on its own.

*Millennial* was estimated with a negative sign and a value of -0.4116236. It is significant at all relevant confidence levels. This implies that the log-odds of *car driver* are 0.4116236 lower for millennials when compared to non-millennials, c.p.. Thus the probability of being a car driver declines when an individual is a millennial.

*Density* was also found to have a significant effect on *car driver*. Both *Average density* and *High density* show significance of 0.000, which means both are significant at all relevant confidence levels. Living in a neighborhood that is classified as *Average density* decreases the log-odds of *car driver* by 0.1709717 c.p. when compared to an individual that lives in an area that is classified as a *Low density area*. For *High density* areas, the magnitude on the log-odds of *car driver* are higher, as the estimated coefficient implies that log-odds are 0.2913609 lower c.p. for someone that lives in a *High density* area, compared to the reference category *Low density*.

As in Model 3, the interaction effect of *millennial* and *Density* was included. For *Millennial and average density,* significance was found to be 0.886, which implies it is not significant at any relevant confidence interval. The interaction between *Density* and *High density,* however is significant at all relevant confidence levels and supplies an additional effect on *car driver* for millennials that live in a high density area. The additional magnitude was estimated to be -0.082729 c.p..

### *4.5.2.2 Land-use variables*

The municipality size and province of residence were included to control for land-use and locational factors. Municipality size, *gemgr*, was found to be significant only for the largest municipality size. This implies that the coefficients between *Less than 5,000 Inhabitants* and *200,000 to 250,000 Inhabitants* are not statistically different from zero. The log-odds of *car driver* are 0.3068119 lower for individuals that live in the largest municipality compared to the reference category, c.p..

Compared to earlier formulated models, the effect of *prov* on *car driver* is estimated to be significant at all relevant confidence levels for all provinces except for *Zeeland* and *Flevoland*. Notably, all significant effects were found to be positive, while the provinces that were found to be not significant were estimated to have a negative effect on *car driver*.

### *4.5.2.3 Trip-related variables*

Both the variable that deals with trip length as the variable for trip purpose were found to be significant at all relevant confidence levels. Trip length was found to have a positive effect on *car driver*, which means that longer distances  increase the chance that an individual will drive a car. For trip purpose, *social and recreational* and *educational* trips were estimated to decrease the log-odds of driving a car by 0.4806513 and 0.9689547 respectively, ceteris paribus.

Model 5 also predicts a lower chance of driving a car for trips in the *weekend*, while *fuel price* is expected to increase the log-odds of *car driver.*

### *4.5.2.4 Socio-demographic variables*

Ageing, as in earlier models, was found to have a significant effect at the 90% and 95% confidence level, and is expected to increase the log-odds of *car driver*, thus making it more likely to drive a car. The coefficient for *gender* was found to have a negative effect, significant at all relevant confidence levels, thus lower log-odds for *car driver* are expected if the respondent is female.

*Education, Income* and *working hours* were also found to have a significant positive effect on *car driver at all relevant confidence levels*, except for *Basic or lower education*, which was found to be not significant at any relevant confidence level. The differences in magnitudes for *income* and *working hours* was expected, as the largest magnitude is predicted for the highest category, while the lowest magnitude is observed at the smallest category. For *education,* however, the largest magnitude was observed for individuals that have the *MBO, HAVO an VWO* level of education.

The year of observation was also found to be significant at all relevant levels, except for *2009*. Compared to *2005*, all other years have negative coefficients, which means the log-odds of *car driver* are lower in all years after *2005*, ceteris paribus. The magnitudes vary between years, but the largest magnitude was found in *2012*.

### *4.5.2.5 Goodness of fit*

The goodness of fit for Model 5 was calculated the same way as for the earlier described models. The PCP was thus calculated using by comparing the predicted values of Model 5 with the actual observations in the dataset. Tabel 24 contains these results.

| Prediction\car driver | No | Yes | Total |
|---|---|---|---|
| **No** | 148.194 | 48.563 | 196757 |
| **Yes** | 35.356 | 85.275 | 120631 |
| **Total** | **183.550** | **133.838** | |

*Table 24. Predictions and observations of car use in Model 5*

As shown in Table 24, 148,194 predictions of not driving a car were made correctly, while 48,563 predictions of not driving a car were made while the respondent actually drove a car. The PCP values for Model 5 were summarized in Table 25.

| | |
|---|---|
| Car driver correctly predicted | 63,72% |
| Non-car driver | 80,74% |
| Total correctly predicted | 73,56% |

*Table 25. PCP values for Model 5*

Car driver is thus more accurately predicted for the instances were the respondent was not the driver, as 80.74% of the observations were correctly predicted. In general, 73,56% of al observations were correctly predicted by Model 5.

## 4.6 Model Comparison

Now that all results of all models have been presented and discussed, this section will supply a brief comparison of all models. The next Chapter will discuss the findings presented in Chapter 4 and formulate an answer on the hypotheses.

The most important comparison is between the unrestricted model, Model 1, and the unrestricted models Model 2 and Model 3. Also, the comparison between Model 2 and Model 3 is of importance, as Model 3 also considers the RSS in terms of an interact term between *millennial* and *density*. A comparison was made in terms of Coefficient and Significance-value and summarized in Table 26, at the end of this section.

Table 26 shows that most coefficients of Model 2 and Model 3 do not differ much in terms of estimated coefficients or significance levels. However, as the interaction term isolates the effect of Residential self-selection from the *millennial* variable, a higher PCP was found.

Model 4 and Model 5 were both formulated to extend on Model 3, where Model 4 included two groups of millennials, *early millennial* and *late millennial*, whereas Model 5 replaced the dependent variable *car use* by the more strict variable *car driver*, decreasing the amount of instances where the dependent variable was true. For Model 4, it was found that the magnitude of the effect of *late millennial* was higher than that of *early millennial*, meaning that individuals that were born between 1995 and 2009 were more likely to not use a car when compared to their non-millennial or early-millennial counterparts. In Model 5, the coefficient for *millennial* was found to have a larger magnitude on *car driver*, implying the direct effect of *millennial* on *car driver* to be bigger than the effect on *car use*. However, it is noted that the estimated coefficients for other variables differ between Model 3 and Model 5, thus a more thorough assessment is needed to be able to draw conclusions

| Coefficient | Value M1 | Significance M1 | Value M2 | Significance M2 | Value M3 | Significance M3 |
|---|---|---|---|---|---|---|
| Intercept | 0,180749 | 0,000 | -16,64218 | 0,011 | -16,32741 | 0,012 |
| Millennial | -0,4936685 | 0,000 | -0,2623367 | 0,000 | -0,2164384 | 0,000 |
| Millennial and Average density | - | - | - | - | 0,0026095 | 0,936 |
| Millennial and High density | - | - | - | - | -0,0973445 | 0,003 |
| Low Density | - | - | 0 | 0,000 | 0 | 0,000 |
| Average Density | - | - | -0,1840493 | 0,000 | -0,1840053 | 0,000 |
| High Density | - | - | -0,3112903 | 0,000 | -0,2961819 | 0,000 |
| 0-5000 Inhabitants | - | - | 0 | 0,000 | 0 | 0,000 |
| 5000-10000 Inhabitants | - | - | 0,1850703 | 0,154 | 0,1846128 | 0,155 |
| 10000-20000 Inhabitants | - | - | 0,1638594 | 0,189 | 0,1629412 | 0,192 |
| 20000-50000 Inhabitants | - | - | 0,1703676 | 0,180 | 0,1691595 | 0,184 |
| 50000-100000 Inhabitants | - | - | 0,1740051 | 0,172 | 0,1740051 | 0,172 |
| 100000-150000 Inhabitants | - | - | 0,0642572 | 0,611 | 0,0629266 | 0,618 |
| 150000-250000 inhabitants | - | - | 0,0230861 | 0,857 | 0,023582 | 0,855 |
| More than 250000 Inhabitants | - | - | -0,3572431 | 0,008 | -0,357119 | 0,008 |
| Groningen | - | - | 0 | 0,000 | 0 | 0,000 |
| Friesland | - | - | 0,0687096 | 0,005 | 0,0673094 | 0,006 |
| Drenthe | - | - | 0,3229762 | 0,000 | 0,3191772 | 0,000 |
| Overijssel | - | - | 0,2802602 | 0,000 | 0,2754339 | 0,000 |
| Flevoland | - | - | -0,137622 | 0,852 | -0,0120801 | 0,870 |
| Gelderland | - | - | 0,3498433 | 0,002 | 0,3431535 | 0,003 |
| Utrecht | - | - | 0,1202816 | 0,007 | 0,1184617 | 0,008 |
| Noord-Holland | - | - | 0,1313505 | 0,134 | 0,1265015 | 0,147 |
| Zuid-Holland | - | - | 0,3787585 | 0,000 | 0,3736024 | 0,000 |
| Zeeland | - | - | -0,0475017 | 0,388 | -0,0464609 | 0,396 |
| Noord-Brabant | - | - | 0,4173092 | 0,000 | 0,4135628 | 0,000 |
| Limburg | - | - | 0,4268073 | 0,000 | 0,4261627 | 0,000 |
| Inkmafstr | - | - | 1,049833 | 0,000 | 1,049731 | 0,000 |
| Work | - | - | 0 | 0,000 | 0 | 0,000 |
| shopping | - | - | 0,8704774 | 0,000 | 0,8706427 | 0,000 |
| Services | - | - | 1,059901 | 0,000 | 1,059668 | 0,000 |
| Social or Recrational | - | - | 0,0514837 | 0,016 | 0,051957 | 0,015 |
| Education | - | - | -1,132622 | 0,000 | -1,139349 | 0,000 |
| Other | - | - | 1,33372 | 0,000 | 1,333955 | 0,000 |
| Weekend | - | - | 0,182982 | 0,000 | 0,1828303 | 0,000 |
| Fuel price | - | - | 10,16642 | 0,034 | 9,931167 | 0,037 |
| Age | - | - | 0,001104 | 0,050 | 0,0011612 | 0,041 |
| Gender | - | - | 0,0610869 | 0,000 | 0,0611177 | 0,000 |
| No education | - | - | 0 | 0,000 | 0 | 0,000 |
| Basic or lower education | - | - | 0,0635422 | 0,287 | 0,0643478 | 0,281 |
| Lower advanced education | - | - | 0,3071179 | 0,000 | 0,3089951 | 0,000 |
| MBO, HAVO, VWO | - | - | 0,440591 | 0,000 | 0,4428963 | 0,000 |
| College or University | - | - | 0,2474949 | 0,000 | 0,2504265 | 0,000 |
| Other education | - | - | 0,22932 | 0,001 | 0,230877 | 0,001 |
| Lowest Income Group | - | - | 0 | 0,000 | 0 | 0,000 |
| Income Group 2 | - | - | 0,2326699 | 0,000 | 0,2308596 | 0,000 |
| Income Group 3 | - | - | 0,3087416 | 0,000 | 0,6062507 | 0,000 |
| Income Group 4 | - | - | 0,3322579 | 0,000 | 0,3293883 | 0,000 |
| Income Group 5 | - | - | 0,4011631 | 0,000 | 0,3974176 | 0,000 |
| No working hours | - | - | 0 | 0,000 | 0 | 0,000 |
| Less than 12 working hours | - | - | 0,083317 | 0,002 | 0,0826848 | 0,002 |
| 12 to 30 working hours | - | - | 0,4150621 | 0,000 | 0,4160671 | 0,000 |
| 30 or more working hours | - | - | 0,5563194 | 0,000 | 0,5570966 | 0,000 |
| 2005 | - | - | 0 | 0,000 | 0 | 0,000 |
| 2006 | - | - | -0,2264085 | 0,030 | -0,2216603 | 0,033 |
| 2007 | - | - | -0,6726089 | 0,029 | -0,6577977 | 0,032 |
| 2008 | - | - | -1,275816 | 0,033 | -1,246208 | 0,037 |
| 2009 | - | - | -0,115338 | 0,000 | -0,1148701 | 0,000 |
| 2010 | - | - | -1,685713 | 0,021 | -1,649745 | 0,023 |
| 2011 | - | - | -3,166596 | 0,023 | -3,09803 | 0,025 |
| 2012 | - | - | -4,351617 | 0,027 | -4,254948 | 0,030 |
| 2013 | - | - | -4,111809 | 0,027 | -4,020725 | 0,030 |
| 2014 | - | - | -3,769768 | 0,023 | -3,688119 | 0,025 |
| 2015 | - | - | -2,313409 | 0,020 | -2,263929 | 0,022 |
| 2016 | - | - | -1,50769 | 0,013 | -1,476867 | 0,014 |
| 2017 | - | - | -2,298261 | 0,017 | -2,2499 | 0,019 |

*Table 26. Comparison of Model 1, Model 2 and Model 3*

# 5. Discussion

This Chapter will discuss the results presented in Chapter 4. These findings will be assessed using the literature that was discussed in Chapter 2. This discussion will then be used to formulate an answer on the Hypotheses and the general Research Question.

## 5.1 The effect of being a millennial on car use

First, Model 1 was formulated with *millennial* being the only independent variable. Following the findings by Jorritsma & Berveling (2014), a negative effect was expected, implying millennials make less use of a car than non-millennial counterparts. The "Millennial Identity", however, connects not only the desire to make the world a better place (Paulin, Ferguson, Jost, & Fallu, 2014) to the "Millennial Identity", also impatience and intolerance for delays (Karakas, Manisaligil, & Sarigollu, 2015) and latent laziness, based on habits and the concept of helicopter parents (Alexander & Sysko, 2012) (Guell, Panter, Jones, & Ogilvie, 2012) were distinguished. This implies that former research has shown that millennials tend to make less use of a car, while literature also shows evidence for more car use among millennials. Based on the findings of Model 1, being millennial was found to decrease the log-odds of car use, thus implying that the dominant factor in the "Millennial Identity" is the sustainability trait. However, latent laziness and especially habits can be foregone when proper intervention is conducted, rendering past behavior insignificant in predicting modal choice (Bamberg, Ajzen, & Schmidt, 2003).

Model 2 controlled for other variables when assessing the effect of millennials on car use. When compared with Model 1, the magnitude of the effect was moderated. Howeverthe coefficient was still estimated to be negative, thus decreasing the log-odds of car use for millennials when compared to non-millennials. When addressing the intercept, Model 1 contains an intercept that has a real-world meaning, as it can be interpreted as the log-odds of car-use for a non-millennial. In Model 2, the intercept was found to be larger, but not interpretable, as two other variables cannot be zero. Both *age*, starting at 16, and *fuel price*, starting at 1.352 and having a maximum value of 1.80192. This means that the value that can be predicted by Model 2 that is closest to the intercept is for an individual that has age 16, has fuel available for 1.352 euro/liter and is further in the lowest category of all other variables. This value can be calculated by using the formula in section 3.3.2. It should be noted that this fuel price was found in the province of *Gelderland* in the year *2005*, which means that a combination of *Groningen, 1.352* and *2005* does not exist in the dataset, but might exist in real life. This combination of variables returns a probability of 0.0531754914, which is rounded to a probability of 5.32%. A millennial, in this case, will have a probability of car use of 0.0413873081, which can be rounded to 4.14%. At this level, being millennial decreases the probability of car use by 1.18%, ceteris paribus. For an individual that lives in a high-density neighborhood, 100,000-150,000 Inhabitants, Limburg, traveled a distance of 2 kilometer (thus lnkmafstr=ln(2)) for other purposes in the weekend, a fuel price of 1.80192, is a 24 year old male with a MBO,HAVO or VWO diploma, who works 30 or more hours a week and lives in 2012 income 4, the probability was calculated to be 0.7624564307, which was rounded to 76.25%. Including the millennial variable in this case yields a probability of 0.7117415421, which is rounded to 71.17%. In this case, the effect of millennial is thus 5.08%, ceteris paribus. As calculated in section 4.1, Model 1 predicts a decrease 12.26% of in the probability of car use.

The estimated effect of *fuel price* on *car use* is positive, which means that if *fuel price* rises, there is a larger probability of car use, controlling for other variables. As *fuel price* was used as a proxy for cost of travel, this finding implies that the demand for car use is inelastic for *fuel price*.

The effect of *income* was expected to have a positive impact on *car use*, as higher incomes increase the mobility flexibility of individuals (Watkins, 2016). Model 2 supports this suggestion, as higher income groups were estimated to have a larger effect on *car use*. Watkins (2016) also argued that individuals with a higher income have more residential locations available, and can therefore are able to minimize commuting costs (Watkins, 2016). However, as commuting costs are different for all individuals, based on their preferences (Kronenberg & Carree, 2012), richer individuals might not consider these commuting costs when choosing their residential location.

As is expected when using a logistic regression, the impact of *millennial* in terms of reduction in probability varies based on the sum of all other coefficients. The effect of *millennial* in terms of probabilities if the sum of all other variables is close to zero. For a sum that exceeds -10 or 10, the effect of millennial is found to be less than 0.01%.

Both Model 1 and Model 2 thus supply evidence for a relationship between the millennial cohort and car use. As formulated before, Hypothesis 1 was formulated as:

*H1. There is a significant difference between millennials and former generations concerning the modal choice of transportation, controlling for other variables.*

Model 1 supplied evidence for a negative relationship between *millennial* and *car use*, without controlling for other variables. Model 2 did control for other variables and also found significant evidence for a negative effect of *millennial* on *car use*, albeit of less magnitude. As *millennial* was found to be significant in both Model 1 and Model 2 at all relevant confidence levels (90%, 95% and 99%), Hypothesis 1 cannot be rejected. This means that there is evidence that millennials differ from earlier cohorts in terms of modal choice of transportation, controlling for other relevant land use and socio-demographic variables.

Hypothesis 1 thus is accepted, implying that an effect of *millennial* on *car use* was proven. However, *millennial* represents a cohort of individuals that are likely to share the same values and preferences. However, these preferences might still differ between persons. Based on the findings in this research, the "sustainability trait" of millennials seems to be the dominant trait, as the effect on car use was found to be negative. This research does thus prove that there is an effect, but it does not explain what underlying factors contribute with which magnitude to this effect. For this, a more extensive research should be conducted, that also includes the preferences and values of respondents. For the purpose of this research, the underlying factors of *millennial* are out of scope, but can be relevant for further research.

## 5.2 Differences between 2006 and 2016

As discussed in section 5.2, *millennial* was found to have a negative effect on *car use*. Model 2a and Model 2b were formulated to test the second hypothesis, which was formulated as

*H2. There is a significant difference between same-aged individuals in 2006 and millennials in 2016 concerning the modal choice of transportation, controlling for other variables.*

This hypotheses aims to find differences between the cohorts, where *millennial* was replaced by *milage2006*, which represents all respondents in 2006 that are of the same age as the millennials are in 2016. As *millennial* was defined as all individuals born between 1982 and 2009 (Alexander & Sysko, 2012), all individuals between ages 7 and 34 were marked as *millennial*. This implies that all respondents born between 1974 and 1999 were included in *milage2006* and thus an overlap exists.

As discussed is Section 4.3.3, some major differences between both models can be found. The most notable of these finding is that the effect of *millennial* in Model 2b was found to be insignificant. This means that for 2016, no conclusions can be drawn on the effect of *millennial* on *car use*, although the coefficient suggests a negative effect. Also notable is the effect of *milage2006* on *car use* in Model 2a, as Model 2a estimates a positive and significant effect. This implies that someone in 2006 that was of the same age as a millennial was in 2016, is more likely to use a car for their trips.

Model 2 does suggest a significant, negative effect of *millennial* in *2016* on *car use*. However, as all observations of the dataset were used in this model, while Model 2a and Model 2b only used observations of a particular year to fit the model and predict the probabilities of car use. Therefore, bias in the chosen year might exist, which can be attributed to economic or political circumstances. Also, it is expected that attitudes towards sustainability and environment were more present in 2016, as these issues are more generally considered by individuals.

In Model 2b, the effect of *income* on *car use* was found to be highest for *Income Group 3*, which contains the average income level. This finding can be part of the inability to live in certain residential locations, which implies more car use is expected (Watkins, 2016).

Relating the findings of Model 2a and Model 2b to Hypothesis 2, a partial answer can be formulated. Model 2 did predict the expected negative effect of millennial on *car* use. However, Model 2b did not show a significant effect of *millennial* on *car use,* which implies no unambiguous conclusions can be drawn. Model 2a did show a significant positive effect for millennial-aged individuals in 2006. Therefore, it can be stated that there is a difference between millennials and their same-aged counterparts, although it remains unclear if this difference is significant. Therefore, we have no evidence to support Hypothesis 2, and thus reject it based on the findings of Model 2a and 2b.

## 5.3 Millennials in urban environments

Model 2 does not control for the effect of RSS on car use. As mentioned by Guan & Wang (2019), people that share the same values and attitudes towards travel behavior are likely to choose the same or similar residential locations (Guan & Wang, 2019). Model 3 used the interaction term between *millennial* and *density* to control for RSS. As it is expected that millennials share the same values and attitudes, RSS is expected for the millennial cohort. The interaction term between *millennial* and *density* isolates the effect of RSS, thus removing bias in both the estimated coefficients of *millennial* and *density*.

Model 3 thus also controls for RSS. This effect was only found to be significant for millennials that live in a high density area. When comparing the estimated coeffeicients of Model 3 with those of Model 2, it was found that the magnitudes of both *millennial* and *high density* were higher in Model 2. However, a millennial that lives in a high density area would have an effect -0.573627 (sum of the coefficients of both *millennial* and *high density*) on the log-odds of *car use*, whereas the total effect would be -0.6099648 in Model 3. Without controlling for the RSS, Model 2 thus underestimates the effect of *density* and *millennial* by 0.0363378 in terms of log-odds. The interaction term of *average density* with *millennial* was not found to have a significant effect. This implies that the RSS is in this case better suited when addressing high-density areas. The reason for this cannot be derived directly from Model 3, but might be attributed to preferences of millennials towards residential location, or the lack of millennials' interest in living in average-density neighborhoods.

The estimated effect of income, when controlling for RSS, does change significantly when compared to Model 2. In Model 2, *Income Group 5* was estimated to have thebiggest magnitude when predicting *car use*. In Model 3, as with Model 2b, *Income Group 3* was found to have the biggest magnitude in predicting *car use*. This, again, can be related to the residential location. It becomes clear that income and RSS do seem to have some relationship. As an individual can only freely choose their residential location if its income is high enough to afford living in such a place (Watkins, 2016), it can be explained why indivuduals with a lower income are estimated to make more use of car, as their commuting distances might be longer.

Model 3 was formulated to formulate an answer to Hypothesis 3, which tests the effect of millennials in high-density environments. Formally, it was formulated as

*H3.* *There is a significant difference between millennials that live in high-density areas and millennials that do not live in high-density areas, controlling for other variables*

Already found in Model 2, both *millennial* and *high density* were found to have a significant negative effect on *car use*, c.p.. This already shows that millennials in a high-density area are predicted to make less use of a car than millennials that live in al *low density* or *average density* area. Model 3 further supports the hypothesis, showing that Model 2 underestimates the effect of millennials in high-density areas, as the effect of RSS was not controlled for in Model 2. Based on the findings in both models, evidence is found for a difference between millennials in high density areas and low- or average- density areas. Therefore, we do not reject Hypothesis 3.

## 5.4 Additional findings

Model 4 and Model 5 were formulated to further explore the relationship between millennials and car use, where Model 4 distinguished between two groups of millennials, while Model 5 replaced car use by driving a car as the dependent variable. This section will discuss the major findings these models add to the outcomes of Model 2 and Model 3.

### 5.4.1 Early millennials versus late millennial

As mentioned, Model 4 extends on Model 3 by making a distinction between *early millennials*, born in or before 1994, and *late millennials*, born in 1995 or after, while *millennial* considered all individuals that were born between 1982 and 2009, as proposed by Alexander & Sysko (2012). Dividing this group in two separate groups might be useful, as it can be expected that the *late millennial* grew up in different times. For example, they grew up during the rise of the internet and are increasingly tech savvy, which might explain their intolerance for delays (Karakas, Manisaligil, & Sarigollu, 2015), as they expect real time information to be present at all time, also when choosing their travel modality. Information about travel times, departing times and even crowdiness of specified public transport modes are available at any time, thus providing reliable information. Driving, however, as mentioned by Verlaan (2018), leads to more delays as congestion rises.

Using the same dataset as was used to test the hypotheses, Model 4 showed evidence for a difference in the effect of late millennials, when compared to early millennials. It was still found that *early millennial* is expected to decrease the probability of *car use*. The effect of *late millennial*, however, was shown to have a bigger magnitude when predicting car use. In absolute terms, the log-odds of *car use* were found to be 0.978321 lower for *late millennial* than for *early millennial*, keeping other variables stable. This difference was moderated, however, by the effect of RSS. Model 4 showed that only the effect of *early millennial aand high density* was significant, thus reducing the difference between early and late millennials in high density areas. Even more interesting to note is the effect of RSS has a positive impact on *car use* for late millennials in high-density areas, implying that the effect of RSS is different for late millennials than it is for early millennials. However, as the coefficient of *late millennial and high density* is not significant, no conclusions can be drawn. On of the reasons that can be underlying at this difference is that the late millennial group in this research were a maximum of 22 years old in 2012 and might not have made their own residential location decision, as it is viable that those individuals still live at their parents' house. This might also explain the magnitude of *late millennial*, as late millennials are not old enough to make use of a car themselves, and thus are dependent on others.

Model 4 thus supplies evidence for a distinction between early millennials and late millennials in terms of the magnitude of the effect of car use. The RSS shows an insignificant, but interesting positive estimated coefficient, which might imly a shift among the millennial population in residential location choice, and thus modal choice of transportation. Further research is needed to validate this finding, which is out of scope of this research.

The percent correctly predicted value of Model 4 exceeds the values of all other models, thus implying it is best suited when addressing the effect of *millennial* on *car use*. Dividing the millennial cohort in multiple groups thus improves the goodness of fit.

### 5.4.2 Active car use versus passive car use

All earlier discussed models consider *car use*, which was specified as all trips undertaken by car. Model 5 extends on this by tightening the scope of car users, including only respondents that stated they drove the car themselves. Model 3 was therefore adjusted to regress on *car driver* instead of *car use*.

Based on the PCP, this model was less accurate than all other models, implying that driving a car differs from using a car. A noteworthy difference is found in the estimated coefficients for *weekend* and *gender*, as Model 5 predicts that trips in the weekend are less likely to be made by car when the respondent is the driver. Also, it is estimated that men make more active use of a car than men do, which was expected following the research of Böcker, van Amen, & Helbich (2017). Trip purpose is estimated to be of less magnitude when compared to the reference category *work* for driving a car than it is for using a car. This decrease in magnitude can most likely be attributed to a decrease in the difference between *work* and other trip purposes. It can be argued that if an individual has to work, it more often drives a car, while for other activities an individual carpools with someone.

The price of fuel might be expected to be negative, as the driver of a car mostly pays for the fuel it needs to get to its destination. However, *fuel price* is estimated to increase the probability of *car driver*, thus again implying an inelastic demand for fuel.

As this research was predominantly interested in the effect of being a millennial on car use, Model 5 does show that there is a significant effect of being a millennial on active car use, implying that millennials also are less likely to drive a car. In this case, further research is needed to properly address this effect, as other factors might become more important, for example car ownership or more detail on the preferences and attitudes individuals have towards driving a car.

# 6. Concluding remarks and further research

This research aimed to find a relationship between the millennial cohort and car use. As millennials are expected to have a different set of attitudes and values towards sustainability and the environment than former generations, it was expected that millennials do make less use of cars. However, the “Millennial Identity” which was introduced in this research, also contains factors of impatience and latent laziness, which might push millennials towards car use.

Based on the dataset of OViN and MON, logistic regression was used to investigate the difference between millennials and former generations in predicting car use. Statistical evidence was supplied that implies that millennials do make less use of a car when compared to other generations, controlling for other variables. This supports Hypothesis 1, which thus was accepted. This research thus shows that millennials do behave different from former generations in terms of car use

As data between 2005 and 2017 was made, the difference between years was also examined. Some evidence was found that would imply that car use was less among millennials in 2016 than it was in 2006. However, fitted models on both 2005 and 2016 did show some difference, although this difference was not statistically supported. This research thus does not provide evidence for a statistical difference between millennials in 2016 and their same-aged equivalents in 2006, which was tested by Hypothesis 2. This hypothesis is thus rejected. Further research can be conducted by including more years to the models, thus controlling for year-specific events. Also, socio-economic, political and attitudinal factors might be added to control for.

Land use and Residential Self Selection were found to influence car use. As not controlling for RSS underestimated the predicted car use of millennials, it was found to be an important factor when addressing car use. Especially in high density areas, an effect of RSS was found. As the effect of *millennial* and *density* was underestimated when not controlling for RSS, statistical evidence was found that millennials in high density areas are less likely to use a car than millennials in lower-density areas. This supports Hypothesis 3, which was thus accepted.

One of the main additional findings was found in a distinction among millennials, as it was found that the effect on car use is larger when considering late millennials, compared to early millennials. However, the effect of RSS was only found to be significant for early millennials, which might imply a shift in the dominant underlying factor of millennial. More information on the built environment, land use or location of respondents can be used to further examine this relationship.

In this research, millennials were all assumed to share the same traits, values and preferences, as they were only considered as a cohort. However, as all individuals are different, differences among millennials are likely t be present. This research thus shows that the millennial cohort acts different relating to the modal choice of transportation. However, more personal information should be included in the model in order to assess the underlying factors in the millennial cohort. This factors were not present for this study, but can be of interest when further assessing the effect of the millennial cohort on car use. Millennials are thus proven to be different than former generations in terms of modal choice of transportation, while further research is needed to show what causes this difference.

## References

Aditjandra, P. T. (2013). The impact of urban development patterns on travel behaviour: Lessons learned from a British Metropolitan region using macro-analysis and micro-analysis in addressing the sustainability agenda. *Research in Transportation Business & Management, 7*, 69-80.

Alexander, C. S., & Sysko, J. M. (2012). A STUDY OF THE COGNITIVE DETERMINANTS OF GENERATION Y'S ENTITLEMENT MENTALITY. *Academy of Educational Leadership Journal, 16*(2), 63-68.

Bamberg, S., Ajzen, I., & Schmidt, P. (2003). Choice of Travel Mode in the Theory of Planned Behavior: The Roles of Past Behavior, Habit, and Reasoned Action. *Basic and Applied Social Psychology, 25*(3), 175-187.

Böcker, L., van Amen, P., & Helbich, M. (2017). Elderl travel frequencies and transport mode choices in Greater Rotterdam, The Netherlands. *Transportation, 44*, 831-852.

Cao, X., Mokhtarian, P. L., & Handy, S. L. (2009). Examining the impacts of residential self-selection on travel behaviour: a Focus on empirical Findings. *Transport Reviews, 29*(3), 359-395.

CBS. (2019). *Pompprijzen motorbrandstoffen.* CBS.

Centraal Bureau voor de Statistiek (CBS); Rijkswaterstaat (RWS). (2017). *Onderzoek Verplaatsingen in Nederland 2016 - OViN 2016.* DANS. doi:https://doi.org/10.17026/dans-293-wvf7

Centraal Bureau voor de Statistiek; Rijkswaterstaat. (2017). *Data Files.* Retrieved from DANS: https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:73441/tab/2

DANS. (n.d.). *Home*. Retrieved March 18, 2019, from DANS: https://dans.knaw.nl/nl

Garcia-Sierra, M., Miralles-Guasch, C., Martínez-Melo, M., & Marquet, O. (2018). Empirical analysis of travellers' routine choice of means of transport in Barcelona, Spain. *Transportation Research Part F, 55*, 365-379.

Guan, X., & Wang, D. (2019). Residential self-selection in the built environment-travel behavior connection: Whose self-selection? *Transportation Research Part D: Transport and Environment*(67), 16-32.

Guell, C., Panter, J., Jones, N. R., & Ogilvie, D. (2012). Towards a differentiated understanding of active travel behaviour: Using social theory to explore everyday commuting. *Social Science & Medicine, 75*, 233-239.

Harms, L. (n.d.). *Mobility among Ethnic Minorities in the Urban Netherlands*. Retrieved 1 29, 2019, from Deutsches Institut für Urbanistik: https://difu.de/publikationen/mobility-among-ethnic-minorities-in-the-urban-netherlands.html#2

Heinen, E. (2016). Identity and travel behaviour: A cross-sectional study on commute mode choice and intention to change. *Transportation Research Part F*, 238-253.

Hubbard, J. (2013). Meet the Millennials. *Finweek*, 43.

Jorritsma, P., & Berveling, J. (2014). *Niet autoloos, maar auto later.* Den Haag: Kennisinstituur voor Mobiliteit.

Karakas, F., Manisaligil, A., & Sarigollu, E. (2015). Management learning at the speed of life: Designing reflective, creative, and collaborative spaces for millennials. *The International Journal of Management Education, 13*(3), 237-248.

Kennisinstituut Mobiliteit. (2018). *Kerncijfers Mobiliteitsbeleid 2018.* Den Haag: Ministerie van Infrastructuur en Waterstaat.

Kennisinstituut Mobiliteit. (2018). *Stabiele beelden: trends in beleving en beeldvorming van mobiliteit.* Den Haag: Ministerie van Infrastructuur en Waterstaat.

Klocker, N., Toole, S., Tindale, A., & Kerr, S.-M. (2015). Ethnically Diverse Transport Behaviours: An Australian Perspective. *Geographical Research, 53*(4), 393-405.

Kronenberg, K., & Carree, M. (2012). On The Move: Determinants of Job and Residential Mobility in Different Sectors. *Urban Studies, 49*(16), 3679-3698.

Kultalahti, S., & Viitala, R. (2015). Generation Y – challenging clients for HRM? *Journal of Managerial Psychology, 30*(1), 101-114.

Meurs, H., & Haaijer, R. (2001). Spatial structure and mobility. *Transportation Research Part D, 6*, 429-446.

Murtagh, N., Gatersleben, B., & Uzzell, D. (2012, September). Multiple identities and travel mode choice for regular journeys. *Transportation Research Part F: Traffic Psychology and Behaviour, 15*(6), 514-524.

Newbold, K. B., Scott, D. M., Spinney, J. E., Kanaroglou, P., & Páez, A. (2005). *Journal of Transport Geography, 13*, 340-351.

Noth, M., Borning, A., & Waddel, P. (2003). An extensible, modular architecture for simulating urban development, transportation and environmental impacts. *Computers, Environment and Urban Systems, 27*(2), 181-203.

Omroep Flevoland. (2018, June 13). *Benzine is het duurst in Flevoland*. Retrieved March 25, 2019, from Omroep Flevoland: https://www.omroepflevoland.nl/nieuws/161202/benzine-is-het-duurst-in-flevoland

Paulin, M., Ferguson, R. J., Jost, N., & Fallu, J.-M. (2014). Motivating millennials to engage in charitable causes through social media. *Journal of Service Management, 25*(3), 334-348.

Ramezani, S., Pizzo, B., & Deakin, E. (2018). An integrated assessment of factors affecting modal choice: towards a better understanding of the causal effects of built environment. *Transportation, 45*, 1351-1387.

RET. (n.d.). *Homepage OVmiles*. Retrieved 1 24, 2019, from OVmiles.nl: https://m.ovmiles.nl/

Siren, A., & Haustein, S. (2013). Babyboomers' mobility patterns and preferences: What are the implications for future transport. *Transport Policy, 29*, 136-144.

Sociaal Cultureel Planbureau. (2019). *Mobiliteitsonderzoek Nederland (MON)*. Retrieved March 18, 2019, from Sociaal Cultureel Planbureau: https://www.scp.nl/Onderzoek/Bronnen/Beknopte_onderzoeksbeschrijvingen/Mobiliteitsonderzoek_Nederland_MON

Sociaal Cultureel Planbureau. (2019). *Onderzoek verplaatsingen in Nederland (OViN)*. Retrieved from Sociaal Cultureel Planbureau: https://www.scp.nl/Onderzoek/Bronnen/Beknopte_onderzoeksbeschrijvingen/Onderzoek_verplaatsingen_in_Nederland_OViN

Stradling, S. G. (2004). Readiness for modal shift in Scotland. *Scottish Geographical Journal, 120*(4), 265-275.

Stryker, S. (2007). Identity Theory and Personality Theory: Mutual Relevance. *Journal of Personality, 75*(6), 1083-1102.

Van Wee, B. (2002). Land use and transport: research and policy challenges. *Journal of transport geography, 10*, 259-271.

Verlaan, J. (2018, November 12). *Helpt meer asfalt deze keer wel?* Retrieved November 13, 2018, from nrc.nl: https://www.nrc.nl/nieuws/2018/11/12/asfalt-blijft-het-wapen-tegen-de-file-a2754886

Versteegh, K. (2017, October 2017). *NRC checkt: 'Jaarlijks kosten files maximaal 3,7 miljard'*. Retrieved November 13, 2018, from NRC.nl: https://www.nrc.nl/nieuws/2017/10/30/jaarlijks-kosten-files-maximaal-37-miljard-13760546-a1579293

Watkins, A. R. (2016). Commuting Flows and Labour Market Structure: Modelling Journey to Work Behaviour in an Urban Environment. *Growth and Change, 47*(4), 612-630.

Wooldridge, J. M. (2014). *Introduction to Econometrics.* Hampshire: Cengage Learning EMEA.

# A. Appendix A: Statistics

| Variable | Variable Type | Mean | Standard deviation | Min | Max |
|---|---|---|---|---|---|
| Millenial | Dummy | 0,169194 | 0,3749234 | 0 | 1 |
| Millenial and Average density | Dummy | 0,030688 | 0,1724712 | 0 | 1 |
| Millenial and High density | Dummy | 0,080926 | 0,2727222 | 0 | 1 |
| Low Density | Dummy | 0,371908 | 0,4833147 | 0 | 1 |
| Average Density | Dummy | 0,198854 | 0,3991389 | 0 | 1 |
| High Density | Dummy | 0,429238 | 0,4949682 | 0 | 1 |
| 0-5000 Inhabitants | Dummy | 0,00138 | 0,037123 | 0 | 1 |
| 5000-10000 Inhabitants | Dummy | 0,018154 | 0,13351 | 0 | 1 |
| 10000-20000 Inhabitants | Dummy | 0,117522 | 0,3220415 | 0 | 1 |
| 20000-50000 Inhabitants | Dummy | 0,386158 | 0,4868684 | 0 | 1 |
| 50000-100000 Inhabitants | Dummy | 0,185722 | 0,3888829 | 0 | 1 |
| 100000-150000 Inhabitants | Dummy | 0,093576 | 0,2912388 | 0 | 1 |
| 150000-250000 inhabitants | Dummy | 0,106945 | 0,3090436 | 0 | 1 |
| More than 250000 Inhabitants | Dummy | 0,090542 | 0,2869574 | 0 | 1 |
| Groningen | Dummy | 0,052337 | 0,2227052 | 0 | 1 |
| Friesland | Dummy | 0,053294 | 0,2246203 | 0 | 1 |
| Drenthe | Dummy | 0,051042 | 0,220083 | 0 | 1 |
| Overijssel | Dummy | 0,065028 | 0,2465751 | 0 | 1 |
| Flevoland | Dummy | 0,045919 | 0,2093088 | 0 | 1 |
| Gelderland | Dummy | 0,104292 | 0,3056393 | 0 | 1 |
| Utrecht | Dummy | 0,067671 | 0,2511811 | 0 | 1 |
| Noord-Holland | Dummy | 0,140409 | 0,347411 | 0 | 1 |
| Zuid-Holland | Dummy | 0,175671 | 0,380541 | 0 | 1 |
| Zeeland | Dummy | 0,048118 | 0,2140154 | 0 | 1 |
| Noord-Brabant | Dummy | 0,128556 | 0,3347079 | 0 | 1 |
| Limburg | Dummy | 0,067665 | 0,2511702 | 0 | 1 |
| lnkmafstr | Continuous | 1,464381 | 1,561998 | -2,30259 | 5,991465 |
| Work | Dummy | 0,299926 | 0,4582261 | 0 | 1 |
| shopping | Dummy | 0,223134 | 0,416348 | 0 | 1 |
| Services | Dummy | 0,041489 | 0,1994178 | 0 | 1 |
| Social or Recrational | Dummy | 0,342105 | 0,4744153 | 0 | 1 |
| Education | Dummy | 0,036265 | 0,1869485 | 0 | 1 |
| Other | Dummy | 0,057082 | 0,2319987 | 0 | 1 |
| Weekend | Dummy | 0,223329 | 0,4164779 | 0 | 1 |
| Fuel price | Continuous | 1,548027 | 0,1409419 | 1,352 | 1,80192 |
| Age | Continuous | 48,00348 | 1725434 | 16 | 99 |
| Gender | Dummy | 1,501238 | 0,4999993 | 1 | 2 |
| No education | Dummy | 0,005331 | 0,072819 | 0 | 1 |
| Basic or lower education | Dummy | 0,06499 | 0,2465083 | 0 | 1 |
| Lower advanced education | Dummy | 0,245841 | 0,4305854 | 0 | 1 |
| MBO, HAVO, VWO | Dummy | 0,365039 | 0,4814419 | 0 | 1 |
| College or University | Dummy | 0,308641 | 0,4619334 | 0 | 1 |
| Other education | Dummy | 0,010158 | 0,1002734 | 0 | 1 |
| Lowest Income Group | Dummy | 0,11917 | 0,3239885 | 0 | 1 |
| Income Group 2 | Dummy | 0,212191 | 0,4088603 | 0 | 1 |
| Income Group 3 | Dummy | 0,213742 | 0,4099471 | 0 | 1 |
| Income Group 4 | Dummy | 0,184232 | 0,3876738 | 0 | 1 |
| Income Group 5 | Dummy | 0,270666 | 0,4443043 | 0 | 1 |
| No working hours | Dummy | 0,364954 | 0,481418 | 0 | 1 |
| Less than 12 working hours | Dummy | 0,041618 | 0,1997146 | 0 | 1 |
| 12 to 30 working hours | Dummy | 0,168122 | 0,3739754 | 0 | 1 |
| 30 or more working hours | Dummy | 0,425306 | 0,4943901 | 0 | 1 |
| 2005 | Dummy | 0,0949 | 0,2930767 | 0 | 1 |
| 2006 | Dummy | 0,079288 | 0,2701879 | 0 | 1 |
| 2007 | Dummy | 0,075082 | 0,2635234 | 0 | 1 |
| 2008 | Dummy | 0,062611 | 0,2422625 | 0 | 1 |
| 2009 | Dummy | 0,052995 | 0,224024 | 0 | 1 |
| 2010 | Dummy | 0,085403 | 0,2794814 | 0 | 1 |
| 2011 | Dummy | 0,081745 | 0,2739768 | 0 | 1 |
| 2012 | Dummy | 0,084499 | 0,2781353 | 0 | 1 |
| 2013 | Dummy | 0,082234 | 0,2747209 | 0 | 1 |
| 2014 | Dummy | 0,083204 | 0,276191 | 0 | 1 |
| 2015 | Dummy | 0,072501 | 0,2593163 | 0 | 1 |
| 2016 | Dummy | 0,071855 | 0,2582485 | 0 | 1 |
| 2017 | Dummy | 0,073683 | 0,2612542 | 0 | 1 |

*Appendix A Table 1. Descriptive statistics*

| Caruse | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No | 1.193 | 1.064 | 1.128 | 998 | 908 | 2.885 | 3.104 | 3.170 | 3.267 | 3.393 | 3.169 | 3.289 | 3.449 | 31.017 |
| Yes | 797 | 810 | 938 | 870 | 805 | 1.946 | 2.042 | 2.192 | 2.362 | 2.321 | 2.257 | 2.539 | 2.804 | 22.683 |
| Total | 1.990 | 1.874 | 2.066 | 1.868 | 1.713 | 4.831 | 5.146 | 5.362 | 5.629 | 5.714 | 5.426 | 5.828 | 6.253 | 53.700 |

*Appendix A Table 2. Car use of millennials per year*

| Caruse | Mean distance |
|---|---|
| No | 5,16 |
| Yes | 20,66 |

*Appendix A Table 3. Mean Trip distance*

| Correlation | milage2006 | density | gemgr | prov | lnkmafstr | doel6 | weekend | brandstofprijs | leeftijd | geslacht | education | incomeproxy | werkuren |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Milage2006** | 1 | | | | | | | | | | | | |
| **density** | 0,0529 | 1 | | | | | | | | | | | |
| **gemgr** | 0,0567 | 0,7287 | 1 | | | | | | | | | | |
| **prov** | -0,0097 | 0,1251 | 0,0171 | 1 | | | | | | | | | |
| **lnkmafstr** | 0,0476 | -0,0438 | -0,0404 | -0,0178 | 1 | | | | | | | | |
| **doel6** | -0,0229 | -0,0249 | -0,0253 | -0,0072 | -0,1401 | 1 | | | | | | | |
| **weekend** | -0,0079 | -0,0117 | -0,0117 | -0,0121 | -0,0081 | 0,2357 | 1 | | | | | | |
| **brandstofprijs** | 0,0059 | -0,159 | -0,1205 | 0,0303 | 0,004 | 0,0135 | -0,0019 | 1 | | | | | |
| **leeftijd** | -0,7388 | -0,0357 | -0,042 | 0,0102 | -0,1063 | 0,0919 | 0,0107 | -0,0083 | 1 | | | | |
| **geslacht** | 0,0401 | 0,0339 | 0,0294 | 0,0049 | -0,1558 | 0,0898 | 0,0069 | -0,001 | -0,0239 | 1 | | | |
| **education** | 0,1787 | 0,0972 | 0,0902 | -0,0171 | 0,1049 | -0,0357 | -0,0057 | -0,0167 | -0,285 | -0,0382 | 1 | | |
| **incomeproxy** | -0,093 | 0,0628 | 0,0466 | 0,0222 | 0,1928 | -0,1489 | -0,0047 | -0,0297 | 0,0032 | -0,3893 | 0,4082 | 1 | |
| **werkuren** | 0,2519 | 0,0294 | 0,028 | 0,0066 | 0,2044 | -0,2779 | -0,0175 | -0,0054 | -0,5514 | -0,1913 | 0,3275 | 0,4845 | 1 |

*Appendix A Table 4. Correlation Matrix for Model 2a*

| Correlation | millennial | density | gemgr | prov | lnkmafstr | doel6 | weekend | brandstofprijs | leeftijd | geslacht | education | incomeproxy | werkuren |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **millennial** | 1 | | | | | | | | | | | | |
| **density** | 0,0555 | 1 | | | | | | | | | | | |
| **gemgr** | 0,0868 | 0,6798 | 1 | | | | | | | | | | |
| **prov** | -0,0343 | 0,1866 | 0,0342 | 1 | | | | | | | | | |
| **lnkmafstr** | -0,0059 | -0,0941 | -0,0946 | -0,0264 | 1 | | | | | | | | |
| **doel6** | 0,0522 | -0,0175 | -0,0204 | -0,0028 | -0,0808 | 1 | | | | | | | |
| **weekend** | -0,001 | -0,0087 | -0,0103 | 0,0002 | 0,0155 | 0,1745 | 1 | | | | | | |
| **brandstofprijs** | -0,0068 | -0,1443 | -0,1094 | -0,0251 | 0,0143 | 0,0074 | -0,0058 | 1 | | | | | |
| **leeftijd** | -0,7693 | -0,061 | -0,0981 | 0,0417 | -0,0388 | -0,0026 | 0,0018 | 0,0082 | 1 | | | | |
| **geslacht** | -0,0167 | 0,004 | 0,0102 | -0,0091 | -0,1364 | 0,0694 | -0,0018 | -0,0027 | 0,0104 | 1 | | | |
| **education** | -0,0177 | 0,0818 | 0,0932 | -0,0196 | 0,0736 | -0,061 | 0,0071 | -0,0335 | -0,0821 | -0,0271 | 1 | | |
| **incomeproxy** | -0,0002 | -0,0849 | -0,1151 | 0,0172 | 0,1326 | -0,038 | 0,0048 | -0,0146 | -0,1659 | -0,0365 | 0,2324 | 1 | |
| **werkuren** | 0,0414 | 0,022 | 0,0294 | 0,0015 | 0,1904 | -0,2815 | 0,0091 | -0,0134 | -0,321 | -0,1742 | 0,3108 | 0,2911 | 1 |

*Appendix A Table 5. Correlation Matrix for Model 2b*

| Correlation Matrix | millennial | middens | highdens | earlym_dmid | earlym_dhigh | latem_dmid | latem_dhigh | gemgr | prov | lnkmafstr | doel6 | weekend | brandstofprijs | leeftijd | geslacht | education | incomeproxy | werkuren | jaar |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| millenial | 1 | | | | | | | | | | | | | | | | | | |
| middens | -0,0198 | 1 | | | | | | | | | | | | | | | | | |
| highdens | 0,0447 | -0,432 | 1 | | | | | | | | | | | | | | | | |
| earlym_dmid | 0,3513 | 0,3182 | -0,1375 | 1 | | | | | | | | | | | | | | | |
| earlym_dhigh | 0,5926 | -0,1332 | 0,3084 | -0,0424 | 1 | | | | | | | | | | | | | | |
| latem_dmid | 0,1746 | 0,1582 | -0,0683 | -0,0125 | -0,0211 | 1 | | | | | | | | | | | | | |
| latem_dhigh | 0,2659 | -0,0598 | 0,1384 | -0,019 | -0,0321 | -0,0095 | 1 | | | | | | | | | | | | |
| gemgr | 0,0588 | -0,1579 | 0,7021 | -0,0471 | 0,2506 | -0,0231 | 0,0981 | 1 | | | | | | | | | | | |
| prov | -0,006 | 0,0319 | 0,1543 | 0,0068 | 0,0294 | 0,0084 | 0,0205 | 0,0501 | 1 | | | | | | | | | | |
| lnkmafstr | -0,0102 | 0,0005 | -0,0633 | 0,0044 | -0,0251 | -0,0125 | -0,0323 | -0,0632 | -0,0242 | 1 | | | | | | | | | |
| doel6 | 0,0661 | 0,0019 | -0,0117 | 0,0079 | 0,0122 | 0,0455 | 0,0676 | -0,0176 | 0,0097 | -0,1118 | 1 | | | | | | | | |
| weekend | 0,0007 | 0,0016 | 0,0293 | -0,0018 | 0,0085 | 0,0012 | 0,0043 | 0,0155 | 0,0664 | 0,0042 | 0,1952 | 1 | | | | | | | |
| brandstofprijs | 0,1284 | -0,0139 | -0,0127 | 0,0319 | 0,0689 | 0,0415 | 0,0459 | 0,0221 | -0,0455 | 0,0071 | 0,0315 | -0,0358 | 1 | | | | | | |
| leeftijd | -0,6511 | 0,0161 | -0,0479 | -0,2202 | -0,3578 | -0,1385 | -0,2088 | -0,0676 | 0,0122 | -0,0648 | 0,0292 | 0,0036 | -0,008 | 1 | | | | | |
| geslacht | 0,0038 | 0 | 0,0122 | -0,0016 | 0,0121 | -0,0002 | -0,0006 | 0,014 | -0,0044 | -0,1479 | 0,0836 | 0,0017 | 0,037 | -0,0061 | 1 | | | | |
| education | -0,0185 | -0,0079 | 0,0846 | 0,01 | 0,0708 | -0,0683 | -0,0941 | 0,0904 | -0,0116 | 0,0892 | -0,0528 | 0,0014 | 0,0499 | -0,1576 | -0,039 | 1 | | | |
| incomeproxy | -0,053 | 0,0192 | -0,0198 | -0,0234 | -0,1003 | 0,0416 | 0,032 | -0,0359 | 0,0365 | 0,1446 | -0,0757 | 0,0043 | 0,0235 | -0,1039 | -0,1635 | 0,3036 | 1 | | |
| werkuren | -0,0268 | -0,0013 | 0,0211 | 0,0241 | 0,025 | -0,067 | -0,1056 | 0,026 | 0,0078 | 0,1975 | -0,2932 | -0,0071 | -0,0691 | -0,3951 | -0,1927 | 0,3083 | 0,3324 | 1 | |
| jaar | 0,1819 | -0,0333 | 0,0663 | 0,0253 | 0,1045 | 0,0782 | 0,1302 | 0,0781 | 0,012 | 0,0041 | 0,0443 | 0,0124 | 0,6104 | 0,0174 | 0,0407 | 0,0712 | 0,0946 | -0,0821 | 1 |

*Appendix A Table 6. Correlation Matrix for Model 4*

# B. Appendix B: Results

| Prediction\car use | No | Yes | Total |
|---|---|---|---|
| **No** | 8.236 | 2.582 | 10.818 |
| **Yes** | 2.950 | 11.397 | 14.347 |
| **Total** | **11.186** | **13.979** | |

*Appendix B Table 1. Predictions and observarions of car use in Model 2a*

| | |
|---|---|
| Car use correctly predicted | 81,53% |
| Non-car use | 73,63% |
| Total correctly predicted | 78,02% |

*Appendix B Table 2. PCP values for Model 2a*

| Prediction\car use | No | Yes | Total |
|---|---|---|---|
| **No** | 8.827 | 2.680 | 11.507 |
| **Yes** | 2.606 | 8.693 | 11.299 |
| **Total** | **11.433** | **11.373** | |

*Appendix B Table 3. Predictions and observarions of car use in Model 2b*

| | |
|---|---|
| Car use correctly predicted | 76,44% |
| Non-car use | 77,21% |
| Total correctly predicted | 76,82% |

*Appendix B Table 4. PCP values for Model 2b*